CrossMark

Assisted-selection of naturally caffeine-free coffee cultivars—characterization of SNPs from a methyltransferase gene

Patrícia Favoretto · Carla Cristina da Silva · Aline Gomes Tavares · Gabriela Giatti · Patrícia Favoretto Moraes · Mary Tulia Vargas Lobato · Maria Bernadete Silvarolla · Guerreiro Oliveiro-Filho · Mirian Perez Maluf

Abstract Breeding of caffeine-free coffee cultivars require tools for an early selection of progenies bearing this later trait. Genes from caffeine synthesis and degradation represent major targets for the development of molecular markers for assisted selection. In this study, we characterized SNPs identified on the caffeine synthase gene from AC1 mutant, a naturally caffeine-free arabica coffee plant. Molecular analysis of normal and mutant sequences indicates the occurrence of SNPs in protein domains, potentially associated with caffeine synthesis in coffee. Progenies F_2, F_1BC_1 and BC from crosses of AC mutants and elite cultivars were evaluated regarding caffeine content in grains and genomic segregation profile of selected SNPs. Genotyping analysis allowed the discrimination between homozygous and heterozygous plants. Quantification of caffeine content indicated a significant variability among progenies and a low frequency of caffeine-free plants. Statistical analyses of genotyping and phenotyping results showed significant association between presence of selected SNPs and reduced caffeine content. Moreover, this association occurs through all evaluated genetic backgrounds and generations, indicating an inheritance stability of both trait and markers. The molecular markers described here represent a successful case of assisted-selection in coffee, indicating their potential use for breeding of caffeine-free cultivars.

Keywords Assisted-selection · Caffeine content · Caffeine synthase · Coffee · Molecular markers

P. Favoretto · A. G. Tavares · G. Giatti · P. F. Moraes · M. T. V. Lobato · M. B. Silvarolla · G. Oliveiro-Filho · M. P. Maluf (✉)
EMBRAPA and Coffee Center Agronomic Institute/ IAC, Campinas, Brazil
e-mail: mirian.maluf@embrapa.br

C. C. da Silva
Molecular Biology and Genetic Engineering Center/UNICAMP, Campinas, Brazil

Introduction

Caffeine is a widely known and investigated compound from coffee grains, due to its effects on both human health and on coffee plants physiology. As some of those effects upon human health are not desirable, breeding programs worldwide pursue the development of naturally caffeine-free cultivars, while retaining other sensorial and quality attributes.

Caffeine content, on most of *Coffea arabica* cultivars is around 1% (percentage of dry weigh), and *C. canephora* cultivars have around 2%. One known variety of *C. arabica* with low caffeine content is *laurina*, with around 0.6% of caffeine in fruits (Tango and Carvalho, 1963). Other *Coffea* species, however, exhibited a lower caffeine content in grains, such as *C. dewevrei* (1.0%), *C. eugenioides* (0.4%), *C. salvatrix* (0.7%), and *C. racemosa* (0.8%) (Mazzafera and Carvalho 1992). In addition, species such as *C. pseudozangebariae* and

C. richardii have no caffeine in grains (Campa et al. 2005). Studies demonstrated that absolute content of caffeine results from a balance between synthesis and degradation. Activity quantification of biosynthetic enzyme using radioactive compounds characterized caffeine content of several *C. arabica* varieties, including the mutant *laurina* (Mazzafera et al. 1994). The authors concluded that the rates of caffeine synthesis and degradation in grains change during fruit maturation and among cultivars. In other similar study, Ashihara and Crozier (1999) demonstrated that low caffeine levels in *C. eugenioides* result from a slow synthesis associated with a rapid degradation.

One breeding strategy to develop low caffeine coffee cultivars is to transfer this trait from *Coffea* species to *C. arabica*. However, this strategy was unsuccessful due to limitations on inter-specific crosses (Charrier 1978; Mazzafera and Carvalho 1992). In another strategy, Silvarolla et al. (2000) focused on the natural variability of caffeine content in *C. arabica* wild accessions collected at Ethiopia. Caffeine quantification on grains revealed a variability directly associated with geographical origin (Silvarolla et al. 2000). Later, Silvarolla et al. (2004) identified three plants (namely AC1, AC2 and AC3) with an average content of 0.076%, which opened the perspective for a naturally caffeine-free coffee cultivar. Despite the lack of caffeine, grains from AC1 have a similar chemical composition and overall cup quality as those of regular coffee (Benatti et al. 2012, Silvarolla, unpublished results). In this context, crosses between accession AC1 and cultivar Mundo Novo (MN) initiated a breeding program aiming to transfer the low caffeine trait to elite cultivars.

