Journal of Advances in Biology & Biotechnology

Volume 27, Issue 4, Page 27-49, 2024; Article no.JABB.113926
ISSN: 2394-1081

From Code to Crop: How Bioinformatics is Transforming Crop Genomics in Modern Agriculture and Bettering Environment

Prajval V [a++], A.Krishnamoorthi [b#*], Riya Thakur [c†], Ruchitha T [d#], Manish Kapoor [e‡], Anushi [f#], Abhishek Singh [g#] and Karthik Chittibomma [h^]

[a] GPS Institute of Agricultural Management, Bengaluru, India.
[b] NBPGR Pusa campus, IARI New Delhi -110012, India.
[c] KVK, Chhindwara, MP, India.
[d] Department of Fruit science, University of Horticultural Sciences, Bagalkot, India.
[e] Department of Botany, Punjabi University Patiala, India.
[f] Department of Fruit Science, Chandra Shekhar Azad University of Agriculture and Technology, Kanpur, 208002, India.
[g] Department of Agricultural Economics, Chandra Shekhar Azad University of Agriculture and Technology, Kanpur, 208002, India.
[h] Department of Genetics and Plant Breeding, College of Agriculture, Central Agricultural University, Imphal, Manipur-795004, India.

Author's contribution

The sole author designed, analysed, interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/JABB/2024/v27i4737

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/113926

Review Article

Received: 09/01/2024
Accepted: 13/03/2024
Published: 16/03/2024

[++] Assistant Professor;
[#] Ph.D. Scholar;
[†] Scientist Horticulture;
[‡] Professor;
[^] Research Scholar;
**Corresponding author: E-mail: krishnavivaan007@gmail.com;*

ABSTRACT

Advancements in bioinformatics have ushered in a new era in modern agriculture by seamlessly integrating computational methodologies with crop genomics. This paper explores the transformative impact of bioinformatics on crop genomics and its consequential benefits for both agriculture and the environment. Bioinformatics tools and techniques are instrumental in decoding the intricate genetic information of crops, leading to the development of tailored solutions for improved crop yield, resistance to diseases, and environmental sustainability. The first aspect of this transformation involves the utilization of high-throughput sequencing technologies, which generate vast amounts of genomic data. Bioinformatics algorithms play a crucial role in processing and analysing these data sets, enabling researchers to unravel the complex genetic codes of crops efficiently. This genomic information serves as a foundation for precision breeding programs, allowing scientists to identify desirable traits and accelerate the development of crops with enhanced resilience and productivity. Furthermore, bioinformatics facilitates the identification of molecular markers associated with key agronomic traits. This information enables the implementation of marker-assisted breeding, a technique that expedites the selection of desired traits in crops, reducing the time and resources required for traditional breeding methods. The result is the development of crop varieties that are not only more productive but also more resistant to pests and diseases, contributing to global food security. Beyond enhancing crop performance, bioinformatics-driven crop genomics also addresses environmental concerns. By deciphering the genetic basis of stress tolerance and resource- use efficiency, researchers can develop crops that require fewer inputs such as water, fertilizers, and pesticides. This not only benefits farmers economically but also reduces the ecological footprint of agriculture, promoting sustainable practices and minimizing environmental impact.

Keywords: Bioinformatics; agronomic; productivity; resource use efficiency; genomic.

1. INTRODUCTION

In recent years, bioinformatics has evolved as an indispensable instrument for the organization, administration, and dissemination of biological data. Since then, it has brought about a revolution in the sequencing of nucleic acids by giving tools for both the analysis and interpretation of data as well as modelling [1]. The fast proliferation of omics approaches has resulted in a more comprehensive perspective on the structure and operation of systems. This has presented bioinformatics with a challenge in terms of the amount of the data and the need for integrative efforts.

The introduction of Next-Generation Sequencing (NGS) technology has made a substantial contribution to the agricultural industry by reducing the amount of time required for the execution of experiments and providing much higher resolution [2]. As a result of this, scientists have shown an unanticipated interest in the topic, which may be attributed to the increased affordability of experimental techniques and the economic needs. NGS technologies have been crucial in the creation of innovations in agriculture, such as the discovery of agronomically significant variation and the fast growth in crop genomic data. These novelties have been made possible by the introduction of NGS technology [3].

There has been a long-standing reliance on cycles of phenotypic selection and crossing in crop breeding. These cycles yield better genotypes via the process of genetic recombination. Genomic sequencing makes it possible to identify all of the genes and genetic variations that contribute to agronomic qualities, as well as to evaluate the changes that occur at the genotype level throughout the breeding process. Quantitative trait loci (QTL) mapping and genome-wide association studies (GWAS) are two examples of the ways in which genomic sequencing of crop populations might enable gene-level resolution of agronomic variation. Genomic sequencing is playing an increasingly essential role in all areas of crop breeding [4].

The processing and analysis of huge genomic datasets, as well as the acquisition of functional insights into plant genomes, are both very important applications of bioinformatics.

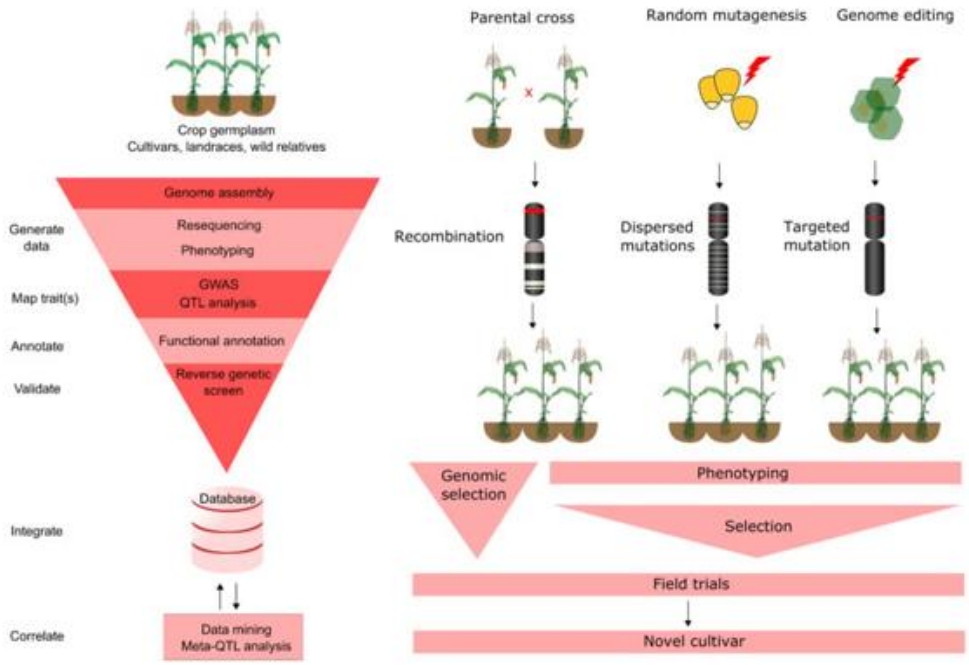

Fig. 1. Quantitative trait loci (QTL) mapping and genome-wide association studies (GWAS)

However, the algorithms that are necessary for genome assembly, sequence alignment, and variant calling are not straightforward, and there are a variety of competing computational techniques, each of which has its own set of biases. It is challenging for assembly and alignment tools to mix multiple data types in order to lessen the effect of distinct biases since many methods that were created for short reads perform poorly when aligning lengthy reads [5].

Providing complete information that may help crop development can be accomplished by downstream analyses like as comparative genomic analysis, variant calling, and genome-wide association studies (GWAS). When it comes to addressing the impacts of population structure, software made specifically for the purpose of carrying out GWAS may utilize models of varied complexity. It is of the utmost importance to use the appropriate bioinformatics tool for the appropriate application [6].

For plant breeding in the 21st century, a methodology that takes into account several disciplines is required in order to identify and address problems associated with breeding and to enhance crop output. There is a significant contribution that genomics and bioinformatics make to the process of boosting the pace of creation of better agricultural cultivars [7]. However, the massive volumes of genotypic and phenotypic data that are now accessible provide a significant obstacle in the way of integrating multiple data outputs for breeding purposes. Addressing this difficulty and assisting in the delivery of breeding goals may be accomplished by the incorporation of phenotypes, genomics, and bioinformatics techniques and resources into public and private breeding pipelines simultaneously [8].

To summarise, developments in genomics and bioinformatics have made major contributions to the field of agriculture. These advancements have made it possible to conduct precision breeding and have helped overcome bottlenecks in crop improvement.

2. HOW TO IMPROVE PLANT GENOMICS

However, there are drawbacks to next-generation sequencing (NGS), including as intrinsic biases and imprecise alignment of repetitive regions. Genome sequencing and resequencing have become common practices in several subfields of plant biology. PacBio single-molecule real-time sequencing and Oxford Nanopore Technologies (ONT) sequencing are examples of third-generation sequencing techniques that have made it possible to generate lengthy reads and produce genome assemblies that are more precise and

contiguous. Breeders are able to uncover genes that are associated with agronomic features, ascertain the location and function of these genes, and build molecular markers that are present throughout the whole genome [9].

The combination of long-read sequencing, long-range mapping methods, and chromosomal conformation capture has made it possible for even smaller labs and non-model crop species to get their hands on highly contiguous chromosome-level crop genome assemblies. Recent developments in optical mapping technologies, such as BioNano Genomics, make it possible to rapidly label long DNA molecules that are more than 250 kilobases in length [10]. This enables the identification of structural variations and the construction of scaffolding of high quality at a cheap cost. For instance, the assembly of the desiccation-resistant grass species Oropetium thomaeum achieved a contig N50 of 2.4 Mb with over 99.5% genome coverage. This contiguity is comparable to that of model plant genomes such as Arabidopsis (TAIR10), rice (V 7) and Brachypodium distachyon (V 2.1). PacBio sequencing and optical mapping from BioNano were utilized in this process [11].

In addition, long-read sequencing reveals repeated areas with a high degree of precision, which assists in the identification of 18 telomeric regions, nine centromeric satellites, and 3247 extended terminal repeats that are intact across 358 families. The chromosomal conformation capture sequencing (Hi-C) technique is an additional breakthrough in the third generation of mapping. This technique is based on the naturally occurring physical tight ligation of DNA segments. Further enhancement of chromosomal phasing and scaffolding is made possible with the integration of high-content data and optical mapping results [12].

When it comes to breeding, the most potent use of third-generation sequencing is the construction of enhanced highly contiguous crop genomes. The size of the genome, the ploidy of the genome, the levels of repetitive content, and the amount of money that is available are all crucial factors to take into consideration when choosing a sequencing strategy for a crop genome assembly project [13]. At the moment, the most important option is to choose between PacBio, ONT, and NGS. These three technologies may be used in conjunction with one another and complemented with additional long-range technologies [14].

