Genet Resour Crop Evol
https://doi.org/10.1007/s10722-022-01498-0

RESEARCH ARTICLE

SNP markers found in non-coding regions can distinguish among low-variant genotypes of Arabica and other coffee species

Alejandro Bolívar-González · Ramón Molina-Bravo · William Solano-Sánchez · Emanuel Araya-Valverde · Suzana T. Ivamoto-Suzuki · Luiz F. P. Pereira · Andrés Gatica-Arias

Received: 6 May 2022 / Accepted: 20 November 2022

Abstract Development of efficient and scalable methods for molecular identification of *Coffea* spp. are necessary to accelerate studies related to the characterization of germplasm for both conservation or breeding purposes, and the validation of coffee germplasm. The low genetic diversity of coffee hinders the establishment of protocols that facilitate the molecular characterization of a given genotype. In this study, nucleotide variability was analyzed at 22 loci in the genome of 19 coffee accessions using de novo primer sets and high-resolution melting (HRM). Single nucleotide polymorphisms (SNPs) variants were studied in coding regions of genes implicated in sucrose accumulation in the seed, *Sucrose synthase 2* (*SUS2*), *Ent-kaurene oxidase 1* (*CaKO1)*, and *Caffeoyl-coenzyme A 3-O-methyltransferase* (*CcOAOMT).* The non-coding *Internal transcribed spacer 2* (*ITS2)* region was also studied. Variability was shown at 103 positions both at the interspecies level (15 loci) and among varieties of *Coffea arabica* L. (4 loci). The HMR technique for identification of variants in genes *CaKO1*, *SUS2*, *CcoAOMT,* as well as in the *ITS2* region proved to be a robust technique for germplasm characterization. More important this technique can be used for fingerprinting and traceability of coffee grain exports which is an increasing market-consumer demand.

Keywords Coffee · Single-nucleotide polymorphism · High-resolution melting analysis · HRM

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10722-022-01498-0.

A. Bolívar-González · A. Gatica-Arias (✉)
Laboratorio Biotecnología de Plantas, Escuela de Biología, Universidad de Costa Rica, San Pedro, Costa Rica
e-mail: andres.gatica@ucr.ac.cr

A. Bolívar-González · A. Gatica-Arias
Programa de Posgrado en Ciencias Agrícolas y Recursos Naturales con Énfasis en Biotecnología (PPCARN), Universidad de Costa Rica, San Pedro, Costa Rica

R. Molina-Bravo
Programa de Biotecnología Vegetal y Recursos Genéticos para el Fitomejoramiento (BIOVERFI), Universidad Nacional, Heredia, Costa Rica

W. Solano-Sánchez
CATIE (Centro Agronómico Tropical de Investigación y Enseñanza), Turrialba, Cartago, Costa Rica

E. Araya-Valverde
Centro Nacional de Innovaciones Biotecnológicas, CENIBiot-CeNAT, San José, Costa Rica

S. T. Ivamoto-Suzuki · L. F. P. Pereira
Area of Genetic Improvement and Plant Propagation, Paraná Rural Development Institute, Londrina, Paraná 86047-902, Brazil

Published online: 10 December 2022

Introduction

Grown in tropical countries and subtropical regions (van der Vossen et al. 2015), coffee is an important agricultural product as it represents one of the most traded commodities in international markets (Cheng et al. 2016; Tran et al. 2018). There are more than 100 species that belong to *Coffea*, but only two are grown extensively (Sousa et al. 2017): *Coffea arabica* L. (Arabica coffee), which represents 71.5% of the world production, while the remaining 28.5% corresponds to *C. canephora* Pierre ex Froehner (Robusta coffee) (ICO 2022). Consumer preference for Arabica coffee is due to its greater acidity, low bitterness and better overall flavor compared to Robusta (Combes et al. 2018).

Varieties of Arabica coffee are highly genetically similar. Low genetic variability among Arabica species has been attributed to the processes of domestication and dissemination and the primarily autogamous mode of reproduction (Anthony et al. 2002). Arabica coffee is the only tetraploid species (2n = 4x = 44) of *Coffea*, thus there are high levels of incompatibility with other diploid coffee species (Bertrand et al. 2003). The low genetic diversity extends through genetically improved lines of Arabica coffee and is recognized as a highly relevant limitation to improve coffee plants tolerance against to abiotic and biotic stress (Merot-L'anthoene et al. 2019).

Genetic improvement of *Coffea* spp. has produced four large groups of plants with characteristics in common: (*i*) the traditional varieties Typica and Bourbon, and genotypes derived from their crosses; (*ii*) local Ethiopian varieties, generally isolated from other commercial varieties; (*iii*) introgressions, such as Sarchimores and Catimores lines that include interspecific crosses; and (*iv*) F1 hybrids, first generation descendants of genetically distinct parental Arabica lines (WCR 2018).

To date, the majority of studies that have determine parentage and genetic diversity among commercial cultivars of *C. arabica* that have been used are microsatellites (SSR) (Anthony et al. 2001; Cubry et al. 2008; Motta et al. 2014; Sánchez et al. 2020), amplified fragment length polymorphisms (AFLP) (Anthony et al. 2001) and, more recently, sequence-related amplified polymorphisms (SRAP) (Jingade et al. 2019), and single nucleotide polymorphisms (SNP) (Zhou et al. 2016; Scalabrin et al. 2020; Zhang et al. 2020, 2021). In general, these studies have shown the low genetic variability that predominates among Arabica coffee species, which can be to a single spontaneous event of hybridization of *C. canephora* and *C. eugenioides* that gave rise to all the germplasm of *C. arabica* (Scalabrin et al. 2020). SNPs represent one of the most abundant variations in genomes (Huq et al. 2016). In some cases, SNPs can arise in protein-coding regions and modify the plant phenotype (Ganal et al. 2009; Huq et al. 2016), and they are considered as a model for the study of the function of certain genes in a desired phenotype (Huq et al. 2016).

In this sense, a DNA-based protocol for the detection and quantification of adulterations in ground coffee based on the identification of SNPs at chloroplast loci between *C. arabica* and *C. canephora*, and subsequent high resolution melting genotyping (HRM) has been reported (Combes et al. 2018). Moreover, through next generation sequencing data analysis of different coffee genotypes genomes, 11 SNPs associated with genes encoding enzymes involved in the conversion of substrates of the caffeine biosynthesis pathway have been identified (Tran et al. 2018). A similar study in tea plants found that SNPs located in the coding regions of the *TCS1* (Tea caffeine synthase), *TIDH* (inosine-5′-monophosphate dehydrogenase), and *sAMS* (S-adenosyl-L-methionine synthetase) genes were positively correlated with varying caffeine and theobromine contents (Li et al. 2014). On the other hand, the sequence of *DREB1D* genes from 38 genotypes of coffee and found variations associated with the differential drought tolerance of each genotype has been analyzed (Alves et al. 2018). Therefore, sequencing of genomic regions related to the metabolism of sugars, aromatic compounds, and defense responses in a greater number of genotypes could contribute to the discovery of genetic variants that lead to quantitative variations in these characteristics (Tran et al. 2018). In addition, identification of these variants would allow differentiation among genotypes or establishment of hybrid parentage using molecular techniques for detection (Zhou et al. 2016).

High resolution melting analysis (HRM) is an easy and rapid technique that can be adapted for SNPs or small INDEL variations analyses (Cruz et al. 2013; Simko 2016). Variations in nucleotide sequences affect the stability of short DNA sequences and modify the melting temperature (Tm) of the molecule (Clevenger 2015). The presence of alleles in a heterozygous condition can also be detected by HRM. Therefore, the objective of the present study was to determine the effectiveness of HRM analysis for detection of

variations in nucleotide sequences of genes related to important agronomic traits (cup quality, disease resistance, and morphology) from different accessions of coffee (*Coffea* spp.) and to estimate nucleotide variability in coding and non-coding regions of these genes.

