BIODIVERSITAS
Volume 26, Number 11, November 2025
Pages: 5906-5919

ISSN: 1412-033X
E-ISSN: 2085-4722
DOI: 10.13057/biodiv/d261146

Revealing genetic markers and evolutionary insights within Piperaceae in chloroplast genome architecture of Indonesian *Piper betle*

ENDANG RAHMAT[1,♥], LAURA ANINDITA[1], NESTI FRONIKA SIANIPAR[1,2], KHOIRUNISSA ASSIDQI[1,2], YOUNGMIN KANG[3,4], KENNETH HAPPY[3,4], ADHITYO WICAKSONO[5]

[1]Department of Biotechnology, Faculty of Engineering, Universitas Bina Nusantara. Jl. Jalur Sutera Barat Kav. 21, Alam Sutera, Tangerang 15143, Banten, Indonesia. Tel.: +62-21-534-5830, ♥email: endang.rahmat@binus.ac.id
[2]Food Biotechnology Research Center, Universitas Bina Nusantara. Jl. Jalur Sutera Barat Kav. 21, Alam Sutera, Tangerang 15143, Banten, Indonesia
[3]University of Science and Technology. 217 Gajeong-ro Yuseong-gu, Daejeon 34054, Republic of Korea
[4]Herbal Medicine Resources Research Center, Korea Institute of Oriental Medicine. 111 Geonjae-Ro, Naju-Si, Jeollanam-Do 58245, Republic of Korea
[5]Genomik Solidaritas Indonesia Lab. Jl. Sultan Agung No. 29, South Jakarta 12980, Jakarta, Indonesia

Manuscript received: 9 September 2025. Revision accepted: 19 November 2025.

Abstract. *Rahmat E, Anindita L, Sianipar NF, Assidqi K, Kang Y, Happy K, Wicaksono A. 2025. Revealing genetic markers and evolutionary insights within Piperaceae in chloroplast genome architecture of Indonesian* Piper betle. *Biodiversitas 26: 5906-5919. Piper betle* is one of the most significant medicinal plants in Southeast Asian tradition, yet its genomic resources remain limited compared to other members of the genus. In this study, we report the first complete chloroplast genome of an Indonesian isolate of *P. betle*, providing a much-needed reference for future molecular work. The genome was assembled into 161,313 bp with a typical quadripartite structure consisting of an LSC (88,995 bp), an SSC (18,201 bp), and two nearly identical IRs (27,057 and 27,060 bp). A total of 113 genes were annotated, including 79 protein-coding, 30 tRNA, and 4 rRNA genes. Repeat analysis revealed 70 cpSSRs, markedly fewer than those described in other *Piper* species, along with 104 long repeats dominated by palindromic types. Sliding-window analysis detected variable regions in *ndhG* and *ndhI* in addition to the well-known hotspots *ycf1* and *rpl32-ndhF*, two novel loci that may serve as species-specific markers. Phylogenetic reconstruction clustered the Indonesian isolate with *P. hancei, P. nigrum*, and *P. kadsura*, yet comparison with a Chinese isolate indicated paraphyletic placement across two clades which may indicate a role of geographic divergence in shaping plastome evolution. The results of these genomic resources reveal reduced SSR content, novel divergence hotspots in *ndhG* and ndhI, and the paraphyletic relationship with the Chinese isolate: creating potential in marker development and evolutionary studies, while also laying the groundwork for future applications in conservation and biotechnology.

Keywords: Chloroplast genome, comparative genomics, marker development, Piperaceae, *Piper betle*

Abbreviations: cp: Chloroplast, cpSSRs: Chloroplast Simple Sequence Repeats, CTAB: Cetyltrimethylammonium Bromide, GC: Guanine-Cytosine, gDNA: genomic DNA, HAC: High Accuracy, IR: Inverted Repeat, LR: Long Repeat, LSC: Large Single-Copy, MISA: MIcroSatellite, ML: Maximum Likelihood, ONT: Oxford Nanopore Technologies, PCG: Protein-Coding Genes, rRNA: ribosomal RNA, SSC: Small Single-Copy, SSR: Simple Sequence Repeat, TBE: Tris-Borate-EDTA, tRNA: transfer RNA

INTRODUCTION

Piper betle L., locally referred to as "*daun sirih*" in Indonesia, is one of many traditionally medicinal and economically significant plants of the Piperaceae family, widely cultivated over many years across South and Southeast Asia (Nayaka et al. 2021; Biswas et al. 2022). While *P. betle* is also cultivated in other Southeast Asian areas such as India, Bangladesh, Malaysia, and Sri Lanka, the Indonesian variant plays a dominant role in local herbal medicine industries and ethnopharmacological practices (Widowati et al. 2020; Nayaka et al. 2021). The diverse applications of *P. betle* have given rise to a wide range of consumer products such as oral hygiene formulas, feminine hygiene pads, supplements, and household cleaners, all driven by the plant's documented antibacterial, antifungal, anti-inflammatory, and antioxidant properties (Nayaka et al. 2021; Biswas et al. 2022). Studies have reported compounds such as hydroxychavicol and chavicol, creating a strong efficacy against microbial pathogens and oxidative stress (Lubis et al. 2020). Despite that, the crucial data regarding its genomic resources have yet to be revealed, a gap that still stands.

Accessing these comprehensive resources, especially the complete chloroplast (cp) genome, is a breakthrough for reliable molecular authentication, cultivar discrimination, and traceability in supply chains (Biswas et al. 2022; Li et al. 2022). Cp genomes in *Piper* species have common quadripartite structures of around 160 kb and include conserved regions such as a Large Single-Copy (LSC), Small Single-Copy (SSC), and Inverted Repeats (IRs) (Geng et al. 2022; Li et al. 2022; Gaikwad et al. 2023). The number of genes found in the cp genome is around 113 unique genes, consisting of 79 protein-coding, 30 tRNA, and 4 rRNA, all of which have been the result of multiple *Piper* cp genomes (e.g., *P. sarmentosum*, *P. nigrum*, and *P. hancei*) (Zhang et al. 2021; Geng et al. 2022; Gaikwad et al. 2023). Nevertheless, no complete cp genome sequence exists for an Indonesian isolate of *P. betle* to date,

indicating a significant gap in genomic information for this paramount medicinal plant.

Cp genome data also facilitate phylogenetic reconstruction and comparative genomics in the Piperaceae taxa, which aids in the selection of guide markers and understanding of evolutionary relationships (Li et al. 2022; Yang et al. 2022). Comparative cp analysis across eight *Piper* species revealed highly conserved gene content and order, with repeat structures and SSRs varying, beneficial for species-level distinction (Li et al. 2022). Such molecular markers drawn from cp genomes are critical for bioauthentication, to prevent mislabeling or substitution of *P. betle* with other *Piper* species (Geng et al. 2022). Beyond taxonomy and quality control, cp genomes create a platform for cp engineering, targeted insertion of foreign genes with high expression, often for medicinal plants (Daniell et al. 2016). This has been proposed to enhance production of therapeutic metabolites to improve stress tolerance and confer pathogen resistance (Daniell et al. 2016; Biswas et al. 2022). For *P. betle*, harnessing cp transformation could encourage a higher yield in producing its compounds crucial for pharmaceutical, nutraceutical, and consumer goods markets (Biswas et al. 2022).

In this study, the first complete cp genome of an Indonesian isolate of *P. betle* is presented, sequenced using a long-read sequencing method using the Oxford Nanopore Technology (ONT). ONT can span entire chloroplast genomes in a few reads, simplifying assembly by crossing repetitive regions common in cp genomes and resolving large structural variations. Although many studies have focused on the phytochemical, pharmacological, and functional biomolecules of *P. betle*, the complete genetic information of this plant remains underrepresented. This report aims to present the cp genome's gene content and variable regions for future authentication or marker development. As this data serves as a strong foundation for genomic resources, it is expected to support future studies that focus on authentic sourcing, cultivar development, and biotechnology-driven innovation in *P. betle* product development and conservation.

MATERIALS AND METHODS

Plant material procurement and isolation of cpDNA

Fresh *Piper betle* leaves were obtained from herbal market located in Jakarta, Indonesia, that carries a large selection of traditional herbs and their derivative products. The collected specimens were deposited in Herbarium Bandungense, School of Life Science and Technology, Institut Teknologi Bandung (Voucher number FIPIA-DEP137).

Chloroplast DNA was extracted from ~100 mg of leaf tissue using the CTAB method (Doyle and Doyle 1990). Samples were snap-frozen in liquid nitrogen, ground with a pre-chilled mortar and pestle, and DNA was eluted in 50 µL buffer and stored at -20°C. Purity was assessed by NanoDrop (Thermo Scientific, USA) (A260/280, A260/230), and integrity confirmed by 1% agarose gel electrophoresis (TBE, GelRed). Concentration was quantified with the Qubit dsDNA HS Assay Kit (Thermo Scientific, USA), and fragment profiles evaluated using the Agilent 4150 TapeStation. Only intact, high-quality DNA was used for downstream analyses and preserved at -80°C.

Chloroplast genome sequencing

Genomic DNA samples were prepared for sequencing using the Oxford Nanopore Technologies Library Preparation Kit (SQK-NBD114-96). DNA fragments underwent end-repair with an end-prep enzyme mixture to produce 5' phosphorylated and 3' dA-tailed ends, followed by adapter ligation for ONT sequencing of the cp genome. Library concentration and quality were measured using a Qubit Fluorometer before loading onto the Flow Cell. Sequencing was performed on the PromethION platform (ONT, UK) with real-time monitoring via MinKNOW software until target data output was reached. Dorado (Oxford Nanopore Technologies 2025) was used for high-accuracy base calling.

