the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2022) **112,** 897–918 doi: 10.1111/tpj.15952

Insights into breeding history, hotspot regions of selection, and untapped allelic diversity for bread wheat breeding

Heike Lehnert[1,*], Thomas Berner[1], Daniel Lang[2,†], Sebastian Beier[3,‡], Nils Stein[4,5], Axel Himmelbach[4], Benjamin Kilian[6] and Jens Keilwagen[1]

[1]*Institute for Biosafety in Plant Biotechnology, Julius Kuehn Institute, Quedlinburg, Germany,*

[2]*PGSB, Helmholtz Center Munich, German Research Center for Environmental Health, Plant Genome and Systems Biology, Neuherberg, Germany,*

[3]*Research Group Bioinformatics and Information Technology, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany,*

[4]*Research Group Genomics of Genetic Resources, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany,*

[5]*Center of integrated Breeding Research (CiBreed), Department of Crop Sciences, Georg-August-University, Göttingen, Germany,*

[6]*Global Crop Diversity Trust, Bonn, Germany*

Received 22 February 2022; revised 17 August 2022; accepted 22 August 2022; published online 8 September 2022.

*For correspondence (e-mail heike.lehnert@julius-kuehn.de).

†Present address: Bundeswehr Institute of Microbiology, Microbial Genomics and Bioforensics, Munich, Germany

‡Present address: Forschungszentrum Jülich GmbH, Institute of Bio- and Geosciences, Bioinformatics (IBG-4), Jülich, Germany

SUMMARY

Breeding has increasingly altered the genetics of crop plants since the domestication of their wild progenitors. It is postulated that the genetic diversity of elite wheat breeding pools is too narrow to cope with future challenges. In contrast, plant genetic resources (PGRs) of wheat stored in genebanks are valuable sources of unexploited genetic diversity. Therefore, to ensure breeding progress in the future, it is of prime importance to identify the useful allelic diversity available in PGRs and to transfer it into elite breeding pools. Here, a diverse collection consisting of modern winter wheat cultivars and genebank accessions was investigated based on reduced-representation genomic sequencing and an iSelect single nucleotide polymorphism (SNP) chip array. Analyses of these datasets provided detailed insights into population structure, levels of genetic diversity, sources of new allelic diversity, and genomic regions affected by breeding activities. We identified 57 regions representing genomic signatures of selection and 827 regions representing private alleles associated exclusively with genebank accessions. The presence of known functional wheat genes, quantitative trait loci, and large chromosomal modifications, i.e., introgressions from wheat wild relatives, provided initial evidence for putative traits associated within these identified regions. These findings were supported by the results of ontology enrichment analyses. The results reported here will stimulate further research and promote breeding in the future by allowing for the targeted introduction of novel allelic diversity into elite wheat breeding pools.

Keywords: genomic regions under selection, private alleles, plant genetic resources, genebank, population genetics, coverage analysis, crop improvement, genotyping-by-sequencing (GBS), iSelect chip, *Triticum*.

INTRODUCTION

Plant breeding is as old as the first domesticated crops (Venske et al., 2019). However, the nature of plant breeding has changed over time. Pre-domestication cultivation and domestication of wild wheat occurred in the Fertile Crescent (Zohary et al., 2012). About 11 000 years ago, early farmers began selecting plants suitable for agriculture and cultivating diploid and tetraploid wild wheat species (Faris, 2014; Peng et al., 2011; Zeibig et al., 2021). Spontaneous hybridization between domesticated tetraploid wheat and a wild diploid donor of the D genome (*Aegilops tauschii*) resulted in hexaploid wheat (*Triticum aestivum*, $2n = 6x = 48$, BBAADD) (Levy & Feldman, 2022; Sharma et al., 2021). Selection pressure and

genetic bottlenecks during crop improvement reduced genetic diversity within domesticated crops compared with their wild relatives (Peng et al., 2011; Smýkal et al., 2018). Humans have consciously or unconsciously selected against alleles associated with unfavorable traits during cultivation, and new alleles, allele combinations, or genes have emerged in farmers' fields through mutation, hybridization, and recombination (Belderok, 2000; Peng et al., 2011; Reif et al., 2005). Over time, natural and artificial selection have resulted in regional landraces consisting of heterogeneous mixtures of different genotypes, which were well adapted to their growing area and farmers' preferences (Belderok, 2000; Faris, 2014; Lupton, 2014). Since the second half of the 19th century, new scientific findings have paved the way for systematic plant breeding and genetics (Belderok, 2000; Lupton, 2014). This has resulted in more uniform and high-yielding traditional cultivars essentially representing only one genotype (Faris, 2014). Since then, breeding progress has accelerated (Venske et al., 2019). In the 1950s and 1960s, the introduction of semi-dwarfing genes (reduced height genes, *Rht*) from Japanese wheat varieties into global elite breeding pools and improvements in agricultural practices (Belderok, 2000; Hedden, 2003; Vergauwen & De Smet, 2017) led to a production increase of 250% over the last 50 years (222 M tonnes in 1961, 776 M tonnes in 2019; FAOSTAT, 2019). In the same period, the global wheat production area increased by only 5% (204 M ha in 1961, 215 M ha in 2019; FAOSTAT, 2019).

The human population is continually growing and is estimated to reach 10 billion by 2050 (Hickey et al., 2019). Consequently, the worldwide demand for food will steadily increase (Tilman et al., 2011). However, trends in temperature and precipitation have changed in global wheat growing areas since 1980 (Lobell et al., 2011). It is understood that the global climate change in combination with an increasing number of extreme weather events negatively impacts wheat yield, in some regions such as Northern and Western Europe, Australia, and Africa (Ray et al., 2019; Trnka et al., 2015, Zhu & Troy, 2018).Today, it is assumed that the uniformity and high-yielding characteristics of today's modern wheat cultivars increase the risk of genetic vulnerability to abiotic and biotic stresses (Fu, 2015; Rahman et al., 2020).

Therefore, there is a strong need for new, well-adapted genotypes with increased yield stability. Systematic plant breeding is one of the most promising approaches to meet these future challenges (Hickey et al., 2019) and to respond to emerging opportunities, e.g., utilization of genetic differences in yield response to increased CO_2 concentration (Marcos-Barbero et al., 2021; Uprety et al., 2009). Genetic diversity is the fundamental prerequisite to select new well-adapted and high-yielding genotypes (Voss-Fels et al., 2015). However, it is questionable whether the genetic diversity of the current elite breeding pool of wheat is broad enough to cope with these challenges (Cavanagh et al., 2013; Venske et al., 2019). By contrast, *ex situ* genebanks are assumed to be rich reservoirs of unexploited plant genetic diversity (Keilwagen et al., 2014). Worldwide, the largest collections of *Triticum* and *Aegilops* accessions are held by the CIMMYT genebank (more than 140 000) or the Lieberman Germplasm Bank (more than 7500), respectively (Sharma et al., 2021). In Germany, the Leibniz Institute of Plant Genetics and Crop Plant Research in Gatersleben holds a collection of wheat (genera *Triticum* and *Aegilops*) plant genetic resources (PGRs) consisting of approximately 30 000 accessions of wild and primitive wheats, traditional cultivars/landraces, and advanced/improved cultivars (Börner et al., 2010). These PGRs are assumed to be a valuable source of genes and alleles that are not yet available in plant breeding pools, e.g., genes or alleles associated with pathogen resistance or abiotic stress tolerance (Lopes et al., 2015; Winfield et al., 2018). However, it is difficult to identify this useful genetic diversity available in the PGRs and transfer it into elite breeding pools (Mascher et al., 2019; Mondal et al., 2016). This is mainly because of the huge linkage drag that is expected in crosses between elite breeding materials and PGRs (Rasheed et al., 2018; Singh et al., 2018).

Much progress has been made in wheat genetics and genomics in the last decade (Chung et al., 2017; Jia et al., 2018). Advances in next-generation sequencing (NGS) technologies have provided whole genome sequencing data and genome-wide single nucleotide polymorphisms (SNPs) for genotyping (Chung et al., 2017; Kilian & Graner, 2012). This has paved the way for the production of wheat reference genomes (IWGSC, 2018; Walkowiak et al., 2020) and high-density SNP marker sets (generated from SNP chips, i.e., Illumina or Affymetrix [Cavanagh et al., 2013; Qaseem et al., 2018; Soleimani et al., 2020; Wang et al., 2014; Winfield et al., 2016], or using NGS-based approaches, e.g., genotyping-by-sequencing [GBS] and exome capture [Poland & Rife, 2012; Winfield et al., 2012]). Such advances have made it possible, for example, to investigate the genetic diversity of diverse collections of bread wheat, to trace the ancestry of modern wheat, and to identify genomic regions associated with signatures of artificial selection (e.g. Cavanagh et al., 2013; GAO et al., 2017; Jordan et al., 2015; Joukhadar et al., 2019; Liu et al., 2019; Pont et al., 2019).

To advance breeding in the future, it is important to have detailed knowledge of genetic diversity, population structure, sources of new allelic diversity, and genomic regions affected by previous and current breeding activities. This will allow for the targeted introduction of new allelic diversity into the elite wheat breeding pool (Fu, 2015; Lopes et al., 2015).

Here, we analyzed a diverse winter wheat collection consisting of 81 modern cultivars (MC) and 209 genebank accessions (GA) from the *ex situ* Genebank in Gatersleben, Germany, with the following objectives: (i) to characterize genetic diversity within the collection; (ii) to determine the population structure and the level of genetic diversity; and (iii) to identify private alleles and genomic regions under selection as a proxy for future sustainable breeding strategies.

RESULTS AND DISCUSSION

Genotyping

The whole collection was genotyped using the GBS approach and the bread wheat 15K+5K iSelect chip. These analyses yielded a final combined marker dataset consisting of 49 181 high-quality SNP markers, of which 37 914 and 11 140 SNPs were exclusively detected by GBS and iSelect markers, respectively. Only 127 markers were identified by both approaches.

The minor allele frequency (MAF) ranged between 0.05 and 0.50 and the polymorphism information content (PIC) ranged between 0.09 and 0.375 (Figure S1a,b). However, markers with a low MAF (0.05–0.10) were enriched (27%) in the dataset, especially among the GBS markers. In total, 32% of GBS markers and 9% of iSelect markers had a low MAF value (Figure S2). Therefore, GBS markers were considered to be less informative, because of the lower proportion of highly informative markers (PIC > 0.30: 36%) and lower average expected heterozygosity (H_e: 0.29) compared with the iSelect markers (PIC > 0.30: 58%; H_e: 0.37) (Figure S2). Overall, observed heterozygosity (H_o) and H_e ranged between 0.0 and 0.12 and between 0.09 and 0.50, respectively (Figure S1c,d). The majority of markers (87%) showed $H_o < 0.05$, as expected for a self-pollinating species.

The markers were unequally distributed across subgenomes and chromosomes (Figure 1). The number of markers per chromosome ranged from 247 (4D) to 4285 (2B) (Table S2a). The D genome had fewer markers (7434), poorer marker coverage (1.8 markers per Mbp), and a larger average marker interval (0.74 Mbp) compared with the A genome (19 115 markers, 3.8 markers per Mbp, 0.26 Mbp) and the B genome (22 632 markers, 4.2 markers per Mbp, 0.25 Mbp) (Figure S3). For all chromosomes, it was demonstrated that combining both GBS and iSelect markers increased the number of markers per chromosome (Table S2a; Figures S3 and S4) and decreased the average marker interval per chromosome (Table S2a; Figures S3 and S4). Nevertheless, the centromeric chromosomal regions remained insufficiently covered by markers (Figure 1).

Based on the annotations of the reference genome (IWGSC, 2018), 43% of the SNP markers were located in genomic regions associated with a high-confidence (HC) gene. Of those markers, 62% were located in the coding sequence (CDS) of the gene (Figure S5). Interestingly, 32% of the GBS markers and 80% of the iSelect markers were located within a region associated with an HC gene (Figure S5).

Currently, SNP markers are widely used to assess the genetic diversity and population structure of diverse collections (e.g., Alipour et al., 2017; Maccaferri et al., 2019; Rufo et al., 2019). It is known that SNP markers obtained by different marker platforms differ in their suitability to evaluate genetic diversity in collections of GA, but did not have a significant effect on population structure (Chu et al., 2020). In this study, two genotyping approaches were used to obtain a high-density and informative SNP marker set. It could be shown that the overlap between the two marker sets was low, so combining both sets improved the marker coverage per chromosome. Furthermore, the combination of both sets made it possible to exploit the advantages of both genotyping approaches, i.e., the better suitability of GBS markers to detect rare alleles, which is an advantage when analyzing GA, and the frequent association of iSelect markers with HC genes. Nevertheless, it is also known that the use of array-based markers can be associated with an increased risk of ascertainment bias (Albrechtsen et al., 2010; Chu et al., 2020), due to the design of the array. Therefore, the iSelect and GBS markers were also separately analyzed. For the genotype panel under investigation, it could be demonstrated that the ascertainment bias is negligible and that population structure and genetic diversity were not influenced by the single or the combined use of iSelect and GBS markers (Text S1). Based on these findings, the combined marker dataset was used for all further analyses.

Population structure and genetic diversity

To reveal genotypic diversity and to determine population structure, the combined marker dataset was linkage disequilibrium (LD)-pruned, resulting in a reduced marker set consisting of 6116 markers (Text S2).

In a principal coordinate analysis (PCoA), 4 and 2% of variance was explained by the first and second principal coordinates (PCos), respectively. On the PCoA scatterplot, the MC group was separated from the GA group, with some overlap of the two groups at their peripheries (Figure 2a). Genetic diversity in the GA group compared to the MC group was assessed. Compared with the MC group, the GA group showed a higher average number of alleles per marker and higher H_e (Table S3a).

