Genet Resour Crop Evol (2023) 70:353–372
https://doi.org/10.1007/s10722-022-01512-5

REVIEW

Next-generation sequencing technology: a boon to agriculture

Balakrishnan Marudamuthu · Tamanna Sharma · Supriya Purru · S. K. Soam · Ch. Srinivasa Rao

Received: 19 April 2022 / Accepted: 3 December 2022 / Published online: 16 December 2022

Abstract Next-generation sequencing (NGS) is a massively parallel sequencing technology that has revolutionized genomic research. The goal of genomics is outlining the genetic composition, collective characterization, and quantification to describe a particular phenotype. NGS has enabled the sequencing of thousands of genomes to study diversity within or between germplasm pools. Advancement in sequencing has also transformed the enhancement strategies in agriculture. The introduction of new platforms employing long-read sequencing ensured more efficient sequencing of large and complex genomes. NGS has replaced the traditional methods of genotyping and plant breeding and simplified the genetic enhancement and modification approaches with the availability of diverse genetic variants information. This review summarizes the available next-generation sequencing platforms and their utilization in agricultural genomics. Agri-genomics has become more efficient after the introduction of NGS technologies with broad applications in genome selection, parentage, genotyping, marker-assisted breeding, and genome editing. Data analysis of high throughput sequencing data allows the exploration of the horizons of evolution and diversity, molecular breeding, soil, and agricultural metagenomics. NGS has proved to be a boon for researchers and breeders, enabling cost-efficient identification of markers for desirable traits that have led to insect-resistant and more productive crops and livestock.

Keywords Agri-genomics · Breeding · Genomics · Genotyping · Next generation sequencing

Abbreviations

NGS	Next-generation sequencing
SBS	Sequencing by synthesis
GS	Genome sequencer
SOLiD	Sequencing by oligo ligation and detection
PGM	Ion personal genome machine
ISFET	Ion-sensitive field-effect transistor
SMRT	Single-molecule real-time
ZMW	Zero mode waveguide
SMS	True single molecule sequencing
WGR	Whole genome resequencing
CNV	Copy number variations
PAV	Presence/absence variation
SNP	Single nucleotide polymorphism
ONT-	Oxford nanopore technologies
WGBS	Whole-genome bisulfite sequencing
ChIP-seq	Chromatin immunoprecipitation sequencing
NGG	Next generation genotyping

B. Marudamuthu (✉) · T. Sharma · S. Purru · S. K. Soam · C. S. Rao
ICAR-National Academy of Agricultural Research Management, Rajendranagar, Hyderabad 500030, India
e-mail: balakrishnan@naarm.org.in

C. S. Rao
e-mail: chsrao_director@naarm.org.in

GBS	Genotyping by sequencing
PCR	Polymerase chain reaction
eDNA	Environmental DNA
ZFNs	Zinc-finger nucleases
TALENs	Transcription activator-like effector nucleases
CRISPR	Clustered regularly interspaced short palindromic repeats
SSNs	Sequence-specific nucleases
RILs	Recombinant inbred lines
QTL	Quantitative trait loci
GS	Genomic selection
GWAS	Genome-wide association studies
BSA	Bulked segregant analysis
GMO	Genetically modified organisms
TILLING	Targeting induced local lesions in genomes
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid

Introduction

Agriculture provides food and other products to sustain the global human population. With the growing population, limited land availability, and water shortage, there is a need to enhance the crop yield and improve the crop traits. This can be accomplished by having a better understanding of the physiological and metabolic pathways of crop production. The evolution of sequencing has made it possible to have a thorough knowledge of genes influencing crop enhancement. Sequencing technologies have progressed from traditional to next-generation sequencing. Traditional sequencing methods include Sanger dideoxy synthesis and the Maxam-Gilbert chemical cleavage method. However, Sanger Sequencing by Synthesis (SBS) dideoxy method became the method of choice for sequencing after incorporating some modifications, such as the use of fluorescent (terminator) dyes, thermal-cycle sequencing, and the introduction of capillary electrophoresis software to interpret and analyze the sequences (Slatko et al. 2018). Human Genome Project, International Rice Genome Sequencing Project, and Arabidopsis Genome Initiative were performed using sanger sequencing only followed by the sequencing of many plant species (Sharma et al. 2018). This is the first-generation sequencing method and is still used nowadays for sequencing a few genes where high throughput sequencing is not required. Advancements in sequencing technologies known as next-generation sequencing include second-generation sequencing such as Illumina or Ion Torrent platforms and third-generation sequencing such as Pacific Biosciences or Oxford Nanopore technology platforms (Taishan et al. 2021). Next-generation sequencers follow the massively parallel mode to sequence whole genomes or transcriptomes and produce large sequence data. Next-generation pyrosequencing includes steps: DNA fragmentation, adapter ligation, amplification, and sequencing.

Genomics is necessary for the improvement of crops to fulfill the rising agricultural demands. Advancements in sequencing technologies have led to the expansion of genomics in plants, animals, and microorganisms, as well as, it has modernized the landscape of sciences. Large, complex, and repetitive crop genomes are also easily sequenced using NGS techniques. Genomes of polyploidy crops such as *Triticum aestivum Gossypium hirsutum*, *Brassica oleracea*, and *Brassica napus* were sequenced using 454 and Illumina sequencing (Choulet et al. 2010; Li et al. 2015; Chalhoub et al. 2014). Further, sequencing of complex polyploidy plant genomes such as *Chenopodium quin*oa genome and *Solanum pennellii* was performed using Single Molecule Real-Time (SMRT) and Oxford Nanopore Technologies, respectively (Jarvis et al. 2017; Schmidt et al. 2017). NGS defines new sequencing methods with reduced cost, high speed, and better coverage than traditional sequencing methods. The current review summarizes the next-generation sequencing technologies and their applications in the field of agriculture. Various methods employing these NGS platforms have been developed and utilized in the agricultural field to enhance crop productivity. One of the high-throughput sequencing-based approaches, Genotyping by Sequencing (GBS) has been used for the identification of markers in crop genetics (Poland and Rife 2012). In this review, we have focused on how next-generation sequencing has revolutionized the existing agricultural research by introducing different techniques utilizing next-generation sequencing technologies.NGS platforms

NGS platforms

As mentioned above, next-generation sequencing technologies are classified into two categories: the second generation and third-generation sequencing. This section evaluates different NGS platforms used in the field of agriculture and how genomic information extracted from these technologies is enhancing agricultural research.

Second generation sequencing platforms

Roche 454 genome sequencer FLX

454 sequencing utilizes a large-scale parallel pyrosequencing system in which a pyrophosphate particle is released after the incorporation of polymerase-dependent deoxynucleotide triphosphate to operate a downstream set of luciferases oxyluciferin responses that ultimately generates light. In this technology, DNA is fragmented into 800 to 1000 bp in length followed by adapter ligation for library preparation using small DNA-capture beads. Emulsion PCR is used for the clonal amplification of DNA molecules bound to these beads followed by enrichment of clonally amplified DNA molecules. These DNA-bound beads are transferred on a Pico Titre Plate with the addition of DNA polymerase, ATP sulfurylase, and luciferase and then kept for sequencing. 454 Genome Sequencer (GS) FLX Titanium platform delivers about 100,000 sequences in a single run duration of 10 h, with an average read length of 330–500 bp. However, the latest GS FLX platform can generate reads of length 700–1000 bp (Gupta and Gupta 2014). This platform has been used to sequence comparatively long stretches of repetitive DNA in *Hordeum vulgare*, one of the most important cereal crops. 454 Roche GS FLX platform was used in this study to obtain reads with average lengths between 200 and 250 bp and assemblies achieved High Throughput Genome Sequences phase I quality with N50 (Steuernagel et al. 2009). In another study, transcriptomes of *Jatropha curcas* L., an important non-edible oilseed crop, have been sequenced using 454 pyrosequencing enabled GS FLX titanium platform. 197.7 Mb data was generated from 383,918 reads with an average read length of 515 bases (Natarajan and Parani 2011). Whole-genome sequencing of *Solanum tuberosum*, the third most important food crop by the Potato Genome Sequencing Consortium (PGSC) was done using both 454 GS FLX and Illumina GA2 instruments (Visser et al. 2009).

Illumina genome sequencing analyzer

The Illumina Genome Analyser was introduced in 2006 by the name of Solexa which was purchased by Illumina in 2007. It is reversible terminator-based sequencing in which reversible terminators are used instead of dideoxy terminators in sequencing by synthesis approach. Illumina sequencers use a glass flow cell coated with millions of oligonucleotides that are complementary to the sequencing adaptors. The flow cell is divided into eight distinct routes, and the internal surface of each flow cell is covalently bound and linked to the fragments of the library. Hybridization of each library fragment with the primers is followed by amplification to generate millions to billions of clonal clusters. Then, fluorescently labeled nucleotides are used to synthesize a complementary strand for each fragment. The flow cell is imaged after the addition of each tagged nucleotide, and the emission from each cluster is recorded. Based on fluorescent emission intensity and wavelength, sequences of the templates are identified. Different Illumina sequencing platforms are available depending on the number and length of reads such as benchtop sequencers: ISeq, MiniSeq, MiSeq, NextSeq, and production-scale sequencers: NextSeq, HiSeq, and NovaSeq (Kulski 2016). Due to the low cost and improved assembly programs of Illumina, it is the most preferred platform for sequencing crop genomes. Many crop genomes have been sequenced by Illumina technology such as *Hordeum vulgare, Triticum aestivum, Oryza sativa, Zea mays*, and *Glycine max*.

Applied biosystems SOLiD sequencer

Sequencing by Oligo Ligation and Detection (SOLiD) platform is based on the sequencing by ligation technology in which DNA ligase is used to detect and incorporate bases in a particular pattern. This technique includes oligo adaptors attached to DNA fragments through magnetic beads connected to complementary oligos and each bead-DNA complex is amplified by PCR emulsion. Hybridization of adapter sequences attached to amplified DNA fragments is

allowed with specific primers that deliver a free 5′ phosphate group for ligation to fluorescently labeled probes. Following ligation, generated fluorescence corresponds to the probe ligated. This process of ligation, detection, and cleavage is repeated multiple times which corresponds to the read length (Gupta and Gupta 2014).

