AGRIVITA Journal of Agricultural Science. 2023. 45(3): 430-442

AGRIVITA
Journal of Agricultural Science
www.agrivita.ub.ac.id

Generating Long-Read Sequences of Balsa (*Ochroma pyramidale* (Cav. ex Lam.) Urb.) Using Minion Oxford Nanopore Technology and Utilization for Phylogenetic Study

Siska Nurfajri[1)], Fifi Gus Dwiyanti[1)], Rahadian Pratama[2)], Muhammad Majiidu[3)], Iskandar Z. Siregar[1*)]

[1)] Department of Silviculture, Faculty of Forestry and Environment, IPB University, Bogor, West Java, Indonesia
[2)] Department of Biochemistry, Faculty of Mathematic and Natural Science, IPB University, Bogor, West Java, Indonesia
[3)] Advanced Research Laboratory (ARLab), IPB University, Bogor, West Java, Indonesia

ARTICLE INFO

Keywords:
matK
MinION
Ochroma pyramidale
Phylogeny
rbcL

Article History:
Received: September 22, 2022
Accepted: 6 July, 2023

*) Corresponding author:
E-mail: siregar@apps.ipb.ac.id

ABSTRACT

Balsa (*Ochroma pyramidale*) is fast-growing forest plant species introduced to Indonesia with limited genetic information. Genetic information can be obtained through molecular assessment which is now feasible due to sequencing technology development. This is supported by the third-generation sequencer technology, which has been developed using long-read sequencing technology. MinION Oxford Nanopore Technology is one of the long-read sequence-based sequencers with a real-time process and portable. This study aims to generate genomic data and analyze the phylogenetic relationship of balsa (*O. pyramidale*) based on long-read sequences with MinION Oxford Nanopore Technologies. Balsa long-read sequencing generated a partial chloroplast genome (cpDNA) sequence of 155,430 bp, which can be used for further DNA barcode-based phylogenetic analysis from the chloroplast genome. Phylogenetic analysis showed that the balsa species (*O. pyramidale*) was genetically grouped in one clade with other *O. pyramidale* species in phylogeny analysis based on *rbcL*, *matK*, and a combination of *rbcL* and *matK* genes indicated that *those genes* were a suitable marker for phylogenetic analysis in balsa species (*O. pyramidale*).

INTRODUCTION

Forests are an important source of diversity of native tree species that produce goods and services. Many forestry plant species in Indonesia are well-known to foreign countries due to their high economic and ecological value and are the object of intensive research. In addition to native species, there are also species originating from outside Indonesia, known as introduced or exotic species, where research on this species is equally important, especially related to aspects of adaptation, growth, and wood quality. One of the prospective tree species is balsa (*Ochroma pyramidale* (Cav. ex Lam.) Urb from Malvaceae family. Balsa has several synonyms names including *Ochroma bicolor*, *Ochroma lagopus*, and *Ochroma tomentosa*. Balsa is native to the Americas and has a natural distribution in Latin America covering the West Indies, southern Mexico, Central America, Venezuela, Colombia, Brazil, Ecuador, Peru, and Bolivia (Howcroft, 2002), however this species is widely planted on the Java Island in Indonesia (Pertiwi et al., 2017; Rachmat et al., 2019), especially in private-owned forest or forest plantation industry (Lisytianto et al.,2021). Balsa is known as the lightest tree species that grows fast so it is included in the fast-growing species (FGS). Balsa can grow to 20 m in 5-6 years with a diameter of 40 cm (Wijoyo et al., 2018).

Commercial balsa wood has a density ranging from 100 kg/m^3 to 170 kg/m^3, however, it can vary from 50 kg/m^3 to 410 kg/m^3 (Wijoyo et al., 2018) with

ISSN: 0126-0537

Cite this as: Nurfajri, S., Dwiyanti, F. G., Pratama, R., Majiidu, M., & Siregar, I. Z. (2023). Generating long-read sequences of balsa (*Ochroma pyramidale* (Cav. ex Lam.) Urb.) using minion oxford nanopore technology and utilization for phylogenetic study. *AGRIVITA Journal of Agricultural Science, 45*(3), 430-442.http://doi.org/10.17503/agrivita.v41i0.3930

an average modulus of elasticity (MOE) of 1,222 - 2,037 MPa and a modulus of rupture (MOR) 9.83–16.63 MPa (Kotlarewski et al., 2016). With such a density, balsa wood is widely used for surfboards, canoes, buoys, aeromodelling, and even bridges (Nuryamah, 2017). In addition, balsa wood can also be used as an insulator of heat, sound, and vibration, boat buoys, swimming buoys, sports equipment, aircraft buoys (hydroplane), lightweight short fiber pulp (Zanzibar, 2017), interior product (Wijoyo et al., 2018), and as one of the potential plants for reforestation (Istiqomah et al., 2017). Balsa species have not been studied much in Indonesia regarding their genetic aspects in the midst of technological advances 4.0 to support tree breeding programs.

The availability of genetic information can provide many benefits, especially to support genetic conservation and tree breeding program (Ide, 2021). In plants, this information is obtained from the chloroplast genome which contains genes that can be used as a study material for evolution at a high taxonomic level (Taberlet et al., 1991). Several genes located on cpDNA are known as universal markers and has been widely used to study genetic kinship relationships at high taxonomic levels including the large subunit of the *rbcL* (ribulose-bisphosphate carboxylase) (Mursyidin et al., 2021) and *matK* (Maturase K), which is a coding region with a high level of amino acid gene substitution and nucleotide level compared to *rbcL* (Mursyidin & Makruf, 2020). This genetic resource information is obtained by DNA sequencing, which is a technique used to sequence nucleotides in a DNA molecule (Roslim et al., 2015). DNA sequencing technology has been known since four decades ago and has entered the third generation sequencing technology, that is sequencers capable of producing long-read sequencing (long read DNA) (Dumschott et al., 2020). Long-read sequencing technology has the advantage to produce long and good-quality nucleotide sequence reads with a cheaper and faster analysis process (Bhalerao et al., 2003; Lu et al., 2016). In addition, it can solve the gap problem that occurs in short-read sequencing (Amarasinghe et al., 2020; Lee et al., 2019).

Therefore, long-read sequencing technology is an alternative solution for several researchers in plant DNA study, especially in trees, most of which are still limited regarding the availability of genetic information, especially endangered and high economic value of wood species (Vallée et al., 2016). One of the long-read sequence technologies is MinION Oxford Nanopore Technologies (https://nanoporetech.com) (Mikheyev & Tin, 2014). This technology has successfully generated genetic information in several tree species such as *Diospyros celebica* (Siregar et al., 2021), *Diospyros rumphii* (Salindeho et al., 2023), and *Dryobalanops aromatica* (Wahyuni et al., 2021). In general, this technology can be applied to balsa trees (*Ochroma pyramidale*). Furthermore, the specific objectives of this study were to generate genomic data and analyze the phylogenetic relationship of balsa (*Ochroma pyramidale*) based on long-read sequences with Oxford Nanopore Technologies (ONT). This study was expected to provide genomic information on balsa (*O. pyramidale*) in the form of chloroplast genome and species relationships obtained from a long-read-based sequencing, which is beneficial for further genetic variation and functional genes studies.

