

Genome-wide detection of SH3-associated molecular markers in *Coffea* cultivars using the Johara R-based bioinformatics pipeline

Joachim Xavier Po¹, Hanna Nicole Portento¹, Raissa Adellaide Sison^{1*}, and Jane Abigail Santiago^{1,2,3}

¹De La Salle University, 2401 Taft Ave, Malate, Manila, 1004 Metro Manila, Philippines

²Western Michigan University, 1903 W. Michigan Ave, Kalamazoo, Michigan, United States

³University of Perpetual Help Delta-System, Alabang-Zapote Rd, Pamplona, 3 Las Pinas City, Philippines

Abstract. Coffee leaf rust (CLR), caused by *Hemileia vastatrix*, is a major constraint to global coffee production. This study screened for SH3-linked molecular markers associated with CLR resistance in the genomes of *Coffea canephora* DH200-94, *C. eugenoides* Bu-A, and *C. arabica* ET-39 using a custom R-based workflow (Johara pipeline). Ten molecular markers commonly linked to the SH3 resistance locus were analyzed via exact primer mapping, revealing varying patterns of marker presence across species. *C. arabica* ET-39 showed the highest number of matches, particularly in the Chromosome 3 regions, suggesting potential SH3-associated resistance, whereas *C. eugenoides* Bu-A had no detectable markers. This work highlights the utility of computational pipelines for rapid genome-scale detection of resistance-associated loci and provides insights for future breeding programs aimed at enhancing CLR resistance in coffee cultivars.

Keywords: *Coffea* cultivar; genome; SH3; coffee leaf rust; R software

1. Introduction

Globally, *Coffea* species are among the most economically important crops, forming the basis of coffee beverages consumed worldwide. The most widely cultivated species include *Coffea arabica* and its progenitors, *C. canephora* and *C. eugenoides* [1]. *C. arabica* accounts for approximately 65% of global coffee production due to its favorable aroma and flavor, while *C. canephora* represents the remaining 35%, prized for its higher yield and increased resistance to coffee diseases [2,3]. Recently, *C. eugenoides* gained attention following its

* Corresponding author: raissa_adellaide_sison@dlsu.edu.ph

success in the 2021 World Barista Championship, leading to a surge in demand and efforts to cultivate this species commercially [4].

Coffee production is continuously threatened by multiple biotic factors, notably coffee leaf rust (CLR), caused by the biotrophic fungus *Hemileia vastatrix*. CLR infection results in the formation of orange-yellow powdery pustules on the undersides of leaves, leading to premature leaf drop and substantial yield losses, sometimes reaching up to 50% [5]. While fungicides are available for disease management, durable resistance is often best achieved through the use of naturally occurring genetic resistance. Among the nine known SH resistance loci (SH1–SH9), the S_H3 locus, originally derived from *Coffea liberica*, is recognized for conferring strong resistance against CLR [6]. Introgression of the S_H3 locus into commercially important coffee species has become a key strategy in breeding programs aimed at improving resistance, with cultivars exhibiting genetic variations that reflect their differential capacity for CLR resistance.

A previous *in silico* analysis detected S_H3 -linked molecular markers in *Coffea* genomes, supporting the localization of the S_H3 resistance region on chromosome 3 [9]. However, this workflow required multiple bioinformatics steps and lacked an automated framework. To address these limitations, the present study introduces the Johara pipeline, an integrated R-based approach for genome-wide detection and visualization of SH3-associated markers.

Due to the limited availability of fully characterized S_H3 sequences, molecular markers linked to this locus—such as BA-42-21B-r, BA-48-21O-f, BA-124-12K-f, Sat160, Sat244, Sat281, SP-M5-SH3, SP-M8-SH3, SP-M16-SH3, and SP-M18-SH3—are commonly used to detect its presence in coffee genomes [7]. Using the Johara pipeline, this study aimed to: (1) assess the presence of S_H3 -linked markers in *C. canephora* DH200-94, *C. eugenioides* Bu-A, and *C. arabica* ET-39; (2) evaluate the *in silico* annealing of these markers within the genome; and (3) compare the prevalence and distribution of S_H3 -associated regions across species. The findings provide insights into the genetic basis of CLR resistance in coffee and support the validation of markers for breeding programs targeting CLR-resistant cultivars.

2. Methodology

Figure 1 illustrates the workflow used in this study to detect S_H3 -linked molecular markers in the genomes of *C. canephora* DH200-94, *C. eugenioides* Bu-A, and *C. arabica* ET-39 cultivars using the Johara pipeline implemented in R, with the source code available on GitHub (<https://github.com/raissasison/Johara>). The pipeline integrates multiple R packages—rentrez, adegenet, ape, ggtree, ggplot2, stats, ips, msa, Biostrings, seqinr, and primerTree—to streamline the detection and visualization of S_H3 -linked molecular markers.

