Journal of Natural Sciences Research
ISSN 2224-3186 (Paper) ISSN 2225-0921 (Online)
Vol.15, No.1, 2024
www.iiste.org

Importance of Genomics and Bioinformatics Resources for Plant Breeding

Werkissa Yali
Chiro National Sorghum Research and Training Centre, P.O.BOX 190, Chiro, Ethiopia
Email: workissayali@gmail.com

Abstract

With the worldwide human population growing rapidly, agricultural production must increase to satisfy crop demand. Improving crops through breeding may be a sustainable approach to extend yield and yield stability without intensifying the utilization of fertilizers and pesticides. Current advances in genomics and bioinformatics provide opportunities for accelerating crop improvement. The increase of third-generation sequencing technologies helps overcome challenges in plant genome assembly caused by polyploidy and frequent repetitive elements. Recent remarkable innovations in platforms for omics-based research and application development provide crucial resources to market research in model and applied plant species. A combinatorial approach using multiple omics platforms and integration of their outcomes is now an efficient strategy for clarifying molecular systems integral to improving plant productivity. Furthermore, crop databases that integrate the growing volume of genotype and phenotype data provide a valuable resource for breeders and a chance for data processing approaches to uncover novel trait-associated candidate genes. As knowledge of crop genetics expands, genomic selection and genome editing hold promise for breeding disease-resistant and stress-tolerant crops with high yields. Furthermore, the promotion of comparative genomics among model and applied plants allows us to understand the biological properties of every species and to accelerate gene discovery and functional analyses of genes. Bioinformatics platforms and their associated databases also are essential for the effective design of approaches making the simplest use of genomic resources, including resource integration. This paper reviews recent advances in research platforms and resources in plant omics alongside related databases and advances in technology.

Keywords: breeding, crops, genomics, bioinformatics, database, omicsresource
DOI: 10.7176/JNSR/15-1-02
Publication date: January 31st 2024

1. INTRODUCTION

Humans depend upon crops for over two thirds of their daily energy intake (Ulijaszek, 1991). Because the global human population grows, agriculture (crop cultivation) is under increasing pressure to supply higher crop yields (Tilman, 2011). Additionally global climate change, limited availability of land and water shortages are posing further agricultural challenges. To extend crop yields while reducing the environmental impact of agriculture, genomics is accelerating crop breeding by helping systematically leverage the genetic components of agronomic traits (Abberton *et al.,* 2016). Crop genome sequences provide a crucial foundation for identifying agronomically relevant variation. During the last decade, the decreasing cost of DNA sequencing has led to a rapid rise within the size of crop genomic data, which represents a considerable opportunity for breeders (Edwards *et al.,* 2013).

Although plant genome assembly (generating a genome sequence from fragmented sequencing reads) remains hampered by frequent long repetitive regions, large genome sizes and frequent polyploidy, advances in sequencing technologies and bioinformatics tools have allowed rapid progress since the sequencing and assembly of the rice genome (IRGSP, 2005). Sustainable agricultural production is an urgent issue in response to global climate change and population increase (Brown and Funk, 2008). Furthermore, the recent increased demand for biofuel crops has created a replacement marketplace for agricultural commodities. One potential solution is to extend plant yield by designing plants that supported a molecular understanding of gene function and the regulatory networks involved in stress tolerance, development and growth (Takeda and Matsuoka, 2008). Recent progress in plant genomics has allowed us to get and isolate important genes and to research functions that regulate yields and tolerance to environmental stress.

Bioinformatics may be a new and emerging science that mixes the facility of computers, mathematical algorithms, and statistics with concepts within the life sciences to unravel biological problems. it's an emerging interdisciplinary area of science and technology encompassing a scientific development and application of data technology solutions to handle biological information by addressing biological data collection and warehousing, data processing, database searches, analyses and interpretation, modeling and merchandise design (Xiong, 2009; Jayaram and Dhingra, 2010). Thus bioinformatics is that the interdisciplinary science of interpreting biological data using information technology and computing. Bioinformatics was invented by Paulien Hogeweg and Ben Hesper in 1970 as "the study of informatics processes in biotic systems". Paulien Hogeweg may be a Dutch theoretical biologist and sophisticated systems researcher studying biological systems as dynamic information science systems at many interconnected levels.

The whole-genome sequencing of Arabidopsis thaliana was completed in 2000 (The Arabidopsis Genome Initiative 2000). Subsequently, the National Science Foundation (NSF) Arabidopsis 2010 project within the USA was launched with the stated goal of determining the functions of 25,000 genes of Arabidopsis by 2010 (Somerville and Dangl, 2000). Technological advances in each omics research area became essential resources for the investigation of gene function in association with phenotypic changes. A number of these advances include the event of high throughput methods for profiling expressions of thousands of genes, for identifying modification events and interactions within the plant proteome, and for measuring the abundance of the many metabolites simultaneously. Additionally, large scale collections of bio resources, like mass produced mutant lines and clones of full-length cDNAs and their integrative relevant databases, are now available (Brady and Provart, 2009, Kuromori *et al.,* 2009). In conjunction with the rice genome sequence and its related genomic resources, advanced development of mapping populations and molecular marker resources has allowed researchers to accelerate the isolation of agronomically important quantitative trait loci (QTLs) (Konishi *et al.,* 2006, Ma *et al.,* 2006, Kurakawa *et al.,* 2007).

