Received: 12 November 2021 | Revised: 16 February 2022 | Accepted: 25 March 2022

DOI: 10.1002/ppp3.10268

Plants People Planet PPP Open Access

RESEARCH ARTICLE

Variability and genetic determinants of cocoa aromas in trees native to South Ecuadorian Amazonia

Kelly Colonges[1,2,3,4] | Rey Gastón Loor Solorzano[5] | Juan-Carlos Jimenez[5] | Marie-Christine Lahon[3,4] | Edward Seguine[6] | Darío Calderon[5] | Cristian Subia[5] | Ignacio Sotomayor[5] | Fabián Fernández[5] | Marc Lebrun[3,4] | Olivier Fouet[1,2] | Bénédicte Rhoné[1,2] | Xavier Argout[1,2] | Pierre Costet[7] | Claire Lanaud[1,2] | Renaud Boulanger[3,4]

[1]Cirad, UMR AGAP, Montpellier, France

[2]AGAP Institut, Univ Montpellier, Cirad, INRAE, Institut Agro, Montpellier, France

[3]Cirad, UMR Qualisud, Montpellier, France

[4]Qualisud, Univ Montpellier, Avignon Université, Cirad, Institut Agro, IRD, Université de La Réunion, Montpellier, France

[5]Instituto Nacional de Investigacion Agropecurias, INIAP, Quito, Ecuador

[6]Seguine Cacao/Guittard Chocolate Co, Arroyo Grande, California, USA

[7]Valrhona, Tain-l'Hermitage, France

Correspondence
Kelly Colonges and Renaud Boulanger, Qualisud, Univ Montpellier, Avignon Université, Cirad, Institut Agro, IRD, Université de La Réunion, Montpellier, France.
Email: kcolonges@gmail.com; renaud.boulanger@cirad.fr

Claire Lanaud, AGAP Institut, Univ Montpellier, Cirad, INRAE, Institut Agro, Montpellier, France.
Email: claire.lanaud@cirad.fr

Funding information
Agence Nationale de la Recherche (ANR, the French National Research Agency), Grant/Award Number: ANR-16-IDEX-0006; U.S. Department of Agriculture (USDA-ARS); Valrhona; I-Site MUSE

Societal Impact Statement

Recent surveys conducted on Amazonian cocoa trees in their home range are a unique opportunity to assess the aromas, diversity and potential of the Ecuadorian Amazon to create new aromatic cocoa varieties. Our results reveal informations about the diversity and genesis of aromas in Ecuadorian fine cocoa. The great aromatic diversity could enrich cocoa flavour selection programmes and provide Amazonian populations with new income linked to aromatic varieties, which could lead to a '*grand cru*' chocolate. Until now, breeding programmes have been mainly focused on improving characteristics for production, but in recent years cocoa quality has been increasingly examined to meet market expectations.

Summary

- Ecuador is known worldwide for its fine or flavour cocoa from the Nacional variety. Currently, farmers mainly grow the modern Nacional variety of cocoa trees (hybrids between the ancestral Nacional and Trinitario) while the ancestral Nacional variety tends to be neglected. In order to enlarge the genetic resources related to this ancestral variety, several surveys were carried out in its area of origin located in the South Ecuadorian Amazonia.
- The 202 trees resulting from these surveys were genotyped using GBS (Genotyping By Sequencing) markers and their dried-fermented beans were characterised for both their aromatic volatile compounds and sensorial quality. A genome wide association study (GWAS) was carried out with the aim to study the genetic and biochemical bases of the aroma traits of this population and to better exploit them in breeding programmes.
- Some association areas and candidate genes related to spicy and woody aromas were found for the first time in *Theobroma cacao*. Some association areas and

candidate genes related to floral and fruity aromas common to other studies were identified.
- Our results support the hypothesis that aroma formation may be related to a defence reaction to biotic and abiotic stresses generated by the fermentation process.

KEYWORDS
aroma, aromatic compound, Ecuadorian Amazonia, genetic resources, genome wide association study, *Theobroma cacao*

1 | INTRODUCTION

Theobroma cacao belongs to the Malvaceae family (Bayer & Kubitzki, 2003) and originated in the humid tropical regions of the north of South America, and mainly from the Amazonian basin. *T. cacao* is a tree of great agronomic and economic interest. Indeed, it is the only worldwide source of chocolate, whose consumption is constantly increasing and is studied for its health benefits (Yeh et al., 2016). It provides a large panel of various aromas (Andújar et al., 2012; Tuenter et al., 2018) highly appreciated, but also offers benefits for human health thanks to its richness in polyphenols (Andújar et al., 2012). Cacao can be classified into two main classes: bulk cocoa characterised by strong notes of cocoa and fine or flavour aromatic cocoa characterised by fruity and/or floral aromas (Cook & Meursing, 1982; Sukha et al., 2008). Fine or flavour cocoa represents about 12% of the current world production. However, its consumption has continued to rise for decades and sought after by chocolate-makers looking for new flavours. It represents important economic niches for tropical countries producing this type of cocoa.

The fine or flavour cocoa varieties currently mostly cultivated are (1) the Criollo, well known for its fruity aromas but yet less cultivated due to its low vigour and low resistance to diseases (Cheesman, 1944); (2) the Trinitario, which are hybrids between from both the Amelonado and Criollo genetic groups, and (3) the ancestral and modern Nacional known for its floral notes, which are called Arriba (Loor et al., 2009; Luna et al., 2002).

The latter is endemic to the Ecuador and is only grown in Ecuador. The trees currently cultivated belong to the modern Nacional variety, which are hybrids between the ancestral Nacional variety and Venezuelan cultivars corresponding to Trinitario genotypes (Bartley, 2005). Using molecular markers, Loor et al. (2009) confirmed this hybrid type.

Although it is currently mainly cultivated along the pacific coast, the modern Nacional variety originated from the Southern part of the Ecuadorian Amazonia (Loor et al., 2012).

Surveys were undertaken in the putative domestication centre of the ancestral Nacional variety, in the South part of the Ecuadorian Amazonia, to search for cocoa trees related to the ancestral Nacional variety (Loor et al., 2009, 2015). These expeditions aimed to collect and save genetic diversity related to the ancestral Nacional, offering new genetic resources to improve the modern Nacional variety or create new aromatic varieties.

Previous recent studies aimed to decipher the genetic determinant of aromatic flavours of cocoa from the modern Nacional variety. In a study focusing on floral aromas, Colonges, Jimenez, Saltos, Seguine, Loor Solorzano, Fouet, Argout, Assemat, Davrieux, Cros, et al. (2021a) was able to identify two main biosynthetic pathways involved in the synthesis of floral aromas: the monoterpene biosynthetic pathway and the L-phenylalanine degradation pathway. Then investigating the genetic determinism of fruity aromas, Colonges, Jimenez, Saltos, Seguine, Solorzano, Fouet, Argout, Assemat, Davrieux, Morillo, et al. (2021c) identify five main biosynthetic pathways involved: the monoterpene biosynthetic pathway, the L-phenylalanine, the fatty and sugar degradation pathways and the synthesis of Maillard precursors (for the production of pyrazines).

To study the variability and genomic determinants of the cocoa aromas of the Amazonian population, with a broader genetic base than the previously studied modern Nacional population, a complete characterisation of the genetic and biochemical traits of the cocoa trees collected on Amazonia was undertaken. To this end, volatile compounds of the fermented and dried beans of each Amazonian tree were evaluated by GC–MS (Gas chromatography–mass spectrometry). Sensory analyses of the liquors (dried fermented beans roasted and ground) were also carried out. The molecular characterisation of this population, obtained by GBS, was used to conduct a GWAS (Genome-Wide Association Study) on all these traits to decipher their genetic determinants.

T. cacao has a small genome (from 411 to 494 Mb), now fully sequenced and annotated from accessions of two varieties: Criollo (Argout et al., 2011, 2017) and Amelonado (Motamayor et al., 2013). The two available genome sequences allowed identifying candidate genes within the genomic regions, potentially involved in these aromas.

2 | MATERIALS AND METHODS

2.1 | Vegetal material

Two hundred and two cocoa trees were used for this study. They came from surveys carried out in the South part of Ecuadorian Amazonia corresponding to the Zamora Chinchipe province identified as the putative domestication origin of the ancestral Nacional variety in Ecuador (Loor et al., 2012, 2015). The same collected trees were planted in two INIAP experimental centres: in Pichilingue (EET-P) and

Domono; and other collected trees were planted in an experimental plot of an agricultural college in Pangui.

2.2 | Harvest and micro fermentations

The matured pods were harvested from cacao trees grown in the three different locations. Fresh beans extracted from the pods were fermented and dried in the place of Domono. Fermentations were carried out within 24 h of harvest under the most homogeneous conditions possible following a micro-fermentation process. For that, the cocoa beans of each genotype were placed in delicate laundry nets. They were then distributed over four floors in the middle of the mass of modern Nacional cocoa beans. At 24 and 72 h of fermentation, stirring was performed. At each stirring, the bags of beans at the bottom were placed at the top and those in the middle-low position were placed in the middle-high and vice versa. During the entire fermentation process, button pills recorded the temperature to check the homogeneity of the transitions of the different stages of fermentation in the different cases. The beans used to make up the mass of fermentation were all chosen from the same variety (modern Nacional variety). In order to verify the uniformity of the fermentation batches, we carried out on each batch of mass: a follow-up of the temperatures during the fermentation, a determination of the volatile compounds (SPME GC–MS) and a sensory analysis. The temperature curves between the different cases are similar. The PCA analyses of the SPME GC–MS volatile compounds results showed that 7 out of 10 masses were very close in their volatile compounds profile. The two different masses differed mainly in three volatile compounds (acetic acid, methyl acetate and 3-methylbutyl acetate) were removed from the study.

After 4.5 days of fermentation, the beans were taken out of the net and dried separately in a greenhouse until a moisture content of at least 8%. The fermented-dried beans were then placed under vacuum until biochemical and sensorial analyses.

The micro-fermentation period lasted 4 months and some clones were fermented several times. At the end of the micro-fermentation period all samples of the same clone were mixed. Biochemical and sensory analyses were done on a sample of the mixture.

2.3 | Biochemical analysis of volatile compounds

Solid Phase MicroExtraction (SPME) coupled to GC–MS analyses were conducted according to the condition described by Assi-Clair et al. (2019) on the 202 genotypes of cacao. We used the average of the three replicates for the phenotype of the GWAS analysis.

2.4 | Sensorial analysis

One hundred and fifty-nine genotypes were characterised by sensory analyses based on blind tastings carried out on three repetitions per sample. The tastings were carried out on cocoa liquor. The cocoa liquor corresponds to merchant cocoa (dried fermented beans) which have been roasted and crushed. Sensorial notes were judged with a score ranging from 0 (no notes detected) to 10 according to ISCQF (2020) protocol. We used the average of the three replicates for the phenotype of the GWAS analysis.

2.5 | DNA extraction and SNP genotyping

DNA was extracted following the protocol of Risterucci et al. (2000) protocol. DNA samples were genotyped by sequencing (GBS) using DArTseq (Diversity Arrays Technology Sequencing) technology (Kilian et al., 2012) and carried out by the DArT company. The resulting raw reads were recovered and prossessed as follow: the adapter sequences and low quality scores extremities (−q 20) were removed from the reads using Cutadapt (v2.10). The reads with a length smaller than 20 bases were filtered. The remaining reads were mapped to the V2 sequence of the Criollo reference genome (Argout et al., 2017) using BWA (Burrows-Wheeler Aligner, v 0.7.15) with the MEM algorithm and standard parameters. SNP calling was performed using HaplotypeCaller in the Genome Analysis Toolkit (GATK v 4.1.9.0) and the final SNP set was established using the GATK VariantFiltration tool with stringent criteria (bi-allelic SNPs only with at least three reads of the alternate allele to be called as an SNP, depth coverage >6 and <40 per accession and no missing data). SNP Markers with unknown locations were discarded for analysis. This protocol was used for the control cacao trees and clones studied.

2.6 | Diversity and structure analysis

Fifty-one cacao tree accessions representative of 11 distinct genetic groups of the *T. cacao* species (identified by Motamayor et al., 2008) were taken as controls for the calculation of diversity and population structure (Dataset S1).

