Jurnal Manajemen Hutan Tropika, *30*(3), 326–335, December 2024
EISSN: 2089-2063
DOI: 10.7226/jtfm.30.3.326

Scientific Article
ISSN: 2087-0469

A Comparative Study of the Organellar Genome of *Gyrinops versteegii* and *Aquilaria malaccensis*

Hartati Hartati[1,5*], Imam Civi Cartealy[2], Supatmi Supatmi[3], Syamsidah Rahmawati[3], N Sri Hartati[3], Ulfah Juniarti Siregar[4], Iskandar Zulkarnaen Siregar[4, 5]

[1]Research Center for Applied Botany, National Research and Innovation Agency, Jl. Raya Jakarta-Bogor Km 46, Bogor, Indonesia 16911
[2]Research Center for Computation, National Research and Innovation Agency, Jl. Raya Jakarta-Bogor Km 46, Bogor, Indonesia 16911
[3]Research Center for Genetic Engineering, National Research and Innovation Agency, Jl. Raya Jakarta-Bogor Km 46, Bogor, Indonesia 16911
[4]Department of Silviculture, Faculty of Forestry and Environment, IPB University, Academic Ring Road, Campus IPB Dramaga, Bogor, Indonesia 16680
[5]Collaborative Research Center, Faculty of Forestry and Environment, IPB University, Academic Ring Road, Campus IPB Dramaga, Bogor, Indonesia

Received May 25, 2024/Accepted October 29, 2024

Abstract

Gyrinops versteegii and Aquilaria malaccensis are two important species of the Aquilarieae tribe. The main problem of this tribe is the challenge of species identification that is strongly dependent on the presence of flowers and fruit, which are not always available. The availability of whole genome information is expected to address the problems of species identification. This research aims to construct and compare the chloroplast and mitochondrial genomes of G. versteegii and A. malaccensis from short-read data using the NOVOplasty and GetOrganelle assembler. The chloroplast genome assembly revealed a full-length quadripartite circular structure with sizes of 174.814 bp (*G. versteegii*) *and 174.821–174.822 bp* (*A. malaccensis*)*, with highly conserved gene and organization. Meanwhile, the mitochondrial genome is multipartite with a size of 400.012 bp* (*G. versteegii*) *and 400.000 bp* (*A. malaccensis*)*, with highly variable genes and organization due to the presence of gene cluster repeats. The LSC/IR/SCC region borders and phylogenetic analysis in chloroplasts indicate variations between the genomes of these two species. The investigation of nucleotide diversity in the chloroplast genome revealed that the trnL-rpl32 region had the highest nucleotide diversity* (*Pi = 0.03*)*. This information will be useful in the future for a variety of downstream analyses.*

Keywords: *chloroplast, GetOrganelle, mitochondria, NOVOplasty, short-read*

**Correspondence author, email*: *hartati007@brin.go.id, hartati72lipi@gmail.com*

Introduction

Gyrinops versteegii and *Aquilaria malaccensis* are well-known as the agarwood-producing trees of the *Aquilarieae* tribe. Agarwood is a non-timber forest product (NTFP) derived from secondary metabolites, which are produced as plant defense products in response to microbial infections induced by physical injury (Naziz et al*.,* 2019). Agarwood sapwood has a considerable economic value and is widely used in incense, perfume, traditional medicine, carvings, decorations, and religious rites (Akter et al*.,* 2013; Naziz et al*.,* 2019). *G. versteegii* is an Indonesian endemic species native to Papua, Maluku, and Nusa Tenggara. Whereas, *A. malaccensis* has a broad distribution and can be found in Sumatra, Borneo, Malaysia, Singapore, the Philippines, Myanmar, and India (Lee et al., 2018).

Identification of *Aquilarieae tribe* species remains challenging due to a lack of distinctive traits between species. However, the presence of flowers and fruit are the key morphological characteristics (Lee & Mohamed, 2016). Therefore, a genetic identification strategy is required, particularly in the absence of flower or fruit parts. The use of several DNA barcodes, such as *matK* and *rbcL,* or non-coding chloroplast DNA segments, such as *trnL-trnF, psbC-trnS,* and internal transcribed spacer (ITS), has proven to be insufficient for differentiating these taxa (Farah et al., 2018; Hartati et al., 2023). Moreover, *Aquilarieae* species are also found paraphyletic (Lee et al., 2022; Hartati et al., 2023). The chloroplast (cp) and mitochondrial (mito) genomes contain numerous genes that can be used to identify species. The high copy number of cp- and mitogenomes in a single cell increases the possibility of obtaining enough reads from whole genome sequencing (WGS) data to construct complete organelle genomes (Twyford & Ness, 2016).

Chloroplast genomes are generally circular quadripartite, with one large single copy (LSC) region and one small single copy (SSC) region flanked by two inverted repeats (IR) regions (Hishamuddin et al., 2020). They range in size from 120 to 160 Kbp (Palmer, 1985). Mitogenomes have a

Jurnal Manajemen Hutan Tropika, *30*(3), 326–335, December 2024
EISSN: 2089-2063
DOI: 10.7226/jtfm.30.3.326

Scientific Article
ISSN: 2087-0469

complex structure with reported mitochondrial genomes in plant species ranging from 66 to 11.000 kb (Fauron et al., 2004; Wu et al., 2020). The whole mitochondrial genome of *A. malaccensis* and *G. versteegii* has never been reported. Complete plant cp- and mitogenome sequences have been proposed to provide a more detailed perspective on a researched species in terms of taxonomy, diversity, phylogenetics, and evolutionary pattern (Farah et al., 2018; Hishamuddin et al., 2020; Lee et al., 2022). Furthermore, it is also essential for barcoding/meta-barcoding for biotechnology applications and genetic engineering, as well as for comparative studies and species identification, which is essential for effective conservation strategies (Tonti-Filippini et al., 2017; Liu et al., 2023).

