Chuka University
info@chuka.ac.ke | 020 231 0512 / 020 231 0518

GENOME-WIDE VARIATION AND OCCURRENCE OF BROAD-BASED RESISTANCE TO COFFEE BERRY DISEASE (*Colletotrichum kahawae,* Waller&Brige) WITHIN *Coffea arabica* CULTIVAR, BATIAN

James M. Gimase[1*] Wilson M. Thagana[2] and Chrispine O. Omondi[3]

1. *Kenya Agricultural and Livestock Research Organization (KALRO) - Coffee Research Institute, P. O. Box 4-00232, Ruiru, Kenya. jgimase@gmail.com; +254736865790*
2. *Department of Agricultural Science and Technology, School of Agriculture and Enterprise Development, Kenyatta University, P. O. Box 43844-00100, Nairobi, Kenya. thagana.wilson@ku.ac.ke; +254722697576.*
3. *KALRO - Sugar Research Institute, P. O. Box 44- 40100, Kisumu, Kenya. Email: chrispine.omondi@kalro.org; +254722967180.*

How to cite:
Gimase, James M**.** Wilson M. Thagana and Chrispine O. Omondi. (2021). GENOME-WIDE VARIATION AND OCCURRENCE OF BROAD-BASED RESISTANCE TO COFFEE *BERRY DISEASE (Colletotrichum kahawae, Waller&Brige) WITHIN Coffea arabica CULTIVAR, BATIAN In: Isutsa, D.K. (Ed.) Proceedings of the 7th International Research Conference held in Chuka University from 3rd to 4th December2020, Chuka, Kenya, p.112-121*

.

ABSTRACT

Coffea arabica varieties are characterized by a narrow genetic base, leading to their susceptibility to disease including Coffee Berry Disease (CBD) and hindered development of molecular tools for improvement. CBD epidemics destroys up to 100% of the crop on a susceptible variety when no control measure is applied. Resistance to CBD is conferred by the dominant T and R genes in Hibrido De Timor (HDT) and Rume Sudan (RS) respectively. The cultivar Batian has three phenotypically uniform crosses that may carry T, R or both genes. 27 families from the three crosses were evaluated for genome-wide variation and occurrence of Tand R genes. Genomic DNA was extracted using standard protocols. Genome-wide single nucleotide polymorphism (SNP) markers were obtained through genotyping by sequencing (GBS), which were used in the analysis of variation and occurrence of the R gene. The variation was analyzed by Principal Component Analysis and hierarchical clustering. The occurrence of the T gene was confirmed through agarose gel electrophoresis using Microsatellite primer, Sat 235 while the R gene was by marker sequence search within the GBS result files. The first two PCs accounted for 37% of the total variation. Both the PCA and hierarchical clustering revealed that majority of the genotypes were closely related to SL 28, while two genotypes, CR30EX809 and CR8EX155 were closely related to RS and HDT. All the genotypes were confirmed to carry the T gene, where 17 were homozygous while 12 were heterozygous. Three genotypes were confirmed to carry both T and R genes, therefore with broad-based resistance to CBD and with relatively high diversity. The genotypes with broad-based resistance to CBD are recommended for multiplication, use in breeding of varieties with multiple gene resistance and improvement of diversity in Arabica varieties.

Keywords: Genetic diversity**,** Genotyping by sequencing, T gene, R gene, SL 28.

INTRODUCTION

Coffee belongs to the family Rubiaceae and the genus *Coffea* with over 124 species that have been characterized (Davies et al., 2011), with only two species, *C. arabica* L. and *C. canephora* P. that are of economic importance (Setotaw et al., 2020). The *C. arabica* L. is an allotetraploid species (2n = 2x = 44) that exhibit a diploid-like meiotic behavior (Lashermes and Combes 2018). The *C. arabica* is believed to have been formed as a result of spontaneous hybridization between two diploid species, *C. canephora* and *C. eugenioides* (Lashermes et al., 1999, Lashermes et al. 2011). The species is autogamous with about 10% out-crossing (Bikila et al., 2017).

On the other hand, *C. canephora* is diploid (2n = 2x = 22), highly diverse (Bertrand et al., 2003), with resistance to common disease and thus a good source of genes for disease resistance (Ky et al., 2001). Next-generation sequencing (NGS) technologies, such as genotyping-by-sequencing (GBS), provide markers that are widely used in the genome-wide analysis (Spiniso-Castillo et al., 2020). The GBS approach is more informative than predesigned single nucleotide polymorphism (SNP) arrays especially on wild germplasm as it is unbiased and provides information on rare alleles (Pailles et al., 2017). nGenetic variation is controlled by segregation of multiple genes where, the variances

of individual loci are so small that they cannot be investigated individually and thus the need to analyze sets of many loci (Bikila et al., 2017). The GBS based SNP markers were successfully utilized in the determination of genetic variations in the Coffea genus by Garavito et al. (2016) and Spiniso-Castillo et al. (2020).

The coffee berry disease (CBD) caused by a fungal pathogen *Colletotrichum kahawae* (Waller and Bridge) (Waller et al., 1993), is a key constraint in Arabica coffee production in Africa (Hindorf and Omondi, 2011). The CBD epidemics can destroy up to 100% (Giddisa, 2016) of the developing berries, on a susceptible variety when no control measure is applied (Gichuru et al., 2012). The control of CBD using intensive fungicide spray programs increases the cost of production by up to 40% (Van der Vosen and Walyaro, 2009) and contributes to environmental pollution (Gichuru et al., 2008), hence use of resistant for CBD management. Resistance to CBD is governed by three genes in the varieties Rume Sudan (R genes), HDT (T gene) and K7 (k gene) where R and T are dominant

while k is recessive (Van Der Vossen and Walyaro, 1980). The breeding program for resistance to CBD in Kenya using conventional approaches, led to the introduction of genes for resistance to CBD to *C. arabica* coffee varieties that are susceptible by crossing with donor varieties and backcrossing to standard varieties to restore desirable attributes (Walyaro, 1983). However, this approach takes a long time to develop a coffee variety due to the long juvenile nature of the Coffea genus (Moncada et al., 2016).

