BIOINFORMATICS IN AGRICULTURE: TOOLS AND SOFTWARES : A REVIEW

[1]NANDHA ABHIJEETA K. AND *[2]MEHTA D. R.

DEPARTMENT OF BIOTECHNOLOGY
JUNAGADH AGRICULTURAL UNIVERSITY
JUNAGADH, GUJARAT - 382 001, INDIA

EMAIL: drmehta@jau.in

[1]Ph.D. Scholar, Department of Biotechnology, J.A.U., Junagadh, Gujarat, India
[2]Professor & Head, Dept. of Genetics & Plant Breeding, J.A.U., Junagadh, Gujarat, India

ABSTRACT

In the era of omics, bioinformatics is striking key role is in fundamental support in describing the multi-faceted aspects of various biological functionalities. It is also blooming worldwide in introducing various novel and additional aspects of agriculture science and make its place in front side in research field. For the improvement of various traits in plants, animals and microbes and to improve various nutritional values and quantity of product to which changes at molecular level using various software and databases available. The knowledge of various mechanism and specific phenotypic traits as well as specific response toward various biotic and abiotic stress, is boon of bioinformatics. We describe various database and software which is highly used in agriculture for development and improving quality purpose. Bioinformatics have great impact on development of Next generation sequencing as in genomics, transcriptomics and proteomics to fulfil various cutting edge and to resolve some unsolved questions.

KEY WORDS: Biological data, bioinformatics tools, BLAST, FASTA, Agriculture

INTRODUCTION

In today's era of omics, bioinformatics plays very key and significant role in various sectors, such as agriculture, biotechnology, forensic science, medical, drug designing and development etc. Bioinformatics is new emerging science, which is combination of statistics which have ability to solve various biological problems and mathematical algorithms along with powers of computer science. Bioinformatics is a form of interdisciplinary of science and technology which adjoining a systematic development and various applications for solution regarding biological information by addressing biological data collections and warehousing, data mining, on the basis of requirement research and usage of various stored data, their analysis and interpretation, design new model or drugs for highly specific usage (Degrave *et al*., 2002; Xiong, 2009; Jayaram and Dhingra, 2010) with the help of these technologies advances have been accelerated at genome scale development in various model plants such as *A. theliana*, soybean, rice, switch grass etc. (Mochida and Shinozaki, 2010). Such technologies make more feasible to next generation technologies, for example; whole genome sequencing for determination of variation analysis, RNA sequencing from transcriptome analysis, Chip-seq analysis for DNA-protein interaction (Lister *et al*., 2009;

Noman and Aqeel, 2017). Apart from that other various approaches are as, phytohormone-mediated cellular signalling approach for analysis of various hormone (Kojima *et al.*, 2009; Ali *et al*., 2017), metabolome analysis approach for determination of various known and unknown metabolic system of plants (Saito and Matsuda, 2010), Protein-protein interaction (Arabidopsis Interactome Mapping Consortium., [AIMC] 2011). For omics-based research, various technologies can be very crucial which helps to facilitate the exchange of knowledge and help to improve present data (Shinozaki and Sakakibara, 2009; Noman *et al.*, 2017b). Moreover, such research innovations help to disclosed that genomics, transcriptomics, metabolomics and proteomics are important shift but out of these all, genomics and proteomics are more crucial than others. As genomics helps to discovering various novel genes which can ultimately beneficial to update the agriculture via biotechnological programs (Noman *et al.*, 2017a,b).

Bioinformatics was invented by Paulien Hogeweg and Ben Hesper in 1970 as "the study of informatic processes in biotic systems". A Dutch theoretical biologist and complex system researcher, Paulien Hogeweg, studied biological systems as dynamic information processing systems at many interconnected levels. It is too difficult to process, analysed and stored all the data by single person or team of few people. Thus, there is arising need of various alternative methods as computational methods to process and contextualized such data. Various studies at genomic and proteomic level promoted the discovery of new genes and their functions.

