

RESEARCH ARTICLES

Chloroplast SNAP Markers Reveal Cytoplasmic Genetic Diversity in Indonesian Sugar Palm (*Arenga pinnata*)

Rizqia Nurul Alfiyanita, Dewi Sukma^{ORCID}, Megayani Sri Rahayu^{ORCID}, and Sudarsono*^{ORCID}

Plant Breeding and Biotechnology Study Program, Department of Agronomy and Horticulture, Faculty of Agriculture, IPB University

*Corresponding author; email: sudarsono_agh@apps.ipb.ac.id

Abstract

Sugar palm (*Arenga pinnata*) is widely distributed across Indonesia, yet its genetic diversity remains poorly characterized. This economically important species is valued for neera sap and young fruits. In this study, the genetic diversity of 60 Indonesian *A. pinnata* accessions from 15 regions was evaluated using chloroplast genome (cpDNA) SNP-based SNAP markers. The markers were developed from informative SNPs identified in complete chloroplast genomes of dwarf and tall *A. pinnata* types, together with one reference accession from NCBI. Haplotype network and phylogenetic analyses resolved 15 distinct haplotypes grouped into three clusters, with HAP15 identified as the most dominant lineage. Cluster III, though represented by only three accessions, occupied a distinct evolutionary position. Geographic mapping revealed that Banten and Bogor harbored the highest haplotype diversity, underscoring their importance for germplasm exploration and conservation. These findings demonstrate the effectiveness of cpDNA SNP-based SNAP markers in uncovering cytoplasmic genetic variation and provide a robust molecular basis for conservation, germplasm management, and future *A. pinnata* improvement programs.

Keywords: *Arenga pinnata*, chloroplast DNA (cpDNA), germplasm conservation, haplotype diversity, SNP-based SNAP markers

Introduction

Sugar palm (*Arenga pinnata* Merr.), a perennial species of the Arecaceae family, is widely distributed across Indonesia and plays a vital role in local economies through its neera sap and fruits (Mujetahid et al., 2023; Zuhud et al., 2020). It grows across diverse altitudes and regions, including Langkat (Fadhilla et al., 2023), Medan (Azhar et al., 2021), Bangka, and Lampung in Sumatra (Rinawati et al., 2021); Lebak, Bogor, Cianjur, Banten, Tasikmalaya, Brebes, and Jember in Java (Rinawati et al., 2021; Terryana et al., 2020; Zuhud et al., 2020); Gowa, Bombana, and Muna in Sulawesi (Rinawati et al., 2021); and West Nusa Tenggara (Haryoso et al., 2020). Despite its broad distribution and economic importance, molecular studies on the genetic diversity and phylogenetic relationships of Indonesian *A. pinnata*—particularly those using chloroplast DNA (cpDNA)—remain limited.

The Indonesian Ministry of Agriculture has released 'Tall', 'Dwarf', and 'Semi tall' *A. pinnata* varieties, distinguished by stem height and flowering time. The 'Tall' variety is characterized by trunk circumference of 106.8 ± 25.2 cm, stem height up to 17.5 m, and a juvenile period of approximately 8–12 years (Ministry of Agriculture of the Republic of Indonesia, 2014). Meanwhile, the 'Dwarf' variety is characterized by a trunk circumference of 139.2 ± 8.90 cm, stem height up to 1.6 m, and a juvenile period of approximately 4–6 years (Ministry of Agriculture of the Republic

of Indonesia, 2011). The ‘Semi-tall’ variety is characterized by trunk circumference of 143.5 ± 13.51 cm, stem height up to 6.8 m, and a juvenile period of approximately 6–8 years (Ministry of Agriculture of the Republic of Indonesia, 2018). These phenotypic differences suggest underlying genetic variations, yet the molecular basis of the diversity remains poorly understood.

Molecular markers are powerful tools for assessing genetic diversity, species identification, and breeding (Özbek, 2024). Among them, cpDNA markers are particularly valuable due to their maternal inheritance, low recombination, and conserved structure (Feng et al., 2024; Wang et al., 2022). CpDNA is also present in high copy numbers in leaf cells (Morley & Nielsen, 2016). Single-nucleotide polymorphisms (SNPs) represent the predominant type of cpDNA variation, occurring in both coding sequences (CDS) and noncoding regions such as intergenic spacers (IGS). In CDS regions, SNPs are typically classified as synonymous or non-synonymous, enabling efficient and precise detection of genetic diversity (Wang et al., 2024).

Single nucleotide amplified polymorphism (SNAP) markers, first developed by Drenkard et al. (2000), are SNP-based markers that employ a modified allele-specific PCR strategy. They use primers with a SNP-specific mismatch at the 3’ terminal base and an additional mismatch near the 3’ end to enhance allelic discrimination during PCR amplification, targeting single-base substitutions at defined genomic positions. Instead of sequencing the entire cpDNA, the development of cpDNA SNP-based SNAP markers targeted to informative SNPs offers significant potential to advance population studies and conservation programs. The SNAP markers enable rapid, efficient, and precise identification of genetic variation, even in species with limited diversity (Heo et al., 2024). For Indonesian *A. pinnata*, which is widely distributed yet under-characterized, cpDNA-based SNAP markers are an effective tool for haplotype determination, phylogenetic analysis, and germplasm conservation strategies (Rahmawati et al., 2022). Therefore, establishing and applying cpDNA-based SNAP markers can

serve as a pivotal starting point for building a robust genomic framework to support *A. pinnata* breeding initiatives.

Accordingly, this study aimed to develop cpDNA-based SNAP markers from informative SNPs and apply them to evaluate haplotype diversity and phylogenetic relationships among 60 *A. pinnata* accessions collected from 15 regions across Indonesia. The findings provide new insights into genetic variation and establish a molecular foundation for germplasm conservation and future cultivar improvement in *A. pinnata*.

Materials and Methods

Research Materials

This study utilized previously assembled cpDNA of *A. pinnata* cv. ‘Akel Toumuung Tall’ and ‘Kutai Timur Dwarf’ from the Plant Molecular Biology Laboratory (PMB Lab.) collections (Alfiyanita, 2025), along with the reference cpDNA sequences of *A. pinnata* (NC_045907.1) obtained from the NCBI repository (https://www.ncbi.nlm.nih.gov/nuccore/NC_045907.1/). This study also used plant samples comprising 60 *A. pinnata* accessions collected from 15 regions across Indonesia: Tanah Datar, Bengkulu, Lahat, Banten, Bogor, West Bandung, Tasikmalaya, Pangandaran, Gunung Kidul, Gresik, Malang, Probolinggo, North Lombok, East Kutai, and BauBau.

SNP and InDel Analysis

SNPs and insertions–deletions (InDels) in *A. pinnata* cpDNA were identified by aligning sequences from three accessions (‘Akel Toumuung Tall’, ‘Kutai Timur Dwarf’, and NC_045907.1) using MAFFT (Yamada et al., 2016) on the Galaxy platform (<https://usegalaxy.eu/>). SNP and InDel counts were manually enumerated in Microsoft Excel 2016.

