Journal of Plantation Crops 2026, 54(1): 16-24
https://doi.org/10.25081/jpc.2026.v54.i1.10079

Research Article

Microsatellite-inferred genetic architecture of oil palm germplasm

H. P. Bhagya[1,2]*, B. Kalyana Babu[2]

[1]*ICAR-Directorate of Cashew Research, Puttur, Dakshina Kannada-574202, Karnataka, India*
[2]*ICAR-Indian Institute of Oil Palm Research, Pedavegi-534450, Andhra Pradesh, India*

(Received: April 02, 2025; Revised: July 07, 2026; Accepted: July 10, 2026)

*Corresponding author: H. P. Bhagya, E-mail: bhagya509@gmail.com

Abstract

The oil palm is a plant that produces vegetable oil and is pollinated by insects. It can grow for many years. To study the population structure trends and genetic diversity of 150 oil palm lines from Indian indigenous collections. A total of 54 SSR loci were used to amplify DNA from 150 oil palm samples. The amplification results revealed a considerable degree of genetic heterogeneity among oil palm individuals based on 54 polymorphic microsatellite loci. The average PIC value for all polymorphic loci across oil palm genotypes was 0.44, ranging from 0.19 to 0.78. However, the mixing of different genotypes was better explained by the SSR markers. Of the 54 SSRs examined in this study, mEgCIR0246, mEgCIR3358, mEgCIR0782, and mEgCIR0779 show a respectably high degree of polymorphism. The mean of Heterozygosity (Ho) was 0.28 and 150 genotypes were split into two main groups using unweighted pair-group technique. The oil palm lines were found to have originated from two separate sources. The SSR method was found to be effective and reliable for measuring the genetic diversity in oil palm. Knowing about genetic diversity and how different populations are structured is very important for managing genetic resources in a way that helps future breeding programs.

Keywords: Genetic diversity, Oil palm genotypes, Polymorphic information content, Population Structure, SSR marker

Introduction

The oil palm belongs to Arecaceae family and is scientifically called *Elaeis guineensis* Jacq. (2n=32). It is also referred to as the oil king of the world (Babu *et al.,* 2021). This is the most edible vegetable oil yielding crop, producing up to 4.0-6.0 metric tons per hectare per year with proper agricultural management (Henson & Harun, 2005). It has 16 pairs of chromosomes and the estimated haploid genome size is about 1.8 gigabytes, as reported by Singh *et al.* (2013). The Dutch imported seeds of oil palm from African country during 1848, and four seedlings were planted at Java Bogor Botanical Gardens, Indonesia. Commercial oil palm farming started in Malaysia in 1917 (Corley & Tinker, 2003). Malaysia, Nigeria, Indonesia, Ecuador, Thailand, Colombia, Papua New Guinea, Cote d'Ivoire, Congo, and India are the top countries in the world for producing oil palm. Indonesia is the top producer of palm oil, with a total output of 32.60 million tons and an average yield of 3.57 metric tons per hectare. Malaysia ranks second with about 18.93 million tons produced and a productivity rate of 3.83 metric tons per hectare, according to Bhagya *et al.* (2020). India brought in 12.71 million tons of edible oil during the 2014-15 period, and out of that, 70% was palm oil, according to Bhagya *et al.* (2020). There hasn't been enough study on the oil palm variety using DNA markers like isozymes, RAPDs, and RFLP as mentioned in the references by Hayati *et al.* (2004), Maizura *et al.* (2006) and Barcelos *et al.* (2000). In genetic research, it is crucial to examine DNA sequence variation.

The great diversity of oil palm germplasm can be investigated for desirable features in order to produce new types using molecular breeding methods. Genome-based tests, such as molecular markers, are seen as useful tools for studying ecosystems genetically, understanding how populations are structured, measuring genetic diversity, and helping in selecting plants with good genetic diversity through markers. These methods are supported by studies from Zaki *et al.* (2012), Zhou *et al.* (2015), Babu *et al.,* (2023) and Bhagya *et al.* (2024)., Simple sequence repeats are considered one of the best and most effective molecular markers available today. This is because they have strong inheritance patterns, can have many different forms, are located in specific places on chromosomes, mutate quickly in genetics, and are easy to analyze. Additionally, SSRs contain a lot of genetic information.

Research on molecular markers in oil palm has offered insights into a multi-omics approach and offers a foundation for more detailed future studies on oil palm (Xu *et al.*, 2024). Recently, molecular characterization of indigenously collected oil palm genotypes of exotic sources was carried out in India (Bhagya *et al.*, 2018, 2019, 2020, and 2024; Babu *et al.,* 2021). Babu *et al.* (2019) created the first microsatellite database for oil palm, called OpSatdb. They worked on establishing and validating the entire genome and specific microsatellite locations in the oil palm species, *Elaeis guineensis.* The polymerase chain reaction, in particular, facilitates their detection (Heckenberger, 2002; Zhou *et al.,* 2015). SSR markers have become one of the most important and dependable tools for studying genetic diversity and population structure in oil palm (Zhou *et al.,* 2015; Singh *et al.,* 2008).

Material and methods

Planting material

The study included 150 oil palm germplasm lines collected from different states in India, such as Tamil Nadu,

Andhra Pradesh, Karnataka, Maharashtra, and Andaman Nicobar. These samples were kept at the field gene bank of ICAR-IIOPR, which is the Indian Institute of Oil Palm Research located in Pedavegi.

Extraction of DNA

DNA was taken from young, unwrapped oil palm leaves that hadn't been opened yet. A special method called modified CTAB was used, which was first described by Murray and Thomson (1980). To check the DNA's quality, it was run on a gel made from 0.8 percent agarose. The amount of DNA was measured using a UV spectrophotometer and then brought up to a concentration of 50 ng per microliter. The DNA was stored in a freezer set at -20 degrees Celsius for future use.

Amplification of PCR and electrophoresis

For the PCR reaction, 0.2 micromolar concentrations of both forward and reverse primers, 2 microliters of 2 mM dNTPs, 0.2 microliters of 0.2 U Taq DNA polymerase (Invitrogen, USA), and approximately 50 ng of template DNA were added to make up a total volume of 20 microliters. The amplification of the genomic SSR loci was performed using a thermocycler (MJ Research, USA) programmed for 35 cycles, each consisting of 30 seconds at 94 °C, 30 seconds at varying annealing temperatures specific to each primer, 1 minute at 72 °C, followed by a final extension step of 10 minutes at 72 °C and holding at 4 °C.

