DOI: 10.22620/agrisci.2021.29.007

USE OF CLUSTER ANALYSIS AND ANALYSIS OF THE MAIN COMPONENTS FOR EVALUATION OF TRITICALE SAMPLES

Evgeniy Dimitrov[1], Krasimira Uzunova[2*], Zlatina Uhr[1]

[1]Institute of Plant Genetic Resources "K. Malkov "- Sadovo

[2]Agricultural University- Plovdiv

***E-mail: uzunova@au-plovdiv.bg**

Abstract

The study was conducted in the period 2015-2017 in the experimental field of IRGR "K. Malkov" Sadovo. The elements of productivity of 24 varieties and lines of triticale (*X. Triticosecale Wittmacks*) were studied. The degree of variation of each of the indicators of productivity is determined by calculating a coefficient of variation. The results obtained show that the most variable indicator is the weight of grains in the central spike whereas the slightest variation is observed in the parameters of test weight and 1000 grains weight. Hierarchical cluster analysis and analysis of the main components (PC-analysis) are applied to assess the genetic similarity and distance between the different varieties. As a result of the clustering the studied samples were divided into six groups. The varieties Trit 32/6 and KS 60 are characterized with the greatest genetic proximity, followed by line BGR 30071 with the Bulgarian variety Rozhen. The largest genetic difference was reported between line BGR 30816 and variety KT 81. The applied analysis of the main components shows that the components PC 1, PC 2 and PC 3 explain 70.6% of the total variation of all traits by genotypes. The largest numbers of samples (9) belong to component 3, as four of them are located in the negative values of PC 3, and the other five genotypes are in the positive values of PC 3. Component 1 is represented by eight samples, and component 2 includes seven triticale genotypes.

Keywords: triticale, elements of productivity, genetic distance, cluster analysis, PC analysis.

INTRODUCTION

Triticale is the newest artificially created cereal-forage crop, obtained by interspecific hybridization between wheat (*Triticum*) and rye (*Secale*) (Yankov et al., 2002). This new crop combines grain quality, productivity and resistance to wheat diseases and the viability and durability of rye. The increased attention to the crop is due to the following advantages: high productive potential, which in many cases reaches and exceeds that of wheat, barley and rye, very good resistance to diseases, high adaptability to grow in soils with high acidity or high content of aluminum in percentage terms (Kirchev, 2019). The areas for growing triticale are being expanded every year due to their undemanding soil and climatic conditions. Globally, the standard used in this culture is the Lasko variety while the group of varieties created in Bulgaria includes: Attila, Zaryad, Vihren, Rakita, AD 7291, Rozhen, Sadovets and others. (Baichev, 2009; Stankov et al., 2014). Worldwide, many triticale breeding programs focus on improving economically important parametars, such as grain yield and biomass yield, grain quality, disease and pest resistance. Various studies have shown that the genetic diversity in triticale germplasm is relatively low and needs to be expanded (Mergoum et al., 2019). Proper breeding of parental forms to

increase yield potential in recombinant genotypes can be made by determining genetic distance (Islam, 2004). Most often, genetic distance is measured as phenotypic distance (Arriel et al., 2007; Kabir et al., 2009). It is believed that if genotypes are phenotypically different in many respects, they are also genetically distant in their genomes. Various researchers have successfully applied the methods of cluster analysis and PC analysis to determine the genetic distance in the selection process (Bhatt, 1970; Mohammadi & Prassana, 2003; Eivazi et al., 2007).

The aim of this paper is to study the elements of productivity and to establish the genetic distance of Triticale samples, with a view to their use in the breeding process as initial source material for the creation of new and highly productive lines and varieties.

MATERIALS AND METHODS

During the period 2017-2019, in an experimental field of the Institute of Plant Genetic Resources - Sadovo, the elements of productivity of 24 varieties and lines of triticale (X. *Triticosecale* Wittmacks) were studied, and the Rakita variety was used as a standard. Biometric measurements were made on 10 randomized plants of each genotype in 3 replicates. The following productivity parameters are reported: **PH** – plant height (cm), **SL** – central spike length (cm), **NSS** – number of spikelets per central spike, **NGS** – number of grains per central spike, **WGS** – grain weight per central spike (g), **HI** - harvest index, **PT m^2** -productive tillering per m^2 (number), **1000 GW** - 1000 grain weight, **TW**-test weight (kg/hl).

