© NIAB 2015
ISSN 1479-2621
Plant Genetic Resources: Characterization and Utilization (2017) **15**(1); 29–35
doi:10.1017/S1479262115000325

Phylogenetic analysis of wheat cultivars in Kazakhstan based on the wheat 90 K single nucleotide polymorphism array

Yerlan Turuspekov[1]*, Joerg Plieske[2], Martin Ganal[2], Eduard Akhunov[3] and Saule Abugalieva[1]

[1]*Institute of Plant Biology and Biotechnology, Almaty 050040, Kazakhstan,* [2]*TraitGenetics Gmbh, 06466 Gatersleben, Germany and* [3]*Kansas State University, Manhattan, KS, USA*

Received 8 April 2015; Revised 25 May 2015; Accepted 19 June 2015 – First published online 30 July 2015

Abstract

The recent introduction of Illumina single nucleotide polymorphism (SNP) arrays is an important step towards comprehensive genome-wide studies of genetic diversity in wheat. In this study, 90 cultivars of hexaploid spring wheat growing in Kazakhstan were genotyped using the high-density wheat 90 K Illumina SNP array. The analysis allowed the identification of 30,288 polymorphic SNPs. A subset of 3541 high-quality SNPs were used for a comparison of 690 wheat accessions representing landraces and varieties, including those from Asia, Australia, Canada, Europe, Kazakhstan, USA and other parts of the world. Phylogenetic analysis showed a clear separation of wheat cultivars according to their geographic origin. In the phylogenetic tree, accessions from Kazakhstan and the USA formed two neighbouring clusters with a common node, and they were distinct from accessions from other regions of the world, including Europe. The results provide important new insights into the genetic relationships between diverse wheat accessions.

Keywords: genetic diversity; genotyping; Illumina array; wheat

Introduction

Hexaploid wheat (*Triticum aestivum* L.) is one of the three major cereal crops of the world and is important in Kazakhstan agriculture. Spring types of wheat in Kazakhstan grow annually on more than 13 million hectares, and wheat is an important product for consumption inside the country. Owing to environmental stresses, the average grain yield of spring wheat in Kazakhstan is only 1.1 ton per hectare (FAO, 2013). However, the grain quality is high and, therefore, it is an important commodity for export to neighbouring countries (Martynov *et al.*, 2005). Traditionally, since Kazakhstan was a part of the former Soviet Union, most of wheat cultivars were developed in collaboration with Russian breeders and using Russian wheat genetic resources (Martynov *et al.*, 2005). According to Udachin and Shahmedov (1984), the entire history of wheat growth in Kazakhstan can be divided into three stages. The first stage covers the period of 18th and 19th centuries, before the establishment of Turkestan district governance by Russia in 1867. The second stage is considered to be relatively short, starting in 1867 and ending at the time of Russian revolution in 1917. The third stage is framed by the time when Kazakhstan was a part of Soviet Union (Udachin and Shahmedov, 1984). As for the first stage, there is not much information that can be extracted from the available literature sources. Perhaps, due to the nomadic way of life of Kazakhs, there was limited practical interest to settled crop agriculture. Some indication on positive changes can be attributed to the second stage when first reports by Flaksberger (1909, 1910) provided

* Corresponding author. E-mail: yerlant@yahoo.com

https://doi.org/10.1017/S1479262115000325 Published online by Cambridge University Press

detailed description of several landraces introduced from other Central Asian republics in southern and south-eastern parts of Kazakhstan. Introduction to real wheat breeding is basically associated with the third stage and well documented at the Vavilov All-Russian Research Institute of Plant Production (Udachin and Shahmedov, 1984). Currently, there are 601 accessions, including 34 landraces, collected in Kazakhstan from 1920 to 2000 that registered in the Focused Identification of Germplasm Strategy Project (http://www.figs.icarda.net). During Soviet time, right after the declaration of virgin lands campaign in 1957, the wheat production area was significantly expanded and shifted to the northern regions of Kazakhstan. At that time, main wheat-related breeding programmes were conducted in close cooperation with the Russian breeders (Martynov *et al.*, 2005). This trend of collaboration is still in place as the two countries share their expertise and genetic resources based on bilateral projects and international activities under the CIMMYT umbrella (Morgounov *et al.*, 2007; Gomez-Becerra *et al.*, 2010).

