

# Genetic diversity of old bread wheat germplasm from the Black Sea region evaluated by microsatellites and agronomic traits

Svetlana Landjeva<sup>1\*</sup>, Ganka Ganeva<sup>1</sup>, Viktor Korzun<sup>2</sup>, Dean Palejev<sup>3</sup>, Sabina Chebotar<sup>4</sup> and Alexander Kudrjavev<sup>5</sup>

<sup>1</sup>Institute of Plant Physiology and Genetics, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria, <sup>2</sup>KWS LOCHOW GMBH, Ferdinand-von-Lochow-Straße 5, 29303 Bergen, Germany, <sup>3</sup>Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria, <sup>4</sup>Plant Breeding and Genetics Institute – National Center of Seed and Cultivar Investigations, Ovidiopolskaya doroga 3, 65036 Odessa, Ukraine and <sup>5</sup>Vavilov Institute of General Genetics, Russian Academy of Sciences, Gubkina Street 3, 119991 Moscow, Russia

Received 28 March 2014; Accepted 21 June 2014 – First published online 14 July 2014

## Abstract

Old germplasm is an important genetic resource for enhancing modern crops with new alleles. In the present study, the genetic diversity of 52 historic varieties and landraces of bread wheat originated from the Western (Bulgaria) and Northeastern (Ukraine, Russia and Georgia) regions of the Black Sea basin was assessed based on microsatellite markers and agronomic characteristics. A set of 24 markers detected a total of 263 alleles at 25 microsatellite loci, with an average number of 10.5 alleles per locus and an average polymorphic information content (PIC) of 0.74. A total of 63 alleles at 22 loci were unique, being specific to a particular accession. Half of the alleles (132) were regionally specific, and the rest were common between the Western and Northeastern accessions. The latter group was characterized with greater total and private allelic richness, a higher number of unique alleles and a higher average PIC. The population was found to be very heterogeneous (average heterogeneity 41%), with the Northeastern pool (52.8%) being more diverse than the Western pool (30.9%). Most of the accessions of the Western group clustered together, and the rest were distributed among the subclusters of the Northeastern germplasm. Large inter-group differences in the frequencies of alleles ranging from 3.1 at *Xgwm294-2A* to 16.7 at *Xgwm333-7B* were observed. This variation might partly account for the differences in certain yield-related traits. The Northeastern accessions had significantly longer spikes with more number of spikelets. Some issues related to germplasm preservation in seed genebanks are discussed herein. The large molecular variation observed could be utilized by breeders for the selection of diverse parents, or by researchers for the production of mapping populations.

**Keywords:** agronomic traits; historic/traditional variety; landraces; molecular diversity; *Triticum aestivum* L.

## Introduction

Landraces and historic varieties are a rich portion of the bread wheat gene pool (Newton *et al.*, 2010). On the one hand, while modern varieties are a result of intended

---

\* Corresponding author. E-mails: s\_landjeva@mail.bg; slandjeva@gmail.com

breeding, landraces lack formal crop improvement and are considered dynamic, genetically diverse and locally adapted populations with distinct identity; on the other hand, historic varieties are regarded as relatively homogeneous selections made within landraces, or early breeding releases, that were once on a variety registration list but are no longer in use (Camacho Villa *et al.*, 2005; Serpolay *et al.*, 2011). Generally, both types of 'exotic' wheats have evolved from a broader germplasm pool and therefore are a valuable and yet underutilized resource for breeding purposes.

Since N. I. Vavilov and H. Harlan started to collect seeds in the beginning of the last century, a number of seed genebank collections have been assembled and reproduced (Börner, 2006). The extent of the genetic diversity of this germplasm needs to be described and evaluated to safeguard natural genetic variability and genomic integrity, and to provide a wider basis for the choice of parents for breeding purposes. In addition to traditional morphological and agronomic traits (Denčić *et al.*, 2000; Dotlačil *et al.*, 2010), and biochemical and physiological characteristics (Gregová *et al.*, 2006), the developments in DNA marker technologies made the genetic variation available in the old germplasm more accessible. Studies on the molecular diversity of landraces and old varieties from France (Roussel *et al.*, 2004), Russia (Khlestkina *et al.*, 2004), Ukraine (Chebotar, 2004), Portugal (Ribeiro-Carvalho *et al.*, 2004) and Oman (Al Khanjari *et al.*, 2007) have been reported.

Bread wheat is the most important crop in the countries of the Black Sea (BS) region. Old Bulgarian germplasm grown to the west of the BS basin was noted for its stable yield, drought tolerance, high protein content, or good bread-making quality (Majdrakov, 1945). The gene pool of wheat grown on the territory of Ukraine, including its southern regions, has a long history, as evidenced by archaeological excavations and historical documents (Khvoiko, 1909). High-yielding Crimean hard red wheat varieties have been shown to participate in the pedigrees of popular European and American varieties (Smale, 1996). The Georgian soft free-threshing landraces are all ancient wheats of various flowering habits, increased tillering, yield stability, disease resistance and grain quality (Jorjadze *et al.*, 2014). Most of this germplasm is not cultivated any longer and is nearly extinct. Seed samples have been maintained and reproduced by the large European seed genebanks in Germany, Russia, Czech Republic and Ukraine.

This study presents the comparative analysis of molecular genetic diversity based on microsatellite markers, and the agronomic performance of landraces and historic varieties from the BS basin. The genetic relationships within this old germplasm as well as some issues related to diversity preservation in seed genebanks are discussed herein.

## Materials and methods

### Plant material

A total of 52 wheat (*Triticum aestivum* L.) landraces and historic varieties from the countries of the BS region were used in the present study (Table S1; Fig. S1, available online). All accessions were collected or developed before 1950; thereafter, they have been maintained and reproduced by seed genebanks at the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany, and the Vavilov Institute for Plant Industry (VIR), St Petersburg, Russia. Seed samples were kindly provided by the two genebanks. The accessions were classified into two groups according to their geographic origin: BS-West, including the accessions from Bulgaria (28), and BS-Northeast, comprising the accessions from Ukraine, Russia and Georgia (24).

### DNA extraction and microsatellite analysis

For each accession, four to six (on average five) seedlings were pooled and their DNA was isolated by a mini-prep method adapted from Rogowsky *et al.* (1991). A total of 24 primer pairs flanking wheat microsatellites and amplifying loci on 14 linkage groups (Table 1) were selected according to their chromosome arm location, level of polymorphism detected and feasibility of allele scoring, including the absence of stutter fragments. Microsatellite designation, composition, primer sequences and chromosome location of the amplified loci have been reported by Devos *et al.* (1995) and Röder *et al.* (1998). All the primer pairs amplified a single locus, except for one that amplified orthologous loci, one on each of chromosomes 5A and 5D. The polymerase chain reaction (PCR) was performed according to Röder *et al.* (1998), with 35 instead of 45 cycles of denaturation for 1 min at 94°C, annealing for 1 min at 50°C (55 or 60°C depending on the primer) and extension for 1 min at 72°C followed by a final extension step for 5 min at 72°C. PCR products were separated using the 3130xl Genetic Analyzer (Applied Biosystems, Carlsbad, CA, USA) according to the manufacturer's instruction. The Genotyper® software (Applied Biosystems, Carlsbad, CA, USA) was used to assess the fragment size of the alleles at each microsatellite locus.

