Research Note Electronic Journal of Plant Breeding, Vol.17(3):

Statistical dissection of quality traits for identifying genetic diversity in tobacco (*Nicotiana tabacum* L.)

P. Pulli Bai, S. Rama Devi*, K. Sathish Babu, K. Prabhakar, N.C. Venkateswarlu and M. Johonson

https://doi.org/10.37992/2026.1703.035

Abstract

In the study, 50 *bidi* and 30 *natu* tobacco genotypes were evaluated for quality characters along with cured leaf yield and analysed for principal component analysis and hierarchical clustering. The principal components with eigen values more than one *i.e.* PC_1 and PC_2 described 45 % and 26 % variation, respectively. Based on the biplot analysis, cured leaf yield showed positive association with nicotine content and reducing sugars, while negative association with chlorides. Clustering was done as per the Wards method, which clustered the eighty genotypes into twelve clusters, in which cluster III contained the highest number of genotypes i.*e.* eleven, whereas, cluster VIII showed the solitary genotype. Cluster VIII recorded high cluster mean for reducing sugars and cured leaf yield whereas, low for chlorides and moderate for nicotine content. Overall, the results indicated that the genotypes *viz.,* ABD 128, ABD 54, ABD 131, NyBD 82, NyBD 56 and NyBD 79 can be utilized in yield improvement programmes to develop best tobacco varieties with good quality traits for market preference.

Keywords: Tobacco, Bidi, Natu, genetic diversity, Hierarchical cluster analysis, Principal components, Quality characters

Tobacco (*Nicotiana tabacum* L.) is a cash crop which is cultivated for its medicinal and commercial value worldwide. India is the second-largest producer in the world, occupied 0.45 M ha of area with 750 M kg production (Subbiah, 2022), which is mainly concentrated in Gujarat and Andhra Pradesh. Different types of tobacco are cultivated across India, including bidi tobacco in Gujarat, Andhra Pradesh, Uttar Pradesh, Bihar, and Madhya Pradesh; natu and burley types in Andhra Pradesh; Flue-Cured Virginia (FCV) tobacco, grown mainly in Andhra Pradesh, Karnataka, Uttar Pradesh and Bihar; chewing tobacco in Maharashtra, Tamil Nadu, and West Bengal; and cigar wrapper, filler, hookah, and cheroot tobaccos in certain parts of Tamil Nadu, catering to both domestic and export markets. Tobacco plays a significant economic role by providing livelihoods to millions of people, particularly in rural areas, even though there are potential opportunities to shift towards alternative crops. Genetically, cultivated tobacco is an allotetraploid species (2n = 4x = 48), primarily valued for its quality traits such as nicotine content, reducing sugars, and chlorides, which are key determinants for marketing and commercial purposes. Hence, understanding the genetics of these traits, the diversity present within germplasm, and their association with cured leaf yield is crucial for developing high-yielding cultivars with superior quality attributes.

Principal Component Analysis (PCA) serves as an effective multivariate technique that analyses complex data by transforming correlated variables into a set of uncorrelated principal components, which collectively explain most of the variation within the dataset (Mishra *et al.,* 2013, Abdi and Williams, 2010). Similarly, cluster analysis groups the accessions based on their genetic relationships and distances (Mellingers, 1972). In this way the combined use of PCA and hierarchical cluster analysis provides an effective approach to determine genetic divergence among accessions, which is highly beneficial for breeding programs (Chaudhary *et al*., 2015). Understanding the genetic diversity enables breeders to identify suitable parental lines for hybridization and the subsequent development of improved cultivars (Kalivas *et al*., 2016).

The present study involved 50 bidi and 30 natu tobacco genotypes, evaluated at the Regional Agricultural Research Station (RARS), Nandyal, under Acharya N. G. Ranga Agricultural University, Andhra Pradesh, India, during rabi, 2023–24 for three important quality traits along with cured leaf yield. The experiment was

Regional Agricultural research Station, Nandyal, Acharya N. G. Ranga Agricultural University, Andhra Pradesh.

