



Original Article

Exploring genetic diversity of wild and related tetraploid wheat species *Triticum turgidum* and *Triticum timopheevii*

Inderjit S. Yadav^a, Narinder Singh^b, Shuangye Wu^c, Jon Raupp^c, Duane L. Wilson^c, Nidhi Rawat^a, Bikram S. Gill^c, Jesse Poland^{d,*}, Vijay K. Tiwari^{a,*}

^a Department of Plant Sciences and Landscape Architecture, University of Maryland, College Park, MD 20742, USA

^b Bayer R&D Services LLC, Kansas City, MO 64153, USA

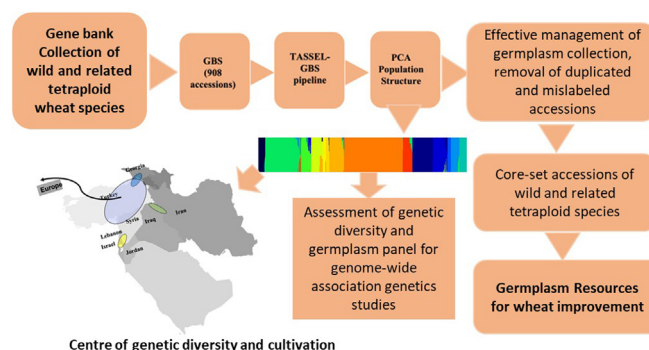
^c Department of Plant Pathology and Wheat Genetics Resource Center, Kansas State University, 1712 Claflin Road, Manhattan, KS 66506, USA

^d Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, 4700 KAUST, Thuwal 23955-6900, Saudi Arabia

HIGHLIGHTS

- Systematic evaluation of genetic diversity exists in wild and related tetraploid wheat species *T. turgidum* and *T. timopheevii*.
- Using genotypic datasets, we found duplicated accessions in *T. timopheevii* (~65 %) and in *T. turgidum* (47 %).
- We combined genotypic and phenotypic datasets to identify core set accessions to introduce novel genetic diversity in wheat breeding programs.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 20 June 2022

Revised 25 August 2022

Accepted 31 August 2022

Available online 7 September 2022

Keywords:

Tetraploid wheat

Nei's index

F-statistics: (F_{ST})

Core collection

Triticum

Population structure

Genetic diversity

Genotyping by sequencing (GBS)

Sequencing

Phenotype

Genome wide association study (GWAS)

ABSTRACT

Introduction: The domestication bottleneck has reduced genetic diversity in wheat, necessitating the use of wild relatives in breeding programs. Wild tetraploid wheat are widely used in the breeding programs but with morphological characters, it is difficult to distinguish these, resulting in misclassification/mislabeling or duplication of accessions in the Gene bank.

Objectives: The study aims to explore Genotyping by sequencing (GBS) to characterize wild and domesticated tetraploid wheat accessions to generate a core set of accessions to be used in the breeding program.

Methods: TASSEL-GBS pipeline was used for SNP discovery, fastStructure was used to determine the population structure and PowerCore was used to generate a core sets. Nucleotide diversity matrices of Nei's and F-statistics (F_{ST}) index were used to determine the center of genetic diversity.

Results: We found 65 % and 47 % duplicated accessions in *Triticum timopheevii* and *T. turgidum* respectively. Genome-wide nucleotide diversity and F_{ST} scan uncovered a lower intra and higher inter-species differentiation. Distinct F_{ST} regions were identified in genomic regions belonging to domestication genes: non-brittle rachis (*Btr1*) and vernalization (*VRN-1*). Our results suggest that Israel, Jordan, Syria, and Lebanon as the hub of genetic diversity of wild emmer; Turkey, and Georgia for *T. durum*; and Iraq, Azerbaijan, and Armenia for the *T. timopheevii*. Identified core set accessions preserved more than 93 % of the available genetic diversity. Genome wide association study (GWAS) indicated the potential chromosomal segment for resistance to leaf rust in *T. timopheevii*.

Peer review under responsibility of Cairo University.

* Corresponding authors.

E-mail address: vktiwiari@umd.edu (V.K. Tiwari).

<https://doi.org/10.1016/j.jare.2022.08.020>

2090-1232/© 2023 The Authors. Published by Elsevier B.V. on behalf of Cairo University.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Conclusion: The present study explored the potential of GBS technology in data reduction while maintaining the significant genetic diversity of the species. Wild germplasm showed more differentiation than domesticated accessions, indicating the availability of sufficient diversity for crop improvement. With reduced complexity, the core set preserves the genetic diversity of the gene bank collections and will aid in a more robust characterization of wild germplasm.

© 2023 The Authors. Published by Elsevier B.V. on behalf of Cairo University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Wheat is one of the most important crop plants for the human population. Two major wheat species grown throughout the world are *Triticum aestivum*, a hexaploid species usually called “common” or “bread” wheat, and *T. turgidum* ssp. *durum*, a tetraploid species adapted to hot and dry conditions surrounding the Mediterranean Sea and similar climates in other regions. Tetraploid wheat (*T. turgidum* L.) is an important species within the genus *Triticum*, and its numerous subspecies including wild forms harbor many desirable agronomic traits for durum as well bread wheat improvement. Hexaploid bread wheat has evolved through two distinct past hybridization events, which included three diploid progenitor species. The wheat genome comprises three sub-genomes namely A, B and D ($2n = 6x = 42$). A genome of wheat was contributed by *T. urartu* (AA), while *Aegilops speltoides* (SS) is thought to be closest match for B- genome donor. The D-genome was contributed by *Aegilops tauschii* (DD). *T. urartu* contributed A genomes to the tetraploid wheat species *T. turgidum* (BBAA; $2n = 4x = 28$) and *T. timopheevii* (GGAA; $2n = 4x = 28$) [1,2], and *Ae. speltoides* (SS) as maternal parent contributed to B and G genomes [3]. These two tetraploid wheat species (*T. turgidum* and *T. timopheevii*) are reproductively isolated (crosses between these are sterile and show reduced chromosome pairing) and belong to distinct evolutionary lineages [4].

Evidence for the first cultivation of wild emmer dates back to ~10,300 to 9,500 years ago in the southern Levant region, while the cultivation of domesticated emmer began ~9500–9000 years ago and was grown as a mixture along with the wild emmer for a longer period [5]. Over the years the durum wheat i.e. *T. turgidum* ssp. *durum* had remained an important economic crop majorly used for making pasta, leading to the whole-genome sequencing of a modern-day cultivar [6].

Wild *T. turgidum* represented by *T. dicoccoides* is the only true wild polyploid wheat of the lineage, found in dry and saline regions of the middle east [7]. The genetic diversity present in *T. dicoccoides* is a rich source for agronomic traits related to grain quality and abiotic/biotic stress [8]. Domesticated lineages of *T. dicoccoides* consist of subspecies *T. turgidum* spp. *dicoccon*, *T. turgidum* spp. *dicoccon*, *T. turgidum* spp. *durum*, *T. turgidum* spp. *turgidum*, *T. turgidum* spp. *polonicum*, *T. turgidum* spp. *carthlicum*, *T. turgidum* spp. *uranicum* and *T. turgidum* spp. *palaeocolchicum* [9]. *T. timopheevii* belongs to the secondary gene pool of wheat and its domesticated form is grown as a minor crop in Georgia. *T. timopheevii* is an important resource for the improvement of durum and bread wheat [10]. *T. timopheevii* consists of two sub-species, wild i.e. *T. timopheevii* spp. *armeniicum* and cultivated *T. timopheevii* spp. *timopheevii* [9]. The two lineages show a spread in the Transcaucasian region at the border of Europe and Asia [11].

Maestra and coworkers (1999) studied chromosomal structure in wild and cultivated populations of *T. timopheevii*, durum, and hexaploid wheat, and reported similar homologous pairing between A-genome and lesser association between chromosomes of the B-G genome. Structural similarities were observed for the A^t-genome chromosomes (1, 2, 5, and 7) and G-genome chromo-

somes (2,3,5, and 6) with their respective A and B genome counterparts in durum and hexaploid wheat [12]. The higher similarity between the A-genome of the two species was attributed to their common ancestor *T. urartu* [1]. The B-genome of *T. turgidum* showed higher divergence from the S-genome of *Ae. speltoides* compared to the G-genome of *T. timopheevii* [9]. *T. turgidum* and *T. timopheevii* share similar morphological characters and are generally mislabeled or misclassified. Molecular markers are routinely applied for whole-genome characterization. Through the use of barcoded multiplexing, low-cost high throughput techniques like GBS can produce a significant number of SNP markers for large populations. The method involves a set of restriction enzymes to digest the genome into fragments, which are then sequenced using Illumina single or paired-end sequencing [13]. SNP calls can be made from the sequencing data using bioinformatics pipelines i.e., UNEAK, TASSEL-GBS, and Fast-GBS [14–16]. Czajkowska and coworkers have reclassified 10 *T. turgidum* accessions into *T. timopheevii* using the gene-based PCR and GBS markers [17].

Change of seed dispersal mode (reduced shattering) is considered as a major mutation leading to the domestication of wheat and a key trait for domestication syndrome [18,19]. The non-brittle rachis was considered a major domestication marker for cereals, where the spike shattering is an observable trait in wild relatives compared to the intact non-shattered rachis of domesticated cereals [20,21]. In barley, two tightly linked genes *Brittle rachis 1* (*Btr1*) and *Brittle rachis 2* (*Btr2*) were associated with the brittle rachis trait, where genetic mutation leading to non-functional genes results in non-brittle spikes [22,23]. In wild emmer, *Btr1* and *Btr2* gene homologs on chromosomes 3A and 3B were identified, where mutation in the *Btr1* gene was found to be associated with the non-shattering trait [24]. Single point mutation of *Btr1* in domesticated einkorn was associated with the non-shattering trait [25]. The adaptability of wheat to diverse environments was attributed to the genetic changes in vernalization-related genes *VRN-1* mapping to the long arm of chromosome 5 [26,27]. *VRN-1* gene encodes MADS-box transcription factor controlling the flowering time and regulates plant's transition from vegetative to reproductive phase [28,29].

There is an urgent need for the preservation of the wild germplasm from the danger of genetic erosion. Wild germplasm holds immense potential for the enhancement of genetic traits like mineral content, grain protein, and resistance to biotic/abiotic stress [30]. For efficient utilization of the germplasm in the breeding program, it is necessary to characterize and classify the available genetic diversity to a small set of accessions representing the maximal diversity. With the advent of parallel sequencing technologies and methods like genotyping-by-sequencing, cost-effective strategies can be developed for the curation of diverse germplasm present in gene banks [31,32]. The present study was undertaken to utilize the low-cost GBS technology to create a reference core set for *T. turgidum* and *T. timopheevii* accessions present in Wheat genetics resource center (WGRC) gene bank for the effective utilization of wild and related tetraploid *Triticum* species in the breeding programs.

Material and methods

Plant material

The present study included 908 accessions of tetraploid wheat from two different species *T. turgidum* and *T. timopheevii* present in the Wheat genetics resource center (WGRC) at Kansas State University (K-State), Manhattan, KS, USA. Table S1 presents passport data and collection site; Fig. 1B shows the geographic location and distribution of the accessions.

Plant tissue collection and genotyping-by-sequencing

For the DNA extraction, plant tissue collection was done from a single plant from each accession after growing them in greenhouse for 2–3 weeks. A small piece of young leaf from a single plant of each accession was cut and collected in 96-well tissue collection box. Plant samples were then lyophilized, followed by genomic DNA extraction using Qiagen BioSprint 96 DNA Plant Kit (QIAGEN, Hilden, Germany). Extracted DNA was quantified with QuantiTTM PicoGreen[®] dsDNA Assay Kit (Thermo Fisher Scientific, Waltham, MA, United States). After DNA quantification and normalization, the DNA samples were transferred to 96-well DNA collection plates with one random well per plate left blank for quality control and library integrity. Genotyping of the DNA samples was performed using genotyping-by-sequencing (GBS) as described [13].

GBS libraries were prepared in 96 plexing using two restriction enzymes—a rare cutter *Pst*I (5'-CTGCAG-3'), and a frequent cutter *Msp*I (5'-CCGG-3') with a common reverse adapter ligated. GBS libraries were sequenced using Illumina HiSeq2000 (Illumina, San Diego, CA, United States) platform at the University of Missouri (UMC; Columbia, Missouri) generating 100 bp long reads.

SNP genotyping and data filtering

TASSEL-GBS V5.0 pipeline was used for SNP discovery [15]. The pipeline was modified for SNP discovery without reference genome. More than 199 k SNPs were discovered that were used for further analysis. Population level SNP filtering was performed and SNPs with minor allele frequency (MAF) less than 0.01 and missing data of more than 20 % were removed. Further, SNPs with heterozygosity greater than 5 % were removed because all the used tetraploid wheat accessions were highly inbred. Fisher's exact test at alpha 0.001 with Bonferroni correction was performed to determine if the putative SNPs were from allelic tags as described [13]. Individual samples with more than 80 % missing SNP calls and more than 5 % heterozygosity were also removed. Retained markers and samples were used for further analyses. After all the filtering steps, 65,535 tags were eventually used for the diversity study.

