



Review

Genome-Wide, High-Density Genotyping Approaches for Plant Germplasm Characterisation (Methods and Applications)

Sirine Werghi , Brian Wakimwayi Koboyi , David Chan-Rodriguez and Hanna Bolibok-Brągoszewska *

Department of Plant Genetics Breeding and Biotechnology, Institute of Biology, Warsaw University of Life Sciences, 02-776 Warsaw, Poland; sirine_werghi@sggw.edu.pl (S.W.); brian_koboyi@sggw.edu.pl (B.W.K.); david_chan_rodriguez@sggw.edu.pl (D.C.-R.)

* Correspondence: hanna_bolibok-brągoszewska@sggw.edu.pl

Abstract

Germplasm collections are a treasure trove of humanity. The accessions constituting those collections (wild crop relatives, landraces, cultivars, etc.) contain genes and allelic variants, which evolved prior to or post domestication, in the course of adaptation and selection, and can be used in breeding to address current and future needs. Precise characterisation of genetic diversity is essential for the efficient conservation of genetic resources and their effective utilisation in crop improvement. Detailed genetic profiles resulting from DNA genotyping constitute a basis for establishing the level of genetic diversity of a collection, analysing population structure, identifying redundancies, performing genome-wide association scans (given the availability of phenotypic information), detecting loci under selection, and many other applications. To obtain an accurate picture of genetic diversity (at the DNA sequence level), robust, high-density, high-throughput, and cost-effective methods are needed. With the advances in the next-generation sequencing, new genotyping approaches emerged (such as genotyping-by-sequencing, whole genome resequencing), which provide excellent genome coverage and low cost per datapoint (with tens of thousands to millions of loci analysed in a single assay). Crop-specific, custom, microarray-based genotyping solutions were also developed. The aim of this review is to provide a comparative description of the genome-wide, high-density genotyping technologies that are most frequently used nowadays, comprising their advantages and drawbacks, as well as factors that determine, which of the methods will best suit the particular germplasm characterisation project. Further, we characterise the current role of these methods in addressing the challenges related to the effective management and use of genetic resources and present recent examples of their application in selected crop plant groups. Finally, we briefly describe constraints to germplasm characterisation and future prospects.

Keywords: genotyping; plant genetic resources; genetic diversity; GBS; reduced representation sequencing; whole genome resequencing; DNA-genotyping arrays



Academic Editor: Andrés J. Cortés

Received: 15 October 2025

Revised: 30 November 2025

Accepted: 3 December 2025

Published: 8 December 2025

Citation: Werghi, S.; Koboyi, B.W.; Chan-Rodriguez, D.; Bolibok-Brągoszewska, H. Genome-Wide, High-Density Genotyping Approaches for Plant Germplasm Characterisation (Methods and Applications). *Int. J. Mol. Sci.* **2025**, *26*, 11833. <https://doi.org/10.3390/ijms262411833>

Copyright: © 2025 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

Over 5.9 million plant accessions (predominantly crop plants and their wild relatives) are maintained in ex situ germplasm collections worldwide, according to the III Report on the State of the World's Plant Genetic Resources for Food and Agriculture [1]. Together these accessions constitute a living library of natural genetic variation—gene variants and specific allelic combinations that came into existence over the course of the evolution

of plant species, adaptation to specific geoclimatic conditions, in response to natural or artificial selection pressure.

Crucially, plant genetic resources (PGR) contain gene variants, which could be useful for crop improvement, to address current needs related to, among others, to climate change (crop varieties more resilient to hostile conditions, such as drought, heat are needed), and population growth (higher yields have to be achieved to provide enough food for the constantly growing human population, for example, by introduction of genes increasing nutrient use or/and uptake efficiency). A well-known and striking example of the realisation of the potential of PGR are the Green Revolution genes [2] originating from landraces of wheat and rice. Semi-dwarf, high-yielding varieties carrying these genes significantly contributed to improving food security in the developing countries between 1960 and 2000 [3]. More recent examples of PGR potential in crop improvement include the discovery of *S. pennellii* genome segments that enhanced the productivity of a top market variety by more than 50% [4]. In rice, a gene enhancing phosphorus deficiency tolerance, *Pstol1*, was discovered in a traditional variety, Kasalath. This gene is absent from the reference genome of the variety Nipponbare and other modern phosphorus-deficiency-sensitive rice varieties [5].

Unfortunately, there are considerable constraints to the effective use of PGR in plant improvement. The vast size of PGR collections often means that little is known about the extent and the distribution of the diversity, and it is very challenging to pinpoint an accession or accessions that could contribute variation suited to the particular breeding target. Specific challenges resulting from the biology of the species in question (such as outcrossing nature, high heterozygosity, heterogeneity of accessions), complicate the issue further [6,7]. Specific schemes (such as development of NAM, MAGIC, or RCSL populations [8,9] were devised to facilitate inclusion of useful exotic variation into breeding programmes. Nevertheless, the accurate choice of the PGR accession for each of these approaches is fundamental to the success of the endeavour and can be facilitated by the availability of the genome-wide, high-density genotypic information for the PGR collection in question.

The aim of this review is to provide a comparative description of the genome-wide, high-density genotyping technologies that are most frequently used nowadays, summarising their advantages and drawbacks, as well as discussing factors that determine, which of the methods will best suit the particular germplasm characterisation project. Further, we characterise the current role of these methods in addressing the challenges related to the effective management and use of genetic resources and present recent examples of genome-wide high-density genotyping application in selected crop plant groups. Finally, we briefly describe constraints to germplasm characterisation and future prospects.

2. Plant Germplasm Genotyping Approaches

Various methods were developed over the decades to detect polymorphism at the DNA level, starting from restriction enzyme-based methods (RFLP), various PCR-based methods (RAPD, SSR), to methods combining both digestion and PCR (such as AFLP). In the first decade of the XXI century, microarray-based methods, especially those independent from the availability of sequence information for the species in question, such as DArT [10] became very popular. From the above-mentioned methods, SSRs still enjoy considerable popularity [11–13] as a reliable and affordable tool for detecting DNA polymorphism, particularly when funding is limited. SSRs have the advantages of being co-dominant, highly polymorphic and multiallelic, as well as sequence specific and therefore easily transferable between experiments. On the other hand, the development of new SSR markers is usually a laborious task and the throughput is low, since typically a single

SSR locus is targeted by a single assay [14]. Though the number of loci that can be analysed in a single study is not very high (usually around 15–30), due to a relatively high labour-intensiveness of the assay, a number of carefully chosen SSR markers can provide reliable information on the genetic diversity level and its distribution in a given collection [15–18]. SSRs also remain an important tool in specific applications, e.g., in varietal identification. In apple a Malus UNiQue genotype coding system (MUNQ) based on 15 SSR [19] is used for germplasm identification.

The advances in next-generation sequencing technologies gave rise to a new type of genotyping methods, described as reduced-representation sequencing methods [20–24]. With steady improvements in sequencing capacity, sinking prices, and ongoing progress in bioinformatic analysis tools and pipelines, detection of polymorphism via low-coverage whole genome resequencing (WGRS) became feasible [25–29]. The constantly growing availability of sequence data and the discovery of numerous single nucleotide polymorphisms (SNPs) permitted the development of microarray-based methods aimed at simultaneous detection of polymorphism at thousands of preidentified, selected SNPs such as Illumina Infinium Beadchip, Affymetrix Axiom or Affymetrix GeneChip [30]. These three groups of genotyping methods prevail nowadays and will be characterised in the Section 2.1 of this review. An overview of these methods is presented in Table 1.

Table 1. Overview of genome-wide high-density genotyping approaches.

Genotyping Approach	Platform	Polymorphism Detection Capacity	Type of Polymorphism Detected	Reference Genome Required	Advantages	Disadvantages
RRS *	GBS DArTseq RAD-seq SLAF-seq	Thousands to hundreds of thousands of SNPs	SNPs PAV (some platforms)	No	Cost-effective. No need for prior SNP information. Works across any genome size and any species. Ideal for non-model and orphan species.	Demands higher computational skills and QC than arrays. Mapping bias. Difficulty distinguishing homologous vs. homoeologous loci in high-ploidy species. Reproducibility is protocol-dependent.
WGRS	Low-coverage-Illumina based Long-read-sequencing based	Millions to hundreds of millions of SNPs	SNPs InDels CNV SVs PAV	Yes	Discovery of various types of polymorphism (SNPs, InDels, CNV, SVs). Survey of polymorphism across the whole genome.	Requires a reference genome sequence (or transcriptome). Demands specialised bioinformatic skills. More expensive compared to RRS and SNP arrays.
SNP arrays	Illumina Infinium (iSelect/BeaChip) ThermoFisher Axiom	iSelect HD: 3 k to 90 k iSelect HTS: 90 k to 700 k Up to 2.6 million	SNPs InDels CNV SVs PAV	Yes	Accurate genotyping in polyploid species. Straightforward downstream analysis. High reproducibility and accuracy of genotyping calls. Same SNP calls remain stable across breeding programmes and years	No novel variant discovery. Reduced power to detect marker-trait association Skewed allele frequencies Biased representation of genetic variations

* Abbreviations used in Table 1: CNV—copy number variation, DArTseq—diversity array technology sequencing, GBS—genotyping-by-sequencing, InDel—insertion/deletion, PAV—presence/absence variation, RAD-seq—restriction site-associated DNA sequencing, RRS—reduced representation sequencing, SLAF-seq—specific-locus amplified fragment sequencing, SNP—single nucleotide polymorphism, SV—structural variation, WGRS—whole genome resequencing.

Specific methods of genotyping a moderately sized set of selected SNPs (for instance, Kompetitive Allele-Specific PCR (KASP) [31] and Fluidigm genotyping assays [32]) or analysing sequence diversity of selected target genes (for example, by amplicon sequencing [33]) are also in frequent use, but since these methods do not provide high-density genome-wide coverage, they will not be described here).

2.1. Reduced Representation Sequencing (RRS) Methods

Reduced representation sequencing (RRS) methods, such as GBS [20], DArTseq [21], RAD-seq [23], SLAF-seq [34], sometimes also collectively called GBS [24] provide high-density, low-cost genotypic data. Since a reference genome sequence is not required, these methods are particularly useful for non-model plants and orphan crops. Generally, these methods permit simultaneous discovery and genotyping of variants.

RRS workflow includes several key steps: (1) genome complexity reduction using restriction enzymes; (2) barcoding genomic DNA with indexed adaptors; (3) high-throughput sequencing of barcoded fragments; (4) bioinformatics analysis to identify genetic variants. Genome complexity reduction is a central concept of the reduced representation sequencing methods. In this analysis step, also called genome sampling, a subsection of the genome, usually containing gene-rich regions, is selected. This is achieved in most cases by using a methylation-sensitive restriction endonuclease for DNA digestion, which initiates the sample preparation. Then, ligation of adaptors, size selection, and PCR amplifications (depending on the approach) can be used to achieve the final set of fragments representing a given DNA sample, which will be subjected to NGS [22]. Bioinformatic pipelines are available for analysis of resulting reads and include both reference-based and de novo solutions, when a reference genome sequence is not yet available for the species being analysed [35].

Reduced-representation approaches typically begin with raw read demultiplexing and adapter/quality filtering, followed by alignment to a reference genome (or pseudo-reference) and variant calling. Several software packages are commonly used for GBS and related reduced-representation data analysis. The TASSEL-GBS pipeline [36] was originally created for maize and allows efficient SNP discovery and genotype calling in large sample sets with low computational requirements. Within TASSEL 3.0, the UNEAK module [37] provides a reference-free option for species lacking a genome assembly. However, the UNEAK software module assumes diploidy, which can limit its performance in polyploids. Another popular toolkit, Stacks (<http://catchenlab.life.illinois.edu/stacks/manual/>, 2 December 2025) [38], has both de novo and reference-based workflows for SNP discovery. It has been applied across many plant species, especially when high-quality reference genomes are not available. For variant calling in complex or polyploid species, FreeBayes (<https://github.com/freebayes/freebayes>, access date: 2 December 2025) [39] provides a Bayesian framework. This allows users to specify ploidy level and model allele dosage more accurately. For deeper or higher-coverage datasets, generic variant-calling tools like GATK (<https://gatk.broadinstitute.org>, access date: 2 December 2025) [40] are also frequently used.

Most imputation algorithms were initially created for the human diploid genome and then adapted to plant genome analysis. In these genomes, each locus has two alleles, and researchers can infer haplotype phase from linkage disequilibrium patterns. The newest popular software successor is Beagle 5.0 [41]. Beagle performs haplotype-based imputation efficiently in large datasets and has been used with crops like soybean [42]. Other tools, such as IMPUTE5 [43] or its versions [44], also depend on pre-phased reference panels. They are best suited for diploid or highly inbred populations where high-quality reference genotypes are available [44]. The module FILLIN created for plant breeding within the TASSEL 5 framework [36] offers a method based on the haplotype library. It efficiently imputes missing genotypes in diploid datasets but assumes biallelic loci. These tools perform well in diploids because estimating allele dosage and phasing is relatively simple compared to polyploids.

In polyploid species, imputation is more difficult because multiple homologous chromosome sets make it harder to estimate allele dosage, phase haplotypes, and model segregation [45]. To address this, several tools have been created specifically for or adapted

to polyploids, such as Imputef [45] and polyRAD (<https://github.com/lvclark/polyRAD>, access date: 2 December 2025) [46]. In TASSEL5, FILLIN [47] can be used with higher ploidy levels through haplotype libraries, although accuracy may decrease when allele dosage is uncertain. Ongoing development of polyploid-aware imputation and variant-calling tools is still limited because realistic models of complex genomes (featuring multiple chromosome copies, allele dosage variation, homoeologous sequence similarity, and sometimes mixed inheritance patterns) require high computing power [48].

Quality control is mostly performed using VCFtools (<https://github.com/vcftools/vcftools>, access date: 2 December 2025) [49] or PLINK (<https://www.cog-genomics.org/plink>, access date: 2 December 2025) [50], filtering markers by call rate, minor allele frequency, and Hardy–Weinberg equilibrium before conducting downstream analyses such as GWAS, diversity, or population-structure inference.

Depending on the species, GBS delivers thousands to hundreds of thousands of SNPs per sample [51]. In a recent study on faba bean, the GBS cost per sample (including library preparation and sequencing) has been reported at approximately EUR 21 per sample [52], which translates to a very low cost per data point.

Since GBS methods do not rely on a set of pre-selected SNPs, they are consequently less prone to an ascertainment bias, a problem associated mostly with SNP arrays, which will be described below. However, GBS may introduce biases related to the choice of the restriction enzyme [22,52]. Wickland et al. [53], showed that SNP call sets from various GBS pipelines differed considerably, highlighting that comparability across studies is not guaranteed unless the method is exactly the same. Moreover, the reproducibility of GBS is dependent on the experimental protocol: Zamalutdinov et al. compared different restriction enzyme combinations (*HindIII-NlaIII*, *PstI-MspI*, and *ApeKI*) and different SNPs calling pipelines in 12 soybean varieties. They found that the enzyme selection and the choice of the right pipeline inherently influence the number of SNPs and their quality [54].

Although RRS methods enable deep sequencing of small genome portions in many samples, certain genomic regions, which might be critical for reaching very specific study aims (such as loci targeted by selection or associated with adaptive divergence) may be missed by the given assay design [55]. Further drawbacks of these methods include uneven coverage, higher rates of missing data, and under-calling of heterozygotes, especially in heterozygous or polyploid species. These issues can be mitigated through the use of genotype imputation and improved bioinformatic pipelines [53].

Noteworthy, it was demonstrated recently that the capacity of the polymorphism detection based on short sequence reads can be extended beyond primarily SNPs with the help of long-read sequencing data [56]. In this study, low coverage (~12×) Oxford Nanopore data from a moderately sized set of soybean accessions (17) was used for SV discovery and those variants could then be genotyped with high accuracy using Illumina reads (WGRS) in a population consisting of 102 Canadian soybean cultivars.

Despite certain limitations, RRS is an immensely popular genotyping approach, widely used for germplasm characterisation and genomic studies in a broad range of plant species. For many applications, such as phylogeny and population structure analysis, GWAS, and genomic prediction, the density and quality of GBS data are sufficient to obtain very reliable results [57] comparable with the outcomes of WGR [58], but in a far more cost-effective way. Examples of plant germplasm characterisation studies involving RRS are listed in Table 2.

Table 2. Examples of RRS application in plant germplasm characterisation studies.