The genetic diversity of wheat cultivars growing in Kazakhstan has been previously studied based on microsatellite simple sequence repeat (SSR) markers (Khlestkina *et al.*, 2004; Abugalieva *et al.*, 2012). The results indicated that genetic diversity in wheat in Kazakhstan is very high (Abugalieva *et al.*, 2012), which is not a surprise as wheat grows in many different ecological niches of this country, the ninth largest country in the world. However, in order to efficiently study the phylogeny of local accessions in comparison with other wheat samples from around the world, it is necessary to use better and more powerful tools. These are now available with the successful introduction of a 9 K (Cavanagh *et al.*, 2013) and a 90 K (Wang *et al.*, 2014) Illumina SNP array. The first report of an assessment of genetic variation in hexaploid wheat accessions from Kazakhstan was published based on the 9 K SNP array (Shavrukov *et al.*, 2014). The drawback of that work was the limited number of accessions (ten cultivars), which could lead to misleading conclusions. Therefore, the purpose of this study was to assess genetic diversity and phylogenetic relationship of a larger set of accessions from Kazakhstan in comparison with 600 wheat cultivars and landraces from around the world based on the use of the 90 K SNP Illumina array.

Materials and methods

The spring wheat panel studied herein consisted of 90 released and prospective cultivars from Kazakhstan (Table S1, available online). Currently, 40 cultivars from Kazakhstan and 21 cultivars from Russia in this genetic panel have been registered through the State Seed Trials Commission of the Republic of Kazakhstan (2013), and grown officially in Kazakhstan. The panel also included 19 and 10 prospective cultivars developed in Kazakhstan and Russia, respectively (Table S1, available online). DNA samples were extracted and purified from single seeds of individual cultivars using commercial kits (Qiagene, CA, USA). The DNA concentration for each sample was adjusted to 50 ng/μl. Accessions were genotyped using the wheat 90 K Illumina iSelect SNP array (Wang *et al.*, 2014). The Illumina Infinium procedure was performed according to the manufacturer's protocol. SNP genotype analysis was carried out using an optimized cluster file for hexaploid wheat using the diploid version of the Illumina Genome Studio software (GS V2011.1). Data from 10,000 polymorphic SNPs for 600 accessions of a world wheat collection were used for comparison. These worldwide wheat accessions were listed in Table S14 in Wang *et al.* (2014). These hexaploid wheat cultivars represented Asia, Australia, Canada, Europe and USA, and landraces from around the world. A total of 3541 overlapping, without missing reads, broadly distributed over the genome and mostly haplotype-specific SNPs, of the two data sets (Kazakhstan and world wheat collection), were selected for further phylogenetic analyses of 690 wheat accessions. Phylogenetic trees were constructed by the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) method using the NTSYS software package (Rohlf, 2008). Population genetic analysis and principal coordinate analysis were performed using GenAlEx 6.5 (Peakall and Smouse, 2006; Peakall and Smouse, 2012).

Results

Characterization of wheat accessions from Kazakhstan using the 90 K SNP array

The 90 spring wheat accessions from Kazakhstan and Russia were genotyped using the 90 K SNP iSelect array containing 81,587 SNPs. The genotyping allowed the generation of 66,223 scorable SNPs (81.2%) with 83.5% variants being transitions and 16.6% transversions. *Blastn* search allowed the identification of 55,165 SNP hits in the chromosome survey sequencing project of IWGSC (http://www.wheatgenome.org) allowing the chromosomal location of the markers. The numbers of SNPs per chromosome ranged from 1419 in chromosome 4D to 4338 in chromosome 2B. In total, 30,288 (37%) of variants were polymorphic, which is comparable with the results obtained by Wang *et al.* (2014), where the authors identified 35,684 (44%) assays for nearly 2500 accessions.

