Molecular marker-based characterization in candidate plus trees of *Pongamia pinnata*, a potential biodiesel legume

Vigya Kesari¹, Vinod Madurai Sathyanarayana², Ajay Parida² and Latha Rangan¹*

¹ Department of Biotechnology, Indian Institute of Technology, Guwahati, Assam 781039, India
² M S Swaminathan Research Foundation, 3rd Cross Street, Institutional Area, Taramani, Chennai 600113, India

Received: 7 August 2010; Revised: 14 September 2010; Accepted: 5 October 2010; Published: 11 October 2010

Citation details: Kesari V, Madurai Sathyanarayana V, Parida A, Rangan L. 2010. Molecular marker-based characterization in candidate plus trees of *Pongamia pinnata*, a potential biodiesel legume. AoB PLANTS 2010: plq017, doi:10.1093/aobpla/plq017

Abstract

**Background and aims**
*Pongamia pinnata*, a legume tree, has many traditional uses and is a potential biodiesel plant. Despite its importance and the availability of appropriate molecular genetic tools, the full potential of *Pongamia* is yet to be realized. The objective of this study was to assess genetic diversity among 10 systematically characterized candidate plus trees (CPTs) of *P. pinnata* from North Guwahati.

**Methodology**
The application and informativeness of polymerase chain reaction-based molecular markers [random amplified polymorphic DNA (RAPD), inter-simple sequence repeat (ISSR) and amplified fragment length polymorphism (AFLP)] to assess the genetic variability and relatedness among 10 CPTs of *P. pinnata* were investigated.

**Principal results**
Polymorphism rates of 10.48, 10.08 and 100% were achieved using 18 RAPD, 12 ISSR and 4 AFLP primer combinations, respectively. Polymorphic information content (PIC) varied in the range 0.33–0.49, 0.18–0.49 and 0.26–0.34 for RAPD, ISSR and AFLP markers, respectively, whereas the corresponding average marker index (MI) values for the above markers were 7.48, 6.69 and 30.75. Based on Nei's gene diversity and Shannon's information index, inter-population diversity (hₒₒ) was highest when compared with intra-population diversity (hₚₚ) and the gene flow (Nₑₑ) ranged from a moderate value of 0.607 to a high value of 6.287 for the three DNA markers. Clustering of individuals was not similar when RAPD- and ISSR-derived dendrogram analyses were compared with that of AFLP. The Mantel test cophenetic correlation coefficient was higher for AFLP (r = 0.98) than for ISSR (r = 0.73) and RAPD (r = 0.84). Molecular markers discriminated the individuals efficiently and generated a high similarity in dendrogram topologies derived using unweighted pair-group arithmetic average, although some differences were observed. The three-dimensional scaling by principal coordinate analysis supported the result of clustering.

**Conclusions**
Comparing the results obtained with the three DNA markers, AFLP indicated higher efficiency for estimating the levels of genetic diversity and proved to be reliable for fingerprinting, mapping and diversity studies in *Pongamia* in view of their suitability for energy production purposes.

* Corresponding author’s e-mail address: latha_rangan@yahoo.com
Introduction

Pongamia pinnata also referred to as Pongam, is a biodiesel tree legume well adapted to arid zones and has many traditional uses (Meera et al. 2003; Brijesh et al. 2006; Scott et al. 2008; Muktai et al. 2009). The pongam tree has the rare property of producing seeds with high oil content (Sarma et al. 2005; Kesari et al. 2010). Despite the importance of this versatile plant and the availability of appropriate molecular genetic tools, the full potential of Pongamia is far from being realized. The natural constraints that limit its large-scale production and availability, to meet the demand for biodiesel production, are its long gestation period (4–7 years), plant height, seed storage behaviour, insect pests, and the seed oil yield and quality (NOVOD 2010). Until recently, the identification of elite individuals of P. pinnata was mainly described in terms of morphological and agronomic traits which are known to be deeply influenced by environmental factors (Kaushik et al. 2007; Kesari et al. 2008). The cross-pollinating nature of P. pinnata contributes to its wide germplasm biodiversity. Thus, it becomes an important step to examine the genetic variations among naturally growing elite individuals of P. pinnata at inter- and intra-population levels, and to prepare strategies for its specific exploitation by plant breeders in promoting it as a versatile biodiesel plant.

With the large array of molecular analytical techniques available, it has become possible to provide an accurate and unambiguous tool for the evaluation of genetic diversity and identification of germplasm (Meudt and Clarke 2006; Simmons et al. 2007; Li et al. 2008). In the last decade, several polymerase chain reaction (PCR)-based DNA marker classes, viz. random amplified polymorphic DNA (RAPD), inter-simple sequence repeat (ISSR) and amplified fragment length polymorphism (AFLP), contributed to this goal, and may extend and complement the assessment of genetic variation within the species based on morphological and polypeptide descriptions, providing more speed, accurate and detailed information (Williams et al. 1990; Zabeau and Vos 1993; Zietkiewicz et al. 1994; Surekha and Larson 2005). In addition, it is independent of the growth, stage, season, location and agronomic practices, which gives them a high value (Lombard et al. 2001). RAPD markers have been widely used in plant research for phylogenetic studies, genome mapping and genetic variation analysis both at intra- and inter-population levels (Li et al. 2008). The technique has several advantages including simplicity, low cost, speed and lack of requirement for DNA sequence information (Williams et al. 1990; Lopes et al. 2007). However, RAPD technology has several limitations including dominance, uncertain locus homology, sensitivity and reliability. Inter-simple sequence repeat represents the marker of choice for varietal identification studies as they are transferable, hypervariable, highly polymorphic, multiallelic dominant markers, relatively simple to interpret and show high information content (Souframanien and Gopalkrishna 2004). The AFLP technique has been used for DNA fingerprinting in a number of tree species such as almond (Sorkheh et al. 2007), cashew (Archak et al. 2003) and Jatropha (Tatikonda et al. 2009). The choice of primer pairs used in AFLP amplification is critical for obtaining high levels of polymorphism. Therefore, it would be useful if few informative markers are identified such that they amplify markers with high levels of polymorphism, which are representative of the whole genome.

