Volume 17 Issue 2
Originally issued as D1

Global Journal of Science Frontier Research: D
Agriculture and Veterinary
Volume 17 Issue 1 Version 1.0 Year 2017
Type : Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 2249-4626 & Print ISSN: 0975-5896

Role of Bioinformatics in Crop Improvement

By Ujjawal Kumar Singh Kushwaha, Indra Deo, Jai Prakash Jaiswal & Birendra Prasad

Govind Ballabh University of Agriculture and Technology

Abstract- Bioinformatics plays a significant role in the development of the agricultural sector, agro-based industries, agricultural by-products utilization and better management of the environment. With the increase of sequencing projects, bioinformatics continues to make considerable progress in biology by providing scientists with access to the genomic information. It is believed that we will take on another giant leap in bioinformatics field in next decade, where computational models of systems wide properties could serve as the basis for experimentation and discovery. Agricultural bioinform -atics areas that need focus would be are data curation and need for the use of restricted vocabularies. Being an interface between modern biology and informatics it involves discovery, development and implementation of computational algorithms and software tools that facilitate an understanding of the biological processes with the goal to serve primarily agriculture and healthcare sectors with several spinoffs.

Keywords: agricultural bioinformatics, crop improve ment, bioinformatics, biology, information-technology.

GJSFR-D Classification : FOR Code: 300299

Strictly as per the compliance and regulations of :

© 2017 Ujjawal Kumar Singh Kushwaha, Indra Deo, Jai Prakash Jaiswal & Birendra Prasad. This is a research/review paper, distributed under the terms of the Creative Commons Attribution-Noncommercial 3.0 Unported License http://creativecommons.org/licenses/by-nc/3.0/), permitting all non commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Role of Bioinformatics in Crop Improvement

Ujjawal Kumar Singh Kushwaha [*α], Indra Deo [σ], Jai Prakash Jaiswal [ρ], Birendra Prasad [ω]

Abstract- Bioinformatics plays a significant role in the development of the agricultural sector, agro-based industries, agricultural by-products utilization and better management of the environment. With the increase of sequencing projects, bioinformatics continues to make considerable progress in biology by providing scientists with access to the genomic information. It is believed that we will take on another giant leap in bioinformatics field in next decade, where computational models of systems wide properties could serve as the basis for experimentation and discovery. Agricultural bioinform -atics areas that need focus would be are data curation and need for the use of restricted vocabularies. Being an interface between modern biology and informatics it involves discovery, development and implementation of computational algorithms and software tools that facilitate an understanding of the biological processes with the goal to serve primarily agriculture and healthcare sectors with several spinoffs. Bioinformatics is more often a tool than a discipline, the tools for analysis of biological data. The ability to represent high resolution physical and genetic maps of plants has been one of the great applications of bioinformatics tools.

Keywords: agricultural bioinformatics, crop improve ment, bioinformatics, biology, information-technology.

I. Introduction

Bioinformatics is a new and emerging science that combines the power of computers, mathematical algorithms, and statistics with concepts in the life sciences to solve biological problems. It is an emerging interdisciplinary area of science and technology encompassing a systematic development and application of information technology solutions to handle biological information by addressing biological data collection and warehousing, data mining, database searches, analyses and interpretation, modeling and product design (Degrave et al., 2002; Xiong, 2009; Jayaram and Dhingra, 2010). Thus bioinformatics is the interdisciplinary science of interpreting biological data using information technology and computer science. Bioinformatics was invented by Paulien Hogeweg and Ben Hesper in 1970 as "the study of informatic processes in biotic systems". Paulien Hogeweg is a Dutch theoretical biologist and complex systems researcher studying biological systems as dynamic information processing systems at many interconnected levels.

*Author *α: Department of Genetics and Plant Breeding Govind Ballabh University of Agriculture and Techno-logy, Pantnagar, India.*
e-mail: kushujjwal@gmail.com
Author σ ρ ω: Department of Genetics and Plant Breeding Govind Ballabh University of Agriculture and Techno-logy, Pantnagar, India.

Over the past few decades, major advances in the field of molecular biology, coupled with advances in genomic technologies, have led to an explosive growth in the biological information generated by the scientific community (Vassilev et al., 2006).

In recent years, rapid developments in genomics and proteomics have generated a large amount of biological data. These data requires sophisticated computational analyses for drawing conclusions. The importance of this new field of inquiry will grow as we continue to generate and integrate large quantities of genomic, proteomic, and other data. 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. Bioinformatics plays an essential role in today's plant breeding with regards to the development of new plant diagnostic tools.

Bioinformatics have huge applications in various fields. Bioinformatics helps scientists to apply different tools and technology to facilitate and analyse the research work. Some major fields where bioinformatics is widely used are as follows.