Next, analysis of molecular variance (AMOVA) was conducted to estimate the amount of genetic variance explained by differences between and within the MC and GA groups. Genetic differentiation between the MC group and the GA group was moderate with a pairwise fixation

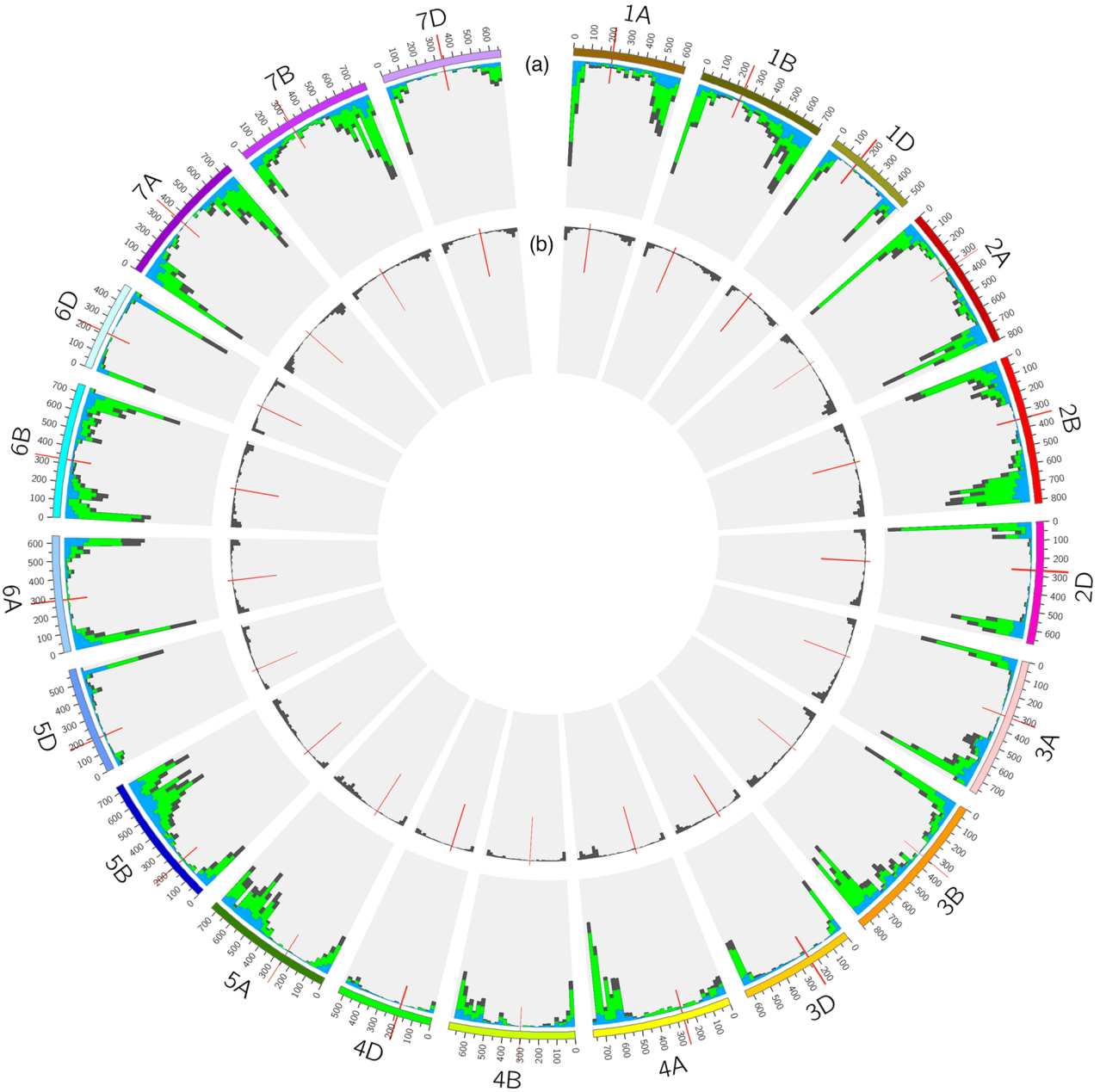

Figure 1. Distribution of markers across the wheat genome. (a) Gray bars indicate the distribution of 49 181 single nucleotide polymorphism (SNP) markers included in the combined marker dataset. Green bars indicate the distribution of 38 041 SNP markers identified by genotyping-by-sequencing (GBS) analysis. Blue bars indicate the distribution of 11 267 SNP markers identified by iSelect chip. Both marker sets (GBS and iSelect) include the 127 markers, which were identified by both approaches. (b) Gray bars indicate the distribution of 6116 SNP markers in the reduced marker set. Red vertical lines highlight centromeres.

index value (F_{st}) of 0.12. The genetic variation among genotypes within groups was greater than the genetic variation between groups (Figure 3a).

Furthermore, a STRUCTURE analysis was conducted. ΔK values revealed an optimal number of $K = 2$ subpopulations (Figure S6). Based on membership coefficients, 157 and 133 genotypes were assigned to STRUCTURE groups I and II, respectively. STRUCTURE group I comprised all MC and 76 GA, which mainly originated from Germany or Western Europe and, to a lesser extent, from South-Eastern Europe, Asia, and America (Figure 2c). STRUCTURE group II consisted of GA that mainly originated from South-Eastern Europe and Asia (Figure 2c).

To compare the results of the STRUCTURE analysis and PCoA, the STRUCTURE grouping was projected onto the results of the PCoA. The first PCo clearly separated

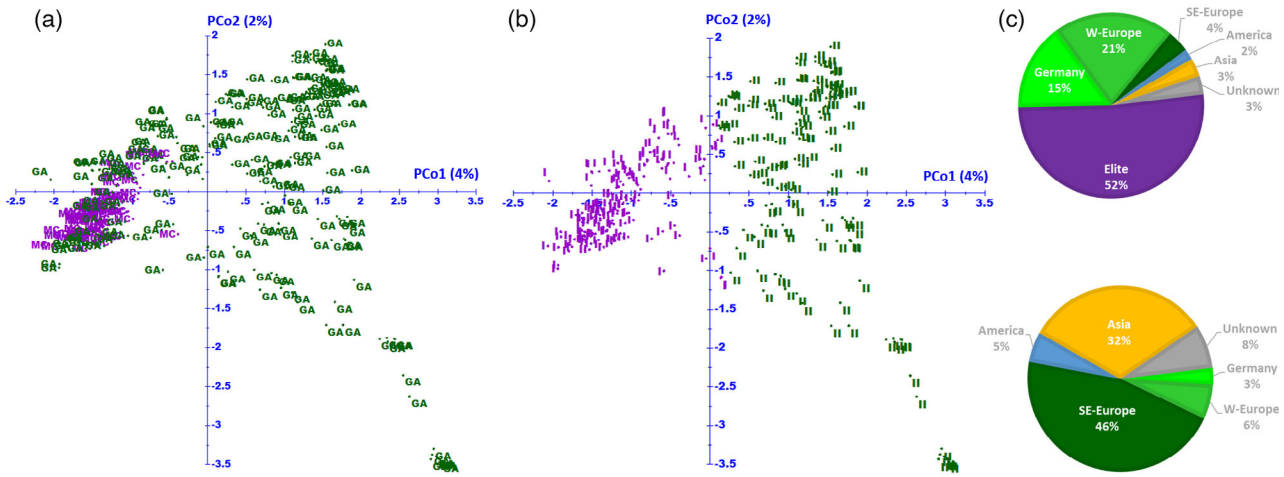

Figure 2. Population structure analysis of a panel of 290 winter wheat genotypes based on reduced set of 6,116 single nucleotide polymorphism (SNP) markers. First two principal coordinates explain 4% and 2% of total variance. (a) Principal coordinate analysis (PCoA) scatterplot according to group of modern cultivars (MC; $n = 81$, shown in purple) and group of genebank accessions (GA; $n = 209$, shown in dark green). (b) PCoA plot according to STRUCTURE grouping for the uppermost level of population structure ($K = 2$). STRUCTURE group I is shown in purple; STRUCTURE group II is shown in dark green. (c) Geographical origin of genotypes assigned to STRUCTURE groups I (top right) and II (bottom right).

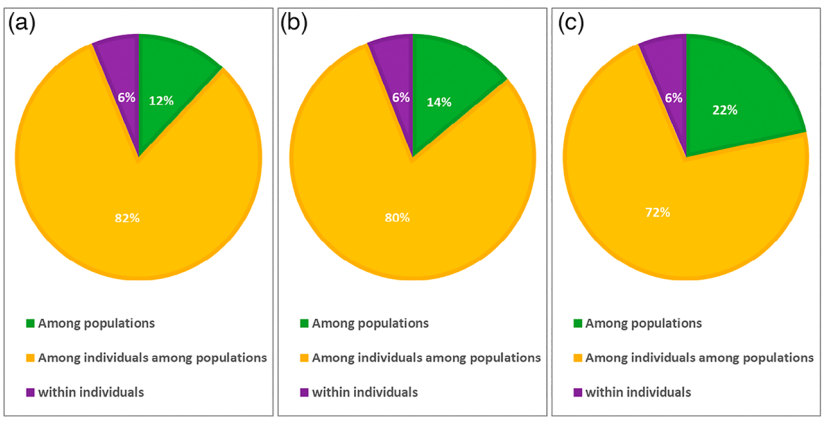

Figure 3. Analysis of molecular variance (AMOVA) results showing the amount of molecular variance explained by differences between and within (a) group of modern cultivars and group of genebank accessions; (b) STRUCTURE group I and STRUCTURE group II ($K = 2$); and (c) STRUCTURE groups 1 to 7 ($K = 7$).

STRUCTURE groups I and II (Figure 2b), thus confirming the STRUCTURE results. The average number of alleles per marker did not differ between the two STRUCTURE groups. However, H_e was higher for STRUCTURE group II than for STRUCTURE group I (Table S3b), suggesting differences in terms of genetic diversity. Genetic differentiation between the two STRUCTURE groups was moderate with a pairwise F_{st} of 0.14. The genetic variation among genotypes within groups (80%) was greater than genetic variation between the two STRUCTURE groups (14%) (Figure 3b).

The STRUCTURE analysis revealed the upper level of subpopulations. However, because there were diverse genotypes in the collection, a more complex subpopulation structure was expected. In this regard, there were two less prominent ΔK peaks at $K = 5$ and $K = 8$, suggesting a lower level of subpopulations (Figure S6). Therefore, a model-free discriminant analysis of principal components (DAPC) was conducted to reveal the lower level of subpopulations.

The optimal number of subpopulations was determined using the K-means clustering algorithm. On the basis of the lowest Bayesian information criterion (BIC) value, the optimal number of subpopulations was $K = 7$ (Figure S7). A DAPC was run to describe the subpopulation groups (Figure S8), and the results of DAPC and STRUCTURE for $K = 7$ were compared. The assignment of genotypes to groups according to STRUCTURE or DAPC was highly correlated, and most genotypes (82%) were assigned to identical groups in the two analyses. The exceptions were mainly genotypes identified as admixed in the STRUCTURE analysis. The results of the STRUCTURE analysis are described in detail below.

In total, 254 of the genotypes showed a membership coefficient of >0.5 to one of the STRUCTURE groups and were assigned to STRUCTURE groups 1 to 7, whereas 36 genotypes were considered to be admixed. The MC were grouped together in STRUCTURE group 5 except for one

cultivar. The GA were mainly assigned to six STRUCTURE groups (1–4, 6, and 7), which differed in the number of genotypes included (14–65) and the origin of genotypes. In accordance with the STRUCTURE grouping for $K = 2$, the first PCo separated the Western European geographical groups (STRUCTURE groups 4–6) from the South-Eastern and Asian geographical groups (STRUCTURE groups 1–3 and 7) (Figure 4). The genotypes included in the Western European geographical groups differed in their geographical origin and their breeding history.

STRUCTURE group 5 included two GA (TRI 11938, TRI 3808) and all the MC except for cv. Solehio (membership coefficient for STRUCTURE group 5: 0.45), which was assigned to the admixed group. Whereas most of the MC in this study were bred for cultivation under German growing conditions (e.g., BSL, 2015), cv. Solehio is an early flowering French cultivar (Arenas-Corraliza et al., 2019; Schittenhelm et al., 2020) that is well adapted to Southern European growing conditions. Furthermore, it is known that the traditional cv. Hadmerslebener VIII (TRI 3808) is included in the pedigree of more than half of the MC investigated in this study (at least 43 out of 81) (Martynov & Dobrotvorskyi, 2012). In contrast, there are no hints that also cv. Maris Mardler (TRI 11938) is prominently represented in the pedigrees of these MC. However, this accession shows a large proportion of the traditional cv. Capelle-Desprez and cv. Wisconsin-245 in its pedigree (Martynov & Dobrotvorskyi, 2012), which are also prominently represented in pedigrees of these MC, as reported previously (Fradgley et al., 2019).

STRUCTURE groups 4 and 6 mainly consisted of genotypes from Germany (4: 54%, 6: 4%) and Western Europe (4: 24%, 6: 86%), respectively (Figure 4). As far as known, most of these accessions are traditional cultivars that were bred before the introduction of semi-dwarfing genes (*rht-B1b* and *rht-D1b*) into the European breeding pools in the 1950s and 1960s by line selection from landraces or hybridization of favorable genotypes since the end of the 19th century (Lupton, 2014; Martynov & Dobrotvorskyi, 2012; Vergauwen & De Smet, 2017).

STRUCTURE group 4 mainly included traditional cultivars developed in Germany before 1950 and other traditional cultivars bred at the same time originating from Europe, the USA, or Russia. Some of these traditional cultivars ranked among the most important cultivars at that time and were widely grown, i.e., cv. Heinrichs von Hindenburg (TRI 994, Germany), cv. Grundmanns Wotan (TRI 4798, Germany), cv. Nordost Sandomir (TRI 243, Germany) (Pronin et al., 2020), cv. Svalöfs Grenadiere (TRI 5108, Sweden) (Nilsson-Ehle, 1913), and cv. Mansholts Weißer Dickkopf I (TRI 5042, Netherlands) (Zeven, 1990).

STRUCTURE group 6 included traditional cultivars that mainly originated from France and could be traced back to one common ancestor represented in all pedigrees, i.e., the traditional cv. Noe (Martynov & Dobrotvorskyi, 2012). This traditional cultivar is described as a selection from the Russian landrace Odessa (Martynov et al., 2006). Because of its favorable characteristics, cv. Noe and its descendants were frequently used in French winter wheat breeding. Therefore cv. Noe is present in the pedigrees of most traditional cultivars developed in France until 1950 (Belderok et al., 2000; Bonnin et al., 2014). The descendants of cv. Noe included in STRUCTURE group 6, i.e., cv. Vilmorin 27 (TRI 6726), cv. Hybride 40 (TRI 7343), cv. Tadepi (TRI 6870), cv. Vague d'Epis (TRI 6874), and cv. Cappelle Desprez (TRI 5164), were grown on large scale in France at that time (Belderok et al., 2000). The cv. Cappelle Desprez was the outstanding cultivar of its time and combined disease resistance, adequate baking quality, and high grain yield (Belderok et al., 2000). Therefore, cv. Cappelle Desprez was frequently used in European breeding programs (Agenbag et al., 2012). Consequently, it is frequently included in pedigrees of today's European MC (Fradgley et al., 2019). Accordingly, 88% of the MC investigated in this study had cv. Cappelle Desprez in their pedigrees. It is presumed that STRUCTURE groups 4 and 6 represent the breeding history of Germany and France, as the traditional cultivars developed in these two countries form the leitmotif of these STRUCTURE groups.

STRUCTURE groups 1–3 mainly included GA from South-Eastern Europe (1: 81%, 2: 55%, 3: 36%) and Asia (1: 6%, 2: 23%, 3: 36%), whereas STRUCTURE group 7 contained only GA from Asia (94%) or with unknown origin (6%) (Figure 4).

