MULTIVARIATE ANALYSIS OF QUANTITATIVE TRAITS CAN EFFECTIVELY CLASSIFY RAPESEED GERMPLASM

Mirjana JANKULOVSKA[1*], Sonja IVANOVSKA[1], Ana MARJANOVIC-JEROMELA[2], Snjezana BOLARIC[3], Ljupcho JANKULOSKI[1], Zoran DIMOV [1], Dane BOSEV[1], Biljana KUZMANOVSKA[1]

[1]Faculty ofAgricultural Sciences and Food, Ss. Cyril and Methodius University, Skopje, Macedonia
[2]Institute of Field and Vegetable Crops, Novi Sad, Serbia
[3]Faculty of Agriculture, University of Zagreb, Zagreb, Croatia

Jankulovska M., S. Ivanovska, A. Marjanovic-Jeromela, S. Bolaric, LJ. Jankuloski, Z. Dimov, D. Bosev, B. Kuzmanovska (2014): *Multivariate analysis of quantitative traits can effectively classify rapeseed germplasm.* Genetika, Vol 46, No. 2, 545-559.

In this study, the use of different multivariate approaches to classify rapeseed genotypes based on quantitative traits has been presented. Tree regression analysis, PCA analysis and two-way cluster analysis were applied in order todescribe and understand the extent of genetic variability in spring rapeseed genotype by trait data.

The traits which highly influenced seed and oil yield in rapeseed were successfully identified by the tree regression analysis. Principal predictor for both response variables was number of pods per plant (NP). NP and 1000 seed weight could help in the selection of high yielding genotypes. High values for both traits and oil content could lead to high oil yielding genotypes. These traits may serve as indirect selection criteria and can lead to improvement of seed and oil yield in rapeseed. Quantitative traits that explained most of the variability in the studied germplasm were classified using principal component analysis. In this data set, five PCs were identified, out of which the first three PCs explained 63% of the total variance. It helped in facilitating the choice of variables based on which the genotypes' clustering could be performed. The two-way cluster analysissimultaneously clustered genotypes and quantitative traits. The final number of clusters was determined using bootstrapping technique. This approach provided clear overview on the variability of the analyzed genotypes. The genotypes that have similar performance regarding the traits included in this study can be easily detected on the heatmap.

_____

Corresponding author: Mirjana Jankulovska, Department of Genetics and Plant Breeding, Faculty of Agricultural Sciences and Food, Ss Cyril and Methodius University in Skopje, blvd. Aleksandar Makedonski bb, 1000 Skopje, Republic of Macedonia, tel: ++389 2 3115 277, fax: ++389 2 3134 210, e-mail: mjankulovska@zf.ukim.edu.mk

Genotypes grouped in the clusters 1 and 8 had high values for seed and oil yield, and relatively short vegetative growth duration period and those in cluster 9, combined moderate to low values for vegetative growth duration and moderate to high seed and oil yield. These genotypes should be further exploited and implemented in the rapeseed breeding program. The combined application of these multivariate methods can assist in deciding how, and based on which traits to select the genotypes, especially in early generations, at the beginning of a breeding program.

Key words: multivariate techniques, principal component analysis, rapeseed tree regression analysis, two-way cluster analysis,

,

INTRODUCTION

Oilseed rape (*Brassica napus* L.) is the most important oilseed crop in Europe (BUS *et al*., 2011) and the second most important oilseed crop in the world after soybean (BASUNANDA *et al*., 2010). Ensuring its efficient production has forced rapeseed breeders to create cultivars that combined high yield, high oil content and specific quality traits (low levels of erucic acid in oil and glucosinolates in seeds). This trend has narrowed the genetic basis of elite oilseed rape breeding material and, as a consequence, genetic variability in this crop is limited regarding many characters of value for breeders (COWLING, 2007; ANANGA *et al*., 2008). Rapeseed cultivars used in Europe are generally of very high quality; however some desirable traits are missing in European gene-pool (HU *et al*., 2003).

The information on the genetic variation in *B. napus* could help breeders and geneticists to understand the structure of their germplasm and may help them to predict which combinations would produce the best offspring. There are various techniques available for evaluation of genetic variation, such as morphological, biochemical and molecular markers (MARJANOVIC-JEROMELA *et al*., 2009). Morphological characterization continues to be the first step in the description and classification of germplasm. The application of appropriate statistical methods is a useful tool for the initial description and classification of genotypes, since it enables plant breeders to identify and select valuable genetic resources for direct use by farmers or for implementation in a breeding program to improve different agronomic traits.

Seed yield is a complex trait that includes various components and finally results in a highly plastic yield structure (DIEPENBROCK, 2000). It is a result of physiological and morphological processes that occur during the phenological development. Yield components have significant role in the final yield performance of the plant, meaning that the indirect selection for traits that influence yield can be effective approach in a rapeseed breeding program (SADAT *et al*., 2010; MARJANOVIC-JEROMELA *et al*., 2011). One of the important factors for understanding yield constraints is the type of model used to analyze the data.

