International Journal of Agriculture and Biosciences 2024 13(2): 157-166.
https://doi.org/10.47278/journal.ijab/2024.097

RESEARCH ARTICLE
eISSN: 2306-3599; pISSN: 2305-6622

Phylogenetic and Genetic Variation of Sugarcane (*Saccharum* spp.) from Island of Java, Indonesia based on *trn*K Chloroplast Gene

Ganies Riza Aristya [1]*, Salfa Athallah Agtari Nabillah [1], Rina Sri Kasiamdari [2], Janis Damaiyani [3] and Heri Prabowo [4]

[1]Laboratory Genetics and Breeding, Faculty of Biology, Universitas Gadjah Mada, Yogyakarta, Indonesia
[2]Laoratory of Plant Systematics, Faculty of Biology, Universitas Gadjah Mada, Yogyakarta, Indonesia
[3]Research Center for Applied Botany, National Research and Innovation Agency (BRIN), Jl. Raya Jakarta - Bogor Km 46. Cibinong, Indonesia 16911
[4]Center of Standard Testing for Sweetener and Fiber Plant Instruments, Malang, Indonesia
*Corresponding author: ganies_riza@ugm.ac.id

ABSTRACT

Sugarcane (*Saccharum officinarum* L.) belongs to the Poaceae family, which is the largest family in the Poales order. Sugarcane plays a crucial role as the primary ingredient in the production of sweeteners such as sugar on a large scale. The sugarcane plants extensively cultivated in Indonesia are hybrids resulting from the crossbreeding of *Saccharum officinarum* with *Saccharum spontaneum*. Phylogenetic analysis and genetic diversity of sugarcane cultivars have been proposed to identify specific potential groups and to optimize hybridization procedures and selection to evaluate superior genotypes. This study aims to determine the genetic and phylogenetic diversity of *S. officinarum* based on the DNA barcoding molecular approach using the *trn*K gene. In this research, the *trn*K gene was amplified through PCR method on five *S. officinarum* cultivars originating from Java Island. The amplified results were then sequenced and aligned using MegaX software. Ambiguous DNA nucleotides were manually edited using GeneStudio software. Genetic variations including the number of haplotypes (h), haplotype diversity (Hd), number of polymorphic sites, and number of parsimony sites were analyzed using DnaSP software and confirmed using GenAlEx 6 software. The phylogenetic tree was reconstructed using the Maximum-Likelihood (ML) method. The sugarcane cultivars POJ, JR03, Pringu, KK, and NX04 showed trnK gene amplification with a 252bp fragment and formed a single clade in phylogenetic analysis (ML bootstrap value of 100; BI posterior probability of 1). Haplotype analysis revealed they do not belong to the same group. Indonesian sugarcane has distinct trnK genetic flow compared to those in Japan, China, South Africa, Brazil, and Australia.

Keywords: DNA Barcoding, Genetic distance, *trn*K

Article History
Article # 24-604
Received: 03-May-24
Revised: 20-May-24
Accepted: 22-May-24
Online First: 04-Jun-24

INTRODUCTION

Sugarcane (*Saccharum officinarum* L.) is one of the plants used as the primary ingredient in sugar production. Sugarcane plants are predicted to contribute about 80% to the global production of white sugar (Raj et al., 2016). The increasing sugar production due to population growth impacts makes sugarcane a vital crop in sugar production in Indonesia. However, the increase in sugarcane consumption does not match national sugar production. Over the past decade (2010–2019), sugarcane production in Indonesia experienced fluctuations and a notable overall decline, reaching its highest level in 2013. In that year, the country produced 35.5 million tons of sugarcane, but by 2019, production had decreased to 27.7 million tons (Widyasari et al., 2022). Currently, sugarcane plants extensively cultivated as an industrial crop are hybrids resulting from the crossbreeding of *Saccharum officinarum* with *Saccharum spontaneum* (Hemaprabha et al., 2022). The majority of commercial sugarcane germplasm comes from the crossbreeding of *S. officinarum* × *S. spontaneum*, with repeated backcrossing to *S. officinarum*. Current sugarcane clones contain about 90% *S. officinarum* and

Cite this Article as: Aristya GR, Nabillah SAA, Kasiamdari RS, Damaiyani J and Prabowo H, 2024. Phylogenetic and genetic variation of sugarcane (*Saccharum* spp.) from Island of Java, Indonesia based on *trn*K chloroplast gene. International Journal of Agriculture and Biosciences 13(2): 157-166. https://doi.org/10.47278/journal.ijab/2024.097

A Publication of Unique Scientific Publishers

10% *S. spontaneum* germplasm (Kaffka and Grantz, 2014). *S. officinarum* is reported to have a sugar content of 18-20 degrees Brix, which is higher compared to *S. spontaneum* (Pompidor et al., 2021).

Sugarcane is a naturally occurring plant with heterozygous traits and complex polyploidy, resulting in genetic variations. The sugarcane improvement process involves hybridization followed by clonal propagation (Hemaprabha et al., 2022). Genetic variations for sugarcane productivity within sugarcane genetic collections in subtropical regions are abundant and can be utilized to create commercial varieties with high yields and optimal sugar content (Ranjan and Kumar, 2017). Research on genetic variation based on phylogenetics in sugarcane in Indonesia aims to enhance productivity and resistance to pests. Phylogenetics is used to study sugarcane genetic diversity by reconstructing the evolutionary history of sugarcane species and their relationships with each other. This is done by analyzing genetic variations within and among different sugarcane cultivars and comparing them with genetic diversity in other related plant species (Kasiamdari et al., 2019).

Several studies have adopted approaches using barcode genes as tools to analyze genetic relatedness among organisms. A study using the molecular approach of the *trn*L-F gene has provided in-depth insights into the genetics and phylogenetics of *Alocasia longiloba* accessions in Indonesia (Asih et al., 2023). Additionally, research utilizing the *mat*K gene has been conducted to identify, characterize, and reconstruct phylogenetic trees to unveil the phylogeny of 24 sugarcane cultivars in Indonesia (Aristya et al., 2020). Analysis using the *rbc*L gene has revealed the level of genetic diversity of *Crassocephalum crepidioides* (Omonhinmin et al., 2023).

In plant chloroplasts, there is the *trn*K gene encoding *trn*A for Lysine (UUU) with two exons separated by an intron. This intron contains a unique section, *trn*KI1, which contains an open reading frame (ORF) called *mat*K (Hausner et al., 2006). Research on the phylogenetics of a species based on the *trn*K gene is still underexplored. Recent studies have suggested that the trnK intron undergoes rapid genetic changes like the *mat*K gene (Crawley and Hilu, 2012) .This argument suggests not using the *trn*K intron sequence in phylogenetic reconstruction because genes undergoing slow evolution are typically favored for phylogenetic reconstruction at higher taxonomic levels (Klopfstein et al., 2017).

