 DOI: 10.35382/TVUJS.14.2.2024.3909

CHARACTERIZATION OF THE CHLOROPLAST GENOME OF *AQUILARIA CRASSNA* PIERRE EX LECOMTE DISTRIBUTED IN VIETNAM USING LONG-READ DATA (OXFORD NANOPORE TECHNOLOGIES)

Nguyen Nhat Nam[1], Do Hoang Dang Khoa[2*]

Abstract – *Aquilaria crassna Pierre ex Lecomte is a critically endangered species that is famous for its fragrance compounds for perfume and incense production. In this study, the complete chloroplast genome of A. crassna was sequenced using a long-read sequencing method of Oxford Nanopore Technologies. The complete chloroplast genome of A. crassna was 174,766 bp in length, which comprised a large single copy region of 87,231 bp, a small single copy region of 3,343 bp, and two inverted repeat regions of 42,096 bp. This genome contained 95 protein-coding genes, 38 transfer RNA genes, and eight ribosomal RNA genes. The analysis of small single repeats revealed 39 mononucleotide and seven dinucleotide repeats, which are located mainly in non-coding regions. Comparative genomic analysis exhibited high conservation of the chloroplast genome within the Aquilaria genus regarding gene content, gene order, and genomic structure. This study provides useful genomic data for further studies examining the genetic population, genetic conservation, molecular markers, and phylogeny of Aquilaria crassna and related species within Thymelaeaceae.*

Keywords: comparative genomics, Do Bau, genomic evolution, Oxford Nanopore Technologies, plastome, Thymelaeaceae.

I. PREFACE

The high-throughput sequencing technologies enabled a significant increase in genomic studies of various organisms such as plants, animals, bacteria, and fungi. The genomic data have provided essential information about the evolutionary history of different species, especially vulnerable organisms. Additionally, the genomic data are useful sources for exploring genetic populations, mining molecular markers, and reconstructing phylogeny. Therefore, it is necessary to build up a genomic database of different species. In this study, the complete chloroplast genome of *Aquilaria crassna*, a critically endangered species that is a source of high-economic agarwood, will be sequenced using a long-read sequencing method developed by Oxford Nanopore Technologies. The features of the chloroplast genome will be characterized regarding genome structure, genome size, and gene content. Additionally, the comparative chloroplast genomic analysis among *Aquilaria* species will be conducted to identify genomic events. These results will provide fundamental genomic data for exploring the genomic evolution of *A. crassna* distributed in Vietnam compared to those in other regions.

II. INTRODUCTION

Aquilaria Lam. is a member of Thymelaeaceae and consists of 27 species distributed in Southeast Asia, China South-Central, China Southeast, East Himalaya, Hainan, New Guinea, Assam, and Bangladesh [1]. The *Aquilaria* species is a source of fragrant resinous wood, called agarwood, with high commercial value, resulting from wounds [2]. Previous studies revealed different chemical compounds in *Aquilaria* species, such as Neopetasone, Dihydroagarofuran-15-al, Jinkocramol [3, 4]. The extract of *Aquilaria* species

[1]Tra Vinh University, Vietnam
[2]Nguyen Tat Thanh University, Vietnam
*Corresponding author: dhdkhoa@ntt.edu.vn
Received date: 25th February 2024; Revised date: 07th June 2024; Accepted date: 24th June 2024

exhibited properties of antioxidant, antiglycation, and anti-inflammatory [5, 6]. The results revealed potential medicinal values that can be developed for high economic products. Therefore, further studies examining the genomic data of *Aquilaria* are needed.

The chloroplast genome (cpDNA) is an essential part of land plants, ranging from 17 kb to 200 kb in length and consisting of a large single copy (LSC), a small single copy (SSC), and two inverted repeats (IR) regions [7]. The cpDNA contains genes that are responsible for photosynthesis. Therefore, cpDNA is a useful source for elucidating the evolutionary history of plants. Previously, 3654 cpDNA sequences were used to explore the plant phylogeny [8]. Additionally, cpDNA data have been used for developing molecular markers for different plants [9]. Within the *Aquilaria* genus, the complete cpDNAs of different species have been sequenced and reported [10–13]. These data provided crucial genomic data for further studies about the evolution of *Aquilaria* species.

