Journal of Tropical Biodiversity and Biotechnology
Volume 09, Issue 02 (2024): jtbb87932
DOI: 10.22146/jtbb.87932

Short Communications

The Complete Chloroplast Genome of *Medinilla tapete-magicum* (Melastomataceae) from Sulawesi, Indonesia

Arief Priyadi[1], Paramita Cahyaningrum Kuswandi[2], Evy Yulianti[2], Risha Amilia Pratiwi[1], Ni Putu Sri Asih[3]*

1)Research Center for Applied Botany, Research Organization for Life Sciences and Environment; National Research and Innovation Agency (BRIN), Jl. Raya Jakarta - Bogor KM. 46 Cibinong, Bogor 16911, West Java, Indonesia.
2)Department of Biology Education, Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta, Jl. Colombo Yogyakarta No.1, Karangmalang, Depok, Sleman, Yogyakarta. Indonesia 55281
3)Research Center for Biosystematics and Evolution; Research Organisation for Life Sciences and Environment; National Research and Innovation Agency (BRIN), Jl. Raya Jakarta - Bogor KM. 46 Cibinong, Bogor 16911, West Java, Indonesia
* Corresponding author, email: nipu002@brin.go.id

Keywords:
Annotation
Assembly
NGS
Phylogeny
Super-barcode
Submitted:
12 August 2023
Accepted:
21 December 2023
Published:
24 May 2024
Editor:
Furzani Binti Pa'ee

ABSTRACT
In this study, the genome of an endemic Sulawesi's plant, *Medinilla tapete-magicum* was sequenced using Illumina NextSeq 500 and assembled the whole chloroplast genome. Results showed that the cpGenome is 155,602 bp in size with typical quadripartite structure of a large single copy (LSC) region (85,409 bp), a short single copy (SSC) region (16,629 bp), and a pair of inverted repeats (IRs) regions (26,782 bp). The cpGenome is composed of 132 genes, which consists of 87 protein coding genes, 37 tRNAs, and 8 rRNAs. The sliding window analyses showed that *psbB-psbH* and *ndhF-rpl32* can potentially be used as markers. Microsatellite motifs of mononucleotide A and T dominated in the cpGenome. The phylogenetic trees from the concatenated 76 shared protein coding gene sequences showed the *Medinilla* clade was monophyletic and *M. tapete-magicum* is a sister species in the SE Asian clade which contain *M. magnifica* and *M. speciosa*.

Copyright: © 2024, J. Tropical Biodiversity Biotechnology (CC BY-SA 4.0)

Medinilla Gaudich. is an epiphytic, climber, or terrestrial shrubs in the family Melastomaceae (Kartonegoro 2022) that consist of 380 species (POWO 2023). Its habitat ranges from tropical Africa to Madagascar, India, Sri Lanka, Southeast Asia, Taiwan, Melanesia, northern Australia, and the Pacific Islands (Kartonegoro 2022).

Medinilla belongs to the Sonerileae/Dissochaeteae clade in the Melastomaceae family. However, the understanding of the generic boundaries and phylogenetic relationships within this clade is limited, and there is uncertainty regarding the taxonomic validity of many of the genera (Zhou et al. 2019). The limited sampling of genera or species in most analyses causes the discussion of generic limits to be hindered (Zhou et al. 2022), especially in its molecular aspect. The utilization of multiple markers in nuclear and chloroplast DNA barcoding has not yet resolved the phylogenetic issues in *Medinilla*, as observed in Kartonegoro et al. (2021). Furthermore, Kartonegoro et al. (2021) and Veranso-Libalah et al. (2022) reported that *Medinilla* was polyphyletic, thus requiring further intensive molecular research. The complete sequences of chloroplast ge-

nomes offer a high resolution in reconstructing phylogenetic between species (Li et al. 2022) and proven effective in addressing challenging phylogenetic inquiries, as an example the genus *Pseudodissochaeta* (Zhou et al. 2019). *Pseudodissochaeta* was then suggested to be kept as a synonym of *Medinilla* (Chen & Renner 2007; Zhou et al. 2019). Kartonegoro et al. (2021) have reclassified *Pseudodissochaeta* to a distinctive genus from *Medinilla* based on molecular, anatomical, and morphological evidence.

