



Contents lists available at ScienceDirect

Current Plant Biology

journal homepage: www.elsevier.com/locate/cpb



Development and characterization of genomic microsatellite markers in *Prosopis cineraria*

Shashi Shekhar Anand, Sapna Thakur, Madhuranjana Gargi, Shruti Choudhary, Pankaj Bhardwaj*

Molecular Genetics Laboratory, Centre for Plant Sciences, Central University of Punjab, Bathinda, India

ARTICLE INFO

Article history:

Received 8 December 2016

Received in revised form 3 March 2017

Accepted 7 March 2017

Keywords:

Genetic diversity

Leguminosae

Microsatellite markers

Polymorphic information content

Prosopis cineraria

ABSTRACT

Characterization of genetic diversity is a must for exploring the genetic resources for plant development and improvement. *Prosopis cineraria* is ecologically imperative species known for its innumerable biological benefits. Since there is a lack of genetic resources for the species, so it is crucial to unravel the population dynamics which will be very effective in plant improvement and conservation strategies. Of the 41 genomic microsatellite markers designed from (AG)_n enriched library, 24 were subsequently employed for characterization on 30 genotypes of Indian arid region. A total of 93 alleles with an average 3.875 could be amplified by tested primer pairs. The average observed and expected heterozygosity was 0.5139 and 0.5786, respectively with 23 primer pairs showing significant deviations from Hardy-Weinberg equilibrium. Polymorphic information content average to 0.5102 and the overall polymorphism level was found to be 93.27%. STRUCTURE analysis and DARwin exhibited the presence of 4 clusters among 30 genotypes.

© 2017 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Prosopis cineraria is a thorny, irregularly branched flowering tree of family Leguminosae. It is one of the most important natural resources of arid regions of India because of its economic values (fuel, fodder), ecological role in preventing soil erosion, multipurpose utility in local medication besides the ability to grow in difficult environmental conditions. The plant is adapted to survive in highly saline and alkaline soils and as a nitrogen fixer, it is grown aside the millet crop at many places in India. It occupies a special place in the life of desert dwellers of many rural communities as a sacred tree. Considering the dry forest habitat of this species (one of the most threatened ecosystems) that are being converted to agricultural land at a much faster rate and its over-exploitation, a knowledge of genetic diversity of the species would support the conservation or restoration strategies in future programmes [1,2].

Microsatellite markers are tandem repeats of 2–7 nucleotides thus termed or simple sequence repeats (SSRs). The traits such as co-dominance, high polymorphism, locus specificity and ubiquitous presence along the genome make them suitable for genetic diversity analysis. Given the lack of genetic resources for this eco-

logically important species, the present study was undertaken to develop microsatellite for its diversity estimation. Microsatellite markers are extremely efficient in genetic diversity estimation studies. Microsatellite markers can be developed directly from genomic library or from specifically enriched library. In this study, (AG)_n enriched library was used to developed 41 genomic microsatellite markers which were subsequently characterized on the genotypes collected from the arid regions. The (AG)_n repeats are highly abundant in plants, hence could be very useful for developing a comprehensive range of SSR markers [3,4].

2. Materials and methods

A total of 30 genotypes of *P. cineraria* (voucher issued by Punjab University, India; Accession no. 21119) were collected from the arid regions of North India encompassing Delhi and Punjab (Abohar, Bathinda, Malout). The complete information regarding geographical locations of sampling sites is given in Table 1. Genomic DNA from young leaves was isolated by CTAB method [5] with some modifications. Quantity and quality of DNA were estimated on 0.8% agarose gel as well as spectrophotometrically on NanoDrop 2000.

A single SSR library enriched for dinucleotide repeats (AG)₁₀ was constructed from genomic DNA using biotin-streptavidin capture protocol with the help of EcoRI and MseI restriction endonuclease with minor modifications [6]. The restriction enzyme

* Corresponding author.

E-mail address: pankajihbt@gmail.com (P. Bhardwaj).

Table 1
Geographical locations of 30 genotypes.

