Research Article

STATUS OF BIOINFORMATICS APPLICATIONS IN AGRICULTURAL RESEARCH AND EDUCATION IN INDIA

BALAKRISHNAN M.*, SRINIVASA RAO CH. AND SOAM S.K.

ICAR-National Academy of Agricultural Research Management, Hyderabad, 500030, India
*Corresponding Author: Email - mbkrishnan72@gmail.com

Abstract: Agriculture has been, and continues to be, the backbone of national economy for many countries like India. The growth of various innovative technologies including improved varieties, agro-techniques and precision agriculture forms an important component of the strategies for improved production and productivity in crops and animals. The bioinformatics tools are playing important role in providing the information about the genes present in the genome of these species. These tools also have made it likely to predict the function of different genes and factors affecting these genes. Agricultural genomics leads to the global understanding of plant/animal and pathogen biology, and its application would be beneficial for agriculture approaches in the era of Next-Generation Sequencing for its influence in genomics, transcriptomic, and metagenomics efforts, describing their role in agriculture and allied sectors.

Keywords: *Agriculture, Crops, Animals bioinformatics, Tools, Genomics, proteomics, metagenomics, Animals, Fisheries and Databases*

Introduction

Agriculture has been, and continues to be, the backbone of national economy for many countries like India. The development of various innovative technologies including improved varieties, agro-techniques and precision agriculture forms an important component of the strategies for improved production and productivity in any crop. Most of the crop improvement strategies are based on conventional tools for creation of variability or exploitation of existing variability present in the genome. But as the information about various metabolic pathways and other processes taking place in the organisms accumulate, several new challenges arise and this has necessitated the use of cutting-edge technologies like biotechnology. With the advent of new tools and techniques and the improvement in the existing technologies, there is a quantum jump in the volume of data generated from the molecular and biotechnological experiments. The handling of such a voluminous data needs very efficient tools. The biological studies have become increasingly cross-disciplinary, involving a large number of disciplines like biology, physics, chemistry, mathematics, computer applications, *etc*. The ability to make global measurements and the need to integrate this information from different levels is the need of today. All these factors necessitated the use of some mechanism to effectively store, retrieve and manage the data in an effective manner leading to the development of Bioinformatics [1]. In a very broad term, Bioinformatics refers to the application of informatics involving computers, mathematics, statistics and related disciplines to generate and store biological data and to effectively organize, analyze and to interpret the information generated, on a large scale. It encompasses structural biology, genomics, expression studies and a large number of related subjects [2]. The "Omics" research in crop plants has added a novel dimension to the application of bioinformatics tools in crop research.

Importance of Bioinformatics in Agriculture

The ultimate challenge facing the molecular biology community today is to make sense of the wealth of data that has been produced by the genome sequencing projects. Today, biologists work in association with scientists from a broad spectrum of disciplines to unravel how complex biological systems work.

The genome sequencing of a number of organisms has led to the discovery of many fascinating things. Today, the world feels the need of this discipline to save resources and time. A number of applications of bioinformatics in agriculture in view of functional genomics, data mining techniques, genome-wide association studies, high-performance computing facilities in agriculture and various bioinformatics tools/databases important for breeders, biotechnologists and pathologists.

Applications of Bioinformatics in Agriculture and Allied Sectors
Methods/ Tools of bioinformatics (protocols)
Field Crops

Field crops of maize, wheat, rice, sorghum, barley, sugarcane, soybean, *etc*. are primary source of food and feed. Field crops are also a rich source of cellulosic biomass and carbohydrates for ethanol production. One of the major challenges facing agriculture today is improving the productivity of field crops in an environmentally sustainable manner. Annual climate variation causes temperature extremes, floods, and droughts which all exacerbate the vulnerability of field crops to pests and diseases. Conventional plant breeding has evolved into a molecular breeding and these modern breeding methods have enhanced the pace of crop improvement work. Plant breeders now use molecular and genetic techniques to selectively identify phenotypes and genotypes that are associated with traits of interest. Such functional genomics studies help plant breeders efficiently utilize the germplasm. Cutting edge molecular tools are now available in economically important crops as well as model plant systems. The large volume of data generated in field crops like rice, wheat, maize, pulses, Solanaceous crops, Sugar crops like sugarcane and sugar beet, oilseed crops like Brassica, have led to the use of several tools of bioinformatics in these crops. The Genome Sequencing Projects in various crops like rice, Solanaceous crops, Sorghum, chickpea, pigeon pea, sugarcane and other crops have generated large data which have invariably contributed to the drastic developments in bioinformatics in the field crops. The preliminary tools have been the various databases like nucleic acid databases and protein databases, which are freely available online (GenBank, DDBJ, EMBL, Swiss-Prot, to name a few).

