ORIGINAL ARTICLE

CrossMark

Association Between SSR Molecular Markers and Chemical and Sensory Traits of Cacao Samples Using Multiple Regression Analysis

Alfredo Vázquez-Ovando[1] · **Héctor Escalona-Buendía**[2] ·
Isidro Ovando-Medina[1] · **Miguel Salvador-Figueroa**[1]

Abstract

The sensory properties of cacao beans are linked to the chemical composition of the seeds, and both characteristics are the partial results of its allelic composition. Therefore, it is useful to search for molecular markers associated with these traits. We perform multiple regression analysis to associate previously generated data of alleles generated with 12 SSR (of cultivated cacao trees) with data obtained from chemical and sensory characterization (of beans) of plants grown in the southern region from Mexico. When the association was significant, the mathematical models for predictive purposes were proposed. All phenotypic traits evaluated showed equations with setting values $R^2 > 0.5$. All chemical characters tested have a significant association with at least two alleles ($P < 0.05$). In addition, the fat content was associated with six molecular markers (mTcCIR03$_{209}$, mTcCIR12$_{188}$, mTcCIR19$_{286}$, mTcCIR07$_{150}$, mTcCIR19$_{310}$). The most common allele was mTcCIR12$_{188}$, which was associated with the contents of eicosanoic acid, moisture, fat and total polyphenols content. The mTcCIR28$_{362}$ allele is associated with sensory characters bitterness, musty odor, and roasted odor. These alleles could be useful as molecular markers of chemical and sensory characteristics of cacao samples.

Keywords Eicosanoic acid · Fat content · Polyphenols · Flavor · Odor

✉ Miguel Salvador-Figueroa
miguel.salvador@unach.mx

1 Instituto de Biociencias, Universidad Autónoma de Chiapas. Boulevard Príncipe Akishino sin número, Colonia Solidaridad 2000, C.P. 30798, Tapachula, Chiapas, Mexico

2 Departamento de Biotecnología, Universidad Autónoma Metropolitana, Unidad Iztapalapa, 09340 Ciudad de México, Mexico

Introduction

The production and marketing of the cacao beans that are obtained from the tropical tree of *Theobroma cacao* L. occupy an important place in the economy of several countries of America, Africa, and Asia. The chemical composition and sensory quality of the cacao beans depends on several factors. The mains include genetic pool of plants, environmental conditions where the tree is grown (Loureiro et al. 2017), as well as the fermentation, drying and roasting conditions of the beans (Kongor et al. 2016). Therefore, these conditions influence the characteristics that will exhibit the cacao products or derivatives.

During the almost 4000 years of human management of this plant (Powis et al. 2011), genetic improvement has been made, however, it was until the second half of the twentieth century when the major efforts were started. These activities sought to obtain desirable characteristics such as large trees, higher productivity, pest tolerance, disease tolerance, and high fruit and seed quality (Monteiro et al. 2009).

Traditional breeding resources that were very useful have been replaced step by step by others supported by molecular markers (Livingstone et al. 2015). Thus, genetic mapping-based analyses focused on the detection of quantitative character-linked loci (QTL) are a tool that has become popular for many plant species. However, experimental crosses are inherently time-consuming and logistically expensive (Stack et al. 2015). In fast-growing plants, the search of QTL is a very profitable option; however, for perennial tree species, this approach it takes several years of work that in some regions of the world is not very cost-effective. For *T. cacao*, several research groups have mapped and located several QTL candidates. The main objectives have been to locate markers for agronomic traits interest or for diseases resistance. Feltus et al. (2011) sequenced and assembled a QTL-rich region of ~3 Mbp and represented by 27 BACs, on linkage group 5 of the *Theobroma cacao* cv. Matina 1–6 genome and identified several genes potentially responsible for black pod disease resistance, bean shape index, and pod weight. Motilal et al. (2016) using an association mapping approach identified several markers SSR and SNP associated with number of filled seeds, black pod resistance and witches' broom disease resistance. Other reports use QTL mapping (Brown et al. 2007), meta-QTL analysis (Lanaud et al. 2009) and genetic linkage map (Royaert et al. 2016) to identify loci associated to disease tolerance/resistance. The search for markers associated with phenotypic characteristics involved in sensory quality has been limited (Vázquez-Ovando et al. 2012). Lanaud et al. (2009) identified QTL's for bitterness, fruits taste, floral taste, among others. Also, Araújo et al. (2009) using QTL mapping identified one QTL associated to butter content at linkage group 9 and two QTLs for butter hardness at linkage groups 9 and 7. The complexity of the sensory traits (aroma, flavor, texture) may be the main reason of the low number of researches in this respect.

It has been shown that the sensory characteristics of cacao beans are related to the chemical composition of the seeds (Luna et al. 2002), so it may be useful to locate molecular markers associated with chemical or sensory characteristics.

Therefore, the studies of associations between molecular markers and phenotypic characters using statistical tools such as multiple regression analysis (MRA) (Virk et al. 1996; Khadivi-Khub 2014; Khadivi-Khub et al. 2014; Körber et al. 2016) may be a useful means for reveal data (alleles) helpful as a previous step to the selection of candidate genomic regions to be QTLs; to predict chemical composition or sensory quality; or alone as true molecular markers for breeding programs. Based on the above, the objective of this work was to search for the association of alleles generated with SSR molecular markers and some chemical and sensory traits of cacao beans using multiple regression analysis.

Materials and Methods

Data Collection

Three data sets were collected from previous studies conducted. First, data from fragments size (indicated as subscripts in the nomenclature of SSR marker) generated from the molecular characterization of cacao trees grown in farms at southern of Mexico with microsatellites markers (mTcCIR1, mTcCIR3, mTcCIR6, mTcCIR7, mTcCIR8, mTcCIR11, mTcCIR12, mTcCIR15, mTcCIR19, mTcCIR21, mTcCIR25, mTcCIR28) obtained by Vázquez-Ovando et al. (2014) were used. These loci markers were chosen because they are distributed throughout the *T. cacao* genome (TropGene DB, Hamelin et al. 2013), they are part of the international set of SSR markers for cacao tree characterization (Saunders et al. 2004) and also have good discriminatory power (Motilal et al. 2009). Details of these loci markers are shown in Supplementary File 1. The cacao trees included as source of DNA (44) belong to cultivated farms of a geographic population or region comprising an approximate area of 4600 km^2 (Soconusco at southern of Mexico), 6 geographic subpopulations (without evidence of genetic grouping by distance) and two genetic groups (Criollo and non-Criollo) and Admixtures individuals, all characterized by 12 SSR markers. The trees were selected based on at least one trait of fruits (pod) or seeds, which resembled those of the Criollo-type (Table 1). Pods were elongated, deeply grooved, pointed at the end, had a lumpy surface with a warty appearance outside, white or slightly pigmented seeds, and sweet micilage as described by Engels (1983) and reported in Vázquez-Ovando et al. (2014).

Second, data from the chemical characterization of cacao beans belonging to the same samples (trees) were obtained (Vázquez-Ovando et al. 2015a); for this, the contents of moisture, ash, fat, protein, fatty acids (C14:0, C16:0, C16:1, C18:0, C18:1, C18:2, C20:0, C22:0), polyphenolic compounds and antioxidant capacity were used. Finally, the numerical data of sensory descriptors reported by Vázquez-Ovando et al. (2015b) obtained by trained panelists, which includes nine odors (chocolate, nutty, hazelnut, acidity, sweet, spicy, roasted, musty, and off-odor) as well as four taste traits (bitterness, acidity, astringency and sweetness). Data from 44 samples were included for the molecular marker-phenotypic characteristic association analysis. Likewise, data from seven composite samples (means of the samples groups) was included for the sensorial characters.

