



Genetic diversity and population structure analysis of Indian blackberry (*Syzygium cumini* L.) using CAAT box-derived polymorphism (CBDP) and start codon targeted polymorphism (SCoT) markers

Ajay Kumar^{a,b}, Kanhaiya Singh^{b,*}, Amit Kumar Singh^{c,*}, Jai Prakash^b,
Amit Kumar Goswami^b, Gyan Prakash Mishra^d, Vishaw Bandhu Patel^e, Suman Lata^f,
Anshuman Singh^g

^a The Graduate School, ICAR-Indian Agricultural Research Institute, New Delhi 110012, India

^b Division of Fruits and Horticultural Technology, ICAR-Indian Agricultural Research Institute, New Delhi 110012, India

^c ICAR-National Bureau of Plant Genetic Resources, New Delhi 110012, India

^d Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi 110012, India

^e Horticultural Science Division, ICAR-Krishi Anusandhan Bhawan-II, New Delhi 110012, India

^f Division of Vegetable Science, ICAR-Indian Agricultural Research Institute, New Delhi 110012, India

^g ICAR- Central Institute for Subtropical Horticulture, Lucknow 226101, India

ARTICLE INFO

Keywords:

AMOVA

CBDP

Genetic diversity

Gene-targeted markers

Population structure

SCoT

Syzygium cumini (L.)

ABSTRACT

Indian blackberry (*Syzygium cumini* L.) also known as jamun is a very important underutilized fruit crop with notable medicinal and economic value. However, its genetic improvement has been constrained by limited knowledge of the genetic diversity within existing collections. Therefore, a comprehensive characterization of genetic diversity in this species, using molecular tools, is essential to support effective germplasm management and application in breeding programs. In this investigation, a total of 32 jamun genotypes consisting of 30 seedling-origin genotypes, one improved cultivar CISH J-37 and one wild genotype (*Syzygium frutescens*) were analysed using the two gene-targeted markers, CBDP and SCoT. In total, 29 primers (22 CBDP and 7 SCoT primers) detected genetic polymorphism across the genotypes. The CBDP markers amplified a higher polymorphism percentage, 94.85% across 291 bands, than the SCoT markers, 92.75% across 69 bands. The mean PIC values for CBDP and SCoT were 0.28 and 0.31, respectively. MI values were higher for CBDP (3.21) than for SCoT (2.88). Cluster analysis using UPGMA identified six clades, which grouped genotypes into seedling-origin, improved and wild categories. The PCoA based on molecular profiling data of CBDP, SCoT and both together explained 26.65%, 38.39% and 23.22% of the variation respectively. AMOVA results revealed that 85–90% of genetic variation existed within populations. Bayesian STRUCTURE analysis grouped genotypes into two major populations confirming genetic divergence between seedling-origin, improved and wild genotypes. This study is the first to integrate CBDP and SCoT markers for genetic diversity analysis of the Indian blackberry. The results highlight the utility of these markers in genetic variation assessment and would help design germplasm conservation and breeding strategies in this crop.

1. Introduction

Indian blackberry (*Syzygium cumini* L.) is an underutilized fruit crop of the Myrtaceae family. It is known globally by various names, such as jamun, jambul, danson plum, jambolana, and black plum.^{1–2} It is native to India and has a wide distribution across the Indian subcontinent and many regions of South Asia, including India, Burma (Myanmar), Nepal,

Sri Lanka, Bangladesh and Pakistan.³ Additionally, it has been successfully domesticated in numerous countries across Asia, America and Africa which shows its adaptability and horticultural significance.⁴ As one of the hardiest fruit crops, it can be easily cultivated in marshy and neglected fields where other fruit plants are difficult to grow commercially.⁵ The purplish berry-like fruits contain a single seed and are consumed as fresh and processed into various products like juice, jam,

* Corresponding authors.

E-mail addresses: kanhaiya.singh@icar.gov.in (K. Singh), amit.singh5@icar.gov.in (A.K. Singh).

<https://doi.org/10.1016/j.jgeb.2025.100468>

Received 19 November 2024; Received in revised form 23 January 2025; Accepted 4 February 2025

1687-157X/© 2025 The Author(s). Published by Elsevier Inc. on behalf of Academy of Scientific Research and Technology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

jelly, vinegar, wine, pudding ice cream and jamun strips.⁶⁻⁷ The fruits, seeds, bark and leaves of this crop are traditionally utilized in Ayurvedic, Unani and Chinese medicinal systems as a natural treatment for hyperglycemia, dysentery, glycosuria, asthma, bronchitis and ulcers.⁸ Notably, it has been found as an effective remedy for diabetes.⁹ The plant's various parts possess significant health activities including anti-anemic, antioxidant, antibacterial, antiallergic, hypolipidemic, hepatoprotective and antipyretic.¹⁰ Due to their potential for use as both food and healthcare, as well as their outstanding biological activities, the genus *Syzygium* is recognized for its economic significance.¹¹

Leveraging genetic diversity and understanding the relationships within crop germplasm collections is essential for their effective management, application in breeding programs, and enhancing the success of breeding efforts.¹²⁻¹⁴ A greater range of molecular markers are utilized to detect diversity at the genetic level across different plant species. In underutilized crops, genetic diversity studies have primarily employed non-locus specific markers, such as RAPD and ISSR which have several disadvantages such as low reproducibility, non-specific binding, lower genetic resolution and limited insight into trait association. To overcome these challenges many novel marker techniques were developed which unravel high polymorphism from the genic or functional region of the genome and thus could provide more comprehensive information on genomic level genetic diversity present in crop germplasm collections.¹⁵⁻¹⁸ Two key gene-based markers CAAT box-derived polymorphism (CBDP) and start codon targeted (SCoT) are widely utilized in genetic diversity studies of crops. Of these two, CBDP was developed by our team and utilizes promoter-targeted primers to take advantage of the CAAT box consensus sequence within promoters of plant genes.¹⁹ On the other hand, SCoT markers emerge as another promising technique that draws from the conserved regions of DNA adjacent to the ATG start codon within the genes of plants. Hence, it can capture genetic differences within a particular gene linked to specific traits.²⁰ These techniques are gaining popularity for molecular marker-based studies in crops due to their cost-effectiveness, high polymorphism and reproducibility over traditional dominant markers. Additionally, since genomic sequence information is not prerequisite for primer designing, these markers can be applied to any crop species.^{19,21-22} The CBDP marker technique was first time validated in jute, linseed and cotton cultivars.¹⁹ Subsequently, it has been utilized for genetic diversity and gene mapping studies in many crop species.²³⁻²⁵ Additionally, SCoT markers have proven effective in assessing genetic diversity, DNA fingerprinting and determining population structure in several horticultural plant species.²⁶⁻³²

In recent years many studies have combined these CBDP and SCoT gene-targeted markers for generating comprehensive insights into genetic diversity and population structure in different plant species including *Simmondsia chinensis*,³³ *Andrographis paniculata*,³⁴ *Triticum durum*,¹⁶ *Prosopis cineraria*,³⁵ *Bauhinia racemosa* Lam.,³⁶ *Clerodendrum serratum* L.,³⁷ *Aegilops triuncialis*,³⁸ *Triticum aestivum* L.,³⁹ *Mangifera indica* L.,⁴⁰ *Hordeum vulgare* L.,⁴¹ *Cannabis sativa* L.,⁴² *Centaurea species*,⁴³ *Abelmoschus esculentus* L.,⁴⁴ and *Senna alexandrina* Mill.⁴⁵ These markers efficiently determined genetic relationships, clustering and population structures.^{39,42} Their combined application demonstrated comprehensive patterns of genetic diversity, making them more valuable than single markers application in the study of crop diversity.⁴⁰ Positive markers correlations have been found reliable and accurate for the analysis of genetic variability.³⁷ SCoT and CBDP markers were effective tools for evaluating variability and relatedness, even under complex classification conditions.⁴³

Despite advances in molecular marker technology, their use in underutilized crops like jamun remains limited, creating a huge knowledge gap in the genetic diversity of this crop. Previous studies have mainly used RAPD,⁴⁶⁻⁴⁸ ISSR⁴⁹⁻⁵⁰ and RAPD + ISSR.⁵¹⁻⁵⁴ To date, just one report has utilized CBDP markers to explore genetic diversity in this crop.⁵⁵ This necessitates characterizing the genetic diversity of existing jamun collections with different sets of reliable markers such as

gene-targeted markers. In this context, we hypothesize that the combination of CAAT box-derived polymorphism (CBDP) and start codon targeted (SCoT) polymorphism markers will significantly enhance the detection of genetic diversity and population structure in jamun collections. Therefore, in the current study, we first time combined both gene-targeted markers (CBDP and SCoT) to evaluate their effectiveness in assessing genetic diversity and a comprehensive understanding of the genetic relationships among jamun genotypes.

2. Material and methods

2.1. Sampling

The study was conducted on a large natural population of jamun spread across the Pusa campus, New Delhi, situated in the semi-arid tropics of India. A total of 31 genotypes have been selected from different 11 sites (Fig. 1). Besides, one improved genotype namely CISH J-37 (Jamwant) was also included in our study. Two to four samples were randomly obtained at each site, with a minimum distance of 20 m between all sampled individuals. Table 1 contains detailed information about the longitudes and latitudes of each genotype.