Bioinformatics approaches have made it possible to *de novo* assemble organellar genomes of plant species from whole genome sequencing (WGS) data generated with short-read technology. NOVOplasty (Dierckxsens et al., 2017) and GetOrganelle (Jin et al., 2020) are two assemblers designed specifically for *de novo* assembly of organellar genomes from whole genome sequencing (WGS) data. NOVOplasty combines both approaches by starting with a related or distant single seed sequence and simultaneously constructing genomes from reads based on k-mers (Dierckxsens et al., 2017; Freudenthal et al., 2020). GetOrganelle, on the other hand, uses a modified baiting and iterative mapping strategy to identify organelle-associated reads before performing *de novo* assembly and creating all possible configurations of circular organelle genomes (Jin et al., 2020). Our study aims a) to *de novo* assembly of both the cp- and mitogenomes of *G. versteegii* and *A. malaccensis* using two assemblers, NOVOplasty and GetOrganelle, based on short-read sequences and b) to conduct a comparative genome study of both the mitochondrial and chloroplast genome sequences of *G. versteegii* and *A. malaccensis.*

Methods

Research procedure Our study was conducted in several stages: a) genomic DNA isolation, b) genomic library preparation and short-read sequencing following the Illumina manufacturers' protocol, c) organellar genome assembly, and d) downstream analysis, including the comparative and phylogenetic study of the cp- and mitogenomes of *G. versteegii and A. malaccensis*

Plant materials and total genomic DNA extraction *G. versteegii* and *A. malaccensis* were used as genomic DNA sources. Fresh leaf of *A. malaccensis* was collected from a selected tree grown in Science Technology Park Sukarno-Cibinong, and a fresh leaf of *G. versteegii* was taken from farmland in Cilodong, West Java Province, Indonesia. Genomic DNA was extracted using Cetyltrimethyl-ammonium bromide (CTAB) following the Doyle and Doyle (1990) protocol with a few minor modifications. The quantity and quality of the DNA samples were evaluated and measured using gel electrophoresis and the Qubit dsDNA BR assay (Life Technologies, Carlsbad, CA, USA) following the manufacturer's instructions.

Short-read sequencing A genomic library with a 300 bp insert size was prepared according to the Illumina library preparation manufacturer's protocols, followed by sequencing with the Illumina Novaseq 4500 (Illumina, San Diego, CA) technology.

Genome assembly About 14,1 GB of raw data from *G. versteegii* and 13,1 GB of raw data from *A. malaccensis* were obtained. After cleaning the adaptor sequences with Cutadapt (Martin, 2011) and trimming the low read quality with FastX toolkit (Hannon, 2010), the raw reads were *de novo assembled* using the NOVOplasty v4.3.1 from usegalaxy.edu (https://github.com/ndierckx/NOVOPlasty) and GetOrganelle v1.7.7.0 (https://github.com/Kinggerm/GetOrganelle) to construct the mitochondrial and chloroplast genomes (Dierckxsens et al., 2017; Jin et al., 2020). QUAST v.5.0.2 (Mikheenko et al., 2018) was used to generate the assembled statistics. The *matK* sequences of *A. malaccensis* (GenBank accession KY927320.1) and *G. versteegii* (GenBank accession LC467531.1) were utilized as the seed sequences to construct the cp genome, while for mitochondria, the seed sequence from *matR* of *Aquilaria sinensis* (GenBank accession AF520171.1) was used. After assembly, the genomes were annotated with GeSeq (Tillich et al., 2017). The circular cp genome structures were visualized using OGDRAW v1.3.1 (Tillich et al., 2017) that linked to usegalaxy.edu.

Comparative analysis of cp- and mitogenomes The GC content was determined using Geneious Prime v2023.2.1 (www.geneious.com). To assess nucleotide diversity in cp genomes, sequences were initially aligned using MAFFT v7 (Rozewicki et al., 2019) using the default settings (FFT-NS-2). The resulting alignments were then imported into DNASP version 5.10.1 (Librado & Rozas, 2009). The polymorphic sites and nucleotide variability (Pi) in MAFFT-aligned cp genomes were assessed using a sliding window approach with a step size of 200 bp and a window length of 600 bp. To determine length variation in cp genomes, the borders between the IR and SC regions were manually assessed with SnapGene Viewer v6.0.2 (www.snapgene.com).

Phylogenomic analysis of *G. versteegii* and *A. malaccensis*
Phylogenomic studies were carried out using multiple cp genomes to determine the relationship of both *G. versteegii* and *A. malaccensis* species to their sister and outgroup. The sister group was chosen from the same family, Thymelaeaceae, however, the tribes were different. *Daphne tangutica* belongs to the Daphneae tribe, while *Gonystylus affinis* is in the *Gonystylus* genus. The outgroup was selected from a different family. *Eucalyptus grandis* comes to the Myrtaceae family. The genomic sequences of the outgroup *E. grandis* (GenBank accession Nc014570), sister group *G. affinis* (GenBank accession NC052860.1), and *Daphne tangutica* (GenBank accession MK557323.1) were obtained from the NCBI GenBank databases. *A. sinensis* (GenBank NC 029243.1), *A. crassna* (GenBank NC 043844.1), and *Gyrinops walla* (GenBank MW455800.1) were also utilized as Aquilarieae species and included for phylogenomic analysis. The sequences were aligned with MAFFT v7

Jurnal Manajemen Hutan Tropika, *30*(3), 326–335, December 2024
EISSN: 2089-2063
DOI: 10.7226/jtfm.30.3.326

Scientific Article
ISSN: 2087-0469

(Rozewicki et al., 2019). Several evolutionary models were evaluated using maximum likelihood statistical approaches to select the best model for constructing a phylogenetic tree. The construction of a phylogenetic tree from the chloroplast genome data sequence was carried out using MEGA7 software (Kumar et al., 2016). The Tamura-3 was selected as a model with a maximum likelihood parameter based on the best fitting model analysis to construct a phylogenetic tree. A 1,000× bootstrap replication was used to provide consistent phylogenetic tree estimation with lower error. This analysis includes all positions with missing data or gaps. The tree was rooted using *E. grandis*.

Results and Discussions

Characteristics of the cp- and mitogenomes of *G. versteegii* and *A. malaccensis* We successfully assembled complete cp genomes of *G. versteegii* and *A. malaccensis* from short-read data using both the NOVOplasty v4.3.1 and the GetOrganelle v1.7.7.0. Both assemblers produced complete cp genomes of similar size. The full cp genomes of *G. versteegii* and *A. malaccensis* showed a quadripartite structure, with the length of the *G. versteegii* cp genome obtained being 174.814 bp and the *A. malaccensis* cp genome being 174.821 with NOVOplasty or 174.822 bp with GetOrganelle (Table 1, Figure 1). QUAST analysis confirmed that the cp genomes of *G. versteegii* and *A. malaccensis* obtained were complete. This is indicated by the genome fraction or coverage obtained being 100%, and the N50 length was the same as the length of a contig. The cp genome of *G. versteegii* obtained with GetOrganelle is composed of two inverted repeats (IRs) regions of 42.146 bp each, a large single copy region of 87.282 bp, and a small single copy region of 3.240 bp. The *A. malaccensis* cp genome is composed of a pair of inverted repeats (IRs) of 42.091 bp, separated by a small single copy (SSC) of 3.347 bp and a large single copy (LSC) of 87.293 bp. The chloroplast genome consists of several genes, which are categorized based on their function in the plastid as described