The degree of variation of each of the signs of productivity is determined by calculating a coefficient of variation. The degree of variation of each of the signs of productivity is determined by calculating a coefficient of variation. It is accepted that the variation is considered insignificant when the coefficient of variation is up to 10%, medium variation - values in the range from 10% to 20%, strong variation- when it is over 20% (Shanin, 1977). Hierarchical cluster analysis and PC analysis based on mean values over the study period were used to determine the genetic distance between the individual genotypes. The mathematical processing of the results was performed using the statistical processing programs SPSS 13 and Microsoft EXCEL 10 for Windows.

RESULTS AND DISCUSSION

The results of the biometric measurements of the productivity elements in the studied triticale samples are presented in Table 1. The table shows the values of the following statistical indicators: arithmetic mean, minimum, maximum, standard deviation, coefficient of variation and error of the mean. In the case of plant height, the results show that in eighteen genotypes the measured height is over 100 cm. The highest value of the indicators length of the central spike, number of spikes in the central spike and number of grains in the central spike was shown by line BGR 30814. The weight of the grains in the central spike in the studied materials varies from 1.2 to 3.42 g, considering the fact that only the line BGR 30816 and the variety Belitsa are characterized by a value of the indicator over 3.0 g. The lowest value of the harvest index is observed in sample KC 20, and the largest number of productive tills has been achieved by variety KT 81. For the trait weight per 1000 grains, the data from the table show that fourteen selection materials exceed the Rakita standard in terms of the trait value, with the highest result being reported for variety KT 81 - 48.1 g. The lowest value of the indicator is characterized by number BGR 30812 - 33.5. The largest percentage (54.1%) of the total number of tested samples weight per 1000

grains is in the range from 35.0 to 40.0 g. The test weight of the studied breeding materials is in the range from 56.4 kg / hl (4047 TH 1) to 76.2 (Sofia 3) kg / hl. There are three triticale genotypes above the standard level. The reported test weight is over 65.0 kg / hl in twenty-two samples. From the results obtained, shown in the table below, we can assume that there is a proven genetic diversity in the studied genotypes on the studied traits of productivity. For the parameters test weight, harvest index and 1000 grains weight, low values of the coefficient of variation were reported, which define these traits as weakly variable. High variation was observed in the weight of the grains in the central spike. The variation of the traits plant height, length of the central spike, productive yield per m^2, number of spikelets and number of grains in the central spike, are evaluated as medium variation.

Table 1. Results of biometric anlysis of the structural elements of productivity of triticale samples for the period 2017-2019

№	Variety, line	PH	SL	NSS	NGS	WGS	HI	PT m^2	1000 GW	TW
1	KC 20	94.	9.1	21.8	50.1	1.84	0.4	454.0	34.8	62.9
2	Trit.32/6	104	9.1	23.7	58.6	2.24	0.4	397.8	39.1	65.2
3	KT 81	113	11.9	26.3	56.9	2.75	0.4	519.1	48.1	67.5
4	130 TM 3-1	105	12.4	31.3	51.0	2.29	0.4	371.4	40.1	66.7
5	4047 TH 1	109	10.9	31.6	54.4	2.76	0.4	397.2	40.6	56.4
6	BGR 30071	101	9.2	27.4	45.4	2.06	0.4	408.7	43.8	67.6
7	BGR 30078	133	13.1	32.5	55.4	2.21	0.4	426.2	37.5	65.7
8	BGR 30812	110	11.6	30.7	53.5	2.22	0.5	397.4	33.5	66.4
9	BGR 30813	89.	8.8	23.4	38.0	1.57	0.4	461.2	39.5	64.6
1	BGR 30814	112	13.4	36.0	72.4	2.78	0.5	416.6	37.1	68.6
1	BGR 30815	114	12.7	35.9	64.4	2.76	0.4	507.9	36.3	65.7
1	BGR 30816	104	11.0	25.7	61.0	3.42	0.4	247.1	44.6	67.0
1	Oak Treiwel	113	10.4	22.9	40.6	1.84	0.4	420.3	42.2	65.6
1	KS 60	105	9.9	25.2	58.9	2.31	0.4	392.2	40.2	66.7
1	Coorong	91.	9.7	22.3	58.0	2.20	0.5	375.7	35.6	61.9
1	Gama 05209	97.	12.3	25.2	60.4	2.61	0.4	462.9	38.2	70.2
1	Vronti	90.	8.4	20.6	45.4	1.77	0.5	431.3	39.3	65.0
1	Sofia 3	101	7.8	21.6	31.3	1.20	0.4	378.4	36.7	76.2
1	Vihren	103	9.7	23.5	49.6	2.31	0.5	361.8	40.0	66.4
2	Persenk	94.	10.4	23.9	43.9	1.95	0.4	314.6	38.4	65.0
2	Zaryad	104	10.9	28.0	61.0	2.68	0.4	326.0	38.7	63.2
2	Belitsa	133	12.1	31.3	57.0	3.09	0.4	329.9	44.8	65.1
2	Rozhen	104	11.3	24.6	53.5	2.52	0.4	406.8	44.9	63.3
2	***Rakita-st.***	110	12.1	32.4	56.2	2.45	0.5	417.0	38.0	68.5
Mean		106	10.8	27.0	53.2	2.33	0.4	400.9	39.7	65.9
Minimum		89.	7.8	20.6	31.3	1.20	0.4	247.1	33.5	56.4
Maximum		133	13.4	36.0	72.4	3.42	0.5	519.1	48.1	76.2
Std. deviation		11.	1.6	4.6	9.1	0.5	0.0	60.3	3.6	3.5
Coef. var., %		10.	14.7	17.1	17.1	21.4	7.0	15.0	9.0	5.3
Standard error		2.3	0.3	0.9	1.9	0.1	0.0	12.3	0.7	0.7

