Trees (2016) 30:1483–1494
DOI 10.1007/s00468-016-1381-x

ORIGINAL ARTICLE

Genetic diversity and population structure of *Melia azedarach* in North-Western Plains of India

Sapna Thakur[1] · Shruti Choudhary[1] · Amandeep Singh[1] · Kamal Ahmad[1] · Gagan Sharma[1] · Aasim Majeed[1] · Pankaj Bhardwaj[1]

Received: 6 September 2015 / Accepted: 29 February 2016 / Published online: 12 March 2016

Abstract

Key message **Genetic structure among *M. azedarach* populations was detected and two subpopulations were present among them. A significant 'isolation by distance' was found in *M. azedarach* population in North-Western Plains of India.**

Abstract *Melia azedarach* is an important forest tree with pharmaceutical, insecticidal, pesticidal, and commercial significance. It is a good reforestation tree because of its fast growth and drought hardy nature. Genetic variation in a species allows itself to adapt, evolve and respond to environmental stress. It provides the basis for survival of a species and critically influences its evolutionary potential. Assessment of genetic diversity is necessary for improvement and conservation of a species. For this, microsatellite markers are of particular interest given the attributes like co-dominance, reproducibility, hyper variability and abundance throughout the genome. In the present study, we analyzed the genetic diversity and population structure of *M. azedarach*, an ecologically imperative species growing in the North-Western Plains of India. We developed 43 microsatellite markers, of which 20 were subsequently employed for analysis of diversity and population structure among 33 populations encompassing 318 genotypes representing North-Western Plains of India. A moderate level of diversity (Na = 5.1, Ho = 0.506, He = 0.712, I = 1.386) was assessed. The highest value of ΔK estimated using STRUCTURE indicated 2 subpopulations ($K = 2$). AMOVA exhibited 73 % variation within populations and 12 % variation was found among regions. Significant positive correlation between geographical and genetic distance was found ($Rxy = 0.365$, $P = 0.010$). The present study lays a foundation on a better understanding of genetic dynamics of the species and reveals its diversity and population structure in North-Western Plains of India.

Keywords North-Western Plains of India · Genetic diversity · Population structure · Microsatellite marker · *Melia azedarach*

Introduction

Melia azedarach (a member of Meliaceae family) also known as Chinaberry, or Persian lilac is an ecologically imperative species. Morphology of its different cultivars varies throughout the world but typically, it attains 10–15 m height with about 60 cm diameter, though wild trees of Asia can reach up to 40 m of height. It is native to Southeast Asia, especially India and China. Both wild and cultivated forms of *M. azedarach* are found in India, well distributed in arid and semi-arid regions of the North-Western Plains. It is native to several North-Indian states viz., Jammu and Kashmir, Himachal Pradesh, Haryana, Punjab and Uttar Pradesh. It has naturalized other parts of the world also and is regarded as highly invasive species in several tropical and subtropical areas (Tourn et al. 1999). It is cultivated for timber, shade, ornamental purposes, and has remained a source of several folk medicines in India.

Communicated by J. Carlson.

Electronic supplementary material The online version of this article (doi:10.1007/s00468-016-1381-x) contains supplementary material, which is available to authorized users.

✉ Pankaj Bhardwaj
pankajihbt@gmail.com; pankajbhardwaj@cup.ac.in

[1] Molecular Genetics Laboratory, Centre for Biosciences, Central University of Punjab, Bathinda 151001, India

Traditionally different parts of the plant were used for specific purposes i.e., bark as an antidiarrhoeal, stem for treating asthma, leaves for skin diseases and brushing teeth, flower for pyogenic infection, fruit as vermifuge while the whole plant is known to stimulate hair growth (Sharma and Paul 2013). Its medicinal potential has been validated in the scientific literature. Leaf extracts of *M. azedarach* has antibacterial effect against several gram-positive and negative pathogens viz., *Enterococcus faecalis, Staphylococcus aureus*, *Pseudomonas aeruginosa,* and *Escherichia coli* (Adamu et al. 2014). Its extract has been shown to control mosquito, lice and ticks under laboratory conditions (AL-Rubae 2009; Maciel et al. 2006). Meliartenin, a limonoid obtained from the tree, has antifeedant and insecticidal properties that are comparable to Azadirachtin function (Carpinella et al. 2003). Its seed extract is considered a potent rodenticide because of its ability to inhibit folliculogenesis in albino rats (Roop et al. 2005).

Genetic diversity and population structure analysis of wild populations is important for understanding the process of evolution. Besides association studies and tree breeding, knowledge of genetic structure is important in phylogeography, quantitative genetics, conservation and management of species, and provides evidence for gene flow, genetic drift, mutation and migration (Doligez and Joly 1997; Fischer et al. 2000). Restricted gene flow aided by the formation of climatic zones and geographical barriers to breeding enhances genetic differentiation and makes a population prone to random genetic drift (Fischer et al. 2000). Genetic variations in a species are the basis for its evolution and play a vital role in its survival and adaptibility in the changing environment.

Genetic diversity and population structure of this ecologically significant species hitherto has not been investigated. Microsatellite markers were previously not available for this valuable species neither adapted from other species. This scenario underscores the importance of the present study. We analyzed genetic diversity and population structure of *M. azedarach* growing in North-Western Plains of India using genomic-SSR markers. In this study,we also tested if ecogeographical regions influence genetic structure. Novel SSR markers were developed from $(AG)_n$ enriched genomic library and subsequently characterized across a large set of individuals (318) pertaining to 33 natural populations. For genetic diversity and population structure analysis, SSR markers are of particular interest. Given the attributes like co-dominance, high polymorphism, and reproducibility, microsatellite markers serve as an important tool to gather fine-scale ecological information (Gupta and Varshney 2000). The present study will help in better understanding of genetic dynamics and to realize the adaptability of the species and reveal if ecogeographical viz., semi-arid and arid regions have any effect on diversity or the genetic structure of North-Western Plain's populations.

Materials and methods

Plant material

Melia azedarach samples were collected from a total of 33 geographically isolated wild populations from arid and semi-arid zones representing the plains of North-Western India. Sampling locations included the states of Delhi, Punjab, Rajasthan and Haryana. These areas, characterized by sparsely populated vegetation cover, have been categorized under arid or semi-arid climatic zones based on the insufficient amount of annual rainfall. Complete information regarding sampling locations is given in Table 1. Young leaves were kept inside multiple layers of blotting sheets for transportation to the laboratory. Genomic DNA was isolated by CTAB method (Doyle 1990) with minor modifications. The quality and quantity of the extracted DNA were estimated on 0.8 % agarose gel and Nano Drop 2000 Spectrophotometer. A total of 355 samples were collected, out of which 318 were selected for further analysis based on the quality of DNA. Among analyzed genotypes, 243 (26 populations) represented the semi-arid zone, while 75 (7 populations) were from the arid zone (Fig. 1).

