Genomic Analysis of Spanish Wheat Landraces Reveals Their Huge Potential for Breeding

CURRENT STATUS: ACCEPTED

Laura Pascual
Universidad Politécnica de Madrid

Magdalena Ruiz
Plant Genetic Resources Center. INIA

Matilde López-Fernández
Universidad Politécnica de Madrid

Helena Pérez-Peña
Universidad Politécnica de Madrid

Elena Benavente
Universidad Politécnica de Madrid

José Francisco Vázquez
Universidad Politécnica de Madrid

Patricia Giraldo
Universidad Politecnica de Madrid

patricia.giraldo@upm.es Corresponding Author
ORCiD: https://orcid.org/0000-0003-4369-1078

DOI: 10.21203/rs.2.9775/v1

SUBJECT AREAS
Epigenetics & Genomics

KEYWORDS
Wheat landraces, GBS, DArTseq markers, SNP, genetic diversity, population structure, breeding
Abstract

Background

One of the main goals for the XXI century breeding is the development of crop cultivars that can maintain current yields under unfavorable environments. Landraces that have been grown under varied local conditions include genetic diversity that will be essential to achieve this objective. The Center of Plant Genetic Resources of the Spanish Institute for Agriculture Research (CRF-INIA) holds a wide collection of wheat landraces. These accessions, locally adapted to a really wide diversity of eco-climatic conditions, represent a highly valuable material for breeding. However, their efficient use requires an exhaustive genetic characterization. The overall aim of this study was to assess the diversity and population structure of a selected set of 380 Spanish landraces and 52 reference varieties of bread and durum wheat by high-throughput genotyping.

Results

DArTseq GBS approach generated 10K SNPs and 40K DArT high-quality markers that were mapped against the currently available bread wheat reference genome. The markers with known location were distributed in all the chromosomes, having a relatively well-balanced genome-wide coverage. The genetic analysis showed that Spanish wheat landraces are clustered in different groups, thus representing genetic pools capable to provide different allelic variation. The subspecies had a major impact on the population structure of durum wheat landraces, identifying three different clusters that corresponded to subsps. durum, turgidum and dicoccon. The population structure of bread wheat landraces was more biased by geographic origin.

Conclusions

The results showed a wider genetic diversity in landraces when compared to a reference set that included commercial varieties, and a higher divergence between landraces and the reference set in durum wheat than in bread wheat. Some genomic regions with patterns of variation that differ between landraces and reference varieties could be detected, pointing out loci under selection during crop improvement that could help to target breeding efforts. The results obtained from this work will be highly valuable for future GWAS analysis.
Background
Wheat is a cereal that belongs to the Poaceae family. Because of its high agronomic adaptability, it is the most widely grown crop in the world. Wheat also occupies a central place in human nutrition providing 20% of the daily protein and food calories. Current cultivated wheat has originated from natural hybridization events between different species (1). Roughly 90 to 95% of the wheat produced in the world is common or bread wheat (*Triticum aestivum* L.; 2n= 6x= 42, 17Gb, AABBD genomes). The rest of the wheat world’s production includes about 35–40 million tonnes of durum wheat (*T. turgidum* var. *durum*; 2n= 4x= 28, 13Gb, AABB genomes), which is cultivated mainly in the Mediterranean region (http://www.fao.org/faostat/en/).

Advances in molecular biology and high-throughput genotyping technologies have significantly improved the field of molecular plant breeding, moving from a phenotype-based to a genotype-based selection (2). The integrated used of genomic and molecular tools in conventional phenotype selection programs has allowed the development of new breeding strategies such as genome-wide association studies (GWAs) and genomic selection. However, in wheat, its large complex genome with over 85% repetitive DNA content has hampered these molecular breeding approaches compared to other crops, i.e. the presence of two or three separate but closely related subgenomes hinders the analysis of homoeologous gene sequences (3). The recently published bread wheat reference genome (4) provides at last high quality information that will help to physically locate molecular markers, and will increase the potential of GWAs analysis by facilitating the identification of key genes underling the associations eventually detected. This is also enabling ‘genomic selection’ to be possible in wheat breeding programs to predict superior traits based on DNA markers (5).

The successful genomics assisted breeding of any crop requires a thorough understanding of the species genetic diversity. As in other crops, genetic diversity of wheat has declined as a consequence of bottlenecks encountered during polyploidization and domestication (6). The modern plant breeding practices, in which only a small number of elite cultivars are included on breeding programs, has further narrowed the genetic base of wheat throughout the world. The consequent erosion of genetic
diversity limits the pool of alleles in which to search for new traits of agronomic interest and adaptability to address novel threats such as climate change. Fortunately, an enormous number of genetically different, locally well-adapted wheat landraces were created by natural and farmer selection in the previous century. These collections represent an invaluable material for genetic studies and crop improvement. A deep knowledge of their genetic/genomic diversity is thus essential to deal with future plant breeding challenges (7).

A large number of studies have been performed to estimate genetic diversity employing different methodologies in diverse plant species (8,9) including wheat (10). It is accepted that molecular markers are the best option to address genetic variation studies, and among them, Single Nucleotide Polymorphisms (SNPs) distributed across the entire genome are the most frequently employed to assess genome wide diversity. It is nowadays possible to detect thousands of SNPs in a large number of accessions in a reasonable time with high-throughput technologies like SNP arrays (11) or genotyping-by-sequencing (GBS) (12).

Assessing genome-wide diversity by GBS provides a robust estimate of the diversity and it has been increasingly adopted as a fast, high-throughput and cost effective tool for whole-genome genetic diversity analysis in large germplasm sets (13). The DArTseq markers, based on GBS (14), efficiently target low copy sequences by a complexity reduction method and have been successful applied for genetic diversity studies in different species (15-17) (18). Compared to SNP-based genomic profiling, the DArTseq method is much more comprehensive in terms of the molecular variation underlying the polymorphisms but also provides data at a more affordable cost, especially in complex polyploid species as wheat (https://www.diversityarrays.com).

High throughput genotyping also provides essential information for the design of high-power genome-wide association studies (GWAs), which enable the identification of agriculturally important genes and ease their transfer from wild or local germplasm into modern cultivars through marker-assisted selection and marker-assisted breeding and/or genomic selection. For GWAs analysis, the optimum diverse panel must be genotyped with a set of molecular markers, covering as much as possible the
genome of the species (19) the population structure needing to be investigated to avoid false associations between phenotypes and markers (20).

The Spanish wheat landraces conserved at National Plant Genetic Resources Center (CRF-INIA) and maintained in the national collection, were collected in the first half of the 20th century. Several studies have shown the great variability of these materials when compared to other germplasm collection (21-24). However, their high throughput genomic characterization has not been yet fully addressed.

The aim of the present study was to characterize two collections of durum and bread wheat landraces from CRF-INIA by using the DArTseq-GBS approach. Specific objectives of the present investigation were: (1) to assess the genomic diversity of a set of durum wheat accessions comprising 191 Spanish landraces and 23 reference varieties, (2) to assess the genomic diversity of a set of bread wheat accessions comprising 189 Spanish landraces and 29 reference varieties, and (3) to compare the genetic diversity of landraces and modern cultivars in both wheat species.

Results
Wheat genotyping

We have characterized at genomic level a set of 380 landraces and 52 reference varieties (Additional file 1). The DArTseq approach allowed us to detected around 100K DArTs (presence/absence markers) and 50K SNPs in each analyzed species.

In the tetraploid durum wheat, a total of 98,983 DArTseq markers and 51,751 SNP markers were obtained for 214 accessions (Additional files 2 and 3). Markers, mapped against the genome of the close hexaploid species *T. aestivum*, were detected along the whole genome (Table 1). In raw data, around 6% and 37% DArTs, and 7 % and 20% SNPs were located in the *T. aestivum* D genome (not present in the tetraploid) or not mapped, respectively. In summary, 57% DArTs and 73% SNPs could be located. After filtering to obtain highly informative markers 38,700 DArTs and 9,324 SNPs were selected for further analysis. In this set of markers, the percentage of markers mapped was similar to the raw data (83% in SNPs and 66% in DArTs). As shown in Table 1, the filters applied did not affect the markers distribution along the genome. The A and B genomes presented a comparable number of
markers, either before or after filtering. The chromosome 4 of the homoeologous group B had the lower density of both types of markers.

