

Communication

# AgroGenome: Interactive Genomic-Based Web Server Developed Based on Data Collected for Accessions Stored in Polish Genebank

Jerzy H. Czembor <sup>1,\*</sup>, Elzbieta Czembor <sup>1</sup>, Marcin Krystek <sup>2</sup> and Juliusz Pukacki <sup>2</sup><sup>1</sup> Plant Breeding and Acclimatization Institute—National Research Institute, Radzikow, 05-870 Blonie, Poland<sup>2</sup> Poznan Supercomputing and Networking Center, ul. Jana Pawla II 10, 61-139 Poznan, Poland

\* Correspondence: j.h.czembor@ihar.edu.pl

**Abstract:** New intensive farming systems have resulted in a narrowing of the genetic diversity used in breeding programs. Breeders are looking for new sources of variation of specific traits to make genetic progress in adaptation to changing environmental conditions. Genomics-based plant germplasm research seeks to apply the techniques of genomics to germplasm characterization. Using these new methods and obtained data, plant breeders can increase the rate of genetic gains in specific breeding programs. Due to the complexity of heterogeneous sources of information, it is necessary to collect large quantities of referenced data. Molecular platforms are becoming increasingly important for the development of strategic germplasm resources for more effective molecular breeding of new cultivars. Following this trend in plant breeding, the AgroGenome portal for precise breeding programs was developed based on data collected for accessions stored in the Polish Genebank. It combines passport data of genotypes, phenotypic characteristics and interactive GWAS analysis visualization on the Manhattan plots based on GWAS results and on JBrowse interface. The AgroGenome portal can be utilized by breeders or researchers to explore diversity among investigated genomes. It is especially important to identify markers for tracking specific traits and identify QTL. The AgroGenome portal facilitates the exploitation and use of plant genetic resources stored in the Polish Genebank.



**Citation:** Czembor, J.H.; Czembor, E.; Krystek, M.; Pukacki, J. AgroGenome: Interactive Genomic-Based Web Server Developed Based on Data Collected for Accessions Stored in Polish Genebank. *Agriculture* **2023**, *13*, 193. <https://doi.org/10.3390/agriculture13010193>

Academic Editors: Vijay Gahlaut, Vandana Jaiswal and Sandeep Kumar

Received: 30 November 2022

Revised: 2 January 2023

Accepted: 3 January 2023

Published: 12 January 2023



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

**Keywords:** AgroGenome portal; biodiversity; crop design; big data; genomic selection; genomics; plant genetic resources; barley; wheat; soybean; pea

## 1. Introduction

Plant germplasm is crucial for crop genetic improvement and to achieve food security globally. Thousands of germplasm accessions have been collected and conserved ex situ and in situ all around the world, and the major challenge for plant scientists is how to exploit and utilize this crucial resource. There are about 1800 genebanks worldwide, including more than 600 in Europe [1]. About 7.4 million accessions are stored globally [2]. However, it is estimated that only 25–30% of these accessions are genetically unique [3–5]. The Polish Genebank (National Centre for Plant Genetic Resources—NCPGR) at the Plant Breeding and Acclimatization Institute—National Research Institute (PBAI-NRI) has collected more than 90,000 accessions of crops and their wild relatives. Based on the data available on the FAO WIEWS—World Information and Early Warning System on Plant Genetic Resources for Food and Agriculture (<http://www.fao.org/wiews/map-test/en/> accessed on 15 November 2022)—it can be concluded that the Polish Genebank ranks seventeenth in the world among institutions collecting plant genetic resources in terms of number of accessions. It is the third largest genebank in Europe, after Germany-IPK and Russia-VIR, and the second in the European Union [6].

In 2010, the EGISET information system was implemented at Polish Genebank as a central database on plant genetic resources in Polish Genebank. The system consists of several databases. The main function of this system is to collect and store passport data

of crops as well as their characterization and evaluation data. The information stored in databases dates back to the 1970s based on the old documentation. The passport data module includes about 91,000 accessions but only for about 50,000 of them specific data are available—totaling in about 204,000 data records. EGISET is also a repository of photos, documents and expedition data. The EGISET system has a link for studying and ordering seed samples (<https://wyszukiwarka.ihar.edu.pl/pl> accessed on 15 November 2022), which allows one to obtain seeds as well as data on the phenotype characteristics of accessions provided by curators of each species [7,8].

New intensive farming systems promoted by the green revolution have resulted in a narrowing of the genetic diversity in the collections used in breeding programs. In many countries, the high-yielding varieties are preferred and cultivated on large areas, while less-yielding cultivars, though often showing high genotypic diversity and greater yield stability, are eliminated from cultivation and often irretrievably lost from national breeding programs [1,9–11]. This increasing genetic uniformity of crops has resulted in changes in pathogen populations, causing them to be more virulent with increased severity of the diseases they cause [10,12]. Therefore, breeders are looking for new sources of variation of specific traits, especially for sources of resistance to diseases and pests. It is very important to make genetic progress in breeding of major crops in adaptation to rapidly changing environmental conditions. This is essential for the improvement of crops and for providing food security in rapidly changing environmental conditions [13].

In many cases, the intensification of production in agriculture has a negative impact on the biodiversity of the agricultural landscape. This type of agriculture needs new, more diverse cultivars suitable for the potential use in sustainable and climate-smart agriculture. Plant breeders in particular are focused on improving plant tolerance and resilience to abiotic and biotic stresses connected with climate change. Because of plant breeders' demand, there is a need for characterization of germplasms in terms of their agronomic potential and for establishing associations between molecular markers and phenotypes. This kind of information is necessary for the practical use of specific accessions in breeding programs [14–21]. Because the genetic studies were conducted on a still-limited number of accessions, the concept of creation of core subsets of crop germplasm collections was developed. The idea behind this concept is to more efficiently utilize the genetic diversity present within the larger collection present in genebank. The creation of core collections has proven to be a very successful way for plant scientists representing many disciplines (plant genetics, plant physiology, plant pathology) to help refine exploration of the larger germplasm collection in genebanks worldwide. Recently, the study of genetic diversity for both germplasm management and breeding has been described. These studies showed that in practice, even the core collection approach was not sufficient to fully characterize genetic diversity [22]. Old varieties and landraces are well-adapted to the environmental conditions in the area where they grew and are a very important source of genetic variability of important agronomic traits [23–25], including resistance to biotic stresses [26,27].

The development of crop varieties using conventional breeding methods has been effective. However, these methods are time-consuming and labor-intensive. Recently, new breeding approaches were developed, aiming to reduce breeding time and provide more precise selection and more efficient use of genetic variation. Genomics-based plant germplasm research (GPGR), or “Genoplasmics”, is a novel, cross-disciplinary research field that seeks to apply the principles and techniques of genomics to germplasm research. Using these new methods and obtained genomics data, plant breeders can substantially increase the rate of genetic gains in breeding programs [15,28–31]. Especially important for plant breeding are Next Generation Sequencing (NGS) technologies. By using them, it was possible to substantially reduce the cost of genotyping and sequencing [31,32]. The development of these methods has enabled plant breeders to use the high-throughput and cost-effective high-density genotyping in practice. The low-cost genotyping platforms were developed, and they have accelerated the use of molecular markers in the breeding programs. In recent years, a more and more important breeding goal is not only the yield,

but also to improve complex traits such as yield quality and adaptation to changing climatic conditions. To improve complex traits, modern breeding approaches, such as genomic selection (GS), are commonly used [24,32,33]. Another modern method used for breeding purposes is the Diversity Arrays Technology (DArT) in combination with next-generation sequencing platforms [34,35]. It is very useful for analyzing phenotypic and genotypic data in large and very diverse germplasm collections. Using this combination of methods can result in the low-cost identification of a relatively large number of polymorphic markers. This advantage resulted in the use of DArTseq-derived markers in more than 400 species (<http://www.diversityarrays.com/> accessed on 15 November 2022).

