



OPEN

Complete chloroplast genome of novel *Adinandra megaphylla* Hu species: molecular structure, comparative and phylogenetic analysis

Huu Quan Nguyen¹, Thi Ngoc Lan Nguyen¹ , Thi Nhung Doan², Thi Thu Nga Nguyen¹, Mai Huong Phạm², Tung Lam Le², Danh Thuong Sy¹, Hoang Ha Chu² & Hoang Mau Chu¹ 

Adinandra megaphylla Hu is a medicinal plant belonging to the *Adinandra* genus, which is well-known for its potential health benefits due to its bioactive compounds. This study aimed to assemble and annotate the chloroplast genome of *A. megaphylla* as well as compare it with previously published cp genomes within the *Adinandra* genus. The chloroplast genome was reconstructed using de novo and reference-based assembly of paired-end reads generated by long-read sequencing of total genomic DNA. The size of the chloroplast genome was 156,298 bp, comprised a large single-copy (LSC) region of 85,688 bp, a small single-copy (SSC) region of 18,424 bp, and a pair of inverted repeats (IRa and IRb) of 26,093 bp each; and a total of 51 SSRs and 48 repeat structures were detected. The chloroplast genome includes a total of 131 functional genes, containing 86 protein-coding genes, 37 transfer RNA genes, and 8 ribosomal RNA genes. The *A. megaphylla* chloroplast genome indicated that gene content and structure are highly conserved. The phylogenetic reconstruction using complete cp sequences, *matK* and *trnL* genes from Pentaphragaceae species exhibited a genetic relationship. Among them, *matK* sequence is a better candidate for phylogenetic resolution. This study is the first report for the chloroplast genome of the *A. megaphylla*.

The chloroplast (cp) acts as a vital and essential organelle playing an indispensable role in several crucial biochemical processes and photosynthesis of plants¹. The cp genome is uniparental inheritance and generally has a quadripartite structure including one large single-copy (LSC) region, one small single-copy (SSC) region, and two inverted repeat regions (IRs) of the same length². In terms of gene structure and composition, the cp genome is more conserved, compared with nuclear and mitochondrial genomes³. These chloroplast DNA features were used by scientists to construct chloroplast DNA phylogenies, demonstrating to be greatly beneficial in the exploration of plant phylogenetic studies and more clarified taxonomic levels^{4,5}. The whole chloroplast genome was reported as circular and its genes were a single evolutionary unit^{6,7}. The chloroplast DNA genome sequencing in complex genome plants has been proven to be comparatively inexpensive and easy^{8,9}.

Adinandra genus has been consumed as a traditional health tea beverage in China, and it has been described to have many curative effects, such as reduction of blood pressure, anti-inflammatory, antibacterial, antitumor, antitoxic, and analgesic effects^{10,11}. Surprisingly, to date, limited reports have been published about *Adinandra* genus. *Adinandra megaphylla* Hu, belonging to the *Adinandra* genus, was first published in Bull. Fan Mem. Inst. Biol., Bot. 6: 172 (1935). The related information is about morphological descriptions¹² and bioactivity assays¹³. A few species were used by molecular phylogenetics, including *A. dumosa*¹⁴; *A. elegans*, *A. formosan*, *A. lasiostyla*, *A. millettii*, *A. yaeyamensi*¹²; *A. millettii*, *A. angustifolia*¹⁵. To date, there have not been any studies on the genomes of *A. megaphylla* Hu, especially the chloroplast genome, this leads to the lack of information for estimating the phylogenetic relationships within *Adinandra* genus.

¹Thainguyen University of Education, Thai Nguyen University, Thai Nguyen 250000, Vietnam. ²Institute of Biotechnology, Vietnam Academy of Science and Technology, Hanoi 100000, Vietnam. ✉email: ntnlan.dhsptn@tnu.edu.vn; chuhoangmau@tnu.edu.vn

