USE OF ARTIFICIAL INTELLIGENCE AND BIOINFORMATICS FOR CROP IMPROVEMENT TO ENSURE FUTURE FOOD SECURITY

Abstract

Modern plant breeders face a tough insurmountable challenge to feed the world's rapidly expanding population. Insect pest attacks, disease severity, and nutrient deficiency have all reduced agricultural crop yield in recent years. Every day, we face an increasing challenge in satisfying the demands of the expanding population. To accomplish the same, it is vital to use multidisciplinary techniques to find answers to current problems. We've recently witnessed a paradigm change toward employing omics information, methods, and technology to boost agricultural productivity. Crop genotype and phenotypic data provided by the omics era has opened several doors. This will prove to be a vital tool for enhancing agricultural output and farmer incomes. A recent development in agricultural technology is the use of artificial intelligence (AI). High-performance, precise, and cost-effectiveness are only few of the advantages of AI in agriculture. Identifying, cloning, and sequencing genes that help plants tolerate harmful environmental impacts should be made easier with a better understanding of plant genomics. Agricultural and food industries, in particular, are rapidly evolving, and machine learning has recently been recognized as a feasible multidisciplinary technique for enhancing and upgrading such industries. Bioinformatics and artificial intelligence might be used to find the genome and its variations, which could then be used to genetically edit crops in the future, according to this paper's focus on plant multi-omics.

Keywords: artificial intelligence, bioinformatics, deep learning, genomics, metabolomics, proteomics, transcriptomics,

Authors

Sandip Debnath
Department of Genetics and Plant Breeding
Palli Siksha Bhavana
Visva-Bharati University
West Bengal, India
sandip.debnath@visva-bharati.ac.in

Sourish Pramanik
Department of Genetics and Plant Breeding
Palli Siksha Bhavana
Visva-Bharati University
Birbhum, West Bengal, India
sourishpramanik2002@gmail.com

Dibyendu Seth
Department of Genetics and Plant Breeding
Palli Siksha Bhavana
Visva-Bharati University
Birbhum, West Bengal, India
deep032002@gmail.com

Biswajit Pramanik
Department of Genetics and
Plant Breeding
Palli Siksha Bhavana
Visva-Bharati University
Birbhum, West Bengal, India
biswajit1996pramanik@gmail.com

I. BACKGROUND

Sustainable agricultural productivity and food security are critical challenges in light of growing populations, environmental degradation, and climate change [1]. Crops provide more than two-thirds of the energy we use each day as individuals. As the world's population continues to rise, agriculture is under increasing pressure to provide more food. Further agricultural concerns are posed by climate change, land scarcity, and water limits. Additional food security issues have been exacerbated by the recent rise in demand for biofuel crops, which has created a new market for agricultural products. A number of genetic applications have provided several chances for integrating the benefits of subsystems biology, integrative biology, and large-scale systematic functional genomic programmes in order to tackle these issues. The area of plant molecular biology is progressing thanks to the discovery of important gene sequences and their functions. Genes that have been attributed to crop yields, quality, and resistance to biotic and abiotic challenges have been identified [2, 3].

Arabidopsis thaliana's whole genome has been available to scientists since 2000. P. 149) of the International Arabidopsis Genome Initiative. There has been a complete genome sequencing of rice (Oryza sativa cv. japonica) since 2005 (International Rice Genome Sequencing Project 2005) A combination of 454 sequencing and Sanger sequencing was used for the first time to sequence the grape genome, whilst rice was still sequenced using BAC and Sanger sequencing. Agriculture accounted for 40 of the 55 plant genomes sequenced as of 2013. There are 237520318 sequences in Genbank as of April 2022 (https : / / www . ncb . nlm.nih.gov/genbank/statistics/).

Plant genome research needs bioinformatics, which is required in order to handle and analyze the massive volumes of genomic data. Third-generation sequencing data presents a challenge for many algorithms designed for short reads. Crop improvement may benefit from comprehensive data provided by GWAS, variant calling, and comparative genomic analysis. Genomic sequencing of crop populations may provide gene-level resolution of agronomic variation, quantitative trait locus (QTL) mapping, and more in many areas of crop breeding, including genome-wide association studies (GWAS). A result of the ease with which breeders may now get genetic information. Crop improvement has never had it so well thanks to recent advances in multi-omics. [6] The "omic space," a conceptual paradigm that ranges from the "genome" to the "phenome," has been proposed [7]. Plant phenotypic changes are linked to changes in the structure and function of genes. Gene-specific molecular breeding and the interplay between the genome, proteome, and metabolome have led to the development of various web-based databases that can hold massive amounts of data. Genome sequencing is useful in improving agronomic qualities so that genetic potential may be harnessed to boost productivity, as a genomics technique. Deoxyribonucleic acid (DNA) sequencing has become more affordable in the recent decade, which has led to a rise in crop genome sequencing, giving breeders an excess of possibilities.

Using machine learning to forecast and categorize data is an alternative to traditional statistical methods (ML). Mathematical and statistical approaches are used to train models without the need for direct programming in machine learning. In order to produce predictions, machine learning develops a number of algorithms that learn from both training data and sample data. Non-parametric machine learning (ML) research on plants and animals have employed support vector machines (SVM), boosting, random forests, and Reproducing

Kernel Hilbert Space (RKHS). Because they discover patterns from data without any previous assumptions, ML models for genomic selection (GS) are particularly advantageous because they take into account all of the variations, their interactions, and environmental factors [7]. Each nucleotide has an impact on a plant's phenotypic, and a scientist is interested in this as well. It is possible that deep learning might make very accurate predictions, but the models themselves are frequently quite complex, making it difficult to use inference to study biological processes. Therefore, academics haven't given much attention to deep learning (DL).

The flow of biological information underpinning complex characteristics necessitates an alternative systems biology approach that includes the integration of various omics data, modelling, and prediction of cellular processes. This technique provides a full understanding of the dynamic system in which various levels of biological structure interact with the external environment in order to exhibit phenotype. One of the most popular omics approaches in plant science is genomics. This is due to the fact that the cost of sequencing is decreasing and the degree of knowledge is increasing. It's possible to identify new alleles regardless of whether or not the genome sequence is available, owing to the sequencing and re-sequencing information gleaned for various crops. Plant network biology may help boost sustainable agricultural yields, but a systematic approach is needed (Figure 1). For the improvement of agriculturally important plants, the most current advances in bioinformatics and artificial intelligence are emphasized in this article.