Previous studies have shown that the *trn*K gene has the potential to clarify the classification of Curcuma plants. The results of study indicated that the *trn*K gene could group Curcuma plants into different clusters (Cao et al., 2001). For instance, genetic variations in the *trn*K gene of *Curcuma kwangsiensis* are phylogenetically closer to *Curcuma wenyujin* than to purple cloud varieties. Although the *trn*K intron exhibits dynamic evolutionary rates, its potential in phylogenetic analysis to explore deeper evolutionary relationships between species remains relevant.

Based on the above information, phylogenetic analysis and genetic diversity of sugarcane cultivars have been proposed to identify specific potential groups and to optimize hybridization procedures and selection to evaluate superior genotypes. One potential gene for analyzing genetic relatedness and diversity in sugarcane plants is the *trn*K gene. However, research on the phylogenetics of a species based on the *trn*K gene is still underexplored. The rapidly evolving *trn*K intron still holds potential for phylogenetic analysis to understand the evolutionary relationships among *S. officinarum* species at a deeper level. This study will further explore the use of the *trn*K gene to analyze the genetic diversity of *S. officinarum* cultivars.

MATERIALS & METHODS

Sample Collection and Species Identification

Five sugarcane cultivar samples, POJ, JR03, Kidang Kencana (KK), NX04, and Pringu (Table 1), were obtained from the collection plantation of Madubaru Factory in Bantul, Yogyakarta Indonesia and the Indonesian Center for Estate Crops Research and Development, Surabaya, East Java, Indonesia (Fig. 1; Table S1). The third to fifth leaf samples from the sugarcane plant tops were taken by cutting the base of the leaf with scissors, then divided into 3-4 pieces and stored in an icebox to maintain a low temperature. The samples were then stored at -20°C in the Genetics and Breeding Laboratory, Faculty of Biology, UGM, for further analysis. Additionally, 17 *trn*K sequences from various collections from countries including South Africa, Australia, Brazil, China, and Japan (Table S1) were also included to enhance understanding of genetic variation and relatedness among sugarcane cultivars worldwide.

DNA Extraction and Quantitative-Qualitative Analysis

DNA isolation was conducted using the Geneaid DNA Kit. Genomic DNA was extracted from 100g of sugarcane leaves that had been separated from their leaf veins. The extraction results were then tested quantitatively and qualitatively. Quantitative testing was performed using Nanodrop to determine the concentration of genomic DNA and DNA purity (A260/280) was measured. A DNA purity ratio is considered good if this falls within the range of 1.8 to 2.0 (Aristya et al., 2020). Qualitative analysis of genomic DNA was conducted using electrophoresis techniques to ensure that the isolated DNA has appropriate size and band patterns. The electrophoresis

Table 1: Sample collection of Tebu (*Saccharum* spp.)

Cultivars	Sample code	Collections time	Collection time	Latitude	Longitude
Kidang Kencana	KK	November 2023	Madubaru Factory	7° 49' 51.0456'' S	110° 20' 44.5128'' E
NX04	NX04				
POJ	POJ	November 2023	The Indonesian Center for Estate Crops	7° 33' 10.1196'' S	112° 20' 3.3792'' E
Pringu	Pringu	December 2017	Research and Development, Surabaya, East		
JR03	JR03		Java		

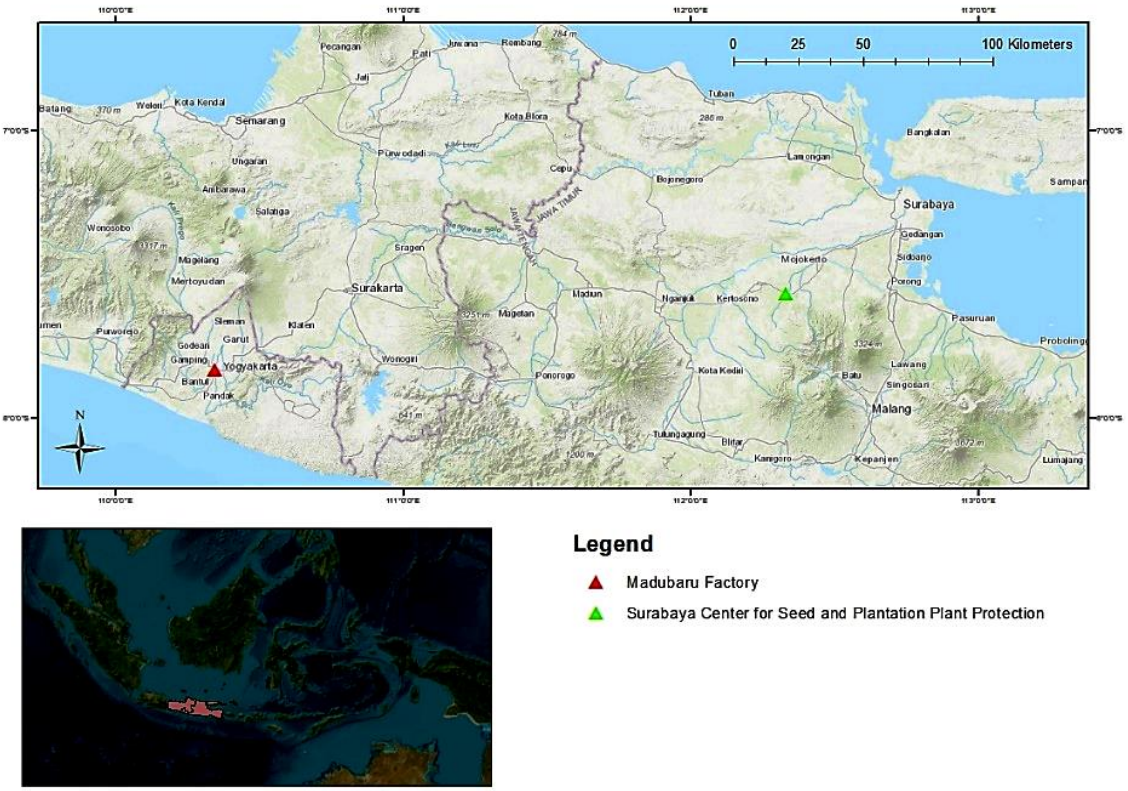

Fig. 1: Sample location sites.

process utilized 1% agarose gel and was run at a voltage of 100V for 45min using 1x TBE buffer. Observation of the electrophoresis results was conducted using Gel Doc and an optilab connected to a computer system.

