Australian Journal of Crop Science AJCS

Aust J Crop Sci. 19(06):620-632 (2025) | ISSN:1835-2707
https://doi.org/10.21475/ajcs.25.19.06.p261

Molecular identification and phylogenetic reconstruction of sugarcane (*Saccharum officinarum* L.) cultivars from Indonesia based on *rbc*L chloroplast gene

Ganies Riza Aristya[1*], Tiara Putria Judith[1], Rina Sri Kasiamdari[2], Janis Damaiyani[3], Muhammad Fauzi Arif[4]

[1]Laboratory of Genetics and Breeding, Department of Tropical Biology, Faculty of Biology, Universitas Gadjah Mada, Yogyakarta, Indonesia
[2]Laboratory of Plant Systematics, Department of Tropical Biology, Faculty of Biology, Universitas Gadjah Mada, Yogyakarta, Indonesia
[3]Research Center for Plant Conservation, Botanic Gardens and Forestry, National Research and Innovation Agency, Bogor, Indonesia
[4]Laboratory of Biology, Department of Biology, Faculty of Mathematics and Natural Sciences, Samarinda, Indonesia

***Corresponding author: Ganies Riza Aristya** ✉
ORCID ID:
Ganies Riza Aristya https://orcid.org/0000-0001-9251-5076
Tiara Putria Judith https://orcid.org/0000-0001-9274-5728
Rina Sri Kasiamdari https://orcid.org/0000-0003-4125-1490
Janis Damaiyani https://orcid.org/0000-0003-0667-1908
Muhammad Fauzi Arif https://orcid.org/0000-0001-9964-1294

Submitted:* *01/11/2024

Revised:* *26/03/2025

Accepted:* *26/03/2025

Abstract: Sugarcane (*Saccharum officinarum* L.) presents high genetic complexity, posing challenges in breeding efforts aimed at developing superior hybrids. This research utilizes DNA barcoding, specifically leveraging the ribulose-1,5-biphosphate carboxylase-oxygenase large subunit (*rbc*L) gene as a DNA barcode. This research aims to assess the effectiveness of the *rbc*L gene in distinguishing local sugarcane cultivars and construct a phylogenetic tree to clarify their evolutionary relationships. The *rbc*L gene was amplified in 19 samples collected from several regions in Indonesia through PCR methods. DNA alignment, sequencing, and analyses were conducted using Maximum Likelihood (ML) and Bayesian Inference (BI) evolutionary models. Additionally, validation was performed using genetic distance, haplotype network, and principal coordinate of analysis (PCoA). The results established successful amplification and identification of samples within the *Saccharum* genus. Notably, samples BKS2, BKS7, OP3, NX04, Pringu, and Kidang Kencana displayed distincy characteristic in both ML and BI analyses, substantiated by polymorphic site variations and genetic distance assessments. However, the limitations associated with the *rbc*L gene constrained comprehensive analysis of other local sugarcane samples, hindering precise phylogenetics placement of the other samples.

Keywords; Bayesian Inference; bootstrap; clade; genetic; Maximum Likelihood.
Abbreviations: BI - Bayesian Inference; MCMC - *Markov Chain Monte Carlo*; ML - Maximum Likelihood; PCoA - Principal Coordinate of Analysis.

Introduction

Sugarcane is one of the economically significant plantation crops cultivated in more than 90 subtropical and tropical countries, contributing nearly 80% of global sugar production growth (Putra et el., 2020; Raj et al., 2016; Shabbir et al., 2021; Yang et al., 2016). Sugarcane holds substantial potential for converting solar energy into harvestable chemical energy, stored as accumulated sucrose and biomass (Mahadevaiah et al., 2021; Thorat et al., 2018). In Indonesia, sugarcane has been commercially cultivated as an industrial crop for over 100 years and considered a strategic national food commodity. However, Indonesia currently encounters challenges regarding sugarcane productivity capacity, which only fulfills about

38% of the total national demand (Widyasari et al., 2022). The average annual decrease in sugarcane production in Indonesia was 60.600 tons from 2016 to 2020 (Ditjenbun, 2022).
The present-day sugarcane cultivars are hybrid, resulting from interspesific hybridization, wherein approximately 80% of the chromosomes derived from *S. officinarum*, 10-15% from *S.spontaneum*, and 5-10% being recombinant chromosomes. These hybrid cultivars were developed through early sugarcane breeding efforts in Java and India at the end of the 19th century (Setta et al., 2014; Wang et al., 2023). The interspecific hybridization of sugarcane was subsequently accompanied by extensive backcrossing with noble cultivars (*S. officinarum*) and selective breeding, leading to highly polyploid hybrid sugarcane (ploidy levels of 10 or more). The total genome size of approximately 10 Gb, along with high complexity, also contributes to the challenges in breeding superior hybrids (Setta et al., 2014). Commercial hybrid cultivars such as PS 864, PSBM 901, PS 092, and PSBK 051 have been recognized, but still remain inadequate to fulfill national demands (Widyasari et al., 2022). A comprehensive understanding of the phylogenetic relationships among sugarcane cultivars, along with the strategic utilization of gene pools from wild relatives, could provide a foundation for selecting ancestral with potential genetic backgrounds, thereby enhancing the genetic diversity of sugarcane cultivars and improving sugarcane productivity in Indonesia (Govindaraj et al., 2021).
In recent decades, DNA-based techniques have been widely applied as an alternative to morphological and phytochemical identification (Peng et al., 2016). DNA barcoding is a technique developed to simplify and standardize the rapid characterization and classification of species using relatively small DNA fragments as markers, which are universally present in the target lineage (Gao et al., 2019; Peng et al., 2016). Short genomic fragments, known as DNA barcodes, have been successfully applied to the genus level in various organisms, including algae, fungi, plants, and animals (Pečnikar and Buzan, 2014; Sawarkar et al., 2021). The exploration of potential barcode genes in plants has proven to be more complex compared to animals (Guo et al., 2022; Singh and Banerjee, 2018). The chloroplast genome is considered a viable alternative to the mitochondrial genome in animals, contains barcode genes found in the majority of plant species (Pečnikar and Buzan, 2014). Numerous studies have utilized DNA barcoding techniques to describe the phylogenetic relationships among organisms. The Consortium for Barcode of Life Plant Working Group has proposed several potential candidate regions, including four coding regions: *mat*K, *rbc*L, rpoC1, and rpoB (CBOL, 2009). The chloroplast genes *mat*K and *rbc*L, utilized as single markers, have been widely applied as barcodes in species such as rice, nutmeg, sago palm, seaweed, dan flowering plants (Abbas et al., 2020; Alshehri et al., 2019; Liu et al., 2019; Singh and Banerjee, 2018; Tallei and Kolondam, 2015). The exploration of these barcode genes holds significant potential for their application in local sugarcane in Indonesia.
The *rbc*L gene, as a candidate barcode gene, has advantages due to its ease of amplification, sequencing, and alignment in most plants, as well as its effective utility in identification from family to genus levels (Mir et al., 2021; Zhu et al., 2022). The *rbc*L gene is located in the large single-copy (LSC) region of the chloroplast genome and encodes the enzyme ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO), which plays a role in carbon fixation during photosynthesis (Mir et al., 2021; Sarvananda, 2018; Xie et al., 2018). The high universality and conserved priming sites of this gene make it particularly useful in interspecies studies (Kress and Erickson, 2007). Studies on the *Acacia* species describe that the *rbc*L gene can reveal distant relationships, such as those between *A. albida*, *A. coriacea*, and *A. tortilis*, compared to *A. catechu* (Ismail et al., 2020). In Mentha species, phylogenetic analysis shows clade separation between *M. arvensis* and *M. spicata* from *M. citrata* (Thakur et al., 2016). However, the utilization of this gene in Indonesian sugarcane species, particularly at the cultivar level, has not been explored. Thus, the potential of this gene to elucidate evolutionary relationships at the intraspecific level remains largely undiscovered. Molecular studies involving sugarcane have been limited due to the crop's highly complex and polyploid genome. Therefore, this study aims to assess the effectiveness of the ribulose-1,5-bisphosphate carboxylase-oxygenase large subunit (*rbc*L) gene in local sugarcane cultivars and construct a phylogenetic tree to examine their evolutionary relationships. The outcomes of this study are projected to expand genetic diversity data for local sugarcane and encourage further research by sugarcane scientists in Indonesia, establish a strong foundation for selecting high-potential genotypes, thereby enhancing strategies for effective hybridization and selection.

