



OPEN ACCESS

EDITED BY

Elisabetta Mazzucotelli,
Council for Agricultural and
Economics Research (CREA), Italy

REVIEWED BY

Yuan-Ming Zhang,
Huazhong Agricultural University,
China
Maria Celeste Gonçalves-Vidigal,
Universidade Estadual de Maringá,
Brazil

*CORRESPONDENCE

Luiz Filipe Protasio Pereira
filipe.pereira@embrapa.br

SPECIALTY SECTION

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

RECEIVED 08 July 2022

ACCEPTED 22 September 2022

PUBLISHED 18 October 2022

CITATION

Ariyoshi C, Sant'ana GC, Felicio MS,
Sera GH, Nogueira LM,
Rodrigues LMR, Ferreira RV,
da Silva BSR, Resende MLV,
Destéfano SAL, Domingues DS and
Pereira LFP (2022) Genome-wide
association study for resistance to
Pseudomonas syringae pv. *garcae*
in *Coffea arabica*.
Front. Plant Sci. 13:989847.
doi: 10.3389/fpls.2022.989847

COPYRIGHT

© 2022 Ariyoshi, Sant'ana, Felicio, Sera,
Nogueira, Rodrigues, Ferreira, da Silva,
Resende, Destéfano, Domingues and
Pereira. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Genome-wide association study for resistance to *Pseudomonas syringae* pv. *garcae* in *Coffea arabica*

Caroline Ariyoshi^{1,2}, Gustavo César Sant'ana³,
Mariane Silva Felicio^{2,4}, Gustavo Hiroshi Sera²,
Livia Maria Nogueira^{1,2}, Lucas Mateus Rivero Rodrigues⁵,
Rafaelle Vecchia Ferreira^{1,2},
Bruna Silvestre Rodrigues da Silva^{1,2},
Mário Lúcio Vilela de Resende⁶,
Suzete Aparecida Lanza Destéfano⁷,
Douglas Silva Domingues⁸ and Luiz Filipe Protasio Pereira^{1,2,9*}

¹Programa de pós-graduação em Genética e Biologia Molecular, Universidade Estadual de Londrina (UEL), Centro de Ciências Biológicas, Londrina, Brazil, ²Área de Melhoramento Genético e Propagação Vegetal, Instituto de Desenvolvimento Rural do Paraná (IDR-Paraná), Londrina, Brazil, ³Tropical Melhoramento & Genética (TMG), Londrina, Brazil, ⁴Programa de pós-graduação em Ciências Biológicas (Genética), Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP), Instituto de Biociências, Campus de Botucatu, Botucatu, Brazil, ⁵Centro de Café Alcides Carvalho, Instituto Agronômico (IAC), Campinas, Brazil, ⁶Departamento de Fitopatologia, Universidade Federal de Lavras (UFLA), Lavras, Brazil, ⁷Instituto Biológico (IB), Campinas, Brazil, ⁸Departamento de Genética, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo (USP), Piracicaba, Brazil, ⁹Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA-Café), Brasília, Brazil

Bacteria halo blight (BHB), a coffee plant disease caused by *Pseudomonas syringae* pv. *garcae*, has been gaining importance in producing mountain regions and mild temperatures areas as well as in coffee nurseries. Most *Coffea arabica* cultivars are susceptible to this disease. In contrast, a great source of genetic diversity and resistance to BHB are found in *C. arabica* Ethiopian accessions. Aiming to identify quantitative trait nucleotides (QTNs) associated with resistance to BHB and the influence of these genomic regions during the domestication of *C. arabica*, we conducted an analysis of population structure and a Genome-Wide Association Study (GWAS). For this, we used genotyping by sequencing (GBS) and phenotyping for resistance to BHB of a panel with 120 *C. arabica* Ethiopian accessions from a historical FAO collection, 11 *C. arabica* cultivars, and the BA-10 genotype. Population structure analysis based on single-nucleotide polymorphisms (SNPs) markers showed that the 132 accessions are divided into 3 clusters: most wild Ethiopian accessions, domesticated Ethiopian accessions, and cultivars. GWAS, using the single-locus model MLM and the multi-locus models mrMLM, FASTmrMLM, FASTmrEMMA, and ISIS EM-BLASSO, identified 11 QTNs associated with resistance to BHB. Among these QTNs, the four with the highest values of association for resistance to BHB are linked to *g000* (Chr_0_434_435) and *g010741* genes, which are predicted to encode a serine/threonine-kinase

protein and a nucleotide binding site leucine-rich repeat (NBS-LRR), respectively. These genes displayed a similar transcriptional downregulation profile in a *C. arabica* susceptible cultivar and in a *C. arabica* cultivar with quantitative resistance, when infected with *P. syringae* pv. *garcae*. However, peaks of upregulation were observed in a *C. arabica* cultivar with qualitative resistance, for both genes. Our results provide SNPs that have potential for application in Marker Assisted Selection (MAS) and expand our understanding about the complex genetic control of the resistance to BHB in *C. arabica*. In addition, the findings contribute to increasing understanding of the *C. arabica* domestication history.

KEYWORDS

coffee domestication, bacterial halo blight, plant disease, single-locus genome-wide association, multi-locus genome-wide association

Introduction

Coffee (*Coffea* sp.) is one of the most important crops in the world, especially for tropical countries. Although the genus *Coffea* contains 124 species, only two have economic importance, *C. arabica* L. and *C. canephora* Pierre, corresponding to 59.96 and 40.04% of the world production, respectively (Davis et al., 2011; International Coffee Organization, 2020). *C. arabica* is an autogamous and tetraploid species ($2n = 4x = 44$) with two diploid ancestors: *C. canephora* and *C. eugenioides* (Lashermes et al., 1999; Clarindo and Carvalho, 2008). It was originated in South-West Ethiopia and South Sudan region and traditionally is cultivated at high altitudes. On this conditions, *C. arabica* is known to provide better beverage quality (Bertrand et al., 2006; Davis et al., 2012; Worku et al., 2018).

Disease incidence is one of the main problems in the cultivation of *C. arabica*, affecting both production and quality. Among the bacterial diseases, we highlight the bacterial halo blight (BHB) caused by *Pseudomonas syringae* pv. *garcae* (Amaral et al., 1956). This disease is prevalent at high altitudes, due to cold winds (Maciel et al., 2018). Worldwide, there are reports of its occurrence in Kenya, Ethiopia, China, and Brazil (Ramos and Shavdia, 1976; Adugna et al., 2012; Xuehui et al., 2013; Rodrigues et al., 2017a). BHB management, based on windbreaks, is not always possible on coffee farms and management based on chemicals is environmentally undesirable. Both solutions increase the production costs and, generally, are not highly efficient in field conditions. Thus, the use of cultivars resistant to BHB is a better alternative for coffee producers.

During the plant-pathogen interaction, once the pathogens overcome the mechanical defense barriers of the plants, resistant plants activate receptors that can recognize the pathogen,

activating signaling pathways that drive the defense response gene expression. These defense genes can reduce pathogen infection and colonization through their antimicrobial and cell wall enforcement properties (Andersen et al., 2018). At the first level of recognition, pathogen conserved molecular patterns, known as pathogen-associated molecular patterns (PAMPs), can be detected by host proteins, known as pattern recognition receptors (PRRs), which are capable of activating PAMP-triggered immunity (PTI). The plasma membrane localized plant receptors, such as receptor-like kinases (RLKs) and receptor-like proteins (RLPs) are examples of PRR proteins (Macho and Zipfel, 2015). However, some specialized pathogens produce effectors that can inhibit the PTI (Boller and Felix, 2009).

The second level of recognition involves R proteins capable of detecting effectors and activates effector-triggered immunity (ETI) (Jones and Dangl, 2006; Yuan et al., 2021). Among the R proteins, the class with the greatest representativeness is the nucleotide binding sites leucine-rich repeats (NBSs-LRRs) (Coll et al., 2011). The NBS-LRR and effector interaction can occur directly or indirectly. In the indirect way, the NBS-LRR monitors the binding of the effector with a “decoy” protein (Cui et al., 2015). Both ETI and PTI can activate the hypersensitive response (HR) (Kushalappa et al., 2016). Furthermore, the ETI robustness depends on the PTI (Yuan et al., 2021). The HR is considered as qualitative resistance controlled by few genes that have major effect (R genes) (Jones et al., 2016), and the qualitative resistance is defined by two categories: resistant and susceptible plants (Corwin and Kliebenstein, 2017).

The third line of defense is the absence or weakening of PTI and/or ETI, without HR (Niks et al., 2015). In this type of resistance, some antimicrobial and cell wall enforcement products are unavailable or available in reduced amounts, which allows the advance of pathogens (Boyd et al., 2012).

This type of resistance can occur in a constitutive or an induced way and correspond to quantitative resistance (Kushalappa and Gunnaiah, 2013). In the quantitative resistance, the individuals can be divided into a continuous distribution between the resistance and susceptibility, which is related to the number and amount of defense products available (Kushalappa and Gunnaiah, 2013; Corwin and Kliebenstein, 2017). The genes controlling quantitative resistance are located at multiple loci with variable effects (Pilet-Nayel et al., 2017).

Due to its reproductive biology and relatively recent origin, *C. arabica* has low genetic diversity (Lashermes et al., 2000). In addition, most *C. arabica* cultivars are derived from only two varieties Typica and Bourbon, a factor responsible for a genetic bottleneck during the cultivar development (Anthony et al., 2001). Consequently, several cultivars are susceptible to diseases and pests (van der Vossen et al., 2015).

In contrast to the lack of genetic diversity in *C. arabica* cultivars, greater diversity is found in *C. arabica* Ethiopian accessions. Due to this diversity, a survey organized by the Food and Agriculture Organization (FAO) between 1964 and 1965, collected these accessions in Ethiopia. The collection took place in two regions, the East and West of the Great Rift Valley (FAO: Food and Agriculture Organization, 1968). During the *C. arabica* domestication process, there was, probably, a movement of plants from the West to the East of the Great Rift Valley (Montagnon and Bouharmont, 1996; Scalabrin et al., 2020).

Several studies, based on this *C. arabica* Ethiopian germplasm, identified sources of desirable traits, such as resistance to BHB (Mohan et al., 1978); resistance to nematodes *Meloidogyne paranaensis*, *M. incognita*, and *M. exigua* (Anzueto et al., 2001; Fatobene et al., 2017; Holderbaum et al., 2020); resistance to coffee berry disease (van der Vossen and Walyaro, 2009); drought tolerance (Queiroz-Voltan et al., 2014); biochemical compounds which influence beverage quality (Bertrand et al., 2006); low caffeine content (Silvarolla et al., 2004); and other chemical compounds (Scholz et al., 2016).

A large number of genetic recombination historical events present among unrelated individuals allows high-resolution mapping through Genome-Wide Association Studies (GWAS) (Migicovsky and Myles, 2017). Recent association maps, using *C. arabica* Ethiopian accessions, identified alleles associated with the chemical composition of green grains (chlorogenic acids levels, caffeine, total sugars, sucrose, lipids, proteins, and tannins) (Sant'ana et al., 2018).

In this work, to explore the genetic diversity of *C. arabica* Ethiopian accessions as well as of selected cultivars, we perform a GWAS for *C. arabica* and *P. syringae* pv. *garcae* interaction. We also identify and validate candidate genes involved in qualitative resistance to BHB by expression analysis and discuss whether the QTNs associated with resistance to BHB may be inserted in genomic regions that were relevant during the *C. arabica* domestication history.

Materials and methods

Plant materials

A total of 132 *C. arabica* genotypes were used, including 120 Ethiopian accessions from the FAO collection (FAO: Food and Agriculture Organization, 1968), the BA-10 genotype, and 11 cultivars: IPR 99, IPR 100, IPR 101, IPR 102, IPR 103, IPR 104, IPR 105, IPR 107, Catuaí Vermelho IAC 99, Bourbon Vermelho, and IAPAR 59 (Supplementary Table 1).

Phenotype data, for resistance to BHB, were accessed from two previous works conducted at the Instituto de Desenvolvimento Rural do Paraná (23° 22' S, 51° 10' W, 585 m asl) in Londrina, Paraná State, Brazil (Mohan et al., 1978; Ito et al., 2008). The phenotyping of *C. arabica* Ethiopian accessions, Bourbon Vermelho cultivar, and BA-10 genotype was carried out in the field with two replicates in plants that were approximately 1-year-old. The isolate number 586 of *P. syringae* pv. *garcae* was used and evaluations were carried out 17 and 35 days after inoculation (dai). The two evaluations obtained the same results (Mohan et al., 1978). The phenotyping of IPR 99, IPR 100, IPR 101, IPR 102, IPR 103, IPR 104, IPR 105, IPR 107, IAPAR 59, and Catuaí Vermelho IAC 99 cultivars were also evaluated in the field, with nine replications in plants approximately 10-months-old, with natural infections (Ito et al., 2008). In those studies, a scale with a score from 0 to 5 was used, where 0 corresponds to highly resistant plants, 1 to resistant, 2 to moderately resistant, 3 to moderately susceptible, 4 to susceptible, and 5 to highly susceptible (Supplementary Figure 1).

The current work used the GBS data from DNA samples of these 132 *C. arabica* accessions (Sant'ana et al., 2018). This single-end sequencing was performed using the restriction enzyme *Pst*I, through the platform Illumina HiSeq2000 in partnership with the Genomic Diversity Facility LIMS, at Cornell University (Ithaca, NY - USA). This work used the GBS data (1,083,002 tags) aligned to the *C. arabica* Et39 reference genome (Salojärvi et al., 2021), performed by Felicio (2020) which allowed identification of 159,000 SNPs.

SNP filtering and linkage disequilibrium analyses

The 159,000 SNPs were filtered using TASSEL software version 5.2.53 (Bradbury et al., 2007). The filtering was performed with the parameters of the minimum allele frequency greater than 0.05 (MAF > 0.05), call rate > 0.8, and removal of monomorphic and indel sites. The data were also imputed by LD kNNi software, which is based on the k-nearest neighbor method (Money et al., 2015). The SNPs obtained after this filter quality were used for the population structure study and for the GWAS.