Material and methods

Plant material and sampling permits

A total of 19 accessions of the genus *Coffea* were selected to make up a representative group of coffee plants used to analyze the natural variability in the nucleotide sequences of several genes (Supplementary Table 1). These accessions included 15 cultivars of *C. arabica* L. and individuals of *C. canephora* Pierre ex Froehner, *C. liberica* Hiern, *Coffea zanguebariae* Lour., and *C. eugenioides* A. Cunn. Each accession was represented by one plant. Data on the collection site of the material, the genetic group of the accessions and their family are shown in Supplementary Table 1.

Plant material collected in the field consisted of young leaves located near to the first or second node below the apical meristem. The leaves were immediately stored in air-tight bags in a cooler and later stored at – 80 ºC. Leaves were lyophilized for 48 h at a condenser temperature of – 50 °C using an Alpha 1–2 LDplus freeze-dryer (CHRIST, Osterode am Harz, Germany).

All the needed permits for sampling leaf tissue from coffee were requested by Comisión de Biodiversidad-UCR at Universidad de Costa Rica (under permit # 323–2021), Centro Agronómico Tropical de Enseñanza e Investigación (CATIE) and Centro de Investigación del Instituto del Café de Costa Rica (CICAFE).

Selection of genes of agronomic interest

Regions of 13 genes [(*ClpP*; NC_008535.1), (*CCD1*; evm.TU.Scaffold_616.20), (*CaWRKY1*; evm.TU.Scaffold_539.575), (*LOX1_5*; evm.TU.Scaffold_632.448), (*Pto*; evm.TU.Scaffold_2093.51), (*DXMT*; evm.TU.Scaffold_1614.62), (*CTgS1*; evm.TU.Scaffold_588.56), (*SUS2*; evm.TU.Scaffold_2016.44), (*CaKO*; evm.TU.Scaffold_522.857), (*CCoAOMT1*; evm.TU.Scaffold_634.1045), (*CaPOP1*; evm.TU.Scaffold_214.1162), (*ERF1*; evm.TU.Scaffold_624.950)] implicated in coffee agronomic characteristics and one internal transcribe spacer (*ITS2*; EU650386.1) were selected for evaluation of nucleotide sequence variation among accessions (Supplementary Table 2). Genomic resources for *Coffea* available in the NCBI sequence database (GenBank) and Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html) were used to design primers to amplify regions between 900 and 1200 bp. The regions included exons and introns of the sequences (Supplementary Table 3). The software Primer3Plus (Untergasser et al. 2012) was used for primer sequence design using the following parameters: primer size (18–20–25 bp), Tm (55–60–65), %GC (40–50–60), maximum Tm difference (3ºC), maximum Poly-X (3), GC clamp (1), maximum GC in 3' end (2), thermodynamic parameters (Santa Lucia 1998), formula for salt correction (Santa Lucia 1998).

Establishment of an amplification strategy for the selected loci

DNA was extracted using CTAB extraction method (Doyle and Doyle 1987) with modifications (Bolívar-González et al. 2018). Reactions contained 70 ng of DNA, 1X Taq Buffer, 1.5 mM $MgCl_2$, 0.5 mM dNTPs (Promega™, Madison, USA), 0.2 µM primers, 0.01 mg BSA, and 0.5 U Taq DNA Polymerase in a total volume of 25 µl. All reagents, unless otherwise stated, were purchased from Thermo Scientific™ (Delaware, USA). A Touchdown-PCR (TD-PCR) was performed in a Professional Thermocycler (Biometra, Göttingen, Germany) as indicated in Supplementary Table 4. PCR products were analyzed by electrophoresis in 1.2% (m/v) agarose gels to confirm its suitability for sequencing.

Analysis of sequence data

Twenty-two purified PCR products corresponding to the genes described in Supplementary Table 2 were sequenced once with the Sanger method using the forward primer in a commercial laboratory (Macrogen, South Korea).

FASTA files and chromatograms were analyzed manually to detect possible patterns of sequencing errors. The Sequencing Analysis 5.3.1 (Applied Biosystems™, California, USA) program was

used for base calling and removal of sections with low sequencing signals using the analysis protocol 3730BDTv3-KB-DeNovo_v5.2, base calling (default parameters), assign N (NO), mixed-heterozygous bases (25%). The first 20 bases and all bases after the 900 bp position were removed.

The homology of the sequences obtained was corroborated by BLAST results as well as multiple alignment. The generated sequences were used to perform the multiple sequence alignment between the accessions and the reference sequence reported in Supplementary Table 2. In cases where high-quality processed sequences were not generated, they were discarded. The alignments were performed with the MAFFT algorithm (https://www.ebi.ac.uk/Tools/msa/mafft/), the results of the alignments were visualized with the CLC sequence viewer 8.0.0 (Qiagen Bioinformatics, Aarhus C, Denmark).

Identification of variants from FASTA sequences

Polymorphic sites were identified using the program NovoSNP 3.0.1 (http://novosnp.bioinf.be/). The files (.ab1) obtained from the sequencing of each PCR product were compared with reference sequences of each *Coffea* gene available at GenBank or Phytozome using default parameters (Weckx et al. 2005). Complete list of the variable sites was filtered by *F*-Score value (> 15). This parameter considers the three bases contiguous to the site of the variation and identifies false positives based on the quality of the chromatogram. Chromatograms were then manually inspected case by case to validate the polymorphisms. Each polymorphism was categorized according to the number of genotypes discriminated and its location in coding or non-coding regions of the genes analyzed. For the analysis of results, the statistics of the sequences of the two PCR products corresponding to each gene were combined.

HRM-PCR analysis of SNP identified in coffee accessions

A panel of 20 polymorphisms (INDELs and SNPs) was defined based on the potential of these sites to discriminate different alleles by melting curve analysis of specific PCR products. The nature of the polymorphism (its effect on the change in melting temperature of the region) and the number or type of distinguishable genotypes determined. PCR products were obtained using specifications reported for HRM analysis using a nested PCR protocol with initial products used as template DNA (Supplementary Table 3).

A MeltDoctor HRM Master Mix kit (TermoFischer Scientific ™, Waltham, Massachusetts, USA) was used following a previously reported protocol (Szurman-Zubrzycka et al. 2017) with minor modifications: 1 µL PCR product (1:1000 dilution), 5 µL MeltDoctor™ HRM Master Mix, 1 µL of each primer (4 µM), 0.8 µL $MgCl_2$ (25 mM), and 1.2 µL PCR water to obtain a reaction volume of 10 µL. A StepOne ™ Real-Time PCR instrument (Applied Biosystems™, California, USA) was used with the following amplification conditions: initial denaturing at 95 ºC (10 min), 30 cycles of 95 ºC (10 s), hybridization (variable) (15 s) and elongation at 72 ºC (20 s). This was followed by a melting profile: formation of homo and heteroduplexes at 95 ºC (1 min), 40 ºC (1 min), and melting interval at 65–95 ºC with constant fluorescence acquisition. Each sample was analyzed in triplicate and a negative control without template DNA was included. Results were analyzed with HRM Software (Applied Biosystems™, California, USA) and curves were normalized. The region of the melting curve that gave the most information for discriminating between melting profiles of the PCR products was selected. The reproducibility of the melting patterns was evaluated for each sample to eliminate false positives that would affect the correct discrimination of different genotypes.

Results

Amplification of the selected loci

PCR products of the expected size (900 to 1200 bp) were obtained for the 22 selected loci in the 13 genes of the 19 accessions of *Coffea*. Specific and unique PCR amplification fragments were verified by agarose electrophoresis and by analyzing the absence of multiple peaks in the melting curves (data not shown).

The locus for a kinase/serine protein implicated in the response to biotic stress (*Pto*) (GenBank accession number DQ124059.1; Phytozome accession name evm.TU. Scaffold_2093.51) showed species-specific amplification. This product amplified in 84%

of the 19 accessions analyzed but not in Geischa, *C. liberica* or *C. canephora* samples. A similar situation occurred with the amplification of the loci *CaWRKY1_1* and *DXMT2* which amplified in 95%of the accessions but not in *C. eugenioides.* Similarly, the *LOX1* loci amplified in 89% of the accessions analyzed, but not in Geischa and Caturra (Supplementary Table 5).