The phylogenetic tree analysis for the 90 wheat accessions from Kazakhstan and Russia was performed using

https://doi.org/10.1017/S1479262115000325 Published online by Cambridge University Press

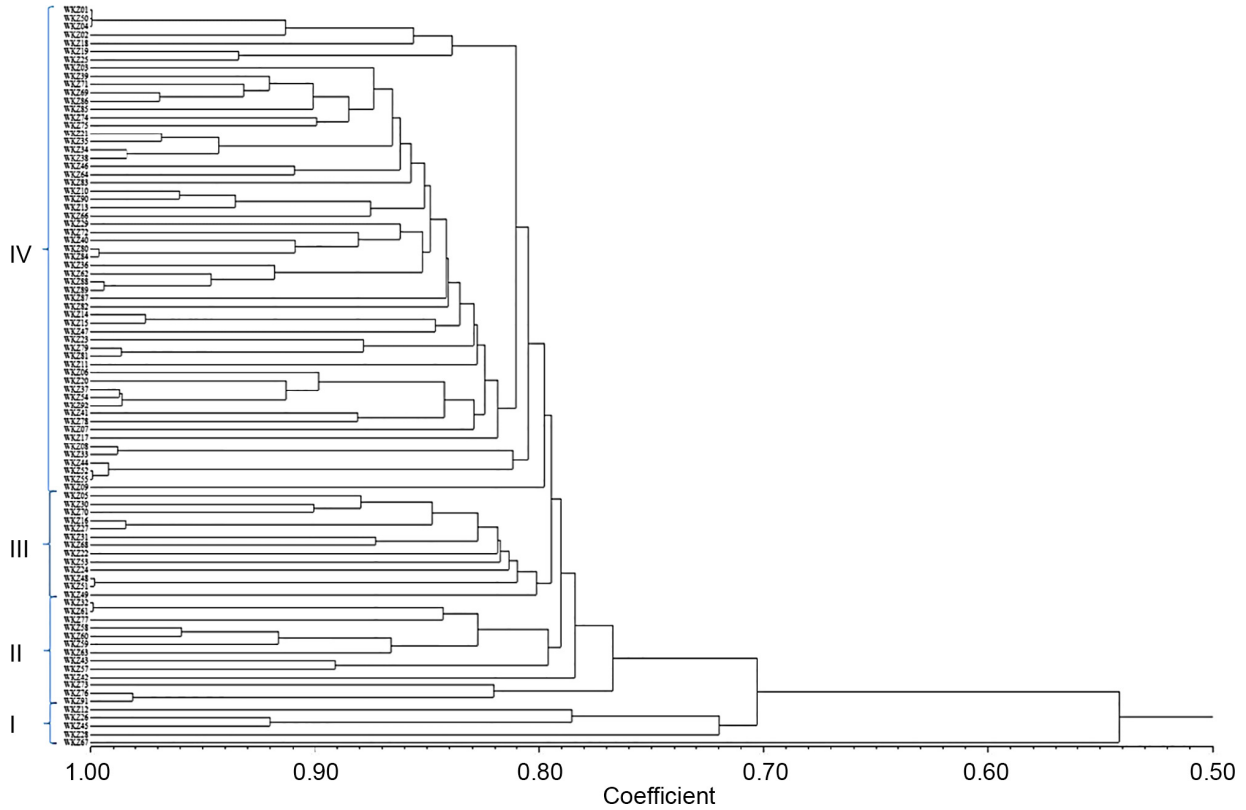


Fig. 1. UPGMA phylogenetic tree of 90 spring wheat accessions growing in Kazakhstan based on 30 K polymorphic SNP markers. Scale is Dice coefficient. Roman numbers are suggested clusters.

the UPGMA method (Fig. 1). The accessions clustered into four major groups and were clearly separated from the out-group accession WKZ67 (line name, Samad). WKZ67 is a cultivar, where one of the parents (Kazakhstanskaya 17) has *Triticum carthlicum* and *Agropyron glaucum* in its pedigree (Martynov *et al.*, 2006). In general, all the clusters were formed by accessions growing in different regions of Kazakhstan. As accessions growing in the southern and south-eastern regions were missing in clusters II and IV, and accessions growing in West Kazakhstan were restricted only to clusters III and IV (Table S1, available online and Fig. 1), this suggests that the clusters are, to some extent, related to the geographical cultivation areas for wheat in Kazakhstan. The association of genetic variation with geography was confirmed by principal coordinates analysis (PCoA) for the 61 officially released cultivars from Kazakhstan and Russia that split them into six geographical regions within Kazakhstan (Table S1, available online and Fig. 2) in accordance to the official list of the State Seed Trial