For the linkage disequilibrium (LD) analysis, the panel of 159,000 SNPs was filtered through the software TASSEL, version 5.2.53. The filters used were: MAF > 0.05, call rate > 0.25, and removal of monomorphic and indel sites. The SNPs corresponding to scaffolds that did not have a defined position during the genome assembly (Chr zero) were also removed. The squared correlation coefficients (r^2) were calculated on sliding windows with 50 adjacent SNPs, in TASSEL version 5.2.53, and used to evaluate the extent of LD decay. The value of LD distance (in bp) was evaluated by a non-linear regression, at $r^2 = 0.2$, in R studio software (RStudio Team, 2020).

Population structure study

A dissimilarity matrix was built and used for hierarchical analyses, with the complete linkage method and the Euclidean distance, by the R package cluster, and visualized graphically through the factoextra package (Maechler et al., 2019; RStudio Team, 2020; Kassambara and Mundt, 2020).

A Principal Component Analysis (PCA) was performed using TASSEL software, version 5.2.53 (Bradbury et al., 2007). The dispersion plot from the first and second Principal Components (PCs) was conducted using the R package cluster and visualized graphically using the package ggfortify (Horikoshi and Tang, 2016).

Population structure was also analyzed using the Bayesian method, implemented in STRUCTURE 2.3.4 (Pritchard et al., 2000). We predefined the number of genetic clusters K from 2 to 10 in the population. The program was set on 10,000 as burn-in iteration, followed by 10,000 Markov chain Monte Carlo (MCMC) replications with 20 runs per K. We used the ΔK criterion in Structure Harvester software to estimate the upper-most level of structure (Evanno et al., 2005; Earl and von Holdt, 2012).

GWAS

The associative mappings were conducted using a single-locus model and four multi-locus models. The single-locus model used was the mixed linear model (MLM), through the software TASSEL 5.2.53 (Yu et al., 2006; Bradbury et al., 2007). For MLM, the association threshold established, to reduce false positive associations, was Bonferroni. For the multi-locus approach, we used four models within the R package mrMLM: “random-SNP-effect MLM (rMLM) and a multi-locus mrMLM” (mrMLM) (Wang et al., 2016), FASTmrMLM (Tamba and Zhang, 2018), “fast multi-locus random-SNP-effect EMMA” (FASTmrEMMA) (Wen et al., 2017), and “integrative sure independence screening EM-Bayesian LASSO” (ISIS EM-BLASSO) (Tamba et al., 2017).

These multi-locus models have two stages. In the first stage, a single-locus model and a modified Bonferroni association

significance threshold are used, producing an intermediate result. In the second stage, for the multi-locus model implementation, it is necessary to reduce the data dimensionality. Thus, only the SNPs with the greatest potential for association selected in the first stage composed the data for the second stage. The critical values for SNPs to integrate data for the second stage were: p-value \leq 0.01, 0.01, 0.005, and 0.01 for mrMLM, FASTmrMLM, FASTmrEMMA, and ISIS EM-BLASSO, respectively. In the second stage, the association threshold for all models was the LOD score = 3. The QTNs associated above LOD = 3 corresponded to the final result.

The single-locus model was adjusted using the first five PCs based on the dissimilarity matrix and kinship matrix (K matrix). To verify the best adjusted performance for the multi-locus models, these were corrected by K matrix, K matrix + the first five PCs based on the 11,290 SNPs and K matrix + Q matrix. The Q matrix was based on the highest level of population structure (K = 2) identified from software STRUCTURE. For the single-locus MLM model, the K matrix was calculated using TASSEL 5.2.53 software, and for the multi-locus models, using the mrMLM package.

From the HapMap genotyping file, the presence or absence of the QTNs identified in the GWAS was performed in the 132 *C. arabica* accessions, with its respective resistance to BHB phenotype score. With this identification, the coefficient of correlation between the QTNs associated with resistance to BHB was calculated. The coefficient of correlation between each QTN identified in GWAS and the resistance to BHB, in the 132 *C. arabica* accessions, was also calculated. The coefficient of correlation calculation and the graphical visualization were made using the R packages Hmisc and corrplot, respectively (Harrell and Dupont, 2020; Wei and Simko, 2021).

Identification of candidate genes

To understand the molecular basis of the *C. arabica* defense to BHB, we investigated the genomic regions of associated QTNs for possible candidate genes. A search for candidate genes was performed using the *C. arabica* Et039 genome functional annotation (Salojärvi et al., 2021), according to the LD threshold, upstream and downstream to the associated QTNs positions. Functional annotation and genomic position of the *C. arabica* genes was retrieved and examined from the *C. arabica* Et039 genome functional annotation. Subsequently, proteins with biotic stress response function were literature mined from annotated information.

Plant materials for RT-qPCR

Plants of 5-months-old of *C. arabica* cultivars Catuaí Vermelho IAC 99 (susceptible), IAPAR 59 (quantitative

resistance), and IPR 102 (qualitative resistance) were submitted to infection with *P. syringae* pv. *garcae* through the method established by Rodrigues et al. (2017b). The inoculum of *P. syringae* pv. *garcae* grown on agar for 48h (28 °C), was diluted in saline (0.85% NaCl) and standardized by spectrophotometry ($A_{600} = 0.25$) containing 10^8 UFC/ml. Two pairs of the first expanded leaves of seedlings were inoculated with punctures through needles, previously dipped in the inoculum, at two points on each side of the central rib.

Leaves were collected before the inoculation at 0 hours (h) and, 10 minutes (min), 30 min, 1 h, 3 h, and 6 h, time after inoculation (tai), for RNA extraction. The experimental design for RT-qPCR was conducted with 3 plants of each cultivar for each collection time.

Two inoculated leaves, in each plant, were preserved to verify the inoculum efficiency and the disease evaluation was performed at 21 (dai). The damage in the inoculated plants was estimated using a scale from 0 to 5, where 0 = resistant, absence of symptoms; 1 = moderately resistant, initial water-soaking, up to 10% of leaf area inoculated with symptoms; 2 = susceptible, water-soaking up to approximately 25% of affected area; 3 = up to 50% of the area inoculated with symptoms; 4 = up to 75% of the affected inoculated area; 5 = more than 75% of the inoculated area necrotic from the disease (Rodrigues et al., 2017b).

Expression profile of candidate genes by RT-qPCR

The RNA extraction was carried out following the protocol of Chang et al. (1993). The cDNAs were synthesized from 2.5 µg of total RNA using the reverse transcriptase enzyme from the SuperScript[®] III First-Strand Synthesis SuperMix kit (Invitrogen), according to the manufacturer's instructions, with a final volume of 20 µL.

Specific pairs of primers were designed for the cDNA corresponding to *g000* (Chr_0_434_435) and *g010741* genes using Primer Express software (Thermo Fisher Scientific) (Supplementary Table 5). Primer efficiencies were performed through the standard curve of cDNA dilutions: 1:2, 1:4, 1:8, 1:16, 1:32, and 1:64. Every measurement of dilutions was performed in triplicate. The corresponding qPCR efficiencies (E) were calculated for primer pairs with the 7500 Fast Real-Time PCR System (Applied Biosystems), according to the equation $E = (10^{-1/\text{slope}} - 1) \times 100$ (Rutledge and Stewart, 2008).

The transcriptional profile was evaluated using qPCR (7500 Fast Real-Time PCR System, Applied Biosystems). The reaction volumes was 25 µL containing 12.5 µL of 2x SYBR[®] Green/ROX qPCR Master Mix (Applied Biosystems), 0.5 µL of each primer (10 µM), 10.5 µL of treated DEPC water, and 1 µL of cDNA diluted to a concentration of 20 ng/µL.

The reactions were prepared in technical triplicates and the thermocycling parameters used were: 10 min at 95 °C, followed

by 40 cycles of amplification of 95 °C for 30 seconds and 60 °C for 60 seconds. Dissociation curves were analyzed to verify the specificity of the primers in the amplification reaction.

The relative gene transcription was assessed using $\Delta\Delta Ct = \Delta Ct (\text{sample}) - \Delta Ct (\text{normalizer})$, where the *GAPDH* gene was used as housekeeping to normalize the data (Barsalobres-Cavallari et al., 2009). The calibrator used to compare the transcriptional activities pattern, between the times after inoculation, was the 0 h for each cultivar. The transcript data were subjected to statistical analysis of variance (ANOVA) and the Tukey test, both at the level of 5% significance, through R studio software (RStudio Team, 2020).

Haplotype analysis

From the genomic region corresponding to the *g010741* gene (CC-NBS-LRR) of *C. arabica* Et039 and the respective homologous genomic regions of *C. arabica* Caturra and, *C. arabica* var. Geisha, a haplotype analysis was performed. For this, the information available in the genome databases of *C. arabica* Et039 (Salojärvi et al., 2021), *C. arabica* Caturra (NCBI BioProject PRJNA506972), and *C. arabica* var. Geisha (Phytozome genome ID 453), was used.

The genomic region corresponding to the *C. arabica* Et039 *g010741* gene (Chr2_sg_E, position 32,109,397 bp – 32,111,166 bp) was used to identify the homologous genomic regions in Caturra and Geisha, through the BLASTn tool. Multiple alignment between the nucleotide sequences of the *C. arabica* Et039 *g010741* gene and the homologous nucleotide sequences of Caturra, and Geisha was carried out to identify SNPs, using MUSCLE (Edgar, 2004). The respective protein sequences were also aligned to identify non-synonymous SNPs (nsSNPs). SNPs inserted into conserved domains were investigated. For this, a protein sequence of 1,280 aa corresponding to the *g010741* gene was retrieved from the *C. arabica* Et039 functional annotation and used to the identification of conserved domains and their respective genomic positions, through the InterPro (Blum et al., 2021).

Near to the *g010741* gene are the QTNs Chr_2_sg_E_32049720_G and Chr_2_sg_E_32049728_G associated with resistance to BHB. Thus, in addition to the haplotype analysis described above, the identification of standard alleles or alternative alleles for the QTNs Chr_2_sg_E_32049720_G and Chr_2_sg_E_32049728_G, in the three *C. arabica* genotypes, was made. From the QTNs Chr_2_sg_E_32049720_G and Chr_2_sg_E_32049728_G positions in the Et039 genome (reference genome used to align the GBS data in this work), the identification of standard alleles or alternative alleles was made. Using the nucleotide sequence (60 bp) surrounding this QTNs in Et039, a search for homology in Caturra and Geisha genotypes, was performed. The search for homology was performed using the BLASTn tool.

Relationship between the loci involved in resistance to BHB and the *C. arabica* domestication process

The frequency of QTNs identified in the GWAS was determined in the clusters “wild”, “domesticated”, and “cultivar” identified in the Bayesian approach population structure at $K = 3$. For this, *C. arabica* Ethiopian accessions classified as admixed were disregarded. The frequency of genotypes resistant to the BHB, based in a scale with a score from 0 to 5, was also determined in the clusters

Results

SNP characteristics

After the quality filters and imputation, we selected 11,290 SNPs for population structure analysis and GWAS. From this total, 5,024 are distributed in the *C. canephora* subgenome, 5,011 in the *C. eugenioides* subgenome, and 1,255 in chromosome (Chr) zero. The Chr zero corresponds to scaffolds that did not have a defined position during the genome assembly. The SNPs/Megabase (Mb) frequency was 13.57 and 11.63 for the *C. canephora* and *C. eugenioides* subgenome, respectively. Among the chromosomes a large difference was observed in the SNP/Mb frequency. The Chr2 of the *C. canephora* subgenome showed the highest density of SNPs (17.81) and the Chr9 of the *C. eugenioides* subgenome the lowest (8.41) (Supplementary Figure 2 and Supplementary Table 2).

Population structure

The population used in this work is composed of 111 *C. arabica* Ethiopian accessions collected at the West of the Great Rift Valley, nine collected at the East of the Great Rift Valley, 11 *C. arabica* cultivars, and the BA-10 genotype.

The clustering based on the dissimilarity matrix of genetic distance of the 132 *C. arabica* accessions (Supplementary Figure 4), through the hierarchical approach, defined 3 clusters (Figure 1A and Supplementary Table 1). The cluster in blue is composed of 32 accessions, most of them collected in forests, which represents a lower domestication level, and the BA-10 genotype, which has a *C. liberica* introgression. The cluster in red is composed of 88 accessions collected, predominantly, in places with the description of naturalized, domesticated, farm open field and shadow. Naturalized, domesticated, and, mainly, farm open field represents a high domestication level. The cluster in green is composed of the *C. arabica* cultivars. Thus, the blue cluster was named “wild”, the red cluster “domesticated”, and

the green cluster “cultivar”. Among the nine accessions collected at the East of the Great Rift Valley, one is in the “wild” cluster and eight are in the “domesticated” cluster.

The PCA scatter plot also showed three clusters, with accessions grouped according to the domestication level. The “wild” cluster in blue is composed of 33 accessions, predominantly, collected in forests, and the BA-10 genotype. The “domesticated” cluster in red is composed of 85 accessions, which were predominantly characterized as domesticated. However, different from the hierarchical analysis result, the “cultivar” cluster in green is composed of cultivars and more two *C. arabica* Ethiopian accessions collected in “farm open field”. Among the accessions collected at the East side of the Great Rift Valley, one is in the “wild” cluster, seven are in the “domesticated” cluster, and one is in the “cultivar” cluster (Figure 1B and Supplementary Table 1).

