NORDIC JOURNAL OF
BOTANY

*Research Article*

# Chloroplast genome of the critically endangered Ginger *Zingiber odoriferum* Blume from Java, Indonesia: characterization, comparison, and conservation insights

**Irfan Martiansyah[1], Risha Amilia Pratiwi[1], Muhammad Rifqi Hariri[2], Mahat Magandhi[1], Asih Perwita Dewi[2] and Andik Wijayanto[3]**

[1]Research Center for Applied Botany, National Research and Innovation Agency, Cibinong, Indonesia
[2]Research Center for Biosystematics and Evolution, National Research and Innovation Agency, Cibinong, Indonesia
[3]Biology Department, Faculty of Mathematics and Natural Sciences, State University of Malang, Malang, Indonesia

**Correspondence: Irfan Martiansyah (irfa009@brin.go.id)**

**Nordic Journal of Botany**
**2025: e04902**
doi: 10.1002/njb.04902
Subject Editor: Torbjörn Tyler
Editor-in-Chief: Torbjörn Tyler
Accepted 30 July 2025
Published 6 September 2025

*Zingiber odoriferum* (Zingiberaceae), a critically endangered ginger species endemic to Java, Indonesia, is ecologically vital yet genomically understudied. We present the first complete chloroplast (cp) genome (163 538 bp), revealing a quadripartite structure with a large single-copy region (88 033 bp), a small single-copy region (15 885 bp), and a pair of inverted repeat regions (29 810 bp each), with a total GC content of 36.04%. The genome encodes 133 genes, including protein-coding genes, tRNA genes, rRNA genes, and pseudogenized *ycf*1 – a conserved Zingiberaceae feature. Leucine (10.33%) and cysteine (1.14%) are the most and least abundant amino acids, respectively. Codon usage favors A/T-rich codons (e.g. AGA [Arg], TTA [Leu]), while C/G-ending codons (e.g. AGC [Ser]) are underutilized. Hypervariable loci in non-coding regions and repeats are identified as molecular markers. Phylogenetic analysis places *Z. odoriferum* sister to *Z. teres* and *Z. smilesianum* (100% bootstrap support). Despite high genome-wide conservation, intergenic spacers exhibit 5–12% divergence, aiding species delineation. However, the limited cp diversity across *Zingiber* underscores the need for nuclear genomic integration to resolve cryptic diversity and enhance conservation strategies. This study advances genomic resources for *Z. odoriferum*, highlights cp genome evolutionary constraints, and provides actionable markers for biodiversity monitoring. It emphasizes the urgency of multi-omics approaches to safeguard this species and its threatened tropical habitat.

Keywords: chloroplast genome, comparative plastid analysis, endangered ginger species, plant conservation genetics, Zingiberaceae

## Introduction

*Zingiber odoriferum* Blume (Zingiberaceae) is an endemic ginger species from Java, Indonesia (Fig. 1), primarily inhabiting tropical forests where it contributes to

© 2025 The Author(s). Nordic Journal of Botany published by John Wiley & Sons Ltd on behalf of Nordic Society Oikos.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.



Figure 1. Habit and inflorescence of *Zingiber odoriferum* (Photo credit: AKA Syafiq).

ecological balance. Classified as 'Critically Endangered' (CR) by the IUCN Red List due to severe threats from habitat destruction, deforestation, and agricultural land conversion (Sabu et al. 2019), this species holds both ecological and pharmacological significance. It produces bioactive compounds with notable antimicrobial, anti-inflammatory, and antioxidant activities, underscoring its potential for medicinal applications, as commonly observed in other *Zingiber* species (Rellstab et al. 2021, Deng et al. 2022)

Research on *Z. odoriferum* has primarily focused on taxonomic and morphological characterization, with limited attention given to its genetic diversity. This has left critical gaps in understanding its anatomy, pharmacology, phylogenetics, and ecological adaptations (Syafiq 2024). These knowledge gaps hinder efforts to elucidate its evolutionary history, population structure, and resilience to environmental stress factors that are vital for effective conservation planning. Although molecular tools are recognized as essential in modern conservation biology, they remain underutilized for this species, despite their proven effectiveness in addressing conservation challenges for small and isolated populations (Rellstab et al. 2021).

The chloroplast (cp) genome is a cornerstone resource for plant evolutionary studies, phylogenetics, and population genetics (Allendorf 2017). Public databases like NCBI host complete cp genomes for multiple *Zingiber* species, enabling insights into evolutionary relationships, adaptive traits, and species authentication (Cui et al. 2019, Li et al. 2020, Miao et al. 2021, Jiang et al. 2023, Xia et al. 2024). However, *Z. odoriferum* lacks any published cp genome data, limiting resolution of its phylogenetic position within Zingiberaceae and obstructing evidence-based conservation strategies (Brandies et al. 2019, Kahraman and Lucas 2019).

This study aims to assemble the first complete cp genome of *Z. odoriferum* to address these knowledge gaps. By comparing it with existing *Zingiber* cp genomes, this research seeks to elucidate its phylogenetic relationships, identify evolutionary divergence patterns, and develop molecular markers for future genetic diversity assessments and DNA barcoding. These findings will advance conservation strategies for *Z. odoriferum* and enrich genomic resources for Indonesia's endangered flora, aligning with global efforts to integrate genomics into biodiversity preservation (Rellstab et al. 2021, Jiang et al. 2023).

## Material and methods

### Plant material, DNA extraction, and sequencing

Leaf samples of *Zingiber odoriferum* were collected from the Bogor Botanic Garden in West Java, Indonesia (collection number XI.B.III.121-121a), under permits issued by the Directorate of Scientific Collection Management – National Research and Innovation Agency (BRIN) (B-752/II.6.2/DI.03/2/2023). A voucher specimen was deposited at Herbarium Bogoriense (BO) for taxonomic verification and future reference.

Total genomic DNA (gDNA) was isolated from fresh leaf tissue using the cetyltrimethylammonium bromide (CTAB) method following Doyle and Doyle (1987). DNA quality was evaluated via 1% agarose gel electrophoresis to confirm high molecular weight and minimal degradation. Quantification and purity were assessed using a Nanodrop 2000 spectrophotometer (Thermo Scientific, MA, USA) for absorbance ratios (A260/A280 and A260/A230) and a Qubit 4.0 Fluorometer with dsDNA HS Assay Kits (Thermo Scientific, MA, USA) for precise concentration measurements. DNA integrity was further verified using the Agilent 4150 TapeStation system (Agilent Technologies, CA, USA) to ensure suitability for downstream sequencing.

Sequencing libraries were constructed using the xGen™ DNA Library Prep EZ UNI kit (Integrated DNA Technologies, IA, USA). Enzymatic fragmentation of gDNA yielded inserts averaging 350 bp in size. Libraries were quantified and validated for size distribution using the Agilent 4150 TapeStation. Paired-end sequencing (150 bp read length) was performed on the Illumina NextSeq 500 platform (Illumina, CA, USA) at Genetika Science Laboratory (Tangerang, Indonesia), generating approximately 10 Gb of raw sequencing data to ensure sufficient coverage for cp genome assembly.

### Cp genome assembly and annotation

Raw sequencing reads were initially assessed for quality using FastQC ver. 0.11.9 (Babraham Institute, Cambridge, UK) (Andrews 2010) to evaluate base quality scores, GC content, and sequence duplication levels. Adapter sequences and low-quality reads (Phred score $< 20$) were trimmed using Trimmomatic ver. 0.39 (Usadel Lab, Aachen, Germany) (Bolger et al. 2014) with the following parameters: SLIDINGWINDOW:4:20, MINLEN:50. Post-trimming, FastQC was rerun to confirm retention of high-quality reads.

The trimmed reads were mapped to the reference cp genome of *Z. officinale* Roscoe (NC_044775) using Bowtie2 ver. 2.5.3 (Langmead Lab, Baltimore, MD, USA) (Langmead and Salzberg 2012) on the Galaxy platform (https://usegalaxy.eu/) (The Galaxy Community 2024). with default alignment settings. The output SAM file was converted and sorted into BAM format using SAMtools ver. 1.9 (Danecek et al. 2021). A complete circular cp genome was reconstructed from the mapped reads, ensuring coverage uniformity and absence of gaps.

Automated annotation was initially performed using CPGAVAS2 (Shi et al. 2019) with *Z. officinale* as the reference. Ambiguous nucleotides, gene boundaries, and intron-exon structures were manually corrected using Unipro UGENE ver. 50.0 (Unipro, Novosibirsk, Russia) (Okonechnikov et al. 2012) and NCBI Genome Workbench ver. 3.8.2 (Kuznetsov and Bollin 2021). Cis- and trans-splicing genes were identified using CPGView (Liu et al. 2023), and the final circular cp genome map was visualized with OGDRAW ver. 1.3.1 (Greiner et al. 2019). The annotated chloroplast genome sequence of *Z. odoriferum* was submitted to NCBI via the BankIt portal (accession number PQ550047).

### Amino acids frequency, and codon usage analysis

Codon usage bias in *Z. odoriferum* was assessed by analyzing amino acid frequency and relative synonymous codon usage (RSCU) exclusively from its protein-coding sequences (CDS). Amino acid frequencies were determined using MEGA 11 (Tamura et al. 2021) by translating CDS into protein sequences and employing the software's statistical module to tabulate occurrences. Codon usage patterns were visualized using 'codon plot' (Stothard 2000), a tool within the sequence manipulation suite, to generate genome-wide codon distribution profiles. RSCU values were calculated using the formula established by Sharp and Li (1987) and updated by Zeng et al. (2023), where the observed codon frequency was divided by the expected frequency under neutral synonymous usage. Codons were categorized as neutrally used (RSCU = 1.0), preferred (RSCU $>$ 1.0), or less preferred (RSCU $<$ 1.0) to quantify bias in synonymous codon selection.

### Repeats identification

Simple sequence repeats (SSRs) in the *Z. odoriferum* cp genome were identified using Krait ver. 1.5.1 (Du et al. 2018) with thresholds set to a minimum of 10 repeats for mononucleotide motifs, 5 for dinucleotide, and 3 for tri-, tetra-, penta-, and hexanucleotide motifs. The frequency of SSR types and motif abundance were visualized using Microsoft Excel to highlight dominant repeat patterns. Genomic distribution of SSRs was categorized into the four major cp regions: the large single copy (LSC), small single copy (SSC), and inverted repeats (IRa and IRb), providing insights into their localization. Long repeats (forward, palindromic, reverse, and complementary) were detected using REPuter ver. 1.0 (https://bibiserv.cebitec.uni-bielefeld.de/reputer, Kurtz et al. 2001) with parameters set to a Hamming distance of 3, a maximum repeat length of 5000 bp, and a minimum repeat size of 30 bp. This analysis aimed to identify structural features and repetitive elements potentially linked to genome stability or evolutionary divergence.

### Comparative analysis of genomic general feature, gene arrangement, and IR junction structure

The cp genome of *Z. odoriferum* was compared to those of other *Zingiber* species from diverse geographical origins to assess structural and compositional variation. Total genome size, GC content, lengths of LSC, SSC, IRa/b regions, and gene counts (protein-coding, tRNA, and rRNA) were analyzed

using Chloroplot (Zheng et al. 2020). For species with multiple publicly available cp genome accessions, all sequences were included to capture intraspecific diversity. Genome-wide structural alignments were performed with Mauve (Darling et al. 2004) and mVISTA (Frazer et al. 2004) in mLAGAN mode, using the annotated *Z. odoriferum* genome as the reference to identify conserved and divergent regions. Inverted repeat (IR) junction boundaries were examined using IRScope (Amiryousefi et al. 2018), focusing on a subset of 10 representative sequences to clarify gene arrangement patterns at IR-LSC/SSC interfaces. These comparative analyses aimed to resolve phylogenetic relationships, structural evolution, and genomic divergence within the genus *Zingiber*.

### Phylogenetic analysis

Bioinformatics analyses were conducted on the Mahameru high performance computer (HPC) at BRIN. Chloroplast genome sequences of 32 *Zingiber* species and *Kaempferia elegans* (outgroup) were retrieved from NCBI using Batch Entrez. Annotations were standardized across all sequences using the PGA annotator (Qu et al. 2019) to ensure consistency. Prior to alignment, the second IR region was removed to avoid redundancy, and whole-genome alignment was performed using Mauve (Darling et al. 2004). Concatenated aligned sequences were generated in MEGA 11 (Tamura et al. 2021), and the best-fit nucleotide substitution model (GTR+I+G4) was selected via Modeltest-NG (Darriba et al. 2019). Phylogenetic trees were reconstructed using maximum likelihood (ML) in IQ-TREE2 ver. 2.3.6 (Minh et al. 2020) with 100 000 ultrafast bootstrap replicates (Minh et al. 2013) and Bayesian inference (BI) in MrBayes ver. 3.2.7 (Ronquist et al. 2012). For BI, two independent runs of 1 000 000 Markov Chain Monte Carlo (MCMC) generations were executed, each with four chains and a sampling frequency of 1000; the first 25% of samples were discarded as burn-in. Resulting trees from ML and BI analyses were combined and visualized using TreeGraph 2 (Stöver and Müller 2010) to resolve topological congruence and branch support.

### Substitution marker identification from nucleotide diversity calculation

To assess genetic variation across the *Zingiber odoriferum* cp genome, nucleotide diversity ($\pi$) was quantified using sliding window analysis in DnaSP ver. 6.12 (Rozas et al. 2017). This approach identifies regions of high or low sequence divergence, which may reflect evolutionary constraints or adaptive variation. The analysis was configured with a window length of 600 base pairs (bp) and a step size of 200 bp, generating overlapping windows to smooth noise and enhance resolution of diversity patterns. A Pi threshold of 0.01 was applied to distinguish conserved regions ($Pi < 0.01$) from hypervariable loci ($Pi \geq 0.01$), a cutoff selected based on preliminary data and aligned with thresholds used in comparable cp genome studies to ensure biological relevance.

