



Comparative assessment of ISSR, RAPD, and SCoT markers for genetic diversity in *Clerodendrum* species of North East India

Barbi Gogoi^{1,2} · S. B. Wann³ · S. P. Saikia¹

Received: 25 May 2020 / Accepted: 28 August 2020 / Published online: 3 September 2020
© Springer Nature B.V. 2020

Abstract

Clerodendrum belonging to the family of Lamiaceae is used in indigenous systems of medicine to treat various life-threatening diseases. The genus has complex morphological variations which lead to limits in its precise taxonomic classifications. Genetic diversity study could enhance taxonomic authentication and evolutionary relationship among the species of *Clerodendrum*. In this study, nine species of *Clerodendrum* collected from different regions of North East India were screened using ISSR, RAPD, and SCoT molecular markers. The markers of ISSR, RAPD, and SCoT generated a total of 79, 126, and 145 amplicons with an average of 6.58, 7.86, and 8.53 amplicon per primer. The polymorphism information contents (PIC) for ISSR, RAPD, and SCoT ranged from 0.28 to 0.37, 0.39 to 0.69, and 0.30 to 0.62 with resolving power (Rp) varying from 5.26 to 11.11, 4.04 to 9.67, and 4.54 to 8.65, respectively. Unweighted Pair Group Method with Arithmetic Mean (UPGMA) based clustering methods grouped 94 genotypes into 6 clusters for ISSR and 3 clusters each for RAPD and SCoT markers. Similarly, population structure-based analysis divided 94 genotypes into 6 populations for ISSR and RAPD and 4 populations for SCoT markers. AMOVA analysis revealed that SCoT markers generated maximum genetic variations within and among genotypes, contrary to ISSR and RAPD markers. Results in this study, suggest that the competence of three markers was relatively the same in genotypes fingerprinting, but SCoT was more efficient in the detection of polymorphism for *Clerodendrum* species. Further, these results could be integrated in the exploration of diverse *Clerodendrum* species and germplasm utilization.

Keywords *Clerodendrum* · ISSR · RAPD · SCoT · Structure-based

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11033-020-05792-x>) contains supplementary material, which is available to authorized users.

✉ S. P. Saikia
spsaikia@gmail.com

S. B. Wann
wannsb@neist.res.in

¹ Medicinal Aromatic and Economic Plants Group, Biological Sciences & Technology Division (BSTD), CSIR-North East Institute of Science & Technology, Jorhat 785006, Assam, India

² Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India

³ Biotechnology Group, Biological Sciences & Technology Division (BSTD), CSIR-North East Institute of Science & Technology, Jorhat 785006, Assam, India

Introduction

Clerodendrum L. (Lamiaceae) is a large diversified genus of small trees or shrubs, predominantly distributed in tropical and subtropical regions of Asia, Africa, America, Egypt, and Madagascar. Globally about 580 species were recorded, and 23 species were reported in India, of which 18 species were documented in North East India [1]. The genus had entrenched pharmacological properties and practice in folklore medicines to treat disease disorders like inflammations, dyspepsia, bronchitis, tumours, dropsy, asthma, hypertension, fever, leprosy, and cancer [2]. The genus had a higher degree of morphological and cytological variation ($2n = 24$ to $2n = 184$) within the species and depicts paraphyletic or polyphyletic origin of the genus [3]. Phylogenetic analysis of *Clerodendrum* sp. based on chloroplast DNA (cpDNA) and ITS sequences revealed polyphyletic nature of the genus [4]. Based on morphological variations, taxonomic classification of the

genus remains incoherent as many researchers classified the genus into different subgenera and nomenclature by more than one author which results in species dualism [5]. Numerous species of this genus are used in the traditional systems of medicine by various tribes of North East India to treat asthma, vermifuge, skin diseases, tumors, cancer, diabetes, obesity and cancer using plant parts such as roots and leaves extract of different *Clerodendrum* species [6]. The leaves of *C. colebrookianum* are commonly used by Khasi, Naga, Mishing and Assamese tribes of North East India to treat high blood pressure [7]. North East India covers 50% of India's biodiversity and it is crucial to evaluate the genetic structure and genetic diversity of the *Clerodendrum* species, which is widely distributed in this region, to eventually utilize its valuable germplasm for further classifying species into clusters according to their genetic similarities.

Assessment of genetic variability could uplift optimum improvement in the genetic and effective conservation of any species [8]. Generally, genetic diversity was accessed using morphological traits, biochemical and molecular markers [9]. However, the advancement of molecular systemic methods suggests molecular markers as reliable tools compared to morpho-chemical characterizations as these markers are not influenced by environmental stress, natural selections and disease susceptibility [10]. Widely used molecular markers like Amplified Fragment Length Polymorphism (AFLP), Randomly Amplified Polymorphic DNA (RAPD) and Inter Simple Sequence Repeats (ISSR) collectively known as Arbitrarily Amplified Dominant (AAD) are used to study genus diversity based on different banding patterns of the species. Additionally, new alternative DNA markers such as Sequence Related Amplified Polymorphism (SRAP), CAAT Box Derived Polymorphism (CBDP), Targeted Region Amplification Polymorphism (TRAP), Conserved Region Amplification polymorphism (CRAP), and Start Codon Targeted Polymorphism (SCoT) are used in genetic fingerprinting, gene population studies and qualitative trait loci (QTL) mapping. The markers ISSR and RAPD have arbitrarily amplified markers based on the non-coding region of genome whereas SCoT primers are gene-targeted markers derived either from the gene itself or instant flanking regions. SCoT had several advantages over RAPD, ISSR, and AFLP because it is easy to design, stable and amplifies reproducible bands based on the conserved region of the translational initiation codon (ATG) and resembles a functional gene with its correlating characters [11]. To our knowledge, a limited number of molecular marker-based studies on *Clerodendrum* sp with SSR [12] and RAPD [13] were reported till date. So, this study is the first attempt to analyze genetic diversity in *Clerodendrum* sp. of North East India using the comparative efficiency of RAPD, ISSR and SCoT markers.

Methodology

Sample collection and extraction of genomic DNA

A total of 94 genotypes from 9 *Clerodendrum* species were collected from different locations of North East India. The details of the collected *Clerodendrum* genotypes were listed in (Support information Table S1). All the collected leaf samples were stored at $-20\text{ }^{\circ}\text{C}$ for 24 h and lyophilized (Scanvac, CoolSafe™) at $-110\text{ }^{\circ}\text{C}$ for 48 h. This process extracts the water content from the frozen samples at low temperature using sublimation process (ice convert directly to vapor state without passing through intermediate liquid phase) under high vacuum and preserved the samples for long term storage. For extraction of genomic DNA, 4 gm of lyophilized leaves grinded and total genomic content was isolated using the modified CTAB method. Quality of DNA was checked on 0.8% agarose gel and visualized in gel documented system (G: BOX, Syngene, U.K.). The total yields of DNA (ng/ μl) as well as DNA purity at A260/A280 ratio were estimated using nanodrop (Eppendorf, Germany) (Support information Table S1). Further, a working concentration of 15 ng/ μl DNA stock was prepared and stored at $4\text{ }^{\circ}\text{C}$ until further use.

Molecular markers genotyping

Clerodendrum species were screened with 3 dominant molecular markers, ISSR, RAPD, and SCoT. Initially 20 primers for each marker were screened using 9 *Clerodendrum* samples. A total 12 ISSR primers, 16 RAPD primers, and 17 SCoT primers produced clear, reproducible banding patterns and were selected for further analysis. The primers that generated weak, no or complex banding patterns were discarded (Support information Table S2). PCR amplification was performed in a total volume of 20 μl reaction mixture in 96-well plates (Abdos) with 15 ng/ μl of DNA template, 10X PCR buffer, 2.5 mM dNTP, 5 pmol primer, 1U Taq DNA polymerase (Sigma) and final volume was adjusted with nuclease-free water. The PCR reactions were amplified in a Thermal Cycler (Applied Biosystem). Further, the amplified products were analyzed in 1.5% agarose gel at 95 V for 1.5 h and visualized in the gel documentation system (G:BOX, Syngene, U.K.). Low range DNA ladder of 100 bp to 3 kb (Genei) was used as a reference marker for each marker analysis.

