scientific reports

Check for updates

OPEN

Genetic diversity and population structure of Bael [*Aegle marmelos* (L.) Correa] genotypes using molecular markers in the North-Western plains of India

Prabhdeep Singh[1]✉, Akash Sharma[1], Vikas Tandon[2], Romesh Kumar Salgotra[3], Manish Sharma[4], Vishal Gupta[5] & Devinder Sharma[6]

Bael is a fruit crop that is extensively distributed throughout South-East Asia and is underutilized in medicine. The potential applications of bael's therapeutic and nutritional qualities in diverse ethnic communities are enormous. This study focuses on evaluating the morpho-pomological and molecular characteristics, utilizing SSR markers, of 80 wild bael genotypes alongside the NB-5 and NB-9 cultivars, derived from the North Western plains of India. Based on the evaluated morpho-pomological features, substantial variations were found between all genotypes. The fruit's inner diameter and pulp weight varied from 4.41 to 11.54 cm and 34.63 to 786.41 g, respectively. Numerous variations in the genotypes were observed in the shell weight/fruit, fruit skull thickness and fruit yield/plant. The bael fruit mucilage's total soluble solids (TSS) and total sugar content varied from 40.10 to 49.60 °brix and 8.11 to 21.17%, respectively. Using ward cluster analysis, the genotypes were divided into two primary clusters. Among the bael genotypes, the population structure analysis identified three subpopulations. SSR markers are used to measure genetic variety; of the 27 polymorphic markers, 17 show allelic diversity between genotypes. Molecular genetic diversity analysis, on the other hand, highlighted the genotypes genetic distinctiveness by classifying them into three major clusters. These findings offer valuable insights into the rich diversity and intricate interactions among the bael genotypes under investigation, paving the way for more strategic future breeding and selection efforts to elevate the quality of this remarkable fruit.

Keywords *Aegle marmelos* (L) Correa, Morpho-pomological diversity, Clustering, SSR markers

Wild and underutilized fruit-bearing plants in the northern Himalayas hold vast potential for rural development by creating diverse commercial products[1]. A diverse range of underutilized crops, not extensively cultivated or traded on a large scale, are mainly grown, marketed, and consumed locally, offering advantages like ease of cultivation and resilience to climate variations[2]. Hence, the fruits of wild and underutilized plants exhibit significant potential nutritional value, serving as rich sources of protein, fat, carbohydrates, and a plethora of macronutrients and micronutrients, while also containing phytochemical compounds with diverse therapeutic

[1]Division of Fruit Science, Sher-E-Kashmir University of Agricultural Sciences and Technology of Jammu, Main Campus, Chatha, Jammu 180009, India. [2]Advanced Centre for Horticulture Research, Udheywalla, Sher-E-Kashmir University of Agricultural Sciences and Technology of Jammu, Chatha, Jammu 180018, India. [3]School of Biotechnology, Sher-E-Kashmir University of Agricultural Sciences and Technology of Jammu, Main Campus, Chatha, Jammu 180009, India. [4]Division of Statistics and Computer Sciences, Sher-E-Kashmir University of Agricultural Sciences and Technology of Jammu, Main Campus, Chatha, Jammu 180009, India. [5]Division of Plant Pathology, Sher-E-Kashmir University of Agricultural Sciences and Technology of Jammu, Main Campus, Chatha, Jammu 180009, India. [6]Division of Entomology, Sher-E-Kashmir University of Agricultural Sciences and Technology of Jammu, Main Campus, Chatha, Jammu 180009, India. ✉email: dhaliwaldeep08@gmail.com

properties. Extracts from underutilized fruits are utilized in traditional medicine for ailments like colds, fevers, and diabetes, showcasing their significant medicinal value[3–7]. Bael [*Aegle marmelos* (L.) Correa], belonging to the Rutaceae family, holds significant prominence as an underutilized fruit crop in India and is well-known for its nutritional and therapeutic qualities[8]. It is mainly found in tropical and subtropical locations[9] and it is remarkably adaptable to difficult soil and environmental circumstances[10]. It thrives in alkaline, rocky, and shallow soils and can tolerate temperatures between − 7 to 50 ºC[11].

Bael is a great option for areas with scarce water supplies because it's high in vitamins and minerals. Given the current state of the fast growing worldwide market for natural antioxidants and functional foods. The bael fruit serves as a rich reservoir of riboflavin, offering therapeutic benefits in combating beriberi[12]. Additionally, its unripe counterpart is often recommended for the treatment of diarrhea and dysentery, while the presence of marmelosin in the fruit underscores its efficacy in addressing various stomach ailments[11]. Beyond the fruit, all parts of the bael plant contain a diverse array of bioactive compounds, including coumarins, alkaloids, sterols, and essential oils, renowned for their medicinal properties[13]. These compounds exhibit a spectrum of health-promoting effects, including analgesic, antipyretic, anti-inflammatory, antifungal, hypoglycemic, wound healing, insecticidal, and anti-fertility activities[2]. In the market, bael fruit is predominantly consumed in processed forms such as jams, squash, murabba, powder, preserves, nectar, and toffee[14]. Particularly during the COVID-19 pandemic, these products have witnessed heightened demand due to their perceived ayurvedic medicinal values, resulting in elevated market prices. Consequently, bael cultivation is emerging as a lucrative venture for farmers, especially in arid and semi-arid regions.

Morpho-pomological and biochemical profiling effectively discerns genetic diversity, conserves germplasm, and evaluates agronomic traits in endangered plants and commercial crops, with morpho-pomological characteristics serving as vital determinants for taxonomic classification and assessing genetic diversity within germplasm[15]. Debbarama and Hazarika[16] study on thirty bael accessions from eight districts of Tripura, India, identified distinct clustering into two groups, revealing underlying patterns of genetic variation. Similarly, an investigation into the genetic diversity of bael genotypes in north-western India has confirmed greater genetic diversity in its native range[15]. Furthermore, Dhakar et al.[17] conducted a thorough evaluation of fruit characteristics among bael genotypes from Ranchi, Jharkhand, India, revealing significant variability and distinct differences in fruit traits. These studies highlight the substantial morphological diversity present in the bael germplasm, which is valuable for cultivar identification and genetic improvement programs.

Morpho-pomological traits have historically been used for bael identification and characterization. On the other hand, substantial genetic variability frequently makes it possible to distinguish between individual trees with greater accuracy. When using morphological traits to evaluate the diversity and relationships between different plant species, environmental influences may have an insufficient effect. Because of this, scientists have looked into the possibility of using molecular markers as a more accurate way to describe and differentiate between different bael species[18]. The taxonomic classification and agronomic evaluation of plants depend heavily on morphological parameters[19]. Because these criteria are simple to apply and reasonably priced, plant breeders prefer to use them when assessing genetic materials[20]. Morphological traits offer a way to evaluate diversity in response to environmental variations, even though they can be sensitive to phenotypic plasticity[21]. The development of plant breeding programmes and the identification of desirable traits depend heavily on morphological investigations. These classifications are useful tools that help plant breeders and gene bank managers to achieve their goals, such as the introduction of commercial cultivars with superior fruit quality and the identification of dwarf and resistant rootstocks[22,23].

The 1990s genomics revolution made great strides towards our understanding of the genetic makeup of many organisms, including plants. During this time, molecular marker technology arose, leading to the creation of various marker types, including sequence characterised amplified region (SCARs), cleaved amplified polymorphic sequences (CAPS), microsatellite or simple sequence repeat (SSR), amplified fragment length polymorphism (AFLP), fragment length polymorphism (RFLP), random amplification of polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), sequence characterised amplified region (SCARs), single nucleotide polymorphism (SNP), and diversity arrays technology (DArT) markers[24]. Direct genetic material comparison is made possible by the use of DNA-based markers, which are unaffected by environmental factors. To improve the quantity and quality of bael, molecular markers can be used to evaluate the viability and purity of accessions[25].

When it comes to precisely identifying and describing closely related trees at the intra-specific level, molecular characterization is especially helpful. The degree of banding pattern similarity provides insight into the genetic similarity and connections between the samples under study. This strategy's efficacy is contingent upon a number of variables, such as the selection of markers, their genomic distribution, the loci they target, the degree of polymorphism, and the repeatability of the outcomes. In order to study genetic diversity and relationships, a number of molecular markers have been used, including inter simple sequence repeats (ISSR) and RAPD markers[8]. For comparative genome mapping, simple sequence repeats (SSRs) are very useful genetic markers that help with genotype classification and germplasm resource optimisation, resource utilization, and enhancing breeding programs. SSRs are among the various marker types that are helpful for evaluating genetic diversity. SSR generators are useful and have a lot of potential as genetic markers. SSR markers are highly esteemed and show great promise as genetic markers. Their capacity to transfer across various genetic backgrounds, multiallelic nature, co-dominance, ease of reproducibility, and random and widespread distribution throughout the genome have made them the preferred option in genetic studies[26]. The primary aim of this survey was to identify promising selections among a varied spectrum of bael genotypes and to examine the diversity present in their morpho-pomological traits. The primary aim of this survey was to identify promising selections among a varied spectrum of bael genotypes and genetic diversity among wild genotypes in the North Western Himalayas to enhance conservation management and future utilization in bael breeding programs.

Results

Morpho-pomological traits

The study examined the morpho-pomological traits of 80 wild bael genotypes and two commercial cultivars (NB-5 and NB-9). The results showed significant variability across multiple traits, including tree height varied from 6.50 to 17.20 m whereas, the tree height recorded in commercial cultivars NB-5 and NB-9 were 8.10 m and 14.30 m respectively. The greenness index ranged from 17.77 to 34.87 SPAD units with the commercial cultivars NB-5 at 28.63 SPAD units and NB-9 at 30.67 SPAD units. Inner diameter spanned from 4.41 to 11.54 cm wherein, NB-5 observed 8.72 cm and NB-9 measured 9.03 cm. Pulp weight ranged from 34.63 to 746.81 g, in contrast to 452.81 g for NB-5 and 574.92 g for NB-9. The total soluble solids (TSS) of mucilage varied from 40.10 to 49.60°brix, with the commercial cultivars NB-5 at 43.50°brix and NB-9 at 49.10°brix. Total sugars ranged from 8.11 to 21.17%, in addition to NB-5 and NB-9 (13.57% and 14.62%) respectively. Additional variability was observed in floral parameters, leaf characteristics, seed and yield attributes, as well as biochemical properties. The majority of investigated bael genotypes exhibited high variability for most traits, with fruit yield/plant (89.52%) having the highest coefficient of variation (CV in %), followed by leaf base (84.82%) and trunk color (69.72%), while TSS mucilage (6.66%) displayed the lowest. Significant variation in a trait across different germplasm individuals is generally indicated by a coefficient of variation larger than 10%[27]. Coefficient of variation analysis revealed high variability (> 10%) in 34 out of 40 traits, indicating the rich genetic diversity within the germplasm. Furthermore, Skewness and kurtosis were computed to explore genetic divergence among genotypes, revealing attributes with high positive skewness were styler end cavity (4.27), stem end cavity (3.34), and fruit yield/plant (3.23) and negative skewness were non-reducing sugars (− 1.94), fruit skull thickness (− 0.88), and filament width (− 0.85). Additionally, attributes with high platykurtic distribution included styler end cavity (16.62), followed by fruit yield/plant (16.06), and stem end cavity (9.38), while those with high leptokurtic distribution encompassed trunk colour (− 1.67), style width (− 1.42), and stigma length (− 1.44) (Table 1).