Chloroplast genome alignments were curated to reduce artifacts prior to analysis. Sliding window analysis revealed nucleotide diversity ($\pi$) peaks across intergenic and coding regions. These hotspots were mapped against gene annotations to determine their functional context. Highly variable loci were flagged as potential markers for population or phylogenetic studies, while conserved regions likely represent essential genomic elements.

## Results and discussion

### Characteristic of *Z. odoriferum* cp genome

The chloroplast genome of *Zingiber odoriferum* spans 163 538 bp and exhibits the canonical quadripartite structure typical of angiosperms, consisting of an LSC region (88 033 bp), an SSC region (15 885 bp), and two IR regions (29 810 bp each) (Fig. 2). The overall GC content is 36.04%, with the IR regions showing the highest GC enrichment (41.1%), followed by the LSC (33.82%) and SSC (29.39%), consistent with patterns observed across land plants (Palmer 1985, Wicke et al. 2011, Dobrogojski et al. 2020).

The cp genome encodes 133 genes, including 87 protein-coding genes, 38 tRNAs, and 8 rRNAs (Table 1). Among these, 21 genes are duplicated in the IR regions, including all rRNA genes (*rrn16S*, *rrn23S*, *rrn4.5S*, *rrn5S*), nine tRNA genes (*trnS-GGA*, *trnA-UGC*, *trnI-GAU*, *trnN-GUU*, *trnR-ACG*, *trnL-CAA*, *trnI-CAU*, *trnH-GUG*, *trnV-GAC*), and eight protein-coding genes (*rps7*, *rps12*, *rps19*, *rpl2*, *rpl23*, *ndhB*, and pseudogenes *ycf1* and *ycf2*). Pseudogenization of *ycf1* is a common feature in Zingiberaceae, reflecting lineage-specific structural evolution (Wu et al. 2018, Li et al. 2021). Nonstandard start codons, such as *petB* (TCA) and *rpl2* (GCG), further highlight codon usage plasticity, as seen in related species (Huang et al. 2018, Jiang et al. 2023).

A total of 23 genes contain introns: 19 genes possess a single intron (e.g. *atpF*, *ndhA*, *petB*, *trnK-UUU*), while 4 genes (*rps12*, *clpP*, *ycf3*) have two introns. Introns in genes such as *trnK-UUU* (harboring *matK*) and *rpoC1* are critical for transcript maturation and post-transcriptional regulation (Liu et al. 2018, Li et al. 2021). Genes with dual introns, like *clpP* and *ycf3*, are often essential for stress adaptation, requiring precise regulation under environmental fluctuations (Wang et al. 2022, Jiang et al. 2023).

The cp genome of *Z. odoriferum* utilizes two distinct RNA splicing mechanisms to ensure accurate transcript processing. The primary mechanism, cis-splicing (Fig. 3A), involves the removal of introns from a single precursor RNA molecule, a process critical for producing mature mRNA for translation. In contrast, the *rps12* gene uniquely employs trans-splicing (Fig. 3B), a rare mechanism where exons from physically separated regions of the genome are joined post-transcriptionally (Jacobs et al. 2010) This exceptional process, likely inherited from the cyanobacterial ancestors of chloroplasts (Horiuchi and Aigaki 2006), underscores the evolutionary complexity of chloroplasts and reflects their endosymbiotic origin (Qu et al. 2023). While cis-splicing dominates most gene processing, the retention of trans-splicing in *rps12* highlights the functional and evolutionary constraints shaping

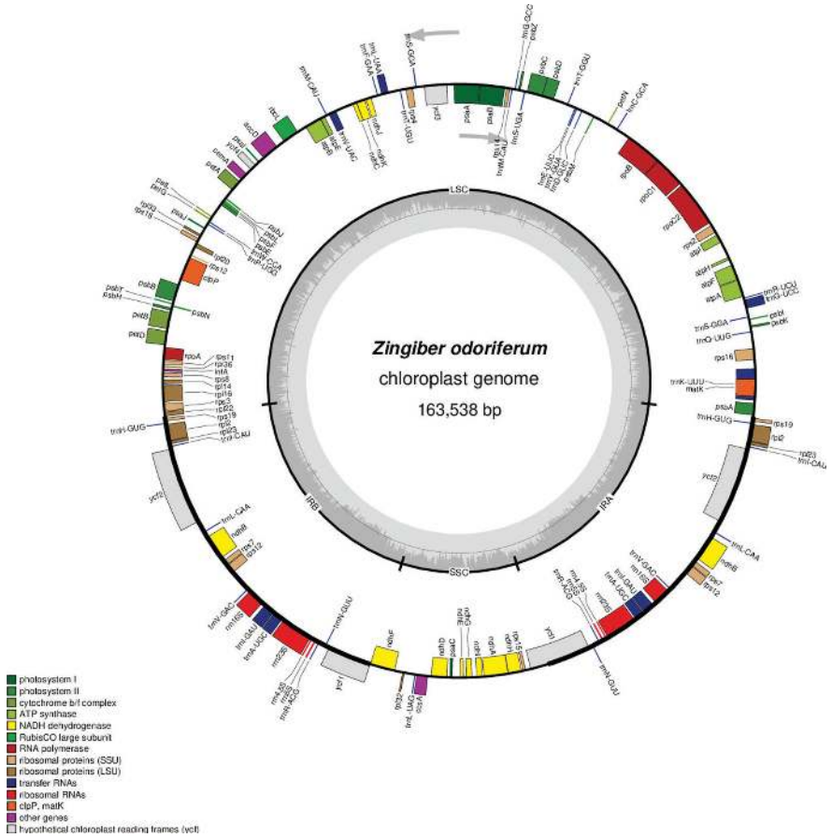



Figure 2.. circular gene map of the cp genome from *Zingiber odoriferum* was generated using OGDRAW. Genes located on the inner side of the circle are transcribed in a counterclockwise direction, whereas those on the outer side are transcribed clockwise. Different functional gene groups were highlighted using distinct colors. The inner circle also shows the GC content in grey, while the AT content is displayed in light grey.

cp genome architecture in land plants (Wicke et al. 2011, Dobrogojski et al. 2020).

*ycf1* and *ycf2* pseudogenes, although nonfunctional in translation, are believed to contribute to chloroplast genome stability and structural integrity (Drescher et al. 2000). The duplication of these genes within the IR regions is a conserved feature among land plants, supporting ribosome assembly and enhancing genomic robustness (Wang et al. 2022). In *Zingiber*, *ycf1* retains a full-length sequence; however, its expression at the IRa/SSC boundary remains unverified, warranting further investigation (Vaughn et al. 2014).

## Amino acid frequency and codon usage pattern of Z. odoriferum cp genome

The cp genome of *Z. odoriferum* encodes 87 protein-coding genes, comprising 27 670 codons. Amino acid frequencies revealed a composition consistent with evolutionary trends observed across plant cp genomes, where hydrophobic residues dominate to support structural stability and membrane-associated functions (Björkholm et al. 2015). Leucine (Leu, 10.33%) and isoleucine (Ile, 8.84%) were the most abundant amino acids (Fig. 4), reflecting their critical roles in protein folding and hydrophobic core formation (Panja et al. 2020). In contrast, cysteine (Cys, 1.14%) and tryptophan (Trp, 1.73%) exhibited the lowest frequencies, a pattern attributed to their biochemical rarity and reduced utilization in plastid proteins due to metabolic constraints (Yang et al. 2023a, b). Moderate frequencies of glycine (Gly, 6.43%) and lysine (Lys, 5.51%) underscored their roles in enzymatic activity and protein flexibility, while polar residues like arginine (Arg, 6.06%) highlighted adaptations to the cp environment, where charged residues mediate protein interactions and redox functions (Trovato et al. 2021).

Codon usage bias, quantified through RSCU analysis, further elucidated translational optimization in *Z. odoriferum*. Of the 61 sense codons, 29 showed RSCU >1.0 (preferred), while 32 were underutilized (RSCU < 1.0) (Fig. 5). Strong preferences were observed for A/T-rich codons, such as AGA (Arg, RSCU = 1.94), TTA (Leu, 1.91), and GCT (Ala, 1.80), aligning with the A/T-biased nucleotide composition of cp genomes and the abundance of corresponding tRNA pools (Li et al. 2021, Yang et al. 2023a, b, Baeza et al. 2024). Conversely, C/G-ending codons like AGC (Ser, 0.30), GGC (Gly, 0.31), and GCG (Ala, 0.33) were strongly disfavored,

Table 1. List of annotated genes of *Zingiber odoriferum* cp genome. *=gene containing a single intron; **=gene containing two introns; (2x)=two copies of gene in IRs; +=trans-splicing gene.

| Category | Group of genes | Name of genes |
|---|---|---|
| RNA genes | ribosomal RNAs | *rrn16S*(2x), *rrn23S*(2x), *rrn4.5S*(2x), *rrn5S*(2x) |
| | transfer RNAs | *trnA-UGC**(2x), *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnG-UCC**, *trnG-GCC*, *trnH-GUG*(2x), *trnI-CAU*(2x), *trnI-GAU**(2x), *trnK-UUU**, *trnL-CAA*(2x), *trnL-UAA**, *trnL-UAG*, *trnM-CAU*(2x), *trnN-GUU*(2x), *trnP-UGG*, *trnQ-UUG*, *trnR-ACG*(2x), *trnR-UCU*, *trnS-GCA*(2x), *trnS-UGA*, *trnT-GGU*, *trnT-UGU*, *trnV-GAC*(2x), *trnV-UAC**, *trnW-CCA*, *trnY-GUA* |
| Transcription and translation related genes | large subunit of ribosome | *rpl14*, *rpl16**, *rpl2**(2x), *rpl20*, *rpl23*(2x), *rpl32*, *rpl33*, *rpl36* |
| | small subunit of ribosome | *rps11*, *rps12***+(2x), *rps14*, *rps15*, *rps16**, *rps18*, *rps19*(2x), *rps2*, *rps3*, *rps4*, *rps7*(2x), *rps8* |
| | DNA dependent RNA polymerase | *rpoA*, *rpoB*, *rpoC1**, *rpoC2* |
| | translational initiation factor | *infA* |
| Photosynthesis related genes | Rubisco Large subunit | *rbcL* |
| | photosystem I | *psaA*, *psaB*, *psaC*, *psaI*, *psaJ* |
| | C-type cytochrom synthesis gene | *ccsA* |
| | photosystem II | *psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*, *ycf3*** |
| | cytochrome b/f complex | *petA*, petB*, petD*, *petG*, *petL*, *petN* |
| | ATP synthase | *atpA*, *atpB*, *atpE*, *atpF*, *atpH*, *atpI* |
| | NADH-dehydrogenase | *ndhA**, *ndhB**(2x), *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK* |
| Other genes | maturase | *matK* |
| | chloroplasts envelop membrane protein | *cemA* |
| | subunit of Acetyl-CoA-carboxylase | *accD* |
| | ATP-dependent protease subunit Proteolysis | *clpP*** |
| | hypothetical open reading frames | *ycf2*(2x), *ycf4* |
| | pseudogenes | *ycf1*(2x) |

likely due to limited tRNA availability or selection against energetically costly base-pairing (Yang et al. 2023a, b). Notably, methionine (Met) and tryptophan (Trp) exhibited no codon bias (RSCU=1.00), as their single codons preclude synonymous alternatives.

The RSCU patterns reflect a balance between mutation pressure, natural selection, and translational efficiency (Błażej et al. 2017). The prevalence of A/T-rich codons aligns with the cp genome's nucleotide composition and tRNA adaptation, optimizing translation rates for essential photosynthetic and metabolic proteins (Daniell et al. 2016). Such insights into codon usage and amino acid preferences not only advance understanding of cp genome evolution but also inform strategies for transgene design and synthetic biology applications in plastid engineering (Andargie and Congyi 2022).

## Repeats identified in cp genome of Z. odoriferum as potential molecular markers

The cp genome of *Z. odoriferum* harbors 238 SSRs, dominated by mononucleotide (47) and trinucleotide (145) motifs, with fewer dinucleotide (26), tetranucleotide (18), and pentanucleotide (2) repeats. Hexanucleotide repeats were absent. SSRs are unevenly distributed: the LSC region contains the highest density, particularly of trinucleotide repeats (70), followed by the IR region (26) and SSC (23) regions (Fig. 6A). Predominant motifs, such as AAA and AAT, reflect the genome's pronounced AT-rich composition (Fig. 6B). REPuter analysis further identified 61 long repeats, including 22 forward, 33 palindromic, and 6 reverse repeats, with no complementary repeats detected (Fig. 6C).

SSR distribution aligns with broader plant genomic trends: AT-rich mononucleotide repeats dominate non-coding regions, mirroring patterns in dicots like *Arabidopsis thaliana* and *Carica papaya*, while GC-rich repeats are rare, contrasting with monocots like *Oryza sativa* (Delseny 2003, Wang et al. 2015, Dobrogojski et al. 2020). The abundance of trinucleotide repeats in *Z. odoriferum* parallels findings in *Nicotiana* species, likely influenced by lineage-specific mutation biases or selective pressures. Their stability, less prone to frameshift mutations than dinucleotide repeats, may explain their prevalence. Notably, hypervariable intergenic spacers (e.g. *trnN-ndhF*, *rpl32-trnL*) exhibit elevated polymorphism, ideal for assessing genetic diversity and population structure in fragmented habitats.

Long repeats – concentrated at structural hotspots like IR junctions – drive recombination events, contributing to IR expansion/contraction and genome size variability. Palindromic repeats near *rps19* and *ndhF* correlate with IR boundary shifts, critical for phylogenetic resolution within *Zingiber*. While less polymorphic than SSRs, conserved long repeats in coding regions (e.g. *rbcL, psbA*) serve as stable anchors for synteny-based evolutionary comparisons.