SNAP primers were designed using WebSnapper (<https://pga.mgh.harvard.edu/cgi-bin/snap3/websnapper3.cgi>) using the following parameters: primer length 18–24

nucleotides, melting temperature (T_m) 50–65 °C, and GC content 40%–60% (Siswanto et al., 2022). Candidate primers were prioritized by shorter length and GC content near 50% to maximize amplification efficiency and thermal stability. Primers flagged for secondary structure formation were excluded during *in silico* screening. Mismatch position relative to the 3' end was evaluated, and primers with mismatches within 1–3 bases of the terminus were retained depending on SNP context.

DNA Isolation and Integrity Testing

Genomic DNA was extracted from 0.4 g of *A. pinnata* leaf tissue using the cetyl trimethyl ammonium bromide (CTAB) method (Doyle & Doyle, 1990), that have been widely used for DNA isolation from palm species (Arshad et al., 2024; Rahmawati et al., 2022). DNA was dissolved in 100 µL TE buffer and stored at -20°C. DNA integrity was assessed by electrophoresis on 1.0% agarose gel in sodium borate buffer. Samples with excessively intense bands compared to the DNA marker signal were diluted to achieve comparable band intensities prior to PCR amplification.

Effectiveness Testing of SNAP Primers

Amplification efficiency of SNAP primers was evaluated by PCR. Each 15 µl PCR reaction contained 7 µl MyTaq HS Red Mix, 2 µl of each forward (Reference or Alternate) and a common reverse primer, and 4 µl of genomic DNA. PCR was performed on a BioRad T100 TM thermocycler under the following conditions: initial denaturation at 95 °C for 5 min, 35 cycles of 95 °C for 30 s, annealing at 50–53 °C for 30 s (primer-dependent), and extension at 72 °C for 30 s; followed by a final extension at 72 °C for 10 min. PCR products were separated on 1.0% agarose gels containing 0.5 µL GelRed™ in 1× SB buffer and run at 100 V for 30 min using a Mupid-ex Umini-gel Basic System (MEX002). A 50 bp DNA ladder was used for size estimation, and bands were visualized under UV illumination. *A. pinnata* samples showing no PCR amplicon

were validated with two additional PCR reactions (i.e., a total of 3 PCR reactions) and were scored negative if all three PCR reactions were negative.

Data Analysis

Banding patterns were scored in binary format, with “1” indicating presence and “0” absence (Rahmawati et al., 2022). Each SNAP marker locus employed two forward primers (REF or ALT) with a single common reverse primer, yielding four binary combinations: 0–0 (both PCR amplicons was absent), 1–0 (PCR amplicon from REF and common reverse primers is present), 0–1 (PCR amplicon from ALT and common reverse primers is present), and 1–1 (PCR amplicon from REF or ALT and common reverse primers is present). These were converted into a four-allele scoring system for downstream analysis:

- 1-1 (two amplicons from PCR using REF or ALT and common reverse primers): indicates two copies of target SNPs, and the scores were converted to 1 0 0 1 (REF = 1 0; ALT = 0 1).
- 1-0 (one amplicon from PCR using REF and common reverse primers): the scores were converted to 1 0 1 0 (both REF = 1 0; ALT absent).
- 0-1 (one amplicon from PCR using ALT and common reverse primers): the scores were converted to 0 1 0 1 (both ALT = 0 1, REF absent).
- 0 0 (no amplicon detected from PCR using either REF or ALT and common reverse primers)

The four-allele scores were used to construct haplotype and phylogenetic networks. Haplotype networks were generated in Network v10.2.0.0 (<https://www.fluxus-engineering.com/sharenet.htm>) using the median-joining method (Bandelt et al., 1999). Phylogenetic relationships were inferred with the weighted neighbor joining method in DARwin v6.0.8 (Perrier & Jacquemoud Collet, 2006), with node support assessed by 1,000 bootstrap replicates.

Results and Discussions

SNP and InDel Analysis of *A. pinnata* cpDNA

Comparative analysis of the chloroplast genomes (cpDNA) from three *A. pinnata* identified a total of 64 SNPs (Table 1 and Table 2) and 46 InDels. Of the 64 SNPs, 48 were located in intergenic spacer (IGS) regions, three in introns, and 13 in the coding region (CDS). Eight of the 13 SNPs in the CDS were non-synonymous, and five were synonymous SNPs, consistent with the general tendency of chloroplast genomes to accumulate variation in regulatory rather than protein-coding sequences (Arshad et al., 2024).

The highest SNP frequency was observed in noncoding regions (Figure 1). Non-synonymous SNPs were detected in *rpoB*, *psaA*, and *ycf1*, while synonymous SNPs were found in *rpoC2* and *ycf1*. Specifically, *rpoB* harbored two non-synonymous SNPs, *psaA* one non-synonymous SNP, *rpoC2* two synonymous SNPs, and *ycf1* five non-synonymous and three synonymous SNPs. In addition, two SNPs were located in the intron of *trnKUUU* and one in the intron of *trnVUAC*.

For InDels, the majority were located in IGS regions, consistent with the high variability of noncoding sequences. Within protein-coding regions, InDels were detected only in *rpoA* and *ycf1*, each occurring within the CDS. Additional

InDels were found in the introns of *trnKUUU*, *rps16*, *atpF*, *rpoC*, and *clpP1*. Overall, the predominance of mutations in noncoding regions reflects the relatively lower selection pressure acting on these regions compared to protein-coding sequences (IñiguezMuñoz et al., 2024).

The distribution of SNPs and InDels highlights several points of biological significance. Noncoding variants, particularly in IGS and intronic regions, provide abundant polymorphisms that can influence transcriptional regulation or RNA splicing efficiency, thereby contributing to subtle differences in chloroplast gene expression (Borsch & Quandt, 2009). Non-synonymous SNPs in *rpoB* and *rpoC2*, which encode core subunits of the plastid-encoded RNA polymerase, may potentially influence plastid transcriptional activity (Liere & Maliga 2001), while the SNP in *psaA* could potentially influence photosystem I function and photosynthetic performance (Azarin et al., 2020). The high density of SNPs and InDels in *ycf1* underscores its role as a hotspot of variation, consistent with its established utility as a phylogenetic marker and its functional importance in protein translocation across the chloroplast membrane (Neubig et al., 2009).

From an evolutionary perspective, the combined presence of SNPs and InDels in both coding and noncoding regions suggests that *A. pinnata* chloroplast genomes harbor a mixture of neutral and potentially adaptive variation.

Table 1
Distribution of SNPs and InDels Across cpDNA of Three *Arenga pinnata* Accessions

Region	SNP count	Remarks	InDel count	Remarks
Intergenic spacers (IGS)	48	Noncoding SNPs (regulatory variation)	36	Predominantly in IGS regions
Introns	3	<i>trnKUUU</i> (2 SNPs), <i>trnVUAC</i> (1 SNP)	8	<i>trnKUUU</i> , <i>rps16</i> , <i>atpF</i> , <i>rpoC</i> , <i>clpP1</i>
Coding sequences (CDS)	13	<i>rpoB</i> (2 nonsyn), <i>psaA</i> (1 nonsyn), <i>rpoC2</i> (2 syn), <i>ycf1</i> (5 nonsyn, 3 syn)	2	<i>rpoA</i> (1 InDel), <i>ycf1</i> (1 InDel)
Total	64	8 nonsyn, 5 syn, remainder in noncoding regions	46	Majority in noncoding, few in CDS and introns

Note. syn - synonymous, nonsyn - non-synonymous.