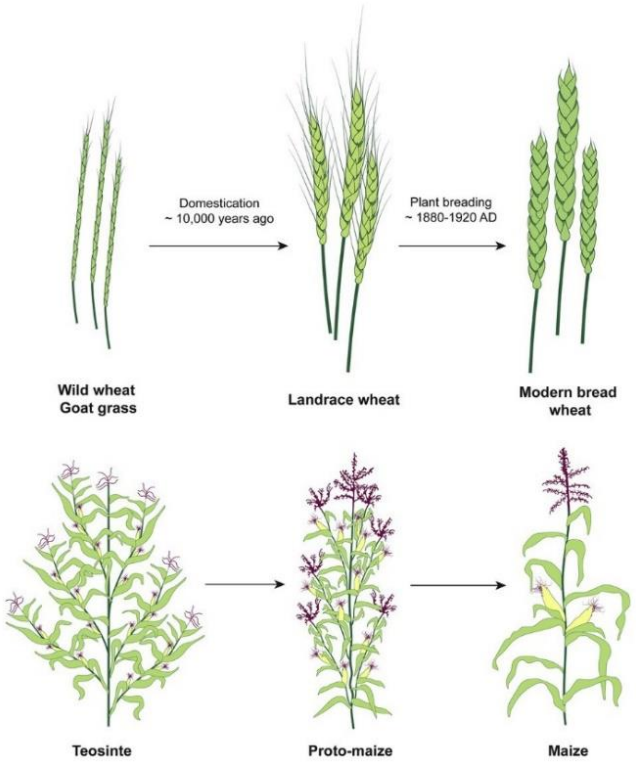

Fig. 2. Plant genomics

It is now feasible to explore agricultural attributes at all levels, from the gene level all the way up to the population level, thanks to such like 'omics' and sequencing technologies of the third generation. On the other hand, key sequence repositories like as Genbank, European Molecular Biological Laboratory (EMBL), PlantGDB, and Phytozome are primarily concerned with the storage and management of genomic data [15]. They do not include variation or phenotypic data from other sources. Because of this, it is more difficult for plant biologists and breeders to establish a connection between genotype and phenotype, which often necessitates the collection of information on genomes, epigenomics, phenotypes, and environments [16].

It is necessary to conduct intelligent mining of large-scale agricultural databases in order to overcome this gap in major repositories. This will allow for the merging of complicated data resources, which will in turn allow for gene discovery and crop improvement. KnetMiner is an intelligent mining program that is web-based and has been used to develop integrative databases for significant crops such as barley and wheat [17]. These databases have provided insights into indirect correlations between distant features and biological processes. Wheat and rice are now undergoing various stages of development in order to further build single information systems that are accessible for both commodities [18].

The assessment of genetic areas that are associated to quantitative phenotypic characteristics is made possible via the analysis of quantitative trait loci (QTL), which helps to bridge the gap between genomics and the domain of study. On the other hand, because of the growing number of quantitative trait loci (QTL) studies that are being carried out and published in plants, a new issue that has arisen in order to find high-quality candidate loci and further enhance crop breeding is to integrate information from various QTL research [19]. Meta-analysis, which is a method that can pool the results of a variety of research and predict the position of QTL with more precision than individual studies, is necessary in order to make full use of the resources that are now available. For the purpose of doing meta-QTL analysis in an effective manner, there are a number of bioinformatics tools available, such as RASQUAL, solQTL, and MetaQTL [20].

In maize, cotton, soybeans, and wheat, meta-QTL studies have been carried out in order to map out characteristics that are associated with crop growth as well as biotic and abiotic responses. Meta-QTL analysis, for instance, has been used to discover five groups of yield and yield-related candidate genes in wheat. This was accomplished by using 195 molecular markers and 197 ESTs that were reported from 55 wheat QTL investigations that were conducted over the course of the last 14 years [21]. In a similar manner, twenty consensus QTLs and the markers that are associated with them were narrowed down by meta-QTL from a mixture of QTL research conducted over the course of the last twenty years. This laid the groundwork for gene mining and crop improvement in soybean.

Genetic Genomics Analysis (GWAS) is a strong approach that involves linking an observable agronomic feature to a genomic area [22]. This method has a high detection power, which allows for the scanning of the whole crop genome and the identification of uncommon alleles using a comparatively small number of genetic markers. Utilizing natural populations, genome-wide association studies (GWAS) provide a better resolution, allowing for the identification of many recombination events and the investigation of natural variants that are related with phenotypical differences. In order to enhance crop yields, association studies are more likely to discover particular candidate genes than QTL analysis. These genes may be directly introduced into crop germplasm in order to improve crop yields [23].

Additional options for performing genome-wide association studies (GWAS) have become available as a result of developments in bioinformatics tools. These tools include PLINK (standard regression analysis) and TASSEL (mixed linear model including population and family structure in the methodology). Other upgraded GWAS bioinformatics tools, such as GAPIT, have also been created in order to handle a huge dataset that contains over one million SNPs among 10,000 people in a computationally efficient manner. This was accomplished by using the compressed mixed linear model and the model-based prediction and selection technique [24].

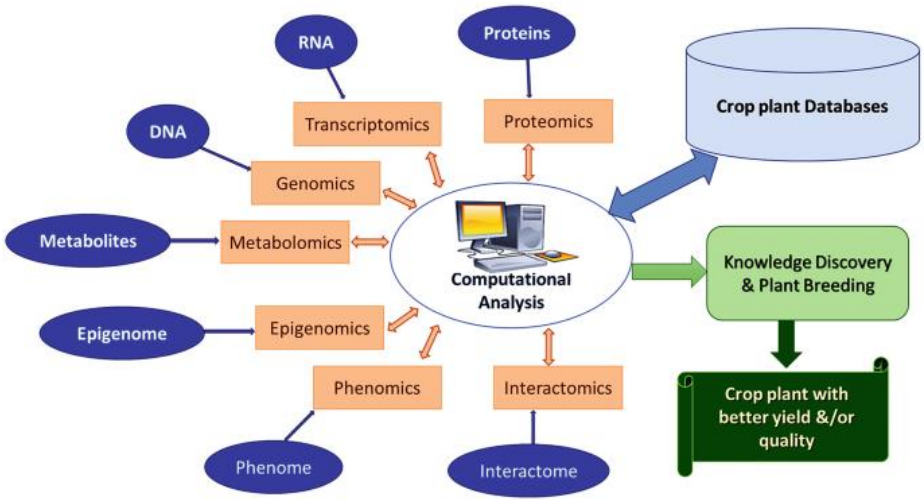

Fig. 3. Knowledge discovery and plant breeding

In the field of breeding, forward genetic screening is a technique that is commonly used to identify and define genes based on a trait that is already known. It is possible to evaluate the phenotypic impact of changed sequences of certain genes or regulatory regions using the process of reverse genetic screening. It is vital to detect functional variation related with agronomic features such as tolerance to abiotic and biotic stressors, disease resistance, enhanced yield, and better nutritional quality [25]. This may be accomplished via the use of both forward and reverse genetic screening.

While forward genetic screening may help enhance gene cloning and marker creation, reverse genetic screening enables selective screening of coding areas while avoiding intergenic sequences. Both types of screening are used in genetic research. When it comes to rice, it has been shown that it is sufficient to sequence 20 megabytes of the 389 megabyte genome in order to recover induced mutations [26]. In the field of crop functional genomics and breeding, reverse genetic screening has been used. Targeted Induced Local Lesions IN Genomes (TILLING) is an example of a reverse genetic methodology that is able to make use of both standard mutation induction and high-throughput mutation techniques [27].

Discovering features using quantitative trait loci (QTL) analysis, genome-wide association studies (GWAS), or reverse genetics is not required for genomics-based breeding, especially when aiming for polygenic agronomic variables like yield. When presented with complicated characteristics that are difficult to introgress in a systematic manner, genotyping-by-sequencing (GBS) provides an alternative breeding strategy that may be used [28]. Calculating the genomic estimated breeding values (GEBV) for sets of variations based on a genotyped and phenotyped training population is the foundation of genetic selection (GS). By combining genetic selection with automated phenotyping approaches, it is possible to further improve the accuracy of GEBV prediction, hence reducing the length of the breeding cycle [29].

For example, promoters and enhancers are examples of cis-regulatory elements (CREs), which are responsible for regulating gene expression and may include about half of all variations that influence phenotypes. When it comes to crops, domesticated features are often brought about by variations in CREs. The use of CREs as breeding targets might be beneficial in situations when the objective is not to completely eliminate a gene but rather to decrease or raise its expression [30]. Through the use of DNase I hypersensitivity mapping, ATAC-seq, and ChIP-seq assays, open chromatin may be discovered, which assists in the prediction of putative central regulatory elements (CREs). Laboratory methods that have been developed relatively recently combine the detection of chromatin signatures with genome editing in order to enable prediction, confirmation, and functional evaluation of CREs over the whole genome [31].