The phylogenetic tree was generated using DARwin software (Perrier & Jacquemoud-Collet, 2006). The genetic distances were calculated using the Dice coefficient and the Neighbour-Joining method.

2.7 | Linkage disequilibrium calculation

Linkage disequilibrium (LD) between pairs of SNPs was calculated with Haploview 4.2 (Barrett et al., 2005) and the LD decay was plotted using the 'ggplot2' R package according to Sardos et al. (2016) protocol.

The LD decay between pairs of SNPs plotted for each chromosome (Figure S1) show that the r^2 value decreases by half around 600 kb. Given that 1 MB corresponds to around 2 cM (Loor, 2007), the LD decay of this population is around 1.2 cM. We decided to use this 600-kb limit to determine the confidence interval of associations. For each positive marker we reported an association area of plus or

minus 300 kb, that is, an association area of 600 kb. If two or more associated markers have overlapping confidence intervals, they are grouped in a single association zone. The lowest and highest number of bp of the grouped markers represents the confidence intervals of this zone.

2.8 | GWAS analysis

SNPs with no missing data and a MAF greater than 5% were selected for the GWAS. The final dataset is composed of 5337 SNP markers. The data set is available in the database tropgene (http://tropgenedb.cirad.fr/tropgene/JSP/interface.jsp?module=COCOA) at the study named 'Cocoa_Amazonie_Ecuador_aroma'.

A GWAS analysis was performed to associate SNP markers with biochemical traits (202 phenotyped accessions × 5337 markers) and with sensory (159 phenotyped accessions × 5337 markers) traits using TASSEL v5.

For all the traits, a mixed model (MLM) was carried out with a structure matrix, determined by running a PCA (principal component analyses integrated with TASSEL v5 software), considered as a fixed effect, and also with a kinship matrix considered as a random effect as covariates to control the false-positive rate. The option of not compressing and re-evaluating the components of variance for each marker was chosen. The kinship matrix was inferred using the Identity by State (IBS) pairwise method proposed by Tassel v5.

The threshold was determined using the R package Simple M based on the Bonferonni correction (Gao et al., 2008, 2010). The threshold corresponds to a p value of 1676×10^{-5}. The confidence interval of the associations was chosen according to the LD.

The physical maps plotting the association areas were created using SpiderMap v1.7.1 software (Rami, 2017).

Candidate genes were identified in a region of 300 Kbp (around the associated markers) using the *T. cacao* genome sequence V2 (Argout et al., 2017). This window size of 600 kb was chosen based on the decrease of LD depending on physical distance along chromosomes estimated from the studied cacao Ecuadorian population. Genes were identified as candidate genes when they fell within the association zone and were annotated as having a particular function that may be involved in the biosynthetic pathway of the associated compound.

3 | RESULTS

3.1 | Phenotypic traits studied

3.1.1 | Biochemical traits

To study the biochemical determinism of cocoa flavours, the volatile compounds contained in the fermented-dried beans were determined by GC–MS. All the compounds identified and their already known aromas are shown in Dataset S2.

Strong positive correlations (between .8 and 1) were identified between several volatile compounds. In contrast, no strong negative correlations (between −.8 and 1) were identified (Figure S2). The strong correlations were observed between biochemical compounds with similar structures.

PCA were performed. These analyses show a continuous variation within the population for the concentrations of volatile compounds (Figure S3a). Axis 1 is mainly influenced by the concentration of cis-2,6-dimethyl-2,6-octadiene, 2,6-dimethyl oct-2-ene, 1,2-dihydrolinalool. Axis 2 is mainly influenced by the concentration of pentan-2-ol, 2-methylpropyl acetate, 2-methyl-but-3-en-2-ol.

3.1.2 | Sensorial traits

A sensory study of the liquors of each cocoa genotype was conducted in which 33 criteria were evaluated (Dataset S3). The three notes, savoury/umami, mouldy and smoky, were not detected in any of the genotypes.

Three strong positive correlations (between .8 and 1) could be observed between the different sensory characteristics. In contrast, no strong negative correlations (between −.8 and 1) were identified (Figure 1).

The results of the PCA performed on the sensory data also show a continuous variation within the population (Figure S3b). Axis 1 is mostly influenced by the roast degree score, the cocoa score and the bitterness score. Axis 2 is mostly influenced by the note 'fruit-tropical', the note 'Grassy, Green, vegetal, Herbal' and the note 'Fruit-Dried'.

3.2 | Genetic diversity and population structure

Phylogenetic analyses performed from SNP markers show that the cocoa population native from the Zamora Chinchipe province is mostly close to the Curaray and Nacional genetic groups. Some accessions are also close to the Caqueta, Contamana and Iquitos genetic groups (Figure 2). A subgroup composed of different trees, all originated from the Pangui region (PAN survey at the North of the region and PGI survey more around the city of El Pangui; Loor et al., 2015) appears more distant from the other accessions (Figure 2).

3.3 | Genome-wide association study

The final dataset is composed of 5337 SNP markers. The SNPs are relatively well distributed along the 10 chromosomes of *T. cacao*. However, a decrease in marker density is observed in the centromeric and pericentromeric regions (Figure 3).

All areas of significant associations are shown in Dataset S4. The number of markers associated and the total number of associations for each trait are reported in Dataset S5.

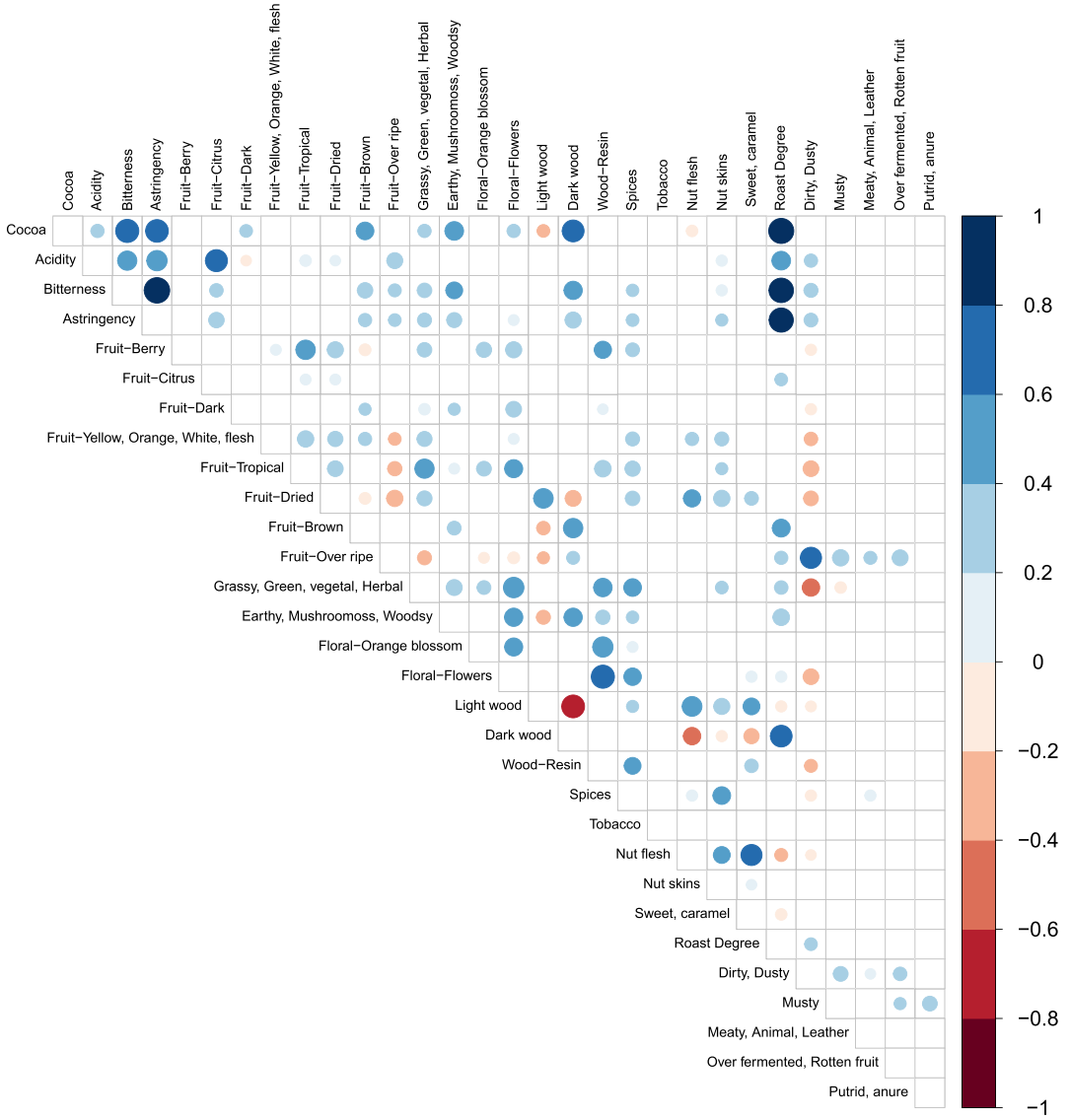

FIGURE 1 Significant correlation matrix of sensorial profiles of cocoa beans from *Theobroma cacao*. Correlation matrix between the sensorial profiles determined in cocoa liquor. The correlations were calculated by the Pearson method. The white boxes represent no significant correlations. The colour of the circles corresponds to Pearsons correlation coefficient. The areas of circles correspond to a *p* value of correlation coefficients. The *p* value threshold for a significant correlation is .05. The different shades of blue represent a positive correlation coefficient while the different shades of red represent a negative correlation coefficient. The intensity of the colour depends on the strength of the R2 correlation coefficient. The scale on the right indicates the interpretations of different colours

3.3.1 | Identification of significant association areas in relation to floral notes

A total of 39 association areas were detected to compounds known to have a floral note and 9 other associations were detected linked to the floral notes, detected by sensory analysis (Figures S4 and S5). Associations were detected on all chromosomes except chromosome 4 (Dataset S4). Seven areas of co-locations between different biochemical compounds as well as between biochemical compounds and sensory notes were observed (Table 1). Two volatile compounds seem to be more related to the presence of orange blossom notes: the cis-beta-ocimene and the o-xylene (they are both known to have floral notes; Dataset S1).

3.3.2 | Areas of association detected in relation to fruity notes

A total of 194 association zones were detected in relation to volatile compounds known to have a fruity note and 39 linked to fruity notes detected by sensory analysis (Figures S4–S7). Areas of association were detected on all chromosomes (Dataset S4). Among the 56 areas of co-locations between the different traits, 26 of them display co-locations between volatile compounds and sensory notes (Table 2):

Diethyl butanedioate, nonan-2-one, 2-pentylfuran, ethyl lactate, trimethyl-oxazole, pentyl acetate, heptan-2-ol, 2,3,5-trimethylpyrazine and ethyl 2-hydroxy-4-methylpentanoate seems to be more related to over fermented/rotten fruit notes.

