# DATA NOTE

# **Open Access**

# Genome-wide DNA polymorphisms in two peatland adapted *Coffea liberica* varieties



Tisha Melia<sup>1\*</sup>, Fatayat<sup>1</sup>, Ninik Nihayatul Wahibah<sup>2</sup>, Siti Fatonah<sup>2</sup>, Dewi Indriyani Roslim<sup>2</sup> and Arisman Adnan<sup>3</sup>

# Abstract

**Objectives** *Coffea liberica* is one of the species within the *Coffea* genus known for its distinctive flavor and resistance to leaf rust disease. Through breeding approaches, two superior varieties of *C. liberica*, designated as Liberoid Meranti 1 (Lim 1) and Liberoid Meranti 2 (Lim 2), were introduced in 2015. These varieties are known for their high adaptability in peatlands. The genetic basis of plant adaptability to peatlands remains largely unknown. It is therefore essential to identify genome-wide DNA polymorphisms in Lim 1 and 2 in order to gain insights into its capacity for adaptation in peatlands.

**Data description** Whole genome sequencing was performed on three plants from each variety (Lim 1 and 2), resulting in 430 million sequencing reads. The mean depth of sequencing for each sample was 36.90x. The reads were mapped to the *Coffea canephora* genome, with an average mapping rate of 96.34%. The sequencing data revealed the presence of 3,766,805 single-nucleotide polymorphisms (SNPs) and 1,123,683 insertion-deletions (indels) in all six plants. Among the SNPs, there was a notable prevalence of transitions, with a ratio of approximately twofold compared to transversions. The generated data offers invaluable genomic resources for marker development, with significant implications for understanding peatlands adaptability.

Keywords Coffea liberica, DNA polymorphism, Indel, Sequencing, SNP

# Objective

The majority of coffee types cultivated globally is arabica (*Coffea arabica*), followed by robusta (*Coffea canephora*) [1, 2]. In addition to these two species, the genus *Coffea* comprises several other species that produce coffee beans. *Coffea liberica* is a species from Liberia, which was then introduced to the rest of the world during the

\*Correspondence:

tisha.melia@lecturer.unri.ac.id

<sup>1</sup>Computer Science Department, Faculty of Mathematics and Natural

Sciences, Universitas Riau, Pekanbaru, Riau, Indonesia

<sup>2</sup>Biology Department, Faculty of Mathematics and Natural Sciences,

Universitas Riau, Pekanbaru, Riau, Indonesia

<sup>3</sup>Mathematics Department, Faculty of Mathematics and Natural Sciences, Universitas Riau, Pekanbaru, Riau, Indonesia nineteenth century, offering the advantage of resistance to the leaf rust disease caused by the fungus *Hemileia vastatrix* [3, 4]. Following further development, several genotypes of *C. liberica* were identified as being well adapted to peatlands. In 2015, the Indonesian Ministry of Agriculture released two superior varieties of *C. liberica* that are well-suited to cultivation in wetland, which are Liberoid Meranti 1 (Lim 1) and Liberoid Meranti 2 (Lim 2) [5]. The estimated coffee production potential for these two varieties is 1.69 and 1.98 tons per hectare for Lim 1 and Lim 2, respectively [5].

Tropical peatlands represent a distinctive ecosystem, occurring across a vast expanse from Southeast Asia, Central and South America to Africa [6, 7]. Plants in peatlands must adapt to a number of challenges, including lower nutrient levels and high water saturation [8].



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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 ine 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-nc-nd/4.0/.

Tisha Melia

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data set 1	Illumina Sequencing files for three Lim 1 and three Lim 2 samples.	Raw sequencing (fastq)	NCBI Sequence Read Archive (https://identifiers.org/ncbi/ins dc.sra:SRP537683 ) [17]
Data file 1	List of SNPs identified in Lim 1 and Lim 2 varieties.	Vcf files (vcf)	Figshare (https://doi.org/10.6084/m9.figshare.27201303.v 3) [18]
Data file 2	List of indels identified in Lim 1 and Lim 2 varieties.	Vcf files (vcf)	Figshare (https://doi.org/10.6084/m9.figshare.27925983.v 1) [19]
Data file 3	Pictures of plants used in this study	Picture (png)	Figshare (https://doi.org/10.6084/m9.figshare.28192670) [20]