Table 2

Nucleotide Substitutions among Identified SNPs across cpDNA of Three Arenga pinnata Accessions

Nucleotide position	NC_04 5907	Ap_KTD	Ap_ATT	Additional information	Nucleotide position	NC_04 5907	Ap_KTD	Ap_ATT	Additional information
1,348	A	A	T	Intergenic	65,783	A	A	T	Intergenic
3,544	C	C	T	Intergenic	65,839	A	A	C	Intergenic
4,043	T	T	G	Intergenic	66,291	C	C	A	Intergenic
4,240	T	T	A	Intergenic	68,144	C	C	T	Intergenic
6,724	C	C	A	Intergenic	69,493	A	A	G	Intergenic
8,392	A	A	G	Intergenic	75,161	A	A	G	Intergenic
14,072	C	C	A	Intergenic	80,772	T	T	A	Intergenic
14,073	A	A	T	Intergenic	103,117**	C	C	T	Intergenic
14,494	A	A	C	Intergenic	116,903	A	A	T	Intergenic
18,492**	A	A	G	<i>rpoC2</i> gene	117,545	T	T	A	Intergenic
20,741	T	T	G	<i>rpoC2</i> gene	117,546	C	C	T	Intergenic
26,325**	A	A	C*	<i>rpoB</i> gene	117,547	T	T	A	Intergenic
27,042**	T	T	C*	<i>rpoB</i> gene	117,548	T	T	G	Intergenic
28,926	C	C	A	Intergenic	117,549	A	A	G	Intergenic
31,923	G	G	A	Intergenic	117,552	C	C	T	Intergenic
32,246	A	A	C	Intergenic	117,553	C	C	A	Intergenic
36,379	G	G	T	Intergenic	117,554	T	T	A	Intergenic
36,380	A	A	T	Intergenic	117,555	A	A	G	Intergenic
36,381	A	A	C	Intergenic	117,556	T	T	A	Intergenic
37,464	A	A	C	Intergenic	121,882	T	T	G	Intergenic
41,051**	A	A	G*	<i>psaA</i> gene	121,885	T	T	A	Intergenic
48,063	A	A	T	Intergenic	124,127	C	C	T	Intergenic
49,910	A	A	T	Intergenic	124,142	G	G	C	Intergenic
53,499	A	A	G	Intergenic	128,841	C	C	A*	<i>ycf1</i> gene
54,512	C	C	A	Intergenic	129,317	T	T	G*	<i>ycf1</i> gene
65,776	A	A	T	Intergenic	129,369	A	A	T*	<i>ycf1</i> gene
65,777	T	T	G	Intergenic	130,745	G	G	A	<i>ycf1</i> gene
65,778	T	T	G	Intergenic	132,226	T	T	A*	<i>ycf1</i> gene
65,779	T	T	A	Intergenic	132,227	T	T	A*	<i>ycf1</i> gene
65,780	T	T	A	Intergenic	132,228	T	T	A	<i>ycf1</i> gene
65,781	C	C	A	Intergenic	132,459**	T	T	C	<i>ycf1</i> gene
65,782	C	C	A	Intergenic	144,118	G	G	A	Intergenic

Notes: Ap_KTD = 'Kutai Timur Dwarf'; Ap_ATT = 'Akel Toumuung Tall'. ** = potential for SNAP marker primer design, * = Non-synonymous SNP.

Noncoding mutations provide sensitive markers for population differentiation, while coding mutations in photosynthesis and transcription related genes may underlie physiological differences among tall, dwarf, and semi-tall varieties.

Importantly, these polymorphisms form a robust foundation for SNAP marker development. Noncoding SNPs and InDels enhance haplotype resolution, while coding SNPs and InDels in *rpoB*, *psaA*, *rpoA*, and *ycf1* link genetic variation to functional traits, enabling markers that capture both neutral diversity and adaptive potential. By targeting these loci, SNAP markers can efficiently distinguish *A. pinnata* accessions, clarify phylogenetic relationships, and support germplasm conservation and breeding initiatives.

The single non-synonymous SNP in *psaA*, a core component of Photosystem I, could alter photosynthetic efficiency and contribute to adaptation across diverse light environments. The high density of SNPs in *ycf1*—a gene widely recognized as a phylogenetic marker—underscores its role as a hotspot of variation, with potential impacts on protein translocation across the chloroplast membrane and overall organelle performance.

The identification of 64 SNPs and 46

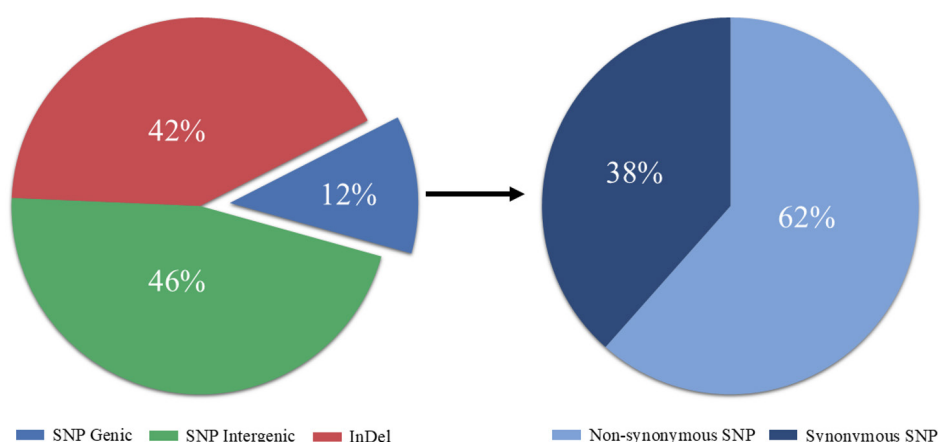
InDels across the chloroplast genomes of the three *A. pinnata* varieties provided a comprehensive catalog of genetic variations. While InDels contributed valuable insights into structural diversity, SNPs offered higher resolution for distinguishing allelic differences and tracing phylogenetic relationships. Building on this foundation, the next step focused on translating these sequence polymorphisms into practical molecular tools. Specifically, SNPs were screened for their suitability in SNAP primer design, enabling the development of targeted markers that could reliably detect genetic variation among *A. pinnata* accessions.

SNP-Based Primer Development

Three categories of SNPs were identified in the chloroplast genomes: non-synonymous SNPs, synonymous SNPs, and SNPs located in noncoding regions (IGS). Non-synonymous SNPs are particularly valuable for marker development because they alter amino acid sequences and may affect protein structure and function (Teng et al., 2008). Although synonymous SNPs do not change amino acids, they can still influence protein properties, gene expression, and molecular stability (Hoda et

Figure 1

Distribution of Sequence Variations in cpDNA of Three Sugar Palm (Arenga pinnata) Accessions



Notes. The left pie chart illustrates the proportion of SNPs and InDels located in coding (genic) versus noncoding (intergenic) regions. The right pie chart shows the relative proportions of synonymous and non-synonymous SNPs within coding regions.

al., 2023). Likewise, SNPs within IGS regions are useful for plant identification, as they can effectively distinguish varieties within a species (Enan, 2020).

