

RESEARCH NOTE

Open Access



Chloroplast genome draft assembly of *Falcataria moluccana* using hybrid sequencing technology

Vilda Puji Dini Anita¹, Deden Derajat Matra² and Ulfah Juniarti Siregar^{3*}

Abstract

Objectives *Falcataria moluccana*, known locally as Sengon, is a fast-growing legume tree that is commonly planted in community forests of Java Island, Indonesia. However, the plantations face attacks of Boktor stem borer (*Xylocopa festiva*) and gall-rust disease (*Uromycladium falcatariae*) as major threats to its productivity. To control those pest and disease, it is necessary to grow resistant sengon clones, which are developed through tree improvement program, of which needs genetic and genomic information. This dataset was created to construct draft of sengon chloroplast genome and to study the evolution of sengon based on *matK* and *rbcl* barcode genes.

Data description Genomic DNA was extracted from leaf samples of one individual healthy tree in a private plantation. The DNA was sequenced using Illumina Novaseq 6000 (Novogen AIT, Singapore) for short-reads data, and MinION of Nanopore following manufacture's protocols SQK-LSK110 for long-reads data. The 66,3 Gb short-reads and 12 Gb long-reads data were hybrid assembled and used to construct a 128.867 bp of *F. moluccana* chloroplast genome with a quadripartite structure, containing a pair of inverted repeats, a large single-copy and a small single-copy region. Phylogenetic tree constructed using *matK* and *rbcl* showed monophyletic origin of *F. moluccana* and other legume trees.

Keywords Draft Chloroplast Genome, *Falcataria moluccana*, Long-reads, Short-reads

Objective

Falcataria moluccana, locally known as Sengon, is main timber commodity in Indonesia, of which total production in 2019 reached 5.468.716,76 m³ [1], increased by 1.817.237,27 m³ from 2018 total production [2]. However, *F. moluccana* plantations have obstacles, especially from Boktor stem borer (*Xylocopa festiva*) and gall-rust (*Uromycladium falcatariae*) disease. These specific pest and disease also attack other tree species from Fabaceae family, such as from genus *Acacia* and *Archidendron*, with exception that in *F. moluccana* has caused more severe losses [3]. Since effective control methods are not available, it is necessary to develop resistant *F. moluccana* from these pest and disease.

*Correspondence:

Ulfah Juniarti Siregar
ulfahjs@apps.ipb.ac.id

¹Tropical Silviculture Program, Department of Silviculture, Faculty of Forestry and Environment, IPB University (Bogor Agricultural University), Bogor, Indonesia

²Department of Agronomy and Horticulture, Faculty of Agriculture, IPB University (Bogor Agricultural University), Bogor, Indonesia

³Department of Silviculture, Faculty of Forestry and Environment, IPB University (Bogor Agricultural University), Bogor, Indonesia



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Table 1 Overview of data files/data sets

| Label | Name of data file/data set | File types (file extension) | Data repository and identifier (DOI or accession number) |
|-------------|---|------------------------------------|---|
| Data file 1 | Statistic of Short-read and Long-read Data of <i>Sengon</i> (<i>Falcataria moluccana</i>) | Document file (.docx) | Figshare https://doi.org/10.6084/m9.figshare.21626951.v1 [21] |
| Data file 2 | Circular map of <i>F. moluccana</i> chloroplast genome | Picture file (.PNG) | Figshare https://doi.org/10.6084/m9.figshare.21627005.v1 [22] |
| Data file 3 | List gene on <i>sengon</i> chloroplast genome | Document file (.docx) | Figshare https://doi.org/10.6084/m9.figshare.21626993.v1 [23] |
| Data file 4 | Phylogenetic tree of <i>matK</i> and <i>rbcL</i> | PNG file in compressed file (.rar) | Figshare https://doi.org/10.6084/m9.figshare.21627014.v1 [24] |
| Data set 1 | Raw Short-reads data | Fastq files (.fastq) | DNA Data Bank of Japan (DDBJ) accession number DRA012508 (https://trace.ddbj.nig.ac.jp/DRASearch/submission?acc=DRA012508) [25] |
| Data set 2 | Raw Long-reads data | Fastq files (.fastq) | DNA Data Bank of Japan (DDBJ) accession number DRA015209 (https://trace.ddbj.nig.ac.jp/DRASearch/submission?acc=DRA015209) [26] |

F. moluccana improvement program has been conducted; however, progress is slow considering the complexity of the resistant traits. In such case genomic approach could assist the selection program by providing information on important genes related to resistance to pests and diseases. Some genes related to resistance to biotic and abiotic stress, as well as adaptation could be located in the cytoplasm, such as in the chloroplast genome [4]. The host range of Boktor stem borer pest and gall-rust disease among trees from Fabaceae family posed an interesting evolutionary relationship among those tree species in the Fabaceae family. Chloroplast genome is relatively small in size and very conservative that it becomes popular subject for studying genetic and evolutionary relationship among plant species [5]. This study aimed at constructing a complete and high quality of *F. moluccana* chloroplast draft genome utilizing the advance of sequencing technology such as Next-generation Sequencing (e.g. Illumina) and Third-generation Sequencing (e.g. Oxford Nanopore) with bioinformatics approach [6], also to find out the evolutionary relationship of *F. moluccana* with several other tree species from Fabaceae family using *matK* and *rbcL* genes, which are commonly used in DNA barcoding.