*Zingiber odoriferum*'s elevated SSR and long repeat density compared to congeners underscores its genomic uniqueness, potentially linked to adaptive divergence in Java's isolated

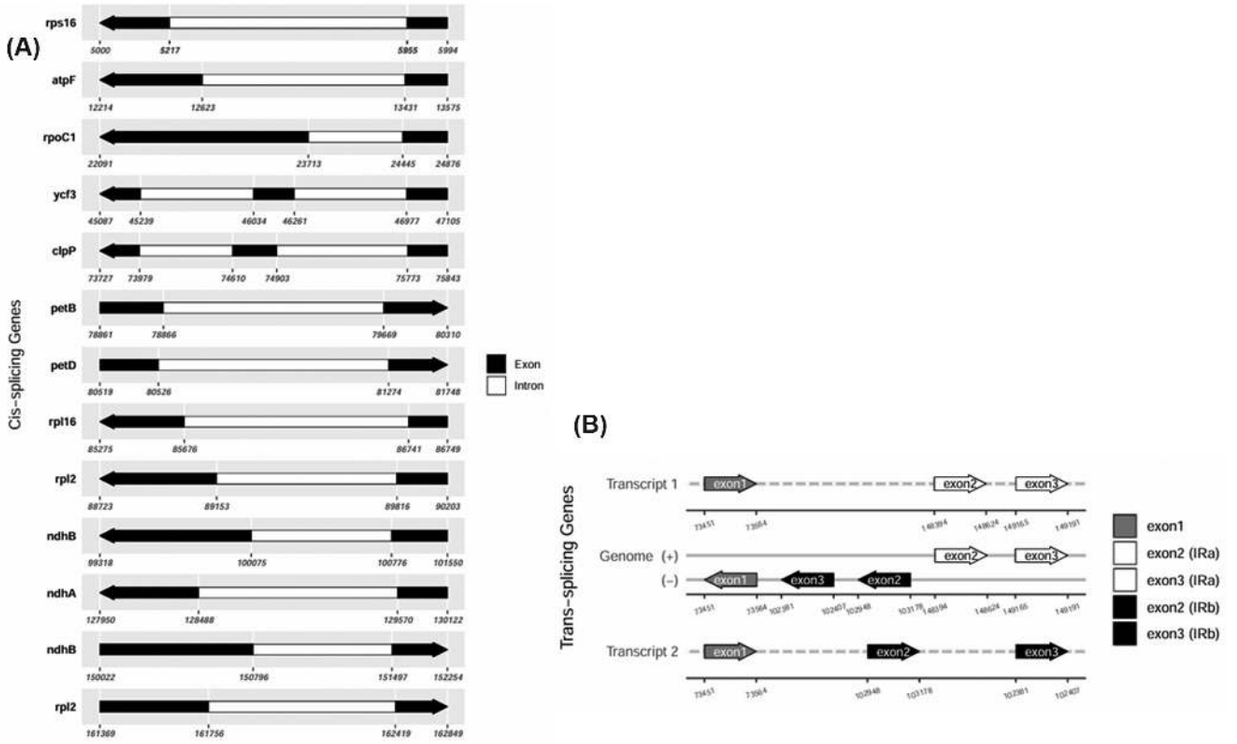



Figure 3. Cis- and trans-splicing genes in the cp genome of *Zingiber odoriferum*. (A) Cis-splicing genes, where exons (black) and introns (gray) are depicted, (B) the trans-splicing gene *rps12*.

forests. Automated annotations, however, require manual validation to address homopolymer overestimation or boundary misidentification. The dual utility of repeats – SSRs for population-level monitoring and long repeats for structural evolution – highlights their value in conservation and phylogenetics. Future studies integrating cross-species comparisons could unravel the functional significance of these motifs, bridging genomic plasticity with ecological resilience in this critically endangered species (Gao et al. 2013).

## General feature of genome, gene arrangement, and IR junction comparison

The cp genomes of *Zingiber* species show considerable structural variation, with total lengths ranging from 155 890 bp in *Z. spectabile* to 165 488 bp in *Z. citriodorum* (Table 2). This variation is mainly due to differences in the SSC region, which spans from 13 855 bp in *Z. citriodorum* to 19 488 bp in *Z. corallinum*. The LSC region varies moderately (86 171–88 205 bp), while the IR regions are relatively conserved (29 700–30 900 bp), except in *Z. spectabile*, where the IRs are shorter (~25 600 bp). Despite these differences in genome size and structure, GC content remains stable at around 36% across all species, indicating strong compositional conservation.

Total gene counts in *Zingiber* species genomes range from 131 to 135, with most species encoding 132–133 genes. The number of protein-coding genes varies slightly (85–89), while tRNA (38) and rRNA (8) counts remain constant, reflecting a conserved translational machinery (Jiang et al. 2023, Li et al. 2023). *Zingiber odoriferum*, the only Indonesian species included in this analysis, displays a typical gene distribution, with a genome size of 163 538 bp, an LSC of 88 033 bp (one of the longest among the species), and IRs measuring 29 810 bp. These structural similarities suggest shared evolutionary constraints, whereas minor variations in gene content and genome length may indicate lineage-specific adaptations (Palmer 1985, Drescher et al. 2000).

Whole-genome alignments show over 95% sequence similarity across *Zingiber* cp genomes (Fig. 7–8). Coding regions such as *rbcL* and *psbA* exhibit near-perfect conservation, consistent with their essential roles in photosynthesis and gene regulation (Dobrogojski et al. 2020). In contrast, 2–5% divergence is observed in intergenic spacers (e.g. *psbA–trnH*, *rpl16* intron) and several genes (e.g. *ycf1*), driven by higher mutation rates in non-coding regions (Zhang 2002, Jeffares et al. 2006). Synonymous substitutions dominate the coding regions, occurring at rates of 0.002–0.005 substitutions/site/million years (Lynch and Jarrell 1993, Wickett et al. 2011). However, insertions/deletions and nonsynonymous changes in genes such as *atpB* and *rps12* suggest the presence of lineage-specific evolutionary pressures. Hypervariable loci such as *ycf1*, with divergence rates reaching 5–12%, represent valuable markers for improving phylogenetic resolution within the genus.

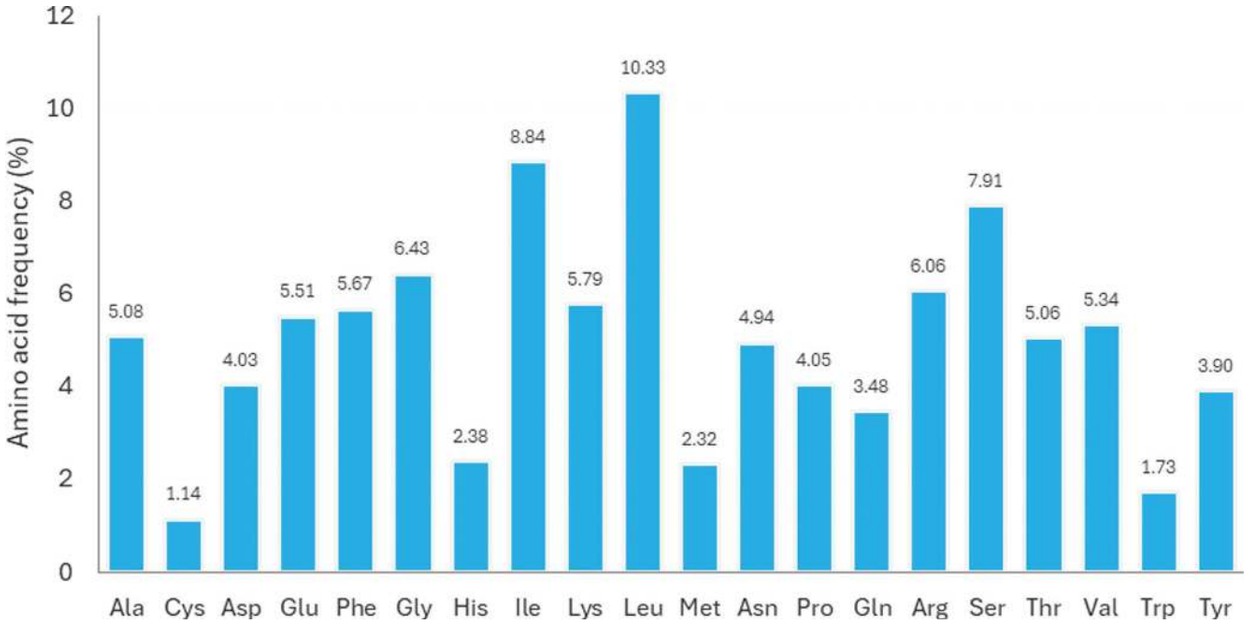



Figure 4. Amino acid frequency of protein-coding *Zingiber odoriferum* cp genome.

Regions with elevated divergence (5–12%), visualized as blue/red peaks in alignments (Fig. 8), coincide with known hypervariable loci such as the *psb*A-*trn*H spacer, *rpl*16 intron, and *ycf*1 gene. These loci, critical for phylogenetic and population genetic studies, exhibit divergence rates sufficient to resolve closely related species. For example, the *ycf*1 gene, a recombination-prone region, shows pronounced variability linked to lineage-specific adaptations, such as responses to

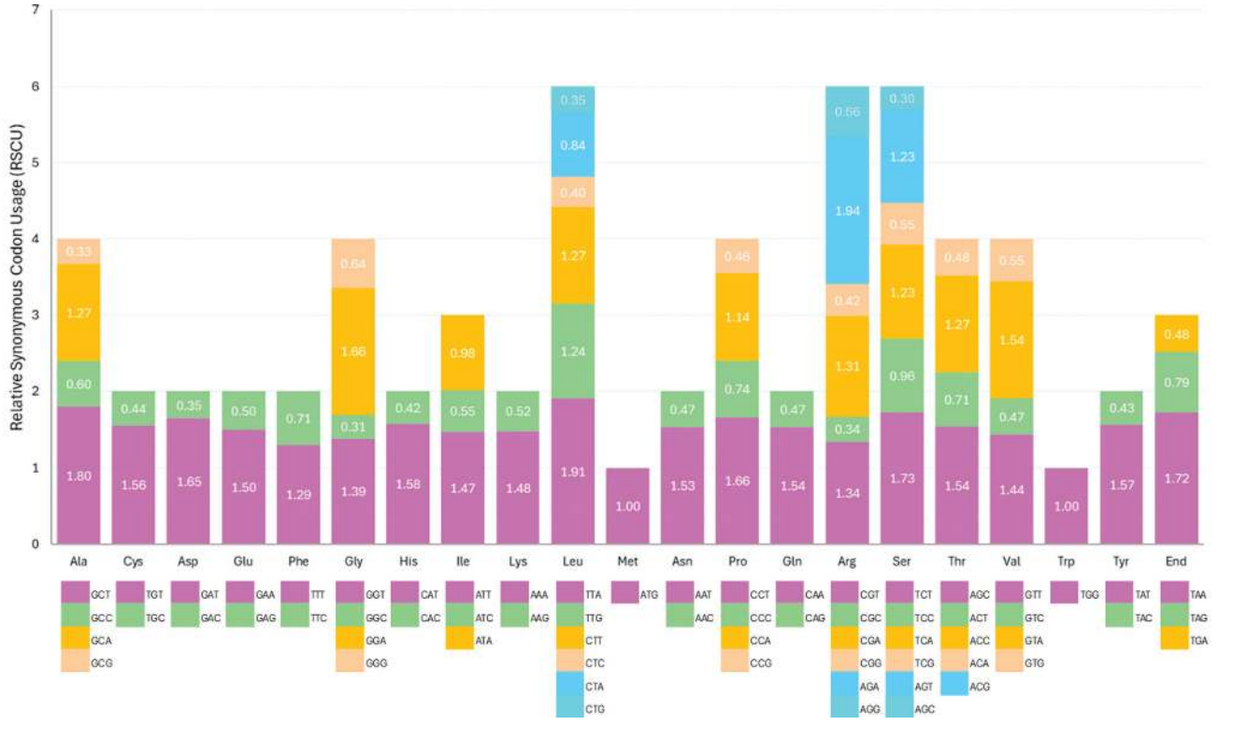



Figure 5. RSCU value of protein-coding *Zingiber odoriferum* cp genome. The bar plot illustrates the RSCU values for codons encoding each amino acid. Different colors represent individual codons within each amino acid group.