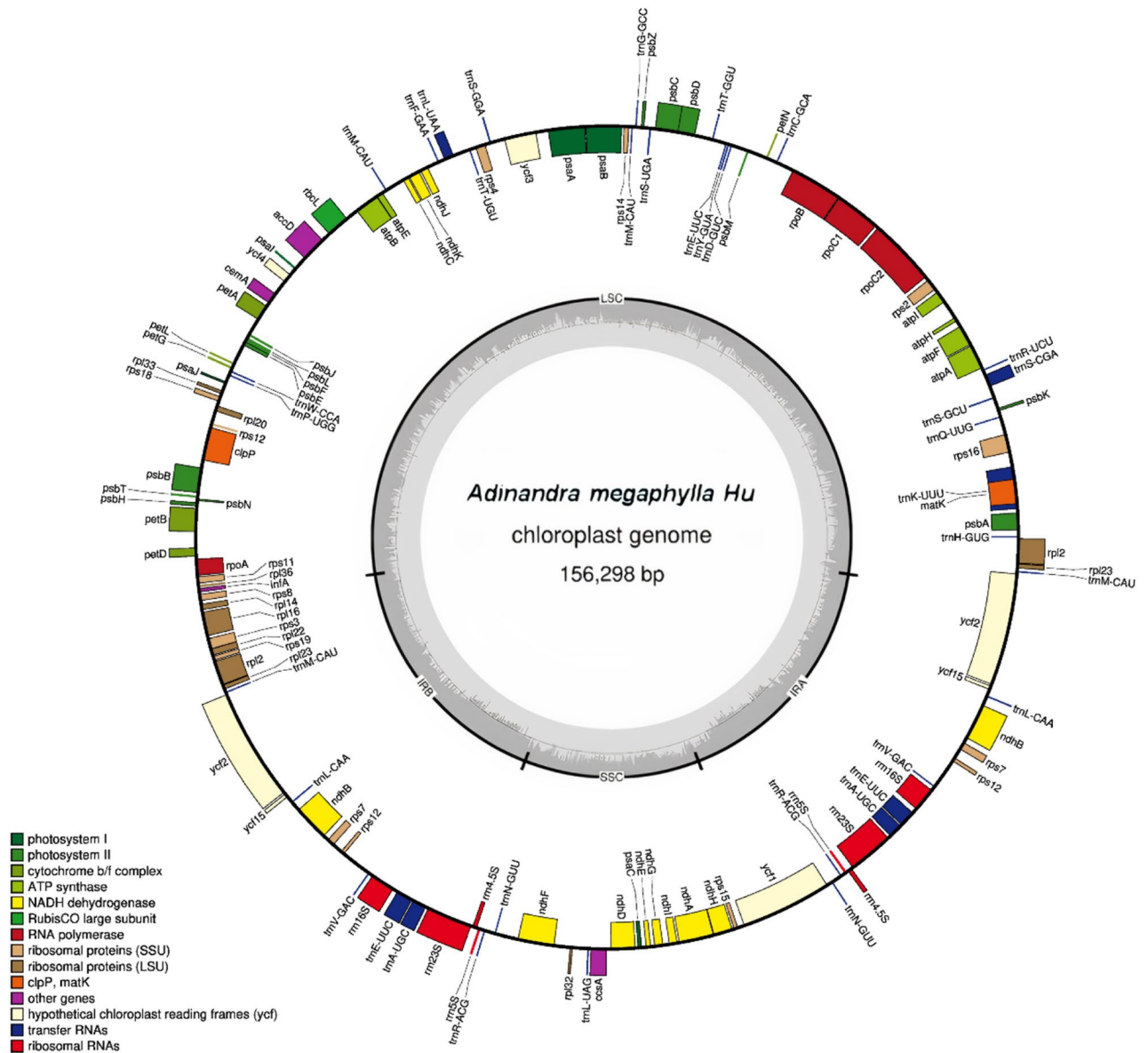


Figure 1. Chloroplast map of *A. megaphylla* Hu in Vietnam. Genes shown inside the circle are transcribed clockwise, whereas genes outside are transcribed counterclockwise. The light gray inner circle shows the AT content, the dark gray corresponds to the GC content.

For the first time, we reported a new complete chloroplast genome of *A. megaphylla* Hu, and combined it with previously published Pentaphragaceae complete chloroplast genomes data to visualize and evaluate the genome organization and phylogenetic relationships.

Results

Chloroplast genome assembly and annotation. Using the PacBio SEQUEL system, 29,815,452 bp of raw sequence data of the whole genome were generated from *A. megaphylla* Hu (Fig. 1). The mean read length is 2938 bp, the N50 contig size is 3594 bp and approximately 5% of the genomic genome belongs to the cp genome with 188× coverage. The cp genome size of 156,298 bp of *A. megaphylla* Hu was derived from the assembly. As shown in most cp genomes, the assembled *A. megaphylla* Hu plastome exhibited the typical quadripartite structure comprising of the four regions, a pair of inverted repeats (IRs 26,093 bp), LSC (85,688 bp), and SSC (18,424 bp). Besides, the cp genome of *A. megaphylla* Hu contains 131 genes, and the percent of the GC content of the cp genome was 37.4% (Table 1).

Chloroplast genome annotation. The cp genome of *A. megaphylla* Hu includes 37 tRNA genes, 8 rRNA genes (16S, 23S, 5S, and 4.5S), and 86 protein-coding genes (Table 1). There were 123 genes assigned into three groups based on their functions. Regarding the photosynthesis-related gene category, there are 43 genes, containing genes encoding the large subunit of RubisCO related to the photosynthetic electron transport chain and

| Genome features | <i>A. megaphylla</i> Hu |
|------------------|-------------------------|
| Genome size (bp) | 156,298 |
| LSC size (bp) | 85,688 |
| SSC size (bp) | 18,424 |
| IR size (bp) | 26,093 |
| GC content (%) | 37.4 |
| No. of genes | 131 |
| No. of PCGs | 86 |
| No. of tRNA | 37 |
| No. of rRNA | 8 |

Table 1. Summary of the chloroplast genome of *A. megaphylla* Hu species.

| Category of genes | Group of genes | Name of genes |
|-------------------------------|--|--|
| Photosynthesis | Subunits of ATP synthase | <i>atpA, atpB, atpE, atpF^a, atpH, atpI</i> |
| | Subunits of NADH-dehydrogenase | <i>ndhA^a, ndhB (×2)^a, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i> |
| | Subunits of cytochrome b/f complex | <i>petL, petB^a, petG, petA, petD, petN</i> |
| | Subunits of photosystem I | <i>psaJ, psaC, psaA, psaI, psaB</i> |
| | Subunits of photosystem II | <i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i> |
| | Subunit of rubisco | <i>rbcL</i> |
| Transcription and translation | Large subunit of ribosome | <i>rpl14, rpl16^a, rpl2 (×2)^a, rpl20, rpl22, rpl23 (×2), rpl32, rpl33, rpl36</i> |
| | DNA dependent RNA polymerase | <i>rpoB, rpoA, rpoC1^a, rpoC2</i> |
| | Small subunit of ribosomal proteins | <i>rps11, rps12, rps14, rps15, rps16^a, rps18, rps19 (×2), rps2, rps3, rps4, rps7 (×2), rps8</i> |
| | rRNA genes | <i>rrn23S (×2), rrn16S (×2), rrn5S (×2), rrn4.5S (×2)</i> |
| | tRNA genes | <i>trnA-UGC (×2)^a, trnC-ACA^a, trnC-GCA, trnD-GUC, trnE-UUC (×2)^a, trnF-GAA, trnG-GCC, trnH-GUG, trnK-UUU^a, trnL-CAA (×2), trnL-UAA^a, trnL-UAG, trnM-CAU (×2), trnN-GUU (×2), trnP-UGG, trnQ-UUG, trnR-ACG (×2), trnR-UCU, trnS-CGA^a, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC (×2), trnW-CCA, trnY-GUA</i> |
| | Translational initiation factor | <i>infA</i> |
| Other genes | Subunit of acetyl-CoA-carboxylase (fatty acid synthesis) | <i>accD</i> |
| | c-type cytochrome synthesis gene | <i>ccsA</i> |
| | Envelope membrane protein (carbon metabolism) | <i>cemA</i> |
| | Protease | <i>clpP^b</i> |
| | Maturase (RNA processing) | <i>matK</i> |
| | Conserved open reading frames | <i>ycf1, ycf2 (×2), ycf3^b, ycf4, ycf15 (×2)</i> |

Table 2. Gene composition of *A. megaphylla* Hu chloroplast genome. Genes marked with the sign are the gene with a single (a) or double (b) introns and duplicated genes (×2).

putative NADPH dehydrogenase genes. Also, 68 genes were functioning in the transcription and translation processes. The majorities are tRNA genes, and the others are rRNA genes and genes encoding subunits of RNA polymerase and ribosome proteins. The remaining twelve genes with different functions are classified in the category of other genes, including five genes with known functions in fatty acid synthesis (*accD*), c-type cytochrome synthesis (*ccsA*), carbon metabolism (*cemA*), (*clpP*) proteolysis, and RNA processing (*matK*). Otherwise, five genes encoding for the conserved reading frames (*ycf1, ycf2, ycf3, ycf4, ycf15*) with unknown functions were annotated in the plastome. Eighteen genes (all 4 rRNA genes, 7 tRNA genes, 4 ribosomal protein-coding genes, 1 NADH-dehydrogenase protein-coding gene and 2 other genes) were annotated with two copies located in IR regions (Table 2). There are 18 cp genes harbored introns, among which 13 genes (*atpF, rpoC1, rpl2 (×2), ndhB (×2), ndhA, petB, rpl16, trnA-UGC (×2), trnC-ACA, trnE-UUC, trnK-UUU, trnL-UAA* and *trnS-CGA*) contained one intron, while 2 genes (*ycf3, clpP*) contained 2 introns (Table 2).