The aforementioned recent high throughput technological advances have provided opportunities to develop collections of sequence based resources and related resource platforms for specific organisms. Such comprehensive models often provide a superb start line for designing experiments, generating hypotheses, or conceptualizing supported the integrated knowledge found within the omic space of a specific organism. Furthermore, the development of such omic resources and data sets for various species allows the comparison of omic properties among species, which promises to be efficient thanks to finding collateral evidence for conserved gene functions which may be evolutionarily supported. Bioinformatics platforms became essential tools for accessing omics data sets for the efficient mining and integration of biologically significant knowledge. Therefore the target of this paper is to review the role of genomics and bioinformatics resources for crop improvement.

2. LITERATURE REVIEW

2.1 Scopes of Bioinformatics

Bioinformatics is a crucial discipline of biological sciences that permits scientists to decipher and manage the vast quantities of knowledge available to them (Ojo and Maxwell, 2010). It consists of two subfields: the event of computational tools and databases. Therefore the application of those tools and databases in generating biological knowledge to understand living systems (Xiong, 2009). These two subfields are complementary to every other. The tool development includes writing software for sequence, structural and functional analysis, also the construction and crating of biological databases. These tools are utilized in three areas of genomic and molecular scientific 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 macromolecule structure analysis, comparison, classification and prediction. The functional analysis includes organic phenomenon profiling, protein–protein interaction prediction, protein sub cellular localization prediction, metabolic pathway reconstruction and simulation (Rao *et al.,* 2008; Xiong, 2009).

2.2 Importance of Bioinformatics

Bioinformatics apply the principles of data sciences and technologies to complex bioscience data (Ojo and Maxwell, 2010). Therefore, the sector of bioinformatics has evolved such the foremost pressing task now involves the analysis and interpretation of varied sorts of data, including nucleotide and aminoalkanoic acid sequences, protein domains, and protein structures (Wales, 2009).The analysis of genetic and DNA sequences like macromolecule sequences, protein structure/function relationships, genome organization, regulation of organic phenomenon , interaction of proteins and mechanisms of physiological functions, can all enjoy a bioinformatics approach.

Macromolecule and protein sequence data from many various species and from population samplings provides a foundation for studies resulting in new understandings of evolution and therefore the explanation of life (Tramontano, 2009). The most tasks of bioinformatics involve the analysis of sequence information that involves identification of genes within the DNA sequences from various organisms, identification of families of related sequences and therefore the development of models, aligning similar sequences and generating phylogenetic trees to look at evolutionary relationships. Finding all the genes and proteins of a genome from a given sequence of amino acids and predicting active sites within the protein structures to connect drug molecules (Rao *et al.,* 2008).

Table1: Classification of databases within the 2004 edition of the biology Database Collection.

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: Vassilev *et al.,* 2005)

2.3 Bioinformatics Tool

2.3.1 Biological databases

Biological databases are archives of consistent data that are stored during a uniform and efficient manner. These databases contain data from a broad spectrum of biology areas. An easy database could be one file containing many records, each of which incorporates an equivalent set of data (Xiong, 2009). Databases are composed of hardware and software for data management. The chief objective of the event of a database is to arrange data during a set of structured records to enable easy retrieval of data. Each record, also called an entry, should contain variety of fields that hold the particular data items. To retrieve a specific record from the database, a user can specify a specific piece of data, called value, to be found during a particular field and expect the pc to retrieve the entire data record. This process is named making a question. There are two sorts of database; the first 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 study on the first resources including information on sequence patterns or motifs, variants and mutations and evolutionary relationships (Rao *et al.,* 2008). Although data retrieval is that the main purpose of all databases, biological have a better level of requirement, referred to as knowledge discovery, which refers to the identification of connections between pieces of data that weren't known when the knowledge was first entered (Xiong, 2009).

2.3.2 Software and Tools

Bioinformatics uses different software that range from simple instruction tools to more complex graphical programs and standalone web-services available from various bioinformatics companies or public institutions. BLAST (Basic Local Alignment Sequence Tool) is that the one that's used for doing sequence alignment. It remains the fastest means by which to spot specific sequences in large datasets and enables the rapid annotation of novel sequences. Although BLAST is that the standard tool for identifying sequence similarities in large datasets, there are several options for assembling sequence datasets, the selection of which depends on hardware availability, dataset size, format, structure and therefore the genetic structure of the organism (Edwards and Batley, 2004).

2.4 Uses of knowledge in Bioinformatics

The first step to creating sense of all the biological sequences and structures is to formulate a way to manage the info, also as the way to process and maintain it. Data management is that the first and most fundamental task of bioinformatics, and bioinformaticians do that by assembling information into databases. A database may be a collection of data stored during a systematic way. Bioinformatics database may contains DNA sequences, RNA sequences, or maybe protein sequences. These sequences could also be organized consistent with their function, or consistent with the species from which they came, or the journal articles which reported them first. A database can also contain journal articles and abstracts. With the info assembled, bioinformaticians can find means by which to mine, retrieve, and use the info. This is often usually done through computer programs, which may search databases and retrieve information, counting on a scientist's needs.

Table 2: 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 and Narayan, 2015)

2.5 Applications of Bioinformatics in Agriculture

Bioinformatics is widely applied in agricultural research. Since agricultural data are of various types and large in collection, its interpretation is difficult; thus Bioinformatics play big role to research the info properly. Collection and storage of plant genetic resource and wisely application of bioinformatics help to supply stronger, more drought, disease and bug resistant crops and improve the standard of livestock making them healthier, more disease resistant and more productive. Comparative genetics consists of the model and non-model plant. Species can reveal a corporation of their genes with reference to one another which further use for transferring information from the model crop systems to other food crops. Arabidopsis thaliana (water cress) and cultivated rice (rice) are samples of available complete plant genomes (Proost *et al.,* 2009).