Genome-Wide Association Study (GWAS) was described as an effective method to identify alleles associated with traits of many crop species in [35]. It is used in many economically important crops to identify genomic regions connected with yield-related traits [20,21,36–38]. In comparison to biparental QTL mapping, the advantage of GWAS is to capture more loci responsible for the traits [39]. The use of diverse and unstructured germplasms in GWAS results in the accumulation of a larger number of recombination events. This, in combination with high-resolution markers, increases the accuracy of mapping. It is very important that, by using GWAS, candidate genes may be identified. However, for this purpose, a large association mapping population is needed to be used. There are examples that when using NGS technology and the reference genomes, GWAS identified many loci and candidate genes which were not been detected by QTL linkage mapping.

Molecular platforms are becoming increasingly important for the development of strategic germplasm resources for more effective molecular breeding of new cultivars which are well-adapted to changing climate conditions [40]. The good examples for such platforms are: the BRIDGE portal, developed for the collection of barley stored in the IPK genebank [41]; the SnpHub platform for wheat [42,43]; and platforms for soybean [44]; and SoyBase [45].

The project AGROBANK refers to the “Creation of bioinformatic management system about national genetic resources of useful plants and development of social and economic resources of Poland throughout the protection and use of them in the process of providing agricultural consulting services” (1/394826/10/NCBR/2018) and was financed by the National Center for Research and Development as part of the first round of competitive research grants under the strategic research and development program, GOSPOSTRATEG “Social And Economic Development Of Poland In The Context Of Globalizing Markets”. The main objective was to develop and implement a national management system for crop plant genetic resources at the Polish Genebank (NCPGR) and to address global trends in the creation of molecular platforms for germplasm resources.

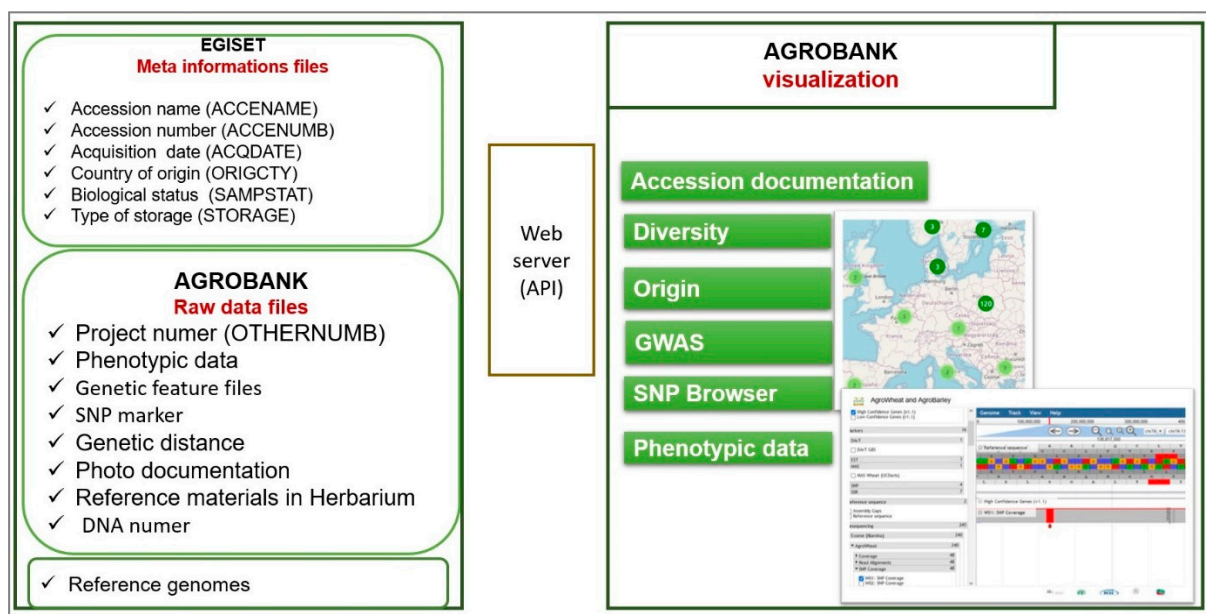
The aim of the AGROBANK project was to create using friendly interactive genomic-based web server using data collected for accessions stored in Polish Genebank. Web server was developed for the most important agronomic crops: barley, wheat, soybean and pea. It combines passport data, phenotypic characteristics, interactive GWAS analysis visualization on the Manhattan plots on the JBrowse interface.

## 2. Materials and Methods

### 2.1. AgroGenome Portal Application Architecture

The AgroGenome portal application architecture is visualized on the Figure 1.

The Polish Genebank EGISSET database is a source of information about accessions passport data. Large accession collections representing the population of the most important species were selected for phenotypic and molecular studies. They were characterized phenotypically in 2018 and 2019 and molecularly by the DArT method. Based on the GWAS analyses, SNPs associated with the described traits were identified.



**Figure 1.** Design schema of the AGROBANK AgroGenome portal. Once the files and information tables are provided as indicated in the “Prepare” step, they are preprocessed for building basic database files. Then, efficient analysis and visualization are performed interactively.

As wheat and barley occupy a particularly large area in the structure of crops, it was decided to develop reference genomes of Polish materials additionally, based on fully sequenced genomes or their large fragments, i.e., exomes, as the second stage of research. Based on the results of bioinformatic analyses of genetic diversity determined by the DArTseq method, 18 individuals were selected for the second stage of the study within barley and 48 individuals within wheat. Sequencing of full genomes of spring barley accessions using the NGS method and sequencing of wheat exomes (using the exome capture method) were carried out using the newest NovaSeq600 platform (Illumina, Inc., 5200 Illumina Way, San Diego, CA 92122, USA). The molecular data were presented by Nathan S. Watson-Haigh (South Australian Genomics Centre (SAGC), SAHMRI, Adelaide, SA 5000, Australia) in the form of the SNP Browser with all the functionalities of the DAWN portal [43].

For almost all accessions, DNA samples for the DNA genebank and reference materials in the form of seeds for the Herbarium were collected. Moreover, cereal spikes and soybean pods were included into the Herbarium collections. Photographic documentation was carried out.

## 2.2. Collecting Data to Develop AgroGenome portal

### 2.2.1. Plant Material

Accessions belonging to the 4 most important crop species were investigated (Table 1). The Polish accessions were selected to cover their diversity held at the Polish Genebank, with priority given to old varieties with key phenotypic traits in Polish breeding programs. This was then supplemented with modern varieties and non-Polish accessions from countries where a particular trait is most frequent. Barley, common wheat and other species, such as emmer and durum, soybean, pea and rape, were evaluated.

**Table 1.** Number of genotypes evaluated under AGROBANK project.

Common Name	Species	Accessions Number	
		Total	Polish Origin
Barley	<i>Hordeum vulgare</i> L.	461	146
Common wheat	<i>Triticum aestivum</i> L.	428	118
Durum and dicoccum wheat	<i>T. dicoccum</i> (Schrank) Schuebl., <i>T. durum</i> Desf.	75	11
Soybean	<i>Glycine max.</i>	196	80
Pea	<i>Pisum sativum</i> L.	184	115
Total		1344	470

### 2.2.2. Passport Data

In 2010, the EGISET information system was implemented at the Polish Genebank (National Centre of Genetic Resources) at the Plant Breeding and Acclimatization Institute—National Research Institute (PBAI—NRI; Polish name: Instytut Hodowli i Aklimatyzacji Roślin—Państwowy Instytut Badawczy, IHAR—PIB). This system is a central database on plant genetic resources in the Polish Genebank. For the AGROBANK portal, it is a source of the passport data of accessions such as accessions name (ACCENAME), accession number (ACCENUMB), acquisition date (ACQDATE), country of origin (ORIGCTY), type of storage (STORAGE) and status MLS (MLS).