To separate the PCR products, a 2.5% superfine resolution (SFR) agarose gel was used. Electrophoresis was completed after three hours at room temperature with a voltage of 100 volts. After the gel was stained with ethidium bromide, the bands were analyzed using the Bio Imaging System (Bio-Rad, USA). For data analysis, only the bands that showed the highest clarity and consistency were considered. To ensure the reliability of the results, two individuals independently repeated the PCR and scored the bands. The molecular weight of each band was determined using a 100-bp DNA ladder.

DNA analysis

The 54 SSR loci used in this study. Using Power Marker V 3.0 (Liu & Muse, 2005), these SSRs were analyzed across 150 germplasm lines to generate basic statistical data, including PIC value, allelic richness, heterozygosity (Ho), gene diversity (He), and inbreeding coefficient (f). The tree was constructed using the CS Chord 1967 frequency matrix and the Unweighted Pair Group Method Analysis (UPGMA).

The gene flow and population structure among oil palm accessions were examined using a model-based clustering method available in the Structure v. 2.3.4 software package.

This method assumes that the sample contains several subpopulations. According to the admixture model (ALPHAPROPSD=0.20), an accession can belong to more than one subgroup. The number of subgroups (K) in the entire population was determined by running the method at different K values ranging from 1 to 10, and performing 5 independent runs for each K value. During the burn-in phase, 100,000 replications were conducted.

Results

SSR polymorphism and genetic diversity

The SSR data was analyzed with a 100-bp DNA ladder, and the findings were then used for further studies on population structure and genetic diversity. Figure 1 shows the banding pattern observed in 150 oil palm genotypes at the SSR locus mEgCIR3358. Power markers were used to evaluate the PIC value, allelic richness, heterozygosity, inbreeding coefficient, and gene diversity (He) (Table 1). The chance that two alleles picked at random from the population are different is called gene diversity. Anticipated heterozygosity (He) is a common term used to describe it. The diversity of genes ranged between 0.21 and 0.81, with an average of 0.52. In terms of heterozygosity, or observed heterozygosity (Ho), an average of 0.28 was recorded, with a range of 0.00 to 0.87.

The largest main allele frequency was discovered in the SSR locus mEgCIR3232; the second-highest was found at mEgCIR0192. Both of the two alleles (170 bp and 180 bp) that distinguish the SSR locus mEgCIR3232—which has amplified 150 oil-palm genotypes—appear at an 88% frequency. The least common major frequency of alleles was observed at the mEgCIR0246 SSR loci (260 bp at 24%), mEgCIR3400 (120 bp at 34%), SMG00217 (210 bp at 39%), and SMG00155 (260 bp at 39%). There is a theoretical range of -1 to 1 for the inbreeding coefficient (F). For landraces that are genetically heterozygous, there is only one. At a single SSR locus, the inbreeding coefficient ranged from -0.32 to 1.00.

Population structure

The LnP (D) value was used to group all the genotypes and figure out the exact population structure (K). We tested K values from one to ten, with ten iterations each. Figure 2 shows that K=2 had the highest change in lnP (D) value. The

Table 1: Summary of SSR marker polymorphism and the most informative loci identified among the 54 SSR markers

Parameter	Value
Number of SSR loci analysed	54
Total alleles detected	169
Mean alleles per locus	3.13
Allele range per locus	2-6
Mean major allele frequency	0.58
Major allele frequency range	0.24-0.88
Mean expected heterozygosity (He)	0.52
Expected heterozygosity range	0.21-0.81
Mean observed heterozygosity (Ho)	0.28
Observed heterozygosity range	0.00-0.87
Mean polymorphism information content (PIC)	0.44
PIC range	0.19-0.78
Mean inbreeding coefficient (*f*)	0.47
Inbreeding coefficient range	-0.32-1.00

