



**HAL**  
open science

## **Integration of GWAS, metabolomics, and sensorial analyses to reveal novel metabolic pathways involved in cocoa fruity aroma GWAS of fruity aroma in *Theobroma cacao***

Kelly Colonges, Juan-Carlos Jimenez, Alejandra Saltos, Edward Seguine, Rey Gastón Loor Solorzano, Olivier Fouet, Xavier Argout, Sophie Assemat, Fabrice Davrieux, Emile Cros, et al.

### ► **To cite this version:**

Kelly Colonges, Juan-Carlos Jimenez, Alejandra Saltos, Edward Seguine, Rey Gastón Loor Solorzano, et al.. Integration of GWAS, metabolomics, and sensorial analyses to reveal novel metabolic pathways involved in cocoa fruity aroma GWAS of fruity aroma in *Theobroma cacao*. *Plant Physiology and Biochemistry*, 2022, 171, pp.213-225. 10.1016/j.plaphy.2021.11.006 . hal-03654533

**HAL Id: hal-03654533**

**<https://hal.inrae.fr/hal-03654533>**

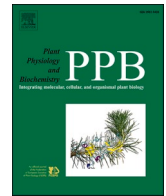
Submitted on 28 Apr 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



# Integration of GWAS, metabolomics, and sensorial analyses to reveal novel metabolic pathways involved in cocoa fruity aroma GWAS of fruity aroma in *Theobroma cacao*

Kelly Colonges<sup>a,b,c,d,\*</sup>, Juan-Carlos Jimenez<sup>e</sup>, Alejandra Saltos<sup>e</sup>, Edward Seguine<sup>f</sup>, Rey Gastón Loor Solorzano<sup>e</sup>, Olivier Fouet<sup>a,b</sup>, Xavier Argout<sup>a,b</sup>, Sophie Assemat<sup>c,d</sup>, Fabrice Davrieux<sup>c,d</sup>, Emile Cros<sup>c,d</sup>, Claire Lanaud<sup>a,b,\*\*</sup>, Renaud Boulanger<sup>c,d,\*\*\*</sup>

<sup>a</sup> Cirad, UMR AGAP, F-34398, Montpellier, France

<sup>b</sup> AGAP Institut, Univ Montpellier, Cirad, INRAE, Institut Agro, Montpellier, France

<sup>c</sup> Cirad, UMR Qualisud, F-34398 Montpellier, France

<sup>d</sup> Qualisud, Univ Montpellier, Avignon Université, Cirad, Institut Agro, IRD, Université de La Réunion, Montpellier, France

<sup>e</sup> Instituto Nacional de Investigación Agropecuarias, INIAP, Ecuador

<sup>f</sup> Guittard Chocolate Co./Seguine Cacao, USA

## ARTICLE INFO

### Keywords:

GWAS  
Cocoa  
Aroma  
Fruity note  
Genetic bases  
Biochemical bases

## ABSTRACT

Nacional is a variety of cocoa tree known for its "Arriba" aroma characterised mainly by fruity, floral, and spicy aromatic notes. In this study, the genetic basis of the fruity aroma of modern Nacional cocoa was investigated. GWAS studies have been conducted on biochemical and sensorial fruity traits and allowed to identify a large number of association zones. These areas are linked to both the volatile compounds known to provide fruity flavours and present in the beans before and after roasting, and to the fruity notes detected by sensorial analysis. Five main metabolic pathways were identified as involved in the fruity traits of the Nacional population: the protein degradation pathway, the sugar degradation pathway, the fatty acid degradation pathway, the monoterpene pathway, and the L-phenylalanine pathway. Candidate genes involved in the biosynthetic pathways of volatile compounds identified in association areas were detected for a large number of associations.

## 1. Introduction

*Theobroma cacao* is a tree species belonging to the Malvaceae family (Bayer and Kubitzki, 2003). Native to the tropical rainforests of northern South America, it is the world's only source of cocoa products obtained after fermentation, drying, and roasting of cocoa beans. *T. cacao* is a diploid plant ( $2n = 2x = 20$ ). Its small genome has been sequenced with 96.7% of the assembly anchored on all 10 chromosomes (Argout et al., 2011, 2017; Motamayor et al., 2013).