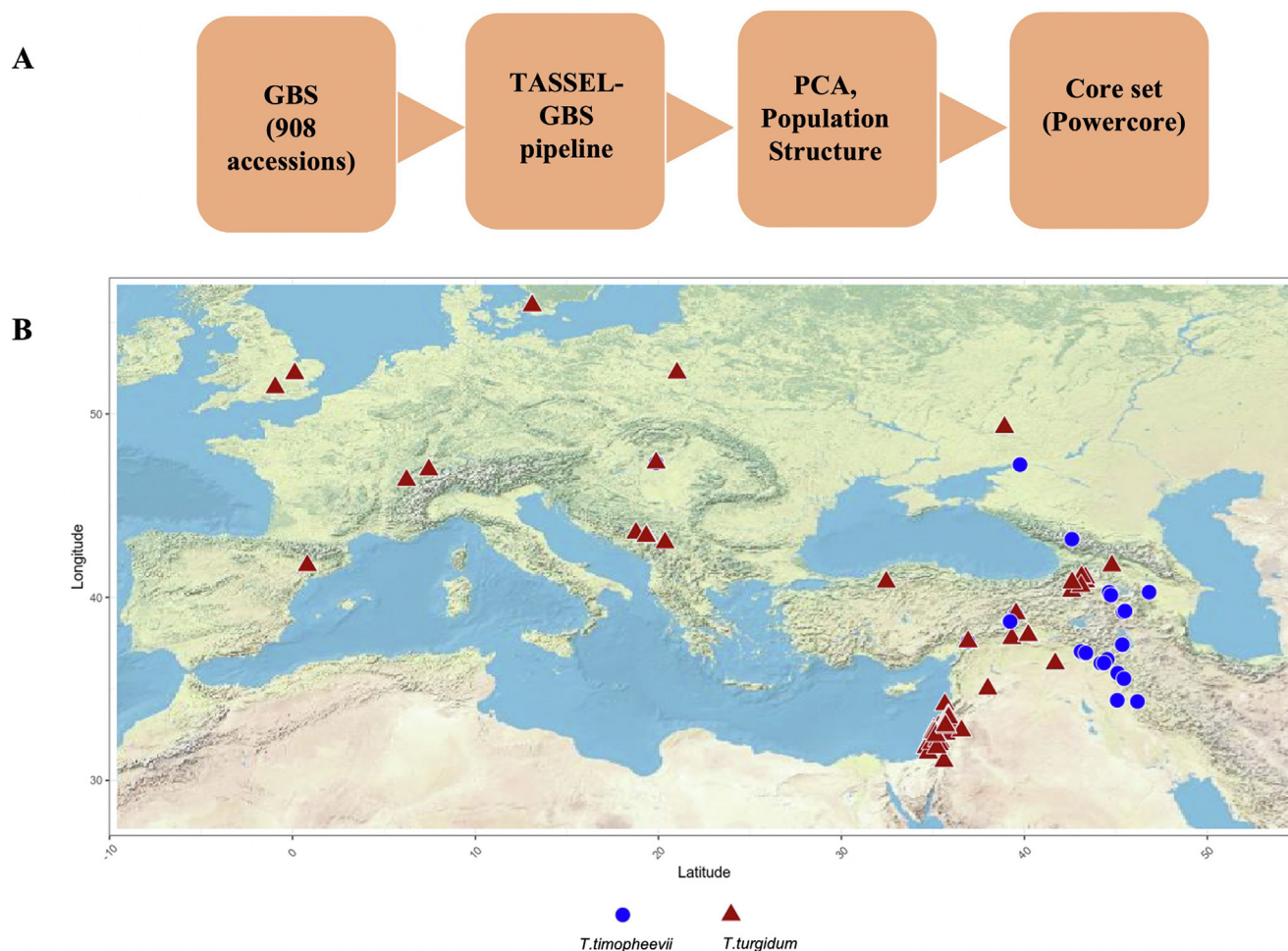


Fig. 1. A) Flowchart of the method used. B) Geographical distribution of tetraploid wheats collection of WGRC. Red stars represent *T. turgidum* and blue stars represent the distribution of *T. timopheevii*. accessions from the same region are clustered together. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Genetic diversity analysis and population structure

GBS tags were blast searched against the *T. turgidum* spp. *durum* genome considering a maximum of three mismatches and no-gap for locating the genomic position. As a measure of average heterozygosity over multiple SNPs in a given population, Nei's diversity index [33] was computed for the whole population using VCFtools [34]. Pairwise F_{ST} between subpopulations and subspecies were computed and plotted using custom R scripts. An R-package *rrBLUP* was used for missing data imputation using the 'mean' method for PCA. Principal component analysis was performed in R. Eigenvalues and eigenvectors were computed with 'e' function using 'A' matrix output of the *rrBLUP* package [35]. Package *ape* was used to draw neighbor-joining tree for all accessions [36]. Species wise Patterson's D (ABBA-BABA) statistic was calculated using Dsuite [37]. Population structure was analyzed at different values of sub-population runs ranging from K = 2 to K = 15 using fastStructure for *T. turgidum* and *T. timopheevii* at both inter and intra-species levels [38]. The best population substructure was selected using *chooseK.py* script of fastStructure. FastStructure output was graphically visualized using R package pophelper [39]. PAUP4.0a163 was used to conduct coalescent SVDquartets analysis [40]. Genome wide association study (GWAS) for leaf rust in *T. timopheevii* using GAPIT [41]. We used the models FarmCPU, GLM, MLM, MLMM and SUPER of gapit. For significant SNP associations FDR less than 0.05 was used.

Diversity of domestication genes

We used the *T. aestivum* *VRN-A1* and *VRN-1B* (AY747600 and AY747606) gene sequence in blast search to detect the corresponding homologs in the *T. turgidum* spp. *durum* genome, *T. dicoccoides* and *T. aestivum*. Similarly, *H. vulgare* homologs of domestication gene *non-brittle rachis 1 (btr1)* and *non-brittle rachis 2 (btr2)* were used for locating the homologous segment for brittle rachis in tetraploid genome. For *photoperiod response locus (Ppd-1)* gene *T. aestivum* (XP_037475580.1) and for *Q-locus* gene *T. turgidum* (AY702955) sequences were used in the blast search.

Genetically diverse representative core set selection

The selected set of SNPs was used to select a representative core-set from the *T. turgidum* and *T. timopheevii* collections. The core sets of tetraploid wheat accessions were identified in two steps using the integration of genotypic and phenotypic datasets. In the first step, identified SNPs were first used with the software package PowerCore to identify core set using default settings [42 43], selecting the lines to retain the most diverse alleles by implementing advanced M (maximization) strategy. One of the limitations of this software was its inability to use more than 15,000 SNPs so we included only SNPs with less than 10–15 % missing data. In the second step the number of selected accessions was further reduced by phenotypically guided selection using the available phenotypic data for growth patterns, leaf rust composite, stem rust (race TTKSK) [44] and hessian fly biotype D resistance. The entire genetic diversity captured by the Mini-Core was assessed by the percent segregating SNPs present in the selected accessions relative to the whole collection.

Results

In the present study, we explored the genetic diversity confined in a diverse collection of tetraploid wheat accessions available at the Wheat genetics resource center (WGRC) at Kansas State University, Manhattan, USA (Table 1). Table S1 shows species pass-

Table 1
Tetraploid germplasm collection of *T.turgidum* and *T. timopheevii* at WGRC gene bank.

Species	Genome	Chromosome (2n)	Number of Accessions
<i>Triticum turgidum</i>	BBA ^U A ^U	28	124
<i>Triticum turgidum</i> spp. <i>dicoccoides</i>	BBA ^U A ^U	28	384
<i>Triticum turgidum</i> spp. <i>turgidum</i>	BBA ^U A ^U	28	2
<i>Triticum turgidum</i> spp. <i>carthlicum</i>	BBA ^U A ^U	28	81
<i>Triticum turgidum</i> spp. <i>dicoccon</i>	BBA ^U A ^U	28	10
<i>Triticum turgidum</i> spp. <i>turanicum</i>	BBA ^U A ^U	28	1
<i>Triticum turgidum</i> spp. <i>polonicum</i>	BBA ^U A ^U	28	1
<i>Triticum turgidum</i> spp. <i>durum</i>	BBA ^U A ^U	28	1
<i>Triticum turgidum</i> spp. <i>paleocolchicum</i>	BBA ^U A ^U	28	2
<i>Triticum timopheevii</i>	GGA ^M A ^M	28	11
<i>Triticum timopheevii</i> spp. <i>timopheevii</i>	GGA ^M A ^M	28	7
<i>Triticum timopheevii</i> spp. <i>armeniicum</i>	GGA ^M A ^M	28	284
Total			908

port information, where 17 duplicated accessions were removed based on the passport and phenotypic data. A total of 908 accessions from *T. turgidum* (primary gene pool) and *T. timopheevii* (secondary gene pool) were used in the study. Fig. 1 shows the flowchart of the methodology used and geographical locations of germplasm collection. *T. turgidum* was represented by eight subspecies: including the wild form *T. turgidum* ssp. *dicoccoides* and 8 other domesticated subspecies (Fig. 2A). The domesticated *T. turgidum* accessions included in the study were *T. turgidum* ssp. *durum*, *T. turgidum* ssp. *paleocolchicum*, *T. turgidum* ssp. *turanicum*, *T. turgidum* ssp. *polonicum*, *T. turgidum* ssp. *dicoccon*, *T. turgidum* ssp. *carthilicum*, and *T. turgidum* ssp. *turgidum*. Wild *T. timopheevii* was represented by *T. timopheevii* ssp. *armeniicum*, whereas *T. timopheevii* ssp. *timopheevii* was the domesticated *timopheevii* subspecies. A set of 10 wheat lines were included in the study to demonstrate the relationship between tetraploid and hexaploid wheat. Fig. 2B and 2C displays the country-wise distribution of the accessions used in the study.

Principal component and cluster analysis

After filtering the SNPs with a minor allele frequency (MAF) of less than 0.01 and missing data greater than 20, a set of 65,535 SNPs were used for the PCA. The first two components represented 27 % and 7 % of the total variability, respectively (Fig. 3A). Accessions were grouped into two major clusters: the first cluster comprised of *T. turgidum* and *T. aestivum* accessions while the second cluster consisted of *T. timopheevii* with an intermix of a few *T. turgidum* accessions. PC1 separated the complexes primarily based on the species type (cluster1 and cluster2), while the PC2 exhibited differentiation of the *T. turgidum* subspecies into 5 sub-clusters along the Y-axis. Further, PCA of tetraploid collection of *T. turgidum* and *T. timopheevii* separated the accessions based on their origin, where PC1 and PC2 contributed 28 % and 7 % of sample variance, respectively (Fig. 3B). Loose clustering was observed for the accessions not belonging to middle eastern origin. *T. timopheevii* ssp. *armeniicum* and domesticated *T. timopheevii* ssp. *timopheevii* grouped into one cluster attributed to their Middle Eastern origin (Armenia, Azerbaijan, Iran, Turkey).

