



RESEARCH ARTICLE

Morpho-nutritional characterization and molecular diversity assessment of pumpkin using SSR and SRAP markers

Deepak D Sharma¹, Chirag P Chandramaniya^{2*}, Sushil Kumar³, Hetal P Patel³, D J Parmar⁴, Ankit Chaudhary⁵ & Trivima Sharma¹

¹K. K. Institute of Agricultural Sciences and Research, Ganpat University, Mehsana 384 012, Gujarat, India

²Department of Genetics and Plant Breeding, Anand Agricultural University, Anand 388 110, Gujarat, India

³Department of Agricultural Biotechnology, Anand Agricultural University, Anand 388 110, Gujarat, India

⁴Department of Agricultural Statistics, Anand Agricultural University, Anand 388 110, Gujarat, India

⁵Kishorbhai Institute of Agriculture Sciences and Research Center, Uka Tarsadia University, Surat 394 350, Gujarat, India

*Correspondence email - chiragpatel91745@gmail.com

Received: 26 April 2025; Accepted: 11 December 2025; Available online: Version 1.0: 09 February 2026

Cite this article: Deepak DS, Chirag PC, Sushil K, Hetal PP, Parmar DJ, Ankit C, Trivima S. Morpho-nutritional characterization and molecular diversity assessment of pumpkin using SSR and SRAP markers. Plant Science Today (Early Access). <https://doi.org/10.14719/pst.9076>

Abstract

Pumpkin (*Cucurbita moschata* Duch. Ex Poir.) is incredibly useful and nutritionally rich vegetable crop having numerous industrial uses regarding seed, flesh and flesh flour, but it is still underutilized in India. Hence, it is necessary to introduce some potential selections with high yield and nutrition content. The present investigation was elucidated the morpho-nutritional potential of pumpkin for 28 morpho-biochemical characters estimated and assessment of molecular diversity using Simple Sequence Repeats (SSR) and Sequence-Related Amplified Polymorphism (SRAP) marker in 34 diverse genotypes of pumpkin collected from different regions of India. The study was conducted at the Main Vegetable Research Station, Anand Agricultural University (AAU), Anand, during the kharif season of 2018. PCA explained 82.72 % total variation across traits, while multi-trait genotype-ideotype distance index (MGDI) identified three high-performing genotypes; Anand Pumpkin 1, GPPK 95 and GPPK 59. A set of five SSR and SRAP polymorphic primers were used to estimate genetic diversity among the genotypes. The similarity matrix generates dendrogram with UPGM based on Jaccard's coefficient implemented in NTYSIS. The clustering grouped 34 genotypes into six main clusters viz. I, II, III, IV, V and VI with 25, 4, 1, 1, 2 and 1 genotypes, respectively. The maximum genetic distance (0.75) was recorded between the genotype pairs GPPK 59 and Arka Chandan, as well as GPPK 90 and Arka Chandan. These findings highlight the potential of specific genotypes for breeding programs aimed at enhancing yield and nutritional value in pumpkin.

Keywords: dendrogram; MGDI; molecular diversity; morpho-nutrition; pumpkin; SRAP; SSR

Introduction

Pumpkin (*Cucurbita moschata* Duch. Ex Poir.) is a diploid plant with $2n = 2x = 40$, a highly nutritious but underutilized vegetable crop. Primary centers of origin are possibly Northern and Southern America (1). The genus *Cucurbita* consists of 27 species of which, five are cultivated viz., *C. maxima*, *C. moschata*, *C. pepo*, *C. ficifolia* and *C. mixta*. Among these species, *C. moschata* is the most widely cultivated species and found to be cross compatible with *C. pepo*, *C. mixta* and *C. maxima* (2).

Pumpkins are often regarded as remarkable wonders of the vegetable kingdom due to their diverse and striking characteristics (3). In India total area under cultivation is 94000 ha and production 2043000 million tons. Every part of the pumpkin fruit is having significant uses; the mature and immature fruits are used as vegetable, fully matured fruits used for preparing candy or fermented into beverages, sweets, supplement cereal flours in bakery products, sauces, soups, spices, instant noodles, flour mixes and natural coloring agent in pasta. Sweet delicacies such as "Halwa" various confections and jams are made using the mashed

pulp of fully ripened fruit. Pulp is also mixed with tomato sauce and ketchup (4), glucose tolerance factor (GTF) pumpkin milk powder (5) which can be used as a diabetic food and for the preparation of pumpkin ice-cream. Its young leaves, flowers and tender shoots are also utilized as cooked vegetables in various culinary preparations.

Pumpkin is a nutritious food, providing a moderate amount of energy and carbohydrates (5.31 %), along with protein (0.98 %). It is also a rich source of vitamins, particularly carotenoid pigments (171 µg/100g) and essential minerals (6). Additionally, phytochemicals such as trigonelline and nicotinic acid, extracted from pumpkin, have been shown to help lower blood cholesterol and glucose levels. Pumpkin flour can be used to supplement the conventional flour contain nutrients and minerals in concentrated form carbohydrate (72.41 %), protein (7.81 %) carotenoid pigments (272 µg/100g) compared to the fruit as such. Pumpkin seeds are a valuable source of nutrients, containing 40–50 % oils, 30 % proteins, 22 % carbohydrates, along with essential minerals and vitamins (6, 7). This vegetable holds significant potential in addressing malnutrition, particularly in meeting vitamin A requirements (8).

Any breeding program begins with an assessment of the genetic diversity within the available germplasm, which serves as the foundation for developing new and improved varieties. It is said that genetic variability is the “sine quanon” of any such programme. A higher degree of variability within a population increases the likelihood of effective selection for desirable traits. Direct selection based on fruit yield performance may not be highly effective; however, selecting for yield-related component traits has been found to be a more reliable approach, as observed in other plant species (9). As a result, multivariate analytic methods such as principal component analysis (PCA) and MGDI can be used as a model instrument for testing and identifying the causes of variance (10, 11). PCA, for example, reduces the dimensionality of a data set by decreasing the number of variables while preserving as much information as possible. It applies an orthogonal transformation to convert a set of potentially correlated variables into a set of uncorrelated variables, referred to as principal components. Breeders frequently seek to generate an ideotype, which is a genotype that combines many traits for optimal performance. The goal of ideotype design is to improve crop performance by taking into account multiple attributes at the same time while selecting genotypes (12).

The analysis of genetic diversity and kinship between or within different species, populations and individuals is a precondition towards effective utilization and protection of plant genetic resources (13). Simple Sequence Repeats (SSR) are generally most reliable and highly reproducible among molecular markers. Certainly, SSRs are now extensively acknowledged as the foundation for many framework linkage maps. This marker system has played a crucial role even in merging linkage maps, since they define specific locations in the genome unequivocally (14). Recently developed Sequence-Related Amplified Polymorphism (SRAP) markers are found to be robust, technically less demanding, highly variable and easy to use (15). Taking in mind the importance of pumpkin crop and to generate more information on above stated aspects, the present investigation was undertaken for the estimation of morphological and nutritionally important biochemical traits from fruit pulp, seed, pulp flour and assessment of genetic divergence using molecular marker system.

Materials and Methods

Experimental material and field evaluation

The present investigation was carried out in well drained sandy loam soil at Main Vegetable Research Station, Anand Agricultural University (AAU), Anand during *kharif* season of the year 2018. The experimental material consists of 34 diverse genotypes of pumpkin (Supplementary Table S1). Evaluation was carried out in three replications in randomized complete block design (RBD) with 10 plants/genotype in each replication with Inter- and intra-row space of 2.0 × 1.0 m, respectively.