According to the Bayesian approach, using STRUCTURE software and Evanno’s criterion, the upper level of the population subdivision were $K = 2$. In addition, smaller peaks were observed at $K = 3$, and at $K = 5$ (Supplementary Figure 5), which might indicate another informative population structure. Therefore, the STRUCTURE results at $K = 2$, $K = 3$, $K = 4$, and $K = 5$ were subject population genetics analyses. The population genetics analyses were based on the membership coefficient (≥ 0.6), so that the accessions could be assigned to a specific group. The membership coefficients of the *C. arabica* cultivars and *C. arabica* Ethiopian accessions at $K = 2$, $K = 3$, $K = 4$, and $K = 5$ are available in the Supplementary Material 1.

At $K = 2$, the 11 cultivars and 30 *C. arabica* Ethiopian accessions, most of them characterized as domesticated, are grouped in the same cluster (in red). The second cluster (in green) is composed of the 32 most wild *C. arabica* Ethiopian accessions and 24 domesticated accessions. For this sub-populations structure, 34 *C. arabica* Ethiopian accessions and the BA-10 genotype (26,51%) were categorized as admixed. Among the accessions collected at the East of the Great Rift Valley, seven are in the red cluster, and two were classified as admixed (Figure 1C and Supplementary Figure 6).

A clearer separation of the domesticated *C. arabica* accessions from the cultivars, and of the domesticated *C. arabica* accessions from the most wild accessions could be observed at $K = 3$. The green cluster is composed of 11 cultivars and 5 domesticated accessions. The cluster in red is composed of 19 domesticated accessions and the BA-10 genotype. The blue cluster is composed of 33 accessions, predominantly, collected in forest. Thus, the green cluster corresponds to the “cultivar” cluster, the red to the “domesticated” cluster, and the blue to the “wild” cluster. At this population structure level, 63 (47,72%) accessions were categorized as admixed. Among the accessions collected at the East of the Great Rift Valley, one is in the “domesticated” cluster, three in the “cultivar” cluster, and five were classified as admixed

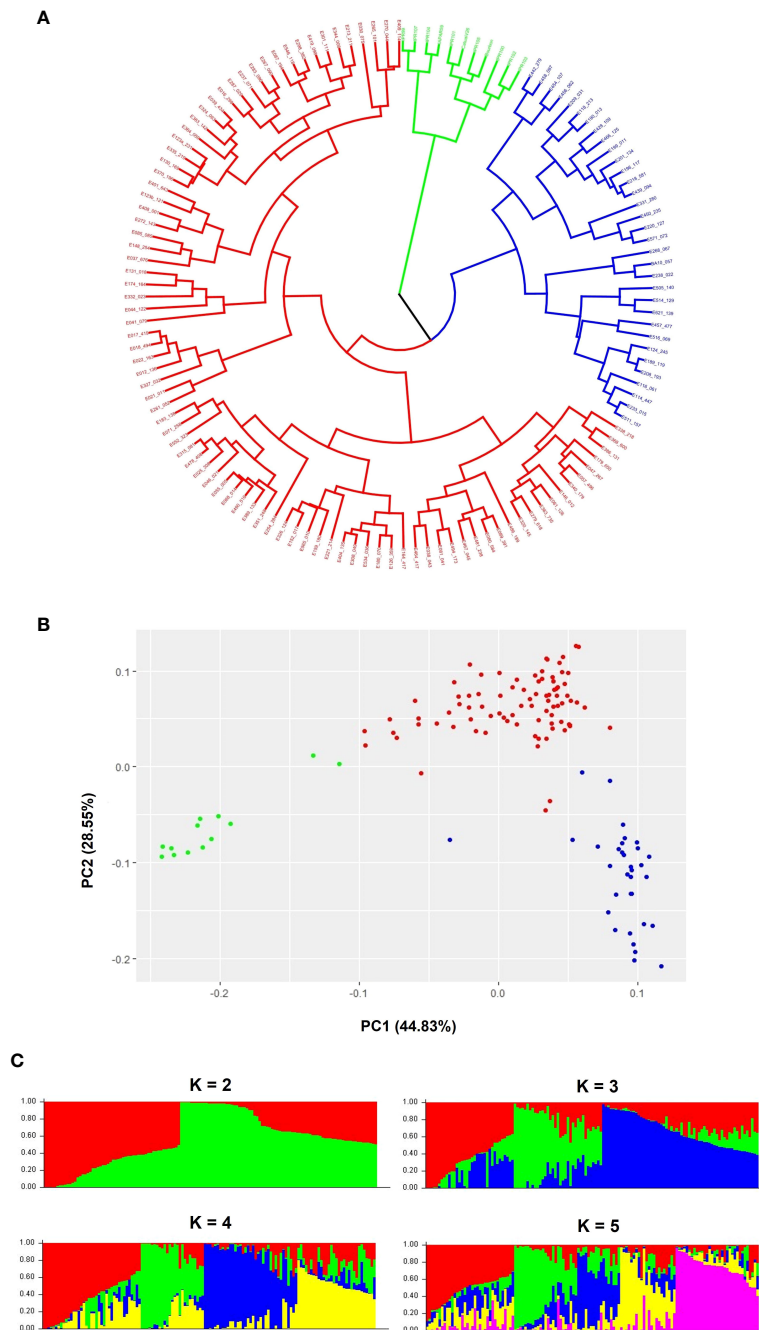


FIGURE 1

Genetic population analysis among 132 *Coffea arabica* genotypes using 11,290 SNPs. (A) Dendrogram; (B) principal coordinate analysis (PCA). Each point represents an accession, $n = 132$. In the PCA the accessions were colored according to the cluster: red “domesticated”, blue “wild”, and green “cultivar”. (C) Bar plots of the estimated membership coefficient of the 132 *C. arabica* accessions for $K = 2$, $K = 3$, $K = 4$ and, $K = 5$, using STRUCTURE.

(Figure 1C and Supplementary Figure 6). Although five accessions collected in the East were classified as admixed, it is possible to observe in three of these accessions a membership coefficient > 0.5 of the “cultivar” cluster. This result reinforces

the genetic proximity between the cultivars and the accessions collected at the East of the Great Rift Valley.

The clustering pattern obtained at $K = 4$ revealed a “wild” cluster in blue with the BA-10 genotype and 28 accessions,

predominantly, collected in forest, and a “cultivar” cluster in green with 11 cultivar and three domesticated accessions. However, at $K = 4$, were identified two “domesticated” clusters in red and in yellow, with 22 and 8 domesticated accessions, respectively. We did not identify any characteristic that could differentiate the accessions grouped in different “domesticated” clusters. For this four sub-populations, 59 (44.69%) accessions were classified as admixed. Among the accessions collected at the East of the Great Rift Valley, one is in the “cultivar” cluster, and eight were classified as admixed (Figure 1C and Supplementary Figure 6). In the same way as was observed at $K = 3$, three accessions collected at the East and classified as admixed have a membership coefficient > 0.5 of the “cultivar” cluster.

At $K = 5$, a “wild” cluster is represented in purple with 26 accessions, predominantly, collected in forest and the BA-10 genotype, and a “cultivar” cluster is represented in green with 11 cultivars and two domesticated accessions. Three “domesticated” clusters, in red, blue and yellow, with 11, 8, and 8 accessions, respectively, were identified. Here, we also did not identify characteristics that could distinguish the accessions grouped in the three different “domesticated” clusters. At $K = 5$, 63 (49.24%) accessions were classified as admixed. Among the accessions collected at the East of the Great Rift Valley, one is in the “domesticated” cluster in blue, and eight were classified as admixed (Figure 1C and Supplementary Figure 6). For this population structure level, only one accession collected in the East of the Great Rift Valley and classified as admixed has a membership coefficient > 0.5 of the “cultivar” cluster.

Although the higher delta K value was identified at $K = 2$, the population structure identified at $K = 3$ was suitable for grouping the *C. arabica* accessions according to their domestication level characteristics. This is in agreement with the hierarchical analysis and PCA results. In addition, the population structure level identified by the Bayesian approach at $K = 3$, was suitable to show the most genetic similarity between accessions collected at the East of the Great Rift Valley and the cultivars. A large difference was observed in the number of accessions that compose the “domesticated” cluster, between the PCA analysis and the Bayesian approach. This is mainly because many accessions were classified as admixed in the Bayesian approach.

GWAS

From 11,290 SNPs obtained after quality filters, 11 QTNs associated with the *C. arabica* response to BHB, through the single-locus MLM model and the four multi-locus models, mrMLM, FASTmrMLM, FASTmrEMMA, and ISIS EM-BLASSO, were identified (Table 1 and Figures 2 and 3A). The single-locus MLM model identified 4 QTNs. For the multi-locus GWAS approaches, a better fit of the data on Q-Q plot was observed in the model corrected with K matrix + PCA (Figure 3B). The multi-locus models mrMLM, FASTmrMLM, and ISIS EM-BLASSO,

adjusted with K matrix + PCA, identified 6, 4, and 5 QTNs, respectively. The FASTmrEMMA did not identify associated QTNs.

The QTNs Chr_2_sg_E_32049720_G, Chr_2_sg_E_32049728_G, Chr_0_434_435_15461_A, and Chr_0_434_435_15529_A have the highest association $-\log_{10}$ (p-value) and a high correlation with resistance to BHB (Table 1, Figure 4). While the other QTNs, identified only through the multi-locus models: Chr_1_sg_C_27081268_C, Chr_5_sg_C_29867225_G, Chr_7_sg_C_1131696_C, Chr_10_sg_C_439213_G, Chr_11_sg_C_11719181_C, Chr_5_sg_E_19112445_C, and Chr_7_sg_E_13418072_C showed a lower correlation to the level of resistance to BHB (Table 1, Figure 4).

Furthermore, only the IPR 102 cultivar and some Ethiopian accessions like E287 with high resistance to BHB have the QTNs: Chr_2_sg_E_32049720_G, Chr_2_sg_E_32049728_G, Chr_0_434_435_15461_A, and Chr_0_434_435_15529_A (Supplementary Table 3).

Identification of candidate genes

From the *C. arabica* Et039 functional annotation, a total of 31 *C. arabica* genes were identified near to 11 QTNs associated with resistance to BHB (Supplementary Table 4). To consider the gene linked to the QTN, distances approximately based on the LD decay result, were used. The LD decay at $r^2 = 0.2$ was 66,291 bp (Supplementary Figure 3). No genes were identified near to the QTNs Chr_11_sg_C_11719181_C and Chr_5_sg_E_19112445_C. Among the *C. arabica* genes identified near to the QTNs associated with resistance to BHB, we found eight genes with functional annotation related to response to biotic stress described in the literature (Table 2).

Interestingly, near to the QTN with the highest $-\log_{10}$ (p-value) of association, Chr_2_sg_E_32049720_G, we found the *g010741* gene with functional annotation for a disease resistance RPP13. This protein is a NBS-LRR, which acts on specific recognition of pathogen and hypersensitivity response (HR). The QTN Chr_0_435_15529_A, also with high $-\log_{10}$ (p-value) of association, is linked to a serine threonine- kinase D6PKL2. Another protein of this class, a serine threonine- kinase endoribonuclease IRE1a-like, was identified near to the QTN Chr_10_sg_C_439213_G.

Regarding defense compounds synthesized by plants, we identified *C. arabica* genes with functional annotation for hydrolase (*g009091*), glycosidase (*g001147*), RNA-binding protein linked to callose deposition (*g009946*), and enzyme involved in alkaloid biosynthesis (*g009954*). It is worth mentioning that the QTN Chr_1_sg_C_27081268_C is near to *g090102*, a gene with functional annotation for the Arabinogalactan protein family.

The detailed descriptions of the genes roles in the plants defense mechanism against biotic stresses, as well as their respective literary references, are presented in discussion topic.

TABLE 1 Description of QTNs associated with resistance to BHB in *C. arabica* through the MLM, mrMLM, FASTmrMLM, and ISIS EM-BLASSO models.