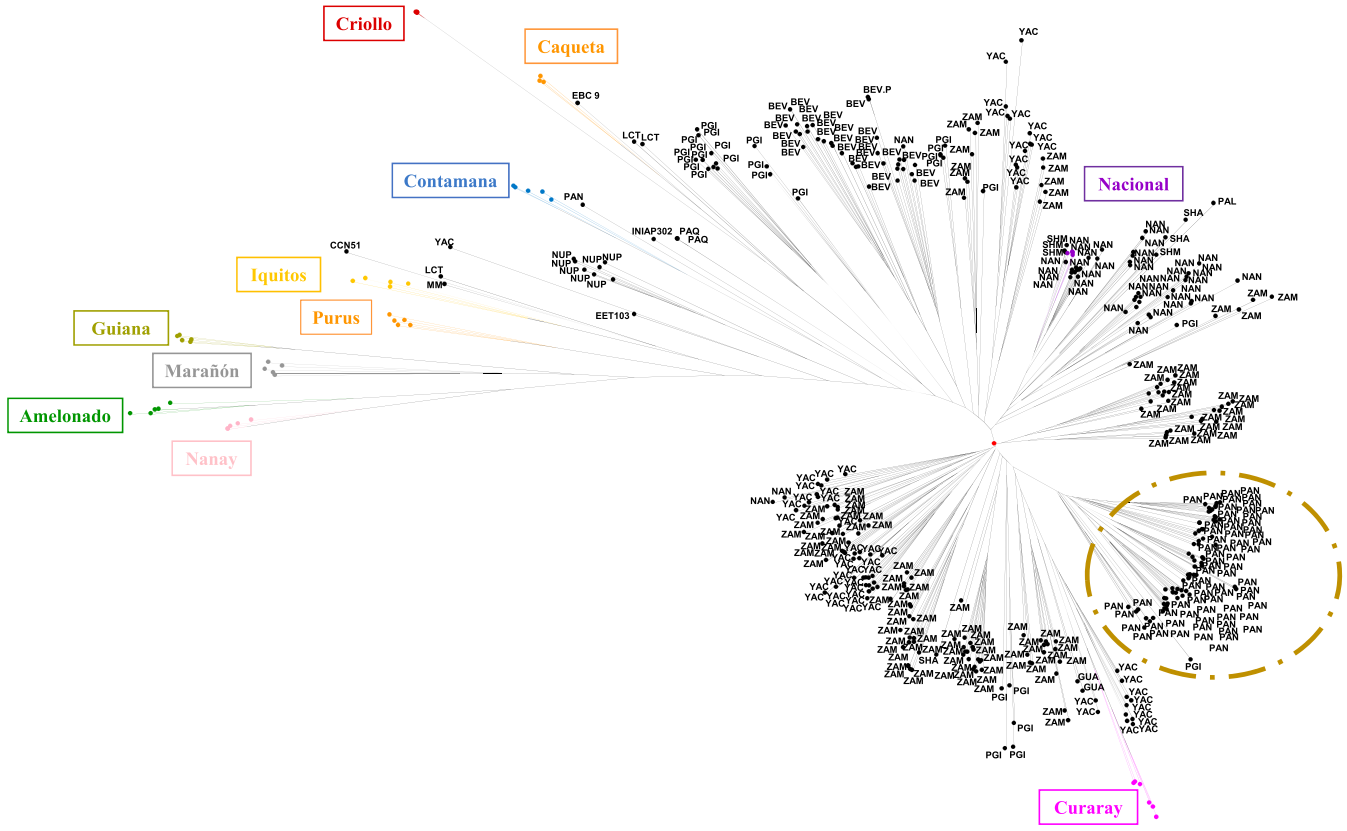

FIGURE 2 Phylogenetic tree representing the genetic diversity of the studied population of *Theobroma cacao*. The phylogenetic tree, established with the Darwin software, represents the genetic diversity of the individuals in the studied population (shown in black), with the known cocoa genetic groups (in colour) taken as a control. The name of each Amazonian location of the collection is indicated as follow, and according to Loor et al. (2015) BEV: BEVI, GUA: GUAI, LCT: LCTEEN, NAN: NANK, NUP: NUPA, PGI: PANGUI 2, PAL: PALANDA, PAN: PANGUI 1, SHA: SHAI, SHM: SHAM, YAC: YACU, ZAM: ZAMORA. Abbreviation: P:Pichilingue. The brown dashed circle shows cocoa trees that appears more distant from the other accessions

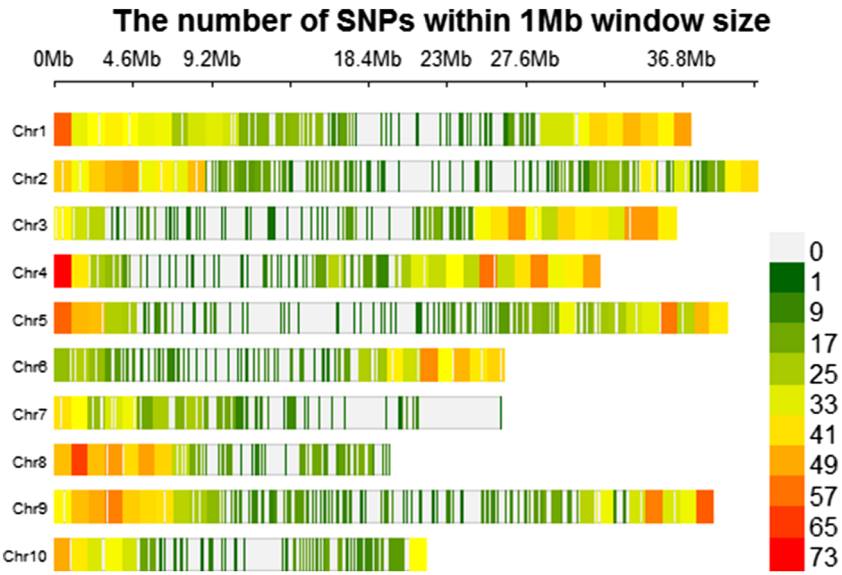

FIGURE 3 Distribution of markers along the 10 chromosomes of *Theobroma cacao*. Distribution of markers along the 10 chromosomes of *T. cacao*. The graph shows the distribution of markers along the 10 chromosomes. The density is calculated on a 1-Mb window. The areas without markers are shown in grey. The weakly marked areas are in green and the strongly marked areas are in red. A colour gradient between green and red represents the marking gradient. Chr: chromosome, SNP: single nucleotide polymorphism

Nonan-2-one, 2,3,5-trimethylpyrazine, trimethyl-oxazole, pentyl acetate, 2,3,5,6-tetramethylpyrazine, hexyl acetate, 2-pentylfuran, butyl butanoate, diethyl butanedioate, ethyl 2-hydroxy-4-methylpentanoate (known to have a fresh blackberry note) and 1,2-propanediol diacetate (known to have a fruity acetic note) seems to be more related to fruit berry notes.

TABLE 1 Co-locations between traits related to floral notes detected in *Theobroma cacao*

Chromosome	Position (bp)	Traits
1	30,025,697–30,625,697	Cis-beta-ocimene
	30,129,690–30,729,690	Dihydromyrcen
5	0–799,633	Dihydromyrcen
	0–659,385	1,2-dihydrolinalool
	0–574,697	Allo ocimen
5	33,897,871–35,986,881	Floral Orange blossom note
	34,315,551–34,915,551	Cis-beta-Ocimen
6	21,655,681–22,484,229	Dihydromyrcen
	22,391,993–22,991,993	Benzyl acetate (Phenylmethyl acetate)
7	3,129,126–3,729,126	1,2-dihydrolinalool
	3,722,709–4,322,709	Benzyl acetate (Phenylmethyl acetate)
9	93,947–693,947	Floral Orange blossom note o-xylene
9	29,047,112–29,647,112	Allo ocimen
	29,194,885–29,794,885	o-xylene

Note: Bp: base pair, in red are represented the sensorial notes, in black the volatile compounds.

TABLE 2 Co-location of fruity traits detected in *Theobroma cacao*

Chromosome	Position (bp)	Traits
1	227,662–827,662	Heptan-2-ol
	227,699–827,699	Pentyl acetate
1	10,403,809–11,003,809	2-pentylfuran
	10,403,809–11,003,809	Ethyl 2-hydroxy-4-methylpentanoate
1	26,058,831–26,658,831	Diethyl butanedioate
	26,551,903–27,151,903	Trimethyl-oxazole
1	28,523,447–29,123,447	2-pentylfuran
	28,523,447–29,123,447	Ethyl 2-hydroxy-4-methylpentanoate
1	30,025,697–30,625,697	Heptan-2-ol
	30,129,690–30,729,690	2-pentylfuran
	30,143,772–30,743,772	Heptan-2-ol
1	35,179,334–35,779,334	Ethyl 2-hydroxy-4-methylpentanoate
	35,664,409–36,264,409	Diethyl butanedioate
1	36,759,829–37,359,829	Over fermented rotten fruit
	36,806,538–37,406,538	Diethyl butanedioate
2	507,611–1,107,611	Nonan-2-one
	703,576–1,303,576	Fruit Berry
2	3,329,347–3,929,347	Diethyl butanedioate
	3,329,347–3,929,347	Limonen
2	9,227,909–9,827,909	Ethyl 2-hydroxy-4-methylpentanoate
	9,688,881–10,376,829	2-pentylfuran
	9,688,881–10,376,829	Ethyl 2-hydroxy-4-methylpentanoate
2	30,661,408–31,261,408	Diethyl butanedioate
	31,061,092–31,661,092	Fruit Dark
	31,061,092–31,661,092	Hexyl acetate

(Continues)

TABLE 2 (Continued)

Chromosome	Position (bp)	Traits
2	32,066,719–34,886,549	Fruit Dark
	32,066,719–33,281,856	Hexyl acetate
2	33,157,986–33,757,986	Diethyl butanedioate
	33,254,963–33,854,963	1,2-propanediol diacetate
2	36,057,073–36,999,677	Limonen
	36,286,693–36,886,693	Fruit Dark
2	36,286,693–36,886,693	Hexyl acetate
	36,987,390–37,587,390	1,2-propanediol diacetate
	37,493,639–38,093,639	Pentyl acetate
	37,618,909–38,574,817	Hexyl acetate
	38,134,117–38,734,117	Nonan-2-one
	38,212,257–38,812,257	Diethyl butanedioate
	38,649,264–39,798,805	Hexyl acetate
	39,198,805–39,802,081	Phenylmethyl butanoate
	39,202,081–39,802,081	Hexyl acetate
	39,500,916–40,100,916	Butyl butanoate
	39,555,481–41,295,090	Diethyl butanedioate
	40,591,750–41,191,750	Heptan-2-ol
3	1,367,283–1,967,283	2,3,5,6-tetramethylpyrazine
	1,536,416–2,136,416	Pentyl acetate
3	21,404,737–22,004,737	2-pentylfuran
	21,404,737–22,004,737	Ethyl 2-hydroxy-4-methylpentanoate
3	25,587,254–26,187,254	Nutty Nut flesh
	25,686,659–26,286,659	Nonan-2-one
3	28,711,521–29,311,521	Diethyl butanedioate
	28,876,174–29,476,174	Nutty Nut flesh
3	31,255,106–31,855,106	Ethyl 2-hydroxy-4-methylpentanoate
	31,784,376–32,384,376	3-methylbutyl 3-methylbutanoate
3	33,124,726–33,724,726	Ethyl 2-hydroxy-4-methylpentanoate
	33,459,519–34,059,519	Heptan-2-ol
4	1523–601,523	2-pentylfuran
	1523–601,523	Nutty Nut flesh
4	22,858–1,000,690	Diethyl butanedioate
	400,690–1,813,500	2,3,5-trimethylpyrazine
4	2,474,103–3,074,103	2,3,5-trimethylpyrazine
	2,858,463–3,458,463	1,2-propanediol diacetate
4	18,763,963–19,363,963	Nutty Nut flesh
	18,887,669–19,487,669	1,2-propanediol diacetate
4	21,039,263–21,639,263	Trimethyl-oxazole
	21,043,930–22,066,066	2,3,5-trimethylpyrazine
4	22,371,100–23,469,866	2,3,5-trimethylpyrazine
	23,017,920–24,748,366	Hexyl acetate
	23,512,495–24,112,495	2,3,5-trimethylpyrazine

TABLE 2 (Continued)

