

Relevance of Bioinformatics in the Era of Omics Driven Research

Dinesh Yadav*

Department of Biotechnology, D.D.U Gorakhpur University, Gorakhpur, India

*Corresponding author: Dinesh Yadav, Department of Biotechnology, D.D.U Gorakhpur University, Gorakhpur (U.P.) 273 009, India, Tel: 09411793038; E-mail: dinesh_yad@rediffmail.com

Rec Date: April 20, 2015, Acc Date: April 21, 2015, Pub Date: April 23, 2015

Copyright: © 2015 Yadav D. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Introduction

Bioinformatics, a well established multidisciplinary branch of sciences, often known as computational biology, is gaining immense importance in the era of omics marked by generation of huge biological data constantly. The advances in molecular biology led to the genesis of bioinformatics for the sole purpose of storing, retrieving and analyzing nucleotide and protein sequences to get an insight into life processes. The bioinformatics primarily deals with data-curation, developing tools, assisting data interpretation and analysis using web-based resources in a biologically meaningful manner. The computational expertise along with a good understating of biological processes are associated with developing appropriate algorithms for sequence comparisons, phylogenetic/evolutionary tree construction, specific pattern recognitions, sequence-structure-function elucidation, annotating sequences, deciphering metabolic pathways, gene regulation and expression, drug designing etc.

The innovations in sequencing technologies over the years, conceptualized as four generations of sequencing namely first, second, third and fourth facilitated by conventional capillary electrophoresis based Sanger's method, pyrosequencing (454 sequencing), sequencing by synthesis and the latest nanopore exonuclease sequencing method respectively has drastically expanded the biological data. The Next Generation Sequencing techniques encompassing second, third and fourth generations sequencing has led to deciphering of genome sequences at unprecedented rate. The outcome of these genome sequencing projects resulted in a huge accumulation of sequences, which needs proper storage, retrieval, processing, annotation and validation for its application in diverse fields of research and thus tools of bioinformatics becomes more relevant.

The science of 'omics' reflects characterization and quantification of pools of diverse biological molecules associated with the structure, function, and dynamics of organisms. The omics based approaches in diverse fields led to the emergence of plethora of new terminologies for specific areas of research like proteomics, metabolomics, cytomics, metagenomics, lipidomics, transcriptomics and many more. The new approach for real time understanding of biology is system biology which combines the information of different field to simulate and analyze the networks, pathways, the spatial and temporal relations that exist in biological systems. The extensive data being generated by experimentation related with diverse fields of omics is being successfully managed by bioinformatics experts through the development of appropriate user-friendly biological databases with provision for open access to researchers globally. The data management and data mining are two important bottlenecks for omics based research and this demands bioinformatics intervention. Bioinformatics aims to establish standard formats by using algorithms based on mathematical and statistical models and developing efficient methods for storing, retrieving and sharing high-throughput data in

the era of omics. The data analysis, molecular modeling, predictions, simulation, phylogenetic analysis, sequence comparison are in the purview of bioinformatics and accordingly there has been development of appropriate in-silico tools.

The 'Agri-genomics' driven crop improvement needs direct intervention of bioinformatics in developing appropriate tools for cost-effective and efficient breeding strategies to be adopted by molecular breeders. Advances in Bioinformatics has led to the development of in-silico tools and web-based resources for identifying genome-wide molecular markers, harnessing genomic technologies for pathogen characterization, diseases diagnosis, getting an insight into the complexity of gene expression and regulation and extrapolating the genomics information of model crops to orphan crops by comparative genomics approach. The crop specific databases are being developed owing to the availability of whole genome sequences of important crops like rice, sorghum, maize, soybean, potato, Brassica rapa, pigeon pea, barley, watermelon, chickpea, melon, wheat and many more for providing enriched genomic resources for crop improvement. For efficient processing and utilization of biological data, bioinformaticians are constantly developing newer web tools like GBrowse genome viewer, Ensembl system to mention a few.

In recent years, sequencing of plant genomes has led to genome wide identification and characterizations of diverse type of plant specific Transcription Factor (TFs) gene families providing key insights into their structural and functional diversity. TFs are pivotal in the regulation of plant development, reproduction, intercellular signaling, response to environment, cell cycle and metabolism. Till date, more than 100 different DNA binding domains have been recognized and based on their DNA binding domain, TFs are classified into different families, some of which are found in all eukaryotes whilst others are kingdom specific. A vast repertoire of transcription factors have been reported in several crop plants, which has subsequently led to the compilation of a large number of crop specific transcription factor databases. The assessments of these TFs using bioinformatics tools are constantly performed by plant biotechnologists to reveal its potential application in transgenic technology.

Drug discovery is another important priority area of research, which has gained momentum with the advent of science of omics. The availability of 3D X-ray or NMR structures of biomolecules supported by docking tools and computational tools has revolutionized drug discovery. The key steps of drug discovery namely target identification, lead identification, lead optimization and pre-clinical tests needs bioinformatics intervention. The Computer-Aided Drug Designing (CADD) seems to be promising in drug discovery processes by substantially reducing the overall cost and time required. The in-silico tools like molecular docking, virtual high throughput screening, QSAR (Quantitative-Structure Activity Relationship), Pharmacophore mapping and Fragment based screening are vital for drug discovery processes. The growing number of microbial genome sequences gives a

possible insight into the pathogenicity and helps in understanding the molecular basis of diseases with the aid of tools of bioinformatics. The bioinformatics assisted web based tools and databases have substantially influenced cancer genomic studies globally.

The Microarray or DNA Chip technology and Real Time Quantitative PCR, which are two major developments in the biotechnological fields extensively used for comprehensive gene expression studies, needs direct intervention of bioinformatics. The microarray is amalgamation of wet-lab experimentation, statistical analysis and bioinformatics assisted data analysis. It basically involves fabrication of chips, sample preparation and labeling, hybridization, washing and image acquisition. The appropriate selection of sequence databases for microarray, designing of oligonucleotide probes, image processing, normalization, analyzing microarray variability, accessing differentially expressed genes, principal component analysis and multidimensional scaling associated with genes, tissues and treatments, experimental design, data standard, storage and sharing are efficiently performed with the aid of bioinformatics. The Real time PCR or qPCR is considered to be one of the most sensitive and reliable quantitative method for gene expression extensively applied for microarray verification, gene copy number determination, quantifying pathogens, in drug therapy studies and also in cancer biology research. Bioinformatics based software/tools are routinely being used for qPCR data processing and analysis.

The relevance of bioinformatics is also being realized in enzyme research especially for in-silico based 3D structural predictions and catalytic site determination. The microbial enzymes are extensively used in diverse industries with more than 500 products and several innovations have been attempted to enhance the enzyme activity, production of enzymes and manipulating the enzymes for desired features. Designing and engineering enzyme activity, enzyme selectivity and enzyme stability along with metagenome screening, microbial genome mining, and exploring diversity of extremophiles have been successfully managed by tools of bioinformatics. The experimental tools for manipulation of enzymes like gene cloning, sequencing, directed evolution methods and gene synthesis are strongly supported by the bioinformatics, molecular modeling and de novo design.

The science of omics have the generic goal to unravel the mystery of living system by identifying the components in totality that make the living system and to analyze the interaction among these components which might lead to understand the functioning of the system. Bioinformatics has truly established its unique position by virtue of developing user-friendly web based resources, analysis tools and softwares to expedite the pace of omics driven research.