There is a strong need to assess and document the extent of genetic diversity in naturally growing, systematically characterized P. pinnata populations to expedite its use in different germplasm-related studies and breeding programmes. In continuation of our studies on candidate plus trees (CPTs), the present study was undertaken with the objective of screening the untapped genetic variability or relatedness among the 10 CPTs of P. pinnata tagged from North Guwahati, Assam, at the inter- and intra-population levels using PCR-based DNA markers (RAPD, ISSR and AFLP). Candidate plus trees are individual trees of P. pinnata possessing superior morphological and reproductive characters compared with other individuals of the same species identified on the basis of morphometric markers (Kesari et al. 2008). The study further involves comparing the informativeness of RAPD, ISSR and AFLP assays employed at the level of populations and individuals in order to analyse the variability studies in the P. pinnata genome. So far, this is the first study that deals with the systematic characterization of elite individuals for further genetic improvement of important biodiesel tree species.

Materials and methods

Plant material

The germplasm used in this study for genetic diversity studies were young seeds, collected during the month of September 2007 from 10 different tagged CPTs of P. pinnata [North Guwahati Pongamia pinnata (NGPP) 26–30 and NGPP 46–50]. The 10 CPTs identified, based on morphometric traits, belong to two different naturally growing populations, viz. 6 and 10 occurring at Sila Forest Range, North Guwahati, Assam (latitude 26°14′6″N and longitude 91°41′28″E) (Kesari et al. 2008). All the seed samples were appropriately stored at −20 °C for further analysis.
Genomic DNA extraction

Genomic DNA extraction for genetic diversity studies was performed by modified sodium dodecyl sulphate method (Kesari et al. 2009). The quality and quantity of the extracted DNA were confirmed to be consistent both spectrophotometrically and by running the extracted DNA in a 0.8% agarose gel stained with ethidium bromide (0.5 μg mL⁻¹). At least two independent DNA preparations for each individual were made, and the quantity and quality of DNA samples were estimated.

RAPD analysis

For RAPD fingerprinting, PCR amplification of the genomic DNA was carried out using 20 arbitrary decamer oligonucleotide primers (Operon Tech, USA). The primer sequence information is provided in the Additional information, Table 1. Each 20 μL of reaction mixture contained 50 ng μL⁻¹ template DNA, 1 × assay buffer (100 mM Tris sulfonic acid, pH 8.8, 15 mM MgCl₂, 500 mM KCl and 0.1% gelatin), 0.2 mM each deoxyribonucleotide triphosphate (B’LGenei, India), 5 pM each primer and 0.5 U of Taq polymerase (B’LGenei, India). The reaction was performed in 0.2-mL microfuge tubes (Dialabs, UK). PCR amplification was carried out in a Mini Thermal Cycler (Applied Biosystems 9700) programmed for 35 cycles. The first amplification cycle consisted of an initial denaturation step of 5 min at 94 °C, followed by 34 cycles of 45 s at 94 °C, annealing temperature of 1 min at 32 °C, and extension at 72 °C for 1 min 30 s. An additional cycle of 5 min at 72 °C was used for primer extension. The amplification products were electrophoresed on 1.5% agarose gels in 0.5× Tris borate EDTA (10× stocks contained 0.8 M Tris, 0.8 M boric acid, 0.5 M EDTA). A 1 kb DNA ladder (Life Technologies, USA) was included as a size marker, and the amplified products were detected with ethidium bromide staining (0.5 μg mL⁻¹). The gels were photographed under a UV transilluminator.

ISSR analysis

PCR amplification was carried out for 20 ISSR primers (Additional information, Table 1), using a similar method to that used for RAPD analysis with the final volume of the reaction mixture being 20 μL. The steps of temperature cycling were as follows: 94 °C for 4 min, followed by 35 cycles of 45 s at 94 °C, 1 min with varied temperatures as per the melting temperature of the ISSR primers used, 1 min 30 s at 72 °C and 10 min final extension step at 72 °C. The amplified products were visualized in a 1.5% agarose gel containing ethidium bromide and photographed for further analysis.

AFLP analysis

Reagents from the Applied Biosystems plant mapping kit consisting of the ligation/preselective amplification module and selective primer kits were used, and assay performed as per the manufacturer’s protocol followed by Vos et al. (1995) with minor modifications. DNA (500 ng) was double-digested with restriction endonucleases EcoRI/Msel (New England Biolabs) for 2 h at 37 °C and ligated to adapters using T4 DNA ligase (New England Biolabs). The digested DNA mixture was diluted (10-fold) and used for pre-selective amplification with pre-selective primer pair. The pre-selective amplification was performed at 72 °C for 2 min followed by 20 cycles of denaturation (at 94 °C for 20 s), primer annealing (at 56 °C for 30 s) and primer extension (at 72 °C for 2 min). The reaction mixture was finally maintained at 60 °C for 30 min. An aliquot of the pre-selective PCR product was electrophoresed on a 1.5% agarose gel and checked for amplification. It was then diluted (1:20) with DNase-free water and used as a template for the selective amplification.

The selective amplification was performed using primers from the AFLP selective primer kit. The selective amplification involved the following thermal cycling conditions: denaturation at 94 °C for 2 min followed by 11 cycles of 94 °C for 20 s, 66 °C for 30 s and 72 °C for 2 min. The annealing temperature was reduced by 1 °C every cycle till it reached 56 °C. This was followed by another 20 cycles of amplification at 94 °C for 20 s, 56 °C for 30 s and 72 °C for 2 min, with a final extension at 60 °C for 30 min.

Selectively amplified product (1 μL) was mixed with 0.5 μL of the GeneScan 500 ROX internal size standard (Applied Biosystems P/N 402985) and 8.5 μL of Hi-Di Formamide (Applied Biosystems P/N 4311320). The mixture was then denatured prior to separation by capillary gel electrophoresis on an automated DNA sequencer (ABI 3130; Applied Biosystems). The electropherograms generated by the sequencer were interpreted with GeneScan® analysis software (PE Applied Biosystems, Foster City, CA). Genotyper® analysis software (PE Applied Biosystems) was then used to create a list of fragments detected in each lane by fragment size. Fragments sized from 50 to 500 bp with a peak height >50 bp in the electropherogram were retained for subsequent analysis.

