THE IDENTIFICATION OF DROUGHT TOLERANT MAIZE ACCESSIONS BY TWO-STEP CLUSTER ANALYSIS

Vojka Babic[1*], Jelena Vancetovic [1], Slaven Prodanovic[2], Violeta Andjelkovic[1], Milosav Babic[1], and Natalija Kravic[1]

[1]Maize Research Institute, Zemun Polje, Slobodana Bajica 1, 11185 Belgrade, Serbia.
*Corresponding author. E-mail: vbabic@mrizp.rs

[2]University of Belgrade, Faculty of Agriculture, Nemanjina 6, 11080 Belgrade, Serbia

ABSTRACT

Water deficit is one of the main limiting factors leading to maize yield reduction in the temperate zone. Global environmental changes additionally put breeding for drought tolerance in the focus of research programmes. In improving drought tolerance of elite hybrids, the first step is identification of drought tolerance sources within the existing gene bank collections. The complete gene bank collection of the Maize Research Institute, Zemun Polje, Serbia, was field tested under extreme water deficit in Egypt, with the aim of identifying tolerant genotypes. Six quantitative traits and a general estimation of the plant habitus during the growing season were estimated. Genotypes were classified by two-step cluster analysis. In the first phase of data analysis the classification was done with four quantitative and one qualitative trait, genotypes being classified into two clusters. In the second phase when the analysis included the duration of the growing season, genotypes were classified into three clusters. The first phase of data analysis pointed out to 231 drought tolerant genotypes, while the second phase data analysis further added a certain number of early and medium early genotypes, giving a total of 558 accessions. Further work on the selected accessions will be continued for the formation of the core collection and direct use in breeding.

Key words: drought tolerance, maize, two-step cluster analysis.

INTRODUCTION

Drought can be considered as a permanent and significantly high deficit of water necessary for plants, which limits life processes of plants in an agricultural or a forest region (Monneveux and Belhassen, 1996). It adversely affects environments, limits food and feed production and it is a serious problem for world's agriculture. The prediction of global climatic changes indicates warming up of the atmosphere in the 21st century. The expected mean air temperature at the end of this century would be higher by 1.4^{o}C to 5.8^{o}C. According to the initial estimations, in the region of southern Europe (to which Serbia belongs), not only the increase in air temperature and evaporation, but also the decrease in precipitations in the warm half of the year, coupled with the reduction in water flow, soil moisture and available water have been expected. In addition, beside changes in mean values of climatic elements, frequent occurrences of climatic extremes (storms, accompanied by floods, drought, extremely high air temperatures, heat waves, etc.) have been predicted. Hence, due to unfavourable effects on the production of food, energy, water supply, human health and biological diversity, the region of southern Europe has been classified into the regions severely endangered by climatic changes (Dodig et al., 2002).

Maize is third crop in the world according to distribution and utilisation. Drought drastically reduces its grain yield. For a successful breeding for drought tolerance in maize it is necessary to have a high degree of expression of a target trait, but also it is important to have a high degree of adjustment to the local growing conditions. Drought tolerance is a polygenically inherited complex trait and is related to many physiological and morphological traits. With regard to this, intensive studies within the field of physiology, agronomy, breeding, and lately biotechnology, have been carried out (Ribaut and Rigot, 2007). Significant progress in yield increase under drought conditions has been achieved by conventional breeding, but this

process is long-term, difficult and requires strictly controlled field conditions. Moreover, stress reduces genetic variance, heritability of yield components (Ribaut et al., 1997) and makes breeding process more difficult.

Gene banks around the world hold collections of the genetic resources of crop plants with a large number of accessions. These accessions are often described according to morphological descriptors or by other traits of interests for breeders. Screening, characterization and evaluation of genetic resources are considered priorities in breeding, as such information is crucial in choosing material for the incorporation into breeding activities (Tucak et al., 2009).

