

Comparative Chloroplast Analysis of *Kjellbergiodendron celebicum*, an Endangered Plant from Sulawesi

Muh Restu

Hasanuddin University

Tri Retnaningsih Soeprbowati

Diponegoro University

Syarif Prasetyo

National Research and Innovation Agency

Nurul Hikma

Hasanuddin University

Lukaz Kruszewski

Polish Academy of Sciences

Asrianny Asrianny

Hasanuddin University

Sri Wahyuni Jufri

Hasanuddin University

Irfan Martiansyah

National Research and Innovation Agency

Siti Halimah Larekeng

sittihalimah@unhas.ac.id

Hasanuddin University

Research Article

Keywords: *Kjellbergiodendron celebicum*, chloroplast genome, Myrtaceae, phylogenetics, endemic species, conservation

Posted Date: October 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10988348/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Kjellbergiodendron celebicum (Koord.), locally known as Tombeuwa, is an endemic tree of the family Myrtaceae restricted to Sulawesi, Indonesia, and is valued as a high-quality timber. Increasing exploitation, a naturally small population size, and the complete absence of genomic resources place the species at risk of decline and potential extinction. This study aimed to characterize the complete chloroplast (cp) genome of *K. celebicum* and to clarify its phylogenetic position within Myrtaceae.

Results

We sequenced the complete chloroplast genome of *K. celebicum* on an Illumina platform and assembled it into a single circular molecule of 161,149 bp, with an average sequencing depth of 4,082.67× and no uncovered bases. The genome exhibited the typical quadripartite structure, comprising a large single-copy (LSC) region of 89,189 bp, a small single-copy (SSC) region of 18,902 bp, and a pair of inverted repeats (IR) of 26,529 bp each, with an overall GC content of 37%. Annotation identified 85 protein-coding genes, 37 tRNA genes, and 8 rRNA. Codon usage analysis revealed a clear bias toward A/T-ending codons, with leucine the most frequently encoded amino acid. Dispersed-repeat analysis detected 73 palindromic, 20 forward, 4 reverse, and 3 complementary repeats and identified 182 simple sequence repeats (SSRs), predominantly mononucleotide A/T motifs. Comparative analysis with nine additional Myrtaceae chloroplast genomes (156,129–161,149 bp) showed highly conserved genome structure and gene content, with only minor expansion and contraction at the IR/SC boundaries. Phylogenetic reconstruction placed *K. celebicum* in a strongly supported clade (bootstrap = 100) with *Lophostemon confertus* and *Xanthostemon chrysanthus*.

Conclusion

The chloroplast genome reported here is the first genomic resource for the genus *Kjellbergiodendron*, providing valuable data for evolutionary studies, molecular-marker development, and conservation strategies for this endemic Sulawesi species.

Introduction

Sulawesi is one of Indonesia's biodiversity centers, harboring numerous endemic species of high ecological and genetic value [1]. One such endemic species is *Kjellbergiodendron celebicum* (Koord.), known locally as Tombeuwa. This shrub-to-tree plant of the family Myrtaceae is native to the humid tropical biome of the Sulawesi region [2]. Efforts to develop and conserve *K. celebicum* require baseline information on its genetic diversity, given that the species has a relatively small population size and is therefore vulnerable to future losses in genetic diversity [3]. Genetic variation within a population is a key

indicator of a species' adaptability and resilience to environmental pressures [4]. This factor plays an important role in conservation, as it determines a population's capacity to persist and recover from environmental disturbance and habitat degradation [5]. Populations with low levels of genetic diversity tend to be more susceptible to declines in individual numbers and face a high risk of extinction. Therefore, information on genetic diversity is essential for supporting sustainable species management and conservation strategies [6]. The level of genetic diversity within a population can also reflect the condition and sustainability of plant species in the wild. Populations with high genetic diversity generally have a greater chance of survival because they can adapt better to environmental change [7]. Conventional evaluation of plant genetic diversity has generally relied on morphological approaches, such as progeny trials, provenance trials, and other phenotype-based assessments [3]. However, such approaches are often influenced by environmental factors. Integrating molecular techniques into biodiversity studies not only enhances understanding of genetic variability and phylogenetic relationships but also provides a stronger scientific basis for formulating conservation strategies and sustainable resource use [8].

K. celebicum found in the Lake Matano area is an endemic species that requires conservation efforts as well as genetic improvement. Previous studies on the genetic diversity of Tombeuwa used RAPD and ISSR markers. However, next-generation sequencing (NGS) approaches can provide higher resolution for revealing genomic variation more thoroughly and comprehensively [9]. Plant genetic data can be obtained from the nuclear, mitochondrial, and chloroplast genomes. The nuclear genome contains a large amount of genetic information in a DNA sequence of nearly 5 million base pairs [10]. Inherited biparentally, the nuclear genome is important for understanding the genes that regulate various cellular functions, such as gene expression, DNA repair, and genome stability. The chloroplast (cp) and mitochondrial genomes consist of relatively stable circular DNA molecules that contain essential genes, including those involved in photosynthesis, and provide important information regarding phylogenetic relationships and the evolutionary dynamics of species [11]. The chloroplast genome is maternally inherited and has a relatively conserved structure and gene content among related species, making it highly useful for comparative analyses across various taxonomic groups, even cryptic ones [12]. The plant mitochondrial genome is also maternally inherited and is involved in complex interactions with the nuclear and chloroplast genomes, including through horizontal gene transfer events. Mitochondrial genome studies have been carried out extensively across various plant species, including *Neolamarckia macrophylla* [13]. To date, however, neither partial nor comprehensive genomic data, whether nuclear, chloroplast, or mitochondrial, are available for the genus *Kjellbergiodendron*, which represents a critical knowledge gap that constrains effective conservation planning and heightens the risk of extinction for this endemic species.

Advances in bioinformatics tools and methods have facilitated the efficient and accurate assembly and annotation of DNA sequences [14]. Characterization of the nuclear, chloroplast, and mitochondrial genomes can serve as a basis for developing molecular markers that contribute substantially to the validity of identifying genetic diversity and variation in plants [15]. Markers commonly used to detect genetic diversity include SSR (Simple Sequence Repeat) and SNAP (Single Nucleotide Amplified

Polymorphism) markers. In recent years, SNAP and SSR markers have been widely used to examine genetic diversity in species such as salak (snake fruit), sugar palm, and coconut [11, 16, 17]. Research on the genetic diversity of *K. celebicum* is therefore important and will greatly assist breeding and conservation efforts, as such genetic diversity data are among the factors used to select accessions with high genetic variation for producing superior accessions. However, in the absence of a reference chloroplast genome, marker development and genetic diversity assessment for *K. celebicum* cannot yet be reliably undertaken. This study therefore aimed to characterize the complete chloroplast genome of *K. celebicum* and to clarify its phylogenetic position within Myrtaceae, providing the first genomic resource for the genus *Kjellbergiodendron* and a foundation for future molecular marker development and conservation strategies for this endemic Sulawesi species.

