

The complete mitochondrial genome sequence of *Aquilaria sinensis*

Zheng-Feng Wang^{a,b,c}  and Hong-Lin Cao^{a,b,c}

^aKey Laboratory of Vegetation Restoration and Management of Degraded Ecosystems, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China; ^bCenter for Plant Ecology, Core Botanical Gardens, Chinese Academy of Sciences, Guangzhou, China; ^cSouthern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou, China

ABSTRACT

Aquilaria sinensis, endemic to China, is an economically important evergreen tree species and a source of agarwood. Due to the high market demand for agarwood, this species is heavily overexploited in the wild and is now listed as an endangered species. Although its nuclear and chloroplast genomes have been previously reported, little is known about its mitochondrial genome. Using the paired-end short reads generated by the Illumina sequencing platform, we assembled and herein report the mitochondrial genome of *A. sinensis* for future phylogenetic, evolutionary, and preservative studies. The length of the *A. sinensis* mitochondrial genome was found to be 341,829 bp and the GC content was 45.01%. A total of 32 protein-coding genes, 19 tRNA genes, and three rRNA genes were annotated. The phylogenetic tree indicated that *A. sinensis* is most closely genetically related to *Vigna radiata*.

ARTICLE HISTORY

Received 7 October 2020
Accepted 19 December 2020

KEYWORDS

Aquilaria sinensis; genome assembly; high-throughput sequencing; mitochondrion

Aquilaria (Thymelaeaceae) species are evergreen trees that are distributed from South to Southeast Asia. These trees are best known for the production of agarwood, which is a resinous heartwood used in traditional medicine, incense, and perfume, as well as for decorations and art pieces. There are 21 species in this genus, 13 of which are able to produce agarwood (Wang et al. 2020). An agarwood is produced in the trunk and main branches as a result of wounding, such as by worm, microbial, or fungal infections, and is harvested by felling or cutting down the tree. Therefore, natural agarwood is becoming more rare due to illegal logging. All species of *Aquilaria* are currently listed as vulnerable to critically endangered on the International Union for Conservation of Nature (IUCN) Red List. *Aquilaria sinensis* (Lour.) Spreng is endemic to South China. It is the only species to produce agarwood used as a crude medicine in China (Wang et al. 2020). Although its nuclear and chloroplast genomes have been previously reported (Wang et al. 2016, 2020), no complete mitochondrial genomic resources are currently available.

The study of plant mitochondria has a long history. The first research dates back to 1904 (Ernster and Schatz 1981), with the first plant mitochondrial genome reported in 1992 (Oda et al. 1992) for *Marchantia polymorpha*. In recent decades, with the development of next generation sequencing technologies, whole genomes, including nuclear and organelle genomes in variable organisms, were able to be fully sequenced and assembled (Wang et al. 2016, 2020; Chen et al. 2020; Ding et al. 2020; Kim et al. 2020; Liu et al. 2020; Miao et al. 2020; Shao et al. 2020; Yang et al. 2020). However, because plant mitochondrial genomes are much

more variable in structure and size (Mackenzie and McIntosh 1999) than chloroplast genomes, fewer plant mitochondrial genomes have been assembled and reported compared to chloroplast genomes and their animal counterparts. In plants, the mitochondrion is an important organelle involved in respiration and metabolism (Mackenzie and McIntosh 1999). Therefore, in this study, we report the complete mitochondrial genomic sequence of *A. sinensis* to provide a resource for future effective conservation and evolutionary studies.

Leaf samples were collected from Dianbai County, China (21°38'27.21"N, 111°15'14.52"E) and a voucher specimen was deposited in the Herbarium of the South China Botanical Garden (no. IBSC-T-201709Y14). From this sample, approximately 160 Gb whole genome sequencing reads were generated with the Illumina HiSeq X Ten (Illumina, San Diego, CA) platform using the paired-end (2 × 150 bp) strategy. Using 14 previously reported mitochondrial genomes (Figure 1) as references, reads related to mitochondrial sequences were extracted using GetOrganelle 1.7.1 (Jin et al. 2020). The extracted reads were used to assemble the mitochondrial genome by NOVOPlasty 4.2.1 (Dierckxsens et al. 2017) using the parameter 'mito_plant'. After assembly, the genome was polished by NextPolish 1.3.0 (Hu et al. 2020) and then annotated by GeSeq (Tillich et al. 2017) and PGA (Qu et al. 2019). The annotated mitochondrial genome was deposited in GenBank with the accession number MW057771. Microsatellites across the mitochondrial genome were determined by MISA-web (Sebastian et al. 2017) using its default parameter setting. The phylogenetic tree for *A. sinensis* and another 14 species of the Malvales order (Figure 1) was