Scoring of data

The scoring of ISSR, RAPD and SCoT markers were done manually in a binary format based on the presence (1) or

absence (0) of bands by visual inspection. Only the distinctive and reproducible bands were calculated. To discriminate the markers, parameters like the total number of bands, number of polymorphic and monomorphic bands, percentage of polymorphism, polymorphic information content (PIC), resolving power (RP), and marker index (MI) were evaluated using Microsoft Excel. The value of PIC was calculated as $PIC = 1 - \sum p_i^2$, where p_i is the frequency of i^{th} allele (band present) and summation extends over n alleles [14]. Resolving power (Rp) of each primer was estimated with $Rp = \sum I_b$ where I_b (informative fragments) = $1 - [2 \times (0.5 - p_i^2)]$, where p_i^2 is the proportion of accession containing bands [15].

Distance matrix and cluster analysis of the datasets were generated using NTSYS (Numerical taxonomy and multivariate analysis system) software v2.1 [16]. The genetic relationship among all the *Clerodendrum* species was calculated using SIMQUAL (Similarity for qualitative data program in NTSYS) module of Jaccard's similarity coefficient [17]. UPGMA with Sequential agglomerative hierarchical and nested clustering method program in NTSYS (SAHN) was used to generate a dendrogram representing the genetic association as revealed by similarity coefficient.

The allelic data matrix of 1 or 0 was used to calculate the population genetic analysis using POPGENE v 1.32 [18] which consist of number of observed alleles (N_a), and number of effective alleles (N_e). Nei's genetic diversity (H), polymorphism index (PI) and Shannon's index (I) were computed for each *Clerodendrum* species. Genetic diversity measure such as H_t = total gene diversity and G_{st} = coefficient of gene differentiation was calculated using Nei gene diversity statistics [19].

Bayesian model-based population structure was studied to infer population structure and assign individuals to population based on ISSR, RAPD and SCoT using STRUCTURE software version 2.3.4 [20]. The membership of each accession was run for a range of genetic clusters from the value of $K = 2-10$ with the admixture model and correlated allele frequency. To obtain optimum K value, the length of the burning period was 5000 iterations followed by 50,000 Monte Carlo Markov Chain replicates. For each K it was replicated 3 times. $\ln(PD)$ derived for each K and plotted to find the plateau of the ΔK values [21]. Population structure was calculated with ΔK using "structure harvester" (<https://taylor0.biology.ucla.edu>) based second order rate of change of Likelihood distribution mean L ("K") and with respect to K . The partitioning of genetic variation among and within different genotypes was also investigated with analysis of molecular variance (AMOVA) and Principal Coordinate Analysis (PCoA) using GenAlEx 6.501 software [22]. The cophenetic correlation coefficients were calculated with COPH and MXCOMP procedures for each marker. Mantel tests between two matrices were compared in NTSYS v2.1

software using the MXCOMP option and the matrices were obtained from the cophenetic values of each marker.

Results

To estimate the genetic diversity in *Clerodendrum*, three different PCR-based molecular markers (ISSR, RAPD, and SCoT) were used. The statistical details of each marker were as follows:

ISSR analysis

Out of 20 ISSR tested primers, 12 primers amplified 79 clear bands for *Clerodendrum* sp. depending upon their simple sequence repeat motifs. The number of amplified fragments varied from 5 to 8, with an average of 6.58 bands per primer (Table 1). Among 79 bands, 56 bands were polymorphic and the average polymorphic percentage was 71.40% across all the genotypes. The maximum numbers of polymorphic bands were observed in UBC809 and UBC810. The average PIC was 0.34 ranging from 0.28 to 0.37 and highest in UBC811 and UBC836 while lowest in UBC810. MI and Rp of the ISSR varied from 0.27 to 1.72 and 5.26 to 11.11. The maximum MI and Rp were found in UBC809 and UBC842. Nei's gene diversity ranged from 0.13 ± 0.06 to 0.26 ± 0.10 with a mean of 0.18 ± 0.09 while the effective number of allelic frequencies spanned from 1.09 ± 0.23 to 1.30 ± 0.37 with a mean of 1.19 ± 0.31 . The genetic diversity parameters such as N_a , N_e , H and I were found to be highest in *C. infortunatum* ($N_a-1.53 \pm 0.50$, $N_e-1.30 \pm 0.37$, $H-0.26 \pm 0.10$, and $I-0.28 \pm 0.08$) and lowest in *C. inerme* ($N_a-1.16 \pm 0.39$, $N_e-1.09 \pm 0.23$, $H-0.13 \pm 0.06$, and $I-0.20 \pm 0.09$) respectively (Table 2). The genetic variation in ISSR markers revealed that the mean value of total gene diversity (H_t), gene diversity within population (H_s), coefficient of gene differentiation (G_{st}) and level of gene flow (N_m) were $H_t-0.24 \pm 0.04$, $H_s-0.11 \pm 0.01$, $G_{st}-0.33$, and $N_m-0.38$, respectively (Support information Table S3). The G_{st} value 0.33 indicates about 33% of total genetic divergence among the populations and the remaining 67% within the populations.

RAPD analysis

The PCR amplification of *Clerodendrum* sp. with 16 RAPD primers yielded 126 scoreable bands with 111 polymorphic fragments (Table 1). The RAPD primers produced a greater number of polymorphic and scoreable bands as compared to ISSR primers. The PIC value ranged from 0.39 (OPB-1) to 0.69 (OPB-17), with an average of 0.55 per loci. The percentage of polymorphic fragments ranged from 71.42 to 100%, with an average

Table 1 Details characteristic of ISSR, RAPD, and SCoT markers with total number of bands amplified, number of polymorphic bands, %polymorphism, PIC, MI and Rp value

Markers	Sequence	Total no. of bands	Total no. of polymorphic bands	Percentage of polymorphism	PIC ^a	MI ^b	Rp ^c	
ISSR	UBC-807 (AG) ₈ T	7	5	71.43	0.33	1.19	9.90	
	UBC-808 (AG) ₈ C	7	4	57.14	0.32	0.74	11.09	
	UBC-809 (AG) ₈ G	6	6	100.00	0.29	1.72	8.55	
	UBC-810 (AG) ₈ T	6	6	100.00	0.28	1.69	8.70	
	UBC-811 (GA) ₈ C	6	5	83.33	0.37	1.52	5.26	
	UBC-834 (AG) ₈ YT	5	4	80.00	0.35	1.12	7.37	
	UBC-835 (AG) ₈ YC	7	5	71.43	0.36	1.28	8.28	
	UBC-836 (AG) ₈ YA	7	5	71.43	0.37	1.30	8.43	
	UBC-841 (GA) ₈ YC	7	4	57.14	0.34	0.79	11.03	
	UBC-842 (GA) ₈ YG	8	5	62.50	0.34	1.06	11.11	
	UBC-846 (CA) ₈ RT	5	2	40.00	0.34	0.27	6.74	
	UBC-856 (AC) ₈ YA	8	5	62.50	0.34	1.05	10.86	
	Mean		6.58	4.67	71.41	0.34	1.14	8.94
	Total		79	56	856.90	4.03	13.73	107.32
	Range		5–8	2–6	40.0–100.0	0.28–0.37	0.27–1.72	5.26–11.11
RAPD	OPA-2 TGCCGAGCTG	8	7	87.50	0.43	2.66	9.04	
	OPA-4 AATCGGGCTG	6	5	83.33	0.59	2.46	4.93	
	OPA-5 AGGGGTCTTG	9	8	90.00	0.58	4.89	7.52	
	OPA-7 GAAACGGGTG	9	9	100.00	0.55	4.13	8.14	
	OPA-10 GTGATCGCAG	9	8	88.89	0.52	3.69	8.66	
	OPA-11 CAATCGCCGT	9	8	88.89	0.59	4.19	7.39	
	OPA-18 AGGTGACCGT	10	9	90.00	0.52	5.17	9.67	
	OPB-1 GTTTCGCTCC	7	5	71.42	0.39	1.40	8.44	
	OPB-5 TGCGCCCTTC	5	5	100.00	0.59	2.99	4.04	
	OPB-6 TGCTCTGCCC	8	7	87.50	0.49	2.51	7.98	
	OPB-9 TGGGGGACTC	7	6	85.71	0.49	3.00	7.22	
	OPB-10 CTGCTGGGAC	7	6	85.71	0.54	2.80	6.39	
	OPB-12 CCTTGACGCA	8	7	87.50	0.62	3.77	6.16	
	OPB-15 GGAGGGTGTT	8	7	87.50	0.63	3.88	5.86	
	OPB-17 AGGGAACGAG	9	7	77.78	0.69	3.07	5.46	
	OPB-18 CCACAGCAGT	7	7	100.00	0.59	4.16	5.67	
	Mean		7.86	6.93	88.23	0.55	3.42	7.04
	Total		126	111	1411.73	8.81	54.77	112.57
	Range		5–10	5–9	71.42–100	0.39–0.69	1.40–5.17	4.04–9.67