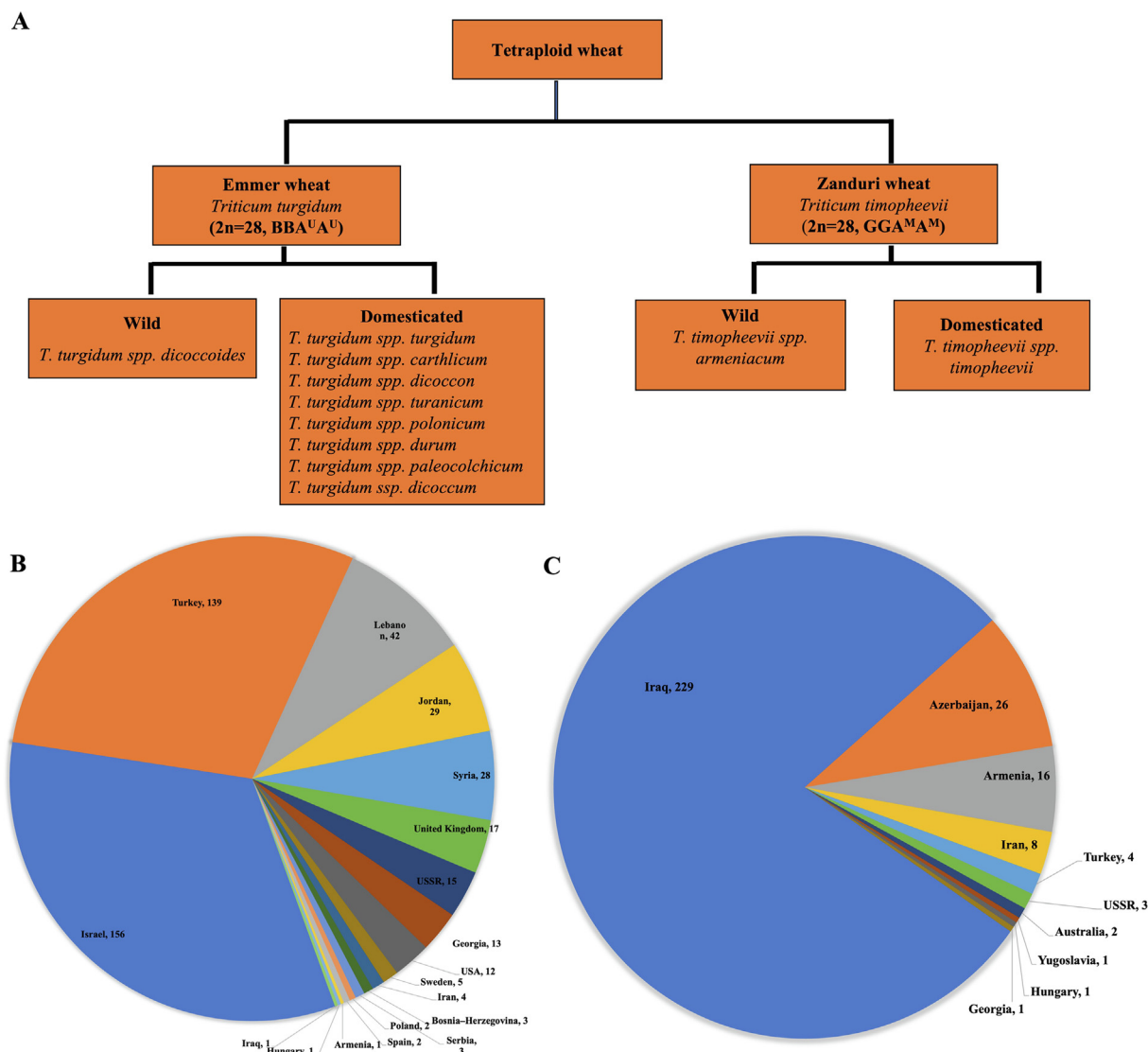


Fig. 2. A) Tetraploid wheat accessions used in the study. Country wise distribution of B) *T. turgidum* and C) *T. timopheevii* accessions.

For *T. turgidum*, a huge variance was observed among the accessions belonging to *T. turgidum* spp. *dicoccoides* (dark green) from the Middle East (Israel, Lebanon, Turkey, Syria, Jordan, and Iraq), Europe (England and Sweden), and the USA. *T. turgidum* spp. *carthlicum* clustered out with the largest variance on PC2. While the accessions of *T. turgidum* spp. *durum*, *T. turgidum* spp. *paleocolchicum*, *T. turgidum* spp. *turanicum*, *T. turgidum* spp. *polonicum*, *T. turgidum* spp. *dicocum*, *T. turgidum* spp. *turgidum* displayed loose clustering. The genotype-based phylogenetic tree formed two distinct clusters; the largest cluster consisted of *T. turgidum* separated from *T. timopheevii* accessions (Fig. S1), wheat accessions formed a subgroup within the largest *T. turgidum* cluster. Five accessions were reclassified based on the clustering pattern in the PCA plot and their proximity to different clusters in the population sub-structure. These include three *T. timopheevii* accessions (TA9, TA875, and TA105) identified as belonging to *T. turgidum* and two *T. turgidum* (TA58 and TA987) accessions were reclassified to *T. timopheevii*. Chromosome-wise distribution of SNPs belonging to *T. turgidum* and *T. timopheevii* are shown in Fig. 3C and 3D, respectively.

Population structure of tetraploid wheat accessions

The fastStructure based population assignment of *T. turgidum* and *T. timopheevii* accessions was performed at different sub-population levels ranging from $K = 2$ to $K = 15$. Using the choose.py script associated with fastStructure, the optimal sub-population level was observed at $K = 14$, indicating the best available population sub-structure for the accessions. (Table 2, Fig. 4). At optimal K -value (14), we observed 11 and 3 species-specific subclusters for *T. turgidum* and *T. timopheevii*, respectively. The highest level of admixture was found in four *T. turgidum* accessions from England and Turkey (TA1002, TA1003, TA1009, and TA1148). *T. timopheevii* accessions were grouped into three distinct clusters (cluster5, cluster8, and cluster14, highlighted with thick line in Fig. 4B), of which *T. timopheevii* spp. *armeniaceum* group consisted of cluster8 and cluster14, while cluster5 majorly consisted of *T. timopheevii* spp. *timopheevii* accessions. Cluster8 constituted the largest group with the majority of accessions hailing from Iraq and a few from Azerbaijan and Turkey. Accessions from Azerbaijan were grouped into cluster14, whereas cluster5 was found to be

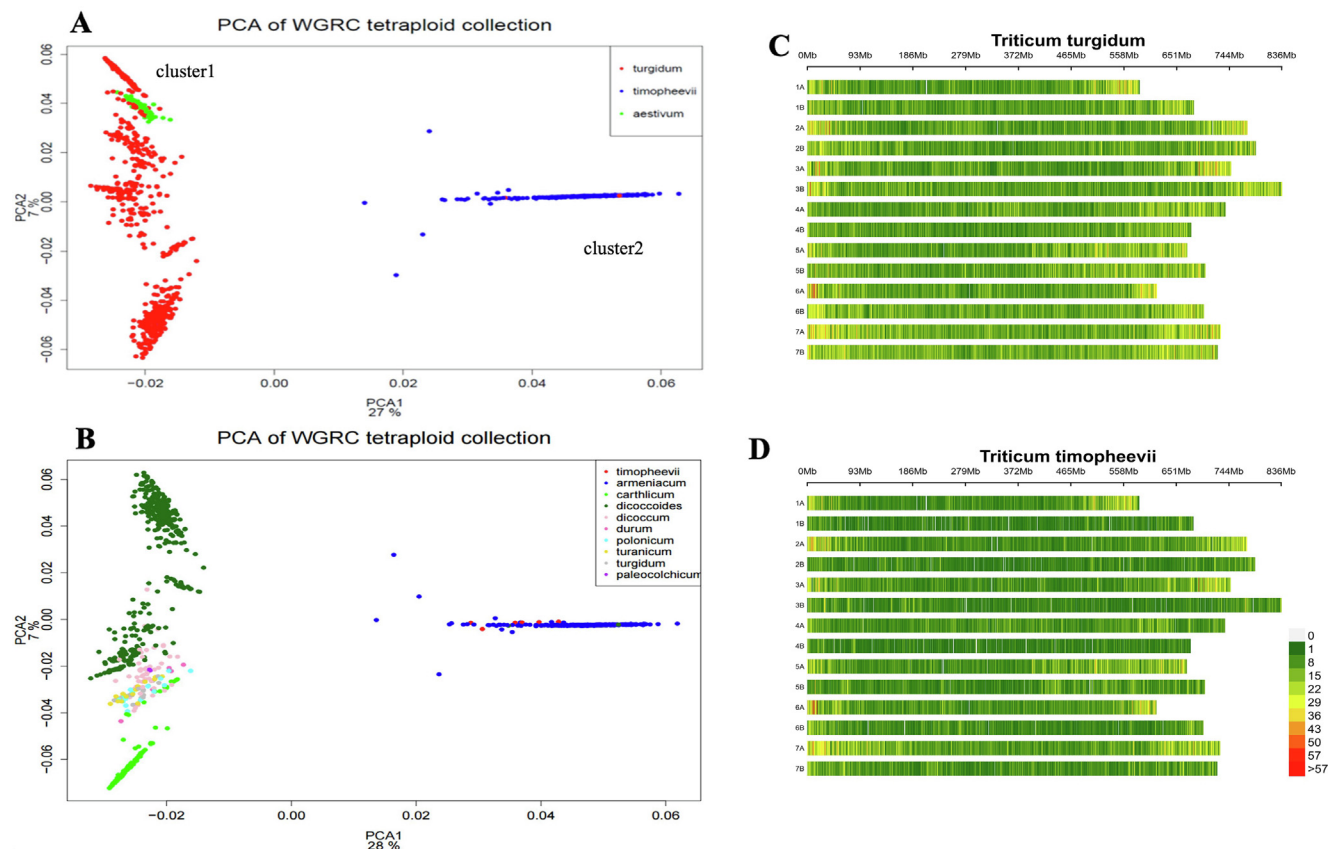


Fig. 3. Principal Component Analysis A) tetraploid collection of three distinct *Triticum* species *T. aestivum* (green dots), *T. turgidum* (red dots) and *T. timopheevii* (blue dots). B) Tetraploid subspecies collection of wild and cultivated subspecies of *T. turgidum* and *T. timopheevii*. Distribution of the SNPs C) SNPs belonging to *T. turgidum* accessions and D) SNPs belonging to *T. timopheevii* accessions mapped to *T. turgidum* genome. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Population sub-clustering of *Triticum turgidum* and *Triticum timopheevii* accessions at optimal sub-population (K = 14).

Population group	Sub-species	Number of accessions	Country of Collection	Nei's genetic diversity index
<i>Triticum turgidum</i>				
Cluster1	<i>dicoccoides</i>	30	Israel, Syria, Iraq	0.059
Cluster2	<i>dicoccoides</i>	156	Israel, Syria, Iraq, Lebanon, Jordan	0.126
Cluster3	<i>dicoccoides</i> (minor: <i>dicoccon</i> , <i>paleocolchicum</i>)	42	England, Sweden, Bosnia–Herzegovina, Serbia, Georgia, Turkey	0.106
Cluster4	<i>dicoccoides</i>	50	Turkey	0.028
Cluster6	<i>dicoccon</i>	15	Serbia	0.043
Cluster7	<i>dicoccoides</i>	64	Turkey, Iraq (1 accession)	0.085
Cluster9	<i>dicoccoides</i>	42	Lebanon	0.082
Cluster10	<i>carthlicum</i>	89	Major (Iraq, Turkey, Armenia), Georgia, Russian Federation, Hungary, Iran, Poland, Spain, Sweden, United Kingdom(UK), United State of America(USA)	0.024
Cluster11	<i>turgidum</i> , <i>dicoccon</i> , <i>turanicum</i> , <i>polonicum</i> , <i>durum</i> , <i>carthlicum</i>	65	England, USA, Spain	0.084
Cluster12	<i>dicoccoides</i> , <i>dicoccon</i>	14	Not available	0.033
Cluster13	<i>dicoccoides</i>	37	Israel	0.063
<i>Triticum timopheevii</i>				
Cluster5	<i>timopheevii</i>	13	Turkey, Iraq, Georgia, Hungary, Russian Federation, Australia	0.016
Cluster8	<i>armeniicum</i>	256	Iraq, Iran, Azerbaijan, Turkey	0.028
Cluster14	<i>armeniicum</i>	32	Israel, Azerbaijan, Armenia, Russian Federation	0.010

composed of a mixture of accessions from the Middle East, European countries (Yugoslavia, Hungary, and Russian Federation), and Australia (Table 2, Fig. 4). *T. turgidum* was grouped into three species-specific classes, *T. turgidum* spp. *dicoccoides* formed clusters 1, 2,3,4,6,7,9,12, and 13, *Triticum turgidum* spp. *carthlicum*

(cluster10), while cluster11 was found to be composed of a mixture with other domesticated sub-species (Table 2). Cluster2 consisted of *T. turgidum* spp. *dicoccoides* accessions from the middle east (Israel, Syria, Iraq, Lebanon, and Jordan), while the majority of *T. turgidum* spp. *carthlicum* accessions in cluster10 were from

