



Estimation of genetic diversity and population structure in *Tinospora cordifolia* using SSR markers

Suchita Lade^{1,2} · Veena Pande² · Tikam Singh Rana^{3,4} · Hemant Kumar Yadav^{1,4}

Received: 8 February 2020 / Accepted: 11 June 2020 / Published online: 16 June 2020
© King Abdulaziz City for Science and Technology 2020

Abstract

Thirty polymorphic SSRs, derived from RNA sequencing of *Tinospora cordifolia* (willd.), were utilized for genetic diversity and population structure evaluation among 96 accessions collected from ten different geographical regions of India. A total of 7611 SSRs were identified from 268149 transcripts. Of all SSR loci, 69.07% of them were tri-nucleotide repeat motifs, followed by di-nucleotide repeat motifs (12.82%). A total of 230 alleles were generated by 30 SSRs with an average of 7.67 alleles per locus with comparatively higher polymorphic information content (average 0.68). The expected (H_e) and observed (H_o) heterozygosity means were 0.71 and 0.12, respectively. All the loci showed significant deviation from Hardy–Weinberg Equilibrium (HWE). The neighbor joining clustering based on jaccard's coefficient grouped all the 96 accessions into three major cluster which was also in congruence with model-based structure plot. The result of molecular variance (AMOVA) revealed higher genetic variance within populations than among populations. The result reflects an existence of high level of genetic diversity in the collected accessions of *T. cordifolia*. The accessions Tc131, Tc31, Tc129, Tc38, Tc16, Tc59, Tc60, Tc17, Tc106 and Tc130 was found to be potential and diverse in nature and the SSRs TCSSR-18, TCSSR-37, TCTSSR-59, TCTSSR-92, TCTSSR-123 and TCTSSR-126 as potential markers. These accessions and newly developed SSR markers provide valuable resource and could be strategically utilized for further genetic improvement of *T. cordifolia*.

Keywords *Tinospora cordifolia* · Simple sequence repeat (SSR) · Genetic diversity · AMOVA

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13205-020-02300-7>) contains supplementary material, which is available to authorized users.

✉ Hemant Kumar Yadav
h.yadav@nbri.res.in

Suchita Lade
suchitalade.sl@gmail.com

Veena Pande
veena_kumaun@yahoo.co.in

Tikam Singh Rana
ranats@nbri.res.in

- ¹ Plant Genetic Resources and Improvement Division, CSIR-National Botanical Research Institute, Rana Pratap Marg, Lucknow, Uttar Pradesh 226001, India
- ² Department of Biotechnology, Kumaun University, Bhimtal Campus, Nainital, Uttarakhand, India
- ³ Plant Diversity, Systematics and Herbarium Division, CSIR-National Botanical Research Institute, Rana Pratap Marg, Lucknow, Uttar Pradesh, 226001, India
- ⁴ Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India

Introduction

Tinospora cordifolia (Willd.) Hook.f. & Thomsom (Menispermaceae, $2n = 22$) is one of the important traditional medicinal plants described in the Ayurveda system of medicine. It is commonly known as Giloy (Hindi) and Guduchi (Sanskrit), and is distributed throughout the tropical parts of Indian subcontinent (Sri Lanka and Bangladesh) and also in China (Kumawat et al. 2019; Spandana et al. 2013). It is a large, glabrous, deciduous and succulent climber with greenish-yellow flower. The leaves are simple, alternate, heart-shaped and half way round with long petiole. Fruits are yellowish-red in color, fleshy, in clusters of 1-3 and rounded on dense stalk with sub-terminal scars. The major active compounds of *T. cordifolia* includes alkaloids, diterpenoids, lactones, glycosides, steroids, sesquiterpenoids, phenolics, aliphatic compounds and polysaccharides (Sharma et al. 2019; Kumawat et al. 2019; Kumar et al. 2018; Mittal et al. 2014; Saha and Ghosh 2012; Gupta and Sharma 2011; Patel and Mishra 2011). *T. cordifolia* is one of the most frequently used ayurvedic medicines to treat a diverse range of health disorders (Dhama et al. 2016; Thomas

et al. 2016; Srivastava 2011; Ghosh et al. 2011) and is designated as miracle herb of the 21st century according by the Ministry of Ayurvedic, Yoga, Naturopathy, Unani, Siddha and Homeopathy (AYUSH), Government of India. In spite of the immense medicinal properties and application of *T. cordifolia*, little information is available about its breeding system, genetic variability and population genetic structure. The knowledge of genetic variability of any plant is considered as prerequisite for its efficient exploitation and utilization, to develop suitable conservation strategies and deploy suitable breeding technique for its genetic improvement (Fu 2015). Various molecular markers have been utilized for understanding the domestication, evolution, distinctiveness, population structure and extent of genetic variation within and between species (Bakatoushi and Ahmed 2018). Molecular markers such as RAPD and ISSR (Malik et al. 2019; Nazneen et al. 2019, Lade et al. 2018; Rana et al. 2012; Shinde and Dhalwal 2010) ITS and cpDNA markers (Ahmed et al. 2006), isozyme (Kalpesh and Mohan 2009), and also gene targeted SCoT markers (Paliwal et al. 2013) have been utilized to unravel the genetic diversity and distinctiveness among limited accessions of *T. cordifolia* considering only small geographical area. Albeit, there is information available, where SSR markers have been used to estimate the genetic diversity in small number of accessions collected mostly from local or small geographical areas, which do not reflect a comprehensive account of the actual genetic diversity (Gargi et al. 2017; Paliwal et al. 2016; Singh et al. 2014, 2016). Therefore, it is important to estimate the genetic potential in diverse and large number of accessions from a wider range of phyto-climatic zones. Among various marker systems, SSRs is considered as highly potential to estimate the genetic variability due to wide coverage of the genome, reproducibility, high level of polymorphism and co-dominant inheritance (Kalia et al. 2010). Considering the medicinal significance and restricted marker-based studies in *T. cordifolia*, more SSR markers needs to be developed, which could substantially contribute towards estimation of genetic diversity more comprehensively and identify the diverse populations for future prospection. The present study was, therefore, undertaken to enrich the SSR resources of *T. cordifolia* by extracting novel SSRs from transcriptome sequences developed at CSIR-National Botanical Research Institute (CSIR-NBRI), Lucknow (India), and utilize a subset of newly developed SSRs to analyse the genetic diversity and population genetic structure in 96 accessions of *T. cordifolia*.

