Review Article

Bioinformatics for Crop Breeding and Sustainable Agriculture

Mohammad Imran[1], Mohammad Rizwan[2], S.A. Kerkhi[1], Nazim Ali[3]

[1]Department of Genetics and Plant Breeding, Sardar Vallabhbhai Patel University of Agriculture and Technology, Meerut-250110
[2]Department of Entomology, Sardar Vallabhbhai Patel University of Agriculture and Technology, Meerut-250110
[3]Department of Animal Husbandry, Sardar Vallabhbhai Patel University of Agriculture and Technology, Meerut-250110

ABSTRACT

Bioinformatics deals with computer aided study of plant species. It includes gene mapping, gene sequencing, varietal information data base, plant genetic resources (PGRs) data base. It also deals with the study of genomics, proteomics and metabolomics in various plant species. It can be further divided as agricultural bioinformatics, horticultural bioinformatics, medicinal plants bioinformatics, forest plants bioinformatics, etc. The goal of modern crop breeding is to understand the genetic and molecular basis of all biological processes in plants that are relevant to the specie. This understanding is fundamental to allow efficient exploitation of plants as biological resources in the development of new cultivars with improved quality, resistant to various biotic and abiotic stresses and reduced economic and environmental costs. Thus, the key objectives for plant bioinformatics include: to encourage the submission of all sequence data into the public domain, through repositories, to provide rational annotation of genes, proteins and phenotypes, and to elaborate relationships both within the plants' data and between plants and other organisms. Bioinformatic studies are useful in pedigree analysis of various cultivars and hybrids. Information about the parentage of cultivars and hybrids is entered into the computer memory which can be retrieved any time. International Plant Genetic Resources Institute (IPGRI), Rome Italy has developed descriptors and descriptor states for various crop plants. Such descriptors help in uniform recording of observations on germplasm of crop plants all over the world. Thus, huge data is collected on crop genetic resources for several years which can be used for further crop improvement programme. Bioinformatics plays an important role in systematic management of this huge data.

Keywords: Bioinformatics; Crop Breeding; Genomics; Sustainable Agriculture

Introduction

Bioinformatics is the application of information technology to the study of living things, usually at the molecular level, it involves the use of computers to collect, organize and use biological information to answer questions in fields like evolutionary biology (Tao et al., 2004). The computer aided study of biology particularly genetics and molecular biology is called bioinformatics. Bioinformatics is the field of science in which biology, computer science, and information technology merge to form a single discipline. Bioinformatics is also defined as the application of computer technology to the management of biological information. It involves the development of software tools for the management and treatment of genetical information (Kane and Brewer, 2007). The definition of bioinformatics is not universally agreed upon. Generally speaking, it is defined as the creation and development of advanced information and computational technologies for problems in biology. Sustainable agricultural production and food security are two important issues of concern in response to population increase, environmental degradation and climate change. According to the United Nations, the world population increases by 70–75 million people

annually, an average of more than two persons every second; and over 95% of these will live in developing countries. It will be difficult satisfying the needs of this growing population and avoid serious food shortages or even famine from the limited arable land and natural resources available.

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. At the beginning of the "genomic revolution", a bioinformatics concern was the creation and maintenance of a database to store biological information, such as raw nucleotide and amino acid sequences. Bioinformatics is an interdisciplinary field mainly involving molecular biology and genetics, computer science, mathematics, and statistics. Data intensive, large-scale biological problems are addressed from a computational point of view. Now the science of bioinformatics is gaining increasing importance in life science specially in the field of molecular biology and plant genetic resources. Evolutionary biology looks at the molecules of different organisms and determines whether they share a common evolutionary history. This process has the potential to uncover relationships between life forms never considered before. By using bioinformatics to track this data, evolutionary biologists can gain new insights into the causes of and cures for various diseases.

This understanding is fundamental to allow efficient exploitation of plants as biological resources in the development of new cultivars of essential crops with improved quality and reduced economic inputs and environmental costs. This knowledge is also vital for the development of new plant diagnostic tools. Traits considered of primary interest are, pathogen and abiotic stress resistance, quality traits for plant, and reproductive traits determining yield. A genome programme can now be envisioned as a highly important tool for plant improvement. Current genome programs generate a large amount of data that will require processing, storage and distribution to the multinational research community (He et al., 2017). The data include not only sequence information, but information on mutations, molecular markers, maps, functional discoveries, etc. Therefore, the field of bioinformatics has evolved as the most pressing task now involves the analysis and interpretation of various types of data, including nucleotide and amino acid sequences, protein domains, protein structures and expression patterns. The actual process of analyzing and interpreting data is referred to the development and implementation of tools that enable efficient access to, and use and management of various types of information. There are newly developed fields in the science, related to bioinformatics and genetics.