Repeat sequences and codon analysis. The total number of identified simple sequence repeats (SSRs) in the chloroplast genome of *A. megaphylla* Hu was 40. All repeats were mono repeats composed of A or T (size of 10–19) (Fig. 2A). There were no di-, tri-, tetra-, penta-, and hexa-nucleotide SSRs in the *A. megaphylla* Hu (Fig. 2B).

The cp genome of *A. megaphylla* Hu was identified with 49 repeats consisting of 26 palindromic repeats, 19 forward and 4 reverse repeats. There were no complement repeats (Fig. 3). The smallest unit size of the repeat was 22 bp while the largest unit size was 62 bp. Most of the size of the repeats (72%) was higher than 30 bp.

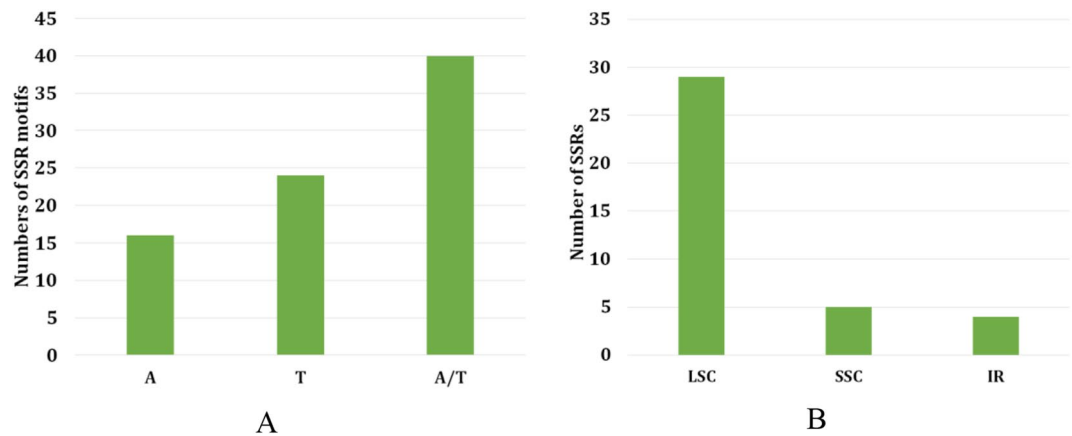


Figure 2. Analysis of single sequence repeats of plastome in *A. megaphylla* Hu. (A) Number of identified SSR sequence motifs; (B) Frequency of repeat types in LSC, SSC, and IR regions.

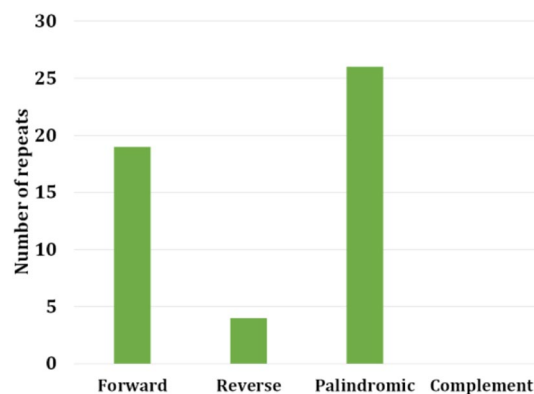


Figure 3. Repeat analysis on a genomic scale in *A. megaphylla* Hu.

The codon usage frequency of 64 protein-coding genes for three *Adinandra* species was evaluated. The total number of codons for protein-coding genes was 52,076 in those coding regions. G- and C-ending are found to be more frequent than their counterparts A and U (Table 3). Among the 20 amino acids, serine was the most abundant (number of codons encoding serine = 4975, 9.55%), leucine ranked second (number of codons encoding leucine = 4883, 9.37%), while the rarest one is tryptophan (677 codons, approximately 1.3%). Thirty codons were observed to be used more frequently than the expected usage at equilibrium ($RSCU > 1$) and thirty-one codons showed the codon usage bias: ($RSCU < 1$). Moreover, the frequency of use for the start codons AUG and UGG (methionine and tryptophan), as well as AUA (isoleucine) showed no bias ($RSCU = 1$).

Comparative chloroplast genomic analysis. To characterize genome divergence, the annotation of *A. megaphylla* Hu was taken as references. The comparison revealed that three chloroplast genomes were highly similar (Fig. 4). The plastome sequences were fairly conserved across the three data with a few regions with a variation. The results exhibited the divergence in LSC and SSC regions were higher than in IR regions. Besides, the sequences in the coding regions tended to be more conserved whereas most of the variations detected were found in conserved non-coding sequences (CNS). The sequences of exons were nearly identical throughout the three taxa. Among the coding genes, the highly disparate regions included *matK*, *rpoC2*, *ndhK*, *ndhD*, *ycf1*.

The sliding window analysis showed that the average pi value of the LSC ($Pi = 0.001569$) and SSC ($Pi = 0.001339$) regions was much higher than that in the IR ($Pi = 0.000219$) regions, which showed that LSC and SSC regions contained the most of the variation (Fig. 5). Among the 3 *Adinandra* species, the average value of nucleotide diversity (Pi) was 0.00119.