Processing of PCR product sequencing data

The length of read (LOR) statistic was used to estimate sequencing quality and the informative potential of each processed sequence (Table 1). The maximum LOR value was 19 for five genes (*CaKO1, CaPOP1, ClpP1_2, ITS2,* and *SUS2*), when the 19 analyzed sequences maintained sufficient information to be used in further analyses. Other PCR products also showed high LOR values, but no sequence variations were observed among accessions (*ClpP1_1, Pto* and *CCD1_2*) (data not shown). In contrast, LOR values for PCR products for *CaKO1_2*, *CaPOP1_2*, *CaWRKY1_2, CCD1*, *CTgS1_2, ERF1,* and *LOX_2* were lower than 5, which demonstrated that the majority of these chromatograms were composed of low-quality sequences.

Loci located in the sequences *CaKO1_1*, *CaPOP1_1, CcoAOMT, ClpP1_1*, *ClpP1_2, CTgS1_1, DXTM1_2, ITS2, LOX1, Pto, SUS2,* and *SUS2_2* genes, showed low percentages of sequencing errors among the 19 accessions studied. The LOR values were sufficiently high for the sequences to be informative (Table 1). Other PCR products showed problems associated with the nature of the amplified region or possible duplications in the genome that were not further considered.

Identification and validation of variants

Possible variations were identified in more than 150 sites using the program novoSNP 3.0.1 (Weckx et al. 2005) (http://novosnp.bioinf.be/). Nevertheless, a total number of highly reliable variants of 103 sites were obtained after manually removing false positives (Fig. 1A).

The *CCD1, ClpP1, ERF1,* and *Pto* loci showed no variations, while the *CaKO1, CcoAOMT, ITS2* and *SUS2* genes were the regions with the highest number of polymorphic sites (Fig. 1A). A total of 88 SNPs and 15 INDELs were detected among the genes with variants with an *F*-Score > 15. Among the SNPs identified, there were more transitions (57) than transversions (31) (Fig. 1B). The majority of the polymorphisms, 69.6%, were identified in non-coding regions of the genes studied (Fig. 2). An important exception was observed in the gene *CaWRKY1*, in which the majority of the polymorphisms, 83.3%, were in the coding region (Fig. 2). The analysis of the first two coding regions of the *SUS2* gene showed that some of these polymorphisms resulted in a change to a biochemically distinct amino acid (Supplementary Fig. 1). These variations occurred mainly among Centroamericano (CAM) and Casiopea (CSP) with *C. canephora* ancestry introgression. The highest

Table 1 Informative potential of each locus based on total sequence length (LOR) and number of genotypes in legible chromatograms after processing with Sequencing Analysis software

PCR product	Size (bp)	Length of read (LOR)	Accessions with sequencing errors	Informative value
CaKO1_1	1105	19	0/19	+
CaKO1_2	1065	4	15/19	–
CaPOP1_1	1074	19	0/19	+
CaPOP1_2	1118	5	17/19	–
CaWRKY1_1	896	6	13/19	+
CaWRKY1_2	910	2	19/19	–
CCD1	1158	4	19/19	–
CCD1_2	1176	10	19/19	–
CcoAOMT	1139	18	1/19	+
ClpP1_1	1185	16	3/19	+
ClpP1_2	986	19	0/19	+
CTgS1_1	1198	16	3/19	+
CTgS1_2	957	5	14/19	–
DXMT1_1	949	10	9/19	+
DXMT1_2	1046	15	4/19	+
ERF1	803	2	17/19	–
ITS2	468	19	2 /19	+
LOX_1	1157	15	4/19	+
LOX_2	1120	5	14/19	–
Pto	1061	16	3/19	+
SUS2_1	951	19	0/19	+
SUS2_2	945	19	0/19	+

(+): informative; (−): non-informative

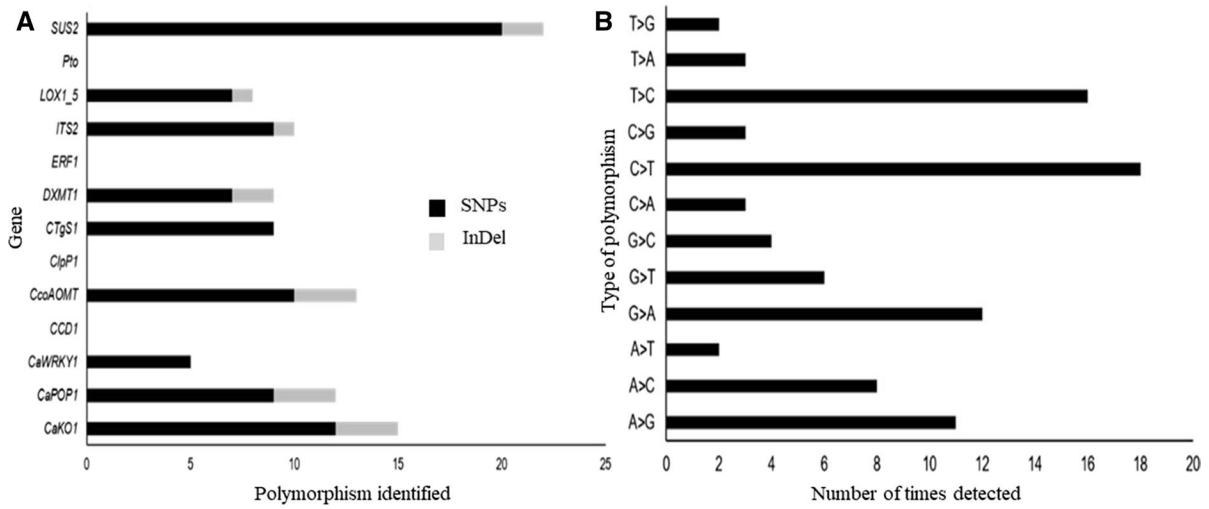

Fig. 1 Classification of the nucleotide variations identified in regions of 13 genes of the genome of 19 accessions of *Coffea* spp. **A** Types of polymorphisms validated between 13 genes of 19 accessions of *Coffea* spp. depending on the number of base pairs involved **B** Types of SNPs validated among 13 genes of 19 *Coffea* accessions

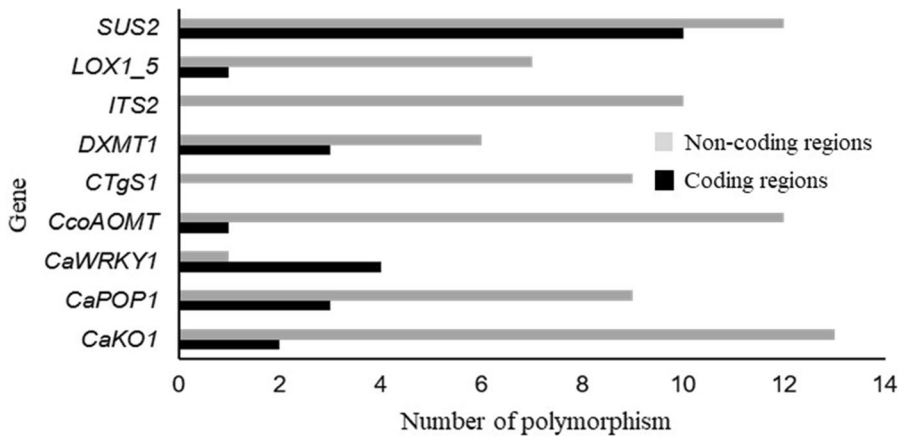

Fig. 2 Location of the variations in genetic sequences of the genome of 19 accessions of *Coffea* spp

number of variants were observed in non- *C. arabica* species (Fig. 3). We detected 61 polymorphic sites in *C. liberica* (CL), 58 in by *C. pseudozanguebariae* (PZB), 52 in *C. canephora* (CCTE), and 50 in *C. eugenioides* (EUG). A few polymorphisms (< 5 SNPs) were observed among *C. arabica* accessions. Among varieties of Arabica coffee, the hybrid genotypes Centroamericano, and Casiopea and the Sarchimor IAPAR 59 presented the highest number of polymorphic sites (Fig. 3).