The Maize Research Institute "Zemun Polje" (MRI) collection encompasses over 6000 accessions, very divergent by their origin. The aim of this study was to select genotypes from the collection that were drought tolerant, which would be the first step for the selection material for creating a core collection.

The two-step cluster analysis was used for data analysis. The algorithm used in this procedure is advantageous compared to a more often applied hierarchical cluster analysis: 1) it has ability to form clusters based on categorical and continuous variables; 2) selection of the number of clusters is done automatically; 3) it has ability to successfully analyse enormous data bases.

MATERIAL AND METHODS

The study was performed with the collection of 6328 maize genotypes: 2217 landraces, collected from all agroecological habitats in former Yugoslavia and 4165 introduced genotypes (2830 inbred lines and 1335 populations). The field experiments were set up in the location of Setz, in the Nile valley, 120 km south of Cairo, in Egypt. Based on the Zemun Polje data, the genotypes were classified into the five maturity groups: I - Extra-early, II - Early, III - Medium, IV - Medium-late and V - Late. There were 264, 1386, 2817, 1425 and 490 genotypes in the I, II, III, IV and the V group, respectively. Each genotype was planted in one row, 10

hills per row with two plants per hill. Distance between hills was 40 cm and 70 cm between rows.

The annual precipitation in the location of Setz is 0 mm, and available moisture is controlled by irrigation (as far as moisture is concerned, the sole unknown parameter was the level of underground waters). The irrigation was applied not later than 15 days prior to flowering within each maturity group.

The material was estimated twice - during the growing season and at the harvest. So-called secondary traits, significant for the maize grain yield increase under stress conditions, were monitored (Banzinger et al., 2000). During the growing season, the following parameters were observed and recorded: general estimation of the habitus (GEH, on the 1 to 3 scale, where 1 is damaged, 2 is medium damaged and 3 is normal), dates of anthesis and silking (on this basis the number of days from planting to anthesis (NDA) and the number of days from planting to silking (NDS) were calculated, as well as, the anthesis-silking interval (ASI). The following traits were monitored at harvest: ear setting percentage (ESP calculated as the number of ears/number of plants per row), seed setting percentage (SSP) and kernel fill percentage (KFP).

The two-step cluster analysis (Everitt, 1980) (SPSS Windows Evaluation Version 15.0) was used for the statistical analysis. Out of observed traits, GEH was considered as a categorical variable, while traits ASI, ESP, SSP, KFP, NDA and NDS were considered as continuous variables. Non-hierarchical clustering methods are very sensitive to extreme values. Due to the presence of extreme values in ESP, SSP and KFP variables, they were transformed by the *ln* (logarithm) transformation, and then all continuous variables were standardized, because of different units of measurement.

RESULTS AND DISCUSSION

Description of genotypes
Most entries of the collection did not survive until the evaluation, under conditions of such extreme drought stress. Only 785

genotypes had complete data that could be included into the statistical analysis. As a result of exceptionally high temperatures and short day length, the period from planting to flowering was extremely shortened. Under conditions of temperate climate in Serbia the period from planting to flowering is from 40 to 80 days, depending on the maturity group. The length of period from planting to flowering in Egypt was from 26 (maturity group I) to 65 days (maturity group V).

Out of 785 observed genotypes, 549 (69.9%) genotypes were estimated as damaged (1 for GEH), 38 (4.8%) were estimated as medium damaged (2 for GEH), while only 198 (25.2%) had the highest estimate of GEH (3, meaning normal).

The average values of 785 maize genotypes and their measures of variations for five observed traits are presented in Table 1.

