Heliyon 10 (2024) e31248

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Characterization and phylogenetic analysis of the complete chloroplast genome of *Curcuma comosa* and *C. latifolia*

Bussarin Wachananawat [a], Bobby Lim-Ho Kong [b], Pang-Chui Shaw [b], Bhanubong Bongcheewin [c,d], Sunisa Sangvirotjanapat [d], Pinidphon Prombutara [e], Natapol Pornputtapong [f], Suchada Sukrong [a,*]

[a] *Center of Excellence in DNA Barcoding of Thai Medicinal Plants, Department of Pharmacognosy and Pharmaceutical Botany, Faculty of Pharmaceutical Sciences, Chulalongkorn University, Bangkok, 10330, Thailand*
[b] *Li Dak Sum Yip Yio Chin R & D Centre for Chinese Medicine and Institute of Chinese Medicine, The Chinese University of Hong Kong, Shatin, Hong Kong, N.T., China*
[c] *Department of Pharmaceutical Botany, Faculty of Pharmacy and Center of Excellence in Herbal Medicine and Natural Products, Faculty of Pharmacy, Mahidol University, Bangkok, 10400, Thailand*
[d] *Sireeruckhachati Nature Learning Park, Mahidol University, Nakhon Pathom, 73170, Thailand*
[e] *Faculty of Science, Omics Science & Bioinformatics Center, Chulalongkorn University, Bangkok, 10330, Thailand*
[f] *Department of Biochemistry and Microbiology, Faculty of Pharmaceutical Sciences, Chulalongkorn University, Bangkok, 10330, Thailand*

ARTICLE INFO

Keywords:
Curcuma comosa
Curcuma latifolia
Chloroplast genome
DNA barcode
Nucleotide diversity
Phylogenetic analysis

ABSTRACT

Members of the *Curcuma* genus, a crop in the Zingiberaceae, are widely utilized rhizomatous herbs globally. There are two distinct species, *C. comosa* Roxb. and *C. latifolia* Roscoe, referred to the same vernacular name "Wan Chak Motluk" in Thai. *C. comosa* holds economic importance and is extensively used as a Thai traditional medicine due to its phytoestrogenic properties. However, its morphology closely resembles that of *C. latifolia*, which contains zederone, a compound known for its hepatotoxic effects. They are often confused, which may affect the quality, efficacy and safety of the derived herbal materials. Thus, DNA markers were developed for discriminating *C. comosa* from *C. latifolia*. This study focused on analyzing core DNA barcode regions, including *rbc*L, *mat*K, *psb*A-*trn*H spacer and ITS2, of the authentic *C. comosa* and *C. latifolia* species. As a result, no variable nucleotides in core DNA barcode regions were observed. The complete chloroplast (cp) genome was introduced to differentiate between the two species. The comparison revealed that the cp genomes of *C. comosa* and *C. latifolia* were 162,272 and 162,289 bp, respectively, with a total of 133 identified genes. The phylogenetic analysis revealed that *C. comosa* and *C. latifolia* exhibited a very close relationship with other *Curcuma* species. The cp genome of *C. comosa* and *C. latifolia* were identified for the first time, providing valuable insights for species identification and evolutionary research within the Zingiberaceae family.

1. Introduction

Curcuma, a rhizomatous genus, belongs to the Zingiberaceae family, which is widely spread across tropical Asia, Australia, and the South Pacific Islands [1,2]. Many species of *Curcuma*, such as *C. longa* L., *C. aromatica* Salisb., and *C. zedoaria* (Christm.) Roscoe has

* Corresponding author.
E-mail address: suchada.su@chula.ac.th (S. Sukrong).

https://doi.org/10.1016/j.heliyon.2024.e31248
Received 2 November 2023; Received in revised form 23 April 2024; Accepted 13 May 2024
Available online 15 May 2024

been used in natural food additives, cosmetics and traditional medicine [3]. Apart from those species, *C. comosa*, or "Wan Chak Motluk" in Thai, is a phytoestrogen-producing herb, has been traditionally used for treatment estrogenic hormone deficits in women and sold in the market [4–6]. According to information from the customs department of the Kingdom of Thailand, *C. comosa* is the top eight most exported herbal plants [7] to many countries around the world such as Pakistan, Japan, Malaysia, the United States of America, and Netherlands, etc. [8]. However, there are plants in the genus *Curcuma* sharing the common name "Wan Chak Motluk" [9] including *C. comosa*, *C. latifolia* [10] and *C. elata* [9] that are difficult to distinguish by appearance (Fig. S1). Since the rarity of *C. elata* [11,12] and its absence in the IUCN report of Thailand [13] under the genus *Curcuma*, the cp genome of two other plants, *C. comosa* and *C. latifolia*, were explored in this study. *C. comosa* contains numerous diarylheptanoids as major active compounds. It possesses phytoestrogenic properties for treating uterine and ovarian abnormalities [12,14–17]. *C. latifolia* contains many compounds including a sesquiterpenoid compound, zederone, that can cause hepatotoxicity [12,18]. Therefore, the misidentification of *C. latifolia* as *C. comosa* has become a major concern for consumers and the herbal industry. The discrimination of *C. comosa* from *C. latifolia* is important for consumer safety.

Many identification methods have been used to authenticate *Curcuma* species. The traditional method is to use taxonomic keys based on morphological data. In particular, the inflorescence is suitable for the identification of *Curcuma* spp. However, inflorescences are not always available in a complete form due to the short-lived and highly seasonal flowering of these species. The limited inflorescence availability makes it challenging to distinguish between different *Curcuma* species [3,19]. Furthermore, the large variation in rhizome morphology among Zingiberaceae can lead to confusion in usage (Fig. S1 and Fig. S2). Thai Herbal Pharmacopoeia (THP), an official national standard compendium, provides standard quality control for many herbal drugs marketed in Thailand; however, there are no data available on *C. comosa* and *C. latifolia*. Thus, other methods for differentiating these plants are needed, such as chemical profiling, molecular cytogenetics, and molecular marker analysis. However, chemical patterns also have limitations, as the chemical contents can be influenced by factors such as cultivation, weather conditions, and harvesting time [20]. A few molecular cytogenetics studies have reported that *C. comosa* cultivars have chromosome numbers 2n = 63 and 42 [21,22], whereas *C. latifolia* has chromosome numbers 2n = 63 and 84 [9]. Recently, several DNA molecular techniques have been emerged as methods for the differentiation of various plants in *Curcuma*, including RAPD [23], ISSR [24], and AFLP techniques [25]. DNA barcoding is a molecular technique that can be differentiated between various species. The standard DNA barcodes for identifying *Curcuma* species are based on specific regions of chloroplast DNA (*mat*K, *rbc*L, and *psb*A-*trn*H intergenic spacer) and a nuclear DNA region (ITS). These DNA barcodes have been established as reliable markers for distinguishing between different *Curcuma* species [26,27]. Other chloroplast DNA regions, such as *rpo*B, *rpo*C1, *acc*D, *ndh*J, *trn*L-F, *rps*36-*rps*8, and *trn*S-*trn*fM, have been evaluated as barcode loci in Zingiberaceae [28,29]. Although traditional DNA barcoding markers have been extensively studied and utilized for species identification, they may have limitations in distinguishing closely related species [30,31] and we have not found any useful polymorphic sites in published DNA barcode sequences including *rbc*L, *mat*K, *ITS*2, and *psb*A-*trn*H spacer for differentiating between *C. comosa* and *C. latifolia*. Because of the lack of adequate variations in DNA barcode sequences, a new method is needed to improve species identification. Cp genome sequencing is one such tool and has recently been shown to successfully discriminate closely related species in the genera *Boesenbergia*, *Curcuma*, *Kaempferia*, and *Pyrgophyllum* [27,32,33]. So far, no research has undertaken the cp genome sequencing of *C. comosa* and *C. latifolia*. To bolster the potential for identification, plant classification, and quality control of these two common medicinal herbs in Thailand, it is imperative that the cp genomes of these two species be sequenced. The objective of this study is to analyze and compare the cp genomes of *C. comosa* and *C. latifolia*. The phylogenetic relationships of these two species were also explored using these genomes.

2. Materials and methods

2.1. Plant materials

Two cultivated authentic species, *C. comosa* Roxb. and *C. latifolia* Roscoe, were collected from Zingiberaceae Collection, Sireeruckhachat Nature Learning Park, Nakhon Pathom, Mahidol University, Thailand and identified by the taxonomist Dr. Bhanubong Bongcheewin at Mahidol University. The samples were collected following the appropriate guidelines and regulations. Voucher specimens of *C. comosa* and *C. latifolia* were deposited at Sireeruckhachat Nature Learning Park and assigned as PBM005645 and PBM005639, respectively.

2.2. PCR amplification for DNA barcoding

The genomic DNA (gDNAs) from authentic *C. comosa* and *C. latifolia* were isolated using a DNeasy Plant Mini Kit (QIAGEN, Germany) and utilized as a template for generating standard DNA barcodes including *rbc*L, *mat*K, ITS2, and *psb*A-*trn*H intergenic spacer. The DNA barcodes were amplified using specified primers in Table S1. PCR amplifications were carried out in a 50 μL reaction mixture containing 20–50 ng template gDNA, 1X PCR buffer, 3 mM $MgCl_2$, 0.2 mM dNTP mix, 0.5 μM of each primer and 0.5 U of Platinum *Taq* DNA polymerase (Invitrogen, USA). The cycling conditions included an initial denaturation at 94 °C for 3 min, followed by 30 cycles of denaturation at 94 °C for 30 s, annealing at 58 °C for 30 s, and extension at 72 °C for 40 s (for ITS2 and *psb*A-*trn*H) or 1 min 30 s (for *rbc*L and *mat*K), with a final extension at 72 °C for 5 min. PCR amplicons were analyzed on a 2 % agarose gel and visualized under a UV transilluminator. After purification with Exonuclease I-Shrimp Alkaline Phosphatase (ExoSAP) (New England Biolabs, USA), the PCR products were subjected to bi-directional sequencing with the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA) using the Sanger sequencing method at Macrogen, South Korea. The sequencing results were analyzed by MAFFT v.7.0 [34].

