



**ИЗСЛЕДВАНЕ НА ГЕНЕТИЧНАТА ОТДАЛЕЧЕНОСТ НА ГЕНОТИПИ  
ТЮТЮН ВИРЖИНИЯ СПОРЕД НЯКОИ БИОМЕТРИЧНИ ПОКАЗАТЕЛИ  
STUDY OF GENETIC DISTANCE OF VIRGINIA TOBACCO GENOTYPES  
ACCORDING TO SOME BIOMETRIC INDICATORS**

**Нели Керанова<sup>1\*</sup>, Марина Друмева-Йончева<sup>2</sup>, Йонко Йончев<sup>2</sup>  
Neli Keranova<sup>1\*</sup>, Marina Drumeva-Yoncheva<sup>2</sup>, Yonko Yonchev<sup>2</sup>**

<sup>1</sup>Аграрен университет – Пловдив

<sup>2</sup>Институт по тютюна и тютюневите изделия – Марково

<sup>1</sup>Agricultural University – Plovdiv

<sup>2</sup>Tobacco and Tobacco products Institute – Markovo

**\*E-mail: nelikeranova@abv.bg**

**Abstract**

The subject of research in the present work is four hybrid combinations of Virginia tobacco. The three-year experiment conducted in the experimental base of the Institute for Tobacco and Tobacco Products in the village of Markovo took into account the change in the following biometric indicators: plant height, number of leaves, length and width of the twelfth leaf. The genetic distance of the hybrids according to the above-mentioned signs has been studied. A combination of mathematical and statistical approaches was applied. The genotypes are grouped in clusters by hierarchical cluster analysis. Factor analysis determines the extent of impact of each indicator on the change in overall dispersion. It was found that for each year of research the hybrids are grouped into three clusters, where in 2013 clustering is influenced by the number of leaves, in 2014 – by the size of the twelfth leaf and in 2015 – by the length of the twelfth leaf. For 2013, the most distant are the X33 and X27, due to the number of leaves and the plant height. For 2014 and 2015, the largest genetic differences were found between X51 and X27 due to the size of the twelfth leaf.

**Keywords:** Virginia tobacco, biometric indicators, cluster analysis.

**ВЪВЕДЕНИЕ**

Тютюнопроизводството е важен отрасъл за икономиката на всяка държава. Генетичното разнообразие от хибридни линии тютюн представлява предизвикателство пред селекционерите в тази област от гледна точка на подобряване на техните качествени и количествени характеристики.

Клъстерният анализ е класически метод за групиране на *n* на брой обекти от дадено множество в *k* на брой групи, наречени *кълъстери*, според

степен на сходство по  $p$  на брой признаци. Чрез него се получава информация за генетичните сходства и различия между изследваните генотипи.

Съществуват редица научни разработки, свързани с изследвания върху различни характеристики на сортове тютюн чрез клъстерен анализ (Vanderauwera et al., 2005; Zago et al., 2006; Chen Yi-qiang et al., 2007).

Оценката на генетичното разнообразие на култивираните сортове тютюн има съществено значение за дългосрочното подобряване на качествата му. Davalieva et al. (2010) изследват десет култивирани линии тютюн, разпространени на територията на Република Македония. Установяват, че те се класифицират в три отделни групи. Проучванията са извършени по 30 показателя, но се доказва, че само двадесет и четири от тях са достатъчни за оценката на тези линии.

Nejad and Ahmadikhah (2010) анализират сортове тютюн според DNA и PARS маркери за митохондриален геном на тютюна. Те доказват, че откритото опрашване не засяга цитоплазмата и че PARS-маркерите са ефективно средство за изследване на полиморфизъм и генетично разнообразие при тютюна.

Тютюневите листа, получени от различни географски райони в Китай, са анализирани чрез газова хроматография, съчетана с многовариантни анализи на данни (Zhang et al., 2013). Йерархичният клъстерен анализ и анализът на основните компоненти показват, че върху качествата на тютюна оказват влияние както географските фактори, така и климатичните условия като температура и валежи. В резултат се установява, че метаболитното профилиране може да се използва за разпознаване на географския произход на тютюневи листа.

Целта на настоящото изследване е да се групират четири хибридни форми тютюн тип Виржиния според различни биометрични показатели и по този начин да се оцени тяхната генетична отдалеченост.