Table 1. Average values of maize genotypes and their measures of variations

	GEH	ASI	ESP	SSP	KFP
Mean	1.55	1.56	59.12	79.51	92.61
Std. Error of Mean	0.03	0.06	0.78	0.42	0.33
Std. Deviation	0.87	1.62	21.87	11.72	9.13
Variance	0.75	2.62	478.31	137.29	83.42
Range	2	16	166.67	70.00	55.00
Minimum	1	-5	0.00	30.00	45.00
Maximum	3	11	166.67	100.00	100.00

Cluster analysis based on five traits

Sixteen genotypes with zero ears setting percentage were rejected from data processing. During the first phase of data analysing five variables were considered (GEH, ASI, ESP, SSP and KFP). The cluster analysis singled out two clusters as the best solution. First cluster group together 231 genotypes, while the rest of 538 genotypes belonged to the second cluster. The first group encompassed 111 local landraces, 62 introduced inbreds and 54 introduced populations, while the second group included 232 local landraces, 122 introduced inbreds and 147 introduced populations. The grouping of this material according to the length of growing season is given in Table 2.

Table 2. Content of 2 clusters according to origin and maturity group

Origin/homogeneity	Maturity group	Cluster		Total (initial number)
		1	2	
Landraces	I	25	19	44 (159)
	II	43	139	182 (726)
	III	42	73	115 (1134)
	IV	0	1	1 (193)
	V	1	0	1 (5)
	Total	111	232	343 (2217)
Introduced inbreds	I	0	1	1 (46)
	II	32	13	43 (412)
	III	25	81	106 (1132)
	IV	0	19	16 (843)
	V	6	41	19 (406)
	Total	63	155	218 (2830)
Introduced populations	I	4	12	14 (59)
	II	15	26	40 (248)
	III	15	64	79 (551)
	IV	18	33	49 (398)
	V	5	16	19 (79)
	Total	57	151	208 (1335)
Total		231	538	769 (6382)

Extremely early (I) local landraces, as well as early (II) introduced inbreds were significantly more present in the first cluster (11 and 6.5%, respectively), while materials belonging to the III, IV and V maturity group were represented in the first cluster with 3%, 1.3% and 2.4% of the initial number, respectively. The frequency of categorical variable GEH (Table 3) points out that the first cluster encompassed all genotypes (38 and 193) estimated 2 and 3, respectively. All genotypes (538) evaluated 1 were grouped together in the second cluster.

Table 3. Genotype percentage within clusters for the general estimation of habit (GEH)

Cluster	GEH 1		GEH 2		GEH 3	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
1	0	0	38	100	193	100
2	538	100	0	0	0	0
Combined	538	100	38	100	193	10

Each continuous trait is presented with a plot of the means for each group and simultaneous confidence intervals for the cluster means (Figure 1). Average SSP, KFP and ESP are largest for the first cluster, while ASI is smaller. However, there were two, more or less overlapping clusters considering the interval of variation of certain continuous variables. It indicates that genotypes from different clusters can have similar values for any of four continuous traits, particularly for ASI.
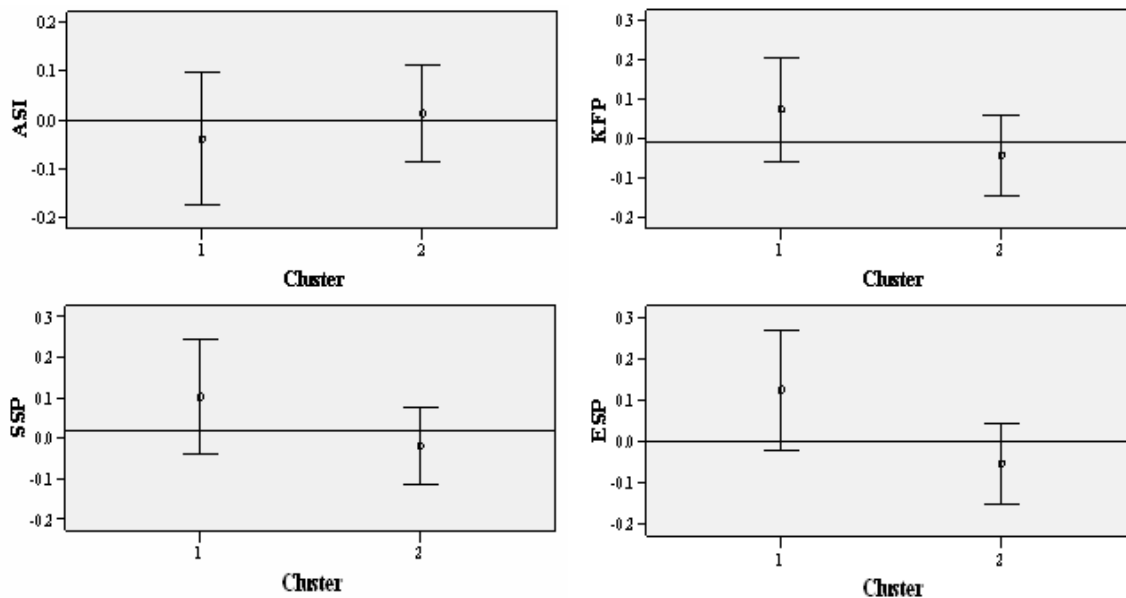