Materials and methods

Plant materials

The plant materials used in the present investigation were collected from various phyto-geographical zones of India.

The stem cuttings of almost equal length and width were planted in the experimental field at CSIR-NBRI, Lucknow and support was provided with the help of iron pipes (Fig. S1). A total of 96 accessions were selected for the present study collected from ten different states namely Uttar Pradesh, Madhya Pradesh, Jammu & Kashmir, Maharashtra, Kerala, Gujarat, Andhra Pradesh, Karnataka, Chhattisgarh and Haryana (Table S1). The total genomic DNA was isolated from the young leaves using CTAB (Cetyl trimethylammonium bromide) method following Porebski et al. (1997). The DNA quality was checked by agarose gel electrophoresis and quantified by Nanodrop (Thermo Nano drop ND 1000).

RNA isolation and sequencing

Approximately 100 mg of fresh young leaves was grounded into a fine powder with liquid nitrogen. Total RNA was extracted and cleaned using Spectrum™ Plant Total RNA Kit (SIGMA-ALDRICH) and RNase-free DNase I, respectively, for transcriptome sequencing. The RNA concentration and purity were determined using Nanodrop (Thermo Nano drop ND 1000) and agarose gel electrophoresis (1.2%), respectively. High-quality RNA samples with concentration of ≥ 600 ng/ μ L; RIN value ≥ 6.8 ; OD A260/280 ≥ 2.10 was used to construct complementary DNA (cDNA) libraries, which were then sequenced using the Illumina HiSeq™ 2000 platform (SciGenom Labs Pvt Ltd).

SSR identification and primer designing

The assembled transcriptome sequence data was subjected to SSR identification using MicoSATellite tool (MISA; <https://webblast.ipk-gatersleben.de/misa/>) (Thiel et al. 2003). The major criteria for SSR identification was set as: minimum number of repeat units: ten for di-nucleotides, five for tri- and tetra-nucleotides, and four for penta- and hexa-nucleotides. The primer pairs were designed using Primer 3 v2.23 (<http://primer3.sourceforge.net>) with the following parameters: primer length of 18–27 bp, GC content of 40–60%, annealing temperature of 55–60 °C, and target PCR product size of 100–300 bp.

PCR amplification and data acquisition

A random set of 135 SSR primers was commercially synthesized (Eurofins Genomics India Pvt. Ltd.) for genotyping. First, all the 135 SSR primers were PCR amplified with 4 accessions of *T. cordifolia* to identify polymorphic SSRs. Furthermore, the polymorphic SSRs were used to amplify the 96 accessions. The PCR reactions was carried out in 10 μ L reaction volume containing 20 ng of template DNA, 1X PCR buffer, 2.0 mM MgCl₂, 0.25 mM each dNTPs, 1U Taq DNA

polymerase (Qiagen), 1.0 pmol forward and reverse primers. The PCR amplification was performed using C1000 Touch™ Thermal Cycler (BIO-RAD, USA) with an initial denaturation of 5.0 min at 95 °C followed by 36 cycles of denaturation for 30 s at 95 °C, annealing for 45 s at 52–55 °C (primer specific), extension for 1.0 min at 72 °C and final extension for 10.0 min at 72 °C. The PCR products were first confirmed on 2.5% agarose gel and subsequently, separated on 6% polyacrylamide gel electrophoresis and visualized by silver staining. The size of PCR products were measured against the 100 bp DNA ladder (Promega, USA) and amplicon size was calculated using online tool fragment size calculator (biotoools.nubic.northwestern.edu/SizeCalc.html).

SSR data analysis

All the SSR fragments were manually scored as allele size and also transformed into binary matrix data, i.e., 0 for the absence and 1 for the presence of the band. The statistical analysis of SSR allelic data was performed using Power Marker (Liu and Muse 2005) to measure observed heterozygosity (H_o), allele frequency, gene diversity or expected heterozygosity (H_e) and polymorphic information content (PIC) value. To identify the genetic relationship among the accessions, pairwise genetic dissimilarities across all the 96 accessions was calculated using Jaccard's coefficient with using DARwin 5.0.128 software (Perrier et al. 2003). The calculated dissimilarity matrix was then used to construct a neighbor joining tree with a bootstrap value of 1000 replicate. Furthermore, the Bayesian clustering was also performed using STRUCTURE v 2.3.3 software to determine the number of subpopulations. The presumed population (K) was set to 2 to 15 with five separate runs per K and 30,000 burn-in period and 100,000 iterations for each run. The optimal K value was calculated using structure harvester (Earl and Von Holdt 2012) and analysing by ΔK statistic and $L(K)$ according to Evanno et al. (2005). The molecular variance (AMOVA) and confirmation of the Hardy–Weinberg equilibrium (HWE) was performed using GenAlEx 6.502 (Excoffier et al. 2009). In addition, the relative discriminatory power of genetic loci/combinations of loci for better population assignment was determined using WHICHLOCI software (Banks et al. 2003). The analysis was done with criteria of Use Critical Population Method at LOD score 3.0 and the threshold of minimum percentage of assigned to population was set to 95.0.

Results

Frequency and distribution of SSRs

The Illumina Hiseq 2500 sequencing generated > 246.0 million reads which were assembled into 268,149 transcripts. In total, 155,821 unique contigs were obtained and

a total of 7611 SSR loci identified were distributed in 7457 sequences. Among the identified SSR, 154 were found to be in compound form, while others were of ideal one-repeat type (Table 1). Out of 7611 SSRs, the most abundant were tri-nucleotide repeats (69.1%), followed by di-nucleotide (12.8%), tetra-nucleotide repeats (8.2%), penta-nucleotide (6.6%) and hexa-nucleotide repeats (1.4%) (Fig. S2). The most common motifs were AAG/CTT repeats (10.7%) followed by AAT/ATT (8.2%) for tri-nucleotides and AG/CT (3.7%) repeats for di-nucleotides (Fig. S3).

