Scientific Papers. Series A. Agronomy, Vol. LXV, No. 1, 2022
ISSN 2285-5785; ISSN CD-ROM 2285-5793; ISSN Online 2285-5807; ISSN-L 2285-5785

APPLICATION OF CLUSTER ANALYSIS AND PRINCIPAL COMPONENT ANALYSIS FOR THE STUDY OF AGRONOMIC CHARACTERISTICS OF *Virginia tobacco* HYBRIDS COMBINATION

Neli KERANOVA[1], Marina DRUMEVA-YONCHEVA[2], Yonko YONCHEV[3]

[1]Agricultural University - Plovdiv, 12 Mendeleev Blvd, 4000, Plovdiv, Bulgaria
[2]Institute for Tobacco and Tobacco Products-Markovo, Bulgaria
[3]Fruit Growing Institute - Plovdiv, 12 Ostromila, 4004, Plovdiv, Bulgaria

Corresponding author email: nelikeranova@abv.bg

Abstract

Agro-morphological traits of five Virginia tobacco hybrid combinations and standard Virginia 0514 were analyzed. The aim of the study was to group and evaluate hybrid combinations by stem height (SH), leaf number (LN), length of 12th leaf (LL), width of the 12th leaf (WL), days to 50% flowering (D50F), dry leaf yield. Hierarchical cluster analysis and analysis of the main components were applied. The grouping of hybrids in clusters found that Hybrid 27, Hybrid 33 and Hybrid 126 with a common parent component Virginia 385 have larger leave size and higher dry leaf yield, which determines not only their differentiation into a separate cluster, but also their strong distance from the other two hybrids and the Virginia 0514 standard. The Hybrid 126 has the longest growing season. The studied traits were transformed into two factors. The first includes length of 12th leaf and width of the 12th leaf, days to 50% flowering and dry leaf yield and explains 42.8% of the total variance. The second main component consists of plant height and leaf number, explains 32.5%. As a result of the study it was proved that the most effective would be the selection activity, aimed at the selection in the population of the second cluster on the grounds - length and width of the 12th leaf, dry leaf yield and days to 50% flowering.

Key words*: Virginia tobacco, hybrid combinations, morphological traits, dry leaf yield, cluster analysis, Principal Component Analysis (PCA).*

INTRODUCTION

Tobacco (*Nicotiana tabacum* L.) is a crop that is important for the Bulgarian economy. There are four varietal groups of tobacco in our country: Oriental - Basma and Kaba Kulak and large-leaved - Virginia Flue-cured and Burley - Air Cured varieties.

Virginia Tobacco is a typical cigarette type and is the most significant ingredient in American and Virginia blend cigarettes (Campbell, 1989). High productivity is the ultimate goal of any selection program. Knowledge of the relation between plant characteristics is crucial for improving tobacco yield (Ahmed and Mohammad, 2017).Tobacco is produced for its vegetative parts. The value of the yield is determined by the yield of the leaves, the number of harvested leaves, the length, width and shape of the leaves (Butorac, 2004). The study of tobacco genetic diversity is of interest for the conservation of genetic resources, the expansion of the genetic base and practical applications in selection programs (Darvishzadeh et al., 2013).

There are a number of studies on the morphological traits of tobacco (Xiao et al., 2007; Maleki et al., 2011; Mitreski et al., 2018; Risteski I. and Karolina Kočoska, 2014; Ali et al., 2014, etc.). The cluster analysis can significantly increase the efficiency of the selection process (Rosseeva et al., 2012). Ivanov and Dimova (2015) perform analysis and evaluation of maize hybrids by joint application of cluster analysis and analysis of the main components as accompanying methods for more efficient breeing work. Jain S. K. and P. R. Patel (2016) report on the effectiveness of PCA analysis in deciding which agronomic traits should be emphasized in the breeding program.

Tobacco characteristics studies are important for the full utilization of tobacco germplasm resources (MeilingXu et al., 2013).