Ion personal genome machine (PGM) sequencer

This platform is based on the detection of hydrogen ions released after the incorporation of a nucleotide in a newly synthesized strand by a polymerase. The pH of the solution is changed by the charge of the released ion which is detected by the proprietary ion sensor. Massively parallel detection of sequencing reactions is possible with the help of an ion torrent chip that contains millions of ion-sensitive field-effect transistor (ISFET) sensors. It is the first commercial sequencing procedure without the use of fluorescence and camera scanning (Pereira et al. 2020).

Third-generation sequencing platforms

Also known as single-molecule sequencing because it doesn't include the amplification step during sequencing library preparation as exists in second-generation sequencing. Also, the read lengths obtained are much longer than the read lengths from second-generation sequencing techniques, hence called long-read sequencing (Christoph 2016).

Pacific biosciences

It is based on Single-Molecule Real-Time (SMRT) sequencing in which a single-stranded molecule of DNA is attached to a polymerase enzyme. Polymerase activity is monitored while incorporating fluorescently labeled nucleotides in newly synthesized strands. This process is enabled with the use of Zero Mode Waveguide (ZMW), extremely small wells made in a metal film deposited on a glass surface. This sequencing is comparatively fast and takes about 4 h per SMRT cell. Read lengths obtained are about 10 kbp, where reads of 54 kbp can also be generated (Christoph 2016).

Oxford nanopore sequencing

This sequencing technology is based on biological nanopores, tiny holes made up of certain transmembrane cellular proteins. Each nanopore is linked to its electrode associated with a channel and sensor chip, which measures the electric current flowing through the nanopore. When a moving nucleotide passes through a nanopore, the current is interrupted and this minute change in the current is decoded to determine the sequence of the passing nucleotide. Every nucleotide disrupts the nanopore to a different level and hence the change in electric current is also different to enable the detection of each nucleotide (Mehdi et al. 2017).

Helicos single molecule sequencing

This sequencing technology is also based on single molecule sequencing, hence called Helicos True Single Molecule Sequencing (tSMS). Firstly, single molecules of template DNA are attached to the specific proprietary surface with the addition of polymerase and one of the fluorescently labeled nucleotides (C, G, A, or T), which is incorporated into the growing complementary strands on all the templates in a sequence-specific manner. After the washing step, imaging of incorporated nucleotides is performed, and their positions are recorded. Subsequently, the fluorescent group is removed in a highly efficient cleavage process, leaving behind the incorporated nucleotide. This process continues through each of the other three bases in multiple cycles, providing a 25–55 bp read (average 35 bp) from each of those individual templates. From 600 million to 1 billion DNA strands, a total of 21 to 35 GB of sequence data is generated per run with 99.995% accuracy (Gupta and Gupta 2014).

NGS technologies used in agricultural research

Next-generation sequencing has become an important genomics tool for the enrichment of genetic gain in plant species. It has enabled the sequencing of plant genomes that are of economic importance resulting in a lot of genomic information that can

be extracted and utilized for trait improvement. It includes whole-genome sequencing, transcriptome sequencing, and metagenomics which led to the discovery of functional genes and markers for desired traits through molecular assisted breeding and to improve crop production and conservation (Fig. 1).

Whole-genome sequencing

For agriculture, the generation of genomic data provides new ways to improve food security, reduce poverty, and introduce new reforms in species conservation programs. NGS has made it possible to sequence more complex genomes in a short period and in a cost-effective way. Crop genomes are more complex due to the repetitive sequences and polyploidy that is problematic for gene identification and understanding the function of the gene. Further, the presence of repetitive sequences, whole-genome duplication, and polyploidy led to the large genome size of crops. However, sequencing has made it easier to get information on all genes present in the genome of a specific crop of interest. The first crop to have its genome sequenced was rice (Delphine et al. 2012). Two subspecies of rice i.e., *Oryza sativa* L. species *japonica* and *indica* were sequenced using a combination of hierarchical clone-by-clone and whole-genome shotgun sequencing (Goff et al. 2002). After that, sequencing of bigger genomes, such as *Glycine max* (Schmutz et al. 2010), *Populus trichocarpa* (Tuskan et al. 2006), and *Vitis* (Jaillon et al. 2007; Velasco et al. 2007) was also performed by the whole genome shotgun sequencing method. Next-generation sequencing enabled the sequencing of complex plant genomes in a time and cost-effective manner. These technologies have been applied to sequence the genomes of different crops such as *Malus domestica, Cucumis melo, Theobroma cacao, Cajanus cajan, Cicer arietinum, Musa, Solanum tuberosum, Brassica oleracea, Citrullus lanatus, Citrus sinensis*, and *Picea* (Michael et al. 2015). Different platforms of short-read sequencing such as Illumina and 454 pyrosequencing were used to sequence the genomes and the assembled short reads were mapped to a reference genome for studying species diversity. Also, the up-gradation in second-generation sequencing technologies has facilitated the generation of complete and contiguous assemblies *de novo*, but the complexity of transposable elements and repetitive sequences resulted in partial genome assemblies

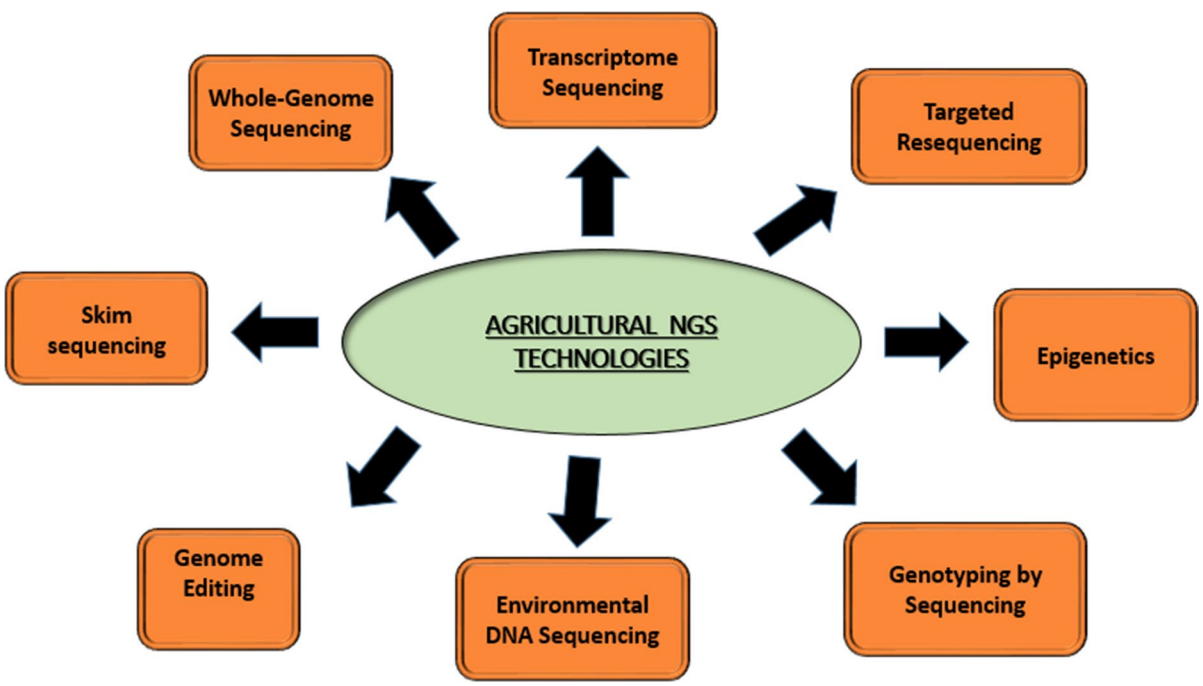

Fig. 1 NGS Technologies employed in agricultural research

in these regions. Third-generation sequencing (long-read sequencing) has proven a boon for these genomes giving complete and contiguous assemblies of the genomes. Next-generation platforms, Illumina HiSeq and NextSeq, Pacific Biosciences, and Nanopore Sequencing technology have been employed for sequencing most crop genomes recently (Table 1).

Whole-genome sequencing is categorized into two sequencing methods:

- Whole-genome resequencing.
- *De novo* sequencing.

Whole-genome resequencing

It is the method of sequencing the entire genome when a reference species genome is available for mapping. Whole-genome resequencing (WGR) of crop genomes enables the identification of different markers such as copy number variations (CNV), Indels, Single Nucleotide Polymorphisms (SNPs), and presence/absence variations (PAV) which provides a deeper understanding of genetic variations in crops species (Xu and Bai 2015). Advancement in NGS technologies has resulted in the sequencing of many genomes which has increased the list of available reference genomes for performing whole-genome resequencing to study genomic variations and specific signature discovery in different species of a particular crop. WGR has been applied to the genomes of *Oryza sativa* (Huang et al. 2009), *Zea mays* (Lai et al. 2010) *Glycine max* (Lam et al. 2010), and many other crops. WGR of crop genomes has revealed various CNVs and PAVs that have agronomic importance. In a study by Tong et al. whole-genome sequencing of 445 *Lactuca* accessions from 47 countries, was performed which provided a comprehensive variation map. SNPs, Indels, and structural variants were identified to reveal the phylogenetic relationship and the domestication history of cultivated lettuce (Wei et al. 2021). Also in a recent study, WGR of *Cucurbita pepo* identified genes controlling early flowering which is the prerequisite for selecting varieties for cultivation. Indels present in the promoter regions of genes were utilized to discover markers to distinguish between cultivars (Abbas et al. 2022). WGR of *Sorghum* discovered various SNPs and Indels, which further resulted in the development of an SNP marker that could be used in molecular breeding to improve aphid resistance in *Sorghum* (Wei et al. 2021). Resequencing of 588 diverse *Brassica napus* accessions from 21 countries uncovered different SNPs, Indels that are involved in improving phenotypic traits (Lu et al. 2019).

