

Development of genome-wide SSR markers in horsegram and their use for genetic diversity and cross-transferability analysis

Rahul Kaldate · Maneet Rana · Vikas Sharma · Hideki Hirakawa · Rahul Kumar · Gagandeep Singh · Rakesh Kumar Chahota · Sachiko N. Isobe · Tilak Raj Sharma

Received: 18 December 2016 / Accepted: 4 July 2017 / Published online: 1 August 2017
© Springer Science+Business Media B.V. 2017

Abstract Horsegram [*Macrotyloma uniflorum* (Lam.) Verdc.] commonly known as *kulthi* or Madras gram is an important drought tolerant legume crop used as food and fodder in India and across the globe. Horsegram is tolerant to many biotic and abiotic stresses and considered a potential future food legume. Despite being a

multiutility crop, insufficient genomic information is available in this species, which is otherwise required for genetic improvement. Hence, in the present work we used next-generation sequencing (NGS) technology for genome-wide development and characterization of novel simple sequence repeat (SSR) markers in horsegram. In all, 2458 SSR primer pairs were designed from NGS data and 117 SSRs were characterized in 48 diverse lines of horsegram. Cross-transferability of these markers was also checked in nine related legume species. The polymorphic SSRs revealed high diversity measures such as mean values of expected heterozygosity (H_e ; 0.54), observed heterozygosity (H_o ; 0.64), and polymorphism information content (PIC; 0.46). Analysis of molecular variance (AMOVA) revealed high degree of genetic variance within the populations. Dendrogram based on Jaccard's similarity coefficient and principal component analysis (PCA) revealed two groups in the analyzed accessions. This observation was further confirmed by Bayesian genetic STRUCTURE analysis. The SSR markers developed herein can be used in diverse genetic analysis including association mapping in this crop and also in related legume crops with limited marker resources. Hence, this new SSR dataset can be useful for molecular breeding research in this underutilized pulse crop. In addition, genetic diversity estimates of analyzed germplasm can be important for devising future breeding programmes in horsegram.

Electronic supplementary material The online version of this article (doi:10.1007/s11032-017-0701-1) contains supplementary material, which is available to authorized users.

R. Kaldate · R. K. Chahota · T. R. Sharma
Department of Agricultural Biotechnology, CSK Himachal Pradesh Agricultural University, Palampur 176062, India

M. Rana
Division of Crop Improvement, ICAR-Indian Grassland & Fodder Research Institute, Jhansi 284128, India

V. Sharma
Department of Botany, Punjabi University Patiala, Punjab 147002, India

H. Hirakawa · S. N. Isobe
Kazusa DNA Research Institute, Kisarazu, Chiba, Japan

R. Kumar
Department of Agriculture, DAV University Jalandhar, Punjab, India

G. Singh
Biotechnology Division, CSIR-Institute of Himalayan Bioresource Technology, Palampur, India

T. R. Sharma (✉)
ICAR-Indian Institute of Agricultural Biotechnology, C/O ICAR-IINRG PDU Campus, Namkum, Ranchi 834010, India
e-mail: sharmat88@yahoo.com

Keywords Horsegram · Illumina sequencing · SSR markers · Polymorphism information content (PIC) · Cross-transferability

Introduction

Horsegram [*Macrotyloma uniflorum* (Lam.) Verdc.] is an important pulse crop of many Southeast Asian countries including India. It is also cultivated as forage crop in some other semi-arid regions of the world and grown as green manure in many tropical countries, especially in Australia and Africa. It is an underutilized warm season food legume with little genetic and genomic information. Horsegram is a diploid plant species with chromosome number $2n = 2 \times = 20, 22, 24$. The estimated genome size of horsegram is 400 Mbps (Bhardwaj et al. 2013). The name *Macrotyloma* is derived from the Greek words macros, meaning large, tylos, meaning knob, and loma, meaning margin, in reference to knobby structures on the pods (Blumenthal and Staples 1993). Genus *Macrotyloma* consists of nearly 25 species indigenous to Africa and Asia. Of these, *Macrotyloma* var. *uniflorum* is the only cultivated species (Allen and Allen 1981; Lackey 1981). In India, Horsegram is cultivated as a summer crop in Karnataka, Tamil Nadu, and Andhra Pradesh and in northwestern and central Himalayan regions of Himachal Pradesh, Jammu & Kashmir, and Uttarakhand. It is also grown in Punjab, Bihar, Uttar Pradesh, Madhya Pradesh, Rajasthan, Maharashtra, and Gujarat. The fodder is rich in protein content and free from digestive inhibitors; it is widely used as a feed for milch animals and horses (Prakash et al. 2008). Horsegram is a nutritious food legume and provides many essential dietary components. The protein content in seeds of cultivated horsegram is reported to be 16.9–30.4% (Patel et al. 1995). It has high lysine content, an essential amino acid (Gopalan et al. 1989). Horsegram seeds are a rich source of phosphorus, iron and vitamins such as carotene, thiamine, riboflavin, niacin and vitamin C (Sodani et al. 2004). It is known to contain many medicinal and therapeutically active principles, thus suggested as an Ayurvedic medicine for a variety of ailments such as edema, piles, and renal stones. It exhibits high antioxidant properties and contains molybdenum, which regulates calcium and iron, which helps in transporting oxygen to cells and forms part of hemoglobin in blood (Ramesh et al. 2011; Murthy et al. 2012). Horsegram also helps in lowering blood cholesterol level (Mehra and Upadhyaya 2013). Chaitanya et al. (2010) proved that the seeds of *M. uniflorum* are endowed with significant anti-urolithiatic activity. Certain tests have proven that lipids extracted from

