Journal of Natural Resources and Environmental Management
http://dx.doi.org/10.29244/jpsl.14.2.399

RESEARCH ARTICLE

OPEN ACCESS

Check for updates

Microsatellite Identification Based on Genome Assembly Reveals Potentials Marker of Macassar Ebony (*Diospyros celebica* Bakh.)

Muhammad Majiidu[a], Deden Derajat Matra[b], Fifi Gus Dwiyanti[c], Iskandar Z. Siregar[c]

[a] Tropical Silviculture Study Program, Graduate School, IPB University, IPB Darmaga Campus, Darmaga, Bogor, 16680, Indonesia
[b] Department of Agronomy and Horticulture, Faculty of Agriculture, IPB University, IPB Darmaga Campus, Dramaga, Bogor, 16680, Indonesia
[c] Department of Silviculture, Faculty of Forestry and Environment, IPB University, IPB Darmaga Campus, Dramaga, Bogor, 16680, Indonesia

Article History
Received 05 October 2023
Revised 08 December 2023
Accepted
10 December 2023

Keywords
Diospyros celebica, genomics, macassar ebony, Sulawesi

ABSTRACT

Macassar ebony (*Diospyros celebica* Bakh.) is an endemic tree species native to Sulawesi. To date, there are limited comprehensive publications regarding its genome assembly. In this study, we employed paired-end libraries of HiSeq 4000 Illumina, generating 141.2 million paired-end reads (42.4 Gigabases). Our analysis revealed the presence of 950,081 Scaffolds, with an N50 value of 6,023. Notably, BUSCO analysis identified 183 (12.7%) complete and single-copy BUSCOs (S), as well as 9 (0.6%) complete and duplicated BUSCOs (D). Furthermore, we identified 12,890 microsatellites within the Macassar Ebony genome. These microsatellites encompass 14 dinucleotide SSR motifs, 12,090 trinucleotide SSRs, 780 tetranucleotide SSR motifs, and 6 pentanucleotide SSRs. This dataset represents a valuable resource for assessing Macassar Ebony genetic makeup in its natural habitats and for subsequent analyses of the Macassar Ebony genome.

Introduction

Macassar Ebony (*Diospyros celebica* Bakh.), a native tree species of the Sulawesi region in Indonesia, belongs to the Ebenaceae Family and is well known for its exceptional production of high-value, luxurious timber [1–3]. This timber has remarkable versatility and useful applications in the construction, furniture, handicrafts, and manufacturing of musical instruments [4]. Notably, according to Koichi Mimura (Personal communication, 19 Februari 2021) a well-known Japanese musical instrument company has lauded Macassar Ebony wood as unparalleled due to its unique acoustic properties, intricate patterns, and exceptional durability. Despite its economic significance, unsustainable exploitation, illegal logging practices, and land conversion have collectively contributed to Sulawesi's substantial degradation of natural forests [5]. This degradation is intricately linked to the diminishing natural habitat of the Macassar Ebony [6]. This species has been classified as vulnerable by the International Union for Conservation of Nature (IUCN) Red List of Threatened Species since 1998 [7]. However, in 2018, Macassar Ebony was conspicuously omitted from the list of protected plant species by the Indonesian government, raising concerns that inadequate management could exacerbate its conservation status through illicit activity.

Due to its distinctive characteristics and ecological importance, concerted efforts are imperative to conserve the Macassar Ebony. Conservation strategies encompass both in-situ and ex-situ approaches, aiming to preserve their ecological, economic, and sociocultural roles within ecosystems. One notable endeavor involved forest restoration and conservation with a focus on Macassar ebony within its natural habitat or in other suitable areas, aligning with the government's commitment to reduce greenhouse gas emissions as articulated in Indonesia's 2030 Forestry and Other Land Uses (FOLU) net sink operational plan [8]. However, the success of ecosystem restoration and conservation initiatives, including extensive planting efforts, depends on several factors. Among these factors, the genetic background of seedlings plays a pivotal role in determining outcomes. A previous study [9], revealed that approximately 40% of reproductive materials for

Corresponding Author: Iskandar Z. Siregar siregar@apps.ipb.ac.id Department of Silviculture, Faculty of Forestry and Environment, IPB University, IPB Darmaga Campus, Darmaga, Bogor, Indonesia.

forest plants originate from fragmented populations with unknown genetic quality. Consequently, to ensure robust growth and bolster the effectiveness of ecosystem restoration and conservation, it is imperative to secure high-quality Macassar ebony seeds or seedlings with sound genetic foundations.