Out of 380 species of *Medinilla* sequences, only 9 species with complete plastome are now available in the GenBank, indicating that chloroplast genome data for *Medinilla* is still limited (Veranso-Libalah et al. 2022). *Medinilla tapete-magicum* Cámara & Leret & Veldk. is a terrestrial shrub that is endemic to Sulawesi, Indonesia. This species has a unique characteristic that its flowers are supported on many leafless branches incrassate both on and above the ground surface forming a dense mat around the plant's base (Cámara-Leret & Veldkamp 2011). This study aims to characterise the complete chloroplast genome of *M. tapete-magicum* and assess its phylogenetic position based on the chloroplast genome sequence. The results may be used as a foundation for further studies of this species, including but not limited to phylogeny, biogeography, conservation genetics, bioprospecting, and may serve as the super-barcode for this endemic species.

Leaf sample was collected from a living collection (Collection number NS. 191, plot no XV.B.319) in Bali Botanic Garden which originated from Batusetan Hill, Enrekang, South Sulawesi, Indonesia. This plant was propagated via stem-cutting from XV.B.242 described in Cámara-Leret & Veldkamp (2011) in which the holotype collection was stored at the Naturalis Biodiversity Center, Leiden (van Balgooy 7557). The leaf was silica-gel dried and transported for DNA extraction.

The genomic DNA was extracted from the silica-gel dried leaf sample by using Quick-DNA HMW MagBead Kit Zymo D6060 (Zymo Research, CA, USA). The quality control procedures were DNA visualization in Agarose Gel Electrophoresis, quantification, and purity assessment by Nanodrop 2000 (Thermo Scientific, MA, USA), Qubit dsDNA HS Assay (Thermo Scientific, MA, USA), and DNA integrity Quality checked by TapeStation (Agilent, CA, USA). The gDNA was enzymatic sheared to match insert size of 350 bp. The sequencing library was prepared by PCR free xGen™ DNA Library Preparation (IDT, IA, USA). The sequencing was performed using an Illumina NextSeq 500 (PT Genetika Science, Tangerang, Indonesia) to generate approximately 10 Gb total output of 150 bp paired-end (PE) raw reads.

The raw reads were subjected to FastQC v. 0.11.9 analysis (Andrews 2010). Low quality reads and those with sequencing adapter residues were trimmed by Trimmomatic v. 0.39 (Bolger et al. 2014), by these following settings: ILLUMINACLIP:TruSeq3-PE-2.fa:2:26:10 SLIDINGWINDOW:5:26 LEADING:26 TRAILING:26 HEADCROP:5 MINLEN:36 AVGQUAL:26. The complete chloroplast genome sequence was *de novo* assembled from the trimmed raw reads by NOVOPlasty v. 4.3.1 (Dierckxsens et al. 2016) with default settings. The full cpGenome sequence of *M. speciosa* NC_068172 was used as the reference and the extracted *rbcL* sequence as the seed. To assess the coverage, trimmed raw reads were mapped back to the sequence of final assembly by Bowtie2 v.2.4.5 (Langmead & Salzberg 2012). The SAM output was converted to BAM with Samtools v.1.9 (Danecek et al. 2021) and visualized in Unipro UGENE v.45.1 (Okonechnikov et al. 2012) for exporting the coverage table. Automatic annotation of the circularized assembly was performed in CPGAVAS2 (Shi et al. 2019) followed by manual checks using Unipro

UGENE v. 45.1 (Okonechnikov et al. 2012) and NCBI Genomic Workbench v. 3.8.2 (Kuznetsov & Bollin 2021). CPGView was used (Liu et al. 2023) to check the annotation completeness. The sequence was deposited in the genbank with accession no. OQ831429. The raw reads supporting this study were deposited in the NCBI Sequence Read Archive (SRA) with BioProject accession number PRJNA979173, BioSample SAMN35566527, and SRA SRR24805950 respectively. OGDRAW v. 1.3.1 was used (Greiner et al. 2019) to visualize the cpGenome.

