Genetic differentiation and diversity analysis of medicinal tree *Syzygium cumini* (Myrtaceae) from ecologically different regions of India

Suphiya Khan¹, Vaishali² and Vinay Sharma¹

¹Department of Bioscience and Biotechnology, Banasthali University, P.O. Banasthali Vidyapith - 304 022, Rajasthan, India

²Department of Biotechnology, SVBP University of Agriculture and Technology, Meerut - 250 004, UP, India

ABSTRACT

This study represents the agro-ecological zone wise surveys of molecular variation of important medicinal tree *Syzygium cumini* Linn. (Jamun) which is native to India. It is used world wide in treatment of diabetes. Despite of its diverse medicinal properties no molecular data is available about the pattern of variation in its natural range. Populations of *S. cumini* in India are located in different habitats which differ from each other with regard to ecological factors. In this study, random amplified polymorphic DNA (RAPD) markers were used to detect inter and intra levels of genetic variations of sixteen *S. cumini* genotypes collected from three major agro-ecological zones of India. A total of 220 amplification products were scored of which 87.50 % were polymorphic. The level of polymorphism ranged from 47.69 % to 74.87 % polymorphic bands per population and was correlated with population size. Different measures of diversity: Shannon's index of phenotypic diversity (I) = 0.451 ± 0.230 ; Nei's genetic diversity (h) = 0.300 ± 0.172 ; effective number of alleles per locus (N_e) = 1.51 ± 0.347 ; total species diversity (H_{sp}) = 0.315 ± 0.031 and within population diversity (H_{pop}) = 0.158 ± 0.104 showed high genetic diversity at species level. Coefficient of genetic differentiation (G_{st} = 0.498; N_m = 0.503) revealed significant genetic differentiation among the populations. Most of the genetic variations are contained among the populations. The results of cluster analysis and principal component analysis (PCA) give only little evidence for an ecotypic differentiation of *S. cumini* populations. Present genetic structure of population suggests *ex situ* conservation in seed banks in which seeds from at least five populations need to be collected and conserved. Secondly, our study provides practical information to herbal drugs manufacturers who use Jamun as a raw material. [Physiol. Mol. Biol. Plants 2010; 16(2) : 149-158] E-mail : dimpy_26@hotmail.com

Keywords : Coefficient of genetic differentiation (G_{st}), PCA, RAPD, *Syzygium cumini*

INTRODUCTION

Assessing genetic diversity and understanding how diversity is structured is not only a prerequisite in designing suitable conservation strategies, but, furthermore, this knowledge helps to resolve taxonomic, phylogenetic, demographic, and ecological questions of great relevance for conservation. Genetic diversity and structure of a species cannot always be predicted on the basis of distribution size and life history and that, therefore, they must be independently studied for each species (Toress *et al.*, 2003). For long-lived tree species

that have to cope with high temporal and spatial environmental heterogeneity, genetic diversity is of prime importance for species persistence. Detrimental anthropogenic impact on the gene pool of trees calls for conservation of tree genetic resources (Geburek, 1997). But relatively few investigations have been undertaken on the conservation genetics of tree species, despite their importance for ecosystem structure and function. Recent surveys indicate that more than 8000 tree species are currently threatened with extinction worldwide (Allnut *et al.*, 2003).

Conservationist need to make choices in matters such as the area that needs to be protected, the populations that should be preserved, the minimum number of individuals required to avoid inbreeding

Correspondence and reprint requests : Suphiya Khan
Tel.: +911438-228302, Fax: +911438228365

problems and to sustain quantitative genetic variation, or the way germplasm should be collected from a population in order to capture most of the genetic diversity for *ex situ* conservation. Therefore, preservation of genetic diversity has become a common objective of most conservation programmes and defining genetic diversity within and among natural populations is a necessary first step towards achieving this goal (Holsinger and Gottlieb, 1991).

Jamun (*Syzygium cumini*) is a large evergreen multipurpose tree of family Myrtaceae upto 15 to 30 m in height (Samba - Murthy and Subrahmanyam, 1989). The Jamun tree which is native to India thrives easily in tropical climate and is found in many parts of Asia and Eastern Africa. *S. cumini* is widely used world wide in treatment of diabetes by the traditional practitioners over many centuries and its various plant parts possess many pharmacological properties. Not only, it has wonderful antihyperglycaemic properties, but it has also proven good anti-oxidant, anti-bacterial, antigenotoxic anti-inflammatory and anti-HIV properties (Sagrawat *et al.*, 2006). In India, populations of *S. cumini* are located in ecologically different habitats. Furthermore, the plant grows on different geological materials and under a magnitude of different climatic conditions. As a consequence, the habitats in which the populations of *S. cumini* are located differ from each other with regard to ecological factors such as nutrient, light and water etc. Genetic variation within and among populations is strongly affected by ecological factors and microgeographic differentiation, caused by these factors, is reported for many species (Reisch *et al.*, 2003). Correlation between the genetic diversity and eco-geographical distribution of individuals in a population may be used to indirectly infer the selective forces shaping the structure of that population (Nevo *et al.*, 1988).