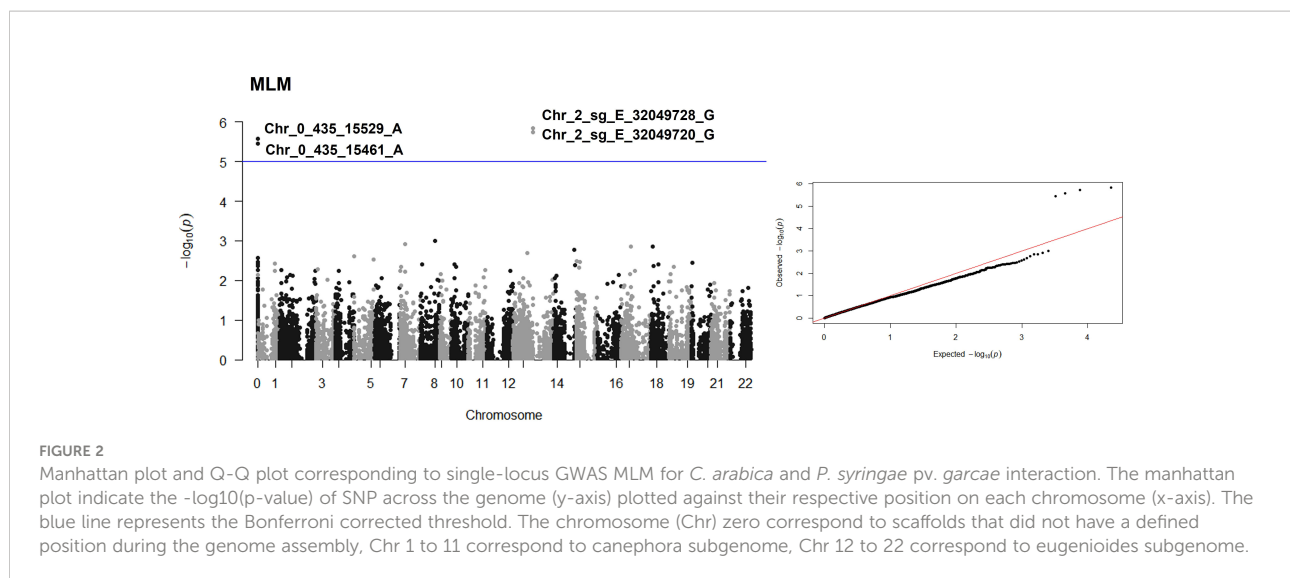
QTN ¹	Chr ²	position (pb) ³	allele standard/ alternative	MLM			mrMLM				FASTmrMLM				ISIS EM-BLASSO			
				-log10 (p- value)	QTN effect	r ² (%) ⁴	-log10 (p- value)	LOD SCORE	QTN effect	r ² (%) ⁴	-log10 (p- value)	LOD SCORE	QTN effect	r ² (%) ⁴	-log10 (p- value)	LOD SCORE	QTN effect	r ² (%) ⁴
Chr_0_435_15461_A	Sca_434HRSCAF=435	15461	G/A	5.45	1.48	20.16												
Chr_0_435_15529_A	Sca_434HRSCAF=435	15529	G/A	5.58	1.35	20.16	6.34	5.52	0.86	13								
Chr_1_sg_C_27081268_C	1 sg C	27081268	T/C				4.59	3.84	-1.29	15.94								
Chr_5_sg_C_29867225_G	5 sg C	29867225	A/G											4.04	3.33	-0.74	7.48	
Chr_7_sg_C_1131696_C	7 sg C	1131696	T/C				4.42	3.69	0.97	9.31				4.66	3.91	0.72	6.76	
Chr_10_sg_C_439213_G	10 sg C	439213	A/G											4.15	3.43	-0.63	8.82	
Chr_11_sg_C_11719181_C	11 sg C	11719181	T/C				7.4	6.55	1.17	21.02	5.26	4.48	0.86	16.3				
Chr_2_sg_E_32049720_G	2 sg E	32049720	C/G	5.73	2.52	21.64	4.07	3.35	0.59	6.27	3.77	3.07	0.51	6.67	11.3	10.35	0.9	19.4
Chr_2_sg_E_32049728_G	2 sg E	32049728	A/G	5.83	2.21	21.17												
Chr_5_sg_E_19112445_C	5 sg E	19112445	T/C				3.98	3.27	0.77	9.97	4.59	3.84	0.66	10.5				
Chr_7_sg_E_13418072_C	7 sg E	13418072	T/C								3.95	3.24	-0.57	7.64	4.71	3.96	-0.65	9.29

¹The name of the QTNs is based on its position in the genome of *C. arabica* Et039 genome and on the allele associated with resistance to BHB.

²chromosome and subgenome (sg).

³QTN positions.

⁴r²(%), phenotypic variation of traits explained by each QTN.



Transcriptional profile of candidate genes by RT-qPCR

The genes *g000* (Chr_0_434_435) and *g010741* (Chr_2_sg_E), predicted for encoding serine/threonine-kinase protein and CC-NBS-LRR, respectively, showed no difference in transcription levels in the absence of the pathogen for the Catuaí Vermelho IAC 99, IAPAR 59, and IPR 102 cultivars (Supplementary Figures 8A, B). However, after inoculation with the pathogen, different transcription levels were observed (Figures 5A, B).

The three cultivars showed repression of *serine/threonine-kinase* transcripts from 10 min (tai). However, the repression in Catuaí Vermelho IAC 99 and IAPAR 59 was higher than observed in IPR 102 during the time course of infection. The cultivar Catuaí Vermelho IAC 99 showed a progressive decrease in repression of *serine/threonine-kinase* transcripts from 30 min to 6 h (tai). The cultivar IAPAR 59 also showed a repression decrease at 30 min (tai) but remained stable until 6 h (tai). The *serine/threonine-kinase* transcript repression in IPR 102 remained stable until 6 h (tai) when it showed an upregulation.

Regarding the *CC-NBS-LRR* gene, the highly resistant cultivar IPR 102 showed no change in the level of transcription at 10 min (tai) compared to that observed in the absence of the pathogen, but at 30 min (tai) an upregulation peak approximately 30 times greater than that found in the cultivar in the absence of the pathogen was observed. However, at 1 h and 3 h (tai) the IPR 102 returned to show the same level of transcription that was found in the absence of the pathogen for the *CC-NBS-LRR* gene. The IPR 102 showed another peak of upregulation at 6 h (tai), approximately 50 times higher than the level found in the absence of the pathogen and with statistical difference from the first peak.

The cultivar Catuaí Vermelho IAC 99 showed a pattern of progressive downregulation between 10 min (tai) and 6 h (tai) for *CC-NBS-LRR* transcripts. The cultivar IAPAR 59 also showed a pattern of downregulation of this gene from 10 min (tai), but similar to the serine/threonine-kinase the level of repression remained stable until 6 h (tai).

For the RT-qPCR analyses, the choice of collection times is based on the results of Cheval et al. (2013) and Nobori et al. (2020). According to these authors, the interaction of *Arabidopsis thaliana* vs. *P. syringae* in the early stages, before 6 h (tai), are essential to determine the plant resistance or susceptibility.

All seedlings of Catuaí Vermelho IAC 99, IAPAR 59, and IPR 102 used for qPCR analyses, showed leaves with the score 4 (susceptible), 2 (moderately resistant), and 0 (highly resistant) at 21 (dai), respectively (Figure 5C).

Haplotype analyses

The *C. arabica* Et039 *g010741* gene showed homology with the genomic region 31,924,108 bp – 31,927,645 bp (Chr2_sg_E) of Caturra, and with the genomic region 1,417,586 bp – 1,421,123 bp (Scaffold_627 | HRSCAF=12830) of Geisha genotype. The homologous genomic region identified in Caturra, corresponds to the locus LOC113732450, and in Geisha to the locus Os08g43010. Both LOC113732450 and Os08g43010 have functional annotation for disease resistance RPP13-like protein 1. From the multiple alignment between *g010741* (Et039), LOC113732450 (Caturra) and Os08g43010 (Geisha), it was possible to identify 143 SNPs in a sequence region of 3,412 bp. Caturra and Geisha showed the same alleles for these 143 SNPs. The position of each SNP identified in the

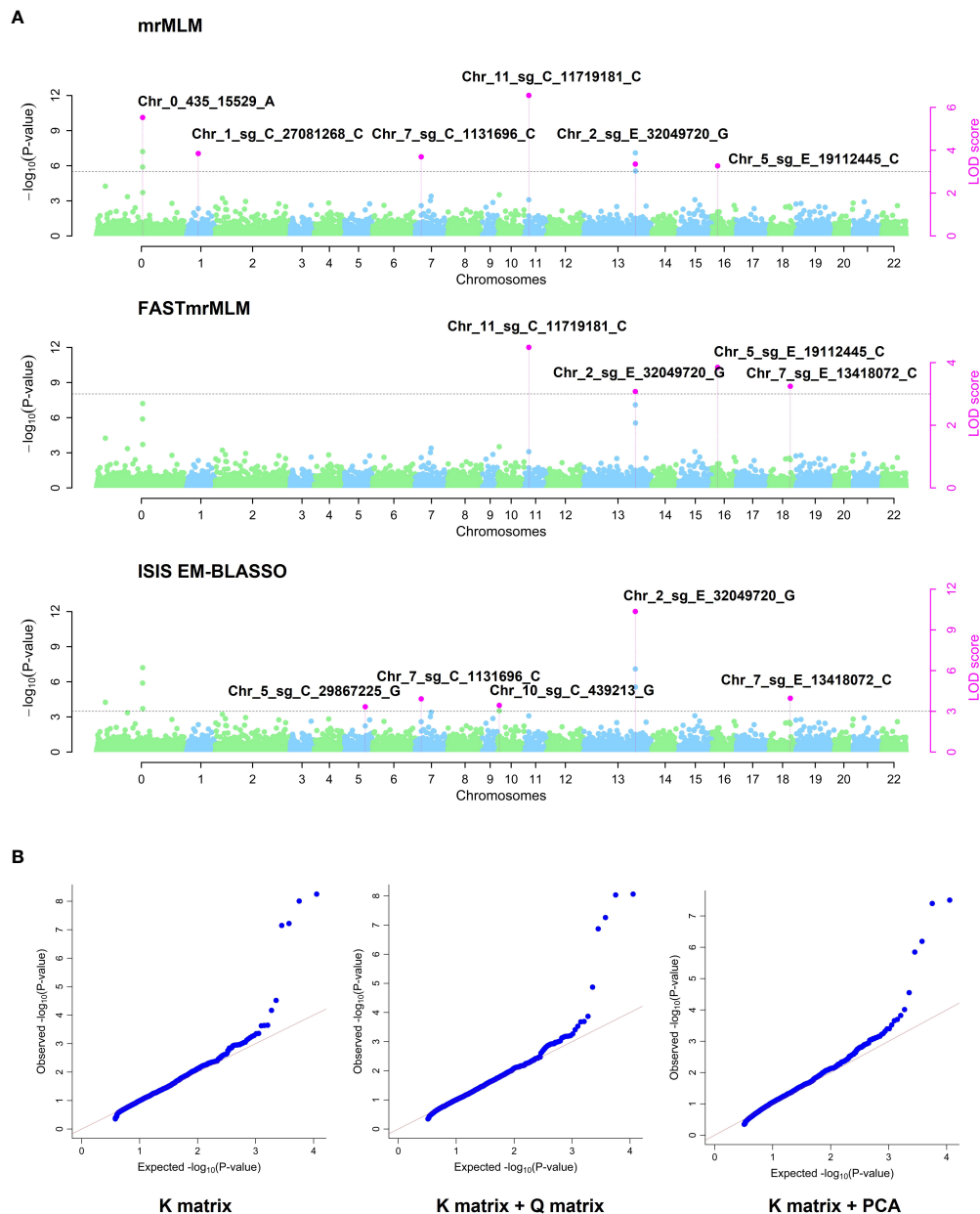


FIGURE 3

Multi-locus GWAS. (A) Manhattan plots corresponding to mrMLM, FASTmrMLM, and ISIS EM-BLASSO for *C. arabica* and *P. syringae* pv. *garcae* interaction. The manhattan plots indicate the $-\log_{10}(p\text{-value})$ and the LOD score of SNP across the genome (y-axis) plotted against their respective position on each chromosome (x-axis). The chromosome (Chr) zero correspond to scaffolds that did not have a defined position during the genome assembly. Chr 1 to 11 correspond canephora subgenome, Chr 12 to 22 correspond to eugenioides subgenome. (B) Q-Q plots from multi-locus models adjusted by different population structure approaches. Kinship matrix (K matrix), principal component analysis (PCA), and the population structure $K = 2$ identified from software STRUCTURE (Q matrix).

haplotype analysis is available in [Supplementary Material 2](#). Thus, the alleles identified in Et039 compose the haplotype 1 (Hap1) and the alleles identified in Caturra and Geisha compose the haplotype 2 (Hap 2) (Figure 6B). Among the 143 SNPs identified, 105 correspond to nsSNPs (Figure 6B). The amino

acid change corresponding to each nsSNPs is available in [Supplementary Material 3](#).

From the protein sequence predicted by the gene *g010741*, it was possible to identify the conserved domains N-terminal coiled-coil (ID IPR041118), NB-ARC (ID IPR002182), and

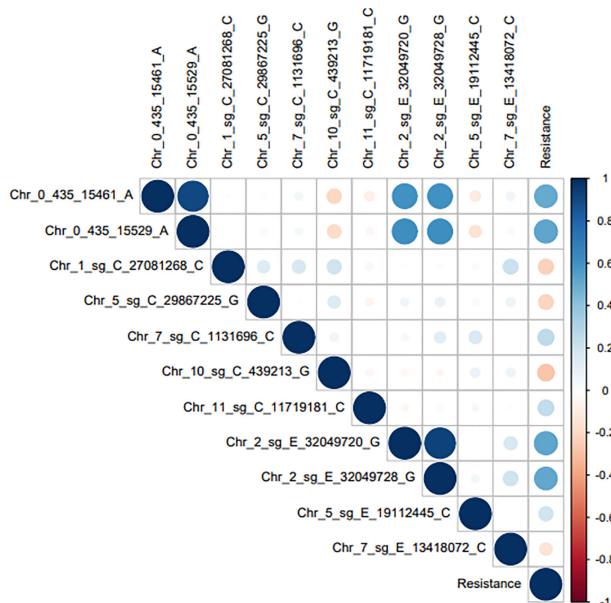


FIGURE 4 Correlation between the 11 QTNs identified in the GWAS and the resistance to BHB phenotype. The closer to zero, the lower the correlation, positive (blue) or negative (red), which depends on the associated QTN effect.

LRR (ID IPR032675). The positions, in bp, of the conserved domains identified for this gene in the *C. arabica* Et039 chromosome 2 eugenioides subgenome, was represented in Figure 6A. Among the 143 SNPs that compose the haplotypes, 13 are inserted in the CC domain, 37 in the NB-ARC domain and 71 in the LRR domain (Figure 6B). Among the 105 nsSNPs, 7 are inserted in the CC domain, 24 in the NB-ARC domain and 60 in the LRR domain (Figure 6B).

C. arabica Et039, which represents Hap1, showed the alternative allele G both for the QTN Chr_2_sg_E_32049720_G

and for the QTN Chr_2_sg_E_32049728_G. Caturra and Gueisha, which represents Hap2, showed the standard allele C for the QTN Chr_2_sg_E_32049720_G and the standard allele A for the QTN Chr_2_sg_E_32049728_G (Figure 6C). In the Caturra genome the QTNs Chr_2_sg_E_32049720_G and Chr_2_sg_E_32049728_G were located in Chr2_sg_E at the positions 31,851,359 bp and 31,851,367 bp, respectively. In the Geisha genome the QTNs Chr_2_sg_E_32049720_G and Chr_2_sg_E_32049728_G were located in Scaffold_627|HRSCAF=12830 at the positions 1,493,868 bp and 1,493,859 bp, respectively.

