Complete Chloroplast Genomes of Malaysian Banana Cultivars Musa acuminata cv Berangan and Musa balbisiana cv Nipah Reveals Lineage Specific Domestication Signatures

Maria Ulfa Md-Shah
Institute of Biological Sciences
Rofina Yasmin Othman
Centre for Research in Biotechnology for Agriculture
Sze Looi Song
Institute for Advance Studies
Jennifer Ann Harikrishna
Institute of Biological Sciences
Yusmin Mohd-Yusuf
yusmin_y@um.edu.my

Institute of Biological Sciences

Research Article

Keywords: Maternal lineage, conservation, comparative analysis, phylogenetic analysis

Posted Date: February 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8421706/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Plant Molecular Biology Reporter on May 25th, 2026. See the published version at https://doi.org/10.1007/s11105-026-01728-0.

Abstract

Banana (*Musa* spp.) represents a critical staple crop with complex domestication history involving interspecific hybridization between *Musa acuminata* (A-genome) and *Musa balbisiana* (B-genome) progenitors. Despite Malaysia's role as a major centre of origin and domestication, chloroplast genomic resources for Malaysian cultivated varieties remain critically underrepresented. In this study, we sequenced and assembled chloroplast genomes of Malaysian accessions, *Musa acuminata* cv. Berangan and *Musa balbisiana* cv. Nipah. Genome features including structure, GC content, repeat organization, and codon usage were characterized. The chloroplast genomes measured 170,035 bp (*M. acuminata* cv. Berangan) and 170,308 bp (*M. balbisiana* cv. Nipah), both exhibiting the typical quadripartite structure. Both genomes contain 113 genes comprising 79 protein coding genes, 30 transfer RNA (tRNA) genes and 4 ribosomal RNA (rRNA) genes. A total of 76 and 92 simple sequence repeats (SSR) reported. The number of long repeats varied with 331 repeats in *Musa acuminata* cv. Berangan and *M. balbisiana* cv. Nipah and exhibited higher repeat numbers (664 repeats). Comparative analysis revealed contraction and expansion of IR regions resulting in duplication of the *ycf1, rps15, ndhH* and *ndhA* gene. Highly variable regions were detected mainly found in noncoding intergenic regions, such as in *rps16-psbK, atpH-atpI, rps4-ndhJ, psbM-psbD, psbI-atpA.* Divergence hotspots were detected mainly in intergenic regions, as well as coding genes such as *accD*, *ycf1* and *ycf2.* Phylogenetic analysis indicated clear maternal lineage separation, *M. acuminata* cv. *Berangan* grouped with *M. acuminata* subspecies but closer to *M. acuminata* subsp. zebrina, whereas *M. balbisiana* cv. Nipah clustered with *M*. *balbisiana*, suggesting maternal plastid inheritance. These complete chloroplast genome sequences provide a valuable molecular foundation for the conservation, molecular breeding and understanding maternal contributions to banana domestication in Malaysia.

Introduction

Bananas and plantains, *Musa* L. are perennial crops cultivated year-round in the tropical and subtropical regions. In 2023, global banana production reached approximately 150 million tonnes, with India, China and Philippines being the top producers. Ecuador emerged as the leading exporter, accounting for 29% of the global market and generating USD$3.78 billion in export revenue. The United States, as the world's largest importer, represented 20% of global imports, valued at USD$3.14 billion. Although banana production has grown steadily over the past decade, the rate of the growth has slowed in recent years (FAO, 2023).

Cooking bananas and plantains play a significant role in global food security as well as national and household economies, as they account for about 80% of global banana production. Bananas (*Musa* spp.) provide essential nutrients to populations in both producing and importing countries. They are particularly important in some of the world's least developed and low-income countries, where they can contribute both to household food security and income generation as a cash crop. Bananas are a rich source of vitamins, minerals, and antioxidants, offering numerous health benefits (Ashokkumar et al., 2018). Research indicates that they possess antioxidant, anti-ulcerogenic, and anti-hypertensive

properties, while also contributing to gut health and potentially mitigating neurodegenerative diseases (Bashmil et al. 2021; Borges et al., 2018; Borges et al., 2019).

The development of chloroplast biotechnology has gained attention as a means of adding value to banana plants beyond their naturally occurring traits. Through molecular farming, the chloroplast genome can be engineered to produce specific characteristics such as enhanced nutritional content, improved stress tolerance, and the synthesis of high-value biopharmaceuticals (Daniell et al., 2021, 2022). Importantly, the diversity of chloroplast genomes across cultivars and species necessitates accurate identification and characterization of each plastome before engineering modifications, ensuring regulatory compatibility and minimizing unintended structural or functional effects (Hua et al., 2022).

As metabolic organelles responsible for photosynthesis and the biosynthesis of amino acids, nucleotides, fatty acids, phytohormones, vitamins and other metabolites, chloroplasts play a vital role in plant physiology and development (Daniell et al., 2016; Schwenkert et al., 2022). In most terrestrial plants, chloroplast genomes exhibit highly conserved circular DNA structure, organized into a quadripartite structure consisting of a two inverted repeat sequences (IR), as well as a large single copy region (LSC) and a small single copy region (SSC) (Oldenburg & Bendich, 2016; Morley, 2019a). Although angiosperms chloroplast genomes are generally conserved, mutational events occur, such as structural rearrangement, insertions, and deletions, inversions, translocations do occur. This variations provide valuable information for studies on population genetics, phylogeny, species identification, and conservation. Furthermore, analysis of chloroplast genomes can reveal codon usage bias, offering insights into translational preferences and evolutionary adaptation across species.

In this study, we present the complete chloroplast genome sequences of two Malaysian *Musa* cultivated varieties of *M. acuminata* cv. Berangan and *M. balbisiana* cv. Nipah obtained using Illumina sequencing technology. A comparative analysis of these chloroplast genomes with those closely related species within the Musaceae family is reported. The results provide valuable information on genome organization, gene content, and structure variation in *M. acuminata* cv. Berangan and *M. balbisiana* cv. Nipah, as well as insight into their phylogenetic relationships, contributing to future genetic and evolutionary studies in *Musa* and related taxa.

Materials and methods

Plant Materials, Chloroplast DNA Extraction and Genome Sequencing

Fresh, healthy leaf samples of *M. balbisiana* cv. Nipah were collected from Rimba Ilmu Botanical Garden, Universiti Malaya, Kuala Lumpur and *M. acuminata* cv. Berangan from the farm in Sekinchan, Selangor. Chloroplast DNA was extracted using sucrose step gradient centrifugation (Jansen et al. 2005) and a modified high salt method (Shi et al. 2012). DNA quality was assessed by electrophoresis on a 1%

agarose gel, and quantification was performed using a Nanodrop 2000 spectrophotometer (Thermo Scientific, USA).

Chloroplast Genome de novo Assembly and Annotation

High quality chloroplast DNA was used to construct paired-end sequencing libraries, which were sequenced using an Illumina HiSeq 2000 platform. Raw reads containing only adaptor fragments were removed to generate clean reads, and sequence quality was evaluate using the FastQC program. *Musa acuminata* ssp. malaccensis (GenBank accession: HF677508) chloroplast genome was used as reference for assembly with GetOrganelle v1.7.6.1 pipeline (Jin et al. 2020). The assembly process incorporated SPAdes v3.11.1 (Prjibelski et al. 2020), bowtie2 v2.4.2 (Langmead and Salzberg, 2012), and BLAST + v2.11 (Camacho et al. 2009) with default parameters. Gene annotations of the protein-coding genes, transfer RNAs (tRNAs), and rRNA genes was performed using GeSeq web server (Tillich et al. 2017) with default settings. The chloroplast genome maps were visualized using OGDRAW 1.3.1 (Greiner et al. 2019).

Chloroplast Genome Structural Analysis

Microsatellite simple sequence repeats (SSRs) were identified using the MISA Perl script (http://pgrc.ipk-gatersleben.de/misa/) with the following parameter settings: ten repeat units for mono-nucleotides, five repeat units for di-nucleotides, four repeat units for tri-nucleotides, and three repeat units for tetra-, penta-, and hexanucleotide repeats. Tandem repeats were detected using the Tandem Repeat Finder web server with parameters set as follows: match, mismatch, and indel parameters of 2, 7, 7 respectively; minimum alignment score of 80, and maximum period size of 500. Redundant or nested repeats were excluded, and repeats located within inverted repeats (IR) regions were counted only once. Relative synonymous codon usage (RSCU) was analyzed using CodonW (Galaxy version 1.4.4). IR-SSC expansion and contraction was identified using Geneious Prime version 2024.0.7 software (http://www.geneious.com, Kearse et al., 2012).

Genome Comparison

Basic chloroplast genome features, of chloroplast genomes, including length, GC content, and junction position of LSC, SSC, and IR regions, were compared using Geneious software. The mVISTA program (http://genome.lbl.gov/vista/mvista/submit.shtml) was used to compare the complete chloroplast genomes of *M. acuminata* cv. Berangan and *M. balbisiana* cv. Nipah with *M. acuminata ssp. malaccensis* (HF 677508) and *M. balbisiana* (NC 028439), using default parameters. Alignment was performed with the LAGAN program (Brudno et al., 2003), and sequence conservation profiles were visualized using mVISTA plot (Frazer et al., 2004).

Phylogenetic Analyses

Phylogenetic analysis was conducted using two newly sequenced chloroplast genomes (*M. acuminata* cv Berangan and *M. balbisiana* cv Nipah) along with 63 other *Musa* accessions using a concatenated alignment of 70 protein-coding genes from each chloroplast genome. Multiple sequence alignment was conducted by the MAFFT software (Katoh et al. 2002). All *Musa* accession and outgroup chloroplast genomes (*Alpinia pumila* and *Zingiber officinale*) were retrieved from Genbank.

