CrossMark

ORIGINAL PAPER

Identification of diverse genotypes with high oil content in *Madhuca latifolia* for further use in tree improvement

B. N. Divakara[1,4] · H. D. Upadhyaya[2] · Ananth Laxmi[3] · Rameshwar Das[1]

Abstract *Madhuca latifolia* Macb. commonly known as Indian butter tree, is an open-pollinated plant. Improvement in seed and oil yields depends on the progress in the desired characters in the base material and the genetic variability available in the collected germplasm. We evaluated 23 genotypes of *M. latifolia* to understand genetic variability, character association and divergence in seed traits and oil content for use in breeding programs. Variation was recorded in seed length (27.3–38.6 mm), seed breadth (15.6–19.1 mm), two dimensional (2D) surface area (328.3–495.4 mm^2), 100 seed weight (216.8–285.3 g), acid value (13.4–25.8 mg KOH/g), iodine number (62.4–78.6) and oil content (37.8–51.0 %). High estimates of genotypic coefficient of variation, broad sense heritability and genetic gain were observed for seed oil content. Variability studies for seed traits revealed that genotype CPT-16 had the highest 100-seed weight (281.5 g) and oil content (51 %). Highly significant genotypic and phenotypic correlations were observed. The 100-seed weight was positively and significantly correlated with oil content at both phenotypic ($r = 0.57$) and genotypic ($r = 0.60$) levels. Cluster analysis of the scores of the first three principle components (80.83 %) resulted in four clusters, consisting of 4, 7, 3 and 9 genotypes in the first, second, third and fourth clusters, respectively. Cluster 3 was distinguished from others based on significantly higher means for most seed traits except seed breadth, acid value, iodine number and oil content. Cluster 1 appeared more divergent as it had significantly higher means for acid value and iodine number. A comparative assessment of means of the four clusters for 100-seed weight and oil content suggested that cluster 3 would be useful for higher 100-seed weight and oil content. Hence these genotypes, CPT-3, CPT-6 and CPT-15 in cluster 3 can be used for direct selection and utilization in breeding programs.

Keywords Heritability · Genetic advance · Association · Genetic divergence

✉ B. N. Divakara
bndsira@gmail.com; diwakarbn@icfre.org

[1] Institute of Forest Productivity, Indian Council of Forestry Research and Education, Ranchi 835 303, Jharkhand, India

[2] Genetic Resources, Global Theme on Crop Improvement, International Crops Research Institute for the SemiArid Tropics, Patancheru, Hyderabad 502 324, Andhra Pradesh, India

[3] Institute of Forest Genetics and Tree Breeding, Coimbatore, Tamil Nadu, India

[4] Present Address: Institute of Wood Science and Technology, Indian Council of Forestry Research and Education, 18th cross Malleshwaram, Bangalore 560 003, Karnataka, India

Introduction

Madhuca latifolia Macb. Syn *M. indica, Bassia latifolia* (Sapotaceae) commonly known as Indian butter tree is a large, highly branched, deciduous tree. Indigenous to the Indian subcontinent and grown predominantly in dry tropical and sub-tropical regions of the continent, it is common in deciduous forests and dry sal plain forests in scattered distribution. It is planted on large scale and

extensively cultivated in northern India and the Deccan peninsula for sweet fleshy corollas and ripe fruits, used as a major source of industrial alcohol as well as liquor. The seeds contain valuable edible oil ranging from 38 to 57 % of seed weight that is used by forest dwelling tribal and local people for cooking and as medicine (rheumatism). The fatty oil is known in commerce as mahua butter, mahua fat, Illeppe butter or Basia fat. Historically mahua has been the single largest indigenous source of natural hard fat in soap manufacture both by the small-scale and organized sectors. Seed cake is used as organic manure, fish poison and exported as effective wormicide to control earthworms in turf, lawns and golf courses. Mahua has several medicinal uses. Seed paste is used to cure muscle fatigue and improve vigour of skin. Bark essence is used in curing bleeding gums and ulcers. In addition to these benefits and advantages, mahua seed oil, because it has properties similar to those of diesel fuel has gained importance as bio-diesel and is fast emerging as a viable alternative to fossil fuel.

The successful adoption of bio-fuels is reliant on the supply of feedstock from non-food crops with the capacity to grow on marginal land that is not destined for the cultivation of food crops (Hill et al. 2006). *M. latifolia* meets both of these criteria and is therefore a potential candidate to contribute significant amounts of fuel feedstock. Thus, it can form the basis of a highly promising, profitable, and self-sustaining platform for small-scale entrepreneurship and self-employment in rural areas, ensuring optimum utilization of wasteland resources and unemployed manpower. Although *M. latifolia* is well known as an oil-yielding tree with wide adaptability and many uses, little effort has been directed to improve it as a crop plant because of its long gestation period and slow growth. Even though *M. latifolia* supports the livelihood of tribal populations in India, no information is available regarding the total number of mahua trees, seed yield of seeds per tree and actual production. Hence, it is important to screen the naturally available *M. latifolia* genetic resources and to select the best planting material with high oil content for higher productivity. The selection of superior trees based on seed morphology and oil content may have greater impact than conventional breeding. There is no available information on geographical variation and its influence on seed quality and quantity with regard to Mahua. Mahua is a backbone for the livelihood of tribal communities in Jharkhand. Genetic improvement through selection, evaluation and breeding will definitely impact tribal socio-economics and improve livelihoods. For these reasons, we investigated genetic variation, association and divergence in seed traits and oil content among 23 accessions of *M. latifolia* collected from Jharkhand, India.