Microsatellite-enriched library construction

A microsatellite $(AG)_n$ enriched genomic library was constructed from a single *M. azedarach* sample using biotin-streptavidin capture method as described by Bhardwaj et al. (2013) with some modifications. For this, genomic DNA was digested with restriction endonucleases *Eco*R1 and *MSe*1. The restriction digestion was confirmed by 1.2 % agarose gel and size selection in the range of 200–1000 bp was done subsequently. DNA fragments in the desired range were purified from gel (NucleoSpin Gel and PCR Clean-up, MN) and proceeded for ligation with sticky adapters, *Eco*R1 (5′-CTCGTAGACTGCGTACC-3′/5′-AATTGGTACGCAGTC-TAC-3′) and *MSe*1 (5′-GACGATGAGTCCTGAG-3′/5′-TACTCAGGACTCAT-3′). Ligated fragments were later amplified using the corresponding primers (5′-GACTGC GTACCAATTC-3′/5′-GATGAGTCCTGAGTAA-3′) and purified (NucleoSpin Gel and PCR Clean-up, MN) before downstream processing. The purified product was heat denatured and hybridized to biotinylated $(AG)_{10}$ oligonucleotides using 6X SSC for 4 h at 65 °C. SSR region containing fragments hybridized to biotinylated $(AG)_{10}$ oligonucelotides were selected by streptavidin-coated paramagnetic beads (New England Biolabs, Inc., NEB, USA) and non-hybridized DNA

Table 1 Geographical locations of *M. azedarach* populations from North-Western Plains of India

Population	Sample size	Latitude (N)	Longitude (E)
Sirsa (SIR)	8	29.44	74.67
Fatehabad (FAT)	7	29.52	75.42
Hisar (HIS)	9	29.18	75.71
Jind (JID)	8	29.31	76.32
Kaithal (KAT)	8	29.62	76.40
Kurukshetra (KAS)	8	29.97	76.83
Karnal (KAR)	13	29.69	76.99
Sonipat (SOP)	13	29.40	76.97
Delhi cantonment (DEC)	9	28.59	77.16
Budha (BUG)	8	29.00	77.01
Shankar road(SHR)	10	28.64	77.18
Badarpur (BAB)	10	28.49	77.30
JNU campus (JNU)	7	28.55	77.17
Tikari Border (TIB)	9	28.69	76.96
Kapashera (KPB)	7	28.52	77.08
Gurgaon (GUR)	9	28.45	77.03
Ring Road (RGR)	9	28.48	77.07
Noida (NO)	9	28.58	77.30
Barnala (BNL)	10	30.36	75.54
Sangrur (SGR)	9	30.25	75.83
Sunam (SMM)	10	30.12	75.82
Bhawanigargh (BIG)	10	30.26	76.04
Patiala (PTL)	10	30.33	76.39
Samana (SMN)	9	30.24	75.84
Bathinda (BTI)	17	30.14	74.79
Mansa (MAN)	7	30.00	75.39
Jaisalmer (JAS)	7	26.87	70.36
Hanumangarh (HAN)	10	29.61	74.29
Sri ganganagar (SRI)	13	29.92	73.86
Ghasana (GHA)	10	29.02	73.08
Raisinghnagar (RAI)	9	29.54	73.45
Padampur (PAD)	17	29.71	73.63
Lunkaransar (LAK)	9	28.50	73.76
Total	318		

was removed by washing with different concentrations of SSC buffer (2× SSC; 0.1 % SDS,1× SSC; 0.1 % SDS) at room temperature and 60 °C, consecutively. Hybridized single stranded DNA was later eluted with nuclease free water, amplified using adapter primers and remaining product was stored at −20 °C for additional usage. Amplified-enriched fragments were ligated to pGEM-T Easy (Promega) vector, later competent *E. coli* were transformed with these recombinant plasmids. Clones positive for insert were confirmed by interrupted β-galactosidase gene and secondary enrichment through PCR amplification. Positive clones of secondary enrichment were grown overnight in LB broth supplemented with 100 μg/ml ampicillin and subsequently recombinant plasmids from these clones were isolated using NucleoSpin Plasmid miniprep kit (Macherey–Nagel) and sequenced with ABI 3730 xl DNA Analyzer using BigDye terminator cycle sequencing kit v3.1 (Applied Biosystems) as per the manufacturer's procedure.

Primer designing and genotyping

SSRs containing sequences were identified using SSR Identification Tool (SSRIT) (Temnykh et al. 2001). Primers were designed from region flanking SSRs using Primer3 web version 4.0.0 software (Untergasser et al. 2012) after removing plasmid chimeras from sequenced fragments. Initially, all the primers were amplified using pooled genomic DNA and primer pairs showing good

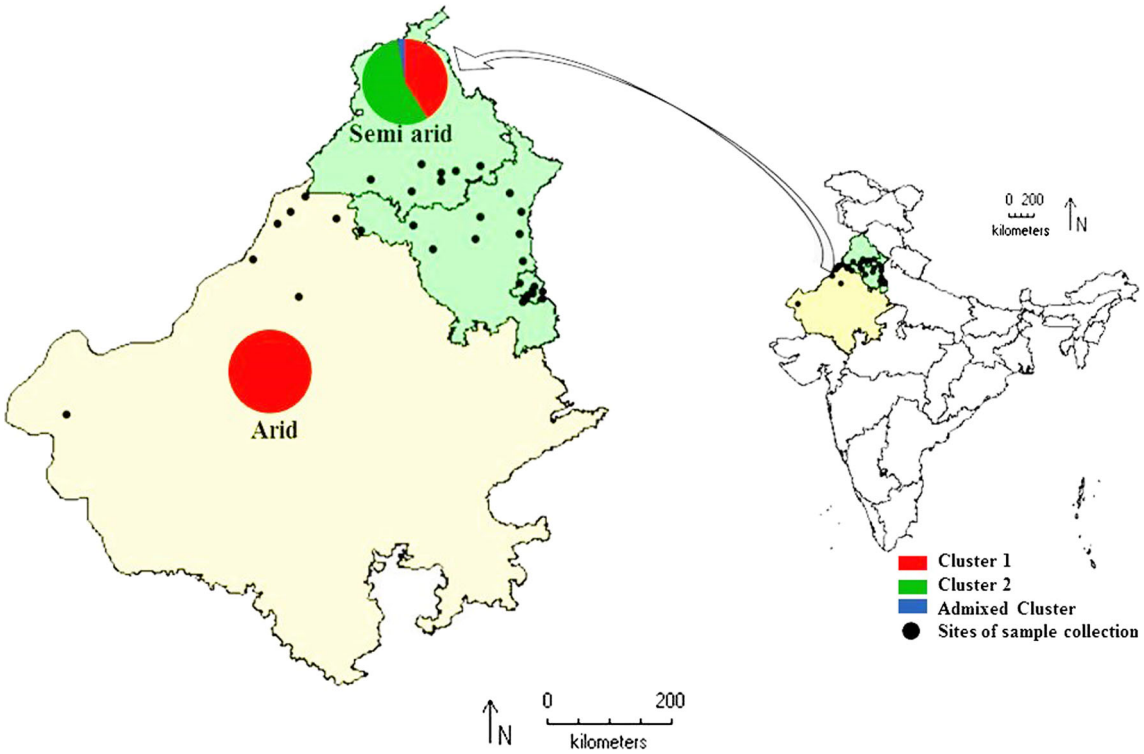

Fig. 1 Sites of sample collection (*solid circles*) in North-Western Plains of India. *Pie charts* represent percentage of two original genetic stocks and admixed population (see Table 1 for full details of locations)

amplification were selected for genotyping. For genotyping, PCR reactions were carried out in a total volume of 10 µl, containing genomic DNA (25 ng), *Taq* DNA polymerase (0.3 U), PCR buffer (1×), dNTPs (2.5 mM, each), forward and reverse primer (5 pM) in T100 Thermal Cycler (BIORAD). PCR profile comprised of an initial denaturation step at 94 °C for 3 min followed by 35 cycles of denaturation at 94 °C for 1 min annealing at appropriate T_m (Table 2) for 1 min and, 72 °C for 1 min for elongation and a final elongation at 72 °C for 8 min.