Plant based biomass is one among the simplest resource for obtaining energy by converting it into biofuels like ethanol which might be wont to drive the vehicles and fly the planes. Biomass based crop species like maize (corn), Panicum virgatum and lignocellulosic species like bagasse, and straw are widely used for biofuel production. We could detect sequence variants in biomass-based crop species to maximise biomass production and recalcitrance. Recently, genome of eucalyptus grandis has been released which is additionally one among major resource of biomass components and every one the genes participate in conversion of sugars into biomass components have already been deciphered, therefore bioinformatics provides great insight into mechanisms and pathways liable for this conversion in order that in future we will enhance production of biomass components in eucalyptus and other relevant plants (Kumar *et al.,* 2017). Thus, the utilization of genomics and bioinformatics together with breeding would likely increase the potential of breeding crop species to be getting used as biofuel feedstock and consequently keep increasing the utilization of renewable energy in modern society (Boyle, 2004).

The goal of plant genomics is to know the genetic and molecular basis of all biological processes in plants. This understanding is prime to permit efficient exploitation of plants as biological resources within the development of latest 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 power to look at organic phenomenon allows us to know how plants answer and interact with the interior and external stimuli. These data may become crucial tool of future breeding decision management systems (Officer *et al.,* 2019).

Bioinformatics helps to know the genetic architecture of microorganism and pathogens to see how these microbes affect the host plant by using Meta genomics and transcriptomics approach, in order that we could generate pathogen resistant crop and would identify those microbes which are beneficial for host (Berg, 2009, Schenk, 2012). Scientists have recently succeeded in transferring genes into rice to extend levels of vitamin A, iron and other micronutrients. Bioinformatic tool helped to supply such golden rice which will 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 therefore the result's a plant whose fruit stays longer on the vine (Fraser *et al.,* 2009).

The change in climate and increase in population will increase pressure on our ability to supply sufficient food. The breeding of novel crops and therefore the adaptation of current crops to the new environment are required to make sure 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 to be used in breeding remains an enormous challenge, and one which can require coordination of diverse skills and expertise. Bioinformatics, when combined with genomics has the potential to assist maintain food security within the face of global climate change through the accelerated production of climate ready crops (Batley and Edwards, 2016).

Pathogen trait is taken into account as a primary interest of plant bioinformatics. The contribution of bioinformatics advances made possible the mapping of the whole genomes of the many organisms in only over a decade. The present efforts to work out gene and protein functions, have improved the power to know the basis causes of plant diseases and find new cures. Furthermore, many future bioinformatics innovations will likely be spurred by the info and analysis demands of the life sciences. Bioinformatics have many practical applications in current disease management with reference to the study of host- pathogen interactions, understanding the disease genetics and pathogenicity factor of a pathogen which ultimately help in designing best management options (Kumar *et al.,* 2017).

2.6 Sequence resources in plants

Comprehensively collected sequence data provide essential genomic resources for accelerating molecular understanding of biological properties and for promoting the appliance of such knowledge. The recent accumulation of nucleotide sequences of model plants, also as of applied species like crops and livestock, has provided fundamental information for the planning of sequence-based research applications in genomics. During this section, we describe recently developed plant sequence resources. Species specific nucleotide sequence collections also provide opportunities to spot the genomic aspects of phenotypic characters supported genome wide comparative analyses and knowledge of model organisms (Flicek *et al.,* 2008, Paterson, 2008, Tanaka *et al.,* 2008).

2.6.1 Genome sequencing projects

The first genome sequence of a plant was completed for A. thaliana, which is now used as a model species in plant biology thanks to its small size, short generation time and high efficiency of transformation. The Arabidopsis genome sequence project was performed as a cooperative project among scientists in Japan, Europe and therefore the USA (Bevan, 1997). The genome sequencing was completed and published in 2000 by the Arabidopsis Genome Initiative (AGI) (The Arabidopsis Genome Initiative 2000). The draft genome sequence of rice, both japonica and indica, a crucial staple food also as a model monocotyledon, was published in 2002 (Goff *et al.,* 2002, Yu *et al.,* 2002). There are variety of providers for plant genome sequences and annotations. Phytozome may be a Web-accessible information resource providing genome sequences and annotations of varied plant species. This resource may be a joint project of the Department of Energy's Joint Genome Institute (DOE–JGI) and therefore the Center for Integrative Genomics, and is meant to facilitate comparative genomic studies among green plants (http://www.phytozome.net/Phytozome_info.php). The present version of Phytozome (ver. 5.0, January 2010) consists of 18 plant species that were sequenced by JGI and other sequencing projects. Gramene (http://www.gramene.org/) is an information resource established as a portal for grass species, and it provides various sorts of information associated with grass genomics, including genome sequences (Ware, 2007).

2.7 Platforms and resources in proteomics

As genome sequencing projects for several organisms are completed, proteome analysis, which is that the detailed investigation of the functions, functional networks and 3D structures of proteins, has gained increasing attention. Large-scale proteome data sets also are a crucial resource for the higher under- standing of protein functions in cellular systems, which are controlled by the dynamic properties of proteins. These properties reflect cell and organ states in terms of growth, development and response to environmental changes. The first objective of functional proteomics was the high-throughput identification of all of the proteins appeared in cells and/or tissues. Recent, rapid technical advances in proteomics (e.g. protein separation and purification methods, advances in mass spectrometry equipment and methodological developments in protein quantification) have allowed us to reach the second generation of functional proteomics, including quantitative proteomics, subcellular proteomics and various modifications and protein-protein interactions (Rossignol *et al.,* 2006 , Yates *et al.,* 2009).

