

# APPLICATION OF BIOINFORMATIC TOOLS IN SUSTAINABLE AGRICULTURAL DEVELOPMENT

\*Rakesh Ranjan, \*Saket Vinayak and #Shanker Kumar Pandey

University Centre of Bioinformatics (Sub-DIC), T. M. Bhagalpur University, Bhagalpur- 812007

# Department of Botany, Jamtara College, Jamtara, Jharkhand-815351.

## Abstract

The emerging science Bioinformatics plays a significant role in the development of the agricultural sector, agro-based industries, agricultural by-products utilization and proper management of the environment. With the increase of sequencing projects, it continues to make considerable progress in biology by providing the genomic information. Bioinformatics is more often the tools for analysis of biological data than a discipline. The ability to represent high resolution physical and genetic maps of plants has been one of the great applications of bioinformatic tools. As the amount of data grows exponentially, there is a parallel growth in the demand for tools and methods in data management, visualization, integration, analysis, modeling, and prediction. Bioinformatic resources and web databases are essential for the most effective use of genetic, proteomic, metabolomic and phenome information important in increasing agricultural crop productivity. Innovations in web based platforms for omics based research, and application of such information has provided the necessary platform to promote molecular based research in model plants, as well as important crop plants. The information is set out under the molecular biology divisions of; DNA based resources and sequencing, RNA and variation analysis, proteomics, structural proteins, and post translation modifications and plant comparative analyses. The high degree of synteny among diverse plant species, commonality in traits, the availability of expression and function information of sequences has enabled the discovery of many useful traits for crop improvement. Genome sequencing of several important plants species has enabled researchers to identify 'chromosome' and 'difference' factor in sequences. This in turn has been used to identify value traits for crop improvement. Overall, a long list of future perspectives and research still to be attempted is detailed, which in the fullness of time should enable the full potential of bioinformatics and use in crop improvement programs to be achieved.

**Key Words:** Bioinformatic Tools, Sustainable Agriculture, Productivity, Population, Food Security.

**Introduction:** Bioinformatics refers to the study of biological information using concepts and methods in computer science, statistics, and engineering. It can be divided into two categories: biological information management and computational biology. The boundaries of these categories are becoming more diffuse and other categories will no doubt surface in the future as this field matures. Bioinformatics apply the principles of information sciences and technologies to complex life science data (Ojo and Maxwell, 2010). Sustainable agricultural production and food security are two important issues of concern in response to population increase, environmental degradation and climate change (Sinha, 2015). 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. These factors have already resulted in food deficiency and malnutrition, which have become serious health problems. Additionally, recent increased demand for biofuel crops has created a

new market for agricultural commodities, causing even more stress on food security (Wiebe, 2001). In order to try to resolve these problems and increase crop yields, breeding plants based on a better molecular understanding of gene function, and on the regulatory mechanisms involved in crop production appears to be necessary. Plant molecular biology continues to progress, and important gene sequences and their function have been described; many of which are related to crop yields (production), crop quality (protein and carbohydrate), and tolerance to biotic and abiotic stresses. There are legal, social and political barriers to the full potential use of crop biotechnology and transgenic plants, nevertheless advances in these fields have lead to improvements in agriculture and human life. One vital tool of biotechnology is 'bioinformatics', which is commonly used to genetically type and identify genotypic and phenotypic changes in plants, and this information is important for improvement in performance of crop plants (Mingsheng, et al. 2012).

**Tools for Agricultural Development:** There are so many tools used for crop improvement, however, some of the most common are as follows:

**Genome Sequencing Projects:** Initially, the publication and accumulation of nucleotide sequences for model plants only provided fundamental information, however now these base sequences form the fundamentals of research in functional plant genetics in applied species such as crops and domestic animals. Furthermore, DNA sequence data continues to be central in providing the genomic basis for accelerating molecular level understanding of basic biological mechanisms, and the application of such information to crops.