Table 1 Overview of data files/data sets

Many plants can flourish naturally on peatlands, some of which have high economic value [9, 10]. The genetic underpinnings of this high adaptability in tropical peatlands have yet to be explored, which is essential for the development of other peat-adapted plant varieties. A genomic approach has been employed in a number of crops to identify DNA mutations associated with environmental adaptations, including variations in flower color in *Mimulus lewisii* [11], flowering time in *Arabidopsis thaliana* [12], and drought resistance in *Cedrus atlantica* [13]. In this study, we employed genomic approaches to identify DNA mutations in Lim 1 and Lim 2 with the aim of gaining insight into its adaptability in peatlands.

# **Data description**

Six young C. liberica plants were planted in the Botanical Garden of the Biology Department at Universitas Riau, Indonesia. The six plants were divided into two varieties: three Lim 1 and three Lim 2. These plants were obtained from a certified coffee breeding program on Meranti Island, overseen by the Indonesian Ministry of Agriculture. The plants were propagated from seeds collected from select parent trees within their tree improvement program. Specimens from all six plants were deposited in a herbarium (Herbarium Riauensis) and assigned the following voucher identifiers: CL-NNW-L1-202501 for Lim 1 plant 1, CL-NNW-L1-202502 for Lim 2 plant 2, CL-NNW-L1-202 503 for Lim 1 plant 3, CL-NNW-L2-202504 for Lim 2 plant 1, CL-NNW-L2-202505 for Lim 2 plant 2, and CL-NNW-L2-202506 for Lim 2 plant 3.

A sample of young leaves, comprising 50–100 mg, was taken from each of the six plants. The DNA was extracted using the Genomic DNA Mini Kit (Plant) Geneaid (catalog no. GP100), in accordance with the instructions provided in the kit. The quality of the extracted DNA was evaluated using a 1% agarose gel and a Nanodrop 2000 (Thermo Scientific, MA, USA). Subsequently, the extracted DNA was sent to an external service provider for library preparation and Illumina sequencing. The sequencing process yielded 430,270,734 151-bp paired-end reads (Table 1), with an average of 71.71 million paired-end reads per sample. In the absence of a

sequenced genome for *C. liberica*, we assume that its genome size is similar to that of other diploid *Coffea* sp. genomes, which is approximately 600 Mb. Our assumption is consistent with the estimated average total nuclear DNA content of *C. liberica* (1.59 picograms (pg)), which is similar to *C. canephora* at 1.46 pg, compared to the tetraploid *C. arabica* at 2.47 pg [14]. This allows us to conclude that the average depth of our sequencing per sample is 36.09, which is sufficient to discover DNA mutations [15]. Subsequently, all raw reads underwent quality control analysis using FastQC v0.11.8 [16], wherein 96.16% of the reads exhibited an average perbase quality score of at least Q30.

The raw sequencing reads from each sample were mapped to the C. canephora genome [1] using BWA-MEM [21] with the default parameters. The mapping of sequencing reads to the C. canephora genome was successful for an average of 96.34% of the reads per sample, which is indicative of the high degree of similarity between the genomes of C. canephora and C. liberica. Prior to utilizing the mapped reads for the identification of DNA mutations, we employed the MarkDuplicates tool to remove duplicated reads that were likely the result of PCR [22]. Subsequently, we followed the recommended workflow to identify germline short variants using GATK [19]. Furthermore, low-quality SNPs were removed based on several criteria, specifically QD < 5, QUAL<30, SOR>3, FS>10, and MQ<50. For indels, those with QD<2, QUAL<30, FS>200, and ReadPos-RankSum < -20 were removed. Following the removal of SNPs and indels that were not present in all samples, a total of 3,766,805 SNPs and 1,123,683 indels were identified (Table 1). 1.02% (38,354) of SNPs and 10.10% (113,439) of indels are multiallelic, indicating the presence of at least three alleles at each position. Among the SNPs, there were almost twice as many transitions (2,503,452) as transversions (1,302,755), which is consistent with the expectation that transitions require fewer changes to the DNA structure than transversions [23]. The number of transitions and transversions were cumulative across all six samples, and therefore may exceed the total number of SNPs. A total of 22.6% of the SNPs (853,209) were found to be overlapping with genes, with 5.5% (207,478) of the SNPs located within exon regions.