A total of 64 SNP sites were initially identified through comparative cpDNA analysis. Since the primary objective was to develop informative SNAP markers for genetic diversity analysis rather than to assess functional effects, SNPs were not pre-selected based on genomic location or predicted biological impact. Instead, candidate SNPs were screened for suitability for primer design. Screening followed established parameters: primer length of 18–24 bp, GC content of 40%–60%, melting temperature (T_m) between 50–65 °C, and a maximum T_m difference of 5 °C between forward and reverse primers (Siswanto et al., 2022). SNP sites flanked by regions that met these criteria and lacked secondary structures were retained. Based on this filtering, six SNP loci were selected for SNAP primer development (Table 3).

Six sets of SNAP primers were successfully designed to target SNP18492, SNP26325, SNP27042, SNP41051, SNP103117, and SNP132459. Among these, SNP18492 and SNP132459 are synonymous substitutions in *rpoC2* and *ycf1*, respectively; SNP26325, SNP27042, and SNP41051 are non-synonymous substitutions in *rpoB* and *psaA*; and SNP103117 is located in the IGS region. Each SNP site requires two primer pairs: one forward/reverse pair for the reference allele (REF) and another for the alternative allele (ALT). Although two distinct forward primers were designed for each allele, a single reverse primer was used for both REF and ALT alleles, as the reverse sequences were identical across variants (Balladona et al., 2020). Table 3 lists the six SNAP primer pairs developed, all of which met the established design criteria.

The development of six SNAP primer sets provided a practical toolkit for detecting chloroplast SNP variation in *A. pinnata*.

Table 3

SNAP Primers Developed from Selected SNPs in the cpDNA of Three Arenga pinnata Accessions

Primer name	Primer sequence	Primer size (bp)	T _m	Expected amplicon (bp)
18492-Reference	TTATAATCCCGATCCGTGC	19	59.1 °C	354
18492-Alternate	GATTTATAATCCCGATCCACGT	22	62.0 °C	
18492-Reverse	TGTTCCACTTTTGGGAAGACCT	21	60.1 °C	
26325-Reference	CGATATCGAGCTGGAGCAC	19	61.7 °C	341
26325-Alternate	CGATATCGAGCTGGAGCAA	19	61.3 °C	
26325-Reverse	CATGGGCGAAAATTGAGTTA	20	58.9 °C	
27042-Reference	GTCGACCAATCCTTCCTACTC	21	61.4 °C	368
27042-Alternate	TCGACCAATCCTTCCTGATT	20	6.01 °C	
27042-Reverse	CGTAATTTCTGTCTATACCAGCACT	25	62.5 °C	
41051-Reference	GAAGACATGATCCCATGCG	19	60.4 °C	367
41051-Alternate	GAAGACATGATCCCAGCCA	19	61.4 °C	
41051-Reverse	TTCAGATACCGCTATACAATTACAAC	26	61.4 °C	
103117-Reference	CTCTATCCCTATCCCATAGGTGT	23	62.0 °C	340
103117-Alternate	CTCTATCCCTATCCCATAGGAAC	23	60.6 °C	
103117-Reverse	GAAGTGTCTTTTCTGTATACTTTCCC	26	61.7 °C	
132459-Reference	CCTTTATTGAGTGGGCCG	18	59.9 °C	373
132459-Alternate	TCCTTTATTGAGTGGGCCA	19	60.9 °C	
132459-Reverse	TATTTCTAGTTGTTGTTTGATCATTTG	27	60.0 °C	

To validate their utility, these primers were subsequently tested across 60 accessions under standardized PCR conditions. This effectiveness testing ensured that the designed primers not only met theoretical design criteria but also produced reliable amplification patterns in practice, thereby confirming their suitability for downstream diversity and haplotype analyses.

Effectiveness Testing of SNAP Primers

Six pairs of SNAP primers were successfully developed and tested on 60 *A. pinnata* accessions using PCR amplification. Given the haploid nature of cpDNA, each accession was expected to yield a single amplification product per primer pair, consisting of a forward primer (REF or ALT) combined with a common reverse primer. Under optimized PCR conditions and standardized DNA input, most accessions produced the expected single-band pattern. Nevertheless, previous studies have reported sequence duplications in certain cpDNA regions, particularly within inverted repeats and specific genes (Lee et al., 2007), which can occasionally result in multiple visible bands.

PCR amplification with the SNAP primers yielded variable amplification patterns among accessions. For example, results obtained with primer pair SNAP18492 on 12 accessions (Figure 2) showed that most samples yielded positive amplification for both primer combinations (REF-reverse and ALT-reverse). In contrast, a subset of accessions failed to amplify with the REF-reverse primer but showed positive amplification with the ALT-reverse primer. These variations confirm that the developed SNAP primers are effective in detecting SNPs within *A. pinnata* cpDNA and can differentiate allelic variants among accessions. The PCR amplification data was subsequently converted into binary format (presence/absence of bands) for downstream diversity and haplotype analysis (Figure 3). This transformation enabled the integration of SNAP marker data into phylogenetic and population-structure analyses, providing a robust framework for distinguishing tall, dwarf, and semi-tall varieties.

The varied amplification patterns observed across accessions demonstrate that the SNAP primers are not only reliable for detecting cpDNA SNPs but also sensitive enough to capture allelic differences among *A. pinnata* varieties. The consistent single band outcomes in most samples confirm primer specificity, while the occasional REF/ALT amplification differences highlight genuine sequence variation rather than technical artifacts. By converting these PCR results into binary data, the primers provided a robust dataset for downstream haplotype resolution and phylogenetic clustering, thereby validating their effectiveness as practical tools for genetic diversity analysis in *A. pinnata*. The application of these six SNAP primer sets resolved haplotype diversity and phylogenetic relationships among *A. pinnata* accessions, providing the foundation for subsequent network and tree analyses.

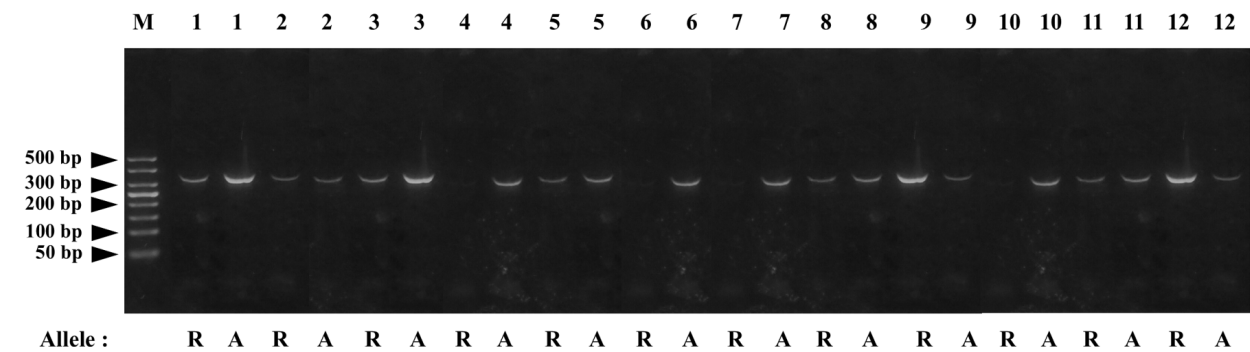
Haplotype Diversity Analysis

Genetic diversity analysis using SNAP markers identified 15 haplotypes among the 64 *A. pinnata* accessions analyzed (Figure 4a). The most prevalent haplotype, HAP15, occurred in 36 accessions and was widely distributed across multiple geographic regions, including Probolinggo (8), Bengkulu (7), Lahat (4), Bogor (3), Gunung Kidul (3), Lebak (2), Gresik (2), West Sumatra (2), Southeast Sulawesi (2), and single accessions from Malang, West Bandung, and Pangandaran. Comparative analysis revealed that HAP15 differed from HAP13 by two loci. HAP13, detected in West Bandung and Bengkulu, shared sequential evolutionary relationships with HAP12 and HAP11, each differing by two loci. HAP12 was found in Bogor, Pangandaran, and Bengkulu, while HAP11 occurred in Banten. Further mutational steps from HAP13 produced median vector MV4, which evolved into HAP14, detected in Bogor.