2.2. Extraction of genomic DNA

Fresh, disease-free leaves were explicitly collected to extract total genomic DNA from each sample following cetyltrimethylammonium bromide (CTAB) procedure by Doyle and Doyle,⁵⁶ with some modifications. The main modification was the addition of 3 % of polyvinyl pyrrolidone (PVP) in the extraction buffer to remove phenolic content from leaf samples. The extracted DNA's integrity and purity were assessed using 1 % agarose gel and its concentration was measured through spectrophotometry. Finally, a working genomic DNA stock of each genotype was prepared with a final DNA concentration of 50 ng/μl.

2.3. PCR of CBDP markers

The study used a total of 40 CBDP primers (Table S1) including 25 previously known and another 15 newly designed by following the criteria.¹⁹ Initially, these primers were screened in a set of four different jamun genotypes and those generating polymorphic bands were selected for further study (Table 2). The polymerase chain reaction (PCR) was performed in a total volume of 20 μl reaction containing 50 ng of template DNA, 1X PCR buffer with 1.5 mM MgCl₂ (Sigma Aldrich, India), 250 mM of each dNTPs, 0.5 μM CBDP primer and 1 U of Taq DNA polymerase (Sigma Aldrich, India). The amplification program consisted of an initial five cycles of denaturation at 94°C for 1 min, annealing at 35°C for 1 min and extension at 72°C for 1 min. Following these initial cycles, the annealing temperature was then increased to 50 °C for 35 cycles, followed by a final extension at 72 °C for 10 min.

2.4. PCR of SCoT markers

The SCoT primers (Table S2)²⁰ were synthesized from Eurofins Genomics (USA) and their sequences are provided in Table 3. PCR was set up in a reaction volume of 20 μl, containing 1X PCR buffer with 1.5 mM MgCl₂ (Sigma Aldrich, India), 250 μM of each dNTP, 0.8 μM of primer, 1 U of Taq DNA polymerase (Sigma Aldrich, India) and 50 ng of DNA template. The amplification was performed in a thermocycler (Biometra Professional Thermocycler, Germany). The PCR program consisted of 35 cycles, with each cycle including denaturation at 94 °C for 1 min, annealing at 50 °C for 1 min and a 2-minute extension step. At the end, a final extension was carried out at 72°C for 5 min to ensure complete synthesis of all PCR fragments.

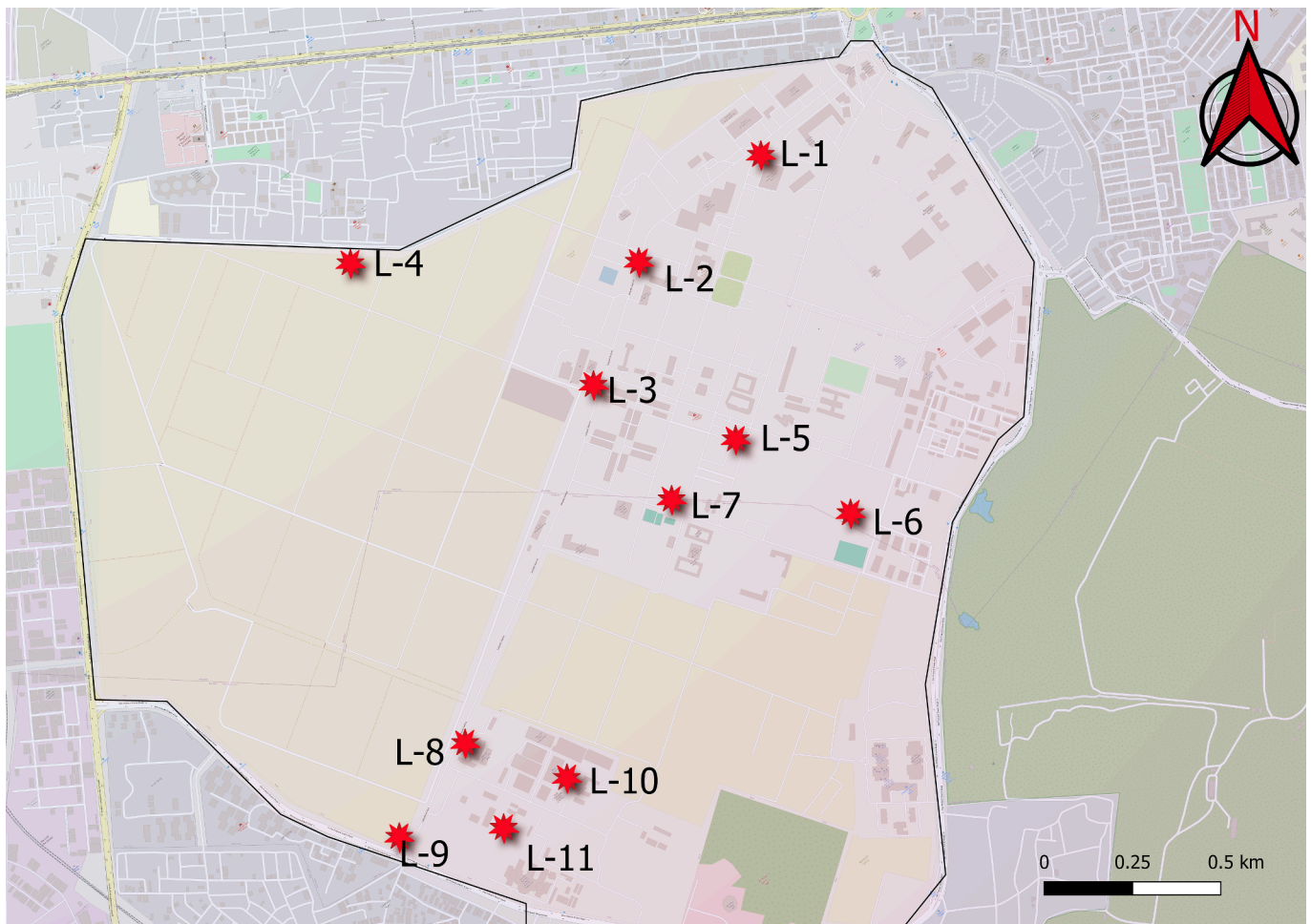


Fig. 1. Map illustrating jamun sample collection sites.

2.5. Electrophoresis and gel documentation

The amplified bands generated with both CBDP and SCoT markers were analysed on a 1 % agarose gel using electrophoresis in 1X TBE buffer. The amplified fragments were stained with ethidium bromide, and a standard molecular size marker (Thermo Scientific, USA) was used to accurately determine their sizes. The images of gels were recorded in a gel imaging system (Syngene, UK).

2.6. Statistical analysis of molecular data

The PCR gel images of CBDP and SCoT markers were visually analysed and reproducible bands were focused for scoring. A binary matrix was generated for the presence (1) and absence (0) of bands and used for various statistical and molecular analyses. To find out the effectiveness of the marker, many parameters such as polymorphic information content (PIC), marker index (MI) and resolving power (Rp) were used. PIC was computed by the following formula $PIC = 2\sum f_i(1 - f_i)$ ⁵⁷. The MI was analysed as suggested.⁵⁸ The Rp describes a primer's ability to detect variability between genotypes and was determined using the method.⁵⁹ Genetic diversity indices Nei's gene diversity index (h), the number of effective alleles (ne) and Shannon diversity index (I) were estimated using the software GenAlEx v.6.⁶⁰ Phylogenetic trees were obtained after carrying out hierarchical clustering. This was done with the Un-weighted Pair Group Method with Arithmetic Mean (UPGMA), as per Jaccard similarity coefficients. The GenAlEx v.6.5 program was used to perform PCoA and AMOVA analyses.⁶⁰ Population structure was estimated by the STRUCTURE software program (version 2.3.4).⁶¹ To

obtain the exact estimate of the population a total of 10 runs of STRUCTURE were performed for each k population ($K = 1-10$). For all runs, we set a burn-in of 10,000 iterations after which 100,000 MCMC repetitions were run. The optimal number of clusters (k) was further determined using the ΔK method.⁶² and this output was processed through a web-based platform using STRUCTURE HARVESTER v0.6.94.

3. Results

3.1. CBDP and SCoT markers polymorphism statistics

The genetic diversity among the jamun genotypes was examined by 22 CBDP and 7 SCoT markers. The representative amplification profiles of 32 studied jamun genotypes with some CBDP and SCoT markers are presented in Fig. 2.

CBDP primers amplified a total of 291 bands out of which 276 were polymorphic and broadly revealed 94.85 % polymorphism among the studied jamun genotypes. The total number of bands (TNB) per primer ranged from 7 (CBDP-19) to 16 (CBDP-7 and CBDP-14), whereas the percentage polymorphic bands (PPB) ranged from 85.71 % for CBDP-19 to 100 % for most of the primers. The Rp values ranged from 3.75 in the case of CBDP-19 to 11.43 for CBDP-6, indicating efficiency in differentiating the genotypes by these primers. The PIC values ranged from 0.145 for CBDP-6 to 0.343 for CBDP-14, with an average PIC value of 0.277. They also reflected a mean marker index of 3.21, out of which the highest, 5.498, was contributed by CBDP-14 (Table 2).

SCoT primers generated 69 bands, out of which 64 were polymorphic, hence giving the percentage of PPB to 92.75 %. The number of

Table 1
Detailed information about the longitudes and latitudes of 32 Indian blackberry genotypes.

