Application of Bioinformatics Tools and Software in Plant Breeding

Manoj Kumar[1]*, Sandhya[1], K.M. Sharma[2] and Pawan Kumar[3]

[1]Assistant Professor, Department of Genetics and Plant Breeding, Agricultural Research Station Ummedganj, Agriculture University, Kota, Rajasthan

[2]Associate Professor, Department of Agronomy, Agricultural Research Station Ummedganj, Agriculture University, Kota, Rajasthan

[3]Scientist, Central Institute for Arid Horticulture, Bikaner, Rajasthan

SUMMARY

Bioinformatics is the field of science in which biology, computer science, and information technology merge to form a single discipline. The ultimate goal of the field is to enable the discovery of new biological insights as well as to create a global perspective from which unifying principles in biology can be discerned. The main components of bioinformatics are the collection and analysis of database and the development of software tools and algorithm as a tool for interpretation of biological data including nucleotide and amino acid sequences, protein domains and structure as well as expression patterns from various organisms. In the context of plant breeding, bioinformatics can empower genetic and genomic selection to determine the optimal combination of genotypes that will produce a desired phenotype and help expedite the isolation of these new varieties. Bioinformatics is also instrumental in collecting and processing plant phenotypes, which facilitates plant breeding.

INTRODUCTION

Over the past decades, the term 'bioinformatics' has become a buzzword in all areas of research in biological science. With the continuous development and advancement in molecular biology, the explosive growth of biological information required a more organized, computerized system to collect, store, manage, and analyse the vast amount of biological data generated in the experiments from all fields (Gomez *et al.,* 2018) Bioinformatics was invented by Paulien Hogeweg and Ben Hesper in 1970 as "the study of informatic processes in biotic systems". Bioinformatics is a new and emerging computer based scientific field which applies mathematics, biology, and computer science to form into a single discipline for the analyses and interpretation of genomics and proteomics data (Tiwari *et al*., 2020). In short, the main components of bioinformatics are (a) the collection and analysis of database and (b) the development of software tools and algorithm as a tool for interpretation of biological data (Zhang SY and Liu SL, 2013). Bioinformatics played a crucial role in many areas of biology as its applications provide various types of data, including nucleotide and amino acid sequences, protein domains and structure as well as expression patterns from various organisms.

Importances of Bioinformatics

The introduction of bioinformatics and computational biology into the area of plant biology is drastically accelerating scientific invention in life science. In order to study how normal cellular activities are altered in different disease states, the biological data must be combined to form a comprehensive picture of these activities. Therefore, the field of bioinformatics has evolved such that the most pressing task now involves the analysis and interpretation of various types of data. Bioinformatics apply the principles of information sciences and technologies to complex life science data (Ojo and Maxwell, 2010). The primary goal of bioinformatics is to increase the understanding of biological processes. What sets it apart from other approaches, however, is its focus on developing and applying computationally intensive techniques to achieve this goal. The analysis of genetic and DNA sequences such as nucleic acid sequences, protein structure/function relationships, genome organization, regulation of gene expression, interaction of proteins and mechanisms of physiological functions, can all benefit from a bioinformatics approach. Nucleic acid and protein sequence data from many different species and from population samplings provides a foundation for studies leading to new understandings of evolution and the natural history of life (Tramontano, 2009). Major research efforts in the field include sequence alignment, gene finding, genome assembly, drug design, drug discovery, protein structure alignment, protein structure prediction, prediction of gene expression and protein–protein interactions, genome-wide association studies, and the modelling of evolution.

Bioinformatics Tools & software

Bioinformatics tools refer to software programs specifically designed for extracting meaningful information from the extensive molecular biology and biological databases. These tools are utilized to conduct sequence or structural analyses. Within this domain, there exist both standardized tools and customized products tailored to meet the specific needs of individual projects. Data-mining software that retrieve data from genomic sequence databases and also visualization tools to analyze and retrieve information from proteomic databases. These can be classified as homology and similarity tools, protein functional analysis tools, sequence analysis tools and miscellaneous tools.

Homology and Similarity Tools:

Homologous sequences are sequences that are related by divergence from a common ancestor. Thus, the degree of similarity between two sequences can be measured while their homology is a case of being either true of false. This set of tools can be used to identify similarities between novel query sequences of unknown structure and function and database sequences whose structure and function have been elucidated.

Protein Function Analysis:

This group of programs allow you to compare your protein sequence to the secondary (or derived) protein databases that contain information on motifs, signatures and protein domains. Highly significant hits against these different pattern databases allow you to approximate the biochemical function of your query protein.

Structural Analysis:

This set of tools allow you to compare structures with the known structure databases. The function of a protein is more directly a consequence of its structure rather than its sequence with structural homologs tending to share functions. The determination of a protein's 2D/3D structure is crucial in the study of its function.

Sequence Analysis:

This set of tools allows you to carry out further, more detailed analysis on your query sequence including evolutionary analysis, identification of mutations, hydropathy regions, CpG islands and compositional biases. The identification of these and other biological properties are all clues that aid the search to elucidate the specific function of your sequence.

Examples of Bioinformatic Tools:

BLAST:

BLAST (Basic Local Alignment Search Tool) comes under the category of homology and similarity tools. It is a set of search programs designed for the Windows platform and is used to perform fast similarity searches regardless of whether the query is for protein or DNA. Comparison of nucleotide sequences in a database can be performed. Also a protein database can be searched to find a match against the queried protein sequence. NCBI has also introduced the new queuing system to BLAST (Q BLAST) that allows users to retrieve results at their convenience and format their results multiple times with different formatting options. Depending on the type of sequences to compare, there are different programs:

blastp :-Compares an amino acid query sequence against a protein sequence database.