This study employed the long-read sequencing method of Oxford Nanopore Technologies to sequence the complete chloroplast genome of *A. crassna* collected in Binh Phuoc Province, Vietnam. The newly sequenced cpDNA was characterized regarding genomic structure, gene content, and small single repeat (SSR) information. Additionally, the comparative chloroplast genomic analysis of *Aquilaria* species was conducted to explore the variations of genomic features such as genome size, gene content, GC content, LSC/SSC/IR junctions, and pairwise similarity.

III. MATERIALS AND METHODS

Fresh leaves of an *A. crassna* individual were collected at Binh Phuoc Province ($11^o55'19.6"$ N, $107^o01'01.6"$ E). The samples were then dried using silica gel beads and stored at 80^oC for further experiments. Total genomic DNA was extracted using the DNeasy Plant Mini Kit (Qiagen, USA). The quality of the extracted DNA was checked using agarose gel electrophoresis and a Nanodrop OneC spectrophotometer (Thermo Fisher Scientific, USA). The high-quality DNA (showing a clear band on the gel and having a minimum concentration of 100 ng/μL with an A260/A280 ratio of $\sim$ 1.7–2.0 and A260/A220 ratio of $\sim$ 2.0–2.2) was used for preparing the library using Ligation Sequencing Kit V14 (SQK-LSK114) following the manufacturer's instructions. Consequently, the library was sequenced using MinION Mk1B with flowcell R9.4.1, following the manufacturer's instructions.

The raw long-read data was base-called using Guppy v6.5.7 [14] to generate approximately 2,210,063 reads that ranged from 18 bp to 204,287 bp. The reads were then assembled to complete the chloroplast genome using the Raven tool with the default setting [15]. The complete chloroplast genome was annotated using Geseq [16]. The OGDraw program was used to illustrate the map of cpNDA [17]. The small single repeats (SSR) were identified using Phobos embedded in Geneious Prime v2023.1 with a minimum length of 10 bp, 12 bp, 15 bp, 16 bp, 20 bp, and 24 bp for mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats, respectively [18]. The complete chloroplast genomes of other *Aquilaria* species were searched and downloaded from the GenBank database. Consequently, 12 complete cpDNAs were used, including *A. crassna* (MN125348), *A. rugosa* (MZ145049), *A. yunnanensis* (MG656407), *A. agallochum* (MZ145047), *A. cumingiana* (MZ145048), *A. malaccensis* (MH286934), *A. beccariana* (MN125347), *A. subintegra* (MN147871), *A. microcarpa* (MN125350), *A. hirta* (MN125349), *A. rostrata* (MN125351), and *A. sinensis* (MN720647). These sequences were imported to Geneious Prime v2023.1 to compare the features of genome size, gene content, GC content, and the boundaries among LSC, SSC, and IR regions using the MAUVE alignment option with the default setting.

IV. RESULTS AND DISCUSSION

The assembly process resulted in the complete chloroplast genome of *A. crassna*, which was