Crop-specific databases evolved gradually with the formation specific to a group of crops (AgBase, Gramene, GrainGenes), or to specific crops (Oryzabase, Maize db, Sorghum db *etc.*) [3]. Apart from the very basic collection of nucleotide sequences, amino acid sequences, EST sequences *etc.*, several databases have been developed providing information of Single Nucleotide Polymorphism (SNP), transposable elements, small RNAs, micro RNAs, databases for comparative genomics (eg., PlantGDB) and other applications [3,4]. Several other databases have been developed for genomic data, on a range of platforms and to suite a variety of different purposes. For sugarcane, a highly complex and polyploidy crop, the SUCEST FUN database was developed from the large collection of sugarcane ESTs developed in a collaborative project involving a consortium of Brazilian researchers from approx. 238000 ESTs developed from 26 diverse libraries. SymGRASS is a database of sugarcane orthologous genes involved in vesicular arbuscular mycorrhiza and root nodule symbiosis [6]. Several analytical tools have also been developed for different crops. These include tools for diversity analyses like PoPGene, Arlequin [7], molecular marker identification and related tools, tools for linkage/QTL mapping (JoinMap, MapMaker, MstMap, QTL IciMapping [8], PERGOLA for polyploids [9], marker trait association studies (TASSEL), comparative genomics studies using different genomes, small RNA and miRNA identification. Next Generation Sequencing techniques have resulted in sequence alignment and annotation tools for different applications like RNA – Seq. Different softwares for analyses of different components have resulted in easier and quicker handling of BIG data with more efficient interpretation. Genome Wide Association (GWAS) like TASSEL, PLINK, PERGOLA, NAM, treeWAS, GWASTools, GAPIT and others are also being widely used by researchers in various crops. Expression studies using different techniques has lately assumed great significance in crop plants. Differential expression analyses using microarrays and other advanced techniques have developed a large volume of data. Bioinformatics tools for organizing, retrieving and utilizing this data to understand the functional aspects of different genes/factors have been developed. It facilitates the use of commonalities in plant biology for a comparative approach to functional genomics, through use of large-scale expression profiling data sets (http://www.plexdb.org). AraPLEX, CottonPLEX, SugarPLEX, TomPLEX and many other expression databases are available here. Apart from the use of these tools in agriculture, for understanding the population diversity, for expression studies, understanding the functions of genes, for comparative studies among different crops *etc.*, with an ultimate aim to improve the production and productivity, these bioinformatics tools have been of immense use in human medicine, diagnostics and therapeutics. The most important among them are the expression studies, protein modelling and interaction tools *etc.*, which have been effectively utilized in drug discovery. The tools for protein structural analyses, protein-protein interaction studies, helps in understanding the binding sets, protein folding and other structural peculiarities thereby facilitating development of new drugs or suitably modifying the existing ones.

DNA based signature of plant variety, Example-Basmati rice
Difficulty in differentiating genuine traditional basmati from pretenders and the significant price difference between them has led fraudulent traders to adulterate traditional basmati. The kit uses a PCR amplification technique based on Simple Sequence Repeats (SSR) that provides the single most discriminating assay for Basmati genotyping.

Plantation crops
The Microbial Information Systems provides the role of microbes in plantation crops and research related data on bio- inoculants, bio-control agents, decomposers, edible mushrooms and photo-gallery. Information on microorganisms obtained from PHYLOBASE. The vegetable oil database provides information for nutritional, health, value addition and quality control using the interface. Information is available on coconut, corn, olive, palm, palm kernel, peanut, safflower, sesame, soybean and sunflower oil. The DECOMPOSERS suite provides information on cellulose and lignin degraders[10, 11]. Agriculture has been, and continues to be, the backbone of national economy for many countries like India. The development of various innovative technologies including