Molecular analyses of AC1 revealed several aspects that affect the transfer of low caffeine trait. Benatti et al. (2012) verified that AC1 fruits are smaller than fruits from MN, and have reduced activity of theobromine synthase and caffeine synthase. Expression analyses indicated that in AC1 fruits exhibited a lower accumulation of transcripts from genes encoding those methyltransferases compared to normal MN fruits (Maluf et al. 2009). Besides this, the genomic sequence of caffeine synthase gene from AC1 revealed several nucleotide polymorphisms, including a base substitution that results in a replacement of a valine for an isoleucine at the enzyme active site (Maluf et al. 2009). Thus, the low caffeine content of AC1 plants results from a combination of a reduced gene expression and a possible knockout mutation of caffeine synthase.

The selection of naturally caffeine-free plants in segregating populations is an arduous task. High selection cost, limited effectiveness of analytical methods for caffeine content and specially the long juvenile period associated with a poor correlation between levels detected in leaves and grains are major challenges for breeders.

Molecular markers have been used with success for assisted selection in several plant species, and then could aid to overcome difficulties on selection of low caffeine coffee progenies. Therefore, we tested several polymorphisms observed on mutant *caffeine synthase* gene for their potential use as molecular markers in the coffee breeding program. Among those, we validated two neighbor SNPs in progenies F_2 and F_1BC_1, from crosses involving AC mutants and elite *C. arabica* cultivars. The validation comprised SNPs genotyping and phenotyping regarding caffeine content of individual plants. These analyses indicated a high correlation between genotype and phenotype, and demonstrated the potential of those SNP markers to select caffeine-free plants.

Material and methods

Plant material

Coffee plants have different origin and belong to the IAC *Coffea* Germplasm Collection, Campinas, Brazil (Table 1). The *C. arabica* accesions AC1, AC2, and AC3 have reduced levels of caffeine. Elite *C. arabica* cultivars Mundo Novo (MN IAC 376–4), Catuaí Vermelho (CV 81 and CV 144), IAC Ouro Verde (OV), Bourbom Vermelho (BV IAC), and Obatã (OB IAC 1669–20) have regular caffeine content. Selected progenies from reciprocal crosses between ACs and elite cultivars, F_2 from both open pollination and self-pollination, and progenies from diverse F_2 and back-crosses have variable caffeine content. Table 1 shows detailed information regarding plant identity, origin, and caffeine content. Young mature leaves of each plant were collected and maintained at -80 °C for DNA extraction.

In silico analyses

For identification of single-nucleotide polymorphisms (SNPs) GeneBank sequences of *caffeine synthase CCS1* (AB086414) and *CaDMXT1* (AB084125) were aligned

Table 1 Origin of evaluated germplasm with corresponding caffeine content in grains

Genotypes	Origin	Description	Caffeine
AC$_1$	Turrialba introduction	Acession	Traces
AC$_2$	Turrialba introduction	Acession	Very low to regular
AC$_3$	Turrialba introduction	Acession	Low to regular
C. arabica Mundo Novo IAC 376–4	C. arabica Sumatra X C. arabica Bourbon Vermelho	Cultivar	Regular
C.arabica Catuaí Vermelho IAC 81	C. arabica Caturra Amarelo X C. arabica Mundo Novo	Cultivar	Regular
C. arabica Catuaí Vermelho IAC 144	C. arabica Caturra Amarelo X C. arabica Mundo Novo	Cultivar	Regular
C. arabica Ouro Verde IAC H 5010–5	C.arabica Catuaí Amarelo IACH2077–2–12-70 X C. arabica Mundo Novo IAC 515–20	Cultivar	Regular
C. arabica Obatã Vermelho IAC 1669–20	C. arabica Villa Sarchi (CIFC H 361) X Híbrido de Timor (CIFC 832/2)	Cultivar	Regular
Bourbon Vermelho	Bourbon introduction	Cultivar	Regular

with genomic sequences from *C. arabica* cv Mundo Novo (MN) and AC1 using the software BioEdit (Maluf et al. 2009). Genomic sequences were amplified by PCR using gene specific primers designed upon available caffeine synthase sequences at the time. Two different sequences were identified: one present only on AC1 and the other on both MN and AC1 genomes (Online resource 1). The polymorphisms among those two sequences were in silico characterized regarding their putative position on protein sequence, relevance of amino acid changes on protein sequence, and relative position in exons. Based on these analyses three SNPs were selected for genotyping score. Primers and MGB-probes were designed using *software PRIMER Express 3.0 (Applied Biosystems)*. All sequences shown here are under legal protection for intellectual property issues and their commercial use is restricted (INPI BR1020130323179).

