

Article

# Development of a Multipurpose Core Collection of Bread Wheat Based on High-Throughput Genotyping Data

Laura Pascual <sup>1</sup>, Mario Fernández <sup>1</sup>, Nieves Aparicio <sup>2</sup> , Matilde López-Fernández <sup>1</sup>,  
Rosario Fité <sup>3</sup>, Patricia Giraldo <sup>1</sup>  and Magdalena Ruiz <sup>3,\*</sup> 

<sup>1</sup> Department of Biotechnology-Plant Biology, School of Agricultural, Food and Biosystems Engineering, Universidad Politécnica de Madrid, 28040 Madrid, Spain; laura.pascual@upm.es (L.P.); mario.fernandez.gonzalez@alumnos.upm.es (M.F.); matilde.lopez@upm.es (M.L.-F.); patricia.giraldo@upm.es (P.G.)

<sup>2</sup> Instituto Tecnológico Agrario de Castilla y León (ITACyL), 47071 Valladolid, Spain; apagutni@itacyl.es

<sup>3</sup> Plant Genetic Resources Centre, National Institute for Agricultural and Food Research and Technology, 28800 Alcalá de Henares, Spain; rfite@inia.es

\* Correspondence: mruiz@inia.es; Tel.: +34-915026681

Received: 12 March 2020; Accepted: 4 April 2020; Published: 8 April 2020



**Abstract:** Modern plant breeding practices have narrowed the genetic base of wheat throughout the world, increasing crop vulnerability. Therefore, there is clearly a need for introducing new germplasm in breeding programs to search for variability related to traits of agronomic interest for wheat improvement. The existence of subsets of accessions (core collections) that represent the diversity conserved in germplasm collections is a favored approach for breeders to explore novel variation and enhance the use of germplasm. In this study, a core collection of Spanish landraces of bread wheat has been created using high-throughput genotyping technologies (DArTseq), which yielded more than 50 K molecular markers. This marker system not only provides a robust estimate of the diversity, but also information about its distribution in the genome. Two core collections of 94 entries were created by using two common sampling strategies: the maximization strategy and the population structure-based method. Both core collections showed high geographic, phenotypic and genetic representativeness, but the collection obtained with the maximization strategy captured better the diversity displayed by the initial collection. This core collection, which includes a broad range of adapted genotypes, can be efficiently utilized for mining new alleles for useful traits in wheat breeding.

**Keywords:** DArTseq markers; GBS; genetic diversity; *Triticum aestivum*

## 1. Introduction

Bread wheat (*Triticum aestivum* L.) is a major staple food crop that is widely grown throughout the world. Currently, in order to meet the increasing requirements of a growing population and tackle the challenges of global climate change, the genetic improvement of this crop must achieve several goals, including higher yield, adaptation to specific environments, tolerance to biotic stresses and quality enhancement. Modern plant breeding practices, in which only a small number of elite cultivars are included in breeding programs, have narrowed the genetic base of wheat throughout the world, increasing crop vulnerability. Therefore, there is clearly a need for introducing new germplasm in breeding programs so as to broaden the gene pool in which to search for new traits of agronomic interest necessary for wheat improvement. Wheat landraces are among the most suitable germplasm resource where the genetic variation required to that end can be searched [1] and utilized through a

pre-breeding process. These locally adapted varieties, traditionally grown with less artificial resource inputs, are genetically diverse repositories of unique traits that have evolved in local environments, which cover a wide range of biotic and abiotic conditions [2,3]. Different studies have shown that Mediterranean wheat landraces represent a particularly important group of genetic resources, where extensive genetic variability as well as tolerance to drought, resistance to diseases and adaptability to low-input farming systems have been documented (see review [3]).

However, the use of wheat landraces in breeding is limited because the accessions preserved in Genebanks have not been globally characterized, leading to a scarcity of genotypic and phenotypic information. In particular, the lack of genome-wide genotypic information due to the characteristics of the wheat genome (large size and polyploidy-related complexity), along with phenotyping costs for specific traits, represent the main limiting factors for the use of such germplasm in breeding programs. It is, therefore, essential to generate subsets of accessions of a suitable size to represent the diversity conserved in germplasm collections, facilitating the availability of fine phenotyping for breeders. This approach is currently being pursued in some international breeding programs, as in the CIMMYT Seeds of Discovery program [4]. In this context, core collections, defined as a limited set of accessions chosen to represent at least 70% of the genetic variation of an entire collection with minimal redundancy, can be a powerful tool for increasing the efficiency of utilization of the germplasm stored in Genebanks [5,6].

Although conservation of allelic variation is important, the challenge of preserving quantitative genetic variation in conjunction with marker variation should be considered. Indeed, the final overall objective of a core collection is its effective use in breeding programs. The evaluation of the quality of the collection with phenotypic and genotypic data is, therefore, essential to confirm that prevalent variation types are preserved in the core collection [7]. There are two basic requirements for setting up useful core collections. The first concerns the use of efficient marker systems for unravelling the diversity of the collection. The second relates to appropriate sampling strategies to retain maximum diversity. To estimate the diversity of a collection, genotypic values are preferred over phenotypic traits to minimize genotype  $\times$  environmental (GE) interactions [8]. The assessment of genome-wide diversity by genotyping by sequencing (GBS) methods provides a robust estimate of diversity and has been increasingly adopted as a fast, high-throughput cost-effective tool for whole-genome genetic diversity analysis in large germplasm sets [9]. Moreover, this approach may reveal new alleles in wheat germplasm that might exhibit a high value for prebreeding [10–12]. DArTseq markers, based on GBS [13], efficiently target low-copy-number sequences via a complexity reduction method and provide data at a more affordable cost, especially in complex polyploid species such as wheat [14] where they have been extensively used [15].

Regarding sampling strategies, the most common approaches using molecular markers are the M (maximization) strategy [16], and stratified sampling [6,16,17]. In the M-strategy, accessions are directly selected from the whole collection by maximizing the probability of retaining all observed alleles in order to construct cores with high allelic richness [16]. This strategy, based on the existence of correlations (shared coancestry) among marker and target loci, reduces the degree of redundancy in the core collection and leads to a more effective capture of localized, high frequency alleles [18]. On the other hand, stratified sampling requires a previous knowledge of the genetic structure of the collection; all genetic groups should contribute to the core collection with the goal of optimizing the representativeness of the genetic diversity in the core collection. Different allocation strategies can be implemented to decide the number of accessions to be selected per group. The H methodology determines the size of the sample per group in proportion to their within group genetic diversity, whereas the D method determines the size in proportion to a genetic distance and/or allele diversity index within the group [19]. The use of a diversity index seems to be more effective in maximizing allele richness, especially the expected heterozygosity, which leads to core subsets less likely to be homozygous for a number of different loci [20,21]. The studies that compare the maximization and the stratified sampling strategies have reported that stratified sampling used in conjunction with the D

method formed core collections that had higher average genetic distances between genotypes, whereas the M-strategy captured more allelic diversity [20,21].

There are several computer programs to aid in core collection design. The Core Hunter algorithm has proved to be a fast and powerful method for designating core collections with increased genetic diversity, which can be applied with or without stratification of the whole collection [21–23]. Core Hunter has additional advantages: repeating the selection process produces a consistent solution, it is freely available and it is less time-consuming than other algorithms [21,22].

The objectives of the research were (a) to apply DArTseq GBS technology to provide a molecular basis for the design of a core collection of Spanish wheat landraces using the two most common approaches, M-strategy and stratified sampling, and (b) to evaluate the quality of the resulting core collections in order to select the most appropriate for a more efficient use of this germplasm in breeding. The use of the GBS technology has allowed the selection of a core collection based on more than 50K molecular markers distributed along the whole genome, and the detection of the presence of genomic regions where the different sampling methods employed performed differently. The results showed that the core collection created with the M-strategy using the Core Hunter algorithm performed better at retaining the diversity available in the initial collection.