Amplification reactions were diluted with equal volume of denaturing dye and subsequently heat denatured at 94 °C for 5 min, snap-cooled and kept on ice till loading for PAGE. The SSR-PCR products were resolved or quantified by electrophoresis on 6 % urea polyacrylamide gel in 1X TBE buffer at 60 W for desired time depending on the expected size of the amplified DNA fragment. Afterward, the gel was subjected to 2–3 initial washing steps prior to staining with 2 % silver nitrate for 45 min. Then followed counterstaining with developer solution (30 % sodium carbonate and 0.01 % sodium thiosulfate) and finally the gel was immediately fixed in 10 % glacial acetic acid solution in water for visualization of bands. 100 bp ladder was used to estimate the fragment size.

Data analysis

The allelic data matrix was generated for each polymorphic primer pair and subsequently CONVERT (version 1.31) a window based program, was used to build input files for the various population genetics software used in the present study (Glaubitz 2004). Genetic diversity estimation for the parameters viz., number of alleles, observed Heterozygosity (H_o), expected Heterozygosity (H_e), Shannon's information index (I), F-statistics and gene flow (Nm) was done using POPGENE software version 1.32 (Yeh et al. 1999). Polymorphic Information Content (PIC) and deviation from Hardy–Weinberg equilibrium (HWE) for each primer pair were calculated using Cervus (Version 3.0.7) (Kalinowski et al. 2007). Population differentiation or discriminatory power of developed primer pairs was investigated using WHICHLOCI v.1.0 (Banks et al. 2003). The relative discriminatory score was estimated using allele frequencies by conducting 10,000 iterations over a constant sample size of populations ($N = 500$) with minimum 95 % assignment accuracy.

Population wise diversity, Principal Coordinates Analysis (PCoA), Analysis of Molecular Variance (AMOVA) and Mantel's test to measure the correlation between geographical and genetic distance was applied using GeneAlEx version 6.5 (Peakall and Smouse 2012). The

Table 2 Features and evaluation details of 20 microsatellite markers

Locus name	Primer sequence	Repeat motif	Ann temp (°C)	No. of alleles	Heterozygosity		PIC	I	Approx. size range (bp)	Fis	Fit	Fst	Nm	Accession no.
					H_0	H_e								
MAZ-1	F5′AGGAAGAATGCCGCTGACTA R5′GGAAATGAAAACCGAAAGCA	$(CT)_5$	59	2	0.469	0.530	0.419	0.811	214–216	−0.273	0.055	0.258	0.719	KJ996073
MAZ-2	F5′GGGGAAGAGGGTCCAAGTT R5′TGAAAAACAATTATGTGATTTAGAAGA	$(CT)_{18}$	56	6	0.104*	0.831	0.807	1.859	264–280	0.813	0.878	0.349	0.466	KJ996074
MAZ-3	F5′ACAATTGGGGAAGTCTGTGC R5′GCGCGAACTTCACTCTCTCT	$(AG)_{11},(AG)_6$	57	4	0.286*	0.706	0.653	1.316	388–392	0.429	0.606	0.311	0.554	KJ996075
MAZ-5	F5′TCGTCATAACGCGAGAGTCA R5′CTTCGGCTTCTTCTGATTGG	$(CT)_{21}$	59	4	0.277*	0.786	0.757	1.666	140–160	0.449	0.634	0.335	0.497	KJ996076
MAZ-7	F5′TTCTGGAAAACCAACCAACC R5′CCTGAGTAAAAGCTACTCTGAATGG	$(CT)_{16}$	58	5	0.733*	0.757	0.715	1.488	176–186	−0.222	0.035	0.210	0.940	KJ996078
MAZ-10	F5′TTAGGCATGGATCACAGAAAA R5′CAGATTGCTGCAAATTGGTAAA	$(AG)_{27}$	57	6	0.349*	0.791	0.759	1.734	94–112	0.261	0.593	0.449	0.307	KJ996079
MAZ-12	F5′GGAAAGAGAGAAATGGTGCAA R5′GACGCGACTTGAACTCAAAA	$(AG)_5,(AG)_{10}$	58	7	0.704*	0.835	0.812	1.863	120–144	−0.017	0.171	0.185	1.101	KJ996080
MAZ-13	F5′GGGTGTCTTTGGACGTGATT R5′CAACGCATGAAAGAGGAAAA	$(AG)_9,(AG)_{15}$	58	3	0.440*	0.595	0.518	0.984	180–190	0.078	0.228	0.163	1.285	KJ996081
MAZ-15	F5′AGCTCGAATCCATCCAGAAC R5′CCCTCTCTGTCTCTGACGCTAT	$(AG)_8$, $(AG)_4$, $(AG)_4$	58	4	0.481*	0.598	0.527	0.997	186–200	−0.057	0.143	0.189	1.075	KJ996082
MAZ-17	F5′AATTGGTCTGATTTGACTCTCTCT R5′CCATGCCTCTATCTTGCTCTC	$(CT)_{17}$	56	4	0.613*	0.590	0.510	0.972	188–198	−0.180	−0.062	0.100	2.251	KJ996083
MAZ-18	F5′TGGGCTACAAAATGAGAAAGG R5′TCAGGGGTATTCATTCATAGGG	$(AG)_{13},(AG)_{13}$	58	4	0.525*	0.701	0.652	1.297	146–160	0.074	0.220	0.158	1.335	KJ996084
MAZ-22	F5′TGAGTCCTGAGTAATGAAAAACAA R5′GTCCAAGTTGGCTCCTGGT	$(AG)_8$, $(AG)_7$	56	7	0.069*	0.719	0.679	1.461	196–212	0.803	0.905	0.515	0.236	KJ996085
MAZ-26	F5′TCCAAAGGCCACAATAACCA R5′TCACCAATGGAGAGGAAGCT	$(AG)_{16}$	59	5	0.462*	0.720	0.669	1.324	172–190	−0.043	0.338	0.365	0.435	KJ996086
MAZ-29	F5′CCAGGCAGATAACGCAGAGA R5′GAGCCTAAACCCCACTTCCT	$(AG)_{11}$, $(AG)_5$	59	3	0.302*	0.632	0.553	1.039	170–180	0.203	0.520	0.397	0.379	KJ996088
MAZ-30	F5′GGAAAGAAGAAGGTGGGTTCC R5′ACAAATAGAGGCGGCTGAGA	$(CT)_{14}$	58	5	0.654*	0.601	0.525	0.994	185–200	−0.200	−0.072	0.107	2.088	KJ996089
MAZ-35	F5′CCCCCTATCAAGACAAGCAA R5′GCTTGTCCTTTACTTTTGTTTTCC	$(AG)_{21}$	58	4	0.343*	0.715	0.662	1.313	100–120	0.330	0.505	0.260	0.710	KJ996090