2.7.1 Proteome profiling

The typical experimental work flow of protein profiling are often summarized as protein sample preparation, separation and detection, then identification. Various technical advances for every step of the method have greatly increased the general performance of plant proteomics. Sample preparation is that the most crucial step in any proteomics experiment. The tactic that uses trichloracetic acid (TCA) and acetone is that the most ordinarily used procedure for protein precipitation. A way using phenol and NH 4 OAC/ MeOH is additionally popular for plant tissues. Sample fractionation effectively improves protein detection and increases proteome coverage by reducing sample complexity. Sequential solubilization is an efficient method for fractionating protein samples supported solubility, molecular mass and isoelectric point (Mochida and Shinozaki, 2010).

In addition to standard gel electrophoresis-based separation, the gel-free separation method is usually used, particularly within the 'shotgun proteomics' approach. Within the gel-free method, the protein mixture is directly digested into peptides and separated by the multidimensional separation method. The multi- dimensional separation method may be a combination of various online separation methods including multidimensional protein identification technology. The shotgun approach is suitable for the analysis of proteins that are difficult to separate by 2-DE also as for high-throughput analysis by automated analytical instruments (Yates *et al.,* 2009).

2.7.2 Quantitative proteomics

Comprehensive quantification of every protein's abundance is sort of important for a far better understanding of the protein dynamics regulated in response to cellular state and environmental changes. A quantitative proteome approach also plays an important role within the discovery of key proteomic changes, including expression, interaction and modification, that are related to genetic variations and/or visible phenotypic changes (Gstaiger and Aebersold, 2009).

Difference gel electrophoresis may be a popular method for differential display of proteins for quantitative protein comparison. In Difference gel electrophoresis protein samples are labelled with different fluorescent dyes before 2-D electrophoresis, enabling accurate analysis of differences in protein abundance between samples (Rossignol *et al.,* 2006). This method is an efficient thanks to remove gel to gel variation while significantly increasing accuracy and reproducibility. Isotope coded affinity tags, isobaric tags for relative and absolute quantitation and stable isotope labeling with amino acids in cell culture are widely used methods for protein differential display using stable isotope labelling (Mochida and Shinozaki, 2010).

2.7.3 Subcellular proteomics

Large-scale proteome analysis of cell organelles is important for understanding the enzymatic inventory of a cell organelle; the compartmentalization of metabolic pathways; cellular logistics like protein targeting, trafficking and regulation; and proteomic dynamics at the organelle level caused by changes in cellular systems (Andersen and Mann 2006, Baginsky 2009). variety of approaches are applied to research the proteome of organelles or subcellular compartments of plant cells like chloroplasts, etioplasts, amyloplasts, chromoplasts, mitochondria, vacuoles, plasma membranes, nucleus, peroxisomes, cytosolic ribosome and cell membrane (Baginsky 2009). Proteomic analyses of chloroplasts, mitochondria and further fractionations are administered to work out detailed localizations of protein in several sub organelle compartments. Techniques for quantitative proteomics, like the ICAT and iTRAQ methods described above, also are effective for acquiring quantitative data on proteomes in each organelle. In Arabidopsis, rice and alga, differential proteome profiles of plant plasma membranes were monitored to spot those proteins differentially expressed in response to environmental factors like cold acclimation, salt stress and bacterial elicitor (Benschop *et al.,* 2007). Post-translational protein modifications Comprehensive approaches to research various sorts of post-translational protein modifications also play a key role within the current study of proteomics (Mochida and Shinozaki, 2010).

Platform advances in structural proteomics Large-scale data sets of protein 3D structures also are crucial information resources for elucidating relationships between protein functions and structures or for analyzing molecules in protein complexes. The International genomics Organization (http://www.isgo.org) was formed to facilitate global genomics research efforts (Stevens *et al.,* 2001). The key centers for genomics are the RIKEN Structural Genomics/Proteomics Initiative in Japan, the Protein Structure Initiative within the USA and therefore the genomics centers of Europe (Yokoyama, 2003). International efforts to work out protein structures have contributed to increases within the number of solved protein structures. Thus, the amount of solved protein structures appearing within the protein data bank (http://www.pdb.org/pdb/home/ home.do), which is that the hottest resource for biomolecule structure data sets, has dramatically increased during the past decade (Kouranov *et al.,* 2006).

2.8 Mutant resources for phenome analysis

Analysis of mutants is an efficient approach for investigation of gene function (Springer, 2000). Comprehensive collections of mutant lines also are essential bioresources for radically accelerating forward and reverse genetics. The available mutant resources for phenome analysis in plant species are well described during a recent review by Kuromori et al., (2009). Various analytical platforms have rapidly evolved, allowing us to get genes involved especially phenotypic changes. Alongside these analysis platforms, the stress for comprehensive collections of mutants and related information resources have dramatically increased, encouraging high-throughput and genome-wide phenome analysis in plant species (Alonso and Ecker 2006).

