Elucidating SNP-based genetic diversity and population structure of advanced breeding lines of bread wheat (*Triticum aestivum* L.)

Vipin Tomar[1,2,3], Guriqbal Singh Dhillon[4], Daljit Singh[5], Ravi Prakash Singh[6], Jesse Poland[7], Arun Kumar Joshi[1,3,6], Budhi Sagar Tiwari[2] and Uttam Kumar[1,3,6]

[1] Borlaug Institute for South Asia, New Delhi, Delhi, India
[2] Department of Biological Sciences and Biotechnology, Institute of Advanced Research, Gandhinagar, Gandhinagar, Gujarat, India
[3] International Maize and Wheat Improvement Centre, New Delhi, Delhi, India
[4] Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala, Punjab, India
[5] The Climate Corporation, Bayer Crop Science, Creve Coeur, MO, USA
[6] Global Wheat Program, International Maize and Wheat Improvement Centre, Texcoco, Mexico
[7] Department of Plant Pathology, Kansas State University, Manhattan, KS, United States of America

Corresponding authors
Vipin Tomar, viomics@gmail.com
Uttam Kumar, u.kumar@cgiar.org,
kumaruttam@hotmail.com

ABSTRACT

Genetic diversity and population structure information are crucial for enhancing traits of interest and the development of superlative varieties for commercialization. The present study elucidated the population structure and genetic diversity of 141 advanced wheat breeding lines using single nucleotide polymorphism markers. A total of 14,563 high-quality identified genotyping-by-sequencing (GBS) markers were distributed covering 13.9 GB wheat genome, with a minimum of 1,026 SNPs on the homoeologous group four and a maximum of 2,838 SNPs on group seven. The average minor allele frequency was found 0.233, although the average polymorphism information content (PIC) and heterozygosity were 0.201 and 0.015, respectively. Principal component analyses (PCA) and population structure identified two major groups (sub-populations) based on SNPs information. The results indicated a substantial gene flow/exchange with many migrants (Nm = 86.428) and a considerable genetic diversity (number of different alleles, Na = 1.977; the number of effective alleles, Ne = 1.519; and Shannon's information index, $I = 0.477$) within the population, illustrating a good source for wheat improvement. The average PIC of 0.201 demonstrates moderate genetic diversity of the present evaluated advanced breeding panel. Analysis of molecular variance (AMOVA) detected 1% and 99% variance between and within subgroups. It is indicative of excessive gene traffic (less genetic differentiation) among the populations. These conclusions deliver important information with the potential to contribute new beneficial alleles using genome-wide association studies (GWAS) and marker-assisted selection to enhance genetic gain in South Asian wheat breeding programs.

INTRODUCTION

Bread wheat (*Triticum aestivum* L.) is an allohexaploid species originating from successive rounds of hybridization in the Fertile Crescent during the Neolithic time, ~8,000 to 10,000 years ago (*Sansaloni et al., 2020*). Wheat is the largest contributor to world grain production, with nearly 30% of total grain production and 50% of the world grain trade (*Akter & Rafiqul Islam, 2017*). Wheat grain is among the most consumed grains worldwide, providing 15% of calories every day, covering more than 220 million hectares with almost 750 metric megatons production every year (*Balfourier et al., 2019*). The global demand for wheat has increased due to population growth, changing food consumption habits, and socio-economic environments, specifically in African and Asian countries (*Mondal et al., 2016*).

During the domestication process, a substantial loss of diversity resulted in genetic bottlenecks. Researchers have been interested in using the genetic diversity of *Triticeae* species. Some of the vital gene pools include *Agropyron, Aegilops, Elymus, Leymus, Hordeum, Secale, Triticum,* and *Thinopyrum*. Above mentioned rich gene pools could improve various traits such as tolerance to biotic and abiotic stresses and micronutrient availability. Novel alleles from around fifty-two species have been introgressed, pointing out the significance of exotic introgressions in the breeding (*Wulft & Moscou, 2014*).

Population structure, genetic diversity, and relationships among genotypes are vital in scheming appropriate breeding plans (*Peterson et al., 2014*; *Singh, Agarwal & Yadav, 2019*). The population structure information aids in estimating the accurate association between phenotypic and genotypic variation (*Knowler et al., 1988*). The population structure information allows researchers to utilize natural diversity to detect vital genes/QTLs using current genetic technologies (*Zhu et al., 2008*). Genetic diversity studies have advanced from mere detection of distinct morphological to molecular traits investigations of DNA variation (*Zhang, Maroof & Kleinhofs, 1993*). Determining the current genetic diversity of crops has paramount importance for the selection and conservation of parents with the various genetic framework, thus providing well-organized crop enhancement (*Mengistu, Kiros & Pè, 2015*).

The first generation of morphological markers could not present the actual picture of diversity because of their limited number, lack of information about environmental and epistatic interaction. Slowly the paradigm shifted to investigate genetic diversity and population structure using PCR (polymerase chain reaction) based markers, which gave a better view of underlying diversity markers due to their abundance and environmentally neutral nature (*Farahani et al., 2019*). Among various molecular markers, the SNP marker is the preferred molecular marker in several crops due to extensive genome coverage, chromosome-specific location, low cost, co-dominant inheritance, and fast-tracking compared to other PCR-based markers. SNP are evolutionarily stable due to lower rates of recurrent mutation. SNPs are considered first-rate to understand genomic evolution and to study complex traits. SNPs have been extensively used in genetic resource characterization, genome-wide association studies (GWAS), genomic selection (GS), and marker-assisted selection (MAS) (*Ganal, Altmann & Röder, 2009*).

Microarrays have been used as a pre-eminent answer to develop SNPs in polyploid wheat genomes (*Wang et al., 2014*). Once an all-inclusive SNP data is accessible for species, a cost-effective microarray may be formed, and the process is relatively convenient. Microarray evades the miscalling risk of diversity on homoeologous genomes and lately with amplified 100 fold in wheat moving from 9 K (*Cavanagh et al., 2013*) to 820 K (*Winfield et al., 2016*). The 90 K wheat SNP array (*Wang et al., 2014*) has been effectively utilized for genetic diversity investigation, the building of high-density maps, and GWAS (*Maccaferri et al., 2015*; *Chen et al., 2016*; *Mengistu et al., 2016*; *Wen et al., 2017*). The low cost of genotyping by sequencing (GBS) makes it a robust approach for discovering and genotyping SNPs in various crops. GBS has been efficiently applied in genomic diversity, genetic linkage studies, and genomic selection in largescale plant breeding programs. GBS is found to be a perfect platform from single-gene to whole-genome sequencing and suited to genetic analysis and marker development (*Heffner, Sorrells & Jannink, 2009*; *Heffner et al., 2010*; *Jannink, Lorenz & Iwata, 2010*; *Fu & Peterson, 2011*; *Poland & Rife, 2012*; *Poland et al., 2012*; *Thomson et al., 2012*; *Lu et al., 2013*; *Fu, Cheng & Peterson, 2014*). The GBS has been established with maize and barley inbred populations with roughly 200,000 and 25,000 sequence tags, respectively (*Elshire et al., 2011*).

In the current study, GBS was used for genotyping 141 elite advanced breeding lines of spring wheat from CIMMYT (Mexico). The objectives were to illustrate the population structure and genetic diversity within and among subgroups. The present study not only defines the population structure and genetic diversity in these elite advanced wheat breeding lines also places a groundwork for genome-wide association study in this panel.

MATERIALS AND METHODS

Plant genetic material

One hundred forty-one advanced breeding lines (ABLs) developed at CIMMYT, Mexico using adapted cultivars from various worldwide breeding programs were used in studying genetic diversity. These lines have been selected from a more extensive set of advanced breeding lines sent to South Asia as part of the CIMMYT wheat breeding program to develop high-yielding varieties suitable for the region. The lines were introduced in 2016 as the 4th cohort of the South Asia Bread Wheat Genomic Prediction Yield Trial (SABWGPYT). The lines were developed by making crosses with diverse and high yielding parental lines as part of the CIMMYT Global Wheat Program (GWP). A set of approx. thirty-eight thousand lines were planted in small plots in Obregon during the 2015-16 crop season. The selection was made based on yield and agronomic traits. Then lines from different genetic backgrounds were carefully selected as candidate varieties suitable for further testing in South Asia. The pedigree details data of those 141 advanced breeding lines are given in Table S1.

