Int.J.Curr.Microbiol.App.Sci (2020) Special Issue-11: 3964-3969

International Journal of Current Microbiology and Applied Sciences
ISSN: 2319-7706 **Special Issue-11 pp. 3964-3969**
Journal homepage: http://www.ijcmas.com

Review Article

Application of Bioinformatics in Plant Breeding System

Akanksha Tiwari[1]*, Piyusha Singh[1] and Sarla Kumawat[2]

[1]Department of Agriculture, Azamgarh, ANDUAT, Ayodhya, U.P., India
[2]JNKVV, Jabalpur, M.P., India
**Corresponding author*

A B S T R A C T

Bioinformatics is an interdisciplinary area of the science composed of biology, mathematics and computer science. The key role of bioinformatics is acquiring striking importance in the era of outstanding advances in omics technologies for its fundamental support in describing the multifaceted aspects of biological functionalities. The manifold omics efforts flourishing worldwide are also contributing fundamental novelties in many aspects of agricultural sciences and, as a consequence, bioinformatics is acquiring a crucial role also in these research fields. Nucleic acid sequences and information from a wide range of genomes become possible through genomics with an unprecedented pace. Genomics made this information accessible to further analysis and experimentation.The rise of third generation sequencing technologies is helping overcome challenges in plant genome assembly caused by polyploidy and frequent repetitive elements. 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 of Plant.

Keywords

Bioinformatics, mathematical, statistical, computing methods

Introduction

A wide range of definitions have been assigned to bioinformatics. The definition used by most people is narrower because, bioinformatics to them is a synonym for "computational molecular biology"- the use of computers to characterize the molecular components of living organisms. In classical definition of bioinformatics, it is the mathematical, statistical and computing methods aimed at resolve biological problems using DNA, RNA and amino acid sequences and related information. According to National Center for Biotechnology Information (NCBI) definition, bioinformatics is the field of science that merge biology, computer science and information technology into a single context (Dayhoff, 1969). Bioinformatics is generally referred to as the application of computer technology to the processing and managing the data generated in biological experiments. The term bioinformatics was originally coined for the application of information technology to large volumes of biological, particularly, genomic data.

The field of bioinformatics has come to be intermingled with traditional computational biology and biostatistics, which are strictly concerned not only with how to handle the

information itself, but rather how to extract biological meaning from it. This new knowledge could have profound impacts on different fields, such as human health, agriculture, environment, energy and biotechnology. Bioinformatics has emerged as a tool to facilitate biological discoveries more than a decade ago. The publication of the completed Arabidopsis thaliana genome sequence (AGI, 2000) and draft sequence for rice genome (Goff *et al.,* 2002) the plant research and industry has step over the threshold of the genomics era. The numerous applications of genomic information opened wide opportunities to start integrating rich rewards from sub-systems biology, integrative biology and large scale systematic functional genomics projects. With the development of Human Genome Project, the data of biology increased fabulously and marvellously. The ability to capture, manage, process, analyse and interpret data became more important than ever. Bioinformatics is a new field of science but it is making progress in every field of biotechnology very rapidly. As it has its application in the medicine by providing the genome information of various organisms, similarly the field of agriculture has also taken advantage of this field because microorganisms play an important role in agriculture and bioinformatics provides full genomic information of these organisms. The genome sequencing of the plants and animals has also provided benefits to agriculture.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.

Bioinformatics emerged from the initial requirement of suitable informatics for biological data organization, management, and distribution [Dayhoff, 1965], but soon it revealed also fundamental in providing tools for data analysis, interpretation, and modeling. The fast spreading of omics techniques, with its growing power and more accessible costs, drastically increased the amount of molecular data collections from different levels of organization of an organism or an environmental sample. This favored a holistic view on systems organization and functionality, further challenging bioinformatics with data size and the need of integrative efforts (Chiusano *et al*., 2008, Bostanand Chiusano, 2015). The goal of plant genomics is to understand the genetic and molecular basis of all biological processes in plants that are relevant to the species. 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. In this review, our objective is to discuss an updated synthetic view of how bioinformatics can effectively improve the efficiency of breeding programs and overcome bottlenecks in crop improvement.

Why is bioinformatics important?