(https://www.ncbi.nlm.nih.gov/genome/browse#!/organelles/musaceae accessed on 12 December 2023). Maximum likelihood (ML) phylogenetic inference was performed using IQ-TREE 2.2.2.6. (Minh et al., 2020) with the best-fit substitution model (TVM + F + I + R6) selected based on the Bayesian Information Criterion (BIC) using ModelFinder (Kalyaanamoorthy et al., 2017). Prior to phylogenetic analysis, the 70 protein-coding genes were concatenated using Geneious software. Branch support values were estimated using the 10,000 bootstrap method implemented in IQ-TREE 2 (Hoang et al., 2018). The resulting phylogenetic tree was visualized using FigTree version 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree).

Results

Chloroplast genome features of Musa species

Next generation sequencing generated 1 Gb of clean reads for *M. acuminata* cv Berangan and 2.13 Gb of clean reads for *M. balbisiana* cv Nipah using Illumina sequencing platform. The complete chloroplast genome sizes were 170,035 bp for *M. acuminata* cv. Berangan and 170,308 bp for *M. balbisiana* cv. Nipah (Fig. 1) Both genomes exhibited the typical quadripartite structure, consisting of a large single copy (LSC) region, a small single copy (SSC) region and two inverted repeat (IR) regions. As shown in Table 1, the chloroplast genome of *M. acuminata* cv Berangan composed an LSC region of 88,711 bp, an SSC region of 10,828 bp, and an IR region of 35,248 bp. The overall GC content is 36.8%, with regional GC contents of 35.2% (LSC), 31.1% (SSC), and 39.8% (IR). Similarly, the chloroplast genome of *M. balbisiana* cv Nipah contained an LSC region of 87,906 bp, an SSC region of 11,451 bp, and an IR region of 35,476 bp (Table 1). The overall GC content was 36.7%, with regional GC contents of 35.1% (LSC), 30.9% (SSC), and 39.5% (IR). The list of genes are shown in Table 2.

Table 1
The structure and length of the chloroplast genome sequences in *M. acuminata* cv. Berangan and *M. balbisiana* cv. Nipah.

Genome characteristics		***M. acuminata* cv. Berangan**	***M. balbisiana* cv. Nipah**
Genome size		170,035 bp	170,308 kb
LSC length		88,711 bp	87,906 bp
SSC length		10,828 bp	11, 451 bp
IR length		35,248 bp	35, 476
GC content	Total	36.8%	36.7%
	LSC	35.2%	35.2%
	SSC region	31.1%	30.9%
	IR region	39.8%	39.4%
Total no. of genes		113	113
Protein-coding genes		79	79
tRNA		30	30
rRNA		4	4

Gene annotation shows 113 genes in *M. acuminata* cv. Berangan and *M. balbisiana* cv. Nipah cp genomes respectively, including 79 protein-coding genes (69 single copy protein-coding genes, 12 repeated/IR genes, and 2 copies of *ndhA, ndhH, ycf1, rps12, rps7, ndhB, ycf2, rpl23, rpl2, rps19, psbL*), 28 tRNA genes (10 repeated/IR genes and 2 copies of *trnH-GUG, trnM-CAU, trnI-CAU, trnL-CAA, trnV-GAC, trnI-GAU, trnE-UUC, trnA-UGC*, *trnR-ACG, trnN-GUU*), and 4 rRNA genes (4 repeated/IR genes and 2 copies of *rrn23*, *rrn4.5*, *rrn5*, and *rrn16*) (Table 2). Among the annotated genes, 16 genes from both species had introns, of which, there were two introns for *paf1* and *clpP1*, and one intron for the other 14 genes (*rps16, atpF*, *rpoC1*, *ndhA*, *petB, petD, rpl16, trnI-GAU, trnE-UUC, tRNA-UGC, trnL-UAA*, *trnC-UAA*). Seven genes within the inverted repeat regions had an intron. Interestingly, the *ndhA* gene at IRa had one intron but with trans-splicing, while one of its exons located in the SSC region and the other duplicated by the IRs.

Table 2
List of genes in *M. acuminata* cv. Berangan and *M. balbisiana* cv. Nipah chloroplast genomes

Category	Group	Name of Gene
Photosynthesis	Photosystem I	*psaA, B, C, I, J*
	Photosystem II	*psbA, B, C, D, E, F, H, I, J, K, L, M, N, T, Z*
	Cytochrome b6/f	*petA, B*[b], *D*[b], *G, L, N*
	ATP synthase	*atpA, B, E, F*[b], *H, I*
	Rubisco	*rbcL*
	NADH oxidoreductase	*ndhA*[b,c], *B*[b,c], *C, D, E, F, G, H*[c], *I, J, K*
	Large subunit ribosomal proteins	*rpl2*[b,c], *14, 16*[b], *20, 22, 23*[c], *32, 33, 36*
Replication	Small subunit ribosomal proteins	*rps2, 3, 4, 7*[c], *8, 11, 12*[b,c], *14, 15*[c], *16*[b], *18, 19*[c]
	RNAP	*rpoA, rpoB, C1*[b], *C2, pafI*[a], *pafII*
	Ribosomal RNAs	*rrn23*[c], *16*[c], *5*[c], *4.5*[c]
	Transfer RNAs	*trnA-UGC*[b,c], *trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnfM-CAU, trnG-GCC, trnG-UCC, trnH-GUG*[c], *trnI-CAU*[c], *trnI-GAU*[b,c], *trnK-UUU*[b], *trnL-CAA*[c], *trnL-UAA*[b], *trnL-UAG, trnM-CAU, trnN-GUU*[c], *trnP-UGG, trnQ-UUG, trnR-ACG*[c], *trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC*[c], *trnV-UAC*[b], *trnW-CCA, trnY-GUA*
Other genes	Other proteins	*accD, ccsA, cemA, clpP*[a], *matK, infA*
Unknown function	Proteins of unknown function	*ycf1*[c], *ycf2*[c]

[a] Gene containing two introns.

[b] Gene containing a single intron.

[c] Two gene copies in the IRs.

Codon Usage Bias in M. acuminata cv. Berangan and M. balbisiana cv. Nipah

Relative synonymous codon usage (RSCU) analysis was conducted using CodonW software to examine codon usage bias in the chloroplast genomes of *M. acuminata* cv. Berangan and *M. balbisiana* cv. Nipah.

The analysis, encompassing 79 protein-coding genes from each cultivar, identified a total of 23,134 codons in *M. acuminata* cv. Berangan and 23,124 codons in *M. balbisiana* cv. Nipah. The results, illustrated in Fig. 2 and Fig. 3, depict the distribution and patterns of codon usage across the protein-coding regions of both chloroplast genomes.

In *M. acuminata* cv. Berangan, the most abundant amino acids are leucine represented by 3,151 codons (13.3%) and cysteine is the least abundant amino acid represented by 581 codons (2.5%) as shown in Fig. 2. Methionine and tryptophan have no biased usage (RSCU = 1). There are 27 preferred codon with RSCU > 1 and 26 of them are A/U-ending. The highest value of RSCU is AGA encoded Arg (2.24), followed by GGA encoded glycine (1.53). Among the codons, there are 34 least preferred codon (RSCU < 1), CGC encoded arginine has the lowest preference index (0.35)

In *M. balbisiana* cv. Nipah, leucine is the most abundant amino acids, represented by 2,335 codons (10.1% of all codons), while cysteine is the least abundant amino acids, represented by 262 codons (1.1% of all codons). From the data shown in Fig. 3, it was revealed that *M. balbisiana* cv. Nipah chloroplast genome has 30 codons showed high preference (RSCU > 1) and 27 codons carried A and U bases in their third position. Among these codons, the UUA encoded leucine is the most preferred codon with the highest RSCU value (1.94), followed by AGA encoded for arginine (1.58). While 32 codons showed weak bias (RSCU < 1) more often carried C and G in their third position. Besides, two codons which are AUG and UGG had no bias (RSCU = 1) or codon usage preferences. Consistent with the AT-rich third position, the lowest RSCU values were G/C ending codons, led by AGC encoded serine (0.30)

Analysis of Repeat Sequences

The distribution and characteristics of short sequence repeats (SSRs) and long repeats within the chloroplast genomes of *M. acuminata* cv. Berangan and *M. balbisiana* cv. Nipah was investigated. Misa analysis shows the *M. acuminata* cv. Berangan has 76 SSRs, with mononucleotide repeats constituting the most abundant type (44.7%). Hexanucleotide repeats were the least frequent (2.6%). Other SSR types detected included 12 dinucleotide (15.8%) 10 trinucleotide (13.3%,) 13 tetranucleotide (17.1%) and 5 pentanucleotide (6.6%) repeats. While the chloroplast genome of *M. balbisiana* cv. Nipah chloroplast genome to harbour 92 SSRs, with mononucleotide repeats being the most abundant (42.4%) (Fig. 4a). There are 15 trinucleotide repeats (16.3%), 14 tetranucleotide repeats (15.2%), 11 trinucleotide repeats (12%), 10 dinucleotide repeats (10.9%), and 3 pentanucleotide repeats (3.3%). In both species, A/T-rich repeats were the most abundant among all repetitive sequences (Fig. 4b). The following repeat types of AG/CT, AT/AT, AAG/CTT, AAT/ATT, AAAC/GTTT, AAAT/ATTT,AATC/ATTG, AAAAT/ATTTT, AATAT/ATATT are common to both *Musa* species. Certain repeat types such as AAGAT/ATCTT, AAGCAG/CTGCTT are specific to *M. acuminata* cv. Berangan. While C/G, AAATT/AATTT, AATTAT/AATTAT occur once in *M. balbisiana* cv. Nipah. Notably, the repeat type AAAATC/ATTTTG is found only in *M. balbisiana* cv. Nipah with total 12 occurrence. These results highlight species-specific differences in SSR profiles. Most of the SSRs were found in the LSC region.