### 2.2.3. Phenotypic Data

The phenotypic characterization of the genotypes included in the research was carried out in-house and in field conditions as part of experiments established on PBAI-NRI plots in 2018 and 2019. For wheat, barley and soybean, the phenotypic characterization took into account almost the entire list of the descriptors given in the recommendations of the IPGRI (International Plant Genetic Resources Institution), which associates most of European countries, and the assessment methodology is respected all over the world. For wheat and barley descriptors (traits), which are crucial for final yield, disease resistance and agricultural traits were described. A total of sixteen agricultural traits were described: days to plants' heading stage (DH), days to plants' milky-waxy stage (DMW), days to plants' maturity stage (DM), days to plants' harvest stage (DPH), plants' height (PH), plants' lodging tendency (LT), row number (RN) for barley, spike density (SD), glume color (GGC1), grain awn type (GAT), spike length (SL), grain per spike (NGS), grain type/covering (GT) only for barley, grain color (pericarp) (GC) and 1000-grain weight (TGW).

For selected traits important in terms of agricultural value, some modifications were made based on consultations with experts for evaluated species. It was mostly conducted for disease resistance for most important pathogens.

- Descriptors used for barley: (<https://croptgenebank.sgrp.cgiar.org/index.php/learning-space-mainmenu-454/manuals-and-handbooks-mainmenu-533/descriptors-mainmenu-547> (accessed on 29 October 2021));
- Descriptors used for wheat: [https://old.vurv.cz/Ewdb/asp/IPGRI\\_descr\\_1985.pdf](https://old.vurv.cz/Ewdb/asp/IPGRI_descr_1985.pdf) (accessed on 29 October 2021);
- Descriptors used for soybean: [https://www.biodiversityinternational.org/fileadmin/\\_migrated/uploads/tx\\_news/Descriptors\\_for\\_soyabean\\_252.pdf](https://www.biodiversityinternational.org/fileadmin/_migrated/uploads/tx_news/Descriptors_for_soyabean_252.pdf) (accessed on 29 October 2021).

GWAS results for barley and soybean were already published [20,21,41].

### 2.2.4. Molecular Data Using DArTseq and GWAS Analysis

#### A. DNA extraction and quantification

One g of young leaf tissue from the 3rd to 4th node of each seedling was excised, frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ . Genomic DNA was extracted from frozen



leaves using a modified cetyltrimethylammonium bromide (CTA B)/chloroform/isoamyl alcohol method [46]. DNA quantification was performed by agarose gel electrophoresis (0.8%), and it was adjusted to 50 ng/ $\mu$ L for genotyping using DArTseq.

## B. Genotyping using DArTseq and GWAS analysis

### B.1. Data Filtering Process

Genotypes were genotyped by Diversity Arrays Technology Pty Ltd, Building 3, Level D, University of Canberra, Monana Street, Bruce, ACT, 2617, Australia, using DArTseq [30]. SNP calls were made against: *Hordeum vulgare* Morex v2, *T. aestivum* Chinese Spring (CS) IWGSC RefSeq v1.0 (<https://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies> - accessed on 8 September 2021) and soybean available in phytozome (<https://phytozome-next.jgi.doe.gov/> - accessed on 10 October 2021) [47]. DArT data were handled in the same manner for all crops. That is, we used the DArTR v1.1.11 package [47] in the R programming language. SNPs and genotypes were removed if SNP markers contained > 5% missing data and genotypes contained > 10% missing data. SNPs with a reproducibility score of (RepAvg) < 100% were removed. Where SNPs originated from the same fragment, a random SNP was retained while the others were discarded. Noninformative monomorphic SNPs were removed, as were rare SNPs with a minor allele frequency of <1%.

### B.2. Genome-wide association studies (GWAS)

GWAS analysis was conducted using the GAPIT v2018.08.18 R package [48,49]. We used the recently developed Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) model, which has been shown to produce fewer false positives, identify more true positives and scale to very large data sets [50]. Physical genome positions of markers were derived from the DArTseq SNP genotype file. Since GAPIT can only handle complete data, only markers with a physical position on one of the chromosomes and zero missing data were used as input to the GWAS analysis. Bonferroni and FDR thresholds were used. DArTseq markers with FDR and Bonferroni  $p = 0.01$  thresholds were taken as significantly associated with the evaluated trait. In order to show the distribution of SNPs over the chromosome, Manhattan plots were also generated. The significance levels for GWAS analysis on the Manhattan plots were as follows: solid line represented the Bonferroni FDR multiple test threshold ( $p = 0.01$ ), and dashed green line represented the FDR threshold (FDR adjusted  $\leq 0.05$ ). In order to show the distribution of SNPs over the chromosome, Manhattan plots were also generated.

## 2.3. SNP Browser

### 2.3.1. Barley SNP Browser

Based on the results of bioinformatic analyses of genetic diversity determined by the DArTseq method, 18 individuals were selected for the second stage of the study within barley. Sequencing of full genomes of spring barley genotypes using the NGS method and sequencing were carried out using the newest NovaSeq600 platform (Illumina), generating  $2 \times 150$  PE reads. DNA extraction and quantification was as for the DArTseq method.

#### A. Barley Genotype Selection for WGS

Distance matrices, provided by Diversity Arrays, were partitioned (clustered) into  $k$  clusters around medoids using the pam () method available in the cluster package [51,52] and performed in the R statistical programming language [53]. The pam-algorithm searchers for  $k$  representative genotypes/clusters were used such that the sum of dissimilarities between genotypes in a cluster and its representative genotype was minimized. Therefore, the number of clusters,  $k$ , was set to 18, the number of genotypes to be selected for whole-genome sequencing (WGS). Some selections were made to ensure a preference for Polish genotypes if one was close to the medoid genotype.

#### B. Sequencing and read processing

Sequencing of full genomes of 18 spring barley accessions using the NGS method was carried out using the newest NovaSeq600 platform (Illumina), generating  $2 \times 150$  PE reads. Raw reads were preprocessed through trimmomatic v0.39 (<http://www.usadellab.org/cms/?page=trimmomatica>, accessed on 10 October 2021) to remove adapters, low-quality bases and short reads. Specifically, the following command line argument was used: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:3:true LEADING:2 TRAILING:2 SLIDINGWINDOW:4:15 MINLEN:36. Reads were processed using the approach described in Watson-Haigh et al. [43]. Briefly, QC reads were aligned to the Barley Morex v2 genome assembly [54] using Minimap v2.17 [55], and variants were called using a SAMtools v1.9 [56] and BCFtool v1.9 calling pipeline, which required a minimum mapping quality of 20 and minimum base call quality of 30. Processing was parallelized per chromosome to facilitate timely completion of the analysis. Read alignment coverage and variant density (variants per 10 kbp) files were generated in BigWig format. All data have been made available as visualization tracks within a JBrowse [57] instance ([http://62.3.171.115/jbrowse/?data=data%2Fbarley\\_morex\\_v2](http://62.3.171.115/jbrowse/?data=data%2Fbarley_morex_v2), accessed on 10 October 2021).

### 2.3.2. Wheat SNP Browser

Based on the results of bioinformatic analyses of genetic diversity determined by the DArTseq method, 48 individuals were selected for the second stage of the study within wheat. Sequencing of wheat exomes (using the exome capture method) were carried out using the newest NovaSeq600 platform (Illumina). DNA extraction and quantification was the same as in the DArTseq method.

#### A. Genotype Selection for WES

Distance matrices, provided by Diversity Arrays, were partitioned (clustered) into  $k$  clusters around medoids using the pam () method available in the cluster package [51,52] and performed in the R statistical programming language [55]. The pam-algorithm searchers for  $k$  representative genotypes/clusters were used such that the sum of dissimilarities between genotypes in a cluster and its representative genotype was minimized. Therefore, the number of clusters,  $k$ , was set to 48, the number of genotypes to be selected for whole-exome sequencing (WES). Some selections were made to ensure a preference for Polish genotypes if one was close to the medoid genotype.