IR contraction and expansion in the chloroplast genome. The IR and SC boundaries of the three *Adinandra* were compared. Overall, the results indicated that the size, organization and gene content of the chloroplast genomes were highly similar among the three species. The size of IR ranges from 26,089 bp (*A. megaphylla* Hu) to 26,095 bp (*A. millettii*). And the size of IR of *A. angustifolia* was 26,092 bp. The *ndhF* gene was situated within the LSC region with a 5 bp overlap with the IRA for all three *Adinandra* species. Similarly, the

| Codon | AA | ObsFreq | RCSU | Codon | AA | ObsFreq | RCSU | Codon | AA | ObsFreq | RCSU |
|-------|----|---------|------|-------|----|---------|------|-------|----|---------|------|
| UAG | * | 672 | 0.73 | AUC | I | 1173 | 0.77 | CGC | R | 230 | 0.41 |
| UGA | * | 914 | 0.99 | AUU | I | 1853 | 1.22 | CGG | R | 361 | 0.65 |
| UAA | * | 1177 | 1.28 | AAA | K | 2071 | 1.33 | CGU | R | 401 | 0.72 |
| GCA | A | 382 | 1.06 | AAG | K | 1049 | 0.67 | AGC | S | 559 | 0.67 |
| GCC | A | 347 | 0.96 | CUA | L | 661 | 0.81 | AGU | S | 795 | 0.96 |
| GCG | A | 210 | 0.58 | CUC | L | 640 | 0.79 | UCA | S | 870 | 1.05 |
| GCU | A | 505 | 1.4 | CUG | L | 475 | 0.58 | UCC | S | 908 | 1.1 |
| UGC | C | 477 | 0.81 | CUU | L | 1019 | 1.25 | UCG | S | 638 | 0.77 |
| UGU | C | 697 | 1.19 | UUA | L | 934 | 1.15 | UCU | S | 1205 | 1.45 |
| GAC | D | 426 | 0.55 | UUG | L | 1154 | 1.42 | ACA | T | 680 | 1.12 |
| GAU | D | 1131 | 1.45 | AUG | M | 889 | 1 | ACC | T | 628 | 1.04 |
| GAA | E | 1336 | 1.39 | AAC | N | 798 | 0.6 | ACG | T | 412 | 0.68 |
| GAG | E | 588 | 0.61 | AAU | N | 1863 | 1.4 | ACU | T | 701 | 1.16 |
| UUC | F | 1579 | 0.83 | CCA | P | 790 | 1.28 | GUA | V | 654 | 1.18 |
| UUU | F | 2235 | 1.17 | CCC | P | 605 | 0.98 | GUC | V | 450 | 0.81 |
| GGA | G | 816 | 1.44 | CCG | P | 386 | 0.62 | GUG | V | 377 | 0.68 |
| GGC | G | 347 | 0.61 | CCU | P | 694 | 1.12 | GUU | V | 741 | 1.33 |
| GGG | G | 543 | 0.96 | CAA | Q | 1033 | 1.4 | UGG | W | 677 | 1 |
| GGU | G | 562 | 0.99 | CAG | Q | 445 | 0.6 | UAC | Y | 695 | 0.65 |
| CAC | H | 375 | 0.58 | AGA | R | 1136 | 2.04 | UAU | Y | 1452 | 1.35 |
| CAU | H | 926 | 1.42 | AGG | R | 624 | 1.12 | | | | |
| AUA | I | 1523 | 1 | CGA | R | 582 | 1.05 | | | | |

Table 3. Relative synonymous codon usage (RSCU) for protein-coding genes in *A. megaphylla* Hu. *stop codon.

rps19 gene was positioned within the LSC region with a 6 bp overlap with the IRb. The border across IRa and SSC was located in the region of the *ycf1* gene with 1067 bp tail section of the gene placed in the IRa (Fig. 6). Results of the IR analysis witnessed neither expansion nor contraction of IR regions in the three species.

Phylogenetic inference. As shown in Fig. 7A, the phylogenetic analysis was based on *matK* sequences recovered good resolution among genera. In the Pentaphragaceae, *Euryodendron* and *Adinandra angustifolia* separates outside other genera. The clade of genus *Ternstroemia* and *Anneslea* were sisters to the clade of *Adinandra* and *Eurya* genera. Indeed, all six *Adinandra* species are grouped in one clade, which is divided into three subclades with 95% support; *A. millettii* stood alone in one subclade, *A. integerrima* and *A. dumosa* formed the second subclade, three other species separated into the third one. These results were different from the previous study¹⁵, in which phylogenetic analysis of *A. angustifolia*, *A. millettii*, *Anneslea fragran* and *Ternstroemia gymnanthera* inferred from the LSC dataset indicated that they belong to one clade (bootstrap values = 100%). This difference might be due to the shortcoming of indicates in phylogenetic analysis when only these four species were representatives of the Pentaphragaceae appearing in Zhang et al.'s study. In contrast with the *matK* sequence, the *trnL* region dataset yielded less phylogenetic resolution than the bootstrap value was 59% at the clade of the genus *Adinandra* (Fig. 7B). Additionally, the *Adinandra* was separated into six subclades; one constructed by the studied *A. megaphylla* Hu, *A. formosana* and *A. lasiostyla* constructed two distinct subclades. *A. hirta* and *A. glischroloma*; *A. millettii* and *A. hainanensis*; *A. angustifolia* and *A. dumosa* formed three separated subclades, respectively (Fig. 7B). In the case of barcoding among the Pentaphragaceae family, the *matK* sequence is suggested for better phylogenetic resolution.

Discussion

Pentaphragaceae is a family of flowering plants and contains 12 genera including approximately 345 species over the world¹⁶. A total of 8 cp genomes in the Pentaphragaceae family have been published currently, 2 of which belong to *Adinandra*. The genus *Adinandra* consists of about 85 species mainly distributed in Bangladesh, Cambodia, China, India, Indonesia, Southern Japan, Laos, Malaysia, Myanmar, New Guinea, Philippines, Sri Lanka, Thailand, and the African tropical forest¹⁷. Because of bioactive compounds, many species in the genus *Adinandra* are of interest^{18–23}.

In the present study, we recently sequenced whole cp genomes for one Vietnamese.

Adinandra megaphylla Hu and implemented comparative analyses on three *Adinandra* cp genomes to explore the structure of cp genomes in the taxa. Gene organization together with codon usage patterns was characterized and results indicated the high conservation, which can be helpful for phylogenetic and population genetics studies.