HRM-PCR analysis of SNP identified in coffee accessions

Twenty inter and intraspecific variable sites defined by the primers were selected for HRM analysis (Supplementary Table 6). Of the 20 molecular markers, 16 identified at least one genotype by HRM analysis (Table 2). Real time PCR products of these genotypes had significantly different melting temperatures and

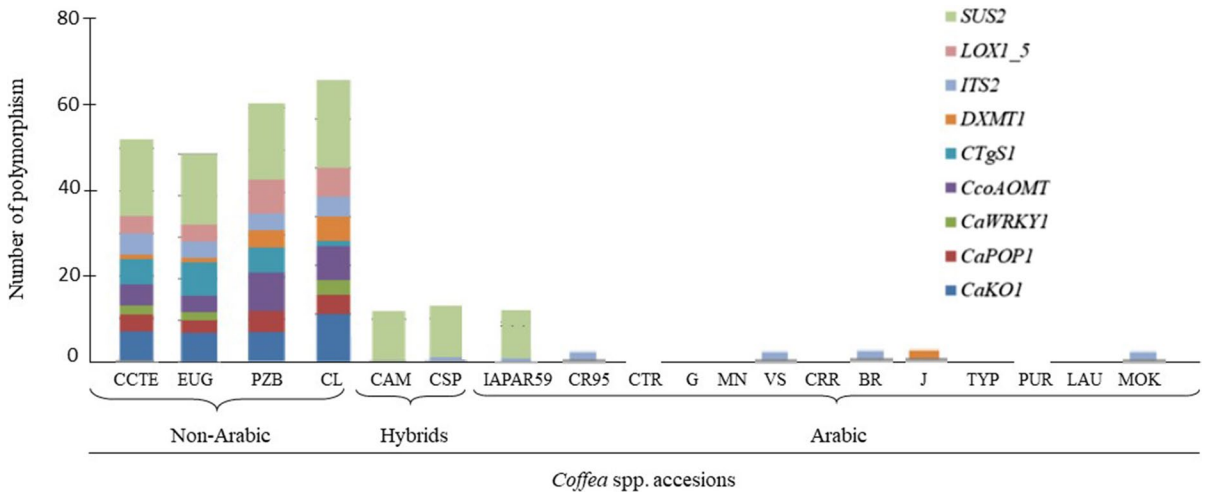

Fig. 3 Unique polymorphisms detected for each *Coffea* accession that allow their identification by PCR-HRM. *CAM* Centroamericano, *CL C. liberica*, IAPAR59 IAPAR59, *CR95* Costa Rica 95, *CTR* Catuaí Rojo, *G* Geisha, *MN* Mundo Novo, *VS* Villa Sarchí, *CRR* Caturra, *CCTE C. canephora*, *BR* Bourbon Red, *J* Java, *CSP* Casiopea, *EUG C. eugenioides*, *TYP* Typica, *PZB C. pzeudobangueriae*, *PUR* Purpurascens, *LAU* Laurina, *MOK* Mokka

could be distinguished from the others by their melting profiles (Supplementary Table 6).

Using HRM analysis, some of the selected regions were validated, and the expected melting profiles were produced for 16 of the 20 loci. The PCR products of CaWRKY1_HRM, SUS21_HRM_2, SNP_Ca032, and SNP_Ca073 could not be applied to distinguish genotypes because of non-specific amplification errors. For example, analysis of the PCR product *ITS2_HRM_1* clearly distinguished the two haplotypes (G/G and A/G) previously identified by sequencing. Accessions with a homozygous G/G haplotype generated a PCR product with a higher melting temperature, seen as a horizontal displacement of the melting curve during HRM analysis. The heterozygous G/A haplotype can be seen in IAPAR59, CR95, Villa Sarchí, Bourbon Red, and Mokka as two signals of equal intensity. This differs from the chromatograms of Centroamericano, Geisha and Mundo Novo, where the signal of a possible adenine is too low for a base call change (Fig. 4A). These results are consistent with the HRM analysis. Although they possess the same haplotype (G/G), *C. liberica,* and *C. pseudozanguebariae* generated PCR products with melting profiles that differed from those of the Arabica accessions (Fig. 4B). The genotypes for a variant identified in the *ITS* gene were corroborated by sequencing (Fig. 4C).

The varieties Centroamericano, IAPAR59, and Casiopea were differentiated from the other Arabica varieties due to the A/G haplotype at two loci in the *SUS22_HRM_1* locus. However, the form of the curves was similar, comparison among varieties with A/A and G/G haplotype showed that between the two loci, a total of two "A" alleles and two "G" alleles are present (Fig. 5A). HRM analysis of the PCR product SUS22_HRM2 showed that the varieties Centroamericano and Casiopea possess the C/T, A/G and G/G, while the variety IAPAR59 shares alleles with *C. canephora* haplotypes T/T, A/A and G/G. The other Arabica varieties possess the C/C, G/G and G/G haplotypes (Fig. 5A), and their PCR products had the highest melting temperature (Fig. 5B). The genotypes for a variant identified in the *SUS2* gene was corroborated by sequencing (Fig. 5C).

Finally, using this group of characterized SNPs from the representative sample of 19 coffee accessions, it was possible to establish the identity of an unknown sample with a certain grade of resolution. A two-step process of amplification and quantification of fluorescence of the melting curve can be used to discriminate among 10 genotypes of *Coffea* in approximately 8 h. In this sense, our results showed that HRM analysis were able to distinguish important cultivars, such as Geisha, Bourbon Red, and Typica (Fig. 6 and Supplementary Table 7). Specifically,

Table 2 Discriminative potential of markers analyzed by HRM in 19 genotypes of *Coffea* spp

PCR product	SNPs	Genotypes differentiated by HRM-PCR
CaKO_1_HRM_1	ins CA del ATAT	*C. arabica/C. canephora/C. eugenioides*
CaKO_1_HRM_2	T > C G > T/A A > G	*C. arabica/C. canephora/C. eugenioides*
CaKO_2_HRM_1	A > T/G	*C. arabica/C. liberica/C. canephora/C. pseudozanguebariae*
CaKO_2_HRM_2	C > T	*C. liberica-C. canephora/C. arabica*
CaPOP1_HRM	C > A G > T	*C. arabica* from others
CaWRKY1_HRM	A > T	None
CCoAOMT1_HRM	T > A/G A > G T > C	*C. eugenioides-C. pseudozanguebariae* from others
ITS2_HRM_1	G > A G/G	Genotypes A/G (IAPAR59, Costa Rica 95, Villa Sarchí, Bourbon Red and Mokka) from others Centroamericano, *C. liberica*, Geisha, Mundo Novo, *C. canephora; C. eugenioide*s, and *C. pseudozanguebariae* from others
ITS2_HRM_2	G > A C > T	*C. eugenioides* and *C. pseudozanguebariae* from others. Mundo Novo from other Arabicas
LOX1_HRM	C > T	Among *C. arabic*a (Geisha)
SUS21_HRM_1	A > C	*C. arabica/C. canephora/C. eugenioides*
SUS21_HRM_2	A > C A > C	Not informative, does not distinguish heterozygous genotypes
SUS22_HRM_1	G > A A > G	*C. arabica* hybrids from others
SUS22_HRM_2	C > T G > A G > A	Differentiates between Arabica hybrids and pure genotypes
SUS22_HRM_3	G > C	Centroamericano and Casiopea from other Arabicas
SNP_Ca012	C/A	Between *C. arabica* (Bourbon Red from Typica), *C. arabica/C. liberica*
SNP_Ca032	C/T	Error, non-specific amplification of larger than expected size
SNP_Ca073	C/T	Does not distinguish between different genotypes
SNP_Ca171	A/G	*C. arabica/C. canephora*
SNP_Ca346	A/G	*C. arabica-C. canephora/C. liberica*

the HRM analysis of the PCR product of CaKO2_HRM_1 distinguished the characteristic profile of 5 species (*C. liberica, C. canephora, C. eugenioides*, *C. arabica,* and *C. pseudozanguebariae*). The distinction of Arabica genotypes from other non-Arabica genotypes was achieved using the primer CaKO_HRM and ITS2_HRM_1. On the other hand, the LOX1_HRM primer indicated a specific profile of the Geisha variety. Using the SNP_Ca012 the two large genetic groups of Arabica (Bourbon and Typica) were distinguished. While the SNP_Ca171 distinguished *C. canephora* from the other four species (*C. arabica, C. liberica, C.pseudozanguebariae,* and *C. eugenioides*) (Fig. 6).