## 2. Materials and Methods

### 2.1. Materials

The Spanish National Plant Genetic Resources Centre, CRF-INIA (Centro de Recursos Fitogenéticos, INIA, Madrid), maintains the national collection composed of 522 Spanish landraces of *Triticum aestivum* subsp. *vulgare* (Vill.). From this collection, a total of 189 genotypes were selected based on their collection site data (altitude, longitude, latitude [24]) and morphological spike traits (see Supplementary Tables S1 and S2) to represent the available diversity [25]. Homozygous lines were derived from these selected genotypes by collecting single bagged spikes from single selected plants during three generations. These 189 genotypes constituted the primary subset collection (PS) from which the entries for the final core collection were selected.

In the present study, the term “accessions” refer to genotypes that constitute the PS and “entries” are genotypes of the core collection [26].

### 2.2. Genetic and Phenotypic Characterization

High-throughput genotyping data for the PS accessions were obtained by DArTseq GBS technology at SAGA (Genetic Analysis Service for Agriculture, Mexico City, Mexico) as described in Pascual et al. [27]. This genotyping technology produces two different sets of markers: SNPs (Single Nucleotide Polymorphisms) and PAVs (Presence Absence Variants), from now on referred as DArTs (Diversity Arrays Technology markers). For this study, we selected a total of 59,276 DArTs and 14,830 SNPs, which were obtained after filtering out the markers that presented the same allelic profile or more than 10% missing data, as described by [27]. In that study, the genetic structure of the 189 accessions of the present research were analysed based on DArT markers, and the allelic profiles for the vernalization gene *Vrn-A1* and the *Glu-1* homoeoloci, determinants of wheat quality [28], were obtained. Both winter and spring landraces are included in the PS [27].

For phenotypic characterization, the accessions were sown in an augmented design during the season 2016–2017 at Alcala de Henares (Madrid). Seven qualitative (growth habit, awnedness, awn color, spike density, glume hairiness, glume color and seed color) and five quantitative agromorphological traits (days to heading and to maturity, plant height, spike length and spikelets per spike) were recorded according to the International Board of Plant Genetic Resources (IBPGR) [29] from five different plants in each accession.

### 2.3. Creation of the Core Collections

In order to determine the optimal collection size, simulations for sizes ranging from 5% to 100% of accessions included in the PS were performed with the DArT markers in the software Bio-R [30], which provides a graphical interface for the Core Hunter algorithm [21]. For simulations, the heterozygosity of the selected collections ( $HE = 1$ ) was maximized, while the default values for the rest of parameters were maintained. Finally, the relationship between collection size and genetic diversity, quantified as the number of polymorphic markers retained, was examined. The optimal size was established as the point where the number of polymorphic markers increased asymptotically.

Two core collections were constructed based on the DArT markers. The maximization core collection (MCC) was obtained with the M-strategy using the Core Hunter algorithm and the expected heterozygosity ( $HE = 1$ ) as the criteria of maximization (as described for the simulations). The stratified core collection (SCC) was created using the stratified sampling strategy. First, accessions were grouped based on populations. Then, inside each population, a number of accessions proportional to the genetic diversity ( $H_s$ ), calculated with the DArTs as described by Nei [31], was selected to maximize the expected heterozygosity ( $HE = 1$ ). Finally, a random core collection (RCC), where accessions were sampled randomly from the PS, was created to serve as reference.

### 2.4. Evaluation of the Core Collections

The quality of the different core collections created was evaluated using geographic, agromorphological and genetic data. Statistical analyses were performed with the software R version 3.5.2 [32].

For qualitative agromorphological traits and allelic profiles for the *Vrn-A1* and *Glu-1* loci, significant differences between the frequencies in the core collections and the PS were checked by Fisher's Exact Test ( $p$ -value < 0.05) [33]. For quantitative characters, the mean, variance, range and coefficient of variation were calculated for the PS, and for each one of the core collections. A homogeneity test (F-test) for variances and a t-test for means ( $p$ -value < 0.05) were used to compare the core collections and PS. The following evaluation parameters were calculated as described by Hu et al. [8]: mean difference percentage (MD), variance difference percentage (VD), coincidence rate of range (CR) and variable rate of coefficient of variation (VR). According to these parameters, a core collection can be considered representative if the percentage of traits with significant differences in their means is less than 20% ( $MD \leq 20$ ) and the coincidence rate of the range retained by the core collection is greater than 80% ( $CR \geq 80\%$ ) [8].

The genetic diversity captured in each core collection was assessed with SNP markers. Different approaches were followed to evaluate the collections. First, the genetic diversity ( $H_s$ ; [31]) was calculated for the PS and each of the core collections. Second, SNPs markers were classified according to their MAF (Minimum Allele Frequency) in: >0.1 (present in at least 19 accessions),  $\geq 0.05$  (9 accessions),  $\geq 0.03$  (6 accessions), >0.01 (2 accessions) and  $\leq 0.01$  (only in one accession). The number of markers fixed for each category in the core collections was calculated. Third, accessions selected and non-selected in each core collection were plotted in the Principal Coordinate Analysis (PCoA) performed by [27] to detect areas not sufficiently covered by the different core collections. Fourth,  $H_s$  along the bread wheat genome in the PS and different core collections was calculated based on SNP markers located in the bread wheat genome as described by [27]. Results were analyzed in order to detect genomic regions in which the core collections failed to retain the available diversity.

Finally, in order to quantify the degree of dissimilarity, Gower's genetic distances [34] between accessions were computed using the agromorphological data, SNP markers and *Glu-1* and *Vrn-A1* alleles. For each core collection, entry to entry mean-distance (E-E), the distance between each accession in the PS and the nearest entry in the core collection (A-NE) and the distance between each entry in the core collection and the nearest neighboring entry (E-NE) were calculated and averaged over all entries as described in [26]. A-NE represents the selection of entries close to each accession in the PS; thus, lower values are obtained for greater representativeness in the core entries. The E-NE distance

indicates the presence of groups of similar entries in the core; thus, both the mean and minimum E-NE reach a maximum when all the entries are far apart.

### 3. Results

High-throughput genotyping provides genetic information that can guarantee the full inclusion of the available genetic diversity when creating a core collection. In order to avoid constraints in terms of budget and time, a primary set of accessions with low genetic redundancy, representing the entire collection, was selected before genotyping. From the full collection of 522 *T. aestivum* subsp. *vulgare* accessions, we selected a PS of 189 landraces covering the full collection range for latitude, longitude and altitude (Supplementary Table S1). This PS included landraces collected in the first half of the 20th century from all Spanish regions (including the two Spanish archipelagos), in which nine agroecological growing zones have been described [35]. The PS also covers the variability for six traits currently used in wheat accession characterization [29] (Supplementary Table S2). According to a previous study [27], the PS was subdivided into 4 populations, with Pop 2 having the highest number of accessions. The genetic diversity values ( $H_s$ ) in each population ranged from 0.13 to 0.32 (Table 1).

#### 3.1. Creation of the Core Collections

The final size of the core collection was determined based on simulations using the 59,276 polymorphic DArT markers present in the PS. For sizes larger than 94 entries, the genetic gain (estimated as the number of polymorphic markers) increased asymptotically (Figure S1). Thus, the size of the core collection was established in 94 genotypes, which captured 56,451 of the polymorphic DArT markers.

DArT markers were also used to select entries of the core collections. The MCC was obtained with the M-strategy by maximizing the expected heterozygosity. The SCC was created using the stratified sampling strategy based on the genetic structure of the PS and the genetic diversity within each population. Thus, those populations with higher diversity contributed a higher proportion of entries to the core collection. Finally, the RCC collection was established by randomly selecting entries from the PS. The final number of entries from each population included in the MCC, SCC and RCC are indicated in Table 1.