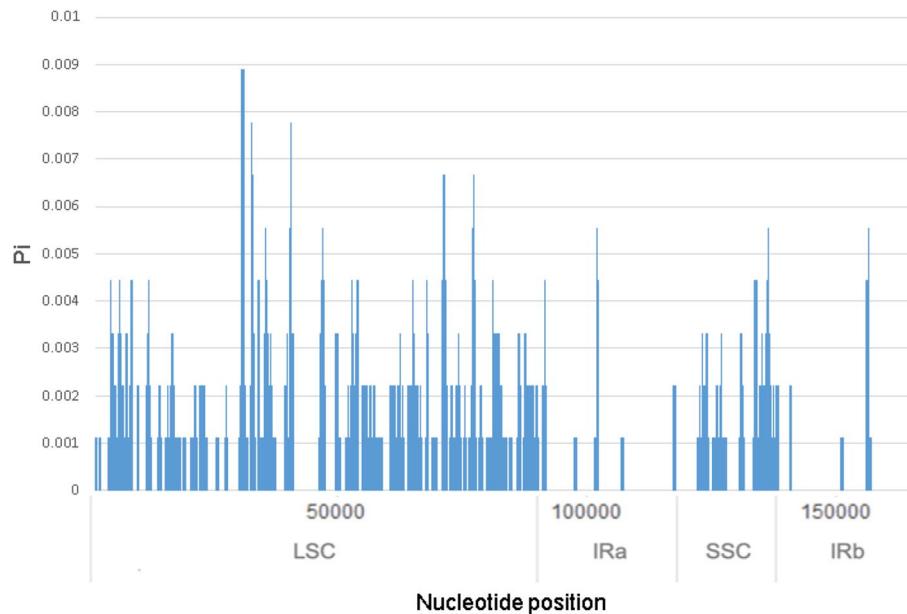


Figure 5. Comparative analysis of nucleotide diversity (Pi) values among the three *Adinandra* species cp genome sequences.

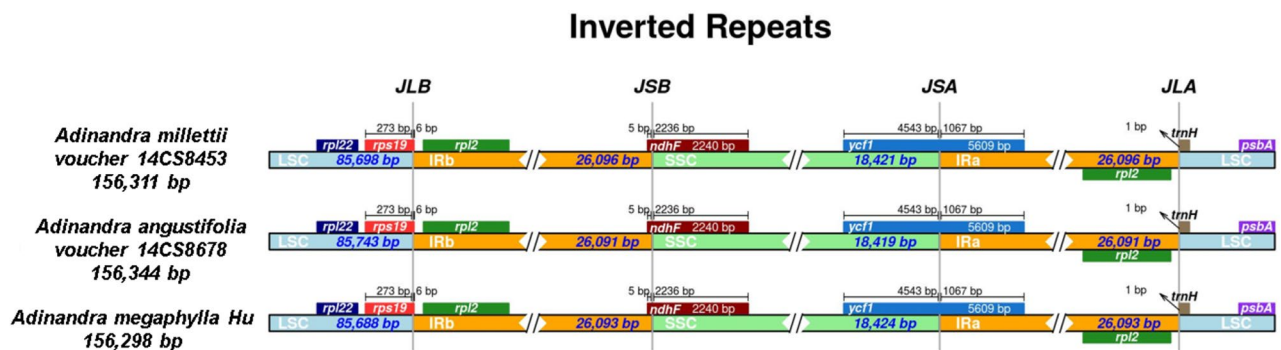


Figure 6. Comparison of LSC, IR and SSC junction positions among the three chloroplast genomes. JLB (junction IRb/LSC), JSB (junction IRb/SSC), JSA (junction IRa/SSC), JLA (junction IRa/LSC).

considered as an indication of recombination³⁰. Due to the potential to generate secondary structures, repeated sequences can act as recognition signals during the recombination process³¹. It is supposed that recombination rarely occurs in angiosperms because of the predominance of uniparental inheritance. Nevertheless, evidence of intermolecular homologous recombination in flowering plants has been reported^{32,33}. To date, studies screening plastome recombination in the taxa are entirely lacking. There was no research demonstrating the presence of plastome recombination in Pentaphragaceae. In this study, the higher number of repeats in comparison with previous estimates might not be substantiation for inter- and intra-specific plastome recombination.

In terms of constructing phylogenetic relationships of plants, complete chloroplast genomes contribute adequate information and have proven their effectiveness in the capability of classification in lower taxonomic levels^{34,35}. *matK* is one of the common DNA barcodes used in plants³⁶. However, the phylogeny results indicated that using only a single gene for species classification may generate different results from different genes. The combination of these barcodes can lead to better species identification.

Conclusion

In this study, three complete chloroplast genomes of *Adinandra* were investigated, including one firstly sequenced chloroplast genomes (*A. megaphylla* Hu) comparatively analyzed with other published genus in the family of Pentaphragaceae for the first time. We assemble the complete chloroplast genome of *A. megaphylla* Hu with 156,298 bp. The structure and gene content of the chloroplast genome of three *Adinandra* were similar and appeared highly conserved. Finally, the phylogenetic relationships built for species of Pentaphragaceae, in terms of comparison public date with our novel sequence of *Adinandra* species. This study provides the potential of chloroplast genome sequences for enhancing species classification and phylogenetic research for in-depth study within Pentaphragaceae.