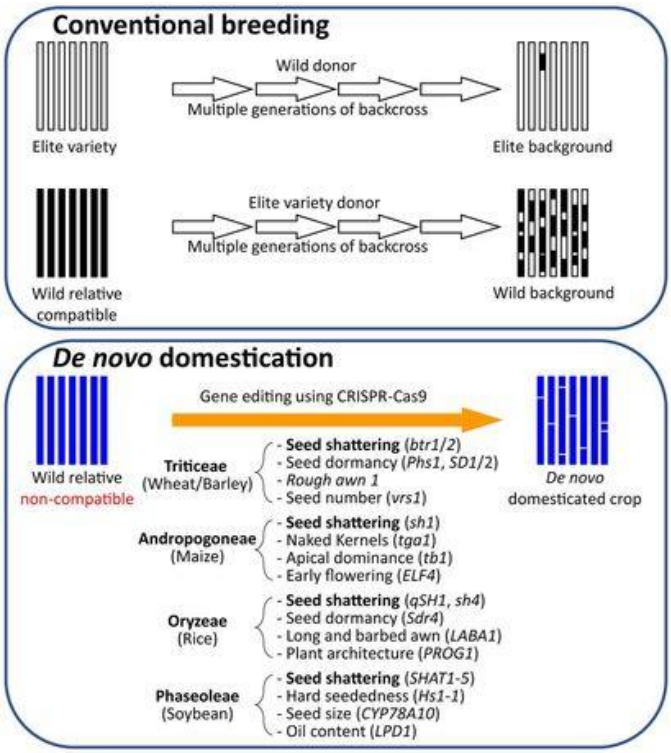

Fig. 4. Conventional breeding

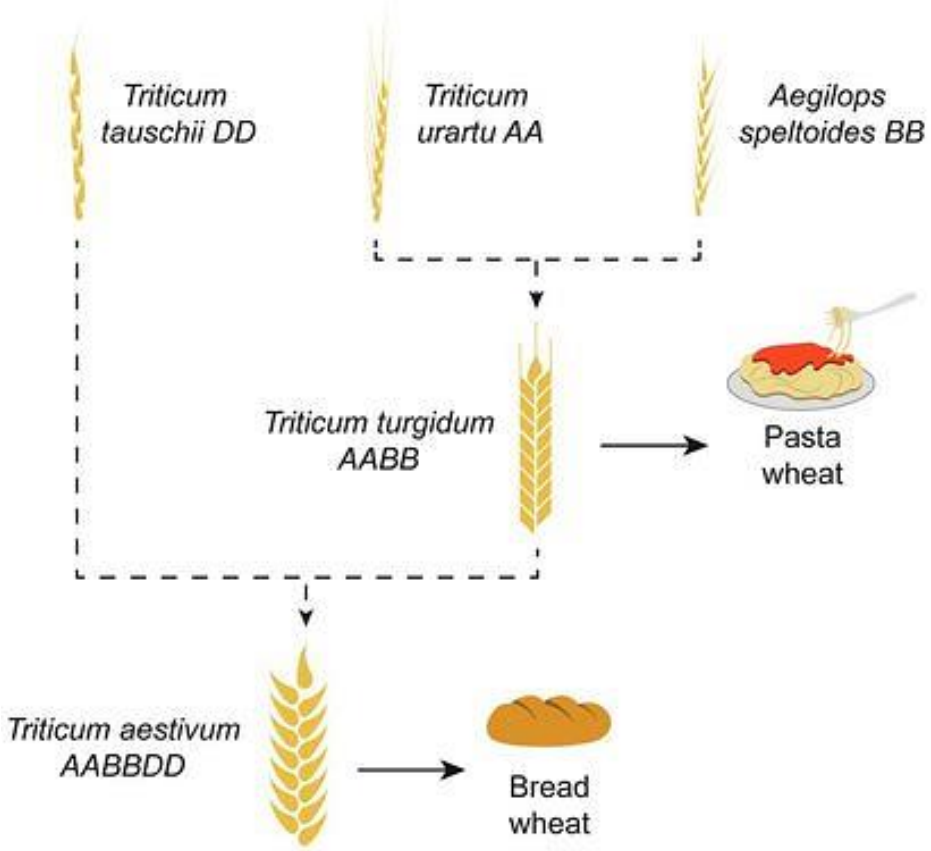

Fig. 5. trait-associated CREs

As our capacity for high-throughput CRE detection increases, the problem that breeders face is determining which CRE in particular to focus on. Only a little amount of information is known about the functional influence that CREs have on plants, and the functions that certain CREs play in regulatory networks are mostly unclear. The experiments that are required to characterize CREs in this manner have been carried out on genes in rice, which has resulted in the creation of a combined mutant library consisting of over 100,000 different lines [32]. It is possible to identify CREs that are related with agronomic features by first creating a CRE mutant library using a similar method and then acquiring expression data from the mutant lines. Once trait-associated CREs have been identified, an allelic series that has been established by genome editing has the potential to quickly generate stepwise variation in a target characteristic [33].

3. CROP BREEDING AND BIOINFORMATICS

However, there are drawbacks to next-generation sequencing (NGS), including as intrinsic biases and imprecise alignment of repetitive regions. Genome sequencing and resequencing have become common practices in several subfields of plant biology. PacBio single-molecule real-time sequencing and Oxford Nanopore Technologies (ONT) sequencing are examples of third-generation sequencing techniques that have made it possible to generate lengthy reads and produce genome assemblies that are more precise and contiguous [34]. Breeders are able to uncover genes that are associated with agronomic features, ascertain the location and function of these genes, and build molecular markers that are present throughout the whole genome [35].

The combination of long-read sequencing, long-range mapping methods, and chromosomal conformation capture has made it possible for even smaller labs and non-model crop species to get their hands on highly contiguous chromosome-level crop genome assemblies. Recent developments in optical mapping technologies, such as BioNano Genomics, make it possible to rapidly label long DNA molecules that are more than 250 kilobases in length [36]. This enables the identification of structural variations and the construction of scaffolding of high quality at a cheap cost. For instance, the assembly of the desiccation-resistant grass species Oropetium thomaeum achieved a contig N50 of 2.4 Mb with over 99.5% genome coverage. This contiguity is comparable to that of model plant genomes such as Arabidopsis (TAIR10), rice (V 7) and Brachypodium distachyon (V 2.1). PacBio sequencing and optical mapping from BioNano were utilized in this process [37].

In addition, long-read sequencing reveals repeated areas with a high degree of precision, which assists in the identification of 18 telomeric regions, nine centromeric satellites, and 3247 extended terminal repeats that are intact across 358 families. The chromosomal conformation capture sequencing (Hi-C) technique is an additional breakthrough in the third generation of mapping. This technique is based on the naturally occurring physical tight ligation of DNA segments. Further enhancement of chromosomal phasing and scaffolding is made possible with the integration of high-content data and optical mapping results [38].

When it comes to breeding, the most potent use of third-generation sequencing is the construction of enhanced highly contiguous crop genomes. The size of the genome, the ploidy of the genome, the levels of repetitive content, and the amount of money that is available are all crucial factors to take into consideration when choosing a sequencing strategy for a crop genome assembly project [39]. At the moment, the most important option is to choose between PacBio, ONT, and NGS. These three technologies may be used in conjunction with one another and complemented with additional long-range technologies [40].

It is now feasible to explore agricultural attributes at all levels, from the gene level all the way up to the population level, thanks to such like 'omics' and sequencing technologies of the third generation. On the other hand, key sequence repositories like as Genbank, European Molecular Biological Laboratory (EMBL), PlantGDB, and Phytozome are primarily concerned with the storage and management of genomic data [41]. They do not include variation or phenotypic data from other sources. Because of this, it is more difficult for plant biologists and breeders to establish a connection between genotype and phenotype, which often necessitates the collection of information on genomes, epigenomics, phenotypes, and environments [42].

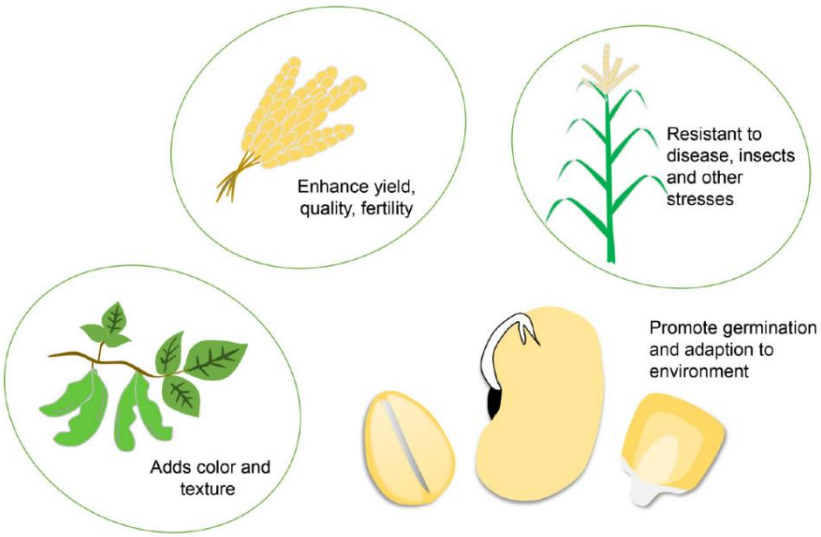

Fig. 6. Crop Breeding and bioinformatics

It is necessary to conduct intelligent mining of large-scale agricultural databases in order to overcome this gap in major repositories. This will allow for the merging of complicated data resources, which will in turn allow for gene discovery and crop improvement. KnetMiner is an intelligent mining program that is web-based and has been used to develop integrative databases for significant crops such as barley and wheat [43]. These databases have provided insights into indirect correlations between distant features and biological processes. Wheat and rice are now undergoing various stages of development in order to further build single information systems that are accessible for both commodities. both systems will enable wide querying across integrated databases [44].

The assessment of genetic areas that are associated to quantitative phenotypic characteristics is made possible via the analysis of quantitative trait loci (QTL), which helps to bridge the gap between genomics and the domain of study. On the other hand, because of the growing number of quantitative trait loci (QTL) studies that are being carried out and published in plants, a new issue that has arisen in order to find high-quality candidate loci and further enhance crop breeding is to integrate information from various QTL research [45]. Meta-analysis, which is a method that can pool the results of a variety of research and predict the position of QTL with more precision than individual studies, is necessary in order to make full use of the resources that are now available. For the purpose of doing meta-QTL analysis in an effective manner, there are a number of bioinformatics tools available, such as RASQUAL, solQTL, and MetaQTL [46].

In maize, cotton, soybeans, and wheat, meta-QTL studies have been carried out in order to map out characteristics that are associated with crop growth as well as biotic and abiotic responses. Meta-QTL analysis, for instance, has been used to discover five groups of yield and yield-related candidate genes in wheat. This was accomplished by using 195 molecular markers and 197 ESTs that were reported from 55 wheat QTL investigations that were conducted over the course of the last 14 years [47]. In a similar manner, twenty consensus QTLs and the markers that are associated with them were narrowed down by meta-QTL from a mixture of QTL research conducted over the course of the last twenty years. This laid the groundwork for gene mining and crop improvement in soybean.

Genetic Genomics Analysis (GWAS) is a strong approach that involves linking an observable agronomic feature to a genomic area. This method has a high detection power, which allows for the scanning of the whole crop genome and the identification of uncommon alleles using a comparatively small number of genetic markers [48]. Utilizing natural populations, genome-wide association studies (GWAS) provide a better resolution, allowing for the identification of many recombination events and the investigation of natural variants that are related with phenotypical differences. In order to enhance crop yields, association studies are more likely to discover particular candidate genes than QTL analysis. These genes may be directly introduced into crop germplasm in order to improve crop yields [49].

Additional options for performing genome-wide association studies (GWAS) have become available as a result of developments in bioinformatics tools. These tools include PLINK (standard regression analysis) and TASSEL (mixed linear model including population and family structure in the methodology). Other upgraded GWAS bioinformatics tools, such as GAPIT, have also been created in order to handle a huge dataset that contains over one million SNPs among 10,000 people in a computationally efficient manner. This was accomplished by using the compressed mixed linear model and the model-based prediction and selection technique [50].

In the field of breeding, forward genetic screening is a technique that is commonly used to identify and define genes based on a trait that is already known. It is possible to evaluate the phenotypic impact of changed sequences of certain genes or regulatory regions using the process of reverse genetic screening. It is vital to detect functional variation related with agronomic features such as tolerance to abiotic and biotic stressors, disease resistance, enhanced yield, and better nutritional quality [51]. This may be accomplished via the use of both forward and reverse genetic screening.