Table 1 (continued)

Markers	Sequence	Total no. of bands	Total no. of polymorphic bands	Percentage of polymorphism	PIC ^a	MI ^b	Rp ^c	
SCOT	S2	CAACAATGGCTACCACCC	9	8	88.89	0.58	4.15	7.48
	S3	CAACAATGGCTACCACCG	6	5	83.33	0.49	2.06	7.84
	S4	CAACAATGGCTACCACCT	10	9	90.00	0.57	4.60	6.15
	S6	CAACAATGGTCACCACGC	6	5	83.33	0.58	2.40	4.54
	S7	CAACAATGGTCACCACGG	6	6	100.00	0.58	3.49	4.79
	S8	ACCATGGCTACCACCGAG	10	8	80.00	0.59	3.80	6.97
	S12	ACGACATGGCGACCAACG	7	6	85.71	0.59	3.05	5.79
	S13	ACGACATGGCGACCATCG	7	6	85.71	0.60	3.06	5.46
	S16	ACCATGGCTACCACCGAC	9	8	88.89	0.59	4.22	7.49
	S17	ACCATGGCTACCACCGAG	14	13	92.86	0.60	7.35	8.65
	S21	ACGACATGGCGACCCACA	9	8	88.89	0.61	4.30	7.77
	S23	CACCATGGCTACCACCAG	9	8	88.89	0.62	4.38	4.63
	S33	CCATGGCTACCACCGCAG	10	9	90.00	0.61	4.97	8.33
	S34	CATGGCTACCACCGGCC	9	8	88.89	0.61	4.35	7.39
	S39	GCAACAATGGCTACCACC	8	8	100.00	0.61	4.85	7.67
	S41	CAATGGCTACCCTGACA	5	4	80.00	0.30	0.95	7.03
	S43	CAATGGCTACCATTAGCG	11	10	90.91	0.62	5.61	5.65
	Mean		8.53	7.59	88.60	0.57	3.97	6.68
	Total		145	129	1506.3	9.75	67.52	113.63
	Range		5–14	4–13	80–100	0.30–0.62	0.95–7.35	4.54–8.65

PIC polymorphic information content, MI Marker index, Rp resolving power

of 88.23% polymorphism. The MI and RP of RAPD ranged from 1.40 (OPB-1) to 5.17 (OPA-18) and 4.04 (OPB-5) to 9.67 (OPA-18). The level of genetic diversity measured with Na, Ne, H and I ranged from 1.06 ± 0.23 to 1.74 ± 0.48 , 1.04 ± 0.16 to 1.36 ± 0.36 , 0.09 ± 0.01 to 0.24 ± 0.11 , and 0.13 ± 0.03 to 0.33 ± 0.17 with their standard deviation from 0.23 to 0.44, 0.16 to 0.36, 0.01 to 0.11 and 0.03 to 0.17 respectively (Table 2). The genetic diversity in *C. infortunatum* was identified to be the highest with effective Na, Ne, H and I value of 1.74 ± 0.48 , 1.36 ± 0.36 , 0.24 ± 0.11 , and 0.33 ± 0.17 while lowest in *C. thomsoniae* with Na, Ne, H and I value of 1.06 ± 0.23 , 1.04 ± 0.16 , 0.09 ± 0.01 , and 0.13 ± 0.03 . The genetic variation accessed using RAPD markers indicated that the mean values of Ht, Hs, Gst, and Nm were 0.33 ± 0.02 , 0.14 ± 0.01 , 0.57 and 0.44 respectively (Support information Table S3). The Gst value of 0.57 indicates about 57% of the total genetic divergence among the population and the remaining 43% within the populations.

SCoT analysis

In the case of SCoT, 145 clear and scoreable bands were amplified with 17 screened primers. The numbers of amplified bands were higher than compared of ISSR and RAPD markers. A total of 129 polymorphic scoreable bands were revealed by each primer with an average of 7.59 for each individual (Table 1). The PIC of SCoT ranged from 0.30 (S41) to 0.62 (S23, S43) with a mean of 0.57 per primer. Similarly, the values of MI and Rp for the genotypes varied from 0.95 (S41) to 7.35 (S17) and 4.54 (S6) to 8.65 (S17), respectively. The genetic diversity values of Na, Ne, H and I ranged 1.26 ± 0.44 to 1.83 ± 0.39 , 0.16 ± 0.29 to 1.50 ± 0.37 , 0.17 ± 0.06 to 0.29 ± 0.16 and 0.25 ± 0.05 to 0.43 ± 0.25 with an average of 1.58 ± 0.46 , 1.35 ± 0.36 , 0.23 ± 0.11 , and 0.34 ± 0.16 (Table 2). These SCoT primers reveal highest genetic diversity in *C. infortunatum* population (Na— 1.83 ± 0.39 , Ne— 1.50 ± 0.37 , H— 0.29 ± 0.16 , I— 0.43 ± 0.25) compared to other populations while lowest

Table 2 Genetic diversity parameters of *Clerodendrum* sp. based on ISSR, RAPD, and SCoT markers using POPGENE software

Markers	Species	Na ^a (mean ± SD)	Ne ^b (mean ± SD)	H ^c (mean ± SD)	I ^d (mean ± SD)
ISSR	<i>C. colebrookianum</i>	1.46 ± 0.50	1.24 ± 0.36	0.19 ± 0.11	0.27 ± 0.17
	<i>C. infortunatum</i>	1.53 ± 0.50	1.30 ± 0.37	0.26 ± 0.10	0.28 ± 0.08
	<i>C. philippinum</i>	1.46 ± 0.50	1.25 ± 0.33	0.18 ± 0.09	0.27 ± 0.07
	<i>C. inerme</i>	1.16 ± 0.39	1.09 ± 0.23	0.13 ± 0.06	0.20 ± 0.09
	<i>C. indicum</i>	1.48 ± 0.50	1.25 ± 0.33	0.18 ± 0.08	0.27 ± 0.12
	<i>C. serratum</i>	1.22 ± 0.41	1.14 ± 0.29	0.16 ± 0.08	0.24 ± 0.04
	<i>C. japonicum</i>	1.19 ± 0.37	1.09 ± 0.27	0.20 ± 0.11	0.21 ± 0.11
	<i>C. paniculatum</i>	1.21 ± 0.41	1.13 ± 0.27	0.16 ± 0.07	0.23 ± 0.13
	<i>C. thomsoniae</i>	1.30 ± 0.46	1.19 ± 0.33	0.18 ± 0.06	0.26 ± 0.06
	Average	1.33 ± 0.45	1.19 ± 0.31	0.18 ± 0.09	0.25 ± 0.09
RAPD	<i>C. colebrookianum</i>	1.49 ± 0.50	1.28 ± 0.37	0.19 ± 0.09	0.28 ± 0.18
	<i>C. infortunatum</i>	1.74 ± 0.48	1.36 ± 0.36	0.24 ± 0.11	0.33 ± 0.17
	<i>C. philippinum</i>	1.60 ± 0.49	1.33 ± 0.33	0.20 ± 0.10	0.31 ± 0.15
	<i>C. inerme</i>	1.21 ± 0.41	1.12 ± 0.26	0.15 ± 0.07	0.22 ± 0.12
	<i>C. indicum</i>	1.66 ± 0.44	1.33 ± 0.33	0.21 ± 0.12	0.32 ± 0.24
	<i>C. serratum</i>	1.29 ± 0.45	1.19 ± 0.33	0.18 ± 0.08	0.27 ± 0.14
	<i>C. japonicum</i>	1.36 ± 0.48	1.20 ± 0.32	0.18 ± 0.10	0.26 ± 0.06
	<i>C. paniculatum</i>	1.40 ± 0.49	1.31 ± 0.40	0.21 ± 0.11	0.31 ± 0.15
	<i>C. thomsoniae</i>	1.06 ± 0.23	1.04 ± 0.16	0.09 ± 0.01	0.13 ± 0.03
	Average	1.41 ± 0.44	1.24 ± 0.32	0.18 ± 0.09	0.27 ± 0.14
SCoT	<i>C. colebrookianum</i>	1.81 ± 0.40	1.47 ± 0.36	0.28 ± 0.14	0.42 ± 0.25
	<i>C. infortunatum</i>	1.83 ± 0.39	1.50 ± 0.37	0.29 ± 0.16	0.43 ± 0.25
	<i>C. philippinum</i>	1.70 ± 0.46	1.41 ± 0.36	0.25 ± 0.12	0.37 ± 0.14
	<i>C. inerme</i>	1.62 ± 0.49	1.36 ± 0.36	0.21 ± 0.10	0.32 ± 0.18
	<i>C. indicum</i>	1.73 ± 0.45	1.39 ± 0.33	0.24 ± 0.11	0.36 ± 0.25
	<i>C. serratum</i>	1.48 ± 0.50	1.33 ± 0.38	0.21 ± 0.10	0.30 ± 0.10
	<i>C. japonicum</i>	1.48 ± 0.50	1.32 ± 0.39	0.20 ± 0.09	0.30 ± 0.13
	<i>C. paniculatum</i>	1.35 ± 0.48	1.25 ± 0.37	0.20 ± 0.09	0.29 ± 0.09
	<i>C. thomsoniae</i>	1.26 ± 0.44	1.16 ± 0.29	0.17 ± 0.06	0.25 ± 0.05
	Average	1.58 ± 0.46	1.35 ± 0.36	0.23 ± 0.11	0.34 ± 0.16

