Determination of Population Structure of Wheat Core Collection for Association Mapping

L. Brbaklić[1]*, D. Trkulja[1], A. Kondić-Špika[1], S. Mikić[1], M. Tomičić[1] and B. Kobiljski[2]

[1]Institute of Field and Vegetable Crops, 30 Maksima Gorkog, 21000 Novi Sad, Serbia
[2]Biogranum, Novi Sad, Serbia

The microsatellites, as one of the most robust markers for identification of wheat varieties, were used for assessment of genetic diversity and population structure to promote effective use of genetic resources. In this study, the set of 284 wheat varieties were genotyped using 30 microsatellite markers. The chosen SSR markers were located among almost all linkage groups and covered all three genomes. The genotypes used originate from 24 different breeding centers worldwide and are included in an extensive core collection of the Institute of Field and Vegetable Crops in Novi Sad, Serbia. The total number of detected alleles was 349 at all analyzed loci. The average number of detected allelic variant per locus was 11.5. The mean value of polymorphic information content was 0.68. According to the probability of data obtained by program Structure, the results have shown presence of 6 subpopulations within the studied set of genotypes. The population structure positively correlated to some extent with geographic origin. The available pedigree data were included for additional explanation of population structure. The results of this study should provide valuable information for future association studies using the diverse wheat breeding material.

Keywords: *Triticum aestivum* L., structure, SSR markers, genetic diversity

Introduction

Wheat is one of the most important cereals worldwide. The most commonly grown species are *Triticum aestivum* spp. (bread wheat) and *Triticum turgidum* var. *durum* (durum wheat), which make 95% of the total wheat consumption (Randhawa et al. 2013). Intensifying of wheat breeding and selection processes during the last century has had a great impact on genetic diversity in the cultivated genotypes. Understanding genetic variability in existing plant material is important as it may indirectly reflect the level of genetic progress in future cultivars (Chen et al. 2012). Selection of crossing parents based on known genetic diversity and estimated population structure could be important for breeding and genetic resource conservation programs (Zhang et al. 2011).

* Corresponding author; E-mail: ljiljana.brbaklic@nsseme.com; Phone: +381 21 4898 223;
Fax: +381 21 4898 222

Exchange and exploitation of the wheat germplasm from different countries for implementation into existing breeding programs has led to the formation of very complex genetic structure in the newly generated elite lines. The characterization of population structure is critical to determine and correctly interpret the associations between functional and molecular diversity (Chen et al. 2012), i.e. finding important regions responsible for expression of the agronomically important traits (Couviour et al. 2011). In the last few years, application of association analysis directly on diverse germplasm (Neumann et al. 2011) for detection of QTLs has taken priority over using specially designed populations, which was time consuming and provided a low level of polymorphisms (Kobiljski et al. 2009).

The existence of structure in a certain population could cause apparent LD (linkage disequilibrium) between loci which are not linked. This seeming LD is a result of artificial selection through the breeding process, genetic drift and mating systems of specific species (Flint-Garcia et al. 2003; Yu and Buckler 2006). The main issue is to avoid incorrect population structure and thereby false associations, which is accomplished by developing and improving appropriate statistical methods and tools (Zhang et al. 2010).

Utilization of molecular markers represents reliable tool for assessment of genetic diversity and population structure in hexaploid wheat genotypes. Although microsatellites have been in use for decades, they are still one of the most powerful and robust molecular markers due to their higher level of genetic diversity in comparison to other codominant markers (Röder et al. 2002; Varshney et al. 2007).

The aim of this paper was to estimate genetic diversity and population structure based on molecular data by using model-based algorithm.

Materials and Methods

The set of 284 varieties were genotyped using 30 microsatellite markers. The chosen genotypes originate from 24 different countries and are a part of the biggest wheat core collection in Serbia, located at the Small Grains Department of the Institute of Field and Vegetable Crops in Novi Sad. The SSR markers located near to important agronomical QTLs were chosen according to the relevant literature data (Table 1). The primer sequences were taken from the data base – GrainGenes (http://wheat.pw.usda.gov/GG2/index.shtml). The microsatellite markers were positioned among all three genomes and almost all linkage groups.