Organism	RRS Variant	No. of Accessions	Institution, Country, Where Germplasm Is Kept	No. of Polymorphic Loci	Analyses	Reference
African mahogany	GBS	115	forest plantations in the Reserva Natural Vale and Viveiro Origem, Brasil	3.3 k	diversity assessment, population structure	[59]
amaranth	GBS	192	National Bureau of Plant Genetic Resources, India	42 k	phylogeny, population structure	[60]
apricot	RAD-seq	168	Luntai National Fruit Germplasm Resources Garden; Yingjisha County Apricot National Forest Germplasm Bank; Xiongyue National Germplasm Resources Garden, China	418 k	population structure, gene flow selection scan	[61]
avocado	GBS	384	Colombian Germplasm Bank; Seedling Rootstocks (SR) (n = 240) of commercial orchards from the northwest Andes; Colombia	4.9 k	diversity assessment, population structure, phylogeny	[62]
banana	DArTseq	856	International Institute of Tropical Agriculture, Nigeria, Tanzania, Uganda; National Agriculture Research Organization, Uganda; Embrapa, Brasil; National Research Centre for Banana, India; International Transit Centre, Belgium	6.1–19.7 k	diversity assessment, population structure	[63]
barley	GBS	22.6 k	Leibniz Institute of Plant Genetics and Crop Plant Research, Germany; National Crop Genebank of China, China; Agroscope, Switzerland	170 k	population structure GWAS redundancy	[64]
blueberry	GBS	195	Philip E. Marucci Center for Blueberry & Cranberry Research and Extension, State University of New Jersey, USA	60.5 k	Population structure, gene flow, section scan	[65]
canola	GBS	433	Kansas State University, USA	251.5 k	population structure	[66]
<i>Capsicum</i>	GBS	283	AGROSAVIA La Selva Research Station, Colombia	68.5 k; 30 k	population structure, GWAS	[67]
cassava	DArTseq	5.3 k	International Center for Tropical Agriculture, Colombia	7 k	redundancy	[7]
common bean	GBS	78	International Center for Tropical Agriculture, Colombia	23.3 k	kinship, population structure, GWAS	[68]
<i>Crotolaria</i>	GBS	80	Genetic Resources Research Institute of Kenya, Kenya	9.8 k	diversity assessment, phylogeny, population structure	[69]
faba bean	GBS	217	ICARDA genebank, Lebanon	40 k	linkage mapping GWAS	[52]
foxtail lily	GBS	96	wild Eremurus populations in Iran	3 k	phylogeny, population structure	[70]

Table 2. Cont.

Organism	RRS Variant	No. of Accessions	Institution, Country, Where Germplasm Is Kept	No. of Polymorphic Loci	Analyses	Reference
melon	GBS	755	National Agriculture and Food Research Organization, Japan	39.3 k	diversity assessment, population structure, core subset selection	[71]
oat	GBS	9112	Multiple institutions	19.9 k	population structure, structural rearrangements	[72]
oil palm	GBS	478	Malaysian Palm Oil Board Research Station, Malaysia	7 k	population structure, core subset selection	[73]
peas	DArTseq	325	Instituto de Agricultura Sostenible, Spain	35.8 k	phylogeny, population structure LD scan	[74]
phalaenopsis	GBS	116	National Cheng Kung University, China	113.5 k	GWAS	[75]
potato	GBS	730	US potato genebank, Sturgeon Bay, USA	7.8 k	ploidy estimation, population structure, core subset selection	[76]
rye	DArTseq	478	Several genebanks, universities and breeding companies	12.8 k	phylogeny, population structure, selection scan	[77]
sesame	GBS	501	US Department of Agriculture sesame collection, USDA-ARS Plant Genetic Resources Conservation Unit, USA	24.7 k	phylogeny, population structure, LD scan	[78]
sunflower	RAD seq	135	Active Germplasm Bank of Instituto Nacional de Tecnología Agropecuaria Manfredi, Argentina	11.8 k	diversity assessment, population structure, LD scan	[79]
quinoa	GBS	136	Germplasm Resources Information Network of the US Department of Agriculture, USA	5.7 k	phylogeny, population structure, LD scan	[80]
wheat	DArTseq	80 k	International Maize and Wheat Improvement Center, Mexico; International Center for Agricultural Research in the Dry Areas, Morocco	40 k	population structure, redundancy, core subset selection, selection scan, GWAS,	[81]
yam	DArTseq	100	International Institute of Tropical Agriculture, Nigeria	7 k	population structure	[82]

2.2. Whole Genome Resequencing (WGRS)

WGRS involves sequencing the entire genome of individual samples using next-generation sequencing (NGS) and/or third-generation sequencing (TGS), typically at $\sim 5\times$ to $15\times$ coverage [83], without the genome sampling step typical to RRS methods. This approach requires a reference genome sequence for the species under study, or at least for a close relative. The use of a reference transcriptome is also a solution. The possibilities of application of WGRS to many non-model species, especially orphan

crops, lacking a reference genome, are limited by this prerequisite. Performing WGRS in such situations requires creating a genome assembly *de novo* [55]. WGRS provides a very detailed and in-depth insight into the genetic diversity of the analysed germplasm set, since by aligning the sequencing reads of the investigated individuals to the reference genome sequence, all types of genetic variations can be identified, including SNPs, structural variations (SVs), copy number variations (CNVs), presence–absence variations (PAV) and insertion-deletions (InDels). Typically, millions of variations across the genome are identified and scored [55,84]. The ability of variant detection can be limited by the incomplete and inaccurate genome assembly used in data analysis as well as the lack of complete annotation. Although WGRS is highly informative, it is relatively expensive due to the extensive sequencing and data analysis involved [55]. Based on the 2025 fee list from a European core facility CNAG (Centro Nacional de Análisis Genómico, (Barcelona, Spain)) WGRS of a 1 GB genome at 10× coverage typically costs EUR 140–EUR 220 net per sample (library preparation and sequencing).

Typical WGRS bioinformatics analysis starts with checking the quality and trimming reads using tools like FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>, access date: 2 December 2025) [85] and Trimmomatic (<http://www.usadellab.org/cms/?page=trimmomatic>, access date: 2 December 2025) [86] to eliminate adapters and low-quality bases. Cleaned reads are then aligned to a reference genome with mappers such as BWA-MEM (<https://deepwiki.com/lh3/bwa/3.1-bwa-mem-algorithm>, access date: 2 December 2025) [87] or Bowtie2 [88]. After alignment, processing steps like sorting, indexing, and marking duplicates are carried out using SAMtools (<https://www.htslib.org/>, access date: 2 December 2025) [89] or Picard Tools (<https://broadinstitute.github.io/picard/>, accessed 7 November 2025) before variant calling.

For diploid species, calling variants usually involves GATK HaplotypeCaller [40] or bcftools (<https://www.htslib.org/>, access date: 2 December 2025) [87]. These tools assume there are two alleles per locus and work well when the sequencing depth is sufficient. These workflows are common for crop species like rice, maize, and soybean [90–92], where reference genomes are well established. Variants are filtered and annotated using VCFtools [49] and SnpEff (<http://pcingola.github.io/SnpEff/>, access date: 2 December 2025) [93] to estimate their potential effects.

When analysing polyploid plants, numerous challenges arise, as mentioned above in the RRS Section 2.1. Preferred variant callers, like FreeBayes [39] and GATK4 [94], allow users to define ploidy level and more accurately model allele dosage. For instance, FreeBayes has been effectively used in tetraploid alfalfa to find dosage-sensitive variants and improve genotype accuracy [95].

Genotype imputation is often used to improve SNP density and accuracy in WGRS analysis. Common imputation software options include Beagle (v5.0) [96], IMPUTE2/IMPUTE5 [43,97], practical haplotype graph (PHG) (v2 [98], AlphaPlantImpute (<https://github.com/AlphaGenes/AlphaPlantImpute2>, access date: 2 December 2025)) [99], and LinkImpute [100]. These tools were developed particularly in non-model organisms and plant datasets with low resources. Success relies on having a well-matched, high-depth reference panel, enough marker overlap between the reference and target samples, and accurate tuning of software settings for plant population structures. It is important to point out that the selection of tools should match the species' ploidy and population structure (diploid or polyploid; inbred or outbred).

Researchers should recognise that while the imputation software used for WGRS and GBS may be the same, the parameter settings and processing methods differ. For instance, GBS datasets often experience issues with missing genotypes and uneven marker distributions, making it preferable to use imputation software specifically designed to

handle missing data. In contrast, for WGRS, the focus should be on variant calling accuracy, depth of coverage, and haplotype representation. These factors should influence the choice of pipelines for SNP calling. Some service providers offer these processing steps, which can alleviate the need for extensive raw data processing on the user's part.

WGRS has been successfully applied in crops for high-resolution genome-wide association studies, delivering millions of SNPs, capturing rare variants and structural variation [24,83]. Examples of plant germplasm characterisation studies involving WGRS are listed in Table 3.

Table 3. Examples of WGRS application in germplasm characterisation studies.

Organism	Coverage (Approx.)	No. of Accessions	Institution, Country, Where Germplasm Is Kept	No. of Polymorphic Loci Detected/(Used)	Analyses	Reference
amaranth	not specified	108	US Department of Agriculture Agricultural Research Service genebank, USA	1.4 M	gene flow, selection scan	[101]
avocado	4.69×	205	Avocado 'Plus Tree' Collection; Arangro Plant Nursery; Colombian Germplasm Bank, Colombia	64 M	phylogeny, population structure, racial tracing	[102]
carrot	not specified	630	Germplasm Resources Information Network of the US Department of Agriculture, USA	5.4 M (168 k)	population structure, selection scan, GWAS	[103]
chickpea	12×	3366	International Crops Research Institute for the Semi-Arid Tropics, India; International Center for Agricultural Research in the Dry Areas, Lebanon	23.5 M	GWAS, LD scan, selection scan	[29]
coffee	not specified	90	Choche germplasm bank of the Ethiopian Biodiversity Institute, Etiopia	11 M	phylogeny	[104]
common bean	not specified	144	International Centre for Tropical Agriculture, Colombia; Leibniz Institute of Plant Genetics and Crop Plant Research, Germany; JungleSeeds, Betchworth, UK; Beans and Beans, Horningsham, UK	20.2 M	population structure, phylogeny, GWAS	[105]
cotton	10.85×	240	Zhejiang University, China	3.8 M	phylogeny, population structure, GWAS	[106]
durian		114	cultivations sites in Hainan and Yunnan, China	39 M	diversity assessment, population structure, LD scan, selection scan, core subset selection	[107]
einkorn	not specified	219	Wheat Genetics Resource Center, USA	121 M	phylogeny, population structure	[108]
ginkgo	6.3×	525	Trees growing in multiple locations in China, Japan, Korea USA and Europe	160 M	phylogeny, population structure, selection scan	[109]
grapevine	15.5×	472	Chinese Academy of Sciences; Chinese Academy of Agricultural Sciences, China; Karlsruhe Institute of Technology, Germany	38.7 M	phylogeny, population structure, LD scan, pedigree analysis, selection scan, GWAS	[110]
hemp	10×	110	Vavilov Institute of Plant Genetic Resources, Russia; various companies	12 M	phylogeny, population structure, selection scan	[111]
lettuce	18.8×	445	Centre for Genetic Resources, the Netherlands	208 M	phylogeny, population structure, selection scan, GWAS	[112]

Table 3. Cont.

Organism	Coverage (Approx.)	No. of Accessions	Institution, Country, Where Germplasm Is Kept	No. of Polymorphic Loci Detected/(Used)	Analyses	Reference
napier grass	15–20×	450	International Livestock Research Institute, Ethiopia; Embrapa, Brasil; US Department of Agriculture, USA; Kenya Agricultural and Livestock Research Organization, Kenya; Lanzhou University, China	170 M (1 M)	diversity assessment, GWAS	[113]
pepper	14.7×	500	U.S. National Plant Germplasm System, USA; Hunan Academy of Agricultural Science, China	1005 M (29 k)	phylogeny, population structure, selection scan	[114]
<i>Populus cathayana</i>	32.3×	438	Chinese Academy of Forestry, China	12.3 M	population structure, selection scan, GEA analysis	[115]
rye	10×	116	Germplasm Resources Information Network, USA; Institute of Crop Science, Chinese Academy of Agricultural Sciences, and other collections	908.6 k	phylogeny, population structure, selection scan	[116]
rye	not specified	94	Several genebanks, universities and breeding companies	2.5 M	gene variants	[26]
soybean	not specified	684	Institute of Plant Biology and Biotechnology, Kazakhstan; Guangzhou University, China	8 M (81 k)	phylogeny, population structure	[117]
tobacco	13×	437	Yunnan Academy of Tobacco Agricultural Sciences, China	2.2 M	phylogeny, population structure, gene flow, GWAS	[118]
tomato	not specified	295	Polytechnic University of Valencia, Spain	28 M (18 M; 8.8 M; 162 k)	phylogeny, population structure, selection scan, LD scan, GWAS	[119]
quinoa	7.8×	303	Leibniz Institute of Plant Genetics and Crop Plant Research, Germany; U.S. National Plant Germplasm System, USA	2.9 M	phylogeny, population structure, LD scan, GWAS	[25]

2.3. SNP Genotyping Arrays

High-density DNA genotyping arrays are a hybridisation-based method. DNA samples hybridise to matching oligonucleotide probes attached to the array surface. The detection of a genotype at a given SNP (AA, BB, or AB) is based on emitted fluorescence. SNP arrays allow for genotyping at a fixed set of selected, previously discovered SNPs. Array designs of various densities are available, such as 6 K, 15 K, 60 K, 90 K, or even 1.4 M [51,120–122]; therefore, suitable genotyping density can be chosen to match the specific project objectives. While the DNA genotyping arrays were strictly developed to detect a single type of polymorphisms—SNPs—detection of copy number variation (CNV) and presence–absence variation (PAV), as well as analyses of ploidy and aneuploidy are also possible [122–125].

Design of a genotyping array is a considerable effort, since very large amounts of sequencing data from multiple accessions of a species are necessary to select informative SNPs, whereas factors such as their genomic distribution, minor variant frequency and location relative to genes and non-coding regions should be taken into account [121,126]. For example, to obtain sufficient data for the development of Citrus Genotyping arrays, first, whole-genome sequencing of 41 diverse accessions of *Citrus* and its close relatives was carried out. Then, predominantly genic SNPs from both the nuclear and chloroplast genomes, which were accurately called in *Citrus* and in related species, were selected for

the 1.4 M SNP Axiom HD Citrus Genotyping array. Following validation of the HD array, a subset of SNPs was selected for the creation of a 58 K SNP array for more cost-effective genotyping in studies where this lower density is sufficient [121]. Similarly, in pigeonpea, the array development was preceded by WGRS of 105 accessions. This resulted in the discovery of 2.0 M variants from which 56 K were selected for an Axiom array [120].

For SNP arrays, genotype calling is usually performed using specialised software like GenomeStudio from Illumina (https://support.illumina.com/array/array_software/genomestudio/downloads.html, access date: 2 December 2025) or Axiom Analysis Suite Software 4.0 from Thermo Fisher (Waltham, MA, USA). After the initial calling, quality control and filtering using tools like PLINK [50] or VCFtools [49] are performed.

The same tools used for imputation in GBS and WGRS datasets are also applicable for SNP array data [41,97,99]. For instance, the Infinium Wheat-Barley 40 K SNP array was specifically created to make downstream imputation easier and demonstrated over 99% genotype concordance and less than 5% missing data [127]. The Plant-ImputeDB project [128] compiled multi-species reference panels for plants and showed that genotype imputation can increase SNP density. In rice, an imputation platform combining both SNP-array and whole-genome resequencing datasets, improved the integration of genetic resources and improved the number of usable SNPs for downstream analysis [129].

To date, various genotyping arrays have been developed, mainly for crops with higher economic importance, such as maize (MaizeSNP50 BeadChip [130]), wheat (Wheat 660K SNP array [131]), rice (Rice3K56 [132]), soybean (SoySNP50K [133]), and pigeonpea (Axiom Cajanus 56K [120]).

However, for many plant species, especially for the orphan crops, the requirement of the very large initial financial and computational investment necessary for array development is too prohibitive, and these genotyping options remain unavailable [51]. An important issue to consider when using genotyping arrays is ascertainment bias, which is caused by the selection of SNPs based on analyses of relatively small and not sufficiently diverse populations. The degree of the ascertainment bias is influenced by the size of the population used to discover and select SNPs for the array development (ascertainment or discovery panel) [134,135]. The ascertainment bias can limit the usefulness of SNP arrays in the detection of rare variants in diverse germplasm and identification of introgressions from distant relatives [51]. Furthermore, it was shown that the use of an array with an ascertainment bias results in distorted outcomes of population diversity and structure analyses [134,136]. Other problems, associated mostly with the earlier SNP arrays (developed at a time when the information on the genome sequence, and especially genomic locations of SNPs in the crop in question was not sufficiently detailed) are uneven genomic distribution and redundancy of the SNPs selected for the array design [124]. To mitigate these issues, newer versions of genotyping arrays were developed for various species based on SNP discovery in diverse and large germplasm sets, taking into account the growing knowledge about the genomic distribution of the SNPs and their performance in germplasm characterisation studies. For example, in wheat the TaNG SNP array was developed to address shortcomings of the earlier wheat SNP genotyping arrays: 35 K Wheat Breeders' array and the Illumina 90 K array—related to the use of small and not sufficiently diverse discovery panels and lack of precise information of genomic SNP location when those arrays were developed—and, at the same time, to take advantages of the advances in the array technology, allowing for a substantial increase in the number of probes in the array. To reach this purpose, skim sequencing data from a total of 315 elite wheat lines and landraces were generated to identify novel SNPs and a haplotype optimisation approach was used to select SNPs for inclusion in the new array. Validated SNPs from previous arrays were also incorporated into the new array. Improved genomic coverage was achieved based on the

markers' positions in the wheat reference genome (IWGSC RefSeq v1.0). The final version of the TaNG SNP array was demonstrated to provide superior results in GWAS compared to the 35 K Wheat Breeder's array [124] (more details in Section 3).

