HAPLOTYPE DIVERSITY IN GENES RESPONSIBLE FOR DROUGHT STRESS RESPONSE IN MAIZE

Ana NIKOLIĆ[1]*, Jelena VANČETOVIĆ[1], Violeta ANĐELKOVIĆ[1], Ksenija MARKOVIĆ[1], Elena TODOROVSKA[2], Borislav ASSENOV[2], Dragana IGNJATOVIĆ-MICIĆ[1]

[1]Maize Research Institute Zemun Polje, S. Bajića 1, 11180, Belgrade, Serbia
[2]AgroBioInstitute, 8 Dragan Tsankov, 1164 Sofia, Bulgaria
*Corresponding author: anikolic@mrizp.rs

Abstract

Great efforts have been made over the past several decades for development of improved cultivars adapted to different agro-ecological areas due to the on-going climatic changes. In order to increase further the selection gain and to accelerate breeding processes in maize, a profound knowledge is required regarding genes and genomic regions encoding for agronomically important traits. In this context, the use of haplotypes could improve selection of quantitative traits with a low heritability due to strong environmental influence. Twenty temperate drought tolerant (15) and sensitive (5) maize inbred lines from Maize Research Institute Zemun Polje (MRIZP) were subjected to SNP genotyping. Additionally, 17 maize (13) and teosinte (4) genotypes were selected for comparison from Panzea database to represent the functional diversity of maize. For SNP identification direct PCR sequencing of eight abiotic responsive candidate genes was done. A small number (3 to 8) of distinct and highly diverse haplotypes were observed in all eight (8) marker genes. Haplotype analysis based on the SNPs revealed the highest haplotype diversity in *MYBR96* (0.817) and the lowest in *MYB8* (0.3235) gene. Network analysis showed a linear relationship between haplotypes for some genes, while for the rest of genes the network graphs reflected more complex relationships between a large numbers of haplotypes. The deployment of the identified haplotypes could be a powerful complementary tool to improve accuracy and efficiency of modern breeding strategies such as marker assisted selection and genomic selection for developing drought tolerant maize genotypes.

Keywords: maize, drought, SNPs, haplotypes.

Introduction

Global crop production is challenged by severe climatic changes like drought (Lesk *et al.*, 2016) and food demands of exponentially growing world population. It has been predicted that yields of important commodity crops need to be increased by almost 40% by the middle of 21st century (Tester and Langridge, 2010). Achieving sustainable yields of commodity crops in inconsistent environments represents a demanding task for the breeders. However, cutting-edge advances in genome analysis technologies provide high resolution molecular information which can perfect advanced quantitative genetic approaches. Low-cost genotyping tools that can capture sequence variation are now available for all agronomically important plant species (Huang and Han, 2014) and they provide an effective means for crop

genetic research studies (Ganal *et al.*, 2012), such as providing a basis for genomic selection (GS) or prediction of hybrid performance.

High-throughput single nucleotide polymorphism (SNP) marker arrays or SNPs detected by DNA sequencing are the genotyping markers of choice for crop genomic research. However, their major limitation is that they provide only bi-allelic information at a locus, hence their information content compared to multi-allelic markers is low, limiting the resolution at which SNP–trait relationships can be delineated. An effective approach to surpass the biallelic limitations of SNPs is to construct haplotypes based on linkage disequilibrium (LD), one of the most important features influencing genetic analysis of crop genomes (Qian *et al*., 2017). A haplotype is described as „two or more SNP alleles that tend to be inherited as a unit" (Bernardo, 2010).

Marker-assissted selection (MAS), which has succesfully been used for mono- or oligogenic inherited traits, failed for highly quantitative traits with a low heritability due to strong environmental influence. MAS strategies proved inadequate due to statistical overestimation of QTL linked markers or complex genetic architectures for most important traits (Bernardo, 2008). That is why the concept of GS with densely spaced genome-wide markers is presently being adopted for many crop breeding programs. Genomic selection methods apply the concept that the breeding value of an individual which has not been phenotyped can be estimated purely on the basis of its genome wide marker profile. In the context of both MAS and GS, the use of haplotypes could be a powerful complementary tool to improve their accuracy and efficiency. Because of their increased information content compared to bi-allelic SNP markers, fitting haplotypes with statistically significant trait associations to phenotypes as fixed effects in GS models could further improve prediction accuracies. The use of haplotype-assisted GS should more accurately depict the complex relationships between genotypic information and phenotypes than single SNPs alone are able to do; hence, this approach could ultimately help to further increase selection gain per unit of time (Qian *et al*., 2017).