Bioinformatics has become an important part of many areas of biology. In experimental molecular biology, bioinformatics techniques such as image and signal processing allow extraction of useful results from large amounts of raw data. In the field of genetics and genomics, it aids in sequencing and annotating genomes and their observed mutations. It plays a role in the textual mining of biological literature and the development of biological and gene ontologies to organize and query biological data. It plays a role in the analysis of gene and protein expression and regulation. Bioinformatics tools aid in the comparison of genetic and genomic data and more generally in the understanding of evolutionary aspects

of molecular biology. At a more integrative level, it helps analyze and catalogue the biological pathways and networks that are an important part of systems biology. In structural biology, it aids in the simulation and modeling of DNA, RNA, and protein structures as well as molecular interactions. Researchers affiliated with our program conduct research in systems biology, genomics, and proteomics.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 pharma industry is to apply bioinformatics tools to reduce time and cost in molecular marker development, drug development [Untergasser *et al*., 2007, Kumari *et al*., 2011]

Plant Breeding in relation to bioinformatics

An interdisciplinary approach is needed for plant breeding in the 21st century to identify and resolve breeding challenges and improve crop production (Moose *et al.,* 2008). Together with novel glasshouse technologies to accelerate plant development (Watson *et al.,* 2018), genomics and bioinformatics play an important role in increasing the production rate of improved crop cultivars. Nevertheless, the vast amounts of genotypic and phenotypic data available create an enormous challenge to integrate diverse data outputs for breeding (Santos *et al.,* 2017). Integrating phenotypes, genomics and bioinformatics tools and resources in public and private breeding pipelines will address this challenge and help deliver breeding targets (Evans *et al.,* 2013).

Crop breeding has long relied on cycles of phenotypic selection and crossing, which generate superior genotypes through genetic recombination. When genome sequences are available, all genes and genetic variants contributing to agronomics traits can be identified and changes made during breeding processes can be assessed at the genotype level. Because of the ready availability of genomic data for breeders today, genomics plays an increasingly important role in all aspects of crop breeding, such as quantitative trait loci (QTL) mapping and genome-wide association studies (GWAS), where genomic sequencing of crop populations can allow gene-level resolution of agronomic variation. For example, advances in genomics-based breeding allow the identification of genetic variation in crop species, which can be applied to produce climate resilient crops [Perez-de-Castro *et al*., 2012, Mousavi-Derazmahalleh *et al*., 2018, Dwivedi *et al*., 2017].The complex biological processes that make up the mechanisms of pathogen resistance and provide quality to our crops are now open for a systematic functional analysis. These analysis are made with specific software on the high amounts of data generated in databases and is the field of plant bioinformatics. (Neerincx and Leunissen, 2005, Meyerand Mewes, 2002).

Genomics approaches are particularly useful when dealing with complex traits, as these traits usually have a multi-genic nature and an important environmental influence. Thanks to these technological improvements it is now feasible for a small laboratory to generate in a short time span (e.g., several months) enough molecular data to obtain a set of mapped quantitative trait loci (QTLs), even in a species lacking any previous genomic information [Varshney *et al*., 2010]. Genomic tools are thus facilitating the detection of QTLs and the identification of existing favourable alleles of small effect, which have frequently remained unnoticed and have not been included in the gene pool

used for breeding [Morgante *et al.*, 2003, Vaughan *et al*., 2007].

Molecular information and plant breeding – a bioinformatics approach

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. 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 (Walsh, 2001, Deckers, 2002).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 (Hospital *et al.*, 2002). In the recent years an increasing amount of information for the DNA polymorphism and sequencing was accumulated in different plant varieties and cultivars. Most of this information was used for the purpose of recognition of different cultivars as well as for their comparison – distances and similarities (Reif *et al.*, 2005).