Table 1 Gene composition of the complete chloroplast genomes of *Gyrinops versteegii* and *Aquilaria malaccensis* obtained from annotation with Geseq

Category and genef unctions	*G. versteegii* and *A. malaccensis* genes content
Gene for photosynthesis	
Subunits of photosystem II	*psbA , psbB, psbC, psbD, psbE, psbF, psbH, psb1, psbJ, psbK, psbL, psbM, psbT, psbZ*
Subunits of Cytochrome b/f complex	*petA, petB, petD, petG, petL, petN,*
Subunit of ATP synthase	*atpA, atpB, atpE, atpF*, atpH, atpI,*
Subunit of rubisco	*rbcL*
ATP-dependent protease subunit P	*clP1*
Maturase	*matK*
Self-replicating system	
Ribosomal RNA genes	*rrn4.5, rrn4.5/ rrn4.5S, rrn5S, rrn5, rrn5S/rrn5, rrn16/rrn16S (2), rrn23/ rrn23S (2)*
Transfer RNA genes	*trnA-UGC*(2), trnC-ACA*, trnC-GCA, trnD-GUC(2), trnE-UUC*, trnF-GAA , trnG-UCC, trnG-GCC, trnG-GCC|trnG-UCC, trnH-GUG (2), trnI-CAU (2), trnI-GAU* (2), trnK-UUU*, trnL-UAG (4), trnL-CAA (2), trnL-UAA*, trnM-CAU (2), trnM-CAU (4), trnN-GUU (2), trnP-UGG (2), trnQ-UUG, trnR-ACG (4), trnR-UCU, trnS-CGA*, trnS-GCU (2), trnS-UGA, trnS-GGA, trnT-UGU, trnT-GGU (2), trnV-GAC(2), trnV-UAC*, trnW-CCA (2), trnY-GUA*
Smal subunit of ribosome	*rps2, rps3, rps4, rps7 (2), rps8, rps11, rps12 (3),rps14, rps15 (2), rps16, rps18, rps19*
Large subunit of ribosom	*rpl2*, rpl14, rpl16, rpl20, rpl22, rpl23 (2), rpl32, rpl33, rpl36,*
Subunit of NADH dehydrogenase	*ndhA*(2), ndhB* (2), ndhC, ndhD (2), ndhE (2), ndhF, ndhG (2), ndhH (2), ndhI (2), ndhJ, ndhK,*
DNA dependent RNA polymerase	*rpoA, rpoB, rpoC1*, rpoC2*
Subunit of photosystem I	*psaA, psaB, psaC (2), psaI, psaJ*
Other genes	
Photosystem assembly/RNA polymerase associated factor	*paf1*, pafII*
Photosystem biogenesis factor 1	*pbf1*
Subunit of acetyl-CoA carboxylase	*accD*
C-type cytochrome synthesis gene	*ccsA (2)*
Envelove membrane protein	cemA
Genes of unknown function	
Conserved open reading frames	*ycf1 (2), ycf2 (2), ycf15 (2)*

Note: *means that the genes at least have two segments, number in the bracket means the number of gene repeats present in LSC and IR.

Jurnal Manajemen Hutan Tropika, *30*(3), 326–335, December 2024
EISSN: 2089-2063
DOI: 10.7226/jtfm.30.3.326