Result and Discussion

In order to provide a foundation for sugarcane breeding in Indonesia, the advantages of DNA barcoding as modern molecular diagnostic technique, utilizing small DNA segments, are applied in this study to assess the effectiveness of the *rbc*L gene in characterizing the evolutionary relationships of sugarcane in Indonesia.

Quantification test result of genomics DNA

Genomic DNA successfully extracted from fresh leaf samples was quantified to confirm DNA presence and assess its concentration and purity. The concentrations ranged from 23.5 ng/µL to 180.97 ng/µL, while the purity (A260/A280) values ranging from 1.73 to 1.82. The Kidang Kencana sample exhibited the highest concentration at 180.97 ng/µL, whereas the JR01 sample had the lowest concentration at 23.5 ng/µL. The concentration values provide useful information for DNA dilution in PCR premix preparation (Aycan and Yildiz, 2024). The PSDK sample had the highest purity at 1.82, while the Pringu sample exhibited the lowest purity at 1.73. A260/A280 ratios between 1.8 and 2.0 are typically regarded as indicating high DNA purity (Abdel-Latif and Osman, 2017; Sophian et al., 2021). A ratio lower than 1.7 indicate contamination by proteins, phenols, or polysaccharides, whereas ratios above 2.0 indicate the presence of RNA. Contaminations may hinder the amplification and annealing efficiency (Irsyadi et al., 2024; Tiwari et al., 2017).

Table 1. Genetic distance of *Saccharum* spp. included other sugarcane species (%).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1																															
2	0.86																														
3	0.16	0.68																													
4	0.24	0.77	0.08																												
5	0.32	0.50	0.16	0.24																											
6	0.16	0.68	***	0.08	0.16																										
7	0.24	0.77	0.08	0.16	0.24	0.08																									
8	0.24	0.77	0.08	0.16	0.24	0.08	***																								
9	0.32	0.86	0.16	0.24	0.32	0.16	0.08	0.08																							
10	0.16	0.68	***	0.08	0.16	***	0.08	0.08	0.16																						
11	0.16	0.68	***	0.08	0.16	***	0.08	0.08	0.16	***																					
12	0.16	0.68	***	0.08	0.16	***	0.08	0.08	0.16	***	***																				
13	0.16	0.68	***	0.08	0.16	***	0.08	0.08	0.16	***	***	***																			
14	0.16	0.68	***	0.08	0.16	***	0.08	0.08	0.16	***	***	***	***																		
15	0.16	0.68	***	0.08	0.16	***	0.08	0.08	0.16	***	***	***	***	***																	
16	0.16	0.68	***	0.08	0.16	***	0.08	0.08	0.16	***	***	***	***	***	***																

17	0.16	0.68	***	0.08	0.16	***	0.08	0.08	0.16	***	***	***	***	***	***	***																	
18	0.16	0.68	***	0.08	0.16	***	0.08	0.08	0.16	***	***	***	***	***	***	***	***																
19	0.49	0.49	0.32	0.41	0.16	0.32	0.44	0.40	0.49	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32																
20	0.16	0.68	***	0.08	0.16	***	0.08	0.08	0.16	***	***	***	***	***	***	***	***	***	0.32														
21	0.16	0.68	***	0.08	0.16	***	0.08	0.08	0.16	***	***	***	***	***	***	***	***	***	0.32	***													
22	0.16	0.68	***	0.08	0.16	***	0.08	0.08	0.16	***	***	***	***	***	***	***	***	***	0.32	***	***												
23	0.16	0.68	***	0.08	0.16	***	0.08	0.08	0.16	***	***	***	***	***	***	***	***	***	0.32	***	***	***											
24	0.16	0.68	***	0.08	0.16	***	0.08	0.08	0.16	***	***	***	***	***	***	***	***	***	0.32	***	***	***	***										
25	0.16	0.68	***	0.08	0.16	***	0.08	0.08	0.16	***	***	***	***	***	***	***	***	***	0.32	***	***	***	***	***									
26	0.16	0.68	***	0.08	0.16	***	0.08	0.08	0.16	***	***	***	***	***	***	***	***	***	0.32	***	***	***	***	***	***								
27	0.16	0.68	***	0.08	0.16	***	0.08	0.08	0.16	***	***	***	***	***	***	***	***	***	0.32	***	***	***	***	***	***	***							
28	0.32	0.87	0.16	0.24	0.32	0.16	0.24	0.24	0.32	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.50	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16						
29	0.24	0.77	0.08	0.16	0.24	0.08	0.16	0.16	0.24	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.41	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.24				
30	0.24	0.67	0.08	0.16	0.24	0.08	0.16	0.16	0.24	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.40	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.24	0.16			
31	0.16	0.68	***	0.08	0.16	***	0.08	0.08	0.16	***	***	***	***	***	***	***	***	***	0.32	***	***	***	***	***	***	***	***	***	0.16	0.08	0.08		

Table 2. Polymorphic sites of sugarcane inferred from *rbc*L gene (site 3 to 1324).