Materials and methods

Plus/superior tree selection

Reconnaissance survey was conducted to identify high yielding candidate plus trees (CPTs) of *M. latifolia* at fruiting stage from different predominantly naturalized locations in Jharkhand, India (Table 1; Fig. 1). We selected sample trees by using a single tree selection method based on phenotypic assessment of characters of economic interest, such as yield potential, crown spread, total height, diameter at breast height (dbh), age of the tree, and apparent resistance to pests and disease. We selected a total of 23 CPTs at latitudes from 22° to 24°50′N and longitudes from 83°30′ to 87°E. Using Microsoft Encarta (2005) edition, latitude and longitude of each CPT were recorded as listed in Table 1.

Seed collection and recording of observations

A few kg of mature seeds were collected following a random sampling procedure from all four directions of the crown of each sampled tree during June-July 2005. The total seeds collected were randomly divided into three replications consisting of 100 seeds for measurement of morphometric traits, viz. seed length, seed breadth, aspect ratio, 2D surface area, seed perimeter, equivalent diameter and roundness using Image Analyzer (Leica). Seeds were spread on a glass platform of macro-viewer for each replication and images were captured using a charge-coupled device (CCD) camera and uploaded into the software called Quantimet 500+ or Qwin. Qwin identified the objects based on our specification for seed colour and calibrated captured images to the actual scale. The various 2D measurements of the detected images were measured as follows:

a. Seed length—Length of the object at longest side in mm
b. Seed breadth—Length of the object at shortest side in mm
c. Aspect ratio—Ratio of length divided by breadth.
d. 2D surface area—2D surface area of the seed in the direction of measurement in mm^2
e. Seed perimeter—Total length of the boundary of the feature
f. Equivalent diameter—Diameter of the seeds measured in different directions and the mean in cm
g. Roundness A shape factor that gives a minimum value of unity for a circle. This is calculated from the ratio of perimeter squared to area (Mahadevan et al. 1999)

$$\text{Roundness} = \frac{\text{Perimeter}^2}{4 \times P_i \times 2D \text{ area} \times 1.064}$$

Table 1 Locational details of *Madhuca latifolia* Candidate Plus Trees (CPTs) selected in Jharkhand, India

CPTs	District	Location/village	Latitude	Longitude	Altitude (m)	Age (years)	Height (m)	DBH (cm)	Seed yield (kg/years)	Crown area (m²)
CPT-1	Lohardaga	Chingri Nawatoli	23°26′19N	84°18′25E	590	95	10	80	300	294.19
CPT-2	Simdega	Pharsabera Kusumtoli	22°39′17N	84°31′56E	410	65	13.6	90	100	191.21
CPT-3	Gumla	Samra Velluwa toil	23°03′03N	84°52′28E	520	80	15.3	53	120	123.75
CPT-4	Simdega	Piosokra	22°36′06N	84°43′01E	370	80	17.5	119	180	331.81
CPT-5	Latehar	Hotwag	23°47′28N	84°27′08E	340	70	15.3	70	120	247.55
CPT-6	Daltonganj	Sarja Polpol	23°54′17N	84°31′21E	290	45	13.8	70	70	166.34
CPT-7	Daltonganj	Akhra	23°55′59N	84°10′23E	250	100	14.4	110	200	481.30
CPT-8	Garhwa	Singakala	23°55′18N	83°44′15E	385	100	13.6	110	200	452.57
CPT-9	Garhwa	Garbandh	24°21′24N	83°28′39E	510	60	10.3	85	150	166.34
CPT-10	Daltonganj	Simarbar batawa	24°27′39N	84°13′04E	215	70	12	74	100	247.55
CPT-11	Chatra	Pitiz	24°15′19N	85°06′50E	375	40	10.3	55	30	133.81
CPT-12	Latehar	Nachna	23°51′54N	84°49′44E	540	120	9.8	103	200	216.51
CPT-13	Ranchi	Piska Mahua Toli	23°28′03N	85°26′24E	610	50	14	60	150	203.67
CPT-14	Hazaribag	Ahoanhe moraha	23°49′37N	85°24′55E	510	40	10.2	54	70	166.34
CPT-15	Koderma	Banpok	24°34′36N	85°42′47E	390	80	10.2	70	100	125.73
CPT-16	Giridih	Madhgopali	23°56′56N	86°00′44E	340	70	11	78	100	212.62
CPT-17	Chaibasa	Khas Jamda	22°09′08N	85°2954E	550	70	12.8	80	100	229.75
CPT-18	Saraikela	Palobera	22°15′52N	85°59′14E	300	45	15.4	65	60	243.38
CPT-19	Jamshedpur	Golkatta	22°39′02N	86°27′09E	190	50	12.7	52	70	128.73
CPT-20	Chaibasa	Junko	22°45′19N	85°26′25E	450	50	17.7	70	100	308.03
CPT-21	Ranchi	PaniMahua Chamma	23°32′27N	85°06′17E	650	50	14	60	150	203.67
CPT-22	Dhanbad	Kokra	23°59′50N	86°30′45E	240	90	14.3	97	100	274.76
CPT-23	Latehar	Chandwa	23°41′42N	84°43′36E	520	80	13.8	106	120	335.05