Table 1 Number and information of the cacao trees included in this study

Code	Geographic subpopulation	Genetic group[a]	% membership to Criollo group[b]	Morphological traits and sensory quality
TASG07	1	Non-Criollo	9.6	Leaves: intense green color, acute apex Flowers: red peduncle, white sepals Fruits: elliptical shape, absent basal constriction, slight surface rugosity, yellow color Seeds: oblong shape, intermediate purple color, white pulp
TASG08	1	Non-Criollo	2.0	Leaves: intense green color, acuminate apex Flowers: red peduncle, red sepals Fruits: elliptical shape, absent basal constriction, slight surface rugosity, reddish-yellow color Seeds: elliptic shape, dark purple color, yellow pulp Moderately resistance to black rot pod
TASG09	1	Admixture	89.1	Leaves: intense green color, acuminate apex Flowers: light red peduncle, white sepals Fruits: elliptical shape, acute apex, absent basal constriction, yellow-green color Seeds: ovate shape, white color, yellow pulp Good sensory quality
TASG10	1	Admixture	83.3	Leaves: intermediate green color, acuminate apex Flowers: red peduncle, green–red sepals Fruits: oblong shape, absent basal constriction, slight surface rugosity, reddish-green color Seeds: ovate shape, light purple color, yellow pulp High productivity Good sensory quality
TASG12	1	Admixture	92.6	Leaves: intermediate green color, acute base, acuminate apex Flowers: light red peduncle, white sepals Fruits: elliptic shape, slight basal constriction, slight surface rugosity, yellow color Seeds: elliptic shape, intermediate purple color, yellow pulp Moderately resistance to frosty pod rot Good sensory quality
TASG13	1	Admixture	88.1	Leaves: intense green color, acuminate apex Flowers: light red peduncle, white sepals Fruits: elliptic shape, acuminate apex, slight surface rugosity, reddish-yellow color Seeds: elliptic shape, dark purple color, yellow pulp

Table 1 (continued)

Code	Geographic subpopulation	Genetic group[a]	% membership to Criollo group[b]	Morphological traits and sensory quality
TASG15	1	Admixture	62.3	Leaves: intense green color, obtuse base, acuminate apex Flowers: light red peduncle, white sepals Fruits: elliptic shape, acute apex, slight basal constriction, intermediate surface rugosity, yellow color Seeds: elliptic shape, white and purple color, yellow pulp Good sensory quality
TASG16	1	Admixture	85.8	Leaves: intense green color, acuminate apex Flowers: green with red peduncle, white sepals Fruits: elliptic shape, acute apex, absent basal constriction Seeds: elliptic shape, white color, yellow pulp Good sensorial quality
TASG17	1	Admixture	53.0	Leaves: intense green color, obtuse base, acuminate apex Flowers: red peduncle, white sepals Fruits: elliptic shape, acute apex, slight basal constriction, slight surface rugosity, yellow color Seeds: elliptic shape, intermediate purple color, yellow pulp Good sensorial quality
TASG21	1	Non-Criollo	12.5	Leaves: intense green color, acuminate apex Flowers: light red, white sepals Fruits: elliptic shape, acuminate apex, absent basal constriction, intermediate surface rugosity, yellow color Seeds: elliptic shape, dark purple color, yellow pulp Good sensorial quality
TASG23	1	Admixture	73.4	Leaves: intermediate green color, acuminate apex Flowers: red peduncle, white sepals Fruits: oblong shape, absent basal constriction, intermediate surface rugosity, yellowish green color Seeds: ovate shape, light purple color, yellow pulp

Table 1 (continued)

Code	Geographic subpopulation	Genetic group[a]	% membership to Criollo group[b]	Morphological traits and sensory quality
CAAM01	2	Non-Criollo	2.9	Leaves: intermediate green color, acute base, acuminate apex Flowers: red peduncle, white sepals Fruits: elliptic shape, acute apex, slight basal constriction, intermediate surface rugosity, yellow color Seeds: elliptic shape, light purple color, yellow pulp
CAAM06	2	Non-Criollo	3.7	Leaves: intense green color, acute base, acuminate apex Flowers: red peduncle, white sepals Fruits: obovate shape, absent basal constriction, slight surface rugosity, yellow color Seeds: ovate shape, white color, yellow pulp
CAAM07	2	Admixture	51.5	Leaves: intense green color, obtuse base, acuminate apex Flowers: red peduncle, white sepals Fruits: elliptic shape, intermediate basal constriction, slight surface rugosity, yellow color Seeds: elliptic shape, dark purple color, yellow pulp
CAAM08	2	Non-Criollo	2.2	Leaves: intense green color, obtuse base, acuminate apex Flowers: light red peduncle, white sepals Fruits: elliptic shape, intermediate basal constriction, intermediate surface rugosity, yellowish green color Seeds: elliptic shape, light purple color, yellow pulp
CAAM09	2	Non-Criollo	14.8	Leaves: intense green color, acuminate apex Flowers: reddish brown peduncle, white sepals Fruits: elliptic shape, acute apex, absent basal constriction, slight surface rugosity, yellow color Seeds: ovate shape, dark purple color, yellow pulp
CAAM11	2	Non-Criollo	19.8	Leaves: light green color, acute base, acuminate apex Flowers: light red peduncle, white sepals Fruits: elliptic shape, acute apex, intermediate basal constriction, fused ridge pairs, intense surface rugosity, reddish yellow color Seeds: ovate shape, intermediate purple color, yellow pulp

Table 1 (continued)

Code	Geographic subpopulation	Genetic group[a]	% membership to Criollo group[b]	Morphological traits and sensory quality
CAAM14	2	Non-Criollo	6.4	Leaves: intense green color, acute base, acuminate apex Flowers: red peduncle, whitish green sepals Fruits: elliptic shape, acute apex, intermediate basal constriction, intermediate furrow depth, reddish-green color Seeds: ovate shape, light purple color, yellow pulp High productivity
TCHR05	2	Criollo	99.0	Leaves: intense green color, acute base, acuminate apex Flowers: red peduncle, light purple sepals Fruits: elliptic shape, acute apex, intermediate basal constriction, intense surface rugosity, reddish-green color Seeds: ovate shape, dark purple color, yellow pulp
FHSA03	3	Admixture	55.5	Leaves: intermediate green color, acute base, acuminate apex Flowers: green peduncle, white sepals Fruits: oblong shape, absent basal constriction, slight surface rugosity, yellow color Seeds: ovate shape, light purple color, yellow pulp High productivity
FHSA04	3	Non-Criollo	6.2	Leaves: intense green color, acute base, acuminate apex Flowers: light red peduncle, white sepals Fruits: oblong shape, absent basal constriction, slight surface rugosity, yellow color Seeds: ovate shape, intermediate purple color, yellow pulp Moderately resistance to frosty pod rot
FHSA05	3	Non-Criollo	15.4	Leaves: intense green color, acute base, acuminate apex Flowers: green peduncle, white sepals Fruits: oblong shape, absent basal constriction, slight surface rugosity, yellow color Seeds: ovate shape, white color, yellow pulp High sensory quality

Table 1 (continued)