Figure 1. Plot of the means and simultaneous confidence intervals for cluster means for ASI, KFP, SSP and ESP

According to the significance, based on the t test, in the cluster 1 continuous variables were sequenced: ear setting percentage, kernel fill percentage, seed setting percentage and the anthesis-silking interval (Figure 2). Negative t value for a particular variable indicated that genotypes had a lower value of that trait than the overall mean value and vice versa. In our research, genotypes distributed in the first cluster had values of ear setting percentage, kernel fill percentage and seed setting percentage above the collection average, while the values of parameters of the anthesis-silking interval were below the average. In the cluster 2, significance of traits, based on the t test, is ranged: ear setting percentage, seed setting percentage, kernel fill percentage and the anthesis-silking interval. The values of ear

setting percentage, seed setting percentage and kernel fill percentage in this cluster were below the collection average, while the values of parameters of the anthesis-silking interval were above this average.
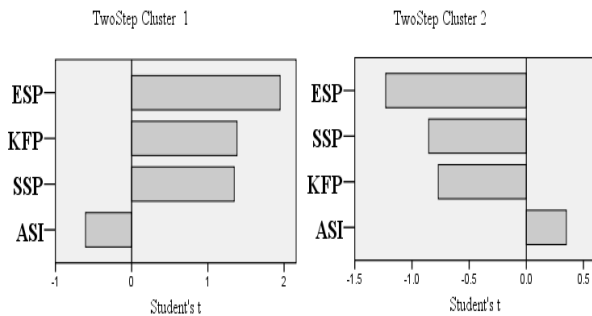

Figure 2. Importance of continuous variables within a cluster

This can lead to the conclusion that 231 genotypes of the total 785 analysed, belonging to the first cluster, can be used for further studies for additional characterisations and evaluation and for the formation of a core collection for tolerance to drought stress (Babic et al., 2009). However, these genotypes represent only 3.62% of the total initial material. According to Brown (1989), a core collection should encompass around 10% of accessions, which would represent a total of 70% of the whole collection's diversity. The MRI gene bank was tested for drought tolerance under extreme conditions and 3.62% is lower than the recommended percentage. Although certain parameters point to genotypes of the first clusters, which should be used in further studies, some results point out that this conclusion should be taken with caution. The Student's *t* test was not statistically significant at the 95% significance

level for any of the clusters. Moreover, it is not good that the intervals of variations of continuous variables overlap for different clusters, especially for values of ASI, whose low values indicate increased level of drought tolerance. Because of that it is decided, in the second phase of the data analysis, in addition to the already used variables, to take into consideration the numbers of days from planting to tasseling and silking. Although it was considered that data on the number of days from planting to anthesis and silking, obtained under conditions of Egypt, should be taken with caution, the aim was to observe whether a better distribution of observed units between clusters would be obtained in a way that values of continuous variables would not greatly overlap.