174,766 bp in length with an average coverage of 2905x. This genome consisted of a large single copy (LSC, 87,231 bp), a small single copy (SSC, 3,343 bp), and two inverted repeats (IR, 42,096 bp) regions (Figure 1). Additionally, there were 95 protein-coding genes, 38 tRNA genes, and 8 rRNA genes in the cpDNA of *A. crassna* (Table 1). Among these genes, 16 protein-coding genes (including *ndhA, ndhB, ndhD, ndhE, ndhG, ndhH, ndhI, psaC, rpl23, rp2, rps7, rps12, rps15, ccsA, ycf1, and ycf2*), eight tRNA (including *trnA-UGC, trnI-CAU, trnI-GAU, trnL-CAA, trnL-UAG, trnN-GUU, trnR-ACG, and trnV-GAC*), and four rRNA genes (including *rrn4.5, rrn5, rrn18, and rrn26*) were duplicated in IR regions. The *pafI* and *clpP1* contained two introns, whereas *trnA-UGC, trnG-UCC, trnI-GAU, trnK-UUU, trnL-UAA, trnV-UAC, atpF, ndhA, ndhB, petB, petD, rpl2, rpl16, rps16*, and *rpoC1* had one intron. The *rps12* gene was a trans-splicing gene. The GC content of the *A. crassna* cpDNA was 36.7%. The boundary of LSC/IR regions was in intergenic space (IGS) between *rps19* and *rpl2*, whereas that of SSC/IR regions was within the *ndhF* coding region (Figure 1). The repeat analysis revealed 46 records of small single repeat (SSR) that ranged from 10 bp to 16 bp (Table 2). Although six types of SSRs were screened, only mononucleotide and dinucleotide SSRs were found in the cpDNA of *A. crassna*. The mononucleotide SSR (39 records) was more abundant than the dinucleotide SSR (7 records). However, the length of dinucleotide SSR (12–16 bp) was longer than that of mononucleotide (10–13 bp). Most SSRs were located in non-coding regions. Some SSRs were found in coding regions such as *ycf1, rpoC2, rpoB, psbF*, and *cemA*. The content of SSR was mostly composed of A and T nucleotides. Only one mononucleotide and one dinucleotide of SSR contained G and C nucleotides (Table 2).

The comparative genomic analysis revealed high conservation of the chloroplast genome among *Aquilaria* species (Table 3, Table 4). Particularly, the genome size ranged from 174,693 bp (*A. rostrata*) to 175,761 bp (*A. hirta*) and

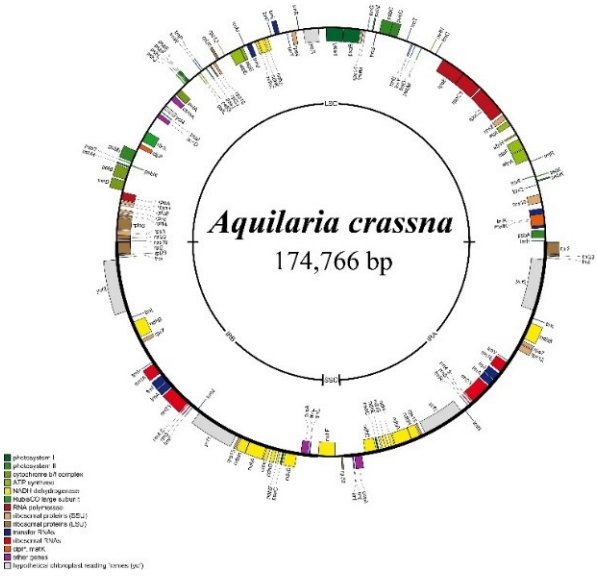

Fig. 1: Chloroplast genome map of *Aquilaria crassna*. Groups of genes with sharing functions would be presented in the same colour. LSC: large single copy; SSC: small single copy; IR: inverted repeat.

contained four parts, including an LSC, an SSC, and two IR regions. The LSC regions were approximately 87 kb in length, whereas the size of the IR regions was 42 kb. The SSC regions were the smallest regions in the cpDNAs of *Aquilaria* species, with 3 kb in length. The GC content of the *Aquilaria* species was constant at 36.7%. Similarly, the same number of genes was found, which included 95 protein-coding genes, 38 *tRNA* genes, and eight *rRNA* genes. Among cpDNAs of *Aquilaria* species, the SSC/IR junction was found within the *ndhF* gene, whereas that of LSC/IR regions was within the *rps9* gene, except for *A. crassna* (in this study, within IGS between the *rps19* and *rpl2* genes) and *A. agallochum* (within the *rpl2* intron). Further investigation of the pairwise identity of cpDNAs among *Aquilaria* species exhibited high similarity, ranging from 99.54% to 99.99% (Table 4). The cpDNA of *A. subintegra* was 99.99% similar to that of *A. crassna* (previously published [12]). Meanwhile, the similarity between *A. rostrata* and *A. hirta*

was 99.54%. The newly sequenced cpDNA of *A. crassna* showed the highest similarity (99.87%) to previously published cpDNAs of *A. crassna* and *A. subintegra*. Although belonging to *A. crassna*, two chloroplast genomes had different locations of the SSC/IR junction (Table 3).