Landraces from China, Tibet, India, Nepal, and Pakistan clustered together in one group (STRUCTURE group 7), distinct from the other landraces and traditional cultivars from Asia and South-Eastern Europe. The majority of the landraces in STRUCTURE group 7 were collected during two expeditions to the Himalayan region in the last century (Knüpffer et al., 2001; Witcombe, 1975). It is known that landraces from the Himalayan region are distinct from those originating from other Asian and European regions because of their specific environmental adaptions (Guo et al., 2020; Stodart et al., 2008). The remaining landraces and traditional cultivars originating from Asia and South-Eastern Europe were distributed in three groups according to their breeding history. In detail, STRUCTURE groups 1 and 3 mainly included short-statured and early flowering traditional cultivars from the Mediterranean region or Russia and Eastern Europe, respectively. In contrast, landraces and traditional cultivars from America, Asia, and South-Eastern Europe, which were mainly tall and early flowering, clustered together in the diverse STRUCTURE group 2. These findings highlight that differentiation among groups is associated with the presence or absence of dwarfing genes, particularly *rht8* and the closely linked photoperiod insensitivity (*Ppd-D1*) gene (Martynov & Dobrotvorskyi,

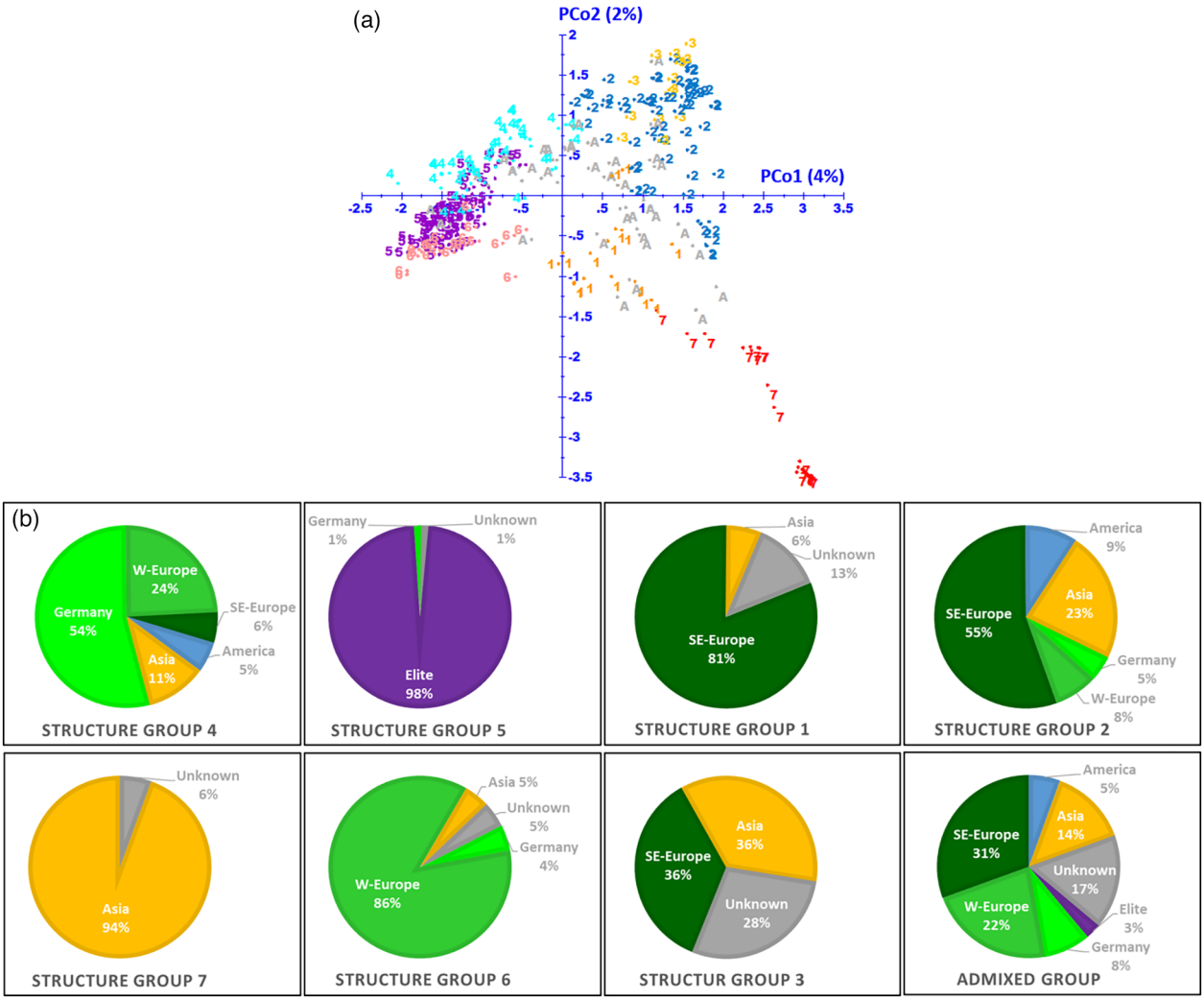

Figure 4. Population structure analysis of a panel of 290 winter wheat genotypes based on the reduced set of 6116 SNP markers. (a) Principal coordinate (PCo) plot according to the STRUCTURE grouping at $K = 7$. The first two principal coordinates explain 4 and 2% of total variance. STRUCTURE groups are highlighted in color, as follows: orange: STRUCTURE group 1, blue: STRUCTURE group 2, yellow: STRUCTURE group 3, turquoise: STRUCTURE group 4, purple: STRUCTURE group 5, pink: STRUCTURE group 6, red: STRUCTURE group 7, gray: admixed group. (b) Geographical origin of genotypes assigned to STRUCTURE groups 1 to 7 and the admixed group.

2012). The *rht8* gene was the first dwarfing gene which was introduced to Europe (Borojevic & Borojevic, 2005). Today, it is widely distributed among South-Eastern European and Russian cultivars (Worland et al., 1998). In 1913, it was introduced into Italian breeding pools by crosses with the Japanese donor cv. Akakomugi (carrying *rht8*, *rht9*, and *Ppd-D1*) (Pujol Andreu, 2011; Salvi et al., 2013; Sanchez-Garcia & Bentley, 2019), which resulted in a number of new short-statured and early flowering cultivars (e.g. cv. Villa Glori, cv. Ardito, cv. Mentana, and cv. Damiano) (Borojevic & Borojevic, 2005). It is assumed that STRUCTURE group 1 reflected these early breeding activities, as cv. Villa Glori (TRI 5400) and descendants of the other early Italian cultivars were mainly included in this group. Later, these early Italian carriers of *rht8* and the descendants thereof were included in pedigrees of wheat cultivars developed in Argentina, the former USSR, and Eastern Europe, and were the founders of short-statured and high-yielding cultivars (Borojevic & Borojevic, 2005). Therefore, it is supposed that STRUCTURE group 3 reflects the breeding activities in the former USSR, that is, the introduction of the *rht8* gene and the widespread use of the semi-dwarf cv. Bezostaya 1 (Borojevic & Borojevic, 2005). The cv. Bezostaya 1 was the main donor of the *rht8* gene in breeding programs in South-Eastern Europe and South Russia (Divashuk et al., 2013) and is therefore present in the pedigrees of many modern Russian cultivars (Martynov et al., 2006). Accordingly, cv. Bezostaya 1 (TRI 6747) and its descendants were dominant in STRUCTURE group 3.

To conclude, the geographical origin was the main factor causing population differentiation in this study, and it divided the collection in two geographical groups (STRUCTURE groups I and II). A more complex subpopulation structure further separated the two geographical groups (STRUCTURE groups I and II) into seven groups, consisting of three subgroups (STRUCTURE groups 4–6), four subgroups (STRUCTURE groups 1–3 and 7), and an admixed group. The MC were assigned to a single group (STRUCTURE group 5), while most of the GA were distributed among six groups (STRUCTURE groups 1–4, 6, and 7). These findings provide further evidence that geographical origin, biological status, and breeding history are the main drivers for population differentiation in diverse collections (Cavanagh et al., 2013; Pont et al., 2019; Rufo et al., 2019).

Genetic differentiation among the seven STRUCTURE groups was moderate to high and ranged between 0.17 (STRUCTURE group 4 versus STRUCTURE group 5) and 0.51 (STRUCTURE group 6 versus structure group 7) (Table 1). The degree of differentiation between the MC group (STRUCTURE group 5) and the groups including GA (STRUCTURE groups 1–4, 6, and 7) was also moderate to high (Table 1). In detail, genetic differentiation was moderate between the MC group and STRUCTURE group 4 (F_{st}: 0.17) and group 6 (F_{st}: 0.18), but high between the MC group and STRUCTURE group 2 (F_{st}: 0.22), group 1 (F_{st}: 0.25), group 3 (F_{st}: 0.31), and group 7 (F_{st}: 0.42) (Table 1). According to the AMOVA results, the genetic variation was greater among genotypes within STRUCTURE groups (72%) than among the STRUCTURE groups (22%) (Figure 3c).

As expected, there was a closer relationship between the MC group and the two groups containing GA of Western European origin (STRUCTURE groups 4 and 6) than between the MC group and the groups containing GA of South-Eastern and Asian origin (STRUCTURE groups 1–3 and 7). This is not surprising, because the local landraces from Northern and Western Europe and the traditional cultivars derived from them were used as founder lines for the development of the current elite wheat breeding pools (Balfourier et al., 2019). For instance, it is well known that traditional German cultivars trace back to regional German landraces and plant materials introduced from France, England, Russia, Sweden, and the USA (Lupton, 2014). This largely explains the composition of STRUCTURE group 4.

On the basis of our results and information in the literature, it is assumed that the genetic differentiation within the Western European genebank groups (STRUCTURE groups 4 and 6) and the South-Eastern and Asian genebank groups (STRUCTURE groups 1–3 and 7) has been caused by: (i) old selection patterns associated with the adaptation of genotypes to the prevailing environmental conditions in different growing regions (Balfourier et al., 2019), (ii) the specific adaptation of landraces to the Himalayan region (Guo et al., 2020), and (iii) breeding activities associated with the introduction of the *rht8* gene (Körmöczi et al., 2019; Nielsen et al., 2014).

There is much debate about the genetic diversity of elite breeding pools (Alipour et al., 2017; Cavanagh et al., 2013; Liu et al., 2019; Pont et al., 2019; Rufo et al., 2019). In general, it is assumed that the genetic diversity of elite breeding pools has followed spatial and temporal trends during the last century (Rauf et al., 2010). Due to the intensive exchange of breeding material, pedigrees of modern elite cultivars are complex (Fradgley et al., 2019). However, it should also be noted that only a limited number of founder lines was used for the development of today's breeding pools (Lopes et al., 2015). Although it is understood that strong and extensive selection during breeding activities in the last century has narrowed the genetic diversity of today's breeding pools (Rauf et al., 2010), systematic breeding activities have introduced novel allelic diversity into elite breeding pools (Balfourier et al., 2019; Mondal et al., 2016; Sharma et al., 2021). In this study, genetic diversity was higher for the worldwide group of GA compared to the MC group released in Western Europe (mainly Germany) during the last two decades. These findings initially indicate that there is reduced genetic diversity in the elite breeding pool, which is in accordance with findings of Alipour

Table 1 Pairwise genetic differentiation (F_{st}; down left diagonal) and gene flow (Nm; upper right diagonal) for comparisons between STRUCTURE groups 1–7 and the admixed group based on 49 181 SNP markers

F_{st}/Nm	STRUCTURE group 5	STRUCTURE group 2	STRUCTURE group 4	Admixed group	STRUCTURE group 6	STRUCTURE group 7	STRUCTURE group 1	STRUCTURE group 3
STRUCTURE group 5	–	1.8	2.37	3.42	2.32	0.68	1.53	1.14
STRUCTURE group 2	0.22	–	2.13	7.06	1.32	1.15	1.9	2.3
STRUCTURE group 4	0.17	0.19	–	4.33	1.43	0.62	1.63	1.08
Admixed group	0.13	0.07	0.10	–	2.40	1.37	3.77	2.35
STRUCTURE group 6	0.18	0.28	0.26	0.17	–	0.49	1.21	0.72
STRUCTURE group 7	0.42	0.30	0.45	0.27	0.51	–	0.85	0.55
STRUCTURE group 1	0.25	0.21	0.23	0.12	0.29	0.37	–	1.27
STRUCTURE group 3	0.31	0.18	0.32	0.18	0.41	0.47	0.28	–

et al. (2017), Liu et al. (2019), and Pont et al. (2019). However, if we consider the population structure when evaluating genetic diversity, the picture looks rather different. The spatial and temporal trends in genetic diversity become quite clear, confirming the results of previous reports (Balfourier et al., 2019). Surprisingly, the genetic diversity of some STRUCTURE groups containing GA (groups 3, 4, 6, and 7) was lower than the genetic diversity of STRUCTURE group 5 containing the MC (Table 2). To exclude bias due to group size, 14 genotypes per STRUCTURE group were randomly selected. However, group size had no influence on the genetic diversity within STRUCTURE groups (Table S4). STRUCTURE groups 3, 4, 6, and 7 mainly included traditional cultivars or landraces, and therefore only represent the allelic diversity after or before the early breeding bottleneck. Therefore, on the one hand, it is suggested that the lower genetic diversity of STRUCTURE groups 3, 4, and 6 compared with that of the MC group is a result of the genetic bottlenecks associated with early breeding activities (Cavanagh et al., 2013; Pont et al., 2019), i.e., the development of the first traditional cultivars from landraces or the introduction of superior breeding material into the early breeding pools (Lupton, 2014). On the other hand, it is considered that the lower genetic diversity in STRUCTURE group 7 compared to the MC group is likely because the landraces in group 7 are from geographically restricted Himalayan regions (Knüpffer et al., 2001; Witcombe, 1975) and are adapted to the specific environmental conditions in these regions (Guo et al., 2020). In contrast, the genetic diversity of STRUCTURE groups 1 and 2 was comparable to, or higher than, that of STRUCTURE group 5 containing the MC (Table 2). It is assumed that there are different reasons for this. STRUCTURE groups 1 and 2 included landraces and traditional cultivars from particular geographical regions. Therefore, these two groups reflect the allele diversity before and after the early breeding bottleneck. Furthermore, the early Italian breeding history reflected in STRUCTURE group 1, i.e., crosses between Italian and Asian material (Borojevic & Borojevic, 2005; Salvi et al., 2013), is assumed to be associated with an increase in genetic diversity. In this context, Voss-Fels et al. (2015) and Balfourier et al. (2019) reported that PGRs from Asia are a valuable and largely unexploited resource to improve genetic diversity in terms of quantitative disease resistance in European breeding pools.

To gain detailed knowledge about genetic diversity, it is of vital importance to be aware of population structure (Nielsen et al., 2014). Both population structure and genetic diversity can accelerate progress in plant breeding by an effective use of PGRs and the selection of promising parents carrying novel and beneficial allelic diversity (Dempewolf et al., 2017; Kilian et al., 2021; Nielsen et al., 2014).

Genomic regions under selection and private alleles

Both the loss of genetic diversity due to selection or genetic bottlenecks and the introduction of novel genetic variation caused genomic signatures (Joukhadar et al., 2019). Thus, genomic signatures of selection describe genomic regions that are, or have been, targets of breeding efforts. In this study, five different approaches were used to identify loci under selection. In total, 512, 506, 53, 867, and 2764 markers were detected as loci under selection by using the non-hierarchical finite island model (Arlequin), the hierarchical island model (Arlequin), the Bayesian model (BayeScan), the FLK test, and the Pcadapt approach, respectively (Figure 5). Of these, 563 markers were identified by at least three approaches (Figure 5). These loci were defined as putative candidates associated with genomic signatures of selection. In total, 43% of the 563 markers were located in genomic regions associated with an HC gene (Figure S9).

The 563 markers were assigned to 57 genomic regions under selection on 10 wheat chromosomes (Figure 5; Table S5). Nine of the detected genomic regions under selection have been reported in the literature more than twice; 36 genomic regions have been reported in one or two previous studies; and 12 of the detected regions (on chromosomes 2A, 2B, 4A, 4B, 6A, and 7B) did not overlap with a previously described genomic region under selection and were assumed to be novel (Afzal et al., 2019; Cheng et al., 2019; Gao et al., 2017; Jordan et al., 2015; Liu et al., 2019; N'Diaye et al., 2018). The nine frequently

Table 2 Average number of alleles, expected heterozygosity (H_e), and number of polymorphic loci determined for STRUCTURE groups 1 to 7 and admixed group based on 49 181 single nucleotide polymorphism (SNP) markers

Group	STRUCTURE group 5	STRUCTURE group 2	STRUCTURE group 4	Admixed	STRUCTURE group 6	STRUCTURE group 7	STRUCTURE group 1	STRUCTURE group 3
Number of individuals	82	65	37	36	22	18	16	14
Average number of alleles	1.90	1.96	1.80	1.98	1.74	1.67	1.74	1.68
H_e	0.25	0.28	0.21	0.31	0.20	0.18	0.25	0.20
Number of polymorphic loci	44 255	46 977	39 426	48 254	36 207	33 024	36 315	33 506

detected genomic regions under selection contained known quantitative trait locus (QTL) regions (LOD ≥ 5; Blake et al., 2019) associated with the flag leaf stay green period, flag leaf senescence, grain hardiness, duration of grain filling, number of seeds per head, grain yield, lodging, plant height, stem solidness, heading date, and severity of leaf rust, stem rust, and stripe rust (Tables S5 and S6). In contrast, the majority of the remaining regions under selection contained few previously reported QTLs (Tables S5 and S6). This indicates that these regions are associated with traits that are not frequently targeted, but appear to have been actively or passively modified by breeding activities in recent decades.