For long repeats analysis (> 30bp), the *M. acuminata* cv. Berangan chloroplast genome contains 331 repeats, which include 198 forward, 125 palindromic and 8 reverse repeats, with no complementary repeats identified. In contrast, *M. balbisiana* cv. Nipah chloroplast genome displays a significant notable difference of 664 long repeats, consisting predominantly of forward repeats (338) and palindromic repeats (315), whereas only 7 reverse and 8 complementary repeats are present (Fig. 4c). These findings highlight a clear distinction between two species with *M. balbisiana* cv. Nipah showing a greater abundance and diversity of long repeats. Among the four repeat types analyzed, forward and palindromic repeats dominate in both *Musa* chloroplast genomes. This suggests their prominent role in the genomic structure of both species.

Expansion of inverted repeats (IR)

A comparative analysis was conducted to explore the expansion and contraction of IR regions of *M. acuminata* cv. Berangan *and M. balbisiana* cv. Nipah with wild *Musa* accessions, *Musa acuminata* subspecies malaccensis DH Pahang (HF 677508) and *Musa balbisiana* (NC 028439) as depicted in Fig. 5. *M. acuminata* cv. Berangan exhibits an IR region length of 35,248 bp. This size is slightly shorter than the wild progenitor, *M. acuminata* ssp. malaccensis (35,433 bp). In contrast, the IR region in *M. balbisiana* cv. Nipah reveals significant expansion (35,476 bp) as compared to wild progenitor, *M. balbisiana* (35,064 bp).

The IRb/SSC junction (JSB) showing remarkable structural diversity across both cultivated and wild accessions. Both wild and cultivated *Musa* accessions exhibit a substantial expansion of the IRb into the SSC region. This structural characteristic not commonly observed in other monocot species. This expansion leads to the full duplication of several genes that are typically located in the SSC of most monocots, including *ycf1, rps15, ndhH* and *ndhA*. Remarkably in *M. balbisiana* cv. Nipah and both wild accessions, the *ndhF* gene spans across the junction of the IRb and SSC regions. Specifically, a portion of *ndhF* is located in the IRb while the remaining segment resides within the SSC region. This unique arrangement contrasts with *M. acuminata* cv. Berangan, where the *ndhF* gene is entirely confined to the SSC, positioned 53 bp away from the JSB. At the SSC/IRa junction (JSA), the *ndhA* gene consistently spans the boundary between the SSC and IRa regions in all *Musa* accessions. This results in partial duplication into the IRa, with the remaining segment in the SSC. While this structural arrangement is conserved across the genus, the length of *ndhA* within IRa varies. Nevertheless, the consistent positioning of the JSA within *ndhA* highlights its conservation across *Musa* species.

Comparative Analysis of Chloroplast Genomes of M. acuminata cv. Berangan and M. balbisiana cv. Nipah

To elucidate the genome divergence and highly variable regions, the chloroplast genomes of *M. acuminata* cv. Berangan and *M. balbisiana* cv. Nipah was compared with their respective wild accessions, *M. balbisiana* and *M. acuminata* subsp. malaccensis DH Pahang. It reveals a generally highly conserved genome structure, with minor variations observed between the accessions. The coding regions are more

conserved compared to the non-coding regions, while the SSC and LSC regions shows more divergence than the IRa and IRb regions (Fig. 6). Notably, significant variation was observed in the intergenic spacers regions between several gene pairs. Particularly in non-coding intergenic regions, such as in *rps16-psbK, atpH-atpI, rps4-ndhJ, psbM-psbD, psbI-atpA.* Additionally, the intergenic region around *ycf1-rps12* intergenic region displayed a notable divergence in *M. balbisiana* and *M. balbisiana* cv. Nipah. In the SSC region, slight variations were observed in the non-coding regions of *rpl32-ccsA* and *psaC-ndhE*. Among the coding region, the *accD, ycf1* and *ycf2* exhibited substantial divergence, highlighting areas of genomic variability that may contribute to functional differences between cultivated and wild species (Fig. 6).

Phylogenetic Relationships of M. acuminata cv. Berangan and M. balbisiana cv Nipah Among Musa species.

For a comprehensive understanding of *Musa* evolutionary relationships, a Maximum Likelihood (ML) phylogenetic was reconstructed using 70 protein-coding chloroplast genes from 63 plant accessions. The resulting phylogeny (Fig. 7) clearly resolved the genus *Musa* into two well-supported clades, corresponding to the sections *Musa* and *Callimusa*. Members of section *Musa*, including the *acuminata* and *balbisiana* groups clustered together within the same clade. In contrast, accession belonging to section *Callimusa*, such as *Musa beccarii*, *Musa gracilis* and *Musa borneensis* formed a distinct and well-supported lineage. The genus *Musella*, represented by a single species, was positioned as a sister to *Ensete*, consistent with previous taxonomic classifications. *Zingiber officinale* and *Alpinia pumila* were designated as the outgroup.

The ML phylogenetic analysis reveals that the Malaysian cultivated accessions, *M. acuminata* cv. Berangan clustered closely with, *M. acuminata* subspecies, *M. acuminata* ssp. zebrina and *M. acuminata* ssp. malaccensis DH Pahang, separated by very short branch lengths. Other subspecies such as *M. acuminata* ssp. microcarpa, *M. acuminata* ssp. banksii and *M. acuminata* ssp. halabanensis, were also grouped within the *acuminata* clade. This reflects a high degree of genetic similarity among these lineages. Previously sequenced cultivated accessions, including *M. acuminata cv.* Pisang Lilin, *M. acuminata* cv. Kluai Khai, *M. acuminata* cv. Williams clustered within the same well-supported acuminata clade. In addition, cultivated accession, *M. balbisiana* cv. Nipah and the wild progenitor, *M. balbisiana*, forms monophyletic clade, indicate closely related and share a common evolutionary ancestor.

Discussions

Comparison of the chloroplast genomes in the Malaysian cultivated Musa species

The chloroplast genomes of *M. acuminata* cv. Berangan and *M. balbisiana* cv. Nipah exhibit a typical quadripartite structure, comprising two inverted repeat (IR) regions separating the large single copy (LSC) and small single copy (SSC) regions. With total genome sizes 170,035 and 170,308 bp, the chloroplast genomes are larger than the average size of 154.2 kb reported for most flowering plants (Kwon et al., 2020). These sizes are consistent with those previously described for other *Musa* species

(Song et al. 2022; Fu et al., 2022). The LSC regions of both genomes are comparable in size to those observed across angiosperms, bryophytes and pteridophytes, indicating overall conservation in this region. In contrast, the SSC regions are notably smaller than the typical angiosperms range (14–47 kb), representing significant contraction relative to most higher plants (Kwon et al. 2020). This likely due to the expansion of the IR regions into SSC.

The *Musa* chloroplast genomes exhibit low GC content, ranging from 36.7% to 36.8%, which is typical of organelle genomes. GC content in plant chloroplast genomes can vary widely, ranging from 25% to 51% across different plant taxa (Kwon et al. 2020). Previous studies on *Musa* species have reported a narrow range of GC content, from 36.45% to 36.82% (Song et al., 2022; Liu, 2018). These minor variations offer valuable insight into the evolutionary divergence and phylogenetic relationships within the genus, further enhancing our understanding of genetic differences among *Musa* species (Liu et al., 2021; Fu et al., 2022). While many plant species tend to have lower GC content, exceptions exist, such as certain *Selaginella* species in lycophytes, which display higher GC content (over 50%). This higher GC content has been linked to increased RNA editing rates (Mower et al. 2019). In addition, the gene order in the chloroplast genomes of *M. balbisiana* cv. Nipah, *M. acuminata* cv. Berangan, and *M. acuminata* ssp. malaccensis DH Pahang is highly conserved. While gene order is generally conserved in land plants, some seed plants exhibit significant rearrangements due to gene losses and intron losses (Frailey et al., 2018). In terms of gene content, both cultivated accessions consistently retained the same core group of 79 protein-coding genes, 30 tRNA genes, and 4 rRNA genes, highlighting the strong evolutionary conservation of chloroplast gene content. Most land plant chloroplast genomes possess a conserved set of genes, with minor variations observed in *Musa* species.

The chloroplast genome of *M. acuminata* cv. Berangan and *M. balbisiana* cv. Nipah display highly conserved codon usage patterns and compositional features, consistent with earlier findings in *Musa* and other non-grass monocots (Martin et al. 2013;Song et al. 2022).The total codon usage varied minimally, highlighting the evolutionary stability of chloroplast coding capacity. The most frequent codons, AUU (Ile), GAA (Glu) and AAA (Lys) are consistent with other studies, where A- and U- ending codons are favoured. This pattern reflects translation efficiency, optimised by abundant corresponding tRNAs in chloroplasts, as observed in Miscanthus and other non-grass monocots (Sheng et al. 2021; Mazumdar et al. 2017). The weak codon usage bias observed in both accessions indicate that their chloroplast genomes are not under significant translational selection pressure, with both species showing a largely random codon usage pattern. Future studies may focus on exploring the functional significance of these codon preferences and their impact on protein synthesis efficiency in these species.