Once a genotyping array is developed, it constitutes an attractive tool, since in comparison to GBS and WGRS, the computational requirements for data processing are lower, there is also lower missing data and error rate [121,122]. Furthermore, the use of a fixed genotyping array facilitates comparisons across studies. Examples of plant germplasm characterisation studies involving high-density DNA genotyping arrays are listed in Table 4.

Table 4. Examples of high-density SNP-genotyping array applications in plant germplasm characterisation studies.

Organism	Array Name	No. of Accessions	Institution, Country, Where Germplasm Is Kept	No. of Polymorphic Loci	Analyses	Reference
camellia	Camelia21K	69	Camellia Germplasm 479 Resource Conservation Center of the Research Institute of Subtropical Forestry, China	19.3 k	phylogeny, population structure, GWAS	[137]
citrus	1.4 M SNP Axiom® HD Citrus genotyping array	196	Citrus Variety Collection, USA	729 k	population structure	[121]
citrus	58 K Axiom® Citrus genotyping array	871	Citrus Variety Collection, USA	43 k	phylogeny, population structure	[121]
cowpea	Illumina Cowpea iSelect Consortium Array	2201	International Institute of Tropical Agriculture, Nigeria	48 k	population structure, LD scan, GWAS	[138]
maize, teosinte	Illumina MaizeSNP50 BeadChip	1172	maize breeding programs of the International Maize and Wheat Improvement Center (Mexico), China, USA, Thailand, and Peru	42.2 k	phylogeny, population structure, selection scan, GWAS	[130]
oat	iSelect 6 K-beadchip	288	USDA National Small Grain Collection, USA	2213	population structure, LD scan, GWAS	[139]
pigeonpea	Axiom Cajanus SNP array	103	International Crops Research Institute for the Semi-Arid Tropics, India	51.2 k	phylogeny, population structure	[120]
rice	Rice3K56	192	Anhui Agricultural University, China	not specified	phylogeny, varietal identification, GWAS	[132]
soybean	SoySNP50K	286	United States Department of Agriculture, USA	47 k	selection scan	[133]
strawberry	Istraw35	891	The strawberry breeding program at Fresh Forward B.V., Huissen, The Netherlands	30 k	core subset selection	[140]
wheat	TaNGv1.1	908	Germplasm Resources Unit at the John Innes Centre, UK; USDA Germplasm Resource Information Network, USA; Nations BioResource Project-Wheat genebank, Japan	42.5 k	linkage mapping, CNV analysis, GWAS	[124]

2.4. Choosing the Right Genome-Wide Genotyping Platform and Other Considerations

With a plethora of methods currently available, selecting the appropriate approach is far from trivial. Various factors—such as the scientific question, budget, species ploidy, and genome complexity—play a decisive part in the decision-making process (Figure 1). In this context, Table 5 presents the main factors to consider and how suitable each method is. In this section, we will discuss some of these factors.

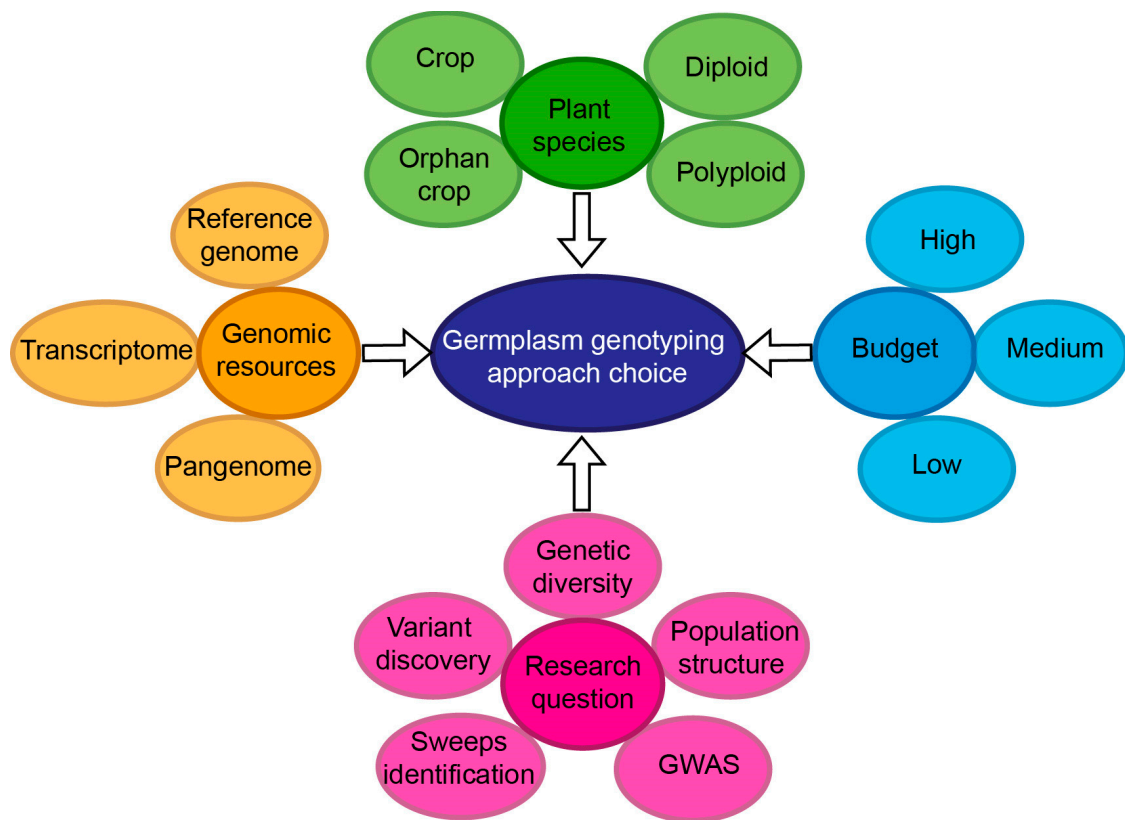


Figure 1. Factors influencing the choice of the genotyping approach.

Table 5. Decision table for choosing an appropriate genotyping method.

Factors	Methods			Comments
	RRS	WGRS	SNP arrays	
Main research objectives				
1. Diversity / population structure	++++	++++	++ *	* May not work for genetically diverse populations. * RRS enables partial
2. Novel variant discovery	++ *	++++	+ **	variant discovery. ** SNP arrays enable none. * Too expensive because a large number of samples are needed for GWAS.
3. GWAS/Genome prediction	+++	++ *	++++ **	** The best choice due to high reproducibility and low missing data.
4. Selection scans	++ *	++++	+ **	* Weak for haplotype-based scans. ** Ascertainment bias.
Available resources				
1. Reference genome	++++	++++	++++	* No reason to choose it over arrays unless you want more discovery.
2. SNP arrays	++ *	+++ **	++++	** Choose it only when full genome resolution is required.
3. None	++++	++ *	-	* A de novo genome assembly is required

Table 5. Cont.

Factors	Methods			Comments
What is the available budget?				
1. Low	++++	-	++++ *	* If SNP arrays are available * Lower coverage WGRS.
2. Medium	++++	++ *	++++ **	** Cost depends on marker density and whether the array is commercial or custom.
3. High	++++	++++	++++	When budget is not a concern, the choice of method depends primarily on the study purpose and sample size.
What is the species' ploidy?				
1. Diploid	++++	++++	++++	Diploid species tolerate all methods well * RRS is only reliable with fully aware polyploidy pipelines.
2. Polyploid	+++ *	++++	++ **	** SNP arrays are reliable only if designed for that ploidy level.
Cross-study comparability	++ *	+++ **	++++ ***	* Missing data causes low overlap between SNPs in different studies. ** Only if using the same reference and pipeline. *** Data is consistent across labs, years, and experiments.

++++ = Best option with the highest performance; +++ = Very good option with minor limitations; ++ = Good/acceptable option but not optimal; + = Usable but least preferred; - = Not recommended/poor. *, **, *** refer to the information in the 2nd, 3rd and 4th column.

The availability of genomic resources will vary depending on the researcher's plant species of interest. Thus, the selection of plant species often dictates the options of genotyping methods at hand.

For instance, genotyping rice—a diploid plant with a small, high-quality, well-annotated genome (~450 Mb) and several genomic resources available—germplasm might be a less complicated task than genotyping rye, which has limited genomic resources [141]. A scientist evaluating the population structure of a rice germplasm bank could select SNP arrays, RRS, or WGRS approach to identify polymorphisms. On the other hand, rye scientists might be constrained to RRS genotyping and a few available SNP arrays (rye5 K and 600 K arrays) [142,143], since the large, repetitive rye genome restricts the use of WGRS in this plant species. Pseudocereals and orphan crops with limited genomic resources might face a similar scenario. For example, RRS genotyping was the most suitable approach to evaluate Barnyard millet (*Echinochloa* spp.)—a popular crop in southern and eastern Asia—germplasm diversity since no reference genome or SNP array was available [144,145]. The polyploid nature of Barnyard millet, similar to other orphan crops such as little millet (*Panicum sumatrense*), has held back the development of reference genomes [146]. The previous examples illustrate that the genome complexity and ploidy of a plant species strongly influence the choice of genotyping method. Polyploidy also occurs in several economically important crops such as wheat, oats, potatoes, cotton, and sugarcane [147]. A major challenge in genotyping polyploid species is to distinguish between true SNPs (allelic variation within individuals) and homeologous SNPs (polymorphism among subgenomes within the same individual) [147]. Also, the presence of paralogous loci—duplicated loci resulting from whole-genome duplication—further complicates SNP calling [148,149]. A

researcher can increase the correct variant calling rate in a polyploid species either by bioinformatic or sequencing means. Sections 2.1 and 2.2 of this review discuss some of the tools utilised in SNP calling for polyploid plants.

Long-read sequencing platforms, such as Pacific Biosciences (PacBio) and Oxford Nanopore Technology (ONT), can address the SNP-calling accuracy issue from two distinct angles: reference genome assembly and polymorphism discovery [48,147]. An incomplete or low-quality genome reference containing gaps or sequencing errors leads to poor read mapping [48]. The average read size obtained from both sequencing platforms (25 Kb and 500 Kb for PacBio and ONT, respectively) facilitates the assembly of repetitive regions, generating highly contiguous genomes [150]. Unlike short-read-based reference genomes, long-read genome assemblies provide a more precise landscape of paralogous loci, increasing the accuracy of variant calling. Furthermore, a haplotype-resolved genome assembly can be obtained by combining long-read sequencing and optical mapping or chromatin conformation capture (Hi-C) techniques [151].

The researcher should keep in mind that a single reference genome does not provide the complete picture of genetic diversity within a plant species, even with an available high-quality chromosome-level genome assembly [48,150,152]. The use of pan-genomes—a catalogue of core and accessory sequences from a single species—can capture the complete and accurate genetic variation comprised within a plant (either diploid or polyploid) germplasm [153–155]. From the polymorphism discovery standpoint, long-read sequencing permits the identification of SVs that would otherwise go undetected by short-read genotyping approaches [156].

The computational infrastructure should also be taken into account when selecting the genotyping strategy. For instance, determining the population structure of an orphan crop germplasm through WGRS will demand a higher computation capability than an RRS approach [48].

Irrespective of the platform choice, one of the key aspects of any genotyping project is data sharing in conformance with the community accepted, standardised data type formats, to make the data available for reuse in research by the global scientific community. Adherence to FAIR (Findable, Accessible, Interoperable, Reusable) principles [157] is recommended. As a rule, a deposition of the generated data in a recognised publicly available repository is requested before publication of the study results in a scientific journal. An example list of recommended repositories for different data types can be found at <https://journals.plos.org/plosone/s/recommended-repositories> (accessed on 14 November 2025). For instance, the NCBI Sequence Read Archive (SRA) repository is used for deposition of raw sequencing data, while information on specific genetic variants in a given sample can be submitted (as a VCF file) to the European Variation Archive (EVA). Multiple other repositories, also cross-disciplinary ones, are available.

3. Applications of Genome-Wide Genotyping Data in Germplasm Characterisation

Genotyping has become an essential tool in germplasm characterisation and management of genetic resources [6]. Depending on the density and the availability of other datasets (such as results of phenotypic evaluations), genome-wide genotyping data can constitute a basis for a wide range of applications. Genome-wide, high-density genotyping methods deliver very detailed genotypic profiles (composed of hundreds and often tens of thousands of markers covering each chromosome of the genome). Based on such detailed genotypic profiles an evaluation of the genetic diversity levels and their distribution in a given germplasm set can be performed with unprecedented precision to clarify and understand the genetic relationships among accessions. It is possible to determine the population

structure—to detect the presence of genetically distinct groups and indicate accessions belonging to each of these groups. When additional data (such as a least passport data) is at hand, the primary determinants of the population structure can be identified, for instance domestication/improvement status, geographic origin [64,77,81], or specific phenotype (for example row type or grain cover in barley [64]).

Application of high-density genotyping methods plays a tremendous role in addressing the challenge of redundancy and error correction in germplasm collections by providing a sufficient number of datapoints to confidently identify groups of duplicates or spot discrepancies between the genotypic information and passport data. It is estimated that unique accessions constitute around 37% of genebank collections worldwide and the unintentional redundancy remains a problem to be addressed to optimise the use of often limited resources available for maintenance and characterisation of genebank collections [1,7]. Recently, based on the GBS genotyping of the barley collection of the Gatersleben genebank, the level of redundancy was found to be around 33% and higher than previously estimated based on passport data [64], while a study using DArTseq successfully identified groups of potentially redundant cassava accessions in the International Center for Tropical Agriculture (CIAT) genebank [7]. These examples underscore the suitability of affordable high-density genotyping (RRS) in the identification of duplicates in both self-pollinated and highly heterozygous crops.

Depending on the study design (if multiple individuals per accession are analysed), the level of homo/heterogeneity of the accessions can be assessed [6]. The within-accession diversity can have a considerable extent, even exceeding the intra-accession diversity, especially in outcrossing species (such as rye or pearl millet [33,158]), but even in self-pollinators, such as barley [64]. Results of such assessments have important implications for the choice of suitable sample sizes and protocols for the periodical regeneration of accessions during gene bank preservation to maintain their genetic integrity.

The comparative analyses of genetic diversity and population structure outcomes and passport data help identify and correct sample tracking errors, for instance, when accessions labelled as “wild” cluster with improved germplasm and are genetically distant from other wild accessions, as was the case in large scale studies of wheat and barley genebank collections, comprising tens of thousands of accessions [64,81]. High-density genotyping data are also used in genetic purity testing, a process that is aimed at the detection of mislabelling, contamination, or residual heterogeneity in seed lots, breeding lines, and hybrid progeny and helps to ensure that these materials are genetically true to their claimed identity [31,159].

Genome-wide high-density genotyping methods are also crucial in supporting the creation of core collections. Core collections represent a subset of germplasm that captures the maximum genetic diversity with minimal redundancy and are essential for efficient conservation, management, and utilisation of genetic resources [140,160].

Another application of high-resolution genotyping data is a selective sweep identification, i.e., detection of genome regions that were subjected to strong positive selection pressure during domestication or plant improvement [77,161]. It provides insights into the evolutionary history of the species and helps to identify loci targeted by selection, which are usually associated with key traits, such as fruit/seed size or seed dispersal mechanisms [77]. The use of high-density genotyping led to a revision of the earlier assumptions about the influence of domestication on the plant genome. While early studies, relying mostly on biparental population and QTL mapping approach, identified few major effect domestication loci [162,163], later population genetics studies using high-density genome-wide genotyping (RRS, WGRS, SNP arrays) revealed that selection pressure impact numerous regions dispersed across the genome and containing hundreds of candidate

genes [81,109,164]. Although, as already mentioned in Section 2.1, RRS approaches miss certain genome regions, which might contain genes targeted by specific selection pressures, many population studies employing RRS methods, (such as [77,81,165,166] successfully identified selective sweep regions and candidate genes). However, an important consideration when planning selective sweep detection, is that, similarly like in many other population genetics endeavours, the outcomes depend strongly not only on the type and number of the markers used but also on the composition and size of the germplasm set and on the sweep/outlier detection method/algorithm, hence the results of independent selection scans carried out in the same species often differ [119,167,168]. For example, in tomato, both WGRS and SNP array-based studies [169,170] on domestication identified two stages in the process: domestication stage and improvement stage, with *S. pimpinellifolium* L. as the wild ancestor and *S. lycopersicum* L. var. *cerasiforme* as the intermediate species. Razifarad et al. [119] revised this hypothesis using WGRS and a broad collection of *S. pimpinellifolium* L. and *S. l.* var. *cerasiforme* accessions, and found, among others, that the origin of *S. l.* var. *cerasiforme* predated domestication (more details in Section 4).