#### B. Sequencing and read processing

The selected genotypes based on the DArTseq data accessions were sequenced on an Illumina NovaSeq, generating  $2 \times 150$  PE reads. Raw reads were preprocessed through trimmomatic v0.39 to remove adapters, low-quality bases and short reads. Specifically, the following command line argument was used: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:3:true LEADING:2 TRAILING:2 SLIDINGWINDOW:4:15 MINLEN:36. Reads were processed using the approach described in Watson-Haigh et al. [43]. Briefly, QC reads were aligned to the IWGSC RefSeq v1.0 genome assembly [58] using Minimap v2.17 [55], and variants were called using a SAMtools v1.9 [58] and BCFtool v1.9 calling pipeline, which required a minimum mapping quality of 20 and minimum base call quality of 30. Processing was parallelized per chromosome to facilitate timely completion of the analysis. Read alignment coverage and variant density (variants per 10 kbp) files were generated in BigWig format. All data have been made available as visualization tracks within a JBrowse [57] instance ([http://62.3.171.115/jbrowse/?data=data%2Fwheat\\_CS\\_v1.0](http://62.3.171.115/jbrowse/?data=data%2Fwheat_CS_v1.0), accessed on 10 October 2021).

### 2.4. Collecting DNA Samples for Genebank

DNA samples were collected for all genotypes tested in the project. They were the basis for establishing the DNA bank and developing the data flow on the platform. A list of descriptors for the DNA samples was developed.

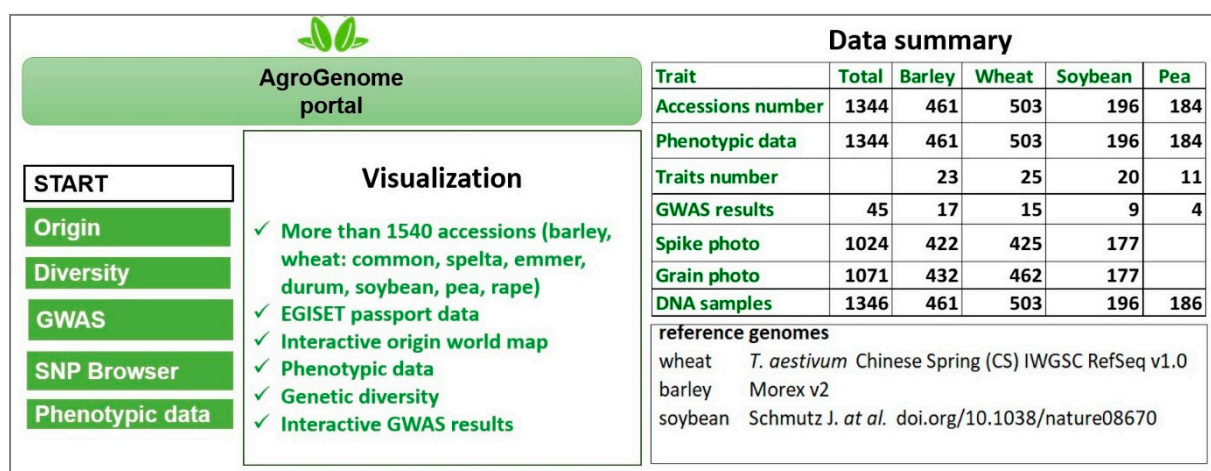
### 2.5. Collecting Reference Materials for Herbarium and Photo Documentation

Ear or pod samples were collected from the field trials and included with the seeds' samples in the Herbarium collection as reference material. Photographic documentation was kept and described in accordance with generally accepted rules in genebanks.

## 3. Results

### 3.1. AgroGenome Portal Summary Presentations

On the start page of the AgroGenome portal, the user can find information about what kind of data were collected to develop the interactive AgroGenome portal. This information is: species, number of accessions for which passport data are available, number of traits for which MTAs are collected based on GWAS analysis of the DArTseq data and phenotypic data and visualized on the Manhattan plots, number of the whole genomes for barley and common wheat number associated with reference of whole genomes as JBrowe interface, and number of photos and reference samples in the Herbarium (Figure 2). A password-protected system for data entry will be included in the next stage of portal development. Data collected during the project after publishing have the status "Open Access: Creative Commons (CC-BY-SA)", and before publishing, they will be available under the prepublication data sharing principle of the Toronto Agreement (<https://www.nature.com/articles/461168a>, accessed on 10 October 2021).




**Figure 2.** Visualization of the start page of the AgroGenome portal created based on phenotypic and genotypic data collected for 1344 accessions of barley, common wheat, durum wheat, emmer wheat, spelt wheat, polonicum wheat, soybean and pea with special emphasis to cover diversity of old cultivars collected in Poland and other European countries [47].

The application was implemented based on the requirements and design proposed by the authors. It is based on the client server architecture. The client's web interface was built with the use of ReactJS framework and is accessible with any web browser application. The server side is based on the Elastic Search service, which provides data storage capabilities as well as convenient API for searching and retrieving data.

### 3.2. AgroGenome Passport Data Presentation

To obtain information about passport data as general filters, users will use species (SPECIES), accession number (ACCENUMB), EGISSET, country of origin (ORIGCTY), acquisition date (ACQDATE) or type of storage (STORAGE). Based on this, users have presentations of accessions and may observe morphology visualization as a photography documentation with additional information if the phenotypic data are available, as status MLS, and if reference materials are in the Herbarium (Figure 3).





**AgroGenome  
portal**

**GENERAL FILTERS**

**Species (SPECIES)**

Accession number (ACCENUMB): EGISET

Accession date (ACQDATE)



Accession name (ACCENAME)

Other number (OTHERNUMB)

Type of storage (STORAGE)

Country of origin (ORIGCTY)

Status MLS (MLS)

ACCENAME			Phenotypic data
▪ OPOLANKA			▪ yes
OTHERNUMB	ACQDATE	STORAGE	Photo documentation
▪ POL003: 59 P	▪ 1971	▪ 13; 50	▪ yes
ACCENUMB	ORIGCTY	MLS	Herbarium
▪ POL003: 20541	▪ POL	▪ 1	▪ yes

**Figure 3.** Visualization of passport data evaluated under AGROBANK project accessions, such as accession number (ACCENUMB), country of origin (ORIGCTY), acquisition date (ACQDATE), type of storage (STORAGE) status MLS, photographic documentation, and if accession is in the Herbarium as a reference sample. Data ACCENUMB, ORIGCTY, ACQDATE, STORAGE and status MLS are collected from the EGISET Polish Genebank data base. Photographic documentation and sample DNA were collected under the AGROBANK project.

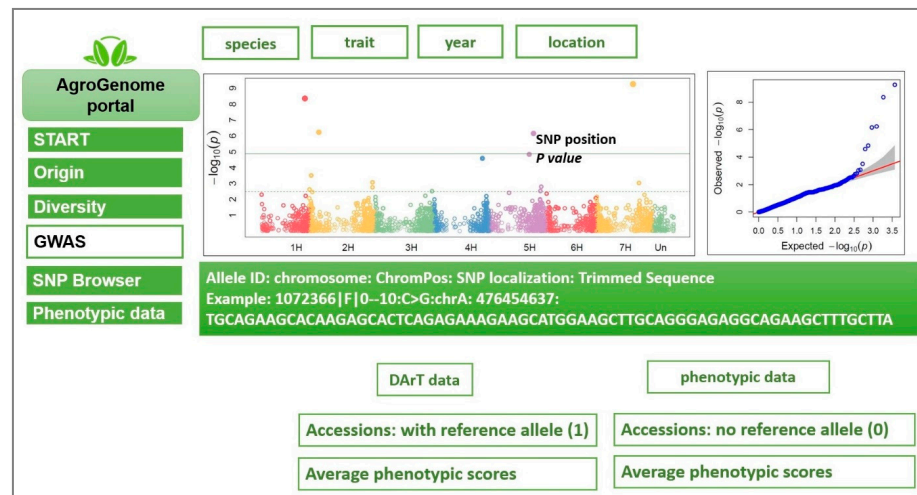
In addition, the collection of DNA samples was used for the development of a data flow diagram related to them and connected with the Polish Genebank EGISET database. The compiled descriptor list was adapted from that used by the United Kingdom National DNA Database. The portal is open and can be developed for new data of tested accessions.