Darvishzadeh and Maleki (2012) apply a cluster analysis to assess genetic diversity for morphological and agronomic characteristics in the gene pool of oriental tobacco.
It was found that of the studied traits, days up to 50% flowering, dry tobacco yield and the number of leaves per plant are most important for distinguishing tobacco genotypes. Samizadeh Lahiji H. et al. (2013) studied the genetic remoteness of 89 varieties of flue-cured tobacco. A cluster analysis based on morphological features was performed. Ning et al. (2009) reported that two groups and four subgroups were formed after a cluster analysis of the agronomic characteristics of fifteen tobacco genotypes.
In the overall assessment of agronomic characteristics, the result shows that compared to group I, plant height, the distance between the nodes, length and width of leaves from the middle area in group II are greater, and the number of leaves and days from transplanting to flowering are less. Porkabiri et al. (2019) report that length, leaf width and leaf area are the effective criteria for creating high-yielding varieties in the selection for high yield tobacco.
Conducting cluster analysis with the help of modern computer programs makes it possible to determine those traits that contribute most to the resulting distribution of genotypes by class (Rachovska et al., 2002).
The aim of the study is to make a grouping and comparative evaluation of hybrid combinations by plant height, leaf number, leaf size - length of middle leaves, and width of middle leaves, dry leaf yield, days to 50% flowering.

MATERIALS AND METHODS

Five-year investigation was carried out in the experimental field of Tobacco and Tobacco Products Institute, Markovo. Five Virginia tobacco hybrid combinations and standard Virginia 0514 were included in a block design with four replications and 27 m^2 experimental plot (Table 1). Each genotype was analyzed according to the following characteristics: stem height (SH), leaf number (LN), leaf length (LL), leaf width (LW), days to 50% flowering (D50F), dry leaf yield (DLY).
A comparative evaluation of the hybrid combinations with the Virginia 0514 standard was performed on each of the traits at a level of statistical significance of 0.05. A variance analysis was applied, as a result of which the average values and standard error for each of the hybrid forms were calculated.

Table 1. F1 hybrid combinations

F1 hybrid combinations	Breeding method	Origin
H27 (Coker 254 x Virginia 385)	F1	Bulgaria
H33 (Virginia 0594 x Virginia 385)	F1	Bulgaria
H51 (Virginia 0594 x L 825)	F1	Bulgaria
H126 (Virginia 385 x L0543)	F1	Bulgaria
H135 (L 0543 x L 0842)	F1	Bulgaria
Virginia 0514 - standart	F1	Bulgaria

For the grouping of hybrid combinations, hierarchical cluster analysis was applied using the intergroup coupling method and measure of similarity the quadratic Euclidean distance. The result is presented by a dendrogram, visualizing the differentiation of hybrids by the degree of similarity in the studied indicators. The analyzed indicators were transformed into two factors by a Principal Component Analysis (PCA), and the rotation of the factors was performed using the Varimax method. It was found that the data meet the necessary prerequisites for its implementation. The data were standardized in advance. The IBM SPSS Statistics 24 software product was used for statistical processing of the experimental data.

RESULTS AND DISCUSSIONS

The data obtained as a result of the experimental activity, as well as the results of their mathematical processing are presented in Table 2.
The results of the single-factor analysis of variance and the comparative evaluation of the studied variants and the Virginia 0514 standard

show that the hybrid combinations have no proven differences with the Virginia 0514 standard based on the indicators of plant height and leaf number.

The highest degree of variation is found in H27 on both indicators. On the basis of the length of the 12th leaf H27 (58.73 cm), H33 (61.16 cm) and H126 (63.64 cm) differ from Virginia 0514 (58.88 cm), which is due to the longer leaves compared to the standard. For the crosses H 51 (33.61 cm), H27 (34.64 cm), H33 (34.01 cm) and H126 (35.71 cm) proven wider leaves than the standard were reported.