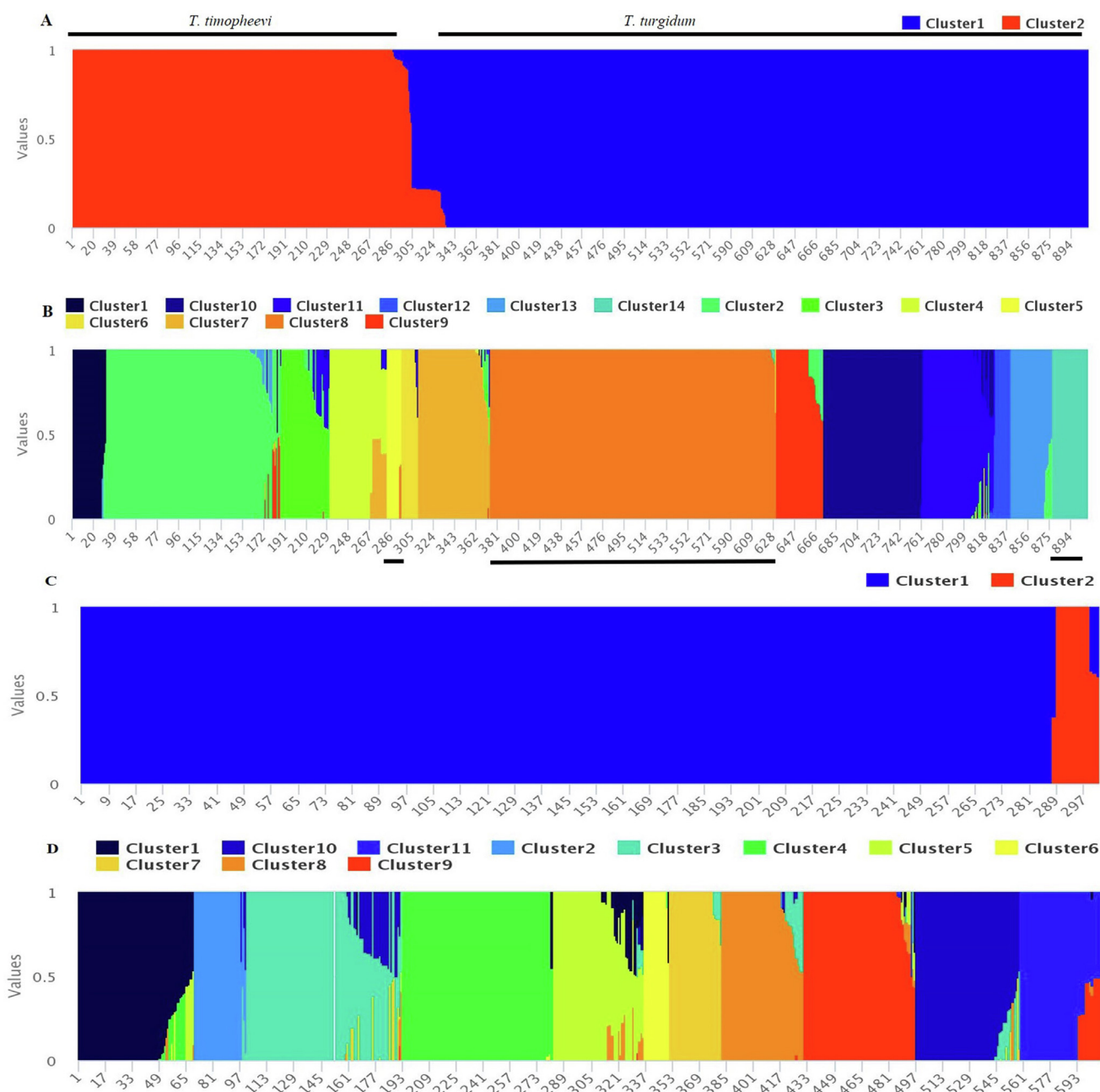


Fig. 4. fastStructure based population sub-clustering of *T. turgidum* and *T. timopheevii* accessions. Using complete set of accessions, A) sub-clustering at K = 2, B) optimal population sub-clustering at K = 14. Species-specific population sub-clustering C) *T. timopheevii* at optimal K = 2 and D) *T. turgidum* at optimal K = 11.

Turkey, Georgia, and Russian Federation, with fewer accessions from other European countries (Armenia, Hungary, Poland, Spain, Sweden, and the United Kingdom).

A slightly different clustering pattern was observed during species-specific population clustering (Table 3, Fig. 4). *T. timopheevii* accessions clustered into two sub-populations at the optimal K-value of 2 compared to the previous 3 clusters. Majority of *T. timopheevii* spp. *armeniicum* accessions (288) formed the largest group, while the *T. timopheevii* spp. *timopheevii* (13) grouped into the second cluster. Cluster1 consisted of timopheevii accession from Azerbaijan, Iran, Armenia, Israel, Turkey, Iraq, and Russian Federation, while cluster2 consisted of accessions from Turkey, Georgia, Hungary, Iraq, Azerbaijan, Russian Federation, and Australia. For *T. turgidum*, cluster3 and cluster4 composed the largest group with 89 accessions each. Cluster1 was found to be the most

diverse and consisted of accessions belonging to *T. turgidum* spp. *dicoccoides*, *T. turgidum* spp. *dicoccum*, *T. turgidum* spp. *carthlicum*, *T. turgidum* spp. *polonicum*, *T. turgidum* spp. 'blue' *dicoccum*, *T. turgidum* spp. 'blue' *durum*, *T. turgidum* spp. 'white' *dicoccum*, *T. turgidum* spp. *turanicum*, *T. turgidum* spp. 'short' *turanicum*, *T. turgidum* spp. *turgidum*, *T. turgidum* spp. *durum*, *T. turgidum* spp. *turgidum*. Cluster6 consisted of *T. aestivum* accession grouped with *T. turgidum* spp. *carthlicum*.

Genome-wide nucleotide diversity, F_{ST} scan, and introgression signal

Nucleotide diversity (Nei's Π) demonstrates the average pair-wise distance between all possible pairs of individuals, calculated based on 76 % (152,769) of 199,349 tags mapping to *T. turgidum* spp. *durum* genome with criteria of 3 mismatches and zero-gap

Table 3Species-specific population sub-clustering of *Triticum timopheevii* (K = 2) and *Triticum turgidum* (K = 11).

Population group	Sub-species	Number of accessions	Country of collection	Nei's genetic diversity index
<i>Triticum</i>	<i>timopheevii</i>			
Cluster1	<i>armeniicum</i>	288	Azerbaijan, Iran, Armenia, Israel, Turkey, Iraq, Russian Federation	0.028764
Cluster2	<i>timopheevii</i>	13	Azerbaijan, Turkey, Georgia, Hungary, Iraq, Russian Federation, Australia	0.0157207
<i>Triticum turgidum</i>				
Cluster1	<i>carthlicum</i> , <i>dicoccoides</i> , <i>dicoccon</i> , <i>durum</i> , <i>polonicum</i> , <i>turanicum</i> , <i>turgidum</i>	68	USA, Spain, England	0.0855099
Cluster2	<i>dicoccoides</i>	30	Iraq, Syria, Jordan, Israel, USA	0.0592321
Cluster3	<i>dicoccoides</i>	89	Israel, Jordan	0.118767
Cluster4	<i>carthlicum</i>	89	USA, Sweden, Russian Federation, Georgia, Spain, United Kingdom, Poland, Turkey, Armenia, Hungary, Iran, Poland	0.0239709
Cluster5	<i>dicoccoides</i> , <i>dicoccon</i> , <i>paleocolchicum</i>	50	England, Sweden, Turkey, Bosnia–Herzegovina, Serbia, Sweden, Georgia	0.111123
Cluster6	<i>dicoccon</i>	15	Serbia	0.0438263
Cluster7	<i>dicoccoides</i>	31	Israel	0.0513653
Cluster8	<i>dicoccoides</i>	49	Iraq, Syria, Lebanon	0.0925869
Cluster9	<i>dicoccoides</i>	64	Turkey	0.0839453
Cluster10	<i>dicoccoides</i>	61	Armenia, Israel, Syria, Jordan, USA, Lebanon, Israel	0.0952499
Cluster11	<i>dicoccoides</i>	50	Turkey	0.0294869

per tag. Higher genetic differentiation of 0.27 was observed between the two tetraploid species *T. turgidum* and *T. timopheevii* (Table 4). Nei's index values for the individual genome (A and B genome) were found in concordance with the global diversity of 0.273 (A genome) and 0.279 (B genome). Within-species comparison resulted in lower nucleotide diversity of 0.18 and 0.04 for *T. turgidum* and *T. timopheevii*, respectively. Species-specific diversity was found to be consistent with the regional distribution of the two species. The pairwise fixation index (F_{ST}) revealed a high level of divergence between the two species. *T. turgidum* and *T. timopheevii* populations demonstrated a high F_{ST} value (0.72). Between the two species, B-genome showed lower F_{ST} (0.70) compared to A-genome (0.74), owing to fewer tags mapping to B-genome from *timopheevii*. Despite regional proximity, two of the *T. timopheevii* sub-populations, *T. timopheevii* spp. *armeniicum* and *T. timopheevii* spp. *timopheevii* displayed a higher F_{ST} (0.60), indicating more sub-species differentiation due to selection and domestication. (Fig. S2). A comparative lower F_{ST} index of 0.44 was observed between *T. turgidum* spp. *dicoccoides* and *T. turgidum* spp. *carthlicum*. A statistically significant introgression signal with positive D-statistics and Z-score of more than 3 was observed

between *T. turgidum* spp. *dicoccoides* and its domesticated accessions with the four-taxon f_d statistic. No significant introgression was observed between the *T. timopheevii* accessions. Relative introgression was detected between the wild and domesticated *T. turgidum* and *T. timopheevii* (Table S2). From the available data, we were unable to detect the introgression signal of tetraploid accession into the wheat.

Identification of duplicated accessions and mini-core set of *T. Turgidum* and *T. timopheevii* accessions

Monoallelic SNPs were filtered using stringent criteria of keeping sites with MAF greater than 0.01 and less than 20 % of missing data. Keeping the diversity, accessions sharing 99 % or more of similar bases in pairwise comparison, were treated as duplicated. Among *T. timopheevii*, 105 (*T. timopheevii* spp. *armeniicum*:94, *T. timopheevii* spp. *timopheevii*:11) unique accessions were identified. Among the *T. turgidum*, 319 (*T. turgidum* spp. *dicoccoides*: 181 and domesticated emmer: 138) unique accessions were identified (Table S3). Based on the Powercore analysis and phenotypic data, 37 core accessions were identified for *T. timopheevii*, 27 accessions for *T. turgidum* spp. *dicoccoides* and 38 accessions for domesticated *T. turgidum* from the WGR collection (Fig. S3, S4, Table S4). Core set preserved ~ 95 %, 93 %, and 98 % of allelic diversity present in wild emmer, domesticated emmer, and *T. timopheevii*, respectively. Lower Nei's diversity values for the core set indicated retention of a rich level of nucleotide diversity for *T. timopheevii* (Π :0.04), wild emmer (Π :0.169) and domesticated emmer (Π :0.12), these were comparable to the species-specific Nei's values. (Table 4).

Center of diversity

Nei's nucleotide diversity for different admixture clusters was found to be lower except for clusters2 and cluster3, which showed a relatively higher value for *T. turgidum* (Table 2). Cluster2 consisted of *dicoccoides* accessions from Israel, Syria, Iraq, Lebanon, and Jordan and may be considered as the epicenter of *T. turgidum* diversity. While cluster 3 also consisted of *dicoccoides* accessions from the European region. Cluster 10 contained the maximum diversity of domesticated accessions. For, *T. timopheevii* spp. *armeniicum* cluster 8 displayed most of the diversity in compar-

Table 4Genome and species specific Nei's diversity indices and pairwise F_{ST} coefficients of tetraploid species and core collection.

Species	Nei's genetic diversity index	F_{ST}
<i>T. turgidum</i> and <i>T. timopheevii</i>	0.270	0.72
A-genome	0.273	0.74
B-genome	0.279	0.70
<i>T. timopheevii</i>	0.040	0.60
Wild	0.028	
Domesticated	0.016	
<i>T. turgidum</i>	0.180	0.44 (<i>dicoccoides</i> & <i>carthilicum</i>)
Wild	0.180	
Domesticated (all accessions)	0.130	
<i>carthilicum</i>	0.030	
Core collection		
<i>T. timopheevii</i>	0.048	
<i>T. turgidum</i> (wild)	0.169	
<i>T. turgidum</i> (domesticated)	0.120	

ison to the other two clusters, predominantly accessions from Iraq and Azerbaijan (Table 2). Comparison of F_{ST} values indicates species differentiation. We calculated the pairwise F_{ST} for each subpopulation cluster/group. Significantly, high F_{ST} values ranging from 0.242 to 0.965 were observed between the different subpopulation clusters (Fig. S9). F_{ST} based clustering grouped cluster 2 and cluster 3 into independent groups of similar diversity profiles. Group I consisted of clusters 1, 2, 9, and 13 from the Mediterranean region (Israel, Syria, Iraq, Lebanon, Jordan), while Group II consisted of clusters 3, 6, 11, and 12 from the European region. Group III consisted of *T. turgidum* ssp. *dicoccoides* (clusters 4,7) and *T. turgidum* ssp. *carthlicum* (cluster10) with many accessions from Turkey and Georgia. Cluster10 was found to be the most diverse group, with a collection of accessions from Central Europe, England, and the USA.

