CROP BREEDING AND APPLIED BIOTECHNOLOGY

ARTICLE

Characterization of the complete chloroplast genome of *Camellia tienii* (Theaceae) and its relatives

Crop Breeding and Applied Biotechnology
24(4): e486824415, 2024
Brazilian Society of Plant Breeding.
Printed in Brazil
http://dx.doi.org/10.1590/1984-70332024v24n4a52

Minh Phuong Nguyen[1] and Viet The Ho[1*]

Abstract: *Camellia tienii, an indigenous plant species in Vietnam, is renowned for its remarkable medicinal properties. However, the current methods for plant identification are not reliable due to shared characteristics with other species in Camellia genus. To address this, the present study aimed to sequence and characterize the complete chloroplast (cp) genome of C. tienii. The resulting cp genome is 162,167 bp in size and encompasses 154 genes, including 98 protein-coding genes, 8 rRNA genes, and 48 tRNA genes. Notably, this genome contains 55 simple sequence repeats, specifically type A and T motifs. It is worth mentioning that the ndhF gene was not detected in this genome, while the trnL gene was present. Phylogenetic analysis revealed that C. tienii shares a close relationship with C. tamdaoensis. These significant findings contribute with valuable insights that can aid in the accurate taxonomy, plant identification, and conservation efforts concerning this herb in Vietnam.*

Keywords: *Chloroplast genome, medicinal plant, next generation sequencing, simple sequence repeat*

INTRODUCTION

The *Camellia* genus, belonging to the Theaceace family, stands as one of the largest genera in East and Southeast Asia, with an estimated number of 120 to 300 documented species (Ly et al. 2022). These diverse *Camellia* species find daily application in various domains, such as the production of beverages like green tea, culinary uses, cosmetics (Le et al. 2023), ornamental purposes, and traditional medicine applications (Lu et al. 2012). Additionally, essential oils derived from *Camellia* plants have been explored (Pereira et al. 2022). Within the *Camellia* genus, numerous medicinal compounds have been identified, including alkaloids, steroids, terpenoids, saponins, carotenoids, and polyphenols (An et al. 2023, Tran et al. 2023). These compounds are commonly employed for the treatment of various ailments caused by bacteria, fungi, viruses, and oxidative stress, as well as for their potential anti-tumour effects in humans (Maslov et al. 2022, An et al. 2023). Furthermore, they have been investigated for their potential to reduce hypertension, high cholesterol levels, and obesity (Le et al. 2023).

To date, the classification of *Camellia* species has primarily relied on morphological characteristics, including leaf anatomy data (Erxu et al. 2009, Jiang et al. 2013), as well as the examination of fruits, flowers, and leaves (Syahbudin et al. 2019, Hoi et al. 2021), or biochemical compounds (Gao et al.

***Corresponding author:**
E-mail: thehv@huit.edu.vn
ORCID: 0000-0003-4863-0530

Received: 07 March 2024
Accepted: 03 September 2024
Published: 01 November 2024

[1] Ho Chi Minh University of Industry and Trade, No. 140 Le Trong Tan, Tay Thanh Ward, Tan Phu District, Ho Chi Minh City, Viet Nam

2022). However, due to the limitations of morphological taxonomy in terms of accuracy, various studies have sought to enhance classification through the application of molecular markers. These markers involve different DNA regions, such as RAPD, AFLP, and ISSR (Guinasekare 2007), ITS (Vijayan et al. 2009), or barcode regions like *mat*K, *rbc*L, and ITS2 (Viet et al. 2019), as well as *mat*K, *rbc*L, *ycf*1, and *trn*L-F (Pang et al. 2022). Nevertheless, these markers have relatively narrow coverage of plant genomes, resulting in limited effectiveness in the classification and identification of plants.

In recent years, there has been a significant increase in the cost and time required to sequence the complete chloroplast (cp) genomes of various plant species, including *Goodyera schlechtendaliana* (Oh et al. 2019) and jackfruit (*Artocarpus heterophyllus*) (Lin et al. 2022), due to advancements in next-generation sequencing (NGS) technology. The available cp genomes have provided useful information to develop molecular marker to identify corresponding plant species (Andrade et al. 2018) or investigate the effectiveness of different DNA barcode regions located on cp genome for plant identification (Inglis et al. 2021). For the *Camellia* genus, there have also been several reports of complete cp genomes from different species, such as *Camellia japonica* (Li et al. 2019), *Camellia chuongtsoensis* (Yu et al. 2020), and *Camellia leyeensis* (Xiao et al. 2022). In Vietnam alone, over 40 species have been documented (Dung et al. 2016), with more than 30 species of yellow-flowered *Camellia* identified, establishing Vietnam as a centre of diversity for this particular type of *Camellia* (Do et al. 2019). The number of newly discovered species has experienced a substantial increase (Do et al. 2019, Ly et al. 2022, Quach et al. 2022). However, only a few cp genomes of *Camellia* species endemic to or originating from Vietnam have been published, such as *Camellia vietnamensis* (Lyu et al. 2019, Chen et al. 2023, Hao et al. 2023).