2.3. Cp genome sequencing

Intact chloroplasts were isolated from 10 g of fresh leaves of authentic *C. comosa* and *C. latifolia* using 40 % (v/v) Percoll solution and a Chloroplast Isolation Kit following the manufacturer's instruction (Sigma–Aldrich, USA). Subsequently, gDNA was extracted from the chloroplasts using a DNeasy Plant Mini Kit (QIAGEN, Germany). The quality of the cp gDNA was assessed by determining the A260/A280 ratio using a NanoDrop One UV–Vis Spectrophotometer (Thermo Scientific, USA) and precisely quantified via a Qubit 3 Fluorometer (Invitrogen, USA). Whole cp genome shotgun libraries were constructed with the SparQ Frag & DNA Library Prep Kit (Quantabio, USA). The average fragment length of the constructed DNA libraries was measured using an Agilent 2100 Bioanalyzer (GMI, USA). Paired-end sequencing was conducted on an Illumina NextSeq 500 (Illumina, USA).

2.4. Cp genome assembly and annotation

The raw Illumina paired-end data for the cp genomes of *C. comosa* and *C. latifolia* were trimmed and removed the low-quality bases using FASTP [35]. Default settings including adapter sequence autodetection, length, and quality filters were applied during this process. The filtered reads were then utilized for the assembly of cp genomes using GetOrganelle with organelle type and SPAdes kmer set with following options: F embplant_pt and -k 21, 45, 65, 85, 105 [36]. Subsequently, all paired-end reads were aligned back to the assembled cp genomes using SAMtools. The cp genome was annotated using the GeSeq automatic annotator, employing BLAST searches (default settings) against related species' cp genomes [37]. The circular map of the cp genome was generated using the online tool OGDRAW [38] with default settings, followed by manual adjustments. Finally, the completed assembly was deposited in GenBank (https://www.ncbi.nlm.nih.gov) along with the corresponding accession numbers.

2.5. Codon usage analysis

The protein-coding genes within the cp genomes of *C. comosa* and *C. latifolia* were examined for relative synonymous codon usage (RSCU) via Mega-X software [39]. In cases where RSCU equals 1, codon usage is considered unbiased. If RSCU exceeds 1, it indicates that the specific codon frequency is higher than that of other synonymous codons; conversely, if RSCU is below 1, the frequency is low. Clustering outcomes derived from RSCU values have the potential to serve as valuable indicators for deducing phylogenetic relationships [40].

2.6. Repeat element analysis

The cp genome sequences (Table S2) belonging to plants from five different genera in the family Zingiberaceae, namely, *Curcuma*, *Zingiber*, *Hedychium*, *Kaempferia*, and *Amomum*, were obtained from the NCBI database for repeat element analysis along with our authentic *C. comosa* and *C. latifolia* cp sequences. The MIcroSAtellite (MISA) [41] tool was employed for analyzing simple sequence repeats (SSRs), with the minimum repeat set at 10, 5, 4, 3, 3, and 3 for mono-, di-, tri-, tetra-, penta-, and hexa-nucleotides, respectively. To identify long repeat sequences, including forward, reverse, complement, and palindromic repeat units, the REPuter software [42] was utilized. The parameters were configured with a Hamming distance of 3, a maximum computed repeats limit of 150, and a minimum repeat size requirement of 30.

2.7. Sequence divergence analysis

The cp genomes of plants within the Zingiberaceae family (Table S2) were compared and analyzed for variable regions by using the mVISTA tool in Shuffle-LAGAN mode [43]. *C. comosa* was used as the reference of the annotated cp genome. Boundaries between LSC, SSC and IR regions were manually defined to detect any differences in gene rearrangement between regions.

The cp genomes (Table S3) belonging to plants of different genera in the family Zingiberaceae including *Curcuma*, *Zingiber*, *Hedychium*, *Kaempferia*, *Amomum*, *Alpinia*, *Wurfbainia*, *Stahlianthus* and *Lanxangia*, in a total of 33 samples were obtained from the NCBI database for sequence divergence analysis. Nucleotide diversity (π) was computed using DnaSP v.6 [44]. The sliding window analysis was conducted with a window length of 600 bp and a step size of 200 bp. In addition, the sequences for each gene and coding region were generated using BioPython [45] based on annotation information provided by GenBank files and then converted into two separate FASTA files. The alignment of gene and coding region FASTA files used MAFFT v.7.0 [34]. Then, the nucleotide diversity (π) and total mutation count (η) of genes and coding regions were calculated for each region by using the nucleotide diversity function provided by DendroPy Library [46] and an in-house Python script, respectively.

2.8. Indel/SNP detection

The cp genomes of plants from various genera within the Zingiberaceae family (Table S2) were aligned with MAFFT v.7.0. [34], and the sequences were edited using Mega-X software [39]. Indels/SNPs were identified using DnaSP v.6 [44], with the cp genome of *C. comosa* serving as the reference.

2.9. Phylogenetic analysis

The cp genomes (Table S4) in the order Zingiberales including *Musaceae*, *Heliconiaceae*, *Strelitziaceae*, *Cannaceae*, *Costaceae* and *Zingiberaceae* were used for phylogenetic analysis. *Typha latifolia* L., a monocotyledon plant, was employed as an outgroup. MAFFT v7.0 [34]. was used to align the cp genome sequences. The maximum likelihood (ML) tree was constructed with IQ TREE v2.0 [47] using default parameters, and the most suitable substitution model was determined with ModelFinder [48]. In addition, Ultrafast Bootstrap (1000 replicates) [49] and the SH-aLRT branch test [50] were performed.

3. Results

3.1. Species discrimination based on standard DNA barcode

The standard DNA barcode regions including *rbc*L, *mat*K, ITS2, and the *psb*A-*trn*H intergenic spacer of *C. comosa* and *C. latifolia* were

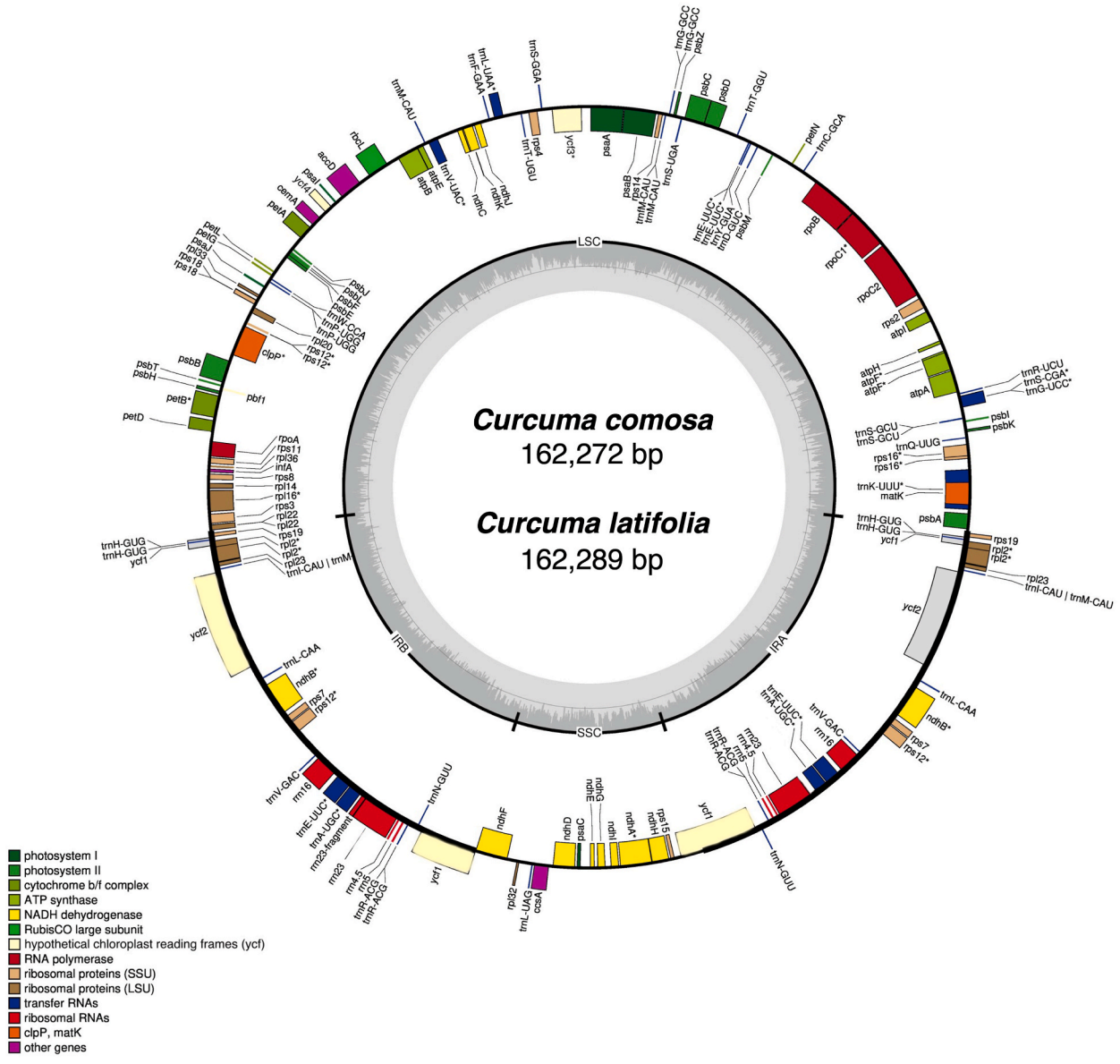


Fig. 1. Gene maps of the *C. comosa* and *C. latifolia* cp genomes. Genes drawn inside and outside of the circles are transcribed clockwise and counterclockwise, respectively. Genes are color coded by functional group. Asterisks indicate intron-containing genes. The darker gray and lighter gray areas in the inner circle represent the GC and AT contents, respectively.

1500, 1300, 500, and 800 bp in length, respectively. Nucleotide sequences in these DNA barcode regions were found to be identical in samples of the same plant species collected from various locations. PCR product size and nucleotide alignment results showed that there were no variations in any of those regions between *C. comosa* and *C. latifolia* (Fig. S3). The nucleotide sequences of all four regions were deposited in GenBank along with their accession numbers (Table S5).