### Diversity analysis

For each geographic region and for the total sample, genomic diversity was evaluated based on (1) allelic richness as a measure of genetic variation (number of alleles

**Table 1.** Microsatellite allelic polymorphism in 52 old bread wheat accessions from the BS region (BS-West, 28 accessions from Bulgaria; BS-Northeast, 24 accessions from Ukraine, Russia and Georgia)

Locus-chromosome	No. of alleles	No. of unique alleles	Range of the fragment size (bp)	PIC	Heterogeneity (%)	Prevailing allele in bp (frequency)			Allele-frequency variation (%)	
						BS-West		BS-Northeast		Total
						BS-West	BS-Northeast			
<i>XTaglut-1A</i>	7	0	137–154	0.79	82.7	142 (0.26)	139 (0.39)	139 (0.30)	7.2	
<i>Xgwm95-2A</i>	7	2	105–124	0.54	11.8	118 (0.70)	118 (0.60)	118 (0.65)	4.3	
<i>Xgwm294-2A</i>	16	3	63–112	0.88	32.7	94 (0.27)	92 (0.33)	84 (0.21)	3.1	
<i>Xgwm312-2A</i>	13	2	191–224	0.83	34.6	193 (0.29)	209 (0.36)	193 (0.28)	6.6	
<i>Xgwm619-2B</i>	12	3	136–174	0.80	14.0	136 (0.54)	136 (0.22)	136 (0.39)	11.7	
<i>Xgwm261-2D</i>	12	5	164–219	0.79	26.9	174 (0.29)	174 (0.46)	174 (0.37)	9.2	
<i>Xgwm155-3A</i>	8	1	130–151	0.74	15.4	143 (0.46)	143 (0.31)	143 (0.39)	8.1	
<i>Xgwm480-3A</i>	6	2	168–192	0.37	21.2	170 (0.84)	170 (0.70)	170 (0.78)	8.0	
<i>Xgwm389-3B</i>	12	4	116–143	0.84	32.7	116 (0.27)	116 (0.26)	116 (0.26)	5.2	
<i>Xgwm160-4A</i>	13	3	165–205	0.86	100.0	183 (0.22)	177 (0.18)	177 (0.18)	4.4	
<i>Xgwm291-5A</i>	15	4	112–177	0.90	27.5	173 (0.23)	167 (0.31)	161 (0.15)	8.7	
<i>Xgwm205-5A</i>	7	1	151–166	0.77	19.6	151 (0.48)	157 (0.39)	151 (0.31)	14.1	
<i>Xgwm205-5D</i>	6	0	137–147	0.67	90.2	137 (0.55)	137 (0.45)	137 (0.50)	5.5	
<i>Xgwm499-5B</i>	14	5	97–176	0.73	28.8	97 (0.43)	97 (0.48)	97 (0.45)	9.2	
<i>Xgwm408-5B</i>	12	4	134–193	0.69	30.8	178 (0.55)	178 (0.47)	178 (0.51)	9.9	
<i>Xgwm272-5D</i>	4	1	127–143	0.64	100.0	127 (0.49)	127 (0.46)	127 (0.48)	4.7	
<i>Xgwm190-5D</i>	10	4	203–225	0.51	17.3	211 (0.71)	211 (0.64)	211 (0.68)	6.9	
<i>Xgwm82-6A</i>	5	0	146–154	0.68	100.0	152 (0.45)	152 (0.37)	152 (0.41)	4.3	
<i>Xgwm570-6A</i>	11	1	93–165	0.72	100.0	93 (0.48)	93 (0.43)	93 (0.46)	6.7	
<i>Xgwm219-6B</i>	14	4	146–194	0.88	26.9	181 (0.29)	177 (0.26)	181 (0.20)	9.0	
<i>Xgwm333-7B</i>	9	4	140–171	0.49	17.6	146 (0.84)	146 (0.50)	146 (0.69)	16.7	
<i>Xgwm400-7B</i>	11	2	134–153	0.74	29.4	145 (0.50)	145 (0.39)	145 (0.45)	13.5	
<i>Xgwm46-7B</i>	11	2	145–179	0.82	26.9	173 (0.39)	175 (0.31)	173 (0.29)	11.9	
<i>Xgwm577-7B</i>	15	4	128–167	0.88	24.0	167 (0.25)	128 (0.27)	143 (0.18)	14.9	
<i>Xgwm44-7D</i>	13	2	160–188	0.89	26.9	164 (0.25)	182 (0.29)	167 (0.16)	12.9	

and average number of alleles per locus); (2) private allelic richness as a measure of genetic distinctiveness (number of specific alleles found in one geographic group but absent in the other); (3) occurrence of unique alleles as another measure of distinctiveness (alleles that occur in one accession in a group or in the whole gene pool); (4) polymorphic information content (PIC) of each marker as a measure of gene diversity, which is calculated by assuming homozygosity according to the formula:

$$\text{PIC} = 1 - \sum p_i^2,$$

where  $p_i$  is the proportion of the population carrying the  $i$ th allele at the locus; (5) locus and variety heterogeneity estimated as the percentage of accessions or loci carrying more than one allele at the corresponding locus or variety, respectively; (6) allele-frequency variation calculated using the formula:

$$A_{XY}(\%) = \frac{\sum |P_{Xi} - P_{Yi}|}{N_{XY}} \times 100,$$

where  $P_{Xi}$  and  $P_{Yi}$  are the proportions of the  $i$ th allele at a given locus in groups  $X$  and  $Y$ , respectively, and  $N_{XY}$  is the total number of alleles detected at the corresponding locus (Khlestkina *et al.*, 2004). The parameter  $A_{XY}$  was calculated only for alleles that were shared by the two geographic groups. To estimate the allelic imbalance at a given locus, its PIC value was compared with the corresponding theoretical  $\text{PIC}_{\max}$  value equivalent to the potentially highest diversity level at the same locus.

### Data analyses

Because of the high non-uniformity of all accessions at a number of loci (i.e. presence of more than one allele per locus in a given accession), the data were initially normalized so that the per-variety sum of counts in each locus was 5 to reflect the fact that the expected number of individuals per accession was 5. In such cases, the detected alleles were considered as equally contributing to the genetic make-up of the accession. For example, if only two alleles were present in a genotype at a given locus, it was estimated that each of them has an expected count of  $5/2$ , and if five alleles were present, each of them has an estimated count of  $5/5 = 1$ .