Genome data assembly

Sequencing read quality was initially inspected using NanoPlot v1.44.1 (De Coster et al. 2018), which provided an overview of general read statistics. Reads were then aligned to a cp reference genome of *Piper hancei* (MZ_046380.1) through minimap2 v2.29 (Li 2018). The aligned reads were filtered and processed further using Samtools v1.22 (Danecek et al. 2021). For de novo genome assembly, Flye v2.9.6 (Kolmogorov et al. 2019) was utilized, followed by iterative correction and optimization with Canu v2.2 (Koren et al. 2017). Sequence polishing was then performed using Medaka v2.3 (Luan et al. 2024) to improve the base-level accuracy. The quality of the assembled chloroplast genome was assessed using two benchmarking tools: Quast v5.0.2 (Gurevich et al. 2013) for structural metrics, and Qualimap v2.3 (Okonechnikov et al. 2015) for further quality insights. Further gene annotation and visualization were conducted via the GeSeq module within the Chlorobox platform (https://chlorobox.mpimp-golm.mpg.de/). To verify taxonomic identity, the final assembled sequences were compared against the NCBI BLASTn database using its algorithm. Raw dataset is available at Zenodo Repository: https://doi.org/10.5281/zenodo.17042752.

Assessment of codon usage frequency

The chloroplast genome was annotated using the GeSeq platform, with Coding Sequences (CDSs) extracted via the GenBank Feature Extractor (Stothard 2000). The cp genome's codon preferences were identified through analyzing the codon usage metrics, such as Relative Synonymous Codon Usage (RSCU), which was computed using MEGA12 (64-bit) (Kumar et al. 2024). For clearer data interpretation, codon usage trends were then visualized through the ggplot2 package in R, providing a graphical representation to support downstream analysis.

Chloroplast simple sequence and long repeat analysis

Chloroplast Simple Sequence Repeats (cpSSRs) were identified using the MIcroSatellite (MISA) (Beier et al. 2017) web tool (https://webblast.ipk-gatersleben.de/misa/),

with thresholds set at ≥10 units for mononucleotide repeats, ≥5 for dinucleotide and trinucleotide repeats, and ≥3 for tetranucleotide, pentanucleotide, and hexanucleotide motifs. To complement this, the REPuter (Kurtz and Schleiermacher 1999) tool (https://bio.tools/REPuter) was used to detect sequences of Long Repeats (LRs), by applying a minimum repeat length of 30 bp and allowing a Hamming distance of 3. REPuter further categorized these repeats into four structural types: forward, reverse, palindromic, and complementary. This combined approach enabled the detection of both simple and complex repetitive elements within the cp genome. The parameters ensured sensitivity to minor mismatches while maintaining specificity in identifying true repeats.

Divergent hotspot identification

Whole chloroplast genomes from multiple *Piper* species were aligned against *P. betle* using MAFFT v7.526 (Katoh and Standley 2013) with default settings to ensure consistent multiple sequence alignment. The resulting alignment was subsequently analyzed in DnaSP v6 (Rozas et al. 2017) to evaluate nucleotide diversity (π) across the entire plastome. A sliding window approach was employed to detect regions of heightened sequence variation, applying a window size of 600 bp with a 200 bp step size. This window size was selected to balance sufficient sequence length for meaningful diversity calculations while maintaining adequate resolution to detect localized variation hotspots. The 200 bp step size ensures smooth transitions and prevents loss of information between adjacent windows. Divergent hotspots, which are potentially useful for species discrimination, were identified through this scan with comparative analysis focused on *P. betle* in relation to nine additional *Piper* taxa: *Piper hancei, Piper sarmentosum, Piper kadsura, Piper nigrum, Piper laetispicum, Piper longum, Piper bambusifolium, Piper bonii*, and *Piper boehmeriifolium.*

Phylogenetic analysis

Eleven complete cp genomes were aligned to construct the phylogenetic tree, including *P. betle* and nine additional *Piper* species: *P. hancei* (MZ_046380.1)*, P. sarmentosum* (MZ_958833.1)*, P. kadsura* (NC_027941.1)*, P. nigrum* (NC_034692.1)*, P. laetispicum* (NC_042254.1)*, P. longum* (NC_047247.1)*, P. bambusifolium* (NC_062129.1)*, P. boehmeriifolium* (NC_066431.1)*,* and *P. bonii* (NC_066434.1)*,* with *Astelia australiana* isolate 13384 (NC_045865.1) used as the outgroup. Multiple sequence alignment of these cp genomes utilized MAFFT v7.526 with default parameters to ensure consistent results across all sequences. The resulting alignment went through model selection analysis using jModelTest v2.1.10 (Darriba et al. 2012) to identify the best-fitting nucleotide substitution model for phylogenetic inference. Maximum Likelihood (ML) was then performed with MEGA 12 with 1000 bootstrap replicates to assess node support. Based on the jModelTest output, the General Time Reversible (GTR) model with a Gamma distribution and a proportion of invariant sites (G + I) was selected. This approach enabled robust reconstruction of the evolutionary relationships among *Piper* species based on their complete chloroplast genomes.

The GTR+G+I model was selected based on jModelTest results as it best captures the evolutionary dynamics of cp genomes by accounting for unequal substitution rates (GTR), rate variation across sites (G), and conserved regions (I) in protein-coding and structural RNA genes. One thousand bootstrap replicates were used to assess phylogenetic confidence, with values ≥70% considered reliable and ≥95% indicating strong clade support. *A. australiana* was chosen as the outgroup because it belongs to a distinct lineage outside Piperaceae while being sufficiently related to properly root the tree and polarize evolutionary changes within *Piper*.

RESULTS AND DISCUSSION

cp genome features of *Piper betle*

Oxford Nanopore sequencing of the *Piper betle* sample generated 32,532 total reads with an average read length of 3,492 bp (N50: 4,585 bp) and a mean quality score of Q15.4. The assembled chloroplast genome achieved a mean coverage depth of 752×, ensuring high sequencing accuracy and reliability. Quality assessment using NanoPlot v1.44.1 and QUAST v5.0.2 confirmed the completeness and structural integrity of the assembly, with no gaps or ambiguous bases detected in the final sequence.

A multitude of genomic features were successfully assembled through the ONT sequencing of this *P. betle* sample isolated from Indonesia, as portrayed in a circular DNA molecule using the GeSeq program (Figure 1). The annotation of multiple regions, genes, and nitrogenous base content was well included within the illustration. Within the entirety of this 161,313 bp plastome, its conserved quadripartite organization has shown the canonical content of a cp genome, structured by the largest spanning area known as the LSC region, followed by the SSC region, and completed by the two IR regions: IRa and IRb, highlighted in yellow in Figure 1. Size-wise, the regions are respectively measured to be 88,995 bp, 18,201 bp, and a near-similar value of 27,057 bp (IRa) and 27,060 (IRb) (Table 1). In terms of genomic architecture, that of *P. betle* is in line with previous cp genome reports from other *Piper* species, such as *P. sarmentosum* (Geng et al. 2022) and *P. nigrum* (Li et al. 2022), as they too exhibit quadripartite structures with comparable region sizes. The Guanine-Cytosine (GC) content, calculated to be 38.37%, also concludes that the sequencing process is indeed viable and high-quality, as it mirrors similar results within other *Piper* plastomes, which range from 38.2% to 38.4% (Wang et al. 2018a; Zhang et al. 2021; Li et al. 2022). In addition, the GC content in this plastome is displayed most in the IR regions (Figure 1).