3.2. Cp genome sequencing and features

The cp genome sequences of *C. comosa* and *C. latifolia* were sequenced by Illumina high-throughput. All paired-end reads showed 87 % of reads mapped and 28X genome coverage. The lengths of the cp genomes of *C. comosa* and *C. latifolia* were determined to be 162,272 bp and 162,289 bp, respectively. Both cp genomes had typical quadripartite structures consisting of a large single-copy (LSC) region of 87,074 bp in *C. comosa* and 87,089 bp in *C. latifolia*, a small single-copy (SSC) region of 15,698 bp in *C. comosa* and 15,700 bp in *C. latifolia*, and two inverted repeat (IR) regions of 29,750 bp in both species. The GC contents of both cp genomes were identical, at 34 % (LSC), 29.7 % (SSC), 41.2 % (IR), and 36.2 % (total). Furthermore, the cp genomes of the two species were anticipated to contain a total of 133 genes, comprising 87 protein-coding genes, 38 tRNA genes, and 8 rRNA genes (Table S6).

3.3. Gene categorization and functional classification

The genes were categorized into 4 main groups based on their functions, including self-replication (4 ribosomal RNA genes, 30 transfer RNA genes, 12 small ribosomal subunit genes, 9 large ribosomal subunit genes, and 4 DNA-dependent RNA polymerase genes), photosynthesis (5 photosystem I genes, 14 photosystem II genes, 6 cytochrome *b*/f complex genes, 6 ATP synthase genes, 1 ATP-dependent protease gene, 1 Rubisco large subunit gene, and 11 NADH dehydrogenase genes), other (maturase, envelope membrane protein, acetyl-CoA-carboxylase, c-type cytochrome synthesis, and translation initiation factor), unknown (4 genes), and intragenic regions (18 genes) (Fig. 1 and Table 1).

3.4. Codon usage

The relative synonymous codon usage (RSCU) was computed for five different genera within the Zingiberaceae family (Table S2), which included *C. comosa* and *C. latifolia*. The usage of the start codons methionine (AUG) and tryptophan (UGG) was no bias (RSCU = 1). The 87 protein-coding genes contained approximately 28,393 codons. Notably, leucine encoded by UUA exhibited the highest RSCU value, approximately 1.94, while alanine encoded by GCG had the lowest RSCU, around 0.39. It was observed that all preferred synonymous codons with A or U at the third position exhibited a higher bias (RSCU >1) compared to those with G or C (Table S7).

Table 1
Gene categorization and functional classification of the *C. comosa* and *C. latifolia* cp genomes.

Gene category	Groups of genes	Names of genes
Self-replication	Ribosomal RNAs	*rrn*4.5(x2), *rrn*5(x2), *rrn*16(x2), *rrn*23(x2)
	Transfer RNAs	**trn*A-UGC(x2), *trn*C-GCA, *trn*D-GUC, *trn*E-UUC, *trn*F-GAA, *trn*G-GCC, **trn*G-UCC, *trn*H-GUG(x2), *trn*I-CAU(x2), **trn*I-GAU(x2), **trn*K-UUU, *trn*L-CAA(x2), **trn*L-UAA, *trn*L-UAG, *trn*M-CAU, *trn*N-GUU(x2), *trn*P-UGG, *trn*Q-UUG, *trn*R-ACG(x2), *trn*R-UCU, *trn*S-GGA, *trn*S-GCU, *trn*S-UGA, *trn*T-GGU, *trn*T-UGU, *trn*V-GAC(x2), **trn*V-UAC, *trn*W-CCA, *trn*Y-GUA, *trn*fM-CAU
	Small ribosomal subunit	*rps*2, *rps*3, *rps*4, *rps*7(x2), *rps*8, *rps*11, **rps*12(x2), *rps*14, *rps*15, **rps*16, *rps*18, *rps*19(x2)
	Large ribosomal subunit	**rpl*2(x2), *rpl*14, **rpl*16, *rpl*20, *rpl*22, *rpl*23(x2), *rpl*32, *rpl*33, *rpl*36
	DNA-dependent RNA polymerase	*rpo*A, *rpo*B, **rpo*C1, *rpo*C2
Photosynthesis	Photosystem I	*psa*A, *psa*B, *psa*C, *psa*I, *psa*J
	Photosystem II	*psb*A, *psb*B, *psb*C, *psb*D, *psb*E, *psb*F, *psb*I, *psb*J, *psb*K, *psb*L, *psb*M, *psb*H, *psb*T, *psb*Z
	Cytochrome *b*/f complex	*pet*A, **pet*B, **pet*D, *pet*G, *pet*L, *pet*N
	ATP synthase	*atp*A, *atp*B, *atp*E, **atp*F, *atp*H, *atp*I
	ATP-dependent protease subunit p gene	**clp*P
	Rubisco large subunit	*rbc*L
	NADH dehydrogenase	**ndh*A, **ndh*B(x2), *ndh*C, *ndh*D, *ndh*E, *ndh*F, *ndh*G, *ndh*H, *ndh*I, *ndh*J, *ndh*K
Other genes	Maturase	*mat*K
	Envelope membrane protein	*cem*A
	Acetyl-CoA-carboxylase	*acc*D
	c-type cytochrome synthesis gene	*ccs*A
	Translation initiation factor	*inf*A
Genes of unknown function	Conserved open reading frames	*ycf*1(x2), *ycf*2(x2), **ycf*3, *ycf*4

Intron-containing genes are labeled with an asterisk. (x2) indicates duplicated genes in IR regions.

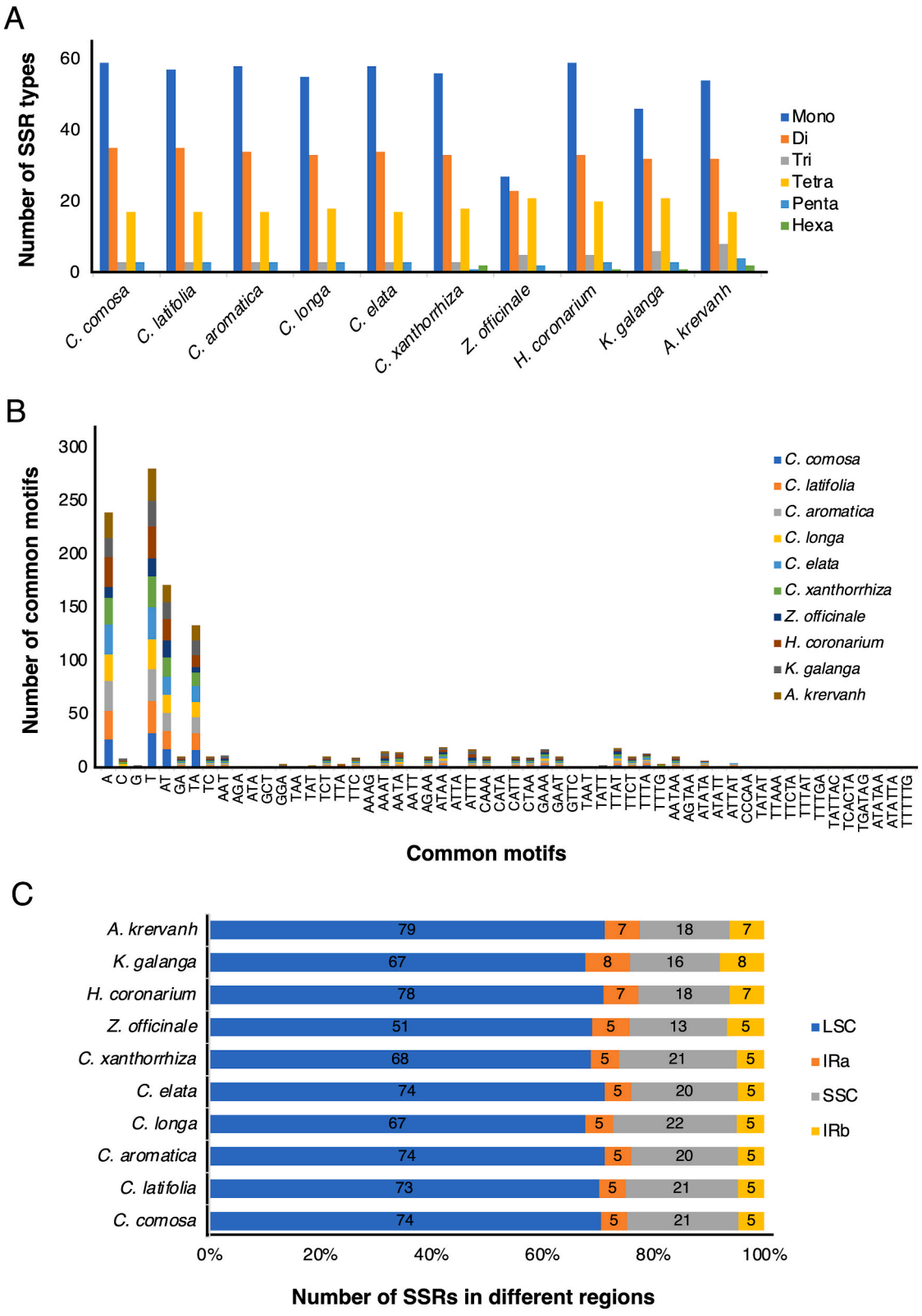


Fig. 2. Simple sequence repeat (SSR) analysis of the ten Zingiberaceae cp genomes. (A) Number of different SSR types. (B) number of common motifs. (C) number of SSRs in the LSC, SSC, and IR regions.

3.5. Repeat structure analysis

The cp genome sequences of the plants in Zingiberaceae (Table S2) were retrieved for analysis of SSR and long repeat. A total of 78–121 SSRs were identified within the cp genomes of plants belonging to five different genera within the Zingiberaceae family (Fig. 2). Mononucleotide repeats were the most prevalent among the various types of SSRs, comprising 27–58 loci, followed by dinucleotide (32–34 loci), tetranucleotide (17–21 loci), trinucleotide (3–8 loci), pentanucleotide (1–4 loci), and hexanucleotide (0–2 loci) repeats (Fig. 2A). Mononucleotide SSRs exhibited a notable abundance of A/T repeats (239–280 loci) (Fig. 2B). The distribution of SSR repeats was mainly observed in the LSC regions (51–79 loci), while only a small portion was found in the SSC regions (13–22 loci) and IR regions (5–8 loci) (Fig. 2C). The long repeat analysis identified a total of 39–79 long repeat sequence types (Fig. 3). Among the various types of long repeats, forward repeats (9–28 loci) were the most prevalent, followed by palindromic (8–28 loci), reverse (4–16 loci), and complement (1–10 loci) repeats (Fig. 3A). Repeat lengths within the range of 30–39 bp were the most frequent among the cp genomes analyzed in this study (Fig. 3B).