Approximately 32 or 47% of the identified genomic regions under selection overlapped with frequently or rarely occurring large chromosomal modifications, respectively (Table S5), i.e., introgressions from wheat wild relatives identified by coverage analysis (Keilwagen et al., 2019; Keilwagen et al., 2022). The connection between genomic regions under selection and genomic regions carrying introgressions has so far only been described for the 2B/2G introgression from *Triticum timopheevii* (Cavanagh et al., 2013; Gaire et al., 2020). To the best of our knowledge, besides the 2B/2G introgression, only for the frequently occurring introgressions on chromosome 2A and 2D information about the origin and the underlying genes are available (Table S5).

Genomic signatures on the short and long arms of chromosome 2B (Region_S_chr2B-1 to Region_S_chr2B-11) detected in this study (Table S5) are assumed to be associated with the 2B/2G introgressions from *T. timopheevii* (Keilwagen et al., 2022; Martynov et al., 2018; Walkowiak et al., 2020). The 2B/2G introgressions were mainly introduced via cv. Wisconsin-245 and its descendants into the Western European breeding pools in the 1950s, and to a lesser content from other sources (Martynov et al., 2018). Therefore, it could not be determined whether these introgressions on chromosome 2B were introduced by a single event and were subsequently broken over time by recombination, or whether they arose from several events (Cheng et al., 2019). In any case, the cv. Maris Fundin (TRI 11510) shows a nearly complete substitution of chromosome 2B, completely enclosing the regions that indicate introgressions on the short and long arms. The cv. Maris Fundin was one of the early donors of the 2B/2G introgression, which is associated with the resistance gene complex *Sr36/Pm6* conferring quantitative resistance to stripe rust and powdery mildew (Martynov et al., 2018). In this study, the resistance gene complex *Sr36/Pm6* was located within *Region_S_chr2B-2* under selection on the short arm of chromosome 2B (Table S5). Furthermore, some genomic regions under selection located within the 2B/2G introgression could be associated with QTLs related to resistance *inter alia* for leaf rust, stripe rust, or stem rust (Table S6). In this context, a recent study has shown that the number of cultivars carrying the 2B/2G introgression has steadily increased since the 1950s and that nearly half of all elite cultivars recently released in Western Europe have ancestors carrying 2B/2G introgressions in their pedigrees (Martynov et al., 2018). Consistent with those findings, the 2B/2G introgression was present in approximately 70% of the MC in this study.

The genomic signature detected on chromosome 2A (approximately 9 Mb – 27 Mb) probably represents an introgression from *Aegilops markgrafii* (Region_S_chr2A-1; Table S5) or the $2N^{V}S$ introgression from *Aegilops ventricosa* (Region_S_chr2A-1 to Region_S_chr2A-4; Table S5) (Keilwagen et al., 2022). It is known that the $2N^{V}S$ introgression is associated with higher grain yield and a resistance gene complex that confers resistance to stripe rust, leaf rust, stem rust, wheat blast, and root knot nematode (Gao et al., 2021; Walkowiak et al., 2020). Initially, the $2N^{V}S$ introgression was introduced into elite breeding pools in the 1990s through the wheat cultivar VPM-1 with an aim to increase disease resistance (Gao et al., Gao et al., 2021). Today, 40–80% of US winter wheat and CIMMYT spring wheat cultivars carry the $2N^{V}S$ introgression (Gao et al., 2021), consistent with our findings. About 40% of the MC and none of the GA investigated here carry the $2N^{V}S$ introgression, which indicates the currency of this introgression event (Keilwagen et al., 2019). Therefore, this genomic region on chromosome 2A was also associated with private alleles in the MC group. The genomic signatures on chromosome 2D (Region_S_chr2D-3 to Region_S_chr2D-5; Table S5) seem to be associated with an introgression from either *A. markgrafii* or *Aegilops umbellulata* (Keilwagen et al., 2022).

In conclusion, the genetic diversity of elite breeding pools has been increased by the introduction of introgressions from wheat wild relatives (Cheng et al., 2019), but decreased by the enrichment or fixation of alleles (Cheng et al., 2019; Fradgley et al., 2019). However, it is presumed that these events are associated with an enrichment of agriculturally important genes or alleles in the elite breeding pools (Cheng et al., 2019). Nevertheless, for most genomic regions under selection and genomic regions associated with private alleles in the MC group, there was no information about functional wheat genes or associated QTLs in the literature. Therefore, for these regions, it is impossible to identify the associated trait that was targeted by breeding efforts without additional mapping populations and phenotyping. However, ontology enrichment analyses can provide new insights into putative traits that may have been actively or passively targeted by former and current breeding activities. Therefore, Gene Ontology (GO), Plant Ontology (PO), and Trait Ontology (TO) analyses were conducted to gain insights into the functions of

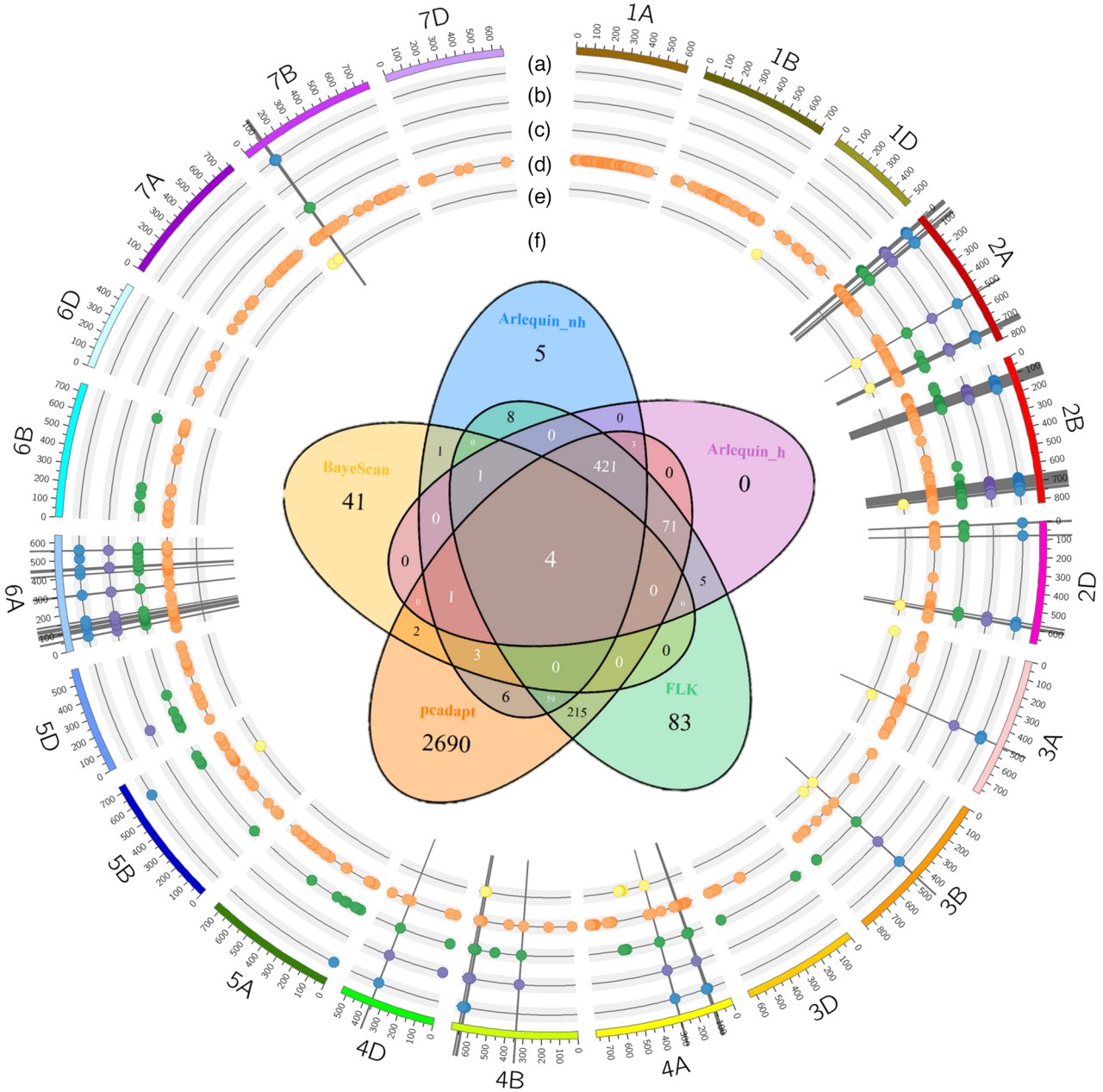

Figure 5. Chromosomal position of single nucleotide polymorphism (SNP) markers detected as loci under selection using (a) a non-hierarchical finite island model (Arlequin), (b) a hierarchical island model (Arlequin), (c) a Bayesian model (BayeScan), (d) an extended Lewontin and Krakauer (FLK) test, or (e) the Pcadapt approach. Gray bars indicate 563 SNP markers identified as outlier loci by at least three approaches. These loci were defined as putative candidates associated with genomic signatures of positive selection between the group of modern cultivars (MC) and the group of genebank accessions (GA). (f) Venn diagram of SNP markers detected as outlier loci using a non-hierarchical finite island model (Arlequin_nh), a hierarchical island model (Arlequin_h), a non-hierarchical Bayesian model (BayeScan), the FLK test (FLK), the Pcadapt approach (pcadapt), or a Bayesian model (BayeScan).

genes located within genomic regions under selection compared to the annotated genomic background.

For genomic regions associated with private alleles in the MC group, four TO terms, 21 PO terms, and 634 GO terms were significantly enriched (false discovery rate [FDR] < 0.01) (Table S7). The enriched TO terms were associated with stem elongation, plant growth, and plant development (Table S7). In total, five genes were associated with these TO terms; these genes were located within Region_C_chr2D-1 on chromosome 2D and were tandem array copies. The 21 identified PO terms were associated with the plant anatomical entity or the plant structure development stage, respectively, and are assumed to be involved in the anatomy or development of roots, ears,

flowers, or the vascular system (Table S7). Of the significantly enriched GO terms, 72% were in the biological process category, and were mainly involved in metabolic processes, hormonal regulation, or abiotic and biotic stress responses (Figure 6b).

For genomic regions under selection, no TO terms, 11 PO terms, and 90 GO terms were significantly enriched (FDR < 0.01) (Table S7). Of the significantly enriched GO terms, 54% were in the biological process category (Figure 6a) and were involved in metabolic processes or abiotic and biotic stress responses. The significantly enriched PO terms were all associated with the plant anatomical entity and are assumed to take part in leaf, root, or ovule anatomy (Table S7). In detail, enriched PO terms were associated with the anatomy of the trichome/trichoblast (PO:0000262, PO:0000282), tracheary elements (PO:0000290), ground meristem (PO:0025594), ground tissue (PO:0025059), cortex (PO:0005708), root epidermal cells (PO:0025164), stipule (PO:0020041), endothelium (PO:0020024), the integument (PO:0020021), and the inner integument (PO:0020022). The genes annotated with one or more of these PO terms were located within 25 genomic regions under selection on chromosomes 2A, 2B, 2D, 3A, 3B, 4B, 6A, and 7B. Ten of these genomic regions were in parts of the genome frequently carrying introgressions of known origin (Table S5).

Leaf- and root-associated traits have to be seen in the context of water and nutrient uptake and transport, tolerance/resistance to stress, and photosynthesis. It is conceivable that breeding activities have actively affected root and leaf anatomy to improve plant performance. For example, the density and length of trichomes can positively affect the biotic and abiotic stress tolerance of wheat (Gupt et al., 2021; Pshenichnikova et al., 2019; Saska et al., 2021) and genotypic differences in these traits exist among di-, tetra-, and hexaploid wheats (Pshenichnikova et al., 2017). In the 1980s, Richards and Passioura (1989, 1981) reported on breeding activities aimed to introduce the characteristic of narrow xylem vessels from landraces into Australian elite breeding materials to improve water-use efficiency under drought conditions. However, it is also conceivable that breeding activities can passively or accidently affect plant anatomy as a side effect of improvements to other important traits. (Lecain et al., 1989). Hoogendoorn et al. (1990) and Miralles et al. (1998) demonstrated that the introduction of semi-dwarfing genes led to changes in leaf anatomy (reduced leaf and cell size and increased total mesophyll surface area/leaf area). Furthermore, it is likely that introgressions have caused unintended changes in plant anatomy. For example, genes associated with leaf pubescence in *T. timopheevii* were transferred together with the actual target genes to bread wheat, causing changes in trichome density and length in two bread wheat populations (Budashkina, 1988; Simonov et al., 2021).

On the basis of the ontology enrichment analyses and the location of known genes, QTL regions, and introgressions, it can be suspected that genomic regions under selection are probably associated with plant anatomy, yield-related traits, and tolerance to abiotic and biotic stresses. This is consistent with the current assumption that breeding activities in the last century have introduced and combined novel and favorable alleles and genes in today's elite breeding cultivars, resulting in improved grain yield, bread making quality, and abiotic and biotic stress tolerance (Lopes et al., 2015; Mondal et al., 2016; Voss-Fels et al., 2019).

To cope with future challenges, it is essential to introduce novel genetic diversity into the elite breeding pools (Johansson et al., 2020). It is predicted that PGRs of wheat carry several useful alleles associated with adaptive traits that are missing from today's elite breeding pools (Lopes et al., 2015; Mascher et al., 2019).

Based on PIC values, 562 and 4721 private alleles were identified in the MC group and the GA group, respectively. In total, 44 of the identified private alleles associated with the MC group also described genomic regions harboring loci under selection. The SNP markers associated with private alleles in the GA group were equally distributed across the chromosomes, whereas those associated with private alleles in the MC group were mainly located on chromosomes 2A, 2B, and 2D (Figure 7). In total, 89 and 11% of private alleles were associated with GBS markers or iSelect markers, respectively (Table 3). Interestingly, 12 and 0.4% of the private alleles within the MC and GA groups, respectively, were associated with iSelect markers (Table 3). It can be assumed that these iSelect markers were actively used for marker-assisted selection in elite breeding pools, because they are generally based on known expressed sequence tags or genes. Furthermore, these findings also indicate that GBS markers are better suited to identify private alleles within the MC and GA groups compared to iSelect markers. One explanation may be the increased risk of ascertainment bias of the array-based iSelect markers (Albrechtsen et al., 2010; Chu et al., 2020), but not exclusively. Since, it was demonstrated that the ascertainment bias is negligible for the genotype panel under investigation.