Chromosome	Position (bp)	Traits
4	24,903,361–25,503,361	2,3,5-trimethylpyrazine
	25,346,030–25,946,030	Trimethyl-oxazole
	25,429,071–26,029,071	Fruit Berry
	25,571,632–26,615,620	Pentyl acetate
	26,026,282–27,175,785	Trimethyl-oxazole
	26,687,962–27,716,239	2,3,5-trimethylpyrazine
	26,744,417–27,908,710	2,3,5,6-tetramethylpyrazine
	27,366,259–28,407,791	Hexyl acetate
	27,459,054–28,503,851	1,2-propanediol diacetate
4	31,091,855–31,691,855	1,2-propanediol diacetate
	31,555,304–32,155,304	2-pentylfuran
	31,555,304–32,155,304	Ethyl 2-hydroxy-4-methylpentanoate
5	0–703,249	Ethyl 2-hydroxy-4-methylpentanoate
	0–354,149	Ethyl lactate
5	2,004,453–2,604,453	2,3,5,6-tetramethylpyrazine
	2,117,256–3,198,438	Diethyl butanedioate
	2,134,166–3,049,302	2,3,5-trimethylpyrazine
	2,134,166–3,049,302	Trimethyl-oxazole
	2,176,544–2,776,544	Heptan-2-ol
5	18,471,512–19,071,512	2-pentylfuran
	18,471,512–19,071,512	Ethyl 2-hydroxy-4-methylpentanoate
5	26,290,212–26,890,212	Diethyl butanedioate
	26,366,869–26,966,869	3-methylbutyl 3-methylbutanoate
5	27,922,159–28,522,159	1,2-propanediol diacetate
	28,319,215–28,919,215	Trimethyl-oxazole
	28,986,415–29,586,415	1,2-propanediol diacetate
	29,310,859–30,773,075	2,3,5-trimethylpyrazine
5	33,394,588–33,994,588	2-pentylfuran
	33,394,588–34,288,858	Ethyl 2-hydroxy-4-methylpentanoate
	33,897,871–35,143,004	Fruit Berry
	34,315,551–34,915,551	2-pentylfuran
	34,315,551–34,915,551	Ethyl 2-hydroxy-4-methylpentanoate
5	36,031,917–36,631,917	Nutty Nut flesh
	36,215,073–36,836,724	2,3,5-trimethylpyrazine
	36,215,073–36,815,073	3-methyl-2-cyclohexen-1-one
6	6,068,335–6,668,335	Fruit Dark
	6,680,133–7,280,133	1,2-propanediol diacetate
6	19,581,377–20,181,377	Ethyl 2-hydroxy-4-methylpentanoate
	19,969,328–20,569,328	2,3,5,6-tetramethylpyrazine
6	22,931,122–23,866,379	2,3,5-trimethylpyrazine
	23,806,262–24,406,262	Nutty Nut flesh
7	0–406,706	2-pentylfuran
	0–363,832	Over fermented Rotten fruit
7	1,249,187–1,849,187	1,2-propanediol diacetate
	1,713,420–2,313,420	Nonan-2-one
	2,216,312–2,816,312	Fruit Dark

(Continues)

TABLE 2 (Continued)

Chromosome	Position (bp)	Traits
8	1,550,3271,850,3272,150,327	D-Limonen
	1,594,5621,894,5622,194,562	2-pentylfuran
	1,827,8242,127,8242,463,216	Heptan-2-ol
8	2,903,311–3,503,311	2-pentylfuran
	2,943,415–3,543,415	Pentyl acetate
	2,982,287–3,582,287	Nutty Nut flesh
8	4,073,313–4,673,313	1,2-propanediol diacetate
	4,671,183–5,271,183	Heptan-2-ol
8	5,430,993–6,030,993	1,2-propanediol diacetate
	5,600,260–6,200,260	Nutty Nut flesh
8	7,259,551–7,859,551	Nutty Nut flesh
	7,853,804–8,453,804	1,2-propanediol diacetate
9	809,358–1,620,668	Ethyl lactate
	1,158,899–1,758,899	Trimethyl-oxazole
	1,237,448–2,322,672	Diethyl butanedioate
	1,607,334–2,207,334	Over fermented Rotten fruit
	1,641,967–2,241,967	Heptan-2-ol
9	3,371,584–3,971,584	Heptan-2-ol
	3,652,480–4,252,480	Diethyl butanedioate
	3,995,514–4,595,514	Pentyl acetate
	4,034,196–4,634,196	Trimethyl-oxazole
	4,254,626–4,854,626	2,3,5-trimethylpyrazine
	4,520,440–5,120,440	Diethyl butanedioate
	4,646,739–5,246,739	Nutty Nut flesh
	4,690,282–5,290,282	Nonan-2-one
	5,184,281–6,071,002	Diethyl butanedioate
	5,298,651–6,284,926	Trimethyl-oxazole
	5,316,557–5,916,557	Over fermented Rotten fruit
	5,985,636–6,585,644	Ethyl 2-hydroxy-4-methylpentanoate
	6,094,380–6,694,380	Diethyl butanedioate
	6,414,969–7,014,974	Nutty Nut flesh
	6,840,479–7,440,479	Diethyl butanedioate
9	8,038,805–8,638,805	Nutty Nut flesh
	8,603,894–9,961,696	Trimethyl-oxazole
	9,660,571–10,260,571	1,2-propanediol diacetate
9	29,047,112–29,647,112	2-pentylfuran
	29,047,112–29,647,112	Nutty Nut flesh
9	33,599,131–34,199,131	3-methylbutyl 3-methylbutanoate
	34,157,432–34,757,432	Trimethyl-oxazole
9	37,706,972–38,306,972	Pentyl acetate
	37,734,657–38,334,657	1,2-propanediol diacetate
	37,850,408–38,890,668	Pentyl acetate
	38,213,689–38,813,689	Nutty Nut flesh
10	0–426,218	Nutty Nut flesh
	0–426,235	Total Nutty
10	769,535–1,417,085	2-pentylfuran
	769,535–1,417,085	Ethyl 2-hydroxy-4-methylpentanoate
	769,535–1,369,535	Over fermented Rotten fruit

TABLE 2 (Continued)

Chromosome	Position (bp)	Traits
10	17,784,779–18,384,779	Trimethyl-oxazole
	18,251,843–18,851,843	Ethyl 2-hydroxy-4-methylpentanoate
10	18,945,761–19,545,761	Trimethyl-oxazole
	19,225,961–19,825,961	Butyl butanoate
	19,405,234–20,005,234	Diethyl butanedioate
	19,652,912–20,252,912	Fruit Berry

Note: Bp: base pair, in red are represented the sensorial notes, in black the volatile compounds.

TABLE 3 Co-location between traits related to green notes detected in *Theobroma cacao*

Chromosome	Position (bp)	Traits
2	40,286,532–0,886,532	2-methyl-but-3-en-2-ol
	40,871,868–1,471,868	Trans-beta-ocimen
3	28,198,827–29,615,250	2-methyl-but-3-en-2-ol
	28,876,174–9,476,174	(Z)-2-hexen-1-ol acetate

Note: Bp: base pair.

Diethyl butanedioate, hexyl acetate, limonen, 1,2-propanediol diacetate and nonan-2-one seems to be more related to fruit dark notes.

The note nutty nut flesh seems to be perceived, thanks to the mixture of three types of compounds: compounds having fresh fruit notes (1,2-propanediol diacetate, ethyl 2-hydroxy-4-methylpentanoate, heptan-2-ol), compounds having a fruity and green note (nonan-2-one, 2-pentylfuran, pentyl acetate, diethyl butanedioate) and compounds having a nutty note (2,3,5-trimethylpyrazine, 3-methyl-2-cyclohexen-1-one, trimethyl oxazole).

3.3.3 | Areas of association detected in relation to vegetal notes

A total of 30 association areas were detected for the volatile compounds involved in vegetal notes (Figures S5–S7). Areas of association were detected on all chromosomes (Dataset S4). Two areas of co-localization were detected between different volatile compounds (Table 3).

3.3.4 | Areas of association detected in relation to woody and spicy notes

A total of 76 association areas were detected in relation to volatile compounds known to have a woody and/or spicy note and 10 linked to the woody and/or spicy notes detected by sensory analysis (Figures S4–S7). Areas of association were detected on all chromosomes (Dataset S4). Eleven areas of co-location were detected between different volatile compounds and between volatile compounds and sensorial notes (Table 4). Camphene and alpha-terpinene seem to be more related to the presence of spice tobacco notes.

TABLE 4 Co-location between traits related to spicy and woody notes detected in *Theobroma cacao*

Chromosome	Position (bp)	Traits
1	10,403,809–11,003,809	Beta-Myrcen
	10,635,523–11,235,523	Camphene
2	34,930,849–35,530,849	Spice Tobacco
	34,930,857–35,530,857	Camphene
2	37,724,506–38,324,511	Alpha-Terpinen
	38,143,664–39,990,777	Camphene
4	0–588,739	Alpha-Terpinen
	22,858–654,737	Camphene
5	29,291,725–29,891,725	Spice Tobacco
	29,626,890–30,226,890	Camphene
7	2,857,538–4,851,587	Camphene
	3,178,085–3,778,085	Spice Tobacco
8	5,024,136–6,200,257	Camphene
	5,600,244–6,857,883	Spice Tobacco
9	4,254,626–5,120,440	Camphene
	4,498,201–5,098,201	Spice Tobacco
9	26,640,576–27,240,576	Camphene
	27,120,722–27,810,582	Spice Tobacco
	27,120,724–27,720,724	Alpha-Terpinen
10	754,940–1,354,940	Spice Tobacco
	842,194–1,442,194	Camphene
10	3,348,574–4,171,361	Camphene
	3,679,144–4,438,496	Total Spice

Note: Bp: base pair, in red are represented the sensorial notes, in black the volatile compounds.

3.3.5 | Areas of association detected in relation to empyreumatic notes

We could identify 66 association zones related to volatile compounds known to have an empyreumatic note (caramel, brown sugar,

TABLE 5 Co-location between traits related to empyreumatic notes detected in *Theobroma cacao*

Chromosome	Position (bp)	Traits
1	5,251,864–5,851,864	Roast Degree
	5,398,258–6,019,937	3-methyl-2-(2-methyl-2-butenyl)-furan
1	7,817,873–8,477,387	Cocoa
	7,817,873–7,748,797	Roast Degree
	8,205,326–8,805,326	Sweet, caramel
1	14,322,116–14,922,116	Sweet, caramel
	14,325,744–14,925,744	2,3-dimethyl-5-ethylpyrazine
1	26,056,034–26,656,034	Sweet, caramel
	26,551,903–27,152,249	2,3-dimethyl-5-ethylpyrazine
1	27,313,384–27,913,384	2,3-dimethyl-5-ethylpyrazine
	27,622,563–28,222,563	2,5-dimethyl-pyrazine
1	28,594,210–29,194,210	1-methylhexyl acetate
	28,865,580–29,465,580	2,5-dimethyl-pyrazine
3	2,502,549–3,145,200	1-methylhexyl acetate
	2,551,310–3,151,310	2,3-dimethyl-5-ethylpyrazine
3	12,353,462–12,953,564	1-methylhexyl acetate
	12,353,462–12,953,462	Roast Degree
3	26,244,291–26,844,291	2,3-dimethyl-5-ethylpyrazine
	26,406,759–27,006,759	Sweet, caramel
	26,656,014–27,256,014	1-methylhexyl acetate
	26,748,027–27,348,027	1-hydroxypropan-2-one
	26,883,974–27,483,974	2,3-dimethyl-5-ethylpyrazine
3	28,145,339–28,745,339	Cocoa
	28,145,339–28,745,339	Roast Degree
4	18,763,963–19,471,045	Roast Degree
	18,871,045–19,471,045	Cocoa
5	515,972815,9721,115,972	1-methylhexyl acetate
	1,281,6291,581,6291,881,629	Sweet, caramel
5	2,569,5752,869,5753,296,902	Cocoa
	2,569,5752,869,5753,296,902	Roast Degree
5	34,094,673–34,694,673	Cocoa
	34,094,673–34,694,673	Roast Degree
6	19,179,147–19,926,195	Roast Degree
	19,581,377–20,181,377	3-methyl-2-(2-methyl-2-butenyl)-furan
6	23,199,939–23,799,955	Roast Degree
	23,454,660–24,406,262	Sweet, caramel
	23,928,498–24,659,115	1-methylhexyl acetate
8	3,449,438–4,049,438	Sweet, caramel
	4,010,117–4,610,117	2,5-dimethyl-pyrazine
	4,040,269–4,640,329	1-methylhexyl acetate
8	4,671,607–5,638,343	1-methylhexyl acetate
	5,523,513–6,123,513	2,5-dimethyl-pyrazine
	5,963,281–6,563,290	1-methylhexyl acetate
10	4,311,469–5,528,191	Roast Degree
	4,928,081–5,528,191	Cocoa
10	13,049,399–13,649,399	Roast Degree
	13,049,399–13,649,399	Cocoa

Note: Bp: base pair, in red are represented the sensorial notes, in black the volatile compounds.