In 2014, *Camellia tienii*, a species endemic to Vietnam, was first discovered in Tam Dao National Park (Ninh and Ninh 2014). Since then, extensive studies have been conducted on various aspects of this species, including the identification of suitable planting areas (Manh et al. 2020), analysis of genotypic and phenotypic diversity (Le et al. 2023), and investigation of its antioxidant activity (Nga et al. 2023). Recently, Anh and colleagues utilized fluorescence microscopy to distinguish *C. tienii* from five other *Camellia* species based on leaf anatomy (Anh et al. 2023). However, this method is time-consuming and costly. Moreover, relying solely on the morphological features for classifying *Camellia* species is not highly accurate (Lu et al. 2012), as these features can vary due to environmental factors and geographic diversity within the *Camellia* genus (Gao et al. 2022, Tran et al. 2023). In our study, we employed next-generation sequencing (NGS) to sequence the chloroplast (cp) genome of *C. tienii*, an endemic species collected in Vietnam. We compared its genome with published cp genomes of *Camellia* species found in or near Vietnam to identify distinct genomic features. The information obtained from this research has significant implications for the taxonomy, botanical identification, breeding, and conservation programs of *C. tienii* in Vietnam.

MATERIAL AND METHODS

Sample collection and DNA sequencing

The specimen of *C. tienii* was obtained from Dong Bua commune, Tam Quan ward, Tam Dao district, Vinh Phuc province, Vietnam. The voucher sample is currently stored at the laboratory of Ho Chi Minh University of Industry and Trade in Vietnam. To extract the total DNA, fresh leaves were processed using the Isolate II Plant DNA Kit from Bioline (UK). The quality of DNA was assessed through 1% gel electrophoresis, while the quantity was measured using Nanodrop from ThermoScientific (Delaware, USA).

First, 500 ng of DNA was fragmented using the S220 Focused-ultrasonicator from Covaris (USA). Subsequently, dA tails were added, adapters were ligated, and the resulting fragments were purified. The library preparation, quality control, cluster generation, and DNA sequencing processes were conducted at Azenta Life Sciences (USA) using the Illumina Novaseq 6,000 sequencer. To assess the quality of the raw reads, FastQC in the Galaxy portal (http://usegalaxy.org) was employed (Nguyen et al. 2024). The NGS data was submitted to the Sequence Read Archive (SRA) database in NCBI under the PRJNA1081179 project and can be accessed at https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1081179.

Chloroplast genome assembly and annotation

The reads from both ends were aligned to the *Camellia vietnamensis* reference sequence with NCBI accession number NC_060778.1 for the assembly process. Subsequently, the Pilon tool (v.1.21) was employed to generate a FASTA file for further analysis (Villanueva-Corrales et al. 2021). The Geseq program (https://chlorobox.mpimp-golm.mpg.de/geseq.

html) was utilized to annotate and locate genes within the chloroplast genomes (Tillich et al. 2017). To identify the counts of protein-coding genes, rRNA, and tRNA in the cp genome, the Chloroplot program (http://irscope.shinyapps.io/Chloroplot/) was utilized (Zheng et al. 2020).

Comparative analysis among *Camellia* cp genomes

The cp genomes were aligned using the MAFFT program (v.7) available at http://mafft.cbrc.jp/alignment/server/. The alignment parameters proposed by Katoh et al. (2019) were used. The junctions of LSC/IRB/SSC/IRA were visualized using IRscope, accessible at http://irscope.shinyapps.io/irapp/ (Hao et al. 2023). The visualization was based on the cp genome annotations of related species available in Genbank, as described by Amiryousefi et al. (2018). To identify SSR motifs, the MIcroSAttelite (MISA) identification tool was utilized. The tool can be accessed at https://webblast.ipk-gatersleben.de/misa/ as described by Beier et al. (2017).

Phylogenetic analysis

The NCBI Genbank (https://www.ncbi.nlm.nih.gov/nucleotide/) was utilized to access additional complete cp genome sequences of various species from the *Camellia* genus. This included retrieving eight cp genomes of *C. vietnamensis* (NC_060778.1; MN078093.1; MN078092.1; MN078084.1; MN078085.1; PP155504.1; PP155501.1; and OL689024.1), five cp genomes of *C. indochinensis* (ON208846; OK135162; ON411685; NC_067091.1; and OM055647.1), one cp genome of *C. tamdaoensis* (NC_069227) and four other cp genomes which are closely related to *C. vietnamensis*, namely *C. oleifera* (MF541730.2), *C. yuhsienensis* (OL689025.1), *C. gauchowensis* (NC_053541.1) and *C. suaveolens* (ON418963.1).

The phylogenetic relationships between the cp genomes were established using the MAFFT alignment mentioned earlier. A phylogenetic tree comprising 14 cp genomes was constructed using Maximum Likelihood (ML) methods with 500 bootstrap replicates, employing MEGA X software. The tree files were then converted to Newick format and annotated using iTOL- Interactive Tree of Life tool (https://itol.embl.de/), as described by Letunic and Bork (2021). Three cp genomes from three genera in the Theaceae family, namely NC_035709.1 (*Anneslea fragrans*), NC_035704.1 (*Pyrenaria diospyricarpa*), and NC_035706.1 (*Ternstroemia gymnanthera*), were utilized as outgroups (Luna and Ochoterena 2004).

