ARTICLE

Comparative sequence analysis of candidate antifungal defense genes within the *S_H3* genomic region of Arabica, Robusta, Liberica, and Excelsa coffee grown in the Philippines

Nick Rainier S. Santos*[1], Monaliza B. Magat[1,2], Jomari P. Urmatan[1], Ernelea P. Cao[1], and Daisy May C. Santos[1]

[1]Plant Molecular Biology and Genetics Laboratory, Institute of Biology, College of Science University of the Philippines Diliman, Quezon City, 1101
[2]Science and Technology Scholarship Division, Science Education Institute, Department of Science and Technology, Bicutan, Taguig City, 1631

ABSTRACT

The *S_H3* genes are a group of well-studied antifungal coffee defense genes, which have been localized to a region in chromosome 3. A previous study screened this region for other antifungal defense genes, particularly, those implicated in defense against coffee leaf rust (CLR) and coffee berry disease (CBD). Three candidate defense gene loci have been identified, namely, an ethylene-responsive transcription factor 1B (*ERF1B*) gene, a chalcone synthase 2-like (*CHS2*) gene, and a major allergen Pru ar 1-like gene. For this study, the main goal is to evaluate and compare the sequences of these selected loci from a sample of Philippine coffee accessions, which consist of the four main varieties utilized in local production: Arabica (*Coffea arabica*), Robusta (*C. canephora*), Liberica (*C. liberica* var. *liberica*) and Excelsa (*C. liberica* var. *dewevrei*). DNA was extracted from young leaves and subjected to PCR amplification using newly designed primers. The PCR products were run through agarose gel electrophoresis after which, gel portions containing the target bands were excised and processed for bidirectional sequencing. Consensus sequences were constructed from the raw sequences using UGENE. Due to probable primer specificity issues, the *CHS2* sequences were excluded from further analysis. In order to analyze the *ERF1B* and Pru ar 1-like gene sequences, MEGA11 was used for multiple sequence alignment, protein sequence prediction, and construction of neighbor-joining trees. The sequences were also cross-referenced against the GenBank database using BLAST. Results show that several of the sequences represent novel orthologs, especially those obtained from the Liberica and Excelsa samples. Furthermore, orthologs which are most likely linked to conferring greater resistance were determined. Most notable among these are the *ERF1B* sequences of a Red Bourbon (A-RB, NSIC-2008-Cf-A-05) and a Mundo Novo (A-MN) sample. Despite originating from Arabica samples, the A-RB and A-MN *ERF1B* orthologs show greater similarity to their counterparts in Liberica/Excelsa and Robusta samples, respectively. These orthologs are recommended for follow-up studies through

*Corresponding author
Email Address: nssantos1@up.edu.ph
Date received: 17 July 2024
Dates revised: 23 September 2024; 02 November 2024; 24 February 2025
Date accepted: 27 April 2025
DOI: https://doi.org/10.54645/2025181XOI-52

KEYWORDS

Coffea, disease, fungi, locus, resistance

additional *in silico* analysis and genotype-phenotype association so they can definitively be considered as target genes in future coffee breeding programs.

INTRODUCTION

The Philippines has lagged in improving local coffee production. To address this concern, the Philippine government crafted the Philippine Coffee Industry Roadmap (PCIR) 2021–2025 (Retrieved from https://pcaf.da.gov.ph/index.php/cir-coffee/ on 15 September 2024) as a guide in all efforts at increasing production. Diseases and pests have been identified as one of the limiting factors in coffee production. One of the long-term goals in the roadmap is to utilize marker-assisted breeding to develop coffee lines with a combination of desirable traits, including resistance against diseases and pests. While resistance is polygenic, some studies have shown that the addition or modification of a few select genes may be enough to improve plant defense greatly (Wang et al. 2018; Sarowar et al. 2019). Therefore, only those coffee defense genes, which contribute the most to resistance, can be considered for gene pyramiding. For example, Nasiro and Teferi (2019) describes a breeding program whereby several theoretical defense genes against the most prevalent fungal diseases, coffee berry disease (CBD, *Colletotrichum kahawae*) and coffee leaf rust (CLR, *Hemileia vastatrix*), are combined into a single coffee lineage. This lineage would also contain theoretical genes for other desirable traits such as tree compactness and beverage quality. However, most coffee defense genes have yet to be fully characterized and identified (Santos et al. 2024). Among the anti-CLR S_H1–9 genes, for example, only the S_H3 genes have been thoroughly investigated (Mahé et al. 2008; Nagaño et al. 2022; da Silva Angelo et al. 2023). These studies noted that the resistant variants of the S_H3 gene are most effective against CLR compared to the resistant variants of other S_H genes. Considering this, other defense genes associated with the S_H3 genes should be evaluated further. A recent review paper (Santos et al. 2024) has identified such genes by synthesizing data from previous studies. Initially, Nagaño et al. (2022) mapped several S_H3-linked genetic markers (Mahé et al. 2008) to a ~10 Mbp region in chromosome 3. Within this region, Santos et al. (2024) then searched for the loci corresponding to defense genes which were identified in differential expression analysis studies (Diniz et al. 2012; Diniz et al. 2017). In the first study, Diniz et al. (2012) investigated genes which are differentially expressed between Hibrido de Timor coffee infected by *H. vastatrix* (CLR) and Hibrido de Timor coffee infected by *Uromyces vignae*. In the second study, Diniz et al. (2017) investigated genes which are differentially expressed between the susceptible Caturra and the resistant Catimor (Caturra X Hibrido de Timor) trees upon infection of CBD. Eventually, Santos et al. (2024) found three candidate defense gene loci within the S_H3 region, namely, an ethylene-responsive transcription factor 1B (*ERF1B*) gene, a chalcone synthase 2-like (*CHS2*) gene, and a major allergen Pru ar 1-like gene. Using as reference a recent study (da Silva Angelo et al. 2023), which performed comparative sequence analysis on the S_H3 genes across different coffee species, this study aimed to conduct a similar analysis of the candidate defense gene loci.

In this study, the coffee samples included accessions from the Arabica (*Coffea arabica* L.), Robusta (*C. canephora* Pierre ex A. Froehner), Liberica or Barako (*C. liberica Hiern* var. *liberica*), and Excelsa (*C. liberica Hiern* var. *dewevrei*) varieties. Additionally, these accessions are owned by the Baguio City branch of the Department of Agriculture - Bureau of Plant Industry (DA-BPI) and the Cavite State University (CvSU) in Indang, Cavite. According to the PCIR 2021–2025, the four coffee varieties are utilized in local coffee production, with DA-BPI and CvSU as among the main Philippine institutions involved with coffee seedling distribution and research. Furthermore, some of these accessions have already been screened with genetic markers in previous studies (Cao et al. 2014; Panaligan et al. 2020; Santos et al. 2023) and the results can be used as additional reference data in this study. From these accessions, selected *ERF1B*, *CHS2*, and Pru ar 1-like gene loci were sequenced. For each gene, the genetic differences were analyzed on how they are distributed across the accessions, and on how they might affect the predicted protein product.

MATERIALS AND METHODS

Plant Materials and DNA Extraction

Philippine accessions of Arabica, Robusta, Liberica, and Excelsa varieties were selected for this study. These include nine Arabica (A-Cu, A-RB, A-YB, A-Cm, A-SR, A-TP1, A-TP2, A-MN, A-Mo), six Robusta (Ro1–6), three Liberica (Lb1–3), and nine Excelsa (Ex1–9) trees, located at either DA-BPI-Baguio or CvSU (Table 1). From these trees, young leaves were collected at the shoot tips. DNA extraction was performed on the leaf samples using a modified version of the cetyltrimethylammonium bromide (CTAB) protocol, which was described by Healey et al. (2014). There were additional modifications to the protocol in this study, but these simply involve adjusting amounts towards smaller tubes due to specifications of the available centrifuge (Santos et al. 2023).

Primer Design

Based on a previous study by Santos et al. (2024), reference loci in chromosome 3 were selected. Using the positions of these loci within the *C. arabica* genome (GCA_003713225) in GenBank as basis (Table 2, Figure 1), PrimerBlast was then used to design new primers for each gene (Tables 3 and 4). The primers were synthesized and ordered from Macrogen, Inc., South Korea through Kinovett Scientific, Philippines. For each gene, two sets of primers were designed.

The first set of primers were used for PCR amplification (Table 3). For each target region, up to three primer pairs were designed. Inside the box for "Enter accession, gi, or FASTA sequence" under "PCR Template", the accession number of the corresponding chromosome was inputted. The boundaries shown in Figure 1 were used as reference for "Range" (Appendix I). Under "Primer Pair Specificity Checking Parameters", "Specificity check" was selected. "Database" was set to "Refseq representative genomes" and "Organism" was set to "*Coffea arabica* (taxid:13443)". The primer pairs were tested first for PCR amplification on some of the Arabica samples. For each gene, the primer pair, which yields the optimal result, was chosen for subsequent PCR amplification of the remaining coffee samples (Table 3).

The second set of primers was used as additional primers during DNA sequencing (Table 4). Macrogen, Inc. states that their standard sequencing features "high quality results and normal read length (1,050bp)" (Retrieved from https://www.macrogen.com/en/business/research/ces on 15 September 2024). Approaching this nucleotide length, the quality of the generated chromatograms in this study start to drop. Therefore, internal primers were designed for amplification products greater than 1,000bp such that they would begin near the middle of the gene or exon (Appendix I). In some cases, additional internal primers were designed in order to bypass mononucleotide repeats which cause stuttering in the chromatograms (Appendix II).