The aim of the research presented herein was to identify SNP mutations in drought candidate responsive genes in 20 drought tolerant and susceptible maize inbred lines from Maize Research Institute Zemun Polje (MRIZP) and based on the SNP information identify conserved haplotypes. Also, the nucleotide diversity in these genes between the inbreds from MRIZP collection and Panzea database was analysed.

Material and Methods

Twenty temperate drought tolerant (15) and sensitive (5) maize inbred lines from MRIZP were subjected to SNP genotyping. Four sensitive and two tolerant inbreds are commercial lines used as parental lines for development in MRIZP breeding programs, while the remaining 14 are from MRIZP genebank. Additionally, a total of 17 foreign maize genotypes including tropical maize inbreds (5), teosinte lines (4), African inbred M37W and temperate maize inbreds (7) were selected from Panzea database to represent the functional diversity of maize. Most of them (without teosinte lines) have been used in crosses to develop the Nested Association Mapping (NAM) population consisting of 302 lines designed to capture the diversity of maize and to preserve historic linkage disequilibrium.

For SNP identification in 20 inbred lines from MRI direct PCR sequencing approach was applied on eight abiotic responsive candidate genes - *MYBR96* (MYB-related-transcription factor 96)*, SDG110a* (histone-lysine N-methyltransferase, H3 lysine-36 and H4 lysine-20 specific), *IDP507* (protein kinase G11A)*, MYB67* (myb domain protein 67) *MYB8, PCO089553b (PSF2), GA20OX1* and *DHN1*. PCR products were subjected to double stranded sequencing on ABI 3130xl platform. The raw sequencing data were base called and

assembled in contigs using phred (Ewing *et al*., 1998) and phrap (Green, 1996) with default parameters. The contigs were aligned and analysed for SNP detection using BioLign software (Hall, 2001). Nucleotide sequences were converted into amino acid sequences and were compared with protein sequences (http://www.ncbi.nlm.nih.gov; http://www.panzea.org/) of the corresponding genes in maize to verify if the identified SNPs were able to produce functional mutations with amino acid changing. The sequence information is given in forward direction.

Conserved haplotypes, representing DNA sequences containing identical allelic variants at all polymorphic sites at particular locus, but originating from separate individuals, were identified visually or by application of NETWORK 5.0 software. Haplotype networks for the whole set of 35 maize lines but separately for each gene were drawn using the same software. The network figures show the number of haplotypes observed for each gene and the SNP position which separates each haplotype from each other. Haplotype frequency is depicted by circles. Circle sizes correspond to the frequency of the corresponding haplotypes. The bigger the circle, the more genotypes are represented by that haplotype. The color in the circles corresponds to the affiliation of genotypes to the relevant collection.

Results and Discussion

Haplotype analysis of MRIZP and Panzea collections based on the SNPs from the small window size, i.e. gene window (all markers within a gene) is presented in Fig. 1. Number of identified SNPs and haplotypes, as well as haplotype diversity, for each gene is given in Table 1.

The highest haplotype diversity was found in *MYBR96* (0.817) and the lowest in *MYB8* (0.3235) in the whole set of maize inbred lines (Table 1). Comparison analysis between both collections showed slightly higher level of haplotype diversity in *MYB8* (0.4902) and *MYBR96* (0.817) in MRIZP collection consisting of only temperate inbreds. The level of haplotype diversity in the rest of genes was higher in Panzea collection represented by an extremely diverse set of temperate, tropical and teosinte lines. Removal of tropical inbreds and teosinte lines led to some reduction in the level of haplotype diversity in Panzea collection, but it still remained higher in comparison to the MRIZP collection in genes *DHN1, IDP 507, MYB67* and *GA20OX1* (data not shown).

Table 1. Haplotype diversity of eight abiotic responsive candidate genes in MRIZP and PANZEA maize collections

genes	Collection (inbred lines)					
	MRIZP			PANZEA		
	SNP number	haplotype number	haplotype diversity	SNP number	haplotype number	haplotype diversity
DHN1	7	3	0.3856	18	7	0.8088
IDP507	5	3	0.3922	5	4	0.6544
MYB67	5	4	0.3987	14	7	0.7206
MYB8	4	6	0.4902	2	3	**0.3235**
MYBR96	13	6	**0.817**	16	6	0.8015
PCO089553B	3	3	0.6013	4	4	0.6397
GA20OX1	3	3	0.3856	7	7	0.6618
SDG110a	10	6	0.6993	4	5	0.7721

In this study a small number of 3 to 8 distinct and highly diverse haplotypes were observed in all eight genes in the set of 35 genotypes from both collections. These results are in accordance with those observed by Ching *et al*. (2002) for 36 elite US maize inbreds, but

lower than the identified 14 haplotypes in the RTCS (rootles concerning crown and seminal roots) in 73 Chinese elite inbreds from five temperate heterotic groups and some tropic germplasm.

