Journal of Agricultural Sciences (JAS) 32(4):1188-1207
DOI: 10.15832/ankutbd.1904275

J Agr Sci-Tarim Bili
e-ISSN: 2148-9297
jas.ankara.edu.tr

Machine Learning-Based Assessment of Genetic Diversity and Population Structure in Hollyhock (*Alcea rosea*): A Comprehensive Morphological, Biochemical, and Molecular Approach

Akife Dalda Sekerci[a*]

[a]*Erciyes University, Faculty of Agriculture, Department of Horticulture, Kayseri, TÜRKIYE*

ARTICLE INFO
Research Article
Corresponding Author: Akife Dalda Sekerci, E-mail: akidal_@hotmail.com
Received: 06 March 2026 / Revised: 05 June 2026 / Accepted: 23 June 2026 / Online: 12 October 2026

Cite this article
Sekerci A D (2026). Machine Learning-Based Assessment of Genetic Diversity and Population Structure in Hollyhock (*Alcea rosea*): A Comprehensive Morphological, Biochemical, and Molecular Approach. *Journal of Agricultural Sciences (Tarim Bilimleri Dergisi),* 32(4):1188-1207. DOI: 10.15832/ankutbd.1904275

ABSTRACT
Hollyhock (*Alcea rosea*) is an important ornamental and medicinal plant. This study aimed to evaluate the genetic diversity and population structure of 28 endemic genotypes collected from various regions of Türkiye using an integrated approach combining morphological, biochemical, and molecular markers. Significant morphological variation was observed, and image analysis techniques were employed to classify flower and leaf colors. Biochemical assays revealed substantial diversity in bioactive compound levels, with total phenolic content ranging from 0.18 to 4.65 mg GAE/kg. Antioxidant activities also varied: FRAP (5.8–22.7%), CUPRAC (26.06–78.26%), and DPPH (13.6–70%). Molecular diversity was assessed using iPBS retrotransposon-based markers, with similarity indices ranging from 0.35 to 0.74, aligning with morphological and biochemical patterns. Principal Component Analysis (PCA) indicated dispersed distribution among genotypes in both two- and three-dimensional plots, consistent with phenotypic variation. STRUCTURE analysis identified four distinct subpopulations based on genetic structure. Furthermore, machine learning (ML) models-Multilayer Perceptron (MLP), Support Vector Machines (SVM), and Random Forest (RF)-were employed to predict key biochemical and morphological traits. Model performance varied, with different R^2 values obtained for antioxidant-related traits. This is the first comprehensive study of *Alcea rosea* integrating molecular data and ML tools to characterize genetic diversity and trait prediction.

Keywords: Antioxidant Activity, iPBS, Retrotransposon

1. Introduction

Understanding the genetic diversity and population structure of plant species is essential for their conservation, sustainable use, and potential improvement. The Hollyhock (*Alcea rosea*) is a perennial plant belonging to the family Malvaceae, native to Europe and Northern Asia (Wang et al. 2023). The family Malvaceae includes about 80 genera and 1,000 species worldwide, with South America being particularly rich in species diversity. These plants are distributed worldwide, except for cold regions. The genus *Alcea* includes about 70 species, mainly distributed in Europe (except northern regions), North America, North Africa, the Caucasus, southern Russia, and regions extending from Anatolia to Afghanistan (Zohary 1963; Hutchinson 1973; Heywood 1978; Cullen 1967). The highest diversity of *Alcea* species is seen in Iran, followed by Russia and Türkiye (Riedl 1976). *Alcea* species, known for their vividly colored ornamental flowers, are generally biennial or perennial herbaceous plants and are widely distributed in Asia and Europe (Wang et al. 2023).

Hollyhock is widely cultivated worldwide as both a medicinal and ornamental plant. The species exhibits remarkable morphological and physiological diversity, with differences in essential oil composition and antioxidant content (Elmastas 2004). This dual use highlights the importance of the species' broad morphological characteristics. Historically, hollyhock has been used in traditional medicine to treat sore throats, coughs, stomach ulcers and wounds (Bahmani et al.2014; Kianitalaei et al. 2019). Due to its high content of antioxidants and phenolic compounds, hollyhock is widely used as an additive in various medical and pharmaceutical applications (Tabarsa et al.2017; Tahmouzi & Nejat 2020). The composition of the plant is quite rich and contains 11% pectin, 25–35% starch, 10% mono- and disaccharides, 5% mucilage, flavonoids, phenols, tannins, coumarin, amino acids and asparagine (Bradley1992; Blumenthal 2000). As an edible medicinal plant, the roots and leaves of hollyhock are widely used in teas (Ulbricht et al. 2003) and salads (Elmastas 2004), increasing the versatility and uses of the plant.

This wide variation is critical for breeding programs and makes it advantageous to use morphological and biochemical markers together with molecular and genetic markers. Characterization of genetic material allows the study of genetic diversity that can be used in various breeding activities (Gramazio et al.2018; Guliyev et al. 2018). Recent advances in molecular marker

technologies have improved breeding programs by allowing detailed characterization of genetic material and discovery of new genetic variations. Various molecular markers differ in their effectiveness and applications (Barut et al. 2019). The limitations associated with LTR and non-LTR retrotransposons have led to the development of a new marker system called inter-primer binding site (iPBS). iPBS is a versatile marker system suitable for molecular characterization of both plant and animal species. This PCR-based marker system (iPBS-retrotransposon) uses tRNA as the primer binding site for reverse transcriptase (Kalendar et al.2020). In addition, iPBS primers are important molecular markers frequently used in the study of ornamental and medicinal-aromatic plants (Dalda-Sekerci 2023). This marker system has been successfully used to study the genetic diversity and population structure of various agricultural products such as pea (Baloch et al. 2015), common bean (Aydin & Baloch 2019; Nemli et al.2015), laurel (Karik et al. 2019), quinoa (Hossein-Pour et al. 2019), pepper (Yildiz et al. 2020), chrysanthemum (Dalda-Sekerci 2023), daisy (Ozsoy et al. 2023), tulip (Salybekova et al. 2025) and rose (Kırbay et al. 2025). Although morphological, biochemical, and molecular marker systems provide valuable information for assessing genetic diversity and population structure, the integration and interpretation of large multidimensional datasets generated from these approaches remain challenging using conventional statistical methods alone. Therefore, advanced computational approaches such as machine learning have gained increasing attention for their ability to identify hidden patterns, improve classification accuracy, and support predictive analyses in complex biological datasets (Simsek et al. 2024).

Machine learning (ML) applies advanced data science approaches to solve complex problems across many scientific disciplines. Despite its widespread use in other fields, adoption of ML techniques in plant and agricultural sciences has been relatively limited (Aasim et al. 2023). However, several studies have reported significant successes using ML in different branches of plant sciences, especially in the field of plant breeding (Van Dijk et al. 2021). RF, SVM, and MLP models were selected because they represent complementary machine learning approaches with different prediction architectures and learning behaviors. RF is highly effective in handling nonlinear interactions and high-dimensional biological datasets, SVM provides strong generalization ability in relatively limited datasets, whereas MLP enables modeling of complex nonlinear relationships between phenotypic and genotypic variables. Therefore, the combined use of these algorithms allows a broader evaluation of predictive performance across different biological data structures. (Simsek et al. 2024; Norouzi et al.2024; Benelli et al. 2025; Yaman et al. 2025). On the other hand, SVM works as a learning algorithm for regression-based prediction tasksand aims to minimize classification and regression errors by optimally fitting the data (Norouzi et al.2024; Isak et al. 2024). Lastly, RF is particularly effective in solving the overfitting problem, handling noisy data, and processing high-dimensional datasets. (Simsek 2024; Tarraf et al. 2024).

Variation is a critical step in the selection process in plant breeding. This study aims to determine the morphological and biochemical characteristics of hollyhock genotypes sampled from different provinces of Türkiye and to reveal their genetic similarities and differences. In the study, genotypes with different morphological characteristics were sampled. Moreover, the genotypes were evaluated biochemically for total phenolic content and antioxidant properties. All morphological and biochemical characteristics were thoroughly examined and correlated with genetic similarity and diversity using molecular markers. Also, the comparative evaluation of different machine learning models revealed varying degrees of effectiveness, with MLP, SVM, and RF demonstrating different R^2 values for plant biochemical traits. In this context, the present study aimed to comprehensively evaluate the genetic diversity and population structure of 28 endemic hollyhocks (*A. rosea*) genotypes collected from different provinces of Türkiye through an integrated approach combining morphological, biochemical, and retrotransposon-based molecular markers together with machine learning algorithms. We hypothesized that the integration of multidimensional datasets with machine learning-based predictive analyses would improve the resolution of genetic diversity assessment and enable more accurate discrimination of population structure among hollyhock genotypes. Therefore, this study provides a valuable scientific framework and essential roadmap for future hollyhock breeding, conservation, and genetic improvement programs.

2. Material and Methods

2.1. Plants material

Hollyhock genotypes were sampled from different regions of Türkiye (Figure 1, Table 1). Ten plants were propagated and grown in pots containing a substrate mixture (1:1:1, v/v, peat, sand, perlite) under natural daylight for each genotype. Subsequently, the plants were transplanted into open field conditions with a row spacing of 1 meter and an inter-row spacing of 1.5 meters, arranged in a randomized complete block design with three replications. The graphical abstract of the study is presented in Figure 2.



Figure 1- Map of provinces where Hollyhock genotypes were sampled

Table 1- Data of the locations where 28 hollyhock genotypes were sampled

Genotype	*Location*	*Altitude (m)*	*Habitat*
1	Yozgat/ Center	1300	Grows naturally in urban settlements
2	Aksaray/ Center	978	Grows naturally in urban settlements
3	Kayseri/Yahyalı	1330	Roadside, scattered forest trees
4	Kayseri/Yahyalı	1330	Roadside, scattered forest trees
5	Kayseri/Yahyalı	1345	Roadside, scattered forest trees
6	Kayseri/Yahyalı	1345	Roadside, scattered forest trees
7	Kayseri/Yahyalı	1330	Roadside, scattered forest trees
8	Nevşehir/ Center	1150	Open field, meadow
9	Kayseri/Yahyalı	1330	Roadside, scattered forest trees
10	Kayseri/Yahyalı	1330	Roadside, scattered forest trees
11	Kayseri/Yahyalı	1340	Roadside, scattered forest trees
12	Kayseri/Yahyalı	1330	Roadside, scattered forest trees
13	Nevşehir/Derinkuyu	1124	Open field, meadow
14	Nevşehir/ Center	1178	Open field, meadow
15	Kayseri/ İncesu	1093	Grows naturally in urban settlements
16	Kayseri/ İncesu	1093	Grows naturally in urban settlements
17	Kayseri/ İncesu	1094	Grows naturally in urban settlements
18	Ankara/Etimesgut	1100	Grows naturally in urban settlements
19	Kahramanmaraş/Ekinözü	1200	Roadside, scattered forest trees
20	Kahramanmaraş/Ekinözü	1200	Roadside, scattered forest trees
21	Adana/Tekir yaylası	2124	Forest boundary
22	Adıyaman/ Center	669	Open field, meadow
23	Mardin/Silopi	1083	Open field, meadow
24	Konya/Meram	1037	Grows naturally in urban settlements
25	Franchi/ Sementi (France)	-	Commercial variety
26	Konya/ Meram	1038	Open field, meadow
27	Konya /Selçuklu	1016	Grows naturally in urban settlements
28	Yalova/ Center	35	Roadside, scattered forest trees

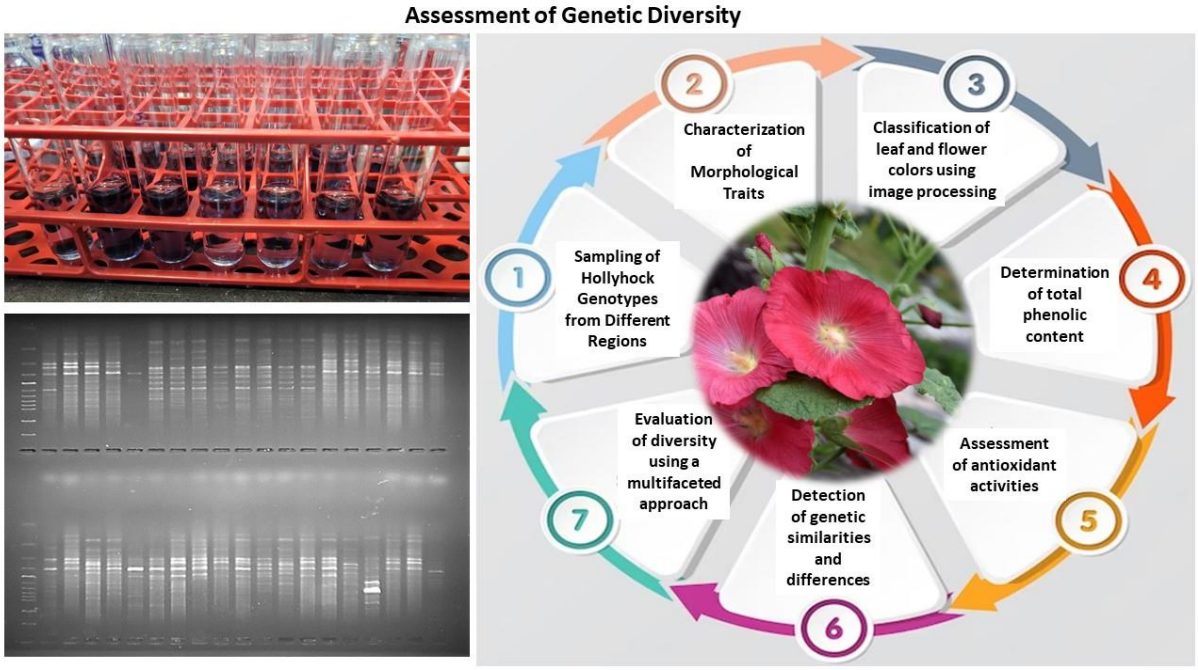

Figure 2- Graphical abstract showing comprehensive assessment of genetic diversity in 28 hollyhock genotypes

2.2. Morphological characterization method

The 28 hollyhock genotypes were evaluated for various phenotypic traits such as plant height, leaf length, leaf width, leaf depth, leaf shape, pedicel anthocyanin content, pedicel hairiness, flower color, flower color variation and pistil color. In addition, color measurements for flowers and leaves were determined using an image acquisition system and image processing techniques. The experiments were carried out according to UPOV (Union for the International Protection of New Varieties of Plants, United Kingdom) guidelines. Using the image acquisition methods described by Gunaydın et al. (2024), the color features of hollyhock flowers and leaves were identified. This system's components include a digital single-lens reflex (DSLR) camera, a tripod that can be adjusted to a 90º angle, and a lighting configuration. Background removal was performed using Adobe® Photoshop® CS6 software, then processed using Image J software version 1.53t and converted into L, a, b, color channels (Cetin et al. 2024; Gunaydın et al. 2024).