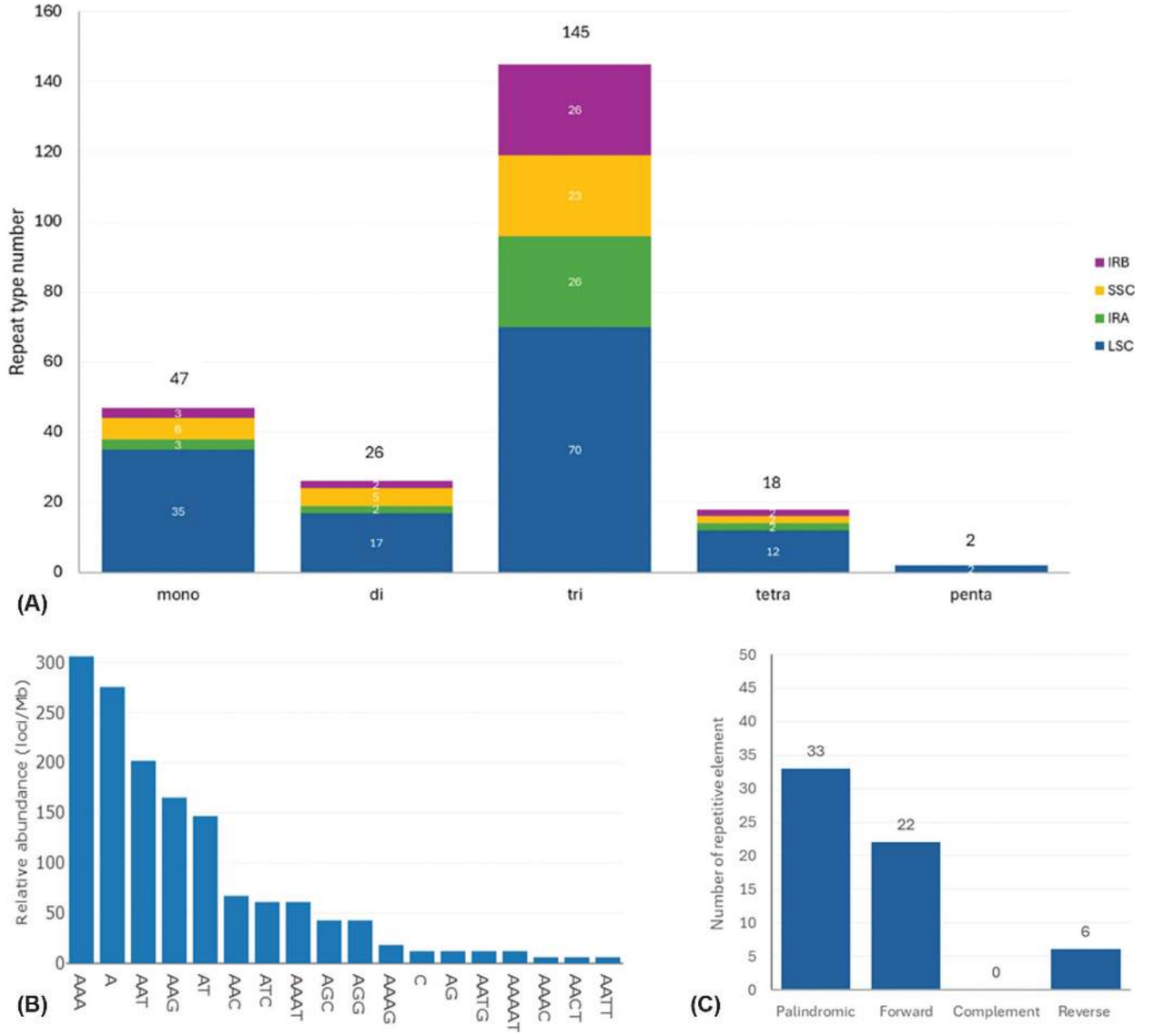



Figure 6. The identified SSR in *Zingiber odoriferum* cp genome. (A) SSR repeat profile and distribution in LSC, SSC, IRa, and IRb, (B) SSR motifs and abundance, and (C) long repeat profile.

environmental stressors in tropical ecosystems. Such hotspots are ideal for developing molecular markers to delineate species boundaries or track population dynamics within the genus.

IR-LSC/SSC junctions in *Zingiber* species display conserved gene order but subtle structural shifts (Fig. 9). At the IRb-LSC boundary, *rps*19 extensions vary (29 752–29 907 bp), influencing adjacent LSC lengths (87 430–88 167 bp). The IRb-SSC junction shows differential *ndh*F overlap (7–63 bp), reflecting IR contraction/expansion. Notably, *Z. odoriferum* exhibits a 19 bp *ndh*F extension into IRb, contrasting with broader genus trends. At the SSC-IRa boundary, partial *ycf*1 duplication spans 5456–5504 bp, with *Z. odoriferum* displaying distinct partitioning (3877 bp in IRa; 1595 bp in SSC).

The comparison of the IR junctions among *Z. odoriferum* and other closest related *Zingiber* species shows relatively conserved arrangement with minor variations in gene positioning and segment sizes (Fig. 9). Genes such as *rps*3, *rpl*22, *rps*19, *rpl*2, *trn*H, *ndh*F, *ycf*1, and *psb*A were typically harbour at the four main junctions: IRb-LSC, IRb-SSC, SSC-IRa, and IRa-LSC with slightly varied configurations. These variations, alongside shifts in *psb*A positioning (97–135 bp from IRa), highlight IR boundary plasticity as a driver of genome size variation and structural evolution (Wang et al. 2024).

## Phylogenetic analysis

Phylogenetic reconstruction using maximum likelihood (ML) and Bayesian inference (BI) yielded congruent

Table 2. Comparative cp genome characteristics of *Zingiber* species. The data presented in Table was processed from publicly available NCBI data and analyzed using the Chloroplot online platform (https://irscope.shinyapps.io/Chloroplot/).

| Species | Origin | Accession number | Total length (bp) | Total GC content | LSC length (bp) | IRA length (bp) | SSC length (bp) | IRB length (bp) | Coding gene | tRNA | rRNA | Total genes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Zingiber odoriferum** | Indonesia | PQ550047 | 163 538 | 36 | 88 033 | 29 810 | 15 885 | 29 810 | 87 | 38 | 8 | 133 |
| *Zingiber citriodorum* | Thailand | PP542025 | 165 488 | 36 | 87 118 | 32 291 | 13 855 | 32 224 | 87 | 38 | 8 | 133 |
| *Zingiber cochleariforme* | China southeast | NC_072310 | 163 665 | 36 | 88 167 | 29 855 | 15 788 | 29 855 | 87 | 38 | 8 | 133 |
| *Zingiber corallinum* | China southeast, Hainan, Laos, Thailand | NC_063565 | 160 957 | 36 | 87 835 | 26 817 | 19 488 | 26 817 | 87 | 38 | 8 | 133 |
| *Zingiber densissimum* | China south-central, Laos, Myanmar, Thailand, Vietnam | NC_072300 | 163 607 | 36 | 87 981 | 29 890 | 15 846 | 29 890 | 87 | 38 | 8 | 133 |
| *Zingiber ellipticum* | China south-central | NC_072301 | 163 455 | 36 | 87 946 | 29 869 | 15 771 | 29 869 | 87 | 38 | 8 | 133 |
| *Zingiber flavomaculosum* | China south-central, Myanmar, Thailand | NC_072311 | 163 298 | 36 | 88 124 | 29 765 | 15 644 | 29 765 | 85 | 38 | 8 | 131 |
| *Zingiber fragile* | China south-central, Thailand | OR337869 | 163 381 | 36 | 88 106 | 29 771 | 15 733 | 29 771 | 86 | 38 | 8 | 132 |
| *Zingiber guangxiense* | China southeast | OR337870 | 163 050 | 36 | 87 720 | 29 782 | 15 766 | 29 782 | 86 | 38 | 8 | 132 |
| *Zingiber koshunense* | Taiwan | NC_072302 | 163 394 | 36 | 87 785 | 29 887 | 15 835 | 29 887 | 87 | 38 | 8 | 133 |
| *Zingiber leptorrhizum* | China southeast | NC_072309 | 162 956 | 36 | 87 430 | 29 902 | 15 722 | 29 902 | 87 | 38 | 8 | 133 |
| *Zingiber longiglande* | China southeast | OR337871 | 163 125 | 36 | 87 693 | 29 906 | 15 720 | 29 806 | 86 | 38 | 8 | 132 |
| *Zingiber mioga* | China south-central, China southeast, Japan, Nansei-shoto, Taiwan, Vietnam | NC_057615 | 163 541 | 36 | 88 035 | 29 810 | 15 886 | 29 810 | 87 | 38 | 8 | 133 |
| *Zingiber mioga* | | OR337872 | 163 511 | 36 | 88 026 | 29 807 | 15 871 | 29 807 | 86 | 38 | 8 | 132 |
| *Zingiber montanum* | Cambodia, China south-central, China southeast, Hainan, Laos, Thailand, Vietnam | NC_053656 | 164 464 | 36 | 87 868 | 30 349 | 15 803 | 30 444 | 87 | 38 | 8 | 133 |
| *Zingiber montanum* | | OP869985 | 163 476 | 36 | 87 797 | 29 842 | 15 995 | 29 842 | 86 | 38 | 8 | 132 |
| *Zingiber montanum* | | MW801386 | 161 483 | 36 | 87 925 | 27 035 | 19 488 | 27 035 | 87 | 38 | 8 | 133 |
| *Zingiber neotruncatum* | Assam, China south-central, East Himalaya, Laos, Thailand | NC_072303 | 162 481 | 36 | 87 001 | 29 834 | 15 812 | 29 834 | 87 | 38 | 8 | 133 |
| *Zingiber officinale* | Assam, China south-central, east Himalaya, India | OR337873 | 162 921 | 36 | 87 486 | 29 929 | 15 577 | 29 929 | 86 | 38 | 8 | 132 |
| *Zingiber officinale* | Assam, China south-central, East Himalaya, India | MW602894 | 162 721 | 36 | 87 486 | 29 829 | 15 577 | 29 829 | 87 | 38 | 8 | 133 |
| *Zingiber officinale* | | NC_044775 | 162 621 | 36 | 87 486 | 29 779 | 15 577 | 29 779 | 87 | 38 | 8 | 133 |
| *Zingiber officinale* | | MN736958 | 161 519 | 36 | 86 304 | 29 809 | 15 577 | 29 829 | 85 | 38 | 8 | 131 |
| *Zingiber orbiculatum* | China south-central, Laos, Myanmar, Thailand | NC_072304 | 163 527 | 36 | 88 032 | 29 833 | 15 829 | 29 833 | 87 | 38 | 8 | 133 |

(*continued*)

Table 2. Continued.

| Species | Origin | Accession number | Total length (bp) | Total GC content | LSC length (bp) | IRA length (bp) | SSC length (bp) | IRB length (bp) | Coding gene | tRNA | rRNA | Total genes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Zingiber purpureum* | Assam, Bangladesh, east Himalaya, India, Myanmar | NC_072305 | 163 135 | 36 | 87 726 | 29 807 | 15 795 | 29 807 | 87 | 38 | 8 | 133 |
| *Zingiber recurvatum* | China south-central, Laos | MT473712 | 163 151 | 36 | 87 780 | 29 792 | 15 787 | 29 792 | 87 | 38 | 8 | 133 |
| *Zingiber recurvatum* | | OR337876 | 163 129 | 36 | 87 825 | 29 793 | 15 718 | 29 793 | 86 | 38 | 8 | 132 |
| *Zingiber roseum* | Bangladesh, east Himalaya, India, Myanmar, Thailand | OR337877 | 163 529 | 36 | 88 034 | 29 833 | 15 829 | 29 833 | 86 | 38 | 8 | 132 |
| *Zingiber simaoense* | China south-central, Thailand | OR337878 | 163 551 | 36 | 88 072 | 29 792 | 15 895 | 29 792 | 86 | 38 | 8 | 132 |
| *Zingiber smilesianum* | China south-central, Laos, Thailand | NC_072306 | 163 640 | 36 | 88 142 | 29 858 | 15 782 | 29 858 | 87 | 38 | 8 | 133 |
| *Zingiber spectabile* | Malaya, Thailand | NC_020363 | 155 890 | 36 | 86 171 | 25 616 | 18 508 | 25 595 | 86 | 38 | 8 | 132 |
| *Zingiber striolatum* | China south-central, China southeast | NC_065159 | 163 711 | 36 | 88 026 | 29 907 | 15 871 | 29 907 | 86 | 38 | 8 | 132 |
| *Zingiber striolatum* | | OM868073 | 163 459 | 36 | 88 205 | 29 752 | 15 750 | 29 752 | 87 | 40 | 8 | 135 |
| *Zingiber teres* | China south-central, Laos, Thailand | NC_062457 | 163 428 | 36 | 88 142 | 29 752 | 15 782 | 29 752 | 87 | 38 | 8 | 133 |
| *Zingiber wandingense* | China south-central | OR337879 | 163 398 | 36 | 87 917 | 29 855 | 15 771 | 29 855 | 86 | 38 | 8 | 132 |
| *Zingiber xishuangbannaense* | Laos, Myanmar, Vietnam | NC_072307 | 163 487 | 36 | 88 199 | 29 895 | 15 498 | 29 895 | 87 | 38 | 8 | 133 |
| *Zingiber yingjiangense* | China south-central | NC_072308 | 163 623 | 36 | 87 957 | 29 934 | 15 798 | 29 934 | 87 | 38 | 8 | 133 |
| *Zingiber yunnanense* | China south-central, Thailand | OR337880 | 163 772 | 36 | 88 082 | 29 892 | 15 906 | 29 892 | 86 | 38 | 8 | 132 |
| *Zingiber zerumbet* | Andaman Is., Assam, Bangladesh, Borneo, Cambodia, China south-central, China southeast, east Himalaya, India, Laos, Malaya, Myanmar, Nicobar Is., Philippines, Taiwan, Thailand, Vietnam | MK262726 | 163 589 | 36 | 89 161 | 29 353 | 15 722 | 29 353 | 89 | 38 | 8 | 135 |

