

# Insights into *Aquilaria* phylogenetics through comparative plastomic resources

Junhu Kan<sup>1,2,3#</sup>, Liyun Nie<sup>2,3#</sup>, Zenglu Mi<sup>2#</sup>, Xiaojin Liu<sup>1</sup>, Daping Xu<sup>1</sup>, Luke R Tembrock<sup>4</sup>, Zhiqiang Wu<sup>2,3\*</sup> and Zhou Hong<sup>1\*</sup>

<sup>1</sup> Research Institute of Tropical Forestry, Chinese Academy of Forestry, Guangzhou 510520, China

<sup>2</sup> Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China

<sup>3</sup> School of Medical, Molecular and Forensic Sciences, Murdoch University, Murdoch, WA 6149, Australia

<sup>4</sup> Department of Agricultural Biology, College of Agricultural Sciences, Colorado State University, Fort Collins, Colorado 80523, United States

# Authors contributed equally: Junhu Kan, Liyun Nie, Zenglu Mi

\* Corresponding authors, E-mail: [wuzhiqiang@caas.cn](mailto:wuzhiqiang@caas.cn); [hzhou1981@sina.com](mailto:hzhou1981@sina.com)

## Abstract

The plastid is an essential organelle for its role in photosynthesis and energy production and its genomic information is always employed as important evolutionary markers to explore the relationship among species. Agarwood (*Aquilaria*), prized for its aromatic blend, finds extensive use in various cultures as incense and perfume. Despite its high economic importance, the phylogenetic status among *Aquilaria* based on plastomes remains ambiguous due to the lack of available plastomic resources. To bridge this knowledge gap, 22 *Aquilaria* plastomes were newly sequenced, similar variation patterns in this genus were determined, including a shared 16 bp extension of the *rps19* gene and seven highly variable regions. The analysis highlighted the highest prevalence of the A/T motif among simple sequence repeats in these plastomes. Further phylogenetic analysis revealed *Aquilaria*'s phylogenetic implications with an expanded dataset. This comprehensive plastomic resource not only enhances our understanding of *Aquilaria* evolution but also presents potential molecular markers for DNA barcoding.

**Citation:** Kan J, Nie L, Mi Z, Liu X, Xu D, et al. 2024. Insights into *Aquilaria* phylogenetics through comparative plastomic resources. *Forestry Research* 4: e030 <https://doi.org/10.48130/forres-0024-0028>

## Introduction

Agarwood, a precious aromatic blend, is produced through natural or induced injury to plants within the *Aquilaria* or *Gyrinops* genus of the Thymelaeaceae family<sup>[1]</sup>. As sap seeps from tree wounds mixed with the white wood, a gradual transformation occurs, leading to the development of a distinctive yellow-brown or black-brown wood-oil mixture, thus giving rise to agarwood<sup>[2]</sup>. Renowned for its extraordinary fragrance, agarwood is widely utilized as both incense and perfume across numerous cultures<sup>[3,4]</sup>. Agarwood from various varieties exhibits distinct medicinal properties, boasting significant pharmacological effects, clinical efficacy, and health benefits<sup>[5–8]</sup>. Its exorbitant value has spurred numerous international trades, with annual trading volumes reaching hundreds of tons and trading scales exceeding millions of dollars<sup>[9–11]</sup>. However, its heightened demand has fueled illegal logging and non-selective felling, exacerbated by the scarcity of naturally occurring agarwood trees in the wild<sup>[12,13]</sup>. At present, Appendix II of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) lists 23 species of *Aquilaria* plants and eight species of *Gyrinops* plants, distributed across countries such as China, Indonesia, Malaysia, India, the Philippines, Cambodia, Vietnam, Laos, Thailand, Papua New Guinea, and Singapore<sup>[14]</sup>. Currently, the agarwood circulating in the market is primarily sourced from several agarwood plants, notably *A. sinensis*, *A. malaccensis*, and *A. crassna*<sup>[15]</sup>.

When the morphological differences are evident, utilizing such markers offer advantages in terms of easy recognition and understanding, as well as simplicity and intuitiveness. However, since commercially traded agarwood primarily exists in the form of traditional Chinese patent medicines, simple preparations, and wood, often devoid of key morphological components (especially flowers and fruits), there arises a need for an effective method to distinguish between different agarwood species. Plastids, organelles within plant cells, play important roles in photosynthesis and energy generation<sup>[16–18]</sup>. Typically, the plastome presents a conserved quadripartite circular structure, consisting of two single-copy (SC) regions [large single-copy (LSC) and small single-copy (SSC) regions] and two inverted repeat (IR) regions<sup>[19–22]</sup>. Plastomes serve as effective tools for unveiling phylogenetic relationships among plants<sup>[23–27]</sup>. Despite the rapid advancement of high-throughput sequencing technologies, which has resulted in the release of thousands of plastomes in public databases, the number of plastomes from the *Aquilaria* genus remains limited. This scarcity of plastomic data has impeded progress in understanding the evolutionary status of *Aquilaria*.

Although many previous studies have attempted to determine the phylogenetic status of *Aquilaria*, the lack of comprehensive genetic data has limited their efficacy. For instance, Farah et al.<sup>[28]</sup> utilized five chloroplast loci (*matK*, *rbcl*, *trnL* intron, *trnL-trnF*, and *psbC-trnS*), along with the internal transcribed spacer (ITS) region, in their analysis of *Aquilaria*.

However, discrepancies emerged between the phylogenetic trees constructed from these two datasets. Notably, while *A. agallocha* was identified as the sister species to *A. sinensis* based on chloroplast loci, the ITS data revealed a different scenario, with *A. sinensis* clustering together with *A. yunnanensis*. A recent study<sup>[29]</sup> further revealed a distinct phylogenetic relationship among *Aquilaria* species, which expanded the analysis by including the plastome of *A. rugosa*, resulting in notable changes to the relationships among *A. sinensis*, *A. yunnanensis*, and *A. agallocha*. Additionally, the phylogenetic status of *A. crassna* appeared ambiguous, as different varieties of this species did not cluster together. Specifically, *A. crassna* (MK779998) exhibited a closer phylogenetic relationship with *A. subintegra* (MN147871) than to *A. crassna* (MN125348) varieties. Given these debates and uncertainties, there is a pressing need for a more comprehensive phylogenetic study to provide novel insights into the evolutionary relationships within the *Aquilaria* genus.

In this study, novel sequencing of 22 complete plastomes of *Aquilaria* was conducted, which were combined with 15 published *Aquilaria* plastomes. Subsequently, this dataset was utilized to conduct comparative genomic and phylogenetic analyses, aiming to achieve three primary objectives: (1) addressing the gap of insufficient plastomic resources within the *Aquilaria* genus and gain deeper insights into its evolutionary patterns; (2) identification of highly variable regions (HVRs) suitable for DNA barcode-based identification; and (3) performing a phylogenetic analysis to infer the relationships among *Aquilaria* species using an expanded dataset.

## Materials and methods

### Material collection, DNA extraction, and sequencing

Fresh leaves from five *A. sinensis* samples were collected from the Xishuangbanna Tropical Botanical Garden of the Chinese Academy of Sciences (21°55'55" N, 101°15' 40" E). Additionally, eight *A. sinensis* samples were collected from Jianfeng, Hainan, China (19°47'48" N, 109°48'37" E), while three samples of *A. sinensis* were collected from the Guangdong Academy of Forestry (23°11'38" N, 113°22'52" E). Moreover, four *A. crassna* and two *A. rugosa* samples were obtained from the Agricultural Genomics Institute, Chinese Academy of Agricultural Sciences (22°32'44" N, 114°03'10" E). Genomic DNA extraction from plant samples was conducted using the cetyl trimethylammonium bromide (CTAB) method<sup>[30]</sup>, followed by the sequencing process by BGI Genomics. For library construction, the Illumina TruSeq DNA PCR-Free Library Prep Kit was utilized, with subsequent sequencing carried out using Illumina HiSeq X Ten system.