^aNa—number of observed alleles^bNe—number of effective alleles^cH—Nei's genetic diversity^dI—Shannon's index

genetic diversity was found in *C. thomsoniae* (1.26 ± 0.44, 1.16 ± 0.29, 0.17 ± 0.06, 0.25 ± 0.05). The mean values of genetic variation (Ht, Hs, Gst, and Nm) accessed with SCoT markers were 0.34 ± 0.03, 0.21 ± 0.11, 0.59 and 0.79, respectively (Support information Table S3). According to the Gst value of 0.59 indicates that about 59% of total genetic divergence among the populations and the remaining 41% within the populations.

Genetic diversity and cluster analysis in *Clerodendrum*

Jaccard's similarity coefficient was used to estimate the genetic diversity and relationships among the *Clerodendrum* sp. Based on ISSR, RAPD, and SCoT data, the genetic

similarity ranged from 0.98 (Cine6 with Cine5, Cine7 with Cine4, Cine9 with Cine8, Cind4 with Cind2, Cind3, Cind5, Cind8 with Cind9 and Cind10, Cjap4 with Cjap2, Cjap3, Cpan3 with Cpan5) to 0.57 (Cinf2 and Cphi6) for ISSR; 0.99 (Cinf12 with Cinf13 and Cine8 with Cine9) to 0.50 (Cinf9 and Cjap4) for RAPD and 0.99 (Ctho6 and Ctho7) to 0.41 (Ccol14 and Cine5) respectively.

From the dendrogram constructed by NTSys using the UPGMA method for ISSR markers, 6 major clusters were generated (Fig. 1a). The cluster 1 includes 41 genotypes that divided into 2 sub-clusters, sub-cluster 1a with 23 genotypes of *C. colebrookianum* (72.2%) and *C. philippinum* (90.9%) and sub-cluster 1b with 18 genotypes *C. indicum* (33.3%), *C. japonicum* (100%) and *C. paniculatum* (100%). Cluster 2 consist of 20 genotypes, 10 each of *C. inerme* (100%) and

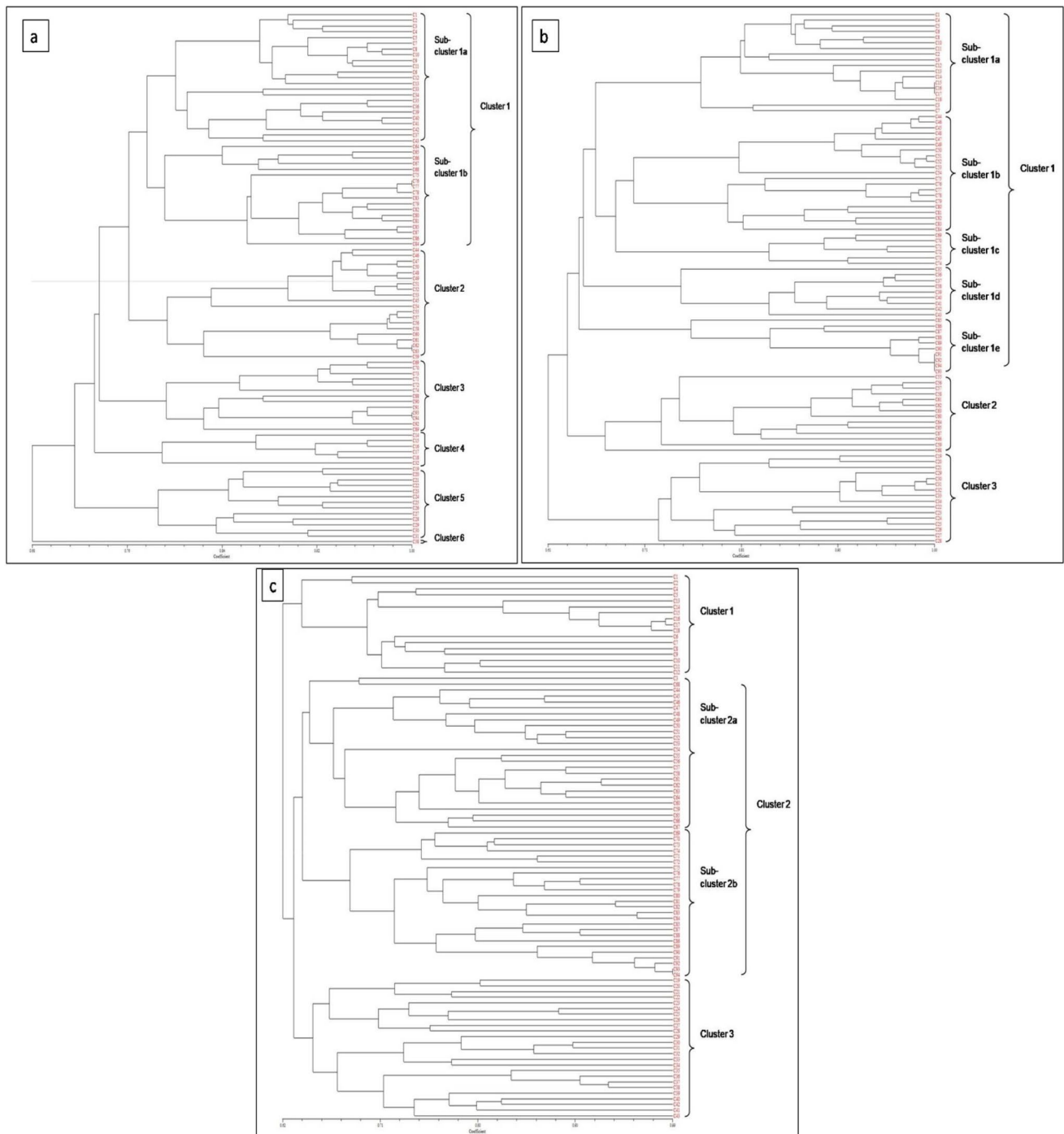


Fig. 1 UPGMA tree-based grouping of 94 genotypes of *Clerodendrum* sp. based on a-ISSR, b-RAPD and c-ScoT markers. In the figure *Clerodendrum* sp. were represented as C1–C18 = *C. colebrookianum*; C19–C32 = *C. infortunatum*; C33–C43 = *C. philippinum*;

C44–C53 = *C. inerme*; C54–C68 = *C. indicum*; C69–C74 = *C. serratum*; C75–C82 = *C. japonicum*; C83–C87 = *C. paniculatum*; C88–C94 = *C. thomsoniae*

C. indicum (66.7%), Cluster 3 with 13 genotypes, 6 genotypes of *C. serratum* (100%) and 7 genotypes of *C. thomsoniae* (100%), Cluster 4 with 6 genotypes, 5 genotypes of *C. colebrookianum* (27.8%) and 1 genotype of *C. infortunatum* (7.1%), Cluster 5 and Cluster 6 contain separate clusters with

13 genotypes of *C. infortunatum* (92.9%) and 1 genotype of *C. philippinum* (9.1%).

