

## Article

# Exploration of the Genetic Diversity of Solina Wheat and Its Implication for Grain Quality

Riccardo De Flaviis <sup>1</sup>, Giorgio Tumino <sup>2</sup>, Valeria Terzi <sup>3</sup>, Caterina Morcia <sup>3</sup>, Veronica Santarelli <sup>1</sup>,  
Giampiero Sacchetti <sup>1,\*</sup> and Dino Mastrocola <sup>1</sup>

<sup>1</sup> Faculty of Bioscience and Technology for Food, Agriculture and Environment, University of Teramo, Via R. Balzarini 1, 64100 Teramo, Italy; rdeflaviis@unite.it (R.D.F.); vsantarelli@unite.it (V.S.); dmastrocola@unite.it (D.M.)

<sup>2</sup> Plant Breeding, Wageningen University and Research, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands; giorgio.tumino@wur.nl

<sup>3</sup> Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria—Centro di Ricerca Genomica e Bioinformatica (CREA-GB), Via San Protaso 302, 29017 Fiorenzuola d'Arda, Italy; valeria.terzi@crea.gov.it (V.T.); caterina.morcia@crea.gov.it (C.M.)

\* Correspondence: gsacchetti@unite.it

**Abstract:** Different Solina wheat accessions ( $n = 24$ ) collected in the Abruzzo region (Italy) were studied using 45,000 SNP markers generated from the DArTseq platform. The structure of genetic data was analyzed by Principal Component Analysis and Hierarchical Cluster analysis that revealed the existence of two main clusters (Clu1 and Clu2) characterized by samples with different geographical origin. The Solina genetic dataset was further merged and analyzed with a public genetic one provided by CIMMYT containing 25,963 genotypes from all over the world. The Solina accessions occupied a vast space, thus confirming a high heterogeneity of this landrace that, nevertheless, is considerably unique and placed quite far from other clusters. Clu1 and Clu2 divergence were clearly visible. Solina clusters were genetically closer to landraces from Turkey and the central fertile crescent than to the Italian genotypes present in the dataset. Selected commercial quality traits of accessions of the two Solina clusters were analyzed (yield, thousand kernel weight, test weight, and protein content), and significant differences were found between clusters. The results of this investigation did not highlight any relationships of Solina with Italian genotypes, and confirmed its wide genetic diversity by permitting to identify two genetic groups with distinct origin and quality traits.

**Keywords:** landrace; local germplasm; single nucleotide polymorphism; DArTseq markers; population structure; commercial quality; genetic diversity; wheat relationships



**Citation:** De Flaviis, R.; Tumino, G.; Terzi, V.; Morcia, C.; Santarelli, V.; Sacchetti, G.; Mastrocola, D.

Exploration of the Genetic Diversity of Solina Wheat and Its Implication for Grain Quality. *Plants* **2022**, *11*, 1170. <https://doi.org/10.3390/plants11091170>

Academic Editor: Pasquale Tripodi

Received: 14 March 2022

Accepted: 25 April 2022

Published: 26 April 2022

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



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Solina is a historically cultivated landrace in Abruzzo (central Italy) and strongly linked with the territory and culture of mountain rural areas. It is not clear from where and when the Solina came out. It is surely possible to affirm that Solina is sown at least since the sixteenth century when it is mentioned in some notary deeds of sale and purchase stipulated at the Lanciano Fair [1].

From an agronomic point of view, Solina is very rustic, well suited to poor soils and those consisting of stone, very resistant to cold weather (it can resist for a long time covered by snow), and to the spring–summer drought, of limited productivity (on average 2 t/ha) but offering constant production; these traits favored its wide use on a territorial scale [2]. Seed circulation for such a long time span with mixing and selection, adjuvated by natural selection, could have promoted the recombination between different Solina genotypes and the fixation of neutral and adaptive mutations [3]. Despite its diversity, Solina is in fact a considerable Distinct and Stable landrace since it showed a phenotypic diverge similar to that of pure lines [4].

In the early twentieth century, Solina wheat has been extensively used in breeding programs carried out in Italy [5] but scientific interest has gradually grown only in recent years, as demonstrated by several recent publications [2–12]. Given the importance that such landrace constitutes both from a historical and economic standpoint [2,13,14], in the year 2007, it was inserted between the target landraces in a European study aimed at seed conservation, breeding, and production [3,4,9], thus it is of utmost importance to collect and genotype it.

The Green Revolution (1940–1960) allowed the doubling of wheat production, responding to the doubling of the world population in the last 40 years. The spread of highly productive varieties, together with the massive use of fertilizers and crop protection products, permitted a spectacular increase in agricultural production, in particular wheat and rice. At an environmental level, this intensive agriculture has been considered, in accordance with “the Borlaug hypothesis”, a tool to conserve the biodiversity of ecosystems, allowing land sparing in favor of natural environments [15].

Green Revolution and, in general, modern breeding, are involved in the erosion of genetic diversity in wheat: the use of a limited number of parental genotypes in breeding reduced the genetic basis of this crop, causing a genetic bottleneck. The genetic heritage represented by many old cultivars and landraces was put aside in favor of a few High Yielding Varieties (HYV), which were disease- and lodging-resistant, highly responsive to input, and characterized by Distinctness, Uniformity and Stability (DUS). Since modern cultivars were implemented primarily under modern agronomic practices using high agricultural inputs, several landraces (traditional crop varieties) remained protected and continuously cultivated in marginal areas, thanks also to their high adaptability in unfavorable growing conditions (e.g., wet, cold soils, high altitudes, and poor soils) [16].

In this context, landrace seeds’ management became typically an in situ on-farm conservation conducted by local farmers themselves, and at the same time, allowed to avoid genetic erosion and to obtain new alleles, deriving from casual mutation or from other neighboring populations [17]. Genetic diversity was preserved and increased by traditional and empirical agronomic practices used by local farmers, and nowadays, many scientific efforts are aimed to understand and exploit it [18].

In order to unravel the genetic structure and breeding history lying behind landraces, a genome-wide approach is needed; thanks to modern high-throughput technologies, such as DarTseq (Diversity Arrays Technology sequence), researchers can efficiently genotype thousands of genomic loci in parallel with a relative low cost [19]. Recently, DarTseq has been applied in complex polyploid crops, such as wheat, to find unexplored loci in landraces and elite wheats [20–24]. Using this technology, Sansaloni and co-workers [25] analyzed the genetic structure of 56,342 hexaploid wheat accessions with 85,531 single nucleotide polymorphism (SNP) markers, proving that the modern cultivars occupied only a little portion of the total genetic diversity space; thus landraces show a high unexplored biodiversity worthwhile of being studied and protected.