Beside these image parameters, the following estimates were also recorded for each genotype.

h. 100-seed weight—Weight of 100 pure seeds in gm
i. Total oil content—Estimated following the procedure of Sadasivam and Manickam (1992)
j. Acid value—Estimated following the procedure of Cox and Pearson (1962)
k. Iodine number—Estimated following the procedure of Horowitz (1975)

Data analysis

Best Linear unbiased predictors (BLUPs) were obtained for each trait. BLUPs were subjected to a significance test at 5 % critical difference to quantify differences between CPTs. Also pod traits of 23 genotypes of *M. latifolia* were analysed using analysis of variance (ANOVA) to understand the significance of differences between genotypes for various pod traits (Gomez and Gomez 1984). The phenotypic variation for each trait was partitioned into components due to genetic (hereditary) and non-genetic

(environmental) factors and estimated using the following formula (Johanson et al. 1955):

$$V_{\mathrm{p}} = {MSG}/{r}; \; V_{\mathrm{g}} = {(MSG - MSE)}/{r}; \; V_{\mathrm{e}} = MSE \qquad (1)$$

where, MSG, MSE and r are the mean squares of CPTs, mean squares of error and number of replications, respectively.

The phenotypic variance (V_{p}) is the total variance among phenotypes when grown over the range of environments of interest. The genotypic variance (V_{g}) is the part of the phenotypic variance that can be attributed to genotypic differences among the phenotypes. The error variance (V_{e}) is the portion of the phenotypic variance attributed to environmental effects. To be able to compare the variation among traits, phenotypic coefficients of variation (PCV) and genotypic coefficients of variation (GCV) were computed according to the method suggested by Burton (1952):

$$PCV = \sqrt{Vp/X} \times 100; \; GCV = \sqrt{Vg/X} \times 100 \qquad (2)$$

where V_{p}, V_{g} and X are the phenotypic variance, genotypic variance and grand mean for each pod and seed-related trait, respectively.

Fig. 1 Distribution of *Madhuca latifolia* candidate plus trees in Jharkhand, India mapped using Microsoft Encarta 2005. Details of number representation is in Table 1

Broad sense heritability (h^2b) was calculated according to Allard (1999) as the ratio of the genotypic variance (V_g) to the phenotypic variance (V_p). Genetic advance (GA) expected and GA as a percentage of the mean assuming selection of the superior 5 % of the genotypes were estimated following Johanson et al. (1955) as below:

$$GA = K \cdot h^2b \cdot \sqrt{V_p}; \text{Genetic gain} = \left(GA/X\right) \times 100 \quad (3)$$

where K is the selection differential (2.06 for selecting 5 % of the genotypes).

Phenotypic (r_p) and genotypic (r_g) correlations were further computed to examine inter-character relationships among seed and seedling traits following Goulden (1952) as:

$$r_p = \text{Cov}_p^{(x_1, x_2)} \Big/ \left[V_p(x_1) \cdot V_p(x_2)\right]^{1/2} \quad (4)$$

$$r_g = \text{Cov}_g^{(x_1, x_2)} \Big/ \left[V_g(x_1) \cdot V_g(x_2)\right]^{1/2} \quad (5)$$

where Cov$_p$ and Cov$_g$ are phenotypic and genotypic co-variances for any two traits x_1 and x_2, respectively, and Vp

and Vg are the respective phenotypic and genotypic variances for those traits.

The mean observations for all traits for each season were standardized by subtracting from each observation the mean value of the character and subsequently dividing it by its respective standard deviation. These standardized values, with average 0 and standard deviation 1, were used for principal component analysis (PCA) on Genstat 10 to quantify the importance of different traits in explaining multivariate polymorphism. Cluster analysis was performed using the scores of the first three principal components (PCs) following Ward (1963). Mean, range and variance were computed for each trait and cluster. Means of clusters were compared using the Newman-Keuls (Newman 1939; Keuls 1952) procedure. The homogeneity of variances among the clusters was tested using Levene's test (Levene 1960).

Results and discussions

BLUPs differed significantly among genotypes for all traits (Table 2; Fig. 2). Variability studies for seed traits revealed that genotype CPT-16 had the highest 100 seed weight (281.5 g) and oil content (51.0 %), followed by CPT-3 (263.9 g, 50.2 %). Maximum values for seed length (38.6 mm) and aspect ratio (2.2) were observed in CPT-15. Lowest 100 seed weight, oil content, aspect ratio, acid value, and Iodine number were recorded for genotype CPT-4 (216.8 g), CPT-21 (37.8 %), CPT-10 (1.6), CPT-13 (13.4) and CPT-11 (62.6), respectively. Though range is a crude measure of variability in genotypes, it does give an idea of the spread of variation for a particular character. Wide variation was observed for seed length (27.3–38.6 mm), seed breadth (15.6–19.1 mm), 2D surface area (328.3–495.4 mm^2), 100 seed weight (216.8–285.3 g),

