

Review

# The Progression in Developing Genomic Resources for Crop Improvement

Pradeep Ruperao <sup>1,\*</sup>, Parimalan Rangan <sup>2</sup>, Trushar Shah <sup>3</sup>, Vivek Thakur <sup>4</sup>, Sanjay Kalia <sup>5</sup>, Sean Mayes <sup>1,\*</sup>  
and Abhishek Rathore <sup>6,\*</sup>

- <sup>1</sup> Center of Excellence in Genomics and Systems Biology, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad 502324, India
  - <sup>2</sup> ICAR-National Bureau of Plant Genetic Resources, PUSA Campus, New Delhi 110012, India; r.parimalan@icar.gov.in
  - <sup>3</sup> International Institute of Tropical Agriculture (IITA), Nairobi 30709-00100, Kenya; tm.shah@cgiar.org
  - <sup>4</sup> Department of Systems & Computational Biology, School of Life Sciences, University of Hyderabad, Hyderabad 500046, India; vivek22@uohyd.ac.in
  - <sup>5</sup> Department of Biotechnology, Ministry of Science and Technology, Government of India, New Delhi 110003, India; sanjay.kalia@nic.in
  - <sup>6</sup> Excellence in Breeding, International Maize and Wheat Improvement Center (CIMMYT), Hyderabad 502324, India
- \* Correspondence: Pradeep.ruperao@icrisat.org (P.R.); sean.mayes@icrisat.org (S.M.); abhishek.rathore@cgiar.org (A.R.)

**Abstract:** Sequencing technologies have rapidly evolved over the past two decades, and new technologies are being continually developed and commercialized. The emerging sequencing technologies target generating more data with fewer inputs and at lower costs. This has also translated to an increase in the number and type of corresponding applications in genomics besides enhanced computational capacities (both hardware and software). Alongside the evolving DNA sequencing landscape, bioinformatics research teams have also evolved to accommodate the increasingly demanding techniques used to combine and interpret data, leading to many researchers moving from the lab to the computer. The rich history of DNA sequencing has paved the way for new insights and the development of new analysis methods. Understanding and learning from past technologies can help with the progress of future applications. This review focuses on the evolution of sequencing technologies, their significant enabling role in generating plant genome assemblies and downstream applications, and the parallel development of bioinformatics tools and skills, filling the gap in data analysis techniques.

**Keywords:** sequencing technologies; plant genomes; pan-genomes; assemblies; bioinformatics tools; databases; big data; artificial intelligence; machine learning



**Citation:** Ruperao, P.; Rangan, P.; Shah, T.; Thakur, V.; Kalia, S.; Mayes, S.; Rathore, A. The Progression in Developing Genomic Resources for Crop Improvement. *Life* **2023**, *13*, 1668. <https://doi.org/10.3390/life13081668>

Academic Editor: Kousuke Hanada

Received: 15 June 2023  
Revised: 21 July 2023  
Accepted: 25 July 2023  
Published: 31 July 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

With more than 40 years of remarkable DNA sequencing improvements, today, the development of cost-reducing and higher throughput sequencing technologies, along with relevant bioinformatics tools, have made it possible to produce high-quality genome assemblies in a much-reduced timeline, which has subsequently led to the mapping of the genetic variations in thousands of individuals, providing genetic insights into population histories and domestication events. The multinational and multi-institutional consortium the Earth BioGenome Project (EBP) aims to unify the phylogenetic networks across all eukaryotic life derived from their complete de novo genomes [1,2]. This illustrates how far the advancement and standardization of genome data generation, assemblies, storage, retrieval, and analysis have developed, with more expected and required with the generation of massive genomic data from species bridging the phylogenetic gaps between currently sequenced genomes.

Complete reference genome assemblies of the entire plant kingdom will open new scientific views on the evolution and speciation events on earth and genetic control of plant traits, both at intra- and inter-species levels. They will also enhance the understanding of how plants function in ecosystems, lead to the discovery of natural botanical compounds for human medicine, and will aid an increase in food production to curb global hunger while respecting planetary boundaries and adapting to climate change.

Here, we provide an overview of the improvements in sequencing technologies, the development of the associated bioinformatics tools, and advancements in plant genomics. We also outline the progress achieved in assembling plant genomes, sequence technologies, and assemblers used to contribute towards crop improvement.

## 2. Genome Sequencing Milestones

Over 40 years of consistent development of reliable sequencing technology emerging to make considerable progress in accuracy, cost, and reduced sequencing time has been improved. From first-generation to third-generation sequencing, the combined technologies developed have significantly increased the read length, improved quality, and provided massive increases in throughput with significant cost reductions. However, currently, second-generation (also called next-generation) sequencing technology dominates.

### 2.1. First-Generation Sequencing (FGS)

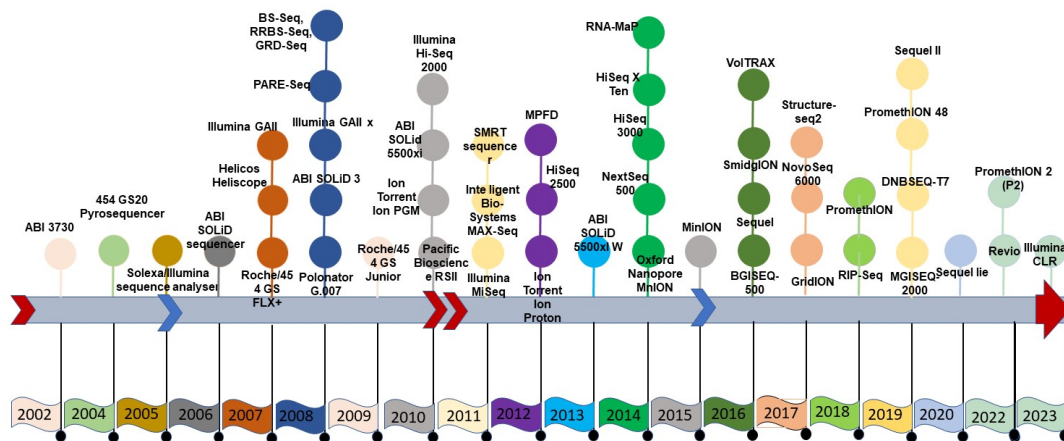
It is during this phase that the sequencing process advanced as a technology to help understand the genetic basis behind the phenotype. This first-generation sequencing technology is based on the dideoxynucleotide chain termination method developed by Sanger and Coulson in 1975, commonly known as the Sanger method [3] and nucleobase-specific partial chemical modification of DNA in Maxam–Gilbert sequencing [4]. The first genome sequence for Phage X174 was generated in 1977 using a variant of this method. The automated Sanger method (through capillary electrophoresis in 1980) was an essential improvement and aided the completion of the Human Genome Project in 2001. The merit of this technology was that it produced a read length of around 1 kb with 99.999% accuracy, but the drawback was its high cost, short run length, and low throughput.

### 2.2. Second-Generation Sequencing (SGS)/Next-Generation Sequencing (NGS)

While the Sanger method was continuously popular, particularly for accurate sequencing of specific sequences, such as genes, many other sequencing technologies emerged at around the same time, such as (i) the pyrophosphate sequencing used by Roche for the 454 sequencing platform (the first major successful commercial SGS technology), (ii) the ligation enzyme method used for the SOLID technique by ABI sequencing company, (iii) single-molecule sequencing with HeliScope from Helicos Biosciences, and (iv) DNA colony sequencing technology from Illumina.

SGS sequencing was conducted in either a stepwise iterative process or in a real-time manner, producing a combination of qualitative and quantitative sequence information, which was not possible with FGS data. The second generation of sequencing technology was symbolized by several approaches, all fundamentally based on parallel data production with individual sequences identified by position on a flow cell or microarray. Roche's 454, Illumina's Solexa, Hiseq technology, and ABI's Solid technologies not only reduced sequencing costs but also increased the speed of sequencing [5]. The thirteen-year duration of the human genome project using Sanger sequencing would now take just one week with SGS technologies to generate the raw sequence data—although assembly remained a significant computational problem. Sequencing throughput has increased with SGS technologies, but the read length is often much shorter than in the first generation. The specific technologies of SGS include the Ion Torrent technology that directly produces digital nucleotide sequence information on a semiconductor chip [6]. It is possible to generate such sequence information with several versions of the Ion Torrent platform, such as the Ion Personal Genome Machine (PGM) System, Ion Proton System and Ion S5 system,

and ION S5 XL systems. The Roche/454 Life Sciences introduced several sequencers in the form of GS, GS 20 run, and GS FLX titanium. Similarly, Illumina sequencing supports a variety of protocols with varying levels of throughput, including MiniSeq [7], MiSeq [8], NextSeq (<https://doi.org/10.48550/arXiv.1711.11004>), HiSeq [9], and NovaSeq models (Figure 1) (Table 1).



**Figure 1.** Timeline of the advancement of sequencing technology.

### 2.3. Third-Generation Sequencing

Third-generation sequencing (TGS) technologies have focused on increasing the read length while maintaining the sequencing throughput. The single-molecule real-time sequencing-by-synthesis (SMRT) technology from Pacific Biosciences introduced read lengths of single DNA fragments exceeding 10 Kb, and long sequences are particularly useful for *de novo* genome assemblies, especially where genomes are large or contain repetitive DNA, as is the case with many plants [10]. Pacific Biosciences (PacBio) has commercialized two sequencing systems, RSII model and Sequel II, producing high-fidelity (HiFi) reads with more than 99% accuracy [11], and Revio is an advanced platform to generate HiFi reads at a higher throughput (15X). SMRT now enables the generation of very long reads of lengths over 30 Kb to 50 Kb.

With continual upgrades in sequencing chemistry and technology, it is possible to generate longer reads of over 100 Kb in length. Nanopore is a technology that takes a different approach to sequencing via synthesis adopted by PacBio. This technology (proposed in 1990 and commercialized by Oxford Nanopore Technologies, ONT) decodes the DNA molecule by detecting electrical fluctuations as a nucleic acid molecule passes through a small diameter biological “pore”. Continuous sequences from single molecules up to 500 Kb have been reported, although generally, a single molecule sequence is likely to average around 20 Kb [12]. By having multiple parallel pores and very rapid processing, it is possible to obtain hundreds of gigabases of nucleotide sequences at a low cost. Early iterations of the technology had relatively poor accuracy, and while the accuracy has improved (partly through the development of software specifically for interpreting nanopore signals), a nanopore is often corrected in practice by using highly accurate short Illumina reads before being used for genome assembly. Nanopore technology comes in different versions, including MinION, benchtop GridION, VolTRAX, and high-throughput PromethION [13]. Next, 10× Genomics is another long-read sequence technology ([www.10xgenomics.com](http://www.10xgenomics.com)) integrated with GemCode technology supplied with the Supernova2 genome assembler. This technology was specifically designed for diploid and low-complex genomes, such as *Corylus avellana*, where its implementation produced a chromosome-level genome assembly [14].

Similarly, optical mapping and Dovetail Hi-C technologies are useful to complete the ordering of various DNA contigs in a genome by creating a visual physical map along large DNA molecules, which assist in correlating a DNA sequence with a physical location [15,16].

This technology was further improved by using nanofluidic methods, and image capture and processing have further improved optical mapping [17,18]. The Bionano technology was commercially developed and made available to process samples through Bionano Genomics (San Diego, CA, USA) (<https://bionanogenomics.com/products/>) and OpGen (<http://www.opgen.com/about-us/opgen-overview/>).

**Table 1.** Sequencing technologies and applications in the SGS era.

Sequencer/Technology	Applications	Reference
ChIP-Seq	Protein-DNA interactions (using chromatin immunoprecipitation)	[19]
DNA-Seq	A genome-derived sequence	[20]
RIP-Seq, CLIP-Seq, HITS-CLIP	Protein-RNA interactions	[21]
RNA-Seq	RNA (that is, the transcriptome)	[22]
RAD-seq	Restriction site-associated DNA sequencing	[23]
TRAP	Genetically targeted purification of polysomal mRNAs	[24]
Global run-on sequencing (GRO-Seq)	Transcript analysis	[25]
Reduced representation bisulphite sequencing (RRBS-Seq)	Genome methylation	[26]
Bisulfite sequencing (BS-Seq)	Genome methylation	[27]
Parallel analysis of RNA ends sequencing (PARE-Seq)	microRNA target discovery	[28]
Targeted DNA-Seq	A subset of a genome (for example, an exome)	[29]
Methyl-Seq	Sites of DNA methylation, genome-wide	[30]
Targeted methyl-Seq	DNA methylation in a subset of the genome	[31]
Hi-C	Three-dimensional genome structure	[32]
Chia-PET	Long-range interactions mediated by a protein	[33]
Ribo-Seq	Ribosome-protected mRNA fragments (that is, active translation)	[34]
Synthetic saturation mutagenesis	Functional consequences of genetic variation	[35]
MAINE-Seq	Histone-bound DNA (nucleosome positioning)	[36]
FRT-Seq	Amplification-free, strand-specific transcriptome sequencing	[37]
PARS	Parallel analysis of RNA structure	[38]
Deep protein mutagenesis	Protein binding activity of synthetic peptide libraries or variants	[39]
Repli-Seq	Replication	[40]
DNase-Seq, Sono-Seq, and FAIRE-Seq	Active regulatory chromatin (that is, nucleosome-depleted)	[41]
NET-Seq	Nascent transcription	[42]
Immuno-Seq	The B-cell and T-cell repertoires	[43]
PhIT-Seq	Relative fitness of cells containing disruptive insertions in diverse genes	[44]
Nacent-Seq	Transcription	[45]
ChIRP-Seq	Genome localization	[46]
Massively parallel functional dissection sequencing (MPFD)	Enhancer assay	[47]
Assay for transposase-accessible chromatin using sequencing (ATAC-Seq)	Open chromatin	[48]
Structure-Seq	RNA structure	[49]
RNA on a massively parallel array (RNA-MaP)	RNA-protein interactions	[50]
SEQ-500	Genome sequencer	[51]
RNA immunoprecipitation sequencing (RIP-Seq)	RNA-protein interactions	[52]
HiSeq 2000/2500/4000/X10	Genome sequencer	<a href="http://www.illumina.com">www.illumina.com</a>
MGISEQ-2000	Genome sequencer	<a href="http://www.en.mgi-tech.com">www.en.mgi-tech.com</a>
NovaSeq 6000	Genome sequencer	<a href="http://www.illumina.com">www.illumina.com</a>
PacBio Sequel/II/HiFi	Genome sequencer	<a href="http://www.pacb.com">www.pacb.com</a>
Nanopore PromethION/MinION	Genome sequencer	<a href="http://www.nanoporetech.com">www.nanoporetech.com</a>
MiSeq	Genome sequencer	<a href="http://www.illumina.com">www.illumina.com</a>
TruSeq	Genome sequencer	<a href="http://www.illumina.com">www.illumina.com</a>
DNBSEQ-T7	Genome sequencer	<a href="http://www.en.mgi-tech.com">www.en.mgi-tech.com</a>
MeDip-Seq/DIP-Seq	Methylated DNA immunoprecipitation sequencing	<a href="http://www.illumina.com">www.illumina.com</a>

### 3. Plant Genomic Resources (Big Data Generation)

Sequencing technologies, mainly using high-throughput NGS sequencers, generate significant amounts of data. For example, the recent sequencer from Illumina (NovaSeq 6000) has a higher output than the earlier generation of sequencing machines producing

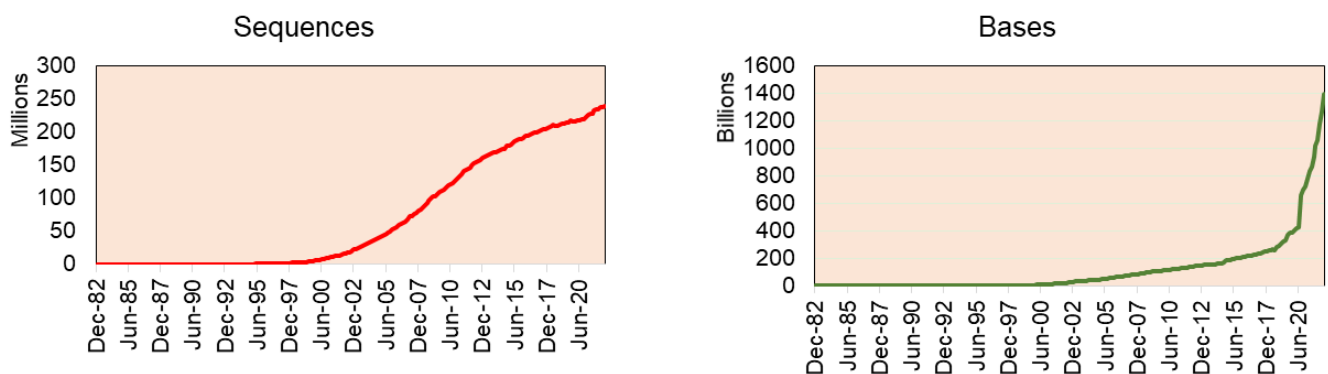
between 1300–20,000 million reads (65 Gb to 3 Tb). The long reads from PacBio reach up to a maximum of 300 Kb, and the data generated with Sequel I, II (CLR), II (HiFi) range from 0.5 million to 400 million reads (15 Gb to 100 Gb), with the nanopore sequencing technology (Minion and Promethion) sequencing ranging from 2.5–12 million reads (40 Gb to 180 Gb).

With this capacity, sequencing land plants having a wide range of genome size DNA content can, in theory, possibly generate good coverage of the entire genome sequence data. For example, the corkscrew plant *Genlisea margaretae* with a 1C value of 0.07 pg (65 Mb) and the canopy plant *Paris japonica* with a 1C value of 152.2 pg (148.9 Gb) are equally accessible in terms of raw sequence generation and coverage [53] (<https://cvalues.science.kew.org/>). Generating several-fold coverage of genomic data produces potentially massive datasets, ranging from Gb to Tb of sequence information. Depending on the scope of the project, handling such large datasets is a major concern for small (or even big) research labs. Decades ago, geneticists were mostly involved in lab work; now, the most limiting factor is the analysis of the data to derive meaning or interpretation out of it using computational tools. Understanding the algorithms and processing the data are a crucial part of genetics and genomics data analysis when searching for biological meaning.

Genomic sequencing is a field where handling big data and its processing requires a suitable storage and data transfer platform, such as is present in cloud technologies. These are extensively applied to enhance the availability of the data to all researchers in a project and indeed researchers worldwide. The genome sequence data generated for a crop genome project are immense; for example, a single *Sorghum* genome sequence contains over 50 gigabytes of raw data (depending on the data format generated), and processing the data for large population-wide studies, such as finding deeper scientific insights, marker–trait association, analyzing diversity, domestication, and assessing data from gene-editing technologies, requires robust storage and computing capacities.

To maintain the uniformity of the data in the global databases, the members' databases (GenBank, EMBL, DDBJ, CNGBdb, IBDC) of the International Nucleotide Sequence Database Collaboration (INSDC) [54] share and update genomic data periodically.

The recent stats release of GenBank reports having 16.7 trillion nucleotide bases for 1.7 million whole genome sequences (as of June 2022) (GenBank and WGS Statistics ([ncbi.nlm.nih.gov](https://ncbi.nlm.nih.gov))) (Figure 2). Of which, green plant data (Viridiplantae) alone have 93.8 million sequences from 2324 genomes (including variants of the same plant species genome), including genomic DNA/RNA for 33.4 million sequences, mRNA for 41.5 million sequences, and rRNA for 80,709 sequences.