In this context, important is the work of Khan and co-workers [3] who studied European populations (in situ and ex situ varieties) of common wheat. Using 41 SNP markers located in neutral regions, the authors found that the genetic complexity increases from modern varieties to on-farm landraces, and that Solina wheat (Italic landrace) shows the highest genetic and haplotypic diversity among few European varieties [3]. The genetic diversity of Solina is difficult to explain for a self-pollinated species; the authors supposed that a very large effective size, caused by an on-farm collective management, could have led to this phenomenon. Seed circulation with mixing and selection, adjuvated by natural selection, had promoted recombination between different genotypes and the fixation of neutral and adaptive mutations.

Nevertheless, the pioneering studies above reported presented two major limitations: the number of SNPs analyzed and the number of accessions collected per landrace (farmer populations). According to Negri and co-workers [17], “on-farm conservation should be reformulated as the management of genetic diversity of locally developed crop varieties

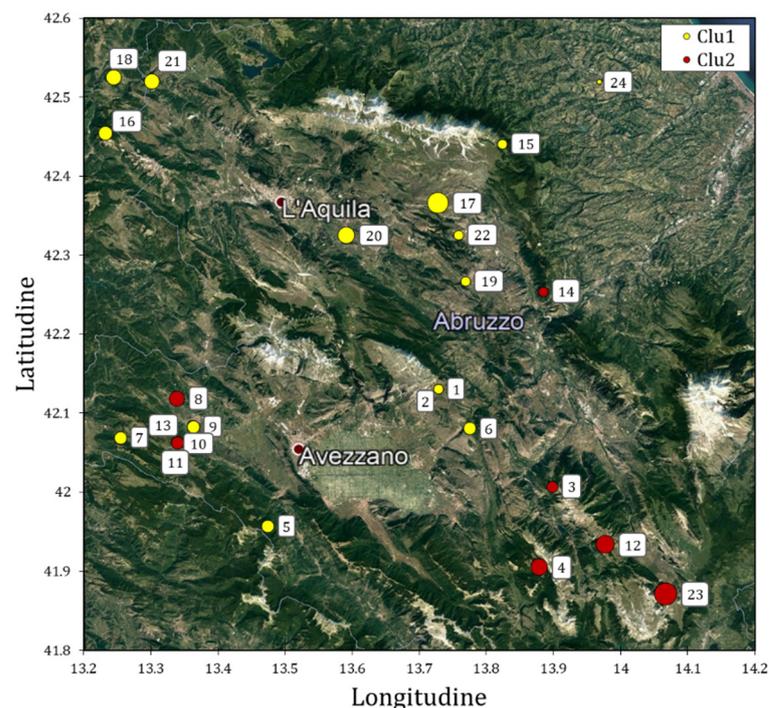
(landraces) by farmers within their own agricultural, horticultural or agri-silvicultural systems". Agricultural systems, understood as the set of agronomic practices and soil characteristics, can be different among different farms, potentially impacting on the genetic variability of the landrace. As a matter of fact, most studies in the field of on-farm landrace genetic characterization are more focused on within-accession diversity than in the between-farmer populations diversity. This study attempted to overcome these weaknesses, focusing on Solina wheat (*Triticum aestivum*, L.) and increasing the number of accessions collected and SNPs analyzed. Solina accessions, in situ conserved by different farmers, were characterized for both intra- and inter-farmer accessions genetic diversity, and also evaluated in relation with some quality parameters.

Moreover, in order to identify any genetic relationships within a vast common wheat germplasm and therefore shed light on the Solina history, the Solina genetic dataset was merged with a public genetic one (CIMMYT) obtained by 25,963 different common wheat genotypes from all over the world and further analyzed. To the authors' knowledge this is the first attempt to merge a landrace genotyping dataset with the international CIMMYT dataset to gain information on landrace genetic variability and relationships.

## 2. Materials and Methods

### 2.1. Grain Samples

Twenty-four accessions of Solina wheat were collected in several locations, representative of Solina cultivation area in the Abruzzo region (Supplementary Table S1). Figure 1 shows the Solina IDs projected on the Abruzzo region map, based on the collection sites.



**Figure 1.** Geographical location of 24 Solina accessions collected in 2018 in the Abruzzo region colored on the basis of their genetic clusters identified by DArT markers analysis (see Sections 3 and 4). The diameters of yellow and red circles are correlated to the altitude of the site of seed collection. Dark red dots with a white border represent the most important cities of the area.

The sampling was conducted in order to obtain the greatest number of accessions representative of the typical areal of Solina cultivation. The sites were distributed mainly in the province of L'Aquila (only two accessions were sampled in the province of Pescara); their average elevation was  $755 \pm 263$  m above sea level and they were characterized by different types of pedo-climatic conditions, distinctives of the Abruzzo region.

The accessions were sampled from grains harvested in the year 2018 in different farms in which Solina wheat was cultivated for at least ten years under organic management. In order to evaluate the genetic variability within accessions and among accessions, four single seeds and one seed bulk (50 seeds) were sampled per accession and named A, B, C, D, and E, respectively.

Seeds of accessions 2 and 3 (Supplementary Table S1) supplied by Az. Agricola Cipolla, (Castelvecchio Subequo, Italy) and Az. Agricola De Santis (Introdacqua, Italy), respectively, were used for field trials conducted in randomized plots in the Abruzzo region (Italy). The experimental fields were located in two different farms coded as F1 and F2 sited in Corropoli (42°49' N, 13°51' E, 70 m a.s.l.) and Castelvecchio Subequo (42°07' N, 13°44' E, 500 m a.s.l.), respectively, that presented different soil and climatic characteristics. The two accessions were cultivated in the 2018/2019 and 2019/2020 seasons. The sowing was done between the end of October and the beginning of December 2018 and 2019, in a 500 m length plot area. The seeding rate was of 250 kg/ha, under rainfed condition. Plot width was 50 m and the distance between plots was 1.5 m. Plants were grown under the same organic conditions, and no pesticides (bio- or chemical) and fertilizers (organic or chemical) were used. All the grains were harvested between the starting and the middle of July 2019 and 2020, depending on the elevation of the fields. The phytosanitary quality of crops was periodically controlled during the two years of cultivation and plants were healthy field-grown plants. The productivity of Solina wheat at 70 and 500 m a.s.l. was of about 20 q/ha in both years and was similar to those observed in the area of cultivation under organic regimes.

## 2.2. Genotyping

The samples were pre-treated following the instructions provided by Diversity Arrays Technology (DARt) Pty. Ltd. (<http://www.diversityarrays.com>, accessed on 1 December 2021, Canberra, Australia): single seeds were pounded separately avoiding DNA cross-contamination, and contemporary 50 seeds for every accession were milled; 96 pounded Seeds and 24 milled Bulk (20 mg per accession) were carefully placed in two 96-well plates and shipped with silica gels to DARt laboratories for SNP analysis (DNA extraction, DARtseq library preparation, sequencing on Illumina HiSeq2500/Novaseq using 2.5 million reads per sample, marker data extraction, and alignment to publicly accessible sequence of wheat (Wheat\_ChineseSpring20 and Wheat\_ConsensusMap\_version\_4)); 45,055 SNP markers were obtained and used for further statistical analyses.