Allelic richness, PICs, heterogeneity, allele frequencies and allele-frequency variations were computed using the software environment for statistical computing and graphics R (R Core Team, 2013). To detect the significance of the difference between the relative PIC values (observed  $\text{PIC}/\text{PIC}_{\max}$ ) averaged over the loci with low and high variations, Welch's  $t$  test was performed in R. To test the significance of the correlations between

gene diversity and allelic richness, correlation analysis was performed using the software package Statistica Statsoft version 7 (StatSoft, 2005).

The genetic distance coefficients between each pair of the accessions were calculated as  $1 - \text{GS}$  (genetic similarity) from the data of 21 microsatellite loci displaying seven or more alleles using the formula:

$$\text{GS} = 2Q_{ij}/(Q_i + Q_j),$$

where  $Q_{ij}$  is the number of alleles common to the varieties  $i$  and  $j$ , and  $Q_i + Q_j$  is the sum of the number of alleles in each of the two varieties. The hierarchical clustering of the accessions was performed using the R `hclust` function to generate a Ward dendrogram (Ward, 1963).

### Agronomic evaluation

Seven agronomic traits were evaluated in a 2-year (2009, 2012) field trial in the experimental field of the Institute of Plant Physiology and Genetics, Bulgarian Academy of Sciences, Sofia (42°41'N, 23°19'E). In both seasons, sowing was carried out in October at the optimum time for this climatic zone. A preliminary field trial aiming at seed multiplication did not reveal any within-accession variations in spike morphological traits (spike colour, waxiness, presence/absence of awns, awns colour, etc.), which allowed each accession to be sown in small plots. Each accession was sown as space-planted two-row plots (1 m in length) in two replications (5 cm between the plants within the row and 20 cm between the rows). To cover the genotypic heterogeneity, seeds from a bulk of spikes for each accession were used for sowing. The following characteristics were measured on five representative plants per replication at the harvest time: final plant height (from the soil surface to the bottom of the spike); spike length; spikelet number; grain number and grain mass per spike; 1000-grain mass. Heading time was determined as days from 1 May when approximately half of the ears in a plot have half emerged from the flag leaf. Trait values averaged over the years and accessions are presented for each group of accessions. To assess the effect of the year (accounting for climatic conditions) and the geographic origin of the accessions, two-way analysis of variance (ANOVA) was performed using Statistica Statsoft version 7.

## Results

### Diversity at microsatellite loci

The applied set of 24 mapped microsatellite markers detected a total of 263 different alleles at 25 microsatellite

**Table 2.** Comparison of molecular genetic diversity and agronomic characterization (mean  $\pm$  standard error; CV in brackets, in %) between the old bread wheat accessions from Bulgaria (BS-West, 28) and from Ukraine, Russia and Georgia (BS-Northeast, 24)

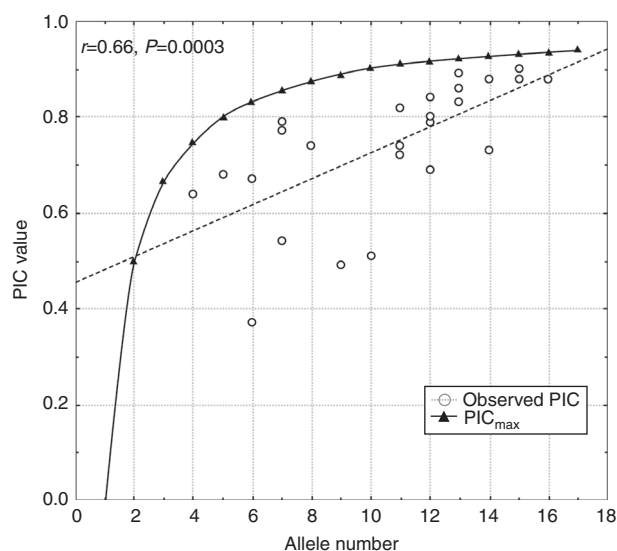
Characteristics	BS-West	BS-Northeast	Total/average
Molecular diversity			
Allelic richness	173	221	263
Average no. of alleles per locus	6.9	8.8	10.5
Private allelic richness	42	90	
No. of loci containing private alleles	18	23	
No. of unique alleles within the group/whole population	49/19	69/44	63
No. of loci containing unique alleles within the group/whole population	18/12	24/20	22
No. of accessions carrying unique alleles within group/whole population	17/12	18/17	29
Average PIC	0.68	0.75	0.74
Average variety heterogeneity	30.9	52.8	41.0
Agronomic characterization			
Plant height (cm)	130.6 $\pm$ 1.66 (6.7)	133.1 $\pm$ 1.36 (4.8)	131.7 $\pm$ 1.11 (6.0)
Spike length (cm)	11.9 $\pm$ 0.25 (11.3)	13.2 $\pm$ 0.33 (11.9)***	12.5 $\pm$ 0.22 (12.6)
Spikelets/spike	20.9 $\pm$ 0.48 (12.2)	22.9 $\pm$ 0.40 (8.2)***	21.8 $\pm$ 0.35 (11.3)
Grain number/spike	35.0 $\pm$ 1.17 (17.6)	37.1 $\pm$ 1.65 (20.9)	35.9 $\pm$ 0.98 (19.2)
Grain mass/spike (g)	1.6 $\pm$ 0.06 (20.5)	1.6 $\pm$ 0.07 (21.9)	1.57 $\pm$ 0.05 (21.0)
1000-grain mass (g)	43.9 $\pm$ 0.84 (10.1)	41.8 $\pm$ 0.88 (9.9)	43.0 $\pm$ 0.62 (10.2)
Heading (days from 1 May)	24.1 $\pm$ 0.83 (18.2)	26.1 $\pm$ 0.38 (6.8)	25.0 $\pm$ 0.51 (14.4)

\*\*\*Significantly different at  $P < 0.001$ .

loci on 14 chromosomes within the 52 accessions (Tables 1 and 2). Allelic richness ranged from 4 (*Xgwm272-5D*) to 16 (*Xgwm294-2A*) alleles per locus with an average number of 10.5 alleles. Allelic polymorphism was manifested by the occurrence of unique alleles appearing only once in the whole pool of entries, and by varying frequency of the prevailing alleles. The number of unique alleles varied between loci spanning from 0 (*XTaglut*, *Xgwm205-5D* and *Xgwm82-6A*) to 5 (*Xgwm261-2D* and *Xgwm499-5B*). In total, 63 unique alleles were detected at 22 loci, representing 24% of all alleles. Gene diversity expressed as PIC values ranged from 0.37 (*Xgwm480-3A*) to 0.90 (*Xgwm291-5A*), with an average of 0.74. The frequency of the prevailing alleles varied considerably between the loci with different polymorphism degrees. For instance, the 161 bp allele at locus *Xgwm291-5A* displaying a total of 15 alleles occurred at a frequency of 0.15, whereas the 170 bp allele at locus *Xgwm480-3A* having six alleles appeared with a frequency of 0.78 (Table 1).

