

# BIOINFORMATICS: AN OVERVIEW AND ITS APPLICATION

[Article ID: SIMM0236]

**Thaarani M<sup>1</sup>, Mohamed Yaseen S.K<sup>2</sup>, Katta Sravan Naga Parimala Kumar<sup>2</sup>**

<sup>\*1</sup> Don Bosco College of Agriculture, Sagayathottam.

<sup>2</sup> G.B. Pant University of Agriculture and Technology, Pant Nagar.

## INTRODUCTION

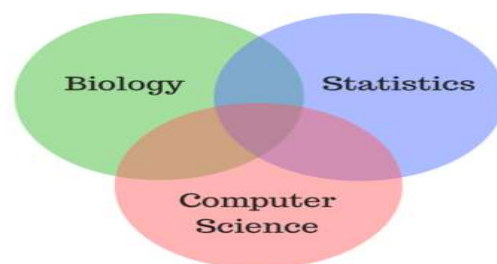
Bioinformatics is a hybrid science that connects biological data with information storage, dissemination, and analysis techniques to serve numerous areas of scientific research. Bioinformatics is fed by high throughput data – generating experiments, including genome sequence determination and measurement of gene expression pattern. It plays a significant role in providing information about the genes present in the genome of species. These tools have also made it possible to predict the function of different genes and factors affecting those genes Luscombe *et al* (2001). The information provided about the genes by the tools makes the scientists to produce enhanced species of plants which have drought, herbicide, and pesticide resistance. The development and implementation of tools that enable efficient access and management of various types of information, the development of new algorithms and statistics with which to assess relationships among members of large data sets, such as methods to locate a gene within a sequence, predict protein structure or function, and

cluster protein sequences into families of related in data analysis and interpretation.

## THE DATA OF BIOINFORMATICS:

The classic data of bioinformatics include DNA sequence of gene or genome, amino acids, sequence of proteins, nucleic acid and protein complex Baxevanis *et al*, (2020). The scientific field associated with measuring biological molecules in a high throughput way is called as “omics”. Transcriptomics (pattern of RNA synthesis from DNA), proteomics (distribution of protein in cells), interactomics (pattern of protein – protein and protein – nucleic acid interaction). Bioinformatics has been driven by the great acceleration in data – generation process in biology.

## Bioinformatics



## GOALS OF BIOINFORMATICS:

The development of efficient algorithms for measuring sequence similarities is an important goal of bioinformatics. The Major approach in bioinformatics, Needleman – Wunsch algorithm, which is based on dynamics programming, guarantees finding the optimal alignment of pair of sequence.

Although the Needleman-Wunsch algorithm is effective, it is too slow for probing a large sequence database. Therefore, much attention has been given to finding fast information-retrieval algorithms that can deal with the vast amounts of data in the archives. An example is the program BLAST (Basic Local Alignment Search Tool). A development of BLAST, known as position-specific iterated- (or PSI-) BLAST, makes

use of patterns of conservation in related sequences and combines the high speed of BLAST with very high sensitivity to find related sequences.

Another goal of bioinformatics is the extension of experimental data by predictions. A fundamental goal of computational biology is the prediction of protein structure from an amino acid sequence. The spontaneous folding of proteins shows that this should be possible. Progress in the development of methods to predict protein folding is measured by biennial Critical Assessment of Structure Prediction (CASP) programs, which involve blind tests of structure prediction methods.

Bioinformatics is also used to predict interactions between proteins, given individual structures of the partners. This is known as the “docking problem.” Protein-protein complexes show good complementary in surface shape and polarity and are stabilized largely by weak interactions, such as hydrophobic surface, hydrogen bonds, and Van Der Waals force. Computer programs simulate these interactions to predict the optimal spatial relationship between binding partners.

#### BRANCH OF BIOINFORMATICS:

Bioinformatics is broadly divided into two groups' viz., Animal bioinformatics and Plant bioinformatics Hagen et al, (2020).

1. **ANIMAL BIOINFORMATICS:** It deals with computer aided study of genomics in various animal species. It includes study of gene mapping, gene sequencing, animal breeds, and animal genetic resources. It further divided into bioinformatics of mammals, reptiles, insect, bird, fish, etc.
2. **PLANT BIOINFORMATICS:** It deals with computer aided study of plant species. It includes gene mapping, gene sequencing, and plant genetic resources and so on. It is further divided into agricultural,

horticultural, medicinal plants and forest bioinformatics.