When the phenotyping data are available, high-density genotyping data of a diverse germplasm set can be used to perform Genome Wide Association Studies (GWAS) to identify causative genes/polymorphisms for a given relevant phenotype and to identify genetic markers associated with key agronomic traits, facilitating marker-assisted selection and crop improvement [83]. Multiple examples of successful GWAS experiments carried out with the use of high-density genotyping data were published to date and several of them were mentioned in Section 4 of this review. However, sometimes the outcomes of the approach are not satisfactory. In some cases, the lack of success can be explained by the shortcomings of the set of markers used for the genotyping. For example, in a comparative analysis, GWAS involving genotyping data generated with the already mentioned (Section 2.3) 35 K Wheat Breeder's array failed to identify a significant association for any of the three traits: heading data, response to leaf rust, response to stem rust, while the application of the TaNG array-generated genotyping data resulted in the identification of significant associations for each of the analysed traits. While SNPs for the 35 K Wheat Breeder's array were identified in exome capture data of 43 wheat accessions of various ploidy and, in consequence, a strong ascertainment bias due to the small size of the discovery panel and preferential location of identified SNP in the genic regions was found in subsequent studies involving the use of this array [124,171], the TaNG array was developed based on WGRS data of 315 diverse accessions and provided a better genome coverage [124]. On the other hand, over the last decades it has become apparent that PAV and SV contribute considerably to the genetic and phenotypic diversity of the plant species [172]. In consequence, the use of a single reference genome during GWAS is frequently associated with 'missing heritability'—a situation when identified marker-trait associations explain only a fraction of phenotypic variation [173]. A study comparing three single-genome references representing three different maize germplasm pools demonstrated that the choice of a reference genome impacts the GWAS outcomes [174]. The availability of a pangenome is a tremendous advantage in such situations. In tomato, the use of a graph pangenome resulted in a 24% increase in estimated heritability in comparison to the single reference genome [173].

4. Examples of High-Resolution Genotyping Data Applications in Selected Plant Groups

4.1. Cereals

A very exhaustive and comprehensive characterisation of wheat germplasm was carried out by Sansaloni et al. [81] based on GBS (DArTseq) genotyping. The study involved

80,000 accessions—hexaploid and tetraploid wheats as well as wild relatives from CIMMYT (Centro Internacional de Mejoramiento de Maíz y Trigo—International Maize and Wheat Improvement Center) and ICARDA (International Center for Agricultural Research in the Dry Areas) genebanks. In total, ca. 40 to 50 k polymorphic SNP loci (filtering criteria: missing rate ≤ 0.5 , minor allele frequency (MAF) ≥ 0.001), were identified within each germplasm group. Data analysis revealed the presence of distinct clusters within the hexaploid wheats and indicated landraces with diversity that is yet to be explored in modern breeding. In tetraploid wheat, a group of ca. 1000 accessions was identified, which were probably misclassified as tetraploids and are in fact hexaploids. The genotyping data were also used to select accessions for core collections constituting ca. 20% of the original germplasm sets, discovery of selection footprints, and, with inclusion of phenotyping data for 3787 accessions, for identification of loci associated with grain protein content and sodium dodecyl sulphate sedimentation (indicative of gluten quality) via GWAS.

WGRS study of an einkorn (*T. monococcum*) diversity panel comprising 219 accessions resulted in the identification of over 121 M of high-quality SNPs. It revealed the existence of three genetic clusters within wild einkorn, corresponding to races α , β and γ , a fairly high diversity within domesticated einkorn and provided support for the hypothesis that einkorn was domesticated from a population related to β race accessions [108].

A breakthrough study by Milner et al. [64] reports on the genetic characterisation via GBS of the whole barley germplasm collection of the IPK Gatersleben genebank and is an excellent example of genebank genomics and its potential. The genotyping of the collection comprising both wild and domesticated barleys (22,626 accessions in total) identified over 170 k polymorphic SNPs (missing rate $< 10\%$, MAF $< 1\%$) and revealed that population structure in barley is largely influenced by the geographic origin, and also by the winter or spring growth habit and morphological characters related to end-use quality (number of rows, and the presence or absence of grain cover). Furthermore, a higher-than-expected level of genetic redundancy in the collection (arising from the presence of duplicate accessions), the extent of within-accession diversity, and errors in passport data were discovered.

A very diverse set of 478 rye (*Secale cereale* L.) accessions was genotyped using DArT-seq [77]. As a result, phylogenetic relationships within the *Secale* genus were revealed based on 12.8 k high-quality SNPs (missing data $< 10\%$, MAF > 0.01 , reproducibility $> 95\%$), as well as the presence of genetic clusters within cultivated ryes. The study indicated the potential of landraces as a source of new variation for rye breeding and identified putative loci under selection in cultivated germplasm. WGRS of 116 rye accession of worldwide origin and various domestication status resulted in the unravelling of the domestication history of rye, based on 908.6 k SNPs [116], while in another study [26] involving resequencing of 94 diverse rye accessions, which focused mainly on phosphate transporter genes, 820 SNPs within these gene families members, including 12 putatively deleterious variants were identified (out of the total number of 2.5 M SNPs).

In oat (*Avena sativa* L.), iSelect 6 K-beadchip was used to analyse a germplasm set comprising almost 290 accessions (mostly landraces) of worldwide origin. Based on 2213 high-quality SNPs, a presence of a relatively strong population structure reflecting geographic origins was discovered, conversely to earlier studies of oat genetic diversity. GWAS involving phenotypic data, available in the Germplasm Resources Information Network (GRIN), identified nine SNPs significantly associated with grain hull type and lemma colour [139]. An interesting approach was used to investigate diversity and population structure in a global collection of cultivated and wild hexaploid oat accessions. GBS data from 15 published and unpublished studies were combined, curated and reanalysed which resulted in the identification of 19,928 SNPs (missing data $< 20\%$, MAF $\leq 1\%$ and

heterozygosity $\geq 5\%$) segregating in 8816 oat taxa. Numerous distinct subpopulations were identified in the analysed set: four subpopulations of wild species *A. sterilis*, a subpopulation of cultivated *A. byzantine* and 16 subpopulations comprising mostly cultivated *A. sativa* accessions. The study also provided support for the role of large-scale chromosome translocations and inversions in shaping population structure and in local adaptation [72].

The Illumina MaizeSNP50 BeadChip, comprising ca. 56 k SNPs, was used for characterisation of 982 maize inbred lines and 190 accessions of teosinte—the wild progenitor of maize. The study revealed phylogenetic relationships of teosinte species, ca. 400 selective sweeps related to maize domestication and 360 adaptive sweeps, resulting from the cultivation of domesticated maize in regions with different environmental conditions (tropical vs. temperate) [130].

Recently, a new and improved rice SNP genotyping array Rice3K56 was developed to address limitations of several previous rice SNP genotyping arrays, such as insufficient marker density or genome coverage. The design of this array was based on WGRS data of over 3 k rice accessions of worldwide origins and included 56,606 high-quality SNPs. The performance of the new array was tested on 192 rice accessions representing both *indica* and *japonica* types and diverse geographic origins. The array turned out to be suitable for varietal identification even when closely related accessions were studied. Similarly, satisfactory outcomes were achieved when GWAS was performed with the new array—over 100 highly significant SNPs associated with 13 traits were identified [132].

4.2. Pseudocereals

In recent years, quinoa inbred lines, diverse panels, and wild relatives have been genotyped through GBS and WGRS. For instance, Mizuno et al., [80] analysed 136 inbred lines—using a GBS approach—and identified 5763 SNPs (missing rate < 0.2 , MAF > 0.05) showing that these inbred lines fall into three subpopulations (High—Northern and Southern—and lowlands). In another study, Patiranage et al. [25], genotyped a diverse panel of 303 quinoa accessions and seven wild relatives using the WGRS strategy. The study revealed 2.9 M SNPs and associated 600 SNPs with 17 agronomically important traits.

Amaranth is the pseudocereal with genomic tools available, including dedicated genomic databases such as AmaranthGDB [175] and AmaranthGRD [176]. In *Amaranthus*, WGRS of 108 domesticated accessions and wild relatives identified 1.4 M SNPs and aided in elucidating gene flow patterns between domesticated, locally adapted, and wild relatives [101]. Chauhan et al. [60] analysed the population structure of an *Amaranthus* diverse panel, comprising 192 accessions from different parts of the world, using a GBS approach. The study generated 41,931 SNPs (missing rate $< 20\%$, MAF $> 5\%$) and identified three subpopulations in the accession panel.

4.3. Tuberous Plants

A collection of 730 accessions from the US Potato genebank, representing various species, was interrogated using GBS [76]. Sets of segregating SNPs ranging from 4.7 to 7.8 k (missing data $< 10\%$, MAF > 0.02) were used for analyses depending on the germplasm set. Higher heterozygosity levels were found in tetraploid potatoes compared to diploid potatoes. Although an overall low population structure was observed, a distinction between wild and cultivated accessions was apparent. Also, indications of introgressions from *S. bolivense* into the cultivated potatoes were found. To support breeding efforts, core subsets consisting of 329 accessions were identified using two algorithms.

DArTSeq genotyping was performed on a collection of 100 winged yam breeding lines [82], and as a result, almost 7 k segregating SNPs (no missing data, MAF = 0.05)

were identified. In parallel, phenotyping evaluation involving 24 traits was carried out. Diversity analyses indicated a considerable variation level in the analysed set and the presence of three distinct groups, which is expected to support the selection of parental lines for maximising heterosis in yam breeding.

The largest collection of cassava (*Manihot esculenta*) germplasm, safeguarded at the International Center for Tropical Agriculture (CIAT) cassava genebank was genotyped using the DArTseq method to optimise the detection of genetic redundancy. The quality of DNA samples and of DNA markers ($MAF \geq 0.001$, call rate ≥ 0.8) was taken into consideration while defining genetic distance thresholds to identify subsets of genetically distinct accessions. In the characterised set of 5301 accessions (95% of the whole CIAT collection) ca. 2500 (47%) distinct genotypes were identified, with clusters of putatively redundant accessions counting as many as 87 entries [7].

4.4. Legumes

A comprehensive chickpea genomic diversity study was performed based on WGRS of 3366 wild and cultivated accessions. A total of 23.51 M SNPs were identified. One of the many outcomes of the study is the identification of 205 SNPs associated with 11 traits. This was achieved by combining phenotypic data of almost 3000 cultivated accessions with genotypic information from almost 4.0 M SNPs. Based on the genomic locations of these SNPs, 79 genes with potential roles in the determination of seed size and development and 24 superior haplotypes for 20 of these genes were identified. Selective sweep detection was also performed, and resulted, among others, in the identification of 37 genes potentially involved in adaptation of chickpea to different cultivation environments. Notably, the study involved the construction of a chickpea pangenome comprising over 3000 individuals [29].

Genotyping data from 2201 accessions from the cowpea core collection, held at the International Institute of Tropical Agriculture (IITA), generated using the Illumina Cowpea iSelect Consortium Array and comprising almost 50 k segregating SNPs, was used to analyse genetic diversity and population structure. In total, 130 groups of putatively identical accessions (comprising up to 15 accessions) were identified. The presence of two major clusters corresponding to geographic origins (West and East Africa) was discovered, with the West African germplasm exhibiting the highest diversity. A successful confirmatory GWAS of seed coat pigmentation patterning was also performed to demonstrate the utility of the collection and accompanying genotyping data for candidate gene identification [138].

The DArTseq method was applied to characterise an existing *Pisum* core collection established at the Instituto de Agricultura Sostenible (Córdoba, Spain), consisting of 325 pea accessions (landraces, wild species, breeding lines, and commercial varieties), and it yielded a total of 11,511 SNP and 24,279 SilicoDArT (presence-absence) markers (missing data < 20%, $MAF > 5\%$ and heterozygosity < 10%). Phylogeny analysis revealed six distinct genetic clusters in the collection and provided support for the classification of *Pisum* into two species *P. flavum* and *P. sativum*. High admixture levels were observed, which inferred continuous genetic exchange among populations. *P. sativum* subspecies: *jomardii* and *arvense* were found to act as introgression channels of wild alleles into cultivated peas during domestication [74].

Another RRS variant, GBS, was used in faba bean to analyse, among others, a diversity panel consisting of 217 accessions, mostly domesticated, representing various geographic origins, and to perform GWAS. Almost 40 k high-quality SNPs were identified (missing rate < 10%, heterozygosity < 10%), and significant marker-trait associations for each of the three traits analysed were found, confirming the suitability of GBS data for this approach [52].

In soybean, WGRS was performed to provide a comparative characterisation of accessions from Kazakhstan and a global germplasm set (684 accessions in total). Almost 81 k high-quality SNPs (missing data < 10%, MAF > 5%) were identified. Population structure and genetic diversity analyses clearly separated wild accessions from the domesticated ones, revealed proximity of Kazakhstani accessions to cultivars from Europe and North America, and, at the same time, a narrow genetic base of Kazakhstani germplasm, providing important cues for local breeding programs [117].

4.5. Oil Plants

A study of genetic diversity in South-Central African oil palm germplasm (478 individuals) was performed using GBS. In total, 7048 high-quality SNPs (missing data < 20%, MAF > 0.01) were identified. Population structure analysis indicated the presence of six populations; among them, the Nigerian subpopulation was found to be the most diverse. Finally, 96 palms (individuals), capturing the majority of present alleles, were selected to form a core collection [73].

To aid GWAS in canola (*Brassica napus* L.), a diversity panel consisting of 433 primarily winter populations originating predominantly from North America and Europe was assembled and genotyped using GBS. In total, 251.5 k SNPs (MAF > 0.05), which mapped to the canola reference genome were identified. A considerable genetic diversity of the population was revealed, as well as the presence of genetic clusters. Phenotypic evaluation confirmed ample variation for the several traits evaluated in the panel, suggesting it is suitable for the detection of marker-trait associations [66].

4.6. Vegetables

In carrot, WGRS of 630 diverse accessions (wild carrots, cultivars, landraces, and outgroups) was performed to analyse the influence of domestication and improvement. Population structure analyses based on over 168 k high-quality SNPs indicated the existence of five subpopulations corresponding to improvement status and geographic origin. Levels of diversity were found to coincide with the improvement status, with the highest diversity in wild carrots. Indications of a bottleneck, probably related to domestication and recent expansion, were found in each subpopulation, except for the wild carrots. In total, 18 selective sweeps related to domestication and improvement were found, bearing genes related to, among others, photoperiodism, control of flower development and flowering time [103].

A very comprehensive study of pepper (*Capsicum* ssp.) genome diversity, based on resequencing of a core collection comprising 500 accessions representing various wild and domesticated species, provided insights into genetic diversity structure, differentiation, and domestication in this genus. Analyses involving 29 k high-quality SNPs (missing data rate ≤ 0.3 , MAF ≥ 0.05) indicated, among others, that the domestications of five cultivated *Capsicum* species occurred independently. Furthermore, selective pressure toward enlarged fruit size and elongated fruit shape targeted distinct genomic regions in different *Capsicum* species [114].

Similarly, in tomato, the WGRS data of 295 *S. pimpinellifolium* L. (wild), *S. lycopersicum* L. var. *cerasiforme* (semidomesticated), and *S. lycopersicum* L. var. *lycopersicum* (cultivated) were a basis to revise the theory regarding its domestication. The population genomics analyses based on over 18 M of SNPs indicated that *S. lycopersicum* L. var. *cerasiforme* (considered to be the semidomesticated intermediate in the evolution of cultivated tomato) exhibited traits associated with cultivated tomatoes, but these traits were lost during the expansion of *S. lycopersicum* L. var. *cerasiforme* forms towards the North and then regained, before the geographic expansion of the cultivated tomato. The study also highlighted the impor-

tance of sufficiently broad sampling and carefulness in defining populations in analyses of domestication history [119].

4.7. Ornamentals

Analyses based on WGRS (at an average depth of ca. 6.3×) of 525 *Ginkgo biloba* genomes from 51 populations across the world resulted in the identification of over 160 M of high-quality SNPs and revealed multiple cycles of population expansions and reductions. Four subpopulations of ginkgo were identified, as well as population subdivision corresponding to geographical distribution within the Chinese germplasm. Selective sweeps, likely resulting from adaptation to environmental changes, and multiple candidate genes were also detected [109].

Iranian Foxtail lily (*Eremurus*) germplasm consisting of 96 accessions representing seven *Eremurus* species was analysed using GBS, which resulted in 3002 high-quality SNPs (missing data < 50%, MAF > 0.1 and < 0.9). Presence of genetic clusters corresponding mostly with the classification into species was found. However, support for the reclassification of two species into subspecies was also provided. Another very useful outcome of this study is the identification of private alleles that can be used to classify foxtail lily at the species level [70].

In *Phalaenopsis*, over 113.5 k of SNPs resulting from GBS genotyping were used to perform GWAS for floral aesthetic traits on a F1 progeny of the cross *P. aphrodite* × *P. equestris* comprising 116 individuals. Ten SNPs for colour-related traits were successfully identified [75].