Role of Genomics in Agriculture

Agricultural genomics, or agrigenomics (the application of genomics in agriculture), has and will continue to drive sustainable productivity and offer solutions to the mounting challenges of feeding the global population. Using modern technology, farmers, breeders, and researchers can easily identify the genetic markers linked to desirable traits, informing cultivation and breeding decisions. Agricultural genomics is a rich field that has been contributing to advances in crop development for decades. From sequencing reference genomes to genotyping for genome-wide association studies to genomic prediction, advances in technology and applications have led to breakthroughs in crop improvement. These innovations have resulted in elite cultivars that have been selected for agriculturally desirable traits, including high yield, stress tolerance and pest resistance

Comparative genetics of the plant genomes has shown that the organization of genes has remained more conserved over evolutionary time than was previously believed [Mahalakshmi and Ortiz, 2001, Matthews *et al.*, 2003, Caetano-Anolles. 2005, Jaiswal, 2006]. The complete sequencing of the Arabidopsis thaliana genome has been regarded as a landmark in plant sciences (Dennis and Surridge, 2000).

These findings suggest that information obtained from the model crop systems can be used to suggest improvements to other food crops. Arabidopsis thaliana (water cress) Oryza sativa (rice), *Triticum aestivum* (wheat) and *Zea mays* (Maize) are examples of available complete land plant genomes (Paterson et. al., 2005, Varshney *et al.*, 2006).Bioinformatics is thoroughly involved with the completion and assessment of a multitude of different complete genome sequences (Claverie and Notredame, 2003). The sequencing of the genome of a species is not the end by itself but is just the beginning of a new venture to unravel genetic information and to gain better insight into the genetics of other species under investigation. The numerous applications of genomic information opened wide opportunities to start integrating rich rewards from sub-systems biology, integrative biology and large scale systematic functional genomics projects. With this accumulation of various

types of data is possible freely to enter the universe of "genomic understanding".

Significance of genome after sequencing

"We've sequenced the genome, put it back together and identified the genes, but now we need to find out what this genome can tell us and how it compares to other genomes."

The sequence tells scientists the kind of genetic information that is carried in a particular DNA segment. For example, scientists can use sequence information to determine which stretches of DNA contain genes and which stretches carry regulatory instructions, turning genes on or off. In addition, and importantly, sequence data can highlight changes in a gene that may cause disease.

A simple comparison of the general features of genomes such as genome size, number of genes, and chromosome number presents an entry point into comparative genomic analysis.

Genome Comparison Tools

MegaBlast is NCBI BLAST based algorithm for large sequence similarity search (Hesslop-Harrison, 2000). MegaBlast implements a greedy algorithm for the DNA sequence gapped alignment search. MegaBlast is used to compare the raw genomic sequences to a database of contaminant sequences (including the UniVec database of vector sequences, the Escherichia coli genome, bacterial insertion sequences, and bacteriophage databases).

The rate of improvement of genetic yield potential has to be increased beyond the rates currently achieved in ongoing breeding programs to protect global food security in times of rapid population growth and climate change. Thus, new or different approaches are needed to accelerate the crop breeding process. Agriculture faces substantial challenges in harnessing the deluge of genomic data of diverse origins and formats for crop improvement. To overcome these challenges, novel breeding methods and bioinformatics tools must be used to translate genomic data into gains in crop yield and yield stability. Recent advances in bioinformatics application for plant genomes not only provide huge potential for large-scale genomic research among plant species but also many technical challenges. Despite these exciting achievements, there remains a critical need for effective tools and methodologies to advance plant biotechnology, to tackle questions that are hardly solved using current approaches, and to facilitate the translation of this newly discovered knowledge to improve plant productivity.

References

AGI. 2000. Nature, 408, 796-815.

Bostan H, Chiusano M L. 2015. NexGenEx-Tom: a gene expression platform to investigate the functionalities of the tomato genome. BMC Plant Biol.15(1):48

Caetano-Anolles. 2005. Evolution of Genome Size in the Grasses. Crop Science, 45, 1809-1816.

Chiusano M L, D'Agostino N, Traini A, Licciardello C, Raimondo E, Aversano M, Frusciante L, Monti L. 2008. ISOL@: an Italian SOL Anaceae genomics resource. BMC Bioinform. 9(Suppl 2):S7

Claverie J-M., Notredame C. 2003. Bioinformatics for Dummies. Willey Publ. Inc. N.Y., USA, p. 452.

Dayhoff M O. 1965. Atlas of protein sequence and structure. Silver Spring: National Biomedical Research Foundation.

Dayhoff M O. 1969. Computer analysis of protein evolution.Sci Am 221 (1):86-95.