In the hexaploid bread wheat, a slightly higher number of markers, 130,899 DArTseq markers and 58,660 SNPs, were generated (Additional files 4 and 5). As in durum wheat, markers were also detected throughout the whole genome (Table 1). The stricter mapping parameters used for bread wheat reduced the chance of homoeologous cross mapping, but also the percentage of located markers when compared to durum wheat. So, around 64% of raw DArTs and 41% of raw SNPs were not mapped in the bread wheat reference genome. After filtering, 44,241 DArTs and 8,238 SNPs were selected for further analysis. In this set of markers, the percentage of mapped markers was similar to the obtained in raw data (36% in DArTs and 58% in SNPs). The D genome presented a reduced amount of markers compared to A and B genomes. Again, the chromosome 4 from homoeologous group B had the lowest density of both types of markers.

The marker distribution along the chromosomes was comparable in both species for all the chromosomes from genomes A and B. In general the distributions showed a higher density of markers at both chromosome ends as it is illustrated for chromosome 2A in Figure 1. The same pattern was observed for bread wheat genome D.

In both species, the distribution of PIC values for DArT and SNP data was asymmetrical and skewed towards the lower values (Figure 2). In durum wheat, the 82% of the DArTs and the 75% of the SNP markers showed a PIC value >0.2. In bread wheat the corresponding values were 76% and 70% for DArTs and SNPs, respectively. In both species and type of markers, the average PIC value was between 0.30 and 0.35.

**Genetic structure of the durum wheat collection**

fastSTRUCTURE runs with 38,700 DArT markers divided the tetraploid wheat landraces into seven populations (K=7) (Figure 3A). All the accessions belonging to subsp. *dicoccon* but one (BGE021775) were grouped in Pop5 and all belonging to subsp. *turgidum* were clustered in Pop3. The landraces in both populations came mostly from the North of Spain (Figure 3B). All the subsp. *durum* landraces but one, (BGE013103) that was classified in Pop3, were distributed into five populations (Pop1, Pop2,
Pop4, Pop6 and Pop7), containing from 10 to 80 accessions (Additional file 1). The Pop6 had the higher number of accessions, showed the greatest degree of admixture and was the population with the most diverse eco-geographical origin (Figure 3). However, some subsp. *durum* populations showed a differentiated geographic distribution (Additional file 1). That is, landraces of Pop1 originated mostly from the eastern Spain, whereas those in Pop2 were from the south-western provinces. The Pop4 included landraces from the South and East of Spain, and the Canary Islands.

Genetic diversity parameters were calculated in the fastSTRUCTURE populations based on SNP data (Table 2). The population with the highest genetic diversity value (Hs) was Pop6 (0.272), and the one with the lowest value was Pop2 (0.048). In the whole landraces collection, the Dest value, a measure of population differentiation in collections with several populations, was 0.22. The F<sub>ST</sub> values, related with Genetic differentiation between populations, ranged from 0.743 (Pop2 vs Pop5) to 0.226 (Pop1 vs Pop6). The Pop6 showed the smallest genetic differentiation from the rest of populations, including those of subsp. *turgidum* (Pop3) and *dicoccon* (Pop5) (Table 2). When diversity between the three subspecies was estimated, regardless of the structured populations, *dicoccon* and *durum* landraces showed the highest value of genetic differentiation between subspecies (F<sub>ST</sub> = 0.42) whereas subsp. *turgidum* showed lower values when compared to either *durum* and *dicoccon* (F<sub>ST</sub> = 0.31 and 0.38, respectively).

When we analyzed the distribution of Hs values along the genome, we detected some genomic diversity patterns that were population-specific (Additional file 6). For example Pop2, Pop5 and Pop7 presented a region of low diversity in the central part of chromosome 2A, while for chromosome 2B we only detected a region of low diversity in Pop5. On the other hand, a similar analysis contrasting the Hs values along the genome in the three durum wheat subspecies has showed some common low diversity regions in *turgidum* and *dicoccon* (e. g., chromosomes 2A, 3A, and 6B), while *durum* showed higher diversity values along the genome (Additional file 7).

We have also explored the genomic structure of the durum wheat collection, including landraces and reference varieties, by a principal coordinate analysis (PCoA) based on the 9,324 filtered SNP
markers. The first two principal coordinates explained 21% of the total variation. Three discrete groups corresponding to the three subspecies could be clearly identified (Figure 4A). This was in agreement with the results obtained with fastSTRUCTURE where the three subspecies were grouped into different populations. A fourth group corresponding to the reference varieties also appeared clearly separated from the subsp. *durum* landraces. The subsp. *durum* accessions were differentiated from the others by PCo1, but the difference between *turgidum* and *dicoccon* was due to PCo2, evidencing, as detected in the Hs analysis, that different sets of markers are responsible of the genetic divergence among subspecies. Some landraces of subsp. *durum* were close to subsp. *dicoccon* and *turgidum*. These landraces were from Pop6 and some of them (e. g. BGE019290) come from the North of Spain (Additional file 1).

We further investigated the allelic variability for a functional marker involved in wheat adaptability, the vernalization gene Vrn-A1, in relation to the population structure. Three different alleles were identified in the collection: the winter type allele vrn-A1, and two alleles related to spring growth habit, Vrn-A1b and Vrn-A1c. The representation of allelic variation in the PCoA showed that most of the accessions carrying the winter type allele vrn-A1 were grouped together as corresponded to *dicoccon* accessions. (Figure 4A) Most of the reference cultivars and subsp. *durum* accessions carried the Vrn-A1c allele, and almost all of the subsp. *turgidum* the Vrn-A1b allele. When analysed within the population structure, all *durum* accessions from Pop2, Pop7, and Pop1 but one, presented the Vrn-A1c allele, which was also identified in the 80% of durum wheat landraces clustered in Pop 6. In Pop3 and Pop4 the most frequent allele was Vrn-A1b. According to passport data (see Additional file 1) accessions with the winter type allele vrn-A1 come mostly from the North of Spain.

The allelic variation was also studied for HMW-GS loci Glu-A1 and Glu-B1, related to wheat rheological properties, but no relationship with the population structure could be observed (data not shown).

As the subspecies was the main discriminant factor in the global PCoA, we decided to perform the analysis excluding the *turgidum* and *dicoccon* accessions to gain insight into the variability within the subsp. *durum* (Figure 4B). The populations identified in the previous analysis with DArTs (Figure 3A) were similarly grouped in the SNP-based PCoA, Pop6 being again the one that showed more
dispersion due to its higher intrapopulation variability (Figure 3A). The only *durum* accession clustered in Pop3 (BGE013103) appeared now closely located to Pop1 landraces (Figure 4B). This local variety can be identified in the Figure 3A at the end of Pop3, showing admixture with Pop1, which suggests that it could have a hybrid genotype between *durum* and *turgidum*.

The Pop1, Pop2, and Pop7 were the *durum* populations most differentiated from the reference set. On the other hand, Pop4 was the closest to the reference group. This population includes old local varieties, as ‘Ledesma’ or ‘Lebrija’, obtained by crossing between ‘Senatore Capelli’ and Spanish landraces. One landrace from Pop6 of subsp. *durum* (BGE026954) grouped together with the reference varieties. This accession, collected at the end of 90’s, is characterized by early-maturity and short plants (22), which suggest that probably it is not a true landrace.

**Genetic Structure of the bread wheat collection**

fastSTRUCTURE runs with 44,241 DArT markers divided the hexaploid wheat landrace accessions into four populations (K=4). Compared to durum wheat landraces, a higher level of admixture was detected in bread wheat populations, in special within Pop2, the largest one, which contained 112 accessions (Figure 5A, Additional file 1). Landraces from Pop1 came from Central Spain, and landraces from Pop4 came from the West, including Canary Islands. Pop2 and Pop3 showed a more diverse eco-geographical origin (Figure 5B, Additional file 1).