Twenty-six complete cpGenome sequences of Melastomataceae and those 3 of Myrtaceae as the outgroup were downloaded from the genbank. From these 29 sequences, as many as 76 of protein coding genes loci were extracted for the phylogenetic analyses. Prior to the analysis, multiple sequence alignment (MSA) was processed for each locus by Muscle implemented in Mega 11 (Tamura et al. 2021). The aligned sequences were concatenated (63,674 in length), then subjected to model test-ng (Darriba et al. 2019) and showed the best model of TVM+G4. The maximum likelihood (ML) phylogenetic analysis was performed using IQTree2 v. 2.2.0 (Minh et al. 2020) with 1000 ultrafast bootstrap option (Hoang et al. 2018) and the Bayesian inference (BI) analysis in MrBayes v. 3.2.7 (Ronquist et al. 2012). The generated trees were edited in FigTree v. 1.4.4 (Rambaut 2009). To explore regions the newly assembled chloroplast genome for molecular markers, four sequences (NC_068172, MK994885, NC_049130, and OQ831429) were selected for whole genome alignment (WGA) in Mauve (Darling et al. 2004). The MSA was subjected to DNA Polymorphism analysis (windows length 400, step size 100) in DNASP v.6.12.3 (Rozas et al. 2017). The nucleotide diversity (pi) threshold of 0.14 was determined to select the hotspot regions. The sequence of OQ831429 was also subjected to MISA-web v.2.1 (Beier et al. 2017) analysis with its default settings to characterise the microsatellite/simple sequence repeats (SSR) loci. The reported di- and tri-nucleotides motifs were checked in the previous 4 sequences MSA for polymorphism.

The trimmed data consisted of 75.6 million reads and 2.8 million reads assembled into the cpGenome sequence of 155,602 bp. The coverage analysis showed the assembly depth of 3,562 ± 843. The cpGenome shows typical quadripartite structure of a large single copy (LSC) 85,409 bp, short single copy (SSC) 16,629 bp, and a pair of inverted repeats (IRs) 26,782 bp (Figure 1). The GC content of the plastome was 36.99% which was slightly higher than the LSC (34.77%), while in the SSC was 30.78% and the IRs 42.47%. The annotations showed 132 genes (111 unique), which consisted of 87 protein coding genes (79 unique), 37 tRNA (28 unique), and 8 rRNA (4 unique) (Table 1). Of these 132 genes, 20 genes contained an intron and 2 genes with two introns. The cpGenome size of *M. tapete-magicum* is similar to those reported *M. speciosa*, *M. beamanii*, *M. amplectens*, *M. petelotii*, and *M. fengii* (Zhou et al. 2019), which ranged from 155,084 –155,841 bp. Longer figures of cpGenome's size ranged from 156,420-156,790 bp observed for *Pseudodissochaeta assamica*, *P. septentrionalis*, and *P. lanceata* which in their report, Zhou et al. (2019) listed as *M. assamica*, *M. septentrionalis*, and *M. lanceata*. This was because Zhou et al. (2019) followed a taxonomic treatment by Chen & Renner (2007) which then was revised (Kartonegoro et al. 2021; Zhou et al. 2022).

The DNASP analysis showed 2 (two) regions: *psbB-psbH* and *ndhF-rpl32* (Figure 2), potentially used as markers (nucleotide diversity value above 0.14). Microsatellite motifs of mononucleotide A and T dominated in the cp Genome, followed by dinucleotides AT and TA, and trinucleotides ATT (Table 2). Among the dinucleotides motifs, $(AT)_6$ in base