Fields of bioinformatics

- Agriculture
- Cropimprovement
- Developmentofdroughtresistantvarieties
- Molecularmedicine
- Biotechnology
- Climate changeStudies
- Alternative energysources
- Microbial genomeapplications
- Genetherapy
- Drugdevelopment
- Antibioticresistance
- Evolutionarystudies
- Wastecleanup
- Forensicanalysis
- Bio-weaponcreation
- Insectresistance
- Improve nutritionalquality
- VeterinaryScience

Table 1: Various categories of bioinformatics and numbers of database included in that Category (Vassilev *et al.*, 2005)

Category	Number of Databases
Genomic	164
Nucleotide sequences	59
Protein sequences	87
Human/vertebrate genomes	77
Structures	64
Microarray/gene expression	39
Metabolic and signaling pathways	33
RNA sequences	32
Others	6

Bioinformatics tool

Biological databases and its management

Biological databases are archives of different data that are stored as particular pattern in uniform and efficient manner. These collections of databases contain data from vast spectrum of genetics and molecular biology areas which includes various fields. A simple database might be a single file containing many records, each of which might includes the same set of information (Xiong, 2009), such databases are prepared using computer hardware and software for data management. The primary objective of the development of such database is to organize specific data in a set of structured records which enable easy retrieval of information from any place of the world. Each record, which is also known as entry, should contain a number of fields that hold the actual data items which is stored by individual person from different places. To retrieve a particular record from the database, a user can specify a particular piece of information, called value, to be found in a particular field and expect the computer to retrieve the whole data record. This process is called making a query.

There are two kinds of database; the primary database and secondary database. Primary or archived databases contain information and annotation of DNA and protein sequences, DNA and protein structures and DNA and protein expression profiles. Secondary or derived databases are so called because they contain the results of analysis which are based on the primary resources including information on sequence patterns or motifs, variants and mutations and evolutionary relationships (Rao *et al.*, 2008) (Fig.1).