De novo ***sequencing***

Several species of crops are distant from their sequenced closed species. In such cases, where no reference genome is available for resequencing, *de novo* sequencing must be performed for exploring their genomes. However, large genome size and high repetitive content is the hurdle for the utilization of this strategy. In *de novo* sequencing, reads are assembled to contigs, whose size and continuity determine the coverage quality of sequencing. Short read *de novo* assemblers are not efficient for obtaining high-level contiguity and assembling genomes having high repetitive content (Liao et al. 2019). The growth of long-read sequencing facilitated the development of new assembly algorithms to generate the complete and gapless assembly of crop genomes with heterozygous, polyploidy, and high repetitive content (Gao et al. 2019). *Oropetium thomaeum* was the first plant genome sequenced by Pacific Biosciences single-molecule real-time (SMRT) sequencing. It was the fourth most contiguous genome including 30% complete centromeres (VanBuren et al. 2015). Further, *de novo* genome sequencing of *Solanum lycopersicum, Musa, Sorghum, Brassica*, and many other crops was performed by Oxford Nanopore Technologies (ONT). Oxford Nanopore third-generation sequencing and Hi–C technology were employed for the *de novo* sequencing of (*Olea europaea*) that provided a reference genome for the study of gene function and molecular breeding in olive (Rao et al. 2021).

Transcriptome sequencing

Transcriptome sequencing also known as RNA-Seq, is the sequencing of transcriptionally active elements of a genome in a specific condition. It provides information about gene expression analysis, functional genomic studies, and gene characterization, in the absence of a sequenced genome (Afzal

Table 1 List of crop genomes sequenced recently using next-generation sequencing technologies

S. No.	Species	Common name	Genome size	NGS platform	Traits analyzed	Reference
1.	*Solanum pimpinellifolium*	Tomato	807.6 Mb	PacBio Sequel	Identification of Structural variants and mapping of disease resistance and fruit quality traits	Wang et al. (2020)
2.	*Oryza sativa* L.	Rice	390 Mb	Illumina HiSeq2500	Evaluation of genomic features and genetic variation in rice for genetic studies and marker development	Kishor et al. (2020)
3.	*Triticum* sp.	wheat	14 GB	Illumina PE, MP, 10X Genomics Chromium, and Hi-C	Exploration of genomic diversity	Walkowiak et al. (2020)
4.	*Hordeum vulgare* L.	Barley	4.1 GB	Illumina HiSeq 2500, PacBio Iso-Seq	20 varieties of barley sequenced for estimating PAVs and exploration of structural variants for quantitative genetic analysis	Jayakodi et al. (2020)
5.	*Brassica oleracea*	–	587 Mb	Illumina HiSeq2000 and HiSeq2500	Identification of PAVs associated with agronomic traits, including disease resistance, flowering time, glucosinolate metabolism, and vitamin biosynthesis	Golicz et al. (2016)
6.	*Glycine max*	Soybean	992.3 Mb – 1059.8 Mb	Pacific Biosciences Sequel, Illumina HiSeq2500	Pan-Genome: 26 accessions analyzed to link genetic variations in candidate genes involved in important traits	Liu et al. (2020)
7.	*Sorghum bicolor* L.	Jowar	729.4 Mb	PacBio RS II system	Key changes in signaling molecules and transcriptional regulators that represent new candidates for understanding and improving sugar metabolism in this important crop were identified	Cooper et al. (2019)
8.	*Vigna mungo* L.	Blackgram	475 Mb	Oxford Nanopore	Identification of agronomically important genes for genetic improvement.	Jegadeesan et al. (2021)
9.	*Gossypium raimondii & G. turneri*	Cotton	880 Mb and 910 Mb	PacBio HiC, and Bionano optical mapping	Provided resources for the genetic improvement of cotton and for performing evolutionary analysis.	Udall et al. (2019)

Table 1 (continued)

S. No.	Species	Common name	Genome size	NGS platform	Traits analyzed	Reference
10.	*Aegilops tauschii*	Tausch's goatgrass	5.71 Gb	Illumina MiSeq and Pacific Biosciences	Compared with other sequenced crop genomes for revealing the genomic characteristics and evolution	Luo et al. (2017)
11.	*Apium graveolens* L.	Celery	3.18 Gb	Illumina HiSeq 2000	Identified agronomically important genes involved in disease resistance, flavonoid biosynthesis, terpenoid metabolism, and other important cellular processes	Li et al. (2020)
12.	*Secale cereale*	Rye	7.74 Gb	PacBio Sequel and Illumina X Ten	Identified genomic characteristics and genes, chromosome regions, and loci for evaluating the genetic basis of rye domestication	Li et al. (2021)
13.	*Arachis hypogaea* L.	Groundnut	2.54 Gb	PacBio Sequel and PacBio RS II	Identified mapping and candidate-gene discovery for seed size and color, foliar disease resistance, and other traits for peanut improvement.	Zhuang et al. (2019)
14.	*Zea mays*	Maize	2.16 Gb	PacBio Sequel	Identification and genotyping of structural variants across 521 diverse inbred lines unveiling genetic variation in maize	Yang et al. (2019)
15.	*Ananas comosus* L.	Pineapple	513 Mb	PacBio RSII and Illumina	Elucidation of genomic basis for fiber production, color formation, sugar accumulation, and fruit maturation	Chen et al. (2019)

et al. 2020). The transcriptome can be assembled *de novo* or it can be mapped with the reference genome for studying genetic variation. It is comparatively easy to assemble and functionally annotate the transcriptome *de novo* than the genome (Bryant et al. 2017). The generated transcripts from transcriptome sequencing are reliant on the specific condition, whether it is the organ, stress, developmental stage, or external stimuli. Hence, it doesn't provide information about structural variation (Schreiber et al. 2018). Transcriptomes of various crops have been sequenced and gene mining has been performed to study the regulatory mechanisms, gene expression analysis, and the identification of biosynthetic pathways and key genes associated with the particular trait of interest. Transcriptomes of *Vitis amurensis* (Xu et al. 2014), *Aloe vera* (Pragati et al. 2018), *Paeonia suffruticosa* (Guo et al. 2018), *Trillium govanianum* (Singh et al. 2017), *Polygonum minus* (Loke et al. 2017), *Cornus officinalis* (Hou et al. 2018), *Dracocephalum tanguticum* (Li et al. 2017), *Salvadora oleoides* (Bhandari et al. 2020) and many more have been sequenced to gain a deeper understanding of their molecular mechanisms.

Targeted resequencing

Targeted resequencing provides information on the gene of interest (exome) identified from association mapping studies. This strategy is more efficient and economical as sequences that are present in fixed areas of genetic variation regions over a big number of samples are used to identify distinctive variations (SNPs, CNVs) that provide information about breeding decisions or to characterize disease susceptibility. Compared to whole-genome sequencing approaches, targeted resequencing is a cost-effective approach of studying the particular region of interest. This sequencing method provides good coverage quality and allows the identification of genetic variants that are rare and difficult to sequence through whole-genome sequencing (Bewicke-Copley et al. 2019). These variants reveal the beneficial mutations that help in notifying the breeding choice as well as causative mutations for plant or animal disease, or susceptibility to parasites. In a study by Pankin et al. targeted resequencing of 433 wild and domesticated barley accessions was performed for the identification of candidate domestication genes. Further, phylogenetic and ancestry analyses revealed the origin of domesticated barley haplotypes (Pankin et al. 2018).

Epigenetics

Alterations in the environment (drought stress, food accessibility, etc.) influence the adaptive responses that cause physiological changes, affecting the viability and reproductive fitness of plants and animals. Epigenetic regulation is accomplished by mechanisms such as posttranslational histone modifications, DNA methylation, and noncoding RNAs that alter the chromatin states leading to gene activation or gene silencing, respectively. Next-generation sequencing has enabled the study of epigenetic regulation at the genome level, referred to as epigenomics. Investigating the epigenetic variation at the genome level provides new insights into the phenotypic diversity in species having low genetic variation. Comparative analysis of epigenomes led to the development of biomarkers specific for a particular genotype or condition, which is termed plant epibreeding (Kapazoglou et al. 2018). In a study by Forestan et al. (2018) chromatin data from genomic and transcriptomic sequencing were analyzed to study how *Zea mays* responds to drought and recovers from drought stress. Several studies have been reported on the analysis of epigenetic modifications in *Triticum aestivum* under abiotic stress. In one such study, microRNAs, heterochromatic small interfering RNA, and small regulatory RNAs were found to be involved in wheat drought stress response (Budak et al. 2015). Also, sequencing of *Hordeum vulgare* revealed several differently methylated genes to be induced in leaves than roots, when exposed to drought and salt stress (Chwialkowska et al. 2016). Epigenetic studies have been performed on many important crops, such as *Oryza sativa, Solanum lycopersicum, Brassica napus*, and many legume crops (Varotto et al. 2020). Two technologies have emerged for epigenetic analysis of genome using next-generation sequencing:

- Whole-genome bisulfite sequencing (WGBS).
- Chromatin immunoprecipitation (eChIP) approach.

Whole-genome bisulfite sequencing (WGBS)

This method allows the evaluation of DNA methylation throughout the genome. In this process, genomic DNA is treated with sodium bisulfite that converts unmethylated cytosines into uracil. Subsequently, PCR amplification, library preparation, and sequencing are performed to determine the untreated and treated sequences for the identification of methylated sequences. Genome-wide DNA methylation levels in *Zea mays* and *Hordeum vulgare* have been accessed using this technology (Li et al. 2018).

Chromatin immunoprecipitation sequencing (ChIP-seq)

Chip technology in combination with the sequencing approach allows the identification of transcription factor binding sites and histone modifications at a genomic scale. Chromatin immunoprecipitation is profiling of the chromatin component of epigenomes. ChIP procedure followed by sequencing is called ChIP-seq. In a study by Zhao et al. genome-wide profiling of five histone modifications and RNA polymerase II was performed using the eChIP approach, and epigenome landscapes were constructed in 20 representative rice varieties (2020).