Code	Geographic subpopulation	Genetic group[a]	% membership to Criollo group[b]	Morphological traits and sensory quality
FHSA06	3	Non-Criollo	1.3	Leaves: intermediate green color, acuminate apex Flowers: green peduncle, white sepals Fruits: oblong shape, absent basal constriction, intermediate surface rugosity, yellow color Seeds: ovate shape, light purple color, yellow pulp
SUED01	3	Non-Criollo	2.4	Leaves: intense green color, acuminate apex Flowers: light red peduncle, white sepals Fruits: elliptic shape, acute apex, absent basal constriction Seeds: elliptic shape, dark purple color, white pulp
SUED03	3	Non-Criollo	1.0	Leaves: intense green color, acute base, acuminate apex Flowers: green peduncle, white sepals Fruits: elliptic shape, acute apex, intermediate basal constriction, fused ridge pairs, dark yellow color Seeds: ovate shape, intermediate purple color, yellow pulp Good sensory quality
SUED05	3	Admixture	2.5	Leaves: intense green color, acuminate apex Flowers: green peduncle, white sepals Fruits: oblong shape, absent basal constriction, intermediate surface rugosity, yellow color Seeds: ovate shape, light purple color, yellow pulp Low productivity High sensory quality
SUED06	3	Non-Criollo	1.2	Leaves: intermediate green color, acute base, acuminate apex Flowers: green peduncle, white sepals Fruits: elliptic shape, acute apex, intermediate basal constriction, intermediate surface rugosity, yellow color Seeds: ovate shape, intermediate purple color, yellow pulp Good sensory quality

Table 1 (continued)

Code	Geographic subpopulation	Genetic group[a]	% membership to Criollo group[b]	Morphological traits and sensory quality
MAMG01	4	Non-Criollo	1.2	Leaves: intense green color, acute apex Flowers: green peduncle, white sepals Fruits: elliptic shape, absent basal constriction, slight surface rugosity, yellow color Seeds: elliptic shape, white color, yellow pulp High sensory quality
MAMG02	4	Non-Criollo	2.4	Leaves: intense green color, acute base, acuminate apex Flowers: light red peduncle, white sepals Fruits: elliptic shape, acute apex, slight basal constriction, intense surface rugosity, yellow color Seeds: ovate shape, light purple and white color, yellow pulp Good sensory quality
MAMG03	4	Non-Criollo	1.8	Leaves: intense green color, obtuse base, acuminate apex Flowers: green peduncle, white sepals Fruits: oblong shape, absent basal constriction, absent surface rugosity, yellow color Seeds: ovate shape, white color, yellow and sweet pulp High sensory quality
MAMG04	4	Non-Criollo	1.0	Leaves: light green color, acute base, acuminate apex Flowers: green peduncle, white sepals Fruits: oblong shape, absent basal constriction, absent surface rugosity, yellow color Seeds: ovate shape, dark purple color, yellow pulp High sensory quality
MAMG05	4	Non-Criollo	8.3	Leaves: intense green color, acute base, acuminate apex Flowers: reddish green peduncle, white (pigmented) sepals Fruits: oblong shape, obtuse apex, absent basal constriction, yellow color Seeds: elliptic shape, white color, yellow and sweet pulp Good sensory quality

Table 1 (continued)

Code	Geographic subpopulation	Genetic group[a]	% membership to Criollo group[b]	Morphological traits and sensory quality
MAMG06	4	Non-Criollo	2.5	Leaves: intermediate green color, obtuse base, acuminate apex Flowers: light red peduncle, white sepals Fruits: elliptic shape, acute apex, absent basal constriction, intense surface rugosity, yellow color Seeds: elliptic shape, intermediate purple color, yellow pulp Susceptible to frosty pod rot High productivity
MAMG08	4	Non-Criollo	14.1	Leaves: intermediate, acuminate apex Flowers: light t red peduncle, white sepals Fruits: oblong shape, slight basal constriction, intermediate surface rugosity, pale yellow color Seeds: elliptic shape, light purple color, yellow pulp Good sensory quality
MAMG10	4	Admixture	23.8	Leaves: intermediate green color, obtuse base, acuminate apex Flowers: green peduncle, white sepals Fruits: elliptic shape, acute apex, slight basal constriction, slight surface rugosity, yellow color Seeds: elliptic shape, light purple color, yellow pulp Resistance to frosty pod rot Good sensory quality
MAMG11	4	Non-Criollo	16.0	Leaves: intense green color, acute base, acuminate apex Flowers: light red peduncle, white sepals Fruits: obovate shape, obtuse apex, slight basal constriction, slight surface rugosity, yellow color Seeds: elliptic shape, white color, yellow pulp High productivity
MAMG13	4	Criollo	98.3	Leaves: intermediate green color, obtuse base, acuminate apex Flowers: red peduncle, white (pigmented) sepals Fruits: elliptic shape, obtuse apex, slight basal constriction, intense surface rugosity, yellow color Seeds: elliptic shape, light purple color, yellow pulp

Table 1 (continued)

Code	Geographic subpopulation	Genetic group[a]	% membership to Criollo group[b]	Morphological traits and sensory quality
HUTG03	5	Non-Criollo	1.6	Leaves: intense green color, obtuse base, acuminate apex Flowers: light red peduncle, white sepals Fruits: elliptic shape, acute apex, slight basal constriction, slight surface rugosity, yellow color Seeds: elliptic shape, light purple color, yellow pulp Susceptible to frosty pod rot
HUTG04	5	Non-Criollo	1.4	Leaves: intense green color, acute base, acuminate apex Flowers: green peduncle, white sepals Fruits: elliptic shape, acuminate apex, intermediate basal constriction, intermediate surface rugosity, yellow color Seeds: ovate shape, white color, yellow pulp High productivity High sensory quality
HUJF04	5	Criollo	97.5	Leaves: intense green color, acute base, acuminate apex Flowers: green peduncle, white (pigmented) sepals Fruits: elliptic shape, acuminate apex, slight basal constriction, slight surface rugosity, yellow color Seeds: ovate shape, light purple color, yellow pulp
HUJF05	5	Criollo	98.9	Leaves: intense green color, acute base, acuminate apex Flowers: light red peduncle, white sepals Fruits: elliptic shape, acuminate apex, intermediate basal constriction, intermediate surface rugosity, yellow color Seeds: ovate shape, dark purple color, white pulp High sensory quality
HUJF06	5	Criollo	98.6	Leaves: intense green color, acute base, acuminate apex Flowers: red peduncle, white sepals Fruits: elliptic shape, acuminate apex, intermediate basal constriction, slight surface rugosity, yellow color Seeds: ovate shape, light purple color, yellow pulp High sensory quality

Table 1 (continued)

Code	Geographic subpopulation	Genetic group[a]	% membership to Criollo group[b]	Morphological traits and sensory quality
TUCA02	6	Non-Criollo	2.4	Leaves: intermediate green color, acute base, acuminate apex Flowers: light red peduncle, white sepals Fruits: oblong shape, obtuse apex, absent basal constriction, intense surface rugosity, yellow color Seeds: elliptic shape, light purple color, yellow and odorous pulp
TURA01	6	Admixture	72.7	Leaves: intense green color, acute base, acuminate apex Flowers: red peduncle, white sepals Fruits: elliptic shape, acute apex, intermediate basal constriction, intense surface rugosity, fused ridge pairs, yellow color Seeds: ovate shape, light purple color, yellow pulp

[a]Grouped verified by the assiggnment (100%) method of Paetkau et al. (2004)

[b]Based on membership to one of two genetic clusters (K) determined with the procedure described by Evanno et al. (2005)

Association Analysis

Following the procedure suggested by Virk et al. (1996), a binary matrix of presence (1) and absence (0) was constructed. The SSR loci of each size (bp) of the 44 samples were scored. The matrix was complemented with data from each of the chemical or sensory characteristics. Molecular data were considered as independent variables and chemical or sensory characters as dependent variables. MRA was performed using the "stepwise" method of "linear regression analysis" option with XLStat Pro software v 2013.