A highly significant correlation was found between the observed gene diversity (PIC values) and allelic richness ( $r = 0.66$ ,  $P < 0.001$ ; Fig. 1). The comparison between the PIC value for a given locus and the corresponding theoretical  $PIC_{max}$  value equivalent to the potentially highest diversity level at the same locus presented an estimation of the allelic imbalance at this locus. For the ten loci displaying ten or less alleles (i.e. less than the average of 10.5 alleles/locus), the observed PIC values

were much lower and, therefore, suggestive of larger allelic imbalance compared with the remaining 15 loci having more than ten alleles. The relative PIC values (observed  $PIC/PIC_{max}$ ) averaged over the loci were lower, showing a higher variation for the loci with ten or less alleles ( $0.7 \pm 23.2\%$ ) than those with more



**Fig. 1.** Scatter plot of PIC values at 25 microsatellite loci against the number of alleles analysed within 52 old bread wheat accessions from the BS region. The corresponding theoretical  $PIC_{max}$  values are also presented.



than ten alleles ( $0.9 \pm 7.9\%$ ) ( $t = -2.609$ ,  $P \approx 0.024$ ,  $df = 11.02$ , as determined by Welch's  $t$  test that does not assume equal variances).

A high level of heterogeneity was found at all loci varying from 11.8% for *Xgwm95-2A*, 1) with six accessions being heterogeneous at this locus, to 100% for *Xgwm160-4A*, *Xgwm272-5D*, *Xgwm82-6A* and *Xgwm570-6A*, with all 52 accessions being heterogeneous at these four loci (Table 1).

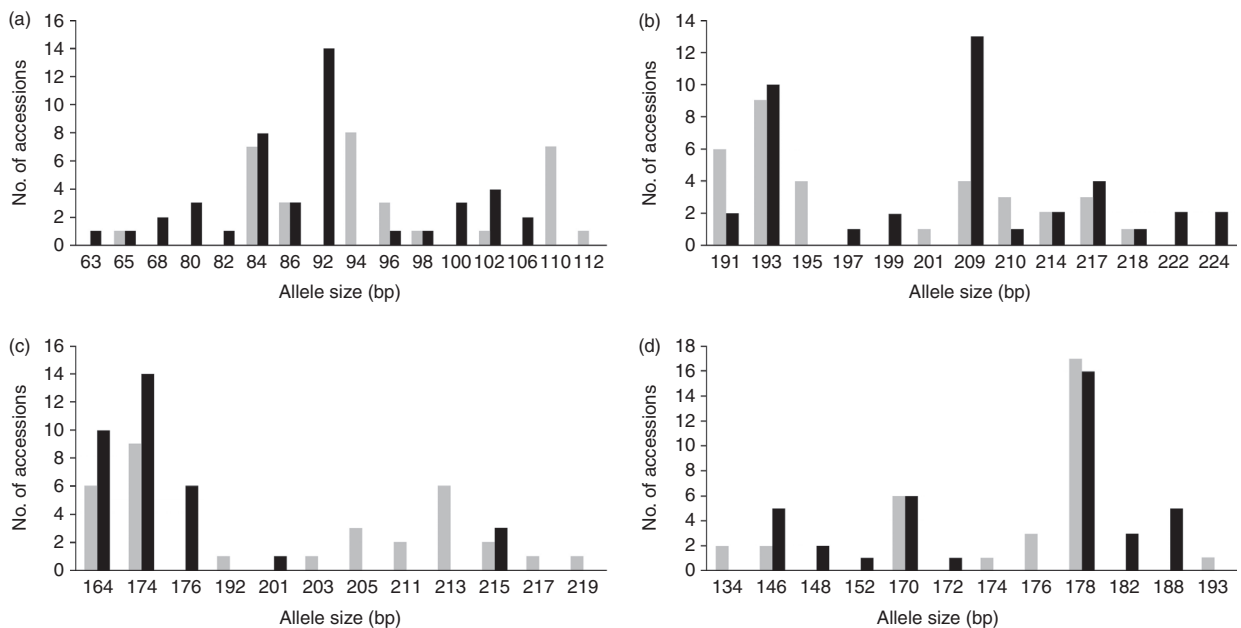
### Diversity in relation to the geographic origin

The size of the two geographic groups differed by only four entries, and since the number of the loci studied was equal for both groups, we directly compared the molecular diversity of the two sets. Genetic diversity was higher for the BS-Northeast population than for the BS-West group. This was evidenced by greater allelic richness (i.e. a higher total and average number of alleles per locus) and greater private allelic richness (i.e. a higher number of group-specific alleles), including a higher number of unique alleles at almost all loci (Table 2). The unique alleles were carried by 18 accessions from the BS-Northeast group and by 17 accessions from the BS-West group, or by 29 accessions when considering the whole sample. More than one-third of the alleles that were unique for the whole population, or 22 alleles, were carried by three Georgian landraces, Ipkli (13), Gomborka (8) and Khulugo (1). The average PIC was also higher for the BS-Northeast pool (Table 2). The correlations between

gene diversity (PIC) and the total number of alleles were found to be highly significant (BS-West group:  $r = 0.83$ ; BS-Northeast group:  $r = 0.64$ ;  $P < 0.001$ ). Neither of the groups had completely homogeneous accessions. Varietal heterogeneity ranged from 16 to 52% (or 4 to 13 heterogeneous loci) for the BS-West group, and from 20 to 100% (or 5 to 25 microsatellite loci) for the BS-Northeast group. The least number of heterogeneous loci (4) was observed in the Bulgarian variety Okerman 804. The Georgian landrace Ipkli was found to be the most heterogeneous with all 25 microsatellite loci having more than one allele.

At 15 of the 25 microsatellite loci, both the BS-West and BS-Northeast groups had the same prevailing allele (Table 1), which was therefore dominant for the whole pool, whereas at the remaining ten loci, the most frequent allele was different for the two groups. The variation in the frequency of shared alleles between the two geographic groups was lowest for *Xgwm294-2A* (3.1%) and highest for *Xgwm333-7B* (16.7%), with an average of 8.7% (Table 1).

Figure 2 shows the allele polymorphism at four microsatellite loci associated with important mapped genes or quantitative trait loci (QTL). Large differences between the two geographic groups were found at loci *Xgwm294* and *Xgwm312* on chromosome 2AL (Fig. 2(a) and (b), respectively) linked to QTL for yield-related traits (Yao *et al.*, 2009). At both loci, the two groups differed with respect to the total and private allele number, and had different prevailing alleles. Locus *Xgwm294* (six shared