Nucleotide Site Number	3	4	6	7	8	12	16	21	22	23	364	392	397	533	574	640	641	642	788	892	1144	1377	1429	1490	1019	1197	1279	1280	1282	1306	1307	1337	1338	1339	1340	1341	1343	1344
BKS1	A	A	T	-	G	-	T	-	C	-	-	-	T	-	-	-	-	G	T	T	-	-	-	-	A	A	-	T	T	C	T	-	G	A	-	G	-	-
BKS2	G	C	G	A	·	·	C	·	T	A	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	A	A	G	·	·	G	A	T	C	A	·	C
BKS3	·	·	·	·	·	·	·	T	·	·	·	·	·	G	·	·	·	·	·	·	·	·	·	·	·	·	A	A	G	·	-	G	·	·	·	·	·	C
BKS5	·	·	·	A	·	·	·	·	·	·	·	·	·	·	A	C	G	A	·	·	·	·	·	·	·	·	A	A	G	·	·	G	·	·	·	·	C	C
BKS7	·	·	G	A	·	A	C	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	A	A	G	·	·	G	·	·	·	·	·	·
POJ	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	A	A	G	·	·	G	·	·	·	·	·	·
Kidang Kencana	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	C	A	A	G	·	·	G	·	·	·	·	·	·
NX04	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	C	A	A	G	·	·	G	·	·	·	·	·	·
Pringu	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	G	G	·	G	·	C	A	A	G	·	·	T	·	·	·	·	·	·
Bulu Lawang	·	·	·	·	·	·	·	·	·	·	T	A	·	·	·	·	·	·	·	·	·	·	·	·	·	·	A	A	G	·	·	G	·	·	·	·	·	C
PS862	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	A	A	G	·	·	G	·	·	·	·	·	·
PS881	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	A	A	G	·	·	G	·	·	·	·	·	·
PSBM	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	A	A	G	·	·	G	·	·	·	·	·	·
PSDK	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	G	·	·	·	·	·	A	A	G	·	·	G	·	·	·	·	·	·
TLH	·	·	·	·	·	·	·	·	·	·	A	A	·	·	·	·	·	·	·	·	·	·	·	·	·	·	A	A	G	·	·	G	·	·	·	·	·	C
JR01	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	-	·	A	A	G	·	·	G	·	·	·	·	·	·
JR02	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	A	A	G	·	·	G	·	·	·	·	·	·
OP2	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	A	A	G	·	·	G	·	·	·	·	·	·
OP3	·	·	G	A	T	A	C	·	T	A	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	A	A	G	·	·	G	·	·	·	·	·	·
BK010677	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	A	A	G	·	·	G	·	·	·	·	·	·
LS975132	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	A	A	G	·	·	G	·	·	·	·	·	·
NC_072525	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	A	A	G	·	·	G	·	·	·	·	·	·
NC_072524	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	A	A	G	·	·	G	·	·	·	·	·	·
MT921804	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	A	A	G	·	·	G	·	·	·	·	·	·
MT721155	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	A	A	G	·	·	G	·	·	·	·	·	·
MZ328080	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	A	A	G	·	·	G	·	·	·	·	·	·
MG685915	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	A	A	G	·	·	G	·	·	·	·	·	·
MF563371	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	C	·	·	·	·	·	·	·	A	A	G	T	·	G	·	·	·	·	·	·
NC_070069	·	·	·	·	·	·	·	·	·	·	·	·	C	·	·	·	·	·	·	·	·	·	·	·	·	·	A	A	G	·	·	G	·	·	·	·	·	·
PP816722	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	A	A	G	·	·	G	·	·	·	C	·	·
KX507245	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	A	A	G	·	·	G	·	·	·	·	·	·

Note: 1. BKS1, 2. BKS2, 3. BKS3, 4. BKS5, 5. BKS7, 6. POJ, 7. Kidang Kencana, 8. NX04, 9. Pringu, 10. Bulu Lawang, 11. PS862, 12. PS881, 13. PSBM, 14. PSDK, 15. TLH, 16. JR01, 17. JR02, 18. OP2, 19. OP3, 20. BK010677, 21. LS975132, 22. NC_072525, 23. NC_072524, 24. MT921804, 25. MT721155, 26. MZ328080, 27. MG685915, 28. MF563371, 29. NC_070069, 30. PP816722, 31. KX507245.

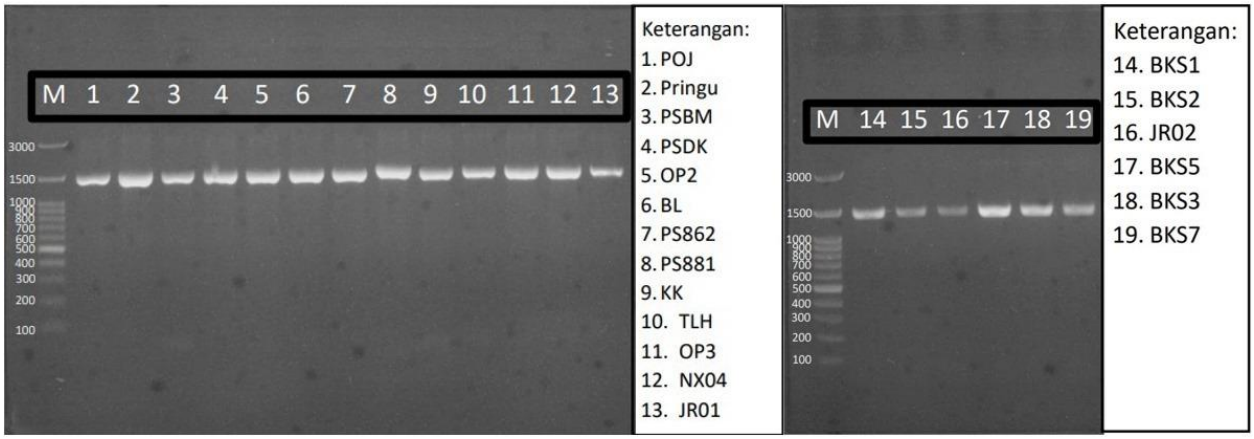

Fig 1. Electrophoregram of the rbcL gene from 19 samples.