3.6. Comparison with the cp genome from zingiberaceae

To compare the sequence divergences of *C. comosa* and *C. latifolia*, the cp genome sequences of selected species in Zingiberaceae (Table S2) were included for comparison, and *C. comosa* was used as the reference. Generally, the coding regions were found to be more conserved compared to the noncoding regions across the species of Zingiberaceae; however, *rpo*C2, *rpo*B, *ycf*1, *ycf*2, and *ndh*F exhibited some degree of variation. The two IR regions exhibited lower divergence compared to the LSC and SSC regions. In contrast, high levels

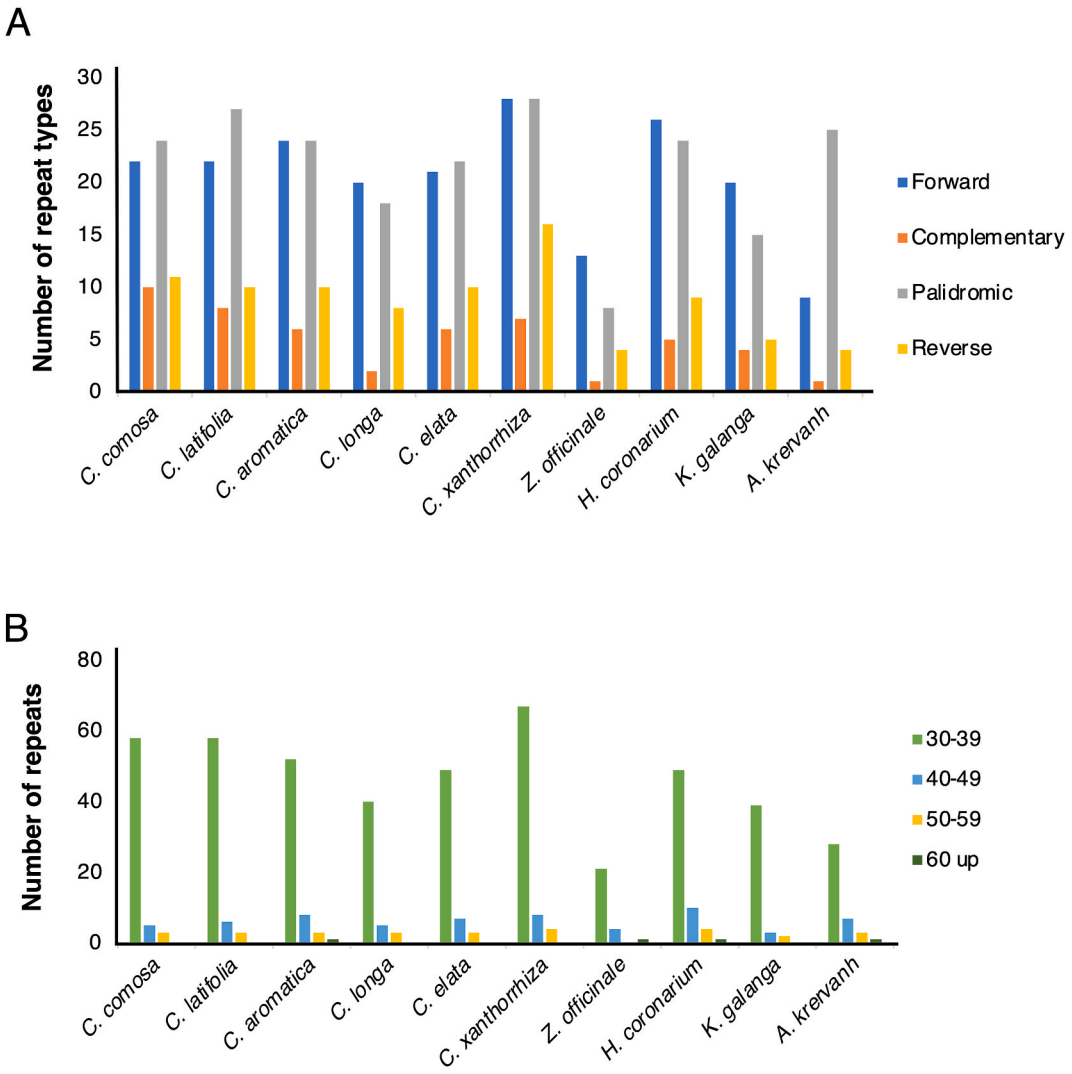
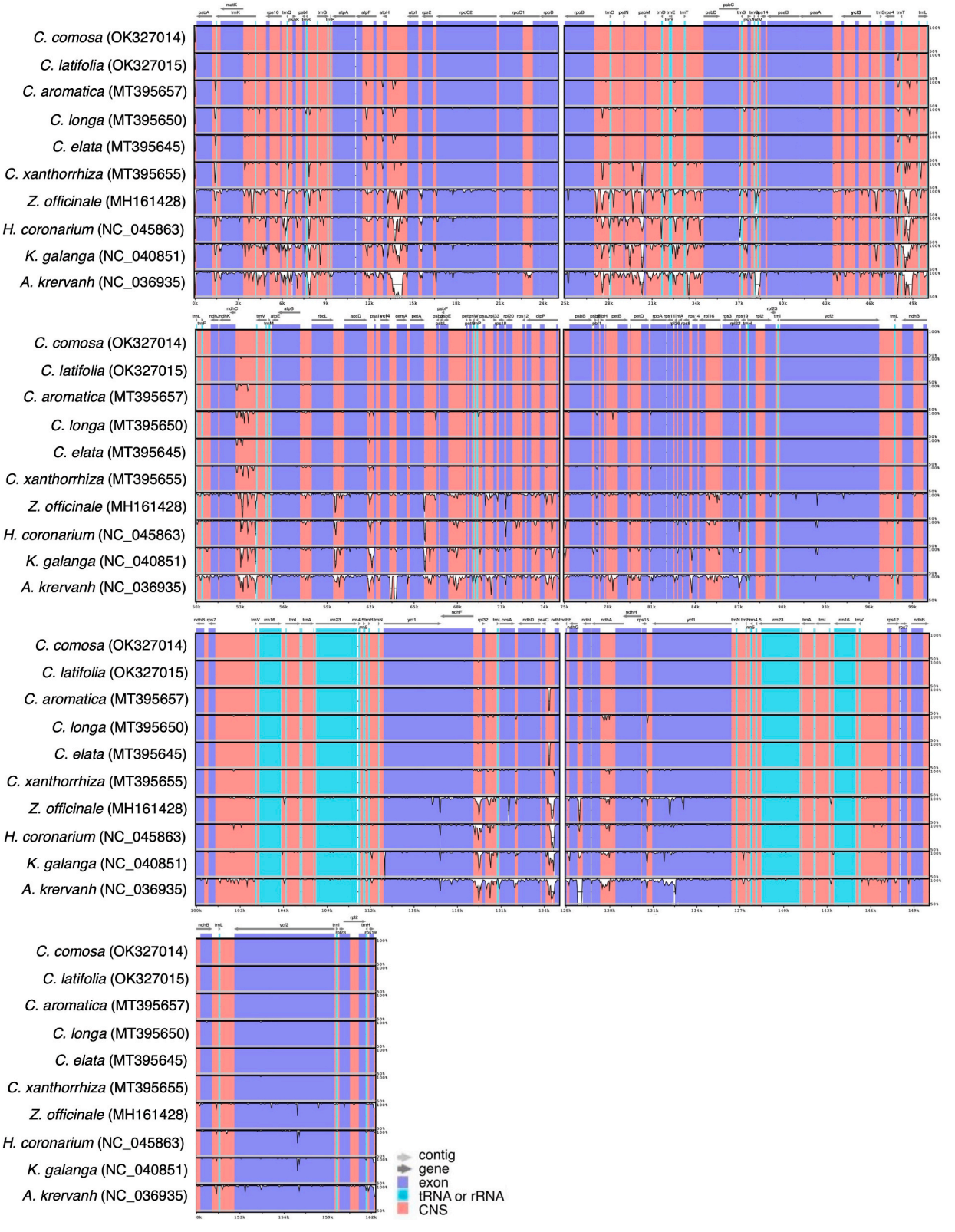


Fig. 3. Long repeat sequence analysis of the ten Zingiberaceae cp genomes. (A) Number of repeat types. (B) Number of repeats.



(caption on next page)

Fig. 4. Sequence alignment of the ten Zingiberaceae cp genomes using mVISTA. The *C. comosa* cp genome was used as a reference. Gray arrows and thick black lines above the alignment indicate the gene orientation. Purple bars represent exons, sky-blue bars represent transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs), and red bars represent noncoding sequences (NCSs). The horizontal axis indicates the coordinates within the cp genome. The vertical scale represents the percent identity, ranging from 50 to 100 %. White peaks represent regions with sequence variation among the ten species.

of divergence were identified in the intergenic regions of *trn*K-*rps*16, *rpo*B-*trn*C, *rps*4-*trn*T, *trn*T-*trn*L, *ndh*C-*trn*V, and *ndh*F-*rpl*32 (Fig. 4). Apart from nucleotide divergence, the contraction and expansion of the border regions were examined for the ten Zingiberaceae species (Table S2). The LSC/IRa, LSC/IRb, SSC/IRa and SSC/IRb, were found to be almost the same (29,642 bp to 29,797 bp). At the boundary of the LSC/IRb region in each cp genome had the *rpl*22-*rps*19 genes. Specifically, the *rpl*22 gene was positioned 21 bp to 48 bp to the left side of the LSC/IRb boundary, while the *rps*19 gene was located 129 bp to 148 bp to the right side of this boundary. Moving on, the *ycf*1-*ndh*F genes were found at the IRb/SSC boundary. The IRb/SSC junction was within the *ycf*1 region, extending between 7 bp to 205 bp into the SSC region. The *ndh*F gene was positioned 8 bp to 218 bp to the right side of the IRb/SSC boundary. The SSC/IRa junctions in the cp genomes were encompassed by the *ycf*1 genes, with a distance of 3705 bp to 3899 bp in the IRa region. Lastly, the *rps*19-psbA genes were located at the boundary of the IRa/LSC region, with the *rps*19 gene was positioned 129 bp to 148 bp to the left side of the IRa/LSC boundary, and the *psb*A gene was situated 109 bp to 125 bp to the right side of this boundary (Fig. 5).