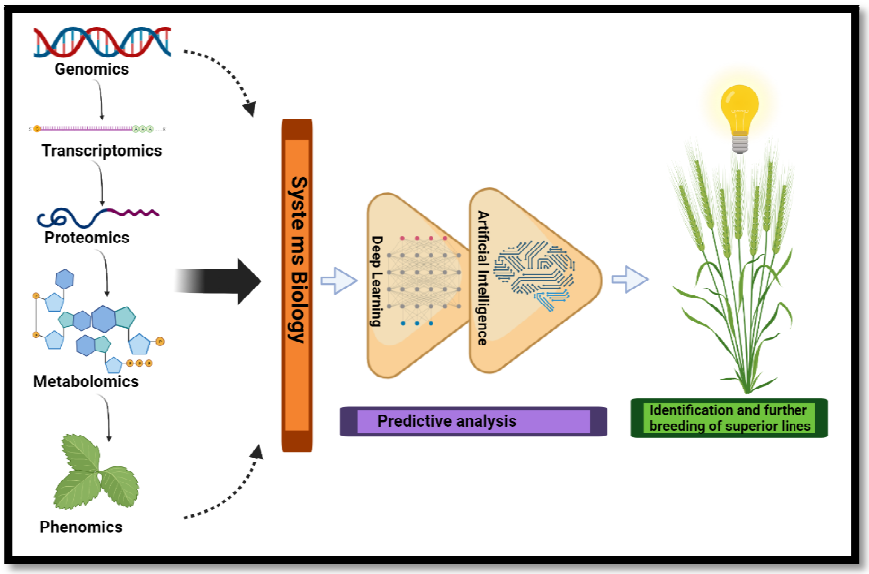

Figure 1: Pictorial Depiction of Genomics, Transcriptomics, Proteomics, Metabolomics And Phenomics' Integrated Application in Crop Improvement

II. NEXT GENERATION SEQUENCING

Short Illumina reads and Sanger sequencing were used to sequence the cucumber genome in 2009, which paved the way for NGS. In order to discover genes and gene families, as well as coding and noncoding areas, regulatory genes, and repetitive sequences, genomic data is employed. More and more plant biologists are routinely sequencing and resequencing

their genomes attributable to next-generation sequencing (NGS). The genomes of 55 plant species, including 40 crop species, have been sequenced as of 2013. It has also been utilised to discover genome-wide molecular phenotypes with several dimensions using low-cost high-throughput techniques. Individual, strain, and/or population differences may be identified using next-generation sequencing technologies and reference genome sequence data. Nucleotide polymorphisms may be consistently identified in genetic research by mapping sequence segments to a specific reference genome data set.

Plant genome assembly is still a challenge because of long repetitive regions, large genome sizes, and frequent polyploidy. However, advances in sequencing technologies (third generation sequencing technologies) and bioinformatics tools have enabled rapid advancements since the rice genome was sequenced and assembled in 2005 [10]. Third-generation sequencing permits the development of high-quality de novo assemblies of the full genome and provides light on the remaining complex of repetitive sequences, including structural variations. In addition, isoform sequencing from third-generation sequencing technologies enables precise investigation of exons, splice sites, and alternatively spliced regions, which helps with genome annotation. It is now possible to get high-quality plant reference genomes using downstream methods such as comparative genomics, variant calling and genome-wide association studies (GWAS). These methods give comprehensive data for crop improvement. Longer reads and more accurate and contiguous genome assemblies have been made possible because to third-generation sequencing, such as single-molecule real-time sequencing (PacBio) and sequencing by Oxford Nanopore Technologies (ONT). Agricultural genome sequencing has become more relevant in recent years because to the development of third-generation sequencing technology capable of producing long reads longer than 10 kilobases (kb).

It is now possible to produce highly contiguous plant genome assemblies even for non-model crop species and smaller facilities because to long-read sequencing, long-range mapping and chromosomal conformation capture. Repetitive sequences may also be found via long-read sequencing. Large DNA molecules surpassing 250 kb may now be labelled quickly and cheaply using new optical mapping methods, such as BioNano Genomics. Hi-C (Chromosome conformation capture sequencing) is a third-generation mapping technology that depends on the physical tightness of DNA segments to be mapped. For example, chromosome phasing and scaffolding may be improved significantly when Hi-C measurements and optical mapping are used together. Reconstruction of the barley genome with a N50 of 1.9 Mb was achieved by Mascher and colleagues using short reads, optical mapping data, and chromatin interaction mapping data. Third-generation sequencing has the potential to improve genomics-based breeding approaches such as trait mapping, because to its improved sequence continuity. Use of third-generation sequencing in crop breeding has been most effective in creating enhanced, highly contiguous crop genomes. Due to intrinsic bias and inadequate repetitive sequence matching in NGS, extremely fragmented partial genome assemblies are created that make it more difficult to find and study hidden In-Dels and structural variations.

It was a common practise in crop breeding to utilise phenotypic selection and cross-breeding cycles to produce improved genotypes. Genetic diversity in agricultural species may now be identified via genomics-based breeding and leveraged to build climate-resistant crops [12]. All genes and genetic variations connected to agronomic traits may be discovered once

the genome sequences are accessible, and breeding modifications can be assessed at the genotype level once they are. Several parts of crop breeding, such as QTL mapping and GWAS, where genomic sequencing of crop populations may offer gene-level resolution of agronomic variation, are becoming more important as breeders now have access to genomic data. Genomics research is the focus of the databases in Table 1.

Table 1: Databases in use of Plant Genomics Research

Database	**URL**
Phytozome v8.0	http://www.phytozome.net/Phytozome_info.php
Gramene	http://www.gramene.org/
Home—BioProject—NCBI	http://www.ncbi.nlm.nih.gov/sites/entrez?db=bioproject
BLAST: Basic Local Alignment Search Tool	http://blast.ncbi.nlm.nih.gov/Blast.cgi
GrainGenes Class Browser	http://wheat.pw.usda.gov/cgi-bin/graingenes/browse.cgi?class=marker
PlantGDB— Resource Plant Comparative Genomics	http://www.plantgdb.org/
TreeView	http://taxonomy.zoology.gla.ac.uk/rod/treeview.html
GenBank	https://www.ncbi.nlm.nih.gov/genbank/
European Molecular Biological Laboratory (EMBL)	https://www.embl.org/
KnetMiner (Knowledge Network Miner)	https://knetminer.com/
LALIGN Server	http://www.ch.embnet.org/software/LALIGN_form.html
PopGene	http://www2.unil.ch/popgen/softwares/fstat.htm
Arlequin 3.11	http://cmpg.unibe.ch/software/arlequin3/
PRIMER-E	http://www.primer-e.com/

III. QTL MAPPING

An organism's genome, made up of all its genes and DNA, is the subject of genomics. Unlike genetics, which focuses on genes and their function in heredity, genomics studies the aggregate description and measurement of an organism's genes [13]. System biology even the most complicated biological systems may now be better understood thanks to advances in genomics. It is possible to get new insights on agricultural plant sustainability by generating genomic resources in a variety of methods, including molecular markers, transcriptome assemblies and biparental population mapping, genetic linkage maps, comparative genome mapping, and functional genomics.

It is now possible to clone QTLs, create genetic maps, and use marker-assisted selection in different segregating populations thanks to many high-throughput genotyping technologies. Genetic markers that cover a large portion of a genome may also be used to research genetic diversity in connection to natural variation, in addition to discovering specific genes linked to complex characteristics. Many species have undergone genome sequencing and extensive Expressed Sequenced Tag (EST) research, which has resulted in excellent sequence resources for the development of molecular markers. All SNP marker sets are integrated in the anticipated model, which is why this is the case. For a number of plants,

such as barley, melon, Brassica, common bean, and sunflower, computational identification of EST base single-nucleotide polymorphisms and/or EST-SNP markers for discovering sequence-tagged site markers has advanced [14, 15, 16, 17, 18]. QTLs may be predicted more accurately with the use of a meta-qtl analysis if more QTLs are detected. MetaQTL decreases the QTL's confidence interval in order to precisely anticipate the QTL's location and impact on a given sample. Low-bias QTL analysis, data visualisation, and interoperability with other genome databases are all features of both SolQTL and RASQUAL. A Meta-QTL analysis has been utilised to discover features linked with crop growth and abiotic and biotic responses in maize [19], cotton [20], soybean [21], or wheat [22]. Two drawbacks to this method are that it is difficult to distinguish between pleiotropic and physically nearby genes because of poor mapping, and only the allelic diversity found in parents of a segregating population can be analyzed.