RESULTS AND DISCUSSION

Assembly and comparison of cp genomes

In this research, we conducted the first-ever sequencing and annotation of the complete cp genome of *C. tienii* in Vietnam. A dataset of 7.318 GB of 150 bp paired-end reads was generated, resulting in a total of 24,396,390 reads with a Phred score of 96.64%, where the majority of reads (greater than 96.64%) had a quality score greater than Q20. The assembled cp genome exhibited a conserved circular structure with a total length of 162.167 bp, following the typical organization of a plant cp genome. It consisted of four main regions: a Large Single Copy (LSC), a Small Single Copy (SSC), and two Inverted Repeat (IR) regions. These IR regions were located between the LSC and SSC regions (Figure 1). While previous studies have reported the size of *Camellia* cp genomes to be approximately 157 mb (Li et al. 2019, Xiao et al. 2022, Hao et al. 2023), the assembled cp genome of *C. tienii* was found to be significantly longer. Nevertheless, other large cp genomes have been reported, such as 161,078 bp in *C. grijsii* (Xie et al. 2021) and 161,958 bp in *C. vietnamensis* (Lyu et al. 2019), indicating considerable variation in cp genome size within the *Camellia* genus.

A total of 13 cp genomes from three *Camellia* species, available in the NCBI GenBank, were downloaded and compared to the cp sequence of *C. tienii*. The gene compositions are provided in Table 1. A notable distinction was observed between the cp genome of *C. tienii* and the other cp genomes, particularly in terms of the number of coding genes and tRNA genes. The obtained cp genome of *C. tienii* was annotated with a total of 154 genes, including 98 protein-coding genes, 8 rRNA genes, and 48 tRNA genes. In *Camellia* cp genomes, the number of genes encoding rRNA remains constant at 8. However, the number of genes encoding proteins and tRNA shows significant variation across different studies (Li et al. 2019, Xu et al. 2023, Ran et al. 2024). These variations in gene content among cp genomes contribute to our understanding of the genetic structure of the *C. tienii* species.

Simple sequence repeats

Previous studies have highlighted the presence of repeated motifs in cp genomes, which have been associated with various genome rearrangements, recombination events, and large inversions. These motifs have proven valuable for phylogenetic studies. In our analysis of 14 cp genomes, a total of 718 simple sequence repeats (SSRs) were identified. The number of SSRs ranged from 48 in *C. tamdaoensis* to 55 in *C. tienii*, with an average of approximately 51 SSRs per cp genome (Figure 2). Three types of SSR motifs were identified: mononucleotide (A), dinucleotide (T), and trinucleotide (AT). The majority of the identified SSR motifs belonged to the mononucleotide type, with A/T repeat units accounting