Among 80 wild genotypes of bael and two commercial cultivars (NB-5 and NB-9), observations revealed that 48.78% exhibited a yellowish-grey trunk color, 42.68% displayed a greyish-yellow trunk colour and 8.54% had a grey trunk colour. Notably, variations in fruit skull color were observed, with 57.32% displaying a greenish-yellow colour, 2.44% presenting a dull white color, 15.85% showing a creamish-yellow colour, 6.10% revealing a russet yellow colour, and 18.29% exhibiting a greenish-yellow coloration. Regarding pulp flavor, 36.59% of genotypes had a mild flavor, 48.78% displayed moderate pulp flavor and 14.63% showcased a strong pulp flavor among all bael genotypes (Table 2).

The genotypic variance (GV), phenotypic variance (PV) and heritability in broad sense (h^2_b) were assessed for morphopomological traits. The traits with the highest genotypic and phenotypic variance were pulp weight (30,294.60 and 30,308.57 respectively), shell weight (3419.67 and 3424.67 respectively) and leaf area (312.69 and 312.80 respectively). The nineteen traits gave higher values for the broad sense heritability (> 98%) (Supplementary Table S3).

The positive and negative Pearson correlations were discovered between the studied 40 morpho-pomological traits. Trunk colour (E-W) and tree spread had a positive correlation (r = 0.24). A positive correlation was observed between bud length and bud width (r = 0.66), petal length and width (r = 0.59 and 0.58), stigma width and length (r = 0.39 and 0.29), filament length and width (r = 0.28), and style width (r = 0.26). The relationship between leaf area and leaf size was positive (r = 0.88). Inner diameter had a negative correlation with fruit maturity group (r = − 0.77) and immature fruit colour (r = − 0.23), but a positive correlation with pulp width (r = 0.91), shell weight (r = 0.85), fruit yield (r = 0.67), number of seed sack per fruit (r = 0.42), total seed weight per fruit (r = 0.41), pulp flavour (r = 0.41), and number of seeds per sack (r = 0.27). TSS mucilage was positively correlated with total sugars (r = 0.88), reducing sugars (r = 0.87) and non reducing sugars (r = 0.69) (Fig. 1) (Supplementary Table S4).

Based on morpho-pomological traits, the 80 wild bael genotypes and two commercial cultivars were grouped into two main clusters with sub clusters (Fig. 2). Thirty-three genotypes made up Cluster I, and forty-nine genotypes made up Cluster II.

Molecular characterization

SSR diversity analysis

The bi-nominal data matrix was created using the banding pattern of every genotype of Bael, with polymorphic bands serving as the basis for the assessment of diversity, and eighty wild bael genotypes and two commercial cultivars (NB-5 and NB-9) compared using 27 citrus-specific microsatellite markers (SSR markers) to amplify genomic DNA. Among the 27 citrus specific microsatellite markers, 17 markers exhibited considerable polymorphism and allelic diversity in bael. These markers, previously noted for their polymorphic nature within the Rutaceae family, produced distinct banding patterns upon amplification, facilitating the assessment of individual genotypes[26,28]. Ten citrus specific microsatellite markers, on the other hand, did not amplify at all, exposing no bands (null allele) in any of the bael genotypes. Sixty-four alleles in total, ranging from 2 to 9, were amplified on all genotypes, with an average of 4 alleles per locus. The major allele frequency ranged from 0.307 (CT02) to 0.784 (CT21), with a value of 0.541. Genic diversity varied from 0.348 (CCSM147) to 0.766 (CT02), with a value of 0.572. Availability ranged from 0.220 (CAGG9) to 0.963 (TAA01), averaging 0.618. Polymorphic information content (PIC) ranged from 0.287 (CCSM147) to 0.729 (CT02), with a value of 0.503, and is strongly dependent on the number of alleles per locus and allele frequencies in the population (Table 3).

Population structure

Following STRUCTURE analysis, among 80 wild bael genotypes with two commercial cultivars (NB-5 and NB-9), three sub-populations were identified. Each genotype was allocated to one of these three sub-populations

S. No	Trait	Abbreviation	Unit	Min	Max	Mean	NB-5	NB-9	SEM	SD	Skewness	Kurtosis	CV (%)
1.	Tree height	TrHe	m	6.50	17.20	10.33	8.10	14.30	0.30	2.69	0.60	−0.73	26.00
2.	Tree spread (E–W)	TrSp	m	0.80	8.80	2.94	2.30	2.50	0.15	1.37	1.25	3.19	46.64
3.	Trunk colour	TrCo	Code	1.00	7.00	3.22	1.00	7.00	0.25	2.24	0.15	−1.67	69.72
4.	Bud length	BuLw	mm	10.18	13.24	11.28	11.43	10.28	0.08	0.76	0.53	0.05	6.71
5.	Bud width	BuWi	mm	7.05	9.85	7.89	8.20	8.15	0.08	0.74	1.16	0.58	9.40
6.	Petal length	PeLe	mm	11.26	18.65	15.84	17.60	11.66	0.20	1.82	−0.77	−0.08	11.47
7.	Petal width	PeWi	mm	7.19	9.80	8.46	9.31	7.39	0.08	0.70	−0.03	−0.92	8.32
8.	Filament length	FiLe	mm	3.54	5.05	4.36	4.88	4.39	0.04	0.36	0.18	−0.53	8.32
9.	Filament width	FiWi	mm	0.48	0.81	0.70	0.77	0.76	0.01	0.07	−0.85	0.51	10.01
10.	Style length	StyLe	mm	1.02	1.52	1.30	1.06	1.34	0.02	0.14	−0.32	−0.96	11.03
11.	Style width	StyWi	mm	1.49	2.53	2.04	1.88	1.97	0.03	0.29	0.27	−1.42	14.33
12.	Stigma length	StiLe	mm	2.25	3.48	2.78	2.50	2.58	0.05	0.43	0.49	−1.44	15.36
13.	Stigma width	StiWi	mm	2.12	2.93	2.44	2.43	2.30	0.02	0.19	0.66	−0.07	7.70
14.	Petiole length	PeLe	cm	1.17	7.17	3.83	2.30	2.27	0.16	1.40	0.55	−0.32	36.68
15.	Leaf area	LeAr	cm^2	52.30	134.47	98.00	102.30	102.20	1.95	17.68	−0.48	−0.17	18.05
16.	Greenness index	GrIn	SPAD unit	17.77	34.87	29.33	28.63	30.67	0.50	4.53	−0.84	−0.23	15.44
17.	Central leaflet shape	CeLeSh	Code	3.00	7.00	4.59	3.00	5.00	0.13	1.21	0.12	−0.41	26.31
18.	Leaf apex	LeAp	Code	3.00	7.00	4.56	5.00	7.00	0.18	1.66	0.43	−1.42	36.48
19.	Leaf base	LeBa	Code	1.00	7.00	1.85	1.00	5.00	0.17	1.57	1.73	1.97	84.82
20.	Leaf size	LeSi	Code	3.00	7.00	4.56	5.00	5.00	0.15	1.37	0.31	−0.84	30.05
21.	Inner diameter	InDi	cm	4.41	11.54	7.75	8.72	9.03	0.20	1.79	−0.02	−1.07	23.16
22.	Pulp weight	PuWe	g	34.63	746.81	260.61	452.81	574.92	19.40	175.66	0.67	−0.34	67.40
23.	Shell weight/fruit	ShWe	g	10.26	225.42	86.54	181.73	225.35	6.47	58.63	0.81	−0.18	67.75
24.	Fruit skull thickness	FrSkTh	mm	1.80	2.99	2.65	2.09	2.62	0.03	0.28	−0.88	−0.07	10.51
25.	Number of seed sack per fruit	NuSeSaFu	number	7.33	16.67	11.95	15.33	15.33	0.24	2.21	0.21	−0.41	18.50
26.	Number of seeds per sack	NuSeSa	number	2.33	17.33	8.49	11.33	9.33	0.36	3.23	0.47	−0.38	38.00
27.	Total seed weight per fruit	ToSeWeFr	g	5.13	46.33	19.36	39.79	29.73	1.03	9.30	0.69	−0.11	48.02
28.	Fruit yield/plant	FrYi	kg	2.68	110.80	17.33	32.58	37.59	1.71	15.51	3.23	16.06	89.52
29.	Fruit maturity group	FrMaGr	Code	3.00	7.00	5.34	7.00	5.00	0.15	1.33	−0.20	−0.72	24.83
30.	Immature fruit colour	ImFrCo	Code	3.00	7.00	3.80	5.00	7.00	0.14	1.29	1.36	0.68	33.92
31.	Styler end cavity	StyEnCa	Code	3.00	5.00	3.10	5.00	3.00	0.05	0.43	4.27	16.62	13.99
32.	Stem end cavity	SteEnCa	Code	3.00	5.00	3.15	3.00	3.00	0.06	0.52	3.34	9.38	16.66
33.	Fruit skull colour	FrSkCo	Code	1.00	5.00	3.22	2.00	5.00	0.11	1.01	0.51	−0.13	31.26
34.	Locule arrangement	LoAr	Code	3.00	7.00	5.05	5.00	7.00	0.08	0.77	0.25	4.12	15.22
35.	Pulp colour	PuCo	Code	1.00	3.00	1.39	1.00	3.00	0.08	0.72	1.54	0.77	51.49
36.	Pulp flavor	PuFl	Code	3.00	7.00	4.56	7.00	5.00	0.15	1.37	0.31	−0.84	30.05
37.	TSS mucilage	TsMu	^{0}brix	40.10	49.60	45.94	43.50	49.10	0.34	3.06	−0.76	−0.93	6.66
38.	Total sugars	ToSu	%	8.11	21.17	15.92	13.57	14.62	0.34	3.06	−0.30	−0.26	19.22
39.	Reducing sugars	ReSu	%	2.51	9.58	5.78	3.61	6.62	0.24	2.14	0.20	−0.96	37.03
40.	Non reducing sugars	NoReSu	%	5.60	11.59	10.14	9.96	8.00	0.13	1.19	−1.94	4.68	11.69

