Received: 31 August 2022 | Accepted: 12 February 2023

DOI: 10.1002/tpg2.20321

The Plant Genome

SPECIAL SECTION: GENOMICS OF ABIOTIC STRESS TOLERANCE AND CROP RESILIENCE TO CLIMATE CHANGE

Genomic-assisted breeding for climate-smart coffee

Maria Amélia G. Ferrão[1,2] | **Aymbire F. A. da Fonseca**[1,2] | **Paulo S. Volpi**[1] | **Lucimara C. de Souza**[1] | **Marcone Comério**[1] | **Abraão C. Verdin Filho**[1] | **Elaine M. Riva-Souza**[1] | **Patricio R. Munoz**[3] | **Romário G. Ferrão**[1,4] | **Luís Felipe V. Ferrão**[3]

[1]Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural—Incaper, ES, Brazil

[2]Empresa Brasileira de Pesquisa Agropecuária—Embrapa Café, Brasília, Brazil

[3]Blueberry Breeding and Genomics Lab, Horticultural Sciences Department, University of Florida, Gainesville, FL, USA

[4]Multivix Group, ES, Brazil

Correspondence
Felipe V. Ferrão, Blueberry Breeding and Genomics Lab, Horticultural Sciences Department, University of Florida, Gainesville, FL 32611, USA.
Email: lferrao@ufl.edu

Funding information
Fapes (Espírito Santo Research Foundation, Brazil), CNPq (National Council for Scientific and Technological Development, Brazil), and CAPES (Brazil)

Assigned to Associate Editor Diego Jarquin.

Abstract

Coffee is a universal beverage that drives a multi-industry market on a global basis. Today, the sustainability of coffee production is threatened by accelerated climate changes. In this work, we propose the implementation of genomic-assisted breeding for climate-smart coffee in *Coffea canephora*. This species is adapted to higher temperatures and is more resilient to biotic and abiotic stresses. After evaluating two populations, over multiple harvests, and under severe drought weather condition, we dissected the genetic architecture of yield, disease resistance, and quality-related traits. By integrating genome-wide association studies and diallel analyses, our contribution is four-fold: (i) we identified a set of molecular markers with major effects associated with disease resistance and post-harvest traits, while yield and plant architecture presented a polygenic background; (ii) we demonstrated the relevance of nonadditive gene actions and projected hybrid vigor when genotypes from different

Abbreviations: Aroma liking, AROMA; Average yield, YIELD; Baker' s ratio, PR; Bayesian Sparce Linear Mixed Model, BSLMM; Bean size, GSIZ; Cercospora leaf spot, CERC; Coffee leaf miner, LMINER; Coffee leaf rust, RUST; Equilibrium, EQUIL; Flavor liking, FLAVOR; Flower time, FL; general combining ability, GCA; General inclination, GL; General scale, GSCE; Genome-wide association studies, GWAS; Genomic selection, GSCE; Leaves blight, LBLIGHT; likelihood ratio test, LRT; Linkage disequilibrium, LD; marker assisted selection, MAS; maturation period, UNIF; maturation time, MAT; number of sparse effect loci involved in determining the phenotype, n_gamma; overall liking, OVLIKING; Perception, HEDONIC; Plant Architecture, PRT; posterior inclusion probability, PIP; Principal Component Analysis, PCA; proportion of genetic variance explained by genetic variants with major effect, rho; proportion of phenotypic variance explained by the sparse effects and random effects, PVE; proportion of PVE explained by the sparse effects only, PGE; Quantitative trait loci, QTL; Retronasal, RETRO; Sieve residual, RES; Sieve size, M15; Sieve size, M13; Sieve size, M10; simple sequence repeat, SSR; Single nucleotide polymorphisms, SNP; Sourness, ACIDITY; specific combining ability, SCA; Sweetness, SWEET; weight (in grams) dry processed and unroasted raw beans, GREEN; weight (in grams) of coffee fruits after used natural dried method (sun-dried beans), CHERRY; Yield 2014–2015, YB1; Yield 2016–2017, YB2; Yield 2018–2019, YB3.

geographically botanical groups are crossed; (iii) we computed medium-to-large heritability values for most of the traits, representing potential for fast genetic progress; and (iv) we provided a first step toward implementing molecular breeding to accelerate improvements in *C. canephora*. Altogether, this work is a blueprint for how quantitative genetics and genomics can assist coffee breeding and support the supply chain in the face of the current global changes.

1 | INTRODUCTION

Coffee is a widely consumed beverage that drives a vibrant industry, which contributes significantly to the economies of several tropical and developing countries. Globally, it is estimated that more than 100 million of growers are benefited from the coffee supply chain and over 2.2 billion cups of coffee are consumed daily (Krishnan et al., 2021). Despite this importance, its sustainability is facing critical challenges, including accelerated climate changes, price volatility, limited access to genetic resources, and the presence of new pests and diseases (Davis et al., 2021; Van der Vossen et al., 2015). Due to these factors, coffee breeding programs have a pivotal role to play by increasing the phenotypic performances and releasing new climate-smart coffee cultivars, that is, a set of plants that combine resilience to biotic and abiotic factors, sustainable mechanization, high-yield, and superior drink quality.

With widespread morphological and physiological variations, coffee domestication has a rich evolutionary history with intense human interventions. The beverage popularly known as coffee is made from roasted and ground beans of two main species: *Coffea arabica*, an autogamous and polyploid species, which mostly contributes to the aroma and sweet flavor; and *Coffea canephora* (also called Robusta coffee), an outcrossing and diploid species that produces a greater yield than Arabica varieties, is more resistant to diseases, and provides approximately double the amount of caffeine (Ferrão et al., 2019b). While Arabica coffee represents ~ 60% of the global production and is considered the main source of drink quality, it has been recently argued that the species does not have the potential to attain the level of climate resiliency required under the existing climate change projections (Davis et al., 2021). For example, it is estimated that *C. arabica* production in Latino America may be reduced in the order of 80% by 2050 (Imbach et al., 2017). At the root of it all is a startling vulnerability: the cultivated Arabica is a delicate crop, quite susceptible to diseases, and with a narrow genetic diversity (Anthony et al., 2002; Cubry et al., 2008; Lashermes et al., 1999; Silvestrini et al., 2007). As an alternative, *C. canephora* is more adapted to higher temperatures and is resilient to biotic and abiotic stresses that raise the species as a potential candidate for more climate-smart cultivars (Ferrão et al., 2019b). However, to make it possible (and sustainable), global efforts have been devoted to improving cup quality, making *C. canephora* more attractive to consumers.

In the recent decades, numerous breeding schemes have been proposed for *C. canephora* species, mostly focused on yield traits. To this end, sexual and asexual breeding strategies are commonly employed in recurrent selection designs, where the best genotypes are either used as parents in future crosses or released as clonal cultivars (Ferrão et al., 2019). To rigorously improve the coffee quality through breeding, understanding the genetic architecture of complex traits is essential. For example, breeders have guided their decisions based on the level of genetic control (i.e., heritability), magnitude of gene action effects, correlations between traits, and dynamics of genotype-by-environment interactions (Mustiga et al., 2018). The use of matting designs provides the means to determine the genetic control of complex traits by estimating combining abilities and the contribution of additive and nonadditive gene actions on the phenotypic variation. While in Arabica coffee, the use of hybrid vigor by exploring the specific combining ability (SCA) effects is a well-documented practice (Cilas et al., 1998; Mohammed, 2011; Walyaro, 1983); combining abilities reported in *C. canephora* remain elusive at this stage, with results restricted to certain genetic backgrounds and being trait-specific (Cilas & Bouharmont, 2005; Cilas et al., 2003; Leroy et al., 1993, 2014).

A more contemporaneous alternative to dissect the genetic architecture of complex traits is using genomic information. Genome-wide association studies (GWAS) provide the means to determine the genetic control of complex traits by measuring a vast number of genetic variants spanning the entire genome and identifying regions affecting the phenotype of interest (Pritchard et al., 2000; Yu et al., 2006). Characteristic features of GWAS include identifying relevant variables that can be used either for marker assisted selection (MAS) or as potential target for gene editing in plant breeding. When compared to traditional QTL analyses, GWAS has the advantage to increase the mapping resolution by using populations with low levels of linkage disequilibrium (LD) and considering a deep history of recombination events. Despite the relevance, genomic association analyses in coffee are still confined to few traits, with most of the investigations performed in small

population sizes, with a restricted genetic diversity and low marker densities.

From a statistical standpoint, most of the GWAS studies reported in the plant literature have been carried out using single-variant association analysis. A more recent line of evidence, however, suggests the use of multilocus (or polygenic) approaches as an alternative to increase power and generate a more robust estimate of genetic variances (Fernando et al., 2017; Zhou et al., 2013). Specifically, polygenic modeling connects trait variation to all molecular markers simultaneously leveraging the ability to estimate effect sizes jointly, by taking LD structure into account (Guan & Stephens, 2011; Nariai et al., 2017). These methods can also parse the relative contributions of genetic variants with measurable versus near-infinitesimal effects to trait genetic variance, that also makes this approach particularly attractive to dissect the genetic architecture of complex trait (Gompert et al., 2019). In this regard, Zhou et al. (2013) proposed the use of a Bayesian sparse linear mixed model (BSLMM), which is a hybrid approach between traditional linear mixed models and Bayesian variable selection regression approaches. Among the key advantages, BSLMM is capable of learning the genetic architecture from the data, yielding good performance across a wide range of scenarios (Bresadola et al., 2019; Comeault et al., 2014; Ferrão et al., 2020; Gompert et al., 2019; Guan & Stephens, 2011; Lloyd-Jones et al., 2017; Nariai et al., 2017).

Aiming to provide new insights into the genetic basis of *Coffea canephora* traits, we integrated diallel and GWAS analyzes with the following main objectives: (i) estimate genetic parameters and decompose the genetic variances into additive and nonadditive genetic effects; (ii) dissect the genetic architecture of important coffee traits; (iii) identify potential genomic regions controlling yield and quality-related traits using a BSLMM, and finally, (iv) propose an alternative for future coffee breeding programs focused on genomic-assisted breeding. Altogether, we argue that improvements focused on *C. canephora* could be a milestone for the coffee industry, with great potential for return on investment because of its genetic diversity, resilience, and productivity. In this study, we present a blueprint on how the use of quantitative genetics and genomics can assist developing climate-smart cultivars.

Core Ideas

- Sustainability of coffee production is threatened by accelerated climate changes.
- *Coffea canephora* is more adapted to higher temperatures and resilient to biotic and abiotic stresses, which make the species a strategic candidate for the development of climate-smart cultivars.
- By integrating genome-wide association analyses (GWAS) and Diallel populations, we described the genetic basis of coffee traits evaluated under drought weather conditions.
- We identified a set of molecular markers with major effects associated with disease resistance and post-harvest traits, while yield and plant architecture presented a polygenic background.
- We emphasize the importance of genomic-assisted breeding by proposing a recurrent selection scheme integrating genomic prediction and marker-assisted selection.

2 | MATERIAL AND METHODS

2.1 | Plant material and experimental design

The populations used in this study were generated as part of the coffee breeding program at the Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural (Incaper), in partnership with Embrapa Café, Brazil. Incaper is one of the main coffee research institutions in Brazil. Since 1985, the institution has an active breeding program of *C. canephora* and has released more than 10 cultivars widely used for Brazilian growers. In this investigation, among thousands of genetic materials maintained in the germplasm collection, we focused on the selection of 10 superior founders (P02, P03, P07, P11, P23, P24, P73, P83, P110, and P149) to compose two complementary breeding populations for genetic analyses. These genotypes were visually selected due their outstanding agronomical traits, including high production, resistance to rust disease (*Hemileia vastatrix*), resilience to drought conditions, and different maturity date (Figure 1a). Recently, the genetic diversity of such materials was described using 18 simple sequence repeat (SSR) molecular markers (L. C. Souza et al., 2021), and the clones were grouped into two main clusters (Figure 1b).

First, a *Diallel population* was established to decompose the observed genetic variance into additive and nonadditive genetic effects. The 10 superior clones were crossed and evaluated in a full diallel fashion design. After a first screening, the best 97 progenies were cloned and installed as a partial diallel, in a randomized block design with three repetitions, and evaluated for six consecutive harvest-production (from 2014 to 2019). In this investigation, the results of the partial diallel were used to estimate combining abilities and genetic parameters.