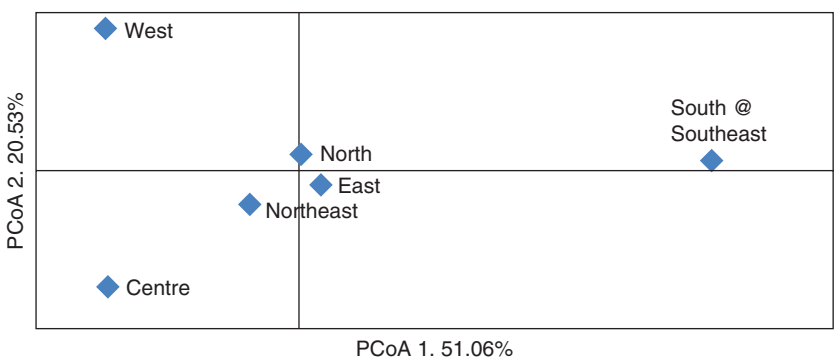


Fig. 2. Principal coordinate analysis (PCoA) for 61 officially released wheat cultivars from Kazakhstan and Russia separated to six groups according to their places of growth in Kazakhstan (regions are given in Table S1, available online).

https://doi.org/10.1017/S1479262115000325 Published online by Cambridge University Press

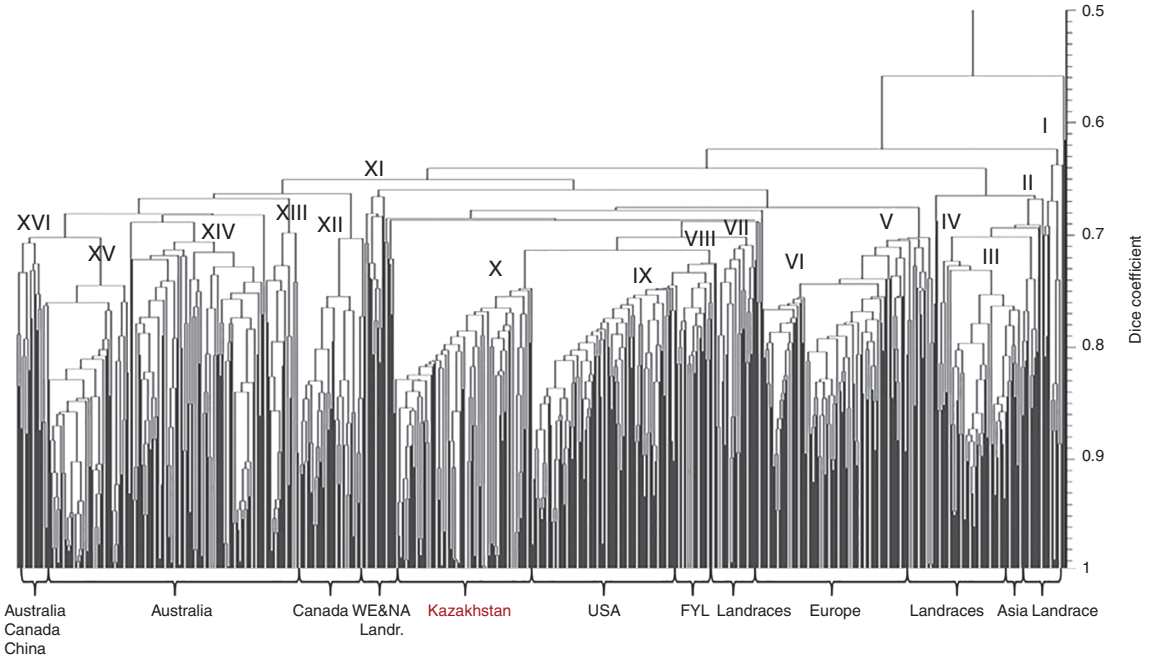


Fig. 3. The phylogenetic tree of 690 hexaploid wheat cultivars and landraces constructed based on 3541 polymorphic SNPs from the 90 K SNP array and UPGMA. WE & NA Landr., West European and North African landraces; FYL, landraces from former Yugoslavia republics. Roman numbers are suggested clusters.

