Physical Mapping of Candidate Resistance and Resistance-associated Genes in the Sat229 region in Chromosome 5 of the *Coffea canephora* Genome

Mara Angela R. Victoria
Terrence Ferdinand S. Nagano
Nick Rainier S. Santos*
Ernelea P. Cao
Daisy May C. Santos
Institute of Biology
University of the Philippines Diliman

ABSTRACT

Diseases and pests hamper Philippine coffee production. Identification of defense-associated genes, which are present in certain plants, can be used to identify which resistant coffee cultivars to propagate. This study aimed to identify *Coffea canephora* candidate resistance (*R*) and resistance-associated genes within a region (19-23 Mb) of chromosome 5, encompassing the Sat229 SSR marker. After using “BioMart” to retrieve data from Ensembl Plants, genes were screened based on their Gene Ontology (GO) annotations. Their predicted protein products were identified using BLASTp. For *R* genes, architecture was further evaluated using hmm scan, PCoils, and LRR search while clustering was analyzed manually. The genes were mapped using the R package “chromoMap”. The results reveal 23 putative *R* genes that encode NLR proteins detecting pathogenic effectors, and 52 resistance-associated genes. Notable ones were predicted to encode an isoform of diacylglycerol kinase 5 (DGK5), programmed cell death-associated proteins, transcription factors containing AP2/ERF domain, two potential novel proteins with C-terminal leucine-rich repeat (LRR) domains involved in defense against fungi, and isoprene synthase. Thus, this study was able to screen the Sat229 region of chromosome 5 of the *C. canephora* genome and identify candidate defense-associated gene loci, which can be used as reference for marker-assisted breeding programs.

Keywords: Coffea canephora, disease, pests, resistance genes, Sat 229

INTRODUCTION

The Philippines is a major consumer of coffee (*Coffea* spp.) but local production is unable to meet its high demand. The country was reported to be only 15% self-sufficient in crop production (Department of Agriculture 2022). One of the major concerns of increasing production is the damaging effects of diseases and pests (Department of Agriculture and Department of Trade and Industry 2019) caused by the coffee leaf rust (CLR) fungus, *Hemileia vastatrix*, and the coffee berry borer (CBB), *Hypothenemus hampei* (Bureau of Plant Industry [date unknown]). Coffee plants could also be affected by other diseases like the coffee berry disease (CBD) caused by the fungus, *Colletotrichum kahawae*, and other pests like the root-knot nematode (RKN), *Meloidogyne* spp.

To minimize infection and infestation, the cost-effective and ecologically appropriate way is to breed resistant coffee cultivars with high yield, high beverage quality, and adaptive to wide climates (van der Vossen 2009). This can be achieved by introgressing resistance genes from resistant cultivars into susceptible cultivars known for other desirable traits (Gichuru et al. 2008). For genes conferring resistance against CLR, the *C. arabica* X *C. canephora* interspecies hybrid, Hibrido de Timor (HDT) has been a source of the S_H6, S_H7, S_H8, and S_H9 genes. Meanwhile, the S_H1, S_H2, S_H4, and S_H5 genes were identified in resistant Ethiopian *C. arabica* varieties. Lastly, the S_H3 gene was derived from *C. liberica* (Noronha-Wagner and Bettencourt 1967; Bettencourt and Noronha-Wagner 1971; Bettencourt et al. 1980). For genes conferring resistance against CBD, the *T* gene is also present in HDT. There are also the dominant *R* gene and the recessive *k* gene, both identified in *C. arabica* var. Rume Sudan. The latter gene was also found independently in *C. arabica* var. K7 (van der Vossen and Walyaro 1980; van der Vossen and Walyaro 2009).

Cui et al. (2020) mapped the locations of genetic markers linked to disease resistance in the *C. canephora* genome. Using the map as the basis, Yu et al. (2021) further studied the *Ck-1* region in chromosome 1. Using the predicted protein products as basis, candidate resistance genes, which may correspond to the putative *Ck-1* gene, were identified. In 2022, Nagaño et al. performed a similar study on the S_H3 region in chromosome 3. However, the identified candidate resistance genes still need to be evaluated using *in vivo* methods in order to determine how they participate in conferring disease resistance.

The majority of putative coffee resistance genes are not yet characterized properly. In addition, there might be major resistance genes that have yet to be discovered. Thus, the biological mechanism underlying the resistance of coffee is still not fully understood as specific functions of resistance genes are not yet identified. Since there is still a lack of knowledge on the function and characterization of resistance-linked genes, breeding more durable, resistant cultivars is still a subject for further studies.

Herrera et al. (2009) identified the locus Sat229 to be associated with CLR resistance upon observing its variable amplification in the second filial generation (F_2), segregating population from the cross between the susceptible male *C. arabica* cultivar Caturra and the *C. arabica* X *C. canephora* hybrid as the resistant female parent. The genetic marker Sat229 was determined to be located in chromosome 5 of the *C. canephora* genome (Cui et al. 2020) and has been the only resistance-associated marker identified in the said chromosome at the time of this study. Thus, this study aimed to identify potential candidates for resistance and resistance-associated genes in the Sat229 region in chromosome 5 of the *C. canephora* genome and construct a physical map showing their location in *C. canephora* chromosome 5.

MATERIALS AND METHODS

Localizing the Genetic Marker

Based on the works of Cui et al. (2020) and Yu et al. (2021), the primers for the CLR resistance-linked SSR marker Sat229 produced a 154 base pair-long product. After running these primers (F: TTCTAAGTTGTTAAACGAGACGCTTA, R: TTCCTCCATGCCCATATTG) through PrimerBlast, the probable location of resistance markers was estimated to be where the forward and reverse primers are attached to, i.e., from 21,184,265 to 21,184,418 bp of chromosome 5 of the *C. canephora* genome. Brodie et al. (2016) reported that the average distance between the causative gene and the linked single nucleotide polymorphism (SNP) marker is 2 Mbp. Extrapolating the data from their study, the selected region for this study was set from 19 Mbp to 23 Mbp of chromosome 5.

Gathering Candidates for Resistance and Resistance-associated Genes

The protein-coding genes within this region were determined using the “BioMart” function (https://plants.ensembl.org/biomart/martview/) of the Ensembl Plants website. The latest “Ensembl Plants Genes 56” database and the *C. canephora* genes (AUK_PRJEB4211_v1) dataset were used. For the filters, the selected region was set to chromosome 5 with coordinates starting from 19 Mbp and ending at 23 Mbp, while GO term accession (GO:0008150 for ‘biological process’) was selected. Information retrieved were “Gene Stable ID”, “Protein Stable ID”, “Gene start (bp)”, “Gene end (bp)”, "GO term accession", "GO term name", "GO term definition", "GO domain", "Interpro ID", "Interpro Short Description", "Interpro Description", and "Peptide" sequence (Yu et al. 2021; Nagaño et al. 2022).

The protein sequence of the resulting shortlist of candidate resistance and resistance-associated genes was run in the Protein Basic Local Alignment Search Tool (BLASTp) (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) using the "non-redundant protein sequences (nr)" database. The metadata of the topmost hits were retrieved, specifically their "Description", "Scientific Name", "Max Score", "Total Score", "Query Cover", "E value", and "Percent Identity" (Yu et al. 2021; Nagaño et al. 2022).

Then, the list of potential resistance genes was refined based on the GO annotations of their predicted encoded proteins and identified functions. Afterward, the filtered genes were classified according to the mechanisms of disease resistance of plants as described by Andersen et al. (2018). The major groups considered were pathogen detection, signal transduction, and defense response.

Evaluating the Architecture and Clustering of Candidate Resistance Genes

The architecture of the identified nucleotide binding-domain leucine-rich repeats (NLRs) from the results of the 'BioMart' function in Ensembl Plants was evaluated by analyzing the protein domains using hmmscan in HMMER 3.3.2 (https://www.ebi.ac.uk/Tools/hmmer/) (Potter et al. 2018; Liu et al. 2021) for visualization. Further evaluation of the presence of coiled-coil (CC) domains was conducted using Pcoils (http://toolkit.tuebingen.mpg.de/pcoils) (Gabler et al. 2020; Gruber et al. 2006; Zimmermann et al. 2017) at a threshold of 0.9 (Die et al. 2018). Meanwhile, the presence of leucine-rich repeat (LRR) motifs was also analyzed using the LRRsearch (http://lrrsearch.com), a webserver detecting the highly conserved segment (HCS) of LRR, with a default false discovery rate of three percent. This segment consisted of a stretch of 11 amino acid residues, i.e., LxxLxLxxNxL, where "L" is either leucine, isoleucine, valine, or phenylalanine, "N" can be asparagine, threonine, serine, or cysteine, and "x" for any amino acid (Bej et al. 2014).

The clustering of NLR genes was then assessed as either being a singleton, in tandem, or in a cluster. A gene cluster was defined as three or more copies of NLR genes having less than 200 kbp distance between two neighboring genes and with less than eight non-NLR genes in between the adjacent genes (van Wersch and Li 2019).

Mapping of the Candidate Resistance and Resistance-associated Genes

To visualize the location of the identified probable resistance and resistance-associated genes, a physical map of the *C. canephora* chromosome 5 was constructed using the R package "chromoMap" (Anand and Rodriguez Lopez 2022) in RStudio where the different categories were color-coded (blue for the genetic marker, red for *R* genes, pink for RLKs or receptor-like kinases, yellow for genes involved in signal transduction, and orange for genes involved in defense response) and the *R* genes in clusters were boxed.

RESULTS AND DISCUSSION

There were 75 identified candidate resistance and resistance-associated genes, and these were physically mapped in Figure 1, along with the SSR Sat229 marker. Each gene was labeled with its gene and protein stable IDs.