A second and more diverse breeding population was used for complementary genomic analyses. Referred here as *GWAS population*, this population was designed by selecting 7 of the 10 founders (P02, P03, P07, P11, P24, P73, and P149) to

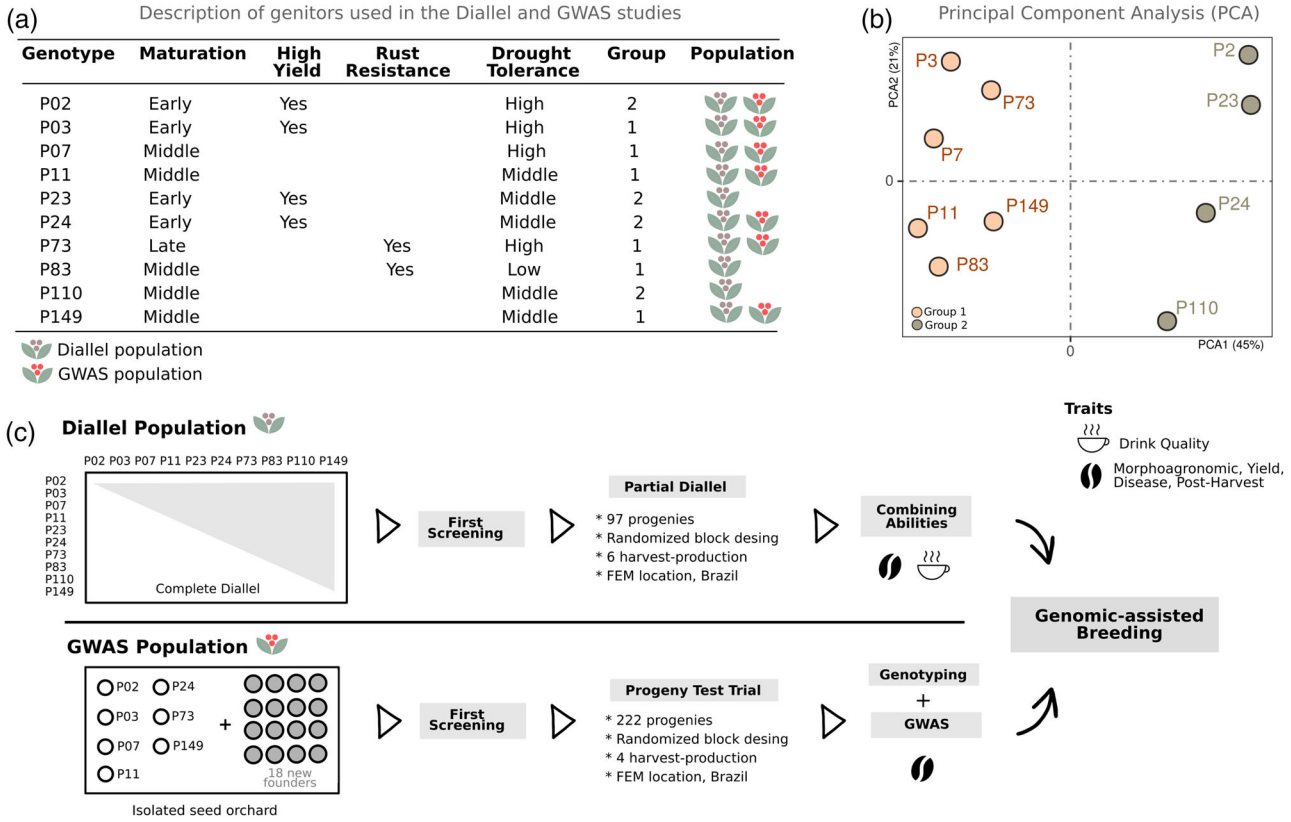

FIGURE 1 Schematic representation of two breeding populations used to dissect the genetic architecture of morphoagronomic, yield, disease resistance, post-harvest, and drink quality traits in *Coffea canephora*. (a) Description of the 10 genitors used as founders; (b) principal component analyses reporting the genetic diversity of the founders computed using 18 simple sequence repeat (SSR) molecular markers population; (c) Diallel population was established from the selection of 10 superior coffee genotypes. After a first screening, the best 97 progenies plus the parental genotypes were established as a partial diallel in a randomized block design and used to estimate combining abilities. The GWAS population is a more diverse breeding population used for genomic analyses. Seven contrasting parents from the Diallel population plus 18 new founders were selected, planted in an isolated field, and allowed to cross pollinate. The 222 best genotypes were cloned and assigned to a randomized complete block design. Please, see the Supporting Information Appendix for additional details on the plant material.

create a base population for recurrent selection. To increase the diversity, we included 18 additional genotypes in an isolated seed orchard and allowed them to cross-pollinate. These 18 new genotypes were visually selected from the germplasm collection, based on their higher production performance and tolerance to biotic and abiotic stresses. After a first screening, 222 progenies were selected, cloned, and assigned to a randomized complete block design with three replications. Progenies were evaluated for four consecutive harvest-production years (2008–2011). This admixture population was genotyped and phenotyped to infer the genomic architecture of complex coffee traits.

2.2 | Traits measurements

In this study, a total of 27 coffee traits encompassing 5 main categories were considered: morphoagronomic, disease resistance, yield, post-harvest, and drink quality. In the morphoagronomic group, we investigated the maturation time (MAT), uniformity of the maturation (UNIF), bean size (GSIZ), plant architecture (PRT), vigor (VIGOR), and the general scale (GSCE)—that is a visual metric assessed by breeders and indicate the overall performance of a genotype. For disease resistance, a total of four traits were visually evaluated, including coffee leaf rust (RUST), coffee leaf mine (LMINER), cercospora leaf spot, also called brown eye spot or berry blotch (CERC), and leave blight (LBLIGHT). All morphoagronomic traits were visually measured by multiple experienced breeders across several years to reduce bias and subjectivity. Further details on each of the traits measured are described in Supporting Information Appendix.

Yield was measured as 60-kg bags per hectare, and the general stability was assessed by dividing the production in biannual periods. The first biennium (Y1) corresponded to the production for 2014 and 2015. Second biennium (Y2), the estimated yield for third and fourth harvest, carried out in 2016 and 2017. Third biennium (Y3) represents the

yield estimated in 2018 and 2019. Finally, an average yield (YIELD) was computed considering all years (from 2014 to 2019).

For post-harvest, a sample of 2 kg of ripened fruits was processed in two stages, herein referred to as CHERRY and GREEN. By CHERRY we mean the weight (in grams) of dried coffee fruits after using a natural method (sun-dried beans). After depulping, by getting rid of the skin and mucilage, we weighted the GREEN beans. For additional post-harvest evaluations, we sampled 300 g of raw beans and classified the bean size and shape using a set of sieves. Namely, we used a round sieve M17 (flat and large beans), round sieve M15 (flat and medium beans), oblong sieve M10 (small "moca" beans), round sieve M13 (medium "moca" beans), and RES (sticks, stones, broken grains, among others residual observed during the post-harvest evaluation).

Finally, sensory traits related to drink quality were evaluated using the hybrids in the *Diallel population*. Samples were subjected to sensory analyses by four Q-Grader judges. The evaluated attributes were aroma liking, flavor liking, retronasal, sourness, sweetness, perception, equilibrium, and overall liking, according to the Coffee Quality Institute (CQI), in collaboration with the Uganda Coffee Development Authority (UCDA) (Uganda Coffee Development Authority, 2012).

Both populations were installed in a representative environment (location) for the Brazilian production: Marilândia Experimental Unit (FEM)—latitude $19^0 24'$ south, longitude $40^0 31'$ west and 70 m altitude. All phenotypes were measured in a system of partial irrigation, or "irrigation under demand." More details about the breeding populations and traits measured are described in the Supporting Information Appendix and Tables S1 and S2.

2.3 | Genotypic data

The *GWAS population* was genotyped using the Genotype-by-Sequencing (GBS) (Elshire et al., 2011). The DNA samples were digested using the ApeKI restriction enzyme, and 96 samples were multiplexed per Illumina flow cell for sequencing. The GBS analysis pipeline was implemented with the TASSEL-GBS software, version 4.3.7 (Glaubitz et al., 2014). Sequenced tags were aligned against the *C. canephora* genome assembly (Denoeud et al., 2014). SNPs were extracted and filtered as follows: (i) triallelic SNPs were removed; (ii) SNPs with minor allele frequency less than 1% were removed; and (iii) SNPs with genotypes that were called in less than 50% of the samples were discarded. After following the quality-control steps, a total of 58,723 SNPs were retained. Further details about the genotypic data are described by Ferrão et al. (2017) and Ferrão et al. (2019a).

2.4 | Combining abilities

For the *Diallel population*, breeding values were predicted using best linear unbiased prediction (BLUP) and restricted maximum likelihood approach (REML) to estimate variance components, as follows: $y_{ijk} = \mu + g_j + g_k + s_{jk} + b_i + e_{ijk}$; where y is the phenotype already pre-corrected for the year effect, μ is the population mean; g_j and g_k are the GCA (general combining ability) effects for the jth and kth parents, respectively; s_{ik} is the SCA (specific combining ability) effect for the cross of the jth and kth parents; b_i is the block effect; and e_{ijk} is the experimental error. GCA and SCA effects were both modeled as random effects, where $g_j \sim N(0, \sigma^2_{g_j})$ $g_k \sim N(0, \sigma^2_{g_k})$ and $s \sim N(0, \sigma^2_s)$, respectively. The choice of treating genotypes as random effects was made due to the highly unbalanced nature of the data. Normality and independence were also assumed for the experimental error, distributed as $e \sim N(0, \sigma^2_e)$. Components of variance were tested using the likelihood ratio test (LRT) and the significance verified by the Chi-square test with 1 degree of freedom.

Broad-sense (H) heritability was calculated at the entry mean level as: $H^2 = (\sigma^2_a + \sigma^2_d)/(\sigma^2_a + \sigma^2_d + \sigma^2_e/r)$, where r is the number of repetitions (blocks); σ^2_a, σ^2_d, and σ^2_e are the additive, dominance, and residual components of variance associated with the GCA, SCA, and experimental errors, respectively. Baker' s ratio (PR) (Baker, 1978) was estimated using the following equation: $PR = 2\sigma^2_a/(2\sigma^2_a + \sigma^2_g)$. PR indicates whether a trait is governed by dominant or additive gene action, where a value below 0.5 indicates that SCA effects were predominant, and the trait would be controlled by nonadditive gene action. All phenotypic analyses were carried out using the ASReml-R software (Butler et al., 2009).

2.5 | Variance components estimated using Bayesian whole-genome regression models

Variance components for additive and nonadditive effects were estimated in the *GWAS population* using a multi-kernel approach. To this end, we considered the following statistical model in a matrix notation: $y = X\beta + W_1 a + W_2 d + W_3 t + W_4 r + e$, where $\mathbf{y}$ is the vector of predicted phenotypes pre-corrected for the blocks and year effects, β is the vector of fixed effects (overall mean), and $\mathbf{e}$ is the vector of the random residual effect as follows: $e \sim N(0, I\sigma^2_e)$, where σ^2_e is the residual variance. The incidences matrix $\mathbf{X}$ and $\mathbf{W}$ relate observations in $\mathbf{y}$ to fixed and random effects, respectively. The random additive effect is defined by $\mathbf{a}$ as follows: $a \sim N(0, G_A\sigma^2_A)$, and was conditioned on the $\mathbf{G_A}$ additive genomic relationship matrix as defined for VanRaden (2008). The random dominance effect is defined by $\mathbf{d}$ as follows:

$d \sim N(0, G_D\sigma^2_D)$, where $\mathbf{G_D}$ is the dominance genomic relationship matrices as defined by Vitezica et al. (2013). Epistatic effect for additive-by-additive and additive-by-dominance are equivalent to the random effects that follow the multivariate Gaussian distributions, whose variance–covariance matrices are proportional to the Hadamard products of the corresponding relationship matrices. Therefore, the random effects for the epistatic effects **t** and **r,** are defined, respectively, as the following: $t \sim N(0, G_{AA}\sigma^2_{AA})$ and $r \sim N(0, G_{AD}\sigma^2_{AD})$. The Hadamard product of genomic relationship matrices has been shown to capture variance from first order epistatic effects by Muñoz et al. (2014). All genomic relationship matrices were computed using the AGHmatrix R-package (Amadeu et al., 2016).