Site	Genotype	Information	Latitude	longitude
Location-1	PCJ-1	Seedling origin	28.64051010	77.16974820
	PCJ-2	Seedling origin	28.64079330	77.16818850
	PCJ-30	Seedling origin	28.64059880	77.16891190
Location-2	PCJ-3	Seedling origin	28.64089260	77.16689040
	PCJ-4	Seedling origin	28.64071590	77.16604070
	PCJ-5	Seedling origin	28.64070790	77.16543900
Location-3	PCJ-6	Seedling origin	28.64011710	77.16475400
	PCJ-7	Seedling origin	28.63910420	77.16301150
	PCJ-8	Seedling origin	28.63865840	77.16185110
Location-4	PCJ-12	Seedling origin	28.63762700	77.16108800
	PCJ-13	Seedling origin	28.63710300	77.16106800
	PCJ-9	Seedling origin	28.64495900	77.15882000
Location-5	PCJ-10	Seedling origin	28.64595600	77.15770400
	PCJ-11	Seedling origin	28.63717270	77.16309700
	PCJ-14	Seedling origin	28.63564700	77.16383240
Location-6	PCJ-15	Seedling origin	28.63530490	77.16373220
	PCJ-16	Seedling origin	28.63145900	77.16514300
	PCJ-17	Seedling origin	28.63242700	77.16281690
Location-7	PCJ-29	Seedling origin	28.63314600	77.16549200
	PCJ-18	Seedling origin	28.63516050	77.16122150
	PCJ-19	Seedling origin	28.63795560	77.16014060
Location-8	PCJ-20	Seedling origin	28.63401590	77.15229990
	PCJ-21	Seedling origin	28.63382300	77.15148900
	<i>S. frutescens</i>	Wild type	28.63448820	77.15331960
Location-9	PCJ-22	Seedling origin	28.63284790	77.14905820
	PCJ-23	Seedling origin	28.63190800	77.15232560
	PCJ-24	Seedling origin	28.63129140	77.15338500
Location-10	PCJ-25	Seedling origin	28.63131490	77.15401320
	PCJ-26	Seedling origin	28.63115300	77.15472030
	PCJ-27	Seedling origin	28.62994280	77.15314920
Location-11	PCJ-28	Seedling origin	28.63023110	77.15256030
	CISH-37	Cultivated type		

bands per primer ranged from 6 (SCoT-21) to 13 (SCoT-13). PPB ranged from 77.78 % for SCoT-6 to 100 % for multiple primers like SCoT-3, SCoT-12, SCoT-13, and SCoT-21. The Rp of SCoT primers ranged from 3.82 (SCoT-21) to 7.21 (SCoT-13) with a mean Rp of 5.28, suggesting their usefulness in distinguishing jamun genotypes. The PIC values for SCoT primers ranged from 0.244 (SCoT-6) to 0.364 (SCoT-12). The marker index varied from 1.547 for SCoT-21 to 4.187 for SCoT-13 with a general mean MI value of 2.88 for all the generated markers (Table 3).

The CBDP and SCoT markers efficiently produced unique amplicons among the genotypes. *S. frutescens* showed unique amplicons by multiple markers, such as 2000 bp (CBDP-3), 1500 bp (CBDP-5), 2500 bp, 1400 bp, 650 bp, 500 bp, 450 bp (CBDP-6), 2000 bp (CBDP-10), 2200 bp, 1600 bp (CBDP-17), 800 bp, 550 bp (CBDP-25). These unique amplicons revealed its distinct genetic profile. CISH J-37 recorded unique bands of 2000 bp (CBDP-4) and 3000 bp (CBDP-13). Seedling-origin genotypes also showed unique amplicons, including 2050 bp, 1700 bp, 1550 bp by CBDP-1 exclusive to PCJ-9, 850 bp by CBDP-9 in PCJ-14, a unique amplicon of 3200 bp in PCJ-21 by CBDP-16 and 790 bp unique amplicon by CBDP-25, exclusive to PCJ-2 (Table 2). In the case of SCoT markers, SCoT-13 generated a unique amplicon of 900 bp in PCJ-8 and two unique amplicons (600 bp and 480 bp) in CISH J-37. Moreover, a unique amplicon of 900 bp was produced by SCoT-21 in PCJ-16 (Table 3).

3.2. Diversity indices of CBDP and SCoT markers

The present study showed significant genetic diversity across the CBDP and SCoT markers, as evidenced by different diversity indices (Table 4), (Table 5). Ne for CBDP markers ranged from a minimum value of 1.25 for CBDP-6 to a maximum of 1.60 for CBDP-25 with a mean value of 1.45 suggesting high genetic diversity. Gene diversity, h averaged between a minimum of 0.1457 for CBDP-6 and a maximum of 0.3484 for CBDP-25 with a mean of 0.2700. These values suggest the existence of

high diversity across loci. Shannon diversity index, I varied from 0.2377 for CBDP-6 to 0.522 for CBDP-25 with an average of 0.4151, which further shows a high level of genetic diversity among loci. Among the CBDP markers, the highest genetic diversity was revealed for CBDP-25 ($h = 0.3484$ and $I = 0.5218$), whereas the lowest values were recorded for CBDP-6 ($h = 0.1457$ and $I = 0.2377$) (Table 4).

For SCoT markers, Ne ranged from 1.42 (SCoT-21) to 1.64 (SCoT-12) with an average of 1.52, a little higher than the allelic diversity of the CBDP markers. Gene diversity (h) was from 0.2446 (SCoT-6) to 0.3645 (SCoT-12) and the average is 0.3081, showing greater diversity within loci. Shannon's index ranged from 0.369 in SCoT-6 to 0.539 in SCoT-12, with a mean of 0.4640, indicating more cumulative diversity in SCoT markers overall. The maximum genetic diversity was seen in the case of SCoT-12 showing $h = 0.3645$ and $I = 0.5393$. However, the lowest was found in SCoT-6 showing $h = 0.2446$ and $I = 0.3690$ (Table 5). The higher mean gene diversity and Shannon's index values of SCoT markers than CBDP indicate wider genetic variability across the loci analysed by SCoT markers.

3.3. Cluster analysis using CBDP, SCoT and combined markers

The genetic relationships among 32 jamun genotypes were evaluated using dendrograms based on CBDP, SCoT, and combined markers (Fig. 3). All analyses revealed six major clades, highlighting genetic diversity among seedling-origin, improved and wild genotypes. The UPGMA dendrogram generated using CBDP data, displayed moderate to high genetic variance with genetic distances ranging from 0.050 to 0.3692. Clade-I and Clade-II grouped most of the seedling origin genotypes like PCJ-16, PCJ-24 and PCJ-12 with distances of 0.0123 to 0.2662 that showed moderate diversity. CISH J-37 formed a separate sub-cluster in Clade-III with a genetic distance of 0.2837 showing divergence because of genetic improvement. Moreover, *S. frutescens* was expected placed as a genetic outlier in the Clade-VI, having the maximum genetic distance of 0.3692.

Similarly, the UPGMA dendrogram constructed with SCoT data showed genetic distance among jamun genotypes from 0.050 to 0.2924, which is similar to CBDP findings. Like CBDP-based clustering, Clade-I in this dendrogram started with seedling-origin genotypes, PCJ-16, PCJ-24 and PCJ-12 clustered at a distance of 0.0125–0.1866 while PCJ-23 and PCJ-8 got placed in Clade-IV with a very low divergence of 0.1429. Clade-VI was assigned to *S. frutescens*, whose wild genotype status was further explained by a maximum genetic distance of 0.2924, while CISH J-37 stayed in Clade-III with a genetic distance of 0.2510.

The combined-set analysis established six clades. Clade-I shows genotypes of seedling origin clustered between 0.1728 and 0.2074 and Clade-IV included PCJ-23 and PCJ-8 at 0.1467 genetic distance. On the other hand, the improved cultivar (CISH J-37), was genetically separated from genotypes of seedling origin by 0.2981 and mapped to Clade-V. *S. frutescens* emerged as the outlier in Clade-VI, showing a genetic distance of 0.3520.

3.4. Principal coordinates analysis (PCoA) and analysis of molecular variance (AMOVA)

Principal Coordinates Analysis (PCoA) was separately performed for CBDP, SCoT and combined both markers data which explained genetic variation of 26.65 %, 38.39 % and 23.22 %, respectively (Fig. 4). PCoA analysis based on CBDP marker data revealed three clusters. The first cluster contained closely related seedling-origin genotypes, CISH J-37 whereas, *S. frutescens* formed different clusters for the CBDP marker. There was a considerable genetic difference among the seedling-origin genotypes, CISH J-37 and *S. frutescens*. Additionally, three groups in the SCoT study display genotypes of seedling origin in cluster 2 (PCJ-19, PCJ-22, and PCJ-28) and cluster 1 (PCJ-6, PCJ-23, and PCJ-13), with minimal intermixing. *S. frutescens*, though, remained alone. The proximity of CISH J-37 to *S. frutescens* reveals its

Table 2
Details of CBDP primers sequences and polymorphism statistics evaluated across markers.