2.8.1 Insertion mutant

With the completion of genome sequencing in plants, insertion mutant resources with index data that document the inserted genomic position became extremely beneficial resources by which to market functional analysis of annotated genes that are disrupted by a reverse genetics approach. Transferred DNA-tagged (T-DNA-tagged) lines and transposon-tagged lines became popular resources for the investigation of insertion mutants in plants. T-DNA-tagged lines have emerged as a well-liked mutant resource thanks to the rapid generation of large- scale populations in Arabidopsis (Krysan *et al.,* 1999). There are variety of resources for insertion mutant populations with insertion site index-tagged data available for various plant species (Kuromori *et al.,* 2009)

2.8.2 Activation tagging

Activation tagging may be a popular method for generating gain-of-function mutant populations. The tactic uses T-DNA or a transposable element containing cauliflower mosaic virus 35S enhancer multimers (Weigel and Mott, 2009). With transcriptional activation of genes near the insertion, novel phenotypes are expected to seem which will identify genes that are redundant or essential for survival. Mutant resources have then been wont to isolate genes from Arabidopsis, rice, petunia and tomato (Mathews *et al.,* 2003). Recently, AT systems employing a transposon of maize En / Spm or Ac/Ds are developed in Arabidopsis and rice, respectively. Variety of AT projects are performed in various plant species like Arabidopsis, rice and soybean (Kuromori *et al.,* 2009).

2.8.3 Chemical and physical mutagenesis

Chemical mutagenic agents, like ethyl methane sulfonate, sodium azide and methylnitrosourea, and physical mutagens, like fast-neutrons, gamma rays and ion-beam irradiation, are wont to generate mutant populations for several years for forward genetics in various plant species. Targeting induced local lesions in genomes (TILLING) was developed as a general reverse-genetic strategy that gives an allelic series of induced point mutations in genes of interest (Till *et al.,* 2004, Till *et al.,* 2006). Because high-throughput TILLING permits the rapid and low-cost discovery of induced point mutations in populations of chemically mutagenized individuals, the tactic has been applied to varied animal and plant species. The TILLING technology also can be wont to explore allelic variations that are appeared in natural variation; this technology is named Eco TILLING (Wang *et al.,* 2006). Several laboratory sites have established TILLING and/or EcoTILLING centers for communities of users as a public service (Barkley and Wang, 2008).

2.8.4 Gene silencing technologies

Although insertion mutagenesis is an efficient method for generating loss of function mutants, it also has limitations within the case of redundant genes and lethal mutants. To beat these limitations, methods to interrupt organic phenomenon are developed and applied to the functional analysis of plant genes. RNA interference may be a popular method for RNA- mediated gene silencing by sequence-specific degradation of homologous mRNA triggered by double-stranded RNA, which is additionally referred to as post-transcriptional gene silencing (Chuang and Meyerowitz, 2000). Constitutive expression of an intron-containing self- complementary hairpin RNA has been an efficient method for silencing target genes in plants. With demands for conditional silencing of target genes (the silencing of which ends up in prevention of plant regeneration or embryonic lethality), conditional RNAi systems employing a chemical-inducible Cre/loxP recombination system or a promoter of warmth shock- inducible genes are recently developed (Guo *et al.,* 2003, Mochida and Shinozaki, 2010).

2.9 Plant comparative genomics and databases

The recent accumulation of nucleotide sequences for agricultural species, including crops and livestock , now allows us to perform genome-wide comparative analyses of model organisms with the goal of discovering key genes involved in phenotypic characteristics (Sato and Tabata, 2006). the mixing of genomic resources derived from various related species, like large- scale collections of cDNAs and data from whole-genome sequencing projects, should facilitate sharing of data about gene function between models and applied organisms. this may also accelerate molecular elucidation of cellular systems associated with agronomically important traits. Variety of data resources for plant genomics accessible on the online have appeared, alongside appropriate analytical tools.

2.9.1 Portal information resources in plants

TAIR is one among the foremost popular and integrated information resource in plant science, and it plays an important role as a portal in Arabidopsis research (http://www.arabidopsis .org/) (Swarbreck *et al.,* 2008). The Salk Institute Genomic Analysis Laboratory is additionally an information resource that integrates various data sets of serious omics results mainly associated with Arabidopsis (http://signal.salk.edu/).The RIKEN Arabidopsis Genome Encyclopedia provides information on various genomic resources built at

RIKEN for Arabidopsis research (http://rarge.gsc.riken.jp/db_home.pl) ((Mochida and Shinozaki, 2010)). Such portal sites have provided gateways for access to comprehensive omics data and/or bio resources. These sites also house cross-referenced data sets built between each annotated gene and its associated instances, like gene–full-length cDNA clones, gene– mutants, gene–expression patterns and gene–homologous genes. Therefore, to see an annotated gene alongside genome sequences and associated information, genome browsers like G browse are implemented on internet sites (Donlin, 2007).

Gramene may be a popular portal that's not only an integrated rice information resource but also a portal for promoting plant comparative genomics (http://www.gramene.org/). Gramene offers integrated genome-oriented data including gene annotation and molecular markers, and also a QTL data- base mainly for Gramineae species. Alongside the launch of genome sequencing projects, portal sites to share the progression of outcomes and to integrate related resources have appeared for various species. The Sol genomics network may be a portal for Solanaceae genome resources that has information on the tomato genome sequencing project (Mueller *et al.,* 2005).

2.9.2 Genome-wide comparisons among plants species

With the completion of genome sequencing during a number of plant species, genome scale comparative analyses are often wont to produce and publish data sets that facilitate identification of con- served and/or characteristic properties among plant species. Using modelled proteome data sets deduced from sequenced genomes in plants, several efforts are completed to construct comprehensive gene families with the aim of building platforms to verify gene content and elucidating the method of gene duplication and functional diversification among species (Sterck *et al.,* 2007). Comprehensive gene family data sets are usually produced by computational procedures including a step that conducts an all-against-all sequence similarity search then a step for building clusters of protein families by methods like Markov Clustering or consideration of protein domain structures. The results of such studies can themselves yield databases that are useful for further phylogenetic studies (Wall *et al.,* 2008).