On the basis of days to 50% flowering, only H126 proves longer days to 50% flowering (74.15 days). For the other hybrid combinations, the values are close to those of the standard, so there are no proven differences with it.

During the five-year study period at the Virginia 0514 standard, lower dry leaf yield was reported compared to H27 (334.55 kg da), H33 (329.15 kg/da) and H126 (320.25 kg/da), which determines the presence of proven differences between them at a level of statistical reliability of 0.05.

Table 2. Comparative evaluation of Virginia hybrid tobacco combinations by morphological characteristics, vegetation periods dry leaf yield

Cluster №	Hybrid	Stem height	Leaf number	Length of 12th leaf	Width of 12th leaf	Days to 50% flowering	Dry leaf yield
I	H51	160.86±2.02$^{n.s.}$	24.40±2.62 $^{n.s.}$	58.73±1.01$^{n.s.}$	33.61±0.4*	69.25±1.43$^{n.s}$	295.25±7.26$^{n.s.}$
	Virginia 0514 st	164.50±3.30	25.58±2.05	58.88±0.96	32.12±0.5	67.65±1.79	296.55±4.49
	H135	157.66±1.85 $^{n.s.}$	25.05±2.10 $^{n.s.}$	57.10±0.92$^{n.s.}$	32.54±0.5$^{n.s.}$	69.3±0.75$^{n.s.}$	292.05±3.66$^{n.s.}$
II	H27	164.30±5.05 $^{n.s.}$	25.12±3.65 $^{n.s.}$	61.56±0.44*	34.64±0.3*	68.1±1.18$^{n.s.}$	334.55±3.54*
	H33	160.02±2.10 $^{n.s.}$	25.73±2.51 $^{n.s.}$	61.15±0.63*	34.01±0.5*	69.2±1.5$^{n.s.}$	329.15±3.61*
	H126	161.50±2.57 $^{n.s.}$	25.65±1.82 $^{n.s.}$	63.64±0.50*	35.71±0.4*	74.15±0.61*	320.25±5.22*
Average		161.47	25.25	60.18	33.77	69.61	311.3
SEM		1.07	0.210	0.970	0.540	0.950	7.710
P-Value		0.590	0.552	0.000	0.000	0.008	0.000

As a result of the conducted hierarchical cluster analysis, it was found that the studied hybrid combinations were grouped in two clusters according to the degree of similarity in the studied indicators (Figure 1).

The first cluster includes hybrid combinations with smaller leaf sizes (length and width of the 12th leaf) and lower yields of dry tobacco: H51, Virginia 0514, H135.

The H27, H33 and H126 hybrid combinations have larger leaf sizes, both 12th leaf length and width, and are more productive.

This determines their differentiation into a separate cluster, which joins the first at a maximum Euclidean distance of 25 units, proving the existence of large differences between the two groups.

All three hybrids have a parent component Virginia 385, which is absent in the other variants, which suggests the impact of this variety on the better characteristics of these hybrid combinations.

As the cluster analysis does not perform tests for statistical reliability of the obtained results, it is expedient to combine this method with another approach.

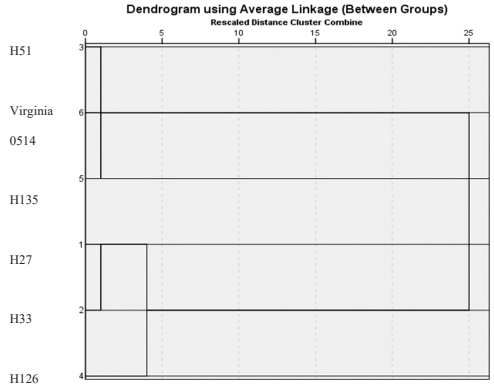

Figure 1. Grouping of hybrid forms of Virginia tobacco according to morphological characteristics, dry leaf yield and days to 50% flowering

For a qualitative description of the factors influencing the obtained clustering, a Principal Component Analysis (PCA) was applied. For this purpose, it was found that the experimental data meet the requirements for PCA: KMO-Test (0.631> 0.5) and Bartlett's test (0.000 <0.05), and the determinant of the correlation matrix is a positive number. The studied traits are transformed into two factors: F_1 and F_2.