**Table 1.** Genetic diversity ( $H_s$ ) of DArT (Diversity Array Technology) markers in each population (Pop) of the primary set and final number of accessions from each population in the core collections.

	Pop 1	Pop 2	Pop 3	Pop 4
$H_s$	0.21	0.32	0.13	0.23
PS	25	112	16	36
MCC	13	50	8	23
SCC	21	37	10	26
RCC	10	55	8	21

PS, primary set; MCC, core collection generated with maximization strategy; SCC, core collection generated with stratified sampling; RCC, core collection generated with random sampling.

#### 3.2. Evaluation of Core Collections

A core collection of germplasm is a useful approach for breeders only when it includes most of the available variation (both genotypic and phenotypic). In this study, each core collection was evaluated considering genotypic and phenotypic data not used for the selection of entries. We evaluated the quality of the three core collections (CCs) at different levels: (1) representativeness of the CC for geographic, phenotypic and allelic (*Glu-1* and *Vrn-A1* loci) variability present in the PS; (2) allelic richness estimated from SNPs; (3) degree of dissimilarity and redundancy according to distances between accessions; and (4) distribution of the genetic variability included in each CC with respect to the PS and along the full genome.

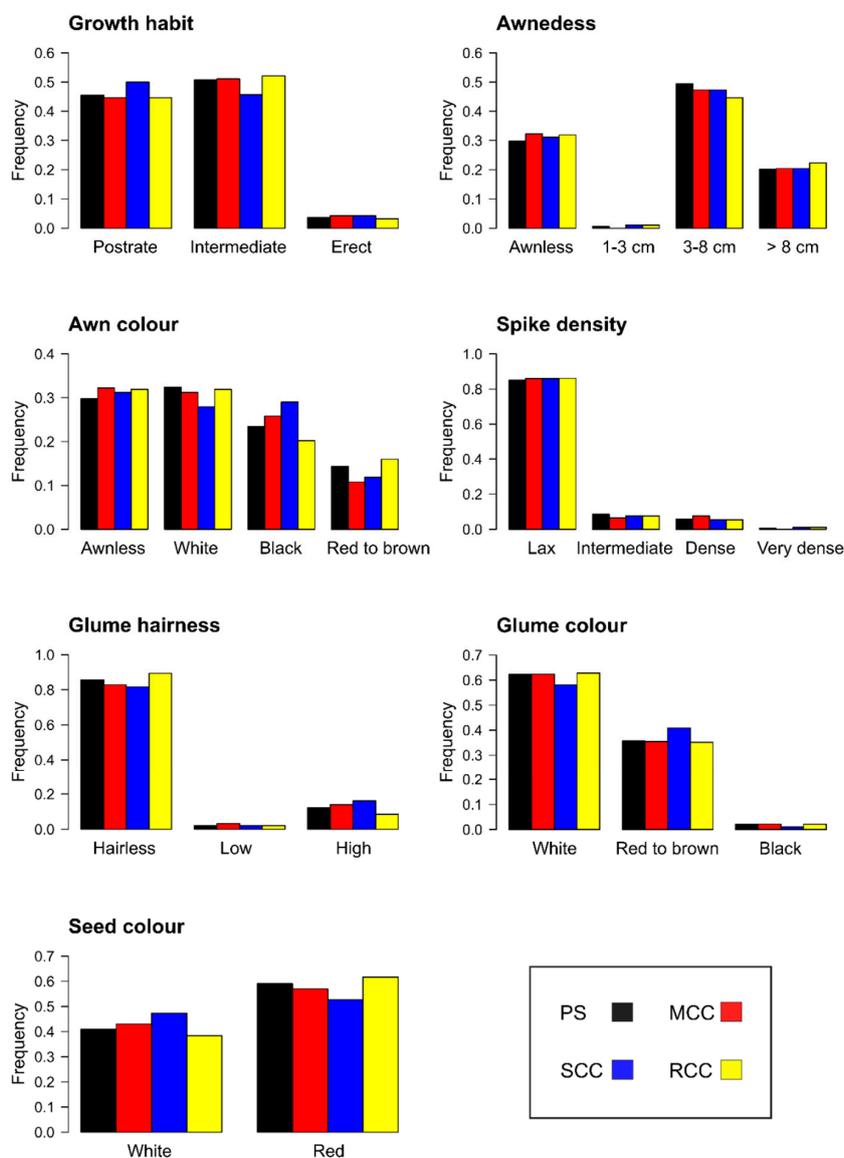
### 3.2.1. Representativeness of the Core Collections

The three CCs represented adequately the geographic diversity and variability of qualitative morphological traits present in the PS (Table 2, Figure 1).

**Table 2.** Latitude, longitude and elevation ranges covered by the primary set and the core collections.

	Latitude	Longitude	Elevation (m)
PS	433310N–281820N	0174928W–0041559E	10–1610
MCC	433310N–281820N	0174928W–0041559E	22–1540
SCC	433310N–281820N	0174928W–0041559E	35–1540
RCC	433310N–281820N	0162421W–0031238E	63–1540

PS, primary set; MCC, core collection generated with maximization strategy; SCC, core collection generated with stratified sampling; RCC, core collection generated with random sampling.



**Figure 1.** Relative frequencies of qualitative agromorphological traits in the primary set (PS) and core collections generated with the maximization strategy (MCC), stratified sampling (SCC) and random sampling (RCC).

The results also showed that the frequency distributions of the qualitative agromorphological traits in the CCs were not significantly different from the PS ( $p$ -values for Fisher tests ranging from 0.37 to 1). For the quantitative traits, no significant differences among means were detected in the CCs (Table 3). Moreover, the null values for the evaluation parameters MD and VD indicated that the PS was properly represented in the core collections (Table 3), and the high VR and CR values in the SCC and MCC revealed the prevalence of diverse entries in these subsets. The MCC generally had the highest coefficient of variation values, higher than those of the PS for some traits.

**Table 3.** Summary statistics for quantitative agromorphological traits in the primary set and core collections.

		Days to Heading (Days)	Days to Maturity (Days)	Plant Height (cm)	Spike Length (mm)	Spikelets Per Spike (Number)	Evaluation Parameter
Mean	PS	171.23	206.86	88.27	117.03	19.11	-
	MCC	171.48	206.81	87.89	118.36	19.31	0
	SCC	172.08	206.9	85.97	115.98	19.02	0
	RCC	171.19	207	88.19	119	19.28	0
Variance	PS	49.16	11.34	137.87	365.54	4.16	-
	MCC	47.48	10.65	160.14	403.72	4.52	0
	SCC	48.83	9.59	137.73	340.78	4.22	0
	RCC	48.97	10.54	159.7	385.03	3.51	0
Coefficient of Variation	PS	4.09	1.63	13.3	16.34	10.68	-
	MCC	4.02	1.58	14.4	16.98	11.01	102.06
	SCC	4.04	1.49	13.58	15.83	10.74	97.92
	RCC	4.09	1.57	14.33	16.49	9.72	99.18
Range	PS	155–188	199–216	53–119	59–168	14–24	-
	MCC	157–188	199–216	53–115	67–168	14–24	91.53
	SCC	157–188	199–215	53–114	59–168	14–24	91.64
	RCC	155–188	200–216	53–119	59–153	14–23	89.46

PS, primary set; MCC, core collection generated with maximization strategy; SCC, core collection generated with stratified sampling; RCC, core collection generated with random sampling; MD, mean difference percentage; VD, variance difference percentage; VR, variable rate of coefficient of variation; CR, coincidence rate of range.