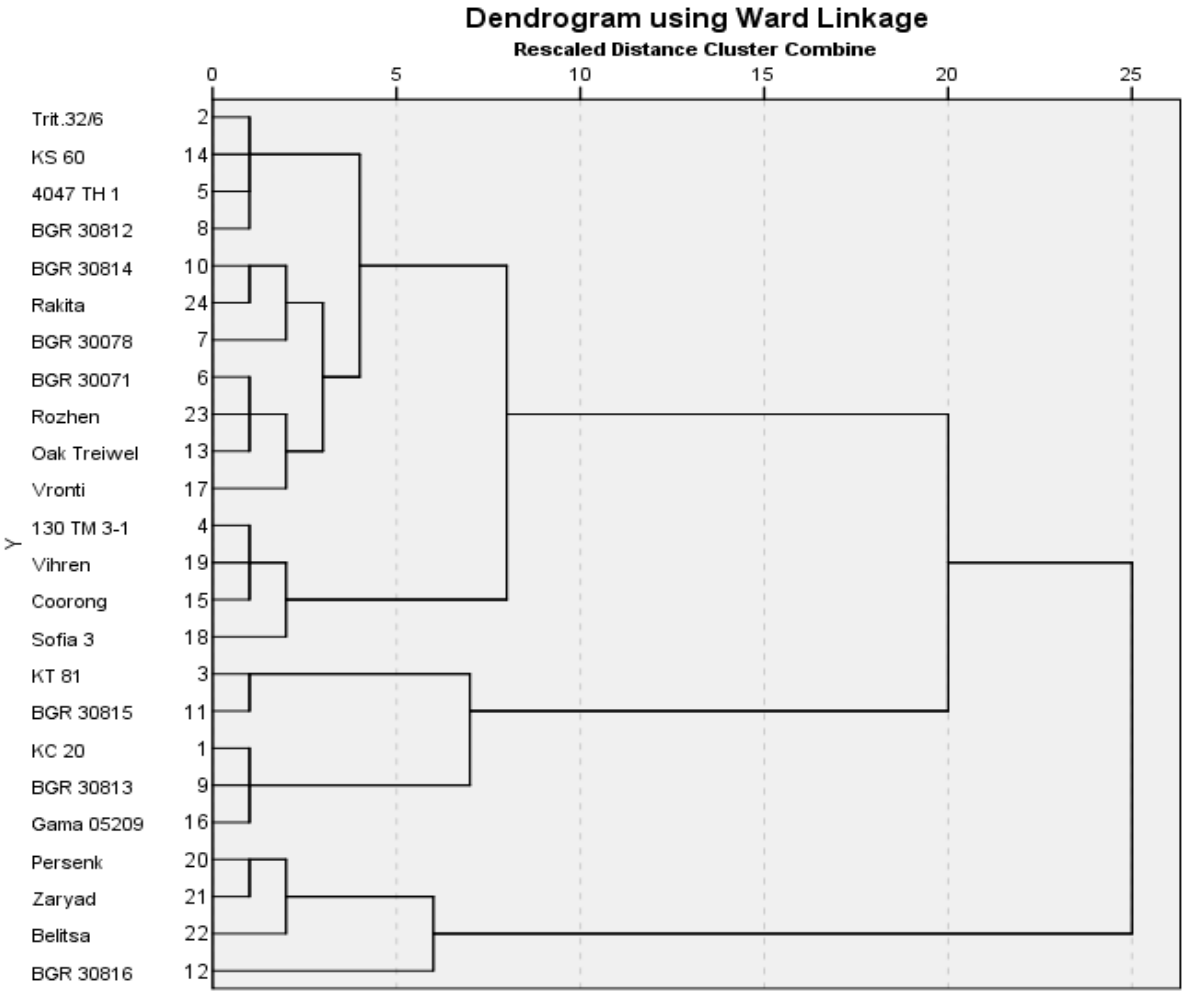

Fig. 1. Dendrogram of cluster analysis

The cluster analysis is a method of classification and hierarchy in which the studied population is divided into a number of groups called clusters. The study of breeding materials through the cluster analysis allows breeders to plan and make more effective decisions for the development of their breeding programs. The samples can be divided by genotype, depending on the phenotypic expression of a particular trait or group of traits, using clustering. The results of the clustering are presented in a dendrogram (Fig. 1). The dotted horizontal line of the dendrogram shows the rescaled distance at which the clusters were formed. From the data presented in Figure 1, it is clear that the studied breeding materials were grouped into six main cluster groups. **The first group** is represented by varieties Trit.32 / 6, KS 60, 4047 TH 1 and line BGR 30812. **The second group** includes the Rakita standard, BGR 30814, as well as BGR 30078 which joins the group at higher rescaled distance. **The third group** includes the samples BGR 30071, Rozhen and Oak Treiwel forming an independent sub-group which the Vronti variety joins. Representatives of **the fourth group** are the varieties 130 TM 3-1, Vihren, Coorong, as well as the variety Sofia 3 which showed the lowest values of the parameters length of the central spike, number of grains and grains weight in the central spike. The fifth cluster group is the most numerous

and consists of two subgroups. The first subgroup includes KT 81 and BGR 30815, and the second subgroup is represented by samples KC 20, BGR 30813 and Gama 05209. The sixth group includes four triticale genotypes, with the Persenk and Zaryad varieties being separated into a separate cluster, joined at greater Euclidean distances by the Belitsa and BGR 30816 varieties, which achieved the highest values of the grain weight index in the central spike.