#### STRUCTURAL GENOMICS:

The increase in the number of new protein sequences arising from genomics and proteomics highlights directly the need for methods to rapidly and reliably determine the molecular and cellular functions of these proteins. One such approach, structural genomics, aims to decode the protein folds, thereby providing three-dimensional portraits for all proteins in a living organism and to infer molecular functions of the proteins. These new structures have revealed many unexpected functional and evolutionary relationships that were hidden at the sequence level. Structural genome is the study that attempts to sequence the whole genome and mapping the genome.

#### PLANT GENOMICS:

The comparison of genome sequences of rice and *Arabidopsis* revealed plenty of useful information for plant genomics because of their extensive but complex designs of synthesis. *Arabidopsis thaliana* has become a well-known model plant for most of the researchers. In spite of the fact that it is a non-commercial plant, it is preferred because of its reproduction, development and reaction to stress and disease in the same way as many crop plants. It has a small genome which does not have the repeated, less-informative DNA sequences that hinder genome analysis performance. Its advantages are that it has large genetic and physical maps of chromosomes, a fast life cycle (around 6 weeks from seed germination to grown seed), productive seed manufacture and simple cultivation in limited space, a huge number of mutant lines and genomic resources and multinational research society of academic, government and industry laboratories.

#### IMPACT OF GENOME SEQUENCING IN AGRICULTURE:

The term genome can be applied particularly to the whole genetic material of an organism

including the full set of nuclear DNA (i.e., nuclear genome) and also to the genetic information stored within organelles, which have their own DNA - the 'mitochondrial genome' or the 'chloroplast genome'. Moreover, the genome can contain non chromosomal genetic elements like viruses, plasmids or transposable elements Gauthier *et al* (2019). Most biological units which are more complex than a virus, have extra genetic material besides that which has in their chromosomes. Therefore 'genome' describes all of the genes and information on non-coding DNA that have the potential to be present.

#### SOFTWARES USED IN BIOINFORMATICS:

The first software that was used in the field of bioinformatics is COMPROTEIN. The commonly used software's follows,

1. *BLAST*
2. *BioPerl*
3. *Clustal*
4. *SAMtools*
5. *Bioconductors*
6. *Biopythons*
7. *BioJava*
8. *Cytoscape*
9. *GROMACS*

#### APPLICATIONS OF BIOINFORMATICS:

1. To understand the genes and their expressions.
2. Provides information about plant's proteins.
3. Many plants have been made insect resistant with the study of bioinformatics ex. Corn and cotton
4. Desired genes can be sorted out for further studies
5. Increased nutritional value of the plant.

6. Its aides in pedigree analysis and provides information about the parentage.
7. Improvement of agronomical characters.
8. Human genome field will profound the effects on the field of medicine.

#### CONCLUSION:

Throughout the past several decades, an extensive amount of scientific data has been generated, particularly within the fields of molecular biology and genomics. It has been essential in determining how the data acquired from biological experiments are used and the studies links so many majors of science and opens up a wider pathway for new innovations. The genomic analysis of eukaryotic species, which is typically confined to information present on chromosomal DNA, has expanded beyond studying the human genome for clinical applications to also sequencing the genomes of both plants and animals for agricultural purposes

#### REFERENCE:

1. Baxevanis, A.D., Bader, G.D. and Wishart, D.S. eds., 2020. *Bioinformatics*. John Wiley & Sons.
2. Can, T., 2014. Introduction to bioinformatics. *miRNomics: MicroRNA biology and computational analysis*, pp.51-71.
3. Hagen, J.B., 2000. The origins of bioinformatics. *Nature Reviews Genetics*, 1(3), pp.231-236.
4. Luscombe, N.M., Greenbaum, D. and Gerstein, M., 2001. What is bioinformatics? A proposed definition and overview of the field. *Methods of information in medicine*, 40(04), pp.346-358.
5. Gauthier, J., Vincent, A.T., Charette, S.J. and Derome, N., 2019. A brief history of bioinformatics. *Briefings in bioinformatics*, 20(6), pp.1981-1996.