SSRs are important molecular markers that play a significant role in plant population genetics and evolution investigations. The analysis revealed variations in the number of cpSSRs with *M. balbisiana* cv. Nipah (92 repeats) exhibiting a higher number compared to *M. acuminata* cv. Berangan (76 repeats). Compared to its wild progenitor, *M. acuminata* ssp. malaccensis has similar SSR number to *M. acuminata* cv. Berangan, though the SSR types differ slightly (Martin et al. 2013). In contrast, *M. balbisiana* (wild

progenitor), exhibits a slightly lower SSR count (80 repeats) compared to the cultivated accession, *M. balbisiana* cv. Nipah (92 repeats). This difference is attributed to a unique hexanucleotide present in *M. balbisiana* cv. Nipah, which is absent in wild accessions (Shetty et al. 2016). Mononucleotide repeats were the most abundant in both accessions and this is also observed in other plant groups (Li et al., 2020). The predominance of certain repeat types in certain accessions within the chloroplast genome suggest potential evolutionary and functional significance. The identified SSRs provide valuable resources for future research, including the development of molecular markers for genetic diversity studies, population genetics, and marker-assisted breeding in *Musa* species.

A significant difference in repeat sequences was observed between the chloroplast genomes of *M. balbisiana* cv. Nipah and its wild progenitor, *M. balbisiana*. *M. balbisiana* cv. Nipah contains over 600 long repeats (> 30 bp) compared to only 212 repeats in the wild accession. In contrast, *Musa acuminata* cv. Berangan harbours a more moderate number of repeats (331) with count closely aligned with its wild progenitor, *Musa acuminata* ssp. malaccensis (321 repeats). These findings suggest that domestication has not uniformly impacted the chloroplast genome repeat content across *Musa* lineages. Specifically, repeat abundance appears to have been more dynamic in *balbisiana*. Whereas *M. acuminata* chloroplast genomes have remained relatively stable during domestication. Previous study reported diverse long repeats in different *Musa* species from 86 to 324 (Song et al. 2022). This uneven distribution is consistent with broader trends in angiosperms where repeats contribute to lineage-specific structural diversity (Cauz-Santos, 2025). The high repeat number in *Musa* may play a functional role in promoting recombination, enabling structural rearrangements, and facilitating adaptive evolution. Furthermore, variability of these repeats suggests they may serve as markers for speciation events, contributing to the phylogenetic diversity. This repeat-rich genome architecture likely plays a key role in the rapid diversification of *Musa* species.

Comparative Genome Analysis of Musa

The IR region size in the cultivated accessions in this study is notably large (35 kb). These sizes are consistent with previous reports of unusually large IR regions in *Musa* species (Martin et al, 2013, Song et al. 2022; Fu et al. 2022). While these values exceed the typical IR size range observed in most monocots (20–30 kb)(Cauz-Santos, 2025), the findings highlight that IR expansion is a significant contributor to chloroplast genome size variation among *Musa* accessions. However, the small increase in IR size in cultivated accessions relative to wild progenitors suggests that structural changes in the IR region are likely genetically stable across lineages, rather than driven by domestication. This indicate these variations have been conserved throughout evolutionary history.

The expansion of the IRb region in *M. balbisiana* cv. Nipah and *M. acuminata* cv. Berangan includes the full duplication of the *ycf1* gene along with *rps15, ndhH* and *ndhA*, leading to multiple copies of gene important for plastid function. Such gene duplications may enhance genome stability by increasing gene dosage and supporting homologous recombination repair pathways, important for maintaining chloroplast integrity (Martin et al., 2013; Fu et al., 2022). The extensive IR expansion observed in *Musa*

represents one of the largest among monocots, a contrast to the more variable or limited IR shifts seen in other plant families like *Araceae* and *Orchidaceae* (Yu et al., 2025; Low & Landrein, 2025). Within the Zingiberales, *Zingiber officinale* retains shorter IR regions with more compact boundaries, highlighting that large IR expansion is a trait unique to *Musa* (Song et al., 2022; Liu et al., 2018). This distinctive chloroplast architecture not only contributes to chloroplast stability during clonal propagation but also offers molecular markers that can aid in lineage differentiation and determining maternal ancestry in hybrid *Musa* species.

The comparative alignment of *M. acuminata* cv. Berangan and *M. balbisiana* cv. Nipah and its wild progenitors revealed overall structural conservation with divergence concentrated in intergenic spacers and a few coding loci. This pattern aligns with typical higher plant genomes, where IR regions are highly conserved, while LSC and SSC regions show greater substitution and indel rates. Divergence in the coding regions was especially prominent in the *accD, ycf1*, and *ycf2* genes. The *accD* gene, which encodes a subunit of acetyl-CoA carboxylase involved in leaf development and fatty acid biosynthesis, is under selection pressure, making it more prone to mutations and variations, reflecting evolutionary adaptation in *Musa*. Conversely, *ycf1* is one of the most variable plastid genes across plants and has been proposed as a universal DNA barcode (Dong et al., 2015). Its high divergence in *Musa* supports its potential as a marker for resolving close phylogenetic relationships and cultivar identification. Both *ycf1* and *ycf2* are large and essential genes for chloroplast function. It shows higher nucleotide diversity due to relaxed functional constraints or adaptive evolution, making them hotspots for polymorphism in *Musa*. Furthermore, the IR/SSC junction in *Musa* chloroplast genome incorporates the entire *ycf1* and part of ndhH/*ndhA* into the inverted repeats. This likely contributes to structural divergence and the elevated polymorphism observed in the *ycf1* region (Martin et al. 201; Wu et al., 2021).

Additionally, hypervariable sites were detected in several intergenic spacers (*rps16–psbK*, *atpH–atpI*, *rps4–ndhJ*, *psbM–psbD*, *psbI–atpA*), supporting previous findings in *Musa* chloroplast genomes (Song et al., 2022). Notably, lineage-specific divergence in the *ycf1–rps12* and *rps16–psbK* regions was observed in *M. balbisiana* and *M. balbisiana* cv. Nipah, but was absent in *M. acuminata* cv. Berangan, highlighting the evolutionary distinctness of the balbisiana plastome. This pattern aligns with phylogenomic analyses that demonstrate a strong separation between the A- and B-genome lineages in *Musa* (Fu et al., 2022). Similar chloroplast genomes divergence patterns have been used in other crops, like rice and wheat, to trace maternal inheritance and hybridization histories (Nock et al., 2011; Gornicki et al., 2014). These divergent hotspots offer practical applications, as hypervariable plastome loci can be developed into cost-effective markers for cultivar identification, which is particularly important for the Malaysian banana industry in breeding and germplasm management.

Phylogenetic anlysis of Musa

The phylogenetic reconstruction of the *Musa* chloroplast genomes provided understanding of the evolutionary relationships between cultivated *M. acuminata* cv. Berangan, *M. balbisiana* cv. Nipah, and their wild progenitors. The analysis clearly separated the *acuminata* and *balbisiana* lineages, with the

cultivated accessions clustering closely with their wild relatives. This finding suggests the maternal lineage is maintained between wild and cultivated accessions, demonstrating the role of chloroplast genome in preserving genetic identity across generations. Within the *acuminata* lineage, *M. acuminata* cv. Berangan was clustered with *acuminata* subspecies, closer to *M. acuminata* subsp. zebrina suggest its maternal lineage. However, it still grouped together with *M. acuminata* ssp. malaccensis, exhibiting minimal divergence. This patterns suggest that the chloroplast genome in *M. acuminata* cv. Berangan highly conserved. In contrast, the *balbisiana* lineage exhibited longer branches as compared to *M. acuminata* reflect deeper divergence in B-genome lineage, while A-genome lineages form a very short branches, consistent with recent radiation. This aligns with previous studies showing that cultivated *Musa* species, exhibit low chloroplast variability (Song et al., 2022). It suggests nuclear genome processes and clonal propagation have contributed more to genetic differentiation than chloroplast genome evolution. The chloroplast genome phylogeny placed *M. acuminata* cv. Berangan within a clade of other *M. acuminata* subspecies, such as *M. acuminata* ssp. microcarpa, *M. acuminata* ssp. banksii and *M. acuminata* ssp. halabanensis, supporting *M. acuminata* as the primary wild ancestor of many cultivated banana varieties. The close relationship between *M. acuminata* and other species such as *M. rubra, M. laterita*, and *M. siamensis* supports previous studies (Li et al., 2010; Fu et al., 2022) and suggests that introgression from *M. acuminata* may have occurred in these species. The high chloroplast genome similarity observed in these species indicates possible hybridization events, which should be explored further.

This phylogenetic analysis clarifies the relationship between cultivated *Musa* varieties in Malaysia and their wild progenitors, providing key insights into maternal inheritance in *Musa* chloroplast genomes. The clustering of *M. acuminata* cv. Berangan with *M. acuminata* ssp. zebrina and *M. balbisiana* cv. Nipah with *M. balbisiana* suggests that chloroplast genomes can track maternal lineage, which complements archaeological and phytolith evidence for the long history of *Musa* domestication in Southeast Asia (Bowdery et al., 1999; Denham et al., 2003).

These findings highlight the utility of chloroplast genomes in tracing maternal inheritance, and their potential for cultivar identification and germplasm management. The conserved structure of the chloroplast genome supports its role as a reliable marker for phylogenetic studies, while the lineage-specific variations identified in this study provide valuable markers for distinguishing genetic lineages. This research not only enhances our understanding of *Musa* domestication in Malaysia but also offers genomic resources to support banana breeding, conservation, and future crop improvement efforts.

