ARTICLE

Bioinformatics survey for candidate resistance genes in the *Coffea canephora Ck-1* gene region

Patricia Louise A. Yu, Nick Rainier S. Santos*, Daisy May C. Santos, and Ernelea P. Cao

Institute of Biology, College of Science, University of the Philippines, Diliman, Quezon City, 1101

ABSTRACT

Coffea sp. is a major crop, both internationally and locally in the Philippines. The Philippines is a major coffee consumer but a minor producer. Once a leading producer in the past, Philippines' coffee plantations dwindled due to disease outbreaks. In order to help prevent infestations of that scale again, the genetic aspect of resistance needs to be elucidated further. Using the genetic markers Sat207 and Sat235, previous studies have putatively identified a *Ck-1* gene linked to resistance against Coffee Berry Disease (CBD) caused by the fungus *Colletotrichum kahawae*. While a more recent study has determined that Sat207 and Sat235 are located in Chromosome 1, *Ck-1* and its protein product remain uncharacterized. In this study, bioinformatics tools and coffee genomes available at GenBank were used to provide a preliminary list of candidates which may correspond to *Ck-1*. The *C. canephora* genome (GCA_900059795.1) was used as the primary reference. The probable locations of Sat207 and Sat235 in the genome were determined and it was found that the region potentially contains other genetic markers linked to resistance against other diseases and pests. To come up with the candidate genes, the first method involved obtaining genes within the hypothetical *Ck-1* region from PlantEnsembl. The Orange Data Mining tool was used to filter them based on their gene ontology annotations: "defense response", "systemic acquired resistance" and "response to biotic stimuli". Ten potential resistance genes were identified. The second method involved using the online Plant Secondary Metabolite Analysis (PlantiSmash) tool to look for biosynthetic gene clusters (BGC), two of which were identified near the same region. One of these two BGCs could be involved in resistance. These candidate resistance genes could be used in marker-assisted selection of disease resistant coffee varieties.

*Corresponding author
Email Address: nssantos1@up.edu.ph
Date received: August 12, 2021
Date revised: September 26, 2021
Date accepted: October 5, 2021

INTRODUCTION

Coffee beverages come from the genus *Coffea*. They are known for containing caffeine, which is a naturally occurring substance that gives coffee its bitter taste and can stimulate the central nervous system (Gibson and Newsham, 2018). Even though *Coffea* possibly has a hundred species, only a few species circulate around the market such as *Coffea arabica* and *Coffea canephora* (Gibson and Newsham, 2018). *Coffea arabica* is an allotetraploid species, which is a hybrid between *C. canephora* and *C. eugenioides* (Geleta et al, 2012). From 2019-2020, the Philippines has consumed around 3,250,000 coffee bags (197,040,000kg: 60-kg per coffee bag) (Retrieved from https://ico.org/historical/1990%20onwards/PDF/1b-domestic-consumption.pdf on 7 July 2021). In the same period, the production from the Philippines is only 307,000 coffee bags (18,420,000kg) (Retrieved from https://ico.org/historical/

KEYWORDS

Coffee, bioinformatics, *Ck-1*, genes, resistance

1990%20onwards/PDF/1a-total-production.pdf on 7 July 2021). This is a small quantity compared to its neighboring countries such as Indonesia and Vietnam who have exported 11,433,000 coffee bags (685,980,000kg) and 30,487,000 (1,829,220,000kg) coffee bags, respectively.

Local private and public groups have now taken initiatives to improve local production. However, there are a lot of hurdles to overcome. According to the 2017-2022 Philippine Coffee Industry Roadmap (Retrieved from https://www.da.gov.ph/wp-content/uploads/2019/06/Philippine-Coffee-Industry-Roadmap-2017-2022.pdf on 8 September 2021), diseases and pests is one of the main concerns. In the context of coffee, diseases refer to bacterial, viral, and fungal infections while pests refer to insect, larval, and worm infestation. According to the Philippine Bureau of Plant Industry (Retrieved from http://bpi.da.gov.ph/bpi/images/Production_guide/pdf/COFFEE.pdf on 8 September 2021), the main disease and pest in the Philippines are the fungal coffee leaf rust (CLR) and coffee berry borer beetle respectively. Meanwhile, other potential concerns include the fungal coffee berry disease (CBD), which primarily affects African plantations, and the root-knot nematode (RKN) which primarily affects Latin American plantations (van der Vossen et al, 2015). According to a model by Cerda et al (2017), diseases and pests can cause as much as 26% to 38% yield losses.

With the help of marker-assisted selection, breeding specific coffee species may give a coffee plant better resistance to diseases and pests while maintaining quality at the same time. Several studies have identified genetic markers which are linked to resistance (Appendix I). In particular, the partial resistance of *C. arabica* to CBD has been studied by van der Vossen and Walyaro since the 1980s (Silva et al, 2006). Eventually, Gichuru et al (2008) described a hypothetical resistance gene called *Ck-1*. They linked *Ck-1* to the Simple Sequence Repeat (SSR) markers, Sat207 and Sat235. Succeeding studies have supported their findings such as those of Alkimim et al (2017) and Silva et al (2018). While it is known that the gene is linked to these two markers, information regarding the gene itself and its potential protein product remains elusive. With the advances in genomics and bioinformatics, it is possible to at least narrow down the list of potential candidates.

This study aims to identify candidates for the *Ck-1* gene by: 1) determining the probable locations of Sat207 and Sat235 as well as other markers within the available coffee genomes in GenBank; and 2) surveying the protein coding genes within the *Ck-1* gene region. There are currently three *Coffea* genomes in the GenBank: *C. arabica* (GCA_900059795), *C. canephora* (GCF_003713225) and *C. eugenioides* (GCF_003713205). The genome of *C. canephora* was used as the primary reference due to the following: 1) this species is a benchmark for resistance; 2) this is widely utilized globally and locally; and 3) it is diploid making it easier to study.

MATERIALS AND METHODS

Mapping the Genomic Locations of Known Resistance Marker Loci

Numerous published literature were reviewed to search for genetic markers linked to coffee resistance against various diseases and pests (Appendix I). Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) was used to determine the probable location of the gathered genetic markers by determining the locations of their primer binding sites within the *Coffea canephora* genome (GCA_900059795.1). The locations in the *C. arabica* (GCA_003713225.1) and the *C. eugenioides* (GCA_003713205.1) genomes were also determined for additional support (Appendix II). Pairwise BlastN comparison was used to determine whether the region defined by the primer binding sites between genomes corresponded to each other (https://blast.ncbi.nlm.nih.gov/Blast.cgi). For genetic marker locations, only those with strong support were considered for further analysis. A strong support must have any of the following: 1) product size is close to its expected size; 2) primer mismatches are few; 3) corresponding counterparts in either *C. arabica* or *C. eugenioides* are present.