Discussion

Previously, HRM analysis has been used to distinguish between *C. arabica* L. and *C. canephora* genotypes in order to identify cases of adulteration of ground coffee blends (Combes et al. 2018). These

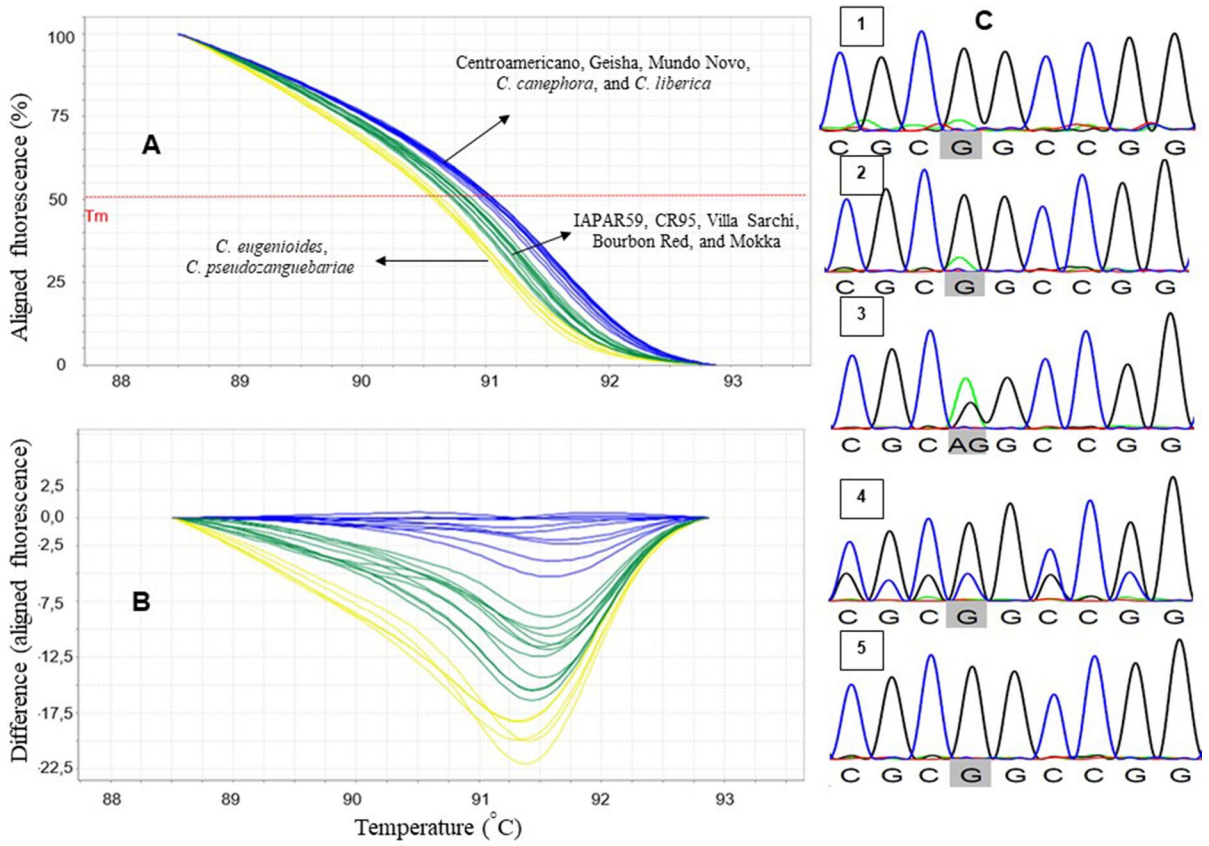

Fig. 4 HRM analysis of a G/A nucleotide variation in the ITS2 sequence of the genome of different coffee accessions. **A** melting profiles of the PCR products defined by the ITS2_HRM_1 primers. Yellow: haplotype A/G of *C. eugenioides*, *C. pseudozanguebariae*. Green: haplotype A/G of *C. arabica* var: IAPAR59, CR95, Villa Sarchí, Bourbon Red and Moka. Blue: haplotype G/G of *C. arabica* var: Centroamericano, Geisha, Mundo Novo, *C. canephora*, *C. liberica*. **B** relative differences of the fusion profiles. **C** haplotype for a variant identified in the *ITS2* gene of nineteen accessions of *Coffea spp*. 1: reference sequence reported in GenBank (EU650386.1); 2: G/G haplotype; 3: A/G (IAPAR59, Bourbon Red, Moka, CR95, Villa Sarchí) haplotype; 4: G/G (*C. canephora*) haplotype; 5: G/G (other Arabicas) haplotype

authors analyzed polymorphic regions of chloroplast DNA of both species and the melting profiles were associated with blend percentages (Combes et al. 2018). In our study, nucleotide variability in distinct genomic regions of *Coffea* spp. were detected by by HRM analysis and sites that differed among closely related accessions of Arabic coffee and also among more genetically distant non-Arabic accessions were revealed. In contrast to Combes et al. 2018, in our study, nuclear genetic variation was detected and revealed the existence of SNPs between Arabica and Robusta nuclear genomes. This demonstrated that HRM is a suitable method for the identification and authentication of *Coffea* species, as suggested by Combes et al. 2018. In addition, HRM analysis has been successfully used to detect polymorphisms and differentiate between genotypes in *Prunus dulcis* (Wu et al. 2009), *Capsicum* (Jeong et al. 2010), *Glycine max* (Cruz et al. 2013), carnaroli rice (Grazina et al. 2022), *Dendrobium* (Chen et al. 2022), *Calanthe* (Buddhachat et al. 2022). HRM analysis is an efficient and potentially scalable technique that can be adapted satisfactorily to genotyping strategies in large populations (Simko et al. 2016).

In the present study, the majority of the polymorphisms were observed in non-coding regions. The locus *ITS2*, considered one of the most informative genomic regions for discriminating at the species level, showed important results within coffee species and Arabica varieties. The *ITS2* region has a reported