2.3. Total phenolic content analysis

One gram of dried hollyhock flower sample was mixed with 30 mL of an 80:20 methanol solution and shaken for 1 hour to extract the total phenolic content. The mixture was then centrifuged at 3900 rpm for 15 minutes. The total phenolic content of 28 hollyhock genotypes was determined using the method outlined by Singleton & Rossi (1965), with three replicates for each genotype. Specifically, 200 µL of hollyhock extract (diluted 1:1 with methanol, 80:20) was mixed with 1800 µL of distilled water, followed by the addition of 1 mL of Folin-Ciocalteu reactant (diluted 1:10 with water). Then, 2 mL of sodium carbonate solution (2% w/v) was added and the mixture was vortexed and incubated for two hours at room temperature (21 ºC) in the dark. The absorbance of the samples was measured at 765 nm by UV-visible spectrophotometer. Total phenolic content was calculated in mg GAE/kg based on the calibration curve prepared with gallic acid standard.

2.4. Antioxidant activity analysis

2.4.1. Antioxidant activity by iron reducing power (TPTZ) method

The antioxidant capacity of the samples was evaluated using the TPTZ method following the protocol of Aktumsek et al. (2013). Briefly, 100 µL of sample was mixed with 2 mL of FRAP reactant containing acetate buffer (0.3 M, pH 3.6), 10 mM TPTZ, and 20 mM $FeCl_3$ at a ratio of 10:1:1. The mixture was incubated at room temperature for 30 min before absorbance was measured at 593 nm. Antioxidant activity was expressed in mg ascorbic acid equivalent (mg AAE/kg) based on the calibration curve established with ascorbic acid standard.

2.4.2. Antioxidant activity by copper reducing power (CUPRAC) method

The antioxidant activity of the samples was assessed using the CUPRAC method, following the protocol of Apak et al. (2006). A reaction mixture was prepared by mixing 1 mL of $CuCl_2$ $2H_2O$ (0.01 M), ammonium acetate (1 M), and neocuproin (0.00075

M) with 0.6 mL of distilled water, respectively. Then, 0.5 mL of sample extract was added and the mixture was incubated at room temperature for 30 min. Absorbance was measured at 450 nm and antioxidant activity was expressed in mM Trolox equivalent (TE)/g based on the calibration curve established with Trolox standards.

2.4.3. Antioxidant activity by 2,2-Diphenyl-1-picrylhydrazyl (DPPH) method

The DPPH radical scavenging activity of hollyhock flower extracts was determined according to the method of Brand-Williams et al. (1995). Briefly, 100 µL of the extract (diluted 1:1 with methanol: water, 80:20) was mixed with 3.9 mL of DPPH solution (0.1 mM in methanol). The mixture was vortexed and incubated in darkness at room temperature for 30 minutes. Absorbance was then measured at 517 nm using a spectrophotometer. The percentage inhibition was calculated using the formula:

% Inhibition=(($A_{Control}$-A_{Sample})/$A_{Control}$) x 100 ($A_{Control}$ is the absorbance value of the sample, and A_{Sample} represents the sample absorbance value).

Morphological and biochemical data were analyzed using SAS software (version 9.00, SAS Institute, USA) with the General Linear Model (GLM) procedure. Treatment means were compared using Duncan's multiple range test at $\alpha = 0.05$ or 0.01. ANOVA was conducted to evaluate morphological parameters, and variance analysis of morphological traits was performed in SAS. Additionally, Principal Component Analysis (PCA) was conducted using PAST software.

2.5. Molecular characterization method

Genomic DNA was extracted from young leaves of 28 hollyhock genotypes using a modified CTAB protocol and subsequently purified (Dalda-Sekerci 2023). Ten iPBS (inter-primer binding site) retrotransposon-based primers were used for PCR amplification (Table 6). Each 15 µl PCR reaction contained 1.5 µl of Taq buffer, 0.33 µl of 2.5 mM dNTPs, 0.2 µl of Taq DNA polymerase, 2.5 µl (20 ng) of template DNA, and 1 µl (5 pM) of each iPBS primer. PCR amplification was performed with an initial denaturation step at 95 °C for 4 min, followed by 35 cycles of denaturation at 95 °C for 1 min, annealing at 52–56 °C for 1 min, and extension at 72 °C for 1 min, with a final extension step at 72 °C for 7 min. Amplified products were separated on 1.5% agarose gel electrophoresis in TBE buffer at 110 V for 3 h, and DNA fragments were visualized under UV light using a gel documentation system (Kodak EL Logic 200, USA). Clear and reproducible bands were scored as present (1) or absent (0) to construct a binary data matrix. Genetic similarity among 28 hollyhock genotypes was estimated using Dice's similarity coefficient (Dice, 1945), and cluster analysis was performed using the UPGMA method based on the SAHN algorithm in NTSYS-pc (version 2.11). Population structure was analyzed using STRUCTURE software (version 2.3.4), with an admixture model and correlated allele frequencies. The number of clusters (K) was set from 1 to 10, with five independent runs per K, a burn-in period of 100,000 iterations, and 100,000 MCMC replications (Pritchard et al.2000; Falush et al. 2003).

2.6. Machine learning analyses

Antioxidant activities (DPPH, TPC, CUPRAC, and FRAP) have been predicted using different machine learning models-MLP, SVM, and RF. The models were trained using three different types of datasets: phenotypic data (P) (leaf width (cm), leaf length (cm) and plant height (m)), genotypic data (G) (containing different markers) and combined phenotypic and genotypic data (PG). Input features included measurements from these datasets, while output variables were antioxidant activities to be predicted. The SVM model was implemented using a linear kernel (svmLinear), whereas RF and MLP models were implemented using the rf and mlp algorithms available in the caret package. Model tuning and parameter optimization were automatically performed during 10-fold cross-validation within the caret framework. Prior to model training, missing values were inspected and datasets were normalized using range transformation. Subsequently, the data were randomly partitioned into training (80%) and testing (20%) subsets using the `createDataPartition` function in the caret package in R. During model training, variables were centered and scaled, and model validation was performed using 10-fold cross-validation to improve model robustness and reduce overfitting. Three main metrics were used to evaluate and compare the accuracy and precision of the models: coefficient of determination (R^2), which evaluates the relationship between the predicted antioxidant values and the true values (Eq. 1); Root Mean Square Error (RMSE), which measures the degree of deviation of the predicted values from the true data points (Eq. 2); and Mean Absolute Error (MAE), which calculates the mean difference between the predicted and observed values (Eq. 3). These metrics provided a detailed analysis of the effectiveness of each model and allowed a comprehensive comparison of the success of the MLP, SVM and RF models in predicting antioxidant activities.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{(\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2)}{n}} \quad (2)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \hat{Y}_i| \quad (3)$$

While $\boldsymbol{Y_i}$ = actual value, $\boldsymbol{\hat{Y}_i}$ = Predicted value, $\boldsymbol{\bar{Y}}$ = mean o the actual values and n = sample count

2.6.1. Antioxidant prediction using MLP

MLP is a widely used type of Artificial Neural Network (ANN) that consists of an input layer, one or more hidden layers, and an output layer. MLP uses supervised learning, which means that the network is trained on a dataset that contains both input and corresponding output variables (Simsek et al. 2024). During training, the weight coefficients of the input nodes are adjusted to optimize performance. These networks consist of basic neural units arranged in layered structures (Norouzi et al. 2024). During training, the MLP model minimizes the prediction error between observed and predicted outputs through iterative optimization of network weights, as represented in Eq. 4. The objective function used in the MLP model represents the prediction error that the network attempts to minimize during training through iterative adjustment of connection weights.

$$E = \frac{1}{n}\sum_{n=1}^{n}(y_s - \hat{y_s}) \quad (4)$$

n = number of observations, ys is the sth observation variable $\hat{y_s}$ is the sth of predicted variable.

To compute the predicted output $\hat{y}$ in an MLP with a hidden layer comprising *p* neurons and *k* output variables, Eq. 5 is employed. The equation describes how the final predicted output is generated through weighted transformations of the input variables across hidden and output layers.

$$\hat{\boldsymbol{y}} = \boldsymbol{f}\left[\sum_{j=1}^{\boldsymbol{p}} \boldsymbol{w_{ji}} \cdot \boldsymbol{g}\left(\sum_{i=1}^{\boldsymbol{k}} \boldsymbol{w_{ji}}\boldsymbol{x_i} + \boldsymbol{w_{j0}}\right) + \boldsymbol{w_o}\right] \quad (5)$$

xi represents the *ith* output variable, *wj* corresponds to the weighted input data entering the *jth* hidden neuron, *f* is the activation function applied to the output neuron, *wji* signifies the weight associated with the direct connection from input neuron *i* to hidden neuron *j*, *wj0* represents the bias specific to the *jth* neuron, *w0* represents the bias linked to the output neuron, and *g* is the activation function utilized for the hidden neuron.

2.6.2. Antioxidant prediction using SVM

SVMs are a flexible class of algorithms that encompass both supervised and unsupervised learning techniques, widely applied in regression, clustering, and classification tasks (Isak et al. 2024). The basic principle of SVMs is to determine a decision boundary that maximizes the margin between support vectors of different classes (Norouzi et al. 2024). In this study, SVM regression was employed to model nonlinear relationships between input variables and antioxidant traits by minimizing prediction error while maximizing generalization ability (Eq.6)

$$f(x) = w\varphi(x) + b \quad (6)$$

2.6.3. Antioxidant prediction using RF

RF is very effective in preventing overfitting, handling noisy data, and processing high-dimensional data sets. The RF model introduces randomness in two ways during the construction of each decision tree, which helps to minimize correlations between the trees while retaining the strengths of individual trees (Simsek, 2024). RF constructs multiple decision trees using randomized subsets of variables and observations, and the final prediction is generated by aggregating outputs from individual trees. , as outlined in Eq. 7 (Tarraf et al. 2017). The RF algorithm combines predictions from multiple independently generated decision trees to improve prediction stability and reduce overfitting.

$$y = \sum_{i=1}^{n}(a_i - a_i^*)k(x, x_i) + b \quad (7)$$

3. Results

3.1. Morphological characterization results

The 28 Hollyhock (*A. rosea*) genotypes naturally occurring in various provinces of Türkiye were collected and characterized based on their morphological traits. Genotypes means were compared using Duncan's multiple range test at α = 0.01. ANOVA was conducted to evaluate morphological parameters. Measurement and observation results regarding the morphological characteristics of the genotypes are presented in Figure 3 and Table 2. Considerable morphological variability was observed among the 28 hollyhock genotypes evaluated in the present study, indicating the presence of substantial phenotypic diversity within the studied germplasm. Plant height ranged from 1.09 m to 1.83 m, while marked differences were also detected in leaf morphology, including leaf size, shape, and anthocyanin pigmentation. Particularly, the pronounced variation in leaf lobing and pigmentation patterns suggests the existence of genetically diverse morphological traits that may contribute to ornamental value and environmental adaptation. In addition, substantial diversity in flower coloration was observed among the genotypes, further

highlighting their potential importance for ornamental breeding programs. Despite the observed phenotypic variation, the consistent yellow pistil color across all genotypes confirmed their taxonomic classification as *Alcea rosea*. Overall, the morphological findings presented in Figure 3 and Table 2 demonstrate a broad level of diversity among the studied genotypes, emphasizing their potential utility for future breeding, selection, and conservation studies.