Table 1: Coffee samples investigated in this study.

Owner[a]	Owner Designation	Coffee Species	Variety	Code
DA-BPI-Baguio	NSIC 2008 Cf-A 05	*C. arabica*	Red Bourbon	A-RB
DA-BPI-Baguio	NSIC 2008 Cf-A 06	*C. arabica*	Caturra	A-Cu
DA-BPI-Baguio	NSIC 2008 Cf-A 07	*C. arabica*	Yellow Bourbon	A-YB
DA-BPI-Baguio		*C. arabica*	Catimor	A-Cm
DA-BPI-Baguio		*C. arabica*	San Ramon	A-SR
DA-BPI-Baguio		*C. arabica*	Typica	A-Tp1
DA-BPI-Baguio		*C. arabica*	Typica	A-Tp2
CvSU		*C. arabica*	Mocha	A-Mo
CvSU		*C. arabica*	Mundo Novo	A-MN
CvSU	NSIC 2007 Cf-R 07	*C. canephora*	Robusta	Ro1
CvSU	CVS-00595	*C. canephora*	Robusta	Ro2
CvSU	"FRT-03"	*C. canephora*	Robusta	Ro3
CvSU	"FRT-08"	*C. canephora*	Robusta	Ro4
CvSU	"FRT-11"	*C. canephora*	Robusta	Ro5
CvSU	"FRT-16"	*C. canephora*	Robusta	Ro6
CvSU	NSIC 2007 Cf-L 01	*Coffea liberica* var. *liberica*	Liberica	Lb1
CvSU	CVS-00472	*Coffea liberica* var. *liberica*	Liberica	Lb2
CvSU	CVS-00590	*Coffea liberica* var. *liberica*	Liberica	Lb3
CvSU	CVS-00577	*Coffea liberica* var. *dewevrei*	Excelsa	Ex1
CvSU	"D7-B3"	*Coffea liberica* var. *dewevrei*	Excelsa	Ex2
CvSU	CVS-00511.1	*Coffea liberica* var. *dewevrei*	Excelsa	Ex3
CvSU	CVS-00545	*Coffea liberica* var. *dewevrei*	Excelsa	Ex4
CvSU	CVS-00551.1	*Coffea liberica* var. *dewevrei*	Excelsa	Ex5
CvSU	CVS-00560	*Coffea liberica* var. *dewevrei*	Excelsa	Ex6
CvSU	CVS-00570	*Coffea liberica* var. *dewevrei*	Excelsa	Ex7
CvSU	CVS-00570.0	*Coffea liberica* var. *dewevrei*	Excelsa	Ex8
CvSU	CVS-00573	*Coffea liberica* var. *dewevrei*	Excelsa	Ex9

[a]The abbreviations stand for the following: DA-BPI-Baguio — Baguio branch of the Department of Agriculture, Bureau of Plant Industry; CvSU — Cavite State University; NSIC — National Seed Industry Council.

PCR Amplification

PCR amplification was performed in a volume of 12.5 µL. The melting temperatures (Table 3) provided by Macrogen, Inc. were used as reference for the computation of annealing temperatures. According to Maddocks and Jenkins (2017), identifying the optimal annealing temperature should start 5°C below the lower melting temperature of a primer pair before making gradual increments after every PCR run. To test each primer pair in this study, only up to three PCR runs were conducted, with 1°C increment in annealing temperature after each trial run. Per reaction tube, the PCR mix contained the following: 2.5 µL 5x MyTaq Reaction Buffer (with dNTPs) (Bioline), 0.5 µL (10 pmoles/µL) of each Primer, 0.25 µL 5 U/µL Taq Polymerase (Bioline), 1 µL of 25 µg/mL DNA, and 7.75 µL double distilled water. PCR was run using the following conditions: initial denaturation at 95°C for 3 min; 44 cycles of denaturation at 95°C for 30 s, annealing temperature (Table 3) for 30 s, extension at 72°C for 60 s per 1,000 bp; and a final extension at 72°C for 10 min. PCR reactions were performed using the T-100 Thermal Cycler (Bio-Rad, Singapore).

Table 2: Differentially-expressed genes during defense of coffee against fungi.

	Previous Literature				GenBank Database			
	Gene[a]		Primers		RNA & Gene[b]	Gene Locus Position[c]		
Diniz et al. 2012	CaRLK	receptor-like kinase	F:	ATGGGAGAAAAGAATGGCAGAAG	XM_027214852	Chr	6e	7,825,693
			R:	GGCCAATTACAGTTTGAAAACACC	LOC113695712		–	7,829,342
	CaWRKY1	tryptophan-arginine-	F:	TGCAACAAGGACAGCACCAG	XM_027229905	Chr	9c	1,611,951
		lysine-tyrosine (TF) 1	R:	CGTGATCGCGGCCGT	LOC113707640		–	1,614,729
	CaPAL	Phenylalanine	F:	GCAGGTCCTACTCATTATACAAGC	XM_027249163	Chr	2c	9,011,522
		ammonia-lyase	R:	CCATTCCACTCTTTCAAACAATCC	LOC113725801		–	9,016,457
	CaCHS	**chalcone synthase**	**F:**	**GTTACAGCAGCCCAAACCAT**	**XM_027260954**	**Chr**	**3c**	**6,583,416**
			R:	**TACCGAGAGGCTCAAATGCT**	**LOC113734419**		**–**	**6,595,921**
	CaLOX13	lipoxygenase 13	F:	GGTCCGCAAGTGTGTGAAC	XM_027269026	Chr	4e	1,537,669
			R:	GCAAGCCAGATGAGAGTAGTC	LOC113741495		–	1,542,545
	CaGT	(salicylic acid)	F:	ACTCCAGCAACAACCACCATTA	XM_027219437	Chr	1e	4,327,914
		glucosyl transferase	R:	GAGACGTCTTGCAAGGTTTTGA	LOC113699229		–	4,330,144
	CaPR1	pathogenesis-	F:	GATTACCTGGACGCCCATAA	XM_027256902	Chr	2e	16,960,631
		related 1	R:	GCTGCCAGGTTTTCTCCATA	LOC113731565		–	16,961,375
	CaPR10	**pathogenesis-**	**F:**	**GCCACCATCCTTGAAGAGAA**	**XM_027263515**	**Chr**	**3e**	**10,864,536**
		related 10	**R:**	**CAACTCTCTGCTTGGCAGTCT**	**LOC113736496**		**–**	**10,874,107**
Diniz et al. 2017	PAL	Phenylalanine	F:	GCAGGTCCTACTCATTGTACAAG	XM_027249163	Chr	2c	9,011,522
		ammonia-lyase	R:	CCATTCCACTCTTTCAAACAATCC	LOC113725801		–	9,016,457
	ICS2	Isochorismate	F:	TGCCATAGTACGAGAAAACA	XM_027268894	Chr	4e	40,582,581
		synthase 2	R:	CCCAGAAAATCGACCATAAA	LOC113741375		–	40,595,672
	NPR1	non-expressor	F:	AGGGCATTGGATTCTGACGA	XM_027209535	Chr	6c	7,325,163
		of PR1	R:	CTCTGTTGTGGTCTTTGCGT	LOC113691407		–	7,330,030
	PR1	pathogenesis-	F:	GCCCGTAAAGTCACCTGT	XM_027250295	Chr	2c	16,975,012
		related 1	R:	AACTACGCTGCCAAAATC	LOC113726532		–	16,975,753
	OPR3	12-oxophytodienoate	F:	ATAACTCCCCACCTTCCAAC	XM_027234844	Chr	10e	9,313,760
		reductase 1-like	R:	ACAGCCTTATCCCAACTCTAT	LOC113711671		–	9,316,264
	COI1	coronatine insensitive 1	F:	CTTAGCATCACCACCCACC	XM_027255045	Chr	2e	5,419,734
			R:	TCCGATCCCCCATACCAAC	LOC113730398		–	5,425,215
	PR10	**pathogenesis-**	**F:**	**GCCACCATCCTTGAAGAGAA**	**XM_027263515**	**Chr**	**3e**	**10,864,536**
		related 10	**R:**	**CAACTCTCTGCTTGGCAGTCT**	**LOC113736496**		**–**	**10,874,107**
	ACS5	1-aminocyclepropane-1-	F:	AGGGCGTCCTGGTCACTAA	XM_027222285	Chr	7e	2,900,512
		carboxylic acid synthase 5	R:	CTCGGCGAGCTAAAAACTGT	LOC113701564		–	2,902,787
	ACO2	1-aminocyclepropane-1-	F:	AAAGTCAGCAATTACCCTCCA	XM_027233429	Chr	9e	595,591
		carboxylic acid oxidase 2	R:	ATCCACCCATTCACCATCCT	LOC113710424		–	597,644
	ETR1	ethylene resistant 1	F:	GCCCCCAAGATATTCCTAAG	XM_027208543	Chr	5c	40,983,814
			R:	TGCAAGACCAAGACCACTAC	LOC113690575		–	40,990,911
	EIN2	ethylene insensitive 2	F:	GTTACTTCTCCAAAACCTACT	XM_027215187	Chr	6e	3,261,308
			R:	TCCCATTTACCACTCTTATCT	LOC113695947		–	3,269,034
	CTR1	constitutive triple	F:	GCAGCTGTGGGTTTCAAGG	XM_027215384	Chr	6e	4,358,091
		response 1	R:	AGTGGGGGAGGGTTTAGTC	LOC113696059		–	4,368,257
	C312112	**ethylene-responsive**	**F:**	**TGGCTGGGCACATTTGAC**	**XM_027261704**	**Chr**	**3c**	**4,367,998**
		factor 1	**R:**	**GGATTGCTGCTTGACCTC**	**LOC113734928**		**–**	**4,368,996**

[a]The genes in bold were localized to Chromosome 3, where the S_H3 genes are located.
[b]In order to localize the differentially-expressed defense genes from previous studies, Santos et al. (2024) ran the primers through PrimerBlast. "Database" was set to "Refseq mRNA" and "Organism" was set to "Coffea arabica (taxid:13443)". Targets with more than 1 primer mismatch were excluded. The top mRNA result and the position of its corresponding gene locus in the *C. arabica* genome (GCA_003713225) are shown.
[c]Gene loci are based on the *C. arabica* genome (GCA_003713225) in GenBank. *C. arabica* is an allotetrapoloid hybrid of *C. canephora* and *C. eugenioides*. Chromosomes 1c–11c represent those derived from the *C. canephora* ancestor while Chromosomes 1e–11e represent those derived from the *C. eugenioides* ancestor.

