Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Comparative Chloroplast Genomics and Maternal Lineage Diversity in Oil Palm (Elaeis guineensis)

Ally Juma Mkude

mkude46@yahoo.com

Tanzanian Forestry Research Institute

Redi Aditama

IPB University

Sintho W. Ardie

IPB University

Willy B. Suwarno

IPB University

Azis Natawijaya

PT Bumitama Gunajaya Agro

Indri Gita Lestari

IPB University

Sudarsono Sudarsono

IPB University

Research Article

Keywords: Oil palm, chloroplast genome, plastome diversity, SSR variation, phylogenetics, maternal lineage

Posted Date: February 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8662850/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Oil palm (*Elaeis guineensis*) is the world's most important source of vegetable oil, yet comparative studies of its chloroplast genomes remain limited, particularly for African accessions. Understanding plastome diversity and maternal lineage variation is critical for tracing domestication history and developing molecular markers for breeding and conservation.

Results

We report the first complete chloroplast genomes from Tanzanian Dura and Pisifera, Cameroonian Dura, and the ancestral Deli Dura from Indonesia, analyzed alongside 44 additional plastomes. Comparative analyses revealed highly conserved genome organization and gene content, with subtle differences in IR–SC boundaries and intergenic spacers. Simple sequence repeat (SSR) profiling confirmed the dominance of A/T mononucleotide repeats but highlighted lineage-specific variation, including a markedly reduced genic SSR density in Pisifera-Tz, while total SSR counts remained stable across accessions (≈ 859–860; Supplementary Table S3). Phylogenetic reconstruction clustered Pisifera-Tz with Dura-Cmr, while Dura-Tz and Dura-BBG formed distinct clades, reflecting multiple maternal lineages. Haplotype network analysis identified 14 distinct haplotypes, revealing moderate but structured plastome diversity. Benchmarking of sequencing platforms demonstrated complementary strengths: MGI provided deep coverage, while PacBio improved assembly contiguity and resolution of repetitive regions.

Conclusion

This study expands plastome resources for *E. guineensis*, generating the first Tanzanian and Cameroonian chloroplast genomes and uncovering lineage-specific SSR variation. Integration of comparative genomics, phylogenetics, and haplotype analysis advances understanding of oil palm domestication and supports the development of plastome-based molecular markers for breeding, conservation, and genetic resource management.

Introduction

Oil palm (*Elaeis guineensis* Jacq.) is the most productive source of vegetable oil worldwide, contributing more than one-third of the global supply (1). Its high oil yield per unit area has made it central to food security, bioenergy, and industrial applications (2,3). However, the rapid expansion of oil palm cultivation has raised concerns about deforestation, biodiversity loss, and sustainability (4,5). In addition, the limited genetic base of current breeding populations constrains progress in developing resilience traits, yield stability, and climate adaptation (6,7).

The narrow genetic base of Southeast Asian oil palm populations largely stems from a small number of African introductions during the colonial era (7). This founder effect has restricted genetic variation in commercially important lines, particularly the Deli Dura population in Indonesia (8). In contrast, African wild populations harbor significant allelic richness and represent untapped reservoirs of genetic diversity (8,9). Expanding cytoplasmic and nuclear variation is therefore critical for sustainable cultivation and long-term breeding success (10).

Chloroplast genomes (plastomes) are powerful tools for studying plant diversity, evolution, and adaptation (11,12). Plastomes play essential roles in photosynthesis, energy metabolism, and biosynthesis of amino acids, fatty acids, and pigments. Their conserved quadripartite structure, uniparental inheritance, and relatively low mutation rate make them valuable for phylogenetics, species differentiation, and lineage tracing (13). In flowering plants, plastomes typically range from 120 to 160 kb and encode 110 to 130 genes involved in transcription, translation, and photosynthesis (14,15). Despite their conservation, single-nucleotide polymorphisms, indels, and structural variations in non-coding regions provide informative markers for evolutionary research, germplasm authentication, and lineage tracing (16,17).

Previous studies of oil palm plastomes revealed highly conserved gene content and organization, consistent with other members of the palm family (Arecaceae) (18). The first complete *E. guineensis* plastome established a baseline structure of ~ 156,973 bp (18), and subsequent work confirmed its stability across the species (19,20). Nevertheless, comparative analyses identified nucleotide variations concentrated in single-copy regions and intergenic spacers, offering opportunities to differentiate closely related lineages (21). Commercial planting materials are derived from hybridization between Dura (thick-shelled) and Pisifera (shell-less) types to produce the high-yielding Tenera hybrid (22). Tracing maternal lineages is therefore essential for germplasm authentication and understanding cytoplasmic genetic effects (23,24)

Given the narrow maternal base of commercially dominant Deli Dura lines (16), resolving plastome diversity across wild and cultivated accessions is critical for breeding programs (25,26). While nuclear genomic studies of E. guineensis and E. oleifera are extensive (27–30), detailed comparative analyses of intraspecific plastome diversity remain limited. Even small numbers of variable sites within plastomes can be powerful for differentiating lineages and mapping cytoplasmic diversity (31). Characterizing plastome variation across diverse African accessions is thus essential for conservation, breeding authentication, and marker development (32).

Tanzania and Cameroon represent centers of origin and diversity for oil palm, harboring ancient lineages and wild populations that could significantly enrich global breeding programs (9,32). In contrast, Indonesian Dura populations descend from a narrow genetic base introduced from Africa in the early 20th century (16). Comparative plastome analyses across African regions, therefore, offer unique insights into domestication history, adaptation mechanisms, and maternal lineage divergence (27)

Recent advances in sequencing technologies have transformed plastome research (33). Short-read platforms such as Illumina and MGI DNBSEQ provide cost-effective accuracy for sequencing and variant detection but struggle with repetitive regions and structural variations (34). Long-read platforms such as PacBio HiFi and Oxford Nanopore generate reads spanning thousands of bases, enabling resolution of repetitive elements and structural variants (35). Long reads improve assembly contiguity, facilitate haplotype phasing, and enable precise detection of complex rearrangements (35). Comparing short- and long-read assemblies provides critical validation for plastome reconstruction, particularly in complex plant lineages such as oil palm (34). However, systematic comparisons of sequencing platform performance in oil palm plastomes remain scarce.

This study addresses these gaps by sequencing and analyzing chloroplast genomes of Tanzanian Dura (Dura-Tz) and Pisifera (Pisifera-Tz) accessions. We integrate these with previously sequenced Cameroonian Dura (Dura-Cmr), the ancestral Deli Dura from Indonesia (Dura-BBG), 29 Indonesian breeding lines, and 15 plastome accessions from public databases. By benchmarking MGI and PacBio sequencing platforms, we highlight their complementary strengths for plastome assembly and variant detection. Collectively, this work provides new insights into plastome diversity in E. guineensis, advances understanding of domestication history, and supports the development of plastome-based molecular markers for breeding, conservation, and genetic resource management.

Methods

Oil palm chloroplast genome samples

Two chloroplast genomes were sequenced in this study, including one Tanzanian Dura type (Dura-Tz) and one Tanzanian Pisifera type (Pisifera-Tz). The selected plant materials was identified by Mr. Basil E. Kavishe, Research Assistant at Tanzania Agricultural Research Institute (TARI). No voucher specimen was deposited in publicly accessible herbarium. However, the plant materials were photographed during sample collection, and they are now monitored by TARI for reference and future studies. Young leaves were collected from the Kwitanga plantation, Kigoma, Tanzania. Samples were cleaned, cut into ~ 20 cm pieces, sterilized with 70% ethanol, air-dried, sealed in airtight plastic bags, and transported in a cold box to PT Genetika Science, Indonesia, for sequencing. Two representative wild oil palm samples, one Cameroonian Dura-type (Dura-Cmr) and one Deli Dura progenitor from Bogor Botanical Garden, Indonesia (Dura-BBG), were sequenced in a previous project at PMB Lab, IPB University (NCBI accession numbers are on Progress). Moreover, as many as twenty-nine (29) samples were also from previously sequenced chloroplast genomes and obtained from Dr. Azis Natawijaya (PT Bumitama Gunajaya Agro Company). Inquiries for these 29 sequences should be directly forwarded to Dr. Azis Natawijaya. Other cpDNAs were obtained from NCBI GenBank DNA Database (https://ncbi.nlm.nih.gov/). All chloroplast genome sequences generated in this study, including their associated genome annotations, are publicly available in the NCBI database under BioProject accession number **PRJNA1416287**.

Chloroplast genome sequencing and assembly

Short-read sequencing of Dura-Tz and Pisifera-Tz total genomes was performed using the MGI DNBSEQ-G400 platform (34) (https://mgi-tech.eu/sequencing-products/dnbseq-g400) (PE150) with MGIEasy FS DNA Library Prep Set. Long-read sequencing of Dura-Tz total genomes was conducted using the PacBio Vega platform with HiFi Plex Prep Kit 96 (PacBio, 103-122-800) (36) (https://www.pacb.com/wp-content/uploads/Insert-HiFi-plex-prep-kit-96).