Blastn:- Compares a nucleotide query sequence against a nucleotide sequence database.

blastx :-Compares a nucleotide query sequence translated in all reading frames against a protein sequence database.

tblastn :-Compares a protein query sequence against a nucleotide sequence database dynamically translated in all reading frames.

tblastx :-Compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

FASTA:

FAST homology search All sequences .An alignment program for protein sequences created by Pearsin and Lipman in 1988. The program is one of the many heuristic algorithms proposed to speed up sequence comparison. The basic idea is to add a fast pre-screen step to locate the highly matching segments between two

sequences, and then extend these matching segments to local alignments using more rigorous algorithms such as Smith-Waterman.

EMBOSS:
EMBOSS (European Molecular Biology Open Software Suite) is a software-analysis package. It can work with data in a range of formats and also retrieve sequence data transparently from the Web. Extensive libraries are also provided with this package, allowing other scientists to release their software as open source. It provides a set of sequence-analysis programs, and also supports all UNIX platforms.

Clustalw:
It is a fully automated sequence alignment tool for DNA and protein sequences. It returns the best match over a total length of input sequences, be it a protein or a nucleic acid.

RasMol:
It is a powerful research tool to display the structure of DNA, proteins, and smaller molecules. Protein Explorer, a derivative of RasMol, is an easier to use program.

PROSPECT:
PROSPECT (Protein Structure Prediction and Evaluation Computer Toolkit) is a protein-structure prediction system that employs a computational technique called protein threading to construct a protein's 3-D model.

Pattern Hunter
Pattern Hunter, based on Java, can identify all approximate repeats in a complete genome in a short time using little memory on a desktop computer. Its features are its advanced patented algorithm and data structures, and the java language used to create it. The Java language version of Pattern Hunter is just 40 KB, only 1% the size of Blast, while offering a large portion of its functionality.

COPIA
COPIA (COnsensus Pattern Identification and Analysis) is a protein structure analysis tool for discovering motifs (conserved regions) in a family of protein sequences. Such motifs can be then used to determine membership to the family for new protein sequences, predict secondary and tertiary structure and function of proteins and study evolution history of the sequences.

Bioinformatics for Plant Breeding
Bioinformatics plays a crucial role in advancing plant breeding by leveraging computational tools and techniques to analyze, interpret, and manage biological data. The goal of plant genomics is to understand the genetic and molecular basis of all biological processes in plants. This understanding is fundamental to allow efficient exploitation of plants as biological resources in the development of new cultivars with improved quality and reduced economic and environmental costs

Crop Improvement
Comparative genetics of the plant genomes has shown that the organization of their genes is more conserved over evolutionary time than was previously believed. These results suggest that information obtained from the model crop systems can be used to suggest improvements to other food crops. At hand the complete genomes of Arabidopsis thaliana (water cress) and Oryza sativa (rice) are available.

Insect Resistance
Bacillus thuringiensis genes control a number of serious pests that have been successfully transferred to cotton, maize and potatoes. This new aptitude of the plants to resist insect attack means that the number of insecticides being used can be reduced and hence the nutritional quality of the crops is increased.

Improve Nutritional Quality
The recent development in agriculture is transferring genes into rice to increase levels of Vitamin A, iron and other micronutrients. This work might have a deep impact in reducing occurrences of blindness and

anaemia which is caused by deficiencies in Vitamin A and iron respectively. Another one is, to insert a yeast gene into the tomato, and the result is a plant whose fruit stays longer on the vine and has an extended shelf life.

Development of Drought Resistance Varieties

Development has been made in developing cereal varieties that have a greater tolerance for soil alkalinity, free aluminium and iron toxicities. These types of varieties will allow agriculture to succeed in poorer soil areas, thus adding more land to the global production base. The development work is in progress on the production of crop varieties capable of tolerating reduced water conditions.

CONCLUSION

Bioinformatics has evolved into a full-fledged multidisciplinary subject that integrates developments in information and computer technology as applied to Biotechnology and Biological Sciences. The role of bioinformatics tools in plant breeding is pivotal and transformative. These tools have revolutionized the traditional methods of plant breeding by offering efficient ways to manage, analyze, and interpret vast amounts of genomic and molecular data. Through advanced algorithms and computational approaches, bioinformatics tools facilitate the identification of desirable traits, marker-assisted selection, and the understanding of complex genetic interactions.

REFERENCES

Gomez-Casati DF, Busi MV, Barchiesi J, Peralta DA, Hedin N, Bhadauria V (2018) Applications of bioinformatics to plant biotechnology. Curr Issues Mol Biol 27:89–104. https://doi.org/10.21775/cimb.027.089

Ojo, O.O. and Maxwell, O. (2010). Incorporating bioinformatics into biological science education in Nigeria: prospects and challenges, Infection, Genetics and Evolution: 10.1016/j.meegid, 11.015

Tiwari A, Singh P, Kumawat S (2020) Applications of bioinformatics in plant breeding system. Int J Curr Microbial App Sci. 11:2825–2831

Tramontano, A.L. (2009). Bioinformatics. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester.DOI: 10.1002/97804700 15902.a0001900.pub2.

Zhang SY, Liu SL (2013) Bioinformatics. In: Maloy S, Hughes K (eds) Brenner's Encyclopedia of Genetics, 2nd edn. Academic Press, London. https://doi.org/10.1016/B978-0-12-374984-0.00155-8