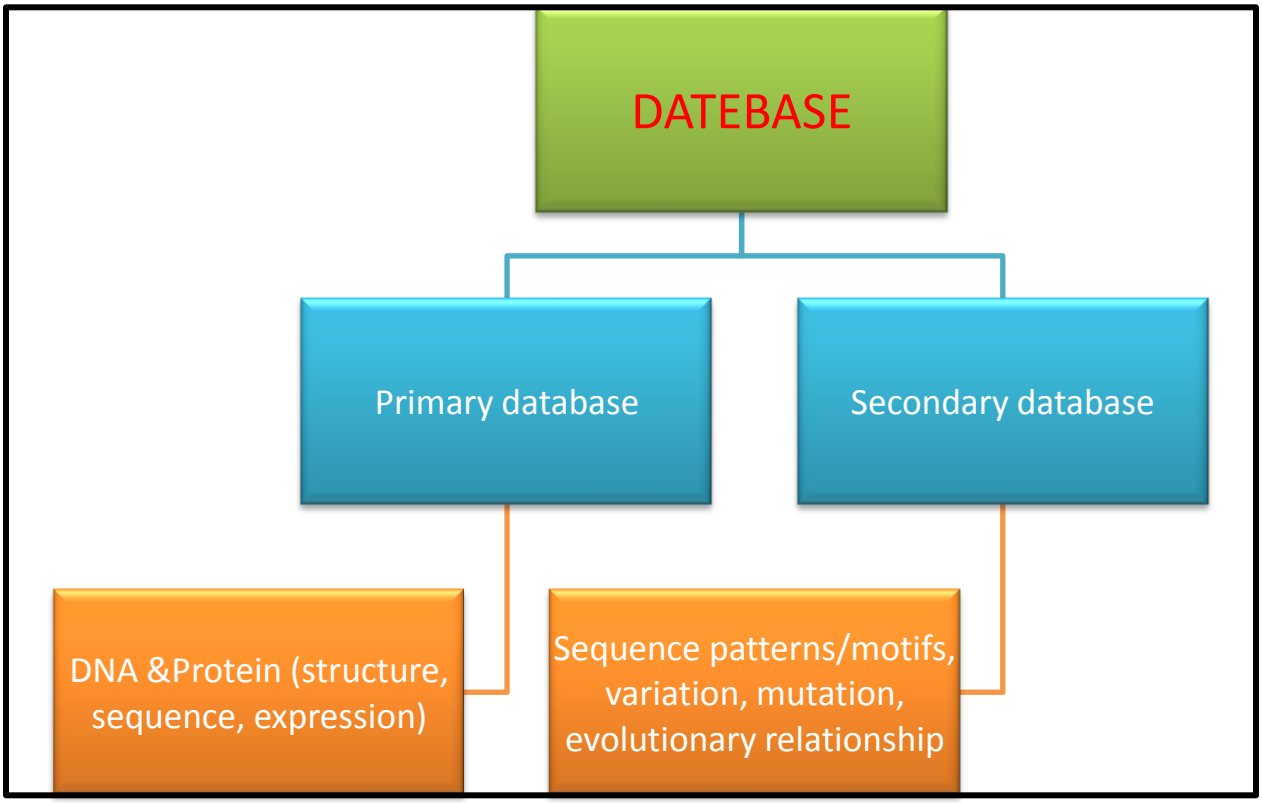

Fig. 1: Types of database

Apart from data retrieval, databases have other purposes too such as biological requirement, known as knowledge discovery, which refers to the identification of connections between pieces of information that were not known when the information was first entered (Xiong, 2009). There are also some other tools which are highly used in bioinformatics which may be categorized into following categories:

1 *Homology and similarity tools*

The term homology means relationship between DNA sequences. Homologous sequences are sequences that are linked by divergence from a common ancestor. Thus, the degree of similarity between two sequences can be measured while their homology is a case of being either true or false. This set of tools can be used to identify connections between novel query sequences of unknown structure and function and database sequences whose structure and function have been elucidated (Kushwaha *et al*., 2017). There are two basic goals for homologous sequence identification, first is to identify homologous sequence and to employ sequence and structure conversation which can help to predict basic biochemical activities and biological functions of various proteins and some non-coding sequences. While second goal is, to get evolutionary reconstruction. For both of these goals, there are two principal types of homologous relationships available which differ in their history of evolution and functional implications. These two were first introduced by Walter Fitch in 1970. Out of them, First one is ***Orthologs***

which is define as evolutionary counterparts derived from a single ancestral gene in the last common ancestor of the given two different species. Second one is ***Paralogs*** which are homologous genes evolved through duplication within the same genome. But these two were remained unknown to molecular biologist until the advent of genomics, after that has become more clear that variation between these two were very pivotal for understanding evolutionary relationships between various genomes and their gene functions. In another term, robust identification of orthologs is indispensable because otherwise any evolutionary scenarios, for example, attempts to reconstruct the gene repertoire and gene order in ancestral genome, are bound to be meaningless. While as per functional analysis, orthologs typically retain the same, ancestral functions in specific species, which makes transfer of functional information within a set of orthologs generally reliable. In opposite of these, paralogs tend to evolve new functions through evolution and study of different paralogus families which may provide means for understanding adaptation. This terminology was first explained by Susumu Ohno in 1970 in his book *Evolution by Gene Duplication.* These duplication helps to reduce the pressure of purification either in original Ohno's model or in new elaborate models, which eventually enables evolution of new functions. In each sequenced genome, a substantial fraction of genes belongs to different families of paralogs which reflects functional diversity due to duplication occurred at various stages of evolution.

2 Computational analysis (software and tools)

After so many years of storage, book-based data need to re-record or transfer into other pages so that can store for longer time. It is not only hardware or book-based record are too hard to store. So different software has been used to align on the basis of user's priority. Such softwares are range from simple command line tools to more complex graphical programs to design drugs or other bio-molecules; standalone web-services available from various bioinformatics companies or public

institutions. Basic Local Alignment Sequence Tool (BLAST) is the one that is used for doing sequence alignment. It is the fastest means by which to identify specific sequences in large datasets in various databases and enables the rapid annotation of novel sequences. Although BLAST is the standard tool for identifying sequence similarities in large data sets, there are several options for assembling sequence datasets, the choice of which depends on hardware availability, dataset size, data format, structure and the genetic structure of the organism (Edwards and Batley, 2004).