Table 1. Descriptive statistics for morpho-pomological traits among the studied bael genotypes.

using a membership probability threshold of 0.8. A genotype's admixture status was determined by looking at probabilities less than 0.8. The genotype distribution looked like this: A mixed genetic composition was seen in 23 individuals in subpopulation 1, 25 in subpopulation 2, 22 in subpopulation 3, and 12 genotypes (Fig. 3) (Supplementary Table S5). The categorization of bael germplasm into three distinct subgroups was confirmed by the Evano test[29], which revealed a distinct peak at delta K when K was equal to 3 (Fig. 4). A maximum mean likelihood value of L (K) = − 1405.81 was obtained from the individual membership coefficients obtained at K = 3 from the STRUCTURE analysis. This value supported the division of the bael germplasm into three (K = 3) subpopulations. In the second subpopulation, expected heterozygosity a measure of the likelihood that two randomly chosen individuals at a particular locus differ from one another (heterozygous) range from 0.4229 to 0.4860, with an average value of 0.4562. The FST values (Supplementary Table S6) show that there was genetic differentiation among the three subpopulations, ranging from 0.2738 to 0.3340, with a mean value of 0.2960 for each subpopulation. Further, AMOVA analysis revealed that 87% variation existed within the populations and 13% variation among the population (Table 4).

Traits	Frequency unit	Categories				
Trunk colour	**No**	**Yellowish grey (40)**	**Yellow (0)**	**Greyish yellow (35)**	**Grey (7)**	
	%	48.78	0	42.68	8.54	
Central leaflet shape	No	Broadly ovate (0)	Lanceolate to ovate (25)	Lanceolate (49)	Ovate (8)	
	%	0	30.49	59.76	9.76	
Leaf apex	No	Acuminate (39)	Acute (22)	Aristate (21)		
	%	47.56	26.83	25.61		
Leaf base	No	Cuneate (60)	Round (11)	Attenuate (9)	Tapering (2)	
	%	73.17	13.41	10.98	2.44	
Leaf size	No	Small (30)	Medium (40)	Large (12)		
	%	36.59	48.78	14.63		
Fruit maturity group	No	Early (12)	Mid (44)	Late (26)		
	%	14.63	53.65	31.7		
Immature fruit colour	No	Light green (56)	Green (19)	Dark green (7)		
	%	68.29	23.17	8.54		
Styler end cavity	No	Shallow (78)	Depressed (4)	Highly depressed (0)		
	%	95.12	4.88	0		
Stem end cavity	No	Shallow (76)	Depressed (6)	Flattened (0)		
	%	92.68	7.32	0		
Fruit skull colour	No	Dull white (2)	Creamish yellow (13)	Greenish yellow (47)	Russet yellow (5)	Greenish (15)
	%	2.44	15.85	57.32	6.1	18.29
Locule arrangement	No	Scattered (5)	Centric (70)	Highly centric (7)		
	%	6.1	85.37	8.54		
Pulp colour	No	Pale yellow (61)	Yellow (10)	Dark yellow (11)		
	%	74.39	12.2	13.41		
Pulp flavour	No	Mild (30)	Moderate (40)	Strong (12)		
	%	36.59	48.78	14.63		

Table 2. Frequency distribution for the qualitative traits in the studied genotypes of bael.

Genetic relationship among genotypes

A dendrogram was created using SSR data to compare the pairwise distances between 80 wild genotypes of bael and two commercial cultivars. With sub-clusters, the genotypes were grouped into three main clusters, I, II, and III (Fig. 5). The seven genotypes in Cluster I was present. Two subclusters, I A and I B, were created out of Cluster I. There are 28 genotypes of bael in Cluster II. Four subclusters, II A, II B, IIC, and IID, were created from Cluster II. There were 47 genotypes of bael in Cluster III. Additionally, cluster III of bael genotypes was split into two subclusters, III A and III B (Supplementary Table S7). To understand the genetic relationships among the studied bael genotypes, principal coordinate analysis (PCoA) was performed and the first three components of PCoAs explained 42.09% of the total genetic variation with 27.10%, 9.26%, and 5.67%, respectively (Fig. 6). The PCoA analysis matched population structure findings in that they revealed the relationships among bael genotypes.

Discussion

Significant morpho-pomological differences have been found in closely related genotypes or populations of bael fruit in previous studies. These differences are explained by the different climates in each of their habitats as well as the relatively low heritability of the vegetative and reproductive characteristics of bael fruit[8,9,16,17,30–36]. In the present study, total sugars ranged from 8.11 to 21.17%, reducing sugars from 2.51 to 9.58% and non-reducing sugars from 5.60 to 11.59%. One study analyzed the biochemical composition of bael and found total sugars to be 14.35%, reducing sugars to be 4.42% and non-reducing sugars to be 9.93% in the fruit pulp[13,37]. Another study reported total sugars in bael fruit pulp to range from 3.08 to 6.94% [38]. Similarly, the studies conducted by Singh et al., [39] reported total sugars content of 7.6 g/100 g, reducing sugars of 6.2 g/100 g, and non-reducing sugars of 1.4 g/100 g in fruit pulp. With morpho-pomological trait variations ranging from 6.66 to 89.52%, the bael population under study had a high potential for reproduction. This variance indicated a noteworthy level of diversity between the individual samples. These studies are confirmative with fruit crops such as bael[16,17] and calamansi[40].

Complementary gene interactions are linked to the following positive skewness traits: immature fruit colour, styler end cavity, stem end cavity, fruit skull colour, locule arrangement, pulp colour, pulp flavour, reducing sugars, central leaflet shape, leaf apex, leaf base, leaf size, pulp weight, shell weight/fruit, number of seed sacks per fruit, number of seeds per sack, total seed weight per fruit, fruit yield/plant, and stem and bud length and width. Conversely, duplicate (additive x additive) gene interactions are linked to negative skewness[41] (petal length, petal width, filament width, style length, leaf area, greenness index, inner diameter, fruit skull thickness, fruit maturity group, TSS mucilage, total sugars, non-reducing sugars). Positive values indicated the presence of gene interaction, whereas negative values (tree height, trunk colour, petal length, width, filament length, style length, style

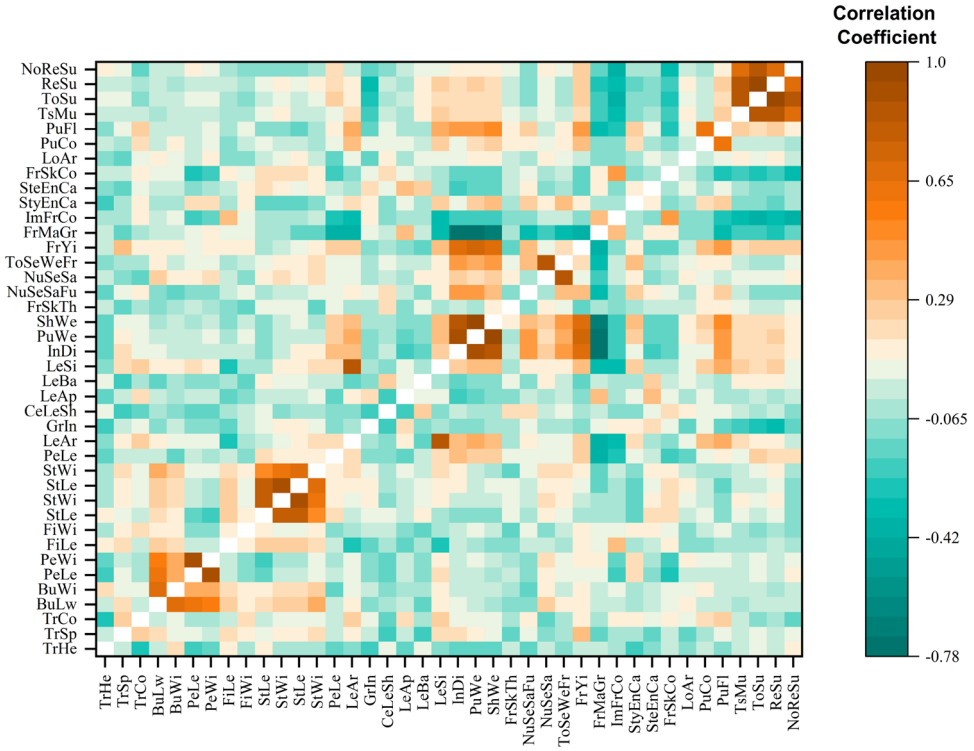

Figure 1. Correlation studies among morpho-pomological traits based on Pearson correlation matrix with heatmap. The blue area indicates a negative correlation between the two traits, and the orange area indicates a positive correlation between the two traits. The darker the color the higher the level of correlation.

width, petiole length, leaf area, greenness index, central leaflet shape, leaf apex, leaf size, inner diameter, pulp weight, shell weight/fruit, fruit skull thickness, number of seed sack per fruit, number of seeds per sack, total seed weight per fruit, fruit maturity group, fruit skull colour, pulp flavour, TSS mucilage, total sugars, reducing sugars) or close to zero kurtosis value indicated the absence of gene interaction suggested that gene interactions exist[41]. Traits with platykurtic and leptokurtic distributions are regulated by different numbers of genes; more genes are required to control platykurtic traits than leptokurtic traits[17]. Previous studies, in bael reported that traits such as fruit weight, skin weight, and acidity displayed positive skewness, while total soluble solids (TSS) and pulp percentage exhibited negative skewness. Additionally, the study reported platykurtic distributions in fruit weight, skin weight, and reducing sugar, whereas leptokurtic distributions were observed in fruit length and pulp weight[17].

Genetic variance and heritability are crucial factors in crop improvement through selective breeding. Higher genotypic variance (GV) and phenotypic variance (PV) indicate greater inheritable genetic variation that can be leveraged to develop improved cultivars. Conversely, lower GV and PV suggest limited potential for selection-based enhancement. In vegetatively propagated crops like Bael, estimating broad sense heritability is particularly valuable, as it accounts for both additive and non-additive genetic components transmitted to offspring. Previous studies on Bael have reported high GV, PV, and heritability for various traits, highlighting the availability of substantial heritable variation that can be effectively exploited in breeding programs to enhance desirable bael characteristics[16]. The availability of substantial heritable variation, as evident from the high genotypic and phenotypic variances, provides a strong foundation for effective selection and breeding programs aimed at enhancing desirable traits in this economically and nutritionally important fruit crop.