IV. GWAS & GENOMIC SELECTION

There is an alternative to QTL mapping called GWAS. Wild populations are the basis for GWAS, whereas biparental populations obtained from controlled crossings lay the foundation for QTL analyses. Multiple recombination events may be found with greater ease, and natural variances linked with phenotypic differences can be examined with more clarity as a result. More precise GWAS mapping than QTL analysis identifies MTAs that may be linked to the amount of linkage disequilibrium (LD) across polymorphic markers across a wide range of genotypes. The breeder's preference for GWAS over QTL analysis is to evaluate a wide genetic base in order to research several potential genes for inclusion in breeding programmes. Genetically-modified organisms were first utilised in the study of complex human traits. This decade has witnessed GWAS were used to several crops including canola, rice and soybeans as well as corn and wheat [23–26]. Polygenic traits make it more difficult to pin down the source of a trait. Based on genetic estimates of breeding values (GEBV) in an individual variation, GS has the upper hand here. Biparental populations may be able to solve the issue of limited QTL translations by using the whole SNP marker collection. Lolium perenne GS-based breeding methods based on computer simulations have been found to shorten the four-year cycle of breeding. On maize breeding lines, GBS has been utilised to discover 55,000 SNP markers [27] and on elite wheat breeding lines to evaluate high yield and stem rust resistance. Using GWAS, a number of traits in five important crops are shown in Figure 2.

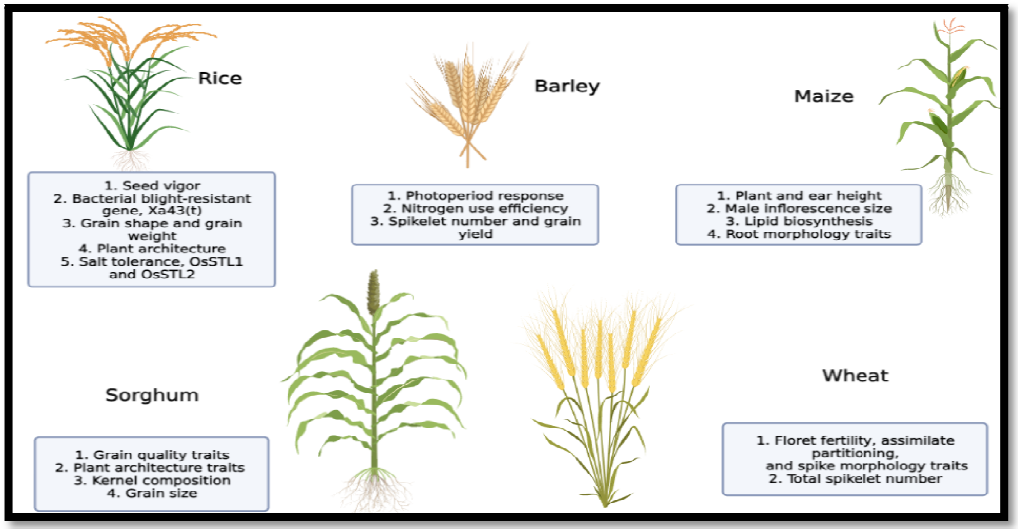

Figure 2: Few Traits or Genes Studied Via Genome Wide Association Study (GWAS) of 5 Major Crops
(Created in Biorender.com) (https://biorender.com)

V. CIS-REGULATORY ELEMENTS (CRES) FOR CROP BREEDING

Gene expressions may be controlled by regulating the cis-regulatory elements (promoters and enhancers) and regulators. CREs are related to chromatin, which binds to proteins, but they are less expressive than genes, making their discovery more difficult. For those who want to control rather than delete the gene, CRE targeting is an excellent choice. Researchers have been able to identify open chromatin regulatory sites by employing bioinformatic approaches like as ChIP-seq [28] and ATAC-seq [29] as well as DNase I hypersensitivity mapping, word-counting, and conservation-based sequence analysis.
CREs are still poorly understood, although contemporary technologies have made it easier to identify regulatory domains, but experimental study is still needed to demonstrate a single CRE's contribution to the target gene's expression. It is termed Plant Cis-Acting Regulatory Elements in Plant CARE's database of plant CREs. The suppression of the gene GRAIN WIDTH 7 due to a mutation in the rice CRE, which resulted in rice with slender grains despite its negative effect on yield, and the variation in tomato seed compartment numbers caused by the regulation of WUSCHEL (WUS) and CLAVATA (CLV3) promoters are just a few examples that have been documented so far. When a mutant library is generated utilising the expression data of mutant lines, it is expected that CREs connected to desired traits would be discovered.

VI. PROTEOMICS

Owing to post-translational modifications (PTM), function, and localization, the genome cannot be linked to mRNA and proteins due to its static nature. To understand the role proteins play in the evolution of plants, it is essential to examine their structure and interactions. Proteomics is a high-performance method for identifying and quantifying protein performance in a given cell or organism. The three basic steps in the majority of proteomics systems are identification or quantification, protein extraction, and separation. As

a consequence of recent, fast technical improvements in proteomics, we have advanced to the second generation of functional proteomics, which comprises quantitative proteomics, subcellular proteomics, different alterations, and protein-protein interactions (e.g., advances in mass spectrometry equipment and methodological developments in protein quantification). The knowledge, resolution, and coverage of the plant proteome are expanded by a variety of means. Several factors, including the availability of resources, facilities, and applications such as global or focused profiling, govern the proteome research approach. It is feasible to separate proteins with excellent reproducibility and resolution using two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), which combines two-dimensional gel electrophoresis (2-DE) with isoelectric focusing (IEF) as the first dimension and SDS-PAGE as the second. In addition, chromatographic separation techniques, including as gel filtration, ion exchange, and affinity chromatography, may be used to separate proteins based on their physicochemical features. Currently, peptide mass fingerprinting is the most used method for identifying proteins. It starts with the breakdown of proteins into peptides, followed by the exact mass determination of the peptides using mass spectrometry (MS). In-gel electrophoresis was developed to avoid the 2D-PAGE restrictions of gel-to-gel variance and restricted repeatability (DIGE). DIGE is used to understand how protein expression changes in response to biotic and abiotic stimuli. Two-dimensional gel electrophoresis is expanded to three dimensions to prevent co-migration interferences. It offers very precise identification of proteins and PTMs using two distinct buffers with different ion carriers [33].

MS identifies proteins based on peptide mass and fragmentation (MS/MS) data using a range of computer techniques. There are three phases in all. In order to convert molecules into gas-phase ions, mass-based ion separation is performed in an electro or magnetic field, followed by measurement of the separated ions with a certain m/z value. Ionizations techniques include electrospray ionization (ESI), surface-enhanced laser desorption/ionization (SELDI), and matrix-assisted laser desorption ionizations (MALDI). Gel-free techniques, such as quantitative approaches, tag-based labelling, metabolic labelling, and label-free methods, may mitigate the disadvantages of gel-based methods, such as their inability to segregate the whole proteome and poor identification of less abundant proteins.