**NCBI—Bio Project:** The NCBI site provides genome sequences and information for many plant species (Viridi plantae) designed to facilitate comparative genomic studies amongst the many other records of plants there. The current version consists of documentation in at least 115 different plants with partial sequences, and about 40,000 Expressed Sequence Tags (ESTs). It also contains separate sites and resources for other web based tools, data banks and other web servers, including agronomically important crops for food or fruit, medicinal plants, a number of green algae, pathogenic bacteria and fungi, viruses and animals.

**Phytozome:** The site includes genome sequences and data sets for various crop species designed to facilitate comparative genomic studies amongst other green plants. The current version consists of 31 plant species wholly or partially sequenced, and is set-up into 10 evolutionary significant nodes.

**Gramene:** An information resource established as a portal for grass species and grass genomics; including genome sequence information. The current version provides data on 24 plants; including 12 wild and domesticated rice genomes. An organelle data bank is also available from this site.

**NCBI-Entrez:** Tracks over 800 whole genome projects from biological organisms, and the 115 species of Viridiplantae; including agronomically important crops for food and fruit, medicine and a number of green algae. The Entrez database can be accessed through the home page of NCBI.

**NCBI-BLAST:** One of the most important sites and tools available, in determining base similarities between nucleotide sequences in databanks. It also contains protein searches and queries. It includes searches for translated nucleotide sequences, conserved domains, multiple alignment tools, evolutionary relationships, and can be applied to all organisms, or limited to specific plants.

**Grain Genes and Plant GDB:** Grain Genes is a specific database for Triticeae and Avena genes, markers, maps and germplasm. Plant GDB contains sequences and a search engine linked to NCIB BLAST for 15 Dicotyledon, 7 Monocotyledon and 3 other plant species. It contains more limited information than the NCBI site over most plants, but is especially useful for agricultural grain species.

**DNA Sequencing and Ultra High-Throughput:** Genome sequence information aids researchers in identify genes and gene families, including the identification of coding or non-coding regions, regulatory genes, and repetitive sequences within the genome (e.g. simple sequence repeats-SSRs); all of these are important in molecular biology. Pyro sequencing, massive parallel DNA sequencing and single molecule sequencing are adaptations of existing methods, which have become available in recent years (Dvorak and Akhunov, 2005). It is expected to accelerate the use of the DNA based web-information considerably in crop plants.

**Whole Genome Sequencing:** Information obtained from whole-genome sequencing in plants allows attempts at chromosome-scale genetic comparisons, thereby identifying conserved genetic areas, which can facilitate identification and documentation of similar genomic sequences in related plant species (Huang X and Yang, 2005). Whole genome comparisons identifying chromosomal duplication of alleles among related species for example can provide comparative evolutionary histories and diversification of species in ecology, taxonomy and plant breeding (Upadhyaya et.al, 2009).

**Molecular (DNA) Markers:** 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. The use of DNA markers for direct selection offers greater potential gains in breeding for QTL and traits with low heritability, and these can be the most difficult to work with in crop breeding. Crop breeders have known the complexity of multiple alleles for decades. However with the advent of molecular markers, genetic diversity and other forms of genetic structure in breeding populations is possible.

**Restriction Fragment Length Polymorphism (RFLP):** RFLP requires hydrolysis of probe DNA from samples. RFLP can provide high quality data but has severe restrictions on throughput because large amounts of DNA are required, and because it is not based on amplification of the target DNA via the polymerase chain reaction (PCR).

**Random Amplified Polymorphic DNA (RAPD):** RAPD is a method based on PCR but uses arbitrary short primers (10 bases long) to identity plant DNA regions. No knowledge of the genome is needed, but by the same token markers can target many places on the genome. Results can be inconsistent and only dominant genes can be identified.

**Simple Sequence Repeats (SSR):** SSR are high quality and consistent DNA markers, but they are the most expensive to develop. SSR markers require extensive band sequencing data for each marker developed, and often the markers are species and even cultivar specific. However they are molecular markers of choice in crop plants.

**Amplified Fragment Length Polymorphism (AFLP):** AFLP requires enzymatic degradation of DNA and careful fragment separation, where only a sub-fraction of the population genetic data is sampled by PCR. It can provide too much information at any time. It is more technically demanding and information can be difficult to interpret. It produces very good high quality data, which is suitable for high output sources and automation.