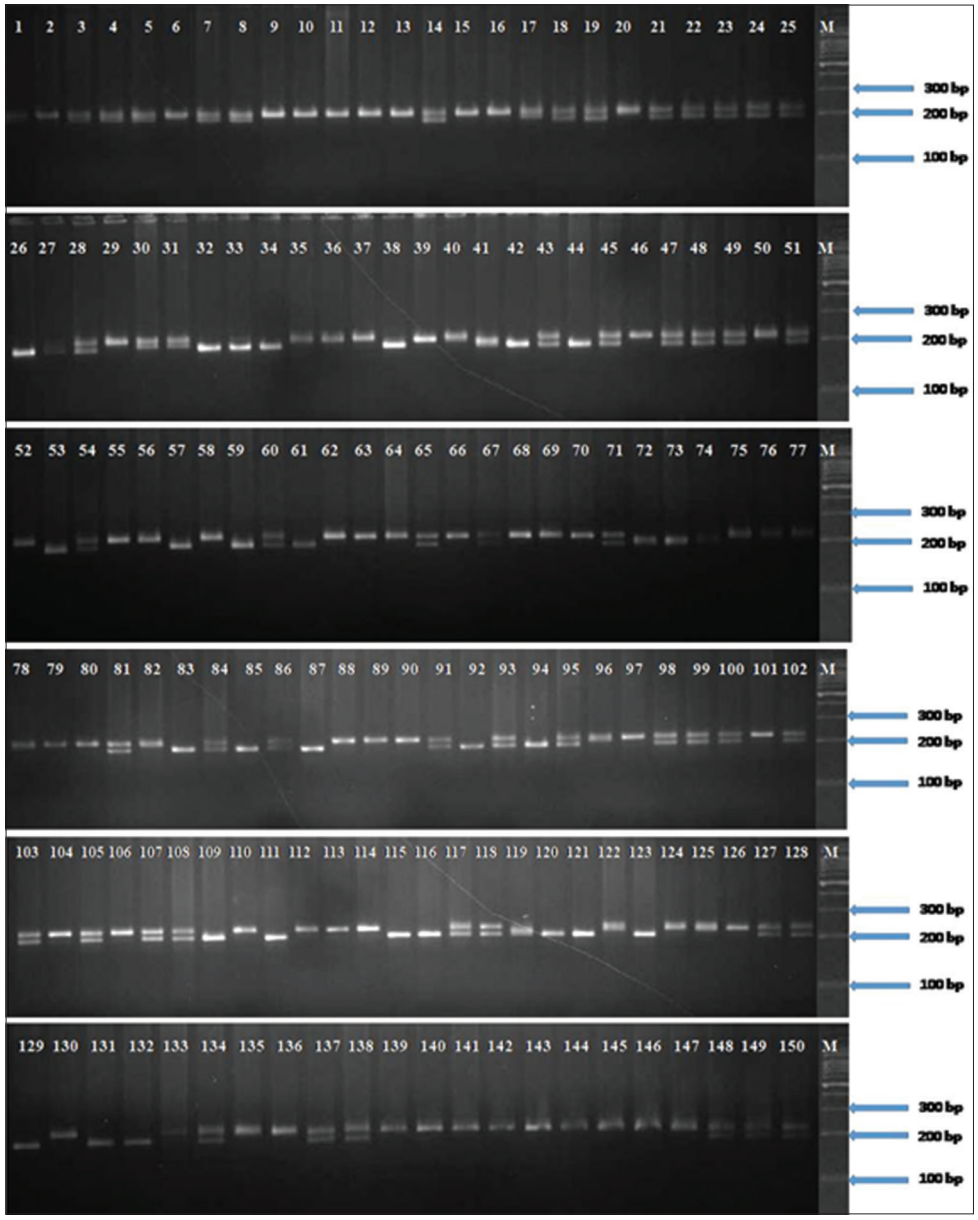

Figure 1: The banding pattern of 150 oil palm genotypes with the genomic SSR locus mEgCIR3358

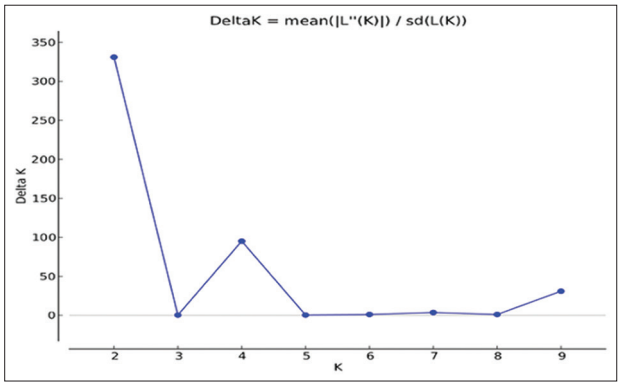

Figure 2: Identification of the appropriate sub-population number (k) against delta K and the maximum K value observed at K=2

oil palm genotypes were divided into two groups based on the ancestry inferred at K=2. The population structure found using the STRUCTURE program (version 2.3.4) closely matched the regional patterns seen in the phylogenetic tree based on germplasm (Figure 3).

The genetic population structure for a native collection that included the oil palm genotype from five states was evaluated using a list of 54 SSR loci distributed across all chromosomes. The phylogenetic study showed a minimum of two demographic groups, depending mostly on their geographic origin. Gene flow between the two population groups has also been detected by the phylogenetic analysis.

Discussion

SSR polymorphism and genetic diversity

In this study, the average He value was 0.52, which is higher than the 0.437 found by Ting *et al.* (2010) using 15 SSR markers, but lower than the 0.644 reported by Bakoumé *et al.* (2015) and the 0.762 from Okoye *et al.* (2016). Even though some selection and breeding activities have been conducted on the materials, the high mean values of He observed in the oil palm samples indicate a significant level of genetic diversity. The goal of oil palm breeding is to make

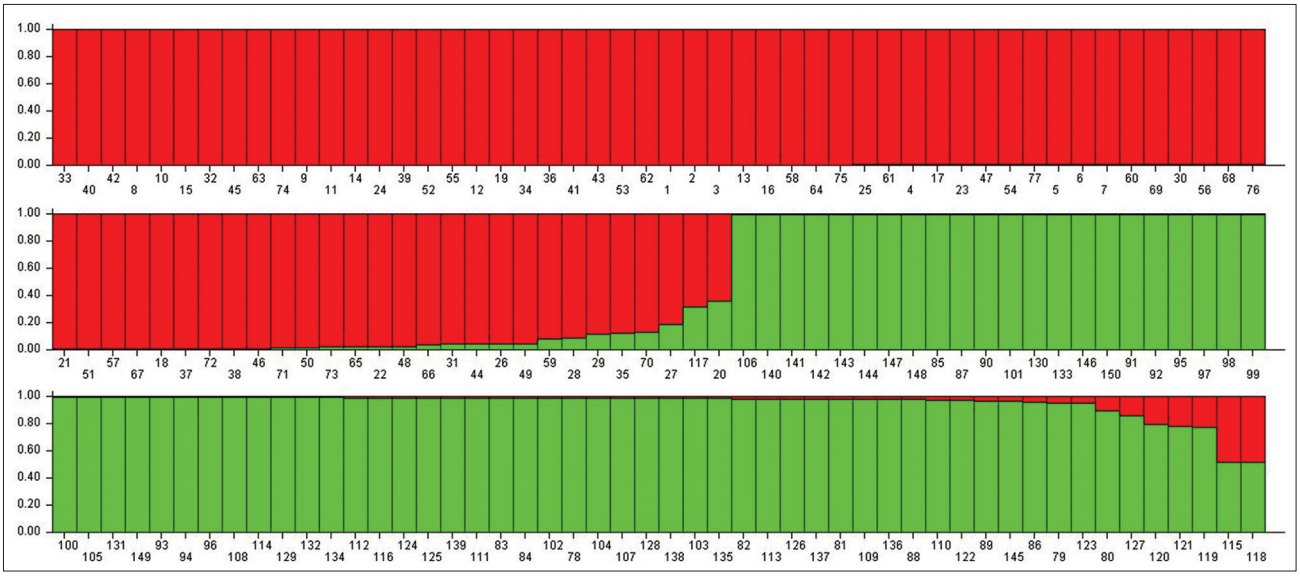

Figure 3: The population structure of oil palm genotypes (a and b) using genomic SSR loci

the best use of different agronomic traits while preventing the negative effects of inbreeding, which can reduce yield. Allelic variety is increased by the varied open-pollinated germplasm found in natural oil palms.