Using WGRS data from 220 varieties, a 21 K SNP genotyping array was developed to support ornamental *Camellia* breeding. The array proved to be useful in the identification of closely related cultivars and in GWAS. The GWAS analysis was performed for five leaf traits and yielded results consistent with the outcomes based on WGRS data. Population structure analysis of 69 accessions identified four major subgroups in accordance with the available information on the cultivars [137].

4.8. Fruit and Nut Trees

RAD-seq was applied to analyse a collection of 168 cultivated and wild apricot accessions. Assessment of genetic diversity and population structure based on over 418 k high-quality SNPs (missing data < 0.5, MAF > 0.05) identified five subpopulations, gene flow between the subpopulations, and selective sweeps caused by domestication [61].

WGRS of 472 accessions representing 48 *Vitis* species from diverse geographic origins was carried out with the aim of providing support for cultivar improvement by, among others, identification of target genes. In total, over 37.8 M of SNPs and 904.4 k of indels (missing calls < 40%, MAF > 0.005) were identified. The population structure of the collection coincided with the improvement status and geographic origin. Genotypic information was also used to perform pedigree analysis and detect genome regions targeted by selection pressure in wild and domesticated grapevines. In total, over 1.3 k candidate genes were found, with only 18 genes in common for both germplasm groups. The study also resulted in the identification of candidate genes for berry shape, panicle type aromatic compounds, and other traits [110].

A collection of 815 common walnut (*Juglans regia* L) accessions, originating predominantly from China, but also from Iran and Pakistan, was analysed using WGRS resulting in the identification of 16.78 M SNPs (MAF > 0.05). Four genetic clusters were identified within the collection, corresponding to the geographic origin of the samples. The level of genetic diversity varied within subpopulations suggesting a genetic bottleneck in Chinese *J. regia* populations. Additionally, numerous signatures of selections related to adaptation

and improvement were discovered. The study also involved GWAS analysis based on phenotypic evaluation of Chinese walnut accessions. This resulted in the identification of genomic loci related to 18 key agronomic traits, such as linoleic acid content [177].

In cultivated strawberry, genotyping data comprising almost 30 k SNPs (missing data < 10%, MAF > 0.05) was used, alongside pedigree data, to support the development of a core collection from a set of 920 genotypes from the current breeding program. The genomic data was compiled from Istraw35 SNP array genotyping (available for 891 genotypes) and resequencing results (for the remaining 29 genotypes). The final core collection consisted of 192 genotypes [140].

4.9. Others

WGRS with 10× coverage was applied to analyse genomes of 240 *Gossypium barbadense* accessions. In total, over 3.6 M high-quality SNPs and over 220 k indels (missing data ≤ 10%, MAF ≥ 0.05) were identified. The germplasm set under study was found not to be very structured; however, five subgroups could be distinguished. Following GWAS, candidate genes for fibre strength and lint percentage were identified, which could be altered by genetic engineering to speed up cotton improvement [106].

Population genetics analyses based on ca 12 M SNPs (missing data < 30%, MAF > 0.05) identified in WGRS data of 110 cannabis accessions revealed the presence of four distinct clusters in the analysed germplasm and provided insights into the domestication history of the crop, indicating, among others, a single domestication origin of *C. sativa* in East Asia. Multiple candidate genes targeted by selection during the divergence of hemp and drug types were also identified [111].

In tobacco, WGRS of 437 accessions yielded over 2.2 M high-quality SNPs (missing data < 0.1, MAF > 0.05). The germplasm set could be divided into eight subgroups based on population structure analyses. Extensive gene flow between tobacco taxa was also found. GWAS analysis using phenotype data of 379 accessions regarding plant height resulted in the identification of three candidate genes, demonstrating the usefulness of the genomic data resource in assisting molecular breeding [118].

5. Conclusions and Outlook

Genome-wide genotyping has become accessible and affordable to the plant community. Researchers and breeders can use WGRS/RRS/SNP to generate high-density marker data for diversity studies, trait mapping, and genomic selection across a wide range of crops, even at large population sizes. The continuous development of bioinformatics tools permits imputation even in large germplasm collections. The reference genome bias and imputation quality in distant germplasm hinder the accuracy of high-density genotyping. Benchmark methods in crops and cross-species comparisons have shown that errors often appear around heterozygous and poorly mapped sequences [178]. Similarly, low-quality and non-contiguous reference genomes of polyploid and orphan crops contribute to poor read mapping. Requirements regarding computational infrastructure and specialised bioinformatic expertise necessary for data analysis and data storage can be limiting for many potential users.

The extent (in terms of numbers of accessions analysed) and scope of germplasm characterisation studies carried out nowadays with the use genome-wide, high-density genotyping vary very much, depending on the crop (its economic importance, genome size and composition, biology), resources available to the research team, the specific study aim, and other factors—from studies exhaustively characterising whole genebank collections comprising tens of thousands of accessions, to very focused analyses of a relatively small set of unique regional germplasm, addressing a specific research question.

Phenotyping is a limiting factor because manual measurements are still the standard for many traits and for calibrating models and are often difficult to repeat across different laboratories [179]. To address this issue, phenomics has emerged, which refers to the high-throughput phenotyping systems (HTP). HTP techniques combine cutting-edge robotics imaging and aerial or spatial platforms of sensor networks with computational tools to capture plant traits across large populations and time scales. This technology enables researchers to detect phenotypic data; for example, water stress or nutrient deficiency, or disease outbreaks at an early stage [180]. Yet, they require careful calibration, standardised workflows, and spatial genotype and environment interaction ($G \times E$) modelling to transform images into interpretable and biologically relevant traits [181,182]. Phenomics, when integrated with machine learning, enables the detection of high-dimensional patterns in phenomics, genomics, and environmental data. These advanced technologies provide accurate prediction of plant performance, trait classification, and automated image analysis [183]. This combination of technologies accelerates the germplasm evaluation and phenotype–genotype association, which strengthens predictive breeding [184].

Increasingly, various emerging innovative technologies are being incorporated into plant germplasm research. Techniques such as single-cell transcriptomics, spatial transcriptomics, and spatial metabolomics are being applied to analyse various aspects of plant biology [185,186] and help understand the impact of environmental stimuli on the phenotypic plasticity [187]. To fully take advantage of the arising wealth of information and apply it effectively for germplasm improvement, an approach called panomics, which integrates genomics, transcriptomics, metabolomics and phenomics and employs deep learning has been proposed [187].

As presented in the current review, genome-wide high-density genotyping powered by advances in NGS technologies has provided the means to significantly increase the information about genetic variation, allele diversity, and gene pool held within the germplasms of important crops. Such information, coupled with the available phenotype data, has identified several agricultural-trait and stress-tolerant-related genes with potential application in Marker-Assisted Selection (MAS) not only in major staple crops but also in underutilised crops. As more germplasms are thoroughly characterised, we will fully tap into the genetic reservoir of such collections and improve our ability to secure food demands in the ever-changing climate conditions.

Author Contributions: Conceptualisation, Supervision, Funding Acquisition, H.B.-B.; Writing—Original Draft Preparation, S.W., B.W.K., D.C.-R. and H.B.-B.; Writing—Review and Editing, S.W., B.W.K., D.C.-R. and H.B.-B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by an OPUS 19 project 2020/37/B/NZ9/00738 granted by the National Science Centre, Poland.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analysed in this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. FAO Commission on Genetic Resources for Food and Agriculture Assessments. *The Third Report on The State of the World's Plant Genetic Resources for Food and Agriculture*; FAO Commission on Genetic Resources for Food and Agriculture Assessments: Rome, Italy, 2025; ISBN 978-92-5-139675-9.
2. Hedden, P. The Genes of the Green Revolution. *Trends Genet.* **2003**, *19*, 5–9. [[CrossRef](#)]
3. Pingali, P.L. Green Revolution: Impacts, Limits, and the Path Ahead. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 12302–12308. [[CrossRef](#)]

4. Gur, A.; Zamir, D. Unused Natural Variation Can Lift Yield Barriers in Plant Breeding. *PLoS Biol.* **2004**, *2*, e245. [\[CrossRef\]](#)
5. Gamuyao, R.; Chin, J.H.; Pariasca-Tanaka, J.; Pesaresi, P.; Catausan, S.; Dalid, C.; Slamet-Loedin, I.; Tecson-Mendoza, E.M.; Wissuwa, M.; Heuer, S. The Protein Kinase Pstol1 from Traditional Rice Confers Tolerance of Phosphorus Deficiency. *Nature* **2012**, *488*, 535–539. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Mascher, M.; Schreiber, M.; Scholz, U.; Graner, A.; Reif, J.C.; Stein, N. Genebank Genomics Bridges the Gap between the Conservation of Crop Diversity and Plant Breeding. *Nat. Genet.* **2019**, *51*, 1076–1081. [\[CrossRef\]](#)
7. Carvajal-Yepes, M.; Ospina, J.A.; Aranzales, E.; Velez-Tobon, M.; Correa Abondano, M.; Manrique-Carpintero, N.C.; Wenzl, P. Identifying Genetically Redundant Accessions in the World's Largest Cassava Collection. *Front. Plant Sci.* **2023**, *14*, 1338377. [\[CrossRef\]](#)
8. Langridge, P.; Waugh, R. Harnessing the Potential of Germplasm Collections. *Nat. Genet.* **2019**, *51*, 200–201. [\[CrossRef\]](#)
9. Gireesh, C.; Sundaram, R.M.; Anantha, S.M.; Pandey, M.K.; Madhav, M.S.; Rathod, S.; Yathish, K.R.; Senguttuvel, P.; Kalyani, B.M.; Ranjith, E.; et al. Nested Association Mapping (NAM) Populations: Present Status and Future Prospects in the Genomics Era. *CRC Crit. Rev. Plant Sci.* **2021**, *40*, 49–67. [\[CrossRef\]](#)
10. Grzebelus, D. Diversity Arrays Technology (DArT) Markers for Genetic Diversity. In *Genetic Diversity and Erosion in Plants; Sustainable Development and Biodiversity*; Ahuja, M.R., Jain, S.M., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 295–309. ISBN 978-3-319-25637-5.
11. Bhattarai, G.; Shi, A.; Kandel, D.R.; Solís-Gracia, N.; da Silva, J.A.; Avila, C.A. Genome-Wide Simple Sequence Repeats (SSR) Markers Discovered from Whole-Genome Sequence Comparisons of Multiple Spinach Accessions. *Sci. Rep.* **2021**, *11*, 9999. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Biswas, M.K.; Bagchi, M.; Nath, U.K.; Biswas, D.; Natarajan, S.; Jesse, D.M.I.; Park, J.I.; Nou, I.S. Transcriptome Wide SSR Discovery Cross-Taxa Transferability and Development of Marker Database for Studying Genetic Diversity Population Structure of Liliun Species. *Sci. Rep.* **2020**, *10*, 18621. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Zhang, J.; Yan, J.; Huang, S.; Pan, G.; Chang, L.; Li, J.; Zhang, C.; Tang, H.; Chen, A.; Peng, D.; et al. Genetic Diversity and Population Structure of Cannabis Based on the Genome-Wide Development of Simple Sequence Repeat Markers. *Front. Genet.* **2020**, *11*, 958. [\[CrossRef\]](#)
14. Rakoczy-Trojanowska, M.; Bolibok, H. Characteristics and a Comparison of Three Classes of Microsatellite-Based Markers and Their Application in Plants. *Cell. Mol. Biol. Lett.* **2004**, *9*, 221–238.
15. Shavvon, R.S.; Qi, H.L.; Mafakheri, M.; Fan, P.Z.; Wu, H.Y.; Vahdati, F.B.; Al-Shmgani, H.S.; Wang, Y.H.; Liu, J. Unravelling the Genetic Diversity and Population Structure of Common Walnut in the Iranian Plateau. *BMC Plant Biol.* **2023**, *23*, 201. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Lyngkhoi, F.; Saini, N.; Gaikwad, A.B.; Thirunavukkarasu, N.; Verma, P.; Silvar, C.; Yadav, S.; Khar, A. Genetic Diversity and Population Structure in Onion (*Allium cepa* L.) Accessions Based on Morphological and Molecular Approaches. *Physiol. Mol. Biol. Plants* **2021**, *27*, 2517–2532. [\[CrossRef\]](#)
17. Lee, K.-J.; Sebastin, R.; Cho, G.-T.; Yoon, M.; Lee, G.-A.; Hyun, D.-Y. Genetic Diversity and Population Structure of Potato Germplasm in RDA-Genebank: Utilization for Breeding and Conservation. *Plants* **2021**, *10*, 752. [\[CrossRef\]](#)
18. Fayaz, H.; Mir, A.H.; Tyagi, S.; Wani, A.A.; Jan, N.; Yasin, M.; Mir, J.I.; Mondal, B.; Khan, M.A.; Mir, R.R. Assessment of Molecular Genetic Diversity of 384 Chickpea Genotypes and Development of Core Set of 192 Genotypes for Chickpea Improvement Programs. *Genet. Resour. Crop Evol.* **2022**, *69*, 1193–1205. [\[CrossRef\]](#)
19. Larsen, B.; van Dooijeweert, W.; Durel, C.E.; Denancé, C.; Rutten, M.; Howard, N.P. SNP Genotyping Dutch Heritage Apple Cultivars Allows for Germplasm Characterization, Curation, and Pedigree Reconstruction Using Genotypic Data from Multiple Collection Sites across the World. *Tree Genet. Genomes* **2024**, *20*, 21. [\[CrossRef\]](#)
20. Elshire, R.J.; Glaubitz, J.C.; Sun, Q.; Poland, J.A.; Kawamoto, K.; Buckler, E.S.; Mitchell, S.E. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* **2011**, *6*, e19379. [\[CrossRef\]](#)
21. Sansaloni, C.; Petrolí, C.; Jaccoud, D.; Carling, J.; Detering, F.; Grattapaglia, D.; Kilian, A. Diversity Arrays Technology (DArT) and next-Generation Sequencing Combined: Genome-Wide, High Throughput, Highly Informative Genotyping for Molecular Breeding of Eucalyptus. *BMC Proc.* **2011**, *5*, P54. [\[CrossRef\]](#)
22. Davey, J.W.; Hohenlohe, P.A.; Etter, P.D.; Boone, J.Q.; Catchen, J.M.; Blaxter, M.L. Genome-Wide Genetic Marker Discovery and Genotyping Using next-Generation Sequencing. *Nat. Rev. Genet.* **2011**, *12*, 499–510. [\[CrossRef\]](#)
23. Peterson, B.K.; Weber, J.N.; Kay, E.H.; Fisher, H.S.; Hoekstra, H.E. Double Digest RADseq: An Inexpensive Method for de Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS ONE* **2012**, *7*, e37135. [\[CrossRef\]](#)
24. Scheben, A.; Batley, J.; Edwards, D. Genotyping-by-Sequencing Approaches to Characterize Crop Genomes: Choosing the Right Tool for the Right Application. *Plant Biotechnol. J.* **2017**, *15*, 149–161. [\[CrossRef\]](#)
25. Patiranage, D.S.R.; Rey, E.; Emrani, N.; Wellman, G.; Schmid, K.; Schmöckel, S.M.; Tester, M.; Jung, C. Genome-Wide Association Study in Quinoa Reveals Selection Pattern Typical for Crops with a Short Breeding History. *Elife* **2022**, *11*, e66873. [\[CrossRef\]](#)