Additional haplotypes were derived through successive mutational changes. HAP10 originated from HAP11 through three locus changes and was represented by one Bogor accession. HAP10 subsequently evolved into

Figure 2

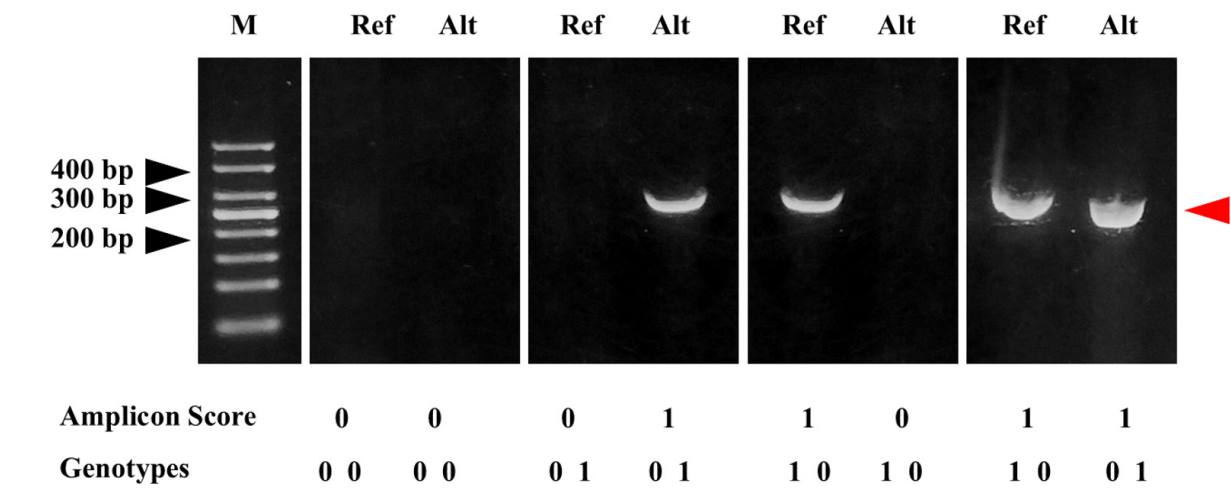
Allele Amplification Profile Using cpDNA-Based SNAP Locus 18492. PCR Products were Resolved on a 1% Agarose Gel in Sodium Borate Buffer and Stained with GelRed™ (0.5 µL). Electrophoresis was Performed at 100 V for 30 Min



Notes. Lanes 1–3 represent sugar palm (*Arenga pinnata*) accessions from Probolinggo, lanes 4–6 from Bogor, lanes 7–9 from Gunung Kidul, and lanes 10–12 from Bengkulu. R = reference allele amplicon, A = alternate allele amplicon, M = 50 bp DNA size markers.

Figure 3

Examples of the Four-allele Scoring and Binary Data Representation for cpDNA-based SNAP Markers and Amplicon Patterns Generated by PCR using Reference (REF–reverse) and Alternate (ALT–reverse) Primer Combinations at a Single SNAP Locus



HAP9, which gave rise to MV3 and subsequent mutations that produced HAP8 and HAP7. HAP9 was found in Banten, HAP8 in Bengkulu and Banten, and HAP7 in Gunung Kidul. Successive mutations in HAP7 generated MV2 and MV1, which in turn produced HAP6 and HAP4 (Bogor and North Lombok). MV1 also gave rise to HAP2, which diverged from HAP1 by two loci. HAP4 evolved into HAP5 and HAP1, while HAP5 produced HAP3. HAP5, HAP3, and HAP2 were

each represented by a single accession from North Lombok, East Kutai, and Tasikmalaya, respectively. HAP1 comprised seven accessions from Bogor (3), Gunung Kidul (1), Malang (2), and West Bandung (1), suggesting multiple independent derivations.

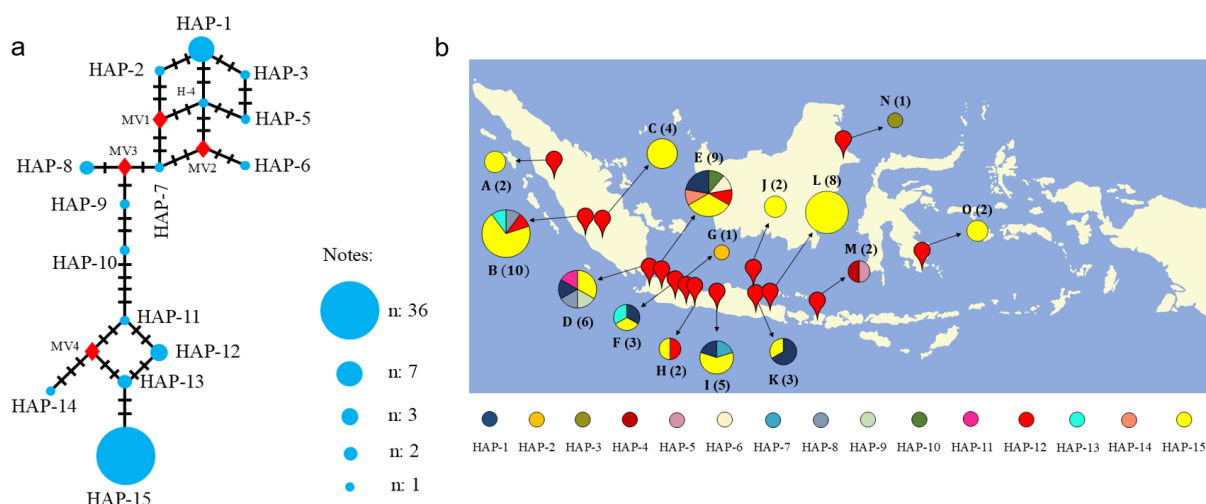
The geographic distribution of haplotypes (Figure 4b) revealed that Banten and Bogor exhibited the highest levels of chloroplast haplotype diversity, underscoring their

importance as priority regions for germplasm exploration and conservation. Chloroplast haplotype diversity is particularly valuable for evolutionary studies due to the conserved structure and low recombination rates of cpDNA (He et al., 2024). Consistent with findings in other crops such as tomato (Xu et al., 2024), ambarella (da Silva et al., 2024), pear (Heo et al., 2024), wheat (Kumar et al., 2020), and melon (Zhang et al., 2023), SNP-based markers proved effective for detecting genetic variation and resolving phylogenetic relationships. As a PCR-based implementation of SNP genotyping, SNAP markers offer a cost effective alternative to high throughput sequencing, particularly in resource limited laboratories.