Table 2 Mean performance of selected *Madhuca latifolia* genotypes for seed traits

| Genotypes	Seed length (mm)	Seed breadth (mm)	Aspect ratio	2D surface area (mm^2)	Perimeter (mm)	Equivalent diameter (mm)	Roundness	100 seed weight (g)	Acid value (mg KOH/g)	Iodine number	Oil content (%)
CPT-1	33.4	17.0	2.0	409.0	84.2	22.8	1.3	236.2	16.1	68.6	44.0
CPT-2	31.7	16.8	1.9	387.4	80.8	22.2	1.3	271.1	15.1	68.2	39.4
CPT-3	37.6	18.1	2.0	486.2	92.9	24.8	1.3	263.9	14.3	66.0	50.2
CPT-4	30.7	15.7	2.0	359.8	78.6	21.3	1.3	216.8	14.6	73.5	40.5
CPT-5	34.3	17.9	1.9	458.4	87.1	24.1	1.3	285.3	17.4	76.4	49.1
CPT-6	38.3	16.8	2.2	474.1	93.5	24.3	1.4	249.7	15.3	62.7	40.9
CPT-7	32.2	17.6	1.8	417.4	83.0	23.0	1.3	222.2	23.9	65.8	40.3
CPT-8	33.6	19.1	1.8	462.3	87.9	24.1	1.3	259.9	17.3	73.6	45.4
CPT-9	35.2	19.1	1.9	495.4	90.8	25.0	1.3	246.0	16.2	72.7	38.8
CPT-10	28.0	18.1	1.6	362.3	73.4	21.4	1.1	227.9	16.4	72.1	45.9
CPT-11	33.6	16.4	2.0	401.9	77.5	22.5	1.1	242.4	17.6	62.6	44.0
CPT-12	33.3	15.6	2.1	383.8	83.8	22.0	1.4	257.6	17.3	68.1	46.1
CPT-13	30.1	15.8	1.9	349.5	66.9	21.0	1.0	254.7	13.4	74.5	47.4
CPT-14	32.3	18.3	1.8	435.5	83.6	23.4	1.2	240.4	21.1	69.6	42.6
CPT-15	38.6	17.0	2.2	483.2	94.6	24.5	1.4	261.3	15.0	71.4	41.1
CPT-16	32.9	18.0	1.8	437.3	84.6	23.6	1.2	281.5	25.3	72.0	51.0
CPT-17	34.3	15.9	2.1	401.8	84.6	22.5	1.3	220.1	19.8	72.6	39.1
CPT-18	32.3	17.0	1.9	407.8	82.2	22.8	1.2	226.5	16.5	69.3	43.8
CPT-19	27.3	16.8	1.6	340.2	72.8	20.8	1.2	236.0	22.5	75.9	45.0
CPT-20	34.6	17.3	2.0	438.9	86.7	23.6	1.3	250.2	17.0	72.9	44.2
CPT-21	32.3	18.8	1.7	456.2	86.7	24.0	1.2	227.1	25.8	77.7	37.8
CPT-22	32.3	16.7	1.9	400.3	83.1	22.5	1.3	264.4	20.3	62.4	48.0
CPT-23	28.8	15.6	1.8	328.3	64.7	20.4	1.0	252.0	18.5	78.6	43.4
Mean	32.9	17.2	1.9	416.4	82.8	22.9	1.3	247.5	18.1	70.7	43.8
SEM	0.9	0.3	0.0	13.6	2.3	0.4	0.0	3.1	2.4	0.9	0.6
CD (5 %)	2.5	0.8	0.1	39.5	6.7	1.2	0.1	8.9	6.9	2.6	1.7
C.V. (%)	4.7	2.7	4.0	5.9	5.0	3.2	5.8	2.2	27.6	2.3	2.3
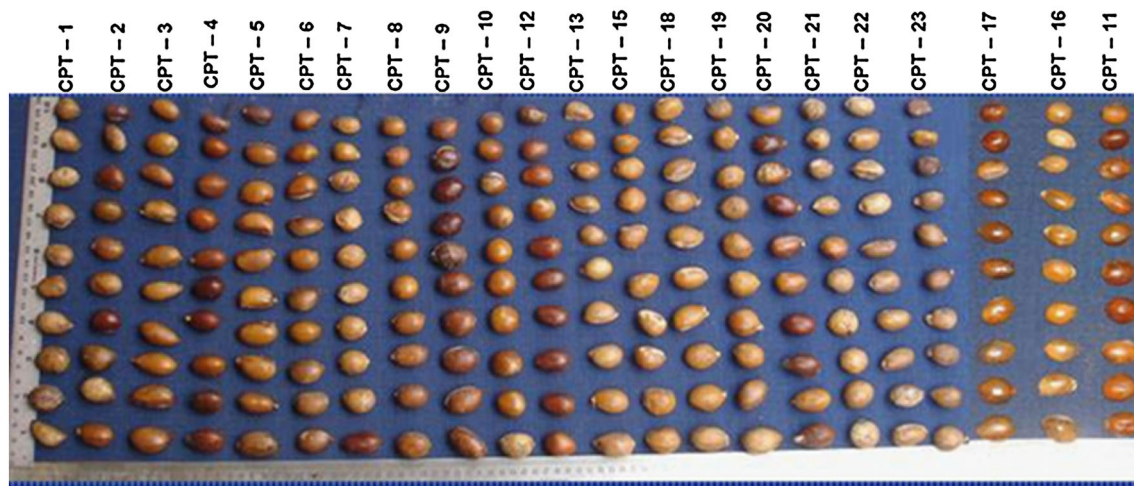
Fig. 2 Seed diversity among selected Candidate Plus Trees of *Madhuca latifolia*

Table 3 Estimates of variance components and other parameter for seed traits in *Madhuca latifolia*