Both the present study's (He=0.21) and Ting *et al.* (2010) findings (He=0.340) indicated that the native oil palm germplasm at Pedavegi had low levels of genetic variation. Similar findings on genetic diversity were observed for He, which ranged from 0.153 to 0.643 (Okoye *et al.,* 2016) and from 0.109 to 0.261 (Hayati *et al.,* 2004). Hayati *et al.* (2004) found that the highest value was observed in the *E. guineensis* population from Cameroon in Africa, with a mean of 0.261, while the lowest was recorded in the *E. guineensis* population groups from Senegal, having a mean of 0.109. Ong *et al.* (2015) found an average of 0.398. In the current investigation, He varies between 0.21 and 0.81, suggesting that a wide range of genotypes, from low to high genetic variation, are available. This implies that there is potential for selecting extremely diverse material from the oil palm germplasm that was indigenously collected by the ICAR-IIOPR. The high mean values of He depicted the high genetic variability in the oil palm samples evaluated in the study, notwithstanding that the materials have undergone some selection and breeding activity. In fact, the principle in oil palm breeding is to exploit complementarities of different agronomic traits while preventing inbreeding effects that depress yield. Open-pollinated germplasm that prevails in the natural oil palm is rich in diversity, thereby increasing allelic diversity. Results of the present study (He = 0.21) and the previous study by Ting *et al.* (2016) (He = 0.340) revealed the presence of low genetic variation among indigenous oil palm germplasm at Pedavegi.

Within a range of 0.00 to 0.87, the mean observed heterozygosity (Ho) was 0.28. The SSR loci with the highest heterozygosity were mEgCIR0555 (0.87), mEgCIR3402 (0.65), and mEgCIR0878 (0.54). Nevertheless, mEgCIR0268 had the lowest heterozygosity (0.00), followed by mEgCIR3649 (0.01), mEgCIR3400 (0.02), and mEgCIR3376. Heterozygosity values above the mean were observed in twenty-eight SSR loci.

The SSR analysis revealed a marked difference between the expected heterozygosity (He=0.52) and the observed heterozygosity (Ho=0.28), indicating a deficit of heterozygous individuals within the evaluated germplasm. Such a pattern may be attributed to several factors, including the narrow genetic base of cultivated oil palm in India, repeated use of a limited number of elite parental lines in breeding programmes, and population subdivision among germplasm originating from different breeding populations (Wahlund effect). The limited number of founder introductions and subsequent selection for superior agronomic traits may have further reduced allelic recombination, resulting in lower observed heterozygosity. Similar reductions in heterozygosity have been reported in cultivated oil palm.

Low observed heterozygosity is due to narrow genetic base and oil palm is selected from the same mother parents unknowingly and due to cross pollinated nature, each individual tree has been considered as single genotype in the study with the same accessions and genotypically they might have originated from same mother palm. According to earlier research on oil palm (Ong *et al.*, 2015), the heterozygosity among the chosen genotypes of the plant was found to be consistent.

The major allele frequency of the SSR locus mEgCIR3232 was highest, whereas mEgCIR0192 was second highest. All 150 oil palm genotypes were amplified by the two alleles (170 bp and 180 bp) that differentiate the SSR locus mEgCIR3232; the frequency of the 180 bp allele was 88 per cent. The SSR loci with the lowest major allele frequency were the mEgCIR0246 (260 bp, 24% frequency), mEgCIR3400 (120 bp, 34% frequency), SMG00217 (210 bp, 39% frequency), and SMG00155 (260 bp, 39% frequency). These results were similar to those of previous investigations on population organization and molecular characterization

(Lanes *et al.,* 2015). Arecaceae is the family that includes the macaw palm (Acrocomia aculeata). Because these alleles preserve the pertinent allelic variety required for plant breeding, this is the explanation for their behaviour (Odong *et al.,* 2011).

Given the aforementioned findings, it was discovered that loci exhibiting high PIC, He, and Ho displayed decreased inbreeding coefficients, and vice versa. In a frequently used population of Deli dura, Hardon (1970) discovered a significantly substantial negative association between the inbreeding coefficient and bunch yield. As Fx increased, there was a noticeable decline in bunch number. Only those traits linked to fitness seem to be affected by inbreeding depression (Hill & Mackay, 1981).

The fixation index (f) varied from −0.32 to 1.00 across loci, reflecting considerable variation in heterozygosity among marker loci. Negative values indicate an excess of heterozygotes at certain loci, whereas positive values indicate heterozygote deficiency. Loci exhibiting an f value of 1.00 showed complete absence of heterozygotes among the evaluated accessions. This pattern is most likely due to allele fixation within the sampled germplasm and may also reflect the presence of null alleles, which can underestimate heterozygosity in SSR analyses. Although null alleles were not specifically evaluated in the present study, this possibility should be considered when interpreting these loci.

Based on the earlier SSR analysis, certain criteria have been established to determine if a primer is highly polymorphic. These criteria can be used in molecular breeding programs for oil palm. They include a higher PIC value, greater heterozygosity, high genetic diversity, a lower frequency of the major allele, and a larger number of polymorphic alleles.

Among the SSR loci evaluated, mEgCIR0246 exhibited the highest level of polymorphism, amplifying six alleles with a polymorphic information content (PIC) of 0.78 and expected heterozygosity (He) of 0.81, indicating excellent discriminatory power for assessing genetic diversity. The relatively high observed heterozygosity (Ho=0.45) further suggests substantial allelic variation within the germplasm. The low major allele frequency (24%) indicates a balanced distribution of alleles, making this marker highly effective for distinguishing among genotypes. The locus mEgCIR3358 detected four alleles and recorded a PIC value of 0.64, He of 0.70, and Ho of 0.44, with a major allele frequency of 41%. These values indicate that this marker possesses a high degree of informativeness and contributes significantly to the assessment of genetic variability. Similarly, mEgCIR0779 and mEgCIR0782 each amplified five alleles, with identical PIC values of 0.61. The expected heterozygosity was 0.68 and 0.67, respectively, while the observed heterozygosity reached 0.42 for mEgCIR0779 and 0.47 for mEgCIR0782. Their moderate major allele frequencies (40% and 43%, respectively) indicate good allelic diversity and balanced allele distribution within the population (Table 2).