3.7. Highly variable sequences

The cp genome sequences belonging to thirty-three species of Zingiberaceae in this study (Table S3) were analyzed for DNA markers. As expected, the sliding window analysis revealed the highest variation in the LSC and SSC regions, with comparatively lower variation in the IR regions (Fig. 6). The average nucleotide diversity (π) value was calculated to be 0.0096 across all Zingiberaceae species included in this study (Table S8). Mutational hotspots were found in six genes, *rps*16-*trn*Q, *ycf*1, *ndh*A, *ndh*I, *ndh*D, and RF19; these sites exhibited remarkable π values higher than 0.03 (Fig. 6A). In addition, the average nucleotide diversity value among twenty species in *Curcuma* was 0.0018 (Table S8), and the mutational hotspots including *rps*16-*trn*Q, *pet*N-*psb*M, and *ndh*A exhibited π values higher than 0.01 (Fig. 6B). Furthermore, the level of divergence among nine different genera in Zingiberaceae (Table S4) varied across different regions of the cp genome. The π value was 0.0080 among all Zingiberaceae species in this study, and no genetic divergence was observed in the rRNA regions. The *ndh*A gene showed the highest π values (0.0490) in coding regions (Table S9). These findings indicated that sequence variability depended on the region. Additionally, there were six SNPs and forty-one indels in *C. comosa* and *C. latifolia* (Table S10).

3.8. Phylogenetic reconstruction with the cp genome sequences of C. comosa and C. latifolia

To investigate the phylogenetic placements of the *C. comosa* and *C. latifolia* species and their relationships within Zingiberales (Table S4), maximum likelihood phylogenetic analyses were carried out utilizing cp genomes from forty species belonging to six families of Zingiberales. In this analysis, six families in Zingiberales were divided into two clades with 100 % bootstrap support (BS) values. One clade was composed of five families, including Musaceae, Strelitziaceae, Heliconiaceae, Cannaceae, and Costaceae, while the other clade included only Zingiberaceae. The clade containing Zingiberaceae was divided into two groups. The first group included four genera (*Alpinia*, *Lanxangia*, *Amomum*, and *Wurfbainia*) (BS = 100 %), and the second group included five genera (*Curcuma*, *Stahlianthus*, *Hedychium*, *Kaempferia*, and *Zingiber*). The second group was further divided into four subgroups (BS = 59–100 %). Subgroup4 was the most complex, with fourteen species, including the species of interest, *C. comosa* and *C. latifolia*, on the same branch as *C. elata* and *C. aromatica* (BS = 61–100 %) (Fig. 7).

4. Discussion

Accurately distinguishing between herbal materials based solely on their morphological traits presents challenges. Consequently, DNA barcoding has emerged as an alternative approach to such discrimination. Despite significant advancements in DNA barcoding technology, none of the identified barcode regions have reliably accomplished the task of identifying all plant species. Furthermore, pinpointing species with closely related genetics remains a daunting task [51]. *C. comosa* and *C. latifolia*, sharing the same vernacular name “Wan Chak Motluk” were identified based on DNA regions including the *rbc*L, *mat*K, *psb*A-*trn*H spacer and ITS2. Unfortunately, there were no variations observed in any of those core barcode regions (Fig. S3). This finding aligns with earlier research indicating low variation in nucleotide sequences within the core barcode regions among closely related species, including species within *Curcuma* [27], *Chrysanthemum* [31], *Cymbidium* [52], *Ligularia* [53], and *Phyllanthus* [54]. Therefore, molecular markers based on the cp genome should be further explored for discriminating *C. comosa* from *C. latifolia*. In previous reports, closely related plant species in *Hedyotis* [55], *Gentiana* [56], *Fritillaria* [57], *Paeoniae* [58], and *Polygonatum* [59] were successfully identified at the species level based on divergent sequences at selected hotspot regions in the cp genome.

Previous research has explored the cp genomes of certain *Curcuma* species, identifying divergent regions suggested for authentication [33]. However, prior investigations have excluded *C. comosa* and *C. latifolia*. In this study, no substantial differences were observed in the total number of genes, protein-coding genes, tRNAs, or rRNAs between *C. comosa* and *C. latifolia*. Similar findings were observed in other plant species in Zingiberaceae [32,33,60–68]. The frequencies of codon usage for each triplet nucleotide sequence

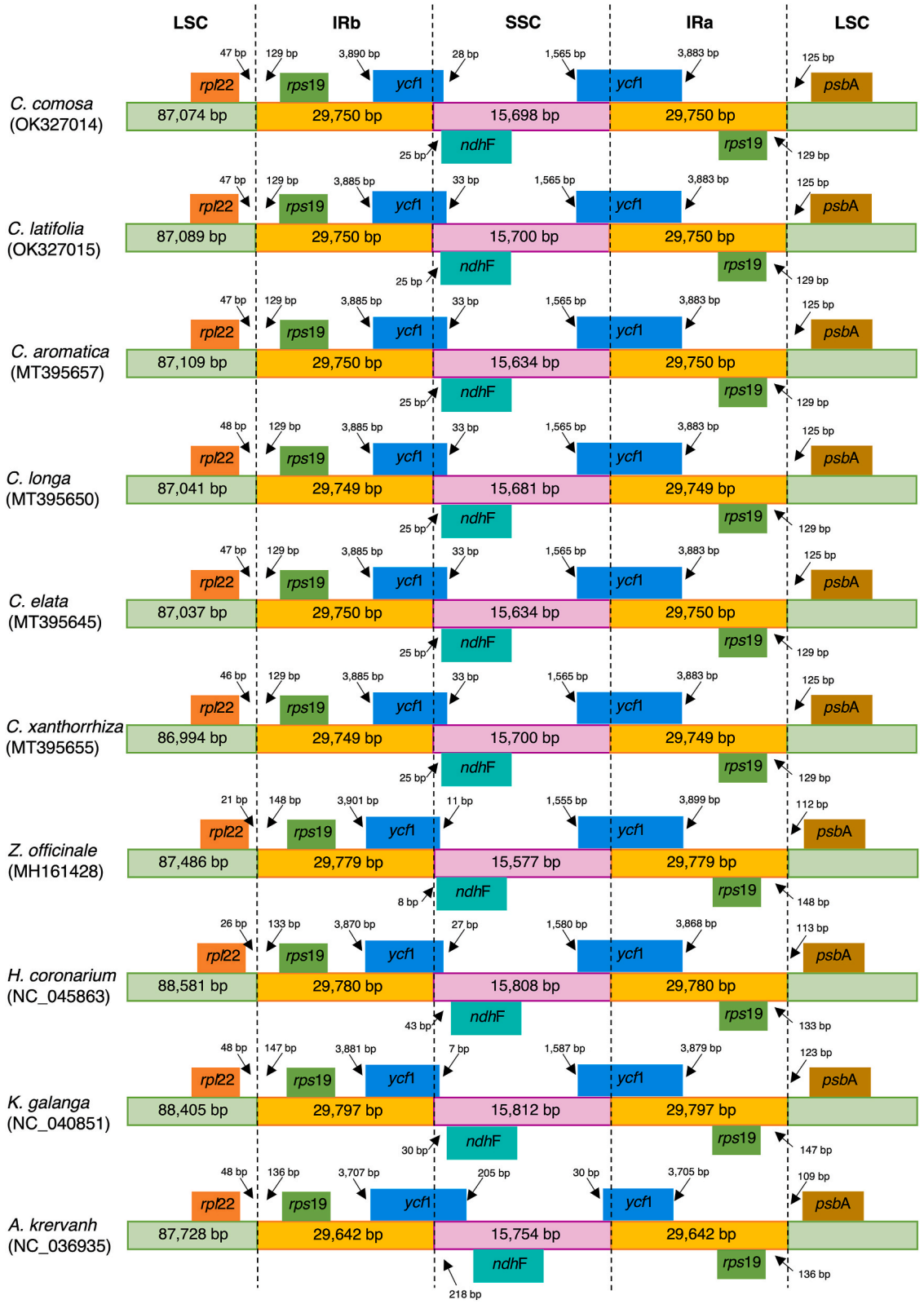


Fig. 5. Comparison of LSC, SSC and IR regional boundaries among the plants in Zingiberaceae. The numbers above the genes denote the distance between the end of the gene and the border sites. The figure is not to scale.