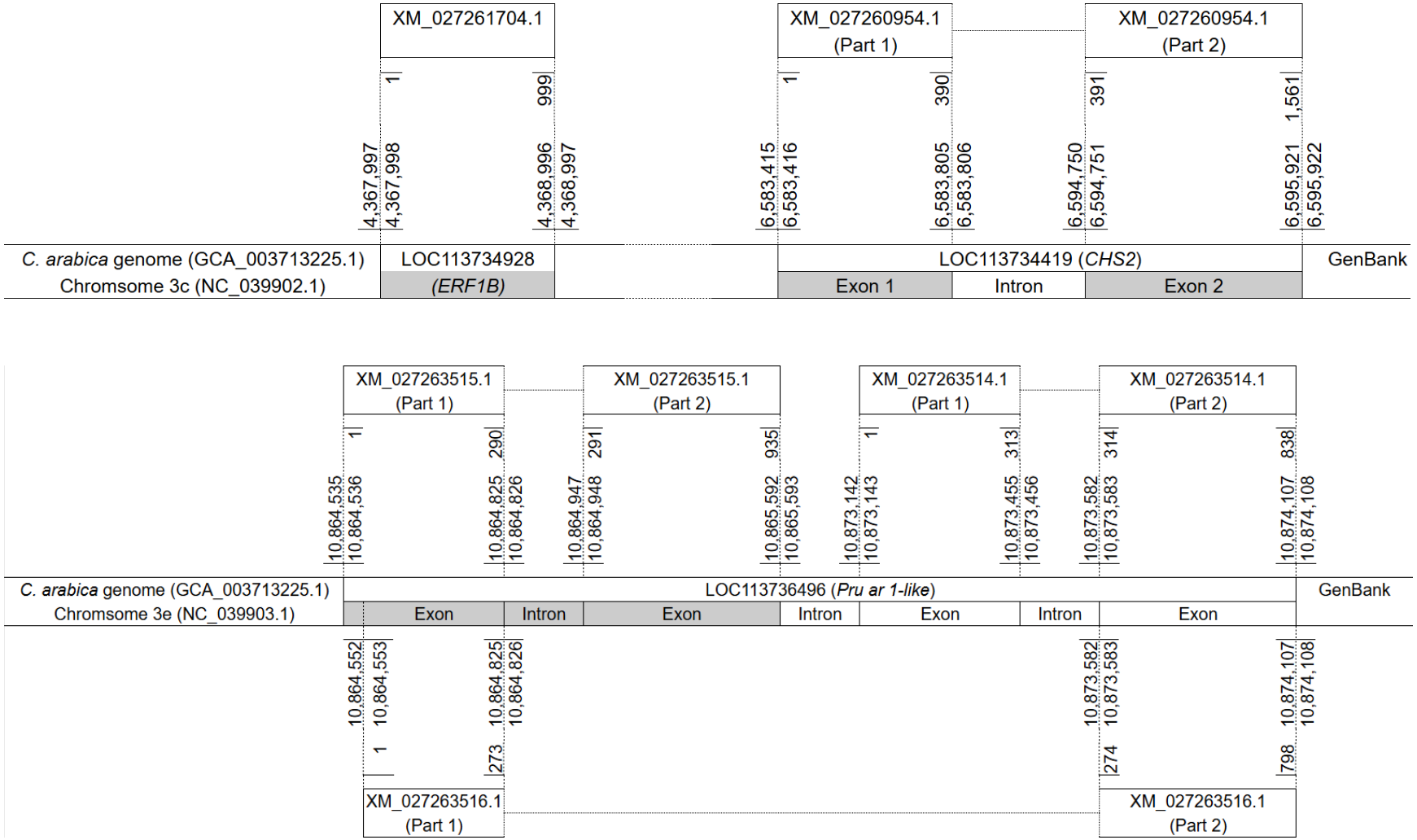


Figure 1. Reference loci at the time of experimentation. The *C. arabica* genome, GCA_003713225, was used as reference. The loci are shown alongside their corresponding mRNA within their respective chromosomes. The ethyelene-responsive factor (*ERF1B*) and chalcone synthase (*CHS2*) genes are located in the chromosome 3 derived from the *C. canephora* ancestor while the Pru ar 1-like gene is located in the chromosome 3 derived from the *C. eugenioides* ancestor. The gray regions represent the target regions which had been sequenced from the samples in this study.

Table 3: Primers utilized in PCR amplification.

Gene Locus[a]		Primers[b]	Product Size	Coverage	Melt. Temp.	Anneal. Temp.[c]	PCR Amplification[d]														
							Arabica	Robusta						Liberica			Excelsa								
							RB	1	2	3	4	5	6	1	2	3	1	2	3	4	5	6	7	8	9
ERF1B LOC113734928	**F1:**	**CTGGCGGCCTAGCTTATTCA**	**1,953bp**	**Whole**	**61.4°C**	**57°C**																			
	R1:	**ATTCTGGCCCACTGGATGAC**		**Gene**	**63.0°C**		/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
	F2:	CGGTCGGCCGTATACATCAT	1,730bp	Whole	61.2°C	56°C																			
	R2:	TTTCTGGTCTGCATGTTGCC		Gene	60.3°C		/	/	/	/	/	/	/	/	X	X	X	X	/	X	X	X	X	/	X
CHS2 LOC113734419	**F1:**	**TGGCAACAAAAATTGAGAGTATGC**	**706bp**	**Exon 1**	**57.3°C**	**52°C**																			
	R1:	**TGCAAAATATTCGAGACCATGC**			**57.0°C**		X	/	/	/	/	X	/	X	X	X	/	/	/	/	/	/	/	X	/
	F2:	ACAAAAATTGAGAGTATGCAGGGA	685bp	Exon 1	58.1°C	54°C																			
	R2:	CCATGCATTTCTAAGCCTCCTAA			58.8°C		/[e]																		
	F3:	AACAAAAATTGAGAGTATGCAGGGA	693bp	Exon 1	58.3°C	54°C																			
	R3:	TTCGAGACCATGCATTTCTAAGCC			61.3°C		/[e]																		
	F1:	**CATTCCAACACGTTGATGCAGTA**	**1,804bp**	**Exon 2**	**59.2°C**	**54°C**																			
	R1:	**CATGTGCGCAACAGAAAGATG**			**58.3°C**		/	/	/	/	/	/	/	/	X	X	X	/	/	/	/	/	/	/	/
	F2:	TTTTGTTTTTGGCCTTCCCTGGAT	1,788bp	Exon 2	62.5°C	57°C																			
	R2:	TGGGATTTTCCTGTGACGAATCA			61.5°C		/	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
	F3:	TACATTCCAACACGTTGATGCAGTA	1,959bp	Exon 2	59.6°C	55°C																			
	R3:	TCTGCATTTGAGCCTTTCAACTC			59.6°C		/	X	/	/	/	X	/	X	X	X	X	/	X	X	/	/	/	/	/
Pru ar 1-like LOC113736496	**F1:**	**TTTCCTGGCAATTGCGTATGTATT**	**1,563bp**	**Exon 1–**	**58.0°C**	**53°C**																			
	R1:	**AAATTTCGTTGGCGACGTGA**		**Exon 2**	**59.3°C**		/	/	/	/	X	X	/	X	X	X	/	/	X	/	X	/	X	/	/
	F2:	TCCTGGCAATTGCGTATGTATTA	1,263bp	Exon 1–	57.2°C	53°C																			
	R2:	CAGGTTTAGCATCGACAGCC		Exon 2	60.5°C		/	X	/	X	/	X	X	X	X	X	/	/	X	X	X	X	X	/	/
	F3:	AATTTCGGGTGTGACGGAGT	1,442bp	Exon 1–	61.7°C	56°C																			
	R3:	TCGACAGCCAATGGCATAACA		Exon 2	60.8°C		/	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

[a]The selected loci in the *C. arabica* genome (GCA_003713225) at GenBank were used as reference.
[b]The primers were developed using PrimerBlast and then ordered from Macrogen Inc., Republic of Korea via Kinovett Scientific, Republic of the Philippines. The primers in bold for each gene or gene portion are those which were used to amplify targets in all samples.
[c]According to Maddocks and Jenkins (2017), looking for the optimal annealing temperature should start 5°C below the lower melting temperature of a primer pair before making gradual increments after every PCR run. To test each primer pair in this study, only up to three PCR runs were conducted, with 1°C increment in annealing temperature after each trial run.
[d]PCR amplification was successful for most Arabica samples. For the other samples, "/" denotes detection of the target band in the agarose gel while "X" denotes absence. In order to rule out experimental error, absence of PCR products is confirmed by at least three separate PCR runs.
[e]When the PCR products were sequenced, the poor quality of the chromatograms suggest that more than one region was amplified.