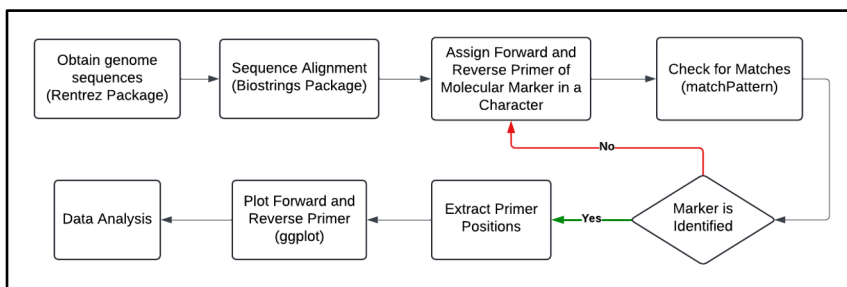


Fig 1. Research Workflow

The Johara pipeline aimed to simplify and enhance the efficiency and reproducibility of gene detection. This pipeline incorporates all the necessary packages within one script in R, making it direct to the point compared with other complex pipelines that need to use multiple software. Furthermore, this pipeline is not only limited to detecting genes in *Coffea*, as it can

also be applied to other species with available genome sequence and molecular marker sequences.

2.1 Data Collection

Nucleotide sequences were obtained from GenBank of the National Center for Biotechnology Information (NCBI) using the `rentrez::entrez_fetch` function. Keywords used to search for the cultivars were “Genome,” “*Coffea*,” “*Canephora*,” “*Arabica*,” “*Eugenioides*,” and “Cultivar.” *C. canephora* DH200-94 (Accession IDs: CM071913.1 to CM071923.1), *C. arabica* ET-39 (Accession IDs: CM072026.1 to CM072047.1), and *C. eugenioides* Bu-A (Accession IDs: CM071880.1 to CM071890.1) with genome sizes of 672.4 Mb, 1.2 Gb, and 644 Mb, respectively, were selected. All sequences of Chromosome 1 to Chromosome 11 of each cultivar were retrieved in FASTA format. Chromosome numbers labelled with “c” and “e” correspond to the *C. canephora* and *C. eugenioides* subgenome of *C. arabica*, respectively.

2.2 Data Processing and Alignment

The acquired sequences were processed using the `Biostrings` package to standardize the data by eliminating extraneous information, such as headers and line breaks for downstream processing. The resulting clean sequences were stored in character vectors for subsequent analyses.

2.3 Identification of S_H3-Linked Molecular Markers

After processing the *Coffea* cultivars sequences, the S_H3-linked molecular markers were identified using the `matchPattern` function that used the S_H3 marker sequences (see Table 1) as pattern to screen for an exact match (`min.mismatch = 0`; `with.indels = FALSE`; `fixed = TRUE`) in the *Coffea* cultivars genomes. Both forward and reverse primer matches were detected and were visualized through plots generated with the `ggplot2` package.

Table 1. Markers for S_H3 detection derived from Mahé et al. [7]

Molecular Marker	Sequence
BA-42-21B-r	Forward: CACACACAGCCTAAGCATCAA Reverse: GGATTGACTCGACTCACCAA
BA-48-21O-f	Forward: ACAGTGAATTCCCCAAGCAC Reverse: ACTTGGCAGGCGTAATTGAA
BA-124-12K-f	Forward: TGATTTTCGCTTGTGTGTCGAG Reverse: TGCAGATTGATGGCACGTTA
Sat160	Forward: TGCTTAGGCACTTGATATAGGA Reverse: CACGTGCAAGTCACATACTTTA

Cont. Table 1. Markers for S_H3 detection derived from Mahé et al. [7]

Molecular Marker	Sequence
Sat244	Forward: GCATGTGCTTTTTGATGTCGT Reverse: GCATACTAAGGAAATTATCTGACTGCT
Sat281	Forward: TCTTCGTCTTTGCTATTGGT Reverse: TATTAACGTCCATCCACACA
SP-M5-SH3	Forward: TTCACGATCCAAGAAGCA Reverse: AGCATGCATTGTAGAAAAA
SP-M8-SH3	Forward: GAATTCAGCGACGATTG Reverse: GATTTGGTGGGAAGGGAGC
SP-M16-SH3	Forward: TTAAGTGGAACTTGGCTTG Reverse: ATCTAGCTTTGGAACATCGT
SP-M18-SH3	Forward: CTATTTGGTGTGGGAAGTAAC Reverse: CTACATCCACGGAGAGAAAC

2.4 Data Analysis

A comparative study was performed to determine the prevalence of S_{H3} -linked molecular markers across different cultivars through the alignment results visualized with the generated plots in R software. The analysis looks into the genetic variations and potential resistance within *C. canephora*, *C. arabica*, and *C. eugenioides* populations.