MATERIALS AND METHODS

Study Site and Sample Collection

This study was conducted from October to December 2020. The samples used in this study were leaves and wood tissues of balsa seedlings grown in the Greenhouse of the Department of Silviculture, Faculty of Forestry and Environment, IPB University, Bogor, West Java. Fresh leaf samples were collected from two individuals of one-month-old seedlings, which a height of about 30 cm and 25 cm, respectively, while the fresh wood tissue or cambium was obtained from one individual of two-month-old seedling which had a height of 93 cm. In total, three samples were used for further DNA extraction and sequencing.

Herbarium specimen of balsa was identified at the Herbarium Center of the Research and Development Center of the Ministry of Environment and Forestry, Gunung Batu, Bogor with the scientific name *Ochroma pyramidale* (Cav. ex Lam.) Urb. In addition, the laboratory works were carried out at the Laboratory of Forest Genetics and Molecular Forestry, Department of Silviculture, Faculty of Forestry and Environment, IPB University and also in the Advanced Research Laboratory, IPB University.

Samples Preparation and DNA Extraction

The wood samples were 100-150 mg, while the leaf samples were 50-100 mg. Genomic DNA of leaves and wood samples were extracted using

the modified CTAB (Cetyl Trimethyl Ammonium Bromide) method (Doyle & Doyle, 1990). Extraction with CTAB buffer was used in DNA extraction in plant genomes that contained many polyphenol and polysaccharide compounds (Sundari, 2018).

DNA Quality and Quantity Test and Sequencing

The quality of extracted genomic DNA was evaluated by 1% (w/v) agarose gel electrophoresis. If the obtained DNA band has fragmentation and only has a little High Molecular Weight DNA, then further DNA purification was carried out using AMPure XP Beads. DNA purity was measured using an Implen NP80 NanoPhotometer. The limit of DNA was stated as pure in molecular analysis when the ratio of A260 / A280 values ranged from 1.7-2.0 (Piskata et al., 2019).

The quantity of extracted DNA was measured using a Qubit 1.0 Fluorometer (Invitrogen). The DNA sequencing process followed the Native Barcoding Genomic DNA protocol from Oxford Nanopore Technologies (version: BE_9065_v109_revZ_14Aug 2019). The sequencing protocol was run using MinION MK1B and flowcell type R9.4.1.

Data Analysis

Data Basecalling

DNA sequencing output from MinION was in the form of Fast5, then Fast5 data from MinION sequencing was processed using Guppy Basecaller (v4.2.3+8aca2af8) for basecalling Fast5 data into Fastq data.

Quality Control (QC) and Filtering

Statistical analysis of the raw reads was performed using the NanoStat program (v1.5.1) to see the distribution of quality scores across reads. The program NanoPlot (v1.31.0) was used to plot read length x quality score. Then, reads were filtered to remove reads of poor quality, i.e., reads with quality score (Q) values below 7 and lengths less than 500 bp (Siregar et al., 2021) using the NanoFilt (v2.7.1) program (De Coster et al., 2018).

Reads Assembly

Reference assembly was performed with the Rebaler program (V0.2.0). The reference species used was *Bombax ceiba* with NCBI accession number NC_037494.1.

Assembly Polishing and Gene Annotation

Assembly contigs were polished using the medaka v1.2.1 (Oosterbroek et al., 2021) to obtain contigs with better accuracy and annotation process of cpDNA assembly through Cloud Server Chlorobox from Max Planck Institute Germany, GeSeq (Tillich et al., 2017). Assembly evaluation was performed with Quast v5.0.2 (Mikheenko et al. 2018).

Phylogenetic Tree Construction

The types of genes or markers used in constructing phylogenetic trees are *rbcL, matK,* (Hollingsworth, 2011) and a combination of both using the SnapGene program (V5.2.3). Nucleotide base sequences were analyzed using the BLASTx (Hall, 2013; Yang et al., 2014) based on Malvaceae family in NCBI database (Schoch et al., 2020). The top 50 results were then constructed into a phylogenetic tree using the MegaX program (v10.2.2). DNA sequences were aligned with ClustalW algorithm and tree constructed with Neighbor-Joining algorithm using a bootstrap of 1000 replications (Wulansari et al., 2015). The obtained phylogenetic tree was then modified using iTOL (Letunic & Bork, 2021) and Inkscape (v1.0.2).

RESULTS AND DISCUSSION

The results of the data quality control (QC), was presented in a histogram comparison between the average read lengths and read quality of the DNA sequencing process (Fig. 1). The longest reads obtained from the sequencing process of balsa (*O. pyramidale*) were 50,000 bp and the average read lengths were mostly in the range below 10,000 bp. MinION sequencing results were qualified using a standardized quality score, the phred score (Q) with the algorithm $Q = -10 \times \log10(P)$, where P was the probability of sequencing error (Delahaye & Nicolas, 2021). The highest quality of MinION sequencing reads based on the quality score was Q27, and most reads fall between Q5 and Q17 (Fig. 1). The complete Q value results of the filtering process are presented in Table 1.

The statistical data of basecalled and filtered reads reads (FastQ) were obtained for several parameters (Table 2). FastQ data was also filtered with the minimum length parameter of 500 bp and the minimum quality length of Q7. A total of 491,403,340 bp (491.403 Mb) raw reads were obtained and it decreased to 466,788,199 bp (466.788 Mb) after filtering.