Phenotyping for morphological and biochemical traits

Data was collected for 14 morphological (Number of First Male Flower, Node Number of First Female Flower, Main Vine Length (m), Fruit Yield per Vine (kg), Average Fruit Weight (kg), Number of Fruits per Vine, Polar Circumference of Fruit (cm), Equatorial Circumference of Fruit (cm), Flesh Thickness (cm), Number of Seeds per Fruit, Seed Weight per Fruit (g) and Seed Index (g)) and 14 biochemical (Soluble Sugar content from Pulp (%), β -Carotene

content from Pulp (mg/100g), Ascorbic Acid content from Pulp (mg/100g), Free Amino acid content from Seed (%), True Protein content from Seed (%), Free Fatty Acid Content (%), Oil Content (%), Soluble Sugar Content from Pulp Flour (%), β -Carotene Content from Pulp Flour (mg/100g), True Protein Content from Pulp Flour (%), Zn content (mg/100g), Fe content (mg/100g), Mn content (mg/100g) and Cu content (mg/100g) traits from five competitive plants of each genotype in each replication. Soluble sugar content (%) from pulp and flour estimated by phenol sulphuric acid method (16), ascorbic acid content from pulp (mg/100g) determined by titrimetric method against KOH, β -carotene content (mg/100g) from pulp and flour determined using method described (17). True protein (%) from seed and flour, free amino acid content (%) from seed determined by colorimetric method (17), oil content (%), free fatty acid content (%) determined (18), micronutrients (Zn, Fe, Mn and Cu) was estimated in previous study (19).

DNA isolation and molecular analysis

The genomic DNA was isolated from leaf samples utilizing the Cetyl Trimethyl Ammonium Bromide (CTAB) extraction technique, a widely recognized and reliable method for plant-DNA extraction (20) of 34 genotypes. The concentration of the extracted genomic DNA was determined using a NanoDrop ND-1000 spectrophotometer (Software V.3.3.0, Thermo Scientific, USA), ensuring accurate quantification (Supplementary Table S2). Working DNA solution of 30 ng/ μ L TE buffer (10 mM Tris-HCl, pH 8.0 and 0.1 mM EDTA, pH 8.0) was prepared from the known quantity of stock DNA solution and stored at 4 °C. 25 SSR marker (Supplementary Table S3) and 30 SRAP marker (Supplementary Table S4) were used for PCR amplification. For amplification 15 μ L of reaction mixture containing genomic 1.5 μ L DNA, 7.5 μ L Mater Mix (2x Genei, Bangalore, India) and 1 μ L of 10 pMol primer (0.5 μ L each forward & reverse) and 5 μ L nuclease free water. PCR amplification was carried out in a PCR tubes of 200 μ L, for SSR marker the DNA amplification condition were as follows: initial denaturation of 94 °C for 5 min, then 35 cycles of 94 °C for 45 s, annealing at ΔT °C (specified primer) for 45 s, extension at 72 °C for 45 s and a final extension at 72 °C for 7 min; for SRAP marker the DNA amplification condition were as follows: initial denaturation of 94 °C for 5 min, then 5 cycles of 94 °C for 30 s, annealing at ΔT °C (specified primer) for 30 s, extension at 72 °C for 1 min, followed by 35 cycles of denaturation of 94 °C for 30 s, annealing at ΔT °C (specified primer) for 30 s, extension at 72 °C for 1 min and a final extension at 72 °C for 7 min in SensoQuest Thermocycler (Germany).

For separation and visualization of PCR products both agarose 3.5 % SSR and 2 % SRAP as well as 6 % non-denaturing polyacrylamide gels (PAGE) were used. The DNA fragments were detected with silver nitrate staining (21) and the gel was scanned under gel scanner. Polymorphism between the genotypes was observed based on length of amplified fragments in terms of number of base pairs by comparing with a 100 bp ladder/marker and the molecular diversity was worked out using NTSYS 2.02 platform.

Statistical analysis of morpho-biochemical and molecular data

The data collected for the traits studied were subjected to analysis of variance (ANOVA) and the critical difference (CD) was calculated to identify significantly different genotypes using the R software. PCA and biplot diagrams developed using GRAPES software (22).

The Multi-Trait Genotype-Ideotype Distance Index (MGIDI) was analysed using the R Package “metan” (23). Normalisation, Factor Analysis, Ideotype Planning and Computing Genotype Distance to Ideotype were the four procedures used to create the MGIDI index (12). Genetic similarity coefficients were computed using Jaccard’s similarity coefficient through the SIMQUAL function. Cluster analysis was conducted using the agglomerative approach with the UPGMA (Unweighted Pair Group Method with Arithmetic Mean) method, implemented via the SAHN clustering function in NTSYS version 2.02. The relationships among pumpkin cultivars were represented through a dendrogram and a genetic similarity matrix.

Results and Discussion

Analysis of variance

The ANOVA results indicated that the mean sum of squares for genotypes was significant across all traits, suggesting substantial genetic variability among the genotypes (Table 1). This extensive variation provides plant breeders with ample opportunities to select superior and desirable genotypes for crop enhancement. The observed morphological and biochemical diversity within the studied germplasm (Table 1) can be effectively utilized in selection and breeding programs to develop high-yielding pumpkin varieties with improved nutritional value. Similar wide and significant variation among the characters in the genotypes studied (24-26).

Character variance analysis

The estimates of the mean, range and coefficient of variation for the various traits analyzed are presented in Table 1.

Mean performance of morphological traits

Descriptive mean value of morpho and biochemical traits were visualized using a box plot (Fig. 1). Fruit yield per plant varied from 0.93 to 6.85 kg (mean: 4.07 kg), with GPPK 115 (6.85 kg) yielding the highest, statistically at par with GPPK 105, GPPK 95, GPPK 143 and GPPK 18. Fruit weight ranged from 0.94 to 6.61 kg (mean: 3.66 kg), with GPPK 115 (6.61 kg) being the highest, comparable to GPPK 105 and GPPK 95. GPPK 56 had the earliest male flowering (40 days), statistically at par with GPPK 33, GP 141 and GPPK 95. GPPK 95 had the earliest female flowering (47.13 days), at par with Pusa Vishwas, Azad Pumpkin 1 and Arka Chandan. First male flower node ranged from 4.00 to 10.53 (mean: 6.65), lowest in Kashi Harit (4.00), at par with GPPK 18, Pusa Vishwas and Azad Pumpkin 1. First female flower node varied from 16.87 to 26.93 (mean: 20.92), with Kashi Harit (16.87) being the lowest, at par with GPPK 2, Arka Chandan and Pusa Vishwas. Main vine length ranged from 2.00 to 6.80 m (mean: 4.16 m), with GPPK 90 (6.80 m) being the longest. Fruits per plant ranged from 0.13 to 1.87 (mean: 0.84), highest at GPPK 113 (1.87), at par with GPPK 115. These findings align with previous studies (26-31). The maximum equatorial circumference was recorded in GPPK 115 (75.61 cm), statistically at par with AP 1, Azad Pumpkin 1, GPPK 69, GPPK 105, GPPK 143 and GPPK 201. The highest polar circumference was observed in GPPK 95 (73.85 cm), found statistically at par with genotypes GPPK 150, GPPK 59, GPPK 105, GPPK 30 and GPPK 56. Flesh thickness ranged from 1.35 to 4.69 cm (mean: 2.89 cm), with GPPK 115 (4.69 cm) having the highest, followed by Saras and GPPK 18.

The maximum number of seeds per fruit was recorded in GPPK 150 (496.27), statistically at par with GPPK 109, GPPK 113, GPPK 155 and GPPK 143, while GPPK 100 had the lowest (168.73). GPPK 109 (78.87 g) had the highest seed weight per fruit, found

Table 1. Estimation of mean performance of 28 traits in 34 pumpkin genotypes evaluated at Anand during kharif, 2018–19