Lately, the inaccuracy and unstable solutions derived from the application of linear regression models when dealing with data sets that have large number of insignificant predictor variables or in the cases when strong interactions between variables exist (HASTIE *et al*., 2001), leaded to development of various alternatives to linear models. The regression trees are non-parametric statistical methods which are used to establish relationships between the predictors and the response variables by splitting the sample using the most influential predictor variable and assigning constant values to each resulting sub-domain in a recursive process (BREIMANN *et al*., 1984). The final separation of the data appears like a tree with two branches derived from

each node (RAWLS and PACHEPSKI, 2002).The most important features of regression trees are their ability to handle nonlinear relationships,wide range of data types, including numerical and categorical data. The obtained results are easy to interpret and the variable which significantly distinguishes the classes is clearly appointed (CRAWLEY, 2007). Additionally, problems caused by multicollinearity are not present as they are dealing with the predictors one by one. Regression trees first became popular in environmental sciences (LEES and RITMAN, 1991; BAKER, 1993), and were later used in studies on land quality assessment and soil properties estimation (MCKENZIE and RYAN, 1999). This model has increasingly been applied in agricultural sciences (LOBELL *et al*., 2005; FERRARO *et al*. 2009; TITTONELL *et al*., 2008) to identify the effect of different management practices on crop yield. However, the information on the use of regression trees in plant breeding programs is lacking. Using regression trees as a breeding tool may help detect the interrelations between variables, describing the variability of germplasm or identification of quantitative traits which influence seed yield.

Statistical methods, such as principal components analysis (PCA), are useful tools for evaluation of germplasm (CANTINI *et al*. 1999; BADENES *et al*. 2000). This technique is an ordination method often used to simultaneously describe the relationships between sets of variables (here, genotypes and quantitative traits). It tends to reduce the dimension of multivariate data by removing inter-correlation among variables and allows a multi-dimensional relationship to be plotted on two or three principal axes (HAYMAN, 1967). As a result, PCA allows visualization of the differences among the individuals, identification of possible groups and finding relationships among individuals and variables (MARTINEZ-CALVO *et al*. 2008).

Cluster analysis is used to describe and represent the structure of the pairwise dissimilarities amongst objects. Clustering methods order objects (genotypes) or variables (environments, traits) in groups that are similar with respect to some measure. Regular clustering, i.e. one-way clustering, aims at finding the best partitioning in one direction of a two-way table or data matrix. However, some modern clustering problems (CIAMPI *et al*., 2005), pose a new challenge: not only to describe dissimilarities relationships among individuals and variables, but also to discover groups of variables and of individuals such that the variables are useful in describing the dissimilarities amongst the individuals and vice versa. To this end, techniques known as two-way clustering and crossed classification clustering have been developed. As opposed to regular one-way clustering, two-way or two-mode clustering aims to find the best partitioning of the data in two directions (both genotypes and environments/traits). The added benefit in comparison with one-way clustering is that it becomes immediately clear why certain objects have been clustered together, since their variables have also been clustered simultaneously (HAGEMAN *et al*., 2012).

In order to increase the existing genetic variability and to develop rapeseed cultivars that will have high seed and oil yield potential, exploitation of the gene pool is of foremost importance. Many researchers have been evaluating the diversity in different collections of rapeseed genotypes using multivariate techniques (CHOUDHARY and JOSHI, 2001; ALEMAYEHU and BECKER, 2002; HU *et al*., 2007; MARJANOVIC-JEROMELA *et al*., 2009; ASGHARI *et al*., 2011). However, very little information is available on the spring rapeseed genotypes created in the South Eastern Europe. Moreover, tree regression analysis and two-way cluster analysis have not been used for characterization of rapeseed germplasm or for identification of superior hybrids. Toward these ends, three multivariate methods were applied todescribe and understand the extent of genetic variation in spring oilseed rape genotypes based on several quantitative traits. The

objectives were a) to assess the extend and patterns of variability within the created material in a plant breeding programme, b) to examine different approaches for classification of rapeseed genotypes based on quantitative traits, c) to identify quantitative traits which explain most of the variability of the genotypes, and d) to establish an effective method that could facilitate the selection of superior genotypes for utilization in field production or as parents in future hybridization program.

MATERIALS AND METHODS

Plant material and experimental design

     In this study, 55 spring rapeseed genotypes were included. Out of them, 10 were registered varieties and advanced breeding lines of different origin, and the other 45 were hybrid combinations obtained as a result of diallel crossing (without reciprocals) of the previous 10 genotypes (Table 1).

Table 1. Spring rapeseed genotypes included in the study

1	Lisora (Germany)	16	Lisora x Ratnik-3	31	Mira x Lisonne	46	Lisora x Jovana
2	Lisora x Jr-ns-6	17	Jr-ns-6 x Ratnik-3	32	Jr-ns-36 x Lisonne	47	Jr-ns-6 x Jovana
3	Jr-ns-6 (Serbia)	18	Mira x Ratnik-3	33	Global x Lisonne	48	Mira x Jovana
4	Lisora x Mira	19	Jr-ns-36 x Ratnik-3	34	Ratnik-3 x Lisonne	49	Jr-ns-36 x Jovana
5	Jr-ns-6 x Mira	20	Global x Ratnik-3	35	Jr-ns-44 x Lisonne	50	Global x Jovana
6	Mira (Serbia)	21	Ratnik-3 (Serbia)	36	Lisonne (Germany)	51	Ratnik-3 x Jovana
7	Lisora x Jr-ns-36	22	Lisora x Jr-ns-44	37	Lisora x Liaison	52	Jr-ns-44 x Jovana
8	Jr-ns-6 x Jr-ns-36	23	Jr-ns-6 x Jr-ns-44	38	Jr-ns-6 x Liaison	53	Lisonne x Jovana
9	Mira x Jr-ns-36	24	Mira x Jr-ns-44	39	Mira x Liaison	54	Liaison x Jovana
10	Jr-ns-36 (Serbia)	25	Jr-ns-36 x Jr-ns-44	40	Jr-ns-36 x Liaison	55	Jovana (Serbia)
11	Lisora x Global	26	Global x Jr-ns-44	41	Global x Liaison		
12	Jr-ns-6 x Global	27	Ratnik-3 x Jr-ns-44	42	Ratnik-3 x Liaison		
13	Mira x Global	28	Jr-ns-44 (Serbia)	43	Jr-ns-44 x Liaison		
14	Jr-ns-36 x Global	29	Lisora x Lisonne	44	Lisonne x Liaison		
15	Global (Sweden)	30	Jr-ns-6 x Lisonne	45	Liaison (Germany)		