Large-scale characterization of plant species in varying geo-climatic conditions can be performed using various parameters such as seed morphometric traits and isozymes (Kundu, 1999). However, environmental factors, as well as the developmental stage of the plants, influence such traits. DNA based markers are popular tools for examining the extent and distribution of the different aspects of genetic diversity in a species. Recently, two meta-analyses showed that diversity surveys made with dominant markers provide comparable data to surveys undertaken with codominant markers (Nybom and Bartish, 2000; Nybom, 2004). To date, no universal marker system is available for codominant markers, which is applicable to every species.

Such markers need to be developed separately for each species and can be extremely costly in resources and time (Kremer *et al.*, 2005). Therefore, random amplified polymorphic DNA (RAPD) (Williams *et al.*, 1993) is the marker of choice in present study. It has been successfully employed to assess phylogenetic relationships in several tree species also (Anand *et al.*, 2004; Vaishali *et al.*, 2008; Mohapatra *et al.*, 2009).

Despite its great economic value, this species has been poorly studied in regards genetic diversity across its distribution. In present study, the genetic diversity of *S. cumini* populations from different ecogeographic regions has been investigated using RAPD markers and evaluating their overall suitability in such studies. Systematic evaluation and quantification of the variability from different geographical regions in the *Syzygium cumini* gene pool will make it easier to exploit this species for tropical agro-forestry.

MATERIALS AND METHODS

Plant material

To investigate the genetic diversity within and among populations from different types of habitat *Syzygium cumini* leaf samples were collected from sixteen different geographical populations representing five States of India *i.e.* Rajasthan, Uttar Pradesh, Uttarakhand, Madhya Pradesh and Karnataka covering 4 major agro-ecological zones *i.e.* arid, semi-arid, dry sub-humid and moist sub-humid and 9 sub-zones (Sehgal *et al.*, 1992) listed in Table 1. Fully mature leaf samples from adult trees of approximately same age (based on trunk size, height etc.) were randomly selected for collection. The samples were shed dried at the site, transferred in plastic bags, sealed and stored at room temperature (Thomson and Henry, 1993) for further experiments.

Total DNA extraction

The total genomic DNA was isolated from various dry leaf samples of *S. cumini* by a cetyl trimethyl-ammonium bromide (CTAB) method of Weising *et al.* (1995) with minor modifications *viz.* extracted aqueous layer was treated with 5M NaCl to remove polysaccharide contamination, incubation for precipitating DNA after adding isopropanol was carried out overnight and washing steps were repeated twice. The extraction buffer contained 100 mM Tris-HCl (pH 8), 1.5 M NaCl, 25 mM EDTA, 2.5 % CTAB (w/v), 2 % β -mercaptoethanol and 1 % polyvinylpyrrolidone. The purified pellet was dissolved in a minimal amount of Tris EDTA (TE) buffer

Table 1. Details of *Syzygium cumini* accessions collected from different agro-ecological regions of India

Major zones	Sub zones	Place	Sample code	State	Latitude	Longitude
Arid	E-1.2	Jodhpur	JD1	Rajasthan	N 26° 23'	E 73° 8'
Semi-arid	D-4.2	Bhilwara	JD2	Rajasthan	N 22° 25'	E 74° 38'
	D-4.2	Kota	JD3	Rajasthan	N 24° 14'	E 75° 49'
	D-4.2	Jhalawar	JD4	Rajasthan	N 24° 40'	E 76° 10'
	D-4.2	Pratapgarh	JD5	Rajasthan	N 24° 20'	E 74° 40'
	D-4.2	Bhopal	JD14	Rajasthan	N 23° 20'	E 77° 30'
	D-4.3	Meerut	JD9	M.P.	N 29° 10'	E 77° 42'
	D-3.3	Banasthali	JD15	U.P.	N 26° 60'	E 75° 54'
	D-4.4	Bangalore	JD16	Rajasthan	N 12° 59'	E 77° 40'
Dry sub-humid	CD-5.4	Saharanpur	JD7	Karnataka	N 29° 58'	E 77° 33'
	CD-5.4	Kanpur	JD8	U.P.	N 26° 28'	E 80° 20'
	CD-5.4	Lucknow	JD11	U.P.	N 26° 50'	E 81° 00'
	CD-4.1	Banaras	JD12	U.P.	N 25° 22'	E 83° 00'
Moist sub-humid	CM-6.2	Pantnagar	JD6	Uttarakhand	N 29° 31'	E 79° 30'
	CD-6.1	Nazibabad	JD10	Uttarakhand	N 29° 36'	E 78° 18'
	CD-6.1	Roorkee	JD14	Uttarakhand	N 29° 52'	E 77° 59'

(Tris-Cl 0.05 M, EDTA 0.01 M). The isolated DNA was checked on 0.8 % agarose gel by staining with ethidium bromide (0.5 µg/ml) and visualized under mini-transilluminator (Bio-Rad, USA). The gel was then photographed and analysed by a Kodak gel documentation system (Model EDAS 290) using lamda DNA double digest (Bangalore Genei Pvt. Ltd., India) as standard.