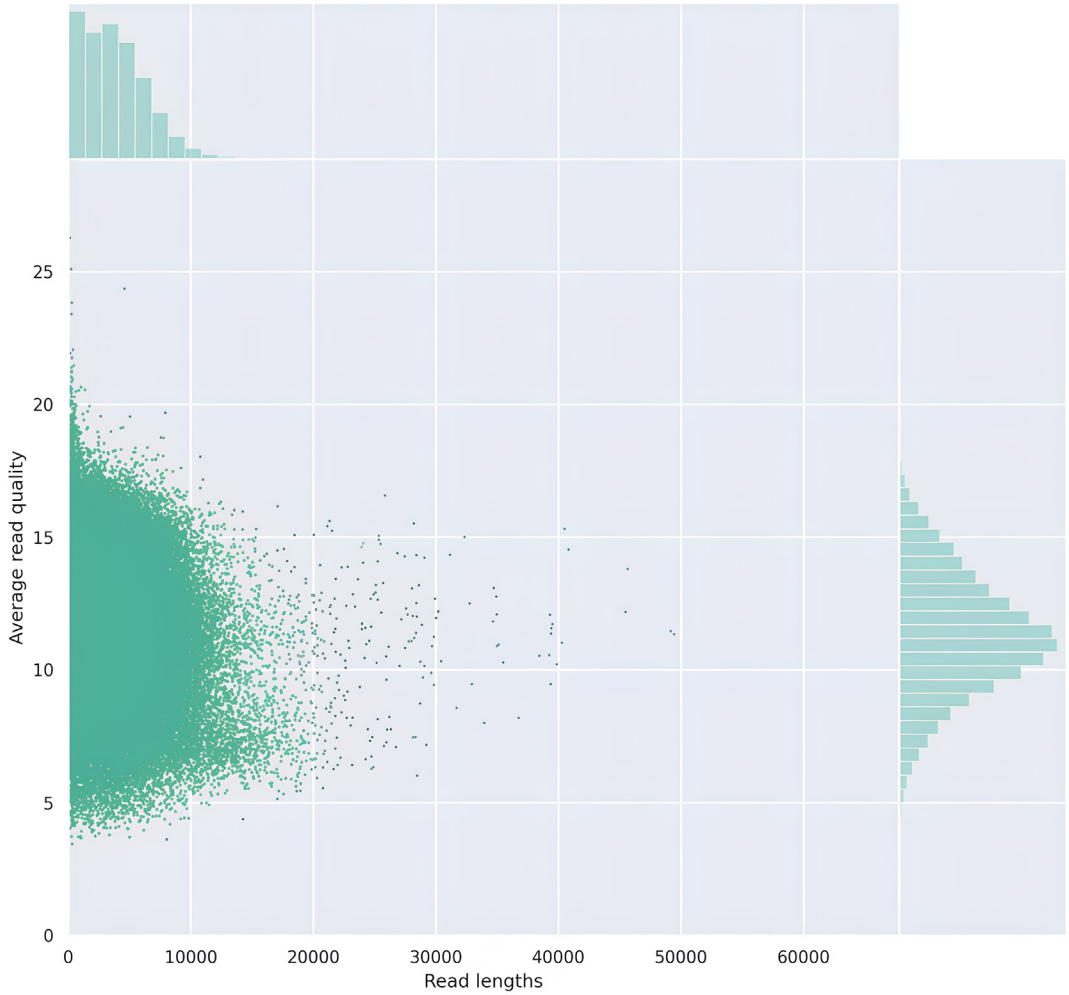


Fig. 1. Histogram of read length comparison and read quality of balsa DNA sequencing results (*O. pyramidale*)

Table 1. Phred score (Q) value of QC data after filtration

Phred score	Number of reads (bp)	Total bases (Mb)	Percentage (%)
>Q5	111,748	466.8	100.0
>Q7	111,731	466.7	100.0
>Q10	84,554	347.0	75.7
>Q12	41,384	162.4	37.0
>Q15	6,712	23.0	6.0

Table 2. Comparison of reads raw data and filtered data sequencing results of balsa (*O. pyramidale*)

No	Parameter	Raw data	Filtered data
1	Mean read length (bp)	3,904	4,177
2	Mean read quality (Q)	11.3	11.5
3	Number of reads (bp)	125,844	111,748
4	Read length N50 (bp)	5,428	5,391
5	Total bases (bp)	491,403,340	466,788,199

The analysis results in Table 2 also showed that the mean read length and mean read quality increased after filtering. This is because after filtering only reads with a minimum length of 500 bp were allowed to pass, so the average read length increased, including the read quality. The quality of reads filtered to a minimum of Q7 also caused the average read quality to increase. This was inversely proportional to the number of reads, N50 read length, and total bases with decreased results after filtering. The minimum parameter specified in the filtering caused the parameters of read count, read length N50 and total bases to decrease due to not all data passing through. Read length N50 represented the smallest contig length necessary to encompass 50% of the genome (Batista et al., 2020). These indicated that 5,391 reads were the minimum reads that could cover the total length of the genome obtained. The total number of reads after filtering was 88.8% of raw reads.

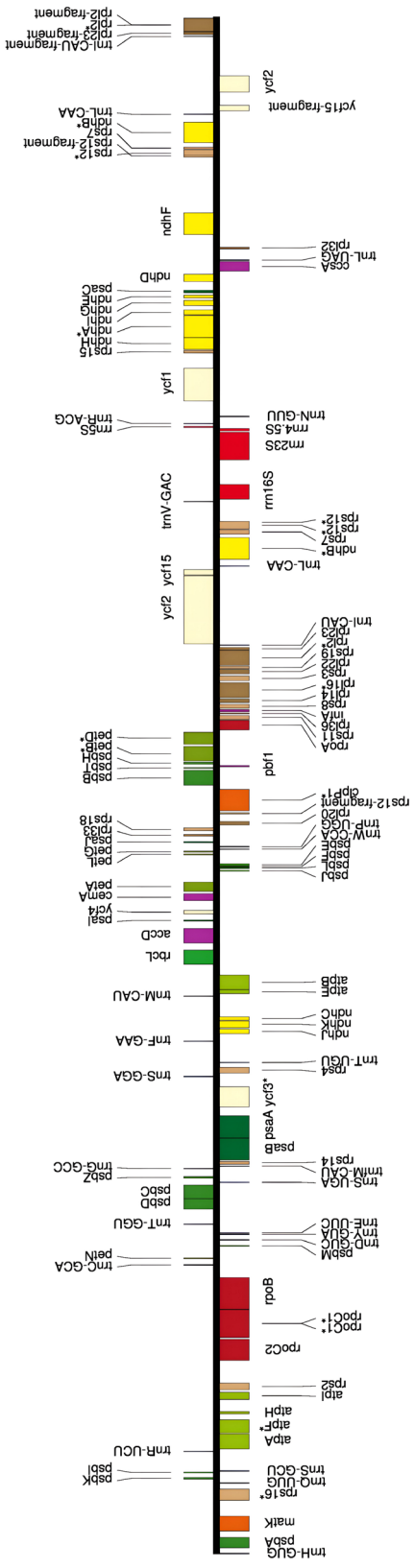
The assembly process in this study referred to reference assembly, an assembly method by using a reference sequence of closely related species that was already known for its structure so that it was used as a template for assembly. The reference used in the study was the genome of *Bombax ceiba* (NC_037494,1) which belongs to the Malvaceae family. This species was used as a reference since balsa (*O. pyramidale*) and *Bombax ceiba* still belong to the same family (Carvalho-Sobrinho et al., 2016), so no major differences were expected. *Bombax ceiba* hold significant importance within the ecosystem of tropical dry deciduous forest ecosystem, known as a multipurpose tree used as food, animal feed, fuel, and medicine (Ju et al., 2015). The complete genome chloroplast of *Bombax ceiba* has been studied (Gao et al., 2018a). In their research Gao et al. (2018b) obtained sequencing results with NGS technology on *Bombax ceiba* around 36.1 Gb raw data and 20.0 Gb after filtering, while the sequencing process in balsa with MinION Oxford Nanopore obtained around 491.403 Mb raw data and 466.788 Mb after filtering.

The next stage of data analysis was gene annotation. The annotation process was performed to obtain the genes identified in the chloroplast genome of balsa (*O. pyramidale*) (Fig. 2). Chloroplast DNA (cpDNA) was characterized by uniparental inheritance (Palmer et al., 1988). Chloroplast DNA form a circular structure with a size from 85-2,000 Kb. cpDNA played a role in controlling the production of ribosomal RNA (rRNA), transfer RNA (tRNA), and some proteins contained in the chloroplast organelle (Cummings et al., 2003).

Based on the annotation results, the size of the balsa chloroplast genome (*O. pyramidale*) was 155,430 bp. The research conducted was basically a whole genome sequencing. However, to obtain the whole genome of a plant species needed a repeated sequencing process and also determined by the sequencing depth. Sequencing depth or sequencing coverage was the coverage of the number of sequence repeats in each particular DNA segment in the genome of the sequenced organism (Tasma et al., 2016). The annotation process was performed using *Bombax ceiba* as the reference. Complete chloroplast genome research on *Bombax ceiba* has been done by Gao et al. (2018a) with the results of a chloroplast genome measuring of 158,997 bp, with 116 genes including 81 protein-encoding genes, 27 tRNA genes, and 8 RNA genes.