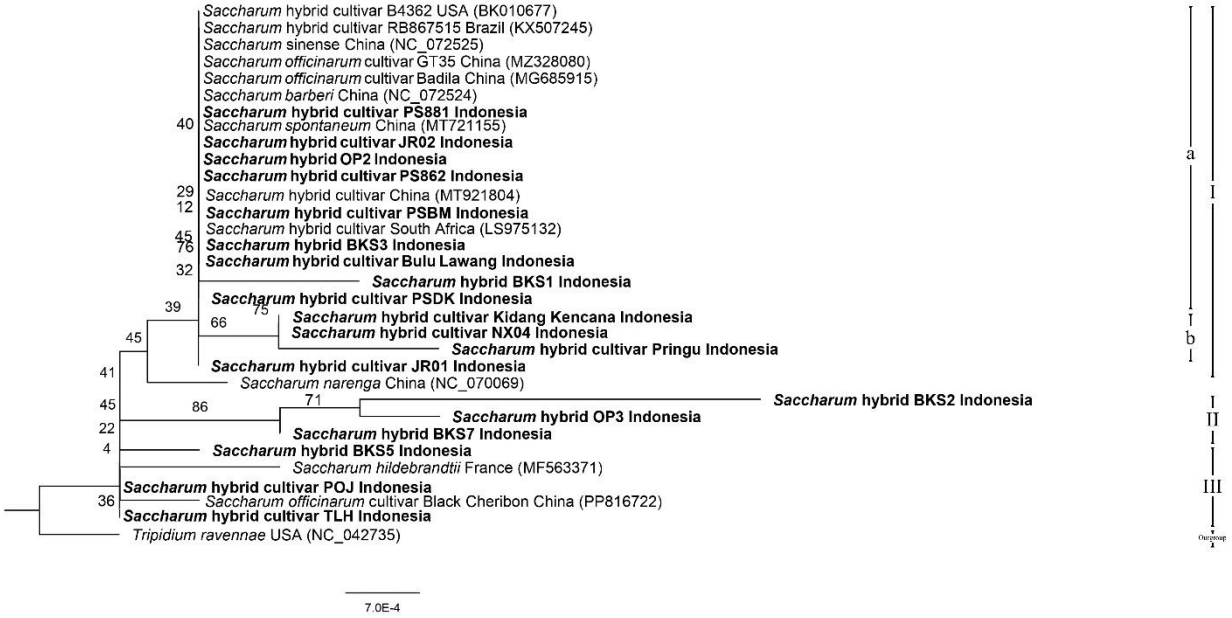

Fig 2. Maximum Likelihood phylogenetic tree of *Saccharum* spp. included other sugarcane species and outgroup from GenBank based on *rbc*L sequence.

Qualitative test result of rbcL gene amplification

Samples were confirmed to be of adequate purity and concentration were subsequently used as templates for amplification via the PCR technique. PCR is widely applied to replicate or amplify small quantities of specific DNA sequences (one or more genes) into thousands to millions of copies, thus providing high efficiency in amplifying the target gene (Semagn et al., 2006; Solanki, 2012). The main steps in the PCR technique include denaturation, annealing, and extension. Denaturation is carried out at 95°C for 30 seconds to separate the double-stranded DNA into single strands. Annealing involves lowering the temperature to 54.4°C, allowing primers to bind to the complementary single-stranded template. Extension requires increasing the temperature to 72°C to activate the DNA polymerase enzyme (Taq Polymerase), which operates optimally at this temperature. In this study, PCR was performed using *rbc*L-F and *rbc*L-R primers. This process aimed to confirm the successful amplification of the *rbc*L gene from sugarcane genomic DNA. The amplification products were subjected to a qualitative analysis using a 1% agarose gel, run at 100V for 30 minutes. Successful amplification of the *rbc*L gene was indicated by the visualization of a band close to the approximately 1500 bp ladder parameter (Fig. 1). The *rbc*L gene has a length of 1431 bp (Wu et al., 2016).

Sequencing result

The samples were aligned using the Sanger sequencing method at LPPT UGM (Applied Biosystem 3500 Genetic Analyzer 2550). The consensus results from 19 samples ranged from 1314 bp to 1346 bp, and were subsequently analyzed using the BLAST algorithm. All samples were confirmed to originate from the *Saccharum* genus, with a query cover of 99% and a percent identity ranging from 99.33% to 99.93%. The query cover and percent identity values indicate a high nucleotide sequence similarity between the samples and those in the database (Newell et al., 2013).

Fig 3. Bayesian Inference phylogenetic tree of *Saccharum* spp. included other sugarcane species and outgroup from GenBank based on *rbc*L sequence.

Reconstruction of phylogenetic tree result

The evolutionary models applied to determine the phylogenetic relationships among sugarcane samples were Maximum Likelihood (ML) and Bayesian Inference (BI). The ML analysis revealed that the open field samples BKS7, OP3, and BKS2 formed a distinct clade with a bootstrap value of 86. Additionally, the cultivars Kidang Kencana, NX04, and Pringu also formed a separate group from the sugarcane cultivars of other countries, although still within the same clade (Fig 2). The bootstrap value reflects the frequency of branch occurrences in data replication process (Shen et al., 2016). The BI analysis provided a similar phylogenetic reconstruction as the ML analysis, with a posterior probability of 1. This evolutionary model further examines the likelihood and prior information (Fig 3). The length of the branches in the phylogenetic tree represents the evolutionary distance between sugarcane species, with longer branches indicating a greater divergence (Binet et al., 2016). The analysis revealed distinctive patterns for the samples BKS2, BKS7, OP3, NX04, Pringu, and Kidang Kencana, highlighting the impact of backcross hybridization and prolonged selection on the DNA structure based on the *rbc*L gene. However, this single gene was not sufficiently powerful to deeply analyze the differences in DNA sequences among other sugarcane samples, as evidenced by the lack of distinct separation between cultivars and sugarcane species in the database. This indicates that the *rbc*L gene has limitations in describing phylogenetic relationships at levels below the family or genus. A study by Sundari et al. (2019) on clove plants demonstrated that the *rbc*L gene has high resolution for identifying phylogenetic relationships at the genus level or higher.

Genetic distance provides valuable information regarding the genetic divergence between organisms, and the understanding can aid in deciphering evolutionary patterns and divergence within species (Dogan and Dogan, 2016; Wei et al., 2020). Based on p-distance analysis, various genetic distances ranging from 0% to 0.87% were constructed (Table 1). P-distance values below 1% suggest minimal genetic variation, while values approaching 1% indicate considerable genetic variation, which could represent significant differences in traits or environmental adaptations (Shen et al., 2016). In this study, BKS2 demonstrated a genetic distance of 0.87%, reflecting considerable genetic variation. Conversely, other samples displayed no genetic distance (0%), indicating a high degree of genetic uniformity. This information provides a critical basis for breeders seeking to develop sugarcane cultivars with either uniform or diverse genetic traits. BKS2, in particular, presents potential as a candidate parent for breeding programs aimed at enhancing genetic diversity.

Polymorphic sites analysis result

Genetic polymorphism refers to the presence of two or more alleles at a single locus due to nucleotide sequence variation within a population. The analysis revealed 38 polymorphic sites (2.86%), which consist of 22 singleton variable sites

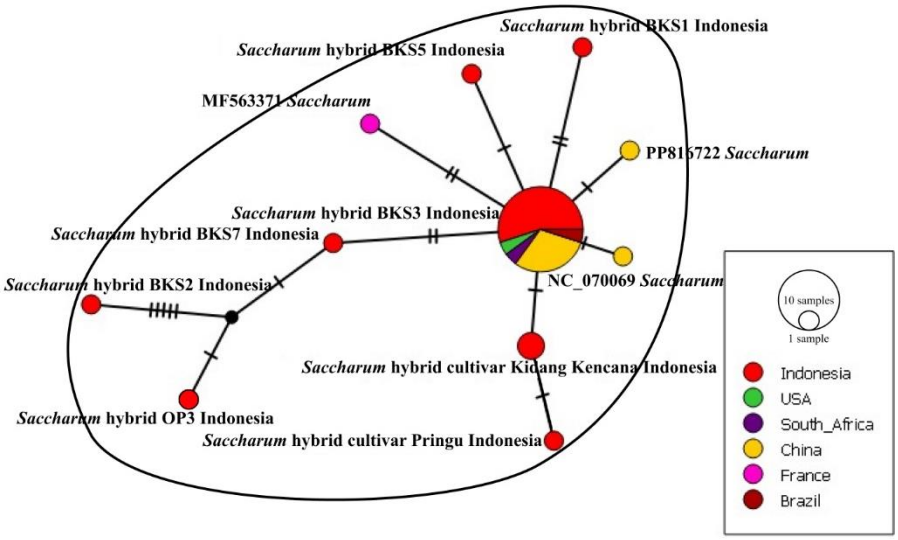

Fig 4. Median-joining haplotype network of sugarcane based on *rbc*L chloroplast gene.