The low genetic diversity of *C. arabica* also hinders the identification and selection of superior genotypes using traditional breeding methods (Sousa et al., 2017). To overcome this constraint, molecular markers have been used as a supporting tool to accurately discriminate genotypes and accelerate coffee breeding programs (Sousa et al., 2017). The DNA marker for the T gene was identified by Gichuru et al. (2008) and linked to Simple Sequence Repeats (SSR) primer locus Sat 235, popularly designated as *Ck*-1. This marker was validated by Alkimim et al. (2017) who confirmed that Sat 235 marker co-segregate with the T-gene. A recent study by Gimase et al. (2020a) identified the putative DNA marker for the R gene using Single Nucleotide Polymorphism (SNP) markers.

The cultivar Batian is a pure line that was selected from a three-way and four-way crosses between resistance donor parents and susceptible Arabica varieties, backcrossed to SL 28 (Omondi et al., 2001) and subjected to several generations of selfing to fix the CBD resistant genes and desirable traits from SL 28 (Gichimu et al., 2014). The cultivar comprises three crosses with several families i.e lines derived from plants selected from the same progeny (Fehr, 1991). SL 28 is a Bourbon type single-tree selection that combines high yield, high quality, and drought tolerance but highly susceptible to CBD (Walyaro, 1983). The main objective of this study was to evaluate the genome-wide diversity within the families of the three Batian crosses and their resistance Donor parent HDT and RS, and recurrent parent SL28 using GBS based SNP markers and identify genotypes within Batian with multiple gene resistance to CBD conferred by the T and R genes.

MATERIALS AND METHODS

Area and Materials

This study was carried out at the Kenya Agricultural and Livestock Research Organization – Coffee Research Institute (KALRO-CRI) in Ruiru, Kenya. The materials were 31 genotypes comprising of 27 families from the three crosses of the cultivar Batian, SL 28, HDT, Rume Sudan and Robusta.

Sample collection, Genomic DNA extraction and Genotyping of SNP Markers

Fresh leaves were randomly picked from each of the 30 coffee genotypes (excluding Robusta, since GBS is species-specific) for DNA sample extraction. The LGC genomics plant sample collection kit (www.lgcgenomics.com) was used in sample collection, where 6 disks were cut and placed in each strip of the 96 deep well sample plate, and sent to the Integrated Genotyping Service and Support (IGSS) platform (https://ordering.igssafrica.org/cgibin/order/login.pl) for DNA extraction and genotyping. Genomic DNA was extracted using a standard cetyltrimethylammonium Bromide (CTAB) protocol of Doyle & Doyle (1978). The quality and quantity of the DNA was evaluated by running it through 0.8% agarose gel electrophoresis and concentration adjusted to 50 ηg/µl. The genomic DNA samples were sent to Diversity Arrays Technology (DArT) Pty Ltd, in Canberra-Australia (http://www.diversityarrays.com) for sequencing and identification of SNP markers. The GBS-SNP was performed following the standard protocol as described by Elshire et al. (2011).

Next-generation sequencing was carried out using the Hiseq2500 Illumina platform. The SNP calling was carried out by the DArT-soft14 algorithm within the KDCompute pipeline developed by Diversity Arrays Technology (https://kdcompute.seqart.net/kdcompute/plugins). identical sequences were collapsed into FASTQ call files that were used in the secondary pipeline for DArT P/L's proprietary SNPs calling algorithms (DArT-soft14) pipeline in the processing of the sequence data (Barilli et al., 2018). Since the allotetraploid *C. arabica* open-access genome assembly, with a reliable sorting of homoeologous sequences, is not yet available (Scalabrin et al., 2020), The filtered sequence reads were aligned against the finer and publicly available diploid *Coffea canephora* genome (http://coffee-genome.org/coffeacanephora) as a reference to find the SNP markers in *C. arabica* genome (Sant'Anna et al., 2018) and to determine their corresponding genomic positions.

The SNP marker quality analysis and Evaluation of the Genome-Wide relations

The SNP loci with >30% missing data and rare SNPs with less than 5% minor allele frequencies (MAF) and heterozygosity (Ho) above 90% were removed (Garot et al., 2018). The genetic relationships within the study genotypes were determined using the Principal Component Analysis (PCA) and hierarchical clustering,

implemented within the clustering analysis and distance matrix plugins components respectively of the KDCompute syste (https://kdcompute.seqart.net/kdcompute/plugins).

Genomic DNA extraction, amplification and electrophoresis using SSR primer locus Sat 235

A total of 31 genotypes (27 Batian families, five control) HDT, Robusta, Rume Sudan and susceptible cultivars SL28 were analyzed. Healthy leaves were picked and genomic DNA extracted (Diniz et al., 2005), with minor modifications in the extraction buffers. About 500 milligrams of fresh leaves were ground and transferred to 2 mL Eppendorf tubes. After grinding, 1 mL extraction solution was added and the tubes shaken vigorously for 5 min and immediately put in a 65^0C water bath for 40 min. After which the samples were centrifuged for 5 minutes at 13000 rpm and the supernatant transferred to a new tube, to which 1 mL CIA (chloroform: isoamyl 24:1) was added and the tubes were shaken for 10 min and centrifuged for 5 min at 12000 rpm.

The supernatant was transferred to a tube and the same volume of frozen Isopropanol was added and maintained at -20^0C for 1 hour. The content was centrifuged at 1300 rpm for 5 min, the supernatant discarded and the pellet washed with 70% ethanol. This step was repeated twice and after drying, the pellets were treated with 190 µL TE (Tris- EDTA buffer plus RNAse 10 mg µL-1) for 30 min at 37^0C and 65^0C for 5 min. The DNA was then purified with the addition of 100 µL TE, 100 µL water, 100 µL NaCl 5 M and 100 µL EDTA 0.5 M.

The samples were homogenized and incubated on ice for 30 min and centrifuged for 5 min at maximum speed and isopropanol added. After drying, the pellet for each genotype was diluted in an appropriate amount of TE buffer as per the amount of DNA quantified using a spectrophotometer and stored at 4°C. The extracted DNA quality was determined by running the samples in 1% agarose gel alongside a lambda standard with a known concentration of DNA fragments for comparison and quantification of the samples.