While forward genetic screening may help enhance gene cloning and marker creation, reverse genetic screening enables selective screening of coding areas while avoiding intergenic sequences. Both types of screening are used in genetic research [52]. When it comes to rice, it has been shown that it is sufficient to sequence 20 megabytes of the 389 megabyte genome in order to recover induced mutations. In the field of crop functional genomics and breeding, reverse genetic screening has been used. Targeted Induced Local Lesions IN Genomes (TILLING) is an example of a reverse genetic methodology that is able to make use of both standard mutation induction and high-throughput mutation techniques [53].

Discovering features using quantitative trait loci (QTL) analysis, genome-wide association studies (GWAS), or reverse genetics is not required for genomics-based breeding, especially when aiming for polygenic agronomic variables like yield. When presented with complicated characteristics that are difficult to introgress in a systematic manner, genotyping-by-sequencing (GBS) provides an alternative breeding strategy that may be used [54].

Calculating the genomic estimated breeding values (GEBV) for sets of variations based on a genotyped and phenotype training population is the foundation of genetic selection (GS). By combining genetic selection with automated phenotyping approaches, it is possible to further improve the accuracy of GEBV prediction, hence reducing the length of the breeding cycle [55].

For example, promoters and enhancers are examples of cis-regulatory elements (CREs), which are responsible for regulating gene expression and may include about half of all variations that influence phenotypes. When it comes to crops, domesticated features are often brought about by variations in CREs. The use of CREs as breeding targets might be beneficial in situations when the objective is not to completely eliminate a gene but rather to decrease or raise its expression [56]. Through the use of DNase I hypersensitivity mapping, ATAC-seq, and ChIP-seq assays, open chromatin may be discovered, which assists in the prediction of putative central regulatory elements (CREs). Laboratory methods that have been developed relatively recently combine the detection of chromatin signatures with genome editing in order to enable prediction, confirmation, and functional evaluation of CREs over the whole genome [57].

As our capacity for high-throughput CRE detection increases, the problem that breeders face is determining which CRE in particular to focus on. Only a little amount of information is known about the functional influence that CREs have on plants, and the functions that certain CREs play in regulatory networks are mostly unclear. The experiments that are required to characterize CREs in this manner have been carried out on genes in rice, which has resulted in the creation of a combined mutant library consisting of over 100,000 different lines [58]. It is possible to identify CREs that are related with agronomic features by first creating a CRE mutant library using a similar method and then acquiring expression data from the mutant lines. Once trait-associated CREs have been identified, an allelic series that has been established by genome editing has the potential to quickly generate stepwise variation in a target characteristic [59].

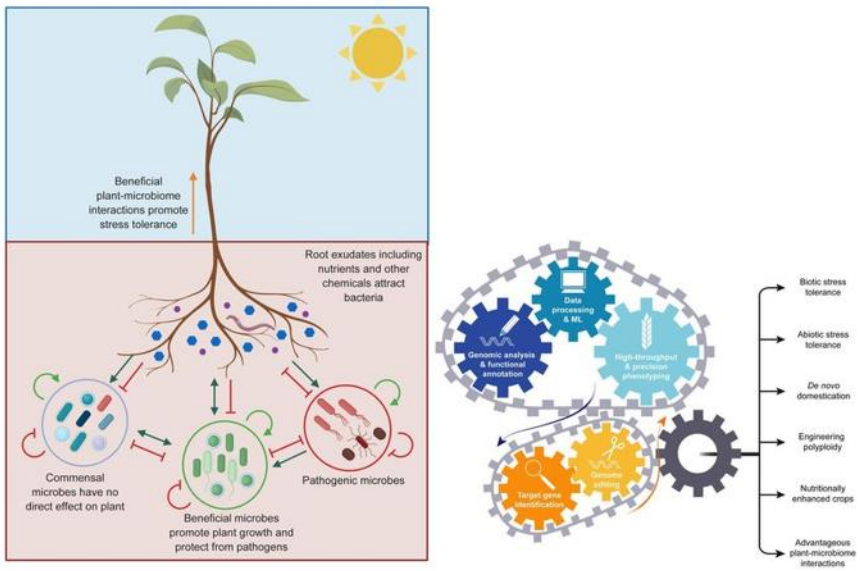

Fig. 7. cis-regulatory elements (CREs), which are responsible for regulating gene expression

4. TOOLS INVOLVED IN BIOINFORMATICS IN AGRICULTURE

In the twenty-first century, there has been a fast development of new sequencing techniques. Second-generation sequencing technologies such as Illumina have made it possible to assemble more than two hundred plant genomes. The short-read duration of these technologies, on the other hand, made it impossible for them to bridge across lengthy sections of repeating sequences, which led to fragmented assemblies. This was the primary obstacle that these technologies presented. On the other hand, the use of third-generation sequencing and the production of long reads by PacBio and Oxford Nanopore has made it possible to assemble plant genomes at the chromosomal level. The methods of long-read sequencing are often paired with optical mapping and conformation capture, which in turn allows for the creation of draft genomes that are unmatched in their consistency [60].

The availability of high-quality genome assemblies at the chromosomal size significantly increases the accuracy of subsequent genomic research. This includes the annotation of genes and regulatory regions, genome-wide association studies (GWAS), measurement of gene expression, and the identification of homologous proteins. The increase in the sequencing of plant genomes has made it very necessary to have accurate gene prediction and functional annotation in order to successfully identify candidate genes with pinpoint accuracy [61].

Regions of the genome that are thought to be translated into messenger RNA (mRNA) that codes for proteins or into one of the classes of non-coding RNAs (ncRNA) are referred to be gene models. Ab initio gene prediction and homology-based approaches, which make use of sequence similarity to known transcripts or proteins, are often used in the construction of gene models. These methods are typically used in conjunction with one another [62].

Although genome annotations are readily available, one of the most significant challenges that molecular breeding processes face is the functional characterisation of genes that have been annotated. One of the most important experimental model plant species, A. thaliana, has been annotated with putative functions for more than ninety percent of its genes, and around fifty percent of its genes have annotations that are supported by experimental data [63]. On the other hand, gene functional annotations for the majority of crop plants are based on homology-based inference. These annotations are carried out by transferring annotations from the majority of genes that are comparable in model plants such as Arabidopsis and rice, with very little direct experimental evidence. Annotation transfer is made much more problematic by the history of plant evolution, which includes many rounds of polyploidy followed by diploidization, which results in gene redundancy, differential loss, and neo- and sub-functionalization [64].

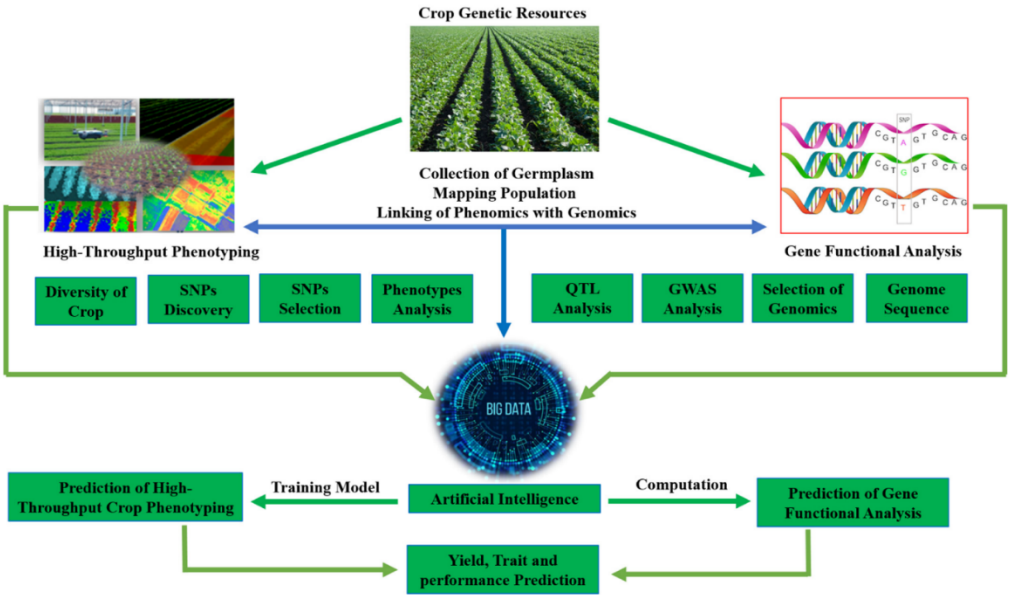

Fig. 8. Tools involved in bioinformatics in agriculture

Rapid advancements in the use of CRISPR/Cas9 genome editing will soon make it possible to generate genome-wide mutant libraries for important crops, which will make a substantial contribution to the efforts being made to annotate their functional characteristics. Integrative genomics methods have also been used to simplify the selection of top candidates [65]. For instance, dedicated databases that integrate genotypic, phenotypic, and association data for rice (SNP-Seek), soybean (SoyBase), and wheat (T3) have been utilized. Using multicriteria decision analysis, methods such as KnetMiner and MCRiceRepGP were created with the intention of ranking candidate genes that are engaged in biological processes of interest. These tools were built in addition to specialized databases [66].

A small percentage of the genomes of the majority of big agricultural plants have protein-coding genes, whereas the remaining portion is composed of sequences that do not code for proteins. It has been discovered via recent technical and conceptual advancements that plant genomes contain hundreds of non-coding RNAs (ncRNAs) that have the capacity to function, as well as the prevalence of remote regulatory elements, such as enhancers [67]. The fact that long noncoding RNAs show a high preference for transcription in reproductive organs suggests that they play a role in the process of sexual reproduction in plants, which is an essential activity that influences blooming, fruit development, and grain formation [68].

Targets for genome editing might include newly described long noncoding RNAs (lncRNAs), which have an effect on critical features. For instance, it was discovered that a rice long noncoding RNA called LDMAR has a role in the regulation of photoperiod-sensitive male sterility (PSMS), which is an essential characteristic that played a role in the creation of hybrid rice [69].

DNA sequences that do not code for a protein, such as cis-regulatory elements (CREs), promoters, and enhancers/silencers, are very important in the process of influencing characteristics that are the focus of artificial selection. The changes that occur in CREs are regarded to be one of the most important evolutionary processes that are responsible for the divergence of cis-regulatory areas that are related with domestication and the emergence of new morphological forms [70]. The expression of genes that are involved in the regulation of essential features, such as the amount of anthocyanin in maize and the amount of time it takes for Arabidopsis to blossom, has been modulated by a number of factors that have been identified as enhancers [71].

Studies conducted on the model plant species Arabidopsis, rice, maize, and cotton have resulted in considerable advancements in the identification of plant CREs over the course of the last several years [72]. The use of DNase-Seq and ATAC-Seq methods in plant research has led to rapid improvements in this area. These approaches evaluate DNA "openness" as a surrogate for the accessibility of DNA to

transcription factors, RNA polymerase, and other protein complexes involved in gene expression. These techniques have been crucial in the rapid advancement of this subject [73].