Scientific Article
ISSN: 2087-0469

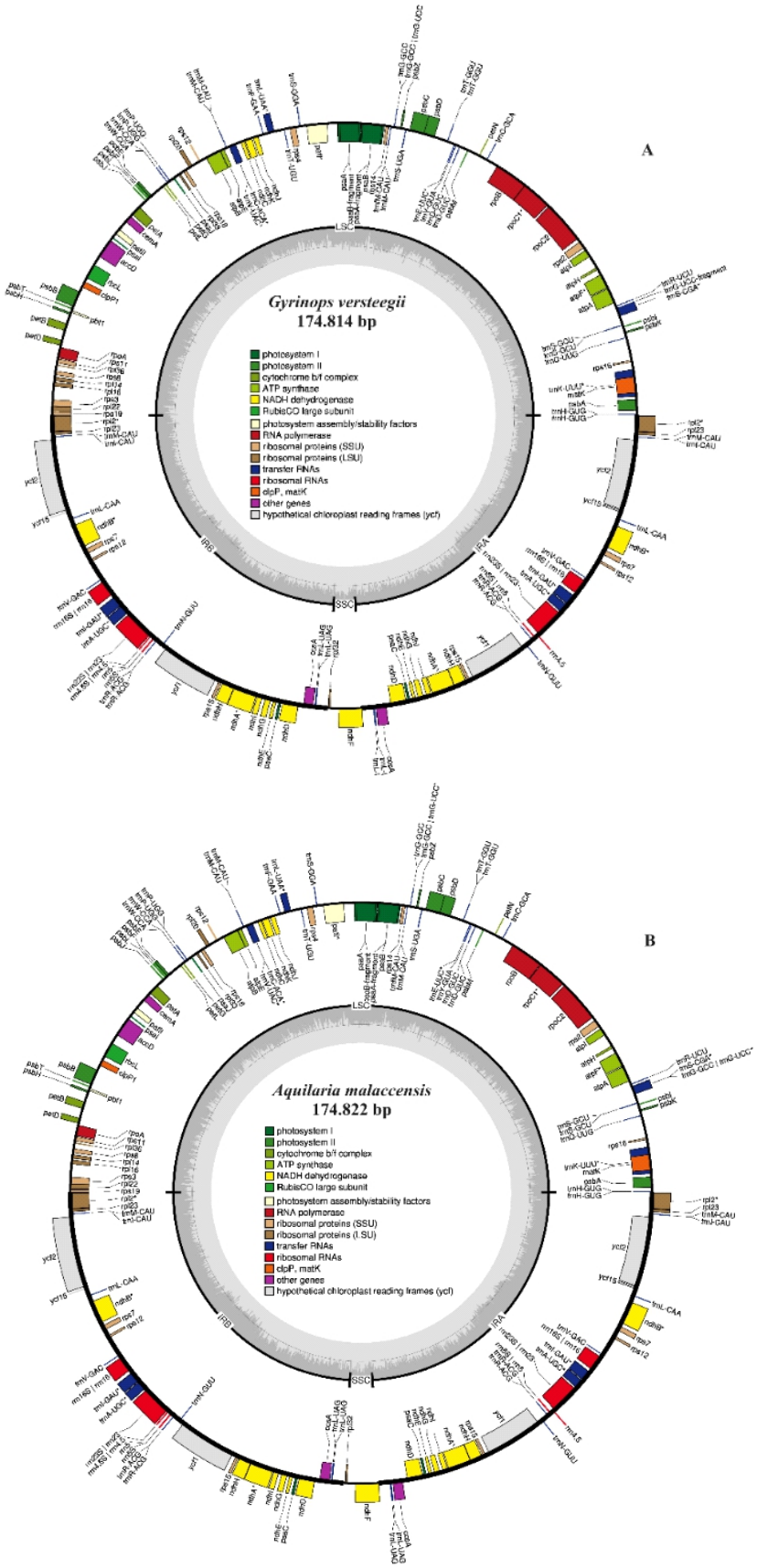

Figure 1 Structure of complete chloroplast genomes of two *Aquilarieae* species obtained with GetOrganelle: A) *Gyrinops versteegii* and B) *Aquilaria malaccensis.*

Jurnal Manajemen Hutan Tropika, *30*(3), 326–335, December 2024
EISSN: 2089-2063
DOI: 10.7226/jtfm.30.3.326

Scientific Article
ISSN: 2087-0469

in Table 1. Geseq annotation revealed that the cp genome of *G. versteegii* and *A. malaccensis* contains 99 protein-coding genes, 10 ribosomal RNA genes, and 66 tRNA with a GC content of 36.71%. Of these, 32 are duplicated in the IR region.

We also constructed the *G. versteegii* and *A. malaccensis* mitogenomes using the NOVOplasty and GetOrganelle assemblers. Both mitochondrial genomes obtained were still partial. The NOVOplasty generated a single large contig in the *A. malaccensis* assembly (Figure 2) and four contigs in the *G. versteegii* assembly. The longest mitogenome of *G. versteegii* obtained by NOVOplasty was 400. 012 bp in length with a GC content of 44.67% (the other three were successively sized from largest to smallest 207.797 bp, 85.509 bp, and 65.004 bp), whereas the total length of *A. malaccensis* mitogenome was 400.000 bp with a GC content of 44,53%. *G. versteegii* showed an 85.343% genome fraction, compared to 52.55% for *A. malaccensis*. Meanwhile, GetOrganelle was unsuccessful in generating large contigs. The only genome reference available for Aquilarieae tribe is the *A. sinensis* mitogenome, which has a length of 341.829 bp with a GC content of 45.01% (Wang & Cao, 2021).

The mitogenome contains numerous genes that are classified based on their function in mitochondria, as shown in Table 2. The mitogenomes of *G. versteegii* and *A. malaccensis* are both larger than the *A. sinensis* reference genome due to the existence of gene repeats forming a multipartite structure (Table 2). The *G. versteegii* mitogenome has a fourfold repeat of gene groups consisting of *orf116, orf 122, nad4, trnD-AUC, trnD-*(*GU*/*GUC*)*, trnD-GUC, trnH- GUG, trnC-GCA, trnS-GGA, nad1, nad4L, atp4, cox2, trnP-UGG, rps7, ccmGC, and trnA- FME*, while *A. malaccensis* has threefold repeats of gene groups consisting of *cox1, rps3, rpl16, sdh4, orf115a, orf119, nad6, nad3, rrn5, rrn18, atp8, rp15, rpl5, rps14, cob, atp1,* and *trnF*-AAA, and twofold repeats of gene groups consisting of *trnP UGG, trnF- GAA, trnS-GCU, rrn26, trnW-CCA, nad9, trnM-CAU, trnI-CAU, trnI-UAU, trnT-GGU, trnE-UUC, nad3, orf100-orf1, nad1, nad4, nad5, nad2, trnY-GUA, trnN-GUU,* and *trnC-GCA*.

Chloroplast genomes are highly conserved in terms of region, gene content, and gene order (Palmer, 1985). In contrast, mitogenomes show remarkable variations in genome size, gene content, gene organization, and structural complexity (Bullerwell, 2011; Wang et al., 2018; Yu et al., 2023). This variation has been identified even within the same species (Fauron et al., 2004). Its size is affected by the expansion of a non-coding sequence and a large repetitive part that forms a multipartite structure (Wu et al., 2022).

Fauron et al. (2004) discovered that the repetitive mitochondrial genome regions differ even among close relatives. Large repeats are considerably more common in larger mitochondrial genomes. For example, approximately

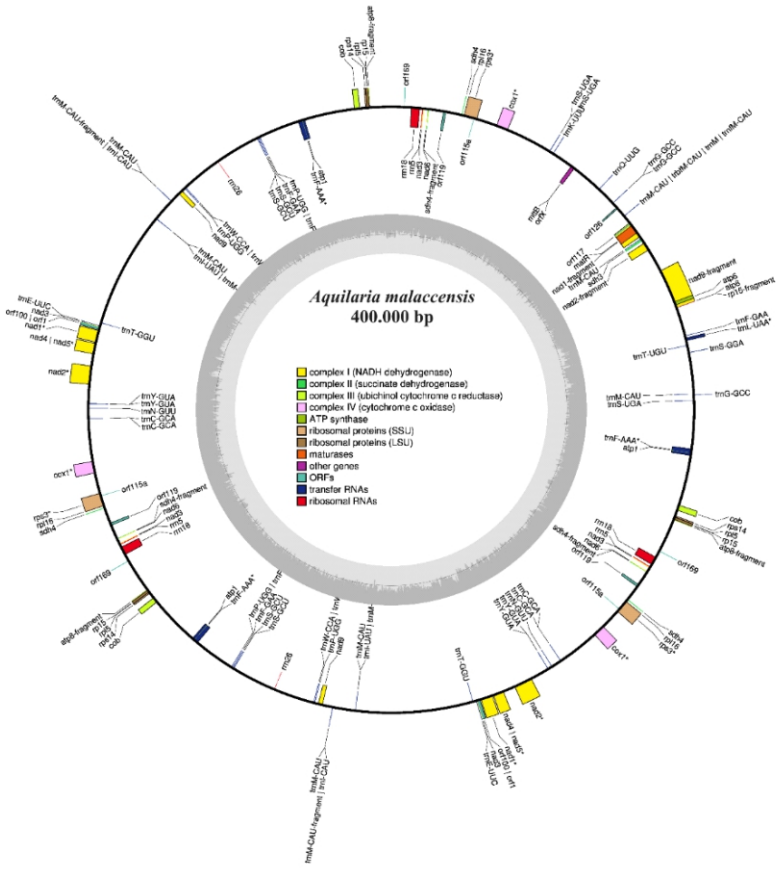

Figure 2 Structure of mitochondria genomes of *Aquilaria malaccensis* species obtain with NOVOplasty.