The number of private alleles associated with GA groups ranged between 111 and 4639, with the highest numbers in STRUCTURE groups 2 (4473) and 7 (3530) and the admixed group (4639) (Figure S10). Furthermore, it was determined if some private alleles were specific to the STRUCTURE groups that mainly included GA. Most private alleles were found in two to seven STRUCTURE groups and were not specific to a single group (Figure S10). However, three private alleles located on chromosomes 3A and 5A were exclusively associated with STRUCTURE group 7 and indicated polymorphisms in

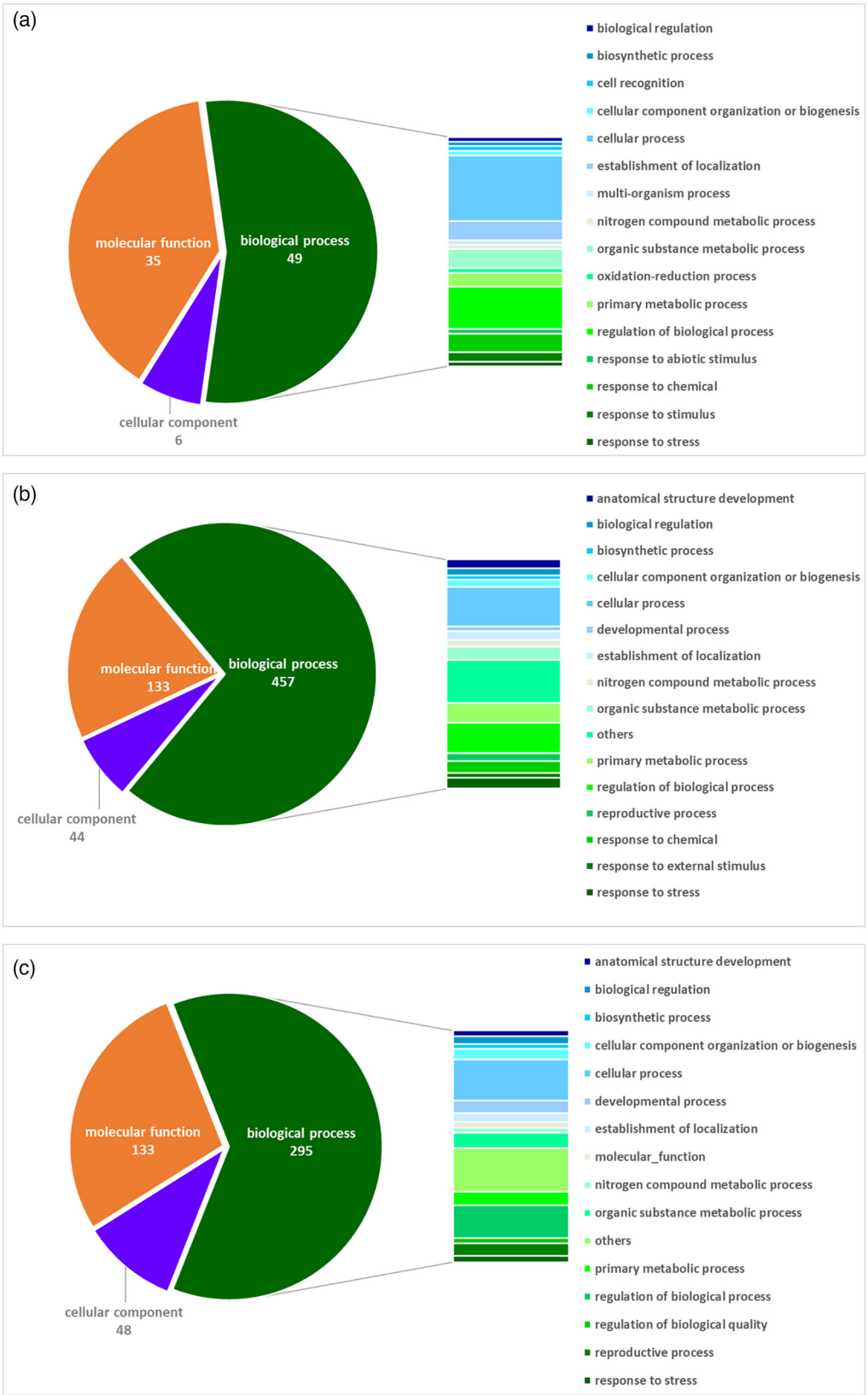

Figure 6. Gene Ontology (GO) enrichment analysis of genes located in (a) genomic regions under selection, (b) genomic regions associated with private alleles in the group of modern cultivars (MC), or (c) genomic regions associated with private alleles in the group of genebank accessions (GA).

the genes *TraesCS3B02G357500* (*TaEREF4*) and *TraesCS5A02G473800* (*TaQ*), respectively (IWGSC, 2018). Both genes encode transcription factors, i.e., the AP2/ERF domain-containing protein and Floral homeotic protein, respectively (IWGSC, 2018; Liu et al., 2020; Zhao et al., 2019). Recently, it has been shown that allelic diversity in the *TaEREF4* gene of Nepalese and Tibetan wheat genotypes is associated with cold acclimation and adaption to harsh environments at high altitudes (Guo et al., 2020). The *TaQ* gene also affects rachis brittleness of wheat (Liu et al., 2020).

According to the positions of the SNP markers on the reference genome (IWGSC, 2018), 36% of private alleles in the MC group and 37% of those in the GA group were located in genomic regions associated with an HC gene (Figure S9). In total, the private alleles associated with the MC group or the GA group were assigned to 63 and 827 genomic regions, respectively (Table S8).

Comparisons of genomic regions associated with private alleles in the GA group with known functional wheat genes revealed that 33 of these regions were associated with one or more known functional wheat genes. Most of these genes were related to grain quality or resistance to biotic stress (Table S8).

For instance, the genomic regions associated with private alleles in the GA group contained genes encoding polyphenol oxidase 1 (*Ppo-A1*, Region_chr2A-62), sucrose synthase II (*Sus-2B*, Region_chr2B-23), waxy proteins (*Wx-A1*, Region_chr7A-5; *Wx-B1*, Region_chr4A-91), genes related to seed dormancy and pre-harvest sprouting resistance (*Sdr-B1*, Region_chr2A-19; *Sdr-B1*, Region_chr2B-27; *Phs1/MFT*, Region_chr3A-1; *Vp-1A*, Region_chr3A-31; *VP-1B*, Region_chr3B-28; *R-A1*, Region_chr3A-38; *R-B1*, Region_chr3B-34), and a gene encoding starch-branching enzyme II (*Sbella*, Region_chr2A-44) (Table S8). Interestingly, some of the detected private alleles indicated polymorphisms within the intron of *Phs1/MFT* and the 3′ untranslated region of *Sbella*.

The *Phs1/MFT* gene locus on chromosome 3A is associated with seed dormancy and pre-harvest sprouting resistance (Liu et al., 2013; Nakamura et al., 2011). There are several polymorphisms and insertions within the promoter and coding regions of the *Phs1/MFT* gene locus (Jiang et al., 2018; Lei et al., 2013; Liu et al., 2013; Liu et al., 2015; Nakamura et al., 2011). For instance, two mutations within the *Phs1/MFT* gene, which result in a mis-splicing site or a premature stop codon, result in decreased seed dormancy and pre-harvest sprouting resistance (Liu et al., 2013; Liu et al., 2015). These two mutations are rare among the wild progenitors of wheat and landraces from the Fertile Crescent, but common in modern cultivars (Liu et al., 2015). In this study, the mutation causing the mis-splicing site was only present in traditional cultivars and landraces mainly originating from South-Eastern Europe and the former USSR

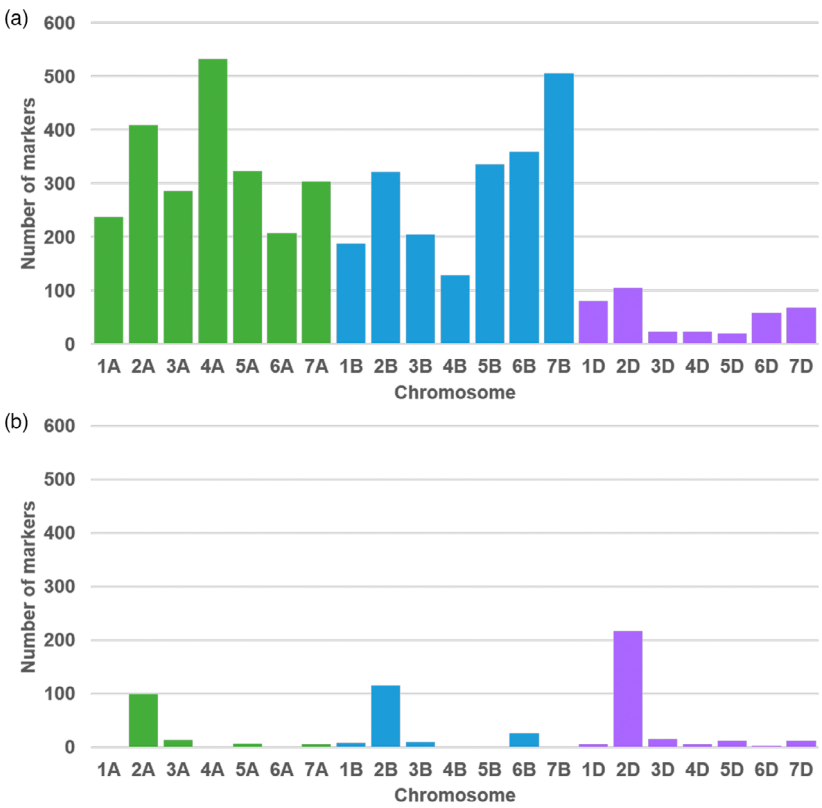

Figure 7. Distribution of markers associated with private alleles (a) in the group of genebank accessions (GA) and (b) the group of modern cultivars (MC).

Table 3 Number of private alleles identified in the group of modern cultivars (MC) and the group of genebank accessions (GA)

	MC	GA
GBS + iSelect	–	7
iSelect	2	550
GBS	560	4164
total	562	4721

(STRUCTURE groups 2 and 3), suggesting that the unfavorable allele leading to mis-splicing is rare in, or absent from, German wheat breeding pools. Wheat breeders may have successfully selected against this unfavorable allele. This is an example of how an unfavorable allele can be actively or passively reduced or eliminated from the MC group.

The *Sbella* gene on chromosome 2A encodes a starch-branching enzyme involved in starch biosynthesis (Tetlow & Emes, 2014). Starch with higher proportions of amylose (i.e., resistant starch) is beneficial for human health (Schönhofen et al., 2016). However, the amylose and resistant starch contents are low in modern wheat cultivars (Li et al., 2020; Schönhofen et al., 2016). Transgenic and non-transgenic approaches clearly demonstrated that an increase in high-amylose starch which is rich in resistant starch can be achieved through simultaneous down-regulation of *Sbella* and/or *Sbellb* isoforms located on group 2 chromosomes (Li et al., 2020; Regina et al., 2006; Regina et al., 2015; Schönhofen et al., 2016; Slade et al., 2012). Wheat germplasm which carries several combinations of mutations in *Sbella* and/or *Sbellb* isoforms is publicly available, but a cultivar carrying mutations in both isoforms on all three homologs has not yet been reported (Li et al., 2020; Schönhofen et al., 2016; Slade et al., 2012). In addition, natural variation at the *Sbella* gene locus has been detected in Russian landraces and cultivars (Konovalov et al., 2012). Both of the polymorphisms within the *Sbella* locus reported here were detected in all Asian landraces, some South-Eastern European landraces, and some traditional cultivars mainly originated from Asia and South-Eastern Europe (STRUCTURE groups 1, 2, and 7). These findings suggest that both private alleles originated from Asian or South-Eastern European landraces. Furthermore, it is likely that both alleles were transferred from Asian landraces into the Italian and South-Eastern European breeding pools during early breeding activities.

Region_G_chr6B-57, associated with private alleles on chromosome 6B, is linked with the *Cre8* gene locus, which confers resistance to cereal cyst nematodes (Williams et al., 2003). The *Cre8* resistance allele was first detected in the Australian cv. Festiguay (Paull et al., 1998) and is now widely distributed in Australian breeding materials (Joukhadar et al., 2019). However, only limited information is available about the presence of the *Cre8* resistance allele in European and Asian wheat germplasm (Imren et al., 2021; Karelov et al., 2019). The private alleles on chromosome 6B detected here might be associated with polymorphisms in *Cre8*, but further research is necessary to explore this genome region in detail. These private alleles were mainly found in landraces and traditional cultivars from South-Eastern Europe and Asia (STRUCTURE group 1).

Finally, Region_G_chr3B-1, indicative of a private allele on chromosome 3B, is linked with the *Yr57* gene locus, which confers stripe rust resistance in wheat (Randhawa et al., 2015). The *Yr57* resistance allele was first reported in a Pakistani wheat landrace from the Watkins Collection, and was recently transferred into Australian breeding material (Randhawa et al., 2014; Randhawa et al., 2015; Randhawa et al., 2019). Initial studies pointed out that the *Yr57* resistance gene allele appears to be absent from wheat cultivars from Australia, India, and Northern Europe (Randhawa et al., 2015; Randhawa et al., 2016). If the detected private allele is associated with a polymorphism in *Yr57*, these results could be helpful to identify new sources of *Yr57* resistance alleles. In this context, the private allele is connected with landraces from Greece, Iran, and Nepal, as well as with traditional cultivars from the USA, former USSR, and Southern Europe.

The three genomic regions that include private alleles in the GA group mentioned above are possibly associated with novel allelic diversity that may useful for breeding in the future.

Furthermore, results of the PO and GO term enrichment analyses indicated a significant enrichment (FDR < 0.01) of 76 PO terms and 435 GO terms in genomic regions associated with private alleles in the group of GA. The PO terms were associated with plant anatomical entity or plant structure development stage, and were assumed to be involved in diverse processes (Table S7). In total, 68% of the GO terms were in the biological process category (Figure 6c), which includes many processes. In accordance with findings of the PO and GO enrichment analyses, some known functional genes and QTL regions (LOD ≥ 5; Blake et al., 2019) associated with plant morphology, plant physiology, grain yield, yield components, grain quality, lodging, and stress tolerance were located within genomic regions associated with private alleles in the group of GA (Table S6).

It was pointed out that allelic diversity of many genes related to various traits was reduced in the MC group compared to the GA group. The reduction of allelic diversity in MC is probably related to the depletion of unfavorable alleles associated with traits of interest and positively highlights the breeding success of the last decades, on the one hand. On the other hand, it might be related to the depletion of alleles that could be useful for further breeding success, but were depleted due to linkage drag or by chance.

CONCLUSIONS

For progress in breeding, it is essential to understand more precisely the targets of selection during wheat breeding

activities in the last century. It is also vital to evaluate whether the private alleles associated with the GA group are linked with traits of future interest. In this regard, phenotyping this genotype collection for agronomically important traits and conducting genome-wide association studies may help to identify QTL regions associated with traits of current or future interest.

A detailed understanding of genetic diversity and population structure, combined with knowledge of past and current selection targets and novel allelic variation for genes of interest in PGRs, will facilitate the introduction of useful alleles into elite breeding pools, for example, through targeted crosses or genome editing. This will promote improved performance of elite breeding material under changing environmental conditions.

EXPERIMENTAL PROCEDURES

Plant materials

A collection of 290 hexaploid winter wheat (*T. aestivum* L.) genotypes was investigated to determine its population structure, genetic diversity, private alleles, and genomic regions under selection. These genotypes represent a diverse collection consisting of 81 MC of winter wheat released in Western Europe (mainly Germany) during the last two decades and a global collection of 209 GA. The GA were classified as traditional landraces (45), a breeder's line (1), traditional or advanced cultivars (146), and accessions with unknown biological status (17) on the basis of the passport data (Table S1) available in the Genebank Information System (GBIS, German Federal ex situ Genebank in Gatersleben; Oppermann et al., 2015). Seeds of the MC were kindly provided by German breeding companies. Seeds of GA were supplied by the German Federal ex situ Genebank at the Leibniz Institute of Plant Genetics and Crop Plant research (IPK) in Gatersleben under Standard Material Transfer Agreements.

For selecting the most diverse GA for flowering time, plant height, and thousand-grain weight representing the whole diversity stored in the German Federal ex situ Genebank, long-term phenotypic data recorded during seed multiplication of the GA were used to determine normalized rank products (Keilwagen et al., 2014). In total, 172 GA were selected corresponding to eight contrasting groups representing the most extreme combinations of these three traits. Additionally, 37 GA were chosen on the basis of one extreme trait characteristic (flowering time, plant height, or thousand-grain weight) (Supplementary Table S1: rank products).

GA used in this study can represent only one genotype (e.g., traditional or advanced cultivars) or can consist of mixtures of different genotypes (e.g., landraces). To reduce heterogeneity and heterozygosity and to ensure genetic purity of GA and MC, the whole panel was grown under single-seed-descent (SSD) selection in controlled greenhouse conditions. Consequently, seeds resulting from SSD selection represent only one genotype. SSD seeds were propagated and genomic DNA was extracted from young leaf tissue according to the protocol of Milner et al. (2018).