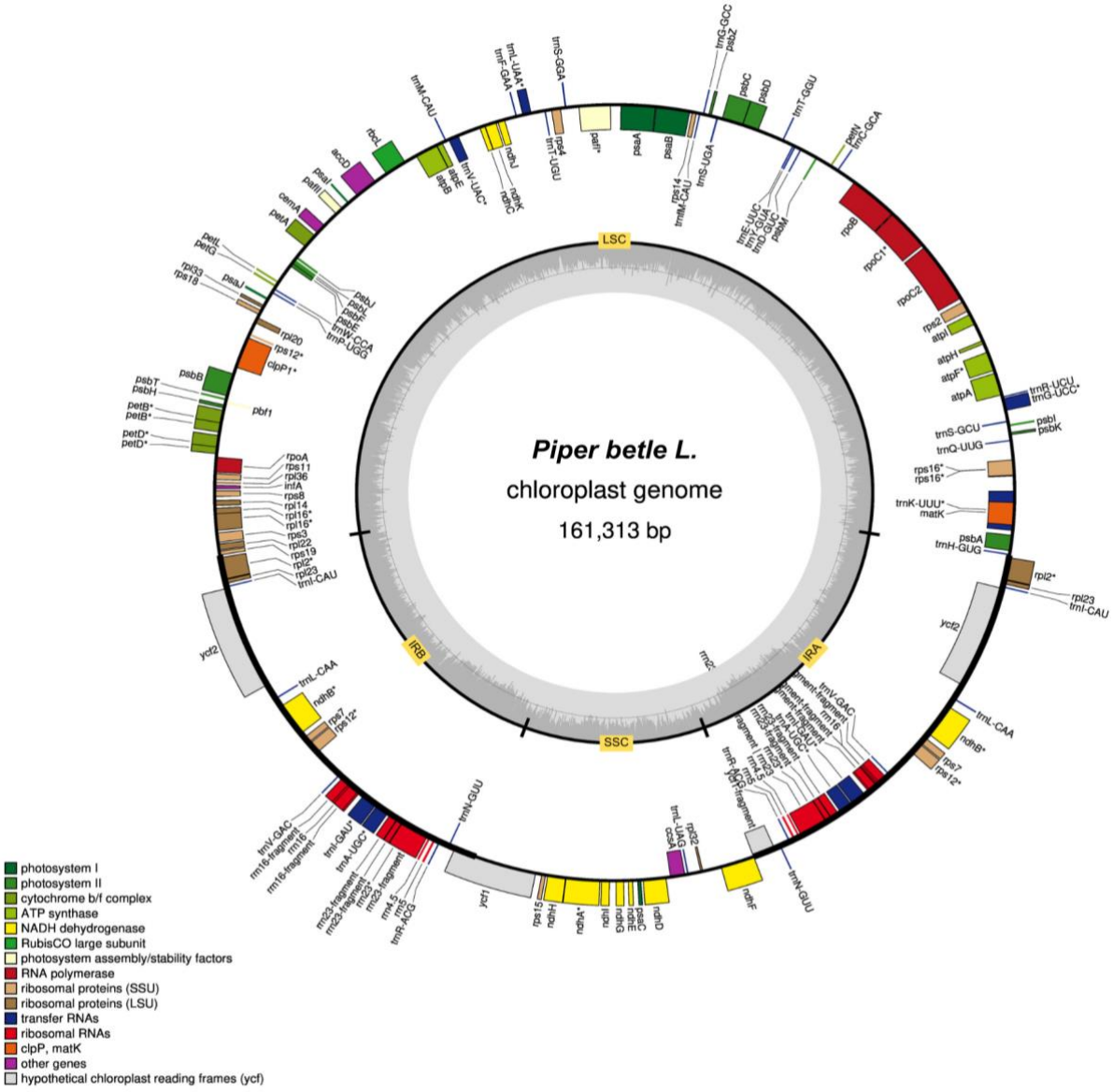

Figure 1. Circular representation of the *Piper betle* chloroplast genome generated using Chlorobox. The inside of the circle shows genes that are transcribed in the clockwise direction, while the outside circle consists of genes transcribed counterclockwise. The innermost ring indicates nucleotide pairing abundance, where darker shading reflects regions with higher GC composition. Genes are color-coded based on their functional classification

Table 1. Gene features with their respective sizes or content that are identified in *Piper betle*

Gene feature	Number
Size (bp)	161,313
LSC (bp)	88,995
IRa (bp)	27,057
SSC (bp)	18,201
IRb (bp)	27,060
Genes	113
CDS	79
tRNA	30
rRNA	4
GC%	38.37%
AT%	61.63%

Piper betle carries a total of 113 genes stored in its plastome, composed of multiple gene types, each with different functions, such as Protein-Coding Genes (PCGs), transfer RNA (tRNA) genes, and ribosomal RNA (rRNA) genes. Respectively, each group contains 79 genes, 30 genes, and 4 genes (Table 2). In Figure 1, some genes are duplicated due to their position in the IR regions, as the two regions harbor essential genes (e.g., tRNAs and rRNAs), allowing for enhanced gene expression through their duplicative character and plastome stabilization through homologous recombination-based repair (Wang et al. 2022; Krämer et al. 2024; Wu et al. 2024). Specifically, an amount of 17 genes were duplicated across IRa and IRb, comprising 5 PCGs (*rpl2, rpl23, rps7, rps12,* and *ycf2*), 1 NADH dehydrogenase gene (*ndhB*), 7 tRNAs (*trnI-CAU,*

trnI-CAA, trnV-GAC, trnI-GAU, trnA-UGC, trnR-ACG, and *trnN-GUU*), and 4 rRNAs (*rrn16, rrn23, rrn4.5,* and *rrn5*) (Table 2).

The *ycf1* gene exists as two incomplete copies: a longer yet truncated version located at the IRb/SSC junction, and a shorter, fragmented duplicate in the IRa region. This asymmetrical duplication reflects a lineage-specific IR expansion that only captured part of the *ycf1* coding sequence without retaining its full-length form, rendering truncated and fragmented genes often pseudogene-like and possibly nonfunctional as it lacks most of the original sequence, though common in angiosperm plastome structures (Vera-Paz et al. 2022). Another similar characteristic is found within the *rps12* gene, known to be trans-spliced: the 5' end (exon 1) found in the LSC region and the 3' end (exon 2 and 3) is duplicated into the IR regions (Wang et al. 2018a; Li et al. 2022). In total, 18 genes in the *P. betle* plastome were found to contain introns. PGCs carry 12 intron-containing genes: *clpP, ycf3*, and *rps12* each harbor two introns, while the remaining 10 harbor one (*atpF, rpoC1, rpl2, rpl16, rps16, ndhA, ndhB, petB, petD*). The last 6 genes are that of tRNAs that each carry a single intron: *trnA-UGC, trnI-GAU, trnG-UCC*, *trnV-UAC, trnL-UAA,* and *trnK-UUU*.

Synonymous codon preference and amino acid distribution

Codon usage bias analysis revealed that all 20 amino acids in the *P. betle* plastome are encoded by 64 total codons, and a stark preference toward A/U-ending codons, mirroring the AT-rich nature of angiosperms (Li et al. 2022; Gaikwad et al. 2023). Codons such as AUU (Ile), UAA (Leu), UCU (Ser), and AGA (Arg) show proof of the highest usage frequencies in comparison to other codons of same amino acid, with relative abundance values of 48.1%, 29.4%, 26.7%, and 28.7%, respectively (Figure 2). The highest codon use is of AUU (1,127 occurrences), followed by GAA (1,030 occurrences), and AAA (1,027 occurrences), whereas the least used codon is the stop codon UGA with only 22 appearances (Supplementary Table 1). Relative Synonymous Codon Usage (RSCU) analysis presented 31 codons with RSCU>1 (strong codon usage bias), predominantly A/U-ending (Supplementary Table 1). Meanwhile, 33 codons exhibited a weak bias (RSCU<1), with the two remaining codons, AUG and UGG, showing no bias (RSCU ≈ 1) (Supplementary Table 1). The complete codon counts and their corresponding RSCU values for *P. betle* are provided in Supplementary Table 1.

Table 2. Annotated genes within their grouped functions in the *Piper betle* plastome

Category	Gene group	Gene name
Replication-associated genes	Small subunit of ribosomal proteins	*rps11, rps12, rps14, rps15, rps16, rps18, rps19, rps2, rps3, rps4, rps7, rps8*
	Large subunit of ribosomal proteins	*rpl14, rpl16, rpl2, rpl20, rpl22, rpl23, rpl32, rpl33, rpl36*
	Ribosomal RNA genes	*rrn16, rrn23, rrn4.5, rrn5*
	DNA-dependent RNA polymerase	*rpoA, rpoB, rpoC1, rpoC2*
	Transfer RNA genes	*trnA-UGC, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnG-UCC, trnH-GUG, trnI-CAU, trnI-GAU, trnK-UUU, trnL-CAA, trnL-UAA, trnL-UAG, trnM-CAU, trnN-GUU, trnP-UGG, trnQ-UUG, trnR-ACG, trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC, trnV-UAC, trnW-CCA, trnY-GUA, trnfM-CAU*
Photosynthetic apparatus genes	Photosystem I	*psaA, psaB, psaC, psaI, psaJ*
	Photosystem II	*pbf1, psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbT, psbZ*
	Cytochrome b/f complex	*petA, petB, petD, petG, petL, petN*
	RUBISCO	*rbcL*
	Subunits of ATP synthase	*atpA, atpB, atpE, atpF, atpH, atpI*
	Subunit of NADH-dehydrogenase	*ndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*
Other genes	Protease	*clpP1*
	Maturase	*matK*
	Envelope membrane protein	*cemA*
	Translation initiation factor	*infA*
	C-type cytochrome synthesis gene	*ccsA*
	Subunit of Acetyl-CoA-carboxylase	*accD*
Genes of undetermined function	Conserved hypothetical chloroplast	*pafI, pafII, ycf1, ycf2*

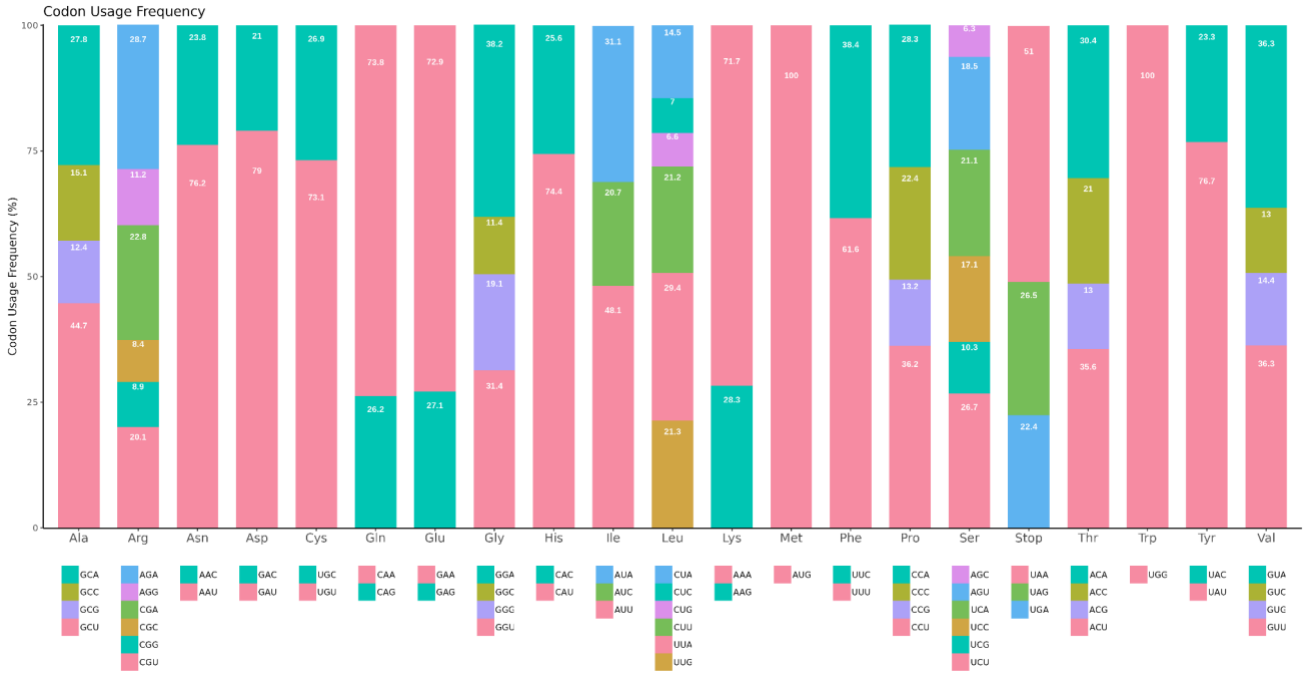

Figure 2. Codon usage frequency in the cp genome of *Piper bettle*. Each stacked bar represents an amino acid arranged alphabetically on the x-axis, with corresponding codon variants labeled below. The y-axis shows the total codon usage frequency (%) for each amino acid. Different colors within each bar represent individual synonymous codons encoding that amino acid, with their relative proportions indicating usage preference