RAPD PCR amplification

A total of 30 random decamer primers (custom synthesized by Bangalore Genei Pvt. Ltd., India) were employed in the experiments for genome screening. The primers with more than 50 % GC content were used for RAPD analysis and of these 17 primers were finally selected for data analysis (Table 2). The RAPD analysis on each primer was repeated twice to check the reproducibility. Moreover, in all the PCR reactions, a negative control is included to avoid the erroneous interpretations. Only those gels which showed consistent amplification were considered in this study. The polymerase chain reaction (PCR) reactions were performed in a final volume of 25 µl containing genomic DNA (25 ng), primer (0.2 µM), Taq DNA polymerase (1.5 U/µl), dNTPs (2.5 mM), Taq polymerase buffer 1 x (10 x buffer contained 100 mM Tris-Cl, 500 mM

KCl, 15 mM MgCl₂ and 0.1 % gelatin). The reaction mixture was overlaid with 40 µl of sterile mineral oil to prevent fluid evaporation. DNA amplification was carried out in a Gene Cyclor (Bio-Rad, USA) with following thermal profile: 4 min at 94 °C (initial denaturation) followed by 45 cycles of 15 sec at 94 °C (denaturation), 45 sec at 40 °C (primer annealing) and 90 sec at 72 °C (primer elongation) and a final step of 4 min at 72 °C (final extension). After completion of the cycle, PCR products were stored at –20 °C until further use. Amplification products were separated on 1.5 % (w/v) agarose gels (migration distance: 10 cm) with 1× tris borate EDTA (TBE) buffer. Electrophoresis was performed as 100 V for 2 hrs. 10 µl of low range DNA ruler (Bangalore Genei, India) was run simultaneously, loaded on gel as a molecular standard, and then the gel was visualized, photographed and analyzed as described in the previous section.

Data analysis

Only reproducible bands were scored as present (1) or absent (0) across all *S. cumini* accessions. The binary character matrix of RAPD data were analyzed using the POPGENE ver 1.31 (Yeh *et al.*, 2000). The following genetic diversity parameters were estimated: Shannon's index of phenotypic diversity ($I = \sum P_i \log P_i$), Nei's

gene diversity ($h=1-\sum P_i^2$); here P_i = the frequency of the i^{th} allele in population, effective number of alleles per locus ($N_e = 1/\sum \alpha_i$ here α_i = the number of alleles at locus i and n = the total number of alleles, polymorphic band percentage (PBP), total species diversity (Hsp), diversity within population (Hpop), coefficient of genetic differentiation ($G_{st} = H_{sp} - H_{pop}/H_{sp}$) and gene flow ($N_m = (1/4)(1-G_{st})/G_{st}$ (Nei, 1973). According to, for G_{st} values greater than 0.01, this measure of differentiation is more appropriate than Weir and Cockerham's F_{st} estimate (1984).

Genetic similarity among collections was calculated using Jaccard's similarity coefficient using the SIMQUAL (Similarity for Qualitative Data) routine of NTSYS-pc (version 2.1) (Rohlf, 1993). Similarity coefficient values were used to construct dendrograms, using the unweighted pair group method with arithmetic averages (UPGMA) from the similarity data matrices.

Finally, a principal component analysis (PCA) was conducted with the same program using the EIGEN and PROJ module to highlight the resolving power of ordination. The numerical taxonomy and multivariate analysis system package for personal computer (NTSYS-pc 2.1) was used for this statistical analysis of data.

Resolving power (Rp) for each primer was calculated following Prevost and Wilkinson's (1999) method for selecting primers that can distinguish maximal number of accessions. Resolving power (Rp) of a primer is $R_p = I_b$ (band informativeness) and was calculated as $1 - [2 \times (0.5 - p)]$, p being the proportion of the 16 accessions containing the bands. The polymorphism information content (PIC) of all polymorphic RAPD loci was calculated per primer using this formula: $PIC = 1 - \sum P_i^2$ (Anderson *et al.*, 1993), where p_i is the frequency of the i^{th} allele in a population.

Table 2. RAPD primers used for the screening of the 16 genotypes of *Syzygium cumini*, together with the scorable and polymorphic fragments, resolving power (Rp) and poly information content (PIC)