SNP-based molecular markers have been widely demonstrated to be effective in detecting genetic variation and resolving phylogenetic relationships in diverse plant species, including tomato (Xu et al., 2024), ambarella (da Silva et al., 2024), pear (Heo et al., 2024), wheat (Kumar et al., 2020), and melon (Zhang et al., 2023). As a PCR-based implementation of SNP genotyping, SNAP markers offer a practical, cost-effective approach for phylogenetic and population genetic analyses, particularly in laboratories with limited access to high-throughput sequencing technologies. Haplotype network analysis enables the visualization of mutation patterns among allele combinations and the relationships between mutations within a population.

Figure 4

(a) Haplotype Network of 64 Sugar Palm (*Arenga pinnata*) Accessions Based on SNP Variation Detected by cpDNA-Based SNAP Markers, Showing 15 Distinct Haplotypes (HAP1 to HAP15) Connected by Mutational Steps. Blue Circles Represent Haplotypes, with Circle Size Proportional to Frequency, Red Diamonds (MV1–MV4) Indicate Median Vectors Representing Hypothetical Ancestral or Intermediate Haplotypes. (b) Geographic Distribution of Haplotypes Across 15 Regions in Indonesia, Visualized as Pie Charts Indicating the Proportion of Haplotypes per Location



Note. The widespread occurrence of HAP 15 highlights its dominance and likely common haplotype status, while the presence of rare haplotypes in regions such as Banten and Bogor underscores their importance for germplasm conservation.

As shown in Figures 4a and 4b, HAP-15 is the most frequent haplotype, widely distributed across multiple geographic regions, and likely represents an ancestral lineage. This interpretation is supported by significant allelic heterogeneity, which contributes to natural genetic variation associated with adaptation to diverse environmental conditions in tropical plant populations. The consistent presence of this dominant haplotype in diverse environments suggests that it may have been subject to selection and preserved due to adaptive advantages (Li et al., 2014). While high-frequency haplotypes are often associated with older lineages, this relationship is not universally applicable (Roodt et al., 2017). Some studies propose that high-frequency haplotypes likely originated in the early Pleistocene; however, estimates of their ages vary considerably, indicating a complex relationship between haplotype frequency and evolutionary age that warrants further investigation (Mráz et al., 2025).

Phylogenetic Analysis

The phylogenetic tree reconstructed from SNP diversity among 60 *A. pinnata* accessions (Figure 5) revealed three distinct clusters, consistent with the groupings observed in the haplotype network. Cluster I contained the largest proportion of accessions (44/60). This cluster included one accession each of HAP14, HAP10, and HAP11; three of HAP12; two of HAP13; and 36 of HAP15. The close positioning of HAP15, HAP13, and HAP12 suggests an evolutionary relationship among these haplotypes, with HAP15 serving as a central lineage.

Cluster II was subdivided into two subgroups. Subcluster IIA comprised five haplotypes (HAP1 through HAP5) totaling 11 accessions, while Subcluster IIB included HAP6 and HAP7, each represented by a single accession. HAP1, the second most frequent haplotype after HAP15, was placed in Cluster II and branched far from the HAP15 group, indicating substantial genetic divergence between these lineages. Cluster III consisted of three accessions (one HAP9 accession and two

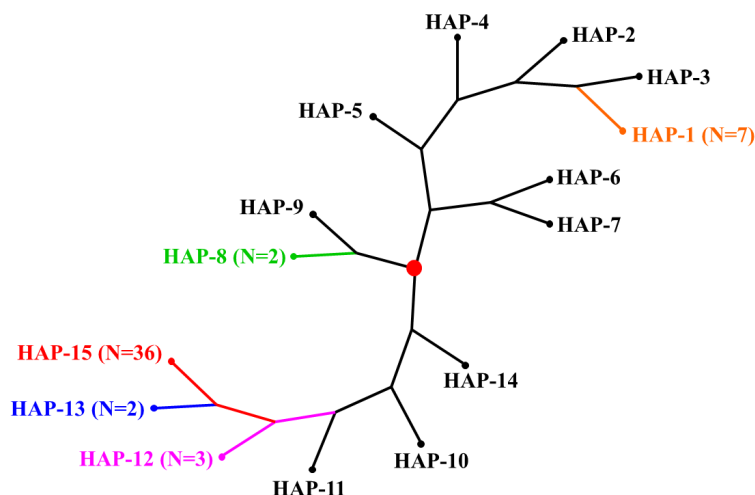
HAP-8 accessions). Despite its small size, this cluster formed a distinct branch near the central node (median vector), suggesting an early divergence from the main lineage.

The phylogenetic structure identified in this study is consistent with previous investigations of *A. pinnata* genetic diversity based on nuclear SSR markers. SSR-based studies conducted in Cianjur, Banten, and other regions of Indonesia reported pronounced geographic structuring and the presence of genetically distinct population clusters, particularly in western Java, which was highlighted as a center of high genetic diversity (Rinawati et al., 2021; Terryana et al., 2020). The concordance between the cpDNA SNAP-based clusters observed here and the nuclear SSR-based population structure reported previously suggests that the genetic differentiation among Indonesian *A. pinnata* populations reflects long-term evolutionary and demographic processes rather than marker-specific effects. However, because cpDNA is maternally inherited and lacks recombination, the haplotype patterns described in this study specifically capture maternal lineage relationships, thereby complementing the biparental inheritance patterns revealed by SSR markers.

From an applied perspective, the identification of distinct haplotype clusters provides valuable guidance for *A. pinnata* germplasm management and pre-breeding strategies. Accessions representing divergent chloroplast lineages, such as those from Clusters I and II or the unique haplotypes within Cluster III, may be prioritized as parental sources to broaden the genetic base of breeding populations. Furthermore, regions harboring rare or phylogenetically distinct haplotypes should be prioritized for germplasm exploration and conservation. Although cpDNA SNAP markers do not directly capture nuclear genetic variation associated with agronomic traits, the phylogenetic framework established in this study offers a robust foundation for integrating molecular diversity data with future phenotypic and genomic analyses to support long-term *A. pinnata* improvement programs.

Figure 5

Phylogenetic Tree of 60 Sugar Palm (Arenga pinnata) Accessions Reconstructed from SNP Variation Detected by cpDNA-Based SNAP Markers



Notes. The tree reveals three major clusters consistent with the haplotype network: Cluster I (44 accessions) dominated by HAP 15 and closely related haplotypes (HAP 12, HAP 13, HAP 14, HAP 10, HAP 11); Cluster II (13 accessions) subdivided into IIA (HAP 1 to HAP 5) and IIB (HAP 6, HAP 7), highlighting divergence from the HAP 15 lineage; and Cluster III (3 accessions) comprising HAP 8 and HAP 9, forming a distinct branch near the central node. The clustering pattern corroborates haplotype network results and reflects maternal lineage differentiation within Indonesian sugar palm populations.