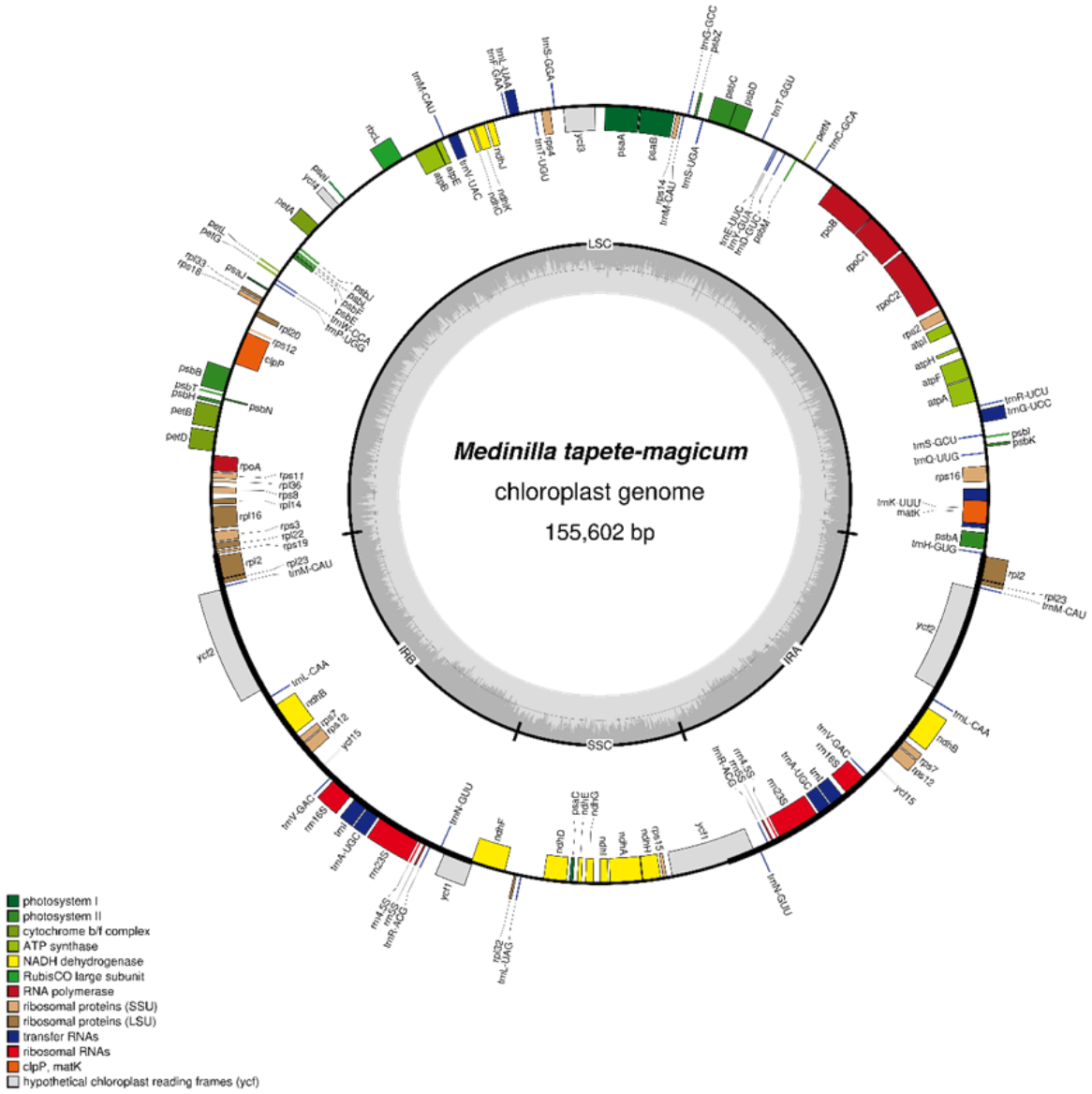

Figure 1. The complete chloroplast genome *of Medinilla tapete-magicum.*

31,931-31,942, $(AT)_7$ in base 51,618-51,631, and $(TA)_9$ in base 51,913-51,930 were polymorphic where as the solely trinucleotide motif $(ATT)_5$ was monomorphic.