2.9.3 Focused database for plant genomics

Databases housing focused data sets alongside rich annotations and well interrelated cross-references also are quite useful for the higher understanding of focused issues especially gene families and/or particular cellular processes. Sequence-specific DNA-binding TFs are key molecular switches that control or influence many biological processes, like development or responses to environmental changes. In plants, the genome-wide identification of repertories of genes encoding TFs of the Arabidopsis genome was reported first, and comparisons with other organisms revealed the properties of plant specific TFs (Riechmann *et al.,* 2000). Within the past decade, with the supply of complete genome sequences, we've been ready to compile catalogs describing the function and organization of TF regulatory systems during a number of organisms. Recently, further integration of knowledge sets of TF-encoding genes has been performed, thus establishing an integrative, knowledge-based resource of TFs across related plant species in terms of comparative genomics of transcriptional regulatory networks (Mochida and Shinozaki, 2010).

3 SUMMARY AND CONCLUSION

Agriculture faces substantial challenges in harnessing the deluge of genomic data of diverse origins and formats for crop improvement. To beat these challenges, novel breeding methods and bioinformatics tools must be wont to translate genomic data into gains in crop yield and yield stability. Bioinformatics is now playing a big role within 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. To accelerate the detection of strong gene trait associations, researchers can apply Meta quantitative trait locci analyses, genomic wide association studies and genetic screens. While genome editing offers a valuable approach to rapidly introduce beneficial mutations into elite cultivars, GS increases selection efficiency without requiring knowledge of underlying genetic drivers.

The high degree of synteny among diverse plant species, commonality in traits, and the supply of expression and performance information of sequences has enabled the invention of the many useful traits for crop improvement. Genome sequencing of several important plants species has enabled researchers to spot ‘chromosome’ and ‘difference’ think about sequences. This successively has been wont to identify value traits for crop improvement. As an example, the barley stem rust resistance gene has been identified from rice barley comparisons and therefore the sugarcane rust resistance gene supported maize-sorghum comparisons. Comparative genomics alongside bioinformatics could help in achieving improvement of yields in rice, maize, and other related grass crops like barley, rye, sugarcane and wheat. The power to represent high resolution physical and genetic maps of plants has been one among the good applications of bioinformatics tools. By applying novel technologies and methods together, future plant breeding are able to do the crop improvement rate required to make sure food security.

4. 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.; et al. Global agricultural intensification during climate change: A role for genomics. Plant Biotechnol. J. 2016, 14, 1095–1098.

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

Alonso, J.M. and Ecker, J.R. (2006) Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in Arabidopsis. Nat. Rev. Genet. 7: 524 – 536.

Andersen, J.S. and Mann, M. (2006) Organellar proteomics: turning inventories into insights. EMBO Rep. 7: 874 – 879.

Baginsky , S. (2009) Plant proteomics: concepts, applications, and novel strategies for data interpretation . Mass Spectrom. Rev. 28: 93 – 120

Barkley, N.A. and Wang, M.L. (2008) Application of TILLING and EcoTILLING as reverse genetic approaches to elucidate the function of genes in plants and animals .Curr. Genomics 9: 212 – 226

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

Benschop , J.J. , Mohammed , S. , O'Flaherty , M. , Heck , A.J. , Slijper , M. and Menke , F.L. (2007) Quantitative phosphoproteomics of early elicitor signaling in Arabidopsis . Mol. Cell. Proteomics 6:1198 – 1214.

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.

Bevan, M. (1997) Objective: the complete sequence of a plant genome. Plant Cell 9: 476 – 478.

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

Brady, S.M. and Provart , N.J. (2009) Web-queryable large-scale data sets for hypothesis generation in plant biology . Plant Cell 21: 1034 – 1051.

Brown, M.E. and Funk, C.C. (2008) Climate. Food security under climate change .Science 319:580 581

Chuang, C.F. and Meyerowitz , E.M. (2000) Specific and heritable genetic interference by double-stranded RNA in Arabidopsis thaliana . Proc. Natl Acad. Sci. USA 97: 4985 – 4990 .

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

Donlin , M.J. (2007) Using the Generic Genome Browser (GBrowse) . Curr. Protoc. Bioinformatics

Edwards, D. and Batley, J. (2004). Plant bioinformatics: from genome to Phenome. Biotechnology22 (5): 232-237.

Edwards, D.; Batley, J.; Snowdon, R.J. Accessing complex crop genomes with next-generation sequencing. Theor. Appl. Genet. 2013, 126, 1–11. [CrossRef] [PubMed]

Flicek , P. , Aken , B.L. , Beal , K. , Ballester , B. , Caccamo , M. , Chen , Y. , et al . (2008) Nucleic Acids Res. 36: D707 – D714.

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.

Goff , S.A. , Ricke , D. , Lan , T.H. , Presting , G. , Wang , R. , Dunn , M. , et al . (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science 296: 92 – 100.

Gstaiger , M. and Aebersold , R. (2009) Applying mass spectrometry- based proteomics to genetics, genomics and network biology . Nat. Rev. Genet. 10: 617 – 627.

Guo , H.S. , Fei , J.F. , Xie , Q. and Chua , N.H. (2003) A chemical-regulated inducible RNAi system in plants . Plant J. 34: 383 – 392.

Huang X. and Madan, A. (1999) A DNA sequence assembly program .Genome Res. 9: 868 – 877.

Hulsen T., de Vlieg , J. and Groenen , P.M. (2006) PhyloPat: phylogenetic pattern analysis of eukaryotic genes . BMC Bioinformatics 7: 398.