Primer	Sequence	TNB	NPB	PPB	PIC	MI	Rp	Band Size (bp)	Unique Amplicon
CAAT-1	TGAGCAGATCCAAT AGC	12	12	100.00	0.25	2.94	7.85	450–2050	3; 1550 bp, 1700 bp, 2050 bp in PCJ
CAAT-2	TGAGCAGATCCAATAAT	12	11	91.67	0.31	3.41	6.37	300–2000	–
CAAT-3	TGAGCAGATCCAAT ACC	11	11	100.00	0.28	3.11	6.24	500–2050	2; 1700 bp in PCJ-9; 20000 bp <i>S. fruitecosum</i>
CAAT-4	TGAGCAGATCCAAT AAG	15	15	100.00	0.28	4.16	9.60	480–1700	1; 2000 bp in CISH-37
CAAT-5	TGAGCAGATCCAAT CTA	8	8	100.00	0.31	2.52	4.54	300–1100	1; 1500 bp in <i>S. fruitecosum</i>
CAAT-6	TGAGCAGATCCAAT CAG	15	14	93.33	0.15	2.04	11.43	350–2500	6; 450 bp, 500 bp, 650 bp, 1400 bp, 2500 bp in <i>S. fruitecosum</i> and 950 bp in PCJ-5
CAAT-7	TGAGCAGATCCAAT CGA	16	15	93.75	0.27	4.05	9.44	350–2400	–
CAAT-9	TGAGCAGATCCAAT GAT	13	12	92.31	0.24	2.90	8.83	450–2200	1; 850 bp in PCJ-14
CAAT-10	TGAGCAGATCCAAT GTT	11	10	90.91	0.26	2.61	6.40	300–2000	1; 2000 bp in <i>S. fruitecosum</i>
CAAT-12	TGAGCAGATCCAATATA	14	13	92.86	0.28	3.69	7.99	400–2000	2; 400 bp, 500 bp in PCJ-6
CAAT-13	TGAGCAGATCCAATGAG	10	10	100.00	0.27	2.70	5.49	450–3000	1; 3000 bp in CISH-37
CAAT-14	TGAGCAGATCCAATGCG	16	16	100.00	0.34	5.50	8.48	700–3000	–
CAAT-16	TGAGCAGATCCAATTGA	14	13	92.86	0.22	2.82	10.06	600–3000	3; 1850 bp, 1900 bp in PCJ-12 and 3200 bp in PCJ-21
CAAT-17	TGAGCAGATCCAATTTG	9	9	100.00	0.23	2.07	5.97	800–2200	2; 1600 bp, 2200 bp in <i>S. fruitecosum</i>
CAAT-18	CTGAGCAGATCCAATAG	10	10	100.00	0.29	2.86	6.09	500–2000	–
CAAT-19	CTGAGCAGATCCAATAC	7	6	85.71	0.29	1.71	3.75	550–2200	1; 2500 bp in PCJ-15
CAAT-20	CTGAGCAGATCCAATAT	8	7	87.50	0.24	1.66	5.52	550–2000	–
CAAT-22	CTGAGCAGATCCAATCG	15	15	100.00	0.27	4.08	9.39	800–2200	–
CAAT-25	CTGAGCAGATCCAATGT	15	15	100.00	0.35	5.23	7.04	400–2100	3; 550 bp 800 bp in <i>S. fruitecosum</i> and 790 bp in PCJ-2
CAAT-28	TGCGTGACACCAATTAG	9	9	100.00	0.24	2.12	6.08	400–1600	–
CAAT-35	TAGACGTGCTACCAATAC	14	14	100.00	0.24	3.29	9.33	500–2050	1; 3000 bp in PCJ-2
CAAT-36	TAGACGTGCTACCAATAT	13	13	100.00	0.24	3.18	8.70	400–1600	–
	Average	12	11	96.00	0.27	3.12	7.48	300–3000	–

TNB, total number of bands; **NPB**, number of polymorphic bands; **PPB**, percentage of polymorphic bands; **PIC**, polymorphism information content; **MI**, marker index; **Rp**, resolving power.

Table 3
Details of SCoT primers sequences and polymorphism statistics evaluated across markers.

Primer	Sequence	TNB	NPB	PPB	PIC	MI	Rp	Band Size (bp)	Unique Amplicon
ScoT-1	CAACAATGGCTACCACCA	9	8	88.89	0.34	2.74	4.64	200–1400	–
ScoT-2	CAACAATGGCTACCACCC	10	8	80.00	0.30	2.39	5.62	700–3000	–
ScoT-3	CAACAATGGCTACCACCG	12	12	100.00	0.32	3.86	6.54	700–3000	–
ScoT-6	CAACAATGGCTACCACGC	9	7	77.78	0.24	1.71	5.66	450–1800	–
ScoT-12	ACGACATGGCGACCAACG	10	10	100.00	0.36	3.64	4.39	500–2500	–
ScoT-13	ACGACATGGCGACCATCG	13	13	100.00	0.32	4.19	7.21	700–3000	3; 480 bp, 600 bp in CISH-37 and 900 bp in PCJ-8
ScoT-21	ACGACATGGCGACCCACA	6	6	100.00	0.26	1.55	3.82	350–900	1; 900 bp in PCJ-16
	Average	10	9	92.00	0.31	2.87	5.41	200–3000	–

TNB, total number of bands; **NPB**, number of polymorphic bands; **PPB**, percentage of polymorphic bands; **PIC**, polymorphism information content; **MI**, marker index; **Rp**, resolving power.

divergence from the seedling genotypes. Furthermore, the combined analysis displayed four clusters in which CISH J-37 constituted a distinct cluster. The genotypes of seedling origin separated into two clusters, alongside *S. fruitecosum* remaining an outlier in cluster 4 due to its extreme genetic variation.

Analysis of Molecular Variance (AMOVA) for CBDP markers showed a 16 % variation among populations (Table 6). At the same time, 84 % of the total variation is seen within populations. In contrast, SCoT markers revealed that there was 90 % molecular variation within the population and 10 % molecular variation among the population. Further, combined analysis of both CBDP and SCoT markers revealed 15 % of the variation among the population and 85 % within the population (Fig. 5).

3.5. Population structure analysis

The population structure of the 32 jamun genotypes was assessed using a combined dataset of CBDP and SCoT markers. Delta K analysis led to an optimal K value of 2. Therefore, the genotypes were divided into two major populations. Consistent with the two major genetic groups, the Mean LnP (K) plot for those corresponds to a very low variation in likelihood for K over two. Again, two distinct populations were seen in the admixture plot (Fig. 6). Here, the first population was mostly represented by seedling-origin genotypes like, PCJ-16, PCJ-24 and PCJ-12. These genotypes showed lower levels of admixture and thus pointed to a significant degree of genetic homogeneity in this

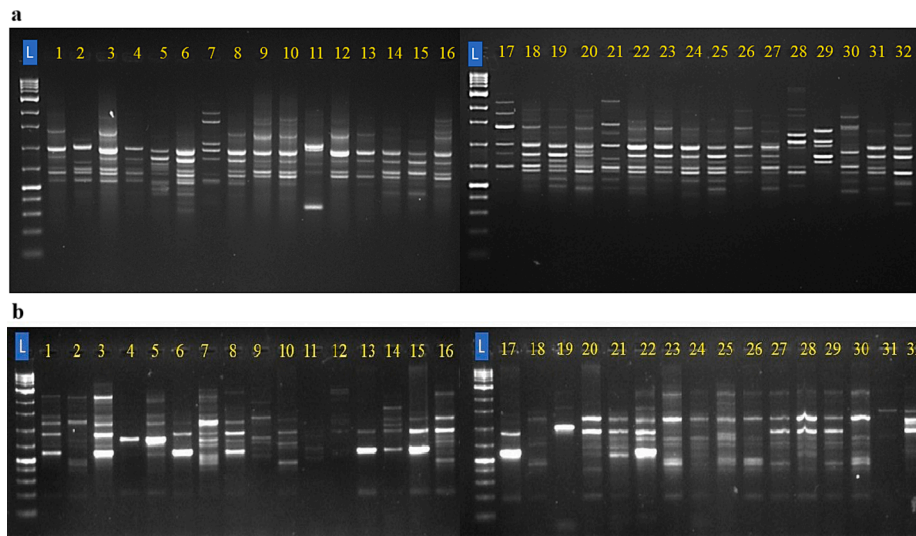


Fig. 2. Representative gel images showing the profile of 32 jamun genotypes by using CDBP-4 (a) and SCoT-12 (b) (L: 1 kb DNA ladder).

Table 4
Different diversity indices analysis of CDBP markers.

Sl. No	Primer name	ne	h	I
1.	CAAT-1	1.3988	0.2451	0.3831
2.	CAAT-2	1.5382	0.3101	0.4627
3.	CAAT-3	1.499	0.2824	0.4239
4.	CAAT-4	1.4195	0.2772	0.4409
5.	CAAT-5	1.4989	0.3144	0.4854
6.	CAAT-6	1.2504	0.1457	0.2377
7.	CAAT-7	1.4556	0.2697	0.4158
8.	CAAT-9	1.3723	0.2415	0.3836
9.	CAAT-10	1.4654	0.2611	0.3925
10.	CAAT-12	1.4819	0.2835	0.4293
11.	CAAT-13	1.4823	0.2700	0.4113
12.	CAAT-14	1.5556	0.3436	0.5220
13.	CAAT-16	1.3155	0.2169	0.3558
14.	CAAT-17	1.3714	0.2303	0.3661
15.	CAAT-18	1.4523	0.2861	0.4488
16.	CAAT-19	1.5086	0.2910	0.4365
17.	CAAT-20	1.3606	0.2373	0.3777
18.	CAAT-22	1.4243	0.2721	0.4314
19.	CAAT-25	1.6030	0.3484	0.5218
20.	CAAT-28	1.3709	0.2360	0.3786
21.	CAAT-35	1.3725	0.2353	0.3772
22.	CAAT-36	1.3778	0.2443	0.3929
23.	Average	1.44	0.27	0.41

ne, effective no. of alleles; h, Nei's gene diversity; I, Shannon index.

Table 5
Different diversity indices analysis of SCoT markers.