Corresponding author: S. Rama Devi, Regional Agricultural research Station, Nandyal, Acharya N. G. Ranga Agricultural University, Andhra Pradesh. *E-Mail: s.ramadevi@angrau.ac.in

How to cite this artilce: Pulli Bai, P., Rama Devi, S., Sathish Babu, K., Prabhakar, K., Venkateswarlu, N.C. and Johonson, M. 2026. Statistical dissection of quality traits for identifying genetic diversity in tobacco (*Nicotiana tabacum* L.). *Electronic Journal of Plant Breeding.* ***17(3):***
doi:10.37992/2026.1703.035

Received: 26.02.2026 **Accepted**: 05.08.2026 **Online**: 06.08.2026

conducted in two replications with spacing of 75 × 75 cm for bidi and 70 × 70 cm for natu tobacco, with each genotype planted in two rows of 7.5 m length. All standard agronomic practices were followed to raise a healthy crop, and data were recorded after harvest. The yield was recorded from five random plants and averaged. The third leaf was collected from the top of the plant for the analysis of quality traits viz., nicotine content, reducing sugars and chlorides.

According to Peeters and Martinelli (1989), both Principal Component Analysis (PCA) (Hotelling, 1936) and cluster analysis are effective tools for identifying the traits contributing most to variation and genetic diversity, respectively. Hence, the recorded data were subjected to statistical analysis using R software, involving PCA and hierarchical cluster analysis. The PCA results were used to develop a biplot based on principal components, while a constellation plot was constructed using Ward's method in hierarchical clustering.

The accessions NyBD 79 and DWFC recorded the highest (8.24 %) and lowest (0.60 %) nicotine content, respectively and the mean was 4.38 % **(Table 1)**. The reducing sugar content mean value was 2.30 % and recorded high in ABD 128 (7.22 %) and low in II 1870 (1.42 %). Chlorides % ranged from 1.85 (II 1870) to 4.76 % (Chilakaluripeta) with a mean of 3.25 % and cured leaf yield was ranged from 125 (5B-154) to 166 g/ plant (NyBD 56) with a mean of 147.80 g/ plant. For the market value medium nicotine content, high cured leaf yield along with reducing sugars and low chlorides are desirable (Prasad *et al.,* 2022). Hence, ABD 128 recorded the desirable quality traits along with high cured leaf yield indicates its potential in using for hybridization programmes for developing best tobacco varieties with good quality.

The results of PCA discloses the significance of first three principal components which exhibited their eigen values near one and explain 96 % of total variation out of four principal components (**Table 2**). Scree plot (**Fig.1**) was also plotted to access the components to explain the variability percentage and to represent the values in descending order. The higher the explained percent of variance, higher information will be contained in those components (Maruthi Prasad *et al.,* 2023). Results of principal component analysis of Mostafa Mehdizadeh *et al.* (2023) stated that the 14 first components could explain 51.49 % of the total variance. Edrisi Maryan *et al.* (2012) analysed 40 tobacco cultivars for PCA and reported first coordinates explain 41.90 % of total variation and clustered the genotypes into five groups based on morphological traits using UPGMA method. The PC1 and PC3 displayed positive weight to chlorides (0.163) and Nicotine (0.117) respectively whereas, PC2 recorded positive loading to all traits except reducing sugars and PC4 shown positive weight for all except cure leaf yield. (**Table 3**). The genotypes in components with low loading could be utilized in hybridization programmes with genotypes in high loading components to achieve genetic gain at its maximum capacity (Maruthi Prasad *et al.,* 2023).

The biplot demonstrated association of 80 accessions for the traits studied (**Fig. 2**). The four characters were

Table 1. Descriptive statistics of 80 tobacco genotypes

S. No.	Characters	Mean	Maximum	Minimum	SD	SE
1	Nicotine content (%)	4.38	8.24	0.60	2.28	0.25
2	Reducing Sugars (%)	2.30	7.22	1.42	0.75	0.08
3	Chlorides (%)	3.25	4.76	1.85	0.55	0.06
4	Cured leaf yield (g/plant)	147.80	166.00	125.00	13.48	1.50

Table 2. Principal component analysis in 80 tobacco genotypes

Components	PC1	PC2	PC3	PC4
Eigen values	1.82	1.05	0.96	1.15
Variance percent	45.70	26.50	24.10	3.80
Cumulative variance percent	45.65	72.11	96.21	100.00