Examples of widely used software in bioinformatics

BLAST

The Basic Local Alignment Search Tool (BLAST) use for analysis of gene and protein sequences against others quarry sequence in public databases. BLASTs are also useful for wide ways, such as for different plant species, microbes interact with plants, human, microbial, and other genomes. This database enables a researcher to compare a query sequence with present sequence in library or specific database and identify library sequences which resemble the query sequence above a certain threshold. The BLAST program can either be downloaded and run as a command-line utility "blastall" or accessed for free over the web. The BLAST web server, hosted by the NCBI, allows anyone with a web browser to perform similarity searches against constantly updated databases of proteins and DNA that include most of the newly sequenced organisms.

This program is based on an open-source format, giving everyone freely access to it and enabling them to have the ability to change the program code as per the researcher's requirement. On the basis of various requirement, which has led to the creation of several BLAST "spin-offs".

There are different BLAST programs available on NCBI, which can be used depending on what one is attempting to do and what they are working with. These different programs vary in query sequence input, the database being searched, and what is being compared. Few examples of different BLAST are given as below:

Nucleotide-nucleotide BLAST (blastn)

This program, given a DNA query, returns the most similar DNA sequences from the DNA database that the user specifies.

Protein-protein BLAST (blastp)

This program, given a protein query, returns the most similar protein sequences from the protein database which is user specifies.

Nucleotide 6-frame translation-protein (blastx)

The usage of this program is to compare six-frame conceptual translation products of a nucleotide query sequence (both strands) against a protein sequence database.

Nucleotide 6-frame translation-nucleotide 6-frame translation (tblastx)

Out of all program, this program is the slowest of the BLAST family. It translates the query nucleotide sequence in all six possible frames and compares it against the six-frame translations of a nucleotide sequence database. The purpose of tblastx is to find very distant relationships between nucleotide sequences, which can help to identify the relatedness or distance between two families or various different species.

Protein-nucleotide 6-frame translation (tblastn)

In the contrast of tblastx, this program compares a protein query against the all six reading frames of a nucleotide sequence database.

Position-Specific Iterative BLAST (PSI-BLAST) (blastpgp)

To identify distant relatives of a protein this program is used. First, a list of all closely related proteins is created. A query sequence against the protein database is then run using this profile which helps to find different larger group of protein. Such group is used to construct another profile, and the process is repeated.As compared to all program for proteins in the search, PSI-BLAST is much more sensitive in picking up distant evolutionary relationships than a standard protein-protein BLAST.

Large numbers of query sequences (megablast)

When comparing large numbers, as more then six, of input sequences via the command-line BLAST, "megablast" is much faster as compared to running BLAST multiple times by using single or six sequences. It concatenates many input sequences together to form a large sequence before searching the BLAST database, then post-analyzes the search results to glean individual alignments and statistical values.

FASTA

FASTA also a search database tool which is highly useful after BLAST. This programme used to compare a nucleotide or peptide sequence with a sequence database based on the rapid sequence algorithm. It was the first and most useful algorithm for database similarity searching. It is developed by developed by David J. Lipman and William R. Pearson in 1985 before BLAST. Initially, they developed FASTP program which was designed for protein sequence similarity searching. A major focus of this program is the calculation of accurate similarity statistics, so that biologists can judge whether an alignment is likely to have occurred by chance, or whether it can be used to infer homology.

EMBOSS

EMBOSS (The European Molecular Biology Open Software Suite) is one of the latest and free open source software analysis packages. It is specially developed to analyse the molecular data. It is combination around hundred programs for sequence alignment, data searching with sequence patterns, protein motif identification and domain analysis, nucleotide sequence pattern analysis, codon usage analysis for small genomes, and much more. It is more preferable nowadays than any other programme as one can compare more than one query type at a time. The main feature of this software is that it automatically copes with data in a variety of formats and even allows transparent retrieval of sequence data from the web.