Gathering Candidates for *Ck-1*

According to Brodie et al (2016), affected genes are usually 2Mbps away on average from the single nucleotide polymorphism (SNP) marker associated with them. Extrapolating that data for the other kinds of markers used in this study, an additional 4Mbps extension was done from the region delineated by the primer binding sites of the resistance markers in the coffee genome (APPENDIX II). Hence for this study, the hypothetical *Ck-1* region was set from ~19.5Mbp to ~35Mbp of chromosome 1. In order to determine potential resistance genes, two methods were used. In the first method, the "Gene Stable ID", "Protein Stable ID", "Gene start (bp)", "Gene end (bp)", "GO term accession", "GO term name", "GO term definition", and "GO domain" of protein coding genes within the *Ck-1* region were obtained from the PlantEnsembl Website via its BioMart function (https://plants.ensembl.org/index.html). At the time of the study, the latest Database, "Ensembl Plants Genes 48", was used. These genes were then screened based on the annotations pertaining to resistance using the program Orange (Demsar et al, 2013), which is a data mining toolkit. Information such as "Gene Stable ID", "Protein Stable ID", "GO term accession", "GO term name", "GO term definition", and "GO domain" were assigned as categorical, while "Gene start (bp)" and "gene end (bp)" were assigned as numerical. Next, these were placed under "Features" for further analysis and the conditions were set to filter the genes if they have the following "GO term name": "defense response", "systemic acquired resistance", and "response to biotic stimulus". The ChromoMap program (Anand, 2019) v0.3.1 in RStudio v1.4 Tiger Daylily was used to visualize their locations within the chromosome.

In the second method, Plant Secondary Metabolite Analysis (PlantiSmash) v1.0.0-beta (Kautsar et al, 2017) was used to look for biosynthetic gene clusters which may be involved in the synthesis of secondary metabolites. Chromosome 1 of *C. canephora* (HG974428.1), where *Ck-1* is located, was run through the web site (http://plantismash.secondarymetabolites.org/) using default settings. Some secondary metabolites have been known to be involved in defense. The proteins within these clusters were then run through BlastP for additional information (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

RESULTS AND DISCUSSION

Cui and co-workers (2020) traced the binding sites of Sat207 and Sat235 to Chromosome 1 of *C. canephora*. This was supported by the PrimerBlast results as the probable primer binding sites for Sat207 and Sat235 are located at 29,996,788 to 29,996,877 and 23,700,199 to 23,700,420, respectively in Chromosome 1 of *C. canephora* (Appendix II). Furthermore, the primer binding site for Sat207 is also located in Chromosome 1 for both subgenomes of *C. arabica* and the genome of *C. eugenioides*. Meanwhile, only the *C. canephora* subgenome of *C. arabica* has a primer binding site for Sat235.

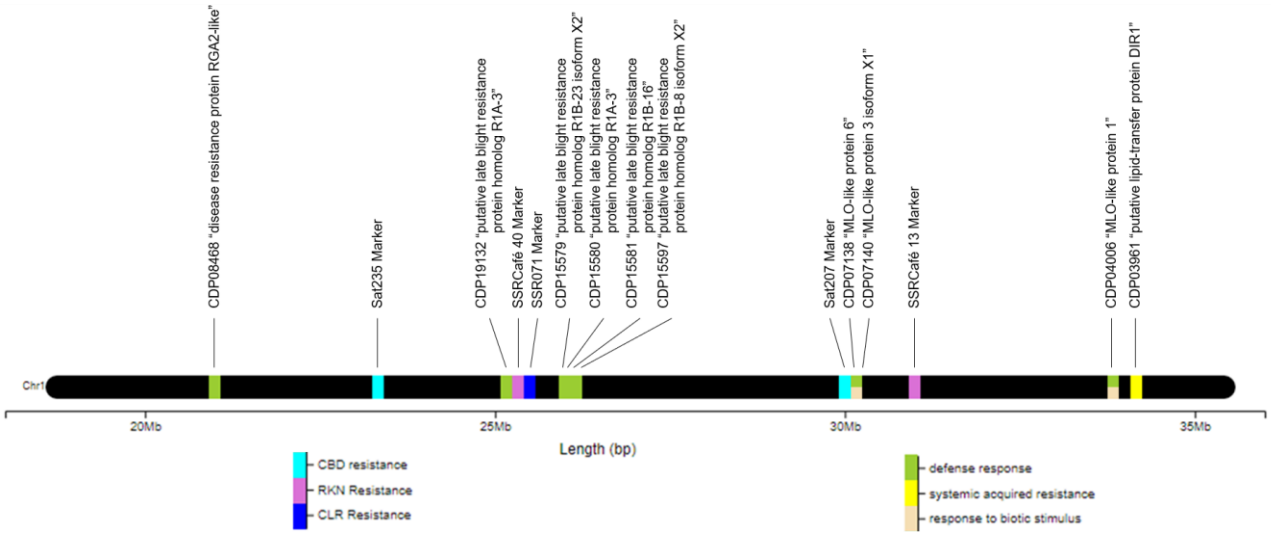

Figure 1: Resistance Genes and Primer Binding Sites of Resistance Markers in the *Ck-1* region of Chromosome 1
In this study, the *Ck-1* region ranges from ~19.5Mbp to ~35Mbp in Chromosome 1 of *C. canephora* containing 1,260 genes. The candidate resistance genes in Table 1 are mapped alongside the primer binding sites of published resistance genetic markers: Sat207 (Gichuru et al, 2008; Alkimim et al, 2017; Silva et al, 2018; Georget et al, 2019) and Sat235 (Gichuru et al, 2008; Gichimu et al, 2014; Alkimim et al, 2017; Silva et al, 2018; Georget et al, 2019) for CBD resistance, SSRCafe 13 (Cristancho & Gaitán, 2008; Pereira et al, 2016) and SSRCafe 40 (López-Gartner et al, 2009; Pereira et al, 2016) for RKN resistance and SSR071 (Pestana et al, 2015) for CLR resistance.

Table 1: Candidate Resistance Genes within the Hypothetical *Ck-1* Region of *Coffea canephora*'s Chromosome 1 (HG974428.1).