Table 3. Factor loading of characters for principal factor in 80 tobacco genotypes

Principal Components	PC1	PC2	PC3	PC4
Nicotine content	-0.675	0.283	0.117	0.670
Reducing Sugars	-0.172	-0.824	-0.473	0.257
Chlorides	0.163	0.487	-0.850	0.107
Cured leaf yield	-0.697	0.043	-0.195	-0.687

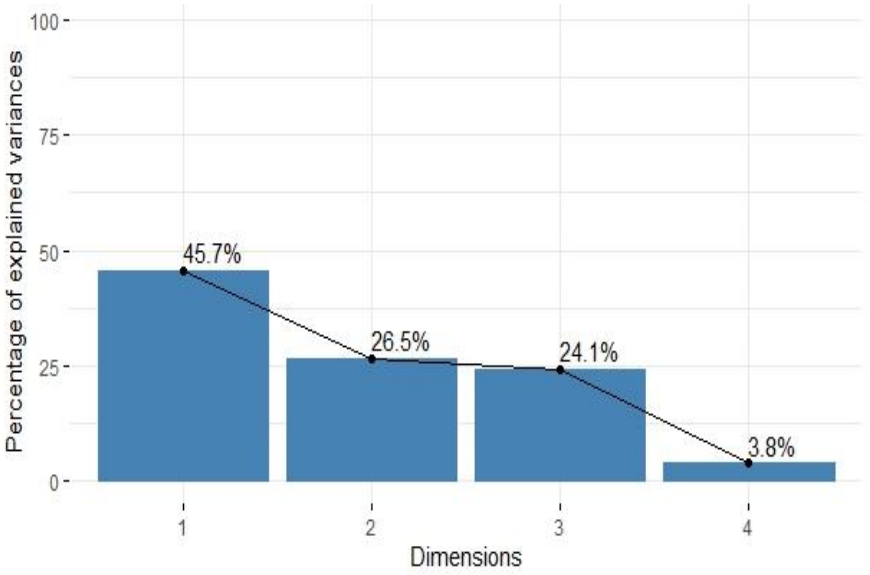

Fig. 1. Scree plot illustrating the percentage of explained variation by each principal component

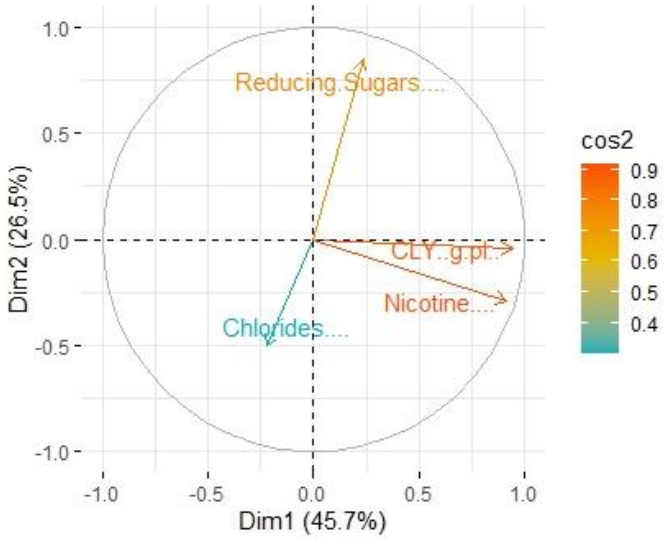

Fig. 2. PCA plot for various traits under study

grouped into three categories. Nicotine and cured leaf yield grouped into same cluster. Other two traits were placed in separate clusters. Cured leaf yield is positively associated with reducing sugars and nicotine but negatively correlated with chlorides (Tiecher *et al.,* 2023; Pace *et al.,* 2020). The accessions with highest score in PC1 should be selected for developing best high yielding types. The genotypes *viz*., ABD 128, NyBD 82, NyBD 75, NyBD 56, ABD 131, NyBD 92, Nandyal Pogaku 1, Chilakaluripeta, Bommidala, Talamariaaku and NyNT 65 were located at the extreme ends of all quadrants of the biplot, indicating that these genotypes are divergent and useful to exploit heterosis by utilizing in the breeding programmes.