Table 2 continued

Locus name	Primer sequence	Repeat motif	Ann temp (°C)	No. of alleles	Heterozygosity		PIC	I	Approx. size range (bp)	Fis	Fit	Fst	Nm	Accession no.
					H_0	H_e								
MAZ-37	F5′CCGTCAGGGACGTAATCCTA R5′CGGCTTCTTCTGATTGGTTC	$(CT)_{24}$	57	8	0.827*	0.847	0.827	1.953	176–220	−0.138	0.028	0.146	1.459	KJ996091
MAZ-38	F5′ATTCCCACCTTTCCCATTTG R5′TTGCTTGAGTGCTTTTGTGG	$(CT)_{15}$	59	5	0.862*	0.706	0.651	1.312	180–200	−0.511	−0.229	0.187	1.090	KJ996092
MAZ-39	F5′AGGTCTTCGGCTTCTTCTGA R5′CCATCCCCGCCAAAATTCAA	$(AG)_{20}$	62	8	0.739*	0.799	0.771	1.768	150–190	−0.162	0.060	0.191	1.060	KJ996093
MAZ-41	F5′ACAACGCATGAAAGAGGAAAAG R5′CTTTGGACGTGATTGTGGGTATA	$(CT)_9(CT)_{13}$	59	5	0.884*	0.786	0.751	1.571	158–180	−0.322	−0.139	0.138	1.555	KJ996095
Mean				5.1	0.506	0.712	0.661	1.386		0.066	0.271	0.251	0.977	
SD				1.9974	0.241	0.094	0.112	0.347		0.080	0.075	0.026	0.127	

H_o observed heterozygosity, H_e expected heterozygosity, *PIC* polymorphic information content, *I* Shannon's information index, *bp* base pairs; significant deviation from Hardy–Weinberg equilibrium at * $p = 0.001$

phylogenetic relationship among populations was illustrated from DARwin version 6 and STRUCTURE version 2.3.4 (Pritchard et al. 2000). Based on multi-locus Bayesian analysis, the hidden population clusters were distinguished. For estimation of optimum number of populations (*K*), a simulation was conducted using parameters; *K*(1–33) with a random start and 6 iterations, a burn-in period of 100,000 and 100,000 number of Markov Chain Monte Carlo (MCMC) repeats after burn-in (Evanno et al. 2005).The number of distinct population clusters (K) was determined using Structure Harvester, a web-based program for visualizing STRUCTURE output by implementing the Evanno method (Earl 2012). Population-level Mantel's test based on the relationship between pairwise *Fst* and geographical distance matrix was carried out by assigning the significance to any correlation between genetic and geographic distances. Distance matrices with matching entries were compared with a null hypothesis of no correlation.

Results

Microsatellite-enriched library statistics

Out of 1912 recombinant clones from the microsatellite $(AG)_n$ enriched genomic library, 573 (30 %) positive clones were identified, 339 (59 %) of which possessed SSRs with a range of 10–100bps. Of these 339 sequences, 48 were non-redundant (NR) and 43 out of which were subsequently utilized for primer designing. Developed primer pairs were used to amplify a pooled genomic DNA of 318 genotypes whereupon 20 (46.52 %) primer pairs were found suitable for further analysis. The polymorphic loci ranged in size from 5 to 24 di-nucleotide repeats, out of which 8 were interrupted, and 12 were simple repeats.

Genetic diversity

The polymorphic potential of 20 primer pairs was evaluated in a test array of 318 genotypes comprising 33 different populations from North-Western Plains of India. Features and evaluation details of 20 successful primers pairs are given in Table 2. A total of 102 alleles (mean 5.1) could be amplified by tested primer pairs in the range of 2–8 alleles per locus. 19 out of 20 (95 %) tested primer pairs showed significant HWE deviations. The observed heterozygosity (Ho) varied from 0.069 to 0.884 with a mean of 0.506, expected heterozygosity (He) varied from 0.530 to 0.847 with a mean of 0.712. PIC varied in the range of 0.419 to 0.827 over all loci with a mean of 0.661 also 19 loci (95 %) exhibited PIC more than 0.500. Shannon's information index (I) varied from 0.811 to 1.953

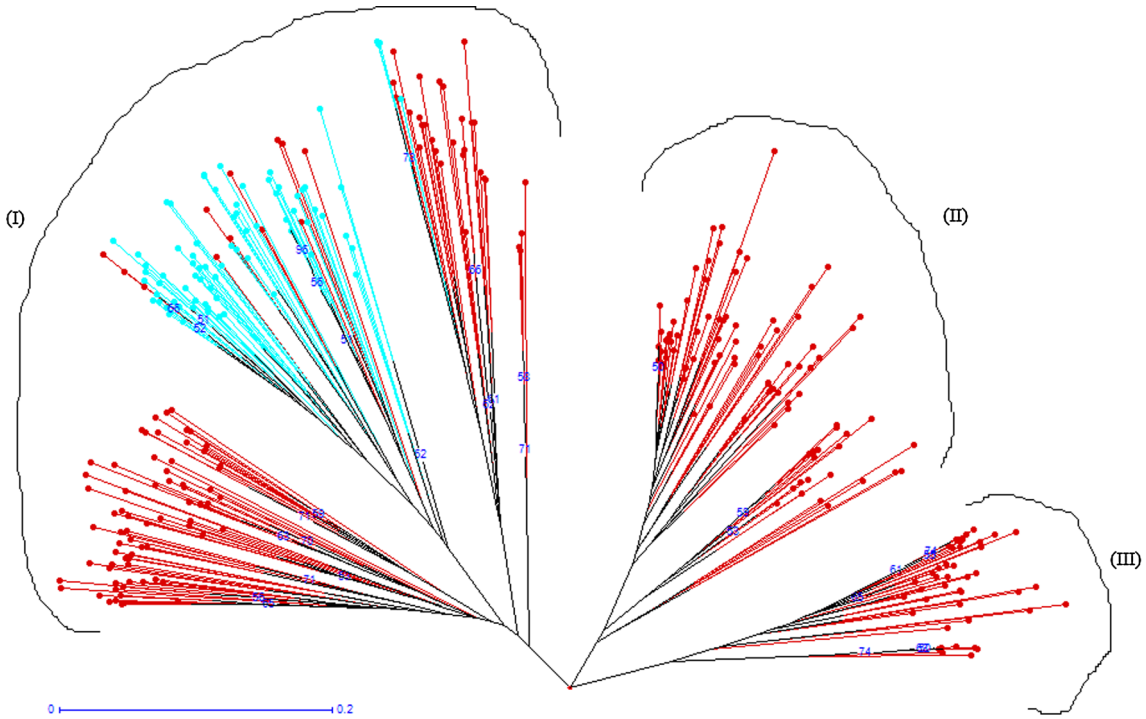

Fig. 2 Unrooted neighbor joining (UNJ) tree constructed from DARwin version 6; each branch represents single individual

with a mean of 1.386. Approximate size of amplified fragments in all test arrays ranged from 90 to 392 bp. *Fis* varied from −0.511 to 0.813 with a mean of 0.066, *Fit* varied from −0.229 to 0.905 with a mean of 0.271, *Fst* varied from 1.00 to 0.515 with a mean of 0.251. Gene flow varied from 0.236 to 2.251 with a mean of 0.977. An empirical analysis based method WHICHLOCI determines the minimum number of primer pairs and their relative efficiency for accurate population assignment. WHICHLOCI analysis revealed a minimum of 20 primer pairs are sufficient for 95.26 % assignment accuracy to 33 populations with a variance of 4.23. The discriminatory score varied from 0.09 to 0.53 for 20 tested loci using minimum 95 % assignment accuracy over 10,000 datasets (Supplementary Table 1).

