



OPEN ACCESS

EDITED BY

Yong Jia,
Murdoch University, Australia

REVIEWED BY

Eknath D. Ahire,
MET Bhujbal Knowledge City, India

*CORRESPONDENCE

Andi Salamah

✉ salamah@sci.ui.ac.id

Imam Civi Cartealy

✉ imam012@brin.go.id

RECEIVED 25 November 2025

REVISED 09 January 2026

ACCEPTED 14 January 2026

PUBLISHED 04 February 2026

CITATION

Cartealy IC, Wulandari DR, Tajuddin T, Salamah A, Bismantoko S, Nugraha I, Bissa SYC and Abinawanto A (2026) Genome survey of nutmeg (*Myristica fragrans*) from Indonesia: genomic resource for Myristicaceae. *Front. Plant Sci.* 17:1753954. doi: 10.3389/fpls.2026.1753954

COPYRIGHT

© 2026 Cartealy, Wulandari, Tajuddin, Salamah, Bismantoko, Nugraha, Bissa and Abinawanto. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Genome survey of nutmeg (*Myristica fragrans*) from Indonesia: genomic resource for Myristicaceae

Imam Civi Cartealy^{1,2*}, Dyah Retno Wulandari³, Teuku Tajuddin³, Andi Salamah^{2*}, Sahid Bismantoko¹, Ihsan Nugraha¹, Stevry Yushady Ch Bissa¹ and Abinawanto Abinawanto²

¹Research Center for Computation, National Research and Innovation Agency (BRIN), Bogor, Indonesia, ²Cellular and Molecular Mechanisms in Biological Systems (CEMBIOS) Research Group, Department of Biology, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Depok, Indonesia, ³Research Center for Applied Botany, National Research and Innovation Agency (BRIN), Bogor, Indonesia

KEYWORDS

Myristica fragrans, genome assembly, genome annotation, chloroplast genome, low-coverage WGS

1 Introduction

Nutmeg (*Myristica fragrans*), which belongs to Myristicaceae family, is a native plant of the Banda islands in Indonesia. Economically, this tree is a source for two distinct commercial spices: nutmeg and mace, which are derived from its seed and seed arils respectively. Nutmeg derived products, such as nutmeg seed extracts and essential oil, are also widely used in the food industry and in the development of pharmaceutical and medicinal products (Al-Rawi et al., 2024). For example, in the food industry, nutmeg has been used for its fragrance, although it has the potential to serve as a natural preservative to prevent foodborne pathogens (Shafi et al., 2025). In addition, nutmeg seed extracts and essential oil have been extensively studied for their medicinal properties, encompassing antibacterial, antifungal, anti-inflammatory, anticancer, antidiabetic, and antioxidant activities (Ashokkumar et al., 2022; Cruz et al., 2024; Valente et al., 2014; Jin et al., 2005; Rengasamy et al., 2018; Al-Rawi et al., 2023; Zhao et al., 2017; Nguyen et al., 2010; Adiani et al., 2015).

Despite the economic status of *M. fragrans* as commercial spices and its potential use in other areas such as medicine, genomic information of *M. fragrans* available on public repositories such as NCBI is very limited. The limitation of genetic information on *M. fragrans* hinders the research development and its sustainable conservation. Thus, information on the genomic resource of *M. fragrans* can be beneficial for further studies of *M. fragrans* that rely on genomic or genetic information. In order to unravel and drafting the genomic landscape of *M. fragrans*, we performed a low-coverage genomic sequencing of *M. fragrans* using Illumina sequencing. The genomic data had been assembled and annotated, and it provides the complete chloroplast genome and partial genome

sequences of *M. fragrans*. Although the genome was partially sequenced, the data had been deposited in public repositories, adding more genomic and genetic information of *M. fragrans* and allowing public access for the benefit of future studies of *M. fragrans*.