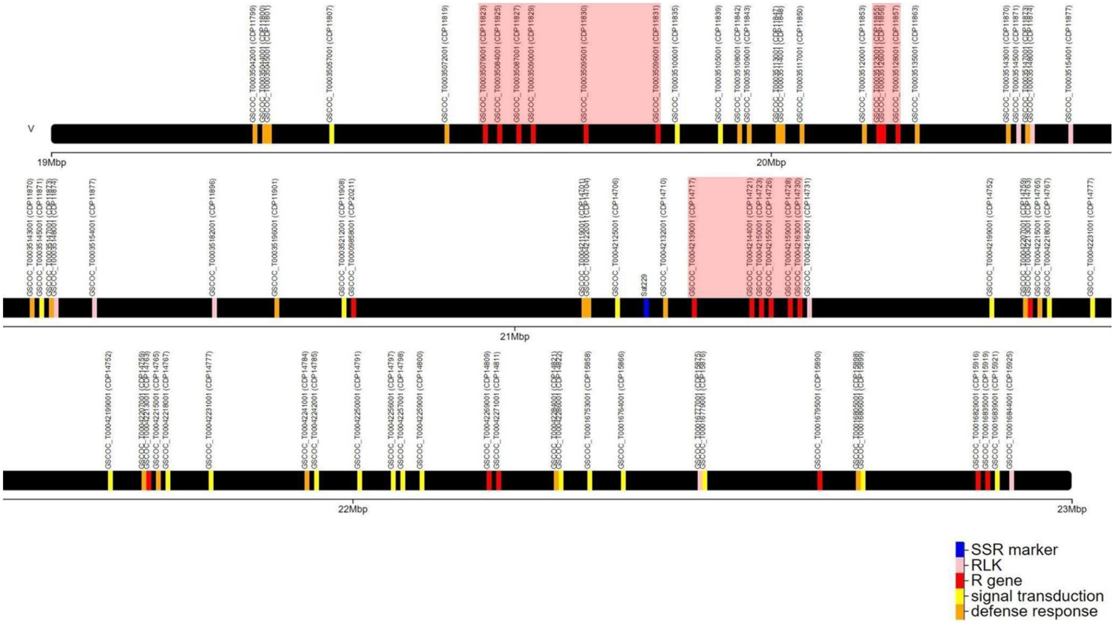

Figure 1. Condensed map of the putative *R* and resistance-associated genes predicted to be involved in *C. canephora* defense response. The physical map was constructed using the "chromoMap" function (Anand and Rodriguez Lopez, 2022) of RStudio. Clustered *R* genes for NLRs are shaded. Gene and protein IDs are based on data from the Ensembl Plants website. Identities of the protein products are listed in Tables 2–6.

To have a better grasp of these genes, they were categorized based on their selected relevant GO entries. The child terms were grouped under their parent terms based on the GO graphs from QuickGO (https://www.ebi.ac.uk/QuickGO/) to group together related genes based on their functions. Table 1 summarized the GO entries under the domain 'biological process' with potential relations to the defense response of plants in the region of interest—19 Mbp to 23 Mbp of *C. canephora* chromosome 5. It was observed that the most abundant genes, i.e., 31 had a 'defense response' (GO:0006952) as the selected related GO annotation. 'Defense response' was defined as biological reactions in order to prevent, mitigate, and recover from the damage that a foreign body or injury inflicted on the organism (Binns et al. 2009). This was followed by 'response to other organism' (GO:0051707), accounting for 14, which are cellular or organismal responses towards a stimulus from another living organism (Binns et al. 2009). Furthermore, 9 functioned in

Table 1. List of GO entries under the domain 'biological process' in the 19–23 Mbp of *C. canephora* chromosome 5 with plausible involvement in plant defense responses

GO term name	Count
defense response	31
response to other organism	14
programmed cell death	1
negative regulation of growth	1
Transport	1
signal transduction	5
Phosphorylation	7
protein dephosphorylation	1
protein ubiquitination	3
cellular response to salicylic acid stimulus	2
regulation of jasmonic acid biosynthetic process	1
DNA-templated transcription	9
epigenetic regulation of gene expression	1
histone H3-K36 methylation	1
cell wall organization or biogenesis	3
polysaccharide biosynthetic process	2
Detoxification	3
response to oxidative stress	4
hydrogen peroxide catabolic process	3
cellular response to DNA damage stimulus	1
RNA phosphodiester bond hydrolysis, endonucleolytic	2
Proteolysis	5
alkaloid metabolic process	7
diterpenoid biosynthetic process	2
phenylpropanoid biosynthetic process	1

'DNA-templated transcription' (GO:0006351). The last major category of genes accounting for 9 was involved in 'DNA-templated transcription' (GO:0006351). Some genes may be categorized under more than one GO annotation.

Out of the 31 identified genes under 'defense response', 23 probable *R* genes were identified as detailed in Table 2 which provided the summary of each gene's chromosome location, topmost BLASTP hit, and GO annotations. These *R* genes have NB-ARC protein domains and, thus, were identified as NLRs which are intracellular receptors that detect pathogen molecules and execute the death of infected cells (Jones et al. 2016).

Table 2. List of identified probable predicted resistance protein-encoding genes in the 19–23 Mbp region of *C. canephora* chromosome 5 based on the topmost BLASTP hit each with their selected related GO entries

Gene stable ID (Protein stable ID)	Chromosome location		Topmost BLASTP hit description	Selected relevant GO entries from the Ensembl Plants search
	Start (bp)	End (bp)		
GSCOC_T00035079001 (CDP11823)	19,601,676	19,604,015	Putative late blight resistance protein homolog R1B-17 [*Coffea eugenioides*] (Max Score=1001; Total Score=1001; Query Cover=94%; E Value=0; Percent Identity=99.39%)	defense response, plant-type hypersensitive response
GSCOC_T00035084001 (CDP11825)	19,624,370	19,628,107	Putative late blight resistance protein homolog R1A-3 [*Coffea arabica*] (Max Score=2178; Total Score=2178; Query Cover=87%; E Value=0; Percent Identity=97.43%)	defense response, plant-type hypersensitive response
GSCOC_T00035087001 (CDP11827)	19,647,177	19,650,164	Probable disease resistance protein At4g27220 [*Coffea arabica*] (Max Score=1671; Total Score=1671; Query Cover=99%; E Value=0; Percent Identity=84.82%)	defense response
GSCOC_T00035090001 (CDP11829)	19,667,891	19,671,658	Putative late blight resistance protein homolog R1A-3 [*Coffea arabica*] (Max Score=2561; Total Score=2561; Query Cover=99%; E Value=0; Percent Identity=99.2%)	defense response, plant-type hypersensitive response

Gene stable ID (Protein stable ID)	Chromosome location		Topmost BLASTP hit description	Selected relevant GO entries from the Ensembl Plants search
	Start (bp)	End (bp)		
GSCOC_T00035095001 (CDP11830)	19,744,457	19,748,721	Putative late blight resistance protein homolog R1A-3 [*Coffea arabica*] (Max Score=1632; Total Score=1632; Query Cover=99%; E Value=0; Percent Identity=90.53%)	defense response
GSCOC_T00035096001 (CDP11831)	19,840,651	19,845,377	Putative late blight resistance protein homolog R1B-14 [*Coffea arabica*] (Max Score=1614; Total Score=1614; Query Cover=93%; E Value=0; Percent Identity=89.3%)	defense response, response to other organism
GSCOC_T00035123001 (CDP11855)	20,149,955	20,154,210	Disease resistance RPP13-like protein 4 [*Coffea arabica*] (Max Score=1711; Total Score=1711; Query Cover=99%; E Value=0; Percent Identity=88.36%)	defense response, response to other organism
GSCOC_T00035126001 (CDP11856)	20,157,092	20,162,086	Disease resistance RPP13-like protein 4 [*Coffea arabica*] (Max Score=1796; Total Score=1796; Query Cover=99%; E Value=0; Percent Identity=99.77%)	defense response, defense response to other organism
GSCOC_T00035128001 (CDP11857)	20,177,334	20,182,779	Disease resistance RPP13-like protein 4 [*Coffea eugenioides*] (Max Score=1648; Total Score=1648; Query Cover=99%; E Value=0; Percent Identity=94.14%)	defense response, response to other organism
GSCOC_T00009858001* (CDP20211)	20,779,538	20,782,638	Putative late blight resistance protein homolog R1A-10 isoform X1 [*Coffea arabica*] (Max Score=1411; Total Score=1411; Query Cover=99%; E Value=0; Percent Identity=98.41%)	defense response
GSCOC_T00042139001* (CDP14717)	21,250,780	21,253,695	Disease resistance protein SUMM2-like isoform X1 [*Coffea arabica*] (Max Score=2009; Total Score=2009; Query Cover=99%; E Value=0; Percent Identity=100%)	defense response, defense response to other organism

Gene stable ID (Protein stable ID)	Chromosome location		Topmost BLASTP hit description	Selected relevant GO entries from the Ensembl Plants search
	Start (bp)	End (bp)		
GSCOC_T00042144001 (CDP14721)	21,327,820	21,330,276	Putative disease resistance protein RGA1 [*Coffea eugenioides*] (Max Score=1365; Total Score=1365; Query Cover=98%; E Value=0; Percent Identity=84.52%)	defense response, response to other organism
GSCOC_T00042150001 (CDP14723)	21,340,041	21,343,475	Putative disease resistance protein RGA3 [*Coffea eugenioides*] (Max Score=1694; Total Score=1694; Query Cover=99%; E Value=0; Percent Identity=78.01%)	defense response, response to other organism
GSCOC_T00042155001 (CDP14726)	21,353,709	21,356,673	Putative disease resistance protein RGA1 [*Coffea eugenioides*] (Max Score=1371; Total Score=1371; Query Cover=87%; E Value=0; Percent Identity=81.95%)	defense response, response to other organism
GSCOC_T00042159001 (CDP14728)	21,386,437	21,390,152	Putative disease resistance protein RGA3 [*Coffea arabica*] (Max Score=1718; Total Score=1718; Query Cover=99%; E Value=0; Percent Identity=78.77%)	defense response, response to other organism
GSCOC_T00042163001 (CDP14730)	21,399,854	21,401,768	Putative disease resistance protein RGA3 [*Coffea eugenioides*] (Max Score=906; Total Score=906; Query Cover=100%; E Value=0; Percent Identity=83.51%)	defense response, response to other organism
GSCOC_} T00042213001† (CDP14763)	21,717,343	21,718,097	Putative disease resistance protein RGA3 [*Coffea eugenioides*] (Max Score=402; Total Score=402; Query Cover=99%; E Value=6e-133; Percent Identity=91.78%)	defense response, response to other organism