Sr. no.	Characters	Mean performance					Mean square genotypes (DF = 33)
		Mean	Range	S.Em	CD (0.05)	CV (%)	
1	Fruit yield per vine	4.07	0.93–6.85	0.30	1.44	14.78	7.069**
2	Average fruit weight	3.66	0.94–6.61	0.19	0.53	14.86	1.635**
3	Days to opening first male flower	44.52	40.0–51.33	0.60	1.68	2.32	28.293**
4	Days to opening first female flower	53.41	47.13–64.20	0.19	2.58	2.97	57.272**
5	Node number of first male flower	6.65	4.0–10.53	0.31	0.89	8.19	5.935**
6	Node number of first female flower	20.92	16.87–26.93	0.58	1.65	4.83	26.034**
7	Main vine length	4.16	2.0–6.80	0.15	0.41	6.05	3.540**
8	Number of fruits per vine	0.84	0.13–1.87	0.07	0.21	15.13	0.656**
9	Equatorial circumference of fruit	59.51	44.29–75.61	3.63	10.24	10.56	148.952**
10	Polar circumference of fruit	61.88	42.63–73.85	4.28	12.08	11.98	168.869**
11	Flesh thickness	2.89	1.35–4.69	0.09	0.25	5.30	1.735**
12	Number of seeds per fruit	360.35	168.33–496.27	23.08	65.18	11.10	22866.790**
13	Seed weight per fruit	45.18	17.36–78.87	3.86	10.91	14.81	866.619**
14	Seed index	12.41	5.73–21.33	0.54	1.53	7.55	30.473**
15	Soluble sugar content from pulp (%)	12.80	2.06–32.13	0.27	0.75	3.61	150.204**
16	Ascorbic acid content from pulp (mg/100g)	4.20	2.92–5.97	0.18	0.50	7.28	1.572**
17	β-carotene content from pulp (mg/100g)	1.90	1.67–3.15	0.04	0.13	4.09	0.273**
18	True protein content from seeds (%)	9.23	6.47–15.97	0.21	0.58	3.88	17.345**
19	Free amino acid content from seeds (%)	3.65	2.25–5.67	0.07	0.19	3.17	2.919**
20	Oil content (%)	31.61	14.2–41.10	0.74	2.08	4.03	112.826**
21	Free fatty acid content	0.85	0.60–1.43	0.02	0.07	4.99	0.113**
22	Soluble sugar content from flour (%)	33.30	20.24–40.84	1.40	3.94	7.26	117.459**
23	β-carotene content from flour (mg/100g)	4.25	3.34–6.44	0.04	0.11	1.56	2.731**
24	True protein content from flour (%)	11.46	8.39–14.31	0.26	0.72	3.88	5.677**
25	Fe content (mg/100g)	8.07	3.75–11.32	0.06	0.16	1.16	8.909**
26	Zn content (mg/100g)	6.43	2.82–11.14	0.06	0.18	1.70	17.213**
27	Mn content (mg/100g)	4.9	1.92–9.27	0.03	0.10	1.19	13.225**
28	Cu content (mg/100g)	1.76	0.64–4.09	0.02	0.05	1.72	2.027**

**Significant at 1 % level. S.Em: standard error of mean; CD: critical difference; CV: coefficient of variation; DF: degree of freedom.

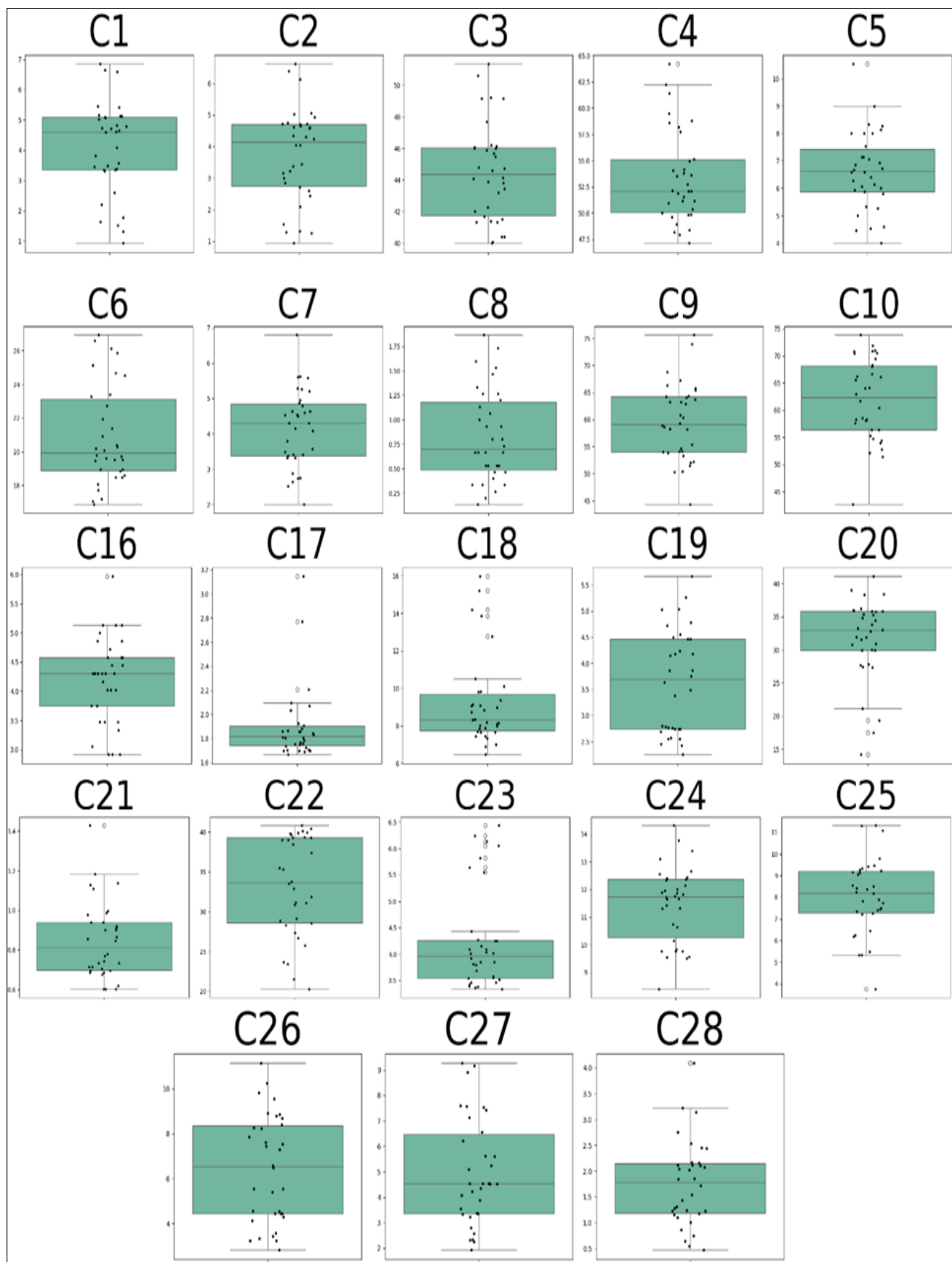


Fig. 1. Box plots displaying mean performance of the traits studied. C1: Fruit Yield per Vine (kg), C2: Average Fruit Weight (kg), C3: Days to Opening of First Male Flower, C4: Days to Opening of First Female Flower, C5: Node Number of First Male Flower, C6: Node Number of First Female Flower, C7: Main Vine Length (m), C8: Number of Fruits per Vine, C9: Equatorial Circumference of Fruit (cm), C10: Polar Circumference of Fruit (cm), C11: Flesh Thickness (cm), C12: Number of Seeds per Fruit, C13: Seed Weight per Fruit (g), C14: Seed Index (g), C15: Soluble Sugar content from Pulp (%), C16: Ascorbic Acid content from Pulp (mg/100g), C17: β -Carotene content from Pulp (mg/100g), C18: True Protein content from Seed (%), C19: Free Amino acid content from Seed (%), C20: Oil Content (%), C21: Free Fatty Acid Content (%), C22: Soluble Sugar Content from Pulp Flour (%), C23: β -Carotene Content from Pulp Flour (mg/100g), C24: True Protein Content from Pulp Flour (%), C25: Fe content (mg/100g), C26: Zn content (mg/100g), C27: Mn content (mg/100g) and C28: Cu content (mg/100g).

statistically at par with genotypes GPPK 95 and GPPK 105, whereas the lowest was in Saras (17.36 g). The highest seed weight per fruit was recorded at GPPK 100 (21.33 g), statistically at par with GPPK 95, while the lowest was in GPPK 50 (5.73 g). These findings align with previous studies (25, 26, 30, 32, 33) which reported similar trends in fruit circumference, flesh thickness, seed count and seed weight across different genotypes.

Mean performance of biochemical parameters

Pulp

Higher sugar content is desirable for pumpkins. The average soluble sugar content was 12.80 % and ranged from 2.06 % to 32.13 %. Arka Chandan had significantly highest sugar content (32.13 %) at par with GPPK 18 (31.17 %). Similar results were reported (24-26). The average ascorbic acid content was 4.20 mg/100g and ranged from 2.92 to 5.97 mg/100mg. The genotype AP 1 had significantly highest ascorbic acid (5.97 %). Similar results were reported (26, 34). A range from 1.67 to 3.15 mg/100g for β -carotene content was depicted with mean of 1.90 mg/100g. AP 1 (3.15 mg/100g) manifested the maximum β -carotene content. The wide variation observed in β -carotene content aligns with earlier reports (25, 26, 32, 35), which also documented significant diversity in carotenoid concentration among pumpkin accessions. The consistency of the present findings with previous research reinforces the role of inherent genetic factors in determining carotenoid biosynthesis and further suggests that high-carotene genotypes like AP 1 can be effectively exploited in varietal improvement to enhance the nutritional quality of pumpkin.