### Plastome assembly and genome annotation

For the *de novo* assembly of the 22 newly sequenced plastomes, GetOrganelle v.1.6.4 was utilized<sup>[31]</sup>, with a complete plastome of *A. sinensis* (MN720647) serving as the reference. Genome annotation was conducted using CPGAVAS2<sup>[32]</sup>, and the annotation results were subsequently verified manually. The annotation files were uploaded in Figshare (doi:10.6084/m9.figshare.25713024). Finally, the circular maps of the *Aquilaria* plastomes were generated using Chloroplast<sup>[20]</sup>, with visualization facilitated by OGDRAW<sup>[33]</sup>.

## Genome comparison and variation analysis

The analysis of the SC and IR borders among the 22 *Aquilaria* plastomes was conducted using CPJSDraw v1.0<sup>[34]</sup> to visualize the expansion and contraction of IRs. Gene rearrangement within these plastomes was examined using Mauve v.2.4.0<sup>[35]</sup>. To facilitate comparative analyses, 15 complete *Aquilaria* plastomes were downloaded from GenBank and reannotated (Supplemental Table S1). Subsequently, the mVISTA<sup>[36]</sup> tool was employed to visualize the similarity among the 37 *Aquilaria* plastomes. Default parameters were utilized for plastome alignment in Shuffle-LAGAN mode. The number of polymorphic sites and nucleotide variability ( $\pi$ ) were further evaluated using DnaSP v.6.12<sup>[37]</sup>, employing a step size of 200 bp and a window length of 600 bp.

## Repeat element analysis

REPuter (<https://bibiserv.cebitec.uni-bielefeld.de/reputer>) online program was employed<sup>[38]</sup> to detect large sequence repeats (LSRs), including forward (F), reverse (R), complement (C), and palindrome (P) sequence repeats. Parameters were set for sequences of  $n \geq 30$  bp, a sequence identity of  $\geq 90\%$ , and a Hamming distance of 3. Additionally, simple sequence repeats (SSRs) were identified using the MISA-web (<https://webblast.ipk-gatersleben.de/misa/>) online program<sup>[39]</sup>. Thresholds of seven for mononucleotide repeats, four for dinucleotide repeats, and three for tri-, tetra-, and penta-nucleotide repeats were utilized.

## Evolutionary rates and phylogenetic analysis

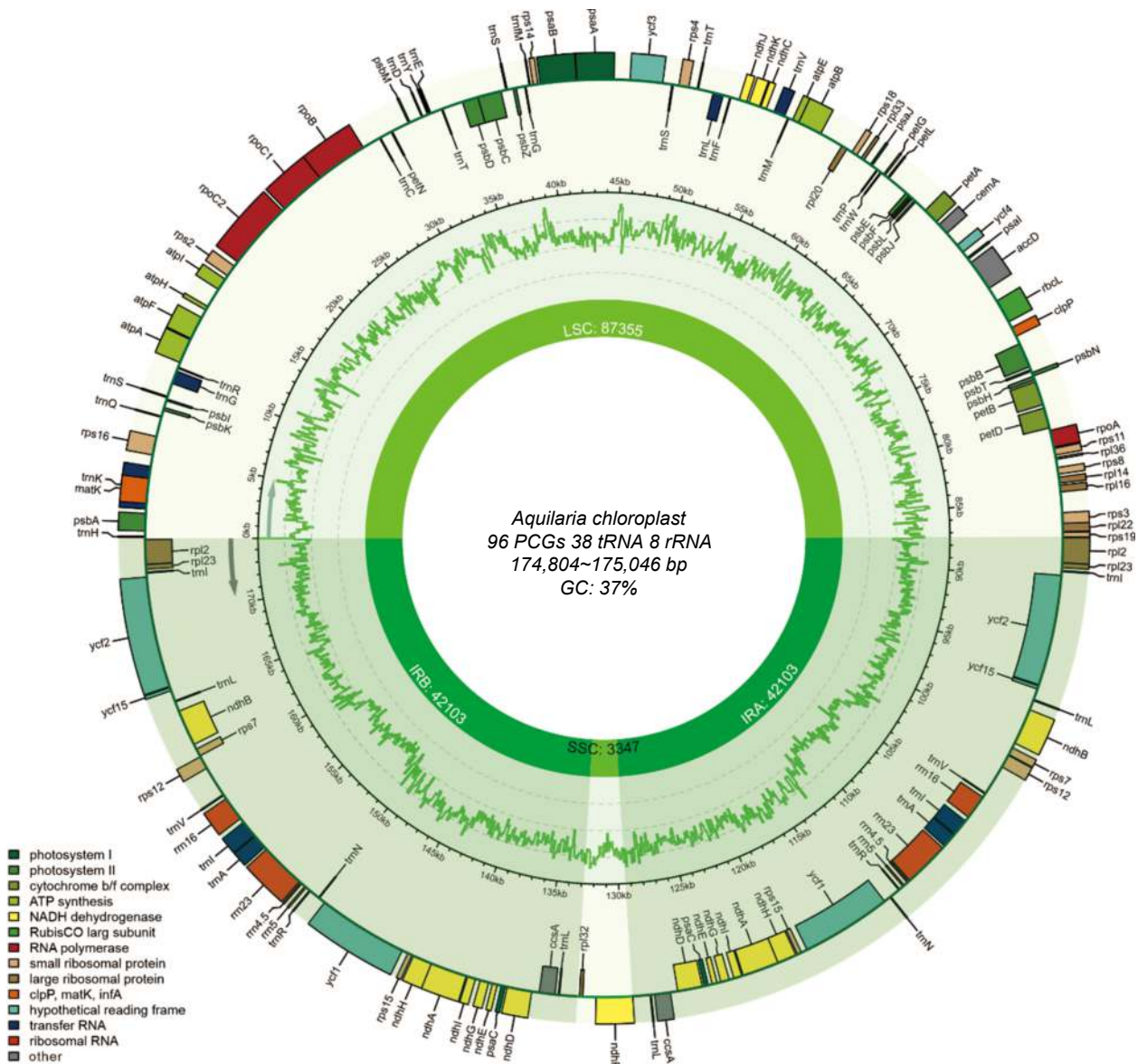
In DAMABE v7.3.11<sup>[40]</sup>, 79 protein-coding genes (PCGs) were extracted from 37 plastomes, followed by alignment using MAFFT software (v.7.487)<sup>[41]</sup>. Subsequently, the CodeML in PAML v.4.9<sup>[42]</sup> was employed to estimate rates of synonymous (Ks) and non-synonymous (Ka) substitution, along with their ratio (Ka/Ks), utilizing a one-ratio model.

Phylogenetic relationships among 37 *Aquilaria* samples were inferred based on complete plastomes with maximum likelihood (ML) and Bayesian inference (BI) methods. To root the trees, three samples from the *Daphne* genus were selected as outgroups. For the ML analysis, RAXML v.8.2.12<sup>[43]</sup> with the GTRCAT model was employed, performing 1,000 bootstrap replicates to assess node support. Convergence was evaluated using the '-l autoMRE' parameter. For the BI analysis, MrBayes 3.2.7a<sup>[44]</sup> was utilized. The best models for nucleotide substitutions were selected using ModelTest-NG<sup>[45]</sup> based on the Bayesian information criterion (BIC). Two trials with four independent Markov chains were run for 10 million generations each, and convergence was verified using Tracer v.1.7.1<sup>[46]</sup>.