UNJ grouped all the genotypes in 3 major clusters (I, II and III) (Fig. 2). Cluster I, contained populations from both the arid and semi-arid regions but the arid region populations (JAS, HAN, SRI, GHA, RAI, PAD, and LAK) were clustered discretely in a well separated sub-cluster from the semi-arid populations. Genotypes from the semi-arid regions formed two more distinct clusters (II and III) in which grouping was largely by geographic distribution. PCoA is an ordination technique that is complementary to UNJ clustering. PCoA supported hierarchical clustering based UNJ tree and first three axes created in it on the basis of *FstP* matrix-explained 66.78 % variation. Three major distinct groups were identified in PCoA and further populations from the semi-arid zone were intermixed while arid populations formed a discrete group (Fig. 3). The relative distinctiveness of sampled populations was determined by assigning individuals on the basis of genetically homogeneous groups rather than on geographical locations of sampling.

Population structure

Hierarchical AMOVA revealed variation and the differentiation among semi-arid and arid regions. Within populations, 73 % variation was accumulated while between populations it was 15 %. A total of 12 % variation was attributed to regions (Table 3). Genotypes from semi-arid region exhibited an average of 5.1 number of alleles (Na) while in the arid region it was 4.45. The observed heterozygosity (Ho) in the semi-arid region was found to be 0.520 and expected heterozygosity (He) was 0.698. In arid region observed heterozygosity, (Ho) was 0.458 and expected heterozygosity (He) was 0.594. Gene flow between regions was found to be 0.679. *F*-statistics of regions indicated a significant genetic variation between regions ($Frt = 0.121$, $p = 0.001$). A significantly great degree of population divergence ($Fst = 0.269$, $p = 0.001$) was revealed (Table 4).

In Bayesian based STRUCTURE analysis, the highest value of ΔK depicted 2 subpopulations or genetically distinct stocks that were present among 33 sampled

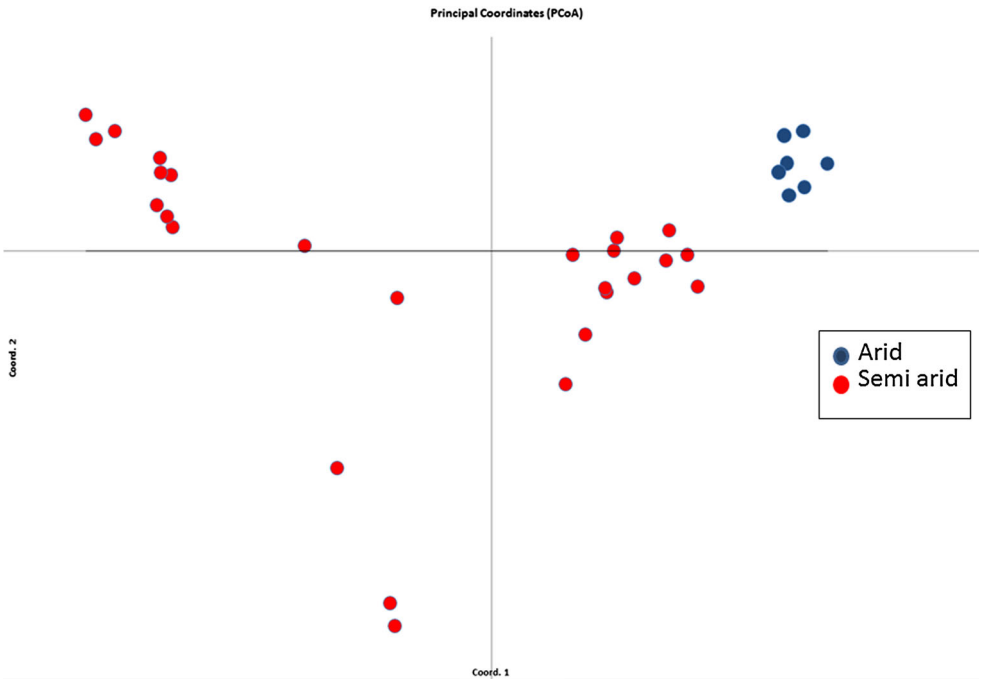

Fig. 3 Principle coordinate analysis (PCoA) of 33 populations from North-Western Plains of India showing the regional differentiation

populations (Fig. 4). Based on an arbitrary value of membership coefficient ($Q \geq 80$ %), first genetic cluster occupied 173 individuals (54.4 %), second genetic cluster occupied 138 individuals (43.4 %) and only 7 individuals (2.2 %) were of mixed origin (Fig. 5). Out of 243 individuals analysed from semi-arid region, 136 (55.9 %) were confined to one cluster, 100 (41.1 %) were confined to second cluster and remaining individuals were of mixed origin however arid region exclusively contained single type of genetic stock. The seven individuals of mixed origin having membership coefficient ($Q \leq 80$ %) were from populations JID (3), HIS (2), KAS (1) and SOP (1).

Genetic distance in terms of *FstP* was compared corresponding geographic distance matrix. *FstP* values varied from 0.006 (SOP, JID) to 0.430 (HAN, PTL) also geographic distance for these two population pairs is 136 and 217 km, respectively. Scatter plots of *FstP* and geographical distance (Fig. 6) showed a significant positive correlation (Rxy = 0.365, $P = 0.010$; $r^2 = 0.133$).

Discussion

Genomic-SSR markers

Melia azedarach is an important forest tree known for its innumerable biological benefits such as antiviral, anthelminthic, and antibacterial, etc. It also holds cultural significance in traditional medicines and incense manufacturing. It is very important to investigate and estimate its genetic diversity in wild to prioritize strategies for its management and apt utilization. There were only a few studies previously regarding the genetic diversity of *M. azedarach*, typically using RAPD marker. Genetic variation of *M. azedarach* was assessed using five RAPD markers on 120 parent trees in community forest of West Java where it is an introduced species. This study reported a moderate level of genetic variation (16–19 %) in sampled populations (Yulianti et al. 2011). The basic limitations of RAPD are irreproducibility and its inability to distinguish heterozygosity.