Genotyping by sequencing (GBS)

Next-Generation Genotyping (NGG) or genotyping by sequencing (GBS), is a less expensive technique of genome screening to explore different plant and animal SNPs for genotyping research (Elshire et al. 2011). This method is a restriction enzyme-mediated complexity reduction followed by sequencing to discover random markers across an entire genome (Jiang et al. 2016). Large numbers of SNPs are generated for genetic analysis and genotyping. It is a low-cost tool with which researchers can accelerate the screening of breeding lines and breeding practices. Through this simple, quick, highly reproducible, extremely specific approach, it is possible to reach significant genomic regions that are unapproachable for other sequence capture-based methods (Poland et al. 2012). The major benefits of this system include reduced handling of samples and fewer PCR and purification steps, inexpensive barcoding, and no size fractionation. Initially, GBS was developed for performing high-resolution association studies in *Zea mays* and then has been used for other species with complex genomes. It is a cost-effective strategy that has been used for discovering and genotyping SNPs in various crop species, performing populations studies, and studying plant genetics and breeding in crop genomes (Truong et al. 2012). Various studies are reported to use the GBS method for genetic analysis and marker development in different crops such as *Lactuca sativa* (Poland et al. 2012), *Brassica napus* (Yang et al. 2012), *Panicum virgatum* (Bus et al. 2012), *Lupinus* (Lu et al. 2013), and *Glycine max* (Sonah et al. 2013). GBS approach has also been shown to optimize cereals crops such as *Triticum aestivum, Oryza sativa, Zea mays, Sorghum, Hordeum vulgare, Avena sativa*, and tuber crops like *Manihot esculenta, Solanum tuberosum*, and *Gossypium*, an industrial crop (Deschamps et al. 2012; He et al. 2014c).

Advantages of GBS include.

- Simple and quick procedure compared to conventional approaches.
- Easy computational data analysis.
- High accuracy of SNP arrays.
- Low cost creates more attractive and availability of large numbers of markers for scientists.
- A small amount of DNA input is sufficient.

Environmental DNA sequencing

Environmental DNA (eDNA) sequencing is a rapidly emerging technique that allows the characterization of microorganisms in the soil, aquatic, systems, and other samples. This approach is utilized for monitoring ecosystem changes, studying biodiversity, and testing water and soil suitability. Organisms shed their DNA into the environment they live in, from which eDNA is characterized and its analysis provides clues about the species present in the environment without disrupting the ecosystem. NGS allows the profiling of thousands of species simultaneously from a single sample. It also provides the required sensitivity for the detection of eDNA, when present at low levels in the environment (Van Poecke et al. 2013). In eDNA sequencing, firstly the eDNA is analyzed by PCR amplification using the single-species approach with specific primers or by using a multiple-species (multiple-taxon) approach with generic primers. After

PCR amplification, DNA sequencing is performed. Evolving NGS technologies have made it possible for performing wide-range biodiversity surveys easily with reduced costs. eDNA metabarcoding is the technique where mass sequencing of species from a complex sample is performed. It is a more powerful and cost-efficient method than single species identification. However, there are some drawbacks of eDNA metabarcoding. With the utilization of generic primers, primer affinity bias results in less efficient amplification of certain sequences than others. Also, data analysis and interpretation are more complex in high throughput sequencing (Thomsen and Willerslev 2015).

Genome editing

Genome editing is defined as the utilization of various molecular biology techniques such as zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) for introducing particular modifications at specific genomic loci. However, with the development of clustered regularly interspaced short palindromic repeats (CRISPR)/Cas systems, there is exceptional progress in genome engineering due to ease and precise DNA manipulation (Zhang et al. 2018). Genome-editing has been applied to a variety of crops and has transformed crop improvement. Insertion, expression, and silencing of certain significant genes are applied to alter or enhance the particular trait of crops. Genomic information is the prerequisite for the application of genomic editing techniques. CRISPR/Cas9 and TALENs genome-editing systems utilize sequence-specific nucleases (SSNs) to create double-stranded breaks (DSBs) at the target locus for targeted insertion, disruption, or replacement of genes in plants. The advantages of the CRISPR/Cas9 system have led to its utilization in genome editing of various crop species (Jiang et al. 2013; Sun et al. 2015; Svitashev et al. 2016; Zhang et al. 2016; Shimatani et al. 2017; Soyk et al. 2017). NGS in combination with CRISPR genome editing technology is applied to resolve the issues related to viruses causing diseases in crops at the genomic level (Mushtaq et al. 2021).

Skim sequencing

Skim sequencing, also known as low coverage genome sequencing, is a less complex approach to genome sequencing for performing high-resolution genotyping of multiple genomes. It omits the complexity of genome reduction and low deep sequencing that occurs in genotyping by sequencing approach. Recombinant inbred lines (RILs) and doubled haploid populations are the most utilized populations for skim sequencing. It provides better coverage than the GBS method and is used for the identification of SNPs, recombination events, quantitative trait loci (QTL) mapping, and genome-wide association analysis. Availability of reference genome is a prerequisite for performing skim sequencing. Firstly, it aligns the reads from the population to the reference genome followed by variant calling. Variants having more than 80% missing alleles are discarded and the variants present in a particular position are called SNPs. This process is known as SNP genotyping. After that, SNP filtering is done to remove the false positive SNPs resulting from sequencing errors. Lastly, SNP imputation for collecting the missing genotype is performed. Skim sequencing has been used for a variety of crops such as *Oryza sativa, Gossypium, Nicotiana tabacum, Cicer arietinum*, and many others (Kumar et al. 2021).

Applications of NGS in agriculture

Trait screening/genomic selection

Genomic Selection (GS) is a marker-based selection technique in which genomic loci, molecular markers, and haplotypes are utilized for developing novel breeding strategies. It evaluates the genomic signatures containing thousands of genetic markers for the prediction of complex traits (Dekkers 2007). Advancements in sequencing technologies have led to the development of genome-wide high throughput markers that are cost-effective and flexible. Hence, NGS-based genotyping has become a promising agrigenomics tool for performing genome selection in both model and non-model crop species as well as for crops with large and complex genomes. Moreover, in this case, genome selection can be performed in the absence of a reference genome. Genetic markers

linked with desired traits are exploited to screen a large population of progeny related to specific features. Trait screening is best for multigene characteristics that are hard to manage with standard breeding or propagation methods. Recurrent population screening enables obvious segregation of the progeny with the required characteristics to be used for breeding practices. Genome selection has been successfully performed in various crop species such as *Triticum aestivum, Oryza sativa, Zea mays, Glycine max*, and *Panicum virgatum* (Bhat et al. 2016). Recently, GS has been attempted in common bean where a panel of elite Andean breeding lines was evaluated for various agronomic traits in two locations under drought, irrigated, and low phosphorous conditions (Keller et al. 2020). Availability of different crop pan genomes has also contributed to the genome selection where the genomic data is linked with the phenotype of interest. *Solanum lycopersicum, Hordeum vulgare*, and *Cajanus cajan* pan genomes have been utilized by the genome selection approach for studying the possible associations of the desired phenotypes. The establishment of pan genomes before employing genome selection is the prerequisite for incorporating the broader extent of markers in the model training process (Bayer et al. 2021).

Evolution and diversity

Compared to other eukaryotic organisms, plant genomes have a high evolution rate that led to greater genome diversity (Murat et al. 2012). Whole-genome duplication is the main mechanism behind the evolution and diversification of plants. Plant genomes have several gene duplicates, some of them are not essential for cell functioning, while some become functional genes via subfunctionalization, neofunctionalization, or non-functionalization. Whole-genome sequencing facilitates the comparison of sequences of the same species (Species Pan-genome) or interprets genomes within the family or genus (Clade Genomes). Whole-genome sequencing has wide applications particularly in population evolution, through the introduction of linkage disequilibrium, phylogenetic analysis, species formation, and genetic structure to investigate mechanisms of biological evolution. Also, mutations such as insertion, deletions, substitutions, and structural variations led to the alteration in the nucleotide sequences of genomes leading to genetic diversity within species. Not only mutations but also different breeding methods cause genetic diversity (Yaman and Uzun 2020; Yaman and Uzun 2021). Cost-efficient sequencing technologies have made it possible to sequence these genomes and to analyze the genetic diversity between or within species of crop genomes (Unamba Chibuikem et al. 2015).

Genome-wide association studies (GWAS)

Also known as genome-wide association (GWA) study, is an association mapping approach in which markers are identified across the complete sets of DNA, or genomes, for a large population to find genetic variations linked with a specific trait. These genetic associations provide information for the enhancement or improvement of crop species. The association mapping study provides various plant and animal whole-genome wide selection applications such as fingerprinting, marker-assisted breeding, and net merit, to improve the significance of commercial crops where millions of genetic variants are read by SNP arrays in a sample of DNA. GWAS has been used in various crops for a better understanding of the genetic architecture of important traits namely days to flowering, resistance, panicle architecture, fertility restoration, and other agronomic traits (Berhe et al. 2021). Agriculturally important traits of various crop species including *Hordeum vulgare, Triticum aestivum, Oryza sativa, Zea mays, Glycine max, Sorghum*, and *Gossypium hirsutum* have been explored by the GWAS approach (Liu and Yan 2019). With next-generation sequencing technologies, fast and cost-efficient SNP discovery has become possible across the genome, leading to the development of high-throughput genome-wide SNP genotyping. Genome-wide SNP genotyping has enabled the use of GWAS study in many significant crops. In a recent study by Yang et al. (2022), GWAS was performed on stems of cucumber. Genes related to cucumber stem diameter were identified and the data is beneficial for the development of new cucumber varieties with strong growth potential and high yield.