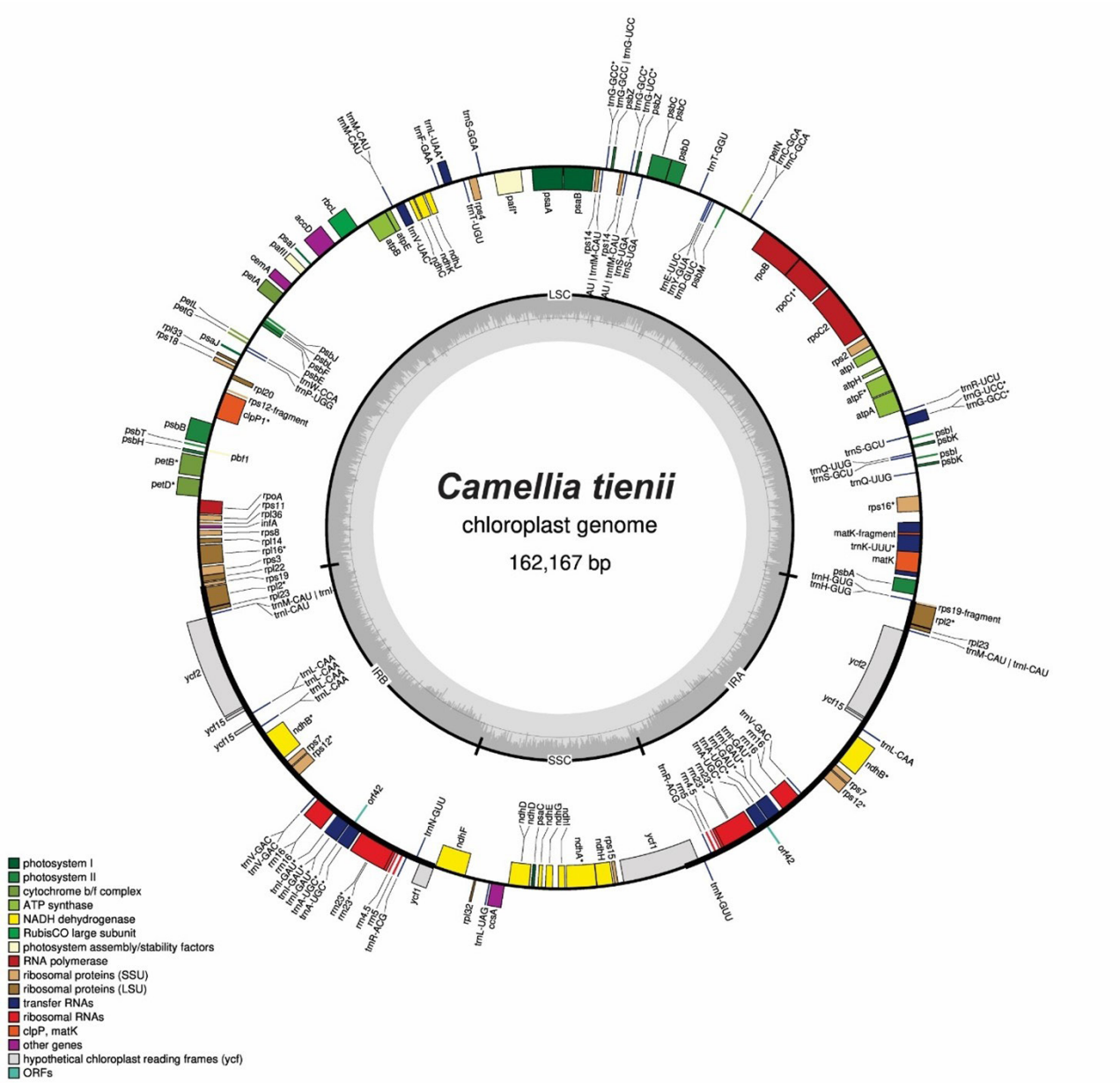

Figure 1. The cp genome map of *Camellia tienii,* generated with http://irscope.shinyapps.io/Chloroplot/, displays the genes transcribed in clockwise and counterclockwise directions, depicted outside and inside of the circle, respectively. The LSC, SSC, IRA, and IRB are labelled as the primary parts of the cp genome. The inner circle's dark and light grey colours represent the GC and AT content, respectively.

Table 1. Size comparison of plastome features of 14 *Camellia* species

Accession	Taxon	Genome size (bp)	Coding genes	rRNA	tRNA
NC_060778.1	*C. vietnamensis*	156,999	91	8	37
MN078093.1	*C. vietnamensis*	156,999	91	8	37
MN078092.1	*C. vietnamensis*	157,004	91	8	37
MN078084.1	*C. vietnamensis*	157,003	91	8	37
MN078085.1	*C. vietnamensis*	157,089	91	8	37
PP155504.1	*C. vietnamensis*	157,000	90	8	37
PP155501.1	*C. vietnamensis*	157,003	90	8	37
OL689024.1	*C. vietnamensis*	156,910	90	8	37
ON208846	*C. indochinensis*	156,621	88	8	37
ON411685	*C. indochinensis*	156,633	92	8	39
NC_067091.1	*C. indochinensis*	156,574	86	8	36
OM055647.1	*C. indochinensis*	156,574	86	8	36
NC_069227	*C. tamdaoensis*	156,633	90	8	35
	C. tienii	162,167	98	8	48

for 43.3% (311 SSRs) and 56.5% (406 SSRs) of the total, respectively. Interestingly, only one dinucleotide type AT motif was found in *C. vietnamensis* (MN078085.1). The abundance of A/T SSR motifs in *Camellia* cp genomes aligns with findings from several previous publications on *Camellia* cp genomes (Wu et al. 2020, Hao et al. 2023, Xu et al. 2023).

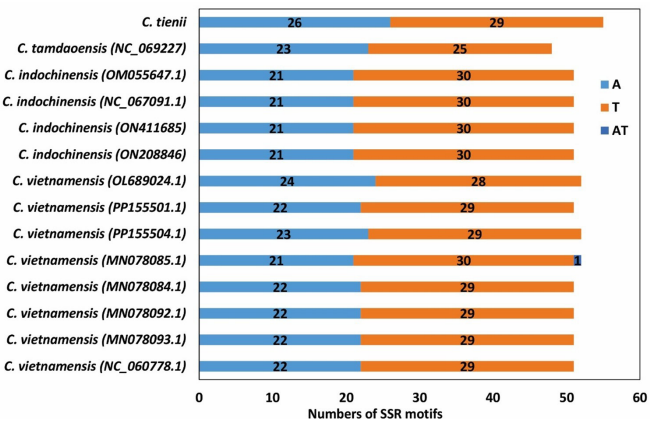

Figure 2. The different simple sequence repeat types in the cp genomes of 14 *Camellia* species.

A notable point is that the number and complexity of SSR in the four *Camellia* species analysed in this study in Vietnam are all lower than in other *Camellia* species. In a study by Li et al. (2019), six *Camellia* cp genomes were analysed, revealing up to four types of SSRs, including mononucleotide, dinucleotide, trinucleotide, and tetranucleotide, with 67 to 74 SSRs per cp genome. Similarly, in a recent study by Ran et al. (2024), a total of 13 *Camellia* species in the *Tubeculata* section were analysed, uncovering up to six types of SSRs, ranging from mononucleotide to hexanucleotide, with SSR numbers ranging from 69 to 75 per genome. The novel and specific microsatellites detected in our study in the cp genome of *C. tienii* hold promise for evolutionary investigations within this species. Additionally, they can aid in the identification and conservation efforts of different species within the *Camellia* genus.

IR contraction and expansion

While the *Camellia* cp genomes generally exhibit a high degree of conservation in terms of genomic structure and size, there are notable variations observed within each species, particularly in the boundary regions between the Inverted Repeat (IR) and Single Copy (SC) regions (Figure 3). However, it is worth mentioning that some variations were also detected within each species. For instance, the *ycf*1 gene was found to be present in only three out of seven cp genomes of *C. vietnamensis*. Similarly, this gene was observed in only one out of three cp genomes of *C. indochinensis*. Although previous publications have reported the presence of *ndh*F genes in all characterized genomes of different *Camellia* species (Wu et al. 2020, Hao et al. 2023, Ran et al. 2024), Figure 3 demonstrates the translocation of this gene, which has also been reported in other studies (Hao et al. 2023). Interestingly, we found that the *ndh*F gene is absent in *C. tienii*. Furthermore, the presence of *trn*L and *ycf*15 genes further contributes to the distinct features of *C. tienii*.