Methods

Sample Collection and DNA Sequencing

Leaf samples of *K. celebicum* (Tombeuwa) were collected from its natural habitat in the Lake Matano area, South Sulawesi, Indonesia. Sample collection was carried out under Contract No. 4600082668 which authorized the collection of plant material within the PT Vale Indonesia concession area at Lake Matano. Formal taxonomic identification of the plant material was verified by T.R. Soeprbowati, S. Prasetyo, and I. Matiansyah. A voucher specimen (No. MTN 001), collected by Jumari et al. on 19 June 2025, was deposited at the Integrated Laboratory, CPalm (Cluster for Paleolimnology), Faculty of Sciences and Mathematics, Diponegoro University. Fresh leaves were transported to PT Genetika Science Indonesia for DNA extraction and sequencing. Total genomic DNA was extracted and sequenced on an Illumina platform to generate 150 bp paired-end reads (Illumina PE150). For comparative and phylogenetic analyses, complete chloroplast genome sequences of additional Myrtaceae species were retrieved from the NCBI GenBank database.

Genome Assembly and Annotation

The quality of raw sequencing reads was assessed using FastQC (v0.12.1) (Wingett dan Andrews 2018). Adapter sequences and low-quality bases were removed using fastp (v0.23.4) with automatic adapter detection for paired end reads, a minimum read length of 50 bp, and a quality threshold of Q15. High-quality clean reads were retained for downstream analyses. De novo assembly of the chloroplast genome was performed using GetOrganelle (v1.8.0.1) pipeline (<https://github.com/Kinggerm/GetOrganelle>) [18]. The assembled chloroplast genome sequences were uploaded to GeSeq within CHLOROBX[19] for annotation, employing 95% sequence similarity thresholds for protein-coding genes, rRNAs, and tRNAs. BLASTX and BLASTN were used to identify coding genes and rRNAs, while tRNAscan-SE v2.0.7. was used for tRNA annotation. The circular genome maps were drawn by the OGDRAW program (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>). The annotated sequences were manually inspected and curated in Geneious Prime to correct potential gene

boundary inconsistencies. BWA (v.0.7.17-r1188) was used to process comparison data, SAMtools (v.1.9) was used to calculate coverage depth, and then visualized using ggplot2 in R.

Codon Usage Bias and Repeat Sequence Analysis

Protein-coding gene (PCG) sequences were extracted from the annotated GenBank (.gb) files. The extracted coding sequences were subsequently used for relative synonymous codon usage (RSCU) analysis, which was calculated using CodonW (v1.4.4); the resulting RSCU values were further processed and visualized using RStudio, employing the ggplot2 dplyr, tidyr, and patchwork packages. Forward (f), reverse (r), complementary (c), and palindromic (p) repeats in the chloroplast genomes were identified using the web-based REPuter tool with default parameters, detecting repeat units ranging from 30 to 100 bp. Simple sequence repeats (SSRs), also referred to as microsatellites, were identified using the online MISA (v2.1) tool. The minimum repeat thresholds were set to 10 for mononucleotide, 5 for dinucleotide, 4 for trinucleotide, and 3 for tetra-, penta-, and hexanucleotide repeats, respectively. Compound SSRs were detected by allowing a maximum interruption length of 50 bp between two adjacent SSRs to be classified as a compound repeat.

Comparative Analysis of the Chloroplast Genome

The expansion and contraction at the boundaries between the single-copy (LSC and SSC) and inverted repeat (IR) regions of the chloroplast genomes were analyzed using the web-based IRscope tool (<https://irscope.shinyapps.io/IRplus/>). IRscope was used to locate, visualize, and compare gene positions at the LSC/IRb, IRb/SSC, SSC/IRa, and IRa/LSC junctions among the analyzed *Kjellbergiodendron* chloroplast genomes. The online tool mVISTA [20] (<https://genome.lbl.gov/vista/mvista/submit.shtml> (accessed on 2 february 2026)) was used to compare the complete chloroplast genomes of the six species in Myrtaceae with the annotation of *K. celebicum* as the reference.

Nucleotide Polymorphism

Complete chloroplast genome sequences of *Kjellbergiodendron celebicum* were aligned using MAFFT implemented in Geneious Prime. The resulting multiple sequence alignment was exported in NEXUS format and used for downstream analyses. After that, sliding window analysis was performed using DnaSP v5 [21] to compute nucleotide diversity (Pi) values across the chloroplast genomes. The analysis employed a window size of 600 bp and a step size of 200 bp, enabling the identification of highly variable regions. We evaluated these regions for their potential as molecular markers for phylogenetic studies.

Phylogenetic analysis

The same aligned dataset was subsequently used for phylogenetic reconstruction. Phylogenetic relationships among Myrtaceae species were inferred using MEGA based on the Maximum Likelihood (ML) method with 1,000 bootstrap replicates to assess branch support. In addition, a median-joining (MJ) network analysis was performed to further examine haplotype relationships among the analyzed chloroplast genomes.

Result

Sequencing Depth and Genome Coverage

To evaluate the quality of the chloroplast genome assembly for *Kjellbergiodendron celebicum*, sequencing depth and genome coverage were analyzed (Fig. 1). The complete chloroplast genome spanned a total length of 161,149 bp and exhibited comprehensive coverage, with an average sequencing depth of 4,082.67×. The sequencing depth ranged from a minimum of 24× to a maximum of 6,692×, indicating robust read support throughout the entire sequence. No uncovered bases were detected, confirming the completeness of the assembly. Although minor local fluctuations in read depth were observed along the genome, these variations did not produce any coverage gaps and are likely attributable to inherent differences in sequencing efficiency or regional sequence composition.

Chloroplast genome features of *Kjellbergiodendron celebicum*

The complete chloroplast genome of *Kjellbergiodendron celebicum* displayed the typical quadripartite structure, consisting of a large single-copy (LSC) region of 89,189 bp, a small single-copy (SSC) region of 18,902 bp, and a pair of inverted repeat (IR) regions of 26,529 bp each, giving a total genome size of 161,149 bp (Fig. 2). The overall GC content was 37%, with the IR regions showing the highest GC content, followed by the LSC and SSC regions. This pattern reflects the enrichment of rRNA genes within the IR regions and is characteristic of angiosperm chloroplast genomes.