A total of 223 and 428 SNPs were identified as overlapping with the start and stop codons, respectively. The aforementioned overlapping statistics were calculated based on genome annotation with the NCBI accession number GCA\_900059795.1, which was subsequently transferred to the GCA\_036785865.1 genome using Liftoff with the default parameters [24].

# Limitations

The present dataset is limited to three samples for each variety (Lim 1 and Lim 2), resulting in a total of six samples (plants). This may be a relatively small number for the purpose of representing the genetic variations inherent to a given variety. Our dataset was sequenced using the Illumina paired-end short reads technology (150 bp), and as a consequence, we anticipate that larger DNA mutations, such as those involving large DNA rearrangements, may be undetected.

## Abbreviations

bp	Base pair
DNA	Deoxyribonucleic acid
indel	Insertion deletion
Lim 1	Liberoid Meranti 1 variety
Lim 2	Liberoid Meranti 2 variety
Mb	Megabase
pg	Picogram
SNP	Single nucleotide polymorphism

#### Acknowledgements

We thank the MAHAMERU BRIN HPC for the computational resources.

#### Author contributions

The study was conceptualized by TM and NNW. Funding was acquired by TM. NNW and SF organized sample collection and wet lab experiments. Bioinformatics analyses were undertaken by TM, F and AA. DIR provided assistance in grant writing and DNA extraction. All authors reviewed and approved the submitted manuscript.

#### Funding

Research funding was provided by the DRTPM Grant from The Directorate General of Higher Education, Research, and Technology.

#### Data availability

The data described in this Data note can be freely and openly accessed on NCBI Sequence Read Archive under accession ID SRP537683 (https://identifier s.org/ncbi/insdc.sra:SRP537683) and Figshare (https://doi.org/10.6084/m9.figs hare.27201303). Please see Table 1 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 no competing interests.

Received: 10 October 2024 / Accepted: 4 February 2025 Published online: 14 February 2025