Phylogenetic relationship

The analysis of phylogenetic relationships among the 18 *Camellia* cp genomes revealed a distinct clustering pattern with a high bootstrap value (Figure 4). In the phylogenetic tree, the three cp genomes used as outgroups formed a single clade, indicating the accuracy of the phylogenetic analysis. The cp genome sequence of *C. tienii* in this study is closely related to those of *C. indochinensis*, highlighting a close relationship between these two species. This is expected, as *C. indochinensis* is native to the Indo-China Peninsula, including Laos, Cambodia, and Vietnam. Although Ho et al. (2023) found genetic relatedness between *C. tienii* and *C. tamdaoensis* using *rbc*L or *trn*H-*psb*A DNA barcodes, our data aligns with the study by Le et al. (2023). When using up to three DNA barcodes (*mat*K, *rbc*L, and *psb*A-*trn*H), Le et al. demonstrated that these two species were separated into distinct clades. The discrepancy in clustering may be attributed to the use of different combinations of DNA barcodes, suggesting the necessity of additional loci to increase discrimination power

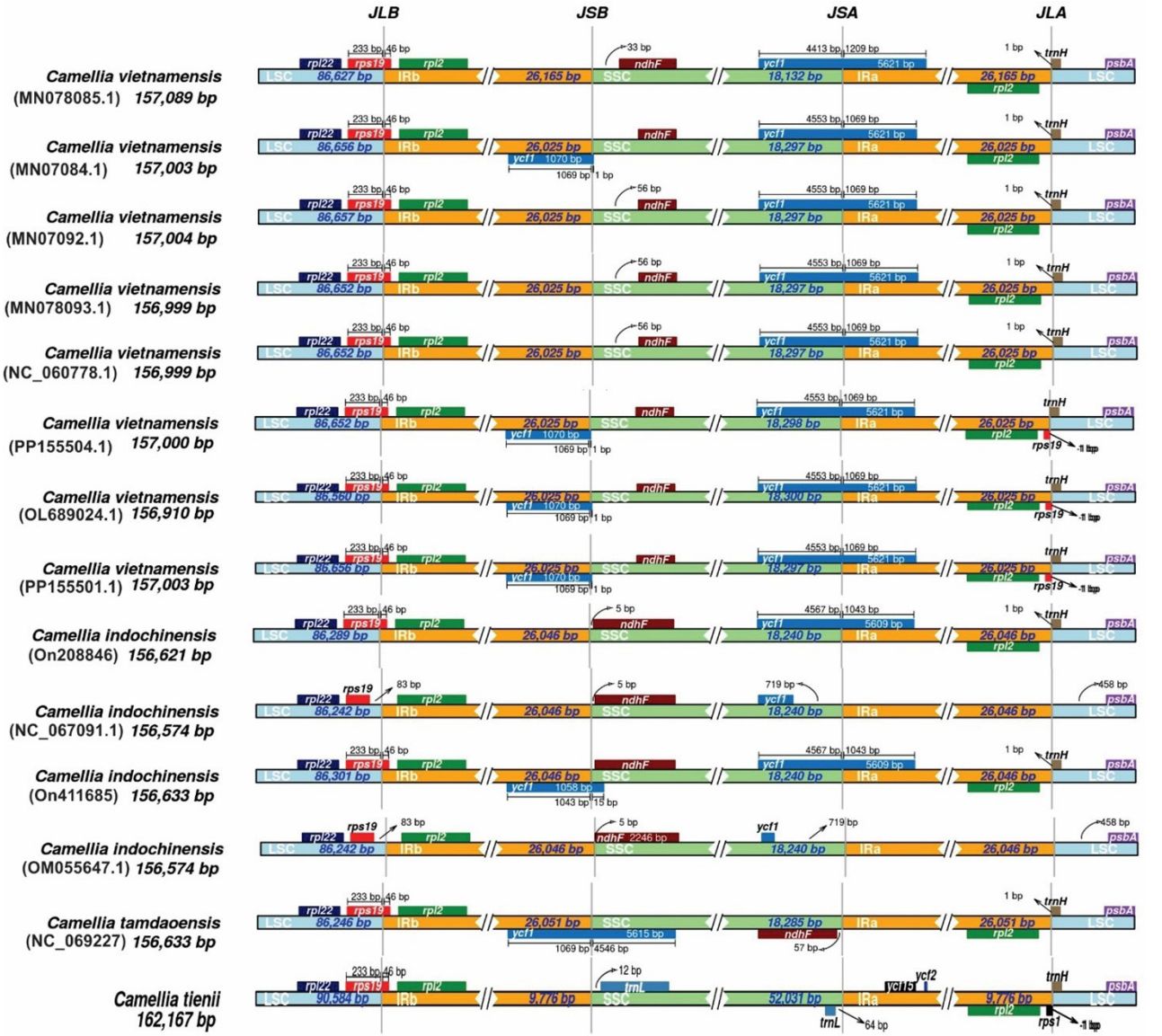

Figure 3. The LSC, IR, and SSC border regions were compared among 14 *Camellia* cp genomes. Genes located at the IR/SC borders are represented by boxes above or below the main lines, with the numbers above the gene indicating the distance in bp from the gene terminal to the boundary region.