Hunter , S. , Apweiler , R. , Attwood , T.K. , Bairoch , A. , Bateman , A. , Binns , D. , et al . (2009) InterPro: the integrative protein signature database. Nucleic Acids Res. 37: D211 – D215.

Hyten , D.L. , Song , Q. , Choi , I.Y. , Yoon , M.S. , Specht , J.E. , Matukumalli , L.K. , et al . (2008) High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. Theor. Appl. Genet. 116: 945 – 952.

Ichikawa , T. , Nakazawa , M. , Kawashima , M. , Iizumi , H. , Kuroda , H. , Kondou , Y. , et al . (2006) The FOX hunting system: an alternative gain-of-function gene hunting technique. Plant J. 48: 974 – 985.

Igawa , T. , Fujiwara , M. , Takahashi , H. , Sawasaki , T. , Endo , Y. , Seki , M. , et al . (2009) Isolation and identification of ubiquitin-related proteins from Arabidopsis seedlings. J. Exp. Bot. 60: 3067 – 3073.

Iida , K. , Seki , M. , Sakurai , T. , Satou , M. , Akiyama , K. , Toyoda , T. , et al . (2004) Genome-wide analysis of alternative pre-mRNA splicing in Arabidopsis thaliana based on full-length cDNA sequences. Nucleic

Acids Res. 32: 5096 – 5103.

Iida , K. , Seki , M. , Sakurai , T. , Satou , M. , K. , Toyoda , T. et al . (2005) RARTF: database and tools for complete sets of Arabidopsis transcription factors. DNA Res.12: 247 - 256.

Iijima , Y. , Nakamura , Y. , Ogata , Y. , Tanaka , K. , Sakurai , N. , Suda , K. , et al . (2008) Metabolite annotations based on the integration of mass spectral information. Plant J. 54: 949 – 962.

Imanishi , T. , Itoh , T. , Suzuki , Y. , O'Donovan , C. , Fukuchi , S. , Koyanagi , K.O. , et al . (2004) Integrative annotation of 21,037 human genes validated by full-length cDNA clones. PLoS Biol. 2: e162.

International Rice Genome Sequencing Project (2005) the map-based sequence of the rice genome. Nature 436: 793 – 800.

IRGSP. The map-based sequence of the rice genome. Nature 2005, 436, 793–800. [CrossRef]

Islam , N. , Lonsdale , M. , Upadhyaya , N.M. , Higgins , T.J. , Hirano , H. and Akhurst , R. (2004) Protein extraction from mature rice leaves for two-dimensional gel electrophoresis and its application in proteome analysis . Proteomics 4: 1903 – 1908.

Itoh , T. , Tanaka , T. , Barrero , R.A. , Yamasaki , C. , Fujii , Y. , Hilton , P.B. , et al . (2007) Curated genome annotation of Oryza sativa ssp. japonica and comparative genome analysis with Arabidopsis thaliana. Genome Res. 17: 175 – 183.

Jaccoud , D. , Peng , K. , Feinstein , D. and Kilian , A. (2001) Diversity arrays: a solid state technology for sequence information independent genotyping . Nucleic Acids Res. 29: E25.

Jaillon , O. , Aury , J.M. , Noel , B. , Policriti , A. , Clepet , C. , Casagrande , A. , et al . (2007) the grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449: 463 – 467.

Jayaram, B. and Dhingra, P. (2010). Bioinformatics for a better tomorrow. Indian Institute of Technology, HauzKhas, New Delhi - 110016, India.

Jiao , Y. , Tausta , S.L. , Gandotra , N. , Sun , N. , Liu , T. , Caly , N.K. , et al . (2009) A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. Nat. Genet. 41: 258 – 263.

Jonsson , P. , Gullberg , J. , Nordstrom , A. , Kusano , M. , Kowalczyk , M. , Sjostrom , M. , et al . (2004) A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS. Anal. Chem. 76: 1738 – 1745.

Konishi , S. , Izawa , T. , Lin , S.Y. , Ebana , K. , Fukuta , Y. , Sasaki , T. and Yano , M. (2006) An SNP caused loss of seed shattering during rice domestication . Science 312: 1392 – 1396.

Kouranov , A. , Xie , L. , de la Cruz , J. , Chen , L. , Westbrook , J. , Bourne , P.E. , et al . (2006) The RCSB PDB information portal for structural genomics. Nucleic Acids Res. 34: D302 – 305.

Krysan , P.J. , Young , J.C. and Sussman , M.R. (1999) T-DNA as an insertional mutagen in Arabidopsis . Plant Cell 11: 2283 – 2290.

Kumar, U.*et al.*(2017) 'Role of Bioinformatics in Crop Improvement Role of Bioinformatics inCropImprovement', (February).

Kurakawa , T. , Ueda , N. Maekawa , M. , Kobayashi , K. , Kojima , M. , Nagato , Y et al . (2007) direct control of shoot meristem activity by a cytokinin-activating enzyme. Nature 445: 652 – 655.

Kuromori , T. , Takahashi , S. , Kondou , Y. , Shinozaki , K. and Matsui , M. (2009) Phenome analysis in plant species using loss-of-function and gain-of-function mutants . Plant Cell Physiol. 50: 1215 – 1231.

Kuromori , T. , Takahashi , S. , Kondou , Y. , Shinozaki , K. and Matsui , M. (2009) Phenome analysis in plant species using loss-of-function and gain-of-function mutants . Plant Cell Physiol. 50: 1215 – 1231.