## **МАТЕРИАЛИ И МЕТОДИ**

Обект на изследване в настоящата работа са четири хибридни форми: *Хибрид 27*, *Хибрид 51*, *Хибрид 33* и стандартът *B0514*. Анализират се експериментални данни, свързани със следните биометрични показатели: височина на стъблото, брой листа, дължина и ширина на дванадесети лист. Тъй като предварително не е известен броят на клъстерите, е използван йерархичен клъстерен анализ. Съществува голямо разнообразие от агломеративни методи за провеждане на този анализ, както и различни мерки за разстояние между отделните клъстери. За определяне на подходящ агломеративен метод за клъстеризация, даващ оптимален резултат, беше приложен методът на едномерното разпределение. Предварително са получени резултати от пет метода: вътрешногрупово свързване, междугрупово свързване, метод на най-близкия и на най-далечния съсед и метод на Ward. След построяване на крос-таблицы и изчисляване на коефициента на контингенция се оказва, че същият е максимален при метода

на междугруповото свързване. Тук разстоянието между два клъстера А и В се дефинира като средната стойност на  $n_A \cdot n_B$  на брой разстояния между  $n_A$  точки от А и  $n_B$  точки от В чрез формулата

$$D(A, B) = \frac{1}{n_A n_B} \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} d(x_i, x_j)$$

където сумата се изменя по всички  $x_i$  от А и всички  $x_j$  от В. Чрез

$$d(x_i, x_j) = \sum_{m=1}^p (x_{im} - x_{jm})^2, \quad i, j = \overline{1, n}$$

означаваме квадратичното евклидово разстояние между два вектора  $x_i(x_{i1}, x_{i2}, \dots, x_{ip})$  и  $x_j(x_{j1}, x_{j2}, \dots, x_{jp})$ . Данните предварително са стандартизирани.

Известно е, че при клъстерния анализ не се извършват тестове за статистическа значимост на резултатите. Поради това е препоръчително да се комбинира с друг метод, при който се правят подобни оценки. В настоящата работа е приложен факторен анализ по метода на главните компоненти (PCA), като предварително са определени корелационните коефициенти между изследваните променливи. Чрез този метод броят на проучваните признаци се редуцира, като корелиращите помежду си се обединяват в общ фактор, а некорелиращите – в отделни. Броят на факторите се определя от броя на по-големите от единица собствени значения на корелационната матрица (Кайзер-трансформация). Прилагайки този метод, се дава качествено описание на клъстерите и се изясняват показателите, влияещи най-силно при разпределението на хибридите в групи.

За математическата обработка е използван програмният продукт IBM Statistics SPSS 24 (Cronk, 2016; Ganeva, 2016).

## РЕЗУЛТАТИ И ДИСКУСИЯ

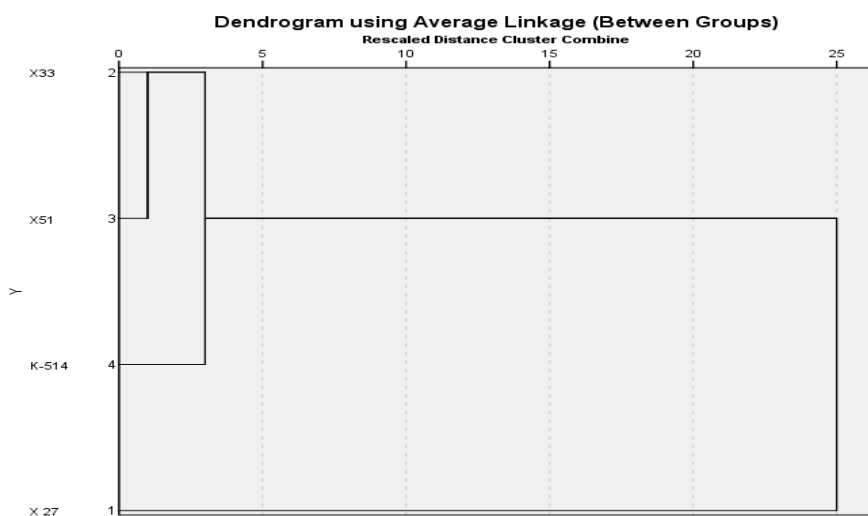
В таблица 1 са дадени последователността на получаването на отделните клъстери при всяка стъпка и междугруповите разстояния при групиране според степента на сходство по биометрични показатели за 2013 г. Четвъртата колона на таблицата (Coefficients) показва разстоянието между клъстерите, които са обединени в тази стъпка. Именно тази таблица дава информация за оптималния брой клъстери.

За 2013 г. Хибрид 33 и Хибрид 51, които са с максимална височина на стъблото, надвишаваща тази на контролния хибрид, формират първия клъстер. Генотипът В0514 е отделен в самостоятелен клъстер поради различие в останалите три показателя (има максимален брой листа и дължина на 12 лист и минимална ширина на 12 лист). Клъстерът, състоящ се от Хибрид 27, се присъединява към останалите на най-голямо евклидово разстояние, което се обяснява със значителните разлики с останалите

хибриди по отношение на височината на растението и броя на листата. Същевременно това е растението с най-голяма ширина на 12-я лист.