The Polymerase Chain Reaction (PCR) was carried out in total volume of 25 µL, containing 10ng/µL template of genomic DNA, 0.4 µM of Sat 235 SSR primer, 75 µM dNTPs (each), 2.5 µM $MgCl_2$, PCR buffer 1x TBE [75 mM Tris-HCl; 0.5 Na_2 EDTA (pH 8.0)], 20 Mm, Boric acid and 1unit Taq DNA polymerase. Amplification was carried out in a Eurogene thermocycler (TECHNE, UK). The amplification program was one cycle of initial denaturation at 94^0 C for 5 minutes followed by 35 cycles of 30 seconds at 94^0C (denaturation), 30 seconds at 55^0C for primer annealing, and 1 minute and 30seconds at 72^0C for elongations with a final extension at 72^0C for 10 minutes. The amplification products with SSR primer Sat 235 were electrophoresed in 2.3% (w/v) agarose gel with a 1x TBE buffer system and then visualized in a UV light trans-illuminator after staining in 60% ethidium bromide solution.

Confirmation of the occurrence of multiple gene resistance to CBD conferred by of the T and R genes

The presence/absence of the T gene was confirmed by observation of the amplified fragment, based on the standard HDT, Robusta, and SL28 while the genotypes carrying DNA marker for R genes were identified by searching the SNP marker sequences within the GBS-based marker result files of the study genotypes.

RESULTS

The analysis of the SNP marker and Evaluation of the Genome-Wide relations

The GBS analysis generated 2280 good quality SNP markers (MAF>5% and Ho <90%), out of which 1575 were aligned on the 11 Chromosomes based on the *C. canephora* reference genome and that were well distributed within the genome (Figure 1). These are the markers that used in further analysis. The PCA results revealed that majority of the Batian genotypes from the three crosses were grouped together and with SL28, apart from eight genotypes, CR8EX149, 158, 419, 420, 423, 760 together with CR22EX759 and 639 that formed a distance group (Figure 2). Two genotypes CR30EX809 and CR8EX155 were closer to RS (Figure 2). HDT didn't show any relationship with any of the genotypes in the study.

The first two PCs accounted for 37% of the total variation. Similar to the PCA, the hierarchical clustering analysis (dendrogram) revealed a close relationship within all the Batian genotypes. Majority of the individual genotypes within the three crosses clustered together (Figure 3). All genotypes in CR30 clustered together except one (CR30EX809). Similarly, in CR8 two genotypes (CR8EX756 &761) deviated from the rest while CR22 formed two sub-groups where one sub-group was clustered together with SL28 and two of the CR8 genotypes (CR8EX756 &761). The diversity index within the Batian and SL28 was 5% except for CR30EX809 that recorded about 10% with RS and CR8EX155 with 15%. HDT had the highest index of more than 25%.

The occurrence of the T (*Ck-1*) gene in Cultivar R11 and Batian

The amplification products revealed that all the 27 Batian families crosses evaluated carry the DNA fragment for the T-gene (*Ck*-1) that is introgressed from the *Coffea canephora* genome (Plate 1), which was present in HDT and Robusta but absent in SL28 and Rume Sudan. Out of the 27 Batian genotypes, 15 genotypes namely CR8-423, CR8-760, CR8-149, CR8-155, CR8-419, CR22-109, CR22-108, CR22-350, CR22-635CR22-357, CR22-639, CR30-807, CR30-813, CR30-244 and CR30-809 were homozygous for the *Ck*-1 gene while 12 genotypes CR8-136, CR8-154, CR8-761, CR8-420, CR22-759, CR22-114, CR22-353,CR22-111, CR30-812, CR30, CR30-242, CR30-233 and CR30-236 were heterozygous for the *Ck*-1 gene (Pate 1).