## Results

### Plastome genome structure and organization

In this study, the structural features and plastome organizations of 22 newly sequenced *Aquilaria* plastomes were investigated (Fig. 1, Supplemental Table S1). The length of these plastomes ranged from 174,804 to 175,046 bp, with an overall GC content ranging between 36.70% and 36.71%. Each plastome exhibited a typical quadripartite structure, comprising IRs ranging from 42,095 to 42,172 bp, an SSC region ranging from 3,344 to 3,347 bp and an LSC region ranging from 87,270 to 87,355 bp in size.



**Fig. 1** Annotation map of the *Aquilaria* plastome. Genes listed inside and outside the circle are transcribed clockwise and counterclockwise, respectively, with genes color-coded based on functional classification.

The plastomes of these *Aquilaria* samples encoded an identical set of 142 genes (Table 1), comprising 96 PCGs, 38 transfer RNA (tRNA) genes, and eight ribosomal RNA (rRNA) genes. Among them, 64 single-copy PCGs were identified, with 16 PCGs located in the IR regions. Additionally, four rRNA genes (*rrn4.5*, *rrn23*, *rrn5*, and *rrn16*) and eight tRNA genes (*trnA-UGC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnL-UAG*, *trnN-GUU*, *trnR-ACG*, and *trnV-GAC*) were found to have two copies in IR regions. Introns were detected in 10 PCGs and six tRNA genes. Notably, two PCGs, *ycf3*, and trans-spliced *rps12* (characterized by the first exon located in the LSC region and the other two in the IR regions) contained two introns. These findings underscore a highly conserved genome structure and gene content across *Aquilaria* plastomes.

**Comparative genomics analysis and identification of highly variable regions**

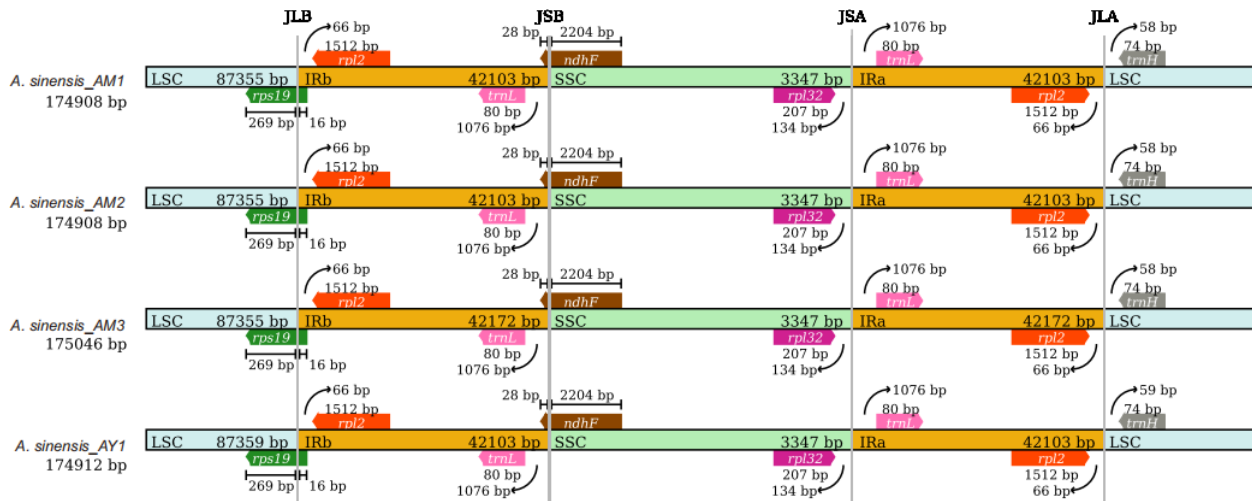
The Mauve alignment results (Supplemental Fig. S1) revealed no gene rearrangements in these newly sequenced plastomes,

with the plastome of *A. malaccensis* serving as the reference. To further explore structure variations within the *Aquilaria* genus, 15 *Aquilaria* plastomes were downloaded from GenBank for comparative analysis. The junction sites on the boundaries of genomic regions in *Aquilaria* species, namely LSC/IRb, SSC/IRb, SSC/IRa, and LSC/IRa, are pivotal in chloroplast genome evolution. These junctions were determined using CPJSDraw. As depicted in Fig. 2 (or Supplemental Fig. S2), the boundaries of genomic regions in *Aquilaria* samples are highly conserved, with only minor variations observed in the junctions and adjacent genes. The *rps19* gene was found across both the LSC and IRb regions, with most of its sequence situated in the LSC region. Furthermore, a 16 bp extension of this gene was consistently observed in the IRb region across all *Aquilaria* plastomes. This 16 bp extension could potentially serve as a genus-special marker for *Aquilaria*. Moreover, it was observed that the IRb/SSC junction was located within the *ndhF* gene. Notably, no genes were found spanning the SSC/IRa and IRa/LSC junctions.

**Table 1.** Genes present within *Aquilaria* plastomes.

Category	Group of genes	Names of genes
Photosynthesis	Subunits of photosystem I	<i>psaA, psaB, psaC(2), psal, psaJ</i>
	Subunits of photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
	Subunits of NADH dehydrogenase	<i>ndhA*(2), ndhB*(2), ndhC, ndhD(2), ndhE(2), ndhF, ndhG(2), ndhH(2), ndhI(2), ndhJ, ndhK</i>
	Subunits of cytochrome b/f complex	<i>petA, petB*, petD*, petG, petL, petN</i>
	Subunits of ATP synthase	<i>atpA, atpB, atpE, atpF*, atpH, atpI</i>
	Large subunit of rubisco	<i>rbcl</i>
	Subunits photochlorophyllide reductase	–
Self-replication	Proteins of large ribosomal subunit	<i>rpl14, rpl16, rpl2*(2), rpl20, rpl22, rpl23(2), rpl32, rpl33, rpl36</i>
	Proteins of small ribosomal subunit	<i>rps11, rps12**(2), rps14, rps15(2), rps16*, rps18, rps19, rps2, rps3, rps4, rps7(2), rps8</i>
	Subunits of RNA polymerase	<i>rpoA, rpoB, rpoC1*, rpoC2</i>
	Ribosomal RNAs	<i>rrn16(2), rrn23(2), rrn4.5(2), rrn5(2)</i>
	Transfer RNAs	<i>trnA-UGC*(2), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnG-UCC*, trnH-GUG, trnI-CAU(2), trnI-GAU*(2), trnK-UUU*, trnL-CAA(2), trnL-UAA*, trnL-UAG(2), trnM-CAU, trnN-GUU(2), trnP-UGG, trnQ-UUG, trnR-ACG(2), trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC(2), trnV-UAC*, trnW-CCA, trnY-GUA, trnM-CAU</i>
	Other genes	<i>matK</i>
Other genes	Maturase	<i>matK</i>
	Protease	<i>clpP</i>
	Envelope membrane protein	<i>cemA</i>
	Acetyl-CoA carboxylase	<i>accD</i>
	c-type cytochrome synthesis gene	<i>ccsA</i>
	Translation initiation factor	–
	other	–
Genes of unknown function	Conserved hypothetical chloroplast ORF	<i>ycf1(2), ycf15(2), ycf2(2), ycf3**, ycf4</i>

Gene\*: Gene with one intron; Gene\*\*: Gene with two introns; Gene(2): Number of copies of multi-copy genes.