Protein	Chromosome	Location		Category	BlastP			Domains
					Query Length	Identity	Percent Identity	
CDP08468	1	21641050	21644176	defense response	916	disease resistance protein RGA2-like (*Coffea arabica*)	92.19%	Rx_N, NB-ARC, RX-CC_like, LRR, PLN03210, RecA-like
CDP19132	1	25445316	25449065	defense response	1249	putative late blight resistance protein homolog R1A-3 (*Coffea arabica*)	98.24%	NB-ARC, RX-CC_like, NB-LRR, Rx_N, PLN03210, CDC6
CDP15579	1	26195001	26199401	defense response	1338	putative late blight resistance protein homolog R1B-23 isoform X2 (*Coffea arabica*)	99.33%	NB-ARC, NB-LRR, RX-CC_like, PLN03210
CDP15580	1	26212619	26216548	defense response	1309	putative late blight resistance protein homolog R1A-3 (*Coffea eugenioides*)	97.40%	NB-ARC, NB-LRR, RX-CC_like, PLN00113
CDP15581	1	26238727	26241468	defense response	913	putative late blight resistance protein homolog R1B-16 (*Coffea eugenioides*)	97.81%	NB-ARC, RX-CC_like, PLN00113
CDP15597	1	26415667	26453362	defense response	2548	putative late blight resistance protein homolog R1B-8 isoform X2 (*Coffea arabica*)	94.44%	NB-ARC, NB-LRR, RX-CC_like
CDP07138	1	30045700	30051164	defense response; response to biotic stimulus	589	MLO-like protein 6 (*Coffea arabica*)	99.83%	Mlo
CDP07140	1	30074771	30080139	defense response; response to biotic stimulus	567	MLO-like protein 3 isoform X1 (*Coffea arabica*)	99.47%	Mlo
CDP04006	1	33563861	33568626	defense response; response to biotic stimulus	468	MLO-like protein 1 (*Coffea arabica*)	95.67%	Mlo
CDP03961	1	33893607	33894704	systemic acquired resistance	111	putative lipid-transfer protein DIR1 (*Coffea eugenioides*)	100%	nsLTP_like, LTP_2

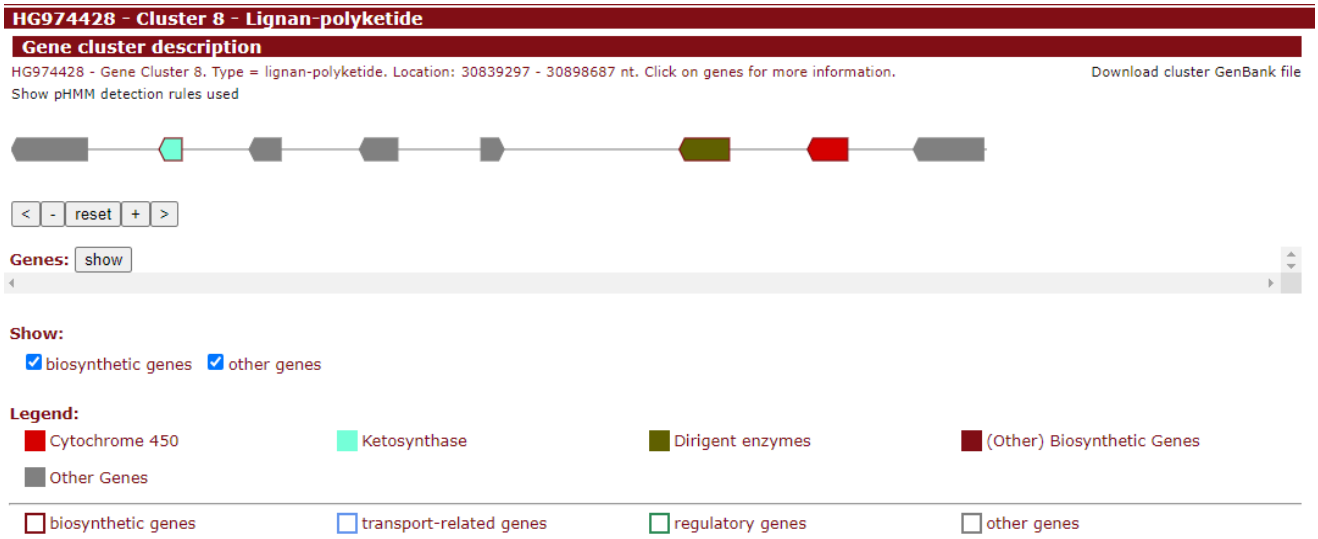

Figure 2: Biosynthetic Gene Cluster 8 in Chromosome 1 of *Coffea canephora*.
PlantiSmash (Kautsar et al, 2017) has identified a BGC within the *Ck-1* region which may be involved in the production of a secondary metabolite it categorized as a lignan based on the presence of a gene whose produce may be a dirigent protein. Lignans have been implicated in conferring resistance against diseases and pests in a variety of ways: inhibition of foreign degradative enzymes, toxicity against herbivores, and enhancement of plant cell wall integrity (Paniagua et al, 2017). It was also categorized as a polyketide because of the ketosynthase gene whose product contains "Chal_sti_synt_C" domain, associated with Chalcone synthase (Yu & Jez, 2008). Chalcone synthase is a polyketide synthase and many of its downstream derivatives are involved in plant defense as well.

Interestingly, the probable primer binding sites of additional resistance SSR markers were found to be within or near the putative *Ck-1* region. These are SSRCafé 40 (López-Gartner et al, 2009; Pereira et al, 2016) and SSRCafe 13 (Cristancho & Gaitán, 2008; Pereira et al, 2016), which are linked to root knot nematode resistance, and SSR071 (Pestana et al, 2015) which is linked to coffee leaf rust resistance. The SSRCafe 13 and SSR071 markers have primer binding sites within the same region for all genomes while the SSRCafé 40 marker does not have a primer binding site within the same region in the *C. eugenioides* subgenome of *C. arabica*. Overall, the relative locations of the primer binding sites of the markers were within Chromosome 1, if present, are consistent across genomes. Lastly, the *Ck-1* region may be involved in broad resistance as the markers are variedly linked to resistance against CBD, CLR, and nematodes.

Candidates based on Gene Ontology Terms

Using the BioMart function of PlantEnsembl, 1,260 genes within the hypothetical *Ck-1* region were obtained. These genes were then screened using Orange based on the gene ontology of their products. Out of the categories present among this batch, 3 were most likely linked to disease resistance, which are "defense response", "systemic acquired resistance", and "response to biotic stimuli". There are ten genes categorized under these based on their potential products: six for defense response exclusively; three for defense response and response to biotic stimuli, and one for systemic acquired resistance. These can be seen in Figure 1 and Table 1.