As shown in Table 3, the population with higher Hs was Pop2 (0.277), and the one with the lowest value was Pop3 (0.101). In the whole landraces collection the Dest value was 0.17, smaller than the differentiation found in the durum collection (0.22). The $F_{ST}$ values between populations ranged from 0.169 (Pop1 vs Pop2) to 0.573 (Pop3 vs Pop4). According to $F_{ST}$ values, Pop4 was the most different population (Table 3). Regarding to the Hs distribution along the genome, we detected some low diversity regions specific of populations (Additional file 8). For instance, Pop3 and showed low diversity regions on chromosomes 1A and 7A, while Pop1 and Pop3 showed low diversity in chromosomes 3A and 2B. This suggests that different genomics regions are responsible of the divergence among populations.
The relationship among the bread wheat accessions was also assessed by a PCoA, based on 8,238 SNPs in the whole bread wheat collection. The total amount of genetic variation explained by the first two principal coordinates was 19.2%. The first two coordinates clearly separated the Pop4 (by PCo1) which formed the most distant group, separated from the other three populations that appeared distributed along PCo2 (Figure 6A). Some degree of overlapping was shown between Pop2 distributed along the PCo2 and both, Pop1 and Pop3 located in the upper extreme of PCo2. This is in agreement with the higher degree of admixture in Pop2 revealed by fastSTRUCTURE analysis (Figure 5A).

Reference varieties were located in a tight space, but overlapping with some Pop2 landraces. As in durum wheat, we investigated the allelic variability of Vrn-A1 gene in relation to the population structure. In bread wheat, three different alleles were identified in the collection, the winter type allele vrn-A1, and two alleles generally related to spring growth habit, Vrn-A1a and Vrn-A1b.

Accessions from Pop4 were characterized by having almost exclusively the Vrn-A1b allele. The accessions presenting the other two alleles could be differentiated along the PCo2 axis, whose upper part corresponded to spring type landraces whereas the reference varieties and winter type landraces were included in the lower part (Figure 6A).

The representation of Glu-B1 alleles specific of Iberian landraces in the PCoA showed that accessions that carried the Glu-B1f allele (HMW subunits 13+16) were clustered at the right part of the PCoA and corresponded to most of Pop4 samples. Again, an interesting tendency could be observed along the PCo2 axis: accessions with the Glu-B1e allele (HMW subunits 20x+20y) appeared grouped in the upper part, while the lower part corresponded to accessions with Glu-B1u or Glu-B1al alleles (HMW subunits 7+8 or 7OE+8) (Figure 6B).

The allelic variation was also studied for HMW-GS loci Glu-A1 and Glu-D1, but no relationship with the population structure could be observed (data not shown).

The overlap between landraces and reference varieties in bread wheat was more remarkable than in durum wheat. Actually, reference varieties fully overlapped with Pop2 accessions. Some of the landraces closer to the references varieties were collected in the 90’s (e. g. BGE025410) and according to their early flowering and lower height phenotype
Divergence between landraces and reference varieties

The PCoA analysis showed a clear genetic divergence between landraces and reference varieties in durum wheat whereas it was not so evident in bread wheat, where landraces and reference varieties overlapped (Figures 4 and 6). Regarding to the overall genetic diversity, landraces showed a higher diversity than the reference varieties in both species (Table 4), the difference being greater in durum wheat.

The divergence between reference varieties and landraces varied among the different populations identified by fastSTRUCTURE. Durum wheat reference varieties showed the highest differentiation with landraces in Pop5 (*dicoccon*) and Pop2 (*durum*), and the smallest with *durum* landraces in Pop4 and Pop6 ($F_{ST}$ values of 0.586, 0.555, 0.372 and 0.214, respectively). In bread wheat, reference varieties showed the highest and lowest differentiation with the landraces in Pop4 and Pop2, respectively ($F_{ST}$ values of 0.416 and 0.08, respectively).

In order to analyze the degree of allele fixation in the reference varieties, we studied the presence of monomorphic markers. Around the 40% of SNP markers were fixed in durum wheat reference varieties, the number of monomorphic markers (3,791) being comparable to the found in subsps. *dicoccon* and *turgidum* (4,079 and 2,119 respectively) and much higher than the found in subsp. *durum* (478 fixed markers). In bread wheat reference accessions, the number of fixed SNP markers was 1,771 (21%) clearly lower than the obtained in durum wheat (Table 4).

We also analyzed the Hs distribution along the genome in landraces and reference varieties (Figure 7). We called low diversity regions in the reference varieties with respect to the landraces, as those might be regions fixed by breeding efforts. In durum wheat we detected 24 genomic regions located in 11 different chromosomes (Figure 7A). All chromosomes except 1A, 1B and 4B presented at least a fixed region. Chromosome 3B, the largest one, with 5 regions spanning a total of 46Mb, presented the greatest number of low diversity regions. On its turn, chromosome 5A presented the widest region (42Mb). In summary, low diversity regions spanned 288MB and include some key genes. Such as, the

([http://webx.inia.es/web_coleccionescrf/CaracterizacionCRFeng.asp](http://webx.inia.es/web_coleccionescrf/CaracterizacionCRFeng.asp)) may not be real landraces.
region located in chromosome 2B, that overlapped with Sr36 gene related with resistance to leaf rust. In bread wheat a similar number (22) of lower diversity regions were called (Figure 7B). Interestingly, no regions were identified in the D genome. In the A genome we detected 12 regions located in chromosomes 1A, 2A, 3A and 7A, finally in the genome B we observed regions in all the chromosomes except 3B. The widest region was in chromosome 2B spanning 110Mb, and chromosome 7A with 6 regions contained the greater number of regions. In bread wheat low diversity regions spanned 291MB and contained the key gene Rht-B1 gene introduced in wheat cultivars during the 20th in chromosome 4B.

Discussion
Due to the new environmental constrains, the need to exploit the full potential of germplasm resources to accelerate crop performance gains has become urgent. Wheat genetic resources from genebanks need to be thorough characterized to transfer useful genetic variation into modern elite gene pools. Cultivated wheat includes cultivars from two polyploid species. Therefore it is necessary to include tetraploid and hexaploid varieties in genetic variability assessment programs. Several recent international initiatives are characterizing local wheat germplasm collections. However, Spanish landraces, locally adapted to a really wide diversity of eco-climatic conditions, are absent or only represented by a small number of accessions in these studies. We decided to explore the genetic diversity of this valuable material, by analyzing at genomic level 380 selected landraces representing the genetic variability of the wheat collection maintained in the Spanish Plant Genetic Resources Center, which is composed by more than 1,600 accessions.

Wheat high throughput genotyping
This is the first report of high throughput genotyping by GBS of Spanish wheat germplasm. The GBS method has the potential to provide in-depth and robust genomic diversity estimates. Moreover, it may reveals new and favorable alleles, which might have a high value for prebreeding (25-27). The GBS-DArTseq methodology has been successfully applied in wheat species, where a standard GBS approach, that requires the genome resequencing, is still challenging due to their huge complex genome (7).
In our study, 50K SNP and 100K DArT markers in each of the wheat species were analyzed, from which we have been able to select around 10K SNP and 40K DArT high quality markers. The availability of the reference wheat genome (4) (WGA v0.4) allowed the marker mapping in both species. The results showed that markers were distributed all over the genome although D genome markers were remarkably less abundant than A and B genome markers in bread wheat. It can be noted that earlier hexaploid wheat maps based on expressed sequence tag loci have showed a fairly even distribution of markers among the three wheat genomes (e. g.,(28)). Among homoeologous sets, group 4 chromosomes were the less covered, especially 4B and 4D. No satisfactory hypothesis has yet been proposed to explain the relative scarcity of markers consistently mapped on 4B compared to 4A (28,29) despite the latter has been the involved in several structural rearrangements during wheat (30).

Interestingly, the percentage of mapped markers was higher in durum wheat than in bread wheat although the reference genome employed is from the latter. This can be due to the fact that two different sets of mapping parameters were used. Parameters were less strict in durum wheat, as the reference genome corresponds to the hexaploid T. aestivum. However, it will be advisable to repeat the mapping procedure in durum wheat once its reference genome becomes available, to avoid any mapping misplacements prior to GWAs analysis. The informativeness of markers was assessed by their PIC values whose distribution and average PIC value was comparable to the previously reported in wheat (17,31-34).