Sl. No	Primer name	ne	h	I
1	SCoT-1	1.5729	0.3426	0.5085
2	SCoT-2	1.5194	0.2986	0.4390
3	SCoT-3	1.5262	0.3217	0.4903
4	SCoT-6	1.4175	0.2446	0.3690
5	SCoT-12	1.6411	0.3645	0.5393
6	SCoT-13	1.5200	0.3221	0.4950
7	SCoT-21	1.4289	0.2629	0.4070
	Average	1.52	0.31	0.46

ne, effective no. of alleles; h, Nei's gene diversity; I, Shannon index.

population. On the other hand, CISH J-37, which was differentiated from *S. frutescens*, appeared in the second graph in green. It showed more important levels of genetic admixture. This graph shows that even though CISH J-37 was an improved variety, it still has some genetic relationship with wild populations. This is seen in the small amount of

mixing between the groups in the plot. The population structure analysis shows the population structure of different seedling-origin genotypes and the wild/improved genotypes revealing genetic divergence in the original lot. Major two populations are genetically well differentiated, with minimum gene flow between them.

4. Discussion

The *Syzygium* genus encompasses a vast number of species, including several cultivable ones such as Indian blackberry (*Syzygium cumini*), Java apple (*Syzygium aromaticum*) and Malay apple (*Syzygium malaccense*). Among these, jamun (*Syzygium cumini*) is native to India, and thus germplasm collection of this crop from India may represent a valuable asset for its genetic improvement and breeding programs. Therefore, a detailed understanding of the extent and patterns of diversity in the jamun gene pool of India is crucial for their effective conservation and utilization in genetic improvement programs. In the past two decades, gene-based molecular markers have proved very useful for characterizing and understanding genetic relationships among germplasm collections of crops.^{40,55} To our knowledge, this study for the first time reports characterization of diversity among jamun genotypes employing two different gene-based marker techniques, CDBP and SCoT.

The present study analysed genetic diversity among the 32 jamun genotypes using 22 CDBP and 7 SCoT primers. The percentage polymorphism (94.85 %) for CDBP markers observed in this study was higher as compared to an earlier study using the same marker.⁵⁵ The SCoT markers also exhibited a similar level of polymorphism (92.75 %) which was comparable to the results of the previous report.²⁶ Both gene-targeted markers showed higher polymorphism percentages than previous studies on conventional markers, RAPD, reported polymorphism ranging from 41.2 % to 87.50 %.^{46–47,52–53} ISSR demonstrated polymorphism 33.3 % to 64.2 %.^{49,51–53}

PIC and MI values are considered very useful marker statistics parameters for characterizing genetic diversity in crops. The averages of the PIC and MI parameters for the SCoT marker were significantly higher than CDBP markers, which was similar to the findings⁶³ However, the PIC value of both these markers was higher than ISSR (0.248), while SCoT markers represented a higher PIC value than RAPD (0.285) markers as reported in the previous study.⁵³ Rp values for the CDBP marker ranged from 3.75 to 11.43, while for the SCoT, they varied from 3.82 to 7.21, showing that these primers were effective in distinguishing the genotypes. The Rp values for markers recorded in this study are greater than earlier study.⁵⁵ Higher Rp and MI values revealed that these

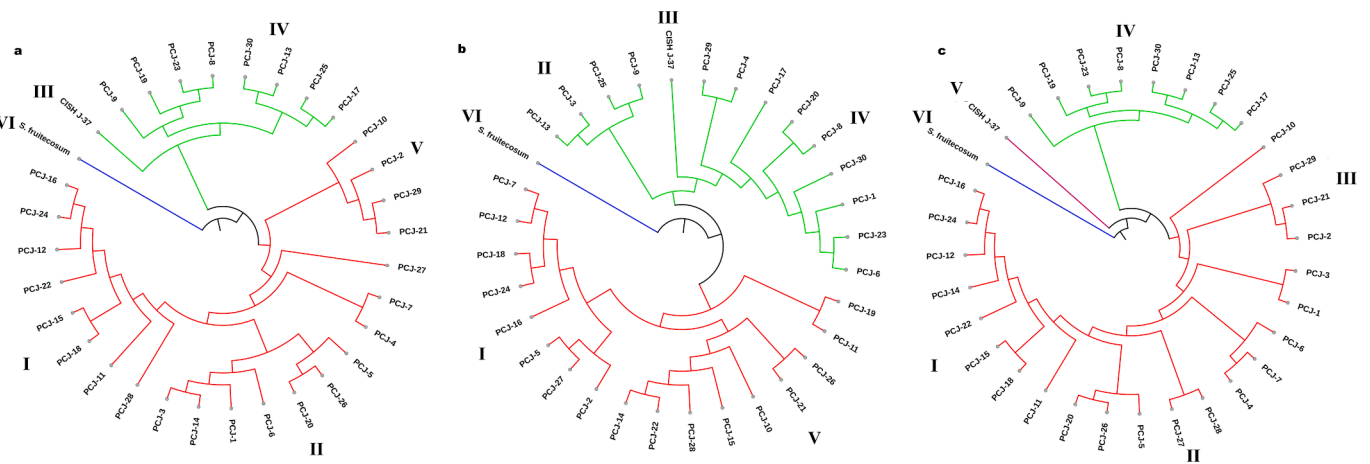


Fig. 3. Dendrograms constructed of 32 jamun genotypes based on CDBP (a), SCoT (b) and combined CDBP + SCoT (c).

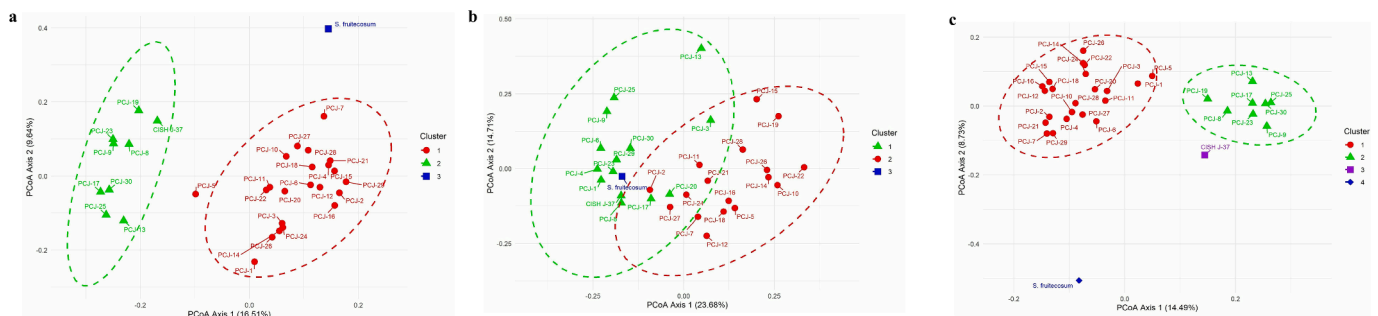


Fig. 4. Principal coordinates analysis (PCoA) based on CDBP (a), SCoT (b) and combined CDBP + SCoT (c).

Table 6

AMOVA results based on CDBP, SCoT and combined (CDBP + SCoT) markers.

Source of variation	df	SS	MS	Est. Var	% Var
CDBP					
Among Pops	1	68.406	68.406	7.571	16 %
Within Pops	30	1200.500	40.017	40.017	84 %
Total	31	1268.906		47.587	100 %
SCoT					
Among Pops	1	15.758	15.758	1.208	10 %
Within Pops	30	336.867	11.229	11.229	90 %
Total	31	352.625		12.437	100 %
CDBP + SCoT					
Among Pops	1	84.117	84.117	8.768	15 %
Within Pops	30	1537.133	51.238	51.238	85 %
Total	31	1621.250		60.005	100 %

Df, degrees of freedom, SS sum of square, MS mean sum of square, % Var percentage variance.

primers possess greater potency and better resolution.⁶⁵ Consistent with our findings, earlier research has also documented the distinctive performance of SCoT and CDBP markers in analysing genetic diversity and

molecular fingerprinting in different plant species.^{35,63–65} Genotype-specific amplicons demonstrate the polymorphism capability of CDBP and SCoT markers and could be useful for genotype identification and genetic diversity studies. Moreover, *S. frutescens* was included as an outgroup in our study and showed unique amplicons by multiple markers.

The genetic diversity index (*h*) quantifies the genetic variation within a population, thus indicating overall genetic diversity and differentiation across genotypes. In comparison, Shannon’s index (*I*) measures both the richness and evenness of genetic features between and within populations.⁶⁶ Higher values of the genetic diversity index (*h*) and Shannon’s index (*I*) indicate a greater level of genetic diversity within and among populations. The *h* and *I* indices obtained in the current study revealed high diversity among loci and findings aligned with many studies.^{44,55} Investigations have demonstrated that markers such as SCoT and CDBP provide greater *h* and *I* value, highlighting their efficacy in capturing genetic heterogeneity among various species, including populations of mango and wheat.^{40,67}

The genetic relationship among jamun genotypes was demonstrated by clustering and PCoA analyses. Cluster analysis is one of the most

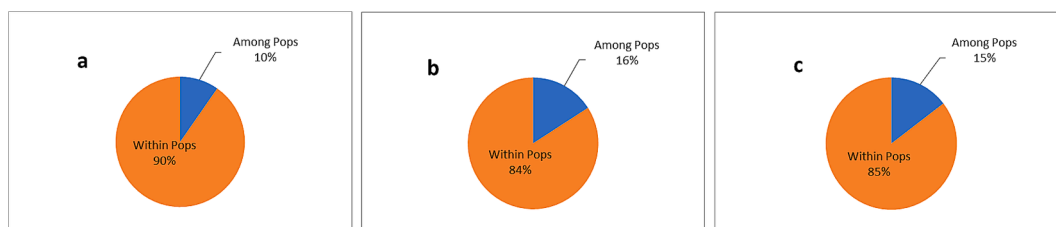


Fig. 5. Molecular Variance Analysis using CDBP, SCoT and combined CDBP + SCoT.