DNA Amplification, Electrophoresis and Sequencing of *trn*K

The amplification of the *trn*K gene was performed through a PCR process using the T100 Thermal-Cycler machine (Biorad), utilizing two primers, Forward (5'-ACCTTTCCCTGCATTAGGCA-3') and Reverse (5'-AGACCACGACTGATCCTCAA-3') (de Cesare et al., 2010). Each reaction had a total volume of 30µL, consisting of 12.5µL of PCR Bioline premix kit, 2µL each of forward and reverse primers with a concentration of 10µM, and 8.5µL of sterile water. The PCR mixture, primers, and water were homogenized and dispensed into microtubes with 25µL per tube. Subsequently, 5µL of DNA template was added to each tube and placed into the PCR machine.

The PCR protocol involved a pre-denaturation step at 95°C for 5min, followed by 35 cycles of denaturation at 95°C for 30s, annealing at 54°C for 45s, and extension at 72°C for 45s, followed by a final extension step at 72°C for 5min. The PCR products were then visualized using electrophoresis technique on a 1% agarose gel run at 100V for 30min using 1x TBE buffer. The GelDoc system was used to document the gel results under UV light. The size of the *trnK* gene DNA was estimated using a 100bp DNA Ladder. Subsequently, the samples were sent to the Unit III UGM Integrated Research and Knowledge Laboratory for Sanger sequencing. The process began with a purification step using ExoSAP-IT, followed by sequencing cycles consisting of initial denaturation at 96°C for 1min, denaturation at 96°C for 10s, annealing at 50°C for 5s for 24-27 cycles, and extension at 60°C for 4min. The products were then purified again using the Bigdye X-Terminator Purification Kit (BDX-T) and analyzed using the ABI 3500 Genetic Analyzer (Applied Biosystems).

Data Analysis

trnK Sequence Editing and Alignment

Editing and alignment of *trn*K sequences were performed using MEGA11 software, while ambiguous DNA nucleotides were manually edited using GeneStudio software. To verify the sugarcane species, the consensus sequencing results were compared with data available on GenBank using the BLAST program (https://blast.ncbi.nlm.nih.gov/Blast.cgi). In this study, we aligned a total of 22 sugarcane *trn*K sequences, resulting in a fragment length of 252bp, using MEGA11 software with the "align by MUSCLE" menu. These sequences may be used in further intraspecific analysis. For phylogenetic analysis, we added two samples as outgroups during the alignment process without affecting the previous fragment length of 252bp.

Phylogenetic and Genetic Distance Analysis

The reconstruction of the phylogenetic tree involves analyzing DNA sequences using the best scheme analysis using PartitionFinder2 software based on Python coding (Lanfear et al., 2016). Next, the alignment results were processed using Nexus program to be run on Maximum-

Likelihood (ML) and Bayesian Inference (BI) programs. The phylogenetic tree was constructed using ML analysis using IQ-Tree v1.6.10 application with 1000 bootstrap replicates (Nguyen et al., 2015). BI analysis was conducted using MrBayes v3.2.6 software (Ronquist et al., 2012) statistically for each node of the phylogenetic tree based on Markov Chain Monte Carlo (MCMC) and repeated for 4,000,000 generations. In this study, we added one samples from GenBank as outgroups, namely *Phaenosperma globose* (GU254935), making a total of 22 sequences. The consensus tree was visualized using FigTree 1.4.4.32.

Genetic Variation and Principal Coordinate of Analysis (PCoA)

The analysis of intraspecific genetic variation involves the number of haplotypes (h), haplotype diversity (Hd), number of polymorphic sites, and number of parsimony sites. The analysis was conducted using DnaSP software, and the results were confirmed using GenAlEx 6.34 software. Additionally, PCoA analysis was also performed using GenAlEx 6.34.

RESULTS & DISCUSSION

A commonly used method for establishing genetic relationships is through DNA Barcoding, which enables the identification of various species based on specific genetic sequences. This research aims to explore the characteristics of *trn*K as a DNA barcode to obtain a deeper understanding of phylogeny and evolution in five sugarcane cultivars.

Quantitative Test Results

Genomic DNA was isolated using Geneaid reagent kits and then quantitatively and qualitatively tested to ensure the presence of sugarcane DNA and assess its purity level. DNA concentration is considered adequate if it falls within the range of more than 25ng/µL and not more than 100ng/µL (Cao et al., 2001). A DNA purity level of 1.8 indicates the presence of pure DNA, while lower values indicate protein contamination. If the value exceeds 2.0, it indicates RNA contamination (Fialova et al., 2020). DNA isolation was performed on five sugarcane cultivars, consisting of 3 samples cultivated by PT Madubaru and 2 samples from the Genetics and Breeding Laboratory collection. All five samples were successfully isolated and showed concentrations ranging from 7.01 to 180.97ng/µL (Table 2). Genomic concentration data show that the Kidang Kencana cultivar has the highest concentration at 180.97ng/µL, while the JR03 cultivar has the lowest concentration value at 7.01ng/µL. The purity of each sample ranged from 1.73 to 1.81 (A260/A280). Variations in DNA concentration and purity among samples may be caused by high levels of polysaccharides, polyphenols, or other secondary metabolites (Nawfetrias and Roswanjaya, 2021). These contaminant compounds can interfere with the DNA extraction and purification processes, resulting in low DNA yields (Friar, 2005).

Qualitative Test Results

Samples that have been quantitatively and qualitatively tested are then subjected to DNA amplification using the Polymerase Chain Reaction (PCR) method. PCR is a technique used to amplify specific DNA segments for various clinical and laboratory applications. PCR exploits the complementary base-pairing properties of DNA, the double-stranded structure of DNA, and the melting temperature of DNA molecules (Putra et al., 2020). The PCR process involves three reaction stages that depend on temperature, including denaturation, annealing, and extension. The denaturation stage begins by heating the reaction to 95°C, aiming to break the hydrogen bonds between the two template DNA strands to separate them. The reaction temperature is then lowered to 58°C, allowing the annealing of complementary base pairs to occur. Annealing primers at too high temperature can result in poor binding and low DNA production, while at too low temperature, it can prevent specific primer binding, inhibiting the desired DNA amplification (Aristya et al., 2020). Finally, the temperature is raised again to 72°C for 2 to 5min for the extension process. These steps are repeated as desired to achieve significant amplification results. The purpose of implementing PCR techniques in this study is to identify the presence of the *trn*K gene in the genomes of five sugarcane cultivars under investigation. The amplification results were qualitatively tested using electrophoresis on 1% agarose gel, run at 100V for 30min. The trnK gene is considered successfully amplified if the Gel Doc visualization shows DNA bands at 250-253bp (de Cesare et al., 2010) without any DNA smearing. DNA smearing indicates contamination by polysaccharides, phenol, RNA, protein, or improper annealing temperature for the target gene (Mainkar et al., 2023).

