UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS AGRÁRIAS E ENGENHARIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA

IASMINE RAMOS ZAIDAN

GERMPLASM CHARACTERIZATION, DIVERSITY STUDY AND GENOME-WIDE ASSOCIATION IN *Coffea canephora*

ALEGRE
ESPÍRITO SANTO – BRASIL
2022

IASMINE RAMOS ZAIDAN

GERMPLASM CHARACTERIZATION, DIVERSITY STUDY AND GENOME-WIDE ASSOCIATION IN *Coffea canephora*

Tese apresentada ao Programa de Pós-Graduação em Agronomia do Centro de Ciências Agrárias e Engenharias da Universidade Federal do Espírito Santo, como requisito parcial para a obtenção do título de Doutor em Agronomia, na linha de pesquisa Produção de Plantas Cultivadas e Nativas

Orientadora: Prof[a]. DSc Marcia Flores da Silva Ferreira

Coorientador: Prof. DSc Adésio Ferreira

ALEGRE

ESPÍRITO SANTO – BRASIL

2022

FICHA CATALOGRÁFICA

Ficha catalográfica disponibilizada pelo Sistema Integrado de Bibliotecas - SIBI/UFES e elaborada pelo autor

---

Zaidan, Iasmine Ramos, 1992-

Z21g Germplasm characterization, diversity study and genome-wide association in *Coffea canephora* / Iasmine Ramos Zaidan. - 2022.

170 f. : il.

Orientadora: Marcia Flores da Silva Ferreira.

Coorientador: Adésio Ferreira.

Tese (Doutorado em Produção Vegetal) - Universidade Federal do Espírito Santo, Centro de Ciências Agrárias e Engenharias.

I. Ferreira, Marcia Flores da Silva. II. Ferreira, Adésio. III. Universidade Federal do Espírito Santo. Centro de Ciências Agrárias e Engenharias. IV. Título.

CDU: 63

IASMINE RAMOS ZAIDAN

GERMPLASM CHARACTERIZATION, DIVERSITY STUDY AND GENOME-WIDE ASSOCIATION IN *Coffea canephora*

Tese apresentada à Universidade Federal do Espírito Santo como parte das exigências do Programa de Pós-Graduação em Agronomia, para a obtenção do título de Doctor Scientiae em Agronomia, na área de concentração em Produção de Plantas Cultivadas e Nativas.

Aprovada em 19 de julho de 2022.

Comissão examinadora:

Márcia Flores da Silva Ferreira

Profª. D.Sc Marcia Flores da Silva Ferreira
Centro de Ciência Agrárias e Engenharias – UFES
Orientadora

Adésio Ferreira

Prof. D.Sc Adésio Ferreira
Centro de Ciência Agrárias e Engenharias – UFES
Coorientador

Milene Miranda Praça Fontes

Profª. D.Sc Milene Miranda Praça Fontes
Centro de Ciências Exatas, Naturais e da Saúde – UFES
Membro Interno

Prof. D.Sc Luiz Antônio dos Santos Dias
Centro de Ciências Agrárias – UFV
Membro Externo

Profª D.Sc Eveline Teixeira Caixeta Moura
Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA
Membro Externo

BIOGRAFIA

IASMINE RAMOS ZAIDAN, filha de Marcos Tadeu Neves Zaidan e Ângela Márcia Ramos Zaidan, nasceu em 24 de fevereiro de 1992, no município de Viçosa, MG. Possui ensino médio (2º grau completo) pelo colégio Nossa Senhora do Carmo, de Viçosa, MG (2009). Graduou-se em Agronomia pela Universidade Federal de Viçosa (2015). Durante a graduação, foi estagiária do CENTREINAR, da Universidade Federal de Viçosa (2013) e da EMATER-MG (2015). Foi bolsista de iniciação científica vinculada ao projeto "Influência da atmosfera modificada e refrigeração sobre propriedades físico-químicas e reológicas da Atemoia (*Annona squamosa* L.) durante o armazenamento", financiado pela PROBIC/FAPEMIG, na Universidade Federal de Viçosa (2013). Foi ainda bolsista de iniciação científica vinculada ao projeto "Estudo e caracterização do processo de produção de graviola em pó através de secagem em leito de espuma", financiado por PIBIC/CNPq, na Universidade Federal de Viçosa (2014). Em agosto de 2016, ingressou no curso de mestrado do Programa de Pós-Graduação em Fitotecnia, pela Universidade Federa de Viçosa, na área de Biotecnologia e, em julho de 2018 concluiu o curso de mestrado com dissertação com título "Potencial biocida de extratos de *Jatropha curcas* L. sobre *Hemileia vastatrix* e *Cercospora coffeicola*". Atualmente é estudante de doutorado da Universidade Federal do Espírito Santo, no Programa de Pós-Graduação em Agronomia, na área de Produção de Plantas Cultivadas e Nativas, com linha de pesquisa voltada para diversidade molecular, relação genética, desenvolvimento e produção de genótipos de *Coffea canephora* (café Conilon).

DEDICATÓRIA

A Deus por ter me abençoado e me guiado durante estes quatro anos de Doutorado; Ao meu pai, Marcos, pelo constante incentivo, por ser meu porto seguro e por ter sido meu apoio incondicional durante esta etapa da minha vida; À Úrsula, minha querida irmã, por ser meu exemplo de responsabilidade, competência, determinação e pelas conversas diárias que tornaram meu caminho mais leve; Ao Samir, meu querido irmão, que diante dos desafios sempre foi meu exemplo de coragem, disciplina, foco e determinação; Ao meu cunhado Rafael, pelo apoio constante, pelo carinho e incentivo durante minha caminhada acadêmica; E, ao meu anjo mãe, Ângela, a quem devo toda minha gratidão para conseguir chegar até aqui, seu amor, sua força e vontade de viver me impulsionaram nesta caminhada e me sustentaram nessa jornada!

AGRADECIMENTOS

Agradeço a Deus por todas as bençãos e graças até hoje recebidas, por ter me sustentado nos momentos desafiadores vividos durante os quatro anos de Doutorado e por toda proteção diante das dificuldades que foram enfrentadas e vencidas.

Ao meu pai, Marcos, por ser meu exemplo de comprometimento com o trabalho, por ser meu porto seguro diante dos desafios da vida, por todo amor, carinho, afeto e proteção, você é minha fortaleza!

À minha querida irmã, Úrsula, minha companheira de vida que dividiu comigo todos os processos desafiadores desta jornada. Gratidão por todo amor, dedicação, cumplicidade e cuidado durante todas as etapas da minha vida, você é a luz que ilumina meus dias!

Ao meu querido irmão, Samir, por ser meu exemplo de persistência, foco, determinação, disciplina e coragem para enfrentar os desafios da vida, você é uma grande inspiração para mim!

Ao meu cunhado, Rafael, por todo carinho, respeito, pelo auxílio nas análises estatísticas e incentivo constante durante minha caminhada acadêmica.

À minha mãe, Ângela (*in memoriam*), a quem eu não tenho palavras para descrever a gratidão por ter sido meu exemplo de força, vontade de viver e grande incentivadora dos meus sonhos. Nos dias mais difíceis, em que os desafios pareciam insuperáveis, eu me lembrava da promessa que lhe fiz, de que um dia eu seria Doutora. Foi por você!

À Universidade Federal do Espírito Santo, ao Centro de Ciências Agrárias e Engenharias e ao Programa de Pós-Graduação em Agronomia, pelo ensino de excelência e por me proporcionar a oportunidade cursar o Doutorado. À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudos.

À minha orientadora Marcia Flores da Silva Ferreira e ao meu coorientador Adésio Ferreira pela orientação, apoio e incentivo constantes, por toda paciência, por todo conhecimento transmitido e pelo suporte durante as análises e escrita da tese. Agradeço à professora Camila Ferreira Azevedo pelo auxílio nas análises estatísticas. Agradeço aos membros da banca, Eveline Teixeira Caixeta Moura, Luiz Antônio dos Santos Dias e Milene Miranda Praça Fontes, pela participação e pelas contribuições com meu trabalho.

Aos produtores que disponibilizaram as sementes para a instalação do experimento, aos colegas Edson, Ramon e Valderson que auxiliaram na coleta das sementes e no plantio das mudas e ao Incaper pela concessão de amostras foliares para extração do DNA.

À minha psicóloga Nena, por ter me auxiliado no meu crescimento pessoal, profissional e por ter me ajudado a superar meus medos e frustrações.

Aos colegas dos Laboratórios de Genética e Melhoramento e Biometria, pela amizade, companheirismo e por toda ajuda durante o processo de execução do experimento, avaliações de campo, parte de bancada do laboratório e análises de dados. Agradeço pelas contribuições durante as análises, mas também pelas boas risadas que tornaram esta caminhada mais leve!

Agradeço aos meus familiares e amigos por sempre estarem na torcida e por acreditarem em mim. O amor e apoio de vocês foi fundamental neste processo!

Gratidão a todos que contribuíram de alguma forma para este sonho se tornasse realidade.

EPÍGRAFE

“Você ganha forças, coragem e confiança através de cada experiência em que você realmente para e encara o medo de frente. Você tem que fazer exatamente aquilo que acha que não consegue fazer.”

(Eleanor Roosevelt)

SUMÁRIO

ABSTRACT

ZAIDAN, Iasmine Ramos. **GERMPLASM CHARACTERIZATION, DIVERSITY STUDY AND GENOME-WIDE ASSOCIATION IN *Coffea canephora*.** Doutorado em Agronomia – Universidade Federal do Espírito Santo, Alegre – ES. Orientadora: DSc Marcia Flores da Silva Ferreira. Coorientador: DSc Adésio Ferreira

The germplasm remaining from ancient seminal crops of Conilon (*Coffea canephora*) in the south of Espírito Santo must be characterized and preserved as they are potential sources of genetic variability and resources for genetic ~~breeding~~ programs. With the release of commercial clones and the use of generally few clonal materials, there may be a narrowing of the genetic base. The aim of this study were (i) to characterize and investigate the genetic diversity of 388 half-sib families of new and promising genotypes of *C. canephora* from old seminal crops in the south of Espírito Santo, based on morpho-agronomic traits; (ii) to evaluate the genetic diversity and population structure of 280 genotypes and 29 commercial clones cultivated in the state by high throughput genotyping SNP data; (iii) to identify chromosomal regions with significant associations with eight phenotypic traits by GWAS. The experiment was installed in April 2018 in Mimoso do Sul - ES in Federer augmented blocks with five commercial clones used as controls. Each of 388 half-sib families had five plants, totaling 2,085 plants. The characteristics evaluated were plant height, stem diameter, average height growth rate, average diameter growth rate, rust incidence, leaf miner infestation, mealybug infestation, drought tolerance. Biotic and abiotic stresses were assessed by visual analysis using a rating scale. Cluster analyzes were performed for the data referring to the average growth rates that were subjected to linear regression and the values of the regression constants and regression coefficients were used in the construction of the dendrogram, a second cluster was made to all data referring to the latest measurements and Pearson's correlation analysis was done to detect the correlation between the traits. A total of 251 genotypes selected in the experimental field, including commercial clones, and 29 genotypes from Incaper were high-throughput genotyping using the DArTseq methodology. SNP were filtering by quality parameters and the data were used for genetic diversity, population structure and molecular analysis of variance. The GWAS study includes 251 genotypes and eight traits. For the phenotypic data, mixed models were used, and, from the sum of the genetic values and the residues of the genotyped materials, the p-values were calculated. The SNP markers with significant associations ($p < 0.05$) along the 11 chromosomes of *C. canephora* had the putative function of the genes verified. In the

morpho-agronomic characterization, eight divergent groups and in the second analysis nine groups detected, shows the germplasm variation. Positive correlations were detected among plant height, stem diameter, average height growth rate and average diameter growth rate. Negative correlations for all stresses with the characters of vegetative development indicated that such adverse conditions impaired the initial development of the coffee plant. A total of 2,542 filtered SNPs revealed six groups by cluster analysis and two genetic groups by STRUCTURE. The high frequency of heterozygotes (He) for some of the groups formed by the cluster analysis indicates the genetic diversity. High F_{ST} values were detected between groups 4 and 2 (0.60), 2 and 5 (0.60), 4 and 3 (0.50) and between the two genetic groups (0.59). In the GWAS analysis, 115 SNPs showed significant associations with seven traits: 48 for height growth rate; 20 for mealybug incidence; seven for plant height and diameter growth rate; 11 for stem diameter; 16 for incidence of rust and five for incidence of leaf mining. Most of these SNPs are located within or close to candidate genes. The putative function of these candidate genes *GSCOC_T00019303001*, *GSCOC_T00022693001*, *GSCOC_T00039643001* and *GSCOC_T00040251001* are related to plant defense mechanisms to protect against pathogens, pests and abiotic stresses. Other candidate genes such as *GSCOC_T00040077001*, *GSCOC_T00028217001* and *GSCOC_T00021883001* presented putative functions related to plant development and plant hormones. All the results show the importance of preserving old crops in the south of Espírito Santo as a source of genetic resources for coffee breeding programs. The phenotypic and genetic diversity detected demonstrate that the old seminal crops in the south of the state constitute a very rich germplasm bank. This valuable source must be maintained and conserved to guarantee the sustainability of the coffee crop to expand the genetic base that was reduced with the replacement of clonal cultivars. In addition, chromosomal regions detected associated with important traits can be used in studies of selection assisted by molecular markers to select plants of favorable attributes for the genetic breeding of coffee.

1. Introduction

The *Coffea canephora* germplasm in the state of Espírito Santo is conserved by public institutions such as Incaper, UFES, IFES and by family farmers (Souza et al. 2013; Ferrão et al. 2019). This germplasm conserved by family farmers has great genetic diversity because they were planted via seed (Ferrão et al. 2019). Planting made from seeds is highly heterogeneous and may have adapted plants with characteristics of agronomic interest. (Ferrão et al. 2019).

Coffea canephora was introduced in Brazil in 1911 in the municipality of Cachoeiro de Itapemirim, southern region of the state of Espírito Santo (Ferrão et al. 2007; Merlo, 2012). Despite the introduction at the beginning of the last century, the increased interest in the species in the country occurred due to the coffee eradication program, with the implementation of the coffee crop renewal plan in the north of Espírito Santo with the recomposition of crops with the new genetic material (Ferrão et al. 2007; Zambolim, 2009), as well as the emergence of soluble coffee in the 1950s and its use in roasted and ground coffee blends (Ferrão et al. 2019).

The wide genetic variability found in *C. canephora* is explained by the species' natural allogamy as a function of gametophytic self-incompatibility governed by the S allele (Lashermes et al. 1996). In addition to this factor, the diversity detected can be explained by the wide demographic distribution of this species, one of the widest within the genus *Coffea* (Maurin et al. 2007) and by the recent domestication process (Musoli et al. 2007). In view of this, natural populations and crop formed by seeds, even when collected from a single matrix plant, can show great genetic variability with a high frequency of heterozygosity (Ferrão et al. 2019).

However, with the release of the first clonal cultivars to Espírito Santo in 1993 (Bragança et al., 1993), the increased use of these clonal cultivars and the adoption of relatively few clones in the renewal of coffee plantations can lead to a loss of diversity and narrowing of the genetic base (Ferrão et al. 2007). Thus, the remaining old crops in the south of the state of Espírito Santo need to be characterized and preserved, as they present the greatest variability of this species in the country (Fonseca et al., 1996), since, currently, most commercial crops in Espírito Santo and Bahia use clonal seedlings (Espíndula and Partelli 2011).

Two botanical varieties of *C. canephora* are commercially cultivated in Brazil (Maurin et al., 2007; Batista-santos et al., 2011) 'Robusta' and 'Conilon' which are divergent heterotic groups with complementary characteristics (Souza et al., 2013, Bilika et al., 2017). The variety Conilon is characterized by shrubby growth, elongated leaves, early maturation, drought tolerance, but it is more sensitive to pests and diseases (Montagnon et al. 2012; Santos et al. 2017; Oliveira et al. 2018). The Robusta variety presents erect growth, larger leaves, medium-high sieve, late maturation, greater resistance to pests and diseases and less tolerance to drought (Montagnon et al. 2012; Santos et al. 2017; Oliveira et al. 2018). The Conilon and Robusta varieties belong to the Congolese group, the conilon belongs to subgroup SG1 (main cultivars in Brazil), while Robusta belongs to the subgroups SG2, B and C (Marraccini et al., 2012).

Studying the morpho-agronomic traits provides the detection of divergent and superior individuals, and these individuals can be used for the development of new cultivars that present greater amounts of attributes related to traits of agronomic interest (Dubberstein et al., 2021). In the *Coffea canephora* breeding programs, one of the main objectives is the selection of highly productive plants and, in addition to the characteristics related to production, the selection of more vigorous genotypes, with greater diameter of the crown and stem, associated with a lower plant height (Alkimim et al. 2017). Other important trait in the selection of superior genotypes is the identification of tolerant/resistant plants to the main stresses that affect the crop, whether biotic or abiotic, such as leaf rust, leaf miner, mealybug and drought (Mohammed 2015; Talhinhas et al. 2017).

Previous studies have shown that *C. canephora* germplasm banks have wide phenotypic variability for traits related to the root system (Silva et al. 2020), morphology of flower traits (Silva et al. 2021), to the concentration of nutrients in leaves (Siva et al. 2021), to fruit production (Partelli et al. 2021) and, to stomatal characteristics (Dubberstein, et al. 2021). Other studies evaluating several morpho-agronomic traits such as traits related to the plant, branches, leaves, cycle, fruits, seeds, response to pests, diseases, drought and harvest, also detected this wide diversity of *C. canephora* (Ferrão et al. 2021).

The genetic diversity and population structure of *C. canephora* has been detected through SNPs (*single nucleotide polymorphisms*) markers (Garavito et al. 2016; Bikila et al. 2017; Alkimim et al. 2018; Anagbogu et al. 2019; Spinoso- Castillo et al. 2020). These

markers are the most abundant type of polymorphism in the genome, being generally biallelic and codominant (Resende et al. 2008; Liao and Lee, 2010) and may be associated with genes that control key traits of agronomic interest (Heffner et al. 2009; Sousa et al. 2017). The DArTseq™ methodology consists of highly informative, high-performance sequencing (Spinoso-Castillo et al. 2020) using a combination of the methodology DArTseq™ and next-generation sequencing for rapid, large-scale discovery of SNPs in a wide variety of non-model organisms (Kilian et al. 2012; Cruz et al. 2013; Raman et al. 2014).

From phenotypic information and molecular data, genome wide association studies (GWAS) are possible, which search for the association between the phenotype of interest and the genotype, enabling the identification of regions of the genome that have the greatest effect on a given trait (Sant' Ana et al. 2018). Through this type of analysis, it became possible to find genes that contain significant SNPs and the identification of genes that participate in the control of the trait and its biological function, facilitating the understanding of the influence of the genotype on the phenotype (Yang et al. 2013).

The diversity of *C. canephora* has been studied and detected from morpho-agronomic characterization and from studies of genetic diversity and population structure through molecular markers. However, with the renewal of crops using clonal cultivars, and the use of few clones, this wide variability detected can be lost, leading to a narrowing of the genetic base of coffee. Therefore, it is extremely interesting to characterize, study and, mainly, preserve these genotypes remaining from old crops in the south of Espírito Santo, which are potential sources of variability and resources for genetic breeding programs.

In *C. canephora*, the GWAS studies are scarce, demonstrating the importance of this work that will provide information of great importance for the genetic improvement programs of the species *C. canephora*. With the detection of chromosomal regions, these can be used in selection studies by molecular markers, helping genetic improvement programs.

2. References

Alkimim ER, Caixeta ET, Sousa TV, Pereira AA, de Oliveira ACB, Zambolim L, Sakiyama NS (2017) Marker-assisted selection provides arabica coffee with genes from other *Coffea* species targeting on multiple incidence to rust and coffee berry disease. Molecular Breeding, 37(1):1–10.

Alkimim ER, Caixeta ET, Sousa TV, da Silva FL, Sakiyama NS, Zambolim L (2018) High-throughput targeted genotyping using next-generation sequencing applied in *Coffea canephora* breeding. Euphytica, 214:1–18. https://doi:10.1007/s10681-018-2126-2

Anagbogu CF, Bhattacharjee R, Ilori C, Tongyoo P, Dada KE, Muyiwa AA, Gepts P, Beckles DM (2019) Genetic diversity and re-classification of coffee (*Coffea canephora* Pierre ex A. Froehner) from South Western Nigeria through genotyping-by-sequencing-single nucleotide polymorphism analysis. Genetic Resourses and Crop Evolution 66:685–696. https://doi.org/10.1007/s10722-019-00744-2

Batista-santos P, Lidon FC, Fortunato A, et al (2011) The impact of cold on photosynthesis in genotypes of *Coffea* spp. Photosystem sensitivity, photoprotective mechanisms and gene expression. J Plant Physiol 168:792–806. https://doi.org/10.1016/j.jplph.2010.11.013

Bikila BA, Sakiyama NS, Caixeta ET (2017) SNPs Based Molecular Diversity of *Coffea canephora*. Journal of Microbiology & Experimentation 5:1–4. https://doi.org/10.15406/jmen.2017.05.00136

Bragança SM, Fonseca AD, Silveira J, Ferrão RG, Carvalho C (1993) "EMCAPA 8111"," EMCAPA 8121"," EMCAPA 8131": primeiras variedades clonais de café conilon lancadas para o Espírito Santo. EMCAPA, Vitória, Brazil

Cruz VMV, Kilian A, Dierig DA (2013) Development of DArT Marker Platforms and Genetic Diversity Assessment of the U.S. Collection of the New Oilseed Crop Lesquerella and Related Species. PLoS One 8:1–13. https://doi.org/10.1371/journal.pone.0064062

Dubberstein D, Oliveira MG, Aoyama EM, Guilhen JH, Ferreira A, Marques I, ..., Partelli FL (2021) Diversity of Leaf Stomatal Traits among *Coffea canephora* Pierre ex A. Froehner Genotypes. Agronomy, 11(6):1126. https://doi.org/10.3390/agronomy11061126

Ferrão RG, Fonseca AFA, da Bragança SM, Ferrão MAG, De Muner LH (2007) Café conilon. Incaper.

Ferrão RG, Fonseca AFA, Ferrão MAG, de Muner LH (2019) Conilon Coffee. Vitória, ES, Brazil

Ferrão MAG, Mendonça RFD, Fonseca AFA, Ferrão RG, Senra JFB, Volpi PS, ..., Comério M (2021) Characterization and genetic diversity of *Coffea canephora* accessions in a germplasm bank in Espírito Santo, Brazil. Crop Breeding and Applied Biotechnology, 21:1–10. https://doi.org/10.1590/1984-70332021v21n2a32

Ferrão R, Volpi P, Senra JDB, Comério M, Ferrão M, Riva-Souza EM, ..., Verdin Filho, AC (2022) Variabilidade de *Coffea canephora* do Banco Ativo de Germoplasma do Incaper: Caracterização dos Acessos com Base em Descritores Mínimos.

Fonseca AD (1996). Propagação assexuada de *Coffea canephora* no Estado do Espírito Santo. In Workshop sobre avanços na propagação de plantas lenhosas, PAIVA, R.(ed.) Univ. Federal de Lavras-UFLA, LAVRAS-MG (pp. 31-34).

Garavito A, Montagnon C, Guyot R, Bertrand B (2016) Identification by the DArTseq method of the genetic origin of the *Coffea canephora* cultivated in Vietnam and Mexico. BMC Plant Biology 16:1–12. https://doi.org/10.1186/s12870-016-0933-y

Heffner EL, Sorrells ME, Jannink JL (2009) Genomic selection for crop improvement. Crop Sci. 49:1–12.

Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, Caig V, Heller-Uszynska K, Jaccoud D, Hopper C, Aschenbrenner-Kilian M, Evers M, Peng K, Cayla C, Hok P, Uszynski G (2012) Diversity arrays technology: A generic genome profiling technology on open platforms. Methods in Molecular Biology 888:67–89. https://doi.org/10.1007/978-1-61779-870-2_5

Lashermes P, Couturon E, Moreau N, et al (1996) Inheritance and genetic mapping of self-incompatibility in *Coffea canephora* Pierre. Theor Appl Genet 93:458–462. https://doi.org/10.1007/BF00223190

Marraccini P, Vinecky F, Alves GSC, et al (2012) Differentially expressed genes and proteins upon drought acclimation in tolerant and sensitive genotypes of *Coffea canephora*. J Exp Bot 63:4191–4212.https://doi.org/10.1093/jxb/ers103

Maurin O, Davis AP, Chester M, Mvungi EF, Jaufeerally-Fakim Y, Fay MF (2007) Towards a Phylogeny for *Coffea* (Rubiaceae): Identifying Well-supported Lineages Based on Nuclear and Plastid DNA Sequences. Annals of Botany 100:1565–1583. https://doi.org/10.1093/aob/mcm257

Merlo PMS (2012) Conilon capixaba: 100 anos de desafios, crescimento e evolução. Vitória, ES: Bumerangue Produção de Comunicação. 100 p.

Mohammed A. (2015) Importance and Characterization of Coffee Berry Disease (*Colletotrichum kahawae*) in Borena and Guji Zones, Southern Ethiopia. J Plant Pathol Microbiol. 06(09)

Montagnon C, Cubry P, Leroy T (2012) Amélioration génétique du caféier *Coffea canephora* Pierre : connaissances acquises, stratégies et perspectives. 21:

Musoli P, Cubry P, Aluka P, Billot C, Dufour M, de Bellis F, Pot D, Bieysse D, Charrier A, Leroy T (2009) Genetic differentiation of wild and cultivated populations: Diversity of *Coffea canephora* Pierre in Uganda. Genome 52:634–646. https://doi.org/10.1139/G09-037

Oliveira LNL de, Rocha RB, Ferreira FM, et al (2018) Selection of *Coffea canephora* parents from the botanical varieties Conilon and Robusta for the production of intervarietal hybrids. 1–7

Partelli FL, Oliosi G, Dalazen JR, da Silva CA, Vieira HD, Espindula MC (2021) Proportion of ripe fruit weight and volume to green coffee: Differences in 43 genotypes of *Coffea canephora*. Agronomy Journal, 113(2):1050-1057.

Raman H, Raman R, Kilian A, Detering F, Carling J, Coombes N, Diffey S, Kadkol G, Edwards D, McCully M, Ruperao P, Parkin IAP, Batley J, Luckett DJ, Wratten N (2014) Genome-wide delineation of natural variation for pod shatter resistance in *Brassica napus*. PloS one 9:e101673. https://doi.org/10.1371/journal.pone.0101673

Resende MDV et al. (2008) Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético, Pesquisa Florestal Brasileira, Vol 56,p 63-77

Santos AV, Rocha RB, Fernandes C de F, et al (2017) Reaction of *Coffea canephora* clones to the root knot nematode, Meloidogyne incognita. https://doi.org/10.5897/AJAR2016.11999

Silva LOE, Schmidt R, Valani GP, Ferreira A, Ribeiro-Barros AI, & Partelli FL (2020) Root trait variability in *Coffea canephora* genotypes and its relation to plant height and crop yield. Agronomy, 10(9), 1394.

Silva CAD, Partelli FL, Aoyama EM, Bonomo R, Vieira HD, Ramalho JC, Ribeiro-Barros AI (2021) Floral morphology of robusta coffee genotypes. Agronomy Journal, 113(4): 3080-3088. https://doi.org/10.1002/agj2.20743

Silva CA da, Santos EA, Viana AP, Dias JRM, Partelli FL (2021) Genetic diversity in *Coffea canephora* genotypes for leaf nutrient concentration. Revista de la Facultad de Ciencias Agrarias UNCuyo, 53(1):22-34. https://doi.org/10.48162/rev.39.003

Sousa TV, Caixeta ET, Alkimim ER, de Oliveira ACB, Pereira AA, Sakiyama NS, Resende Júnior MFR, Zambolim L (2017) Population structure and genetic diversity of coffee progenies derived from Catuaí and Híbrido de Timor revealed by genome-wide SNP marker. Tree Genetics & Genomes 13:1–16. https://doi.org/10.1007/s11295-017-1208-y

Souza F de F, Caixeta ET, Ferrão LFV, Pena GF, Sakiyama NS, Zambolim EM, Zambolim L, Cruz CD (2013) Molecular diversity in *Coffea canephora* germplasm conserved and cultivated in Brazil. Crop Breeding and Applied Biotechnology 13:221–227. https://doi.org/10.1590/s1984-70332013000400001

Spinoso-Castillo JL, Escamilla-Prado E, Aguilar-Rincón VH, Ramos VM, de los Santos GG, Pérez-Rodríguez P, Corona-Torres T (2020) Genetic diversity of coffee (*Coffea* spp.) in Mexico evaluated by using DArTseq and SNP markers. Genetic Resources and Crop Evolution 67:1795–1806. https://doi.org/10.1007/s10722-020-00940-5

Talhinhas P, Batista D, Diniz I, Vieira A, Silva DN, Loureiro A, Tavares S, Pereira AP, Azinheira HG, Guerra-Guimarães L, et al. (2017) The coffee leaf rust pathogen *Hemileia vastatrix*: one and a half centuries around the tropics. Mol Plant Pathol. 18(8):1039–1051.

Yang J, Lee SH, Goddard ME, Visscher PM (2013) Genome-wide complex trait analysis (GCTA): methods, data analyses, and interpretations. In: *Genome-wide association studies and genomic prediction* (pp. 215-236). Humana Press, Totowa, NJ.

Zambolim L (2009) Tecnologias para produção do café conilon. Viçosa: Universidade.

CHAPTER 1

CHARACTERIZATION OF *Coffea canephora* GERMOPLASM BASED ON INITIAL DEVELOPMENT CHARACTERS AND STRESS

ABSTRACT

The germplasm remaining from old seminal crops of Conilon (*Coffea canephora*) in the south of Espírito Santo state is a potential source of genetic variability and a resource for genetic breeding programs, so it must be characterized and preserved. In this study, the genetic diversity of 388 half-sib families of *C. canephora*, from seeds of matrices collected in old seminal crops in Espírito Santo, was assessed for their phenotypic traits. The data were obtained in a trial planted in April 2018 with five commercial clones as controls. Vegetative development was evaluated based on plant height, stem diameter, and average height and diameter growth rates. Abiotic (drought) and biotic (rust, leaf miner, and mealybug) stresses were evaluated by a rating scale. The phenotypic variability of the population was detected for all traits and divergent clusters of families were identified. Negative correlations indicated that adverse conditions hampered the early development of the coffee plants. Groups of families should be investigated for breeding programs according to adequate development, even under stressful conditions. These results show the importance of preserving the germplasm of old crops in the south of Espírito Santo as a source of genetic resources for coffee breeding programs. This valuable resource should be maintained and conserved as a strategy to guarantee the sustainability of coffee growing and expand the genetic base, which has been reduced by the replacement of crops with clonal cultivars.

Keywords: Genetic breeding; genetic variability; Conilon coffee; multivariate analysis

1. Introduction

The genus *Coffea* has 124 cataloged species, but only *Coffea arabica* and *Coffea canephora* have global economic importance (Davis et al., 2011). Brazil is the largest producer and exporter and the second largest consumer of coffee in the world (CONAB, 2022). In 2021, the country produced 47.72 million bags of coffee (31.42 of *C.arabica* variety and 16.29 of *C. canephora*) (CONAB, 2022). The state of Espírito Santo produced 11.22 million processed bags of Conilon coffee in 2021 (CONAB, 2021).

The introduction of coffee cultivation in the state took place in the early 1900s, with Conilon material, but only in the last 40 years did the state begin to commercially produce coffee and clonal cultivars. Today, Espírito Santo has high phenotypic diversity of *C. canephora* in germplasm banks for traits related to the root system (Silva et al., 2020), fruit production (Partelli et al., 2021), flowers' morphologic traits (Silva et al., 2021a), stomatal characteristics (Dubberstein, et al., 2021), nutrient concentration in leaves (Silva et al., 2021b), and variables related to the plant, branches, leaves, growing cycle, fruits, seeds, response to pests, diseases and drought, and harvest and post-harvest factors (Ferrão et al., 2021). However, in the last 29 years, since the launch of the first clonal cultivars (Bragança et al., 1993), crops have been renewed and the genetic variation in seminal crops is being lost (Ferrão et al., 2007). The development and availability of new cultivars involves the selection of clones with high productivity and resistance to abiotic biotic stresses such as drought, as well as stable yields and uniform ripening (Ferrão et al., 2007). Thus, identifying and selecting materials which resist adverse climatic conditions and tolerate the main pests and diseases affecting the coffee plant are fundamental for sustainability of coffee growing.

Currently, crop renewal and the adoption of relatively few clonal cultivars in Espírito Santo and the neighboring state of Bahia (Espíndula and Partelli, 2011) have caused a loss of genetic diversity, a key factor for the development of new varieties via character selection. Thus, the old crops remaining in the south of the state need to be characterized and preserved, since they have useful variability with the matrix plants selected by the farmers themselves or introduced over the years (Ferrão et al., 2019). Currently, the characterization of these genetic resources is still incipient, but trees planted from 15 to even 80 years ago are still found in the south of the state, with high phenotypic variability and without the presence of commercial clones. These genetic reservoirs are highly relevant for the sustainability of the coffee crop (Ferrão et al., 2017),

both for productive aspects and for the selection of plants that are sources of tolerance to important stresses, such as drought, and pest attack and diseases.

Prolonged periods of drought and high temperatures, accentuated by climate change, damage the growth and sustainability of many crops (Fracasso et al., 2016). Drought is the main cause of reduced development, productivity, and quality in the Conilon coffee crop in Espírito Santo (Damatta et al., 2018; Semedo et al., 2018). Coffee rust, caused by the biotrophic fungus *Hemileia vastatrix*, is one of the main diseases of the crop, causing losses of up to 50% (Capucho et al., 2013; Zambolim, 2015), whereas leaf miner attack preferentially occurs in unshaded crops (Leite et al., 2020) in hot and dry climates (De Custódio et al., 2009), causing losses of up to 80% (Parra & Reis, 2013). Green mealybug infestation usually occurs in young plants and can cause losses of up to 60% (Fornazier et al., 2017; Espíndula Júnior et al., 2018).

Yields of rainfed crops can decrease by up to 80% in dry years (DaMatta et al., 2010; Venacio et al., 2020). Water stress affects growth and photosynthesis, and consequently coffee productivity, even in clones considered tolerant (Praxedes et al., 2006; Venancio et al., 2020). Coffee plants are more susceptible to water deficit during the propagation phases of extension (growth), inflorescence, and fruit development, which can interrupt the metabolism and reduce productivity (Naik et al., 2019). Pests and diseases also affect plant development, reducing coffee yields (Hindorf et al., 2011; Naik et al., 2019).

The study of morpho-agronomic characteristics using multivariate techniques can detect divergent and/or superior individuals to develop new and stable cultivars that have homogeneous development and better coffee productivity and quality (Dubberstein et al., 2021). Faced with the need to renew of crops with clonal cultivars and the use of few cultivars, characterizing, and preserving the remaining old crops is imperative for the conservation of the genetic diversity of the species. Thus, the study of these materials, in non-irrigated conditions, is a strategy for the discrimination of individuals with adequate initial vegetative development, based on traits highly correlated with yield and productivity, such as plant height and stem diameter (Martinez et al., 2007; Moncada et al., 2016).

This study was conducted to characterize and investigate the genetic diversity of 388 half-sib families of new and promising genotypes of *C. canephora* from old crops in

southern Espírito Santo, based on morpho-agronomic traits and the correlation between them, as well as to detect genotypes tolerant/resistant to biotic (rust, leaf miner, and mealybugs) and abiotic (drought) stresses for genetic breeding. This is a groundbreaking study of old seminal crops remaining in southern Espírito Santo. We hypothesized that the phenotypic variability present in these crops would make them good genetic resources for breeding program, so their characterization and preservation are fundamental, since the use of only a few clonal cultivars can lead to narrowing and loss of the genetic diversity in the state.

2. Material and Methods

Plant Material

The seeds of the 388 half-sib families of *C. canephora* used to implement the trial came from matrix plants selected in old seminal plantations in the south of the state of Espírito Santo, with 15 to 80 years of age. The sampling and collections of seeds were carried out between April and June 2017 in more than 20 properties in the municipalities of Alegre, Rive (district of Alegre), Cachoeiro de Itapemirim, Jacu (district of Cachoeiro de Itapemirim), Jerônimo Monteiro, and São José do Calçado (Table 1 and Table S1).

To ensure the representativeness of promising, unpublished genotypes and the genetic variability of the study population, old seminal crops were monitored for three years and selected according to the following criteria: composed only of plants from seeds with no commercial clones; no nearby crops with improved materials; and unavailable materials studied previously. Matrix plants were selected by the production ascertained by the farmers and by vegetative vigor, visually analyzed in the field. More than one plant was collected per property. Generally, crops had few cultural practices, only consisting of pruning and fertilization.

Table 1. Characteristics of the municipalities in which the seeds from matrix plants were collected to produce seedlings of the 388 families of half-sibs of *Coffea canephora*.

Origin	Climate	Precipitation	Temperature	Latitude	Longitude	Altitude
Alegre	Aw*	Annual precipitation	Average annual	S 20°45’48’	W 41°32’2’’	117 to 364 meters

		around 1,200 mm	temperature of 23°C			
Cachoeiro de Itapemirim	Aw*	Annual precipitation around 1,197 mm	Average annual temperature of 23,8°C	S 20º50’56’’	W 41º06’46’’	120 to 168 meters
Jerônimo Monteiro	Cwa*	Annual precipitation around 1,293 mm	Average annual temperature of 23.2°C	S 20°47’8’’	W 41°23’52’’	111 to 282 meters
São José do Calçado	Cwa*	Annual precipitation around 1,500 mm	Average annual temperature of 23.3°C	S 21°1’31’’	W 41° 39' 20"	302 to 410 meters

*according to Köppen classification

Experiment Establishment

The trial was implemented in the municipality of Mimoso do Sul (Latitude: 21° 03’01’’ South, Longitude: 41° 30’’44’ West, Altitude 620 meters); with climate classified as Cwb according to yjr Köppen classification; average temperature of 21 °C; annual rainfall of 1,375 mm and with a cool, dry winter and a rainy summer (Freitas et al., 2015). The seedlings were planted in April 2018 at nine months of age (270 days), in a Federer augmented block design (Figure 1). The spacing used was 2.5 m between rows and 1.0 m between plants. The planting of seedlings was carried out as follows: five per family; and for every five families, five plants of one commercial clone. A total of 388 families (2085 plants) were planted. The commercial clones used in the experiment were A1, P2, BRS, RO, and Verdim (Table 2). The crop treatments adopted were weed management and fertilization, according to soil analysis, without the use of irrigation.

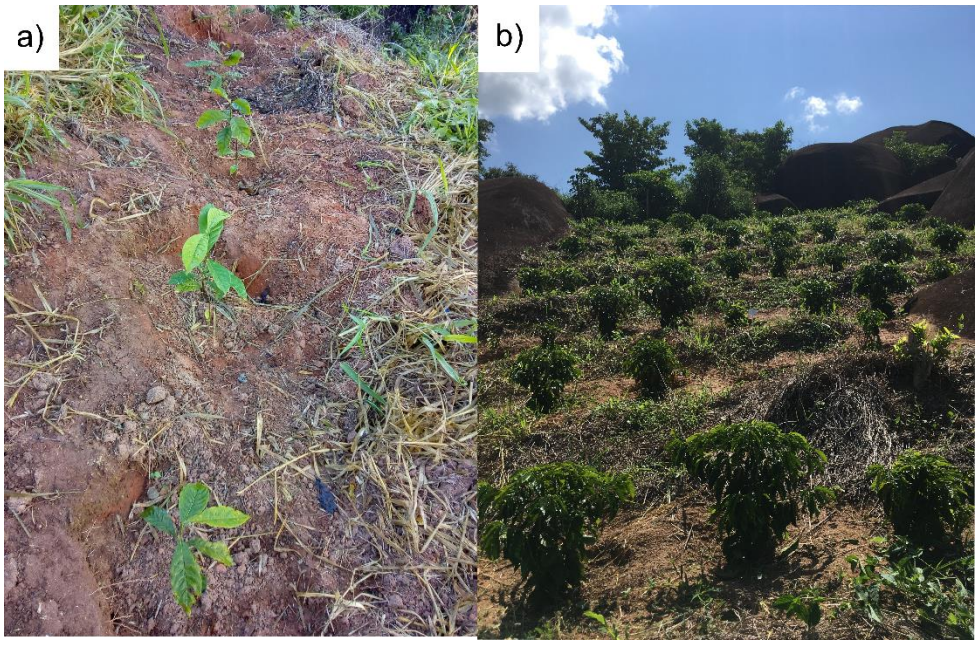

Figure 1. a) Plants with four months of installation in the field (August/2018) and b) plants with one year of installation in the field referring to the beginning of the initial development assessments (April/2019).

Table 2. Characteristics of commercial clones used as controls

					Characteristics				
Commercial cultivar	**Variety/ release year**	**No. of clones of the variety**	**Cultivar type**	**Origin**	**Water déficit***	**Rust***	**Leaf miner***	**Mealybug***	**Reference**
A1	Tributun/2017 Andina/2018	6	Clonal	Genotype initially propagated by Ivan Milanez and Hélio Dadalto, also known as H and H1	2	2	1	1	(Partelli et al. 2020; Partelli et al. 2019)
P2	Monte Pascoal/2020	6	Clonal	Genotype selected by producer Paulo Benacchi, in the municipality of Marilândia – ES	3	1	1	1	(Partelli et al. 2021)
Verdim	-	-	-	-	6	3	3	2	Uninformed
BRS	-	-	-	-	5	2	2	2	Uninformed
RO	-	-	-	-	7	2	1	1	Uninformed

*Average grades according to evaluations carried out in the field

Vegetative Development Assessment

Vegetative development was evaluated by measurements of plant height (HGT) and stem diameter (DIA) during 270 days since planting (Figure 2). Overall, six assessments were carried out for plant height with a ruler, measured in centimeters (cm) from the ground level to the pair of terminal leaves of the orthotropic branch. Stem diameter was measured with a digital caliper millimeters (mm) at five cm above ground level. These assessments were carried out bimonthly between April 2019 and February 2020 in April, June, August, October, December and February.

With the plant height and stem diameter data, the growth rates were obtained in cm.day-¹ and mm.day-¹, respectively, using the formula below, in which GR = growth rate, Lf = final length; Li = initial length; and nd = number of days.

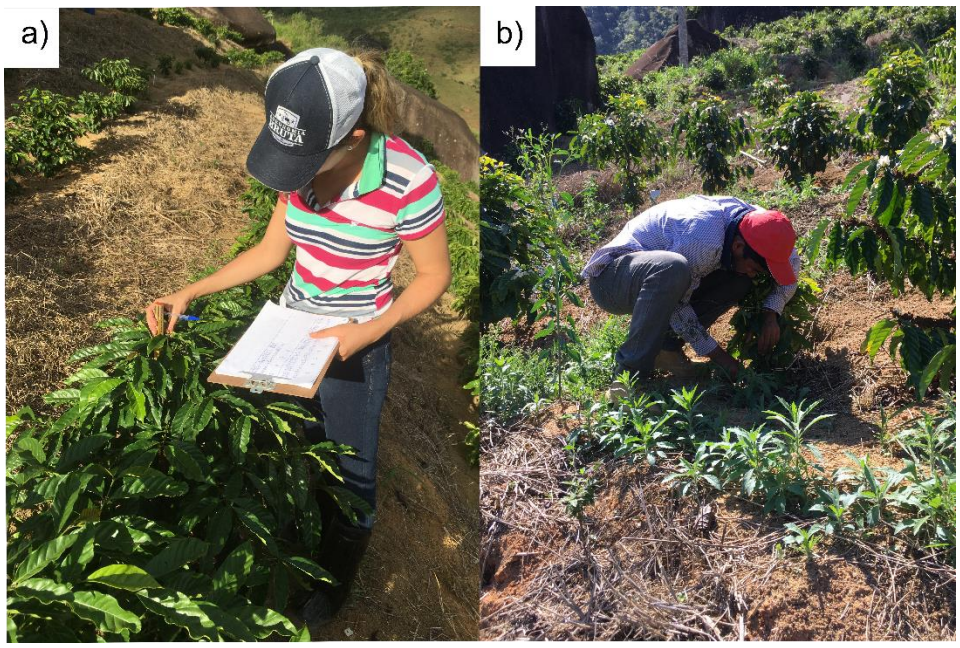

Figure 2. Analysis of plant height (cm) with a graduated ruler (a) and stem diameter (mm) with a digital caliper (b).

With the plant height and stem diameter data, the growth rates were obtained in cm.day-¹ and mm.day-¹, respectively, using the formula below, in which GR = growth rate, Lf = final length; Li = initial length; and nd = number of days.

$$GR = \frac{(Lf - Li)}{nd}$$

The data from the last measurement (February 2020) of plant height and stem diameter were used to plot two histograms using the R Studio software (RStudio Team, 2022) to evaluate the means of families in the population.

Analysis of Tolerance to Abiotic and Biotic Stresses

Drought tolerance was estimated from the visual categorical analysis of the plants using grades from 1 to 9 (Figure 3). This analysis was carried out in February 2019 in 200 families (totaling 1040 plants), in which 1 = vigorous plants without wilting symptoms; 2 = vigorous plants with slightly overhanging leaves; 3 = vigorous plants with some drooping leaves; 4 = dangling leaves; 5 = completely hanging leaves; 6 = completely hanging leaves with onset of discoloration, loss of leaf shine, and slight drying; 7 = completely hanging leaves with discoloration, loss of brightness in all leaves and moderate drying of the leaves; 8 = completely hanging leaves with discoloration and loss of brightness in all leaves, in addition to high drying intensity, with some brown color; and 9 = completely dry plant, showing permanent damage (adapted from Carvalho et al., 2017).

Figure 3. Coffee plants with their respective grades according to the symptoms present. Grade scale from 1 to 9, in which grade 1 = vigorous plants without wilting symptoms and grade 9 = completely dry plant with permanent damage.

The evaluation of biotic stresses was carried out through a graded visual analysis in December 2019, 600 days after planting, in which the rainy season (November-

December) favors the rust epidemic (Zambolim et al., 2015). For rust (*Hemileia vastatrix*) (Figure 4A), the scale ranged from 1 to 5, based on the severity of the disease, in which grade 1 = absence of spots or pustules or formation of spores; 2 = plants with lesions ranging from spots to chlorosis in the infected area but without the formation of urediniospores; grade 3 = low number of pustules per leaf with formation of urediniospores; grade 4 = medium number of pustules per leaf with formation of urediniospores; and grade 5 = high number of pustules per leaf with formation of urediniospores (adapted from Carvalho et al., 2017).

For the analysis of resistance to leaf miners (*Leucoptera coffeaella*) (Figure 4B), lesions were classified as small (0.3 to 0.6 centimeters in diameter); medium (about 0.6 to 1.2 cm in diameter); and large (above 1.2 cm in diameter), adopting the grading scale from 1 to 5, in which grade 1 = plants with less than 1% of leaves with small lesions; grade 2 = plants with 2% to 4% of leaves with lesions; grade 3 = plants with 5% to 19% of leaves with lesions (small, medium and large); grade 4 = plants with 20% to 35% of leaves with lesions (small, medium, and large); and grade 5 = plants with 36 to 100% of leaves with lesions (small, medium, and large) (Andreazi et al., 2015).

The evaluation of mealybug (*Coccus viridis*) (Figure 4C) resistance was also performed using a grading scale ranging from 1 to 5, in which grade 1 = absence of mealybugs; grade 2 = presence of a few individuals per plant; grade 3 = weak infestation (isolated females, colonies or nymphs); grade 4 = moderate infestation (presence of postures, 1st and 2nd instars on some branches); and grade 5 = strong infestation (presence of postures and all stages of development on the organs of all or almost all branches) (adapted from Andrade et al. 2017).

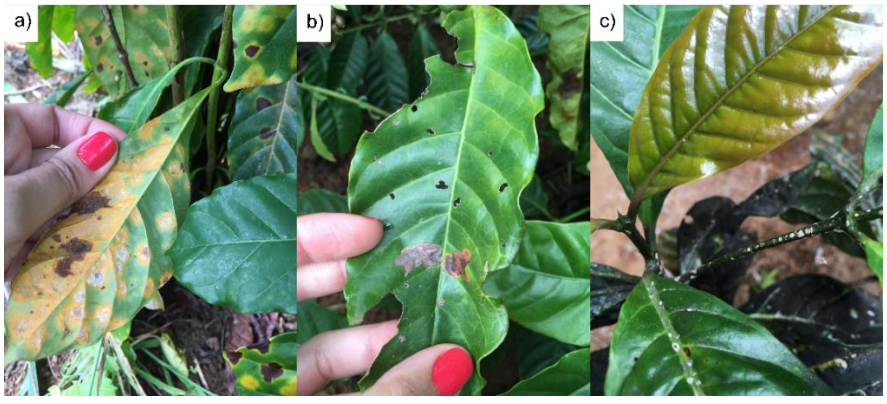

Figure 4. Biotic stresses assessed by visual rating scale. a) Leaf rust (*Hemileia vastatrix*); b) Leaf miner (*Leucoptera coffeaella*) and c) Mealybug.