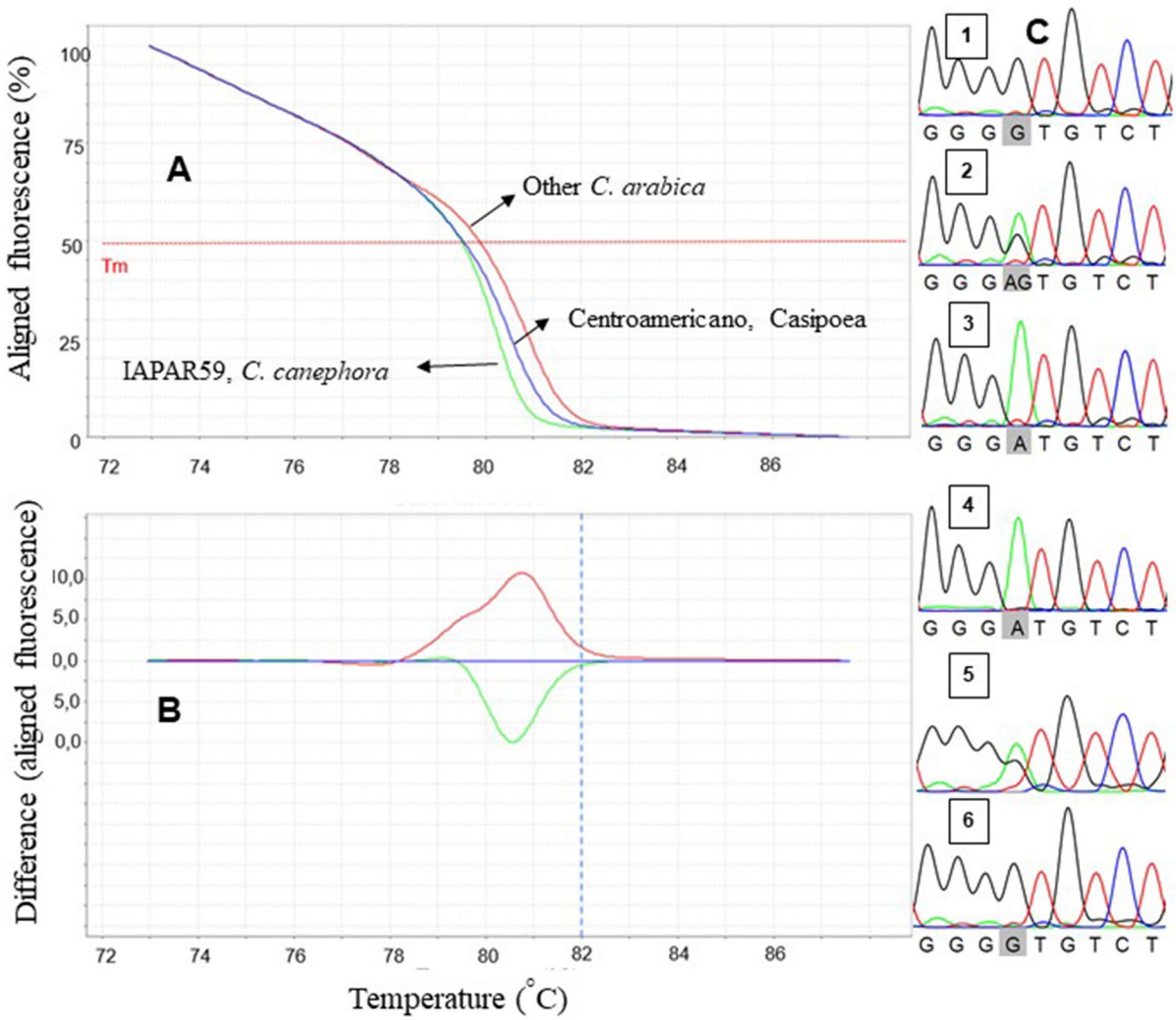

Fig. 5 HRM analysis of a G/A nucleotide variation in the *SUS2* genome sequence of different coffee accessions. **A** melting profiles of the PCR products defined by the primers SUS22_HRM2. Green: *C. arabica* var. IAPAR59, *C. canephora*. Blue: *C. arabica* var. Centroamericano, Casiopea. Red: *C. arabica* (other varieties). **B** relative differences of the fusion profiles. **C** genotypes for a variant identified in the *SUS2* gene of nineteen accessions of *Coffea spp*. 1: reference sequence reported in Phytozome (evm.TU.Scaffold_2016.44); 2: haplotype A/G (Centroamericano); 3: haplotype A/A (IAPAR59); 4: haplotype A/A (*C. canephora*); 5: haplotype A/G (Casiopea); 6: G/G haplotype (other Arabicas)

success rate of 76.1% for discriminating among dicotyledonous species (Yao et al. 2010). Analysis of this region in *Coffea* revealed high intraspecific variability that allowed the genus to be separated into large groups (Lashermes et al. 1997). However, analysis of this region within *C. arabica* had not been explored. Recent studies using SSR (Sánchez et al. 2020) and SNP-GWAS (Scalabrin et al. 2020) methods found very low genetic variability among Arabica coffee cultivars. Our study demonstrated that this locus is very interesting, since it allows the discovery of variations for different cultivars of Arabica coffee.

In polyploids, such as coffee, SNP and point variations analyses can be complicated by covering effects related to the analysis of homologous regions in the sub genomes. *C. arabica* L., a natural allotetraploid, combines two subgenomes (*C. eugenioides* and *C. canephora*) with common homeologuos chromosome exchanges causing genome rearrangements (Combes et al. 2018). Nevertheless, in our study, intra species

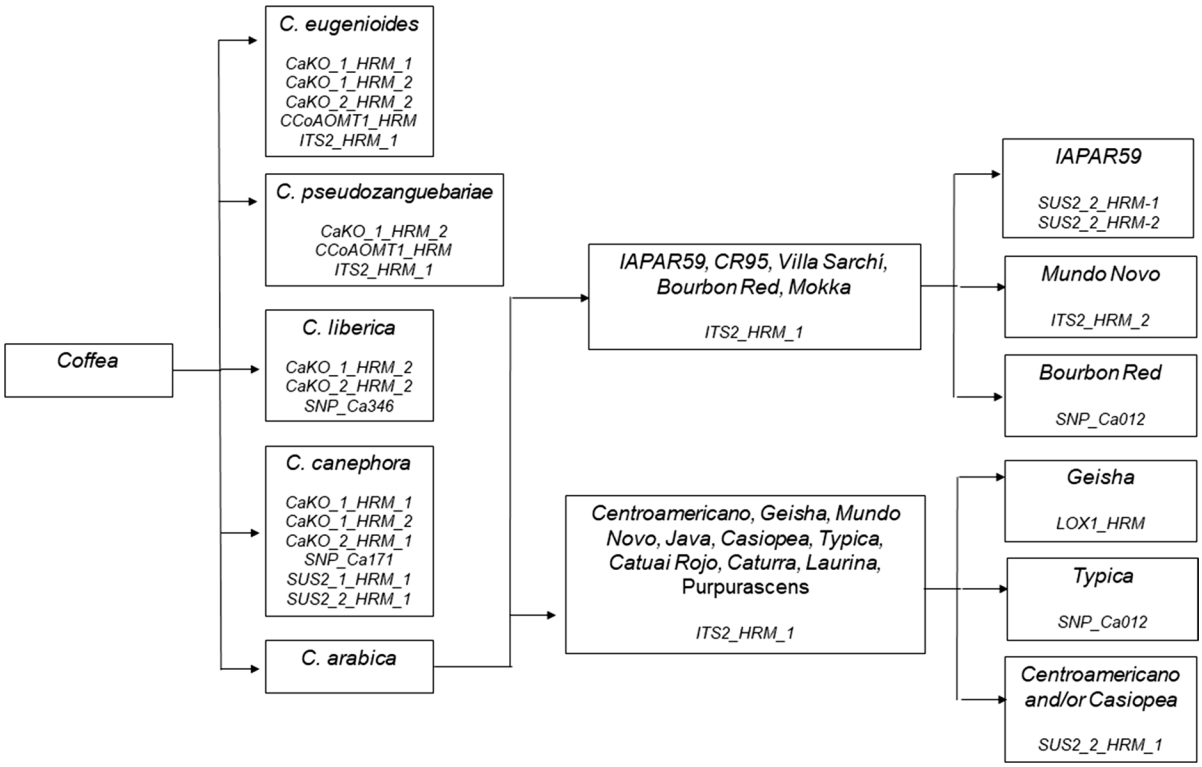

Fig. 6 Discriminative potential of HRM analysis applied to SNP markers identified in 19 *Coffea* spp. genotypes. Primers used to discriminate among *Coffea* spp. genotypes are indicated in each box

SNPs was detected and targeted even the low genetic diversity of *C. arabica* L. On the other hand, HRM analysis was used to separate tetraploid alfalfa plants (*Medicago sativa*) into two groups (Han et al. 2012). Moreover, HRM analysis generated two distinct profiles for tetraploid potato genotypes (*Solanum tuberosum*) that were resistant or susceptible to potato virus Y (PVY) (Nie et al. 2018). In addition, the tetraploid genome of peanut (*Arachis hypogaea* L.) was analyzed in detail to verify 56 SNPs in 12 varieties and two diploid progenitors; this analysis led to the selection of a panel of 33 informative SNP for lines in genetic breeding programs (Hong et al. 2015). HRM analysis has also been used to identify hybrid rice cultivars and separate them from parent lines according to melting profiles (Han et al. 2012; Zhu et al. 2013). Similar results were seen in this study using variants found exclusively in coffee varieties obtained from introgressions. This type of molecular marker allows an efficient management and traceability of inherited loci of interest.