Table 2: Most informative SSR loci

SSR locus	Alleles	PIC	*He*	*Ho*	Major allele frequency (%)
mEgCIR0246	6	0.78	0.81	0.45	24
mEgCIR3358	4	0.64	0.70	0.44	41
mEgCIR0779	5	0.61	0.68	0.42	40
mEgCIR0782	5	0.61	0.67	0.47	43

He=expected heterozygosity, Ho=observed heterozygosity, PIC=polymorphism information content, f=inbreeding coefficient. The four loci listed exhibited the highest PIC values and are therefore the most informative markers identified in this study

Overall, these four loci demonstrated superior informativeness compared with the remaining SSR markers, as evidenced by their high PIC values (>0.60), elevated gene diversity, and relatively low major allele frequencies. According to the classification proposed by Botstein *et al.* (1980), SSR markers with PIC values greater than 0.50 are considered highly informative, making these loci particularly suitable for germplasm characterization, genetic diversity analysis, population structure assessment, cultivar identification, and marker-assisted breeding in oil palm. Among them, mEgCIR0246 emerged as the most powerful marker due to its highest allele number, greatest genetic diversity, and lowest major allele frequency, highlighting its potential for future molecular breeding and conservation studies.

Babu *et al.* (2014) observed that similar criteria were taken into account while choosing the molecular markers present in finger millet. Except for the SSR loci mEgCIR0465 and mEgCIR0425, which were found to have trimer repeat motifs (CCG), all of the polymorphic loci under investigation demonstrated dimer repeat motifs. There were 39 loci out of the 54 with the GA repeat motif. Repeat motifs of GT, CA, and GTAT were found in the remaining SSR loci. Babu *et al.* (2014) found the same results when they studied the finger millet genotypes.

Using SSR markers, the 150 oil palm genotypes were grouped into two main clusters, A and B (Fig. 4). Cluster B contained 75 genotypes, while the primary cluster A also had 75 genotypes. Collections from five separate states make up the A and B clusters. With mixes of several germplasm lines, the main clusters once again created sub-clusters among themselves.

The geographic proximity is likewise relatively close, and the genotype admixture in the main clusters may have resulted from the same parents crossing those genes. The admixture of alleles between genotypes during genotype crossing, breeding, and selection, as well as during the domestication of oil palm accessions from one location to another, may also be related to the four palms—sourced from the Amsterdam Botanical Garden (Bogor)-that gave rise to the modern oil palm industry worldwide (Abimbola *et al.,* 2016). These four palm trees were planted as ornamental plants in Java in 1848 (Jalani, 2012). Rajanaidu (1994) and Abimbola *et al.* (2016) were very worried about this because it showed that the genetic diversity used to grow oil palm in Malaysia was very limited.

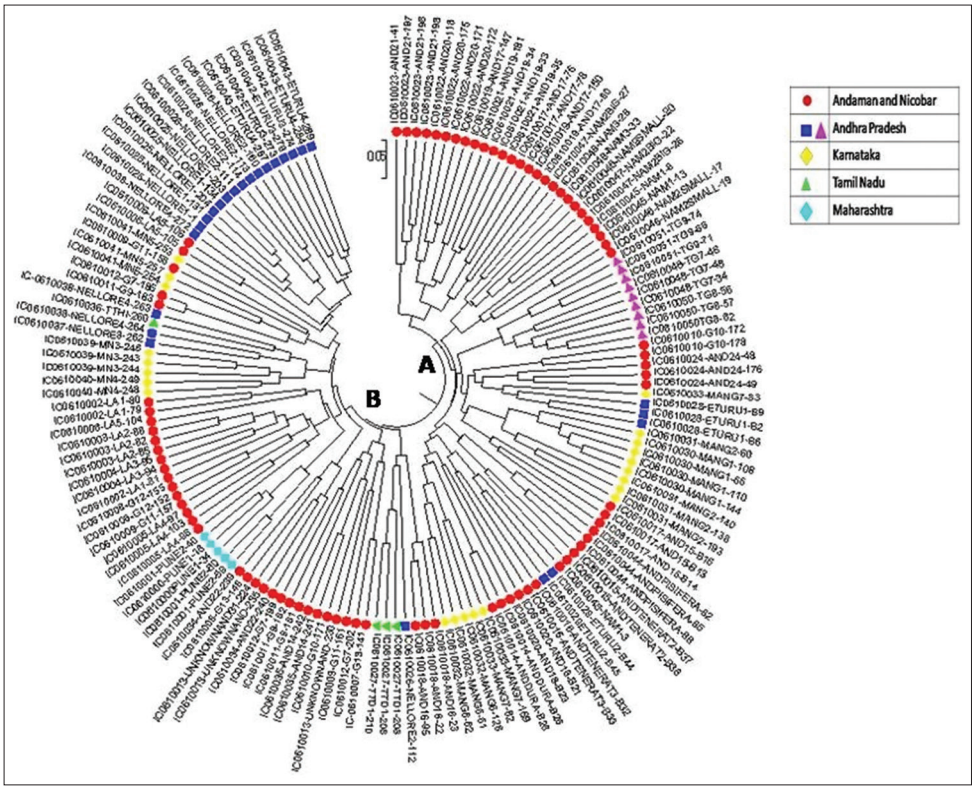

Figure 4: The dendrogram of one hundred and fifty oil palm genotypes as obtained from Power Marker software based on UPGMA analysis

The oil palm genetic material didn't form groups that could be clearly identified through cluster analysis. Some plant types from the same area were grouped together, and others from different areas were also grouped. This indicates there's no clear link between where the plants came from, and similar results were seen in Abimbola *et al.* (2016) and Sapey *et al.* (2017). According to Murthy and Arunachalam (1966), the random way the plant types were spread across different groups suggests that factors other than location, like genetic changes over time, exchange of plant materials, and natural or human-driven selection, are behind the diversity seen.

Population structure

To obtain an accurate population structure (K), all genotypes were grouped according to their LnP (D) values. K values that range from one to ten (with ten cycles) were examined. K 2 yielded the highest ΔK value. According to Ong *et al.* (2015), the K value was 2. Previously, K=8 was reported by Cochard *et al.* (2009), while K=5 was claimed by Bakoumé *et al.* (2015). This implies that the diverse oil palm germplasm from different countries exhibited a robust population pattern. The results showed that, among the 150 oil palm genotypes, there was less germplasm exchange.