horsegram are known to heal peptic ulcers in rats (Jayaraj et al. 2000). Eco-physiologically, horsegram is a drought hardy crop species and exhibits many desirable traits such as tolerance to drought, salinity, and heavy metal stresses. These attributes make it a potential future crop (Sosulski and Young 1979; Kadam et al. 1985; Oram 1990; Smartt 1990; Reddy et al. 2008). With all these features, horsegram holds a great potential as nutraceutical, forage, and food commodity in drought prone areas of the world (Morris 2008).

Although, a crop species of potentially high value, it received little attention of researchers for its genetic improvement. This may be due to limited availability of genetic resources and molecular breeding tools in this crop. Therefore, it is immensely important to develop marker resources in this crop which could be used to study genetic diversity, linkage analysis, QTL and association mapping studies. In addition, marker-assisted selection (MAS) and genomic selection (GS) can also be performed if sufficient genomic resources become available for the purpose. So far, only 232 characterized SSRs along with 13 ILP markers have been developed in this crop (Sharma et al. 2015a). However, more marker resources are required in the species to initiate a genetic improvement programme. Simple sequence repeats (SSRs) are widely used due to their co-dominant inheritance, multi-allelic nature, high reproducibility and transferability, extensive genome coverage, and simple detection methods (Varshney et al. 2005; Agarwal et al. 2008). Thus, in the present study genome-wide SSR markers were developed in horsegram using Illumina sequencing and were used for genetic diversity analysis in horsegram germplasm. The cross-transferability of these markers was also examined across nine related legume species.

Material and methods

Sequencing and assembly of reads

Illumina paired-end libraries were constructed from total genomic DNAs according to the instructions provided by the manufacturer. The developed Illumina libraries were subjected to genome sequencing with an Illumina HiSeq 1000. A total of 375 M paired-end reads of “HPK4” were obtained by an Illumina HiSeq 2000 and subjected to genome assembly with a SOAPdenovo2 r223 (Li et al. 2010). After the assembly, gaps in the

scaffolds were closed by GapCloser 1.10 ($p = 31$). The number of assembled scaffolds were 237,662 (kmer = 81) consisting of 313,833,559 bases (N50 length = 97,711 bases). This sequence information was used to identify genomic SSRs.

SSR mining and primer design

The sequence database obtained as described above was used as a source of potential SSRs with different repeat types. After pre-processing, SSR containing sequences were identified by PERL script-based program, MISA software (MICroSATellite identification tool, <http://pgrc.ipk-gatersleben.de/misa/>). The criteria for identification of SSRs were a minimum of six repeat units for di-nucleotide repeats (DNR) and a minimum of five repeat units for all other types of repeats, i.e. tri-nucleotide repeats (TNR), tetra-nucleotide repeats (TtNR), penta-nucleotide repeats (PNR), and hexa-nucleotide repeats (HNR). All SSR-containing sequences were used for primer design following standard parameters by using the BatchPrimer3 software (<http://probes.pw.usda.gov/cgi-bin/batchprimer3/batchprimer3.cgi>) considering the following criteria: (1) primer length ranging from 18 to 22 with an optimum 20 bp, (2) PCR product size ranging from 100 to 400 with an optimum 300 bp, (3) annealing temperature ranging from 45 to 60 °C, (4) GC content ranging from 40 to 60 with an optimum of 50%. Newly designed SSRs were named as *M. uniflorum* Genomic SSR (MUGSSR).

Plant materials and DNA isolation

The panel of 48 horsegram lines were collected from different horsegram growing regions of the country and maintained at the Department of Agricultural Biotechnology, CSK HP Agricultural University Palampur, India, and were selected for the diversity analysis and characterization of newly developed SSRs. For cross-transferability analysis, nine related legume species namely-*M. sar-gharwalensis*, *Cicer arietinum*, *Vigna unguiculata*, *Lens culinaris*, *Vigna mungo*, *Pisum sativum*, *Trifolium pratense*, *Vigna umbellata*, and *Phaseolus vulgaris* were taken. Genomic DNA was isolated from young leaves of each plant using the CTAB method (Doyle and Doyle 1990) with some modifications. The quantity and quality of DNA was

estimated using 0.8% agarose gel and by comparing with uncut lambda DNA (Fermentas, Lithuania).