Serial No.	Site Name	Longitude	Latitude
City-Bathinda			
1	Near Sushant city	74.99028	30.131825
2	Bhai Matidas nagar	74.963253	30.185764
3	Village Jasii pauwali	74.970558	30.157018
4	Near Hyundia showroom	74.941336	30.176157
5	Housefed colony	74.965901	30.171402
6	Ghudha main campus	74.993491	30.154053
7	Multan road	74.789295	30.12447
8	Rampura phul	74.734391	30.100732
9	Badal road	75.242072	30.275817
10	Behind Redcliff school	74.945475	30.210994
City-Delhi			
11	JNU Purvanchal hostel	77.170082	28.534665
12	IUAC	77.194319	28.499019
13	ICGEB	77.169632	28.529652
14	Ladosarai DDA flats	77.200521	28.550686
15	Safdarjung development area	77.196404	28.522397
City-Abohar			
16	Playground near railway colony	74.194636	30.139233
17	Village sayyd wala ganaga nagar	73.87719	29.90384
18	Kandh wala road abohar	74.189435	30.123679
19	Village Alamgarh, Ganganagar road	74.210595	29.438588
20	Village Sappan wali	74.082278	30.093506
City-Malout			
21	Patti Karam Chand Mehraj	75.19083	30.28611
22	Mehraj link road	75.11694	30.26306
23	Laherbagga	74.93417	30.23639
24	Thermal colony stadium	74.95972	30.46444
25	Faridkot	74.94972	30.44194
26	Rauwla	74.90889	30.38333
27	Goniana Jatti road	74.77806	30.30512
28	Killi Nihal Singhwala	74.79444	30.23667
29	Balluana	74.74528	30.44056
30	Kisanpura	74.78083	30.43111

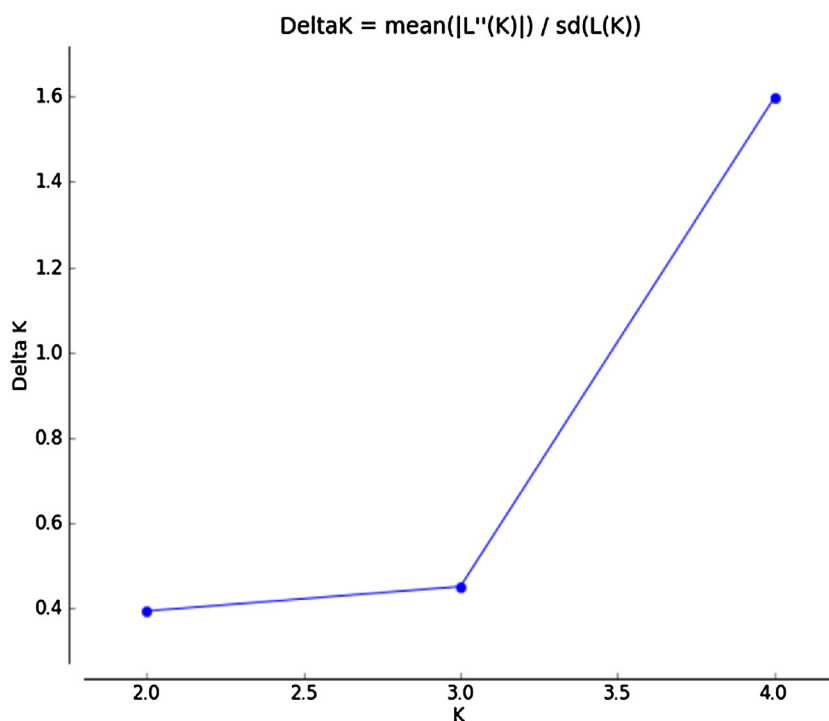
**Fig. 1.** The optimum number of populations (ΔK) using multi-locus Bayesian analysis.

Table 2
Features and evaluation details of 24 microsatellite markers.