A species tree was developed using the SVDquartets coalescent model implemented in PAUP. The species tree demonstrated distinct evolution of *T. timopheevii* from *T. turgidum* and a clear distinction was observed between wild and domesticated emmer wheat. *T. aestivum* was clustered with *T. turgidum* ssp. *carthlicum* owing to its recent evolution (Fig. S10A). A consensus tree based on population structure groups from Table 2, categorized the species clusters according to their geographic location. (Fig. S10B). Clusters 2, 9, and 13 from Group I of F_{ST} clustering were grouped into a single cluster; clusters 4 and 7 from Group III of F_{ST} formed a separate group; and clusters 6, 10, and 11 with domesticated accessions were congregated with hexaploid wheat. Accessions from cluster10 shared 71–81 % of alleles with accessions from cluster 4 and 7, with most of the accessions belonging to Turkey. We grouped all the *T. turgidum* accessions from Turkey from clusters 3, 4, 7, and 10 into one group to calculate the nucleotide diversity, which was found to be significantly higher (0.104) compared to individual groups.

Similarly, a higher genetic distance (F_{ST} : 0.64) was observed between the wild and domesticated accessions in this region. The *T. turgidum* ssp. *carthlicum* accessions from England, the United States, Sweden, and Poland were classified as belonging to cluster 10, except for two accessions from the United States (TA2843 and TA2884) and one accession from Spain (TA2868), which showed admixture between clusters 10 and 11. Cluster10 consisted of accessions majorly from Turkey and with an admixture coefficient of 0.99 indicates Turkey as the center of diversity. *T. turgidum* ssp. *dicoccoides* accessions from England exhibited admixture with samples from clusters 2, 3, and 11, while accessions from the United States were assigned to cluster 2 and 3, accessions from Sweden showed admixture between clusters 3 and 2 (Fig. S11). This distinct pattern of *dicoccoides* accessions may be due to the regional isolation and was also observed in the cluster-specific tree (Fig. S10B). Cluster2, with the maximum diversity of wild emmer, may be considered as the center of species diversity for *T. turgidum* ssp. *dicoccoides* and comprised of the region of Israel, Jordan (West Bank), Syria (Golan Heights) and Lebanon (Al Biqa). Domesticated emmer wheat was found in highest diversity in the regions of Turkey (Kars) and Georgia (Tiflis). With the highest genetic diversity of accession from Turkey, we think that domesticated accessions have found their way into central Europe through Turkey. For *T. timopheevii* spp. *armeniaceum*, the region of higher genetic diversity comprised of Iraq (Arbil, Dahuk, and As Sulaymaniyah), Azerbaijan (Nakhichevanskaya), and Armenia. A comparison of the nucleotide diversity of accessions from these three areas revealed that accessions from Iraq (0.29) had the highest diversity, followed by Azerbaijan (0.19) and Armenia (0.07). F_{ST} comparison revealed lesser divergence (F_{ST} :0.30) between the populations of Iraq and Azerbaijan compared to the Iraq-Armenian population (F_{ST} : 0.41). Based on the higher genetic diversity and

resistant accessions region comprising Arbil, Dahuk, and As Sulaymaniyah in Iraq may be the potential site of origin for the wild type *T. timopheevii*.

Signatures of domestication and adaptation-related genes in tetraploid wheat

VRN-A1 and *VRN-1B* gene sequences of *Triticum aestivum* (NCBI accessions AY747600 and AY747606) were used for the detection of corresponding candidate regions in the *T. turgidum* genome [45]. Blast search identified the gene positions mapping to chromosome 5A (549156384–549152141) and 5B (570844281–570831393). Differential but similar F_{ST} distribution was observed in the candidate region for *VRN-1* on chromosomes 5A and 5B in both *T. turgidum* and *T. timopheevii* accessions. For the domestication gene *Brittle rachis* (*Btr*), two physically linked genes *non-brittle rachis 1* (*btr1*) and *non-brittle rachis 2* (*btr2*) of *H. vulgare* (NCBI accession: KR813335.1) were used in the blast search. [22]. Two chromosomal segments corresponding to 3A and 3B were identified (Table S5). An elevated F_{ST} was observed for both *T. turgidum* and *T. timopheevii* (Fig. S6, S7). *VRN* and *Btr* genes of *T. dicoccoides* and *T. turgidum* showed higher similarity to *T. aestivum* genes. Fig. S8 displays F_{ST} distribution of *Ppd-1* and *Q-locus*. For *Ppd-1A*, single blast hit was observed on chromosome 2A, while no similarity was detected on chromosome 2B of *T. turgidum*. *Q-locus* showed hits on chromosomes 5A and 5B.

Distribution of disease resistant *T. Timopheevii* accessions

Out of 105 unique *T. timopheevii* accessions, phenotypic data was made available for 95 accessions related to leaf rust, Septoria, tan spot, and powdery mildew (data not shown). Fig. S5 shows the distribution of phenotypic data for *T. timopheevii*. Most of the accessions showing resistance to all four diseases were found belonging to Iraq (region of Arbil, Dahuk, and As Sulaymaniyah). Detailed examination revealed accessions showing resistance to Septoria and tan spot belonging to Armenia, Azerbaijan, Iran, and Iraq. Accessions from Armenia and Azerbaijan showed moderate resistance to Powdery mildew and susceptibility to leaf rust, while accessions from Iran were susceptible to both Powdery mildew and leaf rust. Single accessions from Australia and Hungary were resistant to leaf rust only or they were selected by the collectors for being resistant to leaf rust. Genome-wide association study (GWAS) for leaf rust identifies four significant marker-trait associations (MTAs) using MLM and three significant MTAs using FarmCPU on chromosome 2B. One MTA (SNP_ 29001) on distal region of chromosome 2BS was found to be common with both the methods (Table S6; Fig. S12). Therefore, this MTA was considered more important and hence was further examined for putative candidate genes (CGs). For identification of putative CGs, the MTA region was extended to 100 kb in both directions and CGs were identified in this 200 kb region. In this region, four genes were identified which encoded for two NBS-LRR-like resistance proteins (TRITD2Bv1G010030 and TRITD2Bv1G010060), and two Exocyst complex component 2C putative proteins (TRITD2Bv1G010050 and TRITD2Bv1G010090). NBS-LRR-like resistance proteins play a critical role in providing resistance against leaf and stripe rust of wheat. Interestingly, two of the reported leaf rust (*Lr*) resistance genes, *Lr16* [46] and *LrZH2* [47] were also reported earlier in distal region of the chromosome 2BS. Therefore, we believe that the above two NBS-LRR genes may be the potential candidates for either of the two *Lr* genes; however, this certainly needs further validation. The results need to be further examined with the inclusion of more phenotypic data in the future.

Discussion

The WGRC tetraploid wheat collection included germplasm from all over the world, especially from the Fertile Crescent region of the Middle East (Fig. 1). Archaeological evidence suggests that wild emmer and free-threshing tetraploid domesticated emmer were first cultivated in the southern Levant region as early as 10000 years ago [5]. Presently, wild emmer is found growing in the western region of the fertile crescent (Israel, Jordan, and Syria), south-eastern region of Turkey, mountain ranges of east Iraq and west of Iran [48]. Jordan Valley, with the highest genetic diversity, was considered to be the epicenter of wild emmer. [21]. *T. timopheevii* spp. *armeniicum* (*T. araraticum*) displayed a spread in Iran, Iraq, Georgia, Armenia, Azerbaijan, and Turkey [49], whereas *T. timopheevii* spp. *timopheevii* found to be endemic to Georgia [50]. Both the tetraploid species are self-pollinated. Crosses between *T. turgidum* and *T. timopheevii* result in a non-fertile progeny due to failed pairing of chromosomes [51]. These two tetraploid species share highly similar morphological characteristics [17] and to some extent chromosomal similarities [12] that can result in misclassification, mislabeling, and duplication of accessions from different countries. In the absence of passport information, sometimes accession identification becomes difficult. The WGRC gene bank provides germplasm collections to the breeders for its efficient utilization in the breeding program for wheat improvement. Considering the accession duplication, efficient strategies were needed to narrow down these to a core set of accessions representing most of the diversity for their easy utilization in breeding programs. The availability of a core set will reduce the effort needed for the phenotyping of a large number of accessions. An efficient but time-consuming strategy involves using the combination of phenotypic and passport information to limit the number of accessions. Clustering-based methods like PCA can efficiently be used with next-generation sequencing data for their characterization using the genome-wide tags generated through GBS [52]. In the present study, 908 accessions of tetraploid wheat from the WGRC gene bank were used to define a core collection of genotypes covering most of the available genetic diversity.

Tetraploid wheat races and their distribution

Wild emmer germplasm has been explored in many studies using whole-genome sequencing [6,24,53,54] and exome sequencing [55]. Walkowiak and coworkers explored the secondary gene pool species i.e. *T. timopheevii* for the identification of introgression segment on chromosome 2B of wheat variety LongReach Lancer [56]. Cytological methods were widely used for the detection of *timopheevii* introgression signals in the wheat [57,58]. Wild emmer accessions were previously classified into wild race I (*T. turgidum* ssp. *dicoccoides* (Körn. ex Asch. & Graebner) Schweinf.) race II (*T. turgidum* ssp. *dicoccum* Schrank ex Schübler), race IV (*T. turgidum* ssp. *durum* Desf.), race V (*T. turgidum* ssp. *polonicum* L.) and race VI (*T. turgidum* L.) [59]. Races were reclassified as *T. turgidum* and *T. timopheevii* (Zhuk) Zhuk subspecies [60]. Hulled wild tetraploids include *T. timopheevii* spp. *armeniicum* and *T. turgidum* ssp. *dicoccoides*, while *T. timopheevii* spp. *timopheevii* and *T. turgidum* ssp. *paleocolchicum* constitutes members of domesticated hulled subspecies. Domesticated free-threshing emmer species consists of *T. turgidum* ssp. *carthlicum*, *T. turgidum* ssp. *durum*, *T. turgidum* ssp. *polonicum*, *T. turgidum* ssp. *turanicum*, *T. turgidum* ssp. *turgidum* and *T. turgidum* ssp. *turgidum* var *uniaristatum*.

Diversity analysis and population structure

Our results showed higher but almost similar nucleotide diversity between the A-genome (0.273) and B-genome (0.279) of *T. tur-*

gidum and *T. timopheevii*. Although literature reports a higher dissimilarity in the B-genome compared to A-genome between the two species. A potential reason for inconsistency may be the lesser (40 %) mapping of G-genome tags on B-genome used in blast search. Lesser diversity was observed in the species-specific accessions (Table 4). A higher Nei's diversity index was observed for wild accessions (*T. turgidum* spp. *dicoccoides*: 0.180, *T. timopheevii* spp. *armeniicum*: 0.028) compared to domesticated accessions (*T. turgidum*: 0.13, *T. timopheevii*: 0.016). Sub-population cluster3 and cluster4 among the *T. turgidum* ssp. *dicoccoides* showed the highest and lowest Nei's index diversity, respectively (Table 3). Diversity analysis based on fastStructure and F_{ST} estimates grouped two tetraploid species into distinct clusters supporting the species-specific differentiation between the *T. turgidum* and *T. timopheevii*. Rodriguez and coworkers reported higher homologous pairing of B genome of *Ae. speltoides* and G genome of *T. timopheevii* compared to B-genome of *T. aestivum*, suggesting the most recent emergence *T. timopheevii* [61]. In the absence of migration, population divergence increases significantly and results in higher differentiation (F_{ST} values). Different threshold values were reported for significant F_{ST} estimates among the populations, such as greater than 0.25 [62] or greater than 0.15 [63]. We observed intra-species differentiation between wild and domesticated *T. timopheevii* accessions belonging to *T. timopheevii* spp. *armeniicum* and *T. timopheevii* spp. *timopheevii* and also, between wild and domesticated *T. turgidum* accessions i.e., *T. turgidum* ssp. *dicoccoides* and *T. turgidum* ssp. *carthlicum* resulting from domestication (Table 4). *T. timopheevii* spp. *armeniicum*, grows widely from Armenia to Azerbaijan region of the former Russian Federation, while domesticated *timopheevii* remained endemic to a small region in the west of Georgia with cultivation limited to a few villages [64]. The higher genetic differentiation observed within the *timopheevii* accessions may be due to their region-specific cultivation or may be attributed to the lesser diversity captured by fewer accessions of domesticated *timopheevii* under study.