Regarding the genotypic data, all *Glu-1* alleles were included in the MCC and SCC (Table 4), whereas the RCC failed to capture one allele at each of the *Glu-B1* and *Glu-D1* loci. The three CCs included the three *Vrn-A1* alleles identified in the Spanish landraces in a similar proportion, especially in the MCC (Table 4).

**Table 4.** Alleles at the *Glu-1* homoeoloci and gene *Vrn-A1* in the primary set and the core collections.

	<i>Glu-1</i> Homoeoloci			<i>Vrn-A1</i> Alleles (%)		
	<i>Glu-A1</i>	<i>Glu-B1</i>	<i>Glu-D1</i>	<i>Vrn-A1</i>	<i>Vrn-A1a</i>	<i>Vrn-A1b</i>
PS	<i>a,b,c,y</i>	<i>a,al,am,aq,d,e,f,h,i,u,n2,n3,n4,n5,n6</i>	<i>a,c,d,h,j,l,n6</i>	18.52	52.38	29.10
MCC	<i>a,b,c,y</i>	<i>a,al,am,aq,d,e,f,h,i,u,n2,n3,n4,n5,n6</i>	<i>a,c,d,h,j,l,n6</i>	18.09	48.94	32.98
SCC	<i>a,b,c,y</i>	<i>a,al,am,aq,d,e,f,h,i,u,n2,n3,n4,n5,n6</i>	<i>a,c,d,h,j,l,n6</i>	12.77	55.32	31.91
RCC	<i>a,b,c,y</i>	<i>a,al,am,aq,d,e,f,h,i,u,n3,n4,n5,n6</i>	<i>a,c,d,h,j,l</i>	15.96	50.00	34.04

PS, primary set; MCC, core collection generated with maximization strategy; SCC, core collection generated with stratified sampling; RCC, core collection generated with random sampling.

### 3.2.2. Representativeness of the Core Collections

The allelic richness of the CCs was evaluated with 14,830 polymorphic SNPs in the PS. To analyze the degree of allele fixation in the three subsets, we studied the presence of monomorphic markers (Table 5). All the SNPs with predominant alleles (MAF > 0.1) were polymorphic in the three CCs. For the rest of the markers, the MCC had the lowest number of fixed markers, whereas the SCC showed the highest values. Taking into account the number of accessions in the PS, the least frequent allele in

markers with  $MAF \leq 0.01$  was present in only one accession. Thus, it was expected that some of them would not be included. Overall, the MCC possessed the highest gene diversity value, equal to that in the PS.

**Table 5.** Genetic diversity ( $H_s$ ) and distribution of minor allele frequency (MAF) for the SNP (Single Nucleotide Polymorphisms) markers in the primary set and those fixed in the core collections.

MAF	SNP Markers (Number)		Fixed SNP Markers (Number)		
	PS	MCC	SCC	RCC	
>0.1	6376	0	0	0	0
$\geq 0.05$	8394	0	6	2	
$\geq 0.03$	10,092	3	43	8	
>0.01	14,002	534	685	586	
$\leq 0.01$	828	387	419	453	
Total	14,830	921	1104	1049	
$H_s$	0.20	0.20	0.19	0.19	

PS, primary set; MCC, core collection generated with maximization strategy; SCC, core collection generated with stratified sampling; RCC, core collection generated with random sampling.

### 3.2.3. Distances between Entries

The mean Gower's distance between entries (E-E) was higher in the CCs than in the PS, thereby providing a gain of 1% in the RCC and SCC, and 2% in the MCC (Table 6). The three CCs showed higher values for the minimum distances among the entries (mean E-NE distance and minimum E-NE distance) than the PS, especially the MCC. This last subset and the RCC also showed the lowest A-NE distances.

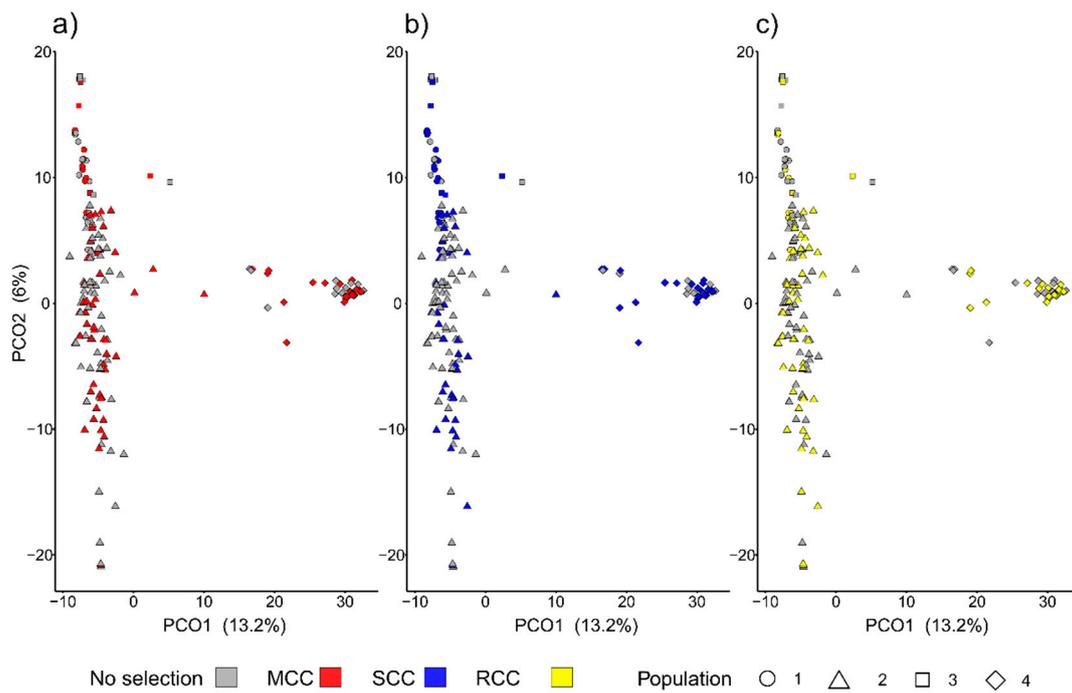
**Table 6.** Gower's genetic distance values between accessions in the primary set and core collections.

	E-E	A-NE	E-NE	Min E-NE
PS	0.189	-	0.073	0.0061
MCC	0.192	0.043	0.090	0.0067
SCC	0.190	0.051	0.081	0.0065
RCC	0.191	0.043	0.081	0.0067

PS, primary set; MCC, core collection generated with maximization strategy; SCC, core collection generated with stratified sampling; RCC, core collection generated with random sampling; E-E, mean distance between entries; A-NE, average distance between each accession in the PS and the nearest entry in the core collection; E-NE, average distance between each entry in the core collection and the nearest neighboring entry; E-NE min, minimum distance between each entry in the core collection and the nearest neighboring entry.

