

Review

Application of Genomic Big Data in Plant Breeding: Past, Present, and Future

Kyung Do Kim ^{1,†} , Yuna Kang ^{2,†} and Changsoo Kim ^{2,3,*} 

¹ Department of Bioscience and Bioinformatics, Myongji University, Yongin 17058, Korea; kyungdokim@mju.ac.kr

² Department of Crop Science, Chungnam National University, Daejeon 34134, Korea; dkw13120@cnu.ac.kr

³ Department of Smart Agriculture Systems, Chungnam National University, Daejeon 34134, Korea

* Correspondence: changsookim@cnu.ac.kr

† These authors contributed equally to this work.

Received: 28 September 2020; Accepted: 26 October 2020; Published: 28 October 2020



Abstract: Plant breeding has a long history of developing new varieties that have ensured the food security of the human population. During this long journey together with humanity, plant breeders have successfully integrated the latest innovations in science and technologies to accelerate the increase in crop production and quality. For the past two decades, since the completion of human genome sequencing, genomic tools and sequencing technologies have advanced remarkably, and adopting these innovations has enabled us to cost down and/or speed up the plant breeding process. Currently, with the growing mass of genomic data and digitalized biological data, interdisciplinary approaches using new technologies could lead to a new paradigm of plant breeding. In this review, we summarize the overall history and advances of plant breeding, which have been aided by plant genomic research. We highlight the key advances in the field of plant genomics that have impacted plant breeding over the past decades and introduce the current status of innovative approaches such as genomic selection, which could overcome limitations of conventional breeding and enhance the rate of genetic gain.

Keywords: next-generation sequencing; single nucleotide polymorphism; marker-trait association; genomic selection; marker-assisted selection

1. Introduction

The history of plant breeding might have begun when humans changed from hunter gatherers to farmers, otherwise known as an agricultural society. The key concept of early plant breeding was “crop domestication”, which implies the adaptation of wild plant species, so that humans could sustainably cultivate food plants to feed an ever-increasing population. Researchers estimate crop domestication dates back 9000 to 11,000 years. Since then, humans have bred plants using selection schemes according to the decisions of farmers. In addition, humans have historically traded crop plants from other countries for hundreds of years, which has diversified the agricultural environment and facilitated plant breeding [1].

The concept of scientific breeding started in the 19th century, triggered by Gregor Mendel’s research with garden peas, for which his plant hybridization led to his laws of inheritance. At the time of publishing his research paper, “Experiments on Plant Hybridization”, his work failed to get the attention of other scientists because biology was always interpreted by quantitative matters not by qualitative approaches [2]. However, his work was re-discovered by William Bateson in 1900. After integration with the Boveri–Sutton chromosome theory, scientists finally conceded his experiment was “the core of classical genetics”. Consequently, plant breeding faced a new era by the advances of

modern genetics with a plethora of newly generated breeding populations in many crops. In other words, the full-scale application of transmission genetics to plant breeding had begun at that point (e.g., introgression of useful traits in other accessions). However, at the beginning, plant breeding was only performed by phenotype evaluations, meaning that a breeder's ability to select desirable individuals was very critical. Soon after this, the breeding environment dramatically changed ever since the Maxam–Gilbert sequencing method was published in the late 1970s. The method was the first to provide nucleotide sequence information that could directly be used for developing molecular markers, which are very objective and have a capability for early stage selection. After the application of this method in plant genetics, first generation (Sanger sequencing), second generation (next-generation sequencing technology mostly referring to Illumina chemistry), and third generation sequencing technologies (Pacific BioScience or Oxford Nanopore technologies) have emerged and had significant impacts on plant breeding in a variety of ways [3]. For example, researchers started to utilize nucleotide information by applying molecular marker systems such as single nucleotide polymorphisms (SNPs), which are theoretically infinite in terms of numbers in plant genomes, enabling expanded marker-assisted selection (MAS) or genomic selection (GS) [4]. Additionally, having the whole genome sequence information is very powerful in plant breeding because the locations of traits and their underlying genes or chromosomes could be helpful for the introgression of interesting traits to other accessions. In addition, transcriptome sequencing (RNAseq) has revealed gene expression changes under various abiotic or biotic stress conditions, leading to profiles of genes associated with many agricultural traits (Figure 1).

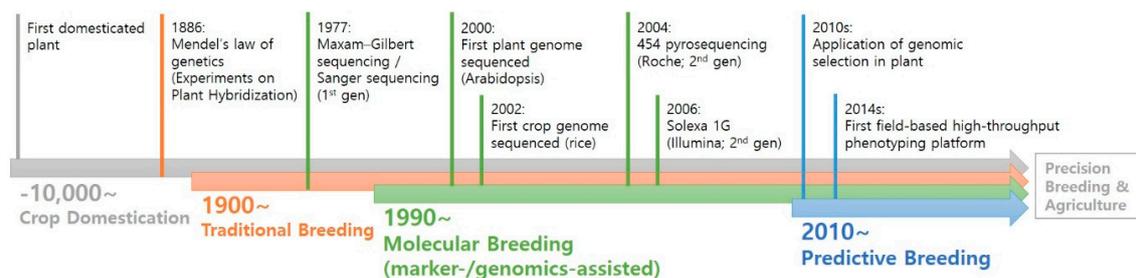


Figure 1. Key technological milestones in plant breeding.

In lectures on “Plant Genetics” at some universities, professors always define genomics as “the study of genetics with massive nucleotide sequence information”. In fact, plant breeding based on genetic tools is to identify genetic variants of interest that can be associated with phenotypic differences. In this respect, genomics provides breeders and geneticists with new opportunities in that modern molecular genetics have offered a relatively limited number of genetic variations whereas whole genome information can theoretically provide an unlimited number of genetic variations based on SNPs. Thus, the power of the whole genome reference becomes a critical tool for this purpose. As stated, sequencing technologies advance quickly, and the cost is decreasing. The massive amount of genetic variants thanks to technological advances has enabled breeders to find marker-trait associations (MTA), which exploit the genotyping of various populations with molecular markers covering the whole genome and analyze the associations between phenotypic variations and genotypic polymorphisms. Thus, researchers may be able to expand their application of conventional MAS techniques to the use of big data, which results in genomics-assisted selection (GAS). The expanded selection scheme has caused more precise dissection of genomes, which may improve breeding programs by reducing linkage drags, false positive markers, etc. [5]. In addition, we now face a new paradigm of plant breeding by predicting segregating models of specific traits based on a training population with the genotype and phenotype data. The established model (expressed as genomics-evaluated breeding values, GEBVs) is used to select target plants in the breeding populations after testing those in a validation population. The concept of this GS is originally based on old statistics and bioinformatics

that use training sets and validation sets, but it is also based on one of the top-notch techniques, machine learning (ML) [6].

In this review, the impact of genomics on modern plant breeding efforts is discussed, covering from sequencing technologies and whole genome references to their applications in breeding such as high-throughput genotyping and GS. We hope this review helps researchers by summarizing the overall history and advances of plant breeding together with the advances in plant genomics.

2. Application of Genomic Tools to Crop Improvement

2.1. History of Nucleotide Sequencing Technologies

Research on the structure of DNA has been conducted by scientists for a long time, but relatively recently, research has been conducted to obtain the genetic information contained in DNA. The development of nucleotide sequencing has greatly contributed to the ability to easily obtain a lot of genetic information. The history of nucleotide sequencing technology can be divided into three major generations [7].

The so-called first-generation sequencing refers to Maxam–Gilbert sequencing and Sanger sequencing. The Maxam–Gilbert method is the principle of decomposing bases using a chemical reaction using formic acid, dimethyl sulfate, or hydrazine on the 5'-end of a DNA fragment with ³²P-labeled, and then confirming it by polyacrylamide gel electrophoresis [8]. The principle of Sanger sequencing is a chain-termination technology developed by Frederick Sanger in 1977. Dideoxynucleotides (ddNTPs) that end the DNA chain during DNA replication are inserted into the DNA polymerase in a limited way. As ddNTPs lack the 3'-OH required to form a phosphate ester bond, the DNA polymerase stops extending the DNA. Using this principle, the method is to add each of the four dideoxynucleotides (ddATP, ddGTP, ddCTP, ddTTP), combined with fluorescent substances, to DNA and polymerize them with polymerase, and then check the base sequence using gel electrophoresis or chromatography [9]. Sanger sequencing was used to read genome sequences from plants such as rice and *Arabidopsis thaliana*, but it was still time-consuming and expensive [10].

Since the mid-2000s, the development of next generation sequencing (NGS) technology has enabled the production of large quantities of sequence data quickly, cheaply, and accurately. First, the 454 technology was developed in 2008. Its principle concept is to make emulsion oil using hydrophilic and hydrophobic properties and then to amplify by emulsion PCR by fixing one DNA strand to one bead. In the amplification process, the two phosphates that occur when a base is combined are turned into ATP, resulting in chemical luminescence as oxyluciferin is generated, which is the signal that identifies the base sequence. This method can read long sequences, but when a homopolymer is produced, the exact sequence cannot be determined, and the sequence is produced in a smaller amount than other second-generation sequencing techniques [11]. Second, it is most commonly used in second-generation technology as sequencing by synthesis (SBS) technology. SBS technology connects a short adapter oligo to DNA fragments and amplifies it in a flow cell containing a complementary oligo to form a cluster. When forming a cluster, dNTPs with fluorescent substances that emit different wave lengths of light depending on the type of base are synthesized one by one by DNA polymerase. At this time, the principle is to read and record the wave length of the fluorescent material, which is combined at each base, and repeat this process about 100 times to read the sequencing [12]. Using this method, it is possible to obtain a large amount of data quickly, accurately, and at a low cost, but it has the disadvantage of mainly obtaining a short sequence. Third, the method of amplification used in ion torrent uses emulsion PCR, such as in the 454 series, but the method of detection uses the semiconductor sequencing method. The method of electrically measuring hydrogen ions is based on the fact that hydrogen ions are generated each time a base is added. However, it is not electrically distinguishable because any additional base produces the same hydrogen ion. Therefore, hydrogen ions are measured after reacting with dATP [13]. Second-generation sequencing has the advantage of sequencing a single DNA molecule into a single base, so it is possible to sequence many DNA fragments simultaneously

in a small amount, but amplification is necessary for signal detection. However, third-generation sequencing can read very long DNA sequences by sequencing single DNA without amplification of DNA molecules.

The third-generation sequencing technology refers to PacBio's single molecule real time (SMRT) technology (www.pacb.com) and Oxford Nanopore technology (ONT) (www.nanoporetech.com). SMRT sequencing is a technology called zero-mode waveguide (ZMW), which performs sequencing by recognizing the fluorescence of complementary binding bases as DNA chains pass through the DNA polymerase fixed at the bottom of a small hole. This method has the advantage of being able to produce a long read of more than 20Kb [14]. The ONT sequencing performs sequencing by forming a nanometer-level channel in a membrane and then passing through a single strand of DNA, changing the potential difference between the membranes according to the base sequence. From the outside of the membrane, the double-stranded DNA is released by a helicase enzyme, and the single strand is configured to move into the channel at a constant rate. Unlike other methods, this method does not undergo any other complementary chemical reaction to the template DNA, and the DNA sequence can be read as it is. In addition, it has the advantage that sequencing is possible with a personal computer or a laptop using a cartridge that can be connected through a USB port, not a special device [15].

2.2. The Whole Genome References

The decreasing cost of nucleotide sequencing has changed the landscape of building the whole genome reference of many plant species. Nonetheless, the sequenced genomes have various different levels in terms of completeness. For example, some references have reached the level of pseudomolecules, which have been anchored to the chromosome scale, while many references have been published as a draft status. For plant breeders, any references with pseudomolecules will be really powerful because all the loci information can be acquired for the traits of interest, leading to the precise development of molecular markers for large-scale MAS and the easy detection of candidate genes. Draft genomes can also contribute to crop breeding thanks to their help in accurate genotyping using nucleotide sequence information.