**Population structure of Spanish wheat landraces**

Ex situ collections preserve valuable genetic resources collected over time that include germplasm from different genepools. The identification of useful alleles is largely improved when the genetic structure of the germplasm collection is known. Moreover, the study of population structure is also important for genomic studies, being a previous and mandatory step to successfully perform further GWAS analysis. In our case, since GWAS will be based on the high quality SNPs identified, we decided to assess the structure with a different set of markers, such as DArTs. The population structure was assessed with FastSTRUCTURE algorithm, developed by the authors of the classical STRUCTURE
software, but with faster runs, and comparable ancestry estimates and prediction accuracies (35). In both species, the analysis of population structure was complemented with a PCoA analysis performed with the SNPs data set. As expected, the results were in close agreement, but some additional information about subgroups and their relation to phenotypic traits could be extracted.

**Durum wheat landraces**

Spanish durum wheat (*Triticum turgidum* L.) landraces belong to three main interfertile subspecies (*dicoccon*, *turgidum* and *durum*). The subsp. *dicoccon*, also known as emmer wheat, is a hulled wheat only grown in the North of Spain that represents the feral situation in this crop. The subsp. *durum* is the most widely cultivated and well adapted to the dry-summer conditions of the South. The subsp. *turgidum*, less common and grown in colder areas than *durum*, corresponds mostly to winter wheats (36). In a previous research, SSRs were used to assess the genetic structure of the collection of durum wheat landraces analyzed here and 9 populations were established. Some of the populations included more than one subspecies and several genotypes could not be classified in any population (37). In the present work we identified seven populations, all of them composed by landraces from one subspecies with only two exceptions (Figure 3). The discrepancies between this and the earlier work can be explain by the different type and number of markers employed for assessing the population structure in both studies (40K DArTs vs 39 SSRs).

Analyses comprising other durum wheat materials have not been able to separate the subspecies *durum* and *turgidum* (38,39). However, we have clearly differentiated the three subspecies. This demonstrates the analytical power of DArT and SNP markers for taxonomical identification and supports the classification of Mac Key (40,41) with *turgidum* and *durum* as separate subtaxa. In our study, subsp. *dicoccon* was closer to *turgidum* than to *durum* (see $F_{ST}$ values, Table 2), indicating that both subspecies share a common allele pool. As both subspecies were grown in similar environmental conditions, this closeness might be due to similar selective pressures during local adaptation. The analysis allowed also the identification of accessions that could be admixtures between subspecies, such as BGE021775 a *dicoccon* landrace closer to *durum*, or BGE04564, a *durum* landrace grouped with the *turgidum* landraces. Those admixtures have been previously reported in durum wheat as not
unusual in ancient local forms (40).

Vernalization genes determine winter/spring growth habit in temperate cereals and are one of the major genetic factors that, affecting their life cycle, enable the crop plant to adapt to a wide range of environments (42). The evolution of spring cultivars from winter habit ancestors is a main event in the post-domestication spread of wheat. However, studies of the major vernalization gene VRN1 have been mostly limited to hexaploid wheat species and very few reports in tetraploid species can be found in the literature (43,44). None of the durum wheat accessions characterized here showed the Vrn-A1a allele described in spring-habit hexaploid wheats, which is according to what has been found in other studies (45,46). All seems to indicate that this allele appeared in wheat evolution after the last polyploidization event. The Vrn-A1c allele has been described as the most frequent determinant of spring habit in tetraploid wheats (45,47,48) but has been described as rare in emmer wheats (38).

In our study, 78% of durum accessions presented this allele, but almost all turgidum accessions (83%) carried the Vrn-A1b, and 7 out of the 11 dicoccon landraces characterized for VRN1 carried the winter vrn-A1 allele (Figure 4). It is noteworthy the presence of alleles associated to spring growth habit in the remaining four accessions from subsp. dicoccon, as in this subspecies, traditionally cultivated in mountainous regions where the vernalisation requirement is of importance, the winter habit is predominant. Our results are consistent with the description of few Spanish spring emmer wheats in the past (36), and remarks the value of the collection to provide favourable alleles for the creation of novel cultivars with improved adaptability to changing eco-climatic conditions.

Low or none influence of the geographic origin on the genetic structure of durum wheat materials, due to the exchange of germplasm, has been reported for landraces of Iran, Central Fertile Crescent and Ethiopia (49-51). In our collection, a restricted geographic distribution exists for the Spanish landraces belonging to the less represented subspecies dicoccon (14 accessions) and to some extent for turgidum landraces (37 accessions). However, landraces of subsp. durum, which were structured into five distinct populations, showed not only greater variability and complexity but also less influence of the origin, although some geographic areas were predominant in some of the populations.
Bread wheat landraces

The great majority of the Spanish bread wheat landraces conserved at CRF belong to *Triticum aestivum* subsp. *vulgare*, and thus our bread wheat set was composed exclusively by this subspecies. The population stratification of the bread wheat panel identified four groups of landraces with a great divergence according to the $F_{ST}$ values. This clustering reflected better the geographic origin of the accessions than in the subsp. *durum*. The genetic differentiation, estimated by $D_{est}$, was smaller in the bread landraces collection than in the whole durum wheat collection, in agreement with the less level of stratification observed. Few studies have simultaneously addressed the variability of hexaploid and tetraploid wheats but a higher genetic diversity in durum than in bread wheat has been previously reported in landraces from other countries (e. g., (52)).

One of the four groups detected (Pop4) was clearly more genetically distant. This group included landraces from western Spain, where there is a prevalence of acidic or neutral soils (53). Most of the accessions from this population show spring growth habit, and carry the $f$ allele at Glu-B1 HMW locus. *Glu-B1f* allele has a low frequency in worldwide collections but has been previously described as characteristic of Iberian landraces, which also holds for the *Glu-B1e* allele (54,55). The latter was predominant in Pop1 and Pop3, being also present in some landraces of Pop2 that were closely grouped by PCoA. The *Glu-B1e* allele is related to poor rheological properties in bread wheat, but the $f$ allele has been associated with a good dough quality (56). The presence of this variant in a discrete group of more differentiated landraces supports their common origin.

In our study, the winter-type allele *vrn-A1* was the most common *Vrn-A1* allele found in *T. aestivum* reference cultivars (22 out of 29 varieties). However, the most frequent within the bread wheat landraces was the *Vrn-A1a* allele, found to be absent in durum wheats. It is the most potent spring-habit allele described, causing complete insensitivity to vernalization (57). Though the measured impact of growth habit on genetic differentiation is limited, spring and winter bread wheat can frequently be distinguished by cluster DAPC analyses (58,59). Two discrete groups regarding growth habit were not clearly defined in our PCoA analysis, but the allelic variability showed some relation to population structure and some biased tendency along the second PCo axis (Figure 6A). On this
matter, it has to be kept in mind that vernalization response is a complex process under polygenic control. Vrn-A1 has been described as the main genotypic determinant of the vernalization requirements in temperate crops, but there are other genes, as Vrn-B1 and Vrn-D1, whose allelic variability has not yet been addressed.

This is the first time that the bread wheat resources maintained in the CRF-INIA Spanish national genebank are characterized at genetic level and compared to durum wheat accessions. The deep knowledge of their genetic structure represents the starting point for the development of a core collection of Spanish T. aestivum wheats, which will allow the efficient management and use of its valuable gene pool.

**Relationship between landraces and modern cultivars**

In diversity studies, wheat landraces usually cluster in a separate group from elite cultivars (50,51,59,60), but in some cases some degree of mixture has also been found (32,58). Concerning the durum wheat materials examined here, it is remarkable the great genetic divergence between the bulk of landraces and the reference set (Figure 4A). Moreover, some *durum* landraces are closer to *turgidum* and *dicoccum* than to the reference varieties (that belong all to subsp. *durum*). However, a great relatedness with the reference varieties has been detected for a reduced group of *durum* landraces that includes ‘Caravaca’ (BGE002869), a Spanish landrace used by CIMMYT in the development of some modern cultivars (see the Genetic Resources Information System for Wheat and Triticale of CIMMYT at http://www.wheatpedigree.net/). It can also be noted that other studies have reported a relation between Spanish and North-African landraces (32,38) and that the reference set included two old cultivars commonly cultivated in Spain in the past (‘Senatore Capelli’ and ‘Bidi-17’) both with a North African origin. Nevertheless, and overall, our results suggest a little involvement of Spanish landraces in the development of modern durum wheat varieties grown in Spain at present.