Conclusions

This study provides a comprehensive analysis of the chloroplast genomes of *M. acuminata* cv. Berangan and *M. balbisiana* cv. Nipah, revealing insights into their genomic structure, gene content, and evolutionary relationships. The complete chloroplast genomes of *M. acuminata* cv. Berangan and *M. balbisiana* cv. Nipah are 170,035 bp and 170,308 bp respectively. The analysis revealed variations in SSR number and distribution between *M. balbisiana* cv. Nipah and *M. acuminata* cv. Berangan, with

mononucleotide repeats being the most abundant in both species, highlighting their importance as potential molecular markers for genetic diversity studies. Comparative analysis revealed a high degree of conservation among the studied *Musa* species, with the inverted repeat regions exhibiting higher sequence conservation than the single-copy regions. Furthermore, the identification of variable regions, including specific genes (*ycf1*, *accD*) and intergenic spacers, has significant implications for developing robust molecular markers for species identification, phylogenetic analysis, and population genetics studies within the *Musa* genus. Phylogenetic analysis provides valuable insight into the evolutionary relationships between cultivated and wild progenitors. The analysis reveals that maternal lineage is effectively preserved in chloroplast genomes with *M. acuminata* cv. Berangan and *M. balbisiana* cv. Nipah maintaining close phylogenetic ties to their respective wild progenitors. The findings emphasize the role of the chloroplast genome in tracking genetic identity across generations and highlight the minimal plastome divergence observed in cultivated accessions, suggesting that nuclear processes and clonal propagation have driven much of the genetic variation in these cultivars. These results not only enhance our understanding of *Musa* domestication but also provide a foundation for future breeding and germplasm conservation efforts, with chloroplast genome data serving as a reliable tool for cultivar identification and evolutionary studies.

Declarations

Conflicts of Interest:

The authors declare no conflict of interest.

Funding:

This research was funded by UMHIR High Impact Research grant (UM.C/625/1/HIR/MOHE/SC/14) and University of Malaya Grants: (RG018-09BIO)

Author Contribution

MUMS designed and carried out the experiment, collected the plant materials, conducted bioinformatic and genome analyses and wrote the first manuscript. RYO conceived the experiments. SZL contributed to the genome analysis and result interpretation YMY, and JAH contributed to the genome analysis, result interpretation and manuscript revision. All authors have read and agreed to the published version of the manuscript.

Data Availability

These data can be found on NCBI under BioProject number PRJNA1003655 (M. acuminata cv. Berangan) and PRJNA1003006 (M. balbisiana cv. Nipah).

References

1. Ashokkumar K, Elayabalan S, Shobana VG, Sivakumar P, Pandiyan M (2018) Nutritional value of banana (Musa spp.) cultivars and its future prospects: A review. 10.5958/2394-4471.2018.00013.8
2. Bashmil YM et al (2021) Screening and Characterization of Phenolic Compounds from Australian Grown Bananas and Their Antioxidant Capacity. Antioxidants 10(10):1521. https://doi.org/10.3390/antiox10101521
3. Borges CV, Amorim EP, Leonel M, Gomez HAG, Santos TPRD, Ledo CADS, Lima GPP (2019) Post-harvest physicochemical profile and bioactive compounds of 19 bananas and plantains genotypes. Bragantia 78(2):284–296. https://doi.org/10.1590/1678-4499.20180252
4. Borges CV, Maraschin M, Coelho DS, Leonel M, Gomez HAG, Belin MAF, Lima GPP (2020) Nutritional value and antioxidant compounds during the ripening and after domestic cooking of bananas and plantains. Food Res Int 132:109061. https://doi.org/10.1016/j.foodres.2020.109061
5. Bowdery D (1999) Phytoliths from tropical sediments: Reports from Southeast Asia and Papua New Guinea. Bull Indo-Pacific Prehistory Association 18:159–167
6. Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, Green ED, NISC Comparative Sequencing Program (2003) LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res 13(4):721–731
7. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. BMC Bioinformatics 10(1):421
8. Cauz-Santos LA (2025) Beyond conservation: the landscape of chloroplast genome rearrangements in angiosperms. New Phytol. https://doi.org/10.1111/nph.70364
9. Daniell H, Jin S, Zhu X-G, Gitzendanner MA, Soltis DE, Soltis PS (2021) Green giant—a tiny chloroplast genome with mighty power to produce high-value proteins: history and phylogeny. Plant Biotechnol J 19:430–447. https://doi.org/10.1111/pbi.13556
10. Daniell H, Nair SK, Guan H, Guo Y, Kulchar RJ, Torres MDT, Shahed-al-Mahmud M et al (2022) Debulking different Corona (SARS-COV-2 delta, omicron, OC43) and influenza (H1N1, H3N2) virus strains by plant viral trap proteins in chewing gums to decrease infection and transmission. Biomaterials 288:121671. https://doi.org/10.1016/j.biomaterials.2022.121671
11. Daniell H, Lin CS, Yu M, Chang WJ (2016) Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. Genome Biol 17(1):1–29. https://doi.org/10.1186/s13059-016-1004-2
12. Denham T, Haberle P, Lentfer SG, Fullagar C, Field R, Therin J, Porch M, N., Winsborough B (2003) Origins of agriculture at Kuk Swamp in the Highlands of New Guinea. Science 301(5630):189–193

13. Dong W, Xu C, Li C, Sun J, Zuo Y, Shi S, Zhou S (2015) ycf1, the most promising plastid DNA barcode of land plants. Sci Rep 5(1):8348
14. Food and Agriculture Organization of the United Nations (2023) – with major processing by Our World in Data. Banana production – FAO [dataset]. Food and Agriculture Organization of the United Nations, Production: Crops and livestock products [original data]. Retrieved January 16, 2025 from https://ourworldindata.org/grapher/banana-production
15. Frailey DC, Chaluvadi SR, Vaughn JN, Coatney CG, Bennetzen JL (2018) Gene loss and genome rearrangement in the plastids of five Hemiparasites in the family Orobanchaceae. BMC Plant Biol 18(1):30. https://doi.org/10.1186/s12870-018-1249-x
16. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics. Nucleic Acids Res 32(suppl2):W273–W279
17. Fu N, Ji M, Rouard M, Yan HF, Ge XJ (2022) Comparative plastome analysis of Musaceae and new insights into phylogenetic relationships. BMC Genomics 23(1):223. https://doi.org/10.1186/s12864-022-08454-3
18. Gornicki P, Zhu H, Wang J, Challa GS, Zhang Z, Gill BS, Li W (2014) The chloroplast view of the evolution of polyploid wheat. New Phytol 204(3):704–714
19. Greiner S, Lehwark P, Bock R (2019) OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res 47:W59–W64. https://doi.org/10.1093/nar/gkz238
20. Hoang DT, Vinh LS, Flouri T, Stamatakis A, Von Haeseler A, Minh BQ (2018) MPBoot: fast phylogenetic maximum parsimony tree inference and bootstrap approximation. BMC Evol Biol 18(1):11. https://doi.org/10.1186/s12862-018-1131-3
21. Hua Z, Tian D, Jiang C, Song S, Chen Z, Zhao Y, Jin Y, Huang L, Zhang Z, Yuan Y (2022) Towards comprehensive integration and curation of chloroplast genomes. Plant Biotechnol J 20(12):2239–2241. 10.1111/pbi.13923Epub 2022 Sep 20. PMID: 36069606; PMCID: PMC9674321
22. Jansen RK, Raubeson LA, Boore JL, DePamphilis CW, Chumley TW, Haberle RC, Cui L (2005) Methods for obtaining and analyzing whole chloroplast genome sequences. In *Methods in enzymology* (Vol. 395, pp. 348–384). Academic Press. https://doi.org/10.1016/S0076-6879(05)95020-9
23. Jin JJ, Yu WB, Yang JB, Song Y, DePamphilis CW, Yi TS, Li DZ (2020) GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol 21(1):241. https://doi.org/10.1186/s13059-020-02154-5
24. Kalyaanamoorthy S, Minh BQ, Wong TK, Von Haeseler A, Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14(6):587–589. https://doi.org/10.1038/nmeth.4285
25. Katoh K, Misawa K, Kuma KI, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30(14):3059–3066

26. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Drummond A (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12):1647–1649
27. Kwon EC, Kim JH, Kim NS (2020) Comprehensive genomic analyses with 115 plastomes from algae to seed plants: structure, gene contents, GC contents, and introns. Genes Genomics 42(5):553–570. https://doi.org/10.1007/s13258-020-00923-x
28. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359
29. Li B, Lin F, Huang P, Guo W, Zheng Y (2020) Development of nuclear SSR and chloroplast genome markers in diverse Liriodendron chinense germplasm based on low-coverage whole genome sequencing. Biol Res 53(1):21. https://doi.org/10.1186/s40659-020-00289-0
30. Li LF, Häkkinen M, Yuan YM, Hao G, Ge XJ (2010) Molecular phylogeny and systematics of the banana family (Musaceae) inferred from multiple nuclear and chloroplast DNA fragments, with a special reference to the genus Musa. Mol Phylogenet Evol 57(1):1–10
31. Liu J, Gao CW, Niu YF (2018) The complete chloroplast genome sequence of flowering banana, Musa ornata. Mitochondrial DNA Part B: Resour 3(2):962–963. https://doi.org/10.1080/23802359.2018.1507647
32. Low SL, Landrein S (2025) Evolutionary dynamics of plastomes within Arisaema (Araceae). *The Nucleus*, 1–12. https://doi.org/10.1007/s13237-025-00554-1
33. Martin G, Baurens FC, Cardi C, D'Hont A, Aury JM (2013) The Complete Chloroplast Genome of Banana (Musa acuminata, Zingiberales): Insight into Plastid Monocotyledon Evolution. PLoS ONE 8(6). https://doi.org/10.1371/journal.pone.0067350
34. Mazumdar P, Binti Othman R, Mebus K, Ramakrishnan N, Harikrishna A, J (2017) Codon usage and codon pair patterns in non-grass monocot genomes. Ann Botany 120(6):893–909. https://doi.org/10.1093/aob/mcx112
35. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 37(5):1530–1534. https://doi.org/10.1093/molbev/msaa131
36. Morley SA, Ahmad N, Nielsen BL (2019) Plant organelle genome replication. Plants 8(10):358. https://doi.org/10.3390/plants8100358
37. Mower JP, Ma PF, Grewe F, Taylor A, Michael TP, VanBuren R, Qiu YL (2019) Lycophyte plastid genomics: extreme variation in GC, gene and intron content and multiple inversions between a direct and inverted orientation of the rRNA repeat. New Phytol 222(2):1061–1075. https://doi.org/10.1111/nph.15650
38. Nock CJ, Waters DL, Edwards MA, Bowen SG, Rice N, Cordeiro GM, Henry RJ (2011) Chloroplast genome sequences from total DNA for plant identification. Plant Biotechnol J 9(3):328–333
39. Oldenburg DJ, Morley, Bendich AJ (2016) The linear plastid chromosomes of maize: terminal sequences, structures, and implications for DNA replication. Curr Genet 62(2):431–442.

