Genetic Diversity and Population Structure of Asian and European Common Wheat Accessions Based on Genotyping-By-Sequencing

CURRENT STATUS: POSTED

Xiu Yang
Sichuan Agricultural University

Ling Xi
Sichuan Agricultural University

Binwen Tan
Sichuan Agricultural University

Wei Zhu
Sichuan Agricultural University

Lili Xu
Sichuan Agricultural University

Yi Wang
Sichuan Agricultural University

Xing Fan
Sichuan Agricultural University

Lina Sha
Sichuan Agricultural University

Haiqin Zhang
Sichuan Agricultural University

Jian Zeng
Sichuan Agricultural University

Guoyue Chen
Sichuan Agricultural University

Jian Ma
Sichuan Agricultural University

Xigui Hu
Sichuan Agricultural University
Yonghong Zhou  
Sichuan Agricultural University

Houyang Kang  houyang.kang@sicau.edu.cn  
Sichuan Agricultural University  
Corresponding Author  
ORCiD: 0000-0002-0561-5413

DOI:  
10.21203/rs.2.17640/v1

SUBJECT AREAS  
Plant Molecular Biology and Genetics  Population Genetics

KEYWORDS  
genetic diversity, population structure, genotyping-by-sequencing, single nucleotide polymorphisms, common wheat
Abstract

Background Availability of information on the genetic diversity and population structure of germplasm facilitates its use in wheat breeding programs. Recently, with the development of next-generation sequencing technology, genotyping-by-sequencing (GBS) has been used as a high-throughput and cost-effective molecular tool for examination of the genetic diversity of wheat breeding lines. In this study, GBS was used to characterize a population of 180 accessions of common wheat originating from Asia and Europe between the latitudes 30° and 45°N.

Results In total, 24,767 high-quality single-nucleotide polymorphism (SNP) markers were used for analysis of genetic diversity and population structure. The B genome contained the highest number of SNPs, followed by the A and D genomes. The polymorphism information content ranged from 0.1 to 0.4, with an average of 0.26. The distribution of SNPs markers on the 21 chromosomes ranged from 243 on chromosome 4D to 2,337 on chromosome 3B. Structure and cluster analyses divided the panel of accessions into two subgroups (G1 and G2). G1 principally consisted of European and partial Asian accessions, and G2 comprised mainly accessions from the Middle East and partial Asia. Molecular analysis of variance showed that the genetic variation was greater within groups (99%) than between groups (1%). Comparison of the two subgroups indicated that G1 and G2 contained a high level of genetic diversity. The genetic diversity of G2 was higher as indicated by the Shannon’s information index (I) = 0.512, diversity index (h) = 0.334, observed heterozygosity (Ho) = 0.226, and unbiased diversity index (uh) = 0.338.

Conclusion The present results will not only help breeders to understand the genetic diversity of wheat germplasm on the Eurasian continent between the latitudes of 30° and 45°N, but also provide valuable information for wheat genetic improvement through introgression of novel genetic variation in this region.
Background

Wheat (Triticum aestivum L.) is an important staple food crop for more than one-third of the world’s population and provides about 20% of calories consumed by humans [1, 2]. As a result of ongoing population growth and climate change, it has been estimated that wheat production must increase by 50% by 2050 [1, 3, 4]. However, wheat productivity is expected to decline by 6–8% (and up to 25% in some tropical regions) in the next few years [5]. Thus, it seems that wheat production cannot meet demand. Therefore, a challenge for wheat breeders is to improve the stability of grain production and grain yield to meet the growing demand, and to improve resistance and tolerance to biotic and abiotic stresses [6]. Analysis of plant genetic diversity is an important aspect of plant breeding, inheritance, conservation, and evolution [7]. However, domestication and strong selection pressure by humans, and the use of modern breeding techniques, have increasingly narrowed the gene pool of wheat [8, 9]. Therefore, it is essential to enrich wheat germplasm resources by introducing favorable mutations into the cultivated gene pool [8, 9, 10].

Morphological traits and molecular markers are two distinct tools for assessment of genetic diversity. However, molecular markers have gained substantial attention because morphological traits are often influenced by the environment [11]. Genetic diversity in wheat has been studied using diverse molecular markers, such as random amplified polymorphic DNA [12], random fragment length polymorphisms [13, 14], amplified fragment length polymorphisms [15, 16], sequence-tagged sites [17] and inter-sequence simple repeats [18]. Single-nucleotide polymorphisms (SNPs) and simple sequence repeats (SSRs) are the most commonly used molecular markers for evaluation of genetic diversity among wheat accessions [11, 19]. Furthermore, SNPs are not only the most abundant type of polymorphism in animal genomes but also exhibit a large number of sequence variants.
in plant genomes [20-22]. At present, SNPs are the marker of choice for plant research and plant breeding, such as analyses of marker-trait association, population structure, genomic selection, quantitative trait loci mapping, and research on plant breeding that particularly requires numerous markers [23]. Use of high-throughput sequencing technology to discover a large number of SNPs has proved to be not time-consuming and cost-effective [24]. With the rapid development of next-generation sequencing (NGS) technologies, a new approach for genotyping-by-sequencing (GBS) has been widely used in plant breeding programs [25]. GBS enormously reduces the complexity of genomes of species that show high levels of genetic diversity by choosing appropriate restriction enzymes (REs) [25], such as wheat with large and complex genomes. An improved GBS protocol that uses two restriction enzymes (PstI/MspI) provides a greater degree of complexity reduction and more uniform libraries for sequencing than a one-enzyme protocol [25]. GBS has been used for genotyping in an increasing number of crops, such as maize, wheat, barley, rice, potato, and cassava. Romay et al. (2013) [26] genotyped a set of 2,815 maize inbred accessions using 681,257 SNP markers and observed that some SNPs were linked to the known candidate genes for kernel color, sweetness, and flowering time. Lam et al. (2010) [27] obtained 205,614 SNPs by resequencing 31 soybean genotypes, which offered a precious genomic resource for soybean breeding programs. The GBS protocol was used to analyze genetic diversity of 369 Iranian hexaploid wheat accessions, in which a total of 566,439,207 sequence reads were generated and 133,039 SNPs were identified [28]. A set of 38,412 GBS-SNPs were identified after sequencing 365 soft winter wheat varieties and F$_5$-derived advanced breeding lines originating from multiple crosses in the Cornell University Wheat Breeding Program using a GBS procedure to analyze genetic diversity [29].