The purpose of the observed correlation between traits is to examine and establish a meaningful and logical relationship between them. It is possible to examine traits that may be challenging to measure by first establishing a relationship between multiple traits. Additionally, choosing correlated traits with a significant correlation can be chosen as appropriate indicators in situations where a trait's appearance is time-specific or necessitates exact measurements for identification. This method works especially well in situations where it is costly, complex, time-consuming, or challenging to measure a trait directly. When two traits exhibit correlation, there is a linear relationship between them that can be utilized in breeding programmes. This relationship can range from − 1 to + 1[42]. Comparable research has been done on fruit crops, such as cornelian cherries[43], Figures [44] and *Pyrus syriaca*[45].

Cluster analysis uncovered significant variations present among bael genotypes. In our study, thirty-three genotypes were present in Cluster I and remaining fifty-nine genotypes in Cluster II. Comparable research has been done on fruit crops; a previous study performed cluster analysis on seventy-five bael genotypes and classified them into two main clusters. These clusters were primarily distinguished based on fruit traits, with the first cluster comprising a total of twelve germplasm and the second cluster containing the remaining sixty-three

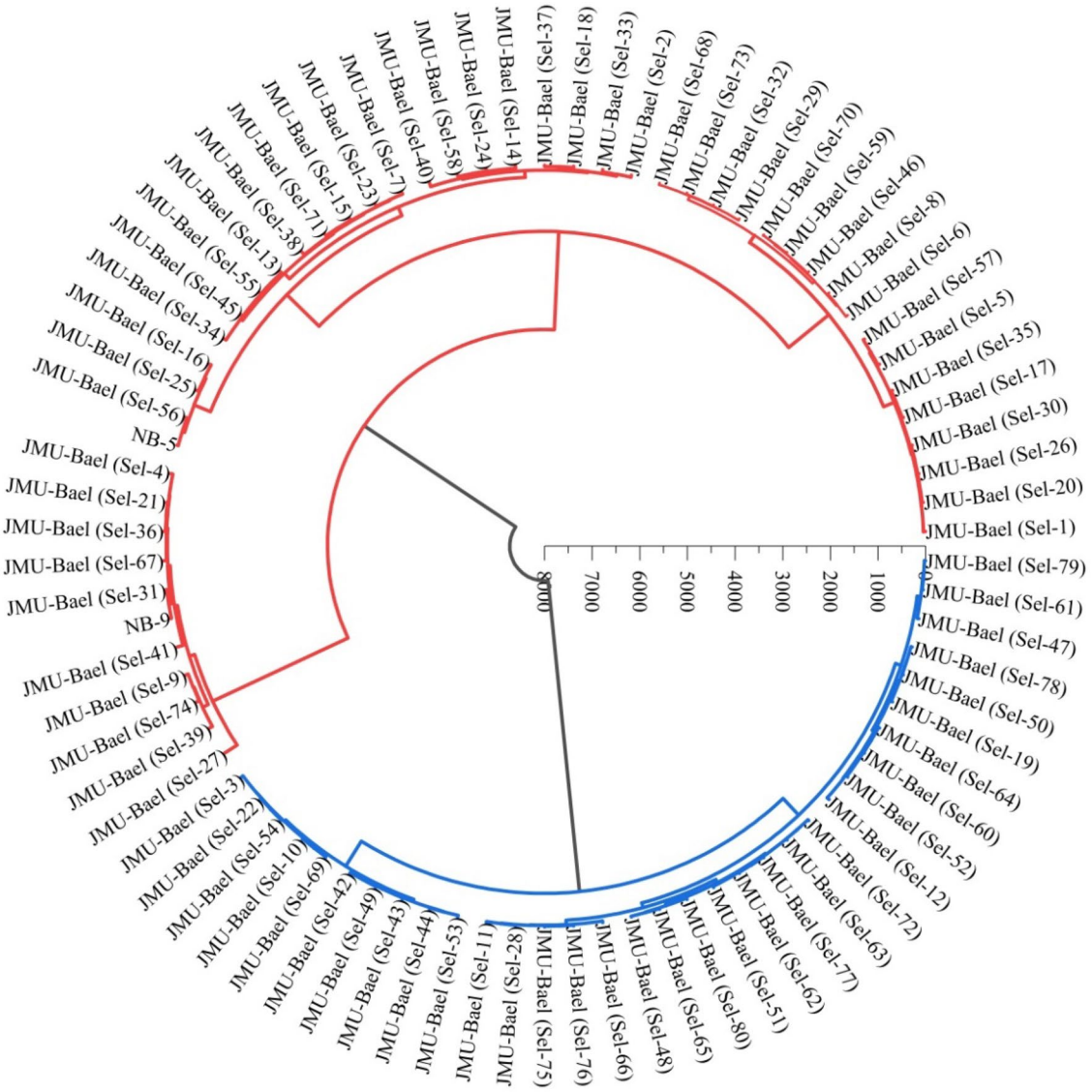

Figure 2. Ward cluster analysis of the studied bael genotypes based on morpho-pomological traits. The results show that the populations were divided into 2 categories, which was indicated by red and blue colors.

germplasm[17]. A comparable outcome was also noted by cluster analysis and studied thirty bael accessions and categorized into two primary clusters, with the first cluster encompassing five accessions and the second cluster comprising twenty-five accessions[16]. This study involving 151 Calamansi individuals, where statistical analysis led to their classification into four groups: the first group comprising 32 individuals, the second group including 7 individuals, the third group consisting of 25 individuals and the fourth group comprises 87 individuals[40].

Eighty wild bael genotypes and two commercial cultivars (NB-5 and NB-9) showing high genetic variation in germplasm. It is necessary for efficient breeding and selection processes. Increased genetic diversity within a given fruit tree increases the likelihood of more successful selection. In the Jammu region where bael is grown, where seeds have been the primary means of propagation for many years, it is anticipated that a wide variety has arisen. On the other hand, little is known about the molecular diversity in this natural gene pool. As such, this aspect needs to be looked into and clarified. The main goal of plant breeders is to improve the quantitative and qualitative characteristics of current cultivars. This has historically been accomplished through traditional breeding techniques that involve using the whole genome and choosing the best recombinants from a large number of segregating individuals. But this method requires a lot of work and time because it requires careful linkage drag, several generations, several crosses, and phenotypic selection[46]. In addition to conventional breeding for crop improvement, DNA-based molecular marker technologies have recently become useful tools for plant breeders. These technologies are used for cultivar identification and genetic diversity assessment. In particular, SSR markers have a traditional use in genetic diversity analysis. We studied the 80 wild bael genotypes with NB-5 and NB-9 cultivars of bael that are continuously important from an agronomic standpoint and are conserved in Jammu. The goal of this study is to improve our knowledge of the morpho-pomological and genetic relationships in bael, a cash crop of considerable economic importance. Recent research of a similar nature revealed the genetic

S. no.	Marker	Major allele frequency	Number of allele	Availability	Gene diversity	Polymorphic information content (PIC)
1.	CAC39	0.613	2	0.915	0.474	0.362
2.	CAG01	0.548	3	0.890	0.586	0.514
3.	CAGG9	0.722	2	0.220	0.401	0.321
4.	CCSM18	0.415	5	0.500	0.685	0.630
5.	CCSM147	0.776	2	0.707	0.348	0.287
6.	CT02	0.307	6	0.537	0.766	0.729
7.	CTT01	0.443	4	0.646	0.672	0.613
8.	CT21	0.784	3	0.451	0.349	0.304
9.	CY01	0.615	4	0.317	0.568	0.526
10.	CY05	0.622	3	0.549	0.535	0.473
11.	CY37	0.379	5	0.805	0.747	0.709
12.	CY48	0.397	4	0.768	0.693	0.635
13.	CY51	0.473	9	0.671	0.697	0.659
14.	SCM05	0.615	2	0.793	0.473	0.361
15.	TAA01	0.532	2	0.963	0.498	0.374
16.	CCSM77	0.538	2	0.317	0.497	0.374
17.	CCSM156	0.419	6	0.451	0.726	0.686
	Mean	0.541	4	0.618	0.572	0.503

Table 3. List of amplified SSR primers along with major allele frequency, number of alleles scored, availability, gene diversity and PIC value.

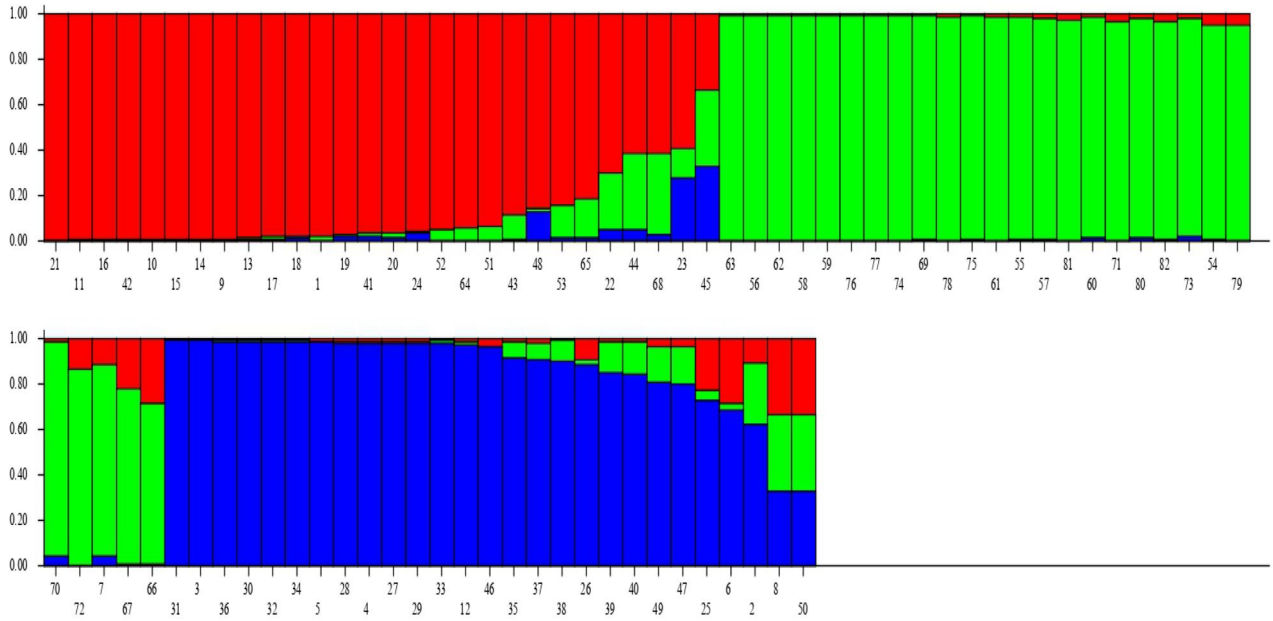

Figure 3. Graphical representation of population structure in the bael genotypes. Each individual is represented by a vertical line and different colours in the same line indicate the individual's estimated membership percentage in K clusters (admixture proportion or Q value): red = cluster 1, light green = cluster 2, blue = cluster 3.

relationships in bael[26,28], persimmon[47,48], all of which were confirmed by SSR data. Our duties extend beyond the simple identification and conservation of bael germplasm to include a detailed analysis of their traits and diversity. To do this, we optimised markers for a more thorough investigation of the relationships and variations among bael germplasm in the plains of North-Western Himalayas by combining the two methods of identification viz., morpho-pomological traits and SSR markers. In this study, we phenotyped and genotyped bael germplasm and evaluated their associations with appropriate reference SSR markers and morpho-pomological features.