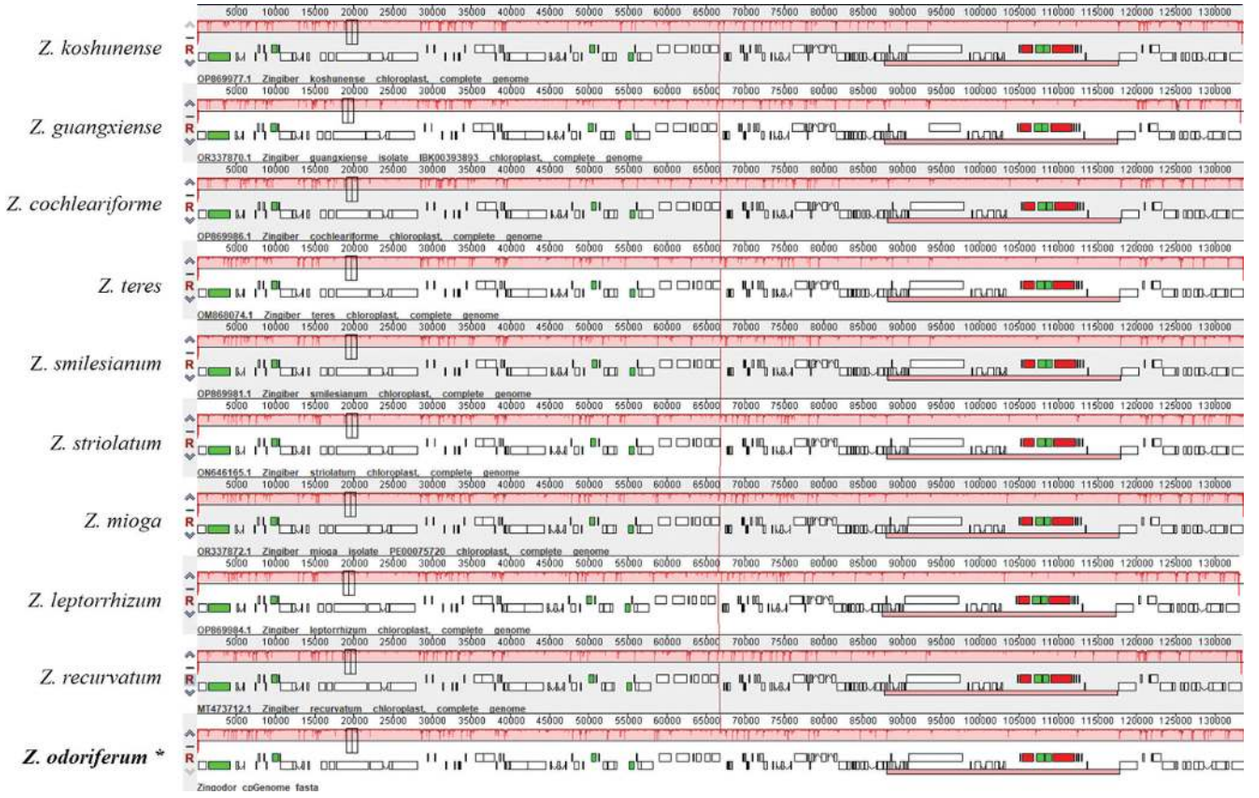



Figure 7. Multiple sequence alignment of cp genomes of *Zingiber odoriferum* (marked by asterisk) with its related species using Mauve. Each horizontal block represents a locally collinear sequence region (LCR) across different *Zingiber* species, with conserved regions shown in the same relative positions. The black lines within blocks indicate gene arrangements, while colored segments highlight syntenic regions. The overall structure suggests a high level of genome conservation among *Zingiber* species, with minor rearrangements or insertions/deletions. The red lines and gaps indicate potential structural variations, which may be useful for phylogenetic and evolutionary studies within the genus.

topologies (Fig. 10), with robust statistical support for tree reliability. Convergence criteria were met, as evidenced by an effective sample size (ESS) of 751 (> 200 threshold), potential scale reduction factor (PSRF) of 0.999 (approaching 1.0), and average standard deviation of split frequencies of 0.0045 (< 0.01) (Harrington et al. 2021), confirming the stability of the inferred tree.

*Zingiber odoriferum* forms a robustly supported clade (100% maximum likelihood bootstrap / 1.0 Bayesian posterior probability) with *Z. teres* S.Q.Tong & Y.M.Xia and *Z. smilesianum* Craib. Notably, *Z. teres* is regarded as a synonym of *Z. smilesianum*, highlighting a taxonomic consensus on their conspecific status. This clade is nested within a larger, well-resolved monophyletic group that includes *Z. koshunense* C.T.Moo (synonym to *Z. kawagoei* Hayata), *Z. guangxiense* D.Fang, *Z. cochleariforme* D.Fang, and both accessions of *Z. recurvatum* S.Q.Tong & Y.M.Xia. These species collectively belong to section *Cryptanthium* (indicated in light blue in the Fig. 10), characterized by their morphological adaptations such as strongly sheathing leaf bases, radical inflorescences, and distinctive bract and bracteole morphology.

The phylogenetic position of *Z. odoriferum* within this group indicates a close evolutionary affinity with other members of section *Cryptanthium*, further supporting its sectional placement. This relationship is congruent with earlier taxonomic frameworks based on morphology (Larsen et al. 1998, Theilade 1999) and is now corroborated by molecular evidence. The shared clade reflects a likely recent divergence from a common ancestor, possibly driven by geographical speciation across southeast Asia.

However, a phylogenetic comparison between nuclear and plastid markers reveals an incongruent placement of *Z. odoriferum*. Ardiyani et al. (2017) found that based on nuclear ITS sequences, *Z. odoriferum* clustered with *Z. montanum* and *Z. purpureum* – species from section *Zingiber*. In contrast, the whole plastid genome analysis in this study placed *Z. odoriferum* with *Z. teres* and *Z. smilesianum* of section *Cryptanthium*. Such incongruence between nuclear and plastid phylogenies is not uncommon and has been reported in several taxa, often due to hybridization, introgression, or plastid capture (Shaw et al. 2005, Wang et al. 2014). In Zingiberaceae, similar conflicts have been discussed

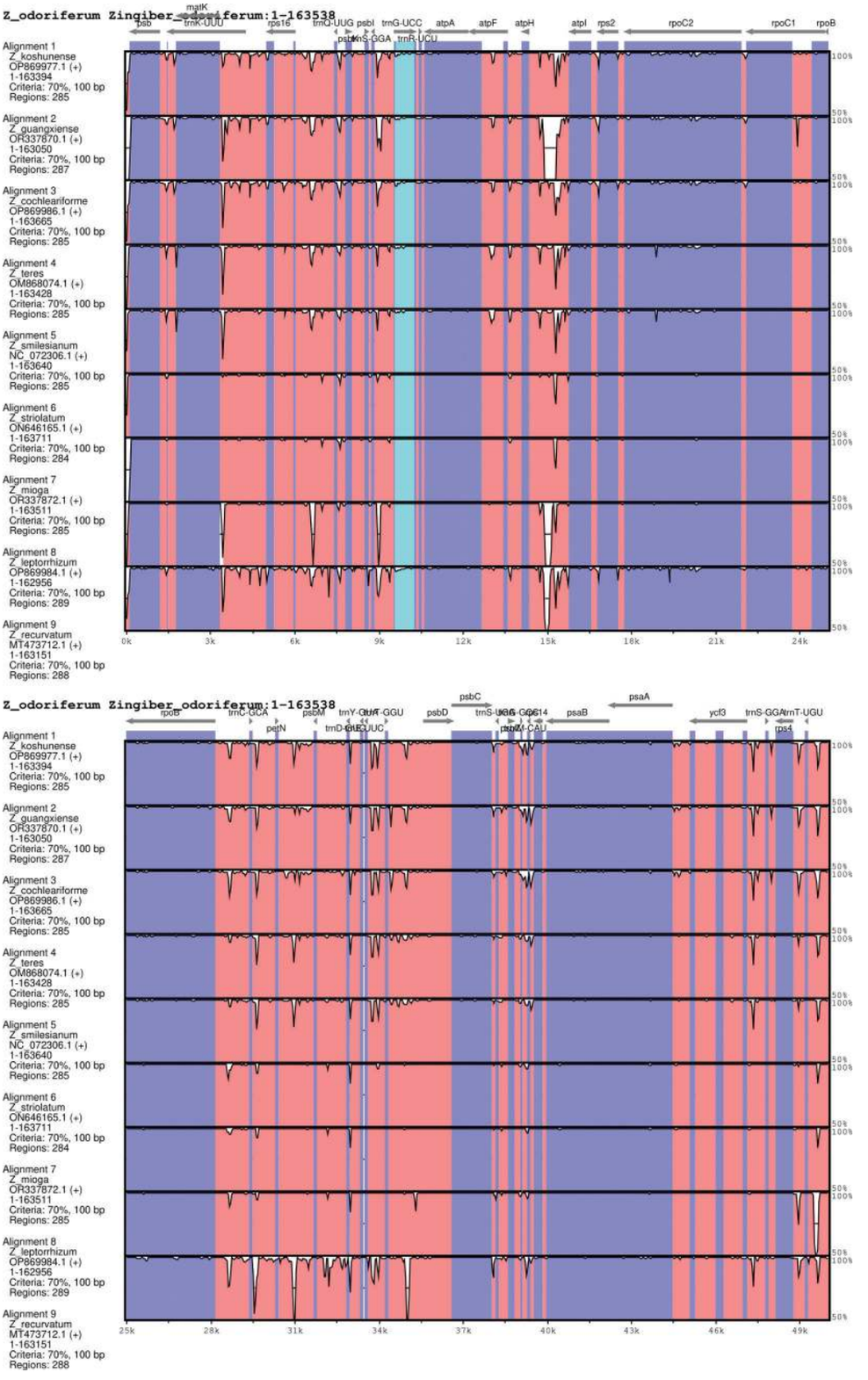



Figure 8. Multiple sequence alignment of cp genomes of *Zingiber odoriferum* and related species was performed using mVISTA, with the *Z. odoriferum* sequence placed in the top row as the reference.

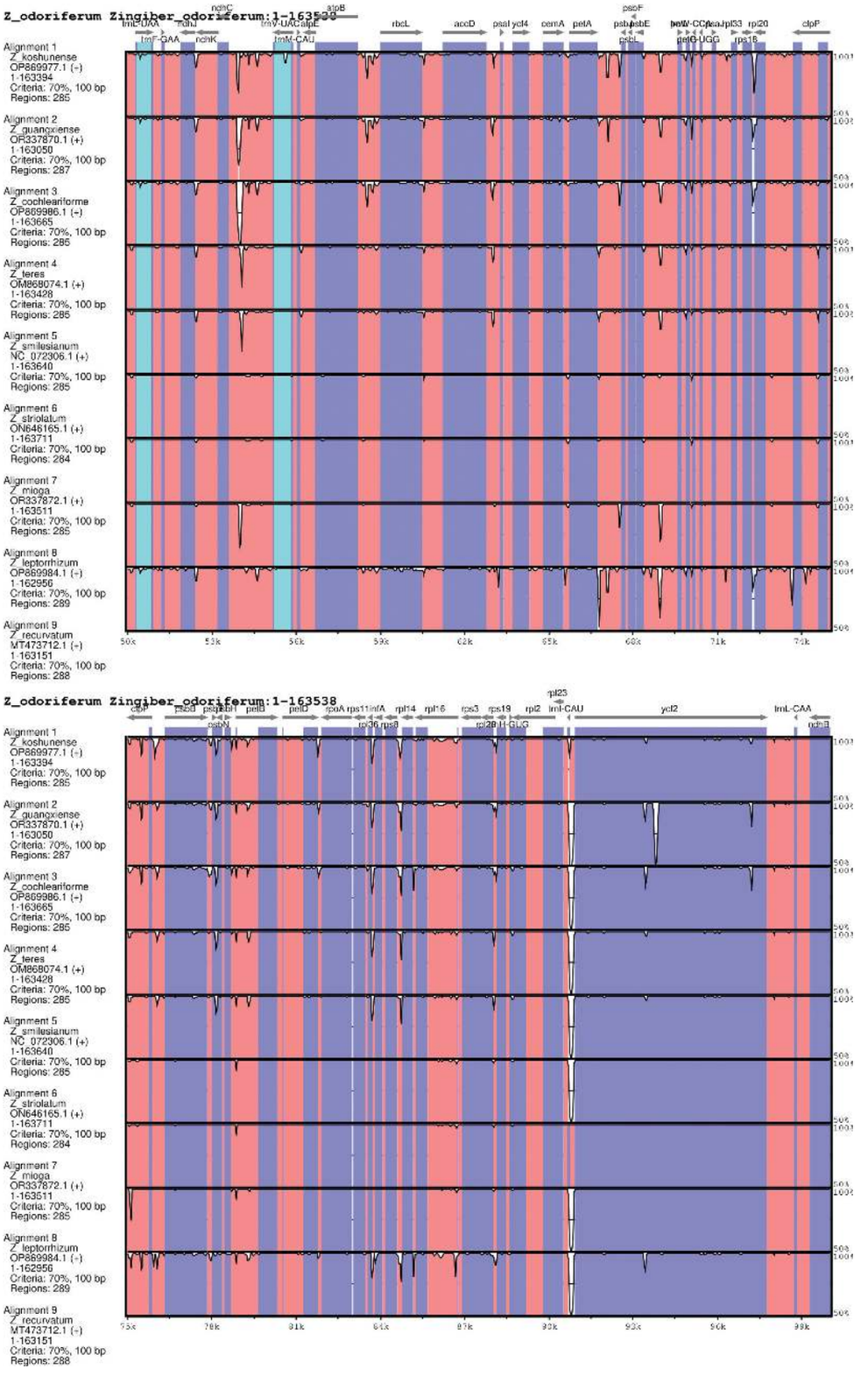



Figure 8. Continued.