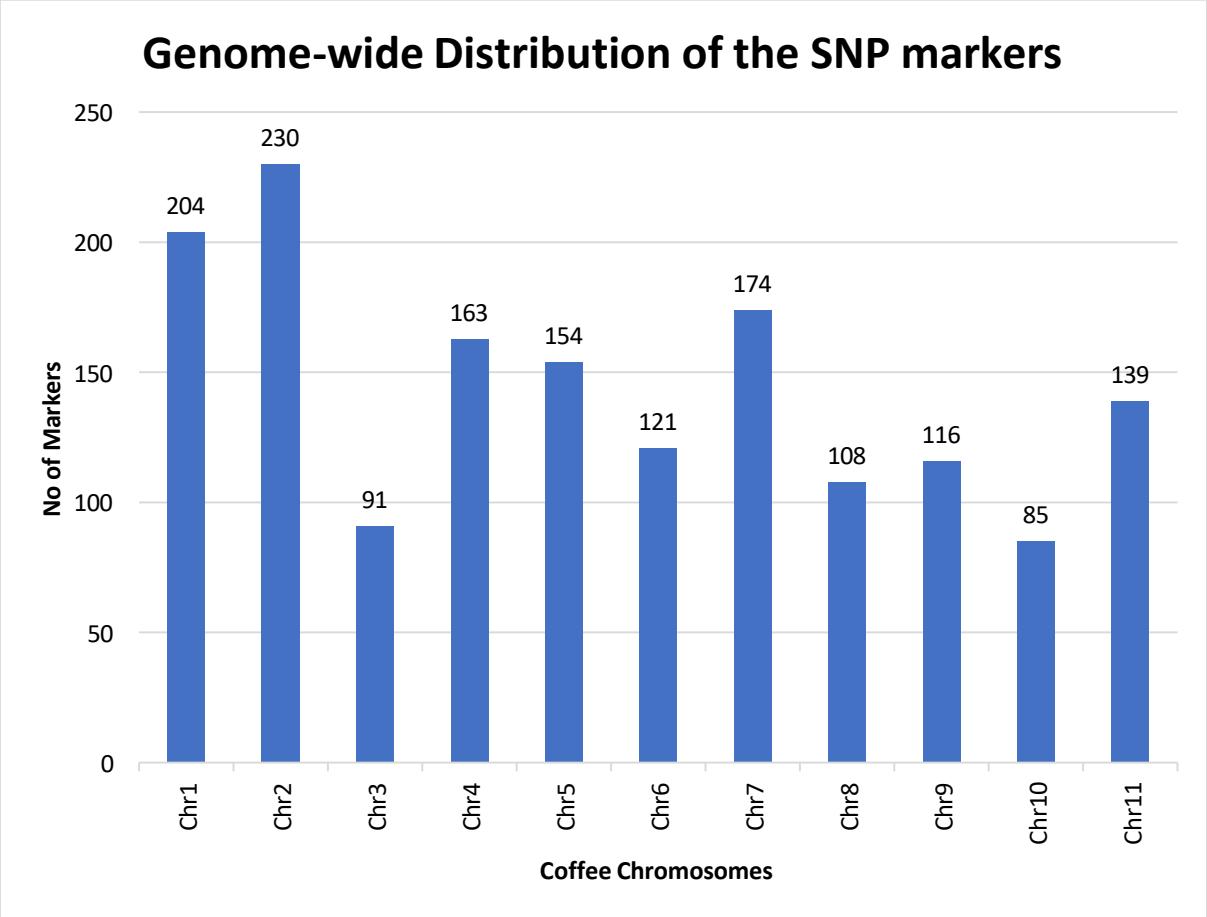


Figure 1: The genome-wide distribution of GBS-based SNP markers

The occurrence of the R-gene within the Cultivar R11 and Batian

A total of 30 genotypes comprising of 27 Batian and three control coffee varieties (Rume Sudan, SL 28 and HDT) were sequenced. The sequences search using the two SNP marker sequences (100025973|F|0-59:T>C-59:T>C and 100034991|F|0-44:C>T-44:C>T), that were significantly associated with CBD resistance through GWAS and QTL mapping (Gimase et al., 2020a&b), was carried out within the sample GBS result files. The search revealed the occurrence of the SNP marker 100034991|F|0-44:C>T-44:C>T in a total of three genotypes (Table 1). The genotypes were CR30-809, CR8-155 and CR8-136 (Table 1). Out of the three genotypes, two genotypes were homozygous, CR30-809 and CR8-155 while CR8-136 was heterozygous for the R-gene marker. This marker was also present in Rume Sudan but absent in SL28 and HDT, therefore polymorphic between the two parents of the mapping genotypes. The SNP marker 100025973|F|0-59:T>C-59:T>C sequence was absent in all the study genotypes. The three genotypes confirmed for the occurrence of the R-gene marker had also been confirmed to carry the T-gene (*Ck*-1).

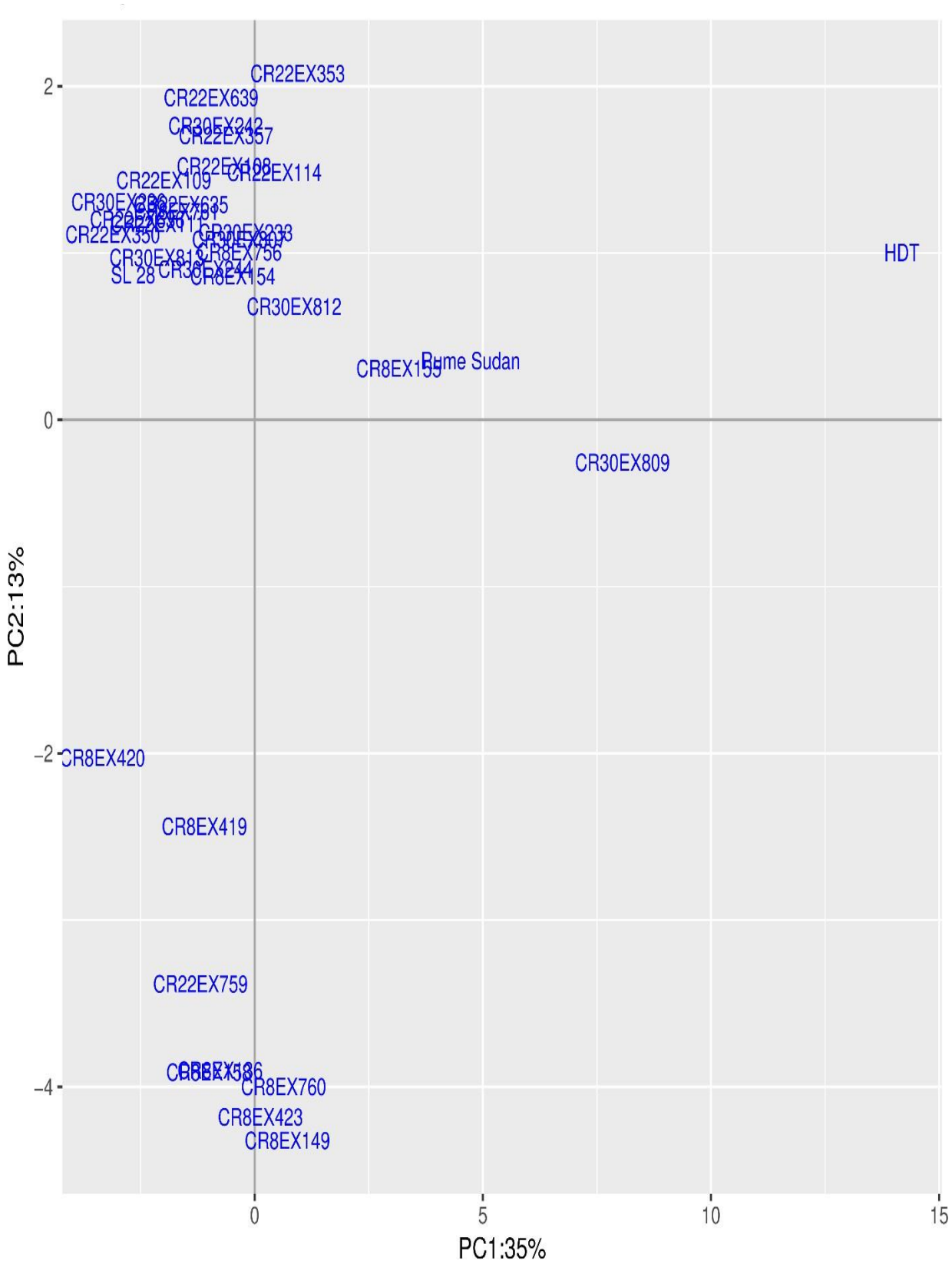


Figure 2: Principal Components Analysis (PCA) plot of the Batian together with the CBD resistance donor parents, HDT, RS the recurrent parent SL28