## 2.3. Solina Dataset Pre-Processing and Filtering

In order to ease data structure representation and statistical elaboration, the DARt dataset (120 samples  $\times$  45,055 markers) was rearranged as trinary values (allelic dosage of alternate alleles) as follows: (i) "0" when only the reference allele homozygote is present; (ii) "1" when both the alleles are present (heterozygosis); (iii) "2" when only the SNP allele homozygote is present; (iv) "null" when both the alleles are absent (absence of fragment with SNP in genomic representation); "null" were considered as missing.

Subsequently, several quality SNP filters were applied to check uninformative or misinformative markers or genotypes: (i) exclusion of SNP markers and genotypes with more than 25% missing values; (ii) exclusion of SNP markers with less than 5% of Minor Allelic Frequency (MAF).

Remaining missing values were imputed by using the marker means.

## 2.4. Public CIMMYT Dataset Importing, Pre-Processing, and Filtering

The study of [25], implemented by CIMMYT (Centro Internacional de Mejoramiento de Maíz y Trigo) in collaboration with ICARDA, INIFAP, NIAB, EI, and DARt is a recent milestone in wheat genetics. Their huge amount of results are publicly available at <https://data.cimmyt.org/> (accessed on 10 January 2020) and were downloaded and rearranged (see Section 2.3) to be merged with the single Seed dataset. From the hexaploid CIMMYT dataset

were imported 9670 SNP markers (only those in common with the single Seeds dataset, using both *AlleleID* and *AlleleSequence* as marker identifiers) and 25,963 genotypes of cultivars and landraces of *Triticum aestivum* L. *aestivum*, which in combination with 96 single Seeds produced a unique matrix of 26,059 genotypes with trinary values. Since significant differences in heterozygosity have been seen between CIMMYT and Solina genotypes, stricter filters were used: (i) exclusion of SNP markers with more than 10% missing values; (ii) exclusion of genotypes with more than 20% missing values; (iii) exclusion of genotypes with more than 10% heterozygosity; (iv) exclusion of SNP markers with less than 5% of MAF.

### 2.5. Commercial Quality

Phenotyping of three commercial quality parameters was done for every accession in duplicate. The total nitrogen content (NC) was determined using the Kjeldahl method (AACC 46-12.01): the sample is digested in sulfuric acid, ammonia is distilled, and then the distillate is titrated with hydrochloric acid. A conversion factor of 5.7 was used for wheat. Test weight (TW) and thousand-kernel weight (TKW) were analyzed after removing all impurities. Moisture content was also analyzed (AACC 44-15.02).

### 2.6. Statistical Analysis

Statistical analyses of genetic data were carried out using R 4.0.4 (R Core Team 2020) and several specific packages. Missing value replacement was done using *na.mean* function (R package *imputeTS*). SNPs density plot were obtained thanks to *CMplot* function (R package *CMplot*). Principal Component Analysis (PCA) based on covariance matrix was computed by the *prcomp* function (R package *stats*), whereas Hierarchical Clustering Analysis (HCA) was performed using Euclidean distances and average-linkage agglomeration method implemented in the *hclust* function (R package *stats*). Moreover, PC1 scores (obtained with single seed dataset) were used in multiple linear regression (MLR) with geographic coordinates (latitude longitude and altitude) as explanatory variables; thus  $\beta$  coefficients were calculated in order to compare the relative weights of the variables.

Statistical analysis of phenotypic data was performed using XLSTAT 2021 software (Addinsoft, Paris, France). Mixed nested ANOVA was carried out on phenotypic data from the 24 accessions of Table S1 by considering the Clu1/Clu2 as fixed effect and the farm of collection as the random nested effect. Repeated measurement ANOVA was carried out on data of accessions 2 and 3 (representative of Clu1 and Clu2, respectively) cultivated in farm F1 and F2 for two consecutive years, and plot was used as random effect.

## 3. Results

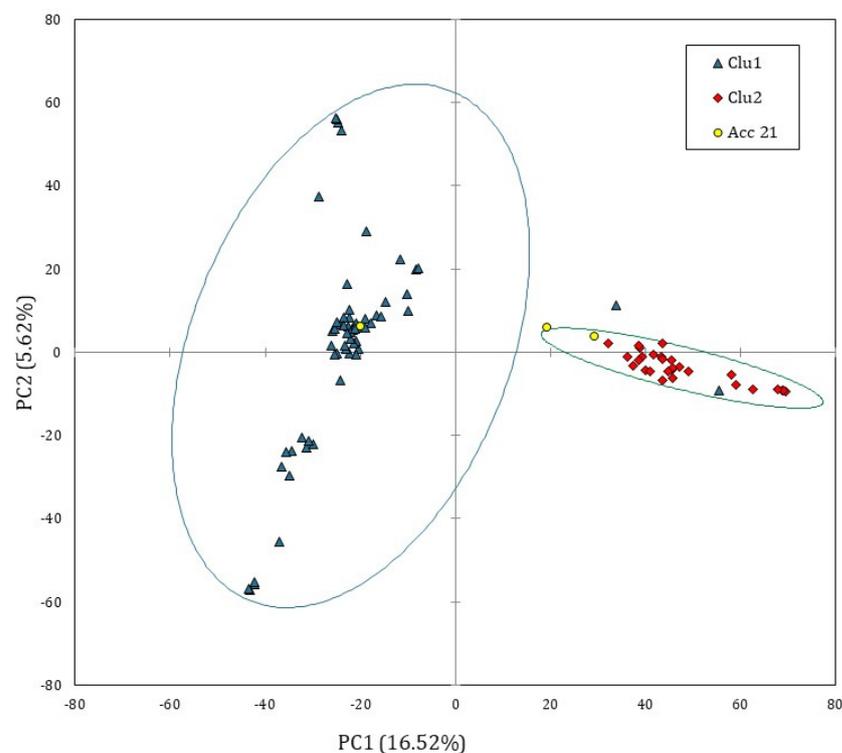
### 3.1. Solina Datasets: DArTseq Marker Characteristics