26. Chan-Rodriguez, D.; Koboyi, B.W.; Werghi, S.; Till, B.J.; Maksymiuk, J.; Shoormij, F.; Hilderlith, A.; Hawliczek, A.; Królik, M.; Bolibok-Bragoszewska, H. Phosphate Transporter Gene Families in Rye (*Secale cereale* L.)—Genome-Wide Identification, Characterization and Sequence Diversity Assessment via DArTseq. *Front. Plant Sci.* **2025**, *16*, 1529358. [\[CrossRef\]](#)
27. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Zhou, Z.; Jiang, Y.; Wang, Z.; Gou, Z.; Lyu, J.; Li, W.; Yu, Y.; Shu, L.; Zhao, Y.; Ma, Y.; et al. Resequencing 302 Wild and Cultivated Accessions Identifies Genes Related to Domestication and Improvement in Soybean. *Nat. Biotechnol.* **2015**, *33*, 408–414. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Varshney, R.K.; Roorkiwal, M.; Sun, S.; Bajaj, P.; Chitkineni, 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 3,366 Genomes. *Nature* **2021**, *599*, 622–627. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Ganai, M.W.; Polley, A.; Graner, E.M.; Plieske, J.; Wieseke, R.; Luerssen, H.; Durstewitz, G. Large SNP Arrays for Genotyping in Crop Plants. *J. Biosci.* **2012**, *37*, 821–828. [\[CrossRef\]](#)
31. Josia, C.; Mashingaidze, K.; Amelework, A.B.; Kondwakwenda, A.; Musvosvi, C.; Sibiya, J. SNP-Based Assessment of Genetic Purity and Diversity in Maize Hybrid Breeding. *PLoS ONE* **2021**, *16*, e0249505. [\[CrossRef\]](#)
32. Park, J.S.; Kang, M.Y.; Shim, E.J.; Oh, J.H.; Seo, K.I.; Kim, K.S.; Sim, S.C.; Chung, S.M.; Park, Y.; Lee, G.P.; et al. Genome-Wide Core Sets of SNP Markers and Fluidigm Assays for Rapid and Effective Genotypic Identification of Korean Cultivars of Lettuce (*Lactuca sativa* L.). *Hortic. Res.* **2022**, *9*, uhac119. [\[CrossRef\]](#)
33. Hawliczek, A.; Bolibok, L.; Tofil, K.; Borzecka, E.; Jankowicz-Cieślak, J.; Gawroński, P.; Kral, A.; Till, B.J.; Bolibok-Bragoszewska, H. Deep Sampling and Pooled Amplicon Sequencing Reveals Hidden Genic Variation in Heterogeneous Rye Accessions. *BMC Genom.* **2020**, *21*, 845. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Sun, X.; Liu, D.; Zhang, X.; Li, W.; Liu, H.; Hong, W.; Jiang, C.; Guan, N.; Ma, C.; Zeng, H.; et al. SLAF-Seq: An Efficient Method of Large-Scale De Novo SNP Discovery and Genotyping Using High-Throughput Sequencing. *PLoS ONE* **2013**, *8*, e58700. [\[CrossRef\]](#)
35. Torkamaneh, D.; Laroche, J.; Bastien, M.; Abed, A.; Belzile, F. Fast-GBS: A New Pipeline for the Efficient and Highly Accurate Calling of SNPs from Genotyping-by-Sequencing Data. *BMC Bioinform.* **2017**, *18*, 5. [\[CrossRef\]](#)
36. Glaubitz, J.C.; Casstevens, T.M.; Lu, F.; Harriman, J.; Elshire, R.J.; Sun, Q.; Buckler, E.S. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS ONE* **2014**, *9*, e90346. [\[CrossRef\]](#)
37. Lu, F.; Lipka, A.E.; Glaubitz, J.; Elshire, R.; Cherney, J.H.; Casler, M.D.; Buckler, E.S.; Costich, D.E. Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel Insights from a Network-Based SNP Discovery Protocol. *PLoS Genet.* **2013**, *9*, e1003215. [\[CrossRef\]](#)
38. Catchen, J.; Hohenlohe, P.A.; Bassham, S.; Amores, A.; Cresko, W.A. Stacks: An Analysis Tool Set for Population Genomics. *Mol. Ecol.* **2013**, *22*, 3124–3140. [\[CrossRef\]](#)
39. Garrison, E.; Marth, G. Haplotype-Based Variant Detection from Short-Read Sequencing. *arXiv* **2012**, arXiv:1207.3907.
40. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernysky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data. *Genome Res.* **2010**, *20*, 1297–1303. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Browning, B.L.; Zhou, Y.; Browning, S.R. A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am. J. Hum. Genet.* **2018**, *103*, 338–348. [\[CrossRef\]](#)
42. Chen, L.; Yang, S.; Araya, S.; Quigley, C.; Taliario, E.; Mian, R.; Specht, J.E.; Diers, B.W.; Song, Q. Genotype Imputation for Soybean Nested Association Mapping Population to Improve Precision of QTL Detection. *Theor. Appl. Genet.* **2022**, *135*, 1797–1810. [\[CrossRef\]](#)
43. Rubinacci, S.; Delaneau, O.; Marchini, J. Genotype Imputation Using the Positional Burrows Wheeler Transform. *PLoS Genet.* **2020**, *16*, e1009049. [\[CrossRef\]](#)
44. Howie, B.; Fuchsberger, C.; Stephens, M.; Marchini, J.; Abecasis, G.R. Fast and Accurate Genotype Imputation in Genome-Wide Association Studies through Pre-Phasing. *Nat. Genet.* **2012**, *44*, 955–959. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Paril, J.; Cogan, N.O.I.; Malmberg, M.M. Imputef: Imputation of Polyploid Genotype Classes and Allele Frequencies. *BMC Genom.* **2025**, *26*, 946. [\[CrossRef\]](#)
46. Clark, L.V.; Lipka, A.E.; Sacks, E.J. PolyRAD: Genotype Calling with Uncertainty from Sequencing Data in Polyploids and Diploids. *G3 Genes | Genomes | Genet.* **2019**, *9*, 663–673. [\[CrossRef\]](#)
47. Swarts, K.; Li, H.; Romero Navarro, J.A.; An, D.; Romay, M.C.; Hearne, S.; Acharya, C.; Glaubitz, J.C.; Mitchell, S.; Elshire, R.J.; et al. Novel Methods to Optimize Genotypic Imputation for Low-Coverage, Next-Generation Sequence Data in Crop Plants. *Plant Genome* **2014**, *7*, plantgenome2014.05.0023. [\[CrossRef\]](#)
48. Phillips, A.R. Variant Calling in Polyploids for Population and Quantitative Genetics. *Appl. Plant Sci.* **2024**, *12*, e11607. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The Variant Call Format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [\[CrossRef\]](#) [\[PubMed\]](#)

50. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; De Bakker, P.I.W.; Daly, M.J.; et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [\[CrossRef\]](#)
51. Rasheed, A.; Hao, Y.; Xia, X.; Khan, A.; Xu, Y.; Varshney, R.K.; He, Z. Crop Breeding Chips and Genotyping Platforms: Progress, Challenges, and Perspectives. *Mol. Plant* **2017**, *10*, 1047–1064. [\[CrossRef\]](#)
52. Zhang, H.; Fechete, L.I.; Himmelbach, A.; Poehlein, A.; Lohwasser, U.; Börner, A.; Maalouf, F.; Kumar, S.; Khazaei, H.; Stein, N.; et al. Optimization of Genotyping-by-Sequencing (GBS) for Germplasm Fingerprinting and Trait Mapping in Faba Bean. *Legume Sci.* **2024**, *6*, e254. [\[CrossRef\]](#)
53. Wickland, D.P.; Battu, G.; Hudson, K.A.; Diers, B.W.; Hudson, M.E. A Comparison of Genotyping-by-Sequencing Analysis Methods on Low-Coverage Crop Datasets Shows Advantages of a New Workflow, GB-EaSy. *BMC Bioinform.* **2017**, *18*, 586. [\[CrossRef\]](#)
54. Zamalutdinov, A.; Boldyrev, S.; Ben, C.; Gentzittel, L. The Evaluation of Different Combinations of Enzyme Set, Aligner and Caller in GBS Sequencing of Soybean. *Plant Methods* **2025**, *21*, 106. [\[CrossRef\]](#)
55. Lou, R.N.; Jacobs, A.; Wilder, A.P.; Therkildsen, N.O. A Beginner's Guide to Low-Coverage Whole Genome Sequencing for Population Genomics. *Mol. Ecol.* **2021**, *30*, 5966–5993. [\[CrossRef\]](#)
56. Lemay, M.A.; Sibbesen, J.A.; Torkamaneh, D.; Hamel, J.; Levesque, R.C.; Belzile, F. Combined Use of Oxford Nanopore and Illumina Sequencing Yields Insights into Soybean Structural Variation Biology. *BMC Biol.* **2022**, *20*, 53. [\[CrossRef\]](#)
57. Wang, N.; Yuan, Y.; Wang, H.; Yu, D.; Liu, Y.; Zhang, A.; Gowda, M.; Nair, S.K.; Hao, Z.; Lu, Y.; et al. Applications of Genotyping-by-Sequencing (GBS) in Maize Genetics and Breeding. *Sci. Rep.* **2020**, *10*, 16308. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Friel, J.; Bombarely, A.; Fornell, C.D.; Luque, F.; Fernández-Ocaña, A.M. Comparative Analysis of Genotyping by Sequencing and Whole-Genome Sequencing Methods in Diversity Studies of *Olea europaea* L. *Plants* **2021**, *10*, 2514. [\[CrossRef\]](#)
59. Faria, J.C.T.; Konzen, E.R.; Caldeira, M.V.W.; de Oliveira Godinho, T.; Maluf, L.P.; Moreira, S.O.; da Silva Carvalho, C.; Leal, B.S.S.; dos Santos Azevedo, C.; Momolli, D.R.; et al. Genetic Resources of African Mahogany in Brazil: Genomic Diversity and Structure of Forest Plantations. *BMC Plant Biol.* **2024**, *24*, 858. [\[CrossRef\]](#) [\[PubMed\]](#)
60. Chauhan, R.; Prabhakaran, S.; Tiwari, A.; Joshi, D.; Chandora, R.; Taj, G.; Jahan, T.; Singh, S.P.; Jaiswal, J.P.; Joshi, R.; et al. Genetic Distances and Genome Wide Population Structure Analysis of a Grain Amaranth (*Amaranthus hypochondriacus*) Diversity Panel Using Genotyping by Sequencing. *Sci. Rep.* **2025**, *15*, 33816. [\[CrossRef\]](#)
61. Li, W.; Liu, L.; Wang, Y.; Zhang, Q.; Fan, G.; Zhang, S.; Wang, Y.; Liao, K. Genetic Diversity, Population Structure, and Relationships of Apricot (*Prunus*) Based on Restriction Site-Associated DNA Sequencing. *Hortic. Res.* **2020**, *7*, 69. [\[CrossRef\]](#) [\[PubMed\]](#)
62. Berdugo-Cely, J.A.; Cortés, A.J.; López-Hernández, F.; Delgadillo-Durán, P.; Cerón-Souza, I.; Reyes-Herrera, P.H.; Navas-Arboleda, A.A.; Yockteng, R. Pleistocene-Dated Genomic Divergence of Avocado Trees Supports Cryptic Diversity in the Colombian Germplasm. *Tree Genet. Genomes* **2023**, *19*, 42. [\[CrossRef\]](#)
63. Akech, V.; Bengtsson, T.; Ortiz, R.; Swennen, R.; Uwimana, B.; Ferreira, C.F.; Amah, D.; Amorim, E.P.; Blisset, E.; Van den Houwe, I.; et al. Genetic Diversity and Population Structure in Banana (*Musa* Spp.) Breeding Germplasm. *Plant Genome* **2024**, *17*, e20497. [\[CrossRef\]](#)
64. Milner, S.G.; Jost, M.; Taketa, S.; Mazón, E.R.; Himmelbach, A.; Oppermann, M.; Weise, S.; Knüpfper, H.; Basterrechea, M.; König, P.; et al. Genebank Genomics Highlights the Diversity of a Global Barley Collection. *Nat. Genet.* **2019**, *51*, 319–326. [\[CrossRef\]](#)
65. Manzanero, B.R.; Kulkarni, K.P.; Vorsa, N.; Reddy, U.K.; Natarajan, P.; Elavarthi, S.; Iorizzo, M.; Melmaiee, K. Genomic and Evolutionary Relationships among Wild and Cultivated Blueberry Species. *BMC Plant Biol.* **2023**, *23*, 126. [\[CrossRef\]](#)
66. Horvath, D.P.; Stamm, M.; Talukder, Z.I.; Fiedler, J.; Horvath, A.P.; Horvath, G.A.; Chao, W.S.; Anderson, J.V. A New Diversity Panel for Winter Rapeseed (*Brassica napus* L.) Genome-Wide Association Studies. *Agronomy* **2020**, *10*, 2006. [\[CrossRef\]](#)
67. Vega-Muñoz, M.A.; López-Hernández, F.; Cortés, A.J.; Roda, F.; Castaño, E.; Montoya, G.; Henao-Rojas, J.C. Pangenomic and Phenotypic Characterization of Colombian Capsicum Germplasm Reveals the Genetic Basis of Fruit Quality Traits. *Int. J. Mol. Sci.* **2025**, *26*, 8205. [\[CrossRef\]](#) [\[PubMed\]](#)
68. López-Hernández, F.; Cortés, A.J. Last-Generation Genome–Environment Associations Reveal the Genetic Basis of Heat Tolerance in Common Bean (*Phaseolus vulgaris* L.). *Front. Genet.* **2019**, *10*, 954. [\[CrossRef\]](#) [\[PubMed\]](#)
69. Muli, J.K.; Neondo, J.O.; Kamau, P.K.; Michuki, G.N.; Odari, E.; Budambula, N.L.M. Genetic Diversity and Population Structure of Wild and Cultivated *Crotalaria* Species Based on Genotyping-by-Sequencing. *PLoS ONE* **2022**, *17*, e0272955. [\[CrossRef\]](#)
70. Hadizadeh, H.; Bahri, B.A.; Qi, P.; Wilde, H.D.; Devos, K.M. Intra- and Interspecific Diversity Analyses in the Genus *Eremurus* in Iran Using Genotyping-by-Sequencing Reveal Geographic Population Structure. *Hortic. Res.* **2020**, *7*, 30. [\[CrossRef\]](#)
71. Shigita, G.; Dung, T.P.; Pervin, M.N.; Duong, T.T.; Imoh, O.N.; Monden, Y.; Nishida, H.; Tanaka, K.; Sugiyama, M.; Kawazu, Y.; et al. Elucidation of Genetic Variation and Population Structure of Melon Genetic Resources in the NARO Genebank, and Construction of the World Melon Core Collection. *Breed. Sci.* **2023**, *73*, 269–277. [\[CrossRef\]](#)

72. Bekele, W.A.; Avni, R.; Birkett, C.L.; Itaya, A.; Wight, C.P.; Bellavance, J.; Brodführer, S.; Canales, F.J.; Carlson, C.H.; Fiebig, A.; et al. Global Genomic Population Structure of Wild and Cultivated Oat Reveals Signatures of Chromosome Rearrangements. *Nat. Commun.* **2025**, *16*, 9486. [CrossRef] [PubMed]
73. Zolkafli, S.H.; Marjuni, M.; Abdullah, N.; Singh, R.; Ithnin, M. Exploring Diversity in African Oil Palm (*Elaeis guineensis* Jacq.) Germplasm Populations via Genotyping-by-Sequencing. *Genet. Resour. Crop Evol.* **2025**, *72*, 6111–6127. [CrossRef]
74. Rispail, N.; Wohor, O.Z.; Osuna-Caballero, S.; Barilli, E.; Rubiales, D. Genetic Diversity and Population Structure of a Wide *Pisum* Spp. Core Collection. *Int. J. Mol. Sci.* **2023**, *24*, 2470. [CrossRef]
75. Hsu, C.C.; Chen, S.Y.; Chiu, S.Y.; Lai, C.Y.; Lai, P.H.; Shehzad, T.; Wu, W.L.; Chen, W.H.; Paterson, A.H.; Chen, H.H. High-Density Genetic Map and Genome-Wide Association Studies of Aesthetic Traits in *Phalaenopsis* Orchids. *Sci. Rep.* **2022**, *12*, 3346. [CrossRef]
76. Tuttle, H.K.; Del Rio, A.H.; Bamberg, J.B.; Shannon, L.M. Potato Soup: Analysis of Cultivated Potato Gene Bank Populations Reveals High Diversity and Little Structure. *Front. Plant Sci.* **2024**, *15*, 1429279. [CrossRef]
77. Hawliczek, A.; Borzęcka, E.; Tofil, K.; Alachiotis, N.; Bolibok, L.; Gawroński, P.; Siekmann, D.; Hackauf, B.; Dušínský, R.; Švec, M.; et al. Selective Sweeps Identification in Distinct Groups of Cultivated Rye (*Secale cereale* L.) Germplasm Provides Potential Candidate Genes for Crop Improvement. *BMC Plant Biol.* **2023**, *23*, 323. [CrossRef] [PubMed]
78. Seay, D.; Szczepanek, A.; De La Fuente, G.N.; Votava, E.; Abdel-Haleem, H. Genetic Diversity and Population Structure of a Large USDA Sesame Collection. *Plants* **2024**, *13*, 1765. [CrossRef] [PubMed]
79. Filippi, C.V.; Merino, G.A.; Montecchia, J.F.; Aguirre, N.C.; Rivarola, M.; Naamati, G.; Fass, M.I.; Álvarez, D.; Di Rienzo, J.; Heinz, R.A.; et al. Genetic Diversity, Population Structure and Linkage Disequilibrium Assessment among International Sunflower Breeding Collections. *Genes* **2020**, *11*, 283. [CrossRef]
80. Mizuno, N.; Toyoshima, M.; Fujita, M.; Fukuda, S.; Kobayashi, Y.; Ueno, M.; Tanaka, K.; Tanaka, T.; Nishihara, E.; Mizukoshi, H.; et al. The Genotype-Dependent Phenotypic Landscape of Quinoa in Salt Tolerance and Key Growth Traits. *DNA Res.* **2020**, *27*, dsaa022. [CrossRef] [PubMed]
81. Sansaloni, C.; Franco, J.; Santos, B.; Percival-Alwyn, L.; Singh, S.; Petroli, C.; Campos, J.; Dreher, K.; Payne, T.; Marshall, D.; et al. Diversity Analysis of 80,000 Wheat Accessions Reveals Consequences and Opportunities of Selection Footprints. *Nat. Commun.* **2020**, *11*, 4572. [CrossRef] [PubMed]
82. Agre, P.; Asibe, F.; Darkwa, K.; Edemodu, A.; Bauchet, G.; Asiedu, R.; Adebola, P.; Asfaw, A. Phenotypic and Molecular Assessment of Genetic Structure and Diversity in a Panel of Winged Yam (*Dioscorea alata*) Clones and Cultivars. *Sci. Rep.* **2019**, *9*, 18221. [CrossRef]
83. Pavan, S.; Delvento, C.; Ricciardi, L.; Lotti, C.; Ciani, E.; D'Agostino, N. Recommendations for Choosing the Genotyping Method and Best Practices for Quality Control in Crop Genome-Wide Association Studies. *Front. Genet.* **2020**, *11*, 447. [CrossRef] [PubMed]
84. Wang, Q.; Lu, Y.; Li, M.; Gao, Z.; Li, D.; Gao, Y.; Deng, W.; Wu, J. Leveraging Whole-Genome Resequencing to Uncover Genetic Diversity and Promote Conservation Strategies for Ruminants in Asia. *Animals* **2025**, *15*, 831. [CrossRef] [PubMed]
85. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010. Available online: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 7 November 2025).
86. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]
87. Li, H.; Durbin, R. Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef] [PubMed]
88. Langmead, B.; Salzberg, S.L. Fast Gapped-Read Alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef]
89. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef]
90. Wu, X.; Heffelfinger, C.; Zhao, H.; Dellaporta, S.L. Benchmarking Variant Identification Tools for Plant Diversity Discovery. *BMC Genom.* **2019**, *20*, 701. [CrossRef]
91. Happ, M.M.; Wang, H.; Graef, G.L.; Hyten, D.L. Generating High Density, Low Cost Genotype Data in Soybean [*Glycine max* (L.) Merr.]. *G3 Genes Genomes Genet.* **2019**, *9*, 2153–2160. [CrossRef]
92. García-Romeral, J.; Castanera, R.; Casacuberta, J.; Domingo, C. Deciphering the Genetic Basis of Allelopathy in Japonica Rice Cultivated in Temperate Regions Using a Genome-Wide Association Study. *Rice* **2024**, *17*, 22. [CrossRef]
93. Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M.; Cingolani, P.; et al. A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff. *Fly* **2012**, *6*, 80–92. [CrossRef]
94. Poplin, R.; Ruano-Rubio, V.; DePristo, M.A.; Fennell, T.J.; Carneiro, M.O.; Van der Auwera, G.A.; Kling, D.E.; Gauthier, L.D.; Levy-Moonshine, A.; Roazen, D.; et al. Scaling Accurate Genetic Variant Discovery to Tens of Thousands of Samples. *bioRxiv* **2017**, *1*, 201178. [CrossRef]

