

Molecular Characterization of *Pongamiapinnata*(L.) Pierre genotypes in respect to seed oil using microsatellite markers

India

Abstract

Twenty-four genotypes selected from different agroclimatic zones of Jharkhand (India) were assessed for seed oil content that varied from 21.86% to 41.86% with 14 genotypes recording above average 32.11 % for the trait which indicates towards efficiency of selection processes. Several genotypes in 'Central and Western Plateau' agroclimatic zone of Jharkhand displayed a good potential for high oil content. The employed 23 polymorphic microsatellite loci exhibited three to twenty one alleles per locus with an average of 12 while total 270 alleles were detected. Two primer set PpSSR21 and PpSSR27 showed 100% polymorphism among the genotypes. The high oil yielding plant K₁₉ showed different band pattern with the locus PpSSR04. From the Nei's analysis it was found that maximum diversity exists between the full sib genotypes K₁₀ and K₂. Thus, the genotypes (K₂ and K₁₀) which are more diverse could be used further in improvement programme. Overall, the genotypes included in the study showed a correlation with their geographical origins such that genotypes from the same region tend to have higher genetic similarity as compared to those from different regions. However, in UPGMA based Nei's analysis, some genotypes were found not to be grouped based on geographical origins possibly due to the exchange of germplasm over time between farmers across the regions. Both molecular and oil content (biochemical) markers appeared useful in analyzing the extent of genetic diversity in *P. pinnata*. The result of these analyses will help to better understand the genetic diversity and relationship among populations constituting a set of useful background information that can be used as a basis for future breeding strategy and improvement of the species.

Key words: amplicons, biofuel, genetic diversity, SSR markers, polymorphism

1. Introduction

With declining reserves of fossil fuels, biofuels can make a substantial contribution to the future energy demands of the domestic and commercial economies. Among many TBOs (Tree Bourne Oilseed species) of India, which can yield oil as a source of energy in the form of biodiesel, Karanz[*Pongamiapinnata*(L) Pierre] has been found to be the most suitable species due to its various favorable attributes like high oil recovery, quality of oil containing high amount of polyunsaturated fatty acids which are an important source of biodiesel and subsequent

Comment [M1]: State the percentage of oil for the crop

processing of biodiesel (Azam et al. 2005; Karmee and Chadha 2005; Sarma et al. 2005; Sharmin et al. 2007). It is a leguminous deciduous tree, commonly known as Indian Beech, Pongam, Honge and Karanj, grows about 15-25 m tall and is well-adapted to semi-arid and humid zones.

P.pinnata is indigenous to the Indian subcontinent and has been successfully introduced to humid tropical regions of the world as well as parts of Australia, New Zealand, China and the United States (Scott et al. 2008). The species easily growing on unproductive land and adaptable to wide agro-climatic conditions, can prove itself to be a promising alternative to the energy requirement in the future especially in a country like India which produces only 22% of its diesel requirement and 78% is imported at a huge cost in foreign currency. However, central to the domestication and sustained use of *P. pinnata*, is an understanding of its molecular genetics including aspects of gene content and structural organization in relation to other plant genomes (Biswas et al. 2011).

In meeting future demands for bio-diesel establishment of extensive commercial-scale *P. pinnata* plantations will be beneficial necessitating identification and selection of *P. pinnata* plants that have genetically linked with optimal production of biofuels, such as high seed oil content. The effectiveness of tree improvement targeting this will depend upon the nature and magnitude of existing genetic variability in the species for which molecular approach is more efficient as it directly quantifies genetic variability at DNA level. Molecular markers have provided a powerful tool for breeders to search for new sources of variation and to investigate genetic factors controlling quantitatively inherited characters. Various PCR based molecular markers have been employed in *Pongamia* for identification, phylogenetic analysis, population studies and genetic linkage mapping (Acharya et al. 2004; Kesari et al. 2010; Thudi et al. 2010; Sharma et al. 2010; Sujathta et al. 2010; Ahlawat et al. 2016; Dandapani et al. 2022; Rajarajan et al. 2022). Simple sequence repeats (SSR) or microsatellite markers have emerged as the markers of choice due to several desirable attributes including their abundance, multiallelic and co-dominant nature, high level of reproducibility, cross-species transferability etc. for such studies.

Keeping these facts in mind present investigation involving microsatellite markers was conducted to study the existing genetic diversity and variation in different genotypes of *P. pinnata* collected from representative parts of Jharkhand and to document their potentiality for exploitation of oil yielding property as a scope for their incorporation in further breeding program.

2. Materials and methods

2.1 Plant material

Germplasm of twenty four genotypes of *P. pinnata* collected from different region of Jharkhand and maintained at Institute of Forest Productivity, Ranchi, Jharkhand (India) as clonal (grafted plants) were utilized (Table 1). Three individual of each of 24 genotype (clonal) were selected for the study.

2.2 Biochemical Analysis

Seed of different genotypes of *P. pinnata* was analyzed for total oil content as per Sadasivam and Manickam (1992). The data obtained were subjected to statistical analysis, employing analysis of variance (ANOVA), 'F'-test for significance at $P \leq 0.05$ and computing LSD values to separate means in different statistical groups using statistical software IBM SPSS version 18.

2.3 DNA extraction and SSR analysis

Total DNA was extracted from fresh leaves by the cetyl tri-methyl ammonium bromide (CTAB) method with slight modification (Murray and Thompson 1980; Doyle and Doyle 1990). The quality and concentration of extracted DNA were estimated by using a UV spectrophotometer (Biophotometer plus, Genetics) and running an aliquot on 0.8% agarose (Sigma-A9539) gels. After visualization the DNA was diluted if required.

2.4 Primer designing

Specific primers were designed for the flanking regions of the SSRs using the software WEBSAT (wsmartins.net/websat/). The sequences would be downloaded from public databases (nucleotide sequences at the NCBI databases (<http://www.ncbi.nlm.nih.gov>)). The designed primers were of 19–25 bp length, annealing temperature was 50–60°C, and product sizes were 100 to 400 bp (Table 2). The criteria adopted to reduce the amplification of nonspecific bands in PCR reactions included a minimum primer annealing temperature (T_m) of 57°C, a maximum difference of 1°C in T_m between the two primers of an SSR locus and a G+C content ranging from 40 to 80°C. The primers were synthesized by SIGMA-ALDRICH (USA). Total twenty nine primer pairs (mono, di and trinucleotide base) designed were tested for DNA amplification on twenty four *P. pinnata* genotypes.

2.5 Standardization of annealing temperature

The best annealing temperatures for SSR primers pair were standardized. For each primer a gradient PCR was performed. Six primer pair did not show any amplification at different annealing temperature. So, only twenty three microsatellite Primer pair was taken for this study (Table 3).

2.6 PCR conditions for SSR primers

Each PCR reaction was carried out in a 20 µl reaction volume containing 10 pmoles each of forward and reverse primer, 1 Unit Red Taq DNA polymerase (Banglore GeNei), 10X Taq buffer A, 200 µM of deoxyribonucleotides and 50 ng of template genomic DNA was added. The temperature profile used for PCR amplification comprised denaturation at 94°C for 4 min, followed by 35 cycles of 92°C for 30 sec, the annealing temperature for 30 sec, 72°C for 1 min and a final extension was given at 72°C for 5min. The annealing temperature was adjusted basing on the specific requirements of each primer combination. The amplicons obtained after PCR was run on 3.5% agarose gel electrophoresis (wide range agarose, Sigma A 7431). Appropriate ladders were also run with the samples. The samples were electrophoresed at constant voltage of 180 Volts (5V/cm) for 3 hours. The gel was then visualized under UV light on a transilluminator and documentation of the samples was done on Gel Documentation System (Syngene).

The amplified products were arranged according to their molecular weight and according to that a particular genotype was assigned to each variety. Each value represents a particular allele. The genotype of all the individuals were fed in the software and on its basis genetic similarity and variation was analyzed. The genotype was assigned a null allele for a microsatellite locus whenever an amplification product could not be detected for a particular genotype-marker combination. The calculations were performed by using Popgene software (Yeh et al. 1999). Percentage of all loci that was polymorphic regardless of allele frequencies was determined using the software.

Nei's analysis of genetic distance and identity was used to study the genetic distance between the populations. A dendrogram were created with UPGMA (unweighted pair group method with arithmetic averaging) based on Nei's genetic distance. Ewens-Watterson test are used to test the neutrality of the locus. The observed F value is compared with the lower and upper 95% confidence interval. If observed F value is within this confidence interval, the locus is neutral otherwise it is not. The distribution of private (unique) alleles was assessed as a mean to describe population distinctiveness (Slatkin 1985) and levels of genetic interogation between populations (Gottelli et al. 1994). Private alleles were assessed manually from the converted data file for allelic frequency with a software Convert1.31 (Glaubitz 2004).

...