II. Fields of Bioinformatics

- Microbial genome applications
- Molecular medicine
- Personalised medicine
- Preventative medicine
- Gene therapy
- Drug development
- Antibiotic resistance
- Evolutionary studies
- Waste cleanup
- Biotechnology
- Climate change Studies
- Alternative energy sources
- Crop improvement
- Forensic analysis
- Bio-weapon creation
- Insect resistance
- Improve nutritional quality
- Development of drought resistant varieties
- Veterinary Science

a) Microbial genome applications

- Genome assembly
- Re-sequencing
- Comparative analysis

© 2017 Global Journals Inc. (US)

- Evolutionary studies
- Antibiotic resistance
- Waste cleanup
- Biotechnology

III. Scopes of Bioinformatics

Bioinformatics is an important discipline of biological sciences that allows scientists to decipher and manage the vast quantities of data available to them (Ojo and Maxwell, 2010). It consists of two subfields: the development of computational tools and databases, and the application of these tools and databases in generating biological knowledge to better understand living systems (Xiong, 2009). These two subfields are complementary to each other. The tool development includes writing software for sequence, structural, and functional analysis, as well as the construction and crating of biological databases. These tools are used in three areas of genomic and molecular biological research, including molecular sequence analysis, molecular structural analysis and molecular functional analysis (Tramontano, 2009; Xiong, 2009). The areas of sequence analysis includes sequence alignment, sequence database searching, motif and pattern discovery, gene and promoter finding, reconstruction of evolutionary relationships and genome assembly and comparison. Structural analyses include protein and nucleic acid structure analysis, comparison, classification and prediction. The functional analysis includes gene expression profiling, protein–protein interaction prediction, protein sub cellular localization prediction, metabolic pathway reconstruction and simulation (Rao et al., 2008; Xiong, 2009). The three aspects of bioinformatics analysis are not isolated, but often interact to produce integrated results. For example, protein structure prediction depends on sequence alignment data; clustering of gene expression profiles requires the use of phylogenetic tree construction methods derived in sequence analysis. Sequence based promoter prediction is related to functional analysis of co-expressed genes.

IV. Importances of Bioinformatics

Bioinformatics apply the principles of information sciences and technologies to complex life science data (Ojo and Maxwell, 2010). Therefore, the field of bioinformatics has evolved such that the most pressing task now involves the analysis and interpretation of various types of data, including nucleotide and amino acid sequences, protein domains, and protein structures (Wales, 2009).

Table1: Classification of databases in the 2004 edition of the Molecular Biology Database Collection (11)

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

(Source: D. Vassilev, J. Leunissen, A. Atanassov, A. Nenov & G. Dimov (2005) Application of Bioinformatics in Plant Breeding, Biotechnology & Biotechnological Equipment, 19:sup3, 139-152).

The analysis of genetic and DNA sequences such as nucleic acid sequences, protein structure/function relationships, genome organization, regulation of gene expression, interaction of proteins and mechanisms of physiological functions, can all benefit from a bioinformatics approach. Nucleic acid and protein sequence data from many different species and from population samplings provides a foundation for studies leading to new understandings of evolution and the natural history of life (Tramontano, 2009). The main tasks of Bioinformatics involve the analysis of sequence information that involves identification of genes in the DNA sequences from various organisms, identification of families of related sequences and the development of models, aligning similar sequences and generating phylogenetic trees to examine evolutionary relationships. Finding all the genes and proteins of a genome from a given sequence of amino acids and predicting active sites in the protein structures to attach drug molecules (Rao et al., 2008).

There are three important sub-disciplines within bioinformatics (Kumor 2009 and Asthana 2009), the

© 2017 Global Journals Inc. (US)

development of new algorithms and statistics with which to assess relationships among members of large data sets; the analysis and interpretation of various types of data including nucleotide and amino acid sequences, protein domains, and protein structures; and the development and implementation of tools that enable efficient access and management of different types of information (Rao et al., 2008).

Table 2: List of plant species in which partial or whole genomes have been sequenced. (Data extracted from the following internet sites: http://www.ncbi.nlm.nih.gov/genomes/PLANTS/PlantList.html; http://www.arabidopsis.org/portals/genAnnotation/other_genomes/index.jsp; http://www.ildis.org/)