Correlation Analysis

For the correlation analysis, data from the last measurements of height (cm) and stem diameter (mm) in February 2020 were used, along with the last two average growth rates, also for height and diameter, and the tolerance grades for drought, rust, leaf miners, and mealybugs. This analysis was performed with the R Studio software (RStudio Team, 2022).

Cluster Analysis

Two cluster analyses were performed. The first involved to data on growth rates of plant height (cm.day-¹) and stem diameter (mm.day-¹). Data were subjected to linear regression analysis, where X was the measurement period and Y was the average growth rate. The values of regression constants and regression coefficients were used to construct a dendrogram. These data were used to verify the grouping of families based on their behavior according to the values of $\beta 0$ and $\beta 1$ found when the growth rates in height and diameter over time were evaluated. After the formation of the clusters, the average growth rates of each cluster was estimated, and graphs of the average growth rates were constructed.

The second grouping was performed with data related to stresses and vegetative growth (height, diameter, average growth rate of height, average growth rate of diameter, drought tolerance, rust incidence, and infestation of leaf miners and insect scale). The data were used to build a stacked bar graph (barplot) with hierarchical grouping (dendrogram) to assess the influence of the stresses on the initial development of the coffee plants and genetic diversity by the clusters that were formed.

For both clusters, the standardized mean Euclidean distance (SMED) was used to calculate the distance between pairs of matrices. Cluster analysis was performed using Ward's method, which consists of a hierarchical clustering in which the similarity measure used for clustering is calculated as the sum of squares between two clusters over all variables (Hair et al., 2009). To determine the number of clusters that were formed, the Mojena statistical criterion (Mojena, 1977) was used in association with the visual analysis of the branches. Statistical analyses were performed using the R Studio software (RStudio Team, 2022).

3. Results

The population of half-sib families of *C. canephora* showed great variation of early developmental characteristics and biotic and abiotic stress resistance (Figure 5). Of the 2,085 plants in the trial (without irrigation) only 352 died (16.88%).

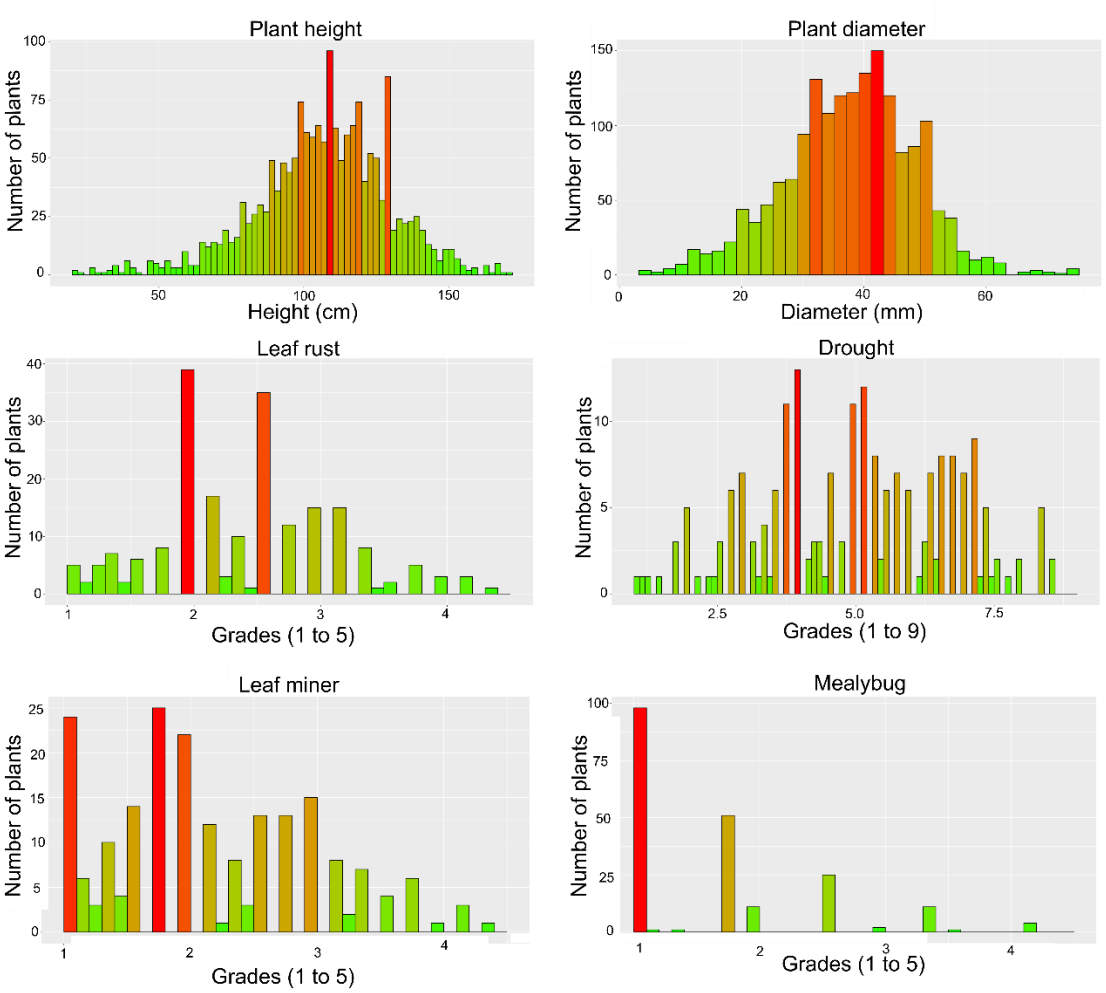

Figure 5. Histograms of mean distributions of plant height (cm), stem diameter (mm), rust incidence, drought tolerance, leaf miner incidence and mealybug incidence of the 388 half-sib families. The red colors represent a greater number of plants, and the green color represents a smaller number of plants.

Table 3 shows the families with the best averages for each trait. Related to the traits of initial development, we found that average height was 106.61 cm, with 32 families averaging above 130 cm. Average plant diameter was 37.67 mm, but some families had diameter greater than 50 mm, and therefore were considered better. The average height growth rate was 0.1713 cm.day-¹ and average diameter growth rate was 0.093 mm.day-¹.

For traits related to tolerance/resistance to diseases, pests and drought, we considered that families with lower average scores were better. For rust, we found families and four commercial clones with an average score of less than 1.5 points, while for leaf miner we found an average score of 2.17 and for cochineal the average score was 1.67 (Table 3). For water stress, some families were considered tolerant, with averages below 2.5 points, and some families were totally susceptible, with total death of plants within the family (Table 3).

Table 3. Amplitude of agronomic, growth, and biotic stress traits evaluated in half-sib families and five commercial clones of *Coffea canephora* and indication of the best varieties in terms of agronomic performance according to analysis of variance (ANOVA) in a simple augmented block design and classification of materials as to their phenological traits.

Variables	Minimum	Maximum	Mean	Families with the best averages for each characteristic
Early development				
Height (cm)	37.10	190.82	106.61	54, G69, 179.6, 162.4, 28.11, 149, 131, 180.7, G144, 179, G5, 162.5, 75, 164.7, 62, 168.2, 123, 179.2, 162.6, G57, 161.8, G62, 136.18, G68, G17, G106, G70, G110, 166.2, G66, G67, and 171.2 (> 130 cm)
Diameter (mm)	18.00	66.84	37.67	G66, 162.4, 168.2, 162.2, 162.7, 167, G162, 166.4, 162.5, G144, G67, 162.3, 162.6, 149, G106, 136.18, G146, G102, 28.11, G72, and 179 (> 50 mm)
Average growth rate in height (cm.dia-¹)	0.0045	0.4214	0.1713	166.2, 166, 146, 54, 144.2, G300, G69, G59, G95, 143.2, 136.8, and 179.2 (> 0.30 cm.dia-¹)
Average growth rate in diameter (mm.dia-¹)	0.0017	0.1918	0.093	G66, 162.4, 166.4, 167, 162.2, 162.7, G162, G144, G78, 149, G67, 162.5, 180.7, 162.3, 166, 146, 166.2, 176.6, G300, 34.2, G64 G106, 111, G105, G102, G69, G72, 179, and G155 (> 0.13 mm.dia-¹)
Biotic and abiotic stresses				
Leaf rust*	0.70	4.62	2.45	**A1**, 100, 61, 98, G22, G77, **Verdim**, 164.7, **BRS**, G79, 151.9, 164.8, and **RO** (< 1.5)

Leaf miner*	0.74	4.26	2.17	**Verdim**, **RO**, 136.3, 171.8, 187, 61, 62, 98, G22, G26, G47, 151.9, 164.1, 164.4, and 164.7 (< 1.1)
Mealybug*	0.35	4.50	1.67	151.5, 151.9, 161.9, 162.1, 162.9, 163.2, 164, 164.1, 164.3, 164.4, 164.5, 164.6, 164.7, 164.8, 165, 170.4, 173.3, 28.11, 28.12, G36, G39, G400, G44, G46, G47, G48, G52, G56, G57, **Verdim,** and **RO** (<0.8)
Drought*	1.05	9.14	4.95	G73, 183.4, 162.1, 28.11, **P2**, 95.1, G22, G57, 187, G20, **A1**, 136.5, 94, 164.1, and G5 (< 2.5)

* 200 half-sib families were evaluated for these characteristics.

We observed the behavior pattern of the families for the vegetative development traits by examining eight clusters obtained with the data of the regression constant and coefficients (Figure 6 and Table 4) and stratify them by the monthly graphs of each cluster formed from the average plant height and stem diameter growth rates (Figure 7A and 7B). In this cluster, three groups represented 76.3% of the families (groups 1, 2, and 3) with 296 half-sib families and four commercial clones (A1, Verdim, RO, and BRS). These clusters showed similar behavior according to average height and diameter growth rates (Figure 7A and 7B), but their growth rates were lower than those in groups 5 (families G107, G95, 144, 143.2, 165.4, and clone P2), 6 (families 146, G300, 144.2, and 166.2) and 7 (families 166, G59, and 143.3), which had the highest initial (0.72, 0.66, and 0.49 cm.day-¹) and final (0.30, 0.25, 0.23 cm.day-¹) height, respectively. For the average diameter growth rate, the families in group 5 had the greatest final growth (0.11 mm.day-¹) and groups 5, 6, and 7 showed the greatest initial diameter growth (0.22 mm.day-¹).

The families in group 4 attained the greatest average height (121.5 cm). The mean diameter of this group was 38.70 mm. Group 6 had the second highest mean height (120.75 cm), followed by group 2, with 114.47 cm. The mean diameters of these two groups were 41.18 and 38.90 mm, respectively.

From the graphs constructed with the height (cm.day-¹) and diameter (mm.day-¹) growth rates (Figure 7A and 7B), we found that the highest growth rate occurred in the first month of evaluation with individuals planted 365 beforehand. We also observed later stability in their vegetative development. Group 5 (families G107, G95, 144, 143.2, 165.4, and clone P2) showed the highest growth rates for both traits, with a decrease over the

measurement periods, followed by groups 6 (146, G300, 144.2, and 166.2) and 7 (166, G59, and 143.3).

Groups 1, 2, 3, 4 and 8 showed similar behavior in terms of vegetative development during the evaluation period (Figure 7). However, group 4 attained the highest growth rates, followed by groups 2, 1, 3 and 8. The commercial clones used as controls were in groups 1 (Verdim and RO), 3 (A1), 5 (P2) and 4 (BR).

Group 8 had only one family, 168.10 (Figure 6 and Figure 7). This family showed the lowest average growth rates in height and diameter in all measurement periods. The average height growth rate was 0.065 cm.day-¹ and the average diameter growth rate r was 0.062 mm.day-¹. The average height of this family were 94.75 cm and 26.67 mm, respectively.

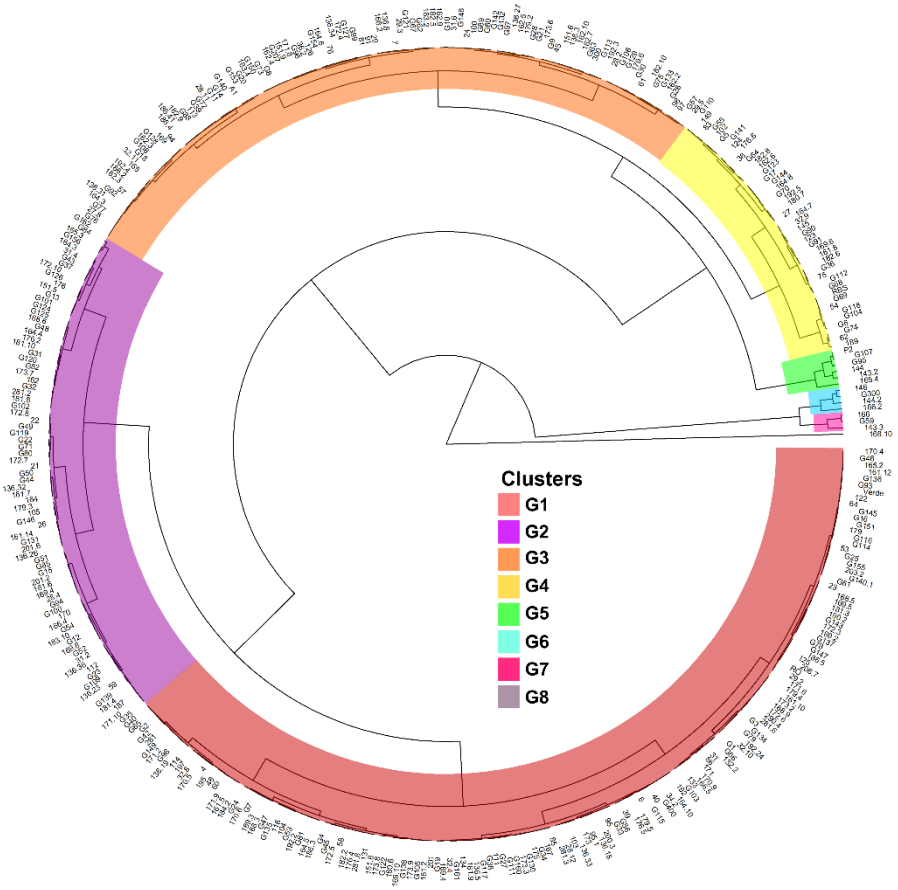

Figure 6. Clustering of 388 half-sib families of *Coffea canephora* and five commercial clones (red) from regression constant data and coefficients for height and diameter, using dissimilarity measures based on the standardized mean Euclidean distance (SMED) and Ward's method. Colors represent the nine groups formed and the cutoff point was determined by the Mojena test (1977).

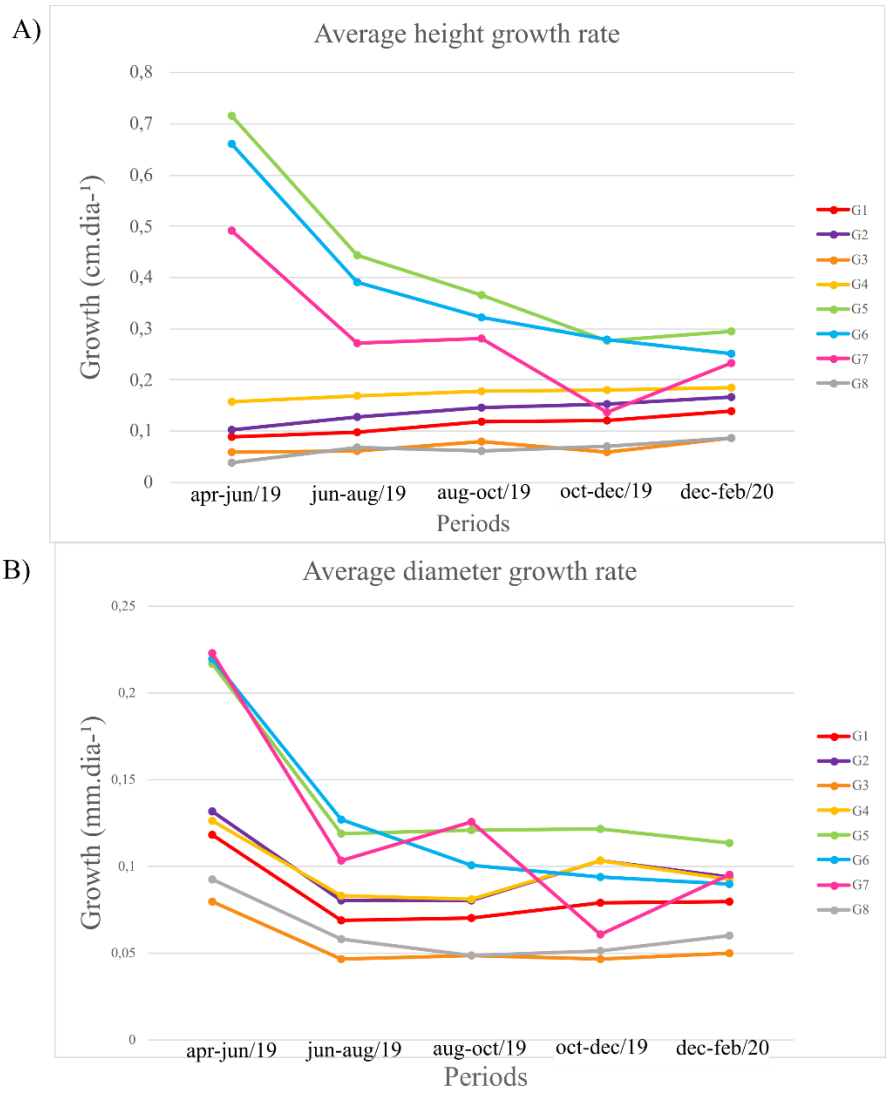

Figure 7. Average growth rates in height (cm.day-¹) (A) and diameter (mm.day-¹) (B) of the eight groups formed from the cluster analysis. Growth rates of the groups and the colors represent the same groups formed in the previous grouping (Figure 6)

Table 4. Groups formed by cluster analysis with the families of each group.

Cluster	Number of families	Families
1	151 and 2 commercial clones	136.5, G1, G2, G3, G4, 134, 133, 32.10, 32.8, 32.4, G7, 95, 95.1, G19, 132.2, 13, G16, 190.2, 189.2, 189.4, 189.3, 4, G27, 6, 58, 60, G25, G24, 168.5, 64, 168.9, 182.2, 168.3, 171.9, 171.6, 103, 104, 173.8, 173.9, 175, 176.4, 176.5, 161.2, G33, 161.5, G34, 169.5, G400, 173, 173.2, 173.3, G38, G40, 151.8, G42, G45, 181.3, 161.10, 161.9, 161.12, G46, G47, 169.10, 170.4, G53, 170.5, 201, G56, 172.5, 172.3, 172.6, 164.5, 164.10, 165.2, 28.12, G81, G79, 171.2, 171, 180.6, G61, 125, 179.5, 179.4, 179, 167, G65, 166.5, G66, 170.6, 170.9, G78, 136.18, 136.33, G98, G103, G93, 203.2, 206.7, G111, 197, 122, 184.2, G122, 281, 29.2, G105, 23, 281.3, 281.6, 281.8, 290.4, G114, 200.3, G115, G116, G117, 53, 186.3, 186.5, 192.24, 136.19, G130, 192, 192.2, G134, G136, 195, 85, 86, G137, G138, 39, 40, G140.1, 46, G135, 111, G143, G149, G147, 116, G145, 114, G151, 31, G155, G160, 34.2, G161, **RO** and **Verdim**
2	104 and 1 commercial clone	136.3, G8, G9, 97, G10, 94, G20, 91, G18, G14, G11, G30, 57, 61, G26, G28, 7, 171.8, 98, 100, G21, 169, 162.9, 162.10, 163.2, 164.3, G39, 173.6, 151.9, 151.6, 172.4, 164.6, G57, 165, 28.11, 126, G63, G62, G60, 179.6, 179.2, 168.2, 162.2, 162.3, 162.4, 162.5, 162.7, G67, G77, G76 G75, 136.8, G73, G72, 136.34, 183.4, 29, G97, G99, G88, G89, G106, G92, G96, G207, G113, G108, 300, G121, 76, 24, 28.2, 186.41, 186.4, 186.2, G127, G128, G129, 136.27, G132, G133, 136.31, 192.3, 81, 192.4, G142, G140, 45, 113, 182.5, 182.3, G148, 182.9, 182.10, 31.6, 29.5, G150, 29.3, G153, G154, 27.8, 32.11, 36.2, G162 and **A1**
3	79	G13, G12, G15, 26, 187, 59, G23, 168.6, G22, 171.10, G31, 105, 176, 176.2, G32, 161.6, 161.7, G35, 169.4, G37, 169.8, 164, 173.7, 151.5, G43, G44, 181.10, 181.8, 181.4, G48, G49, 161.14, 162, G50, 170, G51, G52, 201.2, 201.6, G54, 172.7, 172.10, 172.8, 164.4, 165.3, 28.4, G85, G100, G101, G84, G82, 136.36, G80, 179.3, 166.4, G71, 136.32, G94, G109, G119, G120, G124, G126, 183.10, 184.3, G125, 281.2, G102, 21, 22, G131, 136.23, 136.26, G139, 112, G146, 31.2, 31.4 and G156
4	42 and 1 commercial clone	G5, 32.9, G6, G17, 189, 62, 102, 161.8, 169.6, G36, G55, 164.7, 164.8, 180.7, 176.6, 149, G64, 162.6, G68, G69, G70, G74, G104, G90, G91, G112, G110, 123, 75, G123, 124, 54, G118, 192.5, 83, G141, G144, 182.6, 145, 182.8, 27, 36 and **BRS**
5	5 and 1 commercial clone	165.4, 143.2, 144, G95, G107 and **P2**
6	4	G300, 166.2, 144.2 and 146
7	3	G59, 166 and 143.3
8	1	168.10

Correlation analysis of the vegetative development variables with the drought and biotic stresses variables showed that the stresses, in general, negatively affected the initial growth of the plants in the families (Figure 8). Negative and significant correlations

between drought stress and final height (-0.78), mean height growth rate (-0.66), stem diameter (-0.72), and mean diameter growth rate (-0.67) indicated that drought negatively affected plant development. We found a high and positive correlation between plant height and stem diameter (0.90).

It was possible to verify a high correlation between the severity of rust and the incidence of mealybugs (0.81), as well as between mealybugs and leaf miners (0.79). There was a significant positive correlation of 0.55 between drought and coffee rust.

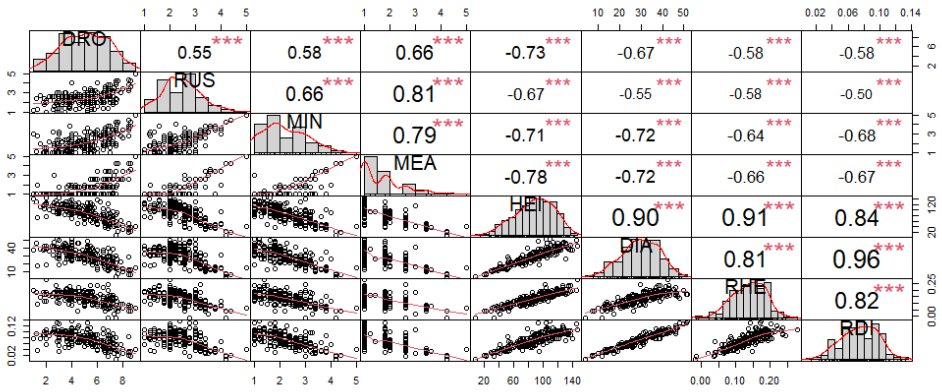

Figure 8. Pearson's correlations for vegetative development data (final height, height growth rate, final diameter, and diameter growth rate) and abiotic (drought) and biotic (rust, leaf miners, and mealybugs) stresses for 200 half-sib families of *Coffea canephora*. The red color refers to significant positive correlations and the green color refers to significant negative correlations. DRO = drought; RUS = leaf rust; MIN = leaf miner; HGT = plant height; DIA = diameter; AHG = average height growth rate; ADG = average diameter growth rate.

In the second cluster analysis of half-sib families, we considered all development data and biotic and abiotic factors. The nine groups formed showed the variability of these families for the traits evaluated (Figure 8 and Table 5).

The families in group 8 had the best vegetative development, with more vigorous and taller plants with larger stem diameters (Figure 8). For mealybugs, most families in this group showed no infestation, the exceptions being families 136.3, G8, 167, 162.4, and G66. This group showed incidences of rust, mealybug infestation, and stress caused by drought (at different levels). However, when compared to other groups, the scores for such stresses were lower, showing that these families had potential vegetative

development even when affected by biotic and abiotic stresses. Commercial clone A1 was in this group.

Group 9, formed by families 103, 168.5, G79, G14, 91, 151.9, 171.8, G26, 98, 61, 164.8, 102, 100, 164.7, 180.7, G6, and clone P2, also had more vigorous plants (taller and with larger stem diameters), no mealybug infestation, low rust severity (most families with a score below 2), and low leaf miner infestation. Drought stress occurred variably in these families. Some families did not show symptoms of water deficit, while others showed symptoms of wilting, yellowing, chlorosis and leaf fall.

In group 1, the clones BRS and RO had adequate initial growth but were affected by all stresses. The Verdim clone was in group 4, in which the plants obtained higher scores for all stresses and the initial development of the coffee plants was more compromised than in groups 8 and 9.

Families 28.11 and G69 had the highest average heights, 144.60 and 145.60 cm, respectively, and average stem expansions of 50 and 42.60 mm, respectively, showing the greatest initial development and a plant architecture favorable to crop management. As for stresses, these families showed some tolerance to drought, leaf miners, and mealybug, but had a rust severity score of 2.

For breeding we can indicate families with important traits, such as family 94 (group 8), which had plant height of 130 cm, stem diameter of 45.82 mm, resistance to scale, and some tolerance to leaf miners and drought. However, it had an average rust severity score of 2.4. The G5 family (3) was tolerant to water stress and mealybugs, showing good tolerance to rust, height of 139.4 cm, and stem diameter of 36.63 mm. However, it was more sensitive to leaf miner infestation, with a score of 3.

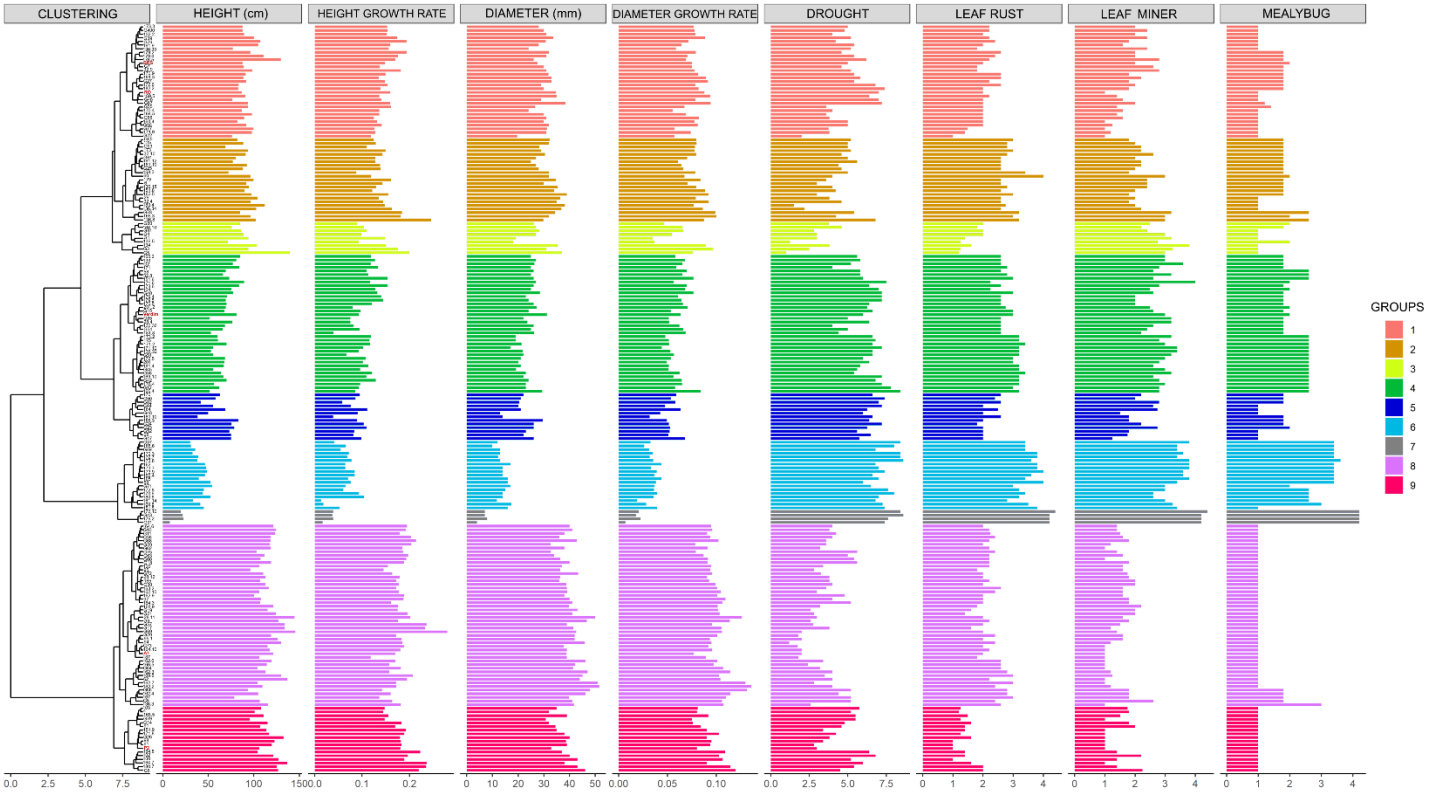

Figure 8. Grouping generated by the Ward method from measurements of the standardized mean Euclidean distance (SMED) and stacked bar graph of 200 families of half-sibs of *Coffea canephora* for the variables drought tolerance, rust severity, leaf miners, mealybugs, final height, height growth rate, final diameter, and diameter growth rate. The colors represent the ten groups formed, and commercial clones are indicated in the grouping by a black dot. The lines of values on the abscissa axis correspond to: values in relation to the dissimilarity in the last fusion level (dendrogram) and values in relation to the evaluated phenotypic traits (barplot).

Table 5. Groups formed by cluster analysis with the families of each group.

Cluster	Number of families	Families
1	29 and 2 commercial clones	173.3, G400, 132.2, G39, G76, 161.9, 136.33, 179.2, 179.6, 176.5, **BRS**, G7, 32.9, 170.9, 181.3, G27, 172.6, 182.2, **RO**, 169.5, G46, G67, G55, 170.4, 164.5, G56, 164.4, G16, G77, 171.9 and G22
2	23	G65, 125, G53, 173, 32.10, G61, 161.12, 161.10, G25, 168.2, 60, 179, 6, 136.18, 151.8, 162.5, 29, 32.4, 183.4, 136.34, G78, 165.3 and 136.8
3	9	G80, 168.10, G81, G4, G1, 136.5, 134, G3 and G5
4	36 and 1 commercial clone	165.2, 133, G2, 171, 58, 32.8, 151.6, 171.6, 173.6, 104, G40, 179.4, 170.5, 201.2, G15, **Verdim**, G85, 28.4, 136.36, G54, 181.8, 173.2, 175, 171.2, 171.10, 136.32, G51, 172.5, 201, 181.4, G45, G38, 169.10, G42, 176.4, G12 and 166.4
5	13	170, G50, G82, G44, 164, G48, 181.10, 168.3, G24, G23, G52, 26 and G47
6	19	G37, 168.6, G49, 179.3, 172.7, 173.8, 162, 173.7, 173.9, 169.4, 105, 59, G71, 172.8, 170.6, 180.6, 161.14, 169.8 and 151.5
7	4	172.10, G43, 176.2 and G31
8	49 and 1 commercial clone	164.6, G10, G11, G36, G68, G70, G62, G75, G63, G60, G18, 126, 95, G72, 28.12, 165, G30, 163.2, 162.10, 162.9, 97, 164.3, 168.9, G19, G21, 28.11, G9, G57, G17, G69, G20, 95.1, 94, G73, 164.10, **A1**, 187, 162.6, 166.5, G64, 162.3, 169.6, 62, 162.7, 162.2, G66, 162.4, 167, G8 and 136.3
9	17 and 1 commercial clone	103, 4, 168.5, G79, G14, 91, 151.9, 171.8, G26, 98, 61, **P2**, 164.8, 102, 100, 164.7, 180.7 and G6

The families in group 9 were good genetic resources for tolerance to the evaluated pests and diseases, and also showed greater vegetative development, along with group 8, and tolerance to infestation by mealybugs. Some families did not show infestation of leaf miners and leaf rust incidence (61, 98, 100, and clone P2), but showed low tolerance to water deficit, and were thus possible sources of resistance genes to the evaluated biotic stresses.

We noted examples where stresses affected the initial development of plants in the families of group 7 (G31, 176.2, G43, and 172.10), which obtained higher scores for all stresses. Their initial development was compromised since they showed lower height, growth, and diameter values than the other groups. Also, the families of group 6 (151.5, 169.8, 161.14, 180.6, 170.6, 172.8, G71, 59, 105, 169.4, 173.9, 173.7, 162, 173.8, 172.7,

179.3, G49, 168.6, and G37) were also more affected by stresses and had less development.

Biotic stresses showed wide variability depending on the group. We found groups of families with the occurrence of pests and diseases, as was the case of groups 8 and 9, both of which showed some type of tolerance to stresses, making them possible sources of genes for resistance. However, there were groups that suffered from stress and failed to properly develop, thus being considered sensitive to pest attacks, rust and drought.

4. Discussion

Half-sib families of *Coffea canephora* showed wide phenotypic variability for the traits of early development and stress resistance. This variability can be explained by the natural allogamy of the species due to gametophytic self-incompatibility and because they came from seeds (Conagin & Mendes, 1961; Berthaud, 1980; Partelli et al., 2020). The variability found in this germplasm is promising for genetic breeding (Rodrigues et al., 2013; Carias et al., 2016; Dubberstein et al., 2020).

Regarding the high positive correlation between height and stem diameter (0.88), Dubberstein et al. (2020) and Avellán et al. (2015) also detected positive and significant correlations for these same characteristics, 0.58 and 0.51, respectively.

We focused on studying the development characteristics in the first year of planting the *C. canephora* crop with the ES germplasm. We identified materials with aptitude for initial development in conditions of absence of irrigation, demonstrating the rusticity present in these materials under stressful conditions in the field. Also, since few crop treatments were used, the detection and selection of these materials is important in the initial phase of development for crops with few technological resourced. This aptitude was expected and was confirmed, given the hypothesis used in this work involving plant matrices in seminal crops. Other studies in the state have shown the wide phenotypic diversity of germplasm collections in the state, evaluating only adult plants with reports of the root system (Silva et al. 2020), fruit production (Partelli et al., 2021), flowering (Silva et al., 2021a), stomatal characteristics (Dubberstein et al., 2021), nutrient concentration in leaves (Silva et al., 2021b) and traits related to the plants, branches, leaves, growing cycle, fruits, seeds, response to pests, diseases, drought, and harvest and post-harvest factors (Ferrão et al., 2021).

We submitted the data referring to the last measurement of plant height (cm) and stem diameter (mm) to linear regression analysis for the construction of the first cluster shown. This statistical method was used to obtain the regression coefficients to verify the grouping of families based on similar behavior according to analyzed $\beta 0$ and $\beta 1$ values due to height and diameter growth rates. From the growth rates, we observed only three groups (5, 6, and 7) with higher vegetative development rates in the first measurement period, with a reduction in growth rates throughout the measurements. The other groups did not show large rate fluctuations during all periods.

We found wide phenotypic variation of drought tolerance, with plants that did not suffer water deficit and plants that showed permanent damage. Visual analysis for drought tolerance has been successfully used for cereals and coffee, proving to be efficient in discriminating genotypes tolerant to water deficit, an important tool for preliminary selection of resistant cultivars (Golabadi et al., 2006). The selection of superior genotypes tolerant to adverse environmental conditions is critical to produce *C. canephora* (Pezzopane et al., 2010; Covre et al., 2018; Thioune et al., 2020; Silva et al., 2020). Although the plants were grown in rainfed conditions, only 16.88% of the plants suffered permanent damage due to drought. Some families showed low scores for this stress, indicating potential for selection of drought-tolerant genotypes, as was the case of families G73, 183.4, 162.1, 28.11, 95.1, G22, G57, 187, G20, 136.5, 94, 164.1, G5 and commercial clones A1 and P2 (< 2.5).

There was an atypical period of drought in January 2019, which may have aggravated the stressful condition at this stage of implementation of the crop. This loss can be due to the water deficit in the soil, which can influence the growth of the plants during their initial development, leading to permanent leaf damage, such as desiccation of young leaves or fully expanded leaves (Araújo et al., 2011; Oliveira et al., 2012; Pizetta et al., 2012; Carvalho et al., 2017).

Drought is the most harmful abiotic stress to coffee production. Periods of extreme drought and super-optimal temperatures (van der Vossen et al., 2015; Rodrigues et al., 2016) are considered the main climatic impediments to successful coffee cultivation (DaMatta & Ramalho, 2006; Ramalho et al., 2014). Covre et al. (2016), in a study evaluating the development of Conilon coffee under irrigated and non-irrigated conditions, reported that irrigated plants showed a total growth of orthotropic branches

31.4% higher than non-irrigated plants. We also found families (Groups 6 and 7) highly affected by drought and biotic field stresses.

Plants considered to be drought tolerant remained vigorous and with turgid leaves, whereas susceptible ones showed yellowing, wilting and leaf drop, with some plants suffering permanent damage. Other studies have shown that water limitations cause a reduction in the turgor of the leaf cells, resulting in reduction in the photosynthetically active area and leaf chlorosis, which can lead to leaf fall, and branch death, affecting growth and productivity (DaMatta & Ramalho, 2006; Damatta et al., 2018).

In the studied population, the qualitative visual analysis for drought tolerance showed that the plants which obtained grades above 5 in the grading scale used (1 to 9) had symptoms of wilting, chlorosis, drying, and dropping of leaves. The 350 plants in the field with a score of 9 all died. Similarly, Carvalho et al. (2017) detected coffee genotypes (Arabica and Conilon) susceptible and tolerant to water deficit through a grading scale.

The groups which obtained higher scores for biotic stresses showed a clear reduction in plant growth. Both leaf miners and rust cause the coffee plant to drop its leaves, with a consequent reduction in photosynthesis, impairing initial development and final yield and quality (Esgario et al., 2020). The factors that favor the occurrence of leaf miner infestation are present mainly in the summer, especially January and February, but a hot and dry climate can provide favorable conditions for the insect throughout the year (Fornazier et al., 2017). We found this condition of drought and high temperatures during the evaluation, which may have influenced the increase in the incidence of this pest.

The biotic stresses (leaf rust, leaf miner and mealybug) are also responsible for the decline in productivity and reduction of the initial growth of the coffee plants, since they are related to foliar disturbances, resulting in reduction of the photosynthetically active area. However, in this study, we observed that some families showed low scores for such stresses, so they can be considered potential families to select for resistance.

We also detected half-sib families that were able to develop well even when affected by disease, pests or water stress, showing better growth averages This was the case of families 164.7, 28.11, G5, 62, and G57 (Table 3). It is also interesting to note that for the development traits (height and diameter), the families showed better averages than the commercial clones. Therefore, they are genetic resources for greater growth and development of coffee plants.

For rust disease, nine families had score below 1.5, while for leaf miners, 13 families had score below 1.1, and for mealybugs, 29 families scored below 0.8. These families can also be considered good genetic resources for resistance to coffee pests and diseases. Regarding drought tolerance, we found 13 families which scored less than 2.5, representing potential resistance to this stress in rainfed conditions.

We detected genotypes that showed adequate vegetative development even under biotic stress and rainfed conditions. We also identified tolerant/resistant genotypes regarding the evaluated stresses. Such results are important for genetic breeding programs, to enable selecting materials resistant to adverse weather and stress conditions. The identification of these materials is also relevant for producers, whose crops suffer from these stresses in the initial development, and for crops grown with few technological resources.

Tolerant genotypes can facilitate crop management by reducing the damage caused by water deficit, incidence of rust, leaf miners, and mealybugs. Besides this, materials that proved to be drought tolerant should be selected to improve resistance in a scenario of climate changes and the adverse conditions changes entail. The use of tolerant genotypes is a viable alternative to assure sustainability of coffee growing. On the other hand, half-sib families were also found that suffered greatly from adverse weather conditions, due to lower resistance to pest infestation and the incidence of rust.

The state of Espírito Santo is globally important for the production and improvement of coffee, specifically the Conilon variety (Ferrão et al. 2019), since it constitutes a genetic reservoir of the species that still remains in old seminal crops in the south of the state. In this study, materials from seminal crops show aptitude for initial development in conditions of no irrigation, demonstrating the rusticity present in these materials under stressful conditions in the field. This aptitude was expected and confirmed, given the hypothesis of collecting matrix plants from seminal crops.

5. Conclusions

The population studied harbors divergent genotypes related to characteristics of vegetative development and tolerance to abiotic and biotic stresses. The phenotypic variability detected can serve as a genetic resource for coffee genetic breeding programs.

The high divergence of the groups was demonstrated in the morphological analyses with specific groups demonstrating substantial differences in the growth rates and in the cluster analyses. In comparison with commercial clones, there was germplasm differential with potential for genotype selection, verified in clusters with examples for possible selection of genotypes with high growth rate and tolerance to the stresses present in the field. The divergent groups detected can also be promising for crosses to improve Conilon coffee in the state of Espírito Santo.

The characterization and preservation of these genotypes is fundamental for the maintenance of the culture, since the use of only a few clonal cultivars can lead to narrowing and eventual loss of the species' genetic diversity.

Moreover, given the climate changes and adverse conditions these changes will likely cause, the selection and use of more tolerant genotypes are important to assure good crop development and yield.

The materials from seminal crops showed aptitude for initial development in conditions without irrigation, demonstrating the rusticity present in these materials under stressful conditions in the field.

6. Acknowledgment

The authors would like to thank Conselho Nacional de Desenvolvimento Científico e Tecnológico [National Council for Scientific and Technological Development] (CNPq, Brasília – DF, Brazil, grant number 311950/2016-7), Fundação de Amparo à Pesquisa e Inovação do Espírito Santo [Research Support Foundation of Espírito Santo] (FAPES, Vitória – ES, Brazil), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior [Coordination for the Improvement of Higher Education Personnel] (CAPES, Brasília, DF, Brazil) – Finance Code 001, for financial support, CAFESUL - Cooperativa dos Cafeicultores do Sul do Estado do Espírito Santo, as well as the producers who allowed the collection of seeds in their crops and all who helped to implement the experiment.

7. References

Andrade A, Borges da Silva E, Pereira AP, Franco JC (2017) Identificação e monitorização de cochonilhas-algodão (Hemiptera, Pseudococcidae) associadas ao

cafeeiro em estufa. Revista Ciências Agrárias, 40:616–629. https://doi.org/10.19084/rca16144

Andreazi E, Sera GH, Faria RTD, Sera T, Shigueoka LH, Carvalho FG ..., Chamlet, D (2015) Desempenho de híbridos F1 de café arábica com resistência simultânea a ferrugem, mancha aureolada e bicho mineiro. Coffee Science, 10:375–382.

Avellán LFP, Loor Solórzano RG, Guerrero Castillo HE, Duicela Guambi L (2015) Caracterización fenotípica del germoplasma de *Coffea canephora* Pierre base para su mejoramiento en Ecuador. Revista Espamciencia,Ecuador, 7(1): 23–35.

Berthaud J (1980) L'incompatibilité chez *Coffea canephora*: méthode de test et déterminisme génétique. Café Cacao Thé, 24:267–274.

Bragança SM, Fonseca AD, Silveira J, Ferrão RG, Carvalho C (1993) "EMCAPA 8111"," EMCAPA 8121"," EMCAPA 8131": primeiras variedades clonais de café conilon lancadas para o Espírito Santo. EMCAPA, Vitória, Brazil

Capucho AS, Zambolim L, Cabral PGC, Maciel-Zambolim E, Caixeta ET (2013) Climate favourability to leaf rust in Conilon coffee. Australasian Plant Pathology, 42(5): 511-514.

Carias CDO, Gravina, GDA, Ferrão MAG, da Fonseca AFA, Ferrão RG, Vivas M, Viana AP (2016) Prediction of genetic gains by mixed models in conilon coffee progenies. Coffee Science, 11(1):39-45.

Carvalho FG, Sera GH, Andreazi E, Sera T, Fonseca IDB, Carducci FC ..., Costa KC (2017) Drought tolerance in seedlings of coffee genotypes carrying genes of different species. Coffee Science, 12(2):156-163.

CONAB – Companhia Nacional de Abastecimento (2021) Acompanhamento da safra brasileira de café: quarto levantamento – Safra 2021. https://www.conab.gov.br/info-agro/safras/cafe/boletim-da-safra-de-cafe?limitstart=0

CONAB – Companhia Nacional de Abastecimento (2022) Acompanhamento da safra brasileira de café: quarto levantamento – Safra 2022. https://www.conab.gov.br/info-agro/safras/cafe/boletim-da-safra-de-cafe?limitstart=0

Conagin CH, Mendes AJT (1961) Pesquisas citológicas e genéticas em três espécies de *Coffea:* auto-incompatibilidade em *Coffea canephora* Pierre ex Froehner. Bragantia 20:788–804. https://doi.org/10.1590/S0006-87051961000100034

Covre AM, Partelli FL, Bonomo R, Braun H, Ronchi CP (2016) Vegetative growth of Conilon coffee plants under two water conditions in the Atlantic region of Bahia State, Brazil. Acta Scientiarum. Agronomy, 38(4):535-545. https://doi.org/10.4025/actasciagron.v38i4.30627

Covre AM, Partelli FL, Bonomo R, Tomaz MA, Ramalho JC (2018) Impacts of water availability on macronutrients in fruit and leaves of conilon coffee. Pesquisa Agropecuária Brasileira, 53:1025-1037. https://doi.org/10.1590/S0100-204X2018000900006

Custódio AADP, Moraes JC, Custódio AADP, Lima LA, Faria MAD, Gomes NM (2009) Incidência do bicho-mineiro do cafeeiro em lavoura irrigada sob pivô central. Coffee Science, 4(1):16-26.

DaMatta FM, Avila RT, Cardoso AA, Martins SC, Ramalho JC (2018) Physiological and agronomic performance of the coffee crop in the context of climate change and global warming: A review. Journal of Agricultural and Food Chemistry, 66(21): 5264-5274. https://doi.org/10.1021/acs.jafc.7b04537

DaMatta FM, Ramalho JDC (2006) Impacts of drought and temperature stress on coffee physiology and production: a review. Brazilian journal of plant physiology, 18:55-81. https://doi.org/10.1590/S1677-04202006000100006

DaMatta FM, Ronchi CP, Maestri M, Barros RS (2010) Café: Ambiente e fisiologia da cultura. In F. M. DaMatta (Eds.), Ecofisiologia de Culturas de Árvores Tropicais (pp. 181–216). Nova Science Publishers, Nova York.

Davis AP, Tosh J, Ruch N, Fay MF (2011) Growing coffee: Psilanthus (Rubiaceae) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and evolutionary history of *Coffea*. Botanical Journal of the Linnean Society, 167(4):357-377. https://doi.org/10.1111/j.1095-8339.2011.01177.x

Dubberstein D, Oliveira MG, Aoyama EM, Guilhen JH, Ferreira A, Marques I, ..., Partelli FL (2021) Diversity of Leaf Stomatal Traits among *Coffea canephora* Pierre ex A. Froehner Genotypes. Agronomy, 11(6):1126. https://doi.org/10.3390/agronomy11061126

Dubberstein D, Partelli FL, Guilhen JHS, Rodrigue WP, Ramalho JC, Ribeiro-Barros A I (2020) Biometric traits as a tool for the identification and breeding of *Coffea canephora* genotypes. Genetics and Molecular Research, 19(2). https://doi.org/10.4238/gmr18541

EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária. (2009). Centro Nacional de Pesquisa de Solos. Centro Nacional de Pesquisas de Solos, Rio de Janeiro, 412p

Esgario JG, Krohling RA, Ventura JA (2020) Deep learning for classification and severity estimation of coffee leaf biotic stress. Computers and Electronics in Agriculture, 169:105-162. https://doi.org/10.1016/j.compag.2019.105162

Ferrão MAG, Mendonça RFD, Fonseca AFA, Ferrão RG, Senra JFB, Volpi PS, ..., Comério M (2021) Characterization and genetic diversity of *Coffea canephora* accessions in a germplasm bank in Espírito Santo, Brazil. Crop Breeding and Applied Biotechnology, 21:1–10. https://doi.org/10.1590/1984-70332021v21n2a32

Ferrão RG, Fonseca AFA, da Bragança SM, Ferrão MAG, De Muner LH (2007) Café conilon. Incaper.

Fornazier MJ, Martins DS, Fanton CJ, Benassi VLRM, Ferrão RG, Fonseca AFA, ..., Muner LH (2017) Manejo de Pragas do Café Conilon. Café conilon, 2:398-433.