Raw reads of Dura-Tz and Pisifera-Tz were processed with **Fastp v0.20.1** (37) to trim low-quality sequences and adaptor contamination. De novo assembly was performed using **NOVOPlasty v4.3.5** (38), and the assemblies were polished using **EnsemblPlants56** (https://plants.ensembl.org).

Sequencing depth and coverage of the Dura-Tz and Pisifera-Tz chloroplast genomes were evaluated by mapping raw reads back to the assembled plastomes. Depth profiles were generated across the entire genome length (0–156 kb) for both Dura-Tz and Pisifera-Tz accessions. Coverage plots were inspected to confirm uniformity across large single copy (LSC), small single copy (SSC), and inverted repeat (IR) regions.

Chloroplast genome annotation

Assembled genomes in FASTA format were annotated using **GeSeq** within CHLOROBOX (39) and the **Dual Organellar Genome Annotator (DOGMA)** with default parameters—annotation employed 95% sequence similarity thresholds for protein-coding genes, rRNAs, and tRNAs. **BLASTX** and **BLASTN** were used to identify coding genes and rRNAs, while **tRNAscan-SE v2.0.7** was used for tRNA annotation (40). Intron boundaries, start, and stop codons were manually curated by comparison with a combination of many related species of *Arecaceae* plastomes. Genome maps were visualized using **OGDRAW** (41).

Comparative analysis of Plastome structure and Synteny

Using assembled chloroplast genomes, the base composition and frequencies of the LSC, SSC, and IR regions were determined in **Geneious Prime v2024.0.7** (42). Two multiple progressive sequence alignments were conducted; the first one included the two new plastomes plus Dura-BBG and Dura-Cmr. Multiple sequence alignments of the plastomes were generated using **MAFFT v7.490** (43). IR boundary variation was visualized using **IRscope** (44). The second analysis was carried out using the two new plastomes plus Dura-BBG and Dura-Cmr and six Indonesian breeding lines (Dura-Id_574, Dura-Id_101128179, Pisifera-Id_221, Pisifera-Id_1169, Tenera-Id_1093, Tenera-Id_15730) and six plastomes that were available on NCBI (ON248756, NC_017602, OR125034, OR125040, OR125048, OR125054) in Mauve v2.4.0 (45).

Simple sequence repeats (SSRs)

Pairwise multiple alignments of cp genomes were performed using **MAFFT v7.490** (43) in Geneious Prime. Simple sequence repeats (SSRs) containing 1–6 nucleotides were identified using the **Phobos v3.3.12** (46). The following configuration was adopted to search for SSR motifs: SSR of one to six nucleotides long, with a minimum repeat number of 10, 5, and 4 units for mono-, di-, and trinucleotide SSRs, respectively, and three units for tetra-, penta-, and hexanucleotide SSRs.

Phylogenetic and haplotype analysis

Phylogenetic analysis included 48 E. guineensis cp genomes. Multiple sequence alignments were generated using **MAFFT v7.490** (43). Phylogenetic trees were constructed using the **Maximum Likelihood** (ML) method with the **Tamura-Nei (TN93/TNG93 + G + I)** genetic distance model, and **1000 bootstrap replications** (47,48). Haplotype networks were constructed from 48 E. guineensis cp genomes, including 19 PT Bumitama Gunajaya Agro accessions, 25 NCBI accessions, and 4 genomes sequenced in the PMB lab at IPB University. Alignments were performed in Geneious Prime, and haplotype networks were generated using **NETWORK software** (49).

Results

Chloroplast Genome Sequencing

Sequencing output is summarized in Table 1. For the Dura-Tz and Pisifera-Tz accessions sequenced with MGI, approximately 242 million reads were generated per sample, corresponding to ~ 36 Gb of total bases. In contrast, PacBio sequencing of Dura-Tz produced only 5.95 million reads but yielded a substantially higher total base output of 63.33 Gb (Table 1).

Table 1
Sequencing data summary for Tanzanian Dura (Dura-Tz) and Pisifera (Pisifera-Tz)

Sample name	Sequencing Platform	Total reads (M.)	Total bases (Gb)
Dura-Tz	MGI	241.78	36.27
	PacBio	5.95	63.33
Pisifera-Tz	MGI	242.00	36.30

These results highlight the complementary strengths of the two platforms. MGI provides deep coverage through massive read counts, ensuring reliable variant detection and uniform base representation. PacBio, despite generating fewer reads, delivers long sequences that enhance assembly contiguity and resolve repetitive regions more effectively. Together, the datasets demonstrate that combining short-read and long-read technologies offers both depth and structural accuracy for plastome reconstruction.

Sequencing Depth and Coverage

Robust sequencing coverage was achieved across the chloroplast genomes of both Dura-Tz and Pisifera-Tz accessions. The Dura-Tz assembly reached depths of up to ~ 3,500 reads (Fig. 1A), while Pisifera-Tz attained nearly 7,000 reads (Fig. 1B), reflecting its substantially higher sequencing throughput. Coverage profiles were generally uniform across the plastome, with expected peaks in the inverted repeat (IR) regions due to duplication and troughs in AT-rich or non-coding regions that are inherently more challenging to sequence.

The overall consistency of coverage confirms the completeness and reliability of both assemblies. Significantly, the greater depth observed in Pisifera-Tz enhances confidence in variant detection and structural resolution, thereby providing a stronger foundation for downstream comparative and phylogenetic analyses.

Chloroplast Genome Assembly and Structure

The complete chloroplast genomes of Dura-Tz (PX957236) and Pisifera-Tz (PX957237) were assembled into circular maps of 156,883 bp, each displaying the typical quadripartite organization comprising large single-copy (LSC), small single-copy (SSC), and inverted repeat (IR) regions (Fig. 2A, 2B). Comparative analysis with previously sequenced accessions (Dura-Cmr and Dura-BBG) revealed highly conserved structural and compositional features (Table 2). These findings are consistent with earlier oil palm plastome characterizations (23,50) and broader plastid genome studies in seed plants (11,13,14, 23).

Table 2
Summary of general features of 4 accessions of *Elaeis guineensis* chloroplast-genomes

Attribute	*Dura-Tz*	*Pisifera-Tz*	Dura-Cmr	Dura-BBG
Genome Size (bp)	156,983	156,983	156,982	156,990
GC content (%)	37.4	37.4	37.4	37.4
LSC size (bp)	85,194	85,194	85,193	85,198
SSC size (bp)	17,643	17,643	17,643	17,644
IRs size (bp)	27,073	27,073	27,073	27,074
LSC GC content (%)	35.5	35.5	35.5	35.5
SSC GC content (%)	31.0	30.9	31.0	31.0
IRs GC content (%)	42.5	42.5	42.5	42.5
Number of genes	115	115	115	115
Number of protein-coding genes	79	80	80	80
Number of tRNA	32	31	31	31
Number of rRNA	4	4	4	4

Chloroplast Genome Annotation and Gene Content

Annotation identified a consistent set of 115 genes across all accessions, comprising 79–80 protein-coding genes, 31–32 tRNAs, and 4 rRNAs. Dura-Tz contained 79 protein-coding genes and 32 tRNAs, compared with 80/31 in others (Table 2). This subtle difference may reflect lineage-specific variation in gene annotation or minor events of tRNA duplication or loss, possibly involving *ycf1* or related loci. Overall, gene composition remains highly stable, underscoring the structural conservation of E. guineensis plastomes. Such uniformity highlights the evolutionary stability of the chloroplast genome, while slight differences in gene counts may serve as useful markers for lineage differentiation and comparative genomics.

Functional categorization revealed a conserved organization across all four accessions, including genes encoding photosystem components, ribosomal proteins, ATP synthase, NADH dehydrogenase, Rubisco, RNA polymerase, and proteins involved in DNA replication and repair. Transfer RNA (tRNA) and ribosomal RNA (rRNA) genes were distributed throughout the plastome, ensuring its translational capacity. Hypothetical chloroplast reading frames (ycf) were also identified, consistent with their conserved presence in angiosperm plastomes. Gene transcription orientation was bidirectional, with genes on the outer ring transcribed clockwise and those on the inner ring transcribed counterclockwise. Core functional categories—photosynthesis, transcription, translation, and genome maintenance—were consistently represented, in line with previous annotations of oil palm plastomes (10,17,19).

Together, these findings confirm the high degree of conservation in gene content and organization among oil palm plastomes. The stability observed between Dura-Tz and Pisifera-Tz, along with other accessions, provides a reliable foundation for comparative and evolutionary analyses. At the same time, minor lineage-specific variations can serve as potential markers of genetic differentiation.