The phylogenetic trees produced by the Maximum Likelihood (ML) and Bayesian Inference (BI) Methods had an identical topology and clades with strong bootstrap support (Figure 3). In this study, the *Medinilla* clade is monophyletic as a sister clade that contains *Blastus* and *Phyllagathis*. This result agrees with those previously reported by Zhou et al. (2019), Zhou et al. (2022) and Kartonegoro et al. (2021). *M. tapete-magicum* is a sister species of a clade contains *M. speciosa* and *M. magnifica*. This species is grouped within the South-East Asian clade and is separated from the Continental Asian counterpart (*M. fengii* and *M. petelotii*). Since *M. tapete-magicum* is the first sample of *Medinilla* genus from Indonesia with a cpGenome resource, other data of the congeneric species sampled from the region is needed to confirm their phylogenetic relationship. Additionally, Figure 3 shows that *M. assamica*, *M. septentrionalis*, and *M. lanceata* are renamed as *Pseudodissochaeta assamica*, *P. septentrionalis*, dan *P. lanceata* as reported by Zhou et al. (2022). This complete cpGenome sequence of *M. tapete-magicum* may serve as the super-barcode for this endemic plant species, as well as a source for DNA barcodes.

Table 1. Gene composition of *Medinilla tapete-magicum* chloroplast genome.

Group of genes	Name of genes	Number
tRNA	*trnA-UGC** (2x), *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnG-GCC*, *trnH-GUG*, *trnK-UUU**, *trnL-CAA* (2x), *trnL-UAA**, *trnL-UAG*, *trnM-CAU* (4x), *trnN-GUU* (2x), *trnP-UGG*, *trnQ-UUG*, *trnR-ACG* (2x), *trnR-UCU*, *trnS-GCU*, *trnS-GGA*, *trnS-UGA*, *trnT-GGU*, *trnT-UGU*, *trnV-GAC* (2x), *trnW-CCA*, *trnY-GUA*, *trnG-UCC**, *trnV-UAC**, *trnI** (2x)	37
rRNA	*rrn16S* (2x), *rrn23S* (2x), *rrn5S* (2x), *rrn4.5S* (2x)	8
Subunits of ATP synthase	*atpA*, *atpB*, *atpE*, *atpF**, *atpH*, *atpI*	6
Subunits of photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*, *ycf3***	16
Subunits of NADH-dehydrogenase	*ndhA**, *ndhB** (2x), *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*	12
Subunits of cytochrome b/f complex	*petA*, *petB**, *petD**, *petG*, *petL*, *petN*	6
Subunits of photosystem I	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*	5
Subunit of rubisco	*rbcL*	1
Large subunit of ribosome	*rpl14*, *rpl16**, *rpl2** (2x), *rpl20*, *rpl22*, *rpl23* (2x), *rpl32*, *rpl33*, *rpl36*	11
DNA dependent RNA polymerase	*rpoA*, *rpoB*, *rpoC1**, *rpoC2*	4
Small subunit of ribosome	*rps11*, *rps12*** (2x), *rps14*, *rps15*, *rps16**, *rps18*, *rps19*, *rps2*, *rps3*, *rps4*, *rps7* (2x), *rps8*	14
Subunit of Acetyl-CoA-carboxylase	*accD**	1
c-type cytochrome synthesis gene	*ccsA*	1
Envelope membrane protein	*cemA*	1
Protease	*clpP***	1
Maturase	*matK*	1
Conserved open reading frames	*ycf1 (2x)*, *ycf15 (2x)*, *ycf2 (2x)*, *ycf4*	7
Total		**132**