Gene stable ID (Protein stable ID)	Chromosome location		Topmost BLASTP hit description	Selected relevant GO entries from the Ensembl Plants search
	Start (bp)	End (bp)		
GSCOC_T00042214001 (CDP14764)	21,718,185	21,721,950	Putative disease resistance protein RGA3 [*Coffea arabica*] (Max Score=1598; Total Score=1598; Query Cover=93%; E Value=0; Percent Identity=82.8%)	defense response, response to other organism
GSCOC_T00042269001 (CDP14809)	22,187,427	22,191,461	Putative disease resistance protein RGA1 [*Coffea eugenioides*] (Max Score=1597; Total Score=1597; Query Cover=99%; E Value=0; Percent Identity=79.44%)	defense response
GSCOC_T00042271001 (CDP14811)	22,200,718	22,204,134	Putative disease resistance protein RGA3 [*Coffea eugenioides*] (Max Score=2000; Total Score=2000; Query Cover=99%; E Value=0; Percent Identity=88.64%)	defense response, response to other organism
GSCOC_T00016795001 (CDP15890)	22,650,181	22,653,450	Putative disease resistance protein RGA3 [*Coffea arabica*] (Max Score=1640; Total Score=1640; Query Cover=99%; E Value=0; Percent Identity=75.63%)	defense response, response to other organism
GSCOC_T00016829001 (CDP15916)	22,866,806	22,870,079	Putative disease resistance protein RGA3 [*Coffea arabica*] (Max Score=1632; Total Score=1689; Query Cover=99%; E Value=0; Percent Identity=86.05%)	defense response, response to other organism
GSCOC_T00016835001 (CDP15919)	22,884,410	22,888,404	Putative disease resistance protein RGA3 [*Coffea arabica*] (Max Score=1724; Total Score=1724; Query Cover=99%; E Value=0; Percent Identity=79.22%)	defense response, response to other organism

* Putative *NLR* genes predicted to lack CC domain: NBS-LRR (NL)

† Putative *NLR* genes predicted to lack LRR domain: CC-NBS (CN)

Further analysis of the architecture of the putative *R* genes was carried out using hmmscan, with the sequence structure of each gene visualized. The positions of the NB-ARC domain, as well as the LRR and CC domains, were marked if present. Results from Ensembl Plants, PCoils analysis, using a threshold of 0.9 (Die et al. 2018), and hmmscan revealed that two putative NLRs–CDP20211 and CDP14717 were predicted to not contain a CC domain in their N-terminals. This domain participates in transducing signals (Prigozhin and Krasileva 2021). Meanwhile, the predicted positions of LRRs were identified using LRRsearch with a three-percent false discovery rate (Bej et al. 2014). All of the putative *R* genes were predicted to contain LRR domains except CDP14763. Thus, only 20 of the 23 genes have the three protein domains of an NLR gene and the other three genes are truncated. The architecture for some of the protein products is shown in Figure 2.

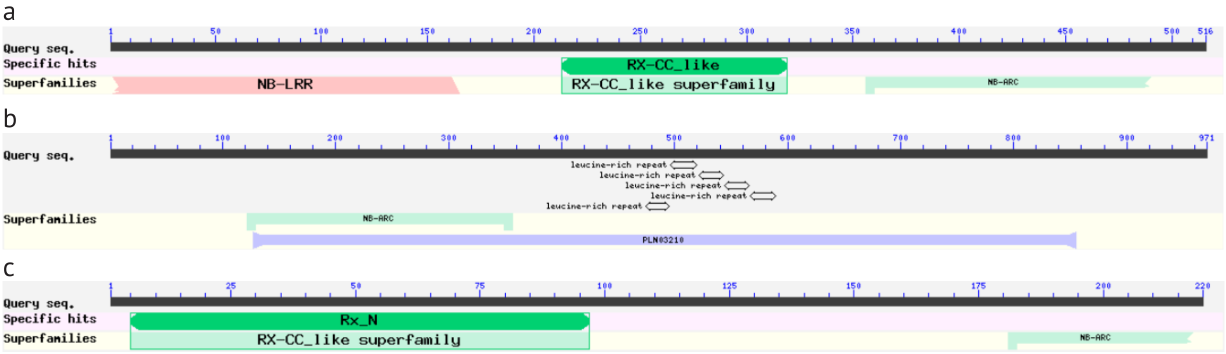

Figure 2. Architecture of select protein products of *R* genes: CDP11823 (a), CDP14717 (b), and CDP14763 (c). Among the *R* genes identified in this study, 20 yield protein products similar to CDP11823 which contains an LRR domain, a CC domain, and an NB-ARC domain. Two yield CDP14717 and CDP20211 which do not have a CC domain. Lastly, one yields CDP14763 which does not have an LRR domain. The images were generated as part of the BlastP results.

Manual gene clustering analysis was done using the criteria. A gene cluster was defined as consisting of three or more copies of NLR genes, with less than 200 kbp distance between two neighboring genes, and with less than eight non-NLR genes in between the adjacent genes (van Wersch and Li 2019). Results revealed that there were three NLR gene clusters in the region of interest as shown in Figure 1. The first cluster consists of the genes encoding CDP11823, CDP11825, CDP11827, CDP11829, CDP11830, and CDP11831. The second gene cluster is made up of three genes encoding the proteins CDP11855, CDP11856, and CDP11857. Lastly, CDP14717, CDP14721, CDP14723, CDP14726, CDP14728, and CDP14730 belong to the third NLR gene cluster. These gene clusters were shaded orange in Figure 1. The other NLR genes exist either in tandems or singletons.

Pathogen Detection

Aside from the NLR genes, other identified resistance-associated genes that also function in detecting pathogens were specified in Table 3. These were seven predicted RLKs that are part of the transmembrane pattern recognition receptors (PRRs) at the surface of cells (Zipfel 2014). Among them, only CDP15875 and CDP15925 were reported to contain LRR domains. This domain confers specificity to the receptors as part of their function in recognizing different molecules of the

Table 3. List of identified probable predicted resistance protein-encoding genes in the 19–23 Mbp region of *C. canephora* chromosome 5 based on the topmost BLASTP hit each with their selected related GO entries

Gene stable ID (Protein stable ID)	Chromosome location		Topmost BLASTP hit description	Protein domains	Selected relevant GO entries from the Ensembl Plants search
	Start (bp)	End (bp)			
GSCOC_ T00035148001 (CDP11874)	20,363,686	20,366,824	G-type lectin S-receptor-like serine/ threonine-protein kinase At2g19130 [*Coffea arabica*] (Max Score=1701; Total Score=1701; Query Cover=99%; E Value=0; Percent Identity=98.57%)	Prot_kinase_dom S_locus_glycoprot_ dom Bulb-type_lectin_dom Pan_app Ser/Thr_kinase_AS Kinase-like_dom_sf Protein_kinase_ATP_ BS SRK-like_kinase Bulb-type_lectin_ dom_sf	phosphorylation, protein phosphorylation
GSCOC_ T00035154001 (CDP11877)	20,416,133	20,417,203	Hypothetical protein M9H77_00820 [*Catharanthus roseus*] (Max Score=162; Total Score=162; Query Cover=97%; E Value=4e-45; Percent Identity=44.86%) Probable serine/ threonine-protein kinase PBL21 isoform X2 [*Elaeis guineensis*] (Max Score=150; Total Score=150; Query Cover=94%; E Value=6e-41; Percent Identity=42.58%)	Prot_kinase_dom Ser-Thr/Tyr_kinase_ cat_dom Kinase-like_dom_sf	protein phosphorylation

Gene stable ID (Protein stable ID)	Chromosome location		Topmost BLASTP hit description	Protein domains	Selected relevant GO entries from the Ensembl Plants search
	Start (bp)	End (bp)			
GSCOC_ T00035182001 (CDP11896)	20,586,375	20,588,306	Probable receptor-like protein kinase At5g18500 isoform X2 [*Coffea eugenioides*] (Max Score=277; Total Score=277; Query Cover=97%; E Value=1e-88; Percent Identity=95.21%)	Prot_kinase_dom Ser-Thr/Tyr_kinase_ cat_dom Kinase-like_dom_sf	protein phosphorylation
GSCOC_ T00042164001 (CDP14731)	21,406,822	21,417,137	Protein STRUBBELIG-RECEPTOR FAMILY 3-like isoform X1 [*Coffea eugenioides*] (Max Score=1624; Total Score=1624; Query Cover=99%; E Value=0; Percent Identity=98.76%)	Prot_kinase_dom Ser-Thr/Tyr_kinase_ cat_dom Leu-rich_rpt Kinase-like_dom_sf LRR_N_plant-typ LRR_dom_sf SRF-like	protein phosphorylation, peptidyl-tyrosine phosphorylation
GSCOC_ T00016777001 (CDP15875)	22,480,014	22,484,039	Receptor-like protein kinase HSL1 [*Coffea arabica*] (Max Score=2008; Total Score=2008; Query Cover=99%; E Value=0; Percent Identity=99.8%)	Prot_kinase_dom Leu-rich_rpt Ser/Thr_kinase_AS Kinase-like_dom_sf LRR_N_plant-typ Protein_kinase_ATP_ BS LRR_dom_sf	phosphorylation, protein phosphorylation
GSCOC_ T00016844001 (CDP15925)	22,916,777	22,919,271	LRR receptor-like serine/threonine-protein kinase GSO2 isoform X2 [*Coffea eugenioides*] (Max Score=981; Total Score=1642; Query Cover=91%; E Value=0; Percent Identity=97.39%)	Leu-rich_rpt Leu-rich_rpt_typical-subtyp LRR_N_plant-typ LRR_dom_sf	defense response to fungus

Gene stable ID (Protein stable ID)	Chromosome location		Topmost BLASTP hit description	Protein domains	Selected relevant GO entries from the Ensembl Plants search
	Start (bp)	End (bp)			
GSCOC_ T0003514500 (CDP11871)	20,342,736	20,348,390	Uncharacterized protein LOC113688897 isoform X1 [*Coffea arabica*] (Max Score=1444; Total Score=1444; Query Cover=99%; E Value=0; Percent Identity=99.86%) Protein phosphatase 1 regulatory subunit pprA like [*Actinidia chinensis* var. *chinensis*] (Max Score=745; Total Score=745; Query Cover=99%; E Value=0; Percent Identity=58.08%)	Leu-rich_rpt Leu-rich_rpt_typical-subtyp LRR_dom_sf	defense response to fungus, negative regulation of growth

pathogenic organism, i.e., either a PAMP or DAMP (Decreux and Messiaen 2005; Zipfel 2014). Furthermore, the analysis of protein domains of the latter predicted that it only contained LRRs as detailed in Table 3. This suggests that CDP15925 may be the C-terminal domain of a novel NLR as its GO annotation proposes its function in the defense response against fungus.

