Utility of P450-Based-Analogue (PBAs) Markers for Diversity Analysis of Linseed (*Linum usitatissimum* L.) Cultivars

MK Rana, Munjal D Vyas, Sonika Singh and KV Bhat
National Research Centre on DNA Fingerprinting, National Bureau of Plant Genetic Resources, New Delhi-110 012

Molecular marker techniques to unravel genome-wide variation scattered in the functional regions are becoming available with the advancements in DNA research. These could be of greater utility for assessing genetic diversity at the molecular level than the neutral marker techniques. Mammalian cytochrome P450 analogues (PBAs) in crop plants are said to be ideal tools in this direction. The utility of these was tested for assessing intra-specific diversity in twenty-one linseed cultivars. The results with fifteen primer-combinations suggested that of the total 136 markers that were generated only 25% were found to be polymorphic. The level of polymorphism in linseed cultivars was quite low, although it was possible to discriminate all the cultivars using combined profiles. The grouping of the cultivars using these markers, however, could be related to the known pedigrees.

Key words: *Linum usitatissimum* L., Cytochrome P450, Genetic diversity, Linseed, DNA markers

Molecular markers, particularly DNA based, have been proving to be very useful in many ways, directly or indirectly, for enhancing the genetic potential of crop plants for improved end-products. Many such markers including restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphisms (AFLPs), simple sequence repeats (SSRs) etc. (Powell *et al.*, 1996; Karp *et al.*, 1997) are now available in the kitty of plant breeders and molecular biologists. Using these markers only genetically neutral regions have been assessed so far in different crop plants for genetic diversity studies. Moreover genetic diversity represented by these markers is not necessarily representative of the whole genome. For appropriate and useful diversity, genome-wide functional regions of the genomes need to be assayed.

Cytochrome P450 are the enzymes involved in polysaccharide biosynthesis, disease related genes, transcription factors, protein kinases and phosphatases, to name a few. A large number of gene families have been observed for these cytochrome P450 in human and other animals. Analogues of these mammalian cytochrome P450s, are commonly observed in angiosperms (Watnabe *et al.*, 2000; Suzuki *et al.*, 2001). Primers corresponding to the functional regions of these human cytochrome P450 can be designed to generate markers and using such markers, it has been observed that these sequences are distributed through out the *Arabidopsis thaliana* genome. Such markers have been termed as P450-based-analogues (PBAs) and have been demonstrated as effective genetic markers for genetic diversity studies reflecting both functional and genome-wide regions (Yamanaka *et al.*, 2003). In the present study we report the utility of these PBAs for assessing intra-specific diversity in linseed (*Linum usitatissimum* L.).

Materials and Methods

Plant material and DNA extraction

Twenty-one linseed (*Linum usitatissimum* L.) cultivars (Table 1) released for different agro-climatic regions of the country over different years were included in this study. DNA from at least ten plants of each of the cultivar was extracted following CTAB method (Saghai-Maroof *et al.*, 1984) with minor changes. The extracted DNA was purified for various contaminants using phenol and was further quantified fluorimetrically. The DNA samples were diluted with Tris: EDTA (1 : 1) buffer to give a final concentration of 10 ng^{-1}. Bulked DNA sample for each cultivar was prepared by pooling equal amount of DNA from each of the plant-wise extracted DNA. These working stocks were kept at 4°C in the refrigerator for regular use whereas the original stocks were stored at -20°C.

PCR amplification and gel analysis

PCR amplification was carried out in 25 μl of total reaction volume containing 1 X PCR buffer (10mM Tris-HCl, pH 9.0; 1.5 mM $MgCl_2$; 50mM KCl; and 0.01% gelatin), 40 ng of genomic DNA, 2.5 mM $MgCl_2$ (M/s Sigma Aldrich), 1 unit of Taq DNA polymerase (M/s Bangalore Genei Pvt. Ltd.), 0.5μM each of forward and reverse primers (Ms Sigma Aldrich) and 0.2mM of dNTP mix (M/S. Bangalore Genei Pvt. Ltd.). PCR amplification for 32 cycles was programmed as: 1 min. at 94°C,