Genotyping-by-sequencing and SNP filtering

Genomic DNA was extracted using the modified CTAB method from the fresh leaves of wheat seedlings (*Dreisigacker et al., 2016*). Genotyping-by-sequencing was performed in Illumina HiSeq 2500 using protocol from *Poland et al. (2012)*. SNP calling was performed

using TASSEL v5.2.6 (*Bradbury et al., 2007*) using the TASSEL-GBSv2 pipeline and aligned to the reference Chinese Spring Wheat Assembly (RefSeq v1.0). Beagle v4.1 with default settings was used to impute missing data. After filter criteria quality control (sample call rate >0.8, MAF ≥ 0.05, SNP call rate >0.7), 14,563 polymorphic SNPs and 141 genotypes were selected for further analysis.

Genetic analysis of SNPs

The genotypic summary of the 14,563 SNPs was obtained using the "geno summary" function of TASSEL v5.2.6 (*Bradbury et al., 2007*). Chromosome-wise genomic SNP distribution, minor allele frequency (MAF), observed heterozygosity, and the polymorphism information content (PIC) were performed using GBS-based SNP markers. The PIC value of each biallelic SNP marker was computed using the following method derived from (*Botstein et al., 1980*):

$$PIC = 1 - (MAF^2 + (1 - MAF)^2) - (2MAF^2(1 - MAF)^2)$$

Population structure analysis

The population structure of the advanced bread wheat breeding lines was inferred using the Bayesian method implemented in STRUCTURE 2.3.4 (*Pritchard, Stephens & Donnelly, 2000*). Ten individualistic evaluations of every k were used. The STRUCTURE was placed on 100,000 burn-in, subsequently 100,000 Markov chain Monte Carlo (MCMC) replications. The best k for the present population was determined using the evanno algorithm (*Evanno, Regnaut & Goudet, 2005*) implemented in Pophelper v2.3.1 (*Francis, 2017*). Optimal K/sub-population was identified by the delta k-value peaks across various k values. Principal component analysis (PCA) based on covariates was performed of the SNP data in Tassel 5.2.6 (*Bradbury et al., 2007*) and plotted using Plotly v4.9.3 (*Sievert, 2020*). The PCAs were used to construct dendrograms using r package ape v5.4-1 (*Paradis & Schliep, 2019*) using the complete linkage method for hierarchical clustering.

Genetic diversity analysis and analysis of molecular variance (AMOVA)

The selected K value from structure output was used to subdivide the advanced wheat lines into sub-populations and was utilized for AMOVA. AMOVA and genetic diversity indices were performed for individual sub-population using GeneAlEx v6.41 (*Peakall & Smouse, 2006*). The percentage of molecular variance among and within subgroups was calculated from AMOVA. The haploid number of migrants (Nm) was calculated using the among-population variance (Va) and within-population variance (Vw) using the formula

$$N_m = \left[\left\{ 1 / \left(\frac{V_a}{V_a + V_w} \right) \right\} - 1 \right] / 2$$

The calculation of Shannon's Information Index (I), effective alleles (Ne), different alleles (Na), number of loci with private alleles, unbiased diversity (uh), and the haploid genetic diversity index (h) were also performed.

Na was calculated by direct count of alleles across subpopulations per loci and averaged by the arithmetic mean across loci per sub-population. Ne was calculated using expected heterozygosity by formula

$$Ne = \frac{1}{1 - He} \text{ where } He = 1 - \sum p^2$$

Here p is the frequency of the allele. I was calculated (per locus and averaged across the number of loci) using the formula.

$$I = -\sum p_i \, ln \, p_i$$

Where ln is the natural logarithm of p_i, i.e., frequency of ith allele, private alleles are the alleles unique to the subpopulation. The haploid genetic diversity index (h) provides the probability that the two individuals would be different and was calculated using the formula.

$$h = 1 - \sum p_i^2$$

The unbiased diversity (uh) was calculated using the allele frequency and sample size (n) with a formula.

$$uh = \frac{n}{n-1} \left(1 - \sum p_i^2 \right)$$

RESULTS

SNPs distribution on the wheat genome

A total of 14,563 bi-allelic SNPs were equitably distributed across the three genomes. The SNP HapMap and SNP positions datasets correspond to the reference Chinese Spring Wheat Assembly (RefSeq v1.0) used for present analysis have been provided through the link (https://doi.org/10.6084/m9.figshare.14273339.v2.). The Chinese Spring Wheat Assembly (RefSeq v1.0) reference genome assembly used in the present study is a widely used reference genome in hexaploid wheat. (https://urgi.versailles.inra.fr/download/iwgsc/IWGSC_RefSeq_Assemblies/v1.0/ and https://urgi.versailles.inra.fr/download/iwgsc/IWGSC_RefSeq_Annotations/v1.0/). The B-genome had a maximum of 7378 SNPs (50.66%), subsequently the A-genome with 5921 SNPs (40.66%) and the D-genome with a minimum number of SNPs, i.e., 1264 (8.68%) (Table 1, Fig. 1A). With 14,563 markers across the genome and genomic coverage of 13.9 GB, the average marker density was one marker per 0.95 Mb. The highest marker density with one marker every 0.54 Mb was observed on chromosome 2B, and the lowest density with one marker per 6.854 Mb on chromosome 4D (Fig. S1). The SNPs per chromosome ranged from 74 (4D) to 1486 (2B). The minimum and maximum SNPs detected on individual chromosomes ranged from 574 (4A) to 1293 (7A), 378 (4B) to 1486 (2B), and 74 (4D) to 255 (1D) at A, B, and D genome. The homoeologous group 7 had a maximum number of SNPs, 2838, followed by group 2 with 2652 SNPs, and group 4 had a minimum number of 1026 SNPs.

Table 1 Chromosome wise distribution of SNP markers across the wheat genomes.

Group	SNPs	First SNP	Last SNP	Coverage	Dist. b/w SNP	Density
1A	686	1.145	593.790	592.645	0.864	1.158
1B	881	1.299	688.328	687.028	0.780	1.282
1D	255	0.079	493.979	493.900	1.937	0.516
2A	912	0.718	771.354	770.635	0.845	1.183
2B	1486	0.019	800.870	800.851	0.539	1.856
2D	254	8.791	649.074	640.282	2.521	0.397
3A	832	0.608	750.539	749.931	0.901	1.109
3B	1027	0.580	829.535	828.955	0.807	1.239
3D	115	24.795	615.062	590.267	5.133	0.195
4A	574	2.705	744.515	741.810	1.292	0.774
4B	378	1.056	666.052	664.996	1.759	0.568
4D	74	3.238	509.798	506.560	6.845	0.146
5A	799	1.294	709.755	708.461	0.887	1.128
5B	1084	8.072	712.941	704.869	0.650	1.538
5D	106	62.140	555.015	492.875	4.650	0.215
6A	825	0.684	617.839	617.155	0.748	1.337
6B	1227	0.196	720.805	720.609	0.587	1.703
6D	210	0.070	473.287	473.217	2.253	0.444
7A	1293	2.083	736.572	734.489	0.568	1.760
7B	1295	1.021	750.125	749.105	0.578	1.729
7D	250	2.145	635.422	633.277	2.533	0.395
Genome A	5921			4915.126	0.830	0.090
Genome B	7378			5156.413	0.699	0.057
Genome D	1264			3830.378	3.030	0.030
Total	14563			13901.92	1.048	0.955

Notes.
  # Position and coverage of SNPs is in megabases (Mb), Distances between SNPs in Mb, Density = No of SNPs per MB