SSR genotyping

The polymorphism survey of 117 newly developed primers pairs in 48 selected horsegram lines was carried out following the method of Sharma et al. (2015a). The polymorphic primer pairs were also used for cross-transferability study in related species. For amplification of genomic DNA, a reaction mixture of 10 µL volume was prepared using 4.8 µL of sterilized distilled water, 2.0 µL genomic DNA (13 ng/µL), 0.5 µL of forward and 0.5 µL of reverse primer (5 µM), 0.5 µL MgCl₂ (25 mM), 1.0 µL 10 X PCR buffer (10 mMTris-Hcl, 50 mMKcl, pH 8.3), 0.5 µL dNTP mix (0.2 mM each of dATP, dGTP, dCTP, and dTTP), and 0.2 µL Taq polymerase (5 U/µL). The PCR conditions were: one cycle of 5 min at 94 °C, 35 cycles of 1 min at 94 °C, 1 min at respective annealing temperature for each primer, as shown in Supplementary Table 2, 1 min at 72 °C, final extension for 7 min at 72 °C. All the PCR reactions were conducted in 96-well Gene Amp PCR System 9700® (Applied Biosystems, CA, USA) and 2720 Thermal Cycler (Applied Biosystems, CA, USA). PCR products were resolved in 6% polyacrylamide gel at a constant current of 65 W at room temperature for 90 min. Gels were prepared and run in 1X TBE buffer and visualization of fragments was done using silver-staining. Size estimation of the alleles was done by using 50 bp DNA size standard (Fermentas, Lithuania).

Data analysis

All unambiguously amplified fragments were scored manually for each SSR locus and included for further analysis. The polymorphism information content (PIC) was calculated according to the formula given by Botstein et al. (1980) and implemented in Cervus version 3.0 (Kalinowski et al. 2007). Other genetic diversity attributes such as expected heterozygosity (H_e), observed heterozygosity (H_o), Shannon information index (I), etc. were calculated with the help of POPGENE version 1.32 (Yeh and Boyle 1997). A neighbor-joining (NJ) tree was constructed using Jaccard's coefficient with the help of DARwin Version 5.0.158 (Perrier and Jacquemoud-Collet (2006). Genetic relationships among the genotypes were also analyzed by principal component analysis (PCA) using DARwin. Genetic

structure analysis was performed by the Bayesian clustering model using software STRUCTURE version 2.3 (Pritchard et al. 2000). An admixture model with correlated allele frequencies was used to infer the value of K with prior population information. All analyses were performed with a burn-in period of 100,000 and a Markov chain Monte Carlo (MCMC) replication number set at 1,000,000. The value of K was estimated using the method described by Evanno et al. (2005) and was obtained using STRUCTURE HARVESTER (Earl and von Holdt 2012). Analysis of molecular variance (AMOVA) was done using GenAlEx (Peakall and Smouse 2012).

Results and discussion

Development and frequency distribution of different SSRs

Microsatellites are tandem repeat motifs of 1–6 bp, which have a frequent occurrence in all prokaryotic and eukaryotic genomes (Hamada et al. 1982; Tautz 1989; Zane et al. 2002). These are found in both coding and non-coding regions and are distributed throughout the nuclear genome. These can also be found in the chloroplast and mitochondrial genomes (Soranzo et al. 1999; Provan et al. 2001; Chung et al. 2006; Rajendrakumar et al. 2007). These properties of SSRs present opportunities and chances of finding SSRs in unexplored species such as horsegram. Moreover, continued advances in next-generation sequencing (NGS) technologies have resulted in generation of large amounts of sequence data and data mining for marker development. Hence, in this work we applied NGS using Illumina sequencing to generate sequence data and to develop new SSR primers. A total of 86,498 sequences were found containing SSRs (supplementary Table 1). Of these, 2458 primer pairs were designed, and 117 were randomly selected for synthesis and validation (Supplementary Table 2). The detailed information about additional 2341 primers designed in the present study is given in Supplementary Table 3. The horsegram sequence data analyzed here comprised 186,445 scaffolds having a total length of 305,302.6 kb (305.30 Mbps). These were found to contain a total of 86,498 SSRs located in 20,966 scaffolds in which 7267 scaffolds had more than one SSR, and 9979 had SSRs of different motifs. DNR were the most abundant (48,202 or 55.73%) followed

by TNR (35,635 or 41.22%) and TtNR (1947 or compound SSRs) (supplementary Table 1). The proportion of different SSR unit size was not evenly distributed. Among DNRs, the AT/AT (85.46%) was the most abundant followed by AG/CT (11.05%), whereas, among TNRs, the AAG/CTT (63.35%) were most abundant followed by AAT/ATT (28.84). Similarly among TtNRs, PNRs, and HNRs the AAAG/ATTT (54.64%), AATAT/ATATT (27.76%), AAAAAT/ATTTTT, and AAATAT/ATATTT (6.71%), were most abundant repeats, respectively. The number of SSR repeats ranged from five to 272, with five repeats 26,284 (30.39%) being the most abundant, followed by six 13,483 (15.59%) and seven repeats 7879 (9.11%) as the next most abundant. Motifs with more than 10 repeats were 20,192 (23.34%).