Hierarchical cluster analysis classified the 80 genotypes into twelve clusters. Cluster III was found to be the largest one with eleven accessions followed by cluster X with ten accessions, whereas cluster VIII was the smallest one which has only one accession *i.e.* ABD 128. (**Table 4**). Dadras *et al.* (2014) grouped 50 tobacco genotypes into eight clusters using cluster analysis based on complete linkage method using Jaccard's genetic distance. Mostafa Mehdizadeh *et al.* (2023) classified 48 tobacco genotypes into five groups by using cluster analysis by the UPGMA method with 5, 12, 10, 6 and 15 genotypes, respectively. The cluster means estimated for four traits indicated the presence of variation among clusters (**Table 5**). The highest and lowest cluster mean values

Table 4. Clustering of eighty tobacco genotypes

Cluster	Number of genotypes	Name of the genotype
I	9	NyBD 79, NyBD 93, GT 4, NyBD 75, NyBD 85, NyBD 86, NyBD 67, NyBD 77, NyBD 71
II	9	NyBD 59, NyBD 78, NyBD 92, NyBD 88, NyBD 68, NyBD 61, Nandyal pogaku 1, GT 7, ABD 129
III	11	ABD 54, NyBD 63, NyBD 56, NyBD 72, NyBD 90, ABD 131, NyBD 60, NyBD 62, NyBD 73, ABD 78, ABD 119
IV	5	ABD 68, GT 5, A119, NyBD 70, NyBD 91
V	8	NyBD 94, NyBD 76, NyBD 66, NyBD 81, NyBD 80, ABD 146, ABD 118, ABD 152
VI	3	NyBD 87, NyBD 89, NyBD 82
VII	4	NyBD 69, NyBD 84, NyBD 74, NyBD 83
VIII	1	ABD 128
IX	7	II 1870, Natu special, Addanki, Natunoonepalli, 5B-154, Solunki, Talamariaaku
X	10	Pencil leaf, II 1872, WAF, Peddavittanam, Oriental tobacco, Dharanikota, Kavali, NyNT64, Kommipaddu vithanam, Jakirampur
XI	8	Rathna(Nandigama), Mutant, Yelamanchali, Sathyenapalli, Bommidala, Singarayakonda, Natuparachure, Chilakaluripeta
XII	5	NyNT65, Nellore, Pryuvittanam, Inkollu, DWFC

Table 5. Cluster mean values of characters under study

Cluster Number	Nicotine content (%)	Reducing sugars (%)	Chlorides (%)	Cured Leaf yield (g/plant)
I	6.97	2.00	2.62	152.22
II	6.06	2.05	3.94	158.00
III	6.45	2.03	3.15	162.18
IV	6.78	2.13	3.39	150.00
V	5.26	2.58	3.07	159.38
VI	4.69	3.73	3.01	159.33
VII	3.86	3.12	3.71	158.25
VIII	3.73	7.22	2.13	160.00
IX	2.17	1.84	2.45	130.14
X	2.02	1.93	3.30	133.50
XI	1.53	2.23	4.06	130.63
XII	1.07	2.59	3.19	132.20

observed for cured leaf yield was (162.18 g/plant) and nicotine content (1.07 %), respectively. Cluster VIII recorded high cluster mean for reducing sugars and low cluster mean for chlorides component, which indicates that it can be useful in hybridization programme to develop best quality tobacco varieties as per the market value through research programmes. Cluster III showed high cluster mean for cure leaf yield, utilizing these accessions will be helpful to develop high yielding varieties. Cluster IX recorded low cluster mean for reducing sugars and yield, which indicates the accessions have poor performance in yield and also quality.

In the constellation plot (**Fig. 3**), most of the NyBD lines were allocated near the origin which indicated that they were less divergence regarding quality traits. It also showed two clear outliers *i.e.* ABD 128 and NyBD 82, were found as strong deviators from the remaining accessions by dropping far outside to the circle, which suggests that, hybridization using these two genotypes will leads to increased heterosis.