     The genotypes were grown in 2010, at experimental field near Skopje, Macedonia. The experimental design was RCBD, with five replications, out of which two replications were irrigated and the other three were non-irrigated. Each plot consisted of 3 rows, 2 m long, with 25 cm distance between rows and 5 cm in the row. Standard crop management practices were applied during the growth season. The data were collected on 10 randomly selected plants from the middle row of each plot. The following traits have been analyzed: plant height (PH), number of primary branches per plant (NB), number of pods per plant (NP), pod length (PL), number of

seeds per pod (NSP), seed weight per pod (WSP), 1000 seed weight (W1000), seed weight per plant (WSPL) and oil content (O). Oil yield per plant (OY) was estimated on the basis of seed weight per plant and oil content. Days to flowering (DF), flowering duration (FD), ripening duration (RD) and vegetative growth duration (VG) have been determined on a plot basis.

Statistical analyses

In order to identify the extend of variability of the genotypes included in the study, the effect that the various analyzed traits have on the expressed variability and to classify the genotypes based on their variability, the following statistical analyses were performed: tree regression analysis, principal component analysis and two-way cluster analysis. The mean values of the genotypes for the analyzed quantitative traits have been submitted to different packages in the R 3.0.3 statistical software.

The basicalgorithm underlying regression trees (BREIMAN *et al*., 1984) is to continuallydivide the data set intohomogeneous groups, assessing the relative''impurity'' of the data before and after a split, in order to establish themost parsimonious tree. The response variable (i.e. yield) is modeled as a piece-wise constant function. Regression trees work by splitting the data first into two subsets based on the predictor variable and its value, whichresults in the greatest increase of the explained variance of the response variable. Each split, called a daughter node, is then treated independently, as its own dataset and the process is repeated recursively. The output can be displayed as a tree-like Figure, with criteria for each split (i.e. node) labeled according the threshold used to define the split. The regression trees were implemented in the ''rpart''package (THERNEAU *et al*., 2014) and were usedfor identification of quantitative traits which influence WSPL in one case and the OY in the second case.

Principal component analysis was performed using the command "principal" from the "psych" package. This application by default standardizes the data matrix and as a result the component scores are standard scores (mean=0, sd = 1) of the standardized input (REVELLE, 2014). Principal components have been extracted until the eigen value > 1.

Quantitative traits that have been identified to have high correlation with the first three principal components were used for two-way clustering. Their standardized values (mean=0, sd = 1) were utilized for estimating the Euclidian distance between the genotypes. The obtained values were normalized according to ROLDAN-RUIZ *et al*. (2001) and clustered with the UPGMA method for creation of a dendrogram. The estimation of optimal number of clusters was performed by multiscale bootstrap resampling for assessing the uncertainty in hierarchical cluster analysis, originallysuggested for DNA microarray data analysis (SUZUKI and SHIMODAIRA, 2013). For each cluster in hierarchical clustering, the *p*-values are calculated (a value between 0 and 1), which indicates how strong the cluster is supported by data. Identified clusters with approximately unbiased (AU) *p*-values larger than 95% after 100 bootstraping replications were considered to be strongly supported by data. The resulting dendrogram was used in the heatmap, as a row dendrogram. The column dendrogram was organized according to the predetermined row dendrogram. "Heatmap.2" command from the "heatmap.plus" package (DAY, 2012) was applied for generating the two-way cluster.

RESULTS AND DISCUSSION

Tree regression analysis

Regression trees, which provide a simple means of capturing nonlinear relationships and variable interactions, appear to be a valuable tool for identifying significant yield constraints. The results of the regression tree model indicated that number of pods per plant was the most important variable determining WSPL (Figure 1) and OY (Figure 2). The number of genotypes and their average seed and oil yield is shown for each terminal node on both Figures. The observations that satisfy the criterion at a given split fall to the left-hand daughter node while those not meeting the criterion continue to the right node. As a result, the genotypes of interest (having high seed and oil yield) are arranged on the right side of the tree. 17 genotypes that have NP<51.45 were positioned on the left side of the tree in both cases. No further splitting on the left side indicates that number of pods per plant is the main constraint for both high seed and oil yield.

The remaining 38 genotypes were again split based on the NP in the case of WSPL. Out of them, 23 genotypes were separated in two nodes: eight genotypes that have W1000 less than 2.36, for which WSPL was 2.466 g and fifteen genotypes with more than 2.36 g W1000, with average WSPL of 3.085 g. The remaining 15 genotypes were positioned on the right side of the tree. They had more than 69.83 pods per plant and average seed yield per plant of 3.639 g (Figure 1).

When analyzing quantitative traits that influence the variability of oil yield per plant, similar pattern can be observed. Namely, after NP, the next differentiation criterion for the genotypes is oil content in the seed. Twelve genotypes that have less than 39.17% oil in the seed and average oil yield per plant of 1.075 g are positioned on the left side of the tree. The W1000 was of tertiary importance, dividing 26 genotypes on two terminal nodes. There are 18 genotypes that had less than 3.045g W1000 and average oil yield of 1.288 g on the left side. Only eight genotypes had more than 3.045 g W1000 and average oil yield of 1.654 g (Figure 2). In both cases, genotypes classified in the furthest right node should be considered for implementation in a breeding program for improving seed or oil yield in rapeseed. Furthermore, for the analyzed genotypes, NP and W1000 were the main constraints for seed yield and NP, O and W1000 for oil yield.