Table 1: List of genes in the chloroplast genome of *Aquilaria crassna*

Gene group	Gene name
RNA ribosome	*rrn4.5 (2x), rrn5 (2x), rrn16 (2x), rrn23 (2x)*
RNA transfer	*trnA-UGC*(2x), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnfM-CAU, trnG-GCC, trnG-UCC*, trnH-GUG, trnI-CAU (2x), trnI-GAU* (2x), trnK-UUU*, trnL-UAA*, trnL-CAA (2x), trnL-UAG (2x), trnM-CAU, trnN-GUU (2x), trnP-UGG, trnQ-UUG, trnR-ACG (2x), trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC (2x), trnV-UAC*, trnW-CCA, trnY-GUA*
ATP synthase	*atpA, atpB, atpE, atpF*, atpH, atpI*
NADH dehydrogenase	*ndhA* (2x), ndhB* (2x), ndhC, ndhD (2x), ndhE (2x), ndhF, ndhG (2x), ndhH (2x), ndhI (2x), ndhJ, ndhK*
Cytochrome	*petA, petB*, petD*, petG, petL, petN*
Photosystem I	*psaA, psaB, psaC (2x), psaI, psaJ, pafI*, pafII*
Photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, pbf1, psbT, psbZ*
Rubisco	*rbcL*
Translational initial factor	*infA*
Large subunit of the ribosome	*rpl2* (2x), rpl14, rpl16*, rpl20, rpl22, rpl23 (2x), rpl32, rpl33, rpl36*
Small subunit of the ribosome	*rps2, rps3, rps4, rps7 (2x), rps8, rps11, rps12* (2x),rps14, rps15 (2x), rps16*, rps18, rps19*
DNA-dependent RNA polymerase	*rpoA, rpoB, rpoC1*, rpoC2*
Acetyl-CoA-carboxylase	*accD*
Cytochrome c	*ccsA (2x)*
Protein capsule	*cemA*
Protease	*clpP1*
Maturase	*matK*
Highly conserved reading frame	*ycf1 (2x), ycf2 (2x)*

**: genes contain intron; (2x): duplicated genes in the IR regions*

The cpDNA sequence of *A. crassna* has a quadripartite structure, as found in other angiosperms [7]. Further comparative analysis revealed similar features regarding genome structure, gene content, and genome size, suggesting a high conservation of cpDNAs within the *Aquilaria* genus. A previous analysis of SSR among eight *Aquilaria* chloroplast genomes revealed the presence of six SSR types located mainly in non-coding regions [19]. A similar result was also found in *A. crassna* in this study, except only two SSR types (mononucleotide and dinucleotide) were found. This finding suggested a unique feature of SSR in *A. crassna* compared to other related taxa. Previously, chloroplast genomes within a genus exhibited different levels of variation. For example, comparative genomic analysis of *Xylocarpus* taxa and related species in Meliaceae revealed gene duplication and gene variations according to different environments, and various SSRs [20]. Similarly, the chloroplast genome of *Sinosenecio* showed differences among LSC/SSC/IR junctions and many highly variable regions that are useful for species identification [21]. In this study, the cpDNAs of twelve *Aquilaria* species exhibited high similarity (above 99.54%). Therefore, further studies covering 27 members of the *Aquilaria* genus are needed to clarify the genetic diversity.

Besides interspecific variations of cpDNAs, the intraspecific differences of cpDNA were also explored. Particularly, genomic data on 17 cpDNAs of *Pseudostellaria heterophylla* revealed seven highly variable regions and ten single nucleotide polymorphisms that can be used for further genomic studies [22]. A similar study about the 12 chloroplast genomes of *Styphnolobium japonicum* provided the necessary information to explore genetic origin and species identification [23]. Although only two cpDNAs of *A. crassna* were analysed in this study, two types of LSC/IR junction were found, suggesting variations of this boundary among *A. crassna* populations. Therefore, further studies based on LSC/IR junctions are needed to develop molecular markers for the origin identification of *A. crassna* from different distributions.