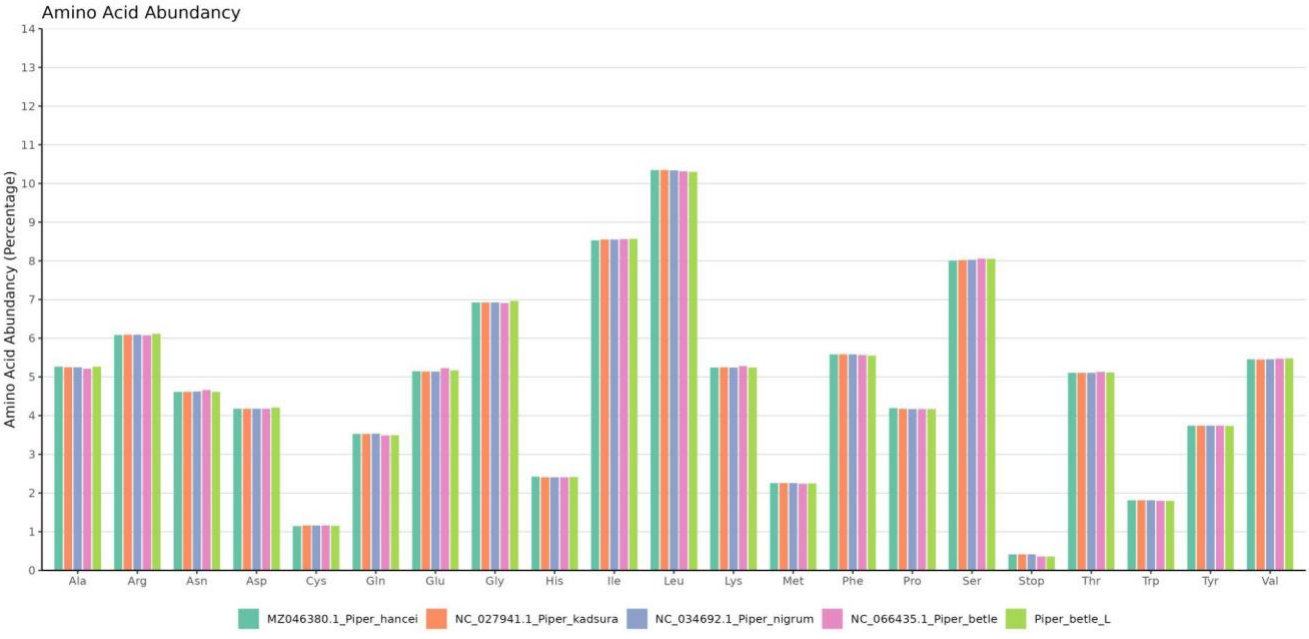

Figure 3. Amino acid abundance across five *Piper* species chloroplast genomes**.** Bars represent the relative percentage (y-axis) of each amino acid type (x-axis) for *Piper kancei* (dark green), *Piper kadsuva* (orange), *Piper nigrum* (blue), *Piper betle* (pink), and this study's *Piper betle* (light green). Leucine shows the highest abundance, while cysteine shows the lowest across all species

Across five cp genomes in the *Piper* genus, amino acid compositions were analyzed, and revealed a broad result of conserved motifs in relative abundance, with *P. betle* itself exhibiting profiles that are nearly identical to those of *P. nigrum, P. hancei,* and *P. kadsura* (Figure 3). In all species, leucine ranked as the most prevalent amino acid, accounting for approximately 10.34% of the total amino acid content in *P. betle.* This was closely followed by isoleucine (8.60%)

and serine (8.08%), both of which are encoded by codons that also displayed high usage frequencies (Figure 2, Supplementary Table 1). On the lower end of the distribution (excluding the stop codon), cysteine was the least abundant, only representing 1.13% of the total. The amino acid usage is relatively uniform across the five Piperaceae taxa, which points to strong conservation through evolution of cp-encoded proteins within the genus, also reflecting the shared selective pressures on plastid function and translation efficiency.

Repeat sequence analysis

Repetitive elements in the *P. betle* cp genome were characterized using MISA for SSRs and REPuter for LRs, following the standard approaches in cp genomics (Daniell et al. 2016; Li et al. 2022). Analysis for repeat motifs totaled 70 SSR motifs across the entire plastome. Mononucleotide repeats with poly-A/poly-T motifs predominate, especially A/T types, which account for most observations. The most frequent motif was A/T with 43 occurrences, with AT/AT coming in second (10 occurrences), and C/G repeats coming in third (4 occurrences) (Figure 4.A). Rare motifs such as AGAT/ATCT, AAG/CTT, and ACT/AGT were each represented, though only by one or two counts.

Types of SSRs are categorized by unit length into mono-, di-, tri-, and tetranucleotide repeats (Shukla et al. 2017). In this *P. betle* plastome, mononucleotides were overwhelmingly dominant (47 out of 70 counts), followed by dinucleotide (10 counts), tetranucleotide (7 counts), and trinucleotide repeats (2 counts) (Figure 4.B). Additionally, LRs were detected at several 104 repeats (≥ 30 bp) by REPuter. These were categorized as palindromic, forward, reverse, or complementary. Palindromic repeats constituted 78.8% in total, while forward repeats made up 18.3%, and reverse/complementary repeats were rare (1.9% and 1.0%, respectively) (Figure 4.C). The Indonesian variety of *P. betle* shows a remarkable reduction in Simple Sequence Repeats (SSRs) in its chloroplast genome, containing only 70 SSRs compared to 215 found in other *Piper* species (Gaikwad et al. 2023). This dramatic decrease indicates that this evolutionary lineage has undergone a specific reduction in repetitive DNA sequences, creating a distinct genetic fingerprint.

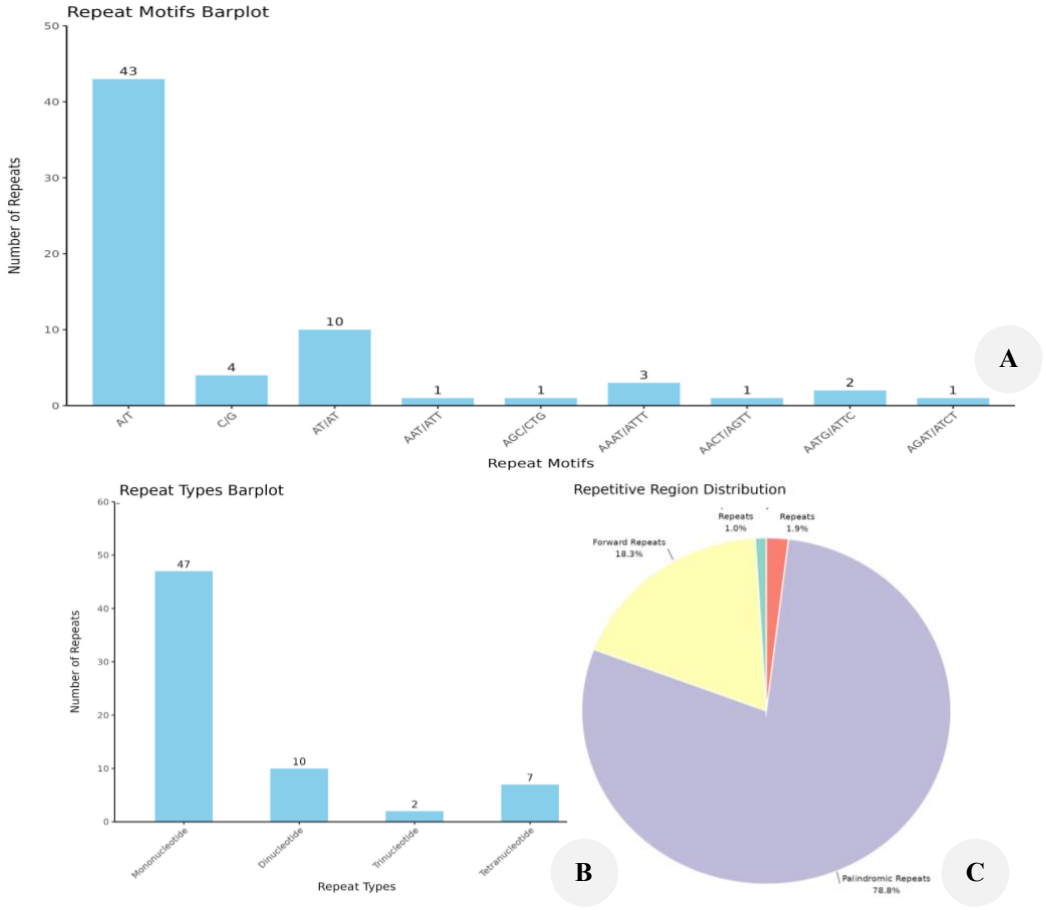

Figure 4. Summary of repeat sequence characteristics in the *Piper betle* cp genome: A. Frequency distribution of specific SSR motifs, where the x-axis represents different repeat motif types (mono-, di-, tri-, tetra-, penta-, and hexanucleotide) and the y-axis indicates the number of occurrences. Mononucleotide repeats are the most abundant. B. Classification of SSRs by repeat unit length, with the x-axis showing repeat types and y-axis showing counts. Forward, palindromic, reverse, and complementary repeats are compared. C. Pie chart showing the proportional distribution of repetitive regions, with Inter-Genic Spacer regions (IGS) comprising the largest proportion, followed by intron regions, Coding Sequences (CDS), and transfer RNA (tRNA) genes