3. Results and Discussion

To further understand the different genetic factors underlying resistance to *H. vastatrix*, S_{H3} -linked molecular markers in *C. canephora*, *C. arabica*, and *C. eugenioides* cultivars were identified within their whole genomes. The S_{H3} region has been mapped in the Chr 3 of CLR-resistant coffee species, including *C. liberica*, *C. canephora*, and *C. arabica var Caturra* [8]. The discovery of the inherent S_{H3} resistance genes of *C. liberica* led to its use in selective breeding with other susceptible coffee species in hopes of introducing the S_{H3} trait into their offspring (see Figure 2). Consequently, various cultivars of commercially demanded coffee species, such as *C. canephora*, *C. arabica*, and *C. eugenioides*, were produced. Furthermore, the S_{H3} region is highly conserved, suggesting that CLR-resistant coffee species should also possess the genetic markers being identified by primer pairs within their genome, most likely in their Chr 3. A bioinformatics tool like R software made it possible to successfully identify the S_{H3} region using established primer pairs *in silico*—indicating the possibility of the cultivars expressing resistance phenotypes against CLR. This method allows for a simpler and more efficient way of identifying resistance genes through the use of a free program that allows pipelines to be saved in scripts for convenient programming. With the Johara pipeline, errors that may occur can be easily identified, and resistance genes can be identified faster.

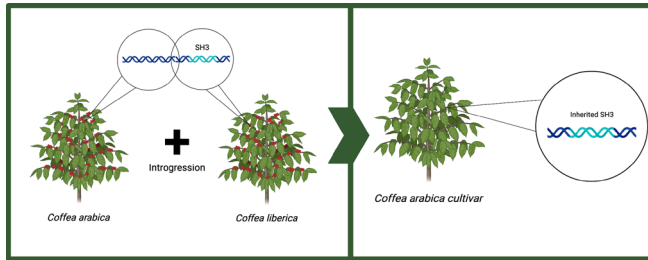


Fig 2. Process of Introgression (Adapted from Po et al. [9])

No markers were detected across the whole genome of the *C. eugenioides* Bu-A cultivar. In contrast, the SP-M18-SH3 primer pair identified two matches in Chr 1 of *C. canephora* DH200-94. Sat160 was also detected in the Chr 3 of *C. canephora* DH200-94. Meanwhile, SP-M18-SH3 also showed one match in both Chr 1c and 10c of *C. arabica* ET-39 cultivar. Moreover, three (3) primer pairs—BA-48-21O-f, Sat160, and Sat281—were found in Chr 3c of *C. arabica* ET 39, while two (2) primer pairs—Sat160 and Sat281—were observed in Chr 3e.

The Johara pipeline in this study showed novel findings in identifying specific markers within various *Coffea* cultivars. In comparing the results of this study with those of others, Table 2 reveals that similar locations and product lengths were detected for some markers. Moreover, the SP-M18-SH3 marker in Chr 1 of *C. canephora* DH200-94, Chr 1c and 10c in *C. arabica* ET-39, and Sat281 in Chr 3e of *C. arabica* ET-39 were detected. These results provide a new set of perspectives on the genomic composition of these *Coffea* cultivars, emphasizing the identification of the SP-M18-SH3 and Sat281 markers in chromosomal regions that had not been documented before.

Table 2. Comparison of Detected Molecular Markers in *Coffea* Genome with Other Studies

<i>Coffea</i> Species	Chr	Marker	Location	Product Length (bp)	References
<i>C. eugenioides</i>	3	BA-48-21O-f	3,244,209 - 3,244,509	301	[10]
		BA-124-12K-f	4,472,200 - 4,472,586	387	
		Sat244	4,494,379 - 4,494,662	284	
		SP-M8-SH3	4,512,389 - 4,512,623	235	
		Sat160	9,337,610 - 9,337,781	172	
			9,360,772 - 9,360,900	129	
		BA-42-21B-r	13,657,499 - 13,657,884	386	

		13,698,643 - 13,699,034	392	
5	BA-42-21B-r	9,034,751 - 9,035,136	386	[10]

Note: * = Molecular Markers Detected in this Study using Johara Pipeline; NA = No specified position

Cont. Table 2. Comparison of Detected Molecular Markers in *Coffea* Genome with Other Studies

<i>Coffea</i> Species	Chr	Marker	Location	Product Length (bp)	References	
<i>C. canephora</i>	1	SP-M18-SH3*	1,078,517 - 1,078,887	371	This study	
			46,946,193 - 46,946,562	370		
		Sat160*	8,457,825 - 8,457,957			This study
			8,384,597 - 8,384,729	133	[11, 10]	
			NA			[8]
			2,709,093 - 2,709,393	301	[10]	
		BA-48-21O-f	3,022,386 - 3,022,700			[11, 10]
			NA	315	[8]	
	3		3,794,524 - 3,794,926			[11]
			NA	403	[8]	
		BA-124-12K-f	3,803,622 - 3,804,022			[11]
			NA	401	[8]	
		3,810,340 - 3,810,766			[11]	
		NA	427	[8]		