Commission of the Republic of Kazakhstan. The graph reveals that genotypes growing in the northern and eastern regions have the highest similarity, and genotypes from the southern, south-eastern, central and western parts of the country are genetically distinct from each other, and from accessions growing in the northern and eastern regions of the Kazakhstan.

Phylogenetic analysis of wheat populations and landraces based on 3541 SNPs

SNP genotyping data for the 90 cultivars growing in Kazakhstan were aligned with data on 10,000 polymorphic SNPs for 600 wheat cultivars and landraces from many regions of the world reported in Wang *et al.* (2014). From these 10,000 SNPs, 3541 high-quality polymorphic markers with non-missing data for both sets of samples were selected for further analyses. An UPGMA phylogenetic tree of the 690 lines is provided in Fig. 3. The accessions grouped into 16 clusters, where most of cultivars showed a clear geographic origin, except cluster XVI that combines cultivars from Australia, Canada and China. It was found that the cluster of accessions from Kazakhstan (cluster X) has a common node with a cluster of samples from the USA (MidWest accessions, cluster IX), indicating that these two groups of samples are closer to each other in comparison with accessions from Europe (clusters V and VI, Fig. 3). The dendrogram shows that cluster VIII (next to the cluster of USA accessions) mostly consists of

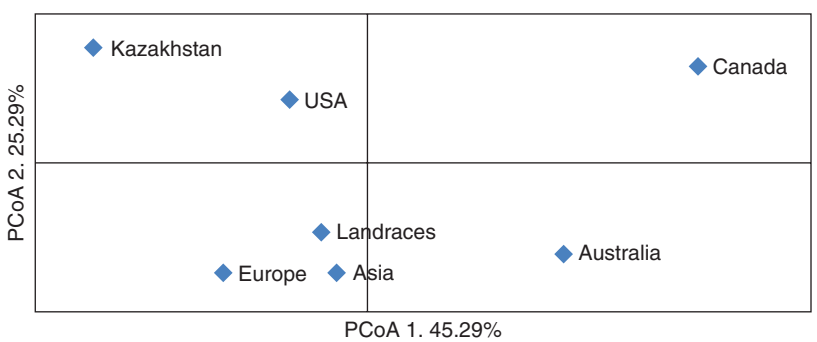


Fig. 4. Principal coordinate analysis (PCoA) of wheat populations and landraces based on 3541 SNPs.

https://doi.org/10.1017/S1479262115000325 Published online by Cambridge University Press

Table 1. Mean genetic diversity indexes in wheat populations and landraces based on 3541 SNPs

Populations	Asia	Australia	Canada	Europe	Kazakhstan	Landraces	USA
N	29	181	46	120	90	124	94
I	0.48	0.46	0.39	0.42	0.37	0.48	0.36
Uh	0.33	0.30	0.26	0.28	0.24	0.32	0.23

SNPs, single nucleotide polymorphisms; *N*, number of accessions; *I*, Shannon index; Uh, unbiased Nei's diversity index.

landraces collected from countries of the former Yugoslavia (Bosnia and Herzegovina, Croatia, Macedonia, Montenegro and Serbia).

Wheat samples from Kazakhstan formed one compact cluster (X) with the exception of five accessions, where one of those accessions was linked to cluster V, three accessions were associated with cluster III, and line WKZ67 was positioned as an outlier for the entire collection (the accession to the far right in Fig. 3). The results of the PCoA (Fig. 4) largely confirmed the trend visible in the phylogenetic tree. Nei's genetic diversity index suggested that the most variable group were the landraces (Table 1).