The analysis was based on the model $Y = a + b_1m_1 + b_2m_2 + \cdots + b_jm_j + \cdots + b_nm_n + d + e$, which related the variation in the dependent variable (Y means value for the chemical or sensory character) as a linear function of the set of independent variables (m_j) represented by the SSR markers (Khadivi-Khub 2014). The b_j terms are the partial regression coefficients that specify the empirical relationships between Y and m_j, d represents the residual values between accessions after regression and e is the random error of Y that includes the environmental variation (Virk et al. 1996; Khadivi-Khub 2014). In order to select the independent variables for the regression equation, F values with 0.045 and 0.099 probabilities were used as "enter" and "removal" criteria, respectively (Affifi and Clark 1984; Khadivi-Khub 2014). The R^2 value denotes the multiple determination coefficients.

In addition, the selected markers were tested independently with linear curve fitting using linear models for confirming the significance of the ß statistic for each allele identified by the analysis. Beta can be defined as the standardized regression coefficient $= BS_x/S_y$, where B is the regression coefficient or slope and S_x and S_y are the standard deviations of the independent (x) and dependent (y) variables (Affifi and Clark 1984; Khadivi-Khub 2014). Student's t test was performed to evaluate the significance between phenotypic trait estimates of genotypes where the specific markers were present and absent. Only those markers that showed significant regression values ($P < 0.05$) were considered as associated with the character under consideration (Khadivi-Khub 2014).

To validate the utility of the association and avoid false positives, the frequency distributions of the phenotypic characteristics included in the study were verified by the Shapiro–Wilks normality test using the software XLStat Pro software v 2013. The histograms of density and statistic values are presented in Supplementary File 2.

Results and Discussion

For all the chemical and sensory characters evaluated, significant associations were found with at least one SSR allele and in all cases, the models are significant (P values < 0.05, Tables 2 and 3). Likewise, the models with the R^2 (phenotypic variation proportion) in the range of 0.204 (linoleic fatty acid) to 0.849 (eicosanoic fatty acid) were found for chemical traits (Table 2). For the sensory data models, the smallest R^2 was 0.545 (sweet odor, Table 2). Tables 1 and 2 highlights in bold the phenotypic traits whose models presented values of

Table 2 Microsatellite markers associated with traits of the chemical composition of cacao beans revealed by multiple regression analysis and their respective coefficients

Trait	SSR markers (alleles)	Correlation coefficient (r)	Multiple determination coefficient (R^2)	Standardized beta coefficient	t value	P value
Myristic acid	**mTcCIR11$_{370}$**	**0.684**	**0.468**	**0.640**	**6.547**	**< 0.0001**
	mTcCIR15$_{228}$	**0.749**	**0.561**	**0.256**	**2.533**	**0.015**
	mTcCIR06$_{335}$	**0.780**	**0.609**	**0.253**	**2.583**	**0.014**
	mTcCIR06$_{209}$	**0.812**	**0.659**	**0.224**	**2.372**	**0.023**
Palmitoleic acid	**mTcCIR28$_{330}$**	**0.417**	**0.174**	**−0.414**	**−3.490**	**0.001**
	mTcCIR06$_{261}$	**0.557**	**0.310**	**−0.265**	**−2.215**	**0.033**
	mTcCIR06$_{251}$	**0.636**	**0.404**	**0.342**	**2.966**	**0.005**
	mTcCIR19$_{328}$	**0.709**	**0.502**	**0.323**	**2.774**	**0.008**
Palmitic acid	**mTcCIR19$_{214}$**	**0.479**	**0.229**	**−0.534**	**−5.056**	**< 0.0001**
	mTcCIR06$_{237}$	**0.568**	**0.323**	**−0.389**	**−3.744**	**0.001**
	mTcCIR19$_{310}$	**0.653**	**0.426**	**0.249**	**2.514**	**0.017**
	mTcCIR11$_{370}$	**0.727**	**0.528**	**−0.393**	**−4.005**	**0.000**
	mTcCIR15$_{248}$	**0.766**	**0.586**	**−0.334**	**−3.203**	**0.003**
	mTcCIR06$_{247}$	**0.796**	**0.633**	**0.350**	**3.007**	**0.005**
	mTcCIR28$_{338}$	**0.823**	**0.677**	**−0.244**	**−2.198**	**0.034**
Linoleic acid	mTcCIR08$_{319}$	0.341	0.116	0.419	2.909	0.006
	mTcCIR12$_{210}$	0.452	0.204	0.307	2.132	0.039
Oleic acid	mTcCIR06$_{211}$	0.370	0.137	0.416	3.232	0.003
	mTcCIR06$_{309}$	0.481	0.231	0.349	2.736	0.009
	mTcCIR03$_{199}$	0.560	0.314	−0.295	−2.342	0.024
	mTcCIR15$_{266}$	0.630	0.397	0.296	2.322	0.026
Stearic acid	**mTcCIR19$_{214}$**	**0.480**	**0.230**	**0.475**	**4.824**	**< 0.0001**
	mTcCIR11$_{370}$	**0.584**	**0.341**	**0.597**	**5.556**	**< 0.0001**
	mTcCIR06$_{237}$	**0.690**	**0.476**	**0.486**	**4.852**	**< 0.0001**
	mTcCIR15$_{228}$	**0.756**	**0.572**	**−0.445**	**−4.125**	**0.000**
	mTcCIR07$_{150}$	**0.786**	**0.618**	**−0.257**	**−2.495**	**0.017**
	mTcCIR28$_{351}$	**0.817**	**0.668**	**−0.236**	**−2.348**	**0.024**
Eicosanoic acid	**mTcCIR11$_{370}$**	**0.583**	**0.340**	**0.273**	**3.461**	**0.002**
	mTcCIR15$_{248}$	**0.658**	**0.433**	**0.340**	**4.289**	**0.000**
	mTcCIR25$_{175}$	**0.733**	**0.537**	**−0.587**	**−7.655**	**< 0.0001**
	mTcCIR07$_{172}$	**0.771**	**0.594**	**−0.544**	**−6.536**	**< 0.0001**
	mTcCIR28$_{351}$	**0.810**	**0.656**	**−0.357**	**−4.564**	**< 0.0001**
	mTcCIR19$_{268}$	**0.838**	**0.702**	**0.220**	**2.669**	**0.12**
	mTcCIR12$_{188}$	**0.862**	**0.743**	**−0.442**	**−5.163**	**< 0.0001**
	mTcCIR08$_{319}$	**0.887**	**0.787**	**−0.434**	**−5.082**	**< 0.0001**
	mTcCIR03$_{227}$	**0.908**	**0.825**	**−0.223**	**−2.949**	**0.006**
	mTcCIR07$_{142}$	**0.921**	**0.849**	**−0.207**	**−2.707**	**0.011**
Behenic acid	mTcCIR19$_{236}$	0.458	0.210	0.381	2.357	0.023
	mTcCIR28$_{322}$	0.553	0.306	0.247	2.477	0.018
	mTcCIR07$_{150}$	0.619	0.383	0.194	2.225	0.032

Table 2 (continued)