3. Results and Discussion

3.1 Biochemical characteristics

Comment [M2]: Adding a table showing some of the physical and chemical properties of the oil

Comment [M3]: Some results require explanation

Seed from different regions contain a lot of variations in morphological and physiological characteristics which could be genetic in nature as a result of adaptation to diverse environmental condition prevailing throughout their distributional range (Mathur et al. 1984). In the present study also significant variation among different genotypes of *P. pinnata* in total oil content was recorded which provides an opportunity for genetic improvement and breeding in the species.

The highest seed oil content (41.87%) was found in the K₁₉ followed by K₂₇ (39.33%) and K₂₂ (38.80%). Average total oil content was 32.11% while seeds of 14 genotypes recorded above average for the trait (Fig.1). The identified promising candidates such as K₁₉, K₂₇, K₂₂, K₀₉ and K₁₅ were can be used for clonal development and utilization in large scale plantation drives of *P. pinnata*. Based on the average oil content at agroclimatic zone, the order of *Pongamia* genotypes from high to low was fifth agroclimatic zone (Central and Western plateau, JH-2) in India < fourth agroclimatic zone in India (Central and North Eastern Plateau , JH-1) < sixth agroclimatic zone in India (South Eastern Plateau, JH-3). Some mixing however was also observed.

Central and Western plateau (JH-2) accounts for 33% natural forests of Jharkhand and characterized by late arrival, early cessation and uneven distribution of monsoon rains; and poor water retention capacity of soil. These conditions might have an effect on enhanced seed oil in the genotypes of the region. However, from overall evaluation of seed oil content, it appears that individual genotypic factors supersede over environmental factors in determining oil yield.

3.2 Molecular characterization through SSR markers

The genetic marker can be used to select wild candidate plus trees collected from different agro-climatic regions of tree improvement programs. Various molecular approaches have been employed for genetic characterization of *P. pinnata* such as RAPD (Acharya et al. 2004; Kesari et al. 2010; Kesari and Rangan, 2011; Kumari et al. 2013), AFLP (Pavithra et al. 2014), ISSR (Jiang et al. 2012; Sharma et al. 2014), TE-AFLP (Kesari et al. 2010; Thudi et al. 2010; Sharma et al. 2010; Pavithra et al. 2014; Sharma et al. 2014) and SSR (Sharma et al. 2014). However there are no reports available on the variability in total seed oil content of *P. pinnata* with the help of SSR markers.

For the use of SSRs as markers, it is necessary to design the primers. The SSRs commonly used for marker development are those belonging to di-, tri- and tetra-nucleotides. The mononucleotides are useful for population genetic analyses of chloroplast genomes and can also be useful in filling gaps in linkage maps created by di-, tri-, and tetra-nucleotide repeats but, at the same time they cause difficulties in accurate sizing of polymorphisms.

In the present study, species specific SSR primers were designed with the help of WEBSAT software. Development of SSRs by data mining from sequence data is a relatively easy and cost-saving strategy for any organisms with enough DNA data. This is a very good example from data to knowledge and from knowledge to basic and applied studies for biology, production, conservation and management of many organisms. The primer sets were selected on the basis of the product length of the amplicons which ranged from 110-400bp. By employing high resolution agarose gel electrophoresis we were successful to clearly separate alleles for the twenty three microsatellite markers studied.

3.3 Amplification profiles and allelic variants

The twenty three polymorphic microsatellite loci identified in this study showed a clear and single or multiple peak for each allele, while remaining six markers showed no clear band in accessions. The amplified products were in agreement with the expected size from sequence data for all 23 microsatellites analyzed except one SSR marker. The distribution of alleles for a single and multiple SSR loci across all genotypes is shown in Fig2-3. A total of 270 alleles were detected by the twenty three markers. The number of alleles per locus varied from 3 to 21. The average number of alleles per locus was 12, indicating a greater magnitude of diversity among the plant materials include in this investigation (Table 4).

By combined analysis of 11 polymorphic loci (more than 70%) it was possible to distinguish almost all genotypes. However, the complete identification of 24 genotypes implied the study of few additional loci. For the twenty three loci studied we detected 270 different alleles, PpSSR04 and PpSSR09 being the most informative loci for genetic identification because number of alleles were present in single genotypes. So specific allele size was used for particular types of genotypes.

The study also revealed that primers PpSSR12 and PpSSR13 generated a maximum of 21 and 20 bands, while primers PpSSR19, PpSSR27, PpSSR02 and PpSSR15 had a minimum of 3, 5 and 6 bands. Many studies have also reported significant differences in allelic diversity among various microsatellite loci (Akagi et al. 1997; Ravi et al. 2003).

The markers employed in the study showed a high level of polymorphism, ranging from 50-100% (as many as 20 markers). AFLP detected higher levels of genetic diversity (100%) in natural populations of *P. pinnata*, whereas RAPD and ISSR showed lesser genetic diversity (approx. 10%) (Kesari et al. 2010). The frequency of genotypes revealing null alleles only contained in single primer set PpSSR19. Two primer set PpSSR21 and PpSSR27 showed 100% polymorphism among the genotypes.

The high oil yielding plant K₁₉ showed different band pattern with the locus PpSSR04.

The amplicon size expected with this primer set was 308bp. The range of the amplicons obtained were 293-322bp. The unique product length 293bp was obtained which is distinct in the genotypes (Fig. 4). So this locus may be used as a marker for high oil yield genotype detection. However, for its documentation, it is necessary to sequence this locus in whole genotype and then differentiate them on the basis of base sequence and further development of SNP.

In addition, there were a number of loci PpSSR08, PpSSR09, PpSSR12, PpSSR13, PpSSR14 and PpSSR16 which generated private allele in K₁₉ genotype (Table 5). These private allele generating loci could be further multiplexed to detect the presence of the mentioned genotype in a single reaction.

The analysis of repeat units under each motif class revealed a varying range of repeat units in each of the classes of repeat motifs. It was observed that, in tri-nucleotide motif, repeat units ranged from 2-24; in dinucleotide motif, from 13-22 and in mononucleotide units ranged from 29-30. Further analysis of the number of repeat units in every class of the SSRs, especially mono, di and tri nucleotides, showed that the number of the microsatellites decreased with increasing repeat unit. Similar result was observed by Varshney et al. (2007). These results clearly indicate the effect of increased stringency of parameters which were maintained during this study to retain the polymorphism level and utility of the SSRs as markers because the probability of polymorphism increases with increasing length of SSRs and, a higher number of repeat followed by shorter stretches would be beneficial for marker development.

Null alleles are those that fail to amplify in a PCR, either because the PCR conditions are not ideal or the primer-binding region contains mutations that inhibit binding. In this study, primers PpSSR19 could not amplify reported products in all genotypes. In the present study, detected null alleles (non-amplifying alleles) that may have been due to mutation at the priming sequence. Similar result was also observed by Callen et al. (1993) in the parents of the 40 CEPH (Centre d'Etude du Polymorphisme Humain) families. In contrast, Moriguchi et al. (2003) detected null alleles at 12 out of 34 loci developed (35.3%) in a segregation analysis, and deduced that the high rate of null allele detection was caused by a high mutation rate at the priming sequences in *C. japonica*. Low rate of null allele detection in the present study suggests that the high mutation rate at priming sequences is not pandemic in this species. However, requires further intensive investigation.

The present study showed newly developed microsatellite markers revealed a lower rate of null allele detection than the microsatellite markers developed by Moriguchi et al. (2003). Therefore, these markers are likely to be valuable tools, not only for genetic mapping, but also for analyses of population genetics and reproduction ecology in natural populations (Derory et

al., 2002; Gonzalez-Martinez et al. 2003).

3.4 Analysis of genetic diversity

Genetic diversity indicated by Shannon's index varies- I (mean value in each population 0.00 - 0.3781) and Nei's mean estimates of diversity- H (from 0.00 to 0.4163) in the present study can be considered high since the values were equal to or higher than 0.40. The values found are similar to those obtained for other tree species *Populus tremuloides* (0.58 to 0.69, Yeh et al. 1995), *Fitzroya cupressoides* (0.42 to 0.56; Allnut et al. 1999) and *Swietenia macrophylla* (0.41 to 0.27, Gillies et al. 1997). The average genetic diversity (H) obtained is somewhat higher than those found for other species e.g. *Trichilia pallida* Swartz, ranged from 0.27 to 0.33 (Zimback et al. 2004) and *Aspidosperma polyneuron*, diversity averaged 0.28 (Torezan et al. 2005).

The estimates of diversity in *Pongamia pinnata* through SSR studies was found to be akin/similar to other important oil yielding species Sesame. In the case of SSR data on Sesame plant, the expected heterozygosity (H_e) and observed heterozygosity (H_o) ranged from 0.12 to 0.48 (mean, 0.29) and 0.00 to 0.30 (mean, 0.15), respectively (Pham et al. 2010).