Jurnal Manajemen Hutan Tropika, *30*(3), 326–335, December 2024
EISSN: 2089-2063
DOI: 10.7226/jtfm.30.3.326

Scientific Article
ISSN: 2087-0469

Table 2 Gene composition of *Gyrinops versteegii* and *Aquilaria malaccensis* mitogenomes obtained from NOVOplasty assembly after annotation with Geseq

Category and gene functions	*G. versteegii species (from four contigs)*	*A. malaccensis (from one contigs)*
Core gene		
ATP synthase	*atp1, atp4 (4), atp6, atp8, atp9*	*atp1 (3), atp6, atp8 (4)*
Cytochrome c biogenesis	*ccmB, ccmC, ccmFN, ccmFc*(4),*	
Ubichinol cytochrome c reductase	*Cob*	*cob (3)*
Cytochrome c oxidase	*cox1*, cox2*(4), cox3*	*cox1*(3)*
Maturases	*matR*	*matR*
Transport membrane protein	*mttB*	*mttB*
NADH dehydrogenase	*nad1 (4), nad2*(2), nad3 (3), nad4* (5), nad4L (4), nad4| nad5*, nad5*,nad6, nad7*, nad9*	*nad1* (3), nad2* (2), nad3 (5), nad4| nad5* (2), nad6 (3),nad9(3)*
Variable genes		
The large subunit of ribosome	*rpl5, rpl10, rpl16, rp15*	*rpl5 (3), rpl16 (3), rp15 (4),*
The small subunit of ribosome	*rps3*, rps4, rps12, rps14, rps7 (4)*	*rps3* (3), rps14 (3),*
Succinate dehydrogenase	*sdh3(2), sdh4(2)*	*sdh3, sdh4 (3)*
Ribosomal RNA (rRNA) genes	*rrn5, rrn18, rrn26*	*rrn5 (3), rrn18 (3), rrn26 (2)*
Transfer RNA (tRNA) genes	*trnA-UGC*, trnC-GCA (4), trnD- GU-trnD-GUC (4), trnD-GUC (4), trnE-UUC* (3), trnF-AAA*,trnF- UGG, trnF-GAA (2), trnH-GUG*, trnG-GCC, trnH-GUG| trnH*, trnI-UAU, trnK-UUU, trnM-CAU(3), trnM-CAU| trnI-CAU, trnM-UAU, trnN- GUU (2), , trnR-ACG, trnS- GCU (2), trnS-UGA (2), trnS-GGA, trnY-GUA (2), trnV-GAC, trnW-CCA| trnW, trnQ-UUG, trnA- FME* (4), trbfM-CAU/trnM-CAU/trnM/trnfM-CAU*	*trnC-GCA (4), trnE-UUC (2), trnF-AAA* (3), trnF-GAA (3), trnG-GCC (2) , ,trnL-UAA, trnI-UAU| trnM-UAU (2), trnK-UUU*, trnL-UAA*, trnM-CAU (6), trnN-GUU (2), trnP-UGG| trnF-UGG (2), trnP-UGG (2), trnS-GCU (4), trnS-CGA (2), trnS-UGA (3), trnS-GGA (2), trnT-UGU (2), trnT-GGU* (2), trnY-GUA*(4), trnW-CCA| trnV trbfM-CAU/trnM-CAU/trnM/trnfM-CAU*

Note: * means that the genes at least have two segments, number in the bracket means the number of gene repeats present in mitochondria.

17% of the 570.000 bp *Zea mays* mitochondrial genome is made up of 8 repeats of 1 kb in length, while in its close relatives, more than 50% of the 360.000 bp *Oriza sativa* mitochondrial genome is made up of 16 repeats of 1 kb in length. BLAST comparisons of all the repeated sequences observed in the mitochondrial genome of *O. sativa* against the entire set of repeated regions found in other plant mitochondrial genomes, including *Z. mays*, revealed that only a very small fraction (less than 5%) of the rice repeated sequences were found to be similar to other plant mitochondrial genomes (Altschul et al., 1990). The presence of these numerous repeat regions makes it challenging to construct plant mitochondrial genomes using short-read sequencing techniques. This likely explains why the number of reported mitogenomes in plant species is still limited.

Studies of *Z. mays, O. sativa, Brassica napus, Beta vulgaris,* and *Marchantia polymorpha* mitogenomes have shown that gene content varies in all of these plant species. Some genes commonly found in plant mitochondria are absent. However, only four genes appear in all known mitochondrial genomes, including *cob, cox1, rRNA26S, and rRNA 18S*. Meanwhile, the presence of other genes varies greatly. As *A. sinensis* is the first and only mitogenome reported from *Aquilarieae,* we can only compare our results with this species. The gene content of *G. versteegii* and *A. malaccensis* mitogenomes are presented in Table 2. The largest contig of *G. versteegii* contains 47 CDS, 38 tRNA, and one rRNA gene, while the *A. malaccensis* mitogenome contains 72 CDS, 55 tRNA, and 10 rRNA. The number of genes is highly affected by the number of repetitions of gene groups. *A. sinensis* consists of 32 protein-coding sequences, 19 tRNA, and 3 rRNA (Wang & Cao, 2021).

GetOrganelle and NOVOplasty are recommended for organelle genome assembly due to their consistent performance (Frudenthal et al., 2020; Georgashvili et al., 2022). In our work, GetOrganelle generated an assembly that was comparable to the reference, but the SSC was flipped. The NOVOplasty assembly did not start with LSC but with IR and SSC, respectively. This result is in agreement with the report by Freudenthal et al. (2020). This finding showed that the software program used for cp genome assembly has a significant impact on the chloroplast genome assembly.