Validation of SSRs and genotyping

Out of 7611, a total of 2981 primer pairs were successfully designed flanking the SSRs motifs. However, the flanking sequences of rest of SSRs were not found suitable for primer designing. A random set of 135 primers were selected and commercially synthesized (Eurofins Genomics India Pvt. Ltd.) for validation (Supplementary File 1). The initial screening of SSR primers with 4 DNA was done to optimize the PCR conditions. After several round of PCR optimization, finally a set of 56 (41.48%) primer pairs were found to be good which generated amplification products with expected size, 46 (34.07%) showed non-specific products and 33 (24.44%) did not amplified any fragment (Supplementary File 1). The 56 SSR primers were then amplified with 96 accessions and amplicon resolved on 6% polyacrylamide gel electrophoresis (PAGE) and visualized through silver staining (Fig. S4). Out of 56 SSRs, 30 were found to be polymorphic among 96 accessions, while the rest 26 were monomorphic. These 30 polymorphic SSRs generated a total of 230 alleles ranged from 2 to 12 alleles with an average of 7.67 ± 2.88 alleles per locus (Table 2). The maximum number of alleles (12) was exhibited by TCTSSR-37 and TCTSSR-59 followed by TCTSSR-123, TCTSSR-104, TCTSSR-60, TCTSSR-92, TCTSSR-18 (11 alleles) and TCTSSR-06 (10 alleles), respectively. Out of 30, 22 SSRs (73%) had alleles range between 6.00 and 12.00, while 8 SSRs (27%) between 2.0 and 5.0. The PIC value was found to be comparatively higher as 80% SSRs showed PIC value in range of 0.61 to 0.85 (Table 2, Fig. S5). The

Table 1 Summary of SSRs mined from transcriptome data of *Tinospora cordifolia*

Parameters used in screening	Numbers
Total number of sequences examined	268149
Total size of examined sequences (bp)	225518219
Total number of identified SSRs	7611
Number of SSR containing sequences	7457
Number of sequences containing more than 1 SSR	521
Number of SSRs present in compound formation	154

observed heterozygosity value ranges from 0.00 to 0.78 with an average of 0.12 ± 0.22 . The gene diversity was in between 0.40 and 0.85 with an average of 0.71 ± 0.13 . Out of 30, 18 SSRs (60%) revealed highly informative gene diversity value range between 0.71 and 0.85, while 8 SSRs (27%) revealed moderate values ranging between 0.61 and 0.70 and 4 SSRs (13%) had lower gene diversity values < 0.60 . The major allele frequency ranged between 0.21 (TCTSSR-92) to 0.76 (TCTSSR-124) with an average of 0.41 ± 0.15 . The SSRs 40% (12) showed major allele frequency between 0.41 and 0.80, while 30% (9 SSRs) ranged between 0.31 and 0.40. All the SSRs showed significant Chi square test for Hardy–Weinberg equilibrium test at $P < 0.001$.

Genetic diversity and population structure

The allelic data was used to calculate a pairwise genetic distance based on Jaccard's coefficient to understand the genetic relationship among the 96 accessions. The pairwise genetic dissimilarity varied from 0.26 to 1.0 with an average of 0.89 ± 0.05 . The maximum genetic dissimilarity of 0.99 was observed among Tc131-Tc31, Tc129-Tc38, Tc16-Tc59, Tc16-Tc60, Tc17-Tc106, Tc4-Tc130 and 0.98 among Tc23-Tc55, Tc16-Tc63, Tc16-Tc90, Tc10-Tc128, Tc15-Tc128, Tc9-Tc112, Tc9-Tc131, Tc7-Tc129 and Tc5-Tc129. The minimum genetic dissimilarity (0.26) was observed between Tc95 and Tc96 followed by Tc53 and Tc52 (0.27), Tc28 and Tc29 (0.28). The neighbor joining (NJ) dendrogram grouped all the 96 accessions into three main clusters: cluster I, cluster II and cluster III (Fig. 1a). The cluster III comprised of 47 accessions (49%), followed by cluster II (34 accessions; 35%) and cluster I (15 accessions; 16%). The cluster I could be subdivided into two sub-clusters namely Ia with 7 accessions and Ib with 8 accessions. Similarly, cluster II was separated into four sub-clusters as IIa, IIb, IIc, II d with 5, 11, 10 and 8 accessions, respectively. Likewise, the cluster III was separated into six sub-clusters IIIa, IIIb, IIIc, III d, IIIe and III f, each consisting of 10, 6, 8, 8, 8 and 7 accessions, respectively. Furthermore, the model-based simulation was performed to investigate the subpopulations among the 96 accessions. The population structure analysis revealed 3 subpopulations ($K = 3$) based on maximum ΔK value obtained through the structure harvester (Fig. S6). The bar plot of model-based clustering revealed that the subpopulation 1 confined 15 accessions (15.62%), among which 10 accessions (66.66%) were found to be pure lines and 5 admixture type (Fig. 1b). The subpopulation 2 comprised 41 accessions (42.70%) with 24 (58.53%) accessions as pure lines, 17 admixture and subpopulation 3 had 40 accessions (41.66%) in which 33 (82.5%) accessions as pure lines, 7 admixtures. The NJ dendrogram and STRUCTURE analysis revealed almost similar patterns of accession compositions. The NJ tree as well as the STRUCTURE plot showed that

the accessions composition of subpopulation 1 was coincided by both the methods. The accessions of subpopulation 2 also corresponded with cluster II, with the exception of seven accessions that grouped in cluster III.

Furthermore, to understand the population differentiation, AMOVA analysis was carried out which revealed that 78% genetic variation was observed within the population and 22% was among the population (Table 3). The average pairwise Φ_{PT} (Similar To f_{st}) was 0.22 which eventually supported the structuring in the populations. The maximum genetic diversity (H_e) was recorded for population 3 (0.12 ± 0.009), followed by population 2 (0.10 ± 0.010) and population 1 (0.11 ± 0.010) with an average of 0.11 ± 0.005 (Table 4). Similarly, population 3 revealed maximum effective number of alleles per locus (N_e), number of alleles per locus (N_a) and Shannon's information index (J). The mean value for effective number of alleles, Nei genetic diversity and Shannon's information index were 0.11 ± 0.005 , 1.17 ± 0.010 and 0.19 ± 0.008 , respectively. The locus score derived from WHICHLOCI analysis showed that the TCTSSR-126, had highest assignment score (2.631), whereas TCTSSR-79 displayed the lowest score (0.701) (Fig. S7, a; Table 5). The "overall" panel without defining the critical population included four SSR markers: TCTSSR-131, TCTSSR-126, TCTSSR-1230 and TCTSSR-85. In this panel, the locus scores varied from 2.27 (TCTSSR-131) to 2.63 for TCTSSR-126, with an average value of 2.44 (Table 5). About 96.87% of the simulated accessions were accurately assigned to their original population using the most effective markers (Table 5). Similarly, in critical population method for population 1, TCTSSR-126 had the highest Locus score (1.92) and TCTSSR-79 the lowest (0.394) (Fig. S7, b). A panel of three marker loci, TCTSSR-126, TCTSSR-06 and TCTSSR-131 had 97.91% correct assignment of accessions to population 1. For Population 2, the Locus score varied from 0.55 for TCTSSR-79 to 1.67 for TCTSSR-126 (Fig. S7, c). The population 2 panel had 96.87% assignment accuracy of accessions and included loci TCTSSR-126 (1.67), TCTSSR-123 (1.60), TCTSSR-85 (1.53) and TCTSSR-46 (1.47). For population 3, the locus score varied from 0.45 for TCTSSR-79 to 2.0 for TCTSSR-85 (Fig. S7, d). The four markers, TCTSSR-85 (2.0), TCTSSR-84 (1.79), TCTSSR-49 (1.79) and TCTSSR-96 (1.67), accounted for 97.91% assignment accuracy of accessions.