### 3.3. AgroGenome GWAS Results Presentation

Interactive Manhattan Plots of Genome-Wide Association Studies (GWAS) were developed with Quantile–Quantile (QQ) plots observed against expected probability values ( $p$ -values) from the GWAS analysis. The significance line is presented. As a filter, the user may use species and trait. On the Manhattan interactive plot, it is possible to find information about SNP ID and  $p$ -value (Figure 4).

By clicking on a data point of a variant  $p$ -value, the application automatically jumps to that variant position in the embedded SNP Browser developed for common wheat and barley. The trait to be displayed can be switched by a select box. The portal gives users the opportunity to observe GWAS results with QQ plots and to visualize indicated marker-trait associations (MTAs) with traits used as general filters for selected species. The BRIDGE portal, which was developed for barley, does not have these functions. It is visualized on the Manhattan figure SNP's ID position and  $p$ -value and, next, on the browse server [43].

On the AgroGenome portal under Manhattan plots, the user can find information about allele ID, chromosome number, chromosome position, SNP localization and trimmed sequence. It is important that evaluated accessions will be divided for two groups—one with a reference allele for the evaluated trait and one without it. The user may compare the average value phenotypic data described for the first and second group and conclude which group is more valuable for his project or breeding program according to the evaluated trait. Moreover, users have access to phenotypic data of each genotype separately.



**Figure 4.** Visualization of the GWAS analysis results on the AgroGenome portal developed for barley, common wheat, durum wheat, emmer wheat, soybean and pea. Results are presented in the form of interactive Manhattan plots, QQ plots, SNP ID, *p*-value, SNP localization, chromosome number, trimmed sequence and DArT data. Accessions are divided for two groups: with reference allele and without it.

### 3.4. AgroGenome SNP Browser Presentation

SNP Browser integrates data for wheat from the *T. aestivum* Chinese Spring (CS) IWGSC RefSeq v1.0 genome with exome data from 48 accessions and for barley from Morex v2 genome with Whole Genome Shotgun (WGS) from 18 accessions (Figure 5). It was developed in cooperation with the main author of DAWN portal, Nathan S. Watson-Haigh, South Australian Genomics Centre (SAGC), SAHMRI, Adelaide, SA 5000, Australia. The broad structure of the data flow and the features of AgroGenome web server was created based on structure developed for DWAN interface and it is described in details [43].



**Figure 5.** Visualization SNP Browser. The tracks from top to bottom are: high-confidence gene, read coverage depth patterns (coverage), individual read alignments and alignment mismatches (read alignment),

mismatches between the read alignments and the CS reference (SNP coverage), variant call density and variant calls. Read coverage depth patterns show the mean coverage (yellow line) as well as 1 and 2 standard deviations (grey background shading). Vertical lines within the read coverage plot indicate the proportion of reads with mismatches to the CS reference, and teardrops shown below the coverage track indicate those positions exceeding 90% alternative bases and at  $\geq 3$  reads coverage. Colored drops hanging off the read coverage profile indicate the presence of putative SNPs/indels to the reference, A-green, T-red, C-blue, G-yellow, indel-grey. (Interface JBrowse developed in cooperation with Nathan S. Watson-Haigh, Genome Informatics and Bioinformatics Training, Flagstaff Hill, SA 5159, Australia, based on the DAWN structure [43]).

For each wheat and barley accession the user can access: (1) “Coverage” tracks for simple visualizing read coverage depth patterns at Kbp to Mbp scales. These show the mean coverage (yellow line) as well as 1 and 2 standard deviations (grey background shading). Regions with read coverage  $> 2 * SD$  from the mean were extracted, then merged if  $\leq 500$  bp apart and reported if  $\geq 5$  kbp (above the mean) or  $\geq 50$  kbp (below the mean) in length. (2) “Read Alignment” tracks for visualizing individual read alignments (when viewing a sufficiently small region) or read coverage depth (when viewing a larger region) and alignment mismatches at the 100’s of bp scale. (3) “SNP Coverage” tracks display of read coverage depth together with the proportion of reads containing mismatches vs. the reference sequence. (4) “Variant Call Density” track display the density of variants in 10 kbp nonoverlapping windows. (5) “Variant Calls” tracks display the variant calls themselves. Vertical lines within the read coverage plot indicate the proportion of reads with mismatches to the CS reference and teardrops shown below the coverage track indicate those positions exceeding 90% alternative bases and at  $\geq 3$  reads coverage. This track is particularly useful for identifying haplotype blocks at the Kbp scale. Most tracks transition to a read coverage depth or variant density plots at the Kbp-Mbp scale when the density of information is too high to be visually meaningful (Figure 5) [43].

Read Alignment tracks for visualizing individual read alignments or read coverage depth and alignment mismatches at the 100’s of bp scale give opportunity: view details, zoom this match, highlight this match, quick-view mate/next location and open mate/next location.

Moreover, single nucleotide polymorphisms (SNPs) detected using the DArTseq method associated with yield traits and phenological or morphological features in barley, wheat, soybean and pea are visualized on the JBrowse interface based on the GWAS results such as chromosome number, chromosome position and SNP localization.

The SNP Browser gives the opportunity to estimate the diversity of accessions at the population level from SNP data and provides information on the amount of genetic diversity between accessions at the whole-genome level but not on its distribution within the genome. However, genetic diversity goes beyond SNPs and includes indels, introgressions and other structural variations such as copy number variation (CNV). These are all known to be important drivers of diversity. The ability to access and visualize genetic diversity in detail, from whole chromosomes to individual genes, will enable a better understanding and utilization of the available diversity in a region of interest, irrespective of scale.

#### 4. Discussion

Currently, the role of the genebank should not only be the storage of germplasm. Until now, in many genebanks, as in the Polish Genebank, information on the phenotypic characteristics of genotypes is often available, but they are in many cases incomplete and outdated due to changing climatic conditions. There is an urgent need for genebank basic works concerning the preservation of plant genetic resources to be carried out with simultaneous, properly conducted phenotypic characterization, considering the applicable descriptors, which at the same time correspond to economically important morphological and physiological traits. Reference exponents in the Herbarium are an additional safeguard for the loss of variability represented by the genotype. With the development of molecular techniques, genebanks should strive to make molecular data available for stored collections.

Genetic diversity analysis, linkage mapping, and association mapping, especially at the whole-genome level, are crucial for modern molecular breeding [15,17–21].

The agriculture sector is of great economic and strategic importance for the European Union. Breeders play a very important role in the practical use of plant genetic resources to meet food security goals in the time of climate change. Due to the complexity of breeding programs, breeders must manage many different and heterogeneous sources of information. Advanced analytics and visualization technologies in the form of portals help bring them benefits of all available data-based solutions, mostly for the results of whole genomics-based plant germplasm research (GPGR) or whole-exome capture (WEC). Whole-genome sequencing is used for mid-sized genomes, such as soybean, to accomplish genotyping-by-sequencing (GBS) [59]. The whole-exome capture (WEC) technologies are used for large-genome species such as wheat [60]. Recent studies showed wheat profile genomic variations on a scale of hundreds or thousands of accessions through WEC [61] or whole-genome resequencing (WGS) [62]. There are many examples of “Genoplasmics” studies which seek to apply the principles and techniques of genomics to the characterization of germplasm available in genebanks. Decreasing sequencing costs are resulting in increased numbers of samples and species which can be sequenced. Using these new methods and vast genomics data, plant breeders can substantially increase the rate of genetic gains in breeding programs and will help to meet food security in many areas of the globe [3,15,21,29].