The DNA was extracted from 284 genotypes (10 plantlets per genotype) using modified CTAB protocol (Doyle and Doyle 1990). The PCR reactions were carried out according to Röder et al. (1998) with specific annealing temperature depending on certain SSR markers. The total amount of PCR reaction was 10 µl, which contained 30 ng of genomic DNA, 1× buffer solution, 2 mM dNTPs, 1.5 mM $MgCl_2$, 10 pmol of each primers and 1 unit of Taq polymerase. The forward primers were labeled with one of the listed fluorescent dyes VIC, 6-FAM, NED or PET. The fragment analysis was performed using capillary electrophoresis at Genetic Analyzer ABI Prism 3130 (Applied Biosystems, Foster City, CA, USA). The total volume of reaction for electrophoresis was 10 µl and consisted of 2 µl of mixed PCR products, 0.2 µl of Gene Scan 500 LIZ standards and 7.8 µl of Hi-Di

Table 1. The chosen microsatellite markers with chromosome position and sequences of forward and reverse primers

No. markers	SSR markers and chromosomes	Sequences of forward primers	Sequences of reverse primers
1	WMC28 (5B)	5' ATC ACG CAT GTC TGC TAT GTA T 3'	5' ATT AGA CCA TGA AGA CGT GTA T 3'
2	CFD65 (1A)	5' AGA CGA TGA GAA GGA AGC CA 3'	5' CCT CCC TTG TTT TTG GGA TT 3'
3	GWM292 (5D)	5' TCA CCG TGG TCA CCG AC 3'	5' CCA CCG AGC CGA TAA TGT AC 3'
4	WMC167 (2D)	5' AGT GGT AAT GAG GTG AAA GAA G 3'	5' TCG GTC GTA TAT GCA TGT AAA G 3'
5	GWM271 (5B)	5' CAA GAT CGT GGA GCC AGC 3'	5' AGC TGC TAG CTT TTG GGA CA 3'
6	CFA2257 (7A)	5' GAT ACA ATA GGT GCC TCC GC 3'	5' CCA TTA TGT AAA TGC TTC TGT TTG A 3'
7	WMC83 (7A)	5' TGG AGG AAA CAC AAT GGA TGC C 3'	5' GAG TAT CGC CGA CGA AAG GGA A 3'
8	GWM148 (2B)	5' GTG AGG CAG CAA GAG AGA AA 3'	5' CAA AGC TTG ACT CAG ACC AAA 3'
9	PSP3200 (6D)	5' GTT CTG AAG ACA TTA CGG ATG 3'	5' GAG AAT AGC TGG TTT TGT GG 3'
10	GWM181 (3B)	5' TCA TTG GTA ATG AGG AGA GA 3'	5' GAA CCA TTC ATG TGC ATG TC 3'
11	CFA2086 (2A)	5' TCT ACT TTC AGG GCA CCT CG 3'	5' TCT CTC CAA ACC TCC CTG TAA 3'
12	WMC170 (2A)	5' ACA TCC ACG TTT ATG TTG TTG C 3'	5' TTG GTT GCT CAA CGT TTA CTT C 3'
13	GWM11 (1B)	5' GGA TAG TCA GAC AAT TCT TGT G 3'	5' GTG AAT TGT GTC TTG TAT GCT TCC 3'
14	WMC44 (1B)	5' GGT CTT CTG GGC TTT GAT CCT G 3'	5' TGT TGC TAG GGA CCC GTA GTG G 3'
15	CFA2114 (6A)	5' ATT GGA AGG CCA CGA TAC AC 3'	5' CCC GTC GGG TTT TAT CTA GC 3'
16	BARC101 (2B)	5' GCT CCT CTC ACG ATC ACG CAA AG 3'	5' GCG AGT CGA TCA CAC TAT GAG CCA ATG 3'
17	WMC216 (1D)	5' ACG TAT CCA GAC ACT GTG GTA A 3'	5' TAA TGG TGG ATC CAT GAT AGC C 3'
18	GWM294 (2A)	5' GGA TTG GAG TTA AGA GAG AAC CG 3'	5' GCA GAG TGA TCA ATG CCA GA 3'
19	WMC166 (7B)	5' ATA AAG CTG TCT CTT TAG TTC G 3'	5' GTT TTA ACA CAT ATG CAT ACC T 3'
20	WMC18 (2D)	5' CTGGGGCTTGGATCACGTCATT 3'	5' AGCCATGGACATGGTGTCCTTC 3'
21	CFD71 (4A)	5' CAATAAGTAGGCCGGGACAA 3'	5' TGTGCCAGTTGAGTTTGCTC 3'
22	GWM157 (2D)	5' GTCGTCGCGGTAAGCTTG 3'	5' GAGTGAACACACGAGGCTTG 3'
23	WMC144 (2D)	5' GGACACCAATCCAACATGAAACA 3'	5' AAGGATAGTTGGGTGGTGCTGA 3'
24	WMC407 (2A)	5' GGTAATTCTAGGCTGACATATGCTC 3'	5' CATATTTCCAAATCCCCAACTC 3'
25	WMC333 (1A)	5' TCAAGCATAGGTGGCTTCGG 3'	5' ACAGCAGCCTTCAAGCGTTC 3'
26	WMC333 (6A)	5' TCAAGCATAGGTGGCTTCGG 3'	5' ACAGCAGCCTTCAAGCGTTC 3'
27	GWM160 (4A)	5' TTCAATTCAGTCTTGGCTTGG 3'	5' CTGCAGGAAAAAAAGTACACCC 3'
28	WMC154 (2B)	5' ATGCTCGTCAGTGTCATGTTTG 3'	5' AAACGGAACCTACCTCACTCTT 3'
29	GWM219 (6B)	5' GATGAGCGACACCTAGCCTC 3'	5' GGGGTCCGAGTCCACAAC 3'
30	GWM368 (4B)	5' CCATTTCACCTAATGCCTGC 3'	5' AATAAAACCATGAGCTCACTTGC 3'