The principal region of common wheat cultivation is located between the latitudes of 30°–
60°N and 27°–40°S, mainly concentrated in the 30°–45°N region [30]. The 180 common wheat accessions used in the present study were collected from 16 countries between the latitudes of 30° and 45°N. The germplasm in this region not only provides novel sources of resistance to biotic and abiotic stresses, but also can enhance the biodiversity of breeding materials. To allow comparison between geographic origin and genotype data, the accessions were grouped into three broad geographical regions, namely Asia, the Middle East, and Europe. The main purpose of this study was to use GBS to evaluate the genetic diversity of the accessions from 16 countries between the latitudes of 30°–45°N and, in addition, to explore the genetic relationship and population structure of these accessions from different regions.

Results

**Chromosome distribution of SNPs**

A total of 24,767 SNPs were identified in the A, B, and D genomes. The highest number of SNPs were identified in the B genome (12,028), followed by the A genome (9741), and the D genome had the lowest number of polymorphic markers with 2998 (Fig 1A, Table 1). In the A genome, chromosome 2A had the highest number of SNPs (1761), and chromosome 6A harbored the lowest number (1154); in the B genome, the highest and lowest number of SNPs were detected on chromosomes 3B and 4B (2337 and 1130, respectively); in the D genome, chromosome 2D had the highest number of SNPs (597), and chromosome 4D harbored the lowest number (243) (Fig 1B, Table 1). The lowest and highest numbers of SNPs identified on an individual chromosome were 243 and 2337 on chromosomes 4D and 3B, respectively (Fig 1B, Table 1). The ratio of number of SNPs in the B to A genomes was 1.23, and that in the B to D and A to D genomes was 4.01 and 3.25, respectively. Thus, the number of SNPs in the A and B genomes exceedingly the number in the D genome, and the number of SNPs in the A genome was only slightly lower than that in the B genome.
Population structure

The 180 common wheat accessions were divided into three broad geographical regions to compare geographic origin and genotype data (Fig 2, Table 2). Using principal component analysis (PCA), the relationship between the wheat accessions based on the broad geographical regions was determined. In the PCA plot the accessions showed a loose distribution (Fig 3). The accessions from Asia and the Middle East were distributed evenly on PC1 but were less evenly distributed on PC2. The majority of Asian accessions were placed in the positive (upper) portion of the plot. European accessions were mainly concentrated on the right side (positive values) of PC1. The PC2 explained very little variation at the regional level. STRUCTURE analysis was used to study the population structure of the 180 accessions, and delta K values obtained were used to determine the optimal number of subpopulations. To determine the optimal k value, the number of clusters (K) was plotted. At k = 2 (Fig 4), a distinct peak was observed, indicating the presence of two subpopulations (Fig 5). Group 1 contained 137 accessions; Group 2 consisted of 43 accessions (Table 3). The degree of genetic differentiation and average distance ($H_e$) in each subpopulation (Table 3) suggested that the highest degree of genetic differentiation was detected in Group 2 ($F_{st} = 0.3968$), whereas the lowest value was observed in Group 1 ($F_{st} = 0.1328$). On the other hand, the lowest $H_e$ was observed in Group 2 and the highest $H_e$ was detected in Group 1. The results of STRUCTURE analyses (Fig 5), PCA, and the UPGMA cluster analysis (Fig 6, Additional file 1: Fig S1) showed a high degree of similarity. It was observed that individuals in Group 2 in the STRUCTURE analysis were separated from individuals in Group 1 on PC1. Individual accessions in Group 2 mainly originated from Asia and the Middle East; Group 1 principally consisted of Asian and European accessions. To further understand the clustering relationships among
countries, we take the country of origin into consideration. Accessions from Spain were divided into two clusters: one portion was clustered with European accessions, and the other portion clustered with Asian/Middle Eastern accessions. Accessions from Kyrgyzstan, Kazakhstan, China, and Japan tended to cluster with accessions from Europe. Half of the accessions from Afghanistan were clustered with European accessions, and half was grouped with Asian/Middle Eastern accessions. The Middle Eastern accessions originating in Armenia mainly clustered separate from the European accessions.