Table 2 presents the genetic similarity between the different triticale samples based on the coefficient according to which the individual cluster pairs are formed between the studied genotypes.

Table 2. Genetic similarity between the studied triticale genotypes

№	Genotype	Genotype	Coeff.	
1	Trit.32/6	KS 60	3.1	genetically similar
2	BGR 30071	Rozhen	8.2	
3	4047 TH 1	BGR 30812	14.5	
4	130 TM 3-1	Vihren	20.9	
5	KC 20	BGR 30813	28.8	
6	BGR 30816	KC 20	207.8	genetically distant
7	BGR 30816	BGR 30813	215.9	
8	BGR 30816	Gama 05209	216.0	
9	BGR 30816	BGR 30815	261.3	
10	BGR 30816	KT 81	272.2	

The results of the table show that the greatest genetic similarity was found in cvs Trit. 32/6 and KS 60, followed by line BGR 30071 with the Bulgarian variety Rozhen. The line BGR 30816 and variety KT 8 (fifth group) are characterized by the greatest genetic difference. A strong genetic difference was also observed between BGR 30816 with the other samples forming a fifth cluster group (KC 20, BGR 30813, Gama 05209 and BGR 30815).

According to several authors (Fang et al., 1996; Khodadadi et al., 2011; Siahbidi et al., 2013), it can be generally accepted that the cluster analysis gives the best estimate for the genetic distance of genotypes and therefore, the cluster analysis is preferably used in genetic diversity research.

The analysis of the main components appears as a supplement to the cluster analysis. The results of the PC analysis (Table 3) show that the three main components PC 1, PC 2 and PC 3 explain 70.6% of the total variation of all traits by genotype, which is large enough.