Seed

The average true protein content was 9.23 %, ranged from 6.47 % to 15.97 %. GPPK 148 had significantly highest protein content (15.97 %) and similar results were also reported (36-38). Free amino acid content (%) varied from 2.25 % to 5.67 % with mean of 3.65 %. Ambili had significantly highest protein content (5.67 %). Same result in pumpkin genotypes was also reported (4, 38, 39). Pumpkin seeds are good source of oil content (%) estimated with mean of 21.29 %, ranged from 14.20 % to 41.10 %. GPPK 95 reported with highest oil content (41.10 %) which was at par with GPPK 115 (39.03 %). Various Researchers (36-38, 40) also reported similar oil content but with narrower range. Lower value of free fatty acid content (%) is desirable for oil to be edible. The range recorded was from 0.60 % to 1.43 % with average of 0.85 %. The minimum free fatty acid content was found in GPPK 201, GPPK 115 and GPPK 48 (0.60 %), which was at par with GPPK 100 (0.62 %). Edible range of free fatty acid content was also reported (36-38, 40). The average Fe, Zn, Mn and Cu content were 8.26 mg/100g, 6.43 mg/100g, 4.98 mg/100g and 1.76 mg/100g respectively. The present findings are in close agreement with earlier studies (36, 38), which also reported comparable ranges for these micronutrients across diverse pumpkin accessions. Such consistency across studies suggests that the mineral composition of pumpkin is predominantly governed by genetic factors, with relatively stable expression across environments.

Pulp flour (%)

The average sugar content in pulp flour was 33.30 %, ranged from 20.24 % to 40.84 %. Azad Pumpkin 1 had highest content (40.84 %) significantly at par with Ambili (40.42%), Pusa Vishwas (40.15 %), Pusa Vikas (39.99%) and GPPK 143 (39.89) similar results were also reported (39). The average β -carotene content was 4.25 mg/100g, ranged from 3.34 to 6.44 mg/100g. The genotype AP 1 had significantly highest β -carotene content (6.44 mg/100g) results were

in accordance with (39). The average protein content from pulp flour was 11.46 %, ranged from 8.39 to 14.31 %. The genotype GPPK 139 had significantly highest protein content (14.31 %) which was at par with GPPK 56 (13.76%). The elevated protein content observed in these genotypes is consistent with earlier reports (36, 39), which similarly documented substantial genotypic variation for seed protein concentration in pumpkin. These findings underscore the strong genetic influence on protein accumulation and highlight the possibility of exploiting high-protein genotypes such as GPPK 139 and GPPK 56 for developing nutritionally enriched cultivars suited for consumer and industrial needs.

Principal component and biplot analysis

PCA, a sophisticated multivariate data analysis tool, was specifically utilized in this study to simplify and interpret complex, high-dimensional datasets. This method enabled the identification of key traits that contributed the most to overall variability, providing deeper insights into trait interactions. Among the 28 principal components (PCs), ten components exhibited Eigenvalues greater than 1, accounting for 82.72 % of the cumulative variability for the traits under investigation (Supplementary Table S5 and Fig. 2A).

The cumulative contribution rate was 82.72 %. Principal Component I (PC I) had an Eigenvalue of 7.11, contributing 25.39 % of the total variability. Germplasm in PC I had the most significant positive impact on fruit yield per vine, average fruit weight, number of fruits per vine, polar circumference of fruit, number of seeds per fruit, seed weight per fruit and seed index (Supplementary Table S5). Principal Component II (PC II) exhibited an Eigenvalue of 2.95, explaining 10.52 % of the variability. Germplasm lines exhibiting maximum positive PC scores and common presence in PC1 to PC10 are lines GPPK 113, GPPK 141, GPPK 105 and GPPK 115 (Supplementary Table S6). Selecting these lines can contribute significantly to the further development of new high yielding with good nutritional varieties. The \cos^2 (squared cosines or squared coordinates) values are used to assess the quality of variable representation on the factor map. A high \cos^2 value signifies a strong representation of the variable on the principal component, whereas a low \cos^2 value indicates that the variable is not well represented by the PCs (Fig. 2B).

The PC (1-2) biplot (Fig. 2C) illustrates trait variability, inter-trait correlations (positively associated characteristics ($<90^\circ$), independent attributes ($=90^\circ$) and negatively associated traits ($>90^\circ$)) and genotype dispersion. Most traits displayed relatively long vector lengths, except for soluble sugar content from pulp flour, true protein content from pulp flour, true protein content from seeds, free amino acid content from seeds, Zn content and ascorbic acid content from pulp suggesting significant variability. In character biplot for the fruit yield exhibited association with days to opening first female flower, average fruit weight, first female flowering node, number of fruits per vine, main vine length, polar circumference of fruit, equatorial circumference of fruit, flesh thickness, number of seed per fruit, seed weight per fruit, seed index, ascorbic acid and oil content as indicated by the very low angle between their corresponding lines (Fig. 2C). Fruit yield showed a marked negative correlation with soluble sugar from pulp, β -carotene flour, Zn, Cu, β carotene from pulp and total free fatty acid as indicated by the angle between their corresponding vectors being greater than 90° . The genotypes Arka Chandan, Kashi Harit, GPPK 126, GPPK 50, GPPK 148, GPPK 150, GPPK 115, GPPK 105 and GPPK 95 exhibited the highest diversity for various