The first includes: length and width of the 12th leaf, days to 50% flower and yield and explains 42.8% of the total variance. In this principal component, the width of the 12th leaf, the length of the 12th leaf, the yield of dry tobacco and days to 50% flowering have the strongest influence. The second principal component consists of: stem height and leaf number and explains 32.5%, with a cumulative percentage of variation of 75.3%. Here the number of leaves and the height of the plant have the highest correlation coefficient, which is evidenced by the maximum values of factor weights (Table 3, Figure 2).

Table 3. PCA for hybrid forms of *Virginia tobacco*

Indicator	Component	
	F_1	F_2
Stem height	0.002	0.892
Leaf number	-0.073	0.921
Length of middle leaves	0.874	-0.281
Width of middle leaves	0.912	-0.131
Days to 50% flowering	0.525	0.409
Dry leaf yield	0.832	0.213
% of Variance	42.8	32.5
Cumulative %	42.8	75.3

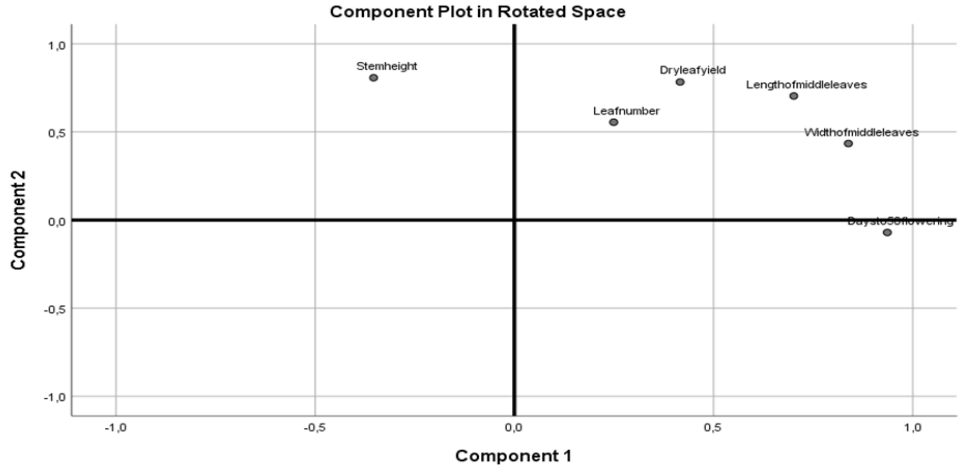

Figure 2. Component plot of the studied indicators on the factorial plane

CONCLUSIONS

The grouping of hybrids in clusters showed that H27, H33 and H126 with a common parent component Virginia 385 have larger leaves and are more productive, which determines not only their differentiation into a separate cluster, but also their strong distance from the other three hybrids in this study. Hybrid H126 has the largest 12th leaf, followed by H27. Hybrid H27 is the most productive. Hybrid H126 has the longest growing season. Hybrid H27 has not only the highest yield, but stands out as relatively stable in this respect, which makes it a recommended variety for cultivation by tobacco growers. The results of the combined application of cluster analysis and principal components analysis show that the most effective would be the selection activity aimed at the populations of the second group based on the traits: length and width of the 12th leaf, dry leaf yield and days to 50% flowering.

REFERENCES

Ahmed, S., Mohammad, F. (2017). Heritability estimates and correlation analysis for production traits in Fcv tobacco, 33, p. 212.