Table 2: Features of SSR in the chloroplast genome of *Aquilaria crassna*

N0	Type	Length	Sequence	Location
1	Dinucleotide	12	TATATATATATA	IGS (*trnR-UCU - atpA*)
2	Dinucleotide	16	TCTCTCTCTCTCTCTC	IGS (*rps2 - rpoC2*)
3	Dinucleotide	14	TATATATATATATA	IGS (*trnY-GUA - trnT-GGU*)
4	Dinucleotide	12	TATATATATATA	IGS (*trnY-GUA - trnT-GGU*)
5	Dinucleotide	12	TATATATATATA	IGS (*petL - psbE*)
6	Dinucleotide	12	TATATATATATA	IGS (*pbfI - psbH*)
7	Dinucleotide	14	TATATATATATATA	IGS (*trnL-UAG - ndhF*)
8	Mononucleotide	10	TTTTTTTTTT	IGS (*trnK-UUU - matK*)
9	Mononucleotide	10	AAAAAAAAAA	IGS (*matK - trnK-UUU*)
10	Mononucleotide	10	AAAAAAAAAA	IGS (*trnK-UUU - rps16*)
11	Mononucleotide	10	AAAAAAAAAA	IGS (*rps16 - trnQ-UUG*)
12	Mononucleotide	10	TTTTTTTTTT	IGS (*rps16 - trnQ-UUG*)
13	Mononucleotide	12	AAAAAAAAAAAA	IGS (*trnQ-UUG - psbK*)
14	Mononucleotide	10	TTTTTTTTTT	IGS (*trnS-GCU - trnG-UCC*)
15	Mononucleotide	10	TTTTTTTTTT	IGS (*trnS-GCU - trnG-UCC*)
16	Mononucleotide	10	TTTTTTTTTT	IGS (*atpF - atpH*)
17	Mononucleotide	11	TTTTTTTTTTT	*rpoC2*
18	Mononucleotide	10	TTTTTTTTTT	*rpoC2*
19	Mononucleotide	10	TTTTTTTTTT	*rpoB*
20	Mononucleotide	10	TTTTTTTTTT	IGS (*petN - psbM*)
21	Mononucleotide	10	TTTTTTTTTT	IGS (*psbM - trnD-GUC*)
22	Mononucleotide	11	AAAAAAAAAAA	IGS (*psbZ - trnG-GCC*)
23	Mononucleotide	10	AAAAAAAAAA	IGS (*trnG-GCC - trnfM-CAU*)
24	Mononucleotide	10	TTTTTTTTTT	IGS (*psaA - pafI*)
25	Mononucleotide	12	AAAAAAAAAAAA	IGS (*psaA - pafI*)
26	Mononucleotide	11	TTTTTTTTTTT	IGS (*psaA - pafI*)
27	Mononucleotide	10	AAAAAAAAAA	IGS (*pafI - trnS-GGA*)
28	Mononucleotide	10	AAAAAAAAAA	IGS (*pafI - trnS-GGA*)
29	Mononucleotide	10	GGGGGGGGGG	IGS (*trnT-UGU - trnL-UAA*)
30	Mononucleotide	10	TTTTTTTTTT	IGS (*trnL-UAA - trnF-GAA*)
31	Mononucleotide	10	AAAAAAAAAA	IGS (*ndhC - trnV-UAC*)
32	Mononucleotide	10	TTTTTTTTTT	IGS (*petG - petL*)
33	Mononucleotide	10	AAAAAAAAAA	IGS (*petL - psbE*)
34	Mononucleotide	10	TTTTTTTTTT	*psbF*
35	Mononucleotide	11	TTTTTTTTTTT	*cemA*
36	Mononucleotide	10	TTTTTTTTTT	IGS (*accD - rbcL*)
37	Mononucleotide	10	TTTTTTTTTT	*petB* intron
38	Mononucleotide	10	TTTTTTTTTT	IGS (*rpl36 - rps8*)
39	Mononucleotide	13	TTTTTTTTTTTTT	IGS (*rps19 - rpl2*)
40	Mononucleotide	12	AAAAAAAAAAAA	IGS (*trnN-GUU - ycf1*)
41	Mononucleotide	10	AAAAAAAAAA	*ycf1*
42	Mononucleotide	10	AAAAAAAAAA	IGS (*ndhF - rpl32*)
43	Mononucleotide	11	AAAAAAAAAAA	IGS (*ndhF - rpl32*)
44	Mononucleotide	10	AAAAAAAAAA	IGS (*ndhF - rpl32*)
45	Mononucleotide	10	TTTTTTTTTT	IGS (*ndhF - rpl32*)
46	Mononucleotide	10	AAAAAAAAAA	IGS (*rpl32 - ndhF*)