The result of annotation using GeSeq tools yielded 317 genes (features). These genes such as *rbcL* and *matK* could be used as DNA barcode markers. The *rbcL* gene had a length of about 1,400 bp and had a role in encoding the RubisCO protein (Nurhasanah et al., 2019). The rbcL gene sequence had a low mutation rate so it has a high degree of similarity amount species (Dobrogojski et al., 2020). This study showed that the *rbcL* gene was located at locations 59,625-61,058 with a length of about 1,434 bp. In addition to the *rbcL* gene, there were also many other types of genes obtained such as *matK*. The *matK* gene was mostly used in taxonomic and phylogenetic studies both intraspecies and interspecies in angiosperms. The *matK* gene was highly conserved gene in plants and had a function in the maturation process which involved in group to intron spicing. Its approximately 1,500 bp in sequence length and was located between the chloroplast intron and the *trnK* gene (Selvaraj et al., 2008). The *matK* gene sequence in this study was located at 2,318-3,832 bp with a length of about 1,515 bp.

Chloroplast DNA was usually used for interspecies relationship analysis, although it was unable to separate intraspecies relationships and had low genetic combination (Taberlet et al., 1991). The *rbcL* and *matK* DNA sequences from the balsa chloroplast genome were processed using BLASTX (Hall, 2013; Yang et al., 2014) on the NCBI website (Schoch et al., 2020), and the top 50 results were assembled into a phylogenetic tree to determine the relationship among balsa species. Phylogenetic trees could describe the relationship between a species and other species (Sindiya et al., 2018).



10kb

Fig. 2. cpDNA annotation results on the balsa (*O. pyramidale*) chloroplast genome

The phylogenetic tree showed that balsa (*O. pyramidale*) in a clade with other *Ochroma pyramidale* species used either *rbcL* (Fig. 3), *matK* (Fig. 4), or a combination of *rbcL* and *matK* (Fig. 5). The same clade in a phylogenetic tree describes a group of descendants of a common ancestor (Dharmayanti, 2011). The algorithm used in constructing the phylogenetic tree was the Neighbour-Joining Tree method (Wulansari et al., 2015), thus, the evolutionary process of balsa plants could not be calculated. The use of Neighbor-Joining Tree method in phylogenetic analysis only formed a phylogenetic tree based on homologous DNA sequences.

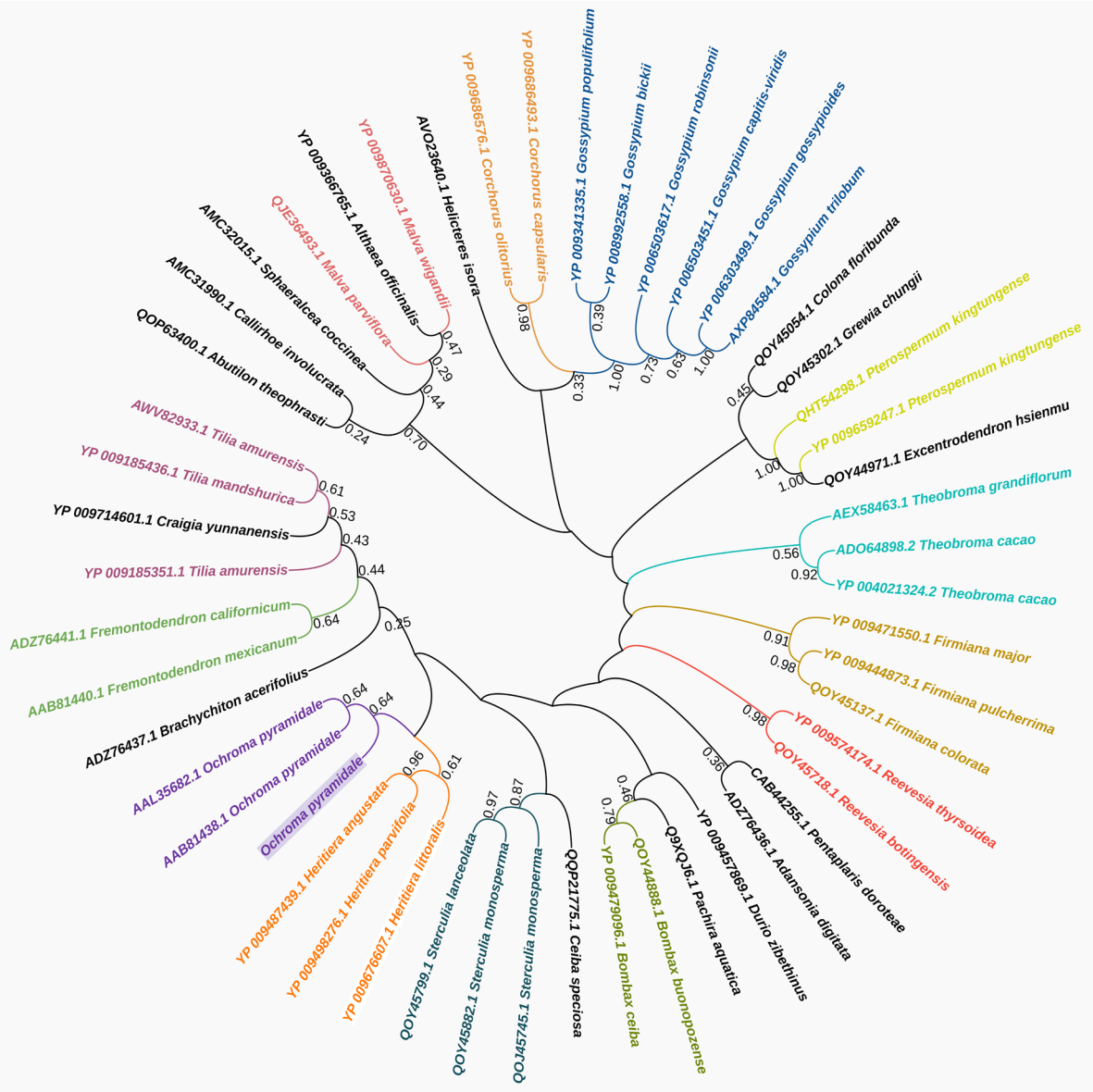


Fig. 3. Phylogenetic tree of balsa (*O. pyramidale*) with *rbcL* gene

The results of phylogenetic tree construction had a node value (bootstrap value) on each branch of the phylogenetic tree. The bootstrap value indicated the strength of the model data set (Dharmayanti, 2011). The results of the bootstrap value obtained for each gene or marker (*rbcL*, *matK*, combination of *rbcL* and *matK*) did not reach a value of 100% even though they were basically the same species. This might be caused by variations in the balsa studied with balsa data obtained from NCBI such as differences in growing places, growth phases, and types of samples used in research due to environmental factors (Sindiya et al., 2018).

Phylogenetic Tree of *rbcL* Markers

There were many types of molecular data that could be used in phylogeny studies, including the *rbcL* gene or marker. The *rbcL* gene had a role in encoding the RubisCO protein (ribulose-1,5-biphosphate carboxylase/oxygenase) which played a role in fixating and reducing CO_2 from the air into organic carbon (Tabita et al., 2008).