Bayesian regression models were fitted using Markov Chain Monte Carlo (MCMC) implementations in the R package BGLR (Pérez & de los Campos, 2014), using the default hyperparameter and prior settings. For the posterior density, we ran the Markov chain for 100,000 time steps, with a burn-in of 10,000. For estimating genetic variance components, we calculated genotypic value in each MCMC sample after burn-in (Alves et al., 2019; Ishimori et al., 2020; Lehermeier et al., 2017) as: $\hat{g}_m = \sum_{z=1}^{L} m_{iz}\hat{u}_z$, where $\hat{g}_m$is the estimated genetic value of the *i*th individual for the additive, dominant, or epistatic effects, $\hat{u}_z$ is the estimated marker effect for a given genetic parametrization (additive or nonadditive) for the marker z, and m_{iz} is the marker score. The total genetic variance and variance components (σ^2_A, σ^2_D, σ^2_{AA}, σ^2_{AD}) were calculated as the variance of estimated values across all genotypes in each MCMC sample.

2.6 | Polygenic genome-wide association studies (GWAS)

We fit BSLMM as implemented in the GEMMA (*option–bslmm 1*). Unlike traditional GWAS analyses, the polygenic GWAS fits a single model with all genetic variants considered simultaneously. In particular, the phenotypic value is modeled as a function of a polygenic term and a vector of measurable effects, assuming that the SNP effects are sampled from a point-normal distribution. In a matrix notation, the model is: $y = 1_n\,\mu + X\beta + u + \varepsilon$, where **y** is an n-vector of phenotype measured in n individuals; **X** is a n x p design matrix of genotypes measured on the same individuals at p genetic markers, relative to additive and dominance effects; β is the SNP effect sampled from a mixture of two distributions, one that expects many small effects and another that generates few strong effects, as follows: $\beta_i \sim \pi N(\sigma^2_k\tau^{-1}) + (1-\pi)\delta_0$, where σ^2_k controls the expected magnitude of nonzero SNP effects and δ_0 denotes a point mass at zero; u is the polygenic term as previously described; and ϵ is a random independent error term. Additionally, we tested the importance of additive and dominance gene actions. We fit BSLMM for each trait with 2 MCMC chains using the default settings implemented in the GEMMA software. Full details about the model formulation are described by Zhou et al. (2013).

The hierarchical structure of the model provides a means to estimate additional parameters that describe aspects of the genetic architecture of each trait. Assuming sparsity-inducing priors on the regression coefficients, we can estimate the amount of phenotypic variance explained either by loci with detectable effects ("sparse effects") or by the polygenic component ("random effects" estimated from the kinship matrix). This set of parameters includes the proportion of phenotypic variance explained by the sparse effects and random effects (PVE) and the proportion of PVE explained by the sparse effects only (PGE). The hyper-parameters *n_gamma* and *rho* are computed from the data and are estimates of the number of sparse effects loci involved in determining the phenotype and the proportion of genetic variance explained by genetic variants with major effects, respectively. The posterior inclusion probability (PIP) for each SNP is the probability that each SNP has a nonzero effect and therefore should be included in the model. We used a conservative threshold of PIP > 0.4 to identify candidate SNPs associated with phenotypes. This threshold is an order of magnitude higher than the widely used PIP > 0.01 or 0.1 (Chaves et al., 2016; Comeault et al., 2014; Gompert et al., 2013), a fact that reduces the probability of uncovering spurious associations.

Neighboring genetic markers in a genomic region are expected to have some redundancy and therefore to have lower individual PIPs (Bresadola et al., 2019; Fernando et al., 2017). To aggregate information from neighboring molecular markers, we used the SNPRelate R package (Zheng et al., 2012), and an LD-based SNP pruning was carried out to remove SNPs within a genomic window of 1 kb. All single nucleotide polymorphisms that exceeded the threshold (PIP > 0.4) were characterized *in silico* for their genomic position and functional effect. The Coffee Genome Hub database (Dereeper et al., 2015) was used to identify *C. canephora* genes located in the interval of 100 Kbp encompassing significant SNPs, as suggested by Sant'Ana et al. (2018).

We compared the BSLMM results for the additive gene action with traditional GWAS analyses, where SNP-trait association are based on a linear mixed model accounting for population structure (Q) and relative kinship (K) matrices. Correction for multiple testing using a Bonferroni test and a threshold of 0.05 was applied to determine significant associations. We approximated the phenotypic variation explained by each candidate SNP using the coefficient of determination (R2), computed as simple linear regression between the phenotypic value and the SNP marker parametrization. Traditional GWAS analyses were carried out using the GWASpoly

R package (Rosyara et al., 2016) and the additive gene action model.

3 | RESULTS

3.1 | Phenotypic dispersion and combining abilities using diallel analyses

We first accessed the phenotypic dispersion observed in the *Diallel population*. The hybrids showed higher phenotypic performances for most traits, when compared to the parents. Regarding the data dispersion, in general, the empirical distribution was reasonably symmetric for all traits (Figure S1). Among the main phenotypes collected here, yield potential is one of the most important. We observed moderate average values (39 bag/hc), a value that sharply contrasts with the yield potential observed in the same Brazilian area --some high-tech coffee farms have reported,on average, 100 bags/hc. Importantly, our experiments were all carried out without regular irrigation, since our main goal is to mimic drought conditions and leverage the selection of more resilient materials.

The genetic architecture of coffee traits was investigated by estimating broad-sense heritability (Table 1 and Table S3). We observed traits with different genetic bases and heritability values ranging from 0.027 to 0.94. While ripening time (MAT) trait recorded the highest heritability value, leaves blight (LBLIGHT) showed the lowest value. Traits related to yield stability presented moderate-to-high heritability values—a result that was also reported in previous coffee studies (Cilas et al., 2003; Leroy et al., 1997; Montagnon et al., 2003). Post-harvest traits also showed large heritability suggesting fast genetic progress when incorporated in breeding designs. Clear exception was observed for the sensory traits. Flavor is a complex and multifactorial trait, highly influenced by the environment. These aspects not only adversely affect the heritability values, but also delay the genetic progress. The proportion of additive and nonadditive genetic variation was estimated using the Baker's ratio (BR). For most traits, values lower than 0.5 were observed suggesting a larger influence of SCA variance compared to GCA variance. The use of matting designs provides valuable information on combining abilities for selecting superior parents and hybrids (Tables S4 and S5). We identified promising hybrids by predicting the random effects associated with the SCA effects (Figure S2).

The relationship between traits was measured using a principal component analysis (PCA) (Figure 2a) (for more detail, see the Supporting Information Appendix and Figure S3). For yield, the most contrasting value was observed in the biennium 2 (YB2) showing the lowest correlation value with the other traits—an indication of a lack of annual production stability in coffee bean production over different years. Yield and quality traits were positively correlated indicating that both traits can be improved simultaneously. Disease traits also showed positive correlation values, with traits grouped in the second quadrant of the PCA. Post-harvest and morphoagronomic traits resulted in a broader dispersion in the PCA plan. For the post-harvest traits, we observed two clear groups in the quadrant II and IV. While M13 and M15 sieve showed a very distinct pattern and positive correlation with bean size (GSIZ), the beans retained at the M10 sieve seem more correlated to processed coffee bean traits (CHERRY, GREEN, and RES traits). Morphoagronomic traits were arranged across the entire plane. Both maturation (MAT) and flowering time (FL) were positively correlated. The traits that mostly contributed to the phenotypic variation per category, accounting information from the PC1 and PC2, were M13, OVLIKING, GSIZ, YIELD, and LMINER (Figure S4).

To further evaluate the opportunity to explore crosses using more divergent materials, we estimated the phenotypic performance of the hybrids as a function of the genetic dissimilarity between the parental genotypes (Figure 2b and Figure S5). Our fundamental hypothesis is that when contrasting parental genotypes are crossed, it could lead to heterosis and, therefore, increase the phenotypic performance of the siblings. For most of the traits, we observed a poor phenotypic performance for hybrids originating from parents with low genetic dissimilarity between them. When paired with the combing abilities previously reported, we observed that traits with an important nonadditive component (e.g., yield stability and quality traits) showed a positive trend and better phenotypic performance when genetically distant parents were crossed.

3.2 | Variance components estimated via whole-genome prediction

After speculating about the genetic architecture using the *Diallel population*, we dissect the genetic basis using a more diverse population and molecular information. When results from combining abilities and multi-kernel results are contrasted in both populations, we observed the importance of nonadditive for most of the traits (Figure 3a). Considering yield stability, for example, variance components estimated for the nonadditive effect contributed significantly to the observed phenotypic variation. Interestingly, the variance components associated with additive variation had a reasonable contribution for post-harvest traits, for RES and M15 traits. In terms of magnitude, disease-related traits showed contrasting values over both populations, with larger residuals estimated in the *Diallel population* when compared to the *GWAS population*. A further assessment was performed in the *GWAS population*, where we used molecular information to explore the importance of epistatic effects (Figure 3b). For traits like PRT and YIELD, the dominance effects showed a

TABLE 1 Descriptive analyses of coffee traits evaluated in the Diallel population for five classes of coffee traits: morphoagronomic (MA); disease resistance (disease), yield stability (yield), post-harvest (PH), and drink quality (Quality).

Trait	Abr.	Direction	Class	Unit	Avg	SD	h2	Baker
Flower time	FL	–	MA	Days	276.888	10.789	0.896	0.369
Maturation	MAT	–	MA	1–7	3.138	0.679	0.947	0.643
Uniformity	UNIF	Low	MA	1–3	1.503	0.243	0.579	0.057
Bean Size	GSIZ	High	MA	1–7	4.281	0.649	0.813	0.536
Plant architecture	PRT	Low	MA	1–3	2.138	0.349	0.785	0.385
Vigor	VIGOR	High	MA	1–9	6.839	0.551	0.792	0.304
General scale	GSCE	High	MA	1–9	6.438	0.471	0.657	0.154
Architecture	GL	Low	MA	1–3	1.541	0.316	0.845	0.346
Rust	RUST	Low	Disease	1–9	2.386	0.516	0.778	0.243
Cercospora	CERC	Low	Disease	1–9	2.367	0.328	0.610	0.107
Leaves blight	LBLIGHT	Low	Disease	1–9	1.224	0.342	0.027	0.012
Leaf mine	LMINER	Low	Disease	1–9	2.972	0.396	0.517	0.178
2014–2015	YB1	High	Yield	bag/hc	32.978	14.565	0.840	0.524
2016–2017	YB2	High	Yield	bag/hc	50.358	18.669	0.694	0.170
2018–2019	YB3	High	Yield	bag/hc	34.840	15.295	0.528	0.256
Average yield	YIELD	High	Yield	bag/hc	39.165	10.862	0.829	0.488
Green beans	GREEN	Low	PH	ratio*	1.915	0.137	0.796	0.185
Dried coffee	CHERRY	Low	PH	ratio*	4.325	0.459	0.759	0.257
Sieve size	M15	Low	PH	%	24.393	14.440	0.955	0.873
Sieve size	M13	High	PH	%	34.930	13.597	0.933	0.819
Sieve size	M10	Low	PH	%	24.183	7.271	0.880	0.467
Sieve residual	RES	Low	PH	%	40.887	14.381	0.907	0.737
Aroma liking	AROMA	High	Quality	1–10	7.285	0.338	0.254	0.023
Flavor liking	FLAVOR	High	Quality	1–10	7.478	0.399	0.370	0.002
Retronasal	RETRO	High	Quality	1–10	7.024	0.303	0.133	0.008
Sourness	ACIDITY	High	Quality	1–10	7.170	0.325	0.263	0.025
Sweetness	SWEET	High	Quality	1–10	7.407	0.403	0.366	0.039
Perception	HEDONIC	High	Quality	1–10	7.381	0.392	0.323	0.403
Equilibrium	EQUIL	High	Quality	1–10	7.259	0.325	0.317	0.266
Overall liking	OVLIKING	High	Quality	1–100	78.284	2.587	0.240	0.413

Abbreviations: Avg, average; SD, standard deviation; h2, broad sense heritability; Baker, Baker's Ratio.

*ratio: ratio between harvested ripened fruits and dried (CHERRY) and processed beans (GREEN). A value of 2 for CHERRY, for example, means a ratio of 2:1, where 2 kg of ripened fruits resulted in 1 kg of dried coffee (CHERRY bean). Analogously, a value of 4 for GREEN means that 4 kg of ripened fruits resulted in 1 kg of raw coffee.

considerably larger importance than epistatic effects. For others, inter and intra-locus variation showed similar projections and less importance on explaining the phenotypic variation, when compared to the additivity.