Ma , J.F. , Tamai , K. , Yamaji , N. , Mitani , N. , Konishi , S. , Katsuhara , M. , et al . (2006) A silicon transporter in rice. Nature 440: 688 – 691.

Mathews , H. , Clendennen , S.K. , Caldwell , C.G. , Liu , X.L. , Connors , K. , Matheis , N. , et al . (2003) Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modifi cation, and transport. Plant Cell 15: 1689 – 1703.

Mochida, K. and Shinozaki, K. (2010) 'Genomics and Bioinformatics Resources for Crop', 51(4), pp. 497–523. doi: 10.1093/pcp/pcq027.

Officer, S. *et al.* (2019) 'Application of Phenomics , Genomic Resources and Bioinformatics Tools for Tea Plant Improvement', 7(6), pp. 622–627.

Ojo, O.O. and Maxwell, O. (2010).Incorporating bioinformatics into biological science education in Nigeria: prospects and challenges, Infection, Genetics and Evolution: 10.1016/j.meegid, 11.015.

Paine JA, Shipton CA, Chaggar S, Howells RM, Kennedy MJ, Vernon G, et al. (2005). Improving the nutritional value of Golden Rice through increased pro-vitamin A content. Nat Biotechnol. 23: 482-487.

Paterson, A.H. (2008) Genomics of sorghum. Int. J. Plant Genomics 2008: 362451.

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.

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.

Riechmann , J.L. , Heard , J. , , Reuber , L. , Jiang , C. , Keddie , J. , et al . (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes .Science 290:2105 2110.

Rossignol , M. , Peltier , J.B. , Mock , H.P. , Matros , A. , Maldonado , A.M. and Jorrin , J.V. (2006) Plant proteome analysis: a 2004–2006 update . Proteomics 6: 5529 – 5548.

Sato, S. and Tabata , S. (2006) Lotus japonicus as a platform for legume research . Curr. Opin. Plant Biol. 9: 128 – 132.

Schenk PM, Carvalhais LC, Kazan K. (2012). Unraveling plant-microbe interactions: can multi- species transcriptomics help? Trends Biotechnol, 30: 177-184.

Somerville, C. and Dangl , J. (2000) Genomics. Plant biology in 2010. Science 290: 2077 – 2078.

Springer, P.S. (2000) Gene traps: tools for plant development and genomics. Plant Cell: 1007 – 1020

Sterck , L. , Rombauts , S. , Vandepoele , K. , Rouze , P. and Van de Peer , Y. (2007) How many genes are there in plants (… and why are they there)? Curr. Opin. Plant Biol. 10: 199 – 203.

Swarbreck , D. , Wilks , C. , Lamesch , P. , Berardini , T.Z. , Garcia-Hernandez , M. , Foerster , H. , et al . (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic Acids Res. 36: D1009 – D1014.

Takeda, S. and Matsuoka, M. (2008) Genetic approaches to crop improvement: responding to environmental and population changes. Nat. Rev. Genet. 9: 444 – 457.

Tanaka , T. , Antonio , B.A. , Kikuchi , S. , Matsumoto , T. , Nagamura , Y. , Numa , Y. , et al . (2008) The Rice Annotation Project Database (RAP-DB): 2008 update. Nucleic Acids Res. 36: D1028 – D1033

Till , B.J. , Colbert , T. , Codomo , C. , Enns , L. , Johnson , J. , Reynolds , S.H. , et al . (2006) High-throughput TILLING for Arabidopsis. Methods Mol. Biol. 323: 127 – 135.

Till , B.J. , Reynolds , S.H. , Weil , C. , Springer , N. , Burtner , C. , Young , K. , et al . (2004) Discovery of induced point mutations in maize genes by TILLING . BMC Plant Biol. 4 : 12

Tilman, D.; Balzer, C.; Hill, J.; Befort, B.L. Global food demand and the sustainable intensification of agriculture. Proc. Natl. Acad. Sci. USA 2011, 108, 20260–20264.

Tramontano, A.L. (2009). Bioinformatics. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester.DOI: 10.1002/97804700 15902.a0001900.pub2.

Ulijaszek, S.J. Human dietary change. Philos. Trans. R. Soc. Lond. B Biol. Sci. 1991, 334, 271–279.

Wales, J.(2009).Bioinformaticshttp://en.wikipedia.org/wiki/BioinformaticsWebservices_inbioinformatics

Wall , P.K. , Leebens-Mack , J. , Muller , K.F. , Field , D. , Altman , N.S. and dePamphilis , C.W. (2008) Plant Tribes: a gene and gene family resource for comparative genomics in plants .Nucleic Acids Res. 36 : D970 – D976 .

Wang, D.K., Sun , Z.X. and Tao , Y.Z. (2006) Application of TILLING in plant improvement . Yi Chuan Xue Bao 33: 957 – 964.

Ware, D. (2007) Gramene: a resource for comparative grass genomics.Methods Mol. Biol. 406: 315 330

Weigel, D. and Mott, R. (2009) the 1001 genomes project for Arabidopsis thaliana .Genome 10: 107.

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

Yates, J.R. , Ruse , C.I. and Nakorchevsky , A. (2009) Proteomics by mass spectrometry: approaches, advances, and applications . Annu. Rev. Biomed. Eng. 11: 49 – 79.

Yokoyama, S. (2003) Protein expression systems for structural genomics and proteomics. Curr. Opin. Chem. Biol. 7: 39 – 43.

Yu , J. , Hu , S. , Wang , J. , Wong , G.K. , Li , S. , Liu , B. , et al . (2002) a draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science 296: 79 – 92