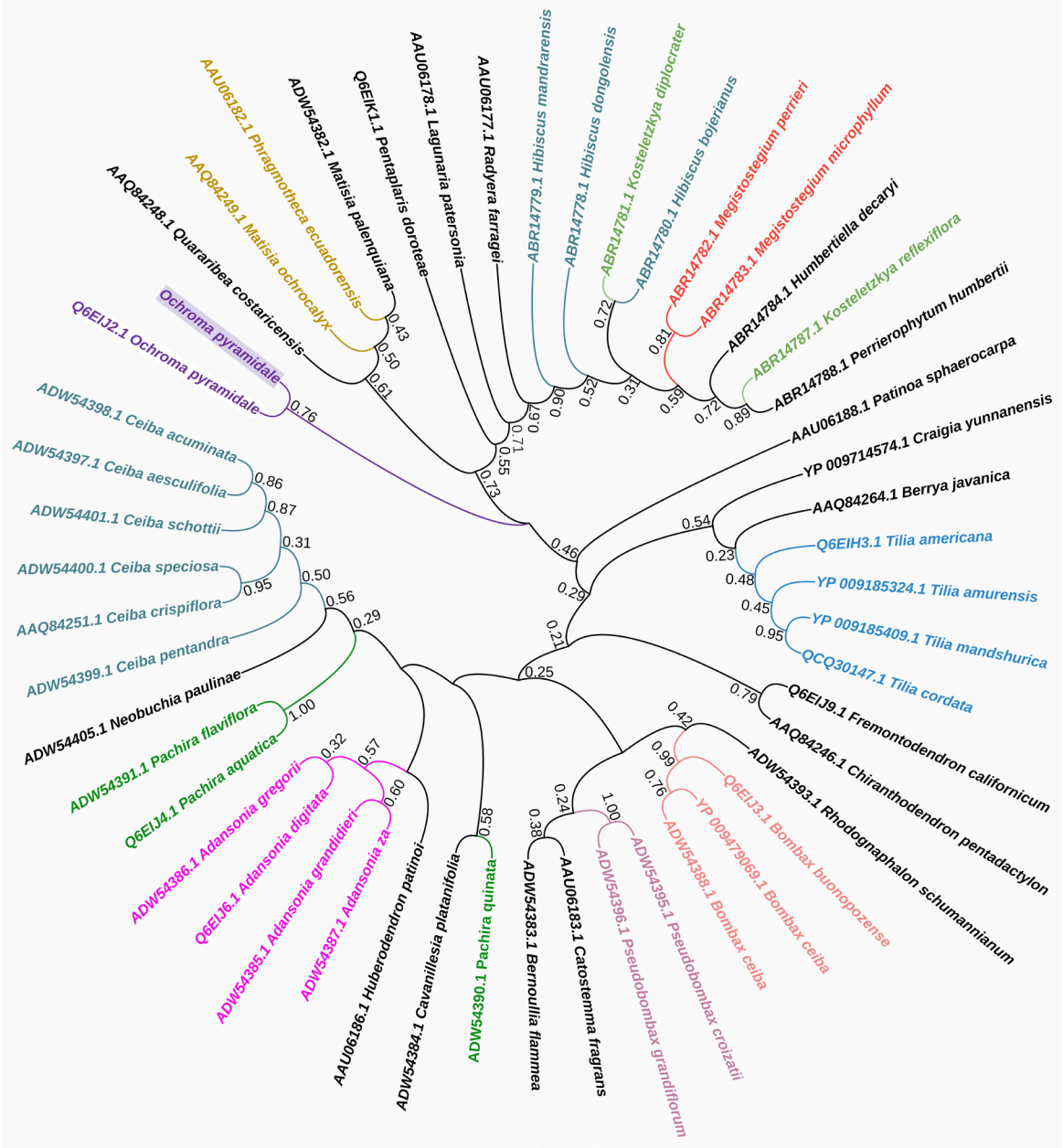


Fig. 4. Phylogenetic tree of balsa (*O. pyramidale*) with *matK* gene

In Fig. 3, the phylogenetic tree of the *rbcL* gene, the bootstrap value of balsa (*O. pyramidale*) was 0.64 or 64%. The bootstrap value was included in the weak category. Bootstrap value had three categories: high (>85%), moderate (70-85%), weak (50-69%), and very weak (<50%) (Kress et al., 2002). The low bootstrap value of the *rbcL* phylogenetic tree in this balsa had the possibility of changes in the arrangement and branching of the phylogenetic tree. The bootstrap value indicated that the use of the *rbcL* gene was not good enough to differentiate between balsa (*O. pyramidale*) species in the Malvaceae family. Kolondam et al. (2012) explained in their research that the *rbcL* gene was only able to distinguish samples up to the family level.

Phylogenetic Tree of *matK* Markers

The bootstrap value on the *matK* marker was classified as moderate (0.76 or 76%). The bootstrap value was higher for the *matK* gene in comparison to *rbcL* gene. It could be concluded that the use of *matK* gene was quite suitable for the analysis of balsa phylogeny. In general, *matK* DNA sequences have better ability to distinguish genera and species than *rbcL* (Ismail et al., 2020). The *matK* gene could evolve faster than the *rbcL* gene and consistently showed a high level of discrimination, especially in Angiosperms (Mursyidin & Makruf, 2020). The *matK* gene played a role in the expression of the K subunit maturase in the plant chloroplast genome (Lian et al., 2022). The *matK* gene had features that formed variation in DNA and amino acids. In both monocots and dicots, the 5' region of the *matK* gene exhibited greater variation than the 3' region. Therefore, *matK* gene at both the nucleic acid and amino acid sequence was often used in resolving kinship relationships in the family and even in species (İnce et al., 2005).