To address this issue, the AgroGenome portal was created to support genebanks and breeding research of the most important agricultural crops such as barley, wheat, soybean and pea. The accessions number in each collection was representative to evaluate genetic diversity, with a minimum of 180 accessions for pea and soybean and more than 450 for wheat and barley. It is important for portal users such as breeders or scientists that there are well-provided passport data from the Polish Genebank EGISET database and AGROBANK project results, such as photography documentation; proper phenotyping characterization, taking into account almost the entire list of the descriptors given in the recommendations of the IPGRI (International Plant Genetic Resources Institution); and interactive visualization of the DArTseq and GWAS results. Based on the results of bioinformatic analyses of genetic diversity determined by the DArTseq method, 48 individuals for wheat and 18 individuals for barley were selected for the second stage to develop the SNP Browser. For wheat sequencing, exomes (using the Exome Capture method), and for barley sequencing, whole genomes were carried out using the newest NovaSeq600 platform (Illumina).

The AgroGenome structure is universal and may be used for many crops. Currently, it was developed for barley, common wheat, durum wheat, emmer wheat, spelt wheat, polonicum wheat, soybean and pea. Moreover, it is as the first portal in which visualized MTAs for yield and associated traits are presented. As described by Wang et al. [42], the SnpHub, Gigwa v2, CanvasDB portals are universal and can be used for many species. However, they do not have the option to present MTAs for evaluated traits such as GWAS results, which are cheap; they are not as efficient as the DArTseq method and phenotypic data; and they cannot be presented on the Jbrowse interface. The AgroGenome portal structure is open and can be developed for new data of evaluated accessions. On the AgroGenome portal under Manhattan plots, the user can find information about allele ID, chromosome number, chromosome position, SNP localization and trimmed sequence.

The AgroGenome portal is end-user friendly and relatively easy to learn. The user may compare the average value phenotypic data described for the first and second group and conclude which group is more valuable for his project or breeding program. Other portals do not have such functionality. Moreover, user have access to phenotypic data of each genotype separately.

## 5. Future Prospects

The functionality of the AgroGenome portal corresponds to most species of crops. Depend on the financial resources' availability, in the future, the portal will be developed also for species and accessions currently not present. Phenotypic characteristics are very

important for germplasm users such as scientists and breeders, and providing such data is a prime responsibility of genebanks. This kind of work in genebanks is the most time- and cost-consuming. However, because of climate change, for accessions evaluated under the AGROBANK project, it requires periodical data updating for many environments, especially for GWAS analysis. At this stage of the project, the so-called core collections representing the total genetic variation of the collection were developed. The authors plan to create core-collection for traits, which is very important for breeders. The next stage of the portal expansion will be the development of trait-specific core collections. In the AgroGenome portal, more accessions will be included to the available collections of the evaluated species, and new economically important crops will be included. Based on the demand of breeders, a larger spectrum of the traits, such as quality characteristics, will be described. An example of this kind of data are markers for physiological traits, such as photosynthesis parameters measured using a PSI II camera (Fo, Fm, Fv, Ft\_Lss, QY\_max, QY\_Lss, NPQ\_Lss) and chlorophyll content, which are already available for visualization on the SNP Browser.

The portal is open to the public, and data for genetic resources generated by different teams from Polish Genebank or other national and international institutions can be included and visualized. Currently, genebanks are establishing more and more close cooperation regarding the exchange of information regarding their collections, removing duplicates and creating base collections. Currently, genebanks are not only “living museums” concentrating mainly on storage of alive plant samples, but more and more are working as specific data repositories on available genetic resources. The Polish Genebank is one of the largest in Europe and in the world; therefore, the use of more and more molecular methods for the characterization of collections will ensure significant progress in field germplasm management and characteristics on more than a national scale.

Creating the AgroGenome platform is a first step for building a global genebanks partnership in the next years and strengthening a rational, efficient and effective global system of ex situ crop diversity. In project AGROBANK and on the AgroGenome portal, plant material, mostly landraces and old cultivars, were included. The number of subcollections from different regions of Europe was sufficient to make them a representative group for the region. This corresponds to the goal of the European Green Deal Programme to increase agrobiodiversity in the EU agricultural system. It helps to develop a new partnership for broader international collaboration and support the essential genebank operations and increase system performance, efficiency and effectiveness.

**Author Contributions:** All authors (J.H.C., E.C., M.K. and J.P.) were conceptually involved in the creation and design of the AgroGenome portal. E.C. and J.H.C.: data preparation—phenotyped plant material and conducted the statistical analysis of phenotypic data; collected the molecular data and bioinformatic analysis data; J.H.C. and E.C.: schema data flow creation; M.K. and J.P.: implementation—conducting programming activities which included supervising application implementation and processing raw data to prepare data for presentation; M.K., J.P. and E.C.: portal visualization; J.H.C.: project coordinator and resources; writing—original draft preparation. All authors reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This portal is a part of the AGROBANK project “Creation of bioinformatic management system about national genetic resources of useful plants and development of social and economic resources of Poland throughout the protection and use of them in the process of providing agricultural consulting services” (1/394826/10/NCBR/2018) financed by the National Center for Research and Development (NCBR, Poland) as part of the 1st round of competitive research grants under the strategic research and development program GOSPOSTRATEG “Social And Economic Development Of Poland In The Context Of Globalizing Markets”.

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

**Informed Consent Statement:** Not applicable.



**Data Availability Statement:** All data generated or analyzed during this study are included in this published article and its supplementary information files. The contact person is Jerzy H. Czembor, Plant Breeding and Acclimatization Institute—National Research Institute, Radzikow, 05-870 Blonie, Poland.

**Acknowledgments:** We thank the Polish Genebank (National Centre for Plant Genetic Resources—KCRZG at the Plant Breeding and Acclimatization Institute—National Research Institute, Radzikow, Poland) for providing seed samples; Radoslaw Suchecki (CSIRO Agriculture and Food, Urrbrae, SA 5064, Australia) for general help in creating the JBrowse interface; Nathan S. Watson-Haigh (South Australian Genomics Centre (SAGC), SAHMRI, Adelaide, SA 5000, Australia) for bioinformatics analysis and creating the JBrowse interface based on the DAWN structure. We would like especially to thanks for Urszula Piechota, Plant Breeding and Acclimatization Institute—National Research Institute, for her cooperation in the development of the data flow for the GWAS data and SNP Browser.