Signal Transduction

After detecting the pathogenic compounds, the signals are then transduced into the intracellular domain of the cell. Genes identified under signal transduction were tabulated in Table 4. These genes were categorized using the recognized components of plant disease response by Andersen et al. (2018). Thus, Table 4 detailed the related resistance-associated genes under the phase of signal transduction specifically those involved in signaling elements like G-protein, ubiquitin, and calcium; hormones such as salicylic acid (SA), jasmonic acid (JA), and brassinosteroid (BR); enzymes like kinases and phosphatases; and epigenetics. Table 4 reported proteins CDP11839, CDP14777, and CDP15866 to be potentially involved in protein ubiquitination and subsequent proteasomal degradation of their target proteins. Components of signaling pathways can be regulated through the activation or repression of protein degradation (Andersen et al. 2018).

Table 4. List of identified resistance-associated genes that function in signal transduction in the 19–23 Mbp region of *C. canephora* chromosome 5 based on the topmost BLASTP hit each with their protein domains and selected relevant GO entries

Gene stable ID (Protein stable ID)	**Chromosome location**		**Topmost BLASTP hit description**	**Protein domains**	**Selected relevant GO entries from the Ensembl Plants search**
	Start (bp)	**End (bp)**			
			Signaling Mechanism		
			G-protein		
GSCOC_T00035212001 (CDP11908)	20,761,548	20,770,617	Diacylglycerol kinase 5 isoform X2 [*Coffea arabica*] (Max Score=988; Total Score=988; Query Cover=98%; E Value=0; Percent Identity=97.73%)	Diacylglycerol_kin_accessory Diacylglycerol_kinase_cat_dom NAD/diacylglycerol_kinase_sf Diacylglycerol_kinase_pln ATP-NAD_kinase_N DGK	defense response, protein kinase C-activating G protein-coupled receptor signaling pathway, signal transduction, phosphorylation
			Calcium		
GSCOC_T00042242001 (CDP14785)	21,949,616	21,968,336	Two pore calcium channel protein 1B-like isoform X2 [*Coffea arabica*] (Max Score=1518; Total Score=1518; Query Cover=99%; E Value=0; Percent Identity=100%)	EF_hand_dom Ion_trans_dom EF-hand-dom_pair Volt_channel_dom_sf TPC1_plant	calcium ion transmembrane transport, calcium ion transport, calcium-mediated signaling, regulation of jasmonic acid biosynthetic process
			Ubiquitin		
GSCOC_T00035105001 (CDP11839)	19,930,929	19,934,482	BTB/POZ domain-containing protein At3g22104 isoform X1 [*Coffea eugenioides*] (Max Score=1091; Total Score=1091; Query Cover=99%; E Value=0; Percent Identity=98.35%)	BTB/POZ_dom SKP1/BTB/POZ_sf NPH3_dom NPH3/RPT2-like	protein ubiquitination
GSCOC_T00042231001 (CDP14777)	21,805,722	21,809,463	E3 ubiquitin-protein ligase RMA1H1-like [*Coffea arabica*] (Max Score=496; Total Score=496; Query Cover=99%; E Value=3e-177; Percent Identity=98.77%)	Znf_RING Znf_RING/FYVE/PHD Znf_RING_CS Znf_C3HC4_RING-type RNF5/RNF185-like	protein ubiquitination, ubiquitin-dependent protein catabolic process

Gene stable ID (Protein stable ID)	Chromosome location		Topmost BLASTP hit description	Protein domains	Selected relevant GO entries from the Ensembl Plants search
	Start (bp)	End (bp)			
GSCOC_ T00016764001 (CDP15866)	22,377,235	22,383,229	DDB1- and CUL4-associated factor 13-like [*Coffea arabica*] (Max Score=940; Total Score=940; Query Cover=99%; E Value=0; Percent Identity=99.34%)	WD40_repeat Sof1 WD40/YVTN_repeat-like_dom_sf WD40_repeat_CS G-protein_beta_WD-40_rep WD40_repeat_dom_sf	protein ubiquitination
Hormones					
Salicylic Acid					
GSCOC_ T00042218001 (CDP14767)	21,741,588	21,744,188	Protein ACCELERATED CELL DEATH 6-like [*Coffea arabica*] (Max Score=1165; Total Score=1165; Query Cover=99%; E Value=0; Percent Identity=93.76%)	Ankyrin_rpt PGG_dom Ankyrin_rpt-contain_sf	regulation of defense response, regulation of salicylic acid mediated signaling pathway, cellular response to salicylic acid stimulus
GSCOC_ T00042219001 (CDP14768)	21,744,248	21,752,130	Protein ACCELERATED CELL DEATH 6-like [*Coffea arabica*] (Max Score=1067; Total Score=1067; Query Cover=85%; E Value=0; Percent Identity=89.02%)	Ankyrin_rpt PGG_dom Ankyrin_rpt-contain_sf	regulation of defense response, regulation of salicylic acid mediated signaling pathway, cellular response to salicylic acid stimulus
Jasmonic Acid					
GSCOC_ T00042242001 (CDP14785)	21,949,616	21,968,336	Two pore calcium channel protein 1B-like isoform X2 [*Coffea arabica*] (Max Score=1518; Total Score=1518; Query Cover=99%; E Value=0; Percent Identity=100%)	EF_hand_dom Ion_trans_dom EF-hand-dom_pair Volt_channel_dom_sf TPC1_plant	calcium ion transmembrane transport, calcium ion transport, calcium-mediated signaling, regulation of jasmonic acid biosynthetic process

Gene stable ID (Protein stable ID)	Chromosome location		Topmost BLASTP hit description	Protein domains	Selected relevant GO entries from the Ensembl Plants search
	Start (bp)	End (bp)			
			Brassinoteroid		
GSCOC_ T00042125001 (CDP14706)	21,145,617	21,154,486	Protein LAZ1 [*Coffea arabica*] (Max Score=1003; Total Score=1003; Query Cover=99%; E Value=0; Percent Identity=99.79%)	Ostalpha/TMEM184C	plant-type hypersensitive response, programmed cell death, negative regulation of brassinosteroid mediated signaling pathway
			Enzymes		
			Kinase		
GSCOC_ T00035212001 (CDP11908)	20,761,548	20,770,617	Diacylglycerol kinase 5 isoform X2 [*Coffea arabica*] (Max Score=988; Total Score=988; Query Cover=98%; E Value=0; Percent Identity=97.73%)	Diacylglycerol_kin_ accessory Diacylglycerol_ kinase_cat_dom NAD/diacylglycerol_ kinase_sf Diacylglycerol_ kinase_pln ATP-NAD_kinase_N DGK	defense response, protein kinase C-activating G protein-coupled receptor signaling pathway, signal transduction, phosphorylation
GSCOC_ T00042250001 (CDP14791)	22,010,496	22,017,553	Probable serine/ threonine-protein kinase At1g54610 isoform X2 [*Coffea arabica*] (Max Score=1445; Total Score=1445; Query Cover=99%; E Value=0; Percent Identity=99.86%)	Prot_kinase_dom Ser/Thr_kinase_AS Kinase-like_dom_sf Protein_kinase_ ATP_BS	phosphorylation, protein phosphorylation
			Phosphatase		
GSCOC_ T00042286001 (CDP14822)	22,292,079	22,294,180	Probable protein phosphatase 2C 58 isoform X1 [*Coffea arabica*] (Max Score=554; Total Score=554; Query Cover=93%; E Value=0; Percent Identity=98.17%)	PPM-type_ phosphatase_dom PP2C PPM-type_dom_sf	protein dephosphorylation

Gene stable ID (Protein stable ID)	Chromosome location		Topmost BLASTP hit description	Protein domains	Selected relevant GO entries from the Ensembl Plants search
	Start (bp)	End (bp)			
			Epigenetics		
GSCOC_ T00035057001 (CDP11807)	19,392,475	19,399,533	Histone-lysine N-methyltransferase ATXR2-like [*Coffea arabica*] (Max Score=1007; Total Score=1007; Query Cover=99%; E Value=0; Percent Identity=98.8%)	SET_dom Znf_MYND ATXR2-like	histone H3-K36 methylation, histone H3-K36 trimethylation
GSCOC_ T00035100001 (CDP11835)	19,868,290	19,882,622	Histone-lysine N-methyltransferase ATXR3-like [*Coffea arabica*] (Max Score=4826; Total Score=4826; Query Cover=99%; E Value=0; Percent Identity=98.28%)	SET_dom LRR_dom_sf F-box-like_dom_sf SDG2_C	epigenetic regulation of gene expression

Also participating in signaling pathways are transcription factors (TFs) listed in Table 5. The TF families particularly involved in plant defense present in the identified genes were APETALA 2/ethylene-responsive element binding factor (AP2/ERF), basic helix-loop-helix (bHLH), TGACG-binding (TGA)/basic leucine zipper (bZIP), myeloblastosis (MYB), and no apical meristem (NAM)-*Arabidopsis* transcription activation factor (ATAF)-cup-shaped cotyledon (CUC2) (NAC). Other TF families that composed other TFs in these genes were homeobox, bromodomain, and Teosinte branched1/Cincinnata/proliferating cell factor (TCP). There exist six TF families notably involved in plant defense that were also present in the region of interest as reported in Table 5—AP2/ERF (in CDP14752 and CDP1585), bHLH (in CDP11834, CDP14797, CDP14798, CDP14799, CDP14800, and CDP15921), TGA/bZIP (in CDP15899), MYB (in CDP15870), NAC (in CDP15876), and WRKY (Tsuda and Somssich 2015). Other TF families found in the analysis were homeobox (in CDP15874), bromodomain (in CDP15907), and TCP (in CDP15905).