The UPGMA analysis revealed some important points in affinity and relationship of the 24 identified superior genotypes from eastern India. According to the UPGMA, the genotypes were divided into two subgroups: - one sub group comprised of ten genotypes i.e. K_3 , K_{20} , K_{22} , K_{23} , K_{27} , K_{12} , K_{21} , K_{19} , K_{25} , K_{26} and the other sub group contained rest of the genotypes (Fig. 5). The highest and lowest oil yielding genotypes separated from each other but within the same sub group. The constructed dendrogram demonstrate that there is a considerable differentiation among the genotypes. Schaal et al. (1998) suggested that historical relationships have been found to contribute populations to have similar genetic structure like that of sharing common ancestry. Therefore, there may be similar gene pools initially established in the demes during their establishment and then similar pattern of selection pressure operates on them.

Clustering of individuals from same agroclimatic in different clusters indicate high genetic variation within population which may be attributed to the use of different seed sources, mutation and/or breeding system. This is expected for out-crossing species and it also agrees with other similar studies which had similar observations (Rosseto et al.1995). Since *P. pinnata* is a predominantly out-crossed species, it was expected to maintain most variations within populations than between populations. Several studies have reported similar observations in other plant species (Padmesh et al. 1998; Virginie et al. 2002) although some other studies were able to find separation of genotypes based on geographic origin (Keskitalo et al. 1998; Desplanque et al. 1999; Zenglu and Randall 2002). Individuals from different geographical locations are expected to be genetically different due to varying selection pressure that allow fixation of a particular

character. The total genetic variation of the species is likely to be distributed among populations as the impact and direction of natural selection varies from one population to another due to environmental variation and genetic drift (Lawrence and Rajanaidu 1985).

Based on the results it can therefore be concluded that, selection of materials for improvement of individuals and conservation of *P. pinnata* genetic resource can be achieved by considering individuals from wide geographical area in eastern India which is the native range of the species.

From the Nei's analysis it was found that maximum diversity exists between the full sib genotypes K₁₀ and K₂ (Table 6). K₁₆ and K₁₄ are the two genotypes which are genetically identical. In plant breeding programmes, the most diverse genotypes are used in the hybridization programme. Thus, the genotypes (K₂ and K₁₀) which are more diverse could be used further in hybridization programme.

Analysis of molecular variance (AMOVA) was performed to assess the variation among and within the genotypes, using the genetic distance matrix generated by SSR markers. The partitioning of the genetic variability among the twenty four genotypes is shown in Table 7. In full sib genotypes, the analysis of molecular variance showed a significant amount of differentiation (77.92%) among the population. But with respect to overall population showed 44.16% of total variability.

Private alleles were also found in these *P. pinnata* genotypes. Private alleles (which are only present in one genotype and absent from all others tested) were suspected to have accelerated the genetic differentiation in these populations. K₁₁ had relatively more private alleles with high frequency. At locus PpSSR20, twenty four private alleles were founded. The locus PpSSR15 did not show any private alleles. In the genotype K19, number of locus showed private alleles. These private alleles were further used for this high oil yielded genotype characterization and unique marker development (Table 5). On the basis of allele size, the high oil yielding and low oil yielding genotypes show significant variation (Table 8).

...

Comment [M4]: If it is possible to add economic feasibility

4. Conclusion

In conclusion, we isolated the first set of twenty three polymorphic microsatellite loci and characterized across twenty four *P. pinnata* accessions. Despite the modest number of genotypes analyzed, the results indicate that the new microsatellite loci developed in this study will serve as a very useful tool for genetic diversity analysis, clonal identification and conservation in *P. pinnata* germplasm. High level of genetic variation was found among different genotypes of *P.*

pinnata. Both molecular and oil content (biochemical) markers appeared useful in analyzing the extent of genetic diversity in *Pongamia* and the result of these analyses will help to better understand the genetic diversity and relationship among populations.

References

Acharya L, Mukherjee AK, Panda PC. 2004. Genome relationship among nine species of Millettieae (Leguminosae: Papilionoideae) based on Random Amplified Polymorphic DNA (RAPD). *J Biosci.* 59: 868–73.

Ahlawat SP, Kumar RV, Pandey SK, Joshi DC, Dhyani SK. 2016. Morphological and molecular level of genetic diversity among *Pongamia* [*Pongamia pinnata* (L.) Pierre] accessions. *Ind J Biotech.* 15: 85-94.

Akagi H, Inagaki YA, Fujimura T. 1997. Highly polymorphic microsatellite of rice consist of AT Repeats, and a classification of closely related cultivars with these microsatellite loci. *Theor Appl Genet.* 94: 61–67.

Allnutt TR, Newton AC, Lara A, Armesto JJ, Premoli A, Veraga R, Gardner M. 1999. Genetic variation in *Fitzroya cupressoides* (alerce), a threatened South American conifer. *Mol Ecol.* 8: 975–987.

Azam MM, Waris A, Nahar NM. 2005. Prospects and potential of fatty acid methyl esters of some nontraditional seed oils for use as biodiesel in India. *Biomass Bioenergy* 29: 293–302.

Biswas B, Scott PT, Gresshoff PM. 2011. Tree legumes as feedstock for sustainable biofuel production: opportunities and challenges. *J Plant Physiol.* 168: 1877-1884.

Callen DF, Thompson AD, Shen Y, Phillips HA, Richards RI, Mulley JC, Sutherland GR. 1993. Incidence and origin of -null alleles in the (AC)_n microsatellite markers. *Am J Hum Genet.* 52:922–927.

Dhandapani T, Mohamed O, Mohamed N, Rajan SP, Suryanarayanan L. 2022. Characterization of assembled clones of *Pongamia pinnata* (L.) Pierre for its genetic variability using ISSR. *Int J Sci Res Arch.* 7: 173-180.

Comment [M5]: Formatting references according to publishing conditions

Comment [M6]: Write the scientific name in italics

- Derory J, Mariette S, Gonzalez-Martinez SC, Chagn D, Madura D, Gerbera S, Brach J, Persyn F, Ribeiro MM, Plomiona C. 2002. What can nuclear microsatellites tell us about maritime pine genetic resources conservation and provenance certification strategies? *Ann For Sci.* 59:699–708.
- Desplanque B, Boudry P, Broomberg K, Saumitou-Lapadre P, Cuguen J, van Dijk H. 1999. Genetic diversity and gene flow between wild, cultivated and weedy forms of *Beta vulgaris* (Chenopodiaceae) assessed by RFLP and microsatellitemarkers. *Theor Appl Gen.* 98: 1194 – 1201.
- Doyle JJ, Doyle JL. 1990. Isolation of plant DNA from fresh tissue. *Focus* 12:13–15.
- Gillies ACM, Cornelius JP, Newton AC, Navarro C, Hernandez M, Wilson J. 1997. Genetic variation in Costa Rican populations of the tropical timber species *Cedrela odorata* L. assessed using RAPDs. *Mol Ecol.* 6:1133–1146.
- Glaubitz JC. 2004. Convert: A User-Friendly Program To Reformat Diploid Genotypic Data for Commonly Used Population Genetic Software Packages. *Mol Ecol Notes.* 4: 309–310.
- Gonzalez-Martinez SC, Gerber S, Cervera MT, Martinez-Zapater JM, Alia R, Gil L. 2003. Selfing and sibship structure in a two-cohort stand of Maritime Pine (*Pinus pinaster* Ait.) using nuclear SSR markers. *Ann For Sci.* 60:115–121.
- Gottelli D, Sillero-Zubiri C, Applebaum GD, Roy MS, Girman DJ, Garcia-Moreno J, Ostrander EA, Wayne RK. 1994. Molecular genetics of the most endangered canid: The Ethiopian wolf *Canis simensis*. *Mol Ecol.* 3:301-312.
- Jiang Q, Yen SH, Stiller J, Edwards D, Scott PT, Gresshoff PM. 2012. Genetic, biochemical, and morphological diversity of the legume biofuel tree *Pongamia pinnata*. *J Plant Genome Sci.* 1:54–67.
- Karmee SK, Chadha A. 2005. Preparation of biodiesel from crude oil of *Pongamia pinnata*. *Biores Tech.* 96: 1425-1429.
- 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.