Genotyping by MGB TaqMan

Genomic DNA was extracted from collected leaves using a CTAB based protocol (Doyle and Doyle 1990). After extraction, DNA was quantified by absorbance under UV light at 260 and 280 nm on a spectrophotometer *PharmaSpec* UV-1700 (UV-*Visible Spectrophotometer* SHIMADZU), according to Sambrook et al. (1989). A conventional PCR was performed to verify overall DNA quality using control actin specific primers as previously described (Maluf et al. 2009).

Genotyping was performed by quantitative PCR using a combination of primers and specific MGB TaqMan probes (Life Technologies). *Caffeine synthase* specific primers and fluorescent probes annealing to a region with two SNPs were used for allelic discrimination using an ABI7300 platform (Applied Biosystems). Sequences and fluorescence used are listed in Online resource 2. For each reaction 100 ng of DNA were used, and amplification conditions were as suggested on the commercial TaqMan® Genotyping Master Mix (Thermo Scientific) and equipment manuals. All reactions were performed in duplicate, and were repeated two times for each genotype. Post-running analyses were performed using the *software 7300 System* SDS (Applied Biosystems).

Caffeine content quantification

Around 40 mature fruits from individual plants were collected and dried until 12% humidity. Afterward, fruits were peeled, grounded, and transferred to extraction tubes. Caffeine was extracted from duplicated samples by adding metanol 70% and incubating at 60 °C for 30 min (Clifford et al. 2003). After cool down, the solutions of Carrez A and B were added to tubes (Ky et al. 1999). Samples were vortexed, centrifuged at 16,000 g, diluted in ultrapure water, filtrated and loaded on an HPLC chromatographer (Shimadzu), with a reversed-phase C18 column and UV detector (272 nm), according to Casal et al. (2000). Concentration was estimated using calibration curves

obtained from known standard caffeine solutions measurements. Extraction and quantification were repeated for all duplicates exhibiting standard deviation higher than 10%.

Statistical analyses and evaluated parameters

We developed a statistical strategy to check the reliability of both methods to determine the caffeine content on progenies under selection. The parameters were:

1 Method accuracy - EM = (a + d)/n: result from the sum of plants classified as low (a) and high (d) caffeine by both genotyping and phenotyping analyses, and divided by the total number of plants.
2 False positive rate - TFP = b/(b + d): result from the division of number of plants genotyped as low caffeine and phenotyped as high caffeine (b) by the total number of plants phenotyped as high caffeine (b + d).
3 False negative rate - TFN = c/(a + c): result from the division of number of plants genotyped as high caffeine and phenotyped as low caffeine (c) by the total number of plants phenotyped as low caffeine (a + c).
4 Overall error rate - TE = (b + c)/n: result from the sum of plants genotyped as low caffeine and phenotyped as high caffeine (b) with plants genotyped as high caffeine and phenotyped as low caffeine (c), divided by the total number of plants.
5 Chi-square with Yates correction - $\chi^2_{McNemar}$ = (|b-c|-0,5)2/b + c: estimate the coincidence rate between the two analytical methods, the genotyping (SNP calling) and phenotyping (HPLC measurement).

Results

In order to verify the association between presence of polymorphism and caffeine-free content in grains we genotyped coffee accessions and progenies under selection. Table 3 shows a list of these genotypes with their corresponding origin and caffeine levels. Although we genotyped three different accessions with reduced levels of caffeine, namely AC1, AC2, and AC3, only the AC1 was used in bi-directional crosses, due to its better breeding potential (Silvarolla, unpublished data). We observed a variation on caffeine content for all evaluated genotypes, and the listed values correspond to the lower and higher values measured for each group of plants. A

complete list with individual caffeine content levels is available (Online resource 3).

The strategy to identify polymorphic SNPs involved alignment of available sequences for *caffeine synthase* (CCS1 and CaDMXT1) with homologous genomic sequences cloned from AC mutants and Mundo Novo cultivar. Among the occurring SNPs, we selected for further analysis those located in previously identified protein sites and potentially associated with reduced caffeine synthase expression (Maluf et al. 2009). From those selected SNPs, the only successful combination of primers and molecular probe for identification of normal and mutant alleles was the one including two SNPs located on exon 2 (Online resource 2). Thus, the genotyping scored the presence/absence of normal SNP1.2 NR and mutant SNP1.2 AC alleles.

Table 2 shows the segregation patterns for selected SNPs and the measured caffeine content. The accession AC1 is homozygous for allele AC, and the accessions AC2, AC3, and the cultivar Mundo Novo are heterozygous. Among the ACs accessions, the AC1 was the only one exhibiting steady reduced caffeine levels (0.03 to 0.27%). On the other hand, content of AC2 (0.06 to 1.44%) and AC3 (0.38 to 1.32%) were similar to those of normal MN grains (0.83 to 1.28%). Based on these analyses, we selected the AC1 accession as the caffeine-free and the cv MN as the high caffeine standards for all further analyses.