Genome annotation identified 85 protein-coding genes, 37 transfer RNA (tRNA) genes, and 8 ribosomal RNA (rRNA) genes (Fig. 2). Genes located on the inner side of the map are transcribed clockwise, whereas those on the outer side are transcribed anticlockwise. Several genes within the IR regions are present as duplicated copies, including all rRNA genes and several protein-coding and tRNA genes, contributing to both the elevated GC content and the structural stability of these regions. Detailed information on gene identity, copy number, and regional distribution is provided in Table 1.

Table 1
Genes identified in the chloroplast genome of *Kjellbergiodendron celebicum*.

Gene related to	Genes' group	Gene names
Photosynthesis	Photosystem I	<i>psaA, psaB, psaJ, psal, psaC</i>
	Photosystem II	<i>psbA, psbB, psbC, psbE, psbF, psbH, psbl, psbL, psbM, psbN, psbT, psbZ</i>
	Cytochrome b6/f complex	<i>petA, petB, petD, petG, petL, petN</i>
	ATP synthase	<i>atpA, atpB, atpE, atpF, atpH, atpI</i>
	RUBISCO	<i>rbcL</i>
	NADH oxidoreductase	<i>ndhA, ndhB×2, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
	Cytochrome c synthesis	<i>ccsA</i>
	Assembly	<i>pafl*, pafII</i>
Transcription and translation	Ribosomal proteins (Large subunit)	<i>rpl2×2, rpl14, rpl16, rpl20, rpl22, rpl23×2, rpl32, rpl33, rpl36</i>
	Ribosomal proteins (Small subunit)	<i>rps2, rps3, rps4, rps7, rps8, rps12×3, rps14, rps15, rps16, rps18, rps19</i>
	Transcription	<i>rpoC1, rpoC2, rpoA, rpoB</i>
	Ribosomal RNAs	<i>rrn16×2, rrn23×2, rrn4.5×2, rrn5×2</i>
	Transfer RNAs	<i>trnH-GUG, trnK-UUU, trnQ-UUG, trnS-GCU, trnG-UCC, trnR-UCU, trnC-GCA, trnD-GUC, trnY-GUA, trnE-UUC, trnG-UCC, trnT-GGU, trnG-GCC, trnS-GGA, trnT-UGU, trnL-UAA×4, trnF-GAA, trnV-UAC, trnW-CCA, trnA-UGC×4, trnR-CCG, trnL-CAA×2, trnV-GAC×2, trnT-GAU×2, trnA-UGC×2, trnR-ACG×2, trnN-GUU×2, trnL-UAG</i>
Other genes	Carbon metabolism	<i>cemA</i>
	RNA processing	<i>matK</i>
	Proteolysis	<i>clpP*</i>
	Fatty acid synthesis	<i>accD</i>
	ORF	<i>ycf1, ycf2×2</i>
	Pseudogene	<i>ψycf1</i>

Repetitive Sequence

Analysis of dispersed repetitive sequences in the *Kjellbergiodendron celebicum* chloroplast genome identified a total of 100 repeats, comprising 73 palindromic, 20 forward, 4 reverse, and 3 complementary repeats (Fig. 3A). Palindromic and forward repeats were therefore the dominant repeat types, whereas reverse and complementary repeats were comparatively rare. Microsatellite (SSR) analysis identified 182 SSRs in total (Fig. 3B). Mononucleotide repeats were by far the most abundant type, accounting for 51.1% of all SSRs, followed by trinucleotide (36.3%), tetranucleotide (7.7%), dinucleotide (3.3%), and pentanucleotide (1.6%) repeats. The mononucleotide SSRs were strongly biased toward A/T, with poly-T (55) and poly-A (38) being the most frequent motifs, consistent with the AT-rich nature of plastid genomes.

Codon usage bias

Codon usage analysis based on relative synonymous codon usage (RSCU) values revealed a conserved codon usage pattern across the protein-coding genes of the *Kjellbergiodendron celebicum* chloroplast genome (Fig. 4). Most amino acids were encoded by multiple synonymous codons with unequal usage frequencies, as reflected by RSCU values deviating from one, whereas methionine (AUG) and tryptophan (UGG), each encoded by a single codon, showed RSCU values of 1.0. A clear preference for A/T-ending codons was observed: codons such as ATT (isoleucine), TTA (leucine), GAT (aspartate), AAT (asparagine), and TAT (tyrosine) had RSCU values greater than one, whereas G/C-ending codons such as GCG (alanine), CGC (arginine), AGC (serine), and TGC (cysteine) generally had RSCU values below one. At the amino acid level, leucine was the most abundant amino acid, followed by isoleucine and serine, while cysteine was the least represented.

IR/SC junction analysis

Figure 5 illustrates the comparison of IR/SC junctions among the six Myrtaceae chloroplast genomes, highlighting the expansion and contraction patterns at the LSC/IRb (JLB), IRb/SSC (JSB), SSC/IRa (JSA), and IRa/LSC (JLA) boundaries. Overall, the positions of genes flanking the junction regions were largely conserved among the examined Myrtaceae species, with only minor variations in gene length and boundary distances. At the JLB (LSC–IRb junction), the *rps19* gene was consistently located at the boundary, partially extending into the IRb region across all Myrtaceae species, indicating a conserved IR expansion pattern. Minor differences in the length of the *rps19* fragment within the IRb region were observed among species, reflecting slight boundary shifts. At the JSB (IRb–SSC junction), the *ndhF* gene spanned the junction, with small variations in its length and overlap between the IRb and SSC regions among Myrtaceae species. Similarly, variation was detected at the JSA (SSC–IRa junction), where the *ycf1* gene crossed the boundary and showed species-specific differences in the length of the *ycf1* fragment extending into the IRa region. At the JLA (IRa–LSC junction), the *trnH* gene was located in the LSC region at a short distance from the IRa boundary.

Nucleotide Polymorphisms

Sequence divergence was assessed from the aligned Myrtaceae chloroplast genomes using sliding-window analysis of nucleotide diversity (π ; window 600 bp, step 200 bp) (Fig. 6). The overall nucleotide diversity was low, consistent with the generally conserved nature of chloroplast genomes, with a mean π of approximately 0.037. However, variation was highly heterogeneous across the genome. A single, pronounced high-diversity region (π up to ~ 0.22 – 0.25) was located within the small single-copy (SSC) region, while the LSC region showed several smaller, scattered peaks, and the IR regions remained highly conserved. This prominent SSC hotspot represents the most variable region of the plastome and is a strong candidate for the development of molecular markers.

Comparative Analysis

We compared sequence divergence and structural conservation using mVISTA, with *K. celebicum* (Sulawesi) as the reference. The alignment revealed a high degree of sequence similarity among the Myrtaceae chloroplast genomes, with the inverted-repeat (IR) regions more conserved than the LSC and SSC regions, and coding regions more conserved than non-coding regions. Several divergence hotspots were observed, primarily in intergenic spacer regions and in the *ycf1* region, which may serve as candidate molecular markers for species identification and population-genetic studies (Fig. 7).