Division	Class	Species
Non Vascular	Algae	Chlamydomonas reinhardtii, Chlorella variabilis, Cocco-myxa sp., Cyanidioschyzon merolae, Ectocarpus siliculosus, Micromonas pusilla, Micromonas sp., Ostreococcus, Lucimarinus, Ostreococcus tauri, Volvox carteri, Zostera marina
Vascular	Bryophytes	Physcomitrella patens, Selaginella moellendorffii
	Dicotyledons	Amborella trichopoda, Aquilegia sp., Arabidopsis lyrata, Arabidopsis thaliana, Arachis hypogaea, Asclepias syriaca, Beta vulgaris, Boechera holboellii, Brassica napus, Brassica napa, Brassica rapa, Caffea caneophora, Cajanus Cajan, Cannabis sativa, Capsella rubella, Carica papaya, Castanea mollissima, Citrullus lanatus, Citrus Clementine, Corchorus olitorius, Cucumis sativus, Eucalyptus grandis, Fragaria vesca, Glycine max, Gossypium hirsutum, Gossypium raimonddi, Hordeum vulgare, Jatropha curcas, Lactuca sativa, Linum usitatissinun, Lotus japonicas, Malus domestica, Manihot esculenta, Medicago truncatula, Mimulus guttatus, Phaseolus vulgaris, Pinus taeda, Populus tremula, Ricinus communis, Theobroma cacao, Populus nigra, Populus trichocarpa, Prunus avium, Prunus persica, Pyrus bretschneideri, Rubus idaeus, Salix purpure, Solanum lycopersicum, Solanum pimpinellifolium, Solanum tuberosum, Spirodella polyrhiza, Thellungiella Parvula, Vitis vinifera
	Monocotyledons	Brachypodium distachyon, Elaeis guineensis, Miscanthus Giganteus, Musa acuminate malaccensis, Oryza sativa, Oryza glaberrima, Panicum hallii, Panicum virgatum, Phoenix dactylifera, Seratia italic, Sorghum bicolor, Triticum aestivum, Zea mays

V. Bioinformatics Tool

a) *Biological databases*

Biological databases are archives of consistent data that are stored in a uniform and efficient manner. These databases contain data from a broad spectrum of molecular biology areas. A simple database might be a single file containing many records, each of which includes the same set of information (Xiong, 2009). Databases are composed of computer hardware and software for data management. The chief objective of the development of a database is to organize data in a set of structured records to enable easy retrieval of information. Each record, also called an entry, should contain a number of fields that hold the actual data items. 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 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 on the primary resources including information on sequence patterns or motifs, variants and mutations and evolutionary relationships (Rao et al., 2008).

Although data retrieval is the main purpose of all databases, biological have a higher level of 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).

b) *Software and Tools*

Bioinformatics uses different software that range from simple command line tools to more complex graphical programs and stand alone web-services available from various bioinformatics companies or public institutions. BLAST (Basic Local Alignment Sequence Tool) is the one that is used for doing sequence alignment. It remains the fastest means by which to identify specific sequences in large datasets and enables the rapid annotation of novel sequences. Although BLAST is the standard tool for identifying sequence similarities in large datasets, there are several options for assembling sequence datasets, the choice

© 2017 Global Journals Inc. (US)

of which depends on hardware availability, dataset size, data format, structure and the genetic structure of the organism (Edwards and Batley, 2004).

VI. Dealing Data With Bioinformatics

Dealing data is an important aspect of Bioinformatics. Bioinformatics deals with the following important biological data:

a) DNA, RNA and Protein Sequences

The sequence of nucleotides in DNA or RNA, and the sequence of amino acids in a protein, can be obtained through laboratory sequencing methods.

b) Molecular Structures

Higher molecular structure can be obtained by combining thermodynamic data and computer modeling with measurements from laboratory techniques, such as x-ray diffraction and nuclear magnetic resonance imaging.

c) Expression Data

Scientists use microarrays in the laboratory to determine when and where genes are expressed. Such microarrays can also measure overall gene expressions in certain cell types, or in specific environmental conditions.

d) Bibliographic Data

The number of scientific articles has increased dramatically in the last few decades, due to the increasing number of research projects and genome sequencing programs.

VII. Uses of Data in Bioinformatics

The first step to making sense of all the biological sequences and structures is to formulate a method to manage the data, as well as how to process and maintain it. Data management is the first and most fundamental task of bioinformatics, and bioinform - aticians do this by assembling information into databases. A database is a collection of information stored in a systematic way. Bioinformatics database may consist of DNA sequences, RNA sequences, or even protein sequences. These sequences may be organized according to their function, or according to the species from which they came, or the journal articles which reported them first. A database may also contain journal articles and abstracts. With the data assembled, bioinformaticians can find means by which to mine, retrieve, and use the data. This is usually done through computer programs, which can search databases and retrieve information, depending on a scientist's needs.