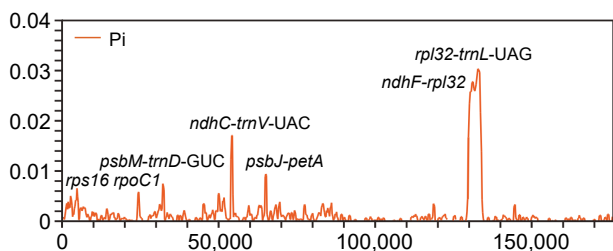


**Fig. 2** Comparison of the large single-copy (LSC), small single-copy (SSC), and inverted repeat (IR) region borders of *Aquilaria* plastomes. The LSC, IR, and SSC regions are depicted with blue, orange, and green blocks, respectively. Gene boxes above the block are transcribed counterclockwise while those below the block are transcribed clockwise. (Representative plastomes from Supplemental Fig. S2).

The overall organization of the plastomes was analyzed using mVISTA, revealing a high degree of conservation among these *Aquilaria* plastomes (Supplemental Fig. S3). Interestingly, the SC regions exhibited higher divergence compared to the IR regions. Furthermore, it was observed that PCGs displayed a lower level of variation when compared to non-coding regions.

Furthermore, we identified some variant hotspot regions, primarily comprising intergenic regions and a few gene coding regions such as *rps16* and *rpoC1*. The nucleotide variability ( $\pi$ ) values across the analyzed coding and intergenic sequences of the 37 chloroplast genomes ranged from 0 to 0.03037 (Fig. 3).

Notably, variation in the SSC region was higher than that in the IR region, consistent with the mVISTA results. Seven regions with  $\pi$  values higher than 0.005, including *rps16*, *rpoC1*, *psbM-trnD-GUC*, *ndhC-trnV-UAC*, *psbJ-petA*, *ndhF-rpl32*, and *rpl32-trnL-UAG*, were detected. These regions were identified as HVRs in this study. Among them, the coding region of the *rps16* gene exhibited a  $\pi$  value exceeding 0.04, while the *rpl32-trnL-UAG* inter-region possessed the most informative sites. These HVRs have the potential to serve as barcodes for identifying *Aquilaria* species.



**Fig. 3** Comparison of nucleotide variability values among 37 plastomes using window sliding analysis. The x-axis indicates the position of the midpoint of the window, while the y-axis indicates the nucleotide diversity of each window.

### Repeat elements analysis

Nucleotide repeats within chloroplast genomes play a crucial role in plant typing and are widely used as genetic molecular markers in population genetics. Large sequence repeats (LSRs) were identified as repeats with a length of  $\geq 30$  bp each. A total of 4,765 LSRs were detected across the 37 plastomes, encompassing palindromic repeats (P), forward repeats (F), reverse repeats (R) and complementary repeats (C) (Fig. 4a). Among these, forward repeats were the most prevalent type, with 2,533 occurrences, while complementary repeats were the least common, totaling 46 occurrences.

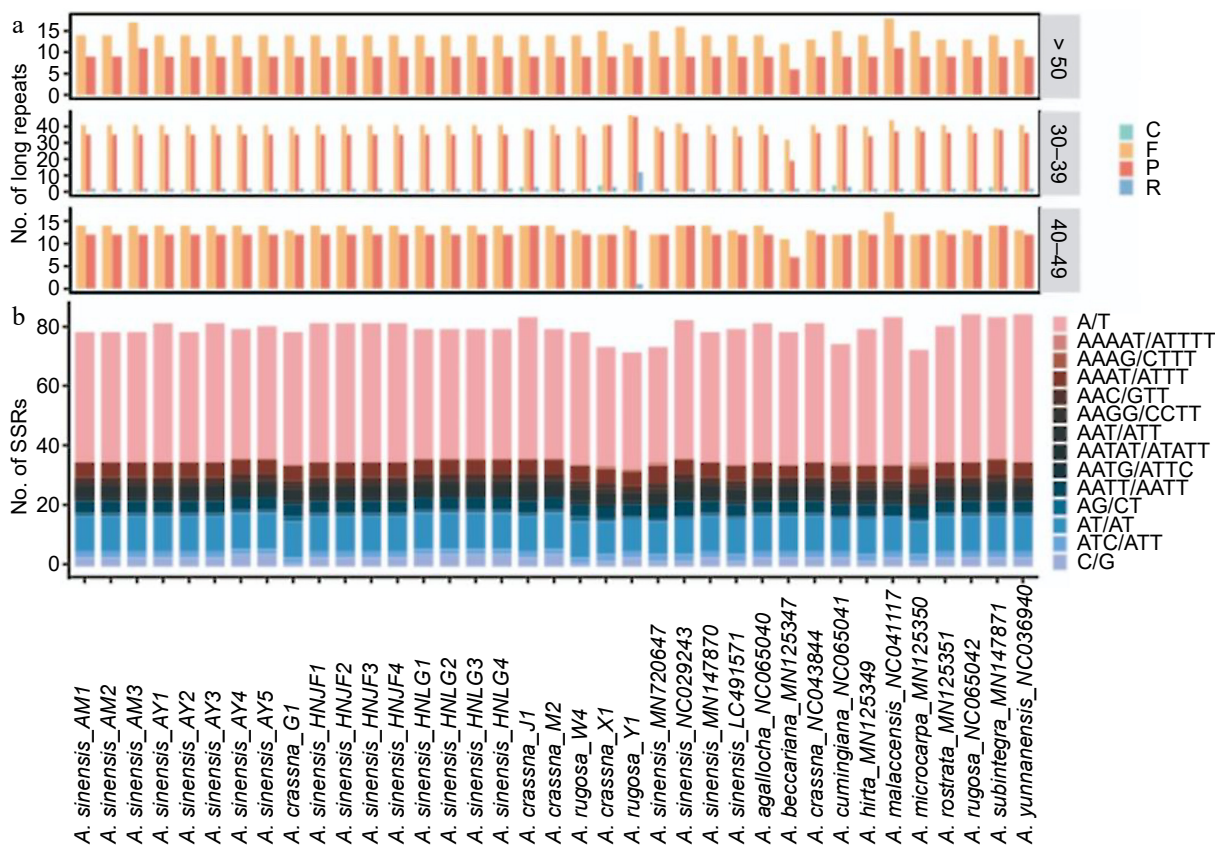
In the MISA-web analysis, five types of simple sequence repeats (SSRs) were identified, with mononucleotide repeats comprising the majority (1,775, 59.91%), followed by dinucleotide repeats (477, 16.1%), trinucleotide repeats (294, 9.92%),

tetranucleotide repeats (409, 13.80%), and pentanucleotide repeats (8, 0.27%) (Fig. 4b). Among these SSRs, repeat units of A/T, AT/AT, AAAT/ATTT, AAAAT/ATTTT, and AAATT/AATTT accounted for 87.55% of the total, indicating a bias towards A/T bases in SSR composition.