Out of the 11 clusters of emmer wheat, *T. turgidum* ssp. *dicoccoides* majorly or solely contributed to the seven clusters (Table 3). Lower nucleotide diversity was observed within the domesticated accession clusters (cluster4 and cluster6) compared to the wild *dicoccoides*. Clustering distinction observed for *T. turgidum* ssp. *carthlicum* and *T. turgidum* ssp. *dicoccon* was an indication of speciation, compared to *T. turgidum* ssp. *durum*, *T. turgidum* ssp. *polonicum*, *T. turgidum* ssp. *turanicum* and *T. turgidum* ssp. *turgidum* (Fig. S10B). Previous studies on durum using molecular markers i.e., SSR, DaRT and SNPs suggested a lesser genetic differentiation within the domesticated *turgidum* accessions [65–67]. Separate sub-population clustering and low nucleotide diversity of *T. turgidum* ssp. *carthlicum* supports its genetic differentiation from other tetraploid wheat sub-species. Riefole and coworkers (2011) reported separate phylogenetic positions for *T. turgidum* ssp. *durum* and *T. turgidum* ssp. *carthlicum* [68]. *T. turgidum* ssp. *carthlicum* (also known as Persian wheat) was supposed to be originated from the cross of domesticated emmer wheat and hexaploid wheat [11,69]. Genetic markers also reported separate phylogenetic clades for *T. turgidum* ssp. *carthlicum* and *T. turgidum* ssp. *dicoccon* [66,70] and provide support for the separate *T. turgidum* ssp. *dicoccon* cluster in population structure analysis and species tree. A recent study reported three distinct lineages for *T. araraticum* based on the genomic and cytogenetic diversity of 862 tetraploid accessions [64]. Two lineages for *T. araraticum* were designated as ARA-0 and ARA-1 and one domesticated *T. timopheevii*. Lineage ARA-0 had wider geographic distribution compared to ARA-1. In our results, population clustering using the whole set of *T. timopheevii* and *T. turgidum* accessions resulted in 3 sub-population clusters (cluster5, cluster8 and cluster14) attributed to the regional specificity of cluster8 to Iraq and cluster 14 to Azer-

baijan and Armenia., but the pattern diminishes during comparison of *T. timopheevii* only accessions, where lower Nei's diversity (0.028) was observed for *T. timopheevii* spp. *armeniicum* and 0.015 for *T. timopheevii* spp. *timopheevii*.

Phylogenetic analysis of wheat chloroplast genomes also showed species-specific distinctions of Poaceae members, specifically the tetraploid wheat [3]. In another GBS based study, two major classes of *T. timopheevii* spp. *armeniicum* and *T. timopheevii* spp. *timopheevii* were reported [70]. These results were consistent with our finding of two major lineages of *T. timopheevii*. More *timopheevii* accessions may be required to support three clade classification. The potential reason for the differential clustering may be the closeness of ARA-1 class to the *T. turgidum* as discussed by Badaeva et al., [64]. Based on nucleotide diversity and differentiation, the region comprising Israel, Jordan (West Bank), Syria (Golan Heights) and Lebanon (Al Biqa) were rich in diversity for *T. turgidum* spp. *dicoccoides*. *Carthilicum* displayed rich diversity in the region of Turkey (Kars) and Georgia (Tiflis). Regions of Iraq (Arbil, Dahuk, and As Sulaymaniyah), Azerbaijan (Nakhichevanskaya), and Armenia can be considered as diversity rich areas for *T. timopheevii* spp. *armeniicum* with the maximum genetic diversity in Iraq. Fig. 6 summarize major identified center of genetic diversity, potential center of domestication and route of cultivation for tetraploid wheat.

F_{ST} signal surrounding the domestication genes

A variable and low *F_{ST}* signal was observed in the potential candidate region for the genes: the non-brittle rachis (*Btr1*), vernalization (*VRN-1*), developmental signaling gene *Ppd-1A* and *Q locus*. However, no clear haplotype was identified in the genic region in the absence of tags mapping on the gene, but a variable and alike *F_{ST}* signal between the wild and domesticated tetraploids on A and B genome provides evidence for the emergence of domestication and environmental adaptation traits in domesticated accessions. Variability in *VRN-1* (MADS-box transcription factor) controls the adaptability of the plant to different environments [26]. *VRN-1* alleles in *T. dicoccoides* and *T. timopheevii* displayed differential origin with insertion in the main gene in *T. timopheevii* and a region of variability in the promotor region of *T. dicoccoides* [71]. In another study of gene-based haplotypes using our core accessions, domesticated *T. turgidum* was found to share haplotypes for the *Btr1-A* and *Btr1-B*. In *T. timopheevii*, novel mutations were identified in the *Btr1-A* allele, displaying partial brittleness of spikes, while no amplification was observed for the allele on the G-genome, suggesting the differential domestication event for *T. timopheevii* [72]. The region surrounding the genes related to domestication traits showed modest differentiation between wild and domesticated accessions, demonstrating the preservation of diversity in the region (Fig. S8).

Core collection

To accelerate the breeding program, we reported a core set of accessions from wild and domesticated accession of *T. timopheevii* and *T. turgidum*. The selected core set of accessions represents more than 93 % of the available allelic diversity, with nucleotide diversity comparable to individual population used in the study (Table 4). From the original number of accessions, a nearly 10x sample reduction was achieved in the core set. Wild emmer consisted of a majority of the accessions from Israel (11) followed by Turkey (6), Jordan (5), and one each from Sweden, Syria, England, and Lebanon, representing the sub-population clusters 1,2,3,4,7, and 13. Domesticated emmer consisted of accessions from Turkey (7), Portugal (3) and one each from Jordan, Sweden, Syria, Afghanistan, Armenia, Former USSR, Greece, Hungary, Kyrgyzstan, Oman, Spain, United States, and Yemen, representing sub-population

clusters 2,3,6,10,11 and 12. *T. timopheevii* spp. *araminicum* core set consisted of accessions from Iraq (21), Armenia(5), Azerbaijan (4), Turkey(2), Russian Federation (2) and Iran(1) from cluster8 and 14, while *T. timopheevii* spp. *timopheevii* consisted of two accessions from Russian federation. Our *T. timopheevii* core set captures a rich source of diversity against fungal diseases like *Septoria*, leaf rust, tan spot, and powdery mildew (Fig. 5). The present core set will serve as a rich source of genetic variations to be utilized in the breeding program and should be coupled with careful phenotyping. In recent years, gene bank curation has been done to explore the novel genetic variations in *T. monococcum* [73], *T. aestivum* [74–77], *Aegilops tauschii* [43], *T. turgidum* L. [78], rye [79], sorghum [80], melon [81], pumpkin [82], capsicum [83] for trait improvement.

GBS is a relatively inexpensive method for genotyping large numbers of samples and provides more SNPs than SNP arrays. Domesticated varieties, although are high yielding but lacks resistance to biotic and abiotic stress. Wild germplasm with rich and unexplored genetic diversity holds the potential for traits improvement related to resistance, mineral content, and protein content [30]. Several resistance genes for powdery mildew resistance *Pm6*[84], *Pm27* [85], *Pm37* [86], *MLG12* [87], leaf rust *Lr50* [88], *LrTt2* [89], stem rust *Sr40* [90] and fusarium head blight resistance [91,92] has been introgressed into the bread wheat from *T. timopheevii*. In recent years, *Pm36*[93], *Pm41*[94], *pm42*[13], and leaf and stripe rust resistance [95] genes were contributed by the wild emmer. The present tetraploid core set will serve as a rich resource of genetic variations. The small set of accessions can effectively be phenotyped by researchers for several traits under field or greenhouse conditions. SNP identified from the collection will facilitate future breeding, association mapping, and introgression

Accession	Septoria	Tan Spot	Powdery Mildew	Leaf Rust
TA105	R	R	5	9
TA1008	R	R	9	8
TA6	R	R	-	8
TA23	-	-	5	8
TA25	R	-	4	2
TA48	-	-	4	8
TA101	R	R	5	9
TA145	R	VR	4	2
TA153	-	-	2	2
TA170	-	-	0	3
TA861	R	R	4	5
TA877	R	R	4	8
TA921	-	-	4	7
TA946	R	R	4	2
TA1486	R	R	1	2
TA1488	R	R	5	8
TA1489	-	-	3	2
TA1499	R	R	7	8
TA1510	-	-	4	9
TA1514	-	-	4	8
TA1520	R	R	3	1
TA1521	-	-	4	1
TA1526	R	R	4	5
TA1536	R	R	5	3
TA1571	R	R	4	9
TA892	-	-	1	2
TA905L1	-	-	5	2
TA44	R	R	4	-
TA49	R	R	4	8
TA1564	R	R	5	5
TA1565	-	-	3	9
TA1569	R	R	4	9
TA2892	R	R	6	8

Fig. 5. Heatmap of selected *T. timopheevii* accessions in response to Septoria, tan spot, powdery mildew and leaf rust. VR: very resistant or immune; R: resistant; Powdery mildew (0:immune,1:resistant,2–4: Moderately resistant,5–9:Susceptible) Leaf rust (0:immune,1–2:resistant,3–4: Moderately resistant,5–9:Susceptible).

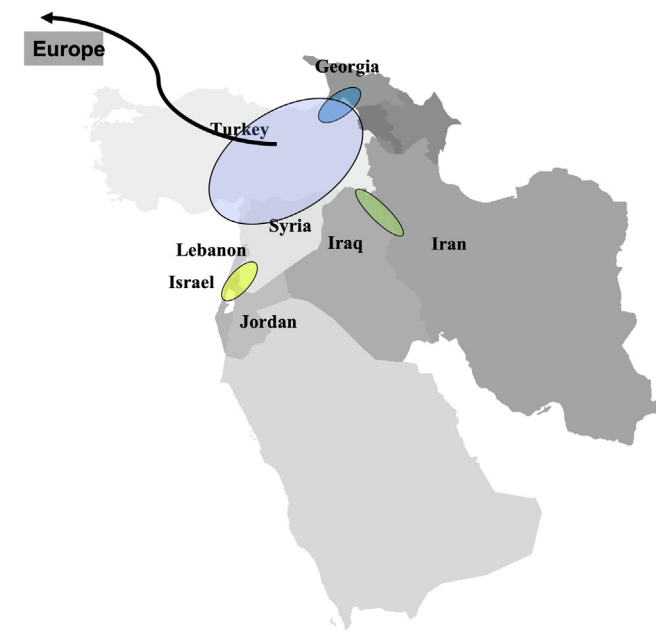


Fig. 6. Centre of rich genetic diversity for tetraploid wheat. Highlighted regions are not on scale, represents major cities. Blue circle: *T. turgidum* spp. *carthlicum* (Kars, Turkey; Tiflis, Georgia); Yellow circle: *T. turgidum* spp. *dicoccoides* (Israel, Jordan, Syria, Lebanon); Green circle: *T. timopheevii* spp. *armeniaceum* (Iraq (Arbil, Dahuk, As Sulaymaniyah)). Purple circle in Turkey may be considered as region of domestication and cultivation of domesticated emmer. Arrow represent the potential route of migration in the Europe. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

studies in wheat. Present core set accessions were already utilized to demonstrate independent domestication of *T. timopheevii* using *Brittle rachis 1* gene [72]. These genetically distinct tetraploid accessions, coupled with quality phenotyping, have the potential to develop elite wheat varieties.

Conclusion

In the present study, we explored the potential of GBS technology in data reduction while maintaining the significant genetic diversity of the species. NGS technology holds the potential for reclassification of species with similar morphological characters that are difficult to distinguish phenotypically. GBS based diversity analysis leads to identification of diversity rich pockets of tetraploid wheat in fertile crescent. Population structure supported the genetic differentiation of both wild and domesticated tetraploids. Wild germplasm showed more differentiation than domesticated accessions, indicating the availability of sufficient diversity for crop improvement. With reduced complexity, a core-set preserves the genetic diversity of the genebank collections and will aid in more robust characterization of wild germplasm. In the present study, we were able to preserve ~ 93 % of the allelic diversity of *T. timopheevii* and *T. turgidum* accessions. GBS proved to be an efficient technique of gene bank curation and can be effectively used to explore the rich diversity for species with huge genomes or higher ploidy not suitable for resequencing. With 10x reduction, and rich diversity of germplasm for disease resistance, the tetraploid core-set will cut expenses associated with storage and maintenance and will act as a resource to speed global breeding initiatives.

Compliance with ethics requirements

This article does not contain any studies with human or animal subjects.

Data availability

The datasets generated during the present study is publicly available in the Sequence Read Archive of NCBI under the Bioproject PRJNA862516.

Terms used

FST: F-statistics.
GBS: Genotyping by sequencing.
WGRC: Wheat genetics resource center.
GWAS: Genome-wide association study.
MAF: minor allele frequency.