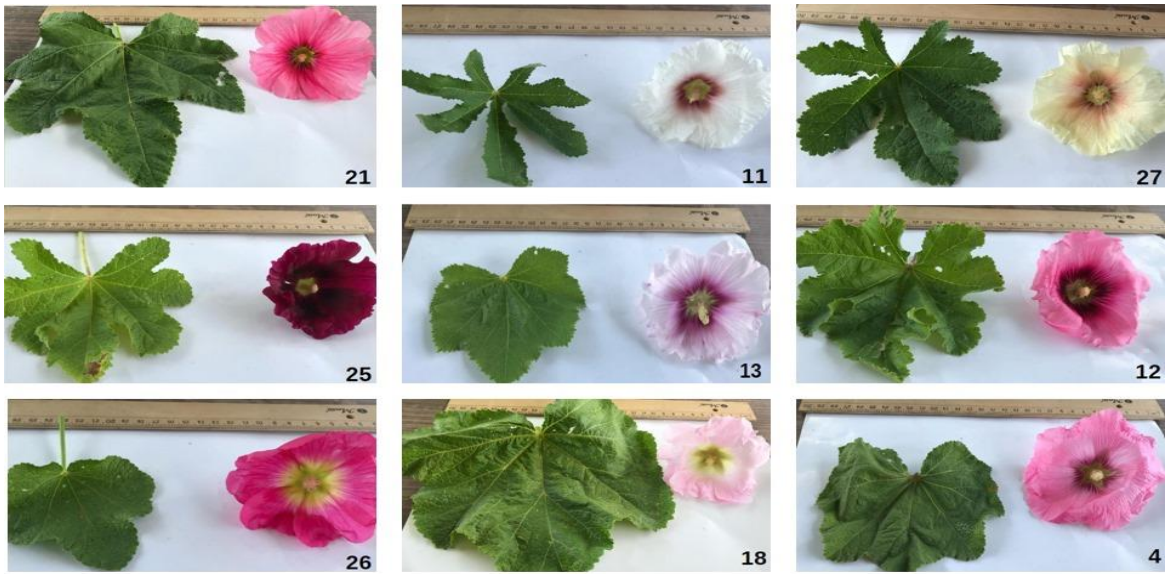

Figure 3- Visuals of some hollyhock genotypes with different leaf and flower characteristics

Table 2- Morphological measurement and observation results of 28 hollyhock genotypes

Genotype No	*Plant high (m)*	*Leaf length (cm)*	*Leaf width (cm)*	*Leaf Depth*	*Leaf Shape*	*Anthocyanin content of pedicel*	*Pedicel hairiness*	*Flower color*	*Color variety of flowers*	*Flower pistil color*
1	1.2a-c	16.0b	17.0b	medium	obtuse	low	low	light pink	monocolor	yellow
2	1.18bc	15.0bc	14.5b-d	shallow	acute	high	high	pink	monocolor	yellow
3	1.34a-c	14.0b-d	12.5b-d	medium	obtuse	low	low	magenta	monocolor	yellow
4	1.59a-c	13.0b-e	11.0cd	shallow	obtuse	medium	moderate	pink	monocolor	yellow
5	1.38a-c	8.0f	15.0bc	medium	acute	medium	moderate	magenta	monocolor	yellow
6	1.40a-c	13.5b-e	9.0d	medium	obtuse	medium	moderate	white-pink	bicolor	yellow
7	1.20bc	13.0b-e	14.0b-d	shallow	rounded	medium	moderate	pink-purple	bicolor	yellow
8	1.42a-c	15.0bc	16.0bc	medium	obtuse	medium	moderate	magenta	monocolor	yellow
9	1.37a-c	15,0bc	13.33b-d	medium	obtuse	low	low	magenta	monocolor	yellow
10	1.84a	13.0b-e	15.0bc	medium	obtuse	medium	moderate	pink-purple	bicolor	yellow
11	1.46a-c	14.73bc	12.0b-d	deep	acute	low	moderate	white-pink	bicolor	yellow
12	1.60a-c	13.0b-e	12.5b-d	medium	acute	high	high	pink-purple	bicolor	yellow
13	1.62a-c	11.0d-f	11.30cd	shallow	rounded	low	low	white-pink	bicolor	yellow
14	1.09c	9.0f	10.5cd	shallow	rounded	high	high	magenta-pink	bicolor	yellow
15	1.36a-c	14.0b-d	13.0b-d	medium	acute	medium	moderate	burgundy	monocolor	yellow
16	1.74ab	14.2b-d	14.5b-d	medium	acute	medium	moderate	pink	monocolor	yellow
17	1.22bc	10.76c-f	11.0cd	shallow	obtuse	medium	moderate	pink	monocolor	yellow
18	1.83a	21.5a	23.0a	shallow	acute	medium	moderate	light pink	monocolor	yellow
19	1.52a-c	10.0ef	10.5cd	shallow	acute	low	low	white-pink	bicolor	yellow
20	1.24bc	9.0f	11.0cd	medium	acute	low	low	white	monocolor	yellow
21	1.85a	9.0f	17.0b	medium	obtuse	medium	moderate	pink	monocolor	yellow
22	1.38a-c	11.66c-f	13.0b-d	medium	rounded	high	moderate	pink-white	bicolor	yellow
23	1.17bc	11.66c-f	15.0cb	medium	acute	low	low	magenta	monocolor	yellow
24	1.38a-c	10.0ef	12.0b-d	deep	acute	high	high	magenta	monocolor	yellow
25	1.15bc	15.0bc	11.17cd	medium	obtuse	low	moderate	burgundy	monocolor	yellow
26	1.36a-c	10.0ef	13.83b-d	shallow	rounded	low	low	pink	monocolor	yellow
27	1.53a-c	15.0bc	16.0bc	deep	acute	low	low	cream	monocolor	yellow
28	1.71ab	15.26bc	14.0b-d	shallow	acute	low	low	white-pink	bicolor	yellow
Means	**1.444****	**13.523****	**13.030****	-	-	-	-	-	-	-

Means indicated with different letters in the same column are significantly different *: (P<0.05), **: significant at P<0.01

In this study, the colors of the flowers and leaves of the genotypes were determined using image processing techniques, including the analysis of color channels and color spaces. The Lab color space comprises L* (lightness), a* (red/green), and b* (yellow/blue). Mean values of the color channels were employed for classification. The mean values of color channels for the flower and leaf samples are presented in Table 3.

In horticulture, the L*, a*, and b* values, crucial criteria for color classification, exhibit significant variation in flower colors. The L value, representing lightness, was found to be high in genotypes with light-colored flowers and low in genotypes with dark-colored flowers, ranging from 90.67 to 24.05. Similarly, the a* value, which indicates redness, was high in dark-colored flowers and lowest in white and light-toned flowers. The a* value of hollyhock genotypes ranged from 68.38 to 1.34. In the same context, the b* value, representing yellowness, exhibited a broad variation, ranging from 10.41 to -24.71.

Considerable variation in leaf color parameters was also observed among the hollyhock genotypes, reflecting substantial phenotypic diversity in foliar pigmentation characteristics. The L* values ranged from 20.60 to 82.63, indicating pronounced differences in leaf brightness among genotypes, with some exhibiting lighter and others darker green coloration. Similarly, wide variation was detected for a* and b* color coordinates, which ranged from −23.91 to 44.29 and from 8.00 to 39.22, respectively. These findings demonstrate significant diversity in leaf color traits, which may contribute to both the ornamental value and genotype differentiation of hollyhock populations.

Table 3- The mean of all color channels for Hollyhock leaves and flowers

	Leaves			*Flowers*		
Genotypes	***L***	***a****	***b****	***L***	***a****	***b****
1	39.61c-h	-16.31b	20.95e	77.76a-c	16.14f	0.04b-e
2	36.99e-ı	-18.63bc	28.69a-e	61.81d-f	63.46ab	4.03ab
3	34.70g-j	-18.94bc	27.78a-e	36.74ı-k	61.91a-c	-1.85b-f
4	39.54c-h	-17.87bc	22.48d-e	54.75e-h	52.29b-e	-7.80d-ı
5	38.48c-h	-19.68bc	26.84a-e	49.59f-ı	62.39a-c	-17.97jk
6	34.65g-j	-17.13bc	22.03d-e	80.71ab	8.87fg	-7.29d-ı
7	37.34d-ı	-18.87bc	25.10b-e	61.48d-f	39.36e	-9.13f-ı
8	43.20c-e	-20.58bc	27.72a-e	57.94d-g	49.32c-e	-5.85c-ı
9	37.19e-ı	-21.74bc	34.22a-e	54.09e-h	62.36a-c	-6.91c-ı
10	36.52e-j	-19.24bc	27.38a-e	58.42d-g	49.84c-e	-12.74ıj
11	42.31c-f	-23.91c	39.22a	55.54e-h	51.11b-e	-5.75c-ı
12	42.91c-e	-22.18bc	35.52a-d	52.66e-h	55.13b-d	-3.54b-g
13	42.46c-f	-22.64bc	31.51a-e	76.36bc	11.36fg	-12.37h-j
14	36.57e-j	-19.67bc	24.23b-e	47.17g-ı	46.32de	-6.18c-ı
15	45.25c	-21.91bc	37.94ab	24.05k	42.85de	10.41a
16	35.48f-j	-15.75b	21.76d-e	45.78g-j	61.29a-c	-12.69h-j
17	71.29b	44.29a	8.00f	71.29b-d	44.29de	-8.04e-ı
18	82.63a	-20.37bc	32.87a-e	82.74ab	21.45f	-4.42c-h
19	20.60jk	-21.47bc	29.74a-e	79.32ab	2.61g	-3.77b-g
20	26.95k	-21.11bc	27.91a-e	84.25ab	1.34g	-6.34c-ı
21	39.84c-h	-21.70bc	28.09a-e	65.10c-e	54.79b-d	0.40b-d
22	35.05g-j	-17.66bc	22.51d-e	75.60bc	10.54fg	-6.48c-ı
23	41.57c-g	-22.11bc	31.88a-e	46.98g-ı	68.38a	-10.38g-j
24	32.74h-k	-16.79bc	36.76ac	43.86e-j	54.18b-d	-24.71k
25	44.49cd	-23.55c	38.58ab	32.53jk	45.86de	1.24bc
26	42.76c-e	-21.98bc	30.40a-e	51.09f-h	60.88a-c	-5.16c-ı
27	30.35ı-k	-19.47bc	23.97c-e	90.52a	2.75g	-3.55b-g
28	41.84c-g	-20.78bc	32.24a-e	90.67a	3.01g	-3.18b-g
Means	**40.801****	**-17.780****	**27.907****	**59.940****	**40.830****	**-6.164****

Means indicated with different letters in the same column are significantly different *: (P<0.05), **: significant at P<0.01

3.2. Bioactive characteristics results

The total phenolic contents and antioxidant activity values of the hollyhock genotypes are detailed in Table 4. Significant variations were observed among the genotypes in terms of both total phenolic content and antioxidant activity. Substantial variation in total phenolic content was observed among the evaluated hollyhock genotypes, with values ranging from 1.18 to 4.37 mg GAE/g dry flower weight. The pronounced differences in phenolic accumulation among genotypes indicate considerable biochemical diversity within the studied population. Notably, genotypes with darker flower coloration generally exhibited higher phenolic content compared to lighter-colored genotypes, suggesting a possible association between flower pigmentation and phenolic compound accumulation. The antioxidant activity of the hollyhock genotypes, expressed as scavenging activity (% inhibition), is also presented in Table 4. According to the antioxidant activity measured by the iron-reducing power (TPTZ)

method, statistically significant differences were detected between the genotypes regarding their antioxidant activities. Antioxidant activity analyses revealed substantial variation among the evaluated hollyhock genotypes, indicating considerable diversity in their bioactive properties. Consistent patterns were obtained across the different antioxidant assays, including CUPRAC and DPPH methods, demonstrating the reliability and compatibility of the analytical approaches used. Certain genotypes exhibited notably higher antioxidant capacities, whereas others showed comparatively low activity levels across all tested methods. The observed variation in antioxidant potential closely paralleled the morphological and biochemical diversity detected among the genotypes, suggesting the presence of genetically distinct materials with differing capacities for bioactive compound accumulation. These findings emphasize the potential value of selected hollyhock genotypes for future breeding programs targeting enhanced phytochemical and antioxidant properties. These findings highlight the potential importance of hollyhock genotypes as valuable genetic resources for ornamental, nutritional, and phytochemical breeding programs.

Additionally, Principal component analysis (PCA) was conducted to examine the relationship between flower color and the bioactive properties of hollyhock genotypes, revealing that the total variation (PC1 and PC2) accounted for 83%. The resulting PCA graph indicated a positive correlation between total phenolic content and antioxidant activity with flower color. Statistically, darker flower colors corresponded with an increase in bioactive components (Figure 4).

Additionally, correlation analysis was conducted between flower color and biochemical properties of hollyhock genotypes. According to the correlation analysis, there was a significant correlation between total phenolic content and flower color, with a considerable increase in total phenolic content observed as flower color darkened. Similarly, a positive correlation was found between total phenolic content and antioxidant activity. However, flower brightness negatively correlated with flower color, indicating that flower brightness decreased as flower color darkened (Figure 5).