Trait	SSR markers (alleles)	Correlation coefficient (r)	Multiple determination coefficient (R^2)	Standardized beta coefficient	t value	P value
Moisture content	mTcCIR28$_{330}$	0.436	0.190	−0.440	−3.707	0.001
	mTcCIR12$_{188}$	0.568	0.323	0.374	2.861	0.007
	mTcCIR15$_{248}$	0.644	0.415	0.401	3.144	0.003
	mTcCIR15$_{228}$	0.691	0.478	0.290	2.176	0.036
Ash content	mTcCIR08$_{318}$	0.369	0.136	0.517	3.726	0.001
	mTcCIR06$_{287}$	0.469	0.220	0.357	2.597	0.013
	mTcCIR08$_{265}$	0.560	0.314	0.319	2.348	0.024
Fat content	**mTcCIR03$_{209}$**	**0.395**	**0.156**	**0.495**	**4.359**	**< 0.0001**
	mTcCIR12$_{188}$	**0.544**	**0.296**	**0.443**	**3.979**	**0.000**
	mTcCIR19$_{286}$	**0.638**	**0.407**	**−0.431**	**−3.712**	**0.001**
	mTcCIR07$_{150}$	**0.694**	**0.482**	**−0.268**	**−2.314**	**0.026**
	mTcCIR19$_{310}$	**0.737**	**0.543**	**−0.253**	**−2.260**	**0.030**
Protein content	**mTcCIR19$_{268}$**	**0.401**	**0.161**	**−0.436**	**−3.835**	**0.000**
	mTcCIR06$_{251}$	**0.550**	**0.302**	**0.449**	**3.879**	**0.000**
	mTcCIR15$_{266}$	**0.630**	**0.397**	**−0.337**	**−2.931**	**0.006**
	mTcCIR15$_{248}$	**0.678**	**0.460**	**−0.267**	**−2.345**	**0.024**
	mTcCIR11$_{304}$	**0.718**	**0.516**	**0.240**	**2.086**	**0.044**
Polyphenols content	mTcCIR08$_{265}$	0.359	0.129	−0.403	−2.885	0.006
	mTcCIR12$_{188}$	0.467	0.218	−0.301	−2.153	0.037
Antioxidant capacity	mTcCIR19$_{268}$	0.276	0.076	−0.255	−1.860	0.070
	mTcCIR15$_{248}$	0.407	0.166	−0.406	−2.852	0.007
	mTcCIR12$_{188}$	0.522	0.272	−0.348	−2.418	0.020

$R^2 > 0.5$, but which were also associated with three or more alleles. With this criterion, 33 alleles were significantly associated (positively or negatively) with 13 phenotypic traits and 14 alleles had an association with more than one character (bold in Tables 2 and 3). The most frequent allele was mTcCIR12$_{188}$, which was found to be associated with six phenotypic traits. The second most frequent allele (mTcCIR15$_{248}$) was associated with five traits. Others alleles that tag at least three traits include mTcCIR06$_{237}$ (three traits), mTcCIR06$_{251}$ (three traits), mTcCIR11$_{370}$ (four traits), mTcCIR15$_{228}$ (four traits), mTcCIR19$_{268}$ (four traits), mTcCIR28$_{351}$ (three traits), and mTcCIR28$_{362}$ (three traits). These alleles may be useful in cacao improvement programs, as it has been reported that markers that associate with more than one character may be highly useful in breeding programs (Khadivi-Khub 2014). The mTcCIR15 and mTcCIR19 SSR markers have previously been associated with agronomic performance traits (Clement et al. 2003) and, one QTL related to pod number and to the average weight of single pods was also detected close to mTcCIR15 (Lanaud et al. 1999).

Table 3 Microsatellite markers associated with sensory traits of cacao beans revealed by multiple regression analysis and their respective coefficients

Trait	SSR markers (alleles)	Correlation coefficient (r)	Multiple determination coefficient (R^2)	Standardized beta coefficient	t value	P value
Sweetness	mTcCIR01$_{147}$	0.838	0.702	0.838	3.429	0.019
Bitterness	**mTcCIR28$_{362}$**	**0.844**	**0.713**	**− 0.651**	**–**	**–**
	mTcCIR12$_{188}$	**0.961**	**0.924**	**− 0.494**	**–**	**–**
	mTcCIR06$_{237}$	**0.992**	**0.985**	**0.181**	**–**	**–**
	mTcCIR15$_{248}$	**0.999**	**0.999**	**− 0.151**	**–**	**–**
	mTcCIR19$_{268}$	**1.000**	**1.000**	**0.057**	**–**	**–**
Acidity	**mTcCIR11$_{292}$**	**0.946**	**0.895**	**− 1.015**	**–**	**–**
	mTcCIR11$_{304}$	**0.987**	**0.975**	**− 0.354**	**–**	**–**
	mTcCIR19$_{268}$	**0.998**	**0.997**	**0.129**	**–**	**–**
	mTcCIR06$_{251}$	**1.000**	**1.000**	**− 0.081**	**–**	**–**
	mTcCIR28$_{351}$	**1.000**	**1.000**	**− 0.020**	**–**	**–**
Astringency	mTcCIR12$_{210}$	0.682	0.465	0.682	2.087	0.091
Chocolate odor	mTcCIR12$_{168}$	0.923	0.852	− 1.094	− 9.085	0.001
	mTcCIR25$_{137}$	0.977	0.955	0.364	3.019	0.039
Nutty odor	**mTcCIR01$_{147}$**	**0.831**	**0.690**	**0.716**	**29.198**	**< 0.0001**
	mTcCIR12$_{188}$	**0.993**	**0.987**	**0.575**	**23.437**	**0.000**
	mTcCIR15$_{228}$	**0.999**	**0.998**	**− 0.112**	**− 4.583**	**0.020**
Hazelnuts odor	mTcCIR06$_{211}$	0.910	0.828	1.144	9.183	0.001
	mTcCIR25$_{137}$	0.978	0.957	0.428	3.433	0.026
Acidity odor	**mTcCIR12$_{168}$**	**0.673**	**0.453**	**0.731**	**16.175**	**0.001**
	mTcCIR12$_{188}$	**0.958**	**0.918**	**− 0.635**	**− 14.040**	**0.001**
	mTcCIR19$_{300}$	**0.997**	**0.994**	**− 0.286**	**− 6.322**	**0.008**
Spicy odor	mTcCIR25$_{137}$	0.828	0.686	0.828	3.302	0.021
Musty odor	**mTcCIR12$_{168}$**	**0.682**	**0.465**	**0.870**	**7.510**	**0.005**
	mTcCIR25$_{137}$	**0.922**	**0.851**	**− 0.807**	**− 7.129**	**0.006**
	mTcCIR28$_{362}$	**0.986**	**0.972**	**0.407**	**3.624**	**0.036**
Off odor	mTcCIR12$_{230}$	0.777	0.603	− 1.142	− 6.553	0.003
	mTcCIR28$_{351}$	0.915	0.915	− 0.668	− 3.830	0.019
Roasted odor	**mTcCIR15$_{228}$**	**0.780**	**0.608**	**− 0.522**	**− 17.904**	**0.003**
	mTcCIR28$_{362}$	**0.943**	**0.889**	**0.702**	**24.108**	**0.002**
	mTcCIR12$_{188}$	**0.990**	**0.981**	**− 0.401**	**− 13.777**	**0.005**
	mTcCIR15$_{248}$	**0.999**	**0.999**	**0.161**	**5.534**	**0.031**
Sweet odor	mTcCIR06$_{251}$	0.738	0.545	0.738	2.446	0.050

Another potential use of the SSR alleles associated with the phenotypic traits evaluated in cacao is with a predictive approach. In this sense, all phenotypic characters evaluated presented equations with adjustment values $R^2 > 0.5$ (Table 3) denoting appropriate adjustments to the model. Although all characters found a model that was significant ($P < 0.05$), only a few characters show a predicted data

distribution against observed data consistent with the fit (Fig. 1) and match the highest R^2 values. Thus, palmitic acid, stearic acid, and eicosanoic acid are the most suitable predicted traits. Also, these phenotypic characters will not be subject to bias, since the data come from a population with normal distribution according to the Shapiro–Wilk test (Supplementary File 2). Of all the phenotypic characters included in the study, only the contents of myristic acid ($P < 0.0001$), behenic acid ($P < 0.0001$), and antioxidant capacity ($P = 0.003$) presented a non-normal distribution. For this reason, although myristic acid was associated with four SSR alleles (Table 2) and the prediction equation can even be described, this association would be subject to bias and is not presented (Table 2) and only the adjustment model is shown for comparative purposes (Fig. 1).