Table 3: List of some published plant genome

Species Name	Size (~Mb)*	#of Chr**
Arabidopsis thaliana (mouse ear cress)	115	5
Bracypodium distachyon	355	5
Brassica rapa (Chinese cabbage)	284	10
Cajanus cajan(pigeonpea)	883	11
Carica papaya (papaya)	372	9
Cucumis sativus (cucumber)	203	-
Fragaria vesca(woodland strawberry)	240	7
Glycine max (soybean)	975	20
Medicago truncatula (barrel medic)	241	8
Malus × domestica(apple)	881.3	-
Oryza sativa (rice, japonica)	372	12
Panicum virgatum (switchgrass)	1,230	-
Populus trichocarpa (poplar)	422.9	19
Ricinus communis (castor bean)	400	-
Pinus taeda(loblolly pine)	22,180	27
Solanum tuberosum(potato)	800	12
Sorghum bicolor (sorghum)	730	10
Theobroma cacao(cacao)	346	-
Vitis vinifera (grapevine)	487	19

(Source: Agarwal R, Narayan J (2015) Unraveling the Impact of Bioinformatics and Omics in Agriculture. Int J Plant Biol Res 3(2): 1039)

VIII. Applications of Bioinformatics in Agriculture

Bioinformatics is widely applied in agricultural research. Since agricultural data are of different types and huge in collection, its interpretation is difficult; thus Bioinformatics play big role to analyze the data properly. 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.

© 2017 Global Journals Inc. (US)

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

Plant based biomass is one of the best resource for obtaining energy by converting it into biofuels such as ethanol which could be used to drive the vehicles and fly the planes. 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. Recently, genome of *eucalyptus grandis* has been released which is also one of major resource of biomass components and all the genes take part in conversion of sugars into biomass components have already been deciphered, therefore bioinformatics provides great insight into mechanisms and pathways responsible for this conversion so that in future we can enhance production of biomass components in eucalyptus and other relevant plants (Bisby et al 1993). 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, Betz et al 2000).

c) Insect resistance

Bacillus thuringiensis 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 Soils 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). Data obtained from such intensive research are huge which are difficult to analyse by a single scientist. Bioinformatics help in a greater amount to solve such problems.

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.

An omics data can now be envisioned as a highly important tool for plant improvement. The ability to examine gene expression allows us to understand how plants respond to and interact with the internal and external stimuli. These data may become crucial tool of future breeding decision management systems (Langridge 2011).

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 (Berg 2009, Schenk 2012).

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. However, relating genomic data to climate related agronomic traits for use in breeding remains a huge challenge, and one which will require coordination of diverse skills and expertise. Bioinformatics, when combined with genomics has the potential to help maintain food security in the face of climate change through the accelerated production of climate ready crops (Batley and Edwards 2016).

i) Bioinformatics in Plant Disease Management

Pathogen trait is considered as a primary interest of plant bioinformatics. The contribution of

© 2017 Global Journals Inc. (US)

bioinformatics advances made possible the mapping of the entire genomes of many organisms in just over a decade. 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. Furthermore, many future bioinformatics innovations will likely be spurred by the data and analysis demands of the life sciences. Bioinformatics have many practical applications in current plant disease management with respect to the study of host-pathogen interactions, understanding the disease genetics and pathogencity factor of a pathogen which ultimately help in designing best management options.

IX. Bioinformatics Impacts on Plant Science

Bioinformatics help to deal with the vast amount of data being generated by the genome sequencing projects. Powerful bioinformatics tools are needed to organize the data and to extend our ability to analyze these complex biological systems. Universally the development of bioinformatics is tightly linked with international collaboration in genome sequencing projects and the pharmaceutical industry in its drive for drug discovery and development. Although in infancy, bioinformatics is already well established. For example, for study of genome regulation and structure, bioinformatics covers many topics including: databases on regulatory sequences; the regulation of gene expression; analysis and recognition of genomic sequences; gene structure prediction; modeling of transcriptional and translational control; and large scale genome analysis. The integration of bioinformatics will influence plant science and lead to crop improvements in the following areas:

- The identification of important genes through genomics, expression analysis and functional genomics. In conjunction with the design and construction of transgenic plants this will allow new target genes to be identified that will improve quantitative and qualitative traits of commercially important crops.
- The design of agrochemicals based on an analysis of the components of signal perception and transduction pathways to select targets, and with cheminformatics, to identify potential compounds that can be used as herbicides, pesticides, or insecticides.
- The utilization of plant genetic resources to preserve genetic diversity in agricultural species. The need for taxonomic data goes far beyond the field of classical taxonomy, and a catalogue of all species, with phenotypic and genotypic attributes is required. The core taxonomic effort gives stability to the work of regulatory, management and conservation bodies. Efficient utilization of biological repositories of clones, cell lines, organisms and seeds. Typically, existing repositories are not linked to each other databases. The commercial databases and repositories are also part of the bioinformatics infrastructure but operate largely outside of the present day cooperative activities.