As of October 2020, a total of 699 genomes in Viridiplantae have been sequenced as many different forms. The details (names of species, forms of assemblies) are well presented in the NCBI's genome database (www.ncbi.nlm.nih.gov/genome/browse#!/overview/viridiplantae) and the database is updated whenever a new genome assembly is added. The information is actively used for a variety of plant sciences. Additionally, other databases, such as Plant GDB (www.plantgdb.org) or Phytozome (www.phytozome.org), are providing plant genome information. Table 1 shows the formats and versions of genomics data for representative crop species deposited in those two databases. Once, assembling plant genomes was very challenging because of some specific features such as being relatively large and having highly repetitive genomes. In addition, plant genomes have abundant polyploidy or have experienced paleopolyploidization, resulting in complicated genomic structures due to paralogous or homologous sequences. Consequently, whole genome plant references in the early 21st century spent a vast amount of resources using first-generation sequencing technology (see previous section for details). However, the quality of the reference genomes was somewhat guaranteed in terms of their completeness. Some model species such as *Arabidopsis* [16], rice [17], maize [18], sorghum [19], populus [20], grapevine [21], papaya [22], or soybean [23] were sequenced at that time. After that, NGS technologies covering the second and third generation sequencing methods dramatically reduced the costs and resolved some assembling issues associated with plant genome structures. As a result, over 250 angiosperm species have been completely sequenced so far, and the number keeps rapidly increasing with or without constructing complete pseudomolecules [24]. Global statistics of sequenced genomes are well-documented in the Genome Database for Angiosperms (GDA, www.angiosperms.org) with appropriate web links. The point is that whole genome resources have facilitated the development of plant breeding programs as these resources enable researchers to provide precise genotyping schemes, which are directly usable for breeders in their field trials in the

form of genomics-assisted breeding (GAB) based on the expansion of the MAS or GAS concepts. We do not discuss the details of the whole genome references in this review because there are a plethora of reviews regarding this subject [24].

Table 1. A representative species of genome served in a database, such as Plant GDB and Phytozome.

Species Name	Version	Data Base Type	Provider	References
<i>Arabidopsis thaliana</i> (Thale Cress)	AtGDB	Chromosome	Plant GDB	[25]
	TAIR10	Chromosome	Phytozome	
	Araport11	Chromosome	Phytozome	[26]
<i>Hordeum vulgare</i> (Barley)	HvGDB	BAC	Plant GDB	[27]
	r1	BAC	Phytozome	[28,29]
<i>Oryza sativa</i> (Rice)	OsGDB	Chromosome	Plant GDB	[30]
	v3.1 (Kitaake rice)	Chromosome	Phytozome	[31]
	v7_JGI	Chromosome	Phytozome	[32]
<i>Sorghum bicolor</i> (Sorghum)	SbGDB	Chromosome	Plant GDB	[33]
	Rio v2.1	Scaffold	Phytozome	[34]
	v3.1.1	Chromosome	Phytozome	[33]
<i>Solanum tuberosum</i> (Potato)	StGDB	Chromosome	Plant GDB	[35]
	v4.03	Chromosome	Phytozome	[36]
<i>Triticum aestivum</i> (Wheat)	TaGDB	BAC	Plant GDB	[37]
	v2.2	Chromosome	Phytozome	[38]
<i>Zea mays</i> (maize)	ZmGDB	Chromosome/BAC	Plant GDB	[39]
	Ensembl-18	EST	Phytozome	[18]
	PH207 v1.1	transcripts	Phytozome	[31]

2.3. High-Throughput Genotyping and the Necessity of Phenotyping

Genotyping is a major process for plant breeding in that accurate plant breeding requires a number of plant individuals with a certain level of genetic variation. At the beginning of Mendelian genetics, genotyping mostly relied on their phenotypic variations, limiting target traits for plant breeding. The advent of the polymerase chain reaction (PCR) method has drastically advanced genotyping technology as a superior tool using PCR-based molecular markers such as random amplification length polymorphism (RAPD) and amplified fragment length polymorphism (AFLP). These marker types were popular choices for many studies because they do not require nucleotide sequence information and are cheap; however, the marker information is not reproducible in different populations. Ever since first-generation sequencing became commercially available, sequence-based PCR markers have become predominant in genotyping procedures. For example, simple sequence repeat (SSR) markers are relatively inexpensive, abundant in plant genomes and more informative than previous PCR-based markers [40]. The most powerful aspect of SSRs were synergized together with the development of expressed sequence tags (ESTs), which capture actively expressed genes. EST-SSRs were once scientists' favorite choice because they can link marker information and genes associated with target traits. Nonetheless, we cannot define SSR as a tool for high-throughput genotyping because gel-based genotyping SSRs are very laborious and time-consuming, and automated fragment analysis systems are relatively generally low-throughput with high cost even if they provide a certain level of multiplexing. The most recent molecular marker is single nucleotide polymorphisms (SNPs), which are theoretically unlimited in plant genomes. The reason why scientists chose SSRs over SNPs from the 1990s to early

2000s was that SNP discovery and genotyping with DNA sequencing was extremely expensive and complicated. However, NGS technologies have made SNPs the primary choice for many breeding studies due to their high flexibility, speed, and cost-effectiveness [41,42]. SNP markers have the potential to be universally used for genotyping from different sources, enabling integrated analysis across different species due to certain levels of similarities in nucleotide sequences. Although there are some ambiguous interpretations in some polyploidy species because of their biallelic nature, SNPs are still the most popular in modern genotyping experiments. The examples of SNP applications are shown in Figure 2 with some citations. As a consequence, a variety of concepts and methods have been adopted and used for SNP genotyping pipelines. The high-throughput application of SNPs for plant breeding can be largely divided into array- or PCR-based SNP genotyping platforms and NGS-based sequencing genotyping platforms. When the number of samples are small, the cost-effectiveness of these high-throughput genotyping methods may not meet our expectation but if one researcher has a fairly good number of samples, array- or PCR-based genotyping platforms can drastically reduce the cost per data point by virtue of their high levels of multiplexing. Of course, an NGS-based method is also high-throughput because it uses an NGS system to find SNPs based on the depth of the sequence information. While array- or PCR-based genotyping platforms require a priori knowledge of the nucleotide sequence information, the NGS-based method does not need it. Therefore, NGS-based genotyping can be widely used for species that do not have a reference genome, but it is not as accurate as array- or PCR-based platforms and not reproducible for different trials. The details of those platforms were previously discussed [41]. We think that the popularity of those platforms for the last decade needs to be checked in the current review article. To date, array- or PCR-based SNP genotyping platforms such as Taqman (Applied Biosystems), SNPlex (Applied Biosystems), BioMark HD (Fluidigm), KASPar (LGC), Axiom Biobank (Affymetrix), Infinium II (Illumina), GoldenGate (Illumina), and iPLEX (Sequenome) are commercially available. Some of those are being actively used in plant sciences based on the number of publications in the NCBI's PubMed database (www.pubmed.gov). Thanks to the flexibility of NGS platforms, a variety of NGS-based pipelines have been applied to plant sciences such as restriction association DNA sequencing (RAD-seq) [43], multiplex shotgun genotyping (MSG) [44] and genotype-by-sequencing (GBS) [45]. Poland and Rife summarized the NGS-based genotyping methods in their review article [46]. Herein, we integrate NGS-based genotyping to GBS because GBS seems the most popular method according to the publication search in the NCBI's PubMed database. Figure 3 shows searchable publications using various genotyping platforms applied to plant sciences from the NCBI's PubMed database from 2011 to present.

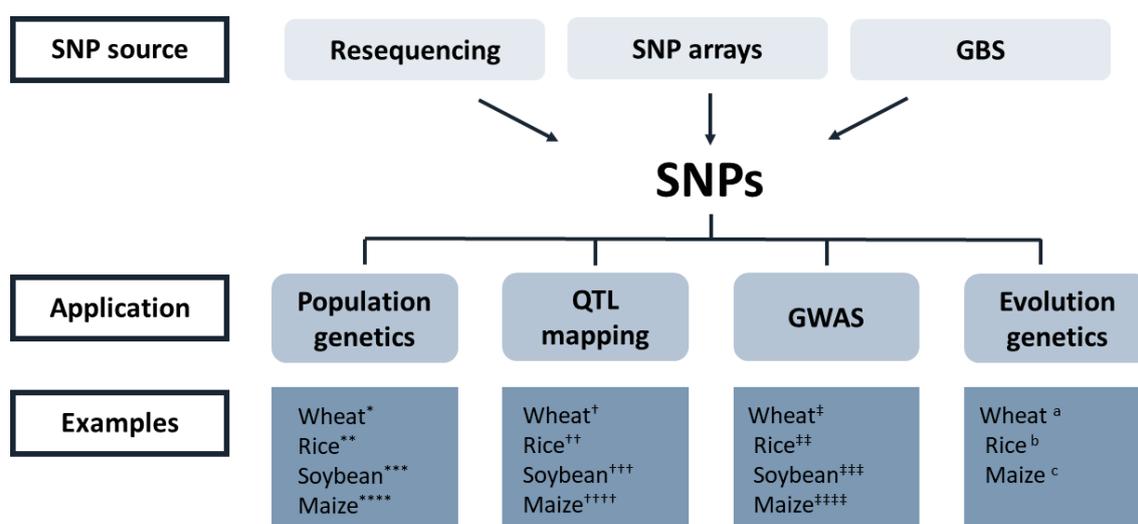


Figure 2. Applications of single nucleotide polymorphisms (SNPs) in plant sciences. Citations are as follows: * [47,48], ** [49], *** [50,51], **** [52–54], † [55–57], †† [58–60], ††† [50,61], †††† [62–64], ‡ [65,66], ‡‡ [67–69], ‡‡‡ [70,71], ‡‡‡‡ [62,72], ^a [73–75], ^b [76,77], ^c [78,79].

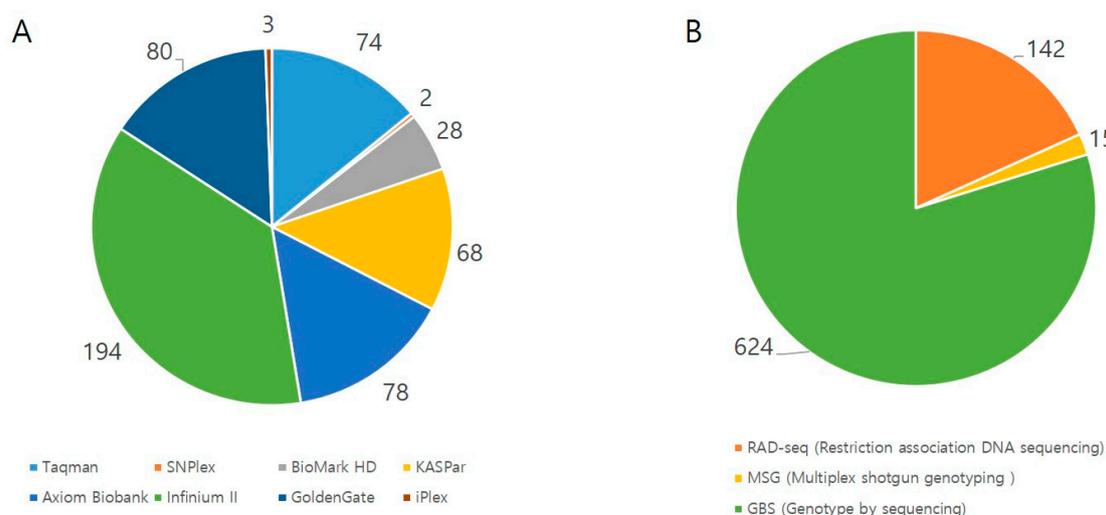


Figure 3. Publication search results (the NCBI's PubMed database, www.pubmed.gov) from 2011 to present using different SNP genotyping methods in plant sciences. (A). Array- or PCR-based genotyping platforms. (B). Next generation sequencing- (NGS)-based genotyping platforms.

For array- or PCR-based platforms, Taqman (multiplexed only) has been widely used for plant sciences. Axiom, GoldenGate, KASPar, and Infinium II have been moderately similar in terms of their use in searchable publications (Figure 3A). In fact, researchers may use the same platforms once they build an array of information, so the tendency may not dramatically change for the next few years.

The whole genome reference is a desirable gadget for calling SNP variants; however, GBS is relatively cost-effective because it generates partial genomic sequences utilizing restriction digestion. Although the reduced cost of nucleotide sequencing has popularized GBS for SNP analysis, genotyping errors, or missing data due to low coverage may be the biggest obstacle for SNP genotyping. In particular, as stated, it is a bit difficult to apply this technique for polyploidy crops, which can complicate SNP calling because of paralogous or homologous sequences. Nonetheless, for those species with no reference information, GBS may be the only option for SNP genotyping and discovery. As researchers already recognize that whole genome deep sequencing is not necessary for only genotyping purpose, reduced representation library (RRL) sequencing has been applied to the SNP genotyping procedure using NGS technology. Largely, three different methods are being used to date as stated, but GBS seems the most popular choice for plant science (Figure 3B). Those three methods are basically the same concept with restriction digestion and the ligation of bar-coded adaptor sequences. M procedures have been published such as double-digested RAD-seq [80], double-digested GBS [81], Ion Torrent GBS [82], and restriction fragment sequencing (REST-seq) [83]. Thus, researchers can choose adequate methods that may be the best fit for their experimental designs. Considering the advance in analysis techniques, genotyping methods would not be an issue for plant breeding.