QTL analysis

Conventional QTL mapping requires genotyping and phenotyping of many progenies from a bi-parental mapping population, which is time-consuming and labor-intensive. However, QTL detection by next-generation sequencing provides an efficient alternative approach to conventional QTL analysis as it significantly reduces the scale and cost of analysis with comparable power to QTL detection using a full mapping population. Various methods are utilized for QTL mapping, most efficient among them are whole-genome sequencing-based bulked segregant analysis (BSA) and genotyping-by-sequencing (GBS). In the NGS-based BSA analysis, members of a segregating population with two different phenotypes are taken and sequenced using NGS platforms. Comparison of the allele frequency must be done to detect differences associated with phenotype which corresponds to QTLs linked with the trait. Subsequently, a confidence interval for the location of QTL is determined using suitable statistical tests. Various tools specific to BSA-based QTL mapping are available such as standalone software package (QTL-seq package), and R package (QTLseqr) for the analysis. This approach has been applied to map QTLs for traits of various crops (Deokar et al. 2019). QTLseq has been successfully applied in *Oryza sativa* and *Cucumis sativus* to identify major genes underlying QTLs linked with blast resistance and seedling vigor (Takagi et al. 2013; Lu et al. 2014). High-throughput QTL-seq approach was employed in an intra-specific F_4 mapping population of *Cicer arietinum* to identify a major genomic region harboring the robust QTL associated with 100-seed weight in chickpea (Das et al. 2015). In combination with differential expression profiling and diversity analysis, QTLseq is utilized in the rapid identification of potential candidate genes. This approach has been applied to crops, such as *Oryza sativa* and *Setaria italica* (Wang et al. 2010; Gedil et al. 2016). Multi-seq is another approach where QTL-seq is applied to several mapping populations derived from crosses with at least one common parent. MQTL-seq has been successfully applied to two F5 mapping populations of chickpea to identify the commonly found major genomic regions (Nguyen et al. 2019).

GBS as already discussed above involves library preparation and sequencing approaches. Restriction site-associated DNA sequencing is a cost-effective GBS method that standardizes the number and coverage of genotyped loci and SNPs by altering the enzymes used and the sequencing depth. GBS has been used for linkage mapping and QTL analysis in a range of crops (Scheben et al. 2020).

Agricultural and soil metagenomics

Metagenomics facilitates the research on large microbial communities directly under their natural and environmental circumstances which involve the study of complex and diverse populations of microbes related to plant and animal development. Soil plays an important part in plant growth, hence understanding of associated microorganisms provides information for soil health as well as enhancement of crop yields. NGS technologies along with PCR and DNA fingerprinting techniques have a significant role in microbial research (Kaushik et al. 2020). The speed and accessibility of NGS technology have improved the field of metagenomics. Next-generation sequencing system provides an effective strategy for screening samples from the environment, and it is fast becoming an essential tool to study microbial diversity. Novel methods such as metabolomics and biotic factors are more important for understanding the microbiome. In a study by Sabale et al. (2019), diazotrophs were identified from the rhizosphere soil of red kidney beans using metagenomics approach by targeting nitrogen fixing gene nifH. Metagenomic Next-Generation Sequencing (mNGS) has been utilized for the identification and characterization of novel viruses from various samples of livestock, including poultry, cattle, pig, and small ruminants (Kwok et al. 2020). Plant fitness and other agricultural features, such as quality traits, soil biogeochemical properties, and crop yield are influenced by the microbiome of soil, plant, and livestock (Iquebal et al. 2022). A recent study has revealed the genome-resolved diversity of Phosphate-solubilizing Bacteria in agricultural soils using a metagenomics approach, which enhances the understanding of targeted engineering and is helpful in management practices for sustainable agriculture (Wu et al. 2022). Hence, metagenomics has its broad application in the field of agriculture by exploring unexploited soil microorganisms, identifying their functionalities and essential genes, and improving crop productivities, nutrient cycling, and

phytopathogen resistance (Chidinma and Oluranti 2022).

Genotyping and marker-assisted breeding

Genome resequencing has allowed genetic variants detection by sequencing the genomes of the same species but with different accessions. Different genotyping platforms have been used for obtaining large-scale marker segregation data on mapping populations for the construction of genetic maps. Various molecular markers for breeding are also identified from the sequenced genome. However, the development of sensitive markers to select desirable ecotypes is critical in plant breeding (Wenqin et al. 2017). The objective of marker-assisted breeding is to develop abiotic and biotic stress-tolerant varieties with high yields. The desired traits are transferred from the donor parent to the offspring. This strategy allows the detection of trait genes and transformants through the identified markers associated with traits for releasing commercially important varieties or breeding stock (Varshney et al. 2005). The markers are very efficient to transfer traits to linked species, known as anchor markers, used for evolutionary studies and comparative genomic analysis.

Parentage

Molecular markers are used for determining parentage based on marker genotypes by constructing pedigrees in outcrossing genotypes. The evaluation of a distinct marker is not sufficient, various markers are used to assess the possibility to detect the real parent. For inline breeding, where there are multiple generations of males or females present in the group, the interpretations of the markers are pooled to determine parentage with the breeder's potential data. In addition to breeding, paternity analysis is also utilized for maintaining paternal balance in poly crosses, tracking pollination events across distances, and evaluating progeny relatedness. Next-generation sequencing provides low-cost single nucleotide polymorphic (SNP) markers to perform paternity analysis in large populations with higher accuracy than other traditional approaches. With this ability of NGS, paternity analysis has been used commonly in breeding programs for selections, maintaining pedigrees, and retaining paternal balance in poly crosses. Paternity analysis has been successfully employed in several crops including *Trifolium pretense* and *Trifolium repens Phleum pretense* and Eucalyptus *urophylla* (Crain et al. 2020).

GMO characterization

Molecular characterization of genetically modified organisms (GMOs) involves obtaining information about the location, copy number, and integrity of the exogenous gene in the plant genomes. It is crucial for obtaining desired traits, safety assessment, and detection of transgenic events. Various methods are utilized for molecular detection of GMOs including digital PCR, southern blotting, quantitative real-time PCR, and PCR-based genome walking. However, all these methods have one or other limitations related to time, effort, and efficiency. These limitations have been overcome with the development of next-generation sequencing leading to the efficient identification of copy number, insertion site, and flanking host DNA sequence of foreign DNA fragments in transgenic crops. NGS provides benefits such as a high degree of automation, accuracy, and standardization with good repeatability for molecular characterization of GMOs as compared to other crops (Wang et al. 2020). Different approaches to GMO characterization using NGS have been developed including whole genome sequencing, targeted sequencing, the combination of NGS with Site Finding PCR, and DNA walking (Debode et al. 2019).

Targeting induced local lesions in genomes (TILLING)

TILLING is a high-throughput technique to identify single nucleotide mutations in a specific region of a gene of interest with a powerful detection method that resulted from chemical-induced mutagenesis. This technology was first established in *Arabidopsis thaliana* and then has been employed in many diploid crops, including *Hordeum vulgare, Oryza sativa, Zea mays, Glycine max, Sorghum, Avena sativa, Solanum lycopersicum*, and *Brassica rapa*. Because of the complex genomes, TILLING is applied to a lesser extent in polyploid crops such as *Triticum aestivum, Arachis hypogaea*, and *Brassica napus* (Gao et al. 2020). Next-generation sequencing technologies have also integrated with the TILLING pipeline for mutation

discovery resulting in the improved efficiency of the mutation detection in polyploidy species such as *Glycine max* (palaeopolyploid), *Triticum aestivum* (allotetraploid), *Arachis hypogaea* (allotetraploid), and *crambe* (hexaploid) (Lakhssassi et al. 2021). Hence, TILLING by Sequencing is the more efficient method for mutation detection in a population, where amplicons from mutagenized plants are assembled and subsequently subjected to high-throughput sequencing (Fanelli et al. 2021). Initially, the TILLING method was developed as a discovery platform for functional genomics, but now it has become the substitute for transgenics for crop breeding (Kurowska et al. 2011). Advancement in next-generation sequencing may result in the development of an *in silico* resource for the assembly of all the mutations from a mutagenized population. The *in silico* TILLING approach allows researchers to immediately get results for the target gene based on available mutations (Till et al. 2018).

Conclusion

Genome sequencing is an integral part of biology that has been used in agriculture for regulating productivity and overcoming the challenges of food security. Recently, the development of NGS technologies has enhanced various strategies for scientific innovation. NGS is a high throughput sequencing that performs massive parallel sequencing of clonally amplified templates on a solid surface with reduced cost of new platform technologies. The advent of long-read sequencing platforms has simplified the sequencing of complex crop genomes having large genome sizes and high repetitive content. Different platforms are available that can be chosen according to the study. NGS has impacted agricultural genomics by introducing transcriptome sequencing, genome resequencing, targeted sequencing, and genotyping by sequencing methodologies. It is an essential genomics tool for many plant breeding and improvement programs. Whole-genome sequencing and genome resequencing provide information on structural variations present within or between the crop species, which is useful in uncovering the population diversity. Transcriptome sequencing and epigenetic analysis are valuable tools for studying the gene expression in a specific environmental condition impacting crop yield or responsible for a particular trait. Variants such as SNPs, CNVs, and markers that are associated with a particular phenotypic effect can be identified by sequencing. In addition, genotyping by sequencing enables the genotyping of multiple genomes in the absence of reference genomes. NGS also play role in various stages of genome editing using CRISPR technology. Whole-genome sequencing is utilized for analyzing CRISPR off-target effects and also CRISPR knockouts and other edits are verified by targeted sequencing. Hence, NGS technologies play a crucial role in the innovations in the agricultural field and provide the information and better understanding for regulation and enhancing the performance of the crop to meet the global demand.

Acknowledgements The authors extend their gratitude to the Director and Joint Director of ICAR—National Academy of Agricultural Research Management, Hyderabad, and the Department of Biotechnology (DBT) for providing the required infrastructure.

Author contributions All the authors contributed equally to this manuscript.

Funding This study was funded by the Director, ICAR-National Academy of Agricultural Research Management, Hyderabad.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Human or animals rights This manuscript does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants.