The percentage apportioning to ancestral groups for each country was determined to examine the geographic distribution of the two STRUCTURE groups, which were projected onto a world map (Fig 7). This analysis revealed the possible origin of the two ancestral populations predicted by STRUCTURE. The accessions from Europe were predominantly assigned to Group 1 (blue segments in Fig 7), and the accessions from Bulgaria (100%) and Portugal (100%) were assigned to Group 1 (Fig 8). For the four countries in the Middle East, except for accessions from Armenia (67% accessions in Group 2), the majority of accessions were assigned to Group 1 (Fig 8). Half of the accessions from Afghanistan were assigned to Group 1 and half were assigned to Group 2. In addition, 53% lines from Korea were placed in Group 2 and 47% were placed in Group 1 (Fig 8).

Genetic diversity

The genetic diversity analysis of the 180 accessions revealed that the mean GD and PIC were 0.32 and 0.26, respectively. The GD ranged from 0.1 to 0.5 and PIC ranged from 0.1 to 0.4 (Fig 9A, B). The \( H_o \) values ranged from 0 to 0.9, but for a considerable number of markers the \( H_o \) value was 0.1 (Figure 9C). The average MAF was 0.24 (Fig 9D). Intra-population genetic diversity analysis revealed that mean observed \( (N_a) \) and effective \( (N_e) \) allele numbers were 1.999 and 1.513 for the two subpopulations, respectively. The highest
\( N_a \) was observed in Group 1 (2.000), and the value of \( N_e \) in Group 2 (1.521) was higher than that in Group 1 (1.505). The mean values of \( I, H_o, h, \) and \( uh \) were 0.512, 0.226, 0.333, and 0.335, respectively. However, the Group 2 population showed higher genetic diversity \( (I = 0.512, H_o = 0.229, h = 0.334, \text{ and } uh = 0.338) \) (Table 4).

AMOVA and genetic diversity indices for the two subpopulations were calculated based on the results of the STRUCTURE analysis. The AMOVA revealed much greater variation within populations (99%) than among the populations (1%). High haploid \( N_m \) was observed (28.124), suggesting extremely high gene flow among subpopulations (Table 5). These results revealed low genetic differentiation among the subpopulations, but high genetic differentiation within subpopulations. The UPGMA cluster analyses based on 7461 markers (Additional file 2: Fig S2) indicated two subgroups, which was consistent with the population structure analysis (based on 24,767 SNPs).

**Discussion**

Wheat germplasm resources are extremely important for breeders. One important wheat-producing area of the world is located between 30° and 45°N latitude, which is rich in wheat germplasm resources. A prerequisite for making full use of these germplasm resources is to assess their genetic diversity [31]. We used high-throughput sequencing technology to discover a large number of SNPs for genotyping hexaploid wheat derived from diverse provenances in this study. Rufo et al. (2019) [32] genotyped a set of 354 Mediterranean wheat accessions using 13,177 SNP markers and observed that the highest number of SNPs were mapped to the B genome and the lowest number to the D genome. Chao et al. (2009) [33] and Berkman et al. (2013) [34] showed that the highest number of SNPs was mapped to the B genome, followed by the A genome and D genome. However, Cavanagh et al. (2013) [35] reported that the majority of SNPs were located in the A
genome and the lowest number in the D genome. Zhang et al. (2017) [10] used 271 SSR markers to detect the highest number of alleles in the D genome, the lowest number were in the A genome. The highest average number of alleles was detected in the B genome, followed by the D genome and the A genome, using 17 SSR markers, as reported by Salem et al. (2015) [36]. In the present study, we obtained 24,767 SNPs markers and observed the lowest frequency of SNPs in the D genomes, whereas the B genome contained the highest frequency of polymorphic markers, which is in agreement with the results of previous studies [1, 19, 28, 32, 37, 38, 39–41]. Furthermore, the fewest SNP markers were located on chromosome 4D, whereas the highest number of SNP markers were located on chromosome 3B, as reported by Saintenac et al. (2013) [42] and Alipour et al. (2017) [28]. Eltaher et al. (2018) [19] obtained 25,566 SNPs by GBS for 270 F$_{3:6}$ Nebraska winter wheat accessions, and observed that the highest number of SNPs were located on chromosome 3B, whereas chromosome 3D carried the lowest number of SNPs. Bhatta et al. (2017) [2] reported that chromosomes 2B and 4D had the highest and lowest numbers of SNPs, respectively. Chromosome 4D had the lowest number of markers and chromosome 1B had the highest number of markers in the study by Sukumaran et al. (2015) [43]. Allen et al. (2017) [44] used 35,143 SNPs and reported that chromosome 2B had the highest number of markers and chromosome 4D had the lowest number of markers. In contrast, the present study showed that chromosome 3B harbors the highest number of SNPs and chromosome 4D has the lowest number.