2x: presents in the IRs, * gene with an intron, ** gene with two introns

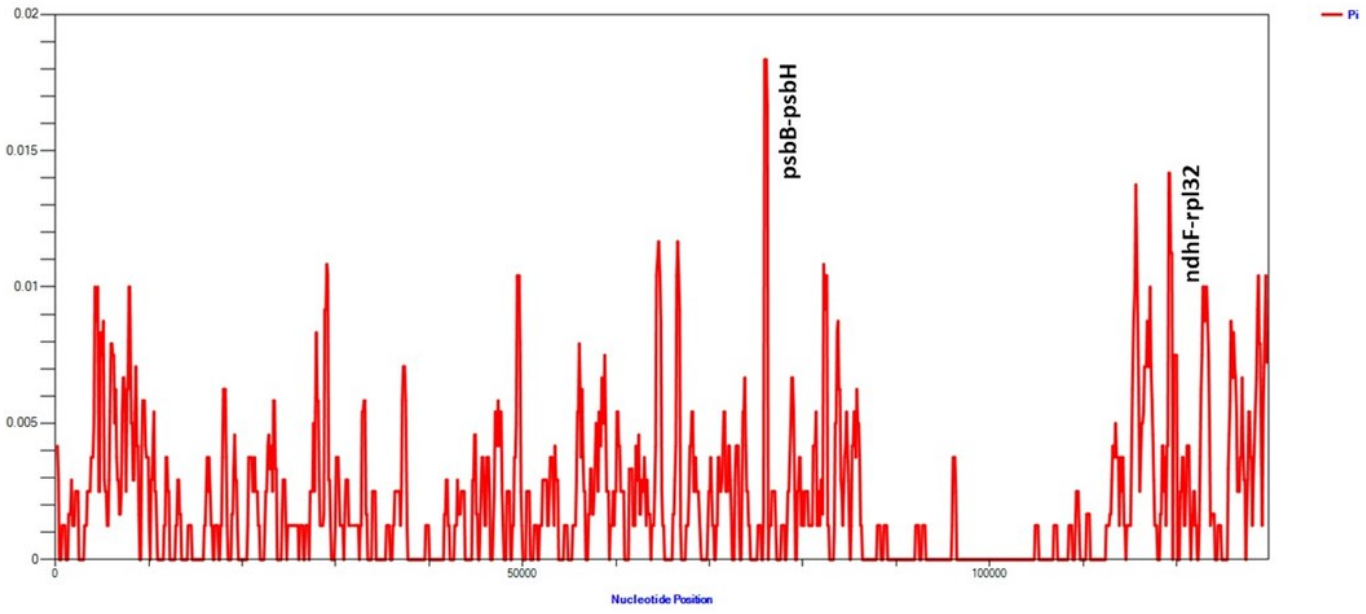

Figure 2. Nucleotide diversity value (pi) in *M. tapete-magicum.*

Table 2. Frequency distribution of the SSR repeat motif in the chloroplast genome of *M. tapete-magicum*.

Repeat motif	Number of repeats of the motif													
	5	6	7	8	9	10	11	12	13	14	15	16	17	total
A	-	-	-	-	-	12	7	-	2	-	-	1	-	22
T	-	-	-	-	-	12	7	4	-	2	1	-	1	27
AT	-	2	1	2	-	-	-	-	-	-	-	-	-	5
TA	-	-	-	-	1	-	-	-	-	-	-	-	-	1
ATT	1	-	-	-	-	-	-	-	-	-	-	-	-	1