Gene products classified under "defense response" exclusively include CDP08468, CDP19132, CDP15579, CDP15580, CDP15581, and CDP15597. Defense response refers to the processes and reactions of the plant to the pathogens. These may include the production of antimicrobial proteins, cell death at the site, and structural alterations (Schnek et al, 2000). Andersen et al (2018) divided these into four categories: 1) hypersensitive response (HR) and reactive oxidative species (ROS), 2) enzymes, 3) defensins and thaumatins, and 4) phytoalexins and symbionts. Among these proteins identified, notable domains include Rx_N (N-terminal of the resistance protein against potato virus X), NB-ARC (nucleotide-binding adaptor shared by Apoptotic protease activating factor-1, certain Resistance gene products, and Cell death protein 4), RX-CC_like (RX-CC: coiled-coil of the resistance protein against potato virus X), LRR (leucine-rich repeats), and NB-LRR (nucleotide-binding site and leucine-rich repeats) (Table 1). In relation to the presence of these protein domains, the UniProt Database (https://www.uniprot.org/) description has these proteins involved with the hypersensitive response. The affected areas of the coffee plant undergo the hypersensitive response, where the cells in contact with the invading fungi (CLR or CBD) undergo programmed cell death in order to block them off (Silva et al, 2006). These protein domains are common amongst plant resistance proteins which employ program cell death as part of their mechanism (McHale et al, 2006).

Meanwhile, the BlastP identities of these "defense response" proteins are basically variations of "late blight resistance protein". Siddappa et al (2014) performed expression analysis in Indian potato upon infection with *Phytophthora infestans*, an oomycete. They observed that defense genes, including "putative late blight resistance protein homolog R1A-3", are up-regulated during infection and this suggests its involvement in host-pathogen interaction. Chen et al (2019) observed that "late blight resistance protein homolog R1B-16" is one of the downstream targets of miR396a-5p and -3p, which are miRNA regulators of defense against fungi in tomatoes. Wang et al (2018) discovered markers, which are located within a "putative late blight resistance protein homolog R1A-3" and have been linked to RKN resistance in pepper. Murphy et al (2021) noted that "late blight resistance R1B-16-like" is among the differentially expressed genes in pepper plants upon infection with different strains of tobacco etch virus (TEV). The involvement of late blight resistance proteins against a wide array of diseases and pests supports the hypothesis that one of the candidate late blight resistance genes may be responsible for the broad-spectrum resistance linked to the *Ck-1* region.

In addition to being classified under "defense response", CDP07138, CDP07140 and CDP04006 are also classified under "response to biotic stimuli". Response to biotic stimuli refers to the plant's tendency to respond to environmental triggers. When met with biotic stress, they may activate morphological,

physiological and biochemical mechanisms. Biotic stress may pertain to either invasion or any factor that may be a hurdle to its survival, growth and development. Plants have mechanisms to sense these factors to be able to respond to them and to prevent further damage (Iqbal et al, 2021). The three proteins have been tentatively identified as "Mildew resistance Locus O" proteins, as they contain the MLO domain (Table 1). According to the review by Acevedo-Garcia et al (2014), the *MLO* gene and its protein product has been associated with plant resistance to diseases caused by fungal mildews. *MLO* has been observed to be expressed in infected plant organs. MLO proteins have also been associated with receptor-like kinases (RLK), making it a good indicator for plant disease resistance.

The last gene product under "systemic acquired resistance" is CDP03961 with a BlastP identity of "putative lipid-transfer protein DIR1". Systemic acquired resistance (SAR) is an induced resistance activated by the exposure to elicitors from virulent, avirulent, non-pathogenic microbes, or chemical stimuli (Hartman et al, 2016). Gao et al (2015) discussed the pathways of SAR, one of which highly involves the production of salicylic acid (SA) and its derivatives, such as methyl SA, SA glucose ester and SA-amino acid conjugates. According to Maldonado et al (2002), DIR1 serves as an encoder for apoplastic lipid transfer proteins and is responsible for long distance mobile signaling during SAR. Salicylic acid, which is significant in plant disease resistance, was found to affect DIR1 to continue its downstream pathway. Mutated DIR1 is found in SAR-defective plants (Maldonado et al, 2002).

Candidate Genes based on being part of a Biosynthetic Gene Cluster (BGC)

PlantiSmash identified 9 BGCs spread throughout Chromosome 1 of *C. canephora* (HG974428.1). Only 2 of these were within the hypothetical *Ck-1* region for this study, BGC 7 and BGC 8. The potential product of BGC 7 is generally classified as a "Saccharide". On the other hand, the potential product of BGC 8 is classified as a "Lignan-polyketide" as seen in Figure 2. PlantiSmash categorized the BGC's probable product as a "Lignan" due to the presence of a gene whose product may be a dirigent protein. Paniagua et al (2017) discusses a variety of ways lignan confer resistance against diseases and pest. Some can inhibit degradative enzymes. Some are toxic against herbivores. Some can enhance cell wall integrity, by increasing hydrophobicity and preventing water leakage which can be used as a method of invasion by some pathogens. Some can also serve as substrates for lignin which strengthens the secondary plant cell wall. As part of the hypersensitive response in coffee, the cells undergoing programmed cell death have their cell walls lignified (Silva et al, 2006). Meanwhile, dirigent proteins primarily regulate their biosynthesis (Paniagua et al, 2017). On the other hand, PlantiSmash also categorized it as "Polyketide" because of the "Ketosynthase" gene whose product may contain the "Chal_sti_synt_C" domain. Chalcone synthase is considered as the archetypal polyketide synthase and serves as the committed step toward flavonoid biosynthesis, many of which are involved in plant defense (Yu & Jez, 2008).

CONCLUSION

This study was able to come up with a shortlist of possible identities of the *Ck-1* gene, originally identified as responsible for conferring resistance against the coffee berry disease (CBD). The location based on the primer binding sites of the resistance markers was traced to a region in chromosome 1, in support of data by Cui et al. (2020). Results suggest that the region confers broad resistance as it may also contain other resistance markers for coffee leaf rust (CLR) and root knot nematodes (RKN). After surveying the 19.5-35Mbp region of chromosome 1 of *C. canephora*, ten resistance-associated genes were identified. Based on Gene Ontology annotation, the ten resistance genes are involved in either "defense response", "systemic acquired resistance" and/or "response to biotic stimulus". Meanwhile, a biosynthetic gene cluster which may produce a secondary metabolite classified as either a lignan or a polyketide has also been identified. Both secondary metabolic groups have been implicated in resistance. Overall, this study provides a list of potential resistance genes that may aid future research to improve disease resistance in coffee.

CONTRIBUTIONS OF INDIVIDUAL AUTHORS

PL Yu is the principal author and conducted the bioinformatic analysis. NR Santos came up with the concept, supervised the study, analysis and manuscript writing. DM Santos assisted in the analysis and manuscript writing. EP Cao is the senior author, who helped and supervised in the design, analysis and writing the manuscript.