Table1. Linseed cultivars included in the study

S.No.	Cultivar Name	National Identity	Pedigree
1	Padmini	IC-426932	Not available
2	Parvati	IC-305053	Not available
3	Rashmi	IC-426929	Not available
4	Shekhar	IC-305055	Not available
5	Sheela	IC-426928	Not available
6	Shubhra	IC-113108	Mukta x K-2
7	Sweta	IC-113107	Mukta xT-1206
8	T-397	IC-113110	T-491 x T-1193
9	NL-97	IC-305109	Not available
10	Neelum	IC-113109	T-1 x NP (RR) 9
11	Meera	IC-427794	Developed through mutation
12	LC-185	IC-113156	NP (RR) 37 x Kangra Local
13	LC-54	IC-113155	K-2 x Kangra Local
14	Laxmi-27	IC-113105	Neelum/kp1/NP(RR)9/ Neelum)/R1/Neelum/ Afg.8
15	KL-210	IC-420772	Not available
16	KL-31	IC-426935	New River x LC 216
17	KL-1	IC-426936	IC-216 x IC-185
18	Kiran	IC-426925	(R-1 x Afg.8) x R-1
19	Jawahar-552	IC-426926	No. 55 x B-67
20	Gaurav	IC-113111	Sel.-3 x EC-1552
21	Garima	IC-426930	T-126 x Neelum

2 min. at each annealing temperature and 3 min. at 72°C in a PTC-200 thermal cycler (M/s MJ Research). Annealing temperature for each primer-set was as described by Yamanaka *et al.*, (2003).

A total of fifteen primer combinations which included three forward [CYP1A1F (5'-GCCAAGCTTTCTAACAATGC-3'), CYP2B6F (5'-GACTC iTGCTACTCCTGGTT-3'), and CYP2C19F (5'-TCCTTGTGCTCTGTCTCTCA-3')] and five reverse [CYP1A1R (5'-AGGGACATGCTCTGACCATT-3'), CYP2B6R (5'-CGAATACAGAGCTGATGAGT-3'), CYP2C19R (5'-CCATCGATTCTTGGTGTTCT-3'), Heme2B6 (5'-ACCAAGACAAATCCGCTTCCC-3') and Heme2C19 (5'-TCCCACACAAATCCGTTTTCC-3')] were used for PCR amplification (Inui *et al.*, 2000, Kiyokawa *et al.*, 1997). After completion of PCR 2.5 µl loading dye was added to each tube and briefly spun for few seconds for mixing. A 1.5% agarose gel with ethidium bromide (10mg/ ml) was prepared and electrophoresis was carried out at 80 volts for 2 hours. The resolved amplification products were visualized under UV light and photographed with black and white Polaroid film 667.

Data analysis

Jaccard's similarity coefficient values (Jaccard, 1908) for each pair-wise comparison between cultivars were calculated and a similarity coefficient matrix was constructed. This matrix was subjected to unweighted pair-group method for arithmetic averages analysis (UPGMA) to generate a dendrogram using average linkage procedure. Free Tree software was used to construct the dendrogram (Pavelicek *et al.*, 1999) Analysis of molecular variance (AMOVA) was applied on the Euclidean distance matrix between cultivars to partition the total genetic variation (Excoffier *et al.*, 1992). Statistical significance of variances was tested by random permutation, with the number of permutations set at 1000. AMOVA was done using computer software Arlequin (Schneider *et al.*, 2000). Robustness of the clustering pattern was tested using 1000 resamplings using Free Tree software.

Results and Discussion

Two parameters, namely, sample DNA concentration and the concentration of $MgCl_2$ were needed to be standardized to obtain optimum amplification. The $MgCl_2$ concentrations that were tried included 3.0 mM, 3.5 mM, 4.0 mM and 4.5 mM, whereas, different DNA sample concentrations were 20 ng, 30 ng, 40 ng and 50 ng. A combination of these two parameters suggested that 40 ng of genomic DNA and 4.0mM of Mg Cl_2 gave optimum amplification and were used for PCR amplification using all the primers tested.