Ali, I., Muhammad, A., Raza, M., Rehman, A., Khaliq, I., Ihtisham, M., Iqbal, A., Anees, M. (2014). Performance of flue cured *Virginia tobacco*, *International Journal of Basic & Applied Sciences, 14*(2), ttp://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.673.3639&rep=rep1&type=pdf

Butorac, J., Beljo, J., Gunjača, J. (2004). Study of inheritance of some agronomic and morphological traits in burley tobacco by graphic analysis of diallel cross. Plant Soil Environ, *50*(4), 162–167.

Campbell, J. (1989). Flue-cured tobacco, suitable for the needs of the industry and satisfying the needs of the consumer. Thematic package, part VI, p. 16–24.

Darvishzadeh, R., Maleki, H. (2012). Analysis of genetic variation for morphological and agronomic traits in Iranian oriental tobacco (*Nicotiana tabaccum* L.) genotypes. *Crop Breeding Journal, 2*(1), 57–61.

Darvishzadeh, Reza, Mirzaei, L., Maleki, H., Laurentin, H., Alavi, S. (2013). Genetic variation in oriental tobacco (*Nicotiana tabacum* L.) by agromorphological traits and simple sequence repeat markers. *Revista Ciência Agronômica, 44*(2), 347–355.

Ivanov, I., Dimova D. (2015). Breeding characteristics of topcross hybrid maize with paternal line 26A by means of cluster analys and analysis of the components. *Agricultural Sciences, VII*(17), 37–42.

Jain, S., Patel, P. (2016). Principal component and cluster analysis in *Sorhum bicolor* (L.) Moench. *Forage Res., 42*(2), 90–95.

Maleki, H., Karimzadeh, G., Darvishzadeh, R., Garrafi, A. (2011). Correlation and sequential path analysis of some agronomic traits in tobacco (*Nicotiana tabaccum* L.) to improve dry leaf yield. *Australian Journal of Crop Science*, 1644–1648.

Meiling Xu, Bingguang Xiao, Fangchan Jiao, Xigfu Wu. (2013). Characterization of 124 Flue-cured tobacco germplasm resources. *Agricultural Science & Tecnology, 14*(2), 286–294.

Mitreski, M., Aleksoski, J., Korubin-Aleksoska, A. (2018). Morphological traits and variability in some flue-cured genotypes. Tutun, *68*. 19–25.

Ning Shang-hui, Luo Xin-bin, Xu Jian-qiang. (2009). Cluster analysis on agronomic characters of 15 tobacco germplasm resources. Crop Research, en.cnki.com.cn

Porkabiri Z, Sabaghnia, N., Ranjbar, R., Maleki, H. (2019). *Scientia agriculturae bohemica, 50*(1), 1–7.

Rachovska, G., Dimova, D., Bozhinov. B. (2002). Application on clustering and principal component analysis for evaluation of breeding materials from winter common wheat. Anniversary scientific session, Sadovo. *Scientific report. Volume III*. 68–72.

Risteski, I., Kočoska. K. (2014). Results of investigations of some bio-morphological traits of *Virginia tobacco* varieties and lines in the producig region of Prillep in 2010 and 2011. Тутун/Tobacco, 64(1), 28–36.

Rosseeva, L., Belan, I., Lozhnikova, L., Blokhina, N., Valueva, L. (2012). Increasing the efficiency of selection by productivity elements in hybrid populations using cluster analysis. *Bulletin of the Altai State Agrarian University, 7*(93), 5–8.

Samizadeh Lahiji, H., Mohsenzadeh Golfazani, M., Edrisi Maryan, K., Shoaeid Deylami, M., Aalami, A. (2013). Assessing the genetic diversity of 89 Flue-Cured Tobacco varieties using morphological traits and inter-simple sequence repeat markers. Crop Breeding Journal, *3*(2), 79–85.

Xiao B.G., J. Zhu, X.P. Lu, Y.F. Bai, Y.P. Li. (2007). Analysis on genetic contribution of agronomic traits to total sugar in flue-cured tobacco (*Nicotiana tabacum* L.). *Field Crops Research, 102*. 98–103.