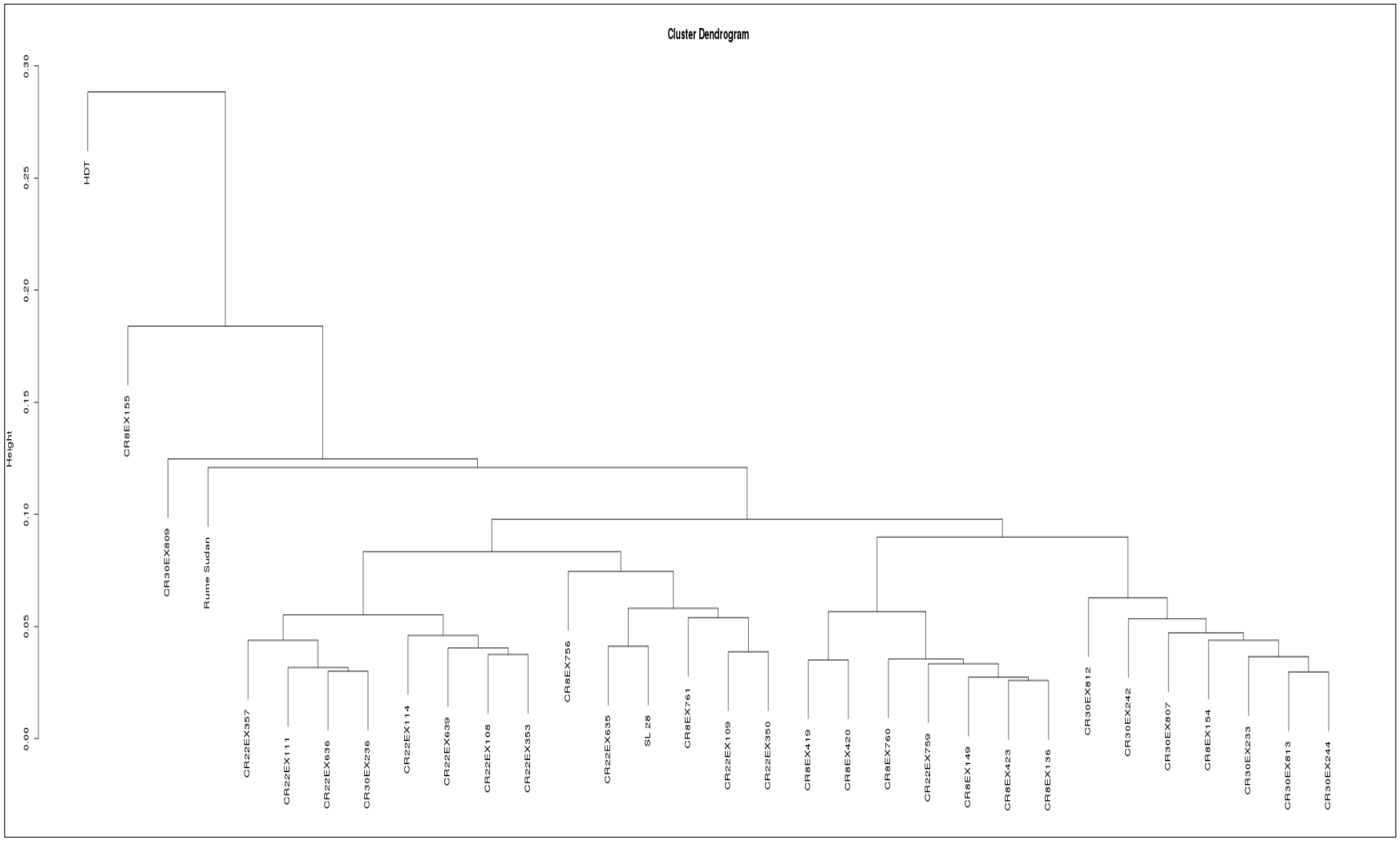
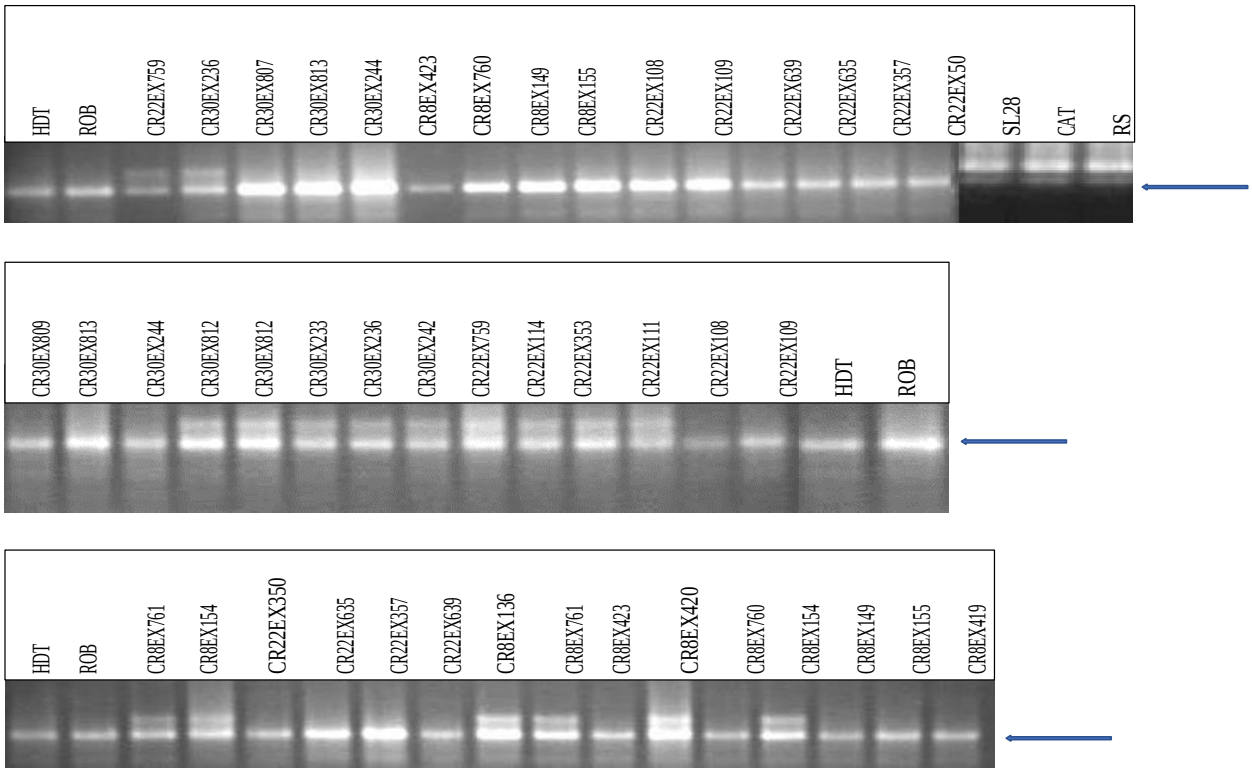