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

## References

- Weise, S.; Lohwasser, U.; Oppermann, M. Document or Lose It—On the Importance of Information Management for Genetic Resources Conservation in Genebanks. *Plants* **2020**, *9*, 1050. [[CrossRef](#)] [[PubMed](#)]
- Commission on Genetic Resources for Food and Agriculture Food and Agriculture Organization of the United Nations. *The Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture*; FAO: Rome, Italy, 2007; ISBN 978-92-5-108820-3.
- Jia, J.; Li, H.; Zhang, X.; Li, Z.; Qiu, L. Genomics-based plant germplasm research (GPGR). *Crop. J.* **2017**, *5*, 166–174. [[CrossRef](#)]
- Diez, M.J.; De La Rosa, L.; Martín, I.; Guasch, L.; Cartea, M.E.; Mallor, C.; Casals, J.; Simó, J.; Rivera, A.; Anastasio, G.; et al. Plant Genebanks: Present Situation and Proposals for Their Improvement. the Case of the Spanish Network. *Front. Plant Sci.* **2018**, *9*, 1794. [[CrossRef](#)] [[PubMed](#)]
- Milner, S.G.; Jost, M.; Taketa, S.; Mazón, E.R.; Himmelbach, A.; Oppermann, M.; Weise, S.; Knüpfper, H.; Basterrechea, M.; König, P.; et al. Genebank genomics highlights the diversity of a global barley collection. *Nat. Genet.* **2018**, *51*, 319–326. [[CrossRef](#)] [[PubMed](#)]
- Czembor, J.H.; Gryziak, G.; Zaczyński, M.; Włodarczyk, S.; Podyma, W. Gromadzenie i zachowanie zasobów genowych roślin użytkowych w Polsce 2015–2017. *Biul. IHAR* **2018**, 15–16.
- Czembor, J.H.; Gryziak, G.; Zaczyński, M.; Puchta, M.; Czembor, E. Gromadzenie i zachowanie zasobów genowych roślin użytkowych w Polsce—Artykuł przeglądowy Część 1. Gromadzenie zasobów genowych roślin użytkowych w trakcie ekspedycji krajowych i zagranicznych. *Agron. Sci.* **2018**, *72*, 135–146. [[CrossRef](#)]
- Czembor, J.H.; Gryziak, G.; Zaczyński, M.; Puchta, M.; Czembor, E. Gromadzenie i zachowanie zasobów genowych roślin użytkowych w Polsce—Artykuł przeglądowy Część 2. Przechowywanie zasobów genowych w formie nasion, prowadzenie herbarium, baz danych i udostępnianie zasobów genowych. *Agron. Sci.* **2018**, *72*, 147–154. [[CrossRef](#)]
- Purugganan, M.D.; Fuller, D.Q. The nature of selection during plant domestication. *Nature* **2009**, *457*, 843–848. [[CrossRef](#)]
- Ingvordsen, C.H. Climate Change Effects on Plant Ecosystems—Genetic Resources for Future Barley Breeding. Ph.D. Thesis, Technical University of Denmark (DTU), Lyngby, Denmark, April 2014.
- Nguyen, G.N.; Norton, S.L. Genebank Phenomics: A Strategic Approach to Enhance Value and Utilization of Crop Germplasm. *Plants* **2020**, *9*, 817. [[CrossRef](#)]
- Marone, D.; Russo, M.; Mores, A.; Ficco, D.; Laidò, G.; Mastrangelo, A.; Borrelli, G. Importance of Landraces in Cereal Breeding for Stress Tolerance. *Plants* **2021**, *10*, 1267. [[CrossRef](#)]
- Ansaldi, B.H.; Franks, S.J.; Weber, J.J. The influence of environmental factors on breeding system allocation at large spatial scales. *AoB PLANTS* **2018**, *10*, ply069. [[CrossRef](#)]
- Singh, D.; Ziem, L.A.; Dracatos, P.M.; Pourkheirandish, M.; Tshewang, S.; Czembor, P.; German, S.; Fowler, R.A.; Snyman, L.; Platz, G.J.; et al. Genome-wide association studies provide insights on genetic architecture of resistance to leaf rust in a worldwide barley collection. *Mol. Breed.* **2018**, *38*, 43. [[CrossRef](#)]
- Hickey, L.T.; Hafeez, A.N.; Robinson, H.; Jackson, S.A.; Leal-Bertioli, S.C.M.; Tester, M.; Gao, C.; Godwin, I.D.; Hayes, B.J.; Wulff, B.B.H. Breeding crops to feed 10 billion. *Nat. Biotechnol.* **2019**, *37*, 744–754. [[CrossRef](#)] [[PubMed](#)]
- Piechota, U.; Czembor, P.C.; Słowacki, P.; Czembor, J.H. Identifying a novel powdery mildew resistance gene in a barley landrace from Morocco. *J. Appl. Genet.* **2019**, *60*, 243–254. [[CrossRef](#)] [[PubMed](#)]
- Rebetzke, G.; Jimenez-Berni, J.A.; Fischer, R.; Deery, D.; Smith, D. Review: High-throughput phenotyping to enhance the use of crop genetic resources. *Plant Sci.* **2019**, *282*, 40–48. [[CrossRef](#)] [[PubMed](#)]
- Volk, G.M.; Byrne, P.F.; Coyne, C.J.; Flint-Garcia, S.; Reeves, P.A.; Richards, C. Integrating Genomic and Phenomic Approaches to Support Plant Genetic Resources Conservation and Use. *Plants* **2021**, *10*, 2260. [[CrossRef](#)]
- Czembor, E.; Czembor, J.H.; Suchecki, R.; Watson-Haigh, N.S. DArT-based evaluation of soybean germplasm from Polish Gene Bank. *BMC Res. Notes* **2021**, *14*, 343. [[CrossRef](#)]

20. Czembor, J.H.; Czembor, E.; Suchecki, R.; Watson-Haigh, N.S. Genome-Wide Association Study for Powdery Mildew and Rusts Adult Plant Resistance in European Spring Barley from Polish Gene Bank. *Agronomy* **2021**, *12*, 7. [[CrossRef](#)]
21. Czembor, J.H.; Czembor, E. Genome-Wide Association Study of Agronomic Traits in European Spring Barley from Polish Gene Bank. *Agronomy* **2022**, *12*, 2135. [[CrossRef](#)]
22. Smykal, P.; Aubert, G.; Burstin, J.; Coyne, C.J.; Ellis, N.T.H.; Flavell, A.J.; Ford, R.; Hýbl, M.; Macas, J.; Neumann, P.; et al. Pea (*Pisum sativum* L.) in the Genomic Era. *Agronomy* **2012**, *2*, 74–115. [[CrossRef](#)]
23. Gilliham, M.; Able, J.A.; Roy, S.J. Translating knowledge about abiotic stress tolerance to breeding programmes. *Plant J.* **2017**, *90*, 898–917. [[CrossRef](#)] [[PubMed](#)]
24. Bailey-Serres, J.; Parker, J.E.; Ainsworth, E.A.; Oldroyd, G.E.D.; Schroeder, J.I. Genetic strategies for improving crop yields. *Nature* **2019**, *575*, 109–118. [[CrossRef](#)] [[PubMed](#)]
25. Raggi, L.; Caproni, L.; Negri, V. Landrace added value and accessibility in Europe: What a collection of case studies tells us. *Biodivers. Conserv.* **2021**, *30*, 1031–1048. [[CrossRef](#)]
26. Cobb, J.N.; Biswas, P.S.; Platten, J.D. Back to the future: Revisiting MAS as a tool for modern plant breeding. *Theor. Appl. Genet.* **2018**, *132*, 647–667. [[CrossRef](#)] [[PubMed](#)]
27. Kumar, A.; Verma, R.P.S.; Singh, A.; Sharma, H.K.; Devi, G. Barley landraces: Ecological heritage for edaphic stress adaptations and sustainable production. *Environ. Sustain. Indic.* **2020**, *6*, 100035. [[CrossRef](#)]
28. Varshney, R.K.; Singh, V.K.; Hickey, J.M.; Xun, X.; Marshall, D.F.; Wang, J.; Edwards, D.; Ribau, J.-M. Analytical and Decision Support Tools for Genomics-Assisted Breeding. *Trends Plant Sci.* **2016**, *4*, 354–363. [[CrossRef](#)]
29. van Bemmelen van der Plaat, A.; van Treuren, R.; van Hintum, T.J.L. Reliable genomic strategies for species classification of plant genetic resources. *BMC Bioinform.* **2021**, *22*, 173. [[CrossRef](#)]
30. 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](#)]
31. Xu, Y.; Zhang, X.; Li, H.; Zheng, H.; Zhang, J.; Olsen, M.S.; Varshney, R.K.; Prasanna, B.M.; Qian, Q. Smart breeding driven by big data, artificial intelligence, and integrated genomic-enviromic prediction. *Mol. Plant* **2022**, *15*, 1664–1695. [[CrossRef](#)]
32. Schaid, D.J.; Chen, W.; Larson, N.B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* **2018**, *19*, 491–504. [[CrossRef](#)]
33. Varshney, R.K.; Roorkiwal, M.; Sorrells, M.E.; Molecular, N.; Strategies, B. *Genomic Selection for Crop Improvement*; Springer: Cham, Switzerland, 2017; ISBN 9783319631684.
34. Wenzl, P.; Raman, H.; Wang, J.; Zhou, M.; Huttner, E.; Kilian, A. A DArT platform for quantitative bulked segregant analysis. *BMC Genom.* **2007**, *8*, 196. [[CrossRef](#)] [[PubMed](#)]
35. Uffelman, E.; Huang, Q.Q.; Munung, N.S.; de Vries, J.; Okada, Y.; Martin, A.R.; Martin, H.C.; Lappalainen, T.; Posthuma, D. Genome-wide association studies. *Nat. Rev. Methods Prim.* **2021**, *1*, 59. [[CrossRef](#)]
36. Desgroux, A.; L'Anthoëne, V.; Roux-Duparque, M.; Rivière, J.-P.; Aubert, G.; Tayeh, N.; Moussart, A.; Mangin, P.; Vetel, P.; Piriou, C.; et al. Genome-wide association mapping of partial resistance to *Aphanomyces euteiches* in pea. *BMC Genom.* **2016**, *17*, 124. [[CrossRef](#)] [[PubMed](#)]
37. Gali, K.K.; Sackville, A.; Tafesse, E.G.; Lachagari, V.R.; McPhee, K.; Hybl, M.; Mikić, A.; Smykal, P.; McGee, R.; Burstin, J.; et al. Genome-Wide Association Mapping for Agronomic and Seed Quality Traits of Field Pea (*Pisum sativum* L.). *Front. Plant Sci.* **2019**, *10*, 1538. [[CrossRef](#)]
38. Tsai, H.-Y.; Janss, L.L.; Andersen, J.R.; Orabi, J.; Jensen, J.D.; Jahoor, A.; Jensen, J. Genomic prediction and GWAS of yield, quality and disease-related traits in spring barley and winter wheat. *Sci. Rep.* **2020**, *10*, 3347. [[CrossRef](#)]
39. Kilian, A.; Huttner, E.; Wenzl, P.; Jaccoud, D.; Carling, J.; Caig, V.; Evers, M.; Heller-Uszynska, K.; Uszynski, G.; Cayla, C.; et al. The fast and the cheap: SNP and DArT-based whole genome profiling for crop improvement. In *The Wake of the Double Helix: From the Green Revolution to the Gene Revolution, Proceedings of the International Congress, Bologna, Italy, 27–31 May 2003*; Tuberosa, R., Phillips, R.L., Gale, M., Eds.; Avenue Media: Bologna, Italy, 2005; pp. 443–461.
40. Brown, H.E.; Huth, N.I.; Holzworth, D.P.; Teixeira, E.I.; Zyskowski, R.F.; Hargreaves, J.N.; Moot, D.J. Plant Modelling Framework: Software for building and running crop models on the APSIM platform. *Environ. Model. Softw.* **2014**, *62*, 385–398. [[CrossRef](#)]
41. König, P.; Beier, S.; Basterrechea, M.; Schüler, D.; Arend, D.; Mascher, M.; Stein, N.; Scholz, U.; Lange, M. BRIDGE—A Visual Analytics Web Tool for Barley Genebank Genomics. *Front. Plant Sci.* **2020**, *11*, 701. [[CrossRef](#)]
42. Wang, W.; Wang, Z.; Li, X.; Ni, Z.; Hu, Z.; Xin, M.; Peng, H.; Yao, Y.; Sun, Q.; Guo, W. SnpHub: An easy-to-set-up web server framework for exploring large-scale genomic variation data in the post-genomic era with applications in wheat. *Gigascience* **2020**, *9*, giaa060. [[CrossRef](#)]
43. Watson-haigh, N.S.; Suchecki, R.; Kalashyan, E.; Garcia, M.; Baumann, U. DAWN: A resource for yielding in-sights into the diversity among wheat genomes. *BMC Genom.* **2018**, *19*, 941. [[CrossRef](#)]
44. Qiu, L.-J.; Xing, L.-L.; Guo, Y.; Wang, J.; Jackson, S.A.; Chang, R.-Z. A platform for soybean molecular breeding: The utilization of core collections for food security. *Plant Mol. Biol.* **2013**, *83*, 41–50. [[CrossRef](#)]
45. Grant, D.; Nelson, R.T.; Cannon, S.B.; Shoemaker, R.C. SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res.* **2009**, *38*, D843–D846. [[CrossRef](#)] [[PubMed](#)]
46. Doyle, A.; Doyle, J.L. Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue. *Phytochem. Bull.* **1987**, *19*, 11–15.