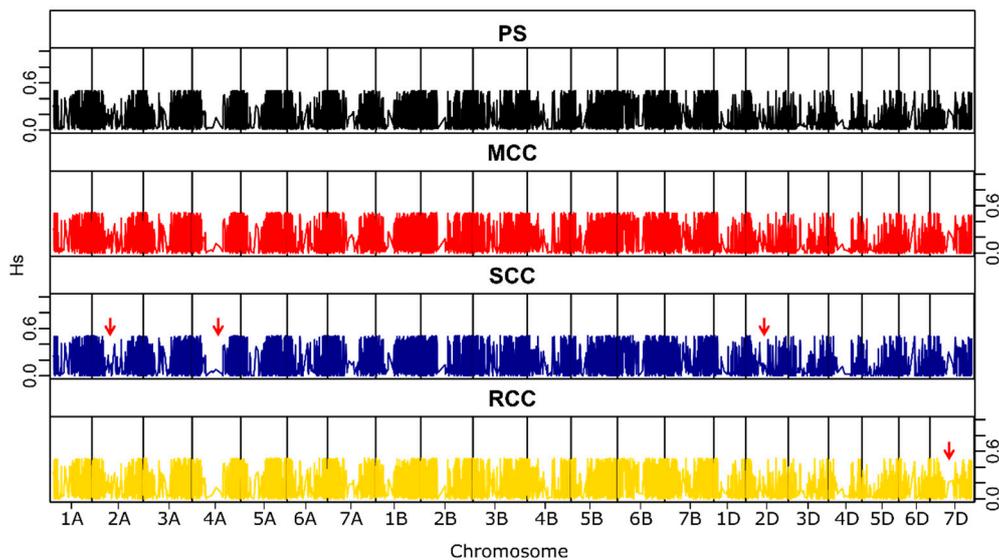
### 3.2.4. Distribution of Genetic Variability

The capture of the available genetic diversity in the three CCs was analyzed by representing the selected accessions in each one of them with respect to the PS by a PCoA analysis explaining 19.2% of the available SNPs diversity (Figure 2). The three subsets well covered the genetic variation, capturing accessions from the four populations of the PS. The RCC, however, failed to include some accessions from Pop 2 placed in the central part of the PCoA graph (Figure 2).

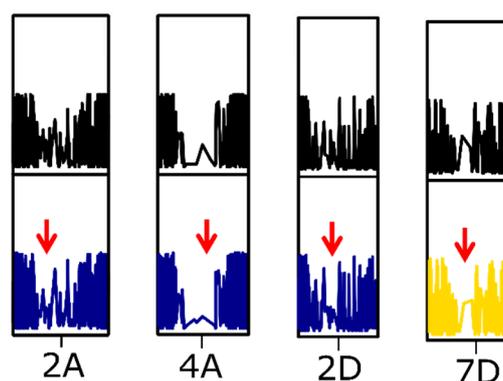


**Figure 2.** Biplot of the first two axes of the principal coordinate analysis according to Pascual et al. [27] showing the relative distribution of the primary set and core collections generated with (a) maximization strategy (MCC), (b) stratified sampling (SCC) and (c) random sampling (RCC). The genetic population is also indicated in the representation of the accessions.

Finally, we analyzed the genetic diversity along the bread wheat genome in the PS and the three CCs (Figure 3). In this case, the SCC failed to capture all the available diversity in the centromeric regions of chromosomes 2A, 4A and 2D. The RCC included less diversity in the 7D, and the MCC captured most of the diversity present in the PS along all the chromosomes. Considering both analyses, the MCC better represented the distribution of genetic variability in the PS.



**Figure 3.** Cont.



**Figure 3.** Genetic diversity ( $H_s$ ) distribution across the genome in the primary set (PS) and the core collections generated with the maximization strategy (MCC), stratified sampling (SCC) and random sampling (RCC). Red arrows indicate lower genetic diversity regions when compared with the PS. Expanded subsections of these regions are shown.

#### 4. Discussion

Several studies have shown the considerable variability among the Spanish wheat landraces compared to other germplasm collections [36,37]. These materials can be an important source of genes for wheat improvement, such as rust resistance [38] and quality traits [37,39]. However, even after the removal of redundant accessions, the collection maintained at CRF-INIA comprising more than 500 landraces is too large for evaluation and use in most breeding programs. Defining a core collection is, therefore, a pre-requisite and valuable tool for utilizing this germplasm, particularly for more complicated phenotypic screens, as has been demonstrated by the Spanish core collections of barley and durum wheat [40,41]. Both core subsets have facilitated detailed study of some difficult to analyze traits such as yield performance, root architecture, disease resistance or characterization of low molecular glutenin subunits [42–46].

On the other hand, the high genetic variability of the Spanish collection complicated the designing of a core collection of suitable size and able to capture the diversity present in the entire collection. To determine the optimal number of entries needed to retain an acceptable proportion of alleles present in the primary set, we tried to find a “point of compromise” between gain of genetic variation and elimination of genetic redundancy in the core collection. The final size of 94 entries captured 96% of polymorphic DArT markers present in the PS, which is in agreement with other studies that have reported values between 70 and 98% using SNPs [47,48] and SSR [40,49,50]. This size represents 18% of the entire collection, and is within the range from 5 to 30% recommended for retaining a great part of the genetic variability with a manageable number of accessions (e.g., [5,40,51]).

Preserving maximum genetic variation with a small number of accessions is a challenging task, and several improvements in sampling strategies have been devised over the last two decades. In order to create a multipurpose collection, we constructed two core collections using the most common approaches: M-strategy [16] and stratified sampling [6,16,17]. According to the objectives of these methodologies, we expected that the MCC maximized the total allelic diversity, and selected more diverse entries, whereas the SCC optimized the representativeness of the genetic diversity, including more representative entries. However, the effective utilization of the resulting core collections in breeding programs depends directly on their quality, which should be correctly evaluated.

Depending on the purpose of a core collection, a variety of metrics can be used to evaluate its quality. In the present study, considering that the aim is a multipurpose collection, the quality was evaluated using different types of variables such as geographic, phenotypic (discrete and continuous) and genotypic data that were not used in the core selection [26,51]. Also, a random core collection was generated to serve as a reference. Both the M-strategy and the stratified sampling selected entries that represented the geographic, phenotypic and genotypic variability of the PS, which validated both

sampling strategies. However, the MCC included higher variation for quantitative agromorphological characters, indicating that this subset maximized the representativeness of the pattern of variation of these traits in the PS [26]. The allelic richness of the collections was analyzed with 14,830 SNPs covering all chromosomes. This genome-wide assay has proved to be a very convenient method for the analysis of the variability in germplasm collections [52]. On average, the CCs captured between 92.6 (in the SCC) and 93.8% (in the MCC) of SNPs present in the PS, and more than 90% of common alleles ( $0.1 \geq \text{MAF} > 0.01$ ). Common-localized alleles may be biologically specialized alleles that enhanced adaptation to different agroecological conditions, and is often the class of alleles most interesting to breeders. Other studies have reported that crop improvement is accompanied by a selective advantage of rare alleles present at low frequency ( $\text{MAF} < 0.05$ ) [53,54]. The percentage of rare allele recovery was 82.9% in the SCC and 85.7% in MCC, which was in agreement with that obtained in other core collections with SNPs [55]. The core collections performed worse in preserving very rare-localized alleles ( $\text{MAF} \leq 0.01$ ) that were present in only one accession, which is in agreement with Wingen et al. [49]. Some studies reported that this type of allele, likely to be maintained by deleterious mutation-selection balance, would be of less interest; they seldom contribute to the improvement of elite varieties and, therefore, their inclusion in the CC might not be worthwhile [7,56,57]. In contrast, other authors have more recently proposed core subsets focused on preserving rare accessions and uncommon alleles, which may have unique genetic potentials for plant breeding [58]. In our case, the high number of accessions possessing specific rare alleles makes it difficult to retain a greater number of very rare alleles without increasing the sample size. Nevertheless, the MCC also maximized the coverage of this type of alleles and included an increased number of the most divergent accessions (56 in MCC, 52 in SCC and 50 in RCC out of the 94 accessions with the highest mean E-E distance in the PS). Moreover, the screening of the genetic diversity in each CC along the genome revealed that the MCC was able to better capture the available diversity from the PS in chromosomes 2A, 4A and 2D than the SCC.