**Single Nucleotide Polymorphism (SNP):** SNP relies on the fact that the vast majority of differences in eukaryotic organisms are surprising point mutations in their DNA. So there are a vast number of polymorphisms that are SNPs. The biggest advantage is automation and techniques that do not require electrophoresis to separate fragments. However it does require DNA sequencing which can be costly. SNPs

are becoming more and more important as molecular markers for genome information and advancement in crop plants.

**Expressed Sequence Tags (EST):** ESTs require cDNA synthesis from RNA, and therefore are the only markers listed which are based on RNA. Preferences for this method should be given to crop species where there is already extensive sequencing, and part or full EST data are present.

**NCBI—Plant Markers:** A genetic marker web database that contains molecular markers such as SNP, SSR and conserved ortholog set cosmid (COS) markers and primers from various plant resources.

**Grain Genes:** The web site for Triticeae genomics provides considerable detail of DNA markers and chromosome linkage map data on wheat, barley, rye and oat.

**Bioinformatics and sustainable development of Agriculture:** Bioinformatics is widely applied in agricultural research. Collection and storage of plant genetic resource and wisely application of bioinformatics help to produce stronger, more drought, disease and insect resistant crops and improve the quality of livestock making them healthier, more disease resistant and more productive.

(a) **Crops:** Comparative genetics consists of the model and non-model plant. Species can reveal an organization of their genes with respect to each other which further use for transferring information from the model crop systems to other food crops. *Arabidopsis thaliana* (water cress) and *Oryza sativa* (rice) are examples of available complete plant genomes (Proost et al 2009).

(b) **Renewable Energy:** Biomass based crop species such as maize (corn), switch grass and lignocellulosic species like bagasse and straw are widely used for biofuel production. We could detect sequence variants in biomass-based crop species to maximize biomass production and recalcitrance. Thus, the use of genomics and bioinformatics in combination with breeding would likely increase the capability of breeding crop species to be being used as biofuel feedstock and consequently keep increasing the use of renewable energy in modern society (Boyle et al 2004,).

(c) **Insect resistance:** *Bacillus thuringiensis* (Bt) genes control a number of serious pests that have been successfully transferred to cotton, maize and potatoes. These crops are known as Bt crops. This new ability of the plants to resist insect outbreak may reduce the amount of insecticides being used.

(d) **Improve Nutritional Quality:** Scientists have recently succeeded in transferring genes into rice to increase levels of Vitamin A, iron and other micronutrients. Bioinformatic tool helped to produce such golden rice that can fight against vitamin A deficiencies. This work could have a profound impact in reducing occurrences of blindness and anemia caused by deficiencies in Vitamin A and iron respectively (Paine et al 2005). Scientists have inserted a gene from yeast into the tomato, and the result is a plant whose fruit stays longer on the vine (Fraser et al 2009).

(e) **Grow in Poorer Soil and Drought Resistant:** Progress has been made in developing cereal varieties that have a greater tolerance for soil alkalinity, free aluminium and iron toxicities. These varieties allow agriculture to succeed in poorer soil areas, thus adding more land to the global production base. Research is in progress to produce crop varieties capable of tolerating reduced water conditions (Wang et al 2004).

(f) **Plant Breeding:** 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.

(g) Agriculturally Important Microorganism: Bioinformatics helps to understand the genetic architecture of microorganism and pathogens to check how these microbes affect the host plant by using meta genomics and transcriptomics approach, so that we could generate pathogen resistant crop and would identify those microbes which are beneficial for host.

(h) Accelerate Crop Improvement in a Changing Climate: The change in climate and increase in population will increase pressure on our ability to produce sufficient food. The breeding of novel crops and the adaptation of current crops to the new environment are required to ensure continued food production. Advances in genomics offer the potential to accelerate the genomics based breeding of crop plants.