Segregation patterns of progenies resulting from crosses of AC1 with other elite cultivars, such as Obatã (OB) and Catuaí Vermelho (CV), also indicate that these have at least one copy of the AC allele. Generally, the presence of a normal allele, SNP1.2 NR, results in synthesis or accumulation of regular caffeine content. Conversely, plants with reduced content contain only the mutated SNP1.2 AC allele, in all tested genetic backgrounds, including Obatã, a cultivar derived from crosses with *C. canephora*. However, a single copy of the AC allele is not sufficient to reduce caffeine content in fruits. The presence of AC alleles in regular caffeine plants, such as cultivar Mundo Novo, confirms this observation. On the other hand, presence of the allele AC in advanced generations such as F_1BC_1, which includes progenies with low caffeine, indicates a reliable association of this allele with caffeine reduction.

The analyses here partially corroborate the inheritance pattern of alleles AC and NR and their association with levels of caffeine in advanced generations. The variable content in grains occurred even in AC1

Table 2 Genetic background, genotype, and phenotype of assessed plants regarding presence of allele NR and AC and caffeine content

Identification[a]	Generation	Polination	Number of plants	With SNP1.2 AC	With SNP1.2 NR	Caffeine (%)
MN	S_2	Open	2	2	2	0.83–1.28
AC1	S_2	Open	2	2		0.06–0.44
AC1	Clone	Open	2	2		0.07–0.08
AC1	S_2	Self	8	8		0.03–0.27
AC2	S_2	Open	3	3	3	0.84–0.92
AC2	S_2	Self	4	4	4	0.06–1.44
AC3	S_2	Open	2	2	2	0.75–0.94
AC3	S_2	Self	6	6	5	0.38–1.32
(MN X AC1)	F_2	Open	38	38	36	0.07–1.24
(MN X AC1)	F_2	Self	20	20	18	0.06–1.19
(AC1 X MN)	F_2	Open	4	4	2	0.25–0.96
(AC1 X MN)	F_2	Self	2	2	2	0.06–0.90
(144 X AC2)	F_2	Open	8	8	7	0.06–1.36
(AC3 X MN)	F_1	Controlled	5	5	5	1.25–1.36
OB X AC3	F_1	Controlled	1	1	1	1.05
OB X AC2	F_1	Controlled	1	1	1	1.20
144 X AC3	F_1	Controlled	1	1	1	1.29
144 X AC2	F_1	Controlled	1	1	1	1.43
(AC2 X AC1)	F_1	Controlled	5	5	5	0.06–1.18
(AC1 X AC2)	F_1	Controlled	9	9	4	0.02–1.20
(AC1 X?)	F_2	Open	26	26	24	0.08–1.22
(OB X AC1)	F_2	Open	11	11	11	0.12–1.22
AC1 X (MN X AC1)	F_1BC_1	Controlled	30	30	15	0.03–1.21
(MN X AC1) X AC1	F_1BC_1	Controlled	10	10	7	0.05–0.97
(MN X AC1) X MN	F_1BC_1	Controlled	7	7	7	0.84–1.19
AC1 X (AC1 X MN)	F_1BC_1	Controlled	1	1	1	0.76
(BV X AC1)	F_1	Controlled	1	1	1	1.06
MN X (AC1 X MN)	F_1	Controlled	1	1	1	1.08
(81 X AC1)	F_2	Open	1	1	1	1.02
(OV X AC1)	F_2	Open	2	2	1	0.03–0.62
(AC3 X 144)	F_1	Controlled	2	2	1	0.77–1.27

[a] MN (Mundo Novo IAC 376–4), 144 (Catuai Vermelho IAC 144), OB (Obatã IAC 1669–20), BV (Bourbom Vermelho), 81 (Catuai Vermelho IAC 81), OV (Ouro Verde IAC H5010–5)

progenies, which does not carry a normal allele but accumulate between 0.06% and 0.44% of caffeine (Table 2). In grains from cv Mundo Novo the variation on caffeine content was between 0.83% and 1.28%.

Results from reciprocal crosses involving AC1 and AC2 indicate that the cross direction may affect gene expression. In this case, progenies from both (AC2 X AC1) and (AC1 X AC2) crossings display similar range of caffeine accumulation, including the typical reduced levels from AC plants. However, all evaluated plants resulting from AC2 X AC1 are heterozygous; indicating that the presence of a normal allele was not sufficient to restore caffeine levels (Table 2).