**Figure 2.** The sequences deposited in Genbank by year indicate that genome sequence projects have increased steadily since 2002, while absolute data have exploded since 2020.

With the increasing complexity of genomic data themselves, the major databases also integrate other genomic features and provide tools to search and retrieve these datasets. The Entrez system of NCBI is one such tool allowing users to search, view, and download the sequences from GenBank. Other modes of data accessibility allow for downloading



from the FTP site (<ftp.ncbi.nlm.nih.gov>) or downloading data programmatically with the provided public API to the Entrez system (<https://eutils.ncbi.nlm.nih.gov>).

Numerous databases have been developed for genomic data to suit a variety of different purposes (Table 2). Based on the data catchment of the database, the database is as big as a global repository holding the sequences of all species, like Ensembl Plants, the National Centre for Biotechnology Information (NCBI), PlantGDB, the Plant Genome Database Japan (PGDBj), to medium size databases hosting only plant genome assemblies/annotations, like Phytozome and the Legume Information System (LIS) (<https://www.legumeinfo.org>), to smaller databases containing crop/plant-specific information, such as for the chickpea SSR database (<https://cegresources.icrisat.org/CicArMiSatDB/index.html>) [55] and chickpea SNP and indel database (<https://cegresources.icrisat.org/cicarvardb/>) [56]. However, the medium to smaller databases are limited to the scope of species-level data, like the LIS and proposed angiosperms database [57], and may do not need to use powerful bioinformatics tools and computational resources to explore the terabytes of genomic data, and many such databases were earlier discussed in [58].

**Table 2.** Recent developments and availability of plant databases.

Database Name	Description	Website	Ref
PlantcircBase	Plant circular RNAs	<a href="http://ibi.zju.edu.cn/plantcircbase/">http://ibi.zju.edu.cn/plantcircbase/</a>	[59]
Fine-Root Ecology Database	Fine root trait database	<a href="http://roots.ornl.gov">http://roots.ornl.gov</a>	[60]
ATTED-II	Coexpression database	<a href="http://atted.jp">http://atted.jp</a>	[61]
Planteome	Plant reference and species-specific ontologies for plants	<a href="http://www.planteome.org">http://www.planteome.org</a>	[62]
PLADIAS	Plant diversity analysis and synthesis	<a href="http://www.pladiaz.cz">www.pladiaz.cz</a>	[63]
TRY plant trait database	Plant trait data	<a href="https://www.try-db.org">https://www.try-db.org</a>	[64]
PmiREN	Small non-coding RNA molecules database	<a href="http://www.pmiren.com/">http://www.pmiren.com/</a>	[65]
Plant DNA C-values	The catalogue of C-value data for land plants and algae	<a href="https://cvalues.science.kew.org/">https://cvalues.science.kew.org/</a>	[53]
PlantPepDB	Phyto-peptides for various therapeutic purposes	<a href="http://www.nipgr.ac.in/PlantPepDB/">http://www.nipgr.ac.in/PlantPepDB/</a>	[66]
MtSSPdb	<i>Medicago truncatula</i> Small Secreted Peptide Database	<a href="https://mtsspdb.noble.org/">https://mtsspdb.noble.org/</a>	[67]
GRooT	A collection of root traits in responses to environmental conditions	<a href="https://groot-database.github.io/GRooT/">https://groot-database.github.io/GRooT/</a>	[68]
MPDB	Medicinal plant database	<a href="https://www.medicinalplantbd.com/">https://www.medicinalplantbd.com/</a>	[69]
GreenPhylDB	Exploration of gene families and homologous relationships among plant genomes	<a href="https://www.greenphyl.org">https://www.greenphyl.org</a>	[70]
PlantscRNAdb	Plant single-cell RNA analysis	<a href="http://ibi.zju.edu.cn/plantscrnadb/">http://ibi.zju.edu.cn/plantscrnadb/</a>	[71]
TarDB	Plant miRNA target sequences	<a href="http://www.biosequencing.cn/TarDB">http://www.biosequencing.cn/TarDB</a>	[72]
<i>Xylella</i> spp.	Host plant species	<a href="https://www.efsa.europa.eu/en/microstrategy/xylella">https://www.efsa.europa.eu/en/microstrategy/xylella</a>	[73]
PlantGSAD	Gene set annotation plant species	<a href="http://systemsbiology.cau.edu.cn/PlantGSEAv2/">http://systemsbiology.cau.edu.cn/PlantGSEAv2/</a>	[74]
CpGDB	Plant chloroplast database	<a href="http://www.gndu.ac.in/CpGDB">http://www.gndu.ac.in/CpGDB</a>	[75]
DBPR	Plant protein, DNA, RNA, Pathway, and Expression Database	<a href="https://www.habdsk.org/dbpr.php">https://www.habdsk.org/dbpr.php</a>	[76]
PtncRNAdb	tRNA-derived non-coding RNAs database	<a href="https://nipgr.ac.in/PtncRNAdb">https://nipgr.ac.in/PtncRNAdb</a>	[77]

#### 4. Plant Genome Assemblies

Genome assembly refers to aligning the small fragments of a DNA sequence to reconstruct the genome sequence in the original order and orientation. High-throughput sequencing through first- and second-generation sequences has enabled the assembly of many plant genomes. The highly fragmented genome assemblies generated with short reads have been improved with long read sequence assemblies, simplifying and improving the ability to generate chromosome-level assemblies with reduced reliance on dedicated research experts.

Thanks to the NGS technology and increased computational power, the standard of the genome assemblies available has improved significantly. Genomics has accelerated its growth in the past decade from draft-level genome assemblies to reference-level genome assemblies [78–80].

The plant genomes assembled in the FGS era faced significant throughput issues and were limited by a read length of around 1 Kb. This necessitated approaches such as

BAC-end reads and BAC barcoding to allow contigs to be linked and positioned throughout the genetic mapping. The plant genomes assembled in the FGS era are far fewer than the genomes assembled in the SGS and TGS sequencing technology era (Figure 3A), primarily due to the lower throughput and high cost of FGS. The situation changed sharply with SGS, as the volume of the sequence (although not the length) was significantly increased. Long-read sequence technologies play a crucial role in genome assembly projects, which helps in scaffolding the contig sequences, and thus many genome projects were initiated with combined SGS and TGS technologies (Figure 3B). With the advent of advanced sequence technologies such as PacBio HiFi sequencing, which produces a 10 to 30 Kb circular consensus sequence, thus reducing error rates (CCS) [11], Oxford Nanopore long-read protocols [81], Hi-C scaffolding [32], and optical mapping technologies, such as Bionano [82], it is possible to assemble complex genomes. The emerging third-generation sequence data have boosted the genome assembly quality to build a chromosome-level assembly by overcoming the limitation of short reads assembly, particularly in plants, where islands of repeat sequences need to be bridged between the gene-rich regions of the chromosomes. With the low-cost and high-throughput sequence data generations, at least 1143 plant reference assemblies have been published ([www.plabipd.de](http://www.plabipd.de)) (Supplementary Table S1). Based on the availability of funds and the feasibility of applying high-volume sequence data generation, multiple individuals of the same species were *de novo* assembled, e.g., potato [83], or the genome assembly of the same varieties improved, such as for chickpea [84,85] and sesame [86]. The development of long-read technologies as part of the TGS allowed for a relatively simple assembly of smaller genomes. With optical and chromatin-based methods, such as Bionano and HiC, far more comprehensive and larger genome assemblies are now possible, which are based on a range of techniques, including the integration of scaffolds into the chromosome through genetic mapping.

In recent years, gold-standard and platinum-standard chromosome-level genome assemblies are being achieved in prominent model crop plants [87–92]. Here, gold-standard assembly refers to cases where the number of superscaffolds matches the number of haploid chromosomes, yielding a chromosome-level assembly; a platinum-standard assembly refers to a telomere-to-telomere (T2T) assembly with the final scaffolds matching the number of haploid chromosomes. This era has led to gold- or platinum-standard assemblies in crop plants, and publications meeting these standards are continuing to appear [93]. The importance of having platinum-standard reference genome assemblies and the importance to compare cultivated species with wild relatives of rice is documented [94].

Chromosome-level genome assemblies were initiated with *Arabidopsis* in 2000 [95] and later with rice in 2005 [96]. These assemblies were generated with the traditional, expensive, and low-throughput Sanger sequencing method. With current third-generation sequencing (such as PacBio, HiFi, Hi-C, and optical mapping methods), it is possible to generate chromosome-level pseudomolecules [97]. With PacBio sequence data, a chromosome-level assembly was first achieved for *Arabidopsis* [98] followed by *Oropetium* [99]. Similar to the PacBio long reads, ONT generates around 200 Kb length reads highly suitable for bacterial genomes assembly [100]. Synthetic long reads (SLR) are long reads generated from Illumina short-read data to assemble long reads [101]. In total, 113 plant species have the chromosome-level genome assemblies published (as of the end of 2022) ([www.plabipd.de](http://www.plabipd.de)) of the total assembly number of 1143 flowering plants, and 125 are non-flowering plants (Supplementary Table S1). Most of these near-complete plant genomes were produced with sequence data generated from multiple technologies. The long-read 10× Genomics with short-read Illumina data were used to assemble the blueberry genome [102]. PacBio and Hi-C sequence technology were used for assembling the octoploid sugarcane genome [103], allotetraploid peanut [104], and teff [105].

Several novel technologies have emerged (such as optical mapping [106]), the Irys system by BioNano Genomics ([www.bionanogenomics.com](http://www.bionanogenomics.com)) and chromosome conformation capture sequencing (Hi-C) [32]) to improve the scaffolding without depending on genetic mapping. However, these advances in genome assembly have recently improved

further to generate the telomere-to-telomere (T2T) assemblies, as first implemented in 2020 for the X chromosome sequence of the human genome [107] and later adapted to plants, such as *Arabidopsis* [108,109], rice [110], and banana [111] (Table 1). The combined integration of PacBio and modified Hi-C protocol as Dovetail Genomics has improved the assembly contiguity for *A. alpina* [112]. The high-resolution gap-free T2T genome assemblies ensure the capture of all the repetitive sequences and genomic variants without any misassemblies.

The greatest bioinformatics challenge for sequencing plant genomes was repetitive sequences, leading to sequencing errors and unrecognizable assembling errors at earlier stages of assembly computation. As the plant genome size and ploidy or repeat content increases, the complexity of assembly of the sequence reads correctly also increases, and thus the assembly programs used in these genome projects needed increasingly sophisticated strategies (such as chromosome flow sorting methods used in wheat) to handle such challenges. Additionally, handling the terabytes of sequence data and storage and managing the computing clusters and complexity of the algorithms also need to be addressed.

In addition to improving the quality of reference genomes to platinum-standard, present-day technologies paved the way for the transformational shift from the representative single genotype's genome sequence to the pan-genome sequence as a reference for a better understanding of the variability present within a species [113]. The advantages of the pan-genome reference are being realized in generating novel insights and the identification of the genes or genomic regions underlying the important agronomical traits and domestication process [86,114–118].

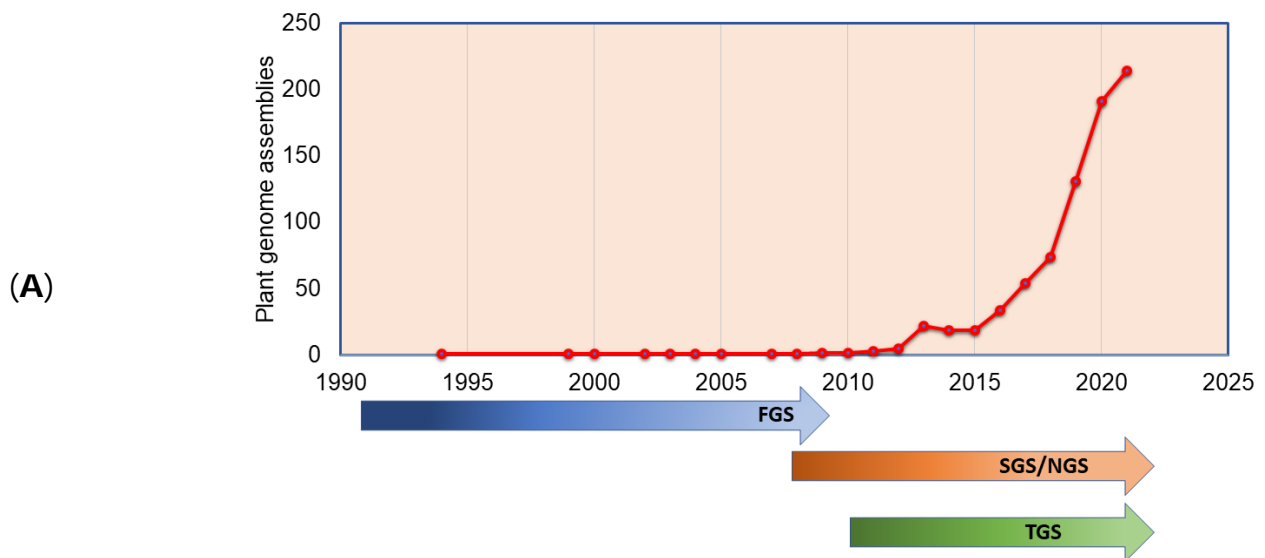
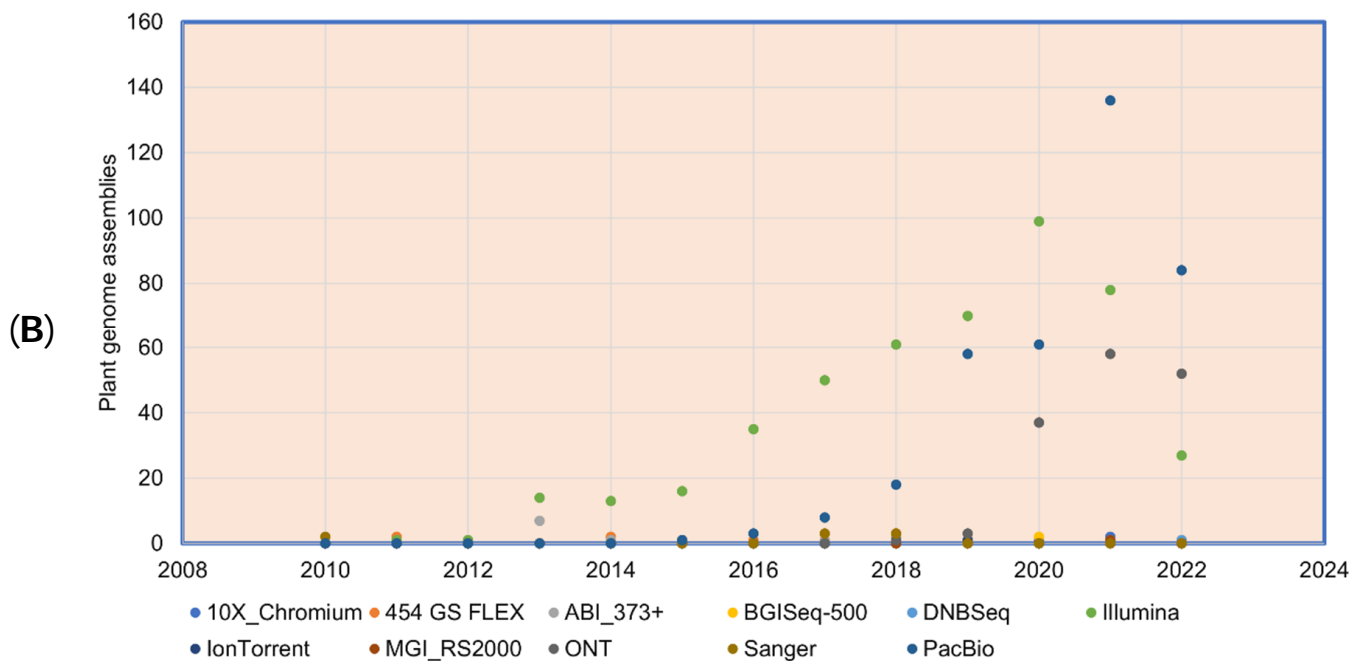


Figure 3. Cont.



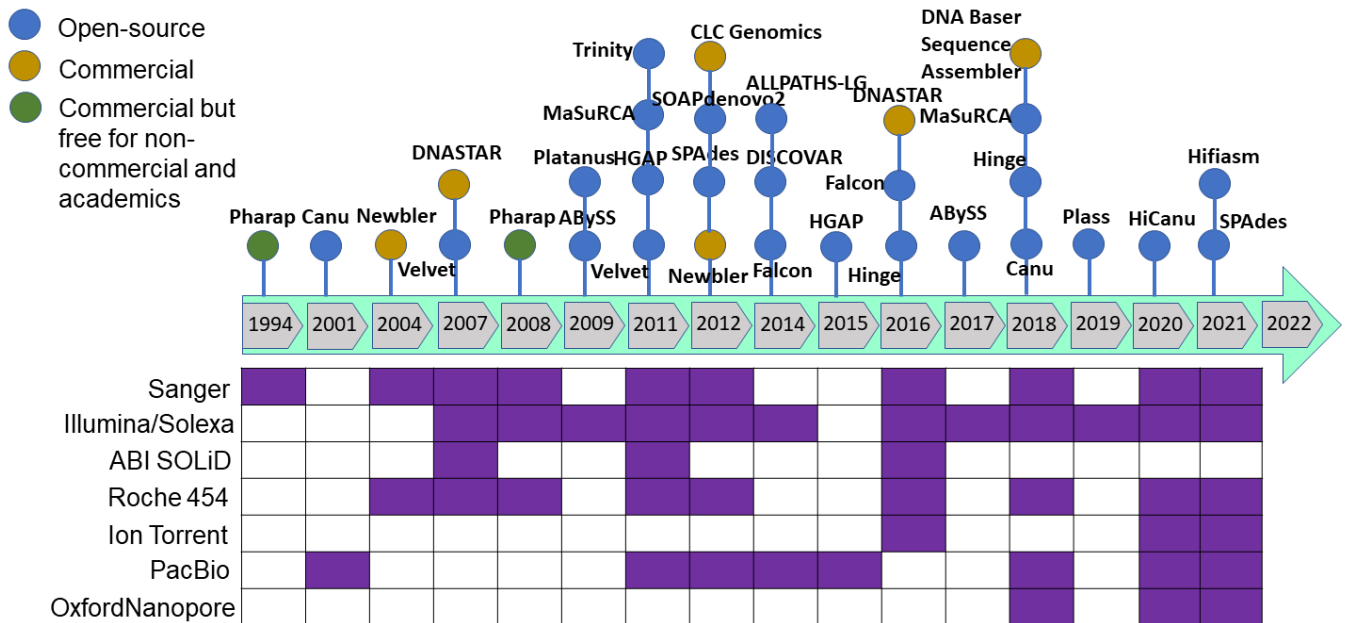


**Figure 3.** Plant genomes assembled in (A) different generations of sequencing technology and (B) sequence technologies used for plant genome projects (the genome assembly projects mostly use multiple sequencing techniques to gain higher accuracy, and the data point indicates the count of the number of times the sequencing technique has been reported in a particular year for genome assembly). The plant genome assemblies stats are derived from [119].