Validation and polymorphic potential of SSRs

Of the 2458 designed primers, 117 SSR primer pairs were used for validation and assessment of the polymorphism among 48 diverse horsegram lines. Of these, 47 exhibited polymorphism and 70 were found monomorphic. Out of 47 polymorphic markers, 31 were DNRs, 10 TNRs, four TtNRs, one each PNR and HNR. Similarly, out of 70 monomorphic markers, 34 were DNRs, 18 TNRs, seven TtNRs, eight PNRs, and three HNRs. A total of 192 alleles were detected by 47 polymorphic SSR markers in which two markers MUGSSR-536 and MUGSSR-537 amplified two loci each MUGSSR-536(A), MUGSSR-536(B) and MUGSSR-537(A), MUGSSR-537(B) with seven, three and four, four alleles, respectively (Table 1). The total number of loci amplified was found to be 49 with an average of 3.90 alleles amplified per locus, with a range of two to 15 alleles, revealing high allelic diversity. Primer pair MUGSSR-549 amplified a maximum of 15 alleles followed by MUGSSR-526 with 13 alleles. Size of alleles ranged from 50 bp to 350 bp. The PIC values ranged from 0.06 in primer MUGSSR-24 to 0.87 in MUGSSR-549 with an average value of 0.46. Sharma et al. (2015a) found all characterized SSR and ILP primers amplified alleles in range of 1–8 with an average of 2.64 alleles per primers and PIC ranged from 0.04 to 0.75 with an average of 0.36. The mean observed number of alleles was 3.19 ± 2.65 in different genotypes used in the study. Highest (0.89) and lowest (0.07) expected heterozygosity (H_E) values were obtained with MUGSSR-549 and MUGSSR-24, respectively, with an

average of 0.54. The observed heterozygosity ranged from 0 to 1 with a mean value of 0.64; Shannon's information index (I) was recorded maximum (2.35) for primer pair MUGSSR-549, whereas the lowest value (0.15) was observed for MUGSSR-24 with an average of 0.94. The genetic diversity revealed by SSR loci was supported by the observed high values of mean allele numbers, heterozygosity and PIC values. However, mean heterozygosity and PIC values were 0.54 and 0.46, respectively. Average Shannon's information index (I) was found to be 0.94 (Table 1). Amplification profiles of 48 horsegram genotypes generated by MUGSSR-530, MUGSSR-532, and MUGSSR-546 markers have been shown in supplementary Figure 1. The estimates of these diversity indices using SSR primers were nearly similar to those obtained by Sharma et al. (2015b) in this species using RAPD and ISSR markers, where the average PIC was reported as 0.34. In another study, using 360 genetically diverse accessions of horsegram, an average of 5.73 alleles per primer pair were observed with mean heterozygosity and PIC of 0.58 and 0.50, respectively (Divya Shikha 2015) and average Shannon's Information Index (I) was found to be 1.07.

Cross-transferability studies

Cross-transferability offers potential for the low-cost generation of microsatellite markers for related or distant species. In order to assess the cross-transferability of novel horsegram SSR markers, the SSR primers were checked for amplification in nine other legume species. The SSR markers showed variable extent of cross-transferability in the analyzed species. Of the 47 primer pairs, 43 produced amplicons in at least one legume species analyzed, whereas four primer pairs did not amplify any amplicons. The amplification profile of nine related species with primer pair MUGSSR-542 is shown in supplementary Figure 1. Of the 43 primer pairs that showed cross-amplification, five primer pairs produced amplicons in all the legumes analyzed, indicating 100% transferability. Among 43 genomic SSR markers, 16 showed amplification each in *M. sargharwalensis* and *P. vulgaris*, 15 in *V. unguiculata*, 17 each in *C. arietinum*, *L. culinaris*, and *V. mungo*, 43 in *P. sativum*, 25 in *T. pratense*, 19 in *V. umbellata*. Surprisingly, highest species-wise transferability was observed in *P. sativum* (91.49%) followed by *T. pratense*

(53.19%), *V. umbellata* (40.43%), *V. mungo*, *C. arietinum*, and *L. culinaris* (36.17% each), *M. sargharwalensis* and *P. vulgaris* (34.04% each), and *V. unguiculata* with lowest value (31.91%). Percentage of cross-transferability in different species is depicted in supplementary Figure 2. There have been many examples of SSR markers transferred within legume genera for example from soybean to peanut (He et al. 2006); from *M. truncatula* to pea, fava bean, and *C. arietinum* (Gutierrez et al. 2005); and from *M. truncatula*, *Pisum sativum*, and *Trifolium pratense* to *Lens culinaris* (Reddy et al. 2010).