Figure 3: The clustering dendrogram indicating the level of genome-wide variation with the study genotypes



Key: Rob – Robusta, CAT – Caturra, RS – Rume Sudan

Plate 1: Occurrence of T (*Ck*-1) gene fragment within the cultivar Batian Families, indicated by the arrow

Table 16: Occurrence of the R-gene marker within the cultivar R11 and Batian.

Genotype	CR30-809	CR8-136	SL 28	CR8-136	Rume Sudan	HDT
Marker(100034991|F|0-44:C>T-44:C>T)	1	2	0	2	1	0

Key: 1 – Homozygous, 2 – Heterozygous, 0 – Absent.

DISCUSSION

The genome-wide relationship within the study varieties

The genome-wide analysis revealed a narrow genetic base within *C. arabica* coffee varieties, Batian and SL 28 and Rume Sudan, with a diversity index of between 5-10%. This is attributed to the fact that these varieties are derived from a few individual collections and whose subsequent dispersal has progressively narrowed further their genetic base (Baruah et al., 2003; Setotaw et al., 2013). The variety of HDT had a higher diversity index of about 25% with one unique Batian genotype CR8EX155, >15%. The variety HDT is a natural interspecific cross between *C. arabica* and *C. canephora* and usually shows a divergence from commercial cultivars in of the most agronomic traits (Agwanda et al., 1997). The expressed diversity in HDT is due to the introgressed genes from the *C. canephora* genome (Lashermes et al., 1999) and due to filial advancement from the original crosses, the existing progenies contain less of the initial *Coffea canephora* genome, of between 5-29% (Gichuru, 2007).

The study revealed uniformity within individuals of the three crosses that make up the cultivar Batian and revealed further that they are closely related to SL 28. The cultivar Batian was obtained from complex crosses between CBD donor parents (HDT, Rume Sudan and K7) and susceptible varieties (SL 28, SL 34, Bourbon and Tanganyika drought-resistant selections) and then backcrossed to SL28 and selfed to restore and fix genes for desirable attributes of superior beverage quality (Gichimu et al., 2014). From this result, it's most likely that this attribute was successfully restored as SL 28 was closely related to the individual genotypes of Batian.

The occurrence of the T-gene (*Ck*-1) within *C. arabica* cultivar Batian

All the 27 Batian genotypes analyzed were confirmed for the occurrence of the T-gene. The cultivar Batian inherited the T-gene from HDT (a spontaneous cross between *C. canephora* and *C. arabica* and Caturra, a *C. arabica* cultivar, highly susceptible to CBD (Gichuru et al., 2008). The *Ck*-1 gene was confirmed by observation of the amplified fragment using the SSR primer, Sat 235, that co-segregates with the gene. A study by Alkimim et al. (2017) using three CBD resistance genotypes in Brazil revealed the occurrence of *Ck*-1, within the genotypes, where two genotypes were homozygous while one was heterozygous and confirmed that Sat 235 marker co-segregates with the gene. Similarly, a study by Mtenga (2016) reported the occurrence of *Ck*-1 in CBD resistant genotypes from Tanzania and Ethiopia accessions. In the study, Sat 235 could not amplify the T-gene fragment in Rume Sudan since the gene for resistance to CBD is this variety is in a different locus. This therefore confirmed further, the findings by Mtenga (2016) as the *Ck*-1 gene was not amplified in Rume Sudan.

The occurrence of the R-gene within *C. arabica* cultivar Batian

Three genotypes were confirmed for the occurrence of the SNP 100034991|F|0-44:C>T-44:C>T. This, therefore, confirmed that the SNP marker 100034991|F|0-44:C>T-44:C>T as a reproducible marker within *C. arabica* genotypes carrying resistance gene inherited from Rume Sudan. The polymorphic occurrence of this locus in Rume Sudan and SL 28 signifies its ability to discriminate variants in terms of resistance to CBD and its suitability for MAS ((Rouet et al., 2019). The polymorphic genomic loci are used as genetic markers in the determination of the co-segregation of genetic alleles with qualitative traits emanating from populations of crosses or naturally occurring populations (Motazedi et al., 2019). The SNP marker 100034991|F|0-44:C>T-44:C>T was found to be linked to two genes by Gimase et al. (2020b), a finding that is further supported by the inheritance study by Van Der Vossen et al., (1980) that reported that the occurrence of two alleles for R gene in Rume Sudan as R1R1.

CONCLUSION

The study confirmed the close relationship within the *C. arabica* coffee varieties as exhibited by a narrow genetic base, with Batian as a very uniform cultivar and genetically very close to SL 28 and that introgression of C. canephora genome in Arabica enhances genetic diversity as revealed by HDT. The study also revealed that all the genotypes within CBD resistant variety Batian carries the T gene while three carries the R gene and therefore with multiple gene resistance to CBD. The study further confirmed the codominant nature of the DNA marker for T gene and R gene due to their ability to discriminate between homozygous and heterozygous variants within the resistant genotypes.

Acknowledgement

This work was supported by KALRO-CRI and International Livestock Research Institute (ILRI) and the National Research Fund (NRF). The authors thank the Director-General, KALRO (through the Institute Director, CRI), the Director-General ILRI and CEO, NRF. This work is published with the permission of the Institute Director, CRI**.**

REFERENCES

Agwanda, C.O., Lashermes. P., Trouslot. P., Combes. M.C. & Charrier. A. (1997). Identification of RAPD markers for resistance to coffee berry disease, *Colletotrichum kahawae*, in arabica coffee. *Euphytica 97*, 241-248.