According to STRUCTURE analysis, there were allele admixtures in 15 genotypes. Among these, the genotype IC0610007-141 was classified as Tenera. IC0610010-171 was identified as Dura, while IC0610010-172 and IC0610010-178 were classified as Tenera. The genotype IC0610012-202 was also identified as Tenera, whereas IC0610041-253 belonged to the Pisifera fruit form. The genotypes IC0610026-114, IC0610019-80, and IC0610017-B13 were classified as Dura. IC0610017-B16 and IC0610020-B21 were identified as Tenera, while IC0610020-B23 was classified as Pisifera. The genotypes IC0610014-B26 and IC0610014-B28 belonged to the Dura fruit form, and IC0610029-B45 was classified as Tenera. Overall, the admixture group comprised seven Tenera, six Dura, and two Pisifera genotypes. The inferred ancestry at K=2 suggested that the oil palm genotypes were grouped into two populations. There was good correspondence between the geographic pattern observed based on accession in the phylogenetic tree and the population structure identified using STRUCTURE software (STRUCTURE v.2.3.4). Although the population groups corresponded largely to geographic origins, there were some notable exceptions.

Population 1 (G1) consisted of 78 indigenous oil palm genotypes collected from Andaman and Nicobar, Andhra Pradesh, Karnataka, Tamil Nadu, and Maharashtra. Among 78, eight genotypes have admixture, viz., IC0610007-141 with 35%, IC0610017-B16 with 34%, IC0610010-171 with 20%, IC0610026-114 with 19%, IC0610012-202 with 19%, IC0610010-178 with 18%, and IC0610010-172 and IC0610041-253 with 10% each. In G1, all the admixtures observed belong to Andaman except IC0610041-253. The population 2 (G2) consisted of 72 oil palm genotypes, and among these, 7 genotypes had more admixtures, which varied from 8% to 45%, viz., IC0610020-B21 and IC0610017-B13 (45% each); IC0610020-B23, IC0610014-B28, and IC0610014-B26 (25% each); IC0610029-B45 (10%); and IC0610019-80 (8%). The admixture observed in G2 belongs to Andaman collections except IC0610029-B45.

Bayesian population structure analysis divided the oil palm germplasm into two major genetic clusters (K = 2). Cluster I mainly consisted of accessions sharing similar pedigree backgrounds and breeding history, whereas Cluster II comprised genetically distinct accessions originating from different breeding populations. The observed clustering reflects historical breeding practices and the use of a limited number of parental sources in developing commercial planting materials. Although comprehensive phenotypic information was not available for all accessions, the identified genetic groups provide useful guidance for selecting genetically divergent parents to maximize heterosis and broaden the breeding population.

K suggested that there are two subpopulations in the oil palm populations (AG001 and AG008), but Ong *et al.* (2015) discovered no strong subpopulation structure in these populations. Andhra Pradesh, Andaman and Nicobar Islands, Karnataka, Tamil Nadu, and Maharashtra were the sources of 78 native oil palm genotypes that made up Population 1 (G1). Of the 78, admixtures of 10% are present in eight genotypes. With the exception of IC0610041-253, all of the admixtures seen in G1 are Andaman. There were 72 oil palm genotypes in Population 2 (G2), and 7 of those genotypes had higher admixtures, ranging from 8 to 45%. The blend in G2 comes through the Andaman Islands. Ong *et al.* (2015) validated the findings, revealing that 7% of the variance occurred between populations.

The observed narrow genetic base in several accessions is consistent with the historical origin of commercial oil palm cultivation. Most cultivated germplasm worldwide traces its ancestry to a very limited number of founder palms introduced from West Africa and subsequently maintained in the Amsterdam Botanical Garden before their distribution to breeding programmes. This founder effect has contributed to reduced genetic variability in cultivated populations and is reflected in the clustering pattern observed in the present study. The introduction of genetically diverse germplasm from different geographical regions is therefore essential for broadening the breeding pool and improving long-term genetic gain.

Conclusion

The polymorphism levels of the mEgCIR0246, mEgCIR3358, mEgCIR0782, and mEgCIR0779 markers in the current study are significantly greater than those of the other 54 SSRs, which makes them useful for mapping and diversification studies in future breeding programs. Among them, mEgCIR0246 emerged as the most powerful marker due to its highest allele number, greatest genetic diversity, and lowest major allele frequency, highlighting its potential for future molecular breeding and conservation studies. The 150 oil palm genotypes were divided into two major clusters and two populations based on the findings for the UPGMA clusters and population structure investigations. We can infer from this result that the germplasm lines in our field gene bank come from two different sources. Fifteen genotypes, namely IC0610007-141, IC0610017-B16, IC0610010-171, IC0610026-114, IC0610012-202, IC0610010-178, IC0610010-172, IC0610041-253, IC0610020-B21, IC0610017-B13, IC0610020-B23, IC0610014-B28, IC0610014-B26, IC0610029-B45, and IC0610019-80, exhibited admixed ancestry, indicating the presence of alleles derived from both genetic clusters. These admixed genotypes represent valuable genetic resources and may serve as promising parental materials for broadening the genetic base and enhancing genetic gain in future oil palm breeding programmes. When used to evaluate the genetic diversity in oil palm, this SSR method has proven to be both accurate and dependable. Creating effective resource management plans will require an understanding of population structure and genetic variation.

Funding

The investigation described in this paper is an integral part of the Ph.D. thesis titled "Morphological and Molecular characterization of oil palm (*Elaeis guineensis* Jacq.) germplasm for productivity traits." No funding was provided during the study or when writing the manuscript.

Author contributions

H. P. Bhagya: Responsible for writing the original draft, software development, methodology, investigation, data curation, and conceptualization. B. Kalyana Babu: Provided supervision, conducted formal analysis, and reviewed and edited the work. The authors state that they have no known competing financial interests or personal relationships that could have influenced the research presented in this paper.

Acknowledgments

We are grateful to Dr. R K Mathur, the former director of ICAR-Indian Institute of Oil Palm Research, Pedavegi, India, and all other staff members, including Research Assistants, Senior Research Fellows, and technical personnel, who contributed to the collection, conservation, and cataloging of oil palm germplasm project at ICAR-Indian Institute of Oil Palm Research, Pedavegi.