Distance based cluster analysis

A neighbor-joining (N-J) tree of 48 accessions showed two major clusters with high (> 50.0%) bootstrap value (Fig. 1). Of the two clusters, the first cluster comprised accessions from Himachal Pradesh. Interestingly, the single accession of wild species *M. axillare* and two accessions each from Madhya Pradesh and Tamil Nadu were also clustered with this group. On the other hand, the accessions belonging to Southern Indian states such as Tamil Nadu and Andhra Pradesh were grouped in the second cluster. Jaccard's similarity matrix showed the highest genetic similarity of 0.93 between HPKM-317 and HPKM-150 indicating that they are genetically the most similar accessions because they are mutants developed from the same parent, validating results obtained in the present study. Accessions TCR-654 and IC-120825 were genetically most dissimilar with the similarity value of 0.25. These four accessions belonged to *M. uniflorum*. When related species *M. axillare* was taken into consideration for similarity estimates it showed maximum dissimilarity with TCR-654 (*M. uniflorum*) with the similarity value of 0.18. Two-dimensional graphical view of genetic diversity in 48 accessions was represented in PCA analysis (Fig 2), which clearly showed the existence of two groups. The clustering pattern showed correspondence with geographical distribution of accessions. This grouping may be attributed to the frequent flow of alleles within a confined population of a geographical region rather than between populations of two geographical regions. Further, there are two main regions of horsegram diversity in the country, one of these includes Andhra Pradesh and other southern states and the other comprises some hilly regions of the north Indian states such as Himachal

Table 1 Diversity statistics for 47 SSR loci studied in 48 horsegram accessions

S. No.	Primer	Allele Size range(bp)	Na	Ne	I	Ho	He	PIC
1	MUGSSR-04	150-200	2	1.972	0.686	0.881	0.499	0.371
2	MUGSSR-06	100-150	4	3.939	1.379	1.000	0.754	0.699
3	MUGSSR-07	100-200	3	2.024	0.740	0.915	0.511	0.388
4	MUGSSR-14	100-150	3	1.649	0.717	0.000	0.402	0.360
5	MUGSSR-24	150-200	2	1.074	0.154	0.000	0.070	0.067
6	MUGSSR-201	100-200	2	1.724	0.611	0.600	0.427	0.332
7	MUGSSR-202	150-200	3	1.352	0.516	0.293	0.263	0.244
8	MUGSSR-203	100-200	2	1.992	0.691	0.938	0.506	0.374
9	MUGSSR-205	150-200	6	5.036	1.696	0.621	0.815	0.773
10	MUGSSR-206	50-150	3	2.571	1.011	1.000	0.623	0.535
11	MUGSSR-208	100-200	2	1.541	0.536	0.455	0.357	0.290
12	MUGSSR-210	100-200	2	1.701	0.602	0.516	0.419	0.327
13	MUGSSR-211	50-150	4	2.272	0.939	0.917	0.568	0.466
14	MUGSSR-213	100-150	2	1.993	0.691	0.939	0.506	0.374
15	MUGSSR-214	150-200	4	2.068	0.800	0.878	0.523	0.406
16	MUGSSR-216	100-150	2	1.998	0.693	0.914	0.507	0.375
17	MUGSSR-218	100-200	2	1.942	0.678	0.759	0.494	0.367
18	MUGSSR-219	100-150	2	2.000	0.693	1.000	0.508	0.375
19	MUGSSR-220	100-150	4	1.294	0.488	0.167	0.230	0.216
20	MUGSSR-235	50-150	4	2.008	0.905	0.604	0.507	0.443
21	MUGSSR-241	100-150	3	1.883	0.716	0.710	0.476	0.373
22	MUGSSR-504	100-200	2	1.999	0.693	0.979	0.505	0.375
23	MUGSSR-506	100-200	3	1.991	0.732	0.830	0.503	0.384
24	MUGSSR-508	150-200	2	1.999	0.693	0.979	0.505	0.375
25	MUGSSR-509	150-200	3	2.036	0.743	0.958	0.514	0.390
26	MUGSSR-512	100-200	3	2.044	0.746	1.000	0.517	0.391
27	MUGSSR-513	100-200	2	1.465	0.497	0.395	0.321	0.267
28	MUGSSR-515	150-200	4	2.085	0.794	1.000	0.526	0.405
29	MUGSSR-520	150-200	2	1.992	0.691	0.938	0.503	0.374
30	MUGSSR-521	150-200	2	1.340	0.421	0.298	0.256	0.221
31	MUGSSR-524	150-200	6	3.731	1.495	0.024	0.741	0.691
32	MUGSSR-525	150-190	2	1.996	0.692	0.957	0.504	0.375
33	MUGSSR-526	250-350	13	6.354	2.102	0.978	0.852	0.825
34	MUGSSR-528	250-350	3	2.080	0.779	0.979	0.525	0.404
35	MUGSSR-530	200-300	7	3.106	1.421	0.222	0.686	0.644
36	MUGSSR-532	250-350	4	2.860	1.131	0.064	0.657	0.581
37	MUGSSR-534	200-300	8	2.492	1.144	1.000	0.605	0.517
38	MUGSSR-536(A)	50-100	7	4.830	1.728	0.771	0.801	0.766
39	MUGSSR-536(B)	200-300	3	2.256	0.885	0.000	0.563	0.458
40	MUGSSR-537(A)	50-100	4	3.459	1.306	1.000	0.719	0.658
41	MUGSSR-537(B)	200-300	4	3.144	1.243	1.000	0.689	0.623
42	MUGSSR-542	250-300	3	1.524	0.637	0.000	0.348	0.314
43	MUGSSR-544	250-300	5	3.316	1.306	0.146	0.707	0.641
44	MUGSSR-545	200-300	6	4.215	1.601	0.000	0.771	0.732