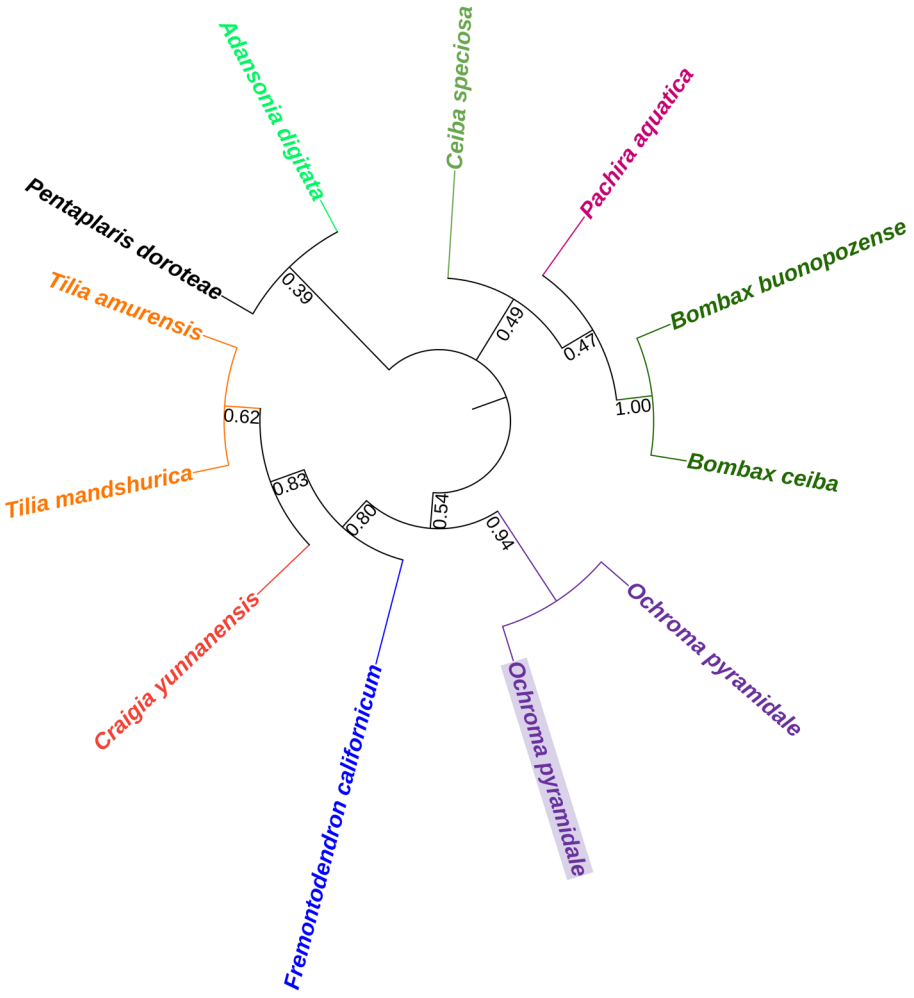


Fig. 5. Phylogenetic tree of balsa (*O. pyramidale*) using combination of *rbcL* and *matK* genes

Phylogenetic Tree with Combination of *rbcL* dan *matK* Marker

The combination of *rbcL* and *matK* genes resulted in a high bootstrap value of 0.94 or about 94%. This indicated that the combination of *rbcL* and *matK* genes was a molecular marker that could be used as a recommendation in phylogeny analysis of balsa (*O. pyramidale*). Based on the three phylogenetic trees with different types of markers, it was shown that the gene or marker combination of *rbcL* and *matK* gave the best results with a higher bootstrap value than the *rbcL* and *matK* genes. This was explained by Wattoo et al., (2016) that the combination of *rbcL* and *matK* could be used to distinguish about 90% of flowering plant species (Angiosperms). Lestari et al. (2018) also reported that using a combination of DNA molecular markers could provide a better and more stable phylogenetic tree topology. These was because the barcode or molecular marker used in plant species had a very labile nature, so the more DNA molecular markers used, the better the identification of a plant species could provide.

CONCLUSION

Long-read sequencing of balsa (*Ochroma pyramidale*) produced reads of 491.403 Mb, as well as partial whole genome sequences of 155,430 bp. This yield which was useful to analyze DNA barcode-based phylogenetic analysis from the chloroplast genome. Phylogeny studies using *rbcL*, *matK*, and a combination of both genes showed that balsa was genetically grouped in one clade with other *Ochroma pyramidale* species. In comparison with *rbcL gene,* the *matK* gene and the combination of the *rbcL* and *matK* genes were greater to analyze relationships in balsa.

ACKNOWLEDGEMENT

This research was funded by the Basic Research Scheme 2021 (No. 1/E1/KP.PTNBH/2021) awarded by the Ministry of Research and Technology/National Research and Innovation Agency of the Republic of Indonesia; the National Competitive Basic Research Scheme 2022 (No. 001/E5/PG.02.00PT/2022) awarded by Ministry of Education, Culture, Research and Technology of Republic of Indonesia; and the National Competitive Basic Research 2022-2023 No: 001/E5/PG.02.00PT/2022 and 001/E5/PG.02.00.PL/2023 awarded by Ministry of Education, Culture, Research and Technology of the Republic of Indonesia entitled “Rintisan Sekuensing 100 Genom Pohon Hutan Asli untuk Mendukung Restorasi Ekosistem”.

REFERENCES

Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biology*, *21*(1), 1–16. https://doi.org/10.1186/s13059-020-1935-5

Batista, F. M., Stapleton, T., Lowther, J. A., Fonseca, V. G., Shaw, R., Pond, C., Walker, D. I., van Aerle, R., & Martinez-Urtaza, J. (2020). Whole Genome Sequencing of Hepatitis A Virus Using a PCR-Free Single-Molecule Nanopore Sequencing Approach. *Frontiers in Microbiology*, *11*(May), 1–9. https://doi.org/10.3389/fmicb.2020.00874

Bhalerao, R., Nilsson, O., & Sandberg, G. (2003). Out of the woods: Forest biotechnology enters the genomic era. *Current Opinion in Biotechnology*, *14*(2), 206–213. https://doi.org/10.1016/S0958-1669(03)00029-6

Carvalho-Sobrinho, J. G., Alverson, W. S., Alcantara, S., Queiroz, L. P., Mota, A. C., & Baum, D. A. (2016). Revisiting the phylogeny of Bombacoideae (Malvaceae): Novel relationships, morphologically cohesive clades, and a new tribal classification based on multilocus phylogenetic analyses. *Molecular Phylogenetics and Evolution*, *101*, 56–74. https://doi.org/10.1016/j.ympev.2016.05.006

Cummings, M. P., Nugent, J. M., Olmstead, R. G., & Palmer, J. D. (2003). Phylogenetic analysis reveals five independent transfers of the chloroplast gene rbcL to the mitochondrial genome in angiosperms. *Current Genetics*, *43*(2), 131–138. https://doi.org/10.1007/s00294-003-0378-3

De Coster, W., D’Hert, S., Schultz, D. T., Cruts, M., & Van Broeckhoven, C. (2018). NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics*, *34*(15), 2666–2669. https://doi.org/10.1093/bioinformatics/bty149

Delahaye, C., & Nicolas, J. (2021) Sequencing DNA with nanopores: Troubles and biases. *PLoS ONE 16*(10): e0257521. https://doi.org/10.1371/journal.pone.0257521

Dharmayanti, N. (2011). Filogenetika Molekuler : Metode Taksonomi Organisme Berdasarkan Sejarah

Evolusi. *WARTAZOA*, *21*(1), 1–10. https://doi.org/10.2307/2799276

Dobrogojski, J., Adamiec, M., & Luciński, R. (2020). The chloroplast genome: a review. *Acta Physiologiae Plantarum*, *42*(6), 1–13. https://doi.org/10.1007/s11738-020-03089-x

Doyle, J.J., & Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. *Focus*, *12*, 13–15. https://cir.nii.ac.jp/crid/1573950400018579968

Dumschott, K., Schmidt, M. H. W., Chawla, H. S., Snowdon, R., & Usadel, B. (2020). Oxford Nanopore sequencing: new opportunities for plant genomics? *Journal of Experimental Botany*, *71*(18), 5313–5322. https://doi.org/10.1093/jxb/eraa263

Gao, Y., Wang, H., Liu, C., Chu, H., Yan, Y., & Tang, L. (2018a). Complete chloroplast genome sequence of the red silk cotton tree (Bombax ceiba). *Mitochondrial DNA Part B: Resources*, *3*(1), 315–316. https://doi.org/10.1080/23802359.2017.1422399

Gao, Y., Wang, H., Liu, C., Chu, H., Dai, D., Song, S., Yu, L., Han, L., Fu, Y., Tian, B., & Tang, L. (2018b). De novo genome assembly of the red silk cotton tree (Bombax ceiba). *GigaScience*, *7*(5), 1–7. https://doi.org/10.1093/gigascience/giy051

Hall, B. G. (2013). Building phylogenetic trees from molecular data with MEGA. *Molecular Biology and Evolution*, *30*(5), 1229–1235. https://doi.org/10.1093/molbev/mst012