2 Materials and methods

2.1 Plant material, DNA extraction, and sequencing

Young leaves of *Myristica fragrans* were collected in Bogor, Indonesia (6°38'29.64"S, 106°46'29.10"E). Genomic DNA was extracted from young leaves using the CTAB protocol (Doyle and Doyle, 1990) with modifications. The sequencing library was prepared using an Illumina DNA Prep Kit according to the manufacturer's protocol. The quantity and quality of the library were assessed using a QubitTM 1X dsDNA High Sensitivity Assay Kit. Finally, the library sample was loaded onto a NovaSeq 6000 (Illumina, San Diego, California) at the Genomic Facility of the National Research and Innovation Agency (BRIN), Indonesia, using a NovaSeq 6000 SP v1.5 kit (300 cycles) to generate 2x150 bp paired-end reads.

2.2 Genome assembly and annotation

The raw reads were processed by Fastp v0.23.2 (Chen, 2023) to remove low-quality sequences and adaptors, and to correct bases in overlapping pair-end reads. The k-mer distribution analysis was performed to estimate the genome size and heterozygosity. The frequency of k-mers was calculated using KMC v 3.2.4 (Kokot et al., 2017) with a k-mer size of 21. Then, the genome size and heterozygosity rate were estimated using GenomeScope v2.0 (Vurture et al., 2017).

The *de novo* genome assembly was performed using multiple genome assembly tools. SPAdes v4.0.0 (Bankevich et al., 2012) and Ray v2.3.1 (Boisvert et al., 2010) were used to assemble the cleaned reads with their default parameters. To produce more contiguous and complete contigs, the assembled contigs from two assemblies were merged using Redundans v2.0.1 (Pryszcz and Gabaldón, 2016). BUSCO (Manni et al., 2021) and Quast (Gurevich et al., 2013) were used to assess the quality of assembly.

Contaminant sequences were removed from the merged contigs using BlobToolKit v4.3.11 (Kumar et al., 2013; Challis et al., 2020). BlobToolKit requires the consensus, coverage, and taxonomy information of the assembly. Thus, as input to BlobToolKit, we provided the contigs, the alignment of reads to the contigs generated by BWA-MEM v0.7.17 (Li and Durbin, 2010), and taxonomy information of the contigs by searching the contigs against the NCBI non-redundant (nt) database using MMseqs2 v17-b804f (Steinegger and Söding, 2017), respectively. We identified alien contigs as any contigs that do not belong to the Streptophyta phylum nor those that do not match to any phyla.

To identify repeats, a custom repeat library for *M. fragrans* was created by combining the *de novo* repeat library generated with

RepeatModeler v2.0.7 (Flynn et al., 2020) and the Viridiplantae repeat library from Dfam database v3.9 (Storer et al., 2021). RepeatMasker v4.1.9 (Tarailo-Graovac and Chen, 2009) was then used to identify and mask the repeat sequences in the assembled contigs. Augustus was run with a pre-trained gene model of *A. thaliana* to predict genes in the contigs using the masked contigs as input. To annotate the genes, MMSeqs v18.8cc5c (Steinegger and Söding, 2017) was used to match the predicted genes with the curated plant protein database from Uniprot. GO term annotations were performed by matching Uniprot IDs with their GO annotations using Uniprot ID mapping.

2.3 Chloroplast genome assembly and annotation

GetOrganelle v1.7.7.1 (Jin et al., 2020) was used to assemble the chloroplast genome with default parameters. Annotations of the chloroplast genome were performed using GeSeq (Tillich et al., 2017). The chloroplast genome was visualized using OGDRAW v1.3.1 (Greiner et al., 2019).

3 Results

3.1 Genome sequence and annotation

The DNA sample of *M. fragrans* generated 250.59 million paired-end reads with total bases of 36.8 Gbp. After the filtration process, a total of 227.23 million of clean paired-end reads were yielded from the raw reads (Supplementary Table 1). K-mer analysis estimated a genome length of 324 Mbp for *M. fragrans* (Figure 1a), which is lower than the flow cytometry-based estimation of *M. fragrans* genome size, i.e., between 691 and 701 Mbp/1C (Kuo et al., 2024). This indicates that the genome of *M. fragrans* was partially sequenced, which is likely due to the low sequencing coverage. We estimated the average read coverage is 52x using the maximum estimated genome size of 701Mbp. The reads fit the GenomeScope v2 model for diploid species with repeat sequences and heterozygosity percentage of 33.5% and 1.24% respectively (Figure 1a).