ClustalW

It is a third-generation program of original version of Clustal which was first created by Des Higgins in 1988. Clustal W, released in 1994, is multiple sequence alignment programs for DNA or proteins identification. It generates biological suitable or meaningful multiple sequence alignments of different sequences, calculates the best match for the selected sequences. ClustalW uses progressive alignment method. For example, the most similar sequences with the best alignment score are aligned first. Then progressively more distant groups of sequences are aligned until a global alignment is obtained. This heuristic approach is helps to find the global optimal solution and such progressive alignment is required due to prohibitive in both memory and time.

RasMol

It is a most powerful research tool. It displays the structure of DNA, proteins and smaller molecules. Protein Explorer, a derivative of RasMol. It is an easier to use this program. Using this programme one can predict the structure of protein and on the basis of that function of that protein also can predicted. It is first developed by Roger Sayle in the early 1990s. analysis in RasMol included various features, for example, it

have specific script language for special run-time environment that automates the execution of tasks which helps to perform many functions such as selection specific protein chains, specific site of protein, color change in protein etc. There are some other software, named Jmol and Sirius, are included with RasMol which incorporate into their commands.

Types of analysis

Protein function analysis

Protein function analysis is carried out to identify all functional elements such as coding and non-coding in a genome. The collection of programs allows you to make comparison between your protein sequence to the alternative or secondary protein databases that cover information on motifs, signatures and protein domains. (Kushwaha *et al*., 2017).

Structural analysis

Using structural analysis one can compare unknown structures with the known structure databases as well as check that if the variation in the known structure may lead to some other functional proteins which may have some additional properties. Thus, the function of protein is directly proportional to the structure of protein. Moreover, protein's 2D/3D structure is most important for various protein-based studies (Jacqueline and Edwards, 2017; Kushwaha *et al.,* 2017).

Structure correlations can embody either generic properties, hydrophobic amino acids with burial in the structure, or more specific properties, particular configuration of residues that form a motif associated with a particular structure or function. The former tend to be approached using statistical methods, of which one of the most general is the Artificial Neural Net (ANN), while the latter are associated more with sequence alignment or pattern matching methods. Both approaches can, of course, be applied together to provide mutual support.

Sequence analysis

Sequence analysis is related to DNA or amino acid sequence of protein to sequence alignment, sequence databases, repeated sequence searches, or other bioinformatics methods on a computer. Many sequence analysis tools allow you to analysis on your query sequence such as identification of mutations, evolutionary analysis, and compositional biases. The identification of biological properties is search based on specific function of your sequence (Siddharthan *et al.,* 2016; Kushwaha *et al.,* 2017).

A sequence alignment is a different way of arranging the sequences of DNA, RNA or protein to identify various regions of similarity or dissimilarity that may be in the form of either functional, structural or evolutionary relationships between the sequences. Such data variation also can be within species or other species. Sequence analysis can be useful identify function to genes and proteins by the study of the similarities between the compared sequences (Stormo, 2000). Nowadays there are numbers of tools and techniques that provide the sequence comparisons (sequence alignment) and analyze that alignment product to understand their biology and other characteristics. The most commonly used similarity search method is the Basic Local Alignment Search Tool (BLAST; Altschul *et al*. 1990). BLAST is a heuristic modification of the Smith–Waterman (1981) algorithm and in practice it is widely used. In BLAST, statistical methods are used to determine the likelihood of a particular alignment between sequences or sequence regions arising by chance given the size and composition of the database being searched (Kushwaha *et al.,* 2008; Kushwaha *et al.,* 2010). The most popular online interface to BLAST is available at NCBI, where a

standalone version is also available for downloading. There are several parameters controlling the behaviour of the BLAST algorithm, and these need to be carefully considered. The set of tools allows us to carry out further, more detailed analysis on our query sequence including evolutionary analysis, identification of mutations, hydropathy regions, CpG islands and compositional biases (Puigbo et al., 2007). Sequence alignments are useful in bioinformatics for identifying sequence similarity, producing phylogenetic trees (Yadav *et al.*, 2009; Dubey *et al*., 2010) and developing homology models of protein structures (Tamura *et al.*, 2007).