SSR Distribution and Motif Composition

Analysis of simple sequence repeats (SSRs) across the four accessions—Dura-BBG, Dura-CMR, Pisifera-Tz, and Dura-Tz—revealed a highly conserved repeat landscape (Fig. 3, Supplementary Table 1). Total SSR counts ranged narrowly from 859 to 860 per plastome, indicating strong structural stability across lineages.

Mononucleotide repeats dominated the SSR profile, averaging 338.5 per accession, with a pronounced bias toward A/T motifs. Each accession contained approximately 317 A/T mononucleotide repeats, accounting for over 93% of the total, while G/C repeats were rare (22 per plastome). This A/T enrichment reflects the low GC content and mutational bias typical of plastid genomes (10,17)

Dinucleotide repeats were equally conserved, with 58 per accession. AT and TA motifs were the most abundant, comprising 144 of the 232 total dinucleotide repeats (62%). Other motifs such as AG, GA, TC, and CT were evenly distributed across accessions, with no lineage-specific enrichment. Trinucleotide and tetranucleotide motifs also showed uniform distribution, with 71 and 95 repeats per accession, respectively. Trinucleotide motifs were dominated by A/T-rich sequences such as TAT, AAG, and AAC, while tetranucleotide repeats included AAAT, AATA, ATTA, and TTTA. These patterns reinforce the A/T bias observed in shorter motifs and suggest that even longer SSRs retain compositional consistency across accessions.

Pentanucleotide repeats were similarly conserved, with 134 motifs per accession. Dominant sequences included AAAAT, AATTT, TTTTA, and TTAAA, all of which are rich in A/T bases. The recurrence of these motifs across all accessions further supports the plastome's mutational bias and structural stability. No accession-specific expansion or contraction of pentanucleotide motifs was observed.

Hexanucleotide repeats showed slightly more diversity, with counts ranging from 157 to 158 per accession. Despite the broader sequence range, dominant motifs remained A/T-rich, including TTCTTT, TTTTAT, and TTTTTT. The consistent counts and motif composition across accessions suggest that even the longest SSRs are evolutionarily stable and not subject to lineage-specific expansion.

Hexanucleotide repeats showed minor variation (157–158), likely reflecting annotation differences or sequencing noise rather than true biological divergence. Although Pisifera-Tz exhibited a marginally higher total SSR count (860 vs. 859 in other accessions), its genic SSR density was substantially reduced (277 vs. ~681–682). This indicates that while overall SSR abundance remains stable, Pisifera-Tz plastomes have undergone lineage-specific contraction in genic regions (see Supplementary Table S3)."

Overall, the uniformity of SSR types and motif composition across all accessions underscores the evolutionary stability of the oil palm plastome. While SSR variability is limited, conserved tetranucleotide, pentanucleotide, and hexanucleotide motifs—particularly those rich in A/T bases—may still serve as neutral molecular markers for population genetics, germplasm authentication, and comparative genomics. This conserved SSR profile highlights plastome stability and mirrors SSR distributions reported in other palm species (8,51,52). SSR marker development in oil palm has similarly emphasized A/T-rich motifs (53,54).

Genic SSRs Across Accessions

Comparative analysis of genic SSRs revealed both broad conservation and striking lineage-specific differences among the four E. guineensis chloroplast genomes (Table 3). Dura-Cmr and Dura-BBG each contained 682 genic SSRs, while Dura-Tz harbored 681, reflecting near-identical SSR abundance and distribution across genes associated with photosynthesis, replication, and genome maintenance. This consistency underscores the evolutionary stability of genic SSRs in most oil palm accessions, with plastome function preserved through balanced representation across essential pathways.

Table 3
Summary of genic SSRs in four *Elaeis guineensis* chloroplast genomes

Genome	Total genic SSRs	Gene categories with SSRs
Dura-Cmr/Dura-BBG/Dura-Tz	681–682	**Photosynthesis**: *atp* (B/E/F/I), *ccsA*, *cemA*, *ndH* (A/B/D/E/H/I/K), *paf* (I/II), *pet* (A/B/D), *psa* (A/B/J), psb (A/B/C/F/I/K/L/M/T/Z), *rbcL.* **Maintenance**: *accD*, *clpP1*, *matK*, *rpl* (2/16/22/32/33), *rps* (3/4/7/11/12/14/15/16/18/19), *ycf* (1/2) **Replication**: *rpo* A/B/C1/C2
Pisifera-Tz	277	**Photosynthesis**: *atp* (B/E/F/I), *cemA*, *ndhK*, *paf* I/II, *pet* A/B/D, *psa* A/B/I/J, *psb* (A/B/C/F/I/K/L/M/T/Z), *rbcL*. **Maintenance**: *accD*, *clpP1*, *matK*, *rpl* (16/22/33), *rps* (3/4/11/12/14/16/18) **Replication**: *rpo* A/B/C1/C2

In contrast, Pisifera-Tz exhibited a markedly reduced genic SSR count (277), representing a substantial contraction relative to the other accessions. Despite this reduction, SSRs in Pisifera-Tz were still distributed across key functional genes, including rbcL, matK, psa, psb, and rpo, ensuring that core

photosynthetic and transcriptional processes remained represented. However, the absence of SSRs in several maintenance-related loci—including ycf1, ycf2, and members of the ndh family—suggests lineage-specific loss or reduced SSR density in regions typically associated with genome stability and stress response.

This pattern highlights two complementary insights: first, the conserved SSR profiles in Dura accessions reflect plastome resilience and functional redundancy across diverse lineages; second, the reduced SSR density in Pisifera-Tz may indicate either structural streamlining or selective pressures that minimized repetitive elements in its plastome. From a practical perspective, these differences provide opportunities for marker development: conserved genic SSRs in Dura accessions can serve as stable lineage markers. At the same time, the unique contraction in Pisifera-Tz offers diagnostic motifs for distinguishing maternal lineages in breeding programs. Despite this contraction, SSRs remained present in key functional genes, including *rbcL*, *matK*, *psa*, *psb*, and *rpo*. The absence of SSRs in *ycf1*, *ycf2*, and *ndh* family genes suggests lineage-specific streamlining, consistent with plastome variation observed in Pisifera lineages (55,56)

Plastome Structure and Synteny

Multiple genome alignment of 16 Elaeis guineensis chloroplast genomes—representing Dura, Tenera, and Pisifera—revealed a high degree of structural conservation across accessions (Fig. 4). Using the Mauve algorithm, plastome sizes ranged narrowly from 153,000 to 158,000 bp, with most accessions exhibiting uniform syntenic block arrangements and stable genome architecture. This consistency reflects strong evolutionary constraints on plastome organization, likely driven by the essential roles of chloroplast genes in photosynthesis and cellular metabolism.

Notably, two Indonesian Pisifera accessions (Pisifera-Id_221 and Pisifera-Id_101128179) displayed structural rearrangements in the 140–145 kb region, characterized by translocations and/or inversions (highlighted in red blocks). These localized disruptions suggest lineage-specific plastome remodeling, potentially arising from recombination events or selective pressures unique to the Pisifera background. While the overall plastome structure remains conserved, such rearrangements may influence gene regulation or plastome stability, warranting further investigation into their functional consequences.

Complementary mapping of the 16 plastomes confirmed the conservation of junction architecture between the large single copy (LSC), small single copy (SSC), and inverted repeat (IR) regions (Fig. 5). All accessions shared nearly identical genome sizes (~ 156.98 kb) and consistent placement of key genes at junctions, including rps19, rpl2, psbA, ndhF, ycf1, and trnN. Minor differences in junction lengths (25–65 bp) were observed but did not affect gene orientation or overall structural layout.

These findings reinforce the evolutionary stability of plastome boundaries in E. guineensis and validate their utility for comparative and phylogenetic analyses. The conserved junctions and syntenic blocks across most accessions provide a robust framework for plastome-based lineage tracing. In contrast, the structural variation in select Pisifera accessions offers a unique opportunity to explore plastome

dynamics and their potential role in fruit form differentiation or adaptation. Such rearrangements have been linked to recombination events in plastomes of palms and other monocots (20,24,45,57) Comparative mapping of LSC, SSC, and IR junctions confirmed consistent placement of key genes (*rps19*, *rpl2*, *psbA*, *ndhF*, *ycf1*, *trnN*) and minor junction length variation (25–65 bp), reinforcing plastome boundary stability [54](44).

Phylogenetic Relationships

Phylogenetic analysis based on complete chloroplast genome sequences revealed distinct clustering patterns among the four E. guineensis accessions (Fig. 6). Pisifera-Tz and Dura-Cmr formed a closely related clade, suggesting high genetic similarity and a likely shared maternal lineage. In contrast, Dura-Tz and Dura-BBG grouped separately, indicating moderate divergence that may reflect differences in evolutionary origin, geographic adaptation, or breeding history. Despite these variations, all accessions clustered within a unified species group, confirming their overall genomic relatedness and supporting their classification within *E. guineensis*. This pattern underscores the utility of chloroplast genomes for resolving maternal lineages while reinforcing the evolutionary cohesion of the species. These results align with previous phylogenetic studies that resolved maternal lineages in oil palm using plastome data (12,17,18).