High-throughput phenotyping (HTP) is also an indispensable methodology and needs to be briefly mentioned, in order to coordinate genomics data with breeding programs. Once, phenotyping required time-consuming and laborious processes. In other words, the number of phenotypic data points could be a stumbling block for combining phenotypic variations with genotypic variants. However, a variety of high-throughput phenotyping platforms (HTPP) have been introduced to plant breeding in the past few years thanks to the advancement of pre-existing technologies such as novel sensors, image analysis, robotics, and remote-sensing data mining (previously reviewed in Araus and Cairns [84]). HTPP enables plant breeders to evaluate numerous agricultural traits in a fast and accurate manner, which could be matched with the huge amount of genotypic data.

2.4. Genomics-Assisted Selection and Breeding

The whole genome reference is obviously omnipotent for crop breeding and genetics research because it enables gene discovery, positioning the gene location, accurate marker assignment, and the development of high resolution maps [85]. However, time, effort, and costs may be the biggest stumbling block for building pseudomolecular levels. Researchers already recognize that the whole genome reference is somewhat overpaid material for accomplishing a specific purpose. In other words, researchers may have to think about cost-effectiveness in order to not waste their resources to achieve their default targets. Plant breeders evaluate a variety of genetic materials such as core collections, bi-parental populations, diversity panels, breeding populations, and mutant lines. Here, we discuss some genomic tools that may be best applied for various breeding materials.

2.4.1. Characterization of Genetic Resources Using Genomics Tools

Plant breeding initially requires genetic variants bearing elite alleles for target traits. Breeders generally use a variety of genetic resources to obtain their desirable alleles such as well-designed bi- or multi-parental populations, diversity panels, core collections, and even intermediate breeding lines. The previously stated genotyping platforms can accelerate connecting genotype information to phenotypes, which is comparable to conventional methods in terms of an experimental size. The capacity to take care of a large number of individuals has enabled breeders to more precisely dissect genomic parts of interest based on the high resolution of the linkage disequilibrium (LD) based on a detailed haplotype analysis. A new concept was created, genomics-assisted breeding (GAB), thanks to their expanded and precise marker-assisted selection, leading to a new revolution in plant breeding, particularly for complex traits. The selection of pre-stated genotyping platform is, therefore, very important for cost-effective breeding programs. All the genetic populations listed above can make use of different SNP genotyping platforms, but there may be some things to consider. We do not discuss a resequencing method in this section due to its cost-effective issue despite its superior strength and accuracy in crop genomics. Commercial SNP-genotyping arrays are available for some major crops (refer to the Table 1 by You et al.) [86], which may be the most accurate and cost-effective way for high-throughput genotyping. If there is a reference genome for a certain crop but no commercial SNP array is available, in-house SNP arrays or reference-based GBS may be good options; however, in order to build an in-house SNP array, a prior knowledge of sequence information is necessary by genome resequencing; therefore, GBS may be the best choice in this case. GBS is relatively accurate and reliable when the whole genome reference is available; however, for those species without any reference genomes, GBS will be the only option for genotyping although the results will not be as accurate as other approaches. The point is that a variety of genotyping platforms enable plant breeders to observe more genomic breakpoints so that the introgression of target traits or genes into the elite cultivars has less deleterious effects, such as linkage drags.

When breeders cannot find desirable alleles, sometimes mutant lines (artificially- or naturally-induced) are investigated. For facilitating the identification of mutant alleles, genomic tools known as target induced local lesions in genomes (TILLING) [87] or ecotype TILLING (EcoTILLING) [88] are useful. These methods use some specialized nucleases, *CELI* or *EndoI*, which recognize and cut mismatched DNA sequences. In terms of excavating useful variants, the methods may be suitable for linking mutant genotypes to phenotypes without sequencing the whole genome. These methods have been applied to some major crops such as rice [89], wheat [90], barley [91], and maize [92].

2.4.2. Marker-Trait Associations for Expanded Marker-Assisted Selection

Conventional plant breeding was dependent upon a breeder's ability to make phenotypic evaluations, meaning that the result of elite selection could be biased by the breeders, and therefore, a bit subjective. The problem arises particularly for complex traits using phenotype-based selection because complex phenotypes show continuous variation in a breeding population, indicating that a

number of breeding accessions must be grown in the field in order to capture desirable variations, which are limited by labor, time, costs, space, environment, and so on. Ever since the PCR technique became predominant in plant science from 1980s, the initial concept of MAS was formed by adopting molecular markers in plant breeding programs with their objectivity and reproducibility, thanks to the denser dissection of genomes. Consequently, capturing genetic variants in various populations associated with agriculturally important traits underlying candidate genes has become a critical field in crop genetics and breeding science. In fact, researchers have worked on an accurate MTA for a successful and efficient MAS.

At the beginning of the era of molecular markers, some tried single-marker analysis (SMA) for MTA. This method consisted of fitting a regression model and running an F test analysis of variance (ANOVA) at each marker locus with phenotypic data based on the appropriate statistical models [93]. In fact, this type of analysis provided the basis for modern association mapping methods. However, the SMA is optimized for common or rare alleles with a minor allele frequency (MAF) greater than 0.01; consequently, it has a very low power for rare alleles and sometimes is inflated by type I errors (false positives) [94]. Nowadays, the SMA is not frequently used for MAS but for a quick screening of genotype and phenotype associations. Two recent main streams for MTA using genomic tools are quantitative trait loci (QTL) analysis and genome-wide association study (GWAS). The detection of QTL is basically performed by linkage mapping with bi-parental populations. Newly applied genomics tools enhance the accuracy of this procedure by enabling increases in the number of markers mapped, securing high-density genetic maps. Although QTL can detect MTA linked to rare alleles because it can be artificially introduced, the disadvantage of QTL mapping is mostly caused by the limited number of crossovers due to short generations of mapping populations. Therefore, QTL mapping gives a low resolution while it provides a high statistical power for detecting major alleles associated with traits having high heritability. To avoid these limitations, the mapping resolution can be drastically improved by adopting newly designed bi-parental populations such as a multi-parent advanced generation intercross (MAGIC) population [95]. High-throughput genotyping partly resolves these advantages by enabling researchers to investigate a large number of mapping individuals. In contrast, GWAS can accommodate many more recombination events that occurred in the history of the analyzing population, improving the resolution of MTA. Diversity panel, core collections, and even nested-association mapping (NAM) populations can be analyzed with this new powerful tool by virtue of high-throughput genotyping. Despite the higher resolution of GWAS compared to QTL mapping, it can generate false positives or true negatives due to subpopulation structures of the analyzing populations, and it requires a relatively larger population size than QTL mapping to capture accurate genotypic and phenotypic variations. Consequently, researchers need to be cautious to select mapping individuals that have large genotypic variations and to detect accurate associations that can be used for downstream usages such as breeding or genetic research. Although a large collection of the population is used for GWAS, the entire process can still be compromised due to an increased genetic heterogeneity if the careful selection of population individuals is not considered [96]. Moderate to high levels of heterozygosity can cause inaccurate genotyping, particularly when using low-coverage genomic sequence data such as GBS. For species with a high level of heterozygosity, researchers need to secure more sequence coverage to avoid any genotyping errors. Polyploidy (or paleopolyploidy) can result in genotyping errors through the confusion between subgenomic (or paralogous) and allelic variations. As stated, there is no resolution for this issue due to the biallelic nature of SNPs unless reference genomes are well established. Another issue for GWAS is that rare alleles are overlooked during filtering genotyping data based on minor allele frequency (MAF). In most of the cases, genotype data are filtered at $MAF \geq 0.05$, meaning that rare alleles can mostly be ignored. Unfortunately, rare alleles represent natural variations in some populations, which could be very important for breeding and genetic research. As stated, QTL may be more useful for rare alleles because bi-parental populations can be precisely designed by artificially introducing traits potentially associated with rare variants.

The final step for genomics-assisted breeding is marker-assisted selection of target accessions. Conventionally, selection was performed by breeders' personal choice based on phenotypic differences. Modern selection schemes, however, mostly depend on molecular markers, which are very objective and abundant in plant genomes, together with the advance of MTA. Since molecular markers have emerged, MAS has been applied to breeding many crops such as rice [97,98], wheat [99,100], sorghum [101,102] and soybean [103,104]. A plethora of genomic data dramatically improved the efficiency of MAS by densely dissecting genomic regions associated with important agricultural traits. MAS combined with genomics has emerged as a crucial tool for plant breeding; however, it is still difficult to select some traits phenotypically, especially for complex traits affected by environmental conditions or developmental stages. Additionally, MAS does not provide solutions for some issues caused by linkage drags. The issue can be resolved if many breeding accessions (or mapping individuals) are grown and investigated, but it is practically impossible. For example, breeders cannot grow over thousands of breeding accessions for each research purpose because the seed-producing ability of crops is not infinite. Even if a breeder can grow many accessions, it is still impossible to dissect genomic regions if favor alleles are tightly linked to deleterious alleles. As an alternative, geneticists started looking at haplotypes based on linkage disequilibrium (LD). Genomics big data combined with haplotype data enable plant scientists to exploit all the genotype and phenotype data as predictors of the genomic estimated breeding values (GEBVs). The new concept created a new selection scheme called GS, which will be further discussed in the next section.

3. Predictive Genomics and Breeding

3.1. Genomic Selection (GS)

Artificial selection of plant individuals or populations with desirable traits is one of the most important processes of plant breeding. An accurate selection would enable plant breeders to maintain a smaller size of breeding population as well as shorten the time needed for developing a new variety. In the early history of plant breeding, selections were made based only on the plant phenotypes, which are largely explained by their genotypes but also could vary in different environments, especially for quantitative traits. Nowadays, with the use of genomic tools, simple traits governed by one or a few genes can be effectively selected before evaluating their phenotypes [105,106].

Although the current MAS methods as described above are very effective for deploying a few major effect genes, the methods are shown to be inaccurate in predicting quantitative traits, mostly the agronomically important traits such as yield, seed weight, and quality [107,108]. This limitation, even with dense markers, is majorly due to the statistical approach, which is inadequate for polygenic traits—it only uses markers that are significantly correlated with a trait and fails to capture many loci of small effect [109,110]. Moreover, a biparental population, commonly used in QTL identification, is insufficient to represent allelic diversity and genetic background of the breeding population. This limits the translation of desirable QTLs directly to the breeding population with the equivalent genetic effect.

To overcome these limitations, Meuwissen et al. [111] proposed GS, which uses whole-genome prediction models for estimating genetic values. GS uses all available markers covering the whole genome to estimate breeding values (referred to as genomic estimated breeding values; GEBV) rather than testing its significance, while traditional MAS uses a predefined subset of significant markers with the rest treated as having zero effect [112,113]. By implementing GS with highly dense markers, which can be obtained by whole-genome sequencing and/or genotyping, both large and small QTL effects can be fully captured [109,110]. Moreover, unlike QTL mapping, there is no need to search for significant QTL-marker associations for each trait. QTL studies are still valuable for understanding the genetic architecture of quantitative traits; however, when considering a breeder's perspective, GS can reduce effort involved in maintaining individual biparental populations and mandatory steps to reassess the effects of QTLs within the breeding population. Therefore, by adapting

GS to plant breeding, the selection of elite genotypes can be achieved more efficiently and faster than with phenotypic selection or MAS approaches, and this can eventually shorten the breeding cycle of many crop species.

To perform GS, two sets of populations are required: a training population (TP) and a breeding (or selection) population (BP). GS builds a prediction model based on the genotypic and phenotypic data of individuals in the TP. The trained models are then applied to predict the breeding values of the genotyped individuals in BP without knowing their phenotypes. Breeders can subsequently select the individuals that are predicted to have superior phenotypes based on the GEBVs. During this selection process, the phenotype evaluations of any desirable traits are not necessary at all, only a genomic profile of the individuals, which can be cost-efficiently evaluated by various genotyping platforms, is required. Thereby, GS provides the ability to select the elite individuals without any phenotypic measurements, and breeders can benefit enormously from this. With GS, breeding becomes faster as the generation intervals between selection cycles become much shorter. Breeders do not need to wait for the population to grow until they can reach the development stages for phenotyping evaluation of an agronomic trait. Elite individuals can be selected even without planting the seeds, and this tremendously reduces the time and cost of the selection process. For example, major seed companies such as Monsanto incorporate GS effectively with their seed-chipping technologies and speed breeding methods to perform selections at least two or three times in a year by advancing the generations back and forth between open fields, greenhouses, and nurseries. Moreover, breeding becomes more intensive as more candidates can be evaluated per selection cycle [114,115]. Breeders can handle and screen a larger size of a breeding population, and even exotic germplasm and wild relatives can be tested to estimate the potential breeding values before introducing them to the breeding population. In this way, the performances of individuals with diverse genetic backgrounds can be assessed with a pre-trained model at a target environment that is either expected or predicted. A prediction model trained at different growing conditions could be able to tell which individuals or genomic profiles would perform best under the changing climate [116–118]. Therefore, GS has great potential to harness genetic gain in many aspects of breeding strategies.