The PIC contributes to a detailed understanding of the level of polymorphism between genotypes. On the basis of previous reports, the PIC can be divided into three categories: (1) when PIC > 0.5, the marker is considered to be highly polymorphic, (2) when 0.25 < PIC < 0.5, the marker is a moderately informative, and (3) when PIC < 0.25, the marker is a low-information marker [45]. Lopes et al. (2015) [46] observed a PIC value of 0.27 using
the 9K SNP array to genotype the WAMI population, and showed that spring wheat contained moderate levels of polymorphism. Novoselović et al. (2016) [47] genotyped a Croatian panel using a set of 1229 Diversity Arrays Technology (DArT) markers and obtained an average PIC value of 0.30 among the populations, which indicated that the accessions from Croatia exhibited moderate polymorphism. Eltaher et al. (2018) [19] analyzed 270 F$_{3:6}$ Nebraska winter wheat accessions, and observed a PIC value of 0.25, which indicated that the population contained moderate genetic diversity. El-Esawi et al. (2018) [48] used 1052 DArT markers to genotype Australian and Belgian wheat accessions, and obtained PIC values of 0.33 and 0.29, respectively, which demonstrated that Australian and Belgian wheat contain moderate genetic diversity. The present results showed that the mean PIC value (0.27) was in agreement with the above-mentioned studies, which indicated that the 180 accessions contained moderate polymorphism. On the other hand, Hao et al. (2011) [49] gemotyped 250 Chinese wheat accessions with 512 SSR markers and observed a PIC value of 0.650, which demonstrated that Chinese wheat shows high genetic polymorphism. Zhang et al. (2010) [50] analyzed 205 elite wheat accessions in the USA, using 245 SSR markers, and obtained a PIC value of 0.54, which indicated that the accessions showed a high level of polymorphism. Relative to SSR markers, the lower PIC value of the SNP and DArT markers may be explained by their bi-allelic nature and slow mutation rate [51, 52].

In the present study, we obtained meaningful information on genetic diversity indices in each subpopulation. High levels of genetic diversity were represented in Groups 1 and 2, with genetic diversity detected in Group 2. The results of AMOVA showed that a high level of genetic diversity was observed within subpopulations, whereas the variation among subpopulations was extremely low (1%). This result may be caused by breeders selecting for specific traits, such as yield, stripe rust resistance, and herbicide tolerance. However,
the low genetic variability among subpopulations is explained by the high gene flow [53]. Wright (1965) [54] showed that when $N_m$ (haploid) values are less than 1, gene exchange among subpopulations is limited. In the present study we observed an extremely high $N_m$ value (28.124), indicating that high gene flow led to low genetic differentiation among subpopulations. The results of this study will not only help breeders to understand the genetic diversity of wheat germplasm on the Eurasian continent between the latitudes of 30° and 45°N, but also provides valuable information for genetic improvement of wheat through inclusion of novel genetic variation from China and certain other countries.

The PCA revealed a degree of broad geographic partitioning of the accessions. A previous study by Winfield et al. (2018) [6] used 32,443 polymorphic markers to genotype 804 hexaploid wheat accessions originating from more than 30 countries around the world, and observed that the majority of accessions from Europe clustered together, separate from the majority of Asian and Middle Eastern accessions. Similarly, in the study of Cavanagh et al. (2013) [35], the European winter wheat population showed the strongest degree of genetic differentiation from the remaining populations. Balfourier et al. (2007) [55] used a set of 38 SSR markers to analyze 3942 accessions originating from 73 countries, and observed that accessions from several Near Eastern and Central Asian areas were grouped in the same subcluster and those from Far Eastern countries clustered together. Strelchenko et al. (2005) [56] analyzed 78 wheat landraces originating from 22 countries and reported that the landraces were separated into European and Asian groups. Chen et al. (2019) [57] reported that West Asian landraces, the majority of European landraces, several South and Central Asian landraces, and the majority of East Asian cultivars clustered together, whereas the majority of East Asian landraces were clustered with several West Asian landraces and the majority of South and Central Asian landraces. Lee et al. (2018) [58] reported that many accessions from Afghanistan, Japan, and Korea
were clustered in the same group, while germplasm from China, the Middle East, and Caucasus clustered in a separate group, and an intermediate group largely consisted mainly of accessions from Afghanistan, Japan, and Korea. In the present study, although there was substantial overlap between clusters, the majority of accessions from Europe clustered together, whereas the accessions from Asia and the Middle East were distributed evenly on PC1 (Fig. 4). However, the relationships of three overlapping subgroups was unclear, which raises the possibility of exchanging adapted germplasm. To obtain useful information on the genetic diversity and population structure of the accessions, they were divided into two subgroups on the basis of the population structure analysis (Fig. 5). In the PCA (Fig. 6), genotypes clustered consistent with the subpopulations identified in the STRUCTURE analysis. Moreover, the UPGMA cluster analysis (Additional file 1: Fig S1) was consistent with the results of the STRUCTURE analysis. The majority of European accessions were divided into Group 1, especially accessions from Bulgaria and Portugal (Fig. 8), whereas portions of the Asian and Middle Eastern accessions were distributed in Groups 1 and 2, respectively. The accessions from Turkey, Syria, Georgia, Armenia, Afghanistan, Kyrgyzstan, and Tajikistan showed complex genetic backgrounds, which is not surprising. The area between the Black Sea and the Caspian Sea, and just south of this region (Iraq), is the assumed location of the center of origin of wheat domestication and seems to be a site of population consolidation. Chen et al. (2019) [57] showed that Chinese wheat accessions were mainly derived from European landraces. In the present study, the accessions originating from China tended to cluster with European accessions (Fig. 8).