Cluster analysis based on seven traits

In this phase of analysis, which took into account 6 continuous traits and one categorical trait, investigated genotypes were classified into three clusters.

This analysis, also, excluded 16 genotypes having values of ear setting percentage zero. A total of 211, 363 and 195 genotypes belonged to the first, second and the third cluster, respectively. A total of 211 genotypes, with an estimated GEH 1 during the growing season, belonged to the first cluster, while 327 genotypes evaluated GEH 1 and 36 genotypes evaluated GEH 2 belonged to the second cluster. The third cluster encompassed 193 genotypes with values of GEH of 3, and two genotypes evaluated GEH 2 (Table 4).

Table 4. Genotype percentage within clusters for the general estimation of habit (GEH)

Cluster	GEH 1		GEH 2		GEH 3	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
1	211	39,2	0	0	0	0
2	327	60,8	36	94,7	0	0
3	0	0	2	5,3	193	100
Combined	538	100	38	100	193	100

Distribution of investigated genotypes, according to clusters, origin and maturity groups, is given in Table 5. More frequent

presence of early genotypes in clusters 2 and 3 is noticeable, while latter genotypes mostly belonged to cluster 1.

Table 5. Content of 3 clusters over origin and homogeneity

Origin/homogeneity	Maturity group	Cluster			Total (initial number)
		1	2	3	
Landraces	I	0	40	4	44 (159)
	II	13	132	37	182 (726)
	III	8	69	38	115 (1134)
	IV	1	0	0	1 (193)
	V	0	0	1	1 (5)
	Total	22	241	80	343 (2217)
Introduced inbreds	I	0	1	0	1 (46)
	II	3	10	32	45 (412)
	III	60	21	25	106 (1132)
	IV	19	0	0	19 (843)
	V	38	4	5	47 (406)
	Total	120	36	62	218 (2830)
Introduced populations	I	0	15	1	16 (59)
	II	5	22	14	41 (248)
	III	21	43	15	79 (551)
	IV	27	6	18	51 (398)
	V	16	0	5	21 (79)
	Total	69	86	53	208 (1335)
Total		211	363	195	769 (6382)

In this case, the variations of continuous variables between clusters were not overlapped. Exception was ear setting percentage, where the values considerably overlapped between the second and the third cluster (Figure 3).
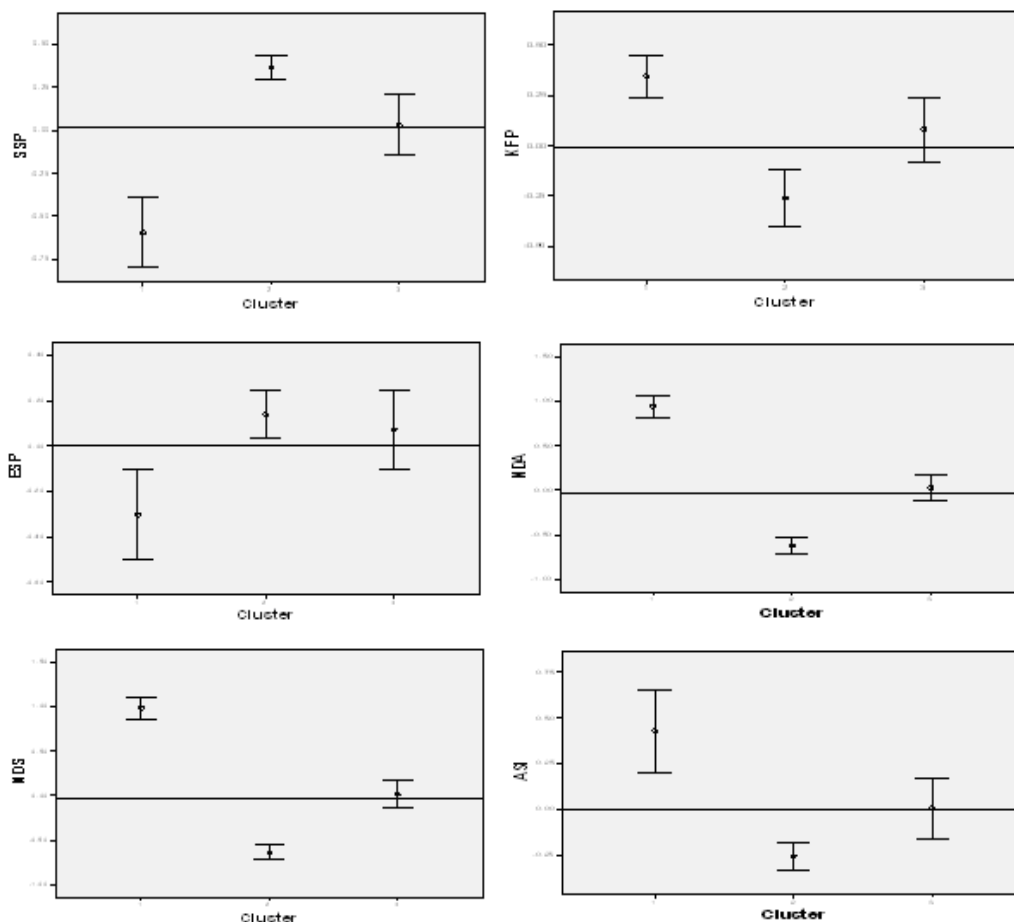