To assemble the *M. fragrans* genome, we used an ensemble method which combines multiple assemblers to generate more contiguous assembly. Individually, Ray generated an assembly of 288 Mbp in 47,091 contigs, while SPAdes generated 324 Mbp in 75,430 contigs. We merged the outputs of Ray and SPAdes, and removed heterozygous contigs using Redundans to produce homozygous scaffolds. These steps generated a more homozygous and contiguous merged assembly of 301 Mbp in 41,225 contigs. However, we detected alien contigs in the 301 Mbp of merged assembly (Figure 1b). Alien contigs contribute to almost half (41.85%) of the total number of contigs, but they only account for 40 Mbp (13.51%) of the assembly length. The top three alien contigs in the merged assembly ranked by the number of contigs are Pseudomonadota (16.80%), Basidiomycota (12.22%), and

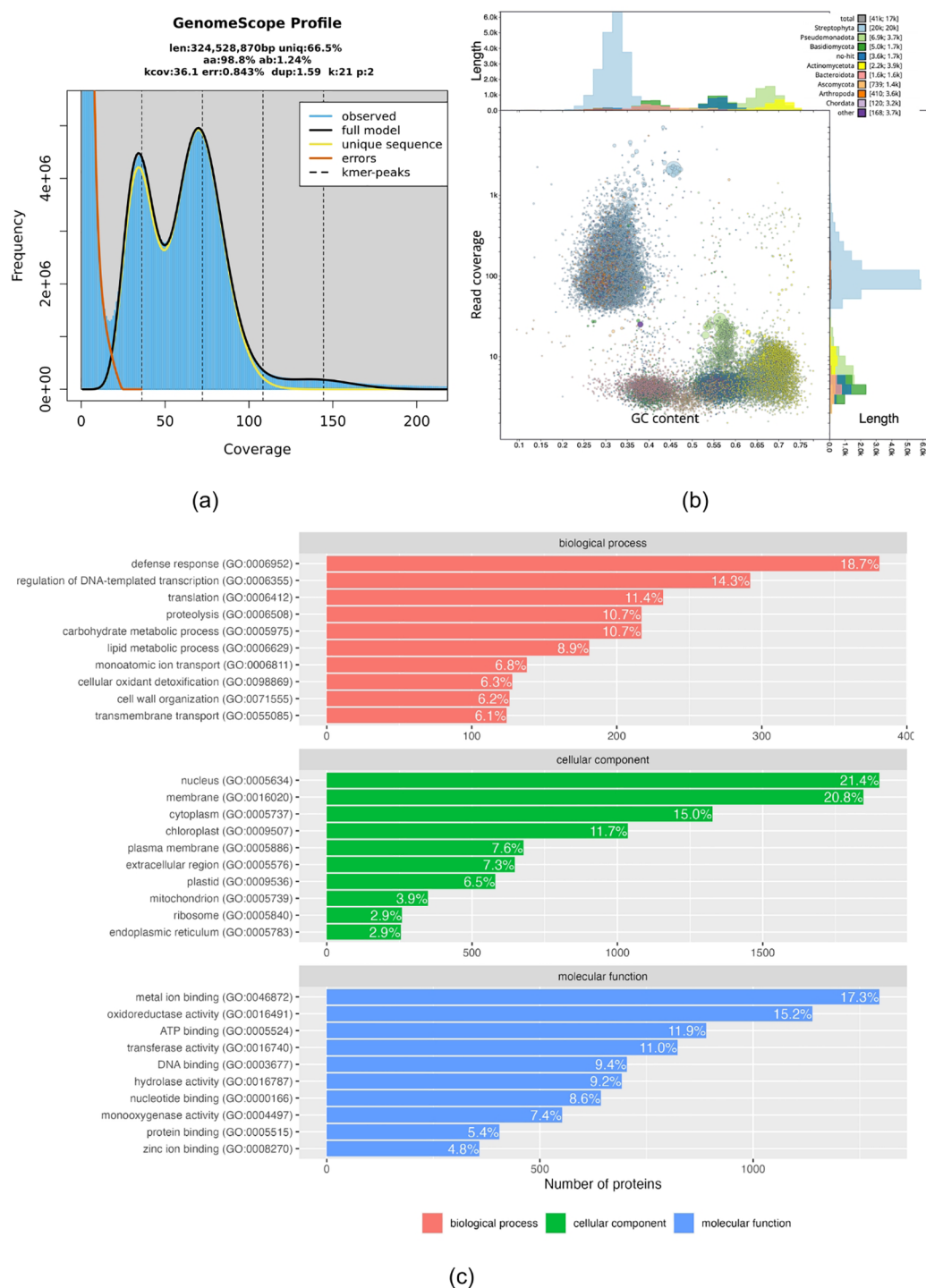


FIGURE 1

(a) The linier plot of k-mer coverage for 21-mer. The left peak around 40x coverage and right peak around 75x coverage are the heterozygosity and homozygosity peak respectively. **(b)** Blobplot of merged assembly from 41,225 contigs are shown. Light and dark blue contigs match to Streptophyta and 'no hit' respectively. Other colors represent contigs from various phyla. The top three alien contigs are shown in light green, dark green, and yellow blobs, which represent contigs from Pseudomonadota, Basidiomycota, and Actinomycetota respectively. **(c)** Bar plot of GO terms annotation of predicted proteins in *M. fragrans*. The GO terms are categorized into three ontologies, i.e. biological process, cellular component, and molecular function.

Actinomycetota (5.4%). After removing the alien contigs, the final assembly yielded a total of 260 Mbp in 23,973 contigs (Supplementary Table 2).