Discussion

The knowledge about the level of genetic diversity provides basic parameters to any plant for its sustainable exploitation and utilization. The genetic diversity could be evaluated based on morphological traits and DNA markers. The

Table 2 Amplification pattern of polymorphism of 30 polymorphic SSR used for diversity analysis in *T. cordifolia*

SSR_ID	Sequence (5'–3') Forward Primer	Sequence (5'–3') Reverse Primer	SSR motifs	MAF	Na	He	Ho	PIC	HWE	χ^2	p value
TCTSSR-37	GAATGGTTACGGTGGATTG	TGACCTGTTAAAAGGGGCAAGT	(AC)10	0.30	12.00	0.79	0.09	0.77	787.31	0.000	***
TCTSSR-59	TGCTGTGTGTGAGAACTCGTG	GGGATGGGAATGGGAGTAGT	(ACT)5	0.26	12.00	0.85	0.00	0.84	1034.00	0.000	***
TCTSSR-70	AAGCTCCAGCAGCAACAAT	TCAGCACGTGTTAAGGACCC	(AGT)5	0.26	7.00	0.81	0.00	0.80	576.00	0.000	***
TCTSSR-75	ATGAAGGAGCCTTGCATTTG	ATGGATTTGCTGTGCTTCC	(AGC)5	0.26	8.00	0.81	0.00	0.79	672.00	0.000	***
TCTSSR-78	TGCATAAACCAATCTGTGTGTG	CCAGTCAATCCCGTAACCTCA	(AAT)5	0.42	5.00	0.70	0.00	0.66	384.00	0.000	***
TCTSSR-79	CTCCAGTTAACCGATCCAGC	ATCTCCGATCTCGATTCTCT	(AGA)5	0.63	4.00	0.54	0.00	0.49	288.00	0.000	***
TCTSSR-93	GGGCAAGTTCTGATTAATGT	CCTTCGATTCAGGGATTCA	(TCT)5	0.39	5.00	0.74	0.00	0.71	376.00	0.000	***
TCTSSR-104	GCCGATGCTTCGTAGATGT	CTCGAAACCCCAATTCTC	(AG)10	0.36	11.00	0.80	0.11	0.79	604.10	0.000	***
TCTSSR-124	AATTCAGGAATAGTGGGGG	GTAAGACCCCGAACCCGTACA	(TGG)5	0.76	4.00	0.40	0.00	0.39	282.00	0.000	***
TCTSSR-131	TCTTGTTCCTACTCCACCCG	AACTCAGCTTGAAGGGCAAA	(ATC)6	0.38	9.00	0.72	0.14	0.69	578.50	0.000	***
TCTSSR-123	CCATTTTGTCTTGTGATTGGAA	AAAGTTGAACCAAAATCCTGCC	(AAT)7	0.49	11.00	0.67	0.22	0.64	400.65	0.000	***
TCTSSR-60	ACGGAAATCTGCTTCTTGT	TTGCTGTGGCAATTTAGGTG	(AAC)5	0.51	11.00	0.69	0.14	0.68	503.31	0.000	***
TCTSSR-6	GAACGAAGCAACAGCAAAAT	GGGGTGAATGAGAGAAGATTG	(AG)10	0.46	10.00	0.70	0.04	0.67	663.45	0.000	***
TCTSSR-84	ATGATGCATGCAATGGAGA	GTATGGGCATGCTGTGAG	(AGT)5	0.39	8.00	0.71	0.07	0.67	374.01	0.000	***
TCTSSR-94	GTAGTCCGACCCCATCTCAT	ATGCCACGATTAACCTAGGCC	(TGG)5	0.38	7.00	0.71	0.00	0.67	576.00	0.000	***
TCTSSR-40	CCATCAGGCCATTTTCAATC	GCCTTCTAGAAAGCCCAACT	(AC)10	0.35	9.00	0.80	0.45	0.78	302.86	0.000	***
TCTSSR-46	CTTCGATCGTCTCTCTCGCT	TGAAGGGCTCAAGGTTATGG	(CT)10	0.31	5.00	0.73	0.00	0.70	360.00	0.000	***
TCTSSR-135	AGATCATTTGGTGGAAACCA	TCATCATGGCAATTTTCCAA	(TTC)5	0.59	4.00	0.54	0.00	0.47	279.00	0.000	***
TCTSSR-133	GCCCAAGTCCAGAAATCAGTC	GAGGAGCCCAACTGTGTAA	(ATA)5	0.48	6.00	0.63	0.08	0.58	280.78	0.000	***
TCTSSR-126	CCTTGAAAATGGAGCTGGAA	GAGGTTTGGGTTTTGGGTT	(CTT)5	0.30	8.00	0.82	0.00	0.82	651.00	0.000	***
TCTSSR-82	GGTGCAGATTGCAATTGTAG	CAATGACCACATTTCTTGCTT	(ATT)5	0.55	5.00	0.64	0.00	0.61	380.00	0.000	***
TCTSSR-96	TTTCACTCCCAATCACACCA	GCTCTTCGAAAGGTGAAAAACG	(AGA)5	0.30	7.00	0.80	0.01	0.78	470.17	0.000	***
TCTSSR-49	TATCACGAAGCGTAGGGACC	GACCACTATTGCAATTTCCGA	(TGC)5	0.46	6.00	0.69	0.25	0.65	211.50	0.000	***
TCTSSR-67	TCCCTACCCCTTTTCTTCAG	CCACTTGGTCACCTGGTTTT	(AAG)5	0.72	2.00	0.40	0.00	0.32	96.00	0.000	***
TCTSSR-61	TGAAATCATCAAGGTGGCAA	TTCTTTACGGCTCCAACGAC	(AGA)5	0.33	9.00	0.79	0.78	0.77	240.55	0.000	***
TCTSSR-85	GCTTCCCAACAATAACTTCGC	CTTGCCACCTACTCGTGTGA	(AAC)5	0.27	9.00	0.78	0.05	0.76	475.72	0.000	***
TCTSSR-69	ATTGCAGATGGCAATGACAA	AACTGACCCTGACCCCAATTTG	(AAT)5	0.38	7.00	0.75	0.17	0.73	363.79	0.000	***
TCTSSR-92	TCAGTGAAGGGGGAAGAACA	TGCTTACCCCAACCAAAATTC	(AAG)5	0.21	11.00	0.85	0.65	0.84	264.70	0.000	***
TCTSSR-18	CAAAGACCAGTCAATTTCTTTG	GCGATTAAGGTGACAAATACC	(AG)10	0.26	11.00	0.82	0.13	0.80	742.41	0.000	***
TCTSSR-90	CTTGGCAGCCTAATTCAAG	TTTGTGCTCAAAAGCTTGC	(ACA)5	0.49	7.00	0.64	0.11	0.59	268.88	0.000	***
Mean ± SD				0.41 ± 0.15	7.67 ± 2.88	0.71 ± 0.13	0.12 ± 0.22	0.68 ± 0.14			