The tree regression analysis of this particular dataset provided insight into the structure of relationship between rapeseed seed and oil yield and analyzed quantitative traits. As a result, multiple traits which should be targeted at the first stages of the selection process were defined. This approach could be used to apply different breeding strategies. The indirect selection for a combination of high values for these traits may be highly effective at the first years of progeny testing in rapeseed, as it could lead to improvement of seed and oil yield in rapeseed.

Although regression tree technique allowed identification of yield determining traits, it has some limitations considering issues that might be of interest at the beginning of a breeding program.Alternative classification techniques should be explored and used in combination with the regression tree to partition the genotypes into groups with similar traits.
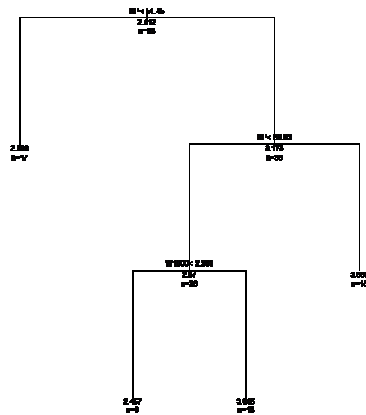
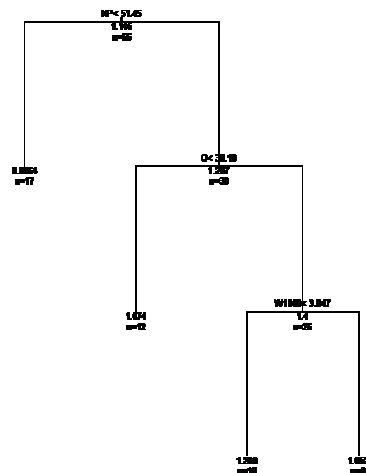*Figure 1. Tree regression analysis for weight of seed pe plant*

Figure 2. Tree regression analysis for oil yield per plant

Principal component analysis

  The principal component analysis identified five PCs with eigen values > 1, which explained 80% of the total genotypes variation (Table 2). BELETE (2011) reported 91.4% variation contributed by first five principal components in Ethiopian mustard, while AFRIN *et al.* (2012) identified that the first three PCs accounted for 68.93% of the variation among the rapeseed genotypes used in their study.

Table 2. PCs scores for the analyzed quantitative traits

Quantitative traits	Factor loadings				
	PC1	PC2	PC3	PC4	PC5
PH	0.20	0.58	0.40	-0.42	-0.10
NB	0.79	0.03	-0.04	-0.39	0.06
NP	0.84	0.10	0.21	-0.27	-0.02
PL	-0.17	0.27	0.77	0.30	0.26
NSP	0.45	0.13	0.49	0.49	0.35
WSP	0.69	-0.02	-0.31	0.47	0.26
W1000	0.71	0.11	-0.50	0.24	0.07
WSPL	0.90	0.34	-0.07	-0.03	-0.03
O	0.05	0.09	0.21	0.39	-0.80
OY	0.88	0.35	-0.02	0.05	-0.19
DF	-0.54	0.45	-0.31	0.14	0.20
FD	0.08	-0.17	0.10	-0.47	0.24
RD	-0.34	0.86	-0.14	-0.03	-0.04
VG	-0.46	0.83	-0.22	-0.05	0.09
SS loadings	4.81	2.34	1.60	1.40	1.07
Proportion of Variance	0.34	0.17	0.11	0.10	0.08
Cumulative Variance	0.34	0.51	0.63	0.73	0.80

  The first PC explained 34% of the total variability. WSPL and OY, which are the main goals in the breeding program, showed the highest correlation with the first PC, followed by NP, NB, W1000 and WSP. The second PC was mostly associated with PH, DF, RD and VG and explained 17% of the total variance. PL and NSP were correlated with the third PC, which accounted for the 11% of the total variance. Oil content had the highest positive correlation with the fourth PC, while FD was positively associated with the PC5. These two PCs explained 10% and 8% of the variance, accordingly. Similar results have been observed by AFRIN *et al.* (2012), DAR *et al.* (2010) and ISLAM and ISLAM (2000). CHOUDHARY and JOSHI (2001) concluded that plant height, secondary branches per plant, days to flowering and 1000-seed weight contributed maximum towards genetic divergence. The factor loadings for the first two PCs were plotted on Figure 3. The orientation of the genotypes according to the values of different analyzed traits can be detected on the Figure 3, but the difference between genotypes and the traits based on which the classification is performed cannot be clearly recognized. Even so, the PC analysis identified traits which contributed most to the variation of the analyzed genotypes and can serve as a useful tool for facilitating the selection of desirable characteristics in rapeseed breeding.
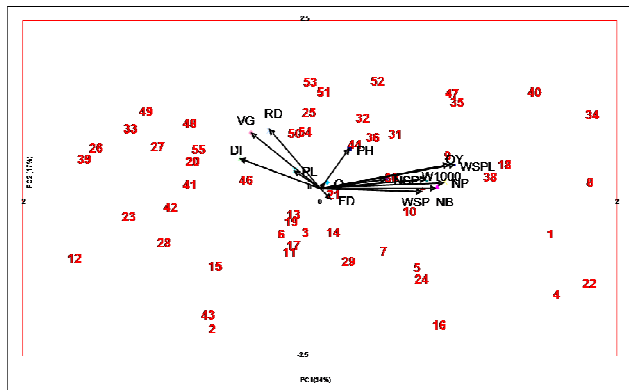
Figure 3. PC1 and PC2 factor loadings for genotypes and quantitative traits