In bread wheat, the situation was quite different, and mixture between landraces and reference varieties was much higher, in special with accessions in Pop2 (Figure 6). Likely, some of them are not true local landraces, but rather old adopted cultivars wrongly classified. Some others are related to
Italian material, such us ‘Ritchela Blanca’, ‘Montnegre’ or ‘Ardito’ (61,62). Clustering of landraces and reference varieties can also be indicating a pedigree relationship. Then, it is possible that some of the landraces characterized in our study were among those utilized by the early breeders to develop pure lines.

Mediterranean wheat landraces have been considered highly valuable due to their huge genetic diversity and the presence of accessions with high resilience to abiotic stresses, resistance to pests and diseases and grain quality (63-65). This fact was patent in our study, where the overall genetic diversity of the reference cultivars was way lower than that of the landraces in both species (Table 2; Figures 4 and 6).

Genomic regions with patterns of variation that differ between landraces and reference varieties can aid to identify loci under selection during crop improvement which is of help to better target breeding efforts (66) (Figure 7). Our analysis allowed the identification of several of such genomic regions studying the distribution of genetic diversity along the genome. The number of regions in durum wheat was higher than in bread wheat, probably due to the higher number of mapped markers in this species. As expected, some of the chromosomes containing fixed regions in both species harbor genes related to agronomically important phenotypes, such as, Sr36 in chromosome 2B in durum wheat, or in bread wheat Rht-B1 (4B) and Vrn-B1 (5B), associated with vernalization (67,68). On their turn, the fixed region in chromosome 1A can be related with the presence determinants of bread quality, as Glu-1 in this chromosomes. Coupling this analysis with future GWAs studies will help to identify the traits underlying each of the fixed regions detected in this work. Moreover deciphering of their gene content will be facilitated by the remapping of the markers against the durum reference genome once it will be available.

Conclusions
The replacement of local landraces by high-yielding wheat varieties that began in the times of the Green Revolution has led to a loss of genetic variation in crop wheats. This depletion has now encouraged the use of genetic resources in wheat breeding programs, but the genetic variability of these resources needs to be exhaustively characterized for their efficient use. This study
demonstrated the relevance of DArTseq technology as a reliable and cost-effective tool for assessing the diversity within and between landraces and for establishing the genetic relationships between them and modern wheat cultivars. The study of the genome-wide diversity will also provide a resource for the design of high-power GWAS experiments, which will help to achieve the overarching goal of improving wheat for different environments, ecosystems and stress situations. The collections of Spanish landraces characterized in the present study are clearly clustered in different groups, thus representing different genepools capable to provide different sources of genes for plant breeding. The panel of genotypes investigated has shown an outstanding degree of diversity compared to the reference counterparts. It then clearly represent a strategic platform and a valuable genetic resource that must be further study to ensure not only its efficient conservation and management but also its useful exploitation in breeding programs.

Methods

Plan material

The plant material analyzed in the present study comprised 432 selected accessions (detailed information is presented in supplementary Additional file 1). This sample included 191 durum and 189 bread wheat landraces and old local cultivars (from now on referred as landraces), representing different ecological and geographical areas from Spain (22,69). Among the durum wheat landraces, 140, 37 and 14 corresponded to the subspecies *durum*, *turgidum* and *dicoccon*, respectively, while all the bread wheat landraces belonged to the subspecies *vulgare*. All these accessions were homozygous lines derived from Genebank accessions provided by CRF-INIA (Centro de Recursos Fitogenéticos, INIA, Madrid). Additionally, the study included a set of 23 durum and 29 bread wheat improved varieties (from now on referred as reference varieties) that comprised the cultivars most widely grown in Spain during the last 50 years, plus some varieties widely used in wheat research such as, the tetraploid ‘Langdon’ and the hexaploid ‘Chinese Spring’.

DNA isolation and genotyping analysis

For each accession, genomic DNA was isolated from young leaves of one plant using the CTAB method (70). Samples were genotyped using the DArTseq GBS technology (71) at Diversity Arrays
Technology Pvt, Ltd (Canberra, Australia) for durum wheat accessions, and SAGA (Genetic Analysis Service for Agriculture, Mexico City, Mexico) for bread wheat accessions.

The marker calling and mapping was performed by the genotyping services. For durum wheat, markers were blasted against the current available *Triticum aestivum* genome IWGSC WGA v0.4 (https://wheat-urgi.versailles.inra.fr) a tightly related species. A marker was mapped when blast E-value < 5e-7 and sequence identity >70%. For bread wheat, markers were blasted against the same genome. However, in this case more strict parameters (blast E-value < 5e-10 and sequence identity >90%) were used, as accessions belonged to the reference genome species. Genotyping services provided two different sets of markers, including their genomic location. DArT markers were scored as binary data (0/1) indicating the presence or absence of a marker in each accession, and SNP markers were scored as 0/1/2 indicating the presence of the reference allele in homozygosis, the alternative allele in homozygosis or an heterozygous genotype, respectively. Raw data are available upon request to the corresponding author.

To contrast with the population structure based on GBS-DArTseq markers, we investigated the allelic variability for functional markers in Vrn-A1 gene, one of the most determinant loci involved on the transition from vegetative to reproductive growth (72,73). It has been described that carrying a dominant allele at Vrn-A1 locus is enough to confer a spring growth habit (48). Three alleles were characterized by PCR, *Vrn-A1a*, *Vrn-A1b* and *Vrn-A1c* according to (45,47), and following the protocols described in https://maswheat.ucdavis.edu/protocols/Vrn/index.htm.

Additionally, the panel of accessions was also genotyped for *Glu-1* homoeoloci. These complex loci encode for the high molecular weight glutenin subunits (HMW-GS), the major determinants of dough quality in wheat. For this purpose, endosperm proteins were extracted from single seeds and fractionated in sodium dodecyl sulphate polyacrylamide gel electrophoresis (12% polyacrylamide gels) according to Payne et al. (1980). HMW-GS alleles classification was performed according to the Catalogue of Gene Symbols for Wheat 2013 (74).

**Data analysis**

Prior to any molecular-based analysis, the set of markers to be considered was filtered employing
homemade R scripts (75) available upon request to the corresponding author. For SNP markers, we did not expect any real heterozygous genotypes, as all the accessions genotyped were homozygous lines. Thus, when the genotypic values for a marker were only 0 or 2, we considered the heterozygous calling an error caused by the presence of the SNP marker in homoeologous genomes. In this case, the genotypes scored as 2 (putative heterozygotes) were recoded as 1. The same procedure was followed when only 1 or 2 genotypic values were present in a sample, but in those cases genotypes scored as 2 were recoded as 0. Finally, when the genotypic values for a marker included 0, 1 and 2 we recoded as missing data the genotypes scored as 2. From this step, we filtered SNP and DART markers following the same procedure. First, when several markers presented the same allelic profile, all but the one with the less missing data were removed. Then, markers with more than 10% missing data or monomorphic (Minimum Allele Frequency <0.05) were excluded.

Genetic substructure inside the durum and bread wheat landrace collections was investigated using the fastSTRUCTURE algorithm (35) and the DArT markers dataset. Default parameters and K values from 1 to 15 were tested. The appropriate number of components that explained structure in the dataset was determined using the chooseK.py function (35). Results for the identified optimal values of K were visualized using DISTRUCT (76). Individual accessions were assigned to the population with the highest proportional membership.

Genetic similarity based on SNP data inside the full sets of accessions (landraces plus reference varieties) was analyzed by Principal Coordinates Analysis (PCoA) using the gl.pcoa function from dartR R package (77).

Gene diversity (Hs) within populations, landraces and reference varieties were calculated based on the SNP dataset according to (78) with the function basic.stats from hierfstat R package (79). Dest, a measure of population differentiation in collections with several populations, was calculated as defined by (80), with this same function. Genetic differentiation between populations was analyzed estimating a pairwise Fixation index (Fst) according to (81) with the stamppFst function from R StAMPP package (82) using the SNP dataset.