A second strategy used to validate the correlation between presence of SNP1.2 AC and reduced levels of caffeine included genotyping of F_2 and BC progenies under selection, which had no caffeine profile from grains at the time. Those progenies had their caffeine levels determined afterward. Table 3 shows the results of those analyses. Control accessions ACs and cv MN

Table 3 SNP validation, segregation patterns, and caffeine levels

Genotypes	Number of plants			Caffeine (%)
	Total	With allele AC SNP1.2	With allele NR SNP1.2	
AC1	10	10		0.03–0.39
AC2	7	7	6	0.06–1.44
AC3	8	8	7	0.38–1.32
MN	10	10	10	0.85–1.30
AC1 ♀ X n ♂	26	26	23	0.09–1.22
AC1 ♀ X AC2 ♂	10	10	6	0.05–1.18
MN ♀ X AC1 ♂	60	60	55	0.07–1.24
AC1 ♀ X MN.1 ♂	2	2	1	0.07
CA144 ♀ X AC1 ♂	8	8	7	0.06–1.36
OB ♀ X AC1 ♂	10	10	9	0.15–1.22
(MN ♀ X AC1♂) X AC1	18	18	14	0.05–1.09
(MN ♀ X AC1♂) X MN	2	2	2	1.05–1.19

exhibit the expected correlation between genotype and phenotype, validating the role of AC allele in caffeine accumulation. The results from advanced generations also validate this correlation, once progenies homozygous for SNP1.2 AC show reduced levels of caffeine. However, the allele NR is dominant and controls final content in most of the genotypes. As observed in the previous analyses, this influence seems determined by the cross direction, and some progenies resulting from crosses with female AC1 plants display reduced caffeine even bearing the NR allele (Table 3, cross AC1 X MN). The progenies exhibited the same variable content observed in previous analyses, with even lower levels of caffeine in some AC1 progenies, such as self-crossed AC1; (AC1 X AC2); AC1 X (MN X AC1); (MN X AC1) X AC1 and (OV X AC1).

In order to estimate this correlation accuracy we devised a statistical analysis including all evaluated accessions. In this analysis, we classified all genotypes according to intervals of caffeine content determined by HPLC analyses, as shown on Fig. 1. The distribution of coffee plants along these intervals indicate the occurrence of two major phenotypic classes: one represented by 45 plants with low caffeine content in fruits (class 1: 0.03 to 0.39 mg g-1); and the second including 171 plants with high content (class 2: 0.63 to 1.44 mg g-1).

The plant distribution along genotypic classes defined by SNP segregation exhibit identical numbers as phenotypic distribution, as 45 plants belong to class 1 (+ allele AC; − allele MN) and 171 plants to class 2 (+ allele AC; + allele MN). However, the correspondence is not absolute once four plants from genotypic class 1 belong to phenotypic class 2, and four plants from genotypic class 2 belong to phenotypic class 1. These observations lead to a false positive rate of 2.34% in the first case, and to a false negative rate of 8.89% in the second case (Table 4).

The method accuracy is different for each evaluated generation. Thus, the standard genotypes AC and MN and in F1 progenies exhibited a 100% accuracy. This accuracy is lower, 83.3%, when considering progenies from AC crosses, but in S2 AC1 the accuracy rises to 94.45%. Advanced F_2 and F_1BC_1 generations have an accuracy of 99.12% and 92.16% respectively. In summary, the method accuracy is 96.30% with a total error rate of 3.70% (Table 4). The non-significant χ^2 McNemar value associated with the estimated P value reveals the strong concordance between both analytical methods.

Discussion

Coffee breeding is currently at a bottleneck, once selection for specific traits face challenges such as low genetic variability of the *C. arabica* species, long life cycle and later expression of traits related with the ultimate coffee product, the cup quality. In addition, as the maintenance of *Coffea* germplasm and segregating populations is high cost, breeding programs seek strategies that
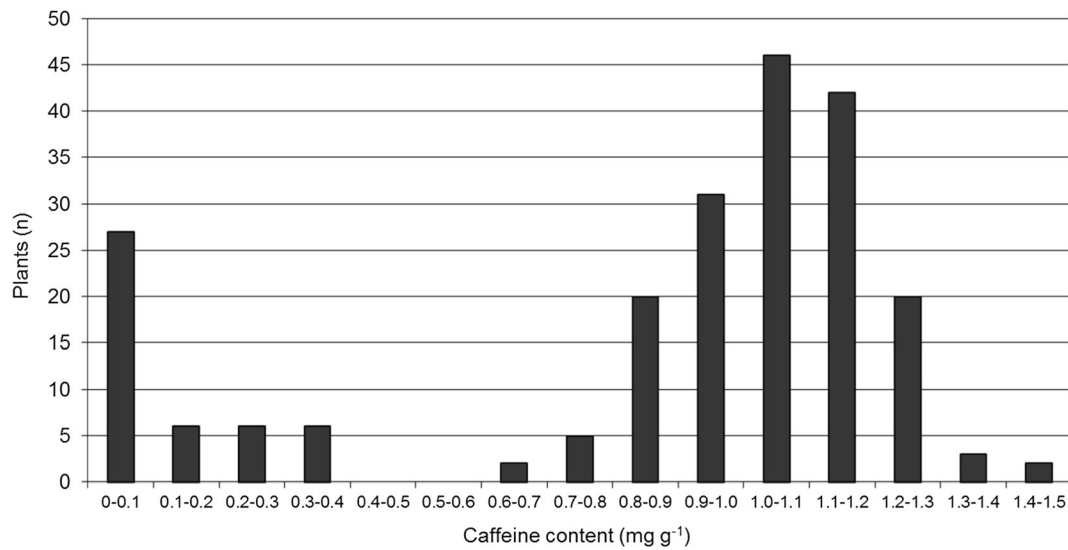
Fig. 1 Number of coffee plants in the intervals defined by caffeine content in grains