Two-way cluster analysis

Most of the variability of the analyzed genotypes has been explained by the first three PCs. The traits that had highest correlation with them were used for the two-way cluster analysis. The multiscale bootstrap resampling identified that the optimal number of clusters for the analyzed data is nine clusters for the genotypes. The clusters determined by the bootstraping technique are presented in Figure 4. Many other researchers were performing cluster analysis based on quantitative traits in rapeseed (ALI *et al*., 1995; DHILLON *et al*., 1999; HU *et al*., 2007; MARJANOVIC-JEROMELA *et al*., 2009) and obtained different clustering based on the material used in their studies, analyzed traits and the environmental conditions in which the observations were performed. The one-way clustering gives an overview on the classification of genotypes in different groups, but the quantitative traits used as a basis for the performed clustering can not be seen. Therefore, additional explanation is needed.
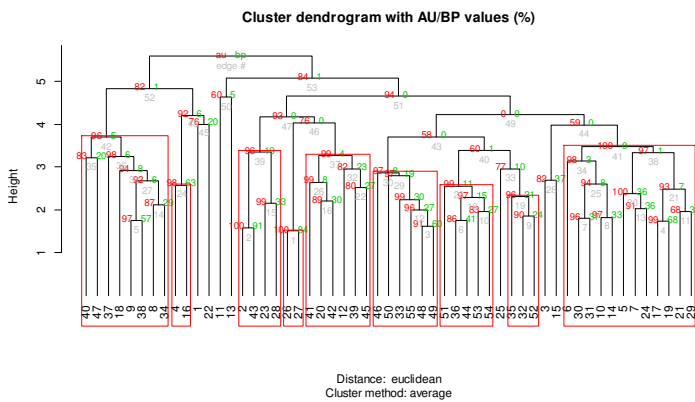

Figure 4.Dendrogram representing clustering of the analyzed rapeseed genotypes
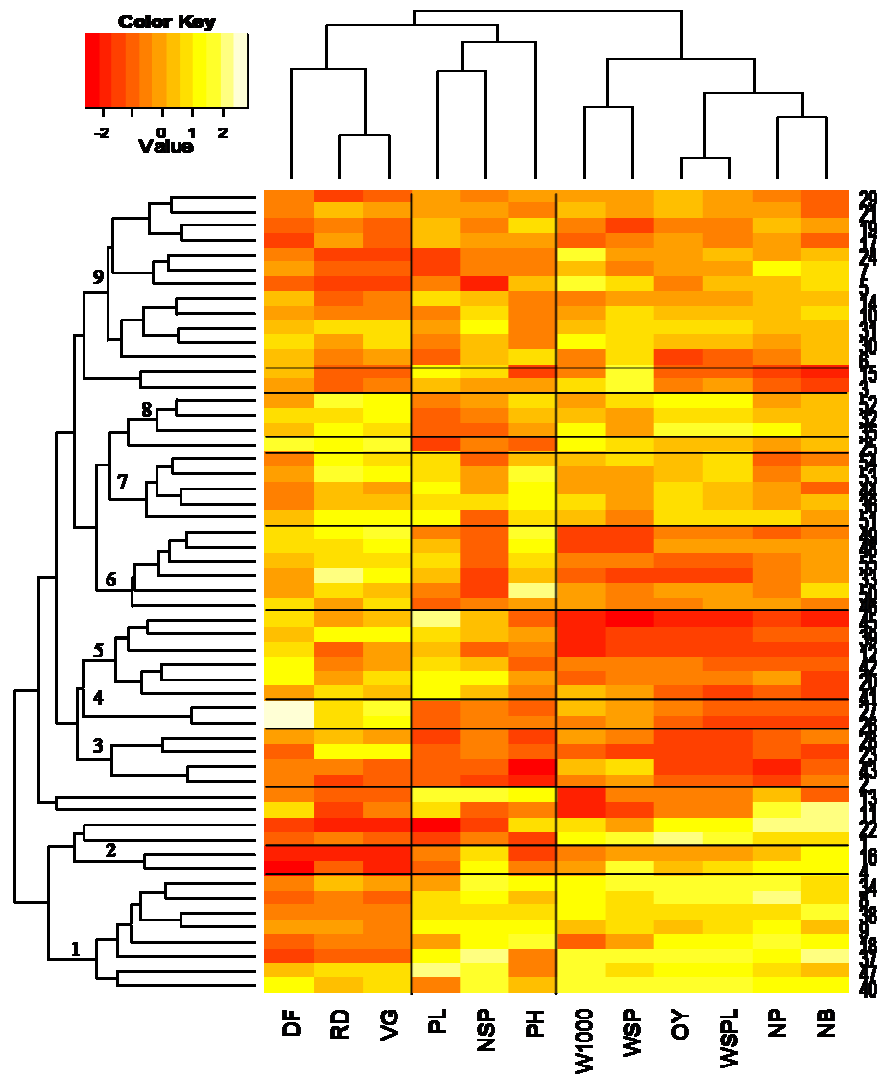
Figure 5. Two-way cluster explaining the effect of analyzed traits on genotypes classification