Data analysis

For all three types of marker system, duplicate samples from each individual were tested, and only clear, unambiguous and reproducible bands amplified in both cases were considered for scoring the data. The number of polymorphic and monomorphic amplification products...
was determined for each primer for 10 individuals. To avoid taxonomic weighing, the intensity of bands was not taken into consideration, and only the presence of band was taken as an indicative. To compare the efficiency of primers, polymorphic information content (PIC), as a marker discrimination power, was computed using the formula $\text{PIC} = 1 - \sum p^2_i$, where $p_i$ is the frequency of the $i$th allele at a given locus (Anderson et al. 1993). The PIC values are commonly used in genetics as a measure of polymorphism for a marker locus using linkage analysis. Marker index (MI) defined as the product of the polymorphism percentage and PIC is used to estimate the overall utility of each marker system and was calculated according to Sorkheh et al. (2007).

The basic parameters for genetic diversity were calculated in the POPGENE application (Yeh et al. 1999). The polymorphism of amplification products ($P$), the number of observed alleles ($n_o$), the mean number of effective alleles ($n_e$), the mean Nei’s gene diversity index ($h$), the Shannon index ($I$) and the level of gene flow ($N_m$) (Slatkin and Barton 1989; McDermott and McDonald 1993) were determined. Within-population diversity ($h_{pop}$), total gene diversity ($h_{sp}$) and inter-population differentiation ($G_{st}$) (Nei 1973) were calculated using the POPGENE software.

Level of similarity among individuals was established as a percentage of polymorphic bands, and a matrix of genetic similarity compiled using Dice’s coefficient (Dice 1945). By applying the unweighted pair-group arithmetic average (UPGMA) method (Sneath and Sokal 1973) using the SHAN subroutine through the NTSYS-pc version 2.2 (Rohlf 2002), dendrograms representing the genetic relationship among the 10 CPTs of *Pongamia* were generated. The correlation between the original similarity indices and cophenetic values was calculated, and Mantel’s test (Mantel 1967) was performed using 250 permutations to check the goodness of fit of the CPTs to a specific cluster in the UPGMA similarity matrix. Further, principal coordinate analysis (PCA) was undertaken for the families with modules STAND, CORR and EIGEN of NTSYS-pc (Rohlf 2002) using the Euclidean distances derived from the standardized values using the NTSYS software.

**RESULTS**

**PCR fingerprinting studies**

**RAPD analysis** Following preliminary screening, 18 primers were chosen for the evaluation of genetic diversity among CPTs of *P. pinnata*. The remaining tested primers did not generate any amplification products or stable band patterns. Specific DNA banding patterns were observed with all the 18 primers used. The number of amplification products and the number of polymorphic fragments for the primers are given in Table 1. A comparison of all the 10 CPTs analysed showed a total of 210 amplified products and the polymorphism index ($P$) value of 10.48% (Table 1). Primers varied in their ability to detect variation both within and between populations as shown in Table 1. Of the 18 RAPD primers, 8 of them produced monomorphic fragments for 10 CPTs (OPO 08, OPAM 20, OPAN 05, OPAA 01, OPAB 05, OPX 20, OPAJ 19 and OPAF 02) and the rest produced polymorphic fragments. Band numbers per primer ranged from 3 (OPAM 20) to 21 (OPX 20) with an average of 11.67 bands per primer. The polymorphism for 10 CPTs ranged from a maximum of 3 per primer (OPC 07, OPAN 01, OPAO 01 and OPAH 13) to a minimum of zero per primer (OPO 8, OPAM 20, OPAN 05, OPAA 01, OPAB 05, OPAF 02, OPAJ 19 and OPX 20) (Table 1). Typical PCR amplification products generated by the representative RAPD primer are presented in Fig. 1A. The amplified product ranged between 0.35 and 1.8 kb in size. The PIC values for the polymorphic RAPD primers ranged from 0.32 (OPL 11) to 0.49 (OPAO 01), whereas MI ranged from 2.91 (OPL 11) to 11.38 (OPAO 01). All the selected polymorphic primers had PIC values greater than 0.32 and the average PIC value for the amplification products was 0.43.

**ISSR analysis** Amplification of ISSR markers was performed using 20 primers, of which 12 primers gave reproducible and good quality banding patterns. The 12 ISSR primers produced a total of 129 fragments with size ranging from 0.2 to 0.85 kb. Out of 129 ISSR loci, 14 bands were polymorphic among 10 CPTs as revealed by six primers (Table 1), whereas the other six primers (HB 14, HB 15, P 6, P 8, 809 and 824) produced monomorphic fragments. The percentage of ISSR polymorphism for these 10 CPTs is thus a maximum of 10.08. The average number of bands per primer was 10.75. A survey of different ISSR primers used (Table 1) indicated that for each primer, the number of visible bands ranged from 6 to 17 (obtained by primer 824 and HB 12, respectively). Primers 825 (ACACACACACA CACACT) and 826 (ACACACACACACACAC) exhibited the highest level of variability, and the polymorphism rate was 25% in both cases. Typical polymorphic ISSR fingerprint using the primer 825 is shown in Fig. 1B. The PIC values for polymorphic ISSR primers ranged from 0.18 (818) to 0.50 (CAG), with an average of 0.39, whereas MI ranged from 2.25 (818) to 12.33 (825).
AFLP analysis DNA fingerprinting profiles for 10 CPTs of *P. pinnata* exhibited a high degree of polymorphism with the AFLP marker. The number of polymorphic fragments generated by each primer pair combinations in AFLP varied from 26 (MseI-CAT/EcoRI-ACA) to 124 (MseI-CAT/EcoRI-ACC), with an average of 62.25 fragments (Table 2). A total of 249 bands were amplified ranging from 50 to 500 bp and were 100 % polymorphic. The highest PIC value (0.34) was observed for the primer combination MSeI-CAA/EcoRI-ACT and the lowest PIC