Table 4- Bioactive characteristics of 28 hollyhock genotypes

Genotypes	*TPC (mg GAE/g)*	*FRAP (%)*	*CUPRAC (%)*	*DPPH (%)*
1	1.67ı	7.06m-n	41.80d	16.808l-n
2	2.10gh	20.20b	41.42d	46.183b
3	2.39d-g	13.08e-g	29.43fg	28.149d-f
4	0.18j	11.78f-h	37.47de	29.720de
5	1.60ı	6.24n	40.44d	15.708mn
6	1.50ı	5.82n	31.27e-g	15.739mn
7	1,84hı	11.57g-ı	28.91fg	24.096f-ı
8	2.17f-h	11.22g-j	50.98c	26.13d-h
9	1.86hı	7.77l-n	26.06g	17.090l-n
10	2.76c-e	10.69h-k	26.92fg	24.253f-ı
11	4.65a	12.04e-h	41.80d	20.766ı-l
12	2.13gh	14.12de	41.42d	30.348d
13	1.64ı	8.89k-m	29.43fg	17.436l-n
14	3.06c	11.87e-h	37.46de	20.860ı-l
15	1.81hı	16.06cd	40.44d	28.306d-f
16	2.39e-g	9.08j-m	31.27e-g	22.714h-k
17	2.06gh	7.82l-n	28.91fg	18.347lm
18	1.80hı	16.76c	50.98c	27.332d-g
19	2.77cd	5.80n	26.06g	13.666n
20	2.74c-e	6.46n	26.92fg	15.614mn
21	2.52 d-f	9.20j-n	31.71e-g	23.656g-j
22	1.84hı	11.60g-ı	39.47d	18.661k-m
23	2.25fg	10.84g-k	32.90ef	25.730e-h
24	2.71c-e	13.88d-f	42.38d	28.023d-g
25	4.37a	22.71a	78.26a	70.028a
26	3.85b	18.15c	53.74c	35.752c
27	3.56b	16.76c	61.96b	40.025c
28	1.55ı	9.37ı-l	29.60fg	19.352j-m
Means	**2.345****	**11.675****	**38.553****	**25.732****

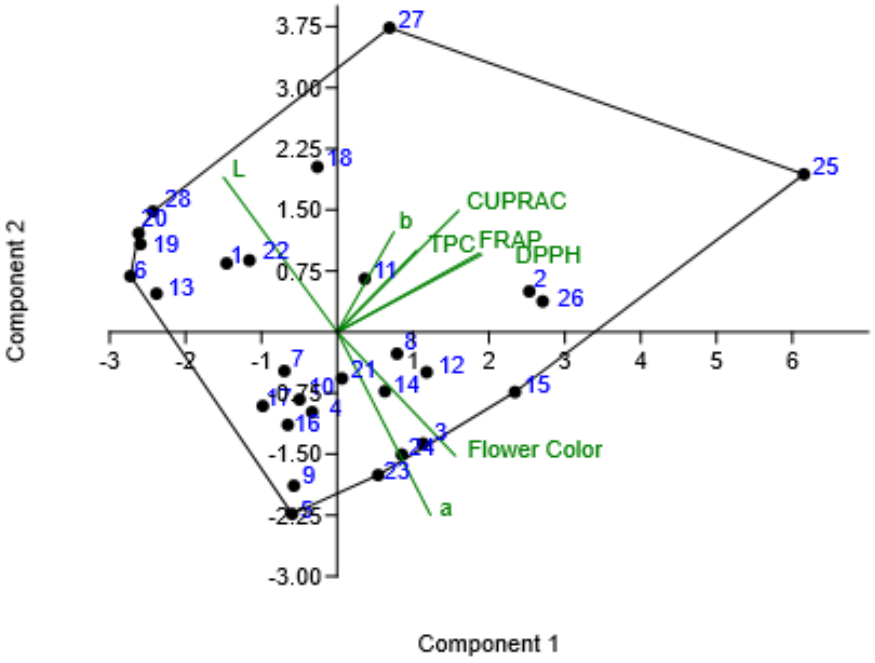

Figure 4- A principal component analysis graph generated to examine the relationship between flower color and bioactive components in 28 hollyhock genotypes

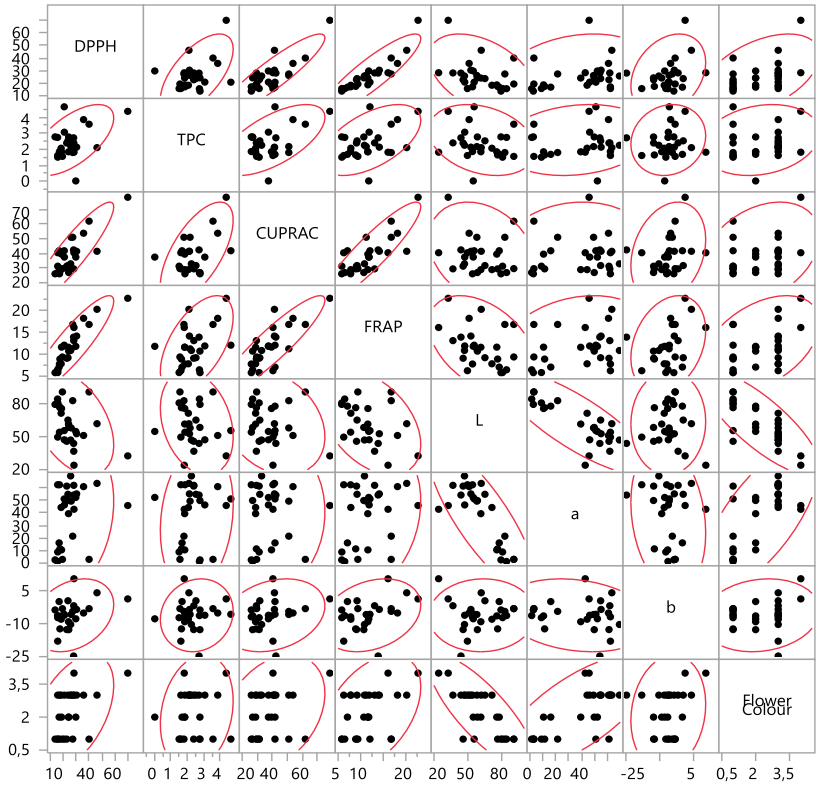

Figure 5- Scatterplot matrix and correlation between flower color and biochemical traits of 28 hollyhock genotypes

3.3. Molecular characterization results

The iPBS marker analysis revealed considerable genetic variation among the 28 hollyhock genotypes. A total of 113 bands were generated using ten iPBS primers, of which 106 were polymorphic, corresponding to a mean polymorphism rate of 94%. The iPBS marker analysis revealed a clear level of genetic variability among the hollyhock genotypes. A total of polymorphic bands ranging from 125 to 1050 bp were generated across the primers, indicating a wide distribution of amplified genomic regions. Among the primers, iPBS-2075 showed the highest amplification efficiency with 18 bands, whereas iPBS-2218 produced the

lowest number of bands (n = 7), reflecting differences in primer informativeness (Tables 5 and 6). At the population genetic level, allele frequency estimates (p and q) varied between 0.252–0.654 and 0.346–0.748, respectively, suggesting an imbalance between allele classes and indicating the predominance of recessive alleles in the studied population. The mean number of effective alleles (Ne = 1.614) further supports a moderate level of genetic diversity among genotypes. Similarly, Shannon's information index (I = 0.512 on average) and heterozygosity estimates (He: 0.253–0.400; uHe: 0.258–0.408) collectively demonstrate a moderate but consistent level of genetic variation within the hollyhock germplasm. Overall, these results confirm the usefulness of iPBS markers in detecting genetic diversity and resolving population-level differences in *Alcea rosea*. These values indicate that there is a significant level of genetic diversity in the population. It can be said that some primers are more successful than others in detecting genetic diversity.

Table 5- Polymorphism values of studied iPBS primers

Primary	*Sequence 5'-3'*	*Bp*	*TNF*	*NPF*	*MP (%)*	*p*	*q*	*Ne*	*I*	*He*	*uHe*
iPBS-2383	GCATGGCCTCCA	175-900	14	12	85.71	0.270	0.721	1.501	0.451	0.299	0.304
iPBS-2387	GCGCAATACCCA	150-1000	9	7	77.78	0.252	0.748	1.448	0.379	0.253	0.258
iPBS-2239	ACCTAGGCTCGGATGCCA	125-700	12	11	91.67	0.375	0.625	1.709	0.553	0.384	0.391
iPBS-2249	AACCGACCTCTGATACCA	200-1000	8	7	87.5	0.330	0.670	1.602	0.480	0.330	0.337
iPBS-2218	CTCCAGCTCCGATTACCA	125-800	7	7	100	0.492	0.508	1.654	0.522	0.360	0.367
iPBS-2228	ACTAGGCTCTGTATACCA	250-1100	8	8	100	0.585	0.415	1.627	0.515	0.351	0.357
iPBS-2230	TCTAGGCGTCTGATACCA	150-1050	15	15	100	0.654	0.346	1.695	0.587	0.400	0.408
iPBS-2075	CTCATGATGCCA	100-1150	18	18	100	0.456	0.544	1.651	0.558	0.379	0.386
iPBS-2398	GAACCCTTTGCCGATACCA	100-900	9	8	88.89	0.263	0.737	1.552	0.502	0.334	0.341
iPBS-2221	ACCTAGCTCACGATCA	125-1050	13	13	100	0.574	0.426	1.702	0.569	0.390	0.397
Total	-	-	113	**106**		-	-	-	-	-	-
Mean	-	**125-1050**			93.81	0.369	0.631	1.614	0.512	0.348	0.354

*TNF: Total Number of Fragments, NPF; Number of Polymorphic Fragments, MP: Mean Polymorphism, p, and q: Allele Frequency, Ne: Number of Effective Alleles, I: Shannon's Information Index, He: Expected Heterozygosity and uHe: Unbiased Expected Heterozygosity

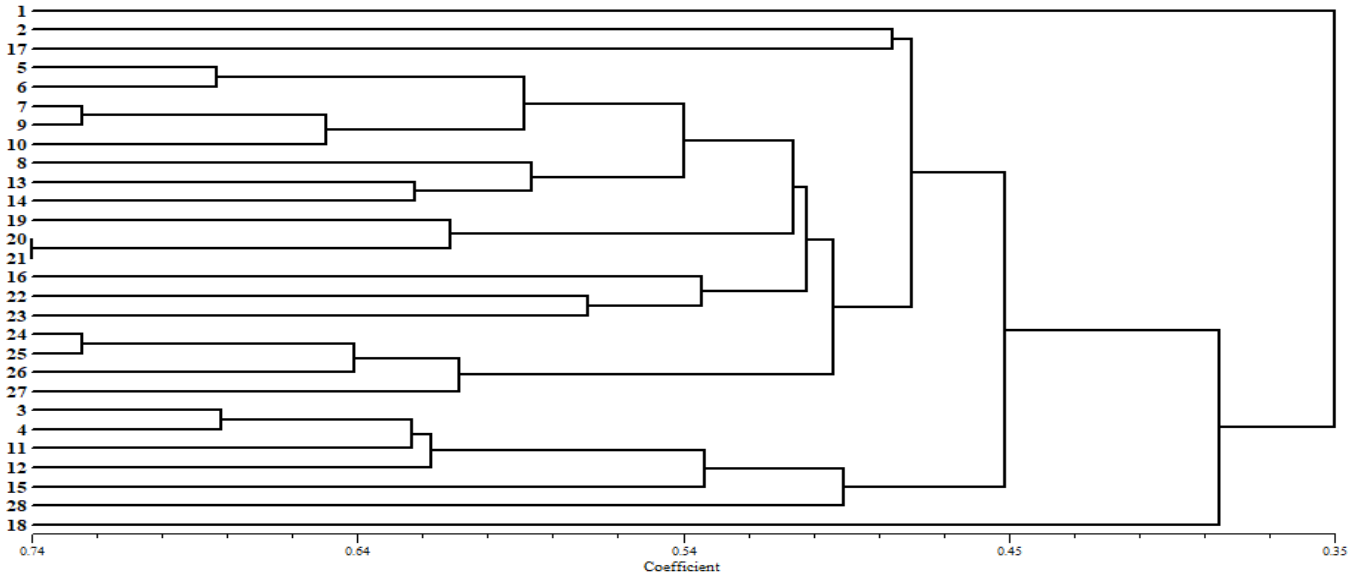

Figure 6- The UPGMA analysis based on Dice coefficients of iPBS markers from the 28 hollyhock genotypes

Cluster analysis based on Dice's similarity coefficient and the UPGMA method revealed genetic similarity values ranging from 0.35 to 0.74. The generated dendrogram reveals significant genetic diversity among all genotypes (Figure 6). Genotype 1 is positioned distantly from the other genotypes, indicating distinct genetic characteristics with only 35% genomic similarity. Similarly, genotype 18 is situated on a separate branch, displaying approximately 38% genetic similarity to the other genotypes. Some genotypes, however, exhibit greater genetic similarity. For instance, genotypes 5 and 6 share about 68% genetic similarity, while genotypes 7 and 9 exhibit 72% similarity. This high similarity could be attributed to the possibility that these genotypes were sampled from the same regions and might have proliferated through seed dispersal. Conversely, genotypes 20 and 21, despite being collected from different provinces, show 74% genetic similarity. Likewise, genotype 24, a commercial variety,

demonstrates high genetic similarity with genotype 25. These findings indicate substantial variation even within this study's limited sample size and geographic scope.