(1.66%), 3 indels (0.22%), and 13 parsimony-informative sites (0.98%) (Table 2). Polymorphic sites represent nucleotide variation positions within a DNA sequence that illustrate genetic diversity among individuals (Dutra et al., 2009; Jiang and Huang, 2015).

Haplotype network and principal coordinate of analysis (PCoA) result

The polymorphic sites identified were grouped into 17 haplotypes. The haplogroup analysis indicated that cultivars PS862, PS881, JR02, OP2, and PSBM clustered together with many sugarcane species from around the world in group 13, while BKS1 and Kidang Kencana were assigned to group 15. The remaining sugarcane samples exhibited distinct alleles, forming their own separate haplogroups (Table 3). This grouping reflects the combination of allele locations on homologous chromosomes (Jin et al., 2023).

The haplotype network revealed that most Indonesian samples share genetic material with samples from the United States, South Africa, China, and Brazil (Fig 4). However, samples BKS2, OP3, Pringu, Kidang Kencana, BKS1, BKS7, and BKS5 did not exhibit a close clustering pattern with other countries, indicating a unique distribution pattern. The genetic similarity among countries can be attributed to the ancestral contributions from various hybridization events, particularly involving *S. officinarum* originating from Indonesia. This results in a genetic exchange phenomenon commonly referred to as gene flow (Babu et al., 2022). The analysis also illustrates variations in circle sizes, which correspond to the number of samples, and the lines linking haplotypes, which represent evolutionary pathways. The short lines connecting haplotypes indicate mutation points. The colors used in the figures correspond to the sample collection locations in this study (Arisuryanti et al., 2024). The Principal Coordinate of Analysis (PCoA) supported the findings from the haplotype network, the coordinates of closely associated haplotype points indicate substantial genetic similarity. Furthermore, the PCoA highlighted that haplotypes H9 (Pringu), H2 (OP3), and H4 (BKS3) did not cluster closely, reinforcing the potential uniqueness of these samples and suggesting that further investigation is necessary to explore their distinct characteristics (Fig 5).

Materials and methods

Sample collection and species identification

This study was conducted at the Genetics and Breeding Laboratory, Faculty of Biology, Universitas Gadjah Mada, Indonesia. Samples for this research were collected from several locations, including the collection of PT Madubaru farm, Special Region of Yogyakarta, the Center for Seeding and Protection of Plantation Crops (BBP2TP) in Surabaya, East Java, as well as open fields in Pakem, Yogyakarta, and Bekasi, West Java (Fig 6; Table 4). Each sugarcane leaf sample was placed in a ziplock bag, stored in a cool box, and promptly transported to the laboratory for storage at -20°C. A total of 19 samples (Table 1) were subsequently used for DNA extraction, amplification, sequencing, and DNA barcoding Additionally, 12 sequences of the *rbc*L gene from related sugarcane species and one outgroup were acquired from GenBank for further analysis (Table 5).

DNA extraction, visualization, and quantification

Fresh leaf samples were ground into powder with the addition of liquid nitrogen using a mortar and pestle. Genomic DNA was extracted from 100 mg of leaf samples using Plant Genomic DNA Mini Kit (GeneAid), following the manufacturer's instructions. The quality of the genomic DNA was visualized on a 0.8% agarose gel. The DNA extract was stored at -20°C and used as a template for DNA barcoding. The DNA quantification and purity were determined using spectrophotometry (NanoDrop Lite Plus Spectrophotometer).

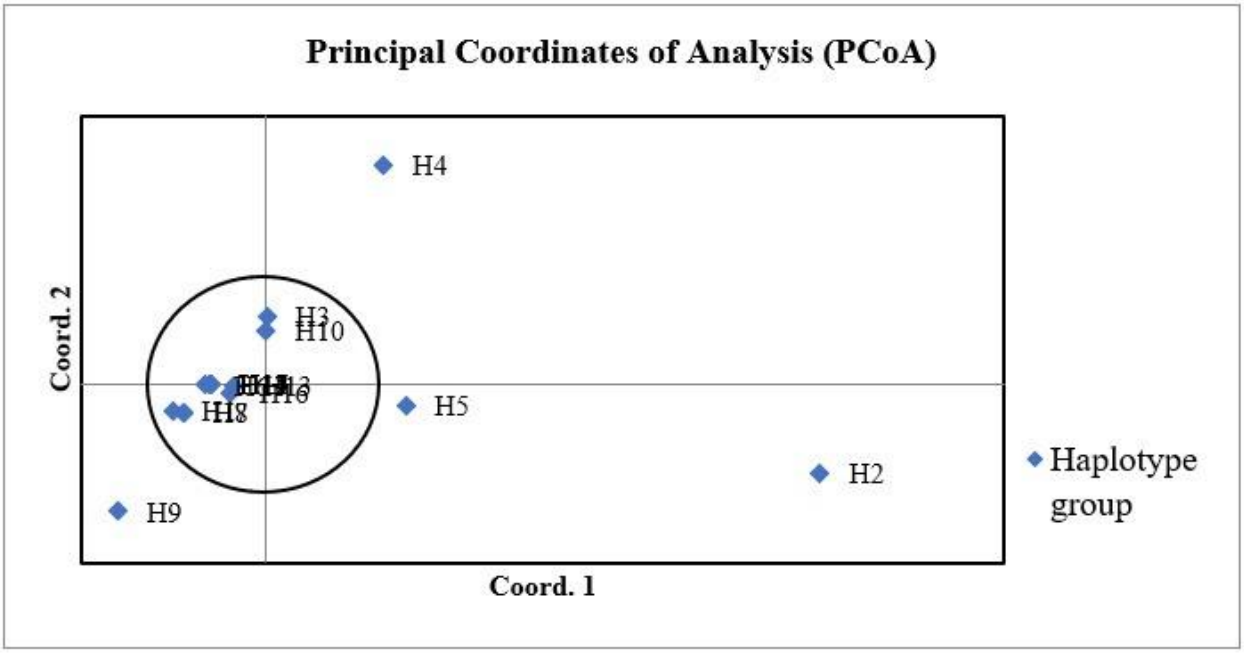

Fig 5. Principal Coordinate of Analysis (PCoA) of sugarcane based on *rbc*L chloroplast gene.

Table 3. Haplogroup of thirty-one sugarcanes based on *rbc*L.