Sl. No.	Locus Name	Primer Sequences	Repeat Motif	Ta	Na	Heterozygosity				Approx size range(bp)	No. of genotypes amplified	Accession no.
						H _o	H _e	PIC	I			
1.	PC.1	F5' AGTGAGATTGATGAGTCCTGA R5' ATACTCAAGCTATGCATCCAA	(CT)10	54	7	0.4000***	0.6836	0.646	1.4715	166	14	KX588495
2.	PC.2	F5' TGGACAGTGTAGAGAGAGAGG R5' GTGCAGGATAAACCTGAAGAT	(GAGAGG)9	53	5	0.3000***	0.5582	0.508	1.0690	122	11	KX588496
3.	PC.4	F5' ACTAGTGATTGACTGCGTACC R5' ATCCCAAAGGCTTTATTCTTA	(TC)10	53	3	0.8333**	0.6356	0.545	1.0282	119	25	KX588498
4.	PC.5	F5' TCAAACCTTATATTCCTTCCA R5' GGGAAATTCGATTAGAGAGAGA	(CT)10	53	5	0.2000***	0.6085	0.530	1.1144	141	14	KX588499
5.	PC.9	F5' ACTAGTGATTGACTGCGTACC R5' GCTACATGCAAACTGCTAGAC	(AG)11	54	2	0.9667***	0.5079	0.375	0.6926	153	30	KX588502
6.	PC.11	F5' ACGTTTACTCAGGACTCTCA R5' GTGTAATGTGGCTTTACGCT	(CT)7	54	3	0.6000***	0.6486	0.565	1.0564	143	23	KX588503
7.	PC.13	F5' AGCAAAGCAAAATTCATGTG R5' CGATTGATGAGTCTGAGTAA	(AG)11	54	3	0.3333***	0.4723	0.419	0.8170	137	26	KX588505
8.	PC.14	F5' CAGTGACAGAGAGAGAGAGA R5' GTACCAATTCCTCTCACCAA	(AG)11	54	3	0.6333***	0.6661	0.581	1.0814	166	22	KX588506
9.	PC.15	F5' GTCTCTTCTTTTCTTTACGC R5' CAGGACTCATCAATCACTAGG	(TC)12	54	3	0.4667***	0.5672	0.496	0.9447	157	25	KX588507
10.	PC.16	F5' TAGTGGGATTGATAAGTCCTG R5' GCGTACCAATTCCTCTCTCT	(AG)10	54	5	0.7000***	0.5554	0.502	1.0506	125	30	KX588508
11.	PC.19	F5' TGAATGCGATTAGAGAGAGAG R5' GTACCAATTCGAACCTGAATC	(AG)10	54	3	0.4667***	0.4588	0.393	0.7698	172	28	KX588511
12.	PC.21	F5' AGTGAGATTGATGAGTCCTGA R5' CTAGTGAATTGACTGCGTACC	(AG)8	54	3	0.1667***	0.2153	0.199	0.4300	133	29	KX588513
13.	PC.23	F5' CATTGTGATGAGTCTGAGT R5' GCGGCCCGCAATTCACATA	(TC)6	57	4	0.5000***	0.6706	0.595	1.1653	161	21	KX588515
14.	PC.25	F5' ACTAGTGAATTGGTACGCAGA R5' CCTACATTTGATGAGTCCTGA	(AG)10	54	2	0.2333 ^{NS}	0.2096	0.185	0.3602	133	30	KX588517
15.	PC.26	F5' ACTAGTGATTGACTGCGTACC R5' CATGAGTGATGAGTCTGAGT	(TC)11	54	2	1.0000***	0.5085	0.375	0.6931	154	30	KX588518
16.	PC.27	F5' CAGGACTCATCAATCACTAGG R5' TTATAAGGATCCAGGGCATA	(AG)10	54	4	0.3667***	0.5689	0.512	1.0384	154	24	KX588519
17.	PC.28	F5' ACTAGTGATTGACTGCGTACC R5' GTCCTGAGTAATGTGAGACCA	(CT)10	54	3	0.4333***	0.6119	0.519	0.9876	142	26	KX588520
18.	PC.29	F5' AATATGGACAGTGCAGAGAGA R5' GTACCAATTCACCCACCAA	(AG)10	54	7	0.7667***	0.7565	0.706	1.5613	113	25	KX588521
19.	PC.30	F5' AGAGAGAGAGAGGAATTGG R5' ATGAAGTGGCTTCTTTTCTT	(AG)12	54	3	0.7333***	0.6486	0.562	1.0521	100	24	KX588522
20.	PC.33	F5' TGAGATTGATGAGTCTGAGT R5' GTACCAATTCACCCACCAA	(AG)10	54	4	0.1667**	0.5418	0.494	1.0138	172	26	KX588525
21.	PC.38	F5' CGTAAACTCGGAATTGGTA R5' CCAATTCCTCTCTCTCTCTCT	(GA)9	53	7	0.2667***	0.7006	0.640	1.4298	151	20	KX588530
22.	PC.39	F5' AGAAAACATGTACGACTCAGC R5' TGTCTCTGAGTAATATGGACAGT	(TC)11	54	3	0.3000***	0.5859	0.506	0.9637	150	13	KX588531
23.	PC.40	F5' TGGTACGCAGTCACTCACTA R5' CGTACCAATTCCTCTCTCTCT	(AG)8	54	4	0.7667***	0.7294	0.667	1.3221	168	25	KX588532
24.	PC.41	F5' AGAGAGAGAGAGACCAATT R5' CCTAGTGATTGATGAGTCTCTG	(AG)20	54	5	0.7333***	0.7774	0.725	1.5137	167	23	KX588533
Mean					3.8750	0.5139	0.5786	0.5102	1.0261			
SD					1.5126	0.2531	0.1585		0.3057			