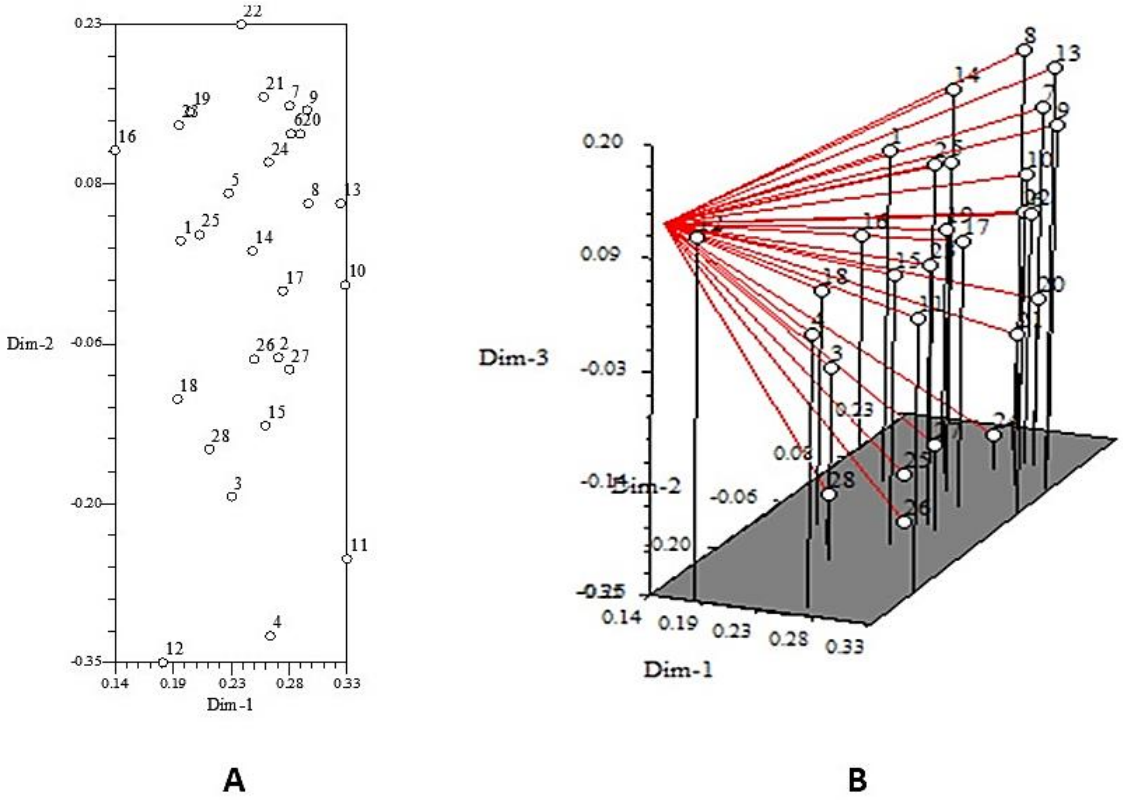

Figure 7- Two-dimensional (A) and three-dimensional (B) graphics obtained as a result of Principal Coordinate Analysis with iPBS data in 28 hollyhock genotypes

The NTSYS program's Mantel test was used to do a correlation analysis on DNA matrix data produced with various similarity matrices utilizing primers. At r = 0.996, the Mantel correlation value was high. The NTSYS program's Principal Coordinate Analysis (PCoA) was used to create two- and three-dimensional images. The cumulative sum of the first three eigenvalues in PCoA was 42.98, accounting for 43% of the total variation. The two-dimensional principal coordinate analysis graph showed that the majority of genotypes were scattered rather than grouped together. However, it was noted that, while being sampled from the same region, genotypes 4, 11, and 12 were positioned independently of the other genotypes (Figure 7). According to this study, hollyhock genotypes sampled from several provinces showed morphological and genetic variations, leading to a dispersed distribution as opposed to grouping in particular ways. This indicates that the phenotypic variation among the sampled genotypes is due primarily to genetic differences rather than morphological and biochemical content changes driven by environmental conditions.

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
2	3	-1837.83333	216.41988	NA	NA	NA
3	3	-2148.16667	445.56266	-310.33333	87.03333	0.19533
4	3	-2371.46667	690.08261	-223.30000	1103.26667	1.59875
5	3	-3698.03333	3234.22174	-1326.56667	1725.03333	0.53337
6	3	-3299.56667	1360.82505	398.46667	1259.56667	0.92559
7	3	-4160.66667	3045.02857	-861.10000	740.30000	0.24312
8	3	-5762.06667	4298.67317	-1601.40000	3805.70000	0.88532
9	3	-3557.76667	2723.79791	2204.30000	1383.13333	0.50780
10	3	-2736.60000	1390.94191	821.16667	NA	NA

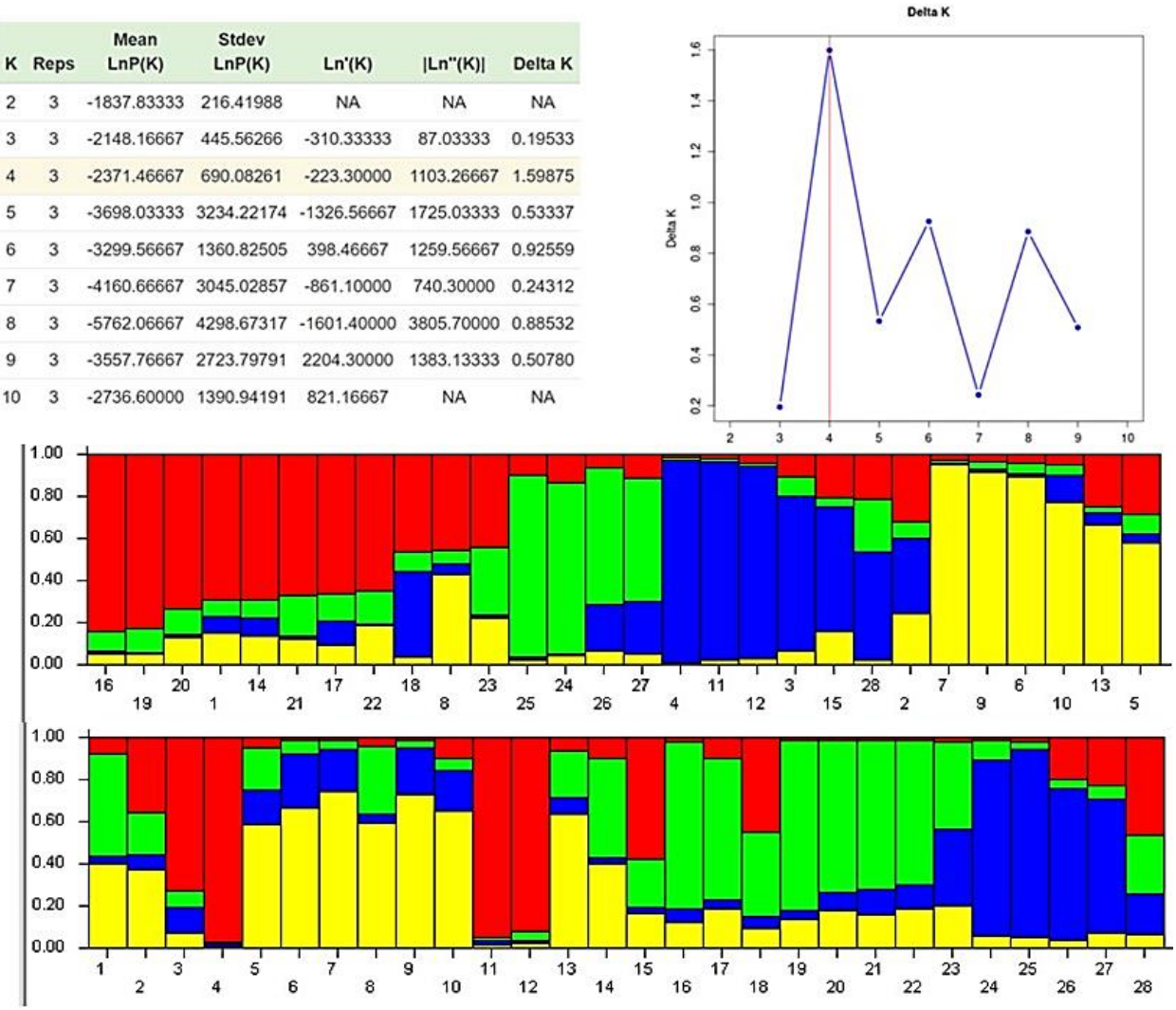

Figure 8- Graphical Representation of Subpopulation Membership Coefficients Obtained from STRUCTURE Software

The STRUCTURE program analyzes the structure of the population, dividing the individuals into subpopulations and creating a Q matrix. The number of subpopulations (K) was determined to be between 1 and 10, with 10 independent calculations performed for each K value. Subsequently, the number of subpopulations was identified using the Structure Selector program. Based on the K value obtained from the iPBS data, the hollyhock genotypes consist of four subpopulations (Figure 8). The analysis with the STRUCTURE program revealed the membership coefficients of the hollyhock genotypes for each population. Individuals with a membership coefficient of 0.80 or higher are considered pure and are a significant resource for breeding programs planned for hybridization. Those with a membership coefficient of 0.79 or lower are considered hybrids. The membership status of hollyhock genotypes to subpopulations is graphically presented in Figure 8.

3.4. Machine learning analysis results

The results in Table 6 provides a detailed analysis of the performance of three machine learning models-MLP, SVM, and RF-in predicting antioxidant activities (DPPH, TPC, CUPRAC, and FRAP) using three different datasets: phenotypic data (P), genotypic data (G), and combined phenotypic and genotypic data (PG). The models' performances are evaluated separately based on three metrics: coefficient of determination (R^2), root mean square error (RMSE) and mean absolute error (MAE).

Table 6- Result of machine learning models with different antioxidants

	R^2				*RMSE*				*MAE*			
	DPPH	***TPC***	***CUPRAC***	***FRAP***	***DPPH***	***TPC***	***CUPRAC***	***FRAP***	***DPPH***	***TPC***	***CUPRAC***	***FRAP***
P												
MLP	0.85	0.85	0.81	0.82	10.4	0.93	12.6	3.63	9.20	0.81	10.8	3.21
SVM	0.74	0.87	0.78	0.76	10.2	1.10	9.46	4.0	8.80	0.94	7.90	3.51
RF	0.87	0.94	0.79	0.83	12.3	0.93	10.4	4.28	10.8	0.81	8.86	3.90
G												
MLP	0.79	0.84	0.85	0.81	14.9	1.26	19.8	4.03	12.8	1.09	17.0	3.20
SVM	0.86	0.72	0.88	0.76	10.1	1.18	13.6	5.64	8.28	1.01	10.9	4.73
RF	0.90	0.73	0.94	0.91	6.42	0.90	9.85	2.45	5.18	0.81	8.15	2.11
PG												
MLP	0.83	0.81	0.85	0.96	14.8	1.50	24.9	7.32	11.8	1.32	22.2	6.12
SVM	0.87	0.85	0.91	0.75	10.5	1.03	13.7	5.63	8.52	0.84	11.3	4.68
RF	0.90	0.85	0.89	0.99	5.69	0.83	7.77	2.23	5.52	0.71	6.65	1.77

MLP:Multilayer Perceptron; SVM: Support Vector Machines; RF: Random Forest; R^2: Coefficient of Determination; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; P: The model trained using phenotypic data; G: the model trained using genotypic data; PG: The model trained using a combined phenotypic and genotypic data.

The R^2 values indicate how well the models explain the variance in antioxidant activities. For models trained on phenotypic data (P), the MLP model shows high R^2 values for DPPH and TPC (0.85 each), suggesting strong predictive accuracy for these antioxidants. However, the R^2 values for CUPRAC (0.81) and FRAP (0.82) are slightly lower, indicating moderate prediction accuracy. The RF model achieves the highest R^2 for TPC (0.94), demonstrating superior explanatory power compared to the other models. For models trained on genotypic data (G), the MLP model's R^2 values are lower for DPPH (0.79) and TPC (0.84), but it performs better for CUPRAC (0.85). SVM shows the highest R^2 for CUPRAC (0.88) and DPPH (0.86), outperforming MLP in these cases, while RF achieves the best R^2 for FRAP (0.91), reflecting its ability to effectively utilize genetic information. With combined data (PG), the MLP model achieves a very high R^2 for FRAP (0.96), indicating that integrating phenotypic and genotypic data enhances predictive accuracy for this antioxidant. SVM achieves the highest R^2 for CUPRAC (0.91), while RF consistently outperforms other models, particularly for FRAP (0.99) and DPPH (0.90), suggesting its robust-ness in capturing complex relationships (Figure 9).