All the 15 primer-combinations could produce successful amplification, however, the number of bands observed with primers CYP1A1F/ CYP1A1R and CYP2B6F/ CYP2B6R was very low (3 bands in each case). One-third of these primer-combinations did not produce any polymorphic amplicons among the linseed cultivars. The maximum level of polymorphism (57.1%) was observed for primer CYP1A1F/ CYP2B6R followed by the primer CYP2C19F/ Heme2B6 (54.5%) (Table 2). The profiling of 21 linseed cultivars using the later (CYP2C19F/ Heme2B6) along with another primer-combination (CYP2C19F/ CYP1A1R) is presented in Figure 1. On the whole, 25% polymorphism was observed using this set of 15 primer-combinations which was quite low as compared to what has been recorded in *Lycopersicon esculentum* (97.8%), *Solanum tuberosum (94.1%) and Capsicum annuum (87.5%) by Yamanaka et al.* (2003). These workers could get large number of polymorphic bands in a very limited set of accessions within a species. For example, in only six *Lycopersicon esculentum*

Table2. PBA primers used and extent of polymorphism revealed in linseed (*Linum usitatissimum* L.) cultivars

Sr. No.	Primer Combinations	Total number of bands	Number of polymorphic bands	Per cent polymorphism
1.	CYP1A1F/ CYP1A1R	3	0	0.00
2.	CYP1A1F/ CYP2B6R	7	4	57.1
3.	CYP1A1F/ CYP2C19R	14	5	35.7
4.	CYP1A1F/ Heme2B6	13	2	15.4
5.	CYP1A1F/ Heme2C19	6	0	0.00
6.	CYP2B6F/ CYP1A1R	10	4	40.0
7.	CYP2B6F/ CYP2B6R	3	1	33.3
8.	CYP2B6F/ CYP2C19R	12	0	0.00
9.	CYP2B6F/ Heme2B6	11	5	45.4
10.	CYP2B6F/ Heme2C19	12	2	16.7
11.	CYP2C19F/ CYP1A1R	13	5	38.5
12.	CYP2C19F/ CYP2B6R	5	1	0.20
13.	CYP2C19F/ CYP2C19R	10	0	0.00
14.	CYP2C19F/ Heme2B6	11	5	54.5
15.	CYP2C19F/ Heme2C19	6	0	0.00
		136	34	25

accessions, ten out of fifteen primer-combinations were cent per cent polymorphic. In the same set of 21 linseed cultivars we have observed 34 and 43% polymorphism using 20 RAPD and 15 ISSR primers, respectively (unpublished results). It is discernible from these comparative estimates that PBA markers are significantly less polymorphic than RAPD and ISSR markers in linseed.

Jaccard's similarity estimates among the cultivars using these markers ranged from 0.819 (Garima-Kiran) to 0.975 (KL210-Laxmi) with an average of 0.913 ± 0.030 (standard deviation) (Table 3). Average genetic similarity of a cultivar across all other cultivars was found to be minimum for Kiran (0.856). In other words, Kiran was found to be most diverse cultivar followed by Garima. Cultivars LS54 and KL1, each with a value of 0.929 shared the maximum genome homology at the investigated loci with other cultivars followed by for cultivars Shekhar and Neelum (0.928 each).

UPGMA based clustering of all 21 cultivars is shown in Figure 2. Four arbitrary clusters are evident from the dendrogram, Cluster I being the largest with two sub-clusters in it, followed by Cluster II. The other two clusters (III and IV) were equal sized each having two cultivars. All four white-flowered cultivars (Sweta, KL1, Shubhra and LC54) were positioned in Cluster Ia. Two of the three yellow-seeded cultivars (KL1 and KL 210) were also placed in Cluster Ia. The third yellow-seeded cultivar in the study (LC185) grouped with cultivars of Cluster II. The two cultivars of the cluster Ib were very recently released ones. The cultivars of the Cluster IV were very distinct from all other cultivars with a very sound bootstrap support (100). Kiran of this cluster has exotic pedigree in its blood which might have led to this distinctness. The pedigree for the other cultivar of this cluster (NL-97) is not available which, otherwise, would have substantiated this observation. Analysis of molecular