Table 4: Additional Primers utilized in DNA sequencing.

Gene	Locus[a]	Primers[b]	Coverage	Melt. Temp.
ERF1B	LOC113734928	F1B: GGGGCAGTCCAAGTTTCCAA	1st Half[c]	64.2°C
		F1C: CTCTATGGTCTCCTGGCACA	2nd Half[d]	61.7°C
		R2 : TTTCTGGTCTGCATGTTGCC	2nd Half	60.3°C
		R1B: TTCAATGCAACGACAGGGGA	1st Half[d]	62.1°C
CHS2	LOC113734419	R1B: AGATTAATCACAAAAGCAAAAGGGT	Exon 1[c]	56.4°C
		F1B: GCGTTACATGTACTTGACAGAGG	Exon 2 (1st Half[c])	59.2°C
		F1C: CGAAGTTGAGAGGCCCTTGT	Exon 2 (2nd Half[d])	62.8°C
		R1B: GGAACGTAAGCCCAACCTCA	Exon 2 (1st Half[d])	62.9°C
Pru ar 1-like	LOC113736496	F1B: AGTTGACCACATTTGTTGATGGT	Exon 2[d]	58.8°C
		R1B: CTTGGCAGTCTTATGCTCGC	Exon 1[d]	60.0°C

[a]The selected loci in the *C. arabica* genome (GCA_003713225) at GenBank were used as reference.
[b]The primers were developed using PrimerBlast and then ordered from Macrogen Inc., Republic of Korea via Kinovett Scientific, Republic of the Philippines.
[c]These primers were designed to bypass mononucleotide repeats which cause stuttering.
[d]These primers were designed for PCR products greater than 1,000bp.

Gel Electrophoresis, DNA Sequencing, and Comparative Sequence Analysis

The amplified PCR products were run through 1% agarose gel electrophoresis. The gels were then viewed under UV light and the gel portions displaying the target band were excised. The excised gels were sent to Macrogen, Inc., South Korea through Kinovett Scientific, Philippines, for gel extraction and bidirectional DNA sequencing. Unipro UGENE v50.0 (Okonechnikov et al. 2012) was used to construct consensus sequences. For each gene, the raw forward and reverse sequences were aligned so that the consensus sequence would have a 2X coverage. The initial consensus sequences were then trimmed further so that only the target regions (Figure 1) remained. MEGA11 (Tamura et al. 2021) was used for multiple sequence alignment, protein sequence prediction, and constructing neighbor-joining (NJ) trees. For the NJ trees, the number of Bootstrap replications was set to 1000, while model/method was set to Maximum Composite Likelihood. BLAST programs were also used for further *in silico* analyses.

RESULTS AND DISCUSSION

Reference Loci, Gene Description, and PCR Amplification

The investigated genes in this study include an ethylene-responsive transcription factor 1B (*ERF1B*) gene, a chalcone synthase 2-like (*CHS2*) gene, and a major allergen Pru ar 1-like gene (Figure 1). These genes have initially been implicated in defense against fungi, such as CLR and CBD, based on differential gene expression analysis (Diniz et al. 2012; Diniz et al. 2017). In another study (Santos et al. 2024), their corresponding loci were determined to be in the S_H3 region of chromosome 3 (Table 2). Aside from the S_H3 loci, this region contained several simple sequence repeats (SSRs) and sequence characterized amplified region (SCAR) markers linked to resistance against CLR (Mahé et al. 2008; Nagaño et al. 2022). Considering these, the candidate gene loci warranted further investigation as carried out in this study. In the next paragraphs, each gene and its corresponding locus are described. Preliminary results regarding PCR amplification are also discussed. The resulting sequences can be retrieved from the National Center for Biotechnology Information (NCBI) website (accession numbers: PV622660–PV622730).

The ERF super-family in plants has been generally implicated in defense against various pathogens and insects (Feng et al. 2020). Going back to the study of Diniz et al. (2017) involving CBD, the ethylene-responsive factor 1 transcript is overexpressed in the susceptible samples relative to their resistant counterparts. Against CLR, however, the role of ERFs is less clear (Silva et al. 2022). The ethylene-responsive factor 1 transcript from Diniz et al. (2017) was localized to the gene locus described as ethylene-responsive transcription factor 1B (*ERF1B*, LOC113734928) in chromosome 3c (Table 2, Figure 1). The locus does not contain any intron. There are similar loci upstream, namely, ethylene-responsive transcription factor ERF098-like (LOC113733775) and ethylene-responsive transcription factor ERF096-like (LOC113733803). In this study, one primer pair based on the *ERF1B* locus successfully amplified its target for all coffee samples (Table 3). Hence, the *ERF1B* genes of all samples in this study were sequenced and analyzed in the next section.

Chalcone synthase is a key enzyme for flavonoid biosynthesis and its expression has been linked to resistance against bacteria and fungi (Dao et al. 2011). The chalcone synthase transcript from Diniz et al. (2012) was localized to the gene locus described as chalcone synthase 2-like (*CHS2*, LOC113734419) in chromosome 3c (Table 2, Figure 1). The locus has two exons separated by a ~10kbp intron. There is a duplicate chalcone synthase 2-like locus (LOC113734418) upstream. In this study, several issues were encountered in obtaining the chalcone synthase 2-like sequences. Firstly, the very long intron of the reference locus necessitated the development of separate primer pairs for each exon (Table 3) and the exclusion of the intronic region from analysis in this study. Secondly, none of the primer pairs could amplify their target regions for all samples. Thirdly, the generated chalcone synthase 2-like sequences contain multiple sites with ambiguous nucleotides, much more than either the *ERF1B* sequences or the Pru ar 1-like sequences do. While nucleotide ambiguity at certain sites could be interpreted as heterozygosity in some cases (Appendix III), it is more likely that the newly designed primer pairs for both exons of chalcone synthase 2-like in this study had specificity issues. During preliminary sequencing, two primer pairs for exon 1 (Table 3) had to be rejected because the resulting poor-quality chromatograms suggested non-specific amplification of highly similar regions. In addition, it should also be noted that the

reference locus and its duplicate are both followed downstream by partial exonic sequences of chalcone synthase (LOC113735466 and LOC113735467), and these were probably confounding factors during primer design and PCR amplification. This puts into question whether the gathered *CHS2* exon 1 and exon 2 sequences belong to the same locus or not. Considering these results, the *CHS2* sequences were excluded from a more in-depth analysis in the rest of the study.

The pathogenesis-related 10 (PR10) transcript from Diniz et al. (2012) and Diniz et al. (2017) was localized to the gene locus described as major allergen Pru ar 1-like (LOC113736496) in chromosome 3e (Table 2, Figure 1). Pru ar 1-like allergens are members of the PR10 family (Longsaward and Viboonjun 2024). PR10 proteins have been implicated in the synthesis of defensive secondary metabolites, phytohormone signaling, and RNase activity against RNA of invading pathogens (Morris et al. 2020; Sinha et al. 2020; Longsaward and Viboonjun 2024). In CLR-resistant coffee, they have been reported to be upregulated (Silva et al. 2022). In the case of the reference Pru ar 1-like locus, it has four exons, representing three differentially spliced mRNA. For this study, however, only the first two exons were considered as these correspond to the XM_027263515 mRNA, which is the most likely target of the *PR10* primers used by Diniz et al. (2012) and Diniz et al. (2017). There is a duplicate locus upstream, major allergen Pru av 1-like (LOC113737776), and two duplicate loci downstream, both major allergen Pru ar 1-like (LOC113737365 and LOC113737366). In this study, the best primer pair based on the Pru ar 1-like locus was amplified in all Arabica samples, some of the Robusta (Ro1, Ro2, Ro3, and Ro6) and Excelsa (Ex1, Ex2, Ex4, Ex6, Ex8, and Ex9) samples, and none of the Liberica samples (Table 3). The resulting sequences from these were analyzed in the next section. Meanwhile, the absence of PCR products for some samples (Ro4, Ro5, Lb1, Lb2, Lb3, Ex3, Ex5, and Ex7) could be due to genetic variations in the primer binding sites or the absence of a direct homolog to the reference locus.