- Kesari V, Rangan L. 2011. Genetic diversity analysis by RAPD markers in candidate plus trees of *Pongamiapinnata*, a promising source of bioenergy. *Biomass Bioenergy*. 35:3123-3128.
- Keskitalo M., Linden A., Valkonen J. 1998. Genetic and morphological diversity of Finnish tancy (*Tanacetum vulgare*, Asteraceae). *Theor Appl Gen*. 96: 1141-1150.
- Kumari K, Sinha A, Singh S, Divakara BN. 2013. Biochemical Characterization and Genetic Diversity of *Pongamia pinnata* (l.) Pierre in Eastern India. *J For Sci*. 29: 200-210.
- Lawrence MJ, Rajanaidu N. 1985. The genetic structure of natural populations and sampling strategies. *Proc Int Workshop on oil palm germplasm and utilization, Selangor, Malaysia*.
- Mathur RS, Sharma KK, Rawat MMS. 1984. Germination behaviour of provenance of *Acacia nilotica* sp. *Indica*. *IndFor*. 110:435-449.
- Moriguchi Y, Iwata H, Ujino-Ihara T, Yoshimura K, Taira H, Tsumura Y. 2003. Development and characterization of microsatellite markers for *Cryptomeria japonica* D. Don, *Theor Appl Genet*. 106: 751-758.
- Murray MG, Thompson WF. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res*. 8: 4321-4325.
- Nei M. 1973. Analysis of gene diversity in subdivided populations. *Proc Natl AcadSci USA*. 70: 3321-3323.
- Padmesh P, Sabu KK, Seeni S, Pushpangadan P. 1998. The use of RAPD in detecting genetic variability in *Andrographispaniculata* Nees: a potent hepatoprotective drug. *Cur Scien*. 76: 833-835.
- Pavithra HR, Shivanna MB, Chandrika K, Prasanna KT, Gowda B. 2014. Genetic analysis of *Pongamiapinnata* (L.) Pierre populations using AFLP markers. *Tree Genet Genomes*. 10:173-188.
- Pham DT, Nguyen TTD, Carlsson AS, Bui MT. 2010. Morphological evaluation of Sesame (*Sesamum indicum* L.) varieties from different origins. *Aust J Crop Sci*. 4: 498-504.
- Rajarajan K, Sakshi S, Rana M, Radhakrishna A, Vishnu R, Anuragi H, Handa AK, Arunachalam A. 2022. Understanding the genetic determinants and population

structure of *Pongamia pinnata* (L.) Pierre for oil yield and its properties using transcriptome derived SSR markers. *Plant Physiol Rep.* 3: 407-418.

Ravi M, Geethanjali S, Sameeyafarheen F, Maheswaran M. 2003. Molecular marker based genetic diversity analysis in Rice (*Oryza sativa* L.) using RAPD and SSR markers. *Euphytica.* 133: 243–252.

Rosseto M, Weaver PK, Dixon KW. 1995. Use of RAPD analysis in devising conservation strategies for the rare and endangered *Grevillea scopigera* (Proteaceae). *Mol Ecol.* 4: 321-329.

Sadasivam S, Manickam A. 1992. Biochemical methods for agricultural sciences. Madras: Wiley Eastern Ltd.

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

Schaal BA, Hayworth DA, Olsen KM, Rauscher JT, Smith WA. 1998. phylogeographic studies in plants, problems and prospects. *Molr Ecol.* 7:465– 474.

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

Sharma SS, Keshaw A, Negi MS, Tripathi SB. 2014. Efficacy of two dominant marker systems, ISSR and TE-AFLP for assessment of genetic diversity in biodiesel species *Pongamia pinnata*. *Curr Sci.* 106:1576-1580.

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 Mol Biol Rep.* 29:12-18.

Sharmin E, Ashraf SM, Ahmad S. 2007. Synthesis, characterization, antibacterial and corrosion protective properties of Epoxies, Epoxy-Polyols and Epoxy-Polyurethane coatings from Linseed and *Pongamia glabrata* seed oils. *Int J Biol Macromol.* 40: 407-422.

Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics.* 139: 457-462.

Sujatha K, Rajwade AV, Gupta VS, Hazra S. 2010. Assessment of *Pongamia pinnata* (L.) - a biodiesel producing tree species using ISSR markers. *Curr Sci.* 99: 1327- 1329.

- Thudi M, Manthena R, Wani SP, Tatikonda L, Hoisington DA, Varshney RK. 2010. Analysis of genetic diversity in *Pongamia* [*Pongamiapinnata* (L) Pierre] using AFLP markers. *J Plant Biochem Biotechnol.* 19: 209-216.
- Torezan JMD, Souza RF, Ruas PM, Ruas CF, Camargo EH, Vanzela ALL. 2005. Genetic variability of pre and postfragmentation cohorts of *Aspidospermapolyneuron* Muell. Arg. (Apocynaceae). *Braz Arch Biol Technol.* 48: 171-180.
- Varshney RK, Mahendar T, Aggarwal RK, Borner A. 2007. Genetic Molecular Markers in Plants: Development and Applications. In: Varshney RK, Tuberosa R, editors. *Genomics-Assisted Crop Improvement: 1: Genomics Approaches and Platforms*, Springer, Berlin; p. 13-29.
- Virginie BC, Jean G, Claude G, Jacques M. 2002. Genetic relationships between cultivated and wild Corsica and Sardinia using RAPD markers. *Euphytica.* 123: 263-271.
- Yeh F, Boyle CT, Ronaghai Y, Xian JM. 1999. Popgene version 3.1 (<http://www.vulberta.ca/~fyeh/fyah>).
- Yeh FC, Chong DKX, Yang RC. 1995. RAPD variation within and among natural populations of trembling aspen (*Populustremuloides*) from Alberta. *J Hered.* 86: 454-460.
- Zenglu L, Randall LN. 2002. RAPD marker diversity among cultivated and wild Soybean accessions from four Chinese provinces. *Crop sci.* 42: 1737-1744.
- Zimback L, Mori ES, Kageyama PY, Veiga RFA, Mello Júnior JRS. 2004. Estrutura genética de populações de *Trichilia Pallid Swartz* (Meliaceae) por marcadores RAPD. *Sci For.* 65: 114-119.

Table 1. Details of *Pongamiapinnata* genotypes used for genetic diversity analysis.

Sl. No.	Code	Forest Division	Latitude	Longitude	Altitude (m)	Agroclimatic zone
1	K ₁₃	Hazaribag North	24°27'10"N	85°31'42"E	370	JH-1
2	K ₂₅	Giridih	24°17'11"N	86°21'55"E	390	Central and North eastern Plateau Zone
3	K ₁₂	Hazaribag West	23°54'19"N	85°19'04"E	610	
4	K ₁₄	Koderma	24°27'21"N	85°46'12"E	380	
5	K ₁₁	Chatra South	23°54'19"N	85°19'04"E	640	
6	K ₁₉	Ranchi	23°33'02"N	84°54'43"E	790	
7	K ₂₇	Ranchi East	23°17'07"N	85°10'35"E	810	
8	K ₀₉	Lohardaga	23°33'02"N	84°54'43"E	640	
9	K ₁₅	Ranchi East	23°27'45"N	85°28'39"E	630	
10	K ₂₃	Khunti	23°03'05"N	85°13'40"E	700	
11	K ₁₀	Garhwa South	23°55'30"N	83°46'11"E	410	
12	K ₀₇	Simdega	22°35'47"N	84°40'49"E	370	
13	K ₀₃	Gumla	23°07'02"N	84°33'21"E	520	
14	K ₀₄	Lohardaga	23°26'17"N	84°38'36"E	590	
15	K ₂₆	Ranchi	23°28'40"N	85°10'17"E	790	
16	K ₁₈	Ranchi East	23°09'04"N	85°28'40"E	500	
17	K ₀₂	Ranchi East	23°28'36"N	85°01'06"E	610	
18	K ₂₄	Ranchi East	23°33'55"N	85°05'05"E	650	
19	K ₀₆	Lohardaga	23°21'06"N	84°39'16"E	570	
20	K ₂₀	Gumla	23°08'22"N	84°46'47"E	800	
21	K ₂₁	Gumla	22°52'39"N	84°53'36"E	500	

22	K ₂₂	Porahat	22°52'35"N	85°18'15"E	690	JH-3
23	K ₁₇	Dhalbum	22°27'10"N	86°37'09"E	350	South Eastern
24	K ₁₆	Saraikela	22°51'42"N	85°56'55"E	390	Plateau Zone