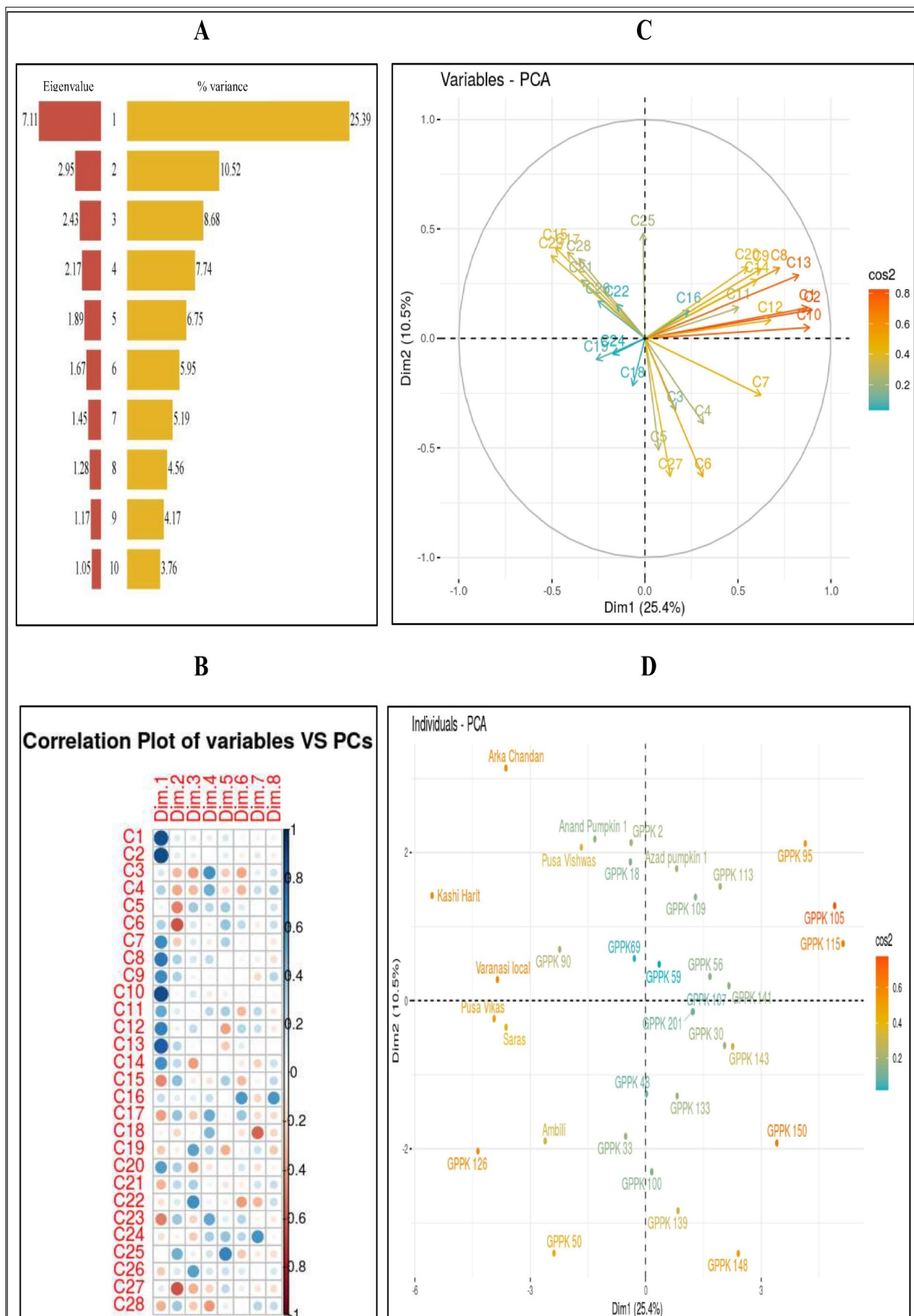


Fig. 2. (A) Depict butterfly bar charts show the variable percentage contribution of each principal component (PC) as well as the eigenvalue. (B) Quality of representation of different traits (cos²). (C, D) Biplots involving PC1 and PC2, illustrating the allocation of 28 traits and 34 genotypes, respectively. Scattered numbers across the plot indicate serial number of lines enlisted in Table 1 and Supplementary Table 1.

traits, as they were positioned far from the origin (Fig. 2D). These highly diverse genotypes have the potential to be valuable in future pumpkin improvement programs.

Selection of high yielding and good grain quality genotypes using MGIDI

There are various drawbacks to PCA that can make it difficult to pick high yielding genotypes. These include subjective interpretation, difficulties managing missing data, inadequate dimensionality reduction, the inability to consider interaction effects and lack of statistical rigor (41). To address these constraints, it is critical to incorporate additional analytical approaches. PCA can be integrated with quantitative indicators such as MGIDI to help identify short duration, high yielding with good grain quality genotypes. MGIDI is an ideal and innovative method for genotypic selection due to its ability to address multicollinearity and eliminate the need for assigning economic weights (23).

Selection of genotypes using MGIDI

The MGIDI index identified three genotypes Anand Pumpkin 1, GPPK 95 and GPPK 59 as high performing for multiple traits, demonstrating significant potential for simultaneously improving the 28 measured traits in pumpkin breeding programs (Fig. 3A). These genotypes were particularly notable for traits such as early flowering, high yielding with high nutritional quality traits. Among them G17, positioned near the cut-off point indicated by the red line, displayed intriguing characteristics warranting further investigation, as suggested (12). Successful applications of this selection index had been demonstrated in evaluating ideal yield and yield-related traits across various crops including wheat (42), brinjal (43) and guar (44). These different studies demonstrated the effectiveness of multivariate selection indices for simultaneous trait selection. MGIDI is the most efficient index for choosing genotypes with desirable features, demonstrating its relevance and usefulness in crop development (12). These selected derivatives serve as the foundation for establishing recombinant populations through judicious crossings, ensuring maximum genetic diversity for the breeding of novel pumpkin lines.

Strength and weakness

Fig. 3B illustrates the relative strengths and weaknesses of the examined genotypes, as determined by total of nine factors (FA1, FA2, FA3, FA4, FA5, FA6, FA7, FA8 and FA9) each factor's contribution to the MGIDI score for each genotype. MGIDI serves as a valuable graphical tool that highlights the strengths and weaknesses of genotypes, offering insights into how they perform in traits that require enhancement. A strength-weakness analysis revealed that FA1, FA2, FA3, FA4 and FA5 had the greatest influence on Anand

Pumpkin 1. FA6, FA7, FA8 and FA9 contributed most significantly to GPPK 95. FA1 FA3 and FA9 made the more contribution to GPPK 59. A similar methodology to evaluate the performance of 13 strawberry cultivars in earlier research (12). In a different study, a system using MGIDI to identify promising guar genotypes with high gum and seed yield over three seasons (44). Likewise, MGIDI is a powerful tool for enhancing selection methods in breeding climate-resilient maize hybrids, assessing their performance under varying moisture and drought conditions (45). The use of MGIDI in quinoa, focusing on different plant spacing strategies (46). In our study, MGIDI is applied to upland cotton, providing a comprehensive framework for identifying genotypes with both high yield and superior quality traits, which are well-suited for hybrid development. The detailed examination of strengths and weaknesses yielded useful insights, emphasising the importance of selecting the best rice genotype with superior quantitative traits. These selected genotypes stood out as promising candidates for future breeding projects, establishing MGIDI as a revolutionary technique for improving pumpkin varieties.

Marker polymorphism and genetic distance

In the present investigation the molecular diversity among 34 genotypes of pumpkin was studied using SSRs or microsatellite markers and SRAP molecular markers. A single sharp band was observed for isolated DNA for all 34 genotypes. The DNA extracted from pumpkin leaves had an average concentration of 1401.09 ng/μL, as quantified using a NanoQuant spectrophotometer. Eventually PCR reaction was carried out with 25 SSR and 30 SRAP primers in order to analyze the genetic diversity in pumpkin genotypes. Out of 25 SSR primers, 10 (40 %) were amplified successfully but only 5 primers (20 %) were recorded polymorphic and for 30 SRAP marker only 5 (16.66 %) gave proper and informative amplification and were polymorphic too (Table 2). All these polymorphic markers were eventually PCR amplified to analyze the genetic diversity among 34 pumpkin genotypes (Fig. 4 and 5).

The Polymorphism Information Content (PIC) values of markers serve as an indicator of their ability to differentiate among accessions by considering both the number of alleles and their relative frequencies (47). In the present study, a total of 38 loci were amplified, of which 35 (95.24 %) exhibited polymorphism. The PIC values ranged from 0.29 (CMTm80) to 0.85 (SRAP 7), with an average of 0.60, indicating a high level of genetic diversity. In this study, SRAP markers were found to be the most informative, as they demonstrated PIC values exceeding 0.5. The average PIC value (0.60) observed here is higher than that reported (48) for RAPD (0.46) and SSR (0.28) markers in *C. pepo*, as well as (49) for AFLP (0.53). However, these values are lower than those reported

Table 2. Characteristics of SSR and SRAP markers and amplified products used for genetic diversity analysis of 34 pumpkin genotypes evaluated at Anand During kharif 2018–19

Sr. No.	Locus Name	Total number of loci	Number of polymorphic loci	Percentage of polymorphism	PIC ^a
1.	CMTm11	2	2	100.00	0.45
2.	CMTm35	2	2	100.00	0.50
3.	CMTm64	2	2	100.00	0.39
4.	CMTm80	2	2	100.00	0.29
5.	CMTm112	2	2	100.00	0.42
6.	SRAP 7 (me2+em1)	7	6	85.71	0.85
7.	SRAP 8 (me2+em2)	4	4	100.00	0.72
8.	SRAP 19 (me4+em1)	6	5	83.33	0.83
9.	SRAP 20 (me4+em2)	5	5	100.00	0.78
10.	SRAP 25 (me5+em1)	6	5	83.33	0.82
	Total	38	35	-	-
	Average	3.8	3.5	95.24	0.60

^aPIC: polymorphism information content.