3.3 | Genetic architecture accessed using polygenic GWAS analysis

Genome-wide association analyses were performed using a BSLMM. A total of 45,989 SNPs distributed across 11 *C. canephora* chromosomes were tested for association using additive and dominance models. Molecular information explained a considerable proportion of the phenotypic variation using both parametrizations (Figure 4a). Point estimates of PVE indicated values ranging from 0.43 for the M10 trait (with a probability interval of 0.16–0.73) to 0.97 for the MAT trait (with a probability interval of 0.90–0.99), both using the additive model. The BSLMM formulation makes it possible to estimate additional parameters that describe aspects related to the trait's genetic architecture. The PGE, for example, is interpreted as the proportion of the PVE due to SNPs with measurable effects and it is an approximation about the importance of markers with major effects on

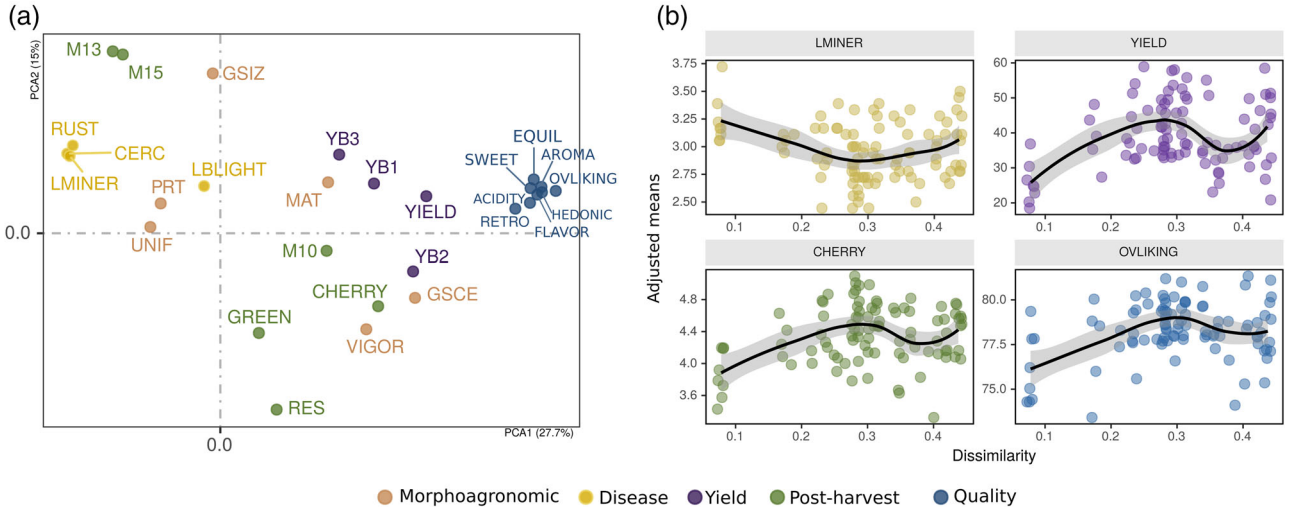

FIGURE 2 (a) Principal component analysis (PCA) displaying the importance of each coffee trait on the phenotypic variation, using the Diallel population. (b) Linear relationship between the phenotypic performance (adjusted means) of the hybrids and their genetic dissimilarity computed using single sequence repeat (SSR) markers for four traits evaluated in the Diallel population.

the expression of the respective phenotypic trait. The MAT trait presented the largest PGE value with a point estimate of 0.88. GREEN and CHERRY, both traits associated with post-harvest, also showed large PGE values and theoretically with simpler genetic architecture. For most of the traits, large PGE were reported for the dominance parametrization.

We also assessed two additional parameters using the BSLMM model. The hyper-parameter *gamma* denotes the posterior samples of number of variants with major effect, while the hyper-parameter *rho* approximates the proportion of genetic variance explained by genetic variants with major effect. Thus, *rho* mass near to zero indicates a highly polygenic genetic basis, while mass values near to 1 means few major effect loci controlling the genetic architecture of the trait. Together, both can be used to shed additional light on the genetic architecture of coffee traits. For most coffee traits, we observed high point estimate values for *gamma* (> 10) indicating a polygenic background (Figure 4b). Accordingly, low *rho* values were also estimated (Figure 4c). Evidence of oligogenic traits were observed for PRT, LBLIGHT, GREEN, and CHERRY traits. Interestingly, results from LBLIGHT disease are largely contrasting when diallel and GWAS results are compared, which might be explained for a larger disease infection in the *GWAS population*.

3.4 | Genotype–phenotype associations and candidate genes associated with yield and quality-related traits

Traits with oligogenic nature are more appropriate to be assayed in MAS designs. To estimate the strength of the association between genotypic variation at individual SNP and phenotypic variation, we inferred the PIP that each genetic variant should be included in the model (see more details in the Supporting Information Appendix S2). Consistent with the genetic projection previously reported, we could map SNPs with large effects (PIP > 0.8) for PRT, LBLIGHT, GREEN, and CHERRY traits using additive and dominance gene parametrizations (Figure 5). In a lower scale (PIP > 0.4), we identified additional associations related to bean production (YIELD) and post-harvest (RES) traits. Remarkably, for GREEN trait, a single maker in chromosome 1 explained by itself more than 30% of the phenotypic variance, while in chromosome 9, a single marker explained more than 25% of the phenotypic variance for LBLIGHT (Table S6). In chromosome 1, we co-localized a marker associated with a QTL explaining more than 10% of the phenotypic variance for CHERRY and GREEN trait. In chromosome 6, using a dominance parametrization, we noticed a single marker explaining 18% of the PRT variation. In lower magnitude, for YIELD, a SNP in chromosome 5 explained 15% of the phenotypic variance. Moreover, as a first assessment, we identified multiple candidate genes flanking SNPs significantly associated with almost all traits based on the annotation of the *C. canephora* genome.

4 | DISCUSSION

4.1 | *Coffea canephora* as an alternative for climate-resilient coffee

Coffee is a popular brewed beverage resulting from a complex and fascinating value chain, that from the seed to the cup needs to be cultivated, harvested, processed, roasted, and brewed to be consumed as our daily coffee. On the basis of this chain, we have plant breeding and its fundamental role on

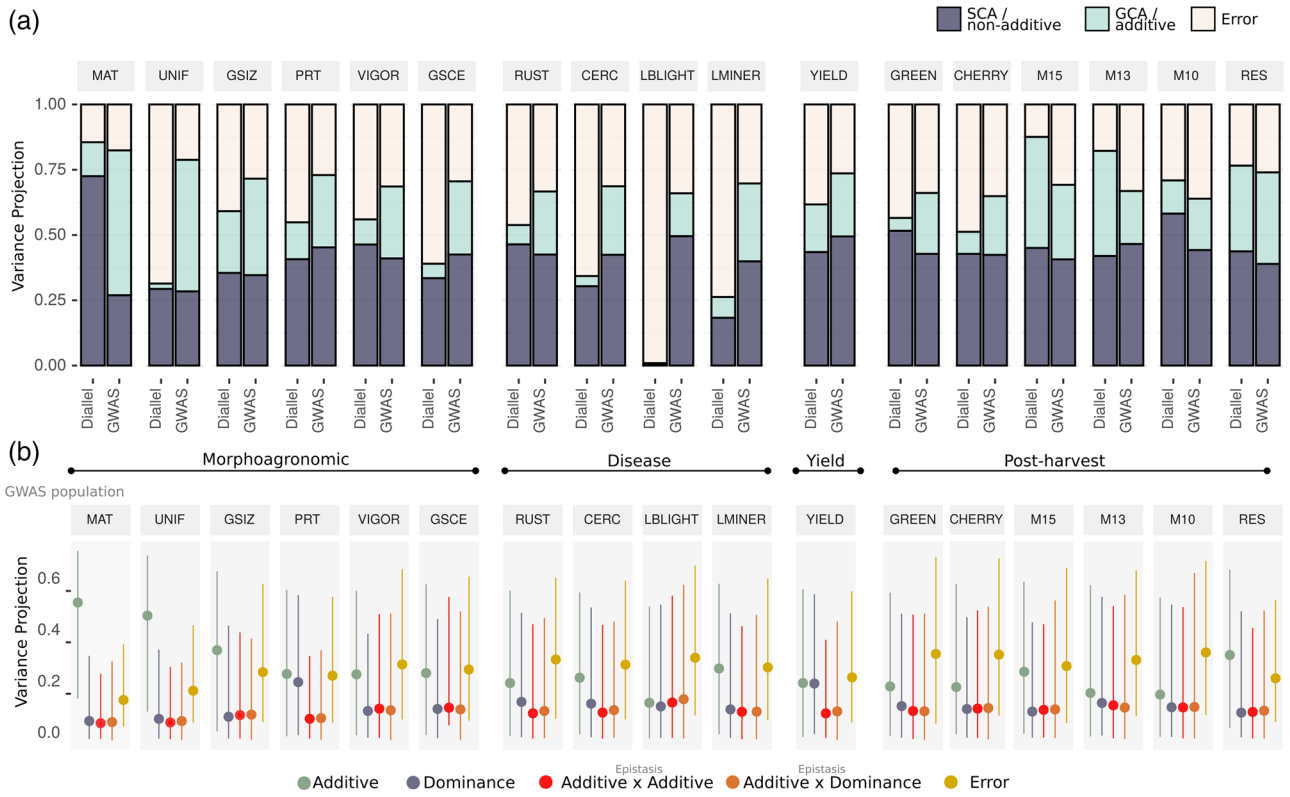

FIGURE 3 (a) Genetic variance projections computed using phenotypic and genomic data using the Diallel and the genome-wide association studies (GWAS) populations, respectively. (b) Variance projections characterizing the genetic architecture for each trait, including additive, dominance, epistasis, and residual variances. Credibility intervals of effect classes for each trait was presented and computed using Bayesian multi-kernel regression models.

releasing new cultivars. A current challenge faced by the coffee community (and breeders) is the projected climate changes and the necessity for developing more climate-resilient cultivars. In a scenario where Arabica production dictates the global market, it has been argued that such a species is not particularly resilient to the projected climate changes. An alarming projection indicates a reduction greater than 50% in coffee production in future years, with losses affecting the livelihoods of 100 million people working in the coffee chain (Davis et al., 2021; Imbach et al., 2017; Ovalle-Rivera et al., 2015). Our fundamental hypothesis is that the climatic unsuitability of coffee farms may be mitigated (or counteracted) by seeking more resilient cultivars. Thus, by providing a better picture of the genetic architecture of important traits, we suggest that genomic-assisted breeding is the best alternative to increase genetic gains and shorten the breeding cycle.

In this work, we introduced the concept of climate-smart coffee and relied on the relevance of *C. canephora* for further improvements. Popularly known as Robusta (and Conilon), the species has high socio-economic importance contributing to the livelihoods of millions of smallholder farmers around the world. Beans from Robusta (and Conilon) have been used in the coffee industry as the main source of instant and espresso coffee. With more caffeine and soluble solids, its brewed coffee has a full bodied, plentiful, and thick crema. Beans from *C. canephora* are also very popular in the *blends* with Arabica coffee, reducing costs of the regular coffee and conferring a strong and decisive body to the beverage (Bozzola et al., 2021). Recently, Robusta (and Conilon) coffee has been raised to be the next milestone in the coffee industry (Bozzola et al., 2021). Tracing a parallel with *C. arabica*, the species is more cost efficient, producing more with less requirement for water and pesticides. In the *Diallel population*, for example, we reported an average production of almost 40 bags per hectare, in a condition of partial irrigation. Comparatively, in 2022, the average yield of Arabica coffee production in Brazil was estimated to be 22 bags (60 kg) per hectare (CONAB, 2022). The incidence of disease was also substantially lower, when compared to Arabica cultivars—a fact that is also well-reported in the coffee literature (Van der Vossen et al., 2015). Such features are particularly interesting in a scenario of rising temperatures, erratic rainfall, and more intense extreme weather events that are projected to render certain producing areas less suitable to the existing technologies.