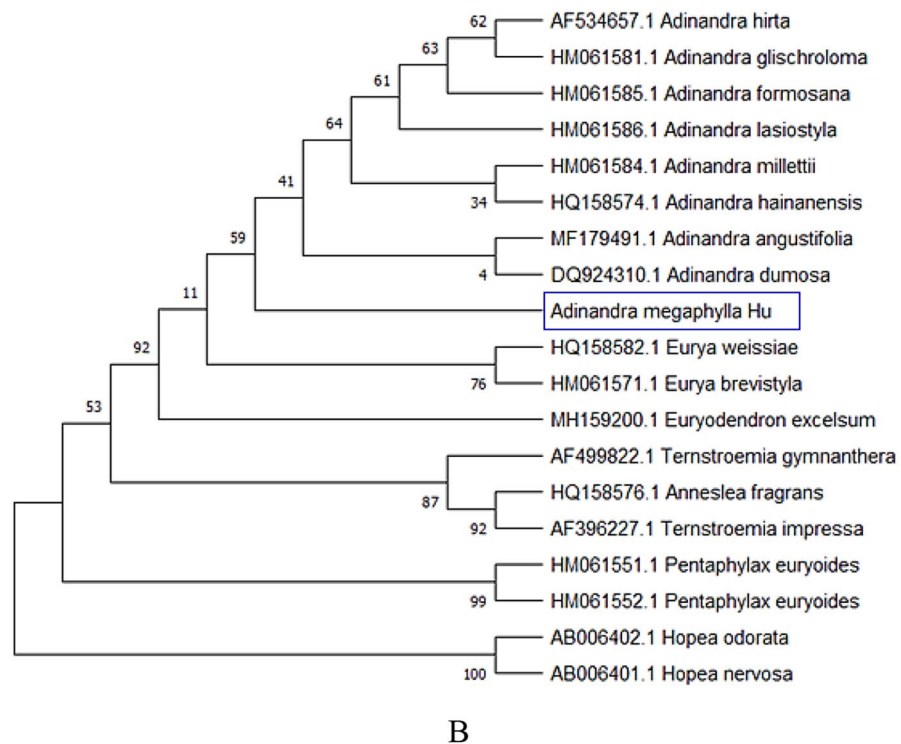
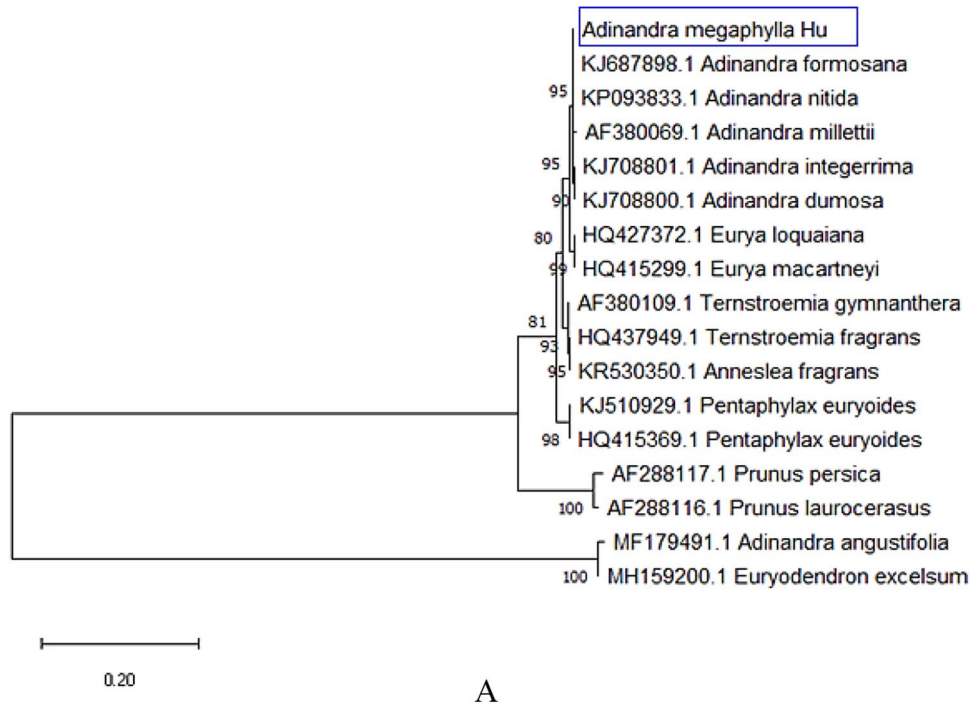


Figure 7. Phylogenetic relationship was inferred using the Maximum Likelihood method based on *matK* (A) and *trnL* (B) genes.

Material and methods

Sample collection. Samples were collected in Hoang Lien—Van Ban Nature Reserve that belongs to Liem Phu Commune, Van Ban District, Lao Cai Province, Vietnam in August 2019 (code number: Nguyen Huu Quan 01), 1200 m, 21°59'15"N; 104°19'28"E. The taxonomic identification is authenticated by Associate Professor Danh Thuong SY, the head of Botany Department in Faculty of Biology, Thai Nguyen University of Education; the voucher specimens were placed in the Herbarium of the Institute of Ecology and Biological Resources (HN), Hanoi, Vietnam. Fresh leaves with the same code number were used to extract genomic DNA (Fig. 8).



Figure 8. Morphological characteristic of *A. megaphylla* Hu. (A) Habit; (B) flowering twig; (C) bud and flower; Photos by Huu Quan Nguyen.

The collection of plant materials has complied with relevant institutional of Hoang Lien—Van Ban Nature Reserve, Vietnamese and international guidelines and legislation.

DNA extraction and chloroplast genome sequencing. Genomic DNA was extracted from young plant leaves using a modified CTAB method³⁷. A260/280 and A260/A230 ratios were measured with the Shimadzu Biospec Nano to assess DNA sample purity. The accurate concentration of double-stranded DNA was determined with Qubit 3 Fluorometer and Qubit HS DNA reagents. Genomic DNA integrity was assessed by agarose gel electrophoresis with 0.8% agarose. Also, DNA libraries were prepared from total genomic DNA using SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, Menlo Park, CA), and adapter ligation was subsequently performed, following the manufacturer's protocol for genomic DNA above 20 kb (Pacific Biosciences). SMRTbell libraries were loaded on one chip and sequenced on a Pacbio SEQUEL system at the Key Laboratory for Gene Technology, Institution of Biotechnology (Hanoi, Vietnam).

Genome assembly and annotation. The total gDNA was sequenced in the PacBio platform by the resequencing method. The sequences derived from the cp genome were identified via the local Blast program³⁸ using *Adinandra angustifolia* (MF179491) cp genomes as the reference¹⁵. Subsequently, the software HGAP4³⁹ was used to assemble the cp genome. The protein-coding, rRNA, and tRNA genes were annotated by the CpGAVAS pipeline⁴⁰. The tRNAscan-SE ver. 1.21 software⁴¹ was applied to verify the tRNA genes with default parameters. The OrganellarGenomeDRAW tool (OGDRAW) ver. 1.3.1⁴² was selected to create the circular gene map. Repeat elements were found using two approaches. Web-based simple sequence repeats finder MISA-web⁴³ was used to detect microsatellites, including 10 repeat units for mono-, 5 repeat units for di-, 4 repeat units for tri-, and 3 repeat units for tetra-, penta-, and hexa-nucleotide SSRs. Among the SSRs of each type, comparing the size of SSRs was employed to count the polymorphic SSRs among the three species. The size and type of repeats in the three *Adinandra* plastomes were investigated using REPuter⁴⁴ with the set parameters as follows: a minimal repeat size of 20 bp, hamming distance of 3 kb, and 90% or greater sequence identity.

Genome comparison. For comparative purposes, we collected two available cp genomes of *A. angustifolia* (#MF179491) and *A. millettii* (#MF179492) from GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>). The overall genome structure, genome size, gene content and repeats across all three *Adinandra* species were compared¹⁵. The whole plastome sequences of the three *Adinandra* plants were aligned with the MAFFT server⁴⁵ and visualized using LAGAN mode in mVISTA⁴⁶. For the mVISTA plot, we used the annotated cp genome of *A. megaphylla* Hu as a reference. The Irscope⁴⁷ was employed to visually display and compare the borders of large single-copy (LSC), small single-copy (SSC), and inverted repeat (IR) regions among the three *Adinandra* species. We also determined the codon usage bias and the sequence divergence among the three *Adinandra* species.

through a sliding window analysis computing pi among the chloroplast genomes in DnaSP ver. 6.12.03⁴⁸. For the sequence divergence analysis, we applied the window size of 600 bp with a 200 bp step size.