Table 4: List of small RNA databases and tools in plants

Database and Tools	URL	Purpose
Arabidopsis small RNA Project	http://asrp.danforthcenter.org/	Provide entry to small RNA data and resources from the Carrington laboratory
Arabidopsis small RNA database	http://mpss.udel.edu/at_sRNA/	Provide sequencing by synthesis based small RNA data
Cereal Small RNAs Database	http://sundarlab.ucdavis.edu/smrnas/	Platform for providing cereals small RNA data and tools for finding small RNAs and their targets
Plant Small RNA Target Analysis Server	http://plantgrn.noble .org/psRNATarget/	It will perform high throughput analysis of next generation data to give a putative list of miRNA and their target pairs
Plant mRNA database (PMRD)	http://bioinformatics.cau.edu.cn/PMRD/	PMRD integrates the available large information of plant mRNA data consisting of mrNA sequence and their target genes, secondary dimension structure, expression profiling, genome browser etc.
UEA snRNA Toolkit	http://srna.tools.cmp.uea.ac.uk/plant/cgi-bin/srna-tools.cgi	This site provide links to tools for the analysis of high through put small RNA data
miRNA Precusor candidates for Arabidopsis thaliana	http://sundarlab.ucdavis.edu/mirna/	It is collection of predicted miRNA and precursor candidates for the Arabidopsis genome predicted by “find MiRNA” method
PsRobot:Plant Small RNA Analysis Toolbox	http://omicslab.genetics.ac.cn/psRobot/	It can identify stem loop shaped smRNAs and their targets

© 2017 Global Journals Inc. (US)

PLncDB: Plant Long noncoding RNA Database	http://chualab.rockefeller.edu/gbrowse2/homepage	Repository of Arabidopsis long non coding RNA
MiSolRNA	http://misolrna.org/about	Provide tomato miRNA data
Phytophthora small RNA Database	http://phytophthora-smallrna-db.cgrb.oregonstste.edu/	Provide access to small RNA data of Phytophthora infestans and P. sojae
miRNA Tools2.0	http://sites.google.com/sites/mirnatools2.0	Provide list of tools to investigate miRNA and their regulatory action

(Source: Agarwal R, Narayan J (2015) Unraveling the Impact of Bioinformatics and Omics in Agriculture. Int J Plant Biol Res 3(2): 1039).

X. Recent Activities in Bioinformatics to Improve Plant Biotechnology

Knowing the complete sequence of a plant's genome can pave the way for all future studies of that organism. For instance, scientists at the United States Department of Agriculture's Agricultural Research Service (USDA-ARS) are now analyzing gene expression patterns in crops such as soybean and barley, in order to determine the function of genes involved in the resistance of plants to environmental stress (www.usda.com/ars). The International Rice Research Institute, Philippines, is working on the complete genome of rice. Brazilian scientists have already completed the gene sequence of *Xylella fastidiosa*, a plant pathogen that infects citrus plants.

Table 5: Examples of genomic databases related to crop improvement

Database Name	Web Link
autoSNPdb	http://autosnpdb.appliedbioinformatics.com.au/
Brachypodiumdatabase	http://www.brachypodium.org/
Brassica rapagenome database	http://brassicadb.org/
DNA Data Bank of Japan (DDBJ)	http://ddbj.sakura.ne.jp/
European bioinformatics institute, EnsEMBL plants	http://plants.ensembl.org/
European Molecular Biology Laboratory (EMBL) nucleotide sequence database	http://www.ebi.ac.uk/embl/
GenBank	http://www.ncbi.nlm.nih.gov/genbank/
Graingenes	http://wheat.pw.usda.gov/
Gramene	http://www.gramene.org/
International Crop Information System (ICIS)	http://www.icis.cgiar.org
International Nucleotide Sequence Database Collaboration (INSDC)	http://www.insdc.org/
Legume Information System (LIS)	http://www.comparative-legumes.org/
MaizeGDB	http://www.maizegdb.org/
Maize sequence database	http://www.maizesequence.org/
Oryzabase	http://www.shigen.nig.ac.jp/rice/oryzabase/
Panzea	http://www.panzea.org/
Phytozome	http://www.phytozome.net/
PlantsDB	http://mips.helmholtz muenchen.de/plant/genomes.jsp
PlantGDB	http://www.plantgdb.org/
The Plant Ontology	http://www.plantontology.org/
Plaza	http://bioinformatics.psb.ugent.be/plaza/
Rice Genome Annotation Project	http://rice.plantbiology.msu.edu/
SSR Primer	http://flora.acpfg.com.au/ssrprimer2/
SSR taxonomy tree	http://appliedbioinformatics.com.au/projects/ssrtaxonomy/php/
SOL Genomics Network (SGN)	http://solgenomics.net/

© 2017 Global Journals Inc. (US)

SoyBase	http://soybase.org/
TAGdb	http://flora.acpfg.com.au/tagdb/
The Crop Expressed Sequence Tag database, CR-EST	http://pgrc.ipk-gatersleben.de/cr-est/
The Triticeae Repeat Sequence Database (TREP)	http://wheat.pw.usda.gov/ITMI/Repeats/
Wheat genome information	http://www.wheatgenome.info