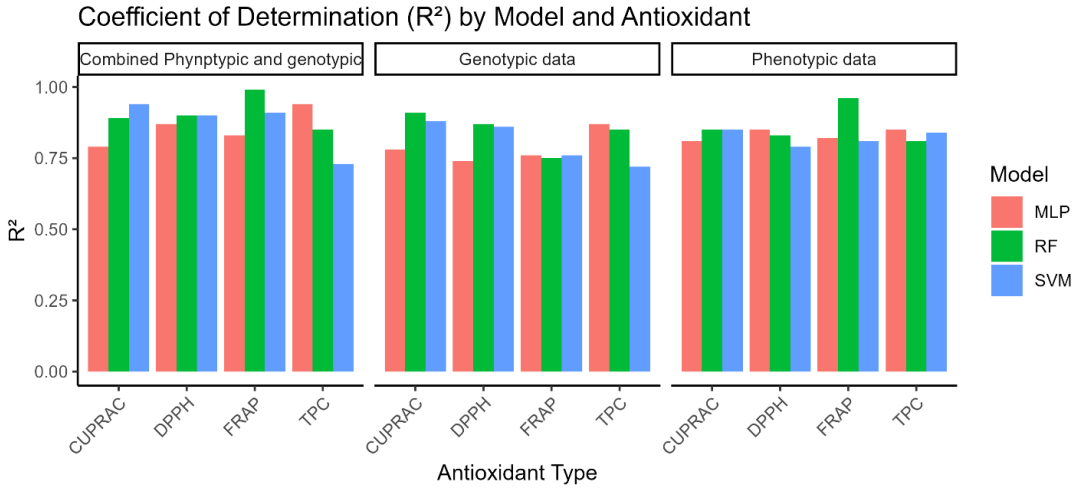

Figure 9- Comparison of MLP, RF, and SVM models in predicting antioxidant-related traits based on R^2 values. Higher R^2 values indicate stronger predictive performance, with RF generally showing the highest explanatory power across combined phenotypic and genotypic datasets.

Better model performance is shown by lower RMSE values, which quantify the average magnitude of prediction mistakes. For phenotypic data (P), the RF model exhibits the lowest RMSE for TPC (0.93) and FRAP (4.28), indicating high precision in its predictions. The SVM model shows a lower RMSE for DPPH (10.2) compared to MLP, which suggests better error handling despite a lower R^2. For genotypic data (G), the RF model again shows strong performance with the lowest RMSE for FRAP (2.45), demonstrating its ability to minimize prediction errors effectively. The SVM model exhibits a lower RMSE for DPPH (10.1), suggesting reasonable error management for this antioxidant. However, the MLP model shows higher RMSE values for CUPRAC (19.8) and FRAP (4.03), reflecting less accurate predictions. With combined data (PG), the RF model achieves the lowest RMSE values across most antioxidants, particularly for FRAP (2.23), confirming its superior predictive accuracy when using integrated data. The MLP model, however, shows a high RMSE for CUPRAC (24.9), indicating poor performance in this scenario (Figure 10).

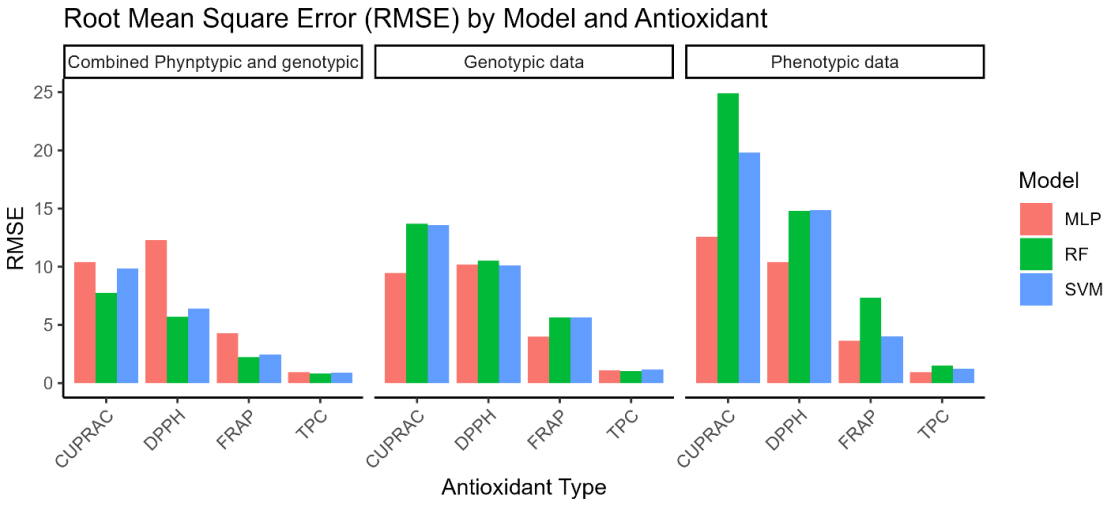

Figure 10- Comparison of MLP, RF and SVM in predicting different antioxidants using RMSE, Lower RMSE values indicate improved prediction accuracy

MAE provides an average measure of the errors between predicted and actual values. Lower MAE values represent more accurate predictions. For phenotypic data (P), the RF model has the lowest MAE for TPC (0.81) and FRAP (3.90), demonstrating the highest ac-curacy and precision in its predictions. The MLP model shows a lower MAE for FRAP (3.21) but higher errors for other antioxidants, particularly CUPRAC (10.8). For genotypic data (G), RF again outperforms with the lowest MAE for FRAP (2.11), highlighting its capacity to make accurate predictions using genetic data. The SVM model demonstrates lower MAE for DPPH (8.28), indicating its effectiveness for certain antioxidants, but shows higher MAE for FRAP (4.73). When using combined data (PG), RF consistently has the lowest MAE across most antioxidants, especially for FRAP (1.77), reinforcing its robust-ness and accuracy. The MLP model, however, has higher MAE for CUPRAC (22.2), suggesting less reliable predictions with combined data (Figure 11).

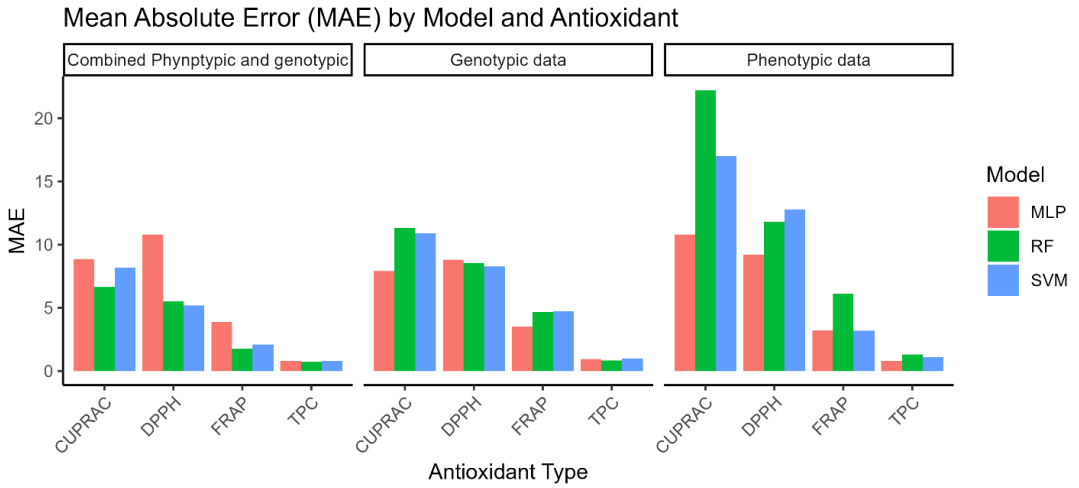

Figure 11- Comparison of MLP, RF and SVM in predicting different antioxidants using MAE, Lower MAE values represent smaller deviations between observed and predicted values

The differences observed among ML models indicate that algorithm selection substantially influences prediction performance in multidimensional biological datasets. The superior performance of RF in several antioxidant traits may be associated with its ability to model nonlinear interactions between phenotypic and molecular variables while maintaining robustness against overfitting. In contrast, the comparatively lower performance of some MLP models may reflect the limited sample size and the sensitivity of neural network architectures to smaller datasets. The superior predictive performance of the RF model may be attributed to its ensemble-based architecture, which combines multiple decision trees and effectively captures complex nonlinear interactions among phenotypic and genotypic variables. In addition, RF is known for its robustness against overfitting and noisy datasets, making it particularly suitable for relatively small but multidimensional biological datasets such as those used in the present study.

From a biological perspective, the improved predictive performance obtained from the combined phenotypic and genotypic dataset suggests that antioxidant activity in hollyhock is influenced by the interaction of multiple morphological, biochemical,

and genetic factors. Phenotypic traits such as plant architecture, leaf morphology, flower characteristics, and color-related parameters may reflect physiological processes associated with secondary metabolite accumulation, whereas the highly polymorphic iPBS markers represent underlying genetic variation among genotypes. The superior performance of the RF model indicates that antioxidant-related traits are governed by complex nonlinear relationships among these variables, highlighting the importance of integrating phenotypic and molecular information to better characterize the biological basis of antioxidant capacity in hollyhock.

Overall, the RF model consistently exhibits the best performance across all antioxidants and data types, particularly when using combined phenotypic and genotypic data, as demonstrated by the highest R^2 values and the lowest RMSE and MAE values. This suggests that RF is highly effective in capturing complex relationships within integrated datasets. Combining phenotypic and genotypic data generally enhances model performance, leading to lower prediction errors and higher explanatory power, underscoring the value of integrative approaches in predictive modeling of antioxidant activities in biological research. The 9-11 figures demonstrate these findings clearly, with RF showing the most consistent and reliable performance across different metrics.

4. Discussion

In this study, 28 endemic *Alcea rosea* genotypes collected from 11 provinces across Türkiye were comprehensively characterized. A wide phenotypic variation was observed across ten morphological traits, complemented by image-based color analysis of flowers and leaves. Biochemical profiling of dried flowers revealed significant diversity in total phenolic content and antioxidant activity, highlighting the pharmacological potential of these genotypes. Molecular marker data supported the morphological and biochemical findings, indicating strong genetic differentiation among populations. Genotypes with high phenotypic and biochemical variation also exhibited considerable genetic divergence. These results align with previous reports on *Alcea rosea*, offering an integrated characterization of the species. Furthermore, machine learning models were employed to successfully predict morphological, biochemical, and molecular traits from genetic data.

The identification of hollyhock genotypes is conducted based on their morphological characteristics. It has been noted that the yellow anthers of flowers in *Alcea* L. and the purple or brownish anthers of flowers in *Althea* L. enable the differentiation of the hollyhock species (Uzunhisarcıklı & Vural 2012). All species included in this study belong to the genus *Alcea* and exhibit high variation. Another study involving a hollyhock population of 162 individuals observed phenotypic variation in 11 morphological traits. This study reported high variation and found a significant positive correlation between some quantitative traits (Wang et al. 2023). Similarly, studies have indicated that hollyhock plants exhibit differences in leaf and flower characteristics, and the morphological traits of hollyhocks have been examined (Ghasemi & Atakishiyeva 2016). It has been noted that hollyhock flowers possess a wide color range, from white and yellow to deep red and purple hues (Ammar et al. 2013), and they have an extended flowering period (Fahamiya et al. 2016; Pullaiah 2006). The high level of morphological variability observed in the present study may therefore be attributed to the underlying genetic diversity within Alcea rosea populations, as well as possible adaptive responses to different ecological conditions across geographical regions. In addition, the congruence between our findings and previous studies supports the robustness of morphological traits as reliable indicators for genotype differentiation in hollyhock; however, such traits alone may be insufficient to fully resolve complex genetic relationships without complementary molecular and biochemical evidence.

Variations detected by morphological and biochemical markers are extremely valuable in breeding studies. In our study, different hollyhock genotypes were classified based on color values and various color channels were determined. Phenotypic classifications in horticultural plants provide guidance in selection. Studies show that color characteristics of agricultural products are directly related to their biochemical characteristics (Gunaydın et al. 2024). This study describes the relationship between the colors of hollyhock flowers and their total phenolic content and antioxidant properties. The wide variation in biochemical properties detected among genotypes is consistent with previous studies. In the study by Ammar et al. (2013), the antioxidant and phenolic contents of the *A. rosea* plant were determined, revealing that hollyhock plants are rich in phenolic compounds and, therefore, exhibit high antioxidant enzyme activity. Fersahoğlu (2016) reported that the total phenolic content of *A. rosea* varied among plants with different flower colors, with the lowest amount being determined as 27.73 mg GA g^{-1}. In another study, the total phenolic content was reported to be 7.92 mg GA g^{-1} (Kordalı et. al. 2021). In this study, antioxidant activity values were examined with three different methods. The activity values changing with different antioxidant methods have been explained in previous studies and it has been stated that DPPH is a color measurement method subject to various interferences (Vinha et al.2024). Some studies have reported that hollyhock genotypes contain pectin, flavonoids, mucilage, starch, sucrose, phenolic acids, asparagine, and tannins and that different flower colors are associated with different biochemical compositions. Consequently, plants with different flower colors possess distinct bioactive characteristics (Lim 2014; Munir et al. 2012; Fahamiya et al. 2016; Dalda-Sekerci 2023). Overall, the observed strong association between flower color variation and biochemical properties suggests that pigmentation-related metabolic pathways in hollyhock are closely linked with the accumulation of phenolic compounds and antioxidant capacity, indicating that color traits may serve as indirect selection criteria in breeding programs. However, the relatively complex and genotype-dependent nature of these relationships highlights the necessity of integrating molecular-level data to more accurately elucidate the underlying genetic mechanisms controlling bioactive compound synthesis.