formamide for denaturation of double-stranded products. The electrophoresis data were collected in Data Collection Software v.3.0. and analyzed using GeneMapper Software v.4.0 (Applied Biosystems, Foster City, CA, USA).

Statistical analysis

Genetic diversity was analyzed by using the PowerMarker software v.3.25 (Liu and Muse 2005). The model-based method was chosen to quantitatively estimate structural membership proportions and account for admixture in individuals that may be assigned to more than one sub-population. Population structure was estimated using the Bayesian algorithm based on molecular data implemented in program Structure v. 2.3.4 (Pritchard et al. 2000). The simulation included admixture model with independent allele frequencies. The length of the burn-in and MCMC (Markov Chain Monte Carlo) were the same for both and appointed as 100,000. The number of iterations was recorded as 3 and the hypothetical number of cluster was set from K = 1 to K = 20. The choice of the numbers of clusters K was determined by comparing log probabilities of data Pr [X|K]. The available pedigree data for genotypes were included (data not shown) and were taken from ECPGR pedigree database (http://genbank.vurv.cz/ewdb/).

Results

A total of 284 genotypes were analyzed with 30 polymorphic SSR markers. The total number of detected alleles was 349 in all analyzed loci. The maximum value of detected allelic form (142) was found in genome B, whereas genome D covered minimum number (73) of allelic variants. The average number of determined alleles per loci ranged from 10.4 in genome D to 12.9 in B genome. The most important parameter of genetic diversity in all investigated loci was the polymorphic information content (PIC). It was calculated for each locus, whereas the average PIC value for the all analyzed genotypes was 0.68 (Table 2).

Table 2. The basic parameters of population genetics obtained with molecular analysis in program PowerMarker v.3.25

Genetic diversity	A genome	B genome	D genome	Average/Total
No. of investigated loci	12	11	7	30
No. of alleles per locus	11.2	12.9	10.4	11.5
PIC value	0.72	0.74	0.57	0.68

The model based method was applied for assessment of population structure of the chosen genotypes. According to Structure output, the probability of data has shown the presence of six subpopulations (Fig. 1). The largest cluster (light blue) consisted of 114 genotypes with approximately 80% varieties originating from Serbia. The second largest subpopulation (pink) contained 45 genotypes almost all originating from France and Great Britain. The next three groups consist of similar number of genotypes (40, 37 and 30)