DArTseq analysis permitted to detect approximately 45 K SNPs for each of the 120 Solina samples investigated. Original Solina dataset was initially divided into Bulk (24 genotypes) and single Seed (96 genotypes) matrices, and were studied always separately. Bulk genotypes may be comparable to a genetic superimposition of 50 seeds, tested to evaluate a larger sample size, and so, the within-accession variability [26]. After the application of several quality check filters, the datasets were reduced to 23,741 SNP markers and 24 genotypes for Bulk dataset, and to 15,959 SNP markers and 93 genotypes for single Seed dataset. Three seeds from three different accessions (20A, 21D, and 24A) were removed from the single Seed dataset since they had more than 25% of missing values. On the other hand, 13,181 and 13,719 markers for Bulk and single Seeds datasets were considered rare alleles and discarded (MAF < 5%). On average, the filtered datasets had 6.5% and 6.7% of null, and 40.4% and 12.8% of heterozygosity for Bulk and single Seeds, respectively (Supplementary Figure S1). Only one Bulk genotype (20E) had less than 25% of heterozygosity whereas only two Seeds (17C and 18D) had more. The sample 20E was also the unique Bulk with more than 15% of missing values (but less than 25%) and, in the same way, the individuals 15A and 14A for single Seeds dataset. Among genomes, the B was always the most represented with 7650 and 5279 markers for Bulk and single Seed

filtered datasets followed by A genome (7385 and 5130) and D genome (4277 and 3055) (Supplementary Table S2). However, several markers (4429 and 2495) were still unassigned to any of the 21 wheat chromosomes. The density plots in Supplementary Figure S2a,b confirmed the wide genome representation for each chromosome in both datasets. The highest SNPs density was always observed in chromosome terminations whereas the lowest was near the centromeres, as already observed by [22]. A clear difference is present between single Seed and Bulk in marker density, as a consequence of different number of markers selected by filters. The homoeologous chromosomes of the fourth group were the least represented and, in fact, only the chromosomes 4D had less than 300 markers for both datasets (Supplementary Table S2).

### 3.2. Solina Population Structure

In order to explore the genomic structure of Solina accessions, a PCA based on covariance matrix was computed only on single Seed dataset; since *prcomp* function is not able to process missing values, they were substituted by marker means. Figure 2 shows the score plot of 93 Solina seeds using the first two principal components (PCs) which explained the 22.14% of the total variance, 16.52% along the PC1 and 5.62% along the PC2.



**Figure 2.** PCA score plot computed by using single Seed dataset (15,959 SNP markers and 93 genotypes). The ellipsis indicate the 95% confidence limit for each Solina cluster. The accession 21 was considered an outlier.

Despite the small, explained variance, the PC1 permitted to identify two main and distinct Solina clusters (Clu1 and Clu2), as highlighted by confidential ellipses in Figure 2. Clu1 is composed by 16 farmer accessions (1, 2, 5, 6, 7, 9, 10, 13, 15, 16, 17, 18, 19, 20, 22, and 24), while in Clu2 by 7 (3, 4, 8, 11, 12, 14, 23), only two seeds of Clu1 from two different accessions, 20B and 13D, deviate from this classification. The accession 21 remained ambiguous; indeed, one seed (21C) was located exactly between the two clusters whereas another seed (21B) was similar to other genotypes in Clu2. Differently, PC2 mostly described genetic variation within the Clu1 without showing further net separation of samples on the basis of their genetic traits. Similar results were obtained with HCA based on Euclidean distances and average-linkage agglomeration method, supporting the two

Solina clusters hypothesis. Moreover, Euclidean distances revealed considerable similarities between accessions (at least within Clu1 or Clu2) but lower within accessions, as shown by the dendrogram in Figure 3.

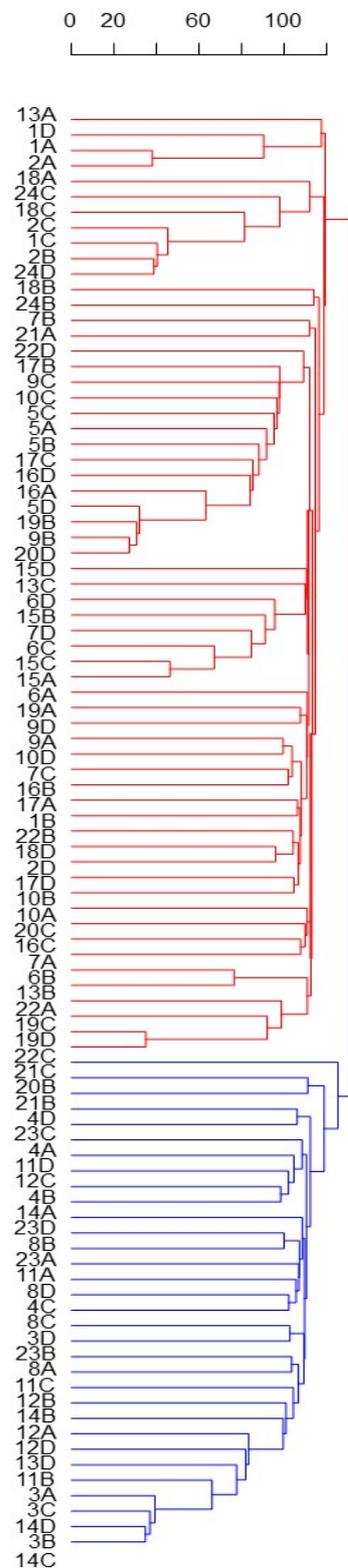
PC1 scores were further used in multiple linear regression (MLR) using latitude, longitude, and altitude (Supplementary Table S1) as explanatory variables. The model showed an  $R^2_{\text{adj}}$  value of 0.231 ( $p < 0.001$ ), and significant  $\beta$ -coefficients ( $p < 0.05$ ) were found with the following values:  $-0.31$ ,  $0.25$ , and  $0.19$ . Clu2, characterized by positive scores in PC1, which tends to localize in the south-eastern area of Abruzzo region, that on average shows a higher altitude value.