UNDER PEER REVIEW

Table 2. Different primers set designed and employed for the study

Sl. No.	Primer code (Locus)	Sequence	Sl. No.	Primer code (Locus)	Sequence
1	PpSSR02 F PpSSR02 R	GCATCACCAACAAAAGCAT CTA AACCCACTTTCTTCTCCCT TTC	13	PpSSR 14 F PpSSR 14 R	GAGACCGCATACAACCTTC AAT CCTTTTGTGGGAAGAAGA TGT
2	PpSSR03 F PpSSR03 R	CACGAAAGGAGGAAGGTA TGTC CACTACCCAGCAAATAGCA GATT	14	PpSSR 15 F PpSSR 15 R	TGTGAATGTGAGTGATGTT GGA CTTTGTGTTGAGTGTGTGT GGA
3	PpSSR04 F PpSSR04 R	CTCGTTGAGACCATGTTAG AAA CACACACACACACACA ATCT	15	PpSSR 16 F PpSSR 16 R	TCGAAGTTGTATGTAACGG TGG TTTGACTCCTCCTCTTCTCT CG
4	PpSSR05 F PpSSR05 R	GTGGAGTCTTGGGATGTGT GTA GACTGAGCTGTCAACCTTG GAT	16	PpSSR 17 F PpSSR 17 R	CTTACCTTTAGGGATAGAA TTTGCC GATCCTTGTTGTTCTTGCCA TT
5	PpSSR06 F PpSSR06 R	GATGATGAGGATACTGCCA TGA AACATAACAATACGTGAGC CCC	17	PpSSR 18 F PpSSR 18 R	GCATGAGCAATAAGGTGGC TCCCCTAAATCAACTCTTC ATAGG
6	PpSSR07 F PpSSR07 R	TGGTATCTTTGGCATTGTGT TG CAACCACTTATCCTCAAGG GAG	18	PpSSR 19 F PpSSR 19 R	TGTTTTGTCCACTGTCAAC CAT CATGATTCGGTTTCTGCAA GT
7	PpSSR08 F PpSSR08 R	GGAATCTCGACACCTAAAC TGG TCAGACCAATTCACCACGA TAC	19	PpSSR 21 F PpSSR 21 R	ATTGGTTGGTTGGTGCTAG AGT CCTTCCTCTTTTCTCCCTCA TT
8	PpSSR09 F PpSSR09 R	CCTTTCATCCCTTGTTTCG T ATTCCCAGTAGTAAGTCGT CGG	20	PpSSR 22 F PpSSR 22 R	CTTATCGGTGCAAGACAAC AAC TCTCTCTCTCTCTCGCTTTG CT
9	PpSSR10 F PpSSR10 R	GGAATCTCGACACCTAAAC TGG TTCAGACCAATTCACCACGA TAC	21	PpSSR 27 F PpSSR 27 R	AGACTCACCACTCAAATGC TTG CCTCATCTTTCCTTCCCTCT TT
10	PpSSR11 F PpSSR11 R	GAGAGAAAGTGTGTAACG AGAGAGG ATCCCCTCCAATTTTCGTA AG	22	PpSSR 28 F PpSSR 28 R	ATGGTTGGTTGGTGCTAGA GTT CCTTCCTCTTTTCTCCCTCA TT
11	PpSSR 12 F PpSSR 12 R	GAGAGAAAGTGTGTAACG AGAGAGG CCCCTTCAAATTCATCATCA TC	23	PpSSR 29 F PpSSR 29 R	CTTATCGGTGCAAGACAAC AAC TCTAACCTCTCTCTCCCT CAA

12	PpSSR 13 F PpSSR 13 R	AGGATGTGGATGTGTCAAA GAG CTGTCCAATCAATTTAC CTA			
----	--------------------------------	--	--	--	--

F=Forward R=Reverse

UNDER PEER REVIEW

Table 3. Particulars of simple sequence repeat primers used

Sl. No.	Primer code (Locus)	Length of the primer	GC%	Repeat Motif	Amplicons size generated by the software (in bp)
1	PpSSR02 (F,R)	22,22	41,45	(GA)10	110
2	PpSSR03 (F,R)	22,23	50,43	(TG)13	247
3	PpSSR04 (F,R)	22,22	41,45	(A)10	333
4	PpSSR05 (F,R)	22,22	50,50	(GAA)15	196
5	PpSSR06 (F,R)	22,22	45,45	(GAG)6	162
6	PpSSR07 (F,R)	22,22	41,50	(GT)9	147
7	PpSSR08 (F,R)	22,22	50,45	(TG)12	273
8	PpSSR09 (F,R)	21,22	43,50	(A)10	165
9	PpSSR10 (F,R)	22,22	50,45	(TG)12	191
10	PpSSR11 (F,R)	25,21	48,42.8	(GAT)6	235
11	PpSSR12 (F,R)	25,22	48,40.9	(GA)15	110
12	PpSSR13 (F,R)	22,22	45.4,40.9	(AGA)6	247
13	PpSSR14 (F,R)	22,22	40.9,45.4	(CA)21	333
14	PpSSR15 (F,R)	22,22	45.4,50	(AG)8	196
15	PpSSR16 (F,R)	22,22	40,40.9	(AG)9	162
16	PpSSR17 (F,R)	25,22	52.6,41.6	(AC)9	147
17	PpSSR18 (F,R)	19,24	40.9,42.8	(ATG)6	273
18	PpSSR19 (F,R)	22,21	45.4,45.4	(AC)10	165
19	PpSSR21 (F,R)	22,22	45.4,50	(AG)10	304
20	PpSSR22 (F,R)	22,22	45.4,45.4	(GA)19	240
21	PpSSR27 (F,R)	22,22	45.4,45.4	(GAA)6	151
22	PpSSR28 (F,R)	22,22	45.4,45.4	(AG)9	299
23	PpSSR29 (F,R)	22,22	45.4,50	(GA)8	378

Table 4: Polymorphism shown by Microsatellite markers among 24 genotypes of *P. pinnata*.

Sl. No.	Locus	Type of repeat (nucleotides)	Product length in WEBSAT software	Amplified product size range in gel (Full sib)	Range of product length	Total number of alleles (Full sib)	Number of Polymorphic alleles	Percentage of polymorphism
1	PpSSR02	Di	146	143-154	11	6	4	67
2	PpSSR03	Di	304	296-309	13	7	5	71
3	PpSSR04	Mono	308	293-322	29	16	5	31
4	PpSSR05	Tri	183	151-191	40	9	6	67
5	PpSSR06	Tri	330	295-335	40	11	10	91
6	PpSSR07	Di	221	214-234	20	9	8	89
7	PpSSR08	Di	307	297-330	33	11	8	73
8	PpSSR09	Mono	400	386-416	30	17	5	29
9	PpSSR10	Di	307	301-318	17	8	6	75
10	PpSSR11	Tri	235	194-261	67	17	12	71
11	PpSSR12	Di	110	87-135	48	21	12	57
12	PpSSR13	Tri	247	241-305	64	20	16	80
13	PpSSR14	Di	333	316-338	22	10	7	70
14	PpSSR15	Di	196	190-203	13	6	3	50
15	PpSSR16	Di	162	141-185	44	17	10	59
16	PpSSR17	Di	147	143-172	29	15	11	73
17	PpSSR18	Tri	273	243-284	41	12	11	92
18	PpSSR19	Di	165	166-170	4	3	1	33
19	PpSSR 21	Di	304	294-309	15	7	7	100
20	PpSSR 22	Di	240	226-261	35	15	9	60
21	PpSSR 27	Tri	151	147-161	14	5	5	100
22	PpSSR 28	Di	299	290-310	20	9	7	78
23	PpSSR 29	Di	378	360-400	40	19	10	53
Total						270	168	62.2

Table 5. Size of the private alleles (base pair) at each locus for *P. pinnata* genotypes.

Sl. No.	locus	Private alleles in genotypes
1	SSR02	K ₁₈ (129), K ₁₅ (136), K ₀₉ (139), K ₀₆ (143), K ₁₂ (151), K ₀₆ (155), K ₂₆ (158), K ₂₃ (160), K ₂₆ (162), K ₂₂ (163), K ₂₂ (164)
2	SSR03	K ₀₆ (296), K ₀₃ (298), K ₁₁ (304), K ₀₂ (308), K ₁₇ (325), K ₁₈ (326), K ₁₅ (331), K ₁₈ (333), K ₁₈ (337)
3	SSR04	K ₁₉ (293), K ₂₄ (298), K ₂₇ (312), K ₁₄ (319), K ₀₄ (320), K ₁₃ (323), K ₁₂ (329), K ₁₂ (331)
4	SSR05	K ₂₀ (150), K ₀₇ (157), K ₁₄ (169), K ₁₆ (173), K ₂₃ (174), K ₁₈ (176), K ₁₈ (177), K ₁₉ (180), K ₁₇ (194), K ₂₃ (198), K ₁₆ (202), K ₁₂ (203), K ₀₆ (206), K ₁₃ (207),
5	SSR06	K ₂₃ (297), K ₀₉ (301), K ₀₆ (303), K ₀₆ (307), K ₀₇ (315), K ₀₃ (327), K ₁₈ (328), K ₁₃ (330), K ₁₁ (333), K ₁₆ (339), K ₂₃ (341), K ₁₂ (345), K ₁₅ (350), K ₁₀ (353),
6	SSR07	K ₂₆ (211), K ₂₂ (212), K ₂₂ (213), K ₂₅ (231), K ₁₅ (238), K ₂₄ (240), K ₁₁ (241), K ₁₁ (242), K ₁₀ (244), K ₁₈ (245), K ₁₅ (251), K ₁₁ (257)
7	SSR08	K ₁₄ (297), K ₀₃ (305), K ₂₂ (308), K ₂₀ (313), K ₂₇ (314), K ₀₂ (316), K ₂₀ (319), K ₁₉ (321), K ₁₁ (326), K ₁₇ (334), K ₁₀ (339), K ₁₁ (347)
8	SSR09	K ₀₃ (386), K ₁₀ (387), K ₁₂ (390), K ₀₉ (392), K ₁₉ (412)
9	SSR10	K ₀₂ (300), K ₂₁ (301), K ₀₇ (306), K ₀₄ (308), K ₂₄ (317), K ₁₈ (321), K ₂₂ (322), K ₁₁ (324), K ₁₁ (328)
10	SSR11	K ₀₂ (190), K ₂₀ (195), K ₂₅ (199), K ₁₇ (201), K ₂₃ (204), K ₁₇ (206), K ₁₁ (208), K ₀₂ (209), K ₀₇ (210), K ₂₅ (213), K ₁₂ (221), K ₂₃ (223), K ₀₂ (224), K ₀₃ (225), K ₀₉ (228), K ₂₀ (232), K ₂₅ (233), K ₁₈ (250), K ₁₇ (252), K ₁₇ (256), K ₁₄ (259), K ₁₃ (261), K ₂₇ (262)
11	SSR12	K ₀₂ (86), K ₀₄ (90), K ₂₅ (92), K ₁₃ (105), K ₁₁ (107), K ₂₆ (108), K ₀₂ (109), K ₀₉ (111), K ₂₄ (114), K ₁₀ (116), K ₁₉ (130), K ₀₃ (131), K ₂₄ (132), K ₁₄ (137),
12	SSR13	K ₁₀ (241), K ₂₃ (242), K ₁₆ (256), K ₁₅ (270), K ₂₁ (272), K ₁₃ (276), K ₀₉ (281), K ₁₉ (284), K ₁₆ (287), K ₂₅ (289), K ₂₃ (294), K ₁₄ (306)
13	SSR14	K ₂₄ (307), K ₂₃ (310), K ₂₆ (311), K ₁₉ (312), K ₂₆ (313), K ₁₉ (317), K ₂₂ (318), K ₂₀ (320), K ₁₆ (328), K ₁₂ (337), K ₁₁ (339), K ₀₄ (342), K ₀₄ (345), K ₀₄ (351), K ₁₂ (357), K ₀₃ (359)
14	SSR15	NO PRIVATE ALLELE
15	SSR16	K ₀₇ (128), K ₀₂ (134), K ₀₆ (135), K ₁₈ (136), K ₁₇ (137), K ₂₅ (141), K ₂₆ (142), K ₁₁ (155), K ₀₇ (158), K ₂₆ (159), K ₁₉ (161), K ₀₇ (162), K ₀₄ (163), K ₀₄ (165), K ₁₉ (181), K ₁₂ (182), K ₁₇ (184), K ₂₅ (185), K ₁₇ (186), K ₁₂ (189), K ₁₅ (191), K ₂₆ (193), K ₀₆ (195)
16	SSR17	K ₀₇ (142), K ₁₆ (155), K ₁₃ (167), K ₁₆ (171)
17	SSR18	K ₂₄ (239), K ₂₃ (241), K ₀₇ (256), K ₀₉ (282), K ₁₀ (288)