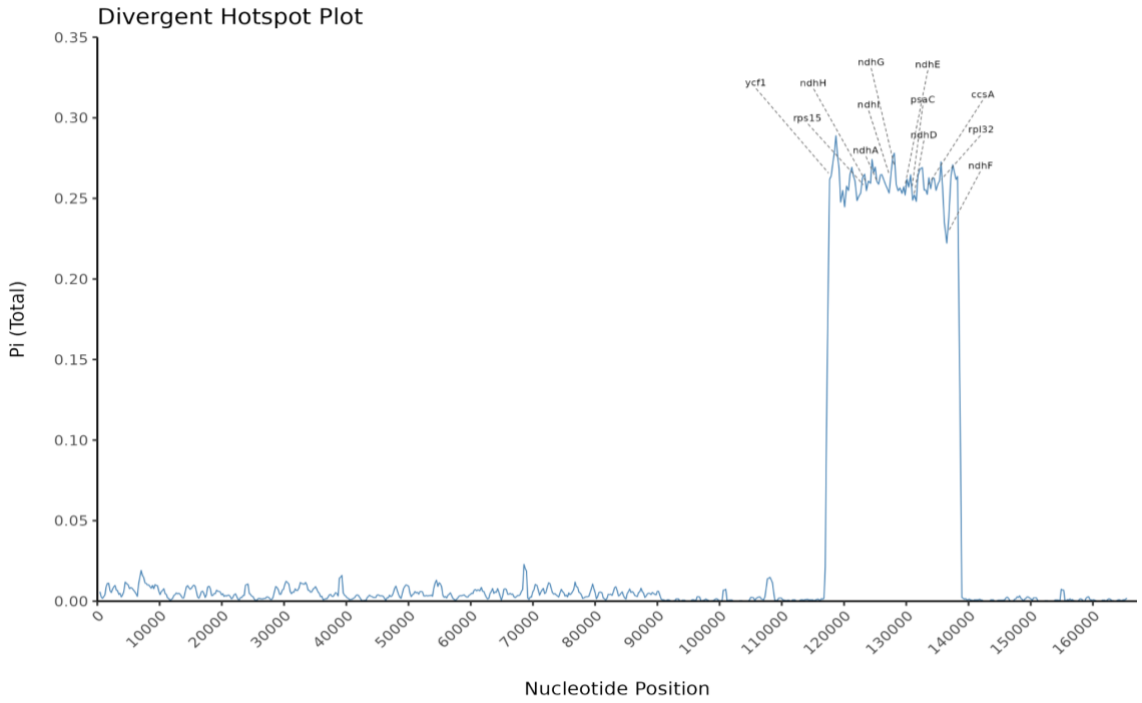

Figure 5. Sliding window analysis of *Piper* species' cp genomes from samples of *Piper kancei*, *Piper capavutianum*, *Piper kadsuva*, *Piper nigrum*, *Piper lasiodegem*, *Piper longum*, *Piper tompubolifum*, *Piper kadsuravorumm*, and *Piper betle*. The x-axis represents nucleotide positions along the chloroplast genome, while the y-axis shows nucleotide diversity (Pi value). Peaks indicate highly variable regions with labeled gene names, revealing divergent hotspots concentrated in specific genomic regions. Analysis parameters: Window length, 600 bp; step size, 200 bp

Divergent hotspot analysis

To identify the *P. betle* plastome's highly variable regions (known as hotspots) in contrast to other Piperaceae taxa, a sliding-window nucleotide diversity (π) analysis was conducted using DnaSP v6.11.01 with a window length of 600 bp and 200 bp in step size (Geng et al. 2022; Li et al. 2022). Firstly, nine additional *Piper* species were used for comparison and underwent multiple sequence alignment beforehand, conducted with MAFFT. The additional *Piper* species were taken from nine published *Piper* cp genomes, including: *P. hancei, P. sarmentosum, P. kadsura, P. nigrum, P. laetispicum, P. longum, P. bambusifolium, P. boehmeriifolium,* and *P. bonii*.

The sliding-window analysis' π graph revealed most of the genome remained below a π value of 0.05, except for a clear cluster of high-divergence peaks with elevated nucleotide diversity (π>0.20) located between positions ~117,000 to ~140,000, spanning the SSC region and extending into the IRb (Figure 5). Loci hotspots were identified, consisting of 12 genes: *ycf1, rps15, ndhH, ndhA, ndhI, ndhG, ndhD, ndhE, ndhF, rpl32, psaC,* and *ccsA.* These genes correspond to those overlapping windows with high π values. Exact per-gene averages were not calculated, and the values represent window-level diversity rather than gene-specific estimates. Though visually, the highest π values were found in *ndhG, ycf1,* and *ndhI* (Figure 5).

Through the identification of highly variable hotspots specific to the plastome of *P. betle* as a Piperaceae taxon, these loci offer strong potential for developing genetic markers. Especially, in understanding the contrast between conserved and rapidly evolving regions across the genus. As several of the highly available genes are associated with essential plastid functions, their divergence can also reflect some subtle evolutionary differences in how *Piper* species adapt to their environments. For a deeper insight into evolutionary relationships, a phylogenetic analysis was conducted using the complete cp genomes of selected Piperaceae taxa.

Phylogenetic analysis

To ensure a resolved evolutionary lineage of *P. betle* in the *Piper* genus, a phylogenetic tree construction was performed using nine complete cp genomes of other Piperaceae taxa and one outgroup (*A. australiana* isolate 13384) (Zou et al. 2024). Specifically, the maximum likelihood phylogenetic tree included the species of *P. kadsura, P. nigrum, P. hancei. P. laetispicum, P. longum, P. bambusifolium, P. sarmentosum, P. boehmerifolium,* and *P. bonii,* along with *P. betle* as this study's isolate (Figure 6). *A. australiana*, as the designated outgroup in Piperaceae phylogeny, serves as a well-characterized monocot distant from the ingroup, thus is reliable to root the overall tree (Amor et al. 2020). By using 1000 bootstrap replicates, this tree provided clades of high quality through the topology seen in Figure 6, which increases the reliability of the resulting relationships' value.

Maximum likelihood analysis placed the Indonesian isolate in a clade with *P. hancei, P. nigrum*, and *P. kadsura*. The bootstrap support for the *P. betle-P. hancei* branch was relatively weak (55), suggesting limited resolution at this node (Figure 6). Other *Piper* taxa formed a separate clade, with *P. bonii* and *P. boehmeriifolium* forming a strongly

supported subgroup (bootstrap = 100), and *P. sarmentosum* branching off earlier with weaker support (bootstrap = 47). These relationships reinforce the divergence trends seen in plastid hotspot loci like *ycf1, ndhG,* and *ndhI*, validating their utility in future marker design and evolutionary studies.

Discussion

This study presents the first complete chloroplast genome of an Indonesian *P. betle*, establishing a novel reference for comparative and evolutionary research in the genus (Nabilla et al. 2021; Li et al. 2022). As a result of ONT sequencing, which has emerged as a powerful modern platform for generating long-read and high-accuracy assemblies, the cp genomic toolkit of this *P. betle* isolate is significantly enriched and made available to the diverse and economically important genus of *Piper* (Li et al. 2022; Simbaqueba et al. 2024). Although plastid genomes of other *Piper* species (e.g., *P. hancei*, *P. nigrum)* have been characterized, *P. betle* remains underrepresented, despite being commonly used in Southeast Asian areas (Zhang et al. 2021; Gaikwad et al. 2023). By providing this complete plastome, the present study creates an essential reference for *P. betle*'s genomic architecture, codon preferences, repeat elements, sequence variability, and phylogenetic placement. Specifically, contrasting profiles in comparison to other *Piper* species, such as divergent hotspots and species-specific characteristics, are highly potent for the current need of *P. betle* molecular markers and DNA barcoding (Wang et al. 2018b; Mondal 2021).

The overall structure of *P. betle* plastome is revealed to maintain the standard quadripartite structure of angiosperm cp genomes, which is akin to other *Piper* species such as *P. hancei, P. kadsura, P. sarmentosum,* and *P. nigrum,* as this reinforces a consistent structure within the genus (Lee et al. 2015; Zhang et al. 2021; Geng et al. 2022; Gaikwad et al. 2023). Its genome size (161,313 kb) spans the entire length of the cp genome, organized into the LSC, SSC, and the a and b IR regions, and with consistent *Piper* gene content, which all falls into the conserved range of plastome features of the *Piper* species (Li et al. 2022; Gaikwad et al. 2023). Nucleotide pairing percentages also show the total GC content to be 38.37%, congruent with the 38.2-38.4% range previously calculated in other Piperaceae taxa, simultaneously characterizing the consistent AT-dominant nature of the genus and angiosperms in general (Wang et al. 2018a; Li et al. 2022).

Within the cp genome, several conserved coding regions are known for universal barcoding of land plants, which reinforce the structural integrity of the plastome, as they code proteins for essential functions (Nagar and Hahsler 2013). They become highly reliable in genera for species authentication and taxonomic discrimination, with regions in Piperaceae taxa being *rbcL, matK, ycf1, ycf2, ndhA-K*, *rps* and *rpl* genes, all found in this study's Indonesian isolate of *P. betle* (Chaveerach et al. 2016; Jiang et al. 2022). Additionally, the *rps12* gene is trans-spliced, with a 5′ exon located in the LSC and duplicated 3′ exons in the IRs; this arrangement is notably the only trans-spliced gene across Piperales plastomes, another conserved gene profile, specific to a splicing architecture within the lineage (Jost and Wanke 2024). These unique genes confirm the completeness of this *P. betle* plastome assembly and its value as a genomic resource for DNA barcode development and other comparative genomics in its related taxa.