#### References

- Salojärvi J, Rambani A, Yu Z, Guyot R, Strickler S, Lepelley M, et al. The genome and population genomics of allopolyploid Coffea arabica reveal the diversification history of modern coffee cultivars. Nat Genet. 2024;56:721–31. https:// doi.org/10.1038/s41588-024-01695-w.
- Mekbib Y, Tesfaye K, Dong X, Saina JK, Hu G-W, Wang Q-F. Whole-genome resequencing of Coffea arabica L. (Rubiaceae) genotypes identify SNP and unravels distinct groups showing a strong geographical pattern. BMC Plant Biol. 2022;22:69. https://doi.org/10.1186/s12870-022-03449-4.
- Wahibah NN, Putri RP, Muflikhah L, Martina A. Analysis of resistance to Fungal Pathogen Hemileia Vastatrix of Liberica Coffee based on functional marker. Int J Phytopathol. 2023;12:01–7. https://doi.org/10.33687/phytopath.012.01.4 371.
- Wahibah NN, Putri RP, Martina A, Arini A, Sidiq Y. Varietal Identification of Liberica Coffee in Kepulauan Meranti Riau using RAPD Marker: A Preliminary Study. Proc. 3rd Int. Conf. Biol. Sci. Educ. IcoBioSE. 2021, Atlantis Press; 2023, pp. 384–91. https://doi.org/10.2991/978-94-6463-166-1\_49
- Martono B, Sudjarmoko B, Udarno L. The potential of liberoid coffee cultivation on the peatlands (a case study: the peatlands in the Meranti island, Riau). IOP Conf Ser Earth Environ Sci. 2020;418:012022. https://doi.org/10.1088/175 5-1315/418/1/012022.
- Hodgkins SB, Richardson CJ, Dommain R, Wang H, Glaser PH, Verbeke B, et al. Tropical peatland carbon storage linked to global latitudinal trends in peat recalcitrance. Nat Commun. 2018;9:3640. https://doi.org/10.1038/s41467-01 8-06050-2.
- Ribeiro K, Pacheco FS, Ferreira JW, De Sousa-Neto ER, Hastie A, Krieger Filho GC, et al. Tropical peatlands and their contribution to the global carbon cycle and climate change. Glob Change Biol. 2021;27:489–505. https://doi.org/10.1 111/gcb.15408.
- Gao S, Song Y, Song C, Wang X, Gong C, Ma X, et al. Long-term nitrogen addition alters peatland plant community structure and nutrient resorption efficiency. Sci Total Environ. 2022;844:157176. https://doi.org/10.1016/j.scitote nv.2022.157176.
- Martins GS, Yuliarto M, Ching Yong W, Melia T, Maretha MV, Sharma M, et al. Estimation of Additive and Dominance effects in an *Acacia crassicarpa* multienvironment progeny trial using genomic Pedigree Reconstruction. Sci. 2024;fxae004. https://doi.org/10.1093/forsci/fxae004.
- Verhoeven JTA, Setter TL. Agricultural use of wetlands: opportunities and limitations. Ann Bot. 2010;105:155–63. https://doi.org/10.1093/aob/mcp172.
- Wu CA, Streisfeld MA, Nutter LI, Cross KA. The genetic basis of a Rare Flower Color Polymorphism in Mimulus lewisii provides insight into the repeatability of evolution. PLoS ONE. 2013;8:e81173. https://doi.org/10.1371/journal.pone. 0081173.
- Corrales A, Carrillo L, Lasierra P, Nebauer SG, Dominguez-Figueroa J, Renau-Morata B, et al. Multifaceted role of cycling DOF factor 3 (CDF3) in the regulation of flowering time and abiotic stress responses in *Arabidopsis*. Plant Cell Environ. 2017;40:748–64. https://doi.org/10.1111/pce.12894.
- Cobo-Simón I, Gómez-Garrido J, Esteve-Codina A, Dabad M, Alioto T, Maloof JN, et al. De novo transcriptome sequencing and gene co-expression reveal a genomic basis for drought sensitivity and evidence of a rapid local adaptation on Atlas cedar (Cedrus Atlantica). Front Plant Sci. 2023;14:1116863. https: //doi.org/10.3389/fpls.2023.1116863.
- Cros J, Gavalda M-C, Chabrillange N, Récalt C, Duperray C, Hamon S. Variations in the total nuclear DNA content in African Coffea species (Rubiaceae). Café Cacao Thé. 1994;38:3–3.
- 15. Koboldt DC. Best practices for variant calling in clinical sequencing. Genome Med. 2020;12:91. https://doi.org/10.1186/s13073-020-00791-w.
- Andrews S, FastQC:. accessed January 3, A Quality Control Tool for High Throughput Sequence Data 2010. http://www.bioinformatics.babraham.ac.u k/projects/fastqc/ (2023).
- Melia T, Fatayat F, Wahibah NN, Fatonah S. WGS data from two Coffea liberica varieties adapted to peatlands (Liberoid Meranti 1 and 2). 2024. https://identi fiers.org/ncbi/insdc.sra:SRP537683
- 18. Melia T, Fatayat F, Wahibah NN, Fatonah S. List of SNPs identified in Lim 1 and Lim 2 varieties 2024. https://doi.org/10.6084/m9.figshare.27201303.v3
- 19. Melia T, Fatayat F, Wahibah NN, Fatonah S. List of indels identified in Lim 1 and Lim 2 varieties 2024. https://doi.org/10.6084/m9.figshare.27925983.v1
- 20. Melia T, Fatayat F, Wahibah NN, Fatonah S. Pictures of plant samples from Lim 1 and Lim 2 varieties 2025. https://doi.org/10.6084/m9.figshare.28192670
- 21. Vasimuddin Md, Misra S, Li H, Aluru S. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. 2019 IEEE Int. Parallel Distrib. Process.

Symp. IPDPS, Rio de Janeiro, Brazil: IEEE; 2019, pp. 314–24. https://doi.org/10.1 109/IPDPS.2019.00041

- 22. Broad Institute. Picard Tools. Broad Inst GitHub Repos; 2024.
- Zou Z, Zhang J. Are nonsynonymous transversions generally more deleterious than nonsynonymous transitions? Mol Biol Evol. 2021;38:181–91. https:// doi.org/10.1093/molbev/msaa200.
- Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. Bioinformatics. 2021;37:1639–43. https://doi.org/10.1093/bioinformatics/btaa 1016.

# **Publisher's note**

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