Table 5. List of identified TFs in the 19–23 Mbp region of *C. canephora* chromosome 5 based on the topmost BLASTP hit each with their protein domains and selected relevant GO entries

Gene stable ID (Protein stable ID)	Chromosome location		Topmost BLASTP hit description	Protein domains	Selected relevant GO entries from the Ensembl Plants search
	Gene start (bp)	Gene end (bp)			
GSCOC_ T00042199001 (CDP14752)	21,662,748	21,665,730	Ethylene-responsive transcription factor ERF054-like [*Coffea arabica*] (Max Score=1003; Total Score=1003; Query Cover=99%; E Value=0; Percent Identity=100%)	AP2/ERF_dom DNA-bd_dom_sf AP2/ERF_dom_sf	defense response, DNA-templated transcription, regulation of DNA-templated transcription
GSCOC_ T00042256001 (CDP14797)	22,058,039	22,058,618	Transcription factor bHLH162-like [*Coffea arabica*] (Max Score=308; Total Score=308; Query Cover=99%; E Value=8e-105; Percent Identity=98.69%)	bHLH_dom MASH1/Ascl1a-like HLH_DNA-bd_sf	regulation of DNA-templated transcription, regulation of transcription by RNA polymerase II
GSCOC_ T00042257001 (CDP14798)	22,070,130	22,071,038	Transcription factor bHLH162-like [*Coffea eugenioides*] (Max Score=409; Total Score=409; Query Cover=99%; E Value=3e-143; Percent Identity=94.52%)	bHLH_dom MASH1/Ascl1a-like HLH_DNA-bd_sf	regulation of DNA-templated transcription, regulation of transcription by RNA polymerase II
GSCOC_ T00042258001 (CDP14799)	22,073,304	22,074,222	Transcription factor bHLH162-like [*Coffea arabica*] (Max Score=432; Total Score=432; Query Cover=99%; E Value=2e-152; Percent Identity=97.69%)	bHLH_dom MASH1/Ascl1a-like HLH_DNA-bd_sf	regulation of DNA-templated transcription, regulation of transcription by RNA polymerase II
GSCOC_ T00042259001 (CDP14800)	22,095,009	22,096,533	Transcription factor bHLH162-like [*Coffea arabica*] (Max Score=462; Total Score=462; Query Cover=99%; E Value=3e-164; Percent Identity=99.55%)	bHLH_dom MASH1/Ascl1a-like HLH_DNA-bd_sf	regulation of DNA-templated transcription, regulation of transcription by RNA polymerase II

Gene stable ID (Protein stable ID)	Chromosome location		Topmost BLASTP hit description	Protein domains	Selected relevant GO entries from the Ensembl Plants search
	Gene start (bp)	Gene end (bp)			
GSCOC_ T00016753001 (CDP15858)	22,332,922	22,333,584	Ethylene-responsive transcription factor 4-like [*Coffea arabica*] (Max Score=447; Total Score=447; Query Cover=99%; E Value=1e-158; Percent Identity=100%)	AP2/ERF_dom DNA-bd_dom_sf AP2/ERF_dom_sf	defense response, DNA-templated transcription, regulation of DNA-templated transcription
GSCOC_ T00016779001 (CDP15876)	22,488,866	22,491,310	NAC domain-containing protein 73-like isoform X1 [*Coffea arabica*] (Max Score=558; Total Score=558; Query Cover=98%; E Value=0; Percent Identity=99.63%)	NAC-dom NAC_dom_sf SOG1-like	regulation of DNA-templated transcription
GSCOC_ T00016806001 (CDP15899)	22,707,627	22,711,257	Basic leucine zipper 61-like [*Coffea arabica*] (Max Score=631; Total Score=631; Query Cover=99%; E Value=0; Percent Identity=100%)	bZIP bZIP_RF2	regulation of DNA-templated transcription
GSCOC_ T00016839001 (CDP15921)	22,898,405	22,900,115	Transcription factor bHLH30-like isoform X1 [*Coffea arabica*] (Max Score=518; Total Score=518; Query Cover=99%; E Value=0; Percent Identity=98.82%)	bHLH_dom HLH_DNA-bd_sf AIG1-like	regulation of DNA-templated transcription

Defense Response

The last phase of disease response mechanism of plants was the defense response itself (Andersen et al. 2018). The determined resistant-associated genes under this stage were summarized in Table 6. Genes under the defense response phase were categorized into cell wall modification, reactive oxygen species (ROS), pathogenesis-related (PR) proteins such as ribonucleases and proteases, vitamins, and secondary metabolites like alkaloids, terpenes, and phenylpropanoids.

In ETI, the activation of NLRs and the subsequent cascade of signals lead to the plant hypersensitive response (HR) (Nguyen et al. 2021), where the infected and surrounding cells are programmed to die, halting the transmission of biotrophic

Table 6. List of identified resistance-associated genes that function in plant "defense response" in the 19–23 Mbp region of *C. canephora* chromosome 5 based on the topmost BLASTP hit each with their protein domains and selected relevant GO entries. The second topmost identified BLASTP hit was given for uncharacterized proteins.

Gene stable ID (Protein stable ID)	Chromosome location		Topmost BLASTP hit description	Protein domains	Selected relevant GO entries from the Ensembl Plants search
	Start (bp)	End (bp)			
Cell Wall Modification					
GSCOC_T00035120001 (CDP11853)	20,131,317	20,134,076	Xyloglucan endotransglucosylase/hydrolase protein 9-like [*Coffea arabica*] (Max Score=592; Total Score=592; Query Cover=99%; E Value=0; Percent Identity=99.64%)	GH16 Beta_glucanase XET_C ConA-like_dom_sf XTH Beta-glucanase/XTH	cell wall biogenesis, cell wall organization, xyloglucan metabolic process, carbohydrate metabolic process, cellular glucan metabolic process
GSCOC_T00042132001 (CDP14710)	21,213,066	21,218,344	Probable galacturonosyltransferase 10 [*Coffea arabica*] (Max Score=1105; Total Score=1105; Query Cover=99%; E Value=0; Percent Identity=99.25%)	Glyco_trans_8 Nucleotide-diphossugar_trans GAUT	cell wall organization, pectin biosynthetic process
GSCOC_T00042241001 (CDP14784)	21,938,666	21,946,066	Callose synthase 12-like [*Coffea arabica*] (Max Score=3703; Total Score=3703; Query Cover=99%; E Value=0; Percent Identity=99.83%)	Glyco_trans_48 FKS1-like_dom1	cell wall organization, (1->3)-beta-D-glucan biosynthetic process
Reactive Oxygen Species					
GSCOC_T00035042001 (CDP11799)	19,281,260	19,283,814	Peroxidase 27-like [*Coffea arabica*] (Max Score=670; Total Score=670; Query Cover=99%; E Value=0; Percent Identity=99.69%)	Peroxidase_pln Haem_peroxidase Haem_peroxidase_sf Peroxidases_heam-ligand_BS Peroxidases_AS Secretory_peroxidase	hydrogen peroxide catabolic process, response to oxidative stress, cellular oxidant detoxification

Gene stable ID (Protein stable ID)	Chromosome location		Topmost BLASTP hit description	Protein domains	Selected relevant GO entries from the Ensembl Plants search
	Start (bp)	End (bp)			
GSCOC_ T00035044001 (CDP11800)	19,295,217	19,297,264	Peroxidase 27-like [*Coffea arabica*] (Max Score=662; Total Score=662; Query Cover=99%; E Value=0; Percent Identity=97.86%)	Peroxidase_pln Haem_peroxidase Haem_peroxidase_ sf Peroxidases_heam-ligand_BS Peroxidases_AS Secretory_ peroxidase	hydrogen peroxide catabolic process, response to oxidative stress, cellular oxidant detoxification
GSCOC_ T00035045001 (CDP11801)	19,302,533	19,304,073	Peroxidase 3-like [*Coffea arabica*] (Max Score=685; Total Score=685; Query Cover=99%; E Value=0; Percent Identity=99.7%)	Peroxidase_pln Haem_peroxidase Haem_peroxidase_ sf Peroxidases_heam-ligand_BS Peroxidases_AS Secretory_ peroxidase	hydrogen peroxide catabolic process, response to oxidative stress, cellular oxidant detoxification
GSCOC_ T00035143001 (CDP11870)	20,326,820	20,330,558	frataxin, mitochondrial-like [*Coffea arabica*] (Max Score=409; Total Score=409; Query Cover=99%; E Value=4e-144; Percent Identity=100%)	Frataxin/CyaY Frataxin Frataxin_CS Frataxin/CyaY_sf	response to oxidative stress, response to hydrogen peroxide
PR Proteins					
Ribonuclease					
GSCOC_ T00035135001 (CDP11863)	20,204,277	20,207,511	Uncharacterized protein LOC113688473 [*Coffea arabica*] (Max Score=521; Total Score=521; Query Cover=99%; E Value=0; Percent Identity=98.04%) Ribonuclease H [*Senna tora*] Max Score=52; Total Score=52; Query Cover=56%; E Value=0.0002; Percent Identity=24.32%)	RNaseH_domain	RNA phosphodiester bond hydrolysis, endonucleolytic