The advantages of GS over traditional MAS have led to its application in plant and animal breeding, especially in livestock breeding [119–121]. In comparison with crop plants, livestock such as cattle, pigs, and chickens have a limited number of offspring that can be produced per generation and have a higher cost per progeny to raise and test their production index. Due to these distinct aspects, livestock sectors adopted GS more rapidly and widely, starting with dairy cattle in 2008 and later with beef cattle and chickens [122–124]. GS has been proven to have great potential to accelerate genetic gain in livestock breeding because the early selection using GS has doubled genetic gain when compared with the conventional progeny testing [125,126]. As shown in livestock breeding, the potential of GS is not questionable in general. However, the GS approaches from livestock breeding cannot be simply applied to complex plant breeding.

On the other hand, in the plant sectors, the potential of GS was explored in trees [127,128] and major crops such as wheat, maize, cassava, and soybean [50,116,129,130]. Although the true gain by GS remains unproven yet in plants, several results show that this approach is very effective, as seen in livestock; for example, proposed GS schemes for wheat breeding has estimated three to five times more genetic gain than that of the classical breeding program [114]. However, the implementation of GS to breeding programs has been limited to major international seed companies and public-funded programs for major cereal crops. This contrasting rate of GS application between plant and animal fields might be due to contrasting tools and resources available for breeders since plant breeders have several additional options in addition to GS to reduce generation times (e.g., double-haploid induction and speed breeding) and to replicate clones and inbreds (e.g., self or vegetative propagation) compared to animal breeders [131]. Nevertheless, to accelerate the wider application of GS in plant breeding programs, plant-specific approaches that differ from GS approaches in livestock breeding should be applied [132]. Recently, Xu et al. [133] suggested several requirements that should be

considered for the extensive use of GS in plants. Although most of the suggestions are similar to those in livestock, the authors suggested that building a better prediction model with high-throughput, cost-effectiveness, and precision phenotyping is the most important consideration for plants. This is because very few plants are produced in a controlled environment while most of the industrialized livestock are produced in environment-controlled facilities. Phenotypic performances of plants are largely affected by genotype-by-environment interaction, which has been continuously emphasized in numerous pieces of literature (see Table 2 for examples) as the most important consideration to improve the prediction accuracy of GS in plants. [114,134–136].

Although improving phenomics and statistical models are the current obstacles to overcome for advancing predictive breeding, GS would still be not possible without cost-effective methods to profile the whole genome. These genotyping methods, either array-based or sequence-based, produce massive amounts of single nucleotide polymorphisms (SNPs), which are commonly used in GS for training models and calculating GBEVs. Instead of SNPs, however, a set of alleles referred to as haplotypes could give more accurate information for the expected phenotype. Previous studies have shown that haplotypes consisting of ten markers have yielded the highest estimation accuracy for breeding values [137,138]. Additionally, even with a low-coverage sequencing of individuals, unobserved genotypes can be imputed to reconstruct the target haplotype using machine learning and/or deep learning frameworks [139,140]. Therefore, haplotype-based models can be a cost-effective option for GS and have the potential to increase the accuracy of GS as well.

Moreover, assembling the haplotypes shared by the group of individuals showing superior performance of the target traits was proposed as a new approach for developing improved crops [141]. An assembled set of the superior haplotypes, which is referred to as a haplotype assembly, can be substantially selected for haplotype-assisted breeding [142]. Presumably, these superior haplotypes consist of beneficial alleles, which can be derived with the help of artificial intelligence (AI). As suggested in a recent review [143], deep learning models can be used to predict molecular phenotypes of given genotypes and further make *ab initio* predictions on unobserved sequence data. Neural networks (NNs) as well as deep learning-based natural language processing (NLP) models, such as bidirectional encoder representations from transformers (BERT) [144] and generative pre-trained transformer (GPT) [145] models, can be applied to predict superior haplotypes of the target traits without knowing the phenotypes.

Table 2. Examples of genomic prediction using genotype by environment (G × E) interaction models in wheat and maize.

Species	Size of Population	Number of Environments	Number of Genotyped Markers	Traits	References
Wheat	599	4	1279	Grain yield	[146–149]
Wheat	693	4	15,744	Grain yield	[149–151]
Wheat	670	4	15,744	Grain yield	[149–151]
Wheat	807	5	14,217	Grain yield	[149–151]
Wheat	557	5	12,083	Grain yield	[152]
Wheat	338	4	7594	Grain yield, days to heading grain volume weight, 1000-kernel weight	[153,154]
Wheat	287	18	~15,000	Grain yield, grain number, thousand-grain weight, thermal time for flowering	[155]
Wheat	297	3	1635	Fusarium head blight	[156,157]
Wheat	250	5	12,083	Plant height, days to heading	[151,158]
Wheat	767	4	2038	Grain yield, plant height, days to heading, days to maturity	[159,160]
Wheat	775	5	2038	Grain yield, plant height, days to heading, days to maturity	[159,160]
Wheat	964	4	2038	Grain yield, plant height, days to heading, days to maturity	[159,160]
Wheat	980	4	2038	Grain yield, plant height, days to heading, days to maturity	[159,160]
Wheat	329	4	7748	14 traits (including grain yield, plant height)	[161]
Wheat	8416 ^a	3	39,758	Days to heading, days to maturity	[162]
Wheat	2374 ^a	3	39,758	Days to heading, days to maturity	[151,158,162]
Maize	504	3	158,281	Grain yield	[116,148,149]
Maize	309	3	158,281	Grain yield, plant height, anthesis-silking interval	[116,151,158,163,164]
Maize	278	3	46,347	Gray leaf spot	[165,166]

^a landrace accessions.

3.2. Predictive Breeding and Agriculture

GS may be considered as the first step of predictive breeding and agriculture. The selection of predicted phenotypes by GS only comprises a small portion of the entire breeding process. A predictive and precisely measurable breeding practice can be applied to not only the selection step but also can be applied throughout the entire breeding process. For example, plant breeders can design an ideal crop plant under future climate conditions [167], and then, the prediction model may show a shortcut for developing a designed variety. The genome profiles of the entire breeding population and their interaction with various environments would suggest ideal decisions throughout the breeding program to reach the designed varieties [168]. These decisions were previously made based on breeders' experience and intuition, which include the ideal combination of parental lines and the conditions for selection (or selective pressure).

This improved decision-making during plant breeding could be expanded to agriculture. Massive advances and transdisciplinary efforts in genomics, phenomics, and artificial intelligence are guiding us into an era of predictive agriculture [169]. Accurate measurement and prediction methods for agriculture would provide appropriate genetic and management solutions at once (genotype by environment by management predictions) to prepare for the future environment of agriculture [168]. Breeders and agronomists will be able to identify the combination of which genetic features and agronomic practices are expected to maximize the potential of target crops. The integration of multi-omics data such as transcriptome, epigenome, and metabolome would help to increase the accuracy of the genotype by environment by management (GxExM) prediction [170,171]. For this improved genome-to-phenome prediction, big data ecosystems for plant breeding need to be built (Figure 4). Diverse germplasm and big data of genomes, phenomes, and environments together with integrative analysis fueled by AI will enable us to precisely identify casual loci and predict breeding values. This will eventually guide us to faster and smarter breeding decisions and management practices.

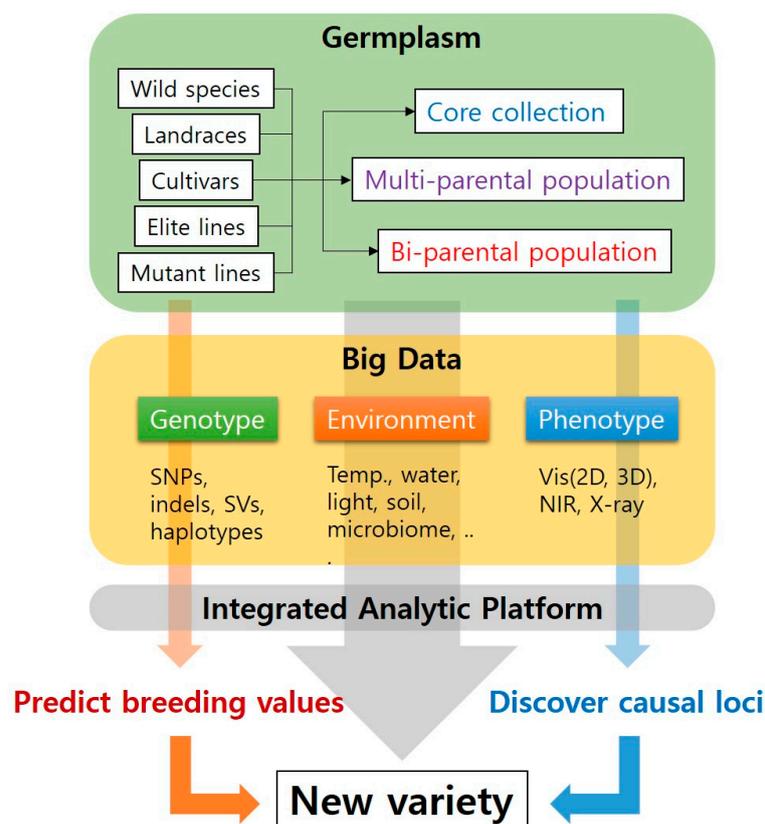


Figure 4. Flowchart of plant breeding in the era of genomics big data.

4. Future Directions

The tremendous advances in sequencing technologies and genomic tools have enabled us to explore vast amounts of genetic data from plant individuals in a cheaper and faster way. Plant breeders have so far successfully introduced these innovations and have developed numerous varieties with higher yield and better quality to ensure the food security of what we consume every day. In the coming decades, however, breeders will face more challenges to meet the global demands for food and feed. Several recent studies have repeatedly reported that climate change has a negative effect on global yields of many vital crops [172–175]. In addition to this, the global production of major crops needs to be increased by 60% in 2050 when considering the current trends in population and diets (e.g., 10 billion people with higher meat consumption) [176]. Even with the help of modern breeding techniques and platforms, the current rate of genetic gain is not enough to reach the projected food demands [177]. Plant breeders, therefore, need to find a more effective path to further enhance genetic gain and develop climate-change-ready varieties. As summarized in this review, many scientists and breeders suggest that genomic prediction and predictive breeding along with the big data of genomes and phenomes are the possible solutions to further increase the rate of genetic gain. Continuous production of genomic big data covering multi-omics data and an interpretation of the multi-dimensional data are also important to harness additional genetic gains. In the same context, private sectors have already incorporated these cutting-edge techniques into their breeding pipeline [178]. Moreover, interdisciplinary efforts could bring us big success as seen throughout the history of plant breeding, and genomics alone is not enough to innovate the current plant breeding paradigm. For instance, the advancements of HTP technologies would enable us to digitalize the detailed performances of plant individuals under different environments, and training these with phenome big data could increase the accuracy of the prediction model for GS [171]. Therefore, genomic prediction combined with technologies such as HTP, speed breeding, and/or genome editing could further boost the rate of genetic gain.

Even though genomic prediction is the most accurate method to select superior individuals, the approach requires profound resources of genomic and phenomic data and computational power. On the other hand, foreground selection with a single marker is still a very attractive method in most plant breeding programs. Therefore, different strategies and approaches should be applied separately for each breeding program after considering the available tools and resources. Most importantly, these applications need to be translated into delivering higher genetic gains in farmers' fields, not just in breeders' fields.

Author Contributions: K.D.K., Y.K. and C.K. wrote the entire manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the “Cooperative Research Program for Agriculture Science and Technology Development (Project No. PJ01315902)”, Rural Development Administration, Republic of Korea.

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

References

1. Doebley, J.F.; Gaut, B.S.; Smith, B.D. The molecular genetics of crop domestication. *Cell* **2006**, *127*, 1309–1321. [[PubMed](#)]
2. Mendel, G. *Experiments in Plant Hybridisation*; Harvard University Press: Cambridge, MA, USA, 1965.
3. Egan, A.N.; Schlueter, J.; Spooner, D.M. Applications of next-generation sequencing in plant biology. *Am. J. Bot.* **2012**, *99*, 175–185. [[PubMed](#)]
4. Collard, B.C.; Mackill, D.J. Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. B Biol. Sci.* **2008**, *363*, 557–572.
5. Hamblin, M.T.; Buckler, E.S.; Jannink, J.-L. Population genetics of genomics-based crop improvement methods. *Trends Genet.* **2011**, *27*, 98–106.