Genotyping

Genotyping of the 290 winter wheat genotypes was conducted by GBS analysis. The extracted DNA was digested with restriction enzymes (*Pst*I and *Msp*I; New England Biolabs, https://international.neb.com/), and GBS library construction and sequencing were implemented according to the protocol of Wendler et al. (2014). The GBS raw reads (available at European Nucleotide Archive: https://www.ebi.ac.uk/ena/browser/view/PRJEB30008) were pre-processed and mapped against the reference genome of Chinese spring wheat (IWGSC, 2018) as described in Keilwagen et al. (2019). SNP calling (SAMtools package including samtools and bcftools; Li, 2011; non-default parameters: --output-tags DP,DPR) was implemented. Only raw polymorphisms with a minimum quality of 40 were retained. Genotype calls with less than two or four reads covering homozygous or heterozygous position per sample were treated as missing. Finally, GBS SNP data were filtered and monomorphic or multiallelic markers, indels, and markers with at least 30% missing values were excluded.

Additionally, at the SGS Institute Fresenius – Trait Genetics Section Gatersleben (Germany), all genotypes were genotyped using the 15K+5K iSelect chip (Illumina Inc., www.illumina.com) (including all markers from the 15K array [Soleimani et al., 2020] and additionally 5000 markers from the 35K array [Allen et al., 2017]) and filtered to remove monomorphic markers, those with missing values (<30%), and duplicated markers, corresponding to the post-processing of SNPs obtained by GBS analysis.

For imputation, the GBS and iSelect marker sets were combined and analyzed together. Duplicates were identified and only one entry was included in the combined marker set. Missing values in the combined marker set were imputed using the software package *Beagle* v.4.1 (Browning & Browning, 2007). The imputed marker set was filtered (heterozygosity $\leq$ 12.5% and MAF $\geq$ 5%), resulting in the final marker dataset.

Determination of population structure

The software tool *PLINK* v.1.07 (LD prune: window size 50, step size 5, r^2 threshold 0.2; Purcell et al., 2007) was used to select a subset of LD-pruned markers, which included only markers in linkage equilibrium. The pruned marker set was used to determine the population structure.

The population structure was determined by PCoA implemented in *DARwin* v.5.0 (Perrier & Jacquemoud-Collet, 2006), Bayesian clustering analysis implemented in STRUCTURE v.2.3.4 (Pritchard et al., 2000), and DAPC using *adegenet* in the R package (Jombart & Ahmed, 2011; Jombart, 2008; R Core Team, 2019) (Text S3).

Genetic diversity, private alleles, and genomic regions under selection

The combined dataset was used to determine genetic diversity and to identify private alleles as well as loci/genomic regions under selection. For this purpose, Arlequin v.3.5.2.2 (Excoffier & Lischer, 2010) was used to: (i) calculate the average number of alleles and the H_e value per marker locus; (ii) estimate the number of polymorphic loci, the H_e value, and the F_{st} value as a measurement of the population differentiation for each subpopulation; and (iii) perform AMOVA to assess the molecular variance among subpopulations. AMOVA was performed between the GA and MC groups as well as between and among the STRUCTURE subpopulations.

The PIC values (Hildebrand et al., 1992) were separately calculated for each marker for the MC group and the GA group to identify the private alleles associated with each of these groups. Loci under selection representing genomic signatures of selection between the GA group and the MC group were identified by population differentiation-based statistics and a principal component analysis (PCA)-based statistic. For this purpose, F_{st}-outlier detection tests based on the FDIST algorithm (Beaumont & Nichols, 1996) were conducted using the software packages Arlequin v.3.5.2.2 (Excoffier & Lischer, 2010) and BayeScan v.2.1 (Foll & Gaggiotti, 2008). In

total, three different approaches were implemented to identify loci under selection, i.e., the non-hierarchical finite island model (Arlequin), the hierarchical island model (Arlequin settings: 20 000 simulations, 100 demes, 10 groups), and the Bayesian model (BayeScan settings: burn-ins 50 000, iterations 100 000, pilot runs 20). Furthermore, a fourth population differentiation-based statistic, i.e., the extended Lewontin and Krakauer (FLK) test (Bonhomme et al., 2010) implemented in hapFLK v1.4 (Fariello et al., 2013), and a PCA-based approach implemented in the R package pcadapt (R Core Team, 2019; Luu et al., 2017) were used to identify loci under selection (Text S4). By default, FDR adjustment for multiple testing (Benjamini & Hochberg, 1995) is implemented in BayeScan (Foll, 2012; Foll & Gaggiotti, 2008). The F_{st} *P*-values, FLK *P*-values, and pcadapt *P*-values were adjusted for multiple testing (FDR, Benjamini & Hochberg, 1995) using the function *p.adjust* (setting: method = 'BH') in the R package (R Core Team, 2019). All markers with FDR-adjusted *P*-values of <0.1 were considered to be under selection.

Markers that were identified as loci under selection by at least three approaches were defined as putative candidates associated with genomic signatures of selection. These markers and those indicating private alleles in the MC group or the GA group were assigned to genomic regions based on the chromosomal position of markers and LD decay. The LD was estimated as described by Lehnert et al. (2018) using the R-based software packages genetics and LDheatmap (R Core Team, 2019; Shin et al., 2006; Warnes et al., 2013). The LD decay was determined as the intersection point of the smooth locally weighted polynomial regression curve with the critical r^2 value ($r^2 = 0.2$), which was 2 Mbp. Therefore, markers within 2 Mbp of each other were assigned to the same genomic region. The genomic regions plus the 1-Mbp flanking regions on each side were defined as putative genomic regions under selection or putative genomic regions associated with private alleles. These genomic regions were matched against known functional wheat genes, publicly available QTLs (LOD $\geq$ 5) listed at the Grain Genes Genome Browser database (Blake et al., 2019), and known large chromosomal modifications identified by coverage analysis (Keilwagen et al., 2019; Keilwagen et al., 2022) to provide clues about the associated traits.

In previous studies, genomic regions under selection were identified in diverse wheat collections using different statistical approaches (Afzal et al., 2019; Cavanagh et al., 2013; Cheng et al., 2019; Fradgley et al., 2019; Gao et al., 2017; Gaire et al., 2020; Jordan et al., 2015; Liu et al., 2019; N'Diaye et al., 2018; Pont et al., 2019). Generally, it is difficult to compare results among studies that have used different marker systems, genotype collections, and maps (genetic versus physical). Therefore, to compare our results with those from other studies, we attempted to remap all markers associated with genomic regions under selection identified in the literature (unless it was already done) to the reference genome of Chinese spring wheat (IWGSC, 2018). However, not all previously published markers associated with genomic regions under selection could be anchored on the reference genome (IWGSC, 2018), either because the markers could not be uniquely mapped or because no flanking marker sequences were available. Therefore, only markers that could be uniquely mapped to the reference genome (IWGSC, 2018) were used for comparison.

Ontology enrichment analyses

In order to study the functional composition of the identified genomic regions, we intersected their genomic coordinates with the coordinates of the HC partition in the wheat IWGSC V1.1 genome annotation (IWGSC, 2018). The resulting protein-coding HC gene loci were then mapped to wheat GO, TO, and PO annotations (release v1.0; https://github.com/PGSB-HMGU/ontology_annotations). Enrichment of specific ontology terms among the given gene sets was tested using the 'Parent–Child-Union' algorithm implemented in the Ontologizer software (Grossmann et al., 2007) using all annotated wheat genes as reference and applying multiple testing correction of *P*-values using the Benjamini–Hochberg method.

ACCESSION NUMBERS

Accession numbers are available in Table S1.

AUTHOR CONTRIBUTIONS

JK and BK designed the work. JK and BK selected the collection of genebank accessions. BK conducted the single seed propagation of genebank material. AH and NS performed GBS analysis. SB, JK, and TB performed bioinformatics analyses. HL performed population genetics analyses. DL performed ontology enrichment analyses. All authors discussed the results. HL wrote the manuscript. All authors read, commented on, and approved the final manuscript.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

ACKNOWLEDGMENTS

We thank the following for providing seeds, DNA, passport data, and fruitful collaborations: the Federal ex situ Genebank of Germany, Gatersleben (IPK, Germany), Erhard Ebmeyer (KWS Lochow GmbH), Johannes Schacht and Anne Starke (Limagrain GmbH), Ludwig Ramgraber and Jens Weyen (Saatzucht Josef Breun GmbH and Co. KG), Carsten Reinbrecht and Stefan Streng (Saatzucht Streng-Engelen GmbH and Co. KG), Mary Ziems (IPK), and Tanja Gerjets (German Federation for Plant innovation [GFPi], proWeizen). We also thank Kerstin Wolf, Ute Krajewski, Peter Schreiber, and Susanne König (all at IPK) for technical assistance. Research to collect GBS and iSelect data for the winter wheat collection and HL were supported by the German Federal Ministry of Food and Agriculture within the GenDiv project (grant no. 2814603813). This work was supported by the (i) European Community's Seventh Framework Programme (FP7, 2014–2019) under grant agreement no. FP7-613556, Whealbi project (http://www.whealbi.eu/project/), (ii) the Leibniz Institute of Plant Genetics and Crop Research (IPK, Gatersleben, Germany), and (iii) the initiative 'Adapting Agriculture to Climate Change: Collecting, Protecting and Preparing Crop Wild Relatives' supported by the Government of Norway and managed by the Global Crop Diversity Trust (https://www.cwrdiversity.org/project/pre-breeding/). Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

All relevant data can be found within the published article and its supplementary material.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Distribution of (a) minor allele frequencies (MAFs), (b) polymorphism information content (PIC) values, (c) observed heterozygosity (H_o) values, and (d) expected heterozygosity (H_e) values for the combined marker set (49 181 markers).

Figure S2. Distribution of (a) minor allele frequencies (MAFs), (b) polymorphism information criterion (PIC) values, and (c) expected heterozygosity (H_e) values for 11 267 iSelect markers (left), 38 041 genotype-by-sequencing (GBS) markers (middle), the combined marker set (49 181 markers), and the reduced marker set (6116 markers) (right).

Figure S3. Relationship between physical chromosome size and (a) number of markers per chromosome, (b) average marker interval per chromosome, and (c) marker coverage per chromosome for the combined marker dataset (49 181 SNP markers).

Figure S4. Relationship between physical chromosome size and (a) number of markers per chromosome and (b) average marker interval per chromosome for 11 267 iSelect markers (left), 38 041 genotype-by-sequencing (GBS) markers (middle), and the reduced marker set (6116 single nucleotide polymorphism markers) (right).

Figure S5. Number of single nucleotide polymorphism (SNP) markers associated with a high-confidence (HC) gene or genomic region (non-gene) not associated with an HC gene. Results are shown for (a) the whole marker set (49 181 markers), (b) markers identified by the genotyping-by-sequencing (GBS) approach (38041), and (c) markers identified by 15K+5K iSelect chip (11267). Small pie charts highlight the number of markers located within a coding or non-coding region of an HC gene.

Figure S6. ΔK plot for the reduced marker dataset (6116 markers).

Figure S7. Bayesian information criterion (BIC) graph showing results of the K-means clustering algorithm implemented in the find.clust function in R. An optimal number of K = 7 subpopulations is indicated based on the lowest BIC value.

Figure S8. Population structure analysis of a panel of 290 winter wheat genotypes based on the reduced set of 6116 single nucleotide polymorphism (SNP) markers. Results of discriminant analysis of principal components (DAPC) are shown as scatterplot. Axes represent the first and second principal components of the DAPC. DAPC groups (1 to 7) are highlighted in color. All modern cultivars grouped together in DAPC group 5.

Figure S9. Number of single nucleotide polymorphism (SNP) markers associated with a high-confidence (HC) gene (gene) or a genome region (non-gene) not associated with an HC gene. Results are shown for (a) SNP markers associated with private alleles in group of modern cultivars (MC; 562 markers), (b) SNP markers associated with private alleles in a group of genebank accessions (GA; 4721 markers), and (c) SNP markers associated with genomic signatures of positive selection (563 markers). Small pie charts show the number of markers located within a coding or non-coding region of an HC gene.

Figure S10. Distribution of private alleles per STRUCTURE group. (a) Number of private alleles associated with a group of genebank accessions per STRUCTURE group. (b) Number of private alleles associated with a group of genebank accessions specific (1) or non-specific (2 to 8) to a STRUCTURE group.

Table S1. Plant materials. List of 290 winter wheat genotypes used in this study and additional information.

Table S2a. Number of markers and average marker interval per chromosome for the combined marker set (49 181 markers), GBS markers, and iSelect markers.

Table S2b. Number of markers and average marker interval per chromosome for the reduced marker set (6116 markers), GBS markers, and iSelect markers.

Table S3a. Average number of alleles and expected heterozygosity (H_e) for a group of modern cultivars (MC), a group of genebank accessions (GA), and a randomly selected group of genebank accessions (GA*) based on 49 181 SNP markers.

Table S3b. Average number of alleles and expected heterozygosity (H_e) shown for all genotypes included in STRUCTURE group I and STRUCTURE group II based on 49 181 SNP markers.

Table S4. Average number of alleles, expected heterozygosity (H_e), and number of polymorphic loci for 14 randomly selected genotypes from STRUCTURE groups 1 to 7 and an admixed group based on 49 181 SNP markers.

Table S5. Genomic regions associated with putative candidate loci under positive selection and functional wheat genes located within these genome regions.

Table S6. Publicly available QTLs (LOD $\geq$ 5) from the Grain Genes Genome Browser database (Blake et al., 2019) located within genomic regions under selection (S, blue) or regions associated with private alleles in a group of modern cultivars (C, purple) or a group of genebank accessions (G, green).

Table S7. Significantly enriched Trait Ontology (TO) and Plant Ontology (PO) terms associated with genomic regions under selection (S) or private alleles in a group of modern cultivars (MC) or a group of genebank accessions (GA).

Table S8. Genomic regions associated with private alleles in group of modern cultivars (MC) or group of genebank accessions (GA) and functional wheat genes located within these genomic regions.

Supplementary Text S1. Population structure and genetic diversity analysis of a panel of 290 winter wheat genotypes based on iSelect or GBS markers, respectively.

Supplementary Text S2. Detailed information about the 6116 markers selected by LD pruning.

Supplementary Text S3. Detailed information about population structure analyses.

Supplementary Text S4. Detailed information about the identification of loci representing genomic regions under selection using population differentiation-based statistics and a principal component analysis-based statistic.

REFERENCES

Afzal, F., Li, H., Gul, A., Subhani, A., Ali, A., Mujeeb-Kazi, A. *et al.* (2019) Genome-wide analyses reveal footprints of divergent selection and drought adaptive traits in synthetic-derived wheats. *G3: genes, Genomes, Genetics*, **9**, 1957–1973.

Agenbag, G., Pretorius, Z., Boyd, L., Bender, C. & Prins, R. (2012) Identification of adult plant resistance to stripe rust in the wheat cultivar Cappelle-Desprez. *Theoretical and Applied Genetics*, **125**, 109–120.

Albrechtsen, A., Nielsen, F.C. & Nielsen, R. (2010) Ascertainment biases in SNP chips affect measures of population divergence. *Molecular Biology and Evolution*, **27**, 2534–2547.

Alipour, H., Bihamta, M.R., Mohammadi, V., Peyghambari, S.A., BAI, G. & Zhang, G. (2017) Genotyping-by-sequencing (GBS) revealed molecular genetic diversity of Iranian Wheat landraces and cultivars. *Frontiers in Plant Science*, **8**, 1293.

Allen, A. M., Winfield, M. O., Burridge, A. J., Downie, R. C., Benbow, H. R., Barker, G. L. *et al.* 2017. Characterization of a Wheat Breeders' Array suitable for high-throughput SNP genotyping of global accessions of hexaploid bread wheat (Triticum aestivum). **15**, 390–401.

Arenas-Corraliza, M.G., Rolo, V., López-Díaz, M.L. & Moreno, G. (2019) Wheat and barley can increase grain yield in shade through acclimation of physiological and morphological traits in Mediterranean conditions. *Scientific Reports*, **9**, 9547.