Quantitative proteomics is also required for the finding of important proteome alterations, such as expression, interaction, and modification that are related with genetic differences and/or observable phenotypic changes. To correctly differentiate between proteins prior to 2-D electrophoresis in DIGE (Differential Gel Electrophoresis), protein samples are tagged with fluorescent dyes. ICAT (Isotope-Coded Affinity Tagging) use in vitro isotopic labelling to quantify protein, with labelled tryptic peptides separated by chromatography and subsequently identified by mass spectrometry. Using isobaric tags, iTRAQ (Isobaric Tagging for Relative and Absolute Quantification) measures proteins. Breeders use this method to identify markers for biotic and abiotic stressors in order to develop genetically modified crops. Stable Isotope Labelling by Amino Acid in Cell Culture (SILAC) employs in vivo labelling of cell populations maintained in N14 or N15 media and has been shown to be successful in identifying proteome anomalies induced by post-translational changes under stress [34]. For complicated multidimensional protein analysis, MudPIT (Multi-Dimensional Protein Identification Technology) is used. After separating digested proteins using biphasic or triphasic microcapillary columns, tandem mass spectrometry is performed. Using this technology, the processes that regulate the quantity of rice tillers have been uncovered.

Figure 3 depicts the approaches used in proteome studies, whereas Table 2 lists the main databases utilized in proteomic research.

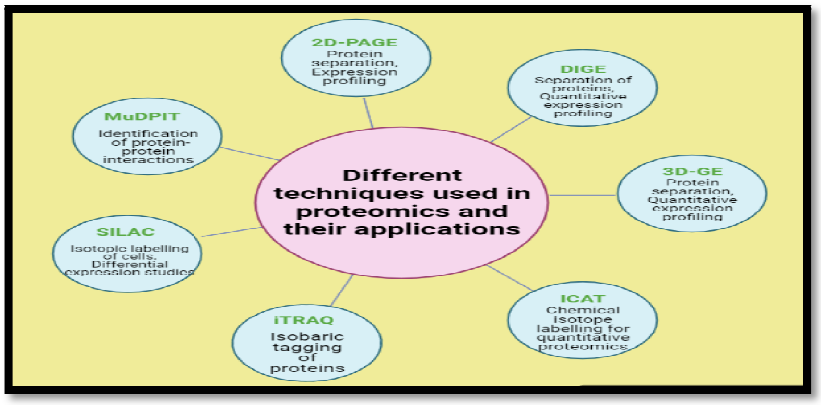

Figure 3: Different Techniques used in Proteomic Studies (Created in Biorender.com) (https://biorender.com)

Table 2: Databases for Proteomic study

Database	Link
Swiss Institute of Bioinformatics' Expasy SWISS-2DPAGE database	http://au.expasy.org/ch2d/
Kazusa DNA Research Institute's Cyano2Dbase	http://bacteria.kazusa.or.jp/cyano_legacy/Synechocystis/cyano2D/index.html
rice proteome database	http://gene64.dna.affrc.go.jp/RPD/
Nottingham *Arabidopsis* Stock Centre (NASC) Proteomics database	http://proteomics.*Arabidopsis*.info/
SUB-cellular location database for *Arabidopsis* proteins (SUBA)	http://suba.plantenergy.uwa.edu.au/
The soybean proteome database	http://proteome.dc.affrc.go.jp/cgi-bin/2d/2d_view_map.cgi
The *Arabidopsis* Protein Phosphorylation Site Database (PhosPhAt)	http://phosphat.mpimp-golm.mpg.de/
Protein data bank, PDB	http://www.pdb.org/pdb/home/home.do
The RIKEN SGPI	http://www.rsgi.riken.go.jp/rsgi_e/index.html
Genomes TO Protein structures and functions (GTOP) database	http://spock.genes.nig.ac.jp/~genome/gtop.html
CATH	http://www.cathdb.info/
Structural Classification of Proteins (SCOP) database	http://scop.mrc-lmb.cam.ac.uk/scop/
PRoteomics IDEntification database (PRIDE)	https://www.ebi.ac.uk/pride/
Peptide Atlas	http://www.peptideatlas.org/
Mass Spectrometry Interactive Virtual Environment (MassIVE)	https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp
Plant Proteomics Database (PPDB)	http://ppdb.tc.cornell.edu/
1001 Proteomes (Discontinued)	https://www.heazleome.org/tools.html
GelMap	https://www.gelmap.de/
Peptide Atlas SRM Experiment Library (PASSEL)	http://www.peptideatlas.org/passel/

VII. TRANSCRIPTOMICS

The genome, as previously said, is fixed and hence unable to reflect the level of gene expression. As a result, the expression level of the genome may be measured using transcriptomic methods. Transcription makes up around 1-2% of the functioning genome. To discover cis-regulatory patterns in gene expression, predict gene function, and screen potential new genes, transcriptomics uses high-throughput gene expression analysis. This expressed genome may be studied using transcriptomics, which studies how genes are expressed in an organism in a variety of situations, tissues (spatial transcriptome), and time periods (temporal transcriptome). These approaches, such as microarrays and GeneChips, may offer complete gene expression profiles for a wide range of species, as is well known. It is becoming more effective to sequence short snippets of expressed RNAs, including sRNAs, in genome-sequenced species. Co-expression and comparative studies may benefit from increased public datasets that have been developed as a result of recent initiatives in the area of transcriptomics.

In the 1970s and 1980s, reverse transcriptase was used to convert cDNA into RNA transcripts in the silk moth [35], and in the 1990s, Sanger sequencing was used to sequence RNA transcripts as expressed sequence tags (ESTs), which are basically used to estimate the gene composition of an organism [36]. After random sequencing in an unbiased cDNA library, ESTs are clustered into groups of transcript sequences using sequence-clustering and/or assembly approaches. Next, the number of ESTs with unique identifiers for each cDNA library and/or sequence cluster is tallied to estimate the quantity of transcripts expressed in each tissue. This concept has also been used in the digital differential display (DDD) tool of the NCBI's UniGene database, which has been utilised in substantial cDNA research for several taxa, including plants. Later, northern blotting and quantitative reverse transcription polymerase chain reaction (qRT-PCR) were utilised to quantify RNA transcripts. Since none of these methods addressed the complete transcriptome, the Sequencing-based Serial Analysis of Gene Expression (SAGE) was developed in 1995 [37]. More than 10 short specific tags (13–15 bp) are concatenated and cloned from each mRNA present in a sample to generate a SAGE library. The sequencing of selected clones from the SAGE library makes the efficient collection of transcript tag sequences feasible. To identify the genes corresponding to each SAGE tag, a dataset of genome sequences or a large collection of expressed sequence tags (ESTs) is required. Several versions of the fundamental protocol (MAGE, SADE, microSAGE, miniSAGE, longSAGE, superSAGE, deepSAGE, 5 ′ SAGE, etc.) have been developed to improve and expand the value of SAGE.

Massive parallel signature sequencing is another sequencing-based technique (MPSS). MPSS uses a 17–20 bp signature sequence near to the 3' end to identify mRNA. Initially, each distinctive sequence is cloned onto microbeads. This approach guarantees that a microbead has just one kind of DNA sequence. For sequencing and measuring, the flow cell comprises an array of microbeads. The signature sequences (MPSS tags) of an MPSS dataset are evaluated, compared to all other signatures, and the number of signatures with similar sequences is counted. Accessible online at http://mpss.udel.edu are databases containing MPSS information on various plant species, including *Arabidopsis*, rice, grapes, and *Magnaporthe grisea* (rice blast fungus). In *Arabidopsis*, high-density TSS mapping was performed utilising the newly published CT-MPSS method for quantitative investigation of the 5 ′ end of transcripts coupled with the cap-trapper strategy for full-length cDNA cloning.