The idea of the pangenome, which was first presented in bacteria, refers to the totality of the genomic sequence and gene content that is contained within a species as opposed to a single person. Rice, soybeans, bread wheat, and oilseed rape are some of the most important crop species for which pangenomes have been recently built [74]. Plant accessory genes have been demonstrated to be over-represented in activities related to signaling, disease resistance, and abiotic stress response. These genes have the potential to contribute to environmental adaptability and phenotypic plasticity, and they also provide intriguing targets for crop development [75].

The pangenome provides a natural alternative to the existing paradigm of utilizing a single reference genome. This is because the selection of the reference has an impact on subsequent genomic investigations, such as genome-wide association studies (GWAS) and the measurement of gene expression and expression levels [76]. Read mapping and variant calling accuracy are both improved when the pangenome is used as a reference. Additionally, the use of the pangenome reference makes it possible to include variants other than SNPs in genome-wide association studies (GWAS) [77]. Recent research conducted on both plants and people has shown that the incorporation of structural variations into association studies has the potential to assist in the identification of causative variants. The discovery of missing quantitative trait locus (QTLs) related with disease resistance in oilseed rape was made possible, for instance, by the use of sequence presence/absence variation [78].

5. UPCOMING FUTURISTIC TECHNOLOGIES FOR ENHANCEMENT IN BIOINFORMATICS

As a result of the fact that machine learning algorithms have the capacity to assist practically all elements of genomic investigations, they are suited for the study of large, multilayer datasets in situations where expert knowledge is either wrong or inadequate [79]. Among the many interesting applications of machine learning to plant genomics are the enhancement of the quality of feature annotation, the discovery of the underlying sequence properties of regulatory areas, and even the prediction of the effect of variants. There is a possibility that analogous investigations in agricultural plants may be delayed due to the restricted availability of large-scale information pertaining to epigenetic alteration and chromatin accessibility [80].

It has been possible to speed up the formation of new crops via the use of speed breeding, which involves altering growing circumstances such as the length of the day and the temperature. The plant production time of some of the most important agri-food crops in the world, including as bread wheat, pasta wheat, barley, and canola, has been effectively decreased thanks to their use of this technique [81]. Additionally, the method has been effectively used for the cultivation of orphan crops, including chickpea, peanut, grass pea, lentil, and quinoa, among others. Using genomic techniques such as precision genome editing by CRISPR in conjunction with speed breeding would enable the domestication of the new crop possible in a short amount of time. This would be possible because of the present knowledge about the genes that are being targeted [82].

The term "high-throughput phenotyping" refers to the process of measuring any morphological or physiological properties of plants. These traits might be the consequence of the intervention of individual genes, interactions between genes, or interactions between genes and the environment. Many agronomically important characteristics, such as yield and its components, as well as tolerance to drought and salt, are regulated by a number of genes that have very little impacts, as well as the interactions between these genes and the environment [83]. In order to cultivate plants and investigate how they react to biotic and abiotic challenges, several research organizations concentrate their efforts on a controlled environment. This is done for practical reasons. The environment and the microclimate, on the other hand, are subject to dynamic changes throughout the day in farming, and these changes have an unequal impact on the plant, for instance because of shadowing. Furthermore, regulated light conditions are not even close to being comparable to the irradiance levels and spectral quality that are characteristic of sunshine situations in the natural environment. It is of the utmost importance to investigate the effects of plant stressors under dynamic environmental settings in order to get a comprehensive understanding of the responses of plants to stress [84].

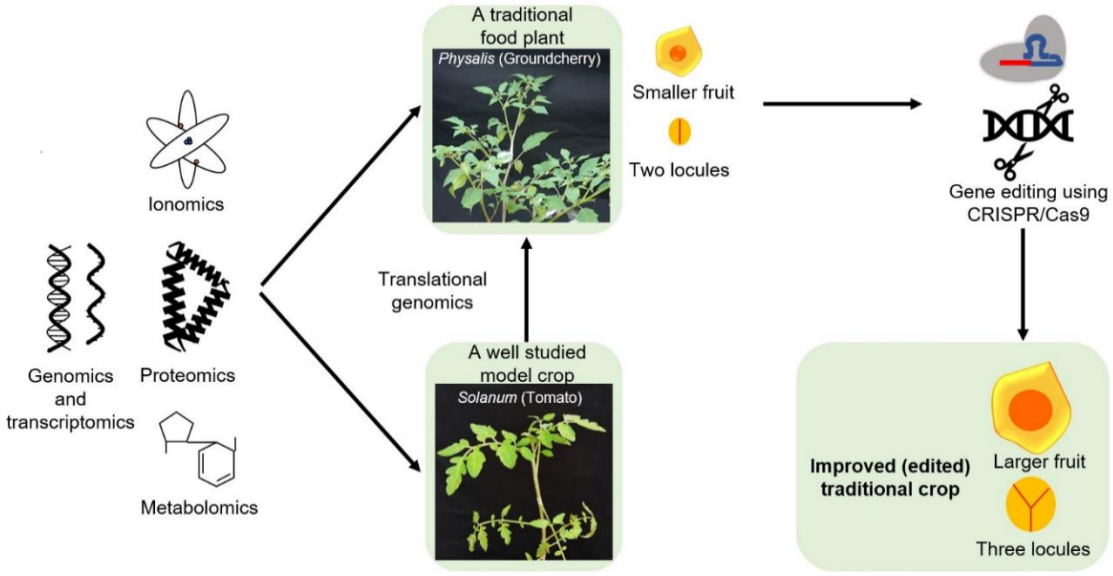

Fig. 9. Upcoming futuristic technologies for enhancement in bioinformatics

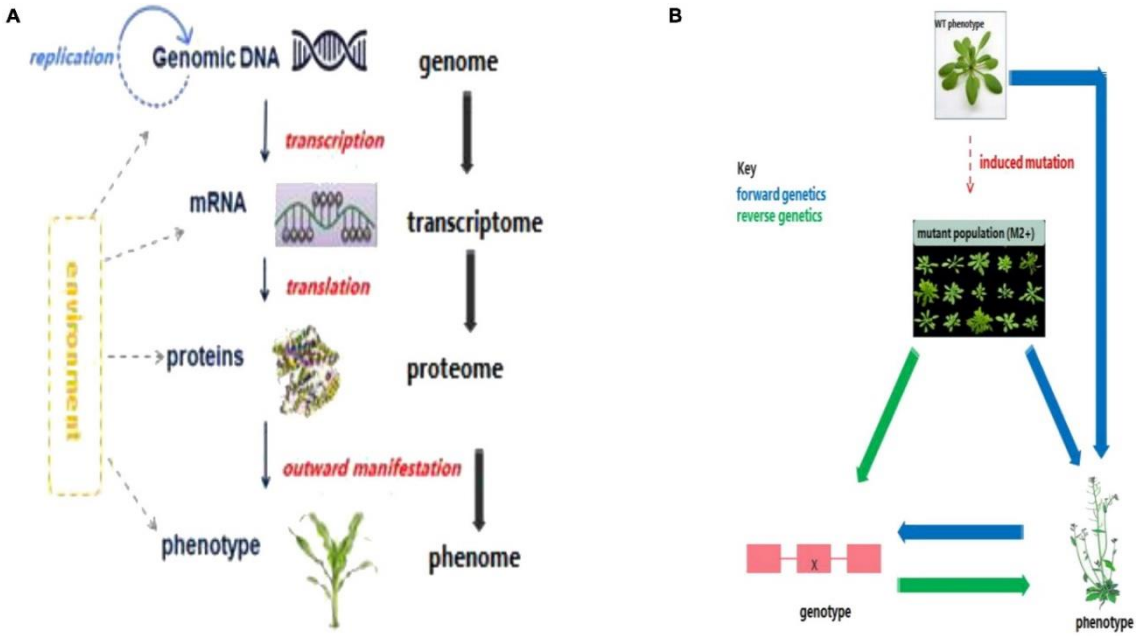

Fig. 10. Plant genomic environment

Platforms for high-throughput phenotyping are both quick and dependable. These platforms make use of robots and imaging technologies that are based on spectral analysis. One of the most significant limitations is the controlled environment, which is distinct from the natural growing conditions that are present in the field [85]. High-throughput in-field phenotyping of attributes, such as canopy temperature, chlorophyll fluorescence, and other biochemical plant parameters, is now possible because to the emergence of hyperspectral imaging technology, which, when paired with drones and manned aircraft, gives the potential for such phenotyping. This technology is starting to become more cost-effective while simultaneously improving the resolution and accuracy of measurements [86].

Using aerial platforms would provide a number of challenges, the most significant of which would be the processing of massive amounts of data in a short period of time. When it comes to high-throughput phenotyping data processing, however, technologies that are based on

machine learning have shown their potential. For the purpose of evaluating complicated physiological characteristics, such as the ability to tolerate abiotic stressors, in-field high-throughput phenotyping is an ideal method. In general, the approaches of machine learning have the ability to contribute a large amount of value to the genetic resources and procedures that are already in existence [87].

6. CLIMATE CHANGE, DISEASE MANAGEMENT, NUTRITIONAL ENHANCEMENT

Due to the continual conflict between plants and pathogens, crop plant pathogens provide a serious risk to contemporary agriculture. This conflict is responsible for developing the genetic diversity of plants. In most cases, diseases are the consequence of a particular relationship between the host and the pathogen. One example of this is the pathogen *Puccinia triticina*, which causes wheat leaf rust. Over cultivation of crops with limited genetic diversity has resulted in an increase in the inoculum of pathogens and has hastened the development of pathogens, which has promoted the spread of these pathogens throughout the world [88].

The epidemiology of infections in particular regions and the geographic dispersion of plant diseases are both impacted by climate change. Through the cultivation of orphan crops and the domestication of new crops, increasing agricultural plant variety will lead to a reduction in the selection pressure exerted on pathogen populations, which will ultimately result in a longer life span for genetic resistance. There is a possibility that this method of disease management is both efficient and environmentally sustainable [89].

The migration of infections to latitudes that are outside of their historical range is another way in which climate change impacts plant diseases. The transfer of pathogens and the spread of illness farther from north to south would be caused by a rise in temperature, which would have an effect on areas not previously susceptible to infection. Recent developments in genomics have offered a snapshot of resistance mechanisms that have formed over the course of the long history of co-evolution. This has led to an enhanced knowledge of the molecular function of these mechanisms and has provided a starting point for research on defence pathways [90].