Balfourier, F., Bouchet, S., Robert, S., de Oliveira, R., Rimbert, H., Kitt, J. *et al.* (2019) Worldwide phylogeography and history of wheat genetic diversity. *Science Advances*, **5**, eaav0536.

Beaumont, M.A. & Nichols, R. (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B: Biological Sciences*, **263**, 1619–1626.

Belderok, B. (2000) Developments in bread-making processes. *Plant Foods for Human Nutrition*, **55**, 1–14.

Belderok, B., Mesdag, J. & Donner, D.A. (2000) *Bread-making quality of wheat; a century of breeding in Europe*. Dordrecht: Kluwer Academic Publishers.

Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 289–300.

Blake, V.C., Woodhouse, M.R., Lazo, G.R., Odell, S.G., Wight, C.P., Tinker, N.A. *et al.* (2019) GrainGenes: centralized small grain resources and digital platform for geneticists and breeders. *Database: The Journal of Biological Databases and Curation*, **2019**, 1–7.

Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J., Blott, S. *et al.* (2010) Detecting selection in population trees: the Lewontin and Krakauer test extended. *Genetics*, **186**, 241–262.

Bonnin, I., Bonneuil, C., Goffaux, R., Montalent, P. & Goldringer, I. (2014) Explaining the decrease in the genetic diversity of wheat in France over the 20th century. *Agriculture, Ecosystems & Environment*, **195**, 183–192.

Börner, A., Neumann, K., Kobiljski, B. & Lohwasser, U. (2010) The maintenance and exploitation of ex situ Genebank collections–association mapping for flowering time in Wheat. *Agriculturae Conspectus Scientificus (ACS)*, **75**, 141–144.

Borojevic, K. & Borojevic, K. (2005) The transfer and history of "reduced height genes" (Rht) in Wheat from Japan to Europe. *Journal of Heredity*, **96**, 455–459.

Browning, S.R. & Browning, B.L. (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-Genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics*, **81**, 1084–1097.

BSL 2015. Beschreibende Sortenliste Getreide, Mais, Öl- und Faserpflanzen, Leguminosen, Rüben, Zwischenfrüchte, Hannover, Bundessortenamt.

Budashkina, E. (1988) Cytogenetic study of introgressive disease-resistant common wheat lines. *Tag. Ber. Acad. Landwirtsch. Wiss. DDR*, **266**, 209–212.

Cavanagh, C.R., Chao, S., Wang, S., Huang, B.E., Stephen, S., Kiani, S. *et al.* (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proceedings of the National Academy of Sciences*, **110**, 8057–8062.

Cheng, H., Liu, J., Wen, J., Nie, X., Xu, L., Chen, N. *et al.* (2019) Frequent intra-and inter-species introgression shapes the landscape of genetic variation in bread wheat. *Genome Biology*, **20**, 1–16.

Chu, J., Zhao, Y., Beier, S., Schulthess, A.W., Stein, N., Philipp, N. *et al.* (2020) Suitability of single-nucleotide polymorphism arrays versus genotyping-by-sequencing for Genebank genomics in Wheat. *Frontiers in Plant Science*, **11**.

Chung, Y.S., Choi, S.C., Jun, T.-H. & Kim, C. (2017) Genotyping-by-sequencing: a promising tool for plant genetics research and breeding. *Horticulture, Environment, and Biotechnology*, **58**, 425–431.

Dempewolf, H., Baute, G., Anderson, J., Kilian, B., Smith, C. & Guarino, L. (2017) Past and future use of wild relatives in crop breeding. *Crop Science*, **57**, 1070–1082.

Divashuk, M.G., Bespalova, L.A., Vasilyev, A.V., Fesenko, I.A., Puzyrnaya, O.Y. & Karlov, G.I. (2013) Reduced height genes and their importance in winter wheat cultivars grown in southern Russia. *Euphytica*, **190**, 137–144.

Excoffier, L. & Lischer, H.E.L. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and windows. *Molecular Ecology Resources*, **10**, 564–567.

Faostat. 2019. *Faostat database collections* [Online]. Available: http://faostat.fao.org/ [Accessed Dezember 2021].

Fariello, M.I., Boitard, S., Naya, H., Sancristobal, M. & Servin, B. (2013) Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics*, **193**, 929–941.

Faris, J.D. (2014) Wheat domestication: key to agricultural revolutions past and future. In: Tuberosa, R., Graner, A. & E.F. (Eds.) *Genomics of plant genetic resources*. Dordrecht: Springer.

Foll, M. (2012) BayeScan v2. 1 user manual. *Ecology*, **20**, 1450–1462.

Foll, M. & Gaggiotti, O. (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–993.

Fradgley, N., Gardner, K.A., Cockram, J., Elderfield, J., Hickey, J.M., Howell, P. *et al.* (2019) A large-scale pedigree resource of wheat reveals evidence for adaptation and selection by breeders. *PLoS Biology*, **17**, e3000071.

Fu, Y.-B. (2015) Understanding crop genetic diversity under modern plant breeding. *Theoretical and Applied Genetics*, **128**, 2131–2142.

Gaire, R., OHM, H., Brown-Guedira, G. & Mohammadi, M. (2020) Identification of regions under selection and loci controlling agronomic traits in a soft red winter wheat population. *The Plant Genome*, **13**, e20031.

Gao, L., Koo, D.-H., Juliana, P., Rife, T., Singh, D., Lemes DA Silva, C. *et al.* (2021) The Aegilops ventricosa 2NvS segment in bread wheat: cytology, genomics and breeding. *Theoretical and Applied Genetics*, **134**, 529–542.

GAO, L., Zhao, G., Huang, D. & JIA, J. (2017) Candidate loci involved in domestication and improvement detected by a published 90K wheat SNP array. *Scientific Reports*, **7**, 44530.

Grossmann, S., Bauer, S., Robinson, P.N. & Vingron, M. (2007) Improved detection of overrepresentation of gene-ontology annotations with parent–child analysis. *Bioinformatics*, **23**, 3024–3031.

Guo, W., Xin, M., Wang, Z., Yao, Y., Hu, Z., Song, W. *et al.* (2020) Origin and adaptation to high altitude of Tibetan semi-wild wheat. *Nature Communications*, **11**, 5085.

Gupt, S.K., Chand, R., Mishra, V.K., Ahirwar, R.N., Bhatta, M. & Joshi, A.K. (2021) Spot blotch disease of wheat as influenced by foliar trichome and stomata density. *Journal of Agriculture and Food Research*, **6**. 100227.

Hedden, P. (2003) The genes of the green revolution. *Trends in Genetics*, **19**, 5–9.

Hickey, L.T., Hafeez, A.N., Robinson, H., Jackson, S.A., Leal-Bertioli, S.C., Tester, M. *et al.* (2019) Breeding crops to feed 10 billion. *Nature Biotechnology*, **37**, 744–754.

Hildebrand, C.E., Torney, D.C. & Wagner, R.P. (1992) Informativeness of polymorphic DNA markers. *Los Alamos Science*, **20**, 100–102.

Hoogendoorn, J., Rickson, J.M. & Gale, M.D. (1990) Differences in leaf and stem anatomy related to plant height of tall and dwarf Wheat (Triticum aestivum L.). *Journal of Plant Physiology*, **136**, 72–77.

Imren, M., Ozer, G., Duman, N. & Dababat, A. (2021) Phenotypic and genotypic characterization of wheat and barley varieties for resistance to cereal cyst nematode (Heterodera latipons). *Genetic Resources and Crop Evolution*, **68**(5), 2131-2141.

IWGSC. (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, **361**, eaar7191.

Jia, M., Guan, J., Zhai, Z., Geng, S., Zhang, X., MAO, L. *et al.* (2018) Wheat functional genomics in the era of next generation sequencing: an update. *The Crop Journal*, **6**, 7–14.

Jiang, H., Zhao, L.-X., Chen, X.-J., Cao, J.-J., Wu, Z.-Y., Liu, K. *et al.* (2018) A novel 33-bp insertion in the promoter of TaMFT-3A is associated with pre-harvest sprouting resistance in common wheat. *Molecular Breeding*, **38**, 69.

Johansson, E., Henriksson, T., Prieto-Linde, M.L., Andersson, S., Ashraf, R. & Rahmatov, M. (2020) Diverse Wheat-alien Introgression lines as a basis for durable resistance and quality characteristics in bread Wheat. *Frontiers in Plant Science*, **11**.

Jombart, T. & Ahmed, I.J.B. (2011) Adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070–3071.

Jombart, T.J.B. (2008) Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Jordan, K.W., Wang, S., LUN, Y., Gardiner, L.-J., Maclachlan, R., Hucl, P. *et al.* (2015) A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. *Genome Biology*, **16**, 48.

Joukhadar, R., Daetwyler, H.D., Gendall, A.R. & Hayden, M.J. (2019) Artificial selection causes significant linkage disequilibrium among multiple unlinked genes in Australian wheat. *Evolutionary Applications*, **12**, 1610–1625.

Karelov, A.V., Pylypenko, L.A., Kozub, N.A., Sozinov, I.A. & Blume, Y.B. (2019) Genetic background of the resistance against parasitic nematodes in Wheat. *Cytology and Genetics*, **53**, 315–320.

Keilwagen, J., Kilian, B., Ozkan, H., Babben, S., Perovic, D., Mayer, K.F.X. *et al.* (2014) Separating the wheat from the chaff - a strategy to utilize plant genetic resources from ex situ genebanks. *Scientific Reports*, **4**, 1–6.

Keilwagen, J., Lehnert, H., Berner, T., Badaeva, E., Himmelbach, A., Börner, A. *et al.* (2022) Detecting major introgressions in wheat and their putative origins using coverage analysis. *Scientific Reports*, **12**, 1–11.

Keilwagen, J., Lehnert, H., Berner, T., Beier, S., Scholz, U., Himmelbach, A. *et al.* (2019) Detecting large chromosomal modifications using short read data from genotyping-by-sequencing. *Frontiers in Plant Science*, **10**, 1133.

Kilian, B., Dempewolf, H., Guarino, L., Werner, P., Coyne, C. & Warburton, M.L. (2021) Crop science special issue: adapting agriculture to climate change: a walk on the wild side. *Crop Science*, **61**, 32–36.

Kilian, B. & Graner, A. (2012) NGS technologies for analyzing germplasm diversity in genebanks. *Briefings in Functional Genomics*, **11**, 38–50.

Knüpffer, H., Filatenko, A., Hammer, K., Grau, M. & Börner, A. (2001) The wheat collection of the Genebank of IPK Gatersleben, Germany11, compilers. 2003. *Report of a Working Group on Wheat*, 40–54.

Konovalov, F., Shaturova, A., Mitrofanova, O. & Kudryavtsev, A. (2012) An approach to DNA polymorphism screening in Sbeiia homeologous genes of polyploid wheat (Triticum L.). *Euphytica*, **183**, 173–184.

Körmöczi, P., Tóth, B., Nagy-György, A., Kocsis, K., Óvári, J., Szabó, B. *et al.* (2019) SNP-based genetic diversity assessment among hungarian bread wheat (Triticum aestivum L.) genotypes. *Cereal Research Communications*, **48**(1), 1–7.

Lecain, D. R., Morgan, J. A. & Zerbi, G. 1989. Leaf anatomy and gas exchange in nearly isogenic Semidwarf and tall winter Wheat. *Crop Science*, **29**, 1246–1251.

Lehnert, H., Serfling, A., Friedt, W. & Ordon, F. (2018) Genome-wide association studies reveal genomic regions associated with the response of Wheat (Triticum aestivum L.) to mycorrhizae under drought stress conditions. *Frontiers in Plant Science*, **9**, 1728.

Lei, L., ZHU, X., Wang, S., Zhu, M., Carver, B.F. & Yan, L. (2013) TaMFT-A1 is associated with seed germination sensitive to temperature in winter wheat. *PLoS One*, **8**, e73330.

Levy, A. & Feldman, M. 2022. Evolution and origin of bread wheat. *The Plant Cell*, **34**, 2549–2567.

Li, H. (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, **27**, 2987–2993.

Li, J., Jiao, G., Sun, Y., Chen, J., Zhong, Y., Yan, L. *et al.* 2020. Modification of starch composition, structure and properties through editing of Tasbeiia in both winter and spring wheat varieties by Crispr/Cas9. *Plant Biotechnology Journal*, **19**(5), 937–951.

Liu, H., Wang, K., Tang, H., Gong, Q., Du, L., Pei, X. *et al.* (2020) Crispr/Cas9 editing of wheat TaQ genes alters spike morphogenesis and grain threshability. *Journal of Genetics and Genomics*, **47**, 563–575.

Liu, J., Rasheed, A., He, Z., Imtiaz, M., Arif, A., Mahmood, T. *et al.* (2019) Genome-wide variation patterns between landraces and cultivars uncover divergent selection during modern wheat breeding. *Theoretical and Applied Genetics*, **132**, 2509–2523.

Liu, S., Sehgal, S.K., Li, J., Lin, M., Trick, H.N., Yu, J. *et al.* (2013) Cloning and characterization of a critical regulator for Preharvest sprouting in Wheat. *Genetics*, **195**, 263–273.

Liu, S., Sehgal, S.K., Lin, M., Li, J., Trick, H.N., Gill, B.S. *et al.* (2015) Independent mis-splicing mutations in TaPHS1 causing loss of preharvest sprouting (PHS) resistance during wheat domestication. *New Phytologist*, **208**, 928–935.

Lobell, D.B., Schlenker, W. & Costa-Roberts, J. (2011) Climate trends and global crop production since 1980. *Science*, **333**, 616–620.

Lopes, M.S., El-Basyoni, I., Baenziger, P.S., Singh, S., Royo, C., Ozbek, K. *et al.* (2015) Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change. *Journal of Experimental Botany*, **66**, 3477–3486.

Lupton, F. (2014) *Wheat breeding: its scientific basis*. Dordrecht: Springer.

Luu, K., Bazin, E. & Blum, M.G. (2017) Pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, **17**, 67–77.

Maccaferri, M., Harris, N.S., Twardziok, S.O., Pasam, R.K., Gundlach, H., Spannagl, M. *et al.* (2019) Durum wheat genome highlights past domestication signatures and future improvement targets. *Nature Genetics*, **51**, 885–895.

Marcos-Barbero, E.L., Pérez, P., Martínez-Carrasco, R., Arellano, J.B. & Morcuende, R. (2021) Screening for higher grain yield and biomass among sixty bread Wheat genotypes grown under elevated CO2 and high-temperature conditions. *Plants*, **10**, 1596.

Martynov, S. & Dobrotvorskyi, D. 2012. Genetic resources information and analytical system (Gris) for wheat and triticale. URL http://wheatpedigree.net/ [Online]. Cimmyt International Maize and What Improvement Center. Available: http://wheatpedigree.net/ [Accessed 2020].

Martynov, S.P., Dobrotvorskaya, T.V. & Krupnov, V.A. (2018) Analysis of the distribution of Triticum timopheevii Zhuk. Genetic material in common Wheat varieties (Triticum aestivum L.). *Russian Journal of Genetics*, **54**, 166–175.

Martynov, S.P., Dobrotvorskaya, T.V. & Pukhalskiy, V.A. (2006) Dynamics of genetic diversity in winter common wheat Triticum aestivum L. cultivars released in Russia from 1929 to 2005. *Russian Journal of Genetics*, **42**, 1137–1147.

Mascher, M., Schreiber, M., Scholz, U., Graner, A., Reif, J.C. & Stein, N. (2019) Genebank genomics bridges the gap between the conservation of crop diversity and plant breeding. *Nature Genetics*, **51**, 1076–1081.

Milner, S.G., Jost, M., Taketa, S., Mazón, E.R., Himmelbach, A., Oppermann, M. *et al.* (2018) Genebank genomics highlights the diversity of a global barley collection. *Nature Genetics*, **51**(2), 319–326.