Visual differentiation between high-quality and low-quality planting stocks (e.g., seeds or seedlings) poses a great challenge. Therefore, developing alternative methods to identify superior planting stock materials is indispensable. Currently, genetic marker-based information regarding natural Macassar ebony populations in Sulawesi remains limited. As an initial step, acquiring genome assembly data and identifying microsatellite markers specific to Macassar ebony are essential prerequisites. The utilization of advanced whole-genome sequencing (WGS) technology presents a promising avenue for providing this crucial information [10]. The data derived from WGS can subsequently be harnessed to create molecular markers tailored for genetic population studies aimed at ecosystem conservation, including developing simple sequence repeat (SSR) markers [11]. Considering the limited genomic information, an experiment was conducted to generate datasets for genome assembly and the development of genetic marker identification.

Material and Methods

Plant Material

In this study, fresh young healthy leaves from a single tree of Macassar Ebony planted inside the IPB Darmaga campus, Darmaga, West Java, Indonesia (6.5568227 S, 106.7293918 E) were collected for DNA extraction. Approximately 2 × 2 cm of leaf was used for DNA extraction.

DNA extraction and Whole Genome Sequencing

Genomic DNA extraction from Maccasar Ebony leaf samples was meticulously carried out using the Zymo Plant & Seeds kit, a specialized kit designed by Zymo Research for the extraction of high-quality DNA from plant and seed samples. The extraction procedure strictly adhered to the manufacturer's protocol, ensuring the consistency and reliability of the results. The Zymo Plant & Seeds kit is renowned for its efficiency in isolating genomic DNA, employing optimized reagents and methodologies to yield pure DNA free from contaminants or inhibitors [12]. Following extraction, a comprehensive evaluation of the genomic DNA was performed using the Agilent Bioanalyzer 5,400 system to ascertain both its quality and quantity. The Agilent Bioanalyzer 5,400 system facilitated a detailed analysis of the DNA sample, providing information on the fragment size distribution and detecting potential issues such as degradation or contamination [13]. Subsequently, the Maccasar Ebony genomic DNA extract, which was confirmed to have optimal quality, was entrusted to Novogene for meticulous Polymerase Chain Reaction (PCR)-free library preparation. Library preparation is a critical step for the preparation of DNA for high-throughput sequencing. The final step involved sequencing the prepared library using an Illumina HiSeq 4000 sequencing system. Sequencing parameters were meticulously configured using 150 base-pair (bp) paired-end reads [10]. This setup allows the sequencing of DNA fragments from both ends, enhancing the accuracy and completeness of the generated genomic data.

Genome Assembly

The read data quality was assessed using FastQC software [14]. The genome assembly process was executed using the Maser Pipeline [15], employing SOAPdenovo (PE) 2.04 with default parameters [16,17]. Following this, the assembly quality was evaluated using Benchmarking Universal Single-Copy Orthologue (BUSCO) v.3.0.2, involving scaffold and contig assessment. Genome assembly was performed using SOAPdenovo software [18]. The draft genome then was evaluated for its quality and completeness using the BUSCO v3.0.2 software [19]. This evaluation included parameters such as complete and single-copy, completed and duplicated, fragmented, and missing BUSCOs. The evaluation process using BUSCO v3.0.2 software was based on expectations from the universal orthologous gene content in an evolutionary context [20]. The results of the assembly process were then annotated using annotation after assembling (PE) software, and homology was established against the UniProt and non-redundant nucleotides from the National Center for Biotechnology Information (NCBI) databases (accessed on 24 August 2020) available in the Maser platform.