			3,816,628 - 3,816,857	230	[10]
		Sat244	3,843,980 - 3,844,281	302	[11, 10]
			NA		[8]
		SP-M8-SH3	3,862,026 - 3,862,267	242	[11, 10]
			NA	242	[8]
<i>C. canephora</i>	3	SP-M5-SH3	9,462,739 - 9,462,898	160	[11, 10]
			NA		[8]
		Sat281	11,972,224 - 11,972,296	73	[11,10]
			NA		[8]
		BA-42-21B-r	12,285,408 - 12,285,785	378	[11, 10]
			NA		[8]

Note: * = Molecular Markers Detected in this Study using Johara Pipeline; NA = No specified position

Cont. Table 2. Comparison of Detected Molecular Markers in *Coffea* Genome with Other Studies

<i>Coffea</i> Species	Chr	Marker	Location	Product Length (bp)	References
<i>C. canephora</i>	5	Sat281	14,529,519 - 14,529,845	327	[11]
	10	SP-M5-SH3	18,467,325 - 18,468,257	933	[11]
	1c	SP-M18-SH3*	18,729,259 - 18,729,628	370	This study
<i>C. arabica</i>	1e	SP-M16-SH3	23,679,794 - 23,680,036	243	[11]
	2c	SP-M16-SH3	47,262,863 - 47,263,124	262	[11]

		3,124,592 - 3,124,906		This study
	BA-48-21O-f*	3,064,088 - 3,064,402	315	[11, 10]
		NA		[8]
		7,938,085 - 7,938,215	131	This study
3c	Sat160*	8,169,877 - 8,170,009	133	[11, 10]
		NA		[8]
		12,250,230 - 12,250,310		[11,10]
	Sat281*	NA	81	[8]
		10,890,805 - 10,890,899	95	This study
		10,836,483 - 10,836,577		[11, 10]
	Sat281	NA	95	[8]
		12,508,841 - 12,509,214		[11, 10]
3c		NA	374	[8]
	BA-42-21B-r	12,571,165 - 12,571,542		[11, 10]
<i>C. arabica</i>		NA	378	[8]
		8,420,170 - 8,420,341		This study
		7,136,657 - 7,136,828	172	[11, 10]
3e	Sat160*	NA		[8]
		7,151,565 - 7,151,667	103	[11, 10]

NA [8]

Note: * = Molecular Markers Detected in this Study using Johara Pipeline; NA = No specified position

Cont. Table 2. Comparison of Detected Molecular Markers in *Coffea* Genome with Other Studies

<i>Coffea</i> Species	Chr	Marker	Location	Product Length (bp)	References
<i>C. arabica</i>		Sat281*	11,086,868 - 11,086,948	81	This study
			3,244,209 - 3,244,509		[11]
		BA-48-21O-f	NA	301	[8]
	3e		10,560,828 - 10,561,219		[11, 10]
			NA	392	[8]
		BA-42-21B-r	10,654,277 - 10,654,680		[11, 10]
			NA	404	[8]
	7c	SP-M8-SH3	270,356 - 271,132	777	[11]
	7e	SP-M5-SH3	1,995,213 - 1,995,838	626	[11]
	10c	SP-M18-SH3*	7,469,045 - 7,469,404	360	This study

Note: * = Molecular Markers Detected in this Study using Johara Pipeline; NA = No specified position

3.1 Detection of S_{H3} Region in *C. eugenoides* cultivar Bu-A

A previous study [12] identified S_{H3} within *C. eugenoides* Chr 3, having four gene variations, along with a highly conserved S_{H3} gene variant in Chr 6. This could explain *C. eugenoides*' capacity to survive against CLR. However, the ten chosen primers in this study could not bind to any site within the *C. eugenoides* cultivar Bu-A genome (see Table 2), suggesting that this specific cultivar lacks an S_{H3} region.

The Johara pipeline did not identify any primers in *C. eugenoides* despite it being the maternal progenitor of *C. arabica*. This finding is contradictory to other studies that have identified S_{H3} in the *C. arabica* genome (see Table 2). To support this finding, a possible

explanation lies in the asymmetrical contributions of the two subgenomes present in the allotetraploid genome of *C. arabica* (subgenome *C. canephora*—CaCc and subgenome *C. eugenoides*—CaCe) [13]. Research suggests that these subgenomes display distinct expression patterns. Specifically, the CaCc subgenome is more actively involved in regulating protein expression, whereas the CaCe is most likely associated with basal processes. This implies that the genetic markers found in *C. arabica* are influenced by contributions from both of its maternal progenitors—rather than being exclusively derived from one.

The absence of S_{H3} markers in *C. eugenoides* is unexpected, given the presence of a highly conserved S_{H3} gene variant on Chr 6 of *C. eugenoides* [12]. This discrepancy may be attributed to several factors, like the genetic variability among the cultivars being tested. Moreover, *C. eugenoides* has traditionally been regarded as rare or experimental in coffee farming until recently [4]. Consequently, it has likely interacted less with other *Coffea* species, which suggests that *C. eugenoides* cultivars have had fewer opportunities to be exposed to CLR-resistant species, potentially reducing their likelihood of possessing genes such as S_{H3} associated with CLR resistance. Additionally, the genetic markers used in this study might not have been optimized for detecting S_{H3} genes in the specific genomic region or cultivar, or they may have limited detection sensitivity. The lack of detectable S_{H3} markers in *C. eugenoides* could be a result of a mismatch in terms of marker specificity. With that being said, further research is warranted to develop molecular markers that are tailored to specific chromosome regions and cultivars.