Fixed genomic regions on the reference varieties were identified by doing a scan of the Hs values
along the different chromosomes. Hs was estimated as described previously. However we did not filter out the SNPs markers with the same allelic profile in order to increase the number of mapped markers to refine the estimations. A region was considered as “fixed” when it contained at least 5 consecutive markers with Hs equal to 0 in the reference varieties and at least 5 markers with Hs > 0.1 in the landraces, and spanned more than 5 Mb.

Acknowledgments
This work was supported by the Ministry of Economy, Industry and Competitiveness projects AGL2016-77149 and RFP2015-00008- C04-01, and Universidad Politécnica de Madrid project VJIDOCUPM18LPB. ML is a recipient of a predoctoral fellowship from Programa Propio of Universidad Politécnica de Madrid.

AUTHOR CONTRIBUTIONS STATEMENT
PG and MR: Conceived and designed the study; EB, ML, LP, MR and JV: Characterized the panel; ML, LP and HP: Performed the bioinformatics analyses; PG, LP and MR: Valorized the results and wrote the manuscript; All authors read and approved the final manuscript.

CONFLICT OF INTEREST STATEMENT
The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References
(1) Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer M, Jakobsen KS, et al. Ancient hybridizations among the ancestral genomes of bread wheat. Science 2014;345(6194):1250092.
(2) Leng P, Lübberstedt T, Xu M. Genomics-assisted breeding—a revolutionary strategy for crop improvement. Journal of integrative agriculture 2017;16(12):2674-2685.
(3) Mochida K, Yamazaki Y, Ogihara Y. Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. Molecular Genetics and Genomics 2004;270(5):371-377.
(4) Appels R, Eversole K, Feuillet C, Keller B, Rogers J, Stein N, et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 2018;361(6403):eaar7191.
(5) Rasheed A, Xia X. From markers to genome-based breeding in wheat. Theor Appl Genet
2019;132(3):767-784.

(6) Haudry A, Cenci A, Ravel C, Bataillon T, Brunel D, Poncet C, et al. Grinding up wheat: a massive loss of nucleotide diversity since domestication. Mol Biol Evol 2007;24(7):1506-1517.

(7) Sehgal D, Vikram P, Sansaloni CP, Ortiz C, Saint Pierre C, Payne T, et al. Exploring and mobilizing the gene bank biodiversity for wheat improvement. PLoS One 2015;10(7):e0132112.

(8) Idrees M, Irshad M. Molecular markers in plants for analysis of genetic diversity: a review. European academic research 2014;2(1):1513-1540.

(9) Govindaraj M, Vetriventhan M, Srinivasan M. Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. Genetics research international 2015;2015.

(10) Khan MK, Pandey A, Choudhary S, Hakki EE, Akkaya MS, Thomas G. From RFLP to DArT: molecular tools for wheat (Triticum spp.) diversity analysis. Genet Resour Crop Evol 2014;61(5):1001-1032.

(11) Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, et al. Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant biotechnology journal 2014;12(6):787-796.

(12) Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS one 2011;6(5):e19379.

(13) Holtz Y, Ardisson M, Ranwez V, Besnard A, Leroy P, Poux G, et al. Genotyping by sequencing using specific allelic capture to build a high-density genetic map of durum wheat. PloS one 2016;11(5):e0154609.

(14) Akbari M, Wenzl P, Caig V, Carling J, Xia L, Yang S, et al. Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theor Appl Genet 2006;113(8):1409-1420.

(15) Valdisser PA, Pereira WJ, Almeida Filho JE, Müller BS, Coelho GR, de Menezes IP, et al. In-depth genome characterization of a Brazilian common bean core collection using DArTseq high-density SNP genotyping. BMC Genomics 2017;18(1):423.

(16) Yang X, Ren R, Ray R, Xu J, Li P, Zhang M, et al. Genetic diversity and population structure of core
watermelon (Citrullus lanatus) genotypes using DArTseq-based SNPs. Plant Genetic Resources 2016;14(3):226-233.

(17) Baloch FS, Alsaleh A, Shahid MQ, Çiftçi V, de Miera, Luis E Sáenz, Aasim M, et al. A whole genome DArTseq and SNP analysis for genetic diversity assessment in durum wheat from central fertile crescent. Plos one 2017;12(1):e0167821.

(18) Robbana C, Kehel Z, Sansaloni C, Bassi F, Amri A. Genome-Wide Genetic Diversity and Population Structure of Tunisian Durum Wheat Landraces Based on DArTseq Technology. International journal of molecular sciences 2019;20(6):1352.

(19) Korte A, Farlow A. The advantages and limitations of trait analysis with GWAS: a review. Plant methods 2013;9(1):29.

(20) Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, et al. An Arabidopsis example of association mapping in structured samples. PLoS genetics 2007;3(1):e4.

(21) Pflüger LA, Martín LM, Alvarez JB. Variation in the HMW and LMW glutenin subunits from Spanish accessions of emmer wheat (Triticum turgidum ssp. dicoccum Schrank). Theor Appl Genet 2001;102(5):767-772.

(22) Ruiz M, Giraldo P, Royo C, Villegas D, Aranzana MJ, Carrillo JM. Diversity and genetic structure of a collection of Spanish durum wheat landraces. Crop Sci 2012;52(5):2262-2275.

(23) Aguiriano E, Ruiz M, Fité R, Carrillo JM. Analysis of genetic variability in a sample of the durum wheat (Triticum durum Desf.) Spanish collection based on gliadin markers. Genet Resour Crop Evol 2006;53(8):1543-1552.

(24) Moragues M, Moralejo M, Sorrells ME, Royo C. Dispersal of durum wheat [Triticum turgidum L. ssp. turgidum convar. durum (Desf.) MacKey] landraces across the Mediterranean basin assessed by AFLPs and microsatellites. Genet Resour Crop Evol 2007;54(5):1133-1144.

(25) Kilian B, Graner A. NGS technologies for analyzing germplasm diversity in genebanks. Briefings in functional genomics 2012;11(1):38-50.

(26) Heslot N, Rutkoski J, Poland J, Jannink JL, Sorrells ME. Impact of marker ascertainment bias on genomic selection accuracy and estimates of genetic diversity. PLoS One 2013 September
(27) 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 biology 2014;14(1):320.

(28) Qi LL, Echalier B, Chao S, Lazo GR, Butler GE, Anderson OD, et al. A chromosome bin map of 16,000 expressed sequence tag loci and distribution of genes among the three genomes of polyploid wheat. Genetics 2004;168(2):701-712.

(29) Miftahudin KR, Ma X, Mahmoud AA, Layton J, Milla MR, Chikmawati T, et al. Analysis of expressed sequence tag loci on wheat chromosome group 4. Genetics 2004;168(2):651.

(30) Devos KM, Dubcovsky J, Dvorak J, Chinoy CN, Gale MD. Structural evolution of wheat chromosomes 4A, 5A, and 7B and its impact on recombination. Theor Appl Genet 1995;91(2):282-288.

(31) Cubizolles N, Rey E, Choulet F, Rimbert H, Laugier C, Balfourier F, et al. Exploiting the repetitive fraction of the wheat genome for high-throughput single-nucleotide polymorphism discovery and genotyping. The plant genome 2016;9(1).

(32) Kabbaj H, Sall AT, Al-Abdallat A, Geleta M, Amri A, Filali-Maltouf A, et al. Genetic diversity within a global panel of durum wheat (Triticum durum) landraces and modern germplasm reveals the history of alleles exchange. Frontiers in plant science 2017;8:1277.

(33) Rimbert H, Darrier B, Navarro J, Kitt J, Choulet F, Leveugle M, et al. High throughput SNP discovery and genotyping in hexaploid wheat. PloS one 2018;13(1):e0186329.

(34) Novoselović D, Bentley AR, Šimek R, Dvojković K, Sorrells ME, Gosman N, et al. Characterizing Croatian wheat germplasm diversity and structure in a European context by DArT markers. Frontiers in plant science 2016;7:184.

(35) Raj A, Stephens M, Pritchard JK. Variational inference of population structure in large SNP datasets. Genetics 2014:genetics. 114.164350.

(36) Gadea M. Trigos españoles. Madrid, España: Instituto Nacional de Investigaciones Agronomicas,; 1954.

(37) Ruiz M, Giraldo P, Royo C, Villegas D, Aranzana MJ, Carrillo JM. Diversity and genetic structure of
a collection of Spanish durum wheat landraces. Crop Sci 2012;52(5):2262-2275.