Microsatellite Identification

Microsatellites were identified using MISA software [21]. This software, known for its accuracy in detecting short parallel repeats, was used with specific parameter settings to ensure complete results. The defined settings cover a wide range of repeat patterns, including 2-6, 3-5, 4-5, 5-5, and 6-5, thus covering a wide

range of repeat lengths and layouts. In addition, to account for the possibility of disruption of the microsatellite chain, a criterion of 100 bases was established.

Results and Discussion

Genome Assembly

Genome assembly is a complex and crucial process in genetics and bioinformatics. This technology enables researchers to study the genomic structure and function of an organism as well as to identify the genes involved in biological processes [10,22,23]. Detailed results of the Macassar Ebony genome assembly are presented in Table 1.

Table 1. Statistics information for genome assembly of *Diospyros celebica* Bahk.

Features	Raw Reads	
Number of reads	141,223,249	
Number of bases (G)	42.4	
GC contents	39.5%	
Q-score	> 30	
	Assembly	
	Scaffold	Contig
Number	950,081	1,816,751
Size includes N	559,317,402	377,972,204
Size without N	416,622,823	377,972,204
Mean Size (bp)	588	208
Median Size (bp)	100	114
Longest Seq (bp)	107,316	7,834
Shortest Seq (bp)	100	100
N50 (bp)	6,023	258
GC content (%)	40.26	40.43

Whole genome sequencing of Macassar ebony was performed using the Illumina HiSeq 4000 sequencing platform with parameters set at 150 bp paired-end reads. The raw sequences have been deposited in the DNA Data Bank of Japan (DDBJ) under the link at https://ddbj.nig.ac.jp/resource/sra-submission/DRA010964. There are 141.2 million paired-end reads with a total of 42.4 billion bases that have been sequenced on Macassar Ebony samples. Reads are small pieces of DNA sequences obtained from the sequencing. As shown in Table 1, 950,081 scaffolds were successfully constructed. The average scaffold size was 588 bp with a median value of 100 bp, indicating that many scaffolds were relatively short. The longest scaffold was 107,316 bp, whereas the shortest scaffold was 100 bp. The N50 scaffold value is 6,023 bp. The GC content was 40.26%. Additionally, 1,816,751 contigs were constructed for this analysis. The average contig size was 208 bp with a median of 114 bp. The difference in size between the average and median indicates significant length variation among the existing contigs. The longest contig was 7,834 bp, while the shortest contig was 100 bp. The N50 value for the ebony contigs was 258 bp, and its GC content was 40.43% (Table 1).

Sequencing efforts have generated substantial amounts of data, providing insights into the genetic composition and structure of the Macassar Ebony genome. The presence of scaffolds and contigs, along with their size distributions and N50 values, offers a valuable resource for further genomic research, including gene identification, functional analysis, and comparative genomics. Moreover, GC content information is relevant for understanding the genome composition and potential functional implications [10]. Furthermore, the assembled genome holds versatile utility, serving purposes such as comparative genomics [24], understanding genetic variation [10,25], functional genomics [26], evolutionary studies [10], disease research [27], phylogenetic [28], functional element prediction [29], biotechnology and genetic engineering [30], and precision medicine [31].