References

Abbas HMK, Zhou YY, Huang HX, Ul Qamar MT, Wang MM, Lu S, Xu C, Liu Z, Wu TQ, Zhong Yu Juan (2022) QTL mapping, whole genome resequencing, and marker-assisted selection provide basics of early flowering in pumpkin. Plant Breed 141(2):266–276. https://doi.org/10.1111/pbr.12996

Bayer PE, Petereit J, Danilevicz MF, Anderson R, Batley J, Edwards D (2021) The application of pangenomics and machine learning in genomic selection in plants. Plant Genome 14:e20112. https://doi.org/10.1002/tpg2.20112

Berhe M, Dossa K, You J et al (2021) Genome-wide association study and its applications in the non-model crop *Sesamum indicum*. BMC Plant Biol 21:283. https://doi.org/10.1186/s12870-021-03046-x

Bewicke-Copley F, Arjun Kumar E, Palladino G, Korfi K, Wang J (2019) Applications and analysis of targeted genomic sequencing in cancer studies. Comput Struct Biotechnol J 17:1348–1359. https://doi.org/10.1016/j.csbj.2019.10.004

Bhandari MS, Meena RK, Shamoon A, Saroj S, Pandey S (2020) First *de novo* genome specific development, characterization and validation of simple sequence repeat (SSR) markers in Genus *Salvadora*. Mol Biol Rep 47:6997–7008. https://doi.org/10.1007/s11033-020-05758-z

Bhat JA, Ali S, Salgotra RK, Mir ZA, Dutta S, Jadon V, Tyagi A, Mushtaq M, Jain N, Singh PK, Singh GP, Prabhu KV (2016) Genomic selection in the era of Next Generation sequencing for complex traits in plant breeding. Front Genet. https://doi.org/10.3389/fgene.2016.00221

Bryant DM, Johnson K, DiTommaso T, Tickle T, Couger MB, Payzin-Dogru D et al (2017) A tissue-mapped axolotl *de novo* transcriptome enables identification of limb regeneration factors. Cell Rep 18:762–776. https://doi.org/10.1016/j.celrep.2016.12.063

Budak H, Hussain B, Khan Z, Ozturk NZ, Ullah N (2015) From genetics to functional genomics: improvement in drought signaling and tolerance in wheat. Front Plant Sci 6:1012

Bus A, Hecht J, Huettel B, Reinhardt R, Stich B (2012) High-throughput polymorphism detection and genotyping in *Brassica napus* using next-generation RAD sequencing. BMC Genomics 13:281. https://doi.org/10.1186/1471-2164-13-281

Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X et al (2014) Plant genetics. Early allopolyploid evolution in the post-neolithic *Brassica napus* oilseed genome. Science 345:950–953. https://doi.org/10.1126/science.1253435

Chen LY, VanBuren R, Paris M et al (2019) The *Bracteatus* pineapple genome and domestication of clonally propagated crops. Nat Genet 51:1549–1558. https://doi.org/10.1038/s41588-019-0506-8

Chidinma NB, Oluranti BO (2022) Metagenomics: a tool for exploring key microbiome with the potentials for improving sustainable agriculture. Front Sustain Food Syst. https://doi.org/10.3389/fsufs.2022.886987

Choulet F, Wicker T, Rustenholz C, Paux E, Salse J, Leroy P et al (2010) Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. Plant Cell 22:1686–1701. https://doi.org/10.1105/tpc.110.074187

Christoph B (2016) Third generation sequencing: technology and its potential impact on evolutionary biodiversity research. Syst Biodivers 14(1):1–8. https://doi.org/10.1080/14772000.2015.1099575

Chwialkowska K, Nowakowska U, Mroziewicz A, Szarejko I, Kwasniewski M (2016) Water-deficiency conditions differently modulate the methylome of roots and leaves in barley (*Hordeum vulgare* L.). J Exp Bot 67:1109–1121

Cooper EA, Brenton ZW, Flinn BS et al (2019) A new reference genome for Sorghum bicolor reveals high levels of sequence similarity between sweet and grain genotypes: implications for the genetics of sugar metabolism. BMC Genomics 20:420. https://doi.org/10.1186/s12864-019-5734-x

Crain J, Larson S, Dorn K et al (2020) Sequenced-based paternity analysis to improve breeding and identify self-incompatibility loci in intermediate wheatgrass (*Thinopyrum intermedium*). Theor Appl Genet 133:3217–3233. https://doi.org/10.1007/s00122-020-03666-1

Das S, Upadhyaya HD, Bajaj D, Kujur A, Badoni S, Kumar V, Tripathi S, Gowda CL, Sharma S, Singh S, Tyagi AK (2015) Deploying QTL-seq for rapid delineation of a potential candidate gene underlying major trait-associated QTL in chickpea. DNA Res 22(3):193–203

Debode F, Hulin J, Charloteaux B, Coppieters W, Hanikenne M, Karim L, Berben G (2019) Detection and identification of transgenic events by next generation sequencing combined with enrichment technologies. Sci Rep 9(1):15595. https://doi.org/10.1038/s41598-019-51668-x

Dekkers JCM (2007) Prediction of response to marker-assisted and genomic selection using selection index theory. J Anim Breed Genet 124:331–41. https://doi.org/10.1111/j.1439-0388.2007.00701.x

Delphine F, Ute B, Peter L (2012) Plant genome sequencing: models for developing synteny maps and association mapping. In: Altman A, Hasegawa PM (eds) Plant biotechnology and agriculture. Academic Press, Cambridge, pp 83–97. https://doi.org/10.1016/B978-0-12-381466-1.00006-7

Deokar A, Sagi M, Daba K, Tar'an B (2019) QTL sequencing strategy to map genomic regions associated with resistance to ascochyta blight in chickpea. Plant Biotechnol J 17(1):275–288. https://doi.org/10.1111/pbi.12964

Deschamps S, Llaca V, May GD (2012) Genotyping-by-sequencing in plants. Biology 1(3):460–483

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K et al (2011) A robust, simple genotyping by sequencing (GBS) approach for high diversity species. PLoS One 6:e19379

Fanelli V, Ngo KJ, Thompson VL et al (2021) A TILLING by sequencing approach to identify induced mutations in sunflower genes. Sci Rep 11:9885. https://doi.org/10.1038/s41598-021-89237-w

Forestan C, Farinati S, Rouster J, Lassagne H, Lauria M, Dal Ferro N, Varotto S (2018) Control of maize vegetative and reproductive development, fertility, and rRNAs silencing by *HISTONE DEACETYLASE 108*. Genetics 208:1443–1466

Gao L, Gonda I, Sun H, Ma Q, Bao K, Tieman DM, Burzynski-Chang EA, Fish TL, Stromberg KA, Sacks GL et al (2019) The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. Nat Genet 51:1044–1051

Gao YL, Yao XF, Li WZ, Song ZB, Wang BW, Wu YP, Shi JL, Liu GS, Li YP, Liu CM (2020) An efficient TILLING platform for cultivated tobacco. J Integr Plant Biol 62:165–180. https://doi.org/10.1111/jipb.12784

Gedil M et al (2016) Perspectives on the application of next-generation sequencing to the improvement of Africa's

staple food crops. In: Kulski J et al (eds) Next generation sequencing – advances. Applications and challenges. InTechOpen, London, pp 287–321

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

Golicz AA, Bayer PE, Barker GC, Edger PP, Kim H, Martinez PA, Chan CK, Severn-Ellis A, McCombie WR, Parkin IA, Paterson AH, Pires JC, Sharpe AG, Tang H, Teakle GR, Town CD, Batley J, Edwards D (2016) The pangenome of an agronomically important crop plant *Brassica oleracea*. Nat Commun 7:13390. https://doi.org/10.1038/ncomms13390

Guo LL, Guo DL, Yin WL, Hou X (2018) Tolerance strategies revealed in tree peony (*Paeonia suffruticosa*, Paeoniaceae) ecotypes differentially adapted to desiccation. Appl Plant Sci 6:e01191. https://doi.org/10.1002/aps3.1191

Gupta AK, Gupta UD (2014) Next generation sequencing and its applications. In: Singh A, AS Verma (eds) Animal biotechnology. Academic Press, Cambridge, pp 345–367. https://doi.org/10.1016/B978-0-12-416002-6.00019-5

He J, Zhao X, Laroche A, Lu Z-X, Liu H, Li Z (2014) Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Front Plant Sci 5:484

Hou DY, Shi LC, Yang MM, Li J, Zhou S, Zhang HX et al (2018) *De novo* transcriptomic analysis of leaf and fruit tissue of *Cornus officinalis* using Illumina platform. PLoS One 13:e0192610. https://doi.org/10.1371/journal.pone.0192610

Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A et al (2009) High-throughput genotyping by whole-genome resequencing. Genome Res 19(6):1068–1076. https://doi.org/10.1101/gr.089516.108

Iquebal MA, Jagannadham J, Jaiswal S, Prabha R, Rai A, Kumar D (2022) Potential use of microbial community genomes in various dimensions of agriculture productivity and its management: a review. Front Microbiol 13:708335. https://doi.org/10.3389/fmicb.2022.708335

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467

Jarvis DE, Ho YS, Lightfoot DJ, Schmöckel SM, Li B, Borm TJ et al (2017) The genome of *Chenopodium quinoa*. Nature 542:307–312. https://doi.org/10.1038/nature21370

Jayakodi M, Padmarasu S, Haberer G et al (2020) The barley pan-genome reveals the hidden legacy of mutation breeding. Nature 588:284–289. https://doi.org/10.1038/s41586-020-2947-8

Jegadeesan S, Raizada A, Dhanasekar P, Suprasanna P (2021) Draft genome sequence of the pulse crop blackgram [*Vigna mungo* (L.) Hepper] reveals potential R-genes. Sci Rep 11(1):11247. https://doi.org/10.1038/s41598-021-90683-9

Jiang W, Zhou H, Bi H, Fromm M, Yang B, Weeks DP (2013) Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. Nucl Acids Res 41:1–12. https://doi.org/10.1093/nar/gkt780