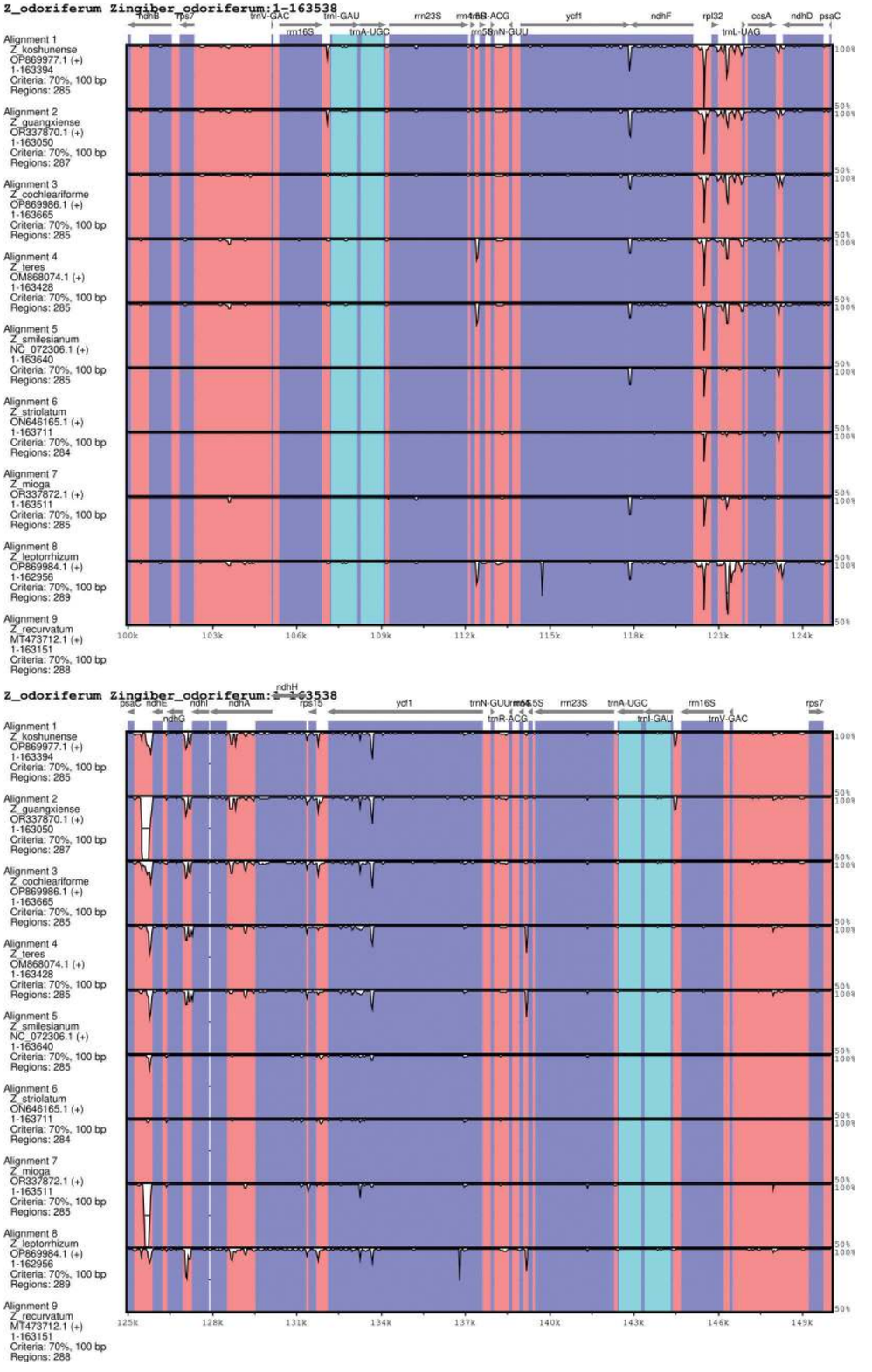



Figure 8. Continued.

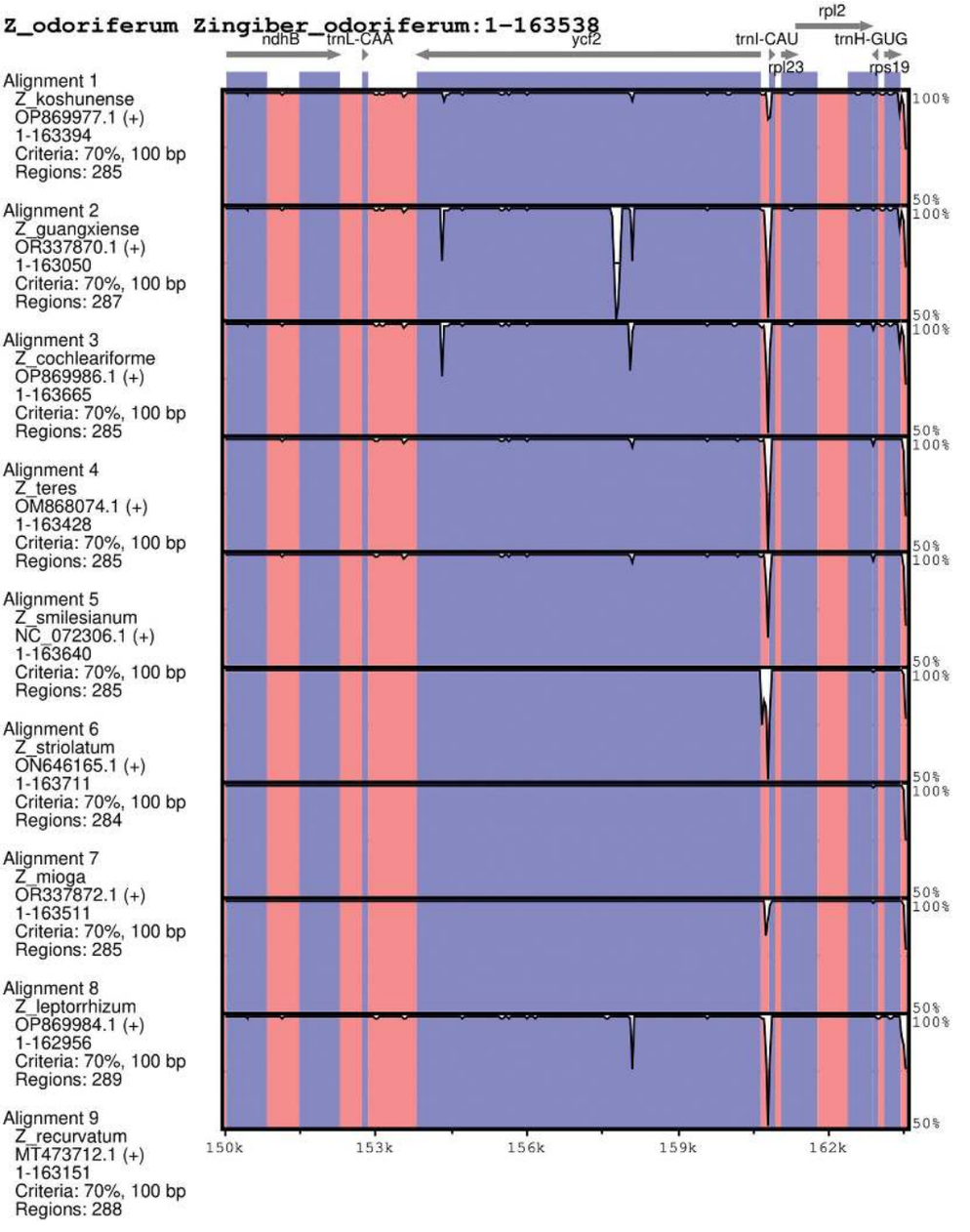



Figure 8. Continued.

(Kress et al. 2002, Theerakulpisut et al. 2012), emphasizing the importance of multilocus analyses to resolve complex evolutionary histories.

In contrast, species such as *Z. officinale* and its intraspecific accessions are part of a moderately supported but clearly separate clade (52–76% bootstrap), representing section *Zingiber* (highlighted in green). This phylogenetic separation underscores the deep genetic divergence between sections *Zingiber* and *Cryptanthium*, which may correspond to long-standing ecological differentiation, such as habitat specificity and floral syndromes. Accordingly, *Z. odoriferum* is unambiguously nested within section *Cryptanthium*, and this has been reflected in the updated Fig. 10 (legend and shading). Its position as sister to *Z. smilesianum* supports both species' inclusion within this section and highlights the need for continued revision and potential synonymization within closely allied taxa.

The broader genus structure reveals well-defined subgroups, with *Z. odoriferum* nested within a lineage including *Z. striolatum* Diels and *Z. mioga* (Thunb.) Roscoe, rather than

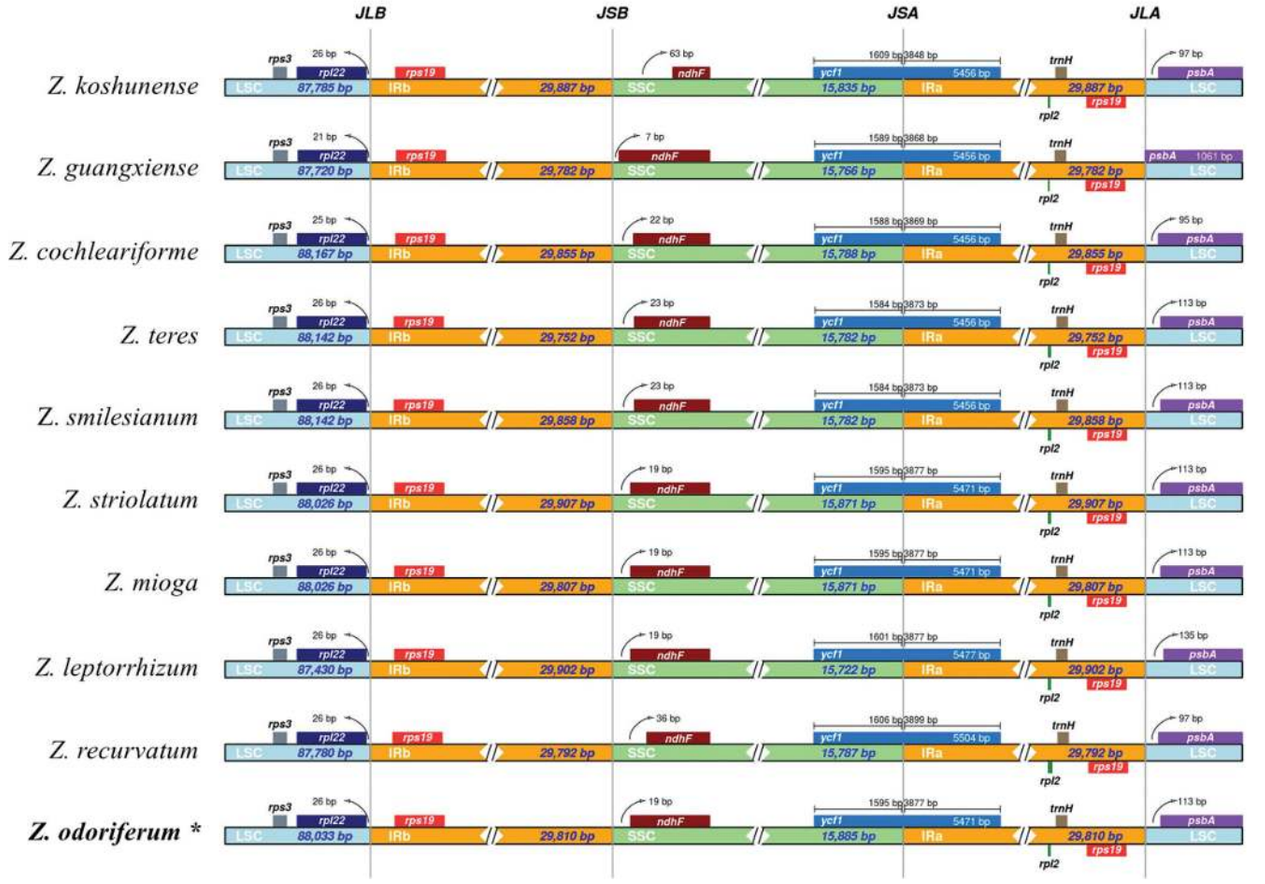



Figure 9. The contraction and expansion of the IR junctions in *Zingiber* species. The diagram illustrates the cp genome of each species along a horizontal track, indicating their respective genome lengths. Four vertical lines represent the junctions between key regions: JLB (LSC/IRb), JSB (IRb/SSC), JSA (SSC/IRa), and JLA (IRa/LSC). Genes positioned above the tracks are transcribed in a clockwise direction, while those below are transcribed counterclockwise. The distances shown are not to scale.

morphologically similar species like *Z. montanum* (J.Koenig) Link ex A.Dietr. or *Z. purpureum* Roscoe. This discordance between chloroplast genome phylogeny and traditional morphology underscores the utility of molecular data in resolving complex evolutionary relationships (Zaman et al. 2025). Highly divergent species such as *Z. xishuangbannaense* and *Z. corallinum* Hance further highlight the genus's genetic diversity, suggesting rapid radiation or niche-specific adaptations.

The strong nodal support across the tree (most clades > 95% bootstrap/1.0 posterior probability) validates the cp genome's utility for phylogenetic inference. However, the limited resolution within the *Z. officinale* clade hints at unresolved intraspecific diversity or hybridization events. These findings refine the evolutionary framework of *Zingiber*, positioning *Z. odoriferum* as a genetically distinct lineage with closer ties to Indo-Malayan species (e.g. *Z. teres*) than to widely cultivated relatives like *Z. officinale*.

The phylogenetic placement of *Z. odoriferum* provides critical insights for conservation prioritization, particularly given its endemic status and threatened classification. The close relationship with *Z. teres* and *Z. smilesianum* suggests shared ecological or adaptive traits, warranting comparative studies on their habitat requirements and stress responses. Conversely, the divergence from *Z. officinale* underscores the need for species-specific conservation strategies. Future work integrating nuclear genomic data and biogeographic modeling could unravel speciation mechanisms and historical migration patterns, while functional assays of variable loci (e.g. *ycf1, psbA-trnH*) may elucidate adaptive evolution in this ecologically vital genus.