allow simultaneous selection of multiple agronomical traits on available populations. Breeding for caffeine-free cultivars is an example of those limitations, not only because this is a later trait but also caffeine levels in leaves do not always correspond to levels in grains. In this scenario, the use of methodologies such as marker-assisted precocious selection that improve precision of selection represent an economy of genetic, human, and financial resources.

The implementation of an assisted-selection strategy requires identification and validation of suitable and reliable molecular markers, which involves genotyping and phenotyping of selected populations. Based on previous analyses of genomic sequences cloned from both Mundo Novo and AC1 plants, we identified two different caffeine synthase sequences: one present exclusively on Mundo Novo genome, which we establish as the normal allele (NR), and the other present on both AC1

Table 4 Estimative of parameters regarding the concordance of analytical methods (TaqMan and HPLC) and classification of coffee plants as low and high caffeine content

Parameters	Population							Total
	Parental		ACs x ACs	ACs S2	F_1	F_2	F_1 RC$_1$	
	MN	ACs						
Plants (n°)	2	9	12	18	11	113	51	216
G*(A:B)	0:2	4:5	3:9	9:9	0:11	15:98	14:37	45:171
F**(A:B)	0:2	4:5	5:7	9:9	0:11	14:99	12:39	45:171
Method accuracy (%)	100	100	83,33	94,45	100	99,12	92,16	96,30
False positive rate (%)	0	0	0	0	0	1,01	7,69	2,34
False negative rate (%)	0	0	40,00	11,11	0	0	8,33	8,89
Overall error rate (%)	0	0	16,67	5,55	0	0,88	7,84	3,70
P value	–	–	0,479	1000	–	1000	0,617	0,724
χ^2 McNemar	–	–	0,500 [ns]	0,000 [ns]	–	0,000 [ns]	0,250 [ns]	0,125 [ns]

*G = Genotypic classes defined by SNP markers analyses: Number of plants with high (A) and low (B) caffeine content; ** F = Phenotypic classes defined by HPLC analyses: number of plants with high (A) and low (B) caffeine content. (−) Values not estimated once the sum of false positive and false negatives equals zero

and MN genome, the mutant allele (AC). Those sequences exhibited a significant number of SNPs, some of which could be potentially associated with the impaired caffeine synthesis detected on AC1 (Maluf et al. 2009). Moreover, a PCR analysis combined with restriction enzyme digestion of partial genomic sequence of caffeine synthase gene indicated that the regular allele comes from the *C. canephora* subgenome and the mutated from the *C. eugenioides* subgenome (Online resource 4). The fact that AC1 has no copy of the regular allele suggests the occurrence of deletion or recombination of this site on the *C. canephora* subgenome, and opens a possibility to use this allele as a molecular-marker for selection of caffeine-free plants. Thus, we genotyped progenies under selection and segregating for no and high caffeine content regarding the presence of two selected SNPs mapped on the *caffeine synthase* gene. The segregation analyses confirmed the previous observations, once regular caffeine MN plants are heterozygous and the AC1 are homozygous for the mutated allele, detected by SNP1.2 AC. Genotyping of other *C. arabica* cultivars evaluated here demonstrated that those also have the mutated allele (data not shown). The molecular analyses also indicated a strong correlation between genotype and phenotype, according to presence or absence of AC (mutant) and NR (normal) allele, once the lack of NR results in reduction of caffeine synthesis. In addition, the analyses indicated that the allele AC is recessive, and one copy only on the genome is not sufficient to block caffeine synthesis. The absence of at least one copy of the NR allele results in reduced caffeine content, in all evaluated plants (Table 2), proving the reliability of these SNPs to separate homozygous mutant plants from heterozygous ones. In addition, the analyses suggest the occurrence of dose effect of the NR allele, once there is a variation on caffeine content in heterozygous progenies, including accessions with low caffeine that are not as reduced as on AC1 plants. Differential expression of genes from the caffeine biosynthetic pathway, differences on genetic background between AC and other coffee plants, and direction of original crosses could also explain differences on the caffeine accumulation observed here.