References

Abimbola, L. M., Din, K. A., Nookiah, R., Hassan, M. S., Ngah, C. W. Z. C. W., & Sukaimi, J. (2016). Genetic variability for yield, yield components and fatty acid traits in oil palm (*Elaeis guineensis* Jacq.) germplasm using multivariate tools. *International Journal of Agriculture, Forestry and Plantation, 2*, 219-226.

Babu, B. K, Mary Rani K. L., Sahu, S., Mathur, R. K., Naveen Kumar, P., Ravichandran, G., Anitha, P., & Bhagya, H. P. (2019). Development and validation of whole genome-wide and genic microsatellite markers in oil palm (*Elaeis guineensis* Jacq.): First microsatellite database (OpSatdb). *Scientific reports, 9*, 1899. https://doi.org/10.1038/s41598-018-37737-7

Babu, B. K., Mathur, R. K., Ravichandran, G., Anitha, P., Bhagya, H. P., & Venu, M. V. B. (2023). A novel QTL linked toasparagine synthetase gene forstem height increment in oil palm (*E. guineensis* Jacq.) identifed and validated through integrated genomic approaches. *Euphytica, 219*, 73. https://doi.org/10.1007/s10681-023-03201-5

Babu, B. K., Dinesh, P., Agrawal, P. K., Sood, S., Chandrashekara, C., Bhatt, J. C., & Kumar, A. (2014). Comparative genomics and

association mapping approaches for blast resistant genes in finger millet using SSRs. *Plos One, 9*(6), e99182. https://doi.org/10.1371/journal.pone.0099182

Babu, B. K., Mathur, R. K., Venu, M. V. B., Sandip, S., Ravichandran, G., Anita, P., & Bhagya, H. P. (2021). Genome-wide association study (GWAS) of major QTLs for bunch and oil yield-related traits in *Elaeis guineensis* L. *Plant Science, 305*,110810. https://doi.org/10.1016/j.plantsci.2020.110810

Bakoumé, C., Wickneswari, R., Siju, S., Rajanaidu, N., Kushairi, A., & Billotte, N. (2015). Genetic diversity of the world's largest oil palm **(***Elaeis guineensis* Jacq**.)** field gene bank accessions using microsatellite markers. *Genetic Resources and Crop Evolution, 62*, 349-360. https://doi.org/10.1007/s10722-014-0156-8

Barcelos, E., Amblard, P., Berthaud, J., & Seguin, M. (2000). The genetic diversity of the American oil palm (*Elaeis oleifera* Kunth.) Corte's revealed by nuclear RFLP markers. In *International symposium on oil palm genetic resources and utilisation,* Kuala Lumpur**.** Proceedings Kuala Lumpur. *The Malaysian Palm Oil Board,* 173-192**.**

Bhagya, H. P., Babu B. K., Naika, M. B. N., Mathur, R. K., Gangadharappa, P. M., Satisha, D., & Naik, R. B. (2018). Identification and utilization of polymorphic SSR markers for genetic diversity studies in oil palm. *International Journal of Current Microbiology and Applied Sciences, 7*(4), 333-341. https://doi.org/10.20546/ijcmas.2018.704.038

Bhagya, H. P., Babu B. K., Gangadharappa P. M., Naika, M. B. N., Satish, D., & Mathur, R. K. (2020). Identification of QTLs in oil palm (*Elaeis guineensis*Jacq.) using SSR markers through association mapping. *Journal of Genetics, 99*, 19. https://doi.org/10.1007/s12041-020-1180-4

Bhagya, H. P., Babu, B. K., Mathur, R. K., Ramajayam, D., Ravichandran, G., & Anitha, P. (2024). Oil palm (*Elaeis guineensis* Jacq.) germplasm genome-wide association analysis for the oil yield traits utilising microsatellite markers. *Industrial Crops and Products, 218*, 118934. https://doi.org/10.1016/j.indcrop.2024.118934

Bhagya, H. P., Mathur, R. K., Ravichandran, G., Ramajayam, D., Kalyana Babu, B., Anitha, P., Murugesan, P., Sunil Kumar, K., Somasundaram, G., & Rahana, S. N. (2021). Identification and selection of elite oil palm (*Elaeis guineensis* Jacq.) genotypes for utilisation in a breeding program. *Journal of Plantation Crops, 49*(3), 162-167. https://doi.org/10.25081/jpc.2021.v49.i3.7449

Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics, 32*(3), 314-331.

Cochard, B., Adon, B., Rekima, S., Billotte, N, de Chenon, R D., Koutou, A., Nouy, B., Omore, A., Purba, A. R., Glaszmann, J.-C., & Noyer, J.-L. (2009). Geographic and genetic structure of African oil palm diversity suggests new approaches to breeding. *Tree Genetics & Genomes, 5*, 493-504. https://doi.org/10.1007/s11295-009-0203-3

Corley, R. H. V., & Tinker, P. B. (2003). *The Oil palm*. (4th ed.). Blackwell Science Ltd. https://doi.org/10.1002/9780470750971

Hardon, J. J. (1970). Inbreeding in populations of the oil palm (*Elaeis guineensis* Jacq.) and its effects on selection. *Oléagineux, 25*, 449-456.

Hayati, A., Wickneswari, R., Maizura, I., & Rajanaidu, N. (2004). Genetic diversity of oil palm (*Elaeis guineensis* Jacq.) germplasm collections from Africa: implications for improvement and conservation of genetic resources. *Theoretical and Applied Genetics, 108*, 1274-1284. https://doi.org/10.1007/s00122-003-1545-0

Heckenberger, M., Bohn, M, Ziegle, J. S., Joe, L. K., Hauser, J. D., Hutton, M., & Melchinger, A. E. (2002). Variation of DNA fingerprints among accessions within maize inbred lines and implications for identification of essentially derived varieties. *Molecular Breeding, 10***,** 181-191. https://doi.org/10.1023/A:1020539330957.

Henson, I. E., & Harun, M. H. (2005). The influence of climatic conditions on gas and energy exchanges above a young oil palm stand in north Kedah, Malaysia. *Journal of Oil Palm Research, 17*, 73-91.