Fracasso A, Trindade L, Amaducci, S (2016) Drought tolerance strategies highlighted by two Sorghum bicolor races in a dry-down experiment. Journal of plant physiology, 190: 1-14. http://dx.doi.org/10.1016/j.jplph.2015.10.009

Golabadi M, Arzani ASAM, Maibody SM (2006) Assessment of drought tolerance in segregating populations in durum wheat. African Journal of agricultural research, 1(5): 162-171. https://doi.org/10.5897/AJAR.9000070

Hair JF, Black WC, Babin BJ, Anderson RE, Tatham RL (2009) Análise multivariada de dados. Bookman editora.

Hindorf H, Omondi CO (2011) A review of three major fungal diseases of *Coffea arabica* L. in the rainforests of Ethiopia and progress in breeding for resistance in Kenya. Journal of advanced research, 2(2):109-120. https://doi.org/10.1016/j.jare.2010.08.006

Leite SA, Guedes RNC, Santos MPD, Costa DRD, Moreira AA, Matsumoto SN, ..., Castellani MA (2020) Profile of coffee crops and management of the Neotropical coffee

leaf miner, *Leucoptera coffeella*. Sustainability, 12(19):8011. https://doi.org/10.3390/su12198011

Mojena R (1977) Hierarchical grouping method and stopping rules: an evaluation. Computer Journal, 20:359-363. https://doi.org/10.1093/comjnl/20.4.359

Naik BJ, Kim SC, Shin MJ, Kim CW, Lim CK, An HJ (2019) Responses to biotic and abiotic stresses and transgenic approaches in the coffee plant. Journal of the Korean Society of International Agriculture, 31(4):359-377. https://doi.org/10.12719/KSIA.2019.31.4.359

Oliveira AC, Pizetta S, Reis E (2012) Análise do desenvolvimento inicial do cafeeiro conilon Cultivar robusta tropical submetido a déficit hídrico. Enciclopédia Biosfera, 8(15):90-100.

Parra JRP, Reis PR (2013) Manejo integrado para as principais pragas da cafeicultura, no Brasil. Visão Agrícola, 8(12):47-50.

Partelli FL, Oliosi G, Dalazen JR, da Silva CA, Vieira HD, Espindula MC (2021) Proportion of ripe fruit weight and volume to green coffee: Differences in 43 genotypes of *Coffea canephora*. Agronomy Journal, 113(2):1050-1057. https://doi.org/10.1002/agj2.20617

Partelli FL, Giles JAD, Oliosi G, Covre AM, Ferreira A, Rodrigues VM (2020) Tributun: a coffee cultivar developed in partnership with farmers. Crop Breeding and Applied Biotechnology, 20. https://doi.org/10.1590/1984-70332020v20n1c21

Pezzopane JR, da Silveira Castro F, Pezzopane JE, Bonomo R, Saraiva GS (2010) Climatic risk zoning for Conilon coffee in Espírito Santo, Brazil. Revista Ciência Agronômica, 41(3):341.

Pizetta S, Oliveira AC, Reis E, Rodrigues R, Olmo B (2012) Influência do déficit hídrico no desenvolvimento inicial do cafeeiro conilon. Enciclopédia Biosfera, 8(15). https://doi.org/10.15809/irriga.2011v16n2p115

Praxedes SC, DaMatta FM, Loureiro ME, Ferrão MA, Cordeiro AT (2006) Effects of long-term soil drought on photosynthesis and carbohydrate metabolism in mature robusta coffee (*Coffea canephora* Pierre var. kouillou) leaves. Environmental and experimental botany, 56(3):263-273. https://doi.org/10.1016/j.envexpbot.2005.02.008

Ramalho JC, DaMatta FM, Rodrigues AP, Scotti-Campos P, Pais I, Batista-Santos P, ..., Leitão AE (2014) Cold impact and acclimation response of *Coffea* spp. plants. Theoretical and Experimental Plant Physiology, 26(1):5-18.

Rodrigues WP, Vieira HD, Barbosa DHSG, Souza Filho GR, Candido LS (2013) Adaptability and genotypic stability of *Coffea arabica* genotypes based on REML/BLUP analysis in Rio de Janeiro State, Brazil. Genetics and Molecular Research, 12(3):2391-2399.

Rodrigues RR, Pizetta SC, Silva NKC, Ribeiro WR, dos Reis EF (2016) Growth initial conilon coffee under water deficit in soil. Coffee Science, 11(1):33-38.

RStudio Team (2022) RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Semedo JN, Rodrigues AP, Lidon FC, Pais IP, Marques I, Gouveia D, ..., Ramalho JC (2021) Intrinsic non-stomatal resilience to drought of the photosynthetic apparatus in *Coffea* spp. is strengthened by elevated air [CO2]. Tree Physiology, 41(5):708-727. https://doi.org/10.1093/treephys/tpaa158

Senra JFDB, Ferrão MAG, Ferreira De Mendonça R, Fonseca AFAD, Ferrão RG, Volpi PS, ..., Da Silva MW (2020) Genetic variability of access of the active germplasm bank of *Coffea canephora* of incaper in southern Espírito Santo. Journal of Genetic Resources, 6(2):172-184. https://doi.org/10.22080/jgr.2020.19162.1194

Silva CAD, Partelli FL, Aoyama EM, Bonomo R, Vieira HD, Ramalho JC, Ribeiro-Barros AI (2021) Floral morphology of robusta coffee genotypes. Agronomy Journal, 113(4): 3080-3088. https://doi.org/10.1002/agj2.20743

Silva CA da, Santos EA, Viana AP, Dias JRM, Partelli FL (2021) Genetic diversity in *Coffea canephora* genotypes for leaf nutrient concentration. Revista de la Facultad de Ciencias Agrarias UNCuyo, 53(1):22-34. https://doi.org/10.48162/rev.39.003

Silva LOE, Schmidt R, Valani GP, Ferreira A, Ribeiro-Barros AI, Partelli FL (2020) Root trait variability in *Coffea canephora* genotypes and its relation to plant height and crop yield. Agronomy, 10(9):1394. https://doi.org/10.3390/agronomy10091394

Silva EA, DaMatta FM, Ducatti C, Regazzi AJ, Barros RS (2004) Seasonal changes in vegetative growth and photosynthesis of Arabica coffee trees. Field Crops Research, 89(2-3):349-357. https://doi.org/10.1016/j.fcr.2004.02.010

Thioune EH, Strickler S, Gallagher T, Charpagne A, Decombes P, Osborne B, McCarthy J (2020) Temperature impacts the response of *Coffea canephora* to decreasing soil water availability. Tropical Plant Biology, 13(3):236-250.

Van der Vossen H, Bertrand B, Charrier A (2015) Next generation variety development for sustainable production of arabica coffee (*Coffea arabica* L.): a review. Euphytica, 204(2):243-256. https://doi.org/10.1007/s10681-015-1398-z

Venancio LP, Filgueiras R, Mantovani EC, do Amaral CH, da Cunha FF, dos Santos Silva FC, ..., Cavatte PC (2020) Impact of drought associated with high temperatures on *Coffea canephora* plantations: a case study in Espírito Santo State, Brazil. Scientific Reports, 10(1):1-21.

Zambolim L, Capucho A, Silva M (2015) Ferrugem do cafeeiro Conilon (*Coffea canephora*). Partelli F, Giles J, Silva M, Eds

CHAPTER 2

DIVERSITY AND STRUCTURE OF *Coffea canephora* FROM OLD SEMINAL CROPS IN ESPÍRITO SANTO, BRAZIL: GENETIC RESOURCES FOR COFFEE BREEDING

ABSTRACT

The *Coffea canephora* germplasm found in old seminal crops in the south of Espírito Santo state represents a valuable genetic resource for coffee improvement programs. The renewal of these crops by a few clonal cultivars can lead to a narrowing of the genetic base. The objective of this study was to characterize the genetic diversity and population structure of 280 genotypes of *C. canephora* from selected matrices in old seminal crops (15–46 years) in southern Espírito Santo using Single nucleotide polymorphism (SNP) molecular markers. Of the 9,491 SNPs obtained with the DArTseq™ technology, 2,542 high-quality ones were used in diversity and population structure analyses. Genetic diversity as expected heterozygosity and population structure using the STRUCTURE function. The cluster analysis revealed six groups and the STRUCTURE analysis detected two gene pools. The high expected frequency of heterozygotes (HE) for some of the groups formed by the cluster analysis indicates the genetic diversity in this population. Excess of heterozygous loci was verified for groups 4, 5, and 6. High FST values were detected between groups 4 and 2 (0.60), 2 and 5 (0.60), 4 and 3 (0.50), and between the two gene pools (0.59). This population had higher genetic diversity indices demonstrating that the seminal old crops in the south of Espírito Santo constitute a very rich germplasm bank. This valuable source must be maintained and conserved to ensure the sustainability of the coffee cultivation in an attempt to expand the genetic basis that has been reduced with the replacement of crops by clonal cultivars.

Keywords: Genetic diversity; SNP markers; Conilon coffee; DArTseq; germplasm

1. Introduction

The state of Espírito Santo (ES) is an important reference to produce *C. canephora* Conilon variety in the world (CONAB, 2021). In 2021, it produced 69% of all *C. canephora* in Brazil, with a production of 673.2 tons (CONAB, 2022). In the state, coffee farming occupies a special place in the history, culture, landscape and economy of more than 80% of the municipalities (Marré and Fonseca, 2021).

The species *C. canephora* is a diplod species (2n =22) and has two botanical varieties of *C. canephora*, Robusta and Conilon, cultivated commercially in Brazil belonging to the Congolese group (Maurin et al., 2007; Cubry et al., 2008; Batista-Santos et al., 2011), correspond to divergent heterotic groups with complementary characteristics (Souza et al., 2013; Bikila et al., 2013; Bikila et al., 2017). The Conilon variety is characterized by bushy growth, elongated leaves, early maturation and drought tolerance; however, it is also more susceptible to pests and diseases (Montagnon et al., 2012; Santos et al., 2017; Oliveira et al., 2018). The Robusta variety has vertical growth, larger leaves, medium to high sieve, late maturity and greater resistance to pests and diseases, although less drought tolerance (Montagnon et al., 2012; Santos et al., 2017; Oliveira et al., 2018).

The *C. canephora* germplasm present in the south of ES is maintained mainly by public institutions such as UFES, IFES, Incaper and by family farmers (Souza et al., 2013; Ferrão et al., 2019). The hypothesis is that there is a useful and diversified germplasm thanks to farmers who, historically, grow their crops via seed (Ferrão et al., 2019). This region probably contains the greatest variability of the species in the country (Fonseca et al., 1996). The old crops in the south of ES, originated from sexual reproduction from matrices selected by the farmers themselves, allowed the establishment of groups with wide genetic variability (Ferrão et al., 2019).

The genetic variability of *C. canephora* is attributed to the natural allogamy of the species due to gametophytic self-incompatibility (Lashermes et al., 1996), as well as the recent domestication process (Musoli et al., 2009) and the species has one of the widest geographic distributions within of the genus *Coffea* (Maurin et al., 2007). Genetic diversity between and within *C. canephora* groups is relevant for coffee breeding programs that seek new varieties (Lashermes et al., 2000; Leroy et al., 2005; Alkimim et al., 2018).

However, with the release of the first cultivars for ES in 1993 (Bragança et al., 1993), the increase in the use of these cultivars since then and the adoption of relatively few cultivars for crop renewal (Ferrão et al., 2007), can lead to the loss of genetic diversity, a key factor for genetic improvement. Therefore, the characterization of genetic diversity is necessary for better conservation and management of available genetic resources (Prakash et al., 2005; Bikila et al., 2017).

Different studies have been carried out to assess the genetic diversity and population structure of *C. canephora*, as well as other species of the genus *Coffea*, using SNPs (Garavito et al., 2016; Bikila et al., 2017; Alkimim et al., 2018; Anagbogu et al., 2019; Spinoso-Castillo et al., 2020), microsatellites (Prakash et al., 2005; Cubry et al., 2008; Souza et al., 2013; Loor Solórzano et al., 2017), ISSR (Yan et al., 2019) and AFLP markers (Prakash et al., 2005).

SNP markers are valuable tools used in genetic diversity studies in species of the genus *Coffea* (Garavito et al., 2016). DArTseq is based on next generation sequencing (NGS) and consists of reducing genome complexity using restriction enzymes (Sansaloni et al., 2011; Kilian et al., 2012). This methodology consists of whole genome sequencing and quickly identifies thousands of quality and highly informative SNPs in gene-rich regions (Kilian et al., 2012; Spinoso-Castillo et al., 2020).In coffee, this methodology has already been used efficiently to identify SNP markers in studies of diversity and population structure (Garavito et al., 2016; Spinoso-Castillo et al., 2020, 2021), detecting SNPs with a tendency to gene-rich regions when mapped to the newly sequenced *C. canephora* genome (Garavito et al., 2016).

In Brazil, a study with *C. canephora*, using microsatellite markers, detected high polymorphism and two main groups of genotypes: one from germplasm banks and one from genotypes collected from crops in Espírito Santo and Rondônia (Souza et al., 2013). However, a narrowing of the genetic base of cultivated *C. canephora* has already been observed, in comparison with the great genetic diversity that is expected from the species (Anagbogu et al., 2019). Therefore, the characterization and, mainly, the preservation of this material remaining from the old seminal crops in the south of ES should be considered for the management and conservation of the genetic resources of *Coffea canephora*.

The objectives of this study are: (i) to evaluate the genetic diversity and population structure of genotypes of ancient seminal cultures in the state of ES – Brazil by SNPs

obtained by the DArTseqTM methodology; (ii) to characterize the diversity of coffee genetic resources in the ancestral cultures that represent the germplasm cultivated in the cultures in ES in comparison with the commercial genotypes cultivated in the state; and (iii) to propose a new collection with new and divergent genotypes for the coffee growing program in Brazil. This is the first genome-wide SNP analysis of *C. canephora* genotypes from ancient seminal plants in ES.

2. Material and Methods

Plant Material

A total of 280 *C. canephora* genotypes were selected and young leaves of these materials were collected for DNA extraction (Table S2). Of these, 251 represent the old genetic resources of Conilon from the state of ES (Table 1). These genotypes were selected based on the vegetative vigor and productive potential from 393 half-sibling families (totaling 2,085 plants) in experimental design implemented since April 2018. Half-sib families originated from seeds from selected matrices in old seminal crops (between 15 and 46 years of implantation). The crops where the seeds were collected for planting the seedlings came from seeds. The collection was carried out in four municipalities in the southern region of the state of ES (Figure 1).

The experiment was implemented in Mimoso do Sul – ES (Latitude 21° 03' 01" South, Longitude 41° 30' 44" West, Altitude 620 meters). As *C. canephora* was initially introduced in the southern region of ES, cultures were sampled in representative municipalities that first received the introduced Conilon material. In addition, 29 commercial clones available and recommended for cultivation in ES were also analyzed in the study (Table 2), including 24 materials provided by Incaper (Centenário, Vitória, Jequitibá, Diamante, Robustão and Marilândia) and another five commercial clones (A1, P2, BRS, RO and Verdim). These clonal materials were selected because they are widely planted in the state of ES, and the five commercial clones (A1, P2, BRS, RO and Verdim) are our controls for the experiment.

Table 1. List of 251 *C. canephora* genotypes and 29 commercial cultivars evaluated with DArTseq™ SNP markers.

Location of matrix plant collection in sampled old crops	No. of individuals	Genotypes
Alegre	5	176.5, 176.6, 201.2, 201.6, 203.2
Cachoeiro de Itapemirim	31	125, **126**, 131.3, 132.2, 133, 134, 136.18, 136.18.1, 136.19, 136.26, 136.27, 136.3, 136.32, 136.33, 136.34, **136.5**, 183.4, 186.2, **186.3**, 186.41, **186.5**, 186.6, 187, 189.2, 189.4, 189, 192.2, 192.24, 192.3, 192.4, 192
Jerônimo Monteiro	65	144.2, 144, 145, 146, 149, 149.1 151.8, 151.9, 161.2, 161.7, 161.9, **162.10**, 162.10.1, 162.2, 162.6, 162.7, **162**, 164.10, **164.4**, 164.7, 164.8, 165.3, 165.4, **166**, 166.2, **166.2.1**, 166.5, 168.10, 168.3, 168.5, **168.6**, 168.9, 169.10, **169.4**, 169.5, 169, 170.6, 170, **171.10**, 171.2, 171.6, **171.9**, 171, **172.3**, 172.4, **172.5**, 172.6, **172.8**, 173.3, 173.6, 173.8, 179.2, 179.2.1, 179.3, 179.5, 179, 180.7, 181.3, 181.4, 181.4.1, 182.10, 182.3, 182.5, 182.8, 182.9
São José do Calçado	17	100, 103, 111, 112, 112.1, 113, 114, 114.1, 83, 85, 86, 91, 95.1, 95.1.1, **95**, 97, 98
South*	133	**06**, 21, 22, 24, 28.2, 281.2, 281.6, 281.8, 282.8, 29.2, 290.4, 300, 31.2, 31.4, 31, 32.10, 32.4, 32.4.1, 32.8, 34.2, 39.2, 40, 46, 53, 54, 58, 59, 60, 60.1,61, 62, 64, 75, G1, G10, G102, G103, G104, **G109**, G11, G110, G111, G115, G117, G118, G12, G120, G120.1, G120.1.1, **G127**, G127.1, G128, G129, G130, G131, G132, G133, G135, G136, G137, G138, G14, G14.1, G140, G141, G143, G144, G145, G146, G147, G148, G149, G15, G151, G153, G154, G155, G156, G16, G17, **G19**, G2, G20, G23, G25, G26, G27, G3, G32, G34, G37, G38, G41, G42, G43, G44, G46, G50, G51, G52, **G55**, G56, G57, **G59**, G59.1, G60, G61, G62, G63, G64, G67, G68, **G69**, G7, G70, G72, G73, G74, G75, G77, G78, G79, G8, **G80**, G9, G91, **G92**, **G93**, G95, G96, G97, G98, G99
Commercial cultivars	29	A1, P2, BRS, RO, Verdim, Centenário P1, Centenário P2, Centenário P3, Centenário P4, Centenário P7, Diamante P4, Diamante P8, **Jequitibá P1**, Jequitibá P2, Jequitibá P3, Jequitibá P4, Jequitibá P6, Marilândia P4, Marilândia P8, Robustão P1, Robustão P2, Robustão P3, **Robustão P4**, Robustão P5, Vitória P1, Vitória P2, Vitória P3, Vitória P4 and Vitória P5

*The "South" genotypes were collected in the municipalities of Alegre, Cachoeiro do Itapemirim, Jerônimo Monteiro and São José do Calçado in the state of Espírito Santo, but without identification in the field by the municipality of origin. In bold, are the 30 most divergent materials found from the genetic distance of Nei (1972).

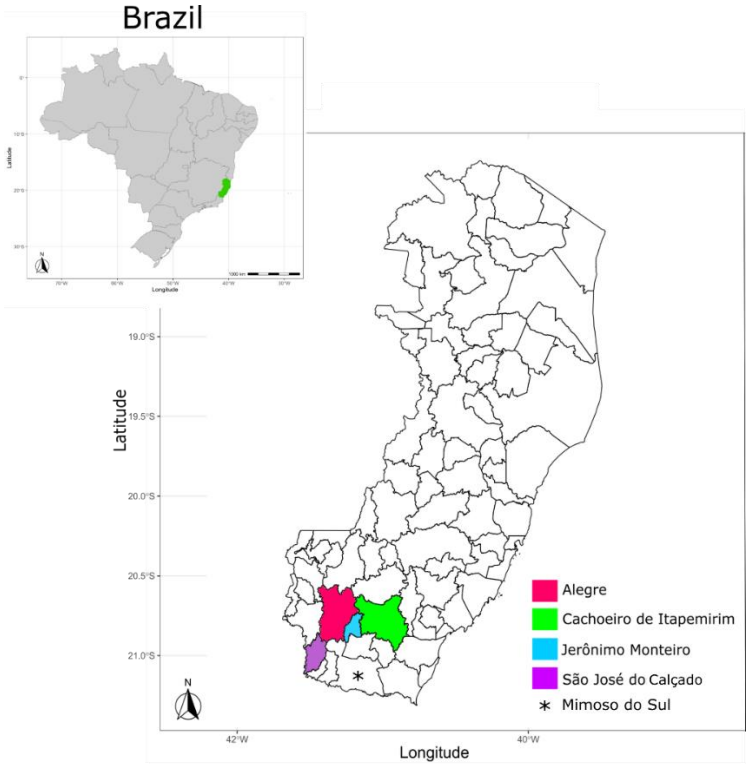

Figure. 1 Map of Espírito Santo indicating the municipalities where the seeds of the selected matrices were collected in the south-central region (Cachoeiro de Itapemirim and Jerônimo Monteiro) and Caparaó region (Alegre and São José do Calçado). The highlight, in green (up map), the location of the state of Espírito Santo on the map of Brazil. *Mimoso do Sul: municipality where the *C. canephora* seedlings from the experimental farm were planted.

Table 2. Characteristics of 29 commercial cultivars used as field control (A1, P2, BRS, RO, Verdim) and of the six commercial *C. canephora* materials provided by Incaper.

	Characteristics									
Commercial cultivar	**Variety/ release year**	**No. of clones of the variety**	**Cultivar type**	**Origin**	**Maturation**	**Harvest concentration**	**Water deficit**	**Rust resistance**	**Productivity (bags.ha-1)**	**Reference**
A1	Tributun/2017 Andina/2018	6	Clonal	Genotype initially propagated by Ivan Milanez and Hélio Dadalto, also known as H and H1	Intermediate	-	-	-	87.03	(Partelli et al. 2020; Partelli et al. 2019)
P2	Monte Pascoal/2020	6	Clonal	Genotype selected by producer Paulo Benacchi, in the municipality of Marilândia – ES	Intermediate	-	-	-	135.60	(Partelli et al. 2021)
Centenário (Centenário P1, Centenário P2, Centenário P3, Centenário P4 and Centenário P7)	Centenária ES8132/2013	9	Clonal	-	Late	July	-	MR*	82.40	(Ferrão et al. 2019)
Diamante (Diamante P4 and Diamante P8)	Diamante ES8112/2013	9	Clonal	-	Early	May	-	MR*	80.70	(Ferrão et al. 2019)
Jequitibá (Jequitibá P1, Jequitibá P2, Jequitibá P3, Jequitibá P4 and Jequitibá P6)	Jequitibá ES8122/2013	9	Clonal	-	Intermediate	June	-	MR*	88.70	(Ferrão et al. 2019)
Marilândia (Marilândia P7 and Marilândia P11)	Marilândia ES8143/2017	12	Clonal	Incaper conilon coffee breeding program, which presents the main characteristics of drought tolerance	Intermediate	May/June	Tolerance	MR*	80.98	(Ferrão et al. 2018)
Robustão (Robustão P1, Robustão P2, Robustão	Emcapa 8141/1999	10	Clonal	Incaper _ The most promising clones from the Incaper breeding program	Intermediate	May/June (with uniformity)	Tolerance	-	112.50	(Ferrão et al. 2000)

P3, Robustão P4 and Robustão P5)				were selected from those with drought tolerance characteristics						
Vitória (Vitória P1, Vitória P2, Vitória P3, Vitória P4 and Vitória P5)	Vitória Incaper 8142/2004	13	Clonal	Superior and selected clones among genetic material considered the “elite” of the Incaper breeding program	May to July (depending on clone)	-	Tolerance	-	70.40	(Ferrão et al. 2019)
Verdim	-	-	-	-	-	-	-	-	-	-
BRS	-	-	-	-	-	-	-	-	-	-
RO	-	-	-	-	-	-	-	-	-	-

* Moderately resistant

DNA Extraction

Young leaves were collected and stored in a freezer at -80°C until DNA extraction. DNA extraction followed the CTAB protocol of Doyle and Doyle (1990), with modifications from the coffee-optimized IAC, which only use MERK's chloroform, isoamyl alcohol and ethanol. The procedures performed were:

- 200 mg of leaf plant tissue (young leaves) were macerated and transferred to 2.0 mL eppendorf tubes;
- 700 µL of extraction buffer were added to the epperdorf with the macerated plant tissue and vortexed;
- The eppendorf tubes were left for 30 minutes at 65°C in a dry bath;
- 650 µL of CIA were added and it was homogenized for 10 minutes until an emulsion was formed;
- The tubes were taken to the centrifuge at 12000 rpm for 10 minutes;
- Then, the aqueous phase (700 µL) was transferred to a new 2.0 mL tube;
- 200 µL of extraction buffer were added and homogenized;
- 650 µL of CIA was added and homogenized again for 5 minutes;
- The tubes were taken to the centrifuge at 12000 rpm for 10 minutes;
- Afterwards, the aqueous phase (700 µL) was transferred to a new 1.5 mL tube and 650 µL of CIA was added
- The tubes were taken to the centrifuge at 12000 rpm for 10 minutes;
- The supernatant was transferred to a 1.5 mL tube;
- The DNA was precipitated with 500 µL of ice-cold isopropanol and homogenized for 5 minutes;
- After homogenization, the tubes were taken to the centrifuge at 12000 rpm for 10 minutes;
- The surface of the precipitate was washed with 250 µL of 70% ethanol;
- The tubes were taken to the centrifuge at 12000 rpm for 3 minutes;
- The two previous steps were repeated two more times;
- After these procedures, the ethanol is removed and taken to dry in a dry bath at 35°C;
- It was resuspended in 40 µL of TE with RNAse (40 µg/mL) and left in a water bath at 37°C for 30 minutes.

DNA concentrations and integrity were estimated using a Nanodrop™ 2000 spectrophotometer (Thermo Scientific). DNA quality was verified on 0.8% agarose gel. DNA genotypes prepared for genotyping using the DArTseq™ methodology were sent to the Service of Genetic Analysis for Agriculture (SAGA) in Mexico for high-throughput genotyping using the DArTseq™ technology.

In this whole genome sequencing methodology, based on next-generation sequencing, the genome complexity was reduced through restriction enzymes. The DArTseq™ methodology consisted of identifying thousands of SNPs along the genome of the species under study, mainly targeting gene regions. And, from it, quality and highly informative SNPs were identified that were used in the analysis of genetic diversity and population structure.

DArTseq™ Analysis Based on SNPs

The genome representation of the 280 *C. canephora* genotypes was obtained from the reduction of DNA complexity using two restriction enzymes, HpaII (frequent cut) and PstI (rare cut), and the ends of the cleaved fragments were linked to a code adapter and a common adapter to identify each sample. The fragments were then subjected to the PCR technique (polymerase chain reaction) where they were heated until denatured. Oligonucleotides and polymerase were added to promote the reaction and taken to the thermocycler which, through temperature variations, allows DNA denaturation. After this step, annealing occurred, which is the union of DNA strands (primers) on each complementary side of the strand. From this moment, the fragments were then amplified, since the template of the new molecule grew due to the combination of these complementation.

Subsequently, equimolar amounts of amplification products from each sample of the 96-well microtiter plate were pooled, purified and quantified, then sequenced on the Illumina Novaseq 6000 System platform. All successful amplifications were pooled and applied to a flow cell for amplification (Kilian et al., 2012). Clusters were sequenced on the Illumina HiSeq2500 sequencing platform, and sequences were processed using proprietary DArT analytical pipelines (Sansaloni et al., 2020). The barcode/sample sequences were identified and used in the label call. Poor quality sequences were filtered out, and identical ones were collapsed into fastqcall files. These files were used in a pipeline for DArT PL's proprietary SNP call algorithms (DArTsoft-seq14), as described

by Sansaloni et al. (2020).The amplified fragments were sequenced and the sequences were processed using the DArTseq™, developed and patented by DArT Pvt. Ltd. (Australia), generating two types of data, (1) codominant markers SNPs (single nucleotide polymorphism) and (2) dominant markers SilicoDArTs ("presence/absence"). SNP markers were filtered by quality parameters to select high quality markers for this study. SilicoDArTs markers were not used in this study

Data Analysis

The *dartR* package of the R software automatically calculated several quality parameters for each SNP marker. The DArTseq™ quality markers were determined by four main marker selection parameters: call rate with a threshold of 0.74 reproducibility parameter of 0.985; filter for monomorphic loci; and minor allele frequency (MAF) of 0.01. A total of 9,491 SNP markers for *C. canephora* were identified using the DArTseq™ methodology. After quality analysis by filtering the data, the 2,542 remaining SNPs were used for the genetic analysis. In addition to these parameters, the statistics of the DArTseq™ markers were verified from the expected and observed heterozygosity (HE and HO), the polymorphic information content (PIC), and the inbreeding fixation index (F), these diversity parameters were calculated from the *dartR* package of the RStudio software. After filtering the SNP markers through the quality parameters, the *CMplot* package of the R software was used to obtain the distribution of the SNP markers along the eleven chromosomes of *C. canephora*.

From the grouping performed, two other groupings were performed for the groups formed. For both groups, the filters used were call rate (74%), reproducibility (98.5%), monomorphic loci and MAF (1%).

Genetic Diversity and Population Structure

For the study of genetic diversity, (i) principal component analysis (PCA) and (ii) cluster analysis was performed using the Ward.D2 method. Additionally, the F statistic (FST) was used to identify the genetic differentiation between the groups formed in the cluster analysis. The population structure of the study population was analyzed using the program STRUCTURE v.2.3.4 (Pritchard et al., 2000). A series of 75,000 Markov Chain Monte Carlo (MCMC) simulations was performed for each K value from 1 to 10 with a 25,000 burn-in length, followed by 20 iterations. The number of hypothetical groups (ΔK) was estimated with the Structure Selector program (Li and Liu, 2018) through application

of grouping by Bayesian approach for the organization of genetically similar genotypes in the same groups.

Molecular analysis of variance (AMOVA) (Excoffier et al., 1992) was performed using the R software package poppr to separate the variation between and within groups belonging to the groups formed by the cluster analysis. The genetic distance was calculated using the StAMPP package of the R software and this function calculates Nei's genetic distance (Nei, 1972) between individuals.

3. Results

The main results found in this study were: i) great genetic diversity in ancient germplasm with six clusters, but few genotypes in the most divergent clusters and a strongly structured germplasm, with two main gene pools, but the prevalence of one; ii) groups of genotypes not represented in the commercial materials analyzed and greater divergence of commercial clones A1 and P2 grouped in group 4; iii) groups showing wide genetic differentiation and genotypes with high divergence highly divergent for mating purposes, suggesting possibilities of outcrossing. iv) the new cluster analyses, with only the 23 genotypes of groups 4, 5 and 6 of the first cluster, and only with the 257 genotypes of groups 1, 2 and 3, demonstrate the formation of 3 and 14 clusters, respectively.

The distribution of filtered SNP markers along the 11 chromosomes of *C. canephora* is shown in Fig.2. These SNPs were preferentially detected in gene regions according to the *C. canephora* reference genome (Denoeud et al., 2014). The number of SNPs per chromosome ranged from 129 to 470, and chromosomes 8 and 9 had the highest number of SNPs (Table S3). The observed heterozygosity (HO) with the eleven chromosomes of *C. canephora* ranged from 0.11 to 0.14, while the expected heterozygosity (HE) ranged from 0.14 to 0.17 (Table S3). The polymorphism information content (PIC) values ranged from 0.18 to 0.22.

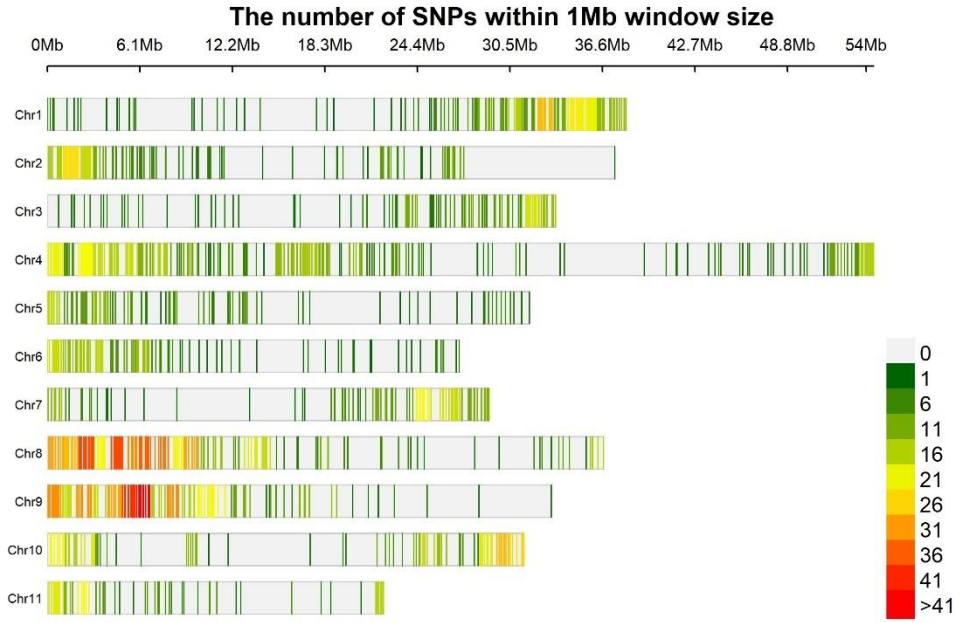

Figure 2. Distribution of 2,542 DArT SNP markers along the 11 chromosomes of *C. canephora* after filtering by call rate (74%), reproducibility (98.5%) and MAF (1.0%). The x-axis represents the position of the chromosome in Mb.

Six groups were obtained by cluster analysis (Fig. 3A). Group 4 presented the highest values of HO and HE (0.36 and 0.24 respectively), negative value of fixation index (-0.50), being considered the most divergent group, with 12 genotypes (Table 3). In group 4 there are two commercial clones: A1 and P2. Groups 5 and 6 can also be considered divergent, as they presented higher values of HO and HE, when compared to groups 1, 2, 3, and also presented negative values of fixation index (Table 3). In group 6 there are three commercial clones: Jequitibá P3, Vitória P1 and Centenário P2. In the PCA analysis, groups 4 and 5 were more distant from the others and the genotypes of group 6 showed greater spatial dispersion in relation to the other groups (Fig. 3B). Group 4 had the highest percentage of heterozygous loci, ranging from 33.7 to 43.7%. Groups 5 and 6 followed, with percentages ranging from 20.8 to 31.0% (Fig. 3C).

Table 3. Distribution of 280 genotypes of *C. canephora* germplasm in groups obtained by the Ward.D2 method. Observed and expected average heterozygosity (HO and HE) and fixation index (F) among genotypes belonging to different hierarchical levels (clusters generated in the dendrogram) and the distribution of the 29 commercial germplasms along the clusters.

Cluster	No. of genotypes	Genotypes	HO	HE	F
1	162	171, G97, 192.24, 136.26, 134, 114.1, G115, G98, 146, G103, 31.4, 31, G78, **Vitória P5**, 111, G145, 164.8, 151.8, G117, 34.2, 64, 131.3, 181.3, 39.2, 32.8, G130, G131, 281.8, 125, G7, G75, 181.4.1, 162.2, G62, 95.1.1, 95.1, 100, G10, **RO**, G11, G91, 300, G77, G46, G154, G96, 144.2, G44, 114, G144, 97, 95, 162.10.1, **BRS**, 181.4, 162, G104, 136.19, **Marilândia P7**, **Centenário P3**, G120, 136.32, 83, G34, 176.6, G19, G153, 172.6, 168.9, 170.6, G64, 168.3, 172.5, 169.4, 168.6, 166, 161.2, 46, **Marilândia P11**, G118, G79, G135, 40, G140, G15, G17, 151.9, 112.1, 113, G141, G14, G14.1, 189.2, 189.4, 179.5, 179.2.1, 179.4, G38, G111, 192, 192.3, 192.2, 166.2.1, 172.3, 166.2, 165.4, 172.4, 32.4.1, **Centenário P4**, 132.2, **Vitória P2**, 186.41, 281.6, 75, G42, G92, **Centenário P1**, 164.7, G63, 126, 149, G12, G51, G68, G146, G43, **Verdim**, 172.8, 173.3, G32, G149, 281.2, 171.9, 170, 176.5, 171.10, G50, 98, G69, G70, G60, G132, G133, 171.6, G37, G156, 173.6, G155, G138, G73, **Jequitibá P4**, 186.5, G127, 136.27, G52, 173.8, G120.1, G120.1.1, G23, 161.9, 180.7 and 22.	0.11	0.14	0.21
2	86	**Robustão P3**, 136.3, G110, 189, 136.34, 162.7, 32.4, 186.6, 182.3, 29.2, G102, 136.5, 136.33, 171.2, 162.6, 162.10, 182.9, G41, 182.8, 182.5, G72, G109, G74, 203.2, G55, **Robustão P2**, G127.1, 201.2, 112, G56, 179, 169, 166.5, 60.1, **Diamante P4**, 60, G2, G129, G3, G95, 183.4, G20, **Jequitibá P2**, **Vitória P3**, G57, 28.2, 179.3, **Vitória P4**, **Centenário P7**, 24, G148, G67, 21, 169.5, 161.7, 32.10, **Jequitibá P1**, **Diamante P8**, G26, 62, 145, G27, **Jequitibá P6**, 164.10, 31.2, G95, 6, 103, 54, 53, 85, 86, 168.10, 186.2, 179.2, 201.6, 144, 58, 186.3, G61, 136.18, 136.18.1, **Robustão P4**, **Robustão P5**, 59 and 61	0.11	0.12	0.08
3	9	G137, G92.1, 192.4, G147, G128, G151, 182.10, G136 and **Robustão P1**	0.10	0.07	-0.43
4	12	**A1**, 165.3, **P2**, G1, G59, 149.1, 290.4, G93, G80, 164.4, G59.1 and G143	0.36	0.24	-0.50
5	6	91, 168.5, 187, G25, G8 and G16	0.24	0.14	-0.71
6	5	G99, 169.10, **Jequitibá P3**, **Vitória P1** and **Centenário P2**	0.27	0.25	-0.08

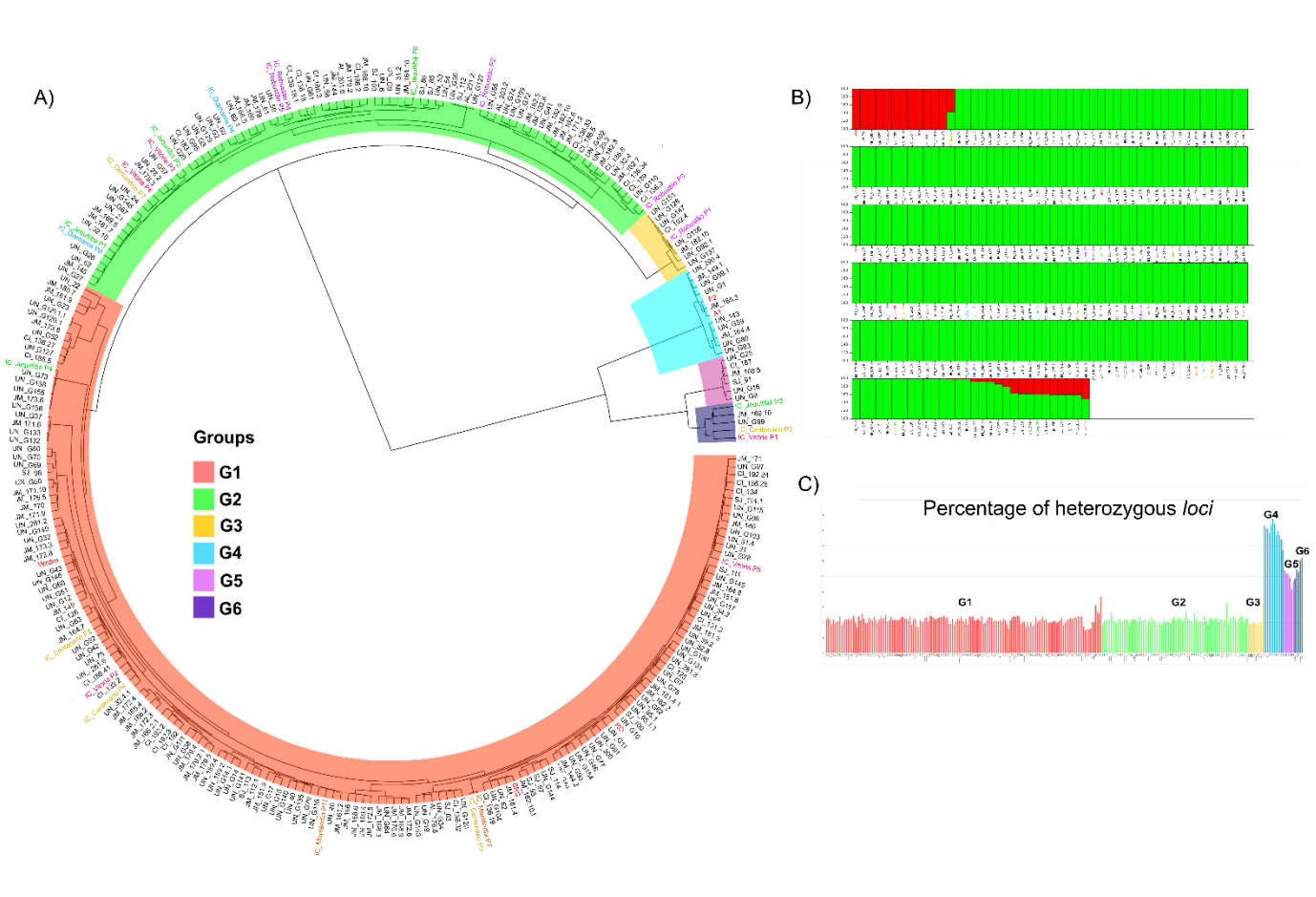

Figure 3. a) Dendrogram of 280 genotypes of *C. canephora* obtained with Euclidean distances calculated with SNPs and Ward.D2 method with proximity criterion between groups. b) STRUCTURE analysis with two gene pools c) Percentage of heterozygous *loci* in 280 *C. canephora* genotypes, on the x axis are the genotypes and on the y axis the percentage of heterozygous *loci*.

Groups 1, 2 and 3 comprise 257 genotypes from the total of genotypes evaluated, including 24 of the 29 commercial clones (Figure 3A and Table 3). The low values of HO and HE detected in these groups are indicators of inbreeding. As can be seen by group 1, which, in addition, presented the highest positive value of F (0.21), indicating an excess of homozygotes in the population (Table 3). Groups 1, 2 and 3 formed a large group in the PCA (Figure S1) and had their percentages of heterozygous loci ranging from 7.28 to 18.4% (Figure 3C). The PCA analysis of the most divergent groups showed greater dispersion of the groups and explained 65.6% of the variation (Figure S1B). For the groups of 257 genotypes, some clusters were further away from the others (3, 9, 10 and 14), but the other clusters did not diverge from the others (Fig. S1C).

Commercial clones were grouped into five of the six groups. Group 1 included eight clones provided by Incaper (Vitória P2, Vitória P5, Marilândia P7, Marilândia P11, Centenário P2, Centenário P3, Centenário P4 and Jequitibá P4) together with the genotypes RO, BRS and Verdim. Group 2 included half of the clones belonging to five different cultivars provided by Incaper (Vitória P3, Vitória P4, Jequitibá P1, Jequitibá P2, Jequitibá P6, Diamante P4, Diamante P8, Robustão P2, Robustão P3, Robustão P4, Robustão P5 and Centenário P7). According to the methodology proposed by Evanno et al. (2005), the highest value of ΔK was obtained for K = 2 and in the STRUCTURE analysis there was the formation of two gene pools (Figure 3B and Figure S2). Group 4 genotypes (Figure 3A) are included in the red gene pool (A1, 165.3, P2, G1, G59, 149.1, 290.4, G93, G80, 164.4, G59.1 and G143) and 88.9% of the genotypes (ie, n = 249) belonged to the green gene pool (Figure 3B and Figure S2). The hybrid genotypes, which showed a mixture of the two gene pools, are represented by 19 genotypes: eleven belonging to groups 5 and 6 (91, 168.5, 187, G25, G8, G16, G99, 169.10, Jequitibá P3, Vitória P1 and Centenário P2), four belonging to group 1 (22, 180.7, 161.9 and G23) and one belonging to group 3 (136.33) (Figure 3B and Table 3).

A new STRUCTURE analysis was performed with the green gene pool only, including 252 genotypes from the previous analyses. According to the methodology proposed by Evanno et al. (2005), the highest value of ΔK was obtained for K = 2 (Figure S3), showing that the genotypes can be separated into two groups. From a total of 29 clones that were used in previous analyses, 23 were grouped into this new group, demonstrating that most commercial clones used in the state share the same gene pool. The other genotypes showed a mixture of the two new gene pools found, with the

exception of genotypes G136, 192.4, Robustão P1, G151 and G128, which had only the green gene pool (Figure S3).

The results of molecular analysis of variance (AMOVA) for two hierarchical levels (groups formed by STRUCTURE and groups formed by hierarchical cluster analysis) revealed that the differences between the groups represented 31.88 and 17.76% of the genetic variation, respectively. So the genetic variation was greater between genotypes, within groups, than between groups, (both for groups formed by cluster analysis: 84.13%; and for those formed by STRUCTURE, 68.32%). A random permutation test indicated that the proportion of variance attributable to all two hierarchical levels was highly significant ($p = 0.001$) for differences between groups and within the sample (Table S4).

The greatest genetic differentiation (F_{ST}) occurred between groups 4 and 2 (0.60), 2 and 5 (0.59), and 4 and 3 (0.59) in the groups formed by clustering. Of the groups formed by STRUCTURE, the greatest genetic differentiation occurred between the green and red gene pools (0.59). The lowest values of genetic differentiation occurred between groups 1 and 3 (0.02) for the cluster analysis and, for STRUCTURE, the lowest value was between the green gene pool and the hybrids (0.12) (Table S5).

Genetic distance was calculated to identify contrasting parents to propose a new collection with divergent genotypes (Figure S4, Figure S5 and Table S6). The genetic distance of Nei (Nei 1972) between the genotypes was calculated, and the highest values of distance found were 0.38 (G93 and 168.6), 0.37 (171.9 and G93; G92 and G93; 172, 5 and G93; G55 and G93), 0.36 (172.8 and G93; 126 and G93; Jequitibá P1 and G93; G80 and 168.6; and 162 and G93). These values indicate potential crosses for coffee breeding programs.

From the previous analyses, two new groupings were made. One for the 23 genotypes of groups 4, 5 and 6, generating another cluster with 3 groups, with the same genotypes as the previous cluster (Figure 4). The diversity parameters of this new cluster showed that groups 1 and 2 had higher values of expected and observed heterozygosity (Table S7). The fixation index of groups 1 and 3 were negative, confirming what we detected in the previous analysis, that there is an excess of heterozygotes in these populations.

In filtering for monomorphic loci, for each of the new clusters, fixed loci between clusters were removed. Many different groups were formed (14 clusters) (Figure 5), when these loci were removed by the new filtering. The initial difference between the groups (Figure 3, first grouping) was explained by loci fixed in homozygosity (within the groups).

In this new approach, some groups showed negative (Table S7). The groups that showed the lowest negative values were cluster 3 (-0.78), 11 (-0.89), 14 (-0.89) and 12 (-0.91). The percentage of heterozygous loci can be seen in Figure 4C. The observed heterozygosity of the groups ranged from 0.15 to 0.21, with most groups having HO = 0.19. The commercial clones were well distributed in 8 of the 14 groups formed.

The new analysis of genetic differentiation, among the 14 clusters, showed that there is a high genetic differentiation between the clusters formed (Table S5). The greatest genetic differentiation occurred between clusters 11 and 12 and the smallest genetic differentiation occurred between clusters 1 and 2 (Table S5).