Gene stable ID (Protein stable ID)	Chromosome location		Topmost BLASTP hit description	Protein domains	Selected relevant GO entries from the Ensembl Plants search
	Start (bp)	End (bp)			
GSCOC_ T00035147001 (CDP11873)	20,353,520	20,355,890	5' exonuclease Apollo-like isoform X1 [*Coffea arabica*] (Max Score=907; Total Score=907; Query Cover=99%; E Value=0; Percent Identity=100%)	DRMBL RibonucZ/ Hydroxyglut_hydro	cellular response to DNA damage stimulus
GSCOC_ T00042284001 (CDP14821)	22,283,813	22,285,216	Ribonuclease 1-like [*Coffea arabica*] (Max Score=481; Total Score=481; Query Cover=99%; E Value=2e-171; Percent Identity=99.57%)	RNase_T2-like Ribonuclease_T2_ eukaryotic RNase_T2-like_sf	RNA phosphodiester bond hydrolysis, endonucleolytic
Protease					
GSCOC_ T00035113001 (CDP11847)	20,007,847	20,009,211	Aspartic proteinase CDR1-like [*Coffea eugenioides*] (Max Score=871; Total Score=871; Query Cover=99%; E Value=0; Percent Identity=94.93%)	Aspartic_ peptidase_AS Peptidase_ aspartic_dom_sf TAXi_C TAXi_N PEPTIDASE_A1 Pepsin-like_plant	proteolysis
GSCOC_ T00035114001 (CDP11848)	20,014,978	20,016,351	Aspartic proteinase CDR1-like [*Coffea eugenioides*] (Max Score=918; Total Score=918; Query Cover=99%; E Value=0; Percent Identity=98.69%)	Aspartic_ peptidase_AS Peptidase_ aspartic_dom_sf TAXi_C TAXi_N PEPTIDASE_A1 Pepsin-like_plant	proteolysis
GSCOC_ T00035117001 (CDP11850)	20,043,564	20,044,931	Aspartic proteinase CDR1-like [*Coffea arabica*] (Max Score=859; Total Score=859; Query Cover=99%; E Value=0; Percent Identity=94.29%)	Aspartic_ peptidase_A1 Aspartic_ peptidase_AS Peptidase_ aspartic_dom_sf TAXi_C TAXi_N PEPTIDASE_A1 Pepsin-like_plant	proteolysis

Gene stable ID (Protein stable ID)	Chromosome location		Topmost BLASTP hit description	Protein domains	Selected relevant GO entries from the Ensembl Plants search
	Start (bp)	End (bp)			
GSCOC_ T00016805001 (CDP15898)	22,700,988	22,702,405	Ervatamin-B-like [*Coffea arabica*] (Max Score=735; Total Score=735; Query Cover=99%; E Value=0; Percent Identity=99.16%)	Pept_cys_AS Peptidase_C1A_C Prot_inhib_I29 Pept_his_AS Pept_asp_AS Papain-like_cys_ pep_sf Peptidase_C1A_ papain-like	proteolysis
Secondary Metabolite					
Alkaloid					
GSCOC_ T00035072001 (CDP11819)	19,548,788	19,550,318	Phospholipase A1-II 1-like [*Coffea arabica*] (Max Score=817; Total Score=817; Query Cover=99%; E Value=0; Percent Identity=99.75%)	Fungal_lipase-like AB_hydrolase PLA	alkaloid metabolic process
GSCOC_ T00042119001 (CDP14701)	21,096,797	21,098,392	Berberine bridge enzyme-like 8 [*Coffea eugenioides*] (Max Score=1093; Total Score=1093; Query Cover=99%; E Value=0; Percent Identity=99.25%)	Oxid_FAD_bind_N BBE FAD-bd_PCMH FAD-bd_PCMH_sub1 FAD-bd_PCMH_sub2 FAD-bd_PCMH-like_sf	alkaloid metabolic process
GSCOC_ T00042122001 (CDP14704)	21,106,561	21,108,165	Berberine bridge enzyme-like 4 [*Coffea arabica*] (Max Score=868; Total Score=868; Query Cover=80%; E Value=0; Percent Identity=96.52%)	Oxid_FAD_bind_N BBE FAD-bd_PCMH FAD-bd_PCMH_sub1 FAD-bd_PCMH_sub2 FAD-bd_PCMH-like_sf	alkaloid metabolic process

Gene stable ID (Protein stable ID)	Chromosome location		Topmost BLASTP hit description	Protein domains	Selected relevant GO entries from the Ensembl Plants search
	Start (bp)	End (bp)			
GSCOC_T00042207001 (CDP14759)	21,708,472	21,709,334	Uncharacterized protein LOC113772426 [*Coffea eugenioides*] (Max Score=266; Total Score=266; Query Cover=85%; E Value=3e-86; Percent Identity=77.5%) Bifunctional epoxide hydrolase 2 [*Capsicum annuum*] (Max Score=247; Total Score=247; Query Cover=85%; E Value=1e-78; Percent Identity=74.38%)	AB_hydrolase	alkaloid metabolic process
GSCOC_T00042208001 (CDP14760)	21,709,411	21,710,006	Hypothetical protein HYC85_026117 [*Camellia sinensis*] (Max Score=91.7; Total Score=91.7; Query Cover=45%; E Value=2e-21; Percent Identity=89.13%) Bifunctional epoxide hydrolase 2 [*Camellia lanceoleosa*] (Max Score=93.6; Total Score=93.6; Query Cover=45%; E Value=1e-20; Percent Identity=89.13%)	AB_hydrolase	alkaloid metabolic process
GSCOC_T00042212001 (CDP14762)	21,714,518	21,716,190	Uncharacterized protein LOC113771758 [*Coffea eugenioides*] (Max Score=295; Total Score=295; Query Cover=66%; E Value=2e-99; Percent Identity=98.61%) Putative F-box/kelch-repeat protein-like [*Capsicum annuum*] (Max Score=256; Total Score=256; Query Cover=98%; E Value=9e-82; Percent Identity=53.93%)	Epox_hydrolase-like AB_hydrolase	alkaloid metabolic process

Gene stable ID (Protein stable ID)	Chromosome location		Topmost BLASTP hit description	Protein domains	Selected relevant GO entries from the Ensembl Plants search
	Start (bp)	End (bp)			
GSCOC_ T00042215001 (CDP14765)	21,728,501	21,730,168	Uncharacterized protein LOC113772426 [*Coffea eugenioides*] (Max Score=628; Total Score=628; Query Cover=99%; E Value=0; Percent Identity=97.78%) Bifunctional epoxide hydrolase 2 [*Capsicum annuum*] (Max Score=508; Total Score=508; Query Cover=99%; E Value=2e-179; Percent Identity=77.22%)	AB_hydrolase_1 Epox_hydrolase-like AB_hydrolase	alkaloid metabolic process
Terpene					
GSCOC_ T00035108001 (CDP11842)	19,959,720	19,964,609	Isoprene synthase, chloroplastic-like [*Coffea arabica*] (Max Score=1269; Total Score=1269; Query Cover=99%; E Value=0; Percent Identity=99.02%)	Terpene_synth_N Terpene_synthase_ metal-bd Terpenoid_cyclase/ PrenylTrfase Isoprenoid_ synthase_dom_sf Terpene_cyclase-like_1_C Terpene_synth_N_ sf Terpene_cyclase_ plant_C1	diterpenoid biosynthetic process
GSCOC_ T00035109001 (CDP11843)	19,970,432	19,976,641	Isoprene synthase, chloroplastic-like [*Coffea arabica*] (Max Score=1242; Total Score=1242; Query Cover=99%; E Value=0; Percent Identity=98.67%)	Terpene_synth_N Terpene_synthase_ metal-bd Terpenoid_cyclase/ PrenylTrfase Isoprenoid_ synthase_dom_sf Terpene_cyclase-like_1_C Terpene_synth_N_ sf Terpene_cyclase_ plant_C1	diterpenoid biosynthetic process

Gene stable ID (Protein stable ID)	Chromosome location		Topmost BLASTP hit description	Protein domains	Selected relevant GO entries from the Ensembl Plants search
	Start (bp)	End (bp)			
Phenylpropanoid					
GSCOC_ T00035196001 (CDP11901)	20,668,169	20,668,735	Dirigent protein 22-like [*Coffea arabica*] (Max Score=382; Total Score=382; Query Cover=99%; E Value=8e-134; Percent Identity=100%)	Dirigent Allene_oxi_cyc_ Dirigent	phenylpropanoid biosynthetic process

pathogens (Andersen et al. 2018). From their protein sequences, the putative R proteins CDP11823, CDP11825, and CDP11829 as receptors and the PCD-associated CDP14706 are predicted to function in plant HR. Other identified resistance-associated genes, CDP11799, CDP11800, and CDP11801, were predicted to encode for peroxidase. Class III plant peroxidases, as PR proteins, are included among the defense proteins induced during the immune response. Aside from participating in the metabolism and production of ROS as well as reactive nitrogen species (RNS), they are also involved in the synthesis of plant bioactive compounds like phytoalexins, which are organic chemicals that hinder the pest's metabolism, development, and reproduction during an infestation. These enzymes also function in reinforcing the plant cell walls by cross-linking cell wall components like extension and ferulic acid and depositing lignin and suberin. Ultimately, the peroxidase products ROS, RNS, and phytoalexins serve as elements in transducing signals along with phytohormones (Almagro et al. 2009; Andersen et al. 2018).

There were three other protein-encoding genes in the region of interest associated with cell wall biogenesis and/or organization. First, CDP11853 may act in cleaving, through xyloglucan endohydrolysis (XEH), and religating polymers of xyloglucan via xyloglucan endotransglycosylation (Hyodo et al. 2003). Xyloglucan is the main hemicellulose component of the primary cell wall in dicots (Wan et al. 2021) such as coffee. Second, CDP14710 is a probable galacturonosyltransferase predicted to be a component of the biosynthesis of pectin, the first barrier to foreign invasion after the breakdown of the outermost cutin layer of the epicuticle. Modifications in the pectin component of cell walls, such as its increased methylesterification, had evidence of increased disease resistance (Wan et al. 2021). Lastly, the region probably has a gene segment coding for the callose synthase protein CDP14784. Callose deposition in areas of probable pathogen attack, in between the plasma membrane and cell wall, inhibits their invasion and spread to nearby regions. At the plasmodesmata, callose deposition controls the permeability of molecules where high levels of callose indicate the thinning and eventual closing of the channels (Wan et al. 2021).