Miralles, D.J., Calderini, D.F., Pomar, K.P. & D'ambrogio, A. (1998) Dwarfing genes and cell dimensions in different organs of wheat. *Journal of Experimental Botany*, **49**, 1119–1127.

Mondal, S., Rutkoski, J.E., Velu, G., Singh, P.K., Crespo-Herrera, L.A., Guzmán, C. *et al.* (2016) Harnessing diversity in Wheat to enhance grain yield, climate resilience, disease and insect Pest resistance and nutrition through conventional and modern breeding approaches. *Frontiers in Plant Science*, **7**, 991.

Nakamura, S., Abe, F., Kawahigashi, H., Nakazono, K., Tagiri, A., Matsumoto, T. *et al.* (2011) A wheat homolog of mother OF FT AND TFL1 acts in the regulation of germination. *The Plant Cell*, **23**, 3215–3229.

N'diaye, A., Haile, J.K., Nilsen, K.T., Walkowiak, S., Ruan, Y., Singh, A.K. *et al.* (2018) Haplotype loci under selection in Canadian durum Wheat germplasm over 60 years of breeding: association with grain yield, quality traits, protein loss, and plant height. *Frontiers in Plant Science*, **9**.

Nielsen, N.H., Backes, G., Stougaard, J., Andersen, S.U. & Jahoor, A. (2014) Genetic diversity and population structure analysis of European Hexaploid bread Wheat (Triticum aestivum L.) varieties. *PLoS One*, **9**, e94000.

Nilsson-Ehle, H. (1913) Über die winterweizenarbeiten in svalöf in den Jahren 1900–1912. *Beiträge zur Pflanzenzucht*, **3**, 62–68.

Oppermann, M., Weise, S., Dittmann, C. & Knüpffer, H. J. D. 2015. Gbis: the information system of the German Genebank. 2015.

Paull, J., Chalmers, K., Karakousis, A., Kretschmer, J., Manning, S. & Langridge, P. (1998) Genetic diversity in Australian wheat varieties and breeding material based on Rflp data. *Theoretical and Applied Genetics*, **96**, 435–446.

Peng, J.H., Sun, D. & Nevo, E. (2011) Domestication evolution, genetics and genomics in wheat. *Molecular Breeding*, **28**, 281.

Perrier, X. & Jacquemoud-Collet, J. P. 2006. DARwin Software. Dissimilarity Analysis and Represntation for windows. URL http://darwin.cirad.fr/.

Poland, J.A. & Rife, T.W. (2012) Genotyping-by-sequencing for plant breeding and genetics. *The Plant Genome*, **5**, 92–102.

Pont, C., Leroy, T., Seidel, M., Tondelli, A., Duchemin, W., Armisen, D. *et al.* (2019) Tracing the ancestry of modern bread wheats. *Nature Genetics*, **51**, 905–911.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Pronin, D., Börner, A., Weber, H. & Scherf, K.A. (2020) Wheat (Triticum aestivum L.) breeding from 1891 to 2010 contributed to increasing yield and Glutenin contents but decreasing protein and gliadin contents. *Journal of Agricultural and Food Chemistry*, **68**, 13247–13256.

Pshenichnikova, T.A., Doroshkov, A.V., Osipova, S.V., Permyakov, A.V., Permyakova, M.D., Efimov, V.M. *et al.* (2019) Quantitative characteristics of pubescence in wheat (Triticum aestivum L.) are associated with

photosynthetic parameters under conditions of normal and limited water supply. *Planta*, **249**, 839–847.

Pshenichnikova, T.A., Doroshkov, A.V., Simonov, A.V., Afonnikov, D.A. & Börner, A. (2017) Diversity of leaf pubescence in bread wheat and relative species. *Genetic Resources and Crop Evolution*, **64**, 1761–1773.

Pujol Andreu, J. 2011. Wheat varieties and technological change in Europe, 19th and 20th centuries: New issues in economic history.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D. *et al.* (2007) Plink: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, **81**, 559–575.

Qaseem, M.F., Qureshi, R., Muqaddasi, Q.H., Shaheen, H., Kousar, R. & Röder, M.S. (2018) Genome-wide association mapping in bread wheat subjected to independent and combined high temperature and drought stress. *PLoS One*, **13**, e0199121.

R Core Team. (2019) *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/

Rahman, S., Islam, S., YU, Z., She, M., Nevo, E. & Ma, W. (2020) Current progress in understanding and recovering the wheat genes lost in evolution and domestication. *International Journal of Molecular Sciences*, **21**, 5836.

Randhawa, M., Bansal, U., Lillemo, M., Miah, H. & Bariana, H. (2016) Postulation of rust resistance genes in Nordic spring wheat genotypes and identification of widely effective sources of resistance against the Australian rust flora. *Journal of Applied Genetics*, **57**, 453–465.

Randhawa, M., Bansal, U., Valárik, M., Klocová, B., DoleŽEL, J. & Bariana, H. (2014) Molecular mapping of stripe rust resistance gene Yr51 in chromosome 4AL of wheat. *Theoretical and Applied Genetics*, **127**, 317–324.

Randhawa, M.S., Bains, N.S., Sohu, V.S., Chhuneja, P., Trethowan, R.M., Bariana, H.S. *et al.* (2019) Marker assisted transfer of stripe rust and stem rust resistance genes into four Wheat cultivars. *Agronomy*, **9**, 497.

Randhawa, M.S., Bariana, H.S., Mago, R. & Bansal, U.K. (2015) Mapping of a new stripe rust resistance locus Yr57 on chromosome 3BS of wheat. *Molecular Breeding*, **35**, 65.

Rasheed, A., Mujeeb-Kazi, A., Ogbonnaya, F.C., He, Z. & Rajaram, S. (2018) Wheat genetic resources in the post-genomics era: promise and challenges. *Annals of Botany*, **121**, 603–616.

Rauf, S., Da Silva, J.T., Khan, A.A. & Naveed, A. (2010) Consequences of plant breeding on genetic diversity. *International Journal of Plant Breeding*, **4**, 1–21.

RAY, D.K., West, P.C., Clark, M., Gerber, J.S., Prishchepov, A.V. & Chatterjee, S. (2019) Climate change has likely already affected global food production. *PLoS One*, **14**, e0217148.

Regina, A., Berbezy, P., Kosar-Hashemi, B., Li, S., Cmiel, M., Larroque, O. *et al.* (2015) A genetic strategy generating wheat with very high amylose content. *Plant Biotechnology Journal*, **13**, 1276–1286.

Regina, A., Bird, A., Topping, D., Bowden, S., Freeman, J., Barsby, T. *et al.* (2006) High-amylose wheat generated by RNA interference improves indices of large-bowel health in rats. *Proceedings of the National Academy of Sciences of The United States of America*, **103**, 3546–3551.

Reif, J.C., Zhang, P., Dreisigacker, S., Warburton, M.L., Van Ginkel, M., Hoisington, D. *et al.* (2005) Wheat genetic diversity trends during domestication and breeding. *Theoretical and Applied Genetics*, **110**, 859–864.

Richards, R. & Passioura, J. (1989) A breeding program to reduce the diameter of the major xylem vessel in the seminal roots of wheat and its effect on grain yield in rain-fed environments. *Australian Journal of Agricultural Research*, **40**, 943–950.

Richards, R. A. & Passioura, J. B. 1981. Seminal root morphology and water use of Wheat II. Genetic Variation1. *Crop Science*, **21**, 253–255.

Rufo, R., Alvaro, F., Royo, C. & Soriano, J.M. (2019) From landraces to improved cultivars: assessment of genetic diversity and population structure of Mediterranean wheat using SNP markers. *PLoS One*, **14**, e0219867.

Salvi, S., Porfiri, O. & Ceccarelli, S. (2013) Nazareno Strampelli, the 'Prophet'of the green revolution. *The Journal of Agricultural Science*, **151**, 1–5.

Sanchez-Garcia, M. & Bentley, A.R. (2019) Chapter 9 - Global Journeys of Adaptive Wheat Genes. In: Miedaner, T. & Korzun, V. (Eds.) *Applications of Genetic and Genomic Research in Cereals*. Duxford: Woodhead Publishing.

Saska, P., Skuhrovec, J., Tylová, E., Platková, H., Tuan, S.-J., HSU, Y.-T. *et al.* (2021) Leaf structural traits rather than drought resistance determine aphid performance on spring wheat. *Journal of Pest Science*, **94**, 423–434.

Schittenhelm, S., Langkamp-Wedde, T., Kraft, M., Kottmann, L. & Matschiner, K. (2020) Effect of two-week heat stress during grain filling on stem reserves, senescence, and grain yield of European winter wheat cultivars. *Journal of Agronomy and Crop Science*, **206**, 722–733.

Schönhofen, A., Hazard, B., Zhang, X. & Dubcovsky, J. (2016) Registration of common wheat germplasm with mutations in Sbeii genes conferring increased grain amylose and resistant starch content. *Journal of plant registrations*, **10**, 200–205.

Sharma, S., Schulthess, A.W., Bassi, F.M., Badaeva, E.D., Neumann, K., Graner, A. *et al.* (2021) Introducing beneficial alleles from plant genetic resources into the Wheat germplasm. *Biology*, **10**, 982.

Shin, J.-H., Blay, S., Mcneney, B. & Graham, J. (2006) LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *Journal of statistical software*, **16**, 9.

Simonov, A., Smirnova, O., Genaev, M. & Pshenichnikova, T. (2021) The identification of a new gene for leaf pubescence introgressed into bread wheat from Triticum timopheevii Zhuk. And its manifestation in a different genotypic background. *Plant Genetic Resources*, **19**(3), 238–244.

Singh, S., Vikram, P., Sehgal, D., Burgueño, J., Sharma, A., Singh, S.K. *et al.* (2018) Harnessing genetic potential of wheat germplasm banks through impact-oriented-prebreeding for future food and nutritional security. *Scientific Reports*, **8**, 12527.

Slade, A.J., Mcguire, C., Loeffler, D., Mullenberg, J., Skinner, W., Fazio, G. *et al.* (2012) Development of high amylose wheat through tilling. *BMC Plant Biology*, **12**, 69.

Smýkal, P., Nelson, M.N., Berger, J.D. & Von Wettberg, E.J. (2018) The impact of genetic changes during crop domestication. *Agronomy*, **8**, 119.

Soleimani, B., Lehnert, H., Keilwagen, J., Plieske, J., Ordon, F., Naseri Rad, S. *et al.* (2020) Comparison between core set selection methods using different Illumina marker platforms: a case study of assessment of diversity in wheat. *Frontiers in Plant Science*, **11**, 1040.

Stodart, B., Mackay, M. & Raman, H. (2008) Assessment of molecular diversity in landraces of bread wheat (Triticum aestivum L.) held in an ex situ collection with diversity arrays technology (DArT™). *Australian Journal of Agricultural Research*, **58**, 1174–1182.

Tetlow, I.J. & Emes, M.J. (2014) A review of starch-branching enzymes and their role in amylopectin biosynthesis. *IUBMB Life*, **66**, 546–558.

Tilman, D., Balzer, C., Hill, J. & Befort, B.L. (2011) Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences*, **108**, 20260–20264.

Trnka, M., Hlavinka, P. & Semenov, M.A. (2015) Adaptation options for wheat in Europe will be limited by increased adverse weather events under climate change. *Journal of the Royal Society Interface*, **12**, 20150721.

Uprety, D.C., Dwivedi, N., Raj, A., Jaiswal, S., Paswan, G., Jain, V. *et al.* (2009) Study on the response of diploid, tetraploid and hexaploid species of wheat to the elevated CO2. *Physiology and Molecular Biology of Plants*, **15**, 161–168.

Venske, E., Dos Santos, R.S., Busanello, C., Gustafson, P. & Costa de Oliveira, A. (2019) Bread wheat: a role model for plant domestication and breeding. *Hereditas*, **156**, 16.

Vergauwen, D. & De Smet, I. (2017) From early farmers to Norman Borlaug—the making of modern wheat. *Current Biology*, **27**, R858–R862.

Voss-Fels, K., Frisch, M., Qian, L., Kontowski, S., Friedt, W., Gottwald, S. *et al.* (2015) Subgenomic diversity patterns caused by directional selection in bread Wheat gene pools. *The Plant Genome*, **8**, 1–13.

Voss-Fels, K.P., Stahl, A., Wittkop, B., Lichthardt, C., Nagler, S., Rose, T. *et al.* (2019) Breeding improves wheat productivity under contrasting agrochemical input levels. *Nature Plants*, **5**, 706–714.

Walkowiak, S., Gao, L., Monat, C., Haberer, G., Kassa, M.T., Brinton, J. *et al.* (2020) Multiple wheat genomes reveal global variation in modern breeding. *Nature*, **588**, 277–283.

Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B.E. *et al.* (2014) Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnology Journal*, **12**, 787–796.

Warnes, G., Gorjanc, G., Leisch, F. & MAN, M. 2013. Genetics: population genetics.R package version *1.3.8.1. URL* http://Cran.R-project.org/package=genetics.

Wendler, N., Mascher, M., NÖH, C., Himmelbach, A., Scholz, U., Ruge-Wehling, B. *et al.* (2014) Unlocking the secondary gene-pool of barley with next-generation sequencing. *Plant Biotechnology Journal*, **12**, 1122–1131.

Williams, K.J., Lewis, J.G., Bogacki, P., Pallotta, M.A., Willsmore, K.L., Kuchel, H. *et al.* (2003) Mapping of a QTL contributing to cereal cyst nematode tolerance and resistance in wheat. *Australian Journal of Agricultural Research*, **54**, 731–737.

Winfield, M.O., Allen, A.M., Burridge, A.J., Barker, G.L., Benbow, H.R., Wilkinson, P.A. *et al.* (2016) High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnology Journal*, **14**, 1195–1206.

Winfield, M.O., Allen, A.M., Wilkinson, P.A., Burridge, A.J., Barker, G.L.A., Coghill, J. *et al.* (2018) High-density genotyping of the A.E. Watkins collection of hexaploid landraces identifies a large molecular diversity compared to elite bread wheat. *Plant Biotechnology Journal*, **16**, 165–175.

Winfield, M.O., Wilkinson, P.A., Allen, A.M., Barker, G.L.A., Coghill, J.A., Burridge, A. *et al.* (2012) Targeted re-sequencing of the allohexaploid wheat exome. *Plant Biotechnology Journal*, **10**, 733–742.

Witcombe, J.R. (1975) Wheat and barley from two Himalayan regions. *Euphytica*, **24**, 431–434.

Worland, A., Korzun, V., Röder, M., Ganal, M. & LAW, C. (1998) Genetic analysis of the dwarfing gene Rht8 in wheat. Part II. The distribution and adaptive significance of allelic variants at the Rht8 locus of wheat as revealed by microsatellite screening. *Theoretical and Applied Genetics*, **96**, 1110–1120.

Zeibig, F., Kilian, B. & Frei, M. (2021) The grain quality of wheat wild relatives in the evolutionary context. *Theoretical and Applied Genetics*, 1–20.

Zeven, A. C. 1990. *Landraces and improved varieties of bread wheat and other wheat forms grown in The Netherlands up to 1944*, Unknown Publisher.

Zhao, Y., Ma, R., XU, D., Bl, H., XIA, Z. & Peng, H. (2019) Genome-wide identification and analysis of the AP2 transcription factor gene family in Wheat (Triticum aestivum L.). *Frontiers in Plant Science*, **10**.

Zhu, X. & Troy, T.J. (2018) Agriculturally relevant climate extremes and their trends in the World's major growing regions. *Earth's Future*, **6**, 656–672.

Zohary, D., Hopf, M. & Weiss, E. (2012) *Domestication of Plants in the Old World: The origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin*. Oxford: Oxford University Press.