Seed traits	Mean ± SD	Range	Variance due to genotype	Heritability (%)	Coefficient of variation (%)		Genetic advance (%) of mean
					Genotypic	Phenotypic	
Seed length (mm)	32.9 ± 2.9	27.3–38.6	9.2	79.5	9.2	10.3	16.9
Seed breadth (mm)	17.2 ± 1.1	15.6–19.1	1.3	85.9	6.6	7.1	12.6
Aspect ratio	1.9 ± 0.2	1.6–2.2	0.0	83.6	9.1	9.9	17.1
2D surface area (mm^2)	416.4 ± 48.9	328.3–495.4	2571.6	81.2	12.2	13.5	22.6
Perimeter (mm)	82.8 ± 7.8	64.7–94.6	65.6	79.3	9.8	11.0	18.0
Equivalent diameter (mm)	22.9 ± 1.3	20.4–25.0	1.9	78.7	6.1	6.8	11.1
Roundness	1.2 ± 0.1	1.0–1.4	0.0	79.4	11.3	12.7	20.8
100 seed weight (g)	247.5 ± 19.2	216.8–285.3	379.1	93.0	7.9	8.2	15.6
Acid value (mg KOH/g)	18.1 ± 3.5	13.4–25.8	18.1	42.1	23.5	36.2	31.4
Iodine number	70.7 ± 4.7	62.4–78.6	22.9	90.1	6.8	7.1	13.2
Oil content (%)	43.8 ± 3.7	37.8–51.0	14.2	93.5	8.6	8.9	17.2

acid value (13.4–25.8 mg KOH/g), iodine number (62.4–78.6) and oil content (37.8–51.0 %) (Table 3). It has been documented that seed weight depends on reserve food material, which is produced as a result of double fertilization (endosperm), is dominated by the maternal traits and is also influenced by the nutrient availability at the time of seed setting and environmental factors (Allen 1960; Johnsen et al. 1989). Embryo development and its physiological function are contributed by the maternal and paternal (pollen grain) traits in the species. The occurrence of *M. latifolia* over a wide range of habitats with diverse geo-climatic conditions was expected to be reflected in the genetic constitution of its populations. In the present study, the significant variation in seed traits exhibited by the genotypes could be attributed to isolations that influence gene flow. Significant variability of seed characters, such as seed size and weight, was observed in seeds of the

selected plus trees (Bagchi and Sharma 1989) and among various provenances of *Santalum album* (Veerendra et al. 1999).

The phenotypic and GCV were also similar for all traits, except acid value, which exhibited striking difference between PCV (36.2) and GCV (23.5), indicating that for most traits genetic control was quite high (Table 3). The magnitude of the error variance was lower than that of the genotypic variance for all traits except acid value (data not given). Estimates of broad sense heritability ranged from 42.1 % (acid value) to 93.5 % (oil content) and for the other seven traits heritability was 60 % or more. Genetic gain (%) ranged between 11.1 and 31.4 % with equivalent diameter giving the lowest value and acid value giving the highest value. Relatively high value of genotypic variance resulted in high estimates of heritability which contributed to the high genetic gains expected in this material. The

genotypic coefficient of variation, heritability and genetic gain were higher for an important trait such as oil content. Our results align with those reported by Kaushik et al. (2007a) for *Jatopha curcas*. Our high estimates of heritability combined with high GA suggest that population means for oil content can be changed considerably by selecting the superior 5 % of the genotypes.

As variation among clones used for estimation of genetic variation and genetic gain, co-variance estimates between traits can be used to estimate genetic correlations between the traits (Foster 1986). In genetic improvement of seed and oil yields of *M. Latifolia*, clear understanding of the relationships among different seed traits is very essential. Correlation establishes the extent of association between seed traits and their attributes so that these components can be used as additional criteria for selection in breeding program. Correlated quantitative traits are of major interest in an improvement program, as the improvement of one character may cause simultaneous correlated changes in other characters. Of the 110 (55 genotypic and 55 phenotypic) correlations, 16 genotypic and 15 phenotypic combinations were significant at 1 % along with 6 genotypic and 4 phenotypic combinations that were significant at 5 % (Table 4). Skinner et al. (1999) suggested that only those correlation coefficients greater than 0.707 or smaller than −0.707 are biologically meaningful so that 50 % of the variation in one trait is predicted by the other. The seed trait pair showing such high

Table 4 Genotypic (G) and phenotypic (P) correlation coefficients for seed traits in *Madhuca latifolia*