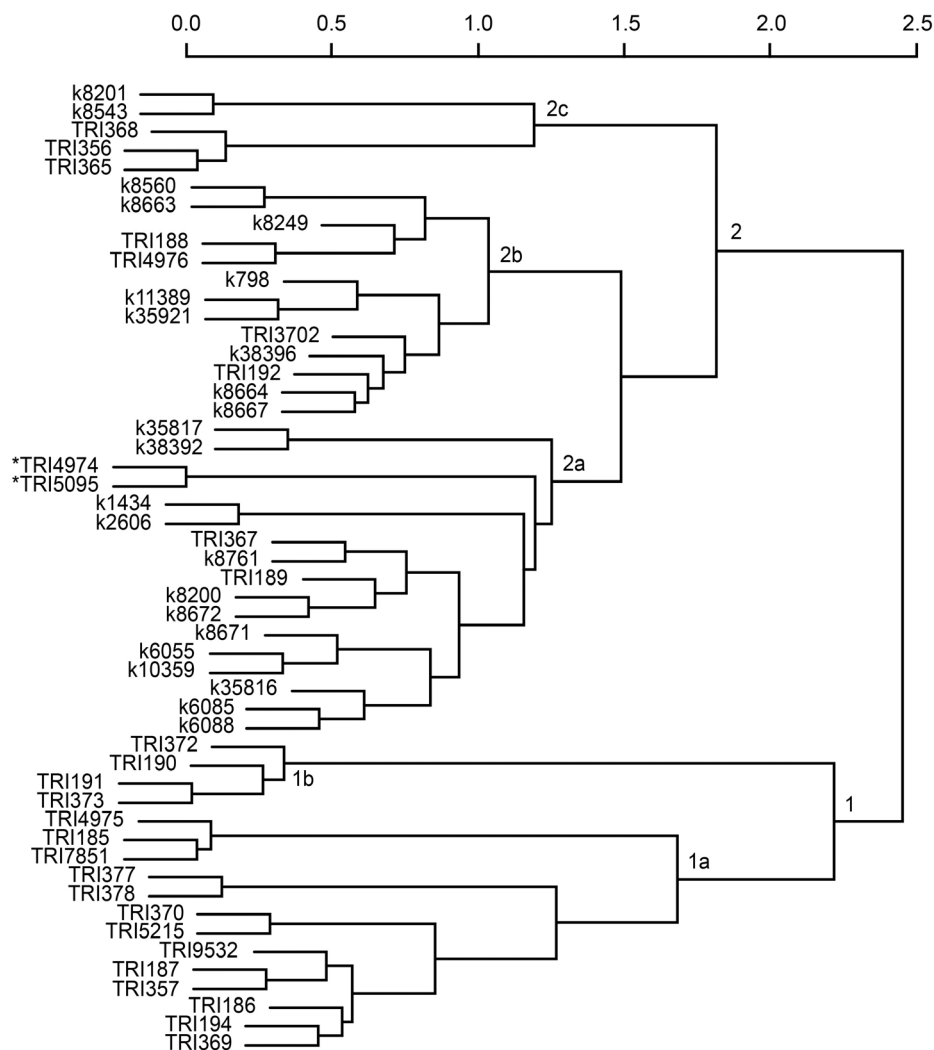
**Fig. 2.** Allele polymorphism at four microsatellite loci observed within 52 old bread wheat accessions from the BS region, including 28 historic varieties from Bulgaria (BS-West; grey bars) and 24 landraces and historic varieties from Ukraine, Russia and Georgia (BS-Northeast; black bars). (a) *Xgwm294-2A*, (b) *Xgwm312-2A*, (c) *Xgwm261-2D* and (d) *Xgwm408-5B*.

and ten private alleles, of which three being unique) was the only one at which the observed prevailing alleles were regionally specific, i.e. allele fragments of size 94 and 92bp were the most frequent for the BS-West and BS-Northeast groups, respectively, and were not present in the other group. At locus *Xgwm261* linked to the semi-dwarfing gene *Rht8* on chromosome 2DS (Korzun *et al.*, 1998), and to QTL for yield components (Heidari *et al.*, 2011; Zhao *et al.*, 2013), ten and five alleles were detected among the BS-West and BS-Northeast accessions, respectively (three common and nine private alleles, of which five being unique) (Fig. 2(c)). Allele *Xgwm261*<sub>174</sub> was dominant in both pools. The marker locus *Xgwm408* in close linkage to the vernalization response gene *Vrn-B1* on 5BL (Leonova *et al.*, 2003) was characterized by large differences between the two

geographic groups (three shared and nine private alleles, of which four being unique) and common dominant allele, *Xgwm408*<sub>178</sub> (Fig. 2(d)). The two sets of accessions differed notably also by the number and frequency of allelic variants at locus *Xgwm400* on 7BS in the vicinity of the *Ppd-B2* gene for photoperiodic response (at 8.8 cM), and a QTL for grain protein content (Khlestkina *et al.*, 2009) displaying four common and seven private alleles, of which two being unique (data not shown).

### Cluster analysis

A dendrogram (Fig. 3) was constructed on the basis of the genetic distance at 21 microsatellite loci displaying seven or more alleles by hierarchical clustering of 1–GS using



**Fig. 3.** Hierarchical clustering dendrogram, showing genetic relationships of 52 bread wheat landraces and historic varieties originated from the Western (denoted by ‘TRI’) and Northeastern (denoted by ‘k’) regions of the BS basin. The genetic distance coefficients were calculated as 1–GS from the data of 21 microsatellite loci displaying seven or more alleles. The asterisk indicates the accessions that are identical at all loci.

Ward's method (Ward, 1963). The genetic distance 1–GS estimates of all possible pairs of accessions ranged from 0.0196 (TRI 191, Dunavka vs. TRI 373, Duravko) to 0.9273 (k8201, a Crimean landrace vs. k38392, the Georgian landrace Lagodehis gruzeltavta). The 1–GS coefficient of the pair TRI 4974 versus TRI 5095 was 0, implying that the two varieties cannot be differentiated. Two clusters were distinguished. Cluster 1 comprised mostly (17 out of 28) accessions from Bulgaria (BS-West group). Within this gene pool, accessions from various agroecoregions grouped together. For instance, subcluster 1a consisted mostly of accessions from Northern Bulgaria, while subcluster 1b included only accessions from the Southern part of the country. The remaining Bulgarian accessions were distributed among the three subgroups of the more diverse cluster 2. Of the 15 Crimean landraces, eight were grouped in subcluster 2a, closer to the major group of Bulgarian wheats. The five Georgian accessions were very diverse and did not form a separate cluster.

### **Agronomic comparison**

A considerable variation in agronomic characteristics estimated by the coefficient of variation (CV, %) was recorded in both geographic groups (Table 2). A relatively high variability (CV > 10%) was established for all traits (except for plant height), with the variation being largest for grain mass per spike (CV > 20%). According to the two-way ANOVA, year and geographic origin, but not their interaction, significantly affected only two yield components (spike length and number of spikelets per spike). Year had more marked effect on both traits ( $P < 0.001$ ,  $df = 1$ ) than did geographic origin ( $P < 0.001$ ,  $df = 1$ ) as indicated by the F-ratios of 109.1 (spike length) and 117.3 (spikelet number) versus 43.1 (spike length) and 19.9 (spikelet number), respectively. Accessions from the BS-Northeast group had longer spikes and more spikelets per spike (Table 2).