PR proteins participate in systemic acquired resistance (SAR) as they are produced when a plant becomes infected by a pathogen (van Loon 1985). Aside from peroxidases that mediate respiratory bursts to produce great quantities of ROS (Lamb and Dixon 1997; Wojtaszek 1997; Andersen et al. 2018), other PR protein families counter and detect effectors from pathogens such as chitinases, β-1,3-glucanases, proteases, protease inhibitors, and ribonucleases (Andersen et al. 2018). Ribonucleases or exonucleases, like CDP11863, CDP11873, and CDP14821, were reported to improve the resistance of *N. tabacum* to cucumber mosaic virus (Sugawara et al. 2016). Meanwhile, aspartic proteases, like CDP11847, CDP11848, and CDP11850, showed antibacterial properties in *A. thaliana* by cleaving the bacterial protein MucD inhibiting growth of *Pseudomonas syringae in planta* and *in vitro* (Wang et al. 2019). Another protease found to exhibit high proteolytic activity is ervatamin B (CDP15898) (Kunduet al. 2000). PCD-associated cysteine proteases, such as ervatamin B, were involved in plant immunity such that their mutation led to increased susceptibility to pathogen infection. For example, *A. thaliana* with mutations in *rd19* and *rd21*, was more susceptible to the bacterium *Ralstonia solanacearum* and fungus *Botrytis cinerea*, respectively (Liu et al. 2018).

Secondary metabolites such as flavonoids, tannins, terpenoids, alkaloids, and phenolics help in fighting off insect herbivory (Andersen et al. 2018). Aside from deterring herbivory, alkaloids also function in storing nitrogen and regulating growth (Ali et al. 2019). Proteins potentially involved in the metabolism of alkaloids in the region of interest include CDP11819, CDP14701, CDP14704, CDP14759, CDP14760, CDP14762, and CDP14765.

Meanwhile, CBD-resistance-associated CDP11842 and CDP11843 proteins similar to isoprene synthase are candidate proteins involved in the biosynthesis of diterpenoids. Isoprene protects photosynthetic membranes by gathering ROS (Fini et al. 2017). Furthermore, isoprene was also reported to have insect-repelling properties (Laothawornkitkul et al. 2008; Loivamäki et al. 2008). It was reported by Idárraga et al. (2011) that there was an upregulation of isoprene synthase 24 hours after *C. liberica* was infested with CBB.

Lastly, a dirigent protein conferring resistance to CLR, CDP11901, was predicted to participate in the biosynthetic pathway of phenylpropanoid. Dirigent proteins are involved in the formation of lignans and lignins from the metabolism of phenylpropanoids. Aside from modifying the cell wall by lignification, lignans participate in plant defense response by disrupting the endocrine system of insects feeding the plant (Paniagua et al. 2017). In CLR resistance, *CaPAL* and *CaCHS* function in producing flavonoids, a phenolic compound like phenylpropanoid (Couttolenc-Brenis et al. 2020).

SUMMARY AND CONCLUSION

The candidate *R* and resistance-associated genes in the Sat229 region in chromosome 5 of the *C. canephora* genome were identified, analyzed, and physically mapped, using their predicted protein products. Web-based bioinformatics tools—Ensembl Plants "BioMart", NCBI BLASTP, hmmscan, PCoils, LRRsearch, and RStudio "chromoMap" were utilized. After filtering the identified *R* and resistance-associated genes using their GO annotations and functions, results revealed that there are 75 potential genes functioning in plant defense response in the 19 to 23 Mbp region of the *C. canephora* chromosome 5. Among these, 23 are candidate *R* genes encoding NLR proteins that recognize pathogen effectors and induce ETI. Other identified 52 genes are resistance-associated genes that function in pathogen detection, signal transduction, and defense response. Notable resistance-associated genes in the region include: CDP11908, involved in the flg22-induced synthesis of PA and ROS and deposition of callose; CDP14767 and CDP14768, proteins similar to ACD6 involved in the positive feedback loop mechanism with SA by complexing with FLS2 leading to the PCD of infected and surrounding cells; and CDP14706 which is associated with ACD11 and HR. Among the identified TFs, the AP2/ERF-containing CDP14752 and CDP1585 were annotated to be involved in 'defense response' by potentially binding to the GCC box in the promoter regions of PR genes, conferring responsiveness to ethylene and upregulation during pathogen invasion.

Potential novel proteins include CDP15925, a predicted RLK from the BLASTP result, and CDP11871, an uncharacterized protein based on the topmost BLASTP hit. Both gene sequences were predicted to encode LRR domains only; thus, they are speculated to be the C-terminal domain of nearby NLR as both function for 'defense response to fungus' based on their GO annotation. There is also evidence that the region containing Sat229, a genetic marker linked to resistance against CLR, also confers *C. canephora* resistance to CBD and CBB. This includes the presence of genes for 'defense response to fungus' as well as genes coding for the CDP11842 and CDP11843 proteins which are similar to isoprene synthase. Isoprene synthase has been linked to CBD resistance as it is found to be upregulated during CBB infestation (Durango et al. 2012).

This *in silico* analysis provided a list of potential *R* and resistance-associated genes that can be used as preliminary information on the resistance brought about by the downstream region of *C. canephora* chromosome 5. Utilizing the generated list and map may aid future research to improve disease resistance in coffee, particularly in identifying resistant cultivars for propagation. On a larger scale, this probably opens new information in breeding resistant cultivars and subsequently increasing the yield of production, especially in the Philippines.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the Institute of Biology, College of Science, UP Diliman and the members of the Plant Molecular Biology and Genetics Laboratory. The authors also declare no conflict of interest.

REFERENCES

Ali AA, Abdelrahman M, El-Sayed MA. 2019. Alkaloid role in plant defense response to growth and stress. In: Jogaiah S, Abdelrahman M, editors. Bioactive Molecules in Plant Defense: Signaling in Growth and Stress. Switzerland: Springer Cham. p. 145–158. https://doi.org/10.1007/978-3-030-27165-7_9.

Almagro L, Ros LVG, Belchí-Navarro S, Bru R, Barceló AR, Pedreño MA. 2009. Class III peroxidases in plant defence reactions. J Exp Bot. 60(2):377–390. https://doi.org/10.1093/jxb/ern277.

Anand L, Rodriguez Lopez CM. 2022. ChromoMap: an R package for interactive visualization of multi-omics data and annotation of chromosomes. BMC Bioinformatics. 23(33). https://doi.org/10.1186/s12859-021-04556-z.

Andersen EJ, Ali S, Byaukama E, Yen Y, Nepal, MP. 2018. Disease resistance mechanisms in plants. Genes. 9(7):339. https://doi.org/10.3390/genes9070339.

Bej A, Sahoo BR, Swain B, Basu M, Jayasankar P, Samanta M. 2014. LRRsearch: An asynchronous server-based application for the prediction of leucine-rich repeat motifs and an integrative database of NOD-like receptors. Comput Biol Med. 53:164–170. https://doi.org/10.1016/j.compbiomed.2014.07.016.

Bettencourt AJ, Noronha-Wagner M. 1971. Genetic factors conditioning resistance of *Coffea arabica* L. to *Hemileia vastatrix* Berk. and Br. Agron Lusit. 31:285–292.

Bettencourt AJ, Rodrigues Jr CJ. 1988. Principles and practice of coffee breeding for resistance to rust and other diseases. In Clarck RJ, Macrae R, editors. Coffee Agronomy Vol. 4. London, UK; New York, NY, USA: Elsevier Applied Science Publishers LTD. p. 199–234.

Bettencourt AJ, Noronha-Wagner M, Lopes M. 1980. Factor genético que condiciona a resistência do clone 1343/269 (Híbrido de Timor) à *Hemileia vastatrix* Berk. and Br. Brotéria Genética. 1(76):53–58.

Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. 2009. QuickGO: a web-based tool for Gene Ontology searching. Bioinformatics. 25(22):3045–3046. https://doi.org/10.1093/bioinformatics/btp536.

Brodie A, Azaria JR, Ofran Y. 2016. How far from the SNP may the causative genes be? Nucleic Acids Res. 44(13):6046–6054. https://doi.org/10.1093/nar/gkw500.

Bureau of Plant Industry (Republic of the Philippines). nd. Organic arabica coffee seedling production. Malate, Manila: BPI - Crop Research and Production Support Division; [updated 2021 Oct 22; accessed 2023 Jul 18]. https://www.buplant.da.gov.ph/index.php/guide/10783-organic-arabica-coffee-seedling-production

Couttolenc-Brenis E, Carrión GL, Villain L, Ortega-Escalona F, Ramírez-Martínez D, Mata-Rosas M, Méndez-Bravo A. 2020. Prehaustorial local resistance to coffee leaf rust in a Mexican cultivar involves expression of salicylic acid-responsive genes. PeerJ. 8:e8345. https://doi.org/10.7717/peerj.8345 .Cui L, Hanika K, Visser RG, Bai Y. 2020. Improving pathogen resistance by exploiting plant susceptibility genes in Coffee (*Coffea* spp.). Agronomy. 10(12):1928. https://doi.org/10.3390/agronomy10121928.

Decreux A, Messiaen J. 2005. Wall-associated kinase WAK1 interacts with cell wall pectins in a calcium-induced conformation. Plant Cell Physiol. 46(2):268–278. https://doi.org/10.1093/pcp/pci026.

Department of Agriculture (Republic of the Philippines). 2022. Philippine Coffee Industry Roadmap 2021–2025. Diliman, Quezon City: Department of Agriculture - Bureau of Agricultural Research (Republic of the Philippines). [accessed: 2022 Jul 12]. http://www.pcaf.da.gov.ph/.php/cir-coffee/.

Department of Agriculture (Republic of the Philippines), Department of Trade and Industry (Republic of the Philippines). 2019. Philippine Coffee Industry Roadmap 2017–2022. [accessed: 2023 May 09]. https://www.da.gov.ph/wp-content/uploads/2019/06/Philippine-Coffee-Industry-Roadmap-2017-2022.pdf.

Die JV, Román B, Qi X, Rowland LJ. 2018. Genome-scale examination of NBS-encoding genes in blueberry. Sci Rep. 8(1):3429. https://doi.org/10.1038/s41598-018-21738-7.