Hill, W. G., & Mackay, T. F. C. (2004). D. S. Falconer and Introduction to Quantitative Genetics. *Genetics, 167*(4), 1529-1536. https://doi.org/10.1093/genetics/167.4.1529

Jalani, B. S. (2012). *Malaysian oil palm industry: contribution, challenges and future prospects.* Nilai, Negeri Sembilan, Malaysia, University Sains Islam Malaysia.

Lanes, É, C. M., Motoike, S. Y., Kuki, K. N., Nick, C., & Freitas, R. D. (2015). Molecular characterization and population structure of the Macaw Palm, *Acrocomia aculeata* (Arecaceae), *Ex Situ* germplasm collection using microsatellite markers. *Journal of Heredity, 106*(1), 102-112. https://doi.org/10.1093/jhered/esu073

Liu, K., & Muse, S. V. (2005). Power Marker**:** an integrated analysis environment for genetic marker analysis**.** *Bioinformatics, 21*, 2128-2129**.** https://doi.org/10.1093/bioinformatics/bti282

Maizura, I**.**, Rajanaidu, N**.**, Zakri, A. H., & Cheah, S. (2006). Assessment of genetic diversity in oil palm (*Elaeis guineensis* Jacq.) using Restriction Fragment Length Polymorphism (RFLP). *Genetic Resources and Crop Evolution*, 53, 187-195**.** https://doi.org/10.1007/s10722-004-4004-0

Murray, M. G., & Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research, 8*(19), 4321-4326. https://doi.org/10.1093/nar/8.19.4321

Murthy, B. R., & Arunachalam, V. (1966). The nature of genetic divergence in relation to breeding system in crop plants. *Indian Journal of Genetics and Plant Breeding, 26*, 188-198.

Odong, T. L., van Heerwaarden, J., Jansen, J., van Hintum, T. J. L., & van Eeuwijk, F. A. (2011). Determination of genetic structure of germplasm collections are traditional hierarchial clustering methods appropriate for molecular marker data. *Theoretical and Applied Genetics, 123*, 195-205. https://doi.org/10.1007/s00122-011-1576-x

Okoye, M. N., Uguru, M. I., Bakoume, C., Singh, R., & Okwuagwu, C. O. (2016). Assesment of genetic diversity of NIFOR oil palm main breeding parent genotypes using microsatellite markers**.** *American Journal of Plant Sciences, 7*(1), 218-237. https://doi.org/10.4236/ajps.2016.71022

Ong, P. W., Maizura, I., Abdullah, N. A. P., Rafii, M. Y., Ooi, L. C. L., Low, E. T. L. & Singh, R. (2015). Development of SNP markers and their application for genetic diversity analysis in the oil palm (*Elaeis guineensis*). *Genetics and Molecular Research, 14*(4), 12205-12216.

Rajanaidu, N. (1994). *PORIM oil palm genebank: collection, evaluation, utilization and conservation of oil palm genetic resources*. Palm Oil Research Institute of Malaysia. http://myagric.upm.edu.my/id/eprint/9403

Sapey, E., Adusei-Fosu, K., Darkwah, D. O., & Agyei-Dwarko, D. (2017). Multivariate analysis of bunch yield and vegetative traits of oil palm germplasm conserved at Oil Palm Research Institute (OPRI) -Ghana. *International Journal of Plant Breeding and Crop Science, 4*(2), 231-236.

Singh, R., Ong-Abdullah, M., Low, E.-T. L., Manaf, M. A. A., Rosli, R., Nookiah, R., Ooi, L. C.-L., Ooi, S.-E., Chan, K.-L., Halim, M. A., Aziz, N., Nagappan, J., Bacher, B., Lakey, N., Smith, S. W., He, D., Hogan, M., Budiman, M. A., Lee, E. K., Sambanthamurthi, R. (2013) Oil palm genome sequence reveals divergence of interfertile species in old and new

Worlds. *Nature,* 500, 335-339. https://doi.org/10.1038/nature12309

Singh, R., Zaki, N. M., Ting, N.-C., Rosli, R., Tan, S.-G., Low, E.-T. L., Ithnin, M., & Cheah, S.-C. (2008). Exploiting an oil palm EST database for the development of gene-derived SSR markers and their expoitation for assessment of genetic diversity. *Biologia, 63*, 227-235. https://doi.org/10.2478/s11756-008-0041-z

Ting, N.-C., Noorhariza, M. Z., Rozana, R., Low, E.-T. L., Ithnin, M., Cheah, S.-C., Tan, S.-G., & Singh, R. (2010). SSR mining in oil palm EST database: application in oil palm germplasm diversity studies. *Journal of Genetics, 89*, 135-145. https://doi.org/10.1007/s12041-010-0053-7

Ting, N.-C., Yaakub, Z., Kamaruddin, K., Mayes, S., Massawe, F., Sambanthamurthi, R., Jansen, J., Low, L. E. T., Ithnin, M., Kushairi, A., Arulandoo, X., Rosli, R., Chan, K.-L., Amiruddin, N., Sritharan, S., Lim, C. C., Nookiah, R., Amiruddin, M. D., & Singh, R. (2016). Fine-mapping and cross validation of QTLs linked to fatty acid composition in multiple independent interspecific crosses of oil palm. *BMC Genomics, 17*, 289. https://doi.org/10.1186/s12864-016-2607-4

Xu, W., John Martin, J. J., Li, X., Liu, X., Zhang, R., Hou, M., Cao, H., & Cheng, S. (2024). Unveiling the secrets of oil palm genetics: A look into omics research. *International Journal of Molecular Sciences, 25*(16), 8625. https://doi.org/10.3390/ijms25168625

Zaki, N. M., Singh, R., Rosli, R., & Ismail, I. (2012). *Elaeis oleifera* genomic-SSR markers: exploitation in oil palm germplasm diversity and cross-amplification in Arecaceae. *International Journal of Molecular Sciences, 13*(4), 4069-4088. https://doi.org/10.3390/ijms13044069

Zhou, L., Xiao, Y., Xia, W., & Yang, Y. (2015). Analysis of genetic diversity and population structure of oil palm (*Elaeis guineensis*) from China and Malaysia based on species-specific simple sequence repeat markers. *Genetics and Molecular Research, 14*(4), 16247-16254.