CRediT authorship contribution statement

Inderjit S. Yadav: Investigation, Methodology, Data curation, Software, Writing – original draft. **Narinder Singh:** Data curation, Software. **Shuangye Wu:** Methodology. **Jon Raupp:** Methodology. **Duane L. Wilson:** Methodology. **Nidhi Rawat:** Data curation, Writing – review & editing. **Bikram S. Gill:** Conceptualization, Writing – review & editing. **Jesse Poland:** Conceptualization, Writing – review & editing. **Vijay K. Tiwari:** Conceptualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors acknowledge the University of Maryland super-computing resources (<http://hpcc.umd.edu>) made available for conducting the research reported in this paper. Authors acknowledge financial support provided by Indian Council of Agricultural Research (ICAR), New Delhi through Netaji Subhas - ICAR International Fellowships to Mr. Inderjit Singh Yadav. Authors also acknowledge support from WGRC I/UCRC which was partially funded by an NSF grant contract (IIP-1338897). VT and JP also acknowledge the support from USDA-NIFA grant 2022-67013-36362.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jare.2022.08.020>.

References

- [1] Dvorak J, Di Terlizzi P, Zhang HB, Resta P. The evolution of polyploid wheats: Identification of the A genome donor species. *Genome* 1993;36:21–31. doi: <https://doi.org/10.1139/g93-004>.
- [2] Dvorak J, McGuire PE, Cassidy B. Apparent sources of the A genomes of wheats inferred from polymorphism in abundance and restriction fragment length of repeated nucleotide sequences. *Genome* 1988;30:680–9. doi: <https://doi.org/10.1139/g88-115>.
- [3] Gornicki P, Zhu H, Wang J, Zhang Z, Gill BS, Li W. The chloroplast view of the evolution of polyploid wheat. *New Phytol* 2014;204:704–14.
- [4] Zohary D, Hopf M, Weiss E. Cereals. Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in Southwest Asia, Europe, and the Mediterranean Basin, 2012, p. 51.
- [5] Feldman M, Kislev ME. Domestication of emmer wheat and evolution of free-threshing tetraploid wheat. *Israel J Plant Sci* 2007;55:207–21. doi: <https://doi.org/10.1560/IIPS.55.3-4.207>.
- [6] Maccaferri M, Harris NS, Twardziok SO, Pasam RK, Gundlach H, Spannagl M, et al. Durum wheat genome highlights past domestication signatures and future improvement targets. *Nat Genet* 2019;51:885–95. doi: <https://doi.org/10.1038/s41588-019-0381-3>.

- [7] Nevo E, Gorham J, Beiles A. Variation for 22Na uptake in wild emmer wheat, *Triticum dicoccoides* in israel: Salt tolerance resources for wheat improvement. *J Exp Bot* 1992;43:511–8. doi: <https://doi.org/10.1093/jxb/43.4.511>.
- [8] Xie W, Nevo E. Wild emmer: Genetic resources, gene mapping and potential for wheat improvement. *Euphytica* 2008;164:603–14. doi: <https://doi.org/10.1007/s10681-008-9703-8>.
- [9] Dvořák J. *Triticum* Species (wheat). *Encyclopedia of Genetics* 2013;2060.
- [10] Mori N, Kondo Y, Ishii T, Kawahara T, Valkoun J, Nakamura C. Genetic diversity and origin of timopheevii wheat inferred by chloroplast DNA fingerprinting. *Breeding Sci* 2009;59:571–8.
- [11] Matsuoka Y. Evolution of polyploid *triticum* wheats under cultivation: The role of domestication, natural hybridization and allopolyploid speciation in their diversification. *Plant Cell Physiol* 2011;52:750–64. doi: <https://doi.org/10.1093/pcp/pcr018>.
- [12] Maestra B, Naranjo T. Structural chromosome differentiation between *Triticum timopheevii* and *T. turgidum* and *T. aestivum*. *Theor Appl Genet* 1999;98:744–50. doi: <https://doi.org/10.1007/s001220051130>.
- [13] 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. doi: <https://doi.org/10.1371/journal.pone.0032253>.
- [14] Torkamaneh D, Laroche J, Bastien M, Abed A, Fast-GBS BF. A new pipeline for the efficient and highly accurate calling of SNPs from genotyping-by-sequencing data. *BMC Bioinf* 2017;18. doi: <https://doi.org/10.1186/s12859-016-1431-9>.
- [15] 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. doi: <https://doi.org/10.1371/journal.pone.0090346>.
- [16] Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD, et al. Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel Insights from a Network-Based SNP Discovery Protocol. *PLoS Genet* 2013;9. doi: <https://doi.org/10.1371/journal.pgen.1003215>.
- [17] Czajkowska BI, Oliveira HR, Brown TA. A discriminatory test for the wheat B and G genomes reveals misclassified accessions of *Triticum timopheevii* and *Triticum turgidum*. *PLoS ONE* 2019;14:1–10. doi: <https://doi.org/10.1371/journal.pone.0215175>.
- [18] Peleg Z, Fahima T, Korol AB, Abbo S, Saranga Y. Genetic analysis of wheat domestication and evolution under domestication. *J Exp Bot* 2011;62:5051–61. doi: <https://doi.org/10.1093/jxb/err206>.
- [19] Abbo S, Pinhasi van-Oss R, Gopher A, Saranga Y, Ofner I, Peleg Z. Plant domestication versus crop evolution: A conceptual framework for cereals and grain legumes. *Trends in Plant Science* 2014;19:351–60. <https://doi.org/10.1016/j.tplants.2013.12.002>.
- [20] Tanno KI, Willcox G. How fast was wild wheat domesticated? *Science* 1979;206(311):1886. doi: <https://doi.org/10.1126/science.1124635>.
- [21] Harlan JR, Zohary D. Distribution of wild wheats and barley. *Science* 1966;153:1074–80. doi: <https://doi.org/10.1126/science.153.3740.1074>.
- [22] Pourkheirandish M, Hensel G, Kilian B, Senthil N, Chen G, Sameri M, et al. Evolution of the Grain Dispersal System in Barley. *Cell* 2015;162:527–39. doi: <https://doi.org/10.1016/j.cell.2015.07.002>.
- [23] Civián P, Brown TA. A novel mutation conferring the nonbrittle phenotype of cultivated barley. *New Phytol* 2017;214:468–72. doi: <https://doi.org/10.1111/nph.14377>.
- [24] Avni R, Nave M, Barad O, Baruch K, Twardziok SO, Gundlach H, et al. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science* 1979;207(97):93–7.
- [25] Pourkheirandish M, Dai F, Sakuma S, Kanamori H, Distelfeld A, Willcox G, et al. On the origin of the non-brittle rachis trait of domesticated einkorn wheat. *Front Plant Sci* 2018;8:1–10. doi: <https://doi.org/10.3389/fpls.2017.02031>.
- [26] Muterko A, Salina E. Origin and distribution of the VRN-A1 exon 4 and exon 7 haplotypes in domesticated wheat species. *Agronomy* 2018;8:1–14. doi: <https://doi.org/10.3390/agronomy8080156>.
- [27] Law CN, Worland AJ, Giorgi B. The genetic control of ear-emergence time by chromosomes 5A and 5D of wheat. *Heredity* (Edinb) 1976;36:49–58. doi: <https://doi.org/10.1038/hdv.1976.5>.
- [28] Shcherban AB, Strygina KV, Salina EA. VRN-1 gene- associated prerequisites of spring growth habit in wild tetraploid wheat *T. dicoccoides* and the diploid A genome species. *BMC Plant Biol* 2015;15:1–13. doi: <https://doi.org/10.1186/s12870-015-0473-x>.
- [29] Shitsukawa N, Ikari C, Shimada S, Kitagawa S, Sakamoto K, Saito H, et al. The einkorn wheat (*Triticum monococcum*) mutant, maintained vegetative phase, is caused by a deletion in the VRN1 gene. *Genes Genet Syst* 2007;82:167–70. doi: <https://doi.org/10.1266/ggs.82.167>.
- [30] Arzani A. Emmer (*Triticum turgidum* spp. *dicoccum*) Flour and Breads. In: Victor R. Preedy RRW and VBP, editor. *Flour and Breads and their Fortification in Health and Disease Prevention*, 2011, p. 69–78.
- [31] Poland JA, Rife TW. Genotyping-by-Sequencing for Plant Breeding and Genetics. *Plant Genome* 2012;5:92–102. doi: <https://doi.org/10.3835/plantgenome2012.05.0005>.
- [32] McCouch SR, McNally KL, Wang W, Hamilton RS. Genomics of gene banks: A case study in rice. *Am J Bot* 2012;99:407–23. doi: <https://doi.org/10.3732/ajb.1100385>.
- [33] Nei M. Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci U S A* 1973;70:3321–3. doi: <https://doi.org/10.1073/pnas.70.12.3321>.
- [34] Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics* 2011;27:2156–8. doi: <https://doi.org/10.1093/bioinformatics/btr330>.
- [35] Endelman JB. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *Plant Genome* 2011;4:250–5. doi: <https://doi.org/10.3835/plantgenome2011.08.0024>.
- [36] Paradis E, Claude J, Strimmer K. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 2004;20:289–90. doi: <https://doi.org/10.1093/bioinformatics/btg412>.
- [37] Malinsky M, Matschiner M, Svavdal H. Dsuite - Fast D-statistics and related admixture evidence from VCF files. *Mol Ecol Resour* 2021;21:584–95. doi: <https://doi.org/10.1111/1755-0998.13265>.
- [38] Raj A, Stephens M, Pritchard JK. FastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics* 2014;197:573–89. doi: <https://doi.org/10.1534/genetics.114.164350>.
- [39] Francis RM. pophelper: an R package and web app to analyse and visualize population structure. *Mol Ecol Resour* 2017;17:27–32. doi: <https://doi.org/10.1111/1755-0998.12509>.
- [40] PAUP (phylogenetic analysis using parsimony). *Encyclopedia of Genetics, Genomics, Proteomics and Informatics*, Dordrecht: Springer Netherlands; 2008, p. 1455. https://doi.org/10.1007/978-1-4020-6754-9_12413.
- [41] Wang J, Zhang Z. GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. *Genom. Proteom Bioinform* 2021;19:629–40. doi: <https://doi.org/10.1016/j.gpb.2021.08.005>.
- [42] Kim KW, Chung HK, Cho GT, Ma KH, Chandrabalan D, Gwag JG, et al. PowerCore: A program applying the advanced M strategy with a heuristic search for establishing core sets. *Bioinformatics* 2007;23:2155–62. doi: <https://doi.org/10.1093/bioinformatics/btm313>.
- [43] Singh N, Wu S, Tiwari V, Sehgal S, Raupp J, Wilson D, et al. Genomic analysis confirms population structure and identifies inter-lineage hybrids in *Aegilops tauschii*. *Front Plant Sci* 2019;10:1–13. doi: <https://doi.org/10.3389/fpls.2019.00009>.
- [44] Rouse MN, Olson EL, Gill BS, Pumphrey MO, Jin Y. Stem rust resistance in *Aegilops tauschii* germplasm. *Crop Sci* 2011;51:2074–8. doi: <https://doi.org/10.2135/cropsci2010.12.0719>.
- [45] Golovkina KA, Kondratenko EY, Blinov AG, Goncharov NP. Molecular characterization of vernalization loci VRN1 in wild and cultivated wheats. *BMC Plant Biol* 2010;10:168. doi: <https://doi.org/10.1186/1471-2229-10-168>.
- [46] McCartney CA, Somers DJ, McCallum BD, Thomas J, Humphreys DG, Menzies JG, et al. Microsatellite tagging of the leaf rust resistance gene Lr16 on wheat chromosome 2BSc. *Mol Breed* 2005;15:329–37. doi: <https://doi.org/10.1007/s11032-004-5948-7>.
- [47] Wang C, Yin G, Xia X, He Z, Zhang P, Yao Z, et al. Molecular mapping of a new temperature-sensitive gene LrZH22 for leaf rust resistance in Chinese wheat cultivar Zhoumai 22. *Mol Breed* 2016;36:1–10. doi: <https://doi.org/10.1007/s11032-016-0437-3>.
- [48] Özkan H, Willcox G, Graner A, Salamini F, Kilian B. Geographic distribution and domestication of wild emmer wheat (*Triticum dicoccoides*). *Genet Resour Crop Evol* 2011;58:11–53. doi: <https://doi.org/10.1007/s10722-010-9581-5>.
- [49] Badaeva ED, Boguslavsky RL, Badaev NS, Zelenin AV. Intraspecific chromosomal polymorphism of *Triticum araraticum* (Poaceae) detected by C-banding technique. *Plant Syst Evol* 1990;169:13–24. doi: <https://doi.org/10.1007/BF00935980>.
- [50] Mitrofanova O, Badaeva E, Salina EA. *T. timopheevii* and *T. zhukovskii*, the bread wheat cousins and their contribution to *T. aestivum* improvement. In: Bonjean Alain P., Angus William J. VJ, Maarten, editors. *The World Wheat Book: A History of Wheat Breeding*. Volume 3, France: Tec & Doc Lavoisier; 2016, p. 1167–228.
- [51] Wagenaar EB. Studies on the genome constitution of *triticum timopheevii* zhuk. li. The T. Timopheevii complex and its origin. *Evolution* (N Y) 1966;20:150–64. doi: <https://doi.org/10.1111/j.1558-5646.1966.tb03351.x>.
- [52] Singh N, Wu S, Raupp WJ, Sehgal S, Arora S, Tiwari V, et al. Efficient curation of genebanks using next generation sequencing reveals substantial duplication of germplasm accessions. *Sci Rep* 2019;9:1–10. doi: <https://doi.org/10.1038/s41598-018-37269-0>.
- [53] Zhu T, Wang L, Rodriguez JC, Deal KR, Avni R, Distelfeld A, et al. Improved Genome Sequence of Wild Emmer Wheat Zavitan with the Aid of Optical Maps 2019;9:619–24. <https://doi.org/10.1534/g3.118.200902>.
- [54] Dvorak J, Wang L, Zhu T, Jorgensen CM, Deal KR, Dai X, et al. Structural variation and rates of genome evolution in the grass family seen through comparison of sequences of genomes greatly differing in size. *Plant J* 2018;95:487–503. doi: <https://doi.org/10.1111/tpj.13964>.
- [55] He F, Pasam R, Shi F, Kant S, Keeble-gagnere G, Kay P, et al. Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. *Nat Genet* 2019;51:896–904. doi: <https://doi.org/10.1038/s41588-019-0382-2>.
- [56] Walkowiak S, Gao L, Monat C, Haberer G, Delorean E, Thambugala D, et al. Multiple wheat genomes reveal global variation in modern breeding. *Nature* 2020;588:277–83. doi: <https://doi.org/10.1038/s41586-020-2961-x>.
- [57] Badaeva ED, Ruban AS, Zoshchuk SA, Kilian B. Molecular cytogenetic characterization of *Triticum timopheevii* chromosomes provides new insight on genome evolution of *T. zhukovskii*. *Plant Syst Evol* 2016;302:943–56. doi: <https://doi.org/10.1007/s00606-016-1309-3>.
- [58] Padmanaban S, Zhang P, Sutherland MW, Martin A. Association between presence of *Triticum timopheevii* introgression and D-genome retention in hexaploid/tetraploid wheat crosses. *Mol Breed* 2018;38:1–8.