3.2 Detection of S_{H3} Region in *C. canephora* cultivar DH200-94

For *C. canephora* cultivar DH200-94, two of the primers successfully aligned with the genome (see Table 2), indicating the potential of CLR resistance in the cultivar. The SP-M18-SH3 primers were able to successfully bind to two regions within Chr 1 of *C. canephora* DH200-94, having product lengths of 371 and 370 bp and matches around 1 Mbp and 46 Mbp, respectively (see Figure 3A & 3B). Moreover, the Sat160 primer pair also matched within the Chr 3 genome of *C. canephora* DH200-94 at around 8Mbp, with a product length of 133 bp (see Figure 3C). As seen in Table 2, other studies have also reported the binding of primer pair Sat160 around 8Mbp of *C. canephora* Chr 3, resulting in the same product length. However, the lack of a significant number of matched markers from other studies suggests that certain S_{H3} regions are not found in the *C. canephora* DH200-94 cultivar.

It is significant that coffee plants exhibiting recombination events in the S_{H3} locus on the *C. canephora* chromosome segment that was introgressed from *C. liberica* were identified despite a reduction in overall frequency [6]. This indicates that the recombination rate within the S_{H3} region is significantly higher than the average recombination rate observed across the genome of *C. canephora*. The elevated recombination rate is located in the subtelomeric region of the Chr 1 of *C. canephora* [6]. These results support the identification of the SP-M18-SH3 marker in the Chr 1 region of *C. canephora* DH200-94, which aligns with the anticipated higher recombination rates associated with the subtelomeric location of S_{H3} .

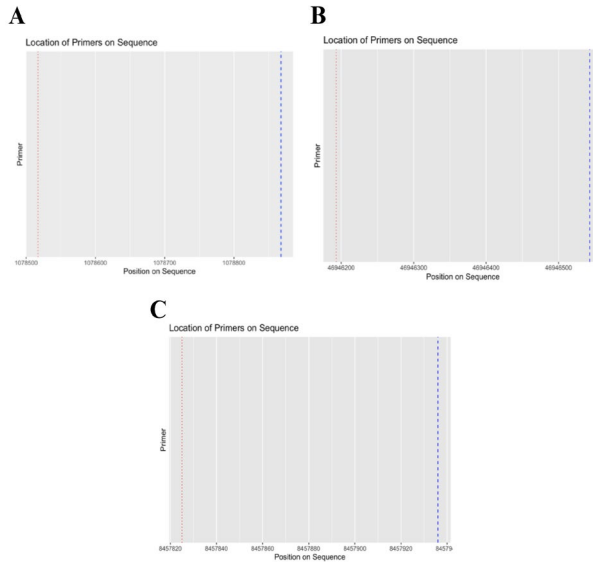


Fig 3. Locations of Primer Pair Matches on *C. canephora* cultivar DH200-94. A & B) SP-M18-SH3 on Chr 1; C) Sat160 on Chr 3. (Note: Dotted line = forward primer; Dash line = reverse primer)

3.3 Detection of S_{H3} Region in *C. arabica* cultivar ET-39

For the *C. arabica* cultivar ET-39, a total of seven primer matches were detected within the genome. Specifically, there was one match for BA-48-21O-f and two matches each for the markers Sat160, Sat281, and SP-M18-SH3. The positions, product length, and which region of the whole genome was the marker positioned can be seen in Table 2.

As the most in-demand coffee species, *C. arabica* is notably more cultivated compared to other *Coffea* species. It has been reported that the S_{H3} gene from wildtype *C. liberica* has been introgressed into *C. arabica*, resulting in cultivars of *C. arabica* that exhibit resistance to CLR [14]. Consequently, the increased exposure and genetic variability in *C. arabica* could have enhanced its likelihood of harboring genes such as S_{H3} associated with CLR resistance.

3.3.1 Detection of S_{H3} Region in *C. arabica* cultivar ET-39 Chr 3c

Among the ten primer pairs used in this paper, the BA-48-21O-f, Sat160, and Sat281 successfully aligned within the Chr 3c (*C. canephora* subgenome) of *C. arabica* cultivar ET-39. The matches recorded using these primer pairs can be used as indicators to confirm the presence of S_{H3} region—most importantly, the possibility of the cultivar being resistant against CLR. The BA-48-21O-f primer pair was able to bind within the Chr 3c of *C. arabica* cultivar ET-39 yielding a product length of 315 bp (see Figure 4A). As summarized in Table 2, previous studies have also recorded the capacity of BA-48-21O-f primer pair to bind around 3 Mbp of the Chr 3c genome of *C. arabica* having the same product length.