S.No.	Primer code	Primer Sequences	Total no. of amplified band (NAB)	No. of polymorphic band (NPB)	% Polymorphic band (PPB)	PIC	Rp
1	70SP10C1	CAGGCCCTTC	12	10	83.33	0.50	11.25
2	71SP10G2	TGCCGAGCTG	12	10	83.33	0.55	11.66
3	75SP10C6	GGTCCCTGAC	13	11	84.61	0.28	8.37
4	76SP10G7	GAAACGGGTG	16	15	93.75	0.61	16.50
5	77SP10G8	GTGACGTAGG	12	11	91.66	0.62	10.65
6	78SP10C9	GGGTAACGCC	8	8	100.00	0.56	9.06
7	79SP10G10	GTGATCGCAG	12	6	50.00	0.59	6.50
8	80SP10T11	CAATCGCCGT	12	12	100.00	0.73	11.25
9	81SP10G12	TCGGCGATAG	14	13	92.85	0.67	7.12
10	82SP10C13	CAGCACCCAC	10	9	90.00	0.62	9.75
11	83SP10G14	TTCCGAACCC	10	9	90.00	0.62	9.75
12	84SP10C15	TTCCGAACCC	16	15	93.75	0.61	16.75
13	85SP10A16	AGCCAGCGAA	20	18	90.00	0.76	13.62
14	86SP10T17	GACCGCTTGT	15	14	93.33	0.65	10.63
15	87SP10T18	AGGTGACCGT	11	8	72.72	0.75	7.62
16	88SP10G19	CAAACGTCGG	10	1	90.00	0.86	5.62
17	89SP10C20	GTTGCGATCC	17	15	88.23	0.85	9.75
Average			12.94	10.88	87.50	0.63	10.34

RESULTS AND DISCUSSION

Understanding population genetic diversity, together with the relationship of ecological and geographical differentiation, is a prerequisite for effective conservation and management of a species.

Levels of polymorphism

RAPD marker system being employed to assess the genetic diversity of *S. cumini* germplasm was quite informative and was able to generate adequate polymorphism and unique DNA fingerprints. In this background, sixteen populations of *S. cumini*, representatives of 4 major agro-ecological zones i.e. from arid, semi-arid, dry sub-humid and moist sub-humid and 9 agro-climatic sub zones covering five states viz. Rajasthan, Uttar Pradesh, Uttarakhand, Madhya Pradesh and Karnataka of India as described by Sehgal (1992), were examined for the extent of genetic variability. Overall, 30 random primers were used to examine the genetic diversity pattern. Of the 30 primers, 20 primers (66 %) resulted in amplification, of which 17 primers gave reproducible and scorable results. A total of 220 amplification products were generated by 17 random primers with an average of 12.94 bands per polymorphic primer (87.50 %) (Table 2). The number of amplification products produced by the primer ranged from as low as 1 (Primer 72SP10C3 and 74SP10G5) to a 16 (85SP10A16). The number of polymorphic primers and fragment generated were not of similar range for tree species. They may vary significantly in different plant species. This is understandable as product amplified depends upon the sequence of random primers and their compatibility within genomic DNA. The number of markers detected by each primer depends on primer sequence and the extent of genetic variation, which is genotype specific (Upadhyay *et al.*, 2004). The percent of polymorphic loci reported was upto 94 % using RAPD markers for *Pinus gerardiana*, which was higher than this study (87.50 %), although lower levels of percent polymorphic loci were also reported viz. 28.8 in *Pinus chiapensis* and *Syzygium travancoricum*. Our results of polymorphism in primers (50-100 %) matched

with earlier studies (Gillies *et al.*, 1997; Kant *et al.*, 2006) and showed that species diversity was not low and it should be able to fit in the environmental variation (Fu *et al.*, 2003). Primer 3 and 5 gave single band, which might be explained as RAPD markers tag specific regions in the genome. Lim and Rao (2005) also reported single band while working with oil palm. Large number of bands per primer obtained from RAPD primers used in present study. A possible reason could be the use of primers with 60 to 70 % GC content in our experiments. In an earlier study on the same genus, *Syzygium travancoricum* a high level of polymorphism (> 81 %) was reported using RAPD markers by Anand *et al.* (2004) which is comparable to our results.

Poly information content (PIC) was calculated for each RAPD primer. PIC values ranged as 0.28 - 0.86 with 0.64 mean gene diversity values. Higher Rp value has a greater capacity to separate different accessions (Prevost and Wilkinson, 1999). Selected random primers used in this study are quite informative as they show much higher value (0.565- 16.75). The higher value of Rp and PIC values (Table 2) revealed that the primers used in present study are most suitable for assaying genetic diversity in *S. cumini* collected from different agro-ecological zones of India.

Genetic differentiation/partitioning of genetic diversity in *S. cumini* populations

To assess the overall distribution of diversity within and among these populations Shannon's diversity index and Nei's genetic diversity analysis was performed using POPGENE software for partitioning of genetic diversity within and among three populations using RAPD data. It revealed high population differentiation with restricted gene flow. High coefficient of genetic differentiation among populations ($G_{ST} = 0.498$) was observed in *S. cumini* populations. The Shannon's diversity index analysis showed that 49 % of genetic variation resides among populations and 51 % within population which are again in broad agreement with the result of genetic differentiation analysis. Most of the variations are contained in among the populations (Table 3). This was further confirmed from the low gene flow (Nm) i.e.

Table 3. Partitioning of the genetic diversity within and among populations of *Syzygium cumini* based on RAPD markers.

Primer	Genetic diversity within population (H _{pop})	Total genetic diversity (H _{sp})	Percentage of diversity within population	Genetic differentiation coefficient among populations (G _{ST})	Gene flow (Nm)
Average	0.158	0.315	51 %	0.498	0.503

0.503. Moreover if $Nm < 1$, then local populations tend to differentiate; if $Nm \geq 1$, then there will be little differentiation among populations and migration is more important than genetic drift (Wright, 1951).