Jurnal Manajemen Hutan Tropika, *30*(3), 326–335, December 2024
EISSN: 2089-2063
DOI: 10.7226/jtfm.30.3.326

Scientific Article
ISSN: 2087-0469

Comparison of LSC/IR/SSC boundary in cp genomes A comparison of the LSC/IR/SC boundary regions of *A. malaccensis* and *G. versteegii using SnapGene Viewer showed that* NOVOplasty and GetOrganelle assembly results have several differences (Figure 3): Firstly, the *rps19* gene (269 bp) was extended by 16 bp across the LSC to the IRA region. Secondly, the position of *ndhF* and *rpl32* in the SSC region obtained from the NOVOplasty and GetOrganelle assemblies was reversed. The *ndhF* gene overlaps the IRA or IRB with the SSC boundary. In *G. versteegii,* this occurs between 52 bp in the IRA or IRB region and 2180 bp in the SSC region. In *A. malaccensis*, it is between 26 bp in the IRA or IRB region and 2212 bp in the SSC region. IRA and IRB are chloroplast inverted repeat regions that have identical gene counts and arrangements. Regional border differences have little effect on gene structure or arrangement. The IR region is one of the primary causes of changes in the size of the cp genome as a result of its expansion and contraction. The IR doubles the gene dosage of many chloroplast genes and controls flip-flop recombination, a defining feature of plastid genomes. The IR regulates the copy number of plastid genomes per cell and promotes their expression through a gene dosage effect (Krämer et al., 2024). Thirdly, the *rpl32* and *trnL* genes are located in the SSC and IR regions, respectively, at different distances from the SSC/IR border. There were no differences identified in the IR/LSC border zone.

Identification of highly variable sequences in the cp genome The nucleotide variability (Pi) values of the cp genomes were determined using an alignment generated by MAFFT v7 and DnaSP software. The Pi value of the *G.*

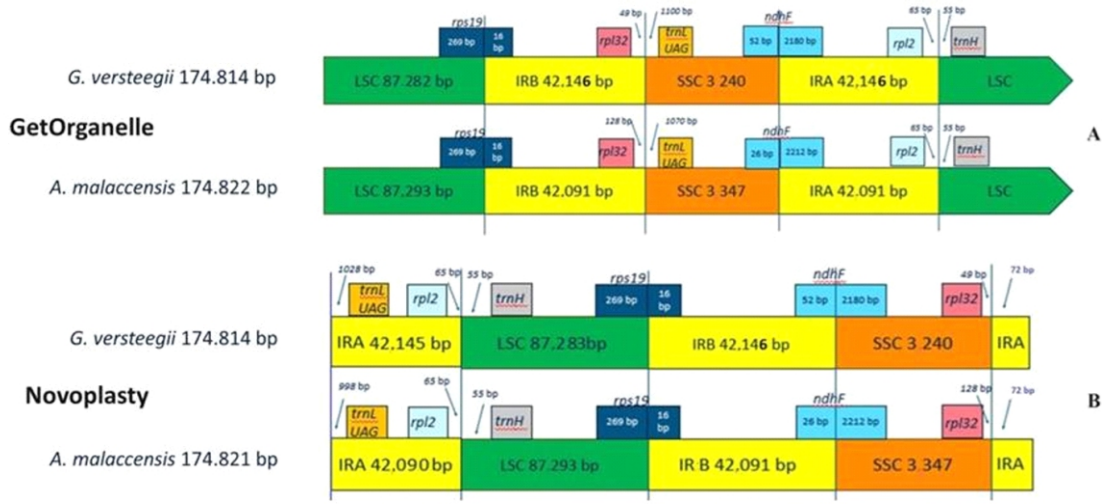

Figure 3 Comparison of the border regions of LSC, IR, and SSC between *Gyrinops verstegii* and *Aquilaria malaccensis* cp genomes obtained with NOVOplasty and GetOrganelle.

Table 3 The regions of highly variable sequences in the cp genomes of *Gyrinops versteegii* and *Aquilaria malaccensis*

Highly variable marker	Length (bp)	Variable sites	Nucleotide diversity (Pi)
matK-rps16	601	7	0.01167
rps16-trnQ	602	7	0.01167
trnL UAG-rpl32	654	18	0.03000
ndhF-rpl32	601	9	0.01500
rpoB-trnC GCA	604	6	0.01000
trnD-trnY	600	5	0.00833
trnT UGU-trnL UAA	606	10	0.01667
trnL-trnF	602	6	0.01000
ndhC-trnV	633	16	0.02667
rps18-rpl33	599	5	0.00833
rpl33-psaJ	599	6	0.01000
petL-psbE	608	6	0.01000
psbJ-petA	604	10	0.01667
petA-cemA	601	5	0.00833
cemA-pafII	599	5	0.00833
rpl36-rps8	603	6	0.01000

Jurnal Manajemen Hutan Tropika, *30*(3), 326–335, December 2024
EISSN: 2089-2063
DOI: 10.7226/jtfm.30.3.326

Scientific Article
ISSN: 2087-0469

versteegii and *A. malaccensis* cp genomes vary from 0 to 0.03. There are 16 highly divergent regions (Pi > 0.005), distributed between the intergenic spacer (IGS) region and the coding sequence (CDS) regions, as shown in Table 3 and Figure 4. We have identified 127 variable sites in 16 regions, with Pi values ranging from 0.0083 to 0.03. The *TrnL UAG-rpl32* has the greatest nucleotide variation (0.03). Our findings align with previous studies conducted by Hishamuddin et al. (2020), who reported that rpl32 showed the most nucleotide variation.

Phylogenetic analysis of chloroplast genomes We conducted a phylogenetic analysis using whole cp genomes to determine the evolutionary position of *G. versteegii* and A. *malaccensis* within the *Aquilarieae species*. The majority of branches have a high node support value (Figure 5). Our cp genomes clustered with *Aquilarieae* species, as expected. All *Aquilarieae* species were split into two classes, with *G. walla* in clade I, and *G. versteegii, A. malaccensis, A. sinensis,* and *A. crassna* in clade II. The evolutionary connection of *Aquilarieae* species suggests that *G. versteegii* is more closely linked to *A. malaccensis*. This finding correlates with