https://doi.org/10.1007/s00294-015-0548-0

40. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A (2020) Using SPAdes de novo assembler. Curr protocols Bioinf 70(1):e102. https://doi.org/10.1002/cpbi.102
41. Schwenkert S, Fernie AR, Geigenberger P, Leister D, Möhlmann T, Naranjo B, Neuhaus HE (2022) Chloroplasts are key players to cope with light and temperature stress. Trends Plant Sci 27(6):577–587. 10.1016/j.tplants.2021.12.004
42. Sheng J, Yan M, Wang J, Zhao L, Zhou F, Hu Z, Diao Y (2021) The complete chloroplast genome sequences of five Miscanthus species, and comparative analyses with other grass plastomes. Ind Crops Prod 162:113248. https://doi.org/10.1016/j.indcrop.2021.113248
43. Shi C, Hu N, Huang H, Gao J, Zhao YJ, Gao LZ (2012) An improved chloroplast DNA extraction procedure for whole plastid genome sequencing. PLoS ONE 7(2):e31468. https://doi.org/10.1371/journal.pone.0031468
44. Song W, Ji C, Chen Z, Cai H, Wu X, Shi C, Wang S (2022) Comparative analysis the complete chloroplast genomes of nine Musa species: genomic features, comparative analysis, and phylogenetic implications. Front Plant Sci 13:832884. https://doi.org/10.3389/fpls.2022.832884
45. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S (2017) GeSeq–versatile and accurate annotation of organelle genomes. Nucleic Acids Res 45(W1):W6–W11. https://doi.org/10.1093/nar/gkx391
46. Wu CS, Sudianto E, Chiu HL, Chao CP, Chaw SM (2021) Reassessing Banana Phylogeny and Organelle Inheritance Modes Using Genome Skimming Data. Front Plant Sci 12:713216. https://doi.org/10.3389/fpls.2021.713216
47. Yu DL, Wei ZQ, Yan RR, Fei SP, Wu W, Hu GX (2025) Comparative Plastomics of Tropidia (Orchidaceae): Unraveling Structural Evolution and Phylogenetic Implications in Epidendroideae. Diversity 17(6):391. https://doi.org/10.3390/d17060391

Figures

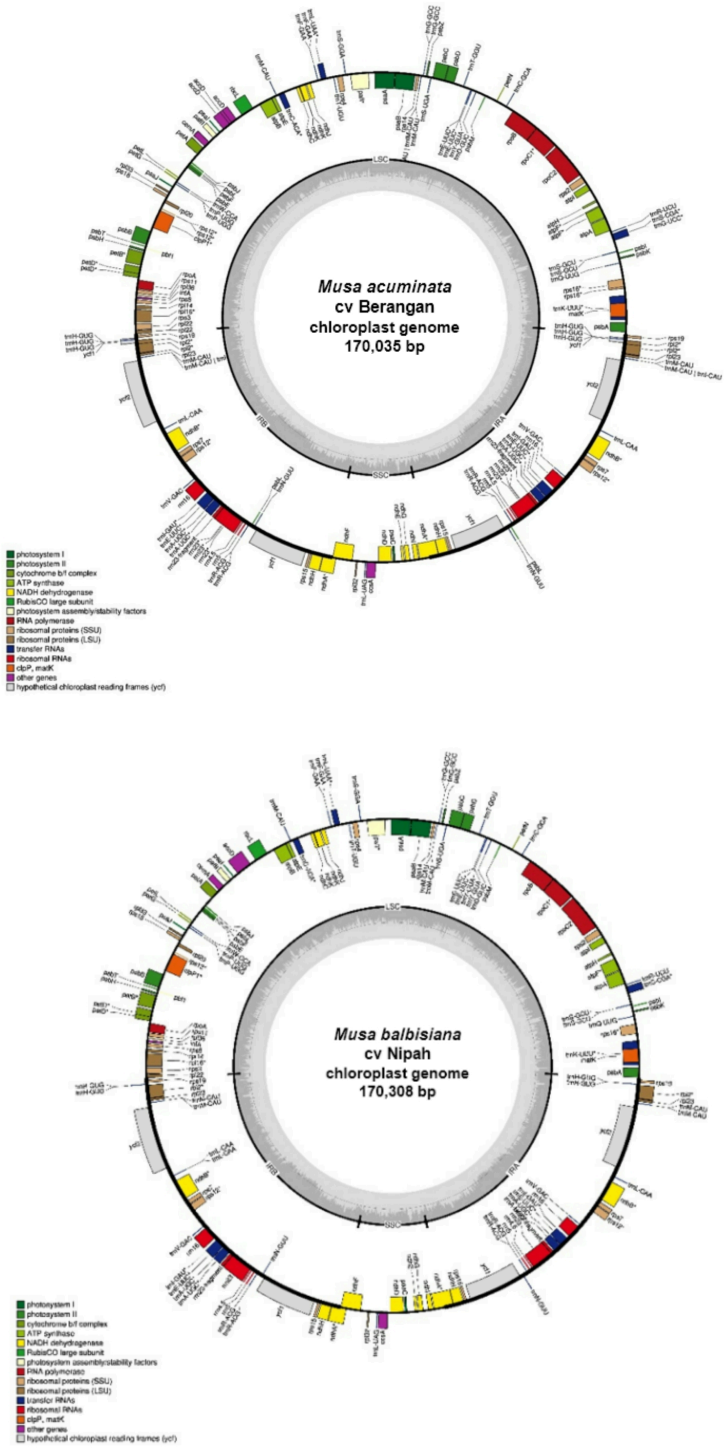

Figure 1

Structure of the assembled and annotated chloroplast genomes of *M. acuminata* cv. Berangan and *M. balbisiana* cv. Nipah. The colors indicate gene function. The genes presented outside the circle are transcribed counterclockwise, whereas the genes presented inside the circle are transcribed clockwise.