Haplotype Diversity

Median-joining haplotype network analysis identified 14 distinct plastome haplotypes (HAP1–HAP14) among 48 accessions (Supplementary table 2). The network displayed short mutational distances and interconnected haplotypes, reflecting low divergence and shared evolutionary origin. Median vectors (mv1–mv5) represented inferred ancestral or intermediate haplotypes, providing insight into mutational pathways and lineage continuity. This moderate but structured plastome diversity supports the evolutionary stability of the plastome and offers practical utility for germplasm management and conservation of cytoplasmic diversity (Fig. 7). The presence of median vectors (mv1–mv5) suggests inferred ancestral haplotypes, consistent with haplotype diversity patterns reported in oil palm germplasm collections (9,28,49,51,52)

Discussion

The assemblies confirm the evolutionary stability of oil palm plastomes, with conserved genome sizes and GC content across accessions. This mirrors findings from broader palm plastome studies (20,24,58) and emphasizes strong selective constraints on chloroplast architecture (10,11,59). Annotation results reinforce this resilience, with highly conserved gene repertoires across accessions, consistent with previous genome sequencing efforts (10,17,60)

SSR analysis revealed conserved motif landscapes across accessions, dominated by A/T mononucleotide repeats. This pattern reflects mutational bias typical of plastid genomes (14,15) and parallels SSR distributions reported in oil palm (53,54). Genic SSR analysis highlighted Pisifera-Tz as an

outlier, with reduced SSR density and absence in *ycf* and *ndh* loci, paralleling reports of cytoplasmic variation in oil palm (61,62).

Comparative synteny analysis confirmed conserved plastome architecture across most accessions, with structural rearrangements observed only in Indonesian Pisifera genomes. These rearrangements are consistent with recombination-driven plastome variation in palms (20,45,57). Phylogenetic analysis resolved maternal lineages, clustering Pisifera-Tz with Dura-Cmr, while haplotype network analysis (Fig. 7) revealed moderate diversity but strong connectivity, consistent with maternal inheritance and evolutionary stability (12,17,49).

Together, these findings highlight a dual narrative: deep conservation of plastome structure and gene content across E. guineensis, coupled with subtle lineage-specific variation in SSR density, Synteny, and haplotype distribution. This balance of stability and divergence provides a robust framework for phylogenetic inference, germplasm authentication, and breeding strategies aimed at conserving cytoplasmic diversity in oil palm (8,28,52,63).

Conclusions

Comparative analysis of four *Elaeis guineensis* chloroplast genomes revealed a highly conserved plastome structure, with only minor variation in junction lengths and genome sizes. Despite this overall stability, Pisifera-Tz exhibited a markedly reduced genic SSR abundance, highlighting its distinct genomic profile. Phylogenetic and haplotype analyses further demonstrated close evolutionary relationships among accessions, while uncovering moderate chloroplast valuable diversity for lineage tracing, population genetics, and applied breeding. Together, these findings provide valuable genomic resources that strengthen evolutionary understanding and support future conservation and crop improvement strategies in oil palm.

Declarations

Authors' contributions:

AJM and SS: Research conceptualization, implementation, data analysis, original manuscript preparation, and reviewing.

RA: Data analysis and manuscript data preparation

SWA and WBS: Supervising and reviewing the manuscript.

AN and GIL: Providing the chloroplast genome of 19 oil palm germplasm accessions from Indonesia and reviewing the manuscript.

Data availability

All data supporting the findings of this study are available within the paper and its Supplementary Information. All chloroplast genome sequences generated in this study, including their associated genome annotations, are publicly available in the NCBI database under BioProject accession number **PRJNA1416287**.

Ethics, Consent to Participate, and Consent to Publish declarations

Not applicable.

Acknowledgments

We appreciate the management of Bogor Botanical Gardens and the National Research and Innovation Agency (BRIN) for giving access to the Dura-BBG sample. This manuscript is part of the AJM's master's (MSc) thesis. We acknowledge the support of Revocatus Petro Mushumbusi, PhD, the Director of Tanzania Forestry Research Institute (TAFORI), and his support for AJM's graduate program at IPB University, Bogor, Indonesia. AI was used to polish the English language quality in the manuscript preparation.

Funding:

- **Ally Juma Mkude, Rare Forest Profession, Tanzania Forest Fund**
- **Redi Aditama, The Oil Palm Research Grant, the Indonesian Palm Oil Fund Management Agency (BPDPKS), GRS_20220227110350**
- **Sudarsono Sudarsono, The Oil Palm Research Grant, the Indonesian Palm Oil Fund Management Agency (BPDPKS), GRS_20220227110350**

References

1. Qaim M, Sibhatu KT, Siregar H, Grass I. Environmental, Economic, and Social Consequences of the Oil Palm Boom. Annu Rev Resour Econ [Internet]. 2020 Oct 6;12(1):321–44. Available from: https://www.annualreviews.org/doi/10.1146/annurev-resource-110119-024922
2. Meijaard E, Brooks TM, Carlson KM, Slade EM, Garcia-Ulloa J, Gaveau DLA, et al. The environmental impacts of palm oil in context. Nat Plants [Internet]. 2020 Dec 7;6(12):1418–26. Available from: https://www.nature.com/articles/s41477-020-00813-w
3. Murphy DJ, Goggin K, Paterson RRM. Oil palm in the 2020s and beyond: challenges and solutions. CABI Agric Biosci [Internet]. 2021 Oct 11;2(1):39. Available from: https://www.cabidigitallibrary.org/doi/10.1186/s43170-021-00058-3
4. Meijaard E, Sheil D. The Moral Minefield of Ethical Oil Palm and Sustainable Development. Front For Glob Chang [Internet]. 2019 May 28;2. Available from: https://www.frontiersin.org/article/10.3389/ffgc.2019.00022/full

5. Vijay V, Pimm SL, Jenkins CN, Smith SJ. The Impacts of Oil Palm on Recent Deforestation and Biodiversity Loss. Anand M, editor. PLoS One [Internet]. 2016 Jul 27;11(7):e0159668. Available from: https://dx.plos.org/10.1371/journal.pone.0159668
6. Soh AC, Mayes S, Roberts J. Introduction to the Oil Palm Crop. In: Oil Palm Breeding [Internet]. Boca Raton : Taylor & Francis, 2017.: CRC Press; 2017. p. 1–6. Available from: https://www.taylorfrancis.com/books/9781498715454/chapters/10.1201/9781315119724-1
7. Corley RHV, Tinker PB. The Oil Palm. Fifth. West Sussex: John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ , UK; 2016. 1–627 p.
8. Magaña-Álvarez A, Pérez-Brito D, Cortés-Velázquez A, Nexticapan-Garcéz Á, Ortega-Ramírez ME, García-Cámara I, et al. Genetic Variability of Oil Palm in Mexico: An Assessment Based on Microsatellite Markers. Agric. 2023 Sep 1;13(9).
9. Hayati A, Wickneswari R, Maizura I, Rajanaidu N. Genetic diversity of oil palm (Elaeis guineensis Jacq.) germplasm collections from Africa: implications for improvement and conservation of genetic resources. Theor Appl Genet [Internet]. 2004 May 16;108(7):1274–84. Available from: http://link.springer.com/10.1007/s00122-003-1545-0
10. Daniell H, Lin C sea, Yu M, Chang W jung. Chloroplast genomes : diversity , evolution , and applications in genetic engineering. Genome Biol. 2016;1–29.
11. Wicke S, Schneeweiss GM, DePamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol [Internet]. 2011 Jul 22;76(3–5):273–97. Available from: http://link.springer.com/10.1007/s11103-011-9762-4
12. Gitzendanner MA, Soltis PS, Wong GK -S., Ruhfel BR, Soltis DE. Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. Am J Bot [Internet]. 2018 Mar 30;105(3):291–301. Available from: https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.1048
13. Jansen RK, Ruhlman TA. Plastid Genomes of Seed Plants. In 2012. p. 103–26. Available from: http://link.springer.com/10.1007/978-94-007-2920-9_5
14. Palmer JD. COMPARATIVE ORGANIZATION OF CHLOROPLAST GENOMES. Annu Rev Genet [Internet]. 1985 Dec;19(1):325–54. Available from: https://www.annualreviews.org/doi/10.1146/annurev.ge.19.120185.001545
15. Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, et al. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J [Internet]. 1986 Sep;5(9):2043–9. Available from: https://link.springer.com/article/10.1002/j.1460-2075.1986.tb04464.x
16. Uthaipaisanwong P, Chanprasert J, Shearman JR, Sangsrakru D, Yoocha T, Jomchai N. Characterization of the chloroplast genome sequence of oil palm (Elaeis guineensis Jacq .). 2012;500:172–80.
17. Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, et al. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am J Bot [Internet].