Assuming Hardy Weinberg equilibrium, the average gene diversity within populations ($H_{pop} = 0.158 \pm 0.02$) while at species level total species diversity ($H_{sp} = 0.315 \pm 0.031$) was observed. Among these three populations semi-arid populations exhibited great level of variability (PBP = 74.87 %; $N_a = 1.748 \pm 0.434$; $N_e = 1.500 \pm 0.390$; $h = 0.282 \pm 0.199$; $I = 0.415 \pm 0.276$) followed by moist sub-humid and dry sub-humid (Table 4). Measures of genetic diversity parameters viz. the observed number of alleles (N_a), effective number of alleles (N_e), Nei gene diversity (h), shannon's index (I) and polymorphic band percentage (PBP) were higher for semi-arid region followed by moist sub-humid and dry sub-humid. This is understandable as there were eight samples incorporated in semi-arid zone in comparison to dry-sub humid and moist sub-humid zone, which contained four and three samples respectively. The level of heterozygosity attributed to population size was reported earlier in *Chamaecyparis* sp. collected from different locations (Hwang *et al.*, 2001).

In the present study, all the estimates of genetic diversity parameters revealed high genetic diversity in *Syzygium cumini* populations. This result corresponds to earlier findings that tropical trees tend to present high levels of genetic diversity (Hamrick and Loveless, 1989). Further, Salvador *et al.* (2000) showed that a range in genetic variation of forest tree from low to very high depending upon specific population. Broader area of sampling from geographically distinct populations is the possible cause of high gene diversity in present study (Premoli *et al.*, 2001).

According to Wright (1951), G_{st} exceeding 0.15 indicates high genetic differentiation among populations than within populations. High G_{st} value, is a classical result for forest trees especially angiosperm (Raspe *et al.*, 2000), because of maternal inheritance of the chloroplast DNA and because seeds are dispersed over shorter distances than with pollen. In this case, the limited gene flow (genetic isolation) probably resulted from the large geographical distances that limit the long distance dispersal of spores, gametes or drifting reproductive fragments. Geographical distance seems to be important factor for genetic divergence as geographical distance and genetic differentiation are positively correlated (Savolainen *et al.*, 2007). In this study the closest geographic distance between the *S. cumini* populations was 50 km that should be beyond the ranges of both seed and pollen dispersal. Therefore, it is reasonable that a greater population differentiation occurs among *S. cumini* population. Another reason of high degree of genetic differentiation implies that populations are reproductively isolated or have been isolated in the past, resulting in differentiation by inbreeding and genetic drift or through local adaptation (Newton *et al.*, 2002). Any genetic differentiations, however was not simply a function of special distance as indicated by the lack of significant relationships. Other factors known to be important for genetic divergence include successional status, life form, ecological and life history traits (Hamrick and Godt, 1996). In another species of *Syzygium* (*S. travancoricum*) Anand *et al.* (2004) less population differentiation within the population, that was contrary to our results and can be best explained by the fact that our samples cover large geographical area that restricts gene flow and thereby may results in high population differentiation among populations.

Table 4. Measures of genetic diversity parameters viz. the observed number of alleles (N_a), effective number of alleles (N_e), Nei gene diversity (h), shannon's index (I) and polymorphic band percentage (PBP) with standard deviations (S.D.) in parenthesis in each population and overall population in *S. cumini* populations assuming Hardy-Weinberg equilibrium.

Population	Measures of diversity				
	N_a (S.D)	N_e (S.D)	PBP (%)	h (S.D)	I (S.D)
Semi arid	1.748 (0.434)	1.500 (0.390)	74.87 %	0.282 (0.199))	0.415 (0.276)
Dry sub- humid	1.476 (0.500)	1.314 (0.384)	47.69 %	0.180 (0.204)	0.267 (0.293)
Moist sub-humid	1.441 (0.497)	1.296 (0.377)	48.95 %	0.170 (0.202)	0.252 (0.292)
Overall population	1.876 (0.329)	1.514 (0.347)	87.50 %	0.300 (0.172)	0.451 (0.230)

Similarity coefficient values and dendrogram analysis for *S. cumini* populations

Binary data matrix of RAPD data were used to make pair wise comparison of the accessions based on shared and unique amplification products to generate a similarity matrix with NTSYS-pc 2.1. Similarity coefficient values ranged from 0.93 (between Kanpur and Meerut) to 0.51 (Pantnagar and Banasthali) (Table 5). The lowest value of similarity coefficient was observed between the genotypes Pantnagar and Banasthali both of which represent climatic as well as ecological contrasts.