**Таблица 1.** Комбиниране на клъстерите и междугруповите разстояния по биометрични показатели за 2013 г.

Етап №	Комбиниране на клъстерите		Коефициент	Етап на първо формиране на клъстер		Етап на следващо обединяване
	Клъстер 1	Клъстер 2		Cluster 1	Cluster 2	
1	2	3	19,828	0	0	2
2	2	4	46,229	1	0	3
3	1	2	240,269	0	2	0



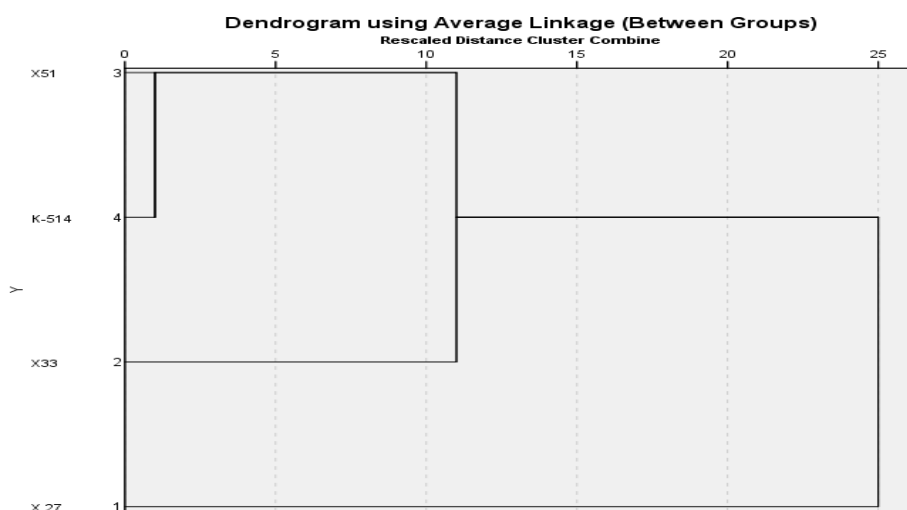
**Фиг. 1.** Дендрограма на клъстеризационната процедура по биометрични показатели за 2013 г.

За всяка от изследваните години са изпълнени условията за прилагане на факторен анализ: КМО-тест, Bartlett's тест за сферичност, случаен характер на експерименталните данни, изследваните показатели са количествени. Факторният анализ установи, че за 2013 г. четирите изследвани показателя се трансформират в два фактора. Първият включва височината и броя на листата и обяснява 49% от общото вариране. Вторият се състои от размерите на 12-я лист и обяснява 36% от общата дисперсия.

Отчитайки резултатите за 2014 г. по биометричните показатели, изследваните хибридни форми се групираха в три клъстера. Най-близки се оказват Хибрид 51 и В0514, които притежават минимални размери на дванадесетия лист. Хибрид 27 е най-отдалечен поради минималната височина на стъблото в сравнение с останалите растения (фиг. 2, табл. 2).

**Таблица 2.** Комбиниране на клъстерите и междугруповите разстояния по биометрични показатели за 2014 г.

Етап №	Комбиниране на клъстерите		Коефициент	Етап на първо формиране на клъстер		Етап на следващо обединяване
	Клъстер 1	Клъстер 2		Клъстер 1	Клъстер 2	
1	3	4	18,278	0	0	2
2	2	3	48,502	0	1	3
3	1	2	92,939	0	2	0



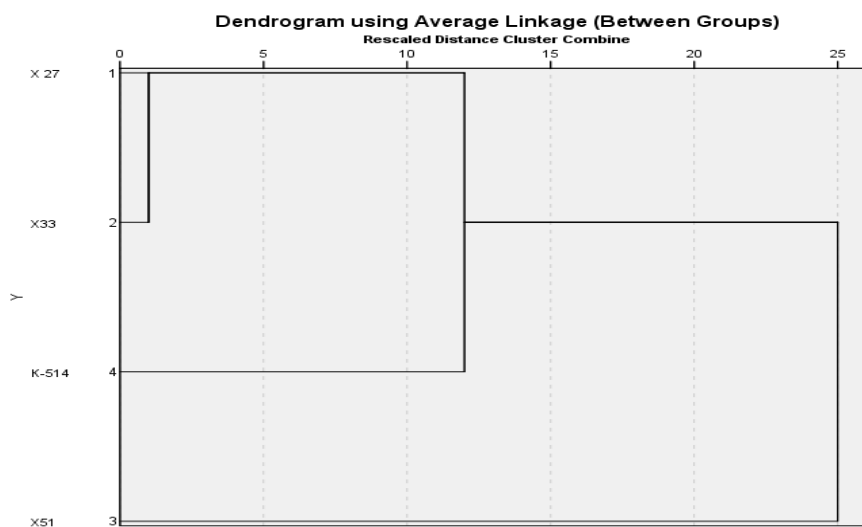
**Фиг. 2.** Дендрограма на клъстеризационната процедура по биометрични показатели за 2014 г.