improved varieties, agro-techniques and precision agriculture forms an important component of the strategies for improved production and productivity in any crop. Most of the crop improvement strategies are based on conventional tools for creation of variability or exploitation of existing variability present in the genome. But as the information about various metabolic pathways and other processes taking place in the organisms accumulate, several new challenges arise and this has necessitated the use of cutting-edge technologies like biotechnology. With the advent of new tools and techniques and the improvement in the existing technologies, there is a quantum jump in the volume of data generated from the molecular and biotechnological experiments. The handling of such a voluminous data needs very efficient tools. The biological studies have become increasingly cross-disciplinary, involving a large number of disciplines like biology, physics, chemistry, mathematics, computer applications, *etc*. The ability to make global measurements and the need to integrate this information from different levels is the need of today. All these factors necessitated the use of some mechanism to effectively store, retrieve and manage the data in an effective manner leading to the development of Bioinformatics [12]. In a very broad term, Bioinformatics refers to the application of informatics involving computers, mathematics, statistics and related disciplines to generate and store biological data and to effectively organize, analyze and to interpret the information generated, on a large scale. It encompasses structural biology, genomics, expression studies and a large number of related subjects [13]. The "Omics" research in crop plants has added a novel dimension to the application of bioinformatics tools in crop research.Fruits: Today the Modern technological development like NGS provides the possibility of cost-efficient whole genome sequencing. Notable developments pertaining to genome sequencing of fruit crops are highlighted and. The information of the sequences has to be used for analyze the functional genomics and biotic and abiotic stress [14].

Livestock
Livestock –meat: Livestock genomics has gone through a paradigm shift since the advent of genome sequencing that includes Genome-Wide Association Studies (GWAS), Whole Genome Predictions (WGP) and Genomic Selection (GS).The Animal QTL database (http://www.animalgenome.org/QTLdb/) reported several thousands of QTLs for major livestock species; however, most of these QTLs were detected using sparse microsatellite markers with large confidence intervals covering several megabases of the genome containing dozens to hundreds of genes and variants. NCBI2R is another R package that annotates lists of SNPs and/or genes, with current information from NCBI, including LD information. R functions in this package will provide annotation of the results from GWAS to provide a broader context of their meaning.

Breeding: Marker assisted selection of high yielding livestock for various production traits. Molecular identification and characterization of indigenous breeds of livestock, Impact of climate changes on disease susceptibility and resistance using genetics and transcriptomic markers [15]. Health: Identification and characterization of SNP markers for disease resistance, epitope mapping, identification of pathogen based on 16srRNA sequencing, Phylogenetic tree for genetic similarity and divergence between isolates for molecular epidemiology, Phytogeography, evolution of pathogens, Host pathogen interaction, identification disease markers, Characterization of vaccine response at genomic and transcriptomic level [15].

Poultry
The recent sequencing and draft assembly of the chicken genome is an exciting development for all interested in avian biology [16]. It is an invaluable information resource that complements the growing list of tools now available in avian genomics. To fully exploit these resources in poultry research and animal breeding is a challenge. The data is freely available through a number of genome browsers world-wide. However, to extract the full benefit of these resources requires much effort and learning new skills in bioinformatics. This is a rapidly developing field, with a huge array of open access software freely available over the internet.

Fishery sectors

Fisheries sector has been an integral component for growth of Indian agriculture. Increased attention on aquatic organisms for food and on the unique genes they may contain, combined with rapid biotechnological development has raised global concern for conservation of aquatic genetic resource. The Fish genetic resources have value in terms of economic, ecological and social uses. Thus, they need to be considered, managed and utilized prudently on sustainable basis and development of fish bioinformatics databases for the country. The coupling of genomics/functional genomics with telemetry has already made a significant impact on our understanding of wild fish migration and established real-time genotyping for fish migration.

Other sectors-microbiology, Computer Science

The most important method for analyzing sequence is alignment. There is pairwise and multiple sequence alignment, local and global ones. Alignment is a way of arranging the sequences of DNA, RNA, or protein, based on identity or similarity of subsequences There is special software for operations with protein spatial structures, their groups and protein-DNA, protein-ligand complexes *etc*. This software allows select structures from data banks and visualize them.