18	SSR19	K ₁₂ (166), K ₂₁ (167), K ₂₅ (168), K ₁₅ (171), K ₂₆ (172), K ₂₃ (174)
19	SSR21	K ₂₀ (296), K ₀₄ (308), K ₁₅ (311), K ₁₁ (313), K ₁₉ (321), K ₁₄ (322), K ₁₄ (323), K ₀₄ (325), K ₁₄ (326), K ₁₂ (330)
20	SSR22	K ₂₀ (222), K ₁₂ (226), K ₂₁ (230), K ₀₉ (232), K ₂₅ (234), K ₂₂ (243), K ₂₃ (270), K ₂₃ (272)
21	SSR27	K ₂₀ (140), K ₂₆ (146), K ₁₃ (170), K ₁₃ (173)
22	SSR28	K ₀₄ (284), K ₁₂ (285), K ₁₂ (286), K ₀₄ (288), K ₁₁ (290), K ₀₇ (301), K ₀₉ (313), K ₁₇ (317)
23	SSR 29	K ₁₂ (360), K ₂₆ (361), K ₂₅ (366), K ₀₆ (385), K ₁₄ (387), K ₁₁ (390), K ₀₄ (391), K ₁₁ (392), K ₂₀ (394), K ₁₃ (395), K ₁₂ (396), K ₀₃ (399), K ₁₈ (400), K ₀₄ (404), K ₁₂ (405), K ₀₃ (406), K ₀₃ (407), K ₀₂ (409), K ₀₄ (411), K ₂₅ (412), K ₀₂ (413), K ₁₁ (414), K ₂₅ (416), K ₁₁ (424)

UNDER PEER REVIEW

Table 6. Nei's analysis of genetic distance and genetic similarity in full sib genotypes. Nei's genetic identity (above diagonal) and genetic distance (below diagonal).

PO	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	*** *	0.16 62	0.17 20	0.14 03	0.22 01	0.16 35	0.28 71	0.02 97	0.07 55	0.16 84	0.07 68	0.17 56	0.28 71	0.11 89	0.06 15	0.13 60	0.21 95	0.23 79	0.35 68	0.26 34	0.18 79	0.00 00	0.07 43	0.09 94
2	1.79 45	*** *	0.18 76	0.12 63	0.10 27	0.13 38	0.15 11	0.08 92	0.25 69	0.22 45	0.21 51	0.16 10	0.24 18	0.14 87	0.15 37	0.12 09	0.02 93	0.05 95	0.10 41	0.05 85	0.24 57	0.05 77	0.04 46	0.15 62
3	1.76 04	1.67 34	*** *	0.15 97	0.25 81	0.30 77	0.29 70	0.26 15	0.21 89	0.26 14	0.20 67	0.12 11	0.10 94	0.07 69	0.06 36	0.15 63	0.10 60	0.32 30	0.15 38	0.22 71	0.25 42	0.04 47	0.15 38	0.07 35
4	1.96 37	2.06 91	1.83 43	*** *	0.32 71	0.22 09	0.18 24	0.16 57	0.19 65	0.37 80	0.31 39	0.13 59	0.22 45	0.15 19	0.22 83	0.11 23	0.21 75	0.06 90	0.19 33	0.20 39	0.13 42	0.21 42	0.09 67	0.13 19
5	1.51 37	2.27 58	1.35 44	1.11 76	*** *	0.34 65	0.27 88	0.07 22	0.13 21	0.14 99	0.13 43	0.17 05	0.05 87	0.04 33	0.22 38	0.19 07	0.31 27	0.18 77	0.28 87	0.31 27	0.12 63	0.19 60	0.15 88	0.12 41
6	1.81 07	2.01 14	1.17 88	1.51 00	1.05 99	*** *	0.41 63	0.17 55	0.25 27	0.16 57	0.25 70	0.30 24	0.16 35	0.19 02	0.13 61	0.28 25	0.21 60	0.16 09	0.23 40	0.23 04	0.25 59	0.08 51	0.24 87	0.16 77
7	1.24 80	1.88 98	1.21 39	1.70 14	1.27 73	0.87 64	*** *	0.16 35	0.04 53	0.09 82	0.18 44	0.23 42	0.15 11	0.16 35	0.15 37	0.40 80	0.29 27	0.28 25	0.40 14	0.23 42	0.31 79	0.11 53	0.25 27	0.12 78
8	3.51 55	2.41 69	1.34 13	1.79 76	2.62 85	1.73 99	1.81 07	*** *	0.23 79	0.23 47	0.39 31	0.18 72	0.31 22	0.13 16	0.18 14	0.16 35	0.11 52	0.13 16	0.04 39	0.20 16	0.18 48	0.14 18	0.13 16	0.13 97
9	2.58 30	1.35 92	1.51 92	1.62 73	2.02 45	1.37 54	3.09 38	1.43 60	*** *	0.26 66	0.21 51	0.19 03	0.16 62	0.19 33	0.19 97	0.01 51	0.16 10	0.04 46	0.16 35	0.24 88	0.13 01	0.11 53	0.16 35	0.24 14
10	1.78 14	1.49 37	1.34 19	0.97 30	1.89 78	1.79 76	2.32 04	1.44 93	1.32 19	*** *	0.34 25	0.19 03	0.22 45	0.12 43	0.15 70	0.02 81	0.02 72	0.06 90	0.12 43	0.14 95	0.08 05	0.10 71	0.06 90	0.09 23
11	2.56 62	1.53 66	1.57 66	1.15 85	2.00 78	1.35 86	1.69 07	0.93 38	1.53 66	1.07 15	*** *	0.38 69	0.27 66	0.18 14	0.10 94	0.06 15	0.07 44	0.06 05	0.07 56	0.13 39	0.11 76	0.08 79	0.13 61	0.18 78