Seed traits	Seed breadth (mm)	Aspect ratio	Area (mm^2)	Perimeter (mm)	Equivalent diameter (mm)	Roundness	100 seed weight (g)	Acid value (mg KOH/g)	Iodine number	Oil content (%)
Seed length (mm)										
G	0.2	0.77**	0.85**	0.88**	0.84**	0.75**	0.33	−0.34	−0.43*	−0.09
P	0.24	0.76**	0.86**	0.89**	0.85**	0.71**	0.28	−0.2	−0.39	−0.04
Seed breadth (mm)										
G		−0.48*	0.67**	0.50*	0.69**	0.15	0.15	0.34	0.12	0.03
P		−0.43*	0.68**	0.48*	0.69**	0.1	0.13	0.25	0.08	0.07
Aspect ratio										
G			0.32	0.47*	0.3	0.59**	0.2	−0.52**	−0.46*	−0.1
P			0.34	0.50*	0.33	0.59**	0.16	−0.39	−0.41*	−0.06
2D surface area (mm^2)										
G				0.93**	0.99**	0.64**	0.32	−0.02	−0.22	−0.09
P				0.91**	0.99**	0.59**	0.27	0.01	−0.2	−0.04
Perimeter (mm)										
G					0.93*	0.88**	0.23	−0.04	−0.35	−0.16
P					0.91**	0.86**	0.2	−0.03	−0.31	−0.09
Equivalent diameter (mm)										
G						0.64**	0.32	0.01	−0.22	−0.07
P						0.58**	0.26	0.02	−0.21	−0.03
Roundness										
G							0.05	−0.08	−0.47*	−0.24
P							0.07	−0.07	−0.37	−0.15
100 seed weight (g)										
G								−0.17	−0.05	0.60**
P								−0.09	−0.05	0.57**
Acid value (mg KOH/g)										
G									0.21	−0.05
P									0.11	−0.04
Iodine number										
G										−0.08
P										−0.07

* Significant at $p = 0.05$

** Significant at $p = 0.01$

correlation was 18 and all were positive. However, 100 seed weight registered positive and highly significant association with oil content at both phenotypic (0.57) and genotypic (0.60) levels is also important as it is the trait of economic interest (Table 4). Similar correlation trend was seen in *Jatropha curcas* (Kaushik et al. 2007a) and *Pongamia pinnata* (Kaushik et al. 2007b). Hence 100 seed weight may be a suitable criterion for selection of plus trees. Aspect ratio recorded negatively significant correlation with acid value (−0.52) at genotypic level and with iodine number both at phenotypic (−0.41) and genotypic (−0.46) levels. Genotypic correlation is an estimated value, whereas, phenotypic correlation is derived from the genotype and environmental interaction (Chaturvedi and Pandey 2004). The genotypic correlation is, therefore, a more reliable estimate for examining the degree of relationship between character pairs. Since mahua seed oil is edible, acid value and iodine number which influence the shelf life of oil are important characteristics. Hence high aspect ratio may be one of the criteria for selecting genotypes with low acid value and iodine number.

Genetic diversity in plant species forms the basis for selection and further improvement. Our information on the genetic structure and diversity relationships of CPTs provides a basis for planning and conducting future collections and efficient utilization of genetic resources to realize the potential for maximizing seed and oil yield. Diversity among the seeds collected from each CPT is presented in Plate 1. A very large proportion of the total variation (80.8 %) was explained by the first 3PCs (data not given).

The first PC alone accounted for 46.2 % of the variation followed by the second PC, which explained 19.8 % of the variation. The third PC accounted for 10.2 % of the variation. Based on loading for the first three PCs, characters such as 2D surface area, aspect ratio, seed breadth, equivalent diameter, 100 seed weight, seed length, perimeter and oil content are important and adequate descriptors in this material.

Cluster analysis of the scores of the first three PCs resulted in four clusters (Fig. 3). The first cluster comprised four genotypes (CPT-10, CPT-13, CPT-19, CPT-23), the second comprised seven genotypes (CPT-5, CPT-7, CPT-8, CPT-9, CPT-14, CPT-16, CPT-21), the third comprised three genotypes (CPT-3, CPT-6, CPT-15), and the remaining nine genotypes (CPT-1, CPT-2, CPT-4, CPT-11, CPT-12, CPT-17, CPT-18, CPT-20, CPT-22) were grouped into the fourth cluster. The clustering pattern proved that geographical diversity need not necessarily be related to genetic diversity. This kind of genetic diversity might be due to differential adoption, selection criteria, selection pressure and environment (Vivekananda and Subramaninan 1993). This indicated that genetic drift produced greater diversity than did geographic diversity (Singh et al. 1996). Absence of any relationship between genetic diversity and geographical distribution is in accordance with the findings of Kaushik et al. (2007a) and Gohil and Pandya (2008) for *J. curcas*. The range, mean, and variances for the four clusters are listed in Table 5. Cluster 3 was distinguished from other clusters based on significantly higher means for most seed traits except seed breadth, acid

Fig. 3 Grouping of 23 *Madhuca latifolia* genotypes based on scores of first three principal components
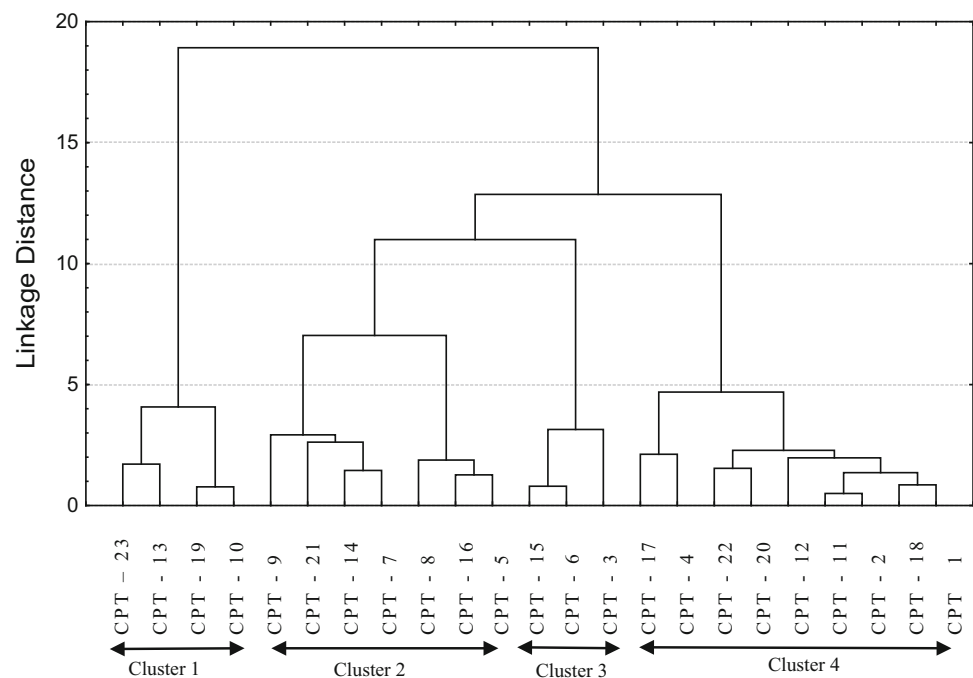
Table 5 Range, mean, and variance for different seed traits in four clusters of selected genotypes of *Madhuca latifolia*