Role of Bioinformatics in Plant Science

Plant informatics deals with computer aided study of plant species. It includes gene mapping, gene sequencing, varietal information data base, plant genetic resources data base. It also deals with the study of genomics, proteomics and metabolomics in various plant species. It can be further divided as agricultural bioinformatics, horticultural bioinformatics, medicinal plants bioinformatics, forest plants bioinformatics, etc. Bioinformatics is growing as an independent discipline and helping immensely to accelerate the growth of biotechnology. Its ultimate goal is to uncover the wealth of biological information hidden in the mass of data and to obtain a clearer insight into the fundamental biology of organisms. Bioinformatics has become a frontline applied science and is of vital importance to the study of new biology which is widely recognised as the defining scientific endeavour of the twenty-first century. The genomic revolution has underscored the central role of bioinformatics in understanding the very basics of life processes. The growth in full genomic sequencing, structural genomics, proteomics, micro-array etc. will be very slow without application of bioinformatics. In fact, usefulness of these areas to solve complex biological problems will be limited without bioinformatics and thus very high importance to bioinformatics.

Protein Modelling

Proteins have specific functions in our bodies determined by DNA sequences. Using bioinformatic techniques, scientists can test theories about how various proteins interact. These tests may help scientists understand how diseases develop in living organisms.

Genome Mapping

Genome mapping is another bioinformatic technique used for scientific research. Computerized genomic maps make it easier for scientists to locate genes, and

this increased efficiency results in higher productivity and greater scientific advancements. Due to this development in bioinformatics, researchers can spend less time on the painstaking mapping process and more time testing their hypotheses.

Bioinformatics and Plant Breeding

Plants are the basis of life on earth. They produce the life-supporting oxygen we breathe, they are essential for our nutrition and health and they provide the environment for the vast biodiversity on earth. For centuries, humans have selected plant varieties that best for their purposes and developed crop plants that have many advantages compared to natural plants in quality, quantity and farming practises. However, quantitative traits involved in resistance to various diseases, insects and quality have proven to be extremely difficult to improve, certainly in combination. The revolution in life sciences signalled by genomics dramatically changes the scale and scope of our experimental enquiry and application in plant breeding. The scale and high-resolution power of genomics enables to achieve a broad as well as detailed genetic understanding of plant performance at multiple levels of aggregation. The complex biological processes that make up the mechanisms of pathogen resistance and provide quality to our economically important crops are now open for a systematic functional analysis. These analyses are made with specific software on the high amounts of data generated in databases and is the field of plant bioinformatics. Genome initiatives are under way for more than 60 different plant species. For this point of view there are some major food crops like maize, rice, wheat, sorghum and barley; and the forage legumes soybean and alfalfa. Several of these genomes are so large that whole genome sequencing is impractical, and efforts have instead been focused on comparative genome methods. Both rice and maize, however, have relatively small genomes and are such key elements of the agricultural economies of the developed world that complete genome sequences have been prioritized.

Arabidopsis thaliana has a small genome (125 Mb total), which already has been sequenced in the year 2000, and it lacks the repeated, less-informative DNA sequences that complicate genome analysis. It has extensive genetic and physical maps of all 5 chromosomes; a rapid life cycle (about 6 weeks from germination to mature seed); prolific seed production and easy cultivation in restricted space; efficient

transformation merits utilising *Agrobacterium tumefaciens* is a large number of mutant lines and genomic resources and multinational research community of academic, government and industry laboratories. *Arabidopsis thaliana* has become universally recognised as a model plant for study. It is a small flowering plant that belongs to the Brassica family, which includes species such as broccoli, cauliflower, cabbage, and radish. Although it is a non-commercial plant, it is favoured among basic scientists because it develops, reproduces, and responds to stress and disease in much the same way as many crop plants. Researchers expect that systematic studies of Arabidopsis will offer important advantages for basic research in genetics and molecular biology and will illuminate numerous features of plant biology, including those of significant value to agriculture, energy, environment, and human health. Because of several reasons Arabidopsis has become the organism of choice for basic studies of the molecular genetics of flowering plants. Tomato Genomics Project includes the development of an integrated set of experimental tools for use in tomato functional genomics. The resources developed will be used to further expand our understanding of the molecular genetic events underlying fruit development and responses to pathogen infection, and will be made available to the research community for analysis of diverse plant biological phenomena.