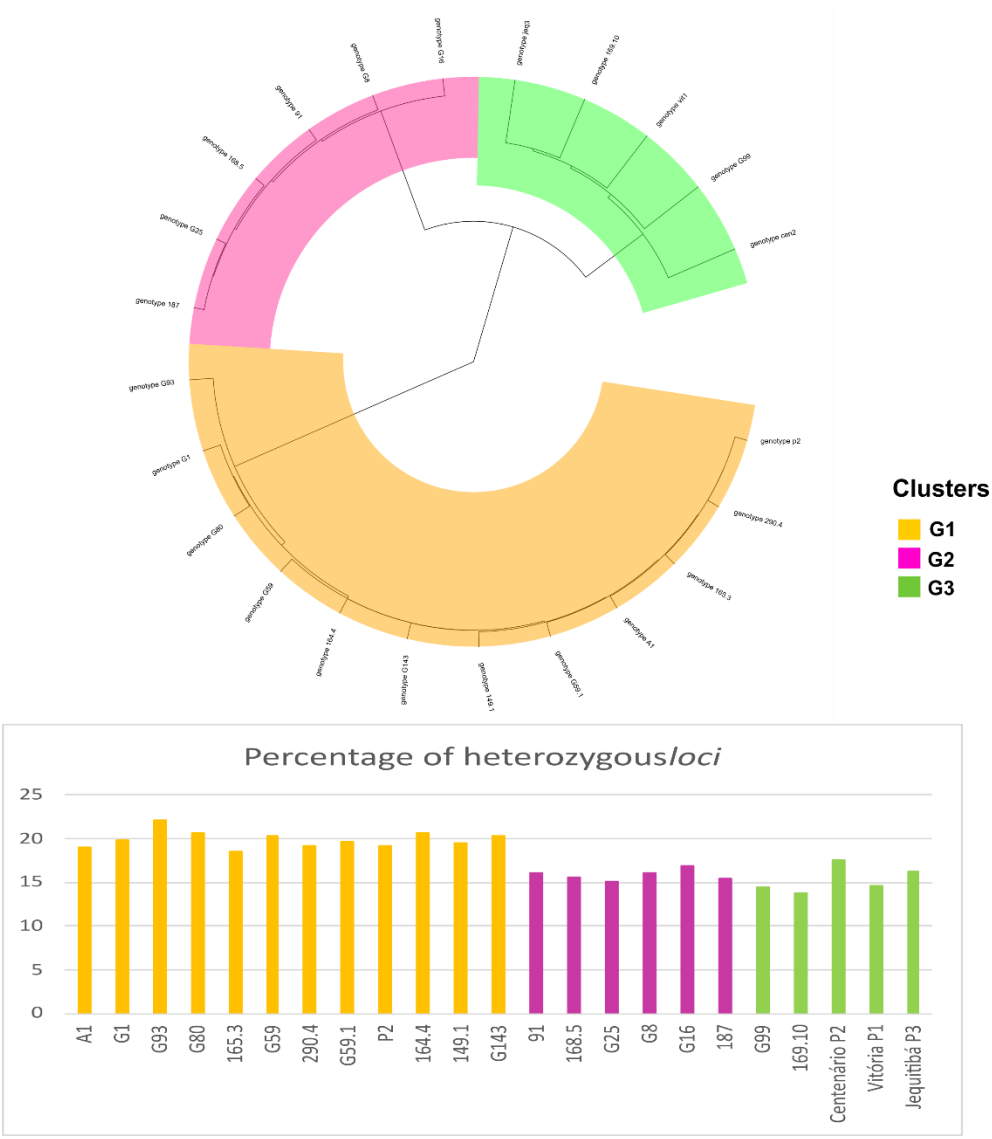

Figure 4. Cluster analysis with the 23 most divergent *Coffea canephora* genotypes from the first cluster, with the formation of three groups. We used the standardized mean Euclidean distance and Ward's method to form the groups; and, percentage of loci in heterozygosity for the group formed. On the x axis we have the genotypes and on the y axis we have the percentage of heterozygous loci.

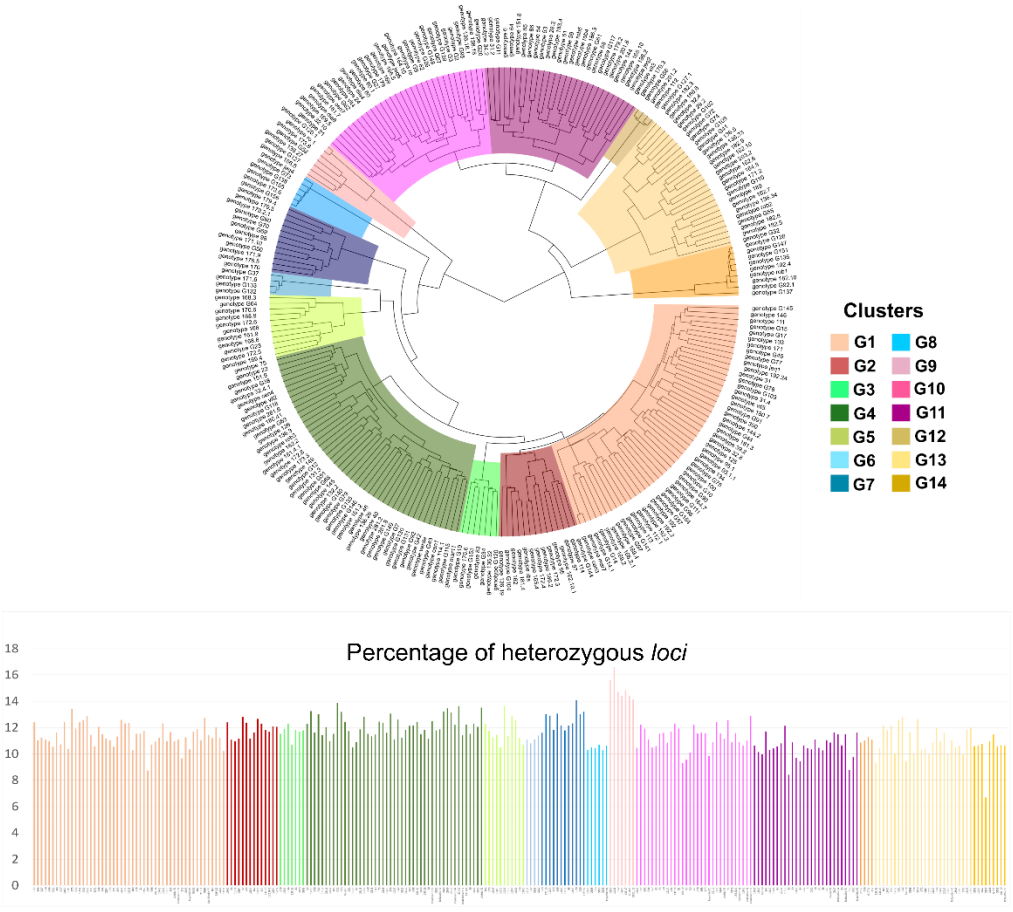

Figure 5. Cluster analysis with the 257 *Coffea canephora* genotypes from the first cluster, with the formation of 14 groups. We used the standardized mean Euclidean distance and Ward's method to form the groups; and, percentage of loci in heterozygosity for the group formed. On the x axis we have the genotypes and on the y axis we have the percentage of heterozygous loci.

4. Discussion

For the first time, the genetic diversity and population structure of ancient seminal crops from southern ES were characterized by SNP markers via DArTseq^{TM} methodology. This methodology has already been efficiently used to characterize the genetic diversity of germplasm within the genus *Coffea* (Garavito et al., 2016; Spinoso-Castillo et al., 2020, 2021).

After filtering the data, there was a 73.2% reduction in the initial number of SNPs. A similar result was previously detected by Garavito et al. (2016) and Spinoso-Castillo et al. (2020) using the same DArTseqTM methodology (62.8 and 89.8%, respectively). Our analyzes detected values of PIC, HE and HO of the chromosomes higher than those detected by previous studies (Garavito et al., 2016; Spinoso-Castillo et al., 2020), demonstrating the potential of ancient seminal crops in southern ES as a source of genetic variability. However, an average PIC value of 0.35 was reported in coffee progenies derived from interspecific crosses between Catuaí and Híbrido de Timor (Sousa et al. 2017), a higher value than that found in our study.

The six groups obtained by cluster analysis can provide direction for crossbreeding between highly divergent genotypes. In the cluster analysis, it is possible to observe that the commercial clones are distributed in five of the six groups formed. This result is extremely relevant for genetic improvement programs to increase the genetic base of the species. It also demonstrates that the improvement strategies that are being adopted for *C. canephora*, corroborate for this to happen. The main breeding methods that are being used in the launch of new cultivars aim to explore the natural genetic variability of *C. canephora*, through the selection of matrices (Ferrão et al., 2019).

The groups are differentiated by HE, HO, F and percentage of heterozygous loci, however, relatively few genotypes were grouped in the most divergent groups. The presence of most genotypes in the groups with lower diversity and low percentage of heterozygous loci, specifically for group 1 (57.9% of genotypes), indicates high inbreeding. This information is of great importance for breeding programs, suggesting the need to expand the narrow genetic base of new cultivars. For this reason, the use of these more divergent materials may favor the expansion of the genetic base of *C. canephora* (Eira et al., 2007).

Our cluster analysis detected the same number of clusters as reported by Bikila et al. (2017) studying clones of *C. canephora* (including Conilon and Robusta) and superior to what was detected by Alkimim et al. (2018), studying *C. canephora* (Conilon, Robusta and hybrids), both works also used SNP markers. With microsatellite markers, the genetic diversity of *C. canephora* cultivated in germplasm banks (IAC and UFV) and of genotypes collected from crops in Espírito Santo and Rondônia, revealed high polymorphism and two main groups of genotypes (Souza et al., 2013).

The low percentage of heterozygous loci in groups 1, 2 and 3 contrasts with the other groups. These groups, which grouped the majority of representatives of the seminal genotypes (83.2%) and 25 of the commercial clones, presented the lowest values of HO and HE. On the other hand, we have groups 4, 5 and 6 (23 genotypes) that showed higher diversity parameters, indicating the potential of these materials as a source of genetic variability. The quality parameters detected were comparable with other studies, being superior to those detected by Spinoso-Castillo et al. (2020) and similar to those found by Garavito et al. (2016).

Low heterozygosity and low percentage of heterozygous loci are not expected for *C. canephora*, as they originate from seeds and the species is allogamous with gametophytic self-incompatibility (Conagin and Mendes, 1961; Berthaud, 1980; Partelli et al., 2020). However, in groups 1 and 2 we observed positive values of F and a higher percentage of homozygous loci. This result is an indication that there may have been cooperation between producers in the south of ES through the exchange, distribution and propagation of these seeds, which is very common among producers from different regions (Fonseca et al., 1996). Allied to this hypothesis, the narrow genetic basis of *C. canephora* that has been detected, compared to the expected diversity for the species, can be explained why the introduction of this material would have been made in a single opportunity, brought from Guinea, Uganda and Angola (Eira et al., 2007). In a study carried out by Ngugi et al. (2019), also with *C. canephora*, most of the F values found were positive, ranging from 0.03 to 0.68.

However, in our work we also observed groups with negative values of F, as was the case of groups 3, 4, 5 and 6. These negative values are expected for allogamous species and are indicative of excess heterozygotes in the population (Wright, 1965). Negative results for the fixation index were also found by Souza et al. (2021) in parental and intraspecific hybrid progenies of Conilon coffee (-0.35 and -0.13, respectively).

It is important to note that two of the commercial clones grown in ES, A1 and P2, belong to the important cultivars Tributun and Monte Pascoal (Partelli et al., 2020, 2021), respectively. These clones were allocated to the red gene pool, while most of the evaluated genotypes belong to the green gene pool. These two clones are in the most divergent group and this result is important, as such clones are being widely used in the state and are a source of genetic variability. They can be used in crosses with other divergent materials to expand the genetic base of the species.

The cultivars released by Incaper were mostly allocated to groups 1 and 2, with 20 clones (83.3% of the total) made available by the institution. These clonal cultivars launched by Incaper have been the basis for the planting and renewal of Conilon coffee plantations in ES, which occurs on the order of 5% per year (Ferrão et al., 2019). However, some clones (Jequitibá P3, Vitória P1 and Centenário P2) were grouped into a more divergent group (group 6). to broaden the genetic base of this species.

The highest percentages of genetic variation occurred within groups (between genotypes) by AMOVA. In another study, the results found for *C. canephora* by Musoli et al. (2009) showed that there is a high percentage of variation (20.3%) explained by the groups formed (7 groups). And the highest percentage of variation occurred between individuals (51.3%), a similar result was detected in our study, with greater variation between genotypes.

The high values of F_{ST} detected are indicative of a greater genetic differentiation between the groups formed (Wright 1978), both for the cluster analysis and for the STRUCTURE analysis. The high genetic differentiation between groups should be exploited in genetic improvement in order to broaden the genetic base of the species. Our values were higher than those detected by Anagbogu et al. (2019) and similar to the mean FST values found by Garavito et al. (2016).

In addition to the analysis of genetic diversity and population structure, we also calculated the genetic distance of Nei (1972) between genotypes to choose divergent parents to generate good hybrids. We were able to detect high values of distances, indicating potential crosses between contrasting genotypes. And, our values were similar to those found by Spinoso-Castillo et al. (2020) who detected genetic distance values ranging from 0.26 to 0.39.

The genetic distance between genotypes reveals that those with higher values may be promising for breeding programs. The perspective of increasing the variability of this group in the coffee germplasm of Espírito Santo, with crosses between divergent parents, must be considered. To propose a new collection with promising, divergent and unpublished genotypes for the coffee breeding program in Brazil, the most contrasting genotypes must be prioritized. When we performed the analyzes separately, only for the 23 genotypes and only for the 257 genotypes, we observed that there is an increase in the number of groups for the grouping with the largest number of individuals. The quality parameters used were the same as in the initial analysis. The most divergent genotype groups remained the same as in the previous analysis. However, for the 257 individuals that were less divergent before, there was a cluster with 14 divergent groups, demonstrating the potential of these ancient materials from the south of the state as a source of genetic variability.

The presence of clones in eight of the 14 groups formed indicates that the breeding strategies used for the species are correct. Since the improved cultivars of *C. canephora*, obtained and recommended by Incaper and its partners for cultivation in ES, have constituted the main basis for the renewal of crops in the state (Ferrão et al., 2019). The mean values of HE (0.18) and HO (0.19) of the new 14 groups were higher than those found by Garavito et al. (2016) and Spinoso-Castillo et al. (2020). For the groups formed by the 23 genotypes, the values of HE and HO were even higher than those detected by previous studies. This demonstrates that the population under study has high values of diversity parameters.

Undoubtedly, the preservation of old seminal crops in the south of ES is fundamental. It is estimated that, currently, about 160 thousand ha, that is, 60% of coffee plantations in Espírito Santo, have been renewed using improved cultivars (Ferrão et al., 2015). The premise is not to stop using clonal cultivars, even because they were divergent in the study carried out. And, using a defined number of clones (together), in addition to guaranteeing the productive potential of the crop, they contribute to maintaining the genetic base (Ferrão et al., 2019). However, care must be taken because, if the crop is conducted with a reduced number of clones, in addition to compromising production, it could lead to disastrous results in the future of conilon coffee farming, due to the reduction of the available genetic base (Ferrão et al., 2019). Therefore, the maintenance

of seminal crops, as sources of divergent alleles, are necessary tools for coffee genetic improvement programs.

5. Conclusions

Our analyzes detected six divergent clusters and a strongly structured germplasm with two gene pools. Groups of genotypes not represented in the commercial materials analyzed were detected and greater divergence of commercial clones A1 and P2, grouped in group 4. Commercial clones are distributed in five of the six clusters formed.

In addition to the use of these clonal materials, it is interesting to characterize, preserve and manage the remaining crops of *C. canephora* in southern ES, using these materials in crosses with the clones that are already being planted, in order to expand the genetic base of this species.

The values of genetic differentiation and genotypes with high divergence suggest possibilities of crosses, aiming to increase the genetic base of the species.

The new analyzes were able to detect: i) the most divergent group, 23 genotypes, continued to show higher values of the diversity parameters (HO and HE), but the clusters formed were the same as in the first cluster analysis; ii) 14 new clusters in the genotypes that were previously in the less divergent groups, with higher values of HO and HE, when compared to previous analyses.

6. Acknowledgment

The authors would like to thank Conselho Nacional de Desenvolvimento Científico e Tecnológico [National Council for Scientific and Technological Development] (CNPq, Brasília – DF, Brazil, grant number 311950/2016-7), Fundação de Amparo à Pesquisa e Inovação do Espírito Santo [Research Support Foundation of Espírito Santo] (FAPES, Vitória – ES, Brazil), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior [Coordination for the Improvement of Higher Education Personnel] (CAPES, Brasília, DF, Brazil) – Finance Code 001, for financial support, CAFESUL - Cooperativa dos Cafeicultores do Sul do Estado do Espírito Santo, as well as the producers who allowed the collection of seeds in their crops and all who helped to implement the experiment.

7. References

Alkimim ER, Caixeta ET, Sousa TV, da Silva FL, Sakiyama NS, Zambolim L (2018) High-throughput targeted genotyping using next-generation sequencing applied in *Coffea canephora* breeding. Euphytica, 214:1–18. https://doi:10.1007/s10681-018-2126-2

Anagbogu CF, Bhattacharjee R, Ilori C, Tongyoo P, Dada KE, Muyiwa AA, Gepts P, Beckles DM (2019) Genetic diversity and re-classification of coffee (*Coffea canephora* Pierre ex A. Froehner) from South Western Nigeria through genotyping-by-sequencing-single nucleotide polymorphism analysis. Genet Resour Crop Evol 66:685–696. https://doi.org/10.1007/s10722-019-00744-2

Batista-Santos P, Lidon FC, Fortunato A, Leitão A, Lopes E, Partelli FL, Ribeiro AI, Ramalho JC (2011) The impact of cold on photosynthesis in genotypes of *Coffea* spp.—Photosystem sensitivity, photoprotective mechanisms and gene expression. J Plant Physiol 168:792–806. https://doi.org/10.1016/j.jplph.2010.11.013

Berthaud J (1980) L'incompatibilité chez *Coffea canephora*: méthode de test et déterminisme génétique. Café Cacao Thé 24: 267–274

Bikila BA, Sakiyama NS, Caixeta ET (2017) SNPs Based Molecular Diversity of *Coffea canephora*. J Microbiol Exp 5:1–4. https://doi.org/10.15406/jmen.2017.05.00136

Bragança SM, Fonseca AD, Silveira J, Ferrão RG, Carvalho C (1993) "EMCAPA 8111"," EMCAPA 8121"," EMCAPA 8131": primeiras variedades clonais de café conilon lancadas para o Espirito Santo. EMCAPA, Vitória, Brazil

CONAB – Companhia Nacional de Abastecimento (2021) Acompanhamento da safra brasileira de café: quarto levantamento – Safra 2021. https://www.conab.gov.br/info-agro/safras/cafe/boletim-da-safra-de-cafe?limitstart=0. Acessed December 2021

CONAB – Companhia Nacional de Abastecimento (2022) Acompanhamento da safra brasileira de café: quarto levantamento – Safra 2022. https://www.conab.gov.br/info-agro/safras/cafe/boletim-da-safra-de-cafe?limitstart=0

Conagin CH, Mendes AJT (1961) Pesquisas citológicas e genéticas em três espécies de *Coffea*: auto-incompatibilidade em *Coffea canephora* Pierre ex Froehner. Bragantia 20:788–804

Cubry P, Musoli P, Legnaté H, Pot D, de Bellis F, Poncet V, Anthony F, Dufour M, Leroy T (2008) Diversity in coffee assessed with SSR markers: Structure of the genus *Coffea* and perspectives for breeding. Genome 51:50–63. https://doi.org/10.1139/G07-096

Denoeud F, Carretero-Paulet L, Dereeper A, Droc G, Guyot R, Pietrella M, Zheng C, Alberti A, Anthony F, Aprea G, Aury JM, Bento P, Bernard M, Bocs S, Campa C, Cenci A, Combes MC, Crouzillat D, Da Silva C, Daddiego L, De Bellis F, Dussert S, Garsmeur O, Gayraud T, Guignon V, Jahn K, Jamilloux V, Jöet T, Labadie K, Lan T, Leclercq J, Lepelley M, Leroy T, Li LT, Librado P, Lopez L, Moñoz A, Noel B, Pallavicini A, Perrota G, Poncet V, Pot D, Priyono, Rigoreau M, Rouard M, Rozas J, Tranchant-Dubreuil C, VanBuren R, Zhang Q, Andrade AC, Argout X, Bertrand B, Kochko A, Graziosi G, Henry RJ, Jayrama, Ming R, Nagai C, Rounsley S, Sankoff D, Giuliano G, Albert VA, Wincker P, Lashermes P (2014) The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. Sci 345(80):1181–1184. https://doi.org/10.1126/science.1255274

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Eira MTS, Fazuoli LC, Guerreiro Filho O, Silvarolla MB, Ferrão MAG, da Fonseca AFA, ... & Souza FDF (2007) Bancos de germoplasma de café no Brasil: base do melhoramento para produtividade e qualidade.

Evanno G, Regnaut S, 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

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genet 131:479–491

Ferrão RG, da Fonseca AFA, Sebastião J, Silveira M, Ferrão MAG, Bragança SM (2000) EMCAPA 8141-Robustão Capixaba, variedade clonal de café conilon tolerante à seca, desenvolvida para o estado do Espírito Santo. Ceres 273:555–559

Ferrão RG, Fonseca AFA da, Bragança SM, Ferrão MAG, De Muner LH (2007) Café conilon. Vitória, ES: Incaper, 702 p

Ferrão RG, Ferrão MAG, Fonseca AFA da, Volpi PS, Verdin Filho AC, Lani JA, Mauri AL, Tóffano JL, Tragino PH, Bravim AJB, Morelli AP (2015) ES8122 - Jequitibá: nova

variedade clonal de café conilon de maturação intermediária para o Espírito Santo. Vitória, ES: Incaper, 220

Ferrão RG, Ferrão L, Volpi P, Ferrão M, Ferrão L, Verdin Filho AC, da Fonseca AFA (2018) Melhoramento genético para obtenção da cultivar Marilândia ES 8143, variedade clonal de café conilon tolerante à seca. Multi-Sci Res 1:1–18

Ferrão RG, Fonseca AFA, Ferrão MAG, de Muner LH (2019) Conilon Coffee. Vitória, ES, Brazil

Ferrão RG, Ferrão MAG, Fonseca AFA, Volpi PS, Verdin-Filho AC, Tóffano JL, Tragino PH, Bragança SM (2019) Cultivars of Conilon coffee In: Ferrão RG, Fonseca AFA, Ferrão MAG, de Muner LH (ed) Conilon Coffee, 3rd edn. Vitória, ES, pp 219–237

Fonseca AFA (1996) Propagação assexuada de *Coffea canephora* no Estado do Espírito Santo. In: Paiva R (ed) Workshop sobre avanços na propagação de plantas lenhosas. Lavras, MG, pp 31–34

Garavito A, Montagnon C, Guyot R, Bertrand B (2016) Identification by the DArTseq method of the genetic origin of the *Coffea canephora* cultivated in Vietnam and Mexico. BMC Plant Biol 16:1–12. https://doi.org/10.1186/s12870-016-0933-y

Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, Caig V, Heller-Uszynska K, Jaccoud D, Hopper C, Aschenbrenner-Kilian M, Evers M, Peng K, Cayla C, Hok P, Uszynski G (2012) Diversity arrays technology: A generic genome profiling technology on open platforms. Methods Mol Biol 888:67–89. https://doi.org/10.1007/978-1-61779-870-2_5

Lashermes P, Couturon E, Moreau N, Paillard M, Louarn, J (1996) Inheritance and genetic mapping of self-incompatibility in *Coffea canephora* Pierre. Theoret Appl Genet 93:458–462. https://doi.org/10.1007/BF00223190

Lashermes P, Andrzejewski S, Bertrand B, Combes MC, Dussert S, Graziose G, Trouslot P, Anthony F (2000) Molecular analysis of introgressive breeding in coffee (*Coffea arabica* L.). Theoret Appl Genet 100:139–146

Leroy T, Marraccini P, Dufour M, Montagnon C, Lashermes P, Sabau X, Piffanelli P (2005) Construction and characterization of a *Coffea canephora* BAC library to study the organization of sucrose biosynthesis genes. Theoret Appl Genet 111:1032–1041

Li YL, Liu JX (2018) StructureSelector: A web based software to select and visualize the optimal number of clusters using multiple methods. Mol Ecol Resour 18:176–177. doi: 10.1111/1755-0998.12719

Loor Solórzano RG, De Bellis F, Leroy T, Plaza L, Guerrero H, Subia C, Calderón D, Fernández F, Garzón I, Lopez D, Vera D (2017) Revealing the Diversity of Introduced *Coffea canephora* Germplasm in Ecuador: Towards a National Strategy to Improve Robusta. The Sci World J 2017:1–12. https://doi.org/10.1155/2017/1248954

Marré WB, da Fonseca AFA (2021) Indicação de Procedência (IP) Espírito Santo para o Café Conilon (*Coffea canephora*). Incaper em Revista 11-12:99–107. doi: 10.54682/ier.v11e12-p99-107

Maurin O, Davis AP, Chester M, Mvungi EF, Jaufeerally-Fakim Y, Fay MF (2007) Towards a Phylogeny for *Coffea* (Rubiaceae): Identifying Well-supported Lineages Based on Nuclear and Plastid DNA Sequences. Ann Bot 100:1565–1583. https://doi.org/10.1093/aob/mcm257

Montagnon C, Cubry P, Leroy T (2012) Amélioration génétique du caféier *Coffea canephora* Pierre : connaissances acquises, stratégies et perspectives. Cahiers Agric 21:143–153. https://doi.org/10.1684/agr.2012.0556

Musoli P, Cubry P, Aluka P, Billot C, Dufour M, de Bellis F, Pot D, Bieysse D, Charrier A, Leroy T (2009) Genetic differentiation of wild and cultivated populations: Diversity of *Coffea canephora* Pierre in Uganda. Genome 52:634–646. https://doi.org/10.1139/G09-037

NCBI – National Center for Biotechnology Information (2021) US, National Library of Medicine. https://www.ncbi.nlm.nih.gov/genome/?term=coffea%20canephora. Acessed October 2021

Nei M (1972) Genetic Distance between Populations. The Am Nat 106:283–292

Ngugi K, Aluka P (2019) Genetic and Phenotypic Diversity of Robusta Coffee (*Coffea canephora* L.). In: Ngugi K, Aluka P (ed) Caffeinated and Cocoa Based Beverages, pp 89–130. https://doi.org/10.1016/B978-0-12-815864-7.00003-9

Oliveira LNL de, Rocha RB, Ferreira FM, Spinelli VM, Ramalho AR, Teixeira AL (2018) Selection of *Coffea canephora* parents from the botanical varieties Conilon and Robusta

for the production of intervarietal hybrids. Ciência Rural 48:1–7. https://doi.org/10.1590/0103-8478cr20170444

Partelli FL, Golynski A, Ferreira A, Martins MQ, Mauri AL, Ramalho JC, Vieira HD (2019) Andina-first clonal cultivar of high-altitude conilon coffee. Crop Breed Appl Biotechnol 19:476–480. http://dx.doi.org/10.1590/1984- 70332019v19n4c68

Partelli FL, Giles JAD, Oliosi G, Covre AM, Ferreira A, Rodrigues VM (2020) Tributun: a coffee cultivar developed in partnership with farmers. Crop Breed and Appl Biotechnol 20:1–4. https://doi.org/10.1590/1984-70332020v20n2c21

Partelli FL, Covre AM, Oliosi G, Covre DT (2021) Monte Pascoal: First clonal conilon coffee cultivar for southern Bahia-Brazil. Funct Plant Breed J 3:107–112. doi: 10.35418/2526-4117/v3n2a9

Partelli FL, Oliosi G, Dalazen JR, da Silva CA, Vieira HD, Espindula MC (2021) Proportion of ripe fruit weight and volume to green coffee: differences in 43 genotypes of *Coffea canephora*. Agron J 1–9. doi:10.1002/agj2.20617

Prakash NS, Combes MC, Dussert S, Naveen S, Lashermes P (2005) Analysis of genetic diversity in Indian robusta coffee genepool (*Coffea canephora*) in comparison with a representative core collection using SSRs and AFLPs. Genet Resour Crop Evol 52:333–343. https://doi.org/10.1007/s10722-003-2125-5

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945–959. doi: 10.1093/genetics/155.2.945.

Santos AV, Rocha RB, Fernandes C de F, da Silveira SF, Ramalho AR, Vieira Júnior JR (2017) Reaction of *Coffea canephora* clones to the root knot nematode, Meloidogyne incognita. Afr J Agric Res 11:916–922. https://doi.org/10.5897/AJAR2016.11999

Sansaloni C, Petroli C, Jaccoud D, Carling J, Detering F, Grattapaglia D, Kilian A (2011) Diversity arrays technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. BMC Proc.5:(S7) –54

Sansaloni C, Franco J, Santos B, Percival-Alwyn L, Singh S, Petroli C, Campos J, Dreher K, Payne T, Marshall D, Kilian B, Milne I, Raubach S, Shaw P, Stephen G, Carling J, Pierre CS, Burgueño J, Crosa J, Li HH, Guzman C, Kehel Z, Amri A, Kilian A, Wenzl P, Uauy C, Banziger M, Caccamo M, Pixley K (2020) Diversity analysis of 80,000 wheat

accessions reveals consequences and opportunities of selection footprints. Nat Commun 11:4572. https://doi.org/10.1038/s41467-020-18404-w

Sousa TV, Caixeta ET, Alkimim ER, de Oliveira ACB, Pereira AA, Sakiyama NS, Resende Júnior MFR, Zambolim L (2017) Population structure and genetic diversity of coffee progenies derived from Catuaí and Híbrido de Timor revealed by genome-wide SNP marker. Tree Genet Genomes 13:1–16. https://doi.org/10.1007/s11295-017-1208-y

Souza F de F, Caixeta ET, Ferrão LFV, Pena GF, Sakiyama NS, Zambolim EM, Zambolim L, Cruz CD (2013) Molecular diversity in *Coffea canephora* germplasm conserved and cultivated in Brazil. Crop Breed Appl Biotechnol 13:221–227. https://doi.org/10.1590/s1984-70332013000400001

Souza LC, Ferrão MAG, Carvalho RD, Ferrão RG, Fonseca AFD, Pinheiro PF, Soares TC (2021) Molecular characterization of parents and hybrid progenies of conilon coffee. An Acad Bras Cienci 93:e20201649 D. https://doi.org/10.1590/0001-3765202120201649

Spinoso-Castillo JL, Escamilla-Prado E, Aguilar-Rincón VH, Ramos VM, de los Santos GG, Pérez-Rodríguez P, Corona-Torres T (2020) Genetic diversity of coffee (*Coffea* spp.) in Mexico evaluated by using DArTseq and SNP markers. Genetic Resour Crop Evol 67:1795–1806. https://doi.org/10.1007/s10722-020-00940-5

Spinoso-Castillo JL, Escamilla-Prado E, Aguilar-Rincón VH, Corona-Torres T, García-de los Santos G, Morales-Ramos V (2021) Quantitative comparison of three main metabolites in leaves of *Coffea* accessions by UPLC-MS/MS. Eur Food Res Technol, 247:375–384. https://dx.doi.org/10.17632/9hjtpksb9v.1

Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 395–420

Wright S (1978) Evolution and the genetics of populations Chicago, University of Chicago

Yan L, Ogutu C, Huang L, Wang X, Zhou H, Lv Y, Long Y, Dong Y, Han Y (2019) Genetic Diversity and Population Structure of Coffee Germplasm Collections in China Revealed by ISSR Markers. Plant Mol Biol Report 37:204–213. https://doi.org/10.1007/s11105-019-01148-3

CHAPTER 3

GENOME WIDE ASSOCIATION FOR INITIAL DEVELOPMENT CHARACTERS AND BIOTIC STRESSES IN *Coffea canephora*

ABSTRACT

The marker assisted selection is an important approach in coffee breeding programs. The GWAS methodology explores the association among molecular markers and genomic regions with the traits of interest. In this study, we aimed to identify chromosomal regions with significant associations, by GWAS, in 251 *Coffea canephora* genotypes for four traits related to early development and four related to incidence of biotic and abiotic stresses. After filtering, 2,542 SNP distributed along the entire genome of *C. canephora* were used. The mixed models were applied to phenotypic data. From the sum of the genetic values (BLUPs) and the residues of the materials that were genotyped, the p-values were calculated and used in the GWAS analyses. A total of 115 SNPs had significant associations: 48 for height growth; 20 for mealybug incidence; seven for plant heigh and growth rate in diameter; 11 for stem diameter; 16 for rust incidence and five for leaf miner incidence. Most of these SNPs are located within or close to candidate genes and it was possible to identify the putative function of these candidate genes for the trait evaluated. Significant associations between molecular markers and the traits, as well as, the identification of candidate genes that could be used for further in-depth studies to carry out the selection assisted by molecular markers in coffee genetic breeding programs.

Key words: GWAS; Molecular breeding; Conilon coffee; Single nucleotide polymorphism

1. Introduction

Coffee genetic breeding programs involve expensive, time-consuming, complex and dynamic processes (Almeida et al., 2021), as it is a perennial crop that takes 25 to 30 years to develop a new variety (Moncada et al., 2016). The use of molecular marker-assisted selection (MAS) is a powerful and efficient tool to accelerate the selection of superior genotypes and the development of new coffee cultivars, increasing the efficiency of breeding programs (Alkimim et al., 2017; Sousa et al., 2019). Molecular markers have the advantage of not being affected by environmental conditions and can be used at any stage of plant development (Almeida et al., 2021).

Molecular studies allow analyzing the genetic structure and similarity between individuals, concomitantly with phenotypic assessments, allowing the selection of superior genotypes (Sousa et al., 2017). With the identification of SNPs (*single nucleotide polymorphisms*), these can be used to identify genes that are associated with traits of interest and identify superior genotypes through genome-wide association (GWAS).

The crop of *Coffea canephora* has high economic importance in the world, and the main objectives of breeding for the species are the selection of highly productive plants, resistant to biotic and abiotic stresses and with sensory quality (Maia et al., 2017). The discrimination and selection of superior genotypes has been carried out based on morpho-agronomic characters of interest. However, the use of only phenotypic characters has limitations such as environmental influence and low selection gain (Souza et al., 2011; Ferrão et al., 2019).

In genetic breeding programs, phenotypic and genotypic evaluations are necessary. And, in addition to production-based selection, other traits have been evaluated to maximize selection gains, including early evaluations of morpho-agronomic traits in order to select the most promising genotypes and determine those traits that are most representative for the selection of genetic materials (Cruz et al., 2014). The characterization is based on descriptors defined for each species and the evaluation will include the measurement of characteristics related to productivity, development, quality attributes, tolerance to biotic and abiotic stresses (Ferrão et al., 2022).

For the genetic breeding of *C. canephora*, in addition to criteria related to productivity, it is sought to select more vigorous plants, with a larger crown diameter that are associated with a lower plant height (Alkimim et al., 2017). Thus, the study of these

materials is a strategy for the discrimination of individuals with good initial vegetative development, based on characters highly correlated with yield and productivity, such as plant height and stem diameter (Martinez et al., 2007; Moncada et al., 2016). Biotic factors also cause damage to the growth and development of Conilon coffee, such as the biotrofic fungus *Hemileia vastatrix*, leaf miner (*Leucoptera coffeella*) and mealybugs. And, although the use of chemical control is efficient in controlling pests and diseases, the most interesting and efficient control would be the use of varieties with genetic resistance (Mohammed, 2015; Talhinhas et al., 2017). In this context, selecting genetic materials that are tolerant to stress and that present adequate vegetative development is fundamental within the strategies of new cultivar development.

SNP markers are highly used in genetic breeding programs. They are the most abundant type of polymorphism in the genome, being generally biallelic and codominant (Resende et al., 2008) and may be associated with genes that control the main characteristics of agronomic interest (Heffner et al., 2009; Sousa et al., 2017). The availability of the *C. canephora* reference genome (Denoeud et al., 2014) and the use of NGS (*next generation sequencing*) technologies provide necessary tools for genotyping, resulting in highly informative and quality SNPs (Andrade et al., 2017). Modern statistical methods, such as the use of mixed models, together with the application of high-thought molecular markers, have been used in plant breeding, allowing the breeder to accurately estimate the genetic value (Kamfwa et al., 2015; Zuiderveen et al., 2016; Perseguini et al., 2016; Resende et al., 2018).

GWAS studies seek the association between the phenotype of interest and the genotype, enabling the identification of regions of the genome that have the greatest effect on a given trait (Sant'Ana et al., 2018). This analysis allows to find genes that contain significant SNPs and the identification of candidate genes that participate in the control of the trait and its biological function, facilitating the understanding of the influence of the genotype on the phenotype (Yang et al., 2013).

In GWAS it is possible to identify SNPs that arise from mutations throughout the evolutionary history of each species and persist with a pattern in a population in disequilibrium for a particular trait, this type of analysis requires a high density of markers such as SNPs (Gimase et al., 2020). SNPs with significant association in overlapping genes or close to them are considered potentially involved with the phenotypic character that was evaluated, and able to identify the genomic sequences in the reference genomes

as occurred for *C. canephora* (González et al., 2017; Visscher et al., 2017; Monteiro et al., 2021; Alquimin et al., 2020).

GWAS studies are scarce for the crop of *C. canephora* and *C. arabica*. For *C. arabica*, the lipid content, including the diterpenes cafestol and kahweol, which are related to beverage quality had SNPs within or near candidate genes related to metabolic pathways of these chemical compounds in grains of coffee (Sant´Ana et al., 2018). For *Colletotrichum kahawae*, the causal agent of the Coffee Berry Disease, two SNP markers were significantly associated with CBD incidence on chromosomes 1 and 2 (Gimase et al., 2020). In another study, Spinoso-Castillo et al. (2022), working with 80 genotypes of some species of the genus *Coffea* (*C. arabica*, *C. canephora* and *C. liberica*), detected 3 SNPs with significant association for important characteristics related to coffee quality related to metabolite content. The only published work for *C. canephora* found a significant association with plant height, canopy projection diameter, vegetative vigor, incidence of rust and incidence of cercosporiosis (De Faria et al., 2022).

This study will provide information relevant for the genetic breedin programs for *C. canephora*, since the chromosomal regions can be used in studies of selection assisted by molecular markers. The objective of this study is to identify chromosomal regions with significant associations in *C. canephora* genotypes for the eight phenotypic traits of importance for this species, using GWAS methodology and to identify SNPs inserted in candidate genes.

2. Material and Methods

Plant Material

A total of 251 *C. canephora* genotypes were evaluated, including 246 genotypes from a breeding population of the UFES and five commercial clones (A1, P2, BRS, RO and Verdim) (Table S2). The breeding population are from seminal seeds from old crops in the south of Espírito Santo (15 – 46 years of implantation). The 246 genotypes were selected for vegetative vigor and productive potential in 388 half-sib families (totaling 2,085 plants). The half-sib families come from seeds from selected matrices in four municipalities in the southern region of the state of Espírito Santo (Alegre, Cachoeiro de Itapemirim, Jerônimo Monteiro and São José do Calçado). The experiment was implemented in Mimoso do Sul – ES (Latitude: 21° 03 '01' 'South, Longitude: 41° 30' 44" West, Altitude 620 meters) in Federer's augmented blocks in April 2018.

DNA Extraction and Genotyping

Young leaves were collected and stored in a freezer at -80°C until DNA extraction. DNA extraction followed the CTAB protocol of Doyle and Doyle (1990), with modifications from the coffee-optimized IAC, which only use MERK's chloroform, isoamyl alcohol and ethanol. The procedures performed were:

- 200 mg of leaf plant tissue (young leaves) were macerated and transferred to 2.0 mL eppendorf tubes;
- 700 µL of extraction buffer were added to the epperdorf with the macerated plant tissue and vortexed;
- The eppendorf tubes were left for 30 minutes at 65°C in a dry bath;
- 650 µL of CIA were added and it was homogenized for 10 minutes until an emulsion was formed;
- The tubes were taken to the centrifuge at 12000 rpm for 10 minutes;
- Then, the aqueous phase (700 µL) was transferred to a new 2.0 mL tube;
- 200 µL of extraction buffer were added and homogenized;
- 650 µL of CIA was added and homogenized again for 5 minutes;
- The tubes were taken to the centrifuge at 12000 rpm for 10 minutes;
- Afterwards, the aqueous phase (700 µL) was transferred to a new 1.5 mL tube and 650 µL of CIA was added
- The tubes were taken to the centrifuge at 12000 rpm for 10 minutes;
- The supernatant was transferred to a 1.5 mL tube;
- The DNA was precipitated with 500 µL of ice-cold isopropanol and homogenized for 5 minutes;
- After homogenization, the tubes were taken to the centrifuge at 12000 rpm for 10 minutes;
- The surface of the precipitate was washed with 250 µL of 70% ethanol;
- The tubes were taken to the centrifuge at 12000 rpm for 3 minutes;
- The two previous steps were repeated two more times;
- After these procedures, the ethanol is removed and taken to dry in a dry bath at 35°C;
- It was resuspended in 40 µL of TE with RNAse (40 µg/mL) and left in a water bath at 37°C for 30 minutes.

DNA concentrations and integrity were estimated using a NanodropTM 2000 spectrophotometer (Thermo Scientific). And the DNA quality was verified in agarose gel (0.8% concentration). DNA genotypes prepared for genotyping using the DArTseq methodology were sent to the Service of Genetic Analysis for Agriculture (SAGA) in Mexico for high-throughput genotyping using the DArTseqTM technology.

For the SNP-Based DArTseq analysis, the genome representation of the 251 genotypes of *C. canephora* was obtained by reducing the DNA complexity using two restriction enzymes, HpaII (frequent cut) and PstI (rare cut) and the ends of the cleaved fragments were ligated to a code adapter and a common adapter to identify each sample. The fragments were amplified and after the PCR reaction, equimolar amounts of amplification products from each sample of the 96-well microtiter plate were pooled, purified and quantified, followed by sequencing on the Illumina Novaseq 6000 system platform and the sequences were processed using the analytical program developed and patented by DArT Unip. Ltd. (Australia), generating SNP marker data.

In the data analysis, the SNP markers were filtered for quality by the dartR package of the R *software* using the parameters: *call rate* with a threshold of 0.74, reproducibility parameter of 0.985, filter for monomorphic loci and minor allele frequency (MAF) of 0.01. Missing data associated with the codominant marker genotypes (denominated missing values of NA) were imputed using the concept of mean or expected value, where the expected value of the indicator variable is 2p. And the percentage of missing values was also verified and there was 5.5% of missing data that were imputed.

The position of each marker in the *C. canephora* reference genome was determined using the Browse Genome tool from the *National Center for Biotechnology Information* website (https://www.ncbi.nlm.nih.gov). From the positions that were provided by the results sent by the DArTseq^{TM} methodology, as well as the chromosome number where the SNPs were located, in the search field of the Browse Genome tool, the chromosome number was entered along with the position of the SNP. From there, SNPs inserted in the genes or close to them were identified. With this information, it was possible to identify the annotations of the mentioned genes that were found.

The *C. canephora* has 11 chromosomes as a basic number. In addition to the SNPs located in the 11 chromosomes of the species, SNPs are identified in “chromosome 0” The chromosome named “0” is not a true chromosome, but a set of sequence *scaffolds*

(Sousa et al., 2017; Sant'Ana et al., 2018; Merot-L'anthene et al., 2019; Gimase et al., 2020), therefore, they were excluded from further analysis. However, a table (Table 3) was built, indicating in these scaffolds, those SNPs that showed a significant association for the evaluated characteristics.

Principal component analysis (PCA) was used to assess the level of population structure of *C. canephora* genotypes using the kinship matrix H (Munoz) using the RStudio software by the "ggfortify" package (RStudio Team, 2022).

Phenotypic Analyses

Eight characteristics were phenotyped, four related to early development and four to incidence to biotic and abiotic stresses. The plant height was evaluated with a graduated ruler and measured in centimeters (cm) from the ground level to the terminal pair of leaves of the orthotropic branch and the stem diameter was verified with a digital caliper and measured in millimeters (mm) at 5 cm above ground level. These two analyses were carried out bimonthly, from April 2019 to February 2020, totaling six measurements.

With the plant height and stem diameter data, the growth rates were obtained in $cm.day^{-1}$ and $mm.day^{-1}$, respectively, using the formula below, in which GR = growth rate, Lf = final length; Li = initial length; and nd = number of days.

$$GR = \frac{(Lf - Li)}{nd}$$

Drought tolerance was estimated from the visual categorical analysis of the plants using grades from 1 to 9. This analysis was carried out in February 2019 in 200 families (totaling 1040 plants), in which 1 = vigorous plants without wilting symptoms; 2 = vigorous plants with slightly overhanging leaves; 3 = vigorous plants with some drooping leaves; 4 = dangling leaves; 5 = completely hanging leaves; 6 = completely hanging leaves with onset of discoloration, loss of leaf shine, and slight drying; 7 = completely hanging leaves with discoloration, loss of brightness in all leaves and moderate drying of the leaves; 8 = completely hanging leaves with discoloration and loss of brightness in all leaves, in addition to high drying intensity, with some brown color; and 9 = completely dry plant, showing permanent damage (adapted from Carvalho et al., 2017).

The evaluation of biotic stresses was carried out through a graded visual analysis in December 2019, 600 days after planting, in which the rainy season (November-December) favors the rust epidemic (Zambolim et al., 2015). For rust (*Hemileia vastatrix*), the scale ranged from 1 to 5, based on the severity of the disease, in which grade 1 = absence of spots or pustules or formation of spores; 2 = plants with lesions ranging from spots to chlorosis in the infected area but without the formation of urediniospores; grade 3 = low number of pustules per leaf with formation of urediniospores; grade 4 = medium number of pustules per leaf with formation of urediniospores; and grade 5 = high number of pustules per leaf with formation of urediniospores (adapted from Carvalho et al., 2017).

For the analysis of resistance to leaf miners (*Leucoptera coffeaella*), lesions were classified as small (0.3 to 0.6 centimeters in diameter); medium (about 0.6 to 1.2 cm in diameter); and large (above 1.2 cm in diameter), adopting the grading scale from 1 to 5, in which grade 1 = plants with less than 1% of leaves with small lesions; grade 2 = plants with 2% to 4% of leaves with lesions; grade 3 = plants with 5% to 19% of leaves with lesions (small, medium and large); grade 4 = plants with 20% to 35% of leaves with lesions (small, medium, and large); and grade 5 = plants with 36 to 100% of leaves with lesions (small, medium, and large) (Andreazi et al., 2015).

The evaluation of mealybug (*Coccus viridis*) resistance was also performed using a grading scale ranging from 1 to 5, in which grade 1 = absence of mealybugs; grade 2 = presence of a few individuals per plant; grade 3 = weak infestation (isolated females, colonies or nymphs); grade 4 = moderate infestation (presence of postures, 1st and 2nd instars on some branches); and grade 5 = strong infestation (presence of postures and all stages of development on the organs of all or almost all branches) (adapted from Andrade et al., 2017).

The frequency distribution of traits related to early development and biotic stresses was demonstrated through histograms for each trait, considering the 251 genotypes evaluated in the genome-wide association analysis. The histogram was built by the "ggplot2" package of the RStudio *software* (RStudio Team, 2022).

Analysis of Phenotypic Data

The phenotypic data were corrected for experimental design (augmented Federer blocks) using the linear mixed models (MLM) methodology, through the RStudio

software using the “sommer” package, the BLUP values (best unbiased linear prediction) and the residual averages. In this model we considered the kinship matrix H for Munoz method (Amadeu et al. 2016) and in this step we find the value of the phenotype corrected for environmental effect with the estimated genetic value (BLUP) plus the residue.

For Federer augmented blocks where there is no repetition of the genotypes being evaluated, we used a dataset that contains the phenotypic data for all eight traits that were measured in this experimental design. In this model we were able to obtain the BLUP values for the genotypes that are not replicated in the field. The presence of control genotypes (commercial clones) allows the adjustment of non-replicated genotypes.

$$y = Xf + Zg + Sb + e \quad (1)$$

where:

y: phenotypes data vector;

f: vector of the mean effect (considered fixed);

g: vector of the genotypic effects of the genotypes (considered random) being $g \sim N(0, H\sigma_g^2)$ where H is a kinship matrix including genotyped and non-genotyped individuals according to Legarra (2009) and Munoz (2014) e σ_g^2 is the additive genetic variance;

b: vector of the environmental effects of the blocks (considered random) being $b \sim N(0, I\sigma_b^2)$ what σ_b^2 is the variance associated with blocks;

e: vector of residual effects (random) being $e \sim N(0, I\sigma_e^2)$ what σ_e^2 is the residual variance. The capital letters X, Z and S represent the incidence matrices of these effects.

The phenotype corrected for design was obtained by means of: $y* = y - X\hat{f} - S\hat{b}$.

Linkage disequilibrium (LD)

Pairwise linkage disequilibrium (LD) of the 251 *Coffea canephora* genotypes were estimated using the square allele frequency correlations (r^2) for the DArTseq SNPs markers. This was done using LD measured by RStudio software. The r^2 was estimated by LD sliding window size of 50000 (bp) and an r^2 threshold set at 0.1. The distribution pattern of the entire LD genome was visualized using graphs generated by the "ggplot2" package by the RStudio software (RStudio Team, 2022).

Genomic Wide Association Study (MLM)

From the equation above (1), we used only the corrected values of the genotyped individuals to perform the GWAS study through the equation model described below (2). The BLUP and residual values were used for further analysis in the RStudio *software* through the "sommer" package to find the "p" values (p-value) to determine which SNP markers had a significant association for each trait that was evaluated. From the "p" values (p-value <0.05) Manhattan plots were generated using the "qqman" package of the RStudio *software* (RStudio Team, 2022). The quantile-quantile (QQ) plot was also generated to assess how well the model adopted in GWAS for this study.

$$y^* = X\beta + M_i\, m_i + e \quad (2)$$

where:

y^*: vector of phenotypes adjusted by the model described by equation 1;

β: vector of the general mean and 2 first principal components of the genomic kinship matrix (fixed effects) with incidence matrix X;

mi: fixed effect of the i-th marker;

$M_{i:}$ incidence matrix of the i-th marker;

e: vector of residual effects (random) being $e \sim N(0, I\sigma_e^2)$ what σ_e^2 is the residual variance.

Pearson's genetic correlation tests were then performed using the BLUPs to verify the correlation between the seven traits of early development and stress tolerance. Correlation distributions were plotted using the "ggplot2" package in *software* RStudio. From the BLUP values for each trait evaluated, boxplots were constructed for the marker with the lowest p-value to verify the behavior of these genotypes.

Candidate Genes

After verifying the SNPs with significant association, a search was performed on the NCBI (*National Center for Biotechnology Information*) website, in the available genome reference of *C. canephora* (Denoeud et al. 2014). Through the positions (pb) of the 115 SNPs and the chromosomes of *C. canephora*, SNPs inserted in or close to candidate genes related to the evaluated traits that had the function related to the trait were selected. The "downstream" and "upstream" distances were considered in the search for

candidate genes considering the shortest distance from the closest flanking gene to the SNP with significant association.