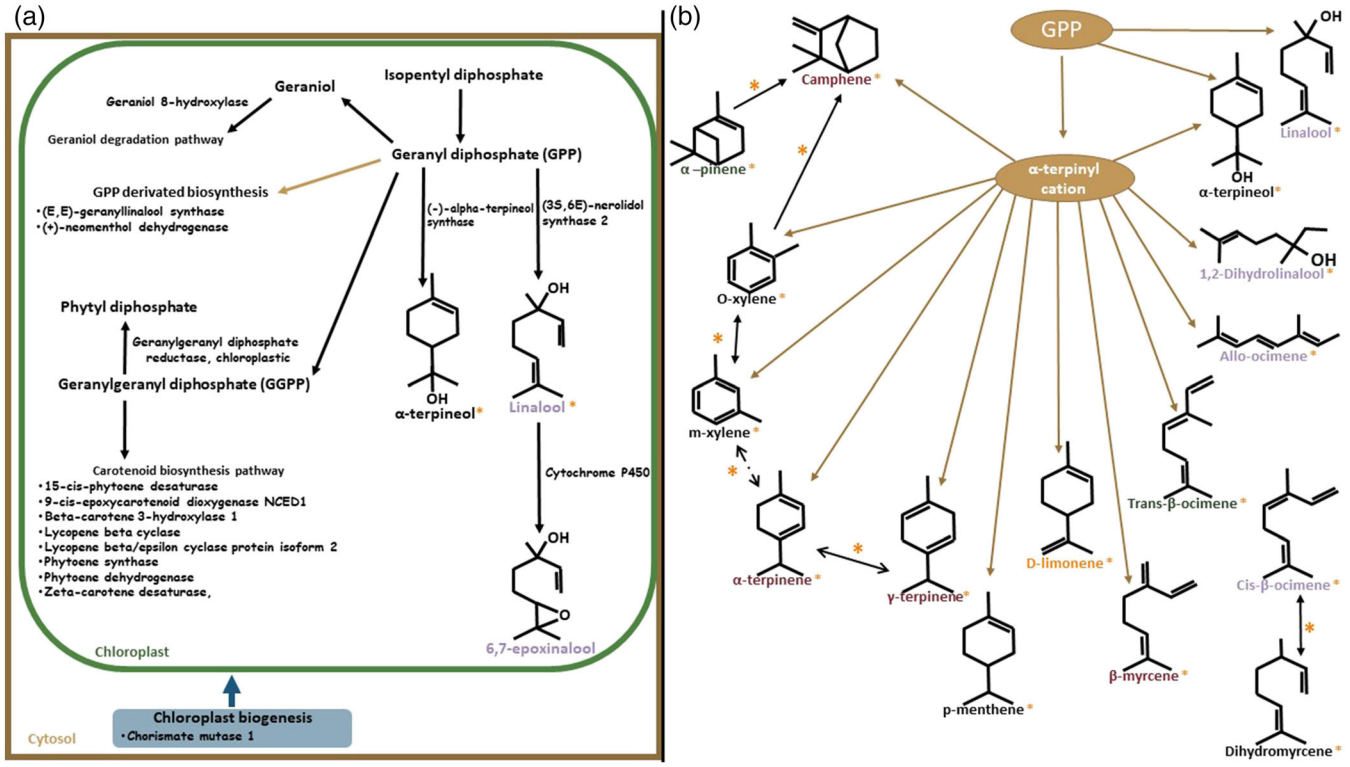

FIGURE 4 Schematic representation of the monoterpene biosynthetic pathway in *Theobroma cacao* adapted to Bohlmann et al. (1998). All compounds marked with an orange star were identified in this study. The volatile compounds are indicated in colour according to their taste: in purple for floral note, in orange for fruity note, in red for spicy or woody note, in green for vegetal and/or herbaceous note, in black are not known to be aromatic. The black arrows represent chemical transformations identified in other organisms. (a) Candidate genes, identified in areas of association with monoterpenes, are shown in black according to their known level of involvement in the monoterpene biosynthetic pathway in other crop plants. (b) Hypothetical pathway in the cocoa tree according to the co-location of association zones. The brown arrows show the hypothetical part of the biosynthetic pathway according to the identified co-locations taking into account that the co-locations are due to a common precursor. The black arrows with an orange star show the hypothetical part of the biosynthetic pathway according to the identified co-locations taking into account that the co-locations between the two compounds are because one is a precursor of the other. GPP: Geranyl diphosphate

roasted...) and 43 linked to empyreumatic notes detected by sensory analysis (Figures S4–S7). Areas of association were detected on all chromosomes (Dataset S4). Twenty areas of co-locations were observed (Table 5). The 2,3-dimethyl-5-ethylpyrazine, 2,5-dimethylpyrazine and 1-methylhexyl acetate seem to be more related to the presence of sweet, caramel notes. The 3-methyl-2-(2-methyl-2-butenyl)-furan and 1-methylhexyl acetate seems to be more related to the roast degree.

3.4 | Identification of candidate genes involved in the different biosynthetic pathways of aromatic compounds detected in the Amazonian population

In each association zone, a search was undertaken for candidate genes involved in the synthesis of the compounds in the association. Candidate genes involved in general plant defences were also observed. The biosynthetic pathways previously described in other plants appear to be valid in cocoa. Schemes summarising these biosynthetic pathways that may be present in cocoa and the involvement of candidate genes in these pathways have been produced. The 1824 candidate genes identified are listed in Dataset S6.

Of the candidate genes detected, 53 are in common with previous GWAS studies on the aroma traits of the modern Nacional population (Dataset S6) (Colonges, Jimenez, Saltos, Seguine, Loor Solorzano, Fouet, Argout, Assemat, Davrieux, Cros, et al., 2021a; Colonges, Jimenez, Saltos, Seguine, Solorzano, Fouet, Argout, Assemat, Davrieux, Cros, & Lanaud, 2021b). The repetition of detection of these candidate genes between two studies on two different populations brings additional weight to the role of these candidate genes.

3.4.1 | Candidate genes involved in the monoterpene biosynthetic pathway

In the genomic regions associated with monoterpenes, candidate genes encoding enzymes involved in their biosynthesis have been identified.

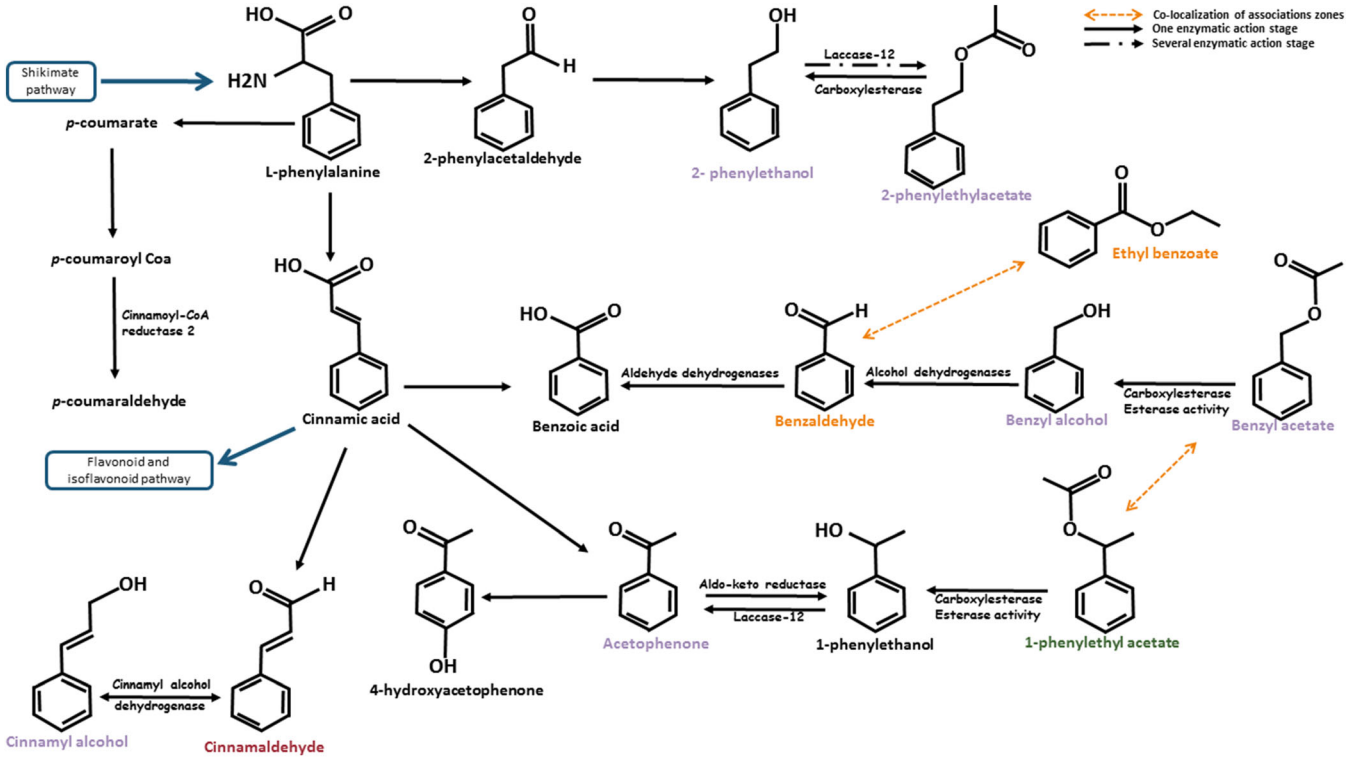

FIGURE 5 Schematic representing the L-phenylalanine degradation pathway according to Lapadatescu et al. (2000) and the hypothetical L-phenylalanine degradation pathway in cocoa. The volatile compounds are indicated in colour according to their already known taste: in purple for floral note, in orange for fruity note, in red for spicy or woody note, in green for vegetal and/or herbaceous note, in black are not known to be aromatic. The black arrows represent chemical transformations identified in other organisms. The orange dotted arrows represent co-locations between the compounds and thus a probable link in the biosynthetic pathway. Candidate genes, identified in the association zones related to compounds included in the L-phenylalanine degradation pathway, are shown in black at their known level of involvement in the monoterpene biosynthetic pathway according to Lapadatescu et al. (2000). In blue, the nearby biosynthetic pathways are shown

The role of each candidate gene identified as being involved in terpene biosynthesis is illustrated in Figure 4a. Co-locations between the association zones of different terpenes can be due to two possibilities either the compounds have the same precursor (like the alpha-terpinyl cation present in *Mentha spicata*) (Srividya et al., 2015), or they follow each other in the biosynthetic pathway (one is necessary to produce the other). Thanks to these possibilities, the hypothesis of a common precursor between camphene, o-xylene, m-xylene, α-terpinene, γ-terpinene, p-menthene, d-limonene, β-myrcene, trans-β-ocimene, allo-ocimene, 1,2-dihydrolinalool and α-terpineol were made (Figure 4b). The other co-locations allowed the hypothesis that some transformations of terpenes into another (such as the transformation of α-pinene into camphene or of o-xylene into m-xylene) already known in the literature are also true in cocoa (Figure 4b).

3.4.2 | Candidate genes involved in the L-phenylalanine degradation pathway

In the areas of the genome associated with the compounds of the L-phenylalanine degradation pathway, candidate genes encoding enzymes involved in their biosynthesis have been identified. The role of each candidate gene identified as being involved in the L-phenylalanine degradation pathway is illustrated in Figure 5.

3.4.3 | Candidate genes involved in the biosynthesis of Maillard reaction precursors

In the areas of the genome associated with pyrazines and furans, which are compounds primarily derived from the Maillard reaction, candidate genes encoding enzymes involved in the biosynthesis of their precursors as well as their biosynthesis have been identified. The role of each candidate gene identified is illustrated in Figure 6.

3.4.4 | Candidate genes involved in the degradation pathways of fatty acids and sugars

In the areas of the genome associated with compounds involved in the fatty acid and/or sugar degradation pathway, candidate genes encoding enzymes involved in their biosynthesis were identified. The role of each identified candidate gene is illustrated in Figure 7.