When *C. canephora* is projected as an alternative to supply the coffee chain in the long term, arguably, drink quality is the main challenge. Since the 60s, the coffee industry

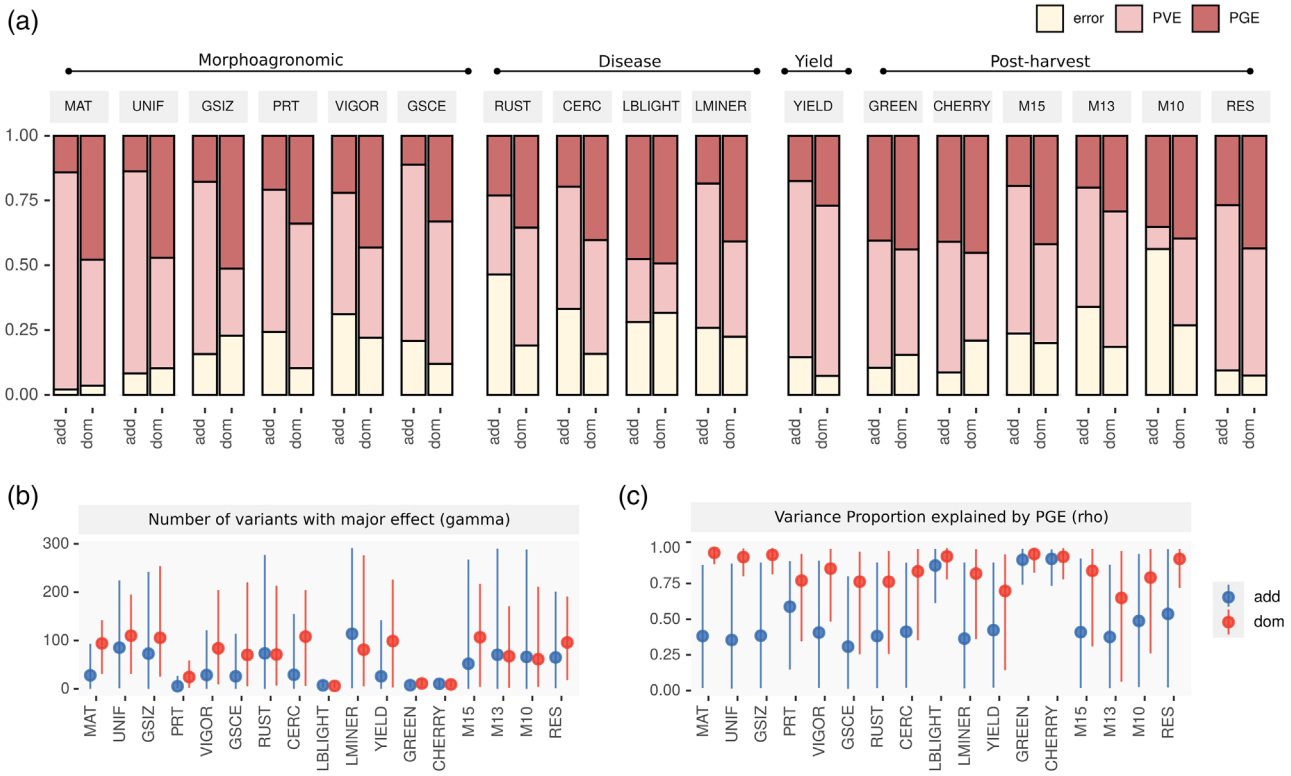

FIGURE 4 (a) Genetic projection using Bayesian sparse linear-mixed model (BSLMM) and the additive (add) and dominance (dom) gene actions models. In BSLMM, the genetic term is divided as: PVE, the proportion of phenotypic variance explained by the polygenic term and PGE, the proportion of the PVE explained by SNPs with a nonzero effect; (b) number of SNPs with measurable associations (gamma) estimated via BSLMM; (c) an approximation to proportion of genetic variance explained by variants with major effect (rho), where rho close to 0 indicates highly polygenic basis, while rho close to unitary values suggests few major effect loci. Vertical lines denote the 97.5% equal-tail probability intervals (ETPIs).

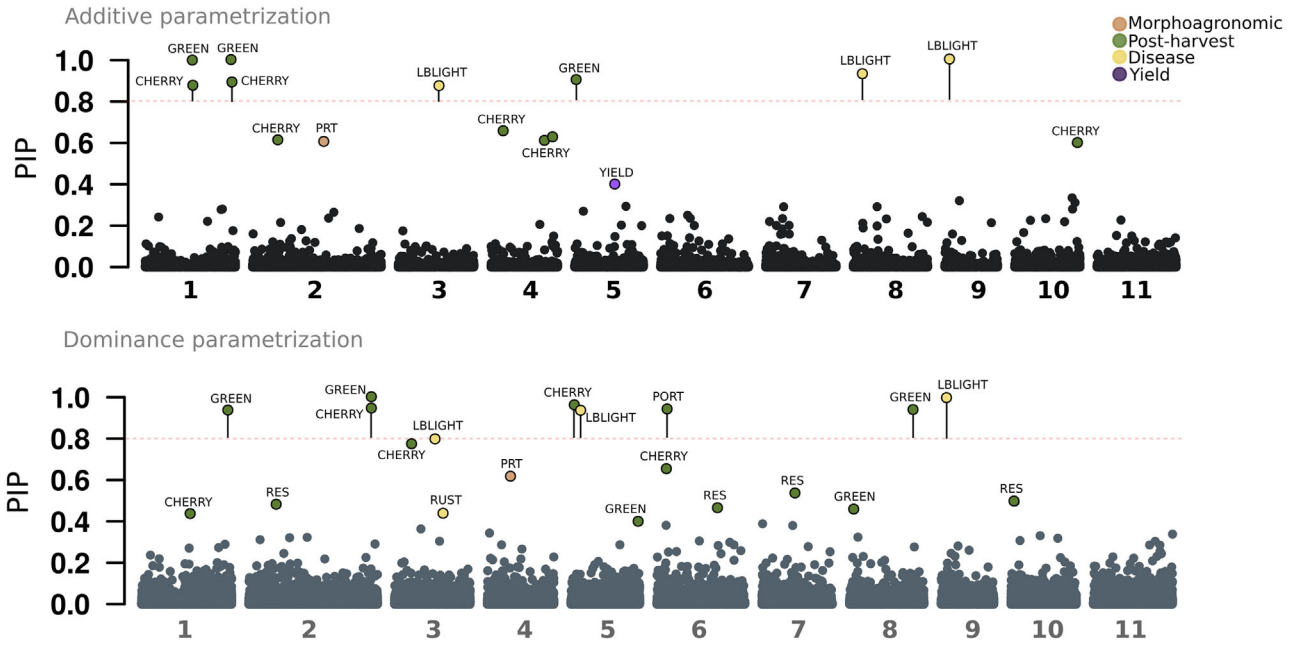

FIGURE 5 Values of posterior inclusion probabilities (PIPs) per 1 kb window for all selected traits in a population of *Coffeea canephora*. Windows with PIP ≥ 0.8 are marked with a vertical line where this threshold is exceeded are indicated. We used a more liberal threshold and indicated traits with PIP > 0.4

has experienced a new movement, in which a new generation of more sophisticated consumers is looking for quality associated with sustainability (Bozzola et al., 2021). So far, few studies have been reported addressing breeding for flavor preference in *C. canephora* (Gamboa-Becerra et al., 2019; Montagnon et al., 1998). Herein, we emphasize that quality can be effectively addressed via plant breeding. Correlation analyses suggested that sensory traits can be improved without compromising other important traits, including yield and disease resistance—a fact that was also reported by Montagnon et al. (1998). In the *Diallel population*, we reported an average value of 78 points for overall liking. As a reference, a minimum score of 80 points has been suggested by specialists to be considered as fine Robusta coffee (Uganda Coffee Development Authority, 2012). We observed some material with values larger than 80 points producing more than 50 bags per hectare, that ultimately justify selection and crosses addressing drink quality and yield performance.

Relying on quality, good coffee has been described as a pleasant sensation, a balanced combination of aroma, flavor, and body in the absence of faults (Seninde & Chambers, 2020; Sunarharum et al., 2014). The *C. canephora* cultivated in Brazil has some unique attributes, including greater bitterness, consistency and astringency, and less acidity and fruity flavor, when compared to Arabica coffee (Ferrão et al., 2019b). Also, it has a distinctive body enriched by chocolate notes and a persistent aftertaste. These aspects make the Robusta beverage genuinely distinct from Arabic coffee. Only in 2010, the CQI created specific standards and protocols to unify the language used to judge Robusta quality. Considering that such standards are relatively recent, modern breeding programs are transitioned to incorporate such evaluations into the breeding pipeline, a fact that is necessary for long-term gains.

4.2 | Nonadditive gene actions can accelerate coffee improvements

Combining ability has long been used by breeders as an important index of hybrid vigor and parental selection. When GCA and SCA are estimated, insights on the relevance of different gene actions on the phenotypic variation are provided. From a practical standpoint, such information is of particular interest because: (i) they can assist in mate allocation; (ii) contribute to improving the selection accuracy when breeding values are predicted on the bases of additive and nonadditive effects; and (iii) can be used to enhance nonadditive genetic variation through the definition of appropriate crossbreeding breeding schemes.

In *C. canephora*, we speculate that the relevance of additive and nonadditive genetic sources is closely related with the domestication history of the species. With a broad scope for regional and historical influences (Cubry et al., 2008; Gomez et al., 2009), the *C. canephora* planted in Brazil retain a large genetic diversity and is mainly structured in two botanical groups: (i) Kouillou (or "Conilon" coffee or SG1 group), that is originated from Central Africa and better adapted to the Brazilian weather and climate conditions; and the (ii) Robusta coffee (SG2 group), originated from the same African region, but with distinct phenotypic characteristics including larger fruits and leaves, resistance to coffee leaf rust, and less resilience to drought conditions (Ferrão et al., 2019b; F. D. E. F. Souza et al., 2013). After investigating the genetic diversity on the parental genotypes (more details in Tables S7 and S8), we hypothesized that genetic gains could be maximized by crossing both pools and obtaining superior crossbred progenies.

Crosses involving genetically distinct pools are a well-reported practice in *C. canephora* (Leroy et al., 1993; Montagnon et al., 2008). Although promising, the relevance of SCA effects (and therefore heterosis) for different traits is not conclusive in the coffee literature. In the Cameroon coffee breeding program, Cilas et al. (2003) and Cilas and Bouharmont (2005) suggested that GCA values were always greater than the SCA effects, when genetic materials from the "Robusta" group were used. More in agreement with our results, Carvalho et al. (2019), Oliveira et al. (2018), and Alkimim et al. (2021) described superior phenotypic performances when genotypes from "Conilon" and "Robusta" groups were crossed. Hybrid vigor was also reported in breeding programs conducted in Côte d'Ivoire, in which the authors reported promising results when crosses between geographically diverse pools were carried out (Montagnon et al., 2008). Collectively, all these results are suggesting that nonadditive sources in coffee, and hence the relevance of SCA effects, might be more prevalent to certain genetic backgrounds and trait-specifics in *C. canephora*.

4.3 | GWAS analysis supports the use of genomic-assisted selection for coffee improvements

Traditionally, genetic architecture of complex traits has been accomplished by scanning recombinant mapping families using a QTL mapping approach. In coffee, QTL have been reported for the incompatibility S locus (Lashermes et al., 1996), pollen viability restoration (Coulibaly et al., 2003), disease resistance (de Almeida et al., 2021), root traits (Achar et al., 2015), fructification time (Akaffou et al., 2003), morphological traits (Michel et al., 2007), and more recently, for yield and quality-related traits (Leroy et al., 2011). Despite the relevance, QTL studies required either a detailed population pedigree or controlled crosses, restricting the conclusions to a certain genetic background (Comeault et al., 2014). Herein, we take advantage of lower levels of LD using a more diverse population and large SNP density to perform

phenotypegenotype associations and infer the genetic basis of complex traits.

Along this study, we emphasized the benefits of estimating genetic parameters using GWAS polygenic models, when compared to single-variant association. Although the implicit genetic model of any complex traits is essentially polygenic, most of the GWAS analyses have been carried out by associating phenotypic variation and each genetic locus independently. Despite the attractiveness of its simplicity, the use of single-SNP regression can be misapplied with respect to our understanding of the underlying genetic mechanism (Guan & Stephens, 2011; Sabatti, 2013; Fernando et al., 2017). In this study, for the first time in the *C. canephora* literature, we reported the use of polygenic GWAS models. When contrasted to conventional single-markers GWAS methods, there is a recent line of evidence supporting its benefits on genetic analyses. For example, Goddard et al. (2016) advocate that it is illogical fitting one SNP at a time as fixed effects, keeping all the rest of the markers as random effects. Similarly, Sabatti (2013) alerted for eventual biased marker effects when relevant regressors are excluded in single-locus GWAS models. In a recent study, Wang and Xu (2019) reported more statistical power when polygenic models were compared to single-SNP versions, while de Los Campos et al. (2023) reported the use of such approach for fine mapping in human studies.