Phylogenetic analysis

The maximum-likelihood (ML) phylogenetic tree based on complete chloroplast genomes is shown in Fig. 8, with *Lagerstroemia speciosa* and *Melastoma candidum* as outgroups. All nodes received maximum bootstrap support (BS = 100). *K. celebicum* was resolved within Myrtaceae as the sister lineage to a strongly supported clade comprising *Lophostemon confertus* and *Xanthostemon chrysanthus*. The remaining taxa formed well-supported lineages, including a *Syzygium* clade and a large clade containing *Eucalyptus*, *Corymbia*, *Angophora*, and related genera, consistent with established relationships within the family. This represents the first placement of the genus *Kjellbergiodendron* in a plastome-based phylogeny of Myrtaceae.

Discussion

Analysis of chloroplast (cp) genome sequences provides extensive information for plant phylogenetic studies, and DNA barcoding based on cp markers enables precise species identification [22]. In this study, the cp genome of *Kjellbergiodendron celebicum* was sequenced, assembled, and annotated using high-throughput sequencing, and its genome structure, simple sequence repeats (SSRs), codon usage, and phylogenetic position were examined. These results provide the first insights into the evolutionary relationships and plastid genomic features of this endemic Sulawesi species, which are essential for its conservation and can serve as a reference for habitat restoration and the preservation of genetic resources. The complete cp genome of *K. celebicum* is 161,149 bp in length, contains 85 protein-coding

genes, and has a GC content of 37%. When compared with other Myrtaceae chloroplast genomes (156,129–161,149 bp; Fig. 5), it shows a highly similar structure, gene content, and GC content, indicating strong evolutionary conservation within the family.

SSRs are widely distributed in plant genomes, predominantly in non-coding regions, and are commonly used for species identification, genetic-diversity analysis, and marker-assisted breeding [23]. In this study, we identified 182 SSR loci in the cp genome of *K. celebicum*, dominated by mononucleotide A/T motifs. This AT bias suggests that plastid SSRs preferentially contain A and T bases, possibly owing to differences in strand-separation stability. Variation analysis also indicated that the IR regions, which have higher GC content, were more conserved than the LSC and SSC regions, supporting the highly conserved structure of the *K. celebicum* cp genome. The identified SSR loci therefore provide a valuable basis for species identification, phylogenetic studies, and the development of molecular markers.

Codons function as essential links between nucleic acids and proteins and play a key role in genetic-information transfer; patterns of codon usage provide reliable information for studies of gene function and evolution [24]. In the cp genome of *K. celebicum*, leucine was the most abundant amino acid, and the genome showed a clear preference for A/T-ending codons. Such codon usage bias in cp genomes arises from both natural selection and mutational pressure [25].

The IR region of cp genomes is regarded as the most conserved region and plays a critical role in genome stability [26], and expansion, contraction, or loss of IR boundaries can produce structural differences among cp genomes [27]. Comparative analysis of the LSC–IR and SSC–IR junctions among the ten Myrtaceae genomes revealed higher variability in the LSC and SSC regions and lower variability in the IR regions, while overall genome organization remained highly conserved. Minor differences were observed in the boundary positions and lengths of genes such as *rps19*, *ycf1*, *ndhF*, and *trnH*, likely reflecting varying degrees of IR expansion and contraction during evolution. The *rps19* gene is essential for ribosome biogenesis and protein synthesis, while *ycf1* encodes a plastid membrane protein involved in transmembrane transport and stress responses; variation in these genes, together with the divergence hotspots detected in the SSC region, provides useful molecular markers for species identification and phylogenetic analysis within Myrtaceae.

Recent advances in cp genome analysis have substantially improved understanding of Myrtaceae evolution. Our plastome-based phylogeny is consistent with established relationships within the family and clarifies, for the first time, the phylogenetic placement of the genus *Kjellbergiodendron*: *K. celebicum* is most closely related to *Lophostemon confertus* and *Xanthostemon chrysanthus* (Fig. 8).

Conclusion

This study reports the first complete chloroplast genome of *Kjellbergiodendron celebicum*, an endemic Myrtaceae species from Sulawesi, and compares it with other Myrtaceae plastomes. The genome (161,149 bp) showed the conserved quadripartite structure, gene content, and nucleotide composition typical of the family, with minor genome-size variation among species driven mainly by small-scale IR/SC

boundary shifts. Sequence variation was unevenly distributed, with the highest nucleotide diversity concentrated in the SSC region, identifying it as a promising source of chloroplast markers. Codon usage showed a strong A/T bias consistent with mutational pressure shaping plastome evolution. Importantly, plastome-based phylogenetic analysis resolved the position of *K. celebicum* within Myrtaceae with full support, demonstrating the utility of complete chloroplast genomes for evolutionary, taxonomic, and conservation studies of this endemic species.

Abbreviations

cp
chloroplast
LSC
large single-copy
SSC
small single-copy
IR
inverted repeat
bp
base pair
GC
guanine-cytosine
NGS
next-generation sequencing
SSR
simple sequence repeat
RSCU
relative synonymous codon usage
PCG
protein-coding gene
ML
maximum likelihood
MJ
median-joining
BS
bootstrap support
SNAP
single nucleotide amplified polymorphism
RAPD
random amplified polymorphic DNA
ISSR

inter-simple sequence repeat.

Declarations

Ethics approval and consent to participate

This study did not involve human participants, human data, or animal experimentation. Collection of *Kjellbergiodendron celebicum* leaf samples from its natural habitat at Lake Matano, South Sulawesi, Indonesia, was authorized under Contract No. 4600082668 ("Pendampingan Teknis untuk Studi Pemanfaatan Lingkungan" / Technical Assistance for Environmental Utilization Study), Amendment No. 1 (26 November 2024), between PT Vale Indonesia Tbk and the Institute for Research and Community Service (LP2M), Universitas Hasanuddin. Formal taxonomic identification of the plant material was verified by T.R. Soeprbowati, S. Prasetyo, and I. Matiansyah, and a voucher specimen (No. MTN 001) was deposited at the Integrated Laboratory, CPalim, Universitas Diponegoro, Semarang, Indonesia. All fieldwork complied with relevant institutional and national regulations governing the collection of plant material in Indonesia.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

The authors would like to express their sincere gratitude to the Ministry of Higher Education, Science, and Technology of Indonesia (Kemdiktisaintek) for the financial support provided through the Riset Kolaborasi Indonesia (RKI) – research grant scheme B, funded via the BIMA platform (Contract No. 01319/UN4.22/PT.01.03/2025, awarded to Muh. Restu).