As a supplement, the heatmap presenting two-way classification of the traits and genotypes is given in Figure 5. The characters used for clustering were divided in three main groups. The first group comprised the same traits that were associated with the PC1. PH, PL and NSP were grouped in the second cluster, while the traits associated with the vegetative development (DF, DR, VG) belonged to the third group. Such classification of the traits allows an effective overview and enables an easy determination of the characteristics based on which

the breeding material has been grouped. As explained by HAGEMAN *et al*. (2012), the two-mode clustering is able to extract relevant features from the data set. In our study, this clustering technique identified genotypes that share a common set of characteristics. The first cluster comprised genotypes 40, 47, 37, 18, 9, 38, 8 and 34, which are characterized by high values for seed and oil yield, and relatively short duration of phenological development. The genotypes in cluster 3 had similar phenological characteristics as those from cluster 1, but they had lower average values for all other analyzed traits. Clusters 3, 4, 5 and 6 comprised genotypes that had low to moderate values for the traits from the first group (WSPL, OY, NP, NB, W1000 and WSP) and high values for the traits from the third group. The genotypes classified in cluster 9 (6, 30, 31, 10, 14, 5, 7, 24, 17, 19, 21 and 29) and cluster 2 (4, 16) are characterized by low values for the duration of vegetative growth stages and high seed and oil yield. The highest average values for weight of seed per plant and oil yield were detected in clusters 1 and 8. The genotypes that belong to these two groups can be considered for future use in the breeding program. Genotypes that do not belong to any of the groups can be used for improvement of certain traits, for which they differ from the genotypes included in the determined clusters. Furthermore, the clusters 1, 2, 4 and 8 did not comprise the parents used for diallel crossing, indicating that as a result of the hybridization more divergent progeny has been obtained. The extend of variation in the breeding material was clearly illustrated and the genotypes were effectively classified on the heatmap. This technique could be used as an efficient tool for selecting genotypes based on desired traits in the early stages of the breeding process. Moreover, two-way cluster analysis may explain the diversity patterns in different germplasm collections.

CONCLUSION

The applied multivariate statistical models were effective for analysis of the variability in rapeseed genotypes, and for their classification in homogenous groups based on several quantitative traits.

Tree regression analysis successfully identified the traits which mostly influenced the seed and oil yield. Number of pods per plant was the principal predictor for both response variables. High seed yielding genotypes could be selected based on this trait and weight of 1000 seeds. The selection of genotypes based on high values for NP, oil content and W1000 may lead to improvement of oil yield in the studied breeding material. This approach can be applied for identification of indirect selection criteria for seed and oil yield in a rapeseed breeding program or for recognition of the characters that have the highest effect on the variability of a certain response variable.

The principal component analysis was effective for classificationof the quantitative traits that explained most of the variability of the studied genotypes. In this data set, the characters were appointed to five PCs, out of which the first three PCs explained 63% of the total variance. It served as a useful tool for detection of traits for which the genotypes expressed the highest differences and facilitated the choice of variables based on which the clustering of the germplasm could be performed.

The two-way cluster analysis, compared to one-way clustering, had the advantage of simultaneously clustering the genotypes and thequantitative traits in a dataset. The genotypes that have similar performance can be easily detected on the heatmap. This model gives a clear overview on the variability of the existing germplasm and can assist in deciding how, and based on which traits to select the genotypes, especially in early generations, at the beginning of a breeding program. The genotypes grouped in the cluster 1 and 8, whichhad high values for seed

and oil yield, and relatively short vegetative growth duration period and in the cluster 9, combining moderate to low values for the duration of vegetative growth and moderate to high seed and oil yield, could be considered for future implementation in a rapeseed breeding program.

The application of these three methods was useful for classification and characterization of rapeseed genotypes. For characterization of different germplasm collections, or a starting material in a breeding program, it should be further exploited.

REFERENCES

AFRIN K.S., F. MAHMUD, MD.S.R. BHUIYAN, MD.A. RAHIM (2012): Genetic Divergence in Advanced Lines of Oilseed Rape (*Brassica napus* ssp. *oleifera* L.). Agriculturae Conspectus Scientificus. *77* (2): 81-85.

ALEMAYEHU N., H. BECKER (2002): Genotypic diversity and patterns of variation in a germplasm material of Ethiopian mustard (*Brassica carinata*A. Braun).Genetic Resources and Crop Evolution *49*: 573-582.

ALI M., L.O. COPENLAND, S.G. ELIAS, J.D. KELLY (1995): Relationship between genetic distanceand heterosis for yield and morphological traits in winter canola (*Brassica napus*L.).Theor Appl. Genet. *91*: 118-121.

ANANGA A., E. CEBERT, K. SOLIMAN, R. KANTETY, K. KONAN, J.W. OSHIENG (2008): Phylogenetic relationships within and among *Brassica* species from RAPD loci associated with blackleg resistance. Afr. J. Biotechnol. *7(9):* 1287-1293.

ASGHARI A., M. SHOKRPOUR, M.H. CHAMANABAD, O. SOFALIAN (2011): Evaluating genetic diversity of canola cultivars using morphological traits and molecular markers. Romanian Biotechnological Letters *16(4):*6305-6312.

BADENES M.L., J. MARTÍNEZ-CALVO, G. LLÁCER (2000): Analysis of a germplasm collection of loquat (*Eriobotrya japonica* Lindl.). Euphytica *114:* 187-194.

BAKER F.A. (1993): Classification and regression tree analysis for asessing hazard of pine mortality caused by *Heterobasidion annosum*. Plant Dis. *77:*136–139.