Haplotype	Number of Samples	Sample Codes
H1	1	BKS7
H2	1	OP3
H3	1	BKS5
H4	1	BKS3
H5	1	TLH
H6	1	Bulu Lawang
H7	1	NC_070069
H8	1	MF563371
H9	1	Pringu
H10	1	PSDK
H11	1	POJ
H12	1	PP816722
H13	14	PS862
		PS881
		JR02
		OP2
		BK010677
		LS975132
		NC_072525
		NC_072524
		MT921804
		MT721155
		MZ328080
		MG685915
		KX507245
		PSBM
H14	1	BKS1
H15	2	Kidang Kencana
		NX04
H16	1	JR01
H17	1	BKS2

Amplification and Sequencing of rbcL

The chloroplast *rbc*L gene was amplified using a T100 Thermal Cycler (Biorad) with two primers: Forward (5'-ATGTCACCACAAACAGAAAC-3') and Reverse (5'-CTATAAGGTATCCATCGCT-3') (Wu et al., 2016). Each reaction mixture had a

Table 4. List of sugarcane cultivars used in this study.

Species	Cultivar/ Sample	Locations site	Collection Years	Latitude (N)	Longitude (E)	Sampling Locations
Saccharum hybrid	POJ	Yogyakarta	2023	7°49'40"S	110°20'14"E	PT Madubaru Collection Farm
Saccharum hybrid	PSBM					
Saccharum hybrid	PSDK					
Saccharum hybrid	BL					
Saccharum hybrid	PS862					
Saccharum hybrid	PS881					
Saccharum hybrid	KK					
Saccharum hybrid	TLH					
Saccharum hybrid	NX04					
Saccharum hybrid	JR01	Jawa Timur	2018	7°34'03"S	112°19'54"E	BBP2TP, Surabaya
Saccharum hybrid	JR02					
Saccharum hybrid	Pringu					
Saccharum hybrid	OP2	Yogyakarta	2024	7°45'17"S	110°21'01"E	*Open Field*
Saccharum hybrid	OP3					
Saccharum hybrid	BKS1	Bekasi	2024	6°09'43"S	107°02'40"E	
Saccharum hybrid	BKS2					
Saccharum hybrid	BKS3					
Saccharum hybrid	BKS5			6°12'27"S	107°01'01"E	
Saccharum hybrid	BKS7			6°12'50"S	107°01'31"E	

Table 5. List of sugarcane samples from the database.

Species	Source	FIPS	Accession Number	Reference
Saccharum hybrid cultivar B4362	USA	US	BK010677	Welker et al., 2019
Saccharum hybrid cultivar	South Africa	SA	LS975132	GenBank
Saccharum hybrid cultivar RB867515	Brazil	BR	KX507245	GenBank
Saccharum spontaneum	China	CH	MT721155	GenBank
Saccharum hybrid cultivar	China	CH	MT921804	GenBank
Saccharum officinarum cultivar GT35	China	CH	MZ328080	GenBank
Saccharum officinarum cultivar Badila	China	CH	MG685915	GenBank
Saccharum sinense	China	CH	NC_072525	GenBank
Saccharum barberi	China	CH	NC_072524	GenBank
Saccharum narenga	China	CH	NC_070069	GenBank
Saccharum officinarum cultivar Black Cheribon	China	CH	PP816722	GenBank
Saccharum hildebrandtii	France	FR	MF563371	Piot et al., 2018
Tripidum ravennae	USA	US	NC_042735	Welker et al., 2019

volume of 30 µl in a PCR tube, with 5 µl of DNA template at a concentration of 50 ng/ µL, 12.5 µl of MyTaq HS Red Mix (Bioline) as the master mix, 2 µl of each Forward and Reverse primer (10 µM), and 8.5 µL of double-distilled water (ddH2O). Amplification was carried out with the following protocol: pre-denaturation at 95°C for 5 minutes, denaturation at 95°C for 30 seconds, annealing at 54-54.4°C for 30 seconds, extension at 72°C for 45 seconds for 35 cycles, and final extension at 72°C for 5 minutes. PCR products were electrophoresed and visualized on a 1% agarose gel with 1X TBE buffer, stained with Florosafe, and run at 100V for 30 minutes. The gel was visualized under UV light using a GelDoc system, and the fragment sizes were compared with a 100 bp DNA ladder for confirmation. Subsequently, samples were sent to the Integrated Research and Testing Laboratory at UGM for Sanger DNA sequencing.

Sequence editing and alignment

The nucleotide sequences were reviewed for base miscalls and ambiguities, and contigs were assembled using GeneStudio software. The consensus sequence obtained was compared with available reference nucleotide sequences from GenBank at

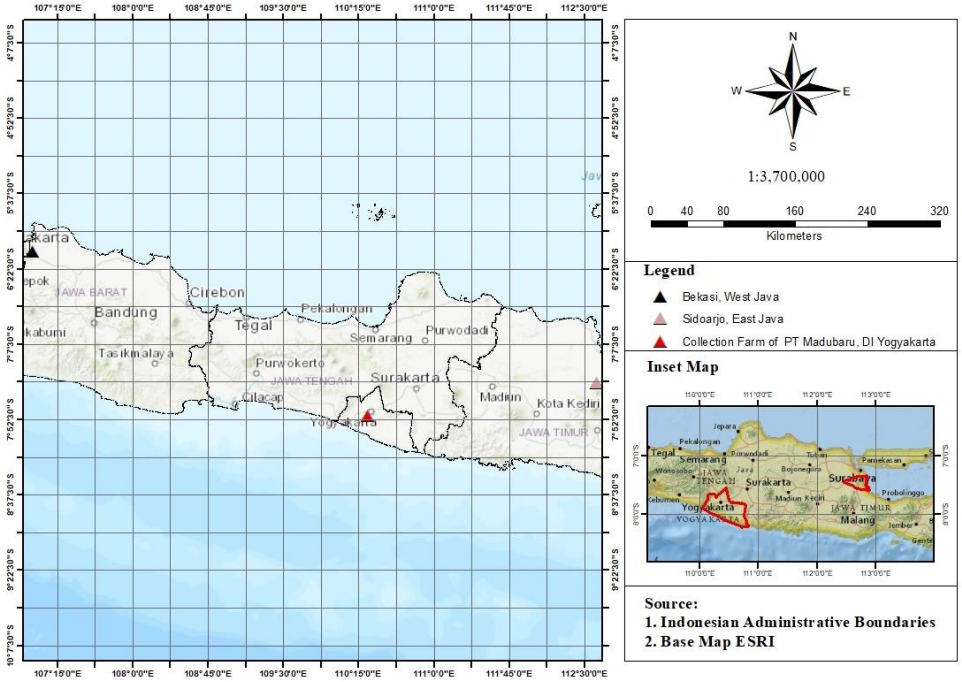

Fig 6. Sampling collection sites.

NCBI using the BLAST algorithm (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The *rbc*L gene sequence files from the 19 study samples were aligned with related sugarcane species and outgroup sequences using MEGA 11.0, utilizing the MUSCLE alignment tool, without affecting the length of the DNA sequences.