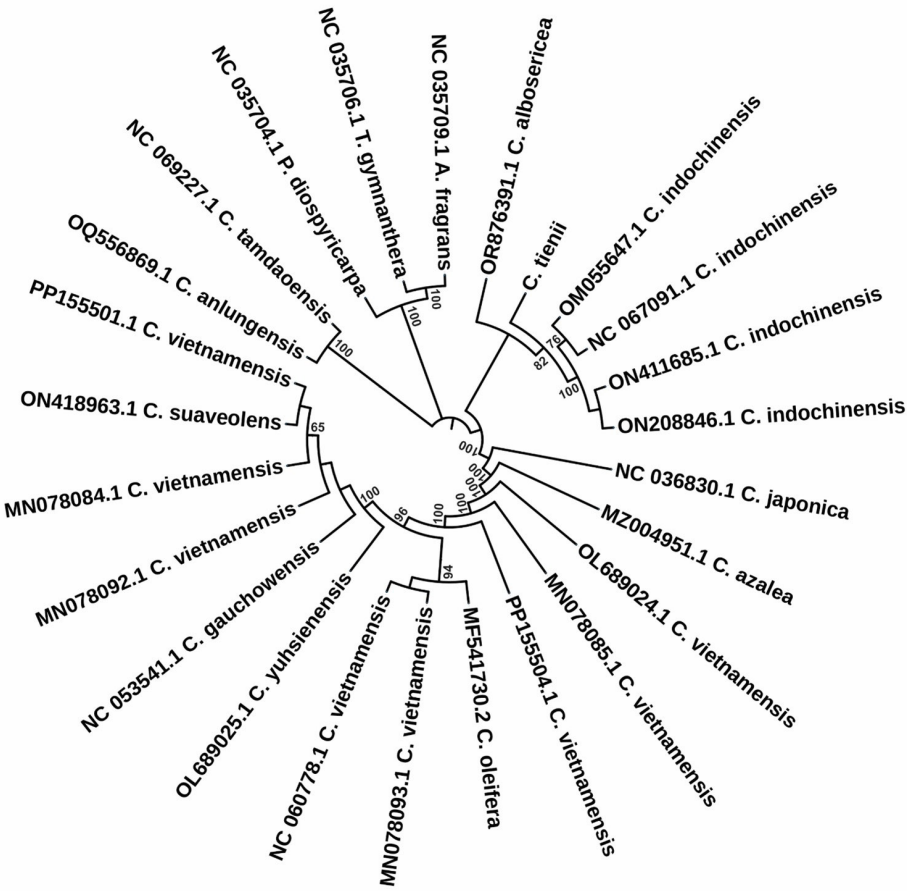

Figure 4. Phylogenetic tree showing the relatedness of *Camellia tienii* with other *Camellia* species using Maximum Likelihood method (Three cp genomes belonging to three genera in Theaceae family consisting of NC_035709.1 (*A. fragrans*), NC_035704.1 (*P. diospyricarpa*), NC_035706.1 (*T. gymnanthera*) were used as outgroups. Numbers near branches are bootstrap values).

among closely related species (Letsiou et al. 2024). Furthermore, Ho et al. (2023) observed disagreements in *Camellia* phylogenetic analysis when comparing the use of a single DNA barcode versus a combination of DNA barcodes. Due to the variability in using DNA barcodes for specimen identification, recent studies emphasize the importance of using next-generation sequencing methods to characterize the whole genome as a super barcode. This approach will enhance our ability to distinguish different species (Coissac et al. 2016, Wu et al. 2021, Ahmed and Zaman 2022).

In this research, we conducted a sequencing and characterization of the complete cp genome of *C. tienii*, an endemic *Camellia* species found in Vietnam. Through comparative analysis with closely related species, distinct features were identified, including variations in genome size, gene numbers, and sequence repeat motifs. The findings from this study offer valuable insights into the typical structure and composition of *C. tienii* cp genomes. These observed differences among cp genomes contribute to our understanding of the genetic makeup within the *Camellia* genus. Additionally, the discovery of unique repeat motifs and highly divergent regions in the cp genome of *C. tienii* presents the potential for the development of molecular markers. These markers could be utilized in future studies focusing on taxonomy and conservation efforts for this valuable herbaceous plant in Vietnam.

DATA AVAILABILITY

The datasets generated and/or analyzed during this research are available from the corresponding author upon request.

REFERENCES

Ahmed F and Zaman MK (2022) A critical review on the challenges and advantages in DNA barcoding for plant identification**. Current Trends in Pharmaceutical Research 9**: 115-139.

Amiryousefi A, Hyvönen J and Poczai P (2018) IRscope: an online program to visualize the junction sites of chloroplast genomes. **Bioinformatics 34**: 3030-3031.

An NTG, Chau DTM, Huong LT, Khoa VV, Hung NH, Thao DT, Trang VTQ, Dai DN and Setzer WN (2023) Lipid peroxidation inhibitory and cytotoxic activities of two *Camellia* species growing wild in Vietnam. **Pharmacognosy Magazine 19**: 385-399.

Andrade MC, Perek M, Pereira FB, Moro M and Tambarussi EV (2018) Quantity, organization, and distribution of chloroplast microsatellites in all species of *Eucalyptus* with available plastome sequence. **Crop Breeding and Applied Biotechnology 18**: 97-102.

Anh LTV, Nga MTP, Linh TP, Anh NM, Linh HTP, Thanh NC, Phuong LTT and Oanh NTK (2023) Application of fluorescence micorscopy for histological discrimination of golden *Camellias* leaves. **Vietnam Journal of Biotechnology 201**: 113-120.