The butter content and quality in cocoa beans are genetically controlled characteristics (Araújo et al. 2009). In our analysis, a significant association was found between fat content and several markers. For this reason, the SSR alleles associated with this trait may be candidates to be used in programs to locate cacao accessions with a higher fat content or indicators of the butter quality, given that the content of stearic acid, which determines the chemical characteristics of cocoa butter, showed strong association with six molecular markers (mTcCIR19$_{214}$, mTcCIR11$_{370}$, mTcCIR06$_{237}$, mTcCIR15$_{228}$, mTcCIR07$_{150}$, mTcCIR28$_{351}$). This set of six markers can be useful in predicting the quality of cocoa. It has also been reported that
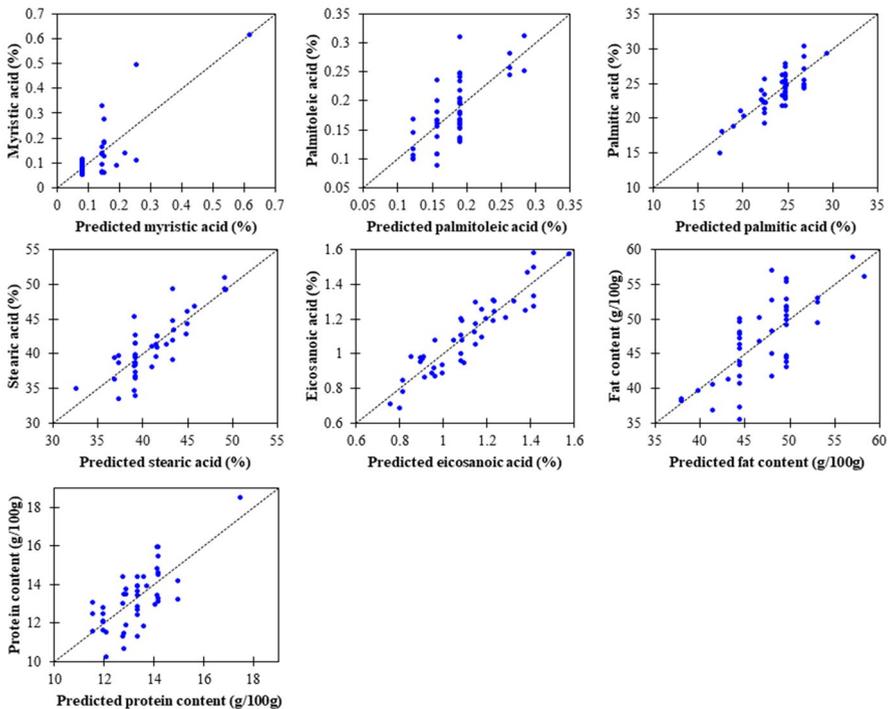
Fig. 1 Correlation drawing between observed data and predicted data for traits of the chemical composition of cacao beans. Obtained from multiple correlation-regressions of 37 alleles

a strong QTL for cocoa butter hardness was found localized in LG 7 near a gene orthologous to fatty acid desaturase 4 (FAD4), which is involved in creating a bond between C2 and C3 of the lipid chain, resulting in lipids with a higher melting point, which, in terms of cocoa butter, represents greater hardness (Argout et al. 2011). The SSR marker mTcCIR07 is locate to LG7.

Bitterness and acidity were associated with five DNA alleles (each of the descriptors) with R^2 values of 1 (Table 4). Another four sensory traits were also associated with several alleles at significance value higher than 0.97. For all phenotypic traits that showed a significant association, a percentage of prediction on the character above 50% even with a single allele (data not shown, R^2 values of 0.5) was achieved. The mTcCIR28$_{362}$ allele was found associated with bitterness taste, musty odor, and roasted odor. Bitterness and musty odor are traits indicators of poor quality of fine cacao. Since all samples were fermented, dried and roasted by the same procedure, the association of the mTcCIR28$_{362}$ allele with these traits could reveal the utility of this fragment as a true molecular marker of undesirable sensory characters in fine cacao. The graphical representations of observed data against the predicted data estimated by the model are shown in Fig. 2. Although the number of samples analyzed for the sensory profile was seven (representative of 45 samples of cacao that were clustered according to their chemical composition), the predictive models fit result in two descriptors with exact-fit (bitterness and acidity).

The physical position of several of the molecular markers used in the study that are associated with chemical or sensorial traits may be an indication of the regions of chromosomes of *T. cacao* that are intrinsically related to phenotypic traits. While SSRs amplify very small alleles, and all those included in the study are associated with non-coding regions as they are di and tetranucleotides (Vieira et al. 2016), the possibility that they are close to coding regions cannot be discounted. The number of SSR allele associated with fatty acids may reinforce this hypothesis. It has been reported in *T. cacao* the presence of a great number of genes encoding for enzymes that have a direct relationship with fatty acid biosynthesis, even higher than of other oleaginous dicotyledons (Argout et al. 2011). Few studies have examined the association of molecular markers with traits different to agronomic yield or disease resistance. It has been reported some molecular markers related to the fat content and cocoa butter hardness (Araújo et al. 2009). Three alleles obtained (in our analysis) with the mTcCIR19 (two) and mTcCIR11 (one) markers located on chromosome 2 associated with fat content coincide with the region reported as QTL marker in the same linkage group (Araújo et al. 2009); however, two additional alleles revealed by the *loci* mTcCIR03 and mTcCIR07 are located on chromosomes 2 and 7, respectively could be useful to explain this phenotypic trait. In the LG 2, 4, and 6, a high number of probable genes involved in fatty acid biosynthesis have been located (Argout et al. 2017; Cocoa Genome Hub 2018). Others alleles revealed with the mTcCIR06 and mTcCIR28 markers may be related with this important trait of cacao beans.

Other researchers analyze the presence of DNA markers related to the color of the cacao pod (Brown et al. 2007; Motamayor et al. 2013) and even, because it is a monogenic trait with a known candidate gene, this trait is used as the positive control to compare the association results with the quantitative traits

Table 4 Multiple regression equations for predicting phenotypic traits of cacao beans