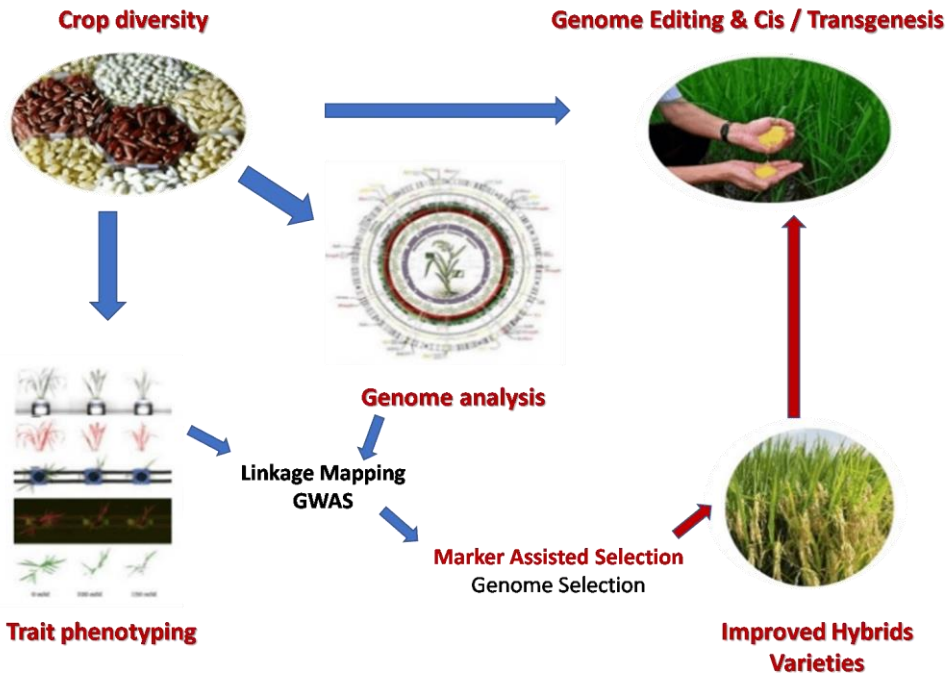

Fig. 11. Climate change, disease management, nutritional enhancement

There is a quick approach for identifying pathogens, following the course of outbreaks, and monitoring the transmission of illness to new sites that can be achieved by genome sequencing. The advancement of third-generation sequencing technologies, particularly Oxford Nanopore, has resulted in the introduction of mobile sequencing equipment that are tiny, inexpensive, and ideal for use in in-field diagnostic systems. The Oxford Nanopore MinION technology has been used for real-time diagnostics of human diseases such as the Ebola and Zika viruses. Protocols for the detection of plant pathogens and pests are also in the process of being developed [91].

In order to guarantee the safety of food supplies across the world, it is essential to increase the nutritional content of crops. There are several staple meals that are poor suppliers of certain macronutrients and vital micronutrients, despite the fact that plants are important sources of both macronutrients and micronutrients. The identification of potential genes involved in plant metabolism has been made possible by recent developments in genome sequencing and annotation, which have given the essential resource [92]. There is the potential for genome editing technologies to be used in the modification of the nutritional profiles of crops. For instance, soybeans might be produced with a high oleic acid content and a low linoleic acid content, and maize could have its anti-nutritional phytic acid content reduced. Furthermore, transgenic technology may be used for the purpose of enhancing the nutritional value of crops [93].

7. CHALLENGES

In the process of breeding, unique combinations of alleles are created either via the natural variety that occurs in germplasm collections or by the processes of producing novel mutations. The employment of advanced breeding techniques, like as irradiation or chemical mutagens, is widespread; yet, these techniques are not without their difficulties because of the high incidence of background mutations. In order to correct characteristics and eliminate harmful background mutations, the CRISPR/Cas method does not need a significant amount of crosses to be performed [94]. In order to direct the Cas protein to specific DNA locations for cleavage, the system makes use of a guide RNA, also known as gRNA. This results in a double-strand break at the target site. Both gene knock-out via mutations that occur during no homologous end joining and gene knock-in through the use of a donor DNA template and homology-directed repair are made possible as a result of this [95].

Over the last five years, the CRISPR/Cas system has been successfully implemented in plants that are crucial for food production. It is anticipated that this will have a significant influence on the agricultural sector. Among the challenges that genome editors face are the enhancement of protoplast transformation, the enhancement of gene targeting efficiency by the use of homology-directed repair, and the optimization of bioinformatics tools for the creation of gRNA with minimum off-target negative consequences [96]. The ideal creation of gRNAs that enable CRISPR/Cas gene editing to be both effective and specific requires the use of bioinformatics tools of the highest quality.

When it comes to the design of gRNA, the two most important characteristics are a high binding affinity to the target site and specificity with a limited likelihood of off-target effects. According to research conducted on human cell lines, it has been shown that the binding effectiveness of guanines at the −1 and −2 PAM-proximal sites is increased, but the binding efficiency of thymines at the +4/−4 PAM-proximal positions is decreased. On the other hand, preliminary investigations suggest that the preferences for bases that have been found in human cells may not be shared by plants [97].

The genomes of plants include a great deal of duplicated information, which makes it challenging to create distinct gRNAs for individual target locations. In the future, bioinformatics tools will be able to design gRNAs to target more genomic locations in crops with greater precision. This will be possible if other endonucleases with varied activities from Cas9 are introduced to the CRISPR/Cas toolbox. Additionally, more empirical data on endonuclease activity in plants will become accessible [98].

8. FUTURISTIC APPROACHES TO COME OVER CHALLENGES

New opportunities to accelerate the application of basic research are being made available as a result of the availability of large-scale sequence and phenotype information at unprecedented scales. These new opportunities include the ability to formulate testable hypotheses regarding

the genetic architecture of quantitative variation, the genes and biological pathways involved, and the causal variants responsible for the inheritance of complex traits in a variety of species [99]. The raw sequence information, on the other hand, needs to be combined with an in-depth understanding of the biology of the species that is being considered, the phenotype or performance of the individuals or population that has been sequenced, and the agroecosystem in which they have been grown, which includes the cultural context and the management practices of the farmers [100].

The capacity to operate at diverse sizes, ranging from molecules to landscapes within a quantitative biology framework, will be necessary in order to realize this promise. Additionally, there will be a need for increased cooperation between breeders, farmers, and the community of biological researchers. Technical impediments to facilitate data-integration and the potential for data-sharing include fragmented and dispersed data across organizations and international borders, inadequate systems for logging and tracking plant genetic resources (PGR), and radically different approaches to data management and sharing within and across public and private sectors due to fundamentally different objectives and low levels of mutual trust [101].

There are a number of projects that have been formed to encourage interoperability in order to overcome these difficulties. Some of these programs are DivSeek, the Research Data Alliance (RDA), the Breeding API, and Global Open Data for Agriculture and Nutrition (GODAN). Through the use of Digital Object Identifiers (DOIs), the Global Information System of the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) operates with the objective of establishing connections between the various information systems that are already in place regarding PGRFA. It is possible that the adoption of standardized procedures would make it easier for organizations to aggregate such data for comparative analysis and collaborative work. Additionally, it might reduce the entrance hurdles that now restrict farmer participation in translational agricultural research and development [102].

There are obvious approaches that may be used to satisfy this requirement for data integration that have been developed from the experience of other fields of biological study. The impact that bioinformatics has had on a wide range of biological concerns is largely attributable to the availability of open-access data. The establishment of three worldwide repositories for the storage of nucleotide sequences occurred in the early 1980s. As a result, huge volumes of nucleotide sequence data were made accessible to the general public without any claims of intellectual property being passed on by the data producers or database administrators. Newer models for the pre-publication of data and manuscripts have been proposed, and alliances of interested parties have been created in order to build data models and suitable structures for the purpose of interface between public and private data [103].

Up until the late 1960s, plant genetic resources were usually considered to be "global public goods." nations that are technologically sophisticated have been pushing for the worldwide recognition of intellectual property protection for living materials. This has caused poor nations, who were the historical origins of a significant amount of the genetic variety of the crops that are being sold and protected, to feel uneasy. Developing countries pushed back through negotiations under the United Nations Environment Programme (UNEP) that led to the Convention on Biological Diversity 1993 (CBD). They insisted on their sovereign rights to regulate access to genetic resources within their borders, with the expectation of negotiating access and benefit sharing agreements with foreign access-seekers [104].

Within the framework of the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization (Nagoya Protocol), which came into effect in 2014, it was agreed to establish mechanisms for the monitoring and enforcement of access and benefit sharing agreements that were negotiated bilaterally. It is not yet possible to make any predictions about the influence that the Nagoya Protocol will have on the willingness of stakeholders to exchange genetic resources for the purpose of agricultural research [105].

A multilateral system of access and benefit-sharing (MLS) was established for contractual parties and international organizations as a result of the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) that was signed in 2004. The Multiple Resource Library (MLS) makes it easier to get access to

the genetic variety of sixty-four different crops and forages, which may then be used for the purposes of conservation, agricultural research, teaching, and plant breeding. Under certain conditions, any commercial users of the content that is accessible via the multilateral system are required to make monetary contributions to an international benefit-sharing fund [106].

On the other hand, accessibility to digital material, such as genetic sequences or phenotypic data, is not particularly regulated by any of the three accords. The fact that digital research data is mentioned as a possible advantage that may be offered in exchange for access to genetic resources is included in all three agreements. The result of this was the emergence of worries over the potential for technical advancements in genomic breeding, gene editing, and gene synthesis to exacerbate the existing technological divide and the resulting economic inequalities between industrialized and poor nations [107].

It has been brought to the attention of stakeholders that the necessary technological capacities are primarily located in prestigious research institutions located in the northern regions of the world. These stakeholders are skeptical that these newly acquired capacities will be utilized to develop technologies that are aimed at resource-poor farmers who are employed in vulnerable agricultural systems. The result of this is that there have been calls for research institutions to cease offering unrestricted free access to genetic sequence data until the problems associated with benefit sharing can be resolved [108].

The cultivation of an open-science culture and the increase of agricultural yields are both dependent on the enhancement of trust and collaboration between many stakeholders. It is important for the United Nations to exercise prudence when it comes to the establishment of legally enforceable solutions. In order to foster trust and transparency among various stakeholders, as well as to encourage the development and utilization of knowledge and technologies that ultimately contribute to the advancement of sustainable development goals, it is necessary to improve governance of the generation and use of genetic sequence data and related information about PGR [109].