Applications of bioinformatics tools in field of agriculture

In the field of Agriculture, applications of bioinformatics are covered carious area, out of few are given as following:

1. *Crop improvement*

As per the increasing crop demand as per the population, one either need to grow more crop or reduce the wastage or add some additional characteristics which helps that crop to fulfil the requirement. Different genetics of the plant genomes has shown that the group of their genes has remained more conserved. These data suggest that information obtained from the model crop systems can be used to propose improvements to other food crops. *Arabidopsis thaliana,* due to their low chromosome number and less genome sequence, was the first plant to be sequenced and considered the model species for examining plant genetics and biology. There are many genes which are similar in all plants or conserved in specific species and the study of genes in a model organism such as *A. thaliana* facilitates understanding of gene expression and function in all the plants species. *Oryza sativa* is very essential staple crop for humans which have been used as staple food more than half of the world population (Gu *et al.*, 2011). *Oryza sativa* is also one of that "model organism" due to the prior work has been done on that. It contains various gene which is present in other species of cereals as well as other crops. It has the smallest genome all the cereals: 430 million nucleotides and it can serve as a model genome for one of the two main groups of flowering plants, the monocotyledons (Brozynska *et al.*, 2015; Hao *et al*., 2016; Kushwaha *et al.,* 2017).

2. Insect Resistance

In country like India, where most of cropping is depend on the rainfall condition have their own flows i.e. a greater number of insect due to favorable condition. Genes from *Bacillus thuringiensis* that can control a number of serious pests have been effectively transferred to cotton, maize and potatoes. This new ability of the plants to resist insect attack and these may also help to increase in the nutritional quality of the crops production (Gu *et al*., 2011). *Bacillus thuringiensis* rod-shaped aerobic bacteria in the genus Bacillus. They are applied to leaves where the insect larvae feed. The toxin genes have also been genetically applied into several crop plants to enhance production of crops. Once they entered in the gut of the pest, they produce toxin which cause the damage in the gut of the pest and ultimately kill that pests (Abberton *et al*., 2015).

3. Improve nutritional quality

Value addition is an important part of human history. It is also applicable in agricultural area. Agriculture scientists and researchers have developed various varieties which have some value addition other than their original one. For example, recently researcher has successfully transferred genes into rice to increase levels of Vitamin A, iron and other micronutrients. This work could have profound influence in reducing occurrences of blindness and anaemia affected by deficiencies in Vitamin A and

iron, respectively. They have successfully inserted a gene from yeast into the tomato and the result is a plant whose fruit stays longer on the vine and has an extended shelf life time, flavour-shaver tomato. Picking tomatoes early means they have chance to develop flavour, colour, and nutrients naturally but in the absence of modification they may rotten and cause the yield loss which can be avoid by using flavour-shaver tomato. By modified the rin gene, agricultural scientists will be able to slow the ripening process, letting the tomato develop on the vine longer - but still keeping it firm enough to ship safely. They hope that their technique may also be applied to other fruits - such as strawberries, bananas, bell peppers, and melons - which suffer from the same shipping and storage complications (Mohammadi and Prasanna, 2003; Abberton *et al.* 2015). Some also have added genes for hepatitis B vaccine into banana, so people may escape from such life threatening or fetal diseases. By using recombination technology gene for this disease has been transferred into plant which is expressed when crop is in ripped stage.

Plant-pathogen interaction

As per the Gene-for-Gene hypothesis, plants contain single dominant resistance R genes that specifically recognize pathogen which contains complementary avirulence genes. These genes can be defined as genes which are present in pathogen that encode a protein product which is directly or indirectly only by those plants that contain the complementary R gene for that product. Several Bioinformatical tools are used for identification, characterization and functional analysis of plant genes involved in the triggering, signalling and response to biotic and abiotic responses. Thus, in-silico biology plays an important role to understand the plant pathogen at gene and genome of plants.