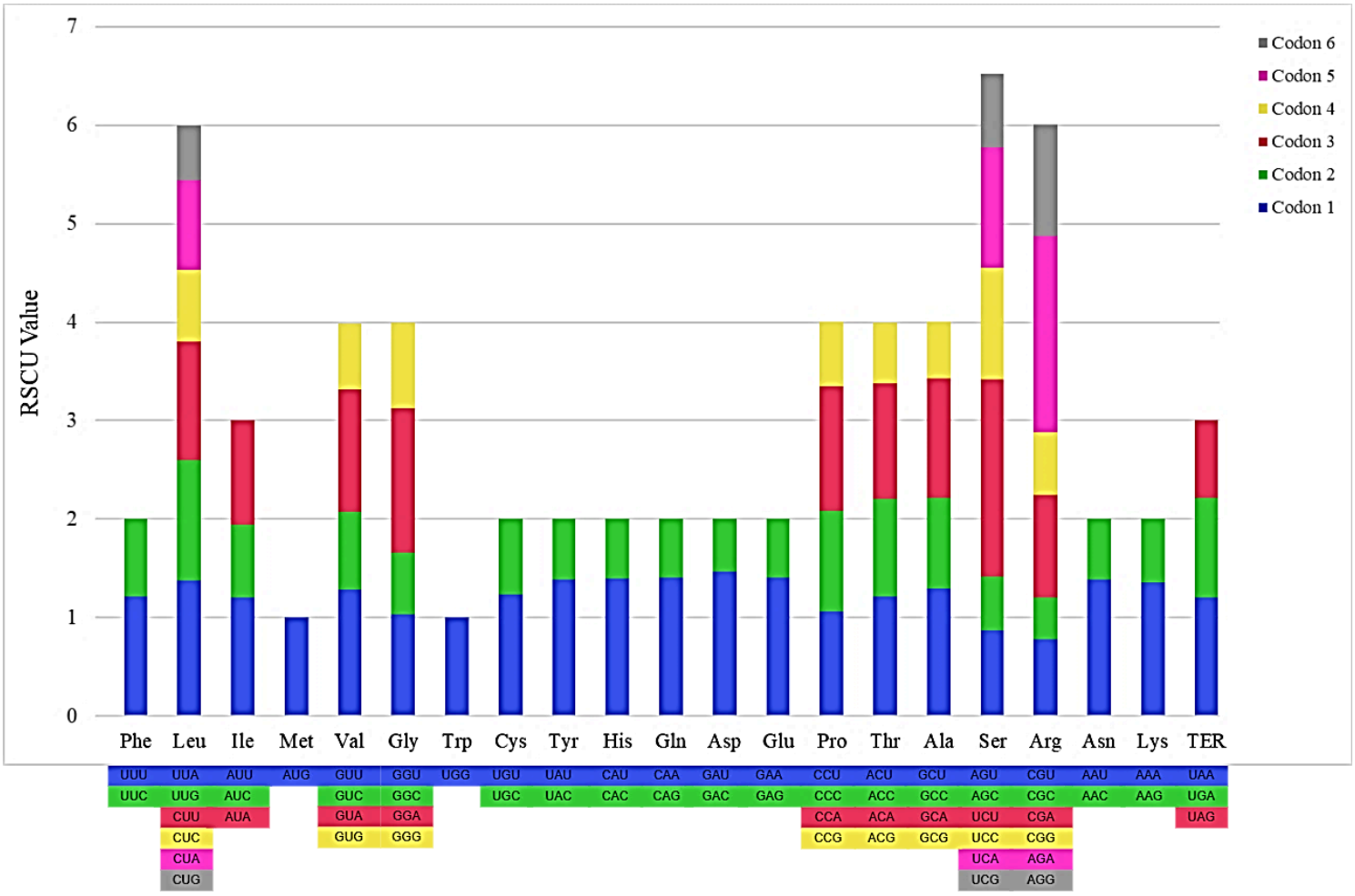

Figure 2

Relative synonymous codon usage (RSCU) analysis of *M. acuminata* cv. Berangan chloroplast genome. The histogram above each amino acid shows codon usage. Colors in the column graph reflect codons in the same colors shown below the figure. RSCU: relative synonymous codon usage; Ala: alanine; Arg: arginine; Asn: asparagine; Asp: asparagine; Cys: cysteine; Gln: glutamine; Glu: glutamic; Gly: glycine; His: histidine; Leu: leucine; Ile: isoleucine; Lys: lysine; Met: methionine; Phe: phenylalanine; Pro: proline; Ser: serine; Thr: threonine; Trp: tryptophan; Tyr: tyrosine; Val: valine.

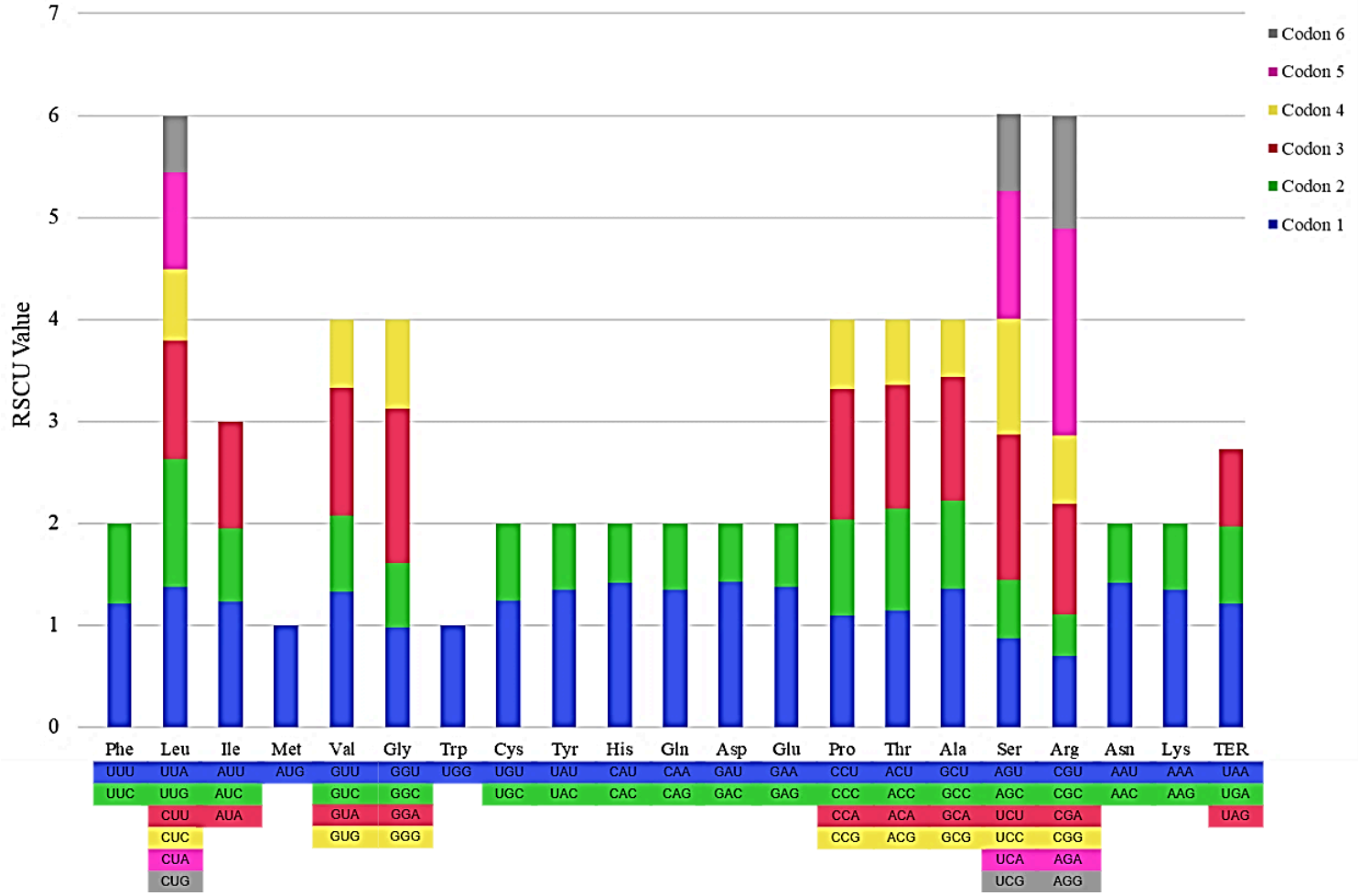

Figure 3

Relative synonymous codon usage (RSCU) analysis in the *M. balbisiana* cv. Nipah chloroplast genomes. The histogram above each amino acid shows codon usage. Colours in the column graph reflect codons in the same colours shown below the figure.

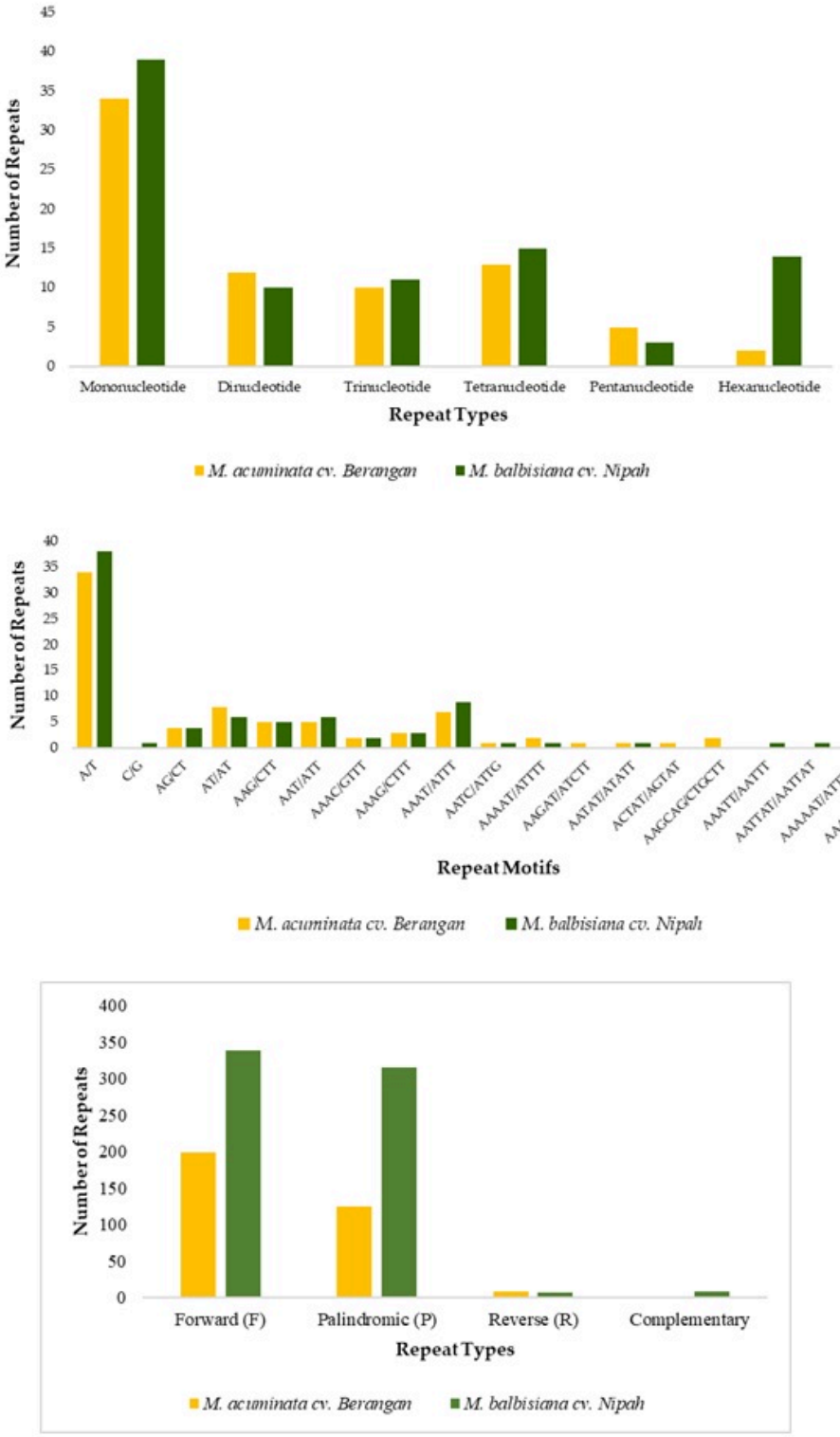

Figure 4

Frequency of simple sequence repeat (SSR) and repetitive regions in the *M. acuminata* cv. Berangan *and M. balbisiana* cv. Nipah chloroplast genomes. (a) Number of different SSRs types (Mononucleotide, dinucleotides, trinucleotides, tetranucleotide, pentanucleotide and hexanucleotide) (b) Number of different SSRs repeat unit types/motifs (c) The number of long repeat types in chloroplast genome.