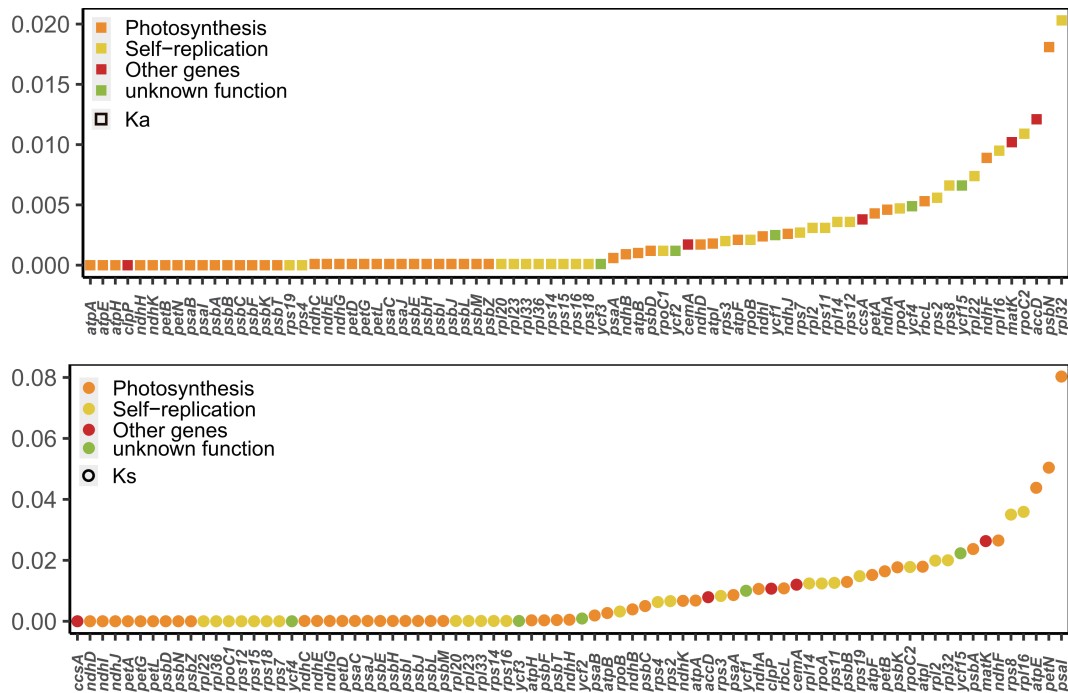
### Selection pressure and phylogenetic analysis

Ka/Ks is a valuable metric for evaluating whether PCGs have undergone adaptive evolution. Typically, synonymous nucleotide substitutions occur more frequently than non-synonymous substitutions in most genes of organisms, resulting in Ka/Ks values typically being less than 1. In this study, the Ka/Ks values for the PCGs of these plastomes were calculated. Most of the PCGs exhibited Ka/Ks values of less than 1, indicating that these species were subjected to strong purifying selection during the long-term evolutionary process (Fig. 5). Notably, the expression-related *accD* gene displayed the highest Ka/Ks value (1.54). This finding suggests that the *accD* gene may have been influenced by positive selection during the evolutionary process.

To elucidate the phylogenetic relationships within the *Aquilaria* genus, 22 newly sequenced *Aquilaria* plastomes were utilized in conjunction with an additional 15 *Aquilaria* plastomes to reconstruct the phylogenetic tree using ML and BI methods. Additionally, three *Daphne* samples were selected as outgroups (Fig. 6). The results revealed the formation of five clades within *Aquilaria*. The newly sequenced *A. sinensis* plastomes clustered together with published *A. sinensis* plastomes, forming a monophyletic clade designated as clade 1 (C1). Clade 2 (C2) comprised three *A. rugosa* plastomes, which clustered



**Fig. 4** (a) Distribution of different types of large sequences repeat (LSRs) in *Aquilaria* plastomes, including complementary (C), forward (F), palindromic (P), and reverse (R) type repeats. (b) Distribution of distinct types of simple sequence repeats (SSRs) in *Aquilaria* plastomes.



**Fig. 5** Values of synonymous (Ks) and non-synonymous (Ka) substitution rates for 79 protein-coding genes in *Aquilaria*.

together and formed a sister group to *A. yunnanensis* with high confidence. Clade 3 (C3) consisted of only one species, identified as *A. agallocha*.

In clade 4 (C4), it was observed that *A. crassna* (NC043844) clustered together with *A. subintegra* (MN147871), which aligns with previous research<sup>[47]</sup>. The newly sequenced *A. crassna* plastomes also clustered together, demonstrating high consistency. This finding suggests the possibility of hybridization events or misidentification of the published plastome sample (NC043844). It is conceivable that hybridization occurred among these species, with plastomes likely following maternal inheritance, leading to significant disturbance in DNA information within the plastomes. Alternatively, the high similarity in morphology between these species may have contributed to misidentification. Lastly, clade 5 (C5) encompassed five species, mirroring the phylogenetic findings of a previous study<sup>[29,47]</sup>.