6. Libbrecht, M.W.; Noble, W.S. Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* **2015**, *16*, 321–332.
7. Kang, Y.; Kang, C.-S.; Kim, C. History of Nucleotide Sequencing Technologies: Advances in Exploring Nucleotide Sequences from Mendel to the 21st Century. *Hortic. Sci. Technol.* **2019**, *37*, 549–558.
8. Maxam, A.M.; Gilbert, W. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 560–564. [[PubMed](#)]
9. Sanger, F.; Nicklen, S.; Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5463–5467.
10. Li, F.W.; Harkess, A. A guide to sequence your favorite plant genomes. *Appl. Plant Sci.* **2018**, *6*, e1030. [[PubMed](#)]
11. Rothberg, J.M.; Leamon, J.H. The development and impact of 454 sequencing. *Nat. Biotechnol.* **2008**, *26*, 1117.
12. Turcatti, G.; Romieu, A.; Fedurco, M.; Tairi, A.P. A new class of cleavable fluorescent nucleotides: Synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic Acids Res.* **2008**, *36*, e25. [[CrossRef](#)]
13. Merriman, B.; Team, I.T.D.; Rothberg, J.M. Progress in ion torrent semiconductor chip based sequencing. *Electrophoresis* **2012**, *33*, 3397–3417. [[CrossRef](#)]
14. Eid, J.; Fehr, A.; Gray, J.; Luong, K.; Lyle, J.; Otto, G.; Peluso, P.; Rank, D.; Baybayan, P.; Bettman, B. Real-time DNA sequencing from single polymerase molecules. *Science* **2009**, *323*, 133–138. [[CrossRef](#)] [[PubMed](#)]
15. Branton, D.; Deamer, D.W.; Marziali, A.; Bayley, H.; Benner, S.A.; Butler, T.; Di Ventra, M.; Garaj, S.; Hibbs, A.; Huang, X. The potential and challenges of nanopore sequencing. In *Nanoscience and Technology: A Collection of Reviews from Nature Journals*; World Scientific: Singapore, 2010; pp. 261–268.
16. Kaul, S.; Koo, H.L.; Jenkins, J.; Rizzo, M.; Rooney, T.; Tallon, L.J.; Feldblyum, T.; Nierman, W.; Benito, M.I.; Lin, X.; et al. *Arabidopsis* Genome Initiative, Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **2000**, *408*, 796–815.
17. Goff, S.A.; Ricke, D.; Lan, T.H.; Presting, G.; Wang, R.; Dunn, M.; Glazebrook, J.; Sessions, A.; Oeller, P.; Varma, H.; et al. A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *japonica*). *Science* **2002**, *296*, 92–100.
18. Schnable, P.S.; Ware, D.; Fulton, R.S.; Stein, J.C.; Wei, F.; Pasternak, S.; Liang, C.; Zhang, J.; Fulton, L.; Graves, T.A.; et al. The B73 maize genome: Complexity, diversity, and dynamics. *Science* **2009**, *326*, 1112–1115. [[CrossRef](#)]
19. Paterson, A.H.; Bowers, J.E.; Bruggmann, R.; Dubchak, I.; Grimwood, J.; Gundlach, H.; Haberler, G.; Hellsten, U.; Mitros, T.; Poliakov, A.; et al. The Sorghum bicolor genome and the diversification of grasses. *Nature* **2009**, *457*, 551–556. [[CrossRef](#)] [[PubMed](#)]
20. Tuskan, G.A.; Difazio, S.; Jansson, S.; Bohlmann, J.; Grigoriev, I.; Hellsten, U.; Putnam, N.; Ralph, S.; Rombauts, S.; Salamov, A.; et al. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **2006**, *313*, 1596–1604. [[PubMed](#)]
21. Jaillon, O.; Aury, J.M.; Noel, B.; Policriti, A.; Clepet, C.; Casagrande, A.; Choisne, N.; Aubourg, S.; Vitulo, N.; Jubin, C.; et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **2007**, *449*, 463–467. [[PubMed](#)]
22. Ming, R.; Hou, S.; Feng, Y.; Yu, Q.; Dionne-Laporte, A.; Saw, J.H.; Senin, P.; Wang, W.; Ly, B.V.; Lewis, K.L.; et al. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* **2008**, *452*, 991–996. [[CrossRef](#)] [[PubMed](#)]
23. Schmutz, J.; Cannon, S.B.; Schlueter, J.; Ma, J.; Mitros, T.; Nelson, W.; Hyten, D.L.; Song, Q.; Thelen, J.J.; Cheng, J.; et al. Genome sequence of the palaeopolyploid soybean. *Nature* **2010**, *463*, 178–183. [[CrossRef](#)]
24. Chen, F.; Dong, W.; Zhang, J.; Guo, X.; Chen, J.; Wang, Z.; Lin, Z.; Tang, H.; Zhang, L. The Sequenced Angiosperm Genomes and Genome Databases. *Front. Plant Sci.* **2018**, *9*, 418. [[CrossRef](#)]
25. Lamesch, P.; Berardini, T.Z.; Li, D.; Swarbreck, D.; Wilks, C.; Sasidharan, R.; Muller, R.; Dreher, K.; Alexander, D.L.; Garcia-Hernandez, M. The *Arabidopsis* Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Res.* **2012**, *40*, D1202–D1210. [[CrossRef](#)] [[PubMed](#)]
26. Cheng, C.Y.; Krishnakumar, V.; Chan, A.P.; Thibaud-Nissen, F.; Schobel, S.; Town, C.D. Araport11: A complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* **2017**, *89*, 789–804. [[CrossRef](#)] [[PubMed](#)]
27. Wicker, T.; Schlagenhauf, E.; Graner, A.; Close, T.J.; Keller, B.; Stein, N. 454 sequencing put to the test using the complex genome of barley. *BMC Genom.* **2006**, *7*, 275. [[CrossRef](#)] [[PubMed](#)]

28. Beier, S.; Himmelbach, A.; Colmsee, C.; Zhang, X.-Q.; Barrero, R.A.; Zhang, Q.; Li, L.; Bayer, M.; Bolser, D.; Taudien, S. Construction of a map-based reference genome sequence for barley, *Hordeum vulgare* L. *Sci. Data* **2017**, *4*, 1–24. [[CrossRef](#)]
29. Mascher, M.; Gundlach, H.; Himmelbach, A.; Beier, S.; Twardziok, S.O.; Wicker, T.; Radchuk, V.; Dockter, C.; Hedley, P.E.; Russell, J. A chromosome conformation capture ordered sequence of the barley genome. *Nature* **2017**, *544*, 427–433. [[CrossRef](#)]
30. Kawahara, Y.; de la Bastide, M.; Hamilton, J.P.; Kanamori, H.; McCombie, W.R.; Ouyang, S.; Schwartz, D.C.; Tanaka, T.; Wu, J.; Zhou, S. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* **2013**, *6*, 4. [[CrossRef](#)]
31. Hirsch, C.N.; Hirsch, C.D.; Brohammer, A.B.; Bowman, M.J.; Soifer, I.; Barad, O.; Shem-Tov, D.; Baruch, K.; Lu, F.; Hernandez, A.G. Draft assembly of elite inbred line PH207 provides insights into genomic and transcriptome diversity in maize. *Plant Cell* **2016**, *28*, 2700–2714. [[CrossRef](#)]
32. Ouyang, S.; Zhu, W.; Hamilton, J.; Lin, H.; Campbell, M.; Childs, K.; Thibaud-Nissen, F.; Malek, R.L.; Lee, Y.; Zheng, L. The TIGR rice genome annotation resource: Improvements and new features. *Nucleic Acids Res.* **2007**, *35*, D883–D887. [[CrossRef](#)]
33. McCormick, R.F.; Truong, S.K.; Sreedasyam, A.; Jenkins, J.; Shu, S.; Sims, D.; Kennedy, M.; Amirebrahimi, M.; Weers, B.D.; McKinley, B. The Sorghum bicolor reference genome: Improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J.* **2018**, *93*, 338–354. [[PubMed](#)]
34. Cooper, E.A.; Brenton, Z.W.; Flinn, B.S.; Jenkins, J.; Shu, S.; Flowers, D.; Luo, F.; Wang, Y.; Xia, P.; Barry, K. A new reference genome for Sorghum bicolor reveals high levels of sequence similarity between sweet and grain genotypes: Implications for the genetics of sugar metabolism. *BMC Genom.* **2019**, *20*, 1–13.
35. Consortium, P.G.S. Genome sequence and analysis of the tuber crop potato. *Nature* **2011**, *475*, 189.
36. Sharma, S.K.; Bolser, D.; de Boer, J.; Sønderkær, M.; Amoros, W.; Carboni, M.F.; D’Ambrosio, J.M.; de la Cruz, G.; Di Genova, A.; Douches, D.S. Construction of reference chromosome-scale pseudomolecules for potato: Integrating the potato genome with genetic and physical maps. *G3 Genes Genomes Genet.* **2013**, *3*, 2031–2047.
37. Šafář, J.; Bartoš, J.; Janda, J.; Bellec, A.; Kubaláková, M.; Valárik, M.; Pateyron, S.; Weiserová, J.; Tušková, R.; Číhalíková, J. Dissecting large and complex genomes: Flow sorting and BAC cloning of individual chromosomes from bread wheat. *Plant J.* **2004**, *39*, 960–968. [[PubMed](#)]
38. Consortium, I.W.G.S. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* **2014**, *345*.
39. Lawrence, C.J.; Dong, Q.; Polacco, M.L.; Seigfried, T.E.; Brendel, V. MaizeGDB, the community database for maize genetics and genomics. *Nucleic Acids Res.* **2004**, *32*, D393–D397.
40. Deschamps, S.; Llaca, V.; May, G.D. Genotyping-by-Sequencing in Plants. *Biology* **2012**, *1*, 460–483.
41. Kim, C.; Guo, H.; Kong, W.; Chandnani, R.; Shuang, L.S.; Paterson, A.H. Application of genotyping by sequencing technology to a variety of crop breeding programs. *Plant Sci.* **2016**, *242*, 14–22. [[CrossRef](#)]
42. Chung, Y.S.; Choi, S.C.; Jun, T.-H.; Kim, C. Genotyping-by-sequencing: A promising tool for plant genetics research and breeding. *Hortic. Environ. Biotechnol.* **2017**, *58*, 425–431.
43. Baird, N.A.; Etter, P.D.; Atwood, T.S.; Currey, M.C.; Shiver, A.L.; Lewis, Z.A.; Selker, E.U.; Cresko, W.A.; Johnson, E.A. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* **2008**, *3*, e3376. [[CrossRef](#)] [[PubMed](#)]
44. Andolfatto, P.; Davison, D.; Erezyilmaz, D.; Hu, T.T.; Mast, J.; Sunayama-Morita, T.; Stern, D.L. Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Res.* **2011**, *21*, 610–617. [[CrossRef](#)] [[PubMed](#)]
45. 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](#)] [[PubMed](#)]
46. Poland, J.A.; Rife, T.W. Genotyping-by-Sequencing for Plant. *Breeding and Genetics. Plant Genome* **2012**, *5*, 92–102. [[CrossRef](#)]
47. Chao, S.; Dubcovsky, J.; Dvorak, J.; Luo, M.-C.; Baenziger, S.P.; Matnyazov, R.; Clark, D.R.; Talbert, L.E.; Anderson, J.A.; Dreisigacker, S. Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). *BMC Genom.* **2010**, *11*, 1–17. [[CrossRef](#)]