Evaluation of 80 tobacco genotypes for major quality characteristics along with yield and was subjected to statistical analysis using R software for multivariate analysis, PCA and clustering methods. The results revealed that, from principal component analysis, first three components contribute 96 % of total variation in tobacco accessions. The PC1 contributed the highest variation (45.7) followed by PC2 (26.5) and PC3 (24.1). The biplot analysis provides the results as there is a positive association of cure leaf yield with nicotine content and reducing sugars and negative association with chlorides. The cluster analysis exhibited diversity

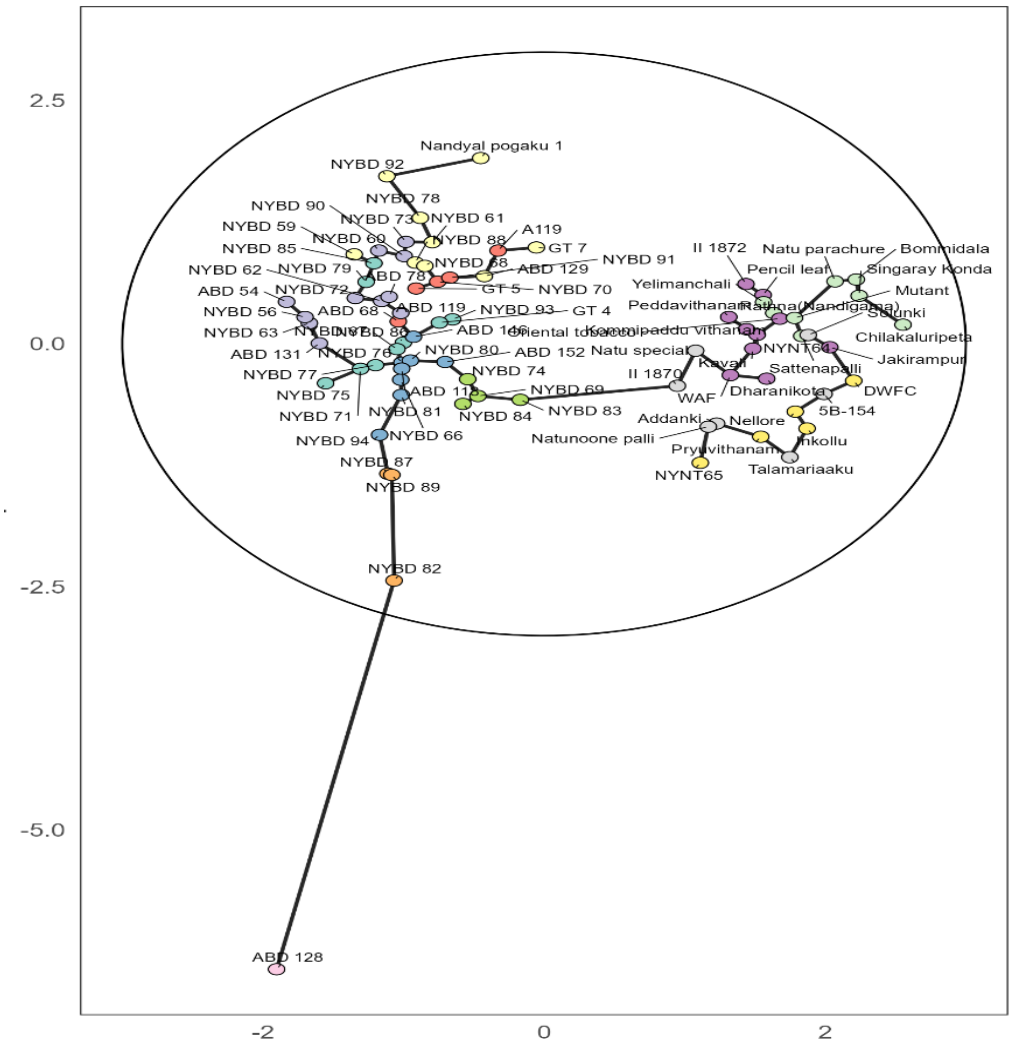

Fig. 3. Constellation plot of eighty tobacco accessions

among the genotypes and classified the genotypes into twelve clusters. Based on constellation plot, clusters, cluster means, it is identified that, the genotypes, *viz.,* ABD 128, NyBD 82, NyBD 89, NyBD 87, NyBD 92, ABD 54, ABD 131, NyBD 56 and NyBD 79 were found to be the promising accessions with wide variation and potential use in breeding programmes in tobacco. Hence, the study of clusters analysis and PCA will be helpful for the selection of genotypes for hybridization.