MAF major allele frequency, Na number of alleles, He expected heterozygosity or gene diversity, Ho observed heterozygosity, PIC polymorphic information content, HWE: deviation from Hardy–Weinberg equilibrium with significant $p < 0.001$; *** $p < 0.001$

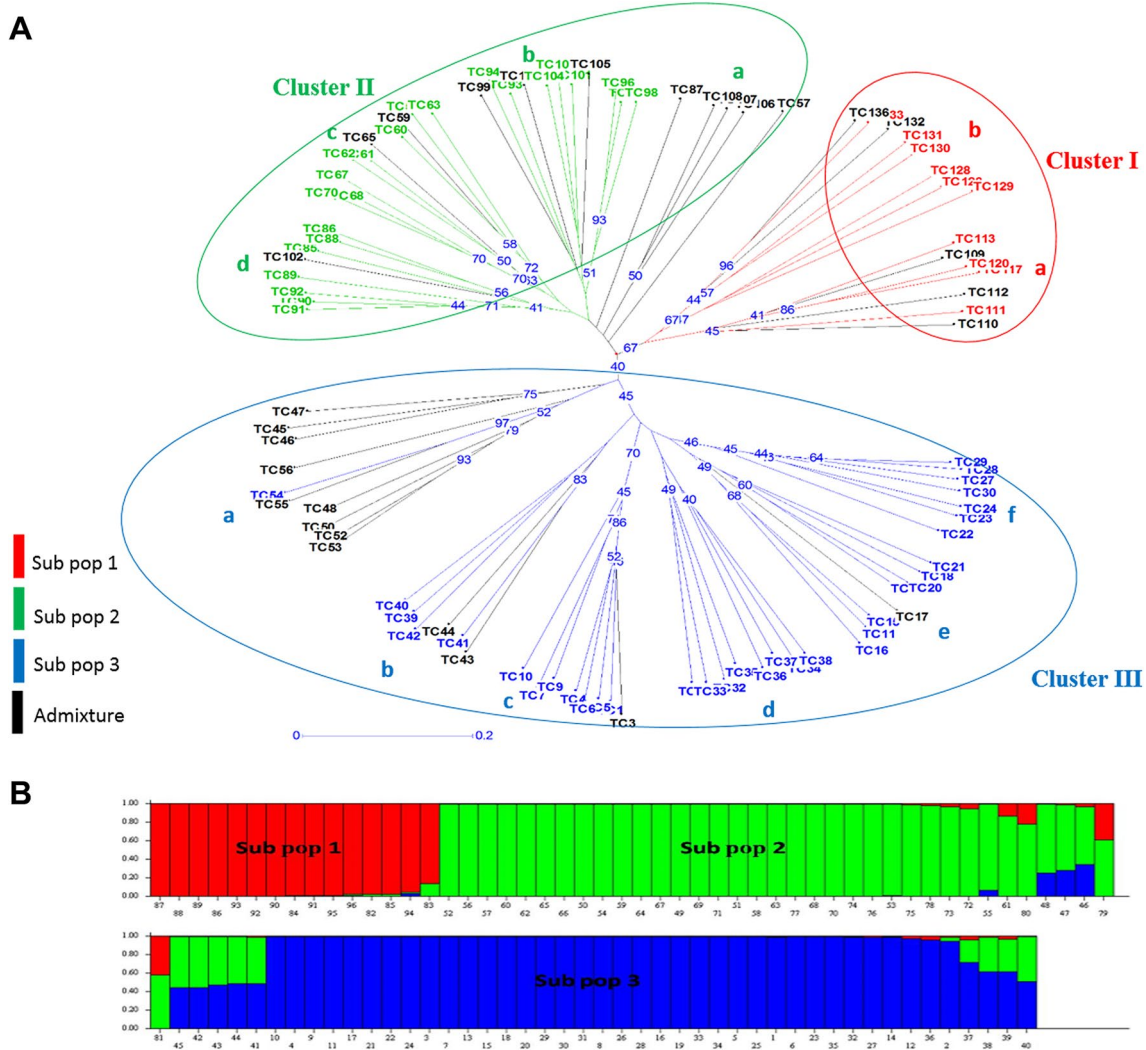


Fig. 1 Distance- and Bayesian Model-based genetic clustering of 96 accessions of *Tinospora cordifolia* using 30 SSR markers. **a** Dendrogram was constructed in DARwin v 5.0.128 based on Jaccard's coefficient using NJ method. Robustness of the node of the phylogenetic tree was assessed from 1000 bootstrap replicates and bootstrap values

of >40% are shown (each color is indicative of a separate group and color coding is according to STRUCTURE analysis data). **b** Population structure analysis of *Tinospora cordifolia* accessions for K=3 subpopulations (pop1–pop3) using STRUCTURE v 2.3.3 application

Table 3 Analysis of molecular variance for 96 accessions from three populations

Source	Degree of freedom	Sum of squares	Variance	% variance
Among pops	2	456.292	6.944	22
Within pops	93	2293.833	24.665	78
Total	95	2750.125	31.609	100