Table 3: Features of chloroplast genomes of *Aquilaria* species

Species	Accession number	Total length (bp)	LSC (bp)	SSC (bp)	IR (bp)	%GC	Genes (Protein -coding/ tRNA/ rRNA)	LSC/IR junction	SSC/IR junction
Aquilaria crassna	This study	174,766	87,231	3343	42,096	36,7	95/38/8	IGS (*rps19/rpl2*)	*ndhF* (26 bp)
A. crassna	MN125348	174,830	87,281	3345	42,102	36,7	95/38/8	*rps19* (16 bp)	*ndhF* (26 bp)
A. rugosa	MZ145049	174,893	87,344	3353	42,098	36,7	95/38/8	*rps19* (16 bp)	*ndhF* (28 bp)
A. yunnanensis	MG656407	174,885	87,508	3351	42,013	36,7	95/38/8	*rpl2* exon2	*ndhF* (28 bp)
A. agallochum	MZ145047	174,866	87,293	3355	42,109	36,7	95/38/8	*rps19* (16 bp)	*ndhF* (27 bp)
A. cumingiana	MZ145048	174,834	87,306	3346	42,091	36,7	95/38/8	*rps19* (16 bp)	*ndhF* (26 bp)
A. malaccensis	MH286934	174,832	87,302	3348	42,091	36,7	95/38/8	*rps19* (16 bp)	*ndhF* (26 bp)
A. beccariana	MN125347	174,831	87,301	3348	42,091	36,7	95/38/8	*rps19* (16 bp)	*ndhF* (26 bp)
A. subintegra	MN147871	174,828	87,279	3345	42,102	36,7	95/38/8	*rps19* (16 bp)	*ndhF* (26 bp)
A. microcarpa	MN125350	174,819	87,298	3349	42,086	36,7	95/38/8	*rps19* (16 bp)	*ndhF* (26 bp)
A. hirta	MN125349	175,761	87,221	3344	42,098	36,7	95/38/8	*rps19* (16 bp)	*ndhF* (26 bp)
A. rostrata	MN125351	174,693	87,255	3234	42,102	36,7	95/38/8	*rps19* (16 bp)	*ndhF* (26 bp)
A. sinensis	MN720647	174,914	87,361	3347	42,103	36,7	95/38/8	*rps19* (16 bp)	*ndhF* (28 bp)

Table 4: Identity percentage of 12 *Aquilaria* species

Species	*A. crassna* (This study)	*A. yunnanensis*	*A. malaccensis*	*A. beccariana*	*A. crassna*	*A. hirta*	*A. microcarpa*	*A. rostrata*	*A. subintegra*	*A. sinensis*	*A. agallochum*	*A. cumingiana*	*A. rugosa*
A. crassna (This study)		99.74	99.69	99.69	99.87	99.69	99.68	99.65	99.87	99.72	99.66	99.68	99.74
A. yunnanensis	99.74		99.66	99.66	99.77	99.64	99.67	99.60	99.76	99.89	99.78	99.65	99.97
A. malaccensis	99.69	99.66		99.99	99.72	99.67	99.92	99.58	99.72	99.63	99.59	99.95	99.64
A. beccariana	99.69	99.66	99.99		99.72	99.67	99.92	99.58	99.72	99.63	99.59	99.95	99.64
A. crassna	99.87	99.77	99.72	99.72		99.70	99.71	99.67	99.99	99.73	99.70	99.71	99.75
A. hirta	99.69	99.64	99.67	99.67	99.70		99.66	99.54	99.70	99.64	99.57	99.66	99.66
A. microcarpa	99.68	99.67	99.92	99.92	99.71	99.66		99.58	99.71	99.64	99.59	99.91	99.65
A. rostrata	99.65	99.60	99.58	99.58	99.67	99.54	99.58		99.67	99.55	99.56	99.56	99.58
A. subintegra	99.87	99.76	99.72	99.72	99.99	99.70	99.71	99.67		99.73	99.69	99.71	99.75
A. sinensis	99.72	99.89	99.63	99.63	99.73	99.64	99.64	99.55	99.73		99.77	99.63	99.91
A. agallochum	99.66	99.78	99.59	99.59	99.70	99.57	99.59	99.56	99.69	99.77		99.59	99.76
A. cumingiana	99.68	99.65	99.95	99.95	99.71	99.66	99.91	99.56	99.71	99.63	99.59		99.64
A. rugosa	99.74	99.97	99.64	99.64	99.75	99.66	99.65	99.58	99.75	99.91	99.76	99.64	