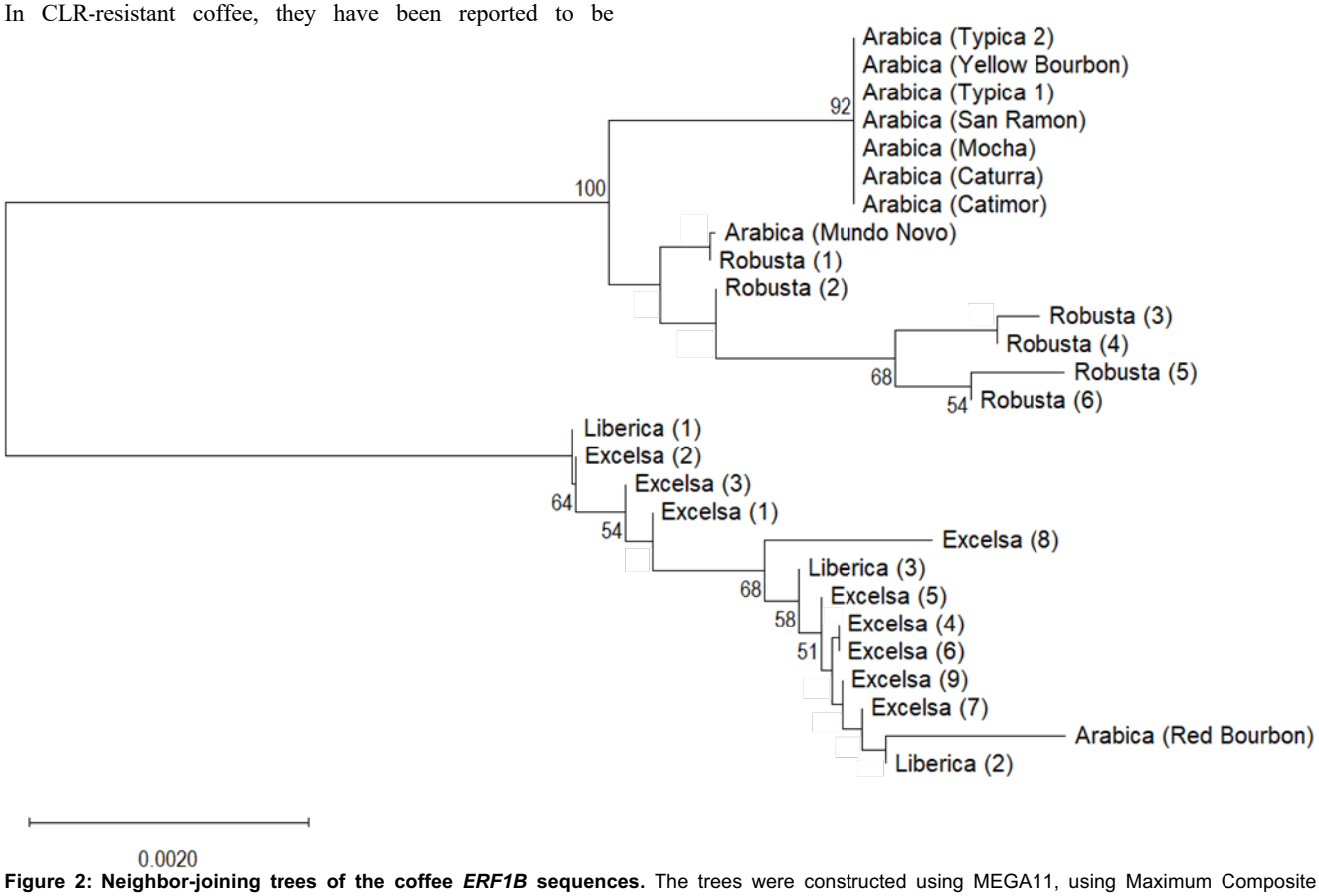


Figure 2: Neighbor-joining trees of the coffee *ERF1B* sequences. The trees were constructed using MEGA11, using Maximum Composite Likelihood. Values at the nodes represent the percentage of 1000 bootstrap samples; values less than 50% are not shown. For the Robusta, Liberica, and Excelsa samples, numbers in parenthesis are identifiers. Scale bar represents 2 nucleotide substitutions for every 1000 nucleotides.

Analysis of Defense Gene Orthologs in Various Coffee Varieties

Novel *ERF1B*, *CHS2*, and Pru ar 1-like gene sequences have been identified in Arabica, Robusta, Liberica, and Excelsa samples. A preliminary survey of these sequences shows that the Arabica samples generally have lower intraspecific variation compared to the Robusta, Liberica, and Excelsa samples. In each gene, for example, several of the Arabica sequences are identical despite belonging to different Arabica cultivars. This is consistent with what has been reported about the Arabica species. *C. arabica* has been observed to have low genetic diversity because of self-fertilization, and this has been associated with its general susceptibility against diseases and pests (Salojärvi et al. 2024). This is why in succeeding analyses, the sequences from Arabica can be presumed to represent the ortholog associated with less effective resistance.

Ethylene-responsive transcription factor 1B (LOC113734928)

As expected, most of the *ERF1B* sequences from the Arabica samples are identical to the reference *ERF1B* locus, LOC113734928. However, the *ERF1B* sequences of the Red Bourbon (A-RB) and Mundo Novo (A-MN) samples show unique nucleotide differences. Based on the NJ trees (Figure 2), the A-RB *ERF1B* sequence shows higher similarity with its counterparts in *C. liberica* (Liberica and Excelsa) samples, while the A-MN *ERF1B* sequence shows higher similarity with its counterparts among the Robusta samples. In particular, the A-RB sequence shares several unique nucleotide mutations with the Liberica and Excelsa sequences (Figure 3). As for the *ERF1B* sequences from the Robusta, Liberica, and Excelsa varieties, they clustered according to species (Figure 2). The sequences from Liberica and Excelsa failed to cluster according to variety. This trend has been observed in previous genetic studies involving the Liberica and Excelsa trees from Cavite

(Cao et al. 2014; Panaligan et al. 2020).

Analysis of the predicted protein sequences also provides interesting insights about variations in the gene. The various *ERF1B* genetic sequences generated by this study yield only four distinct predicted protein sequences because most nucleotide changes are either in the untranslated regions or in the coding region but is a silent mutation. For this part, the predicted protein (XP_027117505.1) of the reference locus (LOC113734928) is used as the reference protein. In this study, the first predicted protein comes from most of the Arabica *ERF1B* sequences, excluding those of A-MN and A-RB. It is identical to the reference protein. The second predicted protein comes from the A-MN and Robusta *ERF1B* sequences. Compared to the reference protein, there is one amino acid replacement, aa64Ile>Thr. The third predicted protein comes from the A-RB, Lb1, Ex1, Ex2, Ex3, Ex5, Ex6, Ex7, and Ex8 *ERF1B* sequences. Compared to the reference protein, there is one amino acid replacement, aa4Thr>Ser. The fourth predicted protein comes from the Lb2, Lb3, Ex4, and Ex9 *ERF1B* sequences. There are two amino acid replacements, aa4Thr>Ser and aa155 Met>Val. It is important to note that among these, the third predicted protein is 100% identical to the XP_027166847 protein, an *ERF1B* protein from *C. eugenioides*. According to a report by Zonneveld and Solano (Retrieved from https://repositorio.catie.ac.cr/bitstream/handle/11554/9780/Coffea_eugenioides_S_Moore_and_Coffea_stenophylla.pdf on 18 September 2024), *C. eugenioides* populations are generally resistant against diseases and pests. Therefore, the third predicted protein could be associated with conferring greater resistance compared to the others.

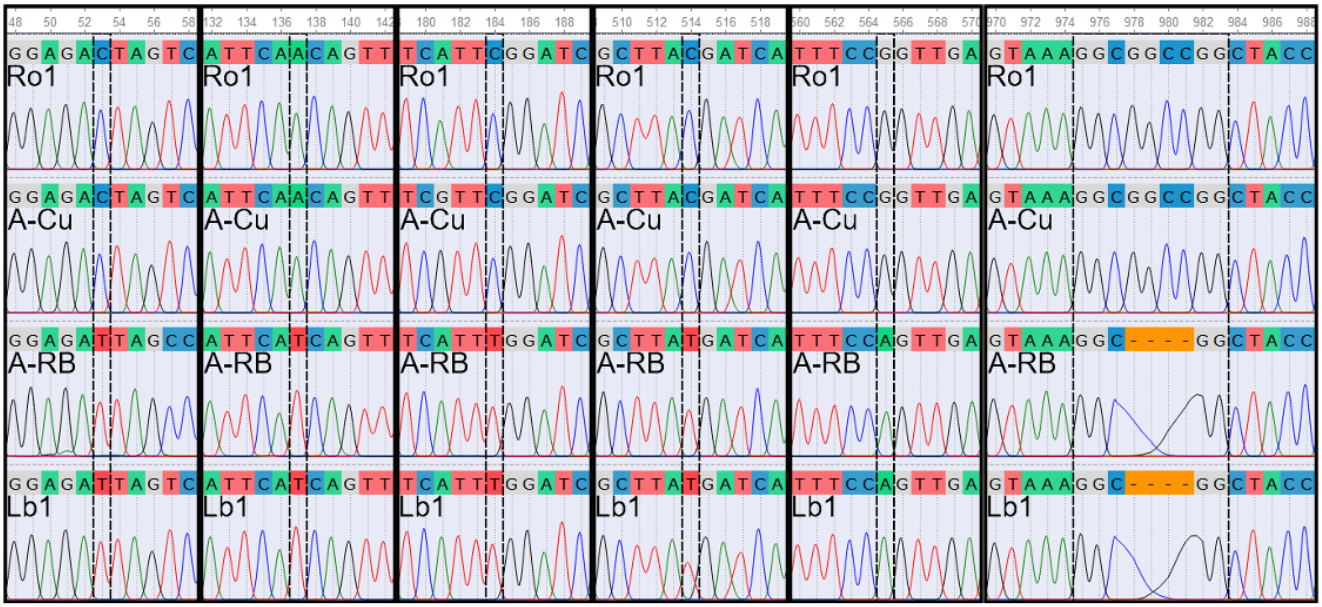


Figure 3: Unique *ERF1B* Mutations shared by A-RB and *C. liberica* samples. The polymorphic sites shown are those where the *ERF1B* of A-RB and *C. liberica samples* (e.g. Lb1) have unique mutations. The sequence from representatives of other Arabica (A-Cu) and Robusta (Ro1) samples are shown in contrast. The sites c53, c137, c184, c514, and c565 have single nucleotide substitutions while site c975–983 has an insertion-deletion (InDel). The locations are based on the XM_027261704.1 mRNA.