In order to analyze the relatedness among the genotype, dendrogram was constructed on the basis of similarity matrix representing Jaccard's coefficient using the UPGMA algorithm. As evident from the dendrogram, 16 genotypes from different sites formed 4 clusters (Figure 1). However, none of the individual primers could cluster these plants into region specific or agro-climatic sub-zone specific cluster. As shown in figure 1, cluster one was represented exclusively as one major agro-ecological zone as it grouped the samples collected from Jodhpur arid-zone and sub zones (E - 1.2). The cluster two was represented by six samples from Bhilwara (D - 4.2), Jhalawar (D - 4.2) samples, Banasthali (D - 3.3), Varanasi (CD - 4.1), Bhopal (D - 4.2) and Bangalore (D - 4.4). In this group all samples except one belonged to semi-arid agro-ecological zone

Fig. 1. UPGMA phenogram of the genetic relationships among 16 *Syzygium cumini* genotypes based on 220 RAPD markers

Table 5. Nei's genetic similarity matrix of 16 genotypes of *Syzygium cumini* based on RAPD analysis

	Jodhpur	Bhilwara	Kota	Jhalawar	Pratapgarh	Pantnagar	Saharanpur	Kanpur	Meerut	Nazibabad	Lucknow	Varanasi	Roorkee	Bhopal	Banasthali	Bangalore
Jodhpur	1.00															
Bhilwara	0.64	1.00														
Kota	0.55	0.62	1.00													
Jhalawar	0.57	0.86	0.64	1.00												
Pratapgarh	0.56	0.65	0.57	0.67	1.00											
Pantnagar	0.53	0.55	0.54	0.60	0.78	1.00										
Saharanpur	0.52	0.65	0.59	0.65	0.78	0.78	1.00									
Kanpur	0.55	0.67	0.59	0.69	0.78	0.74	0.86	1.00								
Meerut	0.55	0.70	0.62	0.71	0.78	0.74	0.89	0.93	1.00							
Nazibabad	0.55	0.64	0.55	0.65	0.82	0.76	0.82	0.89	0.89	1.00						
Lucknow	0.59	0.67	0.61	0.63	0.71	0.67	0.75	0.81	0.80	0.79	1.00					
Varanasi	0.65	0.66	0.56	0.65	0.64	0.63	0.64	0.71	0.67	0.70	0.71	1.00				
Roorkee	0.56	0.59	0.56	0.60	0.73	0.66	0.67	0.70	0.70	0.74	0.68	0.67	1.00			
Bhopal	0.62	0.71	0.64	0.68	0.69	0.61	0.70	0.75	0.75	0.72	0.74	0.77	0.75	1.00		
Banasthali	0.66	0.75	0.63	0.73	0.60	0.51	0.56	0.66	0.63	0.58	0.61	0.70	0.61	0.71	1.00	
Bangalore	0.57	0.66	0.57	0.63	0.62	0.55	0.62	0.65	0.66	0.66	0.65	0.70	0.65	0.75	0.69	1.00

and three agro-climatic sub zones. The exception was the sample from Varanasi (D - 4.1), which belonged to dry sub-humid agro-ecological zone. Inclusion of JD - 16 from Bangalore in this cluster was quite interesting since it is geographically distant from the rest and was expected to reflect genetic adaptation according to semi-arid agro-ecological zone as it was placed in this zone. This association between distantly located genotypes might be attributed to the unique and broad genetic base of the species, which enable it to maintain and exist in different gene combinations. The cluster 3 was the most heterogenous and was represented by eight samples from Pratapgarh (D - 4.2), Saharanpur (CD - 5.4), Kanpur (CD - 5.4), Meerut (CD - 4.3), Nazibabad (CD - 6.1), Lucknow (CD - 5.4), Pantnagar (CM - 6.2), and Roorkee (CD - 6.1). Out of these, Pantnagar (D - 4.2), Roorkee (CD - 6.1) and Nazibabad (CD - 6.1) belonged to moist sub-humid zones and were related to each other as they were grouped in third cluster, and all others except Meerut (CD - 4.3) and Pratapgarh (D - 4.2) belonged to dry sub-humid zone. The cluster four was represented by a single accession Kota (D - 4.2) that is distinct from all other accessions. UPGMA analysis did not show clustering based on the regions or agro-climatic sub-zones, though high variability exists among the samples.

The principal component analysis (PCA) apart from cluster analysis was performed for grouping of accessions. Applying both methods was recommended to extract the maximum amount of information from the matrix data (Messmer *et al.*, 1992). Clustering was useful in detecting relationships among ecotypes, while PCA allowed a view on the relationships between groups. The results of PCA analysis based on RAPD largely corresponded with the result of the cluster analysis (Figure 2). Both methods showed no ecogeographical based isolation.

Cluster analysis results showed no ecogeographic differentiation as mixing of genotypes from different agro-ecological regions occurs. The cluster 1 of Jodhpur (E-1.2) of semi-arid zone forms a separate cluster. As pointed out by Hartl and Clark (1997), ecological stress could also help to select the favorable alleles or to eliminate deleterious mutants, thereby changing the level and pattern of diversity. Inclusion of Bangalore sample in second cluster was quite interesting since it is geographically distant from the rest. This sample was expected to reflect genetic adaptation according to semi-arid agro-ecological zone as it was placed in this zone. This association between distantly located genotypes may be attributed to the unique and broad genetic base

of the species, which enable it to maintain and exist in different gene combinations.