След прилагане на PCA се установи, че четирите показателя се трансформират отново до два, но разпределението им е различно от това за 2013 г. Първият включва броя и размерите на дванадесетия лист и обяснява 60% от варирането, като доминиращи са размерите на листа. Вторият включва само височината и обяснява 28% от общата дисперсия.

В таблица 3 и на фигура 3 са дадени резултатите от клъстеризацията по биометрични показатели за 2015 г. Формирани са три клъстера. Първият включва Хибрид 27 и Хибрид 33 поради максимален брой на листата и максимална ширина на 12-я лист.

**Таблица 3.** Комбиниране на клъстерите и междугруповите разстояния по биометрични показатели за 2015 г.

Етап №	Комбиниране на клъстерите		Коефициент	Етап на първо формиране на клъстер		Етап на следващо обединяване
	Клъстер 1	Клъстер 2		Клъстер 1	Клъстер 2	
1	1	2	104,536	0	0	2
2	1	4	447,274	1	0	3
3	1	3	869,199	2	0	0



**Фиг. 3.** Дендрограма на клъстеризационната процедура по биометрични показатели за 2015 г.

B0514 е с максимална височина, чувствително надвишаваща тази при останалите хибриди, което обуславя отделянето му в самостоятелен клъстер. Най-отдалечен е клъстерът, състоящ се от Хибрид 51, поради минималните размери на височината на стъблото, дължината на 12-я лист и броя на листата. След прилагане на факторен анализ се установи, че четирите показателя се трансформират до един, който изчерпва 73% от общата дисперсия. Преобладаващо е влиянието на броя на листата и дължината на 12-я лист.

## ИЗВОДИ

Извършеното групиране на хибридните линии тютюн тип Виржиния според някои биометрични показатели дава възможност за по-пълна и детайлна оценка на тяхната генетична отдалеченост.

1. В резултат от приложения клъстерен анализ се установи, че и за трите години на изследване хибридните форми се групират в три клъстера, чиито обекти са със сходни биометрични характеристики. За 2013 г. броят на листата оказва най-силно влияние за клъстеризацията, за 2014 г. – размерите на дванадесетия лист, а за 2015 г. – дължината на дванадесетия лист.

2. Резултатите от проведените анализи би следвало да се вземат предвид при бъдещи хибридизации с цел повишаване на качеството на получените генотипи от гледна точка на изследваните биометрични показатели.

## REFERENCES

- Chen Yi-qiang, Shen Xiao-tian, Liu Guo-shun, Zhao Ming-shan, Hu Huan-xin, Zhao Guo-jiao*, 2007. Comprehensive Evaluation of Flue-cured Tobacco Leaves Base on Cluster Analysis and Fuzzy Mathematics, *Acta Agriculturae Universitatis Jiangxiensis*, 04.
- Cronk, B.*, 2016. How to Use SPSS: A Step-By-Step Guide to Analysis and Interpretation, Routledge.
- Davalieva, K., I. Kostovska, K. Filipovski, O. Spiroski, G. Efremov*, 2010. Genetic variability of Macedonian tobacco varieties determined by microsatellite marker analysis, *Diversity*, 2(4), 439–449.
- Ganeva, Z.*, 2016. Discovering Statistics using IBM SPSS Statistics, Elestra, Sofia, (Bg).
- Nejad, A., A. Ahmadikahah*, 2010. Cytoplasmic variation between different tobacco cultivars revealed by mitochondrial-specific markers, *Annals of biological research*, 1(1), 1–9.
- Vanderauwera, S., P. Zimmermann, S. Rombauts, S. Vandenaabeele, C. Langebartels, W. Gruissem, D. Inze, F. Van Breusege*, 2005. Genome-Wide Analysis of Hydrogen Peroxide-Regulated Gene Expression in Arabidopsis Reveals a High Light-Induced Transcriptional Cluster Involved in Anthocyanin Biosynthesis, *Plant Physiology*, 139, 806–821.
- Zago, E., S. Morsa, J. Dat, P. Alard, A. Ferrarini, D. Inze, M. Delledonne, F. Van Breusegem*, 2006. Nitric Oxide- and Hydrogen Peroxide-Responsive Gene Regulation during Cell Death Induction in Tobacco, *Plant Physiology*, 141, 404–411.
- Zhang, L., X. Wang, J. Guo, Q. Xia, Q. Zhao, H. Zhou, F. Xie*, 2013. Metabolic profiling of Chinese tobacco leaf of different geographical origins by GC-MS, *Journal of agricultural and food chemistry*, 61 (11), 2597–2605.