How the copy number of NR on these progenies affects this variation could not be determined here, once the genotyping method did not allow a quantitative allelic discrimination among heterozygous plants. As reviewed by Baumann (2006) the synthesis of caffeine on coffee fruits may occur in three sites: the perisperm, a tetraploid tissue from maternal origin; the mesocarp, also tetraploid but with male and female genomes; and the endosperm, a triploid tissue. This leads to possible different numbers and combinations of alleles in those tissues, which may affect overall caffeine content in fruits. Although there is little information regarding this concept, studies suggest that differences in ploidy explain contrasting results for caffeine content in crosses between *C. arabica* (4n) and several *Coffea* diploid species (Carvalho and Monaco 1968; Berthaud 1978). Here we genotyped the mother plants, but assessed the caffeine from their grains, which may have different genetic composition from the mother plant. These differences in generations could also explain some of the observed inconsistencies between genotype and phenotype. The quantification from isolated tissues and organs during fruit development may also elicit the ploidy effect over caffeine synthesis and accumulation. In addition, for assisted-selection purposes, the use of more stringent methods for genotyping, such as those based on mass array analyses, will allow a proper quantification of the correct number of copies for each allele, and the correlation of ploidy level with caffeine content will be more accurate.

An interesting observation from field monitoring is that on F_2 generations the production of viable coffee plants with reduced caffeine depends on the cross orientation. Therefore, when AC1 is the female parent there is a higher number of progenies without the NR allele and reduced caffeine content. However, as the efficiency of this cross direction is low (Table 2), the observed frequency could result from a low sample analyses. Besides the fact that AC1 is a wild-type accession, with a significant amount of heterozygose, the plant also exhibits altered architecture and reduced flowering. These characteristics, in association with the reduced number of outcross progenies, suggest that the occurrence of other phenotypic differences associated or not with the caffeine-free trait may affect overall AC1 fertility.

The statistical analyses, with a high estimated P value and non-significant $\chi 2$ value for the method concordance, confirmed the positive relation between marker and phenotype (Table 4). In addition, correlation values were as high as 95%, even when considering plants with intermediate caffeine content (Table 4). This means that, regardless of progeny generation, the rate of positively selecting a low caffeine plant by the genotype (class 1: + allele AC; − allele NR) is around 96 in every 100 plants.

Other statistical parameters that validate this marker are the low rates of false positive and negative (Table 4), which measure the chance to incorrectly score a regular or low caffeine plant. The average overall rate, considering all crossing levels, is lower than 5%, and the higher values are for scoring progenies from AC X AC crosses (17.75%). The high rate observed in these crosses reflect the heterogeneity of AC accessions and their wild-type nature. On the other hand, these observed statistical parameters are significant considering the nature of all crosses involved, with different population backgrounds and generations, and the high number of evaluated progenies. Hence, the statistical analysis confirms the co-relation between the mutated allele and the low accumulation of caffeine in plants. To our knowledge, this is the only molecular-marker for this trait in use by coffee breeding programs.

An important aspect revealed by this analysis is that although the overall content varied among the progenies, no plant exhibits an accumulation in the interval between 0.4 to 0.6% of caffeine (Fig. 1). Usually, caffeine content in hybrid plants is the arithmetic mean of both parental content. Studies with intra and inter-specific hybrids indicated that the variability for caffeine on descendants is directly related to the heterozygosity degree of parents (Priolli et al. 2008), probably due to the additive gene effect discussed above. Thus, the lack of plants accumulating those specific caffeine levels suggests that at least one heterozygous combination or genotype class is missing in those crosses. Also, differences of ploidy level of embryo and endosperm tissues may result in dosage gene effects, affecting overall expression of caffeine synthase and absolute amount of caffeine in grains.