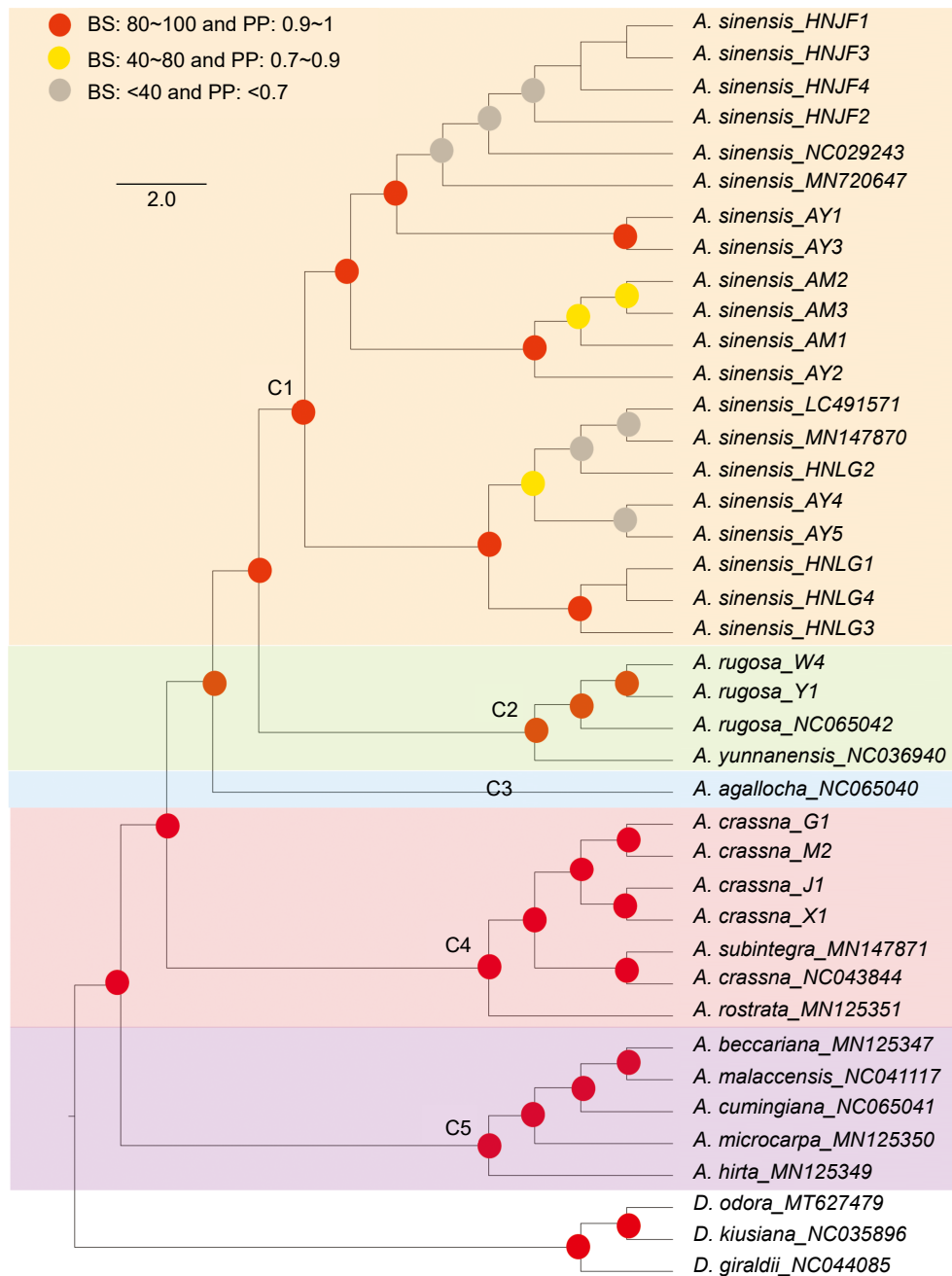
## Discussion

In this study, the complete plastome sequences of 22 *Aquilaria* taxa were presented. By integrating these newly sequenced plastomes with an additional 15 *Aquilaria* published plastomes, a series of comparative analyses within this genus were conducted. The investigation into the location of the SC/IR boundaries offers novel insights into evolutionary studies. The expansion and contraction of IR regions are often associated with various evolutionary events, including gene duplication, loss of gene copies, and the emergence of pseudogenes (such as the *ycf1* pseudogene observed in many plants)<sup>[48,49]</sup>. The substantial variation observed among these boundaries may harbor potential markers for genus/species identification<sup>[50,51]</sup>. The present research has unveiled a consistent structural feature at the IR/SC boundary in *Aquilaria* species. Intriguingly, all *Aquilaria* plastomes analyzed in this study exhibited a 16 bp extension of the *rps19* gene in the IRB

region. This phenomenon is rarely observed in other species within the Thymelaeaceae family, as it has only been detected in a few species across certain genera<sup>[52,53]</sup>. In *Aquilaria*, a consistent 16 bp extension was observed across all examined species, suggesting its potential as a genus-specific marker. Such markers have been utilized for identification previously; for example, the 110 bp expansion of the *rps19* gene in the Crassulaceae family serves as a specific family marker<sup>[19]</sup>. Future studies employing extensive datasets of Thymelaeaceae species may identify additional family-specific identification markers. Furthermore, discrepancies were noted in the plastomes of *A. malaccensis* (NC041117) and other *Aquilaria* species, particularly in the positions and transcriptional directions of *ndhF* and *rpl32*; their transcribed directions were reversed. These anomalies were primarily attributed to assembly errors. Notably, the same misassembly was found in *A. sinensis* (NC029243). The loss of *rps32* and the incorrect positioning of *trnN*, *ndhF*, and *ycf1* genes further underscored the erroneous assembly of the plastome. To elucidate the authentic structural patterns of the plastome genome, these inaccurate annotations were substituted with normal data, thereby eliminating potential sources of misinterpretation.

Furthermore, *accD* displayed the highest Ka/Ks value of 1.54. Previous research has suggested that adaptive evolution may occur at the molecular level, as evidenced by an elevated Ka/Ks value<sup>[54]</sup>. Numerous studies have corroborated that a higher Ka/Ks ratio correlates with a more robust positive selection<sup>[26,55,56]</sup>. Thus, the elevated Ka/Ks ratio observed in *accD* implies its significant contribution in the adaptation and evolutionary processes within *Aquilaria* species.

While previous research has attempted to elucidate the phylogenetic status of the *Aquilaria* genus, the limited availability of plastomic resources for *Aquilaria* has left certain issues unresolved. One such point of interest pertains to the relationship among *A. sinensis*, *A. agallocha*, and *A. yunnanensis*. Farah



**Fig. 6** Phylogenetic tree depicting relationships among *Aquilaria* species. Red circles indicates bootstrap support (BS) values greater than 80% and posterior probability (PP) values exceeding 1. Yellow circles indicates BS values ranging from 40% to 80%, and PP values ranging from 0.7 to 0.9. Grey circles indicates BS values below 40% and PP values less than 0.7.

et al.<sup>[28]</sup> utilized two types of datasets (chloroplast loci and ITS regions) and reported conflicting patterns among these three species. The chloroplast loci data clustered these species into the same clade, with *A. agallocha* being closer to *A. sinensis*. Conversely, the evolutionary tree based on ITS regions separated *A. sinensis* to another clade. Another study<sup>[29]</sup> also examined these species for phylogenetic analysis, yielding different results. In that study, these species were clustered in the same clade, with *A. yunnanensis* exhibiting a closer relationship to *A. sinensis*. Additionally, several studies<sup>[1,47,57]</sup> have explored phylogenetic relationships in *Aquilaria*; however, not all of them included all three species of interest. The phylogenetic

tree reconstructed in this study mirrored the results of the study conducted by Lee et al.<sup>[29]</sup> with high confidence. The structure of the phylogenetic tree appeared to be highly influenced by the volume of the dataset; as the number of *Aquilaria* plastomes were increased, it was anticipated that this result would better elucidate the true relationship among these species.

Furthermore, of interest was a previous study<sup>[47]</sup> which revealed that different varieties of *A. crassna* did not form a monophyletic clade, as *A. subintetgra* (MN147870) clustered with *A. crassna* (MK779998, also coded as NC043844). To explore the underlying reason for this discrepancy, four *A.*

*crassna* plastomes were sequenced. The phylogenetic results unveiled that the newly sequenced *A. crassna* plastomes in this study clustered together. However, the published *A. crassna* (NC043844) also clustered with *A. subintegra* (MN147871), aligning with the findings of the earlier research conducted by Hishamuddin et al.<sup>[47]</sup>. Chloroplasts, organelles present in plants and some protist cells, possess their own DNA distinct from the nuclear genome. The similarity of the chloroplast genome sequences can unveil phylogenetic relationships and evolutionary history among species<sup>[58–62]</sup>, suggesting inheritance from a common ancestor in many cases<sup>[63–65]</sup>, as chloroplasts are typically maternally inherited in plants<sup>[66,67]</sup>. Conversely, similar plastomes may also result from convergent evolution, where analogous chloroplast genome sequences evolve independently on distinct evolutionary trajectories. In the case of *A. crassna* (NC043844), there is a possibility of incorrect assembly, or even misidentification of the specimen as either *A. crassna* (NC043844) or *A. subintegra* (MN147871). Moreover, conducting a pan-plastome study involving *A. subintegra* and *A. crassna* could offer a more comprehensive understanding of the relationship between these two species.

With the augmented dataset of plastomes in this study, a more distinct phylogenetic pattern was unveiled compared to previous research. Besides, although plastomes can provide valuable insights into phylogenetic relationships, the limitations imposed by frequent artificial hybridization are recognized<sup>[8]</sup>. It is anticipated that future studies integrating data from nuclear genomes, mitogenomes, and plastomes will undoubtedly provide a deeper understanding of the phylogenetic implications within the *Aquilaria* genus. This study can serve as a valuable example for species identification in trees, highlighting the significance of understanding evolutionary relationships between species or subspecies, given the high similarity among tree species.

## Conclusions

This study presents the sequencing of 22 *Aquilaria* plastomes, revealing no gene rearrangements among them. Comparative genomics unveiled a consistent genome size (174,804–175,046 bp) and GC content (36.70%–36.75%) across *Aquilaria* plastomes. Additionally, seven HVRs were identified, spanning *rps16*, *rpoC1*, *psbM-trnD-GUC*, *ndhC-trnV-UAC*, *psbJ-petA*, *ndhF-rpl32*, and *rpl32-trnL-UAG*. In examining the IR junction pattern, it was noted that the *rps19* gene spans both the LSC and IRb regions, with a 16 bp expansion of the *rps19* gene in IRb, potentially serving as a genus-specific marker for *Aquilaria*. Furthermore, the present analysis identified four types of LSRs and five types of SSRs with a notable A/T bias. The *ndhD* gene displayed the highest Ka/Ks value of 1.54, indicative of its significant contribution to the adaptation and evolution of *Aquilaria* species. Finally, the present study elucidated the phylogenetic relationships within *Aquilaria* and addressed several issues within the genus through an expanded dataset, providing valuable insights into its evolutionary dynamics.