Table 2: Sample concentration result

Samples	Concentration (ng/µL)	Purification (A_{260}/A_{280})
Kidang kencana	180.97	1.806
NX04	99.48	1.809
POJ	106.3	1.791
Pringu	26.01	1.73
JR03	7.01	1.81

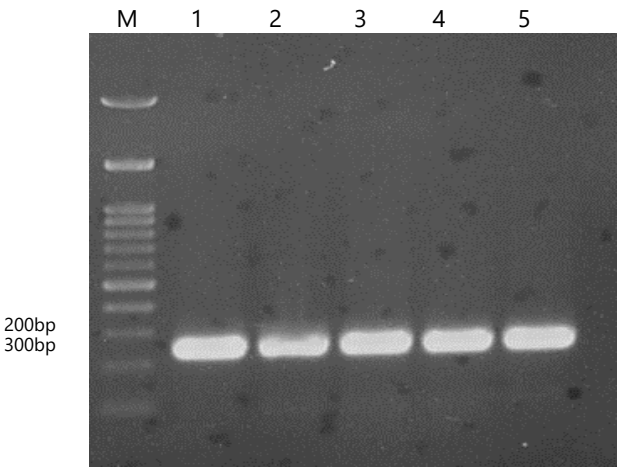

Fig. 2: Genome electropherogram of five sugarcane cultivars. Note: M: Marker 100 Bp, 1. Kidang Kencana; 2. NX04; 3. POJ; 4. Pringu; 5. JR03.

DNA extraction and amplification were performed using two primers, namely *trn*K-F and *trn*K-R primers. Visualization of PCR products was conducted using a UV Transilluminator. The PCR electropherogram indicated that all five samples from the five sugarcane cultivars were successfully amplified, as evidenced by the appearance of the target bands in Fig. 2. The target DNA length from the *trn*K-F and *trn*K-R primers is 236-257bp (de Cesare et al., 2010).

Sequencing Results

The five sugarcane varieties were sequenced at LPPT using the Applied Biosystem 3500 Genetic Analyzer 2550 sequencing machine. The sequencing results were analyzed using GeneStudio software to determine the produced products and the presence or absence of ambiguity in the DNA sequence. To verify the sugarcane species, the consensus sequencing results were compared with data in GenBank. DNA data analysis was conducted using PartitionFinder2 software based on Python coding to determine the best scheme (Lanfear et al., 2016). Subsequently, the phylogenetic tree was constructed using the Maximum-Likelihood (ML) and Bayesian Inference (BI) methods. ML analysis was performed using IQ-Tree v1.6.10 software with 1000 bootstrap replicates (Nguyen et al., 2015) while BI analysis was conducted using MrBayes v3.2.6 software (Ronquist et al., 2012) with the Markov Chain Monte Carlo (MCMC) method for 4,000,000 generations. Genetic variations, including the number of haplotypes (h), haplotype diversity (Hd), number of polymorphic sites, and number of parsimony sites, were analyzed using DnaSP software (Rozas et al., 2017) and confirmed using GenAlEx 6 software (Peakall and Smouse, 2006).

The sequencing results showed that the length of the *trn*K fragment ranged from 240 to 260bp. Analysis using BLAST Nucleotide revealed that the sequences had a similarity of 93.88% with *S. spontaneum*, the Badila cultivar of *S. officinarum*, and various hybrid cultivars of *S. hybrid*. This percentage of similarity after BLAST is below the 98 to 100% range, which is the criterion for categorizing them as the same species (Wu et al., 2021). This result indicates that the samples fall below this range. This is due to the lack of exploration in *trn*K gene-based phylogenetic research on specific species, resulting in insufficient data available in GenBank.

Tree Reconstruction Results

Fig. 3 shows the reconstruction of the phylogenetic tree using the Maximum-Likelihood method with the Kimura-2-Parameter model and 1000 bootstrap replicates (Passari et al., 2017). Maximum likelihood method analysis using the Kimura-2 model builds relationships between species based on branch lengths. Branch lengths on the phylogenetic tree provide important information in evolutionary biology, used for molecular dating, reconciling gene trees with species trees, and measuring biodiversity (Paradis, 2018). Branch length reflects the evolutionary rate of each species, with longer branches indicating greater evolutionary distance, while shorter branches indicate closer distance (Scott and Baum, 2016). Distance-based methods naturally estimate branch lengths and gene rates, as well as identify different selective pressures on various parts of the genome (Binet et al., 2016).

The Maximum Likelihood (ML) method tree assessment is not based on its total length, but rather on how well the joint probabilities predict the data. The tree with highest joint probability (maximum likelihood) of the distribution of the distribution was selected as the most suitable for construction (Nixon, 2001). The reconstruction of the phylogenetic tree shows that the *Saccharum* samples used are grouped in the same clade with a bootstrap value of 100, while the reference samples from NCBI have a bootstrap value ranging from 31 to 70 (Fig. 3). A bootstrap value of 100 is considered good because the criteria for a good bootstrap value are between 80 to 100. The selected reference samples from NCBI have a similarity level of 93.88%, and the tree reconstruction shows that they form two clades, namely a clade with 14 cultivar samples with a bootstrap value ranging from 31 to 71, and another clade with NCBI data with a bootstrap value of 40 and 71.

The Markov Chain Monte Carlo (MCMC) method has been applied in reconstructing the phylogenetic tree for 4 million generations, with sampling every 1,000 generations (Nascimento et al., 2017). This was done to estimate the posterior probability distribution in Fig. 4. The results of the phylogenetic tree reconstruction show that the five *Saccharum* cultivar samples used form the same clade, with a posterior probabilities value of 1, indicating good quality with posterior probabilities values exceeding 0.98.

Genetic distance refers to the proportion of genetic differences between species or populations (Basith and Kusrini, 2021). Using the Kimura 2P model (Table 3), the genetic distance among Saccharum species was assessed. The five sample cultivars showed perfect similarity (*** 0.000), indicating no genetic distance and identifying them as the same species.

Polymorphic Sites

Multiple alignment results of 250 base pairs control region sequences, obtained through partial sequencing, revealed polymorphic sites consisting of 7 singleton variable variants and 32 parsimony informative sites. The total number of mutations is 39 sites, with 6 haplotypes identified (Table 4).

Haplotype and PCoA Results

The haplotype results in Table 5 divide the 22 cultivar samples into nine groups. For instance, cultivar JR03 falls into group H1; Pringu is classified in group H2; POJ is categorized under group H3; NX04 is in group H4, and KK belongs to group H5. These group differences stem from different combinations of alleles located on homologous chromosomes (Jin et al., 2023). Each cultivar exhibits distinct alleles from one another (Table 3).