BASUNANDA P., M. RADOEV, W. ECKE, W. FRIEDT, H.C. BECKER, R.J. SNOWDON (2010): Comparative mapping of quantitative trait loci involved in heterosis for seedling and yield traits in oilseed rape (*Brassica napus* L.). Theor. Appl. Genet. *120:* 271–281.

BELETE Y.S. (2011): Genetic Variability, Correlation and Path Analysis Studies in Ethiopian Mustard *(Brassica carinata*A. Brun) Genotypes. International Journal of Plant Breeding and Genetics *5:* 328-338.

BREIMAN L., J.H. FRIEDMAN, R.A. OLSHEN, C.J. STONE (1984): Classification and Regression Trees. Wadsworth.

BUS A., N. KORBER, R.J. SNOWDON, B. STICH (2011): Patterns of molecular variation in a species-wide germplasm set of *Brassica napus*. Theor. Appl. Gene**t**. *123(8):*1413-23.

CANTINI C., A. CIMATO, G. SANI (1999): Morphological evaluation of olive germplasm present in Tuscany region. Euphytica *109:*173– 181.

CHOUDHARY B. R., P. JOSHI (2001): Genetic diversity in advanced derivatives of *Brassica* interspecific hybrids. Euphytica*121:* 1-7.

CIAMPI A., A.G. MARCOS, M.C. LIMAS (2005): Correspondence analysis and two-way clustering. SORT *29(1):*27-42.

CRAWLEY M.J. (2007): Regression. The R Book, John Wiley & Sons, Ltd, Chichester, UK.

COWLING W.A. (2007): Genetic diversity in Australian canola and implications for crop breeding for changing future environments. Field Crop. Res. *104:* 103-111.

DAR Z., S.A. WANI, A.R.G. ZAFF, A. ISHFAQ, M.A. WANI, M. HABIB, M. KHAN, S. RAZVI (2010): Genetic Divergence Studies in Brown Sarson (*Brassica rapa* L.). Research Journal of Agricultural Sciences *1:* 140-141.

DAY A. (2012): Package "heatmap.plus": Heatmap with more sensible behavior.

http://cran.r-project.org/web/packages/heatmap.plus/heatmap.plus.pdf

DHILLON S.S., K. SINGH, K.S. BRAR (1999): Diversity analysis of highly selected genotypes inIndian mustard (*Brassica juncea* Czern&Coss). 10th International Rapeseed Congress, 26-29 September, Canberra, Australia.

DIEPENBROCK W. (2000): Yield analysis of winter oilseed rape (*Brassica napus* L.): a review. Field Crop Res. *67:* 35-49.

FERRARO D.O., D.E. RIVERO, C.M. GHERSA (2009): An analysis of the factors that influence sugarcane yield in Northern Argentina using classification and regression trees. Field Crop Res., *112:* 149-157.

HAGEMAN J.A., M. MALOSETTI, F.A. VAN EEUWIJK (2012): Two-mode clustering of genotype by trait and genotype by environment data. Euphytica *183:*349–359

HASAN M., F.SEYIS, A.G. BADANI, J. PONS-KÜHNEMANN, W. FRIEDT, W. LÜHS, R.J. SNOWDON (2006): Analysis of genetic diversity in the *Brassica napus* L. gene pool using SSR markers. Genet. Resour. Crop. Evol. *53:* 793-802.

HASTIE T., R. TIBSHIRANI, J. FRIEDMAN (2001): The elements of statistical learning: Data mining, inference, and prediction. Springer, New York.

HAYMAN H.H. (1967): Modern factor analysis. 2nd edition, University of Chicago press, Chicago, p. 474.

HU S.W., J. OVESNÁ, L. KUČERA, V. KUČERA, M. VYVADILOVÁ (2003): Evaluation of genetic diversity of *Brassica napus* germplasm from China and Europe assessed by RAPD markers. Plant, Soil and Environment *49:* 106–113.

HU S.W., C.Y. YU, H.X. ZHAO, G.L. SUN, S.L. ZHAO, M. VYVADILOVA, R. KUCERA (2007): Genetic Diversity of Brassica napus Germplasm from China and Europe Assessed by Some Agronomically Important Characters. Euphytica *154:* 9-16.

ISLAM M.S., M.O. ISLAM (2000): Genetic diversity in rapeseed and mustard (*Brassica sp.*). Bangladesh J. Pl. Breed.Genet.*13:* 25-30.

LEES B.G., K. RITMAN (1991): Decision-tree and rule-induction approach to integration of remotely sensed and GIS data in mapping vegetation in disturbed or hilly environments. Environ. Man-age. *15:* 823–831.

LOBELL B.D., J.I. ORTIZ-MONASTERIO, G.P. ASNER, R.L. NAYLOR, W.P. FALCON (2005): Combining Field Surveys, Remote Sensing, and Regression Trees to Understand Yield Variationsin an Irrigated Wheat Landscape. Agron. J. *97:*241–249

MARJANOVIC-JEROMELA A., A. KONDIC-SPIKA, D. SAFTIC-PANKOVIC, R. MARINKOVIC, N. HRISTOV (2009): Phenotypic and molecular evaluation of genetic diversity of rapeseed (*Brassica napus* L.) genotypes. African Journal of Biotechnology *8 (19):* 4835-4844.

MARJANOVIC-JEROMELA A., R. MARINKOVIC, S. IVANOVSKA, M. JANKULOVSKA, A. MIJIĆ, N. HRISTOV (2011): Variability of yield determining components in winter rapeseed (*Brassica napus* L.) and their correlation with seed yield. Genetika, *43 (1):* 51 -66.