Figure 3. Plot of the means and simultaneous confidence intervals for cluster means for SSP, KFP, ESP, NDA, NDS and ASI

Variables were sequenced according to their significance; hence it is noticeable that all variables were statistically significant in the first cluster at the 95% significance level. Thereby, genotypes in this cluster had values of the number of days from planting to silking, number of days from planting to tasseling, kernel fill percentage and the anthesis-silking interval above the average, while ear setting percentage and seed setting percentage remained below the average (Figure 4).

As far as the second cluster is concerned, all variables were significant at the 95% significance level and were sequenced according to their significance as follows:

number of days from planting to silking, number of days from planting to tasseling, seed setting percentage, anthesis-silking interval, kernel fill percentage and ear setting percentage.

Whereby the values of the number of days from planting to silking, number of days from planting to tasseling, anthesis-silking interval and kernel fill percentage were below the average, while the values for seed setting percentage and ear setting percentage were above the average (Figure 4). Not a single continuous variable was statistically significant at the 95% significance level in the third cluster and all the values were above the average (Figure 4).
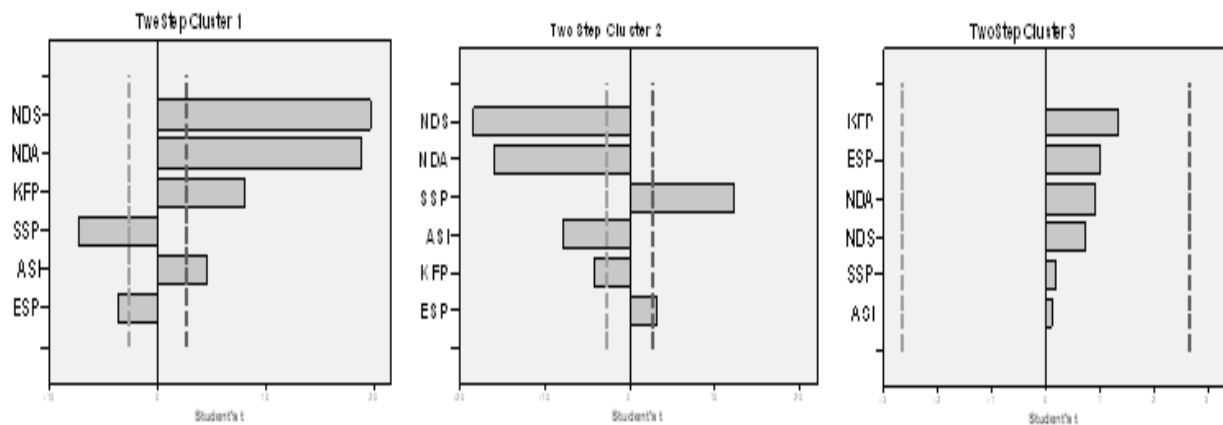