Note: Ta: Annealing temperature; Na: Total number of alleles; Ho: Observed Heterozygosity; He: Expected Heterozygosity; PIC: Polymorphic information content; I: Shannon's Informative index; Significant deviations from Hardy-Weinberg equilibrium at *p < 0.05, ***p < 0.001.

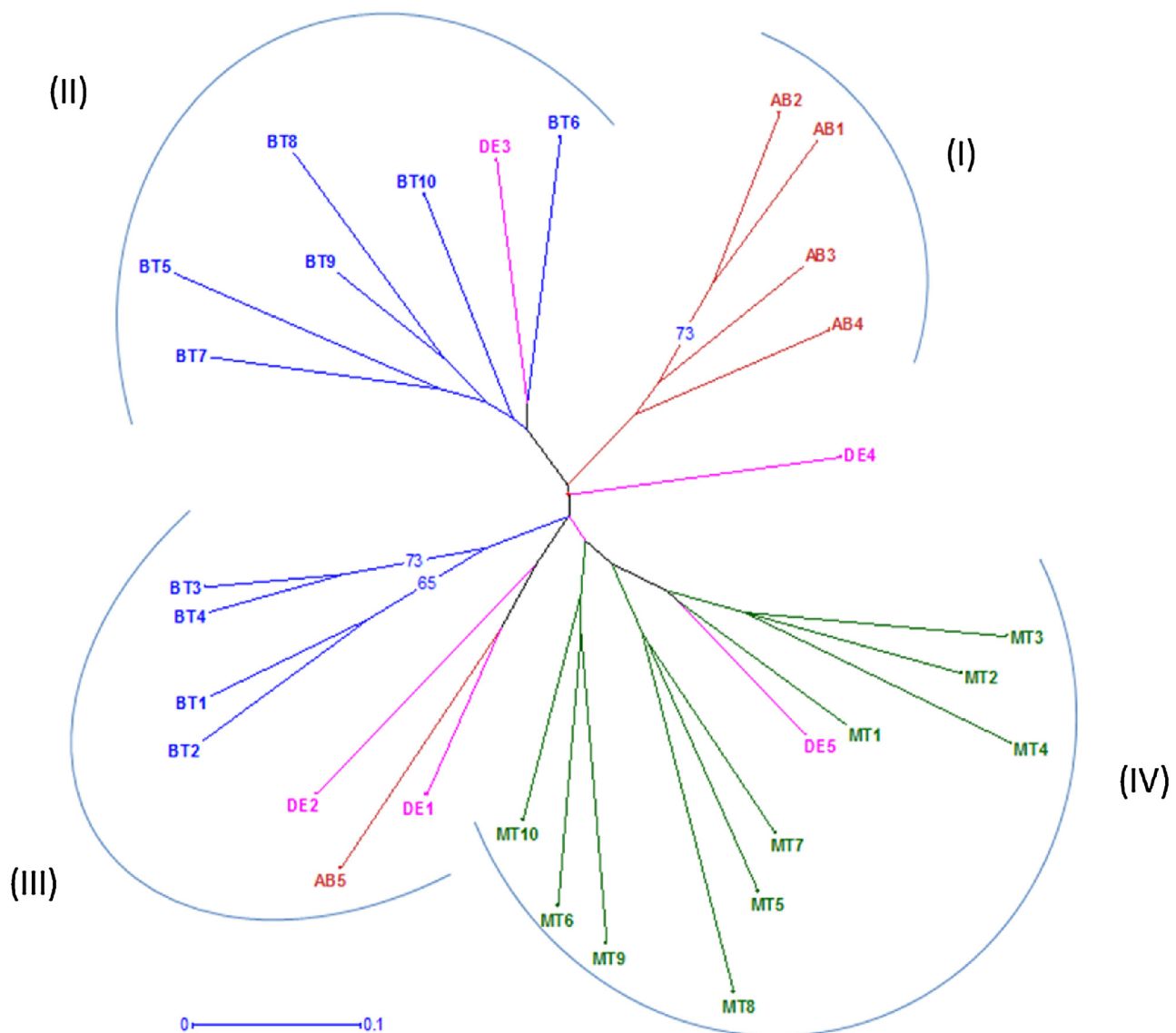


Fig. 2. Unweighted Neighbor-joining tree of 30 genotypes as constructed by DARwin 6.0.14; each branch represents single individual.

digested fragments were ligated to sticky adapters (EcoR1; 5'-CTCGTAGACTGCGTACC-3'/5'-AATTGGTACGCAGTCTAC-3' and Mse1; 5'-GACGATGAGTCTGAG-3'/5'-TACTCAGGACTCAT-3'), size selected (in the range of 200–1000bp) and subsequently amplified using the primers (5'-GACTGCGTACCAATTC-3'/5'-GATGAGTCTGAGTAA-3'). The amplified product thus obtained was heat denatured and hybridized to biotinylated (AG)10 probe at 65 °C for 4 h in the presence of 6× SSC buffer. The hybridized fragments were washed sequentially at room temperature and 60 °C with SSC buffer (2X and 1X SSC; 0.1% SDS) to avoid any non-specific hybridization. The (AG)_n enriched fragments were subsequently eluted by streptavidin-coated paramagnetic beads (New England Biolabs, Inc., NEB, USA) in nuclease free water. The enriched DNA was later amplified and cloned via pGEM-T EASY vector into chemically competent DH-5α (Escherichia coli strain). The clones positive for insert were selected through confirmation of interrupted β-galactosidase gene and PCR amplification (displaying more than one band on 1.2% agarose gel). Plasmids from positive clones isolated using MiniPrep kit (Nucleopore) and