Conclusions

This study identified structural variations in *A. pinnata* cpDNA and successfully developed six SNPbased SNAP primer pairs targeting chloroplast regions. Genetic diversity analysis of 60 accessions from across Indonesia revealed 15 distinct chloroplast haplotypes, reflecting substantial cytoplasmic genetic variation. Phylogenetic analysis further classified these haplotypes into three primary clusters: Cluster I with the highest frequency, and Clusters II and III representing less common but phylogenetically informative maternal lineages. The clear separation of haplotypes and clusters provides a practical basis for identifying genetically distinct cytoplasmic lineages that can be prioritized in germplasm conservation and incorporated into pre-breeding efforts. Overall, the findings confirm that cpDNA-based SNAP markers are effective for evaluating cytoplasmic genetic diversity in *A. pinnata* and offer a robust molecular framework to support germplasm management and the early stages of *A. pinnata* improvement programs.

References

- Alfiyanita, R. N. (2025). *Comparative genomic of cpDNA and diversity of Indonesian sugar palm using molecular marker based on chloroplast genome* [Mater's Thesis, IPB University]. IPB University Scientific Repository. <https://repository.ipb.ac.id/handle/123456789/171471>
- Arshad, A., Aditama, R., Rahayu, M. S., & Natawijaya, A. (2024). Comparative study of chloroplast genomes across seven *Salacca* species. *Biodiversitas Journal of Biological Diversity*, 25(11), 4043–4058. <https://doi.org/10.13057/biodiv/d251104>
- Azarin, K., Usatov, A., Makarenko, M., Kozel, N., Kovalevich, A., & Dremuk, I. (2020). A point mutation in the photosystem I P700 chlorophyll a apoprotein A1 gene confers variegation in *Helianthus annuus* L. *Plant Molecular Biology*. <https://doi.org/10.1007/s11103-020-00997-x>
- Azhar, I., Nasution, Z., Delvian, Agussabti, Aulin, F. R., & Sembiring, M. R. (2021). Utilization

- of sugar palm (*Arenga pinnata* Merr) by the communities around the PT Toba Pulp Lestari. *IOP Conference Series: Earth and Environmental Science*, 782. Article 032017. <https://doi.org/10.1088/1755-1315/782/3/032017>
- Balladona, F. K., Maskromo, I., & Sukma, D. (2020). Development of Molecular Marker Based on SNP sites and Indel in Coconut Chloroplast Genome. *Jurnal Agronida*, 6(1), 10–18. <https://doi.org/10.30997/jag.v6i1.2548>
- Bandelt, H. J., Forster, P., & Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16(1), 37–48. <https://doi.org/10.1093/oxfordjournals.molbev.a026036>
- Borsch, T., & Quandt, D. (2009). Mutational dynamics and phylogenetic utility of noncoding chloroplast DNA. *Plant Systematics and Evolution*, 282, 169–199. <https://doi.org/10.1007/s00606-009-0210-8>
- da Silva, A. V., Garcia, C. B., de Carvalho, I. A. S., do Nascimento, W. F., Ramos, S. L. F., Rodrigues, D. P., Zucchi, M. I., Costa, F. M., Alves-Pereira, A., Batista, C. E., da Silva, E. F., & Veasey E. A. (2024). High genetic structure of *Spondias mombin* in Brazil revealed with SNP markers. *Genetic and Molecular Biology*, 47(4). <https://doi.org/10.1590/1678-4685-GMB-2024-0030>
- Doyle, J. J., & Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. *Focus*, 12, 13–15. <https://doi.org/10.2307/2419362>
- Drenkard, E., Richter, B. G., Rozen, S., Stutius, L. M., Angell, N. A., Mindrinos, M., Cho, R. J., Oefner, P. J., Davis, R. W., & Ausubel, F. M. (2000). A simple procedure for the analysis of single nucleotide polymorphisms facilitates map-based cloning in arabidopsis. *Plant Physiology*, 124(4), 1483–1492. <https://doi.org/10.1104/pp.124.4.1483>
- Enan, M. R. (2020). Molecular identification of date palm varieties using chloroplast barcode atpF-atpH spacer. *Indian Journal of Pure and Applied Biosciences*, 8(3), 1–12. <https://doi.org/10.18782/2582-2845.8075>
- Fadhilla, S., Hakim, L., Agustian, A., Lubis, Y. S., & Siregar, A. W. (2023). The diversity of utilizations, tapping flow discharge, and conservation of sugar palm (*Arenga longipes* Moega) cultivated in Langkat, North Sumatra, Indonesia. *Biodiversitas*, 24(1), 122–132. <https://doi.org/10.13057/BIODIV/D240116>
- Feng, Z., Zheng, Y., Jiang, Y., Pei, J., & Huang, L. (2024). Phylogenetic relationships, selective pressure and molecular markers development of six species in subfamily Polygonoideae based on complete chloroplast genomes. *Scientific Reports*, 14, Article 9783. <https://doi.org/10.1038/s41598-024-58934-7>
- Haryoso, A., Zuhud, E. A. M., Hikmat, A., Sunkar, A., & Darusman, D. (2020). Ethnobotany of sugar palm (*Arenga pinnata*) in the sasak community, Kekait village, West Nusa Tenggara, Indonesia. *Biodiversitas*, 21(1), 117–128. <https://doi.org/10.13057/biodiv/d210116>
- He, S., Xu, B., Chen, S., Li, G., Zhang, J., Xu, J., Wu, H., Li, X., & Yang, Z. (2024). Sequence characteristics, genetic diversity and phylogenetic analysis of the *Cucurbita ficifolia* (*Cucurbitaceae*) chloroplasts genome. *BMC Genomics*, 25(1), 1–15. <https://doi.org/10.1186/s12864-024-10278-2>
- Heo, J. H., Yeon, J., Jung, J. K., Shin, I. S., & Sim, S. C. (2024). Development of cost-effective SNP markers for genetic variation analysis and variety identification in cultivated pears (*Pyrus* spp.). *Plants*, 13(18), 2600. <https://doi.org/10.3390/plants13182600>
- Hoda, A., Bixheku, X., & Lika, M. (2023). Computational analysis of non-synonymous single nucleotide polymorphism in the bovine PKLR gene. *Journal of Biomolecular Structure and Dynamics*, 42(8), 1–14. <https://doi.org/10.1080/07391102.2023.2219315>