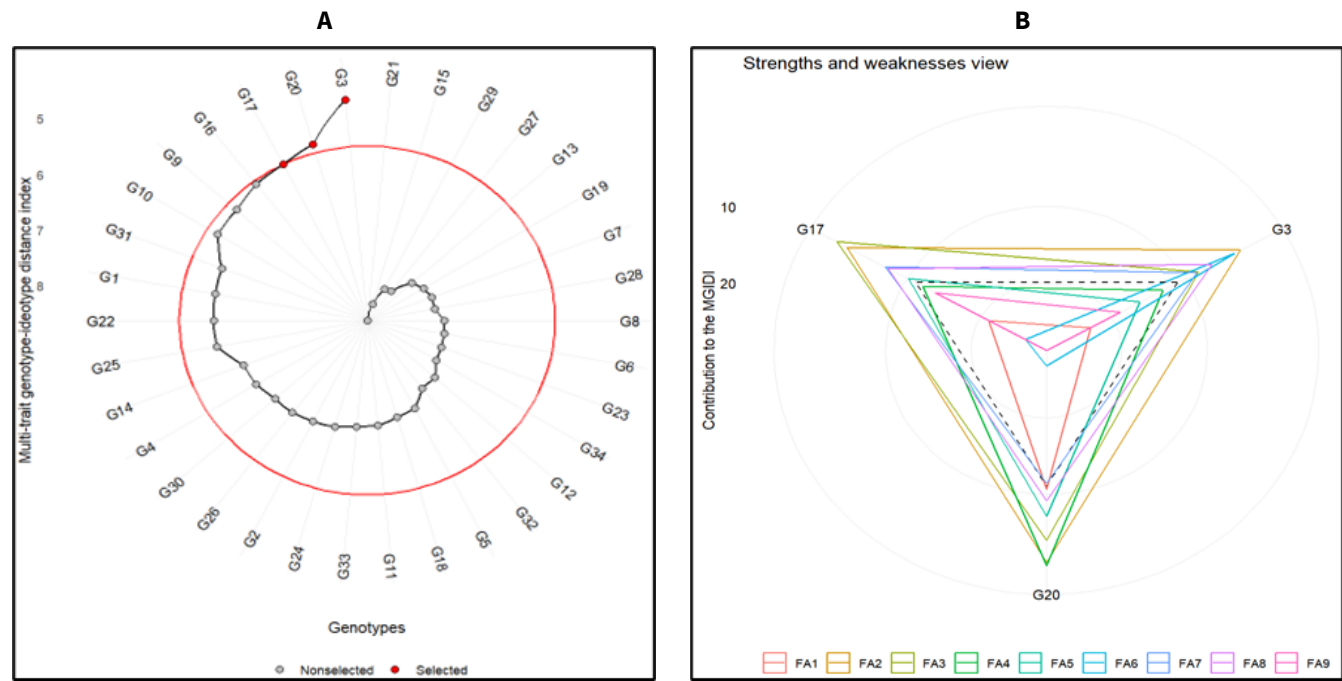


Fig. 3. The MGIDI analysis of lines ordering is presented in ascending order (A, B). The genotypes with the highest rankings and selection are highlighted in red. The central red circle indicates the cut-off point, determined by the selection pressure (A). The percentage contribution of each factor in the generated MGIDI index illustrates the strengths and weaknesses of each line (B). The closer a factor's indices are to the ideotype, the lower the fraction of explanation, indicating proximity to the outer boundary.

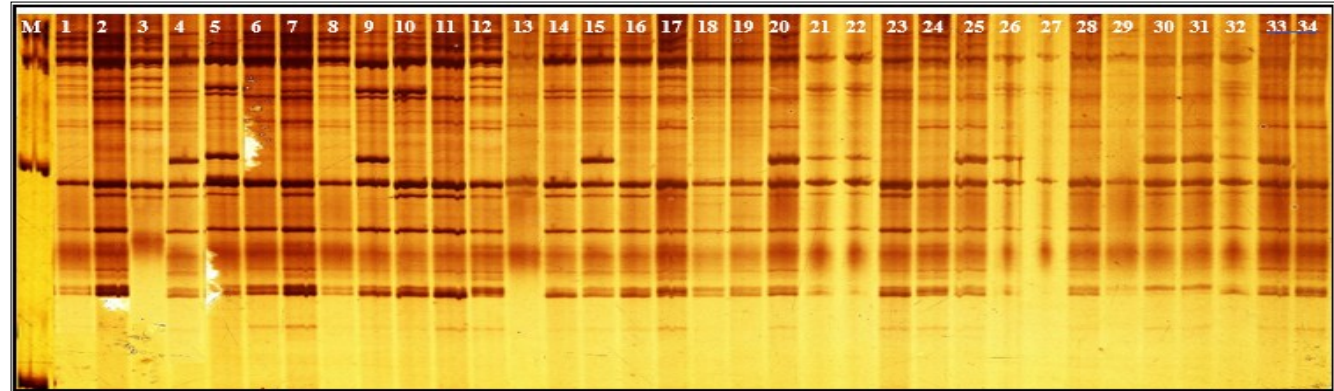


Fig. 4. SRAP profile of SRAP 7 (me2+em1) marker in 34 pumpkin genotypes.

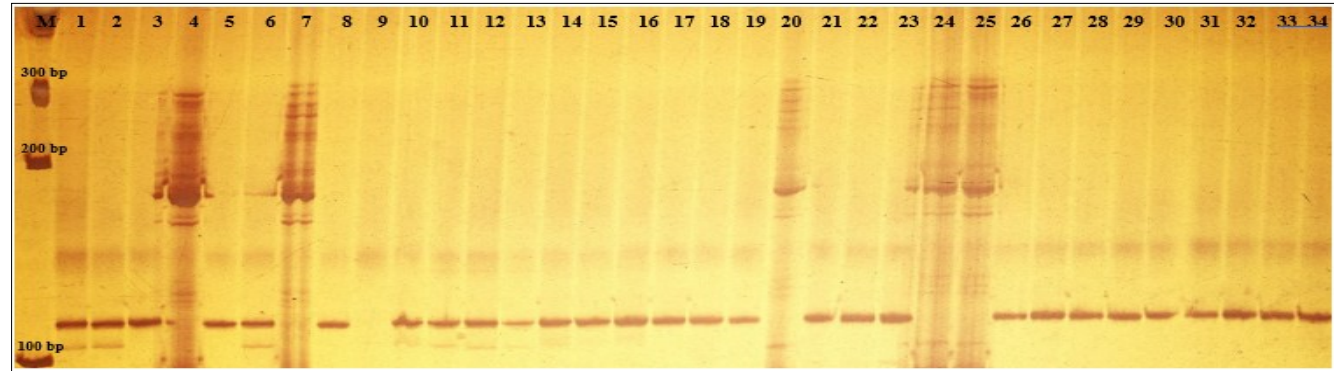


Fig. 5. SSR profile of CMTm64 marker in 34 pumpkin genotypes.

in previous studies for SRAP (0.73) (49) and for AFLP (0.63) and ISSR (0.74) (50). The relatively lower PIC value in this study could be attributed to the limited genotypic diversity among the 34 pumpkin genotypes analyzed. Additionally, the high PIC values and greater allele numbers per marker may also be influenced by the genetic composition of the materials studied (51).

The clustering analysis of accessions based on molecular data revealed that the 34 genotypes were grouped into six main clusters: I, II, III, IV, V and VI, containing 25, 4, 1, 1, 2 and 1 genotypes, respectively (Fig. 6). Main cluster I was further subdivided into three sub-clusters: A (14 genotypes), B (6 genotypes) and C (5 genotypes). Most accessions were grouped in sub-cluster A, suggesting a high degree of genetic similarity among these genotypes. Main cluster II comprised four genotypes, while Cluster V contained two genotypes. The remaining three clusters III, IV and VI each consisted of a single genotype. The Jaccard's similarity coefficient (Supplementary Table S6) among the genotypes varied from 0.25 to 1.00, with an average similarity coefficient of 0.60. The greatest genetic distance (0.75) was observed between the genotypes GPPK 59/Arka Chandan and GPPK 90/Arka Chandan, indicating substantial genomic divergence. This suggests that these genotypes could serve as promising parental lines for biparental mapping populations and genetic enhancement programs aimed at broadening the genetic base of pumpkin. Conversely, the lowest genetic distance (0.00) was recorded between GPPK 100 and GPPK 105, implying that these genotypes likely share a common genetic lineage.

Conclusion

Based MGIDI Anand Pumpkin 1, GPPK 95 and GPPK 59 were identified as elite genotypes and could be used in future breeding programmes for improving yield and nutritional content in pumpkin. The reported resultant molecular diversity can be used to produce high yielding varieties and hybrids, help in solving the emerging need to fight malnutrition in developing countries.

Acknowledgements

The authors acknowledge Anand Agricultural University, Anand for providing experimental land and other necessary facilities for conducting the experiment. We appreciate the facilities, resources and research assistance provided by the Department of Biotechnology, Anand Agricultural University, Anand, Department of Genetics and Plant Breeding, B. A. College of Agriculture, AAU, Anand and Main Vegetable Research Station. We extend our gratitude to all parties who supported this study and contributed to the preparation of this article.