The present study demonstrates that integrating molecular markers with phenotypic and biochemical traits provides a comprehensive framework for evaluating genetic diversity in hollyhock. The genetic similarity values obtained from iPBS markers ranged from 0.35 to 0.74, indicating substantial genetic variation among the studied genotypes despite the relatively limited and geographically restricted sampling. This level of diversity supports the existence of significant intraspecific variation within *Alcea rosea* populations. Similar patterns of genetic variation have been reported in previous studies employing different molecular marker systems. Wang et al. (2023) conducted whole-genome resequencing of 32 randomly selected *Alcea rosea* individuals and identified core single nucleotide polymorphisms (SNPs) that were evenly distributed across the genome, except for chromosome 21. The average nucleotide diversity was determined to be 0.00397 and it was found that these 32 participants could be divided into four subpopulations using principal component analysis. In another study, 156 hollyhock genotypes from 14 species from 16 regions of Iran were examined for genetic diversity using ISSR (Inter Simple Sequence Repeat) molecular markers and genetic similarity was reported to range from 0.63 to 0.99 (Cheng et al. 2021). Similarly, in a study conducted in Iran, genetic diversity of 34 *Alcea* species was assessed using RAPD molecular markers. The study revealed morphological and genetic diversity among and within species (Yin & Khodayari 2022). In another study conducted in Iran, the genetic similarity of 83 genotypes from 5 different *Alcea* species ranged from 0.77% to 0.89% using SCoT primers (Sun et al. 2022). In addition, analyses performed by dividing local populations into subpopulations provide advantages in terms of characterizing the current population based on genetic traits (Sumana et al. 2024). Overall, the level of genetic diversity detected in the present study, despite the relatively limited and geographically constrained sampling, is consistent with previous reports highlighting considerable intraspecific variation within *Alcea* species. This pattern may be attributed to the combined effects of geographic adaptation, gene flow among local populations, and the evolutionary history of the genus. However, the variability observed across different studies also reflects the influence of marker systems and sampling strategies, emphasizing that genetic diversity estimates should be interpreted in a methodological context. Therefore, integrating multiple marker systems with broader population sampling in future studies would be essential to more precisely elucidate the genetic structure and evolutionary dynamics of hollyhock populations.

The MLP model works without requiring any prior knowledge of the structure or connections between input and output signals (Araghinejad et al. 2017). As a result, such models have been recognized as valuable tools for modeling and optimization in the field of plant genetics, including tissue culture applications (Jamshidi et al. 2016; Aasim et al.2022; Eren et al.2023), and molecular marker studies (Sandhu et al. 2021). Previous research has demonstrated the potential of neural networks in genetic improvement, although the use of AI-based techniques in phenotypic and genetic predictions is still in the early stages. For example, Marini et al. (2004) used neural networks to accurately predict corn and soybean yields based on climatic and environmental factors, obtaining coefficients of determination of 0.81 for soybean and 0.77 for corn. Similarly, in a study conducted by Norouzi et al. (2024), rapeseed production was predicted using radical basis function neural networks, SVM and MLP's. Morphological, yield and molecular marker data from 56 hybrids and eight genotypes were used to train these models. In a comparative study of the models, it was stated that the root mean square error (RMSE), mean absolute error (MAE) and coefficient of determination (R^2) values were 226, 183 and 92%, respectively, and the MLP model performed better than the others in terms of predicting hybrid yield. Of the 40 primers evaluated, the ISJ10 primer showed the highest discriminatory power. The study also revealed the superiority of the MLP model when phenotypic and molecular data were used as inputs. In a study conducted by Rosado et al. (2020), a 90% increase in model accuracy was observed when phenotypic and genetic traits were taken into account in bean genetic prediction using artificial neural network-based MLP models. This method leverages quantitative traits to enhance prediction accuracy. These results further highlight how well neural networks predict crop performance, as do findings from other crop species (Shamsabadi et al. 2022; Eren et al. 2023; Hara et al.2023; Huang 2023). In the present study, the improved predictive performance obtained using combined phenotypic and genotypic datasets suggests that antioxidant-related traits in hollyhock are influenced by both observable morphological variation and underlying genetic structure. The integration of these datasets likely enabled the models to capture biologically relevant interactions associated with flower pigmentation, phenolic accumulation, and antioxidant metabolism. In particular, the higher prediction accuracy observed for FRAP and DPPH traits may indicate stronger associations between molecular marker profiles and biochemical antioxidant pathways. Despite these promising applications, the integration of machine learning approaches with molecular marker and phenotypic datasets in ornamental plant genetic studies remains limited, particularly in hollyhock, highlighting the need for more comprehensive predictive frameworks to improve genotype classification and breeding efficiency.

5. Conclusions

In this study, 28 hollyhock genotypes collected from 11 different provinces of Türkiye were examined. These areas are where the hollyhock genotypes used in the study naturally exist. Significant morphological differences were observed among the genotypes based on their morphological characteristics. Image processing analyses performed on leaves and flowers identified classifications based on color characteristics. A high variation was determined in bioactive properties, such as total phenolic content and antioxidant activity. In addition to all morphological and biochemical traits, genetic diversity was also attempted to be genetically determined using iPBS molecular markers. While high genetic variation was detected, genetic similarities were observed among genotypes sampled from the same region, whereas genotypes sampled from different provinces exhibited genetic differences. These genotypes, whose genetic similarities and diversities were determined based on specific plant characteristics, constitute a superior initial population of plants highly adapted to their natural distribution area. The superior agricultural traits of the hollyhock, a valuable ornamental and medicinal plant, such as high flowering tendency, diverse flower

colors, and high biochemical properties, can be transferred to cultivated varieties through breeding efforts. This study serves as a guide for future research. It provides important insights for future breeding programs that could be conducted on larger populations collected from broader geographical regions. Additionally, the relationship between morphological and biochemical traits and the genetic structure was predicted using machine learning methods. In this study, which revealed the genetic diversity of endemic Hollyhock genotypes through various markers, machine learning models were employed to predict phenotypic and biochemical traits influencing genetic diversity. This approach also highlights the potential of machine learning techniques in biotechnology applications and serves as a foundational study for future research. The integration of morphological, biochemical, molecular, and machine learning approaches provides a valuable framework for the characterization and selection of superior hollyhock genotypes with enhanced ornamental and bioactive properties. In particular, the successful application of machine learning models demonstrated the potential of data-driven approaches for predicting antioxidant-related traits and supporting future varietal selection strategies in ornamental and medicinal plant breeding programs. Nevertheless, further validation using larger populations and independent datasets will be necessary to confirm the robustness and generalizability of the machine learning models.

Author Declarations

Authors' contributions: Conceptualization, methodology, data curation, writing, visualization A.D.S

Conflict of interest/Competing interests: The author(s) declare no conflict of interest.

Ethics approval/declarations: No need/ Not applicable.

Funding: I would like to thank the Proofreading & Editing Office for their editorial help and the Office of the Dean for Research at Erciyes University for providing the necessary laboratory equipment and infrastructure at the ArGePark research building. Also, I would like to thank Erciyes University Scientific Research Projects Coordination Office for supporting this study with the project coded 2023-12462.

Consent to participate: No need/ Not applicable.

Consent for publication: No need/ Not applicable.

References

Aasim M, Ali S A, Aydin S, Bakhsh A, Sogukpinar C, Karatas M & Aydin M E (2023). Artificial intelligence–based approaches to evaluate and optimize phytoremediation potential of in vitro regenerated aquatic macrophyte *Ceratophyllum demersum* L. Environmental Science and Pollution Research 30(14): 40206-40217. https://doi.org/10.1007/s11356-022-25081-3

Aasim M, Katırcı R, Akgur O, Yildirim B, Mustafa Z, Nadeem M A & Yılmaz G (2022). Machine learning (ML) algorithms and artificial neural network for optimizing in vitro germination and growth indices of industrial hemp (*Cannabis sativa* L.). Industrial Crops and Products, 181: 114801. https://doi.org/10.1016/j.indcrop.2022.114801

Aktumsek A, Zengin G, Guler G O, Cakmak Y S & Duran A (2013). Antioxidant potentials and anticholinesterase activities of methanolic and aqueous extracts of three endemic *Centaurea* L. species. Food and Chemical Toxicology 55: 290-296. https://doi.org/10.1016/j.fct.2013.01.018

Ammar N M, El-Kashoury E S A, Abou El-Kassem L T & Abd El-Hakeem R E (2013). Evaluation of the phenolic content and antioxidant potential of *Althaea rosea* cultivated in Egypt. *Journal of The Arab Society for Medical Research*, 8(2): 48-52

Apak R, Güçlü K, Özyürek M, Karademir E S & Erçağ E (2006). The cupric ion reducing antioxidant capacity and polyphenolic content of some herbal teas. *International journal of food sciences and nutrition* 57(5-6): 292-304. https://doi.org/10.1080/09637480600798132

Araghinejad S, Hosseini-Moghari S M & Eslamian S (2017). Application of data-driven models in drought forecasting In: Esla-mian S (ed) Handbook of drought and water scarcity. CRC Press, New York, pp 423-440

Aydin M F & Baloch F S (2019). Exploring the genetic diversity and population structure of Turkish common bean germplasm by the iPBS-retrotransposons markers. *Legume Research-An International Journal* 42(1): 18-24 10.18805/LR-423

Bahmani M, Zargaran A & Rafieian-Kopaei M (2014). Identification of medicinal plants of Urmia for treatment of gastrointestinal disorders. Revista Brasileira de Farmacognosia 24(4): 468-480. https://doi.org/10.1016/j.bjp.2014.08.001

Baloch F S, Alsaleh A, de Miera L E S, Hatipoğlu R, Çiftçi V, Karaköy T & Özkan H (2015). DNA based iPBS-retrotransposon markers for investigating the population structure of pea (*Pisum sativum*) germplasm from Turkey. Biochemical Systematics and Ecology 61: 244-252. 10.1016/j.bse.2015.06.017

Barut M, Nadeem M A, Karaköy T & Baloch F S (2020). DNA fingerprinting and genetic diversity analysis of world quinoa germplasm using iPBS-retrotransposon marker system. *Turkish Journal of Agriculture and Forestry* 44(5): 479-491. https://doi.org/10.3906/tar-2001-10

Benelli C, Faraloni C, İzgü T, Şimşek Ö & Tarraf W (2025). Optimizing micropropagation of *Tanacetum balsamita* L.: A machine learning approach to compare semisolid media and temporary immersion system. Horticulturae 11(10): 1173. https://doi.org/10.3390/horticulturae11101173

Blumenthal M, Goldberg A & Brinckmann J (2000). Herbal medicine. Expanded commission E monographs, 519 pp

Bradley P R (1992). British Medical Compendium, Vol. 1. London: British Herbal Medical Association pp. 71-72

Brand-Williams W, Cuvelier M E & Berset C L W T (1995). Use of a free radical method to evaluate antioxidant activity. LWT-Food science and Technology 28(1): 25-30. 10.1016/S0023-6438(95)80008-5

Cheng J, Hu D, Liu Y, Zhang Z & Khayatnezhad M (2021). Molecular identification and genetic relationships among *Alcea* (Malvaceae) species by ISSR Markers: A high value medicinal plant. Caryologia 74: 65-75

Cullen J (1967). *Alcea* L. Flora of Turkey and the East Aegean Islands 2: 411-419

Cetin N, Ropelewska E, Noutfia Y & Günaydın S (2024). Recognition performance of different artificial neural networks for distinguishing banana slices subjected to different combinations of pretreatment and microwave drying. Food Control 163: 110488. https://doi.org/10.1016/j.foodcont.2024.110488

Cetin N, Ropelewska E & Sabanci K (2023). The innovative approach to the assessment of differences in image textures between windfall apple samples dried using non-thermal and thermal techniques without and with ultrasound pretreatment. Expert Systems with Applications 232:120917. 10.1016/j.eswa.2023.120917

Dalda-Sekerci A (2023). Comprehensive assessment of genetic diversity in chrysanthemum germplasm using morphological, biochemical and retrotransposon-based molecular markers. Genetic Resources and Crop Evolution 70(8): 2321-2336. https://doi.org/10.1007/s10722-023-01634-4

Doyle J J & Doyle J L (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical bulletin 19:11-15

Elmastas M, Ozturk L, Gokce I, Erenler R & Aboul-Enein H Y (2004). Determination of antioxidant activity of marshmallow flower (*Althaea officinalis* L.). Analytical letters 37(9): 1859-1869. https://doi.org/10.1081/AL-120039431

Eren B, Türkoğlu A, Haliloğlu K, Demirel F, Nowosad K, Özkan G & Bocianowski J (2023). Investigation of the influence of polyamines on mature embryo culture and DNA methylation of wheat (*Triticum aestivum* L.) using the machine learning algorithm method. Plants 12(18): 3261. https://doi.org/10.3390/plants12183261

Fahamiya N, Shiffa M & Aslam M (2016). A comprehensive review on *Althaea rosea* Linn. *Journal of Pharmaceutical Research* 6(11): 6888-6894

Falush D, Stephens M & Pritchard J K (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164(4): 1567-1587. https://doi.org/10.1093/genetics/164.4.1567

Ferreira da Vinha A, Silva C S, Vilela A, Ferreira J, Medeiros R & Cerqueira F (2024). Potential of portuguese viticulture by-products as natural resources of bioactive compounds-antioxidant and antimicrobial activities. Applied Sciences 14(14): 6278. https://doi.org/10.3390/app14146278

Fersahoğlu H (2016). Farklı renklerdeki gülhatmi çiçeklerinin biyoaktif özellikleri. Yıldız Technical University Institute of Science and Technology Master's Thesis, pp: 54.