(38) Oliveira HR, Campana MG, Jones H, Hunt HV, Leigh F, Redhouse DI, et al. Tetraploid wheat landraces in the Mediterranean basin: taxonomy, evolution and genetic diversity. PLoS One 2012;7(5):e37063.

(39) Sahri A, Chentoufi L, Arbaoui M, Ardisson M, Belqadi L, Birouk A, et al. Towards a comprehensive characterization of durum wheat landraces in Moroccan traditional agrosystems: analysing genetic diversity in the light of geography, farmers’ taxonomy and tetraploid wheat domestication history. BMC evolutionary biology 2014;14(1):264.

(40) MacKey J. Species relationship in Triticum. Hereditas 1966;2:237-276.

(41) Mac Key J. Wheat: its concept, evolution, and taxonomy. Durum Wheat Breeding: CRC Press; 2005. p. 35-94.

(42) Distelfeld A, Cakmak I, Peleg Z, Ozturk L, Yazici AM, Budak H, et al. Multiple QTL-effects of wheat Gpc-B1 locus on grain protein and micronutrient concentrations. Physiol Plantarum 2007;129(3):635-643.

(43) Basualdo J, Díaz ML, Cuppari S, Cardone S, Soresi D, Camargo GP, et al. Allelic variation and differential expression of VRN-A1 in durum wheat genotypes varying in the vernalization response. Plant Breeding 2015;134(5):520-528.

(44) Konopatskaia I, Vavilova V, Kondratenko EY, Blinov A, Goncharov NP. VRN1 genes variability in tetraploid wheat species with a spring growth habit. BMC plant biology 2016;16(3):244.

(45) Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, SanMiguel P, et al. The wheat VRN2 gene is a flowering repressor down-regulated by vernalization. Science (New York, N.Y.) 2004 March 12;303(5664):1640-1644.

(46) Shcherban A, Emtseva M, Efremova T. Molecular genetical characterization of vernalization genes Vrn-A1, Vrn-B1 and Vrn-D1 in spring wheat germplasm from Russia and adjacent regions. Cereal research communications 2012;40(3):351-361.

(47) Fu D, Szűcs P, Yan L, Helguera M, Skinner JS, Von Zitzewitz J, et al. Large deletions within the first intron in VRN-1 are associated with spring growth habit in barley and wheat. Molecular genetics and
genomics 2005;273(1):54-65.

(48) Chu CG, Tan CT, Yu GT, Zhong S, Xu SS, Yan L. A Novel Retrotransposon Inserted in the Dominant Vrn-B1 Allele Confers Spring Growth Habit in Tetraploid Wheat (Triticum turgidum L.). G3 (Bethesda) 2011 Dec;1(7):637-645.

(49) Baloch FS, Alsaleh A, Shahid MQ, Ciftci V, E Saenz de Miera, L., Aasim M, et al. A Whole Genome DArTseq and SNP Analysis for Genetic Diversity Assessment in Durum Wheat from Central Fertile Crescent. PLoS One 2017 Jan 18;12(1):e0167821.

(50) Fayaz F, Sarbarzeh MA, Talebi R, Azadi A. Genetic Diversity and Molecular Characterization of Iranian Durum Wheat Landraces (Triticum turgidum durum (Desf.) Husn.) Using DArT Markers. Biochem Genet 2018:1-19.

(51) Mengistu DK, Kidane YG, Catellani M, Frascaroli E, Fadda C, Pè ME, et al. High-density molecular characterization and association mapping in Ethiopian durum wheat landraces reveals high diversity and potential for wheat breeding. Plant biotechnology journal 2016;14(9):1800-1812.

(52) Zhang P, Dreisigacker S, Buerkert A, Alkhanjari S, Melchinger AE, Warburton ML. Genetic diversity and relationships of wheat landraces from Oman investigated with SSR markers. Genet Resour Crop Evol 2006;53(7):1351-1360.

(53) López-Arias M, Grau-Corbí JM. Metales pesados, materia orgánica y otros parámetros de la capa superficial de los suelos agrícolas y de pastos de la España peninsular. II. Resultados globales.INIA, MMA, MEC, MAPA.Madrid.[In Spanish] 2005.

(54) Rodriguez-Quijano M, Vazquez JF, Carrillo JM. Variation of high molecular weight glutenin subunits in Spanish landraces of Triticum aestivum ssp. vulgare and ssp. spelta. Journal of Genetics & Breeding 1990;44(2):121-126.

(55) Effects of HMW glutenin subunits on some quality parameters of Portuguese landraces of Triticum aestivum ssp. vulgare. Wheat gluten. Proceedings of the 7th International Workshop Gluten 2000, Bristol, UK, 2-6 April 2000: Royal Society of Chemistry; 2000.

(56) Giraldo P, Rodriguez-Quijano M, Simon C, VÃ¡izquez JF, Carrillo JM. Allelic variation in HMW glutenins in Spanish wheat landraces and their relationship with bread quality. Spanish Journal of
(57) Iqbal M, Navabi A, Yang R, Salmon DF, Spaner D. The effect of vernalization genes on earliness and related agronomic traits of spring wheat in northern growing regions. Crop Sci 2007;47(3):1031-1039.

(58) Cavanagh CR, Chao S, Wang S, Huang BE, Stephen S, Kiani S, et al. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proceedings of the national academy of sciences 2013;110(20):8057-8062.

(59) Muller T, Schierscher-Viret B, Fossati D, Brabant C, Schori A, Keller B, et al. Unlocking the diversity of genebanks: whole-genome marker analysis of Swiss bread wheat and spelt. Theor Appl Genet 2018 Feb;131(2):407-416.

(60) Soriano JM, Villegas D, Aranzana MJ, Garcia Del Moral, L. F., Royo C. Genetic Structure of Modern Durum Wheat Cultivars and Mediterranean Landraces Matches with Their Agronomic Performance. PLoS One 2016 Aug 11;11(8):e0160983.

(61) Gadea M. Trigos españoles. : Instituto Nacional de Investigaciones Agronómicas; 1954.

(62) Andreu JP. Las innovaciones biológicas en la agricultura española antes de 1936: el caso del trigo. Agricultura y Sociedad, 1998:163-164.

(63) Lopes MS, El-Basyoni I, Baenziger PS, Singh S, Royo C, Ozbek K, et al. Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change. J Exp Bot 2015;66(12):3477-3486.

(64) Nazco R, Villegas D, Ammar K, Peña RJ, Moragues M, Royo C. Can Mediterranean durum wheat landraces contribute to improved grain quality attributes in modern cultivars? Euphytica 2012;185(1):1-17.

(65) Vázquez JF, Chacón EA, Carrillo JM, Benavente E. Grain mineral density of bread and durum wheat landraces from geochemically diverse native soils. Crop and Pasture Science 2018;69(4):335-346.

(66) Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, et al. Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat Biotechnol 2012;30(1):105.

(67) Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J. Positional cloning of the
wheat vernalization gene VRN1. Proceedings of the National Academy of Sciences 2003;100(10):6263-6268.

(68) Sourdille P, Snape JW, Cadalen T, Charmet G, Nakata N, Bernard S, et al. Detection of QTLs for heading time and photoperiod response in wheat using a doubled-haploid population. Genome 2000;43(3):487-494.

(69) Bread wheat (Triticum aestivum L.) core collection based in Spanish landraces. Proceedings 8th International wheat conference; Junio 2010; : Dzyubenko NI; 2010.

(70) Doyle JJ, Doyle JL, Hortorium LB. ISOLATION OE PLANT DNA FROM FRESH TISSUE. 1990.

(71) Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, et al. Diversity arrays technology: a generic genome profiling technology on open platforms. : Springer; 2012. p. 67-89.

(72) Chen F, Gao M, Zhang J, Zuo A, Shang X, Cui D. Molecular characterization of vernalization and response genes in bread wheat from the Yellow and Huai Valley of China. BMC plant biology 2013;13(1):199.

(73) Zhang XK, Xiao YG, Zhang Y, Xia XC, Dubcovsky J, He ZH. 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(2):458-470.

(74) McIntosh RA, Yamazaki Y, Devos KM, Dubcovsky J, Rogers WJ, Appels R. Catalogue of gene symbols for wheat. Wheat Information Service 2003;97:27-37.