Conclusion

In this study, a GBS protocol was used to investigate the population structure and genetic diversity of wheat accessions originating from the Eurasian continent between the
latitudes of 30° and 45°N. The panel of 180 accessions was divided into two subgroups, which could be identified by their parentage and selection history. Group 1 principally consisted of European and a portion of Asian accessions, and Group 2 predominantly comprised Middle East and a portion of Asian accessions. Groups 1 and 2 showed high values for genetic diversity indices, which were higher for Group 2. The present results demonstrated that the 180 accessions represent high genetic diversity and can be used for future breeding programs to develop new wheat cultivars with desirable characteristics such as high yield potential, tolerance to biotic and abiotic stress, and good end-use quality, while being well-adapted to diverse environments in China and other countries.

Materials And Methods

Plant materials
A total of 180 common wheat accessions from 16 countries situated between the latitudes of 30° and 45°N were used in this study (Additional file 3: Table S1). The seeds were kindly provided by the Triticeae Research Institute, Sichuan Agricultural University, Sichuan, China, the United States Department of Agriculture–Agricultural Research Service (USDA-ARS)–National Plant Germplasm System, United States, and the Xinjiang Academy of Agricultural Sciences, Xinjiang, China.

Genotyping-by-sequencing
Total genomic DNA was extracted from fresh young leaves of approximately 2-week-old seedlings using the Hi-DNAsecure Plant Kit DP350 (TIANGEN, Beijing, China). GBS libraries were constructed following the protocol of Poland et al. (2012) [37]. A single individual was used for genome sequencing on an Illumina HiSeq PE150 platform. SNP calling was performed using TASSEL v. 5.2.40. The GBS analysis pipeline used the default parameters [59]. The wheat ‘Chinese Spring’ reference genome assembly made available by the International Wheat Genome Sequencing Consortium (IWGSC; RefSeq V1.0) in 2017 was
used. The SNPs were filtered based on the criteria minor allele frequency (MAF) threshold $< 5\%$ and missing values $> 10$ [60-62].

**Population structure analysis**

Evolutionary relationships among the 180 wheat accessions were determined using the unweighted pair group method with arithmetic mean (UPGMA) based on genetic distances computed with TASSEL. Dendrograms were constructed using the dendrogram function, and then customized using the dendextend package [63] and circlize package [64] in R. Principal component analysis (PCA) was performed based on genetic distances among the lines computed with TASSEL [65]. Principal components (PCs) were generated using the covariance method. Eigenvalues were generated to determine the proportion of variation explained by each PC. The first and second PCs were plotted using R.

The population structure of all accessions was evaluated using the Bayesian model-based clustering method in STRUCTURE 2.3.4 software [31]. The STRUCTURE analysis was run five times, with $K$ ranging from 1 to 10, using the admixture model, with burn-in of 100,000 generations and a Markov chain Monte Carlo of 100,000 generations. All parameters were set to default values suggested by the software developer [38]. To identify the best fit for the number of clusters ($K$), the Evanno method was utilized [37] using STRUCTURE HARVESTER software [66]. After selection of the optimal $K$, membership (the proportion of the population assigned to each cluster), mean population differentiation (FST), and He were determined for each subpopulation identified. The FST value [67] of each subpopulation provides an estimate of the degree of fixation of alleles within the subpopulation. The $H_e$ (analogous to allelic variation in a random mating population) [68] describes the average distance between individuals within the same population, where values close to 0 indicate that the individuals within the population are genetically identical.
Statistical analysis

Basic summary statistics were calculated using PowerMarker 3.25 software, comprising gene diversity (GD), polymorphism information content (PIC), MAF, and observed heterozygosity (Ho) [69]. On each chromosome, the SNP markers with a PIC value between 0.21 and 0.33 were selected and a total of 7,461 SNPs were used for AMOVA. The number of subpopulations determined on the basis of a STRUCTURE analysis was used for AMOVA. Genetic indices, consisting of number of loci with a private allele, number of effective alleles (Ne), Ho, expected heterozygosity (He), unbiased expected heterozygosity (uhe), and Shannon’s information index (I) were calculated. The AMOVA and estimation of genetic indices were performed using GeneAlEx 6.41 [70].

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and analyzed during the present study are available from the corresponding author on reasonable request.

Author Contributions

HYK and YHZ conceived and designed the research. XY, LX, and BWT conducted the experiments. WZ, LLX, YW, XF, LNS, JZ, and XGH participated in the preparation of the reagents and materials. HQZ, JM, and GYC analyzed the data. XY and HYK wrote the manuscript. All authors read and approved the manuscript.

Funding
This work was funded by the National Natural Science Foundation of China (No. 31771781, 31971883), the National Key Research and Development Program of China (2016YFD0102000 and 2017YFD0100905), and the Science and Technology Bureau of Sichuan Province.

Competing interests
The authors declare that they have no competing interests.

Acknowledgments
We thank Prof. Jianjiang Li, Xinjiang Academy of Agricultural Sciences, Xinjinag, China, for kindly supplying the materials used in this study. We are also grateful to the reviewers for their valuable comments and recommendations.

References
1. Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer M, International Wheat Genome Sequencing Consortium, et al. Ancient hybridizations among the ancestral genomes of bread wheat. Science. 2014;345:1250092.

2. Bhatta M, Regassa T, Rose DJ, Baenziger PS, Eskridge KM, Santra DK, et al. Genotype, environment, seeding rate, and top-dressed nitrogen effects on end-use quality of modern Nebraska winter wheat. J Sci Food Agric. 2017;97:5311-5318.