Based on RAPD cluster analysis, 3 major clusters (Cluster 1, Cluster 2 and Cluster 3) were found (Fig. 1b). In Cluster 1, 64 genotypes were grouped whereas, Cluster 2 and

Cluster 3 comprises 14 genotypes of *C. indicum* (93.4%) and 14 genotypes of *C. infortunatum* (100%) along with 2 genotypes of *C. philippinum* (9.1%). The cluster 1 can be subdivided into 5 sub-clusters as Cluster 1a which consist of 18 genotypes of *C. colebrookianum* (100%); Cluster 1b consist of 20 genotypes of *C. inerme* (100%), *C. japonicum* (100%), *C. indicum* (6.6%), *C. paniculatum* (40%); Cluster 1c consist of 6 genotypes of *C. serratum* (100%); Cluster 1d consist of 9 genotypes of *C. philippinum* (90.9%) and Cluster 1e consist of 10 genotypes of *C. paniculatum* (60%) and *C. thomsoniae* (100%).

In cluster analysis of SCoT markers, the genotypes were divided into 3 major clusters (Fig. 1c). Cluster 1 contained 17 genotypes of *C. colebrookianum* (94.4%). In Cluster 2, 2 sub-clusters were formed as Cluster 2a with 26 genotypes of *C. colebrookianum* (5.6%), *C. indicum* (6.6%) and *C. inerme* (100%) and Cluster 2b with 26 genotypes of *C. indicum* (93.3%), *C. serratum* (100%), *C. japonicum* (100%), *C. paniculatum* (100%), and *C. thomsoniae* (100%). Cluster 3 consists of 25 genotypes of *C. infortunatum* (100%) and *C. philippinum* (100%).

Structure-based study

Population structure was analyzed for 94 genotypes of *Clerodendrum* species with three markers and evaluated with Structure harvester software (Fig. 2). Based on membership fractions, the genotypes with $\geq 80\%$ probability was assigned to corresponding populations while other was categorized as admixture type.

According to Evanno's method, the maximum peak of ΔK was observed at $K=6$. This result indicates that six distinct populations were obtained with ISSR markers similar to the result of UPGMA cluster analysis. The individuals of 14 genotypes were grouped in Cluster 1 that includes all genotypes of *C. japonicum* and *C. paniculatum* along with 1 genotype of *C. colebrookianum*. In Cluster 2 and 3, all genotypes of *C. infortunatum*, *C. indicum*, and *C. thomsoniae* were found. The 17 genotypes of *C. colebrookianum* and 5 genotypes of *C. philippinum* were grouped in Cluster 4. Cluster 5 includes the entire genotypes of *C. inerme* with 1 each genotype of *C. indicum* and *C. philippinum*. Cluster 6 contained genotypes of *C. serratum* (6) and *C. philippinum* (3). 62 genotypes were pure and 32 genotypes were admixture type.

In structure analysis using the RAPD dataset, the maximum peak of ΔK was observed at $K=6$ followed by leveling off and accompanied by an increase in variance. This result indicates 6 distinct populations of *Clerodendrum* sp. Out of 94 genotypes; Cluster 1 contained 18 genotypes among which 14 genotypes were from *C. indicum* and 5 genotypes from *C. serratum*. Cluster 2 (21 genotypes) includes genotypes from *C. philippinum* (9), *C. japonicum* (8), *C.*

paniculatum (3), and *C. serratum* (1) were generated. Cluster 3, 4, 5 and 6 includes all the genotypes of *C. thomsoniae*, *C. infortunatum*, *C. colebrookianum*, and *C. inerme*. Among all the genotypes, 77 were identified as pure and 17 genotypes were found to be as admixture type.

Four distinct populations were obtained with SCoT markers as high value of ΔK was observed at $K=4$. In Cluster 1, 10 genotypes of *C. colebrookianum* were found. Similarly, Cluster 2 consists of 12 genotypes of *C. indicum*. Cluster 3 consists of 44 genotypes which includes 11 genotypes of *C. philippinum*, 10 genotypes of *C. inerme*, 14 genotypes of *C. infortunatum*, 8 genotypes of *C. colebrookianum* and 1 genotype of *C. serratum*. Cluster 4 consists of 28 genotypes of *C. indicum* (8 genotypes), *C. thomsoniae* (7 genotypes), *C. japonicum* (8 genotypes) and *C. serratum* (5 genotypes). 85 genotypes were pure and 9 were found to be admixture genotypes.

Analysis of molecular variance (AMOVA) based study

AMOVA was studied to examine the population differentiation among and within *Clerodendrum* species and considered as statistically significant if $P < 0.001$ (Support information Fig. S1). The ISSR data showed 47% among the population and 53% within the population. Similarly, RAPD based AMOVA analysis showed 53% and 47% variations among and within the populations. The SCoT marker based AMOVA resulted in the least variation among the population (32%) and maximum variation (68%) within the population of *Clerodendrum* species. The present study of three marker systems showed that population variation was partitioning between 32–53% and maximum variation was present within-population ranging from 68–47%. Additionally, SCoT marker showed high genetic variation within and among the populations of *Clerodendrum* species in contrast to ISSR and RAPD markers.

Principle coordinate analysis (PCoA) based study

The Principle coordinate analysis was performed to ensure the spatial representation of genetic distance among the individuals and to analyze the consistency of differentiation among the population denied by cluster analysis. The projection of PCoA was plotted on a two-dimensional scatter plot (Fig. 3). In ISSR based analysis, the percentage of total variations explained by the first three principle components analysis were 15.80%, 14.36%, 12.18%, respectively. The first two principle axes accounted for 30.16 of the total molecular variances. Three species *C. infortunatum*, *C. indicum* and *C. inerme* were clustered with significant distinctive groups.

Based on RAPD data, total percentage of variation with first three axes were 42.27% (axis1-18.86%, axis2-13.08% and axis3-10.33%). The species of *C. colebrookianum*, *C.*

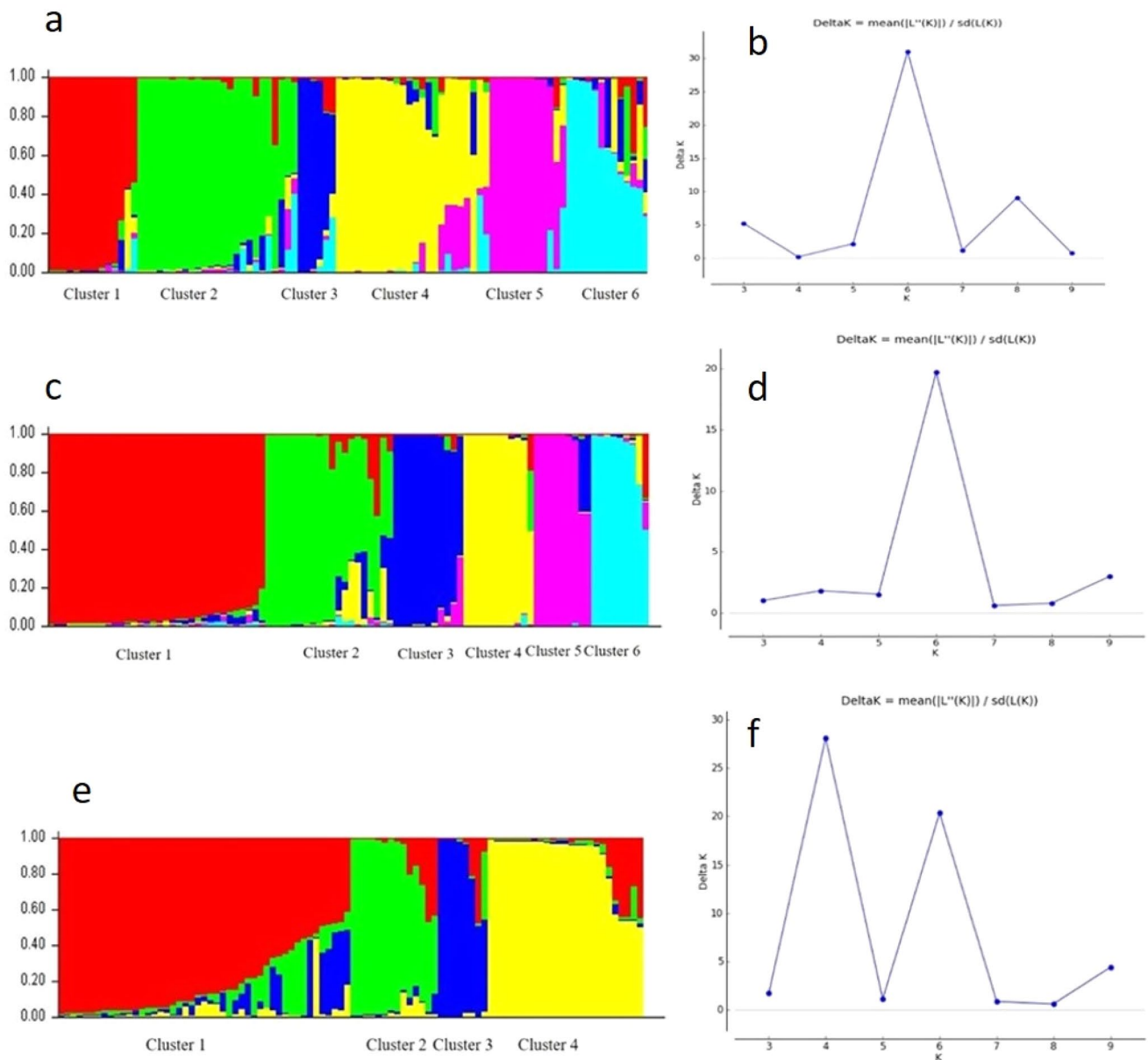


Fig. 2 Population structure of 94 *Clerodendrum* genotypes inferred by STRUCTURE analysis based on **a** ISSR, **c** RAPD and **e** SCoT markers. The vertical bar with different color indicates the estimated membership coefficient for each genotype using Q statistic. Graph