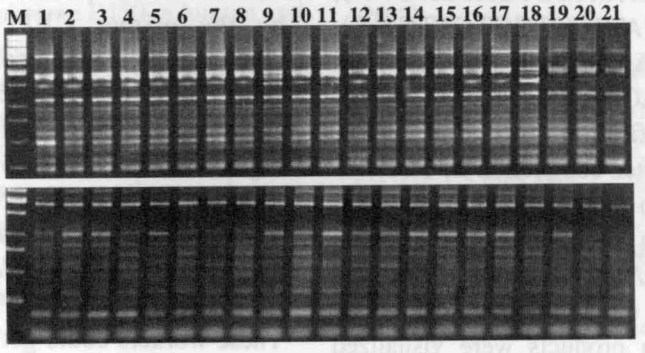

Fig. 1. Profiling of 21 linseed (Linum usitatissimum L.) cultivars using primer combination CYP2C19F/ CYP1A1R (a) and CYP2C19F/ Heme2B6. M is 1 kb molecular weight standard and lanes 1 to 21 are cultivars names as given in Table 1.

Table 3. **Pair-wise Jaccard's (1908) genetic similarity estimates among 21 linseed (*Linum usitatissimum* L.) cultivars**

Padmini	Parvati	Rashmi	Shekhar	Sheela	Shubhra	Sweta	T-397	NL-97	Neelum	Meera	LC185	LC54	Laxmi	KL210	KL3	KL1	Kiran	Jawahar-552	Gaurav	Garima	
Padmini	1.000																				
Parvati	0.921	1.000																			
Rashmi	0.968	0.920	1.000																		
Shekhar	0.952	0.904	0.921	1.000																	
Sheela	0.928	0.926	0.927	0.959	1.000																
Shubhra	0.952	0.919	0.921	0.967	0.943	1.000															
Sweta	0.937	0.873	0.921	0.951	0.942	0.935	1.000														
T-397	0.937	0.903	0.936	0.951	0.942	0.951	0.951	1.000													
NL-97	0.886	0.913	0.886	0.899	0.890	0.899	0.869	0.884	1.000												
Neelum	0.937	0.934	0.936	0.951	0.942	0.967	0.904	0.935	0.913	1.000											
Meera	0.913	0.958	0.898	0.927	0.934	0.927	0.911	0.896	0.921	0.927	1.000										
LC185	0.914	0.927	0.929	0.898	0.888	0.883	0.912	0.927	0.906	0.897	0.904	1.000									
LC54	0.937	0.919	0.936	0.951	0.942	0.935	0.935	0.967	0.898	0.951	0.896	0.943	1.000								
Laxmi	0.921	0.887	0.920	0.951	0.942	0.919	0.934	0.934	0.883	0.950	0.895	0.911	0.967	1.000							
KL210	0.929	0.911	0.928	0.943	0.950	0.927	0.911	0.927	0.891	0.959	0.919	0.889	0.943	0.975	1.000						
KL3	0.913	0.942	0.912	0.911	0.933	0.896	0.911	0.911	0.905	0.926	0.934	0.934	0.942	0.942	0.934	1.000					
KL1	0.953	0.905	0.922	0.968	0.943	0.952	0.952	0.952	0.900	0.936	0.928	0.913	0.952	0.935	0.944	0.927	1.000				
Kiran	0.881	0.864	0.853	0.866	0.828	0.866	0.837	0.837	0.931	0.851	0.871	0.872	0.837	0.822	0.830	0.842	0.867	1.000			
Jawahar-552	0.906	0.903	0.876	0.905	0.895	0.905	0.935	0.919	0.884	0.874	0.911	0.927	0.919	0.888	0.866	0.926	0.921	0.879	1.000		
Gaurav	0.930	0.882	0.929	0.898	0.904	0.913	0.898	0.913	0.878	0.928	0.875	0.906	0.913	0.912	0.920	0.889	0.914	0.859	0.868	1.000	
Garima	0.889	0.917	0.888	0.887	0.908	0.918	0.871	0.902	0.881	0.933	0.909	0.879	0.902	0.901	0.925	0.924	0.903	0.820	0.871	0.926	1.000
Average similarity±	0.925 ±	0.911 ±	0.916 ±	0.928 ±	0.923 ±	0.925 ±	0.914 ±	0.924 ±	0.896 ±	0.928 ±	0.913 ±	0.908 ±	0.929 ±	0.919 ±	0.921 ±	0.918 ±	0.929 ±	0.856 ±	0.899 ±	0.903 ±	0.898 ±
Standard deviation	0.923	0.911	0.025	0.029	0.030	0.026	0.031	0.029	0.015	0.028	0.020	0.019	0.029	0.034	0.032	0.023	0.024	0.025	0.021	0.021	0.026