Table 1 (continued)

S. No.	Primer	Allele Size range(bp)	N _a	N _e	I	H _o	H _e	PIC
45	MUGSSR-546	250-300	4	2.600	1.105	0.089	0.622	0.548
46	MUGSSR-547	250-350	5	2.511	1.159	0.140	0.609	0.552
47	MUGSSR-548	250-300	2	1.999	0.693	0.979	0.505	0.375
48	MUGSSR-549	250-350	15	8.491	2.359	0.698	0.892	0.871
49	MUGSSR-550	250-300	7	4.148	1.605	1.000	0.767	0.722
	Mean		3.92	2.573	0.946	0.643	0.544	0.462

bp Base pair, **N_a** Observed number of alleles; **N_e** Effective number of alleles; **I** Shannon's information index; **H_e** Expected heterozygosity; **H_o** Observed heterozygosity; **PIC** Polymorphism information content

Pradesh, Uttarakhand, and Jammu & Kashmir (Neelam et al. 2014; Sharma et al. 2015a). In PCA, group-I consisted of accessions from Himachal Pradesh with one *M. axillare* accession from Australia and two accessions each from Madhya Pradesh (IC-19432 and IC-23445) and Tamil Nadu (IC-469259 and IC-469271). Group-II had accessions primarily from the southern states of India, which confirmed the results obtained by the neighbor-joining tree. Similar results were also reported by Sharma et al. (2015b) while studying diversity in 51 horsegram accessions using RAPD and ISSR primers.

Bayesian genetic structure

Bayesian clustering method is a powerful computational tool meant for estimating various features of populations. STRUCTURE assumes K (unknown) populations for the given data set, and the value of K can be estimated by posterior probability of the data for a given K. Delta K, which is used to determine the best fit value of K, was computed by STRUCTUREHARVESTER for the given range, i.e. 1–10 and highest value was shown at K = 2. The maximum Delta K mean was 308.06 at K = 2. STRUCTURE analysis indicated two

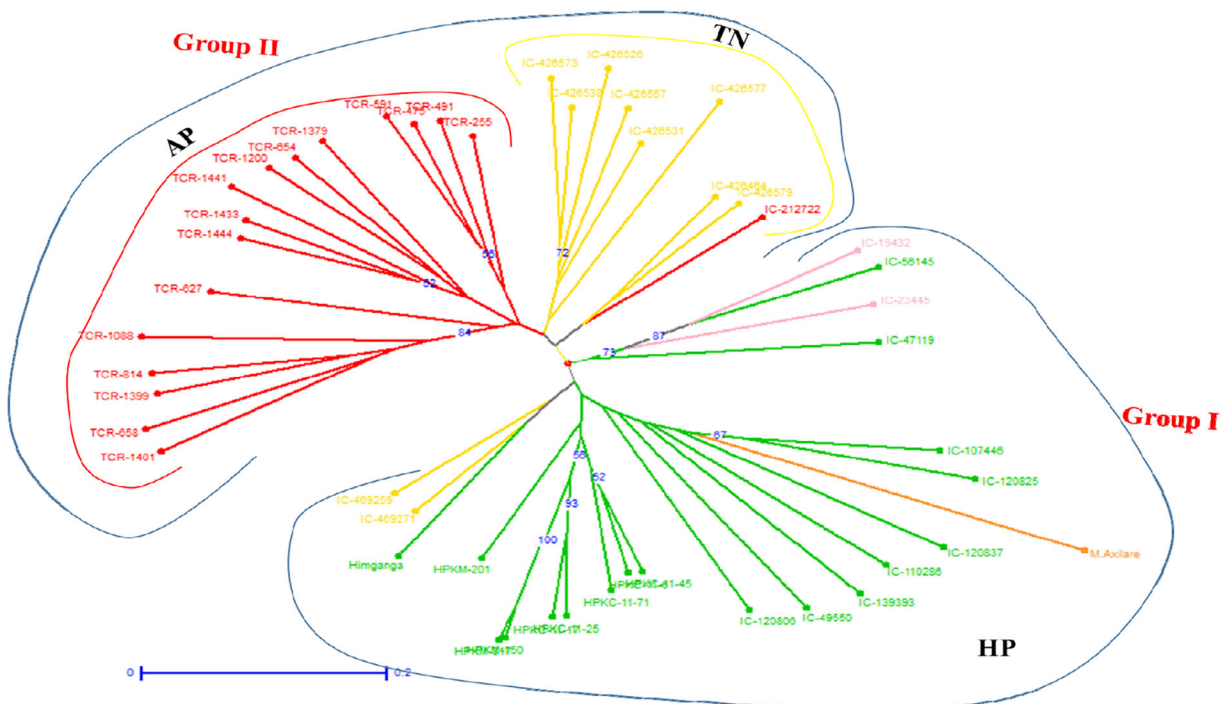


Fig. 1 Radial neighbor-joining tree based on 192 alleles from 47 SSR loci among 48 accessions. Bootstrap values $\geq 50\%$ are shown