## Potential marker from nucleotide diversity calculation

Nucleotide diversity ($\pi$) analysis of the *Z. odoriferum* cp genome reveals distinct patterns of genetic variation, with significant divergence observed in non-coding regions compared to conserved coding sequences (Fig. 11). Coding regions, particularly those involved in essential functions like photosynthesis (*rbcL*, *psaB*) and translation (*rps14*, *rpl20*), exhibit low $\pi$ values ($\pi < 0.005$), reflecting strong purifying selection to maintain

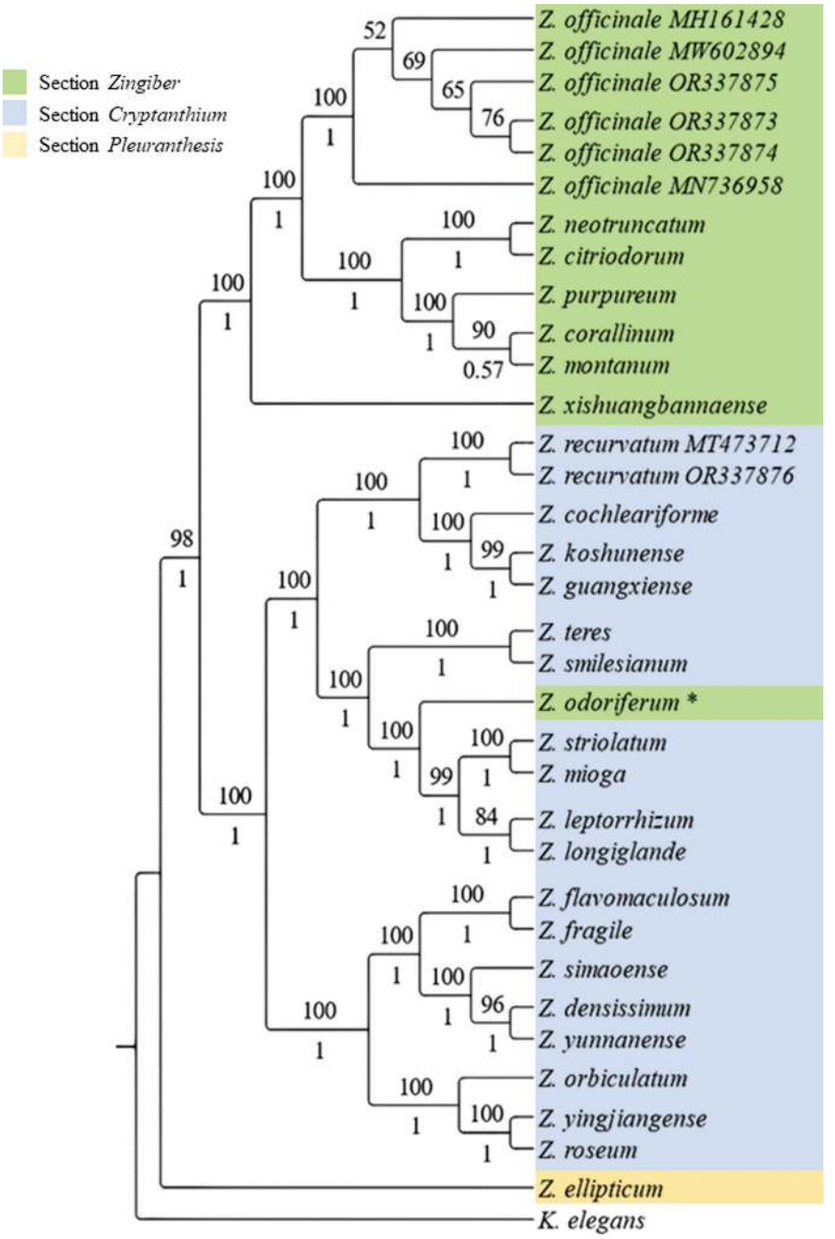



Figure 10. Phylogenetic tree of *Zingiber* species based on cp genome sequences, showing the taxonomic position of *Zingiber odoriferum* within the genus *Zingiber*.

core biochemical processes. In contrast, intergenic spacers and non-coding loci show elevated π values ($\pi > 0.02$), with peaks in regions such as *rpl20–clpP1*, *ndhG–ndhA*, *rbcL–accD*, and *ycf1*, consistent with patterns observed in other angiosperm cp genomes (Jiang et al. 2023, Xia et al. 2024).

The *ycf1* gene emerged as the most polymorphic locus ($\pi > 0.04$), corroborating its role as a hypervariable marker across plant taxa (Xia et al. 2024). This gene, critical for plastid protein import and stress responses, accumulates mutations due to relaxed selection in its non-essential domains, making it a robust tool for species delimitation and phylogenetics. Other hotspots, including *trnN–GUU/ndhF* and *rpl32–trnL*, exhibit similar variability, likely driven by their roles in regulatory elements or recombination-prone regions. These loci align with prior studies identifying *petA–psbJ* and *rpl32–trnL* as divergence hotspots in *Zingiber* (Jiang et al. 2023), underscoring their utility for population genetics and DNA barcoding.

The sliding window analysis of nucleotide diversity (π) across the *Zingiber odoriferum* cp genome, using a threshold

of $\pi > 0.01$, revealed multiple hypervariable regions predominantly located in intergenic spacers and certain protein-coding regions (Fig. 11). The most prominent diversity peaks were identified in *rpoB–trnD-GUC*, *rps14–psaB*, *rbcL–accD*, *psbE–petL*, *rpl20–clpP1* (the highest), *trnN-GUU–ndhF*, *ndhD–ndhE*, *ndhG–ndhA*, and *rps15–ycf1*. These regions exhibit nucleotide diversity values ranging from 0.01 to 0.022, clearly surpassing the average background variation and supporting their potential as mutation hotspots.

This pattern mirrors previous findings in other *Zingiber* species (Jiang et al. 2023, Xia et al. 2024), where intergenic regions such as *rbcL–accD*, *rpl20–clpP1*, and *petA–psbJ*, as well as genes like *clpP1* and *ycf1*, consistently exhibit elevated $\pi$ values. Such regions outperform conventional barcoding loci like *matK* and *rbcL* in resolving close taxonomic relationships, making them valuable candidates for phylogenetic and population genetic analyses (Li et al. 2023).

For example, the *ndhG–ndhA* region ($\pi \approx 0.015$) encompasses genes implicated in the chloroplast NADH dehydrogenase complex, associated with stress responses and fluctuating redox states, suggesting potential roles in adaptation to fragmented and variable environments such as Java's disturbed lowland forests (Palmer 1985).

Collectively, these highly variable loci – including *ycf1*, *rpl20–clpP1*, and *ndhG–ndhA* – emerge as promising molecular markers for future taxonomic resolution, phylogeographic reconstruction, and conservation prioritization within *Zingiber* and closely related genera.

### Insight for conservation efforts

The pronounced genetic conservation between *Z. odoriferum* and its close relatives (*Z. smilesianum*) underscores the value of coordinated, cross-species conservation frameworks. Shared genomic architecture implies that protecting interconnected habitats, particularly Java's dwindling tropical forests – where these species coexist – is critical to mitigating biodiversity loss. However, subtle genomic distinctions, such as species-specific IR border dynamics and hypervariable loci (e.g. ycf1), demand tailored interventions to safeguard unique adaptive traits that may enhance survival under environmental stress. Conservation plans should synergize *ex situ* approaches (e.g. targeted seed banking and curated cultivation in botanic gardens – with *in situ* habitat restoration, prioritizing populations exhibiting high genetic diversity to buffer against ecological disruptions. The hypervariable regions identified here (e.g. *rpl20-clpP1, ndhG-ndhA*) offer actionable tools for real-time genetic monitoring, enabling rapid detection of diversity erosion or inbreeding in wild populations. To maximize impact, these genomic insights should inform policy frameworks advocating for the preservation of fragmented forest ecosystems while integrating Indigenous ecological knowledge into adaptive management practices. By anchoring conservation in genomic evidence, stakeholders can holistically address extinction risks while preserving the evolutionary flexibility essential for long-term species resilience.

## Conclusions

This study presents the first complete, annotated chloroplast genome of *Zingiber odoriferum*, a critically endangered ginger species endemic to Java, Indonesia. The cp genome exhibits a conserved quadripartite structure, with notable features including pseudogenized *ycf1*, codon usage biased toward A/T-rich codons, and a repertoire of SSRs and long repeats enriched in non-coding regions. Comparative analysis reveals high structural conservation across *Zingiber* species, particularly in coding regions and IR junctions, but identifies hypervariable intergenic spacers (*rpl20/clpP1, ndhG/ndhA, ycf1*) as potential molecular markers for species authentication and population studies. Phylogenetically, *Z. odoriferum* forms a well-supported clade with *Z. teres* and *Z. smilesianum*, underscoring their close evolutionary relationship.

Despite this conservation, the limited cp genome diversity within *Zingiber* poses challenges for resolving fine-scale phylogenetic relationships and designing conservation strategies. This highlights the necessity of integrating nuclear genomic data to capture cryptic diversity and adaptive potential. The coexistence of conserved core genes and hypervariable non-coding regions reflects dual evolutionary forces: strong purifying selection on essential functions and relaxed constraints on structural plasticity.

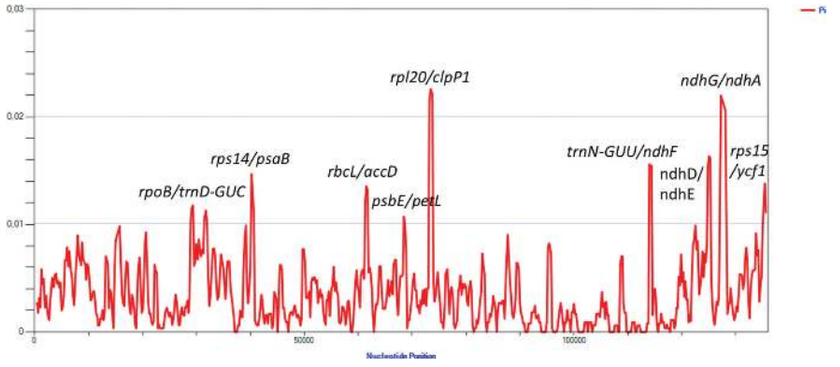



Figure 11. Sliding window analysis of the *Z. odoriferum* cp genome and its close related *Zingiber*. The red line indicates the nucleotide diversity across sample. The threshold was set as 0.01, indicated by blue line.

Future studies should prioritize transcriptomic assays to validate the functional roles of variable loci (e.g. *ycf1* expression) and explore recombination dynamics at IR-SSC junctions to elucidate mechanisms driving genome evolution. By coupling cp genomics with nuclear markers and ecological data, researchers can refine conservation frameworks to preserve *Z. odoriferum*'s genetic diversity and enhance its resilience in rapidly changing tropical ecosystems. This work underscores the urgency of multi-omics approaches to safeguard critically endangered species and their habitats.

*Acknowledgements* – The authors sincerely thank the Research Center for Applied Botany and the Bogor Botanic Garden management for their invaluable support and facilitation throughout this study. We extend our gratitude to Khalif, Rani, Nadhila, and Nisrina for their dedicated efforts in sample collection and data recording. Special acknowledgment is given to MAHAMERU High Performance Computing at BRIN for providing essential computational resources. Finally, we deeply appreciate Mr Arief Priyadi for his expert guidance and insights in bioinformatics analysis.
*Funding* – This research was funded by the 2023 Rumah Program Research: Konservasi Tumbuhan Terancam Kepunahan, organized by OR Hayati dan Lingkungan – BRIN, under contract number 10/III.5/HK/2023.

### Author contributions

**Irfan Martiansyah**: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Supervision (equal); Writing – original draft (equal); Writing – review and editing (equal). **Risha Amilia Pratiwi**: Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Visualization (equal); Writing – review and editing (equal). **Muhammad Rifqi Hariri**: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Supervision (equal); Writing – review and editing (equal). **Mahat Magandhi**: Validation (equal); Writing – review and editing (equal). **Asih Perwita Dewi**: Validation (equal); Writing – review and editing (equal). **Andik Wijayanto**: Validation (equal); Writing – review and editing (equal).

### Data availability statement

All data supporting the findings of this study have been deposited in public repositories. The complete chloroplast genome sequence of *Zingiber odoriferum* has been submitted to GenBank under accession number PQ550047. The raw sequencing data have been submitted to the NCBI Sequence Read Archive (SRA) under BioProject ID PRJNA1243327, BioSample ID SAMN47618771, and SRA accession number SRR32289881.

Data are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.pk0p2nh23 (Martiansyah et al. 2025).

## References

Allendorf, F. W. 2017. Genetics and the conservation of natural populations: allozymes to genomes. – Mol. Ecol. 26: 420–430.
Amiryousefi, A., Hyvönen, J. and Poczai, P. 2018. IRscope: an online program to visualize the junction sites of chloroplast genomes. – Bioinformatics 34: 3030–3031.
Andargie, M. and Congyi, Z. 2022. Genome-wide analysis of codon usage in sesame (*Sesamum indicum* L.). – Heliyon 8: e08687.
Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data. – https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
Ardiyani, M., Newman, M. F. and Poulsen, A. D. 2017. A new species of *Zingiber* (Zingiberaceae) east of Wallace's line. – Gardens' Bull. Singapore 69: 189–199.
Baeza, M., Sepulveda, D., Cifuentes, V. and Alcaíno, J. 2024. Codon usage bias in yeasts and its correlation with gene expression, growth temperature, and protein structure. – Front. Microbiol. 15: 1414422.
Björkholm, P., Harish, A., Hagström, E., Ernst, A. M. and Andersson, S. G. E. 2015. Mitochondrial genomes are retained by selective constraints on protein targeting. – Proc. Natl Acad. Sci. USA 112: 10154–10161.
Błażej, P., Mackiewicz, D., Wnętrzak, M. and Mackiewicz, P. 2017. The impact of selection at the amino acid level on the usage of synonymous codons. – G3 7: 967–981.
Bolger, A. M., Lohse, M. and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. – Bioinformatics 30: 2114–2120.
Brandies, P., Peel, E., Hogg, C. J. and Belov, K. 2019. The value of reference genomes in the conservation of threatened species. – Genes 10: 846.
Cui, Y., Nie, L., Sun, W., Xu, Z., Wang, Y., Yu, J., Song, J. and Yao, H. 2019. Comparative and phylogenetic analyses of ginger (*Zingiber officinale*) in the family Zingiberaceae based on the complete chloroplast genome. – Plants 8: 283.
Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A. and Davies, R. M. 2021. Twelve years of SAMtools and BCFtools. – GigaScience 10: 1–4.
Daniell, H., Lin, C. S., Yu, M. and Chang, W. J. 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. – Genome Biol. 17: 134.
Darling, A. C. E., Mau, B., Blattner, F. R. and Perna, N. T. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. – Genome Res. 14: 1394–1403.
Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B. and Flouri, T. 2019. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. – Mol. Biol. Evol. 37: 291–294.
Delseny, M. 2003. Towards an accurate sequence of the rice genome. – Curr. Opin. Plant Biol. 6: 101–105.
Deng, M., Yun, X., Ren, S., Qing, Z. and Luo, F. 2022. Plants of the genus Zingiber: a review of their ethnomedicine, phytochemistry and pharmacology. – Molecules 27: 2826.
Dobrogojski, J., Adamiec, M. and Luciński, R. 2020. The chloroplast genome: a review. – Acta Physiol. Plant. 42: 98.
Doyle, J. J. and Doyle, J. L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. – Phytochem. Bull. 19: 11–15.