Extending beyond structural conservation, the analysis of codon usage patterns contributes to the molecular evolution of the *P. betle* plastome by revealing evolutionary mutational pressures and how they affect optimal translational selection (Li et al. 2022; Yang et al. 2023; Shi et al. 2024). Codon bias itself is a product of natural selection, with codons that are most efficiently translated showing prevalence in gene expression (Quax et al. 2015; Guan et al. 2018). As seen in *P. nigrum*, codon usage is biased toward A/U-ending codons, indicating an AT-rich *Piper* plastome (Gaikwad et al. 2023).

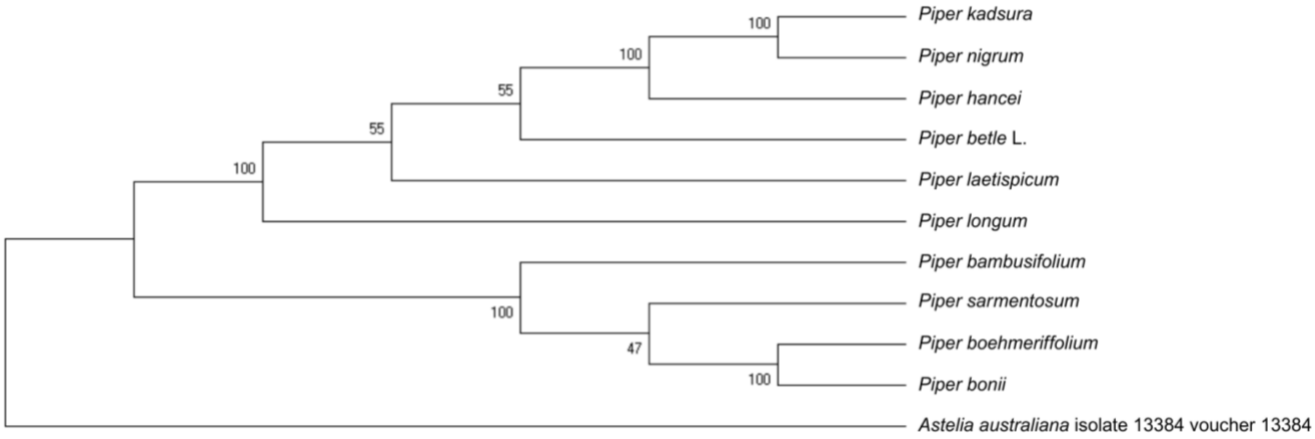

Figure 6. Phylogenetic tree of *Piper betle* with nine other *Piper* species, using an *Astelia australiana* sample as an outgroup. Bootstrap values are indicated at each node. The *Piper betle-Piper hancei* branch shows low support (55), reflecting limited phylogenetic resolution. Phylogeny was visualized by Maximum Likelihood (ML) analysis in MEGA with 1000 bootstrap replicates

In *P. betle*, the same pattern was identified, where AT preference manifests strongly, proving a mutational bias and shared translational constraints between *P. betle* and other *Piper* species (Gaikwad et al. 2023). Functional biases in codon usage, such as in photosynthesis-related genes versus housekeeping genes, were documented in other angiosperms, as the genes with higher expression or essential roles often show optimized codon usage (Zhang et al. 2018; Yang et al. 2023). This analysis would be valuable in future studies to assess whether or not *P. betle* exhibits similar trends, to potentially indicate selective pressures to enhance the plastid proteins' translation efficiency.

Amino acid composition across several *Piper* plastomes portrays a strong evolutionary conservation, strengthening translation efficiency. The highest similarity of amino acid composition across five *Piper* species was found in *P. hancei, P. kadsura,* and *P. nigrum* (Figure 3). The order of the most prevalent amino acids is leucine (10.34%), isoleucine (8.60%), and serine (8.08%), with isoleucine encoded by the highest usage frequency codon, AUU. This pattern is also mirrored in other angiosperm plastomes, such as *Asparagus officinalis* (prevalence in Leu, ~10.63%), with dominance in leucine and isoleucine, while cysteine remains the least frequent, confirming this characteristic found in *P. betle* (Sheng 2022).

Overwhelmingly dominated by A/T mononucleotides, *P. betle* finely reflects a bias for AT dominance in the Piperaceae taxa as angiosperms (Li et al. 2022; Gaikwad et al. 2023). Unlike the 215 SSRs reported in other *Piper* plastomes, the Indonesian isolate contains only 70 SSRs (Gaikwad et al. 2023). This striking contraction suggests a lineage-specific reduction in repeat load, creating a distinctive genomic signature. Such a reduction may imply lower levels of cpSSR-driven variability, but it also provides a unique marker set for identifying *P. betle* within the genus (Xu et al. 2025). Categorization by repeat unit showed a consistent distribution across *Piper* species: prevalence of common and simpler repeats, with less rare and higher-order motifs, providing much for molecular marker development (Li et al. 2022). Lastly, the number of long repeats in *P. betle* (104) and the ubiquity of palindromic repeats (78.8%) also show mirrored results in other *Piper* plastomes, showing potential for facilitating plastome structural variation, possibly through mechanisms like recombination or replication slippage (Li et al. 2022; Gaikwad et al. 2023).

Hypervariable loci were identified across multiple *Piper* species, highly concentrated within the SSC region, seen across other *Piper* plastomes where sequence divergence is unevenly distributed (Zhang et al. 2021; Li et al. 2022). The most prominent hotspots included *ycf1* and the *rpl32-ndhF* intergenic spacer, consistent with previous reports in *P. nigrum* and *P. hancei* (Geng et al. 2022; Gaikwad et al. 2023). In addition to these well-known loci, the Indonesian plastome displayed divergence in *ndhG* and *ndhI*. These two genes have not previously been described as hotspots in *Piper*, marking them as novel features of this isolate (Li et al. 2022; Gaikwad et al. 2023). As both genes encode subunits of the NDH complex, which mediates cyclic electron transport, their divergence could indicate functional adaptations and provide valuable candidates for species-specific barcoding (Hualing 2022). The presence of these two regions suggests a lineage-specific divergence unique to *P. betel*, which, although less suited for genus-wide barcoding, creates potential for species-specific molecular markers in use of authentication, population studies, and cultivar discrimination (Assis 2018; Li et al. 2020). Precise nucleotide diversity (π) values per gene were not provided in the dataset because many of the detected hotspots either occurred in those intergenic regions or spanned multiple loci at once, making it impossible to assign variability scores to a single gene.

Further analysis of phylogeny reveals the position of this Indonesian *P. betle* isolate within the Piperaceae taxa (Li et al. 2022; Gaikwad et al. 2023). In the phylogenetic tree (Figure 6), *P. betle* clustered within a supported clade including *P. kadsura, P. nigrum,* and *P. hancei*, rooted with the help of *A. australiana* as the outgroup. The relationship between *P. betle* and *P. hancei* carried a weak value of 55, lower than the robust values seen between *P. nigrum* and *P. kadsura* (100) and their clade with *P. hancei* (100), likely due to limited divergence and reduced phylogenetic signal at shallow nodes (Zapelloni et al. 2021; Dong et al. 2022). Nevertheless, the close association is consistent with coding region conservation and affirms *P. betle*'s placement in this *Piper* clade.

Importantly, this study reports the first plastome from an Indonesian accession of *P. betle*, whereas Li et al. (2022) sequenced a Chinese isolate cultivated in Hainan, which complements the geographical difference in the data. The addition of the Chinese isolate is seen in a secondary tree in Figure 7 as the two isolates cluster as sister taxa, though embedded within different higher-order branches of the tree. The Indonesian plastome is nested within the clade containing *P. kadsura, P. nigrum,* and *P. hancei,* whereas the Chinese isolate falls into the adjacent clade containing *P. sarmentosum, P. bambusifolium,* and *P. boehmeriifolium*. Despite their sister relationship, this topology renders *P. betle* paraphyletic across the two clades. This pattern is not uncommon in plastid phylogenies, and may be explained by incomplete lineage sorting, plastid capture via hybridization, or region-specific divergence events (Gernandt et al. 2018; Yang et al. 2018).

Taken together, this plastome exhibits three distinctive features: a marked reduction in SSR number, unique divergence hotspots in *ndhG* and *ndhI*, and a paraphyletic relationship to the Chinese isolate. These findings are the evolutionary distinctiveness of the pioneering Indonesian *P. betle* plastome and provide a foundation for species authentication, conservation, and future comparative studies. Future work should prioritize expanded sampling of diverse geographical accessions to map the intraspecific plastome variation, and attempts in cp transformation for high-level expression of transgenes, multi-gene stacking, also transgene containment via maternal inheritance (Daniell et al. 2005). Tailored transformation vectors could then be designed for *P. betle*, for future metabolic engineering, marker development, and even enhanced biosynthetic pathways for its medicinal and commercial value.