MARTINEZ-CALVO J., A.D. GISBERT, M.C. ALAMAR, R. HERNANDORENA, C. ROMERO, G. LLACER, M.L. BADENES (2008): Study of a germplasm collection of loquat (*Eriobotrya japonica* Lindl.) by multivariate analysis. Gene. Res. Crop Evol. *55(5):* 695-703.

MCKENZIE N.J., P.J. RYAN (1999): Spatial prediction of soil properties using environmental correlation. Geoderma*89:*67–94.

R CORE TEAM (2013): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

RAWLS W.J., Y.A. PACHEPSKY (2002): Soil Consistence and Structure as Predictors of Water Retention. Soil Sci. Soc. Am. J. *66:*1115–1126.

REVELLE W. (2014): Package "psych": Procedures for Psychological, Psychometric, and Personality Research. http://cran.r-project.org/web/packages/psych/psych.pdf

ROLDAN-RUIZ I., F.A. VAN EEUWIJK, T.J. GILLILAND, P. DUBREUIL, C. DILLMANN, J. LALLEMAND, M. DE LOOSE, C.P. BARIL (2001): A comparative study of molecular and morphological methods of describing relationships between perennial ryegrass *Lolium perenne* L.) varieties. Theor. Appl. Genet. *103:* 1138—1150.

SADAT H.A., G.A. NEMATZADEH, N.B. JELODAR, O.G. CHAPI (2010): Genetic evaluation of yield and yield components atadvanced generations in rapeseed (*Brassica napus* L.). African Journal of Agricultural Research *5(15):* 1958-1964.

SUZUKI R., H. SHIMODAIRA (2013): Package "pvclust": Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling.

http://www.is.titech.ac.jp/~shimo/prog/pvclust/pvclust.pdf

THERNEAU T., B. ATKINSON, B. RIPLEY (2014): Package "rpart": Recursive Partitioning and Regression Trees. http://cran.r-project.org/web/packages/rpart/rpart.pdf

TITTONELL P., K.D. SHEPHERD, B. VANLAUWE, K.E. GILLER (2008): Unravelling the effects of soil and crop management on maize productivity in smallholder agricultural systems of western Kenya - An application of classification and regression tree analysis. Agric. Ecosyst. Environ. *123:* 137-150

MULTIVARIJACIONE ANALIZE KVANTITATIVNIH SVOJSTAVA MOGU EFIKASNO DA KLASIFIKUJU GERMPLAZMU ULJANE REPICE

Mirjana JANKULOVSKA[1*], Sonja IVANOVSKA[1], Ana MARJANOVIĆ-JEROMELA[2], Snježana BOLARIĆ [3], Ljupčo JANKULOSKI[1], Zoran DIMOV [1], Dane BOŠEV[1], Biljana KUZMANOVSKA[1]

[1]Fakultet za poljoprivredne nauke i hranu, Univerzitet Sv. Kiril i Metodij, Skoplje, Makedonija
[2]Institut za ratarstvo i povrtarstvo, Novi Sad, Srbija
[3]Poljoprivredni fakultet, Univerzitet u Zagrebu, Zagreb, Hrvatska

U ovom istraživanju je prikazana upotreba različitih multivarijacionih metoda za klasifikovanje germplazme uljane repice na osnovu kvantitativnih svojstava. Sa ciljem da se opiše i razume stepen genetske varijabilnosti genotipova za i na osnovu ispitivanih kvantitativnih svojstava kod jare uljane repice, primenjene su: regresiona analizadrveta, PCA analiza i dvostrana klaster analiza.

Regresiona analiza drveta je uspešno identifikovala svojstva koja su jako uticala na prinos semena i ulja kod uljane repice. Glavni indikator za obe zavisne promenjive je bilo svojstvo broj ljuski po biljci. Ovo svojstvo, zajedno sa apsolutnom masom, može olakšati odabir visokoprinosnih genotipova. Visoke vrednosti za ova dva svojstva, kao i za sadržaj ulja, ukazuju na genotipove koje daju visok prinos ulja. Ova svojstva mogu poslužiti kao indirektni kriterijumi za selekciju i doprineti poboljšanju prinosa semena i ulja kod uljane repice.

Kvantitativna svojstva koja su objasnila najveći deo varijabilnost kod ispitivane germplazme su klasifikovana pomoću analize glavnih komponenata. Za ispitivane podatke, identifikovano je 5 glavnih komponenata, od kojih su prve tri objasnile 63% ukupne varijanse. Na ovaj način je olakšan izbor svojstava na osnovu kojih se gentopovi mogu efikasno klasifikovati.

Dvostrana klaster-analiza istovremeno klasifikuje i genotipove i kvantitativna svojstva. Konačan broj klastera određen je butstrap tehnikom. Ovakav pristup omogućuje jasan pregled varijabilnosti analiziranih genotipova. Genotipovi koji imaju sličnu performansu na osnovusvojstava koja su uključena u analizu mogu biti lako detektovani na hitmapi. Genotipovi svrstani u klastere 1 i 8 su se odlikovali visokim vrednostima za prinos semena i ulja, kao i relativno kratkom vegetacijom, dok su oni klasifikovani u klasteru 9 imali kombinaciju srednjih i niskih vrednosti za dužinu vegetacije i srednjih i visokih vrednosti za prinos semena i ulja. Zato bi ove genotipove trebalo dodatno ispitati i uključiti u selekcioni program na uljanoj repici.

Kombinovana primena navedenih multivarijacionih metoda može pomoći u odlučivanju kako i na osnovu kojih svojstava odabirati genotipove, posebno u ranim generacijama, u početnim fazama selekcije.