represent the estimated membership fraction using LnP(D)-derived ΔK with cluster number (K) ranged from 1 to 10 where **b** RAPD, **d** ISSR, **f** SCoT

infortunatum, *C. thomsoniae* and *C. indicum* form distinct clusters. The PCoA analysis based on SCoT markers showed similarity with dendrogram and 24.94% of the total variation was described by three coordinates as PCoA1-9.61, PCoA2-8.66 and PCoA3-6.67 respectively.

Combine data analysis for diversity and population structure

The data from three markers (ISSR, RAPD, and SCoT) were summarized to understand better the diversity and

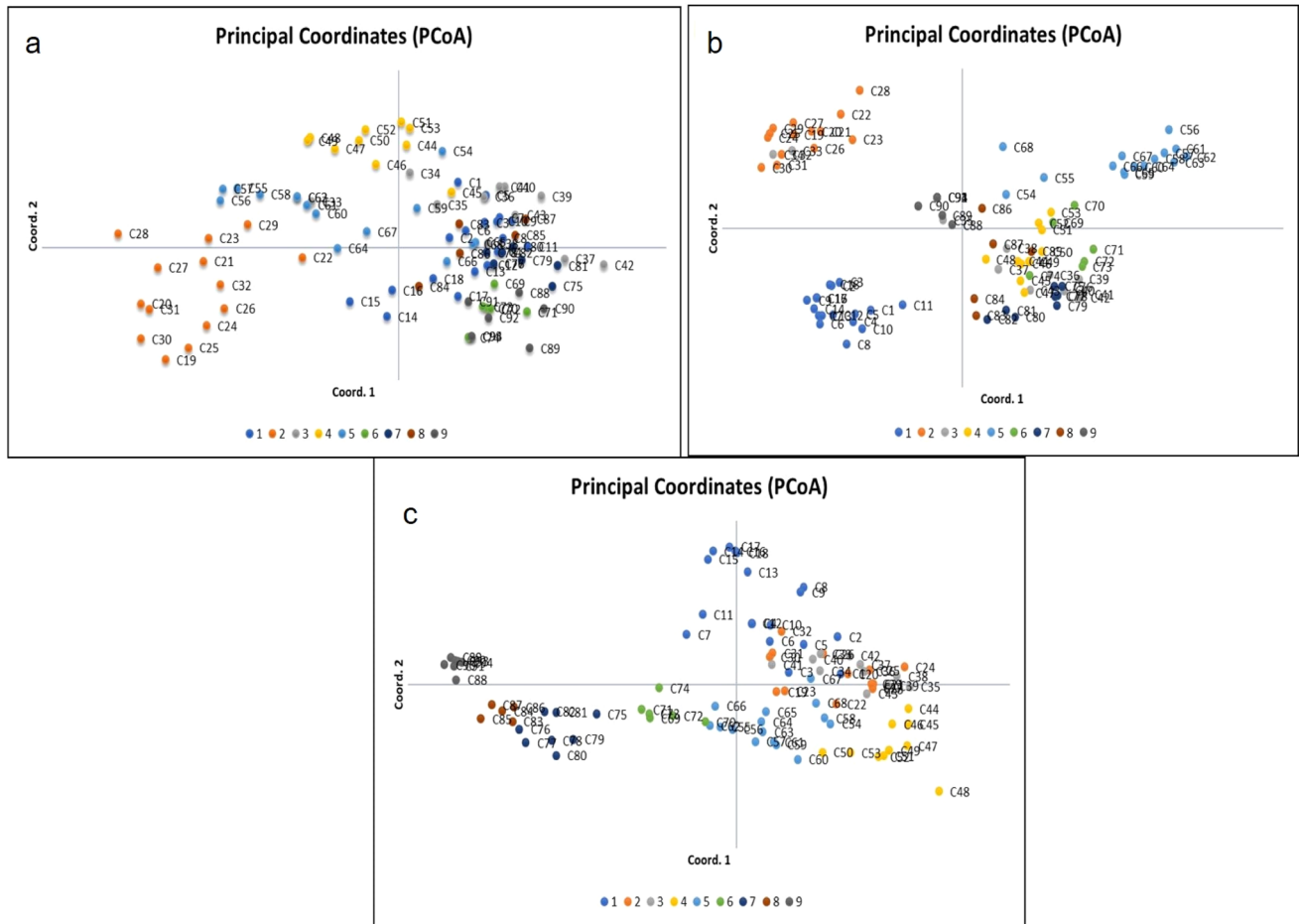


Fig. 3 Principle coordinates analysis of 94 genotypes of *Clerodendrum* sp. based on genetic similarity matrix derived from a- ISSR, b- RAPD and c- SCoT markers. In the figure *Clerodendrum* sp. were represented with coloured dots and sample I.D as: ● C1–C18 = *C.*

colebrookianum; ● C19–C32 = *C. infortunatum*; ○ C33–C43 = *C. philippinum*; ● C44–C53 = *C. inerme*; ● C54–C68 = *C. indicum*; ● C69–C74 = *C. serratum*; ● C75–C82 = *C. japonicum*; ● C83–C87 = *C. paniculatum*; ● C88–C94 = *C. thomsoniae*

population structure. The dendrogram constructed using UPGMA tree showed that all the genotypes were grouped into three clusters (Support information Fig. S2). Cluster 1 consist of 18 genotypes of *C. colebrookianum* (100%); Cluster 2 contains 60 genotypes with two sub-clusters as Cluster 2a with 9 genotypes of *C. philippinum* (81.8%) and Cluster 2b with 51 genotypes of *C. inerme* (100%), *C. indicum* (100%), *C. serratum* (100%), *C. japonicum* (100%), *C. paniculatum* (100%) and *C. thomsoniae* (100%). Clusters 3 consist of 16 genotypes with 14 genotypes of *C. infortunatum* and 2 genotypes of *C. philippinum* (18.2%). The cophenetic coefficient for three molecular markers indicates a good fit for clustering with values of 0.86 for ISSR, 0.89 for RAPD and 0.90 for SCoT. The Mantel test correlation values depicted a positive correlation between the three marker types. The correlation coefficient (r) was 0.37 between ISSR and RAPD, 0.55 between ISSR and SCoT and 0.42 between RAPD and SCoT. The PCoA based on combined markers grouped the genotype similar

to UPGMA clustering method. The first and second principle axes were 9.81 and 8.80% of the variation, respectively. The structures of 94 *Clerodendrum* genotypes were further analyzed with STRUCTURE software using the Bayesian clustering model. The result showed that the highest value of ΔK was observed for $K=4$ which indicates three distinct populations (Support information Fig. S3). Cluster 1 consists of 32 genotypes which include 10 genotypes of *C. inerme*, 15 genotypes of *C. indicum*, 4 genotypes of *C. philippinum* and 3 genotypes of *C. serratum*. Cluster 2 consists of 39 genotypes of *C. infortunatum* (14 genotypes), *C. colebrookianum* (18 genotypes) and *C. philippinum* (7 genotypes). Cluster 3 contains 23 genotypes of *C. thomsoniae* (7 genotypes), *C. japonicum* (8 genotypes), *C. paniculatum* (5 genotypes) and *C. serratum* (3 genotypes). 84 genotypes were pure and 10 genotypes were found to be admixture type. AMOVA analysis also revealed only 42% diversity among populations and 58% within variations within populations. The principal

coordinate analysis showed that the first three coordinates explained 25.56% of total variations.