48. Würschum, T.; Langer, S.M.; Longin, C.F.H.; Korzun, V.; Akhunov, E.; Ebmeyer, E.; Schachschneider, R.; Schacht, J.; Kazman, E.; Reif, J.C. Population structure, genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers. *Theor. Appl. Genet.* **2013**, *126*, 1477–1486. [[CrossRef](#)] [[PubMed](#)]
49. Singh, N.; Choudhury, D.R.; Singh, A.K.; Kumar, S.; Srinivasan, K.; Tyagi, R.; Singh, N.; Singh, R. Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. *PLoS ONE* **2013**, *8*, e84136. [[CrossRef](#)]
50. Jarquín, D.; Kocak, K.; Posadas, L.; Hyma, K.; Jedlicka, J.; Graef, G.; Lorenz, A. Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC Genom.* **2014**, *15*, 740. [[CrossRef](#)]
51. Li, Y.H.; Li, W.; Zhang, C.; Yang, L.; Chang, R.Z.; Gaut, B.S.; Qiu, L.J. Genetic diversity in domesticated soybean (*Glycine max*) and its wild progenitor (*Glycine soja*) for simple sequence repeat and single-nucleotide polymorphism loci. *New Phytol.* **2010**, *188*, 242–253. [[CrossRef](#)]
52. Frascaroli, E.; Schrag, T.A.; Melchinger, A.E. Genetic diversity analysis of elite European maize (*Zea mays* L.) inbred lines using AFLP, SSR, and SNP markers reveals ascertainment bias for a subset of SNPs. *Theor. Appl. Genet.* **2013**, *126*, 133–141.
53. Ganal, M.W.; Durstewitz, G.; Polley, A.; Bérard, A.; Buckler, E.S.; Charcosset, A.; Clarke, J.D.; Graner, E.-M.; Hansen, M.; Joets, J. A large maize (*Zea mays* L.) SNP genotyping array: Development and germplasm genotyping; genetic mapping to compare with the B73 reference genome. *PLoS ONE* **2011**, *6*, e28334. [[CrossRef](#)]
54. Van Inghelandt, D.; Melchinger, A.E.; Lebreton, C.; Stich, B. Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theor. Appl. Genet.* **2010**, *120*, 1289–1299.
55. Zhang, W.; Chao, S.; Manthey, F.; Chicaiza, O.; Brevis, J.; Echenique, V.; Dubcovsky, J. QTL analysis of pasta quality using a composite microsatellite and SNP map of durum wheat. *Theor. Appl. Genet.* **2008**, *117*, 1361–1377. [[PubMed](#)]
56. Mu, J.; Huang, S.; Liu, S.; Zeng, Q.; Dai, M.; Wang, Q.; Wu, J.; Yu, S.; Kang, Z.; Han, D. Genetic architecture of wheat stripe rust resistance revealed by combining QTL mapping using SNP-based genetic maps and bulked segregant analysis. *Theor. Appl. Genet.* **2019**, *132*, 443–455. [[PubMed](#)]
57. Wu, Q.-H.; Chen, Y.-X.; Zhou, S.-H.; Fu, L.; Chen, J.-J.; Xiao, Y.; Zhang, D.; Ouyang, S.-H.; Zhao, X.-J.; Cui, Y. High-density genetic linkage map construction and QTL mapping of grain shape and size in the wheat population Yanda1817 × Beinong6. *PLoS ONE* **2015**, *10*, e0118144.
58. Ye, C.; Argayoso, M.A.; Redoña, E.D.; Sierra, S.N.; Laza, M.A.; Dilla, C.J.; Mo, Y.; Thomson, M.J.; Chin, J.; Delaviña, C.B. Mapping QTL for heat tolerance at flowering stage in rice using SNP markers. *Plant Breed.* **2012**, *131*, 33–41. [[CrossRef](#)]
59. Tiwari, S.; SL, K.; Kumar, V.; Singh, B.; Rao, A.; Mithra SV, A.; Rai, V.; Singh, A.K.; Singh, N.K. Mapping QTLs for salt tolerance in rice (*Oryza sativa* L.) by bulked segregant analysis of recombinant inbred lines using 50K SNP chip. *PLoS ONE* **2016**, *11*, e0153610. [[CrossRef](#)]
60. Famoso, A.N.; Zhao, K.; Clark, R.T.; Tung, C.-W.; Wright, M.H.; Bustamante, C.; Kochian, L.V.; McCouch, S.R. Genetic architecture of aluminum tolerance in rice (*Oryza sativa*) determined through genome-wide association analysis and QTL mapping. *PLoS Genet.* **2011**, *7*, e1002221. [[CrossRef](#)]
61. Iquira, E.; Humira, S.; François, B. Association mapping of QTLs for sclerotinia stem rot resistance in a collection of soybean plant introductions using a genotyping by sequencing (GBS) approach. *BMC Plant Biol.* **2015**, *15*, 5. [[CrossRef](#)]
62. Zhao, X.; Luo, L.; Cao, Y.; Liu, Y.; Li, Y.; Wu, W.; Lan, Y.; Jiang, Y.; Gao, S.; Zhang, Z. Genome-wide association analysis and QTL mapping reveal the genetic control of cadmium accumulation in maize leaf. *BMC Genom.* **2018**, *19*, 1–13. [[CrossRef](#)]
63. Dell’Acqua, M.; Gatti, D.M.; Pea, G.; Cattonaro, F.; Coppens, F.; Magris, G.; Hlaing, A.L.; Aung, H.H.; Nelissen, H.; Baute, J. Genetic properties of the MAGIC maize population: A new platform for high definition QTL mapping in *Zea mays*. *Genome Biol.* **2015**, *16*, 1–23. [[CrossRef](#)] [[PubMed](#)]
64. Chen, J.; Shrestha, R.; Ding, J.; Zheng, H.; Mu, C.; Wu, J.; Mahuku, G. Genome-wide association study and QTL mapping reveal genomic loci associated with Fusarium ear rot resistance in tropical maize germplasm. *G3 Genes Genomes Genet.* **2016**, *6*, 3803–3815. [[CrossRef](#)] [[PubMed](#)]

65. Gurung, S.; Mamidi, S.; Bonman, J.M.; Xiong, M.; Brown-Guedira, G.; Adhikari, T.B. Genome-wide association study reveals novel quantitative trait loci associated with resistance to multiple leaf spot diseases of spring wheat. *PLoS ONE* **2014**, *9*, e108179. [[CrossRef](#)] [[PubMed](#)]
66. Sukumaran, S.; Dreisigacker, S.; Lopes, M.; Chavez, P.; Reynolds, M.P. Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments. *Theor. Appl. Genet.* **2015**, *128*, 353–363. [[CrossRef](#)] [[PubMed](#)]
67. Chen, W.; Gao, Y.; Xie, W.; Gong, L.; Lu, K.; Wang, W.; Li, Y.; Liu, X.; Zhang, H.; Dong, H. Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat. Genet.* **2014**, *46*, 714–721. [[CrossRef](#)] [[PubMed](#)]
68. Huang, X.; Sang, T.; Zhao, Q.; Feng, Q.; Zhao, Y.; Li, C.; Zhu, C.; Lu, T.; Zhang, Z.; Li, M. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* **2010**, *42*, 961. [[CrossRef](#)]
69. Kumar, V.; Singh, A.; Mithra, S.A.; Krishnamurthy, S.; Parida, S.K.; Jain, S.; Tiwari, K.K.; Kumar, P.; Rao, A.R.; Sharma, S. Genome-wide association mapping of salinity tolerance in rice (*Oryza sativa*). *DNA Res.* **2015**, *22*, 133–145. [[CrossRef](#)]
70. Hwang, E.-Y.; Song, Q.; Jia, G.; Specht, J.E.; Hyten, D.L.; Costa, J.; Cregan, P.B. A genome-wide association study of seed protein and oil content in soybean. *BMC Genom.* **2014**, *15*, 1. [[CrossRef](#)]
71. Zhang, J.; Song, Q.; Cregan, P.B.; Nelson, R.L.; Wang, X.; Wu, J.; Jiang, G.-L. Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. *BMC Genom.* **2015**, *16*, 217. [[CrossRef](#)]
72. Tian, F.; Bradbury, P.J.; Brown, P.J.; Hung, H.; Sun, Q.; Flint-Garcia, S.; Rocheford, T.R.; McMullen, M.D.; Holland, J.B.; Buckler, E.S. Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat. Genet.* **2011**, *43*, 159–162. [[CrossRef](#)]
73. Luo, M.-C.; Gu, Y.Q.; You, F.M.; Deal, K.R.; Ma, Y.; Hu, Y.; Huo, N.; Wang, Y.; Wang, J.; Chen, S. A 4-gigabase physical map unlocks the structure and evolution of the complex genome of *Aegilops tauschii*, the wheat D-genome progenitor. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 7940–7945.
74. Peng, J.H.; Sun, D.; Nevo, E. Domestication evolution, genetics and genomics in wheat. *Mol. Breed.* **2011**, *28*, 281. [[CrossRef](#)]
75. Aveni, R.; Nave, M.; Barad, O.; Baruch, K.; Twardziok, S.O.; Gundlach, H.; Hale, I.; Mascher, M.; Spannagl, M.; Wiebe, K. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science* **2017**, *357*, 93–97. [[PubMed](#)]
76. Molina, J.; Sikora, M.; Garud, N.; Flowers, J.M.; Rubinstein, S.; Reynolds, A.; Huang, P.; Jackson, S.; Schaal, B.A.; Bustamante, C.D. Molecular evidence for a single evolutionary origin of domesticated rice. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 8351–8356. [[PubMed](#)]
77. Meyer, R.S.; Choi, J.Y.; Sanches, M.; Plessis, A.; Flowers, J.M.; Amas, J.; Dorph, K.; Barretto, A.; Gross, B.; Fuller, D.Q. Domestication history and geographical adaptation inferred from a SNP map of African rice. *Nat. Genet.* **2016**, *48*, 1083–1088. [[PubMed](#)]
78. Hufford, M.B.; Xu, X.; Van Heerwaarden, J.; Pyhäjärvi, T.; Chia, J.-M.; Cartwright, R.A.; Elshire, R.J.; Glaubitz, J.C.; Guill, K.E.; Kaeppeler, S.M. Comparative population genomics of maize domestication and improvement. *Nat. Genet.* **2012**, *44*, 808–811.
79. Li, Q.; Li, L.; Yang, X.; Warburton, M.L.; Bai, G.; Dai, J.; Li, J.; Yan, J. Relationship, evolutionary fate and function of two maize co-orthologs of rice GW2 associated with kernel size and weight. *BMC Plant Biol.* **2010**, *10*, 143. [[CrossRef](#)] [[PubMed](#)]
80. 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.
81. Poland, J.A.; Brown, P.J.; Sorrells, M.E.; Jannink, J.L. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* **2012**, *7*, e32253.
82. Mascher, M.; Wu, S.; Amand, P.S.; Stein, N.; Poland, J. Application of genotyping-by-sequencing on semiconductor sequencing platforms: A comparison of genetic and reference-based marker ordering in barley. *PLoS ONE* **2013**, *8*, e76925.
83. Stolle, E.; Moritz, R.F. RESTseq—Efficient benchtop population genomics with RESTriCTION FRAGMENT SEQUENCING. *PLoS ONE* **2013**, *8*, e63960. [[CrossRef](#)]