3. Grassini P, Eskridge KM, Cassman, KG. Distinguishing between yield advances and yield plateaus in historical crop production trends. Nat. Commun. 2013;4:2918.

4. Ray DK, Mueller ND, West PC, Foley JA. Yield trends are insufficient to double global crop production by 2050. PLoS ONE. 2013;8:e66428.

5. Schleussner CF, Lissner TK, Fischer EM, Wohland J, Perrette M, Golly A, et al. Differential climate impacts for policy-relevant limits to global warming: the case of 1.5 C and 2 C. Earth system dynamics. 2016;7:327-35

6. Winfield MO, Allen AM, Wilkinson PA, Burridge AJ, Barker GLA, Coghill J, et al. High-
density genotyping of the A.E. Watkins Collection of hexaploid landraces identifies a large molecular diversity compared to elite bread wheat. Plant Biotechnol. J. 2018;16:165-175.

7. Peterson GW, Dong Y, Horbach C, Fu YB. Genotyping-by-sequencing for plant genetic diversity analysis: a lab guide for SNP genotyping. Diversity. 2014;6:665-680.

8. Tanksley SD, McCouch SR. Seed banks and molecular maps: unlocking genetic potential from the wild. Science. 1997;277:1063-1066.

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

10. Zhang HX, Zhang FN, Li GD, Zhang SN, Zhang ZG, Ma LJ. Genetic diversity and association mapping of agronomic yield traits in eighty six synthetic hexaploid wheat. Euphytica. 2017;213:111.

11. Huang X, Börner A, Röder M, Ganal M. Assessing genetic diversity of wheat (Triticum aestivum) germplasm using microsatellite markers. Theor. Appl. Genet. 2002;105:699-707.

12. Joshi CP, Nguyen HT. RAPD (random amplified polymorphic DNA) analysis based intervarietal genetic relationships among hexaploid wheats. Plant Sci. 1993;93:95-103.

13. Siedler H, Messmer MM, Schachermayr GM, Winzeler H, Winzeler M, Keller B. Genetic diversity in European wheat and spelt breeding material based on RFLP data. Appl. Genet. 1994;88:994-1003.

14. Kim HS, Ward RW. Patterns of RFLP-based genetic diversity in germplasm pools of common wheat with different geographical or breeding program origins. Euphytica. 2000;115:197-208.
15. Barrett BA, Kidwell KK. AFLP-based genetic diversity assessment among wheat cultivars from the Pacific Northwest. Crop Sci. 1998;38:1261-1271.

16. Burkhamer RL, Lanning SP, Martens RJ, Martin JM, Talbert LE. Predicting progeny variance from parental divergence in hard red spring wheat. Crop Sci. 1998;38:243-248.

17. Chen HB, Martin JM, Lavin M, Talbert LE. Genetic diversity in hard red spring wheat based on sequence-tagged-site PCR markers. Crop Sci. 1994;34:1628-1632.

18. Nagaoka T, Oghihara Y. Applicability of inter-simple sequence repeat polymorphisms in wheat for use as DNA markers in comparison to RFLP and RAPD markers. Theor. Appl. Genet. 1997;94:597-602.

19. Eltaher S, Sallam A, Belamkar V, Emara, HA, Nower, AA, Salem, KFM, et al. Genetic Diversity and Population Structure of F3:6 Nebraska Winter Wheat Genotypes Using Genotyping-By-Sequencing. Genet. 2018;9:76.

20. Ganal MW, Altmann T, RoÈder MS. SNP identification in crop plants. Curr. Opin. Plant. Biol. 2009;12:211-217.

21. Ganal MW, Polley A, Graner EM, Plieske J, Wieseke R, Luerssen H, et al. Large SNP arrays for genotyping in crop plants. J Biosci. 2012;37:821-828.

22. Rimbert H, Darrier B, Navarro J, Kitt J, Choulet F, Leveugle M, et al. High throughput SNP discovery and genotyping in hexaploid whea. PLoS ONE. 2018;13:e0186329.

23. Kumar S, Banks TW, Cloutier S. SNP discovery through next-generation sequencing and its applications. Int. J. Plant Genomics. 2012;2012.

24. He J, Zhao X, Laroche A, Lu ZX, Liu HK, Li Z. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Frontiers in plant sci. 2014;5:484.

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

26. Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, Casstevens TM, et al. Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol. 2013;14:R55.

27. Lam HM, Xu X, Liu X, Chen W, Yang G, Wong FL, et al. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nat. Genet. 2010;42:1053-1059.

28. Alipour H, Bihamta MR, Mohammadi V, Peyghambari SA, Bai G, Zhang G. Genotyping-by-Sequencing (GBS) Revealed Molecular Genetic Diversity of Iranian Wheat Landraces and Cultivars. Frontiers in plant sci. 2017;8:1293.

29. Heslot N, Rutkoski J, Poland J, Jannink JL, Sorrells ME. Impact of Marker Ascertainment Bias on Genomic Selection Accuracy and Estimates of Genetic Diversity. PLoS ONE. 2013;8:e74612.

30. Nuttonson MY. Wheat-climatic relationships and the use of phenology in ascertaining the thermal and photothermal requirements of wheat. 1957;163

31. Hawkes JG. Germplasm collection, preservation, and use. In 2 Plant breeding symposium. Iowa State University Press, USA, 1981.