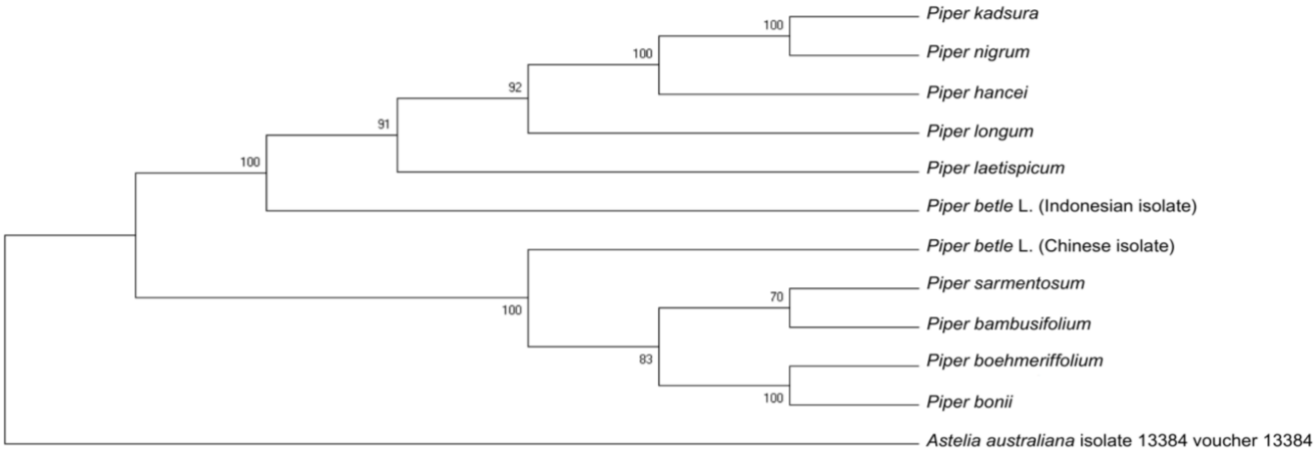

Figure 7. Phylogenetic tree of *Piper betle* isolates and related *Piper* species with *Astelia australiana* as outgroup. Maximum Likelihood tree generated using MEGA software with 1000 bootstrap replicates (values shown at nodes). The tree includes Indonesian and Chinese isolates of *Piper betle* along with nine other *Piper* species, showing their evolutionary relationships and genetic distances

In conclusion, this work reports the first complete chloroplast genome of an Indonesian *P. betle*, providing a much-needed reference for comparative and applied research. The plastome retains the typical quadripartite structure of Piperaceae but shows distinctive features, including reduced SSR content and novel divergence hotspots in *ndhG* and *ndhI*. Phylogenetic analysis placed the Indonesian isolate with *P. hancei, P. nigrum*, and *P. kadsura*, while comparison with the Chinese isolate revealed a paraphyletic relationship, proving the role of geographic divergence in plastome evolution. Through this data, species authentication, barcoding, and conservation are now achievable, while also informing evolutionary studies within the genus. Future work should prioritize broader geographic sampling to capture intraspecific variation and refine phylogeographic patterns. In parallel, the complete plastome opens new opportunities for chloroplast engineering, enabling transgene expression, multi-gene stacking, and metabolic pathway enhancement to support both medicinal and commercial applications of *P. betle*.

ACKNOWLEDGEMENTS

This work was supported by Bina Nusantara University as a part of Bina Nusantara University's PIB Terapan with contract number: 208A/VRRTT/X/2025. ER: conceptualized the research, acquired funding, developed the methodology, provided viable resources, conducted data analysis, and refined the manuscript; LA: managed collection and analysis of the data, and wrote the manuscript; NFS, KA, YK, HK, and AW: took part in manuscript editing and finalization. All the authors agreed on the accountability, accuracy, and integrity of the work. All the authors have approved the final version of the manuscript.

REFERENCES

Amor MD, Holmes GD, James EA. 2020. Characterization of the complete plastid genome of *Astelia australiana* (J. H. Willis) L. B. Moore (Asteliaceae, Asparagales). Mitochondrial DNA B Resour 5 (1): 656-657. DOI: 10.1080/23802359.2019.1711233.

Assis R. 2018. Lineage-specific expression divergence in grasses is associated with male reproduction, host-pathogen defense, and domestication. Genome Biol Evol 11 (1): 207-219. DOI: 10.1093/gbe/evy245.

Beier S, Thiel T, Münch T, Scholz U, Mascher M. 2017. MISA-web: A web server for microsatellite prediction. Bioinformatics 33 (16): 2583-2585. DOI: 10.1093/bioinformatics/btx198.

Biswas P, Anand U, Saha SC, Kant N, Mishra T, Masih H, Bar A, Pandey DK, Jha NK, Majumder M, Das N, Gadekar VS, Shekhawat MS, Kumar M, Radha, Proćków J, de la Lastra JMP, Dey A. 2022. Betelvine (*Piper betle* L.): A comprehensive insight into its ethnopharmacology, phytochemistry, and pharmacological, biomedical and therapeutic attributes. J Cell Mol Med 26 (11): 3083-3119. DOI: 10.1111/jcmm.17323.

Chaveerach A, Tanee T, Sanubol A, Monkheang P, Sudmoon R. 2016. Efficient DNA barcode regions for classifying *Piper* species (Piperaceae). PhytoKeys 70: 1-10. DOI: 10.3897/phytokeys.70.6766.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and BCFtools. GigaScience 10 (2): giab008. DOI: 10.1093/gigascience/giab008.

Daniell H, Kumar S, Dufourmantel N. 2005. Breakthrough in chloroplast genetic engineering of agronomically important crops. Trends Biotechnol 23 (5): 238-245. DOI: 10.1016/j.tibtech.2005.03.008.

Daniell H, Lin C-S, Yu M, Chang W-J. 2016. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. Genome Biol 17 (1): 134-144. DOI: 10.1186/s13059-016-1004-2.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: More models, new heuristics and parallel computing. Nat Methods 9 (8): 772-775. DOI: 10.1038/nmeth.2109.

De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: Visualizing and processing long-read sequencing data. Bioinformatics 34: 2666-2669. DOI: 10.1093/bioinformatics/bty149.

Dong W, Li E, Liu Y, Xu C, Wang Y, Liu K, Cui X, Sun J, Suo Z, Zhang Z, Wen J, Zhou S. 2022. Phylogenomic approaches untangle early divergences and complex diversifications of the olive plant family. BMC Biol 20 (1): 92. DOI: DOI: 10.1186/s12915-022-01297-0.

Doyle JJ, Doyle JL. 1990. Isolation of plant DNA from fresh tissue. Focus 12 (1): 13-15.

Gaikwad AB, Kaila T, Maurya A, Kumari R, Rangan P, Wankhede DP, Bhat KV. 2023. The chloroplast genome of black pepper (*Piper*