(Source: Lai Kaitao, MichaŁT. Lorenc and David Edwards (2012) Genomic Databases for Crop Improvement, Agronomy 2012, 2, 62-73; doi:10.3390/agronomy2010062)

The worldwide Potato Genome Sequencing Consortium, led by the Netherlands Genomics Initiative and the Wageningen University and Research Center is another example. Teams from countries such as Brazil, Chile, Russia, India, China, Peru, and New Zealand are working together to sequence all 840 million base pairs of DNA on potato's 12 chromosomes. All this data may be used by scientists to improve potato, which is the world's fourth most important crop. It can be used to design better plants. Once the genes responsible for certain plant traits are known, scientists can identify the basis for disease resistance and stress tolerance, and thus design methods by which plants can be made hardier and more resilient. Scientists also use bioinformatics to help them design plants with higher quality fruit, or with the ability to survive in extreme environmental conditions.

Australia's Queensland Agricultural Biotechnology Center is studying papaya, an important food crop in the tropics, where it is also used in the cosmetics and pharmaceutical industries. To identify the genes involved in papaya ripening, researchers looked at expressed sequence tags (EST) of the fruit's genome. ESTs are short DNA sequences of expressed genes which have been used as a tool for rapid gene discovery. Researchers were able to pinpoint genes that were highly expressed during the ripening process; once these genes are localized, scientists can produce better papayas which may ripen later, or taste better.

- It can be used to harness genetic diversity

By knowing which plants are closely related, scientists can figure out which sexually compatible species have desirable characteristics (such as longer stalks for rice plants, or larger grains for barley, corn, or wheat). The wild relatives of today's plants may be sources of crop improvement genes. Scientists at the University of Wisconsin, are seeking to improve potatoes by studying the genomes of wild potato species. Researchers at the Weizmann Institute in Israel, on the other hand, are working on understanding the process of gene exchange between crop plants and their wild ancestors, in order to use these processes to incorporate desirable genes from wild relatives into important crop plants.

- It can be used to design new tools to study gene function

Scientists first discovered micro-RNAs (mRNAs), a family of gene sequences in plants. These small RNA molecules control various aspects of plant growth and development. They target traits, DNA sequences, keep certain genes from being active. Mutations in miRNAs can cause faulty floral development, or even plant death. miRNA molecules can be designed to silence whole gene families. As a result, scientists are turning to miRNA technology to develop the next generation of plants. Several projects are now underway in the University of California, Riverside and the Whitehead Institute to predict and identify miRNA families in important crops such as rice.

- It can be used to test, analyze, and identify plants

With more and more microarray profiles online, scientists can learn about and exchange information concerning differences in gene expression. They can also test plants for differences in gene expression or protein profiles under different stress conditions, such as drought, disease, or insect infestation. If certain genes are expressed in high amounts during these stress conditions, then they may hold the key to a plant's survival under stress and they may be used to improve other plants that may not have the same gene. To test if GM plants are comparable to their conventional counterparts, scientists carry out protein or RNA profiling. In a recent research, scientists compared GM potato to conventional potato by analyzing the crops proteome, and found that there were no new proteins unique to individual GM lines. Scientists from the Danish Institute of Agricultural Sciences used microarrays, as well as analysis software, to compare gene expression profiles of transgenic and wild type wheat. They found that there were no significant differences in gene expression in the two wheat types.

XI. Selected Internet Resources for Plant Bioinformatics

1. Arabidopsis Genome Initiative (AGI) http://ge no me- www.stanford.edu/Arabidopsis/AGI
2. Arabidopsis Genome Data Analysis, Cold Spring Harbor Laboratory http://nucleus.cshl.org/protarab
3. Plant Genome and Information Center, USDA http://www.nal.usda.gov/pgdic
4. UK Crop Plant Bioinformatics Network http:// synteny.nott.ac.uk/agr/agr.html
5. The Institute for Genomic Research (TIGR) Database http://www.tigr.org

© 2017 Global Journals Inc. (US)

6. Arabidopsis Genome Center at the University of Pennsylvania http://genome.bio.upenn.edu/ATGCUP.html
7. The Genome Sequencing Center of Washington University in St Louis http://genome.wustl.edu/gsc
8. Grain Genes Database http://wheat.pw.usda.gov Maize Genome Database http://www.agron.missouri.edu
9. ArabidopsisInternal Coding Exon Finder http://clio.cshl.org/genefinder/ARAB/arab.htm
10. NetPlantGene V2.0 Web Prediction Server http://www.cbs.dtu.dk/NetPlantGene.html

XII. Future Perspectives

With the increase of sequencing projects, bioinformatics continues to make considerable progress in biology by providing scientists with access to the genomic information. With cloud based service on internet, scientists are now able to freely access volumes of such biological information, which enables the advancement of scientific discoveries in agriculture. The field of biology has undergone several rounds of transformation in the approaches taken, ranging from theoretical to experimental perturbation to discovering molecular components. In the next decades to come, it is believed that we will take on another giant leap in bioinformatics field, where computational models of systems wide properties could serve as the basis for experimentation and discovery. The ramifications of this will be not only the precise understanding of how plant specified traits, to discover the causality of diseases, and to predict their responses to changes in the environment. This could lead to prevention and targeted treatment of diseases, improved food production, and preservation of the environment.