Our study shows no clear clustering pattern of geographically closer accessions, indicating that the accession between genetic similarity and geographical distance was less significant. This can be explained, since possible ecotypic differentiations are often superimposed by population genetic processes such as founder effects or genetic drift (Nevo *et al.*, 1988). It is more likely that the different levels of intra-population diversity are due to the effects of population size rather than the influence of ecological conditions. Another possible reason of mixing may be attributed to human interventions which make partitioning and distribution of variability, complex. Thus, it would be premature to comment precisely about the correlation of geographic distance and genetic diversity in this case. To confirm the available pattern it is necessary to use more number of accessions from each geographical location.

Results of the present study would be quite informative for conservation planning of *S. cumini*. High genetic diversity among group is higher than within group suggesting regional approach for conservation of *S. cumini* germplasm. Since single or even a few plants will not represent the whole genetic variability, therefore any approach to conserve the full breadth of variation within the species should aim to incorporate as many populations as possible. For a species with limited gene

Fig. 2. Three-dimensional plot of principal component analysis based on a RAPD in 16 *Syzygium cumini* accessions. The numbers plotted represent individual genotypes and correspond to those listed in table 1.

flow and around 50 % variation among populations, it is necessary to collect samples from at least five populations in order to conserve 95 % of the genetic diversity of the species. For a species with only 20 % variation among populations, the samples taken from two populations are enough to get the same results above (Fu *et al.*, 2003). Therefore *ex situ* conservation in seed banks in which seeds from at least five populations need to be collected and conserved is recommended. Secondly, present diversity profile helps herbal drug manufacturers for selecting correct raw material for jamun based herbal drugs. Further work on the levels of secondary metabolites and their possible correlations with different agro-ecological regions along with genetic variation will provide valuable information to herbal drug manufacturers.

ACKNOWLEDGMENTS

Authors acknowledge the financial support provided by Department of Biotechnology, Govt. of India for providing research associateship to Suphiya Khan and also to the Bioinformatics Centre, Banasthali University, India for extensive use of computational facilities.

REFERENCES

- Allnut TR, Newton AC, Premoli A and Lara A (2003). Genetic variation in the threatened South American conifer *Pilgerodendron uviferum* (Cupressaceae) detected using RAPD markers. *Biol. conserv* 114:245-253.
- Anand A, Rao CS, Eganathan P, Kumar NA and Swaminathan MS (2004). Saving an endemic and endangered taxon: *Syzygium travancoricum* Gamble (Myrtaceae): A case study focusing on its genetic diversity and reintroduction. *Physiol Mol Biol Plants* 10:233-242.
- Anderson JA, Churchill GA, Autrique JE, Tanksley SD and Sorrells ME (1993). Optimizing parental selection for genetic linkage map. *Genome* 36: 181-186.
- Fu C, Qiu Y and Kong H (2003). RAPD analysis for genetic diversity in *Changium myrsinoides* (Apiaceae), an endangered plant. *Bot Bull Acad Sinica* 44: 13-18.
- Geburek T (1997). Isozymes and DNA markers in gene conservation of forest trees. *Biodivers and Conserv* 6:1639-1654.
- Gillies ACM, Cornelius JP, Newton AC, Navarro C, Hernandez M and Wilson J (1997). Genetic variation in Costa Rican populations of the tropical timber species *Cedrela odorata* L. assessed using RAPDS. *Mol Ecol* 6: 1133-1145.
- Hamrick JL and Godt M JW (1996). Effect of life history traits on genetic diversity on plant species. *Philos Trans R Soc Land B Biol Sci* 351:1291-1298. doi: 10.1098/rstb.1996.0112.
- Hartl DL and Clark AG (1997). In: Principles of Population Genetics, 3rd edition (Eds. Hartl DL and Clark AG), Sinauer Associates, Sunderland, MA, pp. 682.
- Holsinger KE and Gottlieb LD (1991). Conservation of rare and endangered plants: principles and prospects. In: Genetics and Conservation of Rare Plants, (Eds. Falk DA and Holsinger KE) Oxford University Press: Oxford, pp. 195-223.
- Hwang SY, Lin HW, Kuo YS and Lin TP (2001). RAPD variation in relation to population differentiation of *Chamaecyparis formosensis* and *Chamaecyparis taiwanensis*. *Bot. Bull. Acad. Sinica* 42: 173-179.
- Kant A, Pattanayak D, Chakrabarti SK, Sharma R, Thakur M and Sharma DR (2006). RAPD analysis of genetic variability in *Pinus gerardiana* Wall. in Kinnaur (Himachal Pradesh). *Indian J of Biotech* 5: 62-67.
- Kremer A, Caron H, Cavers S, Colpaert N, Gheysen G, Gribel R, Lemes M, Lowe AJ, Margis R, Navarro C and Salgueiro F (2005). Monitoring genetic diversity in tropical trees with multilocus dominant markers. *Heredity* 95: 274-280.
- Kundu SK (1999). Comparative analysis of seed morphometric and allozyme data among four population of neem (*Azadirachta indica*). *Genet Resour Crop Evol* 46:569-577.
- Lim CC and Rao V (2005). DNA fingerprinting of oil palm-choices of tissues. *J of Oil Palm Res* 17: 136-144.
- Messmer MM, Melchinger AE, Boppenmaier J, Herrmann RG and Jung EB (1992). RFLP analyses of early-maturing European maize germplasm. *Theor and Appl Genet* 83:1003-1012.
- Mohapatra KP, Sehgal RN, Sharma RN and Mohapatra T (2009). Genetic analysis and conservation of medicinal tree species *Taxus wallichiana* in the Himalayan region. *New Forests* 37:109-121.
- Nei M (1973). Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* 70:3321-3323. doi: 10.1073/pnas.70.12.3321.
- Nevo E, Beiles A and Krugman T (1988). Natural selection of allozyme polymorphisms: a microgeographic differentiation in wild emmer wheat (*Triticum dicoccoides*). *Theor and Appl Genet* 75:529-538.
- Newton AC, Allnut TR, Dvorak WS, Del Castillo RF and Ennos RA (2002). Patterns of genetic variation in *Pinus chiapensis*, a threatened Mexican pine, detected by RAPD and mitochondrial DNA RFLP markers. *Heredity* 89:191-198.
- Nybom H (2004). Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Mol Ecol* 13: 1143-1155.
- Nybom H and Bartish I (2000). Effects of life history traits and sampling strategies on genetic diversity estimates obtained with RAPD markers in plants. *Perspect Plant Ecol Evol Syst* 3/2: 93-114.
- Premoli AC, Souto CP, Allnut TR and Newton AC (2001). Effects of population disjunction on isozyme variation in the widespread *Pilgerodendron uviferum*. *Heredity* 87: 337-343.