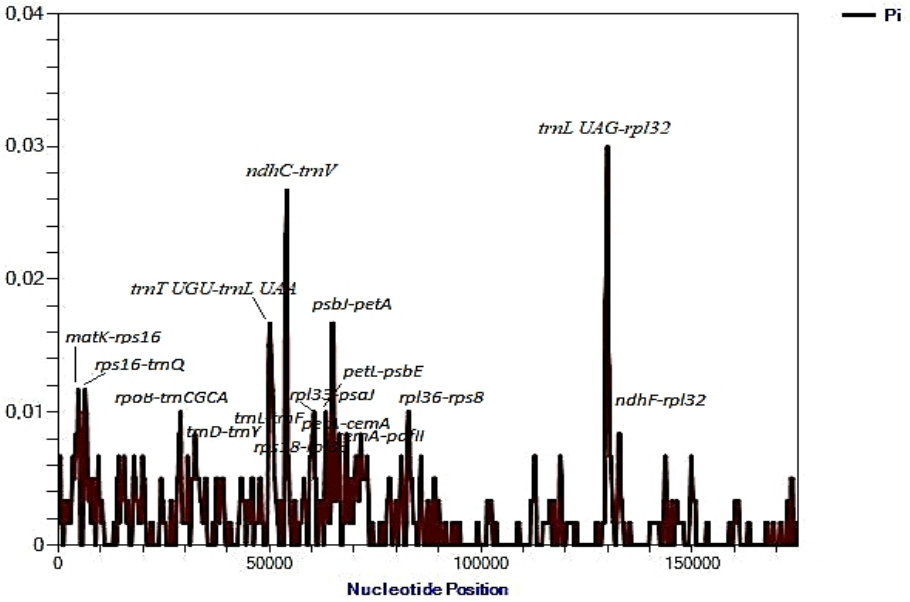

Figure 4 Chloroplast regions with the highest nucleotide diversity.

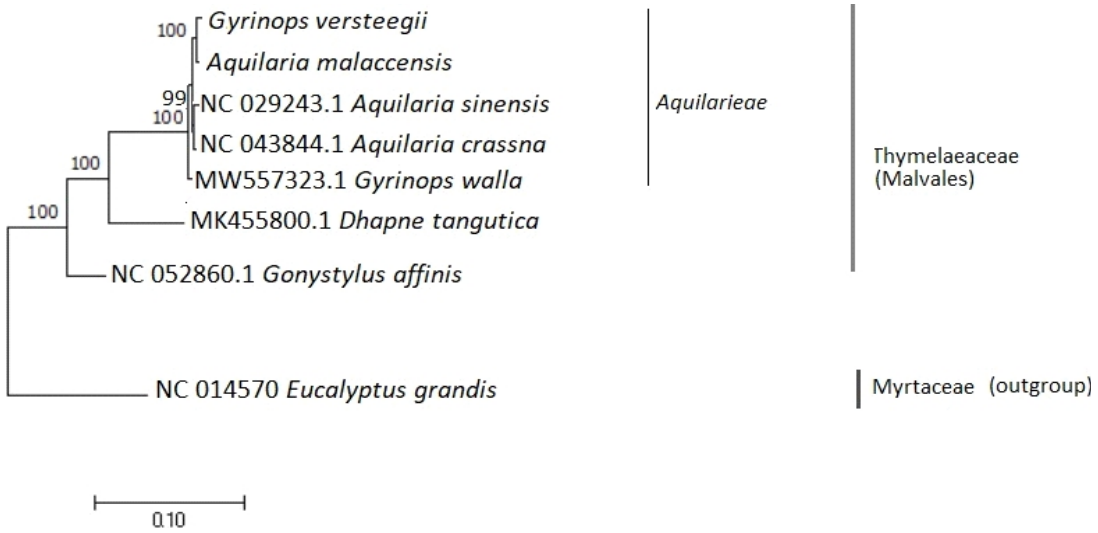

Figure 5 A phylogenetic tree depicting the relationships among *Aquilarieae* was constructed based on complete genome sequences. Maximum likelihood was used to estimate the phylogenetic inference of *Aquilarieae*. *Eucalyptus grandis* was used as the outgroup. The chloroplast tree shows molecular placement between species and within the *Aquilarieae* in Thymelaeaceae.

a previous study by Lee et al. (2022), who suggested that *Aquilaria* and *Gyrinops* should be treated as a single natural group, with *Gyrinops* merging into *Aquilaria* under certain conditions.

Conclusion

The complete cp genome assembly resulted in a circular quadripartite structure of 174.814 bp in *G. versteegii* and 174.822 bp in *A. malaccensis* with conserved genes and genome organization. Meanwhile, the mitogenome generated a multipartite structure of 400.012 bp in *G. versteegii* and 400.000 bp in *A. malaccensis.* with highly variation of genes and genome organization, which was affected by the presence of many repeat gene groups. A comparison of the LSC/IR/SCC region boundaries in chloroplasts shows that the IR region is one of the primary reasons for the change in the size of the cp genome. The nucleotide diversity (Pi) in the chloroplast genome shows that the *trnL UAG-rpl32* coding region has the highest nucleotide diversity (Pi = 0.03). The chloroplast and mitochondrial genome information will be useful in the future for a variety of downstream analyses, including barcoding and evolutionary studies.

Recommendation

Our genome data is currently confined to chloroplasts and mitochondria. To fully investigate the genome *of Aquilarieae*, deep sequencing needs to be performed to obtain the whole genome of the *Aquilarieae* species, which is currently not available in the database. Merging genomic and transcriptomic data will open up new possibilities for examining functional genes, as well as enhancing the capacity to use genome data for species identification, conservation, and a range of downstream analyses.

Acknowledgment

The authors would like to thank the Research Center for Applied Botany-National Research and Innovation Agency for providing permission to use the samples collected under the CITES project. This study was part of the project *Rumah Program Organisasi Riset Hayati dan Lingkungan*, funded by the National Research and Innovation Agency FY 2020 (9/III/HK/2022-RP1WB2-006).

References

Akter, S., Islam, M. T., Zulkefeli, M., & Khan, S. I. (2013). Agarwood production-a multidisciplinary field to be explored in Bangladesh. *International Journal of Pharmaceutical and Life Sciences, 2*(1), 22–32. https://doi.org/10.3329/ijpls.v2i1.15132

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Bullerwell, C. E. (Ed.). (2011). *Organelle genetics: Evolution of organelle genomes and gene expression*. Springer Science & Business Media.

Dierckxsens, N., Mardulyn, P., & Smits, G. (2017). NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research*, *45*(4), e18. https://doi.org/10.1093/nar/gkw955

Doyle, J. J., & Doyle, J. I. (1990). Isolation of plant DNA from fresh tissue. *Focus*, *12*(1), 13–15.