Fig. 5 illustrates the median haplotype network formed from five sugarcane samples and references from NCBI based on the *trn*K gene sequence (252bp). Haplotype networks offer an intuitive way to visualize the

Table 3: Genetic distances of five sugarcane cultivars and reference from NCBI

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1																					
2	8,466																				
3	5,556	6,170																			
4	1,637	6,573	4,649																		
5	3,345	6,545	4,741	1,661																	
6	6,876	14,051	10,894	6,009	6,973																
7	6,876	14,051	10,894	6,009	6,973	***															
8	6,876	14,051	10,894	6,009	6,973	***	***														
9	6,876	14,051	10,894	6,009	6,973	***	***	***													
10	6,876	14,051	10,894	6,009	6,973	***	***	***	***												
11	6,876	14,051	10,894	6,009	6,973	***	***	***	***	***											
12	6,876	14,051	10,894	6,009	6,973	***	***	***	***	***	***										
13	6,876	14,051	10,894	6,009	6,973	***	***	***	***	***	***	***									
14	6,876	14,051	10,894	6,009	6,973	***	***	***	***	***	***	***	***								
15	6,905	14,117	10,943	6,035	7,003	***	***	***	***	***	***	***	***	***							
16	6,905	14,117	10,943	6,035	7,003	***	***	***	***	***	***	***	***	***	***						
17	6,905	14,117	10,943	6,035	7,003	***	***	***	***	***	***	***	***	***	***	***					
18	6,905	14,117	10,943	6,035	7,003	***	***	***	***	***	***	***	***	***	***	***	***				
19	6,905	14,117	10,943	6,035	7,003	***	***	***	***	***	***	***	***	***	***	***	***	***			
20	7,358	14,632	11,433	6,485	7,462	0,406	0,406	0,406	0,406	0,406	0,406	0,406	0,406	0,406	0,406	0,406	0,406	0,406	0,406		
21	6,935	14,183	10,993	6,061	7,034	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	
22	7,358	14,632	11,433	6,485	7,462	0,406	0,406	0,406	0,406	0,406	0,406	0,406	0,406	0,406	0,406	0,406	0,406	0,406	0,406	***	***

Note: 1. POJ; 2. JR03; 3. Pringu; 4. KK; 5. NX04; 6. LS975136; 7. MG685915; 8. LS975132; 9. AB732015; 10. LN896359; 11. KX507 245; 12. NC_029221; 13. OP235371; 14. MZ328080; 15. AB732018; 16. NC_035224; 17. NC_072525; 18. OP896210; 19. NC_070069; 20. AB732016; 21. OP235342; 22. OP235341

Fig. 3: Phylogenetic tree reconstruction of 22 *trnK* samples from members of *Saccharum* spp. and one *Phaenosperma globosa* outgroup sequences using Maximum Likelihood analysis with IQTree 2 program.

relationships between cultivars, providing valuable insights into migration patterns, population structure, and speciation events (Garcia et al., 2021). The circle size reflects the sample count. Lines connecting haplotypes indicate evolutionary pathways among haplotypes, while short lines mark mutation points between haplotypes (Cosme et al., 2020). Each color represents the sample's origin in this study.

Table 4: Polymorphic sites of 22 sugarcane samples based on the *trn*K gene

Sample	3	9	1	1	1	1	2	2	2	2	2	3	3	4	4	5	6	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
			0	2	5	9	0	1	4	5	7	5	8	2	8	3	2	7	7	8	9	9	0	0	0	1	1	2	2	2	2	3	3	3	3	4	4	4
																		2	3	7	4	9	4	7	8	0	2	1	4	5	8	9	0	3	5	7	1	3	6
POJ	T	C	G	A	T	C	C	T	T	T	T	A	T	T	T	A	A	A	A	A	A	T	A	T	T	C	T	T	C	T	A	C	C	T	A	G	A	T	G
JR03	C	C	C	C	A	C	C	A	A	T	A	A	T	A	G	G	-	A	A	A	T	T	A	T	G	T	T	T	T	T	-	T	C	T	G	A	G	G	G
Pringu	C	C	G	C	T	G	-	-	A	T	A	A	T	A	-	G	A	A	A	A	A	T	A	C	G	C	T	T	C	T	C	C	T	G	A	G	A	C	G
KK	T	C	G	C	A	C	C	T	T	T	A	A	T	T	C	-	A	A	A	A	A	T	A	T	T	C	T	T	C	T	A	C	C	T	A	G	A	T	G
NX04	T	C	G	C	A	C	C	T	T	T	A	A	T	T	C	A	-	A	A	A	A	T	A	A	G	C	T	T	C	T	C	C	C	T	G	G	A	T	G
LS975136	T	T	G	A	A	A	C	T	T	C	A	C	C	T	C	A	A	A	A	G	A	A	G	A	T	C	A	A	C	A	A	C	C	T	A	G	A	T	T
MG685915	T	T	G	A	A	A	C	T	T	C	A	C	C	T	C	A	A	A	A	G	A	A	G	A	T	C	A	A	C	A	A	C	C	T	A	G	A	T	T
LS975132	T	T	G	A	A	A	C	T	T	C	A	C	C	T	C	A	A	A	A	G	A	A	G	A	T	C	A	A	C	A	A	C	C	T	A	G	A	T	T
AB732015	T	T	G	A	A	A	C	T	T	C	A	C	C	T	C	A	A	A	A	G	A	A	G	A	T	C	A	A	C	A	A	C	C	T	A	G	A	T	T
LN896359	T	T	G	A	A	A	C	T	T	C	A	C	C	T	C	A	A	A	A	G	A	A	G	A	T	C	A	A	C	A	A	C	C	T	A	G	A	T	T
KX507245	T	T	G	A	A	A	C	T	T	C	A	C	C	T	C	A	A	A	A	G	A	A	G	A	T	C	A	A	C	A	A	C	C	T	A	G	A	T	T
NC_029221	T	T	G	A	A	A	C	T	T	C	A	C	C	T	C	A	A	A	A	G	A	A	G	A	T	C	A	A	C	A	A	C	C	T	A	G	A	T	T
OP235371	T	T	G	A	A	A	C	T	T	C	A	C	C	T	C	A	A	A	A	G	A	A	G	A	T	C	A	A	C	A	A	C	C	T	A	G	A	T	T
MZ328080	T	T	G	A	A	A	C	T	T	C	A	C	C	T	C	A	A	A	A	G	A	A	G	A	T	C	A	A	C	A	A	C	C	T	A	G	A	T	T
AB732018	T	T	G	A	A	A	C	T	T	C	A	C	C	T	C	A	A	-	A	G	A	A	G	A	T	C	A	A	C	A	A	C	C	T	A	G	A	T	T
NC_035224	T	T	G	A	A	A	C	T	T	C	A	C	C	T	C	A	A	-	A	G	A	A	G	A	T	C	A	A	C	A	A	C	C	T	A	G	A	T	T
NC_072525	T	T	G	A	A	A	C	T	T	C	A	C	C	T	C	A	A	-	A	G	A	A	G	A	T	C	A	A	C	A	A	C	C	T	A	G	A	T	T
OP896210	T	T	G	A	A	A	C	T	T	C	A	C	C	T	C	A	A	-	A	G	A	A	G	A	T	C	A	A	C	A	A	C	C	T	A	G	A	T	T
NC_070069	T	T	G	A	A	A	C	T	T	C	A	C	C	T	C	A	A	-	A	G	A	A	G	A	T	C	A	A	C	A	A	C	C	T	A	G	A	T	T
AB732016	T	T	G	A	A	A	C	T	T	C	A	C	C	T	C	A	A	-	C	G	A	A	G	A	T	C	A	A	C	A	A	C	C	T	A	G	A	T	T
OP235342	T	T	G	A	A	A	C	T	T	C	A	C	C	T	C	A	A	-	-	G	A	A	G	A	T	C	A	A	C	A	A	C	C	T	A	G	A	T	T
OP235341	T	T	G	A	A	A	C	T	T	C	A	C	C	T	C	A	A	-	C	G	A	A	G	A	T	C	A	A	C	A	A	C	C	T	A	G	A	T	T