Some of the areas in agricultural bioinformatics that need focus would be are data curation and need for the use of restricted vocabularies. Editing scientific data is important for dissemination of information and therefore highly curated datasets need to be continuously developed through analysis by expert in the field and the results be provided for public use. There is a need for research communities to collaborate and share controlled vocabularies. The efforts of the plant ontology (PO) and gene ontology (GO) consortia would help in a uniform implementation of restricted vocabulary databases. Efforts are being carried out internationally to link existing related databases around the whole world (Kilian et al 1995). This will enable instantaneous transfer of knowledge and information on agriculturally related matters and practices. The linking of agricultural information resources would be helpful.

Being an interface between modern biology and informatics it involves discovery, development and implementation of computational algorithms and software tools that facilitate an understanding of the biological processes with the goal to serve primarily agriculture and healthcare sectors with several spinoffs (Jayaram and Dhingra, 2010). Bioinformatics is more of a tool than a discipline, the tools for analysis of biological data (Kumor, 2009).

XIII. Summaries

Bioinformatics is now playing a significant role in the development of the agricultural sector, agro-based industries, agricultural by-products utilization and better management of the environment. Genomics including sequencing of the model plant and plant pathogen genome has progressed rapidly and opened several opportunities for genetic improvement of crop plants. 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. For instance, the barley stem rust resistance gene has been identified from rice-barley comparisons and the sugarcane rust resistance gene based on maize-sorghum comparisons. Comparative genomics along with bioinformatics could help in achieving improvement of yields in rice, maize, and other related grass crops such as barley, rye, sugarcane and wheat. The ability to represent high resolution physical and genetic maps of plants has been one of the great applications of bioinformatics tools.

References Références Referencias

1. Anonymous, (2005). Pseudomonas versus Arabidopsis: models for genomic research into plant disease resistance.www.actionbioscience.org.
2. Asthana, V. (2009). Bioinformatics at Present: Description and disciplines in bioinformatics.http://bionet.awardspace.info/index.htm
3. Batley Jacqueline and David Edwards (2016) The application of genomics and bioinformatics to accelerate crop improvement in a changing climate. Current Opinion in Plant Biology 2016, 30:78–81
4. Berg G. (2009). Plant-microbe interactions promoting plant growth and health: perspectives for controlled use of microorganisms in agriculture. Appl Microbiol Biotechnol, 84: 11-18.
5. Blum A. (1988). Plant breeding for stress environments1988: CRC Press, Inc.
6. Boyle G. Renewable energy (2004). OXFORD university press. Turner JA. A realizable renewable energy future Science, 285: 687-689.

© 2017 Global Journals Inc. (US)