CONTACT Hong-Lin Cao  caohl@scib.ac.cn  South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, Guangdong 510650, China

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

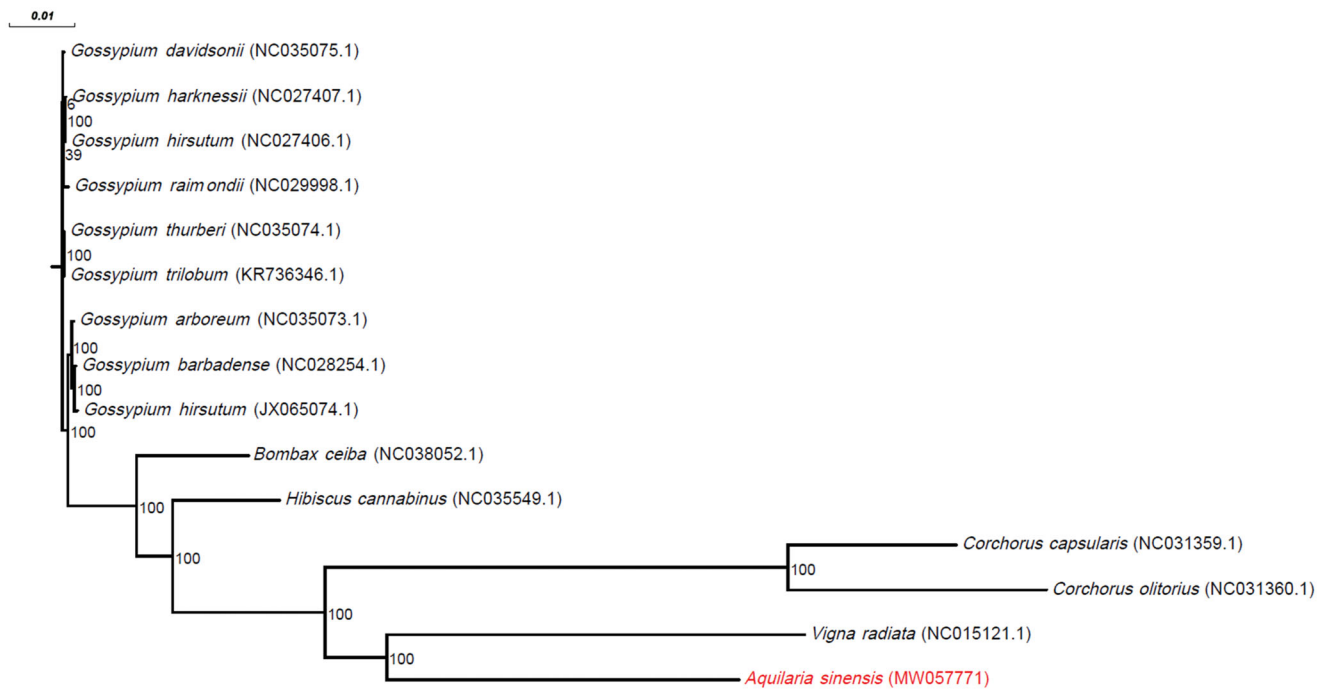


Figure 1. Phylogenetic tree from the complete mitochondrial genome sequences of *Aquilaria sinensis* and 14 additional species. GenBank accession numbers are denoted in parentheses. Bootstrap percentages (1000 replicates) are shown at nodes.

generated using mashtree v1.2.0 (Katz et al. 2019). For those genes shared by *A. sinensis* and another species, the nucleotide diversity (Pi) value was calculated using MEGA-X v10.1.8 (Kumar et al. 2018).

The length of the complete mitochondrial genome of *A. sinensis* was 341,829 bp. The base composition of the mitochondrial genome was 27.45% T, 27.54% A, 22.44% C, and 22.57% G, with a GC content of 45.01%. The base composition of the *A. sinensis* mitochondrial genome was similar to the other 14 Malvales species (Table S1), with the GC content ranging from 42.85% to 45.11%. Annotation analysis revealed a total of 32 protein-coding genes, 19 transfer RNA genes, and three ribosomal RNA genes in the *A. sinensis* mitochondrial genome. The length of protein-coding genes ranged from 81 to 2826 bp, with a mean of 896.938 bp, the exon length ranged from 81 to 1968 bp, with a mean of 720.968 bp, and the intron length ranged from 777 to 1389 bp, with a mean of 1035.667 bp (Table S2). Compared to the other species, *A. sinensis* had the shortest mean gene, exon, and intron lengths. A total of 44 microsatellites were identified in the *A. sinensis* mitochondrial genome (Table S3). Among them, the majority were mononucleotide-repeat microsatellites (79.54%). There were only seven protein-coding genes shared between *A. sinensis* and the other species studied. The Pi values ranged from 0.009437 for the *ccmB* gene to 0.076923 for the *rpl16* gene (Table S4). The phylogenetic tree showed that the mitochondrial genome of *A. sinensis* was closely related to that of *Vigna radiata* (Figure 1).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Science and Technology Project of Guangdong Province under Grant number [2017A030303060]; Overall Planning of 'One Center and Three Bases' for Rare and Endangered Plants Preservation in Guangdong Province and Inventory of Endemic Plants and Important Wild South-China Herbal Medicinal Resources in Three Regions; Key Special Project for Introduced Talents Team of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) under Grant number [GML2019ZD0408].