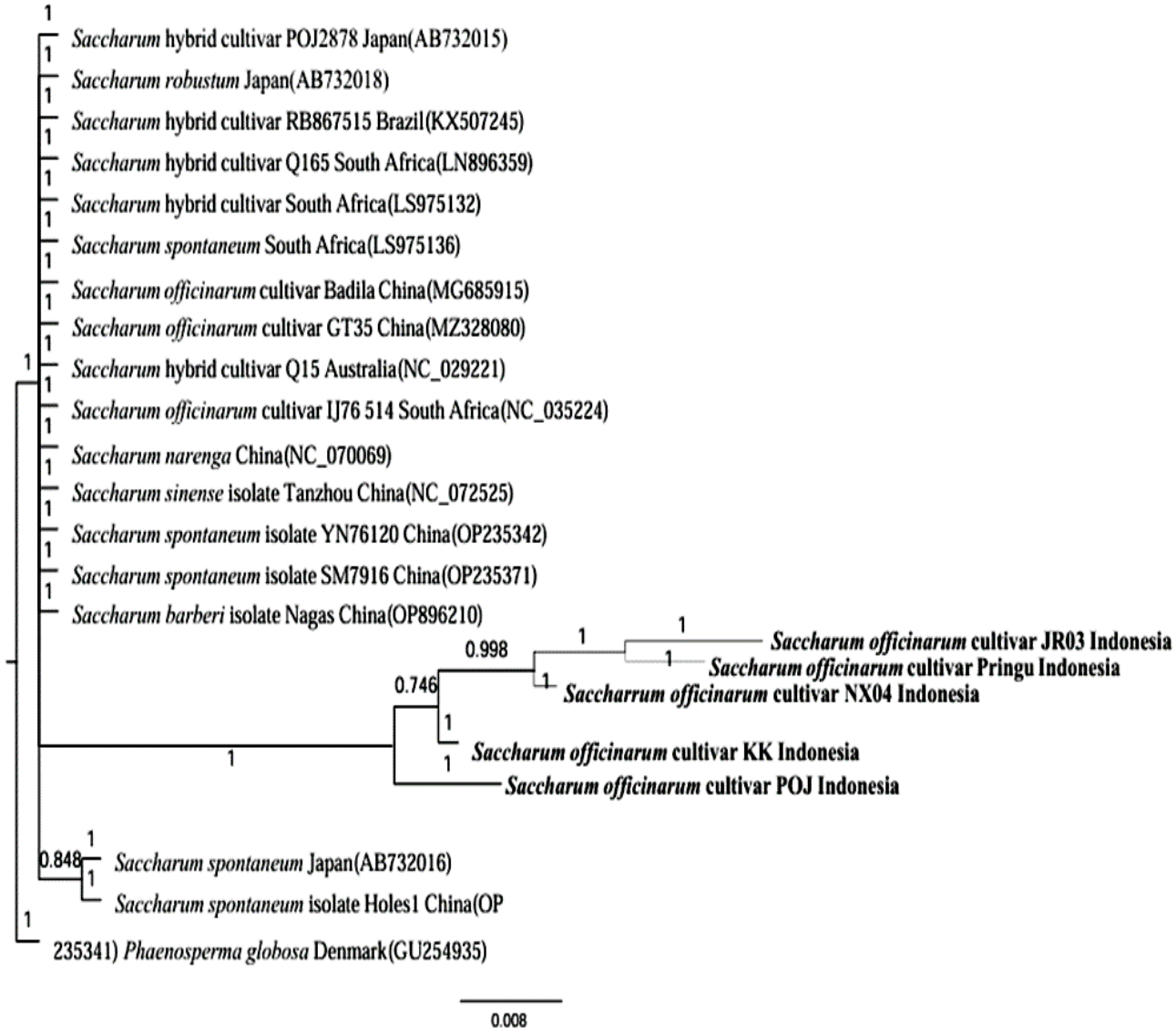

Fig. 4: Reconstruction of the phylogenetic tree of 22 *trn*K samples from *Saccharum* spp. members and one *Phaenosperma globosa* outgroup using MrBayes program.

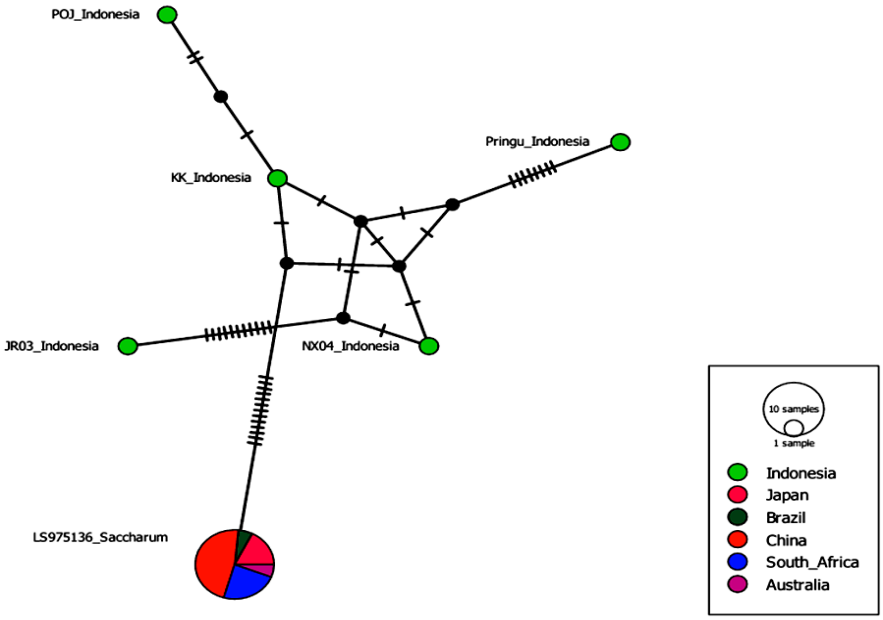

Fig. 5: Median-joining haplotype network of sugarcane based on *trnK* chloroplast gene sequences (252bp).