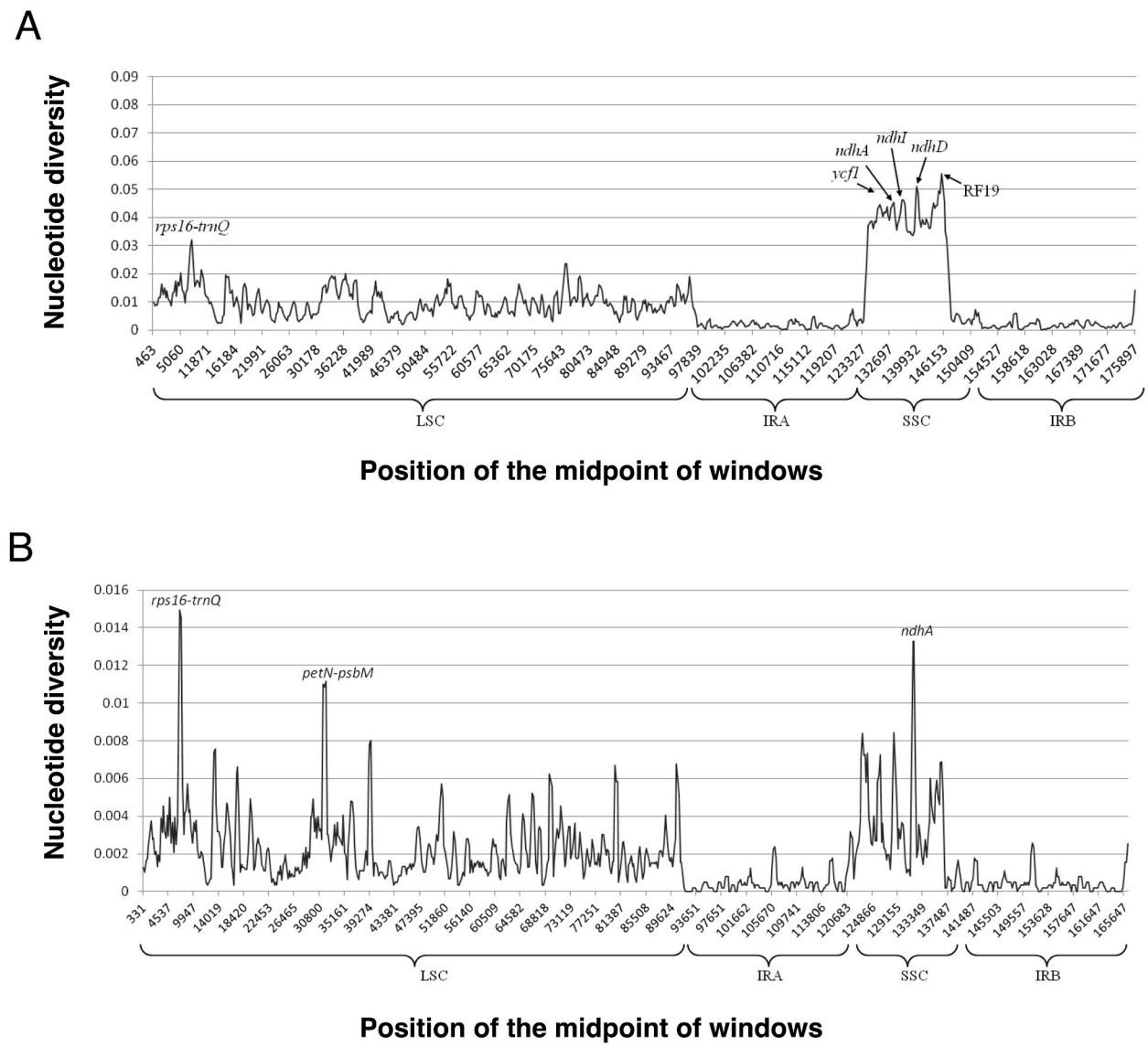


Fig. 6. Sliding window analysis. (A) Pi among thirty-three plant species in Zingiberaceae. (B) Pi among twenty *Curcuma* species. *X*-axis: position of the midpoint of windows, *Y*-axis: nucleotide diversity of each window (Pi).

encoding a specific amino acid were analyzed. Codons play a crucial role in genetic transmission, forming the fundamental units of proteins [69,70]. Codon usage, influenced by mutation biases, is a factor shaping the evolution of chloroplast genomes and varies among species [71]. Analysis revealed a low level of codon usage bias in the cp genomes of *C. comosa* and *C. latifolia*, suggesting their similar evolutionary paths. This result was consistent with those from phylogenetic construction of the twenty *Curcuma* species and other species in Zingiberales using cp genomes. The phylogenetic analysis demonstrated that all *Curcuma* species formed a complex monophyletic clade with robust bootstrap support. As expected, *C. comosa* and *C. latifolia* exhibited a very close relationship in our phylogenetic analysis. Repeat structures are influential in genomic recombination, rearrangement, and sequence divergence within chloroplast genomes [72]. However, significant variations in repeat distribution were not found in the *Curcuma* genus (Figs. 2 and 3). Although the organization of cp genomes remains highly similar across Zingiberaceae species, several regions with interspecific polymorphisms were discovered, mostly located within intergenic regions and at the LSC, SSC and IR regional boundaries of the cp genomes (Figs. 4 and 5).

One of the most important sources of sequence variability is the nucleotide mutation. When comparing nucleotide diversity across various regions of thirty-three cp genomes in Zingiberaceae sourced from the NCBI database, including *C. comosa* and *C. latifolia*, *ndh*A exhibited the highest level of nucleotide diversity. These findings are in line with a previous investigation where such hotspots were identified within the cp genome of other *Curcuma* species [33,66–68]. These markedly divergent sequences hold promise for the development of potential molecular markers. Furthermore, we explored other molecular markers such as SSRs, indels, and SNPs. Upon

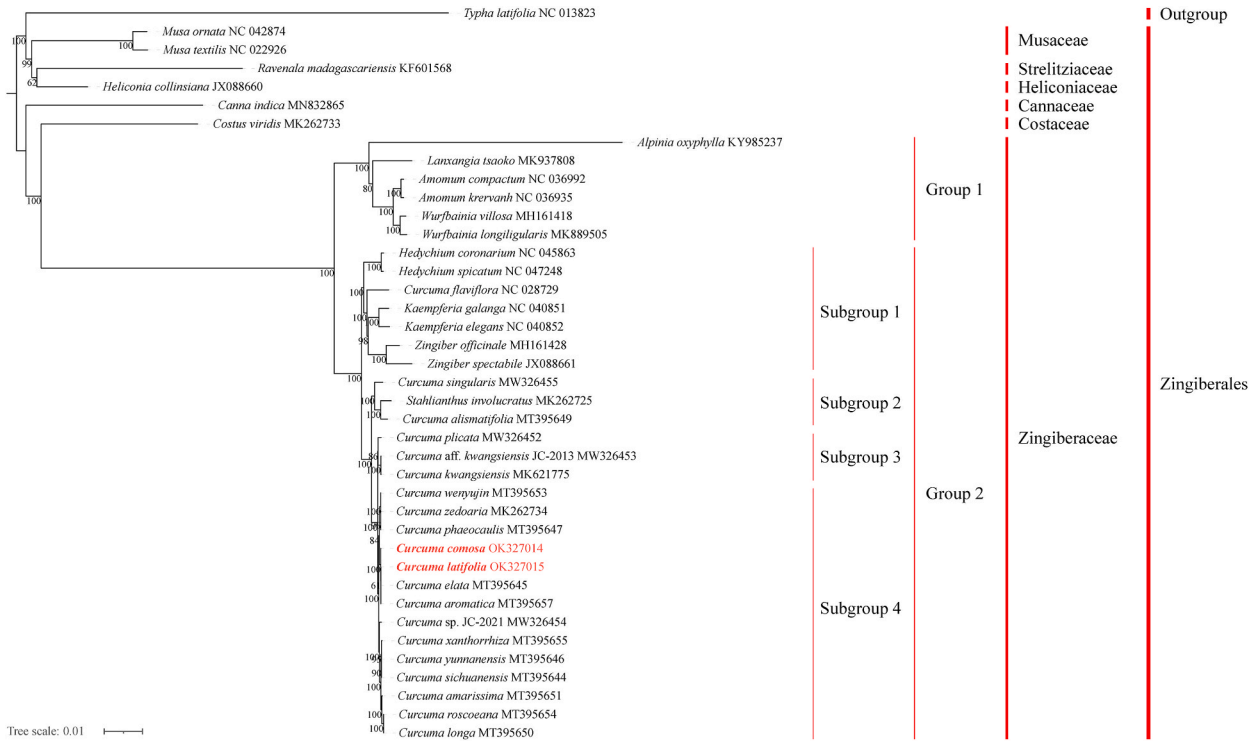


Fig. 7. Phylogenetic relationships of *C. comosa* and *C. latifolia* and other related species in Zingiberales based on maximum likelihood (ML) analysis of the forty cp genome sequences. The branch supports values are reported as Ultrafast Bootstrap (1000 replicates), the accession number of GenBank for each species is listed in the figure. *Typha latifolia* was used as an outgroup.

scrutinizing the chloroplast genomes, we identified indel/SNP variable sites at *trn*T-*trn*L, *ndh*C-*trn*V, and *ndh*A. There was a report recommended the use of the three regions for the identification of plants in the genera *Dendrobium* [73], *Actaea* [74], *Entandrophragma* [75], *Phyllanthus* [54], and *Polygonatum* [59]. For future research, the variable regions identified in the chloroplast genome should be developed as DNA markers for use in the authentication and identification of herbal materials or products.

5. Conclusions

As far as we know, this study marks the first acquisition of the cp genome sequences for *C. comosa* and *C. latifolia*. The constructed phylogenetic tree of *C. comosa*, *C. latifolia*, and related species enhances our comprehension of the relationships among plants in the genus *Curcuma* and the family Zingiberaceae. This resource can be reliably exploited for further research within the Zingiberaceae family. As no variable sites were observed in standard DNA barcode regions, therefore, variation regions within the chloroplast genomes were explored to overcome the limitations of standard DNA barcodes in distinguishing plant species. To improve the herbal pharmacovigilance, it is essential to identify and authenticate herbal medicines before their release into the markets. The information obtained from analyzing the cp genomes, are crucial and useful for developing molecular markers and resolving taxonomic discrepancies at the genus *Curcuma* and family Zingiberaceae levels.

Funding statement

Not applicable.

Data availability statement

Data generated during this study are included in this published article. The chloroplast genome sequences of *C. comosa* and *C. latifolia* were submitted to GenBank (Accession number: OK327014 for *C. comosa*; OK327015 for *C. latifolia*). Accession numbers for additional datasets utilized and analyzed in this research can be located in the Supplementary Tables 2, 3, 4, 5.

CRediT authorship contribution statement

Bussarin Wachananawat: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal

analysis. **Bobby Lim-Ho Kong:** Software, Methodology, Formal analysis. **Pang-Chui Shaw:** Software, Methodology. **Bhanubong Bongcheewin:** Resources. **Sunisa Sangvirotjanapat:** Resources. **Pinidphon Prombutara:** Software, Formal analysis. **Natapol Pornputtapong:** Software, Formal analysis. **Suchada Sukrong:** Writing – review & editing, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research project is supported by the Second Century Fund (C2F), Chulalongkorn University. The authors express gratitude to Mr. Chayapol Tungphatthong for improving the resolution of the figures in this manuscript. Additionally, the authors extend thanks to the Faculty of Pharmaceutical Sciences, Chulalongkorn University, Bangkok, Thailand, for generously providing the necessary equipment.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e31248.