Analogous to the *p* values traditionally presented in the form of Manhattan plots for GWAS analyses, the PIP is a probabilistic metric used for association analyses in the context of polygenic GWAS (Lucas et al., 2018; Wang & Xu, 2019). When traditional single-SNP and polygenic GWAS approaches were compared for the additive effects, we noticed high concordance on the genomic regions pinpointed by both methods for traits with an oligogenic nature (Figures S6–S14). Another important difference on the interpretation of both methods is that polygenic models typically eliminate the tower-like structure observed in traditional Manhattan plots, leaving a single peak standing alone. Wang and Xu (2019) argue that a single peak in association analyses is supposed to be better, because the signal is cleaner and stronger. Comparatively, the absence of multiple testing correction is also a new trend, since the strength of the associations are reported probabilistically using the PIP values. While most studies are reporting a threshold of PIP > 0.1 to declare candidate markers (Armstrong et al., 2018; Chaves et al., 2016; Lucas et al., 2018), herein we used a more conservative value (PIP > 0.4). Another key difference in the model formulation is the use of an extra fixed effects for correction of cryptic relatedness. There is cumulative evidence indicating that the genetic relationships between the individuals can be captured by the markers themselves, without the requirement for additional corrections for the population or sample structure (Kärkkäinen & Sillanpää, 2012).

Using the parameters estimated via BSLMM, we concluded that most of the coffee traits have a complex nature. For traits with a quantitative nature, we argue that genomic-assisted breeding can be more effective when framed in a genomic selection context. Genomic selection relies on high-density genotyping, so that most loci that regulate a trait are in LD with one or more molecular markers. Contrary to genomic selection, a more direct form of marker-assisted selection relies on the use of few molecular markers to predict the genetic merit. For that, markers associated with morphological (PRT), post-harvest (RES, GREEN, and CHERRY), yield stability (YIELD), and disease resistance (RUST and LBLIGHT) showed an oligogenic nature, with few molecular markers accounting for a large portion of the phenotypic variance.

For morphological traits, we highlight a major QTL for the PRT trait identified in the chromosome 2. Expressed later during the coffee development, the full potential of plant architecture is commonly screened in coffee trees after 3 or 4 years of plant development. For the first time reported in the coffee literature, a molecular marker associated with phenotypic variation on PRT could be used to select upright plant architectures in earlier stages and potentially accelerate the breeding process for machine harvest. In a scenario where several countries have faced labor shortages, leveraging the selection of cultivars more prone to the mechanization of harvesting operations can be beneficial for the entire coffee chain.

Also relevant, climate-smart cultivars involve a new generation of cultivars more resilient to infestation of insects and disease. We reported simple genetic architecture for both rust (RUST) and leaf blight (LBLIGHT) —both important coffee diseases. While RUST resistance has long been studied in coffee using family approaches (Alkimim et al., 2017; Pestana et al., 2015), we presented novel evidences of a group of major QTL markers associated with LBLIGHT resistance. The most interesting association was a genomic region mapped in chromosome 3, with a putative gene encoding a disease resistance protein RGA3. Briefly, plants respond to insect and pathogen invasion via pathogen recognition receptors (PRRs) in the cell (Noman et al., 2019). Resistance gene analogs (RGAs) are responsible for intracellular signaling in the cell to activate plant defense genes.

Another region of particular interest was identified in the chromosome 5, with an SNP associated with yield stability. Although it is not technically classified as "major QTL" (PIP < 0.), we encourage future validation studies to target this genomic region, given the importance of the trait for the crop. Among the multiple genes detected in this genomic window, the most interesting was related to protein kinase domain-containing and zinc finger family. Interestingly, both domains were also related to drought tolerance and grain yield in rice, when assayed in an expression profiling trial using

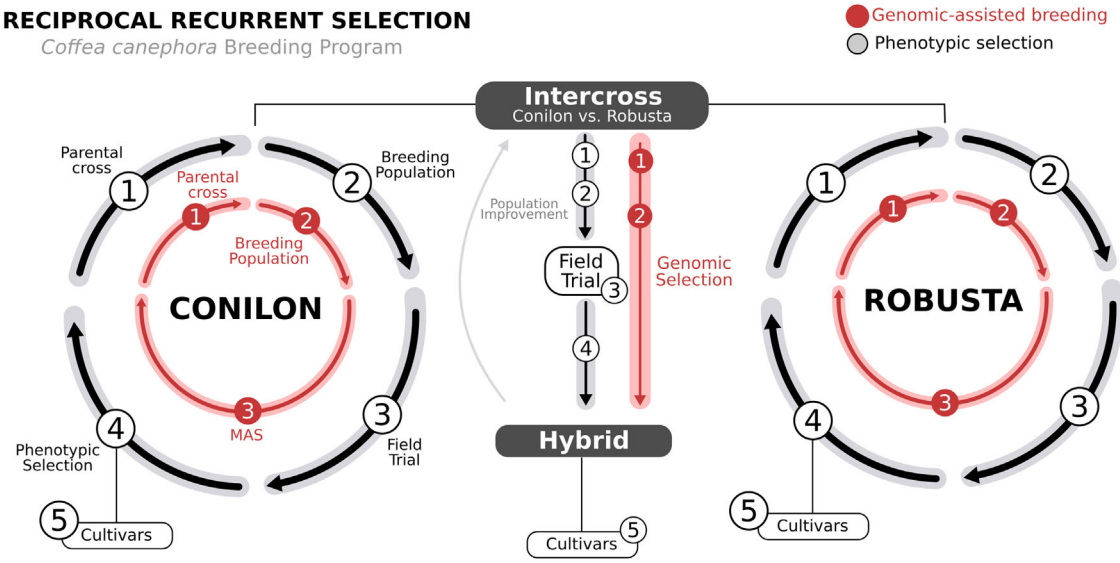

FIGURE 6 Reciprocal recurrent selection assisted by molecular markers applied to *Coffea canephora*. Both botanical groups (Robusta and Conilon) are suggested to be improved independently, following the four major steps of population improvement associated to cyclical breeding design. Rapid-cycle assisted by marker-assisted selection (MAS) is advised. Heterosis is leveraged by selecting the best inter-specific hybrids, based on genomic prediction methodology.

RNAs from stress-treated plants (Jeong et al., 2010). An additional line of evidence was reported by Coelho de Sousa et al. (2022), who also reported markers with large effects in the same chromosome using artificial neural networks for association analyses.

Altogether, for MAS implementation, we have restricted our discussion on three main QTL that can lead to a direct impact on developing climate-smart cultivars by leveraging yield, disease resistance, and harvest mechanization. We also reported other important associations for bean size and outturn index of transforming raw coffee into processed beans (Table S9). Being aware that all biological significances of the associations are primarily speculative, we view the work presented here as just a starting point and believe that our results are sufficiently promising to justify further validations at the molecular level.

4.4 | Genomic-assisted breeding: A new paradigm for fast coffee improvements

Collectively, our results suggest that future coffee efforts should be focused on exploring the hybrid vigor between both botanical groups but assisted by molecular breeding. In *C. arabica*, for example, a first generation of hybrids was developed between 1990 and 2013 by a consortium of research and coffee organizations (Turreira, 2022). Designed as crossbreeding between phylogenetically distant cultivars from American and wild cultivars from Africa, the hybrids gained popularity among growers in Central America because the high production, cup quality, and resistance to pests and diseases (Turreira, 2022; Van der Vossen et al., 2015). For *C. canephora*, we proposed a reciprocal recurrent selection scheme integrated with marker-assisted selection, where crosses between "Robusta" and "Conilon" pools are explored and the best inter-specific hybrids are considered for product development or population improvement (Figure 6). Similar breeding scheme usage was also discussed by Leroy et al. (1993) and Ferrão et al. (2019); here we introduce the novelty of taking into account genomic information, framed in the form of MAS and genomic prediction.

To explore the hybrid vigor, we first stress the importance of identifying contrasting parents in both population pools. To this end, a specific set of molecular markers (e.g., SSR marker system) can be used to classify divergent parents, as suggested in *C. canephora* by Ferrão et al., Rodrigues et al. (2015), and L. C. Souza et al. (2021). After defining the parental crosses, both populations can be improved independently. The selection criteria based on phenotypic selection or via genomic-assisted breeding are notably different. When guided exclusively by visual selection, population improvement is a laborious process that involves large experimental areas and a minimum of four production harvests in coffee to estimate the value per se of a genotype. At this stage, we argue that MAS has its momentum. Markers reported in this study with a large effect on the phenotypic variance of a give trait can assist the selection of promising materials in both genetic pools accelerating the breeding process. As a tangible example, Akpertey et al. (2020) reported SNPs in *C. canephora* assayed using KASP markers. A similar technology could be used in this breeding design.

The second stage in a reciprocal recurrent selection design relies on generating and selecting the best inter-specific hybrids. We propose to use genomic prediction. Unlike the

MAS approach, genomic prediction relies on higher marker density and therefore leads to a larger predictive ability. There are several lines of evidence, including in coffee, supporting the benefits of genomic selection in reducing time and accelerating genetic gains (Alkimim et al., 2020; Carvalho et al., 2019; Ferrão et al., 2017, 2019a; Sousa et al., 2019). However, to be applicable in coffee, we stress the importance of developing a large training population for accurate estimation of genomic estimated breeding values. Future investigations should also consider drink-quality and multi-environmental trails for the success of genomic selection investigation.

A major challenge faced in this study was the fact that the phenotyping *C. canephora* is a perennial crop with a long generation cycle that requires large experimental areas. From a genetic standpoint, the species is subjected to inbreeding depression, late expression of target traits, and susceptibility to seasonal variations. All these aspects make the breeding process strictly challenging and labor-intensive. To circumvent it, we opted to visually assess some of the traits addressed in this study. Although a common practice in plant breeding, visual scores are more prone to errors and subjectivity, even when performed by experienced breeders. This may have been the case of the contrasting results observed in the *Diallel* and *GWAS populations* for the LBLIGHT, MAT, and LMINER traits. We sought to reduce the subjectivity on visual evaluations, by collecting phenotypic data with experienced breeders over multiple years. The use of phenomics is a cutting-edge area prompting further research by the coffee community. Automatic phenotype acquisition based on image analyses has the potential to be included in the breeding design (Figure 6) and improve the throughput and quality of phenotypic data acquisition.

5 | CONCLUSION

Altogether, in this study, we have demonstrated that improvements in *C. canephora* can be accelerated using genomic-assisted breeding. We highlight two main contributions: (i) we draw attention to the importance of nonadditive effects and suggested that future breeding strategies might consider "Robusta" and "Conilon" as independent genetic pools in reciprocal recurrent selection assisted by molecular markers; (ii) we dissected the genetic architecture of coffee traits and highlighted a group of oligogenic and polygenic traits that are more suitable to be assayed in genomic selection and MAS experiments, respectively. Overall, when compared to traditional phenotypic methods, we expect that the methods presented here can maximize future genetic gains and accelerate the breeding of climate-smart coffee cultivars.

AUTHOR CONTRIBUTIONS

Maria Amélia G. Ferrão: Conceptualization; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Writing-original draft; Writing-review & editing. **Aymbire F. A. da Fonseca**: Conceptualization; Project administration; Resources **Paulo S. Volpi**: Data curation. **Lucimara C. de Souza**: Data curation. **Marcone Comério**: Data curation. **Abraão C. Verdin Filho**: Data curation. **Elaine M. Riva-Souza**: Data curation. **Patricio R. Munoz**: Supervision; Writing-review & editing. **Romário G. Ferrão**: Conceptualization; Project administration; Resources. **Luís Felipe V. Ferrão**: Data curation; Formal analysis; Investigation; Methodology; Supervision; Visualization; Writing-original draft.

ACKNOWLEDGMENTS

This work was supported by Fapes (Espírito Santo Research Foundation, Brazil), CNPq (National Council for Scientific and Technological Development, Brazil), and CAPES (Brazil). Additional support was provided by the Instituto Capixaba de Pesquisa, Assitência Técnica e Extensão Rural (Incaper), and Embrapa Cafe. We thank Livia Souza and Anete P. de Souza (CBMEG, Unicamp/Brazil) for their assistance in DNA extraction.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Genotypes in the study belong to the germplasm collection and breeding program of the Incaper institution (ES, Brazil). Phenotypic data used in the diallel analyses are reported in Supporting Information. Genomic and phenotypic traits used in the GWAS analyses are deposited in Dryad https://doi.org/10.5061/dryad.1139fm7.

ORCID

Luís Felipe V. Ferrão https://orcid.org/0000-0002-9655-4838

REFERENCES

Achar, D., Awati, M. G., Udayakumar, M., & Prasad, T. G. (2015). Identification of putative molecular markers associated with root traits in Coffea canephora Pierre ex Froehner. *Molecular Biology International*, *2015*. https://doi.org/10.1155/2015/532386

Akaffou, D. S., Ky, C. L., Barre, P., Hamon, S., Louarn, J., & Noirot, M. (2003). Identification and mapping of a major gene (Ft1) involved in fructification time in the interspecific cross Coffea pseudozanguebariae× C. liberica var. Dewevrei: Impact on caffeine content and seed weight. *Theoretical and Applied Genetics*, *106*, 1486–1490. https://doi.org/10.1007/s00122-003-1207-2

Akpertey, A., Padi, F. K., Meinhardt, L., & Zhang, D. (2020). Effectiveness of Single Nucleotide Polymorphism markers in genotyping

germplasm collections of *Coffea canephora* using KASP assay. *Frontiers in Plant Science*, *11*, 2300.