95. Yu, L.X.; Zheng, P.; Bhamidimarri, S.; Liu, X.P.; Main, D. The Impact of Genotyping-by-Sequencing Pipelines on SNP Discovery and Identification of Markers Associated with Verticillium Wilt Resistance in Autotetraploid Alfalfa (*Medicago sativa* L.). *Front. Plant Sci.* **2017**, *8*, 89. [CrossRef] [PubMed]
96. Pook, T.; Mayer, M.; Geibel, J.; Weigend, S.; Caverio, D.; Schoen, C.C.; Simianer, H. Improving Imputation Quality in Beagle for Crop and Livestock Data. *G3 Genes Genomes Genet.* **2020**, *10*, 177–188. [CrossRef]
97. Howie, B.N.; Donnelly, P.; Marchini, J. A Flexible and Accurate Genotype Imputation Method for the next Generation of Genome-Wide Association Studies. *PLoS Genet.* **2009**, *5*, e1000529. [CrossRef]
98. Long, E.M.; Bradbury, P.J.; Cinta Romay, M.; Buckler, E.S.; Robbins, K.R. Genome-Wide Imputation Using the Practical Haplotype Graph in the Heterozygous Crop Cassava. *G3 Genes Genomes Genet.* **2022**, *12*, jkab383. [CrossRef] [PubMed]
99. Gonen, S.; Wimmer, V.; Gaynor, R.C.; Byrne, E.; Gorjanc, G.; Hickey, J.M. A Heuristic Method for Fast and Accurate Phasing and Imputation of Single-Nucleotide Polymorphism Data in Bi-Parental Plant Populations. *Theor. Appl. Genet.* **2018**, *131*, 2345–2357. [CrossRef] [PubMed]
100. Money, D.; Gardner, K.; Migicovsky, Z.; Schwaninger, H.; Zhong, G.Y.; Myles, S. LinkImpute: Fast and Accurate Genotype Imputation for Nonmodel Organisms. *G3 Genes Genomes Genet.* **2015**, *5*, 2383–2390. [CrossRef]
101. Gonçalves-Dias, J.; Singh, A.; Graf, C.; Stetter, M.G. Genetic Incompatibilities and Evolutionary Rescue by Wild Relatives Shaped Grain Amaranth Domestication. *Mol. Biol. Evol.* **2023**, *40*, msad177. [CrossRef]
102. Cañas-Gutiérrez, G.P.; López-Hernández, F.; Cortés, A.J. Whole Genome Resequencing of 205 Avocado Trees Unveils the Genomic Patterns of Racial Divergence in the Americas. *Int. J. Mol. Sci.* **2025**, *26*, 10353. [CrossRef]
103. Coe, K.; Bostan, H.; Rolling, W.; Turner-Hissong, S.; Macko-Podgórní, A.; Senalik, D.; Liu, S.; Seth, R.; Curaba, J.; Mengist, M.F.; et al. Population Genomics Identifies Genetic Signatures of Carrot Domestication and Improvement and Uncovers the Origin of High-Carotenoid Orange Carrots. *Nat. Plants* **2023**, *9*, 1643–1658. [CrossRef]
104. Mekbib, Y.; Tesfaye, K.; Dong, X.; Saina, J.K.; Hu, G.W.; Wang, Q.F. Whole-Genome Resequencing of *Coffea arabica* L. (Rubiaceae) Genotypes Identify SNP and Unravels Distinct Groups Showing a Strong Geographical Pattern. *BMC Plant Biol.* **2022**, *22*, 69. [CrossRef] [PubMed]
105. Denning-James, K.E.; Chater, C.; Cortés, A.J.; Blair, M.W.; Peláez, D.; Hall, A.; De Vega, J.J. Genome-Wide Association Mapping Dissects the Selective Breeding of Determinacy and Photoperiod Sensitivity in Common Bean (*Phaseolus vulgaris* L.). *G3 Genes Genomes Genet.* **2025**, *15*, jkaf090, Erratum in *G3 Genes Genomes Genet.* **2025**, *15*, jkaf141. <https://doi.org/10.1093/g3journal/jkaf141>. [CrossRef] [PubMed]
106. Yu, J.; Hui, Y.; Chen, J.; Yu, H.; Gao, X.; Zhang, Z.; Li, Q.; Zhu, S.; Zhao, T. Whole-Genome Resequencing of 240 Gossypium Barbardense Accessions Reveals Genetic Variation and Genes Associated with Fiber Strength and Lint Percentage. *Theor. Appl. Genet.* **2021**, *134*, 3249–3261. [CrossRef] [PubMed]
107. Zhong, Y.; Feng, L.; Deng, H.; Ji, X.; Zhang, J.; Sun, Y.; Lin, P.; Qiao, Y.; Xie, S.; Wang, H.; et al. Genomic Resequencing Reveals Genetic Diversity, Population Structure, and Core Collection of Durian Germplasm. *Commun. Biol.* **2025**, *8*, 1273. [CrossRef]
108. Ahmed, H.I.; Heuberger, M.; Schoen, A.; Koo, D.H.; Quiroz-Chavez, J.; Adhikari, L.; Raupp, J.; Cauet, S.; Rodde, N.; Cravero, C.; et al. Einkorn Genomics Sheds Light on History of the Oldest Domesticated Wheat. *Nature* **2023**, *620*, 830–838. [CrossRef]
109. Zhao, Y.P.; Fan, G.; Yin, P.P.; Sun, S.; Li, N.; Hong, X.; Hu, G.; Zhang, H.; Zhang, F.M.; Han, J.D.; et al. Resequencing 545 Ginkgo Genomes across the World Reveals the Evolutionary History of the Living Fossil. *Nat. Commun.* **2019**, *10*, 4201. [CrossRef]
110. Liang, Z.; Duan, S.; Sheng, J.; Zhu, S.; Ni, X.; Shao, J.; Liu, C.; Nick, P.; Du, F.; Fan, P.; et al. Whole-Genome Resequencing of 472 Vitis Accessions for Grapevine Diversity and Demographic History Analyses. *Nat. Commun.* **2019**, *10*, 1190. [CrossRef]
111. Ren, G.; Zhang, X.; Li, Y.; Ridout, K.; Serrano-Serrano, M.L.; Yang, Y.; Liu, A.; Ravikanth, G.; Nawaz, M.A.; Mumtaz, A.S.; et al. Large-Scale Whole-Genome Resequencing Unravels the Domestication History of Cannabis Sativa. *Sci. Adv.* **2021**, *7*, eabg2286. [CrossRef]
112. Wei, T.; van Treuren, R.; Liu, X.; Zhang, Z.; Chen, J.; Liu, Y.; Dong, S.; Sun, P.; Yang, T.; Lan, T.; et al. Whole-Genome Resequencing of 445 Lactuca Accessions Reveals the Domestication History of Cultivated Lettuce. *Nat. Genet.* **2021**, *53*, 752–760. [CrossRef]
113. Teshome, A.; Lire, H.; Higgins, J.; Olango, T.M.; Habte, E.H.; Negawo, A.T.; Muktar, M.S.; Assefa, Y.; Pereira, J.F.; Azevedo, A.L.S.; et al. Whole-Genome Resequencing of a Global Collection of Napier Grass (*Cenchrus purpureus*) to Explore Global Population Structure and QTL Governing Yield and Feed Quality Traits. *G3 Genes Genomes Genet.* **2025**, *15*, jkaf113. [CrossRef]
114. Liu, F.; Zhao, J.; Sun, H.; Xiong, C.; Sun, X.; Wang, X.; Wang, Z.; Jarret, R.; Wang, J.; Tang, B.; et al. Genomes of Cultivated and Wild Capsicum Species Provide Insights into Pepper Domestication and Population Differentiation. *Nat. Commun.* **2023**, *14*, 5487. [CrossRef] [PubMed]
115. Xiang, X.; Zhou, X.; Zi, H.; Wei, H.; Cao, D.; Zhang, Y.; Zhang, L.; Hu, J. Populus Cathayana Genome and Population Resequencing Provide Insights into Its Evolution and Adaptation. *Hortic. Res.* **2024**, *11*, uhad255. [CrossRef] [PubMed]
116. Sun, Y.; Shen, E.; Hu, Y.; Wu, D.; Feng, Y.; Lao, S.; Dong, C.; Du, T.; Hua, W.; Ye, C.Y.; et al. Population Genomic Analysis Reveals Domestication of Cultivated Rye from Weedy Rye. *Mol. Plant* **2022**, *15*, 552–561. [CrossRef] [PubMed]

117. Zatybekov, A.; Genievskaya, Y.; Fang, C.; Abugalieva, S.; Turuspekov, Y. Uncovering the Genetic Landscape of Soybean Accessions from Kazakhstan in Comparison with Global Germplasm Using Whole Genome Resequencing. *BMC Genom.* **2025**, *26*, 802. [\[CrossRef\]](#)
118. Song, Y.; Wang, Y.; Liu, Y.; Li, H.; Ding, J.; Wu, X.; Li, Y.; Jiao, F.; Yang, L. Whole Genome Re-Sequencing in 437 Tobacco Germplasms Identifies Plant Height Candidate Genes. *Sci. Rep.* **2025**, *15*, 4734. [\[CrossRef\]](#)
119. Razifard, H.; Ramos, A.; Della Valle, A.L.; Bodary, C.; Goetz, E.; Manser, E.J.; Li, X.; Zhang, L.; Visa, S.; Tieman, D.; et al. Genomic Evidence for Complex Domestication History of the Cultivated Tomato in Latin America. *Mol. Biol. Evol.* **2020**, *37*, 1118–1132. [\[CrossRef\]](#)
120. Saxena, R.K.; Rathore, A.; Bohra, A.; Yadav, P.; Das, R.R.; Khan, A.W.; Singh, V.K.; Chitikineni, A.; Singh, I.P.; Kumar, C.V.S.; et al. Development and Application of High-Density Axiom Cajanus SNP Array with 56K SNPs to Understand the Genome Architecture of Released Cultivars and Founder Genotypes. *Plant Genome* **2018**, *11*, 180005. [\[CrossRef\]](#)
121. Hiraoka, Y.; Ferrante, S.P.; Wu, G.A.; Federici, C.T.; Roose, M.L. Development and Assessment of SNP Genotyping Arrays for Citrus and Its Close Relatives. *Plants* **2024**, *13*, 691. [\[CrossRef\]](#)
122. Wang, S.; Wong, D.; Forrest, K.; Allen, A.; Chao, S.; Huang, B.E.; Maccaferri, M.; Salvi, S.; Milner, S.G.; Cattivelli, L.; et al. Characterization of Polyploid Wheat Genomic Diversity Using a High-Density 90 000 Single Nucleotide Polymorphism Array. *Plant Biotechnol. J.* **2014**, *12*, 787–796. [\[CrossRef\]](#)
123. LaFramboise, T. Single Nucleotide Polymorphism Arrays: A Decade of Biological, Computational and Technological Advances. *Nucleic Acids Res.* **2009**, *37*, 4181–4193. [\[CrossRef\]](#)
124. Burridge, A.J.; Winfield, M.; Przewieslik-Allen, A.; Edwards, K.J.; Siddique, I.; Barral-Arca, R.; Griffiths, S.; Cheng, S.; Huang, Z.; Feng, C.; et al. Development of a next Generation SNP Genotyping Array for Wheat. *Plant Biotechnol. J.* **2024**, *22*, 2235–2247. [\[CrossRef\]](#) [\[PubMed\]](#)
125. Taniguti, C.H.; Lau, J.; Hochhaus, T.; Arias, D.C.L.; Hokanson, S.C.; Zlesak, D.C.; Byrne, D.H.; Klein, P.E.; Riera-Lizarazu, O. Exploring Chromosomal Variations in Garden Roses: Insights from High-Density SNP Array Data and a New Tool, Qploidy. *Plant Genome* **2025**, *18*, e70044. [\[CrossRef\]](#) [\[PubMed\]](#)
126. Ganai, M.W.; Altmann, T.; Röder, M.S. SNP Identification in Crop Plants. *Curr. Opin. Plant Biol.* **2009**, *12*, 211–217. [\[CrossRef\]](#)
127. Keeble-Gagnère, G.; Pasam, R.; Forrest, K.L.; Wong, D.; Robinson, H.; Godoy, J.; Rattey, A.; Moody, D.; Mullan, D.; Walmsley, T.; et al. Novel Design of Imputation-Enabled SNP Arrays for Breeding and Research Applications Supporting Multi-Species Hybridization. *Front. Plant Sci.* **2021**, *12*, 756877. [\[CrossRef\]](#)
128. Gao, Y.; Yang, Z.; Yang, W.; Yang, Y.; Gong, J.; Yang, Q.Y.; Niu, X. Plant-ImputeDB: An Integrated Multiple Plant Reference Panel Database for Genotype Imputation. *Nucleic Acids Res.* **2021**, *49*, D1480–D1488. [\[CrossRef\]](#)
129. 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. [\[CrossRef\]](#)
130. Xu, G.; Zhang, X.; Chen, W.; Zhang, R.; Li, Z.; Wen, W.; Warburton, M.L.; Li, J.; Li, H.; Yang, X. Population Genomics of Zea Species Identifies Selection Signatures during Maize Domestication and Adaptation. *BMC Plant Biol.* **2022**, *22*, 72. [\[CrossRef\]](#)
131. Sun, C.; Dong, Z.; Zhao, L.; Ren, Y.; Zhang, N.; Chen, F. The Wheat 660K SNP Array Demonstrates Great Potential for Marker-Assisted Selection in Polyploid Wheat. *Plant Biotechnol. J.* **2020**, *18*, 1354–1360. [\[CrossRef\]](#)
132. Zhang, C.; Li, M.; Liang, L.; Xiang, J.; Zhang, F.; Zhang, C.; Li, Y.; Liang, J.; Zheng, T.; Zhang, F.; et al. Rice3K56 Is a High-Quality SNP Array for Genome-Based Genetic Studies and Breeding in Rice (*Oryza sativa* L.). *Crop. J.* **2023**, *11*, 800–807. [\[CrossRef\]](#)
133. Song, Q.; Hyten, D.L.; Jia, G.; Quigley, C.V.; Fickus, E.W.; Nelson, R.L.; Cregan, P.B. Development and Evaluation of SoySNP50K, a High-Density Genotyping Array for Soybean. *PLoS ONE* **2013**, *8*, e54985. [\[CrossRef\]](#) [\[PubMed\]](#)
134. Albrechtsen, A.; Nielsen, F.C.; Nielsen, R. Ascertainment Biases in SNP Chips Affect Measures of Population Divergence Research Article. *Mol. Biol. Evol.* **2010**, *27*, 2534–2547. [\[CrossRef\]](#)
135. Geibel, J.; Reimer, C.; Weigend, S.; Weigend, A.; Pook, T.; Simianer, H. How Array Design Creates SNP Ascertainment Bias. *PLoS ONE* **2021**, *16*, e0245178. [\[CrossRef\]](#)
136. Dokan, K.; Kawamura, S.; Teshima, K.M. Effects of Single Nucleotide Polymorphism Ascertainment on Population Structure Inferences. *G3 Genes Genomes Genet.* **2021**, *11*, jkab128. [\[CrossRef\]](#)
137. Li, J.; Luo, Y.; Zhang, R.; Li, X.; Pan, H.; Yin, H. Decoding Hybrid Origins and Genetic Architecture of Leaf Traits Variation in Camellia via High-Density 21 K SNP Array for Genomic Prediction. *Hortic. Res.* **2025**, *12*, uhaf221. [\[CrossRef\]](#)
138. Fiscus, C.J.; Herniter, I.A.; Tchamba, M.; Paliwal, R.; Muñoz-Amatriaín, M.; Roberts, P.A.; Abberton, M.; Alaba, O.; Close, T.J.; Oyatomi, O.; et al. The Pattern of Genetic Variability in a Core Collection of 2,021 Cowpea Accessions. *G3 Genes Genomes Genet.* **2024**, *14*, jkae071. [\[CrossRef\]](#)
139. Wang, L.; Xu, J.; Wang, H.; Chen, T.; You, E.; Bian, H.; Chen, W.; Zhang, B.; Shen, Y. Population Structure Analysis and Genome-Wide Association Study of a Hexaploid Oat Landrace and Cultivar Collection. *Front. Plant Sci.* **2023**, *14*, 1131751. [\[CrossRef\]](#) [\[PubMed\]](#)