Moreover, similar to the findings of other studies (see Table 2), the Sat160 primer has binding sites around 8Mbp on Chr 3c of *C. arabica* (see Figure 4B) and product lengths of 131-132 bp. Meanwhile, the annealed Sat281 primer has a product length of 95 bp as a result of

binding at around 10 Mbp (see Figure 4C), which is also consistent with previous studies (see Table 2).

The comparable binding sites and product lengths of these primer pairs in this study and previous ones [8, 10, 11] support the claim that the S_{H3} region is conserved among CLR-resistant coffee species. Additionally, the differences in terms of the positions of the binding sites of the primer within Chr 3c indicate that binding sites of different markers are distributed across Chr 3 of the coffee genome, implying that the S_{H3} is not a single gene but rather a region consisting of multiple genes spread throughout the Chr 3 [15]. Furthermore, the results of this study solidifies the findings of previous study [11] that the S_{H3} region spans from 3 Mbp to 16 Mbp of the Chr 3 of *Coffea* species.

The rest of the molecular markers used in this study did not return any matches within the Chr 3c of the *C. arabica* cultivar ET-39. However, in a previous study [11], BA-42-21B-r was able to bind to the Chr 3c of *C. arabica*—suggesting that this specific S_{H3} portion was not conserved in the *C. arabica* cultivar ET-39. Moreover, shown in Figure 4D is what the plot looks like when a primer pair, like SP-M16-SH3, is not located within the region of interest. Consistent with the present results, previous *in silico* analyses (see Table 2) also reported that the SP-M16-SH3 was not able to bind to the Chr 3c of *C. arabica*.

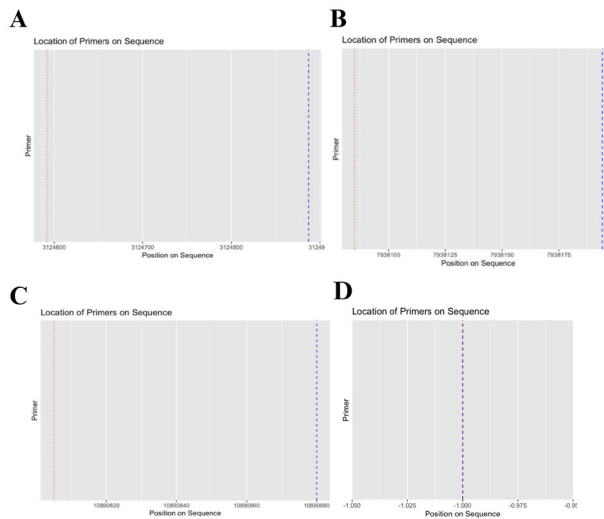


Fig 4. Location of Primer Pair Matches on *C. arabica* cultivar Et-39 Chr 3c. A) BA-48210-f; B) Sat160; C) Sat281; D) SP-M16-SH3. (Note: Dotted line = forward primer; Dash line = reverse primer)

3.3.2 Detection of S_{H3} Region in *C. arabica* cultivar ET-39 Chr 1c and Chr 10c

The primer pair SP-M18-SH3 successfully aligned with the Chr 1c and Chr 10c regions of *C. arabica* ET-39. The novel matches recorded using the primer pair can be employed as markers to verify the existence of the S_{H3} region in Chr 1 and 10 of the ET-39's *C. canephora* subgenome, eventually leading to the confirmation of its contributions to the cultivar being CLR-resistant. Figure 5A shows the match of the SP-M18-SH3 primer pair in Chr 1c, found at around 18 Mbp, with a product length of 370 bp (see Table 2). Moreover, Figure 5B shows

the location of the SP-M18-SH3 primer on Chr 10c of *C. arabica* ET-39 at around 7 Mbp of Chr 10c, with a product length of 360 bp (see Table 2).

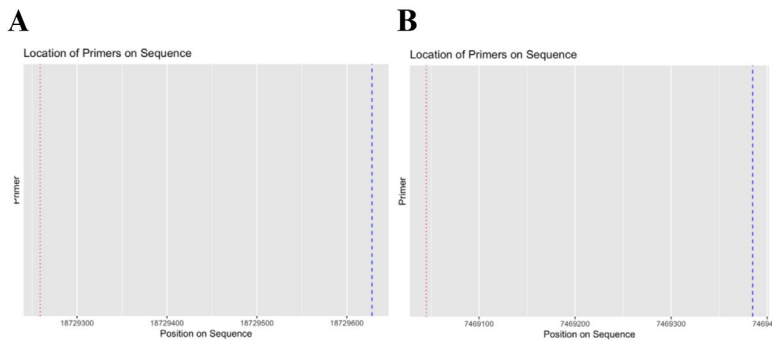


Fig 5. Location of Matched SP-M18-SH3 on *C. arabica* cultivar ET-39. A) on Chr 1c; B) on Chr10c. (Note: Dotted line = forward primer; Dash line = reverse primer)