2005 Jan;92(1):142–66. Available from: https://bsapubs.onlinelibrary.wiley.com/doi/10.3732/ajb.92.1.142

18. Dong W, Liu J, Yu J, Wang L, Zhou S. Highly Variable Chloroplast Markers for Evaluating Plant Phylogeny at Low Taxonomic Levels and for DNA Barcoding. Moustafa A, editor. PLoS One [Internet]. 2012 Apr 12;7(4):e35071. Available from: https://dx.plos.org/10.1371/journal.pone.0035071
19. Poczai P, Varga I, Laos M, Cseh A, Bell N, Valkonen JP, et al. Advances in plant gene-targeted and functional markers: a review. Plant Methods [Internet]. 2013;9(1):6. Available from: http://plantmethods.biomedcentral.com/articles/10.1186/1746-4811-9-6
20. Rahmawati A, Volkaert HA, Dinarti D, Maskromo I, Hatta ANNL, Sudarsono S. Complete Chloroplast Genome Sequences of Coconut cv. Kopyor Green Dwarf and Comparative Genome Analysis to Oil Palm, Date Palm, Sago Palm, and Miniature Sugar Palm. In: Oil Crop Genomics [Internet]. Cham: Springer International Publishing; 2021. p. 189–216. Available from: https://link.springer.com/10.1007/978-3-030-70420-9_10
21. Arshad A, Aditama R, Rahayu MSRI, Natawijaya A. Comparative study of chloroplast genomes across seven Salacca species. 2024;25(11):4043–58.
22. Putu NI, Asih SRI, Pratiwi RA, Prihatini I, Kuswandi PC, Yulianti EVY, et al. Comparative analysis of the complete chloroplast genome of the rare endemic palm Pinanga arinasae Witono from Bali , Indonesia. 2025;49(5):432–46.
23. Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. Vol. 17, Genome Biology. 2016. p. 1–29.
24. Khan AL, Asaf S, Lee IJ, Al-Harrasi A, Al-Rawahi A. First chloroplast genomics study of Phoenix dactylifera (var. Naghal and Khanezi): A comparative analysis. Chen S, editor. PLoS One [Internet]. 2018 Jul 31;13(7):e0200104. Available from: https://dx.plos.org/10.1371/journal.pone.0200104
25. Chan KL, Tatarinova T V., Rosli R, Amiruddin N, Azizi N, Halim MAA, et al. Evidence-based gene models for structural and functional annotations of the oil palm genome. Biol Direct [Internet]. 2017 Dec 8;12(1):21. Available from: https://biologydirect.biomedcentral.com/articles/10.1186/s13062-017-0191-4
26. Mohd Talkah NS, Aziz NAKA, Rahim MFA, Hanafi NFF, Ahmad Mokhtar MA, Othman AS. The chloroplast genome inheritance pattern of the Deli-Nigerian prospection material (NPM) × Yangambi population of Elaeis guineensis Jacq. PeerJ [Internet]. 2024 May 27;12:e17335. Available from: https://peerj.com/articles/17335
27. Malike FA, Marhalil M, Zulkifli Y, Aziz NA, Oyebamiji YO. Genetic Variability of Oil Palm (Elaeis guineensis Jacq .) Deli Dura Inter-Crosses Population in The MPOB Research Station Hulu Paka ,. 2024;53:125–33.
28. Gan ST, Teo CJ, Manirasa S, Wong WC, Wong CK. Assessment of genetic diversity and population structure of oil palm (Elaeis guineensis Jacq.) field genebank: A step towards molecular-assisted

germplasm conservation. Avramidou E V., editor. PLoS One [Internet]. 2021 Jul 29;16(7):e0255418. Available from: https://dx.plos.org/10.1371/journal.pone.0255418

29. Zaki NM, Singh R, Rosli R, Ismail I. Elaeis oleifera Genomic-SSR Markers : Exploitation in Oil Palm Germplasm Diversity and Cross-Amplification in Arecaceae. 2012;4069–88.
30. Osorio-Guarín JA, Garzón-Martínez GA, Delgadillo-Duran P, Bastidas S, Moreno LP, Enciso-Rodríguez FE, et al. Genome-wide association study (GWAS) for morphological and yield-related traits in an oil palm hybrid (Elaeis oleifera x Elaeis guineensis) population. BMC Plant Biol [Internet]. 2019 Dec 3;19(1):533. Available from: https://bmcplantbiol.biomedcentral.com/articles/10.1186/s12870-019-2153-8
31. Corley RHV, Tinker PB. The Oil Palm [Internet]. Wiley; 2015. Available from: https://onlinelibrary.wiley.com/doi/book/10.1002/9781118953297
32. Chung KP. Cytoplasmic inheritance: The transmission of plastid and mitochondrial genomes across cells and generations. Plant Physiol [Internet]. 2025 Apr 30;198(1). Available from: https://academic.oup.com/plphys/article/doi/10.1093/plphys/kiaf168/8123037
33. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet [Internet]. 2016 Jun 17;17(6):333–51. Available from: https://www.nature.com/articles/nrg.2016.49
34. Sun X, Hu Y, Song Z. Library construction of metabarcoding at DNBSEQ-G400 with MGIEasy universal DNA library prep Kit v1 [Internet]. 2020. Available from: https://www.protocols.io/view/library-construction-of-metabarcoding-at-dnbseq-g4-bn95mh86
35. Begum G, Albanna A, Bankapur A, Nassir N, Tambi R, Berdiev BK, et al. Long-Read Sequencing Improves the Detection of Structural Variations Impacting Complex Non-Coding Elements of the Genome. Int J Mol Sci [Internet]. 2021 Feb 19;22(4):2060. Available from: https://www.mdpi.com/1422-0067/22/4/2060
36. Bentz ML, Clines C, Padilla J, Horn C, Lee J, Sheth M. High-throughput PacBio library preparation and sequencing techniques for genomic DNA and TNA. Front Genet [Internet]. 2025 Aug 6;16. Available from: https://www.frontiersin.org/articles/10.3389/fgene.2025.1587691/full
37. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics [Internet]. 2018 Sep 1;34(17):i884–90. Available from: https://academic.oup.com/bioinformatics/article/34/17/i884/5093234
38. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res [Internet]. 2016 Oct 24;gkw955. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw955
39. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq – versatile and accurate annotation of organelle genomes. Nucleic Acids Res [Internet]. 2017 Jul 3;45(W1):W6–11. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkx391
40. Chan PP, Lin BY, Mak AJ, Lowe TM. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. Nucleic Acids Res [Internet]. 2021 Sep 20;49(16):9077–96.

Available from: https://academic.oup.com/nar/article/49/16/9077/6355886

41. Lohse M, Drechsel O, Kahlau S, Bock R. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res [Internet]. 2013 Jul 1;41(W1):W575–81. Available from: http://academic.oup.com/nar/article/41/W1/W575/1092342/OrganellarGenomeDRAWa-suite-of-tools-for
42. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012 Jun;28(12):1647–9.
43. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013 Apr;30(4):772–80.
44. Amiryousefi A, Hyvönen J, Poczai P. IRscope: an online program to visualize the junction sites of chloroplast genomes. Hancock J, editor. Bioinformatics [Internet]. 2018 Sep 1;34(17):3030–1. Available from: https://academic.oup.com/bioinformatics/article/34/17/3030/4961430
45. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. Genome Res [Internet]. 2004 Jul;14(7):1394–403. Available from: http://genome.cshlp.org/lookup/doi/10.1101/gr.2289704
46. Mayer C, Leese F, Tollrian R. Genome-wide analysis of tandem repeats in Daphnia pulex - a comparative approach. 2010;
47. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol [Internet]. 1987 Jul;4(4):406–25. Available from: https://academic.oup.com/mbe/article/4/4/406/1029664/The-neighborjoining-method-a-new-method-for
48. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Battistuzzi FU, editor. Mol Biol Evol [Internet]. 2021 Jun 25;38(7):3022–7. Available from: https://academic.oup.com/mbe/article/38/7/3022/6248099
49. Bandelt HJ, Forster P, Rohl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol [Internet]. 1999 Jan 1;16(1):37–48. Available from: https://academic.oup.com/mbe/article-lookup/doi/10.1093/oxfordjournals.molbev.a026036
50. Li DM, Pan YG, Wu XY, Zou SP, Wang L, Zhu GF. Comparative chloroplast genomics , phylogenetic relationships and molecular markers development of Aglaonema commutatum and seven green cultivars of Aglaonema. Sci Rep [Internet]. 2024;1–17. Available from: https://doi.org/10.1038/s41598-024-62586-y
51. Bakoumé C, Wickneswari R, Rajanaidu N, Kushairi A, Amblard P, Billotte N. Allelic diversity of natural oil palm (Elaeis guineensis Jacq.) populations detected by microsatellite markers: implications for conservation. Plant Genet Resour [Internet]. 2007 Aug 1;5(2):104–7. Available from: https://www.cambridge.org/core/product/identifier/S1479262107710870/type/journal_article