$F_{ST} = 0.220, P < 0.010$

Table 4 Different genetic diversity estimates for three populations of *T. cordifolia* based on SSR data

Population	Na	Ne	I	He
pop 1 (15)	1.05 ± 0.065	1.17 ± 0.017	0.18 ± 0.014	0.11 ± 0.010
pop2 (34)	1.12 ± 0.065	1.16 ± 0.018	0.17 ± 0.014	0.10 ± 0.010
pop3 (47)	1.49 ± 0.057	1.17 ± 0.016	0.21 ± 0.013	0.12 ± 0.009
Mean	1.22 ± 0.037	1.17 ± 0.010	0.19 ± 0.008	0.11 ± 0.005

Na number of alleles, Ne effective number of alleles, I Shannon's information index, He expected heterozygosity

Table 5 Estimation of correct assignments and incorrect assignments of loci to one population at a time using critical population method by WHICHLOCI ver 1.0 Software for three populations of *Tinospora cordifolia*

Marker	Panels (including markers) with the corresponding locus score			
	Overall (whole population)	Population 1	Population 2	Population 3
TCTSSR-06		1.68		
TCTSSR-46			1.47	
TCTSSR-49				1.79
TCTSSR-84				1.79
TCTSSR-85	2.53		1.53	2.00
TCTSSR-96				1.67
TCTSSR123	2.36		1.60	
TCTSSR-126	2.63	1.92	1.67	
TCTSSR-131	2.27	1.63		
Mean of Locus score	2.44	1.74	1.56	1.81
Correct assignment (%)	96.87	97.91	96.87	97.91
Not correct assignment (%)	3.12	2.08	3.12	2.08

No critical population method used for “Overall” panel

DNA markers provide more precise informations on as they are rarely affected by environmental variables (Sharma et al. 2008). Various RAPD based studies in *T. cordifolia* have been reported; however, the results of such studies could not be so reliable due to low reproducibility and low level of polymorphism of RAPD marker (Li et al. 2011; Hamza et al. 2013). Contrary, the SSR markers revealed higher heterozygosity and thus can be attributed to detection of high polymorphism which makes them the most preferred DNA marker (Rana et al. 2015; Sharma et al. 2015). There are, however, few reports available in the public domain on the development and characterization of SSR markers in *T. cordifolia*, wherein limited number of accessions were analyzed (Gargi et al. 2017; Paliwal et al. 2016; Singh et al. 2016). Gargi et al. (2017) extracted SSRs from enriched libraries and utilized 23 SSRs to assess the genetic diversity among 30 accessions collected from very narrow geographical region of India. While, Singh et al. (2016) identified transcriptome derived SSRs and deployed over 24 accessions to understand the level of genetic diversity. All these studies carried out with limited number of accessions mostly collected from narrow geographical region and thus majority of the genetic variation remains unexplored. Therefore, in the present investigation large set of accessions from wider geographical regions was estimated for genetic variability assessment. In addition, SSR marker resource in *T. cordifolia* also enriched with new SSR identified from RNA-seq data. A total of 246 million reads and 268149 assembled transcripts were generated which were greater than the results of previous studies (Singh et al. 2016). This indicates that the transcriptome data was efficiently assembled and appropriate for marker development. The present study revealed that the tri-nucleotide repeats were found to be the most abundant (69.1%) to di-nucleotide (12.8%) and

tetra-nucleotide repeats (8.2%). This was also in agreement with the earlier reports, wherein tri-nucleotide found to be most abundant (57.88%) than the di-nucleotide (36.87%) (Singh et al. 2016). The dominance of tri-nucleotide SSR repeats was also observed in other plants such as *Curcuma alismatifolia* (Taheri et al. 2019), *Elymus* L. (Zhang et al. 2019), *Panicum miliaceum* L. (Hou et al. 2017), *Capsicum annum* (Lu et al. 2012); however, di-nucleotide SSR was found to be dominant in *Notopterygium incisum* (Jia et al. 2019), *Brassica rapa* L. spp. *pekinensis* (Ding et al. 2015). This discrepancy might be due to search criteria, algorithm used and selection constraints of plant species (Shi et al. 2017). The preponderance of tri-nucleotide repeat motifs considered to be usual for SSRs, as insertions or deletions inside translated regions do not interact with open reading frames, whereas frame-shift mutations restrict the expansion of other motif forms (Wang et al. 2018). The AAG/CTT repeats was the most frequent tri-nucleotide repeats (10.7%) in the present data set which was also in corroboration with previous findings, where the most prevalent SSR repeats in *T. cordifolia* was AAG/CTT, (25.60%) (Singh et al. 2016). Among the di-nucleotide, AG/CT was the most frequently observed (3.7%) repeats, which was also in congruence with the earlier studies (Singh et al. 2016; Zheng et al. 2013).

The allele range and average number of alleles in the present investigation was found to be comparatively higher than the previous reports in *T. cordifolia* (Gargi et al. 2017; Paliwal et al. 2016; Singh et al. 2016) which indicated that the accessions used have higher level of genetic variability. The present study also revealed significantly higher average PIC value (0.68) and five SSRs out of 30 showed > 0.8 PIC value indicated high polymorphism. Furthermore, the HWE analysis indicated that the SSRs was randomly distributed in the genome and found suitable to be used

for genetic diversity assessment. Based on allele and PIC value the SSRs, TCTSSR-59, TCTSSR-92, TCTSSR-126, TCSSR-18 and TCSSR-37 was identified as potential marker which could be utilized for genetic studies on other set of accessions of *T. cordifolia*. Likewise, the accessions Tc131, Tc31, Tc129, Tc38, Tc16, Tc59, Tc16, Tc60, Tc17, Tc106 and Tc130 was identified as diverse and potential accessions. These diverse accessions could be further utilized in bio-prospection programme for new biomolecules. The clustering pattern of accessions revealed that the accessions are independent of their geographical affiliations; however, majority of the accessions from Jammu & Kashmir clustered together in cluster II. The Bayesian Model-based structuring and NJ dendrogram revealed similar patterns of population structuring amongst the *T. cordifolia* populations which suggests that any one of these two methods are efficient to differentiate the accessions based on their genetic variations. Furthermore, the importance of genetic structure, represented by the AMOVA analysis indicated that 78% of the genetic variation resided within the populations and 22% among the three populations. The F_{ST} value (0.22) derived by AMOVA analysis supported the strong population stratification among the accessions. The discriminatory power of markers is determined by their efficiency in correct population assignments and inversely correlated with their propensity for causing false assignments (Agrama et al. 2012; Banks et al. 2003). The locus score of a marker is positively associated with the corresponding genetic diversity, PIC, and number of alleles. In the “overall” panel analysis, four markers with high locus scores presented more proximity towards high PICs, genetic diversity and high allele numbers. The SSR markers TCTSSR-79 and TCTSSR-124 had the lowest locus scores, because their polymorphism and diversity were the more towards lowest (PIC = 0.49 and 0.39, diversity = 0.54 and 0.40, allele number = 4, respectively) among the 30 markers. Among the nine SSR markers selected in the four panels, four highly diversified markers (TCTSSR-131, TCTSSR-126, TCTSSR-123 and TCTSSR-85) were common to more than one panel.