(i) Bioinformatics in Plant Disease Management: Pathogen trait is considered as a primary interest of plant bioinformatics. The current efforts to determine gene and protein functions, have improved the ability to understand the root causes of plant diseases and find new cures. Bioinformatics have many applications in current plant disease management with respect to the study of host pathogen interactions, understanding the disease genetics and pathogenicity factor of a pathogen which ultimately help in designing best management options.

**Conclusion:** Bioinformatics is an interdisciplinary area of the science composed of biology, mathematics and computer science. During the last two decades enormous data has been generated in biological science, firstly, with the onset of sequencing the genomes of model organisms and, secondly, rapid application of high throughput experimental techniques in laboratory research. Biological research that earlier used to start in laboratories, fields and plant clinics is now starts at the computational level using computers (In-silico) for analysis of the data, experiment planning and hypothesis development. Bioinformatics develops algorithms and suitable data analysis tools to infer the information and make discoveries. Application of various bioinformatics tools in biological research enables storage, retrieval, analysis, annotation and visualization of results and promotes better understanding of biological system in fullness. This will help in plant health care based disease diagnosis to improve the quality as well as the quantity of Plant. The challenges faced by the bioinformatics community today are the intelligent and efficient storage of huge amount of data generated, and to provide easy and reliable access to this data. Therefore, incisive computer tools must be developed to allow the extraction of meaningful biological information. Emerging trend in agricultural and pharma industry is to apply bioinformatics tools to reduce time and cost in molecular marker and sustainable agricultural development.

### References:

Blueggel M, Chamrad D, Meyer HE (2004): Bioinformatics in proteomics. *Curr. Pharm. Biotechnol.* 5:79–88

Boyle G. Renewable energy (2004). Oxford university press. Turner JA. A realizable renewable energy future *Science*, 285: 687-689.

Dvorak J, Akhunov ED (2005). Tempos of gene locus deletions and duplications and their relationship to recombination rate during diploid and polyploid evolution in the *AegilopsTriticum* alliance. *Genetics* 171: 323–332.

Fraser PD, Enfissi E, Bramley PM. (2009). Genetic engineering of carotenoid formation in tomato fruit and the potential application of systems and synthetic biology approaches. *Archives of Biochemistry and Biophysics*, 483: 196-204

- Huang X, Yang SP (2005). Generating a genome assembly with PCAP. *Curr Protoc Bioinformatics* 11: 11.13
- Mingsheng, F.; Jianbo, S.; Lixing, Y.; Rongfeng, J.; Davies, W.J.; Fusuo, Z. (2012) Improving crop productivity and resource use efficiency to ensure food security and environmental quality. *J. Exp. Bot.*, 63, 13–24 Ansorge 2009
- 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
- Paine JA, Shipton CA, Chaggar S, Howells RM, Kennedy MJ, Vernon G, et al. (2005). Improving the nutritional value of Golden Rice through increased pro-vitamin A content. *Nat Biotechnol.* 23: 482-487
- Proost, S.; van Bel, M.; Sterck, L.; Billiau, K.; van Parys, T.; van de Peer, Y.; Vandepoele, K. Plaza (2009) A comparative genomics resource to study gene and genome evolution in plants. *Plant Cell*, 21, 3718–3731.
- Sinha S. (2015) Role of bioinformatics in climate change studies. *J Science.*;1:1-9.
- Upadhyaya HD, Pundir RPS, Dwivedi SL, Gowda CL, Reddy VG, Singh S (2009) Developing a mini core collection of sorghum for diversified utilization of germplasm. *Crop Sci* 49(5):1769–1780.
- Wang S, Wan C, Wang Y, Chen H, Zhou Z, Fu H et al. (2004). The characteristics of Na<sup>+</sup> and K<sup>+</sup> and free proline distribution in several drought-resistant plants of the Alxa Desert, China. *Journal of arid environments.*, 56: 525-539.
- Wiebe, K. (2001). “Natural Resources, Agricultural Productivity, and Food Security”. *Agriculture Information Bulletin Number 765-3*, Economic Research Service, United States Department of Agriculture (USDA).



\*\*\*\*\*