3.3.3. Detection of S_{H3} Region in *C. arabica* cultivar ET-39 Chr 3e

From the ten primer pairs used in this study, Sat160 and Sat281 successfully aligned with Chr 3e (*C. eugenioides* subgenome) of the *C. arabica* ET-39 cultivar (see Table 2). The alignments of the primer pair confirm the presence of the S_{H3} region, suggesting the great potential of the cultivar to resist CLR. The Sat160 primer, which has a product length of 172 bp, is effectively bound to Chr 3e and is located at around 8 Mbp (see Figure 6A). Meanwhile, Sat281 was observed to be positioned at around 11 Mbp with a product length of 81 bp, which was similar to the findings in Chr 3c of *C. arabica* ET-39 cultivar (see Figure 6B).

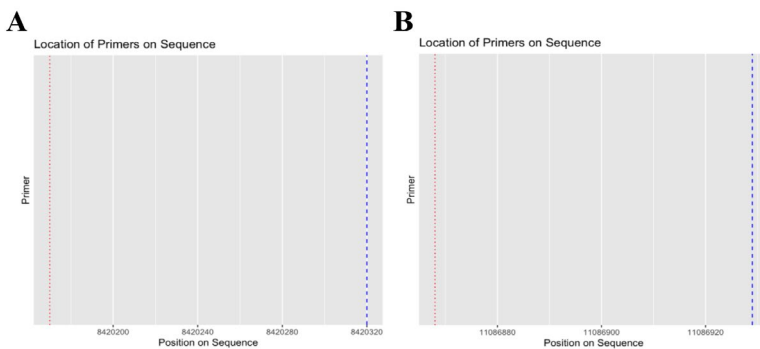


Fig 6. Location of Primer Pair Matches on *C. arabica* cultivar ET-39 Chr 3e. A) Sat160 B) Sat281. (Note: Dotted line = forward primer; Dash line = reverse primer)

As detected by the Johara pipeline, the greater presence of S_{H3} markers in the Chr 3c region of the *C. arabica* ET-39 cultivar compared to Chr 3e is linked to the genetic contributions of its subgenomes. Ingelbrecht et al. [2] highlighted that the distinct genetic differences between *C. canephora* and *C. eugenioides* are due to their hybridization. The *C. canephora* subgenome most likely contains a higher abundance of S_{H3} -related genes given the greater number of matched S_{H3} markers in Chr 3c. Evolutionary adaptations of *C. canephora* may

have led to the development of a stronger defense against pathogens like CLR, enhancing its resistance. In contrast, the Chr 3e region, inherited from *C. eugenoides*, underwent distinct selective pressures in its evolutionary history that brought about the possession of fewer S_{H3} markers, contributing to a lower potential for CLR resistance in that region of the genome.

4. Conclusions

The advancement of bioinformatics tools provides researchers with opportunities to analyze complex biological data efficiently. Using the Johara pipeline in R software, this study detected molecular markers associated with S_{H3} resistance genes across the genomes of *C. canephora*, *C. arabica*, and *C. eugenoides* cultivars. In *C. canephora* DH200-94, the SP-M18-SH3 primer pair aligned to two sites on Chr 1, while Sat160 aligned on Chr 3. In *C. arabica* ET-39, BA-48-21O-f, Sat160, and Sat281 aligned on Chr 3c, SP-M18-SH3 aligned on Chr 1c and Chr 10c, and Sat160 and Sat281 aligned on Chr 3e. No primer pairs successfully aligned with the genome of *C. eugenoides* Bu-A. Among the primers tested, Sat160 exhibited the highest frequency of matches across the cultivars, being present in all but *C. eugenoides* Bu-A. Overall, the Chr 3 region displayed the greatest concentration of marker matches, supporting previous reports that S_{H3}-associated genes are predominantly conserved in this chromosomal region.

Primer pairs that did not produce matches, including Sat244, SP-M16-SH3, SP-M8-SH3, SP-M5-SH3, BA-124-12K-f, and BA-42-21B-r, may reflect the absence of specific S_{H3} motifs in the analyzed cultivars or limitations in primer specificity and detection sensitivity. These results indicate potential genetic variation among the *Coffea* cultivars regarding the presence of S_{H3} resistance genes.

The number of primer matches across a genome is indicative of the likelihood of S_{H3} region presence. Accordingly, the detection of seven markers in *C. arabica* ET-39, three in *C. canephora* DH200-94, and none in *C. eugenoides* Bu-A suggests that S_{H3}-associated resistance genes are more likely to be present in *C. arabica*, consistent with its history of introgression from CLR-resistant species. These findings support the effectiveness of prior breeding efforts aimed at incorporating S_{H3}-mediated resistance into *C. arabica* cultivars and highlight their potential for expanded use in coffee production. Conversely, the absence of detectable S_{H3} markers in *C. eugenoides* underscores the need for further research on this emerging species, including the development of tailored molecular markers and evaluation of environmental and climatic factors that may influence CLR resistance.