References

[1] J. Chen, N. Xia, J. Zhao, J. Chen, R.J. Henny, Chromosome numbers and ploidy levels of Chinese *Curcuma* species, Hortic. Sci. 48 (5) (2013) 525, https://doi.org/10.21273/HORTSCI.48.5.525, 230.
[2] Y. Qu, F. Xu, S. Nakamura, H. Matsuda, Y. Pongpiriyadacha, L. Wu, M. Yoshikawa, Sesquiterpenes from *Curcuma comosa*, J. Nat. Med. 63 (2009) 102–104, https://doi.org/10.1007/s11418-008-0282-8.
[3] B. Sasikumar, Genetic resources of Curcuma: diversity, characterization and utilization, Plant Genet. Resour. 3 (2) (2005) 230–251, https://doi.org/10.1079/PGR200574.
[4] P. Piyachaturawat, S. Ercharuporn, A. Suksamrarn, Uterotrophic effect of *Curcuma comosa* in rats, Int. J. Pharmacogn. 33 (4) (1995) 334–338, https://doi.org/10.3109/13880209509065388.
[5] P. Piyachaturawat, S. Ercharuporn, A. Suksamrarn, Estrogenic activity of *Curcuma comosa* extract in rats, Asia Pacific, J. Pharmacol. 10 (3–4) (1995) 121–126.
[6] J. Weerachayaphorn, A. Chuncharunee, C. Mahagita, B. Lewchalermwongse, A. Suksamrarn, P. Piyachaturawat, A protective effect of *Curcuma comosa* Roxb. on bone loss in estrogen deficient mice, J. Ethnopharmacol. 137 (2) (2011) 956–962, https://doi.org/10.1016/j.jep.2011.06.040.
[7] The SEA-LAC Trade Center, Herbal Products of Thailand, the University of the Thai Chamber of Commerce (UTCC), Bangkok, Thailand, 2012.
[8] Trade MAP. https://www.trademap.org/Country_SelProductCountry. (Accessed 27 January 2024).
[9] P. Soontornchainaksaeng, T. Jenjittikul, Chromosome number variation of phytoestrogen-producing *Curcuma* (Zingiberaceae) from Thailand, J. Nat. Med. 64 (3) (2010) 370–377. https://doi:10.1007/s11418-010-0414-9.
[10] Waan Chak Mod Luk. https://www.pharmacy.mahidol.ac.th/th/knowledge/article/92/ว่านชักมดลูก/. Accessed 30 April 2020.
[11] V. Keeratinijakal, M. Kladmook, K. Laosatit, Identification and characterization of *Curcuma comosa* Roxb., phytoestrogens-producing plant, using AFLP markers and morphological characteristics, J. Med. Plants Res. 4 (24) (2010) 2651–2657. https://doi:10.5897/JMPR09.381.
[12] V. Keeratinijakal, S. Kongkiatpaiboon, Distribution of phytoestrogenic diarylheptanoids and sesquiterpenoids components in *Curcuma comosa* rhizomes and its related species, Rev. Bras. Farmacogn. 27 (3) (2017) 290–296, https://doi.org/10.1016/j.bjp.2016.12.003.
[13] S. Saensouk, T. Boonma, P. Saensouk, Six new species and a new record of *Curcuma* L. (Zingiberaceae) from Thailand, Biodiversitas 22 (4) (2021) 1658–1685. https://doi:10.13057/biodiv/d220410.
[14] A. Suksamrarn, M. Ponglikitmongkol, K. Wongkrajang, A. Chindaduang, S. Kittidanairak, A. Jankam, et al., Diarylheptanoids, new phytoestrogens from the rhizomes of *Curcuma comosa*: isolation, chemical modification and estrogenic activity evaluation, Bioorgan, Med. Chem. 16 (14) (2008) 6891–6902, https://doi.org/10.1016/j.bmc.2008.05.051.
[15] W. Winuthayanon, P. Piyachaturawat, A. Suksamrarn, M. Ponglikitmongkol, Y. Arao, S.C. Hewitt, et al., Diarylheptanoid phytoestrogens isolated from the medicinal plant *Curcuma comosa*: biological actions in vitro and in vivo indicate estrogen receptor-dependent mechanisms, Environ. Health Perspect. 117 (7) (2009) 1155–1161, https://doi.org/10.1289/ehp.0900613.
[16] W. Winuthayanon, K. Suksen, C. Boonchird, A. Chuncharunee, M. Ponglikitmongkol, A. Suksamrarn, et al., Estrogenic activity of diarylheptanoids from *Curcuma comosa* Roxb. requires metabolic activation, J. Agric. Food Chem. 57 (3) (2009) 840–845, https://doi.org/10.1021/jf802702c.
[17] S. Burapan, M. Kim, Y. Paisooksantivatana, B.E. Eser, J. Han, Thai *Curcuma* species: antioxidant and bioactive compounds, Foods 9 (9) (2020) 1219, https://doi.org/10.3390/foods9091219.
[18] P. Pimkaew, A. Chuncharunee, A. Suksamsarn, P. Piyajuturawat, Evaluation on toxicology of *Curcuma latifolia* rosc, Thai J. Toxicol. 23 (2008) 193–196.
[19] J. Maknoi, Taxonomy and Phylogeny of the Genus *Curcuma* L. (Zingiberaceae) with Particular Reference to its Occurrence in Thailand, Prince of Songkla University, 2006. PhD thesis.
[20] P. Dechbumroong, S. Aumnouypol, J. Denduangboripant, S. Sukrong, DNA barcoding of *Aristolochia* plants and development of species-specific multiplex PCR to aid HPTLC in ascertainment of *Aristolochia* herbal materials, PLoS One 13 (2018) e0202625, https://doi.org/10.1371/journal.pone.0202625.
[21] R. Joseph, T. Joseph, J. Joseph, Karyomorphological studies in genus *Curcuma linn*, Cytologia 64 (3) (1999) 313–317, https://doi.org/10.1508/cytologia.64.313.
[22] Y. Paisooksativatana, O. Thepsen, Phenetic relationships of some Thai *Curcuma* species (Zingiberaceae) based on morphological, palynological and cytological evidence, Thai, J. Agric. Sci. 34 (1–2) (2001) 47–57.
[23] T. Boonsrangsom, Genetic diversity of 'Wan Chak Motluk' (*Curcuma comosa* Roxb.) in Thailand using morphological characteristics and random amplification of polymorphic DNA (RAPD) markers, South Afr. J. Bot. 130 (2020) 224–230, https://doi.org/10.1016/j.sajb.2020.01.005.
[24] I. Siriluck, T. Ratchanok, H. Worakij, K. Thitamin, Identification of 24 species of Zingiberaceae in Thailand using ISSR technique, Thai J. Agric. Sci. 47 (1) (2014) 1–6.

[25] A. Sihanat, O. Theanphong, K. Rungsihirunrat, Assessment of phylogenetic relationship among twenty *Curcuma* species in Thailand using amplified fragment length polymorphism marker, \"J. Adv. Pharm. Technol. Research\"\" (JAPTR)\" 11 (3) (2020) 134–141.

[26] W.J. Kress, L.M. Prince, K.J. Williams, The phylogeny and a new classification of the gingers (Zingiberaceae), evidence from molecular data, Am. J. Bot. 89 (2002) 1682–1696, https://doi.org/10.3732/ajb.89.10.1682.

[27] J. Chen, J. Zhao, D.L. Erickson, N. Xia, W.J. Kress, Testing DNA barcodes in closely related species of *Curcuma* (Zingiberaceae) from Myanmar and China, Mol. Ecol. Resour 15 (2) (2015) 337–348, https://doi.org/10.1111/1755-0998.12319.

[28] M. Minami, K. Nishio, Y. Ajioka, H. Kyushima, K. Shigeki, K. Kinjo, et al., Identification of *Curcuma* plants and curcumin content level by DNA polymorphisms in the *trnS-trnf*M intergenic spacer in chloroplast DNA, J. Nat. Med. 63 (2009) 75–79, https://doi.org/10.1007/s11418-008-0283-7.

[29] M.R. Vinitha, U.S. Kumar, K. Aishwarya, M. Sabu, G. Thomas, Prospects for discriminating Zingiberaceae species in India using DNA barcodes, J. Integr. Plant Biol. 56 (8) (2014) 760–773, https://doi.org/10.1111/jipb.12189.

[30] W.J. Kress, K.J. Wurdack, E.A. Zimmer, L.A. Weigt, D.H. Janzen, Use of DNA barcodes to identify flowering plants, Proc. Natl. Acad. Sci. USA 102 (23) (2005) 8369–8374, https://doi.org/10.1073/pnas.0503123102.

[31] Z. Hu, Study on DNA Barcoding and Chloroplast Genome of Medicinal Plants in Compositae, Hubei University of Chinese Medicine, 2012. PhD thesis.

[32] H. Liang, J. Chen, Comparison and phylogenetic analyses of nine complete chloroplast genomes of Zingibereae, Forests 12 (6) (2021) 710, https://doi.org/10.3390/f12060710.

[33] H. Liang, Y. Zhang, J. Deng, G. Gao, C. Ding, L. Zhang, et al., The complete chloroplast genome sequences of 14 *Curcuma* species: insights into genome evolution and phylogenetic relationships within Zingiberales, Front. Genet. 11 (2020) 802, https://doi.org/10.3389/fgene.2020.00802.

[34] K. Katoh, D.M. Standley, MAFFT Multiple sequence alignment software version 7: improvements in performance and usability, Mol. Biol. Evol. 30 (4) (2013) 772–780, https://doi.org/10.1093/molbev/mst010.

[35] S. Chen, Y. Zhou, Y. Chen, J. Gu, Fastp: an ultra-fast all-in-one FASTQ preprocessor, J Bioinform 34 (17) (2018) 884–890, https://doi.org/10.1093/bioinformatics/bty560.

[36] J.J. Jin, W.B. Yu, J.B. Yang, Y. Song, C.W. dePamphilis, T.S. Yi, et al., GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes, Genome Biol. 21 (2020) 241, https://doi.org/10.1186/s13059-020-02154-5.

[37] M. Tillich, P. Lehwark, T. Pellizzer, E.S. Ulbricht-Jones, A. Fischer, R. Bock, et al., GeSeq – versatile and accurate annotation of organelle genomes, Nucleic Acids Res. 45 (1) (2017) 6–11, https://doi.org/10.1093/nar/gkx391.

[38] S. Greiner, P. Lehwark, R. Bock, OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes, Nucleic Acids Res. 47 (1) (2019) 59–64, https://doi.org/10.1093/nar/gkz238.

[39] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, Mega X: molecular evolutionary genetics analysis across computing platforms, Mol. Biol. Evol. 35 (6) (2018) 1547–1549.

[40] X.M. Wu, S.F. Wu, D.M. Ren, Y.P. Zhu, F.C. He, The analysis method and progress in the study of codon bias, Hereditas 29 (4) (2007) 420–426, https://doi.org/10.1360/yc-007-0420.

[41] S. Beier, T. Thiel, T. Münch, U. Scholz, M. Mascher, MISA-web: a web server for microsatellite prediction, Bioinformation 33 (16) (2017) 2583–2585, https://doi.org/10.1093/bioinformatics/btx198.

[42] S. Kurtz, J.V. Choudhuri, E. Ohlebusch, C. Schleiermacher, J. Stoye, R. Giegerich, REPuter: the manifold applications of repeat analysis on a genomic scale, Nucleic Acids Res. 29 (22) (2001) 4633–4642, https://doi.org/10.1093/nar/29.22.4633.