- Prevost A and Wilkinson MJ (1999). A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. *Theor and Appl Genet* 98: 107-112.
- Raspe O, Saumitou-Laprade P, Cuguen J and Jacquemart AL (2000). Chloroplast DNA haplotype variation and population differentiation in *Sorbus aucuparia* L (Rosaceae: Maloideae). *Mol Ecol* 9:1113-1122.
- Reisch C, Poschlod P and Wingender R (2003). Genetic differentiation among populations of *Sesleria albicans* Kit ex Schultes (Poaceae) from ecologically different habitats in central Europe. *Heredity* 91: 579-527.
- Rohlf FJ (1993). NTSYS-pc Numerical Taxonomy and Multivariate Analysis System, Version 2.1. Exeter Software, Setauket, NY, USA.
- Sagrawat H, Mann AS and Kharya MD (2006). Pharmacological potential of *Eugenia Jambolana*: A review. *Phcog Mag* 2: 96-105.
- Salvador L, Alá R, Agundez D and Gil L (2000). Genetic variation and migration pathways of maritime pine (*Pinus pinaster* Ait.) in the Iberian Peninsula. *Theor and Appl Genet* 100: 89-95.
- Samba-Murthy AASS and Subrahmanyam NS (1989). Fruits In: A Text Book of Economic Botany, Wiley, New Delhi, India, pp. 629.
- Savolainen O, Pyhajarvi T and Knurr T (2007). Gene flow and local adaptation in trees. *Annu Rev Ecol Evol Syst* 38:595-619.
- Sehgal JL, Mandal DK, Mandal C and Vadivelu S (1992). Agro-ecological Regions of India. National Bureau of Soil Survey and Land Use Planning (ICAR), Nagpur, India.
- Thomson D and Henry R (1993). Use of DNA from dry leaves for PCR and RAPD analysis. *Plant Mol Biol Report* 11:202-206.
- Toress E, Iriondo JM and Perez C (2003). Genetic structure of an endangered plant *Antirrhinum microphyllum* (Scrophulariaceae): Allozyme and RAPD analysis. *Am. J of Bot* 90(1): 85-92.
- Upadhyay A, Jayadev K, Manimekalai R and Parthasarathy VA (2004). Genetic relationship and diversity in Indian Coconut accessions based on RAPD markers. *Sci Horti* 99:353-362.
- Vaishali, Khan S and Sharma V (2008). RAPD based assessment of genetic diversity of *Butea monosperma* from different agro-ecological regions of India. *Indian J Biotech* 7:320-327.
- Weir BS and Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.
- Weising K, Nyborn H, Wolff K and Meyer W (1995). DNA isolation and purification. In DNA fingerprinting in plants and fungi. CRC Press, Boca Raton, Alorida pp 44-59.
- Williams JGK, Hanafey MK, Rafalski JA and Tingey SV (1993). Genetic analysis using random amplified polymorphic DNA markers. *Met in Enzy* 218:704-740.
- Wright S (1951). The genetical structure of populations. *Annals of Eugen* 15:323-354.
- Yeh F, Yang RC and Boyle T (2000). Popgene (v.1.32), Microsoft Windows-based freeware for Population Genetic analysis. <http://www.ualberta.ca/~fyeh/Pop32.exe>