	Seed length (mm)	Seed breadth (mm)	Aspect ratio	2D surface area (mm²)	Perimeter (mm)	Equivalent diameter (mm)	Roundness	100 seed weight (g)	Acid value (mg KOH/g)	Iodine number	Oil content (%)
Range											
Cluster 1	27.3–30.1	15.6–18.1	1.6–1.9	328.3–362.3	64.7–73.4	20.4–21.4	1.0–1.2	227.9–254.7	13.4–22.5	72.1–78.6	43.4–47.4
Cluster 2	32.2–35.2	17.6–19.1	1.7–1.9	417.4–495.4	83.0–90.8	23.0–25.0	1.2–1.3	222.2–285.3	16.2–25.8	65.8–77.8	37.8–51.0
Cluster 3	37.6–38.6	16.8–18.1	2.0–2.2	474.1–486.2	92.9–94.6	24.3–24.8	1.3–1.4	249.7–263.9	14.3–15.3	62.7–71.4	40.9–50.2
Cluster 4	30.7–34.6	15.6–17.3	1.9–2.1	359.8–438.9	77.5–86.7	21.3–23.6	1.1–1.4	216.8–271.1	14.6–20.3	62.4–73.5	39.1–48.0
Mean[1]											
Cluster 1	28.6^c	16.6^b	1.7^c	345.1^d	69.4^c	20.9^c	1.1^c	242.7^a	17.7^{a,b}	75.3^a	45.4^a
Cluster 2	33.3^b	18.4^a	1.8^c	451.8^b	86.3^b	23.9^a	1.2^b	251.8^a	21.0^a	72.5^{a,b}	43.6^a
Cluster 3	38.2^a	17.3^b	2.1^a	481.2^a	93.7^a	24.5^a	1.4^a	258.3^a	14.9^b	66.7^b	44.1^a
Cluster 4	32.9^b	16.5^b	2.0^b	399.0^c	82.4^b	22.5^b	1.3^b	242.8^a	17.1^{a,b}	68.7^b	43.2^a
Variance[2]											
Cluster 1	1.4	1.4	0.03	207.7	18.4	0.2	0.01	164.5	14.6	7.6	2.8
Cluster 2	1.3	0.3	0	621.7	7.4	0.4	0	620.3	16.7	16.2	26
Cluster 3	0.3	0.5	0.01	40	0.7	0.06	0	57	0.3	19.3	27.9
Cluster 4	1.6	0.4	0.01	461.1	8.8	0.4	0	379.8	3.7	16.7	9
F value	0.67	2.64	6.86	0.69	1.85	0.49	2.81	2.34	4.69	0.38	2.7
Prob > F	0.579	0.079	0.003	0.568	0.172	0.691	0.067	0.106	0.013	0.766	0.075

[1] Difference between means of different clusters were tested using the Newman–Keuls test, Means followed by the same letter are not significantly different at $P = 0.05$

[2] Variance were tested following Levene (1960) test

value, Iodine number and oil content. Cluster 1 also appeared more divergent as it had significantly higher mean values for acid value and Iodine number. A comparative assessment of the means of the four clusters for 100 seed weight vis-à-vis oil content suggested that cluster 3 may be ideal for developing higher yields of seed and oil. Hence these genotypes (CPT-3, CPT-6 and CPT-15) can be used for direct selection and utilization in breeding programs. Earlier studies, of crop plants indicated that inter-mating of divergent groups led to greater opportunity for crossing over which would release latent variation by breaking up predominantly repulsion linkage (Thoday 1960) and the importance of using diverse parents in breeding was also stressed by Singh et al. (1981). Arya et al. (1999) reported significant variation in growth and biomass of tree seedlings at 1 year for *Prosopis cineraria*. Similar records of seedling growth were also reported for *Eucalyptus* and *Casuarina sp.* (Toky and Bisht 1991) and *Bombax ceiba* (Chaturvedi and Pandey 2001).