References

1. E. Balunek, K. Chen, G. Volk, Case study: coffee wild species and cultivars – crop wild relatives and their use in plant breeding. Colorado State University Publishing – Open Educational Resources (2020). <https://colostate.pressbooks.pub/cropwildrelatives/chapter/case-study-coffee-wild-species-and-cultivars/>
2. N. Myhrvold, R. Coste, Coffee production | description, cultivation, process, hulling, & facts | Britannica (2023). [Www.britannica.com. https://www.britannica.com/topic/coffee-production](https://www.britannica.com/topic/coffee-production)
3. A. Simon-Gruita, M. D. Pojoga, N. Constantin, G. Duta-Cornescu, Genetic engineering in coffee. Caffeinated and Cocoa Based Beverages, 447–488 (2019) <https://doi.org/10.1016/b978-0-12-815864-7.00014-3>
4. J. Walbank, *Eugenoides* coffee: a roaster's guide to this unique species. Mtpak.coffee. (2022).
5. P. Talhinhos, D. Batista, I. Diniz, A. Vieira, D. N. Silva, A. Loureiro, S. Tavares, A. P. Pereira, H. G. Azinheira, L. Guerra-Guimarães, V. Várzea, M. do C. Silva, The

- coffee leaf rust pathogen *Hemileia vastatrix*: one and a half centuries around the tropics. *Mol. Plant Pathol.*, **18**(8), 1039–1051 (2017). <https://doi.org/10.1111/mpp.12512>
6. P. Lashermes, M.-C. Combes, A. F. Ribas, A. Cenci, L. Mahé, H. Etienne. Genetic and physical mapping of the SH3 region that confers resistance to leaf rust in coffee tree (*Coffea arabica* L.). *Tree Genetics & Genomes*, **6**(6), 973–980 (2010). <https://doi.org/10.1007/s11295-010-0306-x>
 7. L. Mahé, M.C. Combes, V. M. P. Várzea, C. Guilhaumon, P. Lashermes, Development of sequence characterized DNA markers linked to leaf rust (*Hemileia vastatrix*) resistance in coffee (*Coffea arabica* L.). *Mol. Breeding*, **21**(1), 105–113 (2007). <https://doi.org/10.1007/s11032-007-9112-z>
 8. N. R. S. Santos, M. B. Magat, M. V. Mondragon, E. P. Cao, D. M. C. Santos, Genetic profiling of locally registered Philippine coffee using molecular markers linked to resistance against diseases and pests. *Biodiversitas J. Biol. Divers.*, **24**(7) (2023). <https://doi.org/10.13057/biodiv/d240752>
 9. J. X. Po, H. N. Portento, R. A. Sison, J. Santiago, Detecting the presence of SH3 resistance genes in *Coffea* spp. cultivars using genetic markers, in Proceedings of the DLSU Research Congress 2024, Manila, Philippines, June 20–22 (2024)
 10. T.F. Nagaño, N. R. Santos, D. M. Santos, E. Cao, In silico physical mapping of resistance and resistance-associated genes in the SH3 region of the *Coffea canephora* genome. *Philipp. J. Sci.*, **151**(1) (2021). <https://doi.org/10.56899/151.01.08>
 11. M. L. Merlin, In silico screening of the SH3 resistance locus in *Coffea canephora* and *Coffea arabica* for candidate genes involved in coffee leaf rust resistance. Biology Bachelor's Theses, De La Salle University (2023)
 12. P. C. da S. Angelo, G. H. Sera, L. H. Shigueoka, E. T. Caixeta, Rust resistance SH3 loci in *Coffea* spp. *PMPP*, **127**, 102–111 (2023). <https://doi.org/10.1016/j.pmpp.2023.102111>
 13. O. R. Vidal, J. M. Mondego, D. Pot, A. B. Ambrósio, A. C. Andrade, L. F. Pereira, C. A. Colombo, L. G. Vieira, M. F. Carazzolle, G. A. Pereira, A high-throughput data mining of single nucleotide polymorphisms in *Coffea* Species expressed sequence tags suggests differential homeologous gene expression in the allotetraploid *Coffea arabica*. *Plant Physiol.* **154**(3), 1053–1066 (2010). <https://doi.org/10.1104/pp.110.162438>
 14. E. R. Alkimim, E. T. Caixeta, T. V. Sousa, A. A. Pereira, A. C. B. de Oliveira, L. Zambolim, N. S. Sakiyama, Marker-assisted selection provides arabica coffee with genes from other *Coffea* species targeting multiple resistance to rust and coffee berry disease. *Mol. Breeding*, **37**(1) (2017). <https://doi.org/10.1007/s11032-016-0609-1>
 15. L. Cui, K. Hanika, R. G. F. Visser, Y. Bai Improving pathogen resistance by exploiting plant susceptibility genes in Coffee (*Coffea* spp.). *Agron*, **10**(12), 19–28 (2020). <https://doi.org/10.3390/agronomy10121928>