Discussion

The importance of genetic diversity with informative molecular markers could lead to genetic improvement, germplasm maintenance and identification of unique genotype/individuals [23]. In the present study, three PCR based dominant molecular markers, ISSR (12 primers), RAPD (16 primers), and SCoT (17 primers) were selected to access the genetic diversity among 9 species of *Clerodendrum*. The ISSR, RAPD and SCoT markers generated a high percentage of polymorphism within the genotypes and specify higher informative nature of markers [24]. The level of polymorphism was maximum in SCoT marker (88.6%) followed by RAPD (88.23%) and ISSR (71.41%) markers. Similar results were reported in crops like grape [25], Cicer [26], peanut [27] and mango [28]. Moreover, the higher levels of polymorphism in RAPD compared to ISSR were also reported earlier in rice bean [29] and *Myristica* spp.[30]. The mean of marker index (MI) that evaluates the overall efficacy of a marker system was almost equivalent for both RAPD and SCoT markers as compared to ISSR markers. These closet values of marker index indicate equal efficiency for fingerprinting of *Clerodendrum* species.

The different genetic parameters (Na, Ne, H and I) were found highest in SCoT markers followed by RAPD and ISSR markers. According to Nei's, G_{st} is measured low when G_{st} < 0.05, medium if 0.05 < G_{st} < 0.15, and high if G_{st} > 0.15. The population of genetic differentiation was maximum with SCoT (G_{st}-0.59) followed by RAPD (G_{st}-0.57) and ISSR (G_{st}-0.33). This high level of population variation may occur due to numerous factors such as species breeding system, genetic drift or geographical origin of the populations [31, 32]. The value of Nm < 0.5 indicates a restriction of genetic drift due to an adequate amount of gene flow and Nm > 1.0 indicate initiation of variation as a result of genetic drift in the populations [33]. The observed Nm value in SCoT (Nm-0.79), ISSR (Nm-0.38), and RAPD (0.44) specify extensive gene flow among the populations. Incongruent with RAPD and ISSR marker system, SCoT could generate more information correlated with biological traits and higher genetic polymorphism [9]. SCoT markers are easy to be amplified as compared to RAPD and ISSR due to the presence of elongated oligonucleotides [34]. Various studies reported the efficacy of SCoT marker in diverse plant species such as Potato [35], Tomato [36], Wheat [37], Kalmegh [8], *Vigna unguiculate* [38] and Taxus [39].

Furthermore, the UPGMA dendrogram of ISSR, RAPD, and SCoT generated similar grouping of genotypes with minor deviations in ISSR marker. The deviation in the

clustering of genotypes could occur due to variation in the size of target genome [40]. The comparison of the clustering pattern in this study revealed that the location specificity of SCoT was higher than ISSR and RAPD markers as well as similar to the results of clusters analysis obtain with PCoA. Jaccard's coefficient was analyzed to access the genetic diversity of *Clerodendrum* sp. ISSR showed highest similarity (0.98) among *C. inerme* (Cine6–Cine5, Cine7–Cine4, Cine9–Cine8), *C. indicum* (Cind4–Cind2, Cind3, Cind8–Cind9, Cind10), *C. japonicum* (Cjap4–Cjap2, Cjap3) and *C. paniculatum* (Cpan3–Cpan5) and lowest (0.57) in *C. infortunatum* (Cinf2) with *C. philippinum* (Cphi6). Similarly, RAPD showed maximum similarity (0.99) with *C. infortunatum* (Cinf12–Cinf13) and *C. inerme* (Cine8–Cine9) and lowest (0.50) in *C. infortunatum* (Cinf9) with *C. japonicum* (Cjap4). In SCoT highest similarity (0.99) was observed in *C. thomsoniae* (Ctho6–Ctho7) and lowest (0.41) in *C. colebrookianum* (Ccol14) with *C. inerme* (Cine5). It can be further concluded that *C. infortunatum*, *C. philippinum*, *C. japonicum*, *C. colebrookianum*, and *C. inerme* were distantly related from each other.

The grouping of dendrogram clusters was further supported by Bayesian clustering algorithm using STRUCTURE software that determines the true number of cluster (K) in a given sample of genotypes using an ad hoc statistic ΔK based on the rate of change in the log probability of data between successive values of K [21]. The best interference of delta K (ΔK) was obtained as K = 6 for ISSR, K = 6 for RAPD and K = 4 for SCoT. The AMOVA analysis indicated that genetic diversity by ISSR and SCoT except RAPD markers were distributed more than the diversity among the populations. Further, this could help in strategy development for species collection and estimation. The low variability among the populations was also reported in fig [41] and tetraploid potato [35].

The combinations of molecular markers could analyze various regions of the genome in both functional and neutral regions [42]. In this study, ISSR, RAPD, and SCoT were combined to analyze the genetic diversity of *Clerodendrum* sp. The dendrogram of ISSR + RAPD + SCoT showed similar grouping with the dendrogram of RAPD and SCoT molecular markers. The dendrogram displayed closer relation based on SM similarity coefficients in genotypes of *C. thomsoniae* (C93–C94) and distance relation between *C. infortunatum* (C20) and *C. serratum* (C71). The correlation between genetic distances and geographical origin of *Clerodendrum* genotypes suggests that the natural selection had significantly exaggerated the genome regions that were amplified with ISSR, RAPD and SCoT markers.

Our present study revealed a high level of genetic diversity among the species *Clerodendrum* genotypes and this level of genetic variation could be useful in systemic germplasm management and utilization in plant breeding

programmed. The molecular marker SCoT followed by RAPD and ISSR showed a high level of polymorphism efficiency and could be further used in fingerprinting of *Clerodendrum* species.

Conclusion

Genetic diversity is a crucial means to analyze short- and long- time survival of species in their natural habitats. In this study, 94 genotypes from nine species of *Clerodendrum* were evaluated using ISSR, RAPD and SCoT markers and revealed a high level of genetic diversity. Based on values of PIC, Rp, MI, Ne, H, I, Ht, Hs and Nm, SCoT markers showed high level of discriminatory and informative values in comparison to ISSR and RAPD markers. UPGMA analysis revealed similar clustering of genotypes for RAPD and SCoT marker with minor deviation in ISSR markers and depict similar to results with PCoA analysis. In AMOVA analysis, the maximum difference was 53% among the populations whereas, maximum variation was 68% within the populations. Structure-based population using three markers showed different population number which does not correspond to UPGMA tree except for SCoT based population structure. Therefore, we tentatively could propose that SCoT markers could be useful in differentiation nine species and various other representatives of the *Clerodendrum* genus. Further, this study could provide valuable baseline data on population genetics of *Clerodendrum* species of North East India for better marker-assisted selections, selection of marker technologies for breeding and utilization of comparative genetic studies.

Acknowledgements Barbi Gogoi is thankful to Dr. G.N. Sastry, Director, CSIR-North-East Institute of Science and Technology, Jorhat, Assam, India and Dr. S.P. Saikia and Dr. S.B. Wann for their consistent support in completing the manuscript.

Author contributions BG performed the experiment, analysed data and wrote the manuscript. SPS designed the project and wrote the manuscript and SBW provided technical support.

Funding This study is financially supported by “CSIR-Direct SRF” fellowship under CSIR-HR Division, New Delhi.

Compliance with ethical standards

Conflicts of interest The authors declare no conflict of interests.