representing American (red), the rest of the Serbian varieties with genotypes of various origin (dark blue) and more than a half of Croatian varieties (yellow), respectively. The smallest cluster (green) was composed of 18 mostly American wheat cultivars.
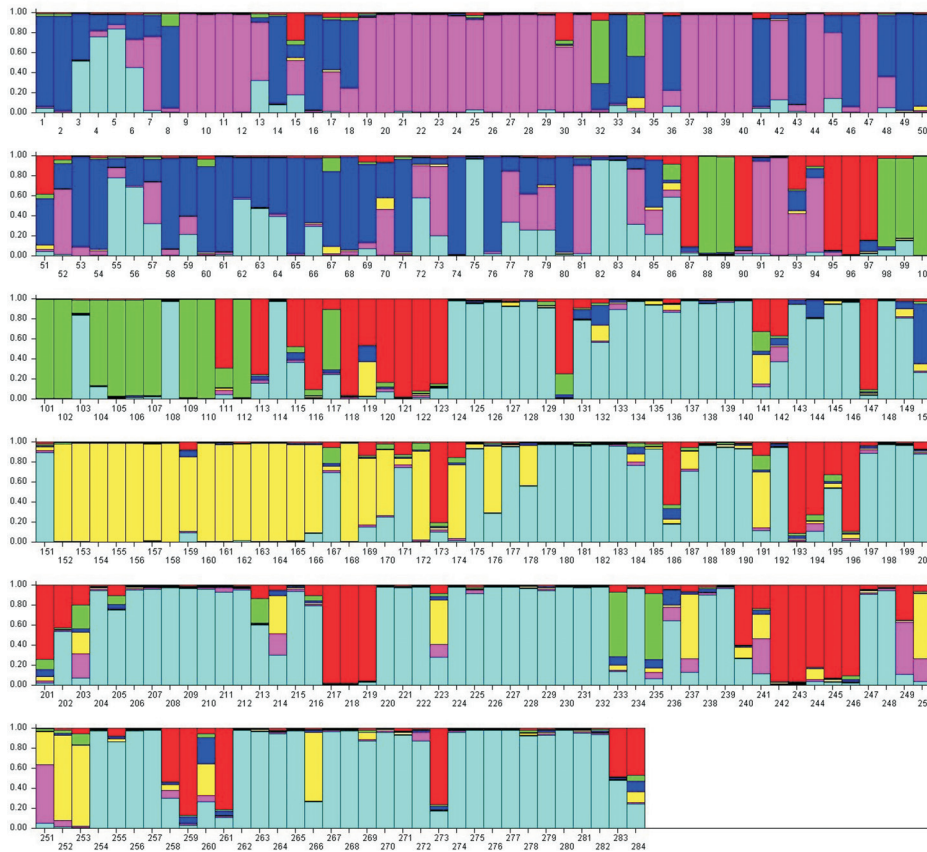

Figure 1. The distribution of genotypes into six subpopulations labeled with different colours (dark blue, pink, green, yellow, light blue, red) estimated based on molecular data in program STRUCTURE

Discussion

The investigation of genetic diversity in materials originating from specific regions and countries (Stachel et al. 2000; Chen et al. 2003) demonstrated extremely low values of mean alleles per locus and level of polymorphism. Wang et al. (2013) reported 6.32 alleles per locus in 92 investigated wheat varieties and related species using 44 SSR markers. Chao et al. (2007) reported on average 7.2 alleles per locus among the US wheat accessions, lower than 11.5 alleles determined in our research. This could be caused by the detection system with low resolution or breeding selection on the investigated materials,

which lead to reduction in genetic diversity. Röder et al. (2002) found 10.5 alleles per locus in 502 European varieties, which indicates that our initial material demonstrates sufficient level of genetic diversity. The PIC value of the loci ranged from 0.57 in D genome to 0.74 in B genome. Wang et al. (2013) found similar PIC values from 0.36 to 0.88 with mean PIC value of 0.75 using more SSR markers. Huang et al. (2002) reported that the highest PIC value was found in B, followed by A and D genomes. The same results were obtained in our material as evidence that D genome possesses the lowest level of polymorphism in hexaploid wheat genotypes. The investigated cultivars of different origin had a great level of polymorphism based on the revealed alleles and PIC values despite relatively small number of SSR markers. The chosen genotypes were adequate and have sufficient level of diversity, which enables further explorations of this set of genotypes for association mapping for traits of interest.

Structure of wheat populations worldwide is still not well known. Our understanding of the population structure is hampered by the increase of germplasm exchange between breeding centers causing changes in the historical structure of genetic diversity (Couviour et al. 2011). Moreover, population structure is greatly affected by many factors such as breeders activities, years of release, environmental conditions etc. In our study, model-based algorithm clustered genotypes into 6 subpopulations which could partially be explained by geography origin. Most of cultivars from the same country were grouped together. Yet, some genotypes with different origin were assigned into the same subpopulations indicating the possible presence of common or related ancestors in their pedigrees. The available information of pedigree data could help to identify frequently used parents for selection in different breeding centers due to extensive exchange of breeding materials. The varieties dominated in the largest group were developed from a Russian variety Bezostaja 1, as a parent, which was often used in breeding programs in East and Central Europe and was carrier of Rht8 and PpD1 genes. Likewise, the genotypes grouped in the fourth cluster with predominant Croatian varieties represented descendants of Mara and Libellula, two Italian varieties that had a great impact on wheat breeding in former Yugoslavia. These cultivars were improved for the local climatic conditions and industry requirements (Borojevic and Borojevic 2005). Our findings confirmed that population structure and genotype relatedness assessed with molecular markers are consistent with genotype origin and available pedigree data.

References

Borojevic, K., Borojevic, K. 2005. Historic role of the wheat variety Akakomughi in Southern and Central European wheat breeding programs. Breeding Sci. **55**:253–256.