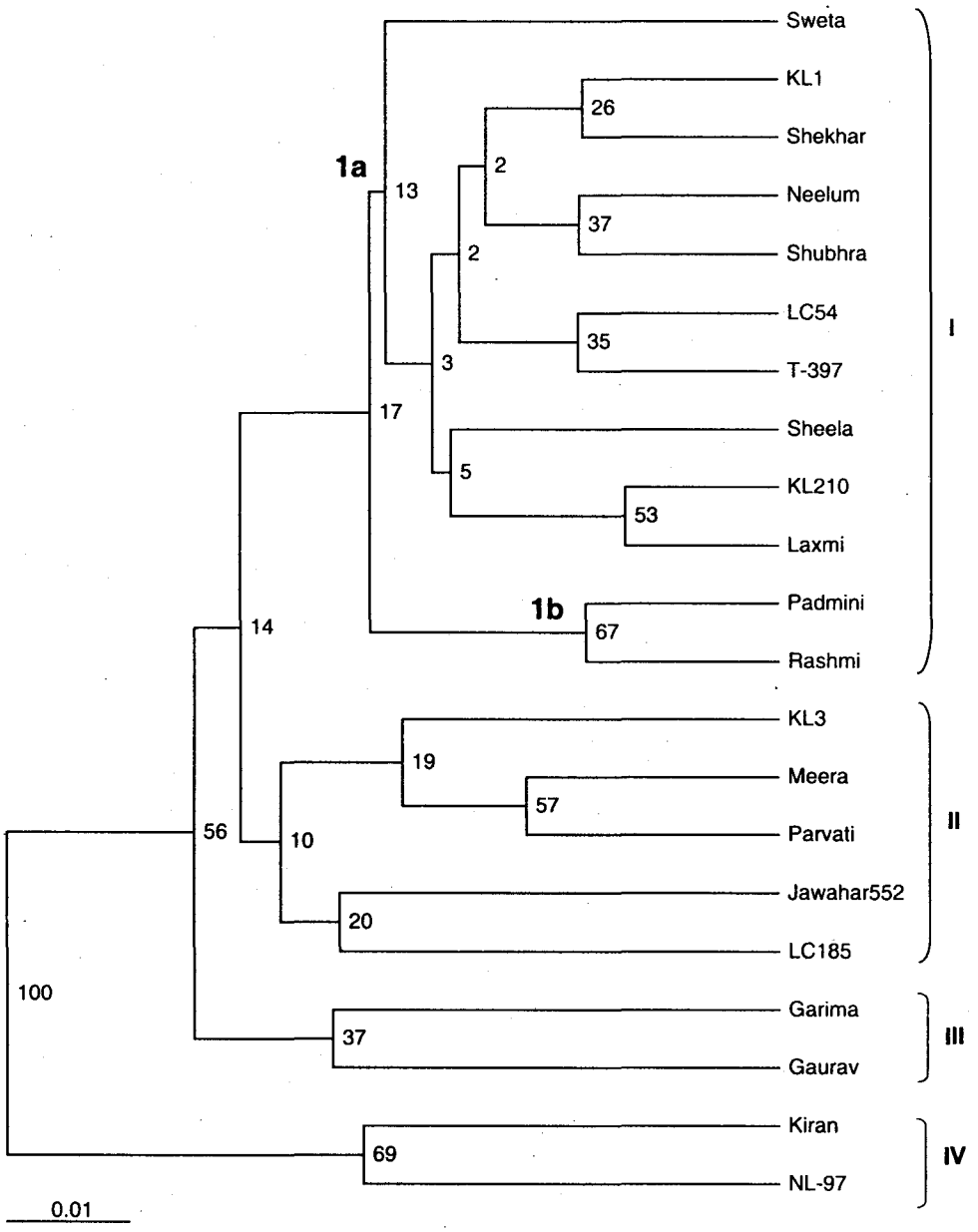

Fig. 2. UPGMA dendrogram for 21 linseed (Linum usitatissimum L.) cultivars using 136 markers generated by 15 PBA primers and Jaccard's similarity estimates. The numbers inside the figure denote per cent bootstrap support for each node.

variance (AMOVA) was done by grouping these cultivars into populations based on year of release (pre-1990 and post-1990), flower colour (white and blue), seed size (large, medium and small), seed colour (yellow and brown) and yield (less than 1000 kg/hectare and more than 1000 kg/hectare). Most of the variation in all these cases could be attributed to within population variation as the variation among these created groups was negligible.

The polymorphism level was quite low in the linseed cultivars using PBAs markers. The utility of these markers need to be assessed on crop-by-crop basis as these have revealed high polymorphism in other crops like chilli, tomato, eggplant etc. Low level of polymorphism revealed by these markers in intra-specific investigation in linseed suggests their utility for higher taxonomic level studies, which needs to be proved in the genus *Linum*.

References

Excoffier L, P Smouse, J Quattro (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479-491.

Inui H, T Komada, Y Ohkawa, H Ohkawa (2000) Herbicide metabolism and cross-tolerance in transgenic potato plants co-expressing human CYP1A1, CYP2B6 and CYP2C19. Pestic. *Biochem. Physiol.* **66:** 116-129.

Jaccard P (1908) Nouvelles recherchés sur la distribution florale. *Bull. Soc. Vaud. Sci. Nat.* **44:** 223-270.