In case of wheat, recent advances in plant genetics and genomics offer unprecedented opportunities for discovering the function of genes and potential for their manipulation for crop improvement. Because of the large size of the wheat genome, it is unlikely that the actual base pair sequences of the DNA molecules will be learned completely in the near future. This project takes an alternative strategy to realise the benefits of new techniques for discovering genes and learning their function (functional genomics). Following the identification of 10,000 wheat ESTs, they will be mapped to their physical location on the chromosomes of wheat (Gupta et al., 2008). This process utilizes a unique feature of the wheat chromosomes, their ability to tolerate deletions of portions of the chromosomes and still produce a viable plant. The mapping logic is direct: if an EST is present in a plant with complete chromosomes, but absent in a plant missing a known part of a single chromosome, then it can be inferred that the DNA sequence that corresponds to that EST is located in that segment of

the chromosome. By the end of the mapping component of this project, a most valuable tool will have been produced: 10,000 unique DNA sequences, likely corresponding to genes, whose physical location in the chromosomes of wheat are known. This sets the stage for the next phase of the project, the analysis of this array of mapped ESTs to determine function.

Molecular Plant Breeding

As the resolution of genetic maps in the major crops increases, and as the molecular basis for specific traits or physiological responses becomes better elucidated, it will be increasingly possible to associate candidate genes, discovered in model species, with corresponding loci in crop plants. Appropriate relational databases will make it possible to freely associate across genomes with respect to gene sequence, putative function, or genetic map position. Once such tools have been implemented, the distinction between breeding and molecular genetics will fade away. Breeders will routinely use computer models to formulate predictive hypotheses to create phenotypes of interest from complex allele combinations, and then construct those combinations by scoring large populations for very large numbers of genetic markers. The vast resource comprising breeding knowledge gathered over the last several decades will become directly linked to basic plant biology, and enhance the ability to elucidate gene function in model organisms. For instance, traits that are poorly defined at the biochemical level but well established as a visible phenotype can be associated by high resolution mapping with candidate genes. Orthologous genes in a model species, such as Arabidopsis or rice, may not have a known association with a quantitative trait like that seen in the crop, but might have been implicated in a particular pathway or signalling chain by genetic or biochemical experiments. The specific phenotypes of commercial interest that are expected to be dramatically improved by these advances include both the improvement of factors that traditionally limit agronomic performance (input traits) and the alteration of the amount and kinds of materials that crops produce (output traits). Examples include: - abiotic stress tolerance (cold, drought, salt) - biotic stress tolerance (fungal, bacterial, viral, chewing and sucking insect attack (feeding)) - nutrient use efficiency - manipulation of plant architecture and development (size, organ shape, number, and position, timing of development, senescence).

Marker Assisted Selection (MAS) for Multiple Disease Resistant

Identification and location of available molecular DNA markers have contributed significantly to marker-assisted studies and selection (MAS) in plant breeding, and in a wider range of research, including species identification and evolution. Indirect markers are closely linked, sometimes they may overlap, with a locus which determine this quantitative trait – QTL. QTLs (Quantitative Trait Loci) are defined as genes or regions of chromosomes which affect a trait. QTLs by themselves are difficult to be recognized. In both cases this information, or as it is called – markers, can be used in further selection purposes. This selection process is named as Marker Assisted Selection (MAS). As selection is based most on markers, higher density of mapping is important. The interval between marker and QTL of about 5 centi Morgans (cM) seemed sufficient for effective selection. Further development and detailed discussion on QTLs includes the statistical aspects of MAS, setting up the threshold of significance of marker effects, overestimation or bias in estimation of QTL effects, optimization of selection programs for several generations with simultaneous utilization of MAS and phenotypic data. Therefore, consequently, efficiency of Marker Assisted Selection will depend on the complexity of species/trait genetic architecture, on the development of the trait in the environment and on their interaction. From the economics point of view the use of markers will cost collection of DNA, genotyping, analyses, detection of QTLs etc. This high price is paid for the genotype building (there is no other way of doing that), for traits that are expensive for evaluation – disease resistance, or traits with low heritability.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Tao, Y., Liu, Y., Friedman, C., Lussier, Y.A. (2004). Information visualization techniques in bioinformatics during the postgenomic era. Drug Discov Today Biosilico 2(6): 237 – 245.

2. Kane, M.D., Brewer J.L. (2007) An information technology emphasis in biomedical informatics education. Journal of Biomedical Informatics 40(1): 67-72.

3. He, K.Y., Ge, D., He, M.M. (2017). Big Data Analytics for Genomic Medicine. Int J Mol Sci 18(2):412. doi:10.3390/ijms18020412

4. Gupta, P.K., Mir, R.R., Mohan, A., Kumar, J. (2008). Wheat genomics: present status and future prospects. Int J Plant Genomics. 2008: 896451.

How to Cite This Article:

Imran, M., Rizwan, M., Kerkhi, S.A., Ali, N. (2018). Bioinformatics for crop breeding and sustainable agriculture. Indian J. Biotech. Pharm. Res. 6(4):1–5.