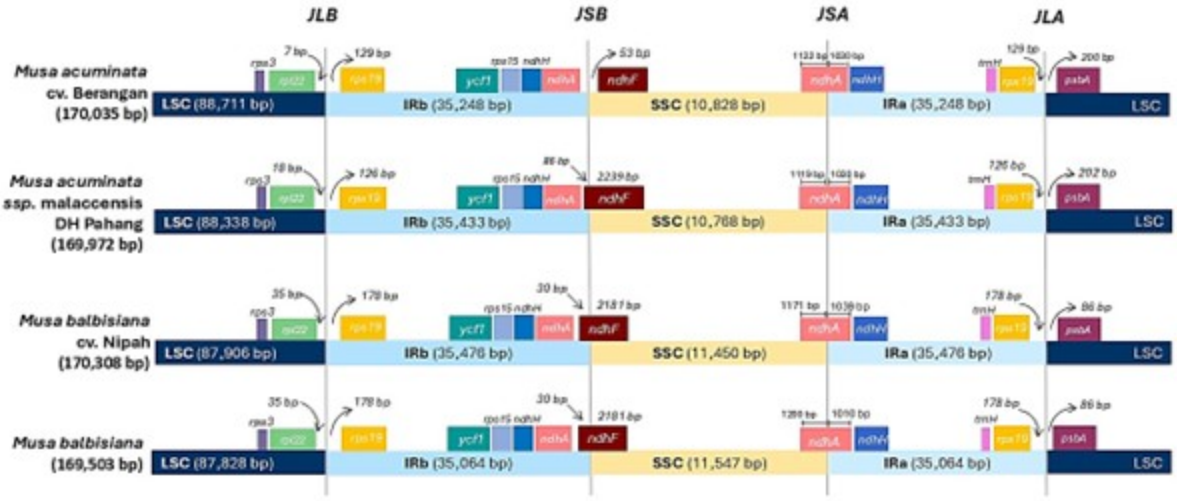

Figure 5

IR boundary junction was compared between the cultivated accession of M. *acuminata*cv. Berangan and *M. balbisiana*cv. Nipah with the wild accessions ofchloroplast genome.

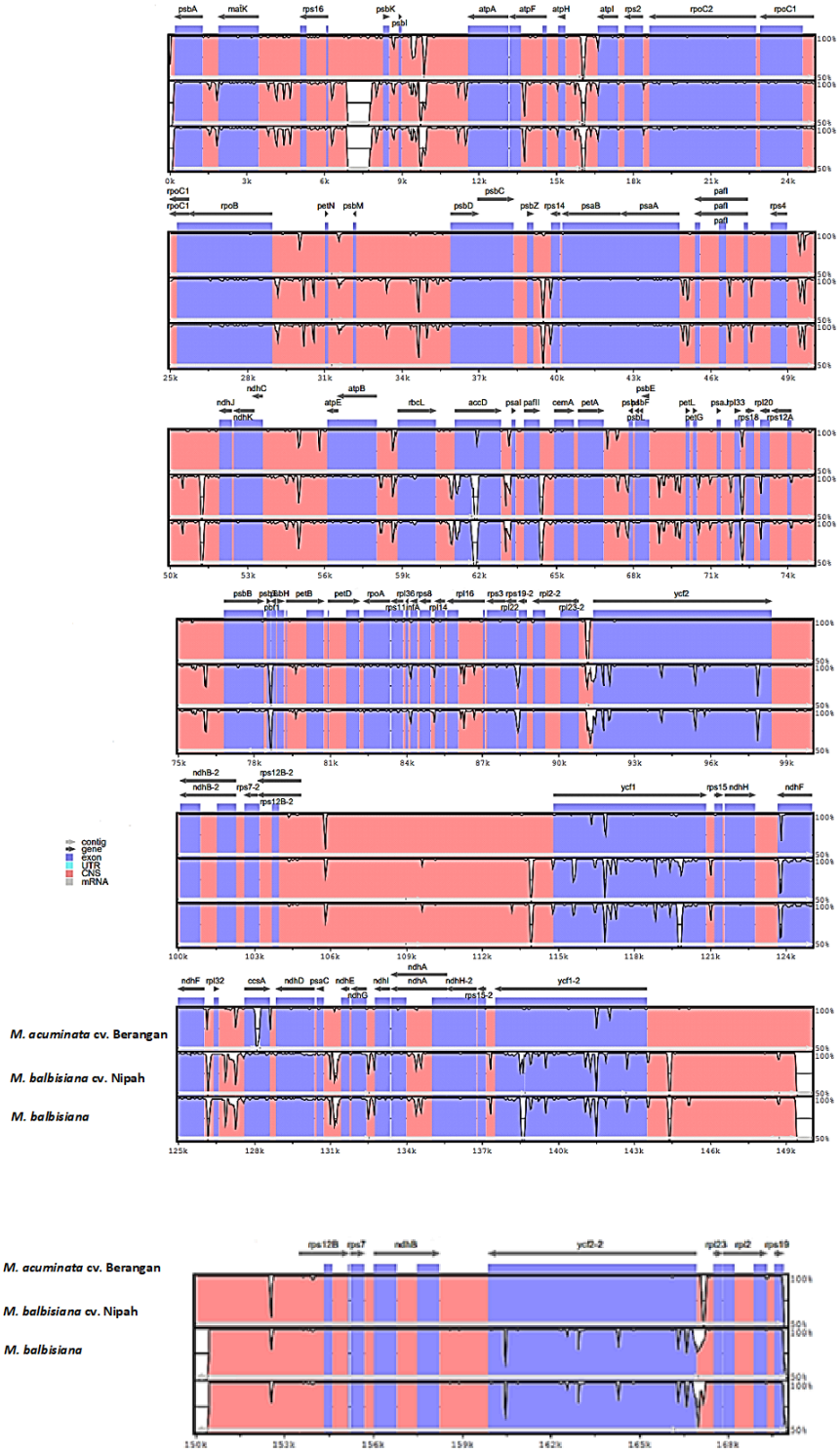

Figure 6

mVISTA identity plot comparing *M. acuminata* cv. Berangan, *M. balbisiana* cv. Nipah and wild accessions, *M. balbisiana* and considering *M. acuminata* ssp. malaccensis as a reference. The top line shows genes in order (transcriptional direction indicated by arrows). The *y*-axis represents the percent identity within 50–100%. The *x*-axis represents the coordinate in the chloroplast genome. Genome regions are color-

coded as protein-coding (exon), tRNAs, or rRNAs, and conserved noncoding sequences (intergenic region). The white block represents regions with sequence variation between two species.

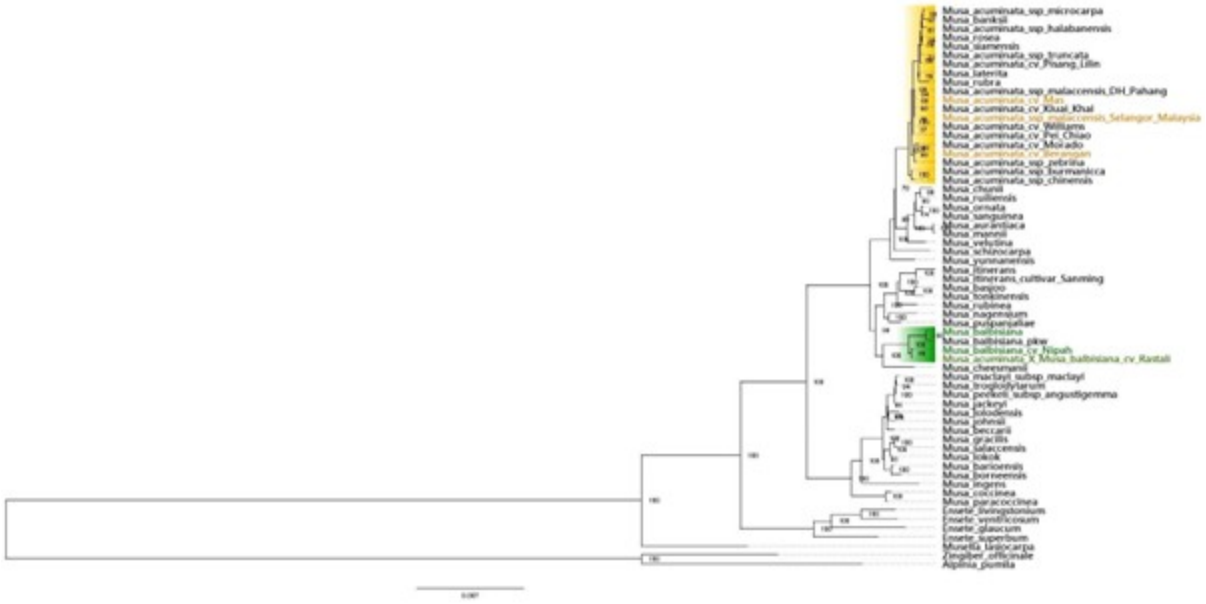

Figure 7

Maximum likelihood phylogenetic inferences of 70 protein coding sequences from 63 *Musa* accessions with *Alpinia* and *Zingiber*as outgroup, with 10,000 bootstrap.