TABLE 2 Functional annotation of genes near to QTNs associated with resistance to BHB in *C. arabica* identified in this study.

gene ¹	QTN	gene position in relation to the QTN ²	functional annotation ³
<i>g000</i>	Chr_0_435_15461_A; Chr_0_435_15529_A	(-)12.090; (-)12.158	Serine/threonine-protein kinase D6PKL2
<i>g009091</i>	Chr_1_sg_C_27081268_C	(-)30.124	probable carbohydrate esterase At4g34215
<i>g090102</i>	Chr_1_sg_C_27081268_C	(+)64.608	Arabinogalactan 1
<i>g009893</i>	Chr_5_sg_C_29867225_G	(-)66.800	UBP1-associated protein 2C
<i>g009954</i>	Chr_5_sg_C_29867225_G	(-)34.821	S-norcochlorine synthase-like
<i>g000156</i>	Chr_10_sg_C_439213_G	(-)52.180	serine threonine- kinase endoribonuclease IRE1a-like
<i>g001147</i>	Chr_10_sg_C_439213_G	(-)49.324	glucan endo-1 3-beta-glucosidase 12-like
<i>g010741</i>	Chr_2_sg_E_32049720_G; Chr_2_sg_E_32049728_G	(+)76.156; (+)76.148	Disease resistance protein RPP13 1

¹The name of the genes was retrieved from the *C. arabica* Et039 genome functional annotation.

²(+) up or (-) downstream position in bp.

³The genes functional annotation was retrieved from the *C. arabica* Et039 genome functional annotation.

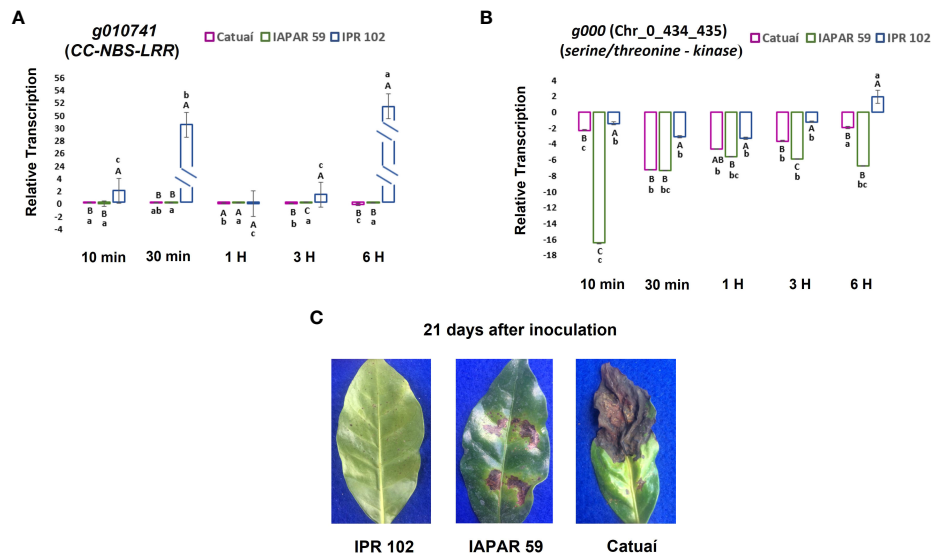


FIGURE 5
(A) Transcription levels of genes related to BHB resistance in leaves of *C. arabica* Catuaí Vermelho IAC 99, IAPAR 59, and IPR 102 inoculated with *P. syringae* pv. *garcae*. **(A)** Serine/threonine-kinase *g000* (Chr_0_434_435) and **(B)** CC-NBS-LRR *g010741* (Chr_2_sg_E). Data represent means \pm standard deviation. Upper case letters compare columns of different genotypes at the same tai and lower-case letters compare columns of the same genotype at different tai. Same letters show no difference by the Tukey test (for $P \leq 0.05$). **(C)** Phenotypic analysis of IPR 102, IAPAR 59, and Catuaí Vermelho IAC 99, 21 days after inoculation.

Discussion

GWAS

In this study, we identify 11 QTNs associated with resistance to BHB in *C. arabica*. Using a multi-locus method, in relation to the single-locus, we identified a greater number of QTNs associated with response to BHB, with a small effect. The

advantages of the multi-locus method over single-locus has already been described in other plants and animal species, such as cotton, maize, and pigs (Li et al., 2018; Su et al., 2018; Xu et al., 2018; Ding et al., 2019). This is possible due to the nature of the multi-locus method, which makes a genome multidimensional scan and a simultaneous estimate of the effect of all SNPs (Cui et al., 2018). The ability to identify a greater number of QTNs and QTNs with small effect associated

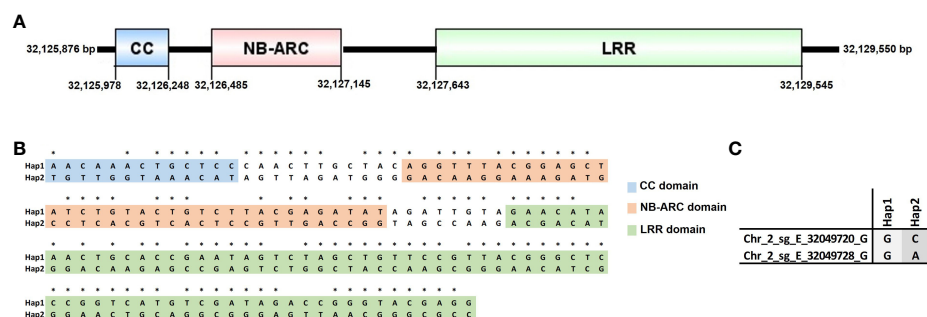


FIGURE 6
Haplotype analysis. **(A)** The *C. arabica* Et039 *g010741* gene (Chr2_sg_E) and their conserved domains with respective positions in pb. **(B)** The 143 SNPs identified between haplotype 1 (Hap1 – Et039) and haplotype 2 (Hap2 – Caturra and Geisha); and the alleles that compose the haplotypes. The SNPs identified between Hap1 and Hap2, inserted in conserved domains, are highlighted. The position of each SNP identified in the haplotype analysis, in the *C. arabica* Et039 chromosome 2 eugenioides subgenome, is available in [Supplementary Material 2](#). The 105 nsSNPs are identified with *. The amino acid change corresponding to each nsSNPs is available in [Supplementary Material 3](#). **(C)** The alleles identified in the QTNs Chr_2_sg_E_32049720_G and Chr_2_sg_E_32049728_G in *C. arabica* Et039 (Hap1) and in Caturra e Geisha (Hap2).

with the trait of interest increase our understanding of the genetic control of complex traits and the estimation of heritability (Ding et al., 2019; Zhang et al., 2019).

Some multi-locus methods were able to overlap the genomic regions associated with resistance to BHB identified through the single-locus method. The QTN Chr_0_435_15529_A was detected using the single-locus MLM and through the multi-locus mrMLM. A complementarity between the results of the single-locus and multi-locus methods was also found for the QTN Chr_2_sg_E_32049720_G, which was identified through the single-locus MLM and the multi-locus mrMLM, FASTmrMLM, and, ISIS EM-BLASSO. Other authors have observed the same ability of multi-locus methods to overlap part of the results obtained by single-locus methods (Misra et al., 2017; Li et al., 2018).

It is worth mentioning that the multi-locus method considers the LD and decreases the number of associated QTNs in the same haplotype (Ding et al., 2019). Probably because of this, the QTN Chr_2_sg_E_32049728_G identified by MLM was not identified in the mrMLM, FASTmrMLM, and ISIS EM-BLASSO, since it is very close to the QTN Chr_2_sg_E_32049720_G. The same was observed for the QTN Chr_0_435_15461_A, in relation to the QTN Chr_0_435_15529_A, in the result of the mrMLM.

Some QTNs were identified through different multi-locus methods, such as the QTN Chr_2_sg_E_32049720_G which was identified by the mrMLM, FASTmrMLM, and ISIS EM-BLASSO, the QTNs Chr_11_sg_C_11719181_C and Chr_5_sg_E_19112445_C identified by the mrMLM, and FASTmrMLM, the QTN Chr_7_sg_C_1131696_C identified by the mrMLM, and ISIS EM-BLASSO, and the QTN Chr_7_sg_E_13418072_C identified by the FASTmrMLM, and ISIS EM-BLASSO, thus, demonstrating the robustness of the association methods as well the reliability of the QTNs.

Identification of candidate genes related to plant defense to BHB

Plant diseases threaten plant production and, consequently, world food security. Identification and understanding of the mechanisms and genetic bases of resistance is essential to achieve greater and more sustainable production. Genes involved in resistance, once identified and validated, can be used to improve the crop based on molecular assisted breeding or using genome editing tools.

Plant resistance is based on metabolites and proteins that directly suppress pathogen development, leading to reduced susceptibility. These metabolites and proteins can be constitutively present or synthesized in plants following pathogen perception (Kushalappa et al., 2016). After the plant proteins recognize the pathogen a hierarchy of downstream genes such as phytohormones (salicylic acid, jasmonic acid,

and ethylene), mitogen-activated protein kinases (MAPKs), and transcription factors (TFs), which regulate downstream genes that produce defense metabolites (phytoanticipins, phytoalexins, and complex conjugates that are deposited to reinforce cell walls), and proteins (pathogenesis-related proteins, detoxify toxins proteins, and enforcing cell wall proteins) (Kushalappa and Gunnaiah, 2013).

Some *C. arabica* genes, near to the QTNs associated with resistance to BHB, identified in this work, have functional annotation for proteins that play roles in the plants defense mechanism against pathogens.

The gene *g010741*(Chr_2_sg_E) is ortholog to the RPP13, a downy mildew resistance protein (NBS-LRR) originally identified in the Niederzenz (Nd-1) ecotype of *Arabidopsis* (Bittner-Eddy et al., 2000). Furthermore, Lewis et al. (2013) identified that RPP13L4/ZAR1 is dependent on a nonfunctional kinase, which functions as a “decoy”, for pathogen recognition.

Other genes that act as defense signal receptor and transmitter proteins were also identified. The gene *g000* (Chr0) has a functional annotation for serine/threonine-protein kinase D6PKL2. In Zhang et al. (2021) this protein was described as specifically expressed after *Fusarium oxysporum* infection in *Vernicia montana* resistant trees. Furthermore, transgenic analysis in *Arabidopsis* and tomato revealed that D6PKL2 significantly enhanced resistance in both species, whereas the *d6pkl2* mutant displayed reduced resistance against *F. oxysporum* (Zhang et al., 2021). The gene *g000156* (Chr_10_sg_C) has orthology with an *Arabidopsis* gene that encodes a serine/threonine-kinase endoribonuclease IRE1a-like protein. In *Arabidopsis*, this gene was transcriptionally induced during bacterial pathogen *P. syringae* pv. *maculicola* infection and plays a predominant role in the secretion of pathogenesis-related proteins during salicylic acid treatment. *Arabidopsis* mutant plants without this gene showed enhanced susceptibility to a bacterial pathogen and were deficient in establishing systemic acquired resistance (SAR) (Moreno et al., 2012).

A protein with hydrolytic function, the probable carbohydrate esterase At4g34215, was identified in the *g009091* (Chr_1_sg_C) gene functional annotation. Balmant et al. (2015) reported that a carbohydrate esterase exhibited a significant fold change in redox modification and a significant fold change in transcriptional level, during the interaction between tomato resistant genotype (PtoR) and *P. syringae* pv *tomato*. The gene *g001147* (Chr10_C) also showed functional annotation for a protein with hydrolytic function, a glucan endo-1|3-beta-glucosidase. This enzyme are grouped in the PR-2 family of pathogenesis-related proteins and catalyze β -1,3-glucans cell wall structural component of fungal and bacteria (Bachman and McClay, 1996; Balasubramanian et al., 2012).

Regarding secondary metabolites of plant defense, the functional annotation of the *g009954* (Chr_5_sg_C) gene corresponds to an S-norcoclaurine synthase-like. S-

norcochlorine is the ultimate precursor to benzyloisoquinoline, an alkaloids that plays a role in the plants defense against herbivores and pathogens (Liscombe et al., 2005; Yogendra et al., 2017).

The *g009893* (Chr_5_sg_C) has functional annotation for a UBP1-associated protein 2C. This protein is a RNA-binding and provides hints to another level of regulation posttranscriptionally. In Bove et al. (2008) is reported that transcripts levels of UBA2 increased following mechanical wounding. In another work, using transient expression, Kim et al. (2008) identified that *Arabidopsis* UBA2 proteins in *Nicotiana benthamiana* leaves induces a programmed-cell-death/senescing phenotype, while constitutive expression causes an early lethality phenotype in *Arabidopsis*.

The gene *g090102* (Chr1_C) has functional annotation for Arabinogalactans proteins family. Arabinogalactans proteins (AGPs) are structural polysaccharides associated with proteins of the plant cell wall. Plants can secrete and accumulate AGPs at infection sites, creating cross-links with pectin. Pathogens can secrete hydrolytic enzymes that damage AGPs. In this way, the AGPs degradation, by the pathogen's hydrolytic enzymes action, can elicit plant defense response through damage-associated molecular patterns (DAMPs) (Villa-Rivera et al., 2021).

Identification of genomic regions probably involved in qualitative resistance

Major genes that recognize the pathogen and trigger PTI and ETI, resulting in HR, are essential to promote qualitative resistance (Jones and Dangl, 2006; Yuan et al., 2021). In the absence of these genes, other genes with minor effects can promote partial resistance to the pathogen and a decrease in symptom severity (St. Clair, 2010). This partial resistance to the pathogen corresponds to quantitative resistance and produces a continuous distribution of phenotypes between the susceptible and the resistant plant (Corwin and Kliebenstein, 2017).