Authors' contributions

DDS, CPC and TS contributed to the writing, including review and editing, as well as the primary drafter of the manuscript. SK and DJP were responsible for design the overall study, set the methodology and supervision. DJP contributed to the data analysis. HPP and AC was involved in the figure and table development. All authors read and approved the final manuscript.

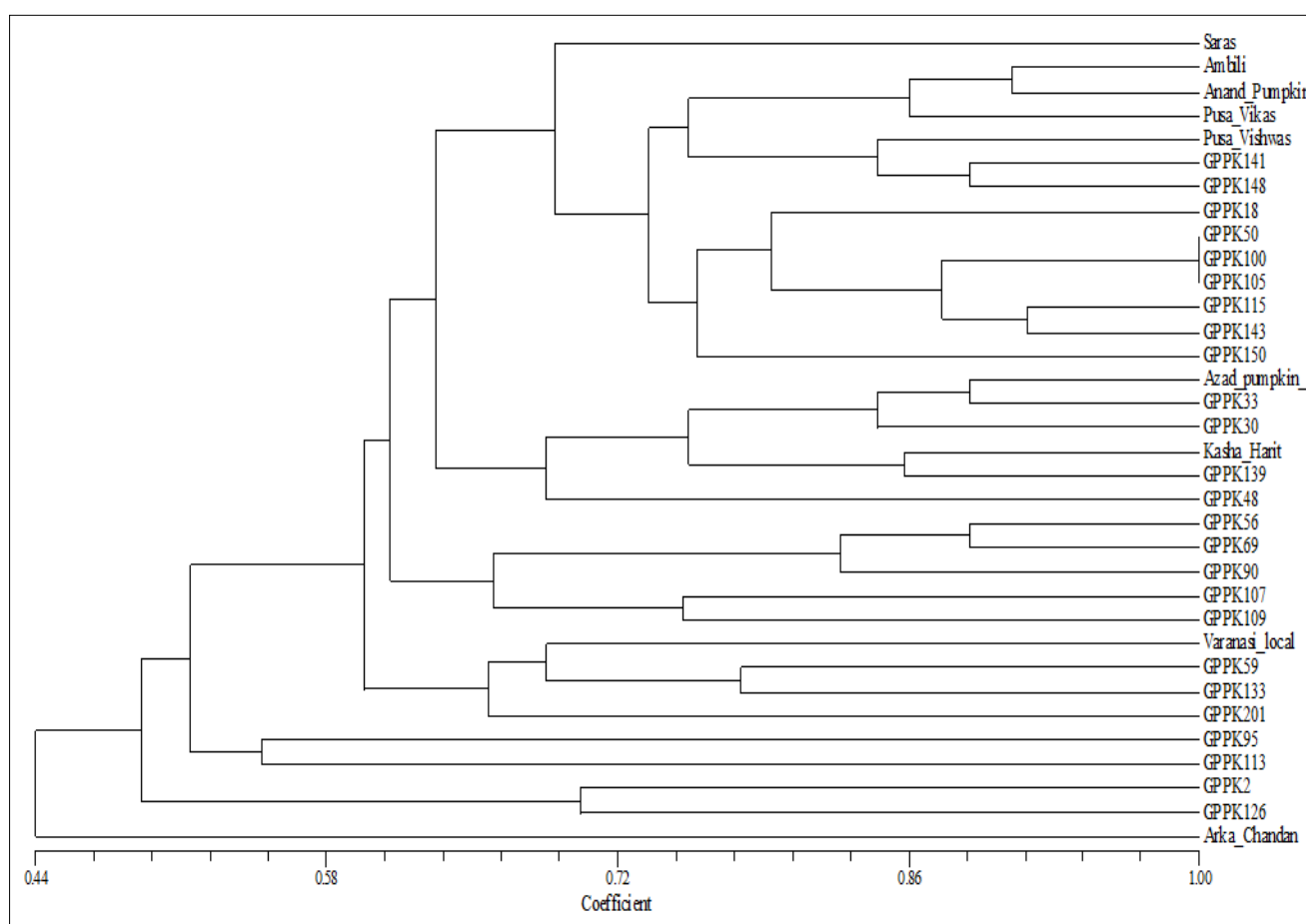


Fig. 6. Dendrogram of 34 pumpkin genotypes constructed using UPGMA cluster analysis based on Jaccard similarity coefficient and molecular marker data.

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interest to declare.