Genetic analysis of SNPs

The heterozygosity among all SNPs ranged from 0.00 to 0.149. Twenty-one SNPs showed high heterozygosity (>0.10), and 3,095 SNPs showed zero heterozygosity (Fig. 1B, Table S2). The heterozygosity was less than 0.05 for around 97% of the SNPs. SNPs in chr7A showed the highest heterozygosity, with SNP S7A_717968190 having heterozygosity of 0.149. The three ABLs, namely GID7396143, GID7399653, and GID7400318, were highly heterologous with heterozygosity equal to 0.16, 0.15, and 0.12, respectively (Table S3). The MAF ranged from 0.05 to 0.50 in the present study. Fifty-five SNPs were found with MAF =0.50 (where both the alleles of these SNPs were equally distributed across the panel). The average observed MAF was 0.23, while most ranged from 0.05 to 0.10, i.e., 2923 SNPs (Fig. 1C). The PIC values ranged from 0.085 to 0.250, with 54 SNPs having a PIC value equal to 0.250 (Fig. 1D). 5353 SNPs (36%) showed a range of 0.24−0.25 for PIC and were the highest among the SNPs.
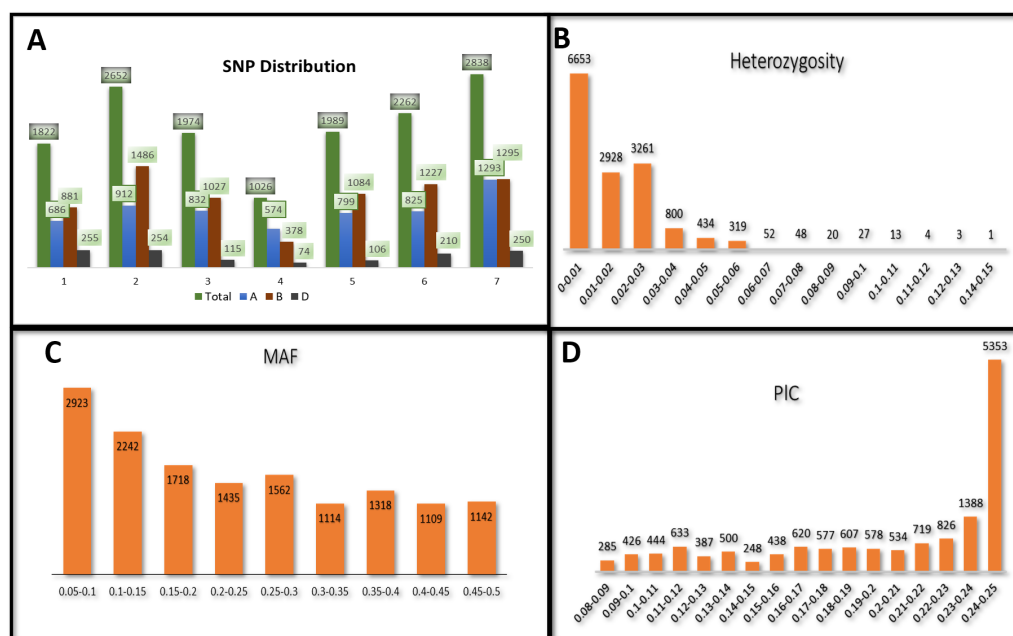
Figure 1 **Distribution of 14,563 SNPs based on chromosomal distribution and markers characteristics.** (A) Distribution of SNP markers across 21 chromosomes of bread wheat for 141 advanced wheat lines, (B) distribution of the percentage of heterozygosity (He), (C) distribution of the minor allele frequency (MAF), and (D) distribution of the polymorphic information content (PIC).

Population structure analysis

The STRUCTURE results for the 141 ABLs showed that ΔK was highest at $K = 2$, indicating the presence of two major subgroups (Fig. 2A). The output of the STRUCTURE for $K = 2$ is given in Fig. 2B. The structure out showed very high ΔK for $K = 2$, compared to other K values, only a few genotypes represented complete distinction between the subgroups. Some genotypes across the subgroups showed near equal distribution of alternative alleles. Similar results were revealed by the PCA analysis where the genotypes of the subgroups could be seen clustered near each other when the first three principal components were observed (Fig. 2C). The dendrogram based on PCA also showed similar results (Fig. 3).

Analysis of molecular variance (AMOVA)

The analysis was performed considering $K = 2$. The variance among groups for $K = 2$ explained only 1% of the total variance and 99% variance within the subgroups. It indicates that the genetic differentiation within subgroups was high. Simultaneously, it was low among the subgroups (Table 2, Fig. 4), which may be due to the low genetic differentiation or excessive gene traffic between the population set. The high gene traffic is supported by the variation trends of both percentages of variance explained and haploid Nm values. The Nm values (haploid number of migrants) was very high, with Nm = 86.428.
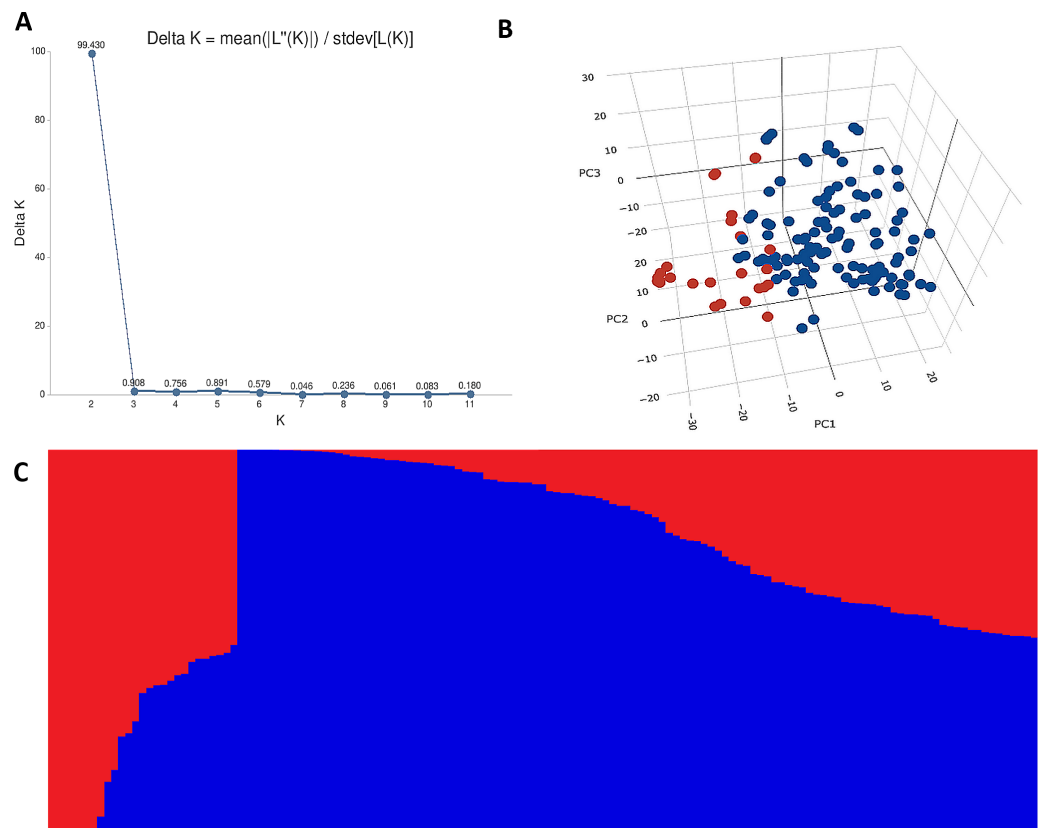
Figure 2 **Population structure analysis of 141 advanced bread wheat lines.** STRUCTURE results based on 14,563 GBS markers. (A) Sharp peak was observed at $K = 2$ with a maximum of ΔK determined by the Evanno method showing the stratification of the population into the minimum number of possible subgroups in the advanced wheat breeding lines. (B) Scatter plot of three principal components (PCs) of the SNP data with lines colored according to the two subgroups. (C) Structure plot for 141 advanced wheat breeding lines was stratified into two distinct clusters, where each color represents one subgroup representing (G1) and (G2).

Allelic pattern covering the populations

The mean Shannon's information index (I), the diversity index (h), and the unbiased diversity index (uh) were high (0.477, 0.313, and 0.320), which indicated its high diversity (Table 3, Fig. 4). The mean of different alleles (Na) and the effective alleles (Ne) of the $K = 2$ population set were 1.977 and 1.519. The different alleles (Na) in the subgroup 1 (G1) and subgroup 2 (G2) were 1.954 and 2.0, respectively, indicating comparatively better diversity in G2 than G1 (Table 3).