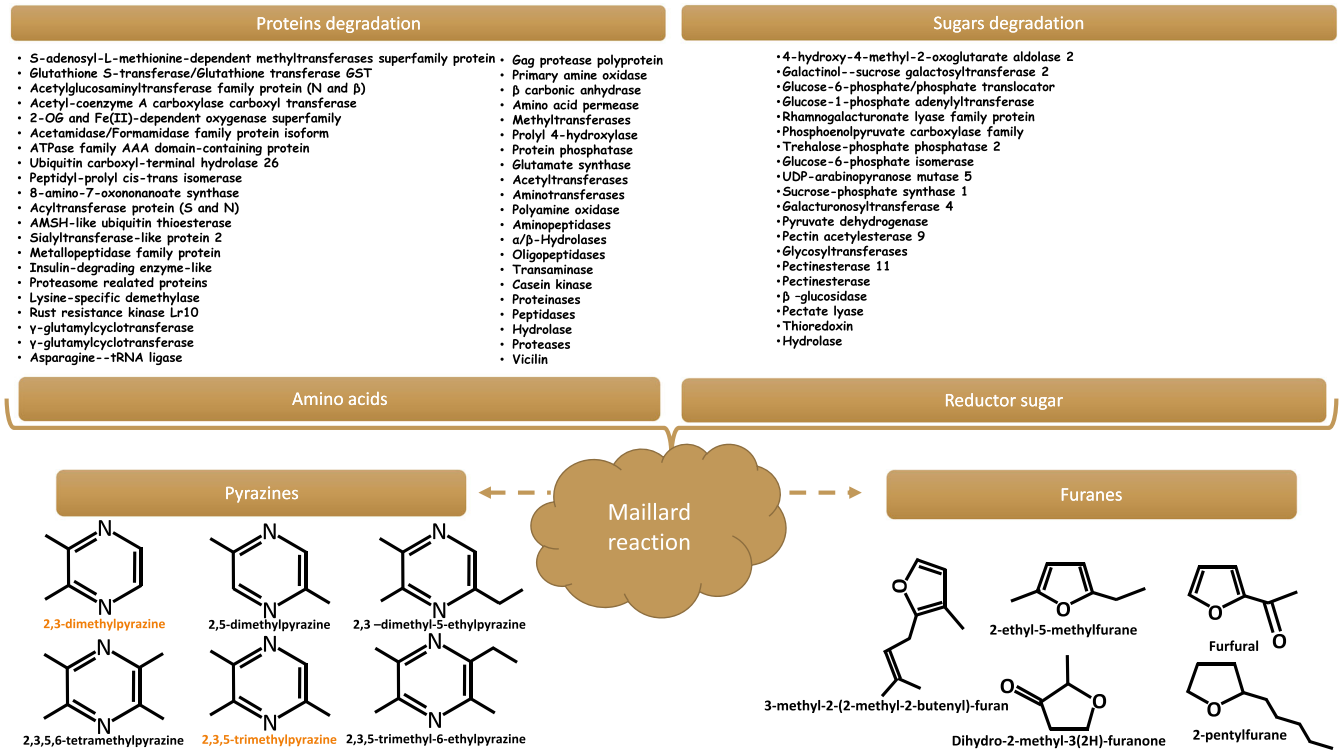

FIGURE 6 Diagram representing the hypothetical biosynthetic pathway of pyrazines and furans in *Theobroma cacao*. The volatile compounds are indicated in colour according to their taste: in orange for fruity note and in black are not known to be aromatic. The orange dotted arrows represent co-locations between the compounds and thus a probable link in the biosynthetic pathway. Candidate genes, identified in the association areas with pyrazines and furans, are shown in black on the left of the figure. 2-OG: 2-oxoglutarate

3.4.5 | Candidate genes involved in general plant defences

Plants synthesise volatile compounds for various purposes including defence (Ponzio et al., 2013). The molecular mechanisms governing these hormonal networks are largely unknown but recent studies have shown their implications in defence responses (Denancé et al., 2013), which is why we have selected, as candidate genes, the genes listed in Table 6. Other candidate genes were found in areas of association. Their function is illustrated in Figure 8, and all genes are included in the Dataset S6.

4 | DISCUSSION

The studied population was collected in the putative area of origin of the ancestral Nacional variety. The collection of these new genotypes was done to safeguard the genetic diversity of native cocoa trees from this Amazonian region and to enlarge the genetic resources available to improve or create new aromatic varieties. Indeed, until now, Ecuadorian aromatic cocoa trees are mainly composed by the modern Nacional variety. The modern Nacional variety is a hybrid population with a narrow genetic base involving a limited number of ancestral Nacional genotypes hybridised with introduced and closely related Trinitario genotypes (Loor et al., 2015).

Our results showed that a larger variability can be observed at the genetic level and at the aromatic level, with common as well as new aromatic volatile compounds identified and segregating in this population. The cocoa trees studied in this paper, widespread in the South of Ecuadorian Amazonia, have a different genetic background than the modern Nacional variety. They are the result of natural evolution from thousands of years giving to this population a low LD favourable to a better precision of localization of associations than the study carried out on the modern Nacional population (for which the hybridizations between the three contrasting ancestors are recent) (Bartley, 2005; Colonges, Jimenez, Saltos, Seguine, Loor Solorzano, Fouet, Argout, Assemat, Davrieux, Cros, et al., 2021a; Loor, 2007). The associations detected are therefore more precise and reliable. Moreover, this population has a higher genetic diversity, increasing the number of marker or gene segregations.

The biosynthetic pathways involved in the synthesis of floral notes are the same as those previously described in the modern Nacional population: the monoterpene biosynthetic pathway and the L-phenylalanine degradation pathway (Colonges, Jimenez, Saltos, Seguine, Loor Solorzano, Fouet, Argout, Assemat, Davrieux, Cros, et al., 2021a). Common and new compounds have been identified as being involved in the variety of floral notes. Associations with five new monoterpenes known to have floral note were identified in this study ((-)-dihydromyrcen, 1,2-dihydrolinalool, allo-ocimene, cis-β-ocimene, dihydromyrcene). A new association with ethylbenzene

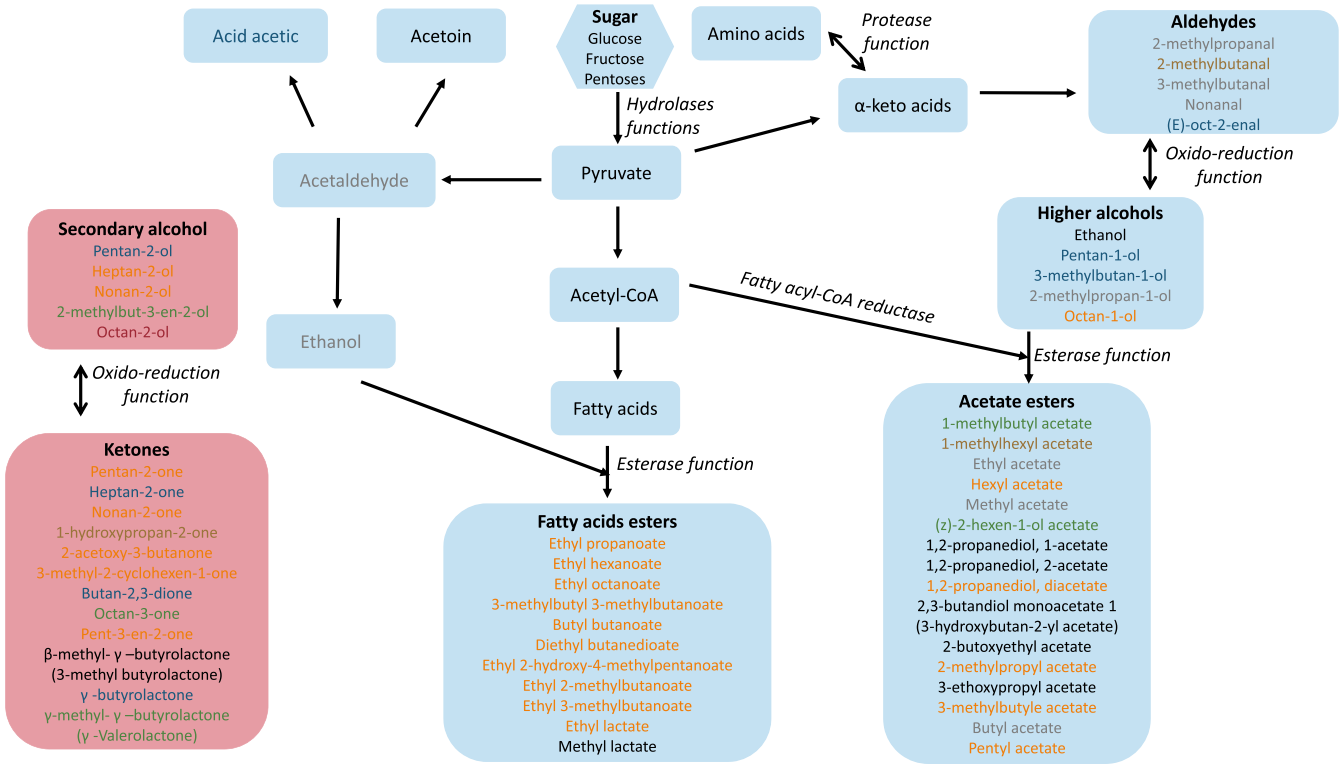

FIGURE 7 Schematic representation of the fatty acid and sugar degradation pathway according to Swiegers et al. (2005) and Dzialo et al. (2017) and hypothetical pathway in *Theobroma cacao*. The volatile compounds are indicated in colour according to their taste: in orange for fruity note, in red for spicy or woody note, in green for vegetal and/or herbaceous note, in blue for cheesy notes, in grey for chemical notes, in brown for chocolate and brown notes, in black are not known to be aromatic. The black arrows represent chemical transformations identified in other organisms. The function of candidate genes, identified in areas of association with compounds involved in the fatty acid and sugar degradation are shown in black italics

TABLE 6 Candidate genes related to growth phytohormones detected in *Theobroma cacao*

Auxine related	ABA related	Gibbereline related
Protein AUXIN SIGNALING F-BOX 2		
Auxin-induced protein	Abscisic acid receptor PYL6	Gibberellin 2-beta-dioxygenase 8
Auxin transport protein BIG	Abscisic acid receptor PYL2	Gibberellin 3-beta-dioxygenase 4
Auxin response factor	Major allergen Pru ar 1	Gibberellin 2-beta-dioxygenase 2
Auxin-responsive protein	Abscisic acid 8′-hydroxylase 4	Dihydroflavonol-4-reductase
Indole-3-acetic acid-amido synthetase GH3.1	Protein phosphatase 2C 16	DELLA protein GAI
Indole-3-acetic acid-induced protein ARG7		
Protein-tyrosine-phosphatase IBR5		

Note: ABA: abscissic acid.

acetate (compound involved in L-phenylalanine degradation) was identified. These new results complement the likely biosynthetic pathways used in cocoa for the synthesis of floral notes.

Associations with new compounds known to have a fruity note were also detected. Two compounds seem to be involved in fresh fruit notes: D-limonene (a monoterpene) and 2-acetoxy-3-butanone (a ketone probably synthetized during fatty acid or sugar degradation). Two other compounds seem to be involved in dried fruit notes: 2-pentylfuran and trimethyl-oxazole (probably synthetized during the Maillard reaction).