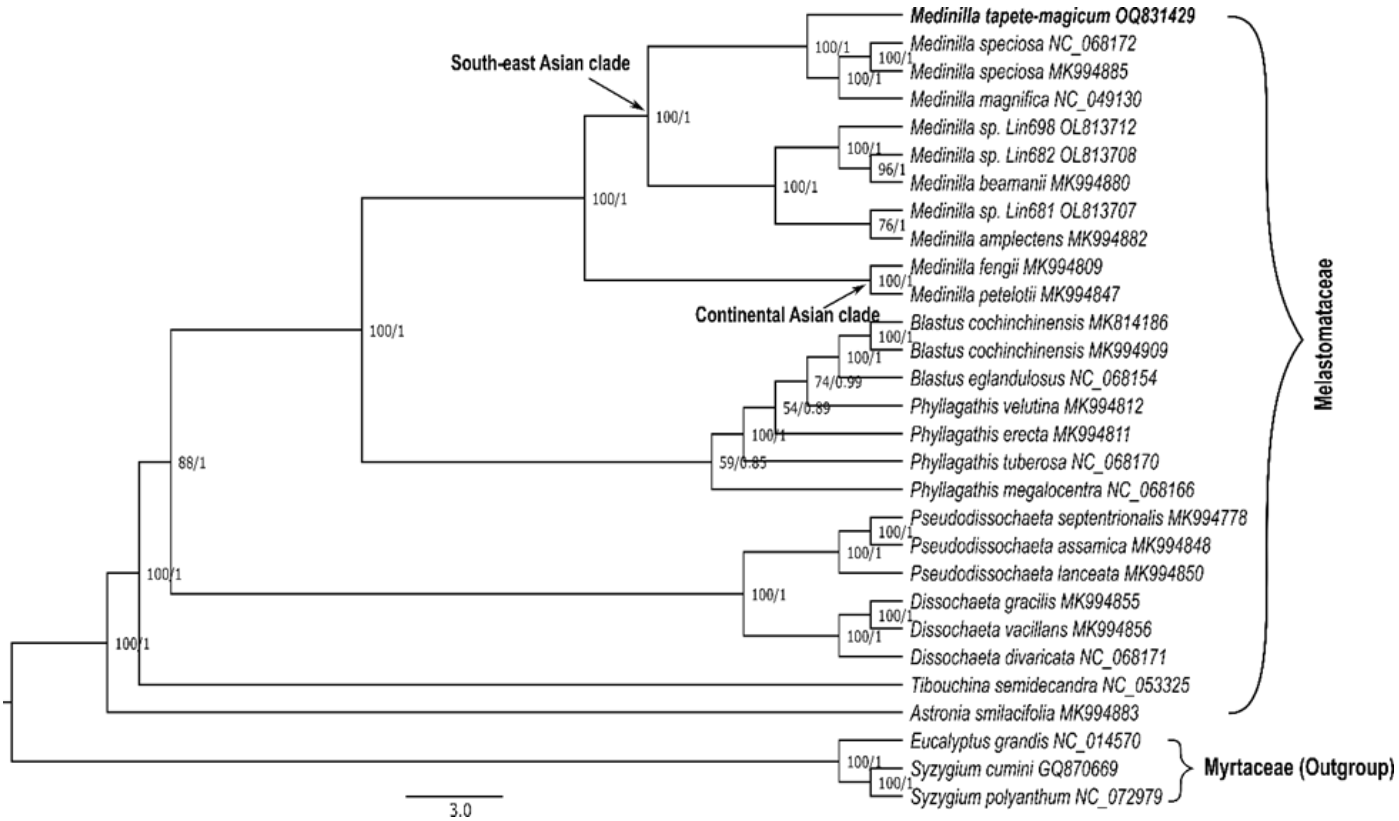

Figure 3. The Maximum Likelihood and Bayesian Inference Phylogenetic Tree of 26 species of Melastomataceae and 3 species of Myrtaceae (outgroup). Numbers in each node are the bootstrap value/posterior probability.

AUTHORS CONTRIBUTION

A.P. collected samples, analysed the data and supervised the manuscript, N.P.S.A. and R.A.P. wrote the manuscript, P.C.K. and E.Y. performed critical review and revision. A.P. is the main contributor, N.P.S.A., R.A.P., P.C.K. and E.Y. are the member contributors.

ACKNOWLEDGMENTS

The research was supported by *Joint Collaboration* Research House Program of the Organization for Biological and Environmental Research (ORHL), National Research and Innovation Agency (BRIN), No.39/III.5/HK/2022. The authors would like to thank the Directorate of Scientific Collection Management (DPKI-BRIN) with permission to collect leaf samples. The bioinformatic pipelines of this work were run in Mahameru High Performance Computing (HPC) facility by BRIN (HPC-BRIN). The authors thank to Chen Feng (Lushan Botanical Garden, PRC) for technical helps in the data analyses.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this study.

REFERENCES

Andrews, S., 2010. *FastQC: a quality control tool for high throughput sequence data*. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

Beier, S. et al., 2017. MISA-web: a web server for microsatellite prediction. *Bioinformatics*, 33(16), pp.2583-2585. doi:10.1093/bioinformatics/btx198.

Bolger, A.M., Lohse, M. & Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), pp.2114-2120. doi: 10.1093/bioinformatics/btu170.

Cámara-Leret, R. & Veldkamp, J., 2011. A remarkable new *Medinilla* (Melastomataceae) from Celebes (Sulawesi), Indonesia. *Gardens' Bulletin Singapore*, 62(2), pp.1-9.