[43] K.A. Frazer, L. Pachter, A. Poliakov, E.M. Rubin, I. Dubchak, VISTA: computational tools for comparative genomics, Nucleic Acids Res. 32 (2) (2004) 273–279, https://doi.org/10.1093/nar/gkh458.

[44] J. Rozas, A. Ferrer-Mata, J.C. Sánchez-DelBarrio, S. Guirao-Rico, P. Librado, S.E. Ramos-Onsins, et al., DnaSP 6: DNA sequence polymorphism analysis of large data sets, Mol. Biol. Evol. 34 (12) (2017) 3299–3302, https://doi.org/10.1093/molbev/msx248.

[45] P.A. Cock, T. Antao, J.T. Chang, B.A. Chapman, C.J. Cox, A. Dalke, et al., Biopython: freely available Python tools for computational molecular biology and bioinformatics, Bioinformation 25 (11) (2009) 1422–1423, https://doi.org/10.1093/bioinformatics/btp163.

[46] J. Sukumaran, M.T. Holder, DendroPy: a Python library for phylogenetic computing, Bioinformation 26 (12) (2010) 1569–1571, https://doi.org/10.1093/bioinformatics/btq228.

[47] L.T. Nguyen, H.A. Schmidt, A.V. Haeseler, B.Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies, Mol. Biol. Evol. 32 (1) (2015) 268–274, https://doi.org/10.1093/molbev/msu300.

[48] S. Kalyaanamoorthy, B.Q. Minh, T.K.F. Wong, A.V. Haeseler, L.S. Jermiin, ModelFinder: fast model selection for accurate phylogenetic estimates, Nat. Methods 14 (2017) 587–589, https://doi.org/10.1038/nmeth.4285.

[49] D.T. Hoang, O. Chernomor, A.V. Haeseler, B.Q. Minh, L.S. Vinh, UFBoot2: improving the Ultrafast bootstrap approximation, Mol. Biol. Evol. 35 (2) (2018) 518–522, https://doi.org/10.1093/molbev/msx281.

[50] S. Guindon, J.F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, O. Gascuel, New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0, Syst. Biol. 59 (3) (2010) 307–321, https://doi.org/10.1093/sysbio/syq010.

[51] A.J. Fazekas, K.S. Burgess, P.R. Kesanakurti, S.W. Graham, S.G. Newmaster, B.C. Husband, et al., Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well, PLoS One 3 (2008) e2802, https://doi.org/10.1371/journal.pone.0002802.

[52] J.B. Yang, M. Tang, H.T. Li, Z.R. Zhang, D.Z. Li, Complete chloroplast genome of the genus *Cymbidium*: lights into the species identification, phylogenetic implications and population genetic analyses, BMC Evol. Biol. 13 (2013) 84, https://doi.org/10.1186/1471-2148-13-84.

[53] X. Chen, J. Zhou, Y. Cui, Y. Wang, B. Duan, H. Yao, Identification of *Ligularia* herbs using the complete chloroplast genome as a super-barcode, Front. Pharmacol. 9 (2018) 695, https://doi.org/10.3389/fphar.2018.00695.

[54] H. Fang, G. Dai, B. Liao, P. Zhou, Y. Liu, Application of chloroplast genome in the identification of *Phyllanthus urinaria* and its common adulterants, Front. Plant Sci. 13 (2023) 1099856. https://doi:10.3389/fpls.2022.1099856.

[55] M.H.Y. Yik, B.L.H. Kong, T.Y. Siu, D.T.W. Lau, H. Cao, P.C. Shaw, Differentiation of *Hedyotis diffusa* and common adulterants based on chloroplast genome sequencing and DNA barcoding markers, Plants 10 (1) (2021) 161, https://doi.org/10.3390/plants10010161.

[56] T. Zhou, J. Wang, Y. Jia, W. Li, F. Xu, X. Wang, Comparative chloroplast genome analyses of species in *Gentiana* section *Cruciata* (Gentianaceae) and the development of authentication markers, Int. J. Mol. Sci. 19 (7) (2018) 1962, https://doi.org/10.3390/ijms19071962.

[57] L. Wu, M. Wu, N. Cui, L. Xiang, Y. Li, X. Li, et al., Plant super-barcode: a case study on genome-based identification for closely related species of *Fritillaria*, Chin. Med. 16 (2021) 52, https://doi.org/10.1186/s13020-021-00460-z.

[58] X. Yang, X. Yu, X. Zhang, H. Guo, Z. Xing, L. Xu, J. Wang, Y. Shen, J. Yu, P. Lv, Y. Wang, M. Liu, X. Tian, Development of mini-barcode based on chloroplast genome and its application in metabarcoding molecular identification of Chinese medicinal material radix *Paeoniae Rubra* (Chishao), Front. Plant Sci. 13 (2022) 819822. https://doi:10.3389/fpls.2022.819822.

[59] Z.N. Xie, J. Lao, H. Liu, W.X. Zhang, W. He, C. Zhong, J. Xie, S.H. Zhang, J. Jin, Characterization of the chloroplast genome of medicinal herb *Polygonatum cyrtonema* and identification of molecular markers by comparative analysis, Genome 66 (4) (2023) 80–90, https://doi.org/10.1139/gen-2022-0087.

[60] Y. Cui, X. Chen, L. Nie, W. Sun, H. Hu, Y. Lin, et al., Comparison and phylogenetic analysis of chloroplast genomes of three medicinal and edible *Amomum* species, Int. J. Mol. Sci. 20 (16) (2019) 4040, https://doi.org/10.3390/ijms20164040.

[61] D.M. Li, C.Y. Zhao, X.F. Liu, Complete chloroplast genome sequences of *Kaempferia galanga* and *Kaempferia elegans*: molecular structures and comparative analysis, Molecules 24 (3) (2019) 474, https://doi.org/10.3390/molecules24030474.

[62] L. Gui, S. Jiang, D. Xie, L. Yu, Y. Huang, Z. Zhang, et al., Analysis of complete chloroplast genomes of *Curcuma* and the contribution to phylogeny and adaptive evolution, Gene 723 (2020) 144355, https://doi.org/10.1016/j.gene.2020.144355.

[63] Q. Yang, G.F. Fu, Z.Q. Wu, L. Li, J.L. Zhao, Q.J. Li, Chloroplast genome evolution in four montane Zingiberaceae Taxa in China, Front. Plant Sci. 12 (2022) 774482. https://doi:10.3389/fpls.2021.774482.

[64] Y.J.M. Liew, K.O. Chua, H.S. Yong, S.L. Song, K.G. Chan, Complete chloroplast genome of *Boesenbergia rotunda* and a comparative analysis with members of the family Zingiberaceae, Braz. J. Bot. 45 (4) (2022) 1209–1222, https://doi.org/10.1007/s40415-022-00845-w.
[65] D. Jiang, X. Cai, M. Gong, M. Xia, H. Xing, S. Dong, S. Tian, J. Li, J. Lin, Y. Liu, H.L. Li, Complete chloroplast genomes provide insights into evolution and phylogeny of *Zingiber* (Zingiberaceae), BMC Genom. 24 (2023) 30, https://doi.org/10.1186/s12864-023-09115-9.
[66] D.M. Li, G.F. Zhu, Y.C. Xu, Y.J. Ye, J.M. Liu, Characterization and phylogenetic analysis of the complete chloroplast genome of *Curcuma zedoaria* (Zingiberaceae), Mitochondrial DNA Part B 5 (2) (2020) 1329–1331, https://doi.org/10.1080/23802359.2020.1734496.
[67] M.K. Kim, W.K. Lee, Y.R. Choi, J. Kim, I. Kang, J. Kang, The complete chloroplast genome sequence of three medicinal species; *Curcuma longa, Curcuma wenyujin*, and *Curcuma phaeocaulis* (Zingiberaceae), Mitochondrial DNA Part B 6 (4) (2021) 1363–1364, https://doi.org/10.1080/23802359.2020.1768917.
[68] J. Wang, X. Liao, Y. Li, Y. Ye, G. Xing, S. Kan, L. Nie, S. Li, L.R. Tembrock, Z. Wu, Comparative plastomes of *Curcuma alismatifolia* (Zingiberaceae) reveal diversified patterns among 56 different cut-flower cultivars, Genes 14 (9) (2023) 1743, https://doi.org/10.3390/genes14091743.
[69] W.H. Campbell, G. Gowri, Codon usage in higher plants, green algae, and cyanobacteria, Plant Physiol. 92 (1) (1990) 1–11, https://doi.org/10.1104/pp.92.1.1.
[70] Q. Liu, S. Dou, Z. Ji, Q. Xue, Synonymous codon usage and gene function are strongly related in *Oryza sativa*, Biosystems 80 (2) (2005) 123–131, https://doi.org/10.1016/j.biosystems.2004.10.008.
[71] B. Li, F. Lin, P. Huang, W. Guo, Y. Zheng, Complete chloroplast genome sequence of *Decaisnea insignis*: genome organization, genomic resources and comparative analysis, Sci. Rep. 7 (2017) 10073, https://doi.org/10.1038/s41598-017-10409-8.
[72] D. Srivastava, A. Shanker, Identification of simple sequence repeats in chloroplast genomes of Magnoliids through bioinformatics approach, Interdiscipl. Sci. Comput. Life Sci. 8 (2016) 327–336, https://doi.org/10.1007/s12539-015-0129-4.
[73] N. Zhitao, Z. Shuying, P. Jiajia, L. Ludan, S. Jing, D. Xiaoyu, Comparative analysis of *Dendrobium* plastomes and utility of plastomic mutational hotspots, Sci. Rep. 7 (2017) 2073, https://doi.org/10.1038/s41598-017-02252-8.
[74] I. Park, J.H. Song, S. Yang, B.C. Moon, Comparative analysis of *Actaea* chloroplast genomes and molecular marker development for the identification of authentic *Cimicifugae* rhizome, Plants 9 (2) (2020) 157, https://doi.org/10.3390/plants9020157.
[75] M. Mascarello, M. Amalfi, P. Asselman, E. Smets, O.J. Hardy, H. Beeckman, et al., Genome skimming reveals novel plastid markers for the molecular identification of illegally logged *African timber* species, PLoS One 16 (2021) e0251655, https://doi.org/10.1371/journal.pone.0251655.