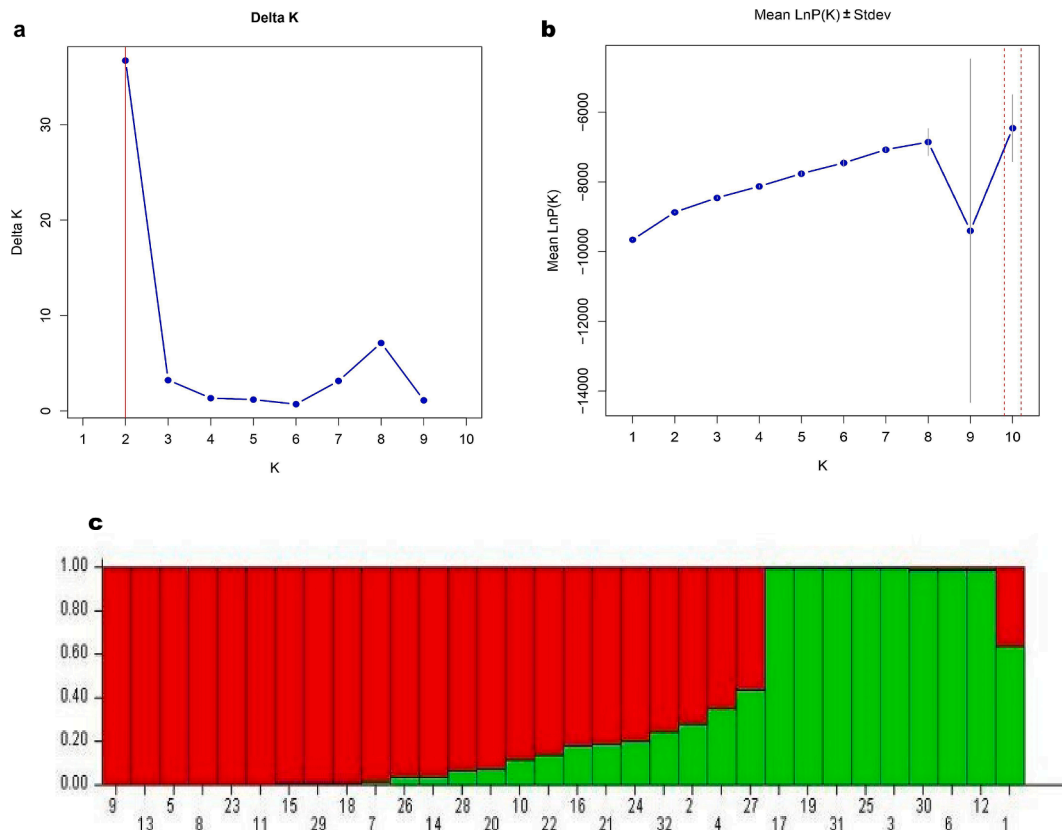


Fig. 6. Population structure analysis based on combined CBDP + SCoT markers. Delta K analysis (a), Mean LnP (K) plot (b), the admixture plot (c).

effective methods for analyzing phylogenetic relationships and classifying genetic groups.⁶⁶ The dendrograms for all 32 genotypes were created using the UPGMA approach. The cluster analysis of CBDP and SCoT markers identified six sub-clusters in which the majority of the genotypes originating from seedling origin, such as PCJ-16, PCJ-24 and PCJ-12, exhibiting moderate diversity, were grouped into Clades I and II. The results obtained in our study were consistent with earlier reports.^{19,44} *S. frutescens* was classified as a genetic outlier by both the markers, while CISH J-37 established a separate sub-cluster showing divergence due to genetic improvement. Similar observations were noted in *Andrographis paniculata*.³⁴ A dendrogram produced with cumulative CBDP and SCoT data yielded many subclusters, similar to results demonstrated in *Prosopis cineraria*.³⁵

PCoA effectively illustrated distinct clustering patterns demonstrating the genetic differentiation among species and populations based on the polymorphic fragments generated by CBDP and SCoT markers. PCoA-based clustering of jamun genotypes largely concurred with the findings of UPGMA analysis. The total genetic variation explained by CBDP, SCoT and combined markers through PCoA analysis was 26.65 %, 38.39 % and 23.22 % and in line with the findings, where they revealed 33 % variation in the top three coordinates for PCoA analysis of SCoT-based markers.³⁴ Also, 20.90 % total variation was observed when both markers were combined.³⁵ The PCoA plot was constructed using the CBDP data and grouped all studied jamun genotypes in three clusters. Cluster 1 contained closely related seedling-origin genotypes. SCoT markers also formed three clusters with little intermixing among genotypes. CISH J-37 and *S. frutescens* formed different clusters for both the markers as well as for combined data, showing higher genetic variations. Our study shows significant genetic variations among wild, improved variety and seedling-origin accessions and the results on similar lines were also observed.⁴⁴ Several studies revealed a higher percentage of variation within the population than among the population, similar to our results, indicating a significant genetic difference across all

accessions within the population.^{68–70,39} demonstrated that CBDP and SCoT data individually showed larger genetic diversity within the population as compared to between populations.³⁵

Utilizing a combined dataset of SCoT and CBDP markers, the population structure of the 32 jamun genotypes was analysed. Population structure-based grouping of jamun genotypes was broadly on the lines obtained using UPGMA and PCoA methods. Many earlier investigations have also demonstrated the utility of gene-based markers in efficiently delineating genetic diversity in crops according to their genomic backgrounds.^{71,72}

5. Conclusion

This study represents a novel attempt to combine CBDP and SCoT markers, enabling a comprehensive genetic diversity assessment across Indian blackberry genotypes. The study evaluated the effectiveness of these gene-based markers in revealing polymorphism, genetic relationships and population structure among seedling-origin, improved and wild genotypes. The results highlighted the greater efficiency of CBDP markers in identifying polymorphisms and discriminating genotypes than SCoT markers. Genotype-specific unique amplicons identified in this research may be useful for genotype identification and diversity studies. Moreover, the combined application of CBDP and SCoT markers proved to be a reliable approach for characterizing genetic diversity in jamun and providing valuable insights into its genomic resources. The genetic diversity revealed in this study can be helpful for the conservation and management of genetic resources of Indian blackberry. Furthermore, it is expected to motivate researchers to explore the genetic diversity of this kind of medicinally important and under-utilized fruit crops using reliable molecular markers.

CRedit authorship contribution statement

Ajay Kumar: Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Data curation. **Kanhaiya Singh:** Writing – review & editing, Supervision, Project administration, Conceptualization. **Amit Kumar Singh:** Writing – review & editing, Resources, Project administration, Conceptualization. **Jai Prakash:** Writing – review & editing, Resources, Conceptualization. **Amit Kumar Goswami:** Writing – review & editing, Resources, Project administration, Conceptualization. **Gyan Prakash Mishra:** Writing – review & editing, Resources, Methodology. **Vishaw Bandhu Patel:** Writing – review & editing, Resources, Investigation, Conceptualization. **Suman Lata:** Writing – review & editing, Resources. **Anshuman Singh:** Writing – review & editing, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors acknowledge the ICAR-Indian Agricultural Research Institute (IARI) for financial assistance in the form of a fellowship crucial for the successful completion of the study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jgeb.2025.100468>.