The data set of *Arabidopsis* CT-MPSS tags is accessible through the plant promoter database ppdb (http://www.ppdb.gene.nagoya-u.ac.jp), which provides rice and *Arabidopsis* promoter annotation. Many databases for plant transcriptome research are included in Table 3.

Table 3: Different Databases for Plant Transcriptomic Study

Database	Link
Ppdb	http://www.ppdb.gene.nagoya-u.ac.jp
ArrayExpress	https://www.ebi.ac.uk/arrayexpress/
ATTED II	http://atted.jp/
Genevestigator	https://www.genevestigator.com/gv/index.jsp
Arabidopsis Gene Expression Database AREX	http://www.arexdb.org/index.jsp
RICEATLAS	http://bioinformatics.med.yale.edu/riceatlas/

VIII. METABOLOMICS

Metabolomics is the comprehensive and multidimensional study of metabolism that identifies metabolites by using a variety of analytical methods and bioinformation. Metabolomics is the study of metabolism. It is possible to compare the metabolomes of different plants, although this is considerably more difficult. Chemical-level phenotyping and diagnostic assessment is inferior to metabolomics since metabolomics are able to simultaneously examine a huge number of metabolites and quantitatively analyse each one of them. Researchers may get a better understanding of how cells react to changes in their internal and external environments by using comprehensive metabolic profile data sets. Genetic variants alter metabolic profiles, and chemical phenotypes may be utilised to identify genes involved in certain pathways. Recent years have seen several technological advancements in metabolomics equipment. Analyzing metabolomics data begins with the collection of metabolic fingerprints from various analytical methods. Some of the sample categorization methods used in conjunction with MS include gas chromatography (GC), high-performance or ultra-performance liquid chromatography (LC), and capillary electrophoresis (CE). CEMS is particularly useful for biological component isolation and analysis because of its high sensitivity [38].

Data processing in metabolomics is essential to determining biological significance. Principal component analysis (PCA), hierarchical cluster analysis (HCA), and self-organization mapping (SOM) are often used to classify samples and/or metabolites. Gene expression profiles of certain genes encoding enzymes engaged in certain pathways are used in combination with the visualisation of metabolic profile on metabolic maps. The study of plants' metabolic processes is a difficult one, yet it is necessary if we are to fully appreciate the growth and development of plants. It is possible to get a better knowledge of plant cell systems via the use of metabolomics. Our understanding of plant cell processes via metabolomics may help us develop molecular breeding to increase plant productivity and functionality in areas such as stress tolerance, pharmaceutical manufacturing, functional meals and biomaterial production.

IX. SYSTEMS BIOLOGY

Several plant species have produced vast amounts of data from sources like the genome, transcriptome, proteome, metabolome, and epigenome in recent years. It was not able to fully comprehend the molecular foundations of complex traits and biological networks due to their independent study. A systems biology technique including the integration of diverse omics data, modelling, and prediction of cellular processes is necessary to comprehend the flow of biological information underlying complex characteristics. This technique permits a comprehensive understanding of the dynamic system in which many levels of biological structure interact with the external environment to exhibit phenotypic. The fundamental objectives of crop biology research are the maximisation of production and the minimization of losses resulting from a range of stress situations. The solution is also complicated due to the complexity of the subject. The integration of transcriptomics, proteomics, and metabolomics considerably simplifies the discovery and investigation of complex plant regulatory networks. Consequently, systems biology emerges as an intriguing multidisciplinary topic of study that combines large amounts of omics data with well-developed mathematical models to test hypotheses and forecast biological systems. The processing, scaling, and analysis of multidimensional datasets in order to extract relevant biological discoveries remains the major barrier to omics data integration. For the integration and analysis of datasets produced by several platforms, it is important to gather, prepare, standardise, and integrate data into a single matrix. Then, clusters of genes, proteins, and metabolites with similar structures were identified. Multiple systems exist to aggregate multidimensional omics data, such as mixOmics, OnPLS modelling, Integromics, sparse Multi-Block Partial Least Squares, and COVAIN. These methods provide the study of plant metabolism and the knowledge of the molecular processes behind agronomically significant plant phenotypes. To identify light-specific metabolic and regulatory markers in rice [40], transcriptomics, metabolomics, and genome-scale computational modelling were used. Transcriptomics, proteomics, and metabolomics data were evaluated in 2020 to supplement information that had previously provided insight on the processes behind the fertility change in a thermosensitive male sterile line of pigeon peas for use in two-line hybrid breeding [41]. Given that phenotypic variance is not just determined by DNA but also by biological regulation in response to the environment, multiomics data are widely used for phenotypic prediction. Reconstructing pathways and networks utilising transcriptome, proteome, and metabolome data may aid in the understanding of these regulatory networks and their functional interaction with biological entities. The right normalisation of omics data yields a similarity matrix, which is then converted into an adjacency matrix and, lastly, a directed graph or network abstraction. Global gene co-expression networks are a potential tool for exploring and predicting specialised metabolite pathways with a high throughput. The last phase in network biology is dynamic modelling, which provides an exhaustive understanding of how gene expression influences protein activity in plants in response to environmental stimuli. By bridging the gap between genotype and phenotype and gaining an understanding of the complexity of multiple traits, systems biology offers tremendous potential for sustainable agriculture. It is useful for modelling and evaluating multigenic traits linked with agricultural production, such as plant architecture, nitrogen use efficiency, water use efficiency, and resistance to biotic and abiotic stress. Due to recent developments in high-throughput experimental analysis and computer capacity, it is now possible to integrate many fields to explain any complicated characteristic. Using well-designed mathematical models

based on time series data, one may develop a systems biology-based breeding strategy by finding possible candidate genes for use in breeding programmes.

X. MACHINE LEARNING

Machine Learning (ML) is a subfield of computer science that uses statistical and mathematical techniques to train models without direct programming [42]. ML develops a variety of algorithms that learn from sample data and train the predictive model. Samuel [43]. ML is the study of programming computers to learn from data. By simplifying functional annotation of genomes and allowing real-time, high-throughput phenotyping of agronomic traits in the greenhouse and field, machine learning helps the discovery of agronomically valuable economic regions. ML is a technique to data analysis that enables computers to learn patterns over time. ML models for GS have the benefit of learning the pattern directly from the data, enabling them to account for all variations, interactions, and environmental factors. For huge, heterogeneous, and formless datasets, such as those produced by optical imaging or sequencing, ML may provide significant benefits over traditional analytic approaches. Crop breeders may use machine learning to rapidly phenotype plants and to examine massive databases for patterns, such as DNA sequence-to-characteristic connections. Machine learning algorithms may employ high-throughput phenotyping and genomic data to automate elements of the gene discovery process that are presently difficult to automate, such as genome labelling and picture interpretation. Figure 4 depicts the fundamental picture interpretation procedure. Although several research have used machine learning (ML) for GS, the subject of deep learning (DL) has yet to be thoroughly investigated.