Table 3. Component analysis of the variance in the studied traits

Component	Total	% of Variance	Cumulative %
1	**3.70**	**41.1**	**41.1**
2	**1.44**	**16.1**	**57.1**
3	**1.21**	**13.5**	**70.6**
4	0.91	10.1	80.7
5	0.86	9.5	90.2
6	0.47	5.2	95.5
7	0.22	2.4	97.9
8	0.14	1.5	99.4
9	0.05	0.6	100.0

The data in Table 4 show that five traits are in the first component and relate positively to it: plant height (0.694), central spike length (0.895), number of spikelets per central spike (0.844), number of grains per central spike (0.839), grain weight per central spike (0.877). The second component contains two features and also relates positively to it productive tillering per m^2 (0.593) and test weight (0.476). The third component includes the parameters 1000 grain weight (0.710), which refers positively and harvest index (-0.468), which refers negatively to PC3. According to Biabani & Pakniyat (2008) and Stamatov & Deshev (2012), the traits found in the individual components are determined by nearby genes in the genome.

Table 4. Explained significant components by indicators in triticale samples

№	Indicators	Component		
		1	**2**	**3**
1	PH	**0.694**	0.341	0.462
2	SL	**0.895**	0.279	0.017
3	NSS	**0.844**	0.354	-0.083
4	NGS	**0.839**	-0.118	-0.306
5	WGS	**0.877**	-0.377	0.101
6	HI	0.400	-0.443	**-0.468**
7	PT m^2	-0.020	**0.593**	-0.280
8	1000 GW	0.190	-0.442	**0.710**
9	TW	-0.173	**0.476**	0.298

The studied triticale samples are related differently to the three main components (Table 5). The first main component includes eight samples, five of which are positive with PC 1 (Trit.32 / 6, BGR 30813, Vronti, Vihren, Persenk) and the other four negative (BGR 30078, BGR 30814, Rakita-st.). The positive values of PC 2 include the varieties 4047 TH 1, KS 60, Coorong and Zaryad, and the negative values of PC 2 include the varieties 130 TM 3-1, Oak Treiwel, Sofia 3. The largest number of samples of studied materials are related to PC 3, as four genotypes (KC 20, BGR 30812, BGR 30815, Gama 05209) refer to the negative part of the component, and the remaining five numbers (BGR 30816, KT 81, BGR 30071, Belitsa and Rozhen) fall into the positive values of PC 3.

Table 5. Explained significant components by triticale samples

№	**Variety, line**	Component		
		1	**2**	**3**
1	KC 20	-1.049	-0.521	**-1.149**
2	Trit.32/6	**-0.496**	0.061	0.074
3	KT 81	0.883	-0.758	**0.886**
4	130 TM 3-1	0.560	**-0.771**	0.103
5	4047 TH 1	0.247	**1.134**	0.368
6	BGR 30071	-0.521	-0.259	**0.585**
7	BGR 30078	**1.655**	-1.559	-0.527
8	BGR 30812	0.393	0.490	**-1.356**
9	BGR 30813	**-1.417**	-0.321	-0.638
10	BGR 30814	**1.703**	0.651	-1.080
11	BGR 30815	1.593	0.249	**-1.652**
12	BGR 30816	0.154	0.903	**2.494**
13	Oak Treiwel	-0.369	**-1.169**	0.533
14	KS 60	-0.239	**0.444**	0.142
15	Coorong	-1.139	**2.027**	-0.700
16	Gama 05209	0.329	0.255	**-0.864**
17	Vronti	**-1.573**	0.528	-0.370
18	Sofia 3	-1.282	**-2.989**	-0.343
19	Vihren	**-0.660**	0.499	0.467
20	Persenk	**-1.013**	0.230	0.318
21	Zaryad	0.068	**0.785**	0.437
22	Belitsa	1.429	-0.317	**1.920**
23	Rozhen	-0.051	0.006	**1.102**
24	Rakita-st.	**0.797**	0.400	-0.750

The graphical manifestation of the analysis of the main components by traits and genotypes is presented in Figure 2 and Figure 3. Figure 2 shows the points and vectors of the studied performance indicators. According to the angles between the vectors of the signs, the correlations between them can be judged. The correlation is stronger and more positive at a sharper angle. At right angles, the correlation is zero, and the obtuse angle determines a negative correlation (Dragov & Dechev, 2016). From the presented data from Figure 3, we can conclude that there is a strong and positive correlation between grain weight per central spike with the signs number of grains per central spike, central spike length and number of spikelets per central spike and plant height. There is a strong, positive correlation between the number of spikelets per central spike with the parameters central spike length, number of grains per central spike and plant height. The correlation between grain weight per central spike with productive tillers per m^2 and test weight is negative. The correlation between 1000 grain weight with number of spikelets per