The distribution of SNPs along the chromosomes of *C. canephora* was verified by the online map Gene2 Chrom web v2 (http://mg2c.iask.in/mg2c_v2.0/) (Jiangtao et al., 2015).

3. Results

The distribution of the phenotypic characteristics evaluated in the population are represented in the histograms (Figures 1 and 2). Characteristics related to vegetative development present a normal distribution typical of quantitative characteristics (Figure 1). The stronger the color of the bar (pink) the more plants have that specific measure, and the green bars are those plants that are at the extremes of the charts. On the other hand, traits related to biotic stresses showed different behavior, due to the classification by scale (Figure 2). For the rust characteristic, we had more than 100 plants with a score equal to 2. For the leaf miner and cochineal trait, most plants obtained a score of 1 (Figure 2B). Generally, the plants presented lower values for all the evaluated stresses, demonstrating the potential of these plants in terms of tolerance to biotic stresses. For the drought tolerance trait, in some genotypes they were tolerant, with a score of 1 and others showed permanent damage with a score of 9 (Figure 2A).

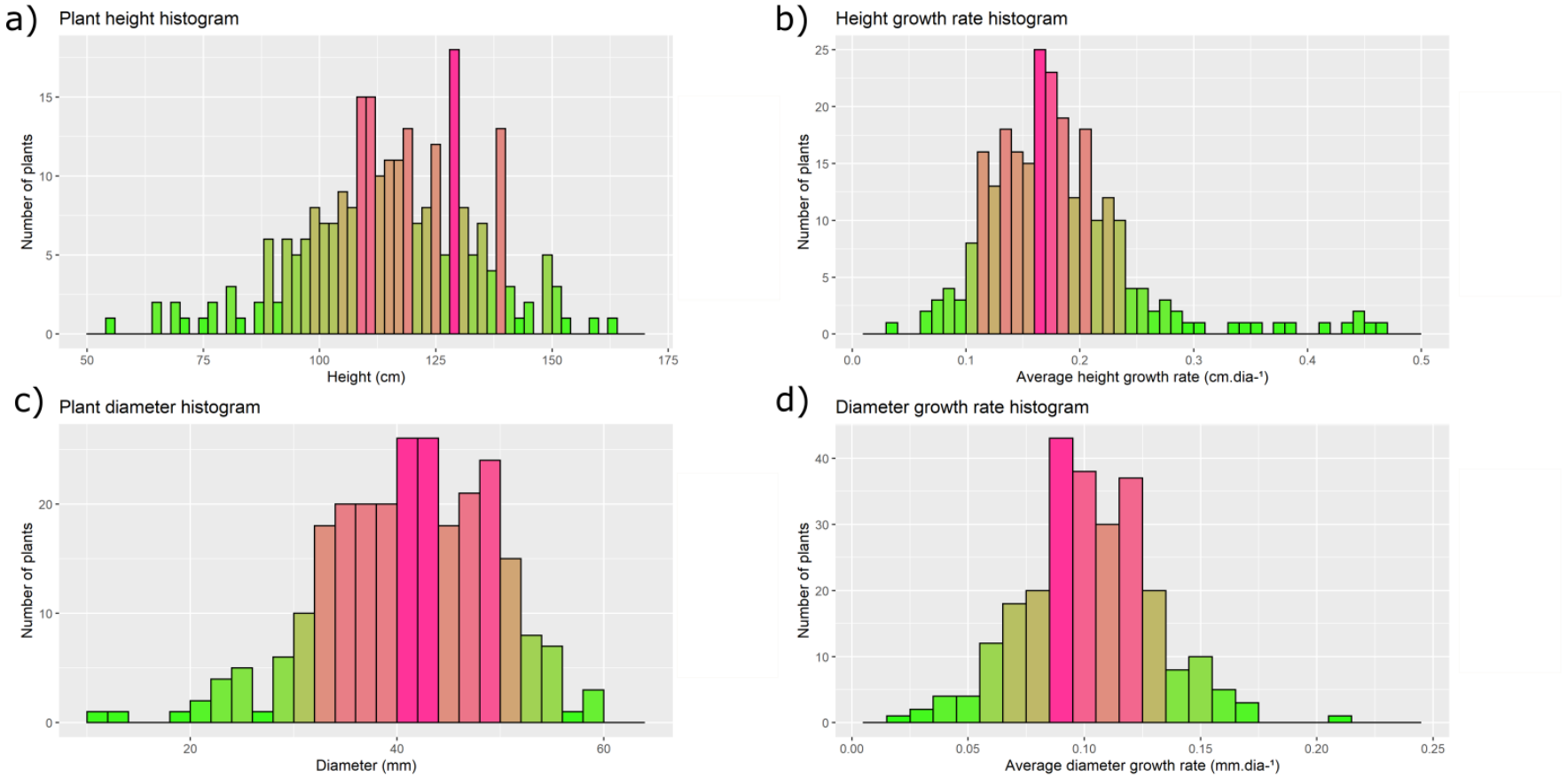

Figure 1. Histograms of phenotypic values distributions of plant height (cm), stem diameter (mm), height growth rate histogram (cm.dia-¹) and diameter growth rate histogram (mm.dia-¹) of the 251 genotypes of *Coffea canephora* of genome-wide association analysis.

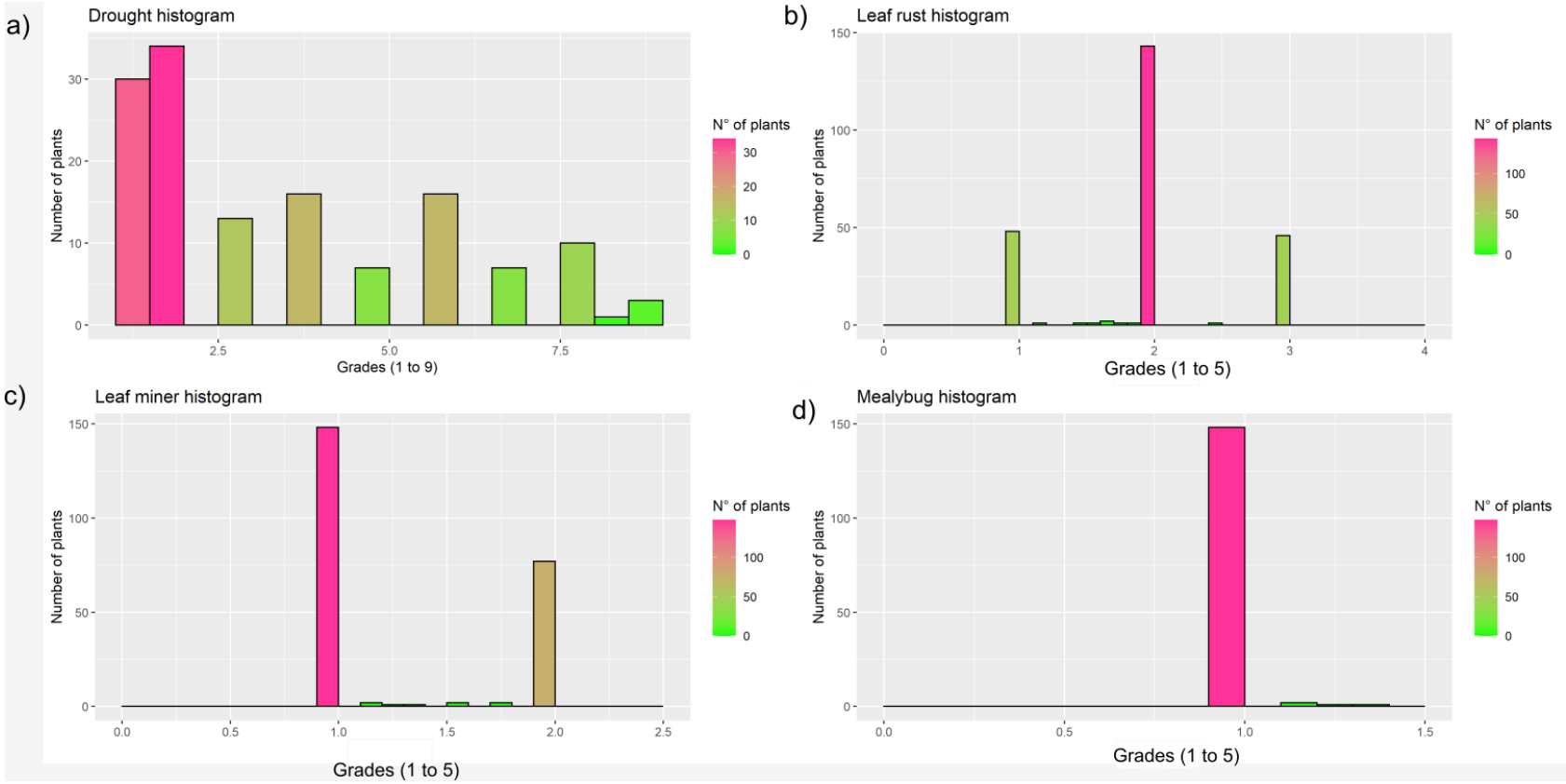

Figure 2. Histograms of phenotypic values distributions of leaf rust, leaf miner and mealybug of the 251 genotypes of *Coffea canephora* of genome-wide association analysis.

Quality analysis across parameters (MAF of 0.01, *call rate* ≥ 74%, monomorphic loci and reproducibility of 98.5%) resulted in 2,542 SNP markers out of the total of 9,491 SNPs identified. In the association analysis of the 2,542 SNPs, 115 showed a significant association for the evaluated characteristics (p-value <0.05) (Table 2). Of the 115 SNPs with significant associations found in this study, 62 were inserted into genes.

The genetic correlation between traits ranged from -0.23 to 0.83 (Figure 3). The lowest positive correlation between early developmental traits was between average height growth rate and stem diameter (0.20). The leaf miner incidence had a negative correlation for all characteristics related to early development. The incidence of rust had a negative correlation for the characteristics related to diameter and positive for the characteristics related to plant height.

The highest values of positive correlations were height and diameter (0.57), height and average growth rate in height (0.42), height and average growth rate in diameter (0.34) and average height of growth height and mean growth in diameter (0.33).

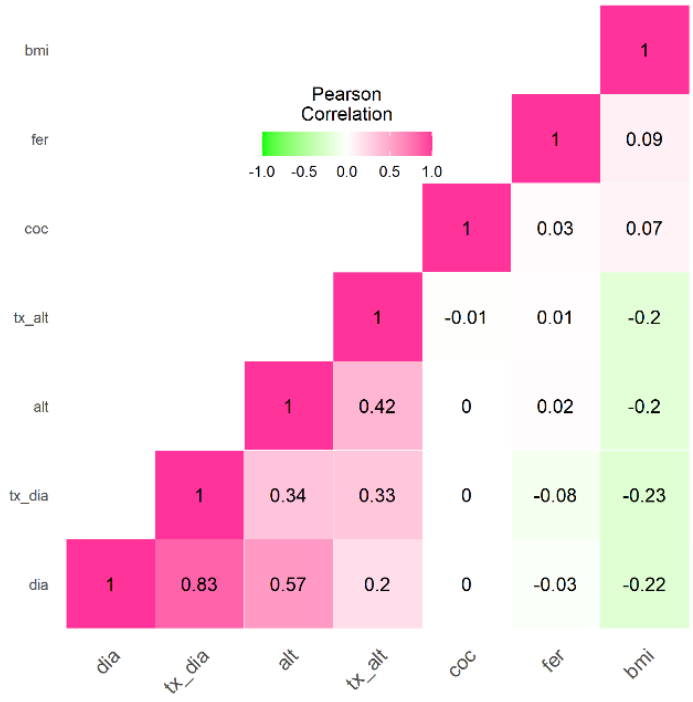

Figure 3. Pearson's genotypic correlation through BLUP values between phenotypic traits that showed a significant association for early development and stress tolerance among the 251 *Coffea canephora* genotypes. alt = height; dia = diameter; tx_alt = average height growth rate; tx_dia = average growth rate in diameter; fer = leaf rust; bmi = leaf miner; coc = mealybug.

The LD analysis was performed from the r^2 (correlation coefficient squared between the allele frequencies at each of the two *loci*) and this paired analysis among the 2,542 SNP markers detected in this analysis generated 444,222 comparisons, of which 54,101 (13.25 %) had an r^2 equal to or greater than 0.1, while a total of 34,237 (8.39%) had an r^2 equal to or greater than 0.5 (Figure S6, Figure S7 and Table 1). The whole genome distribution pattern can be seen in Figure S4 and S5. Some markers were in high LD to perfect LD ($r^2 = 1$), which are more likely to be inherited together. 547 of the total comparisons were in perfect connection, while 177 were completely unlinked ($r^2 = 0$).

Table 1. Genome-wide Pairwise LD distribution for the SNP markers among the 251 genotypes of *Coffea canephora* based on the chromosomes at threshold r^2 equal, or above 0.1 and r^2 equal, or above 0.5

	Pairwase SNP LD			**Percentage**	
Chromosome	**Total pairs**	**$r^2 \geq$ 0.1**	**$r^2 \geq$ 0.5**	**0.1 (%)**	**0.5 (%)**
1	23220	3459	2516	14.90	10.84
2	12246	1847	1172	15.08	9.57
3	10153	1450	964	14.28	9.49
4	58311	10290	4991	17.65	8.56
5	8256	977	1078	11.83	13.06
6	10153	1753	1223	17.27	12.05
7	7626	635	874	8.33	11.46
8	27028	5528	2490	20.45	9.21
9	19701	2728	1184	13.85	6.01
10	7140	757	1330	10.60	18.63
11	2346	199	472	8.48	20.12
12*	222111	24478	15943	11.02	7.178
Total	408291	54101	34237	13.25	8.39

**Coffea canephora* genome *scaffolds*

The eigenvalues of main components indicated that the first component accounted for most of the variation within the population that contributes more than 55.16% of the total variation and the second principal component was responsible for 3.4% of the total variation. There was the formation of a large group with most genotypes and two other smaller groups with few genotypes (Figure 4).

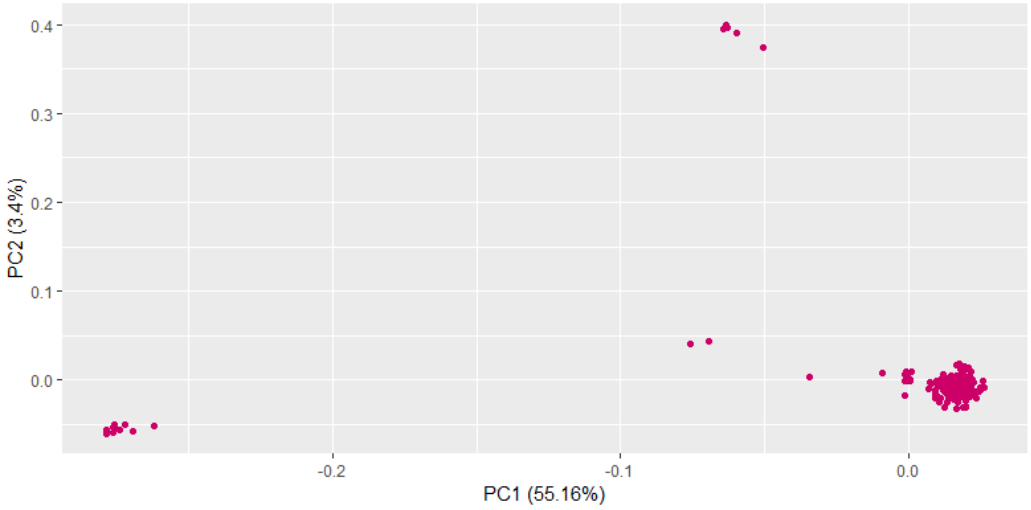

Figure 4. PCA of *Coffea canephora* from SNP markers illustrating the genetic divergence between genotypes, with the two main components explaining 58.56% of the total variability.

Of the eight characteristics evaluated, seven had SNPs with significant associations. For drought tolerance no significant association was detected. The height growth rate presented the highest number of SNPs with significant association (48) which are distributed in eight chromosomes (Figure 5B). For mealybug incidence 20 SNPs with significant association in seven chromosomes (Figure 6C). For height, seven SNPs on six chromosomes (Figure 5A); for stem diameter, 11 SNPs on nine chromosomes (Figure 5C); for diameter growth rate, seven SNPs on four chromosomes (Figure 6D). The rust characteristic with 16 SNPs (Figure 6A) and leaf miner incidence with five SNPs (Figure 6B). Quantile-quantile (QQ) plots were also generated to assess how well the model adopted in GWAS for this study (Figure 9).

The 11 chromosomes of *C. canephora* with the SNPs with significant association were shown in the Figure 8. Chromosome 1 had the highest number of significant SNPs (28), followed by chromosome 9 with 16 (Figure 8). Chromosomes 10 and 7 had the lowest number of SNPs with significant associations, three and four respectively. The dots represent which characteristic is associated with a particular SNP. Markers that showed significance for more than one feature are marked with two colored dots.

Four SNPs had significant association for more than one characteristic (Figure 8 and Table 2). The SNP markers "100064936.13.T.C" and "100074576.62.T.C" (chromosome 8) had a significant association for the mealybug incidence (green dot) and

for height growth rate (red dot). The SNP marker “100064019.44.G.T” (chromosome 9) also had a significant association for these two traits (green and red dots). The SNP marker “100064293.27.T.C” (chromosome 1) had a significant association for the rust incidence trait (black dot) and for the diameter growth rate trait (purple dot). Markers that showed significance for more than one feature are marked with two colored dots (totaling 20 SNPs).

Of the 48 SNPs identified for height growth rate, 35 are embedded in or close to genes. For height, of the seven SNPs detected, five are inserted in or close to genes. Of the five SNPs detected for the leaf miner infestation, two SNPs are inserted into genes. The mealybug infestation, which had 20 SNPs with significant association, has 18 SNPs inserted into genes. Of the 10 SNPs found for the stem diameter, 6 are inserted into or close to the genes. For the rust incidence trait, 14 SNPs inserted in or close to genes were found out of a total of 16 SNPs with significant association. And, of the seven SNPs found for the diameter growth rate, 4 are embedded in or close to genes (Table 2).

Some candidate genes with associates SNP showed functions related to the traits that were evaluated. *GSCOC_T00019303001*, *GSCOC_T00022693001*, *GSCOC_T00039643001* and *GSCOC_T00040251001* genes are related to plant defense mechanisms to protect against pathogens, pests and abiotic stresses. Other candidate genes such as genes *GSCOC_T00040077001*, *GSCOC_T00028217001* and *GSCOC_T00021883001* presented putative functions related to plant development and plant hormones.

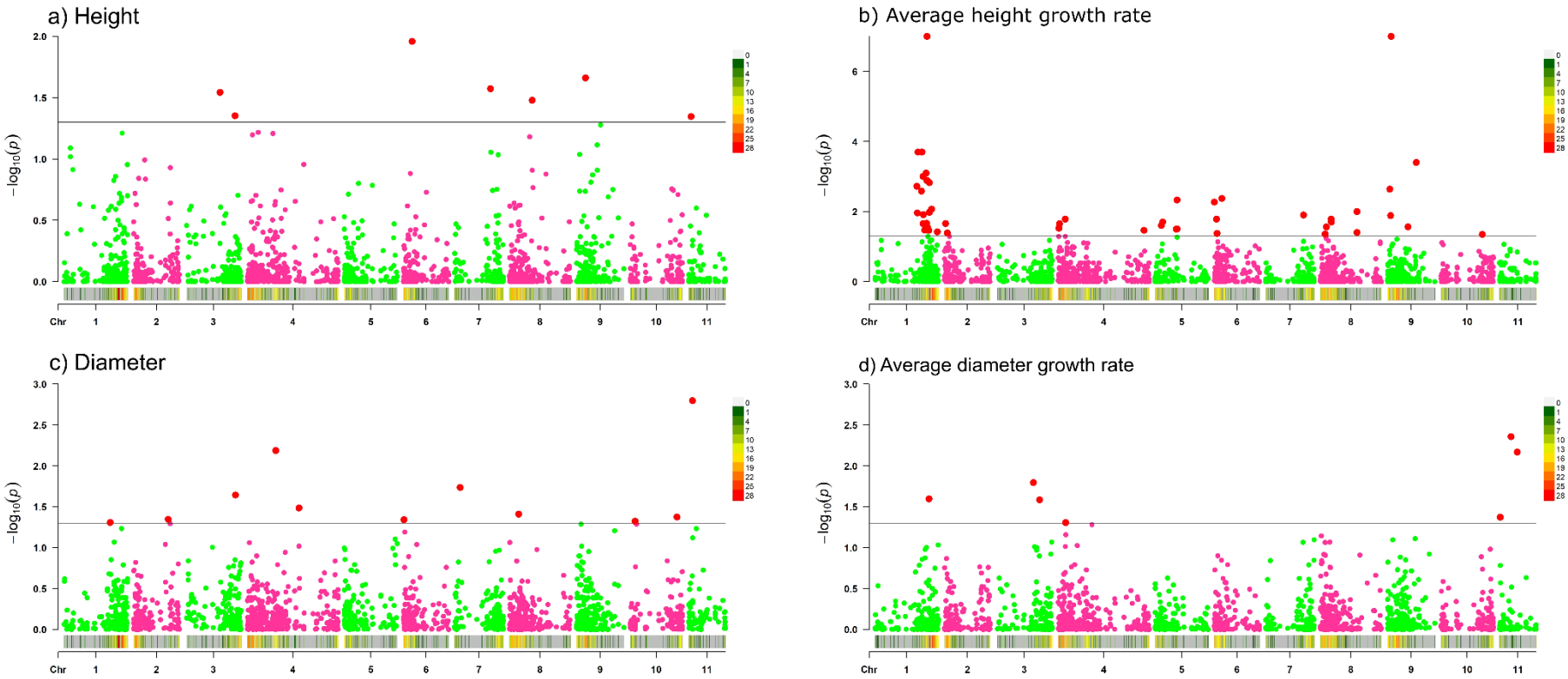

Figure 5. Manhattan plots of the genomic-wide association of single nucleotide polymorphisms (SNPs) of four traits related to early development of 251 *Coffea canephora* genotypes. The -log10(p) threshold of 0.05 (dashed line) was used to select SNPs and identify candidate genes. Significant SNPs are represented by red dots above the threshold line. The distribution of SNPs in the chromosomes is represented by the bar below each chromosome, the red colors indicate higher amounts of SNPs, the yellow colors indicate average amounts of SNPs and the green colors indicate lower amounts of SNPs.

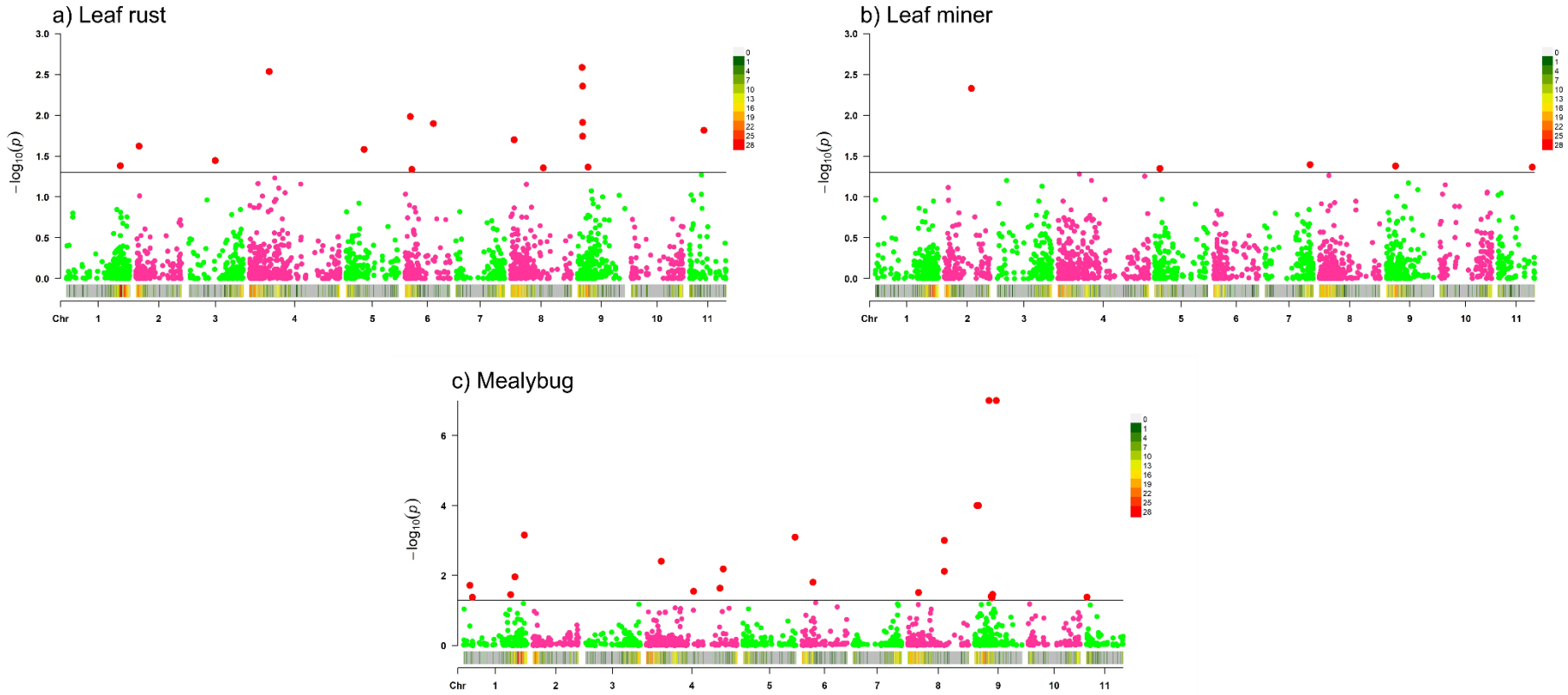

Figure 6. Manhattan plots of the genomic-wide association of single nucleotide polymorphisms (SNPs) of four traits related to biotic stresses of 251 *Coffea canephora* genotypes. The -log10(p) threshold of 0.05 (dashed line) was used to select SNPs and identify candidate genes. Significant SNPs are represented by red dots above the threshold line. The distribution of SNPs in the chromosomes is represented by the bar below each chromosome, the red colors indicate higher amounts of SNPs, the yellow colors indicate average amounts of SNPs and the green colors indicate lower amounts of SNPs.

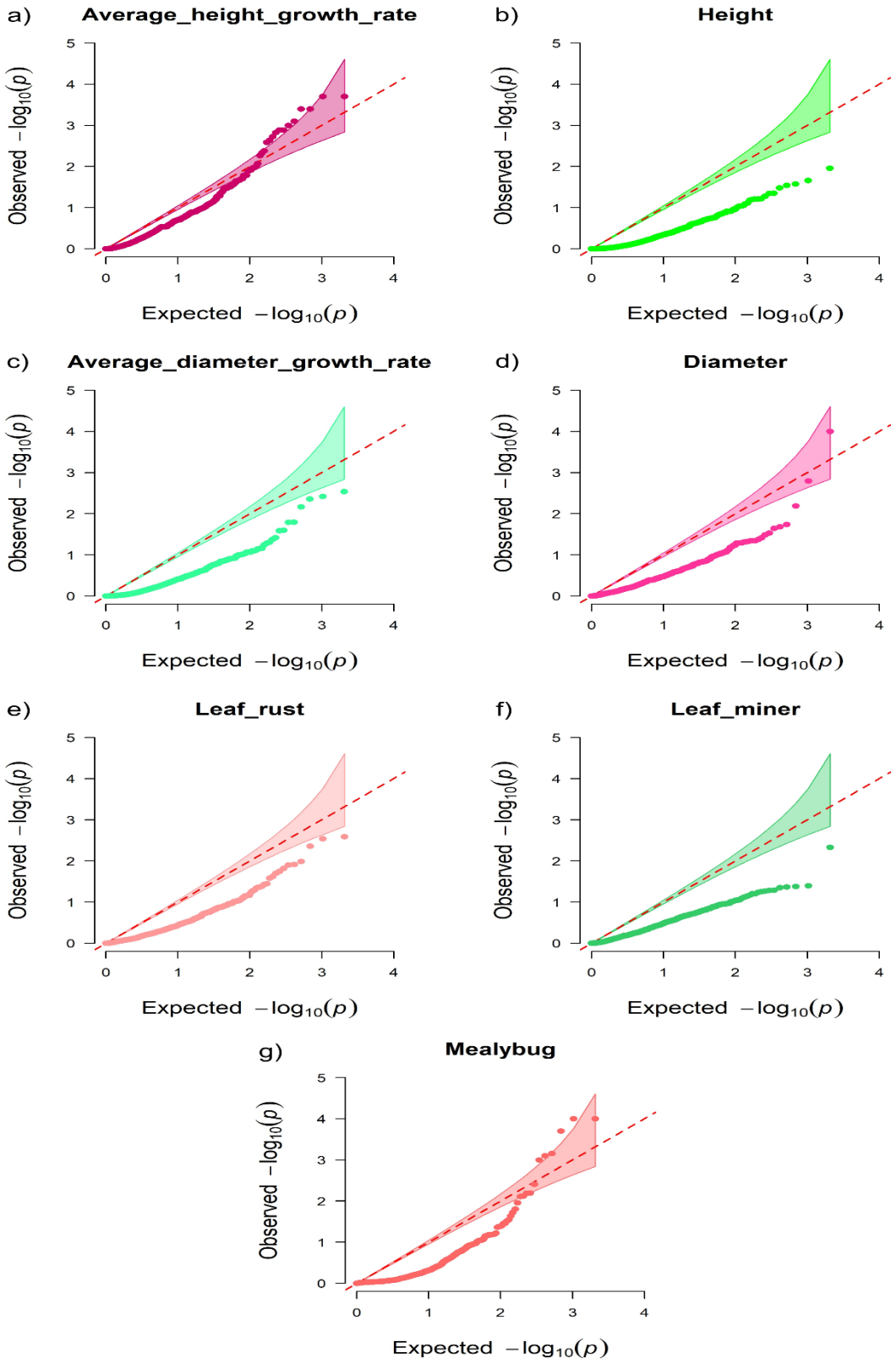

Figure 7. Quantile–quantile (QQ) plots referring to the seven characteristics evaluated demonstrating how the model was adjusted for a given characteristic.

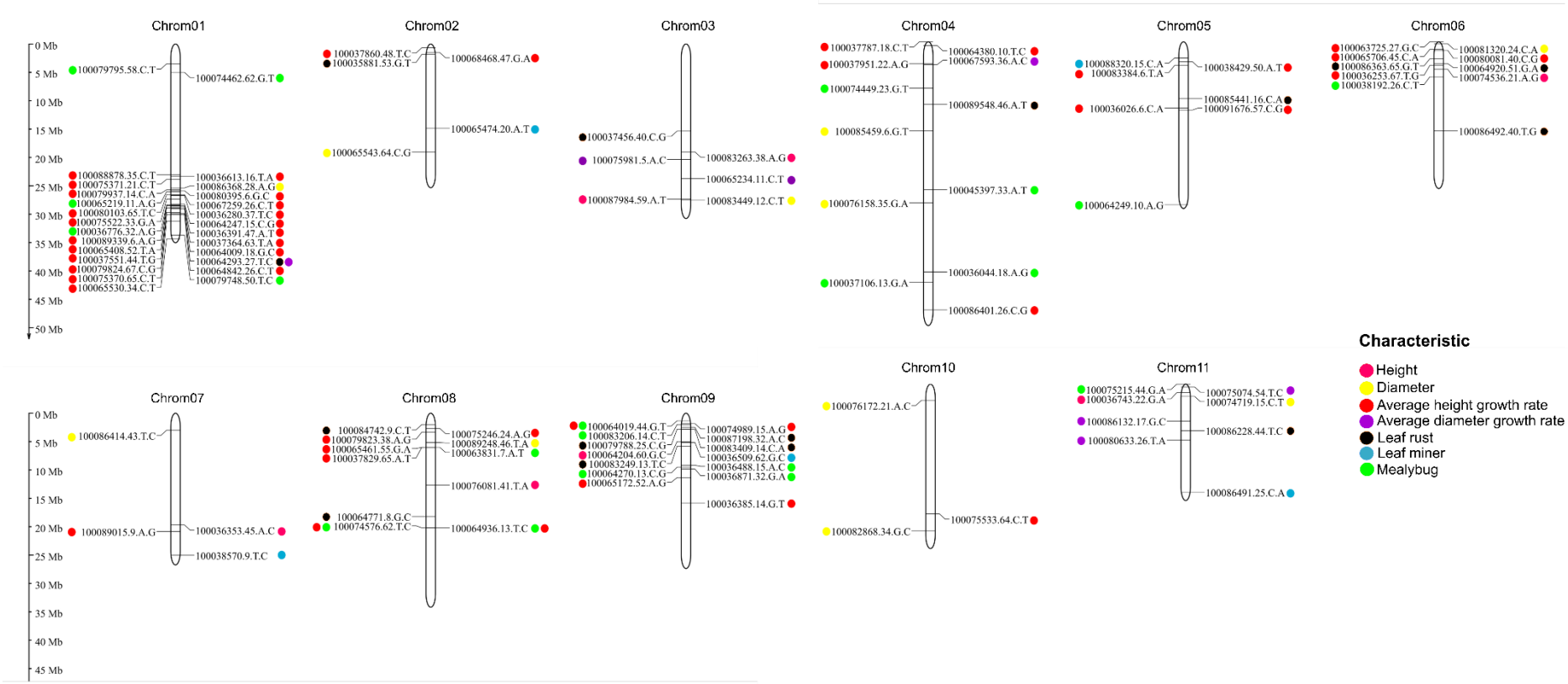

Figure 8. SNP markers with significant association (115) distributed along the 11 chromosomes of *Coffea canephora*. Each colored dot refers to a specific characteristic. pink = height; yellow = diameter; red = average height growth rate; purple = average growth rate in diameter; black = leaf rust; blue = leaf miner and green = mealybug.

Table 2. Positions of single nucleotide polymorphism (SNP) markers with the lowest values of −log10 (p) and *Coffea canephora* genes overlapping or close to them, potentially associated with the evaluated traits (height, diameter, average growth rate in diameter, average growth rate in height, rust tolerance, mealybug incidence, leaf miner incidence).

Characteristic	SNP marker	Chr[a]	Genome position (bp)	p-value	Candidate gene[b]	Distance[c]	Putative function
Height	100087984.59.A.T	3	29,808,608	0.0445	GSCOC_T00032264001	+174	RING_Ubox super family
Height	100083263.38.A.G	3	20,789,026	0.0286	GSCOC_T00020497001	+76	Transmembrane 9 family protein
Height	100074536.21.A.G	6	5,236,036	0.0110	GSCOC_T00022491001	+95	nsLTP2 domain-containing protein
Height	100036353.45.A.C	7	21,452,163	0.0267	-	-	-
Height	100076081.41.T.A	8	13,754,359	0.0332	-	-	-
Height	100064204.60.G.C	9	5,378,240	0.0218	GSCOC_T00040077001	within	PSK domain-containing protein
Height	100036743.22.G.A	11	1,647,113	0.0450	GSCOC_T00017737001	within	Thomboid family protein
Leaf miner	100065474.20.A.T	2	16,164,468	0.0047	-	-	-
Leaf miner	100088320.15.C.A	5	3,142,823	0.0449	-	-	-
Leaf miner	100038570.9.T.C	7	27,303,009	0.0402	-	-	-
Leaf miner	100036509.62.G.C	9	5,654,362	0.0419	GSCOC_T00040122001	within	Hexokinase-like
Leaf miner	100086491.25.C.A	11	20,673,337	0.0432	GSCOC_T00007638001	within	Zinc finger, C3HC4 type (RING finger) / PspA_IM30 super family
Mealybug	100079795.58.C.T	1	3,898,518	0.0191	GSCOC_T00019303001	within	Germin family protein
Mealybug	100074462.62.G.T	1	5,376,425	0.0421	GSCOC_T00013211001	within	Formate--tetrahydrofolate ligase
Mealybug	100079748.50.T.C	1	36,705,213	0.0007	GSCOC_T00015989001	-46	PPR_3 super Family / PPR_2 super family
Mealybug	100065219.11.A.G	1	28,455,365	0.0352	GSCOC_T00024047001	within	DNA_BRE_C super family
Mealybug	100036776.32.A.G	1	31,007,192	0.0110	-	-	-
Mealybug	100036044.18.A.G	4	44,260,989	0.0231	GSCOC_T00016697001	+171	PLN00156 family protein
Mealybug	100045397.33.A.T	4	28,354,539	0.0283	GSCOC_T00018180001	within	S10 family peptidase
Mealybug	100074449.23.G.T	4	8,998,755	0.0039	GSCOC_T00029464001	within	SLC5-6-like_sbd super family
Mealybug	100037106.13.G.A	4	46,184,049	0.0065	GSCOC_T00015399001	+394	-

Mealybug	100064249.10.A.G	5	31,259,098	0.0008	GSCOC_T00026848001	within	Glycosyltransferase_GTB-type super family
Mealybug	100038192.26.C.T	6	6,697,667	0.0156	GSCOC_T00022693001	within	G-type lectin S-receptor-like serine / threonine-protein kinase
Mealybug	100063831.7.A.T	8	6,534,208	0.0307	GSCOC_T00031170001	within	CUE and DUF460 domain-containing protein
Mealybug	**100064936.13.T.C**	8	22,051,563	0.0076	GSCOC_T00022869001	within	Ankyrin repeat domain-containing protein
Mealybug	**100074576.62.T.C**	8	22,051,629	0.0010	GSCOC_T00022869001	within	Ankyrin repeat domain-containing protein
Mealybug	100083206.14.C.T	9	2,236,131	0.0001	GSCOC_T00039555001	+609	PTZ00265 super family
Mealybug	100064270.13.C.G	9	10,367,632	0.0433	GSCOC_T00036697001	within	PLN02586 super family
Mealybug	**100064019.44.G.T**	9	1,489,438	0.0001	GSCOC_T00039398001	within	PLN03014 super family
Mealybug	100036488.15.A.C	9	9,964,738	0.0397	GSCOC_T00036763001	within	TPR_12 / PRK15331 super family
Mealybug	100036871.32.G.A	9	10,799,979	0.0348	GSCOC_T00036653001	within	-
Mealybug	100075215.44.G.A	11	481,190	0.0414	-	-	-
Diameter	100086368.28.A.G	1	27,721,374	0.0492	-	-	-
Diameter	100065543.64.C.G	2	20,833,088	0.0451	GSCOC_T00031874001	within	Trm5 super family
Diameter	100083449.12.C.T	3	29,972,245	0.0227	GSCOC_T00032235001	within	DUF1666 domain-containing protein
Diameter	100085459.6.G.T	4	17,110,256	0.0065	GSCOC_T00014444001	within	DJ-1/PfpI family protein
Diameter	100076158.35.G.A	4	30,896,226	0.0328	-	-	-
Diameter	100081320.24.C.A	6	371,175	0.0455	GSCOC_T00021661001	+159	PLN02785 super family
Diameter	100086414.43.T.C	7	3,297,187	0.0184	GSCOC_T00018230001	within	PLN00113 super family
Diameter	100089248.46.T.A	8	5,635,124	0.0390	GSCOC_T00031312001	within	PLN02805 super family
Diameter	100076172.21.A.C	10	3,103,119	0.0475	-	-	-
Diameter	100082868.34.G.C	10	28,134,955	0.0423	-	-	-
Diameter	100074719.15.C.T	11	2,372,421	0.0016	-	-	-
Leaf rust	**100064293.27.T.C**	1	32,515,469	0.0416	-	-	-
Leaf rust	100035881.53.G.T	2	1,980,982	0.0239	GSCOC_T00024764001	within	glucose-1-phosphate adenylyltransferase
Leaf rust	100037456.40.C.G	3	16,668,913	0.0358	GSCOC_T00034053001	+827	-
Leaf rust	100089548.46.A.T	4	11,989,185	0.0029	GSCOC_T00029089001	within	DnaJ-class molecular chaperone with C-terminal Zn finger domain

Leaf rust	100085441.16.C.A	5	10,969,635	0.0262	-	-	-
Leaf rust	100086363.65.G.T	6	3,257,238	0.0104	GSCOC_T00022154001	-589	MBD domain-containing protein
Leaf rust	100086492.40.T.G	6	17,173,187	0.0126	GSCOC_T00012379001	-175	Rrp4 super family
Leaf rust	100064920.51.G.A	6	4,189,010	0.0462	GSCOC_T00022322001	within	Asp-tRNAAsn/Glu-tRNAGln amidotransferase A subunit or related amidase
Leaf rust	100064771.8.G.C	8	19,812,110	0.0441	GSCOC_T00028935001	within	HAD family hydrolase
Leaf rust	100084742.9.C.T	8	2,219,690	0.0199	-	-	-
Leaf rust	100079788.19.G.C	9	3,053,951	0.0122	GSCOC_T00039694001	+876	STKc_IRAK domain-containing protein / PKc_like super family / Malectin_like super family
Leaf rust	100083409.14.C.A	9	3,054,017	0.0044	GSCOC_T00039694001	+932	STKc_IRAK domain-containing protein / PKc_like super family / Malectin_like super family
Leaf rust	100087198.32.A.C	9	2,710,981	0.0026	GSCOC_T00039643001	+77	CaM_binding domain-containing protein
Leaf rust	100079788.25.C.G	9	3,053,951	0.0181	GSCOC_T00039694001	+876	STKc_IRAK domain-containing protein / PKc_like super family / Malectin_like super family
Leaf rust	100083249.13.T.C	9	6,320,796	0.0432	GSCOC_T00040251001	within	STKc_IRAK / PKc_like super family / GUB_WAK_bind / WAK_assoc
Leaf rust	100086228.44.T.C	11	8,902,368	0.0153	GSCOC_T00036271001	-308	SEC14 family lipid-binding protein
Average growth rate in height	100037364.63.T.A	1	31,246,273	0.0013	GSCOC_T00028217001	within	dof zinc finger protein
Average growth rate in height	100064842.26.C.T	1	32,724,024	0.0015	GSCOC_T00028476001	within	family protein is a DUF640 domain-containing protein that acts as a plant homeotic and developmental regulator
Average growth rate in height	100065408.52.T.A	1	31,431,302	0.0013	GSCOC_T00028261001	+289	protein kinase family protein
Average growth rate in height	100079824.67.C.G	1	32,710,237	0.0105	GSCOC_T00028475001	within	rab3 GTPase-activating protein catalytic subunit
Average growth rate in height	100080395.6.G.C	1	28,294,448	0.0002	-	-	-
Average growth rate in height	100088878.35.C.T	1	25,196,040	0.0019	GSCOC_T00030165001	within	F-box/kelch-repeat protein
Average growth rate in height	100089339.6.A.G	1	31,108,237	0.0216	GSCOC_T00028187001	-214	histone H4
Average growth rate in height	100036391.47.A.T	1	31,075,131	0.0339	GSCOC_T00028181001	within	ADF_gelsolin super family
Average growth rate in height	100080103.65.T.C	1	28,908,232	0.0010	GSCOC_T00024105001	within	PLN02518 family protein

Average growth rate in height	100064247.15.C.G	1	30,769,382	0.0008	GSCOC_T00028127001	within	HAD_like super family
Average growth rate in height	100075370.65.C.T	1	34,070,883	0.0085	-	-	-
Average growth rate in height	100075371.21.C.T	1	25,943,867	0.0002	-	-	-
Average growth rate in height	100036613.16.T.A	1	25,509,292	0.0110	GSCOC_T00009382001	within	PLN02342 family protein
Average growth rate in height	100079937.14.C.A	1	28,019,911	0.0026	GSCOC_T00023992001	within	Peptidases_S8_3 / fn3_6 / PA_subtilisin_like / Inhibitor_I9 / AprE
Average growth rate in height	100067259.26.C.T	1	28,900,964	0.0219	GSCOC_T00024104001	within	-
Average growth rate in height	100075522.33.G.A	1	29,897,383	0.0340	GSCOC_T00024234001	+144	Sua5/YciO/YrdC/YwlC family protein
Average growth rate in height	100064009.18.G.C	1	31,599,843	0.0277	-	-	-
Average growth rate in height	100065530.34.C.T	1	37,418,320	0.0378	-	-	-
Average growth rate in height	100037551.44.T.G	1	32,388,725	0.0341	GSCOC_T00028424001	within	STKc_IRAK domain-containing protein
Average growth rate in height	100036280.37.T.C	1	29,167,076	0.0123	GSCOC_T00024150001	within	HMA domain-containing protein
Average growth rate in height	100068468.47.G.A	2	1,687,561	0.0408	GSCOC_T00024722001	within	TFIIF_beta super family
Average growth rate in height	100037860.48.T.C	2	700,031	0.0221	GSCOC_T00024545001	within	SLC5-6-like_sbd super family
Average growth rate in height	100086401.26.C.G	4	51,432,427	0.0342	-	-	-
Average growth rate in height	100037787.18.C.T	4	429,680	0.0298	GSCOC_T00019839001	within	PTZ00265 super family
Average growth rate in height	100064380.10.T.C	4	753,759	0.0224	GSCOC_T00019890001	+177	IQ and DUF4005 domain-containing protein
Average growth rate in height	100037951.22.A.G	4	4,208,098	0.0165	GSCOC_T00039037001	within	NAD(P)-dependent glycerol-3-phosphate dehydrogenase
Average growth rate in height	100083384.6.T.A	5	4,474,459	0.0200	GSCOC_T00026300001	within	MFS transporter
Average growth rate in height	100038429.50.A.T	5	3,809,142	0.0251	GSCOC_T00026203001	within	CorA family magnesium transporter
Average growth rate in height	100091676.11.A.G	5	13,175,217	0.0315	-	-	-
Average growth rate in height	100091676.20.T.C	5	13,175,217	0.0315	-	-	-
Average growth rate in height	100091676.46.T.C	5	13,175,217	0.0047	-	-	-
Average growth rate in height	100091676.57.C.G	5	13,175,217	0.0315	-	-	-
Average growth rate in height	100036026.6.C.A	5	12,669,938	0.0319	GSCOC_T00031005001	within	Surp and CTD_bind domain-containing protein
Average growth rate in height	100036253.67.T.G	6	4,529,682	0.0042	GSCOC_T00022370001	within	Macoilin super family
Average growth rate in height	100063725.27.G.C	6	7,429	0.0054	GSCOC_T00021599001	within	PEP_TPR_lipo super family
Average growth rate in height	100080081.40.C.G	6	1,676,433	0.0422	GSCOC_T00021883001	within	dof zinc finger protein
Average growth rate in height	100065706.45.C.A	6	1,387,013	0.0165	-	-	-

Average growth rate in height	100089015.9.A.G	7	22,701,270	0.0126	GSCOC_T00016805001	within	C1 family peptidase
Average growth rate in height	100075246.24.A.G	8	2,946,700	0.0438	GSCOC_T00023678001	within	DnaJ super family / PEP_TPR_lipo super family
Average growth rate in height	100079823.38.A.G	8	3,674,494	0.0277	-	-	-
Average growth rate in height	**100064936.13.T.C**	8	22,051,563	0.0100	GSCOC_T00022869001	within	ankyrin repeat domain-containing protein
Average growth rate in height	100037829.65.A.T	8	6,627,944	0.0198	GSCOC_T00031160001	within	-
Average growth rate in height	100065461.55.G.A	8	6,520,970	0.0164	-	-	-
Average growth rate in height	**100074576.62.T.C**	8	22,051,629	0.0396	GSCOC_T00022869001	within	ankyrin repeat domain-containing protein
Average growth rate in height	100074989.15.A.G	9	1,957,693	0.0131	-	-	-
Average growth rate in height	**100064019.44.G.T**	9	1,489,438	0.0023	GSCOC_T00039398001	within	PLN03014 super family
Average growth rate in height	100065172.52.A.G	9	12,350,600	0.0272	GSCOC_T00036462001	within	solute carrier family 26 protein
Average growth rate in height	100036385.14.G.T	9	17,303,493	0.0004	GSCOC_T00018633001	within	ARGLU super family / AvrRxo1 super family
Average growth rate in height	100075533.64.C.T	10	24,876,278	0.0449	GSCOC_T00026964001	within	Ephrin_rec_like super family
Average growth rate in diameter	**100064293.27.T.C**	1	32,515,469	0.0253	-	-	-
Average growth rate in diameter	100065234.11.C.T	3	25,921,554	0.0261	GSCOC_T00032859001	within	PLN02318 super family
Average growth rate in diameter	100075981.5.A.C	3	22,212,357	0.0160	-	-	-
Average growth rate in diameter	100067593.36.A.C	4	4,416,224	0.0494	GSCOC_T00039070001	within	PLN03243 super family
Average growth rate in diameter	100080633.26.T.A	11	10,714,632	0.0068	GSCOC_T00003808001	within	GLTP domain-containing protein
Average growth rate in diameter	100075074.54.T.C	11	620,844	0.0425	-	-	-
Average growth rate in diameter	100086132.17.G.C	11	7,078,471	0.0044	GSCOC_T00011290001	-401	NB-ARC super family

a Chr, chromosome.

b Gene models were identified using the *Coffea canephora* genome.

c Distance (bp) from the closest gene model; within, lies within the candidate gene sequence; +, downstream; −, upstream.