Alkimim, E.R., Caixeta, E.T., Sousa, T.V., Pereira, A.V., Oliveira, A.C.B., Zambolim, L., & Sakiyama, N.S. (2017). Marker-assisted selection provides Arabica coffee with genes from other Coffea species targeting multiple resistance to rust and coffee berry disease. *Mol Breeding 37*(6), 1-10. DOI 10.1007/s11032-016-0609-1

Barilli, E., Cobos, M.J., Carrillo, E.A., Carling, J., & Rubiales, D. (2018). A High-Density Integrated DArTseq SNP-Based Genetic Map of *Pisum fulvum* and Identification of QTLs Controlling Rust Resistance. *Front. Plant Sci*. *9*, 167. DOI: 10.3389/fpls.2018.00167

Baruah, A., Naik, V., Hendre, P., Rajkumar, P., & Aggarwal, R.K. (2003). Isolation and characterization of nine microsatellite markers from *Coffea arabica* L. showing wide cross-species amplification. *Molecular Ecology Notes 3*, 647–650. https://doi:10.1046/j.1471-8286.2003. 00544.x

Bikila, B.A., Sakiyama, N.S., & Caixeta, E.T. (2017). SNPs Based Molecular Diversity of Coffea canephora. *Journal of Microbiology & Experimentation* 5(1): 00136

Davis, A.P., Tosh, J., Ruch, N., & Fay, M. (2011) Growing coffee: Psilanthus (Rubiaceae) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and evolutionary history of Coffea. *Bot J Linn Soc 167*, 357–377. https://doi.org/10.3389/fpls.2017.00138

Diniz, L.E.C., Sakiyama, N.S., Lashermes, P., Caixeta, T.E., Oliveira, A.C.B., & Zambolim, E.M. (2005). Analysis of AFLP markers associated with the *Mex*-1 resistance locus in Icatu progenies. *Crop Breeding and Applied Biotechnology 5*, 387-393.

Doyle, J.J., & Doyle, J.L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull. 19*, 11-15

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E., & Mitchell, S. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE 6*(5), e19379. DOI: 10.1371/journal.pone.0019379

Fehr, R.W. (1991). Principles of cultivar development 1: Theory and technique. Macmillan Publishing Company, 333.

Garavito, A., Montagnon, C., Guyot, R., & Bertrand, B. (2016). Identification by the DArTseq method of the genetic origin of the *Coffea canephora* cultivated in Vietnam and Mexico. *BMC Plant Biology 16*, 242. DOI:10.1186/s12870-016-0933-y

Garo,t E., Joet, T., Combes, M.C., & Lashermes, P. (2018). Genetic diversity and population divergences of an indigenous tree (*Coffea mauritiana*) in Reunion Island: the role of climatic and geographical factors. *Heredity, 122*, 833–847. https://doi.org/10.1038/s41437-018-0168-9

Gichimu, B.M., Gichuru, E.K., Mamati, G.E., & Nyende, A.B. (2014). Occurrence of *Ck*-1 gene conferring resistance to Coffee Berry Disease in *Coffea arabica* cv. Ruiru 11 and its parental genotypes. *Journal of Agricultural and Crop Research 2*(3), 51-61.

Gichuru, E.K. (2007). Characterization of genetic resistance of Coffee berry disease (*Colletotrichum kahawae* Waller and Bridge) in Arabica coffee (*Coffea arabica L.*) that is introgressed from *Coffea canephora* (PhD thesis, University of Nairobi, Kenya).

Gichuru, E.K., Agwanda, C.O., Combes, M.C., Mutitu, E.W., Ngugi, E.C.K., Bertran,d B., & Lashermes P. (2008). Identification of molecular markers linked to a gene conferring resistance to Coffee berry disease (*Colletotrichum kahawae*) in C*offea arabica. Plant Pathology, 57*, 1117–1124. DOI:10.1111/j.1365-3059.2008.01846.x

Gichuru, E.K., Ithiru, J.M., Silva, M.C., Pereira, A.P., & Varzea, V.M.P. (2012). Restructured sampling plan enables the characterization of more virulence genes of *Hemileia vastatrix* in Kenya. 24th International Conference on Coffee Science, San José, Costa Rica.

Giddisa, G. (2016). A Review on the Status of Coffee Berry Disease (Colletotrichum kahawae*)* in Ethiopia. *Journal of Biology, Agriculture and Healthcare 6(*19), 140-151.

Gimase, J.M., Thagana, W.M., Omondi, C.O., Cheserek, J.J., Gichimu, B.M., Gichuru, E.K., Ziyomo, C., & Sneller, C.H. (2020a). Genome-Wide Association Study Identify the genetic loci conferring resistance to Coffee Berry Disease (Colletotrichum kahawae) in Coffea arabica var. Rume Sudan. *Euphytica 216*, 86. https://doi.org/10.1007/s10681-020-02621-x

Gimase, J.M., Thagana, W.M., Omondi, C.O., Cheserek, J.J., Gichimu, B.M., & Gichuru, E.K. (2020b). Quantitative trait Loci (QTL) mapping of Resistance to Coffee Berry Disease (Colletotrichum kahawae. Waller&Bridge) in Coffea arabica variety Rume Sudan. *African Journal of Agricultural Research 16*(8), 1184-1194. https://doi.org/10.5897/AJAR2020.14842

Hindorf, H., & Omondi, C.O. (2011). A review of three major fungal diseases of *Coffea arabica* L. in the rainforests of Ethiopia and progress in breeding for resistance in Kenya. *J. Adv. Res. 2*(2), 109-120. DOI: 10.1016/j.jare.2010.08.006

Ky, C.L., Guyot, B., Louarn, J. Harmon, S., & Noirot, M. (2001). Trigonelline inheritance in the interspecific *Coffea pseudozanguebariae* x *C. liberica* var. Dewevrei cross. *Theor Appl Genet* 1*02*, 630-634.

Lashermes, P., & Combes, M.C. (2018) Diversity and genome evolution in coffee. In: Lashermes P (ed) Achieving sustainable cultivation of coffee. BDS Publ., Cambridge, pp 3–20

Lashermes, P. (1999). Molecular characterization and origin of the *Coffea arabica* L. genome. *Mol. Gen. Genet. 261*, 259–266.