Table 1 Degree of polymorphism and PIC for RAPD and ISSR primers in 10 CPTs of *P. pinnata*

| Markers | Primer code | Total number of bands | Number of polymorphic bands | POL (%) | PIC | MI |
|---------|-------------|-----------------------|----------------------------|---------|-----|----|
| RAPD    | OPC 07      | 14                    | 3                          | 21.43   | 0.42| 9.00|
|         | OPL 11      | 11                    | 1                          | 9.09    | 0.32| 2.91|
|         | OPO 08      | 13 (M)                | 0                          | –       | –   | –  |
|         | OPAH 15     | 9                     | 2                          | 22.22   | 0.42| 9.33|
|         | OPAM 20     | 3 (M)                 | 0                          | –       | –   | –  |
|         | OPAN 01     | 19                    | 3                          | 15.79   | 0.46| 7.26|
|         | OPAO 01     | 13                    | 3                          | 23.08   | 0.49| 11.38|
|         | OPAP 20     | 8                     | 2                          | 25.00   | 0.33| 8.25|
|         | OPAN 05     | 8 (M)                 | 0                          | –       | –   | –  |
|         | OPAP 10     | 15                    | 1                          | 6.67    | 0.48| 3.20|
|         | OPAA 01     | 8 (M)                 | 0                          | –       | –   | –  |
|         | OPAB 01     | 13                    | 2                          | 15.38   | 0.45| 6.92|
|         | OPAB 05     | 5 (M)                 | 0                          | –       | –   | –  |
|         | OPAB 14     | 13                    | 2                          | 15.38   | 0.45| 6.92|
|         | OPAH 13     | 14                    | 3                          | 21.43   | 0.45| 9.58|
|         | OPAF 02     | 5 (M)                 | 0                          | –       | –   | –  |
|         | OPAJ 19     | 18 (M)                | 0                          | –       | –   | –  |
|         | OPX 20      | 21 (M)                | 0                          | –       | –   | –  |
| Total   | 210         | 22                    | 10.48                      | –       | –   | –  |
| Mean    | 11.67       | 2.2                   | –                          | –       | –   | –  |
| ISSR    | HB13        | 12                    | 2                          | 16.67   | 0.49| 8.17|
|         | HB14        | 8 (M)                 | 0                          | –       | –   | –  |
|         | HB15        | 8 (M)                 | 0                          | –       | –   | –  |
|         | P8          | 9 (M)                 | 0                          | –       | –   | –  |
|         | 824         | 6 (M)                 | 0                          | –       | –   | –  |
|         | 825         | 12                    | 3                          | 25.00   | 0.49| 12.33|
|         | 826         | 8                     | 2                          | 25.00   | 0.25| 6.25|
|         | P6          | 10 (M)                | 0                          | –       | –   | –  |
|         | 809         | 10 (M)                | 0                          | –       | –   | –  |
|         | 818         | 16                    | 2                          | 12.50   | 0.18| 2.25|
|         | HB12        | 17                    | 3                          | 17.65   | 0.41| 7.29|
|         | (CAG)5      | 13                    | 1                          | 7.69    | 0.50| 3.85|
| Total   | 129         | 13                    | 10.08                      | –       | –   | –  |
| Mean    | 10.75       | 2.17                  | –                          | –       | –   | –  |

POL, polymorphism; M, monomorphic marker; PIC, average polymorphic information content for polymorphic bands; MI, marker index = POL (%) × PIC.
value (0.26) was recorded for the primer combination MSeI-CAC/EcoRI-ACA (Table 2). Average PIC value per primer combination was 0.31. The MI values ranged from 26 to 34 with an average of 30.75 per primer combination.

**Gene diversity**

Moderate levels of genetic diversity within and between populations were observed based on RAPD and ISSR, whereas higher extent of genetic diversity was observed based on AFLP for 10 CPTs of *P. pinnata*. At the population level, the average values of $n_a$, $n_e$, $h_{pop}$ and percentage of polymorphism ranged from 1.054 to 1.745, 1.042 to 1.370, 0.023 to 0.229 and 5.43 to 74.50 %, respectively, using three molecular markers (RAPD, ISSR and AFLP) (Table 3). Similarly, at the species level, $n_a$, $n_e$, $h_{sp}$ and percentage of polymorphism ranged from 1.101 to 1.992, 1.064 to 1.389, 0.038 to 0.247 and 10.08 to 100 %, respectively. The mean Shannon’s indices ($I$) for two population (POP 6 and POP 10) and

---

**Fig. 1** RAPD and ISSR polymorphic profiles for 10 CPTs of *P. pinnata* belonging to populations 6 and 10. Lane M, 1 kb DNA ladder; lanes 1–5, NGPP 26–NGPP 30; lanes 6–10, NGPP 46–NGPP 50. (A) RAPD (OPAH 13); (B) ISSR (825).
at the species level based on RAPD and ISSR displayed similar estimates of 0.035–0.033 and 0.057–0.061, respectively (Table 3). The highest level of mean Shannon’s indices at the population level (0.356) and at the species level (0.395) was obtained based on the AFLP marker (Table 3). On an average, for all the three markers used in the current study, POP 10 exhibited lowest level of genetic variation based on Nei’s genetic diversity.

The average coefficient of genetic differentiation ($G_{st}$) was 0.377, 0.452 and 0.074, respectively, for RAPD, ISSR and AFLP among the two populations of *P. pinnata* (Table 4). Genetic differentiation between the populations was very low based on AFLP data, suggesting extensive gene flow. The significant differentiation between populations of *P. pinnata* was also revealed in the estimates of gene flow ($N_m$). The value of $N_m$ obtained was lowest for ISSR (0.607), highest for AFLP (6.2874) and moderate for RAPD (0.8275) (Table 4). Thus, there seems to be a moderate gene flow between the two populations of *P. pinnata* studied based on RAPD and ISSR data. However, AFLP data showed higher level of gene flow between populations.