Figure 4. Importance of continuous variables within a cluster

According to these results, genotypes of the third and the second cluster, containing 558 accessions, were selected for the further operations. These are, therefore, early maturity genotypes with seed setting percentage above the average, values of the anthesis-silking interval below the average and ear setting percentage above the average belonging to the second cluster. Also, early to medium maturity genotypes of the third cluster that were evaluated as the best during the growing season, had kernel fill percentage and ear setting percentage above the average (Figure 5).

These 558 genotypes will be further analysed in order to classify them into cores for drought tolerance. The efficient presentation of genetic diversity of the whole collection by the core collection depends

primarily on the distribution of accessions into homogenous groups (van Hintum et al., 2000). Groups should be formed in such a way that the variation among groups is maximal, while the variation within groups is minimal. Stepwise and hierarchical procedures are most often applied for such distribution: first, a great division is made, than these groups are divided into smaller subgroups (van Hintum, 1994; Crossa et al., 1995). The selected material encompasses 321 local populations, 98 introduced inbred lines and 139 introduced populations. This grouping will be a starting point in defining homogenous groups. Introduced inbred lines will be directly included into the testing process with at least three divergent testers in order to observe their combining abilities and to gather information that can be used when selection for possible

ROMANIAN AGRICULTURAL RESEARCH

deriving of synthetic populations will be performed. A special attention will be paid to local populations. As the search is actually for material adapted to agro-climate conditions of the moderate climate, local populations that have been developed for a long period of time in the process of natural crossing and selection, can be of an exquisite importance. Observing the origin of the selected local populations, it can be noticed that the majority of populations originate from Montenegro (29%) and Bosnia and Herzegovina (21%). A larger number of accessions originate from Serbia (25%), but mostly from the southern parts of the country (province Kosovo and southern Serbia). A somewhat smaller number of accessions originate from a coastal part of Croatia (8%), then from Macedonia (9%) and Slovenia (6%). Hence, early and medium maturity populations collected in regions of former Yugoslavia, which are mainly hilly-mountainous regions with certain impacts of the Mediterranean climate and where extensive cultivation has been maintained for a long time, stood out. This material shall be further classified on the basis of morphological properties and information on the origin. The multivariate analysis applied to studies on the germplasm collection allows a better understanding of the structure of the collection, detection of the relationships among accessions, as well as the identification of possible groups (Martines-Calvo et al., 2008). In a later stage, genetic characterisation will be included by the application of genetic markers, with the aim to clearly define homogenous groups of populations according to their relatedness. The more detailed characterisation can enhance selection efficiency of genotypes in the process of the core collection formation.

CONCLUSIONS

The aim of this part of the experiment was to select approximately 10% of genotypes of observed germplasm (6382 accessions stored at the Maize Research Institute, Zemun Polje, gene bank), as potential sources of drought tolerance. As the selection of only the first cluster (the first phase of the analysis,

which did not include the number of days from planting to tasseling and silking) would represent only 3.62% of the initial material, the conclusion was that this phase of the analysis did not provide a desirable selection intensity. The selection of the second and the third cluster (in the analysis taking into consideration the duration of the growing season) represented 8.74% (558 genotypes) of the initial number. Although the number of days from planting to flowering under Egyptian conditions was taken with caution, statistical parameters point to a better distribution of observed genotypes per clusters when the number of days from planting to tasseling and silking is considered. In such a way, a part of genotypes with the estimation GEH 1 (damaged) in the growing season was in fact selected, but that part actually comprised of early genotypes with high seed setting percentage, lower values of the anthesis-silking interval and ear setting percentage above the average. Kernel fill percentage was somewhat lower than the average, but this is understandable, considering extreme conditions to which genotypes were exposed.