Areas of association were identified with five new compounds known to have a vegetal note. There are two monoterpenes: α-pinene and trans-β-ocimene, three compounds probably synthetized during fatty acid or sugar degradation: (Z)-2-hexen-1-ol acetate, 2-methyl-but-3-en-2-ol and octan-3-one. These compounds belong to the same biosynthetic pathways as the key compounds previously identified in

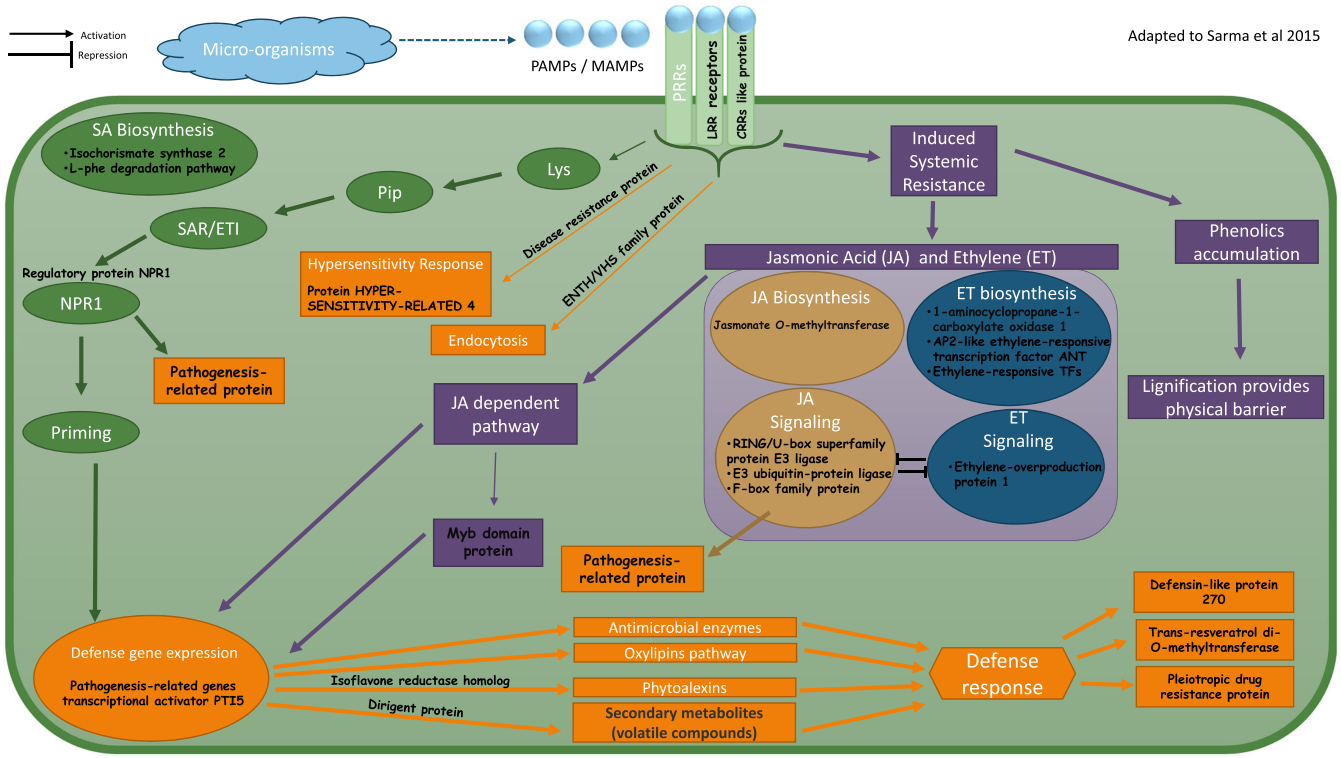

FIGURE 8 Diagram of general plant defences adapted from Sarma et al. (2015). In green are represented the mechanisms associated with the systemic acquired resistance response. In purple are the mechanisms associated with the induced systemic resistance response. In blue are represented the mechanisms associated with ethylene. In brown are represented mechanisms related to jasmonic acid. In orange are represented the general mechanisms. The candidate genes, identified in the association areas linked to volatile compounds, are shown in black at the location of their action. CRR: Cysteine-Rich Receptors; GA: Gibberellic acid (gibberellin); LRR: Leucine-Rich Repeat; MAMPs: Microbe-Associated Molecular Patterns; PAMPs: Pathogens-Associated Molecular Patterns; PRR: Pattern Recognition Receptors; SA: Salicylic Acid, SAR: Systemic Acquired Resistance; TFs: Transcription Factors. In black are represented the candidate genes identified in the area of associations

the synthesis of fruity and floral aromas in the modern Nacional variety (Colonges, Jimenez, Saltos, Seguine, Loor Solorzano, Fouet, Argout, Assemat, Davrieux, Cros, et al., 2021a; Colonges, Jimenez, Saltos, Seguine, Solorzano, Fouet, Argout, Assemat, Davrieux, Morillo, et al., 2021c).

The biosynthesis of all these compounds could therefore be at the origin of new aroma notes in this population. The future confirmation of the functional expression of candidate genes involved in the biosynthesis of these compounds would make it possible to identify diagnostic markers in these genes. These markers could then be used by breeders according to the expectations of chocolate makers or consumers.

The results concerning woody and spicy notes are very different from those observed in the study of the modern Nacional population. This population possesses biochemical compounds that give woody and spicy notes that the modern Nacional does not have, such as the presence of β-myrcene, α-terpinene, or camphene. A combination of the favourable alleles identified in the modern Nacional and those of this wild population for the synthesis of compounds involved in woody and spicy notes could give a unique aroma.

Associations with new compounds known to have an empyreumatic note were also detected. This is the case for the compounds: 1-hydroxypropan-2-one, 2,3-dimethyl-5-ethylpyrazine (not detected in fermented and dried beans), 2,5-dimethyl-pyrazine, 3-methyl-2-(2-methyl-2-butenyl)-furan.

Fifty-three candidate genes detected are common between both modern Nacional and Amazonian populations, which strongly increases the probability that these genes are involved in cocoa flavour synthesis. Furthermore, as the population in this study is a native cocoa population from Amazonia, resulting in low LD, the areas of association detected in this study are smaller and therefore more accurate.

During fermentation, cocoa beans may react to biotic stress, as previously suggested (Colonges, Jimenez, Saltos, Seguine, Loor Solorzano, Fouet, Argout, Assemat, Davrieux, Cros, et al., 2021a; Sabau et al., 2006), and synthesise volatile compounds for defence against micro-organisms involved in fermentation, as previously shown for example in maize, plants synthesise volatile compounds to defend themselves (Block et al., 2019). The genomic areas associated to volatile compounds identified in our study included a large number of genes involved in plant defence pathways. These genes could be at the origin of the synthesis of volatile aromatic compounds (Figure 8). Some candidate genes present in the various association zones are related to phytohormones. Plant hormones interact in complex networks to balance the response to environmental and developmental

signals and thus limit the adaptation costs associated with defence (Denancé et al., 2013). During fermentation, cocoa beans are also subjected to abiotic stresses such as increased temperature around 45°C or increased acidity and thus a decrease in pH (down to 4.75 for well-fermented beans) (Afoakwa et al., 2008). It has also been reported in the literature that abiotic stresses can cause plants to produce volatile compounds (Baldwin, 2010; Dudareva et al., 2006; Vickers et al., 2009). Volatile isoprenoids have been shown to play an important role in protecting against a variety of abiotic stresses including temperature changes (Vickers et al., 2009). As tea plants release volatile compounds such as linalool after cold stress (Zhao et al., 2020), cocoa beans could release volatile compounds in response to temperature stress.

All these new results allow us to broaden our knowledge on the palette of aromatic notes synthesised by the cocoa tree itself and not by the micro-organisms present during fermentation. Even if the aromatic profile of the final chocolate is influenced by the fermentation environment (methods, micro-organisms present, …) as well as by the processing protocols such as roasting (Assi-Clair et al., 2019; Kongor et al., 2016; Owusu et al., 2012), thanks to GWAS we were able to highlight the part of aroma variations depending on the genotype of the cocoa tree. Some volatile aromatic compounds and sensory notes identified were not associated with the genomic variation of cocoa. They certainly constitute the part brought by the 'global' terroir of the cocoa tree (place of cultivation, post-harvest protocols, microflora present throughout the processing, …).

This population shows a large number of aroma traits favourable to the search of flavours and economic niches sought by chocolate makers and potentially appreciated by consumers. It, therefore, constitutes germplasm of greatest interest for plant breeders.

However, this population also shows several unfavourable flavours, and as for all wild plants, domestication of the favourable traits will be necessary to select the desired parts of the genome. The compounds involved in the desired aromas do not appear to be correlated with the compounds responsible for unpleasant aromas. Indeed, no or very low correlations have been observed between these compounds and few areas of association co-localise. The selection of the compounds responsible for the desired aromas combined with a counter-selection of the compounds responsible for the unpleasant aromas seems therefore possible.

This population offers to breeders a wider range of aromas to select for new aromatic niches. The combination of these genotypes between them or with other aromatic varieties could be also promising for the production of new original aromatic varieties adapted to the different Ecuadorian climates.

ACKNOWLEDGEMENTS

We thank the I-Site MUSE, Valrhona and the U.S. Department of Agriculture (USDA-ARS), for their financial support of this project. This work, part of the MUSE Amazcacao project, was publicly funded through ANR (the French National Research Agency) under the 'Investissement d'avenir' programme with the reference ANR-16-IDEX-0006.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

CL, RGLS, conceived the experiment; KC, MCL, ML conducted biochemical analyses; KC, JCJ, DC, CS, IS, FF conducted microfermentation, ES carried out sensorial analyses; KC, OF carried out DNA experiments; KC, BR, XA, PC, RB, CL, analysed data; KC, RB, CL wrote the manuscript.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in the database tropgene at http://tropgenedb.cirad.fr/tropgene/JSP/interface.jsp?module=COCOA, at the study named 'Cocoa_Amazonie_Ecuador_aroma'.

ORCID

Kelly Colonges https://orcid.org/0000-0001-9944-1399
Olivier Fouet https://orcid.org/0000-0001-8547-1474

REFERENCES

Afoakwa, E. O., Paterson, A., Fowler, M., & Ryan, A. (2008). Flavor formation and character in cocoa and chocolate: A critical review. *Critical Reviews in Food Science and Nutrition*, *48*(9), 840–857. https://doi.org/10.1080/10408390701719272

Andújar, I., Recio, M. C., Giner, R. M., & Ríos, J. L. (2012). Cocoa polyphenols and their potential benefits for human health. *Oxidative Medicine and Cellular Longevity*, *2012*, e906252. https://doi.org/10.1155/2012/906252

Argout, X., Martin, G., Droc, G., Fouet, O., Labadie, K., Rivals, E., Aury, J. M., & Lanaud, C. (2017). The cacao Criollo genome v2.0: An improved version of the genome for genetic and functional genomic studies. *BMC Genomics*, *18*, 730. https://doi.org/10.1186/s12864-017-4120-9

Argout, X., Salse, J., Aury, J.-M., Guiltinan, M. J., Droc, G., Gouzy, J., Allegre, M., Chaparro, C., Legavre, T., Maximova, S. N., Abrouk, M., Murat, F., Fouet, O., Poulain, J., Ruiz, M., Roguet, Y., Rodier-Goud, M., Barbosa-Neto, J. F., Sabot, F., … Lanaud, C. (2011). The genome of *Theobroma cacao*. *Nature Genetics*, *43*, 101–108. https://doi.org/10.1038/ng.736

Assi-Clair, B. J., Koné, M. K., Kouamé, K., Lahon, M. C., Berthiot, L., Durand, N., Lebrun, M., Julien-Ortiz, A., Maraval, I., Boulanger, R., & Guéhi, T. S. (2019). Effect of aroma potential of Saccharomyces cerevisiae fermentation on the volatile profile of raw cocoa and sensory attributes of chocolate produced thereof. *European Food Research and Technology*, *245*(7), 1459–1471. https://doi.org/10.1007/s00217-018-3181-6

Baldwin, I. T. (2010). Plant volatiles. *Current Biology*, *20*(9), R392–R397. https://doi.org/10.1016/j.cub.2010.02.052

Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics*, *21*, 263–265. https://doi.org/10.1093/bioinformatics/bth457

Bartley, B. G. D. (2005). *The genetic diversity of cacao and its utilization*. CABI Publishing. https://doi.org/10.1079/9780851996196.0000

Bayer, C., & Kubitzki, K. (2003). Malvaceae. Fam. In *Genera Vasc. Plants Dicotyledons Malvales Capparales non-Betalain Caryophyllales* (pp. 225–311). Springer. https://doi.org/10.1007/978-3-662-07255-4_28

Block, A. K., Vaughan, M. M., Schmelz, E. A., & Christensen, S. A. (2019). Biosynthesis and function of terpenoid defense compounds in maize (Zea mays). *Planta*, *249*(1), 21–30. https://doi.org/10.1007/s00425-018-2999-2

Bohlmann, J., Meyer-Gauen, G., & Croteau, R. (1998). Plant terpenoid synthases: Molecular biology and phylogenetic analysis. *Proceedings of the National Academy of Sciences*, *95*(8), 4126–4133. https://doi.org/10.1073/pnas.95.8.4126

Cheesman, E. (1944). Notes on the nomenclature, classification and possible relationships of cacao populations. *Tropical Agriculture*, *21*, 144–159.