84. Araus, J.L.; Cairns, J.E. Field high-throughput phenotyping: The new crop breeding frontier. *Trends Plant Sci.* **2014**, *19*, 52–61. [[CrossRef](#)] [[PubMed](#)]
85. Lai, J.; Li, R.; Xu, X.; Jin, W.; Xu, M.; Zhao, H.; Xiang, Z.; Song, W.; Ying, K.; Zhang, M.; et al. Genome-wide patterns of genetic variation among elite maize inbred lines. *Nat. Genet.* **2010**, *42*, 1027–1030. [[CrossRef](#)] [[PubMed](#)]
86. 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](#)]
87. Till, B.J.; Reynolds, S.H.; Greene, E.A.; Codomo, C.A.; Enns, L.C.; Johnson, J.E.; Burtner, C.; Odden, A.R.; Young, K.; Taylor, N.E.; et al. Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome Res.* **2003**, *13*, 524–530. [[CrossRef](#)] [[PubMed](#)]
88. Comai, L.; Young, K.; Till, B.J.; Reynolds, S.H.; Greene, E.A.; Codomo, C.A.; Enns, L.C.; Johnson, J.E.; Burtner, C.; Odden, A.R.; et al. Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. *Plant J.* **2004**, *37*, 778–786. [[CrossRef](#)]
89. Kadaru, S.B.; Yadav, A.S.; Fjellstrom, R.G.; Oard, J.H. Alternative ecotilling protocol for rapid, cost-effective single-nucleotide polymorphism discovery and genotyping in rice (*Oryza sativa* L.). *Plant Mol. Biol. Report.* **2006**, *24*, 3–22. [[CrossRef](#)]
90. Wang, J.; Sun, J.; Liu, D.; Yang, W.; Wang, D.; Tong, Y.; Zhang, A. Analysis of Pina and Pinb alleles in the micro-core collections of Chinese wheat germplasm by Ecotilling and identification of a novel Pinb allele. *J. Cereal Sci.* **2008**, *48*, 836–842. [[CrossRef](#)]
91. Caldwell, D.G.; McCallum, N.; Shaw, P.; Muehlbauer, G.J.; Marshall, D.F.; Waugh, R. A structured mutant population for forward and reverse genetics in Barley (*Hordeum vulgare* L.). *Plant J.* **2004**, *40*, 143–150. [[CrossRef](#)]
92. Weil, C.F.; Monde, R.-A. Getting the Point—Mutations in Maize. *Crop Sci.* **2007**, *4*, S60–S67. [[CrossRef](#)]
93. Soller, M.; Brody, T.; Genizi, A. On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. *Theor. Appl. Genet.* **1976**, *47*, 35–39. [[CrossRef](#)] [[PubMed](#)]
94. Konigorski, S.; Yilmaz, Y.E.; Pischon, T. Comparison of single-marker and multi-marker tests in rare variant association studies of quantitative traits. *PLoS ONE* **2017**, *12*, e0178504. [[CrossRef](#)] [[PubMed](#)]
95. Cavanagh, C.; Morell, M.; Mackay, I.; Powell, W. From mutations to MAGIC: Resources for gene discovery, validation and delivery in crop plants. *Curr. Opin. Plant Biol.* **2008**, *11*, 215–221. [[CrossRef](#)] [[PubMed](#)]
96. Alqudah, A.M.; Sallam, A.; Baenziger, P.S.; Börner, A. GWAS: Fast-forwarding gene identification and characterization in temperate Cereals: Lessons from Barley—A review. *J. Adv. Res.* **2020**, *22*, 119–135. [[CrossRef](#)] [[PubMed](#)]
97. Jena, K.; Mackill, D. Molecular markers and their use in marker-assisted selection in rice. *Crop Sci.* **2008**, *48*, 1266–1276. [[CrossRef](#)]
98. Das, G.; Patra, J.K.; Baek, K.-H. Insight into MAS: A Molecular Tool for Development of Stress Resistant and Quality of Rice through Gene Stacking. *Front. Plant Sci.* **2017**, *8*, 985. [[CrossRef](#)]
99. Buerstmayr, H.; Ban, T.; Anderson, J.A. QTL mapping and marker-assisted selection for Fusarium head blight resistance in wheat: A review. *Plant Breed.* **2009**, *128*, 1–26. [[CrossRef](#)]
100. Miedaner, T.; Korzun, V. Marker-assisted selection for disease resistance in wheat and barley breeding. *Phytopathology* **2012**, *102*, 560–566. [[CrossRef](#)]
101. Ejeta, G.; Knoll, J.E. Marker-assisted selection in sorghum. In *Genomics-Assisted Crop Improvement*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 187–205.
102. Madhusudhana, R. Chapter 6—Marker-Assisted Breeding in Sorghum. In *Breeding Sorghum for Diverse End Uses*; Aruna, C., Visarada, K.B.R.S., Bhat, B.V., Tonapi, V.A., Eds.; Woodhead Publishing: Cambridge, UK, 2019; pp. 93–114.
103. Hang, J.; Song, Q.; Cregan, P.B.; Jiang, G.-L. Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*). *Theor. Appl. Genet.* **2016**, *129*, 117–130.
104. Childs, S.P.; Buck, J.W.; Li, Z. Breeding soybeans with resistance to soybean rust (*Phakopsora pachyrhizi*). *Plant Breed.* **2018**, *137*, 250–261. [[CrossRef](#)]
105. Dekkers, J.C.M.; Hospital, F. The use of molecular genetics in the improvement of agricultural populations. *Nat. Rev. Genet.* **2002**, *3*, 22–32. [[PubMed](#)]

106. Lande, R.; Thompson, R. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* **1990**, *124*, 743–756.
107. Bernardo, R. Molecular Markers and Selection for Complex. Traits in Plants: Learning from the Last 20 Years. *Crop Sci.* **2008**, *48*, 1649–1664.
108. Xu, Y.; Crouch, J.H. Marker-Assisted Selection in Plant. Breeding: From Publications to Practice. *Crop Sci.* **2008**, *48*, 391–407.
109. Heffner, E.L.; Sorrells, M.E.; Jannink, J.-L. Genomic Selection for Crop Improvement. *Crop Sci.* **2009**, *49*, 1–12.
110. Goddard, M.E.; Hayes, B.J. Genomic selection. *J. Anim. Breed. Genet.* **2007**, *124*, 323–330.
111. Meuwissen, T.H.E.; Hayes, B.J.; MGoddard, E. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* **2001**, *157*, 1819–1829.
112. Nakaya, A.; Isobe, S.N. Will genomic selection be a practical method for plant breeding? *Ann. Bot.* **2012**, *110*, 1303–1316.
113. Jannink, J.-L.; Lorenz, A.J.; Iwata, H. Genomic selection in plant breeding: From theory to practice. *Brief. Funct. Genom.* **2010**, *9*, 166–177.
114. Bassi, F.M.; Bentley, A.R.; Charmet, G.; Ortiz, R.; Crossa, J. Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). *Plant Sci.* **2016**, *242*, 23–36. [[CrossRef](#)]
115. Cobb, J.N.; Juma, R.U.; Biswas, P.S.; Arbelaez, J.D.; Rutkoski, J.; Atlin, G.; Hagen, T.; Quinn, M.; Ng, E.H. Enhancing the rate of genetic gain in public-sector plant breeding programs: Lessons from the breeder's equation. *Theor. Appl. Genet.* **2019**, *132*, 627–645. [[CrossRef](#)] [[PubMed](#)]
116. Crossa, J.; Beyene, Y.; Kassa, S.; Pérez, P.; Hickey, J.M.; Chen, C.; de los Campos, G.; Burgueño, J.; Windhausen, V.S.; Buckler, E.; et al. Genomic Prediction in Maize Breeding Populations with Genotyping-by-Sequencing. *G3 Genes Genomes Genet.* **2013**, *3*, 1903–1926. [[CrossRef](#)] [[PubMed](#)]
117. Beyene, Y.; Semagn, K.; Mugo, S.; Tarekegne, A.; Babu, R.; Meisel, B.; Sehabiague, P.; Makumbi, D.; Magorokosho, C.; Oikeh, S.; et al. Genetic Gains in Grain Yield Through Genomic Selection in Eight Bi-parental Maize Populations under Drought Stress. *Crop Sci.* **2015**, *55*, 154–163. [[CrossRef](#)]
118. Zhang, X.; Pérez-Rodríguez, P.; Semagn, K.; Beyene, Y.; Babu, R.; López-Cruz, M.A.; San Vicente, F.; Olsen, M.; Buckler, E.; Jannink, J.L.; et al. Genomic prediction in biparental tropical maize populations in water-stressed and well-watered environments using low-density and GBS SNPs. *Heredity* **2015**, *114*, 291–299. [[CrossRef](#)] [[PubMed](#)]
119. Weller, J.I.; Ezra, E.; Ron, M. Invited review: A perspective on the future of genomic selection in dairy cattle. *J. Dairy Sci.* **2017**, *100*, 8633–8644. [[CrossRef](#)] [[PubMed](#)]
120. Hayes, B.J.; Lewin, H.A.; Goddard, M.E. The future of livestock breeding: Genomic selection for efficiency, reduced emissions intensity adaptation. *Trends Genet.* **2013**, *29*, 206–214. [[CrossRef](#)]
121. Meuwissen, T.; Hayes, B.; Goddard, M. Genomic selection: A paradigm shift in animal breeding. *Anim. Front.* **2016**, *6*, 6–14. [[CrossRef](#)]
122. Mehrban, H.; Lee, D.H.; Moradi, M.H.; IlCho, C.; Naserkheil, M.; Ibáñez-Escriche, N. Predictive performance of genomic selection methods for carcass traits in Hanwoo beef cattle: Impacts of the genetic architecture. *Genet. Sel. Evol.* **2017**, *49*, 1. [[CrossRef](#)]
123. Wolc, A.; Zhao, H.H.; Arango, J.; Settar, P.; Fulton, J.E.; O'Sullivan, N.P.; Preisinger, R.; Stricker, C.; Habier, D.; Fernando, R.L.; et al. Response and inbreeding from a genomic selection experiment in layer chickens. *Genet. Sel. Evol.* **2015**, *47*, 59. [[CrossRef](#)]
124. Lu, D.; Akanno, E.C.; Crowley, J.J.; Schenkel, F.; Li, H.; De Pauw, M.; Moore, S.S.; Wang, Z.; Li, C.; Stothard, P.; et al. Accuracy of genomic predictions for feed efficiency traits of beef cattle using 50K and imputed HD genotypes1. *J. Anim. Sci.* **2016**, *94*, 1342–1353. [[CrossRef](#)]
125. Wiggans, G.R.; Cole, J.B.; Hubbard, S.M.; Sonstegard, T.S. Genomic Selection in Dairy Cattle: The USDA Experience. *Ann. Rev. Anim. Biosci.* **2017**, *5*, 309–327. [[CrossRef](#)] [[PubMed](#)]
126. Georges, M.; Charlier, C.; Hayes, B. Harnessing genomic information for livestock improvement. *Nat. Rev. Genet.* **2019**, *20*, 135–156. [[CrossRef](#)]
127. Grattapaglia, D.; Resende, M.D. Genomic selection in forest tree breeding. *Tree Genet. Genomes* **2011**, *7*, 241–255. [[CrossRef](#)]

128. Resende, M.D.V.; Resende, M.F.R., Jr.; Sansaloni, C.P.; Petroli, C.D.; Missiaggia, A.A.; Aguiar, A.M.; Abad, J.M.; Takahashi, E.K.; Rosado, A.M.; Faria, D.A.; et al. Genomic selection for growth and wood quality in Eucalyptus: Capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytol.* **2012**, *194*, 116–128. [[CrossRef](#)] [[PubMed](#)]
129. Poland, J.; Endelman, J.; Dawson, J.; Rutkoski, J.; Wu, S.; Manes, Y.; Dreisigacker, S.; Crossa, J.; Sánchez-Villeda, H.; Sorrells, M.; et al. Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. *Plant Genome* **2012**, *5*, 103–113. [[CrossRef](#)]
130. Ly, D.; Hamblin, M.; Rabbi, I.; Melaku, G.; Bakare, M.; Gauch, H.G., Jr.; Okechukwu, R.; Dixon, A.G.O.; Kulakow, P.; Jannink, J.-L. Relatedness and Genotype × Environment Interaction Affect Prediction Accuracies in Genomic Selection: A Study in Cassava. *Crop Sci.* **2013**, *53*, 1312–1325. [[CrossRef](#)]
131. Hickey, J.M.; Chiurugwi, T.; Mackay, I.; Powell, W.; Hickey, J.M.; Chiurugwi, T.; Mackay, I.; Powell, W.; Eggen, A.; Kilian, A.; et al. Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nat. Genet.* **2017**, *49*, 1297–1303. [[CrossRef](#)]
132. Jonas, E.; de Koning, D.-J. Does genomic selection have a future in plant breeding? *Trends Biotechnol.* **2013**, *31*, 497–504. [[CrossRef](#)]
133. Xu, Y.; Liu, X.; Fu, J.; Wang, H.; Wang, J.; Huang, C.; Prasanna, B.M.; Olsen, M.S.; Wang, G.; Zhang, A. Enhancing Genetic Gain through Genomic Selection: From Livestock to Plants. *Plant Commun.* **2020**, *1*, 100005. [[CrossRef](#)]
134. Desta, Z.A.; Ortiz, R. Genomic selection: Genome-wide prediction in plant improvement. *Trends Plant Sci.* **2014**, *19*, 592–601. [[CrossRef](#)] [[PubMed](#)]
135. Voss-Fels, K.P.; Cooper, M.; Hayes, B.J. Accelerating crop genetic gains with genomic selection. *Theor. Appl. Genet.* **2019**, *132*, 669–686.
136. Crossa, J.; Pérez-Rodríguez, P.; Cuevas, J.; Montesinos-López, O.; Jarquín, D.; de los Campos, G.; Burgueño, J.; González-Camacho, J.M.; Pérez-Elizalde, S.; Beyene, Y.; et al. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends Plant Sci.* **2017**, *22*, 961–975. [[PubMed](#)]
137. Calus, M.P.L.; Meuwissen, T.H.E.; de Roos, A.P.W.; Veerkamp, R.F. Accuracy of Genomic Selection Using Different Methods to Define Haplotypes. *Genetics* **2008**, *178*, 553–561.
138. Villumsen, T.M.; Janss, L.; Lund, M.S. The importance of haplotype length and heritability using genomic selection in dairy cattle. *J. Anim. Breed. Genet.* **2009**, *126*, 3–13.
139. Faux, P.; Geurts, P.; Druet, T. A Random Forests Framework for Modeling Haplotypes as Mosaics of Reference Haplotypes. *Front. Genet.* **2019**, *10*, 562. [[PubMed](#)]
140. Kojima, K.; Tadaka, S.; Katsuoaka, F.; Tamiya, G.; Yamamoto, M.; Kinoshita, K. A genotype imputation method for de-identified haplotype reference information by using recurrent neural network. *PLoS Comput. Biol.* **2020**, *16*, e1008207.
141. Bevan, M.W.; Uauy, C.; Wulff, B.B.H.; Zhou, J.; Krasileva, K.; Clark, M.D. Genomic innovation for crop improvement. *Nature* **2017**, *543*, 346–354. [[PubMed](#)]
142. Abbai, R.; Singh, V.K.; Nachimuthu, V.V.; Sinha, P.; Selvaraj, R.; Vipparla, A.K.; Singh, A.K.; Singh, U.M.; Varshney, R.K.; Kumar, A. Haplotype analysis of key genes governing grain yield and quality traits across 3K RG panel reveals scope for the development of tailor-made rice with enhanced genetic gains. *Plant Biotechnol. J.* **2019**, *17*, 1612–1622.
143. Wang, H.; Cimen, E.; Singh, N.; Buckler, E. Deep learning for plant genomics and crop improvement. *Curr. Opin. Plant Biol.* **2020**, *54*, 34–41.
144. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
145. Adford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I. *Improving Language Understanding by Generative Pre-Training*; Open AI: San Francisco, CA, USA, 11 June 2018.
146. Crossa, J.; Campos, G.d.I.; Pérez, P.; Gianola, D.; Burgueño, J.; Araus, J.L.; Makumbi, D.; Singh, R.P.; Dreisigacker, S.; Yan, J.; et al. Prediction of Genetic Values of Quantitative Traits in Plant Breeding Using Pedigree and Molecular Markers. *Genetics* **2010**, *186*, 713–724. [[CrossRef](#)] [[PubMed](#)]
147. Burgueño, J.; de los Campos, G.; Weigel, K.; Crossa, J. Genomic Prediction of Breeding Values when Modeling Genotype × Environment Interaction using Pedigree and Dense Molecular Markers. *Crop Sci.* **2012**, *52*, 707–719. [[CrossRef](#)]