According to presented results, it can be concluded that the two-step cluster analysis, which considered seven variables, gave results that better responded to the goal of the study. Therefore, it was decided that the genotypes of the second and the third cluster should be selected for further work on the formation of core collections of maize tolerant to water stress.

REFERENCES

Babic, V., Vancetovic, J., Vidakovic, M., Kravic, N., Prodanovic, S., Andjelkovic, V., 2009. *Identification of drought tolerant maize accession by two step cluster analysis.* Book of apstracts (pp 23). 19th EUCARPIA conference, Genetic

resources section. Ljubljana, Slovenia, 26-29 May, 2009.

Banzinger, M., Edmeades, G.O., Beck, D., Bellon, M., 2000. *Stress management*. In: Breeding for drought and nitrogen stress tolerance in maize: From Theory to Practice. Mexico, D.F., CIMMYT: 23-31.

Brown, A.H.D., 1989. *Core collections: a practical approach to genetic resources management*. Genome, 31: 818-824.

Crossa, J., DeLacy, I.H., Taba S., 1995. *The use of multivariate methods in developing a core collection*. In: "Core Collections of Plant Genetic Resources" (Hodgkin, T., Brown, A.H.D., van Hintum, Th.J.L., Morales, E.A.V., eds), John Wiley and Son, UK: 77-92.

Dodig, D., Aleksić, V., Spasov, P., Petrović, R., Miletić, R., 2002. *Climate changes in eastern Serbia and their influence on plant production and ecosystem*. In: Proceedings of ICID International Conference on Drought Mitigation and Prevention of Land Desertification. Bled, 21-25 April, 2002, Slovenia, DOC8, pdf, CD Edition.

Everitt, B.S., 1980. *Cluster analysis*. Quality and quantity, 14: 75-100.

Martines-Calvo, J., Gisbert, A.D., Alaman, M.C., Herhandorena, R., Romero, C., Lacer, G., Badenes, M.L., 2008. *Study of a germplasm collection of loquat (Eriobotrya japonica Lindl.) by multivariate analysis*. Genetic Resources and Crop Evolution, 55 (5): 695-703.

Monneveux, P., Belhassen, E., 1996. *The diversity of drought adaptation in the wide*. Plant Growth Regulation, 20: 85-92.

Ribaut, J.M., Jiang, C., Gonzalez-de Leon, D., Edmeades, G.O., Hoisington, D.A., 1997. *Identification of quantitative trait loci under drought conditions in tropical maize. 2. Yield components and marker-assisted selection strategies*. TAG, 94: 887-896.

Ribaut, J.-M. and Rigot, M., 2007. *Marker-assisted selection to improve drought adaptation in maize: the backcross approach, perspectives, limitations and alternatives*. J. Exp. Botan., 58: 119-128.

Tucak, M., Popović, S., Čupić, T., Šimić, G., Gantner, R., Meglič, V., 2009. *Evaluation of alfalfa germplasm collection by multivariate analysis based on phenotypic traits*. Romanian Agricultural Research, 26: 47-52.

Van Hintum, Th.J.L., 1994. *Hierarchical approaches to the analysis of genetic diversity in crop plants*. In "Core Collections of Plant Genetic Resources" (Hodgkin, T., Brown, A.H.D., van Hintum, Th.J.L., Morales, E.A.V., eds), John Wiley and Son, UK: 23-34

Van Hintum, Th.J.L., Brown, A.H.D., Spillane, C., Hodgkin, T., 2000. *Core collections of plant genetic resources*. IPGRI Technical Bulletin, 3: 11-13.