Beier S, Thiel T, Münch T, Scholz U and Mascher M (2017) MISA-web: a web server for microsatellite prediction. **Bioinformatics 33**: 2583-2585.

Chen J, Zhou K and Hu X (2023) Comparisons of chloroplast genome mutations among 13 samples of oil-tea *Camellia* from South China. **Genes 14**: 1083.

Coissac E, Hollingsworth PM, Lavergne S and Taberlet P (2016) From barcodes to genomes: extending the concept of DNA barcoding. **Molecular Biology 25**: 1423-1428.

Do DN, Luong DV, Nguyen CD, Hoang ST, Le TH, Han JE and Park HS (2019) A new yellow *Camellia* (Theaceae) from central Vietnam. **Korean Journal of Plant Taxonomy 49**: 90-95.

Dung LV, Anna L, Hanh NT and Lieu NT (2016) *Camellia thuongiana* - a new yellow *Camellia* species from Vietnam. **Dalat University Journal of Science 6**: 338-244.

Erxu P, Quifa P, Hongfei L, Jingbo S, Yueqiang D, Failai H and Hui H (2009) Leaf morphology and anatomy of *Camellia* section *Camellia* (Theaceae). **Botanical Journal of the Linnean Society 159**: 456-476

Gao S, Wang B, Liu F, Zhao J, Yuan J, Xiao S, Masabni J, Zou F and Yuan D (2022) Variation in fruit morphology and seed oil fatty acid composition of *Camellia oleifera* collected from diverse regions in Southern China. **Horticulturae 8**: 818.

Guinasekare MTK (2007) Applications of molecular markers to the genetic improvement of *Camellia sinensis* L. (tea) – A review. **Journal of Horticultural Science & Biotechnology 82**: 161-169.

Hao B, Xia Y, Zhang Z, Wang D, Ye H and Ma J (2023) Comparative analysis of the complete chloroplast genome sequences of four *Camellia* species. **Brazillian Journal of Botany 47**: 93-103.

Ho VT, Nguyen TH, Nguyen LBN, Nguyen TTN and Nguyen MP (2023) Comparison of *rbc*L and *trn*H-*psb*A DNA barcodes in diverse *Camellia* species collection in Vietnam. **SABRAO Journal of Breeding and Genetics 55**: 877-885.

Hoi QV, Doudkin RV and Khoi NT (2021) The position of the genus *Camellia* L. (Theacea) in some classification system. **University procedes Volga region Natural Sciences 1**: 33-39.

Inglis PW, Bellon G, Junqueira NTV, Faleiro FG and Ferreira ME (2021) Phylogenetic inference applied to germplasm bank characterization and interspecific breeding in passionfruit**. Crop Breeding and Applied Biotechnology 21**: e362221112.

Jiang W, Ozaktas BB, Mantri N, Tao Z and Lu H (2013) Classification of *Camellia* species from 3 sections using leaf anatomical data with back-propagation neural networks and support vector machines. **Turkish journal of Botany 37**: 1093-1103.

Katoh K, Rozewicki J and Yamada KD (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. **Briefings in Bioinformatics 20**: 1160-1166.

Le TVA, Nga TPM, Nguyen PN and Nguyen TKO (2023) Genotypic and phenotypic diversity of endemic golden *Camellias* collected from the North of Vietnam. **Biochemistry & Biodiversity 20**: e202200843.

Letsiou S, Madesis P, Vasdekis E, Montemurro C, Grigoriou ME, Skavdis G, Moussis V, Koutelidakis AE and Tzakos AG (2024) DNA barcoding as a plant identification method. **Applied Sciences 14**: 1415.

Letunic I and Bork P (2021) Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. **Nucleic Acids Research 49**: W293-W296.

Li W, Zhang C, Guo X, Liu Q and Wang K (2019) Complete chloroplast genome of *Camellia japonica* genome structures, comparative and phylogenetic analysis. **Plos ONE 14**: e0216645.

Lin X, Feng C, Lin T, Harris AJ, Li Y and Kang M (2022) Jackfruit genome and population genomics provide insights into fruit evolution and domestication history in China. **Horticulture Research 9**: uhac173.

Lu H, Jiang W, Ghiasi M, Lee S and Nitin M (2012) Classification of *Camellia* (Theaceae) species using leaf architecture variations and pattern recognition techniques. **Plos ONE 7**: 329704.

Luna I and Ochoterena H (2004) Phylogenetic relationships of the genera of Theaceae based on morphology. **Cladistics 20**: 223-270.

Ly NS, Luong VD, Le TH, Do ND, Tran N, Nguyen AD, Nguyen TL, Uematsu C and Katayama H (2022) *Camellia ananmensis* (Theaceae), a new species from central Vietnam. **Taiwania 67**: 243-249.

Lyu T, Wang Y, Hung Z, Li X, Li J, Fan F, Ye N and Yin H (2019) The complete chloroplast genome of *Camellia vietnamensis*, an economic shrub producing edible seed oil. **Mitochondrial DNA Part B 4**: 3736-3737.

Manh TD, Do TV, Tuan NV, Trung PD, Thang NT, Thuyen DV, Duc DT, Linh MT, Khuong NV, Lam VT, Thinh NH, Son HT, Bon TN, Luong HT, Thuan VV and Phuong NTT (2020) Mapping potential planting areas for golden *Camellias* in North Vietnam. **Walailak Journal of Science and Technology 17**: 1095-1103.