1 2	1.73 94	1.82 64	2.11 08	1.99 57	1.76 88	1.19 60	1.45 17	1.67 56	1.65 94	1.65 92	0.94 95	*** *	0.13 17	0.12 96	0.14 88	0.16 10	0.14 18	0.05 76	0.14 40	0.17 01	0.20 99	0.05 58	0.14 40	0.11 00
1 3	1.24 80	1.41 98	2.21 24	1.49 37	2.83 55	1.81 07	1.88 98	1.16 41	1.79 45	1.49 37	1.28 53	2.02 71	*** *	0.34 19	0.12 29	0.10 58	0.10 24	0.07 43	0.22 30	0.24 88	0.17 34	0.11 53	0.10 41	0.24 14
1 4	2.12 92	1.90 60	2.56 51	1.88 47	3.13 94	1.65 99	1.81 07	2.02 76	1.64 37	2.08 53	1.70 70	2.04 33	1.07 31	*** *	0.10 58	0.02 97	0.14 40	0.11 70	0.21 94	0.12 96	0.18 48	0.11 35	0.05 85	0.26 55
1 5	2.78 94	1.87 31	2.75 52	1.47 70	1.49 69	1.99 46	1.87 31	1.70 70	1.61 07	1.85 17	2.21 30	1.90 50	2.09 62	2.24 60	*** *	0.29 19	0.22 32	0.12 09	0.16 63	0.22 32	0.08 82	0.32 25	0.19 65	0.20 22
1 6	1.99 52	2.11 30	1.85 57	2.18 69	1.65 68	1.26 42	0.89 66	1.81 07	4.19 24	3.57 32	2.78 94	1.82 64	2.24 65	3.51 55	1.23 12	*** *	0.26 34	0.25 27	0.25 27	0.30 73	0.26 01	0.10 09	0.17 84	0.17 04
1 7	1.51 63	3.53 12	2.24 43	1.52 57	1.16 26	1.53 25	1.22 86	2.16 11	1.82 64	3.60 51	2.59 81	1.95 37	2.27 84	1.93 80	1.49 95	1.33 40	*** *	0.18 72	0.31 68	0.34 02	0.19 59	0.12 56	0.14 40	0.06 88
1 8	1.43 60	2.82 23	1.13 00	2.63 1	1.67 30	1.82 69	1.26 42	2.02 76	3.11 00	2.67 31	2.80 56	2.85 43	2.59 92	2.14 54	2.11 24	1.37 54	1.67 56	*** *	0.16 09	0.07 20	0.09 95	0.02 84	0.10 24	0.06 99
1 9	1.03 06	2.26 27	1.87 19	1.64 35	1.24 22	1.45 23	0.91 28	3.12 62	1.81 07	2.08 53	2.58 24	1.93 80	1.50 06	1.51 68	1.79 40	1.37 54	1.14 95	1.82 69	*** *	0.37 44	0.21 33	0.14 18	0.19 02	0.16 77
2 0	1.33 40	2.83 80	1.48 22	1.59 02	1.16 26	1.46 80	1.45 17	1.60 15	1.39 11	1.90 04	2.01 03	1.77 14	1.39 11	2.04 33	1.49 95	1.17 98	1.07 82	2.63 11	0.98 25	*** *	0.30 79	0.15 36	0.10 08	0.19 26
2 1	1.67 21	1.40 38	1.36 97	2.00 84	2.06 92	1.36 29	1.14 60	1.68 83	2.03 98	2.51 92	2.14 09	1.56 09	1.75 21	1.68 83	2.42 85	1.34 67	1.62 99	2.30 74	1.54 52	1.17 79	*** *	0.06 89	0.12 80	0.14 94
2 2	0.00 00	2.85 33	3.10 68	1.54 09	1.62 98	2.46 40	2.16 01	1.95 32	2.16 01	2.23 40	2.43 10	2.88 52	2.16 01	2.17 63	1.13 18	2.29 37	2.07 43	3.56 26	1.95 32	1.87 36	2.67 48	*** *	0.22 69	0.06 77
2 3	2.59 92	3.11 00	1.87 19	2.33 66	1.84 01	1.39 16	1.37 54	2.02 76	1.81 07	2.67 31	1.99 46	1.93 80	2.26 27	2.83 85	1.62 69	1.72 37	1.93 80	2.27 89	1.65 99	2.29 46	2.05 60	1.48 32	*** *	0.16 77
2 4	2.30 84	1.85 64	2.61 08	2.02 57	2.08 65	1.78 57	2.05 71	1.96 80	1.42 11	2.38 24	1.67 26	2.20 68	1.42 11	1.32 61	1.59 85	1.76 94	2.67 68	2.66 11	1.78 57	1.64 72	1.90 11	2.69 21	1.78 57	*** *

UNDER PEER REVIEW

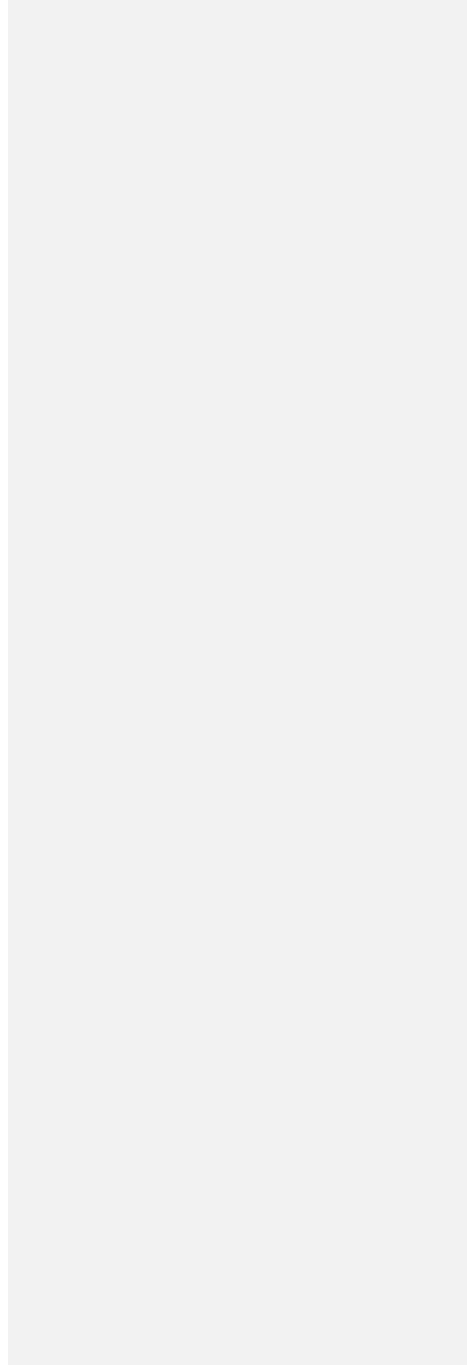


Table 7: Partitioning of genetic variability among the different source of variation (AMOVA design) in *P. pinnata*.

Source of variation	d.f.*	Sum of squares	Variance components	Percentage variation
Among populations	23	1105.938	8.01404 Va	77.92
Among individuals within populations	48	0.000	-2.27083 Vb	-22.08
Within individuals	72	327.000	4.54167 Vc	44.16
Total	143	1432.938	10.28487	

d.f.* = Degree of freedom

Significance test (1023 permutations) at Significance Level = 0.05

P-value = 0.00000 for all variance component.

UNDER PEER REVIEW

Table 8. Comparison of allelic differences in between the highest oil yielding and lowest oil yielding genotypes

Locus	Type of repeat (nucleotides)	Genotypes K ₁₉ (allele size) and types of alleles	Genotypes K ₂₁ (allele size) and types of alleles
SSR02	Di	145/ 145(Homozygous)	145/ 145(Homozygous)
SSR03	Di	302/ 302(Homozygous)	300/ 300(Homozygous)
SSR04	Mono	293/ 293(Homozygous)	309/ 309(Homozygous)
SSR05	Tri	151/ 180(Heterozygous)	155/ 186(Heterozygous)
SSR06	Tri	300/ 322(Heterozygous)	300/ 325(Heterozygous)
SSR07	Di	223/ 223(Homozygous)	217/ 217(Homozygous)
SSR08	Di	317/ 317(Homozygous)	300/ 300(Homozygous)
SSR09	Mono	395/ 395(Homozygous)	407/ 407(Homozygous)
SSR10	Di	310/ 310(Homozygous)	301/ 301(Homozygous)
SSR11	Tri	197/ 246(Heterozygous)	217/ 217(Homozygous)
SSR12	Di	97/ 124(Heterozygous)	101/ 101(Homozygous)
SSR13	Tri	251/ 298(Heterozygous)	247/ 300(Heterozygous)
SSR14	Di	321/ 321(Homozygous)	329/ 329(Homozygous)
SSR15	Di	195/ 195(Homozygous)	195/ 195(Homozygous)
SSR16	Di	146/ 181(Heterozygous)	146/ 167(Homozygous)
SSR17	Di	149/ 169(Heterozygous)	149/ 149(Homozygous)
SSR18	Tri	249/ 275(Heterozygous)	247/ 272(Heterozygous)
SSR19	Di	?/ ?	167/ 167(Homozygous)
SSR21	Di	300/ 300(Homozygous)	294/ 294(Homozygous)
SSR22	Di	237/ 254(Heterozygous)	230/ 240(Heterozygous)
SSR27	Tri	149/ 149(Homozygous)	148/ 148(Homozygous)
SSR28	Di	291/ 291(Homozygous)	295/ 295(Homozygous)
SSR29	Di	369/ 369(Homozygous)	362/ 373(Heterozygous)

Figure Legends

Fig. 1. Total oil content (%) in 24 genotypes of *Pongamiapinnata* recorded above average for the trait.