47. 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](#)]
48. Wang, J.; Zhang, Z. GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. *Genom. Proteom. Bioinform.* **2021**, *19*, 629–640. [[CrossRef](#)]
49. Lipka, A.E.; Tian, F.; Wang, Q.; Peiffer, J.; Li, M.; Bradbury, P.J.; Gore, M.A.; Buckler, E.S.; Zhang, Z. GAPIT: Genome association and prediction integrated tool. *Bioinformatics* **2012**, *28*, 2397–2399. [[CrossRef](#)]
50. Huang, M.; Liu, X.; Zhou, Y.; Summers, R.M.; Zhang, Z. BLINK: A package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience* **2018**, *8*, giy154. [[CrossRef](#)] [[PubMed](#)]
51. Reynolds, A.P.; Richards, G.; de la Iglesia, B.; Rayward-Smith, V.J. Clustering Rules: A Comparison of Partitioning and Hierarchical Clustering Algorithms. *J. Math. Model. Algorithms* **2006**, *5*, 475–504. [[CrossRef](#)]
52. Schubert, E.; Rousseeuw, P.J. Faster k-Medoids Clustering: Improving the PAM, CLARA, and CLARANS Algorithms. In *Similarity Search and Applications, Proceedings of the 12th International Conference, SISAP 2019, Newark, NJ, USA, 2–4 October 2019*; Springer: Cham, Switzerland, 2019; pp. 171–187. [[CrossRef](#)]
53. R Core Team. *A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020. Available online: [www.r-project.org/index.html](http://www.r-project.org/index.html) (accessed on 15 November 2022).
54. Monat, C.; Padmarasu, S.; Lux, T.; Wicker, T.; Gundlach, H.; Himmelbach, A.; Ens, J.; Li, C.; Muehlbauer, G.J.; Schulman, A.H.; et al. TRITEX: Chromosome-scale sequence assembly of Triticeae genomes with open-source tools. *Genome Biol.* **2019**, *20*, 284. [[CrossRef](#)]
55. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [[CrossRef](#)]
56. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
57. Buels, R.; Yao, E.; Diesh, C.M.; Hayes, R.D.; Munoz-Torres, M.; Helt, G.; Goodstein, D.M.; Elvik, C.G.; Lewis, S.E.; Stein, L.; et al. JBrowse: A dynamic web platform for genome visualization and analysis. *Genome Biol.* **2016**, *17*, 66. [[CrossRef](#)]
58. The International Wheat Genome Sequencing Consortium (IWGSC); Appels, R.; Eversole, K.; Feuillet, C.; Keller, B.; Rogers, J.; Stein, N.; Pozniak, C.J.; Stein, N.; Choulet, F.; et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **2018**, *361*, eaar7191. [[CrossRef](#)]
59. Zhou, Z.; Jiang, Y.; Wang, Z.; Gou, Z.; Lyu, J.; Li, W.; Yu, Y.; Shu, L.; Zhao, Y.; Ma, Y.; et al. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* **2015**, *33*, 408–414. [[CrossRef](#)]
60. Chapman, J.A.; Mascher, M.; Buluç, A.; Barry, K.; Georganas, E.; Session, A.; Strnadova, V.; Jenkins, J.; Sehgal, S.; Olliker, L.; et al. A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome Biol.* **2015**, *16*, 26. [[CrossRef](#)] [[PubMed](#)]
61. He, F.; Pasam, R.; Shi, F.; Kant, S.; Keeble-Gagnere, G.; Kay, P.; Forrest, K.; Fritz, A.; Hucl, P.; Wiebe, K.; et al. Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. *Nat. Genet.* **2019**, *51*, 896–904. [[CrossRef](#)]
62. Cheng, H.; Liu, J.; Wen, J.; Nie, X.; Xu, L.; Chen, N.; Li, Z.; Wang, Q.; Zheng, Z.; Li, M.; et al. Frequent intra- and inter-species introgression shapes the landscape of genetic variation in bread wheat. *Genome Biol.* **2019**, *20*, 136. [[CrossRef](#)] [[PubMed](#)]

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