All of the genotypes showed a different number of alleles per locus, highlighting the significant diversity found in the bael genotypes under study. Because of their high degree of diversity, the bael genotypes grown in the North-Western Himalayan region are useful for crop improvement programs. The number of genotypes and SSR markers employed in the study determines how many alleles are present at each locus. The number of alleles

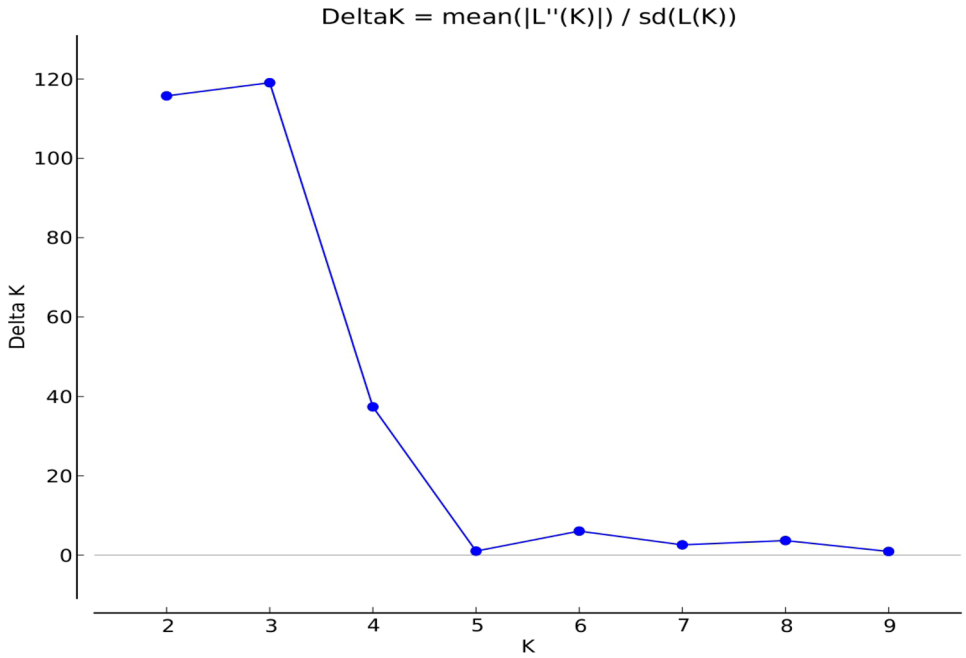

Figure 4. Plot of Delta *K* against subpopulation *K* with three subpopulations in bael genotypes.

Variance component	Degree of freedom	Sum of squares	Mean sum of squares	Estimated variance	Percentage of variance (%)	Stat	Value	P (rand > data)
Among population	3	213.974	71.325	2.639	13	PhiPT	0.126	0.001
Within population	78	1427.904	18.306	18.306	87			
Total	81	1641.878		20.946	100			

Table 4. Analysis of molecular variance (AMOVA).

per locus in the genotypes of bael was, therefore, a good indicator of sufficient polymorphism and suitability for evaluating genetic variation. Polymorphic information content was used to assess a genetic marker's informativeness. PIC ranged demonstrating the fact that PIC is highly influenced by the total number of alleles per locus and the allele frequencies in the population. A high PIC value is indicative of a genetically distant genotype. The findings of the present study align with previous investigations on the genetic diversity of *Aegle marmelos* (bael) using microsatellite markers reported a comparable range of 4 to 7 alleles across loci, with a mean of 4.7 ± 1.059 alleles per locus. Notably, the number of alleles observed for every 10 loci surpassed the effective number of alleles, which ranged from 1.384 to 3.164, with an average value of 1.995 ± 0.11[28]. Similarly, molecular analyses have explored other diversity parameters like polymorphic information content (PIC) (0.234 to 0.998) along with the number of amplified alleles (1 to 4), and gene diversity (0.003 to 0.063)[26]. The concordance between our observations and these previous findings underscores the diverse genetic landscape of bael genotypes. This information emphasizes the importance of comprehensive genetic analyses in elucidating the genetic diversity of bael and informing targeted breeding strategies to develop improved cultivars.

Through the use of DARwin software version 6.0 for clustering analysis, the bael genotypes were divided into three primary clusters, denoted as I, II, and III. The lack of clear differentiation between the bael genotypes according to the locations of each collection highlights the diversity that exists within the bael genotypes. The present findings align with previous studies on construction of the UPGMA dendrogram for forty bael genotypes resulted in the division of the samples into two main clusters. Cluster I comprised eleven genotypes, while cluster II comprised twenty nine genotypes[28]. Similarly, in the investigation involving twenty-four bael genotypes, a similar clustering pattern emerged, with the genotypes segregated into two major clusters A and B. Cluster A encompassed nineteen genotypes, while cluster B comprised five genotypes[26]. These findings suggest distinct genetic groupings within the bael genotypes studied, highlighting underlying patterns of genetic variation and population structure.

The population structure analysis data, shown as a structure matrix, was helpful in lowering the number of false positives. Twelve genotypes indicated mixed ancestry among the three subpopulations that the STRUCTURE analysis divided the genotypes into. Furthermore, the largest percentage of variance has been found between individuals. Three subpopulations were found in a population structure study related to Egyptian citrus rootstock[49], walnuts[50] and persimmons[47]. The model-based method separated the bael genotypes into three subpopulations using PCoA and STRUCTURE. Differences in how people are classified into various categories may

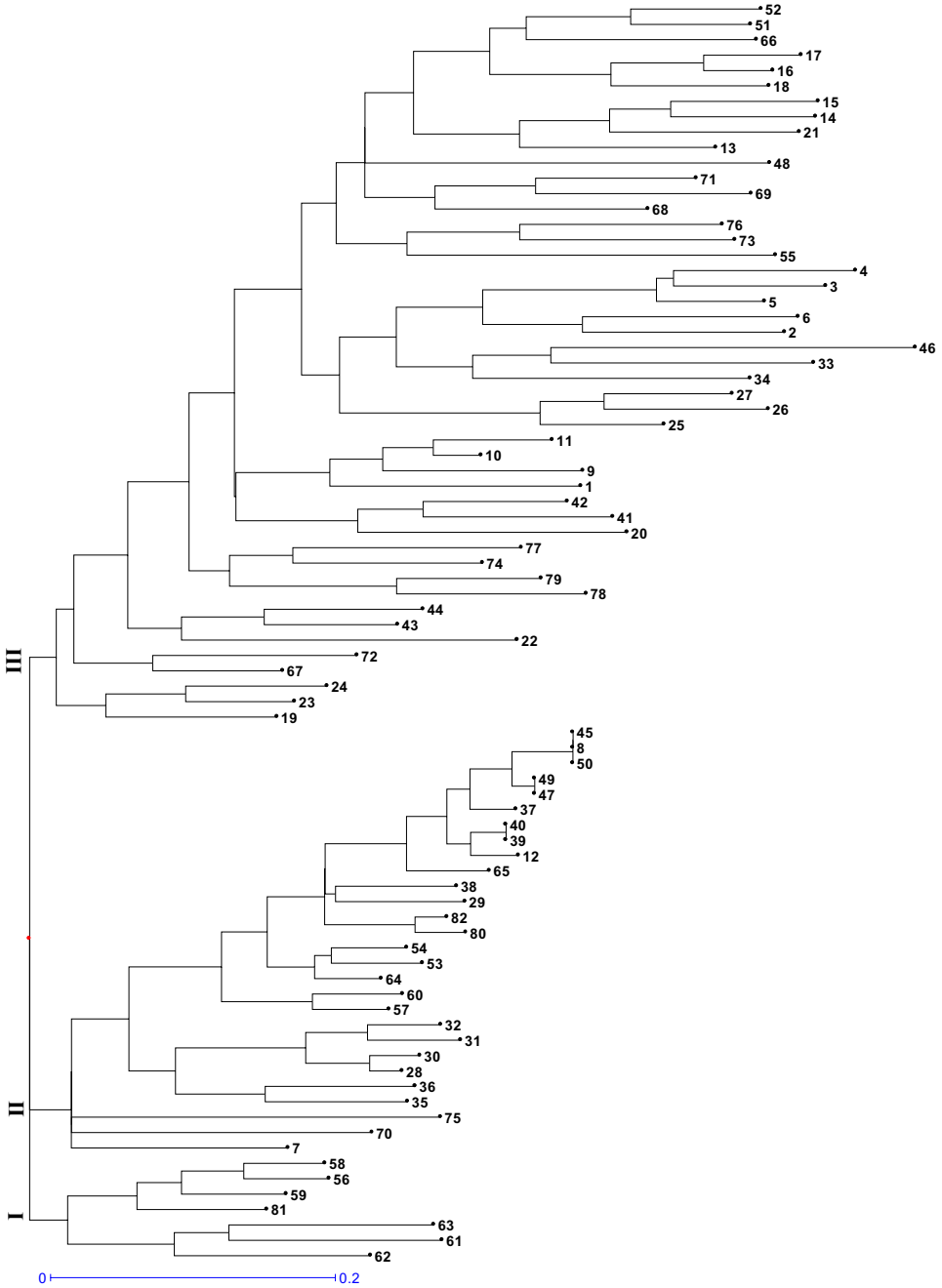

Figure 5. UPGMA dendrogram showing clustering pattern of bael genotypes based on SSR data.

result from variations in the algorithms that the software employs. In contrast to cluster analysis, which assigns genotypes to different groupings or clusters by establishing fixed branch positions for each genotype, structure analysis distributes genotypes to different subpopulations according to their highest membership percentages. The fruit crop diversity studies were also reported in bael[26,28], persimmon[47], walnut[50]. In addition, differences in germplasm, the selection of molecular markers, partial reproductive isolation, and decreased genetic drift can all contribute to differences in genetic diversity and population structure.