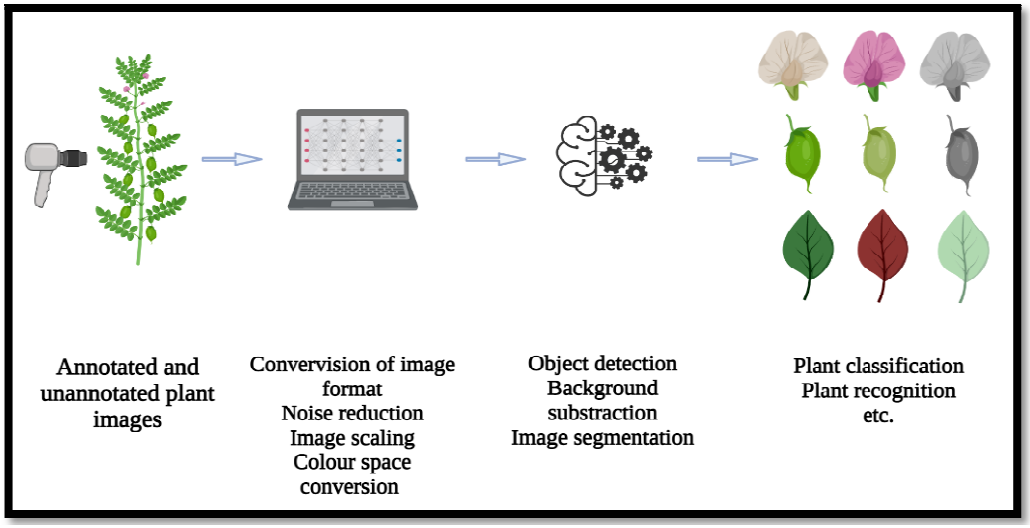

Figure 4: Basic Workflow of Image Interpretation

1. **High throughput crop phenotyping:** It is critical for the study of relationships and crop improvement that plant phenotyping be used to evaluate functional and structural characteristics at the cellular and organism levels. Plant phenotypes are becoming more important in the interpretation of genetic data as genomics research and sequencing technology improve at a rapid pace. Conventional phenotyping is often a bottleneck that limits the number of features and crops that may be assessed since it is subjective, error-

prone, labour expensive and time consuming. High-throughput imaging and automatic sensors, along with machine learning, have enabled robotic high-throughput phenotyping to be established, overcoming the limitations of conventional human-based phenotyping by enabling the rapid generation of phenotypical features and features across large populations [2]. [2]. Image or sensor detection, phenotypic data classification, feature quantification, and forecasting based on specific models or algorithms are four key features of high-throughput phenotyping (Figure 3). High-throughput phenotyping was evaluated in the field by Jose Luis Araus and Jill E. Cairns, respectively. An unsupervised identification technique was used to measure, appraise, and categorise the severity of Glycine max foliar stressors, including bacterial and fungal infections, as well as nutritional deficiency, according to a recent study [44]. Machine learning requires large datasets for training and model building. Non-significant and problematic predictions may be made with a small training set but it is costly and time-consuming to collect large datasets when crops only get measured once a year. A very limited number of research institutes and organisations have the ability to do ML-based phenotyping with high throughput. It will be important to substantially reduce acquisition and operational costs in order to make ML-based phenotyping widely applicable on future farming.

2. **Machine learning in crop genomics research:** Several applications of machine learning include genome assembly, recurrent inference of gene regulatory networks, and identification of genuine Single Nucleotide Polymorphisms (SNPs) in polyploid plants. Optimizing polyploid genomic assemblies with complicated redundancy may be achieved via the application of machine learning. Ma et al. [45] provide a detailed review of machine learning algorithms and associated open-source R tools relevant for plant data analysis. A comprehensive genome assembly and annotation provides the basis for monitoring genetic changes within a plant species and for understanding the shape and function of plant genes, both of which are essential steps in the process of agricultural trait discovery. For interactive inference of gene regulatory networks, ML-based methods that can incorporate diverse types of regulatory signals from multiple data sources have gained popularity. Consequently, inferring regulatory element-gene links is a potential field for uncovering unexplored crop improvement opportunities.

   GWAS is currently one of the most often utilised approaches for detecting MTA in plant species. Traditional GWASs are excellent for identifying SNP markers with considerable effects on complex traits, but they may ignore a variety of interconnected biological processes and mechanisms that influence the phenotypic of complex traits simultaneously. Variable significance values may be used to identify high-resolution variant-trait associations in ML-mediated GWASs. The implementation of this important genetic strategy in practical plant-breeding programmes may be enhanced by using complicated mathematical approaches such as machine learning (ML) algorithms. (ML-based GWAS for Identifying QTL Underlying Soybean Yield and Its Components)

3. **Deep learning:** In the genomics era, multifaceted molecular phenotypes involved in information relay, namely the structure, modification, function, and evolution of elements in DNA, RNA, and protein, along with their interactions, are beginning to be revealed at scale and even at lower cost, allowing fine-grained evaluation of information transfer and transformation along Francis Crick's 1957 "central dogma" [46]. In data mining, it has been shown that deep learning models are very effective in predicting molecular

phenotypes from upstream molecular phenotypes or directly from genomic DNA sequences.

Deep learning is a subfield of machine learning that focuses on densely connected, artificial neural network-trained networks. Artificial Neural Networks (ANN) are well-known strategies for dealing with machine learning issues that have been studied since the 1940s and are based on the nervous systems of animals [45]. A single artificial neural network (ANN) consists of several hidden layers and one input. Deep Neural Network (DNN) is a new machine learning discipline and a type of artificial neural network. DNNs differ from ANNs in that they contain many more hidden layers; hence, the quantity of data needed increases as the DNN's predictive ability increases. In genomics, transcriptomics, proteomics, metabolomics, and systems biology, deep learning has been used to address complicated biological challenges.

Deep learning, which utilises a high number of neurons and models such as CNN, RNN, and MLP, is applicable to GS [47]. The input layer of these models consists of marker data, whereas the output layer contains responses with several hidden layers. The optimal model performance is determined by hyperparameter selection, which is a time-consuming and computationally costly process. The ability of deep learning models to generate ab initio forecasts on unique, previously unknown sequence data (data not within the training set) is perhaps the most notable characteristic, which has numerous important ramifications, whereas the number of high-capacity and trainable characteristics is the most advantageous. Despite the huge number of genetic variations in a real population, deep learning models can only be trained on a small subset of them to predict the effects of all other variants (i.e., the whole mutation space). Knowledge may move from well-studied species (such as *Arabidopsis*) to closely related but less well-studied species (such as *Arabidopsis*) (such as other species in the Brassicaceae). When many variants within a crucial coding region (such as a QTL for a certain trait) are in tight linkage disequilibrium, we may utilise in silico mutagenesis to transfer them from one haplotype to the next, therefore prioritising causative variants. Such a break in linkage disequilibrium would be labor-intensive and difficult to scale up in wet lab research, and practically impossible in nature. Using a large collection of deep learning models, each targeting a different molecular phenotype, or a multi-task learning model addressing multiple molecular phenotypes simultaneously, it is possible to predict not only the causative mutation underlying a QTL, but also its likely molecular mechanism. Importantly, while using the breeding-by-editing approach, we are no longer restricted to the known beneficial natural variations. Instead, we have unrestricted flexibility to design different beneficial alleles based on the 'knowledge' of the biological processes of interest possessed by our deep learning algorithms. Rodrguez-Leal et al. [48] altered the promoter of the tomato CLAVATA3 gene (SlCLV3) to improve fruit size and inflorescence branching. Utilizing generative models in synthetic biology is another way for producing genetic components with defined functionality. Despite the growing interest in generative models like variational autoencoders and generative adversarial networks, their applications in synthetic biology remain restricted. Using GANs to construct synthetic DNA sequences encoding for antimicrobial peptides is one example.