Moreover, it also helps to understand the genetic architecture of different microorganisms as well as pathogens and to identify that how this microorganism affect host plant using various approaches such as metagenomics and transcriptomics. With the help of these result researcher can identify which are beneficial microbes and how to increase it's activity on the other hand, how to develop resistance crop for pathogens which are harmful for plants.

REFERENCES

Abberton, M.; Batley, J.; Bentley, A.; Bryant, J.; Cai, H.; Cockram, J. de Oliveira, A.C.; Cseke, L. J.; Dempewolf, H.; De Pace, C.; Edwards, D.; Gepts, P.; Greenland, A.; Hall, A. E.; Henry, R.; Hori, K.; Howe, G. T.; Hughes, S.; Humphreys, M.; Lightfoot, D.; Marshall, A.; Mayes, S.; Nguyen, H. T.; Ogbonnaya, F. C.; Ortiz, R.; Paterson, A. H.; Tuberosa, R.; Valliyodan, B.; Varshney, R.K. and Yano, M. (2015). Global agricultural intensification during climate change: a role for genomics. *Plant Biotech.*, **14**(4):1095-1098.

Ali, Q.; Javed M. T.; Noman, A.; Haider, M. Z.; Waseem, M.; Iqbal, N.; Waseem, M.; Shah, M. S.; Shahzad, F. and Perveen, R. (2017). Assessment of drought tolerance in mung bean cultivars/lines as depicted by the activities of germination enzymes, seedling's antioxidative potential and nutrient acquisition, *Arch. Agron. Soil Sci.*, DOI:10.1080/03650340.2017.1335393

Altschul S. F.; Gish W.; Miller W.; Myers E. W. and Lipman D. J. (1990). Basic local search alignment tools. *J. Mole. Biol.*, **215**: 403-410.

Arabidopsis Interactome Mapping Consortium, 2011. Evidence for

network evolution in an *Arabidopsis* interactome map. *Sci.*, **333**: 601–607.

Batley, J. and Edwards, D. (2016). The application of genomics and bioinformatics to accelerate crop improvement in a changing climate. *Current Opinion in Plant Biology*, **30**: 78-81.

Brozynska, M.; Furtado, A. and Henry, R. J. (2015). Genomics of crop wild relatives: expanding the gene pool for crop improvement. *Plant Biotech*. **14(**4):1075-1085.

Degrave. W.; Huynh, C.; Roos, D.; Oduola, A. and More, C. M. (2002). Bioinformatics for disease endemic countries: opportunities and challenges in science and technology development for health. *IUPA*. pp: 1-7.

Dubey, A. K.; Yadav, S.; Kumar, M.; Singh, V. K.; Sarangi, B. K. and Yadav, D. (2010). In silico characterization of pectatelyase protein sequences from different source organism. *Enzyme Res*., (Article ID 950230).

Edwards, D. and Batley, J. (2004). Plant bioinformatics: from genome to Phenome. *Biotech.,* **22(**5): 232-237.

Gu H.; Zhu, P.; Jiao, Y.; Meng, Y. and Chen, M. (2011). PRIN: A predicted rice interactome network, *BMC Bioinf.,* **12**: 161.

Hao, L.; Zhang, H.; Zhang, Z.; Hu, S. and Xue, Y. (2016). Information commons for rice (IC4R), *Nucleic Acids Res*., **44**(D1): D1172-D1180.

Jayaram, B. and Dhingra, P. (2010). Bioinformatics for a better tomorrow. Indian Institute of Technology, Hauz Khas, New Delhi - 110 016, India.

Kojima, M.; Kamada-Nobusada, T.; Komatsu, H.; Takei, K.; Kuroha, T. and Mizutani, M. (2009). Highly sensitive and high-throughput analysis of plant hormones using MS-probe modification and liquid chromatography tandem mass spectrometry: an application for hormone profiling in *Oryza sativa*. *Plant Cell Physiol*., **50**: 1201–1214.