In the present study 43 informative microsatellites markers were developed from $(AG)_n$ enriched genomic library that subsequently revealed the level of genetic variation across 33 populations representing North-Western plains of India. Biotin-streptavidin capture method was successfully employed for rapid microsatellite marker development. Out of 573 plasmids sequenced, 59 % fragments enclosed SSRs in the range of 10–100 bp. A huge number of SSRs containing fragments could not be used effectively for primer designing due to insufficient length of sequences flanking SSRs region. 20 polymorphic primer pairs were evaluated across 318 genotypes exhibiting a substantial number of the allele (5.1 per locus) with a maximum of 8 alleles for the locus MAZ-37 (Table 2). The mean number of alleles per locus and mean expected heterozygosity (Na = 5.1, He = 0.712) was greater as compared to *M. azedarach* populations of West Java (Na = 1.52, He = 0.171) analyzed using RAPD markers (Yulianti et al. 2011). This huge difference could be

Table 3 Summary of analysis of variance (AMOVA) for *M. azedarach* populations

Source of variation	*df*	Sum of squares	Variance component	Percentage of variation (%)
Among regions	1	248.028	0.943	12
Among pops	31	874.380	1.146	15
Within pops	285	1790.818	0.611	73

df Degrees of freedom

Table 4 Diversity parameters of eco-geographical regions

Region	Genotypes	Na	Ho	He	Nm	Fst
Semi-arid	243	5.1	0.520	0.698		
Arid	75	4.45	0.458	0.594		
Total	318	–	–	–	0.679	0.269*

Na number of alleles, *Ho* observed heterozygosity, *He* expected heterozygosity, *Nm* gene flow, *Fst* the degree of differentiation within a population among regions

* Significance level was $P = 0.001$

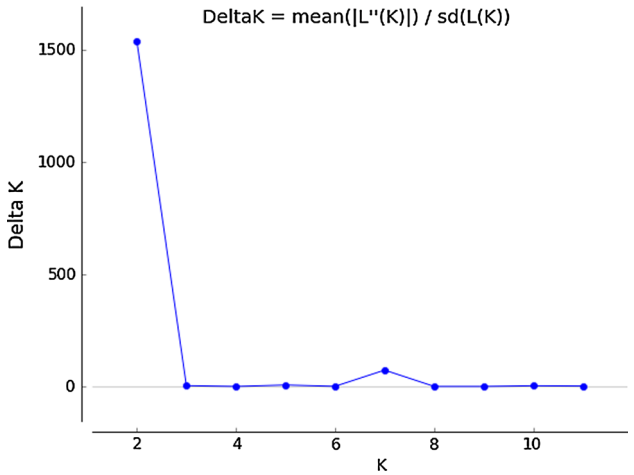

Fig. 4 Estimation of optimum K using the second order statistics (ΔK) given by Evanno et al. (2005)

attributed to the polymorphic nature of microsatellite markers and larger sample size characterized in the present study. On the other hand, studies in other members of Meliaceae family also reported similar levels of genetic diversity in the populations of *Swietenia macrophylla* (He = 0.66) (Novick et al. 2003) and *Cedrelaodorata* (He = 0.820) (Hernández et al. 2008), respectively using polymorphic microsatellites. This high level of genetic diversity in Meliaceae may be related to the occurrence of cross-pollination by insect pollinators (Syamsuwida et al. 2012; Du et al. 2012). With the exception of MAZ-1, all developed primer pairs were highly polymorphism (PIC $\geq$ 0.5) and showed significant ($p = 0.001$) deviation from Hardy–Weinberg equilibrium that is mostly due to loss of heterozygosity (Ho = 0.506, He = 0.712) in the sampled populations (Table 2). The observed deviation from Hardy–Weinberg equilibrium could be attributed to selection, non-random mating or presence of population substructures. WHICHLOCI analysis revealed that all the 20 developed primer pairs are sufficient to meet accuracy specifications using minimum 95 % assignment accuracy over 10,000 datasets.

Population structure

The hierarchical Analysis of Molecular Variance (AMOVA) revealed a large proportion of variation (73 %) within the population. This is in agreement with various studies in which larger variability is present within the population, and proportion of variation among populations and regions is small (Scotti-Saintagne et al. 2013; Du et al. 2012; Yulianti et al. 2011). A significant proportion of the variation observed, was found to be distributed among the populations (15 %) with 12 % variation (Table 3). A moderate gene flow (0.679) and high genetic differentiation based on average *FstP* (0.269) could be due to the seed dispersal mechanism of this species. Since the populations under the study are wild and native to North-Western Plains of India, there is a little scope for human mediated seed dispersal of this species, if any. Only a limited species of birds contribute to seed dispersal of *M. azedarach* due to its large fruit size and presence of secondary compounds that make them indigestible. Furthermore, it is reported that birds drop seeds under the canopy of parent tree or in the proximate zones. Small distance seed dispersal is reported to be carried out by mammals like deer in India (Voigt et al. 2011). It is possible that seed dispersal process has strengthened the population substructure of this species in North-Western Plains of India. The observed heterozygosity in the semi-arid region (Ho = 0.520) was found to be more than the arid region (Ho = 0.458) which ascertains semi-arid region is richer in genetic diversity. The arid climate is also reported to limit its invasiveness in Northern Cape. The frequency of occurrence of *M. azedarach* significantly decreased in the arid region (0.9 %) as compared to the semi-arid region (11.5 %) (Henderson 1991). *Melia volkensii* (Gurke.) is also reported to exhibit an eco-geographical association with genetic variation (Runo et al. 2005). Furthermore, STRUCTURE analysis