Phylogenetic inference and genetic distance

Phylogenetic analysis was reconstructed using two model-based inference methods: Maximum Likelihood (ML) and Bayesian Inference (BI). PartitionFinder v2.1.1 was used to select optimal partitioning schemes based on the Bayesian Information Criterion (BIC) to minimize saturation effects and enhance the phylogenetic resolution of the DNA sequence dataset. ML analysis was performed using IQ-TREE v.2.3.1 with 1000 replications, employing the best scheme generated by PartitionFinder. BI analysis was carried out using MrBayes v.3.2.6 based on *Markov Chain Monte Carlo* (MCMC) with 4,000,000 generations. Reversible-jump MCMC was utilized to allow sampling across all substitution models rather than determining a specific substitution model suggested by PartitionFinder. The outgroup used in this study was *Tripidium ravennae* (GenBank: NC_042735). The phylogenetic tree was visualized using FigTree 1.4.4.0. Genetic distance analysis was performed using the Kimura-2 parameter method (K-2P) with MEGA 11.0 software.

Polymorphic sites, haplotype network, dan principal coordinate of analysis (PCoA)

Intraspecific genetic variation analysis in this study included the number of polymorphic sites and parsimony sites, which were analyzed using DnaSP software and confirmed with GenAIEx 6 software. The haplotype network was constructed using PopART v.1.7, and principal coordinate of analysis (PCoA) was generated using GenAIEx 6.

Conclusion

Based on the *rbc*L gene, all samples were identified as belonging to the *Saccharum* genus, yielding a fragment of 1325 bp. Phylogenetic relationships inferred using ML and BI models revealed unique characteristics in the samples BKS2, BKS7, OP3, NX04, Pringu, and Kidang Kencana, supported by polymorphic sites and genetic distance. PCoA analysis further indicated that these samples did not show genetic flow with other countries, validating that samples originated from genetic assemblies based on selection by sugarcane breeders in Indonesia. However, the *rbc*L gene was not sufficiently robust to deeply analyze and distinguish among other local sugarcane samples. This limitation underscores the need for the use of multiple genetic markers to provide deeper resolution in explaining evolutionary relationships at the species level or lower.

Acknowledgment

This research is funded by The Indonesia Endowment Funds for Education (LPDP) Ministry of Finance No: 30/IV/KS/05/2023 and 2090/UN1/DITLIT/ Dit-Lit/PT.01.03/2023 collaborated with the National Research and Innovation Agency through a Research and Innovation Grant for Advanced Indonesia. The research expresses gratitude to the head and staff of Madubaru Factory for their collaboration and assistance during the research process. The researchers also express gratitude to the Faculty of Biology UGM for granting permission for the research during data collection.

Statement of contributions

GRA: Conceptualization, supervision, result validation, writing the original draft, project and funding administration, reviewing and editing the manuscript. TPJ: Formal analysis, writing the original draft, investigation, and software analysis. RSK: result validation, methodology. JD: formal analysis, writing the original draft, investigation: MFA: Formal analysis, writing the original draft, reviewing and editing the manuscript.

References

Abbas B, Tjolli I, Munarti (2020) Genetic diversity of sago palm (*Metroxylon sagu*) accessions based on plastid cpDNA *mat*K gene as DNA barcoding. Biodiversitas 21(1): 219–225.

Abdel-Latif A, Osman G (2017) Comparison of three genomic DNA extraction methods to obtain high DNA quality from maize. Plant Methods 13(1): 1–9.

Alshehri MA, Aziz AT, Alzahrani O, Alasmari A, Ibrahim S, Osman G, Bahattab O (2019) DNA-barcoding and species identification for some Saudi Arabia seaweeds using *rbc*L gene. Journal of Pure and Applied Microbiology 13(4): 2035–2044.

Arisuryanti T, Waskito Aji K, Nur Shabrina F, Febriyanti D, Setiadi Daryono B, Sendi Priyono D (2024) Phylogenetic and genetic variation of common mudskippers (*Periophthalmus kalolo* Lesson, 1831) from the southern coast of Java, Indonesia inferred from the COI mitochondrial gene. Journal of Genetic Engineering and Biotechnology 22(1): 1–10.

Babu KSD, Janakiraman V, Palaniswamy H, Kasirajan L, Gomathi R, Ramkumar TR (2022) A short review on sugarcane: its domestication, molecular manipulations and future perspectives. Genetic Resources and Crop Evolution 69(8): 2623–2643.

Binet M, Gascuel O, Scornavacca C, Emmanuel EJ, Pardi F (2016) Fast and accurate branch lengths estimation for phylogenomic trees. BMC Bioinformatics 17(23): 1-18.

Consortium Barcode of Life (2009) A DNA barcode for land plants. PNAS 106, 12794–12797.

Ditjenbun (2022) Statistik Unggulan 2020-2022. Kementerian Pertanian, Jakarta, 835-858.

Dogan İ, Dogan N (2016) Genetic distance measures: Review. Turkiye Klinikleri Journal of Biostatistics 8(1): 87–93.

Dutra WO, Moreira PR, Souza PEA, Gollob KJ, Gomez RS (2009) Implications of cytokine gene polymorphisms on the orchestration of the immune response: Lessons learned from oral diseases. Cytokine and Growth Factor Reviews 20(3): 223–232.

Gao Z, Liu Y, Wang X, Wei X, Han J (2019) DNA mini-barcoding: A derived barcoding method for herbal molecular identification. Frontiers in Plant Science 10: 1–11.

Govindaraj P, Gowri R, Mohanraj K, Amalraj VA (2021) SSR marker based molecular genetic diversity analysis among *Saccharum spontaneum* (L.) collected from North Western region of India. Sugar Tech 23(4): 730–740.

Guo M, Yuan C, Tao L, Cai Y, Zhang W (2022) Life barcoded by DNA barcodes. Conservation Genetics Resources 14(4): 351–365.

Irsyadi MB, Sari SK, Oktiastuti E, Rineksane IA (2024) Rapid genomic DNA extraction for Soybean (*Glycine max* L. Merr) using modified CTAB protocol to obtain high-quality DNA. Indian Journal of Biochemistry and Biophysics 61(2): 97–104.

Ismail M, Ahmad A, Nadeem M, Javed MA, Khan SH, Khawaish I, Sthanadar AA, Qari SH, Alghanem SM, Khan KA, Khan MF, Qamer S (2020) Development of DNA barcodes for selected *Acacia* species by using *rbc*L and *mat*K DNA markers. Saudi Journal of Biological Sciences 27(12): 3735–3742.

Jiang Y, Huang C (2015) Polymorphism and cognition: The example of folate and MTHFR C677T. 'Diet Nutr. Dement. Cogn. Decline'. Elsevier Inc. Amsterdam, pp. 357–367.

Jin X, Du H, Zhu C, Wan H, Liu F, Ruan J, Mower JP, Zhu A (2023) Haplotype-resolved genomes of wild octoploid progenitors illuminate genomic diversifications from wild relatives to cultivated strawberry. Nature Plants 9(8): 1252–1266.

Kress WJ, Erickson DL (2007) A Two-Locus Global DNA Barcode for Land Plants: The Coding *rbc*L Gene Complements the Non-Coding *trn*H-*psb*A Spacer Region. PLoS ONE 2(6): 1–10.

Liu Y, Wang K, Liu Z, Luo K, Chen S, Chen K (2013) Identification of medical plants of 24 ardisia species from China using the *mat*K genetic marker. Pharmacognosy Magazine 9(36): 331–337.