The purpose of this study was to validate a molecular-marker for an efficient selection of naturally caffeine-free plants, and to evaluate the inheritance pattern of this trait to normal caffeine cultivars. Besides the complexity on genotyping polyploid species, here the high sequence similarity of N-methyltransferases from caffeine biosynthetic pathway (McCarthy and McCarthy 2007) makes the allelic discrimination a challenge. Studies in tea, Camelia sinensis, also verified an association of presence of selected SNPs in the caffeine synthase gene (TCS1) with variable caffeine content (Li et al. 2014), confirming the potential of this gene as a selecting tool. In our study, using SNPs markers we identified homozygous and heterozygous genotypes, from different genetic backgrounds, and effectively

selected recessive homozygous plants. The positive selection of progenies with reduced caffeine content from advanced generations of intra-specific crosses indicate the stability and adaptability of the novel trait in commercial cultivars. Moreover, the significant correlation between genotype and phenotype observed here opens the possibility for a breeding strategy based on association genetics. Currently, our breeding program advanced one more generation, with a high rate of plants with low caffeine in fruits, selected using the SNPs here, demonstrating the potential of these markers as an efficient tool for the development of naturally caffeine-free cultivars. In addition, a large-scale differential gene expression analysis is in progress to detect metabolic pathways altered by the lack of caffeine. Besides a greater understanding of the caffeine role during plant and fruit development, with this strategy we expect to identify other candidate genes for development of novel markers for assisted-selection.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Ashihara H, Crozier A (1999) Biosynthesis and catabolism of caffeine in low caffeine-containing species of Coffea. J Agric Food Chem 8:3425–3431

Baumann T (2006) Some thoughts on the physiology of caffeine in coffee – and a glimpse of metabolite profiling. Braz. J. Plant Physiol 18:243–251

Benatti LB, Silvarolla MB, Mazzafera P (2012) Characterization of AC1: a naturally decaffeinated coffee. Bragantia 71: 143–154

Berthaud J (1978) L'hybridation interspe'cifique entre Coffea arabica L. et Coffea canephora Pierre. Obtention et comparison des hybrides triploïdes, arabusta et hexaploïdes. Café Cacao Thé 22:87–109

Campa C, Doulbeau S, Dussert S, Hamon S, Noirot M (2005) Qualitative relationship between caffeine and chlorogenic acid contents among wild Coffea species. Food Chem 93: 135–139

Carvalho A, Monaco LC (1968) Relaciones geneticas de especies selecionadas de *Coffea*. Café 9:1–19

Casal S, Oliveira MBPP, Alves MR, Ferreira MA (2000) Discriminate analysis of roasted coffee varieties for trigonelline, nicotinic acid and caffeine content. J Agr Food Chem 48:3420–3424

Charrier A (1978) La structure génétique des caféiers spontanés de la région Malgache (*Mascarocoffea*). Leur relations avec les caféiers d'origene africaine (*Eucoffea*). Mémoires ORSTOM, 87.

Clifford MN, Johnston KL, Knight S, Kuhnert N (2003) Hierarchical scheme for LC-MS^n identification of chlorogenic acids. J Agric Food Chem 51:2900–2911

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

Ky CL, Louarn J, Guyot B, Charrier A, Hamon S, Noirot M (1999) Relations between and inheritance of chlorogenic acid contents in an interspecific cross between *Coffea pseudozanguebariae* and *Coffea liberica* var *dewevrei*. Theor Appl Genet 98:628–637

Li J, Wei Y-L, Li Y-Y, Pang L, Deng W-W, Jiang C-J (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

Maluf MP, Silva CC, Oliveira MPA, Tavares AG, Silvarolla MB, Guerreiro Filho O (2009) Altered expression of the caffeine synthase gene in a naturally caffeine-free mutant of *Coffea arabica*. Genet Mol Biol 32:802–810

Mazzafera P, Carvalho A (1992) Breeding for low seed caffeine content of coffee (*Coffea* L.) by interspecific hybridization. Euphytica 59:55–60

Mazzafera P, Crozier A, Sandberg G (1994) Studies on the metabolic control of caffeine turnover in developing endosperms and leaves of *Coffea arabica* and *Coffea dewevrei*. J Agric Food Chem 42:1423

McCarthy AA, McCarthy JG (2007) The structure of two N-methyltransferases from the caffeine biosynthetic pathway. Plant Physiol 144:879–889

Priolli RHG, Mazzafera P, Siqueira WJ, Möller M, Zucchi MI, Ramos LCS, Gallo PB, Colombo CA (2008) Caffeine inheritance in interspecific hybrids of *Coffea arabica* x *Coffea canephora* (Gentianales, Rubiaceae). Genet Mol Biol 31:498–504

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning — a laboratory manual. New York Cold Spring Harbor Laboratory, New York

Silvarolla MB, Mazzafera P, Lima de MMA (2000) Caffeine content of Ethiopian *Coffea arabica* beans. Genet Mol Biol 23:213–215

Silvarolla MB, Mazzafera P, Fazuoli LC (2004) A naturally decaffeinated arabica coffee. Nature 429:826

Tango JS, Carvalho A (1963) Teor de cafeína e óleo em variedades de café. Bragantia 56:793–798