Discussion

Since the 90 K SNP array was developed using SNPs from selected groups of wheat accessions including durum wheat and *Aegilops tauschii*, one of the main questions for the implications of the array was whether it would be effectively applicable for poorly studied groups of samples from around the world. The work here suggests that genome-wide genotyping of spring wheat accessions from Kazakhstan using the 90 K Illumina SNP array was highly efficient, and the success rate was comparable with the results reported in the original work (Wang *et al*., 2014). One of the advantages of this study was not only the use of a high throughput SNP genotyping array, but also the diverse origins of the wheat accessions, with a large number of samples from many different regions. This eliminates possible misleading results from analysing a limited number of samples. For instance, Shavrukov *et al*. (2014) studied only ten accessions from Kazakhstan and 37 wheat accessions from all over the world, and found that the main group of wheats from Kazakhstan were clustered together with Chinese accession PI 447382, which is grouped in cluster XVI with Australian and Canadian accessions in this work (Fig. 3). Based on our studies, PI 447382 is genetically distinct from the Kazakhstan group of accessions investigated herein. The other conclusion of that report was the identification of the so-called 'unique' SNP markers for wheat accessions from Kazakhstan, which cannot be confirmed after the analysis of a larger number of samples from other regions of Kazakhstan.

The use of high throughput SNP genotyping in this work gives the discrimination of wheat samples according to their region of cultivation. The dendrogram in Fig. 3 is showing that most of the accessions from Australia, Canada, Europe, USA and Kazakhstan are grouped in well-structured separate clusters. Also, clear differences are visible between European, Kazakhstan and USA samples on the one hand, and Australian and Canadian samples on the other hand (Fig. 3). The samples in the Asian group were represented by countries stretching from China to the Middle East, and, therefore, understandably did not form a single cluster. 'Within-region' clustering of wheat accessions from Kazakhstan also underline the value of SNP array genotyping. For instance, the PCoA is suggesting a clear discrimination of groups growing in different regions of Kazakhstan (Fig. 2).

Another interesting outcome of this research is the finding of a close relationship between Kazakhstan and USA (MidWest part of the USA) groups of wheat varieties. In fact, the phylogenetic tree (Fig. 3) is suggesting that wheat growing in Kazakhstan is genetically closer to wheat from the USA than European samples. The explanation for such a close relationship between wheats located in different continents may relate to the historic introduction of the Turkey Red type of wheat from Asia into the USA. One of the key stages of wheat improvement in the USA was the successful introduction of Turkey Red Wheats by Russian Mennonites (ethnically German) starting from 1874 (Quisenberry and Reitz, 1974; Olmstead and Rhode, 2002). The success of Turkey Red Wheat resulted in the importation of wheat varieties into the USA from Russia up to the year 1900, including famous cultivars such as Crimean and Kharkoff (Paulsen and Shroyer, 2008). According to Quisenberry and Reitz (1974), by 1944, the Turkey type of wheat cultivars were grown in 33 states on 30% of the wheat area of the USA. Therefore, it is not surprising that Turkey type cultivars are founders of many modern USA wheat cultivars (Olmstead and Rhode, 2002). On the other hand, the development of wheat cultivars in Kazakhstan

https://doi.org/10.1017/S1479262115000325 Published online by Cambridge University Press

is a result of a close cooperation between breeders from Kazakhstan and the Russian Federation, because both countries were part of the former USSR, and actively shared available genetic resources. The current list of officially released cultivars in Kazakhstan includes 21 cultivars developed in Russia (Table S1, available online) and strongly supports this hypothesis. The phylogenetic tree (Fig. 3) is also suggesting connections between USA cultivars and landraces from the former Yugoslavia. A reasonable explanation for this relationship can be the introduction of USA and former USSR wheat germplasm into the former Yugoslavia breeding activities after World War II (Jost and Cox, 1989).

Apart from phylogeny studies, this work allowed the identification of 30,288 polymorphic SNP markers amongst 90 accessions from Kazakhstan and Russia, which is strong confirmation of the high value and efficiency of the 90 K Illumina SNP array. These results can also contribute significantly to ongoing research related to the identification of quantitative trait loci for important agronomic traits via genome-wide association mapping studies.

Supplementary material

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S1479262115000325

Acknowledgements

This work was funded by the ADAPTAWHEAT project supported by seventh framework programme of the European Union, and the 1784/GF4 project supported by the Ministry of Education and Sciences of the Republic of Kazakhstan. The authors are grateful to John Snape (John Innes Centre, Norwich, UK) for critical reading of the manuscript.