Ghasemi M & Atakishiyeva Y (2016). Investigation of the Antibacterial Effect of Native *Peganum harmala*, *Mentha pulegium* and *Alcea rosea* Hydro-alcoholic Extracts on Antibiotic Resistant *Streptococcus pneumoniae* and *Klebsiella pneumonia* Isolated from Baku, Azerbaijan. Infection Epidemiology and Microbiology 2(2): 12-14. 0.7508/iem.2016.02.004

Gramazio P, Plesa I M, Truta A M, Sestras A F, Vilanova S, Plazas M & Sestras R E (2018). Highly informative SSR genotyping reveals large genetic diversity and limited differentiation in European larch (*Larix decidua*) populations from Romania. *Turkish Journal of Agriculture and Forestry* 42(3): 165-175. https://doi.org/10.3906/tar-1801-41

Guliyev N, Sharifova S, Ojaghi J, Abbasov M & Akparov Z (2018). Genetic diversity among melon (*Cucumis melo* L.) accessionsrevealed by morphological traits and ISSR markers. *Turkish Journal of Agriculture and Forestry* 42(6): 393-401. https://doi.org/10.3906/tar-1707-18

Gunaydın S, Ropelewska E, Sacilik K & Çetin N (2024). Exploration of machine learning models based on the image texture of dried carrot slices for classification. *Journal of Food Composition and Analysis* 129: 106063. https://doi.org/10.1016/j.jfca.2024.106063

Hara P, Piekutowska M & Niedbała G (2023). Prediction of pea (*Pisum sativum* L.) seeds yield using artificial neural networks. Agriculture 13(3): 661. https://doi.org/10.3390/agriculture13030661

Heywood V H (Ed.) (1978). Flowering plants of the world pp: 336

Hossein-Pour A, Haliloglu K, Ozkan G & Tan M (2019). Genetic diversity and population structure of quinoa (*Chenopodium quinoa* Willd.) using iPBS-retrotransposons markers. Applıed Ecology and Envıronmental Research 17(2):1899-1911

Huang Y (2023). Improved SVM-based soil-moisture-content prediction model for tea plantation. Plants 12(12): 2309. https://doi.org/10.3390/plants12122309

Hutchinson J (1973). The families of flowering plants (*Angiospermae*) Dicotyledones. Oxford Univ Press Oxford 1: 311-313

Isak M A, Bozkurt T, Tütüncü M, Dönmez D, İzgü T & Şimşek Ö (2024). Leveraging machine learning to unravel the impact of cadmium stress on goji berry micropropagation. Plos one 19(6): e0305111. https://doi.org/10.1371/journal.pone.0305111

Jamshidi S, Yadollahi A, Ahmadi H, Arab M M & Eftekhari M (2016). Predicting in vitro culture medium macro-nutrients composition for pear rootstocks using regression analysis and neural network models. Frontiers in plant science 7: 274. https://doi.org/10.3389/fpls.2016.00274

Kalendar R, Antonius K, Smýkal P & Schulman A H (2010). iPBS: a universal method for DNA fingerprinting and retrotransposon isolation. Theoretical and Applied Genetics, 121, 1419-1430. https://doi.org/10.1007/s00122-010-1398-2

Karik Ü, Nadeem M A, Habyarimana E, Ercisli S, Yildiz M, Yılmaz A & Baloch F S (2019). Exploring the genetic diversity and population structure of Turkish laurel germplasm by the iPBS-retrotransposon marker system. Agronomy 9(10): 647. https://doi.org/10.3390/agronomy9100647

Kırbay E, Kazaz S, Doğan Meral E & Dalda Şekerci A (2025). Hybridization Efficiency and Genetic Diversity in Cut Chrysanthemum: Integration of Morphological and iPBS Marker Analysis. Horticulturae 11(9): 1101. https://doi.org/10.3390/horticulturae11091101

Kianitalaei A, Feyzabadi Z, Hamedi S, Qaraaty M (2019). *Althaea officinalis* in Traditional Medicine and modern phytotherapy. *Journal of Advanced Pharmacy Education and Research* 9(2-2019): 154-161

Kordalı Ş, Bozhüyük A U, Beyzi E, Güneş A & Turan M (2021). Tıbbi Bitki Olarak Kullanılan *Malva sylvestris* L. ve *Alcea rosea* L. Türlerinin Antioksidant Enzim, Fenolik Madde ve Bitki Besin Element İçerikleri. *Journal of the Institute of Science and Technology* 11(1): 786-794.

Lim T K (2014). *Alcea rosea*. In Edible Medicinal and Non Medicinal Plants Volume 8 Flowers (pp. 292-299). Dordrecht: Springer Netherlands.

Marini F, Zupan J & Magrì A L (2004). On the use of counterpropagation artificial neural networks to characterize Italian rice varieties. Analytica chimica acta 510(2): 231-240. https://doi.org/10.1016/j.aca.2004.01.009

Nemli S, Kianoosh T & Tanyolac M B (2015). Genetic diversity and population structure of common bean (*Phaseolus vulgaris* L.) accessions through retrotransposon-based inter primer binding sites (iPBSs) markers. *Turkish Journal of Agriculture and Forestry* 39(6): 940-948. https://doi.org/10.3906/tar-1505-59

Norouzi M A, Ahangar L, Payghamzadeh K, Sabouri H & Sajadi S J (2024). Investigation of genetic diversity of different spring rapeseed (Brassica napus L.) genotypes and yield prediction using machine learning models. Genetic Resources and Crop Evolution 1:14. https://doi.org/10.1007/s10722-024-01915-6

Ozsoy H, Sekerci A D & Simsek O (2023). Evaluation of Genetic Diversity in Daisy Species Based on Morphological and Molecular Markers. *International Journal of Agricultural and Natural Sciences* 16(3): 322-330. 10.5281/zenodo.1043422

Pritchard J K, Stephens M & Donnelly P (2000). Inference of population structure using multilocus genotype data. Genetics 155(2): 945-959. https://doi.org/10.1093/genetics/155.2.945

Pullaiah T (2006). Encyclopaedia of world medicinal plants. Daya Books, Vol 1.

Riedl I (1976). Malvaceae in Rechinger KH (ed.): Flora Iranica pp:120.

Rosado R D S, Cruz C D, Barili L D, de Souza Carneiro J E, Carneiro P C S, Carneiro V Q & Nascimento M (2020). Artificial neural networks in the prediction of genetic merit to flowering traits in bean cultivars. Agriculture 10(12): 638. https://doi.org/10.3390/agriculture10120638

Salybekova N, Dalda-Sekerci A, Apushev A, Gurcan K, Bulut-Topbas F, Tutuncu M, Sarı O, Yusupov B, Celıkel F G & Turmetova G (2025). Assessing genetic diversity and population structure in *Tulipa* species from Türkiye and Kazakhstan. Notulae Botanicae Horti Agrobotanici Cluj-Napoca 53(1): 14252-14252. https://doi.org/10.15835/nbha53114252

Sandhu K S, Lozada D N, Zhang Z, Pumphrey M O & Carter A H (2021). Deep learning for predicting complex traits in spring wheat breeding program. Frontiers in Plant Science 11: 613325. https://doi.org/10.3389/fpls.2020.613325

Shamsabadi E E, Sabouri H, Soughi H & Sajadi S J (2022). Using of molecular markers in prediction of wheat (*Triticum aestivum* L.) hybrid grain yield based on artificial intelligence methods and multivariate statistics. *Russian Journal of Genetics* 58(5): 603-611. https://doi.org/10.1134/S102279542205009X

Singleton V L & Rossi J A (1965). Colorimetry of total phenolics with phosphomolybdic-phosphotungstic acid reagents. *American journal of Enology and Viticulture* 16(3): 144-158

Sumana S L, Liao Y, Zhang C, Jing X, Zhu J, Tang Y & Su S (2024). Genetic Diversity of the Common Black Carp Strain (*Cyprinus carpio* var. *baisenensis*). Diversity 16(7): 413. https://doi.org/10.3390/d16070413

Sun Y, Jiang H, Zeng F, Pan X, Wu X, Qi Y & Wu X (2022). Species identification and genetic diversity of *Alcea* (Malvaceae) using SCoT molecular markers: medicinal plant. Genetika 54(1): 369-378. https://doi.org/10.2298/GENSR2201369S

Simsek O (2024). Machine learning offers insights into the impact of in vitro drought stress on strawberry cultivars. Agriculture 14(2): 294. https://doi.org/10.3390/agriculture14020294

Simsek, O, Dalda Sekerci A, Isak M A, Bulut F, Izgü T, Tutuncu M & Donmez D (2024). Optimizing micropropagation and rooting protocols for diverse lavender genotypes: A synergistic approach integrating machine learning techniques. Horticulturae 10(1):52 https://doi.org/10.3390/horticulturae10010052

Tabarsa M, Anvari M, Joyner H S, Behnam S & Tabarsa A (2017). Rheological behavior and antioxidant activity of a highly acidic gum from *Althaea officinalis* flower. Food Hydrocolloids 69: 432-439. https://doi.org/10.1016/j.foodhyd.2017.02.009

Tahmouzi S, Nejat M R S (2020). New infertility therapy effects of polysaccharides from *Althaea officinalis* leaf with emphasis on characterization, antioxidant and anti-pathogenic activity. International journal of biological macromolecules 145: 777-787. 10.1016/j.ijbiomac.2019.12.224

Tarraf W, İzgü T, Şimşek Ö, Cicco N & Benelli C (2024). Saffron in Vitro Propagation: An Innovative Method by Temporary Immersion System (TIS), Integrated with Machine Learning Analysis. Horticulturae 10(5): 454. 61. https://doi.org/10.3390/horticulturae10050454

Ulbricht C, Basch E, Ulbricht C, Hammerness P & Vora M (2003). Marshmallow (*Althaea officinalis* L.) monograph. Journal of Herbal Pharmacotherapy 3(3): 71-81. https://doi.org/10.1080/J157v03n03_08

Uzunhisarciklı M E & Vural M (2012). The taxonomic revision of *Alcea* and *Althaea* (Malvaceae) in Turkey. *Turkish Journal of Botany* 36(6): 603-636. https://doi.org/10.3906/bot-1108-11

Van Dijk A D J, Kootstra G, Kruijer W & de Ridder D (2021). Machine learning in plant science and plant breeding. Iscience 24(1): 101890

Wang Y, Zhao S, Chen P, Liu Y, Ma Z, Malik W A & Chang Y (2023). Genetic diversity and population structure analysis of Hollyhock (*Alcea rosea* Cavan) using high-throughput sequencing. Horticulturae 9(6): 662. https://doi.org/10.3390/horticulturae9060662

Yaman M, Palaz E B, Isak M A, Demirel S, İzgü T, Adalı S & Popescu M (2025). Integrating in vitro propagation and machine learning modeling for efficient shoot and root development in Aronia melanocarpa. Horticulturae, 11(8): 886. https://doi.org/10.3390/horticulturae11080886

Yildiz M, Koçak M, Nadeem M A, Cavagnaro P, Barboza K, Baloch F S & Keleş D (2020). Genetic diversity analysis in the Turkish pepper germplasm using iPBS retrotransposon based markers. *Turkish Journal of Agriculture and Forestry* 44(1): 1-14. https://doi.org/10.3906/tar-1902-10

Yin J & Khodayari H (2022). RAPD marker to assess genetic diversity and exchange genes in *Alcea* (Malvaceae). Genetika 54(2): 565-574. https://doi.org/10.2298/GENSR2202565Y

Zohary M (1963). Taxonomical studies in *Alcea* of south-western Asia. Part I. Bulletin of the Research Council of Israel 11: 210-229

Copyright © 2026 The Author(s). This is an open-access article published by Faculty of Agriculture, Ankara University under the terms of the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium or format, provided the original work is properly cited.