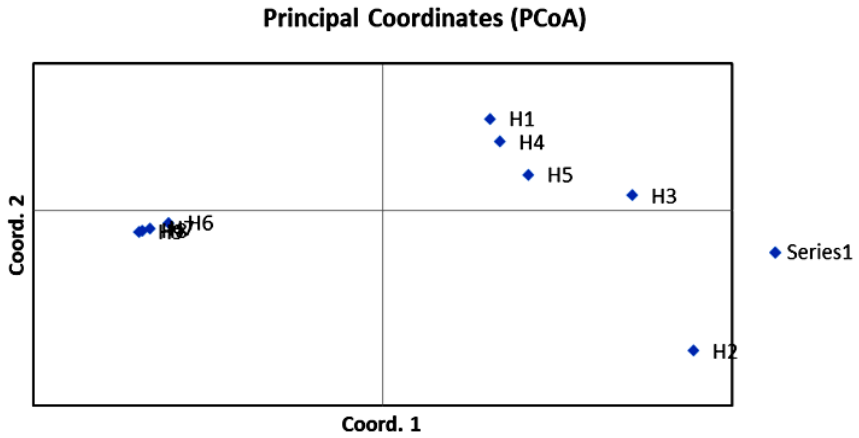

Fig. 6: Principal Coordinate Analysis (PCoA) of five sugarcane cultivars based on the *trn*K gene (252bp).

Table 5: Haplotype of 22 sugarcane samples based on the *trnK* gene

Haplotype	Jumlah Samples	Code
H1	1	JR03
H2	1	Pringu
H3	1	POJ
H4	1	NX04
H5	1	KK
H6	9	LS975136
		MG685915
		LS975132
		AB732015
		LN896359
		KX507245
		NC_029221
		OP235371
		MZ328080
H7	5	AB732018
		NC_035224
		NC_072525
		OP896210
		NC_070069
H8	2	AB732016
		OP235341
H9	1	OP235342

The distance between samples in auxiliary space directly reflects the relationship between them. PCoA is a multivariate approach used to group genotypes based on similarity coefficients, offering more significant insights into major groups compared to cluster analysis (Sharma et al., 2022). The PCoA grouped 22 sugarcane genotypes into four different populations. Samples with similar characteristics will be located close to each other, grouped in the same population, and can be easily classified together (Kurdi et al., 2020). Based on the *trn*k gene analysis, it was found that the evolution of *S. officinarum* did not originate from Indonesia (Fig. 6). Samples from Indonesia exhibit higher polymorphism compared to samples from abroad, which could be attributed to cultivar breeding. Selective crossing aims to combine desired traits from both parents through genetic changes and base sequence variations. This is part of the selective breeding process, where parents with specific characteristics are chosen for crossing to produce offspring with desired traits (D'Ambrosio et al., 2019).

Conclusion

The sugarcane cultivars used in this study, including POJ, JR03, Pringu, KK, and NX04, showed *trn*K gene amplification with a fragment length of 252bp after alignment. Phylogenetic analysis using the Maximum-Likelihood (ML) method indicates that these five cultivars form one clade with a bootstrap value of 100. Additionally,

analysis with Bayesian Inference (BI) also shows the formation of one clade with a posterior probability (pp) value of 1. Haplotype analysis indicates that the five cultivars do not belong to the same haplotype group. Through haplotype network analysis and principal coordinate analysis (PCoA), it was found that sugarcane plants in Indonesia do not share a similar *trn*K genetic flow as those found in Japan, China, South Africa, Brazil, and Australia (Table S1-3).

Acknowledgements

This research is funded by the National Research and Innovation Agency through a Research and Innovation Grant for Advanced Indonesia No: 30/IV/KS/05/2023 and 2090/UN1/DITLIT/Dit- Lit/PT.01.03/2023. 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.

Author's Contribution

GRA developed the theory, devised the project, the main conceptual ideas, and supervised the findings of this work, revised manuscript, and proof outline. SAAN collected data, performed molecular experiments, developed phylogenetic tree analyses, and wrote the manuscript. RSK verified the analytical methods, carried out the experiment and revised the manuscript. JD helped supervise the project, conceived the original idea, and analyzed the data. HP performed the analytic calculations, developed the numerical and designed the model framework. All authors discussed the results and contributed to the final manuscript.

REFERENCES

Aristya, G.R., Putri, F., and Kasiamdari, R.S. (2020). DNA barcoding and phylogenetic analysis sugarcane (*Saccharum officinarum* L.) based on *mat*K (*maturase* K) gene. *Key Engineering Materials*, 840 KEM, 162–170.

Asih, N.P.S., Purnomo, and Daryono, B.S. (2023). Genetic diversity and phylogeny of *Alocasia longiloba* Miq. Indonesia accessions Based on *trnL-F* Intergenic Spacer Region. *Jurnal Sains dan Teknologi,* 12(2), 485-493.

Basith, A., and Kusrini, E. N. I. (2021). Genetic diversity analysis and phylogenetic reconstruction of groupers *Epinephelus* spp. from Madura Island, Indonesia based on partial sequence of CO1 gene. *Biodiversitas* 22(10), 4282–4290.

Binet, M., Gascuel, O., Scomavacca, C., Emmanuel, E. J., and Pardi, F. (2016). Fast and accurate branch lengths estimation for phylogenomic trees. *BMC Bioinformatics*, 17(1), 1–18.

Cao, H., Sasaki, Y., Fyshimi, H., and Komatsu, K. (2001). Molecular analysis of medicinally-used Chinese and Japanese Curcuma based on *18*S rRNA gene and *trnK* gene sequences. *Biological and Pharmaceutical Bulletin*, 24(12), 1389–1394.

Cosme, L.V., GLoria-Soria, A.G., Caccone, A., Powell, J.R., and Martin, D.J. (2020). Evolution of kdr haplotypes in worldwide populations of *Aedes aegypti*: Independent origins of the F1534C kdr mutation. *PLOS,* 14(4),1–18.

Crawley, S.S. and Hilu, K.W. (2012). Caryophyllales: Evaluating phylogenetic signal in *trnK* intron versus *mat*K. *Journal of Systematics and Evolution*, 50(5), 387–410.

D'Ambrosio, J., Phocas, F., Haffray, P., Bestin, A., Brard-Fudulea, S., Poncet, C., Quillet, E., Dechamp, N., Fraslin, C., Charles, M., and Dupont-Nivet, M. (2019). Genome-wide estimates of genetic diversity, inbreeding and effective size of experimental and commercial rainbow trout lines undergoing selective breeding. *Genetics Selection Evolution*, 51(1), 1–15.

de Cesare, M., Hodkinson, T.R., and Barth, S. (2010). Chloroplast DNA markers (cpSSRs, SNPs) for *Miscanthus*, *Saccharum* and related grasses (Panicoideae, Poaceae). *Molecular Breeding*, 26(3), 539–544.