Similarly, the numbers of effective alleles (Ne) was relatively higher in G2 (1.525) than in G1 (1.512). However, only G2 showed a significant percentage of private alleles, i.e., 4.6%. G1 contained 22 members with private allele ranging from 7 in GID7399640 to 149 in GID7400491, while G2 contained 90 members with private alleles ranging from 3 in GID7396143 to 188 in GID7400293 (Table 3 & Table S4). The diversity indexes *I*, h and uh for G1 & G2 were 0.468 and 0.487, 0.307 and 0.318, 0.320 and 0.321, respectively, indicating relatively higher diversity within G2 than G1.
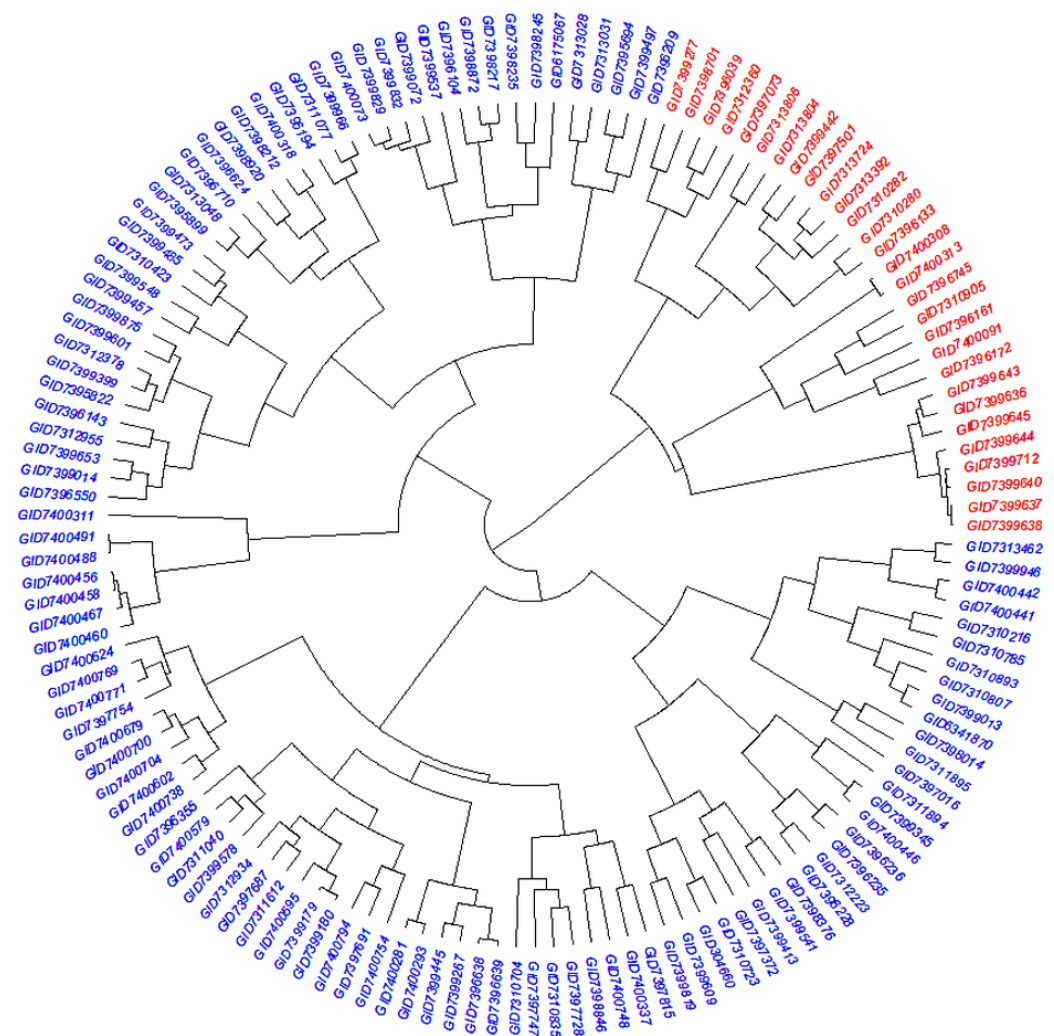
Figure 3 Dendrogram demonstrating the genetic relationships among 141 advanced bread wheat lines based on 14,563 GBS markers. Dendrogram showing the relationship among advanced breeding lines. Phylogenetic network constructed using complete linkage hieratical clustering based on 14,563 SNPs for all 141 advanced breeding lines. Breeding lines labels were color-coded into two subgroups.

Table 2 Analysis of Molecular Variance for *K* = 2 for within population and among-population variation.

K	Source	df	SS	MS	Est. Var.	%	Nm
2	Among Pops	1	3094.516	3094.516	14.292	1%	86.428
	Within Pops	139	343401.995	2470.518	2470.518	99%	
	Total	140	346496.511		2484.810	100%	

Notes.

#K = K value for sub-populations, Source = source of variation, df = degree of freedom, SS = sum of squares, MS = mean sum of squares, Est. Var.= estimated variation, % = percentage of variance explained, and Nm = haploid number of migrants.
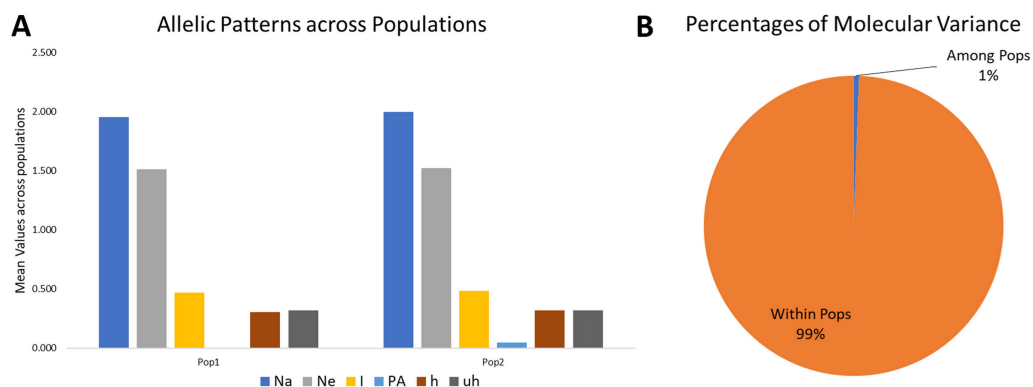
Figure 4 Proportion of Allelic patterns and diversity. (A) The allelic patterns and diversity indexes for $K = 2$, each subpopulation based on number of different alleles (Na), number of effective alleles (Ne), Shannon's information index(I), private alleles (PA), and diversity index (h) and (B) diversity among and within subgroups for $K = 2$.

Table 3 Allelic patterns and diversity indexes for $K = 2$, for subpopulations and mean diversity among subpopulations for each K.

Diversity indexes	Pop1	Pop2	Mean
Na	1.954 ($\pm$0.002)	2.000 ($\pm$0.000)	1.977
Na Freq. $\geq$ 5%	1.870 ($\pm$0.003)	1.973 ($\pm$0.001)	1.921
Ne	1.512 ($\pm$0.003)	1.525 ($\pm$0.002)	1.519
I	0.468 ($\pm$0.002)	0.487 ($\pm$0.001)	0.477
h	0.307 ($\pm$0.001)	0.318 ($\pm$0.001)	0.313
uh	0.320 ($\pm$0.001)	0.321 ($\pm$0.001)	0.320
PA	6.8E−5 ($\pm$0.000)	0.046 ($\pm$0.002)	0.023
Lines with PA	22	90	
Range of PA	3–149	3–188	

Notes.

\# Na = number of different alleles, Ne= number of effective alleles, I = Shannon's information index, PA = private alleles, h = diversity index, and uh = unbiased diversity index, $\pm$ represents standard error

DISCUSSION

A panel of 141 genotypes from an extensive collection of elite advanced bread wheat breeding lines from CIMMYT, Mexico, was used to study the genetic diversity (Table S1). The high-throughput SNP genotypic data obtained through GBS was used to explore population genetics and genetic diversity, supporting future breeding efforts (e.g., GWAS) in the bread wheat breeding program in South Asia.

The majority of the 14,563 SNPs were distributed on A and B genome, and only 8.67% SNPs were on D genome, which is at par with earlier findings (*Chao et al., 2009*; *Akhunov et al., 2010*; *Berkman et al., 2013*; *Würschum et al., 2013*; *Marcussen et al., 2014*; *Shavrukov et al., 2014*; *Edae, Bowden & Poland, 2015*; *Alipour et al., 2017*; *Eltaher et al., 2018*; *Rufo et al., 2019*). The lower SNPs across the D genome indicates its young wheat evolutionary past and less genetic diversity (*Caldwell et al., 2004*; *Alipour et al., 2017*; *Eltaher et al., 2018*),
which could be explained by lower recombination rates and frequency in the D genome (*Chao et al., 2009*). This could further be defined as a larger wild emmer diversity which contributed to hexaploid formation than *Ae. tauschii* (D-genome donor) (*Dubcovsky & Dvorak, 2007*). Significant initial gene movement must have occurred amongst *T. aestivum* and *T. turgidum* (AABB); however, not amongst *Ae. tauschii* (DD) and hexaploid (*Caldwell et al., 2004*; *Dvorak et al., 2006*). It leads to less genetic diversity in the D genome compared to the A and B genomes (*Talbert, Smith & Blake, 1998*; *Caldwell et al., 2004*; *Dvorak et al., 2006*; *Berkman et al., 2013*). The role of A, B, and D genomes to genetic diversity of hexaploid wheat were reported prior via diverse markers systems, i.e., RFLPs, AFLP, SSRs, DArT (*Liu & Tsunewaki, 1991*; *Röder et al., 1998*; *Peng et al., 2000*; *Nielsen et al., 2014*). The smallest number of SNP were on 4D, while the maximum markers were positioned on chromosome 2B, which agrees with *Allen et al. (2017)* and *Bhatta et al. (2017)*. Previous studies also reported the lowest number of SNPs at the 4D (*Saintenac et al., 2013*; *Sukumaran et al., 2014*; *Allen et al., 2017*; *Alipour et al., 2017*; *Bhatta et al., 2017*). While the highest SNP have been on a different chromosome in some of these studies, i.e., 3B (*Saintenac et al., 2013*; *Alipour et al., 2017*) and 1B (*Sukumaran et al., 2014*).