ACKNOWLEDGEMENT

The authors are very thankful to ICAR-NIRCA (CTRI), Rajahmundry and Acharya N. G. Ranga Agricultural University (ANGRAU), Guntur, Andhra Pradesh for providing the facilities required for the present study.

REFERENCES

Abdi, H. and Williams, L.J. 2010. Principal component analysis. Wiley interdisciplinary reviews. *Computational Statistics*, **2**(4):433-459.[Cross Ref]

Chaudhary, S.P., Sagar, B.K., Hooda. and Arya, R.K. 2015. Multivariate analysis of pearl millet data to delineate genetic variation. *Forage Research*, **40**:201-208.

Dadras, AR, Sabouri, H., Nejad, G.M., Sabouri, A. and Shoai-Deylami, M. 2014. Association analysis, genetic diversity and structure analysis of tobacco based on AFLP markers. *Molecular Biology Reports*, **41**(5):3317-29. [Cross Ref]

Edrisi Maryan, K., Samizadeh Lahiji, H. and Shoaei Deylami, M. 2012. Assessing the genetic diversity of tobacco (*Nicotiana tabacum* L.) varieties. *Crop Breeding Journal,* **2**(2): 125-132.

Hotelling, H. 1936. Relations between two sets of varieties. *Biometrica*, **28**(3-4): 321-377.[Cross Ref]

Kalivas, A., Ganopoulos, I., Bosmali, I., Tsaliki, E., Osathanunkul, M., Xanthopoulou, A., Moysiadis, T., Avramidou, E., Grigoriadis, I., Zambounis, A.,

Tsaftaris, A., Nianiou-Obeidat, I. and Madesis, P. 2016. Genetic diversity and structure of tobacco in greece on the basis of morphological and microsatellite markers. *Crop Science*, **56**: 2652-2662. [Cross Ref]

Maruthi Prasad, B. P., Patil, B. R., Geeta, D. and Matiwade. P. S. 2023. Principal Component Analysis (PCA) and hierarchial clustering in tobacco (*Nicotiana tabacum* L.) for yield and yield attributing traits. *Electronic Journal of Plant Breeding*, **14**(2):737-746. [Cross Ref]

Mellingers, J.S. 1972. Measures of genetic similarity and genetic distance studies in genetics. *VII Univ Tex Publ*, **27**(13):145-153.

Mishra, V.K., Singh, T. and Baranwal, D.K. 2013. Genetic diversity based on cluster and principal component analyses for yield and its contributing characters in wheat (*Triticum aestivum* L.). *Madras Agricultural Journal*, **100**(4-6): 320–323. [Cross Ref]

Mostafa Mehdizadeh, Mohammad Mohsenzadeh Golfazani and Habibollah Samizadeh Lahiji. 2023. Assessment of genetic diversity in different types of tobacco using molecular markers. *Journal of Plant Physiology and Breeding*, **13**(2): 81-97.

Pace, C., Vann, M., Fisher, L. and Hardy, D. 2020. Implications of chloride application rate and nitrogen fertilizer source to flue-cured tobacco. *Agronomy Journal*, **112**: 2916–2927. [Cross Ref]

Peeters, J.P. and Martinelli, J.A. 1989. Hierarchical cluster analysis as a tool to manage variation in germplasm collections. *Theoretical and Applied Genetics,* **78**:42-48. [Cross Ref]

Prasad, L.K, Chandrasekhara Rao, C., Damodar Reddy, D., Padmaja, K. and Johnson, N. 2022. A New Extractant for Concurrent Estimation of Nicotine, Reducing Sugars and Chlorides in Tobacco Cured Leaf, Communications in Soil Science and Plant Analysis, DOI: 10.1080/00103624.2022.2063315. [Cross Ref]

Subbiah, R. 2022. Tobacco Industry in India. *International Journal for Multidisciplinary Research.* **4** (5): 1-13. [Cross Ref]

Tiecher, T., Pace, C. R., Gatiboni, L., Vann, M., Hardy, D. and Fisher, L. 2023. *Flue-cured tobacco and Cl rates:* Implications on yield, quality, and nutrient concentration. *Agronomy Journal*, **115**: 896–908. [Cross Ref]