Conclusions

CPT-3, CPT-5 and CPT-16 were found to be superior for 100 seed weight and oil content. Hence seeds of these CPTs can be assigned priority for extensive afforestation programs. Seed traits, viz. 100 seed weight and oil content are under strong genetic control; improvement in these characters may be achieved through further selection. Since 100 seed weight and oil content were positively correlated, while aspect ratio negatively influenced iodine and acid value, 100 seed weight and aspect ratio can be included as potential selection criteria for *M. latifolia*. The divergence among the genotypes based on the first 3 PCs and the subsequent comparative assessment of means of the four clusters for 100-seed weight and oil content suggested that, cluster 3 may be ideal for developing higher 100-seed weight and oil. Hence these genotypes, CPT-3, CPT-6 and CPT-15, can be used for direct selection and utilization in breeding programs.

References

Allard RW (1999) Principles of plant breeding, 2nd edn. Wiley, New York
Allen GS (1960) Factors affecting the viability and germination behaviour of coniferous seed. IV. Stratification period and incubation temperature, *Pseudostuga menziesii* (Mirb.) *Franco*. For Chron 36:18–19
Arya S, Toky OP, Bhist RP, Tomer R, Harris PJC (1999) Variation in growth and biomass production of one year seedling of 30 provenances of *Prosopis cineraria* (L.) Druce in arid India. Ind J For 22:169–173
Bagchi SK, Sharma VP (1989) Biometrical studies on seed characters of *Santalum album* L. Silvae Gene 38:152–153
Burton GW (1952) Quantitative inheritance of grass. In: Proceeding of the 6th International Grassland Congress, Pennsylvania State College, State College, U.S. pp 74–83
Chaturvedi OP, Pandey N (2001) Genetic Divergence in *Bombax ceiba* L. germplasm. Silvae Genet 50:3–4
Chaturvedi OP, Pandey N (2004) Correlation and path analysis studies between biomass and other characters in *Bombax ceiba* L. Silvae Genet 53:5–6
Cox HE, Pearson D (1962) The chemical analysis of Foods. Chemical Publishing Co. Inc, New York, p 420
Foster GS (1986) Provenance variation of eastern cottonwood in the lower Mississippi valley. Silvae Genet 35:32–38
Gohil RH, Pandya JB (2008) Genetic diversity assessment in physic nut (*Jatropha curcas* L.). Int J Plant Prod 2:321–326
Gomez AK, Gomez AA (1984) Statistical procedure for agricultural research. Wiley.
Goulden CH (1952) Some distance properties of latent root and vector methods used in multivariate analysis. Biometrica 53:325–338
Hill J, Nelson E, Tilman D (2006) Environmental, economic and energetic costs and benefits of biodiesel and ethanol biofuels. Proc Natl Acad Sci USA 103:11206–11210
Horowitz W (1975) Official methods of analysis of AOAC, 12th edn. Association of official analytical chemists Washington, Washington
Johanson HW, Robinson HF, Comstock RE (1955) Estimate of Genetic and environmental variability in soybeans. Agron J 47:314–318
Johnsen O, Dietrichson J, Skaret G (1989) Phenotypic changes in progenies of northern clones of *Picea abies* (L.) Karst. grown in a southern seed orchard. III. Climate changes and growth in a progeny trial. Scand J For Res 4:343–350
Kaushik N, Kumar K, Kumar S, Kaushik N, Roy S (2007a) Genetic variability and divergence studies in seed traits and oil content of Jatropha (*Jatropha curcas* L.) accessions. Biomass Bioenergy 31:497–502
Kaushik N, Kumar S, Kumar K, Beniwal RS, Kaushik N, Roy S (2007b) Genetic variability and association studies in pod and seed traits of *Pongamia pinnata* (L.) Pierre in Haryana, India. Genet Resour Crop Evol 54:1827–1832
Keuls M (1952) The use of the studentized range in connection with an analysis of variance. Euphytica 1:112–122
Levene H (1960) Robust tests for equality of variances. In: Olkin I (ed) Contributions to probability and statistics: essays in honour of Harold Hotelling. Stanford University Press, Stanford, pp 278–292
Mahadevan NP, Sivakumar V, Gurudev SB (1999) Relationship of cone and seed traits on progeny growth performance in *Casuarina equisetifolia* Forst. & Forst. Silvae Genet 48:273–277
Newman D (1939) The distribution of range in samples from a normal population expressed in terms of an independent estimate of standard deviation. Biometrica 63:395–400
Sadashivam S, Manikam A (1992) Biochemical Methods for Agricultural Sciences. Wiley Eastern Ltd, New Delhi
Singh YP, Kumar A, Chauhan BPS (1981) Genetic divergence in pearl millet. Indian J Genet 41:186–190
Singh AK, Singh SB, Singh SM (1996) Genetic divergence in scented and fine genotypes of rice (*Oryza sativa* L.). Ann Agric Res 17:163–166
Skinner DZ, Bauchan GR, Auricht G, Hughes S (1999) A method for the efficient management and utilization of large germplasm collections. Crop Sci 39:1237–1242

Thoday JM (1960) Effect of disruptive selection. III. Coupling and repulsion. Heredity 14:35–49

Toky OP, Bisht RP (1991) Provenance trials of eucalyptus and casuarina in arid climate of northwestern India. Indian J For 14:315–317

Veerendra HCS, Ramalakshmi S, Mallesha BB (1999) Variation in seed characteristics in provenances of sandal (*Santalum album* L.). Ind For 125:308–312

Vivekananda P, Subramaninan S (1993) Genetic divergence in rainfed rice. Oryza 39:60–62

Ward JH (1963) Hierarchical grouping to optimize an objective function. J Am Stat Assoc 38:236–244