| Trait	Model equation	Multiple determination coefficient (R^2)	F	$Pr > F$
Palmitoleic acid	$Y = 0.19 + (0.09*mTcCIR06_{251}) - (0.03*mTcCIR06_{261}) + (0.07*mTcCIR19_{328}) - (0.07*mTcCIR28_{330})$	0.502	9.827	<0.0001
Palmitic acid	$Y = 26.77 - (2.42*mTcCIR06_{237}) + (7.01*mTcCIR06_{247}) - (7.88*mTcCIR11_{370}) - (2.03*mTcCIR15_{248}) - (4.65*mTcCIR19_{214}) + (5.00*mTcCIR19_{310}) - (2.30*mTcCIR28_{338})$	0.677	10.768	<0.0001
Stearic acid	$Y = 41.54 + (4.21*mTcCIR06_{237}) - (2.41*mTcCIR07_{150}) + (16.71*mTcCIR11_{370}) - (4.29*mTcCIR15_{228}) + (5.78*mTcCIR19_{214}) - (2.35*mTcCIR15_{228})$	0.668	12.405	<0.0001
Eicosanoic acid	$Y = 1.27 - (0.14*mTcCIR03_{227}) - (0.13*mTcCIR07_{142}) - (0.28*mTcCIR07_{172}) - (0.27*mTcCIR08_{319}) + 0.39*mTcCIR11_{370}) - (0.19*mTcCIR12_{188}) + (0.15*mTcCIR15_{248}) + (0.10*mTcCIR19_{268}) - (0.60*mTcCIR25_{175}) - (0.18*mTcCIR28_{351})$	0.849	17.386	<0.0001
Fat	$Y = 47.92 + (13.90*mTcCIR03_{209}) - (3.52*mTcCIR07_{150}) + (5.19*mTcCIR12_{188}) - (6.54*mTcCIR19_{286}) - (9.92*mTcCIR19_{310})$	0.543	9.029	<0.0001
Protein	$Y = 14.15 + (3.32*mTcCIR06_{251}) + (0.80*mTcCIR11_{304}) - (0.84*mTcCIR15_{248}) - (1.24*mTcCIR15_{266}) - (1.38*mTcCIR19_{268})$	0.516	8.092	<0.0001
Bitterness	$Y = 4.66 + (0.21*mTcCIR06_{237}) - (0.74*mTcCIR12_{188}) - (0.16*mTcCIR15_{248}) + 0.06*mTcCIR19_{268}) - (0.69*mTcCIR28_{362})$	1	–	–
Acidity	$Y = 2.9 - (0.17*mTcCIR06_{251}) - (1.5*mTcCIR11_{292}) - (0.74*mTcCIR11_{304}) + (0.19*mTcCIR19_{268}) - (0.03*mTcCIR28_{351})$	1	–	–
Nutty odor	$Y = 0.56 + (1.115*mTcCIR01_{147}) + (0.90*mTcCIR12_{188}) - (0.18*mTcCIR15_{228})$	0.998	592.74	0.000
Acidity odor	$Y = 2.07 + (0.985*mTcCIR12_{168}) - (0.86*mTcCIR12_{188}) - (0.39*mTcCIR19_{300})$	0.994	173.83	0.001
Musty odor	$Y = 1.02 + (1.29*mTcCIR12_{168}) - (0.84*mTcCIR25_{137}) + (0.43*mTcCIR28_{362})$	0.972	35.183	0.008
Roasted odor	$Y = 1.72 - (0.37*mTcCIR12_{188}) - (0.48*mTcCIR15_{228}) + (0.10*mTcCIR15_{248}) + (0.45*mTcCIR28_{362})$	0.999	432.47	0.002
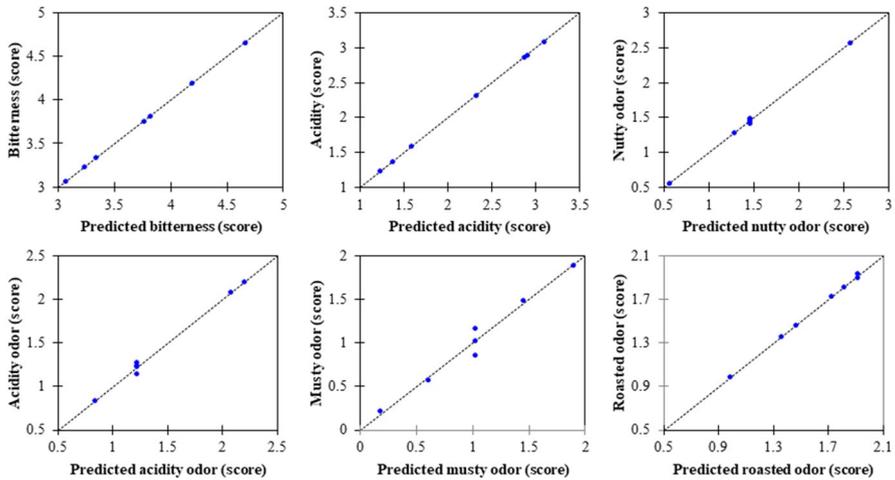
Fig. 2 Correlation drawing between observed data and predicted data for traits of the sensory descriptor of cacao beans. Obtained from multiple correlation-regressions of 37 alleles

(Romero-Navarro et al. 2017). Such markers are located on chromosome 4 of the cacao genome. In our analysis, we did not find molecular markers associated with polyphenolic molecules, which determine the color of the pod; however, in this same chromosome, we locate alleles revealed by the microsatellite mTc-CIR12 (mTcCIR12$_{168}$, mTcCIR12$_{188}$), in addition to other alleles (del LG6, CIR6 y CIR 28) close to genes located in the LG6 (Cocoa Genome Hub 2018) and that are involved in the transformation of phenolic compounds responsible for color and precursors of flavor. This association seems to have the same direction in the correlation (Table 3), that is, the presence of the alleles is related to increase in desirable odors (nutty odor) and a decrease of undesirable odors (musty and acidity odors).

Conclusion

The fit models found in the association study reveal the usefulness of multiple regression analysis to correlate the alleles generated by microsatellite molecular markers and phenotypic traits: palmitoleic acid, palmitic acid, stearic acid, eicosanoic acid, fat content, protein content, bitterness taste, acidity taste, nutty odor, acidity odor, musty odor, and roasted odor. Finally, the empirical validation of the proposed models would be the additional step to convert the alleles proposed in the present analysis into true markers of phenotypic quality.

Compliance with Ethical Standards

Conflict of Interest The authors declare no conflict of interest.

References

Affifi AA, Clark V (1984) Computer-aided multivariate analysis. Van Nostrand Reinhold, New York

Araújo IS, de Souza Filho GA, Pereira MG, Faleiro FG, de Queiroz VT, Guimarães CT, Moreira MA, de Barros EG, Machado RCR, Pires JL, Schnell R, Lopes UV (2009) Mapping of quantitative trait loci for butter content and hardness in cocoa beans (*Theobroma cacao* L.). Plant Mol Biol Rep 27(2):177–183

Argout X, Salse J, Aury J, Guiltinan M, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova S, Abrouk M, Murat F, Fouet O, Poulain J, Ruiz M, Roguet Y, Rodier-Goud M, Barbosa-Neto J, Sabot F, Kudrna D, Ammiraju J, Schuster S, Carlson J, Sallet E, Schiex T, Dievart A, Kramer M, Gelley L, Shi Z, Bérard A, Viot C, Boccara M, Risterucci A, Guignon V, Sabau X, Axtell M, Ma Z, Zhang Y, Brown S, Bourge M, Golser W, Song X, Clement D, Rivallan R, Tahi M, Akaza J, Pitollat B, Gramacho K, D'Hont A, Brunel D, Infante D, Kebe I, Costet P, Wing R, McCombie W, Guiderdoni E, Quetier F, Panaud O, Wincker P, Bocs S, Lanaud C (2011) The genome of *Theobroma cacao*. Nat Genet 43(2):101–109

Argout X, Martin G, Droc G, Fouet O, Labadie K, Rivals E, Aury JM, Lanaud C (2017) The cacao Criollo genome v2.0: an improved version of the genome for genetic and functional genomic studies. BMC Genom 18:730

Brown J, Phillips-Mora W, Power E, Krol C, Cervantes-Martinez C, Motamayor J, Schnell J (2007) Mapping QTLs for resistance to frosty pod and black pod diseases and horticultural traits in *Theobroma cacao* L. Crop Sci 47:1851–1858

Clement D, Risterucci AM, Motamayor JC, N´Goran JN, Lanaud C (2003) Mapping QTL for yield components, vigor, and resistance to *Phytophtora palmivora* in *Theobroma cacao* L. Genome 46:204–212

Cocoa Genome Hub (2018) Access the cocoa Criollo Genome Version 2 with 99% of the genes anchored to the 10 chromosomes. Available at https://cocoa-genome-hub.southgreen.fr/node/4. Accessed: March 3, 2018

Engels JMM (1983) A systematic description of cacao clones. I. The discriminative value of quantitative characteristics. Euphytica 32:377–385