Alkimim, E. R., Caixeta, E. T., Sousa, T. V., Gois, I. B., Lopes Da Silva, F., Sakiyama, N. S., Zambolim, L., Alves, R. S., & Resende, M. D. V. D. e. (2021). Designing the best breeding strategy for *Coffea canephora*: Genetic evaluation of pure and hybrid individuals aiming to select for productivity and disease resistance traits. *PLoS ONE*, *16*, e0260997. https://doi.org/10.1371/journal.pone.0260997

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

Alkimim, E. R., Caixeta, E. T., Sousa, T. V., Resende, M. D. V., da Silva, F. L., Sakiyama, N. S., & Zambolim, L. (2020). Selective efficiency of genome-wide selection in *Coffea canephora* breeding. *Tree Genetics & Genomes*, *16*, 1–11.

Alves, F. C., Granato, Í. S. C., Galli, G., Lyra, D. H., Fritsche-Neto, R., & de los Campos, G. (2019). Bayesian analysis and prediction of hybrid performance. *Plant Methods*, *15*, 14. https://doi.org/10.1186/s13007-019-0388-x

Amadeu, R. R., Cellon, C., Olmstead, J. W., Garcia, A. A. F., Resende, M. F. R., & Muñoz, P. R. (2016). AGHmatrix : R P*ackage to construct relationship matrices for autotetraploid and diploid species* : A blueberry example 1–10. https://doi.org/10.3835/plantgenome2016.01.0009

Anthony, F., Combes, M., Astorga, C., Bertrand, B., Graziosi, G., & Lashermes, P. (2002). The origin of cultivated Coffea arabica L. varieties revealed by AFLP and SSR markers. *Theoretical and Applied Genetics*, *104*, 894–900. https://doi.org/10.1007/s00122-001-0798-8

Armstrong, C., Richardson, D. S., Hipperson, H., Horsburgh, G. J., Küpper, C., Percival-Alwyn, L., Clark, M., Burke, T., & Spurgin, L. G. (2018). Genomic associations with bill length and disease reveal drift and selection across island bird populations. *Evolution Letters*, *2*, 22–36.

Baker, R. J. (1978). Issues in diallel analysis. *Crop science*, *18*, 533–536. https://doi.org/10.2135/cropsci1978.0011183X001800040001x

Bozzola, M., Charles, S., Ferretti, T., Gerakari, E., Manson, H., Rosser, N., & von der Goltz, P. (2021). *The coffee guide*. International Trade Centre.

Bresadola, L., Caseys, C., Castiglione, S., Buerkle, C. A., Wegmann, D., & Lexer, C. (2019). Admixture mapping in interspecific *Populus* hybrids identifies classes of genomic architectures for phytochemical, morphological and growth traits. *New Phytologist*, *223*, 2076–2089. https://doi.org/10.1111/nph.15930

Butler, D. G., Cullis, B. R., Gilmour, A. R., & Gogel, B. J. (2009). *{ASReml}-R reference manual* (Technical Report). Queensland Department of Primary Industries, Australia.

Carvalho, H. F., Silva, F. L. D. A., Resende, M. D. V. D. E., & Bhering, L. L. (2019). Selection and genetic parameters for interpopulation hybrids between kouilou and robusta coffee. *Bragantia*, *78*, 52–59. https://doi.org/10.1590/1678-4499.2018124

Chaves, J. A., Cooper, E. A., Hendry, A. P., Podos, J., De León, L. F., Raeymaekers, J. A. M., Macmillan, W. O., & Uy, J. A. C. (2016). Genomic variation at the tips of the adaptive radiation of Darwin's finches. *Molecular Ecology*, *25*, 5282–5295. https://doi.org/10.1111/mec.13743

Cilas, C., & Bouharmont, P. (2005). Genetic studies on several bean traits of *Coffea canephora* coffee in Cameroon. *Journal of the Science of Food and Agriculture*, *85*, 2369–2374. https://doi.org/10.1002/jsfa.2245

Cilas, C., Bouharmont, P., & Bar-Hen, A. (2003). Yield stability in *Coffea canephora* from diallel mating designs monitored for 14 years. *Heredity*, *91*, 528–532. https://doi.org/10.1038/sj.hdy.6800351

Cilas, C., Bouharmont, P., Boccara, M., Eskes, A. B., & Baradat, P. h. (1998). Prediction of genetic value for coffee production in *Coffea arabica* from a half-diallel with lines and hybrids. *Euphytica*, *104*, 49–59. https://doi.org/10.1023/A:1018635216182

Coelho De Sousa, I., Nascimento, M., De Castro Sant'anna, I., Teixeira Caixeta, E., Ferreira Azevedo, C., Damião Cruz, C., Lopes Da Silva, F., Ruas Alkimim, E., Campana Nascimento, A. C., & Vergara Lopes Serão, N. (2022). Marker effects and heritability estimates using additive-dominance genomic architectures via artificial neural networks in *Coffea canephora*. *PLoS ONE*, *17*, e0262055. https://doi.org/10.1371/journal.pone.0262055

Comeault, A. A., Soria-Carrasco, V., Gompert, Z., Farkas, T. E., Buerkle, C. A., Parchman, T. L., & Nosil, P. (2014). Genome-wide association mapping of phenotypic traits subject to a range of intensities of natural selection in *Timema cristinae*. *The American Naturalist*, *183*, 711–727. https://doi.org/10.1086/675497

CONAB. (2022). No Title. https://www.conab.gov.br/

Coulibaly, I., Revol, B., Noirot, M., Poncet, V., Lorieux, M., Carasco-Lacombe, C., Minier, J., Dufour, M., & Hamon, P. (2003). AFLP and SSR polymorphism in a *Coffea* interspecific backcross progeny [(*C. heterocalyx*× *C. canephora*)× *C. canephora*]. *Theoretical and Applied Genetics*, *107*, 1148–1155. https://doi.org/10.1007/s00122-003-1355-4

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

Davis, A. P., Mieulet, D., Moat, J., Sarmu, D., & Haggar, J. (2021). Arabica-like flavour in a heat-tolerant wild coffee species. *Nature Plants*, *7*, 413–418. https://doi.org/10.1038/s41477-021-00891-4

De Almeida, D. P., Caixeta, E. T., Moreira, K. F., De Oliveira, A. C. B., De Freitas, K. N. P., Pereira, A. A., Rosado, R. D. S., Zambolim, L., & Cruz, C. D. (2021). Marker-assisted pyramiding of multiple disease resistance genes in coffee genotypes (Coffea arabica). *Agronomy*, *11*, 1763. https://doi.org/10.3390/agronomy11091763

de los Campos, G., Grueneberg, A., Funkhouser, S., Pérez-Rodríguez, P., & Samaddar, A. (2023). "Fine mapping and accurate prediction of complex traits using Bayesian Variable Selection models applied to biobank-size data." *European Journal of Human Genetics*, *31*, 113–320. https://doi.org/10.1038/s41431-022-01135-5

Denoeud, F., Carretero-paulet, L., Dereeper, A., Guyot, R., Pietrella, M., Zheng, C., Alberti, A., Aprea, G., Aury, J., Bento, P., Bernard, M., Campa, C., Cenci, A., Combes, M., Crouzillat, D., Daddiego, L., Bellis, F. D. E., Garsmeur, O., Gayraud, T., … Lashermes, P. (2014). The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science*, *345*, 1181–1184.

Dereeper, A., Bocs, S., Rouard, M., Guignon, V., Ravel, S., Tranchant-Dubreuil, C., Poncet, V., Garsmeur, O., Lashermes, P., & Droc, G. (2015). The coffee genome hub: A resource for coffee genomes. *Nucleic Acids Research*, *43*, D1028–D1035. https://doi.org/10.1093/nar/gku1108

Elshire, R. J., Glaubitz, J. C., Sun, Q. I., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, *6*, e19379. https://doi.org/10.1371/journal.pone.0019379

Fernando, R., Toosi, A., Wolc, A., Garrick, D., & Dekkers, J. (2017). Application of whole-genome prediction methods for genome-wide association studies: A Bayesian approach. *Journal of Agricultural, Biological and Environmental Statistics*, *22*, 172–193. https://doi.org/10.1007/s13253-017-0277-6

Ferrão, L. F. V., Johnson, T. S., Benevenuto, J., Edger, P. P., Colquhoun, T. A., & Munoz, P. R. (2020). Genome-wide association of volatiles reveals candidate loci for blueberry flavor. *New Phytologist*, *226*, 1725–1737. https://doi.org/10.1111/nph.16459

Ferrão, L. F. V., Ferrão, R. G., Ferrão, M. A. G., Fonseca, A., Carbonetto, P., Stephens, M., & Garcia, A. A. F. (2019a). Accurate genomic prediction of *Coffea canephora* in multiple environments using whole-genome statistical models. *Heredity*, *122*, 261–275. https://doi.org/10.1038/s41437-018-0105-y

Ferrão, L. F. V., Ferrão, R. G., Ferrão, M. A. G., Francisco, A., & Garcia, A. A. F. (2017). A mixed model to multiple harvest-location trials applied to genomic prediction in Coffea canephora. *Tree Genetics & Genomes*, *13*, 95.

Ferrão, R. G., de Muner, L. H., da Fonseca, A. F. A., & Ferrão, M. A. G. (2019b). *Conilon coffee*. Incaper.

Gamboa-Becerra, Hernández-Hernández, González-Ríos, Suárez-Quiroz, Gálvez-Ponce, Ordaz-Ortiz, & Winkler, (2019). Metabolomic markers for the early selection of coffea canephora plants with desirable cup quality traits. *Metabolites*, *9*, 214. https://doi.org/10.3390/metabo9100214

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q. I., & Buckler, E. S. (2014). *TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline*. *PLOS ONE*, *9*. https://doi.org/10.1371/journal.pone.0090346

Goddard, M. E., Kemper, K. E., MacLeod, I. M., Chamberlain, A. J., & Hayes, B. J. (2016). Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms and genetic architecture. Proceedings of the Royal Society B, *283*(1835), https://doi.org/10.1098/rspb.2016.0569

Gomez, C., Dussert, S., Hamon, P., Hamon, S., De Kochko, A., & Poncet, V. (2009). Current genetic differentiation of Coffea canephora Pierre ex A. Froehn in the Guineo-Congolian African zone: Cumulative impact of ancient climatic changes and recent human activities. *BMC Evolutionary Biology*, *9*, 1–19. https://doi.org/10.1186/1471-2148-9-167

Gompert, Z., Brady, M., Chalyavi, F., Saley, T. C., Philbin, C. S., Tucker, M. J., Forister, M. L., & Lucas, L. K. (2019). Genomic evidence of genetic variation with pleiotropic effects on caterpillar fitness and plant traits in a model legume. *Molecular Ecology*, *28*, 2967–2985. https://doi.org/10.1111/mec.15113

Gompert, Z., Lucas, L. K., Nice, C. C., & Buerkle, C. A (2013). Genome divergence and the genetic architecture of barriers to gene flow between Lycaeides idas and L. melissa. *Evolution; Internation Journal of Organic Evolution*, *67*, 2498–2514. https://doi.org/10.1111/evo.12021

Guan, Y., & Stephens, M. (2011). Bayesian variable selection regression for genome-wide association studies and other large-scale problems. *The Annals of Applied Statistics*, *5*, 1780–1815.

Imbach, P., Fung, E., Hannah, L., Navarro-Racines, C. E., Roubik, D. W., Ricketts, T. H., Harvey, C. A., Donatti, C. I., Läderach, P., Locatelli, B., & Roehrdanz, P. R. (2017). Coupling of pollination services and coffee suitability under climate change. *Proceedings of the national academy of sciences*, *114*, 10438–10442. https://doi.org/10.1073/pnas.1617940114

Ishimori, M., Takanashi, H., Hamazaki, K., Atagi, Y., Kajiya-Kanegae, H., Fujimoto, M., Yoneda, J., Tokunaga, T., Tsutsumi, N., & Iwata, H. (2020). Dissecting the genetic architecture of biofuel-related traits in a sorghum breeding population. *G3: Genes, Genomes, Genetics*, *10*, 4565–4577. https://doi.org/10.1534/g3.120.401582

Jeong, J. S., Kim, Y. S., Baek, K. H., Jung, H., Ha, S. -H., Do Choi, Y., Kim, M., Reuzeau, C., & Kim, J. u. -K. (2010). Root-specific expression of OsNAC10 improves drought tolerance and grain yield in rice under field drought conditions. *Plant Physiology*, *153*, 185–197. https://doi.org/10.1104/pp.110.154773

Kärkkäinen, H. P., & Sillanpää, M. J. (2012). Robustness of Bayesian multilocus association models to cryptic relatedness. *Annals of Human Genetics*, *76*, 510–523.