REFERENCES

ACEVEDO-GARCIA J, KUSCH S, PANSTRUGA R. Magical mystery tour: MLO proteins in plant immunity and beyond. New Phytol 2014; 204(2): 273–281.

ALKIMIM ER, CAIXETA ET, SOUSA TV, PEREIRA AA, OLIVEIRA AC, ZAMBOLIM L, SAKIYAMA NS. Marker-assisted selection provides arabica coffee with genes from other *Coffea* species targeting on multiple resistance to rust and coffee berry disease. Mol Breed 2017; 37(1).

ANAND L. chromoMap: An R package for Interactive Visualization and Annotation of Chromosomes. bioRxiv 2019; 605600.

ANDERSEN E, ALI S, BYAMUKAMA E, YEN Y, NEPAL M. Disease Resistance Mechanisms in Plants. Genes 2018; 9(7): 339.

BALAKIREVA A, ZAMYATNIN A. Indispensable Role of Proteases in Plant Innate Immunity. Int J Mol Sci 2018; 19(2): 629.

BRODIE A, AZARIA JR, OFRAN Y. How far from the SNP may the causative genes be? Nucleic Acids Res 2016; 44(13): 6046–6054.

CERDA R, AVELINO J, GARY C, TIXIER P, LECHEVALLIER E, ALLINNE C. Primary and Secondary Yield Losses Caused by Pests and Diseases: Assessment and Modeling in Coffee. PLoS ONE 2017; 12(1): e0169133.

CHEN L, MENG J, HE XL, ZHANG M, LUAN YS. *Solanum lycopersicum* microRNA1916 targets multiple target genes and negatively regulates the immune response in tomato. Plant Cell Environ 2019; 42(4): 1393-1407.

CRISTANCHO M, GAITÁN Á. Isolation, characterization and amplification of simple sequence repeat loci in coffee. Crop Breed Appl Biotechnol 2008; 8(4): 321-329.

CUI L, HANIKA K, VISSER RGF, BAI Y. Improving Pathogen Resistance by Exploiting Plant Susceptibility Genes in Coffee (*Coffea* spp.). Agronomy 2020; 10(12): 1928.

DEMSAR J, CURK T, ERJAVEC A, GORUP C, HOCEVAR T, MILUTINOVIC M, MOZINA M, POLAJNAR M,

TOPLAK M, STARIC A, STAJDOHAR M, UMEK L, ZAGAR L, ZBONTAR J, ZITNIK M, ZUPAN B. Orange: Data Mining Toolbox in Python. J Mach Learn Res 2013; 14: 2349−2353.

GAO Q-M, ZHU S, KACHROO P, KACHROO A. Signal regulators of systemic acquired resistance. Front Plant Sci 2015; 6: 228.

GELETA M, HERRERA I, MONZÓN A, BRYNGELSSON T. Genetic Diversity of Arabica Coffee (*Coffea arabica* L.) in Nicaragua as Estimated by Simple Sequence Repeat Markers. Sci World J 2012; 1-11.

GEORGET F, MARIE L, ALPIZAR E, COURTEL P, BORDEAUX M, HIDALGO JM, MARRACCINI P, BREITLER JC, DÉCHAMP E, PONCON C, ETIENNE H, BERTRAND B. Starmaya: The First Arabica F1 Coffee Hybrid Produced Using Genetic Male Sterility. Front Plant Sci 2019; 10(1344): 1-10.

GIBSON M, NEWSHAM P. Tea and Coffee. In: Food Science and Culinary Arts. Elsevier Inc., 2018: 353-372.

GICHIMU BM, GICHURU EK, MAMATI GE, NYENDE AB. Occurrence of Ck-1 gene conferring resistance to Coffee Berry Disease in *Coffea arabica* cv. *Ruiru 11* and its parental genotypes. Publ Agric Rural Dev 2014; 2(3): 51-61.

GICHURU EK, AGWANDA CO, COMBES MC, MUTITU EW, NGUGI ECK, BERTRAND B, LASHERMES P. Identification of molecular markers linked to a gene conferring resistance to coffee berry disease (*Colletotrichum kahawae*) in *Coffea arabica*. Plant Pathol 2008; 57(6): 1117–1124.

HARTMAN GL, PAWLOWSKI ML, CHANG H-X, HILL CB. Successful Technologies and Approaches Used to Develop and Manage Resistance against Crop Diseases and Pests. In: Woodhead Publishing Series in Food Science, Technology and Nutrition, Emerging Technologies for Promoting Food Security. Woodhead Publishing, 2016: 43–66.

IQBAL Z, IQBAL MS, HASHEM A, ABD_ALLAH EF, ANSARI MI. Plant Defense Responses to Biotic Stress and Its Interplay with Fluctuating Dark/Light Conditions. Front Plant Sci 2021; 12(631810): 1-22.

KAUTSAR SA, SUAREZ DURAN HG, BLIN K, OSBOURN A, MEDEMA MH. plantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters. Nucleic Acids Res 2017; 45(W1): W55–W63.

LÓPEZ-GARTNER G, CORTINA H, MCCOUCH SR, MONCADA MDP. Analysis of genetic structure in a sample of coffee (*Coffea arabica* L.) using fluorescent SSR markers. Tree Genet Genomes 2009; 5(3): 435–446.

MALDONADO AM, DOERNER P, DIXON RA, LAMB CJ, CAMERON RK. A putative lipid transfer protein involved in systemic resistance signaling in *Arabidopsis*. Nature 2002; 419(6905): 399–403.

MCHALE L, TAN X, KOEHL P, MICHELMORE RW. Plant NBS-LRR proteins: adaptable guards. Genome Biol 2006; 7: 212.

MURPHY JF, HALLMARK HT, RAMARAJ T, SUNDARARAJAN A, SCHILKEY F, RASHOTTE AM. Three Strains of Tobacco etch virus Distinctly Alter the Transcriptome of Apical Stem Tissue in *Capsicum annuum* during Infection. Viruses 2021; 13(5): 741

PANIAGUA C, BILKOVA A, JACKSON P, DABRAVOLSKI S, RIBER W, DIDI V, HOUSER J, GIGLI-BISCEGLIA N, WIMMEROVA M, BUDÍNSKÁ E, HAMANN T, HEJATKO J. Dirigent proteins in plants: modulating cell wall metabolism during abiotic and biotic stress exposure. J Exp Bot 2017; 68(13): 3287–3301.