central spike, number of grains per central spike and productive tillering is weak and negative. Similar correlations between the various elements of productivity have been reported by other authors. Stoyanov (2013) found a high correlation between the number of grains and the mass of grains in the spike. Although the weight of the grains in the spike is a direct component of the yield and there is a high correlation with it, (Rachovska and Uhr, 2010), Stoyanov (2013), gives preference to the trait weight of the grains in the spike, as a more valuable breeding trait, which will be sufficiently reliable in carrying out the breeding process. Zhang et al. (2012) report that a larger number of grains in a spike does not always determine a higher yield. Since the weight of grains in a spike is directly related to the yield as its component, it is important to establish the presence or absence of correlation between the other characteristics of the spike with this parameter. This would make it possible to evaluate newly obtained lines at an early stage of their breeding process.

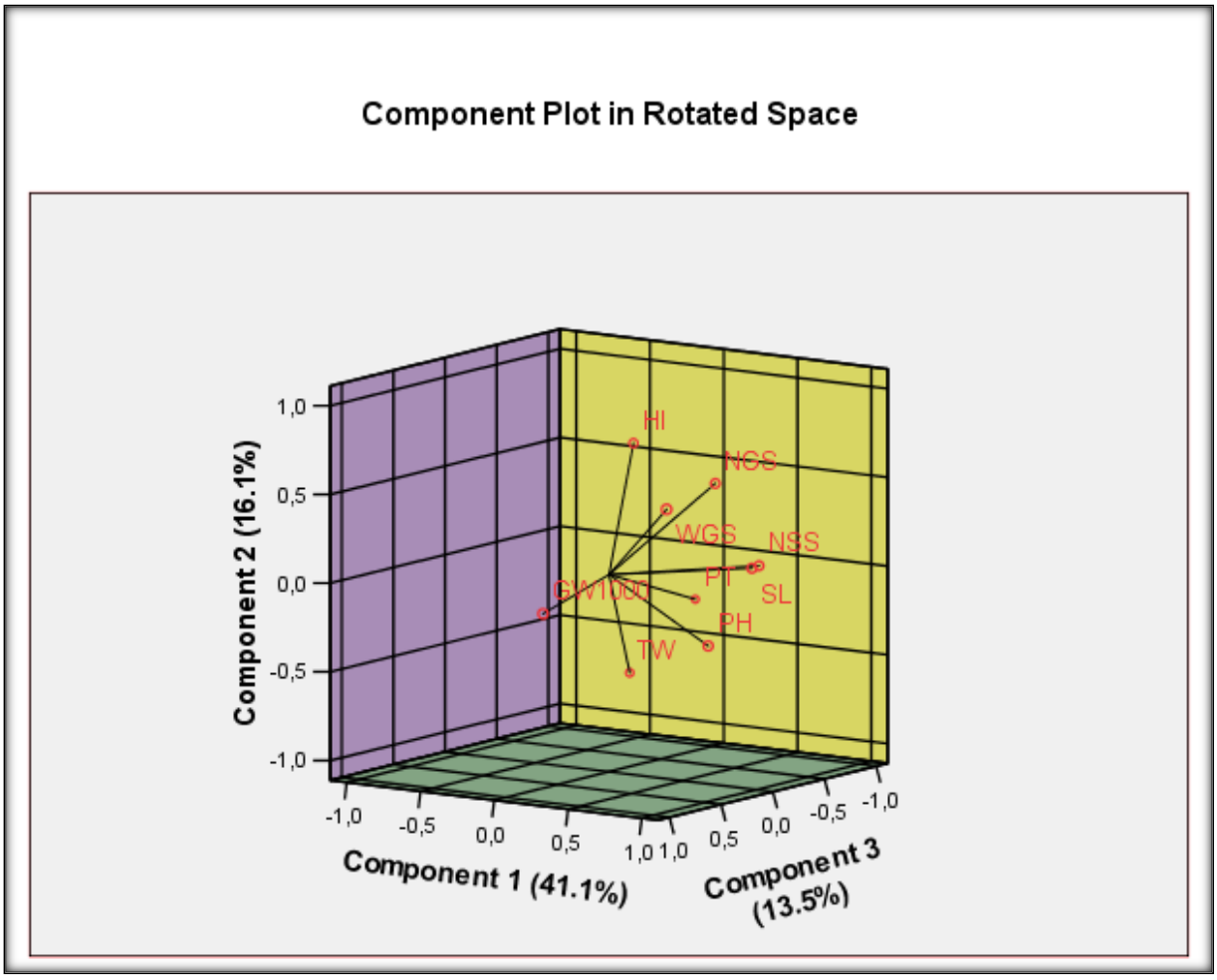

Fig. 2. Projection of the studied features by main components

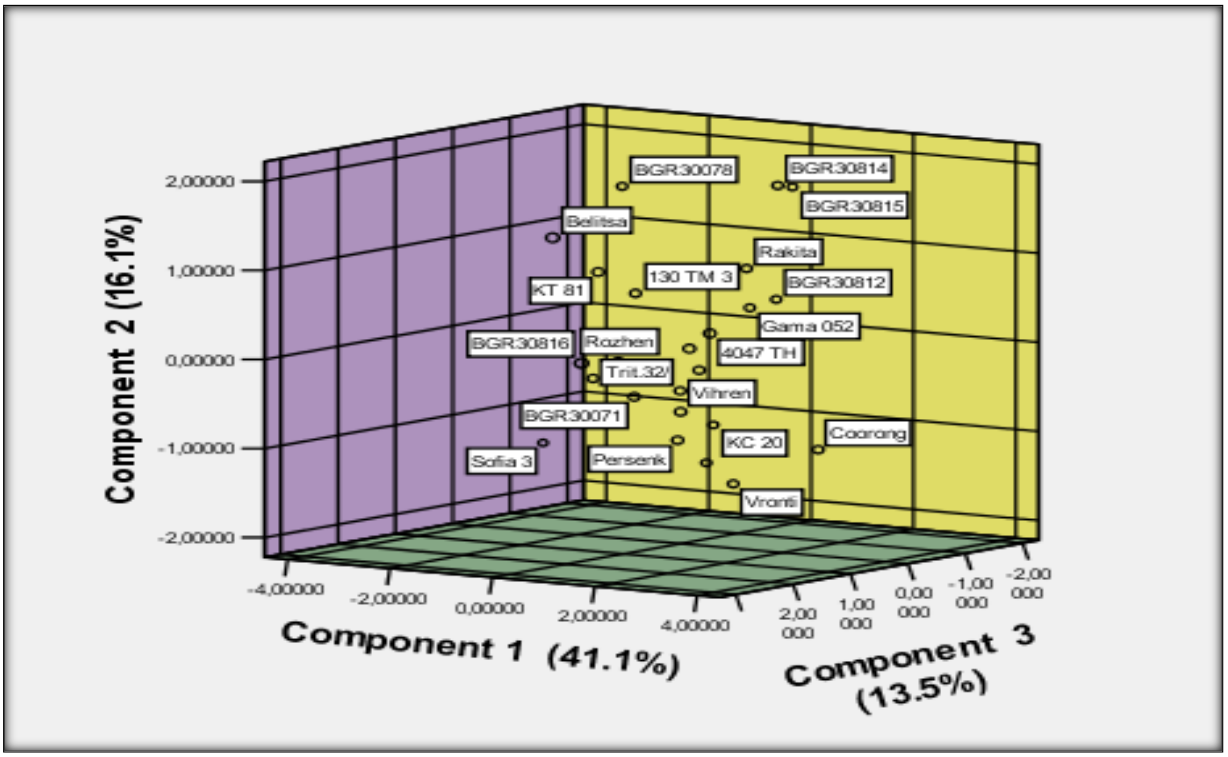

Fig 3. Projection of the studied samples by main components

From the graphical representation of the analysis of the main components by genotypes (Figure 3) we can see the location of the studied samples in the coordinate system. The genotypes Sofia 3, Coorong and BGR 30816, located in the farthest parts of the factorial plane, can be mentioned as sources of variation in order to create a variety of starting material and enrich the gene pool in triticale. These genotypes can be used as parent pairs in selective triticale improvement work.

CONCLUSIONS

The largest variation is determined for the grain weight in the central spike, and the least variable are the test weight and the 1000 grains weight.

A strong and positive correlation is observed between the parameters grain weight per central spike with the number of grains per central spike, followed by the strong correlation between the number of spikelets per central spike with central spike length.

The application of statistical methods, such as the cluster analysis and the PC analysis, is a reliable means of grouping genotypes by their genetic distance.

The studied triticale genotypes are divided into six main cluster groups, with different degrees of genetic distance. The greatest genetic proximity was found between the Trit. 32/6 and KS 60 varieties, and the largest genetic distance - between line BGR 30816 and cv KT 81.

The analysis of the main components shows that the components PC 1, PC 2 and PC 3 explain 70.6% of the total variation of all traits of genotypes.