The low levels of observed heterozygosity (0.00−0.149), with approximately 97% of SNPs having heterozygosity <0.05 and only 21 SNPs having heterozygosity above 0.10, showed the panel had high genetic stability (*Kristensen et al., 2018*; *Rimbert et al., 2018*; *Chu et al., 2020*; *Sun et al., 2020*). Since these lines would/may not segregate further across generation leading to stable phenotypic evaluations. The lines with the low heterozygosity are highly desirables for selection as parental genotypes in any breeding program. Furthermore, the least heterozygous lines (GID7395694, GID7399636, GID7399643, GID7399645, and GID7400337) and highly heterozygous lines (GID7396143, GID7399653, and GID7400318) were observed among ABLs (Table S3). High heterozygosity among these three ABLs indicates either genomic instability or the higher outcrossing ability of these lines. MAF could easily arbitrate the allelic distribution of the SNPs. In this study, 55 SNPs were identified with MAF equal to 0.50. Traits showing distribution patterns similar to these SNPs could be easily associated with such SNPs.

PIC values are signals of informative markers in the crops and reflect through the spreading of informative markers in the genome, which can be used for studying genetic diversity (*Nielsen et al., 2014*; *Salem & Sallam, 2016*). The moderately informative PIC values indicate the SNPs' bi-allelic nature, which is limited to PIC 0.5, where both the alleles have similar occurrences (*Eltaher et al., 2018*). Another reason is the slow nucleotide mutation rate in GBS-SNPs compared to the mutation rate of SSRs (*Thuillet et al., 2002*; *Chesnokov & Artemyeva, 2015*). Most of the SNPs were moderately informative, with PIC values ranged from 0.085 to 0.250. The average PIC value among all sites was 0.201 (Fig. 1D). *Lopes et al. (2015)* used the 9K SNP for the WAMI population, detected a PIC value of 0.27 and disclosed that spring wheat confined moderate polymorphism. In another study, *Novoselović et al. (2016)* obtained an average moderate polymorphism of 0.30 PIC amongst the Croatian population by 1,229 Diversity Arrays Technology (DArT) markers. Furthermore, *El-Esawi et al. (2018)* also found moderate PIC (0.33 and 0.29) in Australian and Belgian wheat, respectively. Intrestingly, *Eltaher et al. (2018)* also detected a moderate

PIC value of 0.25 in 270 F3V6 Nebraska winter wheat. The present study outcome is following the above-mentioned previous studies.

In the present study, the STRUCTURE analysis was identified by ΔK with the highest peak at $K = 2$, which is vital for the elucidation of genetic diversity. *Winfield et al. (2016)* used 32,443 SNP markers and 804 wheat genotypes collected from over 30 countries. They detected that most European wheat accessions were grouped together, divided from the Asian and Middle Eastern accessions. *Cavanagh et al. (2013)* also reported that the winter wheat from the European population displayed robust genetic differentiation in their study. *Chen et al. (2019)* described that West Asian, European, numerous Central and South Asian landraces, and most East Asian cultivars grouped in the same cluster, whiles most of the East Asian landraces were grouped with South, Central and West Asian landraces. *Lee et al. (2018)* described that most Japanese, Korean and genotypes from Afghanistan were grouped in a cluster, while the Middle Eastern, Chinese, and Caucasus germplasm were in a separate group. The genetic diversity and population structure in the current ABLs were not surprising since the genetic composition, despite being variable, is restricted due to common ancestry (Table S1), leading to closely linked clusters. PCA and dendrogram results were in agreement with STRUCTURE results. They exhibited closely related groups, which might be because the selection of lines was based on traits in wheat, e.g., yield, biotic, and abiotic resistance arising from the parental pedigree of genotypes. Besides genetic diversity analysis, genetic structure analysis for subgroups composition is also an essential part of genome-wide association studies (GWAS) to counter false positives arising due to common ancestry among the panel of genotypes (*Yao et al., 2019*). Hence, population interchange and exploitation of global germplasm have become an essential preliminary step to increase the genetic source for wheat breeding (*Zhao et al., 2019*).

Private alleles provide important information identifying distinctive genetic variability at loci and diversified genotypes, which could be employed in crop breeding to enhance the allele affluence in a population (*Borba et al., 2009*; *Salem & Sallam, 2016*). For $K = 2$, G1 and G2 contained 22 and 90 members with private alleles, respectively, illustrating a clear difference in the lines containing private alleles (Table S4). These results indicate G2 being genetically diverse compared with subgroup G1, further supported by slightly higher values of diversity indexes in G2 (0.318) than G1(0.307). Similar results have been previously observed in studies on wheat genotypes using SNP markers where higher values of *I, u* and *uh* in a subgroup are indicative of a higher diversity of the group (*Alipour et al., 2017*; *Eltaher et al., 2018*; *Kumar et al., 2020*; *Mourad, Belamkar & Baenziger, 2020*). A system should be designed to identify private alleles equipped for receiving and harnessing the essential adaptive genes.

The AMOVA results showed high genetic diversity within-subgroups; however, the diversity between subgroups was very low (1%). The result may be due to the common parental backgrounds and selection based on designated agronomic traits resulted in high gene flow levels. This low level of variation among the stratified groups occurs due to increased gene exchange described by *Arora et al. (2014)*. The allelic outlines elucidated valuable evidence on genetic diversity in each subgroup. *Wright (1965)* also described restricted Nm (haploid) gene flow between populations. In the current study, a very

high Nm value (86.428), suggesting a high level of genetic exchange/flow among the subgroups, caused small genetic variation (*Eltaher et al., 2018*). Hence, the high genetic interchange amid subgroups directed to a small genetic variation amongst subgroups. The variation amongst subgroups was noteworthy ($P <0.001$) regardless of being low (1%). The present study results will aid breeders to understand the genetic diversity and perform marker-assisted selection on this panel.

CONCLUSION

In the present study, we applied GBS-based SNP to learn GBS-SNP markers' usefulness for diversity analysis in 141 elite wheat breeding lines. Despite very designated, our advanced breeding lines panel was found to be genetically diverse, which could be instrumental for future South Asian breeding programs to develop new elite wheat varieties of alluring traits, i.e., high yield, biotic and abiotic resistance. Besides, the present study dappled two subgroups that were enlightened by their parentage and selection history. The low heterozygosity detected among elite advanced wheat breeding lines within subgroups and the moderate divergence among subgroups suggested that the elite advanced wheat breeding lines could be used further for GWAS studies.

ACKNOWLEDGEMENTS

The technical support of Dr. Sandesh Shrestha is duly acknowledged.

Funding

The present study is supported by US Agency for International Development (USAID) through "Feed the Future Innovation Lab for Applied Wheat Genomics" (Cooperative Agreement No. AID-OAA-A-13-00051). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
US Agency for International Development (USAID): AID-OAA-A-13-00051.

Competing Interests

Daljit Singh was a postdoc in the Department of Plant Pathology, Kansas State University, Manhattan, KS, USA, at the start of this project. He then joined The Climate Corporation, Bayer Crop Science, Creve Coeur, MO, USA. The Climate Corporation has no role in the project.

Author Contributions

- Vipin Tomar conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

- Guriqbal Singh Dhillon analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Daljit Singh analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Ravi Prakash Singh, Arun Kumar Joshi, Budhi Sagar Tiwari and Uttam Kumar conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Jesse Poland conceived and designed the experiments, performed the experiments, authored or reviewed drafts of the paper, performed the GBS, and approved the final draft.