Chen, J. & Renner, S., 2007. Melastomataceae. *Flora of china*, 13, pp.360-399.

Danecek, P. et al., 2021. Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2), giab008. doi: 10.1093/gigascience/giab008.

Darling, A.C.E. et al., 2004. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Research*, 14(7), pp.1394-1403. doi: 10.1101/gr.2289704.

Darriba, D. et al., 2019. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Molecular Biology and Evolution*, 37(1), pp.291-294. doi: 10.1093/molbev/msz189.

Dierckxsens, N., Mardulyn, P. & Smits, G., 2016. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research*, 45(4), pp.e18-e18. doi: 10.1093/nar/gkw955.

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(W1), pp.W59-W64. doi: 10.1093/nar/gkz238.

Hoang, D.T. et al., 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, 35(2), pp.518-522. doi: 10.1093/molbev/msx281.

Kartonegoro, A. et al., 2021. Molecular phylogenetics of the *Dissochaeta* alliance (Melastomataceae): Redefining tribe Dissochaeteae. *Taxon*, 70(4), pp.793-825. doi: 10.1002/tax.12508.

Kartonegoro, A., 2022. Annotated checklist of the *Medinilla* (Melastomataceae) of Malesia. *Rheedea*, 32, pp.221-279. doi: 10.22244/rheedea.2022.32.04.02.

Kuznetsov, A. & Bollin, C.J., 2021. NCBI Genome Workbench: desktop software for comparative genomics, visualization, and GenBank data submission. In *Multiple Sequence Alignment*. Humana Press, Springer, pp.261-295. doi: 10.1007/978-1-0716-1036-7_16.

Langmead, B. & Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), pp.357-359. doi: 10.1038/nmeth.1923.

Li, B. et. al., 2022. Complete chloroplast genome sequences of three Aroideae species (Araceae): lights into selective pressure, marker development and phylogenetic relationships. *BMC Genomics*, 23(1), 218. doi: 10.1186/s12864-022-08400-3.

Liu, S. et al., 2023. CPGView: A package for visualizing detailed chloroplast genome structures. *Molecular Ecology Resources*, 23(3), pp.694-704. doi: 10.1111/1755-0998.13729.

Minh, B.Q. et al., 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, 37(5), pp.1530-1534. doi: 10.1093/molbev/msaa015.

Okonechnikov, K., Golosova, O. & Fursov, M., 2012. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*, 28(8), 1166. doi: 10.1093/bioinformatics/bts091.

POWO, 2023, *Plants of the World Online*, viewed 21 October 2023, from http://www.plantsoftheworldonline.org/.

Rambaut, A., 2009, 'FigTree v1. 3.1.' in *Molecular evolution, phylogenetics and epidemiology*, from http://tree.bio.ed.ac.uk/software/figtree/.

Ronquist, F. et al., 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, 61(3), pp.539-542. doi: 10.1093/sysbio/sys029.

Rozas, J. et al., 2017. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Molecular Biology and Evolution*, 34(12), pp.3299-3302. doi: 10.1093/molbev/msx248.

Shi, L. et al., 2019. CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Research*, 47(W1), pp.W65-W73. doi: 10.1093/nar/gkz345.

Tamura, K., Stecher, G. & Kumar, S., 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, 38(7), pp.3022-3027. doi: 10.1093/molbev/msab120.

Veranso-Libalah, M.C. et al., 2022. Phylogeny and systematics of the tribe Sonerileae (Melastomataceae) in Africa: a revised taxonomic classification. *Journal of Systematics and Evolution*, 61(4), pp.657-681. doi: 10.1111/jse.12921.

Zhou, Q. et al., 2019. Analyses of plastome sequences improve phylogenetic resolution and provide new insight into the evolutionary history of Asian Sonerileae/Dissochaeteae. *Frontiers in Plant Science*, 10(1477). pp.1-16. doi: 10.3389/fpls.2019.01477

Zhou, Q-J. et al., 2022. Out of chaos: phylogenomics of Asian Sonerileae. *Molecular Phylogenetics and Evolution*, 175, 107581. doi: 10.1016/j.ympev.2022.107581.