## 5. Genome Assemblers

As sequencing technology evolved, assembly approaches also had to evolve. The Celera Assembler and Arachne assemblers were developed to handle genomes of the fruit fly (*Drosophila melanogaster*) and human genome in 2000–2003; later, AMOS was launched under an open-source framework. These assemblers were developed based on overlap–layout–consensus on an overlap graph [120] in which the nodes were the reads and the edges represented the shared sequence between reads. This type of assembler is suitable for assembling FGS technology sequencing reads produced by the dideoxy termination method (Sanger sequencing). As massively parallel high-throughput sequencing technology was developed to produce millions of bases (in SGS), the read size became smaller and more error-prone with higher genome coverage. The leading Illumina technology of SGS/NGS sequencing technology yields 35–150 bp length paired-end reads from fragments with a 200–300 bp insert size. Such high-throughput data required a new approach, and thus *de Bruijn* graph-based assembly was developed [121,122] where the nodes represent fixed-length strings drawn from a larger set of strings, and the edges represent perfect shared sequences. However, *de Bruijn* graph-based assemblers have difficulties handling sequencing errors and need high computational power (100+ Gb of memory). The challenge with uneven genome coverage and reads too short to span repeated regions can be addressed by a combination of many short reads and fewer longer reads or mate–pair reads (Sanger, 454 and Illumina sequencing methods). Multiplex *de Bruijn* graphs automate the assemblies of long HiFi reads [123], and the recently updated Minimap2 version can be used for long read assembly [124]. Newbler was the first assembler released in 2004 to assemble the 454 sequence data followed by a hybrid version of the MIRA assembler for 454 and mixed with Sanger reads. After upgrading the Illumina sequence technology to produce from the initial 36 base-length read to reads over 100 bases in length, the produced sequence was suitable for *de novo* assembly. After the release of the SHARCGS assembler for

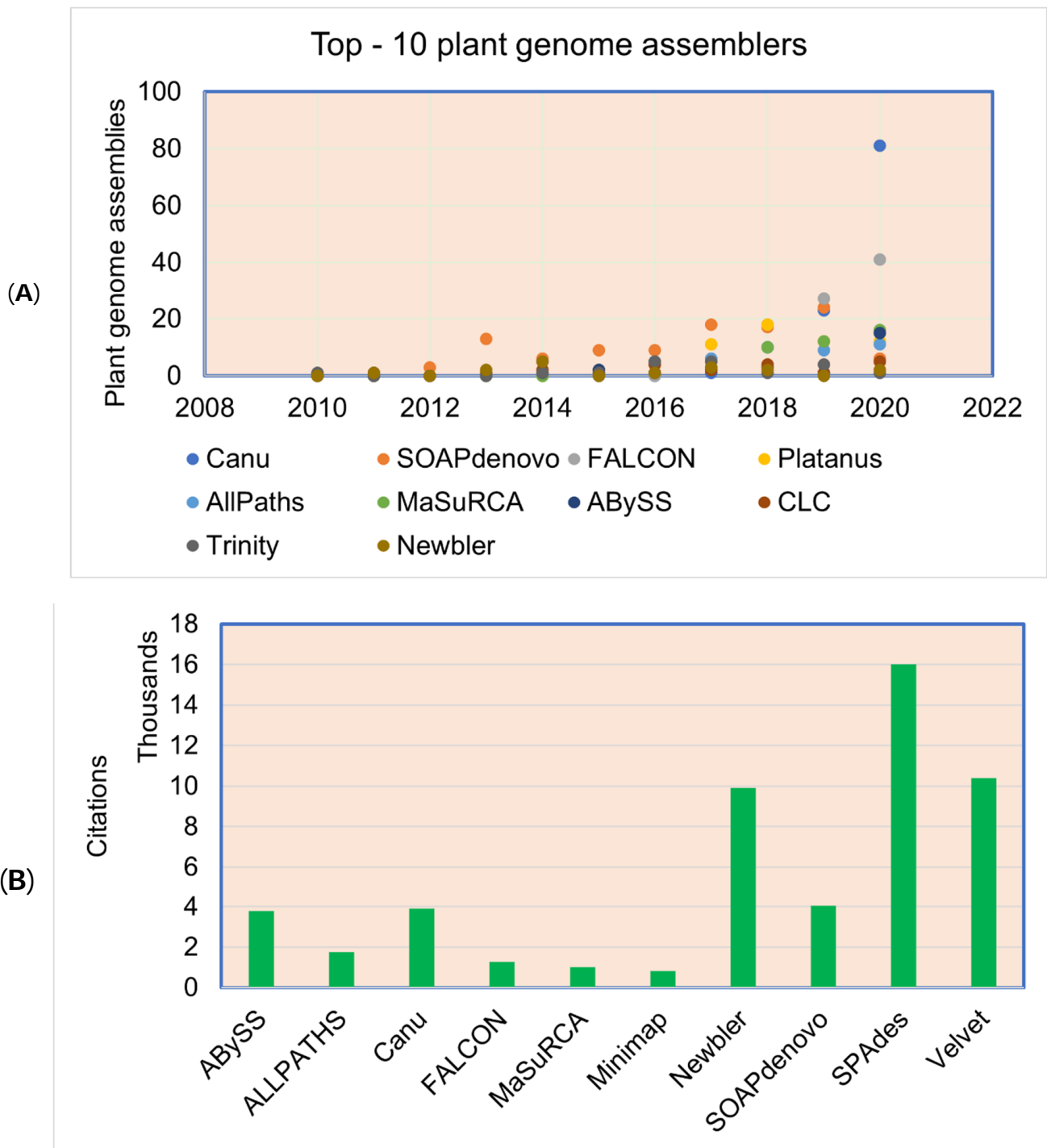
Solexa reads, other assemblers were released and became the most popular assembly tools (Figures 4 and 5).



**Figure 4.** The most commonly used genome assemblers and their release timeline to analyze the sequence generated from variant sequencing technologies.

Plant genome assembly was initiated with *Arabidopsis thaliana* in December 2000 [95] where the approach relied on overlapping bacterial artificial chromosome (BAC) clones which were end sequenced and the same approach was applied to the crop plant rice [125,126]. Later, the emerging whole genome shotgun (WGS) strategy was applied to black cottonwood [127]. This was where more difficulties and challenges were faced to assemble the short sequence reads, which resulted in a more fragmented assembled genome sequence followed by two versions of the grapevine genome sequence in 2007 [128,129]. A hybrid approach was adopted to sequence the cucumber with Illumina and Sanger sequencing technology, indicating the feasibility of using this approach for plant genome sequencing [130]. With the change in technology, 454 combined with the Sanger sequencing approach was applied to the genomes of apple [131], cocoa [132], and muskmelon [133]. In 2011, the first plant genome was sequenced using SGS technology combining 454, Illumina, and the SOLID platform for strawberry [134], Chinese cabbage [135], potato [136], chickpea [137], pigeonpea [138], and watermelon [139].

The advances in sequencing technology (SGS and TGS) and assembly approaches have removed the limitation of genome sequencing for not only the crops with small genome sizes but also enabled sequencing and assembly of large genome crops, like wheat (~17 Gbp) [87,140,141], barley (5.1 Gbp) [142], rye (~7–8 Gbp) [143], and tea (~3.8–4.0 Gbp) [144], which are important for animal feed and human nutrition.



**Figure 5.** (A) The most commonly used genome assemblers for the plant genome assemblies’ projects and the (B) citations of the assemblers.

The genome assembly quality has improved as the sequencing technologies and assembling tools improved (Figures 3–5), especially when combined with the utilization of multiple sequencing technologies of TGS, for example.

The initial assembly version of the sorghum genome assembly released in 2009 [145] with shotgun sequencing and BAC libraries data captured 738.5 Mb of sequences in 12,873 contig sequences (scaffolded to 3304 sequences), which is more fragmented compared to the chromosome-scale assembly of the sorghum genome using nanopore sequencing and

optical mapping data that produced a hybrid assembly made of 29 scaffolds capturing the 661.16 Mbps [146].

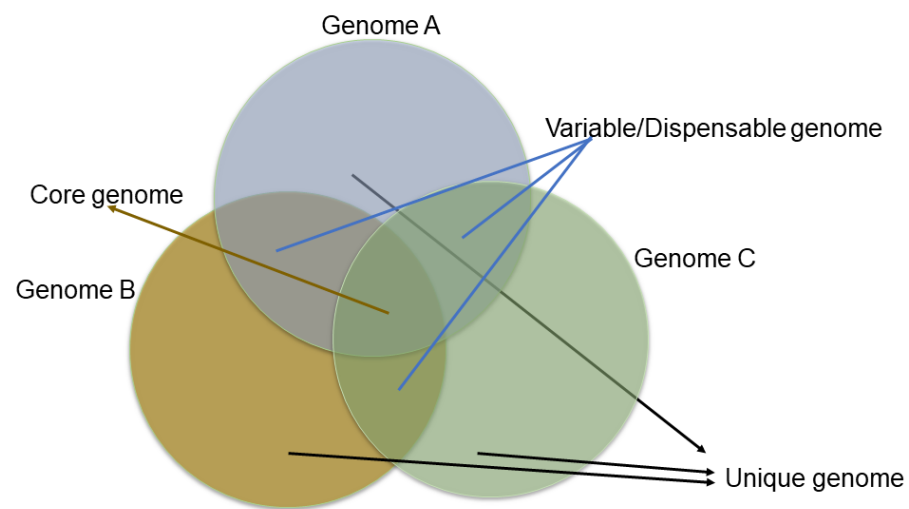
For a large genome (~8 Gb) rye (*Secale cereale*), initially, a virtual linear gene order model (22,426 genes) was established with high-throughput transcript mapping and chromosome survey sequencing [147]; following reference genome assembly with a shotgun, *de novo* genome assembly produced 1.29 million scaffolds, capturing 2.8 Gbp of sequence [148] and later chromosome-scale genome assembly with 10×, HiC, Bionano optical genome mapping, and chromosome-specific shotgun (CSS) reads produced 6.74 Gb (of estimated 7.9 Gb) [149].

In addition to the chromosome-scale assemblies, TGS has enabled the assembly of polyploid genomes, such as bread wheat [87], potato [150], and peanut [151].

## 6. Advancements in Plant Genomics

With the emerging sequence technology and bioinformatics tools, it is possible to assemble a nearly complete genome sequence. With cytogenetic advances to measure the genome size (such as flow cytometry), a genome size estimation is a useful first step in a complete genome sequencing project. The amount of sequencing data required to produce a given level of coverage depends on the 1C amount of DNA per cell (including ploidy level), and for most species, this can be found in the Kew Plant Genome Database. Most plant genome assemblies are smaller than the cytogenetic genome estimation size; this may be because of assembly errors or difficult-to-approach genomic regions, like centromeric and repetitive regions in the plant genome, where assemblers struggle (physical maps, such as Bionano, resolve such issues). Some of the assembled plant genome sizes are quite close to the cytogenetic estimated size, indicating the assembler has captured the majority of the genome content. Assemblies above the estimated size, however, may need refinement to reduce contaminants or alter the assembly parameters.

The genome assembly provides the coordinate system for the gene models and other genomic features, like SNPs, Indels, SSRs, etc. Predicting the gene models with *ab initio* gene findings and supporting evidence in the form of RNA data increases the accuracy. However, this may not list out the complete complement of genes of the species for which resequencing a wide range of diverse accessions will reveal more genes that are genotype-specific. For example, the resequencing of >1000 wild and cultivated rice accessions has predicted the presence of thousands of genes with lower sequence diversity in cultivated rice, indicating a rice domestication genetic bottleneck [114,152]. Moreover, genetic diversity is often reduced during domestication, and resequencing a single individual may not capture the species-wide gene content. Thus, the concept of the pan-genome was developed and adapted to plants' genomes to identify the species-wide gene content. The core genome is usually defined as the housekeeping genes (which must be present for the organism to survive and reproduce) and the variable/dispensable genes (these genes are present or absent in a particular cultivar/accession of a species) that exhibit the gene diversity or variability in a species (Figure 6). Thus, the first plant pan-genomes appeared in 2007, describing the variable genes in rice and maize genomes, and were later adapted to a wide range of plant genomes [153], including banana [154], white lupin [155], barley [156], wheat [156], wheat panache [157], and sorghum [158] (Table 3).



**Figure 6.** Venn diagram showing core and accessory genes.

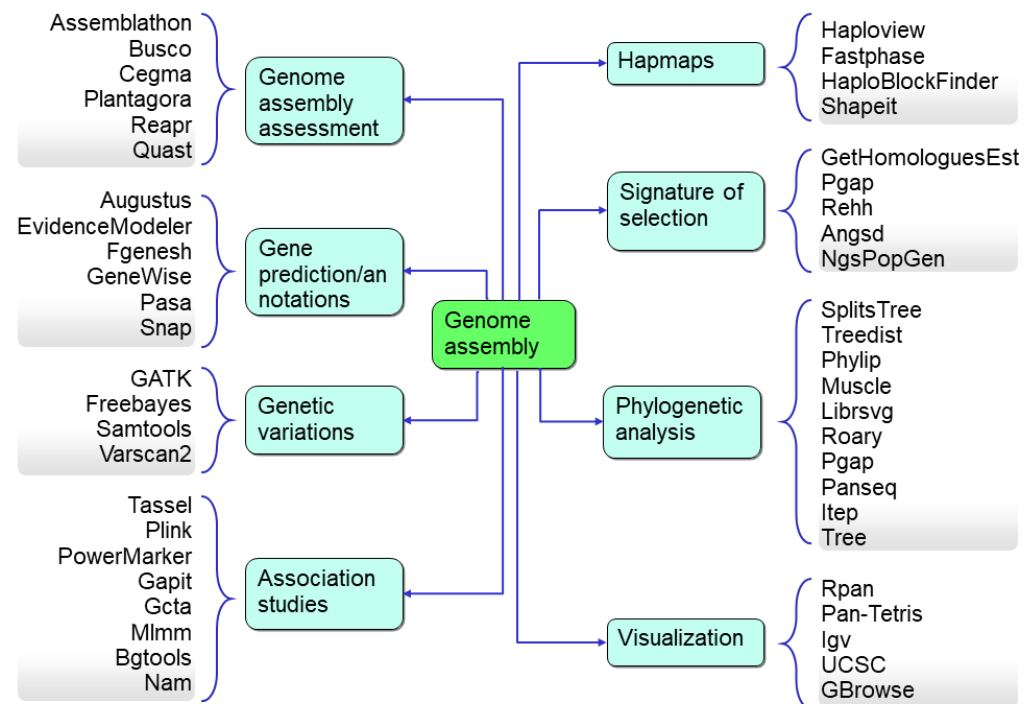
**Table 3.** The pan-genome assemblies.

Approach	Species	Domestication Status	Ploidy	Number of Accessions	Reference
<i>de novo</i>	<i>Brassica rapa</i>	Crop	Diploid	3	[159]
<i>de novo</i>	<i>G. soya</i> (soybean)	Wild	Tetraploid	7	[160]
<i>de novo</i>	<i>O. sativa</i>	Crop	Diploid	3	[161]
<i>de novo</i> transcriptome	<i>Zea mays</i> (maize)	Crop	Diploid	503	[162]
<i>de novo</i> metagenome assembly	<i>O. sativa</i> (indica/japonica)	Crop	Diploid	1483	[163]
Iterative assembly	<i>B. oleracea</i>	Crop	Diploid	10	[164]
Read mapping	Populus (poplar)	Wild	Diploid	7	[165]
<i>de novo</i>	<i>B. distachyan</i>	Wild	Diploid	54	[166]
<i>de novo</i>	<i>Medicago truncatula</i>	Wild	Diploid	15	[167]
Iterative assembly	<i>Triticum aestivum</i> (bread wheat)	Crop	Hexaploid	19	[115]
Iterative assembly	<i>B. napus</i>	Crop	Tetraploid	53	[168]
Iterative assembly	<i>Capsicum</i> (pepper)	Crop	Diploid	383	[169]
Iterative assembly	<i>O. sativa</i> / <i>O. rufipogon</i>	Crop	Diploid	67	[170]
Map-to-pan	<i>O. sativa</i> (rice)	Crop	Diploid	3010	[78]
<i>de novo</i>	<i>Sesamum indicum</i> (sesame)		Diploid	5	[171]
Iterative assembly	<i>Helianthus annuus</i> (sunflower)	Crop	Diploid	493	[172]
Iterative assembly	<i>Solanum lycopersicum</i> (tomato)	Crop	Diploid	725	[116]
<i>de novo</i>	<i>B. napus</i> (oilseed rape)	Crop	Tetraploid	9	[173]
<i>de novo</i>	<i>Juglans</i> (walnut)	Wild	Diploid	6	[174]
<i>de novo</i> , graph	<i>G. max</i> (soybean)	Crop	Diploid	29	[175]
PHG	Sorghum		Diploid	398	[176]
Iterative assembly	<i>B. napus</i>	Crop	Tetraploid	50	[177]
Iterative assembly	Pigeon pea ( <i>Cajanus cajan</i> )		Diploid	89	[113]
<i>de novo</i>	Pecan ( <i>Carya illinoensis</i> )	Tree	Diploid	4	[178]
<i>de novo</i>	White lupin	Crop	Diploid	39	[155]
Iterative assembly	Sorghum	Crop	Diploid	354	[158]
Iterative assembly	<i>Brassica napus, rapa, oleracea</i>	Crop	Diploid, diploid, amphidiploid	87, 77 and 79	[179]
Iterative assembly	Chickpea	Crop	Diploid	3366	[180]
<i>de novo</i>	Sorghum	Crop/Wild relatives	Diploid	16	[181]
Iterative assembly	Eggplant ( <i>Solanum melongena</i> L.)		Diploid	23	[182]
Iterative assembly	Banana ( <i>Musa</i> and <i>Ensete</i> )		Triplod	15	[154]
<i>de novo</i>	Tomato ( <i>Solanum lycopersicum</i> )	Crop	Diploid	838	[183]
<i>de novo</i>	Potato ( <i>Solanum tuberosum</i> L.)	Crop	Diploid	44	[83]
Iterative assembly	Lupin	Crop	Diploid	55	[184]

The most commonly used downstream analysis with pan-genome assemblies is to identify the genetic variation of any DNA segment in a genome or a gene (including gene fragments) that can be used as a marker for genotyping. Bioinformatics resources enhancing crop genomics for downstream analysis include copy number variations (CNV),



identification of variations based on the length (SNP, SSR, Indels), a set of SNPs used as a unit in the form of a haplotype to increase the resolution of GWAS, k-mer analysis, linkage disequilibrium (LD), presence–absence variations, pan-genome-wide association studies (PWAS), genotyping-by-sequencing, reduced representation sequencing, domestication, and diversity analysis (Figure 7). With these bioinformatics tools, the genomic data also assists plant phylogenomic research with useful information, such as genome diversity and speciation events. Therefore, bioinformatics has become a most essential part of plant genomics research.