then sequenced on ABI 3730xl DNA Analyzer (Applied Biosystems) using M13 universal primer.

The SSRs containing sequences were identified with SSR Identification Tool (SSRIT) [7] with self-modified specifications like tetramer as maximum motif length and at least 5 numbers of repeats. Raw sequences were voided from plasmid and adapter sequences and further checked for duplicity by ClustalW (<http://www.genome.jp/tools/clustalw>). Primers were designed only from the unique, sufficiently long sequences flanking the SSR region using Primer3 software [8] keeping in mind the primer size, product size, annealing temperature and GC content. Genetic diversity among *P. cineraria* genotypes was analyzed using SSRs. A total amplification reaction volume of 10 μl constituted of 25 ng of template DNA, 0.3 U Taq DNA polymerase (Bangalore Genei™), 1X Taq buffer (1 mM Tris pH 9.0, 50 mM KCl, 0.01% gelatin, 1.5 mM MgCl₂), 2.5 mM dNTPs, 5–10 ng each of forward and reverse primer in T100 Thermal Cycler (BIORAD). Characterization of developed primer pairs was done on 6% urea polyacrylamide gel under denaturing conditions and visualized using silver stain [9].

The effective number of alleles (N_e), observed and expected heterozygosity (H_o , H_e), Shannon's information index (I) were calculated using POPGENE software v1.31 [10]. Estimation of the hidden population was done by multi-locus Bayesian analysis using STRUCTURE v2.3.4 [11]. The optimum number of populations (ΔK) was calculated using a web-based program Structure Harvester [12]. The Cluster analysis was done by DARwin v6.0.014 [13] for evaluating the phylogenetic relationship among characterized genotypes.