### 3.3. CIMMYT/Solina Dataset: DArTseq Marker and Population Structure

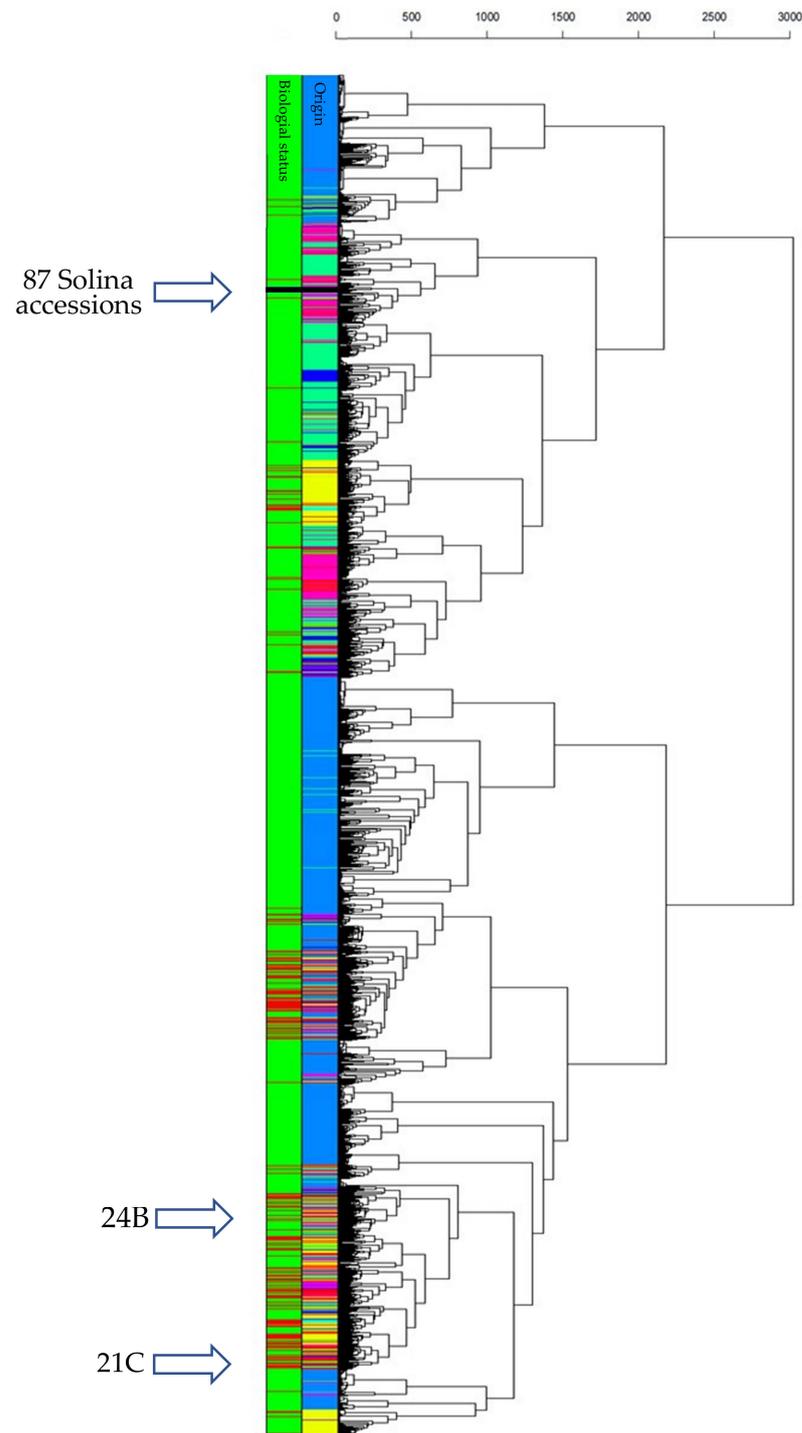
To better analyze the genetics of Solina accessions in an international and multi-varieties context, a public SNP database provided by CIMMYT [25] was imported; 9670 markers and 26,059 genotypes of common wheat assemble the CIMMYT/Solina merged dataset that, after several and more strict filters, was reduced to 3842 markers and 24,868 genotypes of which 91 were Solina seeds. Three Solina seeds (20A, 21D, and 24A) were discarded once again (see Section 3.2), but also 260 CIMMYT genotypes with more than 20% of missing values were removed. Regarding markers, 4872 with more than 10% of missing values and 956 with less than 5% of MAF were omitted from the final dataset. The heterozygosity filter was applied to uniform the heterozygosity between genotypes in the two datasets (initially higher in CIMMYT dataset), and to discard several genotypes that resulted very similar towards anyone else. Thus, 928 CIMMYT genotypes and two Solina seeds (18D and 17C) were removed since they had more than 10% of heterozygosity. The chromosomal distributions of markers follow the one previously described for Solina datasets, although with a reduced number of markers (Supplementary Table S2 and Supplementary Figure S2c).

The filtered merged dataset is constituted of 3058 cultivars and 21,719 landraces; the most represented nations were Mexico (10,614), Iran (3695), and China (2198). Only 28 genotypes were Italian.

To understand how the Solina accessions related with CIMMYT populations, PCA and HCA were computed. Since this dataset was larger than previous ones, a clearer 3D score plot was depicted and shown in Supplementary Figure S3. The first three PCs explained an overall variance of 19.32% (9.10%, 5.35%, and 4.87% for PC1, PC2, and PC3 respectively), showing 24,868 data points corresponding to just as many genotypes. Thanks to confidence 3D ellipsoids and colors point, an intuitive picture of the dataset can be made; the genetic structure depicted by PCA offered interesting information about worldwide genetic relationships and correlations of common wheat that were not further discussed in this work (for an in-depth discussion see [25]). However, landrace genetic diversity was considerably higher than cultivars that were located very close to each other within a portion of the space. The Solina accessions occupied a vast 3D space considerably unique, quite far from other clusters. Despite the limited number of markers, Clu1 and Clu2 divergence is clearly visible along the PC3, which is correlated to the protein content ( $r = 0.594$ ) according to the CIMMYT phenotypic dataset. Going in-depth, a HCA provided further information about genotype linkages through Euclidean distances (Figure 4). In particular, average-linkage agglomeration method classified 89 Solina seeds together at 250 Euclidean distances, while two seeds of two different accessions (21C and 24B) diverge considerably starting from the first dendrogram branch and clustering near to a Switzerland genotype (Blanc-Precoce) and to Baionette I (a cultivar obtained by Nazareno Strampelli, breeding Rieti  $\times$  Prince Albert), respectively. Clu1 and Clu2 seeds join together at 125 and 150 Euclidean distances, respectively, whereas the first CIMMYT genotypes are found at 410 Euclidean distances. Considering the first genetic outsiders, they are 545 landraces and 12 cultivars of which 294 (45%) genotypes are Turkish, 59 (11%) are Iranian, and 43 (8%) are Syrian. The geographic origin of 119 (21%) neighbor genotypes are unknown. No Italian genotypes are met in the dendrogram up to a Euclidean distance of 1710.



**Figure 3.** HCA based on Euclidean distances and average-linkage agglomeration method computed by using single Seed dataset (15,959 SNP markers and 93 genotypes). Labels indicate the number ID of the 24 Solina accessions, and letters correspond to the four different seeds (A, B, C, D) analyzed. Two Solina clusters were highlighted by different colors.



**Figure 4.** HCA based on Euclidean distances and average-linkage agglomeration method computed by using CIMMYT/Solina dataset (3842 markers and 24,868 genotypes of which 91 were Solina seeds). Two rows of labels indicate wheat origin (95 colors for each country: dodger blue for Mexico, spring green for Iran, deep pink for Turkey, yellow for China, black for Solina) and biological status (green for CIMMYT landraces; red for CIMMYT cultivars; black for Solina genotypes).