The following samples can be emitted as sources of variation: Sofia 3, Coorong and BGR 30816.

Genetically distant breeding samples falling into different cluster groups and components can be used as sources of starting

material to achieve genetic diversity in the breeding process when creating new lines and varieties of triticale.

REFERENCES

Arriel, N., A. Mauro, E. Arriel, M. Costa, I. Barbaro , F. Muniz, 2007. Genetic divergence in sesame based on morphological and agronomic traits. Crop Breed Appl Biotechnol 7, 253-261.

Baichev, V., 2009. Economic characteristics of the newly recognized variety of triticale Attila. Field crops studies. Volume V-1, 79-85.

Bhatt, J. M., 1970. Multivariate analysis approach to selection of parents for hybridization aiming at yield components in self-pollination crops. Aust J Agric Rec 21: 1-7.

Biabani, A., H. Pakniyat, 2008. Evaluation of seed yield-related characters in sesame (Sesamum indicum L.) using factor and path analysis. Pakistan Journal of Biological Sciences 11(8):1157-60.

Dragov, R., D. Dechev, 2016. Genetic distance by important economic traits between Bulgarian and foreign durum wheat varieties. Science & Technologies, Volume VI, Number 6: Agrobiological Science, 41-47.

Eivazi, A. R., M. Naghati, M. Hajheidari, S. M. Pirseyedi, M. R. Ghaffari, S. A. Mohhamadi, I. Majidi, G. Salekdeh, M. Mardi, 2007. Assesing wheat (Triticum aestivum L.) genetic diversity using quality traits, Amplified fragment length polymorphisms, simple sequence repeats and proteome analysis. Ann Appl Biol. 152: 81-91.

Fang, X., E. Xiong, W. Zhu, 1996. Cluster analysis of elite wheat germplasm. Jiangsu Agric Sci 4, 14-16.

Islam, M. R., 2004. Genetic diversity in irrigated rice. Pak J Biol Sci 2, 226-229.

Kabir, M., A. Khan, M. Hassain, 2009. Genetic divergence in pointed gourd. J Agric Rural Dev 7 (1&2), 87-92.

Khodadadi, M, M. Fotokian, M. Miransari, 2011. Genetic diversity of wheat (Triticum aestivum L.) genotypes based on cluster and principal component analyses for breeding strategies. Aust J Crop Sci 5(1), 17-24.

Kirchev, H., 2019. Triticale. Monograph.

Mergoum, M., , S. Sapkota, A. ElFatih A. ElDoliefy, S. Naraghi, S. Pirseyedi, M. Alamri, W. AbuHammad, 2019. Advances in Plant Breeding Strategies: Cereals pp 405-451| Cite as Triticale (x Triticosecale Wittmack) Breeding.

Mohammadi, S. A., Prassana, B. M., 2003. Analysis of genetic diversity in crop plants: salient statical tools and considerations. Crop Sci. 43: 1235-1248.

Rachovska, G., Z. Uhr, 2010. Inheritance, heterosis and variability of quantitative traits related to productivity in F1 hybrids of common winter wheat. Field Crop Studies, 6 (3), 361-367.

Shanin, J., 1977. Methodology of field experience. Publishing house of the Bulgarian Academy of Sciences, Sofia, 134-142.

Siahbidi, M., A. Aboughadareh, G. Tahmasebi, M. Teymoori, M. Jasemi, (2013): International journal of Agriculture: Research and Review. Vol., 3 (1), 184-194.

Stamatov, S., M. Deshev, 2012. Evaluation of the possibilities for increasing the number of fruit boxes in sesame (Sesamum indicum L.) using factor analysis. Plant Sciences, Year XLIX No. 6, pp.81-84.

Stankov, I., I. Yanchev, T. Raycheva, 2014. A New Stage of Triticale Breeding and

Production in Bulgaria. Soil Science Agrochemistry and Ecology, Vol. XLVIII, № 3-4, 86-89.

Stoyanov, H., 2013. Correlation between Spike Characteristics of Common Winter Wheat (Triticum aesitvim L.) Varieties. Breeding and genetic studies in field crops, 95-100.

Yankov, B., G. Moskov, J. Terziev, H. Yancheva, 2002. Plant breeding, 39-45.

Zhang, W., A. Li, J. Tian, L. Zhao, 2012. Development of Near Isogenic Lines of Wheat Carrying Different Spike Branching Genes and Their Agronomic and Spike Characters. Journal of Agricultural Science, 4-8, 215-221.