Maslov O, Kolisnyk S, Komisarenko M and Golik M (2022) Study of total antioxidant activity of green tea leaves (*Camellia sinensis* L.). **Herba Polonica 68**: 1-9.

Nga NP, Oanh NTK and Linh ND (2023) Evaluation of the antioxidant activity of two endemic golden *Camellia* species (*Camellia tamdaoensis* Ninh et aHakoda and *Camellia tienii* Ninh) in Vietnam. **TNU Journal of Science and Technology 228**: 333-340.

Nguyen MP, Trinh TH, Dao KD, Luu PL and Ho VT (2024) Complete chloroplast genome of the Jewel Orchid, *Anoectochilus formosanus* (Orchidaceae) and its relatives. **Revista de Biología Tropical 72**: e56423.

Ninh T and Ninh LNH (2014) The yellow *Camellias* of the Tam Dao national park. **Agricultural and Food Sciences**: 122-128.

Oh SH, Suh HJ, Park J, Kim Y and Kim S (2019) The complete chloroplast genome sequence of *Goodyera schlechtendaliana* in Korea (Orchidaceae). **Mitochondrial DNA B Resources 4**: 2692-2693.

Pang Z, Wang YL, Mantri N, Wang Y, Hua ZJ, Quang YP, Zhou X, Jiang ZD, Qi ZC and Lu HF (2022) Molecular phylogenetic relationships and taxonomy position of 161 *Camellia* species in China. **Taiwania 67**: 560-570.

Pereira AG, Garcia-Perez P, Cassani L, Chamorro F, Cao H, Barba FJ, Simal-Gandara J and Prieto MA (2022) *Camellia japonica*: A phytochemical perspective and current applications facing its industrial exploitation. **Food Chemistry: X 13**: 100258.

Quach VH, Hoang TT, Truong QC, Le VH, Luong VD Vo QT, Nong VD and Curry AS (2022) *Camellia phuongchiana* (Theaceae, section Piquetia), a new species from Lam Dong Province, Southern Vietnam. **Brittonia 74**: 339-345.

Ran Z, Li Z, Xiao X, An M and Yan C (2024) Complete chloroplast genomes of 13 species of sect. *Tuberculata* Chang (*Camellia* L.): genomic features, comparative analysis, and phylogenetic relationships. **BMC Genomics 24**: 108.

Syahbudin A, Widyastutu A, Masruri NW and Meinata A (2019) Morphological classification of ta clones (*Camellia sinensis*, Theaceae) at the Mount Lawu forest, East Java, Indonesia. **IOP Conference Series: Earth and Environmental Science 394**: 012014.

Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R and Greiner S (2017) GeSeq – versatile and accurate annotation of organelle genomes. **Nucleic Acids Research 45**: w6-w11.

Tran TT, Takahashi K, Nishikawa H, Tanabe R, Nguyen TTT, Nguyen VHT, Nguyen TMT and Bui HV (2023) Habitat characteristics of *Camellia quephongensis* and adaptation mechanisms in Que Phong district, North-Central Vietnam. **International Journal of Plant Biology 14**: 959-973.

Viet NV, Chung PQ, Trung DQ, Ha TV and Douangmala S (2019) Testing three proposed DNA regions (*mat*K, *rbc*L, and ITS2) for identification of *Camellia euphlebia* and *Camellia chrysantha*. **Journal of Forestry Science and Technology 7**: 19-23.

Vijayan K, Zhang WJ and Tsou CH (2009) Molecular taxonomy of *Camellia* (Theaceae) infferred from rnITS sequences. **American Journal of Botany 96**: 1348-1360.

Villanueva-Corrales S, García-Botero C, Garcés-Cardona F, Ramírez-Ríos V, Villanueva-Mejía DF and Álvarez JC (2021) The complete chloroplast genome of *Plukenetia volubilis* provides insights into the organelle inheritance. **Frontiers Plant Science 12**: 667060.

Wu L, Li J, Gu Y, Zhang F, Li Z and Tan X (2020) Complete chloroplast genome sequences and phylogenetic analysis of three *Camellia* oleifera cultivars. **International Journal of Agriculture & Biology 24**: 743-750.

Wu L, Wu M, Cui N, Li Y, Li X and Chen S (2021) Plant super-barcode: a case study on genome-based identification for closely related species of *Fritillaria*. **Chinese Medicine 16**:52.

Xiao X, Lu J, Yang G and Li Z (2022) The complete chloroplast genome of *Camellia leyeensis* (Theaceae). **Mitochondrial DNA Part B 7**: 735-737.

Xie F, Sun Y, Li X, Nie Z, Yin H and Li J (2021) The complete chloroplast genome of *Camellia grijsii*, an ornamental shrub with floral aroma. **Mitochondrial DNA part B 6**: 748-749.

Xu Y, Liu Y, Yu Z and Jia X (2023) Complete chloroplast genome sequence of the long blooming cultivar *Camellia* 'Xiari Qixin': Genome features, comparative and phylogenetic analysis. **Genes 14**: 460.

Yu B, Sun YB, Huang LL, Xu YC, Zhao CY and Liu XF (2020) The complete chloroplast genome sequence of *Camellia chuongtsoensis*. **Mitochondrial DNA Part B 6**: 247-249.

Zheng S, Poczai P, Hyvonen J, Tang J and Amiryousefi A (2020) Chloroplot: An online program for the versatile plotting of organelle genomes. **Frontiers in Genetics 11**: 576124.