The elucidation of which genes act on qualitative, quantitative, or both types of resistance is important, since the combination of qualitative and quantitative resistance, in the same genotype, is considered a promising strategy in breeding programs (Pilet-Nayel et al., 2017). This is because quantitative resistance can exercise different selection pressures on the pathogen (Quenouille et al., 2014). In addition, it can reduce the effective population size of the pathogen, which increases the genetic drift and delays the emergence of new virulent forms (Brun et al., 2010). Thus, the effectiveness of resistance conferred by major genes is preserved.

The QTNs Chr_2_sg_E_32049720_G, Chr_2_sg_E_32049728_G, Chr_0_434_435_15461_A, and Chr_0_434_435_15529_A can be associated with qualitative resistance to BHB. Four of our results reinforces this hypothesis. Firstly, the QTNs showed a strong correlation with the resistance phenotype (Figure 4). Secondly, only

C. arabica plants with these 4 QTNs showed qualitative resistance to BHB (Supplementary Table 3). Thirdly, the genes close to these QTNs are predicted to encode a serine/threonine-kinase protein and a CC-NBS-LRR protein, which act *via* an indirect interaction with the pathogen effector (Table 2). Finally, fourth, those genes showed a very similar transcriptional profile for cultivars Catuaí Vermelho IAC 99 (susceptible) and IAPAR 59 (quantitative resistance), while the cultivar IPR 102 (qualitative resistance) showed a different profile (Figures 5A, B), with these genes being upregulated in relation to the susceptible cultivars.

In *Arabidopsis* and tomato interaction with *P. syringae* the effector AvrPto binds to pattern recognition receptors (PRRs), including receptor-like protein kinases (RLKs) protein, and blocks pattern triggered immunity (PTI) activation. This block enhances host susceptibility. In resistant plants, a serine/threonine-kinase can mimic the kinase domain of the RLKs and bind to AvrPto. An NB-LRR protein recognizes the bind between the effector and the serine/threonine-kinase and then activates the effector triggered immunity (ETI). Thus, PTI and ETI are activated, resulting in a hypersensitivity response (HR) (Supplementary Figure 9) (Jones and Dangl, 2006; Zipfel and Rathjen, 2008; Zong et al., 2008).

Thus, the genes *g000* (Chr_0_434_435) and *g010741* (Chr_2_sg_E) could probably play a similar role in resistance to BHB in *C. arabica*. These results are an indication that the transcription levels for genes *g000* (Chr_0_434_435) *serine/threonine-kinase* and *g010741*(Chr_2_sg_E) *CC-NBS-LRR*, found in IPR 102, may be necessary to activate qualitative resistance to BHB.

The haplotype analysis performed in this work identified a high number of polymorphisms in the *CC-NBS-LRR* gene, among *C. arabica* genotypes. Interestingly, Et039 (Hap1), showed the alternative alleles associated with resistance to BHB for the QTNs Chr_2_sg_E_32049720_G and Chr_2_sg_E_32049728_G. Meanwhile, Caturra and Geisha (Hap2) showed the standard alleles for the same QTNs (Figure 6C). The nsSNP identified between Hap1 and Hap2, mainly those inserted in conserved domains (Figure 6B), may be related to the defense response functionality of the gene. However, the haplotypes must be validated in a greater number of genotypes with phenotyping for resistance to BHB in order to reinforce the role of this genes in the resistance to BHB.

Relationship between the loci involved in resistance to BHB and the *C. arabica* domestication process

The population structure results showed by hierarchical analysis, PCA, and Bayesian approach at $K = 3$ were consistent for clustering the *C. arabica* Ethiopian accessions

and the cultivars according to their domestication level characteristics. The admixed accessions identification was possible with the Bayesian approach. Furthermore, the Bayesian approach at $K = 3$ was the most suitable in demonstrating the genetic proximity between the accessions collected at the East of the Great Rift Valley and the *C. arabica* cultivars. This result corroborate with other studies that identified the movement of *C. arabica* from West to East of the Great Rift Valley, in direction to Yemen, from where *C. arabica* was disperse worldwide (Montagnon and Bouharmont, 1996; Silvestrini et al., 2007; Scalabrin et al., 2020).

From the frequency analysis of resistance to BHB into the clusters formed in population structure Bayesian approach at $K = 3$, it was possible to identify a higher frequency of resistance to BHB in the “domesticated” cluster (Figure 7A). However, in the *C. arabica* cultivars, there was a decrease in the frequency of accessions resistant to BHB. The QTNs Chr_2_sg_E_32049720_G and Chr_2_sg_E_32049728_G, with the highest $-\log_{10}$ (p-values) of association identified by both the single-locus and multi-locus methods, showed a higher frequency in the “domesticated” cluster (Figure 7B). The distribution frequency of the others QTNs associated with resistance to BHB identified in this work are shown in Supplementary Figure 7.

The highest frequency, in the “domesticated” cluster, of *C. arabica* accessions resistant to BHB and of the QTNs Chr_2_sg_E_32049720_G and Chr_2_sg_E_32049728_G, suggest that resistance to BHB is a trait inserted in loci that may have had an influence in plants selection during the *C. arabica* domestication history (Figures 7A, B). The influence of genomic regions linked to resistance to disease in the domestication process has already been described for other species, such as peanut (*Arachis hypogaea*), tomato (*Solanum pimpinellifolium* L. and *S. lycopersicum* L.), and soybean (*Glycine max* L.) (Zhao et al., 2015; Zhang et al., 2017; Razifard et al., 2020). In addition, genomic regions of resistance to disease showed the highest selection rate, during the divergence between *Coffea arabica*, *C. canephora*, and *C. excelsa* (Huang et al., 2020).

However, the lower frequency in the “cultivar” cluster of both *C. arabica* resistant to BHB and QTNs with the highest $-\log_{10}$ (p-values) identified in this GWAS agrees with the low genetic variability and low resistance to pathogens known in *C. arabica* cultivars (van der Vossen, 1985; Anthony et al., 2002). The lack of resistance to diseases in cultivars, a disadvantage in relation to the diversity found in wild genotypes, is extensively reported in the literature in many different species (Lam et al.,

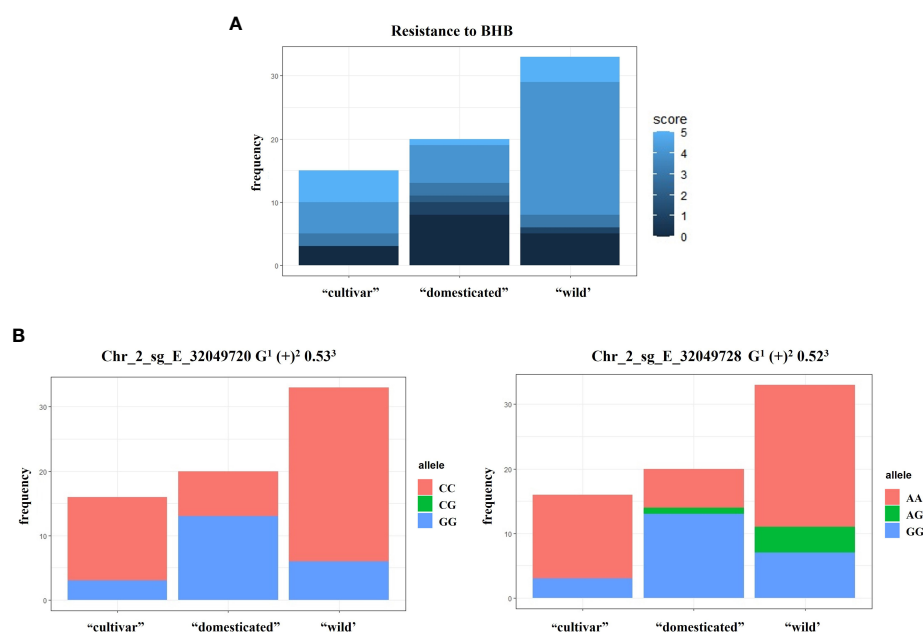


FIGURE 7

Population structure and relation with resistance to BHB. (A) The distribution frequency of accessions resistance to BHB into the “cultivars” ($n = 16$), “domesticated” ($n = 20$), and “wild” ($n = 33$) clusters identified in population structure Bayesian approach at $K = 3$. *C. arabica* Ethiopian accessions classified as admixed were disregarded in this analysis. The resistance to BHB is based in a scale with a score from 0 to 5, where 0 corresponds to highly resistant plants, 1 to resistant, 2 to moderately resistant, 3 to moderately susceptible, 4 to susceptible, and 5 to highly susceptible. (B) The distribution frequency of the QTNs associated with resistance to BHB into the “cultivars”, “domesticated”, and “wild” clusters. QTN associated with resistance to BHB with its respective location in the *C. arabica* Et039 genome¹; QTN effect²; r^2 (%) of the QTN in relation to the resistance phenotype³.

2011; Stalker et al., 2013; Cao et al., 2014). According to these studies, characteristics such as greater palatability, greater production, and fruit size, had greater relevance in the cultivar development over resistance to diseases. In the case of *C. arabica*, most cultivars are derived from only two varieties: Typica and Bourbon (Krug et al., 1939). This fact makes the genetic diversity among commercial *C. arabica* cultivars quite limited (Anthony et al., 2002; Steiger et al., 2002; Vidal et al., 2010). In addition, breeding programs that originated these cultivars prioritized criteria such as high productivity, beverage quality, and drought tolerance (van der Vossen, 1985). As a result, a large genetic bottleneck was formed. It is important to observe that, in this work, the haplotype analysis of a *CC-NBS-LRR* gene showed a greater genetic diversity from an Ethiopian accession (*C. arabica* Et039) than in commercial genotypes (Caturra and Geisha).

These results reinforce the importance of conservation of *C. arabica* wild types, both in their native habitats and in germplasm banks. Since their genetic diversity can be used to identify genes involved in traits of interest, such as resistance to diseases, and to elucidate the *C. arabica* domestication process.

Conclusions

This study provides 11 QTNs associated with resistance to BHB that have potential for application in Marker Assisted Selection and the characterization of a serine/threonine-kinase and a *CC-NBS-LRR*, probably, involved in qualitative resistance to BHB, could help to direct the breeding and could be a gene editing target.

Our results demonstrate that the combination of different GWAS methods can increase the efficiency and accuracy of obtaining a panel of QTNs associated with features of interest.

Furthermore, this work provides an insight into the relationship between the loci involved in resistance to BHB and the *C. arabica* domestication history. This reinforces the importance of maintenance of the Ethiopian coffee collections as a genetic diversity source for breeding programs on the pathway to achieve sustainable agriculture.

This is the first report for the interaction between *C. arabica* and *P. syringae* at the genetics and molecular level. In this way, the results presented are a starting point for future works that may bring greater evidence and validation of the resistance genes identified, such as gene expression analysis and validation in gene knock out mutants or gene overexpressing mutants.

Data availability statement

The GBS data presented in the study are deposited in the National Center for Biotechnology Information repository, accession number PRJNA882422.

Author contributions

CA, LP wrote the manuscript. CA, GCS, LFPP designed the experiment. MSF, RVF, LMN, BSS performed genotyping data. CA, GHS, LMRR, SALD participated in preparation of the materials and phenotype analysis. DSD, LFPP, MLVR provide funds for research. All the authors read, contributed and approved the final version of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by grants from the Consórcio Pesquisa Café (Grant 10.18.20.027.00), INCT Café and FAPEMIG. CA, RVF, LMN, BSS and MSF received a PhD fellowship from CAPES (Finance Code 001). LP and DD received a Research Fellowship from CNPq (311452/2020-5).

Acknowledgments

We gratefully acknowledge the Arabica Coffee Genome Consortium (ACGC) for providing access to *C. arabica* Et039 genome data. We also acknowledge the Consórcio Pesquisa Café, INCT Café, Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for financial support and fellowships.

Conflict of interest

GCS was employed by company Tropical Melhoramento & Genética (TMG).

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.989847/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Histogram of the disease distribution, values of response to Bacterial Halo Blight obtained in field evaluation (Mohan et al., 1978; Ito et al., 2008). The X-axis represents the classes of distribution for the 120 *C. arabica* wild accessions (blue), 11 *C. arabica* cultivars (red) and BA-10 genotype evaluated. The Y-axis shows the count of *C. arabica* genotypes in each category.

SUPPLEMENTARY FIGURE 2

Distribution of 11,290 SNPs, identified in 120 *C. arabica* wild accessions, 11 *C. arabica* cultivars and BA-10 genotype, in the *C. arabica* genome. Each chromosome has its size information in base pairs. The bars correspond to the number of SNPs on each chromosome.

SUPPLEMENTARY FIGURE 3

Linkage disequilibrium (LD) decay for 16,146 SNPs obtained after quality filters and the removal of SNPs from Chr zero. Each point represents a physical distance (bp) between SNPs pairs and their squared correlation coefficient (r^2). The red line represents the line for non-linear regression against physical distance. The green line represents the LD decay value for $r^2 = 0.2$.

SUPPLEMENTARY FIGURE 4

Dissimilarity matrix representing the genetic distance between the 120 *C. arabica* wild accessions, 11 *C. arabica* cultivars and BA-10 genotype. The color intensity is proportional to the value of genetic distance between the accessions. The matrix was constructed based on the 11,290 SNPs identified in this population. The color intensity is proportional to the value of genetic distance between the accessions.

SUPPLEMENTARY FIGURE 5

Structure analysis of 120 *C. arabica* wild accessions, 11 *C. arabica* cultivars and BA-10 genotype. Evolution of ΔK values (y-axis) according to the number of genetic groups (x-axis).

SUPPLEMENTARY FIGURE 6

Classification of 132 accessions of *C. arabica*, using 11,290 SNPs identified in this population, into $K = 2$, $K = 3$, $K = 4$, and $K = 5$ using STRUCTURE 2.3.4. The distribution of the accessions to different populations is indicated by the color code. Numbers on the y-axis show the subgroup membership, and the x-axis shows the different accessions.