Research involving human participants or animals This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Deori C, Roy DK, Talukdar SR, Pagag K, Sarma N (2013) Diversity of the genus *Clerodendrum* Linnaeus (Lamiaceae) in North-east India with special reference to Barnadi Wildlife Sanctuary. *Assam Pleione* 7(2):473–488
- Hu HJ, Liu Q, Yang YB, Yang L, Wang ZT (2014) Chemical constituents of *Clerodendrum trichotomum* leaves. *Zhong Yao Cai* 37:1590–1593
- Steane DA, Scotland RW, Mabberley DJ, Olmstead RG (1999) Molecular systematics of *Clerodendrum* (Lamiaceae): ITS sequences and total evidence. *Am J Bot* 86:98–107
- Yuan YW, Mabberley DJ, Steane DA, Olmstead RG (2010) Further disintegration and redefinition of *Clerodendrum* (Lamiaceae): Implications for the understanding of the evolution of an intriguing breeding strategy. *Taxon* 59(1):125–133
- Shrivastava N, Patel T (2007) *Clerodendrum* and Healthcare: An Overview. *Med Aromat Plant Sci Biotechnol* 1(1):142–150
- Yadav A, Verma PK, Chand T, Bora HR (2018) Ethno-medicinal knowledge of *Clerodendrum* L. among different tribes of Nambor reserve forest, Assam, India. *J. Pharmacogn. Phytochem.* 7(5):1567–1570
- Yonggam D (2005) Ethno Medico-Botany on the Mishing tribe of East Siang district of Arunachal Pradesh. *Arun Forest News* 21:44–49
- Tiwari G, Singh R, Singh N, Choudhury DR, Paliwal R, Kumar A, Gupta V (2016) Study of arbitrarily amplified (RAPD and ISSR) and gene targeted (SCoT and CDBP) markers for genetic diversity and population structure in Kalmegh [*Andrographis paniculata* (Burm. f.) Nees]. *Ind Crops and Prod* 2016 86:1–1.
- Agarwal A, Gupta V, Haq SU, Jatav PK, Kothari SL, Kachhwaha S (2019) Assessment of genetic diversity in 29 rose germplasms using SCoT marker. *J King Saud Univ Sci* 31(4):780–788
- Verma KS, ul Haq S, Kachhwaha S, Kothari SL (2017) RAPD and ISSR marker assessment of genetic diversity in *Citrullus colocynthis* (L.) Schrad: a unique source of germplasm highly adapted to drought and high-temperature stress. *3 Biotech* 7(5):288.
- Satya P, Karan M, Jana S, Mitra S, Sharma A, Karmakar PG, Ray DP (2015) Start codon targeted (SCoT) polymorphism reveals genetic diversity in wild and domesticated populations of ramie (*Boehmeria nivea* L. Gaudich.), a premium textile fiber producing species. *Meta Gene* 3:62–70
- Mizusawa L, Kaneko S, Hasegawa M, Isagi Y (2011) Development of nuclear SSRs for the insular shrub *Clerodendrum izuin-sulare* (Verbenaceae) and the widespread *C. trichotomum*. *Am J Bot* 98(11):e333–6.
- Prasad MP, Sushant S, Chikkaswamy BK (2012) Phytochemical analysis, antioxidant potential, antibacterial activity and molecular characterization of *Clerodendrum* species. *Int J Mol Biol* 3(3):71–76
- Liu P, Que Y, Pan YB (2011) Highly polymorphic microsatellite DNA markers for sugarcane germplasm evaluation and variety identity testing. *Sugar Tech* 13(2):129–136
- Prevost A, Wilkinson MJ (1999) A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. *Theor Appl Genet* 98(1):107–112
- Rohlf FJ (1998) NTSYS: numerical taxonomy and multivariate analysis system version 2.02. Applied Biostatistics Inc., Setauket, NY.
- Jaccard P (1908) Nouvelles recherches sur la distribution florale. *Bull Soc Vaudoise Sci Nat* 44:223–270
- Yeh FC, Yang RC, Boyle T (1999) POPGENE version 1.32: Microsoft Windows-based freeware for population genetic analysis, quick user guide. Center for International Forestry Research, University of Alberta, Edmonton, Alberta, Canada.

19. Nei M (1973) Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* 70(12):3321–3323
20. Pritchard J, Wen X, Falush D (2012) Documentation for structure software: Version 2.3. 4. <https://www.pritchbsd.uchicago.edu/structure.html>
21. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620
22. Peakall RO, Smouse PE. (2012) GenALEX 6.5 genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28(19):2537–2539.
23. Vinceti B, Loo J, Gaisberger H, Van Zonneveld MJ, Schueler S, Konrad H (2013) Conservation priorities for *Prunus africana* defined with the aid of spatial analysis of genetic data and climatic variables. *PLoS ONE* 8(3):e59987
24. Pandey MK, Gautami B, Jayakumar T, Sriswathi M, Upadhyaya HD, Gowda MVC, Radhakrishnan T, Bertoli D, Knapp SJ, Cook DR, Varshney RK (2012) Highly informative genic and genomic SSR markers to facilitate molecular breeding in cultivated groundnut (*Arachis hypogaea*). *Plant Breed* 131:139–147
25. Guo DL, Zhang JY, Liu CH (2012) Genetic diversity in some grape varieties revealed by SCoT analyses. *Mol Biol Rep* 39:5307–5313
26. Amirmoradi B, Talebi R, Karami E (2012) Comparison of genetic variation and differentiation among annual *Cicer* species using start codon targeted (SCoT) polymorphism, DAMD-PCR, and ISSR markers. *Plant Syst Evol* 298:1679–1688
27. Xiong F, Zhong R, Han Z, Jiang J, He L, Zhuang W, Tang R (2011) Start codon targeted polymorphism for evaluation of functional genetic variation and relationships in cultivated peanut (*Arachis hypogaea* L.) genotypes. *Mol Biol Rep* 38:3487–3494
28. Luo C, He XH, Chen H, Hu Y, Ou SJ (2012) Genetic relationship and diversity of *Mangifera indica* L.: revealed through SCoT analysis. *Genet Resour Crop Evol* 59:1505–1515
29. Muthusamy S, Kanagarajan S, Ponnusamy S (2008) Efficiency of RAPD and ISSR markers system in accessing genetic variation of rice bean (*Vigna umbellata*) landraces. *Electron J Biotechnol* 11(3):32–41
30. Sheeja TE, Sabeesh C, Shabna OV, Shalini RS, Krishnamoorthy B (2013) Genetic diversity analysis of *Myristica* and related genera using RAPD and ISSR markers. *JOSAC* 22:38–46
31. Ying-Xiong QI, Xin-Wen ZH, Cheng-Xin FU, Chan YS (2005) A preliminary study of genetic variation in the endangered, Chinese endemic species *Dyosma versipellis* (Berberidaceae). *Bot Bull Acad Sinica* 46.
32. Rafizadeh A, Koohi-Dehkordi M, Sorkheh K (2018) Molecular insights of genetic variation in milk thistle (*Silybum marianum* [L.] Gaertn.) populations collected from southwest Iran. *Mol Bio Rep* 45(4):601–9.
33. Wright S (1969) Evolution and genetics of populations: The theory of gene frequencies. Chicago Univ. Press, Chicago, USA, Dogan NY, Kurt P, Osma E. Determination of genetic diversity of *Sonchuser zincanicus* Matthews (Asteraceae), a Critically Endangered plant endemic to Turkey using RAPD markers. *Biharean Biologist*. 2018 Dec 1;12(2).
34. Collard BCY, Mackill DJ (2009) Start codon targeted (SCoT) polymorphism: a simple, novel DNA marker technique for generating gene targeted markers in plants. *Plant Mol Biol Report* 27:86–93
35. Gorji AM, Poczai P, Polgar Z, Taller J (2011) Efficiency of arbitrarily amplified dominant markers (SCoT, ISSR and RAPD) for diagnostic fingerprinting in tetraploid potato. *Am J Potato Res* 88(3):226–237
36. Shahlaei A, Torabi S, Khosroshahli M (2014) Efficacy of SCoT and ISSR markers in assessment of tomato (*Lycopersicon esculentum* Mill.) genetic diversity. *Int J Biol Sci* 5:14–22
37. Hamidi H, Talebi R, Keshavarzi F (2014) 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* 42(4):558–67.
38. Igwe DO, Afiukwa CA, Ubi BE, Ogbu KI, Ojuederie OB, Ude GN (2017) Assessment of genetic diversity in *Vigna unguiculata* L. (Walp) accessions using inter-simple sequence repeat (ISSR) and start codon targeted (SCoT) polymorphic markers. *BMC Gent* 18(1):98.
39. Hao Q, Lu C, Zhu Y, Xiao Y, Gu X (2018) Numerical investigation into the evolution of groundwater flow and solute transport in the Eastern Qaidam Basin since the Last Glacial Period. *Geofluids* 2018.
40. Costa R, Pereira G, Garrido I, Tavares-de-Sousa MM, Espinosa F (2016) Comparison of RAPD, ISSR, and AFLP molecular markers to reveal and classify orchard grass (*Dactylis glomerata* L.) germplasm variations. *PloS one* 11(4):e0152972.
41. Ikegami H, Nogata H, Hirashima K, Awamura M, Nakahara T (2009) Analysis of genetic diversity among European and Asian fig varieties (*Ficus carica* L.) using ISSR, RAPD, and SSR markers. *Genet Resour Crop Evol* 56:201–209
42. Satish L, Shilpha J, Pandian S, Rency AS, Rathinapriya P, Ceasar SA, Largia MJ, Kumar AA, Ramesh M (2016) Analysis of genetic variation in sorghum (*Sorghum bicolor* (L.) Moench) genotypes with various agronomical traits using SPAR methods. *Gene* 576:581–585

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