- Iñiguez-Muñoz, S., Llinàs-Arias, P., Ensenyat-Mendez, M., Bedoya-López, A. F., Orozco, J. I. J., Cortés, J., Roy, A., Forsberg-Nilsson, K., DiNome, M. L., & Marzese, D. M. (2024). Hidden secrets of the cancer genome: Unlocking the impact of non-coding mutations in gene regulatory elements. *Cellular and Molecular Life Sciences*, 81(1), 37. <https://doi.org/10.1007/s00018-024-05314-z>
- Kumar, D., Chhokar, V., Sheoran, S., Singh, R., Sharma, P., Jaiswal, S., Iquebal, M. A., Jaiswar, A., Jaisri, J., Angadi, U. B., Rai, A., Singh, G.P., Kumar, D., & Tiwari, R. (2020). Characterization of genetic diversity and population structure in wheat using array-based SNP markers. *Molecular Biology Reports*, 47(1), 293–306. <https://doi.org/10.1007/s11033-019-05132-8>
- Lee, H., Jansen, R. K., Chumley, T. W., & Kim, K. (2007). Gene relocations within chloroplast genomes of jasminum and menodora (Oleaceae) are due to multiple, overlapping inversions. *Molecular Biology and Evolution*, 24(5), 1161–1180. <https://doi.org/10.1093/molbev/msm036>
- Li, P., Filiault, D., Box, M. S., Kerdaffrec, E., van Oosterhout, C., Wilczek, A. M., Schmitt, J., McMullan, M., Bergelson, J., Nordborg, M., & Dean, C. (2014). Multiple FLC haplotypes defined by independent cis-regulatory variation underpin life history diversity in *Arabidopsis thaliana*. *Genes & Development*, 28(15), 1635–1640. <https://doi.org/10.1101/gad.245993.114>
- Liere, K., & Maliga, P. (2001). Plastid RNA polymerases in higher plants. In E.-M. Aro & B. Andersson (Eds.), *Regulation of photosynthesis* (pp. 29–49). Kluwer Academic Publishers. https://doi.org/10.1007/0-306-48148-0_2
- Ministry of Agriculture of the Republic of Indonesia. (2011). *Description of Genjah Kutai Timur sugar palm variety*. Decree No. 3879/Kpts/Sr.120/9/2011. Directorate General of Food Crops, Jakarta, Indonesia.
- Ministry of Agriculture of the Republic of Indonesia. (2014). *Description of Akel Toumuung sugar palm variety*. Decree No. 1059/Kpts/Sr.120/10/2014. Directorate General of Food Crops, Jakarta, Indonesia.
- Ministry of Agriculture of the Republic of Indonesia. (2018). *Description of Parasi sugar palm variety*. Decree No. 910/Kpts/KB.310/12/2018. Directorate General of Food Crops, Jakarta, Indonesia.
- Morley, S. A., & Nielsen, B. L. (2016). Chloroplast DNA copy number changes during plant development in organelle DNA polymerase mutants. *Frontiers in Plant Science*, 7, 57. <https://doi.org/10.3389/fpls.2016.00057>
- Mráz, P., Flašková, L., Chrtek, J., Mrázová, V., Puščaš, M., Josefiová, J., & Závěská, E. (2025). Persistence over millennia through extreme clonal longevity: Phylogenomic insight into history of one of the world's rarest plant species. *Journal of Biogeography*, 52(1), 199–212. <https://doi.org/10.1111/jbi.15028>
- Mujetahid, A., Muin, A. V. F., & Wulan, S. (2023). Potential of aren (*Arenga pinnata*) in KHDTK Hasanuddin University Educational Forest, Rompegading Village, Cenrana District, Maros Regency. *IOP Conference Series: Earth and Environmental Science*, 1277, Article 012013. <https://doi.org/10.1088/1755-1315/1277/1/012013>
- Neubig, K. M., Whitten, W. M., Carlsward, B. S., Blanco, M. A., Endara, L., Williams, N. H., & Moore, M. (2009). Phylogenetic utility of ycf 1 in orchids : A plastid gene more variable than mat K. *Plant Systematics and Evolution*, 277, 75–84. <https://doi.org/10.1007/s00606-008-0105-0>
- Özbek, Ö. (2024). Molecular markers used to reveal genetic diversity and phylogenetic relationships in crop plants. *OBM Genetics*, 8(4), 1–25. <https://doi.org/10.21926/obm.genet.2404274>
- Perrier, X., & Jacquemoud-Collet, J. P. (2006). DARwin 6 software: version 6.0.8. Montpellier, France. <http://darwin.cirad.fr/darwin/DocDARwin6.pdf>.
- Rahmawati, A., Dinarti, D., Maskromo, I., Volkaert, H. A., & Sudarsono. (2022). Coconut diversity based on chloroplast

- single nucleotide amplified polymorphism (SNAP) and insertion-deletion (InDel) markers. *Biodiversitas*, 23(8), 4073–4081. <https://doi.org/10.13057/biodiv/d230827>
- Rinawati, D. Y., Refflinur, Dinarti, D., & Sudarsono. (2021). Genetic diversity of sugar palm (*Arenga pinnata*) derived from nine regions in Indonesia based on SSR markers. *Biodiversitas*, 22(9), 3749–3755. <https://doi.org/10.13057/biodiv/d220919>
- Roodt, D., Lohaus, R., Sterck, L., Swanepoel, R. L., Van De Peer, Y., & Mizrachi, E. (2017). Evidence for an ancient whole genome duplication in the cycad lineage. *PLoS ONE*, 12(9), Article e0184454. <https://doi.org/10.1371/journal.pone.0184454>
- Siswanto, T., Sianipar, N. F., Warnars, H. L. H. S., & Prabowo, H. (2022). Algorithm model determination of DNA primer design for the success PCR process. *International Journal of Emerging Technology and Advanced Engineering*, 12(7), 101–107. https://doi.org/10.46338/ijetae0722_11
- Teng, S., Michonova-Alexova, E., & Alexov, E. (2008). Approaches and resources for prediction of the effects of non-synonymous single nucleotide polymorphism on protein function and interactions. *Current Pharmaceutical Biotechnology*, 9(2), 123–133. <https://doi.org/10.2174/138920108783955164>
- Terryana, R. T., Nugroho, K., & Lestari, P. (2020). Genetic diversity of sugar palm populations from Cianjur and Banten revealed by simple sequence repeat (SSR) markers. *IOP Conference Series: Earth and Environmental Science*, 418, 012001. <https://doi.org/10.1088/1755-1315/418/1/012001>
- Wang, Y., Yu, J., Chen, Y. K., & Wang, Z. C. (2022). Complete chloroplast genome sequence of the endemic and endangered plant *Dendropanax oligodontus*: Genome structure, comparative and phylogenetic analysis. *Genes*, 13(11), 2028. <https://doi.org/10.3390/genes13112028>
- Wang, R., Yang, Y., Tian, H., Yi, H., Xu, L., Lv, Y., Ge, J., Zhao, Y., Wang, L., Zhou, S., & Wang, F. (2024). A scalable and robust chloroplast genotyping solution : development and application of snp and indel markers in the maize chloroplast genome. *Genes*, 15(3), Article 293. <https://doi.org/10.3390/genes15030293>
- Xu, W., Gong, C., Mai, P., Li, Z., Sun, B., & Li, T. (2024). Genetic diversity and population structure analysis of 418 tomato cultivars based on single nucleotide polymorphism markers. *Frontiers in Plant Science*, 15, 1445734. <https://doi.org/10.3389/fpls.2024.1445734>
- Yamada, K. D., Tomii, K., & Katoh, K. (2016). Application of the MAFFT sequence alignment program to large data: Re-examination of the usefulness of chained guide trees. *Bioinformatics*, 32(21), 3246–3251. <https://doi.org/10.1093/bioinformatics/btw412>
- Zhang, J., Yang, J., Lv, Y., Zhang, X., Xia, C., Zhao, H., & Wen, C. (2023). Genetic diversity analysis and variety identification using SSR and SNP markers in melon. *BMC Plant Biology*, 23(1), 56. <https://doi.org/10.1186/s12870-023-04056-7>
- Zuhud, E. A. M., Al Manar, P., Zuraida, & Hidayati, S. (2020). Potency and conservation of aren (*Arenga pinnata* (Wurmb) Merr.) in Meru Betiri National Park, East Java, Indonesia. *Jurnal Manajemen Hutan Tropika*, 26(3), 212–221. <https://doi.org/10.7226/jtjm.26.3.212>