SUPPLEMENTARY FIGURE 7

Population structure and relation with resistance to BHB. The distribution frequency of the QTNs associated with resistance to BHB into the "cultivars" ($n = 16$), "domesticated" ($n = 20$), and "wild" ($n = 33$) clusters identified in population structure Bayesian approach at $K = 3$. *C. arabica* Ethiopian accessions classified as admixed were disregarded in this analysis. QTN associated with resistance to BHB with its respective location in the *C. arabica* Et039 genome1; QTN effect2; r^2 (%) of the QTN in relation to the resistance phenotype3.

SUPPLEMENTARY FIGURE 8

Transcribed levels in the absence of the pathogen. (A) The g000 (Chr_0_434_435) and (B) g010741 (Chr_2_sg_E). Data represent the means \pm standard deviation. Upper case letters compare columns of different genotypes. Same letters show no difference by Tukey test (for $P \leq 0.05$).

SUPPLEMENTARY FIGURE 9

Model for PTI-inhibition and ETI-induction by *Pseudomonas syringae* effector. In susceptible plants, the effectors bind to PRRs. This blocks PTI activation and enhances host susceptibility to the bacterium. In resistant plants, serine/threonine – kinase is evolved to mimic the kinase domain of PRRs and binds to the effectors to trigger ETI, resulting in a resistance outcome. Blocking (A) and triggering (B) a hypersensitive response by *Pseudomonas syringae* effector.

References

- Aduagna, G., Jefuka, C., Teferi, D., and Abate, S. (2012). New record and outbreaks of bacterial blight of coffee (*Pseudomonas syringae*) in southern Ethiopia: impact of climate change scenarios. In 24th international conference on coffee science (ASIC) São José, Costa Rica, 85.
- Amaral, J. F., Teixeira, C., and Pinheiro, E. D. (1956). A bactéria causadora da mancha-aureolada do cafeeiro. *Arquivo do Instituto. Biológico SP Brazil* 23, 151–155.
- Andersen, E. J., Ali, S., Byamukama, E., Yen, Y., and Nepal, M. P. (2018). Disease resistance mechanisms in plants. *Genes* 9, 339–368. doi: 10.3390/genes9070339
- Anthony, F., Bertrand, B., Quiros, O., Wilches, A., Lashermes, P., Berthaud, J., et al. (2001). Genetic diversity of wild coffee (*Coffea arabica* L.) using molecular markers. *Euphytica* 118, 53–65. doi: 10.1023/A:1004013815166
- Anthony, F., Combes, M. C., Astorga, C., Bertrand, B., Graziosi, G., and Lashermes, P. (2002). The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. *Theor. Appl. Genet.* 104, 894–900. doi: 10.1007/s00122-001-0798-8
- Anzueto, F., Bertrand, B., Sarah, J. L., Eskes, A. B., Decazy, B., Rica, C., et al. (2001). Resistance to *Meloidogyne incognita* in Ethiopian *Coffea arabica* accessions. *Euphytica* 118, 1–8. doi: 10.1023/A:1003712232325
- Bachman, E. S., and McClay, D. R. (1996). Molecular cloning of the first metazoan beta-1,3 glucanase from eggs of the sea urchin *Strongylocentrotus purpuratus*. *Proc. Natl. Acad. Sci. U S A.* 93 (13), 6808–6813. doi: 10.1073/pnas.93.13.6808
- Balasubramanian, V., Vashisht, D., Cletus, J., and Sakthivel, N. (2012). Plant β -1,3-glucanases: their biological functions and transgenic expression against phytopathogenic fungi. *Biotechnol. Lett.* 34 (11), 1983–1990. doi: 10.1007/s10529-012-1012-6
- Balmant, K., Parker, J., Yoo, M. J., Zhu, N., Dufresne, C., and Chen, S. (2015). Redox proteomics of tomato in response to *Pseudomonas syringae* infection. *Hortic. Res.* 2, 15043. doi: 10.1038/hortres.2015.43
- Barsalobres-Cavallari, C. F., Severino, F. E., Maluf, M. P., and Maia, I. G. (2009). Identification of suitable internal control genes for expression studies in *Coffea arabica* under different experimental conditions. *BMC Mol. Biol.* 10, 1. doi: 10.1186/1471-2199-10-1
- Bertrand, B., Vaast, P., Alpizar, E., Etienne, H., Davrieux, F., and Charmetant, P. (2006). Comparison of bean biochemical composition and beverage quality of arabica hybrids involving Sudanese-Ethiopian origins with traditional varieties at various elevations in central America. *Tree Physiol.* 26, 1239–1248. doi: 10.1093/treephys/26.9.1239
- Bittner-Eddy, P. D., Crute, I. R., Holub, E. B., and Beynon, J. L. (2000). *RPP13* is a simple locus in *Arabidopsis thaliana* for alleles that specify downy mildew resistance to different avirulence determinants in *Peronospora parasitica*. *Plant J.* 21, 177–188. doi: 10.1046/j.1365-3113x.2000.00664.x
- Blum, M., Chang, H. Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., et al. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* 49 (D1), 344–354. doi: 10.1093/nar/gkaa977
- Boller, T., and Felix, G. (2009). A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annu. Rev. Plant Biol.* 60, 379–406. doi: 10.1146/annurev.arplant.57.032905.105346
- Bove, J., Kim, C. Y., Gibson, C. A., and Assmann, S. M. (2008). Characterization of wound-responsive RNA-binding proteins and their splice variants in arabidopsis. *Plant Mol. Biol.* 67 (1-2), 71–88. doi: 10.1007/s11103-008-9302-z
- Boyd, L. A., Ridout, C., Sullivan, D. M. O., Leach, J. E., and Leung, H. (2012). Plant – pathogen interactions: Disease resistance in modern agriculture. *Trends Genet.* 29, 233–240. doi: 10.1016/j.tig.2012.10.011
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308
- Brun, H., Chèvre, A., Fitt, B. D. L., Powers, S., Besnard, A., Ermel, M., et al. (2010). Quantitative resistance increases the durability of qualitative resistance to *Leptosphaeria maculans* in *Brassica napus*. *New Phytol.* 185, 285–299. doi: 10.1111/j.1469-8137.2009.03049.x

- Cao, K., Zheng, Z., Wang, L., Liu, X., Zhu, G., Fang, W., et al. (2014). Comparative population genomics reveals the domestication history of the peach, *Prunus persica*, and human influences on perennial fruit crops. *Genome Biol.* 15, 415–429. doi: 10.1186/s13059-014-0415-1
- Chang, S., Puryear, J., and Cairney, J. (1993). A simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol.* 11, 113–116. doi: 10.1007/BF02670468
- Cheval, C., Aldon, D., Galaud, J., and Ranty, B. (2013). Calcium/calmodulin-mediated regulation of plant immunity. *Biochim. Biophys. Acta* 1833, 1766–1771. doi: 10.1016/j.bbamcr.2013.01.031
- Clarindo, W. R., and Carvalho, C. R. (2008). First *Coffea arabica* karyogram showing that this species is a true allotetraploid. *Plant Syst. Evol.* 274, 237–241. doi: 10.1007/s00606-008-0050-y
- Coll, N. S., Epple, P., and Dangel, J. L. (2011). Programmed cell death in the plant immune system. *Cell Death Differ.* 18, 1247–1256. doi: 10.1038/cdd.2011.37
- Corwin, J. A., and Kliebenstein, D. J. (2017). Quantitative resistance: more than just perception of a pathogen. *Plant Cell* 29, 655–665. doi: 10.1105/tpc.16.00915
- Cui, H., Tsuda, K., and Parker, J. E. (2015). Effector-triggered immunity: from pathogen perception to robust defense. *Annu. Rev. Plant Biol.* 66, 487–511. doi: 10.1146/annurev-arplant-050213-040012
- Cui, Y., Zhang, F., and Zhou, Y. (2018). The application of multi-locus GWAS for the detection of salt-tolerance loci in rice. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01464
- Davis, A. P., Gole, T. W., Baena, S., and Moat, J. (2012). The impact of climate change on indigenous arabica coffee (*Coffea arabica*): predicting future trends and identifying priorities. *Plos One* 7, e47981. doi: 10.1371/journal.pone.0047981
- Davis, A. P., Tosh, J., Ruch, N., and Fay, M. (2011). Growing coffee: *Psilanthus* (Rubiaceae) subsumed on the basis of plastid and nuclear DNA sequences; implication for the size, morphology, distribution and evolutionary history of coffee. *Bot. J. Linn. Soc.* 167, 357–377. doi: 10.1111/j.1095-8339.2011.01177.x
- Ding, R., Yang, M., Quan, J., Li, S., Zhuang, Z., Zhou, S., et al. (2019). Single-locus and multi-locus genome-wide association studies for intramuscular fat in duroc pigs. *Front. Plant Sci.* 10. doi: 10.3389/fgene.2019.00619
- Earl, D. A., and von Holdt, B. M. (2012). Structure harvester: A website and program for visualizing STRUCTURE output and implementing the evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7
- Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.* 5, 113. doi: 10.1186/1471-2105-5-113
- Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x
- FAO: Food and Agriculture Organization (1968). *Coffee mission to Ethiopia 1964–65* (Rome: FAO).
- Fatobene, B. J., dos, R., Andrade, V. T., Aloise, G. S., Silvarolla, M. B., Gonçalves, W., et al. (2017). Wild *Coffea arabica* resistant to *Meloidogyne paranaensis* and genetic parameters for resistance. *Euphytica* 213, 196. doi: 10.1007/s10681-017-1986-1
- Felicio, M. S. (2020). *Estudo de associação genômica ampla aplicada ao conteúdo de macronutrientes em grãos de coffea arabica l. [doctor's thesis]* ([Botucatu (SP)]: Universidade Estadual Paulista).
- Harrell, E. F. Jr., and Dupont, C. (2020). “hmisc: harrell miscellaneous,” in *R package version 4.4-0* (CRAN.R-project.org/package=Hmisc).
- Holderbaum, M. M., Ito, D. S., Santiago, D. C., Harumi, L., Fernandes, L. E., and Sera, G. H. (2020). Arabica coffee accessions originated from Ethiopia with resistance to nematode *Meloidogyne paranaensis*. *Aust. J. Crop Sci.* 14, 1209–1213. doi: 10.21475/ajcs.20.14.08.p1763
- Horikoshi, M., and Tang, Y. (2016). “Ggfortify: data visualization tools for statistical analysis results,” in *R package version 4.4-0* (CRAN.R-project.org/package=ggfortify).
- Huang, L., Wang, X., Dong, Y., Long, Y., Hao, C., Yan, L., et al. (2020). Resequencing 93 accessions of coffee unveils independent and parallel selection during *Coffea* species divergence. *Plant Mol. Biol.* 103, 51–61. doi: 10.1007/s11103-020-00974-4
- International Coffee Organization (2020) *Trade statistics tables 2020*. Available at: <https://www.ico.org/prices/po-production.pdf> (Accessed June 15, 2021).
- Ito, D. S., Sera, T., Sera, G. H., Del Grossi, L., and Kanayama, F. S. (2008). Resistance to bacterial blight in arabica coffee cultivars. *Crop Breed. Appl. Biotechnol.* 8, 99–103. doi: 10.12072/1984-7033.v08n02a01
- Jones, J. D. G., and Dangel, J. L. (2006). The plant immune system. *Nat. Rev.* 444, 323–329. doi: 10.1038/nature05286
- Jones, J. D. G., Vance, R. E., and Dangel, J. L. (2016). Intracellular innate immune surveillance devices in plants and animals. *Science* 354, 1117–1126. doi: 10.1126/science.aaf6395
- Kassambara, A., and Mundt, F. (2020). “Factoextra: extract and visualize the results of multivariate data analyses,” in *R package version 1.0.7* (CRAN.R-project.org/package=factoextra).
- Kim, C. Y., Bove, J., and Assmann, S. M. (2008). Overexpression of wound-responsive RNA-binding proteins induces leaf senescence and hypersensitive-like cell death. *New Phytol.* 180 (1), 57–70. doi: 10.1111/j.1469-8137.2008.02557.x
- Krug, C. A., Mendes, J. E. T., and Carvalho, A. (1939). “Taxonomia de coffea arabica l,” in *Boletim técnico n°C62* (SP, Brazil: Instituto Agrônômico de Campinas), 9–57.
- Kushalappa, A. C., and Gunnaiah, R. (2013). Metabolo-proteomics to discover plant biotic stress resistance genes. *Trends Plant Sci.* 18, 522–531. doi: 10.1016/j.tplants.2013.05.002
- Kushalappa, A. C., Yogendra, K. N., and Karre, S. (2016). Plant innate immune response: qualitative and quantitative resistance. *Crit. Rev. Plant Sci.* 35, 38–55. doi: 10.1080/07352689.2016.1148980
- Lam, H., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F., et al. (2011). Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat. Genet.* 42, 1053–1059. doi: 10.1038/ng.715
- Lashermes, P., Andrzejewski, S., Bertrand, B., Combes, M.-C., Dussert, S., Graziosi, G., et al. (2000). Molecular analysis of introgressive breeding in coffee (*Coffea arabica* l.). *Theor. Appl. Genet.* 100, 139–146. doi: 10.1007/s001220050019
- Lashermes, P., Combes, M. C., Robert, J., Trouslot, P., D’Hont, A., Anthony, F., et al. (1999). Molecular characterization and origin of the *Coffea arabica* l. genome. *Mol. Gen. Genet.* 261, 259–266. doi: 10.1007/s004380050965
- Lewis, J. D., Lee, A. H. Y., Hassan, J. A., Wan, J., Hurley, B., Jhingree, J. R., et al. (2013). The *Arabidopsis* ZED1 pseudokinase is required for ZAR1-mediated immunity induced by the *Pseudomonas syringae* type III effector HopZ1a. *PNAS* 110 (46), 18722–18727. doi: 10.1073/pnas.131552011
- Li, C., Fu, Y., Sun, R., Wang, Y., and Wang, Q. (2018). Single-locus and multi-locus genome-wide association studies in the genetic dissection of fiber quality traits in upland cotton (*Gossypium hirsutum* l.). *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01083
- Liscombe, D. K., Macleod, B. P., Loukanina, N., Nandi, O. I., and Facchini, P. J. (2005). Evidence for the monophyletic evolution of benzyloquinoline alkaloid biosynthesis in angiosperms. *Phytochem. Lett.* 66 (11), 1374–1393. doi: 10.1016/j.phytochem.2005.04.029
- Macho, A. P., and Zipfel, C. (2015). Targeting of plant pattern recognition receptor-triggered immunity by bacterial type-III secretion system effectors. *Curr. Opin. Microbiol.* 23, 14–22. doi: 10.1016/j.mib.2014.10.009
- Maciel, K. W., Destefano, S. A. L., Beriam, L. O. S., de Almeida, I. M. G., Patricio, F. R. A., Rodrigues, L. M. R., et al. (2018). Bacterial halo blight of coffee crop: aggressiveness and genetic diversity of strains. *Bragantia* 77, 96–106. doi: 10.1590/1678-4499.2016267
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2019). “Cluster: cluster analysis basics and extensions,” in *R package version 2.1.0* (CRAN.R-project.org/package=cluster).
- Migicovsky, Z., and Myles, S. (2017). Exploiting wild relatives for genomics-assisted breeding of perennial crops. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00460
- Misra, G., Badoni, S., Anacleto, R., Graner, A., Alexandrov, N., and Sreenivasulu, N. (2017). Whole genome sequencing-based association study to unravel genetic architecture of cooked grain width and length traits in rice. *Sci. Rep.* 7, 12478. doi: 10.1038/s41598-017-12778-6
- Mohan, S. K., Cardoso, R. M. L., and Paiva, M. A. (1978). Resistência em germoplasma de *Coffea* ao crestamento bacteriano incitado por *Pseudomonas garcae*. *Pesquisa Agropecuária Bras.* 13, 53–64.
- Money, D., Gardner, K., Migicovsky, Z., Schwaninger, H., Zhong, G. Y., and Myles, S. (2015). LinkImpute: fast and accurate genotype imputation for nonmodel organisms. *Genes Genomes Genet.* 5, 2383–2390. doi: 10.1534/g3.115.021667
- Montagnon, C., and Bouharmont, P. (1996). Multivariate analysis of phenotypic diversity of coffee arabica. *Genet. Resour. Crop Evol.* 43, 221–227. doi: 10.1007/BF00123274
- Moreno, A. A., Mukhtar, M. S., Blanco, F., Boatwright, J. L., Moreno, I., Jordan, M. R., et al. (2012). IRE1/bZIP60-mediated unfolded protein response plays distinct roles in plant immunity and abiotic stress responses. *Plos One* 7 (2), e319. doi: 10.1371/journal.pone.0031944
- Niks, R., Li, X., and Marcel, T. C. (2015). Quantitative resistance to biotrophic filamentous plant pathogens: concepts, misconceptions, and mechanisms. *Annu. Rev. Phytopathol.* 53, 445–470. doi: 10.1146/annurev-phyto-080614-115928
- Nobori, T., Wang, Y., Wu, J., Stolze, S. C., Tsuda, Y., Finkemeier, I., et al. (2020). Multidimensional gene regulatory landscape of a bacterial pathogen in plants. *Nat. Plants* 6, 883–896. doi: 10.1038/s41477-020-0690-7
- Pilet-Nayel, M.-L., Moury, B., Caffier, V., and Montarry, J. (2017). Quantitative resistance to plant pathogens in pyramiding strategies for durable crop protection. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01838