Table 2 Summary analysis of molecular variance

Source	df	SS	MS	Est. Var.	%
Among Pops	3	215.236	71.745	4.660	18%
Within Pops	44	904.347	20.553	20.553	82%
Total	47	1119.583		25.213	100%

Cluster-I corresponding to the Himalayan region, showed a greater extent of admixture in its accessions. Thus, admixture shown by STRUCTURE may be due to the mixing of accessions across regions. Structuring of accessions into two clusters with low admixture also points towards isolated and conserved genetic background of the crop species. The results obtained confirm the findings of Sharma et al. (2015b) and Divya Shikha (2015), in which they have reported that cluster of hilly region, was showing greater admixture as compared to the cluster comprising of accessions from central and southern India. Result of analysis of molecular variance (AMOVA) from the data revealed 82% genetic variance resided within the population, whereas 18% among populations (Table 2), thus indicating more genetic variance and allelic richness within populations than among populations. This distribution pattern of genetic variance is due to greater chances of mixing of genomic regions of individuals of same populations. Thus, both the southern and northern regions harbored more genetic variance within populations than among different populations.

The novel SSR markers developed in present study can be enormously useful for future genetic improvement of horsegram in terms of large scale diversity analysis, linkage mapping, QTL detection and association mapping studies. In addition they can be valuable for conducting similar genetic analyses in other related legume crops devoid of SSR marker resources.

Acknowledgements Authors gratefully acknowledge the Department of Biotechnology (DBT), Department of Science and Technology (DST), Government of India and Japanese Society for Promotion of Science (JSPS) for providing financial support to conduct this work.

References

Agarwal M, Shrivastava N, Padh H (2008) Advances in molecular marker techniques and their applications in plant. *Sciences. Plant Cell Rep* 27:617–631

- Allen ON, Allen EK (1981) *The Leguminosae: a source book of characteristics, uses, and nodulation*. Univeristy of Wisconsin Press, Madison, WI, USA, p 446
- Bhardwaj J, Chauhan R, Swarnkar MK, Chahota RK, Singh AK, Shankar R, Yadav SK (2013) Comprehensive transcriptomic study on horsegram (*Macrotyloma uniflorum*): de novo assembly, functional characterization and comparative analysis in relation to drought stress. *BMC Genomics* 14:647
- Blumenthal MJ, Staples IB (1993) Origin, evaluation and use of *Macrotyloma* as forage—a review. *Tropical Grassl* 27:16–29
- Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Human Genet* 32(3):314–331
- Chaitanya DAK, Kumar S, Reddy M, Mukherjee AM, Sumanth NSV, Ramesh DA (2010) Anti urolithiatic activity of *Macrotyloma uniflorum* seed extract on ethylene glycol induced urolithiasis in albino rats. *J Innovative Trends Pharma Sci* 1:216–226
- Chung AM, Staub JE, Chen JF (2006) Molecular phylogeny of Cucumis species as revealed by consensus chloroplast SSR marker length and sequence variation. *Genome* 49:219–229
- Divya Shikha (2015) Analysis of genetic diversity and population structure in horsegram (*Macrotyloma uniflorum*) germplasm using SSR markers. MSc Thesis, P 87. Department of Agricultural Biotechnology, CSK Himachal Pradesh Krishi Vishvavidyalaya, Palampur, India
- Doyle JJ, Doyle JE (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:13–15
- Earl DA, VonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4(2):359–361
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14(8):2611–2620
- Gopalan C, Ramashastry BV, Balasubramanyam SC (1989) Nutritive value of Indian foods. Hyderabad, India: National Institute of Nutrition, ICMR P 156
- Gutierrez MV, VazPatto MC, Huguet T, Cubero JJ, Moreno MT, Torres AM (2005) Cross-species amplification of *Medicago truncatula* microsatellites across three major pulse crops. *Theor Appl Genet* 110:1210–1217
- Hamada HM, Petrino MG, Kakunaga T (1982) A novel repeated element with Z-DNA forming potential is widely found in evolutionarily diverse eukaryotic genomes. *PNAS* 79(21):6465–6469
- He G, Woullard FE, Marong I, Guo BZ (2006) Transferability of soybean SSR markers in peanut (*Arachis hypogaea* L.). *Peanut Sci* 33:22–28
- Jayaraj AP, Tovey FI, Lewin MR, Clark GC (2000) Duodenal ulcer prevalence: experimental evidence for possible role of dietary lipids. *J Gastroenterol Hepatol* 15:610–616
- Kadam SS, Salunkhe DK, Maga JA (1985) Nutritional composition, processing, and utilization of horsegram and moth bean. *Critical Rev Food Sci Nutri* 22:1–26
- Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol Ecol* 16:1099–1106