### 3.4. Commercial Quality Parameters

TWK, TW, and proteins were analyzed on the Solina sample set used for DArTseq analysis. Table 1 summarized the ANOVA results. Within the 24 Solina accessions, TWK and TW of the two Solina genetic clusters were significantly different ( $p < 0.05$ ), even though the nested farm effect resulted always in the main effect ( $p < 0.001$ ). No significant

differences among clusters were observed for protein content which showed an extremely high variability (from 9.66 to 15.48% <sub>d.b.</sub>).

**Table 1.** Significant *p*-values from ANOVA F-tests, for each phenotypical trait analyzed from Solina samples collected in 2018 in Abruzzo.

Effect	Type	TKW	TW	Protein %
Clu1/Clu2	Fixed	0.018	0.017	ns
Farm (Clu1/Clu2)	Random	<0.001	<0.001	<0.001

The effect of genetic cluster on TWK and TW was confirmed by the in situ experiment carried out by comparing two Solina accessions cultivated in two different farms over two years, and this experiment highlighted also a significant difference in the protein content of the two accessions (Table 2). On the other hand, the main effect for experimental field data diverged for each variable: farm effect was predominantly in TWK whereas the year effect in TW. Plot was always not significant ( $p > 0.05$ ).

Clu1 was the cluster with the highest TWK and TW in both the sample set used for DArTseq analysis and the sample set obtained by the in situ experiment. In the latter experiment, Clu1 resulted in the genetic cluster with the highest value of protein content. This result reflects the highest score values of Clu1 along the PC3 calculated using the CIMMYT phenotypic dataset (see Section 3.3).

**Table 2.** Significant *p*-values from ANOVA F-tests, for each trait analyzed from in situ experiment samples conducted in 2019 and 2020.

Effect	Type	TKW	TW	Protein %	Yield
Plot	Random	ns	ns	ns	ns
Year	Fixed	ns	<0.001	ns	0.011
Farm	Fixed	<0.001	ns	ns	<0.001
Clu1/Clu2	Fixed	0.003	0.044	0.003	ns

#### 4. Discussion

The current study aimed at obtaining valuable information on Solina wheat, a historically Italian landrace, adapted to the mountain environment of central Italy; as a matter of fact, characterizing genetic population structures is a fundamental step towards landrace protection, conservation and proper utilization in breeding program, agronomic management, and product transformation. The challenge was to draw out useful results from a consistent genetic diversity, typical of landraces [3], which makes it difficult to set clear boundaries around these genotypes.

##### 4.1. Population Structure Analysis

DArTseq analysis permitted to deeply investigate the Solina genome. As previously stated, wide genetic diversity is expected by local and historical landraces, and for this reason also, a bulk was considered in the analysis. The higher heterozygosity in Bulk dataset (Supplementary Figure S1) was the result of 50 milled individuals and of an evident genetic within-accession diversity; this confirms previous results reported by Khan and co-workers [3].

Even though the final datasets were significantly reduced, high-quality SNP markers have remained widely distributed abroad the three genomes, as shown in Supplementary Table S2. Genome D and the homoeologous chromosomes of the fourth group were characterized by the lowest number of markers; this result is in accordance with those of other authors [22–24] who analyzed Spanish, Iranian, and Afghan common wheat landraces with the same DArTseq approach. The limited diversity of D genome may be originated from the “recent” introgression occurred with *Aegilops tauschii* ( $2n = DD$ ), the common wheat D-genome donor, and the relative genetic bottleneck [27].

Since SNP datasets are matrices with a great imbalance in variables/cases ratio, they were analyzed through an unsupervised multivariate approach to reduce dimensionality and better visualize huge datasets. A PCA based on covariance matrix and HCA based on average-linkage agglomeration method permitted to identify clearly two genetic clusters (Clu1 and Clu2). This is clearly shown in the score plot in Figure 2, where a clear separation is visible along the first two principal components, despite that the genotype 21, together with 20 and 13, in minor extent, was revealed to be very heterogenous. A possible explanation for this might be due to the seed management of that specific farmer. Accession 21 could be a landrace mixture of distinct local varieties, Solina and Baionetta included, and accessions 20 and 13, arguably, could be a mixture of Solina(s) Clu1 and Clu2.

Recent works in this field demonstrated how geography was the main effect driving the genetic diversity [22,24], even though others did not find any geographical effect on genetics [20,23]. In this study,  $\beta$ -coefficients, calculated by MLR, clearly demonstrated how a geographic effect existed between Clu1 and Clu2. It is widely known that the environment, especially those with low resources and harsher climate, affects the genome of plants, and leads to genetic adaptation and stability [28]. As a consequence, the genetic difference observed between the two Solina clusters may be originated from different abiotic stresses due to different pedo-climatic characteristics of cultivation areas.

Solina wheat is isolated from all the worldwide cultivars and landraces reported in the CIMMYT database so it could be considered as a distinct landrace but it shows a wide genetic variance, even stressing the results reported by Khan et al. [3] on a single accession.

In an attempt to clarify a possible historical origin of Solina clusters or to find an event that significantly affected its genetic diversity, an analysis of the first genetic outsiders was carried out and revealed that Solina is located neither near nor within the Italian genotypes cluster, and is very close to landraces from the central fertile crescent (mainly Turkish), which is the major center of wheat domestication.

It has been hypothesized that *Triticum aestivum* first arose in Turkey, in which earliest records dated 8600–7800 before present, have been found. However, it is even possible that several allopolyploid speciation processes occurred, still carrying several unanswered questions [29]. After speciation, a continuous process of diversification resulted in the development of numerous landraces with fine adaptation to different agroecological environments, and spreading of common wheat in humid and cold environments of Central and Northern Europe.

The genetic similarity found among Solina and Fertile Crescent germplasm arises the question if Solina can be considered an allochthonous or an autochthonous landrace [30]. Zeven [31] defined autochthonous “a landrace grown for a long period in the farming system concerned. As the environment changes annually and as the landrace becomes ‘contaminated’ with few genotypes of other landrace(s), or cultivar(s) it will continuously adapt itself”. On the contrary, “allochthonous landrace is an autochthonous landrace of a foreign region recently introduced into the region concerned. Depending on the number of generations of after growth and on the frequency of seed change, it may become an autochthonous landrace”. Starting from the facts that Solina is an historical and dynamic population, with distinct identity, very well adapted to the Abruzzo agroecological conditions, it can be categorized as an Italian autochthonous landrace that still maintains strong genetic relationships with the landraces of the major ancient center of diversification of common wheat.

#### 4.2. Commercial Quality Parameters

Phenotypic variation within the Solina landrace was already observed by Bonvicini [13], who, several decades ago, collected 53 Solina accessions and categorized them in 15 groups based on the period of earing, height plant, resistance to rust, and color of the spikes. Noteworthy is that Bonvicini selected two performing Solina lines based on rust resistance and capacity of constant production [13]. What happened to these two selected lines is not known, but it is reasonable to suppose that, in a historical period in which Italy closed the customs and depended only on its own stocks of grain, they were diffused among

local farmers in order to comply with national policies that encouraged the use of varieties having the characteristics selected by Bonvicini [13]. This hypothesis is supported by the fact that, nowadays, constant production is a distinctive characteristic of the Solina landrace that grants its appreciation by local farmers [2].