Hartati, N., Sudarmonowati, E., Fatriasari, W., Hermiati, E., Dwianto, W., Kaida, R., Baba, K., Hayashi, T., Sri Hartati, N., Sudarmonowati, E., Fatriasari, W., Hermiati, E., Dwianto, W., Kaida, R., Baba, ichi, & Hayashi, T. (2010). Wood Characteristic of Superior Sengon Collection and Prospect of Wood Properties Improvement through Genetic Engineering. *Wood Research Journal*, *1*(2), 103–107. http://www.ejournalmapeki.org/index.php/wrj/article/view/173

Hollingsworth, P. M. (2011). Refining the DNA barcode for land plants. *Proceedings of the National Academy of Sciences (PNAS)*, *108*(49), 19451-19452. https://doi.org/10.1073/pnas.1116812108

Howcroft, N. (2002). *The Balsa Manual : Techiniques For Establishiment and The Management of Balsa (Ochroma pyamidale) plantation in Papua New Guinea*. Keravat(PNG): International Tropical Timbers Organisation. https://www.itto.int/files/itto_project_db_input/2015/Technical/pd7-99%20rev2(F)%20e_The%20Balsa%20Manual_e.pdf

Ide, Y. (2021). Genetics and improvement of forest trees. *Forests*, *12*(2), 1–3. https://doi.org/10.3390/f12020182

İnce, A. G., Karaca, M., Onus, A. N., & Bilgen, M. (2005). Chloroplast *matK* gene phylogeny of some important species of plants. *Akdeniz ÜNiversitesi Ziraat FakülLtesi Dergisi*, *18*(2), 157-162. https://www.researchgate.net/publication/228865648_Chloroplast_matK_Gene_Phylogeny_of_Some_Important_Species_of_Plants

Ismail, M., Ahmad, A., Nadeem, M., Javed, M. A., Khan, S. H., Khawaish, I., Sthanadar, A. A., Qari, S. H., Alghanem, S. M., Khan, K. A., Khan, M. F., & Qamer, S. (2020). Development of DNA barcodes for selected *Acacia* species by using *rbcL* and *matK* DNA markers. *Saudi Journal of Biological Sciences*, *27*(12), 3735–3742. https://doi.org/10.1016/j.sjbs.2020.08.020

Istiqomah, F. N., Budi, S. W., & Wulandari, A. S. (2017). Peran fungi mikoriza arbuskula (FMA) dan asam humat terhadap pertumbuhan balsa (*Ochroma bicolor* Rowlee.) pada tanah terkontaminasi timbal (Pb). *Jurnal Pengelolaan Sumberdaya Alam Dan Lingkungan (Journal of Natural Resources and Environmental Management)*, *7*(1), 72–78. https://doi.org/10.29244/jpsl.7.1.72-78

Kress, W.J., Prince, L. M., & Williams, K. J. (2002). The phylogeny and a new classification of the gingers (Zingiberaceae): Evidence from molecular data. *American Journal of Botany*, *89*(10), 1682–1696. https://doi.org/10.3732/ajb.89.10.1682

Ju, M.M., Ma, H.C., Xin, P.Y., Zhou, Z.L., & Tian, B. (2015). Development and Characterization of EST-SSR Markers in Bombax ceiba (Malvaceae). *Applications in Plant Sciences*, *3*(4), 1500001. https://doi.org/10.3732/apps.1500001

Kolondam, B. J., Lengkong, E., J. Polii, M., Pinaria, A., & Runtunuwu, S. (2012). Barcode DNA berdasarkan gen *rbcL* dan *matK* anggrek payus limondok (*Phaius tancarvilleae*) (DNA barcode of payus limondok orchid (*Phaius tancarvilleae*) based on the *rbcL* and *matK* genes). *Jurnal Bios Logos*, *4*(2). https://doi.org/10.35799/jbl.2.2.2012.1041

Kotlarewski, N. J., Belleville, B., Gusamo, B. K., & Ozarska, B. (2016). Mechanical properties of Papua New Guinea balsa wood. *European Journal of Wood and Wood Products*, *74*(1), 83–89. https://doi.org/10.1007/s00107-015-0983-0

Lee, Y. G., Choi, S. C., Kang, Y., Kim, K. M., Kang, C. S., & Kim, C. (2019). Constructing a reference genome in a single lab: The possibility to use

oxford nanopore technology. *Plants*, *8*(8), 1–13. https://doi.org/10.3390/plants8080270

Lestari, D. A., Azrianingsih, R., & Hendrian, H. (2018). Filogenetik Jenis-jenis Annonaceae dari Jawa Timur Koleksi Kebun Raya Purwodadi Berdasarkan Coding dan Non-coding sekuen DNA. *Journal of Tropical Biodiversity and Biotechnology*, *3*(1), 1. https://doi.org/10.22146/jtbb.28308

Letunic, I., & Bork, P. (2021). Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, *49*(W1), W293–W296. https://doi.org/10.1093/nar/gkab301

Lian, C., Yang, H., Lan, J., Zhang, X., Zhang, F., Yang, J., & Chen, S. (2022). Comparative analysis of chloroplast genomes reveals phylogenetic relationships and intraspecific variation in the medicinal plant Isodon rubescens. *PLoS ONE*, *17*(4 April), 1–18. https://doi.org/10.1371/journal.pone.0266546

Listyanto, T., Poedyastanto, E.P.F., Abqoriah, S.M., & Lukmandaru, G. (2021). Specific gravity, extractive content, and natural durability of balsa (*Ochroma pyramidale)* wood at 3 and 4 years old. *IOP Conf. Series: Earth and Environmental Science*. *891*, 012013. https://doi.org/10.1088/1755-1315/891/1/012013

Lu, H., Giordano, F., & Ning, Z. (2016). Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics and Bioinformatics*, *14*(5), 265–279. https://doi.org/10.1016/j.gpb.2016.05.004

Mikheenko A, Prjibelski A, Saveliev V, Antipov D, & Gurevich A. (2018). Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics*, 34(13), i142 - i150. https://doi.org/10.1093/bioinformatics/bty266

Mikheyev, A. S., & Tin, M. M. Y. (2014). A first look at the Oxford Nanopore MinION sequencer. *Molecular Ecology Resources*, *14*(6), 1097–1102. https://doi.org/10.1111/1755-0998.12324

Mursyidin, D.H., & Makruf, M.I. (2020). Keanekaragaman dan kekerabatan genetik artocarpus berdasarkan penanda dna kloroplas *matK* & *rbcL*: kajian in silico. *Floribunda*, *6*(5). https://doi.org/10.32556/floribunda.v6i5.2020.322

Mursyidin, D. H., Ahyar, G. M. Z., Saputra, A. W., & Hidayat, A. (2021). Genetic diversity and relationships of *Phalaenopsis* based on the *rbcL* and *trnL-F* markers: In silico approach. *Biosaintifika: Journal of Biology & Biology Education*, *13*(2), 212–221. https://doi.org/10.15294/biosaintifika.v13i2.29904

Nurhasanah, Sundari, & Papuangan, N. (2019). Amplification and analysis of *rbcl* gene (ribulose-1,5-bisphosphate carboxylase) of clove in Ternate Island. *IOP Conference Series: Earth and Environmental Science*, *276*(1). https://doi.org/10.1088/1755-1315/276/1/012061

Nuryamah. (2017). *Knowledge management* (S. Noviani (ed.); Maret-April, Issue April). PUPR. https://docplayer.info/93250471-Knowledge-management-penerapan-teknologi-konstruksi.html