References

Abugalieva SI, Volkova LA, Ermekbaev KA and Turuspekov EK (2012) Genotyping of commercial cultivars of spring bread wheat from Kazakhstan by microsatellite DNA-markers. Biotechnology. *Theory and Practice* 2: 35–45. doi:http://dx.doi.org/10.11134/btp.2.2012.4.

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, Lopez da Silva M, Bockelman H, Talbert L, Anderson JA, Dreisigacker S, Baenziger S, Carter A, Korzun V, Morrel PL, Dubcovsky J, Morell MK, Sorrells ME, Hayden M and 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.

FAO (2013) FAOSTAT. Available at http://faostat3.fao.org/home/E (accessed 12 December 2014).

Flaksberger K (1909) In relation to variability of wheat in agricultural exhibitions in Poltava, Rostov-na-Donu, and Tashkent in 1909. *Proceedings of the Bureau of Applied Botany* 2: 671–683 (in Russian).

Flaksberger K (1910) The ariability of wheat in Semirechye oblast. *Proceedings of the Bureau of Applied Botany* 3: 62–165 (in Russian).

Gomez-Becerra HF, Abugalieva A, Morgounov A, Abdullaev K, Bekenova L, Yessimbekova M, Sereda G, Shpigun S, Tsygankov V, Zelenskiy Y, Pena RH and Cakmak I (2010) Phenotypic correlations, G × E interactions and broad sense heritability analysis of grain and flour quality characteristics in high latitude spring bread wheats from Kazakhstan and Siberia. *Euphytica* 171: 23–38.

Jost M and Cox TS (1989) History of wheat breeding in Yugoslavia. *Podravka, Koprivnica* 7: 1–18.

Khlestkina EK, Roder MS, Efremova TT, Borner A and Shumny VK (2004) The genetic diversity of old and modern Siberian varieties of common spring wheat as determined by microsatellite markers. *Plant Breeding* 123: 122–127.

Martynov SP, Dobrovotvorskaya TV, Morvounov AI, Urazaliev RA and Absattarova AS (2005) Genealogical analysis of diversity of spring bread wheat cultivars released in Kazakhstan from 1929–2004. *Acta Agronomica Hungary* 53: 261–272.

Martynov SP, Dobrotvorskaya TV, Hon I and Faberova I (2006) Wheat Pedigree and Identified Alleles of Genes on Line. Available at http://genbank.vurv.cz/wheat/pedigree/ (accessed 10 November 2014).

Morgounov A, Gomez-Becerra HG, Abugalieva A, Dzhunusova M, Yessimbekova M, Miminjanov H, Zelenskiy Y, Ozturk L and Cakmak I (2007) Iron and zinc grain density in common wheat grown in Central Asia. *Euphytica* 155: 193–203.

Olmstead A and Rhode P (2002) The Red Queen and the Hard Reds: productivity growth in American wheat, 1800–1940. *The Journal of Economic History* 62: 929–966.

Paulsen GM and Shroyer JP (2008) Early history of wheat improvement in the Great Plains. Celebrate the Centennial. *Agronomy Journal* 100 (Suppl.): S70–S78.

Peakall R and Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288–295.

Peakall R and Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics* 28: 2537–2539.

Quisenberry K and Reitz L (1974) Turkey wheat: the cornerstone of an empire. *Agricultural History* 48: 98–110.

Rohlf FJ (2008) *NTSYSpc: Numerical Taxonomy System, Ver. 2.11X*. Setauket, NY: Exeter Publishing, Ltd.

Shavrukov Y, Sucheski R, Eliby S, Abugalieva A, Kenenbayev S and 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: 258–271.

State Register of Breeding Achievements Allowed to Grow in the Republic of Kazakhstan. Plant Cultivars, (2013) *Official issue 2013 by the State Seed Trial Commission of the Ministry of Agriculture of the Republic of Kazakhstan*. Kazakhstan: Almaty, 118 pp. (in Russian).

https://doi.org/10.1017/S1479262115000325 Published online by Cambridge University Press

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 and 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.

Udachin R and Shahmedov I (1984) *Wheats of Central Asia*. Tashkent: FAN, USSR, 135 pp. (in Russian).

https://doi.org/10.1017/S1479262115000325 Published online by Cambridge University Press