- Pritchard, J. K., Stephens, M., Rosenberg, N. A., and Donnelly, P. (2000). Association mapping in structured populations. *Am. J. Hum. Genet.* 67, 170–181. doi: 10.1086/302959
- Queiroz-Voltan, R. B., Nardin, C. F., Fazuoli, L. C., and Toma-Braghini, M. (2014). Caracterização da anatomia foliar de cafeeiros arábica em diferentes períodos sazonais. *Biotemas* 27, 1–10. doi: 10.5007/2175-7925.2014v27n4p1
- Quenouille, J., Paulhiac, E., Moury, B., and Palloix, A. (2014). Quantitative trait loci from the host genetic background modulate the durability of a resistance gene: a rational basis for sustainable resistance breeding in plants. *Heredity* 112, 579–587. doi: 10.1038/hdy.2013.138
- Ramos, A. H., and Shavdia, L. D. (1976). A die-back of coffee in Kenya. *Plant Dis.* 60, 831–835.
- Razifard, H., Ramos, A., Della Valle, A. L., Bodary, C., Goetz, E., Elizabeth, J., et al. (2020). Genomic evidence for complex domestication history of the cultivated tomato in latin america. *Mol. Biol. Evol.* 37, 1118–1132. doi: 10.1093/molbev/msz297
- Rodrigues, L. M. R., Almeida, I. M. G., Patrício, F. R. A., Beriam, L. O. S., Maciel, K. W., Braghini, M. T., et al. (2017b). Aggressiveness of strains and inoculation methods for resistance assessment to bacterial halo blight on coffee seedlings. *J. Phytopathol.* 165, 105–114. doi: 10.1111/jph.12543
- Rodrigues, L. M. R., Sera, G. H., Guerreiro Filho, O., Beriam, L. O. S., and de Almeida, I. M. G. (2017a). First report of mixed infection by *Pseudomonas syringae* pathovars *garcae* and *tabaci* on coffee plantations. *Plant Prot.* 76, 543–549. doi: 10.1590/1678-4499.2016.399
- RStudio Team (2020). *RStudio: integrated development for r*. RStudio (Boston, MA: PBC). Available at: <http://www.rstudio.com/>.
- Rutledge, R. G., and Stewart, D. (2008). Critical evaluation of methods used to determine amplification efficiency refutes the exponential character of real-time PCR. *BMC Mol. Biol.* 9, 96. doi: 10.1186/1471-2199-9-96
- Salojärvi, J., and Arabica Coffee Genome Consortium. (2021). Chromosome-level assembly of allotetraploid *Coffea arabica* reveals the complex history of a recent allopolyploid. *28th International Conference on Coffee Science (ASIC)*, Montpellier, France S1-PO-19.
- Sant'ana, G. C., Pereira, L. F. P., Pot, D., Ivamoto, S. T., Domingues, D. S., Ferreira, R. V., et al. (2018). Genome-wide association study reveals candidate genes influencing lipids and diterpenes contents in *Coffea arabica* l. *Sci. Rep.* 8, 465. doi: 10.1038/s41598-017-18800-1
- Scalabrín, S., Toniutti, L., Di Gaspero, G., Scaglione, D., Magris, G., Vidotto, M., et al. (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. *Sci. Rep.* 10, 4642. doi: 10.1038/s41598-020-61216-7
- Scholz, S. B. M., Kitzberger, C. S. G., Pagiatto, N. F., Pereira, L. F. P., Davrieux, F., Pot, D., et al. (2016). Chemical composition in wild ethiopian arabica coffee accessions. *Euphytica* 209, 429–438. doi: 10.1007/s10681-016-1653-y
- Silvarolla, M. B., Mazzafera, P., and Fazuoli, L. C. (2004). A naturally decaffeinated arabica coffee. *Nature* 429, 826. doi: 10.1038/429826a
- Silvestrini, M., Junqueira, M. G., Favarin, A. C., Guerreiro-Filho, O., Maluf, M. P., Silvarolla, M. B., et al. (2007). Genetic diversity and structure of Ethiopian, Yemen and Brazilian *Coffea arabica* l. accessions using microsatellites markers. *Genet. Resour. Crop Evol.* 54, 1367–1379. doi: 10.1007/s10722-006-9122-4
- Stalker, H. T., Tallury, S. P., Bertioli, D., and Bertioli, S. C. L. (2013). The value of diploid peanut relatives for breeding and genomics. *Peanut Sci.* 40, 70–88. doi: 10.3146/PS13-6.1
- St. Clair, D. A. (2010). Quantitative disease resistance and quantitative resistance loci in breeding. *Annu. Rev. Phytopathol.* 48, 247–268. doi: 10.1146/annurev-phyto-080508-081904
- Steiger, D. L., Nagai, C., Moore, P. H., Morden, C. W., Osgood, R. V., and Ming, R. (2002). AFLP analysis of genetic diversity within and among *Coffea arabica* cultivars. *Theor. Appl. Genet.* 105, 209–215. doi: 10.1007/s00122-002-0939-8
- Su, J., Ma, Q., Li, M., Hao, F., and Wang, C. (2018). Multi-locus genome-wide association studies of fiber-quality related traits in chinese early-maturity upland cotton. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01169
- Tamba, C. L., Ni, Y. L., and Zhang, Y. M. (2017). Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies. *Plos Comput. Biol.* 13, e1005357. doi: 10.1371/journal.pcbi.1005357
- Tamba, C. L., and Zhang, Y. (2018). A fast mrMLM algorithm for multi-locus genome-wide association studies. *BioRxiv* 341784. doi: 10.1101/341784v1
- van der Vossen, H. A. M. (1985). “Coffee selection and breeding,” in *Coffee: botany biochemistry and production of beans and beverage*. Eds. M. N. Clifford and K. C. Wilson (Westport, London: Croom Helm), 48–96.
- van der Vossen, H., Bertrand, B., and Charrier, A. (2015). Next generation variety development for sustainable production of arabica coffee (*Coffea arabica* l.): a review. *Euphytica* 204, 243–256. doi: 10.1007/s10681-015-1398-z
- van der Vossen, H. A. M., and Walyaro, D. J. (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
- Vidal, R. O., Mondego, J. M. C., Pot, D., Ambrosio, A. B., Andrade, A. C., Pereira, L. F. P., et al. (2010). A high-throughput data mining of single nucleotide polymorphisms in *Coffea* species expressed sequence tags suggests differential homeologous gene expression in the allotetraploid *Coffea arabica*. *Plant Physiol.* 154, 1053–1066. doi: 10.1104/pp.110.162438
- Villa-Rivera, M. G., Cano-Camacho, H., López-Romero, E., and Zavala-Páramo, M. G. (2021). The role of arabinogalactan type II degradation in plant-microbe interaction. *Front. Microbiol.* 12. doi: 10.3389/fmicb.2021.730543
- Wang, S. B., Feng, J. Y., Ren, W. L., Huang, B., Zhou, L., Wen, Y. J., et al. (2016). Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Sci. Rep.* 6, 19444. doi: 10.1038/srep19444
- Wei, T., and Simko, V. (2021). “R package ‘corrplot’: visualization of a correlation matrix (version 0.88),” in CRAN (R-project.org/package=corrplot).
- Wen, Y. J., Zhang, H., Ni, Y. L., Huang, B., Zhang, J., Feng, J. Y., et al. (2017). Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Brief. Bioinform.* 19, 700–712. doi: 10.1093/bib/bbw145
- Worku, M., DeMeulenaer, B., and Duchateau, L. (2018). Effect of altitude on biochemical composition and quality of green arabica coffee beans can be affected by shade and postharvest processing method. *Food Res. Int.* 105, 278–285. doi: 10.1016/j.foodres.2017.11.016
- Xuehui, B., Lihong, Z., Yongliang, H., Guanghai, J., Jinhong, L., and Zhang, H. (2013). Isolation and identification of the pathogen of coffee bacterial blight disease. *Chin. J. Trop. Crop* 34, 738–742. doi: 10.3969/j.issn.1000-2561.2013.04.028
- Xu, Y., Yang, T., Zhou, Y., Yin, S., Li, P., Liu, J., et al. (2018). Genome-wide association mapping of starch pasting properties in maize using single-locus and multi-locus models. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01311
- Yogendra, K. N., Dhokane, D., Kushalappa, A. C., Sarmiento, F., Rodriguez, E., and Mosquera, T. (2017). StWRKY8 transcription factor regulates benzyloisoquinoline alkaloid pathway in potato conferring resistance to late blight. *Plant Sci.* 256, 208–216. doi: 10.1016/j.plantsci.2016.12.014
- Yuan, M., Jiang, Z., Bi, G., Nomura, K., Liu, M., Wang, Y., et al. (2021). Pattern-recognition receptors are required for NLR-mediated plant immunity. *Nature* 592, 105–109. doi: 10.1038/s41586-021-03316-6
- Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208. doi: 10.1038/ng1702
- Zhang, Y. M., Jia, Z., and Dunwell, J. M. (2019). Editorial: The applications of new multi-locus gwas methodologies in the genetic dissection of complex traits. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00100
- Zhang, Q., Wu, L., Yin, H., Xu, Z., Zhao, Y., Gao, M., et al. (2021). D6 protein kinase in root xylem benefiting resistance to *Fusarium* reveals infection and defense mechanisms in tung trees. *Hortic. Res.* 8, 240. doi: 10.1038/s41438-021-00656-2
- Zhang, X., Zhang, J., He, X., Wang, Y., Ma, X., and Yin, D. (2017). Genome-wide association study of major agronomic traits related to domestication in peanut. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01611
- Zhao, S., Zheng, F., He, W., Wu, H., Pan, S., and Lam, H. (2015). Impacts of nucleotide fixation during soybean domestication and improvement. *BMC Plant Biol.* 15, 81. doi: 10.1186/s12870-015-0463-z
- Zipfel, C., and Rathjen, J. P. (2008). Plant immunity: AvrPto targets the frontline. *Curr. Biol.* 18, 218–220. doi: 10.1016/j.cub.2008.01.016
- Zong, N., Xiang, T., Zou, Y., Chai, J., and Zhou, J. M. (2008). Blocking and triggering of plant immunity by *Pseudomonas syringae* effector AvrPto. *Plant Signal. Behav.* 3, 583–585. doi: 10.4161/psb.3.8.5741