BUSCO Analysis

BUSCO is used to evaluate the completeness and quality of genome assembly by comparing genome sequences to a set of genes that are considered universal and conservative. The results of BUSCO analysis are presented in Table 2. A total of 183 scaffolds (12.7%) contained complete and single-copy BUSCOs (S) genes. This indicates that these scaffolds successfully assembled universally considered genes without duplication. Only 9 scaffolds (0.6%) contained complete and duplicated BUSCOs (D) genes. The duplication of BUSCO

genes can occur naturally within the genome. In addition, 136 scaffolds (9.4%) contained fragmented BUSCOs (F) genes (Table 2). This suggests that some BUSCO genes are not found in a single scaffold fragment, but are distributed across multiple fragments. There total of 1,112 scaffolds (77.2%) contained missing BUSCOs (M) genes. This could be due to a number of genes that were not present in the analyzed genome or errors in the genome assembly. In the analysis of ebony contigs, 130 contigs (9%) contained complete and single-copy BUSCOs (S) genes. Only 5 contigs (0.3%) contained complete and duplicated BUSCOs (D) genes. A total of 124 contigs (8.6%) contained fragmented BUSCOs (F) genes, and 1,181 contigs (82%) contained missing BUSCOs (M) genes (Table 2).

Table 2. Statistical information for BUSCO analysis of assembly Scaffold from *Diospyros celebica* Bahk.

	Scaffold	Contig
Complete and single-copy BUSCOs (S)	183 (12.7%)	130 (9%)
Complete and duplicated BUSCOs (D)	9 (0.6%)	5 (0.3%)
Fragmented BUSCOs (F)	136 (9.4%)	124 (8.6%)
Missing BUSCOs (M)	1,112 (77.2%)	1,181 (82%)

BUSCO analysis sheds light on the quality and completeness of the Macassar Ebony genome assembly. This revealed that while certain scaffolds and contigs effectively captured complete, single-copy genes, indicating a reasonably satisfactory assembly, the presence of fragmented and absent BUSCOs indicates room for improvement. Both the scaffold and contig assemblies displayed a high percentage of missing genes, which suggests the need for further refinement of the assembly or complementary genomic analyses to identify and potentially bridge gaps in the genome. Furthermore, the low percentage of duplicated genes suggests that excessive gene duplication was not a significant concern in this genome assembly. This outcome aligns with the findings of Anita et al. [32], who reported a similar result when constructing the genome assembly of *Falcataria moluccana* using the SOAPdenovo assembler. Their assembly comprised more than 82% of missing BUSCOs. It's worth noting that Illumina, a sequencing platform known for its high accuracy, still exhibits gaps within the assembly. To address this issue effectively, it is essential to combine both short-read and long-read technologies to fill these gaps [33–35].

Microsatellite Identification

The results of microsatellite (SSR) identification in the Macassar Ebony genome are detailed in Table 3. Furthermore, this analysis revealed the presence of SSR variations with different motifs. For instance, Macassar Ebony samples have detected SSRs with dinucleotide, trinucleotide, tetranucleotide, and pentanucleotide motifs. For example, there were 14 SSRs with dinucleotide motifs, 12,090 SSRs with trinucleotide motifs, 780 SSRs with tetranucleotide motifs, and six SSRs with pentanucleotide motifs (Table 3). The trinucleotide motif was significantly more abundant than other motifs. Similar results were observed in other dicotyledonous plants, including *Durio testudinarius* [36], *Lansium domesticum* [37], and *Mangifera casturi* [38]. Moreover, monocotyledonous plants such as rice, maize, and wheat have also demonstrated comparable results [39].

Table 3. Summary of identified microsatellites (SSR) from the assembly of *Diospyros celebica* Bahk.

Features	Number
Total number of identified SSRs	12,890
Number of SSR-containing sequences	10,051
Number of sequences containing more than 1 SSR	1,926
Number of SSRs present in compound formation	139
Motifs:	
Dinucleotide	14
Trinucleotide	12,090
Tetranucleotide	780
Pentanucleotide	6

The information obtained from this microsatellite (SSR) analysis has broad potential. Further research on the genetics and characteristics of bony tissues can be deepened based on these findings. Additionally, knowledge of SSR variations in ebony species can be applied in plant breeding. The selection of varieties

containing desired traits becomes more informed owing to a better understanding of their genetics. Apart from the benefits of plant breeding, this information is also relevant to biodiversity conservation efforts. With a deeper understanding of the genetic variation within Macassar Ebony populations, conservation measures can be designed more effectively. Efforts to preserve and restore vulnerable Macassar Ebony populations can be based on knowledge of their genetic structures. For example, a study conducted by Matra et al. [38] used SSR markers to study the genetic diversity of *Mangifera casturi*, which is categorized as an Extinct in the Wild (EW) species by IUCN [40]. Additionally, a study using an SSR marker was also reported by Dwiyanti et al. [41] aimed to understand the population genetic characteristics of *Dipterocarpus littoralis*, which is categorized as a Critically Endangered (CR) species by the IUCN [42].