V. CONCLUSION

In this study, the complete chloroplast genome of *A. crassna* distributed in Vietnam was sequenced and characterized. The structure and gene content of *A. crassna* were similar to those of other *Aquilaria* species, suggesting high conservation in this genus.

The newly sequenced chloroplast genome of *A. crassna* provided essential data for further studies of genetic conservation and molecular markers for this critically endangered species.

ACKNOWLEDGEMENT

The authors thank Nguyen Tat Thanh Univeristy for providing facilities to conduct this study. Additionally, the authors would like to express their sincere gratitude to

Tra Vinh University for providing opportunity to carry out this research. Also, the authors thank anonymous reviewers for meaningful comments to improve the quality of this research.

REFERENCES

[1] POWO. Aquilaria Lam. *Plant of the World Online*. https://powo.science.kew.org /taxon/urn:lsid:ipni.org:names:331359-2 [Accessed 1st January 2024].

[2] Shivanand P, Arbie NF, Krishnamoorthy S, Ahmad N. Agarwood—The fragrant molecules of a wounded tree. *Molecules*. 2022;27(11): 3386. https://doi.org/10.3390/molecules27113386.

[3] Toan TQ, Dinh TTT, Tran TT, Hoang TB, Pham QL, Nguyen QC, et al. Chemical compositions of essential oil of agarwood (Aquilaria crassna) harvested in Phu Quoc island, Vietnam. *Engineering Chemistry*. 2023;4: 51–57. https://doi.org/10.4028/p-oHM1mQ.

[4] Huy BL, Xuan HN, Tan PN, Thanh LT, Minh TL. *Study about agarwood oil (Aquilaria crassna Pierre) in Phu Quoc Island National Park, Vietnam*. Research Aspects in Biological Science Vol. 1. Book Publisher International; 2022. p. 1–38. https://doi.org/10.9734/bpi/rabs/v1/1795B.

[5] Jok VA, Che Radzi N, Ku Hamid KH. A review: pharmacological properties of Aquilaria spp. *Advanced Materials Research*. 2015;1113: 193–197. https://doi.org/10.4028/www.scientific.net/AMR.1113.193.

[6] Wongwad E, Pingyod C, Saesong T, Waranuch N, Wisuitiprot W, Sritularak B, et al. Assessment of the bioactive components, antioxidant, antiglycation and anti-inflammatory properties of Aquilaria crassna Pierre ex Lecomte leaves. *Industrial Crops and Products*. 2019;138: 111448. https://doi.org/10.1016/j.indcrop.2019.06.011.

[7] Dobrogojski J, Adamiec M, Luciński R. The chloroplast genome: a review. *Acta Physiologiae Plantarum*. 2020;42(6): 98. https://doi.org/10.1007/s11738-020-03089-x.

[8] Yang T, Sahu SK, Yang L, Liu Y, Mu W, Liu X, et al. Comparative analyses of 3,654 plastid genomes unravel insights into evolutionary dynamics and phylogenetic discordance of green plants. *Frontiers in Plant Science*. 2022;13: 808156. https://doi.org/10.3389/fpls.2022.808156.