Author Contribution

SHL and MR: Research conceptualization, implementation, data analysis, original manuscript preparation, and reviewing. NH, IM, SP, and SWJ: Data analysis and manuscript data preparation. TRS and A: Supervising and reviewing the manuscript. LK: Providing comparative chloroplast genome data and reviewing the manuscript. All authors read and approved the final manuscript.

Acknowledgement

The authors would like to express their sincere gratitude to the Ministry of Higher Education, Science, and Technology of Indonesia (Kemdiktisaintek) for the financial support provided through the Riset Kolaborasi Indonesia (RKI) – research grant scheme B, funded via the BIMA platform (Contract No. 01319/UN4.22/PT.01.03/2025, awarded to Muh. Restu).

Data Availability

The complete chloroplast genome sequence of *Kjellbergiodendron celebicum* generated in this study is being submitted to the NCBI GenBank database, and the accession number will be provided prior to publication. The comparative chloroplast genome sequences of other Myrtaceae species used in this study were retrieved from the NCBI GenBank database.

References

1. Pitopang R, Saleh MFRM, POPULASI, KERAPATAN DAN POLA DISTRIBUSI. *Sarcotheca celebica Veldkamp* (Oxalidaceae), Tumbuhan Endemik Sulawesi Pada Tanah Ultramafik, Di Kabupaten Tojo Una-Una, Sulawesi Tengah. *Ber Biol.* 2025;24:371–81.
2. Priscilia B, Novita H, Elya B. Antioxidant and antibacterial assay against fish pathogen bacteria of *Kjellbergiodendron celebicum* (Koord.) Merr. leaf extract. *Pharmacognosy J.* 2020;12.
3. Larekeng SH, Restu M, Gusmiaty G, Rismawati R. Polymorphism of simple sequence repeat regions of Sulawesi Ebony (*Diosphyros celebica* Bakh.) in experimental forest of Hasanuddin University Provenance. *Agrotech J Universitas Sembilanbelas November Kolaka.* 2016;1:37–44.
4. Larekeng SH, Restu M, Susilowati A, Rachmat HH. Genetic diversity of parental and offspring population in ebony (*Diospyros celebica* bach) revealed by Microsatellites marker. *Int J Emerg Technol.* 2019;10:178–85.
5. Kardos M, Armstrong EE, Fitzpatrick SW, Hauser S, Hedrick PW, Miller JM et al. The crucial role of genome-wide genetic variation in conservation. *Proceedings of the National Academy of Sciences. National Academy of Sciences;* 2021;118:e2104642118.
6. Makmur MF, Larekeng SH, Restu M. Genetic diversity of eight types of *bamboo* based on random amplified polymorphic DNA (rapd) markers. *Plant Arch.* 2020;20:2333–7.
7. Govindaraj M, Vetriventhan M, Srinivasan M. Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. *Genet Res Int Wiley Online Libr.* 2015;2015:431487.
8. Moura M, Dias EF, Belo Maciel MG. Conservation genetics of the highly endangered Azorean endemics *Euphrasia azorica* and *Euphrasia grandiflora* using new SSR data. *Conservation Genetics.* Springer; 2018;19:1211–22.

9. Panahi B, Mohammadzadeh Jalaly H, Hamid R. Using next-generation sequencing approach for discovery and characterization of plant molecular markers. *Curr Plant Biol Elsevier*. 2024;40:100412. <https://doi.org/10.1016/J.CPB.2024.100412>.
10. Chakravartty N, Neelapu NRR. The de novo genome assembly of lemon grass to identify the genes in essential oil production. *J Appl Biol Biotechnol*. 2024;12:100–49. <https://doi.org/10.7324/JABB.2024.125317>.
11. Arshad A, Aditama R, Rahayu Msri, Natawijaya A, Matra Dd, Sudarsono S. Comparative study of chloroplast genomes across seven *Salacca* species. *Biodiversitas*. 2024;25.
12. Thode VA, Lohmann LG. Comparative chloroplast genomics at low taxonomic levels: a case study using *Amphilophium* (Bignoniaceae, Bignoniaceae). *Front Plant Sci Front Media SA*. 2019;10:796.
13. Xie H, Lv Y-W, Liu X-H, Wu C, Wang Z-Y, Lee SY, et al. Sequencing of mitochondrial genome of *Neolamarckia macrophylla* uncovers divergent structure in genus *Neolamarckia*. *BMC Genomics*. Springer; 2025.
14. Bakker FT, Lei D, Yu J, Mohammadin S, Wei Z, van de Kerke S, et al. Herbarium genomics: Plastome sequence assembly from a range of herbarium specimens using an Iterative Organelle Genome Assembly pipeline. *Biol J Linn Soc Blackwell Publishing Ltd*. 2016;117:33–43. <https://doi.org/10.1111/BIJ.12642>.
15. El-Aziz A, Farid S, Elkomey S. Evaluation of molecular and phenotypic diversity in relation to heterosis in some tomato lines under different climatic conditions. *Journal of Agricultural Chemistry and biotechnology*. Mansoura University, Faculty of Agriculture; 2016;7:141–51.
16. Rinawati DY, Reflinur R, Dinarti D, Sudarsono S. Genetic diversity of sugar palm (*Arenga pinnata*) derived from nine regions in Indonesia based on SSR markers. *Biodiversitas*. 2021;22.
17. Hatta ANN L, Sukma D, Maskromo I, Sudarsono S. Validated SNAP markers based on the CYP P450 87 A3 gene in coconut (*Cocos nucifera*) are associated with yearly stem height increase. *Biodiversitas*. 2023;24.
18. Jin J-J, Yu W-B, Yang J-B, Song Y, DePamphilis CW, Yi T-S, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol Springer*. 2020;21:241.
19. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq - Versatile and accurate annotation of organelle genomes. *Nucleic Acids Res*. 2017;45:W6–11. <https://doi.org/10.1093/nar/gkx391>.
20. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res*. Oxford University Press; 2004;32:W273–9.
21. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Volume 25. *Bioinformatics: Oxford University Press*; 2009. pp. 1451–2.
22. Daniell H, Lin C-S, Yu M, Chang W-J. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol*. 2016;17:134. <https://doi.org/10.1186/s13059-016-1004-2>.
23. Zhu Y, Zhang X, Yan S, Feng C, Wang D, Yang W, et al. SSR identification and phylogenetic analysis in four plant species based on complete chloroplast genome sequences. *Plasmid Acad Press*.

2023;125:102670. <https://doi.org/10.1016/J.PLASMID.2023.102670>.