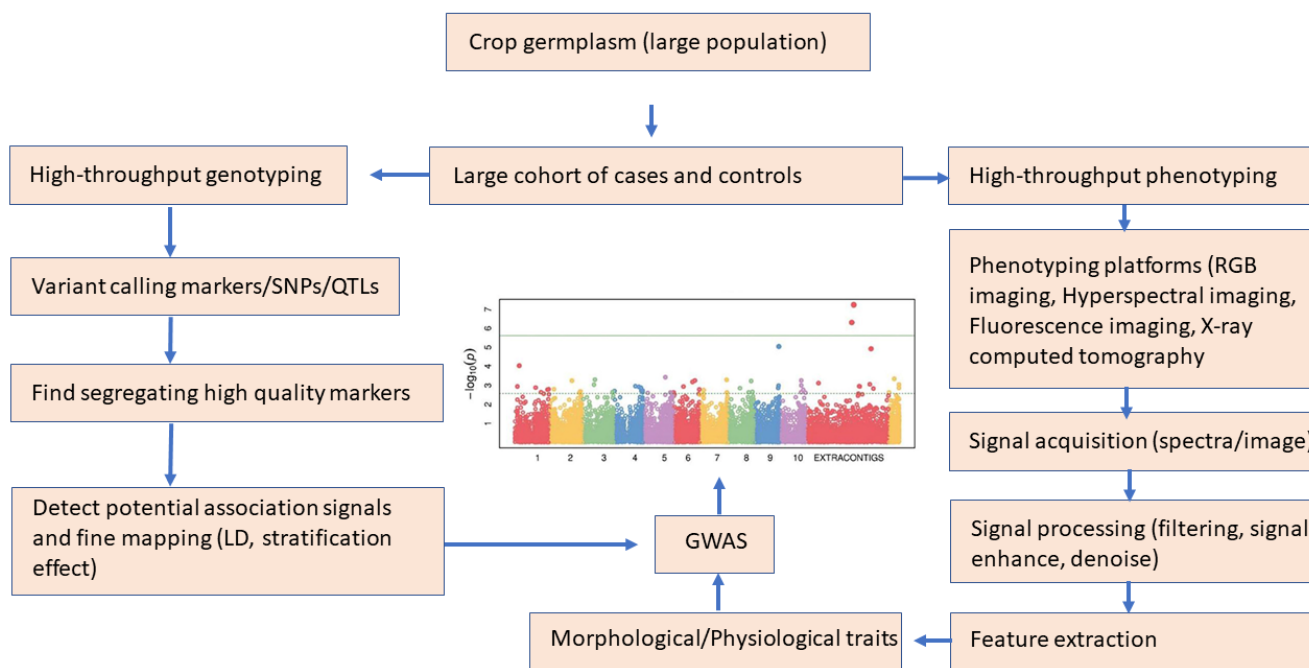


**Figure 7.** The downstream analysis and the associated bioinformatics tools used for the data analysis.

High-throughput genotyping enables the genotyping of thousands of targeted loci (genetic markers) on thousands of samples. Depending on the number of markers and the sample size, different genotyping techniques can call genotypes in different ranges. Some of the technologies include Illumina golden gate, Affymetrix SNP, reduced-representation genome sequencing, exome-seq, Fluidigm (<https://investors.fluidigm.com/node/13686/pdf>), IntelliQube (<https://www.myebpl.com/intelliqube.html>), MassARRAY [185], MassEXTEND, GeneChip [186], APEX-Seq [187], BeadARRAY (<https://www.illumina.com/science/technology/microarray.html>), TaqMan [188], and DArT (<https://www.diversityarrays.com/>). Genotyping by sequencing (GBS) is a highly multiplexed system for constructing reduced representation libraries from the sequencing platform with low-cost, reduced sample handling with no need for a reference genome. GBS (including the single digest RAD and double digest RAD and skim-sequencing) are tools for genomics-assisted breeding in a range of plant species through the applications of SNPs identification, gene/QTL mapping, molecular diversity, GWAS, construction of high-density genome maps, haplotype maps, phylogenetics, identification of candidate genes, genetic linkage analysis, molecular marker discovery, and genome sequencing and selection. Such genetic resources assist in predicting the genetic value of selected candidates based on the genomic estimated breeding values (GEBV) from high-density and quality markers. Genomic selection (GS) is an approach to exploit genetic markers to develop new markers-based models to increase the genetic gain of complex traits for breeding programs. High-throughput marker technologies have

changed the entire scenario of marker applications and enabled the use of GS routine work for crop improvement.

Plant phenotyping through conventional methods relies on manual measurements, which are laborious, error-prone, and time-consuming. Similar to genotyping, high-throughput phenotyping (HTP) (“phenomics”) has unique advantages in facilitating accurate, automated, high-quality data collection techniques, including visible light imaging, X-ray computed tomography, visible and near-infrared spectroscopy, multi-spectral imaging, chlorophyll fluorescence, fluorescence imaging, and nuclear magnetic resonance (NMR) [189] (Xiao et al., 2022). These tools are generally used to obtain high-resolution images of samples from which features are extracted with image processing algorithms. Mostly machine learning algorithms are used to generate robust data processing to produce accurate and time-efficient phenotypes of plants [190]. Highly accurate genotype and phenotypic data need appropriate statistical methods to identify true associations between genetic and phenotypic variation (Figure 8). Plant phenotyping systems, imaging techniques, challenges, and their applications have been reviewed elsewhere, including imaging systems, data collection methods, and analysis techniques and problems [191–193]. GWAS has high efficiency and high resolution and is conducted on a genome-wide scale with statistical programs. Some of the R packages developed for association analysis are GAPIT [194], qqman [195], gwasrapidd [196], eQTpLot [197], Postgwas [198], GWASTools [199], and IntAssoPlot [200].



**Figure 8.** A schematic workflow of genotype and phenotype workflow for GWAS (plot adapted from [153]).

## 7. Data Science and Artificial Intelligence

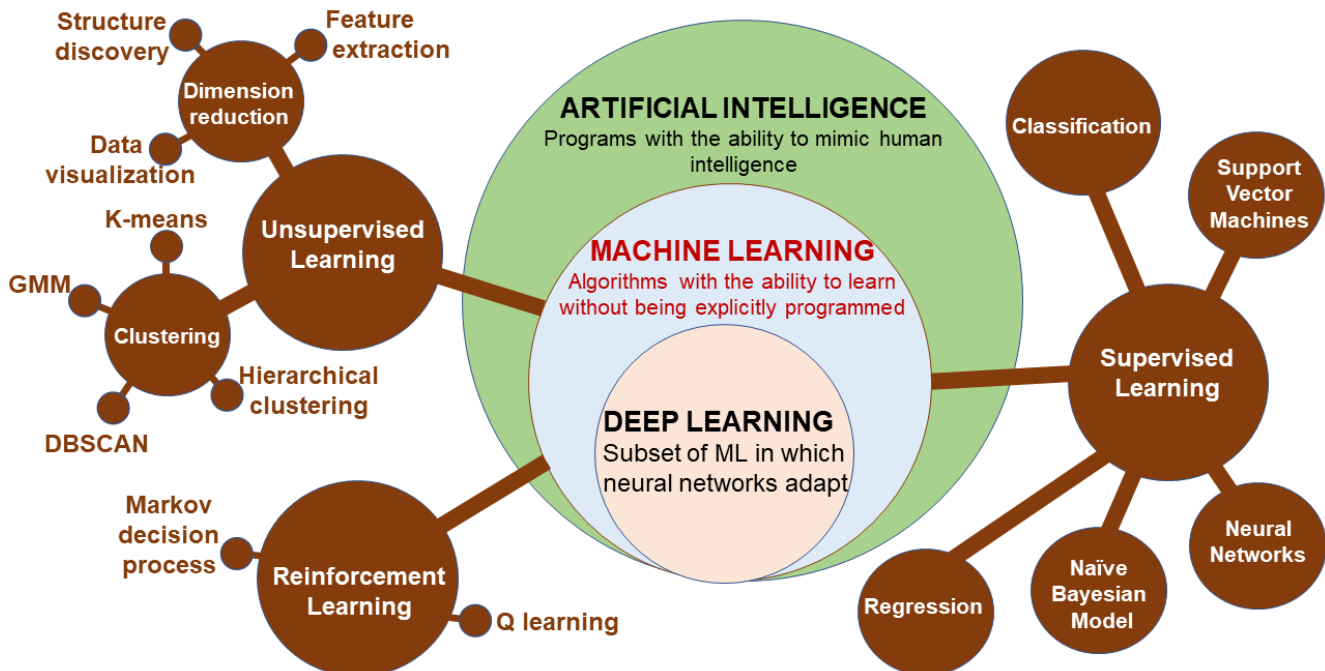
Genomics data science is a field that needs powerful computational and statistical methods to decode the information in plant genomic DNA. Having a better understanding of genomics with these data science tools helps researchers to uncover the differences between the varieties at a DNA level and enhance crop improvement. Bioinformatics has emerged to bring in vivo experimentation and in vitro data analysis with statistical and computational tools to process the data by developing and implementing the algorithms as software tools to make predictions based on the experimental data.

Researchers are now generating more genomic data to understand genome functions and mine genetic information to explore novel insights from the vast amounts of generated

genomic data. Sequencing huge numbers of individuals of a species generates terabytes of data, and processing such large amounts of data needs additional terabytes to petabytes of storage and working computational infrastructure. Researchers need special computational and software tools to mine and interpret hidden biological information through assembling the sequence data, aligning the sequence reads, and mining the variation, association studies and other genomic insights [201].

Artificial intelligence (AI) tools help researchers process vast quantities of genomic sequence data to find patterns in a genome [202]. AI typically contains hidden layers of analysis leading to biases in generating the results and may be undetected [203]. Thus, there is a need to apply human intelligence to validate the prediction/results in other dimensions.

Machine learning (ML) is a subset of artificial intelligence (AI) involving the development of algorithms that learn to perform a specific task based on given inputs. ML is implemented in either supervised learning (predicting output based on the given input features describing the object) or unsupervised learning (seeking patterns comparison and grouping the data) [204]. Supervised learning can be further grouped into two categories of algorithms: classification and regression. Similarly, unsupervised learning categories include clustering and association. Reinforcement learning is a feedback learning method in which the right action has a positive score, while a negative score is for the wrong action. The deep learning (DL) approach involves using layers of neural networks, and DL uses several such layers as artificial neural networks [204]. Convolutional neural networks are effective at image processing, while recurrent neural networks deal with sequential data and support vector machines that can capture nonlinear relations between objects. A better classification of the relationships between ML methods is depicted in Figure 9.



**Figure 9.** The structure of the artificial intelligence and sub-groups of machine learning methods (DB-SCAN: density-based spatial clustering of applications with noise; GMM: Gaussian mixtures model).

ML method implementations are available in the form of a Weka (<https://www.cs.waikato.ac.nz/ml/weka/>) and Orange (<https://orange.biolab.si/>) as user-friendly graphical interfaces, scikit learn (<https://scikit-learn.org/>) [205], Keras [206], and PyTorch [207]. In Advances, in-neural information processing systems and are available as the TensorFlow

package (<https://www.tensorflow.org/overview/>) [208] in Python and the Caret package in R (<https://cran.r-project.org/web/packages/caret/vignettes/caret.html>) [209].

ML is widely used for crop improvement; some of the case studies include plant–pathogen interactions [210], traits, and phenotyping [211], and applications include at the molecular level in plants [212]. The use of ML in plant genomics has increased in the last decade [213]; applications include the classification of genes into active and inactive genes in maize [214], identifying genome crossovers [215], identification of near-complete genetically fixed genomic regions [216], gene regulatory networks in maize [217], gene prediction with deep learning with a variety of architectures [218], diagnosis of pests and disease [219], gene prediction concerning climatic conditions [220], predicted gene expression levels from genomic sequence data [221], identifying variants based on short-read sequence alignments [222], and classifying genes as core and dispensable genes [223].

The applications of ML have been widely used in phenotyping through high-throughput, image-based plant phenotyping which uses a convolutional neural network (CNN) [224] and deep learning [225]. From a recent review [226], the most commonly used genome selection R packages based on the linear mixed model and Bayesian regression model are rrBLUP and Bayesian models rrBLUP [227], BGLR [228], lme4 [229], ASReml [230], and glmnet [231]. For the multiple trait-based genome selection, MTGS (genomic selection using multiple traits) and BMTME (Bayesian multi-trait and multi-environment) [232] packages have been developed. The more detailed approaches and categories in genome selection were discussed in an earlier study [226].

ML can improve plant breeding [204,233], with plant breeders relying on genomic selection [234] to identify the QTLs (quantitative trait loci) (genomic regions associated with traits), assess the genetic architecture of the crop, and predict traits for new genotypes. ML algorithms used for such predictions are random forests [235], support vector machines [236], and gradient tree boosting [237].

Mobile apps have been designed to collect data, record details, predict plant disease, predict weather changes, and other miscellaneous applications. The apps and the underlying algorithms interpret images captured through the devices, thus reporting the health condition of plants, soil color and other phenotypes. The availability of these apps for farmers assists in detecting disorders and suggests suitable measures to protect the crop. Some of the apps are AgSpeak (<https://www.agspeak.in/>), AutoML (<https://www.automl.org/>), aWhere (<https://www.climatehot.earth/awhere>), Farm at Hand (<https://www.farmathand.com/>), Plantix (<https://plantix.net/en/>), Tumaini (<https://ciat.cgiar.org/phenomics-platform/tumaini/>), and Xarvio (<https://www.xarvio.com/global/en.html>).

## 8. Conclusions/Future Aspects

The goal of improving sequencing technology has been to generate genetic information in a faster, cheaper, and more accurate way. The more portable sequencing platforms (such as the Minion from Oxford) require less power, reagents, maintenance, and storage and have an easy processing format. It is also equally important to have advanced and compatible bioinformatics tools to analyze the big data generated from the agriculture sector.

First- and second-generation sequencing technologies generated short-sequence reads resulting in highly fragmented reference genome assemblies (unless coupled with long-range systems of mapping, such as BAC-end sequencing) but were used to generate the first reference genomes for plants. Such low-quality assemblies (compared to third-generation assemblies) have many gaps and do not represent the actual genome structure. On the other hand, combined second- and third-generation sequence data have contributed to generating full chromosome-level (CL) to T2T-level reference sequences. Only in a few plant genomes have high-quality, gapless chromosome levels to T2T quality assemblies been generated; therefore, further improvements are necessary to generate high-quality standards.

T2T-level genome assemblies will provide insights into the genetic diversity, identification of domestication events, and the investigation of the evolutionary history of plant species.

The sequencing of multiple accessions of a plant species is expected to allow the assembly of a pan-genome which represents the collection of core and dispensable genes present in a species [153]. In addition to pan-genome studies, several intensive genome and transcriptome projects have been initiated (10,000 plant genomes and 1000 plant transcriptomes) [238]. Additionally, the Earth BioGenome Project (EBP) is planning to sequence and catalogue the genome of all eukaryotes on Earth.

The recent advances and developments in bioinformatics applications for plant genomes provide huge potential for plant genome research. As sequencing technology has become much more affordable and portable to handle, the importance of bioinformatics tools increases to analyze and manage the data. More plant species genome databases are being established with a variety of analysis methods. Phylogenomics and GWAS now generate more accurate results with the tools developed with newer algorithms. Moreover, high-throughput phenotyping needs to provide results with a high resolution to meet the density of genotype information.

The genetic information in the form of sequence data or optical maps needs to be as error-free as possible, selecting the appropriate informatics tools for *de novo* assembly, scaffolding, annotation, and downstream analysis. This is key for gold- or platinum-standard genome assemblies.

With the rate of the growing world population, there is a constant increase in demand for food, and AI will play a vital role in meeting these demands, coupled with computational power through robotics, smartphone apps, and image processing algorithms. AI provides automation in agriculture. Technology is being developed in agriculture for automated methods, crop improvement, and crop protection. With computational advances, including AI, ML, and DL, the future GAB, including marker-assisted selection (MAS), MABC, marker-assisted recurrent selection (MARS) [239], haplotype-based breeding, speed breeding (SB) [240,241], and genomic selection (GS), are expected to play a key role in breeding more smart crop cultivars with higher production and nutritional value in both a cost- and time-saving manner.

In the past two decades, the parallel advances in sequencing technology and bioinformatics tools have enabled plant researchers to generate genomics resources for economically important plants, which is critical for crop improvement and to develop a greater scientific understanding of the gene underlying critical traits for future agriculture.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/life13081668/s1>, Table S1: The available plant genome assemblies (at the level of scaffold, chromosome and t2t standard quality).

**Author Contributions:** P.R. (Pradeep Ruperao) and A.R. conceived and designed this review. S.M. and A.R. supervised this review. P.R. (Pradeep Ruperao) and P.R. (Parimalan Rangan) collected and analyzed the data. P.R. (Pradeep Ruperao) and P.R. (Parimalan Rangan) wrote the manuscript. V.T., S.K., T.S. and S.M. assessed the review. All authors have read and agreed to the published version of the manuscript.

**Funding:** The authors also acknowledge the supporting funds from AVISA (OPP1198373), ICAR-BMGF (101165), and the Department of Biotechnology, Government of India for supporting this study.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** PR was supported by a grant from the Department of Biotechnology, Government of India and AVISA project. We also acknowledge the support of ICAR-BMGF towards the APC. The study is made possible through data available at [www.plabipd.de](http://www.plabipd.de), we thank the source of the database.