140. Koorevaar, T.; Willemsen, J.H.; Visser, R.G.F.; Arens, P.; Maliepaard, C. Construction of a Strawberry Breeding Core Collection to Capture and Exploit Genetic Variation. *BMC Genom.* **2023**, *24*, 740. [\[CrossRef\]](#) [\[PubMed\]](#)
141. Hamilton, J.P.; Li, C.; Buell, C.R. The Rice Genome Annotation Project: An Updated Database for Mining the Rice Genome. *Nucleic Acids Res.* **2025**, *53*, D1614–D1622. [\[CrossRef\]](#) [\[PubMed\]](#)
142. 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. [\[CrossRef\]](#)
143. Haseneyer, G.; Schmutzer, T.; Seidel, M.; Zhou, R.; Mascher, M.; Schön, C.-C.; Taudien, S.; Scholz, U.; Stein, N.; Mayer, K.F.; et al. From RNA-Seq to Large-Scale Genotyping—Genomics Resources for Rye (*Secale cereale* L.). *BMC Plant Biol.* **2011**, *11*, 131. [\[CrossRef\]](#)
144. Wallace, J.G.; Upadhyaya, H.D.; Vetriventhan, M.; Buckler, E.S.; Tom Hash, C.; Ramu, P. The Genetic Makeup of a Global Barnyard Millet Germplasm Collection. *Plant Genome* **2015**, *8*, plantgenome2014.10.0067. [\[CrossRef\]](#)
145. Maharajan, T.; Krishna, T.P.A.; Krishnakumar, N.M.; Vetriventhan, M.; Kudapa, H.; Ceasar, S.A. Role of Genome Sequences of Major and Minor Millets in Strengthening Food and Nutritional Security for Future Generations. *Agriculture* **2024**, *14*, 670. [\[CrossRef\]](#)
146. Renganathan, V.G.; Vanniarajan, C.; Karthikeyan, A.; Ramalingam, J. Barnyard Millet for Food and Nutritional Security: Current Status and Future Research Direction. *Front. Genet.* **2020**, *11*, 500. [\[CrossRef\]](#)
147. Clevenger, J.; Chavarro, C.; Pearl, S.A.; Ozias-Akins, P.; Jackson, S.A. Single Nucleotide Polymorphism Identification in Polyploids: A Review, Example, and Recommendations. *Mol. Plant* **2015**, *8*, 831–846. [\[CrossRef\]](#)
148. Limborg, M.T.; Seeb, L.W.; Seeb, J.E. Sorting Duplicated Loci Disentangles Complexities of Polyploid Genomes Masked by Genotyping by Sequencing. *Mol. Ecol.* **2016**, *25*, 2117–2129. [\[CrossRef\]](#)
149. Bourke, P.M.; Voorrips, R.E.; Visser, R.G.F.; Maliepaard, C. Tools for Genetic Studies in Experimental Populations of Polyploids. *Front. Plant Sci.* **2018**, *9*, 513. [\[CrossRef\]](#) [\[PubMed\]](#)
150. Pucker, B.; Irisarri, I.; De Vries, J.; Xu, B. Plant Genome Sequence Assembly in the Era of Long Reads: Progress, Challenges and Future Directions. *Quant. Plant Biol.* **2022**, *3*, e5. [\[CrossRef\]](#) [\[PubMed\]](#)
151. Gladman, N.; Goodwin, S.; Chougule, K.; Richard McCombie, W.; Ware, D. Era of Gapless Plant Genomes: Innovations in Sequencing and Mapping Technologies Revolutionize Genomics and Breeding. *Curr. Opin. Biotechnol.* **2023**, *79*, 102886. [\[CrossRef\]](#) [\[PubMed\]](#)
152. Della Coletta, R.; Qiu, Y.; Ou, S.; Hufford, M.B.; Hirsch, C.N. How the Pan-Genome Is Changing Crop Genomics and Improvement. *Genome Biol.* **2021**, *22*, 3. [\[CrossRef\]](#) [\[PubMed\]](#)
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. Shi, J.; Tian, Z.; Lai, J.; Huang, X. Plant Pan-Genomics and Its Applications. *Mol. Plant* **2023**, *16*, 168–186. [\[CrossRef\]](#) [\[PubMed\]](#)
155. Lei, L.; Goltsman, E.; Goodstein, D.; Wu, G.A.; Rokhsar, D.S.; Vogel, J.P. Plant Pan-Genomics Comes of Age. *Annu. Rev. Plant Biol.* **2021**, *72*, 411–435. [\[CrossRef\]](#)
156. Chawla, H.S.; Lee, H.T.; Gabur, I.; Vollrath, P.; Tamilselvan-Nattar-Amutha, S.; Obermeier, C.; Schiessl, S.V.; Song, J.M.; Liu, K.; Guo, L.; et al. Long-Read Sequencing Reveals Widespread Intragenic Structural Variants in a Recent Allopolyploid Crop Plant. *Plant Biotechnol. J.* **2021**, *19*, 240–250. [\[CrossRef\]](#)
157. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3*, 160018. [\[CrossRef\]](#) [\[PubMed\]](#)
158. Allan, V.; Vetriventhan, M.; Senthil, R.; Geetha, S.; Deshpande, S.; Rathore, A.; Kumar, V.; Singh, P.; Reddymalla, S.; Azevedo, V.C.R. Genome-Wide DArTSeq Genotyping and Phenotypic Based Assessment of Within and Among Accessions Diversity and Effective Sample Size in the Diverse Sorghum, Pearl Millet, and Pigeonpea Landraces. *Front. Plant Sci.* **2020**, *11*, 587426. [\[CrossRef\]](#)
159. Campuzano-Duque, L.F.; Bejarano-Garavito, D.; Castillo-Sierra, J.; Torres-Cuesta, D.R.; Cortés, A.J.; Blair, M.W. SNP Genotyping for Purity Assessment of a Forage Oat (*Avena sativa* L.) Variety from Colombia. *Agronomy* **2022**, *12*, 1710. [\[CrossRef\]](#)
160. Berkner, M.O.; Jiang, Y.; Reif, J.C.; Schulthess, A.W. Trait-Customized Sampling of Core Collections from a Winter Wheat Genebank Collection Supports Association Studies. *Front. Plant Sci.* **2024**, *15*, 1451749. [\[CrossRef\]](#)
161. Pavlidis, P.; Alachiotis, N. A Survey of Methods and Tools to Detect Recent and Strong Positive Selection. *J. Biol. Res.* **2017**, *24*, 7. [\[CrossRef\]](#)
162. Li, C.; Zhou, A.; Sang, T. Rice Domestication by Reducing Shattering. *Science* **2006**, *311*, 1936–1939. [\[CrossRef\]](#)
163. Frary, A.; Nesbitt, T.C.; Frary, A.; Grandillo, S.; Van Der Knaap, E.; Cong, B.; Liu, J.; Meller, J.; Elber, R.; Alpert, K.B.; et al. Fw2.2: A Quantitative Trait Locus Key to the Evolution of Tomato Fruit Size. *Science* **2000**, *289*, 85–88. [\[CrossRef\]](#) [\[PubMed\]](#)

164. Maccaferri, M.; Harris, N.S.; Twardziok, S.O.; Pasam, R.K.; Gundlach, H.; Spannagl, M.; Ormanbekova, D.; Lux, T.; Prade, V.M.; Milner, S.G.; et al. Durum Wheat Genome Highlights Past Domestication Signatures and Future Improvement Targets. *Nat. Genet.* **2019**, *51*, 885–895. [\[CrossRef\]](#)
165. Wegary, D.; Teklewold, A.; Prasanna, B.M.; Ertiro, B.T.; Alachiotis, N.; Negera, D.; Awas, G.; Abakemal, D.; Ogugo, V.; Gowda, M.; et al. Molecular Diversity and Selective Sweeps in Maize Inbred Lines Adapted to African Highlands. *Sci. Rep.* **2019**, *9*, 13490. [\[CrossRef\]](#)
166. Ndjioudjop, M.N.; Alachiotis, N.; Pavlidis, P.; Goungoulou, A.; Kpeki, S.B.; Zhao, D.; Semagn, K. Comparisons of Molecular Diversity Indices, Selective Sweeps and Population Structure of African Rice with Its Wild Progenitor and Asian Rice. *Theor. Appl. Genet.* **2019**, *132*, 1145–1158. [\[CrossRef\]](#)
167. Semagn, K.; Iqbal, M.; Alachiotis, N.; N'Diaye, A.; Pozniak, C.; Spaner, D. Genetic Diversity and Selective Sweeps in Historical and Modern Canadian Spring Wheat Cultivars Using the 90K SNP Array. *Sci. Rep.* **2021**, *11*, 23773. [\[CrossRef\]](#) [\[PubMed\]](#)
168. Narum, S.R.; Hess, J.E. Comparison of FST Outlier Tests for SNP Loci under Selection. *Mol. Ecol. Resour.* **2011**, *11*, 184–194. [\[CrossRef\]](#)
169. Lin, T.; Zhu, G.; Zhang, J.; Xu, X.; Yu, Q.; Zheng, Z.; Zhang, Z.; Lun, Y.; Li, S.; Wang, X.; et al. Genomic Analyses Provide Insights into the History of Tomato Breeding. *Nat. Genet.* **2014**, *46*, 1220–1226. [\[CrossRef\]](#)
170. Blanca, J.; Montero-Pau, J.; Sauvage, C.; Bauchet, G.; Illa, E.; Díez, M.J.; Francis, D.; Causse, M.; van der Knaap, E.; Cañizares, J. Genomic Variation in Tomato, from Wild Ancestors to Contemporary Breeding Accessions. *BMC Genom.* **2015**, *16*, 257. [\[CrossRef\]](#) [\[PubMed\]](#)
171. You, Q.; Yang, X.; Peng, Z.; Xu, L.; Wang, J. Development and Applications of a High Throughput Genotyping Tool for Polyploid Crops: Single Nucleotide Polymorphism (SNP) Array. *Front. Plant Sci.* **2018**, *9*, 104. [\[CrossRef\]](#)
172. Zanini, S.F.; Bayer, P.E.; Wells, R.; Snowdon, R.J.; Batley, J.; Varshney, R.K.; Nguyen, H.T.; Edwards, D.; Golicz, A.A. Pangenomics in Crop Improvement—From Coding Structural Variations to Finding Regulatory Variants with Pangenome Graphs. *Plant Genome* **2022**, *15*, e20177. [\[CrossRef\]](#)
173. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
174. Gage, J.L.; Vaillancourt, B.; Hamilton, J.P.; Manrique-Carpintero, N.C.; Gustafson, T.J.; Barry, K.; Lipzen, A.; Tracy, W.F.; Mikel, M.A.; Kaeppler, S.M.; et al. Multiple Maize Reference Genomes Impact the Identification of Variants by Genome-Wide Association Study in a Diverse Inbred Panel. *Plant Genome* **2019**, *12*, 180069. [\[CrossRef\]](#)
175. Gonçalves-Dias, J.; Stetter, M.G. PopAmaranth: A Population Genetic Genome Browser for Grain Amaranths and Their Wild Relatives. *G3 Genes Genomes Genet.* **2021**, *11*, jkab103. [\[CrossRef\]](#)
176. Singh, A.; Mahato, A.K.; Maurya, A.; Rajkumar, S.; Singh, A.K.; Bhardwaj, R.; Kaushik, S.K.; Kumar, S.; Gupta, V.; Singh, K.; et al. Amaranth Genomic Resource Database: An Integrated Database Resource of Amaranth Genes and Genomics. *Front. Plant Sci.* **2023**, *14*, 1203855. [\[CrossRef\]](#) [\[PubMed\]](#)
177. Ji, F.; Ma, Q.; Zhang, W.; Liu, J.; Feng, Y.; Zhao, P.; Song, X.; Chen, J.; Zhang, J.; Wei, X.; et al. A Genome Variation Map Provides Insights into the Genetics of Walnut Adaptation and Agronomic Traits. *Genome Biol.* **2021**, *22*, 300. [\[CrossRef\]](#)
178. Takuno, S.; Ralph, P.; Swarts, K.; Elshire, R.J.; Glaubitz, J.C.; Buckler, E.S.; Hufford, M.B.; Ross-Ibarra, J. Independent Molecular Basis of Convergent Highland Adaptation in Maize. *Genetics* **2015**, *200*, 1297–1312. [\[CrossRef\]](#) [\[PubMed\]](#)
179. Papoutsoglou, E.A.; Faria, D.; Arend, D.; Arnaud, E.; Athanasiadis, I.N.; Chaves, I.; Coppens, F.; Cornut, G.; Costa, B.V.; Ćwiek-Kupczyńska, H.; et al. Enabling Reusability of Plant Phenomic Datasets with MIAPPE 1.1. *New Phytol.* **2020**, *227*, 260–273. [\[CrossRef\]](#) [\[PubMed\]](#)
180. Sheikh, M.; Iqra, F.; Ambreen, H.; Pravin, K.A.; Ikra, M.; Chung, Y.S. Integrating Artificial Intelligence and High-Throughput Phenotyping for Crop Improvement. *J. Integr. Agric.* **2024**, *23*, 1787–1802. [\[CrossRef\]](#)
181. Song, P.; Wang, J.; Guo, X.; Yang, W.; Zhao, C. High-Throughput Phenotyping: Breaking through the Bottleneck in Future Crop Breeding. *Crop J.* **2021**, *9*, 633–645. [\[CrossRef\]](#)
182. Wang, Z.; Hao, J.; Shi, X.; Wang, Q.; Zhang, W.; Li, F.; Mur, L.A.J.; Han, Y.; Hou, S.; Han, J.; et al. Integrating Dynamic High-Throughput Phenotyping and Genetic Analysis to Monitor Growth Variation in Foxtail Millet. *Plant Methods* **2024**, *20*, 168. [\[CrossRef\]](#)
183. Singh, A.; Ganapathysubramanian, B.; Singh, A.K.; Sarkar, S. Machine Learning for High-Throughput Stress Phenotyping in Plants. *Trends Plant Sci.* **2016**, *21*, 110–124. [\[CrossRef\]](#)
184. Parmley, K.A.; Higgins, R.H.; Ganapathysubramanian, B.; Sarkar, S.; Singh, A.K. Machine Learning Approach for Prescriptive Plant Breeding. *Sci. Rep.* **2019**, *9*, 17132. [\[CrossRef\]](#)
185. Yu, X.; Liu, Z.; Sun, X. Single-Cell and Spatial Multi-Omics in the Plant Sciences: Technical Advances, Applications, and Perspectives. *Plant Commun.* **2023**, *4*, 100508. [\[CrossRef\]](#) [\[PubMed\]](#)

186. Ke, Y.; Pujol, V.; Staut, J.; Pollaris, L.; Seurinck, R.; Eekhout, T.; Grones, C.; Saura-Sanchez, M.; Van Bel, M.; Vuylsteke, M.; et al. A Single-Cell and Spatial Wheat Root Atlas with Cross-Species Annotations Delineates Conserved Tissue-Specific Marker Genes and Regulators. *Cell Rep.* **2025**, *44*, 115240. [[CrossRef](#)] [[PubMed](#)]
187. Weckwerth, W.; Ghatak, A.; Bellaire, A.; Chaturvedi, P.; Varshney, R.K. PANOMICS Meets Germplasm. *Plant Biotechnol. J.* **2020**, *18*, 1507–1525. [[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.