Conclusions

The present investigation revealed that the considerable level of genetic variability exists across the *T. cordifolia* accessions. The study has resulted into the identification of several genetically diverse accessions (Tc131, Tc31, Tc129, Tc38, Tc16, Tc59, Tc16, Tc60, Tc17, Tc106 and Tc130), which could be further utilized in bio-prospecting of *T. cordifolia* for its valuable traits. The highly informative SSRs such as TCTSSR-59, TCTSSR-92, TCTSSR-126, TCSSR-18, TCTSSR-131, TCTSSR-123 and TCSSR-37 could be utilized for sorting out the large set of germplasm.

Acknowledgements Authors thank Director, CSIR-NBRI, Lucknow facilities and encouragement. This work was financially supported by the CSIR through HCP 010 and BSC 107 projects.

Author contributions SL performed the experiments and prepared the manuscript, HKY and SL analyzed the data. TSR, HKY and VP designed, coordinated the experiments and improved the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest in the publication.

References

- Agrama HA, McClung AM, Yan W (2012) Using minimum DNA marker loci for accurate population classification in rice (*Oryza sativa* L.). *Mol Breed* 29:413–425
- Ahmed SM, Verma V, Qazi PH, Ganaie MM, Bakshi SK, Qazi GN (2006) Molecular phylogeny in Indian *Tinospora* species by DNA based molecular markers. *Plant Syst Evol* 256:75–87
- Bakatoushi REL, Ahmed DGA (2018) Evaluation of genetic diversity in wild populations of *Peganum harmala* L., a medicinal plant. *J Genet Eng Biotechnol* 16:143–151
- Banks MA, Eichert W, Olsen JB (2003) Which genetic loci have greater population assignment power? *Bioinformatics* 19:1436–1438
- Dhama K, Sachan S, Khandia R, Munjal A, Iqbal Hafiz MN, Latheef SK, Karthik K, Samad HA, Tiwari R, Dadar M (2016) Medicinal and beneficial health applications of *Tinospora cordifolia* (Guduchi): a miraculous herb countering various diseases/disorders and its immunomodulatory effects. *Bentham Sci* 10(2):96–111
- Ding Q, Li J, Wang F, Zhang Y, Li H, Zhang J, Gao J (2015) Characterization and development of EST-SSRs by deep transcriptome sequencing in Chinese Cabbage (*Brassica rapa* L. ssp. *Pekinensis*). *Int J Genomics* 2015:473028
- Earl DA, Von Holdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4:359–366
- 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
- Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity* 103:285–298
- Fu YB (2015) Understanding crop genetic diversity under modern plant breeding. *Theor Appl Genet* 128:2131–2142
- Gargi M, Thakur S, Anand SS, Choudhary S, Bhardwaj P (2017) Development and characterization of genomic microsatellite markers in *Tinospora cordifolia*. *J Genet* 8(96):e25–e30
- Ghosh AK, Martyr CD, Steffey M, Wang YF, Agniswamy J, Amano M et al (2011) Design of substituted bis-tetrahydrofuran (bis-THF)-derived potent HIV-1 protease inhibitors, protein-ligand X-ray structure, and convenient syntheses of bis-THF and substituted bis-THF ligands. *ACS Med Chem Lett* 2:298–302
- Gupta R, Sharma V (2011) Ameliorative effects of *Tinospora cordifolia* root extract on histopathological and biochemical changes induced by aflatoxin-b (1) in mice kidney. *Toxicol Int* 18:94–98
- Hamza H, Abederrahim MAB, Elbekkay M, Ferchichi A (2013) Comparison of the effectiveness of ISSR and SSR markers in determination of date palm (*Phoenix dactylifera* L.) agronomic traits. *Aust J Crop Sci* 7(6):763–769