Phylogenetic identification. The sequences of *matK* and *trnL* from all *Adinandra* species and other members of the family Pentaphragmaceae from Genbank (<https://www.ncbi.nlm.nih.gov/genbank/>) were used to identify the taxonomic position of the studied *A. megaphylla* Hu. These sequences were aligned with ClustalW mode in Unipro UGENE software v36.0⁴⁹ before a maximum likelihood (ML)⁵⁰ phylogenetic tree was constructed using Mega-X software⁵¹ with 1000 bootstraps. The chosen methods followed the previous study of this genus¹⁵.

Received: 12 March 2021; Accepted: 21 May 2021

Published online: 03 June 2021

References

- Neuhaus, H. E. & Emes, M. J. Nonphotosynthetic metabolism in plastids. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **51**, 111–140. <https://doi.org/10.1146/annurev.arplant.51.1.111> (2000).
- Bendich, A. J. Circular chloroplast chromosomes: The grand illusion. *Plant Cell* **16**, 1661–1666. <https://doi.org/10.1105/tpc.160771> (2004).
- Asaf, S. *et al.* Comparative analysis of complete plastid genomes from wild soybean (*Glycine soja*) and nine other *Glycine* species. *PLoS ONE* **12**, e0182281. <https://doi.org/10.1371/journal.pone.0182281> (2017).
- Alwadani, K. G., Janes, J. K. & Andrew, R. L. Chloroplast genome analysis of box-ironbark Eucalyptus. *Mol. Phylogenet. Evol.* **136**, 76–86. <https://doi.org/10.1016/j.ympev.2019.04.001> (2019).
- Gonçalves, D. J. P., Simpson, B. B., Ortiz, E. M., Shimizu, G. H. & Jansen, R. K. Incongruence between gene trees and species trees and phylogenetic signal variation in plastid genes. *Mol. Phylogenet. Evol.* **138**, 219–232. <https://doi.org/10.1016/j.ympev.2019.05.022> (2019).
- Gitzenanner, M. A., Soltis, P. S., Wong, G. K. S., Ruhfel, B. R. & Soltis, D. E. Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. *Am. J. Bot.* **105**, 291–301. <https://doi.org/10.1002/ajb2.1048> (2018).
- Gu, C., Ma, L., Wu, Z., Chen, K. & Wang, Y. Comparative analyses of chloroplast genomes from 22 Lythraceae species: Inferences for phylogenetic relationships and genome evolution within Myrtales. *BMC Plant Biol.* **19**, 281. <https://doi.org/10.1186/s12870-019-1870-3> (2019).
- Guo, X. *et al.* Plastome phylogeny and early diversification of Brassicaceae. *BMC Genomics* **18**, 176. <https://doi.org/10.1186/s12864-017-3555-3> (2017).
- Saarela, J. M. *et al.* A 250 plastome phylogeny of the grass family (Poaceae): Topological support under different data partitions. *PeerJ* **6**, e4299. <https://doi.org/10.7717/peerj.4299> (2018).
- Chen, Y., She, G. & Chen, H. Study on antitumor activities of substances extracted from *Adinandra nitida* Merr. ex H. L. Li. *Nat. Sci.* **12**, 43–45 (1997).
- Yu, J. & Chen, M. Studies on flavonoids extraction from *Adinandra nitida* Merr. ex H. L. Li. and on their antioxidative and bacteriostatic bioactivities. *J. Shantou Univ.* **12**, 52–58 (1997).
- Wu, C. C., Hsu, Z. F. & Tsou, C. H. Phylogeny and taxonomy of Eurya (Ternstroemiaceae) from Taiwan, as inferred from ITS sequence data. *Bot. Stud.* **48**, 97–116 (2007).
- Nguyen, T. N. L. *et al.* Antibacterial, antioxidant and anti-Cancerous activities of *Adinandra megaphylla* Hu leaf extracts. *Biosci. Biotechnol. Res. Commun.* **13**, 1015–1020. <https://doi.org/10.21786/bbrc/13.3/5> (2020).
- Rosea, J. P. *et al.* Phylogeny, historical biogeography, and diversification of angiosperm order Ericales suggest ancient Neotropical and East Asian connections. *Mol. Phylogenet. Evol.* **122**, 59–79 (2018).
- Yu, X. Q. *et al.* Insights into the historical assembly of East Asian subtropical evergreen broadleaved forests revealed by the temporal history of the tea family. *New Phytol.* **215**, 1235–1248. <https://doi.org/10.1111/nph.14683> (2017).
- Steven, P. F. Angiosperm Phylogeny Website. Version 14, July 2017 [and More or Less Continuously Updated Since]. <http://www.mobot.org/MOBOT/research/APweb/> (2001 onwards).
- Min, T. L. & Bruce, B. B. *Theaceae*. In *Flora of China*, Vol. 12 (eds Wu Z. Y. *et al.*) 435–443 (Science Press, Beijing and Missouri Botanical Garden Press, 2007).
- Yuan, C., Huang, L., Suh, J. H. & Wang, Y. Bioactivity-guided isolation and identification of antiadipogenic compounds in Shiya tea (leaves of *Adinandra nitida*). *J. Agric. Food Chem.* **67**, 6785–6791. <https://doi.org/10.1021/acs.jafc.9b01326> (2019).
- Brad, K. & Zhang, Y. Study on extraction and purification of apigenin and the physical and chemical properties of its complex with lecithin. *Pharmacogn. Mag.* **14**, 203–206. https://doi.org/10.4103/pm.pm_159_17 (2018).
- Chen, Y., Chen, G., Fu, X. & Liu, R. H. Phytochemical profiles and antioxidant activity of different varieties of *Adinandra* tea (*Adinandra* Jack). *J. Agric. Food Chem.* **63**, 169–176. <https://doi.org/10.1021/jf503700v> (2015).
- Gao, H., Liu, B., Liu, F. & Chen, Y. Anti-proliferative effect of camellianin A in *Adinandra nitida* leaves and its apoptotic induction in human Hep G2 and MCF-7 cells. *Molecules* **15**, 3878–3886. <https://doi.org/10.3390/molecules15063878> (2010).
- Liu, B., Yang, J., Ma, Y., Yuan, E. & Chen, C. Antioxidant and angiotensin converting enzyme (ACE) inhibitory activities of ethanol extract and pure flavonoids from *Adinandra nitida* leaves. *Pharm. Biol.* **48**, 1432–1438. <https://doi.org/10.3109/13880209.2010.490223> (2010).
- Zuo, W., Xu, J., Zhou, C. & Gan, L. Simultaneous determination of five flavonoid compounds in leaves of *Adinandra nitida* by HPLC-PAD. *China J. Chin. Mater. Med.* **35**, 2406–2409 (2010).
- Daniell, H., Lin, C. S., Yu, M. & Chang, W. J. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol.* **17**, 134. <https://doi.org/10.1186/s13059-016-1004-2> (2016).
- Palmer, J. D. Comparative organization of chloroplast genomes. *Annu. Rev. Genet.* **19**, 325–354. <https://doi.org/10.1146/annurev.ge.19.120185.001545> (1985).
- Jansen, R. K. *et al.* Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci. USA.* **104**, 19369–19374. <https://doi.org/10.1073/pnas.0709121104> (2007).
- Gao, X. *et al.* Comparative chloroplast genomes of Paris Sect. Marmorata: Insights into repeat regions and evolutionary implications. *BMC Genomics* **19**, 878. <https://doi.org/10.1186/s12864-018-5281-x> (2018).
- Meng, J. *et al.* Comparative analysis of the complete chloroplast genomes of four aconitum medicinal species. *Molecules* <https://doi.org/10.3390/molecules23051015> (2018).
- Zhou, J. *et al.* Complete chloroplast genomes of *Papaver rhoeas* and *Papaver orientale*: Molecular structures, comparative analysis, and phylogenetic analysis. *Molecules* <https://doi.org/10.3390/molecules23020437> (2018).
- Jansen, R. K. & Ruhlman, T. A. Plastid genomes of seed plants. In *Genomics of chloroplasts and mitochondria. Advances in Photosynthesis and Respiration (Including Bioenergy and Related Processes)* (eds Bock, R. & Knoop, V.) 103–126 (Springer, 2012) https://doi.org/10.1007/978-94-007-2920-9_5.