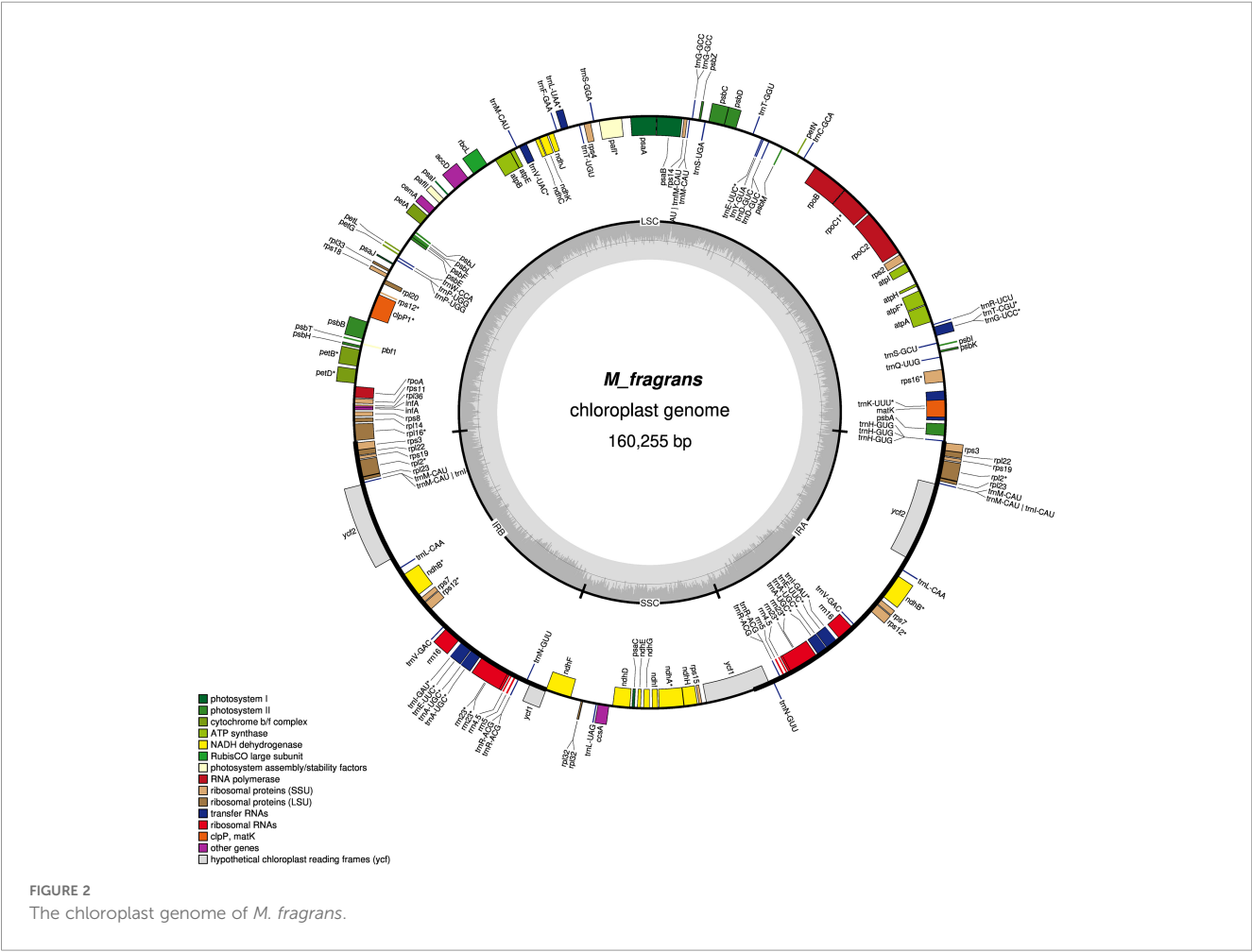
Augustus detected 145,534 exons in the masked assembly, which formed 30,340 proteins. To determine the identity of these proteins, we compare their sequences to the curated Uniprot plant protein sequences. Of the predicted proteins, 20,453 identity matches were found, representing 67.4% of the total predicted proteins. Figure 1c shows the distribution of GO terms of the identified proteins in the *M. fragrans* assembly.

3.2 Chloroplast genome sequence and annotation

The size of the chloroplast genome of *M. fragrans* is 160,255 bp (Figure 2). It has a typical quadripartite chloroplast structure

consisting of a pair of inverted repeats (IRs) of 28,403 bp separated two single copy regions, the large single copy (LSC) of 85,160 bp and small single copy (SSC) of 18,289 bp. The chloroplast genome contains 134 genes, which are categorized based on their function in the plastid as described in Supplementary Table 3. It consists of 8 ribosomal RNA genes, 37 transfer RNA genes, and 89 protein coding genes (PCG). Of these, 21 genes are duplicated in the IR regions.

In comparison to previously published chloroplast genome of *M. fragrans* (Cai et al., 2019), which reported the chloroplast genome size of 155kbp, the size of the chloroplast genome of *M. fragrans* reported in this publication is longer (Supplementary Table 4). We also identified two additional chloroplast genes, which are present in the chloroplast genome reported in this publication, i.e. the t-RNA gene *trnG-GCC* and the large subunit of ribosome gene *rpl22*, but are missing in chloroplast genome reported by (Cai et al., 2019).



Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/SRX30845511> <https://www.ncbi.nlm.nih.gov/genbank/PX562784> <https://figshare.com/> 10.6084/m9.figshare.30585914.