Additionally, the amino acid changes of the second to fourth predicted proteins fall outside the AP2 domain of the reference protein (aa89–152). The AP2 domain is the main feature of ERF proteins (Feng et al. 2020). The preservation of most of the amino acid sequences, despite the multiple changes in the nucleotide sequences, highly suggests the importance of this gene. Pathogens have been reported to produce effectors which target selected transcription factors of the plant hosts (Aerts et al. 2022). Hence, the *ERF1B* genetic mutations can be a defense mechanism against such pathogen effectors, whereby the resulting protein changes are enough to alter interactions with a potential effector but are too small to affect protein function.

Pru ar 1-like (LOC113736496)

Compared to the *ERF1B* sequences, there is much less genetic variation among the Pru ar 1-like sequences. As observed in the NJ tree (Figure 4), the available genetic sequences clustered according to species. Most Arabica Pru ar 1-like sequences are identical to the reference locus. Exceptions to this include the Pru ar 1-like sequences of the Catimor (A-Cm), Mocha (A-Mo), and A-MN samples, in which a single nucleotide substitution (c236A>C) was detected. Meanwhile, all available Robusta sequences (Ro1, Ro2, Ro3, and Ro6) are identical. When compared to the sequences of the other varieties, the main distinguishing feature is a huge deletion in the 3' untranslated region. For the available Excelsa sequences (Ex1, Ex2, Ex4, Ex6, Ex8, and Ex9), there are several nucleotide differences but most are in the untranslated regions, in the intron, or in the coding region but is a silent mutation.

The various Pru ar 1-like genetic sequences generated by this study also yield four distinct predicted protein sequences. For this part, the predicted protein (XP_027119316.1) corresponding to exon 1 and exon 2 of the reference locus (LOC113736496) is used as the reference protein. In this study, the first predicted protein comes from most of the Arabica Pru ar 1-like genetic sequences. It is also identical to the reference protein. The second predicted protein comes from the A-Cm, A-Mo, and A-MN Pru ar 1-like genetic sequences. When compared to the reference protein, there is one amino acid replacement, aa43Ile>Leu. The third predicted protein comes from the Ro1, Ro2, Ro3, and Ro6 Pru ar 1-like genetic sequences. When compared to the reference protein, there is one amino acid replacement, aa7Ser>Asn. In addition, this protein is more similar to XP_027117013, which is the predicted protein of a Pru ar 1-like locus (LOC113734609) in chromosome 3c. This is expected as chromosome 3c represents the chromosome 3 inherited by *C. arabica* from its *C. canephora* ancestor. Lastly, the fourth predicted protein comes from the Ex1, Ex2, Ex4, Ex6, Ex8, and Ex9 Pru ar 1-like genetic sequences. When compared to the reference protein, their predicted protein sequences yield only one amino acid replacement, aa64Leu>His. Interestingly,

the amino acid changes of the second to fourth predicted proteins fall within the region encompassed by the Bet v 1-like domain (aa5–151). The Bet v 1 domain is the main feature of PR10 proteins (Morris et al. 2020; Longsaward and Viboonjun 2024). Therefore, there is a high probability that the amino acid changes affect protein function. Given the currently available data, however, it cannot be determined if the altered protein function leads to an increase or a decrease in the conferred resistance.

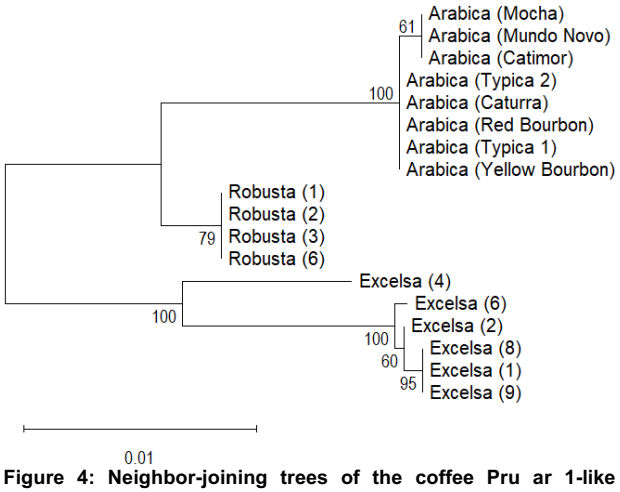


Figure 4: Neighbor-joining trees of the coffee Pru ar 1-like sequences. The trees were constructed using MEGA11, using Maximum Composite Likelihood. Values at the nodes represent the percentage of 1000 bootstrap samples; values less than 50% are not shown. For the Robusta, Liberica, and Excelsa samples, numbers in parenthesis are identifiers. Scale bar represents one nucleotide substitution for every 100 nucleotides.

Unique haplotypes in the A-RB and A-MN trees

When the *CHS2*, *ERF1B*, and Pru ar 1-like sequences are analyzed together, it can be inferred that the A-RB and A-MN trees have unique haplotypes in the S_H3 region of their chromosome 3c. In this study, it should be recalled that the reference loci for the *CHS2* and *ERF1B* genes are located in chromosome 3c of the reference Arabica genome (GCA_003713225) while the reference locus for the Pru ar 1-like gene is located in chromosome 3e. Congruently, the *ERF1B* and *CHS2* sequences from the A-RB and A-MN samples have notable differences from their counterparts in other Arabica samples. Meanwhile, the Pru ar 1-like sequences from the A-RB and A-MN samples have little or no difference from their counterparts in the other Arabica samples. This point is further illustrated by the NJ trees for the *ERF1B* (Figure 2) and Pru ar 1-like sequences (Figure 4). Therefore, other genes within the chromosome 3c region of the A-RB and A-MN trees should be up for investigation as well, as elaborated in the next sections.

As discussed earlier, the A-RB *ERF1B* sequence shows greater similarity to those of the *C. liberica* (Liberica and Excelsa) samples. This reinforces the findings of a previous study, where it is hypothesized that the A-RB tree contains a haplotype introgressed from *C. liberica*. Santos et al. (2023) utilized genetic marker analysis and reported that the A-RB sample also showed higher similarity with *C. liberica* samples for several S_H3-linked markers. These findings about the A-RB tree become even more relevant when additional information about the S_H3 genes is considered. A previous study (Mahé et al. 2008) reported a resistant S_H3 variant, which was introgressed from *C. liberica* into *C. arabica*. Furthermore, another previous study (Gutiérrez-Calle et al. 2021), which utilized genetic marker analysis, reported that there is a high similarity between *C. arabica* samples containing this resistant S_H3 variant and *C. liberica* samples. Considering data from this study and previous ones altogether, there is a high probability that the A-RB tree has the resistant S_H3 variant, reported by Mahé et al. (2008), within its chromosome 3c as well.

Meanwhile, the A-MN *ERF1B* sequence shows greater similarity to those of the Robusta samples. Unlike the A-RB tree however, the A-MN tree was not among the accessions screened with resistance-linked markers in a prior study (Santos et al. 2023). Based on the results of this study nonetheless (Figure 2), it can already be inferred that its unique haplotype in chromosome 3c originated from *C. canephora*. This is also relevant given that the introgression of defense genes from *C. canephora* to *C. arabica* has been reported to improve the resistance of the latter (Salojärvi et al. 2024). Given the distinct origin of its haplotype, the S_H3 genes of the A-MN tree probably differ from those of the A-RB tree and this adds to the value of this accession.

CONCLUSION

This study investigated three gene loci from Arabica, Robusta, Liberica, and Excelsa varieties for comparative sequence analysis. The coffee accessions came from DA-BPI-Baguio and CvSU, which are among the main Philippine institutions involved with the coffee industry. Meanwhile, the three gene loci code for *CHS2*, *ERF1B*, and Pru ar 1-like genes, and all are located within a previously identified S_H3 region of chromosome 3. Using newly designed primers, this study was able to obtain the *ERF1B* sequences from all the samples but not from all samples in the case of the *CHS2* and Pru ar 1-like gene sequences. Several of the sequences were novel, especially those from the Liberica and Excelsa samples. Because of quality issues, however, the *CHS2* sequences were excluded from further analysis. The genetic variation in the *ERF1B* and Pru ar 1-like gene sequences were assessed based on taxonomic relationships and based on their effect on the predicted proteins. Using data from this study in conjunction with data from previous studies, it was determined which orthologs were most likely linked to conferring greater resistance. It is recommended that these sequences be subjected to additional *in silico* analysis and genotype-phenotype association studies in order to determine which among them, if any, should be considered as target defense genes in future breeding programs.