After calculating the p-value, the q-value were calculated. For this analysis, two traits showed markers with significant association, mealybug incidence (5 SNPs) and average height growth rate (2 SNPs) (Table 3). The SNP marker “100083206.14.C.T” was significant for both traits. For these two characteristics (average height growth rate and mealybug), it can be observed that the quantile-quantile graphs presented a better fit to the model (Figures 7A and 7G).

Table 3. Q-values (q-value < 0.09) for two traits that showed SNPs with significant association for 251 *Coffea canephora* genotypes. Markers in bold refer to markers that also had significant p-values in the previous table

Characteristic	Allele ID	Genome position (bp)	q-value
Mealybug	100086321.25.G.C	13,051,391	0
Mealybug	100039204.33.G.C	8,637,761	0
Mealybug	100088083.67.G.T	0	0
Average height growth rate	**100083206.14.C.T**	2,236,131	0
Average height growth rate	100075121.66.T.C	31,314,897	0
Mealybug	100083206.14.C.T	2,236,131	0.05084
Mealybug	**100064019.44.G.T**	1,489,438	0.05084
Mealybug	100090110.40.G.A	2,193	0.08473

In total, 57 SNPs were found, with significant p-values for "chromosome 0" (*scaffolds*) of *C. canephora* (Table S8). Of these, the marker "100076552.7.G.A" was significant for the characteristics height (0.0348) and diameter (0.0541); the marker "100076012.33.C.T" was significant for the diameter (0.0538) and average diameter growth rate (0.0522) and the marker "100084770.47.G.A" was significant for the diameter (0.0460) and average diameter growth rate (0.0163).

The boxplots based on the effects of genetic values (BLUPs), in three *loci* for each characteristic related to vegetative development with significant associations (p-values), are shown in Figure 9. For the height of plants, we used the markers "X100083263.38.A.G" (0.0286) located on chromosome 3, "X100036353.45.A.C" (0.0267) located on chromosome 7 and "X100076081.41.T.A" (0.0332) located on chromosome 8 (Figure 10A and Table 4). The smallest amplitude of variation occurred for the marker "X100036353.45.A.C", for the alternative homozygote genotype (2), where we also had a greater effect when compared to the reference homozygote (0).

The average height growth rate was evaluated for the markers "X100079937.14.C.A" (0.0026), "X100067259.26.C.T" (0.0219) and "X100086401.26.C.G" (0.0342) (Figure 9B and Table 4). For the three markers evaluated, notice that the behavior of the genotypes was similar, only the heterozygous genotype (1) for the marker "X100067259.26.C.T" had a smaller effect when compared to the other genotypes, for the same marker (Figure 9B and Table 4). In the average, growth rate in diameter, the behavior of the genotypes, for the three evaluated markers "X100075981.5.A.C" (0.0160), "X100075074.54.T.C" (0.0425) and "X100086132.17.G.C " (0.0044), also showed a similar behavior, similar to what happens with the average height growth rate (Figure 10D and Table 4).

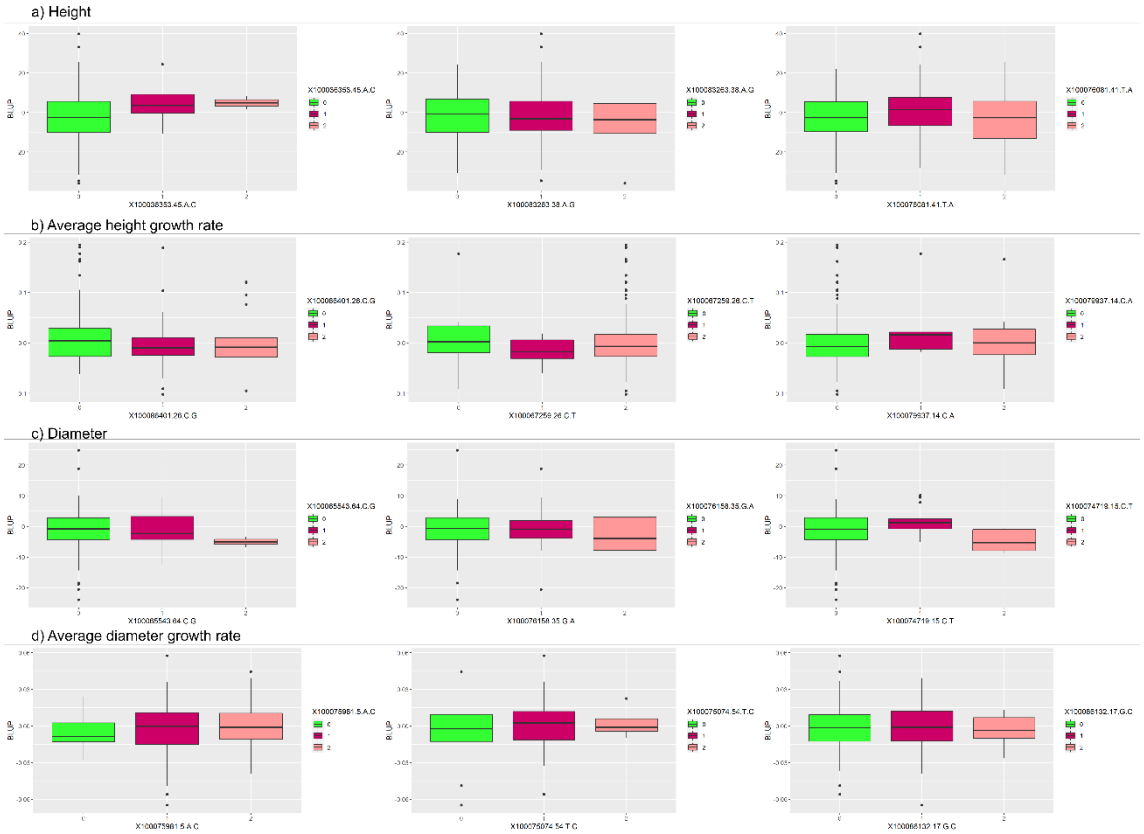

Figure 9. Boxplots for traits related to vegetative development for 251 *Coffea canephora* genotypes from BLUP values. The colors represent the genotypes of the formed groups where: green = reference homozygote (0); dark pink = heterozygote (1) and light pink = alternative homozygote (2).

For stem diameter, the selected markers were "X100065543.64.C.G" (0.0451) located on chromosome 2, "X100076158.35.G.A" (0.0328) located on chromosome 4,

and "X100074719.15.C.T" (0.0016) located on chromosome 11 (Figure 9C). For the markers "X100065543.64.C.G" and "X100074719.15.C.T" there was a greater effect (> BLUP) for the reference homozygous (0) and heterozygous (1) genotypes when compared to the alternative homozygous genotype (2).

In the rust trait, for the marker "X100079788.25.C.G" (0.0181) located on chromosome 9, there is a difference between the effects of BLUP values for reference homozygotes (0) and heterozygotes (1), when compared to the homozygote alternative (2), which presented a lower estimated genetic value. Also for this trait, we have the marker "X100086363.65.G.T" (0.0104) located on chromosome 6, which showed a different behavior from the previous marker, with the alternative homozygous genotype (2) showing higher BLUP values than the other genotypes (Figure 10A).

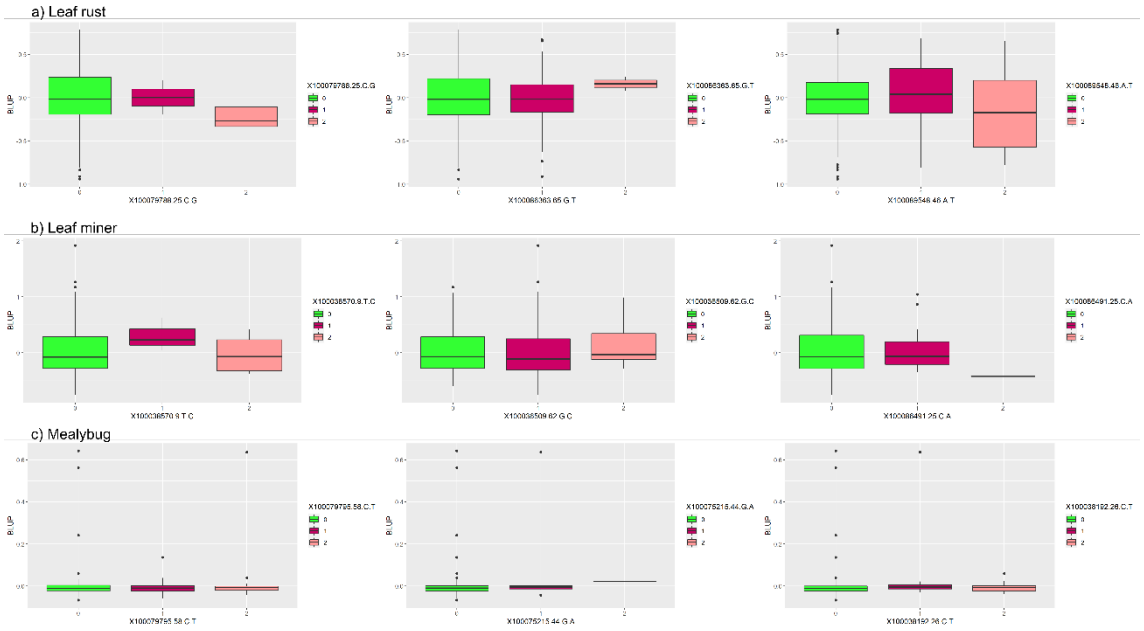

Figure 10. Boxplots for traits related to biotic stresses for 251 *Coffa canephora* genotypes from BLUP values. The colors represent the genotypes of the formed groups where: green = reference homozygote (0); dark pink = heterozygote (1) and light pink = alternative homozygote (2).

For the markers "X100038570.9.T.C" (0.0402) located on chromosome 7, "X100036509.62.G.C" (0.0419) located on chromosome 9, and "X100086491.25.C.A" (0.0432) located on chromosome 11, for the leaf miner characteristic, we have for the marker "X100038570.9.T.C", the heterozygous genotypes with a higher BLUP value than

the other genotypes (Figure 10B). For the marker "X100086491.25.C.A", the alternative homozygous genotypes (2) had a lower BLUP effect.

The effects of BLUP for the cochineal trait, for the three evaluated markers ("X100079795.58.C.T", "X100075215.44.G.A" and "X100038192.26.C.T"), were very similar (Figure 11C). With the exception of the alternative homozygous genotype (2) for the marker "X100075215.44.G.A", which showed a slightly higher BLUP effect than the others, for the same marker (Figure 10C).

The lowest percentage of the reference homozygote (0) was 0.40% for the average height growth rate characteristic at the marker "100065461.55.G.A." and the highest percentage was 98.41% in the marker "100036488.15.A.C." for the cochineal characteristic (Table 4). For the heterozygous genotypes (1) the lowest percentage was 0.40% for the markers "100065530.34.C.T", "100091676.11.A.G", "100091676.20.T.C." and "100091676.57.C.G" for the characteristic average height growth rate. For alternative homozygous genotypes (2), the lowest percentage was 0.40% for the traits rust, cochineal, leaf miner and average height growth rate, the highest percentage was 98.80% for the characteristic average height growth rate, for the flag "100065530.34.C.T" (Table 4).

The markers that showed the highest percentages for the alternative homozygote (above 90%) were "100067259.26.C.T", "100065530.34.C.T", "100037551.44.T.G", "100083384.6.T.A", "100036026.6.C.A" and "100065461.55. G.A", all markers with values above 90% for alternative homozygotes are for the average height growth rate characteristic (Table 4).

Some markers have already presented mean percentage values, that is, there was a similar distribution between the reference, heterozygous and alternative homozygous loci, they are the markers "100086368.28.A.G" and "100086401.26.C.G" (Table 4).

Table 4. Percentages of homozygous reference, alternative homozygous and heterozygous loci for the 115 markers with significant association for seven traits evaluated in 251 *Coffea canephora* genotypes.

Characteristic	Marker	Chromosome	p-value	0 (aa)	1 (Aa)	2 (AA)	0%	1%	2%
Leaf rust	100035881.53.G.T	2	0.0239	-0.0122	0.0050	0.0877	70.91	25.90	3.19
	100037456.40.C.G	3	0.0358	0.0048	-0.0513	0.1184	78.49	19.52	1.99
	100079788.19.G.C	9	0.0122	0.0082	-0.0061	-0.0986	51.00	43.03	5.97
	100083409.14.C.A.	9	0.0044	0.0067	-0.0092	-0.0591	51.00	42.63	6.37
	100085441.16.C.A.	5	0.0262	-0.0231	0.0444	0.0712	72.90	25.50	1.60
	100086228.44.T.C	11	0.0153	-0.0054	-0.0174	0.0952	76.89	19.92	3.19
	100086363.65.G.T	6	0.0104	-0.0020	-0.0257	0.1601	84.06	15.14	0.80
	100087198.32.A.C.	9	0.0026	-0.0140	0.0688	-0.0967	87.25	12.35	0.80
	100089548.46.A.T.	4	0.0029	-0.0076	0.0488	-0.1403	69.72	23.11	7.17
	100086492.40.T.G	6	0.0126	-0.0064	0.0012	-0.0135	62.55	33.07	4.38
	100064293.27.T.C	1	0.0416	0.00002	-0.0852	-	95.22	4.78	-
	100079788.25.C.G	9	0.0181	-0.0018	0.0023	-0.2048	98.01	0.80	1.19
	100083249.13.T.C.	9	0.0432	-0.0499	0.0000	0.0107	17.93	38.64	43.43
	100084742.9.C.T	8	0.0199	0.0031	-0.0837	-0.1199	93.23	3.59	3.18
	100064920.51.G.A	6	0.0462	-0.0051	-0.0039	0.1767	97.21	2.39	0.40
Mealybug	100079795.58.C.T	1	0.0191	0.0062	-0.0099	0.0149	39.84	48.61	11.55
	100074462.62.G.T	1	0.0421	-0.0003	-0.0162	-	97.21	2.79	-
	100079748.50.T.C.	1	0.0007	-0.0028	0.0263	-0.0592	92.03	7.57	0.40
	100065219.11.A.G	1	0.0352	-0.0022	0.0080	0.0059	86.45	13.15	0.40
	100036776.32.A.G	1	0.0110	-0.0008	-0.0016	-	98.01	1.99	-
	100036044.18.A.G	4	0.0231	0.0003	-0.0094	-0.0090	88.84	10.76	0.80
	100045397.33.A.T	4	0.0283	-0.0007	-0.0037	-	96.41	3.59	-
	100074449.23.G.T	4	0.0039	-0.0010	0.0027	-0.0205	71.31	25.10	3.59
	100037106.13.G.A	4	0.0065	-0.0009	0.0024	-0.0057	92.83	5.58	1.59
	100064249.10.A.G	5	0.0008	0.0003	-0.0113	-	91.63	8.37	-
	100038192.26.C.T	6	0.0156	-0.0031	0.0211	-0.0069	80.48	10.36	9.16
	100063831.7.A.T	8	0.0307	0.0004	-0.0263	-0.0001	94.02	3.99	1.99
	100064936.13.T.C	8	0.0076	-0.0018	0.0227	-0.0122	87.65	6.77	5.58
	100074576.62.T.C.	8	0.0010	-0.0021	0.0348	-0.0103	86.06	5.98	7.97
	100083206.14.C.T	9	0.0001	-0.0010	0.0102	0.0109	98.01	1.59	0.40
	100064270.13.C.G	9	0.0433	-0.0005	-0.0093	-	97.21	2.79	-

	100064019.44.G.T	9	0.0001	-0.0030	0.1021	-	98.01	1.99	-
	100036488.15.A.C.	9	0.0397	-0.0007	-0.0020	-0.0243	98.41	1.19	0.40
	100036871.32.G.A	9	0.0348	0.0005	-0.0147	-0.0050	91.63	7.18	1.19
	100075215.44.G.A	11	0.0414	-0.0034	0.0807	0.0227	96.41	3.19	0.40
Average height growth rate	100037364.63.T.A.	1	0.0013	-0.0010	0.0108	-	92.03	7.97	-
	100064842.26.C.T	1	0.0015	-0.0005	0.0074	-	94.42	5.58	-
	100065408.52.T.A.	1	0.0013	-0.0010	0.0108	-	92.03	7.97	-
	100079824.67.C.G	1	0.0105	-0.0001	-0.0124	0.0541	93.63	5.18	1.19
	100080395.6.G.C	1	0.0002	-0.0011	0.0123	-	92.43	7.57	-
	100088878.35.C.T	1	0.0019	-0.0003	0.0036	-	94.42	5.58	-
	100089339.6.A.G	1	0.0216	-0.0016	0.0085	-0.0392	80.87	18.33	0.80
	100036391.47.A.T	1	0.0339	-0.0015	0.0338	-	96.02	3.98	-
	100080103.65.T.C.	1	0.0010	-0.0017	0.0174	0.0415	92.03	7.57	0.40
	100064247.15.C.G	1	0.0008	-0.0010	-0.0098	0.1215	92.83	5.98	1.19
	100075370.65.C.T	1	0.0085	-0.0003	-0.0064	0.1660	92.83	6.77	0.40
	100075371.21.C.T	1	0.0002	-0.0003	0.0797	-0.0091	94.42	0.80	4.78
	100036613.16.T.A.	1	0.0110	-0.0010	0.0040	0.0737	93.22	5.98	0.80
	100079937.14.C.A.	1	0.0026	-0.0011	0.0368	0.0053	93.23	1.99	4.78
	100067259.26.C.T	1	0.0219	0.0162	-0.0156	-0.0002	3.58	2.79	93.63
	100075522.33.G.A.	1	0.0340	-0.0011	0.0507	-	98.01	1.99	-
	100064009.18.G.C	1	0.0277	-0.0010	0.0491	-0.0098	92.03	2.79	5.18
	100065530.34.C.T	1	0.0378	-0.0299	0.0228	0.0001	0.80	0.40	98.80
	100037551.44.T.G	1	0.0341	0.0243	0.0151	-0.0007	1.59	1.19	97.22
	100036280.37.T.C	1	0.0123	-0.0014	0.0152	-0.0418	93.63	4.38	1.99
	100068468.47.G.A.	2	0.0408	0.0002	-0.0014	-0.0353	92.03	7.57	0.40
	100037860.48.T.C	2	0.0221	-0.0005	-0.0051	0.0325	92.83	5.18	1.99
	100086401.26.C.G	4	0.0342	0.0125	-0.0069	-0.0054	32.67	34.26	33.07
	100037787.18.C.T	4	0.0298	0.0010	-0.0166	-0.0178	94.02	5.58	0.40
	100064380.10.T.C.	4	0.0224	0.0011	-0.0555	-0.0193	97.61	1.99	0.40
	100037951.22.A.G	4	0.0165	0.0012	-0.0170	0.0000	95.22	3.19	1.59
	100083384.6.T.A	5	0.0200	-	0.0026	-0.0002	-	5.98	94.02
	100038429.50.A.T	5	0.0251	-0.0002	-0.0115	0.0344	94.82	3.59	1.59
	100091676.11.A.G	5	0.0315	0.0004	-0.0914	-0.0016	95.22	0.40	4.38
	100091676.20.T.C.	5	0.0315	0.0004	-0.0914	-0.0016	95.22	0.40	4.38
	100091676.46.T.C.	5	0.0047	-0.0003	0.0172	-0.0016	94.02	1.59	4.38
	100091676.57.C.G	5	0.0315	0.0004	-0.0914	-0.0016	95.22	0.40	4.38

	100036026.6.C.A	5	0.0319	0.0387	-0.0192	0.0002	1.59	4.38	94.02
	100036253.67.T.G	6	0.0042	0.0009	-0.0139	-	93.23	6.77	-
	100063725.27.G.C	6	0.0054	-0.0063	-0.0025	0.0100	27.09	45.42	27.49
	100080081.40.C.G	6	0.0422	-0.0011	0.0049	0.1771	94.02	5.58	0.40
	100065706.45.C.A.	6	0.0165	-0.0008	0.0001	0.1660	97.61	1.99	0.40
	100089015.9.A.G	7	0.0126	-0.0001	0.0009	-	94.82	5.18	-
	100075246.24.A.G	8	0.0438	0.0191	-0.0058	-0.0036	20.72	52.99	26.29
	100079823.38.A.G	8	0.0277	-0.0063	0.0072	0.0118	57.77	31.47	10.76
	100064936.13.T.C	8	0.0100	-0.0027	0.0094	0.0296	87.65	6.77	5.58
	100037829.65.A.T.	8	0.0198	0.0005	-0.0081	-0.0088	93.63	3.98	2.39
	100065461.55.G.A.	8	0.0164	-0.0353	0.0090	-0.0001	0.40	1.59	98.01
	100074576.62.T.C.	8	0.0396	-0.0036	0.0158	0.0264	86.06	5.98	7.97
	100074989.15.A.G	9	0.0131	-0.0014	0.0207	0.0046	93.63	5.98	0.40
	100064019.44.G.T	9	0.0023	-0.0012	0.0570	-	98.01	1.99	-
	100065172.52.A.G	9	0.0272	-0.0001	-0.0147	0.0559	92.43	5.98	1.59
	100036385.14.G.T	9	0.0004	0.0010	-0.0164	-	93.63	6.37	-
	100075533.64.C.T	10	0.0449	0.0010	-0.0190	-	94.82	5.18	-
	100065474.20.A.T	2	0.0047	0.0315	-0.0838	-	92.43	7.57	-
	100088320.15.C.A.	5	0.0449	0.0228	-0.0108	-	98.01	1.99	-
Leaf miner	100038570.9.T.C.	7	0.0402	0.0194	0.2979	-0.0247	51.59	38.25	9.16
	100036509.62.G.C	9	0.0419	0.0229	0.0041	0.0949	96.81	1.20	1.99
	100086491.25.C.A.	11	0.0432	0.0234	0.0303	-0.4240	89.24	10.36	0.40
	100086368.28.A.G	1	0.0492	-0.7359	-0.3703	-1.6117	29.48	34.26	36.26
	100065543.64.C.G	2	0.0451	-0.8817	-1.0608	-5.0070	91.63	7.57	0.80
	100083449.12.C.T	3	0.0227	-0.6472	-1.8423	-0.6519	76.10	23.50	0.80
	100085459.6.G.T	4	0.0065	-0.7708	-2.1324	-	88.45	11.55	-
	100076158.35.G.A.	4	0.0328	-0.8575	-0.5586	-3.2763	72.91	21.51	5.58
Diameter	100081320.24.C.A.	6	0.0455	-0.8488	-2.5071	-	92.22	4.78	-
	100086414.43.T.C	7	0.0184	-1.2235	-0.2949	-0.8945	66.13	30.68	3.19
	100089248.46.T.A.	8	0.0390	-1.0943	-0.2178	-2.4395	68.92	26.29	4.78
	100076172.21.A.C.	10	0.0475	-0.7257	-2.0106	1.6122	80.88	17.93	1.19
	100082868.34.G.C	10	0.0423	-0.9257	-0.6033	-1.9951	90.43	7.18	2.39
	100074719.15.C.T	11	0.0016	-0.9865	1.5698	-4.7015	92.83	5.18	1.99
	100064293.27.T.C	1	0.0253	-0.0017	-0.0050	-	96.41	3.59	-
Average diameter growth rate	100065234.11.C.T	3	0.0261	-0.0024	-0.0013	-0.0017	23.90	26.30	49.80
	100075981.5.A.C.	3	0.0160	-0.0058	-0.0031	0.0005	15.94	36.65	47.41

	100067593.36.A.C.	4	0.0494	-0.0023	-0.0011	-0.0022	48.61	38.25	13.14
	100080633.26.T.A.	11	0.0068	-0.0016	-0.0012	-0.0113	60.96	35.86	3.18
	100075074.54.T.C.	11	0.0425	-0.0022	0.0001	0.0027	82.47	15.94	1.59
	100086132.17.G.C	11	0.0044	-0.0016	-0.0019	-0.0053	72.11	24.30	3.59
	100087984.59.A.T	3	0.0445	-1.7318	-1.9584	-4.4459	52.59	36.65	10.76
	100083263.38.A.G	3	0.0286	-0.8919	-1.9716	-4.5043	45.42	29.88	24.70
	100074536.21.A.G	6	0.0110	-2.0562	-4.1891	-	97.61	2.39	-
Height	100036353.45.A.C.	7	0.0267	-2.1783	-1.2721	4.8678	97.61	1.59	0.80
	100076081.41.T.A.	8	0.0332	-2.3906	0.5825	-3.3409	71.71	14.34	13.95
	100064204.60.G.C	9	0.0218	-2.0719	-2.7409	-4.1715	96.41	2.79	0.80
	100036743.22.G.A	11	0.0450	-2.0950	-2.2893	-	93.63	6.37	-

4. Discussion

In this work, 115 SNPs with significant associations were detected for important coffee crop traits of initial development such as plant height, stem diameter, height growth rate, diameter growth rate, and for incidence to rust, leaf miner and mealybug. Of the115 SNPs found, 62 were inserted into candidate genes. So far, few works have studied the GWAS in *C. canephora*. De Faria Silva et al. (2022) detected 404 SNPs with significant associations for five traits (plant height, crown diameter, vegetative vigor, incidence of rust and brown eye spot), working with 165 genotypes of *C. canephora*, including Conilon, Robusta and hybrid genotypes. Of the 404 SNPs, 217 were inserted in genes (De Faria Silva et al., 2022).

The leaf rust trait had 16 significant SNPs distributed on chromosomes 1, 2, 3, 4, 5, 6, 8, 9 and 11, of which 14 are inserted in the genes or are close to them. In the work by De Faria Silva et al. (2022), for the same trait, 22 significant SNPs were detected, nine of which were inserted into genes and four genes were found located on chromosomes 2, 8, 10 and 11 and these are potentially associated with rust tolerance.

On chromosome 8, two rust-related SNPs ("100064771.8.G.C", "100084742.9.C.T.") were detected, with the marker "100064771.8.G.C" inserted in the gene of a large group of hydrolase proteins of the HAD family at position 19,812,110. De Faria Silva et al. (2022) detected a SNP on the same chromosome inserted in the Flavone-3-Hydroxylase gene (*Cc08_g11560*), however, the position was detected was different (26,333,822..26,335,804). We also detected a rust-related SNP on chromosome 11 ("100086228.44.T.C") at position 8,902,368 related to SEC14 family lipid-binding protein. And, on chromosome 2, a disease-related SNP ("100035881.53.G.T") located in the GSCOC_T00024764001 gene with putative glucose-1-phosphate adenylyltransferase function was detected, and De Faria Silva et al. (2022) also detected a disease-related SNP on chromosome 11. disease in question, but the mitochondrial component E1 pyruvate dehydrogenase gene (*Cc11_g17430*) was at position 33,430,810..33,440,872.

For the height characteristic, seven significant SNPs were found on chromosomes 3, 6, 7, 8, 9 and 11, and five are embedded in genes or are close to them. De Faria Silva et al. (2022) found 27 SNPs significant for plant height and 13 genes potentially linked to this feature were identified. As can be seen, the only work that studies two of the characteristics evaluated in this study (plant height and rust) is that of De Faria Silva et

al. (2022). Of the SNPs identified in 6 of the 11 chromosomes of *Coffea canephora*, De Faria Silva et al. (2022) also detected SNPs in the chromosomes detected in our study, but at different positions on the chromosomes. The putative functions were related to RING_Ubox super family, transmembrane 9 family protein, nsLTP2 domain-containing protein, PSK domain-containing protein and thomboid family protein. While De Faria Silva et al. (2022) detected serine-rich protein-related functions, probable ribosome biogenesis protein RLP24, putative Adipocyte plasma membrane-associated protein, pentatricopeptide repeat-containing protein At1g61870, mitochondrial, hypothetical protein and putative TPX2 (targeting protein for Xklp2) protein family.

The other studies found with GWAS for coffee assess aspects related to quality (Sant'Ana et al., 2018; Spinoso-Castillo et al., 2022) or other diseases (Gimase et al., 2020). This demonstrates the importance of this study and encourages other studies for these characteristics evaluated to consolidate what was detected in this study.

The SNP marker "100064204.60.G.C" located on chromosome 9 is inserted in the *GSCOC_T00040077001* gene, which has a putative function of plant-specific phytosulfocin precursor proteins. The gene *GSCOC_T00019303001* detected on chromosome 1 for the cochineal trait, belongs to the family of germline proteins (oxalate oxidase) (Lu et al., 2020). It is widespread in fungi and in various plant tissues, having the characteristic of playing a role in plant signaling and defense, which may be related to the response of plants to pests and pathogens (Lu et al., 2020).

Another gene related to pest attack, such as mealybug, is the gene *GSCOC_T00022693001* located on chromosome 6. In this subfamily there are plant receptor-like kinases (RLKs), where BAK1 is linked to plant development regulated by brassinosteroid and in pathways related to plant incidence to pathogens and herbivore attack. For rust disease we found the gene GSCOC_T00039643001 which is located on chromosome 9 of *C. canephora.* It has the CaM_binding protein containing the domain that is a series of calmodulin-binding plant proteins dependent on the presence of calcium ions (Lu et al., 2020). These proteins may be involved in plant defense response processes, a fact that is extremely relevant because the SNP is related to rust (Lu et al., 2020).

The candidate gene *GSCOC_T00040251001,* also for the rust incidence trait, is found on chromosome 9 and is related to receptors linked to the galacturonan-binding

kinase wall. The GUB_WAK_bind domain is rich in cysteine and is the extracellular part of the serine/threonine kinase that will bind to cell wall pectins (Lu et al., 2020). In this context, the cysteine molecule plays a role in the redox signaling of several stress processes (Mieyal et al., 2012).

For the characteristic height growth rate, the genes *GSCOC_T00028217001* and *GSCOC_T00021883001*, located on chromosomes 1 and 6, respectively, of *C. canephora* have the putative function of the dof zinc finger protein. This protein acts as a transcription factor and can play several roles in multiple biological processes that include responses to plant hormones and stress, germination, flowering, among others. In this context, because it is related to the response to plant hormones, there may be a relationship with plant growth hormones, such as auxins and cytokinins (Chakraborty e Akhtar, 2021).

Plants have defense mechanisms against pathogens and pests through signaling cascades (Takahashi et al., 2007; Coll et al., 2011) and these cascades can be regulated by transcriptional proteins such as case of the splicing factor rich in serine and arginine, detected in the candidate gene *Cc06_g03240* (De Faria Silva et al., 2022). Alternative splicing and its regulators are known to be linked to plant responses to biotic and abiotic stresses (Duque, 2011). In this context, two SNPs were detected on chromosome 9, 100079788.19.G.C and 100083409.14.C.A, related to the incidence of rust, which are found in the gene GSCOC_T00039694001. These SNPs are related to the same protein as the serine/specific protein kinase catalytic domains and some plant receptor kinases are found (Lu et al., 2020).

The traits that showed the strongest Pearson genetic correlations are expected because these traits are directly related to the development, height and stem diameter (Martinez et al., 2007). These findings suggest that selection for growth-related traits, such as plant height, could lead to an increase in stem diameter. On the other hand, we have the characteristics related to biotic stresses that presented weak and/or negative correlations, such as, for example, the incidence of the leaf miner pest that presented negative correlations with all traits related to early development.

5. Conclusion

By a GWAS study, we detected 115 SNPs with significant associations located within or close to genes. Four SNPs are associated with more than one trait (100064936.13.T.C, 100074576.62.T.C, 100064019.44.G.T and 100064293.27.T.C).

Through the *National Center for Biotechnology Information* (NCBI) it was possible to identify candidate genes with putative functions related to developmental traits and plant defense mechanisms against pests and diseases.

These results are relevant for coffee genetic breeding programs, since the chromosomal regions that were detected could be used for selection assisted by molecular markers to help and accelerate the coffee improvement process.

6. Acknowledgment

The authors would like to thank Conselho Nacional de Desenvolvimento Científico e Tecnológico [National Council for Scientific and Technological Development] (CNPq, Brasília – DF, Brazil, grant number 311950/2016-7), Fundação de Amparo à Pesquisa e Inovação do Espírito Santo [Research Support Foundation of Espírito Santo] (FAPES, Vitória – ES, Brazil), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior [Coordination for the Improvement of Higher Education Personnel] (CAPES, Brasília, DF, Brazil) – Finance Code 001, for financial support, CAFESUL - Cooperativa dos Cafeicultores do Sul do Estado do Espírito Santo, as well as the producers who allowed the collection of seeds in their crops and all who helped to implement the experiment.

7. References

Alkimim ER, Caixeta ET, Sousa TV, Pereira AA, de Oliveira ACB, Zambolim L, Sakiyama NS (2017) Marker-assisted selection provides arabica coffee with genes from other *Coffea* species targeting on multiple incidence to rust and coffee berry disease. Molecular Breeding, 37(1):1–10.

Almeida DP, Caixeta ET, Moreira KF, de Oliveira ACB, de Freitas KNP, Pereira AA, ..., Cruz CD (2021) Marker-assisted pyramiding of multiple disease resistance genes in coffee genotypes (*Coffea arabica*). Agronomy, 11(9):1763.

Amadeu RR, Cellon C, Olmstead JW, Garcia AA, Resende Jr MF, Muñoz PR (2016) AGHmatrix: R package to construct relationship matrices for autotetraploid and diploid species: a blueberry example. The plant genome, 9(3).

Andrade AC, da Silva Junior OB, Carneiro F, Marraccini P, Grattapaglia D (2017) Towards GWAS and Genomic Prediction in Coffee: Development and Validation of a 26K SNP Chip for *Coffea canephora*. In *Embrapa Café-Resumo em anais de congresso*

(ALICE). In: Plant & Animal Genome, 25., 2017, San Diego. Final program & exhibite guide... San Diego: USDA.

Chakraborty T, Akhtar N (2021) Biofertilizers: Prospects and challenges for future. Biofertilizers: Study and Impact 575–590.

Coll NS, Epple P, Dangl JL (2011) Programmed cell death in the plant immune system. Cell Death Differ, 18:1247–1256.

Cruz CD, Carneiro PSC, Regazzi AJ (2014) Modelos biométricos aplicados ao melhoramento genético. 3.ed. Viçosa: UFV, v.2. 668p.

Denoeud F, Carretero-Paulet L, Dereeper A, Droc G, Guyot R, Pietrella M, Zheng C, Alberti A, Anthony F, Aprea G, Aury JM, Bento P, Bernard M, Bocs S, Campa C, Cenci A, Combes MC, Crouzillat D, Da Silva C, Daddiego L, De Bellis F, Dussert S, Garsmeur O, Gayraud T, Guignon V, Jahn K, Jamilloux V, Jöet T, Labadie K, Lan T, Leclercq J, Lepelley M, Leroy T, Li LT, Librado P, Lopez L, Moñoz A, Noel B, Pallavicini A, Perrota G, Poncet V, Pot D, Priyono, Rigoreau M, Rouard M, Rozas J, Tranchant-Dubreuil C, VanBuren R, Zhang Q, Andrade AC, Argout X, Bertrand B, Kochko A, Graziosi G, Henry RJ, Jayrama, Ming R, Nagai C, Rounsley S, Sankoff D, Giuliano G, Albert VA, Wincker P, Lashermes P (2014) The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. Science 345(80):1181–1184. https://doi.org/10.1126/science.1255274

De Faria Silva L, Alkimim ER, Barreiro PRRM et al. Genome-wide association study of plant architecture and diseases resistance in *Coffea canephora* (2022) Euphytica 218: 92 https://doi.org/10.1007/s10681-022-03042-8

Duque P (2011) A role for SR proteins in plant stress responses. Plant Signal Behav 6:49–54.

Ferrão RG, Fonseca AFA, Ferrão MAG, de Muner LH (2019) Conilon Coffee. Vitória, ES, Brazil

Ferrão R, Volpi P, Senra JDB, Comério M, Ferrão M, Riva-Souza EM, ..., Verdin Filho, AC (2022) Variabilidade de *Coffea canephora* do Banco Ativo de Germoplasma do Incaper: Caracterização dos Acessos com Base em Descritores Mínimos.

Gimase JM, Thagana WM, Omondi CO, Cheserek JJ, Gichimu BM, Gichuru EK, ..., Sneller CH (2020) Genome-Wide Association Study identify the genetic loci conferring incidence to Coffee Berry Disease (*Colletotrichum kahawae*) in *Coffea arabica* var. Rume Sudan. Euphytica, 216(6):1–17.

González AM, Godoy L, Santalla M (2017) Dissection of incidence genes to *Pseudomonas syringae* pv. *phaseolicola* in UI3 common bean cultivar. International Journal of Molecular Sciences, 18, pii: E2503.

Heffner EL, Sorrells ME, Jannink JL (2009) Genomic selection for crop improvement. Crop Sci. 49:1–12.

Jiangtao C, Yingzhen K, Qian W, Yuhe S, Daping G, Jing LV, Guanshan L (2015) MapGene2Chrom, a tool to draw gene physical map based on Perl and SVG languages. *Yi chuan= Hereditas*, 37(1):91–97.

Kamfwa K, Cichy KA, Kelly JD (2015) Genome-wide association study of agronomic traits in common bean. Plant Genome 8:1–12. https://doi.org/10.3835/plantgenome2014.09.0059

Legarra A, Aguilar I, Misztal I (2009) A relationship matrix including full pedigree and genomic information. Journal of dairy science, 92(9):4656–4663.

Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, ..., Marchler-Bauer, A (2020) CDD/SPARCLE: the conserved domain database in 2020. Nucleic acids research, 48(D1):D265–D268.

Maia T, Badel JT, Fernandes MB, et al (2017) Variation in Aggressiveness Components in the *Hemileia vastatrix* Population in Brazil, J Phytopathol 165:174-188.

Martinez HEP, Augusto HS, Cruz CD, Pedrosa AW, Sampaio NF (2007) Vegetative growth of the coffee plant (*Coffea arabica* L.) and its correlation with the production in narrower spacing. Acta Sci. Agron. 29:481–489.

Mieyal JJ, Chock PB (2012) Posttranslational modification of cysteine in redox signaling and oxidative stress: focus on s-glutathionylation. Antioxidants & redox signaling, 16(6): 471–475.

Mohammed A. (2015) Importance and Characterization of Coffee Berry Disease (*Colletotrichum kahawae*) in Borena and Guji Zones, Southern Ethiopia. J Plant Pathol Microbiol. 06(09).

Moncada MDP, Tovar E, Montoya JC, González A, Spindel J, McCouch S (2016) A genetic linkage map of coffee (*Coffea arabica* L.) and QTL for yield, plant height, and bean size. Tree genetics & genomes, 12(1):1–17.

Monteiro ALR, Pantaleão AS, Badel JL, Soares PH, Carneiro VQ, Carneiro PC, Carneiro JES (2021) Genome-wide association study (GWAS) of *Phaseolus vulgaris* incidence to *Xanthomonas citri* pv. *fuscans*. Plant Pathology, 70(7):1733–1744.

Muñoz PR, Resende Jr MF, Gezan SA, Resende MDV, de Los Campos G, Kirst M, ..., Peter GF (2014) Unraveling additive from nonadditive effects using genomic relationship matrices. Genetics, 198(4):1759-1768.

NCBI – National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; (2022) – [cited 2022 Jun 27]. Available from: https://www.ncbi.nlm.nih.gov/

Perseguini JMKC, Oblessuc PR, Rosa JRBF, Gomes KA, Chiorato AF et al. (2016) Genome-wide association studies of anthracnose and angular leaf spot incidence in common bean (*Phaseolus vulgaris* L.). PLoS One 11:e0150506. https://doi.org/10.1371/journal.pone.0150506

RStudio Team. (2022). *RStudio: Integrated Development for R. RStudio*, PBC, Boston,

MA URL http://www.rstudio.com/.

Resende MDV et al. (2008) Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético, Pesquisa Florestal Brasileira, Vol 56,p 63-77

Resende RT, de Resende MDV, Azevedo CF, Fonseca e Silva F, Melo LC, Pereira HS, ..., Vianello RP (2018) Genome-wide association and regional heritability mapping of plant architecture, lodging and productivity in *Phaseolus vulgaris*. G3: Genes, Genomes, Genetics, 8(8):2841–2854.

Sant'Ana GC, Pereira LF, Pot D, Ivamoto ST, Domingues DS, Ferreira RV, ..., Leroy T (2018) Genome-wide association study reveals candidate genes influencing lipids and diterpenes contents in *Coffea arabica* L. Scientific reports, 8(1):1–12.

Sousa TV, Caixeta ET, Alkimim ER, de Oliveira ACB, Pereira AA, Zambolim L, Sakiyama NS (2017). Molecular markers useful to discriminate *Coffea arabica* cultivars with high genetic similarity. Euphytica, 213(3):1–15.

Sousa TV, Caixeta ET, Alkimim ER, Oliveira ACB, Pereira AA, Sakiyama NS, Zambolim L, Resende MDV (2019) Early selection enabled by the implementation of genomic selection in *Coffea arabica* breeding. Frontiers in Plant Science, 9:1934.

Souza FFD (2011) Estudos sobre a diversidade, estrutura populacional, desequilíbrio de ligação e mapeamento associativo em *Coffea canephora* Pierre ex Froehner, Universidade Federal de Viçosa, Viçosa, MG.

Sousa TV, Caixeta ET, Alkimim ER (2017) Molecular markers useful to discriminate *Coffea arabica* cultivars with high genetic similarity. Euphytica, 213:75.

Spinoso-Castillo JL, Paulino PR, Bello-Bello JJ, Esteban EP, Víctor Heber AR, Tarsicio CT, ..., Victorino MR (2022). SNP markers identification by genome wide association study for chemical quality traits of coffee (*Coffea* spp.) Germplasm. Molecular Biology Reports, 1-11.https://doi.org/10.1007/s11033-022-07339-8

Takahashi Y, Nasir KHB, Ito A, Kanzaki H, Matsumura H, Saitoh H et al (2007) A highthroughput screen of cell-death-inducing factors in Nicotiana benthamiana identifies a novel MAPKK that mediates INF1-induced cell death signaling and non-host incidence to Pseudomonas cichorii. Plant J, 49:1030–1040.

Talhinhas P, Batista D, Diniz I, Vieira A, Silva DN, Loureiro A, Tavares S, Pereira AP, Azinheira HG, Guerra-Guimarães L, et al. (2017) The coffee leaf rust pathogen *Hemileia vastatrix*: one and a half centuries around the tropics. Mol Plant Pathol. 18(8):1039–1051.

Visscher PM, Wray NR, Zhang Q (2017) 10 years of GWAS discovery: biology, function, and translation. American Journal of Human Genetics, 101:5–22

Yang J, Lee SH, Goddard ME, Visscher PM (2013) Genome-wide complex trait analysis (GCTA): methods, data analyses, and interpretations. In: *Genome-wide association studies and genomic prediction* (pp. 215-236). Humana Press, Totowa, NJ.

Zuiderveen GH, Padder BA, Kamfwa K, Song Q, Kelly JD (2016) Genome-wide association study of anthracnose incidence in Andean beans (*Phaseolus vulgaris*). PLoS One 11: e0156391. https://doi.org/ 10.1371/journal.pone.0156391

8. Final considerations

This work consists of an unprecedented study in ancient seminal crops remaining in the south of Espírito Santo. The phenotypic and genetic variability that were detected consist of genetic resources, serving as raw material for coffee genetic improvement programs. The characterization and preservation of these genotypes is fundamental for the maintenance of the culture because the use of few clonal cultivars and the insertion of materials of the Robusta variety in the state, can lead to a narrowing and loss of genetic diversity of the species.

In fact, the state of Espírito Santo is of world importance for the production and improvement of coffee, specifically *Coffea canephora* Conilon variety. Because it constitutes a genetic reservoir of the species that still remains in ancient seminal crops in the south of the state. However, there is a concern with the conservation and preservation of this material, due to the fact that these remaining seminal crops in the south of the state are decreasing due to the renewal by commercial clones of clonal propagation. Furthermore, the introduction of Robusta genotypes from Rondônia in the state has been taking place in the last ten years, and the recent wide dissemination of these genotypes in the state.

In this study, materials from seminal crops have aptitude for initial development in conditions of absence of irrigation, demonstrating the rusticity present in these materials under stressful conditions in the field. This aptitude was expected and was confirmed, given the hypothesis for the collection of mother plants in seminal crops. In this study, few cultural treatments were used and the crop was in rainfed condition, so the detection and selection of these materials is extremely important in the initial phase of development, for crops with little technological resource. Materials that proved to be drought tolerant, must be observed and selected in the face of climatic changes and adverse conditions that these changes entail, the use of tolerant genotypes is a viable alternative aiming at the sustainability of the coffee culture. On the other hand, half-sib families were also found that suffered a lot in the face of adverse weather conditions, due to pest infestation and the incidence of rust.

The genetic diversity detected from the DArTseq^{TM} methodology by SNP markers should also be explored. Divergent groups were found, with two groups presenting marker parameter values (fixation index, observed and expected heterozygosity) that indicate wide genetic variability, and therefore should be selected for genetic improvement purposes. The genetic distance values found for some of the genotypes (G93, 168.6, G92, 172.5, G55, 172.8, 126, Jequitibá P1, G80 and 162) serve as a starting point to propose a new collection of genotypes that can promote the expansion of the genetic base of coffee. In the crops of southern ES there are genotypes with around 10% of loci in heterozygosity, but there is also a smaller number of individuals with 50% of loci in heterozygosity, suggesting hybridization between different gene pools, which may be feasible by the introduction of Robusta genotypes in the state in the last ten years, suggesting the need for investigation into this fact.

The high divergence of these groups was demonstrated in the morphological analyzes with specific groups demonstrating expressive differences in growth rates and molecular diversity. Comparing to commercial clones, there is a differential germplasm for both morphological and molecular aspects, demonstrating potential for the selection of genotypes, verified in clusters with examples of possibility of selection of genotypes with high growth rate and with tolerance to the stresses present in field. The divergent groups detected may also indicate promising crosses for the improvement of Conilon coffee in the state of Espírito Santo.

The GWAS study was able to detect, among the 2,542 SNPs, 115 SNPs with significant association, of which 62 were inserted in genes. Of the eight traits evaluated, seven traits of agronomic interest related to early development and biotic stresses (plant height, average height growth rate, stem diameter, average diameter growth rate, rust, leaf miner and mealybug) showed SNPs with significant association. These results must be explored, since this study provides information of relevance for the genetic improvement programs of the species *Coffea canephora*, which generally involve expensive, late, and complex processes, as it is a perennial species that requires 25 to 30 for the launch of a new variety. Therefore, once chromosomal regions have been found, they can be used in genome-wide association studies, which is a powerful and efficient tool to accelerate the selection of superior genotypes that have greater amounts of favorable attributes, thus increasing the efficiency of programs. of coffee improvement.