Allelic richness is an evaluation criterion that ensures the inclusion of restricted alleles, whereas genetic distances between accessions is an evaluation criterion related to the concept of the maximization of the representation of genetic diversity in the whole collection [19,26,57]. In the present study, both phenotypic and genotypic variables were combined to calculate the distance among accessions since they provided complementary information thereby maximizing overall diversity for analysis [55,59]. The mean distance between accessions (E-E) reaches a maximum when diverse entries are sampled, but the presence of similar entries at the extreme ends of the distributions cannot be distinguished by this criterion. Therefore, two additional distances (A-NE and E-NE) recommended to evaluate the quality of multipurpose CCs were calculated [26]. The A-NE distance is a good criterion to evaluate the representativeness of genetic diversity of the PS, whereas the E-NE distance allows evaluation of whether the core collection has entries that are as different as possible from each other [26]. The lower A-NE and the higher E-NE distances in the MCC indicated that this subset maximized the representativeness of the genetic diversity of the PS and that all the entries were far apart genetically. The PCoA of the MCC also demonstrated that this subset was well distributed within the PS, covering the four genetic populations, even though the information on the genetic structure of the collection was not used to constrain the subset extraction. To some extent, this latter result could be related to the small number of populations identified in the collection [60].

Other studies have shown that the CC created based on the M-strategy maximizes the genetic variability index [23,40], whereas the structured method yields subsets that better represent the distribution of the genetic variability of the initial collections [7,20,40,57]. Also, worse values for E-NE distances have been reported in collections developed using the M-strategy [23,40]. In the present study, however, the M-strategy performed better than the stratified method, by increasing genetic diversity and reducing redundancy [26]. The higher quality of the MCC could be due to the lower degree of stratification of our collection. Only Pop 4 was clearly separated, while some overlap was shown among the other three populations, especially within Pop 2, which was the largest population [27]. In contrast, the stratified sampling performed better than the M-strategy in the Spanish durum wheat

collection [40]. Comparisons of the genetic structure of the two wheat collections revealed that bread wheat populations exhibited a higher level of admixture and less genetic differentiation (Population differentiation index  $D_{ST} = 0.22$  in durum and  $0.17$  in bread wheat) [27]. This finding may explain why population-based sampling did not optimize the representativeness of the genetic diversity in bread wheat. Furthermore, the little gain by minimizing A–NE in the MCC compared to RCC could be also caused by the weaker structure of our bread wheat collection [26]. The results presented here have shown that the core collection designed with the M-strategy had superior performance; thus, this subset was selected as the Spanish core collection.

The Spanish wheat core collection constructed in the present study contains genotypes collected from every region in Spain where bread wheat is cultivated. Such coverage is essential because growing regions possess very diverse environmental conditions in terms of climate, altitude and soil characteristics. Wheat is grown from cold sub-humid areas in the northern parts of Spain to warm semi-arid regimes in the southeast [61], in basic or neutral soils in the Centre and East, and acid soils in the western regions [62]. Our core collection also includes all the *Vrn-A1* alleles for the vernalization response identified in Spanish landraces. The *Vrn-A1* gene is one of the most determinant loci involved in the transition from vegetative to reproductive growth [63], and thus for wheat adaptability. Such adaptation of Spanish landraces to different agroecological conditions has resulted in the accumulation of favorable alleles, including for stress tolerance, which can be incorporated into breeding programs [44,64–66]. Furthermore, in the case of wheat, functional quality requirements must also be kept in mind in order to have a multipurpose collection useful for wheat improvement. Our collection covers the allelic variation for 30 alleles at the *Glu-1* loci, the main genetic determinants of gluten quality [28,37].

## 5. Conclusions

The use of high-throughput genotyping technologies has allowed the selection of a core collection based on more than 50K molecular markers distributed across the whole genome. In addition, this approach has enabled us to detect the presence of genomic regions where different sampling methods employed performed differently. The M-method using the Core Hunter algorithm has demonstrated to be a fast and powerful method for designating core collections, especially for non-highly structured collections or in the absence of knowledge of a clear genetic structure in the whole collection. The core collection of Spanish landraces of bread wheat designed in the present study includes a broad range of adapted genotypes, and maximizes the representativeness of the genetic and phenotypic diversity in the initial collection of 522 landraces. This wheat core collection can be efficiently utilized in mining new alleles for useful traits and in broadening the genetic base in the cultivated wheat germplasm pool.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4395/10/4/534/s1>, Figure S1: Relationship between number of polymorphic DArT markers and sample size. The trend line is shown in black. Table S1: Latitude, longitude and elevation ranges covered by the initial collection of 522 accessions and the primary set. Table S2: Qualitative agromorphological traits classes included in the initial 522 accessions and primary set.

**Author Contributions:** Conceptualization, M.R. and L.P.; data curation, L.P., M.F., R.F., M.L.-F. and N.A.; formal analysis, L.P., M.F., N.A. and M.R.; funding acquisition, P.G.; investigation, L.P., M.F., M.L.-F., P.G. and M.R.; methodology, L.P., M.F., M.L.-F., R.F., P.G. and M.R.; project administration, P.G. and L.P.; resources, P.G., L.P., M.R.; supervision, L.P. and M.R.; validation, L.P. and M.R.; visualization, L.P., M.F., M.L.-F. and M.R.; writing—original draft, M.R. and L.P.; writing—review and editing, M.R., L.P. and P.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Spanish Autonomous Community of Madrid under Grants No. AGRISOST-CM S2018/BAA-4330, the National Institute for Agricultural and Food Research and Technology under Grants No. AT2016-006 and RFP2015-00008-C04-01, the Universidad Politécnica de Madrid under Grant No. VJIDOCUPM18LPB, Ministry of Economy, Industry and Competitiveness under Grants No. AGL2016-77149-C2-1P, and Structural Funds 2014–2020 (ERDF and ESF). M.L.-F. is a recipient of a predoctoral fellowship from the Programa Propio of the Universidad Politécnica de Madrid.