- Lackey JA (1981) Phaseoleae. In: advances in legume Systematics. (RM Polhill and PH Raven, eds). Royal Botanic Gardens, Kew, UK, pp 301–327
- Mehra A, Upadhyaya M (2013) *Macrotyloma uniflorum* Lam. Traditional crop of kumaun Himalaya and ethnobotanical perspectives. *Int J Agril Food Sci* 3(4):148–150
- Morris JB (2008) *Macrotyloma axillare* And *M. uniflorum*: descriptor analysis, anthocyanin indexes, and potential uses. *Genet Resour Crop Evol* 55:5–8
- Murthy SM, Devaraj V, Anitha RP, Tejavathi DH (2012) Studies on the activities of antioxidant enzymes under induced drought stress in vivo and in vitro plants of *Macrotyloma uniflorum* (lam.) Verdc. *Recent Res Sci Tech* 4:34–37
- Neelam S, Kumar V, Natarajan S, Venkateshwaran K, Pandravada SR (2014) Evaluation and diversity observed in Horsegram (*Macrotyloma uniflorum* (lam) Verdc.) Germplasm from Andhra Pradesh, India. *Int J Plant Res* 4(1):17–22
- Oram RN (1990) Register of Australian herbage plant cultivars. edition. Melbourne, Australia, The Common wealth Scientific and Industrial Research Organisation
- Patel DP, Dabas BS, Sapraandal RS (1995) Evaluation of Horsegram (*Macrotyloma uniflorum* Lam.) germplasm. In: National Bureau of Plant Genetic Resources Publication, New Delhi, India
- Peakall R, Smouse PE (2012) GenA1Ex 6.5: genetic analysis in excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28:2537–2539
- Perrier X, Jacquemoud-Collet JP (2006) DARwin software. <http://darwin.cirad.fr/> Darwin
- Prakash BG, Guled MB, Bhosale AM (2008) Identification of suitable horsegram varieties for northern dry zone of karnataka. *Karnataka J Agril Sci* 21(3):343–345
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Provan J, Powell W, Hollingsworth PM (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends Ecol Evol* 16:142–147
- Rajendrakumar P, Biswal AK, Balachandran SM, Srinivasarao K, Sundaram RM (2007) Simple sequence repeats in organellar genomes of rice: frequency and distribution in genic and intergenic regions. *Bioinformatics* 23:1–4
- Ramesh CK, Rehman A, Prabhakar BT, Vijay ABR, Aditya SJ (2011) Antioxidant potential in sprout vs. seeds of *Vigna radiata* and *Macrotyloma uniflorum*. *J Appl Pharma Sci* 1: 99–103
- Reddy MRK, Rathour R, Kumar N, Katoch P, Sharma TR (2010) Cross-genera legume SSR markers for analysis of genetic diversity in lens species. *Plant Breed* 129:514–518
- Reddy PCO, Sairanganayakulu G, Thippeswamy M, Reddy PS, Reddy MK, Sudhakar C (2008) Identification of stress induced genes from the drought tolerant semi-arid legume crop horsegram (*Macrotyloma uniflorum* (lam.) Verdc.) through analysis of subtracted expressed sequence tags. *Plant Sci* 175: 372–384
- Sharma V, Rana M, Katoch M, Sharma PK, Ghani M, Rana JC, Sharma TR, Chahota RK (2015a) Development of SSR and ILP markers in horsegram (*Macrotyloma uniflorum*), their characterization, cross-transferability and relevance for mapping. *Mol Breed* 35:102
- Sharma V, Sharma TR, Rana JC, and Chahota RK. (2015b) Analysis of Genetic Diversity and Population Structure in Horsegram (*Macrotyloma uniflorum*) Using RAPD and ISSR. *Agril Res*: 1-10
- Smartt J (1990) Grain legumes: evolution and genetic resources. Cambridge University Press, Cambridge, UK
- Sodani SN, Paliwal RV, Jain LK. (2004) Phenotypic stability for seed yield in rain fed horsegram (*Macrotyloma uniflorum*). Proceedings of the National Symposium on Arid legumes for sustainable Agriculture and Trade; Jodhpur, Rajasthan, India: Central Arid Zone Research institute
- Soranzo N, Provan J, Powell W (1999) An example of microsatellite length variation in the mitochondrial genome of conifers. *Genome* 42:158–161
- Sosulski F, Young CG (1979) Field and functional properties of air classified protein and starch fraction from eight legume flours. *J Am Oil Chem Soc* 56:292–295
- Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucl Acids Res* 17: 6463–6471
- Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. *Trends Biotech* 23:48–55
- Yeh FC, Boyle TJB (1997) Population genetic analysis of codominant and dominant markers and quantitative traits. *Belgian J Bot* 129:157
- Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. *Mol Ecol* 11:1–16