Conclusion

In the present study, genome sequencing of Macassar Ebony (*Diospyros celebica* Bakh.) yielded a substantial dataset, comprising 141.2 million paired-end reads and a total of 42.4 billion sequenced bases, which facilitated the construction of 950,081 scaffolds. Within these scaffolds, we identified various categories of BUSCO genes: 183 scaffolds (12.7%) contained complete and single-copy BUSCOs (S), 9 scaffolds (0.6%) contained complete and duplicated BUSCOs (D), 136 scaffolds (9.4%) contained fragmented BUSCOs (F), and 1,112 scaffolds (77.2%) lacked identified BUSCOs (M). Furthermore, during our analysis of the ebony genome for simple sequence repeats (SSRs), we found that 14 contigs contained dinucleotide motifs, 12,090 contigs featured trinucleotide motifs, 780 contigs exhibited tetranucleotide motifs, and 6 contigs contained pentanucleotide motifs. This finding provides basic information for further assessments of the genetic diversity and structure of Macassar Ebony, which can be used for consideration in future conservation and cultivation strategies for this species.

Acknowledgment

The authors express their gratitude to the World Resources Institute (WRI) Forest Program for providing a grant (Project Code: 04682, Fiscal Year 2019–2020) entitled "Timber Tracking Project IPB-WRI: Collection of physical timber reference material and setting up a reference data building pipeline for DNA of commercial timber species *(Diospyros celebica*. Bakh)". The grant awarded by the Norwegian International Climate & Forest Initiative (NICFI), covers various research aspects, including research design, sample collection, laboratory expenses, sample preparation, sequencing, data collection, and analysis of results. The authors also would like to acknowledge the Ministry of Education, Culture, Research, and Technology for their support of this study through the National Competitive Basic Research Scheme 2023 (*Skema Penelitian Dasar Kompetitif Nasional 2023*) titled "Pilot Sequencing of 100 Native Forest Tree Genomes to Support Ecosystem Restoration (Rintisan Sekuensing 100 Genom Pohon Hutan Asli untuk Mendukung Restorasi Ekosistem)". The study is conducted under contract No.001/E5/PG.02.00.PL/2023 between DIKBUDRISTEK and IPB University, as well as contract No: 15850/IT3.D10/PT.01.02/P/T/2023 between Directorate of Research and Innovation IPB University and the Principal Investigator, Iskandar Z. Siregar.

References

1. Djarwanto, R.; Damayanti, J.; Balfas, E.; Basri; Jasni, I.M.; Sulastiningsih; Andrianto, D.; Martono, G.; Pari, A.; Sopandi. *Pengelompokan Jenis Kayu Perdagangan Indonesia*; FORDA Press: Bogor, ID, **2017**; ISBN: 978-602-6961-23-5.
2. Karlinasari, L.; Noviyanti, N.; Purwanto, Y.A.; Majiidu, M; Dwiyanti, F.G.; Rafi, M.; Damayanti, R.; Harnelly, E.; Siregar, I.Z. Discrimination and determination of extractive content of ebony (*Diospyros celebica* Bakh.) from Celebes Island by Near-Infrared Spectroscopy. *Forests* **2021**, *12*, 1–6, doi:10.3390/f12010006.
3. Siregar, I.Z.; Dwiyanti, F.G.; Pratama, R.; Matra, D.D.; Majiidu, M. Generating long-read sequences using Oxford Nanopore Technology from *Diospyros celebica* genomic DNA. *BMC Res Notes* **2021**, *14*, 1–4, doi:doi.org/10.1186/s13104-021-05484-0.
4. Riswan, S. Kajian biologi eboni (*Diospyros celebica* Bakh.). *Berita Biologi* **2002**, *6*, 211–218, doi:10.14203/beritabiologi.v6i2.1483.