**Acknowledgments:** We thank M.J. Tomás for technical assistance.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. FAO. Harvesting Nature's Diversity. Available online: <http://www.fao.org/3/v1430e/V1430E04.htm> (accessed on 2 March 2020).
2. Ficiciyan, A.; Loos, J.; Sievers-Glotzbach, S.; Tschamtkke, T. More than yield: Ecosystem services of traditional versus modern crop varieties revisited. *Sustainability* **2018**, *10*, 2834. [[CrossRef](#)]
3. Lopes, M.S.; El-Basyoni, I.; Baenziger, P.S.; Singh, S.; Royo, C.; Ozbek, K.; Aktas, H.; Ozer, E.; Ozdemir, F.; Manickavelu, A.; et al. Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change. *J. Exp. Bot.* **2015**, *66*, 3477–3486. [[CrossRef](#)] [[PubMed](#)]
4. CIMMYT. Seeds of Discovery. Available online: <http://seedsofdiscovery.org/en/> (accessed on 5 March 2020).
5. Brown, A. The case for core collections. In *The Use of Plant Genetic Resources*; Brown, A., Frankel, O., Marshall, D., Williams, J., Eds.; Cambridge University Press: Cambridge, UK, 1989; pp. 136–156.
6. Brown, A. Core collections: A practical approach to genetic resources management. *Genome* **1989**, *31*, 818–824. [[CrossRef](#)]
7. Zhang, H.; Zhang, D.; Wang, M.; Sun, J.; Qi, Y.; Li, J.; Han, L.; Qiu, Z.; Tang, S.; Li, Z. A core collection and mini core collection of *Oryza sativa* L in China. *Appl. Genet.* **2011**, *122*, 49–61. [[CrossRef](#)] [[PubMed](#)]
8. Hu, J.; Zhu, J.; Xu, H. Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. *Appl. Genet.* **2000**, *101*, 264–268. [[CrossRef](#)]
9. Holtz, Y.; Ardisson, M.; Ranwez, V.; Besnard, A.; Leroy, P.; Poux, G.; Roumet, P.; Viader, V.; Santoni, S.; David, J. Genotyping by sequencing using specific allelic capture to build a high-density genetic map of durum wheat. *PLoS ONE* **2016**, *11*, e0154609. [[CrossRef](#)]
10. Kilian, B.; Graner, A. NGS technologies for analyzing germplasm diversity in genebanks. *Brief. Funct. Genom.* **2012**, *11*, 38–50. [[CrossRef](#)]
11. Heslot, N.; Rutkoski, J.; Poland, J.; Jannink, J.-L.; Sorrells, M.E. Impact of marker ascertainment bias on genomic selection accuracy and estimates of genetic diversity. *PLoS ONE* **2013**, *8*, e74612. [[CrossRef](#)]
12. Manickavelu, A.; Jighly, A.; Ban, T. Molecular evaluation of orphan afghan common wheat (*Triticum aestivum* L.) landraces collected by Dr. Kihara using single nucleotide polymorphic markers. *BMC Plant Biol.* **2014**, *14*, 320. [[CrossRef](#)]
13. Sansaloni, C.; Petrolini, C.; Jaccoud, D.; Carling, J.; Detering, F.; Grattapaglia, D.; Kilian, A. Diversity arrays technology (DART) and next-generation sequencing combined: Genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. *BMC Proc.* **2011**, *5*, 54. [[CrossRef](#)]
14. Diversity Arrays Technology. Available online: <https://www.diversityarrays.com> (accessed on 5 March 2020).
15. Rasheed, A.; Mujeeb-Kazi, A.; Ogbonnaya, F.C.; He, Z.; Rajaram, S. Wheat genetic resources in the post-genomics era: Promise and challenges. *Ann. Bot.* **2018**, *121*, 603–616. [[CrossRef](#)] [[PubMed](#)]
16. Schoen, D.J.; Brown, A. Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 10623–10627. [[CrossRef](#)] [[PubMed](#)]
17. Erskine, W.; Muehlbauer, F. Allozyme and morphological variability, outcrossing rate and core collection formation in lentil germplasm. *Appl. Genet.* **1991**, *83*, 119–125. [[CrossRef](#)]
18. Bataillon, T.M.; David, J.L.; Schoen, D.J. Neutral genetic markers and conservation genetics: Simulated germplasm collections. *Genetics* **1996**, *144*, 409–417.
19. Franco, J.; Crossa, J.; Taba, S.; Shands, H. A sampling strategy for conserving genetic diversity when forming core subsets. *Crop Sci.* **2005**, *45*, 1035–1044. [[CrossRef](#)]
20. Franco, J.; Crossa, J.; Warburton, M.L.; Taba, S. Sampling strategies for conserving maize diversity when forming core subsets using genetic markers. *Crop Sci.* **2006**, *46*, 854–864. [[CrossRef](#)]
21. Thachuk, C.; Crossa, J.; Franco, J.; Dreisigacker, S.; Warburton, M.; Davenport, G.F. Core hunter: An algorithm for sampling genetic resources based on multiple genetic measures. *BMC Bioinform.* **2009**, *10*, 243. [[CrossRef](#)]
22. Díez, C.M.; Imperato, A.; Rallo, L.; Barranco, D.; Trujillo, I. Worldwide core collection of olive cultivars based on simple sequence repeat and morphological markers. *Crop Sci.* **2012**, *52*, 211–221. [[CrossRef](#)]

23. Krishnan, R.R.; Sumathy, R.; Ramesh, S.; Bindroo, B.; Naik, G.V. SimEli: Similarity elimination method for sampling distant entries in development of core collections. *Crop Sci.* **2014**, *54*, 1070–1078. [[CrossRef](#)]
24. INIA. National Inventory of Plant Genetic Resources. Available online: [http://webx.inia.es/web\\_inventario\\_nacional/Introduccioneng.asp](http://webx.inia.es/web_inventario_nacional/Introduccioneng.asp) (accessed on 5 March 2020).
25. Aparicio, N.; Alvaro, F.; Sillero, J.C.; Ruiz, M.; López, P.; Cátedra, M.; Codesal, P. Bread Wheat (*Triticum aestivum*, L.) Core Collection based in Spanish landraces. In Proceedings of the 8th International Wheat Conference, St. Petersburg, Russia, 1–4 June 2010; Dzyubenko, N.I., Ed.; NI Vavilov Research Institute of Plant Industry (VIR): Saint Petersburg, Russia, 2005; p. 85.
26. Odong, T.; Jansen, J.; Van Eeuwijk, F.; van Hintum, T.J. Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation. *Appl. Genet.* **2013**, *126*, 289–305. [[CrossRef](#)]
27. Pascual, L.; Ruiz, M.; López-Fernández, M.; Pérez-Peña, H.; Benavente, E.; Vázquez, J.F.; Sansaloni, C.; Giraldo, P. Genomic analysis of Spanish wheat landraces reveals their variability and potential for breeding. *BMC Genom.* **2020**, *21*, 122. [[CrossRef](#)]
28. Payne, P.I.; Nightingale, M.A.; Krattiger, A.F.; Holt, L.M. The relationship between HMW glutenin subunit composition and the bread-making quality of British-grown wheat varieties. *J. Sci. Food Agric.* **1987**, *40*, 51–65. [[CrossRef](#)]
29. IBPGR. *Revised Descriptor List for Wheat (Triticum spp)*; International Board for Plant Genetic Resources: Rome, Italy, 1985; p. 12.
30. Pacheco, A.; Alvarado, G.; Rodríguez, F.; Burgueño, J. *BIO-R (Biodiversity Analysis with R for Windows) Version 2.0*; Cimmyt: Sonora, Mexico, 2016.
31. Nei, M. Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA* **1973**, *70*, 3321–3323. [[CrossRef](#)]
32. R Core Team R. *A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018.
33. Fisher, R. *Statistical Methods for Research Workers*; Oliver and Boyd: Edinburgh, Scotland, 1954.
34. Gower, J.C. A general coefficient of similarity and some of its properties. *Biometrics* **1971**, 857–871. [[CrossRef](#)]
35. Ruiz, M.; Giraldo, P.; Royo, C.; Villegas, D.; Jose Aranzana, M.; Carrillo, J.M. Diversity and genetic structure of a collection of spanish durum wheat landraces. *Crop Sci.* **2012**, *52*, 2262–2275. [[CrossRef](#)]
36. Ruiz, M.; Rodriguez-Quijano, M.; Metakovsky, E.V.; Vazquez, J.F.; Carrillo, J.M. Polymorphism, variation and genetic identity of Spanish common wheat germplasm based on gliadin alleles. *Field Crop. Res.* **2002**, *79*, 185–196. [[CrossRef](#)]
37. Giraldo, P.; Rodriguez-Quijano, M.; Simon, C.; Vazquez, J.F.; Carrillo, J.M. Allelic variation in HMW glutenins in Spanish wheat landraces and their relationship with bread quality. *Span. J. Agric. Res.* **2010**, *8*, 1012–1023. [[CrossRef](#)]
38. Martinez, F.; Niks, R.E.; Moral, A.; Urbano, J.M.; Rubiales, D. Search for partial resistance to leaf rust in a collection of ancient Spanish wheats. *Hereditas* **2001**, *135*, 193–197. [[CrossRef](#)]
39. Vázquez, J.F.; Chacón, E.A.; Carrillo, J.M.; Benavente, E. Grain mineral density of bread and durum wheat landraces from geochemically diverse native soils. *Crop Pasture Sci.* **2018**, *69*, 335–346. [[CrossRef](#)]
40. Ruiz, M.; Giraldo, P.; Royo, C.; Carrillo, J.M. Creation and validation of the spanish durum wheat core collection. *Crop Sci.* **2013**, *53*, 2530–2537. [[CrossRef](#)]
41. Igartua, E.; Gracia, M.; Lasa, J.; Medina, B.; Molina-Cano, J.; Montoya, J.; Romagosa, I. The Spanish barley core collection. *Genet. Resour. Crop Evol.* **1998**, *45*, 475–481. [[CrossRef](#)]
42. Silvar, C.; Flath, K.; Kopahnke, D.; Gracia, M.; Lasa, J.; Casas, A.; Igartua, E.; Ordon, F. Analysis of powdery mildew resistance in the Spanish barley core collection. *Plant Breed.* **2011**, *130*, 195–202. [[CrossRef](#)]
43. Yahiaoui, S.; Cuesta-Marcos, A.; Gracia, M.P.; Medina, B.; Lasa, J.M.; Casas, A.M.; Ciudad, F.J.; Montoya, J.L.; Moralejo, M.; Molina-Cano, J.L. Spanish barley landraces outperform modern cultivars at low-productivity sites. *Plant Breed.* **2014**, *133*, 218–226. [[CrossRef](#)]
44. Ruiz, M.; Giraldo, P.; González, J.M. Phenotypic variation in root architecture traits and their relationship with eco-geographical and agronomic features in a core collection of tetraploid wheat landraces (*Triticum turgidum* L.). *Euphytica* **2018**, *214*, 54. [[CrossRef](#)]
45. Ruiz, M.; Bernal, G.; Giraldo, P. An update of low molecular weight glutenin subunits in durum wheat relevant to breeding for quality. *J. Cereal Sci.* **2018**, *83*, 236–244. [[CrossRef](#)]