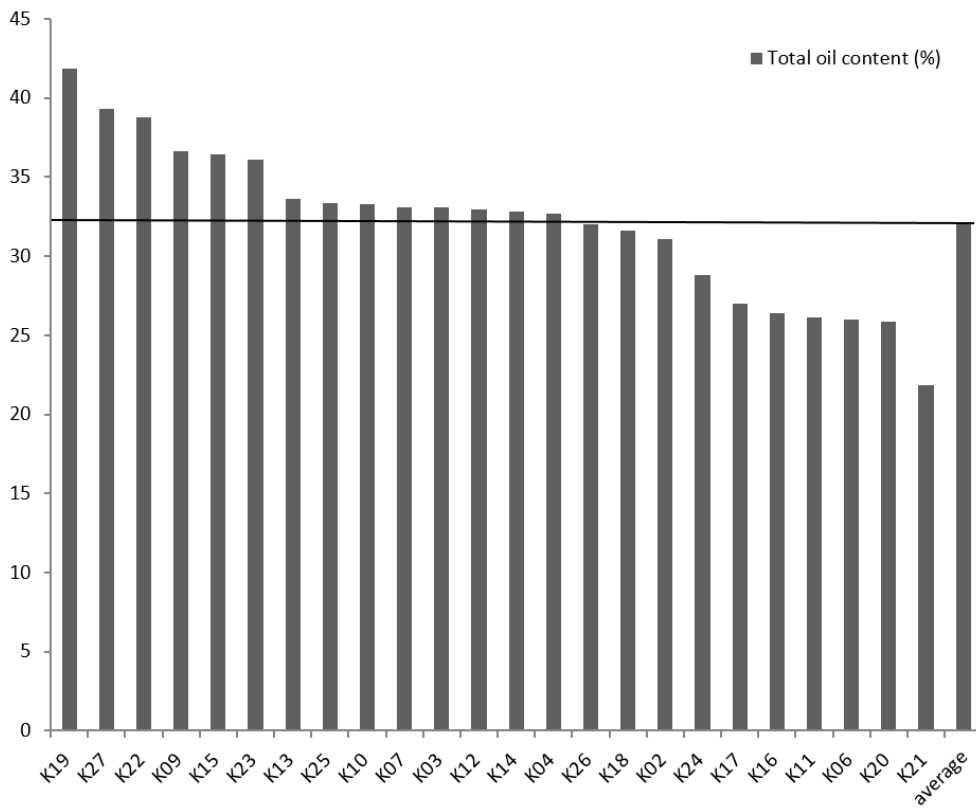


Figure Legends

Fig. 2(a, b, c and d). Gel picture showing the amplicons generated by SSR primer PpSSR5, PpSSR11, PpSSR12 and PpSSR14. Numbers in the top correspond to the *Pongamia* genotypes given in Table 1. M= Molecular markers (100bp ladder, Low range DNA ruler).

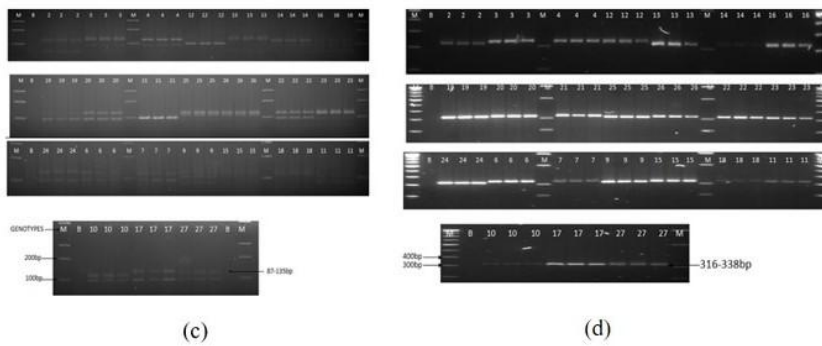
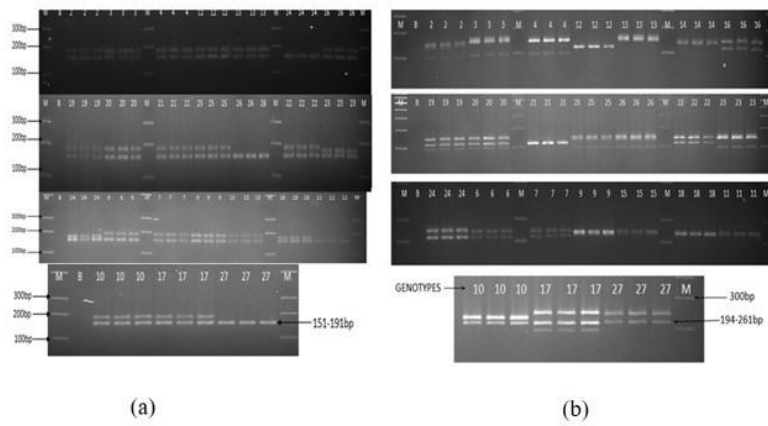


Figure Legends

Fig. 3 (a, b). Gel picture showing the amplicons generated by SSR primer PpSSR16 and PpSSR17. Numbers in the top correspond to the *Pongamia* genotypes given in Table 1. M= Molecular markers (100bp ladder, Low range DNA ruler).

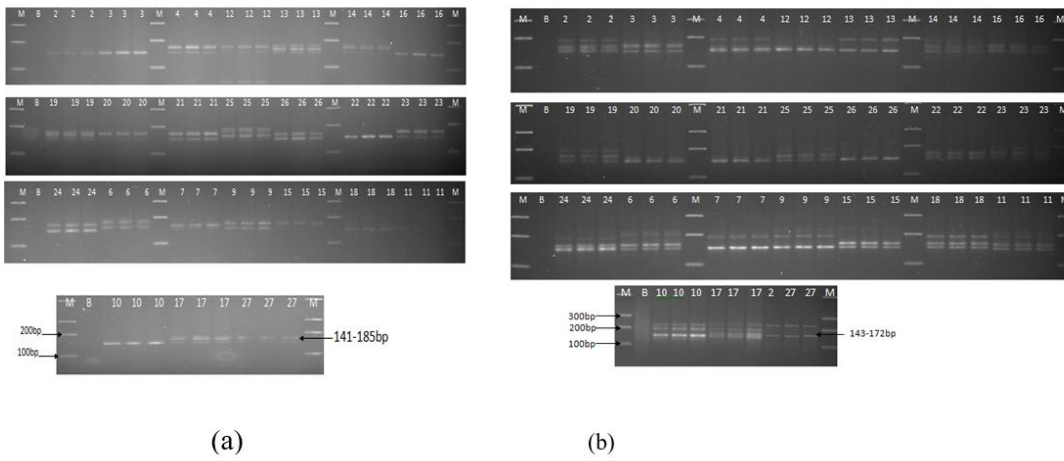
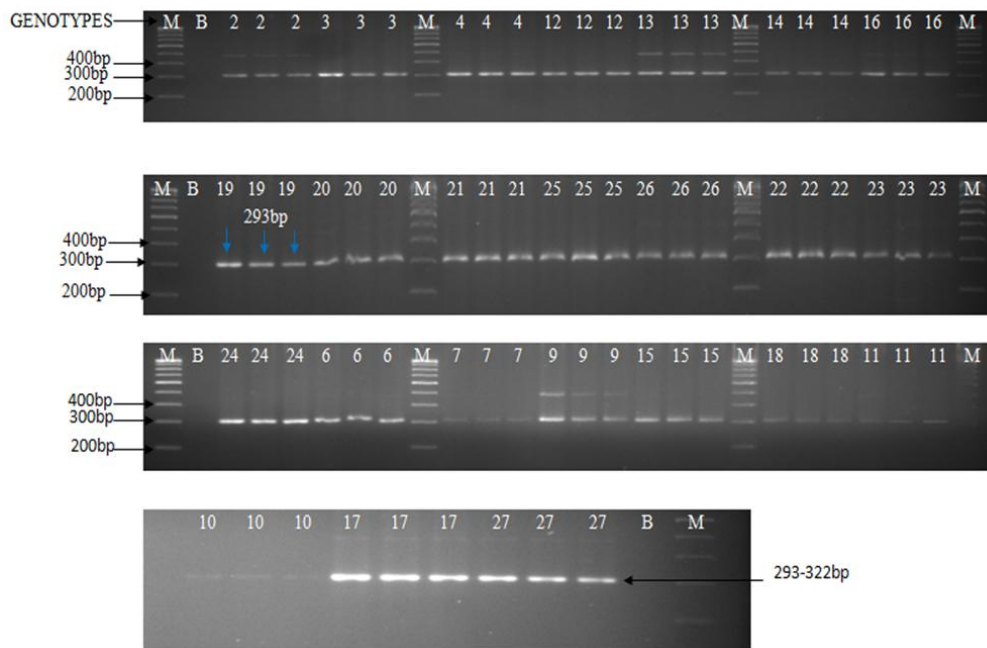


Figure Legends

Fig. 4. Different band pattern obtained by K19 at locus PpSSR04



UNDE

Figure Legends

Fig. 5. Dendrogram based on Nei's analysis of genetic distance and genetic similarity in full sib genotypes.