24. Powell W, Morgante M, McDevitt R, Vendramin GG, Rafalski JA. Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. *Proceedings of the National Academy of Sciences*. 1995;92:7759–63.
25. Li B, Lin F, Huang P, Guo W, Zheng Y. Complete chloroplast genome sequence of *Decaisnea insignis*: Genome organization, genomic resources and comparative analysis. *Sci Rep. Nature Publishing Group UK London*; 2017;7:10073.
26. Fu PC, Zhang YZ, Geng HM, Chen SL. The complete chloroplast genome sequence of *Gentiana lawrencei* var. *farreri* (Gentianaceae) and comparative analysis with its congeneric species. *PeerJ*. PeerJ Inc.; 2016;2016. <https://doi.org/10.7717/PEERJ.2540>
27. Wang W, Chen S, Zhang X. Whole-genome comparison reveals divergent IR borders and mutation hotspots in chloroplast genomes of *herbaceous bamboos* (Bambusoideae: Olyreae). *Molecules*. MDPI; 2018;23:1537.

Figures

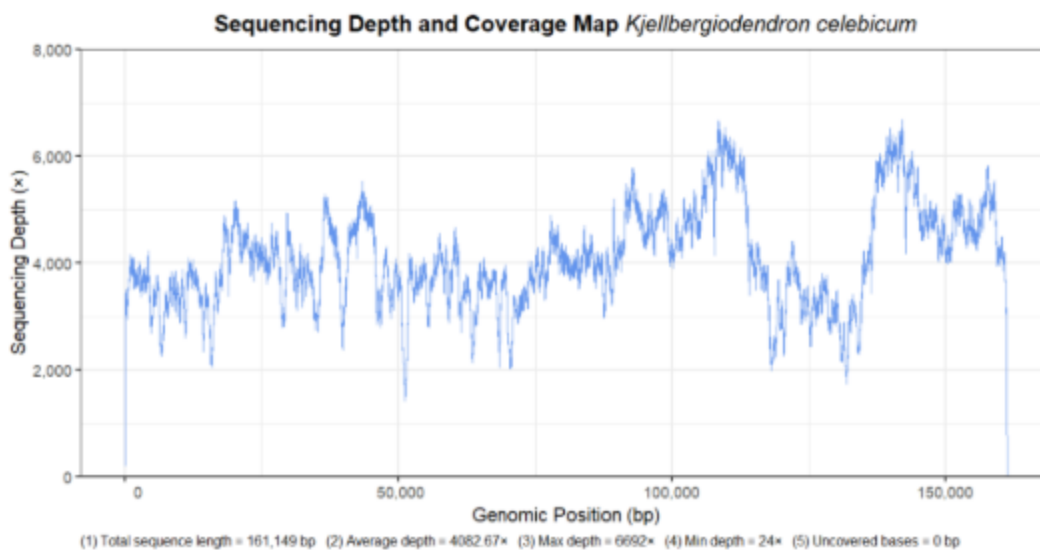


Figure 1

Sequencing depth and coverage map of the assembled chloroplast genome of *Kjellbergiodendron celebicum*.

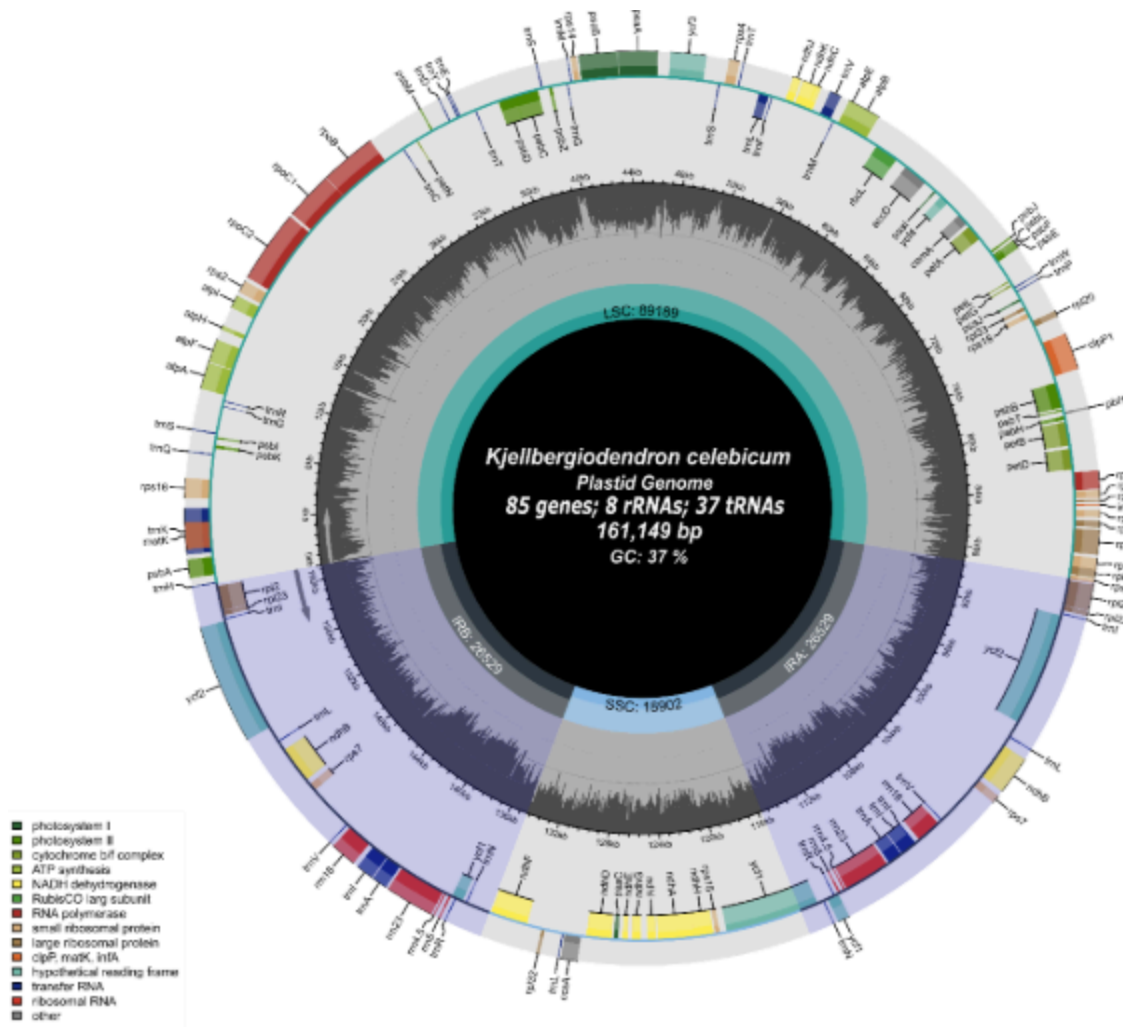


Figure 2

The concentric circles represent the quadripartite structure of the *Kjellbergiodendron celebicum* chloroplast genome, consisting of the large single-copy (LSC), small single-copy (SSC), and inverted repeat (IR) regions. Dark grey shading indicates GC content, and light grey shading indicates AT content. Genes on the inner side of the outer circle are transcribed clockwise, whereas genes on the outer side are transcribed anticlockwise. Genes marked with an asterisk (*) contain introns.