ORCID

Zheng-Feng Wang  <http://orcid.org/0000-0002-7599-7891>

Data availability statement

The complete mitochondria genome sequences of *Aquilaria sinensis* have been deposited in GenBank under the accession numbers MW057771 and is also accessible at <https://doi.org/10.13140/RG.2.2.27460.65925>. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA532686, SRR8892930, and SAMN11412194, respectively.

References

- Chen X, Yuan Z, Li C, Song Y. 2020. Complete mitochondrial genome sequence of *Empoasca dwalata* (Hemiptera: Cicadellidae: Typhlocybinae). *Mitochondrial DNA Part B*. 5(3):2260–2261.
- Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res*. 45(4):e18.
- Ding Z, Cui H, Zhu Q, Wu Y, Zhang T, Qiu B, Hao P. 2020. Complete sequence of mitochondrial genome of *Cucumis melo* L. *Mitochondrial DNA Part B*. 5(3):3194–3195.
- Ernster L, Schatz G. 1981. Mitochondria: a historical review. *J Cell Biol*. 91(3 Pt 2):227s–255s.

- Hu J, Fan J, Sun Z, Liu S. 2020. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics*. 36(7):2253–2255.
- Jin J-J, Yu W-B, Yang J-B, Song Y, dePamphilis CW, Yi T-S, Li D-Z. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol*. 21(1):241.
- Katz LS, Griswold T, Morrison SS, Caravas JA, Zhang S, den Bakker HC, Deng X, Carleton HA. 2019. Mashtree: a rapid comparison of whole genome sequence files. *J Open Source Softw*. 4(44):1762.
- Kim C-K, Jin M-W, Kim Y-K. 2020. The complete mitochondrial genome sequences of *Bupleurum falcatum* (Apiales: Apiaceae). *Mitochondrial DNA Part B*. 5(3):2576–2577.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 35(6):1547–1549.
- Liu F, Movahedi A, Yang W, Xu L, Xie J, Zhang Y. 2020. The complete chloroplast genome and characteristics analysis of *Callistemon rigidus* R.Br. *Mol Biol Rep*. 47(7):5013–5024.
- Mackenzie S, McIntosh L. 1999. Higher plant mitochondria. *Plant Cell*. 11(4):571–585.
- Miao K, Liao M, Li B, Zhang S, Wang Y, Zhang Z, Rong X, Yu Y. 2020. Complete mitochondrial genome sequences of sea slug *Melanochlamys* sp. (Cephalaspidae: Aglajidae). *Mitochondrial DNA Part B*. 5(3):2382–2383.
- Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, Akashi K, Kanegae T, Ogura Y, Kohchi T, et al. 1992. Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. A primitive form of plant mitochondrial genome. *J Mol Biol*. 223(1):1–7.
- Qu XJ, Moore MJ, Li DZ, Yi TS. 2019. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods*. 15:50.
- Sebastian B, Thomas T, Thomas M, Uwe S, Martin M. 2017. MISA-web: a web server for microsatellite prediction. *Bioinformatics*. 33:2583–2585.
- Shao S-C, Luo Y, Yu W-B. 2020. The complete mitochondrial genome sequence of *Lactarius trivialis* (Russulales, Basidiomycota). *Mitochondrial DNA Part B*. 5(3):2078–2079.
- Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Res*. 45(W1):W6–W11.
- Wang Z-F, Cao H-L, Cai C-X, Wang Z-M. 2020. Using genetic markers to identify the origin of illegally traded agarwood-producing *Aquilaria sinensis* trees. *Global Ecol Conserv*. 22:e00958.
- Wang Y, Zhan D-F, Jia X, Mei W-L, Dai H-F, Chen X-T, Peng S-Q. 2016. Complete chloroplast genome sequence of *Aquilaria sinensis* (Lour.) Gilg and evolution analysis within the Malvales order. *Front Plant Sci*. 7:280.
- Yang SJ, Hou R, Wu K, Liu P, Chen Y, Yang W, Chen P. 2020. The complete mitochondrial genome sequence of *Cletus rubidiventris* (Heteroptera: Coreidae). *Mitochondrial DNA Part B*. 5(3):3093–3094.