Author contributions

IC: Data curation, Project administration, Validation, Conceptualization, Methodology, Investigation, Supervision, Funding acquisition, Writing – original draft, Resources, Writing – review & editing, Software, Formal Analysis, Visualization. DW: Writing – review & editing, Resources. TT: Funding acquisition, Conceptualization, Resources, Writing – review & editing. AS: Writing – review & editing. SB: Writing – review & editing, Visualization. IN: Writing – review & editing. SYB: Writing – review & editing. AA: Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This work was partially supported by the Rumah Program Grant of Research Organization for Electronics and Informatics - National Research and Innovation Agency (BRIN) of Indonesia, number 1/III.6/HK/2023, and the Rumah Program Grant of Research Organization for Life Sciences and Environment - National Research and Innovation Agency (BRIN) of Indonesia number 9/III/HK/2022 and number 9/III.5/HK/2023.

Acknowledgments

The computations described in this paper were performed using the High Performance Computing Facility of the National Research

and Innovation Agency, Republic of Indonesia. The authors also acknowledge the Faculty of Mathematics and Natural Sciences, The University of Indonesia for its publication assistance program.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2026.1753954/full#supplementary-material>

References

- Adiani, V., Gupta, S., Chatterjee, S., Variyar, P. S., and Sharma, A. (2015). Activity guided characterization of antioxidant components from essential oil of nutmeg (*Myristica fragrans*). *J. Food Sci. Technol.* 52, 221–305. doi: 10.1007/s13197-013-1034-7
- Al-Rawi, S., Ibrahim, A., Nazari, M., Abdul Majid, A. S., Abdul Majid, A. M. S., Kadir, M., et al. (2023). Antiangiogenic and anticancer potential of supercritical fluid extracts from nutmeg seeds; *in vitro*, *ex vivo*, and *in silico* studies. *J. Angiotherapy* 7, 1–13. doi: 10.25163/angiotherapy.719371
- Al-Rawi, S. S., Ibrahim, A. H., Ahmed, H. J., and Khudhur, Z. O. (2024). Therapeutic, and pharmacological prospects of nutmeg seed: A comprehensive review for novel drug potential insights. *Saudi Pharm. J.* 32, 1020675. doi: 10.1016/j.jsps.2024.102067
- Ashokkumar, K., Simal-Gandara, J., Murugan, M., Dhanya, M. K., and Pandian, A. (2022). Nutmeg (*MYRISTICA FRAGRANS* houtt.) essential oil: A review on its composition, biological, and pharmacological activities. *Phytotherapy Res.* 36, 2839–2515. doi: 10.1002/ptr.7491
- Bankovich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Boisvert, S., Laviolette, F., and Corbeil, J. (2010). Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J. Comput. Biology: A J. Comput. Mol. Cell Biol.* 17, 1519–1335. doi: 10.1089/cmb.2009.0238
- Cai, C.-N., Ma, H., Ci, X., Conran, J., and Li, J. (2019). Comparative phylogenetic analyses of chinese *horsfieldia* (Myristicaceae) using complete chloroplast genome sequences. *J. Systematics Evol.* 59, 504–514. doi: 10.1111/jse.12556
- Challis, R., Richards, E., Rajan, J., Cochrane, G., and Blaxter, M. (2020). BlobToolKit – interactive quality assessment of genome assemblies. *G3 Genes[Genomes]Genetics* 10, 1361–1745. doi: 10.1534/g3.119.400908
- Chen, S. (2023). Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta* 2, e107. doi: 10.1002/imt2.107
- Cruz, A., Sánchez-Hernández, E., Teixeira, A., Oliveira, R., Cunha, A., and Martín-Ramos, P. (2024). Phytoconstituents and ergosterol biosynthesis-targeting antimicrobial activity of nutmeg (*Myristica fragrans* houtt.) against phytopathogens. *Molecules* 29, 4715. doi: 10.3390/molecules29020471

