

Role of Genomics and Bioinformatics Resources for Crop Improvement

Werkissa Yali*

Department of Agriculture, Chiro National Sorghum Research and Training Centre, P.O.BOX 190, Chiro, Ethiopia

*Corresponding author: Dr Werkissa Yali, Department of Agriculture, Chiro National Sorghum Research and Training Centre, P.O.BOX 190, Chiro, Ethiopia, E-mail workissayali@gmail.com

Received date: August 19, 2021; Accepted date: September 02, 2021; Published date: September 09, 2021

Copyright: © 2021 Yali W. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

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; Omics resource

Introduction

Humans depend upon crops for over two thirds of their daily energy intake [1]. Because the global human population grows, agriculture (crop cultivation) is under increasing pressure to supply higher crop yields [2]. Additionally global climate change, limited availability of land and water shortages is 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 [1]. 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 [3].

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 [4]. Sustainable agricultural production is an urgent issue in response to global climate change and population increase [3]. 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 [5]. 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 [6]. 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 [2]. 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 [1-3]. 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 cDNA and their integrative relevant databases, are now available [7]. 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) [2-4].

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.

Literature Review

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 [2-5]. 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 [3-6]. 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 [7]. 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 [7-8].

Importance of Bioinformatics

Bioinformatics apply the principles of data sciences and technologies to complex bioscience data [6]. 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 [4-6]. 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 in Tables 1 and 2.

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

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

Bioinformatics Tools

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 [8]. 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.

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 [9].

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

Table 2: List of some published plant genome.

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 [7-9]. 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, conditional RNAi systems employing a chemical-inducible recombination system or a promoter of warmth shock-inducible genes are recently developed.

Discussion

Plant comparative genomics and data bases

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 [10]. 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.

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 [10]. The Salk Institute Genomic Analysis Laboratory is additionally an information resource that integrates various data sets of serious omics results mainly associated with Arabidopsis [6,7]. The RIKEN Arabidopsis Genome Encyclopedia provides information on various genomic resources built at RIKEN for Arabidopsis research [9,10]. 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 Gbrowse are implemented on internet sites [11].

Gramene may be a popular portal that's not only an integrated rice information resource but also a portal for promoting plant comparative genomics [8]. 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 [7-9].

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 loci 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.

Conclusion

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.

References

1. Abberton M, Batley J, Bentley A, Bryant, J, Cai H, et al.(2016) Global agricultural intensification during climate change: a role for genomics. *Plant Biotechnol J* 14: 1095–1098.
2. Agarwal R, Narayan J (2015) Unraveling the Impact of Bioinformatics and Omics in Agriculture. *Int J Plant Biol Res* 3: 1039.
3. Alonso JM, Ecker JR (2006) Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in Arabidopsis. *Nat Rev Genet* 7: 524 – 536.
4. Andersen JS, Mann M (2006) Organellar proteomics: turning inventories into insights. *EMBO Rep* 7: 874 – 879.
5. Baginsky S (2009) Plant proteomics: concepts, applications, and novel strategies for data interpretation . *Mass Spectrom. Rev* 28: 93 – 120.
6. Barkley NA, Wang ML(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.
7. Batley J, David E (2016) The application of genomics and bioinformatics to accelerate crop improvement in a changing climate. *Curr Opin Plant Biol* 30: 78–81.
8. Benschop JJ, Mohammed S, Flaherty O, Menke FL, Slijper M, et al. (2007) Quantitative phosphoproteomics of early elicitor signaling in Arabidopsis . *Mol Cell Proteomics* 6: 1198–1214.
9. 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.
10. Bevan M (1997) Objective: the complete sequence of a plant genome. *Plant Cell* 9: 476 – 478.
11. Brady SM, Provart NJ. (2009) Web-queryable large-scale data sets for hypothesis generation in plant biology. *Plant Cell* 21: 1034 – 1051.