52. Bakoumé C, Wickneswari R, Siju S, Rajanaidu N, Kushairi A, Billotte N. Genetic diversity of the world's largest oil palm (Elaeis guineensis Jacq.) field genebank accessions using microsatellite markers. Genet Resour Crop Evol [Internet]. 2015 Mar 3;62(3):349–60. Available from: http://link.springer.com/10.1007/s10722-014-0156-8
53. B. KB, K. L. MR, Sahu S, Mathur RK, P. NK, G. R, et al. Development and validation of whole genome-wide and genic microsatellite markers in oil palm (Elaeis guineensis Jacq.): First microsatellite database (OpSatdb). Sci Rep [Internet]. 2019 Feb 13;9(1):1899. Available from: https://www.nature.com/articles/s41598-018-37737-7
54. Zhou L, Yarra R, Zhao Z, Jin L, Cao H. Development of SSR markers based on transcriptome data and association mapping analysis for fruit shell thickness associated traits in oil palm (Elaeis guineensis Jacq.). 3 Biotech [Internet]. 2020 Jun;10(6):280. Available from: https://link.springer.com/10.1007/s13205-020-02269-3
55. Wu H, Li DZ, Ma PF. Unprecedented variation pattern of plastid genomes and the potential role in adaptive evolution in Poales. BMC Biol [Internet]. 2024 Apr 29;22(1):97. Available from: https://bmcbiol.biomedcentral.com/articles/10.1186/s12915-024-01890-5
56. Tu XD, Lin WJ, Fu HH, Lin YZ, Shen J, Chen S, et al. Comparative Analysis of Six Complete Plastomes of Tripterospermum spp. Int J Mol Sci [Internet]. 2024 Feb 22;25(5):2534. Available from: https://www.mdpi.com/1422-0067/25/5/2534
57. SRI ASIH NP, PRATIWI RA, PRIHATINI I, KUSWANDI PC, YULIANTI E, PRIYADI A. Comparative analysis of the complete chloroplast genome of the rare endemic palm Pinanga arinasae Witono from Bali, Indonesia. Turk J Botany [Internet]. 2025 Sep 19;49(5):432–46. Available from: https://journals.tubitak.gov.tr/botany/vol49/iss5/7
58. Id F, Cauz-santos LA, Morales JA, Alencar D, Picanc D, Teresa M, et al. Complete chloroplast genomes and phylogeny in three Euterpe palms (E . edulis , E . oleracea and E . precatoria) from different Brazilian biomes. 2022;1–26.
59. Wang J, Kan S, Liao X, Zhou J, Tembrock LR, Daniell H, et al. Plant organellar genomes: much done, much more to do. Trends Plant Sci [Internet]. 2024 Jul;29(7):754–69. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1360138523004120
60. Singh R, Low ETL, Ooi LCL, Ong-Abdullah M, Ting NC, Nagappan J, et al. The oil palm SHELL gene controls oil yield and encodes a homologue of SEEDSTICK. Nature [Internet]. 2013 Aug 24;500(7462):340–4. Available from: https://www.nature.com/articles/nature12356
61. Mohd Talkah NS, Aziz NAKA, Rahim MFA, Hanafi NFF, Ahmad Mokhtar MA, Othman AS. The chloroplast genome inheritance pattern of the Deli-Nigerian prospection material (NPM) × Yangambi population of Elaeis guineensis Jacq. PeerJ. 2024 May;12:e17335.
62. Singh R, Ong-Abdullah M, Low ETL, Manaf MAA, Rosli R, Nookiah R, et al. Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. Nature [Internet]. 2013 Aug 24;500(7462):335–9. Available from: https://www.nature.com/articles/nature12309

63. Fadila AM, Marhalil M, Zulkifli Y, Shamsudin NAA, Oyebamiji YO. Genetic Variability of Oil Palm (Elaeis guineensis Jacq.) Deli Dura Inter-Crosses Population in The MPOB Research Station Hulu Paka, Terengganu, Malaysia. Malaysian Appl Biol [Internet]. 2024 Jun 30;53(2):125–33. Available from: https://jms.mabjournal.com/index.php/mab/article/view/2975

Figures

Figure 1.

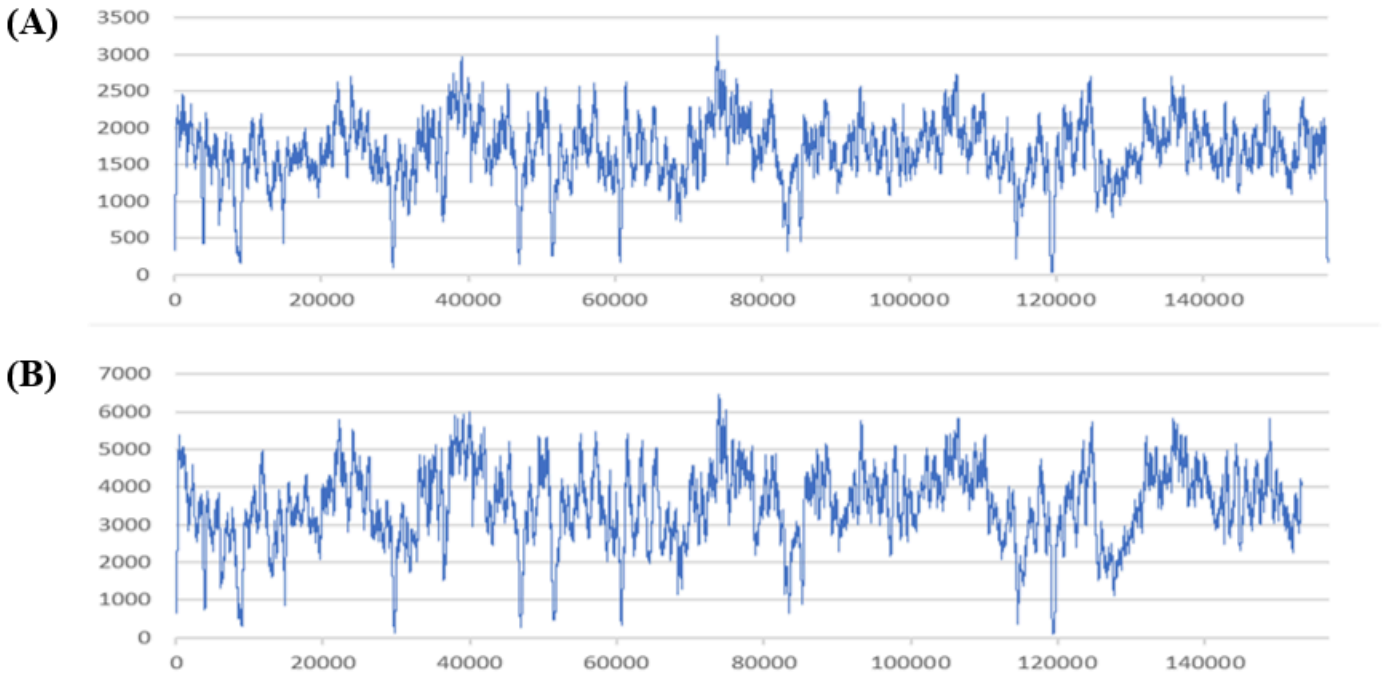


Figure 1

Sequencing depth and coverage profiles of Tanzanian Dura (Dura-Tz) and Pisifera (Pisifera-Tz) plastomes. (A) Dura-Tz assembly showing coverage depths up to ~3,500 reads. (B) Pisifera-Tz assembly reaching ~7,000 reads, reflecting higher sequencing throughput. Coverage is generally uniform, with peaks in inverted repeat (IR) regions and troughs in AT-rich or non-coding regions. Greater depth in Pisifera-Tz enhances confidence in variant detection and structural resolution.