Field applications/ importance of bioinformatics

Over the past few decades, major advances in the field of molecular biology, coupled with advances in genomic technologies, have led to an explosive growth in the biological data generated by the scientific community. And sui generis right was available only to the makers of European Union. In U.S. copyright law is used to protect databases. In 1991, it was accessed whether databases can be protected through copyright or the standards of 'industrious collection' or 'sweat of brow' can be applied to them. The fact that laws in European Union and U.S.A. are so different, they have significant impact and the possibility of collaborative effort between EU and the U.S. seems remote. Therefore, the system of I.P rights can result in unintentional consequences. Because of these contradictions, complex databases such as the bioinformatics databases are not well covered under Database Law. Molecular approaches to disease diagnosis, Control and prevention of diseases through understanding the factors that influence evolution of pathogens. Selection of disease resistant and high yielding livestock. Establishing source of infection and disease transmission

Conclusion

Bioinformatics is the science of managing and analyzing biological information. Whereas all branches of modern biology make some use of computers, molecular biology and especially genetic engineering could not even exist without them. Given a certain research and computer infrastructure, developing countries may have relatively easy access to the products of bioinformatics. However, their future use of this technology hinges on the availability of bioinformatics knowledge in the public domain. India has more trained bioinformaticians than any other country on earth. These people, who are trained to use computers in the life sciences, and systems biologists, and biostatisticians, are the heroes of the coming biotech revolution. They keep it moving forward. Professionals here are uniquely positioned to both contribute to pushing this technology forward, and to make a huge amount of money, for themselves and the economy, while doing it. The development of various innovative technologies including improved varieties, agro-techniques and precision agriculture forms an important component of the strategies for improved production and productivity in any crop in agriculture that serves as an example to the rest of the world where the entrepreneurs and the policymakers have done much figured out how to balance making money with the needs of the people to receive affordable agriculture production care.

Application of review: In this present study the literature reviewed indicated that there is lot of scope to use bioinformatics tools and its applications in agriculture with practical values to improve the yield in future.

Review Category: Agri-informatics

Abbreviations

GWAS-Genome-Wide Association Studies
EST -Expressed Sequence Tag
WGP-Whole Genome Predictions
GS-. Genomic Selection
QTL- Quantitative Trait Locus
NGS. Next generation Sequence

*Principle Investigator or Chairperson of research: Dr M. Balakrishnan
Institute: ICAR-National Academy of Agricultural Research Management, Hyderabad, 500030, India
Research project name or number: Nil

Author Contributions: All author equally contributed

Author statement: All authors read, reviewed, agree and approved the final manuscript

Conflict of Interest: None declared

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] Baxevanis A. and Ouellette B.F.F. (2009) *A Practical Guide to The Analysis of Genes and Proteins. (Third Ed). John Wiley & Sons Inc*., 540.

[2] Luscombe N. M., Greenbaum D., Gerstein M. (2001) *Yearbook of Medical Informatics*, 83-99.

[3] Dhanapal A.P. and Govindaraj (2015) *Genetics Research International*, 1-15

[4] Kaitao Lai, Michał T. Lorenc and David Edwards (2012) *Agronomy,* 2, 62-73.

[5] Mochida K., & Shinozaki K. (2010) *Plant and Cell Physiology*, 51(4), 497–523.

[6] Luis Carlos Belarmino, Roberta Lane de Oliveira Silva, Nina da Mota Soares Cavalcanti, Nicolas Krezdorn, Ederson Akio Kido, Ralf Horres, Peter Winter, Günter Kahl and Ana Maria Benko-Iseppon (2013) *BMC Bioinformatics,* 14(Suppl 1), S2,-S2.

[7] De Filippis F., La Storia A., Villani F., Ercolini D. (2013) *PLoS ONE* 8(7), e70222.

[8] Lei Meng, HuihuiLi, Luyan Zhang and Jiankang Wang (2015) *The Crop Journal*, 3, 269-283.

[9] Fabian Grandke, Soumya Ranganathan, Nikkie van Bers, Jorn R. de Haan and Dirk Metzler (2017) *BMC Bioinformatics*, 18,12.

[10] Dash S., Van Hemert J., Hong L., Wise R.P., Dickerson J.A. (2012) *Nucleic Acids Research*,40 (D1), D1194-D1201.

[11] Rajagopal V., Manimekalai R. and Devakumar K. (2005) *Bioinformation*,1(2),75-77.

[12] Baxevanis A. and Ouellette F. (2001) *John Willey & Sons, Inc. N.Y.*, USA, 518.

[13] Rekha Sharma, Avishek Maitra, Pramod Kumar Singh, Madhu Sudan Tantia (2013) *Springer plus*, 2, 359.

[14] Daniel O'Connor Andrew J. Pollard (2013) *Clinical Infectious Diseases*, 57(6),860–869.

[15] Sharma N., Singh S.K., Lal S., Singh N.K. (2016) *Transcriptomics*, 4, 130.

[16] Warren W.C., Hillier L.W., Tomlinson C., *et al*., (2017). Genes/Genomes Genetics, 7(1), 109-117.