Figure 4

Relative synonymous codon usage (RSCU) of protein-coding genes in the chloroplast genome of *Kjellbergiodendron celebicum*.

Inverted Repeats

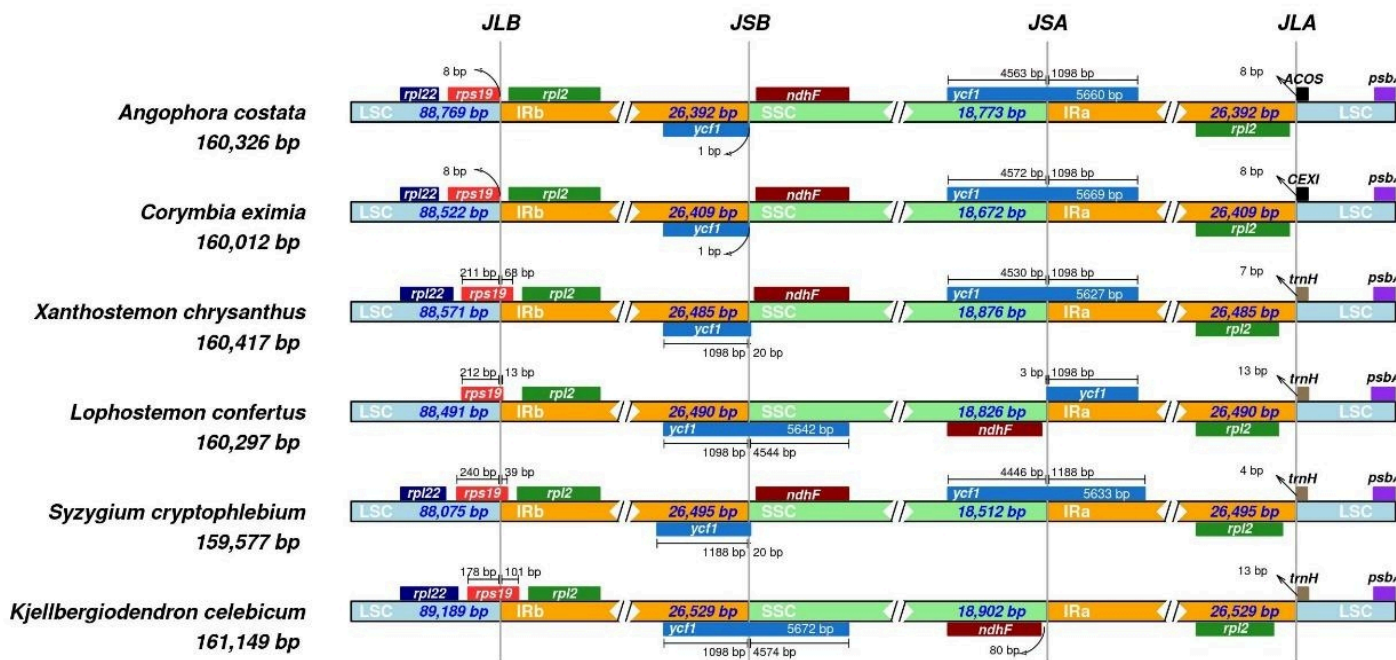


Figure 5

Comparison of the expansion and contraction of the IR/SC boundary regions across six Myrtaceae chloroplast genomes, including *Kjellbergiodendron celebicum*.