5. Golar, G.; Muis, H.; Akhbar, A.; Khaeruddin, C. Threat of forest degradation in ex-forest concession right (HPH) in Indonesia. *Sustainability and Climate Change* **2022**, *15,* 216–223, doi:10.1089/scc.2022.0019.
6. Fatlan, K.; Pamoengkas, P.; Majiidu, M.; Siregar, I.Z. Tree species diversity on ebony habitat with different degradation levels in Sulawesi. In Proceedings of International Symposium on Arboriculture in the Tropics: Trees and Human Health, Bogor, Indonesia (Virtual), 21-22 June 2021.
7. World Conservation Monitoring Centre. *Diospyros celebica*. The IUCN Red List of Threatened Species 1998; e.T33203A9765120. Available online: https://www.iucnredlist.org/species/33203/9765120 (accessed on 04 December 2023).
8. KLHK (Kementerian Lingkungan Hidup dan Kehutanan). Rencana Operasional Indonesia's FOLU Net Sink 2030. 2022. Available on: https://pustandpi.or.id/download/rencana-operasional-indonesias-folu-net-sink-2030/ (accessed on 26 September 2023).
9. Jalonen, R.; Valette, M.; Boshier, D.; Duminil, J.; Thomas, E. Forest and landscape restoration severely constrained by a lack of attention to the quantity and quality of tree seed: insights from a global survey. *Conservation Letters* **2017**, *4*, 1–9, doi:10.1111/conl.12424.
10. Ng, K.K.S.; Kobayashi, M.J.; Fawcet, J.A.; Hatakeyama, M.; Paape, T.; Ng, C.H.; Ang, C.C.; Tnah, L.H.; Lee, C.T.; Nishiyama, T.; et al. The genome of *Shorea leprosula* (Dipterocarpaceae) highlights the ecological relevance of drought in aseasonal tropical rainforests. *Communication Biology* **2021**, *4*, 1–14, doi:10.1038/s42003-021-02682-1.
11. Burke, M.K.; Long, A.D. Perspective: What paths do advantageous alleles take during short-term evolutionary change? *Molecular Ecology* **2012**, *21*, 4913–4916, doi:10.1111/j.1365-294x.2012.05745.x.
12. Garfinkel, A.R.; Otten, M.; Crawford, S. SNP in potentially defunct tetrahydrocannabinolic acid synthase is a marker for Cannabigerolic acid dominance in *Cannabis sativa* L. *Genes* **2021**, *12*, 1–14, doi:10.3390/genes12020228.
13. Butler, T. Comparison of Constant- and Pulsed Field Electrophoresis Technologies for Analysis of High Molecular Weight and Large DNA Fragments. 2023. Available online: https://www.agilent.com/cs/library/applications/application-dna-fragment-size-automated-electrophoresis-5994-6103en-agilent.pdf (accessed on 26 September 2023).
14. Andrews, S. FastQC: a Quality Control Tool for High Throughput Sequence Data. 2010. Available online: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (accessed on 26 September 2023).
15. Kinjo, S.; Monma, N.; Misu, S.; Kitamura, N.; Imoto, N.; Yoshitake, K.; Gojobori, T.; Ikeo, K. Maser: one-stop platform for NGS big data from analysis to visualization. *Database* **2018**, 1–12, doi:10.1093/database/bay027.
16. Luo; Liu, B.; Xie, Y.; Li, Z.; Huang, W.; Yuan, J.; He, G.; Chen, Y.; Pan, Q.; Liu, Y. et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **2012**, *1*, 1–6, doi:10.1186/2047-217x-1-18.
17. Matra, D.D.; Ritonga, A.W.; Natawijaya, A.; Poerwanto, R.; Sobir; Siregar, U.J.; Widodo, W.D.; Inoue, E. Datasets for genome assembly of six underutilized Indonesian fruits. *Data in Brief* **2019**, *22*, 960–963, doi:10.1016/j.dib.2018.12.070.
18. Li, R.; Zhu, H.; Ruan, J.; Qian, W.; Fang, X.; Shi, Z.; Li, Y.; Li, S.; Shan, G.; Kristiansen, K.; et al. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Research* **2010**, *20*, 265–272, doi:10.1101/gr.097261.109.
19. Seppey, M.; Manni, M.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness. In *Gene Prediction. Methods in Molecular Biology*; Kollmar, M. Eds.; Humana Press: New York, USA, 2019; ISBN 978-1-4939-9172-3.
20. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212, doi:10.1093/bioinformatics/btv351.
21. Thiel, T.; Michalek, M.; Varshney, R.K.; Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet*. **2003**, *106*, 411–422, doi:10.1007/s00122-002-1031-0.