Data description

Genomic DNA was extracted from 400 mg fresh leaf samples using CTAB method from [7] with modifications. The leaves were collected from one 7 years-old individual healthy tree, grown at a private plantation in Cikarawang Village, Bogor, West Java. The quality of extracted genomic DNA was evaluated using agarose gel electrophoresis. The purity of the genomic DNA was assessed using NanoPhotometer NP80 Implen and the quantity was measured using Qubit 1.0 Fluorometer with Qubit dsDNA BR (Broad-Range) Assay Kit. Short-reads sequencing was done using Illumina Nova-seq 6000 (Novogen AIT, Singapore), while long-reads

sequences were obtained using MinION from Nanopore, following manufacture's protocols SQK-LSK110. Data can be accessed from DNA Data Bank of Japan (DDBJ) with accession number DRA012508 for short-reads data (Dataset 1) [25] and DRA015209 for long-reads data (Dataset 2) [26].

Hybrid chloroplast genome assembly was performed using the pipeline from <http://github.com/asdcid/Chloroplast-genome-assembly> [8]. The pre-assembly was performed by quality check, following the script from http://github.com/asdcid/Chloroplast-genome-assembly/tree/master/1_pre_assembly. Short-reads data was quality checked using FASTQC [9] and trimmed using BBDukv37.31 [10]. Quality check for long-reads data was also done using FASTQC program. Adapter trimming was performed using Porechop v0.2.1 [11] while quality trimming was done using NanoFilt v1.2.0 [12]. The trimming result were double checked using FASTQC. From this pre-assembly step, the total bases of long-reads data were reduced from 12Gb to 11Gb, while for short-reads data was reduced from 66,3 Gb to 63,4 Gb (Data file 1). These clean-reads were aligned to the reference NC_047364.1 (*F. moluccana*) using Bowtie v2.2.6 [13] for short-reads and Blasrv5.1 for long-reads [14].

Chloroplast-mapped reads were assembled using Unicycler v0.3.1 [15] and corrected using SPAdes in Unicycler with default settings from http://github.com/asdcid/Chloroplast-genome-assembly/tree/master/2_assembly. Afterwards, script from http://github.com/asdcid/Chloroplast-genomeassembly/tree/master/3_post_assembly was performed for post-assembly step. All contigs are combined into a single contigs with the same structure against used reference using Mummer v2.23 [16] and Pilon v1.20.1 to polish the data [17]. Draft of chloroplast contig were annotated using GeSeq [18] towards all Fabaceae reference in NCBI RefSeq and visualized using OGDraw in MPI-MP Chlorobox [19] (Data file 2). The chloroplast genome encoded 95 genes, composed

of 27 tRNA genes, 1 rRNA gene, and 67 protein coding genes (Data file 3). Phylogenetic analysis reconstruction was performed using MEGAX (Molecular Evolutionary Genetic Analysis) v10.2.2 [20] with Maximum Likelihood method, Tamura-3 model and bootstrap value of 10.000 replication. For the phylogenetic analysis *Intsia bijuga* (NC_047336.1) was used as an outgroup. Based on phylogenetic analysis using *matK* and *rbcL* gene markers, the constructed phylogenetics trees indicated a monophyletic topology. The phylogenetic tree using *matK* showed 3 groups (Data files 4, Fig. 2A), of which *F. moluccana* in this study are in the same clade with *Archidendropsis granulosa* in the second group and separated from other *F. moluccana* accessions. In the case of *rbcL* marker, the phylogenetic tree formed 9 groups (Data files 4, Fig. 2B), of which the *F. moluccana* studied are placed in the same group no. 9 with other *F. moluccana* accessions.

Limitations

This study used leaves samples from one individual tree accession in a private plantation, with unknown origin. The tree selected shows resistance to pest and disease attacks.

Acknowledgements

The authors thank to Laboratory of Forest Genetics and Molecular Forestry, Department of Silviculture, Faculty of Forestry and Environment, IPB University and Laboratory Science Molecular in the Advanced Research Laboratory (ARLab), IPB University for facilitating this study.

Author contributions

U.J.S designed the experiment and overall study. V.P.D.A conducted the experiments. D.D.M and V.P.D.A performed the chloroplast genome assembly, analysis, and interpretation. All authors prepared the manuscript.

Funding

This study was supported by Ministry of Education, Culture, Research, and Technology of Indonesia for post graduate research scheme (Skema Penelitian Pasca Sarjana/PTM) entitled "Analisis Genomik Dengan Teknologi Sekuensing Secara Hybrid (Long-Read Dan Short-Read) Pada Sengon (*Falcataria Moluccana*)", with contract No: 082/E5/PG.02.00.PT/2022 between Mendikbudristek and IPB University and contract No: 3868/IT3.L1/PT.01.03/P/B/2022 between LPPM IPB University and Principal Investigator (Ulfah Juniarti Siregar).