Figure 2:

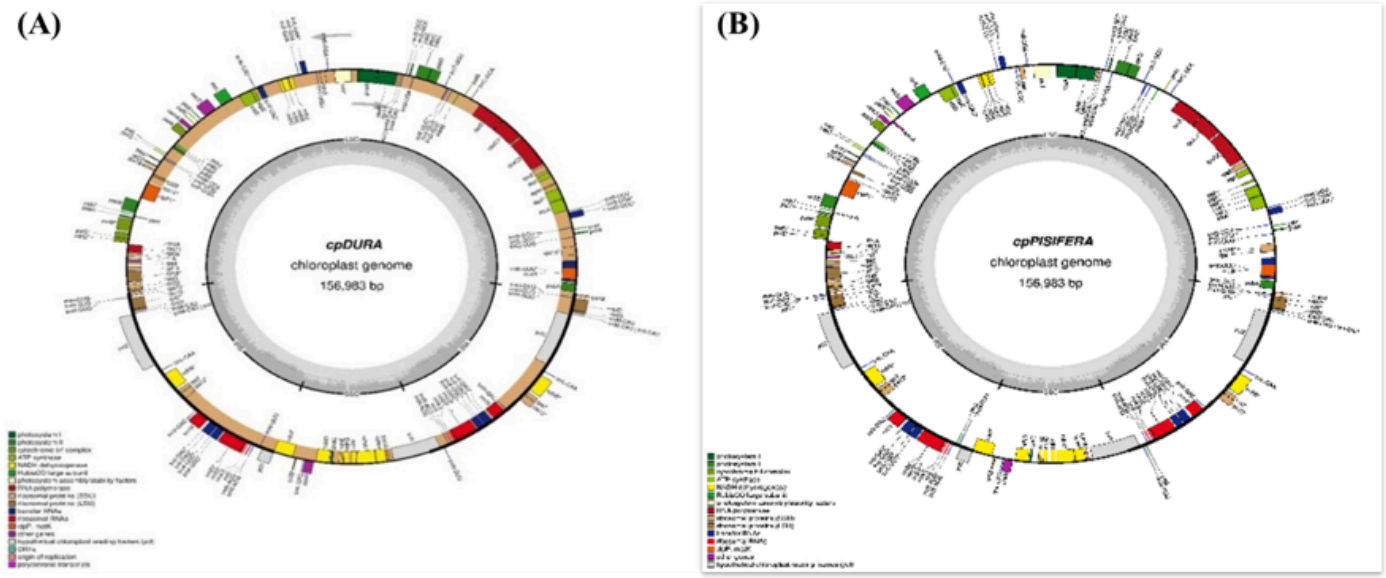


Figure 2

Circular maps of Dura-Tz (PX957236) and Pisifera-Tz (PX957237) chloroplast genomes. Genome organization showing large single-copy (LSC), small single-copy (SSC), and inverted repeat (IR) regions. Gene content and order are highly conserved, with minor differences in intergenic spacers and IR–SC boundaries.

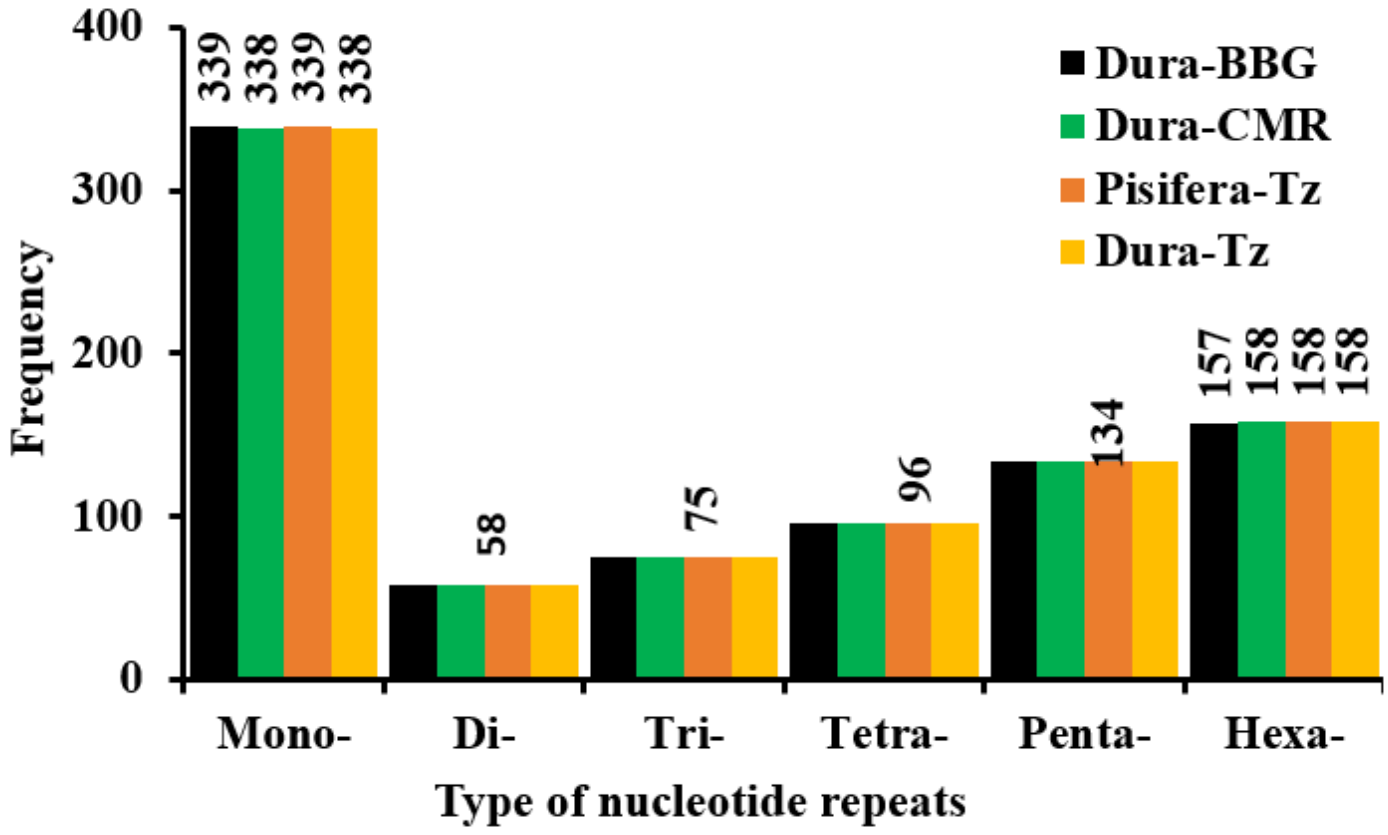


Figure 3

Distribution of simple sequence repeats (SSRs) across chloroplast genomes. Comparison of SSR types and densities among Dura-Tz, Pisifera-Tz, Dura-Cmr, and Dura-BBG. Mononucleotide A/T repeats dominate across all plastomes, with Pisifera-Tz showing markedly reduced genic SSR density.

Figure 4.



Figure 4

Multiple genome alignment of 16 *Elaeis guineensis* plastomes. Showing structural conservation and localize rearrangements. The alignment was performed using the Mauve algorithm to identify Local Collinear Blocks (LCBs). Each horizontal track represents an individual accession, including Dura, Pisifera, and Tenera fruit forms, with the reference genome NC-017602 included for comparison. The green LCB (left) corresponds to the Large Single Copy (LSC) region, while the teal LCB (right) represents the Small Single Copy (SSC) and Inverted Repeat regions. The vertical height of the colored profiles indicates the degree of sequence conservation. Colored lined connecting the tracks denote synteny regions; diagonal red lines and displaced red blocks in accessions Pisifera-Id_221 and pisifera-Id_1109 highlight structural variations, including translocations and inversions relative to the reference. Scale bar at the top indicates genomic position in base pairs.