**Conflicts of Interest:** The authors declare that this study was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

- Blaxter, M.; Archibald, J.M.; Childers, A.K.; Coddington, J.A.; Crandall, K.A.; Di Palma, F.; Durbin, R.; Edwards, S.V.; Graves, J.A.M.; Hackett, K.J.; et al. Why Sequence All Eukaryotes? *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2115636118. [[CrossRef](#)] [[PubMed](#)]
- Lewin, H.A.; Richards, S.; Aiden, E.L.; Allende, M.L.; Archibald, J.M.; Bálint, M.; Barker, K.B.; Baumgartner, B.; Belov, K.; Bertorelle, G.; et al. The Earth BioGenome Project 2020: Starting the Clock. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2115635118. [[CrossRef](#)] [[PubMed](#)]
- Sanger, F.; Nicklen, S.; Coulson, A.R. DNA Sequencing with Chain-Terminating Inhibitors. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5463–5467. [[CrossRef](#)]
- Maxam, A.M.; Gilbert, W. A New Method for Sequencing DNA. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 560–564. [[CrossRef](#)]
- Varshney, R.K.; Pandey, M.K.; Bohra, A.; Singh, V.K.; Thudi, M.; Saxena, R.K. Toward the Sequence-Based Breeding in Legumes in the Post-Genome Sequencing Era. *Theor. Appl. Genet.* **2019**, *132*, 797–816. [[CrossRef](#)]
- Rothberg, J.M.; Hinz, W.; Rearick, T.M.; Schultz, J.; Mileski, W.; Davey, M.; Leamon, J.H.; Johnson, K.; Milgrew, M.J.; Edwards, M.; et al. An Integrated Semiconductor Device Enabling Non-Optical Genome Sequencing. *Nature* **2011**, *475*, 348–352. [[CrossRef](#)] [[PubMed](#)]
- Pichler, M.; Coskun, Ö.K.; Ortega-Arbulú, A.S.; Conci, N.; Wörheide, G.; Vargas, S.; Orsi, W.D. A 16S rRNA Gene Sequencing and Analysis Protocol for the Illumina MiniSeq Platform. *Microbiologyopen* **2018**, *7*, e00611. [[CrossRef](#)]
- Coil, D.; Jospin, G.; Darling, A.E. A5-Miseq: An Updated Pipeline to Assemble Microbial Genomes from Illumina MiSeq Data. *Bioinformatics* **2015**, *31*, 587–589. [[CrossRef](#)]
- Liu, L.; Hu, N.; Wang, B.; Chen, M.; Wang, J.; Tian, Z.; He, Y.; Lin, D. A Brief Utilization Report on the Illumina HiSeq 2000 Sequencer. *Mycology* **2011**, *2*, 169–191.
- Mosher, J.J.; Bowman, B.; Bernberg, E.L.; Shevchenko, O.; Kan, J.; Korlach, J.; Kaplan, L.A. Improved Performance of the PacBio SMRT Technology for 16S rDNA Sequencing. *J. Microbiol. Methods* **2014**, *104*, 59–60. [[CrossRef](#)]
- Wenger, A.M.; Peluso, P.; Rowell, W.J.; Chang, P.C.; Hall, R.J.; Concepcion, G.T.; Ebler, J.; Functammasan, A.; Kolesnikov, A.; Olson, N.D.; et al. Accurate Circular Consensus Long-Read Sequencing Improves Variant Detection and Assembly of a Human Genome. *Nat. Biotechnol.* **2019**, *37*, 1155–1162. [[CrossRef](#)]
- Mikheyev, A.S.; Tin, M.M.Y. A First Look at the Oxford Nanopore MinION Sequencer. *Mol. Ecol. Resour.* **2014**, *14*, 1097–1102. [[CrossRef](#)] [[PubMed](#)]
- Greninger, A.L.; Naccache, S.N.; Federman, S.; Yu, G.; Mbala, P.; Bres, V.; Stryke, D.; Bouquet, J.; Somasekar, S.; Linnen, J.M.; et al. Rapid Metagenomic Identification of Viral Pathogens in Clinical Samples by Real-Time Nanopore Sequencing Analysis. *Genome Med.* **2015**, *7*, 99. [[CrossRef](#)]
- Pavese, V.; Cavalet-Giorsa, E.; Barchi, L.; Acquadro, A.; Marinoni, D.T.; Portis, E.; Lucas, S.J.; Botta, R. Whole-Genome Assembly of *Corylus Avellana* Cv “Tonda Gentile Delle Langhe” Using Linked-Reads (10× Genomics). *G3 Genes Genomes Genet.* **2021**, *11*, jkab152. [[CrossRef](#)] [[PubMed](#)]
- Cai, W.; Aburatani, H.; Stanton, V.P.; Housman, D.E.; Wang, Y.K.; Schwartz, D.C. Ordered Restriction Endonuclease Maps of Yeast Artificial Chromosomes Created by Optical Mapping on Surfaces. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 5164–5168. [[CrossRef](#)] [[PubMed](#)]
- Meng, X.; Benson, K.; Chada, K.; Huff, E.J.; Schwartz, D.C. Optical Mapping of Lambda Bacteriophage Clones Using Restriction Endonucleases. *Nat. Genet.* **1995**, *9*, 432–438. [[CrossRef](#)]
- De Carli, F.; Menezes, N.; Berrabah, W.; Barbe, V.; Genovesio, A.; Hyrien, O. High-Throughput Optical Mapping of Replicating DNA. *Small Methods* **2018**, *2*, 1800146. [[CrossRef](#)]
- Yang, H.; Garcia-Manero, G.; Montalban-Bravo, G.; Chien, K.S.; Kalia, A.; Tang, Z.; Wei, Y.; Nimmakayalu, M.; Rush, D.; Mallampati, S.; et al. High-Throughput Characterization of Cytogenomic Heterogeneity of MDS Using High-Resolution Optical Genome Mapping. *Blood* **2021**, *138*, 105. [[CrossRef](#)]
- Mikkelsen, T.S.; Ku, M.; Jaffe, D.B.; Issac, B.; Lieberman, E.; Giannoukos, G.; Alvarez, P.; Brockman, W.; Kim, T.K.; Koche, R.P.; et al. Genome-Wide Maps of Chromatin State in Pluripotent and Lineage-Committed Cells. *Nature* **2007**, *448*, 553–560. [[CrossRef](#)]
- Ley, T.J.; Mardis, E.R.; Ding, L.; Fulton, B.; McLellan, M.D.; Chen, K.; Dooling, D.; Dunford-Shore, B.H.; McGrath, S.; Hickenbotham, M.; et al. DNA Sequencing of a Cytogenetically Normal Acute Myeloid Leukaemia Genome. *Nature* **2008**, *456*, 66–72. [[CrossRef](#)]
- Licatalosi, D.D.; Mele, A.; Fak, J.J.; Ule, J.; Kayikci, M.; Chi, S.W.; Clark, T.A.; Schweitzer, A.C.; Blume, J.E.; Wang, X.; et al. HITS-CLIP Yields Genome-Wide Insights into Brain Alternative RNA Processing. *Nature* **2008**, *456*, 464–469. [[CrossRef](#)]
- Mortazavi, A.; Williams, B.A.; McCue, K.; Schaeffer, L.; Wold, B. Mapping and Quantifying Mammalian Transcriptomes by RNA-Seq. *Nat. Methods* **2008**, *5*, 621–628. [[CrossRef](#)] [[PubMed](#)]
- Baird, N.A.; Etter, P.D.; Atwood, T.S.; Currey, M.C.; Shiver, A.L.; Lewis, Z.A.; Selker, E.U.; Cresko, W.A.; Johnson, E.A. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE* **2008**, *3*, e3376. [[CrossRef](#)]

24. Heiman, M.; Schaefer, A.; Gong, S.; Peterson, J.D.; Day, M.; Ramsey, K.E.; Suárez-Fariñas, M.; Schwarz, C.; Stephan, D.A.; Surmeier, D.J.; et al. A Translational Profiling Approach for the Molecular Characterization of CNS Cell Types. *Cell* **2008**, *135*, 738–748. [[CrossRef](#)]
25. Core, L.J.; Waterfall, J.J.; Lis, J.T. Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science* **2008**, *322*, 1845–1848. [[CrossRef](#)] [[PubMed](#)]
26. Meissner, A.; Mikkelsen, T.S.; Gu, H.; Wernig, M.; Hanna, J.; Sivachenko, A.; Zhang, X.; Bernstein, B.E.; Nusbaum, C.; Jaffe, D.B.; et al. Genome-Scale DNA Methylation Maps of Pluripotent and Differentiated Cells. *Nature* **2008**, *454*, 766–770. [[CrossRef](#)] [[PubMed](#)]
27. Cokus, S.J.; Feng, S.; Zhang, X.; Chen, Z.; Merriman, B.; Haudenschild, C.D.; Pradhan, S.; Nelson, S.F.; Pellegrini, M.; Jacobsen, S.E. Shotgun Bisulphite Sequencing of the Arabidopsis Genome Reveals DNA Methylation Patterning. *Nature* **2008**, *452*, 215–219. [[CrossRef](#)]
28. German, M.A.; Pillay, M.; Jeong, D.H.; Hetawal, A.; Luo, S.; Janardhanan, P.; Kannan, V.; Rymarquis, L.A.; Nobuta, K.; German, R.; et al. Global Identification of MicroRNA-Target RNA Pairs by Parallel Analysis of RNA Ends. *Nat. Biotechnol.* **2008**, *26*, 941–946. [[CrossRef](#)]
29. Ng, S.B.; Turner, E.H.; Robertson, P.D.; Flygare, S.D.; Bigham, A.W.; Lee, C.; Shaffer, T.; Wong, M.; Bhattacharjee, A.; Eichler, E.E.; et al. Targeted Capture and Massively Parallel Sequencing of 12 Human Exomes. *Nature* **2009**, *461*, 272–276. [[CrossRef](#)]
30. Lister, R.; Pelizzola, M.; Downen, R.H.; Hawkins, R.D.; Hon, G.; Tonti-Filippini, J.; Nery, J.R.; Lee, L.; Ye, Z.; Ngo, Q.M.; et al. Human DNA Methylomes at Base Resolution Show Widespread Epigenomic Differences. *Nature* **2009**, *462*, 315–322. [[CrossRef](#)]
31. Deng, J.; Shoemaker, R.; Xie, B.; Gore, A.; Leproust, E.M.; Antosiewicz-Bourget, J.; Egli, D.; Maherali, N.; Park, I.H.; Yu, J.; et al. Targeted Bisulfite Sequencing Reveals Changes in DNA Methylation Associated with Nuclear Reprogramming. *Nat. Biotechnol.* **2009**, *27*, 353–360. [[CrossRef](#)] [[PubMed](#)]
32. Lieberman-Aiden, E.; Van Berkum, N.L.; Williams, L.; Imakaev, M.; Ragoczy, T.; Telling, A.; Amit, I.; Lajoie, B.R.; Sabo, P.J.; Dorschner, M.O.; et al. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **2009**, *326*, 289–293. [[CrossRef](#)]
33. Fullwood, M.J.; Liu, M.H.; Pan, Y.F.; Liu, J.; Xu, H.; Mohamed, Y.B.; Orlov, Y.L.; Velkov, S.; Ho, A.; Mei, P.H.; et al. An Oestrogen-Receptor- $\alpha$ -Bound Human Chromatin Interactome. *Nature* **2009**, *462*, 58–64. [[CrossRef](#)] [[PubMed](#)]
34. Ingolia, N.T.; Ghaemmaghami, S.; Newman, J.R.S.; Weissman, J.S. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science* **2009**, *324*, 218–223. [[CrossRef](#)]
35. Patwardhan, R.P.; Lee, C.; Litvin, O.; Young, D.L.; Pe’Er, D.; Shendure, J. High-Resolution Analysis of DNA Regulatory Elements by Synthetic Saturation Mutagenesis. *Nat. Biotechnol.* **2009**, *27*, 1173–1175. [[CrossRef](#)] [[PubMed](#)]
36. Pons, N.; Harris, E.Y.; Prudhomme, J.; Wick, I.; Eckhardt-Ludka, C.; Hicks, G.R.; Hardiman, G.; Lonardi, S.; Le Roch, K.G. Nucleosome Landscape and Control of Transcription in the Human Malaria Parasite. *Genome Res.* **2010**, *20*, 228–238. [[CrossRef](#)]
37. Mamanova, L.; Andrews, R.M.; James, K.D.; Sheridan, E.M.; Ellis, P.D.; Langford, C.F.; Ost, T.W.B.; Collins, J.E.; Turner, D.J. FRT-Seq: Amplification-Free, Strand-Specific Transcriptome Sequencing. *Nat. Methods* **2010**, *7*, 130–132. [[CrossRef](#)]
38. Kertesz, M.; Wan, Y.; Mazor, E.; Rinn, J.L.; Nutter, R.C.; Chang, H.Y.; Segal, E. Genome-Wide Measurement of RNA Secondary Structure in Yeast. *Nature* **2010**, *467*, 103–107. [[CrossRef](#)]
39. Fowler, D.M.; Araya, C.L.; Fleishman, S.J.; Kellogg, E.H.; Stephany, J.J.; Baker, D.; Fields, S. High-Resolution Mapping of Protein Sequence-Function Relationships. *Nat. Methods* **2010**, *7*, 741–746. [[CrossRef](#)]
40. Hansen, R.S.; Thomas, S.; Sandstrom, R.; Canfield, T.K.; Thurman, R.E.; Weaver, M.; Dorschner, M.O.; Gartler, S.M.; Stamatoyannopoulos, J.A. Sequencing Newly Replicated DNA Reveals Widespread Plasticity in Human Replication Timing. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 139–144. [[CrossRef](#)]
41. John, S.; Sabo, P.J.; Thurman, R.E.; Sung, M.H.; Biddie, S.C.; Johnson, T.A.; Hager, G.L.; Stamatoyannopoulos, J.A. Chromatin Accessibility Pre-Determines Glucocorticoid Receptor Binding Patterns. *Nat. Genet.* **2011**, *43*, 264–268. [[CrossRef](#)]
42. Churchman, L.S.; Weissman, J.S. Nascent Transcript Sequencing Visualizes Transcription at Nucleotide Resolution. *Nature* **2011**, *469*, 368–373. [[CrossRef](#)] [[PubMed](#)]
43. Logan, A.C.; Gao, H.; Wang, C.; Sahaf, B.; Jones, C.D.; Marshall, E.L.; Buño, I.; Armstrong, R.; Fire, A.Z.; Weinberg, K.I.; et al. High-Throughput VDJ Sequencing for Quantification of Minimal Residual Disease in Chronic Lymphocytic Leukemia and Immune Reconstitution Assessment. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 21194–21199. [[CrossRef](#)] [[PubMed](#)]
44. Carette, J.E.; Guimaraes, C.P.; Wuethrich, I.; Blomen, V.A.; Varadarajan, M.; Sun, C.; Bell, G.; Yuan, B.; Muellner, M.K.; Nijman, S.M.; et al. Global Gene Disruption in Human Cells to Assign Genes to Phenotypes by Deep Sequencing. *Nat. Biotechnol.* **2011**, *29*, 542–546. [[CrossRef](#)] [[PubMed](#)]
45. Khodor, Y.L.; Rodriguez, J.; Abruzzi, K.C.; Tang, C.H.A.; Marr, M.T.; Rosbash, M. Nascent-Seq Indicates Widespread Cotranscriptional Pre-mRNA Splicing in Drosophila. *Genes Dev.* **2011**, *25*, 2502–2512. [[CrossRef](#)] [[PubMed](#)]
46. Chu, C.; Qu, K.; Zhong, F.L.; Artandi, S.E.; Chang, H.Y. Genomic Maps of Long Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions. *Mol. Cell* **2011**, *44*, 667–678. [[CrossRef](#)]
47. Patwardhan, R.P.; Hiatt, J.B.; Witten, D.M.; Kim, M.J.; Smith, R.P.; May, D.; Lee, C.; Andrie, J.M.; Lee, S.I.; Cooper, G.M.; et al. Massively Parallel Functional Dissection of Mammalian Enhancers In Vivo. *Nat. Biotechnol.* **2012**, *30*, 265–270. [[CrossRef](#)] [[PubMed](#)]

48. Buenrostro, J.D.; Giresi, P.G.; Zaba, L.C.; Chang, H.Y.; Greenleaf, W.J. Transposition of Native Chromatin for Fast and Sensitive Epigenomic Profiling of Open Chromatin, DNA-Binding Proteins and Nucleosome Position. *Nat. Methods* **2013**, *10*, 1213–1218. [[CrossRef](#)]
49. Ding, Y.; Tang, Y.; Kwok, C.K.; Zhang, Y.; Bevilacqua, P.C.; Assmann, S.M. In Vivo Genome-Wide Profiling of RNA Secondary Structure Reveals Novel Regulatory Features. *Nature* **2014**, *505*, 696–700. [[CrossRef](#)]
50. Buenrostro, J.D.; Araya, C.L.; Chircus, L.M.; Layton, C.J.; Chang, H.Y.; Snyder, M.P.; Greenleaf, W.J. Quantitative Analysis of RNA-Protein Interactions on a Massively Parallel Array Reveals Biophysical and Evolutionary Landscapes. *Nat. Biotechnol.* **2014**, *32*, 562–568. [[CrossRef](#)]
51. Goodwin, S.; McPherson, J.D.; McCombie, W.R. Coming of Age: Ten Years of next-Generation Sequencing Technologies. *Nat. Rev. Genet.* **2016**, *17*, 333–351. [[CrossRef](#)] [[PubMed](#)]
52. Zeng, Y.; Wang, S.; Gao, S.; Soares, F.; Ahmed, M.; Guo, H.; Wang, M.; Hua, J.T.; Guan, J.; Moran, M.F.; et al. Refined RIP-Seq Protocol for Epitranscriptome Analysis with Low Input Materials. *PLoS Biol.* **2018**, *16*, e2006092. [[CrossRef](#)] [[PubMed](#)]
53. Pellicer, J.; Leitch, I.J. The Plant DNA C-Values Database (Release 7.1): An Updated Online Repository of Plant Genome Size Data for Comparative Studies. *New Phytol.* **2020**, *226*, 301–305. [[CrossRef](#)] [[PubMed](#)]
54. Arita, M.; Karsch-Mizrachi, I.; Cochrane, G. The International Nucleotide Sequence Database Collaboration. *Nucleic Acids Res.* **2021**, *49*, D121–D124. [[CrossRef](#)] [[PubMed](#)]
55. Doddamani, D.; Katta, M.A.V.S.K.; Khan, A.W.; Agarwal, G.; Shah, T.M.; Varshney, R.K. CicArMiSatDB: The Chickpea Microsatellite Database. *BMC Bioinform.* **2014**, *15*, 212. [[CrossRef](#)] [[PubMed](#)]
56. Doddamani, D.; Khan, A.W.; Katta, M.A.V.S.K.; Agarwal, G.; Thudi, M.; Ruperao, P.; Edwards, D.; Varshney, R.K. CicArVarDB: SNP and InDel Database for Advancing Genetics Research and Breeding Applications in Chickpea. *Database* **2015**, *2015*, bav078. [[CrossRef](#)]
57. Chen, F.; Dong, W.; Zhang, J.; Guo, X.; Chen, J.; Wang, Z.; Lin, Z.; Tang, H.; Zhang, L. The Sequenced Angiosperm Genomes and Genome Databases. *Front. Plant Sci.* **2018**, *9*, 418. [[CrossRef](#)]
58. Chen, F.; Song, Y.; Li, X.; Chen, J.; Mo, L.; Zhang, X.; Lin, Z.; Zhang, L. Genome Sequences of Horticultural Plants: Past, Present, and Future. *Hortic. Res.* **2019**, *6*, 112. [[CrossRef](#)]
59. Chu, Q.; Zhang, X.; Zhu, X.; Liu, C.; Mao, L.; Ye, C.; Zhu, Q.H.; Fan, L. PlantcircBase: A Database for Plant Circular RNAs. *Mol. Plant* **2017**, *10*, 1126–1128. [[CrossRef](#)]
60. Iversen, C.M.; McCormack, M.L.; Powell, A.S.; Blackwood, C.B.; Freschet, G.T.; Kattge, J.; Roumet, C.; Stover, D.B.; Soudzilovskaia, N.A.; Valverde-Barrantes, O.J.; et al. A Global Fine-Root Ecology Database to Address below-Ground Challenges in Plant Ecology. *New Phytol.* **2017**, *215*, 15–26. [[CrossRef](#)]
61. Obayashi, T.; Aoki, Y.; Tadaka, S.; Kagaya, Y.; Kinoshita, K. ATTED-II in 2018: A Plant Coexpression Database Based on Investigation of the Statistical Property of the Mutual Rank Index. *Plant Cell Physiol.* **2018**, *59*, e3. [[CrossRef](#)] [[PubMed](#)]
62. Cooper, L.; Meier, A.; Laporte, M.A.; Elser, J.L.; Mungall, C.; Sinn, B.T.; Cavaliere, D.; Carbon, S.; Dunn, N.A.; Smith, B.; et al. The Planteome Database: An Integrated Resource for Reference Ontologies, Plant Genomics and Phenomics. *Nucleic Acids Res.* **2018**, *46*, D1168–D1180. [[CrossRef](#)]
63. Wild, J.; Kaplan, Z.; Danihelka, J.; Petřík, P.; Chytrý, M.; Novotný, P.; Rohn, M.; Šulc, V.; Brůna, J.; Chobot, K.; et al. Plant Distribution Data for the Czech Republic Integrated in the Pladias Database. *Preslia* **2019**, *91*, 1–24. [[CrossRef](#)]
64. Kattge, J.; Bönnisch, G.; Díaz, S.; Lavorel, S.; Prentice, I.C.; Leadley, P.; Tautenhahn, S.; Werner, G.D.A.; Aakala, T.; Abedi, M.; et al. TRY Plant Trait Database—Enhanced Coverage and Open Access. *Glob. Chang. Biol.* **2020**, *26*, 119–188. [[CrossRef](#)]
65. Guo, Z.; Kuang, Z.; Wang, Y.; Zhao, Y.; Tao, Y.; Cheng, C.; Yang, J.; Lu, X.; Hao, C.; Wang, T.; et al. PmiREN: A Comprehensive Encyclopedia of Plant MiRNAs. *Nucleic Acids Res.* **2020**, *48*, D1114–D1121. [[CrossRef](#)]
66. Das, D.; Jaiswal, M.; Khan, F.N.; Ahamad, S.; Kumar, S. PlantPepDB: A Manually Curated Plant Peptide Database. *Sci. Rep.* **2020**, *10*, 2194. [[CrossRef](#)]
67. Boschiero, C.; Dai, X.; Lundquist, P.K.; Roy, S.; de Bang, T.C.; Zhang, S.; Zhuang, Z.; Torres-Jerez, I.; Udvardi, M.K.; Scheible, W.R.; et al. MtSSPDB: The Medicago Truncatula Small Secreted Peptide Database. *Plant Physiol.* **2020**, *183*, 399–413. [[CrossRef](#)]
68. Guerrero-Ramírez, N.R.; Mommer, L.; Freschet, G.T.; Iversen, C.M.; McCormack, M.L.; Kattge, J.; Poorter, H.; van der Plas, F.; Bergmann, J.; Kuyper, T.W.; et al. Global Root Traits (GRooT) Database. *Glob. Ecol. Biogeogr.* **2021**, *30*, 25–37. [[CrossRef](#)]
69. Hussain, N.; Chanda, R.; Abir, R.A.; Mou, M.A.; Hasan, M.K.; Ashraf, M.A. MPDB 2.0: A Large Scale and Integrated Medicinal Plant Database of Bangladesh. *BMC Res. Notes* **2021**, *14*, 301. [[CrossRef](#)]
70. Valentin, G.; Abdel, T.; Gaëtan, D.; Jean-François, D.; Matthieu, C.; Mathieu, R. GreenPhylDB v5: A Comparative Pangenomic Database for Plant Genomes. *Nucleic Acids Res.* **2021**, *49*, D1464–D1471. [[CrossRef](#)]
71. Chen, H.; Yin, X.; Guo, L.; Yao, J.; Ding, Y.; Xu, X.; Liu, L.; Zhu, Q.H.; Chu, Q.; Fan, L. PlantscRNadb: A Database for Plant Single-Cell RNA Analysis. *Mol. Plant* **2021**, *14*, 855–857. [[CrossRef](#)] [[PubMed](#)]
72. Liu, J.; Liu, X.; Zhang, S.; Liang, S.; Luan, W.; Ma, X. TarDB: An Online Database for Plant MiRNA Targets and MiRNA-Triggered Phased siRNAs. *BMC Genom.* **2021**, *22*, 348. [[CrossRef](#)]
73. Delbianco, A.; Gibin, D.; Pasinato, L.; Morelli, M. Update of the Xylella Spp. Host Plant Database—Systematic Literature Search up to 30 June 2021. *EFSA J.* **2022**, *20*, e07039. [[CrossRef](#)] [[PubMed](#)]