22. Matra, D.D.; Adrian, M.; Poerwanto, R. Underutilised Fruit Tree Genomes from Indonesia. In *Underutilised Crop Genomes*, *Compendium of Plant Genomes*; Chapman, M.A., Eds.; Springer: Cham, UK, 2022**;** ISBN 978-3-031-00847-4.

23. Nystedt, B.; Street, N.; Wetterbom, A.; Zuccolo, A.; Lin, Y.C.; Scofield, D.G.; Vezzi, F.; Delhomme, N.; Giacomello, S.; Alexeyenko, A.; et al. The Norway spruce genome sequence and conifer genome evolution. *Nature* **2013**, *497*, 579–584, doi:10.1038/nature12211.

24. The International Brachypodium Initiative. Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature* **2010**, *463*, 763–768, doi:10.1038/nature08747.

25. Gan, X.; Stegle, O.; Behr, J.; Steffen, J.G.; Drewe, P.; Hildebrand, K.L.; Lyngsoe, R.; Schultheiss, S.J.; Osborne, E.J.; Sreedharan, V.T.; et al. Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature* **2011**, *477*, 419–423, doi:10.1038/nature10414.

26. Wang, W.; Mauleon, R.; Hu, Z.; Mansueto, L.; Copetti, D.; Sanciangco, M.; Palis, K.C.; Xu, J.; Sun, C.; Fu, B.; et al. Genomic variation in 3.010 diverse accessions of Asian cultivated rice. *Nature* **2018**, *557*, 43–49, doi:10.1038/s41586-018-0063-9.

27. Xin, X.; He, S.Y. *Pseudomonas syringae* pv. *tomato* DC3000: A model pathogen for probing disease susceptibility and hormone signaling in plants. *Annual Review of Phytopathology* **2013**, *51*, 473–498.

28. Jiao, Y.; Wickett, N.; Ayyampalayam, S.; Chanderbali, A.S.; Landherr, L.; Ralph, P.E.; Tomsho, L.P.; Hu, Y.; Liang, H.; Soltis, P.S.; et al. Ancestral polyploidy in seed plants and angiosperms. *Nature* **2011**, *473*, 97–100, doi:10.1038/nature09916.

29. Kajikawa, M.; Sierro, N.; Kawaguchi, H.; Bakaher, N.; Ivanov, N.V.; Hashimoto, T.; Shoji, T. Genomic insights into the evolution of the nicotine biosynthesis pathway in tobacco. *Plant Physiol.* **2017**, *174*, 999–1011, doi:10.1104/pp.17.00070.

30. Shih, P.M.; Wu, D.; Latifi, A.; Axen, S.D.; Fewer, D.P.; Talla, E.; Calteau, A.; Cai, F.; Tandeau, M.N.; Rippka, R.; et al. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc Natl Acad Sci USA* **2013**, *110*, 1053–1058, doi:10.1073/pnas.1217107110.