A major incidental finding of this study are the unique sequences in two of the Arabica accessions, A-RB and A-MN. Based on analysis, it has been inferred that the two trees contain a unique haplotype in the S_H3 region of their chromosome 3c. Hence, it is also recommended that their S_H3 genes should be considered for analysis in future studies and then compared with the S_H3 sequences gathered by da Silva Angelo and co-workers (2023). These haplotypes can also be screened for more defense genes as well as genes linked to aroma, yield, and secondary metabolites. Furthermore, DA-BPI-Baguio and CvSU should prioritize the A-RB and A-MN trees, respectively, in terms of propagation, field monitoring, and further screening. For example, the A-MN tree should be subjected to the same genetic marker analysis done on the A-RB tree (Santos et al. 2023) in order to check for other possible unique haplotypes. Based on further findings, these trees could very well be chosen as parental sources of the target defense genes in breeding programs.

Issues were encountered during primer design for PCR amplification. Of the best primer pairs in this study, only the one for *ERF1B* successfully amplified its target product in all the samples. Defense gene loci, including those investigated in this study, tend to be duplicated. The presence of duplicate loci is probably why primers designed for them tend to have specificity and cross-species transferability issues. Therefore, it is recommended that in future studies, newly designed primers should be tested further with PrimerBlast against the available coffee genomes at GenBank. At the time of writing this study,

there are now eleven available coffee genomes consisting of four species, *C. arabica*, *C. canephora*, *C. eugenioides*, and *C. humblotiana* (Retrieved from https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=13442 on 3 May 2025). Lastly, it should be noted that updates have been made to some of the reference loci since the start of this study. These changes could make primer design in future studies even more accurate.

ACKNOWLEDGMENT

The authors acknowledge the additional help provided by Miguel Mondragon, Terrence Ferdinand Nagaño, Maria Theresa Tengco, and Danica Pearl Untiveros as research assistants. MT Tengco and DP Untiveros assisted in the procurement of laboratory materials, while M Mondragon and TF Nagaño helped during the sample collection. Furthermore, the authors would like to acknowledge the Baguio City branch of the Department of Agriculture - Bureau of Plant Industry and the Cavite State University in Indang, Cavite for providing leaf samples from their accessions for DNA extraction, and for assisting us during the sampling process. Lastly, the authors acknowledge the Office of the Chancellor of the University of the Philippines Diliman, through the Office of the Vice Chancellor for Research and Development, for funding this study (Outright Research Grant Project No. 232317 ORG). The authors declare no conflict of interest.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

CONTRIBUTIONS OF INDIVIDUAL AUTHORS

NR Santos is the principal and corresponding author. He conceptualized and designed the study, performed the laboratory work and analysis, and wrote the manuscript. The other authors provided assistance all-throughout. MB Magat and JP Urmatan mainly assisted in laboratory work and analysis. EP Cao mainly assisted in the analysis and manuscript writing. DM Santos is the senior author, who supervised in the study design, laboratory work, analysis, and manuscript writing.

REFERENCES

Aerts N, Chhillar H, Ding P, Van Wees SC. Transcriptional regulation of plant innate immunity. Essays Biochem 2022; 66(5):607–620. https://doi.org/10.1042/EBC20210100

Cao EP, Constantino-Santos DM, Ramos LA, Santos BS, Quilang JP, Mojica RM. Molecular and morphological differentiation among *Coffea* (Rubiaceae) varieties grown in the farms of Cavite Province, Philippines. Philipp Sci Lett 2014; 7(2):387–397.

da Silva Angelo PC, Sera GH, Shigueoka LH, Caixeta ET. Rust resistance S_H3 loci in *Coffea* spp. Physiol Mol Plant Pathol 2023; 127:102111. https://doi.org/10.1016/j.pmpp.2023.102111

Dao TT, Linthorst HJ, Verpoorte R. Chalcone synthase and its functions in plant resistance. Phytochem Rev 2011; 10(3):397–412. https://doi.org/10.1007/s11101-011-9211-7

Diniz I, Figueiredo A, Loureiro A, Batista D, Azinheira H, Várzea V, Pereira AP, Gichuru E, Moncada P, Guerra-Guimarães L, Oliveira H, Silva MdC. A first insight into the involvement of phytohormones pathways in coffee resistance and susceptibility to *Colletotrichum kahawae*. PLoS ONE 2017; 12(5):e0178159. https://doi.org/10.1371/journal.pone.0178159

Diniz I, Talhinhas P, Azinheira HG, Várzea V, Medeira C, Maia I, Petitot A-S, Nicole M, Fernandez D, Silva MdC. Cellular and molecular analyses of coffee resistance to *Hemileia vastatrix* and nonhost resistance to *Uromyces vignae* in the resistance-donor genotype HDT832/2. Eur J Plant Pathol 2012; 133:141–157. https://doi.org/ 10.1007/s10658-011-9925-9

Feng K, Hou X-L, Xing G-M, Liu J-X, Duan A-Q, Xu Z-S, Li M-Y, Zhuang J, Xiong A-S. Advances in AP2/ERF super-family transcription factors in plant. Crit Rev Biotechnol 2020; 40(6):750–776. https://doi.org/10.1080/07388551.2020.1768509

Gutiérrez-Calle SA, Sánchez-Díaz RA, Delgado-Silva YB, Montenegro JD, Gutiérrez DL, Maicelo-Quintana JL, Guerrero-Abad JC. Capillary electrophoresis as a tool for genotyping S_H3 mediated coffee leaf rust resistance. Sci Agropecu 2021; 12(1):91–99. https://doi.org/10.17268/sci.agropecu.2021.011

Healey A, Furtado A, Cooper T, Henry RJ. Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. Plant Methods 2014; 10:21. https://doi.org/10.1186/1746-4811-10-21

Longsaward R, Viboonjun U. Genome-wide identification of rubber tree pathogenesis-related 10 (PR-10) proteins with biological relevance to plant defense. Sci Rep 2024; 14:1072. https://doi.org/10.1038/s41598-024-51312-3

Maddocks S, Jenkins R. Chapter 2 - Designing and Ordering Your Polymerase Chain Reaction Primers. In: Jenkins R, Maddocks S, ed. Understanding PCR: A Practical Bench-Top Guide. Cambridge: Academic Press, 2017:11–30.

Mahé L, Combes M-C, Várzea VM, Guilhaumon C, Lashermes P. Development of sequence characterized DNA markers linked to leaf rust (*Hemileia vastatrix*) resistance in coffee (*Coffea arabica* L.). Mol Breed 2008; 21:105–113. https://doi.org/10.1007/s11032-007-9112-z

Morris J S, Caldo K M, Liang S, Facchini P J. PR10/Bet v1-like Proteins as Novel Contributors to Plant Biochemical Diversity. ChemBioChem 2020; 22(2):264–287. https://doi.org/10.1002/cbic.202000354

Nagaño TF, Santos NR, Santos DM, Cao EP. *In silico* physical mapping of resistance and resistance-associated genes in the S_H3 Region of the *Coffea canephora* Genome. Philipp J Sci 2022; 151(1):107–126. https://doi.org/10.56899/151.01.08

Nasiro K, Teferi D. Considerations in breeding and selection for certain important characters in Arabica coffee (*Coffea arabica* L.). J Biol Agric Healthc 2019; 9(11):21–28. https://doi.org/10.7176/JBAH/9-11-03

Okonechnikov K, Golosova O, Fursov M, UGENE team. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics 2012; 28(8):1166–1167. https://doi.org/10.1093/bioinformatics/bts091

Panaligan AC, Baltazar MD, Alejandro GJ. Genetic polymorphism of registered and popularly cultivated coffee (*Coffea* spp.) in the Philippines using inter-simple sequence repeats markers. Biodiversitas 2020; 21(9):4228–4233. https://doi.org/10.13057/biodiv/d210938

Salojärvi J, Rambani A, Yu Z, Guyot R, Strickler S, Lepelley M, Cui Wang, Rajaraman S, Rastas P, Zheng C, Muñoz DS, Meidanis J, Paschoal AR, Bawin Y, Krabbenhoft TJ, Wang ZQ, Fleck SJ, Aussel R, Bellanger L, Charpagne A, Fournier C, Kassam M, Lefebvre G, Métairon S, Moine D, Rigoreau M, Stolte J, Hamon P, Couturon E, Tranchant-Dubreuil C, Mukherjee M, Lan T, Engelhardt J, Stadler P, De Lemos SMC, Suzuki SI, Sumirat U, Wai CM, Dauchot N, Orozco-Arias S, Garavito A, Kiwuka C, Musoli P, Nalukenge A, Guichoux E, Reinout H, Smit M, Carretero-Paulet L, Filho OG, Braghini MT, Padilha L, Sera GH, Ruttink T, Henry R, Marraccini P, de Peer YV, Andrade A, Domingues D, Giuliano G, Mueller L, Pereira LF, Plaisance S, Poncet V, Rombauts S, Sankoff D, Albert VA, Crouzillat D, de Kochko A, Descombes P. The genome and population genomics of allopolyploid *Coffea arabica* reveal the diversification history of modern coffee cultivars. Nat Genet 2024; 56:721–731. https://doi.org/10.1038/s41588-024-01695-w

Santos NR, Cao EP, Santos DM. Genetic approaches to improving coffee resistance against diseases and pests in the Philippines. Philipp Agric Scientist 2024; 107(2):162–174. https://doi.org/10.62550/IU082023

Santos NR, Magat MB, Mondragon MV, Cao EP, Santos DM. Genetic profiling of locally registered Philippine coffee using molecular markers linked to resistance against diseases and pests. Biodiversitas 2023; 24(7):4136–4144. https://doi.org/10.13057/biodiv/d240752