74. Ma, X.; Yan, H.; Yang, J.; Liu, Y.; Li, Z.; Sheng, M.; Cao, Y.; Yu, X.; Yi, X.; Xu, W.; et al. PlantGSAD: A Comprehensive Gene Set Annotation Database for Plant Species. *Nucleic Acids Res.* **2022**, *50*, D1456–D1467. [[CrossRef](#)]
75. Plant Chloroplast Database. Available online: <http://www.gndu.ac.in/CpGDB> (accessed on 25 July 2022).
76. Plant Protein, DNA, RNA, Pathway and Expression Database. Available online: <https://www.habdisk.org/dbpr.php> (accessed on 25 July 2022).
77. tRNA-Derived Non-Coding RNAs Database. Available online: <https://nipgr.ac.in/PtncRNADB> (accessed on 25 July 2022).
78. Wang, W.; Mauleon, R.; Hu, Z.; Chebotarov, D.; Tai, S.; Wu, Z.; Li, M.; Zheng, T.; Fuentes, R.R.; Zhang, F.; et al. Genomic Variation in 3,010 Diverse Accessions of Asian Cultivated Rice. *Nature* **2018**, *557*, 43–49. [[CrossRef](#)] [[PubMed](#)]
79. Ballouz, S.; Dobin, A.; Gillis, J.A. Is It Time to Change the Reference Genome? *Genome Biol.* **2019**, *20*, 159. [[CrossRef](#)]
80. Varshney, R.K.; Sinha, P.; Singh, V.K.; Kumar, A.; Zhang, Q.; Bennetzen, J.L. 5Gs for Crop Genetic Improvement. *Curr. Opin. Plant Biol.* **2020**, *56*, 190–196. [[CrossRef](#)]
81. Dumschott, K.; Schmidt, M.H.W.; Chawla, H.S.; Snowdon, R.; Usadel, B. Oxford Nanopore Sequencing: New Opportunities for Plant Genomics? *J. Exp. Bot.* **2020**, *71*, 5313–5322. [[CrossRef](#)]
82. Belser, C.; Istace, B.; Denis, E.; Dubarry, M.; Baurens, F.C.; Falentin, C.; Genete, M.; Berrabah, W.; Chèvre, A.M.; Delourme, R.; et al. Chromosome-Scale Assemblies of Plant Genomes Using Nanopore Long Reads and Optical Maps. *Nat. Plants* **2018**, *4*, 879–887. [[CrossRef](#)]
83. Tang, D.; Jia, Y.; Zhang, J.; Li, H.; Cheng, L.; Wang, P.; Bao, Z.; Liu, Z.; Feng, S.; Zhu, X.; et al. Genome Evolution and Diversity of Wild and Cultivated Potatoes. *Nature* **2022**, *606*, 535–541. [[CrossRef](#)]
84. Jain, M.; Misra, G.; Patel, R.K.; Priya, P.; Jhanwar, S.; Khan, A.W.; Shah, N.; Singh, V.K.; Garg, R.; Jeena, G.; et al. A Draft Genome Sequence of the Pulse Crop Chickpea (*Cicer arietinum* L.). *Plant J.* **2013**, *74*, 715–729. [[CrossRef](#)] [[PubMed](#)]
85. Parween, S.; Nawaz, K.; Roy, R.; Pole, A.K.; Venkata Suresh, B.; Misra, G.; Jain, M.; Yadav, G.; Parida, S.K.; Tyagi, A.K.; et al. An Advanced Draft Genome Assembly of a Desi Type Chickpea (*Cicer arietinum* L.). *Sci. Rep.* **2015**, *5*, 12806. [[CrossRef](#)]
86. Wang, H.; Yang, J.; Zhang, Y.; Qian, J.; Wang, J. Reconstruct High-Resolution 3D Genome Structures for Diverse Cell-Types Using FLAMINGO. *Nat. Commun.* **2022**, *13*, 2645. [[CrossRef](#)] [[PubMed](#)]
87. Alonge, M.; Shumate, A.; Puiu, D.; Zimin, A.V.; Salzberg, S.L. Chromosome-Scale Assembly of the Bread Wheat Genome Reveals Thousands of Additional Gene Copies. *Genetics* **2020**, *216*, 599–608. [[CrossRef](#)]
88. Zhang, S.V.; Zhuo, L.; Hahn, M.W. AGOUTI: Improving Genome Assembly and Annotation Using Transcriptome Data. *Gigascience* **2016**, *5*, 31. [[CrossRef](#)]
89. Mamidi, S.; Healey, A.; Huang, P.; Grimwood, J.; Jenkins, J.; Barry, K.; Sreedasyam, A.; Shu, S.; Lovell, J.T.; Feldman, M.; et al. A Genome Resource for Green Millet *Setaria Viridis* Enables Discovery of Agronomically Valuable Loci. *Nat. Biotechnol.* **2020**, *38*, 1203–1210. [[CrossRef](#)]
90. Zhou, Y.; Zhao, X.; Li, Y.; Xu, J.; Bi, A.; Kang, L.; Xu, D.; Chen, H.; Wang, Y.; Wang, Y.G.; et al. Triticum Population Sequencing Provides Insights into Wheat Adaptation. *Nat. Genet.* **2020**, *52*, 1412–1422. [[CrossRef](#)]
91. Zhu, H.Z.; Zhang, Z.F.; Zhou, N.; Jiang, C.Y.; Wang, B.J.; Cai, L.; Wang, H.M.; Liua, S.J. Bacteria and Metabolic Potential in Karst Caves Revealed by Intensive Bacterial Cultivation and Genome Assembly. *Appl. Environ. Microbiol.* **2021**, *87*, e02440-20. [[CrossRef](#)]
92. Kille, B.; Balaji, A.; Sedlazeck, F.J.; Nute, M.; Treangen, T.J. Multiple Genome Alignment in the Telomere-to-Telomere Assembly Era. *Genome Biol.* **2022**, *23*, 182. [[CrossRef](#)]
93. Zhang, F.; Xue, H.; Dong, X.; Li, M.; Zheng, X.; Li, Z.; Xu, J.; Wang, W.; Wei, C. Long-Read Sequencing of 111 Rice Genomes Reveals Significantly Larger Pan-Genomes. *Genome Res.* **2022**, *32*, 853–863. [[CrossRef](#)]
94. Mussurova, S.; Al-Bader, N.; Zuccolo, A.; Wing, R.A. Potential of Platinum Standard Reference Genomes to Exploit Natural Variation in the Wild Relatives of Rice. *Front. Plant Sci.* **2020**, *11*, 579980. [[CrossRef](#)] [[PubMed](#)]
95. Kaul, S.; Koo, H.L.; Jenkins, J.; Rizzo, M.; Rooney, T.; Tallon, L.J.; Feldblyum, T.; Nierman, W.; Benito, M.I.; Lin, X.; et al. Analysis of the Genome Sequence of the Flowering Plant *Arabidopsis thaliana*. *Nature* **2000**, *408*, 796–815. [[CrossRef](#)]
96. Matsumoto, T.; Wu, J.; Kanamori, H.; Katayose, Y.; Fujisawa, M.; Namiki, N.; Mizuno, H.; Yamamoto, K.; Antonio, B.A.; Baba, T.; et al. The Map-Based Sequence of the Rice Genome. *Nature* **2005**, *436*, 793–800. [[CrossRef](#)]
97. Michael, T.P.; VanBuren, R. Building Near-Complete Plant Genomes. *Curr. Opin. Plant Biol.* **2020**, *54*, 26–33. [[CrossRef](#)]
98. Berlin, K.; Koren, S.; Chin, C.-S.; Drake, J.P.; Landolin, J.M.; Phillippy, A.M. Erratum: Corrigendum: Assembling Large Genomes with Single-Molecule Sequencing and Locality-Sensitive Hashing. *Nat. Biotechnol.* **2015**, *33*, 1109. [[CrossRef](#)] [[PubMed](#)]
99. Vanburen, R.; Bryant, D.; Edger, P.P.; 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. [[CrossRef](#)] [[PubMed](#)]
100. Loman, N.J.; Quick, J.; Simpson, J.T. A Complete Bacterial Genome Assembled de Novo Using Only Nanopore Sequencing Data. *Nat. Methods* **2015**, *12*, 733–735. [[CrossRef](#)]
101. McCoy, R.C.; Taylor, R.W.; Blauwkamp, T.A.; Kelley, J.L.; Kertesz, M.; Pushkarev, D.; Petrov, D.A.; Fiston-Lavier, A.S. Illumina TruSeq Synthetic Long-Reads Empower de Novo Assembly and Resolve Complex, Highly-Repetitive Transposable Elements. *PLoS ONE* **2014**, *9*, e106689. [[CrossRef](#)]
102. Colle, M.; Leisner, C.P.; Wai, C.M.; Ou, S.; Bird, K.A.; Wang, J.; Wisecaver, J.H.; Yocca, A.E.; Alger, E.I.; Tang, H.; et al. Haplotype-Phased Genome and Evolution of Phytonutrient Pathways of Tetraploid Blueberry. *Gigascience* **2019**, *8*, giz012. [[CrossRef](#)]

103. Zhang, L.M.; Leng, C.Y.; Luo, H.; Wu, X.Y.; Liu, Z.Q.; Zhang, Y.M.; Zhang, H.; Xia, Y.; Shang, L.; Liu, C.M.; et al. Sweet Sorghum Originated through Selection of Dry, a Plant-Specific Nac Transcription Factor Gene. *Plant Cell* **2018**, *30*, 2286–2307. [[CrossRef](#)]
104. Bertioli, D.J.; Jenkins, J.; Clevenger, J.; Dudchenko, O.; Gao, D.; Seijo, G.; Leal-Bertioli, S.C.M.; Ren, L.; Farmer, A.D.; Pandey, M.K.; et al. The Genome Sequence of Segmental Allotetraploid Peanut *Arachis hypogaea*. *Nat. Genet.* **2019**, *51*, 877–884. [[CrossRef](#)]
105. VanBuren, R.; Wai, C.M.; Pardo, J.; Yocca, A.E.; Wang, X.; Wang, H.; Chaluvadi, S.R.; Bryant, D.; Edger, P.P.; Bennetzen, J.L.; et al. Exceptional Subgenome Stability and Functional Divergence in Allotetraploid Teff, the Primary Cereal Crop in Ethiopia. *bioRxiv* **2019**, 580720. [[CrossRef](#)]
106. Lam, E.T.; Hastie, A.; Lin, C.; Ehrlich, D.; Das, S.K.; Austin, M.D.; Deshpande, P.; Cao, H.; Nagarajan, N.; Xiao, M.; et al. Genome Mapping on Nanochannel Arrays for Structural Variation Analysis and Sequence Assembly. *Nat. Biotechnol.* **2012**, *30*, 771–776. [[CrossRef](#)]
107. Miga, K.H.; Koren, S.; Rhie, A.; Vollger, M.R.; Gershman, A.; Bzikadze, A.; Brooks, S.; Howe, E.; Porubsky, D.; Logsdon, G.A.; et al. Telomere-to-Telomere Assembly of a Complete Human X Chromosome. *Nature* **2020**, *585*, 79–84. [[CrossRef](#)] [[PubMed](#)]
108. Naish, M.; Alonge, M.; Wlodzimierz, P.; Tock, A.J.; Abramson, B.W.; Schmücker, A.; Mandáková, T.; Jamge, B.; Lambing, C.; Kuo, P.; et al. The Genetic and Epigenetic Landscape of the Arabidopsis Centromeres. *Science* **2021**, *374*, eabi7489. [[CrossRef](#)] [[PubMed](#)]
109. Wang, B.; Yang, X.; Jia, Y.; Xu, Y.; Jia, P.; Dang, N.; Wang, S.; Xu, T.; Zhao, X.; Gao, S.; et al. High-Quality Arabidopsis Thaliana Genome Assembly with Nanopore and HiFi Long Reads. *Genom. Proteom. Bioinform.* **2022**, *20*, 4–13. [[CrossRef](#)]
110. Song, J.M.; Xie, W.Z.; Wang, S.; Guo, Y.X.; Koo, D.H.; Kudrna, D.; Gong, C.; Huang, Y.; Feng, J.W.; Zhang, W.; et al. Two Gap-Free Reference Genomes and a Global View of the Centromere Architecture in Rice. *Mol. Plant* **2021**, *14*, 1757–1767. [[CrossRef](#)]
111. Belser, C.; Baurens, F.-C.; Noel, B.; Martin, G.; Cruaud, C.; Istace, B.; Yahiaoui, N.; Labadie, K.; Hřibová, E.; Doležel, J.; et al. Telomere-to-Telomere Gapless Chromosomes of Banana Using Nanopore Sequencing. *Commun. Biol.* **2021**, *4*, 1047. [[CrossRef](#)]
112. Jiao, W.B.; Accinelli, G.G.; Hartwig, B.; Kiefer, C.; Baker, D.; Severing, E.; Willing, E.M.; Piednoel, M.; Woetzel, S.; Madrid-Herrero, E.; et al. Improving and Correcting the Contiguity of Long-Read Genome Assemblies of Three Plant Species Using Optical Mapping and Chromosome Conformation Capture Data. *Genome Res.* **2017**, *27*, 778–786. [[CrossRef](#)]
113. Zhao, J.; Bayer, P.E.; Ruperao, P.; Saxena, R.K.; Khan, A.W.; Golicz, A.A.; Nguyen, H.T.; Batley, J.; Edwards, D.; Varshney, R.K. Trait Associations in the Pangenome of Pigeon Pea (*Caajanus cajan*). *Plant Biotechnol. J.* **2020**, *18*, 1946–1954. [[CrossRef](#)]
114. Huang, X.; Kurata, N.; Wei, X.; Wang, Z.X.; Wang, A.; Zhao, Q.; Zhao, Y.; Liu, K.; Lu, H.; Li, W.; et al. A Map of Rice Genome Variation Reveals the Origin of Cultivated Rice. *Nature* **2012**, *490*, 497–501. [[CrossRef](#)] [[PubMed](#)]
115. Montenegro, J.D.; Golicz, A.A.; Bayer, P.E.; Hurgobin, B.; Lee, H.T.; Chan, C.K.K.; Visendi, P.; Lai, K.; Doležel, J.; Batley, J.; et al. The Pangenome of Hexaploid Bread Wheat. *Plant J.* **2017**, *90*, 1007–1013. [[CrossRef](#)]
116. Gao, L.; Gonda, I.; Sun, H.; Ma, Q.; Bao, K.; Tieman, D.M.; Burzynski-Chang, E.A.; Fish, T.L.; Stromberg, K.A.; Sacks, G.L.; et al. The Tomato Pan-Genome Uncovers New Genes and a Rare Allele Regulating Fruit Flavor. *Nat. Genet.* **2019**, *51*, 1044–1051. [[CrossRef](#)]
117. Kou, Y.; Liao, Y.; Toivainen, T.; Lv, Y.; Tian, X.; Emerson, J.J.; Gaut, B.S.; Zhou, Y. Evolutionary Genomics of Structural Variation in Asian Rice (*Oryza sativa*) Domestication. *Mol. Biol. Evol.* **2020**, *37*, 3507–3524. [[CrossRef](#)]
118. Li, H.; Wang, S.; Chai, S.; Yang, Z.; Zhang, Q.; Xin, H.; Xu, Y.; Lin, S.; Chen, X.; Yao, Z.; et al. Graph-Based Pan-Genome Reveals Structural and Sequence Variations Related to Agronomic Traits and Domestication in Cucumber. *Nat. Commun.* **2022**, *13*, 682. [[CrossRef](#)] [[PubMed](#)]
119. Marks, R.A.; Hotaling, S.; Frandsen, P.B.; VanBuren, R. Representation and Participation across 20 Years of Plant Genome Sequencing. *Nat. Plants* **2021**, *7*, 1571–1578. [[CrossRef](#)] [[PubMed](#)]
120. Myers, E.W. The Fragment Assembly String Graph. *Bioinformatics* **2005**, *21*, ii79–ii85. [[CrossRef](#)]
121. Idury, R.M.; Waterman, M.S. A New Algorithm for DNA Sequence Assembly. *J. Comput. Biol.* **1995**, *2*, 291–306. [[CrossRef](#)] [[PubMed](#)]
122. Pevzner, P.A.; Tang, H.; Waterman, M.S. An Eulerian Path Approach to DNA Fragment Assembly. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 9748–9753. [[CrossRef](#)]
123. Bankevich, A.; Bzikadze, A.V.; Kolmogorov, M.; Antipov, D.; Pevzner, P.A. Multiplex de Bruijn Graphs Enable Genome Assembly from Long, High-Fidelity Reads. *Nat. Biotechnol.* **2022**, *40*, 1075–1081. [[CrossRef](#)]
124. Sadasivan, H.; Maric, M.; Dawson, E.; Iyer, V.; Israeli, J.; Narayanasamy, S. Accelerating Minimap2 for Accurate Long Read Alignment on GPUs. *bioRxiv* **2022**, *6*, 13–23. [[CrossRef](#)]
125. Goff, S.A.; Ricke, D.; Lan, T.H.; Presting, G.; Wang, R.; Dunn, M.; Glazebrook, J.; Sessions, A.; Oeller, P.; Varma, H.; et al. A Draft Sequence of the Rice Genome (*Oryza sativa* L. Ssp. japonica). *Science* **2002**, *296*, 92–100. [[CrossRef](#)]
126. Yu, J.; Hu, S.; Wang, J.; Wong, G.K.S.; Li, S.; Liu, B.; Deng, Y.; Dai, L.; Zhou, Y.; Zhang, X.; et al. A Draft Sequence of the Rice Genome (*Oryza sativa* L. Ssp. indica). *Science* **2002**, *296*, 79–92. [[CrossRef](#)]
127. Tuskan, G.A.; DiFazio, S.; Jansson, S.; Bohlmann, J.; Grigoriev, I.; Hellsten, U.; Putnam, M.; Ralph, S.; Rombauts, S.; Salamov, A.; et al. The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **2006**, *313*, 1596–1604. [[CrossRef](#)]