XI. CONCLUSION

It is crucial to adapt plant breeding curriculum to the digital age. Researchers and breeders must find a balance between machine-generated suggestions and farmer desires. Developing information for plant breeding is ineffective if researchers lack the capacity to use that knowledge. Information–action strategies, which integrate additional abilities and viewpoints to facilitate the development of knowledge for enhanced breeding and smarter farming, are necessary. Agriculture will depend on Next-Generation AI techniques to make judgments and recommendations based on massive data that is indicative of the environment and the systems biology of a plant. Breeding will be able to perform at greater levels than ever before because to Next-Gen AI's capacity to utilise diverse and complex data in an effective manner. The use of ML and DL has led to significant phenomics and genomics findings. As promising as these discoveries are, they are not yet adequate to contemplate depending only on technology to accelerate the breeding process, which remains a difficult, time-consuming, and costly endeavour. Despite gains in the efficiency of data generation, the plant research community still confronts difficulties with translational procedures. In isolation, genomes, epigenomics, transcriptomics, proteomics, metabolomics, and phenomics continue to be largely distinct fields of study that provide scant insight. To expedite plant development, it is necessary to concurrently use and integrate multi-omics data. Utilizing enormous quantities of genetic data from a variety of sources and formats for crop development is fraught with considerable difficulties in agriculture. To address these problems, novel breeding tactics and bioinformatics technology must be used to turn genetic data into advances in agricultural production and yield stability. Using meta-QTL analysis, GWAS, and genetic screens, researchers may uncover significant gene-trait connections more quickly. While genome editing is an effective method for rapidly introducing beneficial mutations into champion crops, GS enhances selection efficiency without needing knowledge of genetic drivers. ML algorithms may employ high-throughput phenotyping and genomic data to automate difficult-to-automate aspects of the gene discovery process, such as genome annotation and image interpretation. Combining new technologies and methods will allow future plant breeding to achieve the crop growth rate necessary for food security.

LIST OF ABBREVIATIONS

BAC	Bacterial Artificial Chromosomes
GWAS	Genome-Wide Association Studies
QTL	Quantitative Trait Loci
DNA	Deoxyribonucleic Acid
RNA	Ribonucleic Acid
ML	Machine Learning
GS	Genomic Selection
DL	Deep Learning
NGS	Next-Generation Sequencing
EST	Expressed Sequence Tag
MTA	Marker-Trait Associations
LD	Linkage Disequilibrium
GBS	Genomic Based Selection
CREs	Cis-Regulatory Elements
MS	Mass Spectrometry

cDNA	Complementary DNA
SAGE	Serial Analysis of Gene Expression
PAGE	polyacrylamide gel electrophoresis
SNP	Single Nucleotide Polymorphism
DNN	Deep Neural Network
ANN	Artificial Neural Networks
CNN	Convolutional Neural Networks

REFERENCES

[1] S. Satpathy, S. Debnath, and A. Mishra, "Study on character association *Lens culinaris* Medik.," Electron. J. Plant Breed, vol. 12, pp. 58-65, April, 2021.
[2] Z. An, C. Wang, B. Raj, S. Eswaran, R. Raseed, S. Debnath, et al., "Application of New Technology of Intelligent Robot Plant Protection in Ecological Agriculture," J. Food Qual, vol. 1, pp. 1-7, April, 2022.
[3] N. Rajan, S. Debnath, A.K. Dutta, B. Pandey, A.K. Singh, R.K. Singh, "Elucidation of Nature of Gene Action and Estimation of Combining Ability Effects for Fruit Yield Improvement and Yield Attributing Traits in Brinjal Landraces," J. Food Qual, vol. 12, April 2022.
[4] R. Velasco, A. Zharkikh, M. Troggio, D.A. Cartwright, A. Cestaro, D. Pruss, et al., "A high quality draft consensus sequence of the genome of a heterozygous grapevine variety," PloS one, vol. 2, pp. e1326, December 2007.
[5] T.P. Michael, and S. Jackson, "The first 50 plant genomes," Plant Genome, vol. 6, July 2013.
[6] T. Toyoda, and A. Wada, "Omic space: coordinate-based integration and analysis of genomic phenomic interactions," Bioinformatics, vol. 20, pp. 1759-1765, July 2004.
[7] J.M. González-Camacho, L. Ornella, P. Pérez-Rodríguez, D. Gianola, S. Dreisigacker, and J. Crossa, "Applications of machine learning methods to genomic selection in breeding wheat for rust resistance," Plant Genome, vol. 11, pp. 170104, July 2018.
[8] S. Huang, R. Li, Z. Zhang, L.I. Li, X. Gu, W. Fan, et al., "The genome of the cucumber, *Cucumis sativus* L.," Nat. genet, vol. 41, pp. 1275-1281, December 2009.
[9] H. Wang, E. Cimen, N. Singh, and E. Buckler, "Deep learning for plant genomics and crop improvement," Curr. Opin. Plant Biol, vol. 24, pp. 34-41, April 2020.
[10] T. Sasaki, "The map-based sequence of the rice genome," Nature, vol. 436, pp. 793-800, August 2005.
[11] M. Mascher, H. Gundlach, A. Himmelbach, S. Beier, S.O. Twardziok, T. Wicker, et al., "A chromosome conformation capture ordered sequence of the barley genome," Nature, vol. 544, pp. 427-433, April 2017.
[12] M. Mousavi-Derazmahalleh, P.E. Bayer, J.K. Hane, B. Valliyodan, H.T. Nguyen, M.N. Nelson, et al., "Adapting legume crops to climate change using genomic approaches," Plant Cell Environ, vol. 42, pp. 6-19, January 2019.
[13] S.P. Chand, S. Debnath, M. Rahimi, M.S. Ashraf, P. Bhatt, S.A. Rahin, "Contextualization of Trait Nexus and Gene Action for Quantitative and Qualitative Characteristics in Indian Mustard," J. Food Qual, vol. 2022, pp. 1-24, May 2022.
[14] M.W. Blair, M.M. Torres, M.C. Giraldo, and F. Pedraza, "Development and diversity of Andean-derived, gene-based microsatellites for common bean (*Phaseolus vulgaris* L.)," BMC Plant Biol, vol. 9, pp. 1-4, December 2009.
[15] A. Heesacker, V.K. Kishore, W. Gao, S. Tang, J.M. Kolkman, A. Gingle, et al., "SSRs and INDELs mined from the sunflower EST database: abundance, polymorphisms, and cross-taxa utility," Theor. Appl. Genet, vol. 117, pp. 1021-1029, November 2008.
[16] R.V. Kantety, M. La Rota, D.E. Matthews, and M.E. Sorrells, "Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat," Plant Mol. Biol, vol. 48, pp. 501-510, March 2002.