Data Availability

The GBS SNP Data is available in the Supplemental File. The SNP dataset used in the present study is also available at: Tomar, Vipin (2021): WheatGBS.hmp.txt. figshare. Dataset. https://doi.org/10.6084/m9.figshare.14273339.v2.

SNPs were aligned and assigned to an exact physical location by referencing the Chinese Spring Wheat Assembly v1.0 (International Wheat Genome Sequencing Consortium [IWGSC]), and used for further downstream analysis.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.11593#supplemental-information.

REFERENCES

Akhunov ED, Akhunova AR, Anderson OD, Anderson JA, Blake N, Clegg MT, Coleman-Derr D, Conley EJ, Crossman CC, Deal KR, Dubcovsky J, Gill BS, Gu YQ, Hadam J, Heo H, Huo N, Lazo GR, Luo MC, Ma YQ, Matthews DE, McGuire PE, Morrell PL, Qualset CO, Renfro J, Tabanao D, Talbert LE, Tian C, Toleno DM, Warburton ML, You FM, Zhang W, Dvorak J. 2010. Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. *BMC Genomics* **11**:702 DOI 10.1186/1471-2164-11-702.

Akter N, Rafiqul Islam M. 2017. Heat stress effects and management in wheat. A review. *Agronomy for Sustainable Development* 37 DOI 10.1007/s13593-017-0443-9.

Alipour H, Bihamta MR, Mohammadi V, Peyghambari SA, Bai G, Zhang G. 2017. Genotyping-by-sequencing (GBS) revealed molecular genetic diversity of Iranian wheat landraces and cultivars. *Frontiers in Plant Science* **3**:1293 DOI 10.3389/fpls.2017.01293.

Allen AM, Winfield MO, Burridge AJ, Downie RC, Benbow HR, Barker GLA, Wilkinson PA, Coghill J, Waterfall C, Davassi A, Scopes G, Pirani A, Webster T, Brew F, Bloor C, Griffiths S, Bentley AR, Alda M, Jack P, Phillips AL, Edwards KJ. 2017. Characterization of a Wheat Breeders' Array suitable for high-throughput SNP

genotyping of global accessions of hexaploid bread wheat (*Triticum aestivum*). *Plant Biotechnology Journal* **15**:390–401 DOI 10.1111/pbi.12635.

Arora A, Kundu S, Dilbaghi N, Sharma I, Tiwari R. 2014. Population structure and genetic diversity among Indian wheat varieties using microsatellite (SSR) markers. *Australian Journal of Crop Science* **8(9)**:1281–1289.

Balfourier F, Bouchet S, Robert S, De Oliveira R, Rimbert H, Kitt J, Choulet F, Paux E. 2019. Worldwide phylogeography and history of wheat genetic diversity. *Science Advances* **5**:eaav0536 DOI 10.1126/sciadv.aav0536.

Berkman PJ, Visendi P, Lee HC, Stiller J, Manoli S, Lorenc MT, Lai K, Batley J, Fleury D, Šimková H, Kubaláková M, Weining S, Doležel J, Edwards D. 2013. Dispersion and domestication shaped the genome of bread wheat. *Plant Biotechnology Journal* **11**:564–571 DOI 10.1111/pbi.12044.

Bhatta M, Regassa T, Rose DJ, Baenziger PS, Eskridge KM, Santra DK, Poudel R. 2017. Genotype, environment, seeding rate, and top-dressed nitrogen effects on end-use quality of modern nebraska winter wheat. *Journal of the Science of Food and Agriculture* **97**:5311–5318 DOI 10.1002/jsfa.8417.

Botstein D, White RL, Skolnick M, Davis RW. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **32**:314–331.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**:941–947 DOI 10.1093/bioinformatics/btm308.

Caldwell KS, Dvorak J, Lagudah ES, Akhunov E, Luo MC, Wolters P, Powell W. 2004. Sequence polymorphism in polyploid wheat and their D-genome diploid ancestor. *Genetics* **167(2)**:941–947 DOI 10.1534/genetics.103.016303.

Cavanagh CR, Chao S, Wang S, Huang BE, Stephen S, Kiani S, Forrest K, Saintenac C, Brown-Guedira GL, Akhunova A, See D, Bai G, Pumphrey M, Tomar L, Wong D, Kong S, Reynolds M, Da Silva ML, Bockelman H, Talbert L, Anderson JA, Dreisigacker S, Baenziger S, Carter A, Korzun V, Morrell PL, Dubcovsky J, Morell MK, Sorrells ME, Hayden MJ, Akhunov E. 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 of the United States of America* **110**:8057–8062 DOI 10.1073/pnas.1217133110.

Chao S, Zhang W, Akhunov E, Sherman J, Ma Y, Luo MC, Dubcovsky J. 2009. Analysis of gene-derived SNP marker polymorphism in US wheat (*Triticum aestivum* L.) cultivars. *Molecular Breeding* **23**:23–33 DOI 10.1007/s11032-008-9210-6.

Chen G, Zhang H, Deng Z, Wu R, Li D, Wang M, Tian J. 2016. Genome-wide association study for kernel weight-related traits using SNPs in a Chinese winter wheat population. *Euphytica* **212**:173–185 DOI 10.1007/s10681-016-1750-y.

Chen H, Jiao C, Wang Y, Wang Y, Tian C, Yu H, Wang J, Wang X, Lu F, Fu X, Xue Y, Jiang W, Ling H, Lu H, Jiao Y. 2019. Comparative population genomics of bread wheat (*Triticum aestivum*) reveals its cultivation and breeding history in China. *bioRxiv* DOI 10.1101/519587.
Chesnokov YV, Artemyeva AM. 2015. Evaluation of the measure of polymorphism information of genetic diversity. *Sel'Skokhozyaistvennaya Biologiya* **50**:571–578 DOI 10.15389/agrobiology.2015.5.571eng.

Chu J, Zhao Y, Beier S, Schulthess AW, Stein N, Philipp N, Röder MS, Reif JC. 2020. Suitability of single-nucleotide polymorphism arrays versus genotyping-by-sequencing for genebank genomics in wheat. *Frontiers in Plant Science* **11(42)**:1–12 DOI 10.3389/fpls.2020.00042.

De O Borba TC, Dos A Mendes C, Guimarães ÉP, Brunes TO, Fonseca JR, Brondani RV, Brondani C. 2009. Genetic variability of Brazilian rice landraces determined by SSR markers. *Pesquisa AgropecuáRia Brasileira* **44(7)**:706–712 DOI 10.1590/s0100-204x2009000700009.

Dreisigacker S, Deepmala S, Jaimez RA, Luna BG, Muñoz SZ, Ríos CNC, M J, M S. 2016. *CIMMYT wheat molecular genetics: laboratory protocols and applications to wheat breeding.* Mexico, D.F: CIMMYT.

Dubcovsky J, Dvorak J. 2007. Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science* **316**:1862–1866 DOI 10.1126/science.1143986.

Dvorak J, Akhunov ED, Akhunov AR, Deal KR, Luo MC. 2006. Molecular character-ization of a diagnostic DNA marker for domesticated tetraploid wheat provides evidence for gene flow from wild tetraploid wheat to hexaploid wheat. *Molecular Biology and Evolution* **23**:1386–1396 DOI 10.1093/molbev/msl004.

Edae EA, Bowden RL, Poland J. 2015. Application of population sequencing (POPSEQ) for ordering and imputing genotyping-by-sequencing markers in hexaploid wheat. *G3: Genes, Genomes, Genetics* **5**:2547–2553 DOI 10.1534/g3.115.020362.

El-Esawi MA, Witczak J, Abomohra AEF, Ali HM, Elshikh MS, Ahmad M. 2018. Analysis of the genetic diversity and population structure of Austrian and Belgian wheat germplasm within a regional context based on DArT markers. *Gene* **9**:47 DOI 10.3390/genes9010047.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLOS ONE* **6**:e19379 DOI 10.1371/journal.pone.0019379.

Eltaher S, Sallam A, Belamkar V, Emara HA, Nower AA, Salem KFM, Poland J, Baenziger PS. 2018. Genetic diversity and population structure of F3:6 Nebraska Winter wheat genotypes using genotyping-by-sequencing. *Frontiers in Genetics* **9**:76 DOI 10.3389/fgene.2018.00076.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* **14**:2611–2620 DOI 10.1111/j.1365-294X.2005.02553.x.