References

- Hameed F, Gupta N, Rahman R, Anjum N, Nayik GA, Gull A, eds. *Antioxidants in Fruits: Properties and Health Benefits*. Singapore: Springer; 2020: 615–637. https://doi.org/10.1007/978-981-15-7285-2_32.
- Qamar M, Akhtar S, Ismail T, et al. Phytochemical profile, biological properties, and food applications of the medicinal plant *Syzygium cumini*. *Foods*. 2022;11. <https://doi.org/10.3390/foods11030378>.
- Ayyanar M, Subash-Babu P. *Syzygium cumini* (L.) Skeels: a review of its phytochemical constituents and traditional uses. *Asian Pac J Trop Biomed*. 2012;2: 240–246. [https://doi.org/10.1016/S2221-1691\(12\)60050-1](https://doi.org/10.1016/S2221-1691(12)60050-1).
- Singh S, Singh AK, Saroj PL, Mishra S. Research status for technological development of jamun (*Syzygium cumini*) in India: a review. *Indian J Agric Sci*. 2019;89:1991–1998.
- Hiwale S. Jamun (*Syzygium cumini*). In: Hiwale S, eds. *Sustainable horticulture in semiarid dry lands*, Springer India; 2015: 237–46. https://doi.org/10.1007/978-81-322-2244-6_16.
- Nuengchamnong N, Inkaninan K. On-line characterization of phenolic antioxidants in fruit wines from family myrtaceae by liquid chromatography combined with electrospray ionization tandem mass spectrometry and radical scavenging detection. *LWT*. 2009;42:297–302. <https://doi.org/10.1016/j.lwt.2008.04.012>.
- Venugopal KS, Cherita C, Anu-Appaiah KA. Augmentation of chemical and organoleptic properties in *Syzygium cumini* wine by incorporation of grape seeds during vinification. *Food Chem*. 2018;242:98–105. <https://doi.org/10.1016/j.foodchem.2017.09.029>.
- Chaudhary B, Mukhopadhyay K. *Syzygium cumini* (L.) Skeels: a potential source of nutraceuticals. *Int J Pharm Biol*. 2012;2:46–53.
- Joshi M, Paudel M, Munmun Joshi C, Upreti S. Therapeutic influence of jamun (*Syzygium cumini*): a review. *J Pharmacogn Phytochem*. 2019;8:1056–1059.
- Chhikara N, Kaur R, Jaglan S, Sharma P, Gat Y, Panghal A. Bioactive compounds and pharmacological and food applications of *Syzygium cumini*-a review. *Food Funct*. 2018;9:6096–6115. <https://doi.org/10.1039/c8fo00654g>.
- Ranghoo-Sanmukhiya VM, Chellan Y, Soulange JG, et al. Biochemical and phylogenetic analysis of *Eugenia* and *Syzygium* species from Mauritius. *J Appl Res Med Aromat Plants*. 2019;12:21–29. <https://doi.org/10.1016/j.jarmap.2018.10.004>.
- Poehlman JM. *Breeding field crops*. Springer Netherlands; 1987. <https://doi.org/10.1007/978-94-015-7271-2>.
- Bernard A, Barreche T, Lheureux F, Dirlwanger E. Analysis of genetic diversity and structure in a worldwide walnut (*Juglans regia* L.) germplasm using SSR markers. *PLoS One*. 2018;13. <https://doi.org/10.1371/journal.pone.0208021>.
- Seelam RT, Puram VRR, Nallamothu V, Injeti SK, Madhunanpantula VP. Unlocking the potential of rice for thermotolerance using simple sequence repeat (SSR) marker-based assessment of genetic variability and population structure. *Genet Resour Crop Evol*. 2024. <https://doi.org/10.1007/s10722-024-02022-2>.
- Poczai P, Varga I, Laos M, et al. Advances in plant gene-targeted and functional markers: a review. *Plant Methods*. 2013;9:1–32. <https://doi.org/10.1186/1746-4811-9-6>.
- Heidari P, Etmnan A, Azizezhad R, Khosroshahli M. Genomic variation studies in durum wheat (*Triticum turgidum* ssp. durum) using CBDP, SCoT and ISSR markers. *Indian J Genet*. 2017;77:379–386. <https://doi.org/10.5958/0975-6906.2017.00051.7>.
- Piertney SB, Webster LMI. Characterising functionally important and ecologically meaningful genetic diversity using a candidate gene approach. *Genetica*. 2010;138: 419–432. <https://doi.org/10.1007/s10709-008-9322-2>.
- Hamidi H, Talebi R, Keshavarzi F. Comparative efficiency of functional gene-based markers, start codon targeted polymorphism (SCoT) and conserved dna-derived polymorphism (CDDP) with ISSR markers for diagnostic fingerprinting in wheat (*Triticum aestivum* L.). *Cereal Res Commun*. 2014;42:558–567. <https://doi.org/10.1556/CRC.2014.0010>.
- Singh AK, Rana MK, Singh S, Kumar S, Kumar R, Singh R. CAAT box- derived polymorphism (CBDP): A novel promoter -targeted molecular marker for plants. *J Plant Biochem Biotechnol*. 2014;23:175–183. <https://doi.org/10.1007/s13562-013-0199-5>.
- Collard BCY, Mackill DJ. Start codon targeted (scot) polymorphism: a simple, novel DNA marker technique for generating gene-targeted markers in plants. *Plant Mol Biol Rep*. 2009;27:86–93. <https://doi.org/10.1007/s11105-008-0060-5>.
- Singh SK, Chhajer S, Pathak R, Bhatt RK, Kalia RK. Genetic diversity of Indian jujube cultivars using SCoT, ISSR, and rDNA markers. *Tree Genet Genomes*. 2017; 13:1–14. <https://doi.org/10.1007/s11295-016-1092-x>.
- Rai MK. Start codon targeted (SCoT) polymorphism marker in plant genome analysis: current status and prospects. *Planta*. 2023;257. <https://doi.org/10.1007/s00425-023-04067-6>.
- Etmnan A, Pour-Aboughadareh A, Mohammadi R, Noori A, Ahmadi-Rad A. Applicability of CAAT box-derived polymorphism (CBDP) markers for analysis of genetic diversity in durum wheat. *Cereal Res Commun*. 2018;46:1–9. <https://doi.org/10.1556/0806.45.2017.054>.
- Najafi S, Shoostari L, Mohammadi R, Etmnan A, Mehrabi AM, Ghorbanpour M. Genetic diversity and population structure analysis in durum wheat (*Triticum turgidum* L. var. Durum) germplasm using agronomic characters and CBDP markers. *Genet Resour Crop Evol*. 2024;71:3405–3417. <https://doi.org/10.1007/s10722-023-01854-8>.
- Fabriki-Ourang S, Karimi H. Assessment of genetic diversity and relationships among *Salvia* species using gene targeted CAAT box-derived polymorphism markers. *J Genet*. 2019;98:75. <https://doi.org/10.1007/s12041-019-1121-2>.
- Guo DL, Zhang JY, Liu CH. Genetic diversity in some grape varieties revealed by SCoT analyses. *Mol Biol Rep*. 2012;39:5307–5313. <https://doi.org/10.1007/s11033-011-1329-6>.
- Luo C, He XH, Chen H, Hu Y, Ou SJ. Genetic relationship and diversity of *Mangifera indica* L.: revealed through SCoT analysis. *Genet Resour Crop Evol*. 2012; 59:1505–1515. <https://doi.org/10.1007/s10722-011-9779-1>.
- Tabasi M, Sheidai M, Hassani D, Koohdar F. DNA fingerprinting and genetic diversity analysis with SCoT markers of Persian walnut populations (*Juglans regia* L.) in Iran. *Genet Resour Crop Evol*. 2020;67:1437–1447. <https://doi.org/10.1007/s10722-020-00914-7>.
- Baghizadeh A, Dehghan E. Efficacy of SCoT and ISSR markers in assessment of genetic diversity in some Iranian pistachio (*Pistacia vera* L.) cultivars, 37–43 *PHJ*. 2018;1. <https://doi.org/10.22123/PHJ.2017.54299>.
- Nair AGH, Vidya P, Mohan C. Analysis of genetic variability in sweet potato accessions using Start codon targeted (SCoT) polymorphism. *Int J Biotechnol Biochem*. 2016;12:111–121.
- Kumar J, Agrawal V. Assessment of genetic diversity, population structure and sex identification in dioecious crop, *Trichosanthes dioica* employing ISSR. *SCoT and SRAP Markers. Heliyon*. 2019;5:1346. <https://doi.org/10.1016/j.heliyon.2019.10116>.
- Vanijajiva O. Start codon targeted (SCoT) polymorphism reveals genetic diversity of manilkara in Thailand. *Biodiversitas*. 2020;21:666–673. <https://doi.org/10.13057/biodiv/d210232>.
- Heikrujam M, Kumar J, Agrawal V. Genetic diversity analysis among male and female *Jujube* genotypes employing gene targeted molecular markers, start codon targeted (SCoT) polymorphism and CAAT box-derived polymorphism (CBDP) markers. *Meta Gene*. 2015;5:90–97. <https://doi.org/10.1016/j.mgene.2015.06.001>.
- Tiwari G, Singh R, Singh N, et al. Study of arbitrarily amplified (RAPD and ISSR) and gene targeted (SCoT and CBDP) markers for genetic diversity and population structure in kalmegh [*Andrographis paniculata* (Burm. f.) Nees]. *Ind Crops Prod*. 2016;86:1–11. <https://doi.org/10.1016/j.indcrop.2016.03.031>.
- Shekhawat JK, Rai MK, Shekhawat NS, Kataria V. Exploring genetic variability in *Prosopis cineraria* using two gene targeted CAAT box-derived polymorphism (CBDP) and start codon targeted (SCoT) polymorphism markers. *Mol Biol Rep*. 2018;45:2359–2367. <https://doi.org/10.1007/s11033-018-4400-8>.
- Sharma U, Rai MK, Shekhawat NS, Kataria V. Genetic homogeneity revealed in micropropagated *Bauhinia racemosa* Lam. using gene targeted markers CBDP and SCoT. *PMPP*. 2019;25:581–588. <https://doi.org/10.1007/s12298-018-00639-z>.
- Apana N, Amom T, Tikendra L, Potshangbam AM, Dey A, Nongdam P. Genetic diversity and population structure of *Clerodendrum serratum* (L.) Moon using CBDP, iPBS and SCoT markers, 25:100349 *J Appl Res Med Aromat Plants*. 2021 <https://doi.org/10.1016/j.jarmap.2021.100349>.