Deckers J., Hospital F. 2002.Nature Reviews Genet., 3, 22-32

Dennis, C. and Suggidge, C. 2000. A. thaliana genome. Nature 408:491.

Evans, K.; Jung, S.; Lee, T.; Brutcher, L.; Cho, I.; Peace, C.; Main, D. 2013. Addition of a breeding database in the Genome Database for Rosaceae. Database.

Goff S. A., Ricke D., Lan, Presting T. H., Wang G., Dunn R.M. *et al.*, 2002. Science, 296, 92-100.

Hesslop-Harrison J. S. 2000. Comparative Genome Organization in Plants: From Sequence and Markers to Chromatin and Chromosomes. Plant Cell, 12, 617-636.

Hospital F., Bouchez A., Lecomete L., Causse M., Charcosset A. (2002) 7th WCGALP, Montpellier, France, 22-05

Jaiswal P., Ni J., Yap I., Ware D., Spooner W., Youens-Clark K., Ren L., Liang C., Zhao W., Ratnapu K., Faga B., Canaran P., Fogleman M., Hebbard C., Avraham S., Schmidt S., Casstevens T. M., Buckler E.S., Stein L. and McCouch S. 2006. Gramene: a bird's eye view of cereal genomes. Nucleic Acids Research, 34:717-723.

Kumari N., Singh V. K., Narayan O. P., Rai L. C. 2011. Toxicity of butachlor assessed by molecular docking to NusB and GroES protein. Online Journal of Bioinformatics, 12, 289-303.

Mahalakshmi V. and Ortiz R. 2001. *Plant genomics and agriculture: From model organisms to crops, the role of data mining for gene discovery*. Electronic Journal of Biotechnology, 3.

Matthews D. E., Carollo V. L., Lazo G. R. and Anderson O. D. 2003. Bioinformatics and Triticeae Genomics: Resources and Future Developments. Nucleic Acids Research, 31, 183-186.

Meyer K., Mewes H. W. 2002.Curr.Opin. Plant Biol., 5, 173-177

Moose, S. P.; Mumm, R. H. 2008.Molecular plant breeding as the foundation for 21st century crop improvement. Plant Physiol147, 969–977.

Morgante, M.; Salamini, F. 2003.From plant genomics to breeding practice. Curr.Opin.Biotechnol., 14, 214-219.

Neerincx P., Leunissen J. 2005. Briefings in Bioinformatics, 6(2), 178-188. 16.

Paterson A. H., Freeling M. and Sasaki, T. 2005.Grains of knowledge: Genomics of model cereals. Genome Research, 15, 1643-1650.

Perez-de-Castro, A. M.; Vilanova, S.; Cañizares, J.; Pascual, L.; M Blanca, J.; J Diez, M.; Prohens, J.; Picó, B. 2012. Application of genomic tools in plant breeding.Curr.Genom.13, 179–195

Reif J. C., Melchinger A. A., Frisch M. (2005) Crop Sci., 45, 1-7.

Santos, R.; Algar, A.; Field, R.; Mayes, S. 2017. Integrating GIScience and Crop Science datasets: A study involving genetic, geographic and environmental data. PeerJ Preprints, 5.

Untergasser A., Nijveen H., Rao X., Bisseling T., Geurts R. and Leunissen J. A. M. 2007. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Research, 35, W71-W74.

Varshney R. K., Hoisington A. D. and Tyagi K. A. (2006) Trends in Biotechnology, 24, 1-10

Varshney, R. K.; Glaszmann, J. C.; Leung, H.; Ribaut, J. M. More genomic resources for less-studied crops. Trends Biotechnol., 2010, 28, 452-460.

Vaughan, D. A.; Balász, E.; Heslop-Harrison, J. S. 2007. From crop domestication to super-domestication. Ann. Bot., 100, 893-901.

Walsh B. 2001. Theor. Pop. Biology, 59, 175-184.

Watson, A.; Ghosh, S.; Williams, M. J.; Cuddy, W. S.; Simmonds, J.; Rey, M. D.; Hatta, M. A. M.; Hinchliffe, A.; Steed, A.; Reynolds, D.; *et al.*, 2018. Speed breeding is a powerful tool to accelerate crop research and breeding. Nat. Plants 4, 23–29.