Even though it was not in the purpose of this work to verify if the two different genetic clusters identified in this study resulted from an adaptation of the two lines selected by Bonvicini [13], it was of interest to investigate if phenotypic variations among Clu1 and Clu2 in terms of yield, TW, TKW, and protein content occur. Yield, TW, and protein are of utmost economical interest for local Solina producers; this is because TW and protein content are grade-determining factors for the commercial quality of wheat. Moreover, protein content is an essential criterion for breadmaking characteristics, since higher protein generally produce doughs more resistant to overworking during mixing and breads with greater loaf volume.

Clu1 that was collected in a wide area characterized by a mean altitude of 703 m showed higher TW, TKW, and protein contents than Clu2 that was collected in a restricted area characterized by a mean altitude of 883 m and located in the south-east of the sampling area. Differences in TW and TKW were observed in both the seeds set collected in 2018 and in the samples collected in the in situ experiments conducted in 2019 and 2020, whilst protein content resulted significantly different only in the in situ experiments. This result is due to the fact that the samples collected in 2018 were cultivated in different geographical areas (Figure 1), pedo-climatic conditions, and under different agronomic management conditions, thus presenting an extremely broad range of protein content. These results are in accordance with literature, in fact, protein content is reported to be heavily affected by rainfall, temperature, and soil fertility, which makes environment the most influential effect, even greater than genotype or interaction environment  $\times$  genotype on protein content [32,33].

## 5. Conclusions

The results of this investigation highlighted the genetic relationships of Solina with Turkey and Middle East landraces, and confirmed its wide genetic diversity by permitting to identify two principal genetic groups with distinct origin and quality traits. These findings constitute a first step towards protection and conservation of Solina genetic heritage, but could also have implications for its agronomic management and product end-use. Further experiments (e.g., characterization of frost tolerance, vernalization, and photoperiod genes) need to be carried out in order to study the origin of this genetic differentiation, probably due to adaptation phenomena caused by environment pressures. Despite its between accessions diversity, the Solina wheat remains genetically distant from all the worldwide cultivars and landraces reported in the CIMMYT database, thus it could be also considered as a unique genetic resource for breeding programs.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/plants11091170/s1>, Figure S1: Plot of missing values (%) and heterozygosity (%) for Bulk and single Seeds dataset, Figure S2: Marker density plot for Bulk (a), Single Seed (b), and CIMMYT/Solina (c) datasets along the 21 chromosomes, Figure S3: PCA 3D score plot computed by using CIMMYT/Solina dataset (3842 markers and 24,868 genotypes of which 91 were Solina seeds). The ellipsoids indicate the 95% confidence limit for each Solina cluster. Colors indicate biological status: CIMMYT landraces (green spheres); CIMMYT cultivars (red spheres); Solina genotypes (blue spheres), Table S1: Geographical location of 24 Solina accessions collected in 2018 in Abruzzo region, Table S2: Numbers of SNP markers detected and filtered in the three datasets, also ordered per genome and chromosome.

**Author Contributions:** Conceptualization, R.D.F., G.T., V.T. and G.S.; Data curation, R.D.F., G.T. and G.S.; Formal analysis, R.D.F., G.T., V.S. and G.S.; Funding acquisition, D.M.; Investigation, R.D.F., V.S. and G.S.; Methodology, R.D.F., V.S. and G.S.; Project administration, G.S.; Software, R.D.F. and G.T.; Supervision, G.T. and G.S.; Visualization, R.D.F., G.T., V.T. and G.S.; Writing—original draft, R.D.F. and G.S.; Writing—review and editing, V.T., C.M., G.S. and D.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** The authors wish to acknowledge that this work was part of the Project DM 61316 ‘Miglioramento della produzione e della trasformazione dei frumenti tipici autoctoni’ (Improvement of production and processing of autochthonous and typical wheats) coordinated by Consorzio di Ricerca per l’Innovazione Tecnologica, la Qualità e la Sicurezza degli Alimenti (Research Consortium for Technological Innovation, Quality and Safety of Foods) of the Abruzzo region and funded by MIUR (Ministry of Instruction, University and Research, Italy).

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** This manuscript is dedicated to the late Antonio Michele Stanca, pioneer of modern cereal science in Italy, and passionate promoter of this project.

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

## References

- Manzi, A. Cereali. In *Origine e Storia Delle Piante Coltivate in Abruzzo*; Casa Editrice Rocco Carabba: Lanciano, Italy, 2006; pp. 23–42.
- Porfiri, O.; Silveri, D. La “Solina” e altre varietà locali di cereali ancora coltivate in Abruzzo: I risultati di una campagna di collezione e caratterizzazione promossa dall’ARSSA. In *Atti VI Convegno Nazionale Biodiversità*; Tecnomack: Bari, Italy, 2004; pp. 843–852.
- Khan, A.R.; Goldringer, I.; Thomas, M. Management Practices and Breeding History of Varieties Strongly Determine the Fine Genetic Structure of Crop Populations: A Case Study Based on European Wheat Populations. *Sustainability* **2020**, *12*, 613. [[CrossRef](#)]
- Dawson, J.C.; Serpolay, E.; Giuliano, S.; Schermann, N.; Galic, N.; Chable, V.; Goldringer, I. Multi-trait evolution of farmer varieties of bread wheat after cultivation in contrasting organic farming systems in Europe. *Genetica* **2012**, *140*, 1–17. [[CrossRef](#)] [[PubMed](#)]
- Piergiorganni, A.R. Evaluation of genetic variation and grain quality of old bread wheat varieties introduced in north-western Italian environments. *Genet. Resour. Crop Evol.* **2013**, *60*, 325–333. [[CrossRef](#)]
- Porfiri, O.; Torricelli, R.; Silveri, D.D.; Papa, R.; Barcaccia, G.; Negri, V. The Triticeae genetic resources of central Italy: Collection, evaluation and conservation. *Hereditas* **2004**, *135*, 187–192. [[CrossRef](#)] [[PubMed](#)]
- Osman, A.; Chable, V. Inventory of initiatives on seeds of landraces in Europe. *J. Agric. Environ. Ternational Dev.* **2009**, *103*, 95–130.
- Serpolay-Besson, E.; Dawson, J.C.; Chable, V.; Van Bueren, E.L.; Osman, A.; Pino, S.; Silveri, D.; Goldringer, I. Diversity of different farmer and modern wheat varieties cultivated in contrasting organic farming conditions in western Europe and implications for European seed and variety legislation. *Org. Agric.* **2011**, *1*, 127–145. [[CrossRef](#)]
- Dawson, J.C.; Serpolay, E.; Giuliano, S.; Schermann, N.; Galic, N.; Berthelot, J.-F.; Chesneau, V.; Ferte, H.; Mercier, F.; Osman, A.; et al. Phenotypic diversity and evolution of farmer varieties of bread wheat on organic farms in Europe. *Genet. Resour. Crop Evol.* **2012**, *60*, 145–163. [[CrossRef](#)]
- Durazzo, A.; Casale, G.; Melini, V.; Maiani, G.; Acquistucci, R. Total polyphenol content and antioxidant properties of Solina (*Triticum aestivum* L.) and derivatives thereof. *Ital. J. Food Sci.* **2016**, *28*, 221.
- Tavoletti, S.; Foligni, R.; Mozzon, M.; Pasquini, M. Comparison between fatty acid profiles of old and modern varieties of *T. turgidum* and *T. aestivum*: A case study in central Italy. *J. Cereal Sci.* **2018**, *82*, 198–205. [[CrossRef](#)]
- 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](#)]
- Bonvicini, M. *Genealogical Selection of the Wheat “Solina”*; Istituto di Allevamento Vegetale per la Cerealicoltura: Bologna, Italy, 1936.
- Perrino, P.; Laghetti, G.; Hammer, K. Collection of plant genetic resources in Italy, 1987. *Genet. Resour. Crop Evol.* **1988**, *36*, 377–390. [[CrossRef](#)]
- Salles, J.-M.; Teillard, F.; Tichit, M.; Zanella, M. Land sparing versus land sharing: An economist’s perspective. *Reg. Environ. Chang.* **2017**, *17*, 1455–1465. [[CrossRef](#)]