- Hou S, Sun Z, Li Y, Wang Y, Ling H, Xing G, Han Y, Li H (2017) Transcriptomic analysis, genic SSR development and genetic diversity of Proso Millet (*Panicum miliaceum*; Poaceae). *Appl Plant Sci* 5(7):1600137
- Jia Y, Bai JQ, Liu ML, Jiang ZF, Wu Y, Fang MF, Li ZH (2019) Transcriptome analysis of the endangered *Notopterygium incisum*: cold-tolerance gene discovery and identification of EST-SSR and SNP markers. *Plant Divers* 41(1):1–6
- Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK (2010) Microsatellite markers: an overview of the recent progress in plants. *Euphytica* 177:309–334
- Kalpesh I, Mohan JSS (2009) Assessment of genetic diversity in the medicinal climber *Tinospora cordifolia* (Willd.) Miers (Menispermaceae) from Gujarat India. *Afr J Biotechnol* 8:6499–6505
- Kumar V, Singh S, Singh A, Dixit AK, Srivastava B, Sidhu GK, Singh R, Meena AK, Singh RP, Subhose V, Prakash O (2018) Phytochemical, antioxidant, antimicrobial, and protein binding qualities of hydro-ethanolic extract of *Tinospora cordifolia*. *JBAP* 8(3):192–200
- Kumawat NK, Dushyant K, Kumar VJ, Padmapriya B, Manoj NT (2019) *Tinospora cordifolia*: a wonderful miracle herb of 21st century of India. *Medicinal plants. Int J Phytomed Relat Ind* 11(2):117–122
- Lade S, Sikarwar PS, Ansari A, Khatoon S, Kumar N, Yadav HK, Ranade SA (2018) Diversity in a widely distributed dioecious medicinal plant, *Tinospora cordifolia* (Willd.) Miers. ex. Hook F. and Thomas. *Current Sci* 114(7):1520–1526
- Li L, Wanapu C, Huang X, Huang T, Li Q, Peng Y, Huang G (2011) Comparison of AFLP and SSR for genetic diversity analysis of *Brassica napus* hybrids. *J Agric Sci* 3(3):101–110
- Liu K, Muse SV (2005) Power Marker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21:2128–2129
- Lu FH, Cho MC, Park YJ (2012) Transcriptome profiling and molecular marker discovery in red pepper, *Capsicum annuum* L. TF68. *Mol Biol Rep* 39:3327–3335
- Malik A, Arya A, Kaushik V, Sindhu A (2019) Studies on genetic variability of *Tinospora cordifolia* collected from different agroclimatic zones of Haryana using RAPD markers. *Res Crop* 20(2):407–412
- Mittal J, Sharma MM, Batra A (2014) *Tinospora cordifolia*: a multipurpose medicinal plant—a review. *J Med Plants* 2(2):32–47
- Nazneen HF, Naik BA, Ramesh P, Sekhar AC, Shankar PC (2019) Genetic diversity using random amplified polymorphic DNA (RAPD) and inter-simple sequence repeats (ISSR) markers in *Tinospora cordifolia* from the Rayalseema region in Andhra Pradesh. *Afr J Biotechnol* 18(10):231–241
- Paliwal R, Singh R, Singh AK, Kumar S, Kumar A, Majumdar RS (2013) Molecular characterization of Giloe (*Tinospora cordifolia* (Willd.) Miers ex Hook F & Thoms.) accessions using Start Codon Targeted (SCoT) markers. *Int J Med Arom Plants* 3(4):413–422
- Paliwal R, Singh R, Choudhary DR, Singh AK, Kumar S, Kumar A, Bhatt KC, Singh R, Mahato AK, Singh NK, Singh R (2016) Development of genomic simple sequence repeats (g-SSR) markers in *Tinospora cordifolia* and their application in diversity analysis. *Plant Gene* 5:118–125
- Patel MB, Mishra S (2011) Hypoglycemic activity of alkaloidal fraction of *Tinospora cordifolia*. *Phyto Med* 18:045–052
- Perrier X, Flori A, Bonnot F (2003) Data analysis methods. In: Hamon P, Seguin M, Perrier X, Glaszmann JC (eds) Genetic diversity of cultivated tropical plants. Enfield Science Publishers Montpellier, Enfield, pp 43–76
- Porebski S, Bailey GL, Baum BR (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenols components. *Plant Mol Biol Rep* 15:8–15
- Rana V, Thakur K, Sood R, Sharma V, Sharma TR (2012) Genetic diversity analysis of *Tinospora cordifolia* germplasm collected from north western Himalayan region of India. *J Genet* 91:99–103
- Rana JC, Chahota RK, Sharma V, Rana M, Verma N, Verma B, Sharma TR (2015) Genetic diversity and structure of *Pyrus* accessions of Indian Himalayan region based on morphological and SSR markers. *Tree Genet Genomes* 11:821
- Saha S, Ghosh S (2012) *Tinospora cordifolia*: one plant, many roles. *Anc Sci Life* 31(4):151
- Sharma RK, Gupta P, Sharma V, Sood A, Mohapatra T, Ahuja PS (2008) Evaluation of rice and sugarcane SSR markers for phylogenetic and genetic diversity analyses in bamboo. *Genome* 51:91–103
- Sharma V, Rana M, Katoch M, Sharma PK, Ghani M, Rana JC, Sharma TR, Chahota RK (2015) Development of SSR and ILP markers in horse gram (*Macrotyloma uniflorum*), their characterization, cross-transferability and relevance for mapping. *Mol Breed* 35:102
- Sharma P, Dwivedee BP, Bisht D, Dash AK, Kumar D (2019) The chemical constituents and diverse pharmacological importance of *Tinospora cordifolia*. *Heliyon* 5:e02437
- Shi T, Li R, Chen Q, Li Y, Pan F, Chen Q (2017) De novo sequencing of seed transcriptome and development of genic-SSR markers in common buckwheat (*Fagopyrum esculentum*). *Mol Breed* 37:147
- Shinde VM, Dhalwal K (2010) DNA fingerprinting of *Tinospora cordifolia* using RAPD analysis. *J Global Pharma Technol* 2:38–42
- Singh K, Kadyan S, Panghal M, Yadav JP (2014) Assessment of genetic diversity in *Tinospora cordifolia* by inter simple sequence repeats (ISSR) and expressed sequence tagged- simple sequence repeats (EST-SSR). *Int J Pharm Pharm Sci* 6(10):520–524
- Singh R, Kumar R, Mahato AK, Paliwal R, Singh AK, Kumar S, Marla SS, Kumar A, Singh NK (2016) De novo transcriptome sequencing facilitates genomic resource generation in *Tinospora cordifolia*. *Funct Integr Genomics* 16(5):581–591
- Spandana U, Ali SL, Nirmala T, Santhi M, SipaiBabu SD (2013) A review on *Tinospora cordifolia*. *Int J Curr Pharm Rev Res* 4(2):61–68
- Srivastava P (2011) *Tinospora cordifolia* (Amrita)—A miracle herb and lifeline to many diseases. *Int J med Arom Plants* 1(2):57–61
- Taheri S, Abdullah TL, Rafii MY, Harikrishna JA, Werbrouck SPO, Teo CH, Sahebi M, Azizi P (2019) De novo assembly of transcriptomes, mining, and development of novel EST-SSR markers in *Curcuma alismatifolia* (Zingiberaceae family) through Illumina sequencing. *Sci Rep* 9:3047
- Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet* 106:411–422
- Thomas A, Rajesh EK, Suresh Kuma D (2016) The significance of *Tinospora crispa* in treatment of diabetes mellitus. *Phyto Ther Res* 30:357–366
- Wang H, Lei Y, Yan L, Wan L, Cai Y, Yang Z, Lv J, Zhang X, Xu C, Liao B (2018) Development and validation of simple sequence repeat markers from *Arachis hypogaea* transcript sequences. *Crop J* 6:172–180
- Zhang Z, Xie W, Zhao Y, Zhang J, Wang N, Ntakirutimana F, Yan J, Wang Y (2019) EST-SSR marker development based on RNA-sequencing of *E. sibiricus* and its application for phylogenetic relationships analysis of seventeen *Elymus* species. *BMC Plant Biol* 19:235
- Zheng X, Pan C, Diao Y, You Y, Yang C, Hua Z (2013) Development of microsatellite markers by transcriptome sequencing in two species of Amorphophallus (*Araceae*). *BMC Genomics* 14(1):490