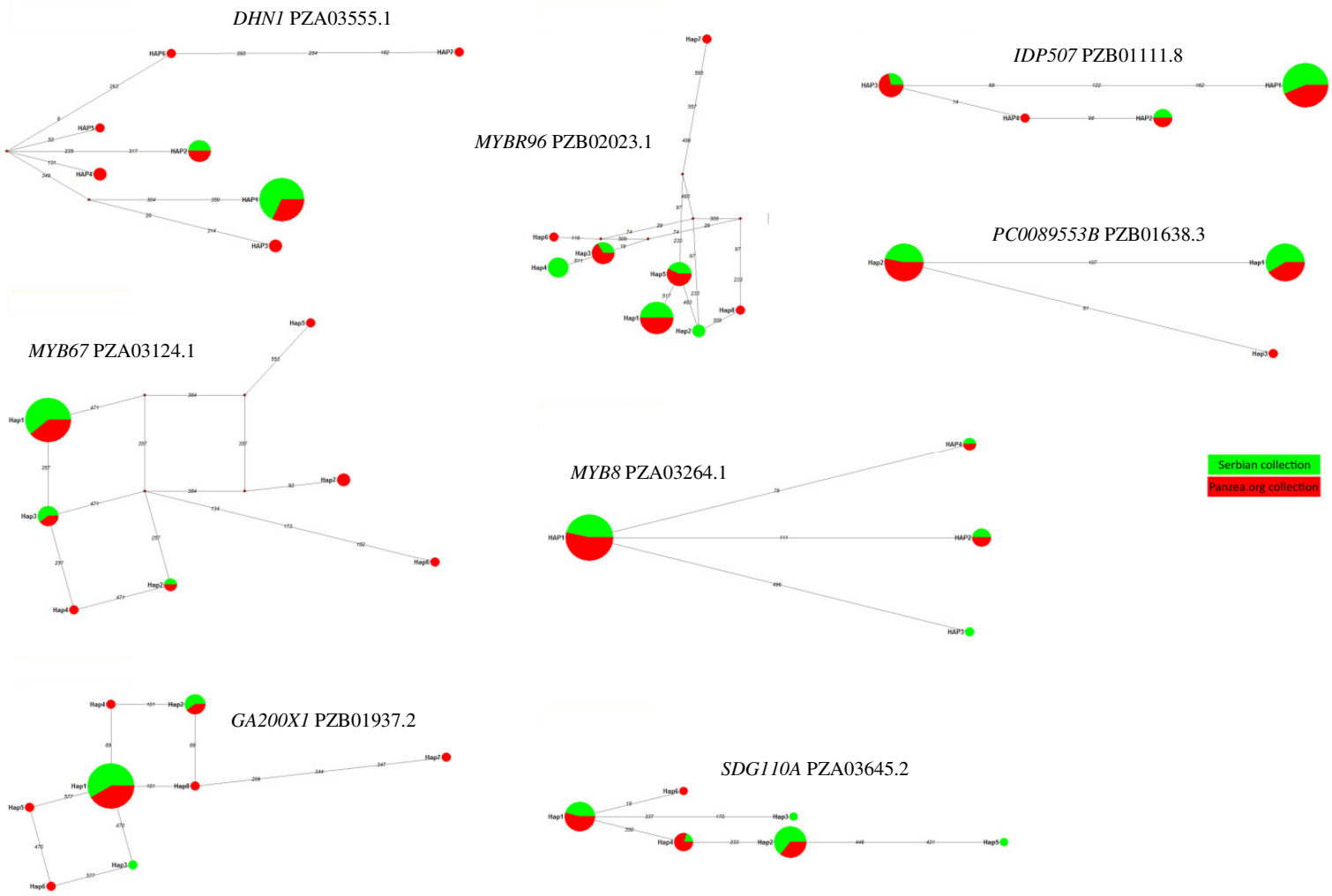

Fig. 1. Haplotype network of the eight marker genes developed on affiliation of maize genotypes to MRIZP and Panzea collections. Each circle represents a haplotype and is labeled accordingly. The two color codes, green and red in the circles represents the genotypes belonging to the two collections, MRIZP and Panzea respectively.