Kushwaha H.; Gupta N.; Singh V. K.; Kumar A. and Yadav D. (2008). In silico analysis of PCR amplified DOF (DNA binding with one finger) transcription factor domain and cloned genes from cereals and millets. *Online J. Bioinformatics*, **9:** 130-143.

Kushwaha, H.; Gupta, S.; Singh, V. K.; Rastogi, S. and Yadav, D. 2010. Genome wide identification of Dof transcription factor gene family in sorghum and its comparative phylogenetic analysis with rice and Arabidopsis. *Molecular Biology Reporter* (OI: 10.1007/s11033-010-0650-9).

Kushwaha, U.; Deo, I.; Jaiswal, J. P. and Prasad, B. (2017). Role of bioinformatics in crop improvement, *Global Journal of Science Frontier Research: D Agriculture and Veterinary*. **17**: 1.

Lister, R.; Gregory, B. D. and Ecker, J. R. (2009). Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. *Curr. Opin. Plant Biol*., **12**: 107–118.

Mochida, K. and Shinozaki, K. (2010). Genomics and bioinformatics resources for crop improvement. *Plant Cell Physiol*., **51**: 497-523.

Mohammadi, S. A. and Prasanna, B. M. (2003). Analysis of genetic diversity in crop plants salient statistical tools and considerations, *Crop Sci.*, **43**:1235-1248.

Noman, A. and Aqeel, M. (2017). miRNA-based heavy metal homeostasis and plant growth. *Environ Sci Poll Res*., **24**:10068–10082. 10.1007/s11356-017-8593-5

Noman, A.; Fahad, S.; Aqeel, M.; Ali, U.; Ullah, A.; Anwer, S.; Khan, S. and Zainab, M. (2017a). miRNAs: Major modulators for crop growth and development under abiotic stresses. *Biotechnology Letters.* DOI 10.1007/s10529-017-2302-9

Noman, A.; M., Aqeel, J.; Deng, N.; Khalid, T.; Sanaullah, H. and Shuilin, H., (2017b). Biotechnological advancements for improving floral attributes in ornamental plants. *Front Plant Sci*., **8**: 530.

Puigbo, P.; Guzma, E.; Romeu, A. and Garcia-Vallve, A. (2007). OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. *Nucleic Acids Res.,* **35:** W126-W131.

Rao, V. S.; Das, S. K.; Rao, V. J and Srinubabu, G. (2008). Recent developments in life sciences research: Role of bioinformatics. *African J. Biotech.,* **7**(5):495-503.

Shinozaki, K. and Sakakibara, H. (2009). Omics and bioinformatics: an essential toolbox for systems analyses of plant functions beyond 2010. *Plant Cell Physiol*., **50**: 1177–1180.

Siddharthan, N.; Raja Prabu, M. and Sivasankari, B. (2016). Bioinformatics in drug discovery: A review. *Int. J. Res. Arts Sci.,* **2**: 2-20.

Smith, T. F. and Waterman, M. S. (1981). identification of common molecular sequences. *J. Mole. Biol.,* **147**: 195–197.

Stormo, G. D. (2000). DNA binding sites: representation and discovery. *Bioinformatics*, **16**(1): 16-23.

Tamura, K.; Dudley, J.; Nei, M. and Kumar, S. (2007). MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mole.Biol. Evolu.,* **24**: 1596-1599.

Vassilev, D.; Leunissen, J.; Atanassov, A.; Nenov, A. and Dimov, G. (2005). Application of bioinformatics in plant breeding. *Biotech. Biotechnol. Equip.,* **19**(3): 139-152

Xiong, J. (2009). Essential bioinformatics: introduction and biological databases. *Cambridge University press*, USA. http://www.cambidge.org /catalogue.

Yadav, P. K.; Singh, V. K.; Yadav, S.; Yadav, K. D. S. and Yadav, D. (2009). *In-silico* analysis of pectin lyase and pectinase sequences. *Biochemistry* (Mosc), **9**: 1049-1055.

[MS received : September 25 , 2019] ***[MS accepted : September 28, 2019]***