- [38]. Khodae L, Azizinezhad R, Etminan AR, Khoshroshahi M. Assessment of genetic diversity among Iranian Aegilops triuncialis accessions using ISSR, SCoT, and CBDP markers. *JGEB*. 2021;19:5. <https://doi.org/10.1186/s43141-020-00107-w>.
- [39]. Ghobadi G, Etminan A, Mehrabi AM, Shoostari L. Molecular diversity analysis in hexaploid wheat (*Triticum aestivum* L.) and two Aegilops species (*Aegilops crassa* and *Aegilops cylindrica*) using CBDP and SCoT markers. 19:56 *JGEB*. 2021 <https://doi.org/10.1186/s43141-021-00157-8>.
- [40]. Jena RC, Chand PK. Multiple DNA marker-assisted diversity analysis of Indian mango (*Mangifera indica* L.) populations, 11:10345 *Sci Rep*. 2021 <https://doi.org/10.1038/s41598-021-89470-3>.
- [41]. Ahmed DA, Tahir NA, razzak, Salih SH, Talebi R. Genome diversity and population structure analysis of Iranian landrace and improved barley (*Hordeum vulgare* L.) genotypes using arbitrary functional gene-based molecular markers. *Genet Resour Crop Evol*. 2021;68:1045–1060. <https://doi.org/10.1007/s10722-020-01047-7>.
- [42]. Pandey Y, Chaturvedi T, Swaroop H, Gupta AK, Shanker K, Tiwari G. Phytochemical and genetic marker (SCoT and CBDP) based study of genetic diversity and population structure in natural populations of *Cannabis sativa* L.: A high-value sustainable biodiversity of North-Indian Himalaya, 200:116892 *Ind Crops Prod*. 2023 <https://doi.org/10.1016/j.indcrop.2023.116892>.
- [43]. Atia MAM, El-Moneim DA, Abdelmoneim TK, et al. Evaluation of genetic variability and relatedness among eight *Centaurea* species through CAAT-box derived polymorphism (CBDP) and start codon targeted polymorphism (SCoT) markers. *Biotechnol Biotechnol Equip*. 2021;35:1230–1237. <https://doi.org/10.1080/13102818.2021.1960891>.
- [44]. Puneeth PV, Lata S, Yadav RK, et al. Exploring the genetic diversity using CAAT box-derived polymorphism (CBDP) and start codon targeted (SCoT) markers in cultivated and wild species of okra (*Abelmoschus esculentus* (L.) Moench). *Genet Resour Crop Evol*. 2023;70:749–761. <https://doi.org/10.1007/s10722-022-01458-8>.
- [45]. Chouksey S, Ashfaq MA, Kaira P, et al. Development of highly discriminatory SCoT- and CBDP-based SCAR fingerprint for authentication of Indian senna (*Senna alexandrina* Mill.) formerly *Cassia angustifolia* Vahl., 15:1424665 *Front Plant Sci*. 2024 <https://doi.org/10.3389/fpls.2024.1424665>.
- [46]. Khan S, Vaishali, Sharma V. Genetic differentiation and diversity analysis of medicinal tree *Syzygium cumini* (Myrtaceae) from ecologically different regions of India. *PMBP*. 2010;16:149–158.
- [47]. Shakya R, Siddiqui SA, Srivatawa N, Bajpai A. Molecular characterization of Jamun (*Syzygium cumini* L. Skeels) genetic resources. *Int J Fruit Sci*. 2010;10:29–39. <https://doi.org/10.1080/15538361003676769>.
- [48]. Singh JP, Singh A, Bajpai A, Zareen AL. Hypothesis comparative analysis of DNA polymorphisms and phylogenetic relationships among *Syzygium cumini* Skeels based on phenotypic characters and RAPD technique. *Bioinformatics*. 2014;10:201–208.
- [49]. Gajera HP, Gevariya SN, Patel SV, Golakiya BA. Nutritional profile and molecular fingerprints of indigenous black jamun (*Syzygium cumini* L.) landraces. *J Food Sci Technol*. 2018;55:730–739. <https://doi.org/10.1007/s13197-017-2984-y>.
- [50]. Khan S, Agarwal S, Singh K, Chaturgoon A, Pareek A. Molecular fingerprinting and phytochemical investigation of *Syzygium cumini* L. From different agro-ecological zones of India, 12:931 *Plants*. 2023 <https://doi.org/10.3390/plants12040931>.
- [51]. Ahmad I, Bhagat S, Simachalam P, Srivastava RC. Molecular characterization of *Syzygium cumini* (wild jamun) from A&N Islands. *Indian J Horti*. 2012;69:306–311.
- [52]. Cheong MLS, Ranghoo-Sanmukhiya VM. Phylogeny of *Syzygium* species using morphological, RAPD and ISSR markers. *Int J Agric Biol*. 2013;15:511–516.
- [53]. Jaisankar I, Subramani T, Singh AK, et al. Genetic diversity among jamun (*Syzygium cumini* L. Skeels) accessions of andaman and nicobar islands revealed through morpho-biochemical and DNA markers. *J And Sci Ass*. 2016;21:35–44.
- [54]. Sharma VK, Sharma D, Sharma R, Sharma SD, Dhiman K, Murkute AA. Morpho-molecular exploration and selection of elite genotypes from indigenous *Syzygium cumini* L. Skeels (jamun) diversity of North-Western Indian Himalayas. *Plant Genet. Reso*. 2023;21:312–321. <https://doi.org/10.1017/S147926212300076X>.
- [55]. Singh A, Verma SK, Prasad G, Kumar A, Sharma PC, Singh AK. Elucidating genetic diversity and population structure in jamun [*Syzygium cumini* L. (Skeels)] using morpho-physiological traits and CAAT box-derived polymorphism. *S Afr J Bot*. 2022;151:454–465. <https://doi.org/10.1016/j.sajb.2022.10.023>.
- [56]. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull*. 1987;19:11–15.
- [57]. Roldan-Ruiz I, Dendauw J, Van Bockstaele E, Depicker A, De Loose M. AFLP markers reveal high polymorphic rates in ryegrasses (*Lolium* spp.). *Mol Breed*. 2000;6:125–134.
- [58]. Kumar A, Mishra P, Singh SC, Sundaresan V. Efficiency of ISSR and RAPD markers in genetic divergence analysis and conservation management of *Justicia adhatoda* L., a medicinal plant. *Pl Syst Evol*. 2014;300:1409–1420. <https://doi.org/10.1007/s00606-013-0970-z>.
- [59]. Prevost A, Wilkinson MJ. Communicated by A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. *Theor Appl Genet*. 1999; 98:107–112.
- [60]. Peakall R, Smouse PE. GenALEX 6.5: genetic analysis in excel. population genetic software for teaching and research—an update. *Bioinformatics*. 2012;28:2537–2539. <https://doi.org/10.1093/bioinformatics/bts460>.
- [61]. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genet*. 2000;155:945–959.
- [62]. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol*. 2005;14:2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>.
- [63]. Bokaei AS, Sofalian O, Sorkhילהlloo B, Asghari A, Pour-Aboughadareh A. Deciphering the level of genetic diversity in some Aegilops species using CAAT box-derived polymorphism (CBDP) and start codon target polymorphism (SCoT) markers. *Mol Biol Rep*. 2023;50:5791–5806. <https://doi.org/10.1007/s11033-023-08488-0>.
- [64]. Yekken MZ, Emirialioğlu O, Çiftçi V, Bayraktar H, Palacioğlu G, Özer G. Analysis of genetic diversity among common bean germplasm by start codon targeted (SCoT) markers. *Mol Biol Rep*. 2022;49:3839–3847. <https://doi.org/10.1007/s11033-022-07229-z>.
- [65]. Etminan A, Pour-Aboughadareh A, Mohammadi R, et al. Applicability of start codon targeted (SCoT) and inter-simple sequence repeat (ISSR) markers for genetic diversity analysis in durum wheat genotypes. *Biotechnol Biotechnol Equip*. 2016;30:1075–1081. <https://doi.org/10.1080/13102818.2016.1228478>.
- [66]. Que Y, Pan Y, Lu Y, et al. Genetic analysis of diversity within a chinese local sugarcane germplasm based on start codon targeted polymorphism. *Biomed Res Int*. 2014;2014, 468375. <https://doi.org/10.1155/2014/468375>.
- [67]. Abouseada HH, Mohamed ASH, Teleb SS, et al. Genetic diversity analysis in wheat cultivars using SCoT and ISSR markers, chloroplast DNA barcoding and grain SEM. *BMC Plant Biol*. 2023;23:193. <https://doi.org/10.1186/s12870-023-04196-w>.
- [68]. Randi E, Lucchini V. Detecting rare introgression of domestic dog genes into wild wolf (*Canis lupus*) populations by Bayesian admixture analyses of microsatellite variation. *Conserve Genet*. 2002;3:29–43. <https://doi.org/10.1023/A:1014229610646>.
- [69]. Moradkhani H, Mehrabi AA, Etminan A, Pour-Aboughadareh A. Molecular diversity and phylogeny of *Triticum-Aegilops* species possessing D genome revealed by SSR and ISSR markers. *Plant Breed Seed Sci*. 2016;71:81–95. <https://doi.org/10.1515/plass-2015-0024>.
- [70]. Lesniowska-Nowak J, Okoń S, Wieremczuk A. Molecular diversity analysis of genotypes from four Aegilops species based on retrotransposon-microsatellite amplified polymorphism (REMAPP) markers. *Cereal Res Commun*. 2021;49:37–44. <https://doi.org/10.1007/s42976-020-00086-1>.
- [71]. Pour-Aboughadareh A, Ahmadi J, Mehrabi AA, Etminan A, Moghaddam M. Insight into the genetic variability analysis and relationships among some Aegilops and *Triticum* species, as genome progenitors of bread wheat, using SCoT markers. *Plant Biosyst*. 2018;152:694–703. <https://doi.org/10.1080/11263504.2017.1320311>.
- [72]. Pour-Aboughadareh A, Pocza P, Etminan A, Jadidi O, Kianersi F, Shoostari L. An analysis of genetic variability and population structure in wheat germplasm using microsatellite and gene-based markers. *Plants*. 2022;11, 11091205. <https://doi.org/10.3390/plants11091205>.