Jiang Z, Wang H, Michal JJ, Zhou X, Liu B, Woods LCS, Fuchs RA (2016) Genome wide sampling sequencing for SNP genotyping: methods, challenges and future development. Int J Biol Sci 12(1):100–108. https://doi.org/10.7150/ijbs.13498

Kapazoglou A, Ganopoulos I, Tani E, Tsaftaris A (2018) Chapter nine - epigenetics, epigenomics and crop improvement. In: Marcel K (ed) Advances in botanical research, vol 86. Academic Press, Cambridge, pp 287–324. https://doi.org/10.1016/bs.abr.2017.11.007

Kaushik P, Sandhu OS, Brar NS, Kumar V, Malhi GS, Kesh H, Saini I (2020) Soil metagenomics: prospects and challenges. In: Radhakrishnan R (ed) Mycorrhizal fungi - utilization in agriculture and forestry. IntechOpen, London. https://doi.org/10.5772/intechopen.93306

Keller B, Ariza-Suarez D, de la Hoz J, Aparicio JS, Portilla-Benavides AE, Buendia HF, Mayor VM, Studer B, Raatz B (2020) Genomic prediction of agronomic traits in common bean (*Phaseolus vulgaris* L.) under environmental stress. Front Plant Sci. https://doi.org/10.3389/fpls.2020.01001

Kishor DS, Seo J, Hyoun CJ, Hee-Jong K (2020) Evaluation of whole-genome sequence, genetic diversity, and agronomic traits of basmati rice (*Oryza sativa* L.). Front Genet. https://doi.org/10.3389/fgene.00086

Kulski JK (2016) Next-generation sequencing - an overview of the history, Tools, and "Omic" applications. In: Kulski JK (ed) Next generation sequencing - advances, applications, and Challenges. IntechOpen, London. https://doi.org/10.5772/61964

Kumar P, Choudhary M, Jat BS, Kumar B, Singh V, Kumar V, Singla D, Rakshit S (2021) Skim sequencing: an advanced NGS technology for crop improvement. J Genet 100:38

Kurowska M, Daszkowska-Golec A, Gruszka D et al (2011) TILLING: a shortcut in functional genomics. J Appl Genet 52(4):371–390. https://doi.org/10.1007/s13353-011-0061-1

Kwok KTT, Nieuwenhuijse DF, Phan MVT, Koopmans MPG (2020) Virus metagenomics in farm animals: a systematic review. Viruses 12:107. https://doi.org/10.3390/v12010107

Lai J, Li R, Xu X, Jin W, Xu M, Zhao H et al (2010) Genome wide patterns of genetic variation among elite maize inbred lines. Nat Genet 42(11):1027–1030. https://doi.org/10.1038/ng.684

Lakhssassi N, Zhou Z, Cullen MA, Badad O, El Baze A, Chetto O, Embaby MG, Knizia D, Liu S, Neves LG, Meksem K (2021) TILLING-by-sequencing to Decipher Oil Biosynthesis pathway in soybeans: a new and effective platform for high-throughput gene functional analysis. Int J Mol Sci 22(8):4219. https://doi.org/10.3390/ijms22084219

Lam H-M, Xu X, Liu X, Chen W, Yang G, Wong F-L et al (2010) Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nat Genet 42(12):1053–1059

Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ et al (2015) Genome sequence of cultivated upland cotton (Gossypium hirsutum TM-1) provides insights into genome

evolution. Nat Biotechnol 33:524–530. https://doi.org/10.1038/nbt.3208
Li H, Fu Y, Sun H, Zhang Y, Lan X (2017) Transcriptomic analyses reveal biosynthetic genes related to rosmarinic acid in *Dracocephalum tanguticum*. Sci Rep 7:74. https://doi.org/10.1038/s41598-017-00078-y
Li Q, Hermanson PJ, Springer NM (2018) Detection of DNA methylation by whole-genome bisulfite sequencing. Methods Mol Biol 1676:185–196. https://doi.org/10.1007/978-1-4939-7315-6_11
Li M-Y, Feng K, Hou X-L, Jiang Q, Xu Z-S, Wang G-L, Liu J-X, Wang F, Xiong A-S (2020) The genome sequence of celery (*Apium graveolens* L.), important leaf vegetable crop rich in apigenin in the Apiaceae family. Hortic Res 7:9. https://doi.org/10.1038/s41438-019-0235-2
Li G, Wang L, Yang J et al (2021) A high-quality genome assembly highlights rye genomic characteristics and agronomically important genes. Nat Genet 53:574–584. https://doi.org/10.1038/s41588-021-00808-z
Liao X, Li M, Zou Y, Wu F-X, Yi P, Wang J (2019) Current challenges and solutions of *De novo* assembly. Quant Biol 7(2):90–109
Liu HJ, Yan J (2019) Crop genome-wide association study: a harvest of biological relevance. Plant J 97:8–18. https://doi.org/10.1111/tpj.14139
Liu Y, Du H, Li P, Shen Y, Peng H, Liu S, Zhou G-A, Zhang H, Liu Z, Shi M, Huang X, Li Y, Zhang M, Wang Z, Zhu B, Han B, Liang C, Tian Z (2020) Pan-Genome of Wild and cultivated soybeans. Cell 182(1):162–176
Loke KK, Rahnamaie-Tajadod R, Yeoh CC, Goh HH, Noor NM (2017) Transcriptome analysis of *Polygonum minus* reveals candidate genes involved in important secondary metabolic pathways of phenylpropanoids and flavonoids. Peer J 5:e2938. https://doi.org/10.7717/peerj.2938
Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD et al (2013) Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. PLoS Genet 9:e1003215. https://doi.org/10.1371/journal.pgen.1003215
Lu H, Lin T, Klein J et al (2014) QTL-seq identifies an early flowering QTL located near flowering locus T in cucumber. Theor Appl Genet 127:1491–1499
Lu K, Wei L, Li X et al (2019) Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement. Nat Commun 10(1):1154. https://doi.org/10.1038/s41467-019-09134-9
Luo MC, Gu YQ, Puiu D et al (2017) Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. Nature 551(7681):498–502. https://doi.org/10.1038/nature24486
Mehdi K, Jean-François G, Mourad E (2017) Generations of sequencing technologies: from first to next generation. Biol Med 9:3. https://doi.org/10.4172/0974-8369.1000395
Michael TP, VanBuren R (2015) Progress, challenges and the future of crop genomes. Curr Opin Plant Biol 24:71–81
Muhammad Afzal SS, Alghamdi, Hussein H, Migdadi MA, Khan, Nurmansyah SB, Mirza, Ehab El-Harty (2020) Legume genomics and transcriptomics: from classic breeding to modern technologies. Saudi J Biol Sci 27(1):543–555. https://doi.org/10.1016/j.sjbs.2019.11.018
Murat F, Van de Peer Y, Salse J (2012) Decoding plant and animal genome plasticity from differential paleo-evolutionary patterns and processes. Genome Biol Evol 4:917–928
Mushtaq M, Dar AA, Basu U, Bhat BA, Mir RA, Vats S, Dar MS, Tyagi A, Ali S, Bansal M, Rai GK, Wani SH (2021) Integrating CRISPR-Cas and next generation sequencing in plant virology. Front Genet 12:735489. https://doi.org/10.3389/fgene.2021.735489
Natarajan P, Parani M (2011) *De novo* assembly and transcriptome analysis of five major tissues of *Jatropha curcas* L. using GS FLX titanium platform of 454 pyrosequencing. BMC Genom 12:191. https://doi.org/10.1186/1471-2164-12-191
Nguyen KL, Grondin A, Courtois B (2019) Next-generation sequencing accelerates crop gene discovery. Trends Plant Sci 24(3):263–274. https://doi.org/10.1016/j.tplants.2018.11.008
Pankin A, Altmüller J, Becker C, von Korff M (2018) Targeted resequencing reveals genomic signatures of barley domestication. New Phytol 218(3):1247–1259. https://doi.org/10.1111/nph.15077
Pereira R, Oliveira J, Sousa M (2020) Bioinformatics and computational tools for next-generation sequencing analysis in clinical genetics. J Clin Med 9(1):132. https://doi.org/10.3390/jcm9010132
Poland JA, T.W (2012) Rife, genotyping-by-sequencing for plant breeding and genetics. Plant Genome 5:92–102
Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7(2):e32253
Pragati C, Muniya R, Sangwan RS, Ravinder K, Anil K, Vinod C (2018) *De novo* sequencing, assembly and characterisation of *Aloevera* transcriptome and analysis of expression profiles of genes related to saponin and anthraquinone metabolism. BMC Genomics 19:427. https://doi.org/10.1186/s12864-018-4819-2
Rao G, Zhang J, Liu X et al (2021) *De novo* assembly of a new Olea europaea genome accession using nanopore sequencing. Hortic Res 8:64. https://doi.org/10.1038/s41438-021-00498-y
Sabale SN, Suryawanshi PP, Krishnaraj P (2019) Soil metagenomics: concepts and applications. In: Hozzein WN (ed) Metagenomics-basics, methods and applications. IntechOpen, London
Scheben A, Severn-Ellis AA, Patel D et al (2020) Linkage mapping and QTL analysis of flowering time using ddRAD sequencing with genotype error correction in *Brassica napus*. BMC Plant Biol 20:546. https://doi.org/10.1186/s12870-020-02756-y
Schmidt MH, Vogel A, Denton AK, Istace B, Wormit A, Geest H et al (2017) De novo assembly of a new *Solanum pennellii* accession using nanopore sequencing. Plant Cell 29:2336–2348. https://doi.org/10.1105/tpc.17.00521
Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Schreiber M, Stein N, Mascher M (2018) Genomic approaches for studying crop evolution. Genome Biol 19:140. https://doi.org/10.1186/s13059-018-1528-8