32. Rufo R, Alvaro F, Royo C, Soriano JM. From landraces to improved cultivars: Assessment of genetic diversity and population structure of Mediterranean wheat using SNP markers. PLoS ONE. 2019;14:e0219867.

33. Chao S, Zhang W, Akhunov E, Sherman J, Ma Y, Luo MC, et al. Analysis of gene-derived SNP marker polymorphism in US wheat (Triticum aestivum L.) cultivars. Mol. Breed. 2009;23:23-33.

34. Berkman PJ, Visendi P, Lee HC, Stiller J, Manoli S, Lorenc MT, et al. Dispersion and
domestication shaped the genome of bread wheat. Plant Biotechnol. J. 2013;11:564-571.

35. Cavanagh CR, Chao S, Wang S, Huang BE, Stephen S, Kiani S, et al. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proc. Natl. Acad. Sci. U.S.A. 2013;10:8057-8062.

36. Salem KF, Röder MS, Börner A. Assessing genetic diversity of Egyptian hexaploid wheat (Triticum aestivum L.) using microsatellite markers. Genetic resources and crop evolution. 2015;62:377-385.

37. Poland JA, Brown PJ, Sorrells ME, Jannink J. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE. 2012;7:e32253.

38. Akhunov ED, Akhunova AR, Anderson OD, Anderson JA, Blake N, Clegg MT, et al. Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. BMC Genom. 2010;11:702.

39. Würschum T, Langer SM, Longin CFH, Korzun V, Akhunov E, Ebmeyer E, et al. Population structure, genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers. Theor. Appl. Genet. 2013;126:1477-1486.

40. Shavrukov Y, Suchecki R, Eliby S, Abugalieva A, Kenebayev S, Langridge P. Application of next-generation sequencing technology to study genetic diversity and identify unique SNP markers in bread wheat from Kazakhstan. BMC plant Biol. 2014;14:258.

41. Edae EA, Bowden RL, Poland J. Application of population sequencing (POPSEQ) for ordering and imputing genotyping-by-sequencing markers in hexaploid wheat. G3 2015;5:2547-2553.
42. Saintenac C, Jiang D, Wang S, Akhunov E. Sequence-based mapping of the polyploid wheat genome. G3: Genes, Genomes, Genetics. 2013;3:1105-1114.

43. Sukumaran S, Dreisigacker S, Lopes M, Chavez P, Reynolds MP. Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments. Theor. Appl. Genet. 2015;128:353-363.

44. Allen AM, Winfield MO, Burridge AJ, Downie RC, Benbow HR, Barker GLA, et al. Characterization of a Wheat Breeders’Array suitable for high-throughput SNP genotyping of global accessions of hexaploid bread wheat (Triticum aestivum). Plant Biotechnol. J. 2017;15:390-401.

45. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 1980;32:314-331.

46. Lopes MS, Dreisigacker S, Peña RJ, Sukumaran S, Reynolds MP. Genetic characterization of the wheat association mapping initiative (WAMI) panel for dissection of complex traits in spring wheat. Theor. Appl. Genet. 2015;128:453-464.

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

48. El-Esawi MA, Witczak J, Abomohra AE, Ali HM, Elshikh MS, Ahmad M. Analysis of the genetic diversity and population structure of Austrian and Belgian wheat germplasm within a regional context based on DArT markers. Genes. 2018;9:47.

49. Hao C, Wang L, Ge H, Dong Y, Zhang X. Genetic diversity and linkage disequilibrium in Chinese bread wheat (Triticum aestivum L.) revealed by SSR markers. PLoS ONE. 2011;6:e17279.

50. Zhang D, Bai G, Zhu C, Yu J, Carver BF. Genetic diversity, population structure, and
linkage disequilibrium in US elite winter wheat. The Plant Genome. 2010;3:117-127.

51. Thuillet AC, Bru D, David J, Roumet P, Santoni S, Sourdille P, et al. Direct estimation of mutation rate for 10 microsatellite loci in durum wheat, Triticum turgidum (L.) Thell. ssp durum desf. Mol. Biol. Evol. 2002;19:122-125.

52. Chesnokov YV, Artemyeva AM. Evaluation of the measure of polymorphism information of genetic diversity. Сельскохозяйственная биология. 2015;5:571-578.

53. Arora A, Kundu S, Dilbaghi N, Sharma I, Tiwari R. Population structure and genetic diversity among Indian wheat varieties using microsatellite (SSR) markers. Aust. J. Crop. Sci. 2014;8:1281-1289.

54. Wright S. The interpretation of population structure by F-statistics with special regard to systems of mating author (s): Sewall wright reviewed work (s): published by: society for the study of evolution stable. Evolution 1965;19:395-420.

55. Balfourier F, Roussel V, Strelchenko P, Exbrayat-Vinson F, Sourdille P, Boutet G, et al. A worldwide bread wheat core collection arrayed in a 384-well plate. Theor. Appl. Genet. 2007;114:1265-1275.

56. Strelchenko P, Street K, Mitrofanova O, Mackay M, Balfourier F, Aleppo S. Genetic diversity among hexaploid wheat landraces with different geographical origins revealed by microsatellites: comparison with AFLP, and RAPD data. Proc. 4th Crop Sci. 2005.