- Doyle, J. J., and Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. *Focus* 12, 13–15.
- Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., et al. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci.* 117, 9451–9457. doi: 10.1073/pnas.1921046117
- Greiner, S., Lehwark, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 47, W59–W64. doi: 10.1093/nar/gkz238
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086
- Jin, D.-Q., Lim, C. S., Hwang, J. K., Ha, I., and Han, J.-S. (2005). Anti-oxidant and anti-inflammatory activities of macelignan in murine hippocampal cell line and primary culture of rat microglial cells. *Biochem. Biophys. Res. Commun.* 331, 1264–1269. doi: 10.1016/j.bbrc.2005.04.036
- Jin, J. J., Yu, W. B., Yang, J. B., Song, Y., dePamphilis, C. W., Yi, T. S., et al. (2020). GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* 21, 241. doi: 10.1186/s13059-020-02154-5
- Kokot, M., Długosz, M., and Deorowicz, S. (2017). KMC 3: counting and manipulating k -mer statistics. *Bioinformatics* 33, 2759–2615. doi: 10.1093/bioinformatics/btx304
- Kumar, S., Jones, M., Koutsovoulos, G., Clarke, M., and Blaxter, M. (2013). Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. *Front. Genet.* 4. doi: 10.3389/fgene.2013.00237
- Kuo, Y. T., Kurian, J. G., Schubert, V., Fuchs, J., Melzer, M., Muraleedharan, A., et al. (2024). The holocentricity in the dioecious nutmeg (*Myristica fragrans*) is not based on major satellite repeats. *Chromosome Res.* 32, 8. doi: 10.1007/s10577-024-09751-1
- Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics* 26, 589–595. doi: 10.1093/bioinformatics/btp698
- Manni, M., Berkeley, M. R., Seppely, M., Simão, F. A., and Zdobnov, E. M. (2021). BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* 38, 4647–4545. doi: 10.1093/molbev/msab199
- Nguyen, P. H., Le, T. V. T., Kang, H. W., Chae, J., Kim, S. K., Kwon, K., et al. (2010). AMP-activated protein kinase (AMPK) activators from myristica fragrans (Nutmeg) and their anti-obesity effect. *Bioorganic Medicinal Chem. Lett.* 20, 4128–4131. doi: 10.1016/j.bmcl.2010.05.067
- Pryszcz, L. P., and Gabaldón, T. (2016). Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res.* 44, e113–e1135. doi: 10.1093/nar/gkw294
- Rengasamy, G., Venkataraman, A., Veeraraghavan, V. P., and Jainu, M. (2018). Cytotoxic and apoptotic potential of myristica fragrans houtt. (Mace) extract on human oral epidermal carcinoma KB cell lines. *Braz. J. Pharm. Sci.* 54, e18028. doi: 10.1590/s2175-97902018000318028
- Shafi, Z., Pandey, V. K., Habiba, U., Singh, R., Shahid, M., Rustagi, S., et al. (2025). Exploring the food safety and preservation landscape of myristica fragrans (L.) against foodborne pathogen: A review of current knowledge. *J. Agric. Food Res.* 19, 101639. doi: 10.1016/j.jafr.2025.101639
- Steinberger, M., and Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* 35, 1026–1285. doi: 10.1038/nbt.3988
- Storer, J., Hubley, R., Rosen, J., Wheeler, T. J., and Smit, A. F. (2021). The dfam community resource of transposable element families, sequence models, and genome annotations. *Mobile DNA* 12, 25. doi: 10.1186/s13100-020-00230-y
- Tarailo-Graovac, M., and Chen, N. (2009). Using repeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinf.*, 4.10.1–4.10.14. doi: 10.1002/0471250953.bi0410s25
- Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., et al. (2017). GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45, W6–11. doi: 10.1093/nar/gkx391
- Valente, V. M. M., Jham, G. N., Jardim, C. M., Dhinra, O. D., and Ghiviriga, I. (2014). Major antifungals in nutmeg essential oil against aspergillus flavus and A. Ochraceus. *J. Food Res.* 4, 515. doi: 10.5539/jfr.v4n1p51
- Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., et al. (2017). GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* 33, 2202–2204. doi: 10.1093/bioinformatics/btx153
- Zhao, Q., Liu, C., Shen, X., et al. (2017). Cytoprotective effects of myristicin against hypoxia-induced apoptosis and endoplasmic reticulum stress in rat dorsal root ganglion neurons. *Mol. Med. Rep.* 15, 2280–2288. doi: 10.3892/mmr.2017.6258