## Author contributions

The authors confirm contribution to the paper as follows: study conception and design: Wu Z, Hong Z; data collection: Liu X, Xu D, Tembrock LR; analysis and interpretation of results:

Kan J, Mi Z, Nie L; draft manuscript preparation: Kan J, Mi Z; organizing and supervising the whole project: Wu Z, Hong Z. All authors reviewed the results and approved the final version of the manuscript.

## Data availability

The annotation files of newly assembled *Aquilaria* plastomes are uploaded to the Figshare website (<https://figshare.com/>, doi:10.6084/m9.figshare.25713024).

## Acknowledgments

This study was funded by Forestry Science and Technology Innovation Project of Guangdong Province (Grant No. 2023KJCX009) and Fundamental Research Funds for the Central Non-profit Research Institution of Chinese Academy of Forestry (Grant No. CAFYBB2021QD003-03). It was also funded by the Science Technology and Innovation Commission of Shenzhen Municipality (Grant No. RCYX20200714114538196) and the Shenzhen Fundamental Research Program (Grant No. JCYJ20220818103212025). This work was also funded by the Guangdong Pearl River Talent Program (Grant No. 2021QN02N792) and the Innovation Program of Chinese Academy of Agricultural Sciences.

## Conflict of interest

The authors declare that they have no conflict of interest.

**Supplementary Information** accompanies this paper at (<https://www.maxapress.com/article/doi/10.48130/forres-0024-0028>)

## Dates

Received 6 May 2024; Revised 6 August 2024; Accepted 19 August 2024; Published online 4 September 2024

## References

1. Eurlings MCM, Gravendeel B. 2005. *TrnL-trnF* sequence data imply paraphyly of *Aquilaria* and *Gyrinops* (Thymelaeaceae) and provide new perspectives for agarwood identification. *Plant Systematics and Evolution* 254:1–12
2. Hashim YZHY, Kerr PG, Abbas P, Mohd Salleh H. 2016. *Aquilaria* spp. (agarwood) as source of health beneficial compounds: a review of traditional use, phytochemistry and pharmacology. *Journal of Ethnopharmacology* 189:331–60
3. Takamatsu S, Ito M. 2020. Agarotetrol in agarwood: its use in evaluation of agarwood quality. *Journal of Natural Medicines* 74:98–105
4. López-Sampson A, Page T. 2018. History of use and trade of agarwood. *Economic Botany* 72:107–29
5. Adam AZ, Lee SY, Mohamed R. 2017. Pharmacological properties of agarwood tea derived from *Aquilaria* (Thymelaeaceae) leaves: an emerging contemporary herbal drink. *Journal of Herbal Medicine* 10:37–44
6. Liu Y, Chen H, Yang Y, Zhang Z, Wei J, et al. 2013. Whole-tree agarwood-inducing technique: an efficient novel technique for producing high-quality agarwood in cultivated *Aquilaria sinensis* trees. *Molecules* 183:3086–106
7. Ahmad Mokhtar AM, Mohd Zain HH, Mohammed Muayad TA. 2021. Overview of medicinal properties and toxicities of



- agarwood species. *Educatum Journal of Science, Mathematics and Technology* 8:1–16
8. Tsuyuki H, Roberts E. 1965. Zone electrophoretic comparison of muscle myogens and blood proteins of artificial hybrids of Salmonidae with their parental species. *Journal of the Fisheries Board of Canada* 22:767–73
  9. Yin Y, Jiao L, Dong M, Jiang X, Zhang S. 2016. Wood resources, identification, and utilization of agarwood in China. In *Agarwood*, ed. Mohamed R. Singapore: Springer. pp. 21–38. doi:10.1007/978-981-10-0833-7\_2
  10. Wyn L, Anak N. 2010. *Wood for the trees: a review of the agarwood (Gaharu) trade in Malaysia*. Malaysia: Traffic Southeast Asia. 108 pp.
  11. Uddin MS, Mukul SA, Khan MASA, Alamgir M, Harun MY, et al. 2008. Small-scale agar (*Aquilaria agallocha* Roxb.) based cottage enterprises in Maulvibazar district of Bangladesh: production, marketing and potential contribution to rural development. *Small-scale Forestry* 7:139–49
  12. Sy EY, Melgar AGB. 2024. The online agarwood trade and seizure analysis in the philippines. *Journal of Nature Studies* 22(2):12–26
  13. Eurlings MCM, van Beek HH, Gravendeel B. 2010. Polymorphic microsatellites for forensic identification of agarwood (*Aquilaria crassna*). *Forensic Science International* 197:30–34
  14. Naef R. 2011. The volatile and semi-volatile constituents of agarwood, the infected heartwood of *Aquilaria* species: a review. *Flavour and Fragrance Journal* 26:73–87
  15. Dahham SS, Hassan LEA, Ahamed MBK, Majid ASA, Majid AMSA, et al. 2016. In vivo toxicity and antitumor activity of essential oils extract from agarwood (*Aquilaria crassna*). *BMC Complementary and Alternative Medicine* 16:236
  16. Wu Z, Gu C, Tembrock LR, Zhang D, Ge S. 2017. Characterization of the whole chloroplast genome of *Chikusichloa mutica* and its comparison with other rice tribe (Oryzaceae) species. *PLoS One* 12:e0177553
  17. Takamatsu T, Baslam M, Inomata T, Oikawa K, Itoh K, et al. 2018. Optimized method of extracting rice chloroplast DNA for high-quality plastome resequencing and *de novo* assembly. *Frontiers in Plant Science* 9:266
  18. Zhao J, Zhang X, Hong Y, Liu Y. 2016. Chloroplast in plant-virus interaction. *Frontiers in Microbiology* 7:1565
  19. Ding H, Han S, Ye Y, Bi D, Zhang S, et al. 2022. Ten plastomes of *Crassula* (Crassulaceae) and phylogenetic implications. *Biology* 11:1779
  20. Zheng S, Poczai P, Hyvönen J, Tang J, Amiryousefi A. 2020. Chloroplast: an online program for the versatile plotting of organelle genomes. *Frontiers in Genetics* 11:576124
  21. Sheng J, Yan M, Wang J, Zhao L, Zhou F, et al. 2021. The complete chloroplast genome sequences of five *Miscanthus* species, and comparative analyses with other grass plastomes. *Industrial Crops and Products* 162:113248
  22. Xie C, An W, Liu S, Huang Y, Yang Z, et al. 2021. Comparative genomic study on the complete plastomes of four officinal *Ardisia* species in China. *Scientific Reports* 11:22239
  23. Yan XL, Kan SL, Wang MX, Li YY, Tembrock LR, et al. 2024. Genetic diversity and evolution of the plastome in allotetraploid cotton (*Gossypium* spp.). *Journal of Systematics and Evolution* 00:Early View
  24. Zhang S, Han S, Bi D, Yang J, Ge W, et al. 2024. Intraspecific and intragenomic genomic variation across three *Sedum* species (Crassulaceae): a plastomic perspective. *Genes* 15:444
  25. Kan J, Zhang S, Wu Z, Bi D. 2024. Exploring plastomic resources in *Sempervivum* (Crassulaceae): implications for phylogenetics. *Genes* 15:441
  26. Wang J, Liao X, Li Y, Ye Y, Xing G, et al. 2023. Comparative plastomes of *Curcuma alismatifolia* (Zingiberaceae) reveal diversified patterns among 56 different cut-flower cultivars. *Genes* 14:1743
  27. Wang J, Kan S, Liao X, Zhou J, Tembrock LR, et al. 2024. Plant organellar genomes: much done, much more to do. *Trends in Plant Science* 29:754–69
  28. Farah AH, Lee SY, Gao Z, Yao TL, Madon M, et al. 2018. Genome size, molecular phylogeny, and evolutionary history of the tribe Aquilarieae (Thymelaeaceae), the natural source of agarwood. *Frontiers in Plant Science* 9:712
  29. Lee SY, Turjaman M, Chaveerach A, Subasinghe S, Fan Q, et al. 2022. Phylogenetic relationships of *Aquilaria* and *Gyrinops* (Thymelaeaceae) revisited: evidence from complete plastid genomes. *Botanical Journal of the Linnean Society* 200:344–59
  30. Porebski S, Bailey LG, Baum BR. 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter* 15:8–15
  31. Jin JJ, Yu WB, Yang JB, Song Y, DePamphilis CW, et al. 2020. GetOrganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. *Genome Biology* 21:241
  32. Shi L, Chen H, Jiang M, Wang L, Wu X, et al. 2019. CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Research* 47:W65–W73
  33. Greiner S, Lehwark P, Bock R. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3. 1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research* 47:W59–W64
  34. Li H, Guo Q, Xu L, Gao H, Liu L, et al. 2023. CPJSDraw: analysis and visualization of junction sites of chloroplast genomes. *PeerJ* 11:e15326
  35. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research* 14:1394–403
  36. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. 2004. VISTA: computational tools for comparative genomics. *Nucleic Acids Research* 32:W273–W279
  37. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, et al. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution* 34:3299–302
  38. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, et al. 2001. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research* 29:4633–42
  39. Beier S, Thiel T, Münch T, Scholz U, Mascher M. 2017. MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33:2583–85
  40. Xia X. 2017. DAMBE6: new tools for microbial genomics, phylogenetics, and molecular evolution. *Journal of Heredity* 108:431–37
  41. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30:772–80
  42. Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 24:1586–91
  43. Stamatakis A. 2014. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–13
  44. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61:539–42
  45. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, et al. 2020. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Molecular Biology and Evolution* 37:291–94
  46. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* 67:901–04
  47. Hishamuddin MS, Lee SY, Ng WL, Ramlee SI, Lamasudin DU, et al. 2020. Comparison of eight complete chloroplast genomes of the endangered *Aquilaria* tree species (Thymelaeaceae) and their phylogenetic relationships. *Scientific Reports* 10:13034