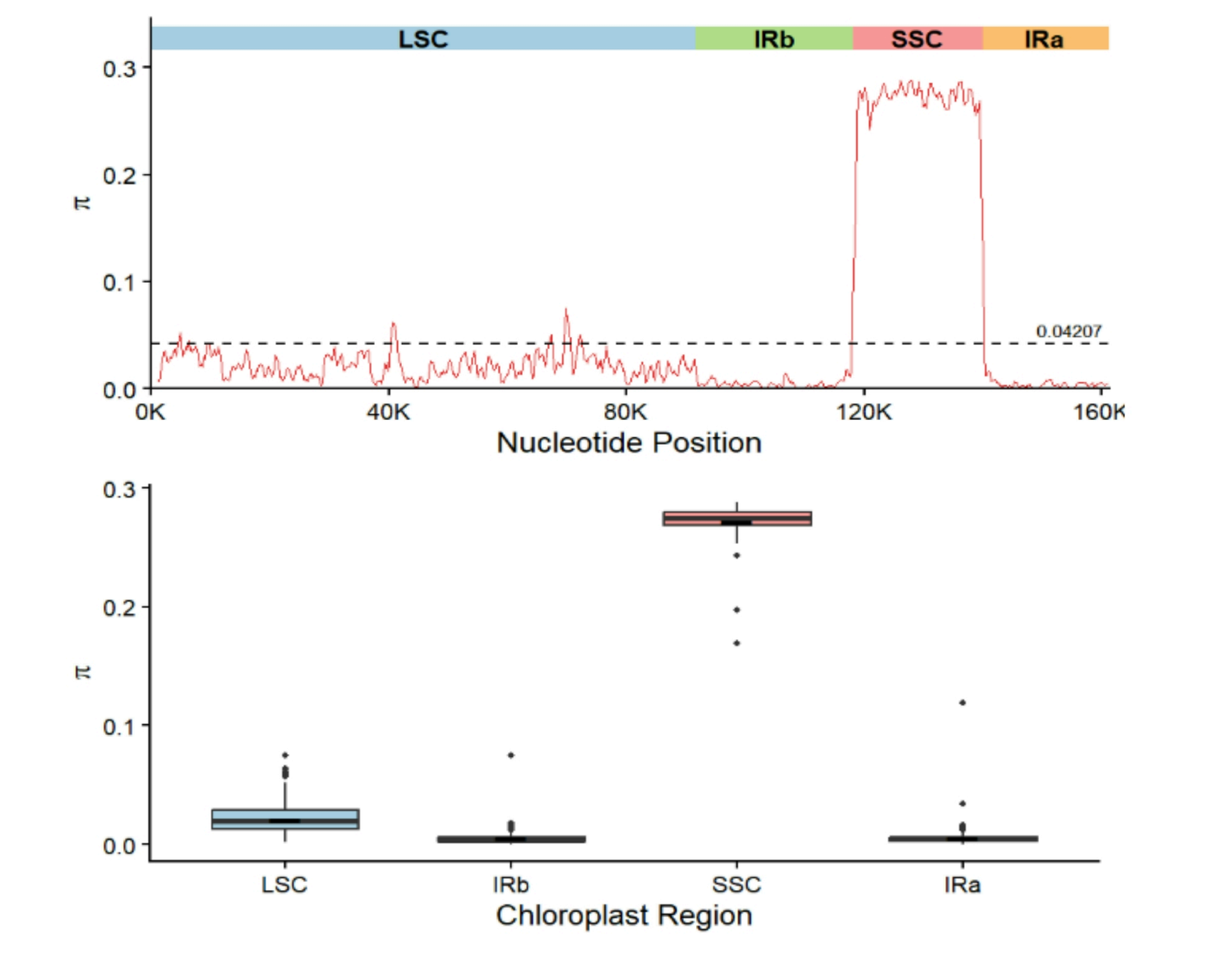


Figure 6

A. Nucleotide diversity (π) determined by sliding-window analysis (window length 600 bp, step size 200 bp) across the aligned Myrtaceae chloroplast genomes. The x-axis and y-axis indicate the position of the midpoint of the nucleotide diversity of each window. Sequences exhibiting relatively high nucleotide diversity represent potential hotspot regions. **B.** The distribution of Pi value differences is shown among the LSC, SSC, IRA, and IRB regions

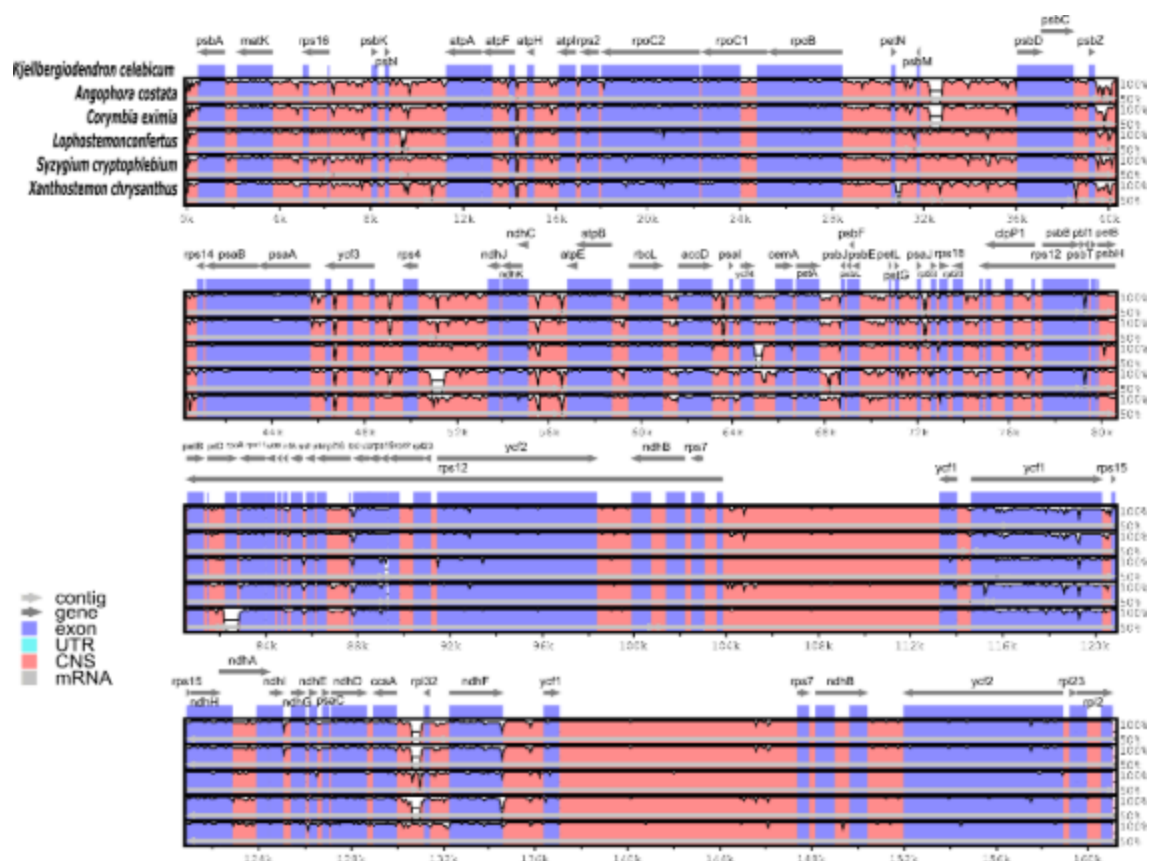


Figure 7

Comparison of *K. celebicum* chloroplast genomes using mVISTA with *K. celebicum* genome as a reference.

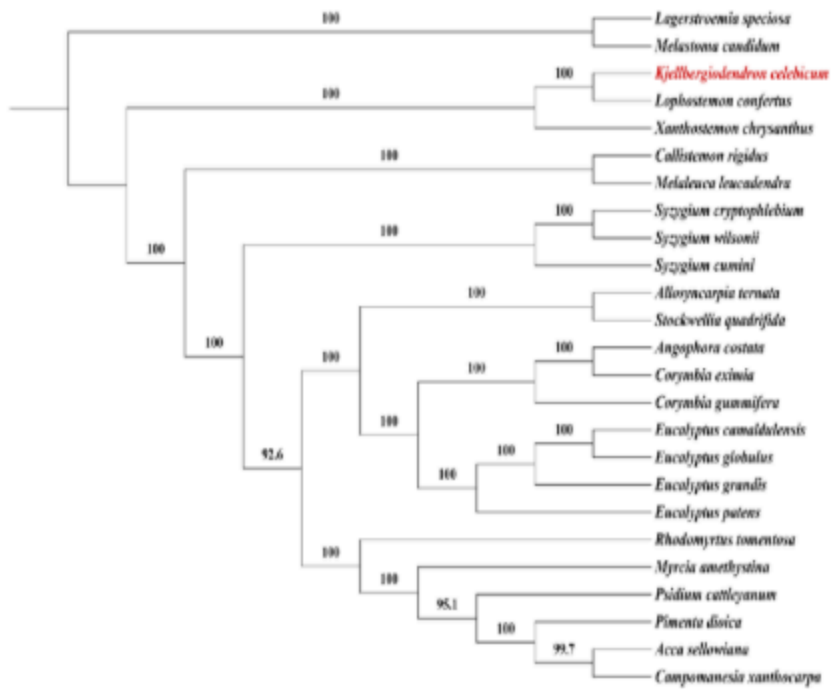


Figure 8

Maximum-likelihood phylogeny based on complete chloroplast genomes of Myrtaceae, showing the position of *Kjellbergiodendron celebicum* (in red). Numbers at nodes indicate bootstrap support.