Chao, S.M., Zhang, W.J., Dubcovsky, J., Sorrels, M. 2007. Evaluation of genetic diversity and genome-wide likage disequilibrium among U.S. wheat (*Triticum aestivum* L.) germplasm representing different market classes. Crop Sci. **47**:1018–1030.

Chen, X., Min, D., Yasir, T.A., Hu, Y.G. 2012. Genetic diversity, population structure and linkage disequilibrium in elite Chinese winter wheat investigated with SSR markers. PLoS one **7**(9):e44510. doi:10.1371/journal.pone 0044510.

Chen, X.M., He, Z.H., Shi, J.R., Xia, L.Q., Ward, R., Zhou, Y., Jiang, G.L. 2003. Genetic diversity of high quality winter wheat varieties (lines) based on SSR marker. Acta Agronomica Sinica **29**:13–19.

Couviour, F.L., Faure, S., Poupard, B., Flodrops, Y., Dubreuil, P., Praud, S. 2011. Analysis of genetic structure in a panel of elite wheat varieties and relevance for association mapping. Theor. Appl. Genet. **123**:715–727.

Doyle, J.J., Doyle, J.L. 1990. Isolation of plant DNA from fresh tissue. Focus **12**:13–15.

Flint-Garcia, S.A., Thornsberry, J.M., Buckler, E.S. 4th 2003. Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. **54**:357–374.

Huang, X.Q., Börner, A., Röder, M.S., Ganal, M.W. 2002. Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. Theor. Appl. Genet. **105**:699–707.

Kobiljski, B., Dencic, S., Kondic-Spika, A., Lohwasser, U., Börner, A. 2009. Locating stable across environment QTL involved in the determination of agronomic characters in wheat. Cereal Res. Commun. **37**:327–333.

Liu, K., Muse, S.V. 2005. PowerMarker: An integrated analysis environment for genetic marker analysis. Bioinformatics **21**:2128–2129.

Neumann, K., Kobiljski, B., Dencic, S., Varshney, R.K., Börner, A. 2011. Genome-wide association mapping: A case study in bread wheat (*Triticum aestivum* L.). Mol. Breed. **27**:37–58.

Pritchard, J.K., Stephens, M., Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics **155**:945–959.

Randhawa, H.S., Muhammad, A., Pozniak, C., Clarke, J.M., Graf, R.J., Fox, S.L., Humphreys, D.G., Knox, R.E., Depauw, R.M., Singh, A.K., Cuthbert, R.D., Hucl, P., Spaner, D. 2013. Application of molecular markers to wheat breeding in Canada. Plant Breed. **132**:458–471.

Röder, M.S., Korzun, V., Wendehake, K., Plaschke, J., Tixier, M-H., Leroz, P., Ganal, M.W. 1998. A microsatellite map of wheat. Genetics **194**:2007–2023.

Röder, M.S., Wendehake, K., Korzun, V., Bredemeijer, G., Laborie, D., Bertrand, L., Isaac, P., Rendell, S., Jackson, J., Cooke, R.J., Vosman, B., Ganal, M. 2002. Construction and analysis of a microsatellite-based database of European wheat varieties. Theor. Appl. Genet. **106**:67–73.

Stachel, M., Lelley, T., Grausgruber, H., Vollmann, J. 2000. Application of microsatellites in wheat (*Triticum aestivum* L.) for studying genetic differention caused by selection for adaptation and use. Theor. Appl. Genet.**100**:242–248.

Varshney, R.K., Thudi, M., Aggarwal, R., Börner, A. 2007. Genic molecular markers in plants: Development and applications. In: Varshney, R.K., Tuberosa, R. (eds), Genomics-assisted Crop Improvement: Genomics Approaches and Platforms. Vol. 1. Springer, Dordrecht, The Netherlands, pp. 13–29.

Wang, Y., Wang, C., Zhang, H., Yue, Z., Liu, X., Ji, W. 2013. Genetic analysis of wheat (*Triticum aestivum* L.) and related species with SSR markers. Genetic Res. and Crop Evol. **60**:1105–1117.

Yu, J., Buckler, E.S. 2006. Genetic association mapping and genome organization of maize. Current Opinion in Biotechnol. **17**:155–160.

Zhang, D., Bai, G., Zhu, C., Yu, J., Carver, B.F. 2010. Genetic diversity, population structure, and linkage disequilibrium in U.S. elite winter wheat. The Plant Genome **3**:117–127.

Zhang, L., Liu, D.C., Guo, X.L., Yang, W.L., Sun, J.Z., Wang, D.W., Sourdille, P., Zhang, A.M. 2011. Investigation of genetic diversity and population structure of common wheat cultivars in northern China using DArT markers. BMC Genetics **12:**42.