Data availability

The data described in this Data note can be freely and openly accessed on DNA Data Bank of Japan (DDBJ) with accession number DRA012508, DRA015209, and figshare. Please see Table 1 and references list [21–26] for details and links to the data.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Published online: 09 March 2023

References

1. BPS-Statistics Indonesia. Statistics of Forestry Production 2019 (Indonesian). Jakarta: BPS-Statistics Indonesia; 2020.
2. BPS-Statistics Indonesia. Statistics of Forestry Production 2018 (Indonesian). Jakarta: BPS-Statistics Indonesia; 2019.
3. Darwati W, Anggraeni I. The boktor and tumor attack at sengon in the plantation of tea ciater (Indonesian). Jurnal Sains Natural Universitas Nusa Bangsa. 2018;8:59–69. <https://doi.org/10.31938/jsn.v8i2.119>
4. Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genom Biol. 2016;17:134.
5. Kim KJ, Lee HL. Wide spreads occurrence of small inversions in the chloroplast genomes of land plants. Mol Cells. 2005;9(1):104–13.
6. Paajanen, P., Kettleborough, G., Lopez-Girona, E., Giolai, M., Heavens, D. & Baker, D. et al. A critical comparison of technologies for a plant genome sequencing project, <https://doi.org/10.1093/gigascience/gjy163> (2019).
7. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bull. 1987;19:11–5.
8. Wang W, Schalamun M, Morales-Suarez A, Kainer D, Schwessinger B, Lanfear R. Assembly of chloroplast genomes with long- and short-read data: a comparison of approaches using *Eucalyptus pauciflora* as a test case. BMC Genomics. 2018;19:977. <https://doi.org/10.1186/s12864-018-5348-8>
9. Andrews S. 2022. FastQC: a quality control tool for high throughput sequencing data (2010). <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. Accessed 12 August 2022.
10. BBTools. 2022. BMAP – Bushnell B. sourceforge.net/projects/bbmap/. Accessed 20 August 2022
11. Wick RR, Judd LM, Gorrie CL, Holt KE. Completing bacterial genome assemblies with multiplex MinION sequencing. Microb Genomics. 2017;3:1–7. <https://doi.org/10.1099/mgen.0.000132>
12. De Coster W, D'Hert S, Schultz DT, Cruets M, Broeckhoven CV. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics. 2018;34:1666–2669. <https://doi.org/10.1093/bioinformatics/bty149>
13. Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9. <https://doi.org/10.1038/nmeth.1923>
14. Chaisson MJ, Tesler G. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. BMC Bioinformatics. 2012;13:1–17. <https://doi.org/10.1186/1471-2105-13-238>
15. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler. Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2016;13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>
16. Marçais G, Delcher AL, Phylippou AM, Coston R, Salzberg SL, Zimin A. MUMmer4: a fast and versatile genome alignment system. PLoS Comput Biol. 2018;14:e1005944. <https://doi.org/10.1371/journal.pcbi.1005944>
17. Walker BJ, Abeel T, Shea T, Priest M, Abouellie A, Sakthikumar S, et al. Pilon: an Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLoS ONE. 2014;9:e112963. <https://doi.org/10.1371/journal.pone.0112963>
18. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq – versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45:W6–W11. <https://doi.org/10.1093/nar/gkx391>
19. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organelle genomes. Nucleic Acids Res. 2019;47:W59–W64.
20. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol. 2018;35:1547–9. <https://doi.org/10.1093/molbev/msy096>
21. Anita VPD, Siregar UJ, Matra DD. 2022. Statistic of Short-read and Long-read Data of Sengon (*Falcataria moluccana*). <https://doi.org/10.6084/m9.figshare.21626951.v1>
22. Anita VPD, Siregar UJ, Matra DD. 2022. Circular map of *F. moluccana* chloroplast genome. <https://doi.org/10.6084/m9.figshare.21627005.v1>
23. Anita VPD, Siregar UJ, Matra DD. 2022. List gene on sengon chloroplast genome. <https://doi.org/10.6084/m9.figshare.21626993.v1>
24. Anita VPD, Siregar UJ, Matra DD. 2022. Phylogenetic tree of *matK* and *rbcL*. <https://doi.org/10.6084/m9.figshare.21627014.v1>

25. DNA Data Bank of Japan [https://. trace.ddbj.nig.ac.jp/DRAsearch/submission?acc=DRA012508](https://trace.ddbj.nig.ac.jp/DRAsearch/submission?acc=DRA012508) (2020). Accessed 12 Des 2022
26. DNA Data Bank of Japan <https://. trace.ddbj.nig.ac.jp/DRAsearch/submission?acc=DRA015209> (2022). Accessed 12 Des 2022

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.