31. Zou, C.; Chen, A.; Xiao, L.; Muller, H.M.; Ache, P.; Haberer, G.; Zhang, M.; Jia, W.; Deng, P.; Huang, R.; et al. A high-quality genome assembly of quinoa provides insights into the molecular basis of salt bladder-based salinity tolerance and the exceptional nutritional value. *Cell Res*. **2017**, *27*, 1327–1340, doi:10.1038/cr.2017.124.

32. Anita, V.; Matra, D.D.; Siregar, U.J. Chloroplast genome draft assembly of *Falcataria moluccana* using hybrid sequencing technology. *BMC Res Notes* **2023**, *16*, 1–4, doi:10.1186/s13104-023-06290-6.

33. Anita, V.P.D.; Siregar, U.J.; Matra, D.D. Analisis genomik dengan teknologi sekuensing secara hybrid (short-reads dan long-reads) pada Sengon (*Falcataria moluccana*). Thesis, Institut Pertanian Bogor Bogor, Indonesia, 2023.

34. Chen, Z.; Erickson, D.L.; Meng, J. Polishing the Oxford Nanopore long-read assemblies of bacterial pathogens with Illumina short reads to improve genomic analyses. *Genomics* **2021**, *113*, 1366–1377, doi:10.1016/j.ygeno.2021.03.018.

35. Chen, D.; Du, Y.; Fan, X.; Zhu, Z.; Jiang, H.; Wang, J.; Fan, Y.; Chen, H.; Zhou, D.; Xiong, C.; et al. Reconstruction and functional annotation of *Ascosphaera apis* full-length transcriptome utilizing PacBio long reads combined with Illumina short reads. *Journal of Invertebrate Pathology* **2020**, *176*, 107475, doi:10.1016/j.jip.2020.107475.

36. Magandhi, M.; Sobir; Kusumo, Y.W.E.; Sudarmono; Matra, D.D. Development and characterization of Simple SequenceRepeats (SSRs) markers in durian kura-kura (*Durio testudinarius* Becc.) using NGS data. In Proceedings of 4th International Conference on Biosciences (ICoBio 2021), Bogor, Indonesia, 11–12 August 2021.

37. Sari, H.P.; Efendi, D.; Suwarno, W.B.; Matra, D.D. Mining and characterization of genomic-based microsatellite markers in duku (*Lansium domesticum*). In Proceedings of The 3rd International Conference on Biosciences, Bogor, Indonesia, 8 August 2019.

38. Matra, D.D.; Fathoni, M.A.N.; Majiidu, M.; Wicaksono, H.; Sriyono, A.; Gunawan, G. Susanti, H.; Sari, R.; Fitmawati, F.; Siregar, I.Z.; et al. The genetic variation and relationship among the natural hybrids of *Mangifera castur*i Kosterm. *Sci. Rep.* **2021**, *11*, 19766, doi:10.1038/s41598-021-99381-y.

39. Morgante, M.; Hanafey, M.; Powell, W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat. Genet.* **2002**, *30*, 194–200, doi:10.1038/ng822.

40. Rhodes, L.; Maxted, N. *Mangifera casturi*. The IUCN Red List of Threatened Species 2016: e.T32059A61526819. Available online: https://www.iucnredlist.org/species/32059/61526819 (accessed on 18 November 2023).

41. Dwiyanti, F.G.; Harada, K.; Siregar, I.Z.; Kamiya, K. Population genetics of the critically endangered species *Dipterocarpus littoralis* Blume (Dipterocarpaceae) endemic in Nusakambangan Island, Indonesia. *Biotropia* **2014**, *21*, 1–12, doi:10.11598/btb.2014.21.1.1.

42. Hamidi, A.; Robiansyah, I. *Dipterocarpus littoralis*. The IUCN Red List of Threatened Species 2018: e.T33376A125628315. Available online: https://www.iucnredlist.org/species/33376/125628315 (accessed on 18 November 2023).