Farahani S, Maleki M, Mehrabi R, Kanouni H, Scheben A, Batley J, Talebi R. 2019. Whole genome diversity, population structure, and linkage disequilibrium analysis of chickpea (*Cicer arietinum* l.) genotypes using genome-wide dartseq-based snp markers. *Gene* **10(9)**:676 DOI 10.3390/genes10090676.

Francis RM. 2017. pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources* **17**:27–32 DOI 10.1111/1755-0998.12509.

Fu YB, Cheng B, Peterson GW. 2014. Genetic diversity analysis of yellow mustard (*Sinapis alba* L.) germplasm based on genotyping by sequencing. *Genetic Resources and Crop Evolution* **61**:579–594 DOI 10.1007/s10722-013-0058-1.

Fu Y-B, Peterson GW. 2011. Genetic diversity analysis with 454 pyrosequencing and genomic reduction confirmed the eastern and western division in the cultivated barley gene pool. *The Plant Genome* DOI 10.3835/plantgenome2011.08.0022.

Ganal MW, Altmann T, Röder MS. 2009. SNP identification in crop plants. *Current Opinion in Plant Biology* **12**:211–217 DOI 10.1016/j.pbi.2008.12.009.

Heffner EL, Lorenz AJ, Jannink JL, Sorrells ME. 2010. Plant breeding with Genomic selection: gain per unit time and cost. *Crop Science* **50**:1681–1690 DOI 10.2135/cropsci2009.11.0662.

Heffner EL, Sorrells ME, Jannink JL. 2009. Genomic selection for crop improvement. *Crop Science* **49**:1–12 DOI 10.2135/cropsci2008.08.0512.

Jannink JL, Lorenz AJ, Iwata H. 2010. Genomic selection in plant breeding: from theory to practice. *Briefings in Functional Genomics and Proteomics* **9**:166–177 DOI 10.1093/bfgp/elq001.

Knowler WC, Williams RC, Pettitt DJ, Steinberg AG. 1988. Gm(3;5,13,14) and type 2 diabetes mellitus: an association in American Indians with genetic admixture. *American Journal of Human Genetics* **43(4)**:520–526.

Kristensen PS, Jahoor A, Andersen JR, Cericola F, Orabi J, Janss LL, Jensen J. 2018. Genome-wide association studies and comparison of models and cross-validation strategies for genomic prediction of quality traits in advanced winter wheat breeding lines. *Frontiers in Plant Science* **9**:69 DOI 10.3389/fpls.2018.00069.

Kumar D, Chhokar V, Sheoran S, Singh R, Sharma P, Jaiswal S, Iquebal MA, Jaiswar A, Jaisri J, Angadi UB, Rai A, Singh GP, Kumar D, Tiwari R. 2020. Characterization of genetic diversity and population structure in wheat using array based SNP markers. *Molecular Biology Reports* **47**:293–306 DOI 10.1007/s11033-019-05132-8.

Lee S, Choi YM, Lee MC, Hyun DY, Oh S, Jung Y. 2018. 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* **65**:1591–1602 DOI 10.1007/s10722-018-0633-6.

Liu YG, Tsunewaki K. 1991. Restriction fragment length polymorphism (rflp) analysis in wheat. II. linkage maps of the rflp sites in common wheat. *The Japanese Journal of Genetics* **46**:617–633 DOI 10.1266/jjg.66.617.

Lopes MS, Dreisigacker S, Peña RJ, Sukumaran S, Reynolds MP. 2015. Genetic characterization of the wheat association mapping initiative (WAMI) panel for dissection of complex traits in spring wheat. *Theoretical and Applied Genetics* **128**:453–464 DOI 10.1007/s00122-014-2444-2.

Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD, Buckler ES, Costich DE. 2013. Switchgrass genomic diversity, ploidy, and evolution: novel insights from a

network-based SNP discovery protocol. *PLOS Genetics* **9**:e1003215 DOI 10.1371/journal.pgen.1003215.

Maccaferri M, Ricci A, Salvi S, Milner SG, Noli E, Martelli PL, Casadio R, Akhunov E, Scalabrin S, Vendramin V, Ammar K, Blanco A, Desiderio F, Distelfeld A, Dubcovsky J, Fahima T, Faris J, Korol A, Massi A, Mastrangelo AM, Morgante M, Pozniak C, N'Diaye A, Xu S, Tuberosa R. 2015. A high-density, SNP-based consensus map of tetraploid wheat as a bridge to integrate durum and bread wheat genomics and breeding. *Plant Biotechnology Journal* **33**:648–663 DOI 10.1111/pbi.12288.

Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer M, Jakobsen KS, Wulff BBH, Steuernagel B, Mayer KFX, Olsen OA, Rogers J, Doležel J, Pozniak C, Eversole K, Feuillet C, Gill B, Friebe B, Lukaszewski AJ, Sourdille P, Endo TR, Kubaláková M, Šíhalíková J, Dubská Z, Vrána J, Šperková R, Šimková H, Febrer M, Clissold L, McLay K, Singh K, Chhuneja P, Singh NK, Khurana J, Akhunov E, Choulet F, Alberti A, Barbe V, Wincker P, Kanamori H, Kobayashi F, Itoh T, Matsumoto T, Sakai H, Tanaka T, Wu J, Ogihara Y, Handa H, Maclachlan PR, Sharpe A, Klassen D, Edwards D, Batley J, Sandve SR, Lien S, Wulff B, Caccamo M, Ayling S, Ramirez-Gonzalez RH, Clavijo BJ, Wright J, Martis MM, Mascher M, Chapman J, Poland JA, Scholz U, Barry K, Waugh R, Rokhsar DS, Muehlbauer GJ, Stein N, Gundlach H, Zytnicki M, Jamilloux V, Quesneville H, Wicker T, Faccioli P, Colaiacovo M, Stanca AM, Budak H, Cattivelli L, Glover N, Pingault L, Paux E, Sharma S, Appels R, Bellgard M, Chapman B, Nussbaumer T, Bader KC, Rimbert H, Wang S, Knox R, Kilian A, Alaux M, Alfama F, Couderc L, Guilhot N, Viseux C, Loaec M, Keller B, Praud S. 2014. Ancient hybridizations among the ancestral genomes of bread wheat. *Science* **345**:1250092 DOI 10.1126/science.1250092.

Mengistu DK, Kidane YG, Catellani M, Frascaroli E, Fadda C, Pè ME, Dell'Acqua M. 2016. High-density molecular characterization and association mapping in Ethiopian durum wheat landraces reveals high diversity and potential for wheat breeding. *Plant Biotechnology Journal* **14**:1800–1812 DOI 10.1111/pbi.12538.

Mengistu DK, Kiros AY, Pè ME. 2015. Phenotypic diversity in Ethiopian durum wheat (*Triticum turgidum* var. durum) landraces. *Crop Journal* **7**:190–199 DOI 10.1016/j.cj.2015.04.003.

Mondal S, Rutkoski JE, Velu G, Singh PK, Crespo-Herrera LA, Guzman C, Bhavani S, Lan C, He X, Singh RP. 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 DOI 10.3389/fpls.2016.00991.

Mourad AMI, Belamkar V, Baenziger PS. 2020. Molecular genetic analysis of spring wheat core collection using genetic diversity, population structure, and linkage disequilibrium. *BMC Genomics* **21**:434 DOI 10.1186/s12864-020-06835-0.

Nielsen NH, Backes G, Stougaard J, Andersen SU, Jahoor A. 2014. Genetic diversity and population structure analysis of European hexaploid bread wheat (*Triticum aestivum* L.) varieties. *PLOS ONE* **9**:e94000 DOI 10.1371/journal.pone.0094000.

Novoselović D, Bentley AR, Šimek R, Dvojković K, Sorrells ME, Gosman N, Horsnell R, Drezner G, Šatović Z. 2016. Characterizing croatian wheat germplasm diversity and structure in a European context by DArT markers. *Frontiers in Plant Science* **7**:184 DOI 10.3389/fpls.2016.00184.

Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**:526–528 DOI 10.1093/bioinformatics/bty633.

Peakall R, Smouse PE. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**:288–295 DOI 10.1111/j.1471-8286.2005.01155.x.

Peng J, Korol AB, Fahima T, Röder MS, Ronin YI, Li YC, Nevo E. 2000. Molecular genetic maps in wild emmer wheat, Triticum dicoccoides: genoma-wide coverage, massive negative interference, and putative Quasi-linkage. *Genome Research* **10**:1509–1531 DOI 10.1101/gr.150300.