- [59] Percival J. *The wheat plant : a monograph*. London: London: Duckworth; 1921.
- [60] Van SM. Wild wheats: a monograph of *Aegilops* L. and *Amblyopyrum* (Jaub. & Spach) Eig (Poaceae). Wageningen Agriculture University Papers; 1994.
- [61] Rodríguez S, Maestra B, Perera E, Díez M, Naranjo T. Pairing affinities of the B- and G-genome chromosomes of polyploid wheats with those of *Aegilops speltoides*. *Genome* 2000;43:814–9. doi: <https://doi.org/10.1139/g00-055>.
- [62] Hartl DL, Clark AG. *Principles of Population Genetics*. 3rd ed. Sinauer Associates; 1997.
- [63] Frankham R, Ballou JD, Briscoe DA, McInnes KH. *Introduction to Conservation Genetics*. Cambridge University Press; 2002.
- [64] Badaeva ED, Konovalov FA, Knüpfner H, Fricano A, Ruban AS, Kehel Z, et al. Genetic diversity, distribution and domestication history of the neglected GGAtAt gene pool of wheat. *Theor Appl Genet* 2022;135:755–76. doi: <https://doi.org/10.1007/s00122-021-03912-0>.
- [65] Oliveira HR, Jacocks L, Czajkowska BI, Kennedy SL, Brown TA. Multiregional origins of the domesticated tetraploid wheats. *PLoS ONE* 2020;15:1–20. doi: <https://doi.org/10.1371/journal.pone.0227148>.
- [66] Laidò G, Mangini G, Taranto F, Gadaleta A, Blanco A, Cattivelli L, et al. Genetic Diversity and Population Structure of Tetraploid Wheats (*Triticum turgidum* L.) Estimated by SSR, DArT and Pedigree Data. *PLoS ONE* 2013;8:e67280. doi: <https://doi.org/10.1371/journal.pone.0067280>.
- [67] Oliveira HR, Hagenblad J, Leino MW, Leigh FJ, Lister DL, Penã-Chocarro L, et al. Wheat in the Mediterranean revisited - tetraploid wheat landraces assessed with elite bread wheat Single Nucleotide Polymorphism markers. *BMC Genet* 2014;15:1–15. doi: <https://doi.org/10.1186/1471-2156-15-54>.
- [68] Riefolo C, Ficco D, Cattivelli L, Vita P. Genetic diversity of gluten proteins in *T. turgidum* L. *Cereal Res Commun* 2011;39:405–14. doi: <https://doi.org/10.1556/CRC.39.2011.3.10>.
- [69] Kuckuck H. On the origin of *triticum carthlicum neyski* (*triticum persicum* Vav.). *Wheat Inf Serv* 1979:1–5.
- [70] Hyun DY, Sebastin R, Lee KJ, Lee GA, Shin MJ, Kim SH, et al. Genotyping-by-Sequencing Derived Single Nucleotide Polymorphisms Provide the First Well-Resolved Phylogeny for the Genus *Triticum* (Poaceae). *Front Plant Sci* 2020;11:1–15. doi: <https://doi.org/10.3389/fpls.2020.00688>.
- [71] Shcherban AB, Salina EA. Evolution of VRN-1 homoeologous loci in allopolyploids of *Triticum* and their diploid precursors. *BMC Plant Biol* 2017;17. doi: <https://doi.org/10.1186/s12870-017-1129-9>.
- [72] Nave M, Tas M, Raupp J, Tiwari VK, Ozkan H, Poland J, et al. The independent domestication of timopheev's wheat: Insights from haplotype analysis of the brittle rachis 1 (btr1-a) gene. *Genes (Basel)* 2021;12:1–10. doi: <https://doi.org/10.3390/genes12030338>.
- [73] Adhikari L, Raupp J, Wu S, Wilson D, Evers B, Koo DH, et al. Genetic characterization and curation of diploid A-genome wheat species. *Plant Physiol* 2022;188:2101–14. doi: <https://doi.org/10.1093/plphys/kiac006>.
- [74] 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;114:1265–75. doi: <https://doi.org/10.1007/s00122-007-0517-1>.
- [75] Hao CY, Dong YC, Wang LF, You GX, Zhang HN, Ge HM, et al. Genetic diversity and construction of core collection in Chinese wheat genetic resources. *Chin Sci Bull* 2008;53:1518–26. doi: <https://doi.org/10.1007/s11434-008-0212-x>.
- [76] Pascual L, Fernández M, Aparicio N, López-Fernández M, Fité R, Giraldo P, et al. Development of a multipurpose core collection of bread wheat based on high-throughput genotyping data. *Agronomy* 2020;10. doi: <https://doi.org/10.3390/agronomy10040534>.
- [77] Hao CY, Zhang XY, Wang LF, Dong YS, Shang XW, Jia JZ. Genetic diversity and core collection evaluations in common wheat germplasm from the Northwestern Spring Wheat Region in China. *Mol Breed* 2006;17:69–77. doi: <https://doi.org/10.1007/s11032-005-2453-6>.
- [78] Ruiz M, Giraldo P, Royo C, Carrillo JM. Creation and validation of the spanish durum wheat core collection. *Crop Sci* 2013;53:2530–7. doi: <https://doi.org/10.2135/cropsci2013.04.0238>.
- [79] Sidhu JS, Ramakrishnan SM, Ali S, Bernardo A, Bai G, Abdullah S, et al. Assessing the genetic diversity and characterizing genomic regions conferring Tan Spot resistance in cultivated rye. *PLoS ONE* 2019;14. doi: <https://doi.org/10.1371/journal.pone.0214519>.
- [80] Cuevas HE, Rosa-Valentin G, Hayes CM, Rooney WL, Hoffmann L. Genomic characterization of a core set of the USDA-NPGS Ethiopian sorghum germplasm collection: Implications for germplasm conservation, evaluation, and utilization in crop improvement. *BMC Genomics* 2017;18:108. doi: <https://doi.org/10.1186/s12864-016-3475-7>.
- [81] Wang X, Ando K, Wu S, Reddy UK, Tamang P, Bao K, et al. Genetic characterization of melon accessions in the U.S. National Plant Germplasm System and construction of a melon core collection. *Molecular. Horticulture* 2021;11. doi: <https://doi.org/10.1186/s43897-021-00014-9>.
- [82] Nguyen NN, Kim M, Jung JK, Shim EJ, Chung SM, Park Y, et al. Genome-wide SNP discovery and core marker sets for assessment of genetic variations in cultivated pumpkin (*Cucurbita* spp.). *Hortic Res* 2020;7:108. doi: <https://doi.org/10.1038/s41438-020-00342-9>.
- [83] Tripodi P, Rabanus-Wallace MT, Barchi L, Kale S, Esposito S, Acquadro A, et al. Global range expansion history of pepper (*Capsicum* spp.) revealed by over 10,000 genebank accessions. *Proceedings of the National Academy of Sciences* 2021;118:e2104315118. <https://doi.org/10.1073/pnas.2104315118>.
- [84] Jorgensen JH, Jensen CJ. Gene Pm6 for resistance to powdery mildew in wheat. *Euphytica* 1973;22:423.
- [85] Järve K, Peusha HO, Tsybalova J, Tamm S, Devos KM, Enno TM. Chromosomal location of a *Triticum timopheevii*-derived powdery mildew resistance gene transferred to common wheat. *Genome* 2000;43:377–81.
- [86] Srnčić G, Murphy JP, Lyster JH, Leath S, Marshall DS. Inheritance and chromosomal assignment of powdery mildew resistance genes in two winter wheat germplasm lines. *Crop Sci* 2005;45:1578–86. doi: <https://doi.org/10.2135/cropsci2004.0530>.
- [87] Maxwell JJ, Lyster JH, Cowger C, Marshall D, Brown-Guedira G, Murphy JP. MIA12: A *Triticum timopheevii*-derived powdery mildew resistance gene in common wheat on chromosome 7AL. *Theor Appl Genet* 2009;119:1489–95. doi: <https://doi.org/10.1007/s00122-009-1150-y>.
- [88] Brown-Guedira GL, Singh S, Fritz AK. Performance and Mapping of Leaf Rust Resistance Transferred to Wheat from *Triticum timopheevii* subsp. *armeniaceum*. *Phytopathology* 2003;93:784–9. doi: <https://doi.org/10.1094/PHYTO.2003.93.7.784>.
- [89] Leonova IN, Budashkina EB, Flath K, Weidner A, Börner A, Röder MS. Microsatellite mapping of a leaf rust resistance gene transferred to common wheat from *Triticum timopheevii*. *Cereal Res Commun* 2010;38:211–9. doi: <https://doi.org/10.1556/CRC.38.2010.2.7>.
- [90] Dyck PL. Transfer of a gene for stem rust resistance from *Triticum araraticum* to hexaploid wheat1. *Genome* 1992;35:788–92.
- [91] Malihipour A, Gilbert J, Fedak G, Brülé-Babel A, Cao W. Mapping the a genome for *qtl* conditioning resistance to Fusarium head blight in a wheat population with *Triticum timopheevii* background. *Plant Dis* 2017;101:11–9. doi: <https://doi.org/10.1094/PDIS-02-16-0144-RE>.
- [92] Malihipour A, Gilbert J, Fedak G, Brülé-Babel A, Cao W. Characterization of agronomic traits in a population of wheat derived from *Triticum timopheevii* and their association with Fusarium head blight. *Eur J Plant Pathol* 2016;144:31–43. doi: <https://doi.org/10.1007/s10658-015-0744-2>.
- [93] Blanco A, Gadaleta A, Cenci A, Carluccio A v., Abdelbacki AMM, Simeone R. Molecular mapping of the novel powdery mildew resistance gene Pm36 introgressed from *Triticum turgidum* var. *dicoccoides* in durum wheat. *Theoretical and Applied Genetics* 2008;117:135–42. <https://doi.org/10.1007/s00122-008-0760-0>.
- [94] Li G, Fang T, Zhang H, Xie C, Li H, Yang T, et al. Molecular identification of a new powdery mildew resistance gene Pm41 on chromosome 3BL derived from wild emmer (*Triticum turgidum* var. *dicoccoides*). *Theor Appl Genet* 2009;119:531–9. doi: <https://doi.org/10.1007/s00122-009-1061-y>.
- [95] Ullah N, Asif M, Badshah H, Bashir T, Mumtaz AS. Introgression lines obtained from the cross between *triticum aestivum* and *triticum turgidum* (Durum wheat) as a source of leaf and stripe (yellow) rust resistance geneswe. *Turkish J Biol* 2016;40:547–53. doi: <https://doi.org/10.3906/ibiy-1501-99>.