Mahadevaiah C, Appunu C, Aitken K, Suresha GS, Vignesh P, Mahadeva Swamy HK, Valarmathi R, Hemaprabha G, Alagarasan G, Ram B (2021) Genomic selection in sugarcane: Current status and future prospects. Frontiers in Plant Science 12: 1–18.

Mir RA, Bhat KA, Rashid G, Ebinezer LB, Masi A, Rakwal R, Shah AA, Zargar SM (2021) DNA barcoding: a way forward to obtain deep insights about the realistic diversity of living organisms. Nucleus 64(2): 157–165.

Newell PD, Fricker AD, Roco CA, Chandrangsu P, Merkel SM (2013) A small-group activity introducing the use and interpretation of BLAST. Journal of Microbiology & Biology Education 14(2): 238–243.

Pečnikar ŽF, Buzan EV (2014) 20 years since the introduction of DNA barcoding: From theory to application. Journal of Applied Genetics 55(1): 43–52.

Peng X, Wu X, Ji Q, Yang R, Li Y (2016) Molecular authentication of *Tetrastigma hemsleyanum* from its adulterant species using ISSR, CAPS, and ITS2 barcode. Molecular Biology Reports 43(8): 785–794.

Putra RP, Ranomahera MRR, Rizaludin MS, Supriyanto R, Dewi VAK (2020) Short communication: Investigating environmental impacts of long-term monoculture of sugarcane farming in Indonesia through DPSIR framework. Biodiversitas 21(10): 4945–4958.

Raj P, Selvi A, Prathima PT, Nair N V. (2016) Analysis of genetic diversity of *Saccharum* complex using chloroplast microsatellite markers. Sugar Tech 18(2): 141–148.

Sarvananda L (2018) Short introduction of DNA barcoding. International Journal of Research 5(4): 673–686.
Sawarkar AD, Shrimankar DD, Kumar M, Kumar P, Kumar S, Singh L (2021) Traditional system versus DNA barcoding in identification of Bamboo species: A Systematic Review. Molecular Biotechnology 63(8): 651–675.
Semagn K, Bjørnstad Å, Ndjiondjop MN (2006) An overview of molecular marker methods for plants. African Journal of Biotechnology 5(25): 2540–2568.
Setta N, Monteiro-Vitorello CB, Metcalfe CJ, Cruz GMQ, Del Bem LE, Vicentini R, Nogueira FTS, Campos RA, Nunes SL, Turrini PCG, Vieira AP, Ochoa Cruz EA, Corrêa TCS, Hotta CT, de Mello Varani A, Vautrin S, da Trindade AS, de Mendonça Vilela M, Lembke CG, Sato PM, de Andrade RF, Nishiyama MY, Cardoso-Silva CB, Scortecci KC, Garcia AAF, Carneiro MS, Kim C, Paterson AH, Bergès H, D'Hont A, de Souza AP, Souza GM, Vincentz M, Kitajima JP, Van Sluys MA (2014) Building the sugarcane genome for biotechnology and identifying evolutionary trends. BMC Genomics 15(1): 1-17.
Shabbir R, Javed T, Afzal I, El Sabagh A, Ali A, Vicente O, Chen P (2021) Modern biotechnologies: Innovative and sustainable approaches for the improvement of sugarcane tolerance to environmental stresses. Agronomy 11:1–20.
Shen Y, Guan L, Wang D, Gan X (2016) DNA barcoding and evaluation of genetic diversity in Cyprinidae fish in the midstream of the Yangtze River. Ecology and Evolution 6(9): 2702–2713.
Singh J, Banerjee S (2018) Utility of DNA barcoding tool for conservation and molecular identification of intraspecies of rice genotypes belonging to Chhattisgarh using *rbc*L and *mat*K gene sequences. Plant Archives 18: 69–75.
Solanki G (2012) Polymerase chain reaction. International Journal of Pharmacological Research 2(3): 98–102.
Sophian A, Purwaningsih R, Muindar M, Igirasa EPJ, Amirullah ML (2021) Short Communication: Analysis of purity and concentration of DNA extracted from intron patho gene-spin extraction on crab processed food product samples. Asian Journal of Tropical Biotechnology 18(1): 28–31.
Sundari, Khadijah, Jayali AM, Sukamto NH (2019) The application of barcode DNA *rbc*L gene for identification of medicinal plants: Red jabon and gofasa. In 'J. Phys. Conf. Ser.': 1–7.
Tallei TE, Kolondam BJ (2015) DNA barcoding of sangihe nutmeg (*Myristica fragrans*) using matK Gene. HAYATI Journal of Biosciences 22(1): 41–47.
Thakur VV, Nehru J, Vishwavidyalaya K, Tiwari S, Tripathi N, Tiwari G, Vishwa M, Thakur V, Tiwar G, Sapre S (2016) DNA barcoding and phylogenetic analyses of mentha species using *rbc*L sequences. Annals of Phytomedicine 5(1): 59–62.
Thorat AS, Sonone NA, Choudhari V V., Devarumath RM, Babu KH (2018) Plant regeneration from direct and indirect organogenesis and assessment of genetic fidelity in *Saccharum officinarum* using DNA-based markers. Bioscience Biotechnology Research Communications 11(1): 60–69.
Tiwari S, Vijayaraje R, Krishi S, Vidyalaya V, Sewak R, Tomar S, Bai RL, Ahuja A, Tomar RS, Tripathi MK (2017) Modified protocol for lant genomic DNA isolation. Indian Res J Genet & Biotech 9(4): 478–485.
Wang K, Zhang H, Khurshid H, Esh A, Wu C, Wang Q, Piperidis N (2023) Past and recent advances in sugarcane cytogenetics. Crop J. 11(1): 1–8.
Wei Y, Lewis R, Naseri A, Zhang S, Zhi D (2020) Genealogical search using whole-genome genotype profiles. 'Responsible Genomic Data Shar. Challenges Approaches'. United States, pp. 51–94.
Widyasari WB, Putra LK, Ranomahera MRR, Puspitasari AR (2022) Historical notes, germplasm development, and molecular approaches to support sugarcane breeding Program in Indonesia. Sugar Tech 24(1): 30–47.
Wu, Jiang Y, Wang L, Gu C, Niu J, Sun B, Li Y, Yang L (2016) Cloning, prokaryotic expression and purification of Rubisco gene from sugarcane cultivar Guitang 11. Journal of Southern Agriculture 47(4): 524–529.
Xie DF, Yu Y, Deng YQ, Li J, Liu HY, Zhou SD, He XJ (2018) Comparative analysis of the chloroplast genomes of the chinese endemic genus urophysa and their contribution to chloroplast phylogeny and adaptive evolution. International Journal of Molecular Sciences 19(7): 1–20.
Yang CF, Yang LT, Li YR, Zhang GM, Zhang CY, Wang WZ (2016) Sequence characteristics and phylogenetic implications of the nrDNA Internal Transcribed Spacers (ITS) in protospecies and landraces of sugarcane (*Saccharum officinarum* L.). Sugar Tech 18(1): 8–15.