Farah, A. H., Lee, S. Y., Gao, Z., Yao, T. L., Madon, M., & Mohamed, R. (2018). Genome size, molecular phylogeny, and evolutionary history of the tribe Aquilarieae (Thymelaeaceae), the natural source of agarwood. *Frontiers in Plant Science*, *9*, 321994. https://doi.org/10.3389/fpls.2018.00712

Fauron, C., Allen, J., Clifton, S., & Newton, K. (2004). Plant mitochondrial genomes. In H. Daniell, & C. Chase, C. (Eds.), *Molecular biology and biotechnology of plant organelles: Chloroplasts and mitochondria* (pp. 151–177). Springer. https://doi.org/10.1007/978-1-4020-3166-3_6

Freudenthal, J. A., Pfaff, S., Terhoeven, N., Korte, A., Ankenbrand, M. J., & Förster, F. (2020). A systematic comparison of chloroplast genome assembly tools. *Genome Biology*, *21*, 254. https://doi.org/10.1186/s13059-020-02153-6

Hannon, G. J. (2010). FastX-toolkit. Retieved from http://hannonlab.cshl.edu/fastx_toolkit

Hartati, H., Pratama, R., Hartati, N., Siregar, U. J., Rahmawati, S., Ardiyani, M., Majiidu, M, & Siregar, I. Z. (2023). Phylogenetic study of *Gyrinops versteegii* (Gilg) Domke, the agarwood-producing tree from Indonesia. *AIP Conference Proceedings*, *2972*, 060011. https://doi.org/10.1063/5.0184218

Hishamuddin, M. S., Lee, S. Y., Ng, W. L., Ramlee, S. I., Lamasudin, D. U., & Mohamed, R. (2020). Comparison of eight complete chloroplast genomes of the endangered *Aquilaria* tree species (Thymelaeaceae) and their phylogenetic relationships. *Scientific Reports*, *10*(1), 13034. https://doi.org/10.1038/s41598-020-70030-0

Jin, J. J., Yu, W. B., Yang, J. B., Song, Y., dePamphilis, C. W., Yi, T. S., & Li, D. Z. (2020). GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology*, *21*, 241. https://doi.org/10.1186/s13059-020-02154-5

Krämer, C., Boehm, C.R., Liu, J. Yin-ting, M. K., Hertle, A. P., Former, J., Ruf, S., Schottler, M. A., Zoschke, R., & Bock, R. (2024). Removal of the large inverted repeat from the plastid genome reveals gene dosage effects and leads to increased genome copy number. *Nature Plants*, *10*, 923–935. https://doi.org/10.1038/s41477-024-01709-9

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7:

Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, *33*(7), 1870–1874. https://doi.org/10.1093/molbev/msw054

Lee, S. Y., & Mohamed, R. (2016). The origin and domestication of *Aquilaria*, an important agarwood-producing genus. In R. Mohamed (Ed.), *Agarwood. Science behind the fragrance* (pp. 1–20). Tropical Forestry. Springer. https://doi.org/10.1007/978-981-10-0833-7_1

Lee, S. Y., Turjaman, M., & Mohamed, R. (2018). Phylogenetic relatedness of several agarwood-producing taxa (Thymelaeaceae) from Indonesia. *Tropical Life Sciences Research*, *29*(2), 13. https://doi.org/10.21315/tlsr2018.29.2.2

Lee, S. Y., Turjaman, M., Chaveerach, A., Subasinghe, S., Fan, Q., & Liao, W. (2022). Phylogenetic relationships of *Aquilaria* and *Gyrinops* (Thymelaeaceae) revisited: Evidence from complete plastid genomes. *Botanical Journal of the Linnean Society*, *200*(3), 344–359. https://doi.org/10.1093/botlinnean/boac014

Librado, P., & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, *25*(11), 1451–1452. https://doi.org/10.1093/bioinformatics/btp187

Liu, H., Hou, Z., Xu, L., Ma, Q., Wei, M., Tembrock, L. R., Zhang, S., & Wu, Z. (2023). Comparative analysis of organellar genomes between diploid and tetraploid *Chrysanthemum indicum* with its relatives. *Frontiers in Plant Science*, *14*, 1228551. https://doi.org/10.3389/fpls.2023.1228551

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*, *17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200

Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D., & Gurevich, A. (2018). Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics*, *34*(13), i142i150. https://doi.org/10.1093/bioinformatics/bty266

Naziz, P. S., Das, R., & Sen, S. (2019). The scent of stress: Evidence from the unique fragrance of agarwood. *Frontiers in Plant Science*, *10*, 840. https://doi.org/10.3389/fpls.2019.00840

Palmer, J. D. (1985). Comparative organization of chloroplast genomes. *Annual Review of Genetics*, *19*(1), 325–354. https://doi.org/10.1146/annurev.ge.19.120185.001545

Rozewicki, J., Li, S., Amada, K. M., Standley, D. M., & Katoh, K. (2019). MAFFT-DASH: Integrated protein sequence and structural alignment. *Nucleic Acids Research*, *47*(W1), W5W10. https://doi.org/10.1093/nar/gkz342

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., & Greiner, S. (2017). GeSeq–versatile and accurate annotation of organelle genomes. *Nucleic Acids Research*, *45*(W1), W6W11. https://doi.org/10.1093/nar/gkx391

Tonti-Filippini, J., Nevill, P. G., Dixon, K., & Small, I. (2017). What can we do with 1000 plastid genomes? *The Plant Journal*, *90*(4), 808818. https://doi.org/10.1111/tpj.13491

Twyford, A. D., & Ness, R. W. (2017). Strategies for complete plastid genome sequencing. *Molecular Ecology Resources*, *17*(5), 858–868. https://doi.org/10.1111/1755-0998.12626

Wang, X., Cheng, F., Rohlsen, D., Bi, C., Wang, C., Xu, Y., Wei, S., Ye, Q., Yin, T., & Ye, N. (2018). Organellar genome assembly methods and comparative analysis of horticultural plants. *Horticulture Research*, *5*, 3. https://doi.org/10.1038/s41438-017-0002-1

Wang, Z. F., & Cao, H. L. (2021). The complete mitochondrial genome sequence of *Aquilaria sinensis*. *Mitochondrial DNA Part B*, *6*(2), 381–383. https://doi.org/10.1080/23802359.2020.1869609

Wu, Z. Q., Liao, X. Z., Zhang, X. N., Tembrock, L. R., & Broz, A. (2022). Genomic architectural variation of plant mitochondria-A review of multichromosomal structuring. *Journal of Systematics and Evolution*, *60*(1), 160–168. https://doi.org/10.1111/jse.12655

Yu, X., Wei, P., Chen, Z., Li, X., Zhang, W., Yang, Y., Liu, C., Zhao, S., Li, X., & Liu, X. (2023). Comparative analysis of the organelle genomes of three Rhodiola species provide insights into their structural dynamics and sequence divergences. *BMC Plant Biology*, *23*(1), 156. https://doi.org/10.1186/s12870-023-04159-1