[17] S. Kaur, N.O. Cogan, G. Ye, R.C. Baillie, M.L. Hand, A.E. Ling, et al., "Genetic map construction and QTL mapping of resistance to blackleg (*Leptosphaeria maculans*) disease in Australian canola (*Brassica napus* L.) cultivars," Theor. Appl. Genet, vol. 120, pp. 71-83, December 2009.
[18] R. Kota, R.K. Varshney, T. Thiel, K.J. Dehmer, and A. Graner, "Generation and comparison of EST-derived SSRs and SNPs in barley (*Hordeum vulgare* L.)," Hereditas, vol. 135, pp. 145-151, December 2001.
[19] R. Liu, H. Zhang, P. Zhao, Z. Zhang, W. Liang, Z. Tian, et al., "Mining of candidate maize genes for nitrogen use efficiency by integrating gene expression and QTL data," Plant Mol. Biol. Rep, vol. 30, pp. 297-308, April 2012.
[20] J.I. Said, Z. Lin, X. Zhang, M. Song, and J. Zhang, "A comprehensive meta QTL analysis for fiber quality, yield, yield related and morphological traits, drought tolerance, and disease resistance in tetraploid cotton," BMC Genom, vol. 14, pp. 1-22, December 2013.
[21] Z.M. Qi, Q. Wu, X. Han, Y.N. Sun, X.Y. Du, C.Y. Liu, et al., "Soybean oil content QTL mapping and integrating with meta-analysis method for mining genes," Euphytica, vol. 179, pp. 499-514, June 2011.
[22] E. Hanocq, A. Laperche, O. Jaminon, A.L. Lainé, J. Le Gouis, "Most significant genome regions involved in the control of earliness traits in bread wheat, as revealed by QTL meta-analysis," Theor. Appl. Genet, vol. 114, pp. 569-584, February 2007.
[23] K. Gacek, P.E. Bayer, I. Bartkowiak-Broda, L. Szala, J. Bocianowski, D. Edwards, et al., "Genome-wide association study of genetic control of seed fatty acid biosynthesis in *Brassica napus*," Front. Plant Sci, vol. 7, pp. 2062, January 2017.
[24] X. Huang, T. Sang, Q. Zhao, Q. Feng, Y. Zhao, C. Li, et al., "Genome-wide association studies of 14 agronomic traits in rice landraces," Nat. Genet, vol. 42, pp. 961-967, November 2010.
[25] H. Sonah, L. O'Donoughue, E. Cober, I. Rajcan, and F. Belzile, "Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean," Plant Biotechnol. J, vol. 13, pp. 211-221, February 2015.
[26] F. Tian, P.J. Bradbury, P.J. Brown, H. Hung, Q. Sun, S. Flint-Garcia, et al., "Genome-wide association study of leaf architecture in the maize nested association mapping population," Nat. Genet, vol. 43, pp. 159-162, February 2011.
[27] J.E. Rutkoski, J.A. Poland, R.P. Singh, J. Huerta-Espino, S. Bhavani, H. Barbier, et al., "Genomic selection for quantitative adult plant stem rust resistance in wheat," Plant Genome, vol. 7, pp. plantgenome2014-02, November 2014.
[28] D.S. Johnson, A. Mortazavi, R.M. Myers, and B. Wold, "Genome-wide mapping of in vivo protein-DNA interactions," Sci, vol. 316, pp. 1497-1502, June 2007.
[29] J.D. Buenrostro, B. Wu, H.Y. Chang, and W.J. Greenleaf, "ATAC-seq: a method for assaying chromatin accessibility genome-wide," Curr. Protoc. Mol. Biol, vol. 109, pp. 21-29, January 2015.
[30] F.M. Pauler, S.H. Stricker, K.E. Warczok, and D.P. Barlow, "Long-range DNase I hypersensitivity mapping reveals the imprinted Igf2r and Air promoters share cis-regulatory elements," Genome Research, vol. 15, pp. 1379-1387, October 2005.
[31] S. Rombauts, K. Florquin, M. Lescot, K. Marchal, P. Rouzé, and Y. Van de Peer, "Computational approaches to identify promoters and cis-regulatory elements in plant genomes," Plant Physiol, vol. 132, pp. 1162-1176, July 2003.
[32] J.V. de Velde, K.S. Heyndrickx, and K. Vandepoele, "Inference of transcriptional networks in *Arabidopsis* through conserved noncoding sequence analysis, Plant Cell, vol. 26, pp. 2729-2745, July 2014.
[33] T. Rabilloud, "When 2 D is not enough, go for an extra dimension," Proteomics, vol. 13, pp. 2065-2068, July 2013.
[34] G. Mastrobuoni, S. Irgang, M. Pietzke, H.E. Aßmus, M. Wenzel, W.X. Schulze, et al., "Proteome dynamics and early salt stress response of the photosynthetic organism *Chlamydomonas reinhardtii*," BMC Genom, vol. 13, pp. 1-3, December 2012.

[35] G.K. Sim, F.C. Kafatos, C.W. Jones, M.D. Koehler, A. Efstratiadis, and T. Maniatis, "Use of a cDNA library for studies on evolution and developmental expression of the chorion multigene families," Cell, vol. 18, pp. 1303-1316, December 1979.
[36] R. Lowe, N. Shirley, M. Bleackley, S. Dolan, and T. Shafee, "Transcriptomics technologies," PLoS Comput. Biol, vol. 13, pp. e1005457, May 2017.
[37] V.E. Velculescu, L. Zhang, B. Vogelstein, and K.W. Kinzler, "Serial analysis of gene expression," Sci, vol. 270, pp. 484-487, October 1995.
[38] R. Ramautar, G.W. Somsen, and G.J. de Jong, "CE-MS in metabolomics," Electrophoresis, vol. 30, pp. 276-291, January 2009.
[39] M.Y. Hirai, M. Yano, D.B. Goodenowe, S. Kanaya, T. Kimura, M. Awazuhara, et al., "Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*," PNAS, vol. 101, pp. 10205-10210, July 2004.
[40] M. Lakshmanan, S.H. Lim, B. Mohanty, J.K. Kim, S.H. Ha, and D.Y. Lee, "Unraveling the light-specific metabolic and regulatory signatures of rice through combined in silico modeling and multiomics analysis," Plant Physiol, vol. 169, pp. 3002-3020, December 2015.
[41] L.T. Pazhamala, P. Chaturvedi, P. Bajaj, S. Srikanth, A. Ghatak, A. et al., "Multiomics approach unravels fertility transition in a pigeonpea line for a two-line hybrid system," Plant Genome, vol. 13, pp. e20028, July 2020.
[42] A. Singh, G. Vaidya, V. Jagota, D.A. Darko, R.K. Agarwal, S, Debnath, et al., "Recent Advancement in Postharvest Loss Mitigation and Quality Management of Fruits and Vegetables Using Machine Learning Frameworks," J. Food Qual, vol. 2022, pp. 1-9, June 2022.
[43] A.L. Samuel, "Some studies in machine learning using the game of checkers," IBM J. Res. Dev, vol. 44, pp. 206-226, January 2000.
[44] S. Ghosal, D. Blystone, A.K. Singh, B. Ganapathysubramanian, A. Singh, and S. Sarkar, "An explainable deep machine vision framework for plant stress phenotyping," PNAS, vol. 115, pp. 4613-4618, May 2018.
[45] H. Wang, E. Cimen, N. Singh, and E. Buckler, "Deep learning for plant genomics and crop improvement," Curr. Opin. Plant Biol, vol. 24, pp. 34-41, April 2020.
[46] F. Crick, "Central dogma of molecular biology," Nature, vol. 227, pp. 561-563, August 1970.
[47] A.K. Singh, I.R. Khan, S. Khan, K. Pant, S. Debnath, and S. Miah, "Multichannel CNN Model for Biomedical Entity Reorganization," Biomed Res. Int, vol. 2022, pp. 1-11.
[48] D. Rodríguez-Leal, Z.H. Lemmon, J. Man, M.E. Bartlett, and Z.B. Lippman, "Engineering quantitative trait variation for crop improvement by genome editing," Cell, vol. 171, pp. 470-480, October 2017.