[9] Dong W, Liu J, Yu J, Wang L, Zhou S. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for dna barcoding. *PLoS One*. 2012;7(4): e35071. https://doi.org/10.1371/journal.pone.0035071.

[10] Hishamuddin MS, Lee SY, Ng WL, Ramlee SI, Lamasudin DU, Mohamed R. Comparison of eight complete chloroplast genomes of the endangered Aquilaria tree species (Thymelaeaceae) and their phylogenetic relationships. *Scientific Reports*. 2020;10: 13034. https://doi.org/10.1038/s41598-020-70030-0.

[11] Deng X, Jiang Z, Jiang Q, Guo W, Li Y, Zhang X. Characterization of the complete chloroplast genome of Aquilaria sinensis, an endangered agarwood-producing tree. *Mitochondrial DNA Part B*. 2020;5(1): 422–423. https://doi.org/10.1080/23802359.2019.1703593.

[12] Li GD, Rao PY, Guo JL, Zhang YH. The complete chloroplast genome of a critically endangered agarwood tree, Aquilaria crassna (Thymelaeaceae). *Mitochondrial DNA Part B*. 2019;4(1): 1810–1811. https://doi.org/10.1080/23802359.2019.1613200.

[13] Wang Y, Zhan DF, Jia X, Mei WL, Dai HF, Chen XT, et al. Complete chloroplast genome sequence of Aquilaria sinensis (Lour.) gilg and evolution analysis within the malvales order. *Frontiers in Plant Science*. 2016;7: 280. https://doi.org/10.3389/fpls.2016.00280.

[14] Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biology*. 2019;20: 129. https://doi.org/10.1186/s13059-019-1727-y.

[15] Vaser R, Šikić M. Time- and memory-efficient genome assembly with Raven. Nature Computational Science. 2021;1(5): 332–336. https://doi.org/10.1038/s43588-021-00073-4.

[16] Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Research*. 2017;45: W6–W11. https://doi.org/10.1093/nar/gkx391.

[17] Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research*. 2019;47: W59–64. https://doi.org/10.1093/nar/gkz238.

[18] Kraemer L, Beszteri B, Gäbler-Schwarz S, Held C, Leese F, Mayer C, et al. STAMP: Extensions to the STADEN sequence analysis package for high throughput interactive microsatellite marker design. *BMC Bioinformatics*. 2009;10: 233–241. https://doi.org/10.1186/1471-2105-10-41.

[19] Hishamuddin M, Samsuddin A, Mohamed R. Mining and analysis of chloroplast simple sequence repeats (SSRs) from eight species of Aquilaria. *Turkish Journal of Botany*. 2021;46(1): 26–36. https://doi.org/10.3906/bot-2108-61.

[20] Tan F, Li W, Feng H, Huang Y, Banerjee AK. Interspecific variation and phylogenetic relationship between mangrove and non-mangrove species of a same family (Meliaceae)—insights from comparative analysis of complete chloroplast genome. *PeerJ*. 2023;11: e15527. https://doi.org/10.7717/peerj.15527.

[21] Peng JY, Zhang XS, Zhang DG, Wang Y, Deng T, Huang XH, et al. Newly reported chloro-

plast genome of Sinosenecio albonervius Y. Liu & QE Yang and comparative analyses with other Sinosenecio species. *BMC Genomics*. 2022;23(1): 639. https://doi.org/10.1186/s12864-022-08872-3.

[22] Zhang W, Zhang Z, Liu B, Chen J, Zhao Y, Huang Y. Comparative analysis of 17 complete chloroplast genomes reveals intraspecific variation and relationships among Pseudostellaria heterophylla (Miq.) Pax populations. *Frontiers in Plant Science*. 2023;14: 1163325. https://doi.org/10.3389/fpls.2023.1163325.

[23] Mu Z, Zhang Y, Zhang B, Cheng Y, Shang F, Wang H. Intraspecific chloroplast genome variation and domestication origins of major cultivars of Styphnolobium japonicum. *Genes (Basel)*. 2023;14(6): 1156. https://doi.org/10.3390/genes14061156.