The analysis of molecular variance (AMOVA) provides further insights into the partitioning of genetic variation within the bael (*Aegle marmelos*) population. The results indicate that 13% of the total variation resides among the populations, while a substantial 87% of the variation is present within the populations. The moderate yet significant PhiPT value of 0.126 ($p < 0.001$) suggests genetic differentiation among the bael populations. These findings point to the presence of distinct genetic groups within the studied bael germplasm, which can be leveraged in breeding programs to develop improved cultivars. The high within-population variation (87%) highlights the availability of diverse genetic resources that can be effectively utilized for selection and hybridization to enhance desirable traits. The AMOVA results, in conjunction with the observed patterns of genetic

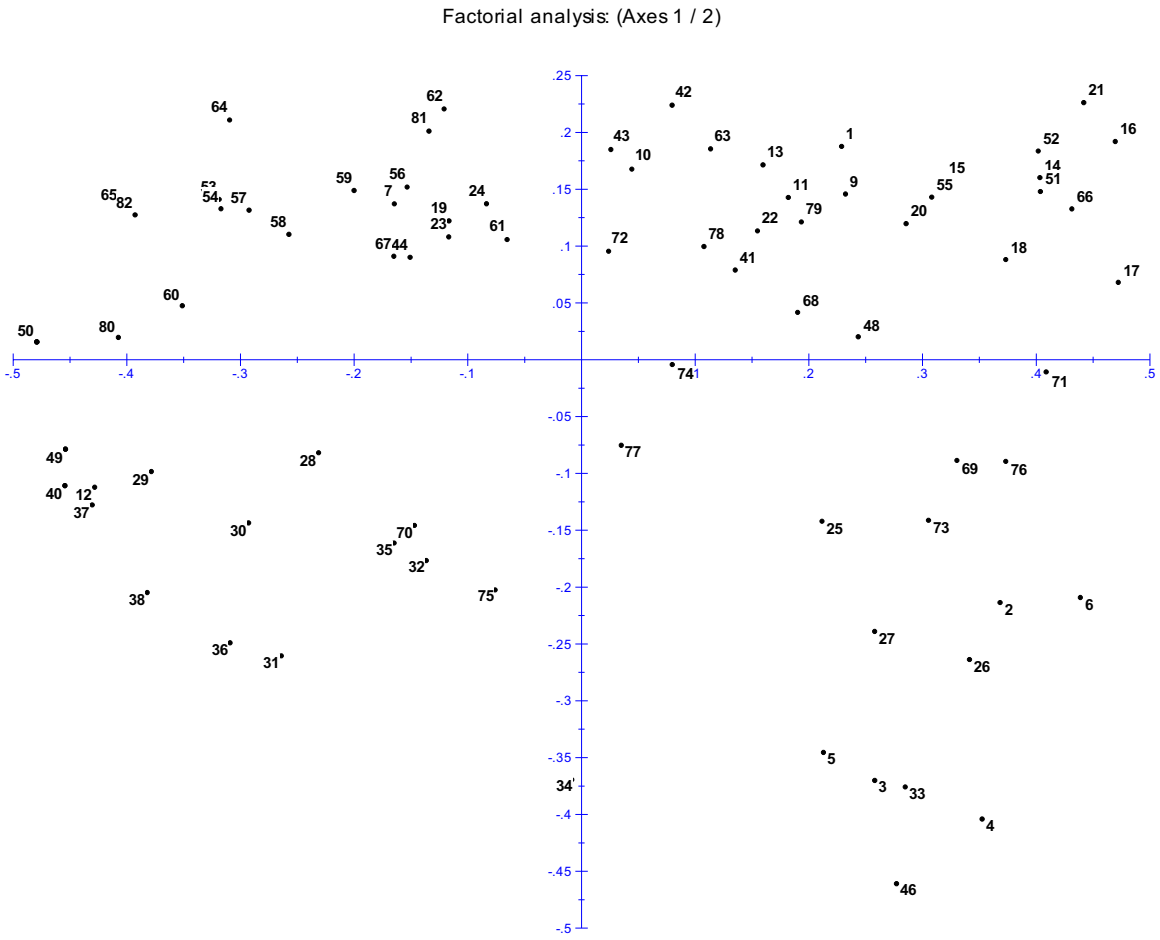

Figure 6. Principal coordinate analysis (PCoA) of bael genotypes.

diversity and clustering, provide valuable insights into the genetic architecture of bael. This information will guide the implementation of targeted conservation and breeding strategies for this economically and nutritionally important fruit crop. In a prior investigation concerning bael, the analysis of molecular variance (AMOVA) indicated that 70% of the overall marker variation was due to interpopulation variance, with the remaining 30% attributed to intrapopulation variance[36].

Conclusion

The Jammu region in North Western Himalayan has a notable diversity of wild bael genotypes, according to the results of the current investigations. Finding synonyms is a crucial tool for bael germplasm management in management studies. The study's eighty wild bael genotypes and two commercial cultivars (NB-5 and NB-9) appeared to have highly variable relationships, according to a combination of morpho-pomological and molecular marker analysis. Additionally, we looked at the morpho-pomological profiles and genetic relationships of two commercial cultivars, NB-5 and NB-9, as well as representative to eighty wild genotypes. These cultivars' distinct characteristics indicated their potential for use in breeding. According to the findings, SSR markers can also be a useful tool for identifying genotypes that exhibit desired morpho-pomological traits. The collective findings imply that a genetic bottleneck has not materialised in bael due to the variety of genetic and morpho-pomological variants in the population. Identifying prospective bael parents with traits of agronomic interest can speed up bael breeding with the help of the genetic and morpho-pomological profiles produced by this study. The diverse genotypes with superior traits such as JMU-Bael (Sel-27) will be crossed for improvement or development of superior bael cultivars. This study provided important information for upcoming breeding and selection programmes targeted at enhancing bael cultivars by demonstrating a significant variation in morpho-pomological traits among bael genotypes. A thorough understanding of bael diversity is essential for the sustainable cultivation and conservation of this significant fruit crop. This understanding is derived from the combination of morpho-pomological data and genetic analysis based on molecular markers.

Materials and methods

Plant materials

In the current study, a total of 80 wild bael genotypes, along with two national recommended cultivated varieties, NB-5 and NB-9, were meticulously chosen from the regions of Jammu (32.73°N, 74.87°E, 300 m above sea

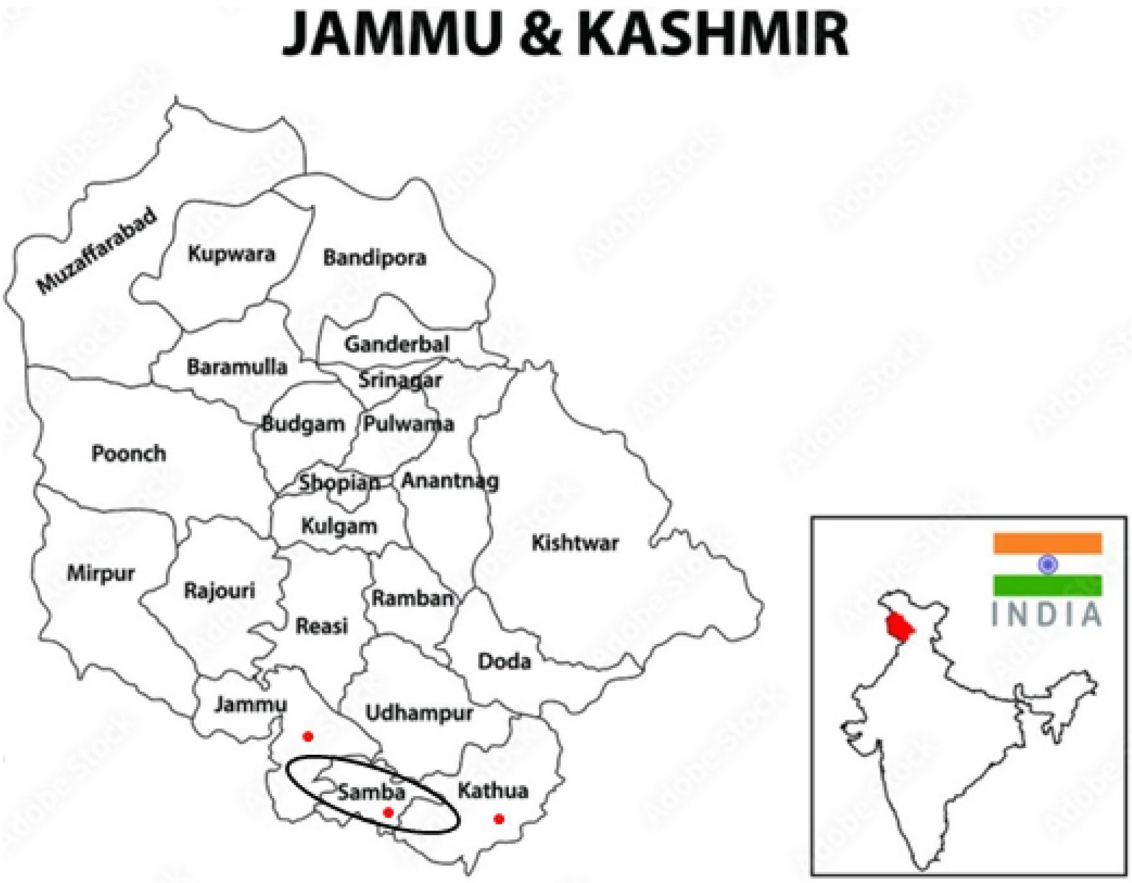

Figure 7. Centre of genetic diversity of bael. Highlights region are selected districts.

level), Samba (32.57°N, 75.12°E, 384 m above sea level) and Kathua (32.37°N, 75.52°E, 393 m above sea level) in the Jammu province of the Jammu and Kashmir Union Territory, India (Supplementary Table S1) (Fig. 7). The collection of plant material was carried out in accordance with relevant institutional, national, and international guidelines and legislation. Voucher specimens were identified by Prabhdeep Singh under the guidance of Dr. Akash Sharma, for breeding to develop the new cultivars. The deposition of voucher specimens of the collected germplasm of bael genotypes in the Division of Fruit Science, Faculty of Horticulture, SKUAST-Jammu (Voucher ID-AUJ/FS/23-24/121).