9. CONCLUSION

For the international community to develop mechanisms to address the issues that have been brought up in this review, it will most certainly take several years. This is especially true in the event that it is collectively decided that new legally binding agreements (or amendments or protocols to existing legally binding agreements) are required. In the meanwhile, there will be chances for organizations and networks that are interested in developing inclusive forms of governance for the deployment of the new technological capabilities that are outlined in this article in order to realize the objectives of sustainable development. In order for wide coalitions of scientists, information technologists, gene bank managers, breeders, farmers, and civil society organizations to be successful, they will need to identify chances to define a set of shared objectives and build methods for working together that are inclusive and transparent. In the event that they are successful, the governance structures, best practices, and benefit-sharing standards that they produce have the potential to favorably impact the tone of continuing intergovernmental discussions as well as the shape and substance of norms that are ultimately formed under the auspices of the United Nations. In order to develop innovation platforms and governance structures that will inspire confidence and encourage the most effective and fair deployment of those technologies, the ball is now in the court of people who are advocates of these new technologies.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Tilman D, Balzer C, Hill J, Befort BL. Global food demand and the sustainable intensification of agriculture. Proc. Natl. Acad. Sci. USA. 2011;108:20260–20264.
2. Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, Pretty J, Robinson S, Thomas SM, Toulmin C. Food security: The challenge of feeding 9 billion people. Science. 2010;327:812–818.
3. Perez-de-Castro AM, Vilanova S, Cañizares J, Pascual L, M Blanca J, J Diez M, Prohens J, Picó B. Application of genomic tools in plant breeding. Curr. Genom. 2012;13:179–195.
4. Abberton M, Batley J, Bentley A, Bryant J, Cai H, Cockram J, de Oliveira AC, Cseke LJ, Dempewolf H, De Pace C et al. Global

agricultural intensification during climate change: A role for genomics. Plant Biotechnol. J. 2016;14:1095–1098.
5. Edwards D, Batley J, Snowdon RJ. Accessing complex crop genomes with next-generation sequencing. Theor. Appl. Genet. 2013;126:1–11.
6. IRGSP. The map-based sequence of the rice genome. Nature 2005;436:793–800.
7. Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, Pruss D, Pindo M, Fitzgerald LM, Vezzulli S, Reid J et al. A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS ONE. 2007;2:e1326.
8. Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P et al. The genome of the cucumber, Cucumis sativus L. Nat. Genet. 2009;41:1275–1281.
9. Goodwin S, McPherson JD, McCombie WR. Coming of age: Ten years of next-generation sequencing technologies. Nat. Rev. Genet. 2016:17:333–351.
10. Michael TP, Jackson S. The First 50 Plant Genomes. Plant Genome. 2013;6:1–7.
11. Mousavi-Derazmahalleh M, Bayer PE, Hane JK, Babu V, Nguyen HT, Nelson MN, Erskine W, Varshney RK, Papa R, Edwards D. Adapting legume crops to climate change using genomic approaches. Plant Cell Environ; 2018.
12. Dwivedi SL, Scheben A, Edwards D, Spillane C, Ortiz R. Assessing and exploiting functional diversity in germplasm pools to enhance abiotic stress adaptation and yield in cereals and food legumes. Front. Plant Sci. 2017;8: 1461.
13. Moore JH, Asselbergs FW, Williams SM. Bioinformatics challenges for genome-wide association studies. Bioinformatics. 2010; 26:445–455.
14. Batley J, Edwards D. The application of genomics and bioinformatics to accelerate crop improvement in a changing climate. Curr. Opin. Plant Biol. 2016;30:78–81.
15. Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. Software for computing and annotating genomic ranges. PLoS Comput. Biol. 2013;9:e1003118.
16. Pabinger S, Dander A, Fischer M, Snajder R, Sperk M, Efremova M, Krabichler B, Speicher MR, Zschocke J, Trajanoski Z. A survey of tools for variant analysis of next-generation genome sequencing data. Brief. Bioinform. 2014;15:256–278.
17. Hwang S, Kim E, Lee I, Marcotte EM. Systematic comparison of variant calling pipelines using gold standard personal exome variants. Sci. Rep. 2015;5:17875.
18. Sedlazeck FJ, Lee H, Darby CA, Schatz MC. Piercing the dark matter: Bioinformatics of long-range sequencing and mapping. Nat. Rev. Genet. 2018;6: 329–346.
19. Ong Q, Nguyen P, Phuong Thao N, Le L. Bioinformatics approach in plant genomic research. Curr. Genom. 2016;17: 368–378.
20. Grierson CS, Barnes SR, Chase MW, Clarke M, Grierson D, Edwards KJ, Jellis GJ, Jones JD, Knapp S, Oldroyd G et al. One hundred important questions facing plant science research. New Phytol. 2011; 192:6–12.
21. Matthews DE, Carollo VL, Lazo GR, Anderson OD. GrainGenes, the genome database for small-grain crops. Nucleic Acids Res. 2003;31:183–186.
22. Tello-Ruiz MK, Naithani S, Stein JC, Gupta P, Campbell M, Olson A, Wei SR, Preece J, Geniza MJ, Jiao YP et al. Gramene 2018: Unifying comparative genomics and pathway resources for plant research. Nucleic Acids Res. 2018;46:D1181–D1189.
23. Scheben A, Batley J, Edwards D. Revolution in genotyping platforms for crop improvement. In Advances in Biochemical Engineering/Biotechnology; Springer: Berlin/Heidelberg, Germany. 2018;1–16.
24. Moose SP, Mumm RH. Molecular plant breeding as the foundation for 21st century crop improvement. Plant Physiol. 2008; 147:969–977.
25. Watson A, Ghosh S, Williams MJ, Cuddy WS, Simmonds J, Rey MD, Hatta MAM, Hinchliffe A, Steed A, Reynolds D et al. Speed breeding is a powerful tool to accelerate crop research and breeding. Nat. Plants. 2018;4:23–29.
26. Santos R, Algar A, Field R, Mayes S. Integrating GIScience and Crop Science datasets: A study involving genetic, geographic and environmental data. PeerJ Preprints. 2017;5:e2248v2244.
27. Evans K, Jung S, Lee T, Brutcher L, Cho I, Peace C, Main D. Addition of a breeding database in the Genome Database for Rosaceae. Database. 2013;2013:bat078.
28. Li C, Lin F, An D, Wang W, Huang R. Genome Sequencing and Assembly by Long Reads in Plants. Genes. 2017;9:6.

29. Vlk D, Repkova J. Application of next-generation sequencing in plant breeding. Czech J. Genet. Plant. 201753:89–96.
30. Chen K, Ji F, Yuan SJ, Hao WT, Wang W, Hu ZH. The performance of activated sludge exposed to arsanilic acid and amprolium hydrochloride in sequencing batch reactors. Int. Biodeterior. Biodegrad. 2017;116:260–265.
31. Jiao W-B, Schneeberger K. The impact of third generation genomic technologies on plant genome assembly. Curr. Opin. Plant Biol. 2017;36:64–70.
32. VanBuren R, Bryant D, Edger PP, Tang H, Burgess D, Challabathula D, Spittle K, Hall R, Gu J, Lyons E et al. Single-molecule sequencing of the desiccation-tolerant grass Oropetium thomaeum. Nature 2015; 527:508–511.
33. Stankova H, Hastie AR, Chan S, Vrana J, Tulpova Z, Kubalakova M, Visendi P, Hayashi S, Luo M, Batley J et al. BioNano genome mapping of individual chromosomes supports physical mapping and sequence assembly in complex plant genomes. Plant Biotechnol. J. 2016;14: 1523–1531.
34. Van Berkum NL, Lieberman-Aiden E, Williams L, Imakaev M, Gnirke A, Mirny LA, Dekker J, Lander ES. Hi-C: A method to study the three-dimensional architecture of genomes. J. Vis. Exp. 2010;39: e1869.
35. Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, Radchuk V, Dockter C, Hedley PE, Russell J et al. A chromosome conformation capture ordered sequence of the barley genome. Nature. 2017;544:427–433.
36. Michael TP, Jupe F, Bemm F, Motley ST, Sandoval JP, Lanz C, Loudet O, Weigel D, Ecker JR. High contiguity Arabidopsis thaliana genome assembly with a single nanopore flow cell. Nat. Commun. 2018; 9:541.
37. Goodwin S, Gurtowski J, Ethe-Sayers S, Deshpande P, Schatz MC, McCombie WR. Oxford Nanopore sequencing, hybrid error correction and de novo assembly of a eukaryotic genome. Genome Res. 2015;25:1750–1756.
38. Med.stanford.edu. Stanford Medicine Sequencing Service Rates. Available:http://med.stanford.edu/gssc/rates.html (accessed on 27 May 2018).
39. Cgrb.oregonstate.edu. Illumina HiSeq 3000 Service Fees. Available:http://cgrb.oregonstate.edu/core/illumina-hiseq-3000/illumina-hiseq-3000-service-fees (accessed on 27 May 2018).
40. Allseq.com. General overview of Illumina Sequencing. Available:http://allseq.com/knowledge-bank/sequencing-platforms/illumina/ (accessed on 27 May 2018).
41. Gordon D, Huddleston J, Chaisson MJP, Hill CM, Kronenberg ZN, Munson KM, Malig M, Raja A, Fiddes I, Hillier LW et al. Long-read sequence assembly of the gorilla genome. Science. 2016;352: aae0344.
42. Washington.edu. University of Washington PacBio Sequencing Services. Available:https://pacbio.gs.washington.edu/ (accessed on 27 May 2018).
43. Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, Studholme DJ. Assessing the performance of the Oxford Nanopore Technologies MinION. Biomol. Detect. Quantif. 2015;3:1–8.
44. Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, Tyson JR, Beggs AD, Dilthey AT, Fiddes IT et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat. Biotechnol. 2018;36:338–345.
45. George S, Pankhurst L, Hubbard A, Votintseva A, Stoesser N, Sheppard AE, Mathers A, Norris R, Navickaite I, Eaton C et al. Resolving plasmid structures in Enterobacteriaceae using the MinION nanopore sequencer: Assessment of MinION and MinION/Illumina hybrid data assembly approaches. Microb. Genom. 2017;3:e000118.
46. Nanoporetech.com. (accessed on 27 May 2018). Available:https://nanoporetech.com
47. Schmidt MHW, Vogel A, Denton AK, Istace B, Wormit A, van de Geest H, Bolger ME, Alseekh S, Mass J, Pfaff C et al. De novo assembly of a new Solanum pennellii accession using nanopore sequencing. Plant Cell. 2017;29:2336–2348.
48. Pareek CS, Smoczynski R, Tretyn A. Sequencing technologies and genome sequencing. J. Appl. Genet. 2011;52:413–435.
49. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. Nucleic Acids Res. 2008;36:D25.
50. Kanz C, Aldebert P, Althorpe N, Baker W, Baldwin A, Bates K, Browne P, van den Broek A, Castro M, Cochrane G et al.