Ethical issues: None

References

- Andres TC. Diversity in tropical pumpkin (*Cucurbita moschata*): A review of intraspecific classifications. In: Progress in Cucurbit Genetics and Breeding Research. Proceedings of Cucurbitaceae; 2020. p. 107–12.
- Tindall HD. Vegetables in the tropics. London: Macmillan Education; 1987.
- Bailey LH. The domesticated cucurbits. J Genet Herb. 1929;2:62.
- Sharma S, Rao TR. Nutritional quality characteristics of pumpkin fruit as revealed by its biochemical analysis. Int Food Res J. 2013;20(5):2309–16.
- Xinyu MX. Study on GTF pumpkin milk powder. China Dairy Ind. 1988;6:6.
- Bose TK, Som MG. Vegetable crops in India. Calcutta: Naya Prakash; 1988. p. 92–5.
- Singh JK, Kumar JC, Sharma JR. Genetic variability and heritability of some economic traits in pumpkin in different seasons. Punjab Hortic J. 1988;28(3–4):238–42.
- Satkar KP, Kulthe AA, Chalke PR. Preparation of bitter gourd ready-to-serve beverage and effect of storage temperature on its keeping quality. Bioscan. 2013;8(1):115–7.
- Fisher RA. The correlation between relatives on the supposition of Mendelian inheritance. Philos Trans R Soc. 1918;52:399–433. <https://doi.org/10.1017/S0080456800012163>
- Al-Ashkar I, Alderfasi A, El-Hendawy S, Al-Suhaibani N, El-Kafafi S, Seleiman MF. Detecting salt tolerance in doubled haploid wheat lines. Agronomy. 2019;9(4):211. <https://doi.org/10.3390/agronomy9040211>
- El-Hendawy S, Al-Suhaibani N, Al-Ashkar I, Alotaibi M, Tahir MU, Solieman T, et al. Combining genetic analysis and multivariate modeling to evaluate spectral reflectance indices as indirect selection tools in wheat breeding under water deficit stress conditions. Remote Sens. 2020;12(9):1480. <https://doi.org/10.3390/rs12091480>
- Olivoto T, Nardino M. MGDI: Toward an effective multivariate selection in biological experiments. Bioinformatics. 2021;37(10):1383–9. <https://doi.org/10.1093/bioinformatics/btaa981>
- Weising K, Atkinson RG, Gardner RC. Genomic fingerprinting by microsatellite-primed PCR: A critical evaluation. Genome Res. 1995;4(5):249–55. <https://doi.org/10.1101/gr.4.5.249>
- Akkaya MS, Shoemaker RC, Specht JE, Bhagwat AA, Cregan PB. Integration of simple sequence repeat DNA markers into a soybean linkage map. Crop Sci. 1995;35:1439–45. <https://doi.org/10.2135/cropsci1995.0011183X003500050030x>
- Roberts DW, Wolfe AD. Sequence-related amplified polymorphism (SRAP) markers: A potential resource for studies in plant molecular biology. Appl Plant Sci. 2014;2(7):1400017. <https://doi.org/10.3732/app.1400017>
- Malik CP, Singh MB. Extraction and estimation of amino acids and keto acids. In: Plant enzymology and histo-enzymology. New Delhi: Kalyani Publishers; 1980.
- Sadasivam S. Biochemical methods. New Delhi: New Age International Publishers; 1970.
- Cox HE, Pearson D. The chemical analysis of foods. New York: Publishing Campus Inc.; 1962. p. 420.
- Sadasivam S, Manickam A. Biochemical methods. New Delhi: New Age International Publishers; 1996.
- Doyle JJ, Doyle JL. Isolation of plant DNA from fresh tissue. Focus. 1990;12:13–5. <https://doi.org/10.2307/2419362>
- Goldman D, Merril CR. Silver staining of DNA in polyacrylamide gels: Linearity and effect of fragment size. Electrophoresis. 1982;3:24–6. <https://doi.org/10.1002/elps.1150030105>
- Gopinath PP, Parsad R, Joseph B, Adarsh VS. GRAPES: General RShiny Based Analysis Platform Empowered by Statistics; 2020.
- Olivoto T, Lúcio AD. Metan: An R package for multi-environment trial analysis. Methods Ecol Evol. 2020;11:783–9. <https://doi.org/10.1111/2041-210X.13384>
- Akter S, Rasul MG, Islam AA, Rahman MM. Genetic variability, correlation and path coefficient analysis of yield and quality traits in pumpkin (*Cucurbita moschata* Duch. ex Poir.). Bangladesh J Plant Breed Genet. 2013;26:25–33. <https://doi.org/10.3329/bjpb.v26i1.19981>
- Chaudhari DJ, Acharya RR, Gohil SB, Patel NA. Genetic divergence study in pumpkin (*Cucurbita moschata* Duch. ex Poir.). J Pharmacogn Phytochem. 2017;6:744–7.
- Kumar V, Mishra DP, Yadav GC, Dwivedi DK. Genetic diversity assessment for morphological, yield and biochemical traits in genotypes of pumpkin. J Pharmacogn Phytochem. 2017;6:14–8. <https://doi.org/10.20546/ijcmas.2017.605.108>
- Kumar R, Rajasree V, Praneetha S, Rajeswari S, Tripura U. Correlation and path coefficient analysis studies in pumpkin (*Cucurbita moschata* Duch. ex Poir.) for yield and quality traits. Int J Curr Microbiol Appl Sci. 2018;7:3067–75. <https://doi.org/10.20546/ijcmas.2018.705.358>
- Mahmud E, Karim MR, Talukder MMR, Hasan GN, Islam MN. Phenotypic variability among pumpkin germplasm (*Cucurbita moschata* Duch. ex Poir.) in the southern part of Bangladesh. Int J Agric Pap. 2016;1:22–6.
- Rasul MG, Akter S, Animul IAKM. Genetic variability, correlation and path coefficient analysis of yield and quality traits in pumpkin (*Cucurbita moschata* Duch. ex Poir.). Bangladesh J Plant Breed Genet. 2013;26:25–33. <https://doi.org/10.3329/bjpb.v26i1.19981>
- Srikanth M, Bharad SG, Thulasiram LB, Potdukhe NR. Studies on genetic variability, heritability and genetic advance in pumpkin (*Cucurbita moschata* Duch. ex Poir.). Int J Curr Microbiol Appl Sci. 2017;6:1416–22. <https://doi.org/10.20546/ijcmas.2017.606.166>
- Sultana S, Kawochar MA, Naznin S, Siddika A, Mahmud F. Variability, correlation and path analysis in pumpkin (*Cucurbita moschata*). Bangladesh J Agric Res. 2015;40:479–89. <https://doi.org/10.3329/bjar.v40i3.25421>
- Muralidhara MS, Narasgowda NC, Narayanaswamy P. Genetic divergence in pumpkin (*Cucurbita moschata* Duch. ex Poir.). Indian J Hortic. 2014;4(3):144–7.
- Cyrl NC, Ayinde DL, Olatunji O. Genetic variability and heritability of vegetative, fruit and seed yield traits in fluted pumpkin (*Telfairia occidentalis* Hook. f.). Afr J Biotechnol. 2014;13:3262–70. <https://doi.org/10.5897/AJB2013.13088>
- Zinash A, Workneh TS, Woldetsadik K. Effect of accessions on the chemical quality of fresh pumpkin. Afr J Biotechnol. 2013;12(51):7092–8.
- Pandey S, Singh J, Upadhyay AK, Ram D. Genetic variability for antioxidant and yield components in pumpkin (*Cucurbita moschata* Duch. ex Poir.). J Veg Sci. 2002;29(2):123–26.
- Tarek AE, Khaled MT. Characteristics and composition of watermelon, pumpkin and paprika seed oils and flours. J Agric Food Chem. 2001;49(3):1253–9. <https://doi.org/10.1021/jf001117+>
- Pandya JB, Rao RTV. Analysis of certain biochemical changes associated with growth and ripening of pumpkin fruit in relation to its seed development. J Pure Appl Sci. 2010;18:34–9.

38. Habib A, Biswas S, Siddique AH, Manirujjaman M, Uddin B, Hasan S, et al. Nutritional and lipid composition analysis of pumpkin seed (*Cucurbita maxima* Linn.). J Nutr Food Sci. 2015;5:374-80.
39. Kulaitiene J, Jariene E, Danilcenko H, Cerniauskiene J, Wawrzyniak A, Hamulka J, et al. Chemical composition of pumpkin (*Cucurbita maxima* Duch.) flesh flours used for food. J Agric Food Environ. 2014;12:61-4.
40. Montesano D, Blasi F, Simonetti M, Santini A, Cossignani L. Chemical and nutritional characterization of seed oil from *Cucurbita maxima* L. var. Berrettina pumpkin. Foods. 2018;7(3):30. <https://doi.org/10.3390/foods7030030>
41. Rupji M, Dwivedi B, Kowalski J. NOJAH: Not Just Another Heatmap for genome-wide cluster analysis. PLoS One. 2019;14:42-5. <https://doi.org/10.1371/journal.pone.0204542>
42. Meier C, Marchioro V, Meira D, Olivoto T, Klein L. Genetic parameters and multiple-trait selection in wheat genotypes. Pesq Agropecu Trop. 2021;51:79-86. <https://doi.org/10.1590/1983-40632021v5167996>
43. Uddin S, Billah M, Afroz R, Rahman S, Jahan N, Hossain M, et al. Evaluation of 130 eggplant (*Solanum melongena* L.) genotypes for future breeding program based on qualitative and quantitative traits and genetic parameters. Horticulturae. 2021;7:376. <https://doi.org/10.3390/horticulturae7100376>
44. Benakanahalli NK. A framework for identification of stable genotypes based on MTSI and MGDI indexes: An example in guar (*Cyamopsis tetragonoloba* L.). Agronomy. 2021;11:1221. <https://doi.org/10.3390/agronomy11061221>
45. Singamsetti A, Zaidi PH, Seetharam K, Vinayan MT, Olivoto T, Mahato A, et al. Genetic gains in tropical maize hybrids across moisture regimes with multi-trait-based index selection. Front Plant Sci. 2023;14:1147424. <https://doi.org/10.3389/fpls.2023.1147424>
46. Ahmed SR, Ali Z, Ijaz I. Multi-trait selection of quinoa ideotypes at different levels of cutting and spacing. Sustainability. 2023;15:11446. <https://doi.org/10.3390/su151411446>
47. Smith JSC, Smith OS. Fingerprinting crop varieties. Adv Agron. 2000;47(1):85-140. [https://doi.org/10.1016/S0065-2113\(08\)60489-7](https://doi.org/10.1016/S0065-2113(08)60489-7)
48. Zraidi A, Stift G, Pachner M, Shojaeiyan A, Gong L, Lelley T. A consensus map for *Cucurbita pepo*. Mol Breed. 2007;20(4):375-88. <https://doi.org/10.1007/s11032-007-9098-6>
49. Ferriol M, Pico B, Nuez F. Genetic diversity of a germplasm collection of *Cucurbita pepo* using SRAP and AFLP markers. Theor Appl Genet. 2003;107:271-82. <https://doi.org/10.1007/s00122-003-1242-z>
50. Paris HS, Yonash N, Portnoy V, Mozes-Daube N, Tzuri G, Katzir N. Assessment of genetic relationships in *Cucurbita pepo* using DNA markers. Theor Appl Genet. 2002;106(6):971-8. <https://doi.org/10.1007/s00122-002-1157-0>
51. Ramu P, Senthilvel S, Upadhyaya HD, Hash CT. Assessment of genetic diversity in the sorghum reference set using EST-SSR markers. Theor Appl Genet. 2013;126(8):2051-64. <https://doi.org/10.1007/s00122-013-2117-6>

Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonpublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc. See https://horizonpublishing.com/journals/index.php/PST/indexing_abstracting

Copyright: © The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited (<https://creativecommons.org/licenses/by/4.0/>)

Publisher information: Plant Science Today is published by HORIZON e-Publishing Group with support from Empirion Publishers Private Limited, Thiruvananthapuram, India.