3. Results and discussion

Plant species have a high frequency of GA/AG repeats, therefore (AG)₁₀ probe was utilized for the enrichment process in developing microsatellites markers. A total of 1131 recombinant clones from microsatellite (AG)_n enriched library were isolated. Out of 292 (26%) positive clones identified, 77 (27%) possessed SSRs. The level of enrichment achieved depends upon type of probe used, optimal wash temperature and buffer concentrations. All the SSR-containing sequences except one, possessed di-nucleotide AG/CT motif which shows the success of enrichment process (Table 2). The approximate genome size for this species is 1252 Mbp [14]. A total of 41 different primer pairs were developed which corresponds to 0.001% genome coverage (~30.5 Mbp/SSR marker). Out of 41 primer pairs, 24 could successfully amplify the pooled genomic DNA. The 17 primer pairs which either showed weak or no amplification in the test set of 30 genotypes, could be useful in different set of genotypes. The detailed features and evaluation details of 24 microsatellite markers on test set of 30 genotypes is presented in Table 2. A cross species BLAST of developed primer pairs using word size 7, expect threshold 1000, and program selection optimized for somewhat similar sequences (blastn) against *Prosopis alba* transcriptome i.e. SRA experiment set-SRX350952 [15] revealed 24–38% query coverage and 90–100% identity (Supplementary Table 1). A total of 93 alleles with the range of 2–7 alleles per locus could be amplified in the test array of 30 genotypes compared to 79 total alleles in the range of 1.322–1.536, reported in a study using ISSR and DAMD markers in the same species from Indian Thar Desert [16]. A higher no. of alleles obtained in the present study could be due to the highly polymorphic nature of microsatellite markers. Approximate size of amplified fragments in all the tested primers ranged from 100 to 172 bp. Polymorphic information content (PIC) varied from 0.177 to 0.725 (average = 0.5102). 16 primer pairs exhibited PIC ≥ 0.5 which signifies their high polymorphic nature [17]. A genetic diversity estimation study conducted on *P. rubriflora* using 13 SSR markers also revealed PIC in the range from 0.073 to 0.791 [18]. The observed heterozygosity was within the range 0.1667–0.9667 (average = 0.5139) and the expected from 0.2153 to 0.7774 (average = 0.5786). Using 12 SSR markers in *P. alba* and *P. chilensis*, 2–7 alleles per loci and heterozygosity in the range of 0.2–0.8 was observed [19] which is comparable to present study. 23 loci showed significant deviations from Hardy-Weinberg equilibrium and (HWE) the observed heterozygosity is found to be less than expected under HWE which indicates towards inbreeding in the arid populations of North India. The Shannon's informative index varied from 0.3602 to 1.5613 (average = 1.0261). STRUCTURE analysis based on multi-locus data assigned 30 genotypes to 4 subpopulations ($\Delta K = 4$) (Fig. 1). Cluster analysis on the basis of genetic similarity using DARwin also suggested the presence of 4 clusters (I, II, III, and IV) which is in agreement with STRUCTURE analysis (Fig. 2). Cluster (I) was comprised of individuals from Abohar, Cluster (II) from Bathinda, Cluster (IV) from Malout and Cluster (III) possessed individuals from Bathinda, Abohar, and Delhi. Using DARwin most of the genotypes clustered according to their geographical location. The present study indicates the presence of

4 subpopulations in the arid region of North India. Presence of subpopulations in *P. cineraria* indicates the genetic differentiation which in turn is governed by balance between selection and gene flow. Geographical barriers, distant spatial patterns, local adaptation and limited gene flow lead to increased genetic differentiation among populations [20,21].

The primer pairs developed in this study will not only add on to the genetic resource for this key species but also to the closely related species of the genus *Prosopis*. Furthermore, the present study will be helpful in planning conservation and restoration strategies for arid regions in the future programmes.

Acknowledgement

Sapna Thakur and Shruti Choudhary acknowledge the fellowship received from ICMR Indian Council of Medical Research towards Ph.D.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.cpb.2017.03.001>.