Drescher, A., Ruf, S., Calsa Jr, T., Carrer, H. and Bock, R. 2000. The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. – Plant J. 22: 97–104.

Du, L., Zhang, C., Liu, Q., Zhang, X. and Yue, B. 2018. Krait: an ultrafast tool for genome-wide survey of microsatellites and primer design. – Bioinformatics 34: 681–683.

Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. and Dubchak, I. 2004. VISTA: computational tools for comparative genomics. – Nucleic Acids Res. 32: W273–W279.

Gao, C., Ren, X., Mason, A. S., Li, J., Wang, W., Xiao, M. and Fu, D. 2013. Revisiting an important component of plant genomes: microsatellites. – Funct. Plant Biol. 40: 645–661.

Greiner, S., Lehwark, P. and Bock, R. 2019. OrganellarGenomeDRAW (OGDRAW) ver. 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. – Nucleic Acids Res. 47: W59–W64.

Harrington, S. M., Wishingrad, V. and Thomson, R. C. 2021. Properties of Markov chain Monte Carlo performance across many empirical alignments. – Mol. Biol. Evol. 38: 1627–1640.

Horiuchi, T. and Aigaki, T. 2006. Alternative trans-splicing: a novel mode of pre-mRNA processing. – Biol. Cell 98: 135–140.

Huang, B. H., Lin, Y. C., Huang, C. W., Lu, H. P., Luo, M. X. and Liao, P. C. 2018. Differential genetic responses to the stress revealed the mutation-order adaptive divergence between two sympatric ginger species. – BMC Genomics 19: 692.

Jacobs, J., Glanz, S., Bunse-Grassmann, A., Kruse, O. and Kück, U. 2010. RNA trans-splicing: identification of components of a putative chloroplast spliceosome. – Eur. J. Cell Biol. 89: 932–939.

Jeffares, D. C., Mourier, T. and Penny, D. 2006. The biology of intron gain and loss. – Trends Genet. 22: 16–22.

Jiang, D., Cai, X., Gong, M., Xia, M., Xing, H., Dong, S., Tian, S., Li, J., Lin, J., Liu, Y. and Li, H. L. 2023. Complete chloroplast genomes provide insights into evolution and phylogeny of *Zingiber* (Zingiberaceae). – BMC Genomics 24: 30.

Kahraman, K. and Lucas, S. J. 2019. Comparison of different annotation tools for characterization of the complete chloroplast genome of *Corylus avellana* cv Tombul. – BMC Genomics 20: 874.

Kress, W. J., Prince, L. M. and Williams, K. J. 2002. The phylogeny and a new classification of the gingers (Zingiberaceae): evidence from molecular data 89: 1682–1696.

Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J. and Giegerich, R. 2001. REPuter: the manifold applications of repeat *American Journal of Botany* analysis on a genomic scale. – Nucleic Acids Res. 29: 4633–4642.

Kuznetsov, D. and Bollin, C. 2021. NCBI genome workbench: an integrated platform for sequence analysis. – Bioinformatics 37: 514–515.

Langmead, B. and Salzberg, S. L. 2012. Fast gapped-read alignment with Bowtie 2. – Nat. Methods 9: 357–359.

Larsen, K., Lock, J. M., Maas, H. and Maas, P. J. M. 1998. Zingiberaceae. – In: Kubitzki, K., (ed.), The families and genera of vascular plants. IV. Flowering Plants· Monocotyledons: Alismatanae and Commelinanae (except Gramineae). Springer, pp. 474–495.

Li, D. M., Ye, Y. J., Xu, Y. C., Liu, J. M. and Zhu, G. F. 2020. Complete chloroplast genomes of *Zingiber montanum* and *Zingiber zerumbet*: genome structure, comparative and phylogenetic analyses. – PLoS One 15: e0236590.

Li, D. M., Li, J., Wang, D. R., Xu, Y. C. and Zhu, G. F. 2021. Molecular evolution of chloroplast genomes in subfamily Zingiberoideae (Zingiberaceae). – BMC Plant Biol. 21: 558.

Li, D. M., Liu, H. L., Pan, Y. G., Yu, B., Huang, D. and Zhu, G. F. 2023. Comparative chloroplast genomics of 21 species in Zingiberales with implications for their phylogenetic relationships and molecular dating. – Int. J. Mol. Sci. 24: 15031.

Liu, M. L., Fan, W. B., Wang, N., Dong, P. B., Zhang, T. T., Yue, M. and Li, Z. H. 2018. Evolutionary analysis of plastid genomes of seven *Lonicera* L. species: implications for sequence divergence and phylogenetic relationships. – Int. J. Mol. Sci. 19: 4039.

Liu, S., Ni, Y., Li, J., Zhang, X., Yang, H., Chen, H. and Liu, C. 2023. CPGView: a package for visualizing detailed chloroplast genome structures. – Mol. Ecol. Resour. 23: 694–704.

Lynch, M. and Jarrell, P. 1993. A method for calibrating molecular clocks and its application to animal mitochondrial DNA. – Genetics 135: 1197–1208.

Martiansyah, I., Pratiwi, R. A., Hariri, M. R., Magandhi, M., Dewi, A. P. and Wijayanto, A. 2025. Data from: Chloroplast genome of the critically endangered ginger *Zingiber odoriferum* Blume from Java, Indonesia: characterization, comparison, and conservation insights. – Dryad Digital Repository, https://doi.org/10.5061/dryad.pk0p2nh23.

Miao, Y., Chen, L., Li, H., Zhou, D., Pan, Y. and Liu, T. 2021. Characterization of the chloroplast genome of Zingiber mioga (Thunb.) Rosc. – Mitochondrial DNA B 6: 318–319.

Minh, B. Q., Nguyen, M. A. T. and Von Haeseler, A. 2013. Ultrafast approximation for phylogenetic bootstrap. – Mol. Biol. Evol. 30: 1188–1195.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A. and Lanfear, R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. – Mol. Biol. Evol. 37: 1530–1534.

Okonechnikov, K., Golosova, O., Fursov, M. and UGENE Team 2012. Unipro UGENE: a unified bioinformatics toolkit. – Bioinformatics 28: 1166–1167.

Palmer, J. D. 1985. Comparative organization of chloroplast genomes. – Annu. Rev. Genet. 19: 325–354.

Panja, A. S., Maiti, S. and Bandyopadhyay, B. 2020. Protein stability governed by its structural plasticity is inferred by physicochemical factors and salt bridges. – Sci. Rep. 10: 1822.

Qu, X. J., Moore, M. J., Li, D. Z. and Yi, T. S. 2019. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. – Plant Methods 15: 50.

Qu, X. J., Zou, D., Zhang, R. Y., Stull, G. W. and Yi, T. S. 2023. Progress, challenge and prospect of plant plastome annotation. – Front. Plant Sci. 14: 1166140.

Rellstab, C., Dauphin, B. and Exposito-Alonso, M. 2021. Prospects and limitations of genomic offset in conservation management. – Evol. Appl. 14: 1202–1212.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L. and Darling, A. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. – Syst. Biol. 61: 539–542.

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E. and Sánchez-Gracia, A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. – Mol. Biol. Evol. 34: 3299–3302.

Sabu, M., Lofthus, Ø., Newman, M. F. and Poulsen, A. D. 2019. *Zingiber odoriferum*. – The IUCN Red List of Threatened Species 2019, p. e.T117468374A124284927.
Sharp, P. M. and Li, W. H. 1987. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. – Nucleic Acids Res. 15: 1281–1295.
Shaw, J., Lickey, E. B., Beck, J. T., Farmer, S. B., Liu, W., Miller, J., Siripun, K. C., Winder, C. T., Schilling, E. E. and Small, R. L. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. – Am. J. Bot. 92: 142–166.
Shi, L., Chen, H., Jiang, M., Wang, L., Wu, X., Huang, L. and Liu, C. 2019. CPGAVAS2, an integrated plastome sequence annotator and analyzer. – Nucleic Acids Res. 47: W65–W73.
Stothard, P. 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. – BioTechniques 28: 1102–1104.
Stöver, B. C. and Müller, K. F. 2010. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. – BMC Bioinformatics 11: 7.
Syafiq, A. K. A. 2024. Analisis keragaman morfologi dan genetik tanaman jahe (*Zingiber odoriferum* Blume.) menggunakan marka ISSR. – Skripsi S1, Universitas Sultan Ageng Tirtayasa.
Tamura, K., Stecher, G. and Kumar, S. 2021. MEGA11: molecular evolutionary genetics analysis version 11. – Mol. Biol. Evol. 38: 3022–3027.
The Galaxy Community. 2024. The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. – Nucleic Acids Res. 52: W83–W94.
Theerakulpisut, P., Triboun, P., Mahakham, W., Maensiri, D., Khampila, J. and Chantaranothai, P. 2012. Phylogeny of the genus Zingiber (Zingiberaceae) based on nuclear ITS sequence data. – Kew Bull. 67: 389–395.
Theilade, I. 1999. A synopsis of the genus Zingiber (Zingiberaceae) in Thailand. – Nord. J. Bot. 19: 389–410.
Trovato, M., Funck, D., Forlani, G., Okumoto, S. and Amir, R. 2021. Amino acids in plants: regulation and functions in development and stress defense. – Front. Plant Sci. 12: 772810.
Vaughn, J. N., Chaluvadi, S. R., Tushar, R., Rangan, L. and Bennetzen, J. L. 2014. Whole plastome sequences from five ginger species facilitate marker development and define limits to barcode methodology. – PLoS One 9: e108581.
Wang, Z., Du, S., Dayanandan, S., Wang, D., Zeng, Y. and Zhang, J. 2014. Phylogeny reconstruction and hybrid analysis of *Populus* (Salicaceae) based on nucleotide sequences of multiple single-copy nuclear genes and plastid fragments. – PLoS One 9: e103645.
Wang, Y., Yang, C., Jin, Q., Zhou, D., Wang, S., Yu, Y. and Yang, L. 2015. Genome-wide distribution comparative and composition analysis of the SSRs in Poaceae. – BMC Genet. 16: 18.
Wang, X., Wang, J., Li, S., Lu, C. and Sui, N. 2022. An overview of RNA splicing and functioning of splicing factors in land plant chloroplasts. – RNA Biol. 19: 897–907.
Wang, J., Kan, S., Kong, J., Nie, L., Fan, W., Ren, Y., Reeve, W., Mower, J. P. and Wu, Z. 2024. Accumulation of large lineage-specific repeats coincides with sequence acceleration and structural rearrangement in Plantago plastomes. – Genome Biol. Evol. 16: evae177.
Wicke, S., Schneeweiss, G. M., dePamphilis, C. W., Müller, K. F. and Quandt, D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. – Plant Mol. Biol. 76: 273–297.
Wickett, N. J., Forrest, L. L., Budke, J. M., Shaw, B. and Goffinet, B. 2011. Frequent pseudogenization and loss of the plastid-encoded sulfate-transport gene cysA throughout the evolution of liverworts. – Am. J. Bot. 98: 1263–1275.
Wu, M. L., Li, Q., Xu, J. and Li, X. W. 2018. Complete chloroplast genome of the medicinal plant Amomum compactum: gene organization, comparative analysis, and phylogenetic relationships within Zingiberales. – Chin. Med. 13: 10.
Xia, M., Jiang, D., Xu, W., Liu, X., Zhu, S., Xing, H., Zhang, W., Zou, Y. and Li, H. L. 2024. Comparative chloroplast genome study of Zingiber in china sheds light on plastome characterization and phylogenetic relationships. – Genes 15: 1484.
Yang, Q., Xin, C., Xiao, Q. S., Lin, Y. T., Li, L. and Zhao, J. L. 2023a. Codon usage bias in chloroplast genes implicates adaptive evolution of four ginger species. – Front. Plant Sci. 14: 1304264.
Yang, S., Li, G. and Li, H. 2023b. Molecular characterizations of genes in chloroplast genomes of the genus *Arachis* L. (Fabaceae) based on the codon usage divergence. – PLoS One 18: e0281843.
Zaman, W., Ayaz, A. and Park, S. 2025. Integrating morphological and molecular data in plant taxonomy. – Pak. J. Bot. 57: 4.
Zeng, Y., Shen, L., Chen, S., Qu, S. and Hou, N. 2023. Codon usage profiling of the chloroplast genome in Juglandaceae. – Forests 14: 378.
Zhang, M. Q. 2002. Computational prediction of eukaryotic protein-coding genes. – Nat. Rev. Genet. 3: 698–709.
Zheng, S., Poczai, P., Hyvönen, J., Tang, J. and Amiryousefi, A. 2020. Chloroplot: an online program for the versatile plotting of organelle genomes. – Front. Genet. 11: 576124.