31. Kawata, M., Harada, T., Shimamoto, Y., Oono, K. & Takaiwa, F. Short inverted repeats function as hotspots of intermolecular recombination giving rise to oligomers of deleted plastid DNAs (ptDNAs). *Curr. Genet.* **31**, 179–184. <https://doi.org/10.1007/s002940050193> (1997).
32. Medgyesy, P., Fejes, E. & Maliga, P. Interspecific chloroplast recombination in a nicotiana somatic hybrid. *Proc. Natl. Acad. Sci. USA* **82**, 6960–6964. <https://doi.org/10.1073/pnas.82.20.6960> (1985).
33. Sullivan, A. R., Schiffthaler, B., Thompson, S. L., Street, N. R. & Wang, X. R. Interspecific plastome recombination reflects ancient reticulate evolution in picea (pinaceae). *Mol. Biol. Evol.* **34**, 1689–1701. <https://doi.org/10.1093/molbev/msx111> (2017).
34. He, S., Wang, Y., Volis, S., Li, D. & Yi, T. Genetic diversity and population structure: Implications for conservation of wild soybean (*Glycine soja* Sieb. et Zucc) based on nuclear and chloroplast microsatellite variation. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms131012608> (2012).
35. Zhang, Y. *et al.* The complete chloroplast genome sequences of five epimedium species: Lights into phylogenetic and taxonomic analyses. *Front. Plant Sci.* **7**, 306. <https://doi.org/10.3389/fpls.2016.00306> (2016).
36. Dong, W. *et al.* Discriminating plants using the DNA barcode rbcLb: An appraisal based on a large data set. *Mol. Ecol. Resour.* **14**, 336–343. <https://doi.org/10.1111/1755-0998.12185> (2014).
37. Souza, H. A. V., Muller, L. A. C., Brandão, R. L. & Lovato, M. B. Isolation of high quality and polysaccharide-free DNA from leaves of *Dimorphandra mollis* (Leguminosae), a tree from the Brazilian Cerrado. *Genet. Mol. Res.* **11**, 756–764. <https://doi.org/10.4238/2012.March.22.6> (2012).
38. Johnson, M. *et al.* NCBI BLAST: A better web interface. *Nucl. Acids Res.* **36**(Web Server Issue), W5–W9. <https://doi.org/10.1093/nar/gkn201> (2008).
39. Chin, C. S. *et al.* Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563–569. <https://doi.org/10.1038/nmeth.2474> (2013).
40. Liu, C. *et al.* CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genomics* **13**, 715. <https://doi.org/10.1186/1471-2164-13-715> (2012).
41. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964. <https://doi.org/10.1093/nar/25.5.955> (1997).
42. Lohse, M., Drechsel, O. & Bock, R. OrganellarGenomeDRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* **52**, 267–274. <https://doi.org/10.1007/s00294-007-0161-y> (2007).
43. Beier, S., Thiel, T., Münch, T., Scholz, U. & Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **33**, 2583–2585. <https://doi.org/10.1093/bioinformatics/btx198> (2017).
44. Kurtz, S. & Schleiermacher, C. REPuter: Fast computation of maximal repeats in complete genomes. *Bioinformatics* **15**, 426–427. <https://doi.org/10.1093/bioinformatics/15.5.426> (1999).
45. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **20**, 1160–1166. <https://doi.org/10.1093/bib/bbx108> (2019).
46. Mayor, C. *et al.* VISTA: Visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* **16**, 1046–1047. <https://doi.org/10.1093/bioinformatics/16.11.1046> (2000).
47. Amiroussi, A., Hyvönen, J. & Poccai, P. IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* **34**, 3030–3031. <https://doi.org/10.1093/bioinformatics/bty220> (2018).
48. Rozas, J. *et al.* DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **34**, 3299–3302. <https://doi.org/10.1093/molbev/msx248> (2017).
49. Okonechnikov, K., Golosova, O., Fursov, M. & Team, U. Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics* **28**, 1166–1167. <https://doi.org/10.1093/bioinformatics/bts091> (2012).
50. John, P. H. & Keith, A. C. Phylogeny estimation and hypothesis testing using maximum likelihood. *Annu. Rev. Ecol. Syst.* **28**, 437–466. <https://doi.org/10.1146/annurev.ecolsys.28.1.437> (1997).
51. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).

Acknowledgements

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 106.02-2018.338.

Author contributions

H.Q.N., T.N.L.N., D.T.S., H.H.C. and H.M.C. conceived and designed the experiments; H.Q.N., T.N.L.N., T.N.D., T.L.L. and M.H.P. performed the experiments; D.T.S., H.Q.N., M.H.P., T.L.L. and T.T.N.N. performed data analyses; H.Q.N., T.N.D., T.N.L.N., D.T.S. and H.M.C. prepared the manuscript; T.N.L.N., H.H.C. and H.M.C. made the Proof-reading. All authors approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to T.N.L.N. or H.M.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021, corrected publication 2022