Peterson GW, Dong Y, Horbach C, Fu YB. 2014. Genotyping-by-sequencing for plant genetic diversity analysis: a lab guide for SNP genotyping. *Diversity* **6(4)**:665–680 DOI 10.3390/d6040665.

Poland JA, Brown PJ, Sorrells ME, Jannink JL. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLOS ONE* **7(2)**:e32253 DOI 10.1371/journal.pone.0032253.

Poland JA, Rife TW. 2012. Genotyping-by-sequencing for plant breeding and genetics. *The Plant Genome* **5(3)** DOI 10.3835/plantgenome2012.05.0005.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155(2)**:945–959.

Rimbert H, Darrier B, Navarro J, Kitt J, Choulet F, Leveugle M, Duarte J, Rivière N, Eversole K, Consortium on behalf of TIWGS, Le Gouis J, Consortium on behalf TB, Davassi A, Balfourier F, Le Paslier M-C, Berard A, Brunel D, Feuillet C, Poncet C, Sourdille P, Paux E. 2018. High throughput SNP discovery and genotyping in hexaploid wheat. *PLOS ONE* **13**:e0186329 DOI 10.1371/journal.pone.0186329.

Röder MS, Korzun V, Gill BS, Ganal MW. 1998. The physical mapping of microsatellite markers in wheat. *Genome* **41(2)**:278–283 DOI 10.1139/g98-009.

Rufo R, Alvaro F, Royo C, Soriano JM. 2019. From landraces to improved cultivars: assessment of genetic diversity and population structure of Mediterranean wheat using SNP markers. *PLOS ONE* **14**:e0219867 DOI 10.1371/journal.pone.0219867.

Saintenac C, Jiang D, Wang S, Akhunov E. 2013. Sequence-based mapping of the polyploid wheat genome. *G3: Genes, Genomes, Genetics* **3**:1105–1114 DOI 10.1534/g3.113.005819.

Salem KFM, Sallam A. 2016. Analysis of population structure and genetic diversity of Egyptian and exotic rice (*Oryza sativa* L.) genotypes. *Comptes Rendus - Biologies* **339**:1–9 DOI 10.1016/j.crvi.2015.11.003.

Sansaloni C, Franco J, Santos B, Percival-Alwyn L, Singh S, Petroli C, Campos J, Dreher K, Payne T, Marshall D, Kilian B, Milne I, Raubach S, Shaw P, Stephen G, Carling J, Pierre CS, Burgueño J, Crosa J, Li HH, Guzman C, Kehel Z, Amri A, Kilian

A, Wenzl P, Uauy C, Banziger M, Caccamo M, Pixley K. 2020. Diversity analysis of 80,000 wheat accessions reveals consequences and opportunities of selection footprints. *Nature Communications* **11**:4572 DOI 10.1038/s41467-020-18404.

Shavrukov Y, Suchecki R, Eliby S, Abugalieva A, Kenebayev S, Langridge P. 2014. Application of next-generation sequencing technology to study genetic diversity and identify unique SNP markers in bread wheat from Kazakhstan. *BMC Plant Biology* **14**:258 DOI 10.1186/s12870-014-0258-7.

Sievert C. 2020. Interactive Web-Based Data Visualization with R, plotly, and shiny. Boca Raton: Chapman and Hall/CRC DOI 10.1201/9780429447273.

Singh N, Agarwal N, Yadav HK. 2019. Genome-wide SNP-based diversity analysis and association mapping in linseed (*Linum usitatissimum* L.). *Euphytica* **215**:139 DOI 10.1007/s10681-019-2462-x.

Sukumaran S, Dreisigacker S, Lopes M, Chavez P, Reynolds MP. 2014. Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments. *Theoretical and Applied Genetics* **128**:353–363 DOI 10.1007/s00122-014-2435-3.

Sun C, Dong Z, Zhao L, Ren Y, Zhang N, Chen F. 2020. The Wheat 660K SNP array demonstrates great potential for marker-assisted selection in polyploid wheat. *Plant Biotechnology Journal* **18**:1354–1360 DOI 10.1111/pbi.13361.

Talbert LE, Smith LY, Blake NK. 1998. More than one origin of hexaploid wheat is indicated by sequence comparison of low-copy DNA. *Genome* **41(3)**:402–407 DOI 10.1139/g98-037.

Thomson MJ, Zhao K, Wright M, McNally KL, Rey J, Tung CW, Reynolds A, Scheffler B, Eizenga G, McClung A, Kim H, Ismail AM, de Ocampo M, Mojica C, Reveche MY, Dilla-Ermita CJ, Mauleon R, Leung H, Bustamante C, McCouch SR. 2012. High-throughput single nucleotide polymorphism genotyping for breeding applications in rice using the BeadXpress platform. *Molecular Breeding* **19**:875–886 DOI 10.1007/s11032-011-9663-x.

Thuillet AC, Bru D, David J, Roumet P, Santoni S, Sourdille P, Bataillon T. 2002. Direct estimation of mutation rate for 10 microsatellite loci in durum wheat, Triticum turgidum (L.) thell. ssp durum desf. [2]. *Molecular Biology and Evolution* **19**:122–125 DOI 10.1093/oxfordjournals.molbev.a003977.

Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, Maccaferri M, Salvi S, Milner SG, Cattivelli L, Mastrangelo AM, Whan A, Stephen S, Barker G, Wieseke R, Plieske J, Lillemo M, Mather D, Appels R, Dolferus R, Brown-Guedira G, Korol A, Akhunova AR, Feuillet C, Salse J, Morgante M, Pozniak C, Luo MC, Dvorak J, Morell M, Dubcovsky J, Ganal M, Tuberosa R, Lawley C, Mikoulitch I, Cavanagh C, Edwards KJ, Hayden M, Akhunov E. 2014. Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnology Journal* **12**:787–796 DOI 10.1111/pbi.12183.

Wen W, He Z, Gao F, Liu J, Jin H, Zhai S, Qu Y, Xia X. 2017. A high-density consensus map of common wheat integrating four mapping populations scanned by the 90k SNP array. *Frontiers in Plant Science* **8**:1389 DOI 10.3389/fpls.2017.01389.

Winfield MO, Allen AM, Burridge AJ, Barker GLA, Benbow HR, Wilkinson PA, Coghill J, Waterfall C, Davassi A, Scopes G, Pirani A, Webster T, Brew F, Bloor C, King J, West C, Griffiths S, King I, Bentley AR, Edwards KJ. 2016. High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnology Journal* **14**:1195–1206 DOI 10.1111/pbi.12485.

Wright S. 1965. The interpretation of population structure by f-statistics with special regard to systems of mating. *Evolution* **19**:395–420 DOI 10.1111/j.1558-5646.1965.tb01731.x.

Wulft BBH, Moscou MJ. 2014. Strategies for transferring resistance into wheat: from wide crosses to GM cassettes. *Frontiers in Plant Science* **5**:692 DOI 10.3389/fpls.2014.00692.

Würschum T, Langer SM, Longin CFH, Korzun V, Akhunov E, Ebmeyer E, Schachschneider R, Schacht J, Kazman E, Reif JC. 2013. Population structure, genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers. *Theoretical and Applied Genetics* **126**:1477–1486 DOI 10.1007/s00122-013-2065-1.

Yao F, Zhang X, Ye X, Li J, Long L, Yu C, Li J, Wang Y, Wu Y, Wang J, Jiang Q, Li W, Ma J, Wei Y, Zheng Y, Chen G. 2019. Characterization of molecular diversity and genome-wide association study of stripe rust resistance at the adult plant stage in Northern Chinese wheat landraces. *BMC Genetics* **20**:38 DOI 10.1186/s12863-019-0736-x.

Zhang Q, Maroof MAS, Kleinhofs A. 1993. Comparative diversity analysis of RFLPs and isozymes within and among populations of *Hordeum vulgare* ssp. spontaneum. *Genetics* **134**(3):909–916.

Zhao J, Wang Z, Liu H, Zhao J, Li T, Hou J, Zhang X, Hao C. 2019. Global status of 47 major wheat loci controlling yield, quality, adaptation and stress resistance selected over the last century. *BMC Plant Biology* **19**:5 DOI 10.1186/s12870-018-1612-y.

Zhu C, Gore M, Buckler ES, Yu J. 2008. Status and prospects of association mapping in plants. *The Plant Genome* **1**(1):5–20 DOI 10.3835/plantgenome2008.02.0089.