57. Chen H, Jiao C, Wang Y, Wang Y, Tian C, Yu H, et al. Comparative Population Genomics of Bread Wheat (Triticum aestivum) Reveals Its Cultivation and Breeding History in China. BioRxiv. 2019;519587.

58. Lee S, Choi YM, Lee MC, Oh, S Jung Y. Geographical comparison of genetic diversity in Asian landrace wheat (Triticum aestivum L.) germplasm based on high-molecular-weight glutenin subunits. Genetic resources and crop evolution. 2018;65:1591-1602.
59. Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, et al. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS ONE. 2014;9:e90346.

60. Li H, Vikram P, Singh RP, Kilian A, Carling J, Song J, et al. A high density GBS map of bread wheat and its application for dissecting complex disease resistance traits. BMC Genom. 2015;16:216.

61. Saint PC, Burgueno J, Crossa J, Dávila GF, Lopez PF, Moya ES, et al. Genomic prediction models for grain yield of spring bread wheat in diverse agro-ecological zones. Sci. Rep. 2016;6:27312.

62. Vikram P, Franco J, Burgueño-Ferreira J, Li H, Sehgal D, Saint PC, et al. Unlocking the genetic diversity of Creole wheats. Sci. Rep. 2016;6:23092.

63. Galili T. Dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics. 2015;31:3718-3720.

64. Gu Z, Gu L, Eils R, Schlesner M, Brors B. Circlize implements and enhances circular visualization in R. Bioinformatics. 2014;30:2811-2812.

65. Bradbury PJ, Zhang Z, Kroon DE, Casstevens, TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 2007;23:2633-2635.

66. Earl DA, vonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet Resour. 2012;4:359-361.

67. Nei M. F-statistics and analysis of gene diversity in subdivided populations. Ann Hum Gen. 1977;41:225-233.

68. Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics. 1978;89:583-590.

69. Liu K, Muse SV. PowerMarker: an integrated analysis environment for genetic marker
analysis. Bioinformatics. 2005;21:2128-2129.

70. Peakall R, Smouse PE. GENALEX 6: genetic analysis in excel. Population genetic software for teaching and research. Mol. Ecol. Notes. 2006;6288-295.

Tables

Table 1. Chromosomal distribution of the SNP markers used in the study.

| Chromosome No. | A    | B    | Genome | Total |
|---------------|------|------|--------|-------|
| 1             | 1442 | 1379 | 426    | 3247  |
| 2             | 1761 | 1829 | 597    | 4187  |
| 3             | 1217 | 2337 | 460    | 4014  |
| 4             | 1358 | 1130 | 243    | 2731  |
| 5             | 1191 | 1666 | 477    | 3334  |
| 6             | 1154 | 1953 | 282    | 3389  |
| 7             | 1618 | 1734 | 513    | 3865  |
| Total         | 9741 | 12028| 2998   | 24767 |

Table 2. Provenance of the 180 common wheat accessions used in the study.

| Region       | Country       | No. Acc |
|--------------|---------------|---------|
| Asia         | Afghanistan   | 12      |
|              | Kyrgyzstan    | 4       |
|              | KA Kazakhstan | 4       |
|              | Tajikistan    | 4       |
|              | China         | 12      |
|              | Korea         | 15      |
|              | Japan         | 10      |
| Middle East  | Turkey        | 10      |
|              | Syria         | 10      |
|              | Georgia       | 14      |
|              | Armenia       | 9       |
| Europe       | Bulgaria      | 20      |
|              | Greece        | 14      |
|              | Italy         | 17      |
|              | Spain         | 13      |
|              | Portugal      | 12      |

Table 3. Results of STRUCTURE analysis of 180 wheat accessions for the fixation index ($F_{st}$; indicating significant divergence), average distances (expected heterozygosity), and number of genotypes in each subpopulation.

| Population | $F_{st}$ | Exp.hetero | No. of Genotypes |
|------------|---------|------------|------------------|
| G1         | 0.1328  | 0.3132     | 137              |
| G2         | 0.3968  | 0.2413     | 43               |

Table 4. Means of genetic parameters for each subpopulation of the 180 wheat accessions. Number of alleles ($N_a$), number of effective allele ($N_e$), Shannon’s index ($I$), observed heterozygosity ($H_o$), diversity index ($h$), and unbiased diversity index ($uh$).
Table 5 Analysis of molecular variance using 24,767 SNP markers of genetic differentiation among and within three subpopulations of the 180 common wheat accessions.

| Source          | df | SS     | MS     | Est. Var. | %  | P value |
|-----------------|----|--------|--------|-----------|----|---------|
| Among Pops      | 1  | 5599.573 | 5599.573 | 31.467 | 1% | 0.001   |
| Within Pops     | 178| 630097.271 | 3539.872 | 99%     | 0.001 |
| Total           | 179| 635696.844 | 3571.339 | 100%    | 0.001 |
| Nm              |    |         |        | 28.124   |    |         |

Additional Files

**Additional file 1: Fig S1.** UPGMA clustering dendrogram generated using 24,767 SNPs and 180 common wheat accessions. Colors reflect groups derived from STRUCTURE analysis.

**Additional file 2: Fig S2.** Cluster analysis based on the genetic distances among the 180 accessions for 7461 SNPs used for AMOVA.

**Additional file 3: Table S1.** Details of the 180 common wheat accessions used in the study.