### Genetic relationship as revealed by RAPD, ISSR and AFLP

To evaluate the effectiveness of different markers for the study of genetic relationship in CPTs, genetic similarity coefficient was calculated. The value derived from the data of RAPD marker analysis varied from 0.111 between NGPP 26 and NGPP 46 to 0.857 between NGPP 27 and NGPP 28, and NGPP 46 and NGPP 47. For the ISSR marker, the values varied from 0.267 between NGPP 28 and NGPP 49 to 0.875 between NGPP 47 and NGPP 48. In case of AFLP, the genetic similarity values varied from a maximum of 0.714 (between NGPP 28 and NGPP 49) to a minimum of 0.047 (between NGPP 26 and NGPP 29). Finally, 10 RAPD, 6 ISSR and 4 AFLP primer

### Table 2 Degree of polymorphism and PIC for AFLP primers applied to 10 CPTs of *P. pinnata*

| Primer combinations | Total number of bands | POL | PIC | MI | Number | % |
|---------------------|-----------------------|-----|-----|----|--------|---|
| Msel-CAA/ EcoRI-ACT | 50                    | 50  | 100 | 0.34 | 100    | 34|
| Msel-CAC/ EcoRI-ACA | 26                    | 26  | 100 | 0.26 | 100    | 26|
| Msel-CAG/ EcoRI-AAC | 49                    | 49  | 100 | 0.30 | 100    | 30|
| Msel-CAT/ EcoRI-ACC | 124                   | 124 | 100 | 0.33 | 100    | 33|

POL, polymorphism; PIC, average polymorphic information content for polymorphic bands; MI, marker index = POL (%) × PIC.

### Table 3 Mean genetic data of two populations having a total of 10 CPTs of *P. pinnata*

| Markers | Populations | $n_a$ | $n_e$ | $h$ | $I$ | $p$ | POL (%) |
|---------|-------------|-------|-------|-----|-----|-----|---------|
| RAPD    | POP 6       | 1.087 | ± 0.283 | 1.016 | ± 0.228 | 0.038 | ± 0.124 | 0.054 | ± 0.177 | 18 | 8.7 |
|         | POP 10      | 1.039 | ± 0.193 | 1.016 | ± 0.092 | 0.010 | ± 0.056 | 0.017 | ± 0.087 | 8  | 3.86 |
|         | Mean        | 1.063 | ± 0.238 | 1.042 | ± 0.160 | 0.024 | ± 0.090 | 0.035 | ± 0.132 | 13 | 6.28 |
|         | At species level | 1.105 | ± 0.307 | 1.064 | ± 0.206 | 0.038 | ± 0.116 | 0.057 | ± 0.171 | 22 | 10.63 |
| ISSR    | POP 6       | 1.054 | ± 0.227 | 1.046 | ± 0.188 | 0.024 | ± 0.101 | 0.034 | ± 0.144 | 7  | 5.43 |
|         | POP 10      | 1.054 | ± 0.227 | 1.040 | ± 0.178 | 0.022 | ± 0.096 | 0.032 | ± 0.137 | 7  | 5.43 |
|         | Mean        | 1.054 | ± 0.227 | 1.042 | ± 0.183 | 0.023 | ± 0.098 | 0.033 | ± 0.141 | 7  | 5.43 |
|         | At species level | 1.101 | ± 0.302 | 1.076 | ± 0.241 | 0.042 | ± 0.129 | 0.061 | ± 0.185 | 13 | 10.08 |
| AFLP    | POP 6       | 1.807 | ± 0.395 | 1.399 | ± 0.325 | 0.248 | ± 0.165 | 0.385 | ± 0.229 | 201| 80.72 |
|         | POP 10      | 1.683 | ± 0.466 | 1.340 | ± 0.334 | 0.210 | ± 0.177 | 0.326 | ± 0.253 | 170| 68.27 |
|         | Mean        | 1.745 | ± 0.431 | 1.370 | ± 0.330 | 0.229 | ± 0.171 | 0.356 | ± 0.241 | 186| 74.50 |
|         | At species level | 1.992 | ± 0.089 | 1.389 | ± 0.308 | 0.247 | ± 0.149 | 0.395 | ± 0.183 | 249| 100 |

$n_a$, observed number of alleles; $n_e$, effective number of alleles; $h$, Nei’s (1973) gene diversity; $I$, Shannon’s information index; $p$, number of polymorphic bands; POL (%), percentage polymorphism.
combinations identified to be polymorphic for 10 CPTs were selected for cluster analysis through the UPGMA dendrogram constructed using SHAN neighbour-joining tree separately. The genetic relationships among 10 CPTs of *P. pinnata* belonging to two populations were analysed on the basis of Dice’s genetic distance (Dice 1945). The UPGMA clustering algorithm from RAPD grouped the individuals into two major clusters at a similarity index value of 0.26 (Fig. 2A). Cluster I consists of individuals belonging to population 6 (NGPP 26–30), whereas individuals belonging to population 10 (NGPP 46–50) were grouped in cluster II. The UPGMA cluster from ISSR marker analysis separated the 10 CPTs similarly into two major groups at a similarity index value of 0.55 (Fig. 2B), which was similar to RAPD analysis. According to the cluster analysis on the basis of RAPD and ISSR markers, 10 CPTs were classified into 2 major clusters, but the intra-positioning of the individuals in the dendrogram was different.

The relationship among individual CPTs belonging to two populations was also visualized in the UPGMA dendrogram generated from AFLP analysis (Fig. 2C). Five distinct clusters were formed. The data obtained based on AFLP analysis in the present study discriminated most individuals and grouped them together, though belonging to two different populations (such as NGPP 26, 48; NGPP 28, 46 and NGPP 49, 30 more closely related), hinting that inter-population differences were more significant compared with intra-population ones (Fig. 2C). The PCA analysis supported the results of clustering (Additional information, Fig. 1A–C). The Mantel method used to compare the similarity matrices produced correlation coefficients that were statistically significant for each of the three marker systems used independently, viz. RAPD, ISSR and AFLP. The cophenetic correlation coefficient between the dendrogram and the original similarity matrix was large and significant for RAPD ($r = 0.84$), ISSR ($r = 0.73$) and AFLP ($r = 0.98$), giving a good degree of confidence in the association obtained for the CPTs.