PEREIRA T, SETOTAW T, SANTOS D, MENDES A, SALGADO S, CARVALHO G, REZENDE R. Identification of microsatellite markers in coffee associated with resistance to *Meloidogyne exigua*. Genet Mol Res 2016; 15(3): gmr.15038054.

PESTANA KN, CAPUCHO AS, CAIXETA ET, DE ALMEIDA DP, ZAMBOLIM EM, CRUZ CD, ZAMBOLIM L, PEREIRA AA, DE OLIVEIRA ACB, SAKIYAMA NS. Inheritance study and linkage mapping of resistance loci to *Hemileia vastatrix* in Híbrido de Timor UFV 443–03. Tree Genet Genomes 2015; 11(4).

SIDDAPPA S, TIWARI J, SINDHU R, SHARMA S, BHARDWAJ V, CHAKRABARTI S, SINGH BP. *Phytophthora infestans* associated global gene expression profile in a late blight resistant Indian potato cv. Kufri Girdhari. Aust J Crop Sci 2014; 8: 215-222.

SILVA MDC, VÁRZEA V, GUERRA-GUIMARÃES L, AZINHEIRA HG, FERNANDEZ D, PETITOT AS, BERTRAND B, LASHERMES P, NICOLE M. Coffee resistance to the main diseases: leaf rust and coffee berry disease. Braz J Plant Physiol 2006; 18(1): 119–147.

SILVA RA, ZAMBOLIM L, CASTRO IS, RODRIGUES HS, CRUZ CD, CAIXETA ET. The Híbrido de Timor germplasm: Identification of molecular diversity and resistance sources to coffee berry disease and leaf rust. Euphytica 2018; 214(9).

VAN DER VOSSEN H, BENOÎT B, BENOÎT B. Next generation variety development for sustainable. Euphytica 2015; 204(2): 243–256.

WANG X, FAZARI A, CAO Y, ZHANG Z, PALLOIX A, MAO S, ZHANG B, DJIAN-CAPORALINO C, WANG L. Fine mapping of the root-knot nematode resistance gene Me1 in pepper (*Capsicum annuum* L.) and development of markers tightly linked to Me1. Mol Breed 2018; 38(4).

YU O, JEZ JM. Nature's assembly line: biosynthesis of simple phenylpropanoids and polyketides. Plant J 2008; 54(4): 750–762.

APPENDIX I. List of coffee resistance markers

The following information were collected from the studies listed below. Most studies involved the use of genetic markers to distinguish resistant coffee trees from the susceptible ones.

Label/s	Type	Primer Sequences		Disease and/or Pest[a]	Linked Gene	References
		Forward	Reverse			
BA-42-21B-r	SCAR	CACACACAGCCTAAGCATCAA	GGATTGACTCGACTCACCAA	CLR	*SH3*	Mahé et al (2007)
BA-48-21O-f	SCAR	ACAGTGAATTCCCCAAGCAC	ACTTGGCAGGCGTAATTGAA	CLR	*SH3*	Mahé et al (2007); Alkimim, et al. (2017)
BA-124-12K-f	SCAR	TGATTTCGCTTGTTGTCGAG	TGCAGATTGATGGCACGTTA	CLR	*SH3*	Mahé et al (2007); Prakash et al. (2011); Alkimim, et al. (2017)
CaRHvII 1	SCAR	CAAGCCGATCATAACTTATC	TGGTGGAGAATTCCTTCAG	CLR	*SH?*	Diola et al. (2011)
CaRHvII 2	SCAR	CATCGCCAGGTCATGTTC	ATCTTCAATCGGAGTAACAC	CLR	*SH?*	Diola et al. (2011); Alkimim, et al. (2017)
CaRHvII 3	SCAR	CTGATACGGGCAATCTTATC	GTAGATCTGGAAGCTCTTC	CLR	*SH?*	Diola et al. (2011); Alkimim, et al. (2017)
CaRHvII 4	SCAR	GATATCGCGCTTATGACAC	CTGATTGGATACTCTGACG	CLR	*SH?*	Diola et al. (2011)
CaRHvII 5	SCAR	CAACACTGGTAGACTCGG	ACTACTGACTTCAGGACAC	CLR	*SH?*	Diola et al. (2011); Alkimim, et al. (2017)
CaRHvII 6	SCAR	CTCAAAGACAATCTCAGTGG	CGTTGGTTGCCGATGATG	CLR	*SH?*	Diola et al. (2011)
EST-SSR009	SSR	TCCGAGTCACATCCCATAAA	GGAGGAAAGTGAAGGAAGAAGA	CLR		Pestana et al. (2015)
EST-SSR032	SSR	AGTCCTTGGCACTTGCTTT	CAGACAACGATCAATACCTTCC	CLR		Pestana et al. (2015)