Fialova, L., Romanovska, D., and Marova, I. (2020). A comparative study of some procedures for isolation of fruit DNA of sufficient quality for PCR-based assays. *Molecules*, 25(18), 1-15.

Friar, E. A. (2005). Isolation of DNA from plants with large amounts of secondary metabolites. *Methods in Enzymology*, 2005(395), 3–14.

Garcia, E., Wright, D., Gatins, R., Roberts, M.B., Pinheiro, H.T., Salas, E., Chen, J., Winnikoff, J.R., and Bernardi, G. (2021). Haplotype network branch diversity, a new metric combining genetic and topological diversity to compare the complexity of haplotype networks. *PLoS ONE*, 16(6), 1–15.

Hausner, G., Olson, R., Simon, D., Johnson, I., Sanders, E.R., Karol, K.G., McCourt, R.M., and Zimmerly, S. (2006). Origin and evolution of the chloroplast *trnK* (*mat*K) intron: a model for evolution of group II intron RNA structures. *Molecular Biology and Evolution*, 23(2), 380–391.

Hemaprabha, G., Mohanraj, K., Jackson, P.A., Lakshmanan, P., Ali, G.S., Li, A.M., Huang, D.L., and Ram, B. (2022). Sugarcane genetic diversity and major germplasm collections. *Sugar Tech*, 24(1), 279–297.

Jin, X., Du, H., Zhu, C., Wan, H., Liu, F., Ruan, J., Mower, J.P., and 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.

Kaffka, S.R. and Grantz, D.A. (2014). Sugar Crops. *Encyclopedia of Agriculture and Food Systems*, 5, 240–260.

Kasiamdari, R.S., Aristya, G.R., Prabowo, H., Ariesta, C., and Arif, M.F. (2019). Genetic variation and characterization of the sucrose synthase 2 gene (Sus2) in sugarcane based molecular markers. *Biodiversitas*, 20(11), 3087–3096.

Klopfstein, S., Massingham, T., and Goldman, N. (2017). More on the best evolutionary rate for phylogenetic analysis. *Systematic Biology*, 66(5), 769–785.

Kurdi, E.R., Kumar, K., and Patra, D. (2020). Introducing principal coordinate analysis (PCoA) assisted EEMF spectroscopic based novel analytical approach for the

discrimination of commercial gasoline fuels. *Journal of Fluorescence*, 30(6), 1583–1589.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., and Calcott, B. (2016). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, 34(3), 772–773.

Mainkar, P., Jayaswal, D., Kumar, D., Jayaswall, K., Jaiswal, S., Singh, A.N., Kumar, S., and Kansal, R. (2023). Development of modified CTAB and Trizol protocols to isolate high molecular weight (HMW) RNA from polyphenol and polysaccharides rich pigeonpea (*Cajanuscajan* L. Millsp). *PloS ONE*, 18(12), 1–8.

Nascimento, F.F., Reis, M.D., and Yang, Z. (2017). A biologist's guide to Bayesian phylogenetic analysis. *Nature Ecology and Evolution*, 1(10), 1446–1454.

Nawfetrias, W., and Roswanjaya, Y.P. (2021) 'Variations on The Concentration and Purity of Oil Palm Flowers (Elais Guineensis Jac .) in Some Extraction Methods for Identification of the Flowering Gene', *Agrotech Journal*, 6(2), pp. 69–77. Available at: https://doi.org/10.31327/atj.v6i2.1658

Nguyen, L.T., Schmidt, H.A., Haeseler, A.V., and Minh B.Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1), 268–274.

Nixon, K.C. (2001). Phylogeny. *Encyclopedia of Biodiversity: Second Edition*, 6(4), 16–23.

Omonhinmin, C.A., Olomukoro, E.E., Onuselogu, C.C., Popoola J.O., and Oyejide, S.O. (2023). rbcL gene dataset on intra-specific genetic variability and phylogenetic relationship of *Crassocephalum crepidioides* (Benth) S. Moore. (Asteraceae) in Nigeria. *Data in Brief*, 48(2023), 1-10.

Paradis, E. (2018). Analysis of haplotype networks: The randomized minimum spanning tree method. *Methods in Ecology and Evolution*, 9(5), 1308–1317.

Passari, A.K., Mishra, V.K., Singh, G., Singh, P., Kumar, B., Gupta, V.K., Sarma, R.K., Saikia, R., Donovan, A.O., and Singh, B.P. (2017). Insights into the functionality of endophytic actinobacteria with a focus on their biosynthetic potential and secondary metabolites production. *Nature,* 8(1), 1–18.

Peakall, R. and Smouse, P.E. (2006). GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, 6(1), 288–295.

Pompidor, N., Charron, C., Hervouet, C., Bocs, S., Droc, G., Rivallan, R., Manez, A., Mitros, T., Swaminathan, K., Glaszmann, J., Garsmeur, O., and D'Hont, A. (2021). Three founding ancestral genomes involved in the origin of sugarcane. *Annals of Botany*, 127(6), 827–840.

Putra, R.P., Ranomahera, M.R.R., Rizaludin, M.S., Supriyanto, R., and Dewi, V.A.K. (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, P.T., and Nair, N.V. (2016). Analysis of genetic diversity of *Saccharum* complex using chloroplast microsatellite markers. *Sugar Tech*, 18(2), 141–148.

Ranjan, R., and Kumar, B. (2017). Study of genetic variability for cane yield and its component traits in early maturing sugarcane. *International Journal of Current Microbiology and Applied Sciences*, 6(10), 1739–1748.

Ronquist, F., Teslenko, M., Mark, P.V., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M.A., and Huelsenbeck, J.P. (2012). Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61(3), 539–542.

Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E., and Sanchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, 34(12), 3299–3302.

Scott, A.D., and Baum, D.A. (2016). Phylogenetic Tree. *Encyclopedia of Evolutionary Biology*, 3, 270–276.

Sharma, A., Sharma, S., Kumar, N., Rana, R.S., Sharma, P., Kumar, P., and Rani, M. (2022). Morpho-molecular genetic diversity and population structure analysis in garden pea (*Pisum sativum* L.) genotypes using simple sequence repeat markers. *PLoS ONE*, 17(9), 1–19. doi: 10.1371/journal.pone.0273499

Widyasari, W.B., Putra, L.K., Ranomahera, M.R.R., and Puspitasari, A.R. (2022). Historical notes, germplasm development and molecular approaches to support sugarcane breeding program in Indonesia. *Sugar Tech*, 24(1), 30–47.

Wu, H.Y., Chan, K.T., But, G.W.C., and Shaw, P.C. (2021). Assessing the reliability of medicinal *Dendrobium* sequences in GenBank for botanical species identification. *Scientific Reports*, 11(1), 1–9.