Krishnan, S., Matsumoto, T., Nagai, C., Falconer, J., Shriner, S., Long, J., Medrano, J. F., & Vega, F. E. (2021). Vulnerability of coffee (*Coffea* spp.) genetic resources in the United States. *Genetic Resources and Crop Evolution*, *68*, 2711.

Lashermes, P., Combes, M. - C., Robert, J., Trouslot, P., D'hont, A., Anthony, F., & Charrier, A. (1999). Molecular characterisation and origin of the Coffea arabica L. genome. *Molecular and General Genetics MGG*, *261*, 259–266. https://doi.org/10.1007/s004380050965

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

Lehermeier, C., de los Campos, G., Wimmer, V., & Schön, C.-C. (2017). Genomic variance estimates: With or without disequilibrium covariances? *Journal of Animal Breeding and Genetics*, *134*, 232–241. https://doi.org/10.1111/jbg.12268

Leroy, T., De Bellis, F., Legnate, H., Kananura, E., Gonzales, G., Pereira, L. F., Andrade, A. C., Charmetant, P., Montagnon, C., & Cubry, P. (2011). Improving the quality of African robustas: QTLs for yield-and quality-related traits in Coffea canephora. *Tree Genetics & Genomes*, *7*, 781–798.

Leroy, T., De Bellis, F., Legnate, H., Musoli, P., Kalonji, A., Loor Solórzano, R. G., & Cubry, P. (2014). Developing core collections to optimize the management and the exploitation of diversity of the coffee Coffea canephora. *Genetica*, *142*, 185–199. https://doi.org/10.1007/s10709-014-9766-5

Leroy, T., Montagnon, C., Charrier, A., & Eskes, A. B. (1993). Reciprocal recurrent selection applied to Coffea canephora Pierre. I. Characterization and evaluation of breeding populations and value of intergroup hybrids. *Euphytica*, *67*, 113–125. https://doi.org/10.1007/BF00022734

Leroy, T., Montagnon, C., Cilas, C., Yapo, A., Charmetant, P., & Eskes, A. B. (1997). Reciprocal recurrent selection applied to Coffea canephora Pierre. III. Genetic gains and results of first cycle intergroup crosses. *Euphytica*, *95*, 347–354. https://doi.org/10.1023/A:1003074716379

Lloyd-Jones, L. R., Robinson, M. R., Moser, G., Zeng, J., Beleza, S., Barsh, G. S., Tang, H., & Visscher, P. M. (2017). Inference on the genetic basis of eye and skin color in an admixed population via Bayesian linear mixed models. *Genetics*, *206*, 1113–1126. https://doi.org/10.1534/genetics.116.193383

Lucas, L. K., Nice, C. C., & Gompert, Z. (2018). Genetic constraints on wing pattern variation in Lycaeides butterflies: A case study on mapping complex, multifaceted traits in structured populations. *Molecular Ecology Resourcers*, *18*, 892–907.

Michel, N., Serge, H., & Valérie, P. (2007). Genetic basis of species differentiation between Coffea liberica Hiern and C. canephora Pierre: Analysis of an interspecific cross. *Genetic Resources and Crop Evolution*, *54*, 1011–1021.

Mohammed, W. (2011). Heterosis and combining ability analysis for coffee quality in diallel crosses of diverse coffee (Coffea arabica L.) parents in origin. *East African Journal of Sciences*, *5*, 12–21.

Montagnon, C., Guyot, B., Cilas, C., & Leroy, T. (1998). Genetic parameters of several biochemical compounds from green coffee, Coffea canephora. *Plant Breeding*, *117*, 576–578. https://doi.org/10.1111/j.1439-0523.1998.tb02211.x

Montagnon, C., Leroy, T., Cilas, C., & Charrier, A. (2003). Heritability of Coffea canephora yield estimated from several mating designs. *Euphytica*, *133*, 209–218. https://doi.org/10.1023/A:1025543805652

Montagnon, C., Leroy, T., Cilas, C., Legnaté, H., & Charrier, A. (2008). Heterozygous genotypes are efficient testers for assessing between-population combining ability in the reciprocal recurrent selection of Coffea canephora. *Euphytica*, *160*, 101–110. https://doi.org/10.1007/s10681-007-9561-9

Muñoz, P. R., Resende, M. F. R., Gezan, S. A., Resende, M. D. V., de los Campos, G., Kirst, M., Huber, D., & Peter, G. F. (2014). Unraveling additive from nonadditive effects using genomic relationship matrices. *Genetics*, *198*, 1759–1768. https://doi.org/10.1534/genetics.114.171322

Mustiga, G. M., Gezan, S. A., Phillips-Mora, W., Arciniegas-Leal, A., Mata-Quirós, A., & Motamayor, J. C. (2018). Phenotypic description of Theobroma cacao L. for yield and vigor traits from 34 hybrid families in Costa Rica based on the genetic basis of the parental population. *Frontiers in Plant Science*, *9*, 808. https://doi.org/10.3389/fpls.2018.00808

Nariai, N., Greenwald, W. W., Deboever, C., Li, H. E., & Frazer, K. A. (2017). Efficient prioritization of multiple causal eQTL variants via sparse polygenic modeling. *Genetics*, *207*, 1301–1312. https://doi.org/10.1534/genetics.117.300435

Noman, A., Aqeel, M., & Lou, Y. (2019). PRRs and NB-LRRs: From signal perception to activation of plant innate immunity. *International Journal of Molecular Sciences*, *20*, 1882. https://doi.org/10.3390/ijms20081882

Oliveira, L. N. L. D. E., Rocha, R. B., Ferreira, F. M., Spinelli, V. M., Ramalho, A. R., & Teixeira, A. L. (2018). Selection of Coffea canephora parents from the botanical varieties Conilon and Robusta for the production of intervarietal hybrids. *Ciência Rural*, *48*. https://doi.org/10.1590/0103-8478cr20170444

Ovalle-Rivera, O., Läderach, P., Bunn, C., Obersteiner, M., & Schroth, G. (2015). Projected shifts in Coffea arabica suitability among major global producing regions due to climate change. *PLoS ONE*, *10*, e0124155. https://doi.org/10.1371/journal.pone.0124155

Pérez, P., & de los Campos, G. (2014). Genome-Wide Regression and Prediction with the BGLR Statistical Package. *Genetics*, *198*, 483–495. https://doi.org/10.1534/genetics.114.164442

Pestana, K. N., Capucho, A. S., Caixeta, E. T., de Almeida, D. P., Zambolim, E. M., Cruz, C. D., Zambolim, L., Pereira, A. A., de Oliveira, A. C. B., & Sakiyama, N. S. (2015). Inheritance study and linkage mapping of resistance loci to Hemileia vastatrix in Híbrido de Timor UFV 443-03. *Tree genetics & genomes*, *11*, 1–13.

Pritchard, J. K., Stephens, M., Rosenberg, N. A., & Donnelly, P. (2000). Association Mapping in Structured Populations. *The American Journal of Human Genetics*, *67*, 170–181. https://doi.org/10.1086/302959

Rosyara, U. R., De Jong, W. S., Douches, D. S., & Endelman, J. B. (2016). Software for genome-wide association studies in autopolyploids and its application to potato. *The Plant Genome*, *9*. https://doi.org/10.3835/plantgenome2015.08.0073

Sabatti, C. (2013). Multivariate linear models for GWAS. In K-A. Doh, Z. S. Qin, & M. Vannucci (Eds.), *Advances in statistical bioinformatics* (pp. 188–207). Cambridge University Press.

Sant'ana, G. C., Pereira, L. F. P., Pot, D., Ivamoto, S. T., Domingues, D. S., Ferreira, R. V., Pagiatto, N. F., Da Silva, B. S. R., Nogueira, L. M., Kitzberger, C. S. G., Scholz, M. B. S., De Oliveira, F. F., Sera, G. H., Padilha, L., Labouisse, J.-. P., Guyot, R., Charmetant, P., & Leroy, T. (2018). Genome-wide association study reveals candidate genes influencing lipids and diterpenes contents in Coffea arabica L. *Scientific Reports*, *8*, 1–12. https://doi.org/10.1038/s41598-017-18800-1

Seninde, D. R., & Chambers, E. (2020). Coffee flavor: A review. *Beverages*, *6*, 44. https://doi.org/10.3390/beverages6030044

Silvestrini, M., Junqueira, M. G., Favarin, A. C., Guerreiro-Filho, O., Maluf, M. P., Silvarolla, M. B., & Colombo, C. A. (2007). Genetic diversity and structure of Ethiopian, Yemen and Brazilian Coffea arabica L. accessions using microsatellites markers. *Genetic Resources and Crop Evolution*, *54*, 1367–1379. https://doi.org/10.1007/s10722-006-9122-4

Sousa, T. V., Caixeta, E. T., Alkimim, E. R., Oliveira, A. C. B., Pereira, A. A., Sakiyama, N. S., Zambolim, L., & Resende, M. D. V. (2019). Early selection enabled by the implementation of genomic selection in Coffea arabica breeding. *Frontiers in Plant Science*, *9*, 1934. https://doi.org/10.3389/fpls.2018.01934

Souza, F. D. E. F., Caixeta, E. T., Ferrão, L. F. V., Pena, G. F., Sakiyama, N. S., Zambolim, E. M., Zambolim, L., & Cruz, C. D. (2013). Molecular diversity in Coffea canephora germplasm conserved and cultivated in Brazil. *Crop Breeding and Applied Biotechnology*, *13*. https://doi.org/10.1590/S1984-70332013000400001

Souza, L. C., Ferrão, M. A. G., Carvalho, R. D., Ferrão, R. G., Fonseca, A. F. A. D. A., Pinheiro, P. F., & Soares, T. C. B. (2021). Molecular characterization of parents and hybrid progenies of conilon coffee. *Anais da Academia Brasileira de Ciencias*, 93. https://doi.org/10.1590/0001-3765202120201649

Sunarharum, W. B., Williams, D. J., & Smyth, H. E. (2014). Complexity of coffee flavor: A compositional and sensory perspective. *Food Research International*, *62*, 315–325. https://doi.org/10.1016/j.foodres.2014.02.030

Turreira-García, N. (2022). Farmers' perceptions and adoption of *Coffea arabica* F1 hybrids in Central America. *World Development Sustainability*, *1*, 100007. https://doi.org/10.1016/j.wds.2022.100007

Uganda Coffee Development Authority. (2012). Uganda coffee. https://ugandacoffee.go.ug/sites/default/files/Resource_center/UCDAAnnualReport_2011-2012_0.pdf

Vanraden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, *91*, 4414–4423. https://doi.org/10.3168/jds.2007-0980

Vitezica, Z. G., Varona, L., & Legarra, A. (2013). On the additive and dominant variance and covariance of individuals within the genomic selection scope. *Genetics*, *195*, 1223–1230. https://doi.org/10.1534/genetics.113.155176

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*, 243–256. https://doi.org/10.1007/s10681-015-1398-z

Walyaro, D. J. A. (1983). *Considerations in breeding for improved yield and quality in arabica coffee (*Coffea arabica *L.)* [Doctoral thesis, Agricultural University Wageningen].

Wang, M., & Xu, S. (2019). A coordinate descent approach for sparse Bayesian learning in high dimensional QTL mapping and genome-wide association studies. *Bioinformatics*, *35*, 4327–4335.

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., Mcmullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S., & Buckler, E. S. (2006). *A unified mixed-model method for association mapping that accounts for multiple levels of relatedness*, *38*, 203–208. https://doi.org/10.1038/ng1702

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*, 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

Zhou, X., Carbonetto, P., & Stephens, M. (2013). Polygenic modeling with Bayesian sparse linear mixed models. *PLoS Genetics*, *9*, e1003264. https://doi.org/10.1371/journal.pgen.1003264

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ferrão, M. A. G., da Fonseca, A. F. A., Volpi, P. S., de Souza, L. C., Comério, M., Filho, A. C. V., Riva-Souza, E. M., Munoz, P. R., Ferrão, R. G., & Ferrão, L. F. V. (2023). Genomic-assisted breeding for climate-smart coffee. *The Plant Genome*, e20321. https://doi.org/10.1002/tpg2.20321