128. Jaillon, O.; Aury, J.M.; Noel, B.; Policriti, A.; Clepet, C.; Casagrande, A.; Choisne, N.; Aubourg, S.; Vitulo, N.; Jubin, C.; et al. The Grapevine Genome Sequence Suggests Ancestral Hexaploidization in Major Angiosperm Phyla. *Nature* **2007**, *449*, 463–467. [[CrossRef](#)]
129. Velasco, R.; Zharkikh, A.; Troglio, M.; Cartwright, D.A.; Cestaro, A.; Pruss, D.; Pindo, M.; FitzGerald, L.M.; 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. [[CrossRef](#)]
130. Huang, S.; Li, R.; Zhang, Z.; Li, L.; Gu, X.; Fan, W.; Lucas, W.J.; Wang, X.; Xie, B.; Ni, P.; et al. The Genome of the Cucumber, *Cucumis sativus* L. *Nat. Genet.* **2009**, *41*, 1275–1281. [[CrossRef](#)]
131. Velasco, R.; Zharkikh, A.; Affourtit, J.; Dhingra, A.; Cestaro, A.; Kalyanaraman, A.; Fontana, P.; Bhatnagar, S.K.; Troglio, M.; Pruss, D.; et al. The Genome of the Domesticated Apple (*Malus × Domestica* Borkh.). *Nat. Genet.* **2010**, *42*, 833–839. [[CrossRef](#)]
132. Argout, X.; Salse, J.; Aury, J.M.; Guiltinan, M.J.; Droc, G.; Gouzy, J.; Allegre, M.; Chaparro, C.; Legavre, T.; Maximova, S.N.; et al. The Genome of Theobroma Cacao. *Nat. Genet.* **2011**, *43*, 101–108. [[CrossRef](#)]
133. Garcia-Mas, J.; Benjak, A.; Sanseverino, W.; Bourgeois, M.; Mir, G.; González, V.M.; Heñaff, E.; Camâra, F.; Cozzuto, L.; Lowy, E.; et al. The Genome of Melon (*Cucumis melo* L.). *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 11872–11877. [[CrossRef](#)]
134. Shulaev, V.; Sargent, D.J.; Crowhurst, R.N.; Mockler, T.C.; Folkerts, O.; Delcher, A.L.; Jaiswal, P.; Mockaitis, K.; Liston, A.; Mane, S.P.; et al. The Genome of Woodland Strawberry (*Fragaria vesca*). *Nat. Genet.* **2011**, *43*, 109–116. [[CrossRef](#)]
135. Wang, X.; Wang, H.; Wang, J.; Sun, R.; Wu, J.; Liu, S.; Bai, Y.; Mun, J.H.; Bancroft, I.; Cheng, F.; et al. The Genome of the Mesopolyploid Crop Species Brassica Rapa. *Nat. Genet.* **2011**, *43*, 1035–1039. [[CrossRef](#)] [[PubMed](#)]
136. Xu, X.; Pan, S.; Cheng, S.; Zhang, B.; Mu, D.; Ni, P.; Zhang, G.; Yang, S.; Li, R.; Wang, J.; et al. Genome Sequence and Analysis of the Tuber Crop Potato. *Nature* **2011**, *475*, 189–195. [[CrossRef](#)]
137. Varshney, R.K.; Song, C.; Saxena, R.K.; Azam, S.; Yu, S.; Sharpe, A.G.; Cannon, S.; Baek, J.; Rosen, B.D.; Tar'an, B.; et al. Draft Genome Sequence of Chickpea (*Cicer arietinum*) Provides a Resource for Trait Improvement. *Nat. Biotechnol.* **2013**, *31*, 240–246. [[CrossRef](#)]
138. Varshney, R.K.; Chen, W.; Li, Y.; Bharti, A.K.; Saxena, R.K.; Schlueter, J.A.; Donoghue, M.T.A.; Azam, S.; Fan, G.; Whaley, A.M.; et al. Draft Genome Sequence of Pigeonpea (*Cajanus cajan*), an Orphan Legume Crop of Resource-Poor Farmers. *Nat. Biotechnol.* **2012**, *30*, 83–89. [[CrossRef](#)]
139. Xu, Y.; Wang, J.; Guo, S.; Zhang, J.; Sun, H.; Salse, J.; Lucas, W.J.; Zhang, H.; Zheng, Y.; Mao, L.; et al. The Draft Genome of Watermelon (*Citrullus lanatus*) and Resequencing of 20 Diverse Accessions. *Nat. Genet.* **2013**, *45*, 51–58. [[CrossRef](#)]
140. Zimin, A.V.; Puiu, D.; Hall, R.; Kingan, S.; Clavijo, B.J.; Salzberg, S.L. The First Near-Complete Assembly of the Hexaploid Bread Wheat Genome, *Triticum aestivum*. *Gigascience* **2017**, *6*, gix097. [[CrossRef](#)] [[PubMed](#)]
141. Sato, K.; Abe, F.; Mascher, M.; Haberer, G.; Gundlach, H.; Spannagl, M.; Shirasawa, K.; Isobe, S. Chromosome-Scale Genome Assembly of the Transformation-Amenable Common Wheat Cultivar ‘Fielder’. *DNA Res.* **2021**, *28*, dsab008. [[CrossRef](#)]
142. Mascher, M.; Gundlach, H.; Himmelbach, A.; Beier, S.; Twardziok, S.O.; Wicker, T.; Radchuk, V.; Dockter, C.; Hedley, P.E.; Russell, J.; et al. A Chromosome Conformation Capture Ordered Sequence of the Barley Genome. *Nature* **2017**, *544*, 427–433. [[CrossRef](#)]
143. Li, G.; Wang, L.; Yang, J.; He, H.; Jin, H.; Li, X.; Ren, T.; Ren, Z.; Li, F.; Han, X.; et al. A High-Quality Genome Assembly Highlights Rye Genomic Characteristics and Agronomically Important Genes. *Nat. Genet.* **2021**, *53*, 574–584. [[CrossRef](#)] [[PubMed](#)]
144. Zhang, X.; Chen, S.; Shi, L.; Gong, D.; Zhang, S.; Zhao, Q.; Zhan, D.; Vasseur, L.; Wang, Y.; Yu, J.; et al. Haplotype-Resolved Genome Assembly Provides Insights into Evolutionary History of the Tea Plant *Camellia sinensis*. *Nat. Genet.* **2021**, *53*, 1250–1259. [[CrossRef](#)] [[PubMed](#)]
145. Paterson, A.H.; Bowers, J.E.; Bruggmann, R.; Dubchak, I.; Grimwood, J.; Gundlach, H.; Haberer, G.; Hellsten, U.; Mitros, T.; Poliakov, A.; et al. The Sorghum Bicolor Genome and the Diversification of Grasses. *Nature* **2009**, *457*, 551–556. [[CrossRef](#)] [[PubMed](#)]
146. Deschamps, S.; Zhang, Y.; Llaca, V.; Ye, L.; Sanyal, A.; King, M.; May, G.; Lin, H. A Chromosome-Scale Assembly of the Sorghum Genome Using Nanopore Sequencing and Optical Mapping. *Nat. Commun.* **2018**, *9*, 4844. [[CrossRef](#)]
147. Martis, M.M.; Zhou, R.; Haseneyer, G.; Schmutzer, T.; Vrána, J.; Kubaláková, M.; König, S.; Kugler, K.G.; Scholz, U.; Hackauf, B.; et al. Reticulate Evolution of the Rye Genome. *Plant Cell* **2013**, *25*, 3685–3698. [[CrossRef](#)]
148. Bauer, E.; Schmutzer, T.; Barilar, I.; Mascher, M.; Gundlach, H.; Martis, M.M.; Twardziok, S.O.; Hackauf, B.; Gordillo, A.; Wilde, P.; et al. Towards a Whole-Genome Sequence for Rye (*Secale cereale* L.). *Plant J.* **2017**, *89*, 853–869. [[CrossRef](#)]
149. Rabanus-Wallace, M.T.; Hackauf, B.; Mascher, M.; Lux, T.; Wicker, T.; Gundlach, H.; Baez, M.; Houben, A.; Mayer, K.F.X.; Guo, L.; et al. Chromosome-Scale Genome Assembly Provides Insights into Rye Biology, Evolution and Agronomic Potential. *Nat. Genet.* **2021**, *53*, 564–573. [[CrossRef](#)]
150. Freire, R.; Weisweiler, M.; Guerreiro, R.; Baig, N.; Hüttel, B.; Obeng-Hinne, E.; Renner, J.; Hartje, S.; Muders, K.; Truberg, B.; et al. Chromosome-Scale Reference Genome Assembly of a Diploid Potato Clone Derived from an Elite Variety. *G3 Genes Genomes Genet.* **2021**, *11*, jkab330. [[CrossRef](#)]
151. Bertoli, D.J.; Cannon, S.B.; Froenicke, L.; Huang, G.; Farmer, A.D.; Cannon, E.K.S.; Liu, X.; Gao, D.; Clevenger, J.; Dash, S.; et al. The Genome Sequences of *Arachis duranensis* and *Arachis ipaensis*, the Diploid Ancestors of Cultivated Peanut. *Nat. Genet.* **2016**, *48*, 438–446. [[CrossRef](#)] [[PubMed](#)]

152. Xu, X.; Liu, X.; Ge, S.; Jensen, J.D.; Hu, F.; Li, X.; Dong, Y.; Gutenkunst, R.N.; Fang, L.; Huang, L.; et al. Resequencing 50 Accessions of Cultivated and Wild Rice Yields Markers for Identifying Agronomically Important Genes. *Nat. Biotechnol.* **2012**, *30*, 105–111. [[CrossRef](#)]
153. Bayer, P.E.; Golicz, A.A.; Scheben, A.; Batley, J.; Edwards, D. Plant Pan-Genomes Are the New Reference. *Nat. Plants* **2020**, *6*, 914–920. [[CrossRef](#)]
154. Rijzaani, H.; Bayer, P.E.; Rouard, M.; Doležel, J.; Batley, J.; Edwards, D. The Pangenome of Banana Highlights Differences between Genera and Genomes. *Plant Genome* **2022**, *15*, e20100. [[CrossRef](#)] [[PubMed](#)]
155. Hufnagel, B.; Soriano, A.; Taylor, J.; Divol, F.; Kroc, M.; Sanders, H.; Yeheyis, L.; Nelson, M.; Péret, B. Pangenome of White Lupin Provides Insights into the Diversity of the Species. *Plant Biotechnol. J.* **2021**, *19*, 2532–2543. [[CrossRef](#)] [[PubMed](#)]
156. Kamal, N.; Lux, T.; Jayakodi, M.; Haberer, G.; Gundlach, H.; Mayer, K.F.X.; Mascher, M.; Spannagl, M. The Barley and Wheat Pan-Genomes. In *Methods in Molecular Biology*; Springer: Berlin/Heidelberg, Germany, 2022; Volume 2443.
157. Bayer, P.E.; Petereit, J.; Durant, É.; Monat, C.; Rouard, M.; Hu, H.; Chapman, B.; Li, C.; Cheng, S.; Batley, J.; et al. Wheat Panache: A Pangenome Graph Database Representing Presence–Absence Variation across Sixteen Bread Wheat Genomes. *Plant Genome* **2022**, *15*, e20221. [[CrossRef](#)]
158. Ruperao, P.; Thirunavukkarasu, N.; Gandham, P.; Selvanayagam, S.; Govindaraj, M.; Nebie, B.; Manyasa, E.; Gupta, R.; Das, R.R.; Odeny, D.A.; et al. Sorghum Pan-Genome Explores the Functional Utility for Genomic-Assisted Breeding to Accelerate the Genetic Gain. *Front. Plant Sci.* **2021**, *12*, 666342. [[CrossRef](#)] [[PubMed](#)]
159. Lin, K.; Zhang, N.; Severing, E.I.; Nijveen, H.; Cheng, F.; Visser, R.G.F.; Wang, X.; de Ridder, D.; Bonnema, G. Beyond Genomic Variation—Comparison and Functional Annotation of Three Brassica Rapa Genomes: A Turnip, a Rapid Cycling and a Chinese Cabbage. *BMC Genom.* **2014**, *15*, 250. [[CrossRef](#)]
160. Li, Y.H.; Zhou, G.; Ma, J.; Jiang, W.; Jin, L.G.; Zhang, Z.; Guo, Y.; Zhang, J.; Sui, Y.; Zheng, L.; et al. De Novo Assembly of Soybean Wild Relatives for Pan-Genome Analysis of Diversity and Agronomic Traits. *Nat. Biotechnol.* **2014**, *32*, 1045–1052. [[CrossRef](#)]
161. Schatz, M.C.; Maron, L.G.; Stein, J.C.; Hernandez Wences, A.; Gurtowski, J.; Biggers, E.; Lee, H.; Kramer, M.; Antoniou, E.; Ghiban, E.; et al. Whole Genome de Novo Assemblies of Three Divergent Strains of Rice, *Oryza Sativa*, Document Novel Gene Space of Aus and Indica. *Genome Biol.* **2014**, *15*, 506. [[CrossRef](#)]
162. Hirsch, C.N.; Foerster, J.M.; Johnson, J.M.; Sekhon, R.S.; Muttoni, G.; Vaillancourt, B.; Peñagaricano, F.; Lindquist, E.; Pedraza, M.A.; Barry, K.; et al. Insights into the Maize Pan-Genome and Pan-Transcriptome. *Plant Cell* **2014**, *26*, 121–135. [[CrossRef](#)]
163. Yao, W.; Li, G.; Zhao, H.; Wang, G.; Lian, X.; Xie, W. Exploring the Rice Dispensable Genome Using a Metagenome-like Assembly Strategy. *Genome Biol.* **2015**, *16*, 187. [[CrossRef](#)]
164. Golicz, A.A.; Bayer, P.E.; Barker, G.C.; Edger, P.P.; Kim, H.; Martinez, P.A.; Chan, C.K.K.; Severn-Ellis, A.; McCombie, W.R.; Parkin, I.A.P.; et al. The Pangenome of an Agronomically Important Crop Plant Brassica Oleracea. *Nat. Commun.* **2016**, *7*, 13390. [[CrossRef](#)]
165. Pinosio, S.; Giacomello, S.; Faivre-Rampant, P.; Taylor, G.; Jorge, V.; Le Paslier, M.C.; Zaina, G.; Bastien, C.; Cattonaro, F.; Marroni, F.; et al. Characterization of the Poplar Pan-Genome by Genome-Wide Identification of Structural Variation. *Mol. Biol. Evol.* **2016**, *33*, 2706–2719. [[CrossRef](#)] [[PubMed](#)]
166. Gordon, S.P.; Contreras-Moreira, B.; Woods, D.P.; Des Marais, D.L.; Burgess, D.; Shu, S.; Stritt, C.; Roulin, A.C.; Schackwitz, W.; Tyler, L.; et al. Extensive Gene Content Variation in the Brachypodium Distachyon Pan-Genome Correlates with Population Structure. *Nat. Commun.* **2017**, *8*, 2184. [[CrossRef](#)] [[PubMed](#)]
167. Zhou, P.; Silverstein, K.A.T.; Ramaraj, T.; Guhlin, J.; Denny, R.; Liu, J.; Farmer, A.D.; Steele, K.P.; Stupar, R.M.; Miller, J.R.; et al. Exploring Structural Variation and Gene Family Architecture with De Novo Assemblies of 15 Medicago Genomes. *BMC Genom.* **2017**, *18*, 261. [[CrossRef](#)] [[PubMed](#)]
168. Hurgobin, B.; Golicz, A.A.; Bayer, P.E.; Chan, C.K.K.; Tirnaz, S.; Dolatabadian, A.; Schiessl, S.V.; Samans, B.; Montenegro, J.D.; Parkin, I.A.P.; et al. Homoeologous Exchange Is a Major Cause of Gene Presence/Absence Variation in the Amphidiploid Brassica Napus. *Plant Biotechnol. J.* **2018**, *16*, 1265–1274. [[CrossRef](#)] [[PubMed](#)]
169. Ou, L.; Li, D.; Lv, J.; Chen, W.; Zhang, Z.; Li, X.; Yang, B.; Zhou, S.; Yang, S.; Li, W.; et al. Pan-Genome of Cultivated Pepper (*Capsicum*) and Its Use in Gene Presence–Absence Variation Analyses. *New Phytol.* **2018**, *220*, 360–363. [[CrossRef](#)]
170. Zhao, Q.; Feng, Q.; Lu, H.; Li, Y.; Wang, A.; Tian, Q.; Zhan, Q.; Lu, Y.; Zhang, L.; Huang, T.; et al. Erratum to: Pan-Genome Analysis Highlights the Extent of Genomic Variation in Cultivated and Wild. *Nat. Genet.* **2018**, *50*, 278–284. [[CrossRef](#)]
171. Yu, J.; Golicz, A.A.; Lu, K.; Dossa, K.; Zhang, Y.; Chen, J.; Wang, L.; You, J.; Fan, D.; Edwards, D.; et al. Insight into the Evolution and Functional Characteristics of the Pan-Genome Assembly from Sesame Landraces and Modern Cultivars. *Plant Biotechnol. J.* **2019**, *17*, 881–892. [[CrossRef](#)]
172. Hübner, S.; Bercovich, N.; Todesco, M.; Mandel, J.R.; Odenheimer, J.; Ziegler, E.; Lee, J.S.; Baute, G.J.; Owens, G.L.; Grassa, C.J.; et al. Sunflower Pan-Genome Analysis Shows That Hybridization Altered Gene Content and Disease Resistance. *Nat. Plants* **2019**, *5*, 54–62. [[CrossRef](#)]
173. Song, J.M.; Guan, Z.; Hu, J.; Guo, C.; Yang, Z.; Wang, S.; Liu, D.; Wang, B.; Lu, S.; Zhou, R.; et al. Eight High-Quality Genomes Reveal Pan-Genome Architecture and Ecotype Differentiation of Brassica Napus. *Nat. Plants* **2020**, *6*, 34–45. [[CrossRef](#)]