Oosterbroek, S., Doorenspleet, K., Nijland, R., & Jansen, L. (2021). Decona: From demultiplexing to consensus for Nanopore amplicon data. *ARPHA Conference Abstracts*, *4*, 10–11. https://doi.org/10.3897/aca.4.e65029

Palmer, J. D., Jansen, R. K., Michaels, H. J., Chase, M. W., James, R., Palmer, D., & Manhart, J. R. (1988). Chloroplast DNA variation and plant phylogeny. *Annals of the Missouri Botanical Garden*, *75*(4), 1180–1206. https://doi.org/10.2307/2399279

Pertiwi, Y.A.B, Ishiguri, F., Aiso, H., Oshima, J., & Yokota, S. (2017). Wood properties of 7-year-old balsa (*Ochroma pyramidale*) planted in East Java. *International Wood Products Journal*, *8*(4),1–6. https://doi.org/10.1080/20426445.2017.1394560

Piskata, Z., Servusova, E., Babak, V., Nesvadbova, M., & Borilova, G. (2019). The quality of DNA isolated from processed food and feed via different extraction procedures. *Molecules*, *24*(6), 1–10. https://doi.org/10.3390/molecules24061188

Rachmat, H.H., Subiakto, A., & Susilowati, A. (2019). Genetic resources of fast-growing tree for rehabilitating upland area of deteriorated Saguling catchment, West Java, Indonesia. *Biodiversitas*, *20*(2), 442–447. https://doi.org/10.13057/biodiv/d200220

Roslim, D. I., Oktavia, S., & Herman. (2015). Analisis sebagian sekuen DNA dari gen MEISA1 pada ubi kayu (*Manihot esculenta* Crantz.) genotipe menggalo dan roti. *Jurnal Dinamika Pertanian*, *30*(2), 109–116. https://journal.uir.ac.id/index.php/dinamikapertanian/article/view/803

Salindeho, R. A., Dwiyanti, F. G., Pratama, R. Matra, D .D., Majiidu, M., Fatlan, K. F., & Siregar, I. Z. (2023). Understanding *Diospyros rumphii* Bakh from North Sulawesi through long-read

sequences analysis using MinION Oxford Nanopore Technologies. *IOP Conf. Series: Earth and Environmental Science*, *1188* (012035), 1–9. https://doi.org/10.1088/1755-1315/1188/1/012035

Schoch, C. L., Ciufo, S., Domrachev, M., Hotton, C. L., Kannan, S., Khovanskaya, R., Leipe, D., McVeigh, R., O'Neill, K., Robbertse, B., Sharma, S., Soussov, V., Sullivan, J. P., Sun, L., Turner, S., & Karsch-Mizrachi, I. (2020). NCBI Taxonomy: A comprehensive update on curation, resources and tools. *Database*, *2020*(2), 1–21. https://doi.org/10.1093/database/baaa062

Selvaraj, D., Sarma, R. K., & Sathishkumar, R. (2008). Phylogenetic analysis of chloroplast *matK* gene from Zingiberaceae for plant DNA barcoding. *Bioinformation*, *3*(1), 24–27. https://doi.org/10.6026/97320630003024

Sindiya, V., Mukarramah, L., Rohimah, S., Al GhifariPerwitasari, D., & Su'udi, M. (2018). Studi in silico potensi dna barcode pada anggrek langka *Paphiopedilum*. *BIOSFER : Jurnal Biologi Dan Pendidikan Biologi*, *3*(1), 20–26. https://doi.org/10.23969/biosfer.v3i1.1250

Siregar, I. Z., Dwiyanti, F. G., Pratama, R., Matra, D. D., & Majiidu, M. (2021). Generating long-read sequences using Oxford Nanopore Technology from *Diospyros celebica* genomic DNA. *BMC Research Notes*, *14*(75), 1–4. https://doi.org/10.1186/s13104-021-05484-0

Sundari. (2018). Teknik isolasi DNA genom tanaman cengkeh dengan menggunakan modifikasi bufer CTAB (DNA isolation technique of clove plant genomes using CTAB buffer modification). *Jurnal Biologi Edukasi Edisi*, *21*, 21–26. https://jurnal.usk.ac.id/JBE/article/download/13928/10504

Taberlet, P., Gielly, L., Pautou, G., Bouvet, J., Biologie, L. De, & Fourier, U. J. (1991). Plant universal primer. *Plant Molecular Biology*, *17*(ii), 1105–1109. https://doi.org/10.1007/bf00037152

Tabita, F. R., Hanson, T. E., Satagopan, S., Witte, B. H., & Kreel, N. E. (2008). Phylogenetic and evolutionary relationships of RubisCO and the RubisCO-like proteins and the functional lessons provided by diverse molecular forms. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1504), 2629–2640. https://doi.org/10.1098/rstb.2008.0023

Tasma, I. M., Satyawan, D., & Rijzaani, H. (2016). Pembentukan pustaka genom, resekuensing, dan identifikasi SNP berdasarkan sekuen genom total genotipe kedelai Indonesia. *Jurnal AgroBiogen*, *11*(1), 7. https://doi.org/10.21082/jbio.v11n1.2015.p7-16

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), W6–W11. https://doi.org/10.1093/nar/gkx391

Vallée, G. C., Muñoz, D. S., & Sankoff, D. (2016). Economic importance, taxonomic representation and scientific priority as drivers of genome sequencing projects. *BMC Genomics*, *17*(782), 125–186. https://doi.org/10.1186/s12864-016-3100-9

Wahyuni, D., Dwiyanti, F. G., Pratama, R., Majiidu, M., Rachmat, H. H., & Siregar, I. Z. (2021). Chloroplast Genome draft of *Dryobalanops aromatica* generated using Oxford Nanopore Technology and its Potential application for phylogenetic study. *Forests*, *12*(1515), 1–14. https://doi.org/10.3390/ f12111515

Wattoo, J. I., Saleem, M. Z., Shahzad, M. S., Arif, A., Hameed, A. & Saleem, M.A. (2016). DNA Barcoding: amplification and sequence analysis of *rbcL* and *matK* genome regions in three divergent plant species. *Advancements in Life Science*, *4*(1), 3–7. https://doi.org/10.1186/s12864-016-3100-9

Wijoyo, S. S., Santosa, A., & P, C. J. (2018). Perancangan furnitur dengan material kayu balsa. *Jurnal Intra*, *6*(2), 105–115. https://publication.petra.ac.id/index.php/desain-interior/article/viewFile/7166/6501

Wulansari, N., Nurilmala, M., & Nurjanah, N. (2015). Detection tuna and processed products based protein and DNA barcoding. *Jurnal Pengolahan Hasil Perikanan Indonesia*, *18*(2), 119–127. https://doi.org/10.17844/jphpi.2015.18.2.119

Yang, Y., Jiang, X. T., & Zhang, T. (2014). Evaluation of a hybrid approach using UBLAST and BLASTX for metagenomic sequences annotation of specific functional genes. *PLoS ONE*, *9*(10). https://doi.org/10.1371/journal.pone.0110947

Zanzibar, M. (2017). The type of dormancy and pre treatment for breaking dormancy of balsa (*Ochroma bicolor* ROWLEE) seed. *Jurnal Perbenihan Tanaman Hutan*, *5*(1), 51–60. https://doi.org/10.20886/bptpth.2017.5.1.51-60