48. Zheng G, Wei L, Ma L, Wu Z, Gu C, et al. 2020. Comparative analyses of chloroplast genomes from 13 *Lagerstroemia* (Lythraceae) species: identification of highly divergent regions and inference of phylogenetic relationships. *Plant Molecular Biology* 102:659–76
49. Abdullah, Mehmood F, Shahzadi I, Waseem S, Mirza B, et al. 2020. Chloroplast genome of *Hibiscus rosa-sinensis* (Malvaceae): comparative analyses and identification of mutational hotspots. *Genomics* 112:581–91
50. Wang J, Fu GF, Tembrock LR, Liao XZ, Ge S, et al. 2023. Mutational meltdown or controlled chain reaction: The dynamics of rapid plastome evolution in the hyperdiversity of Poaceae. *Journal of Systematics and Evolution* 61:328–44
51. Zhou J, Zhang S, Wang J, Shen H, Ai B, et al. 2021. Chloroplast genomes in *Populus* (Salicaceae): comparisons from an intensively sampled genus reveal dynamic patterns of evolution. *Scientific Reports* 11:9471
52. He L, Zhang Y, Lee SY. 2021. Complete plastomes of six species of *Wikstroemia* (Thymelaeaceae) reveal paraphyly with the monotypic genus *Stellera*. *Scientific Reports* 11:13608
53. Qian S, Zhang Y, Lee SY. 2021. Comparative analysis of complete chloroplast genome sequences in *Edgeworthia* (Thymelaeaceae) and new insights into phylogenetic relationships. *Frontiers in Genetics* 12:643552
54. Bakewell MA, Shi P, Zhang J. 2007. More genes underwent positive selection in chimpanzee evolution than in human evolution. *Proceedings of the National Academy of Sciences of the United States of America* 104:7489–94
55. Ai B, Gao Y, Zhang X, Tao J, Kang M, et al. 2015. Comparative transcriptome resources of eleven *Primulina* species, a group of 'stone plants' from a biodiversity hot spot. *Molecular Ecology Resources* 15:619–32
56. Hurst LD. 2002. The *Ka/Ks* ratio: diagnosing the form of sequence evolution. *Trends in Genetics* 18:486–87
57. Zhang YH, Huang Y, Li ZM, Zhang SD. 2019. Characterization of the complete chloroplast genome of the vulnerable agarwood tree, *Aquilaria yunnanensis* (Thymelaeaceae). *Conservation Genetics Resources* 11:161–64
58. Guo L, Guo S, Xu J, He L, Carlson JE, et al. 2020. Phylogenetic analysis based on chloroplast genome uncover evolutionary relationship of all the nine species and six cultivars of tree peony. *Industrial Crops and Products* 153:112567
59. Sun J, Wang Y, Liu Y, Xu C, Yuan Q, et al. 2020. Evolutionary and phylogenetic aspects of the chloroplast genome of *Chaenomeles* species. *Scientific Reports* 10:11466
60. Li E, Liu K, Deng R, Gao Y, Liu X, et al. 2023. Insights into the phylogeny and chloroplast genome evolution of *Eriocaulon* (Eriocaulaceae). *BMC Plant Biology* 23:32
61. Liang H, Zhang Y, Deng J, Gao G, Ding C, et al. 2020. The complete chloroplast genome sequences of 14 *Curcuma* species: insights into genome evolution and phylogenetic relationships within Zingiberales. *Frontiers in Genetics* 11:802
62. Dong WL, Wang RN, Zhang NY, Fan WB, Fang MF, et al. 2018. Molecular evolution of chloroplast genomes of orchid species: insights into phylogenetic relationship and adaptive evolution. *International Journal of Molecular Sciences* 19:716
63. Du F, Li Y, Xu K. 2023. Phylogeny and evolution of *Cocconeopsis* (Cocconeidaceae) as revealed by complete chloroplast and mitochondrial genomes. *International Journal of Molecular Sciences* 25:266
64. Ahmad S, Kaul T, Chitkara P, Raza K. 2023. Comparative insight into rice chloroplasts genome: mutational phylogenomics reveals *Echinochloa oryzicola* as the ongoing progenitor of rice. *Genetic Resources and Crop Evolution* 70:869–85
65. Tomitani A, Okada K, Miyashita H, Matthijs HCP, Ohno T, et al. 1999. Chlorophyll *b* and phycobilins in the common ancestor of cyanobacteria and chloroplasts. *Nature* 400:159–62
66. Li S, Duan W, Zhao J, Jing Y, Feng M, et al. 2022. Comparative analysis of chloroplast genome in *Saccharum* spp. and related members of 'Saccharum Complex'. *International Journal of Molecular Sciences* 23:7661
67. Yan L, Wang H, Huang X, Li Y, Yue Y, et al. 2022. Chloroplast genomes of genus *Tilia*: comparative genomics and molecular evolution. *Frontiers in Genetics* 13:925726



Copyright: © 2024 by the author(s). Published by Maximum Academic Press, Fayetteville, GA. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit <https://creativecommons.org/licenses/by/4.0/>.