Fini A, Brunetti C, Loreto F, Centritto M, Ferrini F, Tattini M. 2017. Isoprene responses and functions in plants challenged by environmental pressures associated to climate change. Front Plant Sci. 8. https://doi.org/10.3389/fpls.2017.01281.

Gabler F, Nam S, Till S, Mirdita M, Steinegger M, Söding J, Lupas AN, Alva V. 2020. Protein sequence analysis using the MPI Bioinformatics Toolkit. Curr Protoc Bioinformatics. 72(1):e108. https://doi.org/10.1002/cpbi.108.

Gichuru EK, Agwanda CO, Combes MC, Mutitu EW, Ngugi E, Bertrand B, Lasherme P. 2008. Identification of molecular markers linked to a gene conferring resistance to Coffee berry disease (*Colletotrichum kahawae*) in *Coffea arabica*. Plant Pathol. 57:1117–1124. https://doi.org/10.1111/j.1365-3059.2008.01846.x

Gruber M, Söding J, Lupas AN. 2006. Comparative analysis of coiled-coil prediction methods. J Struct Biol. 155(2):140–145. https://doi.org/10.1016/j.jsb.2006.03.009.

Herrera J, Alvarado G, Cortina H, Combes M-C, Romero G, Lashermes P. 2009. Genetic analysis of partial resistance to coffee leaf rust (*Hemileia vastatrix* Berk. and Br.) introgressed into the cultivated *Coffea arabica* L. from the diploid *C. canephora* species. Euphytica. 167:57–67. https://doi.org/10.1007/s10681-008-9860-9.

Hyodo H, Yamakawa S, Takeda Y, Tsuduki M, Yokota A, Nishitani K, Kohchi T. 2003. Active gene expression of a xyloglucan endotransglucosylase/hydrolase gene, XTH9, in inflorescence apices is related to cell elongation in *Arabidopsis thaliana*. Plant Mol Biol. 52(2):473–482. https://doi.org/10.1023/A:1023904217641.

Idárraga SM, Castro AMP, Macea E, Gaitan AL, Rivera L, Cristancho M, Góngora CE. 2011. Sequences and transcriptional analysis of *Coffea arabica* var. Caturra and *Coffea liberica* plant responses to coffee berry borer *Hypothenemus hampei* (Coleoptera: Curculionidae: Scolytinae) attack. J Plant Interact. 7(1):56–70. https://doi.org/10.1080/17429145.2011.573151.

Jones JDG, Vance RE, Dangl JL. 2016. Intracellular innate immune surveillance devices in plants and animals. Science. 354(6316). https://doi.org/10.1126/science.aaf6395.

Kundu S, Sundd M, Jagannadham MV. 2000. Purification and characterization of a stable cysteine protease Ervatamin B, with two disulfide bridges, from the latex of *Ervatamia coronaria*. J Agric Food Chem. 48(2):171–179. https://doi.org/10.1021/jf990661j.

Lamb C, Dixon RA. 1997. The oxidative burst in plant disease resistance. Annu Rev Plant Physiol Plant Mol Biol. 48(1):251–275. https://doi.org/10.1146/annurev.arplant.48.1.251.

Laothawornkitkul J, Paul ND, Vickers CE, Possell M, Mullineaux PM, Hewitt CN, Taylor JR. 2008. The role of isoprene in insect herbivory. Plant Signal Behav. 3(12):1141–1142. https://doi.org/10.4161/psb.3.12.7171.

Liu H, Hu M, Wang Q, Cheng L, Zhang Z. 2018. Role of papain-like cysteine proteases in plant development. Front Plant Sci. 9. https://doi.org/10.3389/fpls.2018.01717.

Liu Y, Zeng Z, Zhang Y, Li Q, Jiang X, Jiang Z, Tang J, Chen D, Wang Q, Chen J, et al. 2021. An angiosperm NLR Atlas reveals that NLR gene reduction is associated with ecological specialization and signal transduction component deletion. Mol Plant, 14(12):2015–2031. https://doi.org/10.1016/j.molp.2021.08.001.

Loivamäki M, Mumm R, Dicke M, Schnitzler J. 2008. Isoprene interferes with the attraction of bodyguards by herbaceous plants. Proc Natl Acad Sci USA. 105(45):17430–17435. https://doi.org/10.1073/pnas.0804488105.

Nagaño TFS, Santos NRS, Santos DMC, Cao EP. 2022. In silico physical mapping of resistance and resistance-associated genes in the SH3 region of the *Coffea canephora* genome. Philipp J Sci. 151(1):107–126. https://doi.org/10.56899/151.01.08.

Nguyen Q-M, Iswanto ABB, Son GH, Kim SH. 2021. Recent advances in effector-triggered immunity in plants: New pieces in the puzzle create a different paradigm. Int J Mol Sci. 22(9):4709. https://doi.org/10.3390/ijms22094709.

Noronha-Wagner M, Bettencourt AJ. 1967. Genetic study of resistance of *Coffea* sp. to leaf rust. I. Identification and behaviour of four factors conditioning disease reaction in *Coffea arabica* to twelve physiologic races of *Hemileia vastatrix*. Can J Bot. 45:2021–2031. https://doi.org/10.1139/b67-220.

Paniagua C, Bilkova A, Jackson PA, Dabravolski SA, Riber W, Didi V, Houser J, Gigli-Bisceglia N, Wimmerová M, Budinská E, et al. 2017. Dirigent proteins in plants: modulating cell wall metabolism during abiotic and biotic stress exposure. J Exp Bot. 68(13):3287–3301. https://doi.org/10.1093/jxb/erx141.

Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. 2018. HMMER web server: 2018 update. Nucleic Acids Res. 46(W1):W200–W204. https://doi.org/10.1093/nar/gky448.

Prigozhin DM, Krasileva KV. 2021. Analysis of intraspecies diversity reveals a subset of highly variable plant immune receptors and predicts their binding sites. Plant Cell. 33(4):998–1015. https://doi.org/10.1093/plcell/koab013.

Sugawara T, Trifonova E, Kochetov AV, Kanayama Y. 2016. Expression of an extracellular ribonuclease gene increases resistance to Cucumber mosaic virus in tobacco. BMC Plant Biol. 16(S3), 246. https://doi.org/10.1186/s12870-016-0928-8.

Tsuda K, Somssich IE. 2015. Transcriptional networks in plant immunity. New Phytol. 206(3):932–947. https://doi.org/10.1111/nph.13286.

van der Vossen H. 2009. The cup quality of disease-resistant cultivars of Arabica coffee (*Coffea arabica*). Exp Agric. 45(3):323–332. https://doi.org/10.1017/S0014479709007595.

van der Vossen HAM, Walyaro DJ. 1980. Breeding for resistance to coffee berry disease in *Coffea arabica* L. II. Inheritance of the resistance. Euphytica. 29:777–791. https://doi.org/10.1007/BF00023225

van der Vossen, HAM, Walyaro DJ. 2009. Additional evidence for oligogenic inheritance of durable host resistance to coffee berry disease (*Colletotrichum kahawae*) in Arabica coffee (*Coffea arabica* L.). Euphytica. 165(1):105–111. https://doi.org/10.1007/s10681-008-9769-3.

van Loon LC. 1985. Pathogenesis-related proteins. Plant Mol Biol. 4(2–3):111–116. https://doi.org/10.1007/BF02418757.

van Wersch S, Li X. 2019. Stronger when together: Clustering of plant NLR disease resistance genes. Trends Plant Sci. 24(8):688–699. https://doi.org/10.1016/j.tplants.2019.05.005.

Wang Y, Garrido-Oter R, Wu J, Winkelmüller TM, Agler MT, Colby TV, Nobori T, Kemen E, Tsuda K. 2019. Site-specific cleavage of bacterial MucD by secreted proteases mediates antibacterial resistance in *Arabidopsis*. Nat Commun. 10(1):2853. https://doi.org/10.1038/s41467-019-10793-x.

Wojtaszek P. 1997. Oxidative burst: An early plant response to pathogen infection. Biochem J. 322(3):681–692. https://doi.org/10.1042/bj3220681.

Yu PA, Santos NRS, Santos DMC, Cao EP. 2021. Bioinformatics survey for candidate resistance genes in the *Coffea canephora Ck-1* gene region. Sci Eng J. [accessed: 2022 Jul 13]. 14(2):226–234. https://scienggj.org/2021-226/.

Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. 2017. A completely reimplemented MPI Bioinformatics Toolkit with a new HHpred server at its core. J Mol Biol. 430(15):2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007.

Zipfel C. 2014. Plant pattern-recognition receptors. Trends Immunol. 35(7):345–351. https://doi.org/10.1016/j.it.2014.05.004.

Mara Angela R. Victoria is a summa cum laude BS Biology graduate from the Institute of Biology, University of the Philippines Diliman. This study was her undergraduate thesis under the Plant Molecular Biology and Genetics Laboratory.

Terrence Ferdinand S. Nagano is a BS Biology graduate and an incoming MS Biology student from the Institute of Biology, University of the Philippines Diliman. He is a research associate and a member of the Plant Molecular Biology and Genetics Laboratory. His work focuses on bioinformatics and genetics of coffee.

Nick Rainier S. Santos <nssantos1@up.edu.ph> is a PhD student in Biology and a faculty member from the Institute of Biology, University of the Philippines Diliman. He is a member of the Plant Molecular Biology and Genetics Laboratory whose work focuses on the genetics of coffee resistance.

Ernelea P. Cao is a Professor Emeritus from the Institute of Biology, University of the Philippines Diliman. She was a former director of the Institute of Biology and the Natural Sciences Research Institute. Currently, she is the head of the Plant Molecular Biology and Genetics Laboratory. Her research interests include cyanobacterial biotechnology and plant genetics.

Daisy May C. Santos is an Assistant Professor from the Institute of Biology, University of the Philippines Diliman. She is the co-laboratory head of the Plant Molecular Biology and Genetics Laboratory. Her research interests include plant molecular biology, particularly coffee genetics and metabolomics.