148. Cuevas, J.; Crossa, J.; Soberanis, V.; Pérez-Elizalde, S.; Pérez-Rodríguez, P.; Campos, G.d.I.; Montesinos-López, O.A.; Burgueño, J. Genomic Prediction of Genotype \times Environment Interaction Kernel Regression Models. *Plant Genome* **2016**, *9*, 1–20. [[CrossRef](#)] [[PubMed](#)]
149. Cuevas, J.; Crossa, J.; Montesinos-López, O.A.; Burgueño, J.; Pérez-Rodríguez, P.; de los Campos, G. Bayesian Genomic Prediction with Genotype \times Environment Interaction Kernel Models. *G3 Genes Genomes Genet.* **2017**, *7*, 41–53. [[CrossRef](#)] [[PubMed](#)]
150. Lopez-Cruz, M.; Crossa, J.; Bonnett, D.; Dreisigacker, S.; Poland, J.; Jannink, J.-L.; Singh, R.P.; Autrique, E.; de los Campos, G. Increased Prediction Accuracy in Wheat Breeding Trials Using a Marker \times Environment Interaction Genomic Selection Model. *G3 Genes Genomes Genet.* **2015**, *5*, 569–582. [[CrossRef](#)]
151. Montesinos-López, A.; Montesinos-López, O.A.; Gianola, D.; Crossa, J.; Hernández-Suárez, C.M. Multi-environment Genomic Prediction of Plant Traits Using Deep Learners With Dense Architecture. *G3 Genes Genomes Genet.* **2018**, *8*, 3813–3828. [[CrossRef](#)]
152. Rutkoski, J.; Poland, J.; Mondal, S.; Autrique, E.; Pérez, L.G.; Crossa, J.; Reynolds, M.; Singh, R. Canopy-Temperature and Vegetation Indices from High-Throughput Phenotyping Improve Accuracy of Pedigree and Genomic Selection for Grain Yield in Wheat. *G3 Genes Genomes Genet.* **2016**, *6*, 2799–2808.
153. Milner, S.G.; Maccaferri, M.; Huang, B.E.; Mantovani, P.; Massi, A.; Frascaroli, E.; Tuberosa, R.; Salvi, S. A multiparental cross population for mapping QTL for agronomic traits in durum wheat (*Triticum turgidum* ssp. *durum*). *Plant Biotechnol. J.* **2016**, *14*, 735–748. [[CrossRef](#)]
154. Crossa, J.; de los Campos, G.; Maccaferri, M.; Tuberosa, R.; Burgueño, J.; Pérez-Rodríguez, P. Extending the Marker \times Environment Interaction Model for Genomic-Enabled Prediction and Genome-Wide Association Analysis in Durum Wheat. *Crop Sci.* **2016**, *56*, 2193–2209.
155. Sukumaran, S.; Crossa, J.; Jarquin, D.; Lopes, M.; Reynolds, M.P. Genomic Prediction with Pedigree and Genotype \times Environment Interaction in Spring Wheat Grown in South and West Asia, North Africa, and Mexico. *G3 Genes Genomes Genet.* **2017**, *7*, 481–495.
156. Montesinos-López, A.; Montesinos-López, O.A.; Crossa, J.; Burgueño, J.; Eskridge, K.M.; Falconi-Castillo, E.; He, X.; Singh, P.; Cichy, K. Genomic Bayesian Prediction Model for Count Data with Genotype \times Environment Interaction. *G3 Genes Genomes Genet.* **2016**, *6*, 1165–1177.
157. Montesinos-López, O.A.; Montesinos-López, A.; Crossa, J.; Toledo, F.H.; Montesinos-López, J.C.; Singh, P.; Juliana, P.; Salinas-Ruiz, J. A Bayesian Poisson-lognormal Model for Count Data for Multiple-Trait Multiple-Environment Genomic-Enabled Prediction. *G3 Genes Genomes Genet.* **2017**, *7*, 1595–1606.
158. Montesinos-López, O.A.; Montesinos-López, A.; Crossa, J.; Gianola, D.; Hernández-Suárez, C.M.; Martín-Vallejo, J. Multi-trait, Multi-environment Deep Learning Modeling for Genomic-Enabled Prediction of Plant Traits. *G3 Genes Genomes Genet.* **2018**, *8*, 3829–3840. [[CrossRef](#)] [[PubMed](#)]
159. Juliana, P.; Singh, R.P.; Poland, J.A.; Suchismita, M.; Crossa, J.; Montesinos-López, O.A. Prospects and Challenges of Applied Genomic Selection—A New Paradigm in Breeding for Grain Yield in Bread Wheat. *Plant Genome* **2018**, *11*, 180017. [[CrossRef](#)]
160. Montesinos-López, O.A.; Martín-Vallejo, J.; Crossa, J.; Gianola, D.; Hernández-Suárez, C.M.; Montesinos-López, A.; Juliana, P.; Singh, R. New Deep Learning Genomic-Based Prediction Model for Multiple Traits with Binary, Ordinal, and Continuous Phenotypes. *G3 Genes Genomes Genet.* **2019**, *9*, 1545–1556.
161. Ward, B.P.; Brown-Guedira, G.; Tyagi, P.; Kolb, F.L.; Van Sanford, D.A.; Sneller, C.H.; Griffey, C.A. Multi-environment and Multitrait Genomic Selection Models in Unbalanced Early-Generation Wheat Yield Trials. *Crop Sci.* **2019**, *59*, 491–507. [[CrossRef](#)]
162. Crossa, J.; Jarquín, D.; Franco, J.; Pérez-Rodríguez, P.; Burgueño, J.; Saint-Pierre, C.; Vikram, P.; Sansaloni, C.; Petroli, C.; Akdemir, D.; et al. Genomic Prediction of Gene Bank Wheat Landraces. *G3 Genes Genomes Genet.* **2016**, *6*, 1819–1834. [[CrossRef](#)]
163. Montesinos-López, O.A.; Montesinos-López, A.; Crossa, J.; Toledo, F.H.; Pérez-Hernández, O.; Eskridge, K.M.; Rutkoski, J. A Genomic Bayesian Multi-trait and Multi-environment Model. *G3 Genes Genomes Genet.* **2016**, *6*, 2725–2744. [[CrossRef](#)]
164. Montesinos-López, O.A.; Montesinos-López, A.; Crossa, J.; Montesinos-López, J.C.; Luna-Vázquez, F.J.; Salinas-Ruiz, J.; Herrera-Morales, J.R.; Buenrostro-Mariscal, R. A Variational Bayes Genomic-Enabled Prediction Model with Genotype \times Environment Interaction. *G3 Genes Genomes Genet.* **2017**, *7*, 1833–1853.

165. Montesinos-López, O.A.; Montesinos-López, A.; Pérez-Rodríguez, P.; de los Campos, G.; Eskridge, K.; Crossa, J. Threshold Models for Genome-Enabled Prediction of Ordinal Categorical Traits in Plant Breeding. *G3 Genes Genomes Genet.* **2015**, *5*, 291–300. [[CrossRef](#)]
166. Montesinos-López, O.A.; Montesinos-López, A.; Crossa, J.; Burgueño, J.; Eskridge, K. Genomic-Enabled Prediction of Ordinal Data with Bayesian Logistic Ordinal Regression. *G3 Genes Genomes Genet.* **2015**, *5*, 2113–2126. [[CrossRef](#)]
167. Hammer, G.L.; McLean, G.; van Oosterom, E.; Chapman, S.; Zheng, B.; Wu, A.; Doherty, A.; Jordan, D. Designing crops for adaptation to the drought and high-temperature risks anticipated in future climates. *Crop Sci.* **2020**, *60*, 605–621. [[CrossRef](#)]
168. Cooper, M.; Tang, T.; Gho, C.; Hart, T.; Hammer, G.; Messina, C. Integrating genetic gain and gap analysis to predict improvements in crop productivity. *Crop Sci.* **2020**, *60*, 582–604. [[CrossRef](#)]
169. Messina, C.D.; Cooper, M.; Reynolds, M.; Hammer, G.L. Crop science: A foundation for advancing predictive agriculture. *Crop Sci.* **2020**, *60*, 544–546. [[CrossRef](#)]
170. Washburn, J.D.; Burch, M.B.; Franco, J.A.V. Predictive breeding for maize: Making use of molecular phenotypes, machine learning.; physiological crop models. *Crop Sci.* **2020**, *60*, 622–638. [[CrossRef](#)]
171. Harfouche, A.L.; Jacobson, D.A.; Kainer, D.; Romero, J.C.; Harfouche, A.H.; Scarascia Mugnozza, G.; Moshelion, M.; Tuskan, G.A.; Keurentjes, J.J.B.; Altman, A. Accelerating Climate Resilient Plant breeding by Applying Next-Generation Artificial Intelligence. *Trends Biotechnol.* **2019**, *37*, 1217–1235.
172. Zhao, C.; Liu, B.; Piao, S.; Wang, X.; Lobell, D.B.; Huang, Y.; Huang, M.; Yao, Y.; Bassu, S.; Ciais, P.; et al. Temperature increase reduces global yields of major crops in four independent estimates. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 9326–9331. [[CrossRef](#)]
173. Asseng, S.; Ewert, F.; Martre, P.; Rötter, R.P.; Lobell, D.B.; Cammarano, D.; Kimball, B.A.; Ottman, M.J.; Wall, G.W.; White, J.W.; et al. Rising temperatures reduce global wheat production. *Nat. Clim. Chang.* **2015**, *5*, 143–147. [[CrossRef](#)]
174. Deutsch, C.A.; Tewksbury, J.J.; Tigchelaar, M.; Battisti, D.S.; Merrill, S.C.; Huey, R.B.; Naylor, R.L. Increase in crop losses to insect pests in a warming climate. *Science* **2018**, *361*, 916–919. [[CrossRef](#)]
175. Scheelbeek, P.F.D.; Bird, F.A.; Tuomisto, H.L.; Green, R.; Harris, F.B.; Joy, E.J.M.; Chalabi, Z.; Allen, E.; Haines, A.; Dangour, A.D. Effect of environmental changes on vegetable and legume yields and nutritional quality. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 6804–6809. [[CrossRef](#)]
176. Alexandratos, N.; Bruinsma, J. *World Agriculture Towards 2030/2050: The 2012 Revision*; ESA Working Papers 12-03; Food and Agriculture Organization of the United Nations: Rome, Italy, 2012.
177. Xu, Y.; Li, P.; Zou, C.; Lu, Y.; Xie, C.; Zhang, X.; Prasanna, B.M.; Olsen, M.S. Enhancing genetic gain in the era of molecular breeding. *J. Exp. Bot.* **2017**, *68*, 2641–2666. [[CrossRef](#)] [[PubMed](#)]
178. Waltz, E. Digital farming attracts cash to agtech startups. *Nat. Biotechnol.* **2017**, *35*, 397–398. [[CrossRef](#)] [[PubMed](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).