Morpho-pomological traits

The data of morpho-pomological traits was recorded as per Guidelines for the Conduct of Test for Distinctiveness, Uniformity and Stability of bael (*Aegle marmelos* Correa) Protection of Plant Varieties and Farmers Right's Authority (PPV&FRA) Government of India[51] and bael descriptor of National Bureau of Plant Genetic Resources[52].

Statistical analyses of morpho-pomological traits

The collected data were subjected to statistical analysis to determine the variability and patterns among the morpho-pomological traits and across the selected genotypes. Descriptive statistics, such as mean, standard error, standard deviation, skewness, kurtosis and coefficient of variation (CV), were calculated for each trait. The data obtained from the survey were sorted and analyzed using Microsoft Office Excel 2019, and statistical analysis software[53]. Pearson correlation coefficients among 40 morpho-pomological traits were calculated by statistical analysis software[53]. Cluster analysis was used to group Bael genotypes based on Ward's minimum variance using statistical analysis software[53].

Genotypic variance (GV), Phenotypic variances (PV) and Heritability (H^2_b) of the trait(s) were calculated using the formulae:

Genotypic variance (GV)

$$(\mathrm{GV}) = \frac{\mathrm{EMS} - \mathrm{MSt}}{\text{No. of replications (r)}}$$

Phenotypic variance (PV)

$$PV = \sigma_G^2 + \sigma_E^2$$

Heritability (h^2_b)

$$h2b = \frac{VG}{VP} x\ 100$$

where, or σ^2_G = Genotypic variance, σ^2_P = Phenotypic variance, σ^2_E = Error variance, MS_G = Genotypic mean square value, MS_E = Error mean square of value, r = Number of replication.

Molecular characterization

A set of 27 citrus specific microsatellite primers (SSR markers) were selected for genetic characterization of bael[26,28] (Supplementary Table S2). The concentration of the primers was made up to 10 µM and stored at – 20 °C.

Genomic DNA isolation

The CTAB method was used to extract genomic DNA from young trifoliate leaf of bael samples[54]. After that, each sample's DNA concentration was determined using a Nanodrop spectrophotometer by measuring the absorbance at 260/280 nm and the DNA quality was evaluated using a 0.8% agarose gel. DNA was diluted to a final concentration of 50 ng/µl for subsequent applications.

SSR marker analysis

PCR amplification was conducted in a 96-well Universal Gradient Thermal Cycler, utilizing a 15 µl reaction mixture. The reaction contained 1 µl of genomic DNA (50 ng), 0.2 M of both forward and reverse primers, 0.5 U of Taq polymerase (D1806-Sigma-Aldrich, USA), 1 × PCR buffer containing MgCl2 and 0.2 mM dNTPs. The amplification process followed this protocol: an initial denaturation at 94 °C for 4 min, succeeded by 30 cycles of denaturation at 94 °C for 30 s, annealing at 44–63 °C for 30 s, extension at 72 °C for 30 s, concluding with a final extension at 72 °C for 8 min. Subsequently, the PCR products were combined with 2 µl of 6 × loading dye (Thermo Scientific # R0611, Waltham, Massachusetts, USA) and resolved by electrophoresis on a 3% metaphor agarose gel[55].

SSR data analysis

Bands on the amplified DNA fragments represent the alleles for each SSR locus. Allele size was determined using a 100 bp DNA Ladder (3407A Takara) and allelic variants were estimated based on their relative movement in the gel. A null allele was assumed to exist for a certain genotype if a PCR result was not amplified. Power Marker software version 3.25[56] was used to determine a various parameters, including polymorphism information content[57], number of alleles, gene diversity[58], heterozygosity and the frequency of the major allele. DARwin version 6.0 was used to build a pairwise distance matrix. An unweighted neighbor-joining approach was used to perform cluster analysis after creating a dissimilarity matrix[59]. The dendrogram was generated through bootstrap analysis involving 1000 permutations. Additionally, principal coordinate analysis (PCoA) was conducted using DARwin version 6.0[59].

Population structure analysis

Model- based cluster analysis was performed to infer the genetic structure and to define the number of clusters (gene pools) in the data set using the software STRUCTURE version 2.3.4 software[60]. The number of presumed population (K) was set from 1 to 10, and analysis was repeated 2 times. For each run, burn-in and iterations were set to 1,00,000 and 2,00,000, respectively and a model without admixture and correlated allel frequencies was used. The optimum value of K was determined by calculating the 1K value to estimate the most likely number of groups[29]. STRUCTURE results were processed with the software STRUCTURE version 2.3.4 software[61] to obtain the most likely K value. Further, Analysis of Molecular variance (AMOVA) and Phi-PT values was analysed by using GenAlEx software[62].

Ethical approval

The collection of Bael germplasm resources and research activities has been conducted in compliance with the Regulations on Resident Instructions and duly approved by the Competent Authority of Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu, Main Campus, Chatha, Jammu, India.

Data availability

The original contributions presented in the study are included in the article/Supplementary Files, further inquiries can be directed to the corresponding authors.

Received: 21 March 2024; Accepted: 30 July 2024
Published online: 04 August 2024

References

1. Bachheti, A. *et al.* Bioactive constituents and health promoting compounds of underutilized fruits of the northern Himalayas of India: A review. *Food Prod. Process. Nutr.* **5**(1), 1–21. https://doi.org/10.1186/s43014-023-00140-5 (2023).
2. Meena, V. S. *et al.* Underutilized fruit crops of Indian arid and semi-arid regions: Importance, conservation and utilization strategies. *Horticulturae* **8**(2), 1–29. https://doi.org/10.3390/horticulturae8020171 (2022).