Because of the genetic homogeneity of coffee accessions, preservation and detailed characterization of germplasm is crucial for genetic improvement and conservation. A recent study has demonstrated that all *C. arabica* L. species were derived from a single polyploidization event, which explains the low level of genetic variability (Scalabrin et al. 2020). Arabica coffee represents a species that is particularly vulnerable to climate change and phytosanitary problems (van der Vossen et al. 2015). The use of modern strategies for genetic improvement is essential in order to increase genetic variability and genetic resources for the coffee sector in the future. New strategies have been described to increase genetic variability in coffee populations through the induction of mutations (Bolívar-González et al. 2018; Vargas-Segura et al 2019). Techniques such as HRM analysis can be used to characterize established varieties and to support introgression or hybridization programs by screening for molecular markers in progeny plants.

Conclusions

This study demonstrated the genotypic identification of *Coffea* accessions using HRM analysis in genomic regions of interest. This methodology could be applied to several genes of interest without the need of a prior sequencing. The HMR technique for identification of variants in genes *CaKO1*, *SUS2*, *CcOAOMT,* as well as in the *ITS2* region proved to be a robust technique for germplasm characterization. More important this technique can be used for fingerprinting and traceability of coffee grain exports which is an increasing market-consumer demand.

Acknowledgements The authors would like to thank the "Centro Nacional de Alta Tecnología-Consejo Nacional de Rectores" (CeNAT-CONARE) and "Sistema de Estudios de Posgrado de la Universidad de Costa Rica" (SEP-UCR) for the grant awarded to A.B-G. Moreover, we thank the anonymous reviewers for their helpful comments on this manuscript.

Authors' contributions A.B-G. designed and performed the experiments, analyzed data, and wrote the draft manuscript; A.G-A conceived the project, designed and coordinated the experiments, analyzed data, and edited the final manuscript; E.A-V. provided fundamental feedback of experiments execution, discussed the results, and edited the final manuscript; R.M.-B. provided fundamental feedback of experiments execution, discussed the results, and edited the final manuscript; W.S–S provided the plant material, and edited the final manuscript; S.T.I-S and L.F.P.P revised and edited the final manuscript.

Funding This research was financed by the University of Costa Rica and the "Ministerio de Ciencia, Tecnología y Telecomunicaciones" (MICITT) (project No. 111-B5-140).

Availability of data and material The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

Alves G, Torres L, de Aquino S, Reichel T, Freire L, Vieira N, Vinecky F, This D, Pot D, Etienne H, Paiva L, Marraccini P, Andrade A (2018) Nucleotide diversity of the coding and promoter regions of *DREB1D* a candidate gene for drought tolerance in *Coffea* species. Trop Plant Biol 11:3148

Anthony F, Bertrand B, Quiros O (2001) Genetic diversity of wild coffee (*Coffea arabica* L) using molecular markers. Euphytica 118:53–65

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. Theor Appl Genet 104:894–900

Ashihara H, Sano H, Crozier A (2008) Caffeine and related purine alkaloids: biosynthesis catabolism function and genetic engineering. Phytochemistry 69:841–856

Bertrand B, Guyot B, Anthony F, Lashermes P (2003) Impact of the *Coffea canephora* gene introgression on beverage quality of *C arabica*. Theor Appl Genet 107:387–394

Bolívar-González A, Valdez-Melara M, Gatica-Arias A (2018) Responses of Arabica coffee (*Coffea arabica* L var Catuaí) cell suspensions to chemically induced mutagenesis and salinity stress under in vitro culture conditions. In Vitro Cell Dev Biol Plant 54:576–589

Buddhachat K, Sripairoj N, Punjansing T, Kongbangkerd A, Inthima P, Tanming W, Kosavititkul P (2022) Species discrimination and hybrid detection in terrestrial orchids using Bar-HRM: a case of the Calanthe group. Plant Gene 29:100349. https://doi.org/10.1016/j.plgene.2021.100349

Chen W, Chen X, Xu J et al (2022) Identification of *Dendrobium officinale* using DNA barcoding method combined with HRM and qPCR technology. Food Anal Methods. https://doi.org/10.1007/s12161-021-02194-y

Cheng B, Furtado A, Smyth HE, Henry RJ (2016) Influence of genotype and environment on coffee quality. Trends Food Sci Technol 57:20–30

Clevenger J, Chavarro C, Pearl SA, Ozias-Akins P, Jackson SA (2015) Single nucleotide polymorphism identification in polyploids: a review example and recommendations. Mol Plant 8:831–846

Combes MC, Joët T, Lashermes P (2018) Development of a rapid and efficient DNA-based method to detect and quantify adulterations in coffee (Arabica versus Robusta). Food Control 88:198–206

Cruz MFA, Bueno RD, Souza FB, Moreira MA, Barros EG (2013) Identificação de SNPs para conteúdo de ácidos graxos em soja pela técnica HRM. Pesqui Agropecu Bras 48:1596–1600

Cubry P, Musoli P, Legnate H (2008) Diversity in coffee assessed with SSR markers: structure of the genus *Coffea* and perspectives for breeding. Genome 51:50–63

Dias-Guzzo S, Harakava R, Mui-Tsai S (2009) Identification of coffee genes expressed during systemic acquired resistance and incompatible interaction with *Hemileia vastatrix*. J Phytopathol 157:625–638

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Ferreira T, Farah A, Oliveira T, Lima I, Vitório F, Oliveira E (2016) Using real-time PCR as a tool for monitoring the authenticity of commercial coffees. Food Chem 199:433–438

Ganal MW, Altmann T, Röder MS (2009) SNP identification in crop plants. Curr Opin Plant Biol 12:211–217

Grazina L, Costa J, Amaral JS, Garino C, Arlorio M, Mafra I (2022) Authentication of carnaroli rice by HRM analysis targeting nucleotide polymorphisms in the Alk and