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of cluster of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Feltus F, Saski C, Mockaitis K, Haiminen N, Parida L, Smith Z, Ford J, Staton M, Ficklin S, Blackmon B, Cheng C, Schnell R, Kuhn D, Motamayor J (2011) Sequencing of a QTL-rich region of the *Theobroma cacao* genome using pooled BACs and the identification of trait specific candidate genes. BMC Genom 12:379

Hamelin C, Sempere G, Jouffe V, Ruiz M (2013) TropGeneDB, the multi-tropical crop information system updated and extended. Nucl Acids Res 41:D1172–D1175

Khadivi-Khub A (2014) Regression association analysis of fruit traits with molecular markers in cherries. Plant Syst Evol 300:1163–1173

Khadivi-Khub A, Karimi E, Hadian J (2014) Population genetic structure and trait associations in forest savory using molecular, morphological and phytochemical markers. Gene 546(2):297–308

Kongor JE, Hinneh M, Walle DV, Afoakwa EO, Boeckx P, Dewttinck K (2016) Factors influencing quality in cocoa (*Theobroma cacao*) bean flavour profile—a review. Food Res Int 82:44–52

Körber N, Bus A, Li J, Parkin IAP, Wittkop B, Snowdon RJ, Stich B (2016) Agronomic and seed quality traits dissected by genome-wide association mapping in *Brassica napus*. Front Plant Sci 7:386

Lanaud C, Kébé I, Risterucci AM, Clément D, N'Goran JKA, Grivet L, Tahi M, Cilas C, Pieretti I, Eskes AB, Despréaux D (1999). Mapping quantitative trait loci (QTL) for resistance to *Phytophthora palmivora* in *T. cacao*. In: 12th international conference on cocoa research, 17–23 November 1996, Salvador, Bahia, Brazil. Cocoa Producer's Alliance, Lagos, Nigeria, pp 99–105

Lanaud C, Boult E, Clapperton J, N'Goran JN, Cros E, Chapelin M, Risterucci AM, Allaway D, Gillmore M, Cattaruzza A, Fouet O, Clément D, Petithuguenin P (2005) Identification of QTL related to fat content, seed size and sensorial traits in *Theobroma cacao* L. In: 14th international cocoa research conference. Proceedings: towards a sustainable cocoa economy—what strategies to this end? CEPLAC, CRIG, MCB. Lagos: Cocoa Producers' Alliance, pp 1119–1126. ISBN 978-065-959-5

Lanaud C, Fouet O, Clément D, Boccara M, Risterucci A, Surujdeo-Maharaj S, Legavre T, Argout X (2009) A meta-QTL analysis of disease resistance traits of *Theobroma cacao* L. Mol Breed 24:361–374

Livingstone D, Royaert S, Stack C, Mockaitis K, May G, Farmer A, Saski C, Schnell R, Kuhn D, Motamayor JC (2015) Making a chocolate chip: development and evaluation of a 6K SNP array for *Theobroma cacao*. DNA Res 22(4):279–291

Loureiro GAHA, Araujo QR, Sodré GA, Valle RR, Souza JJO, Ramos EMLS, Comerford NB, Grierson PF (2017) Cacao quality: Highlighting selected attributes. Food Rev Int 37(4):382–405

Luna F, Crouzillant D, Cirou L, Bucheli P (2002) Chemical composition and flavor of Ecuadorian cocoa liquor. J Agric Food Chem 50:3527–3532

Monteiro WR, Lopes UV, Clement D (2009) Genetic improvement in Cocoa. In: Jain SM, Priyadarshan PM (eds) Breeding plantation tree crops: tropical species. Springer, New York, pp 589–626

Motamayor JC, Mockaitis K, Schmutz J, Haiminen N, Livingstone D, Cornejo O, Findley SD, Zheng P, Utro P, Royaert S, Saski C, Jenkins J, Podicheti R, Zhao M, Scheffler BE, Stack JC, Feltus FA, Mustiga GM, Amores F, Phillips W, Marelli JP, May GD, Shapiro H, Ma J, Bustamante CD, Schnell RJ, Main D, Gilbert D, Parida L, Kuhn DN (2013) The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. Genom Biol 14(6):r53

Motilal LA, Zhang D, Umaharan P, Mischke S, Boccara M, Pinney S (2009) Increasing accuracy and throughput in large-scale microsatellite fingerprinting of cacao field germplasm collections. Trop Plant Biol 2(1):23–27

Motilal LA, Zhand D, Mischke S, Meinhardt LW, Boccara M, Fouet O, Lanaud C, Umaharan P (2016) Association mapping of seed and disease resistance traits in *Theobroma cacao* L. Planta 244(6):1265–1276

Paetkau D, Slade R, Burden M, Estoup A (2004) Genetoc assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. Mol Ecol 13:55–65

Powis T, Cyphers A, Gaikwad N, Grivetti L, Cheong K (2011) Cacao use and the San Lorenzo Olmec. Proc Natl Acad Sci 108(21):8595–8600

Romero-Navarro JA, Phillips-Mora W, Arciniegas-Leal A, Mata-Quirós A, Haiminen N, Mustiga G, Livingstone D, Bakel H, Kuhn DN, Parida L, Kasarskis A, Motamayor JC (2017) Application of genome wide association and genomic prediction for improvement of cacao productivity and resistance to black and frosty pod diseases. Front Plant Sci 8:1905

Royaert S, Jansen J, da Silva DV, de Jesus Branco SM, Livingstone DS, Mustiga G, Marelli JP, Araújo IS, Corrêa RX, Motamayor JC (2016) Identification of candidate genes involved in Witches' broom disease resistance in a segregating mapping population of *Theobroma cacao* L in Brazil. BMC Genom 17:107

Saunders JA, Mischke S, Leamy EA, Hemeida AA (2004) Selection of international molecular standards for DNA fingerprinting of *Theobroma cacao*. Theor Appl Genet 110:41–47

Stack J, Royaert S, Gutierrez OA, Nagai C, Schnell RJ, Motamayor JC (2015) Assessing microsatellite linkage disequilibrium in wild, cultivated, and mapping populations of *Theobroma cacao* L. and its impact on association mapping. Tree Genet Genomes 11:19

Vázquez-Ovando A, Molina-Freaner F, Nuñez-Farfán J, Salvador-Figueroa M (2012) Potencial de los marcadores moleculares para el rescate de individuos de *Theobroma cacao* L de alta calidad. BioTecnología 16(1):36–56 **(In Spanish with abstract in English)**

Vázquez-Ovando JA, Molina-Freaner F, Nuñez-Farfán J, Ovando-Medina I, Salvador-Figueroa M (2014) Genetic identification of *Theobroma cacao* L trees with high Criollo ancestry in Soconusco, Chiapas, Mexico. Genet Mol Res 13(4):10404–10414

Vázquez-Ovando A, Molina-Freaner F, Nuñez-Farfán J, Betancur-Ancona D, Salvador-Figueroa M (2015a) Classification of cacao beans (*Theobroma cacao* L.) of southern Mexico based on chemometric analysis with multivariate approach. Eur Food Res Technol 240(6):1117–1128

Vázquez-Ovando A, Chacón-Martínez L, Betancur-Ancona D, Escalona-Buendía H, Salvador-Figueroa M (2015b) Sensory descriptors of cocoa beans from cultivated trees of Soconusco, Chiapas, Mexico. Food Sci Technol 35(2):285–290

Vieira MLC, Santini L, Diniz AL, Munhoz CF (2016) Microsatllite markers: what they mean and why they are so useful. Gent Mol Biol 39(3):312–328

Virk PS, Ford-Lloyd BF, Jackson MT, Pooni HS, Clemenot TP, Newbury HJ (1996) Predicting quantitative variation within rice germplasm using molecular markers. Heredity 76:296–304