---

### Table 4 Genetic population structure and estimate of gene flow within the populations of CPT of *P. pinnata*

|        | $h_{sp}$   | $h_{pop}$  | $G_{st}(1 - h_{pop}/h_{sp})$ | $N_m$  |
|--------|------------|------------|-----------------------------|--------|
| RAPD   | 0.038 ± 0.0137 | 0.024 ± 0.0066 | 0.377 | 0.828 |
| ISSR   | 0.042 ± 0.0166 | 0.023 ± 0.0070 | 0.452 | 0.607 |
| AFLP   | 0.247 ± 0.0222 | 0.229 ± 0.0202 | 0.074 | 6.287 |

$h_{sp}$, total variability; $h_{pop}$, variability within population; $G_{st}$, inter-population differentiation; $N_m$, estimate of gene flow from $G_{st}$ (Slatkin and Barton 1989). $N_m = 0.5(1 - G_{st})/G_{st}$ (McDermott and McDonald 1993).
Discussion

For the effective genetic improvement programme of *P. pinnata*, maintenance of high genetic diversity is one of the most important issues as the species have immense multipurpose importance, yet genetically uncharacterized. A wide range of variability exists in *Pongamia* with respect to seed and oil traits (Kaushik et al. 2007; Kesari et al. 2008; Mukta et al. 2009). Morphological markers for identifying the individuals are limited in number and they do not often reflect genetic relationships because of interaction with the environment, epistasis and the largely unknown genetic control of the traits (Smith and Smith 1989). In contrast, DNA markers are found in abundance and are not influenced by the environment or developmental stage of a plant, making them an ideal tool for genetic relationship studies (Reddy et al. 2002). In the case of *Pongamia*, a few recent studies have been conducted to assess molecular diversity in the germplasm collection using a PCR-based approach. Sahoo et al. (2009) utilized ISSR markers to assess the extent of genetic structure in a representative set of 226 individuals of *P. pinnata* encompassing 7 populations from different eco-geographical regions of Orissa. Sharma et al. (2010) assessed 20 random individuals of the *P. pinnata* using 2 dominant molecular markers, AFLP and 3 endonuclease (TE)-AFLP. In the present study, genetic variability for 10 CPTs of *P. pinnata* from North Guwahati, Assam, was assessed by DNA marker-based analysis. The use of seeds instead of leaves as a starting material for genetic variability studies is time saving and less expensive (McDonald et al. 1994). The plants do not have to be individually grown, and DNA can be extracted directly from seeds with the technical precautions of eliminating polysaccharides by supplementary DNA precipitation. The use of RAPD markers from seeds in a stable manner for varietal identification and genetic diversity evaluation has already been reported (McDonald et al. 1994; Fu et al. 2004). The present study constitutes the first successful attempt at assessment of genetic variability by DNA-based molecular characterization using seeds of *Pongamia* CPTs. The amplification profiles of the three marker systems resulted in differences in the detected banding patterns. The level of polymorphism for each of the three marker systems was quite variable. AFLP analysis revealed a high level of genetic diversity (100 %) in natural populations of *P. pinnata*, whereas only a small amount of variation (approx. 10 %) was detected based on both RAPD and ISSR markers in CPTs of *P. pinnata*. Kaushik et al. (2007) also reported that the distribution of genetic variability is limited in *P. pinnata* on the basis of their morphological characteristics such as pod and seed traits. However, higher level of polymorphism (94.3 %) was reported in seven populations of *P. pinnata* using few ISSR markers, but the samplings were done from three different eco-geographical regions of Orissa (Sahoo et al. 2009). Other woody perennial species such as *Populus tremuloides* and *Isotoma petraea* display lower levels of genetic variation detected with the same RAPD methodology (Yeh et al. 1995; Bussell 1999).

PIC analysis can be used to evaluate markers so that the most appropriate marker can be selected for genetic mapping and phylogenetic analysis (Anderson et al. 1993; Powell et al. 1996). Higher PIC value for the AFLP marker obtained would make its MI much higher than that of the other two dominant makers used in the present study. MI together with PIC value has been used to assess the informativeness of the markers earlier in another biofuel crop *Jatropha* (PIC = 0.26, MI = 25.15; Tatikonda et al. 2009).

From the heterogeneity data obtained it was observed that the total variability (h_tot) was highest when compared with average intra-population diversity (h_pop) based on all three molecular markers used in the current study. Similar trends have been observed in other crop species such as wild species of almond (Blair et al. 2006; Sorkheh et al. 2007). AFLP revealed high levels of genetic diversity at the population as well as the species level based on the value of the Shannon index. The estimate of mean G_st revealed that *P. pinnata* tends to be mainly an outcrossing mating system. According to Hamrick (1989), the tree populations distributed continuously having a longer life span with anemophily or cross-pollinating in nature possess higher genetic diversity, but lower genetic differentiation. Higher N_m value (6.287) based on AFLP reflects ample gene flow among and within populations that might be attributed primarily to the long-distance dispersal of pollen by strong winds (Huang et al. 2008). The results of the current study for *P. pinnata* are, in general, in accordance with what is expected for tropical trees (Li et al. 2008; Przyborowski and Sulima 2010).

The wide variation in genetic similarity among the 10 CPTs by the three DNA-based markers used in the study reflects a high level of polymorphism at the DNA level. RAPD and ISSR markers showed a high degree of similarity in dendrogram topologies, although with some differences, in the positioning of few individuals. Both markers aim to amplify a different region of the genome, and thus it is reasonable that there are some fine differences between the two dendrograms based on an individual data set. Relationships inferred from a neighbour-joining tree
were similar to the ones obtained from morphology, but somewhat better resolutions were achieved. The similar structure of the dendrogram also suggests that both techniques are suitable for genetic polymorphism research on *Pongamia*. According to Hamrick and Loveless (1989), the breeding system of a species is an important determinant of variability at both the species and population levels. The morphological, RAPD and ISSR data led to similar representations of the individual relationships. However, AFLP results revealed that the genetic variability was not according to the population distribution.