## **Discussion**

### **Microsatellite diversity**

The amount of population genetic variation is a fundamental parameter for diversity preservation and for breeders as a basis for crop improvement in response to environmental challenges, and economic and health demands. Important characteristics of genetic diversity are allelic richness, allelic distinctiveness (number of private, rare and unique alleles) and locally prevailing alleles. When compared with the results reported for old germplasm of bread wheat, our study evidenced comparable or higher microsatellite

diversity across the wheat gene pool from the BS basin. For instance, the average number of alleles per locus (10.5) was higher in comparison with the same characteristics obtained for 59 Portuguese landraces 'Barbela' (4.8; Ribeiro-Carvalho *et al.*, 2004), 161 Omani landraces (8.7; Al Khanjari *et al.*, 2007), but much lower than the one calculated for the set of 347 diverse old accessions from 44 countries (18.3; Mitrofanova *et al.*, 2009). The proportion of unique alleles (24%) displaying the qualitative aspect of allele diversity was higher compared with that detected in Omani wheats (2.6% Al Khanjari *et al.*, 2007) and the old diverse germplasm (18%; Mitrofanova *et al.*, 2009). Gene diversity expressed as the average PIC value (0.74) was either comparable with (0.8; Mitrofanova *et al.*, 2009) or higher than the PIC values reported for old Omani (0.5), Portuguese (0.52) or Siberian (0.6; Khlestkina *et al.*, 2004) germplasm. Landraces and old varieties are expected to be genetically non-uniform. The average heterogeneity of 41%, revealed in this study, was very high compared with the available literature data on old germplasm (9.1%; Al Khanjari *et al.*, 2007).

The plant material studied is a product of natural and artificial selection carried out by generations of farmers or is a result of early breeding attempts, and so is characterized by high allelic richness and diversity levels. The loci with a relatively lower number of allelic variants and/or lower PIC and PIC/PIC<sub>max</sub> values could evidence possible diversity reductions due to intended breeding. Indeed, there have been reports demonstrating how breeding for specific targets narrows down the genetic variation, leading to reduced allele frequency in certain groups of genotypes in bread (Stachel *et al.*, 2000; Huang *et al.*, 2002) and durum (Maccaferri *et al.*, 2003) wheat elite varieties, designed for growing under more or less standardized conditions. However, no such trend is observed in our study on the old genetic resources. The loci with a higher number of alleles (> 10) and higher PIC/PIC<sub>max</sub> values (on average 0.9) did include loci with known associations with potential breeding targets, such as traits related to productivity (*Xgwm294-2A*, *Xgwm312-2A* and *Xgwm261-2D*) and flowering time (*Xgwm261-2D*, *Xgwm408-5B*, *Xgwm400-7B* and *Xgwm46-7B*) (Leonova *et al.*, 2003; Khlestkina *et al.*, 2009; Yao *et al.*, 2009; Heidari *et al.*, 2011; Zhao *et al.*, 2013). One possible explanation is that both the farmers' selection and the early breeding activities have been carried out under specific non-standardized agroecological conditions, and, therefore, the high allelic variation at these loci might reflect the broad environmental diversity, resulting in the fixation of various traits and alleles of competitive power. The differences in allelic richness at the various microsatellite loci could also evidence the differences in their mutation rate. In particular, landraces that have been developed at low levels of fertilizers and disease control have been subject to natural selection pressure for adaptability



and survival rather than to purposeful breeding for grain productivity (Zeven, 1998), and, therefore, could accumulate mutations of various rates at different loci.

### Geographical variations

Molecular diversity in the two gene pools differed with respect to allelic richness, distinctiveness and extent of heterogeneity. The accessions from Ukraine, Georgia and Russia (BS-Northeast group) were characterized by a higher total number of alleles, a higher number of private (i.e. group-specific) alleles and a higher number of unique alleles, as well as greater heterogeneity. Large differences in allele frequencies found at certain loci between the two geographic areas could be related to adaptation driving forces on linked genes.

In the case of locus *Xgwm261* on 2DS, this could be explained by selection pressure for genomic regions associated with flowering time and/or yield-related traits. The *Xgwm261*<sub>165–174</sub> alleles often in combination with the *Ppd-D1b* gene for day-length response in later heading genotypes could probably guarantee wider adaptability (Šíp *et al.*, 2010).

Loci *Xgwm261* on 2DS, and *Xgwm312* and *Xgwm294* on 2AL have been linked to highly significant QTL for yield and yield components (Yao *et al.*, 2009; Heidari *et al.*, 2011; Zhao *et al.*, 2013). The variation in the number and frequencies of alleles at these loci could at least partially account for the significant differences in the spike length and spikelet number observed between the two sets of accessions. The significant prevalence of allele *Xgwm408*<sub>178</sub> on 5BL in both groups (average frequency of 0.51, as shown in this study, and 0.67 among old Ukrainian germplasm, as reported earlier by Chebotar (2004)) is suggestive of its adaptive role probably associated with the tightly linked *Vrn-B1* gene determining growth habit and flowering time (Leonova *et al.*, 2003). All but two accessions were of winter type. According to van de Wouw *et al.* (2009), alleles that are regionally specific and dominant are possibly the ones that contribute to the adaptation to the local agroecological conditions and, therefore, are more likely to be useful in achieving the goals of breeders. The old germplasm consisting of landraces and historic varieties is a valuable source of such alleles, so the monitoring of genetic erosion should be focused on these locally common alleles as recommended by van de Wouw *et al.* (2009).

### Genetic relationships

Most of the accessions from Bulgaria (BS-West group) were grouped together into a separate cluster. The

grouping of some varieties could be due to their genealogy relationships. For instance, Okerman 804 (TRI 9532) and Sadovka (TRI 194) both have varieties Noé and Nr.16 (TRI 187) in their pedigrees. The three accessions group together in subcluster 1a. Accessions originated from neighbouring regions and being subjected to comparable selection forces also tended to group together, suggesting that concurrent allele combinations have been selectively fixed to succeed in achieving adequate response to the local environment. For instance, TRI 190, TRI 191, TRI 372, TRI 373 and TRI 4975 (subcluster 1b) were selected from the region of Plovdiv, Southern Bulgaria, while TRI 368, TRI 356 and TRI 365 (subcluster 2c) came from the region of Veliko Turnovo, Northern Bulgaria.

The second major cluster comprised the remaining Bulgarian accessions and the ones of the BS-Northeast group with no clear pattern of grouping. One reason might be that gene flows had probably taken place through the constant migration of people between the Crimean Peninsula and Bulgaria since the 15th century. It is very likely that many crops, including wheat, had been introduced as a result of the migration of people. Also, the market exchange through the major BS ports (see Fig. S1, available online) had possibly contributed to seed gene flows. Grain exports through the Belgorod-Dniester port near Odessa on the Northwestern BS shore had occurred as earlier as 13–14th centuries as evidenced by the report of the florentine trade company representative F.B. Pegolotti (Konovalova, 1994).