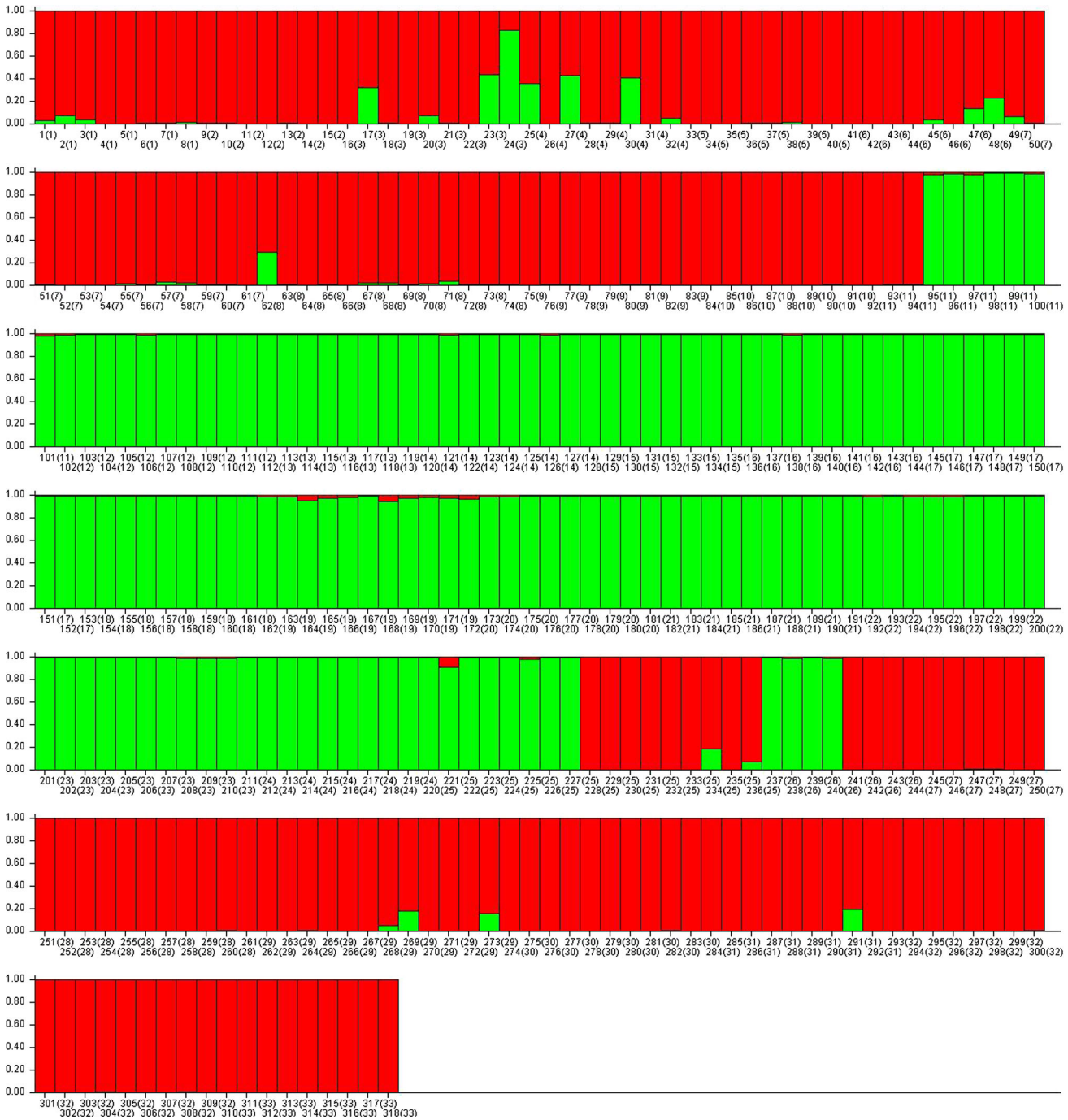

Fig. 5 Bar plot given by STRUCTURE with membership coefficient (Q) of individuals on *y* axis. *Each color* represents different genetic stock

also detected the presence of two genetic subpopulations (clusters) in tested genotypes. Populations from semi-arid region showed presence of two clusters consisting of original genetic stocks along with admixed cluster, while popualtions of arid region were clearly assigned to single subpopulation at optimum ΔK ($K = 2$). The semi-arid region could be the focal point of distribution of *M. azedarach* in North-Western plains of India. The climatic zonation had been associated with genetic differences in many tree species such as poplar (Du et al. 2012); *Elymus* (Bockelmann et al. 2003).

Grouping of populations on the basis of genetic similarity by PCoA and UNJ resulted in the formation of three clusters among them. In both cases, populations from arid region formed a well-separated cluster while the populations from semi-arid region formed two mixed clusters.

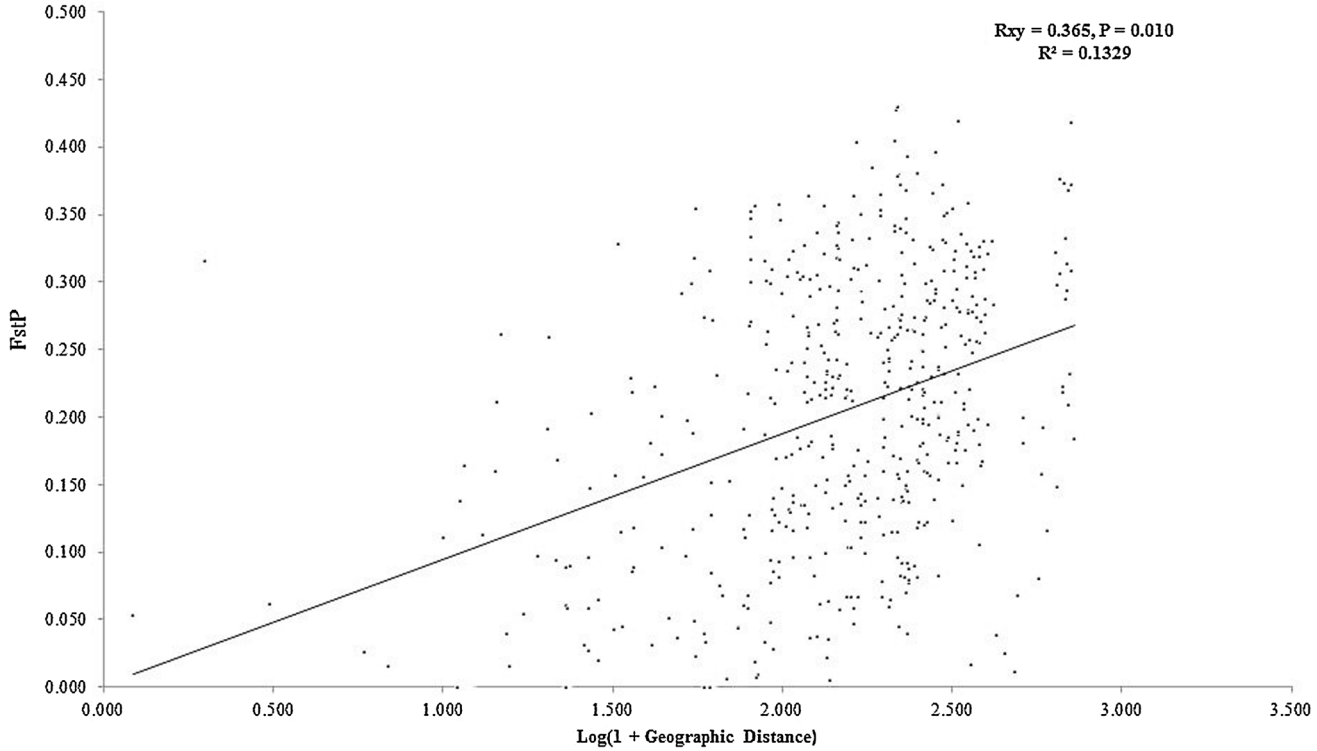

Fig. 6 Relationship between genetic and geographic distances for 33 *M. azedarach* populations. Genetic distance is represented by pairwise F_{st} and geographic distance is given in Kilometers

Though gene flow between these two ecogeographical regions appears possible, it is rather limited by topographic factors and seed dispersal mechanism. The populations of arid region chiefly encompass the Indian Thar Desert which possesses different edaphic and climatic patterns. These results are in accordance with hierarchical AMOVA in which there is a significant splitting of populations on the basis of the climatic regions. Seed dispersal mechanisms and environmental factors could be influencing the gene flow between geographically distant populations.

Isolation by distance (IBD)

A significant differentiation or formation of genetic structure among populations could be an important step for the evolution of a species. Inbreeding, the rate of pollen and seed dispersal, limited gene flow due to geographic barriers and adaptive response of a species to particular climatic conditions can give rise to such variation which over the time accumulates to exhibit speciation (Hamrick et al. 1992). We observed a significant 'isolation by distance' (Rxy $= 0.365$, $P = 0.010$; $r^2 = 0.133$) which also explains genetic differentiation in the population of *M. azedarach*. Both the distance matrices showed congruence i.e. genetic distance (Supplementary Table 2) between populations pairs, that increased with geographic distance (JNU-PTL; 402 km, 0.404; PTL-SMN, 54 km, 0.023) and are in accordance with 'isolation by distance' model. Such correlation indicates gene flow between the *M. azedarach* populations of North-Western plains of India to be dependent on geographic distance. Populations of *S. macrophylla* and *C. lilloi* having wider geographic distributions showed a comparable trend of higher genetic differentiation in independent studies (Lowe et al. 2003; Inza et al. 2012).