Sharma TR, Devanna BN, Kiran K, Singh PK, Arora K, Jain P, Tiwari IM, Dubey H, Saklani BK, Kumari M, Singh J, Jaswal R, Kapoor R, Pawar DV, Sinha S, Bisht DS, Solanke AU, Mondal (2018) TK status and prospects of next-generation sequencing technologies in crop plants. Curr Issues Mol Biol 27:1–36. https://doi.org/10.21775/cimb.027.001

Shimatani Z, Kashojiya S, Takayama M, Terada R, Arazoe T, Ishii H et al (2017) Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. Nat Biotechnol 35:441–443. https://doi.org/10.1038/nbt.3833

Singh P, Singh G, Bhandawat A, Singh G, Parmar R, Seth R et al (2017) Spatial transcriptome analysis provides insights of key gene(s) involved in steroidal saponin biosynthesis in medicinally important herb *Trillium govanianum*. Sci Rep 7:45295. https://doi.org/10.1038/srep45295

Slatko BE, Gardner AF, Ausubel FM (2018) Overview of next-generation sequencing technologies. Curr Protoc Mol Biol 122(1):e59. https://doi.org/10.1002/cpmb.59

Sonah H, Bastien M, Iquira E, Tardivel A, Legare G, Boyle B et al (2013) An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. PLoS One 8:e54603. https://doi.org/10.1371/journal.pone.0054603

Soyk S, Müller NA, Park SJ, Schmalenbach I, Jiang K, Hayama R et al (2017) Variation in the flowering gene SELF PRUNING 5G promotes day neutrality and early yield in tomato. Nat Genet 49:162–168. https://doi.org/10.1038/ng.3733

Steuernagel B, Taudien S, Gundlach H et al (2009) *De novo* 454 sequencing of barcoded BAC pools for comprehensive gene survey and genome analysis in the complex genome of barley. BMC Genomics 10:547. https://doi.org/10.1186/1471-2164-10-547

Sun X, Hu Z, Chen R, Jiang Q, Song G, Zhang H et al (2015) Targeted mutagenesis in soybean using the CRISPR-Cas9 system. Sci Rep 5:10342. https://doi.org/10.1038/srep10342

Svitashev S, Schwartz C, Lenderts B, Young JK, Mark Cigan A (2016) Genome editing in maize directed by CRISPR-Cas9 ribonucleoprotein complexes. Nat Commun 7:1–7. https://doi.org/10.1038/ncomms13274

Taishan Hu, Chitnis N, Monos D, Anh Dinh (2021) Next-generation sequencing technologies: an overview. Hum Immunol 82(11):801–811. https://doi.org/10.1016/j.humimm.2021.02.012

Takagi H, Abe A, Yoshida K et al (2013) QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J 74:174–183

Thomsen PF, Willerslev E (2015) Environmental DNA – an emerging tool in conservation for monitoring past and present biodiversity. Biol Conserv 183:4–18. https://doi.org/10.1016/j.biocon.2014.11.019

Till BJ, Datta S, Jankowicz-Cieslak J (2018) TILLING: the next generation. Adv Biochem Eng Biotechnol 164:139–160. https://doi.org/10.1007/10_2017_54

Truong HT, Ramos AM, Yalcin F et al (2012) Sequence-based genotyping for marker discovery and co-dominant scoring in germplasm and populations. PLoS ONE 7(5):e37565. https://doi.org/10.1371/journal.pone.0037565

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A et al (2006) The genome of black cottonwood, *Populus trichocarpa*. Science 313:1596–1604

Udall JA, Long E, Hanson C et al (2019) *De novo* Genome sequence assemblies of *Gossypium raimondii* and *Gossypium turneri*. G3-Genes Genom Genet 9(10):3079–3085. https://doi.org/10.1534/g3.119.400392

Unamba Chibuikem IN, Nag A, Sharma RK (2015) Next generation sequencing technologies: the doorway to the unexplored genomics of non-model plants. Front Plant Sci. https://doi.org/10.3389/fpls.2015.01074

Van Poecke RM, Maccaferri M, Tang J, Truong HT, Janssen A et al (2013) Sequence-based SNP genotyping in durum wheat. Plant Biotechnol J 11(7):809–817

VanBuren R, Bryant D, Edger PP, Tang H, Burgess D, Challabathula D, Spittle K, Hall R, Gu J, Lyons E et al (2015) Single-molecule sequencing of the desiccation-tolerant grass *Oropetium thomaeum*. Nature 527:508–511

Varotto S, Tani E, Abraham E, Krugman T, Kapazoglou A, Melzer R, Radanovićv A, Miladinović D (2020) Epigenetics: possible applications in climate-smart crop breeding. J Exp Bot 71(17):5223–5236. https://doi.org/10.1093/jxb/eraa188

Varshney RK, Graner A, Sorrells ME et al (2005) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23(1):48–55

Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, Pruss D, Pindo M, Fitzgerald LM, Vezzulli S, Reid J et al (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS One 2:e1326

Visser RGF, Bachem CWB, de Boer JM et al (2009) Sequencing the Potato Genome: outline and first results to come from the elucidation of the sequence of the World's third most important food crop. Am J Potato Res 86:417. https://doi.org/10.1007/s12230-009-9097-8

Walkowiak S, Gao L, Monat C, Haberer G, Kassa MT, Brinton J, Ramirez-Gonzalez RH, Kolodziej MC, Delorean E, Thambugala D, Klymiuk V, Byrns B, Gundlach H, Bandi V, Siri JN, Nilsen K, Aquino C, Himmelbach A, Copetti D, Ban T, Pozniak CJ (2020) Multiple wheat genomes reveal global variation in modern breeding. Nature 588(7837):277–283. https://doi.org/10.1038/s41586-020-2961-x

Wang Z et al (2010) RNA-seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63

Wang X, Gao L, Jiao C et al (2020) Genome of *Solanum pimpinellifolium* provides insights into structural variants during tomato breeding. Nat Commun 11:5817. https://doi.org/10.1038/s41467-020-19682-0

Wang X, Jiao Y, Ma S, Yang J, Wang Z (2020) Whole-genome sequencing: an effective strategy for insertion

information analysis of Foreign genes in transgenic plants. Front Plant Sci 11:1664–462X. https://doi.org/10.3389/fpls.2020.573871

Wei T, Van Treuren R, Liu X et al (2021) Whole-genome resequencing of 445 *Lactuca* accessions reveals the domestication history of cultivated lettuce. Nat Genet 53:752–760. https://doi.org/10.1038/s41588-021-00831-0

Wenqin W, Xuan HC, Mihai M, Jianhong X, Wenwei X (2017) The promise of agriculture genomics. Int J Genomics. https://doi.org/10.1155/2017/9743749

Wu X, Cui Z, Peng J et al (2022) Genome-resolved metagenomics identifies the particular genetic traits of phosphate-solubilizing bacteria in agricultural soil. ISME Commun 2:17. https://doi.org/10.1038/s43705-022-00100-z

Xu X, Bai G (2015) Whole-genome resequencing: changing the paradigms of SNP detection, molecular mapping and gene discovery. Mol Breed 35:33. https://doi.org/10.1007/s11032-015-0240-6

Xu WR, Li RM, Zhang NB, Ma F, Jiao YT, Wang ZP (2014) Transcriptome profiling of *Vitis amurensis*, an extremely cold-tolerant chinese wild *Vitis* species, reveals candidate genes and events that potentially connected to cold stress. Plant Mol Biol 86:527–541. https://doi.org/10.1007/s11103-014-0245-2

Yaman M, Uzun A (2020) Evaluation of superior hybrid individuals with intra and interspecific hybridization breeding in apricot. Int J Fruit Sci 20(sup3):S2045–S2055

Yaman M, Uzun A (2021) Morphological and molecular identification of hybrid individuals obtained by interspecies hybridization (Prunus armeniaca× Prunus salicina). Int J Agric Nat Sci 14(1):7–15

Yang H, Tao Y, Zheng Z, Li C, Sweetingham MW, Howieson JG (2012) Application of next-generation sequencing for rapid marker development in molecular plant breeding: a case study on anthracnose disease resistance in *Lupinus angustifolius* L. BMC Genom 13:318. https://doi.org/10.1186/1471-2164-13-318

Yang N, Liu J, Gao Q et al (2019) Genome assembly of a tropical maize inbred line provides insights into structural variation and crop improvement. Nat Genet 51:1052–1059. https://doi.org/10.1038/s41588-019-0427-6

Yang Y, Dong S, Miao H, Liu X, Dai Z, Li X, Gu X, Zhang S (2022) Genome-wide Association Studies reveal candidate genes related to Stem Diameter in Cucumber (*Cucumis sativus* L.). Genes (Basel) 13(6):1095. doi: https://doi.org/10.3390/genes13061095

Zhang Y, Liang Z, Zong Y, Wang Y, Liu J, Chen K et al (2016) Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. Nat Commun 7:1–8. https://doi.org/10.1038/ncomms12617

Zhang Y, Massel K, Godwin ID et al (2018) Applications and potential of genome editing in crop improvement. Genome Biol 19:210. https://doi.org/10.1186/s13059-018-1586-y

Zhang J, Li W, Lv P, Yang X, Xu W, Ni X, Feng H, Zhao G, Pu M, Zhou S, Zhao Z, Li L, Huang Y, Li JFanYan, Zhu J, Liu G (2021) Whole-genome resequencing and transcriptome analysis provide insights on aphid-resistant quantitative trait loci/genes in Sorghum bicolor. Plant Breed 140(4):618–629. https://doi.org/10.1111/pbr.12946

Zhao L, Xie L, Zhang Q, Ouyang W, Deng L, Guan P, Ma M, Li Y, Zhang Y, Xiao Q, Zhang J, Li H, Wang S, Man J, Cao Z, Zhang Q, Zhang Q, Li G, Li X (2020) Integrative analysis of reference epigenomes in 20 rice varieties. Nat Commun 11(1):2658. https://doi.org/10.1038/s41467-020-16457-5

Zhuang W, Chen H, Yang M et al (2019) The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. Nat Genet 51:865–876. https://doi.org/10.1038/s41588-019-0402-2