The observed grouping of Crimean wheats with varieties from Southern or Northern Bulgaria, namely subclusters 2a, 2b and 2c, complies with the reports on the close relationships between accessions from Ukraine and Bulgaria, based on their position in the neighbour-joining tree (Balfourier *et al.*, 2007; Chebotar, unpublished). The finding that Georgian wheats do not cluster also reflects their large diversity. According to Jorjadze *et al.* (2014), Georgian landraces have been evolved under a broad range of environments and have probably developed various well-adapted ecotypes. Among the five investigated accessions, Ipkli, which is the oldest reported name for bread wheat in Georgia, and presumably a very ancient wheat variety, was found to be the most polymorphic and most 'unique' carrying 13 alleles that were not identified in the rest of the material. Being a part of a larger Caucasus group of accessions, this material has been placed in a different tree branch together with samples from Turkey and Central Asia (Balfourier *et al.*, 2007).

### Genebank aspects

The results of this study raise questions in relation to diversity preservation via seed banking. Some of these

questions have already been brought up by other authors (Börner *et al.*, 2000; Newton *et al.*, 2010; Pinheiro de Carvalho *et al.*, 2013). One of the important challenges for *ex situ* genebanks is to minimize the contamination due to erroneous handling during the reproduction (Börner, 2006). This is believed to aid in maintaining the genetic integrity and purity in the genebank accessions of self-pollinating species. Therefore, the efforts to preserve the old germplasm variation involve not only seed reproduction but also the periodical evaluation of molecular, morphological and agronomic diversity. Comparisons of agromorphological and DNA-based data can further identify the duplications within collections and thus help in their rationalizing (Dobrovolskaya *et al.*, 2005).

The IPK genebank keeps two samples each of five Bulgarian varieties: Nr.7 (TRI 185 and TRI 4974); Nr.84 (TRI 188 and TRI 4976); Professor Ackermann (TRI 192 and TRI 370); Dunavka/Duravko (TRI 191 and TRI 373); Sadovka/Sadorka (TRI 194 and TRI 372). The pairs might represent heterogeneous populations or duplicates, or could be a result of seed mishandling. Although the first pair might have come from a heterogeneous population, and, therefore, is expected to be genetically diverse to some extent, the two accessions are very distant on the dendrogram. In addition, the latter (TRI 4974) appears genetically identical to TRI 5095, another Bulgarian historic variety, which comes from another breeding station designed for growing in a different environment. This might signify incorrect handling during seed multiplication in the case of TRI 4974. The pair Dunavka/Duravko as well as the two accessions under the name Nr.84 are genetically very similar and group together within subclusters 1b and 2b, respectively, suggesting that they are biotypes of a nearly uniform variety. The two pairs under similar names (Professor Ackermann and Sadovka/Sadorka) are genomically detached in addition to their morphological distinctiveness (see Table S1, available online). Thus, they either could be different accessions under wrong names or might have come from highly non-uniform populations that have possibly been further divided into different accessions, thus representing examples of genetic shift. A case of genetic drift as a result of numerous regenerations has already been reported for an originally highly heterogeneous Albanian accession that became more homogeneous after 15 regenerations as revealed by microsatellite analysis (Börner *et al.*, 2000).

The analysis of genealogy of modern varieties shows that, in general, landraces and historic varieties have had limited use for breeding purposes (Martynov *et al.*, 2006). However, the inherent wide genetic variability of this germplasm may have a buffer effect against environmental stresses over time. This is important in the context of current climate fluctuations and constant changes in

pathogen populations. In addition, the recent broadening of organic farming has driven the interest in more diverse germplasm, which is considered to be more suitable for growing under non-standardized environmental conditions in organic farms (Serpoly *et al.*, 2011). Landraces and old varieties are a product of selection and adaptation to specific environments that had been carried out in conditions far too different from the current standardized agricultural practices. In this aspect, they represent a useful source of alleles to improve crop plasticity and their ability to grow in low input or organic agrosystems.

This study demonstrated the presence of high diversity in the old germplasm from the Western and Northeastern regions of the BS basin. Enhancing the wheat crop with 'novel' alleles from old parents requires extensive studies on the agronomic performance, grain quality and stress responses of this exotic germplasm, as well as genetic research to associate traits of interest with genomic regions, followed by appropriate incorporation strategies that could be assisted by molecular marker technologies.

### Supplementary material

To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/S1479262114000781>

### Acknowledgements

This study was a part of a bilateral cooperation between the Bulgarian and Russian Academies of Sciences. The authors thank Dr A. Börner, IPK, Gatersleben for providing the seeds. They also gratefully acknowledge Mrs K. Prokopova and Mrs M. Popova for technical assistance.

### References

- Al Khanjari S, Hammer K, Buerkert A and Röder MS (2007) Molecular diversity of Omani wheat revealed by microsatellites: II. Hexaploid landraces. *Genetic Resources and Crop Evolution* 54: 1407–1417.
- Balfourier F, Roussel V, Strelchenko P, Exbrayat-Vinson F, Sourdille P, Boutet G, Koenig J, Ravel C, Mitrofanova O, Beckert M and Charret G (2007) A worldwide bread wheat core collection arrayed in a 384-well plate. *Theoretical and Applied Genetics* 114: 1265–1275.
- Börner A (2006) Preservation of plant genetic resources in the biotechnology era. *Biotechnology Journal* 1: 1393–1404.
- Börner A, Chebotar S and Korzun V (2000) Molecular characterization of the genetic integrity of wheat (*Triticum aestivum* L.) germplasm after long-term maintenance. *Theoretical and Applied Genetics* 100: 494–497.