Melia azedarach populations of North-Western plains of India exhibited the presence of two types of original genetic stock between them. Gene flow between arid and semi-arid region though possible was hindered due to seed dispersal mechanism and climatic factors. Consequently genetic differentiation among regions was established. Results and data produced in the present study have provided insight into the phylogenetic distribution and evolutionary potential of *M. azedarach* in the North-Western plains of India.

Author contribution statement PB conceived the study, participated in designing coordination data analysis, interpretation, reviewed and improved the manuscript. ST and SC carried out marker development, genotyping, statistical analysis and drafting of the manuscript. AS, GS, KA, AM collected samples, isolated DNA and participated in genotyping. All authors have read and approved the final manuscript.

Acknowledgments Authors thank Vice Chancellor, Central University of Punjab, Bathinda for providing the necessary facilities to carry out the present work. ST and SC acknowledge the fellowship received from ICMR towards PhD.

Data archiving statement The sequence data generated for this study are available under accession numbers from KJ996073 to KJ996095 at NCBI (http://ncbi.nlm.nih.gov). Accession numbers are included in Table 2.

Compliance with ethical standards

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

References

Adamu M, Naidoo V, Eloff JN (2014) The antibacterial activity, antioxidant activity and selectivity index of leaf extracts of thirteen South African tree species used in ethnoveterinary medicine to treat helminth infections. BMC Vet Res 10:52

AL-Rubae AY (2009) The potential uses of *Melia azedarach* L. as pesticidal and medicinal plant review. Am Eurasian J Sustain Agric 3:185–194

Banks MA, Eichert W, Olsen JB (2003) Which genetic loci have greater population assignment power? Bioinformatics 19:1436–1438

Bhardwaj P, Kumar R, Sharma H, Tewari R, Ahuja PS, Sharma RK (2013) Development and utilization of genomic and genic microsatellite markers in Assam tea (*Camellia assamica* ssp. *assamica*) and related Camellia species. Plant Breed 132:748–763

Bockelmann AC, Reusch T, Bijlsma R, Bakker J (2003) Habitat differentiation vs. isolation-by-distance: the genetic population structure of Elymus athericus in European salt marshes. Mol Ecol 12:505–515

Carpinella MC, Defago MT, Valladares G, Palacios SM (2003) Antifeedant and insecticide properties of a limonoid from *Melia azedarach* (Meliaceae) with potential use for pest management. J Agric Food Chem 51:369–374

Doligez A, Joly HI (1997) Genetic diversity and spatial structure within a natural stand of a tropical forest tree species, *Carapa procera* (Meliaceae), in French Guiana. Heredity 79:72–82

Doyle JJ (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Du Q, Wang B, Wei Z, Zhang D, Li B (2012) Genetic diversity and population structure of Chinese white poplar (*Populus tomentosa*) revealed by SSR markers. J Hered 103:853–862

Earl DA (2012) Structure harvester: a website and program for visualizing structure output and implementing the Evanno method. Conserv Genet Resour 4:359–361

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14:2611–2620

Fischer M, Husi R, Prati D, Peintinger M, van Kleunen M, Schmid B (2000) RAPD variation among and within small and large populations of the rare clonal plant *Ranunculus reptans* (Ranunculaceae). Am J Bot 87:1128–1137

Glaubitz JC (2004) Convert: a user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. Mol Ecol Notes 4:309–310

Gupta P, Varshney R (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113:163–185

Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species Population genetics of forest trees. Springer 42:95–124

Henderson L (1991) Invasive alien woody plants of the northern Cape. Bothalia 21:177–189

Hernández G, Buonamici A, Walker K, Vendramin G, Navarro C, Cavers S (2008) Isolation and characterization of microsatellite markers for *Cedrela odorata* L. (Meliaceae), a high value neotropical tree. Conserv Genet 9:457–459

Inza MV, Zelener N, Fornes L, Gallo LA (2012) Effect of latitudinal gradient and impact of logging on genetic diversity of *Cedrela lilloi* along the Argentine Yungas Rainforest. Ecol Evol 2:2722–2736

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

Lowe A, Jourde B, Breyne P, Colpaert N, Navarro C, Wilson J, Cavers S (2003) Fine-scale genetic structure and gene flow within Costa Rican populations of mahogany (*Swietenia macrophylla*). Heredity 90:268–275

Maciel M, Morais SM, Bevilaqua C, Camurca-Vasconcelos A, Costa C, Castro C (2006) Ovicidal and larvicidal activity of *Melia azedarach* extracts on *Haemonchus contortus*. Vet Parasitol 140:98–104

Novick RR, Lemes MR, Navarro C, Caccone A, Bermingham E (2003) Genetic structure of Mesoamerican populations of Big-leaf mahogany (Swietenia macrophylla) inferred from microsatellite analysis. Mol Ecol 12:2885–2893

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28:2537–2539

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Roop J, Dhaliwal P, Guraya S (2005) Extracts of *Azadirachta indica* and *Melia azedarach* seeds inhibit folliculogenesis in albino rats. Braz J Med Biol Res 38:943–947

Runo M, Muluvi G, Odee DW (2005) Analysis of genetic structure in *Melia volkensii* (Gurke) populations using random amplified polymorphic DNA. Afr J Biotechnol 3:421–425

Scotti-Saintagne C, Dick CW, Caron H, Vendramin GG, Guichoux E, Buonamici A, Duret C, Sire P, Valencia R, Lemes MR (2013) Phylogeography of a species complex of lowland Neotropical rain forest trees (*Carapa*, Meliaceae). J Biogeogr 40:676–692

Sharma D, Paul Y (2013) Preliminary and Pharmacological Profile of *Melia azedarach* L.: an Overview. J Appl Pharm Sci 3:133–138

Syamsuwida D, Palupi ER, Siregar IZ, Indrawan A (2012) Flower initiation, morphology, and developmental stages of flowering-fruiting of mindi (*Melia azedarach L*). J Manaj Hutan Trop 18:10

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 11:1441–1452

Tourn G, Menvielle M, Scopel A, Pidal B (1999) Clonal strategies of a woody weed: *Melia azedarach*. Plant Soil 217:111–117

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3—new capabilities and interfaces. Nucl Acids Res 40:e115

Voigt FA, Farwig N, Johnson SD (2011) Interactions between the invasive tree *Melia azedarach* (Meliaceae) and native frugivores in South Africa. J Trop Ecol 27:355–363

Yeh F, Yang R, Boyle T (1999) POPGENE. Microsoft windows-based freeware for population genetic analysis: release 1.31. University of Alberta, Edmonton. Available from http://www.ualberta.ca/~fyeh/popgene_download.html. Accessed 1 May 2014

Yulianti, Siregar IZ, Wijayanto N, Darma IT, Syamsuwida D (2011) Genetic variation of *Melia azedarach* in community forests of West Java assessed by RAPD. Biodivers 12(2):64–69