The EMBL Nucleotide Sequence Database. Nucleic Acids Res. 2005;33: D29–D33.
51. Duvick J, Fu A, Muppirala U, Sabharwal M, Wilkerson MD, Lawrence CJ, Lushbough C, Brendel V. PlantGDB: A resource for comparative plant genomics. Nucleic Acids Res. 2007;36:D959–D965.
52. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N et al. Phytozome: A comparative platform for green plant genomics. Nucleic Acids Res. 2012;40:D1178–D1186.
53. Lai K, Lorenc MT, Edwards D. Genomic databases for crop improvement. Agronomy. 2012;2:62.
54. Hassani-Pak K, Rawlings C. Knowledge Discovery in Biological Databases for Revealing Candidate Genes Linked to Complex Phenotypes. J. Integr. Bioinform. 2017;14.
55. Hassani-Pak K, Castellote M, Esch M, Hindle M, Lysenko A, Taubert J, Rawlings C. Developing integrated crop knowledge networks to advance candidate gene discovery. Appl. Transl. Genom. 2016: 11;18–26.
56. Yuan Y, Scheben A, Chan CK, Edwards D. Databases for wheat genomics and crop improvement. In Methods in Molecular Biology; Springer: Berlin/Heidelberg, Germany. 2017;1679:277–291.
57. Furbank RT, Tester M. Phenomics–technologies to relieve the phenotyping bottleneck. Trends Plant Sci. 2011;16:635–644.
58. Mackay TF, Stone EA, Ayroles JF. The genetics of quantitative traits: Challenges and prospects. Nat. Rev. Genet. 2009;10:565–577.
59. Mace E, Jordan D. Integrating sorghum whole genome sequence information with a compendium of sorghum QTL studies reveals uneven distribution of QTL and of gene-rich regions with significant implications for crop improvement. Theor. Appl. Genet. 2011;123:169–191.
60. Veyrieras JB, Goffinet B, Charcosset A. MetaQTL: A package of new computational methods for the meta-analysis of QTL mapping experiments. BMC Bioinform. 2007;8:49.
61. Kumasaka N, Knights AJ, Gaffney DJ. Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. Nat. Genet. 2016;48:206.
62. Tecle IY, Menda N, Buels RM, van der Knaap E, Mueller LA. solQTL: A tool for QTL analysis, visualization and linking to genomes at SGN database. BMC Bioinform. 2010;11:525.
63. Liu RX, Zhang H, Zhao P, Zhang ZX, Liang WK, Tian ZG, Zheng YL. Mining of candidate maize genes for nitrogen use efficiency by integrating gene expression and QTL data. Plant Mol. Biol. Report. 2012;30:297–308.
64. Said JI, Lin Z, Zhang X, Song M, Zhang J. A comprehensive meta QTL analysis for fiber quality, yield, yield related and morphological traits, drought tolerance and disease resistance in tetraploid cotton. BMC Genom. 2013;14:776.
65. Qi Z-M, Wu Q, Han X, Sun Y-N, Du X-Y, Liu C-Y, Jiang H-W, Hu G-H. Chen Q-S. Soybean oil content QTL mapping and integrating with meta-analysis method for mining genes. Euphytica. 2011;179:499–514.
66. Hanocq E, Laperche A, Jaminon O, Laine A-L, Le Gouis J. Most significant genome regions involved in the control of earliness traits in bread wheat, as revealed by QTL meta-analysis. Theor. Appl. Genet. 2007; 114:569–584.
67. Borevitz JO, Nordborg M. The impact of genomics on the study of natural variation in Arabidopsis. Plant Physiol. 2003;132: 718–725.
68. Bergelson J, Roux F. Towards identifying genes underlying ecologically relevant traits in Arabidopsis thaliana. Nat. Rev. Genet. 2010;11:867–879.
69. Huang X, Han B. Natural variations and genome-wide association studies in crop plants. Annu. Rev. Plant Biol. 2014;65: 531–551.
70. Korte A, Farlow A. The advantages and limitations of trait analysis with GWAS: A review. Plant Methods. 2013;9:29.
71. Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z et al. Genome-wide association studies of 14 agronomic traits in rice landraces. Nat. Genet. 2010;42:961–967.
72. Sonah H, O'Donoughue L, Cober E, Rajcan I, Belzile F. Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. Plant Biotechnol. J. 2015;13:211–221.
73. Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR,

McMullen MD, Holland JB, Buckler ES. Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat. Genet. 2011;43:159–162.

74. Gacek K, Bayer PE, Bartkowiak-Broda I, Szala L, Bocianowski J, Edwards D, Batley J. Genome-wide association study of genetic control of seed fatty acid biosynthesis in Brassica napus. Front. Plant Sci. 2017;7:2062.
75. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 2007;81: 559–575.
76. Ma C, Blackwell T, Boehnke M, Scott LJ. GoT2D Investigators. Recommended joint and meta-analysis strategies for case-control association testing of single low-count variants. Genet. Epidemiol. 2013;37:539–550.
77. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics. 2007;23:2633–2635.
78. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z. GAPIT: Genome association and prediction integrated tool. Bioinformatics. 2012;28:2397–2399.
79. Jankowicz-Cieslak J, Till BJ. Forward and reverse genetics in crop breeding. In Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools; Al-Khayri JM, Jain SM, Johnson DV Eds.; Springer: Heidelberg, Germany, 2015;1:215–240.
80. Sessions A, Burke E, Presting G, Aux G, McElver J, Patton D, Dietrich B, Ho P, Bacwaden J, Ko C et al. A high-throughput Arabidopsis reverse genetics system. Plant Cell. 2002;14: 2985–2994.
81. Henry IM, Nagalakshmi U. Lieberman MC, Ngo KJ, Krasileva KV, Vasquez-Gross H, Akhunova A, Akhunov E, Dubcovsky J, Tai TH et al. Efficient genome-wide detection and cataloging of EMS-induced mutations using exome capture and next-generation sequencing. Plant Cell. 2014;26:1382–1397.
82. Slade AJ, Fuerstenberg SI, Loeffler D, Steine MN, Facciotti D. A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nat. Biotechnol. 2005;23:75–81.
83. Perry JA, Wang TL, Welham TJ, Gardner S, Pike JM, Yoshida S, Parniske M. A TILLING reverse genetics tool and a web-accessible collection of mutants of the legume Lotus japonicus. Plant Physiol. 2003;131:866–871.
84. Jannink JL, Lorenz AJ, Iwata H. Genomic selection in plant breeding: From theory to practice. Brief. Funct. Genom. 2010;9:166–177.
85. Bhat JA, Ali S, Salgotra RK, Mir ZA, Dutta S, Jadon V, Tyagi A, Mushtaq M, Jain N, Singh PK et al. Genomic selection in the era of next generation sequencing for complex traits in plant breeding. Front. Genet. 2016;7:221.
86. Heffner EL, Sorrells ME, Jannink J-L. Genomic selection for crop improvement. Crop Sci. 2009;49:1–12.
87. Lin Z, Cogan NO, Pembleton LW, Spangenberg GC, Forster JW, Hayes BJ, Daetwyler HD. Genetic gain and inbreeding from genomic selection in a simulated commercial breeding program for perennial ryegrass. Plant Genome. 2016;9:1–12.
88. De Oliveira EJ, de Resende MDV, da Silva Santos V, Ferreira CF, Oliveira GAF, da Silva MS, de Oliveira LA, Aguilar-Vildoso CI. Genome-wide selection in cassava. Euphytica. 2012;187:263–276.
89. Scheben A, Batley J, Edwards D. Genotyping-by-sequencing approaches to characterize crop genomes: Choosing the right tool for the right application. Plant Biotechnol. J. 2017;15;149–161.
90. Voss-Fels K, Snowdon RJ. Understanding and utilizing crop genome diversity via high-resolution genotyping. Plant Biotechnol. J. 2016;14:1086–1094.
91. Poland J.; Endelman, J.; Dawson, J.; Rutkoski, J.; Wu, S.Y.; Manes, Y.; Dreisigacker, S.; Crossa J, Sanchez-Villeda H, Sorrells M et al. Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. Plant Genome. 2012;5:103–113.
92. Rutkoski JE, Poland JA, Singh RP, Huerta-Espino J, Bhavani S, Barbier H, Rouse MN, Jannink JL, Sorrells ME. Genomic Selection for Quantitative Adult Plant Stem

Rust Resistance in Wheat. Plant Genome. 2014;7:1–10.
93. Crossa J, Beyene Y, Kassa S, Perez P, Hickey JM, Chen C, de los Campos G, Burgueno J, Windhausen VS, Buckler E et al. Genomic prediction in maize breeding populations with genotyping-by-sequencing. G3 Genes Genomes Genet. 2013;3:1903–1926.
94. Meyer RS, Purugganan MD. Evolution of crop species: Genetics of domestication and diversification. Nat. Rev. Genet. 2013;14:840–852.
95. Swinnen G, Goossens A, Pauwels L. Lessons from domestication: Targeting cis-regulatory elements for crop improvement. Trends Plant Sci. 2016;21:506–515.
96. Wang S, Li S, Liu Q, Wu K, Zhang J, Wang S, Wang Y, Chen X, Zhang Y, Gao C et al. The OsSPL16-GW7 regulatory module determines grain shape and simultaneously improves rice yield and grain quality. Nat. Genet. 2015;47:949–954.
97. Pauler FM, Stricker SH, Warczok KE, Barlow DP. Long-range DNase I hypersensitivity mapping reveals the imprinted Igf2r and Air promoters share cis-regulatory elements. Genome Res. 2005;15:1379–1387.
98. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A method for assaying chromatin accessibility genome-wide. Curr. Protoc. Mol. Biol. 2015;109:21–29.
99. Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of in vivo protein-DNA interactions. Science. 2007; 316:1497–1502.
100. Korkmaz G, Lopes R, Ugalde AP, Nevedomskaya E, Han RQ, Myacheva K, Zwart W, Elkon R, Agami R. Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. Nat. Biotechnol. 2016;34:192–198.
101. Rombauts S, Florquin K, Lescot M, Marchal K, Rouzé P, Van de Peer Y. Computational approaches to identify promoters and cis-regulatory elements in plant genomes. Plant Physiol. 2003; 132:1162–1176.
102. Van de Velde J, Heyndrickx KS, Vandepoele K. Inference of transcriptional networks in Arabidopsis through conserved noncoding sequence analysis. Plant Cell. 2014;26:2729–2745.
103. Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouzé P, Rombauts S. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res. 2002;30:325–327.
104. Meng XB, Yu H, Zhang YF, Zhuang F, Song XG, Gao SS, Gao CX, Li JY. Construction of a genome-wide mutant library in rice using CRISPR/Cas9. Mol. Plant. 2017;10:1238–1241.
105. Lu YM, Ye X, Guo RM, Huang J, Wang W, Tang JY, Tan LT, Zhu JK, Chu CC, Qian YW. Genome-wide targeted mutagenesis in rice using the CRISPR/Cas9 system. Mol. Plant. 2017;10:1242–1245.
106. Rodríguez-Leal D, Lemmon ZH, Man J, Bartlett ME, Lippman ZB. Engineering quantitative trait variation for crop improvement by genome editing. Cell. 2017;171:470–480.
107. Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. Nat. Rev. Genet. 2015;16:321–332.
108. Walter A, Liebisch F, Hund A. Plant phenotyping: From bean weighing to image analysis. Plant Methods. 2015; 11:14.
109. Liu N, Koh ZX, Goh J, Lin Z, Haaland B, Ting BP, Ong MEH. Prediction of adverse cardiac events in emergency department patients with chest pain using machine learning for variable selection. BMC Med. Inform. Decis. Mak. 2014;14: 75.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/113926