Waxy genes. Food Control 135:108829. https://doi.org/10.1016/j.foodcont.2022.108829
Han Y, Khu DM, Monteros MJ (2012) High-resolution melting analysis for SNP genotyping and mapping in tetraploid alfalfa (*Medicago sativa* L). Mol Breed 29:489–501
Hong Y, Pandey MK, Liu Y, Chen X, Liu H, Varshney RK et al (2015) Identification and evaluation of single-nucleotide polymorphisms in allotetraploid peanut (*Arachis hypogaea* L) based on amplicon sequencing combined with high resolution melting (HRM) analysis. Front Plant Sci 6:1068
Huq A, Akter S, Nou IS, Kim HT, Jung YJ, Kang KK (2016) Identification of functional SNPs in genes and their effects on plant phenotypes. J Plant Biotech 43:1–11
International coffee organization (ICO) Monthly coffee market report (2021/22)
Jeong HJ, Jo YD, Park SW, Kang BC (2010) Identification of *Capsicum* species using SNP markers based on high resolution melting analysis. Genome 53:1029–1040
Jingade P, Huded A, Kosaraju B, Mishra M (2019) Diversity genotyping of Indian coffee (*Coffea arabica* L) germplasm accessions by using SRAP markers. J Crop Improv 33(327):345
Lashermes P, Combes M, Trouslot P, Charrier A (1997) Phylogenetic relationships of coffee-tree species (*Coffea* L) as inferred from ITS sequences of nuclear ribosomal DNA. Theor Appl Genet 94:947–955
Lashermes P, Andrzejewski S, Bertrand B (2000) Molecular analysis of introgressive breeding in coffee (*Coffea arabica* L). Theor Appl Genet 100:139–146
Lepelley M, Cheminade G, Tremillon N, Simkin A, Caillet V, McCarthy J (2007) Chlorogenic acid synthesis in coffee: An analysis of CGA content and real-time RT-PCR expression of *HCT HQT C3H1* and *CCoAOMT1* genes during grain development in *C canephora*. Plant Sci 172:978–996
Li J, Wei YL, Li YY, Pang L, Deng WW, Jiang CJ (2014) A correlation study of caffeine content with theobromine content cSNPs and transcriptional expression of three genes in tea plants. Crop Sci 54:1124–1132
Merot-L'anthoene V, Tournebize R, Darracq O, Rattina V, Lepelley M, Bellanger L et al (2019) Development and evaluation of a genome-wide *Coffee* 85 K SNP array and its application for high-density genetic mapping and for investigating the origin of *Coffea arabica* L. Plant Biotechnol J 17:1418–1430
Mizuno K, Matsuzaki M, Kanazawa S, Tokiwano T, Yoshizawa Y, Kato M (2014) Conversion of nicotinic acid to trigonelline is catalyzed by N-methyltransferase belonged to motif B' methyltransferase family in *Coffea arabica*. Biochem Biophys Res Commun 452:1060–1066
Motta LB, Soares TCB, Ferrao MAG (2014) Molecular characterization of arabica and conilon coffee plants genotypes by SSR and ISSR markers. Braz Arch Biol Technol 57:728–735
Nie X, Dickison VL, Brooks S, Nie B, Singh M, De Koeyer DL, Murphy AM (2018) High resolution DNA melting assays for detection of Rx1 and Rx2 for high-throughput marker-assisted selection for extreme resistance to potato virus X in tetraploid potato. Plant Dis 102:382–390
Pereira LF, Ivamoto S (2015) Characterization of coffee genes involved in isoprenoid and diterpene metabolic pathways. In: Preedy V (ed) Coffee in health and disease prevention. Academic Press, London, pp 45–51
Petitot A, Barsalobres-Cavallari C, Ramiro D, Albuqerque F, Etienne H, Fernandez D (2013) Promoter analysis of the WRKY transcription factors *CaWRKY1a* and *CaWRKY1b* homoeologous genes in coffee (*Coffea arabica*). Plant Cell Rep 32:1263–1276
Privat I, Foucrier S, Prins A, Epalle T, Eychenne M, Kandalaft L, Caillet V, Lin C, Tanksley S, Foyer C, McCarthy J (2008) Differential regulation of grain sucrose accumulation and metabolism in *Coffea arabica* (Arabica) and *Coffea canephora* (Robusta) revealed through gene expression and enzyme activity analysis. New Phytol 178:781–797
Romero G, Alvarado G, Cortina H, Ligarreto G, Galeano N (2010) Partial resistance to leaf rust (*Hemileia vastatrix*) in coffee (*Coffea arabica* L): genetic analysis and molecular characterization of putative candidate genes. Mol Breed 25:685–697
Ságio S, Barreto H, Lima A, Moreira R, Rezende P, Paiva L, Chalfun-Junior A (2014) Identification and expression analysis of ethylene biosynthesis and signaling genes provides insights into the early and late coffee cultivars ripening pathway. Planta 239:951–963
Sánchez E, Solano W, Gatica-Arias A, Chavarría M, Araya-Valverde E (2020) Microsatellite DNA fingerprinting of *Coffea* sp germplasm conserved in Costa Rica through singleplex and multiplex PCR. Crop Breed Appl Biot 20:e27812013
Santa Lucia J (1998) A Unified view of polymer dumbbell and oligonucleotide DNA nearest-neighbor thermodynamics. Proc Natl Acad Sci 95:1460–1465
Scalabrin S, Toniutti L, Di Gaspero G, Scaglione D, Magris G, Vidotto M et al (2020) A single polyploidization event at the origin of the tetraploid genome of *Coffea arabica* is responsible for the extremely low genetic variation in wild and cultivated germplasm. Sci Rep 10:1–13
Simkin A, Kuntz M, Moreau H, McCarthy J (2010) Carotenoid profiling and the expression of carotenoid biosynthetic genes in developing coffee grain. Plant Physiol Biochem 48:434–442
Simko I (2016) High-resolution DNA melting analysis in plant research. Trends Plant Sci 21:528–537
Singh R, Irikura B, Nagai C, Albert H, Kumagi M, Paull R, Moore P, Wang M (2011) Characterization of prolyl oligopeptidase genes differentially expressed between two cultivars of *Coffea arabica* L. Trop Plant Biol 4:203–216
Sousa TV, Caixeta ET, Alkimim ER, de Oliveira ACB, Pereira AA, Zambolim L, Sakiyama NS (2017) Molecular markers useful to discriminate *Coffea arabica* cultivars with high genetic similarity. Euphytica 213:75
Szurman-Zubrzycka M, Chmielewska B, Gajewska P, Szarejko I (2017) Mutation detection by analysis of DNA heteroduplexes in TILLING populations of diploid species. In: Jankowicz-Cieslak J, Tai T, Kumlehn J, Till, BJ (eds) Biotechnologies for plant mutation breeding, (Springer International Publishing, Cham), https://doi.org/10.1007/978-3-319-45021-6_18

Tran HTM, Furtado A, Vargas CAC, Smyth H, Slade-Lee L, Henry R (2018) SNP in the *Coffea arabica* genome associated with coffee quality. Tree Genet Genomes 14:72

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3—new capabilities and interfaces. Nucleic Acids Res 40:e115

Vargas-Segura C, López-Gamboa E, Araya-Valverde E, Valdez-Melara M, Gatica-Arias A (2019) Sensitivity of Seeds to chemical mutagens detection of DNA polymorphisms and agro-metrical traits in M1 generation of coffee (*Coffea arabica* L). J Crop Sci Biotechnol 22:451–464

van der Vossen H, Bertrand B, Charrier A (2015) Next generation variety development for sustainable production of Arabica coffee (*Coffea arabica* L): a review. Euphytica 204(2):243–256

Weckx S, Del-Favero J, Rademakers R, Claes L, Cruts M, De Jonghe P, Van Broeckhoven C, De Rijk P (2005) novoSNP a novel computational tool for sequence variation discovery. Genome Res 15:436–442

World coffee research (2018). Las variedades del café arábica. Un catálogo global de variedades que abarca: Costa Rica, El Salvador, Guatemala, Honduras, Jamaica, Kenia, Malawi, Nicaragua, Panamá, Perú, República Dominicana, Rwanda, Uganda, Zambia, Zimbabue. Available at https://varieties.worldcoffeeresearch.org/es/varieties Access on 5th Apr 2022

Wu SB, Tavassolian I, Rabiei G, Hunt P, Wirthensohn M, Gibson JP, Ford CM, Sedgley M (2009) Mapping SNP-anchored genes using high-resolution melting analysis in almond. Mol Genet Genomics 282:273–281

Yao H, Song J, Liu C, Luo K, Han J, Li Y, Pang X, Xu H, Zhu Y, Xiao P, Chen S (2010) Use of ITS2 region as the universal DNA barcode for plants and animals. PLoS ONE 5:e13102

Zhang D, Vega F, Infante F, Solano W, Johnson E, Meinhardt L (2020) Accurate differentiation of green beans of Arabica and Robusta coffee using nanofluidic array of single nucleotide polymorphism (SNP) markers. J AOAC Int 103:315–324

Zhang D, Vega FE, Solano W, Su F (2021) Infante F Meinhardt LW Selecting a core set of nuclear SNP markers for molecular characterization of Arabica coffee (*Coffea arabica* L) genetic resources. Conserv Genet Resour 13:329–335

Zhou L, Vega F, Tan H, Lluch A, Fang ML, Mischke W, Irish S, Zhang B, D, (2016) Developing single nucleotide polymorphism (SNP) markers for the identification of coffee germplasm. Trop Plant Biol 9:82–95

Zhu Y, Wang Q, Hu J, Zhu L, Wang J, Zhu S, Guan Y (2013) High resolution melting analysis; an efficient method for fingerprinting of hybrid rice cultivars and their parental lines. Aust J Crop Sci 7:2048–2053