174. Trouern-Trend, A.J.; Falk, T.; Zaman, S.; Caballero, M.; Neale, D.B.; Langley, C.H.; Dandekar, A.M.; Stevens, K.A.; Wegrzyn, J.L. Comparative Genomics of Six Juglans Species Reveals Disease-Associated Gene Family Contractions. *Plant J.* **2020**, *102*, 410–423. [[CrossRef](#)]
175. Liu, Y.; Du, H.; Li, P.; Shen, Y.; Peng, H.; Liu, S.; Zhou, G.A.; Zhang, H.; Liu, Z.; Shi, M.; et al. Pan-Genome of Wild and Cultivated Soybeans. *Cell* **2020**, *182*, 162–176.e13. [[CrossRef](#)] [[PubMed](#)]
176. Jensen, S.E.; Charles, J.R.; Muleta, K.; Bradbury, P.J.; Casstevens, T.; Deshpande, S.P.; Gore, M.A.; Gupta, R.; Ilut, D.C.; Johnson, L.; et al. A Sorghum Practical Haplotype Graph Facilitates Genome-wide Imputation and Cost-effective Genomic Prediction. *Plant Genome* **2020**, *13*, e20009. [[CrossRef](#)] [[PubMed](#)]
177. Dolatabadian, A.; Bayer, P.E.; Tirnaz, S.; Hurgobin, B.; Edwards, D.; Batley, J. Characterization of Disease Resistance Genes in the Brassica Napus Pangenome Reveals Significant Structural Variation. *Plant Biotechnol. J.* **2020**, *18*, 969–982. [[CrossRef](#)] [[PubMed](#)]
178. Lovell, J.T.; Bentley, N.B.; Bhattarai, G.; Jenkins, J.W.; Sreedasyam, A.; Alarcon, Y.; Bock, C.; Boston, L.B.; Carlson, J.; Cervantes, K.; et al. Four Chromosome Scale Genomes and a Pan-Genome Annotation to Accelerate Pecan Tree Breeding. *Nat. Commun.* **2021**, *12*, 4125. [[CrossRef](#)] [[PubMed](#)]
179. Bayer, P.E.; Scheben, A.; Golicz, A.A.; Yuan, Y.; Faure, S.; Lee, H.T.; Chawla, H.S.; Anderson, R.; Bancroft, I.; Raman, H.; et al. Modelling of Gene Loss Propensity in the Pangenomes of Three Brassica Species Suggests Different Mechanisms between Polyploids and Diploids. *Plant Biotechnol. J.* **2021**, *19*, 2488–2500. [[CrossRef](#)]
180. Varshney, R.K.; Roorkiwal, M.; Sun, S.; Bajaj, P.; Chitikineni, A.; Thudi, M.; Singh, N.P.; Du, X.; Upadhyaya, H.D.; Khan, A.W.; et al. A Chickpea Genetic Variation Map Based on the Sequencing of 3366 Genomes. *Nature* **2021**, *599*, 622–627. [[CrossRef](#)] [[PubMed](#)]
181. Tao, Y.; Luo, H.; Xu, J.; Cruickshank, A.; Zhao, X.; Teng, F.; Hathorn, A.; Wu, X.; Liu, Y.; Shatte, T.; et al. Extensive Variation within the Pan-Genome of Cultivated and Wild Sorghum. *Nat. Plants* **2021**, *7*, 766–773. [[CrossRef](#)]
182. Barchi, L.; Rabanus-Wallace, M.T.; Prohens, J.; Toppino, L.; Padmarasu, S.; Portis, E.; Rotino, G.L.; Stein, N.; Lanteri, S.; Giuliano, G. Improved Genome Assembly and Pan-Genome Provide Key Insights into Eggplant Domestication and Breeding. *Plant J.* **2021**, *7*, 766–773. [[CrossRef](#)]
183. Zhou, Y.; Zhang, Z.; Bao, Z.; Li, H.; Lyu, Y.; Zan, Y.; Wu, Y.; Cheng, L.; Fang, Y.; Wu, K.; et al. Graph pangenome captures missing heritability and empowers tomato breeding. *Nature* **2022**, *606*, 527–534. [[CrossRef](#)]
184. Garg, G.; Kamphuis, L.G.; Bayer, P.E.; Kaur, P.; Dudchenko, O.; Taylor, C.M.; Frick, K.M.; Foley, R.C.; Gao, L.L.; Aiden, E.L.; et al. A pan-genome and chromosome-length reference genome of narrow-leaved lupin (*Lupinus angustifolius*) reveals genomic diversity and insights into key industry and biological traits. *Plant J.* **2022**, *111*, 1252–1266. [[CrossRef](#)]
185. Ellis, J.A.; Ong, B. The MassARRAY<sup>®</sup> System for Targeted SNP Genotyping. In *Methods in Molecular Biology*; Springer: Berlin/Heidelberg, Germany, 2017; Volume 1492.
186. Dalma-Weiszhausz, D.D.; Warrington, J.; Tanimoto, E.Y.; Miyada, C.G. The Affymetrix GeneChip<sup>®</sup> Platform: An Overview. *Methods Enzymol.* **2006**, *410*, 3–28.
187. Fazal, F.M.; Han, S.; Parker, K.R.; Kaewsapsak, P.; Xu, J.; Boettiger, A.N.; Chang, H.Y.; Ting, A.Y. Atlas of Subcellular RNA Localization Revealed by APEX-Seq. *Cell* **2019**, *178*, 473–490.e26. [[CrossRef](#)] [[PubMed](#)]
188. Mealer, M.; Moss, M. TaqMan<sup>®</sup> Small RNA Assays. *Appl. Biosyst.* **2018**, *44*, 4398987.
189. Xiao, Q.; Bai, X.; Zhang, C.; He, Y. Advanced High-Throughput Plant Phenotyping Techniques for Genome-Wide Association Studies: A Review. *J. Adv. Res.* **2022**, *35*, 215–230. [[CrossRef](#)]
190. Mochida, K.; Koda, S.; Inoue, K.; Hirayama, T.; Tanaka, S.; Nishii, R.; Melgani, F. Computer Vision-Based Phenotyping for Improvement of Plant Productivity: A Machine Learning Perspective. *Gigascience* **2018**, *8*, giy153. [[CrossRef](#)] [[PubMed](#)]
191. Tsaftaris, S.A.; Minervini, M.; Scharf, H. Machine Learning for Plant Phenotyping Needs Image Processing. *Trends Plant Sci.* **2016**, *21*, 989–991. [[CrossRef](#)] [[PubMed](#)]
192. Lee, U.; Chang, S.; Putra, G.A.; Kim, H.; Kim, D.H. An Automated, High-Throughput Plant Phenotyping System Using Machine Learning-Based Plant Segmentation and Image Analysis. *PLoS ONE* **2018**, *13*, e0196615. [[CrossRef](#)]
193. Kolhar, S.; Jagtap, J. Plant Trait Estimation and Classification Studies in Plant Phenotyping Using Machine Vision—A Review. *Inf. Process. Agric.* **2021**, *10*, 114–135. [[CrossRef](#)]
194. Wang, J.; Zhang, Z. GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. *Genom. Proteom. Bioinform.* **2021**, *19*, 629–640. [[CrossRef](#)]
195. Turner, S.D. Qqman: An R Package for Visualizing GWAS Results Using Q-Q and Manhattan Plots. *J. Open Source Softw.* **2018**, *3*, 731. [[CrossRef](#)]
196. Magno, R.; Maia, A.T. Gwasrapidd: An R Package to Query, Download and Wrangle GWAS Catalog Data. *Bioinformatics* **2020**, *36*, 649–650. [[CrossRef](#)] [[PubMed](#)]
197. Drivas, T.G.; Lucas, A.; Ritchie, M.D. EQTpLot: A User-Friendly R Package for the Visualization of Colocalization between EQTL and GWAS Signals. *BioData Min.* **2021**, *14*, 32. [[CrossRef](#)]
198. Hiersche, M.; Rühle, F.; Stoll, M. Postgwas: Advanced GWAS Interpretation in R. *PLoS ONE* **2013**, *8*, e71775. [[CrossRef](#)]
199. Gogarten, S.M.; Bhangale, T.; Conomos, M.P.; Laurie, C.A.; McHugh, C.P.; Painter, I.; Zheng, X.; Crosslin, D.R.; Levine, D.; Lumley, T.; et al. GWASTools: An R/Bioconductor Package for Quality Control and Analysis of Genome-Wide Association Studies. *Bioinformatics* **2012**, *28*, 3329–3331. [[CrossRef](#)]



200. He, F.; Ding, S.; Wang, H.; Qin, F. IntAssoPlot: An R Package for Integrated Visualization of Genome-Wide Association Study Results With Gene Structure and Linkage Disequilibrium Matrix. *Front. Genet.* **2020**, *11*, 260. [[CrossRef](#)]
201. Ruperao, P.; Gandham, P.; Odeny, D.A.; Mayes, S.; Selvanayagam, S.; Thirunavukkarasu, N.; Das, R.R.; Srikanda, M.; Gandhi, H.; Habyarimana, E.; et al. Exploring the Sorghum Race Level Diversity Utilizing 272 Sorghum Accessions Genomic Resources. *Front. Plant Sci.* **2023**, *14*, 1143512. [[CrossRef](#)] [[PubMed](#)]
202. Ahmed, I.; Jeon, G. Enabling Artificial Intelligence for Genome Sequence Analysis of COVID-19 and Alike Viruses. *Interdiscip. Sci. Comput. Life Sci.* **2022**, *14*, 504–519. [[CrossRef](#)]
203. León-Novelo, L.G.; McIntyre, L.M.; Fear, J.M.; Graze, R.M. A Flexible Bayesian Method for Detecting Allelic Imbalance in RNA-Seq Data. *BMC Genom.* **2014**, *15*, 920. [[CrossRef](#)] [[PubMed](#)]
204. van Dijk, A.D.J.; Kootstra, G.; Kruijer, W.; de Ridder, D. Machine Learning in Plant Science and Plant Breeding. *iScience* **2021**, *24*, 101890. [[CrossRef](#)]
205. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
206. Chollet, F. Keras: The Python Deep Learning library—NASA/ADS. Available online: [harvard.edu](https://arxiv.org/abs/1609.03226) (accessed on 24 July 2023).
207. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems*; The MIT Press: Cambridge, MA, USA, 2019; Volume 32.
208. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M. Tensorflow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *arXiv* **2016**, arXiv:1603.044672019.
209. Kuhn, M. Building Predictive Models in R Using the Caret Package. *J. Stat. Softw.* **2008**, *28*, 1–26. [[CrossRef](#)]
210. Sun, S.; Wang, C.; Ding, H.; Zou, Q. Machine Learning and Its Applications in Plant Molecular Studies. *Brief. Funct. Genom.* **2018**, *19*, 40–48. [[CrossRef](#)] [[PubMed](#)]
211. Sperschneider, J. Machine Learning in Plant–Pathogen Interactions: Empowering Biological Predictions from Field Scale to Genome Scale. *New Phytol.* **2020**, *228*, 35–41. [[CrossRef](#)] [[PubMed](#)]
212. Wang, H.; Cimen, E.; Singh, N.; Buckler, E. Deep Learning for Plant Genomics and Crop Improvement. *Curr. Opin. Plant Biol.* **2020**, *54*, 34–41. [[CrossRef](#)]
213. Bayer, P.E.; Petereit, J.; Danilevicz, M.F.; Anderson, R.; Batley, J.; Edwards, D. The Application of Pangenomics and Machine Learning in Genomic Selection in Plants. *Plant Genome* **2021**, *14*, e20112. [[CrossRef](#)]
214. Sartor, R.C.; Noshay, J.; Springer, N.M.; Briggs, S.P. Identification of the Expressome by Machine Learning on Omics Data. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 18119–18125. [[CrossRef](#)] [[PubMed](#)]
215. Demirci, S.; Peters, S.A.; de Ridder, D.; van Dijk, A.D.J. DNA Sequence and Shape Are Predictive for Meiotic Crossovers throughout the Plant Kingdom. *Plant J.* **2018**, *95*, 686–699. [[CrossRef](#)]
216. Bourgeois, Y.; Stritt, C.; Walser, J.C.; Gordon, S.P.; Vogel, J.P.; Roulin, A.C. Genome-Wide Scans of Selection Highlight the Impact of Biotic and Abiotic Constraints in Natural Populations of the Model Grass *Brachypodium distachyon*. *Plant J.* **2018**, *96*, 438–451. [[CrossRef](#)]
217. Zhou, P.; Li, Z.; Magnusson, E.; Cano, F.G.; Crisp, P.A.; Noshay, J.M.; Grotewold, E.; Hirsch, C.N.; Briggs, S.P.; Springer, N.M. Meta Gene Regulatory Networks in Maize Highlight Functionally Relevant Regulatory Interactions. *Plant Cell* **2020**, *32*, 1377–1396. [[CrossRef](#)]
218. Pérez-Enciso, M.; Zingaretti, L.M. A Guide for Using Deep Learning for Complex Trait Genomic Prediction. *Genes* **2019**, *10*, 553. [[CrossRef](#)]
219. Gambhir, J.; Patel, N.; Patil, S.; Takale, P.; Chougule, A.; Prabhakar, C.S.; Managanvi, K.; Raghavan, A.S.; Sohane, R.K. Deep Learning for Real-Time Diagnosis of Pest and Diseases on Crops. In *Smart Innovation, Systems and Technologies*; Springer: Berlin/Heidelberg, Germany, 2022; Volume 266.
220. Montesinos-López, A.; Montesinos-López, O.A.; Gianola, D.; Crossa, J.; Hernández-Suárez, C.M. Multi-Environment Genomic Prediction of Plant Traits Using Deep Learners with Dense Architecture. *G3 Genes Genomes Genet.* **2018**, *8*, 3813–3828. [[CrossRef](#)] [[PubMed](#)]
221. Azodi, C.B.; Tang, J.; Shiu, S.H. Opening the Black Box: Interpretable Machine Learning for Geneticists. *Trends Genet.* **2020**, *36*, 442–455. [[CrossRef](#)]
222. Poplin, R.; Chang, P.C.; Alexander, D.; Schwartz, S.; Colthurst, T.; Ku, A.; Newburger, D.; Dijamco, J.; Nguyen, N.; Afshar, P.T.; et al. A Universal Snp and Small-Indel Variant Caller Using Deep Neural Networks. *Nat. Biotechnol.* **2018**, *36*, 983–987. [[CrossRef](#)] [[PubMed](#)]
223. Yocca, A.E.; Edger, P.P. Machine Learning Approaches to Identify Core and Dispensable Genes in Pangenomes. *Plant Genome* **2022**, *15*, e20135. [[CrossRef](#)] [[PubMed](#)]
224. Koh, J.C.O.; Spangenberg, G.; Kant, S. Automated Machine Learning for High-throughput Image-based Plant Phenotyping. *Remote Sens.* **2021**, *13*, 858. [[CrossRef](#)]
225. Arya, S.; Sandhu, K.S.; Singh, J.; Kumar, S. Deep Learning: As the New Frontier in High-Throughput Plant Phenotyping. *Euphytica* **2022**, *218*, 47. [[CrossRef](#)]
226. Tong, H.; Nikoloski, Z. Machine Learning Approaches for Crop Improvement: Leveraging Phenotypic and Genotypic Big Data. *J. Plant Physiol.* **2021**, *257*, 153354. [[CrossRef](#)] [[PubMed](#)]

227. Endelman, J.B. Ridge Regression and Other Kernels for Genomic Selection with R Package RrBLUP. *Plant Genome* **2011**, *4*, 255–258. [[CrossRef](#)]
228. Pérez, P.; De Los Campos, G. Genome-Wide Regression and Prediction with the BGLR Statistical Package. *Genetics* **2014**, *198*, 483–495. [[CrossRef](#)]
229. Bates, D.; Mächler, M.; Bolker, B.M.; Walker, S.C. Fitting Linear Mixed-Effects Models Using Lme4. *J. Stat. Softw.* **2015**, *67*, 48. [[CrossRef](#)]
230. Butler, D.G.; Cullis, B.R.; Gilmour, A.R.; Gogel, B.J.; Thompson, R. ASReml-R Reference Manual Version 4. In *ASReml-R Reference Manual*; VSN International Ltd.: Hertfordshire, UK, 2018.
231. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, 1–22. [[CrossRef](#)] [[PubMed](#)]
232. Montesinos-López, O.A.; Montesinos-López, A.; Luna-Vázquez, F.J.; Toledo, F.H.; Pérez-Rodríguez, P.; Lillemo, M.; Crossa, J. An R Package for Bayesian Analysis of Multi-Environment and Multi-Trait Multi-Environment Data for Genome-Based Prediction. *G3 Genes Genomes Genet.* **2019**, *9*, 1355–1369. [[CrossRef](#)] [[PubMed](#)]
233. Niazian, M.; Niedbała, G. Machine Learning for Plant Breeding and Biotechnology. *Agriculture* **2020**, *10*, 436. [[CrossRef](#)]
234. Ersoz, E.S.; Martin, N.F.; Stapleton, A.E. On to the next Chapter for Crop Breeding: Convergence with Data Science. *Crop Sci.* **2020**, *60*, 639–655. [[CrossRef](#)]
235. González-Recio, O.; Forni, S. Genome-Wide Prediction of Discrete Traits Using Bayesian Regressions and Machine Learning. *Genet. Sel. Evol.* **2011**, *43*, 7. [[CrossRef](#)]
236. Long, N.; Gianola, D.; Rosa, G.J.M.; Weigel, K.A. Application of Support Vector Regression to Genome-Assisted Prediction of Quantitative Traits. *Theor. Appl. Genet.* **2011**, *123*, 1065–1074. [[CrossRef](#)]
237. González-Recio, O.; Jiménez-Montero, J.A.; Alenda, R. The Gradient Boosting Algorithm and Random Boosting for Genome-Assisted Evaluation in Large Data Sets. *J. Dairy Sci.* **2013**, *96*, 614–624. [[CrossRef](#)]
238. Cheng, S.; Melkonian, M.; Smith, S.A.; Brockington, S.; Archibald, J.M.; Delaux, P.M.; Li, F.W.; Melkonian, B.; Mavrodiev, E.V.; Sun, W.; et al. 10KP: A Phylo diverse Genome Sequencing Plan. *Gigascience* **2018**, *7*, gyy013. [[CrossRef](#)] [[PubMed](#)]
239. Varshney, R.K.; Ribaut, J.M.; Buckler, E.S.; Tuberosa, R.; Rafalski, J.A.; Langridge, P. Can Genomics Boost Productivity of Orphan Crops? *Nat. Biotechnol.* **2012**, *30*, 1172–1176. [[CrossRef](#)]
240. Varshney, R.K.; Bohra, A.; Yu, J.; Graner, A.; Zhang, Q.; Sorrells, M.E. Designing Future Crops: Genomics-Assisted Breeding Comes of Age. *Trends Plant Sci.* **2021**, *26*, 631–649. [[CrossRef](#)] [[PubMed](#)]
241. Xu, Y.; Luo, H.; Zhang, H.; Yung, W.S.; Li, M.W.; Lam, H.M.; Huang, C. Feeding the World Using Speed Breeding Technology. *Trends Plant Sci.* **2023**, *28*, 372–373. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.