In the current study, four AFLP primer combinations were sufficient to disclose a proper number of polymorphic DNA fragments in support of all comparisons and of their statistical evaluation. Amplified fragment length polymorphism detected polymorphism more efficiently due to a greater number of loci detected/assay reactions. Based on the marker index value obtained, AFLP is found to be more efficient in detecting diversity among 10 CPTs. Genotype identification should be possible even with few AFLP primer combinations (Karudapuram and Larson 2005; Ozkan et al. 2005). However, to assess the genetic diversity and the genetic structure, a higher number of genotypes and localities are necessary. In this sense, Ozkan et al. (2005) and Karudapuram and Larson (2005) analysed 76 genotypes from 10 ecologically and geographically different locations. Among the three marker systems used, the highest similarity and cophenetic correlations were obtained for AFLP followed by RAPD and ISSR. The reason could be associated with the different nature of the three markers and different types of polymorphism detected (Scariot et al. 2007). Principal coordinate analysis is used to show multiple dimensions of the distribution of the genotypes in a scatter-plot (Keim et al. 1992). Separation of individuals to their respective clusters as evident from the UPGMA dendrogram as well as PCA was observed for all the three markers with the exception of a few rearrangements. This multivariate approach was used to complement the information obtained from the cluster analysis because it is more informative regarding distances among major groups (Taran et al. 2005).

The advantages of using three different marker systems are that they may have different applications in studies of *P. pinnata*, specific to their characteristics, but are unlikely to be similar because of convergent evolution and they can generate information at many different loci. In fact, the ISSR markers target regions between microsatellite loci distributed across the genome, while the RAPD and AFLP markers scan the entire genome and, hence, genome-wide genetic variation could be detected by the use of three DNA-based marker systems (Pamidimarri et al. 2009).

**Conclusion and forward look**

In summary, in this genomic era, the present study is the first attempt at molecular characterization in systematically identified locally grown CPTs of *P. pinnata* with multiple molecular markers. The observations demonstrate the usefulness, limitations and resolution power of several types of molecular markers in comparison to classical morphological descriptors for analysing the *P. pinnata* genome, for which no prior genome sequence information is available. The results presented herein reveal that 10 CPTs of *P. pinnata* belonging to two different populations exhibit large proportions of genetic variation as confirmed by fingerprinting profiles. Amplified fragment length polymorphisms outperform RAPDs and ISSRs in determining genetic variability among CPTs of *P. pinnata* due to their high reproducibility, robustness, informativeness and fewer reported reaction artefacts. The other two markers, RAPD and ISSR, showed less variability and showed congruent results in determining the genetic relationship among elite individuals of the study material. This study has provided a larger number of reliable and reproducible fingerprinting profiles for sustainable management, linkage mapping, developing single-locus sequence-characterized amplified region markers and genetic improvement of *P. pinnata* through conventional breeding methodologies as well as molecular breeding approaches such as marker-assisted selection.

**Additional information**

The following additional information is available in the online version of this article.

Figure S1. Principal co-ordinate map for the first, second and third principal coordinates estimated for RAPD, ISSR and AFLP markers for 10 CPTs of *P. pinnata* belonging to populations 6 and 10. (A) RAPD; (B) ISSR; (C) AFLP.

Table S1. Sequence information of RAPD and ISSR oligonucleotide primers used for amplification and polymorphism study among 10 CPTs of *P. pinnata* belonging to populations 6 and 10.

**Sources of funding**

V.K. thanks Council of Industrial Research (CSIR), Government of India for Senior Research Fellowship (SRF). L.R. acknowledges Department of Science and Technology-Science and Engineering Research Council (DST-SERC), Government of India, for funding the research.
Contributions by the authors
All the authors contributed to a similar extent overall and all authors have seen and agreed to the submitted manuscript.

Acknowledgements
Authors are thankful to the Forest Department officials of Sila Forest, North Guwahati, for kind supply of study material.

Conflict of interest statement
None declared.

References
Anderson JA, Churchill GA, Autrique JE, Tankelsy SD, Sorrells ME. 1993. Optimizing parental selection for genetic linkage maps. Genome 36: 181–186.

Archak S, Ambika B, Gaikwad D, Gautam EV, Rao VB, Swamy KRM, Karikaloo JL. 2003. DNA fingerprinting of India cashew (Anacardiaceae) varieties using RAPD and ISSR techniques. Euphytica 5: 397–404.

Blair MW, Giraldo MC, Buendia HF, Tovar E, Duque MC, Beebe SE. 2006. Microsatellite marker diversity in common bean (Phaseolus vulgaris L.). Theoretical and Applied Genetics 113: 100–109.

Brijesh S, Daswani PG, Tetali P. 2006. Studies on Pongamia pinnata (L.) Pierre leaves: understanding the mechanism(s) of action in infectious diarrhea. Journal of Zhejiang University-Science B 7: 665–674.

Bussell JD. 1999. The distribution of random amplified polymorphic DNA (RAPD) diversity among populations of Isotoma petraea (Lobeliaceae). Molecular Ecology 8: 775–789.

Duce LR. 1945. Measures of the amount of ecological association between species. Ecology 26: 297–302.

Fu YB, Ferrandiz YSN, Phan AT, Coulman B, Richards KW. 2004. AFLP variation in four blue gram seed sources. Crop Science 44: 283–288.

Hamrick JL. 1989. Isozymes and analyses of genetic structure of plant populations. In: Soltis D, Soltis P, eds. Isozymes in plant biology. Portland, OR: Discorides Press, 87–105.

Hamrick JL, Loveless MD. 1989. The genetic structure of tropical tree population: associations with reproductive biology. In: Bock JH, Linhart YB, eds. The evolutionary ecology of plants. Boulder, CO: Westview Press, 129–146.

Huang Y, Ji K, Jiang Z, Tang G. 2008. Genetic structure of Buxus sinica var. parvifolia, a rare and endangered plant. Scientia Horticulturae 116: 324–329.

Karadaparam S, Larson S. 2005. Identification of Hedysarum varieties using amplified fragment length polymorphism on a capillary electrophoresis system. International Journal of Biomolecular Analysis 16: 316–324.