- Camacho Villa T, Maxted N, Scholten M and Ford-Lloyd B (2005) Defining and identifying crop landraces. *Plant Genetic Resources: Characterization and Utilization* 3: 373–384.
- Chebotař S (2004) Microsatellite analysis of winter wheat varieties that were a basis of breeding process at the South of Ukraine. Selection, seed production and cultivation of field crops. *Proceedings of International scientific conference "Problems of agricultural production in the region of Southern Russia (landscape farming system, soil fertility, plant breeding and seed production)"*, 7–9 June 2004, Rostov-on-Don, pp. 126–130.
- Denčić S, Kastori R, Kobiljski B and Duggan B (2000) Evaluation of grain yield and its components in wheat cultivars and landraces under near optimal and drought conditions. *Euphytica* 113: 43–52.
- Devos KM, Bryan GJ, Collins AJ, Stephenson P and Gale MD (1995) Application of two microsatellite sequences in wheat storage proteins as molecular markers. *Theoretical and Applied Genetics* 90: 247–252.
- Dobrovolskaya O, Saleh U and Malysheva-Otto L (2005) Rationalising germplasm collections: a case study for wheat. *Theoretical and Applied Genetics* 111: 1322–1329.
- Dotlačil L, Hermuth J, Stehno Z, Dvoraček V, Bradová J and Leiřová L (2010) How can wheat landraces contribute to present breeding? *Czech Journal of Genetics and Plant Breeding* 46 (Special Issue): S70–S74.
- Gregová E, Hermuth J, Kraic J and Dotlačil L (2006) Protein heterogeneity in European wheat landraces and obsolete cultivars: additional information II. *Genetic Resources and Crop Evolution* 53: 867–871.
- Heidari B, Sayed-Tabatabaei BE, Saeidi G, Kearsey M and Suenaga K (2011) Mapping QTL for grain yield, yield components, and spike features in a doubled haploid population of bread wheat. *Genome* 54: 517–527.
- Huang XQ, Börner A, Röder MS and Ganai MW (2002) Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. *Theoretical and Applied Genetics* 105: 699–707.
- Jorjadze M, Berishvili T and Shatberashvili E (2014) The ancient wheats of Georgia and their traditional use in the southern part of the country. *Emirates Journal of Food and Agriculture* 26: 192–202.
- Khlestkina EK, Röder MS, Efremova TT, Börner 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.
- Khlestkina EK, Giura A, Röder MS and Börner A (2009) A new gene controlling the flowering response to photoperiod in wheat. *Euphytica* 165: 579–585.
- Khvoiko VV (1909) Excavations in the square of the village Kritoborodintsakh, Letichevsk municipality, Podolsk region and near the village of Veremiye, Kiev municipality Tr. Mosk. Arkheolog. Obshch. T. XXII, vip. 2. M., pp. 281–309.
- Konovalova IG (1994) Export of grain from the North-Western Black Sea region in the XIV century and its importance for the economic development of the region. In: Dragnev D (ed.) *Evul Mediu Timpuriu în Moldova: Probleme de Istoriografie și Istoria Urbană*. Chișinău: Romanian Edition, pp. 108–125.
- Korzun VN, Röder MS, Ganai MW, Worland AJ and Law CN (1998) Genetic analysis of the dwarfing gene (*Rht8*) in wheat. Part I. Molecular mapping of *Rht8* on the short arm of chromosome 2D of bread wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics* 96: 1104–1109.
- Leonova I, Pestsova E, Salina E, Efremova T, Röder M, Börner A and Fischbeck G (2003) Mapping of the *Vrn-B1* gene in *Triticum aestivum* using microsatellite markers. *Plant Breeding* 122: 209–212.
- Maccaferri M, Sanguinetti MC, Donini P and Tuberosa R (2003) Microsatellite analysis reveals a progressive widening of the genetic basis in the elite durum wheat germplasm. *Theoretical and Applied Genetics* 107: 783–797.
- Majdrakov P (1945) For wheat from Pavlikeni. *Seed Production* 4: 132–141.
- Martynov S, Dobrotvorskaya T, Hon I and Faberova I (2006) Wheat pedigree and identified alleles of genes on Line. Crop Research Institute, Prague. Available at <http://genbank.vurv.cz/wheat/pedigree/>
- Mitrofanova OP, Strelchenko PP, Konarev AV and Balfourier F (2009) Genetic differentiation of hexaploid wheat inferred from analysis of microsatellite loci. *Russian Journal of Genetics* 45: 1351–1359.
- Newton AC, Akar T, Baresel JP, Bebeli PJ, Bettencourt E, Blade-nopoulos KV, Czembor JH, Fasoula DA, Katsiotis A, Koutis K, Koutsika-Sotiriou M, Kovacs G, Larsson H, Pinheiro de Carvalho MAA, Rubiales D, Russell JR, dos Santos TMM and Vaz Patto MC (2010) Cereal landraces for sustainable agriculture. A review. *Agronomy for Sustainable Development* 30: 237–269.
- Pinheiro de Carvalho MAA, Bebeli PJ, Bettencourt E, Costa G, Dias S, Dos Santos TMM and Slaski JJ (2013) Cereal landraces genetic resources in worldwide GeneBanks. A review. *Agronomy for Sustainable Development* 33: 177–203.
- R Core Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at <http://www.R-project.org/>
- Ribeiro-Carvalho C, Guedes-Pinto H, Igrejas G, Stephenson P, Schwarzacher T and Heslop-Harrison JS (2004) High levels of genetic diversity throughout the range of the Portuguese wheat landrace 'Barbela'. *Annals of Botany* 94: 699–705.
- Rogowsky PM, Guidet FLY, Langridge P, Shepherd KW and Koebner RMD (1991) Isolation and characterisation of wheat-rye recombinants involving chromosome arm 1DS of wheat. *Theoretical and Applied Genetics* 82: 537–544.
- Roussel V, Koenig J, Beckert M and Balfourier F (2004) Molecular diversity in French bread wheat accessions related to temporal trends and breeding programs. *Theoretical and Applied Genetics* 108: 920–930.
- Röder MS, Korzun V, Wendehake K, Plaschke J, Tixier M-H, Leroy P and Ganai MW (1998) A microsatellite map of wheat. *Genetics* 149: 2007–2023.
- Serpolay E, Dawson JC, Chable V, Lammerts Van Bueren E, Osman A, Pino S, Silveri D and Goldringer I (2011) Diversity of different farmer and modern wheat varieties cultivated in contrasting organic farming conditions in western Europe and implications for European seed and variety legislation. *Organic Agriculture* 1: 127–145.
- Smale M, with contributions from Aquino P, Crossa J, del Toro E, Dubin J, Fischer T, Fox P, Khairallah M, Mujeeb-Kazi A, Nightingale KJ, Ortiz-Monasterio I, Rajaram S, Singh R, Skovmand B, van Ginkel M, Varughese G and Ward R (1996) Understanding global trends in the use of wheat diversity and international flows of wheat genetic resources. Economics Working Paper 96-02. Mexico, D.F. CIMMYT.

- Šíp V, Chrpová J, Žofajová A, Pánková K, Užík M and Snape JW (2010) Effects of specific *Rht* and *Ppd* alleles on agronomic traits in winter wheat cultivars grown in middle Europe. *Euphytica* 172: 221–233.
- Stachel M, Lelley T, Grausgruber H and Vollmann J (2000) Application of microsatellites in wheat (*Triticum aestivum* L.) for studying genetic differentiation caused by selection for adaptation and use. *Theoretical and Applied Genetics* 100: 242–248.
- StatSoft (2005) Statistica 7 (StatSoft Inc.: Tulsa, OK). Available at [www.statsoft.com/textbook/](http://www.statsoft.com/textbook/)
- van de Wouw M, Kik C, van Hintum T, van Treuren R and Visser B (2009) Genetic erosion in crops: concept, research results and challenges. *Plant Genetic Resources: Characterization and Utilization* 8: 1–15.
- Ward JH Jr (1963) Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 58: 236–244.
- Yao J, Wang L, Liu L, Zhao C and Zheng Y (2009) Association mapping of agronomic traits on chromosome 2A of wheat. *Genetica* 137: 67–75.
- Zeven AC (1998) Landraces: a review of definitions and classifications. *Euphytica* 104: 127–139.
- Zhao C, Cui F, Fan Z, Li J, Ding A and Wang H (2013) Genetic analysis of important loci in the winter wheat backbone parent Aimengniu-V. *Australian Journal of Crop Science* 7: 182–188.