9. Supplementary Figures

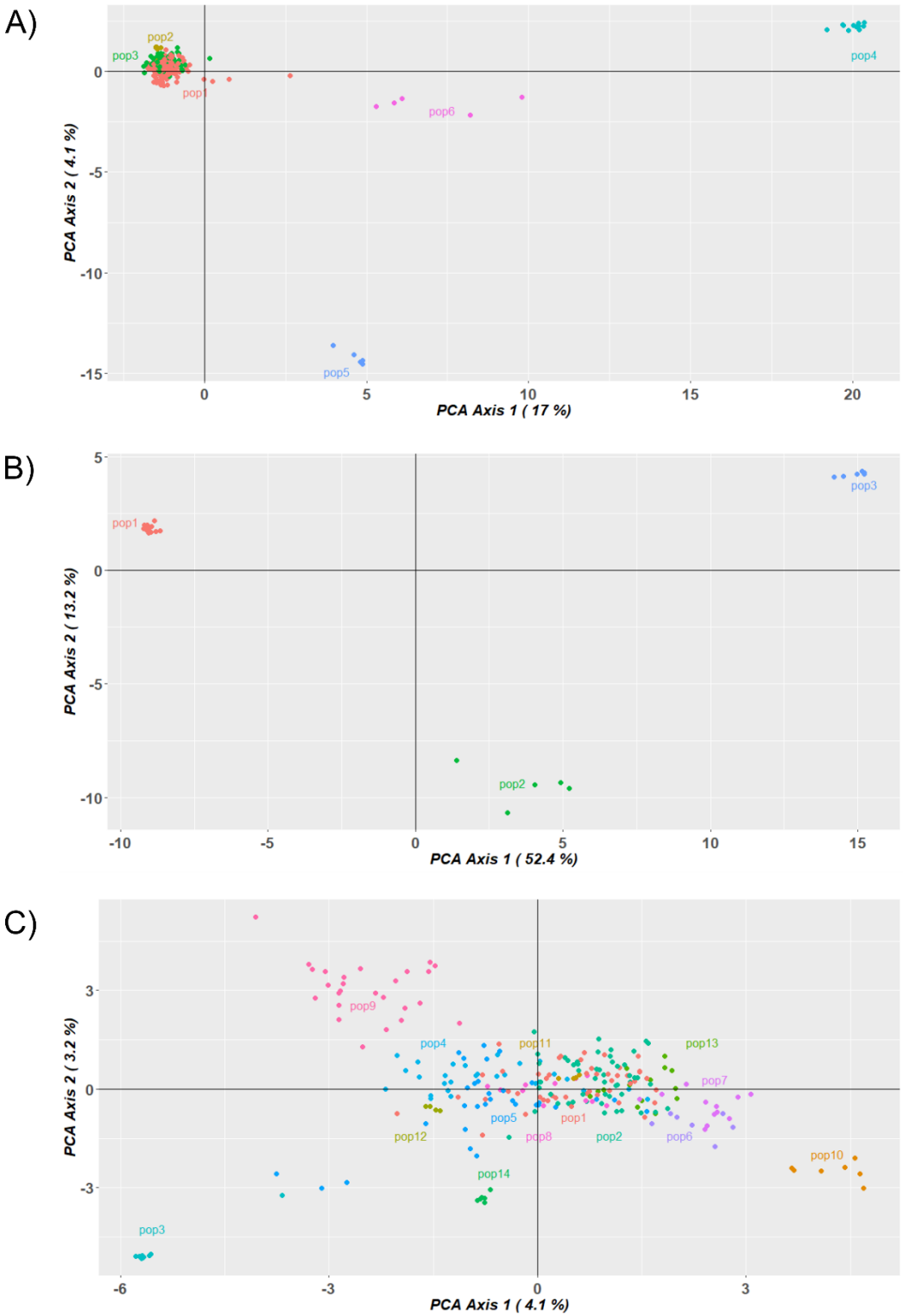

Figure S1. A) Principal components analysis of the six groups formed by the first grouping of the 280 genotypes of *Coffea canephora*. B) Analysis of principal components of the three groups formed by the 23 genotypes of *Coffea canephora*. C) Analysis of principal components of the 14 groups formed by the 257 genotypes of *Coffea canephora*

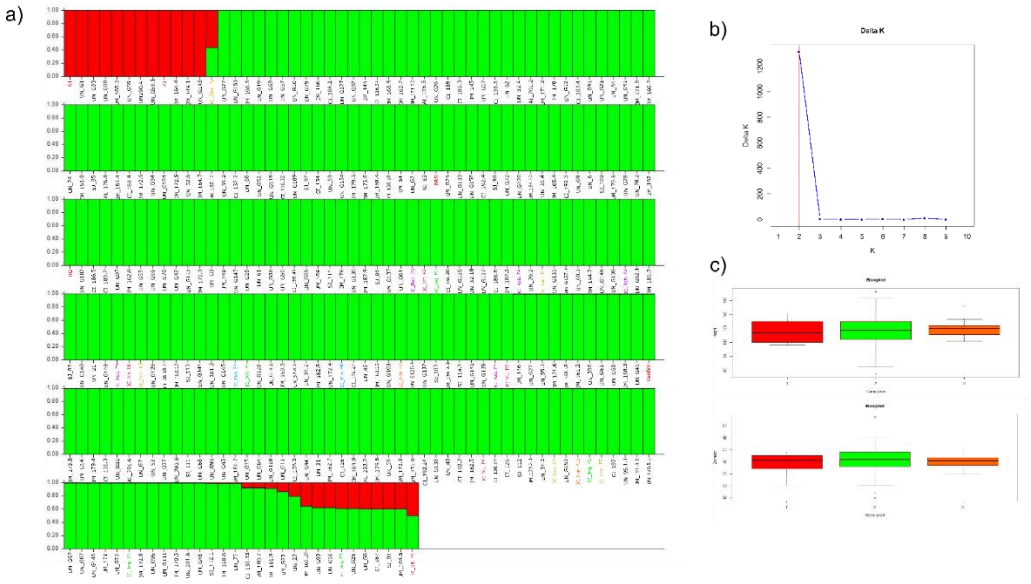

Figure S2. A) Population structure of 280 *C. canephora* genotypes using SNP marker data with the formation of two gene pools. B) Graphic obtained with the values of ΔK for visualization of the best K (K = 2), according to the methodology proposed by Evanno et al. (2005). C) Boxplot of plant height (mm) and stem diameter (mm) data for the three formed gene pools. AL = Alegre; CI = Cachoeiro de Itapemirim; IC = Incaper Cultivars; JM = Jerônimo Monteiro; SJ = São José do Calçado; UN = South

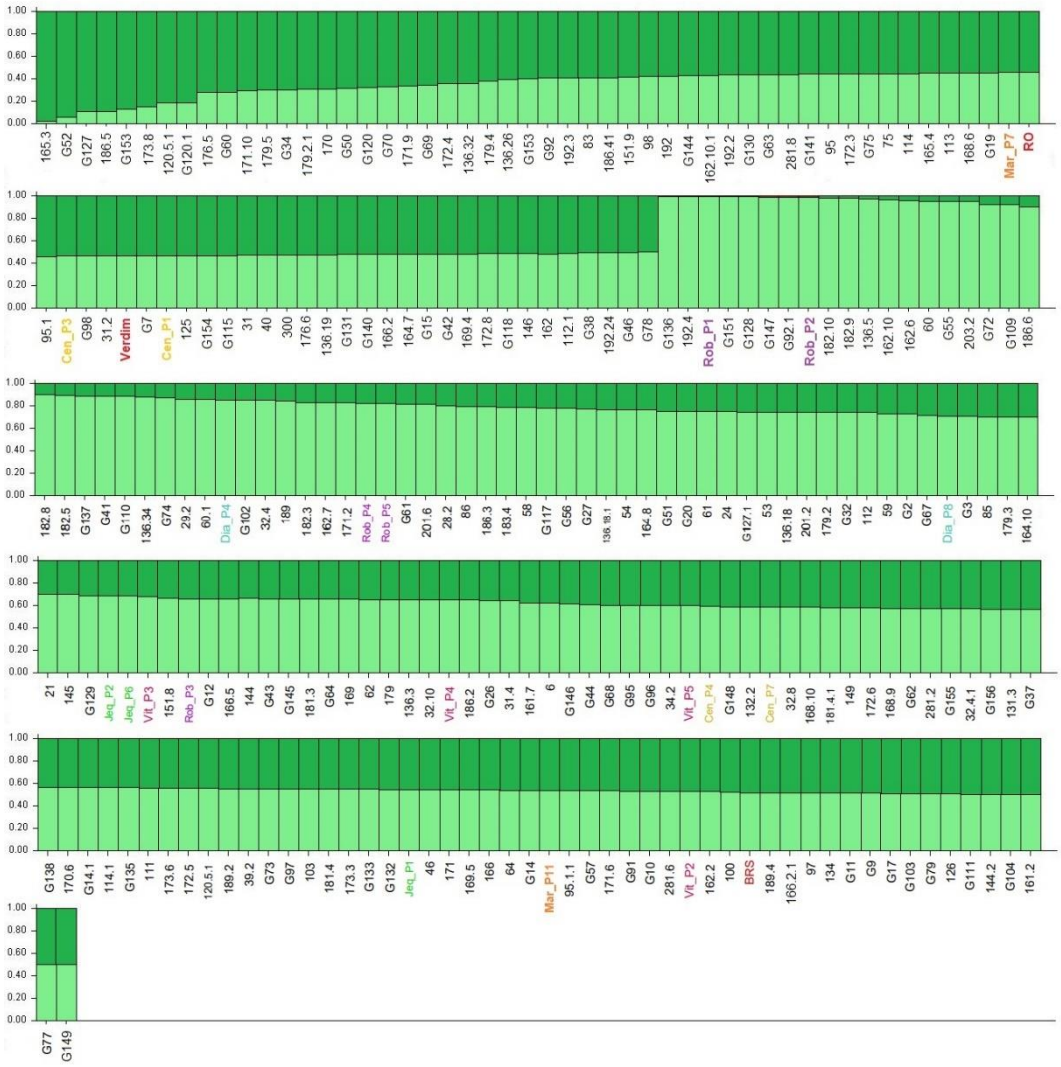

Figure S3. Population structure of 252 *Coffea canephora* genotypes from the green gene pool from the first analysis performed for the 280 genotypes using SNP marker data with the formation of two gene pools for K = 2, according to the methodology proposed by Evanno et al. (2005).

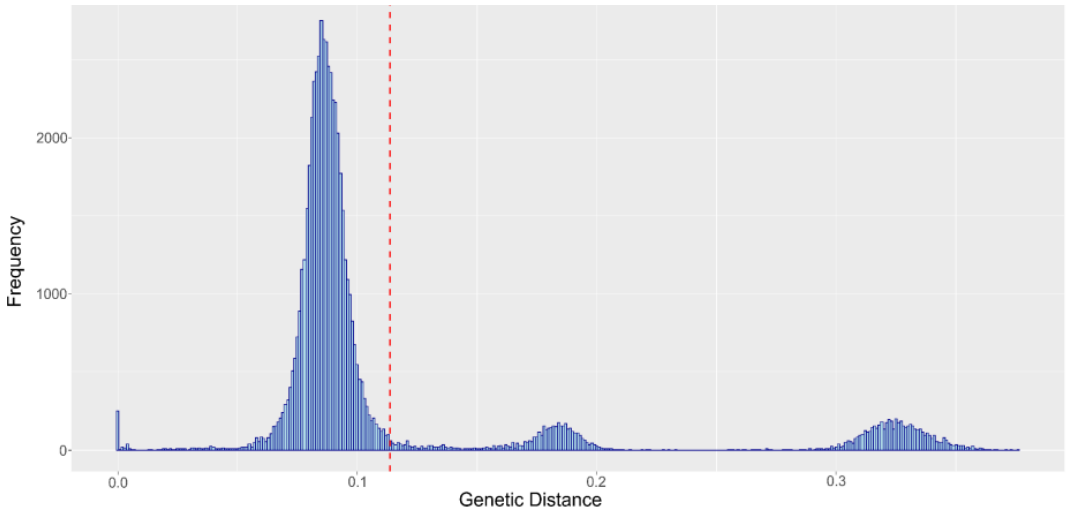

Figure S4. Frequency of genetic distance values between pairs of evaluated genotypes.

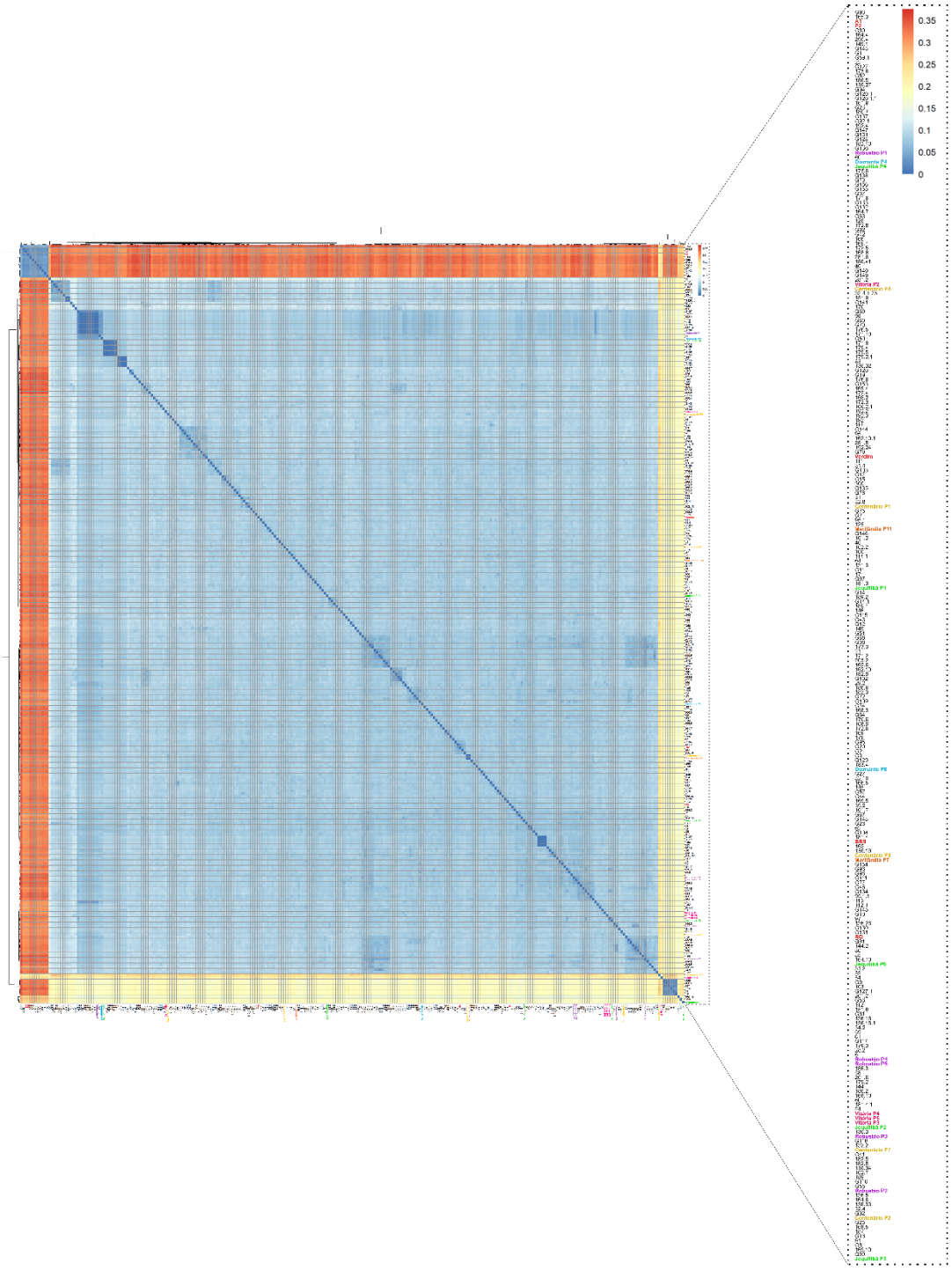

Figure S5. Genetic distance heatmap (Nei, 1972) of the 280 *Coffea canephora* genotypes. The red color represents the most genetically dissimilar genotypes and the blue color represents the most genetically similar genotypes. Commercial clones are represented by the colors red (A1, P2, BRS, RO and Verdim), purple (Robustão), pink (Vitória), green (Jequitibá), blue (Diamante), yellow (Centenário) and orange (Marilândia).

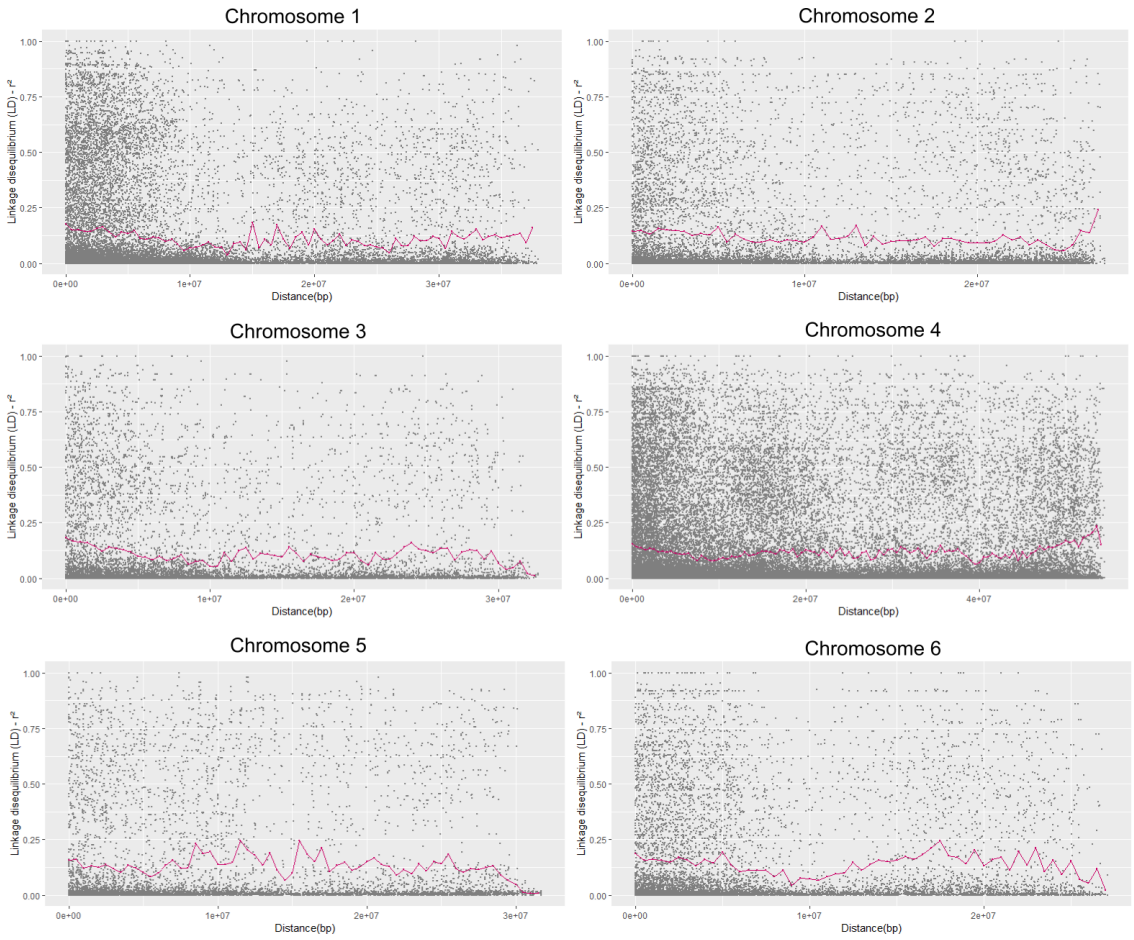

Figure S6. Allele pair linkage imbalance (r^2) in *Coffea canephora* chromosomes for all genotypes, plotted according to genetic distance in base pairs. The pink line refers to the downward trend by combining the SNPs within 500000 bp windows and averaging the r^2 values (continue).

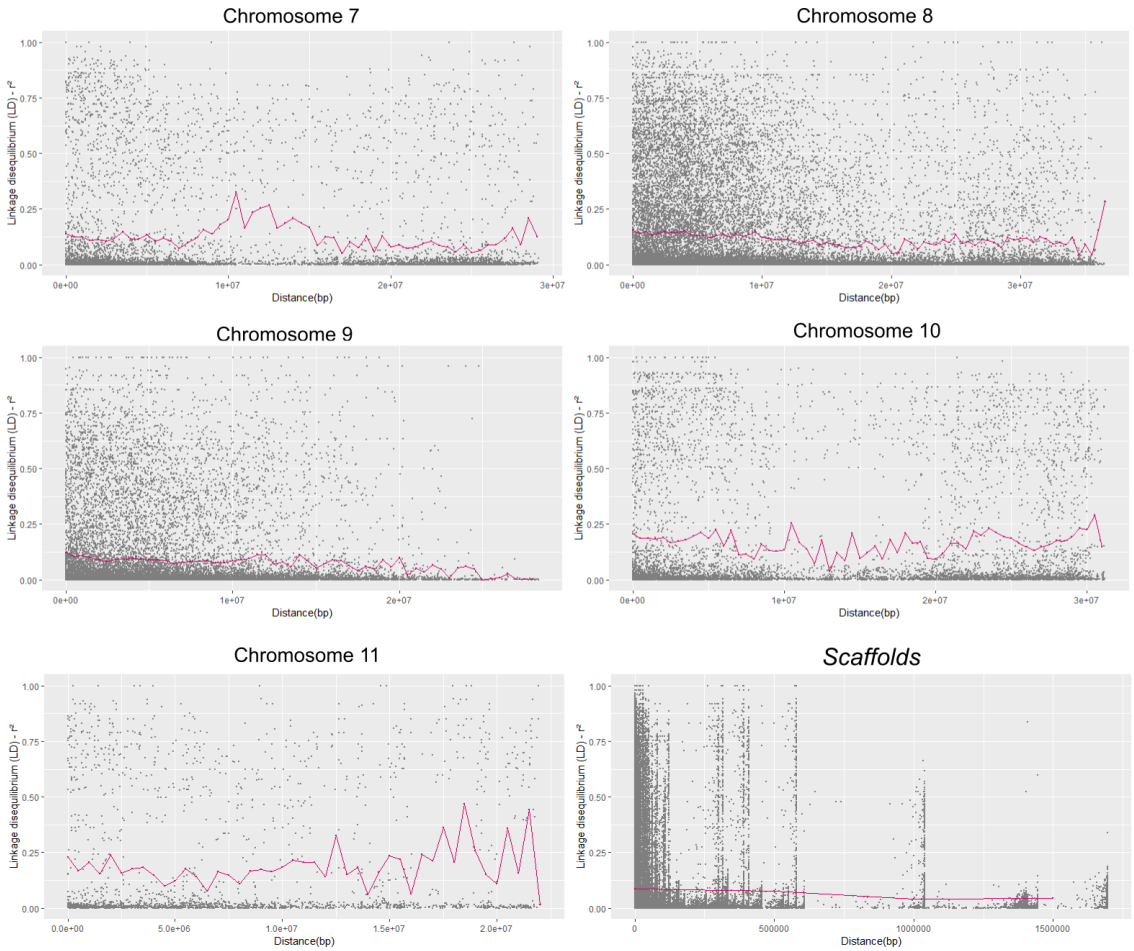

Figure **S7**. Allele pair linkage imbalance (r^2) in *Coffea canephora* chromosomes for all genotypes, plotted according to genetic distance in base pairs. The pink line refers to the downward trend by combining the SNPs within 500000 bp windows and averaging the r^2 values.

10. Supplementary Tables

Table S1. List of 388 half-sib families of *Coffea canephora* and five commercial clones.

Family	Origin	Crop age (years)*	Commercial/Cultivated	Latitude	Longitude	Altitude
176	Alegre	40	Cultivated	S 20° 45' 08.9"	W 41° 25' 49.4"	117
176_2	Alegre	40	Cultivated	S 20° 45' 08.9"	W 41° 25' 49.4"	117
176_4	Alegre	40	Cultivated	S 20° 45' 08.9"	W 41° 25' 49.4"	117
176_5	Alegre	40	Cultivated	S 20° 45' 08.9"	W 41° 25' 49.4"	117
201	Alegre	35	Cultivated	S 20° 43' 47.1"	W 41° 27' 12.7"	364
201_2	Alegre	35	Cultivated	S 20° 43' 47.1"	W 41° 27' 12.7"	364
201_6	Alegre	35	Cultivated	S 20° 43' 47.1"	W 41° 27' 12.7"	364
176_6	Alegre	40	Cultivated	S 20° 45' 08.9"	W 41° 25' 49.4"	117
203_2	Alegre	25	Cultivated	S 20° 43' 33.9"	W 41° 26' 32.2"	243
206_7	Alegre	NA	Cultivated	S 20° 43' 33.9"	W 41° 26' 32.2"	243
197	Alegre	35	Cultivated	S 20° 42' 37.3"	W 41° 23' 53.7"	118
200_3	Alegre	45	Cultivated	S 20° 43' 39.1"	W 41° 26' 36.2"	257
195	Alegre	80	Cultivated	S 20° 42' 07.0"	W 41° 20' 32.2"	105
136_5	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 50.1""	W 41° 19" 40'.3"	153
136_3	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 50.1""	W 41° 19" 40'.3"	153
134	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 50.8"	W 41° 19" 39.2"	152
133	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 51' .0"	W 41° 19" 39.6"	157
132_2	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 51' . 3"	W 41° 19" 39. 8"	157
131	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 51' .7"	W 41° 19" 39.6"	168
190_2	Cachoeiro de Itapemirim	20	Cultivated	S 20° 43' 35.5"	W 41° 20" 58.6"	136
189_2	Cachoeiro de Itapemirim	20	Cultivated	S 20° 43' 35.6"	W 41° 20" 59.6"	133
189_4	Cachoeiro de Itapemirim	20	Cultivated	S 20° 43' 35.6"	W 41° 20" 59.6"	133
189_3	Cachoeiro de Itapemirim	20	Cultivated	S 20° 43' 35.6"	W 41° 20" 59.6"	133

189	Cachoeiro de Itapemirim	20	Cultivated	S 20° 43' 35.6"	W 41° 20" 59.6"	133
187	Cachoeiro de Itapemirim	20	Cultivated	S 20° 43' 35.4"	W 41° 20" 59.7"	130
136_36	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 50.1"	W 41° 19" 40'.3"	146
126	Cachoeiro de Itapemirim	40	Cultivated	S 20° 48" 26.8"	W 41° 18' 43.3"	143
125	Cachoeiro de Itapemirim	40	Cultivated	S 20° 48" 27.0"	W 41° 18' 43.1"	143
136_8	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 50.1"	W 41° 19' 40'.3"	151
136_18	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 50.1"	W 41° 19' 40'.3"	151
136_34	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 50.1"	W 41° 19' 40'.3"	151
183_4	Cachoeiro de Itapemirim	20	Cultivated	S 20° 43' 33.9"	W 41° 20" 59.6"	120
136_33	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 50.1"	W 41° 19' 40'.3"	151
136_32	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 50.1"	W 41° 19' 40'.3"	151
123	Cachoeiro de Itapemirim	40	Cultivated	S 20° 48' 28.3"	W 41° 18' 41.5"	146
183_10	Cachoeiro de Itapemirim	20	Cultivated	S 20° 43' 33.9"	W 41° 20' 59.6"	120
184_2	Cachoeiro de Itapemirim	20	Cultivated	S 20° 43' 34.7"	W 41° 20' 59.7"	125
184_3	Cachoeiro de Itapemirim	20	Cultivated	S 20° 43' 34.7"	W 41° 20' 59.7"	125
124	Cachoeiro de Itapemirim	40	Cultivated	S 20° 48' 28.0 "	W 41° 18' 42.4"	144
186_41	Cachoeiro de Itapemirim	20	Cultivated	S 20° 43' 35.3"	W 41° 20' 59.9"	129
186_4	Cachoeiro de Itapemirim	20	Cultivated	S 20° 43' 35.3"	W 41° 20' 59.9"	129
186_3	Cachoeiro de Itapemirim	20	Cultivated	S 20° 43' 35.3"	W 41° 20' 59.9"	129
186_2	Cachoeiro de Itapemirim	20	Cultivated	S 20° 43' 35.3"	W 41° 20' 59.9"	129
186_5	Cachoeiro de Itapemirim	20	Cultivated	S 20° 43' 35.3"	W 41° 20' 59.9"	129
192_24	Cachoeiro de Itapemirim	20	Cultivated	S 20° 43' 36.3""	W 41° 21' 00.1"	139
136_19	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 50.1"	W 41° 19' 40'.3"	151
136_23	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 50.1"	W 41° 19' 40'.3"	151
136_26	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 50.1"	W 41° 19' 40'.3"	151
136_27	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 50.1"	W 41° 19' 40'.3"	151
136_31	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 50.1"	W 41° 19' 40'.3"	151
192	Cachoeiro de Itapemirim	20	Cultivated	S 20° 43' 36.3"	W 41° 21' 00.1"	137

192_2	Cachoeiro de Itapemirim	20	Cultivated	S 20° 43' 36.3"	W 41° 21' 00.1"	137
192_3	Cachoeiro de Itapemirim	20	Cultivated	S 20° 43' 36.3"	W 41° 21' 00.1"	137
192_5	Cachoeiro de Itapemirim	20	Cultivated	S 20° 43' 36.3"	W 41° 21' 00.1"	137
192_4	Cachoeiro de Itapemirim	20	Cultivated	S 20° 43' 36.3"	W 41° 21' 00.1"	137
168_6	Jerônimo Monteiro	27	Cultivated	S 20° 49' 32.4"	W 41° 21" 30.1"	205
168_5	Jerônimo Monteiro	27	Cultivated	S 20° 49' 32.4"	W 41° 21" 30.1"	205
168_10	Jerônimo Monteiro	27	Cultivated	S 20° 49' 32.4"	W 41° 21" 30.1"	205
168_9	Jerônimo Monteiro	27	Cultivated	S 20° 49' 32.4"	W 41° 21" 30.1"	205
182_2	Jerônimo Monteiro	46	Cultivated	S 20° 46" 18.8""	W 41° 23" 17.4"	115
168_3	Jerônimo Monteiro	27	Cultivated	S 20° 49' 32.4"	W 41° 21" 30.1"	204
172_2	Jerônimo Monteiro	35	Cultivated	S 20° 49' 18.4""	W 41° 21" 23'.8"	194
172	Jerônimo Monteiro	35	Cultivated	S 20° 49' 18.4""	W 41° 21" 23'.8"	194
171_10	Jerônimo Monteiro	35	Cultivated	S 20° 49' 19.4"	W 41° 21" 25.7"	194
171_9	Jerônimo Monteiro	35	Cultivated	S 20° 49' 19.4"	W 41° 21" 25.7"	194
171_8	Jerônimo Monteiro	35	Cultivated	S 20° 49' 19.4"	W 41° 21" 25.7"	194
171_6	Jerônimo Monteiro	35	Cultivated	S 20° 49' 19.4"	W 41° 21" 25.7"	194
173_8	Jerônimo Monteiro	35	Cultivated	S 20° 49' 17.2""	W 41° 21" 26.7"	199
173_9	Jerônimo Monteiro	35	Cultivated	S 20° 49' 17.2""	W 41° 21" 26.7"	199
175	Jerônimo Monteiro	33	Cultivated	S 20° 45' 02.7"	W 41° 24' 42.2"	122
161_2	Jerônimo Monteiro	30	Cultivated	S 20° 49' 25.7"	W 41° 21" 32.6"	252
161_5	Jerônimo Monteiro	30	Cultivated	S 20° 49' 25.7"	W 41° 21" 32.6"	252
161_6	Jerônimo Monteiro	30	Cultivated	S 20° 49' 25.7"	W 41° 21" 32.6"	252
161_7	Jerônimo Monteiro	30	Cultivated	S 20° 49' 25.7"	W 41° 21" 32.6"	252
161_8	Jerônimo Monteiro	30	Cultivated	S 20° 49' 25.7"	W 41° 21" 32.6"	252
169	Jerônimo Monteiro	27	Cultivated	S20° 49' 32.7"	W 41° 21" 30.8"	207
169_4	Jerônimo Monteiro	27	Cultivated	S20° 49' 32.7"	W 41° 21" 30.8"	207
169_5	Jerônimo Monteiro	27	Cultivated	S20° 49' 32.7"	W 41° 21" 30.8"	207
169_6	Jerônimo Monteiro	27	Cultivated	S20° 49' 32.7"	W 41° 21" 30.8"	207

169_8	Jerônimo Monteiro	27	Cultivated	S20° 49' 32.7"	W 41° 21" 30.8"	207
162_9	Jerônimo Monteiro	25	Cultivated	S 20° 49' 24.4"	W 41° 21" 32.5"	256
162_10	Jerônimo Monteiro	25	Cultivated	S 20° 49' 24.4"	W 41° 21" 32.5"	256
163_2	Jerônimo Monteiro	25	Cultivated	S 20° 49' 24.1"	W 41° 21" 33.0"	262
164	Jerônimo Monteiro	25	Cultivated	S 20° 49' 23.3""	W 41° 21' 35.1"	264
164_2	Jerônimo Monteiro	25	Cultivated	S 20° 49' 23.3""	W 41° 21' 35.1"	264
164_3	Jerônimo Monteiro	25	Cultivated	S 20° 49' 23.3""	W 41° 21' 35.1"	264
173	Jerônimo Monteiro	35	Cultivated	S 20° 49' 17.2"	W 41° 21' 36.7"	280
173_2	Jerônimo Monteiro	35	Cultivated	S 20° 49' 17.2"	W 41° 21' 36.7"	280
173_3	Jerônimo Monteiro	35	Cultivated	S 20° 49' 17.2"	W 41° 21' 36.7"	280
173_6	Jerônimo Monteiro	35	Cultivated	S 20° 49' 17.2"	W 41° 21' 36.7"	280
173_7	Jerônimo Monteiro	35	Cultivated	S 20° 49' 17.2"	W 41° 21' 36.7"	280
151_9	Jerônimo Monteiro	> 25	Cultivated	S 20° 45' 25.3"	W 41° 21' 42.6"	111
151_8	Jerônimo Monteiro	> 25	Cultivated	S 20° 45' 25.3"	W 41° 21' 42.6"	111
151_6	Jerônimo Monteiro	> 25	Cultivated	S 20° 45' 25.3"	W 41° 21' 42.6"	111
151_5	Jerônimo Monteiro	> 25	Cultivated	S 20° 45' 25.3"	W 41° 21' 42.6"	111
181_10	Jerônimo Monteiro	46	Cultivated	S 20° 46' 19.0"	W 41° 23' 17.5"	115
181_8	Jerônimo Monteiro	46	Cultivated	S 20° 46' 19.0"	W 41° 23' 17.5"	115
181_4	Jerônimo Monteiro	46	Cultivated	S 20° 46' 19.0"	W 41° 23' 17.5"	115
181_3	Jerônimo Monteiro	46	Cultivated	S 20° 46' 19.0"	W 41° 23' 17.5"	115
161_10	Jerônimo Monteiro	30	Cultivated	S 20° 49' 25.7"	W 41° 21' 32.6"	254
161_9	Jerônimo Monteiro	30	Cultivated	S 20° 49' 25.7"	W 41° 21' 32.6"	254
161_12	Jerônimo Monteiro	30	Cultivated	S 20° 49' 25.7"	W 41° 21' 32.6"	254
161_14	Jerônimo Monteiro	30	Cultivated	S 20° 49' 25.7"	W 41° 21' 32.6"	254
162	Jerônimo Monteiro	25	Cultivated	S 20° 49' 24.4"	W 41° 21' 32.5"	254
169_10	Jerônimo Monteiro	27	Cultivated	S20° 49' 32.7"	W 41° 21' 30.8"	208
170	Jerônimo Monteiro	35	Cultivated	S 20° 49' 19.0"	W 41° 21' 26.3"	194
170_4	Jerônimo Monteiro	35	Cultivated	S 20° 49' 19.0"	W 41° 21' 26.3"	194

170_5	Jerônimo Monteiro	35	Cultivated	S 20° 49' 19.0"	W 41° 21' 26.3"	194
172_5	Jerônimo Monteiro	35	Cultivated	S 20° 49' 18.4"	W 41° 21' 23.8"	192
172_4	Jerônimo Monteiro	35	Cultivated	S 20° 49' 18.4"	W 41° 21' 23.8"	192
172_3	Jerônimo Monteiro	35	Cultivated	S 20° 49' 18.4"	W 41° 21' 23.8"	192
172_6	Jerônimo Monteiro	35	Cultivated	S 20° 49' 18.4"	W 41° 21' 23.8"	192
172_7	Jerônimo Monteiro	35	Cultivated	S 20° 49' 18.4"	W 41° 21' 23.8"	192
172_10	Jerônimo Monteiro	35	Cultivated	S 20° 49' 18.4"	W 41° 21' 23.8"	192
172_8	Jerônimo Monteiro	35	Cultivated	S 20° 49' 18.4"	W 41° 21' 23.8"	192
164_4	Jerônimo Monteiro	25	Cultivated	S 20° 49' 23.3"	W 41° 21' 35.1"	267
164_5	Jerônimo Monteiro	25	Cultivated	S 20° 49' 23.3"	W 41° 21' 35.1"	267
164_6	Jerônimo Monteiro	25	Cultivated	S 20° 49' 23.3"	W 41° 21' 35.1"	267
164_7	Jerônimo Monteiro	25	Cultivated	S 20° 49' 23.3"	W 41° 21' 35.1"	267
164_8	Jerônimo Monteiro	25	Cultivated	S 20° 49' 23.3"	W 41° 21' 35.1"	267
164_10	Jerônimo Monteiro	25	Cultivated	S 20° 49' 23.3"	W 41° 21' 35.1"	267
165	Jerônimo Monteiro	32	Cultivated	S 20° 49' 20.7"	W 41° 21' 37.8"	265
165_2	Jerônimo Monteiro	32	Cultivated	S 20° 49' 20.7"	W 41° 21' 37.8"	265
165_3	Jerônimo Monteiro	32	Cultivated	S 20° 49' 20.7"	W 41° 21' 37.8"	265
165_4	Jerônimo Monteiro	32	Cultivated	S 20° 49' 20.7"	W 41° 21' 37.8"	265
166	Jerônimo Monteiro	32	Cultivated	S 20° 49' 18.6"	W 41° 21' 37.8"	265
166_2	Jerônimo Monteiro	32	Cultivated	S 20° 49' 18.6"	W 41° 21' 37.8"	265
143_2	Jerônimo Monteiro	40	Cultivated	S 20° 45' 36.6"	W 41° 21' 17.0"	142
143_3	Jerônimo Monteiro	40	Cultivated	S 20° 45' 36.6"	W 41° 21' 17.0"	142
144	Jerônimo Monteiro	40	Cultivated	S 20° 45' 36.6"	W 41° 21' 17.0"	142
144_2	Jerônimo Monteiro	40	Cultivated	S 20° 45' 36.6"	W 41° 21' 17.0"	142
146	Jerônimo Monteiro	40	Cultivated	S 20° 45' 36.6"	W 41° 21' 17.0"	142
171_2	Jerônimo Monteiro	35	Cultivated	S 20° 49' 19.4"	W 41° 21' 25.7"	194
171	Jerônimo Monteiro	35	Cultivated	S 20° 49' 19.4"	W 41° 21' 25.7"	194
180_7	Jerônimo Monteiro	46	Cultivated	S 20° 46' 18.3"	W 41° 23' 17.4"	115

180_6	Jerônimo Monteiro	46	Cultivated	S 20° 46' 18.3"	W 41° 23' 17.4"	115
179_6	Jerônimo Monteiro	40	Cultivated	S 20° 51' 02.9"	W 41° 22" 04.9"	131
179_5	Jerônimo Monteiro	40	Cultivated	S 20° 51' 02.9"	W 41° 22' 04.9"	131
179_4	Jerônimo Monteiro	40	Cultivated	S 20° 51' 02.9"	W 41° 22' 04.9"	131
179_3	Jerônimo Monteiro	40	Cultivated	S 20° 51' 02.9"	W 41° 22' 04.9"	131
179_2	Jerônimo Monteiro	40	Cultivated	S 20° 51' 02.9"	W 41° 22' 04.9"	131
179	Jerônimo Monteiro	40	Cultivated	S 20° 51' 02.9"	W 41° 22' 04.9"	131
149	Jerônimo Monteiro	40	Cultivated	S 20° 45' 36.6"	W 41° 21' 17.0"	141
168_2	Jerônimo Monteiro	27	Cultivated	S 20° 49' 32.4"	W 41° 21' 30.1"	205
167	Jerônimo Monteiro	27	Cultivated	S20° 49' 22.4"	W 41° 21' 43.6"	272
166_5	Jerônimo Monteiro	32	Cultivated	S 20° 49' 18.6"	W 41° 21' 37.8"	282
166_4	Jerônimo Monteiro	32	Cultivated	S 20° 49' 18.6"	W 41° 21' 37.8"	282
162_2	Jerônimo Monteiro	25	Cultivated	S 20° 49' 24.4"	W 41° 21' 32.5"	256
162_3	Jerônimo Monteiro	25	Cultivated	S 20° 49' 24.4"	W 41° 21' 32.5"	256
162_4	Jerônimo Monteiro	25	Cultivated	S 20° 49' 24.4"	W 41° 21' 32.5"	256
162_5	Jerônimo Monteiro	25	Cultivated	S 20° 49' 24.4"	W 41° 21' 32.5"	256
162_6	Jerônimo Monteiro	25	Cultivated	S 20° 49' 24.4"	W 41° 21' 32.5"	256
162_7	Jerônimo Monteiro	25	Cultivated	S 20° 49' 24.4"	W 41° 21' 32.5"	256
170_6	Jerônimo Monteiro	35	Cultivated	S 20° 49' 19.0"	W 41° 21' 26.3"	194
170_9	Jerônimo Monteiro	35	Cultivated	S 20° 49' 19.0"	W 41° 21' 26.3"	194
182_6	Jerônimo Monteiro	32	Cultivated	S 20° 46' 18.8"	W 41° 23' 17.4"	115
182_5	Jerônimo Monteiro	32	Cultivated	S 20° 46' 18.8"	W 41° 23' 17.4"	115
182_3	Jerônimo Monteiro	32	Cultivated	S 20° 46' 18.8"	W 41° 23' 17.4"	115
145	Jerônimo Monteiro	40	Cultivated	S 20° 45' 36.6"	W 41° 21' 17.0"	140
116	Jerônimo Monteiro	15	Cultivated	S 20° 47' 34.2"	W 41° 19' 02.9"	117
182_8	Jerônimo Monteiro	32	Cultivated	S 20° 46' 18.8"	W 41° 23' 17.4"	115
182_9	Jerônimo Monteiro	32	Cultivated	S 20° 46' 18.8"	W 41° 23' 17.4"	115
182_10	Jerônimo Monteiro	32	Cultivated	S 20° 46' 18.8"	W 41° 23' 17.4"	115

97	São José do Calçado	> 25	Cultivated	S 21° 00' 29.3"	W 41° 38"' 51.0"	341
95	São José do Calçado	30	Cultivated	S 21° 02' 55.8"	W 41° 39' 17.4"	312
95_1	São José do Calçado	30	Cultivated	S 21° 02' 55.8"	W 41° 39' 17.4"	312
94	São José do Calçado	30	Cultivated	S 21° 02' 55.0"	W 41° 39' 17.7"	307
91	São José do Calçado	35	Cultivated	S 21° 03" 57'.2"	W 41° 40" 05.9"	308
98	São José do Calçado	> 25	Cultivated	S 21° 00' 29.3"	W 41° 38" 51.1"	342
100	São José do Calçado	> 25	Cultivated	S 21° 00' 30.9"	W 41° 38" 50.3"	344
102	São José do Calçado	20	Cultivated	S 20° 59' 13.2"	W 41° 38" 28.6"	360
103	São José do Calçado	20	Cultivated	S 20° 58' 06.4"	W 41° 38" 05.5"	359
104	São José do Calçado	18	Cultivated	S 20° 57" 48'.3"	W 41° 38" 12.1"	390
105	São José do Calçado	18	Cultivated	S 20° 57" 48'.5"	W 41° 38" 12.5"	390
81	São José do Calçado	35	Cultivated	S 21° 03" 58.4"	W 41° 40" 05.6"	304
83	São José do Calçado	35	Cultivated	S 21° 03' 58.6"	W 41° 40' 05.6"	304
85	São José do Calçado	35	Cultivated	S 21° 03' 59.2"	W 41° 40' 05.6"	302
86	São José do Calçado	35	Cultivated	S 21° 03' 59.3"	W0 41° 40' 05.4"	302
113	São José do Calçado	18	Cultivated	S 20° 57' 52.6"	W 41° 38' 13.3"	408
112	São José do Calçado	18	Cultivated	S 20° 57' 52.4"	W 41° 38' 13.3"	405
111	São José do Calçado	18	Cultivated	S 20° 57' 51.8"	W 41° 38' 13.7"	410
114	São José do Calçado	18	Cultivated	S 20° 57' 52.8"	W 41° 38' 12.2"	399
4	-	-	Cultivated	-	-	-
6	-	-	Cultivated	-	-	-
7	-	-	Cultivated	-	-	-
21	-	-	Cultivated	-	-	-
22	-	-	Cultivated	-	-	-
23	-	-	Cultivated	-	-	-
24	-	-	Cultivated	-	-	-
26	-	-	Cultivated	-	-	-
27	-	-	Cultivated	-	-	-

29	-	-	Cultivated	-	-	-
31	-	-	Cultivated	-	-	-
31	-	-	Cultivated	-	-	-
36	-	-	Cultivated	-	-	-
39	-	-	Cultivated	-	-	-
40	-	-	Cultivated	-	-	-
45	-	-	Cultivated	-	-	-
46	-	-	Cultivated	-	-	-
53	-	-	Cultivated	-	-	-
54	-	-	Cultivated	-	-	-
57	-	-	Cultivated	-	-	-
58	-	-	Cultivated	-	-	-
59	-	-	Cultivated	-	-	-
60	-	-	Cultivated	-	-	-
61	-	-	Cultivated	-	-	-
64	-	-	Cultivated	-	-	-
75	-	-	Cultivated	-	-	-
76	-	-	Cultivated	-	-	-
122	-	-	Cultivated	-	-	-
281	-	-	Cultivated	-	-	-
300	-	-	Cultivated	-	-	-
27_8	-	-	Cultivated	-	-	-
28_11	-	-	Cultivated	-	-	-
28_12	-	-	Cultivated	-	-	-
28_2	-	-	Cultivated	-	-	-
28_4	-	-	Cultivated	-	-	-
281_2	-	-	Cultivated	-	-	-
281_3	-	-	Cultivated	-	-	-

281_6	-	-	Cultivated	-	-	-
281_8	-	-	Cultivated	-	-	-
29_2	-	-	Cultivated	-	-	-
29_3	-	-	Cultivated	-	-	-
29_5	-	-	Cultivated	-	-	-
290_4	-	-	Cultivated	-	-	-
31_2	-	-	Cultivated	-	-	-
31_4	-	-	Cultivated	-	-	-
31_6	-	-	Cultivated	-	-	-
32_10	-	-	Cultivated	-	-	-
32_11	-	-	Cultivated	-	-	-
32_4	-	-	Cultivated	-	-	-
32_8	-	-	Cultivated	-	-	-
32_9	-	-	Cultivated	-	-	-
34_2	-	-	Cultivated	-	-	-
36_2	-	-	Cultivated	-	-	-
G1	-	-	Cultivated	-	-	-
G10	-	-	Cultivated	-	-	-
G100	-	-	Cultivated	-	-	-
G101	-	-	Cultivated	-	-	-
G102	-	-	Cultivated	-	-	-
G103	-	-	Cultivated	-	-	-
G104	-	-	Cultivated	-	-	-
G105	-	-	Cultivated	-	-	-
G106	-	-	Cultivated	-	-	-
G107	-	-	Cultivated	-	-	-
G108	-	-	Cultivated	-	-	-
G109	-	-	Cultivated	-	-	-

G11	-	-	Cultivated	-	-	-
G110	-	-	Cultivated	-	-	-
G111	-	-	Cultivated	-	-	-
G112	-	-	Cultivated	-	-	-
G113	-	-	Cultivated	-	-	-
G114	-	-	Cultivated	-	-	-
G115	-	-	Cultivated	-	-	-
G116	-	-	Cultivated	-	-	-
G117	-	-	Cultivated	-	-	-
G118	-	-	Cultivated	-	-	-
G119	-	-	Cultivated	-	-	-
G12	-	-	Cultivated	-	-	-
G120	-	-	Cultivated	-	-	-
G121	-	-	Cultivated	-	-	-
G122	-	-	Cultivated	-	-	-
G123	-	-	Cultivated	-	-	-
G124	-	-	Cultivated	-	-	-
G125	-	-	Cultivated	-	-	-
G126	-	-	Cultivated	-	-	-
G127	-	-	Cultivated	-	-	-
G128	-	-	Cultivated	-	-	-
G129	-	-	Cultivated	-	-	-
G13	-	-	Cultivated	-	-	-
G130	-	-	Cultivated	-	-	-
G131	-	-	Cultivated	-	-	-
G132	-	-	Cultivated	-	-	-
G133	-	-	Cultivated	-	-	-
G134	-	-	Cultivated	-	-	-

G135	-	-	Cultivated	-	-	-
G136	-	-	Cultivated	-	-	-
G137	-	-	Cultivated	-	-	-
G138	-	-	Cultivated	-	-	-
G139	-	-	Cultivated	-	-	-
G14	-	-	Cultivated	-	-	-
G140_1	-	-	Cultivated	-	-	-
G140_1	-	-	Cultivated	-	-	-
G141	-	-	Cultivated	-	-	-
G142	-	-	Cultivated	-	-	-
G143	-	-	Cultivated	-	-	-
G144	-	-	Cultivated	-	-	-
G145	-	-	Cultivated	-	-	-
G146	-	-	Cultivated	-	-	-
G147	-	-	Cultivated	-	-	-
G148	-	-	Cultivated	-	-	-
G149	-	-	Cultivated	-	-	-
G15	-	-	Cultivated	-	-	-
G150	-	-	Cultivated	-	-	-
G151	-	-	Cultivated	-	-	-
G153	-	-	Cultivated	-	-	-
G154	-	-	Cultivated	-	-	-
G155	-	-	Cultivated	-	-	-
G156	-	-	Cultivated	-	-	-
G157	-	-	Cultivated	-	-	-
G16	-	-	Cultivated	-	-	-
G160	-	-	Cultivated	-	-	-
G161	-	-	Cultivated	-	-	-

G162	-	-	Cultivated	-	-	-
G17	-	-	Cultivated	-	-	-
G18	-	-	Cultivated	-	-	-
G19	-	-	Cultivated	-	-	-
G2	-	-	Cultivated	-	-	-
G20	-	-	Cultivated	-	-	-
G207	-	-	Cultivated	-	-	-
G21	-	-	Cultivated	-	-	-
G22	-	-	Cultivated	-	-	-
G23	-	-	Cultivated	-	-	-
G24	-	-	Cultivated	-	-	-
G25	-	-	Cultivated	-	-	-
G26	-	-	Cultivated	-	-	-
G27	-	-	Cultivated	-	-	-
G28	-	-	Cultivated	-	-	-
G3	-	-	Cultivated	-	-	-
G30	-	-	Cultivated	-	-	-
G300	-	-	Cultivated	-	-	-
G31	-	-	Cultivated	-	-	-
G32	-	-	Cultivated	-	-	-
G33	-	-	Cultivated	-	-	-
G34	-	-	Cultivated	-	-	-
G35	-	-	Cultivated	-	-	-
G36	-	-	Cultivated	-	-	-
G37	-	-	Cultivated	-	-	-
G38	-	-	Cultivated	-	-	-
G39	-	-	Cultivated	-	-	-
G4	-	-	Cultivated	-	-	-