nigrum L.) and its comparative analysis with related *Piper* species. Front Plant Sci 13: 1095781. DOI: 10.3389/fpls.2022.1095781.
Geng X, Zhu Y, Ren Z, Chen R, Liu Q. 2022. The complete chloroplast genome of *Piper sarmentosum* Roxburgh, 1820 (Piperaceae). Mitochondrial DNA B Resour 7 (5): 854-855. DOI: 10.1080/23802359.2022.2074805.
Gernandt DS, Dugua XA, Vázquez-Lobo A, Willyard A, Letelier AM, de la Rosa JAP, Piñero D, Liston A. 2018. Multi-locus phylogenetics, lineage sorting, and reticulation in *Pinus* subsection Australes. Am J Bot 105 (4): 711-725. DOI: 10.1002/ajb2.1052.
Guan D-L, Ma L-B, Khan MS, Zhang X-X, Xu S-Q, Xie J-Y. 2018. Analysis of codon usage patterns in *Hirudinaria manillensis* reveals a preference for GC-ending codons caused by dominant selection constraints. BMC Genomics 19 (1): 542. DOI: 10.1186/s12864-018-4937-x.
Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: Quality Assessment Tool for genome assemblies. Bioinformatics 29 (8): 1072-1075. DOI: 10.1093/bioinformatics/btt086.
Hualing M. 2022. Cyanobacterial NDH-1 complexes. Front Microbiol 13: 933160. DOI: 10.3389/fmicb.2022.933160.
Jiang S, Chen F, Qin P, Xie H, Peng G, Li Y, Guo X. 2022. The specific DNA barcodes based on chloroplast genes for species identification of Theaceae plants. Physiol Mol Biol Plants 28 (4): 837-848. DOI: 10.1007/s12298-022-01175-7.
Jost M, Wanke S. 2024. A comparative analysis of plastome evolution in autotrophic Piperales. Am J Bot 111: e16300. DOI: 10.1002/ajb2.16300.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol 30 (4): 772-780. DOI: 10.1093/molbev/mst010.
Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37 (5): 540-546. DOI: 10.1038/s41587-019-0072-8.
Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27 (5): 722-736. DOI: 10.1101/gr.215087.116.
Krämer C, Boehm CR, Liu J, Ting MKY, Hertle AP, Forner J, Ruf S, Schöttler MA, Zoschke R, Bock R. 2024. Removal of the large inverted repeat from the plastid genome reveals gene dosage effects and leads to increased genome copy number. Nat Plants 10 (6): 923-935. DOI: 10.1038/s41477-024-01709-9.
Kumar S, Stecher G, Suleski M, Sanderford M, Sharma S, Tamura K. 2024. MEGA12: Molecular Evolutionary Genetic Analysis version 12 for adaptive and green computing. Mol Biol Evol 41 (12): msae263. DOI: 10.1093/molbev/msae263.
Kurtz S, Schleiermacher C. 1999. REPuter: Fast computation of maximal repeats in complete genomes. Bioinformatics. https://bio.tools/REPuter.
Lee J-H, Choi I-S, Choi B-H, Yang S, Choi G. 2015. The complete plastid genome of *Piper kadsura* (Piperaceae), an East Asian woody vine. Mitochondrial DNA A DNA Mapp Seq Anal 27 (5): 3555-3556. DOI: 10.3109/19401736.2015.1074216.
Li H. 2018. Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics 34: 3094-3100. DOI: 10.1093/bioinformatics/bty191.
Li J, Fan R, Xu J, Hu L, Su F, Hao C. 2022. Comparative analysis of the chloroplast genomes of eight *Piper* species and insights into the utilization of structural variation in phylogenetic analysis. Front Genet 13: 925252. DOI: 10.3389/fgene.2022.925252.
Li Q-J, Su N, Zhang L, Tong R-C, Zhang X-H, Wang J-R, Chang Z-Y, Zhao L, Potter D. 2020. Chloroplast genomes elucidate diversity, phylogeny, and taxonomy of *Pulsatilla* (Ranunculaceae). Sci Rep 10 (1): 19781. DOI: 10.1038/s41598-020-76699-7.
Luan T, Commichaux S, Hoffmann M, Jayeola V, Jang JH, Pop M, Rand H, Luo Y. 2024. Benchmarking short and long read polishing tools for nanopore assemblies: Achieving near-perfect genomes for outbreak isolates. BMC Genomics 25: 679. DOI: 10.1186/s12864-024-10582-x.
Lubis RR, Marlisa, Wahyuni DD. 2020. Antibacterial activity of betle leaf (*Piper betle* L.) extract on inhibiting *Staphylococcus aureus* in conjunctivitis patient. Am J Clin Exp Immunol 9 (1): 1-5.
Mondal B. 2021. Conversion of metabolomic data to genomic marker for genetic characterization of *Piper betle* L. chemotypes: A review. Agric Rev R-2118: 1-8. DOI: 10.18805/ag.r-2118.
Nabilla S, Safira UM, Puspita PJ, Subositi D, Maruzy A, Artika IM. 2021. Genetic diversity analysis of *Piper betle* from eight accessions of Indonesia based on SRAP markers. Biodiversitas 22 (8): 3401-3408. DOI: 10.13057/biodiv/d220837.
Nagar A, Hahsler M. 2013. Fast discovery and visualization of conserved regions in DNA sequences using quasi-alignment. BMC Bioinformatics 14: S2. DOI: 10.1186/1471-2105-14-s11-s2.
Nayaka NMDMW, Sasadara MMV, Sanjaya DA, Yuda PESK, Dewi NLKAA, Cahyaningsih E, Hartati R. 2021. *Piper betle* (L.): Recent review of antibacterial and antifungal properties, safety profiles, and commercial applications. Molecules 26 (8): 2321. DOI: 10.3390/molecules26082321.
Okonechnikov K, Conesa A, García-Alcalde F. 2015. Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32 (2): 292-294. DOI: 10.1093/bioinformatics/btv566.
Oxford Nanopore Technologies. 2025. Nanopore sequencing platform accuracy. Oxford Nanopore Technologies. https://nanoporetech.com/platform/accuracy.
Quax TEF, Claassens NJ, Söll D, Van der Oost J. 2015. Codon bias as a means to fine-tune gene expression. Mol Cell 59 (2): 149-161. DOI: 10.1016/j.molcel.2015.05.035.
Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol 34 (12): 3299-3302. DOI: 10.1093/molbev/msx248.
Sheng W. 2022. The entire chloroplast genome sequence of *Asparagus cochinchinensis* and genetic comparison to *Asparagus* species. Open Life Sci 17 (1): 893-906. DOI: 10.1515/biol-2022-0098.
Shi N, Yuan Y, Huang R, Wen G. 2024. Analysis of codon usage patterns in complete plastomes of four medicinal *Polygonatum* species (Asparagaceae). Front Genet 15: 1401013. DOI: 10.3389/fgene.2024.1401013.
Shukla N, Kuntal H, Shanker A, Sharma SN. 2017. Mining and analysis of simple sequence repeats in the chloroplast genomes of genus *Vigna*. Biotechnol Res Innov 2 (1): 9-18. DOI: 10.1016/j.biori.2018.08.001.
Simbaqueba J, Garzón-Martínez GA, Castaño N. 2024. New variants in the chloroplast genome sequence of two Colombian individuals of the cedar timber species (*Cedrela odorata* L.), using long-read Oxford Nanopore technology. Intl J Plant Biol 15 (3): 865-877. DOI: 10.3390/ijpb15030062.
Stothard PM. 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques 28 (6): 1102-1104. DOI: 10.2144/00286ir01.
Vera-Paz SI, Díaz DDDC, Jost M, Wanke S, Rossado AJ, Hernández-Gutiérrez R, Salazar GA, Magallón S, Gouda EJ, Ramírez-Morillo IM, Donadío S, Mendoza CG. 2022. New plastome structural rearrangements discovered in core Tillandsioideae (Bromeliaceae) support recently adopted taxonomy. Front Plant Sci 13: 924922. DOI: 10.3389/fpls.2022.924922.
Wang M-T, Wang J-H, Zhao K-K, Zhu Z-X, Wang H-F. 2018a. Complete plastome sequence of *Piper laetispicum* (Piperaceae): An endemic plant species in South China. Mitochondrial DNA B Resour 3 (2): 1035-1036. DOI: 10.1080/23802359.2018.1511850.
Wang W, Chen S, Zhang X. 2018b. Whole-genome comparison reveals divergent IR borders and mutation hotspots in chloroplast genomes of herbaceous bamboos (Bambusoideae: Olyreae). Molecules 23 (7): 1537. DOI: 10.3390/molecules23071537.
Wang Z-X, Wang D-J, Yi T-S. 2022. Does IR-loss promote plastome structural variation and sequence evolution? Front Plant Sci 13: 888049. DOI: 10.3389/fpls.2022.888049.
Widowati L, Handayani L, Mujahid R. 2020. The use of betel (*Piper betle*) leaves for maintaining the health of women and children at various ethnic groups in Indonesia. Nusantara Bioscience 12 (2): 120-126. DOI: 10.13057/nusbiosci/n120206.
Wu H, Li D-Z, Ma P-F. 2024. Unprecedented variation pattern of plastid genomes and the potential role in adaptive evolution in *Poales*. BMC Biol 22 (1): 97. DOI: 10.1186/s12915-024-01890-5.
Xu Y, Chen S, Chen S, Wei X, Shang H, Zhang Q, Zhang J. 2025. Genome-wide development of SSR molecular markers for modern sugarcane cultivars. Front Plant Sci 16: 1573967. DOI: 10.3389/fpls.2025.1573967.
Yang L, Abduraimov O, Tojibaev K, Shomurodov K, Zhang Y-M, Li W-J. 2022. Analysis of complete chloroplast genome sequences and insight into the phylogenetic relationships of *Ferula* L. BMC Genomics 23 (1): 643. DOI: 10.1186/s12864-022-08868-z.
Yang Q, Xin C, Xiao Q-S, Lin Y-T, Li L, Zhao J-L. 2023. Codon usage bias in chloroplast genes implicate adaptive evolution of four ginger species. Front Plant Sci 14: 1304264. DOI: 10.3389/fpls.2023.1304264.
Yang Y, Zhu J, Feng L, Zhou T, Bai G, Yang J, Zhao G. 2018. Plastid genome comparative and phylogenetic analyses of the key genera in

Fagaceae: Highlighting the effect of codon composition bias in phylogenetic inference. Front Plant Sci 9: 82. DOI: 10.3389/fpls.2018.00082.

Zapelloni F, Pons J, Jurado-Rivera JA, Jaume D, Juan C. 2021. Phylogenomics of the *Hyalella* amphipod species-flock of the Andean Altiplano. Sci Rep 11 (1): 366. DOI: 10.1038/s41598-020-79620-4.

Zhang L, Hu X, Liu M. 2021. Complete chloroplast genome sequences of the medicinal plant *Piper hancei*. Mitochondrial DNA B Resour 6 (9): 2775-2776. DOI: 10.1080/23802359.2021.1967217.

Zhang R, Zhang L, Wang W, Zhang Z, Du H, Qu Z, Li X-Q, Xiang H. 2018. Differences in codon usage bias between photosynthesis-related genes and genetic system-related genes of chloroplast genomes in cultivated and wild *Solanum* species. Intl J Mol Sci 19 (10): 3142. DOI: 10.3390/ijms19103142.

Zou Y, Zhang Z, Zeng Y, Hu H, Hao Y, Huang S, Li B. 2024. Common methods for phylogenetic tree construction and their implementation in R. Bioengineering 11: 480. DOI: 10.3390/bioengineering11050480.

SUPPLEMENTARY DATA

Supplementary Table 1. Relative Synonymous Codon Usage (RSCU) value of *Piper betle* plastome

Codon	Count	RSCU
UUU(F)	934	1.23
UUC(F)	583	0.77
UUA(L)	828	1.76
UUG(L)	599	1.28
CUU(L)	596	1.27
CUC(L)	198	0.42
CUA(L)	408	0.87
CUG(L)	186	0.4
AUU(I)	1127	1.44
AUC(I)	485	0.62
AUA(I)	729	0.93
AUG(M)	614	1
GUU(V)	544	1.45
GUC(V)	194	0.52
GUA(V)	543	1.45
GUG(V)	216	0.58
UCU(S)	587	1.6
UCC(S)	376	1.02
UCA(S)	465	1.27
UCG(S)	227	0.62
CCU(P)	225	0.9
CCA(P)	322	1.13
CCG(P)	150	0.53
ACU(T)	497	1.42
ACC(T)	294	0.84
ACA(T)	425	1.22
ACG(T)	181	0.52
GCU(A)	644	1.79
GCC(A)	217	0.6
GCA(A)	400	1.11
GCG(A)	179	0.5
UAU(Y)	782	1.53
UAC(Y)	238	0.47
UAA(*)	50	1.53
UAG(*)	26	0.8
CAU(H)	491	1.49
CAC(H)	169	0.51
CAA(Q)	706	1.48
CAG(Q)	250	0.52
AAU(N)	961	1.52
AAC(N)	300	0.48
AAA(K)	1027	1.43
AAG(K)	405	0.57
GAU(D)	908	1.58
GAC(D)	242	0.42
GAA(E)	1030	1.46
GAG(E)	382	0.54
UGU(C)	231	1.46
UGC(C)	85	0.54
UGA(*)	22	0.57
UGG(W)	490	1.0
CGU(R)	335	1.2
CGC(R)	141	0.51
CGA(R)	380	1.37
CGG(R)	148	0.53
AGU(S)	408	1.11
AGC(S)	138	0.38
AGA(R)	479	1.72
AGG(R)	187	0.67
GGU(G)	597	1.25
GGC(G)	216	0.45
GGA(G)	727	1.53
GGG(G)	363	0.76
Total	26887	