Figure 5

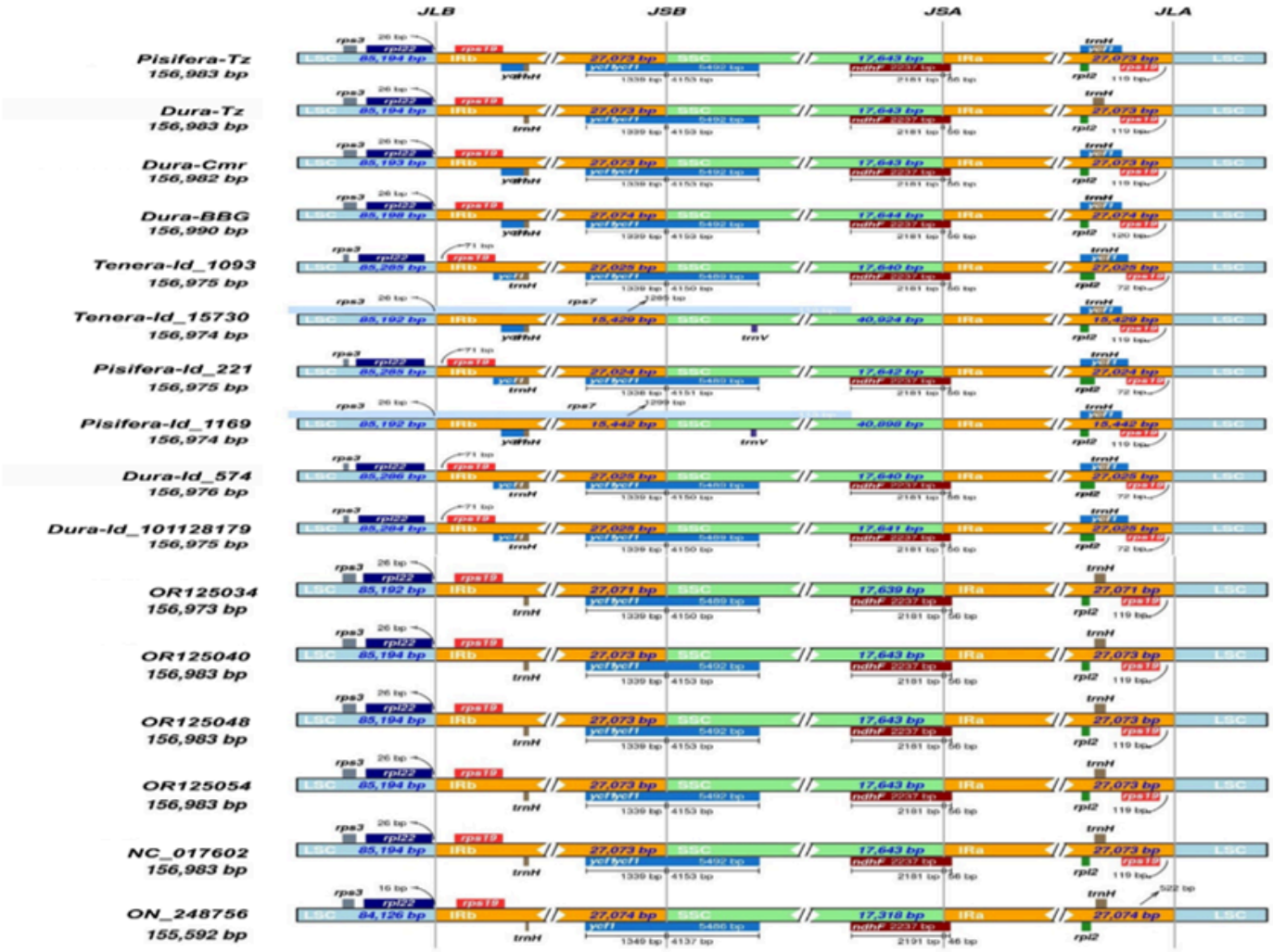


Figure 5

Comparison of inverted repeat (IR) boundary regions among16 oil palm (*Elaeis guineensis*) plastomes using IRscope

Figure 6

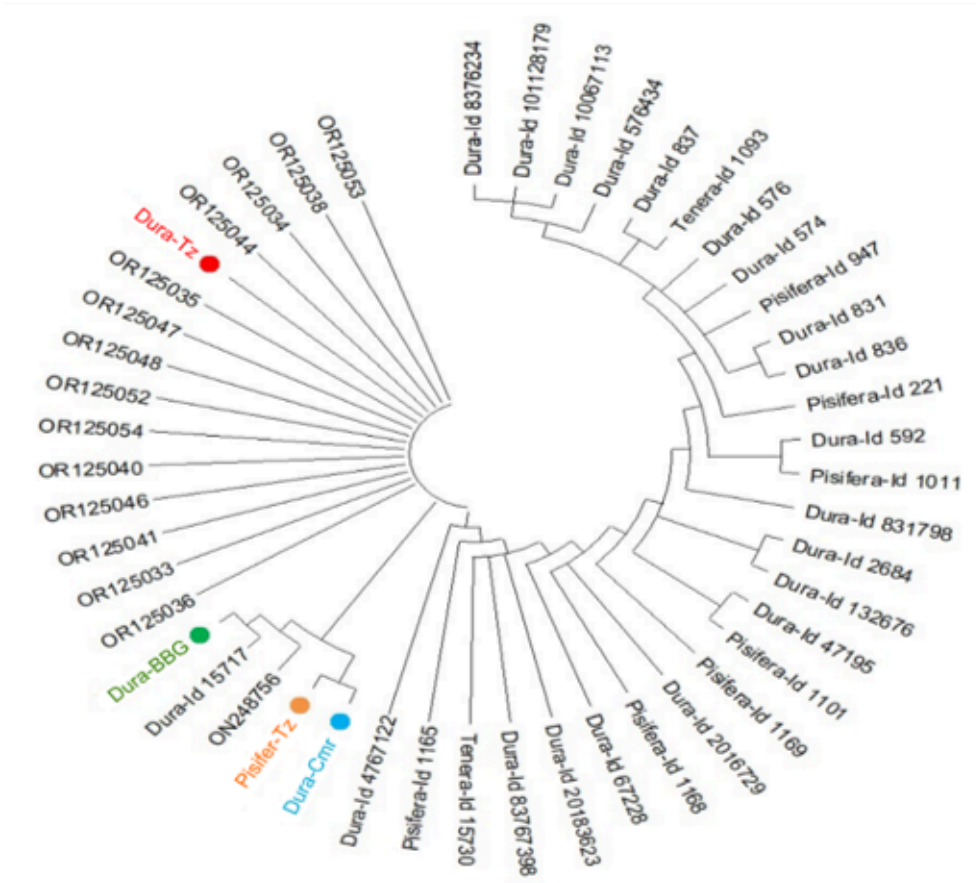


Figure 6

Phylogenetic relationships among oil palm accessions inferred using the Maximum Likelihood method and Tamura -Nei (TN93+G+I) model based on likelihood tree based on complete plastome sequences. Pisifera-Tz clusters with Dura-Cmr, while Dura-Tz and Dura-BBG form distinct clades, reflecting multiple maternal lineages.

Figure 7

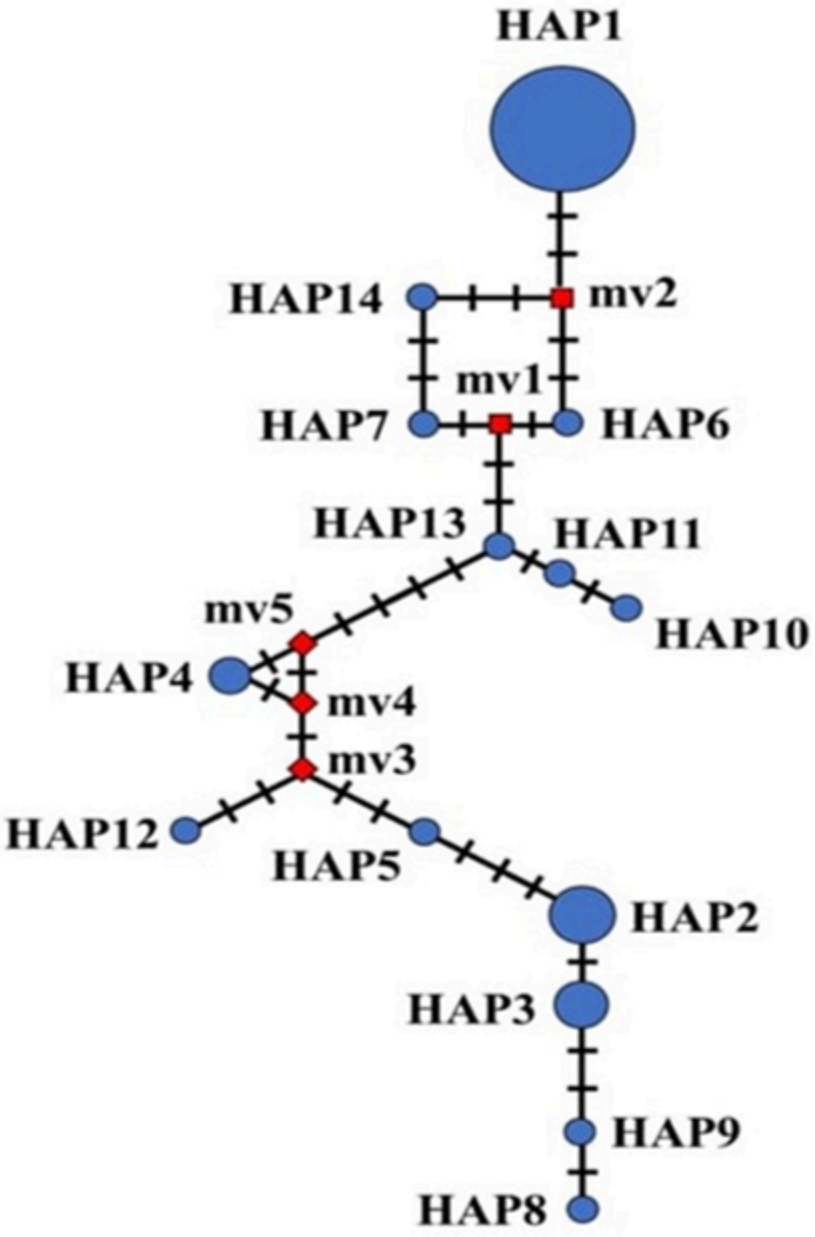


Figure 7

*The fourteen haplotypes from 48 Elaeis guineensis*plastome genomes, with circle sizes proportional to haplotype frequency and red squares (mv1–mv5) denoting inferred median vectors.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AllyMkudeSupplementaryTableS1.docx

- AllyMkudeSupplementaryTableS2.docx
- AllyMkudeSupplementaryTableS3.docx
- AllyMkudeSupplementaryTableS4.docx