Kaushik N, Kumar S, Kumar K, Beniwal RS, Kaushik N, Roy S. 2007. Genetic variability and association studies in pod and seed traits of Pongamia pinnata (L.) Pierre in Haryana, India. Genetic Resources and Crop Evolution 54: 1827–1832.

Keim P, Beavis W, Schupp J, Freestone R. 1992. Evaluation of soybean RFLP marker diversity in adapted germplasm. Theoretical and Applied Genetics 85: 205–212.

Kesari V, Krishnamachari A, Rangan L. 2008. Systematic characterization and seed oil analysis in candidate plus trees of biodiesel plant, Pongamia pinnata (L.). Annals of Applied Biology 152: 357–404.

Kesari V, Sudarshan M, Das A, Rangan L. 2009. PCR amplification of the genomic DNA from the seeds of Ceylon Ironwood, Jatropha and Pongamia. Biomass and Bioenergy 33: 1724–1728.

Kesari V, Das A, Rangan L. 2010. Physico-chemical characterization and microbial assay from seed oil of Pongamia pinnata, potential biofuel crop. Biomass and Bioenergy 34: 108–115.

Li F, Gan S, Weng Q, Zhao X, Huang S, Li M, Chen S, Wang Q, Shi F. 2008. RAPD and morphological diversity among four populations of the tropical tree species Paramichelia bailloni (Pierre) Hu in China. Forest Ecology and Management 255: 1793–1801.

Lombard V, Dubreuil P, Dillman C, Baril C. 2001. Genetic distance estimators based on molecular data for plant registration and protection: a review. Acta Horticulturae 546: 55–63.

Lopes PM, Brito JM, Gomes S, Meirinhos J, Santos L, Pinto HG. 2007. RAPD and ISSR molecular markers in Olea europaea L.: Genetic variability and molecular cultivar identification. Genetic Resources and Crop Evolution 54: 117–128.

Mantel N. 1967. The detection of disease clustering and generalized regression approaches. Cancer Research 27: 209–220.

McDermott JM, McDonald BA. 1993. Gene flow in plant pathosystems. Annual Review of Phytopathology 31: 353–373.

McDonald MB, Elliot LJ, Sweeney PM. 1994. DNA extraction from dry seeds for RAPD analysis in varietal identification studies. Seed Science and Technology 22: 171–176.

Meera B, Kumar S, Kalidhar SB. 2003. A review of the chemistry and biological activity of Pongamia pinnata. Journal of Medicine and Aromatic Plant Science 25: 441–465.

Meudt HM, Clarke AC. 2006. Almost forgotten or latest practice? AFLP applications, analyses and advances. Trends in Plant Science 12: 106–117.

Mukta N, Murthy IYLN, Sripal P. 2009. Variability assessment in Pongamia pinnata (L.) Pierre germplasm for biodiesel traits. Industrial Crops and Products 29: 536–540.

Nei M. 1973. Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences of the USA 70: 3321–3323.

NOVOD. 2010. 5th R&D report on tree borne oilseeds. In: NOVOD Board India. http://www.novodboard.com/5th%20R&D-Report.pdf (2 June 2010).

Ozkan H, Kafkas S, Ozer MS, Brandolini A. 2005. Genetic relationships among South-East Turkey wild barley populations and sampling strategies of Hordeum spontaneum. Theoretical and Applied Genetics 112: 12–20.

Pamidimarri DVNS, Singh S, Mantan SG, Patel J, Reddy MP. 2009. Molecular characterization and identification of markers for toxic and non-toxic varieties of Jatropha curcas L. using RAPD, AFLP and SSR markers. Molecular Biology Reports 36: 1357–1364.

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A. 1996. The unity of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Molecular Breeding 2: 225–238.
The analysis of genetic diversity of Salix viminalis genotypes as a potential source of biomass by RAPD markers. Industrial Crops and Products 31: 395–400.

Reddy MP, Sarla N, Siddiq EA. 2002. Inter simple sequence repeat (ISSR) polymorphism and its application in plant breeding. Euphytica 128: 9–17.

Rohlf FJ. 2002. NTSYS-pc: numerical taxonomy and multivariate analysis system. Version 2.231. New York: Exeter Software.

Sahoo DP, Aparajita S, Rout GR. 2009. Inter and intra-population variability of Pongamia pinnata: a bioenergy legume tree. Plant Systematics and Evolution 285: 121–125.

Sarma AK, Konwer D, Bordoloi PK. 2005. A comprehensive analysis of fuel properties of biodiesel from koroch seed oil. Energy and Fuels 19: 656–657.

Scariot V, De Keyser E, Handa T, De Riek J. 2007. Comparative study of the discriminating capacity and effectiveness of AFLP, STMS and EST markers in assessing genetic relationships among evergreen azaleas. Plant Breeding 126: 207–212.

Scott PT, Pregelj L, Chen N, Hadler JS, Djordjevic MA, Gresshoff PM. 2008. Pongamia pinnata: an untapped resource for the biofuels: industry of the future. Bioenergy Research 1: 2–11.

Sharma SS, Negi MS, Sinha P, Kumar K, Tripathi SB. 2010. Assessment of genetic diversity of biodiesel species Pongamia pinnata accessions using AFLP and three endonuclease-AFLP. Plant Molecular Biology Reporter doi:10.1007/s11105-010-0204-2.

Simmons MP, Zhang LB, Webb CT, Müller K. 2007. A penalty of using anonymous dominant markers (AFLPs, ISSRs, and RAPDs) for phylogenetic inference. Molecular Phylogenetic Evolution 42: 528–542.

Slatkin M, Barton NH. 1989. A comparison of three indirect methods for estimating average levels of gene flow. Evolution 43: 1349–1368.

Smith JSC, Smith OS. 1989. The description and assessment of distances between inbred lines of maize. II. The utility of morphological, biochemical and genetic descriptors and a scheme for testing of distinctiveness between inbred lines. Maydica 34: 151–161.

Sneath PHA, Sokal RR. 1973. Numerical taxonomy: the principles and practice of numerical classification. San Francisco, CA: Freeman.