G40	-	-	Cultivated	-	-	-
G42	-	-	Cultivated	-	-	-
G43	-	-	Cultivated	-	-	-
G44	-	-	Cultivated	-	-	-
G45	-	-	Cultivated	-	-	-
G46	-	-	Cultivated	-	-	-
G47	-	-	Cultivated	-	-	-
G48	-	-	Cultivated	-	-	-
G49	-	-	Cultivated	-	-	-
G5	-	-	Cultivated	-	-	-
G50	-	-	Cultivated	-	-	-
G51	-	-	Cultivated	-	-	-
G52	-	-	Cultivated	-	-	-
G53	-	-	Cultivated	-	-	-
G54	-	-	Cultivated	-	-	-
G55	-	-	Cultivated	-	-	-
G56	-	-	Cultivated	-	-	-
G57	-	-	Cultivated	-	-	-
G59	-	-	Cultivated	-	-	-
G6	-	-	Cultivated	-	-	-
G60	-	-	Cultivated	-	-	-
G61	-	-	Cultivated	-	-	-
G62	-	-	Cultivated	-	-	-
G63	-	-	Cultivated	-	-	-
G64	-	-	Cultivated	-	-	-
G65	-	-	Cultivated	-	-	-
G66	-	-	Cultivated	-	-	-
G67	-	-	Cultivated	-	-	-

G68	-	-	Cultivated	-	-	-
G69	-	-	Cultivated	-	-	-
G7	-	-	Cultivated	-	-	-
G70	-	-	Cultivated	-	-	-
G71	-	-	Cultivated	-	-	-
G72	-	-	Cultivated	-	-	-
G73	-	-	Cultivated	-	-	-
G74	-	-	Cultivated	-	-	-
G75	-	-	Cultivated	-	-	-
G76	-	-	Cultivated	-	-	-
G77	-	-	Cultivated	-	-	-
G78	-	-	Cultivated	-	-	-
G79	-	-	Cultivated	-	-	-
G8	-	-	Cultivated	-	-	-
G80	-	-	Cultivated	-	-	-
G81	-	-	Cultivated	-	-	-
G82	-	-	Cultivated	-	-	-
G83	-	-	Cultivated	-	-	-
G84	-	-	Cultivated	-	-	-
G85	-	-	Cultivated	-	-	-
G88	-	-	Cultivated	-	-	-
G89	-	-	Cultivated	-	-	-
G9	-	-	Cultivated	-	-	-
G90	-	-	Cultivated	-	-	-
G91	-	-	Cultivated	-	-	-
G92	-	-	Cultivated	-	-	-
G93	-	-	Cultivated	-	-	-
G94	-	-	Cultivated	-	-	-

G95	-	-	Cultivated	-	-	-
G96	-	-	Cultivated	-	-	-
G97	-	-	Cultivated	-	-	-
G98	-	-	Cultivated	-	-	-
G99	-	-	Cultivated	-	-	-
G400	-	-	Cultivated	-	-	-
A1	Tributun (variety)	-	Commercial	-	-	-
Verdim	-	-	Commercial	-	-	-
BRS	-	-	Commercial	-	-	-
RO	-	-	Commercial	-	-	-
P2	Monte Pascoal (variety)	-	Commercial	-	-	-

Table S2. List of *Coffea canephora* genotypes evaluated with DArTseq™ SNP markers

Accession	ID in structure	Origin	Crop age (years)*	Commercial/ Cultivated	Latitude	Longitude	Altitude
176.5	38	Alegre	40	Cultivated	S 20° 45' 08.9"	W 41° 25' 49.4"	117
176.6	82	Alegre	40	Cultivated	S 20° 45' 08.9"	W 41° 25' 49.4"	117
201.2	52	Alegre	35	Cultivated	S 20° 43' 47.1"	W 41° 27' 12.7"	364
201.6	77	Alegre	35	Cultivated	S 20° 43' 47.1"	W 41° 27' 12.7"	364
203.2	166	Alegre	25	Cultivated	S 20° 43' 33.9"	W 41° 26' 32.2"	243
125	226	Cachoeiro de Itapemirim	40	Cultivated	S 20° 48" 27.0"	W 41° 18' 43.1"	143
126	163	Cachoeiro de Itapemirim	40	Cultivated	S 20° 48" 26.8"	W 41° 18' 43.3"	143
134	111	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 50.8"	W 41° 19" 39.2"	152
187	159	Cachoeiro de Itapemirim	< 20	Cultivated	S 20° 43' 35.4"	W 41° 20" 59.7"	130
189	43	Cachoeiro de Itapemirim	< 20	Cultivated	S 20° 43' 35.6"	W 41° 20" 59.6"	133
192	259	Cachoeiro de Itapemirim	< 20	Cultivated	S 20° 43' 36.3"	W 41° 21' 00.1"	137
131.3	74	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 51'.7"	W 41° 19" 39.6"	168
132.2	99	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 51'. 3"	W 41° 19" 39. 8"	157
136.18	119	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 50.1"	W 41° 19' 40'.3"	151
136.18.1	236	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 50.1"	W 41° 19' 40'.3"	151
136.19	223	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 50.1"	W 41° 19' 40'.3"	151
136.26	199	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 50.1"	W 41° 19' 40'.3"	151
136.27	262	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 50.1"	W 41° 19' 40'.3"	151
136.3	135	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 50.1""	W 41° 19" 40'.3"	153
136.32	106	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 50.1"	W 41° 19' 40'.3"	151
136.33	21	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 50.1"	W 41° 19' 40'.3"	151
136.34	34	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 50.1"	W 41° 19' 40'.3"	151
136.5	47	Cachoeiro de Itapemirim	32	Cultivated	S 20° 49' 50.1""	W 41° 19" 40'.3"	153
183.4	58	Cachoeiro de Itapemirim	< 20	Cultivated	S 20° 43' 33.9"	W 41° 20" 59.6"	120
186.2	20	Cachoeiro de Itapemirim	< 20	Cultivated	S 20° 43' 35.3"	W 41° 20' 59.9"	129
186.3	44	Cachoeiro de Itapemirim	< 20	Cultivated	S 20° 43' 35.3"	W 41° 20' 59.9"	129
186.41	184	Cachoeiro de Itapemirim	< 20	Cultivated	S 20° 43' 35.3"	W 41° 20' 59.9"	129
186.5	157	Cachoeiro de Itapemirim	< 20	Cultivated	S 20° 43' 35.3"	W 41° 20' 59.9"	129
186.6	204	Cachoeiro de Itapemirim	< 20	Cultivated	S 20° 43' 35.3"	W 41° 20' 59.9"	129
189.2	158	Cachoeiro de Itapemirim	< 20	Cultivated	S 20° 43' 35.6"	W 41° 20" 59.6"	133
189.4	84	Cachoeiro de Itapemirim	< 20	Cultivated	S 20° 43' 35.6"	W 41° 20" 59.6"	133
192.2	215	Cachoeiro de Itapemirim	< 20	Cultivated	S 20° 43' 36.3"	W 41° 21' 00.1"	137
192.24	187	Cachoeiro de Itapemirim	< 20	Cultivated	S 20° 43' 36.3""	W 41° 21' 00.1"	139
192.3	145	Cachoeiro de Itapemirim	< 20	Cultivated	S 20° 43' 36.3"	W 41° 21' 00.1"	137

192.4	133	Cachoeiro de Itapemirim	< 20	Cultivated	S 20° 43' 36.3"	W 41° 21' 00.1"	137
144	30	Jerônimo Monteiro	40	Cultivated	S 20° 45' 36.6"	W 41° 21' 17.0"	142
145	45	Jerônimo Monteiro	40	Cultivated	S 20° 45' 36.6"	W 41° 21' 17.0"	140
146	7	Jerônimo Monteiro	40	Cultivated	S 20° 45' 36.6"	W 41° 21' 17.0"	142
149	178	Jerônimo Monteiro	40	Cultivated	S 20° 45' 36.6"	W 41° 21' 17.0"	141
162	152	Jerônimo Monteiro	30	Cultivated	S 20° 49' 24.4"	W 41° 21' 32.5"	254
166	18	Jerônimo Monteiro	32	Cultivated	S20° 49' 32.7"	W 41° 21" 30.8"	207
169	186	Jerônimo Monteiro	27	Cultivated	S 20° 49' 19.0"	W 41° 21' 26.3"	194
170	54	Jerônimo Monteiro	35	Cultivated	S 20° 49' 19.4"	W 41° 21' 25.7"	194
171	127	Jerônimo Monteiro	35	Cultivated	S 20° 49' 19.4"	W 41° 21' 25.7"	194
179	189	Jerônimo Monteiro	40	Cultivated	S 20° 51' 02.9"	W 41° 22' 04.9"	131
144.2	214	Jerônimo Monteiro	40	Cultivated	S 20° 45' 36.6"	W 41° 21' 17.0"	142
149.1	202	Jerônimo Monteiro	40	Cultivated	S 20° 45' 36.6"	W 41° 21' 17.0"	141
151.8	80	Jerônimo Monteiro	> 25	Cultivated	S 20° 45' 25.3"	W 41° 21' 42.6"	111
151.9	176	Jerônimo Monteiro	> 25	Cultivated	S 20° 45' 25.3"	W 41° 21' 42.6"	111
161.2	29	Jerônimo Monteiro	30	Cultivated	S 20° 49' 25.7"	W 41° 21" 32.6"	252
161.7	113	Jerônimo Monteiro	30	Cultivated	S 20° 49' 25.7"	W 41° 21" 32.6"	252
161.9	94	Jerônimo Monteiro	30	Cultivated	S 20° 49' 25.7"	W 41° 21' 32.6"	254
162.10	95	Jerônimo Monteiro	25	Cultivated	S 20° 49' 24.4"	W 41° 21" 32.5"	256
162.10.1	237	Jerônimo Monteiro	25	Cultivated	S 20° 49' 24.4"	W 41° 21" 32.5"	256
162.2	36	Jerônimo Monteiro	25	Cultivated	S 20° 49' 24.4"	W 41° 21' 32.5"	256
162.6	161	Jerônimo Monteiro	25	Cultivated	S 20° 49' 24.4"	W 41° 21' 32.5"	256
162.7	162	Jerônimo Monteiro	25	Cultivated	S 20° 49' 24.4"	W 41° 21' 32.5"	256
164.10	139	Jerônimo Monteiro	25	Cultivated	S 20° 49' 23.3"	W 41° 21' 35.1"	267
164.4	190	Jerônimo Monteiro	25	Cultivated	S 20° 49' 23.3"	W 41° 21' 35.1"	267
164.7	93	Jerônimo Monteiro	25	Cultivated	S 20° 49' 23.3"	W 41° 21' 35.1"	267
164.8	165	Jerônimo Monteiro	25	Cultivated	S 20° 49' 23.3"	W 41° 21' 35.1"	267
165.3	65	Jerônimo Monteiro	32	Cultivated	S 20° 49' 20.7"	W 41° 21' 37.8"	265
165.4	144	Jerônimo Monteiro	32	Cultivated	S 20° 49' 20.7"	W 41° 21' 37.8"	265
166.2	75	Jerônimo Monteiro	32	Cultivated	S 20° 49' 18.6"	W 41° 21' 37.8"	265
166.2.1	272	Jerônimo Monteiro	32	Cultivated	S 20° 49' 18.6"	W 41° 21' 37.8"	265
166.5	9	Jerônimo Monteiro	32	Cultivated	S 20° 49' 18.6"	W 41° 21' 37.8"	282
168.10	26	Jerônimo Monteiro	27	Cultivated	S 20° 49' 32.4"	W 41° 21" 30.1"	205
168.3	61	Jerônimo Monteiro	27	Cultivated	S 20° 49' 32.4"	W 41° 21" 30.1"	204
168.5	28	Jerônimo Monteiro	27	Cultivated	S 20° 49' 32.4"	W 41° 21" 30.1"	205
168.6	124	Jerônimo Monteiro	27	Cultivated	S 20° 49' 32.4"	W 41° 21" 30.1"	205

168.9	35	Jerônimo Monteiro	27	Cultivated	S 20° 49' 32.4"	W 41° 21" 30.1"	205
169.10	41	Jerônimo Monteiro	27	Cultivated	S20° 49' 32.7"	W 41° 21' 30.8"	208
169.4	117	Jerônimo Monteiro	27	Cultivated	S20° 49' 32.7"	W 41° 21" 30.8"	207
169.5	249	Jerônimo Monteiro	27	Cultivated	S20° 49' 32.7"	W 41° 21" 30.8"	207
170.6	149	Jerônimo Monteiro	35	Cultivated	S 20° 49' 19.0"	W 41° 21' 26.3"	194
171.10	37	Jerônimo Monteiro	35	Cultivated	S 20° 49' 19.4"	W 41° 21" 25.7"	194
171.2	53	Jerônimo Monteiro	35	Cultivated	S 20° 49' 19.4"	W 41° 21' 25.7"	194
171.6	25	Jerônimo Monteiro	35	Cultivated	S 20° 49' 19.4"	W 41° 21" 25.7"	194
171.9	72	Jerônimo Monteiro	35	Cultivated	S 20° 49' 19.4"	W 41° 21" 25.7"	194
172.3	175	Jerônimo Monteiro	35	Cultivated	S 20° 49' 18.4"	W 41° 21' 23.8"	192
172.4	254	Jerônimo Monteiro	35	Cultivated	S 20° 49' 18.4"	W 41° 21' 23.8"	192
172.5	89	Jerônimo Monteiro	35	Cultivated	S 20° 49' 18.4"	W 41° 21' 23.8"	192
172.6	86	Jerônimo Monteiro	35	Cultivated	S 20° 49' 18.4"	W 41° 21' 23.8"	192
172.8	6	Jerônimo Monteiro	35	Cultivated	S 20° 49' 18.4"	W 41° 21' 23.8"	192
173.3	172	Jerônimo Monteiro	35	Cultivated	S 20° 49' 17.2"	W 41° 21' 36.7"	280
173.6	116	Jerônimo Monteiro	35	Cultivated	S 20° 49' 17.2"	W 41° 21' 36.7"	280
173.8	66	Jerônimo Monteiro	35	Cultivated	S 20° 49' 17.2""	W 41° 21" 26.7"	199
179.2	115	Jerônimo Monteiro	40	Cultivated	S 20° 51' 02.9"	W 41° 22' 04.9"	131
179.2.1	238	Jerônimo Monteiro	40	Cultivated	S 20° 51' 02.9"	W 41° 22' 04.9"	131
179.3	8	Jerônimo Monteiro	40	Cultivated	S 20° 51' 02.9"	W 41° 22' 04.9"	131
179.4	68	Jerônimo Monteiro	40	Cultivated	S 20° 51' 02.9"	W 41° 22' 04.9"	131
179.5	169	Jerônimo Monteiro	40	Cultivated	S 20° 51' 02.9"	W 41° 22' 04.9"	131
180.7	48	Jerônimo Monteiro	46	Cultivated	S 20° 46' 18.3"	W 41° 23' 17.4"	115
181.3	225	Jerônimo Monteiro	46	Cultivated	S 20° 46' 19.0"	W 41° 23' 17.5"	115
181.4	83	Jerônimo Monteiro	46	Cultivated	S 20° 46' 19.0"	W 41° 23' 17.5"	115
181.4.1	261	Jerônimo Monteiro	46	Cultivated	S 20° 46' 19.0"	W 41° 23' 17.5"	115
182.10	264	Jerônimo Monteiro	46	Cultivated	S 20° 46' 18.8"	W 41° 23' 17.4"	115
182.3	207	Jerônimo Monteiro	46	Cultivated	S 20° 46' 18.8"	W 41° 23' 17.4"	115
182.5	219	Jerônimo Monteiro	46	Cultivated	S 20° 46' 18.8"	W 41° 23' 17.4"	115
182.8	253	Jerônimo Monteiro	46	Cultivated	S 20° 46' 18.8"	W 41° 23' 17.4"	115
182.9	192	Jerônimo Monteiro	46	Cultivated	S 20° 46' 18.8"	W 41° 23' 17.4"	115
83	126	São José do Calçado	35	Cultivated	S 21° 03' 58.6"	W 41° 40' 05.6"	304
85	227	São José do Calçado	35	Cultivated	S 21° 03' 59.2"	W 41° 40' 05.6"	302
86	193	São José do Calçado	35	Cultivated	S 21° 03' 59.3"	W0 41° 40' 05.4"	302
91	23	São José do Calçado	35	Cultivated	S 21° 03" 57'.2"	W 41° 40" 05.9"	308
95	81	São José do Calçado	30	Cultivated	S 21° 02' 55.8"	W 41° 39' 17.4"	312

97	110	São José do Calçado	> 25	Cultivated	S 21° 00' 29.3"	W 41° 38"' 51.0"	341
98	136	São José do Calçado	> 25	Cultivated	S 21° 00' 29.3"	W 41° 38" 51.1"	342
100	148	São José do Calçado	> 25	Cultivated	S 21° 00' 30.9"	W 41° 38" 50.3"	344
103	271	São José do Calçado	20	Cultivated	S 20° 58' 06.4"	W 41° 38" 05.5"	359
111	97	São José do Calçado	20	Cultivated	S 20° 57' 51.8"	W 41° 38' 13.7"	410
112	229	São José do Calçado	20	Cultivated	S 20° 57' 52.4"	W 41° 38' 13.3"	405
113	239	São José do Calçado	20	Cultivated	S 20° 57' 52.6"	W 41° 38' 13.3"	408
114	188	São José do Calçado	20	Cultivated	S 20° 57' 52.8"	W 41° 38' 12.2"	399
112.1	276	São José do Calçado	18	Cultivated	S 20° 57' 52.4"	W 41° 38' 13.3"	405
114.1	273	São José do Calçado	18	Cultivated	S 20° 57' 52.8"	W 41° 38' 12.2"	399
6	147	Unknown	-	Cultivated	-	-	-
21	230	Unknown	-	Cultivated	-	-	-
22	3	Unknown	-	Cultivated	-	-	-
24	79	Unknown	-	Cultivated	-	-	-
31	155	Unknown	-	Cultivated	-	-	-
40	206	Unknown	-	Cultivated	-	-	-
46	263	Unknown	-	Cultivated	-	-	-
53	90	Unknown	-	Cultivated	-	-	-
54	120	Unknown	-	Cultivated	-	-	-
58	170	Unknown	-	Cultivated	-	-	-
59	112	Unknown	-	Cultivated	-	-	-
60	100	Unknown	-	Cultivated	-	-	-
61	181	Unknown	-	Cultivated	-	-	-
62	49	Unknown	-	Cultivated	-	-	-
64	62	Unknown	-	Cultivated	-	-	-
75	107	Unknown	-	Cultivated	-	-	-
300	39	Unknown	-	Cultivated	-	-	-
28.2	151	Unknown	-	Cultivated	-	-	-
281.2	242	Unknown	-	Cultivated	-	-	-
281.6	32	Unknown	-	Cultivated	-	-	-
281.8	96	Unknown	-	Cultivated	-	-	-
29.2	209	Unknown	-	Cultivated	-	-	-
290.4	121	Unknown	-	Cultivated	-	-	-
31.2	240	Unknown	-	Cultivated	-	-	-
31.4	143	Unknown	-	Cultivated	-	-	-
32.10	201	Unknown	-	Cultivated	-	-	-

32.4	50	Unknown	-	Cultivated	-	-	-
32.4.1	250	Unknown	-	Cultivated	-	-	-
32.8	92	Unknown	-	Cultivated	-	-	-
34.2	251	Unknown	-	Cultivated	-	-	-
39.2	98	Unknown	-	Cultivated	-	-	-
60.1	213	Unknown	-	Cultivated	-	-	-
95.1	19	Unknown	-	Cultivated	-	-	-
95.1.1	260	Unknown	-	Cultivated	-	-	-
G1	2	Unknown	-	Cultivated	-	-	-
G10	15	Unknown	-	Cultivated	-	-	-
G102	154	Unknown	-	Cultivated	-	-	-
G103	265	Unknown	-	Cultivated	-	-	-
G104	88	Unknown	-	Cultivated	-	-	-
G109	108	Unknown	-	Cultivated	-	-	-
G11	134	Unknown	-	Cultivated	-	-	-
G110	130	Unknown	-	Cultivated	-	-	-
G111	142	Unknown	-	Cultivated	-	-	-
G115	174	Unknown	-	Cultivated	-	-	-
G117	203	Unknown	-	Cultivated	-	-	-
G118	104	Unknown	-	Cultivated	-	-	-
G12	55	Unknown	-	Cultivated	-	-	-
G120	138	Unknown	-	Cultivated	-	-	-
G120.1	269	Unknown	-	Cultivated	-	-	-
G127	22	Unknown	-	Cultivated	-	-	-
G127.1	212	Unknown	-	Cultivated	-	-	-
G128	247	Unknown	-	Cultivated	-	-	-
G129	235	Unknown	-	Cultivated	-	-	-
G130	191	Unknown	-	Cultivated	-	-	-
G131	211	Unknown	-	Cultivated	-	-	-
G132	270	Unknown	-	Cultivated	-	-	-
G133	131	Unknown	-	Cultivated	-	-	-
G135	275	Unknown	-	Cultivated	-	-	-
G136	217	Unknown	-	Cultivated	-	-	-
G137	194	Unknown	-	Cultivated	-	-	-
G138	205	Unknown	-	Cultivated	-	-	-
G14	67	Unknown	-	Cultivated	-	-	-

G14.1	248	Unknown	-	Cultivated	-	-	-
G140	228	Unknown	-	Cultivated	-	-	-
G141	274	Unknown	-	Cultivated	-	-	-
G143	218	Unknown	-	Cultivated	-	-	-
G144	241	Unknown	-	Cultivated	-	-	-
G145	243	Unknown	-	Cultivated	-	-	-
G146	109	Unknown	-	Cultivated	-	-	-
G147	179	Unknown	-	Cultivated	-	-	-
G148	216	Unknown	-	Cultivated	-	-	-
G149	231	Unknown	-	Cultivated	-	-	-
G15	123	Unknown	-	Cultivated	-	-	-
G151	5	Unknown	-	Cultivated	-	-	-
G153	252	Unknown	-	Cultivated	-	-	-
G154	114	Unknown	-	Cultivated	-	-	-
G155	200	Unknown	-	Cultivated	-	-	-
G156	132	Unknown	-	Cultivated	-	-	-
G16	73	Unknown	-	Cultivated	-	-	-
G17	46	Unknown	-	Cultivated	-	-	-
G19	11	Unknown	-	Cultivated	-	-	-
G2	122	Unknown	-	Cultivated	-	-	-
G20	180	Unknown	-	Cultivated	-	-	-
G23	31	Unknown	-	Cultivated	-	-	-
G25	70	Unknown	-	Cultivated	-	-	-
G26	40	Unknown	-	Cultivated	-	-	-
G27	14	Unknown	-	Cultivated	-	-	-
G3	177	Unknown	-	Cultivated	-	-	-
G32	76	Unknown	-	Cultivated	-	-	-
G34	87	Unknown	-	Cultivated	-	-	-
G37	91	Unknown	-	Cultivated	-	-	-
G38	182	Unknown	-	Cultivated	-	-	-
G41	63	Unknown	-	Cultivated	-	-	-
G42	173	Unknown	-	Cultivated	-	-	-
G43	105	Unknown	-	Cultivated	-	-	-
G44	140	Unknown	-	Cultivated	-	-	-
G46	17	Unknown	-	Cultivated	-	-	-
G50	183	Unknown	-	Cultivated	-	-	-

G51	69	Unknown	-	Cultivated	-	-	-
G52	103	Unknown	-	Cultivated	-	-	-
G55	164	Unknown	-	Cultivated	-	-	-
G56	185	Unknown	-	Cultivated	-	-	-
G57	13	Unknown	-	Cultivated	-	-	-
G59	78	Unknown	-	Cultivated	-	-	-
G59.1	156	Unknown	-	Cultivated	-	-	-
G60	57	Unknown	-	Cultivated	-	-	-
G61	195	Unknown	-	Cultivated	-	-	-
G62	33	Unknown	-	Cultivated	-	-	-
G63	56	Unknown	-	Cultivated	-	-	-
G64	125	Unknown	-	Cultivated	-	-	-
G67	42	Unknown	-	Cultivated	-	-	-
G68	101	Unknown	-	Cultivated	-	-	-
G69	12	Unknown	-	Cultivated	-	-	-
G7	85	Unknown	-	Cultivated	-	-	-
G70	171	Unknown	-	Cultivated	-	-	-
G72	137	Unknown	-	Cultivated	-	-	-
G73	141	Unknown	-	Cultivated	-	-	-
G74	129	Unknown	-	Cultivated	-	-	-
G75	60	Unknown	-	Cultivated	-	-	-
G77	4	Unknown	-	Cultivated	-	-	-
G78	150	Unknown	-	Cultivated	-	-	-
G79	16	Unknown	-	Cultivated	-	-	-
G8	71	Unknown	-	Cultivated	-	-	-
G80	51	Unknown	-	Cultivated	-	-	-
G9	146	Unknown	-	Cultivated	-	-	-
G91	59	Unknown	-	Cultivated	-	-	-
G92	24	Unknown	-	Cultivated	-	-	-
G92.1	224	Unknown	-	Cultivated	-	-	-
G93	27	Unknown	-	Cultivated	-	-	-
G95	118	Unknown	-	Cultivated	-	-	-
G96	167	Unknown	-	Cultivated	-	-	-
G97	160	Unknown	-	Cultivated	-	-	-
G98	102	Unknown	-	Cultivated	-	-	-
G99	10	Unknown	-	Cultivated	-	-	-

120.5.1	280	Unknown	-	Cultivated	-	-	-
A1	1	Tributun (variety)	-	Commercial	-	-	-
Verdim	64	Tributun (variety)	-	Commercial	-	-	-
BRS	128	-	-	Commercial	-	-	-
RO	153	-	-	Commercial	-	-	-
P2	168	Monte Pascoal (variety)	-	Commercial	-	-	-
Centenário P1	210	Incaper Cultivars	-	Commercial	-	-	-
Centenário P2	222	Incaper Cultivars	-	Commercial	-	-	-
Centenário P3	234	Incaper Cultivars	-	Commercial	-	-	-
Centenário P4	246	Incaper Cultivars	-	Commercial	-	-	-
Centenário P7	258	Incaper Cultivars	-	Commercial	-	-	-
Diamante P4	244	Incaper Cultivars	-	Commercial	-	-	-
Diamante P8	256	Incaper Cultivars	-	Commercial	-	-	-
Jequitibá P1	245	Incaper Cultivars	-	Commercial	-	-	-
Jequitibá P2	257	Incaper Cultivars	-	Commercial	-	-	-
Jequitibá P3	268	Incaper Cultivars	-	Commercial	-	-	-
Jequitibá P4	279	Incaper Cultivars	-	Commercial	-	-	-
Jequitibá P6	198	Incaper Cultivars	-	Commercial	-	-	-
Marilândia P11	255	Incaper Cultivars	-	Commercial	-	-	-
Marilândia P7	266	Incaper Cultivars	-	Commercial	-	-	-
Robustão P1	277	Incaper Cultivars	-	Commercial	-	-	-
Robustão P2	196	Incaper Cultivars	-	Commercial	-	-	-
Robustão P3	220	Incaper Cultivars	-	Commercial	-	-	-
Robustão P4	208	Incaper Cultivars	-	Commercial	-	-	-
Robustão P5	232	Incaper Cultivars	-	Commercial	-	-	-
Vitória P1	267	Incaper Cultivars	-	Commercial	-	-	-
Vitória P2	278	Incaper Cultivars	-	Commercial	-	-	-
Vitória P3	197	Incaper Cultivars	-	Commercial	-	-	-
Vitória P4	221	Incaper Cultivars	-	Commercial	-	-	-
Vitória P5	233	Incaper Cultivars	-	Commercial	-	-	-

Table S3. Chromosomal distribution of 2,542 polymorphic SNPs used to genotype 280 *C. canephora* genotypes, including mean polymorphism information content (PIC), expected heterozygosity (HE), observed heterozygosity (HO), and fixation index (F). [1]National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/genome/?term=coffea+canephora)

Chromosome	PIC	HE	HO	F	N° of SNPs	Chromosome size (Mb)[1]	N° of genes[1]
Chr 1	0.20	0.13	0.16	0.23	216	38.19	2,198
Chr 2	0.19	0.11	0.14	0.21	158	27.62	1,653
Chr 3	0.20	0.12	0.16	0.25	141	33.54	1,753
Chr 4	0.19	0.11	0.15	0.27	342	54.52	4,000
Chr 5	0.20	0.12	0.17	0.29	129	32.03	1,632
Chr 6	0.18	0.12	0.14	0.14	143	28.19	1,727
Chr 7	0.19	0.12	0.15	0.20	165	29.14	1,661
Chr 8	0.19	0.11	0.15	0.27	470	37.29	2,389
Chr 9	0.20	0.13	0.16	0.19	402	28.83	2,146
Chr 10	0.22	0.14	0.17	0.18	238	31.59	1,718
Chr 11	0.20	0.12	0.14	0.14	138	22.35	1,094
Average	0.20	0.12	0.15	0.22	231	33.03	1,997

Table S4. Analysis of molecular variance (AMOVA) for the extraction of SNP variation between and within groups (groups) based on 280 *Coffea canephora* genotypes genotyped with 2.542 polymorphic SNPs.

Category	**Source of variation**	**d.f.**	**Sum of squares**	**Mean square deviations**	**Percentage of variation**	**P-value**
Groups based on STRUCTURE at K = 2	Between groups	2	13712.97	6856.4855	31.88	0.001
	Between genotypes Within groups	277	68717.15	248.0763	-0.20	0.493
	Within genotypes	280	69861.49	249.5053	68.32	0.001
	Total	559	152291.60	272.4358	100.00	
Groups based on cluster analysis	Between groups	5	17138.91	3427.7828	17.76	0.001
	Between genotypes Within groups	274	65291.20	238.2891	-1.89	0.829
	Within genotypes	280	69861.49	249.5053	84.13	0.001
	Total	559	152291.60	272.4358	100.00	

*degrees of freedom

Table S5.1 Pairwise F_{ST} for different hierarchical levels based on 280 *Coffea canephora* accessions genotyped with 2542 SNPs based on six Clusters generated in the dendrogram by the Ward.D2 method.

	Cluster 4	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 6**	**Cluster 5**
Cluster 4	0					
Cluster 1	0.5636057	0				
Cluster 2	**0.5970950**	0.15788759	0			
Cluster 3	**0.5936376**	0.02142376	0.1563861	0		
Cluster 6	0.3362023	0.26468237	0.3902023	0.3028636	0	
Cluster 5	0.5145880	0.41850564	**0.5957493**	0.4604507	0.3498602	0

Table S5.2 Pairwise F_{ST} for different hierarchical levels based on 280 *Coffea canephora* accessions genotyped with 2542 SNPs based on Structure.

	Cluster 2	**Cluster 1**	**Cluster 3**
Cluster 2 (red pool)	0		
Cluster 1 (hybrids)	0.3901153	0	
Cluster 3 (green pool)	**0.5948044**	0.1019733	0

Table S5.3 Pairwise F_{ST} for different hierarchical levels based on 23 *Coffea canephora* accessions genotyped with 2722 SNPs based on three Clusters generated in the dendrogram by the Ward.D2 method.

	Cluster 1	**Cluster 2**	**Cluster 3**
Cluster 1	0		
Cluster 2	0.3490969	0	
Cluster 3	**0.5596507**	0.3559234	0

Table S5.4 Pairwise F_{ST} for different hierarchical levels based on 257 *Coffea canephora* accessions genotyped with 1295 SNPs based on 14 Clusters generated in the dendrogram by the Ward.D2 method.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10	Cluster 11	Cluster 12	Cluster 13	Cluster 14
Cluster 1	0													
Cluster 2	0.007766725	0												
Cluster 3	0.185740294	0.20009244	0											
Cluster 4	0.027687179	0.03887753	0.2137103	0										
Cluster 5	0.010851012	0.02098482	0.1568006	0.03236783	0									
Cluster 6	0.044384306	0.05071526	0.3166646	0.07312560	0.05552992	0								
Cluster 7	0.079609248	0.08078361	0.3325213	0.13079877	0.10451477	0.12569367	0							
Cluster 8	0.073950782	0.07871339	0.3056131	0.10807126	0.08357412	0.13221727	0.1783450	0						
Cluster 9	0.076361762	0.08001336	0.2498082	0.08506713	0.07359072	0.14135429	0.1833105	0.1582378	0					
Cluster 10	0.171693985	0.17484498	0.4808530	0.21234669	0.20122607	0.07148355	0.2458757	0.2723704	0.2876711	0				
Cluster 11	0.182800595	0.18792818	**0.5431535**	0.23559531	0.20480692	0.28142998	0.3194716	0.2685894	0.2736389	0.4598779	0			
Cluster 12	0.179528806	0.19502870	**0.5015144**	0.19636165	0.19502414	0.28196537	0.3159182	0.2887492	0.2712967	0.4443775	**0.5695486**	0		
Cluster 13	0.030775851	0.03084302	0.2742851	0.07411112	0.04712191	0.07849660	0.1066844	0.1015471	0.1234881	0.2082453	0.2290171	0.2550524	0	
Cluster 14	0.206214227	0.21519542	**0.5091226**	0.25469492	0.22976619	0.33148330	0.3602968	0.3189202	0.3099294	0.4837620	**0.5074829**	**0.5667479**	0.2826391	0

Table S6. Genetic distance was calculated in order to identify contrasting parents to propose a new collection with divergent genotypes

Genotype X	Genotype Y	Greater genetic distance
G93	168.6	0.375759
168.6	G93	0.375759
171.9	G93	0.372521
G92	G93	0.371065
172.5	G93	0.366913
G55	G93	0.366401
172.8	G93	0.36351
126	G93	0.362977
Jequitibá P1	G93	0.362857
G80	168.6	0.361678
162	G93	0.361513
171.10	G93	0.361473
6	G93	0.361385
172.3	G93	0.361233
186.3	G93	0.361014
162.10	G93	0.36065
166.2.1	G93	0.360415
186.5	G93	0.3603
G69	G93	0.360233
G19	G93	0.360215
G64	G93	0.360198
G59	168.6	0.360162
169.4	G93	0.35973
Robustão P4	G93	0.359699
136.5	G93	0.359359
G127	G93	0.358998
G109	G93	0.357998
164.4	168.6	0.357979
95	G93	0.357892
166	G93	0.357571
G16	G93	0.357484
186.2	G93	0.357466
145	G93	0.357373
32.10	G93	0.357331
189.4	G93	0.357201
192.3	G93	0.356961
192	G93	0.356914
136.27	G93	0.356828
G143	168.6	0.356813
G11	G93	0.356718
181.3	G93	0.356474
176.5	G93	0.356303
G32	G93	0.356257
114.1	G93	0.356016
144	G93	0.355894
164.7	G93	0.355442
179.2.1	G93	0.355405
G115	G93	0.355352
173.8	G93	0.355096
Robustão P2	G93	0.354903
64	G93	0.354771
G70	G93	0.354589
G60	G93	0.354573
G12	G93	0.354054
169.5	G93	0.354049

G52	G93	0.353811
136.19	G93	0.353715
111	G93	0.353478
G79	G93	0.353351
179.4	G93	0.353318
Robustão P5	G93	0.353202
182.9	G93	0.352995
G130	G93	0.352843
170.6	G93	0.352607
170	G93	0.352551
G20	G93	0.352524
G140	G93	0.35229
G63	G93	0.351825
G77	G93	0.351796
151.9	G93	0.351669
189.2	G93	0.35164
171	G93	0.351597
149	G93	0.351565
G91	G93	0.351556
164.8	G93	0.351376
98	G93	0.351352
G103	G93	0.351078
161.7	G93	0.351
G117	G93	0.350961
G42	G93	0.350767
134	G93	0.350726
165.4	G93	0.350688
G97	G93	0.350617
G59.1	168.6	0.350534
31.4	G93	0.350528
62	G93	0.350514
G120.1.1	G93	0.35044
151.8	G93	0.350369
168.3	G93	0.350356
G44	G93	0.35034
179.2	G93	0.350307
131.3	G93	0.350251
G144	G93	0.350162
166.5	G93	0.350151
60	G93	0.350023
G27	G93	0.349968
162.6	G93	0.349898
G98	G93	0.349722
39.2	G93	0.349438
G14.1	G93	0.349104
G120.1	G93	0.349086
168.9	G93	0.348919
162.2	G93	0.348863
G1	168.6	0.348844
G104	G93	0.348566
146	G93	0.348527
32.4.1	G93	0.348478
281.2	G93	0.348409
G50	G93	0.348402
G67	G93	0.34825
162.7	G93	0.348139
G148	G93	0.348133
161.2	G93	0.347824
G146	G93	0.347507
100	G93	0.347464

192.2	G93	0.347377
176.6	G93	0.347142
290.4	168.6	0.346836
G74	G93	0.346824
95.1	G93	0.346774
G68	G93	0.346766
172.6	G93	0.346724
281.8	G93	0.346723
Vitória P3	G93	0.34671
G141	G93	0.34663
G156	G93	0.346482
40	G93	0.346258
Verdim	G93	0.346257
136.34	G93	0.346249
G111	G93	0.346168
149.1	168.6	0.346163
114	G93	0.34614
136.26	G93	0.346094
166.2	G93	0.346075
187	G93	0.346002
G8	G93	0.345966
75	G93	0.34575
G131	G93	0.345665
Vitória P2	G93	0.34559
183.4	G93	0.345589
186.41	G93	0.345524
181.4	G93	0.345522
Diamante P4	G93	0.345439
168.10	G93	0.345351
203.2	G93	0.345321
162.10.1	G93	0.345243
144.2	G93	0.345132
172.4	G93	0.345062
G41	G93	0.345005
G9	G93	0.344968
173.6	G93	0.344924
G25	G93	0.344904
192.24	G93	0.344903
G120	G93	0.344768
182.3	G93	0.344704
169	G93	0.344702
G110	G93	0.344643
91	G93	0.344549
173.3	G93	0.344505
G135	G93	0.344405
189	G93	0.344349
G73	G93	0.34428
161.9	G93	0.34422
G75	G93	0.344217
G72	G93	0.343906
53	G93	0.343662
136.32	G93	0.343629
P2	168.6	0.343429
A1	168.6	0.343158
G147	G93	0.343052
G57	G93	0.342889
G7	G93	0.342698
G15	G93	0.342694
G127.1	G93	0.342351
G137	G93	0.342331

179.5	G93	0.342226
58	G93	0.342226
112	G93	0.342209
32.8	G93	0.342034
31.2	G93	0.34203
Vitória P5	G93	0.341998
300	G93	0.341973
61	G93	0.341964
171.2	G93	0.341758
168.5	G93	0.341691
192.4	G93	0.34148
103	G93	0.341329
G26	G93	0.341165
179	G93	0.341077
Jequitibá P4	G93	0.341038
G155	G93	0.340995
G149	G93	0.340926
G43	G93	0.340872
G96	G93	0.340781
G151	G93	0.340746
97	G93	0.340564
86	G93	0.340461
G23	G93	0.340457
Centenário P4	G93	0.340383
186.6	G93	0.340325
G138	G93	0.340325
136.18.1	G93	0.340275
31	G93	0.340233
Centenário P3	G93	0.339808
G128	G93	0.339618
Marilândia P7	G93	0.33958
G92.1	G93	0.339534
59	G93	0.339471
112.1	G93	0.339287
165.3	168.6	0.33926
G51	G93	0.339182
G129	G93	0.33911
G61	G93	0.33898
BRS	G93	0.338937
G95	G93	0.338877
21	G93	0.338752
95.1.1	G93	0.33871
G136	G93	0.338344
46	G93	0.338226
G17	G93	0.338146
G154	G93	0.338049
Robustão P1	G93	0.337984
G78	G93	0.337867
32.4	G93	0.337668
G14	G93	0.337435
182.8	G93	0.337219
182.10	G93	0.337168
113	G93	0.336803
125	G93	0.336748
G145	G93	0.33649
G153	G93	0.336265
G56	G93	0.336188
182.5	G93	0.336012
G38	G93	0.335941
G102	G93	0.335919

179.3	G93	0.335903
RO	G93	0.335713
Marilândia P11	G93	0.335705
164.10	G93	0.335241
G46	G93	0.3351
G118	G93	0.334742
136.3	G93	0.334229
G10	G93	0.333482
201.2	G93	0.333237
136.18	G93	0.332934
132.2	G93	0.332933
G2	G93	0.332755
Centenário P1	G93	0.332663
G34	G93	0.332322
Jequitibá P2	G93	0.332056
29.2	G93	0.332029
34.2	G93	0.331811
54	G93	0.331659
Centenário P7	G93	0.331537
83	G93	0.330914
180.7	G93	0.33067
201.6	G93	0.330479
Diamante P8	G93	0.33015
28.2	G93	0.329673
G132	G93	0.329436
60.1	G93	0.329107
171.6	G93	0.328962
Robustão P3	G93	0.328435
24	G93	0.327394
G3	G93	0.326913
G133	G93	0.325133
136.33	G93	0.324501
Jequitibá P6	G93	0.324178
281.6	G93	0.323222
G62	G93	0.323222
181.4.1	G93	0.323212
85	G93	0.322113
G37	G93	0.32121
Vitória P4	G93	0.320143
22	G93	0.307425
G99	G93	0.286445
169.10	G93	0.280762
Jequitibá P3	G93	0.27796
Centenário P2	G16	0.27514
Vitória P1	G93	0.257687

Table S7.1. Distribution of 23 genotypes of *C. canephora* germplasm in groups obtained by the Ward.D2 method. Observed and expected average heterozygosity (HO and HE) and fixation index (F) among genotypes belonging to different hierarchical levels (groups generated in the dendrogram).

Cluster	No. of genotypes	Genotypes	HO	HE	F
1	12	**A1**, G1, G93, G80, 165.3, G59, 290.4, G59.1, **P2**, 164.4, 149.1 and G143	0.34	0.20	-0.70
2	5	G99, 169.10, **Centenário P2**, **Vitória P1** and **Jequitibá P3**	0.27	0.28	0.03
3	6	91, 168.5, G25, G8, G16 and 187	0.21	0.12	-0.75

Table S7.2. Distribution of 257 genotypes of *C. canephora* germplasm in groups obtained by the Ward.D2 method. Observed and expected average heterozygosity (HO and HE) and fixation index (F) among genotypes belonging to different hierarchical levels (groups generated in the dendrogram).

Cluster	No. of genotypes	Genotypes	HO	HE	F
1	47	22, 172.8, G79, G92, 161.2, 281.6, 162.2, 145, G12, G63, **Verdim**, G51, 131.3, G7, 281.8, 132.2, G68, G118, G43, 75, G146, 136.3, 126, 173.3, G42, G115, 151.9, 149, G38, 186.41, G130, 136.26, 40, **Centenário P1**, G131, **Robustão P3**, G140, G149, 281.2, **Centenário P4**, 32.4.1, **Marilândia P11**, 181.4.1, 46, 114.1, G135 and **Vitória P2**	0.19	0.24	0.21
2	51	G77, 146, G57, G10, G46, 95.1, 300, G17, 180.7, G91, G75, G14, 189.4, 32.8, 164.7, 111, 39.2, G98, 134, G154, G15, 171, G44, G111, 31.4, 192.3, 100, G78, 31, 189.2, G97, G96, 192.24, 144.2, 192.2, 181.3, 125, **Vitória P5**, **Centenário P3**, 113, G145, **Jequitibá P1**, G14.1, 192, 95.1.1, G103, **Marilândia P7**, 103, 166.2.1, G141 and 112.1	0.19	0.23	0.17
3	9	G151, 192.4, G147, G137, G136, G92.1, G128, 182.10 and **Robustão P1**	0.19	0.10	-0.78
4	28	179.3, 186.2, 168.10, 144, 186.3, 183.4, 64, 201.6, 151.8, 53, 59, 179.2, 54, G11, 6, 28.2, 58, 61, 86, G61, Vitória P3, G117, **Robustão P4**, 85, **Robustão P5**, 31.2, 34.2 and **Jequitibá P2**	0.19	0.21	0.11
5	31	166.5, G27, G62, G26, G67, 62, 24, 60, 161.7, G95, 136.18, G2, 164.10, G9, **RO**, G3, G20, 169, 179, **Jequitibá P6**, 32.10, 60.1, G148, Vitória P4, 21, G129, 136.18.1, **Diamante P8** and **Centenário P7**	0.19	0.22	0.14
6	7	G19, 176.6, G34, 136.32, 83, G120 and G153	0.20	0.20	0.02
7	12	G69, 171.10, 176.5, 170, G60, 179.4, 171.9, 98, 179.5, G70, G50 and 179.2.1	0.18	0.19	0.04

8	11	166, G23, 168.9, 168.3, 172.6, 172.5, 161.9, 169.4, 168.6, G64 and 170.6	0.19	0.20	0.04
9	26	136.33, 136.34, 189, 136.5, 32.4, 171.2, G41, G32, 162.10, G109, G74, G110, G72, G102, 162.6, 162.7, G55, 164.8, 203.2, 182.9, **Robustão P2**, 186.6, 182.3, 29.2, 182.5 and 182.8	0.20	0.20	0.02
10	7	G127, 173.8, G52, 186.5, 136.27, G120.1 and G120.1.1	0.15	0.14	-0.06
11	4	171.6, G37, G133 and G132	0.20	0.11	-0.89
12	4	201.2, G56, G127.1 and 112	0.19	0.10	-0.91
13	14	166.2, 95, 181.4, G104, 97, **BRS**, 165.4, 162, 172.3, 114, 136.19, 162.10.1, G144 and 172.4	0.19	0.21	0.12
14	6	173.6, G156, G73, G155, G138 and **Jequitibá P4**	0.21	0.11	-0.89

Table S8. P-values (p value < 0.059) for seven traits that showed SNPs with significant association for 251 *Coffea canephora* genotypes for "chromosome 0" (*scaffolds*)

Characteristic	Allele ID	p-value
Height	**100076552.7.G.A**	0.0348
Height	100086221.57.G.A	0.0586
Diameter	**100076012.33.C.T**	0.0538
Diameter	**100076552.7.G.A**	0.0541
Diameter	**100084770.47.G.A**	0.046
Diameter	100087976.25.G.C	0.0296
Diameter	100067022.15.G.A	0.0251
Diameter	100090515.18.T.G	0.0215
Diameter	100088948.5.T.C	0.0449
Diameter	100090515.6.T.C	0.0377
Diameter	100036156.28.G.T	0.0001
Diameter	100091848.44.C.A	0.052
Diameter	100039587.29.C.G	0.0482
Diameter	100091505.58.G.A	0.0207
Diameter	100087714.30.T.G	0.0077
Diameter	100084443.8.A.T	0.0011
Average diameter growth rate	**100076012.33.C.T**	0.0522
Average diameter growth rate	**100084770.47.G.A**	0.0163
Average diameter growth rate	100089844.20.C.A	0.0177
Average diameter growth rate	100090449.23.T.C	0.0245
Average diameter growth rate	100066922.26.G.T	0.0066
Average diameter growth rate	100068294.29.G.A	0.0038
Average diameter growth rate	100065316.18.T.A	0.0573
Average diameter growth rate	100068706.6.G.A	0.0029
Average diameter growth rate	100065136.14.G.A	0.0092
Average diameter growth rate	100065956.19.A.C	0.0425
Average diameter growth rate	100074498.63.T.C	0.0382
Average height growth rate	100036170.5.A.G	0.0516
Average height growth rate	100087322.28.G.A	0.0209
Average height growth rate	100083446.13.T.C	0.0516
Average height growth rate	100064056.24.C.T	0.0256
Average height growth rate	100088839.5.C.T	0.0541
Average height growth rate	100036546.20.G.C	0.0335
Average height growth rate	100087432.9.A.G	0.0068
Average height growth rate	100080126.43.A.G	0.0024
Average height growth rate	100038714.31.T.G	0.0261
Average height growth rate	100063716.48.C.A	0.0004
Average height growth rate	100085234.17.A.G	0.0002
Leaf rust	100088143.16.G.A	0.0382
Leaf rust	100037030.38.A.G	0.0375
Leaf rust	100091851.23.C.G	0.0514
Leaf rust	100064790.66.G.T	0.0011
Leaf rust	100081314.12.G.A	0.0132
Leaf rust	100076012.28.T.G	0.0301
Leaf rust	100085127.48.C.G	0.0376
Leaf miner	100064333.58.A.G	0.0518
Leaf miner	100089967.5.T.A	0.0497
Leaf miner	100063802.10.A.T	0.0582
Leaf miner	100083240.32.G.C	0.0532
Leaf miner	100091178.11.A.T	0.0204
Leaf miner	100063835.51.A.T	0.0467

Mealybug	100067274.5.C.T	0.0184
Mealybug	100080157.63.G.T	0.0167
Mealybug	100090110.40.G.A	0.0002
Mealybug	100090110.26.G.A	0.0064
Mealybug	100065046.24.T.G	0.0204
Mealybug	100064996.16.C.A	0.0078