46. Martínez-Moreno, F.; Ruiz, M.; Blandón, M.; Giraldo, P. Resistance to leaf rust in a core collection of ancient Spanish tetraploid wheats. In Proceedings of the I Spanish Symposium on Cereal Physiology and Breeding, Zaragoza, Spain, 9–10 April 2018; p. 34.
47. Acuña-Matamoros, C.L.; Reyes-Valdés, M.H. Comparison of optimization methods for core subset selection from a large collection of Mexican wheat landraces characterized by SNP markers. *Plant Genet. Resour.* **2018**, *16*, 228–236. [[CrossRef](#)]
48. Kobayashi, F.; Tanaka, T.; Kanamori, H.; Wu, J.; Katayose, Y.; Handa, H. Characterization of a mini core collection of Japanese wheat varieties using single-nucleotide polymorphisms generated by genotyping-by-sequencing. *Breed. Sci.* **2016**, *66*, 213–225. [[CrossRef](#)]
49. Wingen, L.U.; Orford, S.; Goram, R.; Leverington-Waite, M.; Bilham, L.; Patsiou, T.S.; Ambrose, M.; Dicks, J.; Griffiths, S. Establishing the AE Watkins landrace cultivar collection as a resource for systematic gene discovery in bread wheat. *Appl. Genet.* **2014**, *127*, 1831–1842. [[CrossRef](#)]
50. Balfourier, F.; Roussel, V.; Strelchenko, P.; Exbrayat-Vinson, F.; Sourdille, P.; Boutet, G.; Koenig, J.; Ravel, C.; Mitrofanova, O.; Beckert, M. A worldwide bread wheat core collection arrayed in a 384-well plate. *Appl. Genet.* **2007**, *114*, 1265–1275. [[CrossRef](#)]
51. Van Hintum, T.J.; Brown, A.; Spillane, C. *Core Collections of Plant Genetic Resources*; Bioversity International: Rome, Italy, 2000; p. 48.
52. Varshney, R.K.; Nayak, S.N.; May, G.D.; Jackson, S.A. Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol.* **2009**, *27*, 522–530. [[CrossRef](#)]
53. Gale, M.; Marshall, G.A. The chromosomal location of Gai 1 and Rht 1, genes for gibberellin insensitivity and semi-dwarfism, in a derivative of Norin 10 wheat. *Heredity* **1976**, *37*, 283–289. [[CrossRef](#)]
54. Murai, M.; Takamura, I.; Sato, S.; Tokutome, T.; Sato, Y. Effects of the dwarfing gene originating from ‘Dee-geo-woo-gen’ on yield and its related traits in rice. *Breed. Sci.* **2002**, *52*, 95–100. [[CrossRef](#)]
55. Vikram, P.; Franco, J.; Burgueño-Ferreira, J.; Li, H.; Sehgal, D.; Saint Pierre, C.; Ortiz, C.; Sneller, C.; Tattaris, M.; Guzman, C. Unlocking the genetic diversity of Creole wheats. *Sci. Rep.* **2016**, *6*, 23092. [[CrossRef](#)]
56. Marshall, D.; Brown, A. Optimum sampling strategies in genetic conservation. In *Crop Genetic Resources for Today and Tomorrow*; Frankel, O.H., Hawkes, J.G., Eds.; Cambridge University Press: Cambridge, UK, 1975; pp. 53–70.
57. Odong, T.; van Heerwaarden, J.; Jansen, J.; van Hintum, T.J.; van Eeuwijk, F. Statistical techniques for defining reference sets of accessions and microsatellite markers. *Crop Sci.* **2011**, *51*, 2401–2411. [[CrossRef](#)]
58. Reyes-Valdés, M.H.; Burgueño, J.; Singh, S.; Martínez, O.; Sansaloni, C.P. An informational view of accession rarity and allele specificity in germplasm banks for management and conservation. *PLoS ONE* **2018**, *13*, e0193346. [[CrossRef](#)]
59. Wen, W.; Franco, J.; Chavez-Tovar, V.H.; Yan, J.; Taba, S. Genetic characterization of a core set of a tropical maize race Tuxpeño for further use in maize improvement. *PLoS ONE* **2012**, *7*, e32626. [[CrossRef](#)]
60. Corrado, G.; Caramante, M.; Piffanelli, P.; Rao, R. Genetic diversity in Italian tomato landraces: Implications for the development of a core collection. *Sci. Hortic.* **2014**, *168*, 138–144. [[CrossRef](#)]
61. Informe Cereal de Invierno GENVCE 2018. Available online: [www.genvce.org/repositorio/informe-cereal-de-invierno-genvce-2018](http://www.genvce.org/repositorio/informe-cereal-de-invierno-genvce-2018) (accessed on 23 February 2020).
62. Map of Soil pH in Europe. Available online: <https://esdac.jrc.ec.europa.eu/content/soil-ph-europe> (accessed on 23 February 2020).
63. Zhang, X.K.; Xiao, Y.G.; Zhang, Y.; Xia, X.C.; Dubcovsky, J.; He, Z.H. Allelic variation at the vernalization genes Vrn-A1, Vrn-B1, Vrn-D1, and Vrn-B3 in Chinese wheat cultivars and their association with growth habit. *Crop Sci.* **2008**, *48*, 458–470. [[CrossRef](#)]
64. Cantalapiedra, C.P.; García-Pereira, M.J.; Gracia, M.P.; Igartua, E.; Casas, A.M.; Contreras-Moreira, B. Large differences in gene expression responses to drought and heat stress between elite barley cultivar scarlett and a spanish landrace. *Front. Plant Sci.* **2017**, *8*, 647. [[CrossRef](#)]

65. Monteagudo Gálvez, A.; Casas Cendoya, A.M.; Pérez Cantalapiedra, C.; Contreras-Moreira, B.; Gracia Gimeno, M.P.; Igartua Arregui, E. Harnessing novel diversity from landraces to improve an elite barley variety. *Front. Plant Sci.* **2019**, *10*, 434. [[CrossRef](#)]
66. Ruiz, M.; Zambrana, E.; Fite, R.; Sole, A.; Tenorio, J.L.; Benavente, E. Yield and quality performance of traditional and improved bread and durum wheat varieties under two conservation tillage systems. *Sustainability* **2019**, *11*, 4522. [[CrossRef](#)]



© 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/>).