Network analysis showed a linear relationship between haplotypes for some genes such as *IDP507, PCO089553b*, *MYB8* and *SDG110a*, while for *DHN1, MYB67, MYBR96* and *GA20OX1* the network graphs reflected more complex relationships between a large number of haplotypes. This was more definitely pronounced in *MYBR96* gene where the presence of eight haplotypes was identified in the set of 35 maize genotypes. The haplotype network of *MYBR96* gene showed the presence of four major haplotypes (Hap1, Hap3, Hap4 and Hap5), which are connected to four minor ones by 1-2 SNPs. It is interesting to note, that one of the major haplotypes Hap4 is composed only of five temperate inbreds from MRIZP elite mini core collection for drought tolerance. In these lines a unique mutation (A→G transition) was observed in the acidic Ser/Thr – rich area located downstream from the SANT domain (SWI3, ADA2, N-Cor, TFIIIB DNA binding domain). This transition leads to amino acid change from Tryptophan to Alanine at the corresponding position of the protein and despite the fact that it does not affect the binding site of *MYBR96* transcription factor, this replacement might possibly reflect the spatial conformation through changing of its functional activity. To our knowledge the observed mutation is the first one reported for this position of the *MYBR96* gene as compared to the 44 other inbred lines representing the functional diversity of maize from the Panzea database (http://www.panzea.org/) (Assenov *et al*., 2013).
The simplest haplotype network consisting of only three haplotypes was observed for the marker gene *PCO0089553B (PSF2)*, one of the core genes that encode proteins which have an important role in DNA replication machinery in plants. The two major haplotypes include equal number of genotypes and are split by the marker SNP PZB01638.3 (A/T) in position 107, which is located in intron1 of the gene. This marker SNP was associated with different phenotypic traits (anthesis-silking interval, grain yield and selection index of drought tolerance) under water stress conditions at four environments in China (Hao *et al*., 2011). However, in our analysis no obvious differentiation of genotypes according to their stress response was observed.

Conclusion

Our study showed unbalanced distribution of the maize inbred lines in the analyzed haplotypes, most of which consisted of only teosinte and/or CML germplasm. The largest haplotypes consisted of lines from both MRIZP and Panzea collections. The analyzed haplotypes of the selected eight abiotic responsive genes did not comprise lines according to their level of tolerance to the drought stress, except Hap4 in *MYBR96* which was composed exclusively of five MRIZP drought tolerant lines with a unique mutation and it could be used in breeding programs for developing drought tolerant genotypes. A more profound analysis of genes expressed under drought stress in search for informative SNPs and haplotypes valuable for marker assisted or genome selection is underway.

References

Assenov B., Andjelkovic V., Ignjatovic-Micic D., Vancetovic J., Nikolic A., Christov N. K., Tsonev S., Abu-Mhadi N., Vassilev D., Muhovski Y., Ilchovska M., Todorovska E. (2013). Identification of SNP Mutations in MYBE-1 Gene Involved in Drought Stress Tolerance in Maize. Bulgarian Journal of Agricultural Science, 19 (No 2): 181-185.

Bernardo R. (2008). Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Sci., 48:1649.

Bernardo R.N. (2010). Breeding for Quantitative Traits in Plants. Woodbury, MN: Stemma Press.

Ewing B., Hillier L., Wendl M.C., Green P. (1998). Base-Calling of Automated Sequencer Traces UsingPhred.I. Accuracy Assessment. Genome Res., 8: 175-185.

Ching A., Caldwell K. S, Jung M., Dolan M., Smith O.S.H, Tingey S., Morgante M. and Rafalski A. J. (2002). SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genetics, 3:19

Ganal M. W., Polley A., Graner E. M., Plieske,J., Wieseke R., Luerssen H., et al. (2012). Large SNP arrays for genotyping in crop plants. J. Biosci. 37, 821–828.

Green P. (1996). http://bozeman.mbt.washington.edu/phrap.docs/phrap.html/

Hall T. (2001). Biolign alignment and multiple contig. Editor. http://en.bio-soft.net/dna/BioLign.html

Hao Z., Li X., Xie C., Weng J., Li M., Zhang D., Liang X., Liu L., Liu S., Zhang S. (2011). Identification of Functional Genetic Variations Underlying Drought Tolerance in Maize Using SNP Markers. Journal of Integrative Plant Biology, 53 (8): 641–652

Huang X., Han, B. (2014). Natural variations and genome-wide association studies in crop plants. Annu. Rev. Plant Biol., 65: 531–551.

Lesk C., Rowhani P., Ramankutty N. (2016). Influence of extreme weather disasters on global crop production. Nature, 529: 84–87.

Qian L., Hickey L.T., Stahl A., Werner C.R., Hayes B., Snowdon R.J., Voss-Fels K.P. (2017). Exploring and harnessing haplotype diversity to improve yield stability in crops. Front. Plant Sci., 8:1534.

Tester M., Langridge P. (2010). Breeding technologies to increase crop production in a changing world. Science, 327: 818–822.