Karp A, S Kresovich, KV Bhat, WG Ayad, T Hodgkin (1997) Molecular tools in plant genetic resources conservation: a guide to the technologies. International Plant Genetic Resources Institute, Rome.

Kiyokawa S, M Ohbayashi, Y Shimada, Y Kikuchi (1997) PCR-amplification of sequences encoding the heme-binding region of plant cytochrome P450. *Plant Biotech.* **14:** 175-178.

Pavelicek A, S Hrda and J Flegr (1999). FreeTree-Freeware program for construction of phylogenetic trees based on the basis of distance data and bootstrap/ jackknife analysis of the tree robustness. Application in the RAPD analysis of the genus *Frenkelia. Folia Biologica (Praha).* **45**: 97-99.

Powell W, M Morgante, C Andre, M Hanafey, J Vogel, S Tingey and A Rafalski (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol. Breed.* **2:** 225-238.

Saghai-Maroof MA, KM Soliman, RA Jorgensen, RW Allard (1984) Ribosomal DNA spacer-length polymorphism in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci. USA.* **81:** 8014-8018.

Schneider S, D Roessli, L Excoffier (2000) Arlequin ver. 2000: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland. (http://anthro.unige.ch/arlequin).

Suzuki E, JA Watanabe, KN Watanabe (2001) Possibility of genetic diversity study in plants using sequence information of mammalian xenobitics metabolizing cytochrome P450 genes (In Japanese). *Breed. Res.* **3:** 187.

Watanabe KN and M Iwanaga (1999) Plant genetic resources and its global contribution. *Plant Biotech.* **16:** 7-13.

Yamanaka S, E Suzuki, M Tanaka, Y Takeda, JA Watanabe, KN Watanabe (2003) Assessment of cytochrome P450 sequences offers a useful tool for determining genetic diversity in higher plant species. *Theor Appl Genet.* **108:** 1-9.