(75) R Core Team. R: A language and environment for statistical computing. 2014.

(76) Rosenberg NA. DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Notes 2004;4(1):137-138.

(77) Gruber B, Unmack Pj, Berry OF, Georges A. dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. Molecular Ecology Resources 2018;18(3):691-699.

(78) Nei M. Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences 1973;70(12):3321.

(79) Goudet J. Hierfstat, a package for R to compute and test hierarchical F-statistics. Molecular
Ecology Notes 2005;5(1):184-186.

(80) Jost L. GST and its relatives do not measure differentiation. Mol Ecol 2008;17(18):4015-4026.

(81) Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution 1984;38(6):1358-1370.

(82) Pembleton LW, Pembleton ML. Package ‘StAMPP’. 2017.

Tables
Table 1. Number of SNP and DArT markers identified in T. turgidum and T. aestivum accessions. The total number of markers before and after filtering, and their distribution in genomes and chromosomes is presented.
|                | Triticum turgidum |                | Triticum aestivum |                |                |                |                |
|----------------|------------------|----------------|------------------|----------------|----------------|----------------|----------------|
|                | Raw DARTs       | Filtered DARTs | Raw SNPs         | Filtered SNPs  | Raw DARTs      | Filtered DARTs | Raw SNPs       | Filtered SNPs  |
| Total          | 98,983          | 38,700         | 51,751           | 9,324          | 130,899        | 44,241         | 58,660         | 8,238          |
| Mapped         | 61,871          | 25,525         | 41,463           | 7,776          | 46,633         | 16,071         | 34,482         | 4,737          |
| Not mapped     | 37,112          | 13,175         | 10,288           | 1,548          | 84,266         | 28,170         | 24,178         | 3,501          |
| Genome A       | 26,236          | 10,523         | 18,397           | 3,431          | 16,117         | 5,957          | 12,762         | 1,958          |
| Genome B       | 29,818          | 12,696         | 19,569           | 3,733          | 17,730         | 6,982          | 13,623         | 1,962          |
| Genome D       | 5,817           | 2,306          | 3,497            | 612            | 12,786         | 3,132          | 8,097          | 817            |
| 1A             | 2,691           | 1,109          | 1,956            | 433            | 1,841          | 723            | 1,594          | 285            |
| 1B             | 4,070           | 1,734          | 2,676            | 569            | 2,430          | 951            | 1,886          | 278            |
| 1D             | 644             | 252            | 432              | 87             | 1,795          | 384            | 1,046          | 101            |
| 2A             | 4,098           | 1,439          | 2,883            | 549            | 2,551          | 891            | 2,072          | 302            |
| 2B             | 5,175           | 2,092          | 3,485            | 651            | 3,188          | 1,272          | 2,500          | 345            |
| 2D             | 1,074           | 408            | 664              | 105            | 2,499          | 773            | 1,627          | 156            |
| 3A             | 3,664           | 1,434          | 2,627            | 472            | 2,258          | 687            | 1,923          | 316            |
| 3B             | 4,688           | 1,999          | 3,164            | 588            | 2,740          | 1,059          | 2,245          | 330            |
| 3D             | 847             | 315            | 497              | 97             | 1,879          | 406            | 1,277          | 106            |
| 4A             | 3,598           | 1,462          | 2,244            | 368            | 2,223          | 877            | 1,453          | 167            |
| 4B             | 2,396           | 1,009          | 1,519            | 341            | 1,494          | 491            | 1,028          | 145            |
| 4D             | 327             | 126            | 200              | 38             | 1,033          | 170            | 515            | 71             |
| 5A             | 3,594           | 1,345          | 2,738            | 591            | 2,135          | 701            | 1,749          | 330            |
| 5B             | 4,142           | 1,626          | 2,626            | 532            | 2,572          | 1,007          | 1,948          | 299            |
| 5D             | 746             | 262            | 489              | 100            | 1,731          | 383            | 1,068          | 119            |
| 6A             | 3,175           | 1,351          | 2,072            | 369            | 1,909          | 777            | 1,385          | 203            |
| 6B             | 4,517           | 2,036          | 3,002            | 492            | 2,465          | 1,027          | 1,970          | 283            |
| 6D             | 975             | 417            | 457              | 60             | 1,538          | 432            | 1,013          | 118            |
| 7A             | 5,416           | 2,383          | 3,877            | 649            | 3,200          | 1,301          | 2,586          | 355            |
| 7B             | 4,830           | 2,200          | 3,097            | 560            | 2,841          | 1,175          | 2,046          | 282            |
| 7D             | 1,204           | 526            | 758              | 125            | 2,311          | 584            | 1,551          | 146            |

Table 2. Genetic diversity within populations (Hs) and FST values between populations of *T. turgidum* landraces assessed with SNPs.
### Table 3. Genetic diversity within populations (Hs) and FST values between populations of *T. aestivum* landraces assessed with SNPs.

|        | Pop1  | Pop2  | Pop3  | Pop4  | Pop5  | Pop6  | Pop7  |
|--------|-------|-------|-------|-------|-------|-------|-------|
| **Hs** | 0.186 | 0.048 | 0.253 | 0.089 | 0.169 | 0.272 | 0.104 |
| **FST** | 0.476 | 0.709 | 0.474 | 0.653 | 0.692 | 0.238 | -     |

### Table 4. Overall genetic diversity (Hs) and number of monomorphic SNP markers in the set of reference varieties compared to landraces.

|             | *Triticum turgidum* | *Triticum aestivum* |
|-------------|---------------------|---------------------|
|             | Reference | Landraces | Reference | Landraces |
| **Hs**      | 0.196     | 0.323     | 0.250     | 0.300     |
| **Monomorphic markers** | 3791     | 24        | 1771      | 2         |

**Figures**
Markers density along chromosome 2A. (A) T. turgidum raw (blue) and filtered (red) SNP markers. (B) T. turgidum raw (blue) and filtered (red) DArT markers. (C) Filtered SNPs in T. turgidum (purple) and T. aestivum (yellow).
Figure 2

Average PIC distribution in filtered markers. (A) T. turgidum DArTs. (B) T. aestivum DArTs. (C) T. turgidum SNPs. (D) T. aestivum SNPs.
(A) T. turgidum STRUCTURE plot based on DArT markers. The number below the Pop indicates the number of accessions clustered in each population. (B) Collection sites of the different T. turgidum accessions, colored by their STRUCTURE population assignment. When GPS coordinate data were not available, the coordinate of the capital of the province of origin has been used. T. turgidum subsp. durum landraces are shown as circles, subsp. dicoccon as squares and subsp. turgidum as triangles.
Cluster analysis of *T. turgidum* accessions using PCoA. Accessions from subsp. durum are shown as circles, subsp. dicoccon as squares, subsp. turgidum as triangles and reference varieties as asterisks. (A) Graphical representation of PCo1 and PCo2 for the whole collection of durum wheat. Accessions are colored according to their Vrn-A1 alleles. (B) Graphical representation of PCo1 and PCo2 for subsp. durum accessions, which are colored according to their STRUCTURE population assignment.
(A) T. aestivum STRUCTURE plot based on DArT markers. The number below the Pop indicates the number of accessions clustered in each population. (B) Collection sites of the different T. aestivum accessions, colored by their STRUCTURE population assignment. When
GPS coordinate data were not available, the coordinate of the capital of the province of origin has been used.

Figure 6

Cluster analysis of T. aestivum accessions using PCoA. Accessions from STRUCTURE Pop1 are shown as circles, Pop2 as triangles, Pop3 as squares, Pop 4 as rhombus, and reference varieties as asterisks. (A) Graphical representation of PCo1 and PCo2 for the whole collection of bread wheat. Accessions are colored according to their Vrn-A1 allele. (B) Accessions are colored according to their Glu-B1 allele.
Genetic diversity (Hs) distribution along the genome in the reference varieties and landraces. (A) T. durum. (B) T. aestivum. Fixed regions in the reference materials are marked by red bars. Position of ten key genes in the genome is indicated.

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.

Additional file 1.xlsx
Additional file 7.pdf
Additional file 8.pdf
Additional file 6.pdf
Additional file 3.csv
Additional file 5.csv
Additional file 2.csv
Additional file 4.csv