3. Alqasoumi, S. I., Basudan, O. A., Al-Rehaily, A. J. & Abdel-Kader, M. S. Phytochemical and pharmacological study of *Ficus palmata* growing in Saudi Arabia. *Saudi Pharm. J.* **22**(5), 460–471. https://doi.org/10.1016/j.jsps.2013.12.010 (2014).
4. Chauhan, P. K., Somesh, S., Chandrika, H. & Manisha, M. Evaluation of phytochemical and in-vitro antioxidant and antibacterial activities of wild plant species of Bauhinia and Ficus of HP. *World J. Pharm. Pharm. Sci.* **3**(4), 659–668 (2014).
5. Oza, M. J. & Kulkarni, Y. A. Traditional uses, phytochemistry and pharmacology of the medicinal species of the genus Cordia (Boraginaceae). *J. Pharm. Pharmacol.* **69**(7), 755–789. https://doi.org/10.1111/jphp.12715 (2017).
6. Hanan, E., Sharma, V. & Ahmad, F. J. Nutritional composition, phytochemistry and medicinal use of quince (*Cydonia oblonga* Miller) with emphasis on its processed and fortified food products. *Int. J. Food Process. Technol.* **11**, 831 (2020).
7. Zhang, L. *et al.* The UHPLC-QTOF-MS phenolic profiling and activity of Cydonia oblonga Mill. reveals a promising nutraceutical potential. *Foods* **10**(6), 1230. https://doi.org/10.3390/foods10061230 (2021).
8. Amulya, R. N., Adivappar, N., Shivakumar, B. S. & Satish, K. M. Studies on genetic variability and relationship of bael (*Aegle marmelos* (L) Correa) using morphological and molecular markers. *J. Hortic. Sci.* **17**(1), 88–94. https://doi.org/10.24154/jhs.v17i1.846 (2022).
9. Uddin, M. S., Islam, M. S., Alam, M. A. & Hossain, M. M. Study on physico morphological characteristics of 14 Bael (Aegle marmelos Corr) genotypes grown at Chapainawabganj, Bangladesh. *Int. J. Minor. Fruits Med. Arom. Plants* **2**(2), 29–33 (2016).
10. Khanal, A., Dall'acqua, S. & Adhikari, R. Bael (*Aegle marmelos*), an underutilized fruit with enormous potential to be developed as a functional food product: A Review. *J. Food Process. Preserv.* https://doi.org/10.1155/2023/8863630 (2023).
11. Pathirana, C. K. *et al.* Assessment of the elite accessions of bael [Aegle marmelos (L) Corr] in Sri Lanka based on morphometric, organoleptic, and elemental properties of the fruits and phylogenetic relationships. *Plos one.* **15**(5), 1–20. https://doi.org/10.1371/journal.pone.0233609 (2020).
12. Singh, S., Sharma, J. R., Sehrawat, S. K., Jitarwal, O. P. & Gavri, A. Studies on the collection and evaluation of Bael cultivars. *Int. J. Chem. Stud.* **8**(6), 212–214. https://doi.org/10.22271/chemi.2020.v8.i6c.10773 (2020).
13. Sharma, N. *et al.* Aegle marmelos (L.) correa: an underutilized fruit with high nutraceutical values: A review. *Int. J. Mol. Sci.* **23**(18), 10889. https://doi.org/10.3390/ijms231810889 (2022).
14. Singh, A. K. *et al.* Research status of bael (*Aegle marmelos*) in India: A review. *Indian J. Agric. Sci.* **89**(10), 1563–1571. https://doi.org/10.56093/ijas.v89i10.94576 (2019).
15. Singh, P. *et al.* Diversity in morpho-pomological attributes and biochemical profiling of bael (Aegle marmelos (L.) Correa) genotypes of North-Western India. *Heliyon* **10**(4), e26525. https://doi.org/10.1016/j.heliyon.2024.e26525 (2024).
16. Debbarma, P. & Hazarika, T. K. Genetic diversity of Bael [Aegle marmelos (L.) Corr.] accessions from north-east India based on principal component and cluster analysis. *Genet. Resour. Crop Evol.* https://doi.org/10.1007/s10722-023-01619-3 (2023).
17. Dhakar, M. K., Das, B., Nath, V., Sarkar, P. K. & Singh, A. K. Genotypic diversity for fruit characteristics in bael [Aegle marmelos (L.) Corr.] based on principal component analysis. *Genet. Resour. Crop Evol.* **66**, 951–964. https://doi.org/10.1007/s10722-019-00757-x (2019).
18. Benharrat, H., Véronési, C., Theodet, C. & Thalouarn, P. Orobanche species and population discrimination using intersimple sequence repeat (ISSR). *Weed Res.* **42**(6), 470–475. https://doi.org/10.1046/j.1365-3180.2002.00305.x (2002).
19. Jannatabadi, A. A., Talebi, R., Armin, M., Jamalabadi, J. G. & Baghebani, N. Genetic diversity of Iranian landrace chickpea (Cicer arietinum L.) accessions from different geographical origins as revealed by morphological and sequence tagged microsatellite markers. *J. Plant Biochem. Biotechnol.* **23**, 225–229. https://doi.org/10.1007/s13562-013-0206-x (2014).
20. Geleta, N., Labuschagne, M. T. & Viljoen, C. D. Genetic diversity analysis in sorghum germplasm as estimated by AFLP. SSR and morpho-agronomical markers. *Biodivers. Conserv.* **15**, 3251–3265. https://doi.org/10.1007/s10531-005-0313-7 (2006).
21. Mondini, L., Noorani, A. & Pagnotta, M. A. Assessing plant genetic diversity by molecular tools. *Diversity.* **1**(1), 19–35. https://doi.org/10.3390/d1010019 (2009).
22. Hrotko, K., Magyar, L. & Gyeviki, M. Evaluation of native hybrids of Prunus fruticosa Pall. as cherry interstocks. *Acta Agri Serbica.* **13**(25), 41–45 (2008).
23. Khadivi-Khub, A. & Anjam, K. *Prunus scoparia*, a suitable rootstock for almond (*Prunus dulcis*) under drought condition based on vegetative and fruit characteristics. *Sci. Hortic.* **210**, 220–226. https://doi.org/10.1016/j.scienta.2016.07.028 (2016).
24. Salgotra, R. K. & Stewart, C. N. Jr. Functional markers for precision plant breeding. *Int. J. Mol. Sci.* **21**(13), 4792. https://doi.org/10.3390/ijms21134792 (2020).
25. Mujeeb, F., Bajpai, P., Pathak, N. & Verma, S. R. Genetic diversity analysis of medicinally important horticultural crop Aegle marmelos by ISSR markers. *Methods Mol. Biol.* **1620**, 195–211. https://doi.org/10.1007/978-1-4939-7060-5_14 (2017).
26. Kholia, A. Biochemical and molecular characterization of diverse genotypes of Bael (*Aegle marmelos* Correa.) for their nutraceutical properties. [Ph.D. Thesis]: G.B. Pant University of Agriculture and Technology, Pantnagar, India (2018).
27. Shiqie, B. Ecological characteristics and morphological variations of centipedegrass in different populations. *J. Beijing For. Univ.* **24**(4), 97–101 (2002).
28. Sharma, C. K. & Sharma, V. Analysis of Aegle marmelos (L.) Corr diversity using citrus based microsatellite markers. *J. Appl. Hortic.* **17**(3), 217–221. https://doi.org/10.37855/jah.2015.v17i03.41 (2015).
29. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE, a simulation study. *Mol. Ecol.* **14**, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x (2005).
30. Singh, A. K. *et al.* Descriptors for characterization and evaluation of bael (Aegle marmelos (L.) Correa ex Roxb.) germplasm for utilization in crop improvement. *Genet. Resour. Crop. Evol.* https://doi.org/10.1007/s10722-024-01903-w (2024).
31. Pavani, P., Harshavardhan, G., Das, S. N. & Das, B. K. Evaluation of bael genotypes for biochemical characters. *Int. J. Curr. Microbiol. Appl. Sci.* **7**(6), 2930–2934. https://doi.org/10.20546/ijcmas.2018.706.344 (2018).
32. Pavani, P., Kiranmayi, P. & Dash, S. N. Evaluation of bael (*Aegle marmelos* Correa) genotypes for morphological, quality and yield related characters. *Int. J. Basic Appl. Biol.* **4**(3), 164–167 (2017).
33. Singh, A. K., Singh, S., Singh, R. S., Joshi, H. K. & Sharma, S. K. Characterization of bael (*Aegle marmelos*) varieties under rainfed hot semi-arid environment of western India. *Indian J. Agric. Sci.* **84**(10), 1236–1242 (2014).
34. Mani, A., Singh, A., Jain, N. & Misra, S. Flowering, fruiting and physio-chemical characteristics of bael (Aegle marmelos correa.) grown in northern districts of West Bengal. *Curr. J. Appl.* **23**(3), 1–8. https://doi.org/10.9734/CJAST/2017/36310 (2017).
35. Kumar, M., Sharma, S., Kumar, M., Nibhoria, A. & Yadav, B. S. Variability for phenological traits and fruit yield attributes in bael (*Aegle marmelos* Correa) cultivars under Semi-Arid Conditions. *Ekin J. Crop Breed. Genetic.* **9**(2), 119–125 (2023).
36. Chaturvedi, K. *et al.* Exploring the genetic diversity of Aegle marmelos (L) Correa populations in India. *Plant Genet. Res.* **21**(2), 107–114. https://doi.org/10.1017/S1479262123000485 (2023).
37. Kaur, A. & Kalia, M. Physico chemical analysis of bael (*Aegle marmelos*) fruit pulp, seed and pericarp. *Chem. rev. lett.* **6**(22), 1213–1218 (2017).
38. Sarkar, A., Rashid, M., Musarrat, M. & Billah, M. Phytochemicals and nutritional constituent evaluation of Bael (*Aegle marmelos*) fruit pulp at different development stage. *Asian Food Sci. J.* **20**(1), 78–86. https://doi.org/10.9734/afsj/2021/v20i130257 (2021).
39. Singh, U., Kocher, A. & Boora, R. Proximate composition, available carbohydrates, dietary fibres and anti-Nutritional factors in Bael (Aegle marmelos L.) leaf, pulp and seed powder. *Int. J. Sci. Res. Public.* **2**(4), 1–4 (2012).
40. Xin, Y. H. *et al.* Evaluation on the phenotypic diversity of Calamansi (*Citrus microcarpa*) germplasm in Hainan island. *Sci. Rep.* **12**(1), 371. https://doi.org/10.1038/s41598-021-03775-x (2022).

41. Choo, T. M. & Reinbergs, E. Analyses of skewness and kurtosis for detecting gene interaction in a doubled haploid population. *Crop Sci.* **22**(2), 231–235. https://doi.org/10.2135/cropsci1982.0011183X002200020008x (1982).
42. Forde, H. I. Walnuts. In *Advances in Fruit Breeding* (eds Janick, J. & Moore, J. N.) 439–455 (Purdue University Press, West Lafayette, 1975).
43. Rodriguez-Delgado, M. A., Gonzalez-Hernandez, G., Conde-Gonzalez, J. E. & Perez-Trujillo, J. P. Principal component analysis of the polyphenol content in young red wines. *Food Chem.* **78**(4), 523–532. https://doi.org/10.1016/S0308-8146(02)00206-6 (2002).
44. Samec, D. *et al.* Assessment of the differences in the physical, chemical and phytochemical properties of four strawberry cultivars using principal component analysis. *Food Chem.* **194**, 828–834. https://doi.org/10.1016/j.foodchem.2015.08.095 (2016).
45. Murty, B. R. & Arunachalam, V. The nature of genetic divergence in relation to breeding system in crop plants. *Indian J. Genet. Plant Breed.* **26**, 316–321 (1966).
46. Varshney, R. K., Hoisington, D. A. & Tyagi, A. K. Advances in cereal genomics and applications in crop breeding. *Trends Biotechnol.* **24**(11), 490–499. https://doi.org/10.1016/j.tibtech.2006.08.006 (2006).
47. Guan, C. *et al.* Germplasm conservation, molecular identity and morphological characterization of persimmon (Diospyros kaki Thunb) in the NFGP of China. *Sci. Hortic.* **272**, 109490. https://doi.org/10.1016/j.scienta.2020.109490 (2020).
48. Wang, L. *et al.* Genetic diversity among wild androecious germplasms of Diospyros kaki in China based on SSR markers. *Sci. Hortic.* **242**, 1–9. https://doi.org/10.1016/j.scienta.2018.07.020 (2018).
49. El Zayat, M. A. S., Hassan, A. H., Nishawy, E., Ali, M. & Amar, M. H. Patterns of genetic structure and evidence of Egyptian citrus rootstock based on informative SSR, LTR-IRAP and LTR-REMAP molecular markers. *J. Genet. Eng. Biotechnol.* **19**, 1–14. https://doi.org/10.1186/s43141-021-00128-z (2021).
50. Bernard, A., Barreneche, T., Lheureux, F. & Dirlewanger, E. Analysis of genetic diversity and structure in a worldwide walnut (*Juglans regia* L) germplasm using SSR markers. *PLoS One.* **13**(11), 0208021. https://doi.org/10.1371/journal.pone.0208021 (2018).
51. Patel, A. *et al. Guidelines for the conduct of test for distinctiveness, uniformity and stability of bael (Aegle marmelos Correa)* 1–25 (Protection of plant varieties and farmer's right authority, 2011).
52. Mahajan, R. K. *et al.* Minimal descriptors of agri-horticultural crops. Part III: Fruit crops. National Bureau of Plant Genetic Resources. 27–31 (2002).
53. OriginPro 9.1. OriginLab Corporation, One Roundhouse Plaza, Suite 303 Northampton, MA 01060, United States. 1800-969-7720 (2022). www.OriginLab.com.
54. Doyle, J. J. & Doyle, J. L. A rapid DNA isolation procedure to small amounts of fresh leaf tissue. *Phytochemical Bull.* **19**, 11–15 (1987).
55. Gupta, C., Salgotra, R. K., Venegas, R. A., Mahajan, R. & Koul, U. Genetic diversity and marker trait association for yield attributing traits in accessions of common bean (*Phaseolus vulgaris* L.) in India. *Plant Genet Res* **20**(2), 98–107. https://doi.org/10.1017/S147926212200017X (2022).
56. Liu, K. & Muse, S. V. Power marker: An integrated analysis environment for genetic marker analysis. *Bioinformatics.* **21**, 2128–2129. https://doi.org/10.1093/bioinformatics/bti282 (2005).
57. Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**, 314–331 (1980).
58. Nei, M. *Molecular Evolutionary Genetics* (Columbia University Press, 1987).
59. Perrier, X. & Jacquemoud-Collet, J. P. DARwin software (2006). http://darwin.cirad.fr/darwin.
60. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genet.* **155**, 945–959. https://doi.org/10.1093/genetics/155.2.945 (2000).
61. Earl, D. A. & VonHoldt, B. M. Structure harvester: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**, 359–361. https://doi.org/10.1007/s12686-011-9548-7 (2012).
62. Peakall, R. & Smouse, P. E. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research–an update. *Bioinform.* **28**, 2537–2539. https://doi.org/10.1111/j.1471-8286.2005.01155.x (2012).

Author contributions

P.S. performed the experiment and wrote the main manuscript; A.S. and R.K.S. designed the experiment; M.S. analysed the statistical data; V.T., V.G. and D.S. reviewed the manuscript. All authors contributed to the article.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-69030-1.

Correspondence and requests for materials should be addressed to P.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.