Colonges, K., Jimenez, J. C., Saltos, A., Seguine, E., Loor Solorzano, R. G., Fouet, O., Argout, X., Assemat, S., Davrieux, F., Cros, E., Boulanger, R., & Lanaud, C. (2021a). Two main biosynthesis pathways involved in the synthesis of the floral aroma of the Nacional cocoa variety. *Frontiers in Plant Science*, *12*, 2064. https://doi.org/10.3389/fpls.2021.681979

Colonges, K., Jimenez, J. C., Saltos, A., Seguine, E., Solorzano, R. L., Fouet, O., Argout, X., Assemat, S., Davrieux, F., Cros, E., & Lanaud, C. (2021b). Integration of GWAS, metabolomics, and sensorial analyses to reveal novel metabolic pathways involved in cocoa fruity aroma. *Plant Physiology and Biochemistry*. Accepted, *171*, 213–225. https://doi.org/10.1016/j.plaphy.2021.11.006

Colonges, K., Jimenez, J. C., Saltos, A., Seguine, E., Solorzano, R. L., Fouet, O., Argout, X., Assemat, S., Davrieux, F., Morillo, E., Boulanger, R., Cros, E., & Lanaud, C. (2021c). Genetic bases of fruity notes (fresh and dried) of the Nacional cocoa variety. E Guichard JL Quéré Eds Proc 16th Weurman Flavour Res Symp 2021. https://doi.org/10.5281/zenodo.5046157

Cook, L. R., & Meursing, E. H. (1982). *Chocolate production and use*. Harcourt Brace Jovanovich.

Denancé, N., Sánchez-Vallet, A., Goffner, D., & Molina, A. (2013). Disease resistance or growth: The role of plant hormones in balancing immune responses and fitness costs. *Frontiers in Plant Science*. https://doi.org/10.3389/fpls.2013.00155

Dudareva, N., Negre, F., Nagegowda, D. A., & Orlova, I. (2006). Plant volatiles: Recent advances and future perspectives. *Critical Reviews in Plant Sciences*, *25*(5), 417–440. https://doi.org/10.1080/07352680600899973

Dzialo, M. C., Park, R., Steensels, J., Lievens, B., & Verstrepen, K. J. (2017). Physiology, ecology and industrial applications of aroma formation in yeast. *FEMS Microbiology Reviews*, *41*(Supp_1), S95–S128. https://doi.org/10.1093/femsre/fux031

Gao, X., Becker, L. C., Becker, D. M., Starmer, J. D., & Province, M. A. (2010). Avoiding the high Bonferroni penalty in genome-wide association studies. *Genetic Epidemiology*, *34*, 100–105.

Gao, X., Starmer, J., & Martin, E. R. (2008). A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genetic Epidemiology*, *32*(4), 361–369. https://doi.org/10.1002/gepi.20310

ISCQF. (2020). First draft of the Protocol for Cocoa Liquor Sensory Evaluation: Art of the International Standards for the Assessment of Cocoa Quality and Flavour (ISCQF). Compiled by Bioversity International, in collaboration with the members of the ISCQF Working Group. ISBN: 978–92–9255-158-2.

Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., Caig, V., Heller-Uszynska, K., Jaccoud, D., Hopper, C., Aschenbrenner-Kilian, M., Evers, M., Peng, K., Cayla, C., Hok, P., & Uszynski, G. (2012). Diversity Arrays Technology: A Generic Genome Profiling Technology on Open Platforms. In F. Pompanon & A. Bonin (Eds.), *Data prod. Anal. Popul. Genomics methods Protoc* (pp. 67–89). Humana Press. https://doi.org/10.1007/978-1-61779-870-2_5

Kongor, J. E., Hinneh, M., de Walle, D. V., Afoakwa, E. O., Boeckx, P., & Dewettinck, K. (2016). Factors influencing quality variation in cocoa (*Theobroma cacao*) bean flavour profile — A review. *Food Research International*, *82*, 44–52. https://doi.org/10.1016/j.foodres.2016.01.012

Lapadatescu, C., Giniès, C., Le Quéré, J. L., & Bonnarme, P. (2000). Novel scheme for biosynthesis of aryl metabolites from L-phenylalanine in the fungus *Bjerkandera adusta*. *Applied and Environmental Microbiology*, *66*(4), 1517–1522. https://doi.org/10.1128/AEM.66.4.1517-1522.2000

Loor, S. R. G. (2007). Contribution à létude de la domestication de la variété de cacaoyer Nacional dEquateur: Recherche de la variété native et de ses ancêtres sauvages. Corpus ID: 130056754

Loor, R. S. G. L., Fouet, O., Lemainque, A., Pavek, S., Boccara, M., Argout, X., Amores, F., Courtois, B., Risterucci, A. M., & Lanaud, C. (2012). Insight into the wild origin, migration and domestication history of the fine flavour Nacional *Theobroma cacao* L. variety from Ecuador. *PLoS ONE*, *7*, e48438. https://doi.org/10.1371/journal.pone.0048438

Loor, S. R. G., Lachenaud, P., Fouet, O., Argout, X., Peña, G., Macias, J. C., Puyutaxi, F. M. A., Valdez, F., Hurtado, J., & Lanaud, C. (2015). Rescue of cacao genetic resources related to the national variety: Surveys in the Ecuadorian Amazon (2010-2013). *Rev ESPAMCIENCIA* ISSN 1390-8103, *6*, 7–15.

Loor, S. R. G., Risterucci, A. M., Courtois, B., Fouet, O., Jeanneau, M., Rosenquist, E., Amores, F., Vasco, A., Medina, M., & Lanaud, C. (2009). Tracing the native ancestors of the modern *Theobroma cacao* L. population in Ecuador. *Tree Genetics & Genomes*, *5*(3), 421–433. https://doi.org/10.1007/s11295-008-0196-3

Luna, F., Crouzillat, D., Cirou, L., & Bucheli, P. (2002). Chemical composition and flavor of Ecuadorian cocoa liquor. *Journal of Agricultural and Food Chemistry*, *50*(12), 3527–3532. https://doi.org/10.1021/jf0116597

Motamayor, J. C., Lachenaud, P., Da Silva e Mota, J. W., Loor, R., Kuhn, D. N., Brown, J. S., & Schnell, R. J. (2008). Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L). *PLoS ONE*, *3*(10), e3311. https://doi.org/10.1371/journal.pone.0003311

Motamayor, J. C., Mockaitis, K., Schmutz, J., Haiminen, N., Donald Livingstone, O. C. III, Findley, S. D., Zheng, P., Utro, F., Royaert, S., Saski, C., Jenkins, J., Podicheti, R., Zhao, M., Scheffler, B. E., Stack, J. C., Feltus, F. A., Mustiga, G. M., Amores, F., Phillips, W., … Kuhn, D. N. (2013). The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. *Genome Biology*, *14*(6), r53. https://doi.org/10.1186/gb-2013-14-6-r53

Owusu, M., Petersen, M. A., & Heimdal, H. (2012). Effect of fermentation method, roasting and conching conditions on the aroma volatiles of dark chocolate. *Journal of Food Processing & Preservation*, *36*(5), 446–456. https://doi.org/10.1111/j.1745-4549.2011.00602.x

Perrier, X. & Jacquemoud-Collet, J. P. (2006). DARwin software.

Ponzio, C., Gols, R., Pieterse, C. M. J., & Dicke, M. (2013). Ecological and phytohormonal aspects of plant volatile emission in response to single and dual infestations with herbivores and phytopathogens. *Functional Ecology*, *27*(3), 587–598. https://doi.org/10.1111/1365-2435.12035

Rami J-F (2017). Spidermap v1.7.1, a free computer software (version 2017).

Risterucci, A. M., Grivet, L., NGoran, J. A. K., Pieretti, I., Flament, M. H., & Lanaud, C. (2000). A high-density linkage map of *Theobroma cacao* L. *Theoretical and Applied Genetics*, *101*(5–6), 948–955. https://doi.org/10.1007/s001220051566

Sabau, X., Loor, R. G., Boccara, M., Fouet, O., Jeanneau, M., Argout, X., Legavre, T., Risterucci, A.-M., Wincker, P., Da Silva, C., Jimenez, J. C., Cros, E., Boulanger, R., Alary, R., & Lanaud, C. (2006) Preliminary results on linalool synthase expression during seed development and fermentation of Nacional and Trinitario clones. 15th Int Cocoa Res Conf Cocoa Product Qual Profitab Hum Health Environ. Corpus ID: 89720924.

Sardos, J., Rouard, M., Hueber, Y., Cenci, A., Hyma, K. E., van den Houwe, I., Hribova, E., Courtois, B., & Roux, N. (2016). A genome-wide association study on the seedless phenotype in Banana (Musa spp.) reveals the potential of a selected panel to detect candidate genes in a Vegetatively propagated crop. *PLoS ONE*, *11*, e0154448.

Sarma, B. K., Yadav, S. K., Singh, S., & Singh, H. B. (2015). Microbial consortium-mediated plant defense against phytopathogens: Readdressing for enhancing efficacy. *Soil Biology and Biochemistry*, *87*, 25–33. https://doi.org/10.1016/j.soilbio.2015.04.001

Srividya, N., Davis, E. M., Croteau, R. B., & Lange, B. M. (2015). Functional analysis of (4S)-limonene synthase mutants reveals determinants of catalytic outcome in a model monoterpene synthase. *Proceedings of the National Academy of Sciences*, *112*, 3332–3337. https://doi.org/10.1073/pnas.1501203112

Sukha, D. A., Butler, D. R., Umaharan, P., & Boult, E. (2008). The use of an optimised organoleptic assessment protocol to describe and quantify different flavour attributes of cocoa liquors made from Ghana and Trinitario beans. *European Food Research and Technology*, *226*(3), 405–413. https://doi.org/10.1007/s00217-006-0551-2

Swiegers, J. H., Bartowsky, E. J., Henschke, P. A., & Pretorius, I. S. (2005). Yeast and bacterial modulation of wine aroma and flavour. *Australian Journal of Grape and Wine Research*, *11*, 139–173. https://doi.org/10.1111/j.1755-0238.2005.tb00285.x

Tuenter, E., Foubert, K., & Pieters, L. (2018). Mood components in cocoa and chocolate: The mood pyramid. *Planta Medica*, *84*(12/13), 839–844. https://doi.org/10.1055/a-0588-5534

Vickers, C. E., Gershenzon, J., Lerdau, M. T., & Loreto, F. (2009). A unified mechanism of action for volatile isoprenoids in plant abiotic stress. *Nature Chemical Biology*, *5*, 283–291. https://doi.org/10.1038/nchembio.158

Yeh, M.-C., Platkin, C., Estrella, P., MacShane, C., Allinger, D., Elbaum, R., Keeney, M., Brumaru, B., & Wyka, K. (2016). Chocolate consumption and health beliefs and its relation to BMI in college students. *HSOA J Obes Weight Loss*. https://doi.org/10.24966/OWL-7372/1000004

Zhao, M., Wang, L., Wang, J., Jin, J., Zhang, N., Lei, L., Gao, T., Jing, T., Zhang, S., Wu, Y., & Wu, B. (2020). Induction of priming by cold stress via inducible volatile cues in neighboring tea plants. *Journal of Integrative Plant Biology*, *62*(10), 1461–1468. https://doi.org/10.1111/jipb.12937

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Colonges, K., Loor Solorzano, R. G., Jimenez, J.-C., Lahon, M.-C., Seguine, E., Calderon, D., Subia, C., Sotomayor, I., Fernández, F., Lebrun, M., Fouet, O., Rhoné, B., Argout, X., Costet, P., Lanaud, C., & Boulanger, R. (2022). Variability and genetic determinants of cocoa aromas in trees native to South Ecuadorian Amazonia. *Plants, People, Planet*, 1–20. https://doi.org/10.1002/ppp3.10268