7. 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.
8. Edwards, D. and Batley, J. (2004). Plant bioinformatics: from genome to Phenome. Biotechnology22 (5): 232-237.
9. Emmersen, J., Rudd, S. Mewes, H.W. and Tetko, I. V. (2007). Separation of sequences from host–pathogen interface using triplet nucleotide frequencies. Fungal Genetics and Biology 44:231-241.
10. Field D, Tiwari B, Snape J. (2005). Bioinformatics and data management support for environmental genomics. PLoS Biol., 3: e297.
11. 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.
12. Hermosa, M.R., Turra, D., Fogliano, V., Monte, E. and Lorito, M. (2006). Identification and characterization of potato protease inhibitors able to inhibit pathogenicity and growth of Botrytis cinerea. Physiological and Molecular Plant Pathology 68: 138–148.
13. Jayaram, B. and Dhingra, P. (2010). Bioinformatics for a better tomorrow. Indian Institute of Technology, HauzKhas, New Delhi - 110016, India.
14. Kasuga M, Liu Q, Miura S, Yamaguchi-Shinozaki K, Shinozaki K. Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-indu.
15. Kilian A, Kudrna DA, Kleinhofs A, Yano M, Kurata N, Steffenson B, et al. (1995). Rice-barley synteny and its application to saturation mapping of the barley Rpg1 region. Nucleic Acids Res., 23: 2729-2733.
16. Koltai,H. and Volpin, H. (2003). Agricultural genomics: an approach to plant protection. European journal of plant pathology 109: 101-108.
17. Kruglyak, L. (1999). Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nature Genetics, 22: 139-144.
18. Kumor, S. (2009). Introduction to bioinformatics, scope and potential of bioinformatics. http://www.bioinformaticsweb.org/catagory/bioinformatics.html.
19. Lindeberg, M., Stavrinides, J., Chang, J.H., Alfano, J. R., Collmer, A., Dangl, J. L., Greenberg, J. T., Mansfield, J.W., and Guttman, D.S. (2005). Proposed guidelinesfor a unified nomenclature and phylogenetic analysis of type III Hop effector proteins in the plant pathogen Pseudomonas syringae. Mol. Plant-Microbe Interact. 18:275-282.
20. Luscombe, N. M., Greenbaum, D. and Gerstein, M. (2001). What is Bioinformatics? A Proposed Definition and Overview of the Field. New Haven, USA.
21. Michelmore, R. W. (2003). The impact zone: genomicsand breeding for durable disease resistance. Current Opinion in Plant Biology6:397–404.
22. Middleton, H, (2000). From gene-specific tests to pharmacogenetics.Community Genetetics3: 198-203.
23. Mochida K, Shinozaki K. (2010). Genomics and bioinformatics resources for crop improvement. Plant Cell Physiol., 51: 497-523.
24. Nilges, M. and Linge, J.P. (2009). Bio–InformatiqueStructurale. Institut Pasteur, F–75015 Paris, France.
25. 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.
26. 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.
27. Proost S, Van Bel M, Sterck L, Billiau K, Van Parys T, Van de Peer Y, et al. PLAZA (2009). A comparative genomics resource to study gene and genome evolution in plants. Plant Cell., 21: 3718-3731.
28. Rao, V. S., Das, S.K.., Rao, V.J and Srinubabu, G. (2008). Recent developments in life sciences research: Role of Bioinformatics. African Journal of Biotechnology7 (5): 495-503.
29. Roca, W., Espinoza, C. and Panta, A. (2004). Agricultural Applications of Biotechnology and the Potential for Biodiversity Valorization in Latin America and the Caribbean. Journal of Agro Biotechnology Management and Economics 7(1&2): 13-22.
30. Schenk PM, Carvalhais LC, Kazan K. (2012). Unraveling plant-microbe interactions: can multi-species transcriptomics help? Trends Biotechnol, 30: 177-184.
31. Tramontano, A.L. (2009). Bioinformatics. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester.DOI: 10.1002/9780470015902.a0001900.pub2.
32. Vassilev. D., Leunissen, J.A., Atanassov, A., Nenov, A. and Dimov, G. (2005). Application of bioinformatics in plant breeding. Wageningen University, Netherland.
33. Vassilev.D, Nenov, A. Atanassov, A. Dimov, G. and Getov, L. (2006). Application of bioinformatics infruit plant breeding. Journal of fruit and ornamental plant research14:145-162.
34. Vencato, M., Tian, F. Alfano, J. R. Buell, C. R. Cartinhour, S. DeClerck, G. A. Guttman, D. S. Stavrinides, J. Joardar, V. Lindeberg, M. Bronstein, P. A. Mansfield, J. W. Myers, C. R. Collmer, A. and

© 2017 Global Journals Inc. (US)

Schneider, D. J. (2006). Bioinformatics-Enabled Identification ofthe HrpLRegulon and Type III Secretion System Effector Proteins of Pseudomonas syringaepv. phaseolicola1448A. MPMI 19(11): 1193–1206.

35. Wales, J. (2009). Bioinformatics http://en.wikipedia.org/wiki/Bioinformatics#Web_services_in_bioinformatics.
36. Wang S, Wan C, Wang Y, Chen H, Zhou Z, Fu H et al. (2004). The characteristics of Na< sup>+, K< sup>+ and free proline distribution in several drought-resistant plants of the Alxa Desert, China. Journal of arid environments., 56: 525-539.
37. Willis, D.K., and Kinscherf, T.G. (2004). Global regulation in Pseudomonas syringae. Pages 223-238 in: Pseudomonas, Vol. 2. J. L. Ramos, ed. Kluwer Academic/Plenum Publishers, New York, U.S.A.
38. Xiong, J. (2009). Essential bioinformatics: introduct - ion and biological databases. Cambridge University press, USA. http://www.cambidge.org /catalogue.
39. Xu Y. (2010). Molecular plant breeding CABI.Langridge P, Fleury D. Making the most of 'omics' for crop breeding. Trends Biotechnol. 2011; 29: 33-40.

© 2017 Global Journals Inc. (US)

Global Journal of Science Frontier Research (D) Volume XVII Issue I Version I Year 2017
24

This page is intentionally left blank

© 2017 Global Journals Inc. (US)