Lashermes, P., Combes, M.C., Ansaldi, C., Gichuru, E., & Noir, S. (2011). Analysis of alien introgression in the coffee tree (*Coffea arabica* L.). *Mol Breeding 27*, 223–232. DOI: 10.1007/s11032-010- 9424-2.

Mtenga, J.D. (2016). Diversity, combining ability and Coffee Berry Disease (*Colletotrichum kahawae*) resistance among Ethiopian and Tanzanian Arabica coffee genotypes, (PhD thesis, Sokoine University of Agriculture, Tanzania).

Moncada, M.P., Tovar, E., Montoya, J.C., González, A., Spindel, J., & McCouch, S. (2016). A genetic linkage map of coffee (*Coffea arabica* L.) and QTL for yield, plant height, and bean size. *Tree Genetics and Genomes 12*, 5. DOI: 10.1007/s11295-015-0927-1

Motazedi, E., Maliepaard, C., Finkers, R., Visser, R., & Ridder D (2019). Family-Based Haplotype Estimation and Allele Dosage Correction for Polyploids Using Short Sequence Reads. *Front. Genet. 10*, 335. DOI: 10.3389/fgene.2019.00335

Nemli, S., Asciogul, T.K., Ates, D., Esiyok, D., & Tanyola, M.B. (2017). Diversity and genetic analysis through DArTseq in common bean (*Phaseolus vulgaris* L.) germplasm from Turkey. *Turkish Journal of Agriculture and Forestry 41*, 389-404. DOI:10.3906/tar-1707-89

Omondi, C.O. (1998). Genetic diversity among isolates of *C. kahawae* causing Coffee Berry Disease and their interactions with varieties and breeding populations of Arabica coffee (PhD thesis, University of Nairobi, Kenya).

Omondi, C.O., Ayiecho, P.O., Mwang'ombe, A.W., & Hindorf, H. (2001). Resistance of *Coffea arabica* cv. Ruiru 11 tested with different isolates of *Colletotrichum kahawae*, the causal agent of coffee berry disease. *Euphytica 121*, 19–24.

Pailles, Y., Ho, S., Pires, I.S., Tester, M., Negrão, S., & Schmöckel, S.M. (2017). Genetic Diversity and Population Structure of Two Tomato Species from the Galapagos Islands. *Front. Plant Sci*. *8*,138. doi: 10.3389/fpls.2017.00138

Rouet, C., Lee, E.A., Banks, T., O'Neill, J., LeBlanc, R., & Somers, D.J. (2019). Identification of polymorphism within the *Rosa multifora muRdr1A* gene linked to resistance to multiple races of *Diplocarpon rosae* W. in tetraploid garden roses (*Rosa x hybrida*). *Theoretical and Applied Genetics*. https://doi.org/10.1007/s00122-019-03443-9

Sant'Ana, G.C., Pereira, L.F..P, Po,t D., Ivamoto, S.T., Domingues, D.S., & Ferreira, R.V. (2018) Genome-wide association study reveals candidate genes influencing lipids and diterpenes contents in *Coffea arabica* L. *Scientific Reports 8*, 465. DOI:10.1038/s41598-017-18800-1

Scalabrin, S., Toniutti, L., Gaspero, G., Scaglione, D., Magris, G., & Michele, M. (2020). A single polyploidization event at the origin of the tetraploid genome of *Coffea arabica* is responsible for the extremely low genetic variation in wild and cultivated germplasm. *Scientific Reports 10*, 4642. https://doi.org/10.1038/s41598- 020-61216-7

Setotaw, T.A., Caixeta, E.T., Pereira, A.A., Oliveira, A.C.B., Cruz, C.D., Zambolim, E.M., Zambolim, L., Sakiyama, N.S. (2013). Coefficient of Parentage in *Coffea arabica* L. Cultivars Grown in Brazil. *Crop science 53*, 1237–1247. doi: 10.2135/cropsci2012.09.0541

Setotaw, T.A., Caixeta, E.T., Zambolim, E.M., Sousa, T.V., Pereira, A.A., Baião, A.C., Cruz, C.D., Zambolim, L., & Sakiyama, N.S. (2020). Genome Introgression of Híbrido de Timor and Its Potential to Develop High Cup Quality *C. arabica* Cultivars. *Journal of Agricultural Science 12*(4), 64-76. https://doi.org/10.5539/jas.v12n4p64

Sousa, T.V., Caixeta, E.T., Alkimim, E.R., Oliveira, A.C.B., Pereira, T.V., & Sakiyama, N.S. (2017). Population structure and genetic diversity of coffee progenies derived from Catuaí and Híbrido de Timor revealed by genome-wide SNP marker. *Tree Genetics & Genome 13*,124. DOI: 10.1007/s11295-017-1208-y

Spinoso-Castillo. J.L., Escamilla-Prado, E., Aguilar-Rinco´n, V.H., Ramos, VM., Santos, G.G., Pe´rez-Rodrı´guez, P., & Corona-Torres, T. (2020). Genetic diversity of coffee (Coffea spp.) in Mexico evaluated by using DArTseq and SNP markers. *Genet Resour Crop Evol.* https://doi.org/10.1007/s10722-020-00940-5

Van der Vossen, H.A.M., & Walyaro, D.J.A. (1980). Breeding for resistance to coffee berry disease in *Coffea arabica* L. II. Inheritance of the resistance. *Euphytica 29*, 777-791.

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* 165:105–111. DOI 10.1007/s10681-008-9769-3

Van Der Vossen, H.A.M. (2018). Developing varieties of Arabica coffee in: Lashermes, P. (ed.), *Achieving sustainable cultivation of coffee*, Burleigh Dodds Science Publishing, Cambridge, UK, 1-33. http://dx.doi.org/10.19103/AS.2017.0022.05

Waller, J.M., Bridge, P.D., Black, R.L., & Hakiza, G. (1993). Differentiation of the coffee berry disease pathogen.

Mycological Research 97, 989-994.

Walyaro DJA (1983). Considerations in breeding for improved yield and quality in Arabica Coffee (*Coffea arabica*

L.) (PhD thesis, Wageningen, the Netherlands).

****