EST-SSR050	SSR	AAGGAGGAAGAGCGACCAAA	GCGTCAATGTTGGAAGGAAA	CLR		Pestana et al. (2015)
EST-SSR107	SSR	AACAAGAGTTTCTGCCTGTG	TCTCTCGAAGTAGATTGCTCTG	CLR		Pestana et al. (2015)
M24; SSRCafé 20	SSR	GGCTCGAGATATCTGTTTAG	TTTAATGGGCATAGGGTCC	CBD, RKN	*R or k*	Combes et al. (2001); Kiguongo et al. (2014); Pereira et al. (2016)
Sat11, SSR016	SSR	ACCCGAAAGAAAGAACCAAG	CCACACAACTCTCCTCATTC	CLR		Kiguongo et al. (2014); Pestana et al. (2015)
Sat27	SSR	AGGAGGGAGGTGTGGGTGAAG	AGGGGAGTGGATAAGAAGG	CLR		Combes et al. (2001); Motta et al. (2014); van der Vossen et al. (2015)
Sat47; SSRCafé 12; SSRCafé 23,	SSR	TGATGGACAGGAGTTGATGG	TGCCAATCTACCTACCCCTT	RKN		Combes et al. (2001); Poncet et al. (2004); Motta et al. (2014); van der Vossen et al. (2015); Georget et al. (2019)
Sat160	SSR	TGCTTAGGCACTTGATATAGGA	CACGTGCAAGTCACATACTTTA	CLR	*SH3*	Mahé et al (2007); Al-Murish et al. (2013)
Sat207	SSR	GAAGCCGTTTCAAGCC	CAATCTCTTTCCGATGCTCT	CBD	*Ck-1*	Gichuru et al. (2008); Alkimim, et al. (2017); Silva et al. (2018); Georget et al. (2019)
Sat225	SSR	CATGCCATCATCAATTCCAT	TTACTGCTCATCATTCCGCA	CLR	*SH?*	Herrera et al. (2008); Pereira et al. (2011); Georget et al. (2019)
Sat227	SSR	TGCTTGGTATCCTCACATTCA	ATCCAATGGAGTGTGTTGCT	CBD	*R or k*	Kiguongo et al. (2014)
Sat229	SSR	TTCTAAGTTGTTAAACGAGACGCTTA	TTCCTCCATGCCCATATTG	CLR	*SH?*	Herrera et al. (2008); Pereira et al. (2011)
Sat235	SSR	TCGTTCTGTCATTAAATCGTCAA	GCAAATCATGAAAATAGTTGGTG	CBD	*Ck-1*	Gichuru et al. (2008); Gichimu et al. (2014); Alkimim, et al. (2017); Silva et al. (2018); Georget et al. (2019)
Sat244	SSR	GCATGTGCTTTTTGATGTCGT	GCATACTAAGGAAATTATCTGACTGCT	CLR	*SH3*	Mahé et al (2007); Alkimim, et al. (2017); Georget et al. (2019)
Sat281, SSRCafé 24	SSR	TCTTCGTCTTTGCTATTGGT	TATTAACGTCCATCCACACA	CLR	*SH3*	Rovelli et al. (2000); Mahé et al (2007)
SP-M5-SH3	SCAR	TTCACGATCCAAGAAGCA	AGCATGCATTGTAGAAAAA	CLR	*SH3*	Mahé et al (2007)
SP-M8-SH3	SCAR	GAATTCAGCGACGATTG	GATTTGGTGGAAGGGAGC	CLR	*SH3*	Mahé et al (2007)
SP-M16-SH3	SCAR	TTAACTGGAAACTTGGCTTG	ATCTAGCTTTGGAACATCGT	CLR	*SH3*	Mahé et al (2007); Alkimim, et al. (2017)
SP-M18-SH3	SCAR	CTATTTGGTGTGGGAAGTAAC	CTACATCCACGGAGAGAAAC	CLR	*SH3*	Mahé et al (2007)
SSR071	SSR	GCTAAGTTCAATTGCCCCTGT	GGGTTAATTTGATTGCGTGA	CLR		Pestana et al. (2015)
SSR100	SSR	ACCCTTTACTACTTATTTACTCTC	ACATCCCCTTGCCATTTCTTC	CLR		Pestana et al. (2015)
SSRCa034	SSR	TGGACAAGAAATTGAAGTGG	GGGTTTAAATTATCGGGTGT	CLR		Pestana et al. (2015)
SSRCafé 13	SSR	GATCAGAACTTTGAGCTCAGCA	AATGTGGCACGCTAGAAGTG	RKN		Cristancho & Gaitán (2008); Pereira et al. (2016)
SSRCafé 15	SSR	GGGAAGGAATTCTTTCAACTCT	CTTGGAAATTACCATGCAACC	RKN		Cristancho & Gaitán (2008); Pereira et al. (2016)
SSRCafé 19	SSR	TTAAGACATCGGTGCATTCA	TGTGTACTGGGTTTTTTGATGT	RKN		Cristancho & Gaitán (2008); Pereira et al. (2016)
SSRCafé 32	SSR	CCTATCAAACGCATCATGT	CTGTAGGATTGGGTCATTTC	RKN		Pereira et al. (2016)
SSRCafé 39	SSR	ACCAATTCCTGTCAGTCAGG	TGGCCATGAGAATAGGCATC	RKN		López-Gartner et al. (2009); Pereira et al. (2016)
SSRCafé 40	SSR	TAAAGTGGATGCGTCTCCCA	GGATAAGCAAGGAGCTGCAA	RKN		López-Gartner et al. (2009); Pereira et al. (2016)
SSRCafé 41	SSR	CCATTCTAACCAAACCTGTCC	CTCAAACACTTGGGTGTGCA	RKN		López-Gartner et al. (2009); Pereira et al. (2016)