16. Boukid, F.; Folloni, S.; Sforza, S.; Vittadini, E.; Prandi, B. Current trends in ancient grains-based foodstuffs: Insights into nutritional aspects and technological applications. *Compr. Rev. Food Sci. Food Saf.* **2018**, *17*, 123–136. [[CrossRef](#)]
17. Negri, V.; Maxted, N.; Veteläinen, M. European Landrace Conservation: An Introduction. In *European Landraces: On Farm Conservation, Management and Use*; Biodiversity Technical Bulletin no. 15; Veteläinen, M., Negri, V., Maxted, N., Eds.; European Cooperative Programme for Plant Genetic Resources: Rome, Italy, 2009.
18. 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](#)] [[PubMed](#)]
19. Akbari, M.; Wenzl, P.; Caig, V.; Carling, J.; Xia, L.; Yang, S.; Uszynski, G.; Mohler, V.; Lehmensiek, A.; Kuchel, H.; et al. Diversity arrays technology (DARt) for high-throughput profiling of the hexaploid wheat genome. *Theor. Appl. Genet.* **2006**, *113*, 1409–1420. [[CrossRef](#)]
20. Baloch, F.S.; Alsaleh, A.; Shahid, M.Q.; Çiftçi, V.; de Miera, L.E.S.; Aasim, M.; Nadeem, M.A.; Aktaş, H.; Özkan, H.; Hatipoğlu, R. A Whole Genome DARtseq and SNP Analysis for Genetic Diversity Assessment in Durum Wheat from Central Fertile Crescent. *PLoS ONE* **2017**, *12*, e0167821. [[CrossRef](#)]
21. Edet, O.U.; Gorafi, Y.S.; Nasuda, S.; Tsujimoto, H. DARtseq-based analysis of genomic relationships among species of tribe Triticeae. *Sci. Rep.* **2018**, *8*, 16397. [[CrossRef](#)]
22. 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](#)]
23. Mahboubi, M.; Mehrabi, R.; Naji, A.M.; Talebi, R. Whole-genome diversity, population structure and linkage disequilibrium analysis of globally diverse wheat genotypes using genotyping-by-sequencing DARtseq platform. *3 Biotech* **2020**, *10*, 48. [[CrossRef](#)]
24. Tehseen, M.M.; Istiqlil, D.; Kehel, Z.; Sansaloni, C.P.; da Silva Lopes, M.; Kurtulus, E. Genetic diversity and population structure analysis of *Triticum aestivum* L. Landrace panel from Afghanistan. *Genes* **2021**, *12*, 340. [[CrossRef](#)]
25. Sansaloni, C.; Franco, J.; Santos, B.; Percival-Alwyn, L.; Singh, S.; Petroli, C.; Campos, J.; Dreher, K.; Payne, T.; Marshall, D.; et al. Diversity analysis of 80,000 wheat accessions reveals consequences and opportunities of selection footprints. *Nat. Commun.* **2020**, *11*, 4572. [[CrossRef](#)] [[PubMed](#)]
26. Tumino, G.; Voorrips, R.E.; Rizza, F.; Badeck, F.-W.; Morcia, C.; Ghizzoni, R.; Germeier, C.U.; Paulo, M.J.C.; Terzi, V.; Smulders, M.J. Population structure and genome-wide association analysis for frost tolerance in oat using continuous SNP array signal intensity ratios. *Theor. Appl. Genet.* **2016**, *129*, 1711–1724. [[CrossRef](#)] [[PubMed](#)]
27. Chao, S.; Dubcovsky, J.; Dvorak, J.; Luo, M.-C.; Baenziger, S.P.; Matnyazov, R.; Clark, D.R.; Talbert, L.; Anderson, J.; Dreisigacker, S.; et al. Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). *BMC Genom.* **2010**, *11*, 727. [[CrossRef](#)]
28. Chapin, F.S.; Autumn, K.; Pugnaire, F.I. Evolution of Suites of Traits in Response to Environmental Stress. *Am. Nat.* **1993**, *142*, S78–S92. [[CrossRef](#)]
29. Matsuoka, Y. Evolution of polyploidy *Triticum* wheats under cultivation: The role of domestication, natural hybridization and allopolyploid speciation in their diversification. *Plant Cell Physiol.* **2011**, *52*, 750–764. [[CrossRef](#)]
30. Casañas, F.; Simó, J.; Casals, J.; Prohens, J. Toward an evolved concept of landrace. *Front. Plant Sci.* **2017**, *8*, 145. [[CrossRef](#)]
31. Zeven, A.C. Landraces: A review of definitions and classifications. *Euphytica* **1998**, *104*, 127–139. [[CrossRef](#)]
32. Williams, R.M.; O'Brien, L.; Eagles, H.A.; Solah, V.; Jayasena, V. The influences of genotype, environment, and genotype × environment interaction on wheat quality. *Aust. J. Agric. Res.* **2008**, *59*, 95–111. [[CrossRef](#)]
33. Ashraf, M. Stress-Induced Changes in Wheat Grain Composition and Quality. *Crit. Rev. Food Sci. Nutr.* **2014**, *54*, 1576–1583. [[CrossRef](#)]