References

- [1] E. Khurana, J.S. Singh, Ecology of seed and seedling growth for conservation and restoration of tropical dry forest: a review, *Environ. Conserv.* 28 (2001) 39–52, <http://dx.doi.org/10.1017/S0376892901000042>.
- [2] N. Kaushik, V. Kumar, Khejri (*Prosopis cineraria*)-based agroforestry system for arid Haryana, India, *J. Arid Environ.* 55 (2003) 433–440.
- [3] X.W. Wang, A. Kaga, N. Tomooka, D.A. Vaughan, The development of SSR markers by a new method in plants and their application to gene flow studies in azuki bean [*Vigna angularis* (Willd.) Ohwi & Ohashi], *Theor. Appl. Genet.* 109 (2004) 352–360.
- [4] Z. Zhao, C. Guo, S. Sutharzan, P. Li, C.S. Echt, J. Zhang, C. Liang, Genome-wide analysis of tandem repeats in plants and green algae, *G3: Genes Genomes Genet.* 4 (2014) 67–78.
- [5] J.J. Doyle, Isolation of plant DNA from fresh tissue, *Focus* 12 (1990) 13–15.
- [6] P. Bhardwaj, R. Kumar, H. Sharma, R. Tewari, P.S. Ahuja, R.K. Sharma, Development and utilization of genomic and genic microsatellite markers in Assam tea (*Camellia assamica* ssp. *assamica*) and related *Camellia* species, *Plant Breed.* 132 (2013) 748–763.
- [7] S. Temnykh, G. DeClerck, A. Lukashova, L. Lipovich, S. Cartinhou, S. McCouch, Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential, *Genome Res.* 11 (2001) 1441–1452.
- [8] A. Untergasser, I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm, S.G. Rozen, Primer3—new capabilities and interfaces, *Nucleic Acids Res.* 40 (2012) e115.
- [9] S. Thakur, S. Choudhary, A. Singh, K. Ahmad, G. Sharma, A. Majeed, P. Bhardwaj, Genetic diversity and population structure of *Melia azedarach* in North-Western Plains of India, *Trees* 30 (2016) 1483–1494, <http://dx.doi.org/10.1007/s00468-016-1381-x>.
- [10] F. Yeh, R. Yang, T. Boyle, POPGENE. Microsoft Windows based Freeware for Population Genetic Analysis: Release 1.31, University of Alberta, Edmonton, 1999, Accessed 1 May 2014, Available from <http://www.ualberta.ca/~fyeh/popgene.download.html>.
- [11] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, *Genetics* 155 (2000) 945–959.
- [12] D.A. Earl, STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method, *Conserv. Genet. Resour.* 4 (2012) 359–361.
- [13] X. Perrier, J.P. Jacquemoud-Collet, DARwin Software: Dissimilarity Analysis and Representation for Windows, 2006, Accessed 1 November 2016 Website <http://darwin.cirad.fr/darwin>.
- [14] M.D. Bennett, I.J. Leitch, Nuclear DNA amounts in angiosperms: progress, problems and prospects, *Ann. Bot.* 95 (2005) 45–90.
- [15] S.L. Torales, M. Rivarola, M.F. Pomponio, S. Gonzalez, C.V. Acuña, P. Fernández, S.N.M. Poltri, De novo assembly and characterization of leaf transcriptome for the development of functional molecular markers of the extremophile multipurpose tree species *Prosopis alba*, *BMC Genom.* 14 (2013) 705.
- [16] S.K. Sharma, S. Kumar, D. Rawat, S. Kumaria, A. Kumar, S.R. Rao, Genetic diversity and gene flow estimation in *Prosopis cineraria* (L.) Druce: a key stone tree species of Indian Thar Desert, *Biochem. Syst. Ecol.* 39 (2011) 9–13.
- [17] A. Khar, K. Lawande, K. Negi, Microsatellite marker based analysis of genetic diversity in short day tropical Indian onion and cross amplification in related *Allium* spp, *Genet. Res. Crop Evol.* 58 (2011) 741–752.

- [18] F.M. Alves, M.I. Zucchi, A.M. Azevedo-Tozzi, Â.L. Sartori, A.P. Souza, Characterization of microsatellite markers developed from *Prosopis rubriflora* and *Prosopis ruscifolia* (Leguminosae-Mimosoideae), legume species that are used as models for genetic diversity studies in Chaquénian areas under anthropization in South America, *BMC Res. Notes* 7 (2014) 375.
- [19] C.F. Bessega, C.L. Pometti, J.T. Miller, R. Watts, B.O. Saidman, J.C. Vilardi, New microsatellite loci for *Prosopis alba* and *P. chilensis* (Fabaceae), *Appl. Plant Sci.* 1 (2013) 1200324.
- [20] M. Fischer, R. Husi, D. Prati, M. Peintinger, M. van Kleunen, B. Schmid, RAPD variation among and within small and large populations of the rare clonal plant *Ranunculus reptans* (Ranunculaceae), *Am. J. Bot.* 87 (2000) 1128–1137.
- [21] Y. Tsumura, K. Uchiyama, Y. Moriguchi, M.K. Kimura, S. Ueno, T. Ujino-Ihara, Genetic differentiation and evolutionary adaptation in *Cryptomeria japonica*, *G3: Genes Genomes Genet.* 4 (2014) 2389–2402.