[a]The following abbreviations refer to the corresponding disease/pest: CBD – coffee berry disease (*Colletotrichum kahawae*), CLR – coffee leaf rust (*Hemilea vastatrix*), RKN – root knot nematode (*Meloidogyne* spp.).

APPENDIX II. Determining the probable locations of resistance markers within the available coffee genomes in GenBank

The primers of resistance markers (APPENDIX I) were run through PrimerBlast (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) to determine the locations of their primer binding sites as shown below. Primer binding sites on the same row mean the region delineated by them have high degree of similarity between genomes based on pairwise BlastN comparison (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

Marker	*Coffea canephora*					*Coffea arabica (Coffea canephora)*[a]					*Coffea arabica (Coffea eugenioides)*[a]					*Coffea eugenioides*				
	Chromosome		Location and Size (bp)			Chromosome		Location and Size (bp)			Chromosome		Location and Size (bp)			Chromosome		Location and Size (bp)		
CaRHvII 3	HG974428	1	1,529,849	1,530,289	441	NC_039898	1	2,023,329	2,023,762	434						NC_040035	1	2,047,322	2,047,769	448
SSRCafé 39	HG974428	1	5,390,628	5,390,814	187	NC_039898	1	4,835,609	4,835,780	172	NC_039899	1	3,095,985	3,096,165	181	NC_040035	1	5,596,381	5,596,547	167
Sat235	HG974428	1	23,700,199	23,700,420	222	NC_039898	1	36,409,488	36,409,712	225										
SSRCafé 40	HG974428	1	25,594,253	25,594,829	577	NC_039898	1	38,492,676	38,492,965	290						NC_040035	1	36,001,043	36,001,307	265
SSR071	HG974428	1	25,834,117	25,834,340	224	NC_039898	1	40,304,541	40,304,760	220	NC_039899	1	36,820,804	36,821,023	220	NC_040035	1	40,709,677	40,709,896	220
Sat207	HG974428	1	29,996,788	29,996,877	90	NC_039898	1	43,559,340	43,559,433	94	NC_039899	1	40,297,338	40,297,421	84	NC_040035	1	44,158,696	44,158,779	84
	HG974428	1	30,907,685	30,907,865	181	NC_039898	1	44,494,357	44,494,538	182	NC_039899	1	41,228,713	41,228,891	179	NC_040035	1	52,598,055	52,598,231	177
SSRCafé 13	HG974428	1	30,907,685	30,907,865	181	NC_039898	1	44,494,357	44,494,538	182	NC_039899	1	41,264,112	41,264,290	179	NC_040035	1	52,598,055	52,598,231	177
	HG974428	1	30,907,685	30,908,080	396															
EST-SSR050	HG974429	2	52,846	53,010	165	NC_039900	2	48,348	48,512	165	NC_039901	2	62,043	62,204	162	NC_040036	2	55,624	55,785	162
EST-SSR032	HG974429	2	739,434	739,635	202						NC_039901	2	750,570	750,769	200	NC_040036	2	715,425	715,624	200
											NC_039901	2	6,179,684	6,179,981	298					
CaRHvII 2	HG974429	2	6,266,794	6,267,091	298	NC_039900	2	6,073,548	6,073,845	298	NC_039901	2	6,252,190	6,252,487	298					
SSRCafé 41	HG974429	2	29,304,755	29,304,876	122	NC_039900	2	30,860,959	30,861,084	126	NC_039901	2	34,114,241	34,114,352	112	NC_040036	2	42,648,315	42,648,426	112
	HG974429	2	39,345,702	39,345,788	87	NC_039900	2	56,261,704	56,261,801	98										
SSRCafé 32											NC_039901	2	61,585,160	61,585,234	75	NC_040036	2	68,983,095	68,983,171	77
											NC_039901	2	61,885,343	61,885,417	75					
BA-48-21O-f	HG974430	3	2,709,093	2,709,393	301											NC_040037	3	3,244,209	3,244,509	301
	HG974430	3	3,022,386	3,022,700	315	NC_039902	3	3064088	3064402	315										
BA-124-12K-f	HG974430	3	3,816,628	3,816,857	230											NC_040037	3	4,472,200	4,472,586	387
Sat244	HG974430	3	3,843,980	3,844,281	302											NC_040037	3	4,494,379	4,494,662	284
SP-M8-SH3	HG974430	3	3,862,026	3,862,267	242											NC_040037	3	4,512,389	4,512,623	235
	HG974430	3	8,384,597	8,384,729	133	NC_039902	3	8,169,877	8,170,009	133										
Sat160											NC_039903	3	7,136,657	7,136,828	172	NC_040037	3	9,337,610	9,337,781	172
											NC_039903	3	7,151,565	7,151,667	103	NC_040037	3	9,360,772	9,360,900	129
SP-M5-SH3	HG974430	3	9,462,739	9,462,898	160															
						NC_039902	3	10,836,483	10,836,577	95										
Sat281, SSRCafé 24	HG974430	3	11,972,224	11,972,296	73															
						NC_039902	3	12,250,230	12,250,310	81										
						NC_039902	3	12.508.841	12.509.214	374										
	HG974430	3	12,285,408	12,285,785	378	NC_039902	3	12.571.165	12.571.542	378	NC_039903	3	10,560,828	10,561,219	392	NC_040037	3	13,657,499	13,657,884	386
BA-42-21B-r																NC_040037	3	13,698,643	13,699,034	392
											NC_039903	3	10,654,277	10,654,680	404					
																NC_040039	5	9,034,751	9,035,136	386
	HG974432	5	21,184,265	21,184,418	154															
Sat229						NC_039907	5	37,697,241	37,697,397	157										
						NC_039907	5	37,701,306	37,701,459	154										
											NC_039912	8	3,873,710	3,873,866	157					

CaRHvII 1											NC_039906	5	24710521	24711053	533	NC_040039	5	34,241,287	34,241,819	533
M24, SSRCafé 20	HG974433	6	34,249,062	34,249,209	148	NC_039908	6	37,272,554	37,272,719	166						NC_040040	6	50,300,479	50,300,643	165
Sat11, SSR016	HG974434	7	20,138,071	20,138,214	144											NC_040041	7	34,889,426	34,889,551	126
											NC_039911	7	30,796,441	30,796,578	138					
CaRHvII 6	HG974435	8	1,507,198	1,507,654	457						NC_039912	8	1,522,557	1,523,013	457	NC_040042	8	1,686,655	1,687,111	457
	HG974435	8	7,333,729	7,333,935	207															
CaRHvII 4	HG974435	8	7,364,266	7,364,475	210															
	HG974435	8	7,421,641	7,421,847	207						NC_039912	8	11,919,902	11,920,108	207					
CaRHvII 5	HG974436	9	11,835,818	11,836,076	259	NC_039914	9	10,351,181	10,351,439	259						NC_040043	9	10,019,270	10,019,528	259
	HG974436	9	11,835,818	11,836,076	259	NC_039914	9	10,351,181	10,351,439	259						NC_040043	9	10,257,447	10,257,705	259
EST-SSR107						NC_039918	11	3,952,890	3,953,048	159	NC_039919	11	6,048,674	6,048,842	169	NC_040045	11	1,621,587	1,621,741	155
Sat227	HG974438	11	7,119,637	7,119,934	298															
						NC_039918	11	18,959,212	18,959,391	180										
	HG974438	11	11,245,230	11,245,375	146	NC_039918	11	21,566,779	21,566,924	146	NC_039919	11	28,214,626	28,214,771	146					
SSRCafé 19	HG974438	11	11,289,648	11,289,691	44															
	HG974438	11	11,289,648	11,289,797	150	NC_039918	11	21,611,680	21,611,847	168	NC_039919	11	28,373,374	28,373,543	170	NC_040045	11	28,096,782	28,096,951	170
EST-SSR107	HG974438	11	17,569,119	17,569,285	167															
SSR100						NC_039918	11	27,156,009	27,156,193	185	NC_039919	11	35,056,367	35,056,551	185					
EST-SSR009	HG974438	11	23,602,789	23,602,940	152	NC_039918	11	27,540,662	27,540,821	160	NC_039919	11	35,382,384	35,382,529	146	NC_040045	11	36,670,118	36,670,267	150
Sat27	HG974438	11	25,129,130	25,129,264	135	NC_039918	11	27,851,237	27,851,373	137	NC_039919	11	35,760,481	35,760,621	141	NC_040045	11	37,480,191	37,480,331	141
SSRCafé 19																NC_040045	11	38,127,884	38,128,029	146
CaRHvII 1	HG974438	11	30,633,611	30,634,143	533	NC_039918	11	33,486,942	33,487,474	533						NC_040045	11	44,309,043	44,309,575	533
	HG974438	11	31,211,933	31,212,063	131															
Sat47, SSRCafé 12, SSRCafé 23						NC_039918	11	34,189,403	34,189,522	120										
											NC_039919	11	40,107,790	40,107,941	152					
																NC_040045	11	44,891,734	44,891,867	134
											NC_039919	11	42,075,301	42,075,597	297					
Sat225	HG974438	11	31,710,498	31,710,774	277	NC_039918	11	36,001,243	36,001,506	264	NC_039919	11	42,095,810	42,096,073	264	NC_040045	11	45,619,510	45,619,796	287

[a]For the *C. arabica* genome, the *C. canephora* and the *C. eugenioides* subgenomes were treated as if they are separate genomes.