Sarowar S, Alam ST, Makandar R, Lee H, Trick HN, Dong Y, Shah J. Targeting the pattern-triggered immunity pathway to enhance resistance to *Fusarium graminearum*. Mol Plant Pathol 2019; 20(5):626–640. https://doi.org/10.1111/mpp.12781

Sinha RK, Verma SS, Rastogi A. Role of pathogen-related protein 10 (*PR 10*) under Abiotic and Biotic Stresses in Plants. Phyton-IJEB 2020; 89(2):167–182. https://doi.org/10.32604/phyton.2020.09359

Silva MdC, Guerra-Guimarães L, Diniz I, Loureiro A, Azinheira H, Pereira AP, Tavares S, Batista D, Várzea V. An overview of the mechanisms involved in Coffee-*Hemileia vastatrix* Interactions: Plant and Pathogen Perspectives. Agronomy 2022; 12(2):326. https://doi.org/10.3390/agronomy12020326

Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis version 11. Mol Biol Evol 2021; 38:3022–3027. https://doi.org/10.1093/molbev/msab120

Wang J, Zhou L, Shi H, Chern M, Yu H, Yi H, He M, Yin J, Zhu X, Li Y, Li W, Liu J, Wang J, Chen X, Qing H, Wang Y, Liu G, Wang W, Li P, Wu X, Zhu L, Zhou J-M, Ronald P C, Li S, Li J, Chen X. A single transcription factor promotes both yield and immunity in rice. Science 2018; 361(6406):1026–1028. https://doi.org/10.1126/science.aat7675

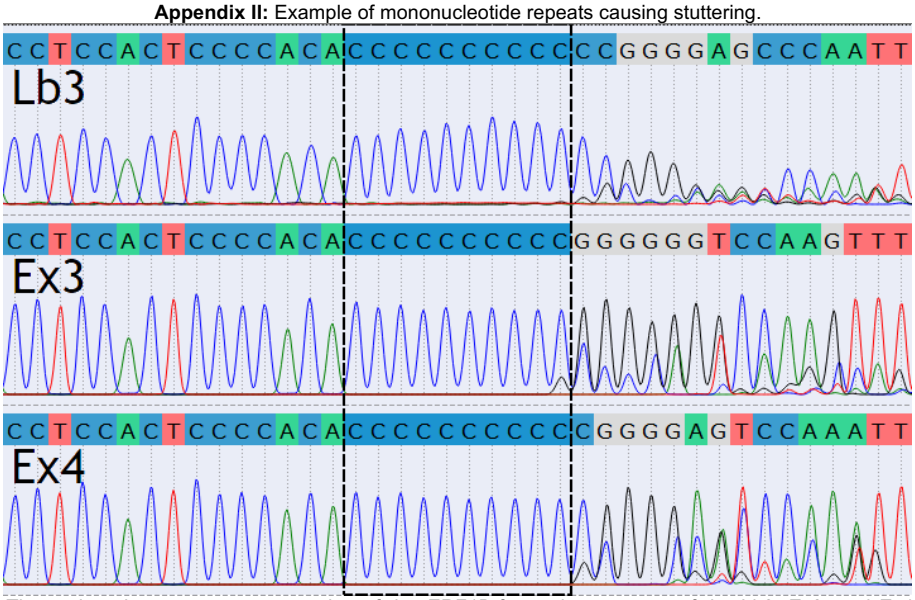
Waterhouse A M, Procter J B, Martin D M, Clamp M, Barton G J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 2009; 25(9):1189–1191. https://doi.org/10.1093/bioinformatics/btp033

Appendix I: PrimerBlast parameters during primer design.

Gene[a]	Chromosome	Primers		Range				PCR product size (Max)
				Forward		Reverse		
				From	To	From	To	
ERF1B	3c	F1		4,366,997	4,367,997	4,368,997	4,369,997	2,000
(LOC113734928)	(NC_039902.1)	F1B		4,367,997	4,368,097	4,368,997	4,369,997	2,000
		F1C		4,366,997	4,367,997	4,368,497	4,369,497	800
		F2		4,366,997	4,367,997	4,368,997	4,369,997	2,000
		R1		4,366,997	4,367,997	4,368,997	4,369,997	2,000
		R1B		4,368,497	4,369,497	4,368,997	4,369,997	800
		R2		4,366,997	4,367,997	4,368,997	4,369,997	2,000
CHS2	3c		F1	6,582,415	6,583,415	6,583,805	6,584,805	800
(LOC113734419)	(NC_039902.1)		F2	6,582,415	6,583,415	6,583,805	6,584,805	800
		Exon 1	F3	6,582,415	6,583,415	6,583,805	6,584,805	800
			R1	6,582,415	6,583,415	6,583,805	6,584,805	800
			R1B	6,582,415	6,583,415	6,583,805	6,583,855	1,000
			R2	6,582,415	6,583,415	6,583,805	6,584,805	800
			R3	6,582,415	6,583,415	6,583,805	6,584,805	800
			F1	6,593,750	6,594,750	6,595,922	6,596,922	2,000
			F1B	6,594,751	6,594,801	6,595,922	6,596,922	2,000
			F1C	6,594,336	6,595,336	6,595,922	6,596,922	800
		Exon 2	F2	6,593,750	6,594,750	6,595,922	6,596,922	2,000
			F3	6,593,750	6,594,750	6,595,922	6,596,922	2,000
			R1	6,593,750	6,594,750	6,595,922	6,596,922	2,000
			R1B	6,593,750	6,594,750	6,595,336	6,596,336	800
			R2	6,593,750	6,594,750	6,595,922	6,596,922	2,000
			R3	6,593,750	6,594,750	6,595,922	6,596,922	2,000
Pru ar 1-like	3e	F1		10,863,535	10,864,535	10,865,593	10,866,593	2,000
(LOC113736496)	(NC_039903.1)	F1B		10,863,948	10,864,948	10,865,593	10,866,593	1,000
		F2		10,863,535	10,864,535	10,865,593	10,866,593	2,000
		F3		10,863,535	10,864,535	10,865,593	10,866,593	2,000
		R1		10,863,535	10,864,535	10,865,593	10,866,593	2,000
		R1B		10,863,535	10,864,535	10,864,948	10,865,948	800
		R2		10,863,535	10,864,535	10,865,593	10,866,593	2,000
		R3		10,863,535	10,864,535	10,865,593	10,866,593	2,000

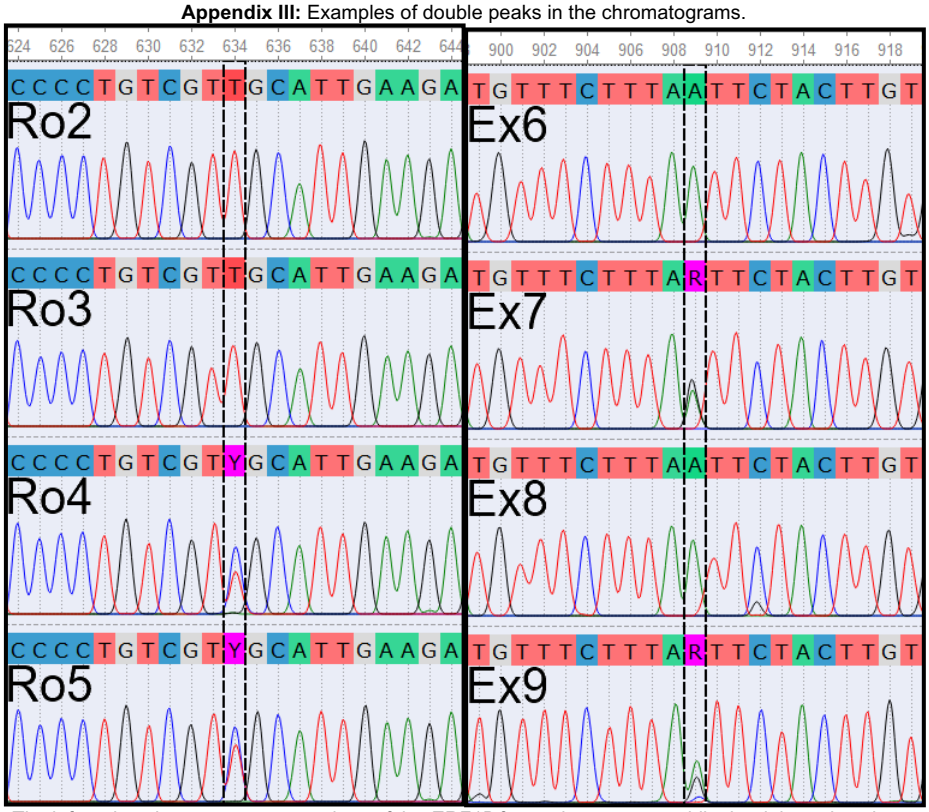
[a]The selected loci of the *C. arabica* genome (GCA_003713225) in GenBank were used as reference.

Appendix II: Example of mononucleotide repeats causing stuttering.



The region shown represent a portion of the *ERF1B* forward sequences of the Lb3, Ex3, and Ex4 samples. There is a notable decrease in quality after the "C" mononucleotide repeat.

Appendix III: Examples of double peaks in the chromatograms.



The left region shown represent portions of the *ERF1B* forward sequences while the right region shown represent portions of the *ERF1B* reverse sequences. Double peaks have been identified for the Ro4 and Ro5 samples at site c634 and for the Ex7 and Ex9 samples at site c909. A double peak for a particular polymorphic site can be interpreted as heterozygosity, but only if it is present in both forward and reverse sequences for a given sample.