Cocoa is classified into two types of products: so-called standard or bulk cocoa, which has a pronounced cocoa taste, and so-called fine aromatic cocoa, which is characterised by floral and fruity notes (Sukha et al., 2008). The production of fine aromatic cocoa, therefore, represents about 5% of the global production but is no less important. Some

Latin American countries produce almost exclusively fine cocoa, which is a significant source of income for them.

The *T. cacao* showed a high genetic diversity. Currently, ten genetic groups have been identified (Motamayor et al., 2008). The most widely cultivated varieties providing fine aromatic cocoa are the Nacional, Criollo, and Trinitario. Trinitarios are hybrids between the Criollo and the Amelonado. The Amelonado variety is a variety producing mainly "standard" cocoa. The Criollo variety produces cocoa beans with mainly fruity aromas (Lachenaud and Motamayor, 2017) but is not widely cultivated because of its low vigour and increased susceptibility to disease (Cheesman, 1944).

The Nacional variety is native to Ecuador. The Nacional variety trees currently cultivated (called modern Nacional in this paper) are the result of several generations of crossbreeding between the ancestral Nacional

\* Corresponding author. TA A-108 / 03 Avenue Agropolis, 34398, Montpellier, Cedex 5, France.

\*\* Corresponding author. Cirad, UMR AGAP, F-34398, Montpellier, France.

\*\*\* Corresponding author. Cirad, UMR Qualisud, F-34398 Montpellier, France.

E-mail addresses: [kcolonges@gmail.com](mailto:kcolonges@gmail.com) (K. Colonges), [renaud.boulanger@cirad.fr](mailto:renaud.boulanger@cirad.fr) (R. Boulanger).

<https://doi.org/10.1016/j.plaphy.2021.11.006>

Received 17 September 2021; Received in revised form 3 November 2021; Accepted 6 November 2021

Available online 22 November 2021

0981-9428/© 2021 The Authors.

Published by Elsevier Masson SAS. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

and the Trinitarios (Loor S. et al., 2009).

The Nacional variety is well known for its floral and spicy flavour, denominated "Arriba" flavour. It is for this reason that it is sought after by chocolate makers. It is characterised by floral and woody notes (Luna et al., 2002). In addition, Nacional is known for its low astringency and low bitterness (International Cocoa Organization, 2017). The floral aroma of Nacional was studied and two biosynthetic pathways were highlighted as being mainly responsible for this floral aroma: the terpene biosynthesis pathway and the L-phenylalanine degradation pathway (Colonges et al., 2021; Ziegler, 1990). The aroma of the modern Nacional probably contains floral and fruity aromas that could be the legacy of crossbreeding with Trinitarios (hybrid trees between Criollo and Amelonado) (Loor S. et al., 2009; Rottiers et al., 2019). Loor et al. (2009) (Loor S. et al., 2009), have demonstrated this hybrid nature using molecular markers. This genetic mixing has led to a dilution of the Arriba flavour (Beckett et al., 2017; Boza et al., 2014; Loor S. et al., 2009). From 1940 onwards, surveys were carried out on the Ecuadorian coast to collect Nacional-type cocoa trees to preserve their genetic resources. These collections were placed in two main experimental stations: the Experimental Tropical Station of Pichilingue (EET-P) of INAP (Instituto Nacional de Investigaciones Agropecuarias) and the Cocoa Flavour Centre of Tenguel (CCAT). The fine aromas of the modern Nacional are therefore a blend of the aromas of the ancestral Nacional and the aromas of the Criollo and Amelonado ancestors. Here, this study will be focused on fruity aromas. The fruity aromas are composed of two main classes of aromas: dried fruit aromas and fresh fruit aromas. Dried fruit aromas are mainly due to pyrazines, which mainly appear after roasting. During this stage of transformation, heating induces a large number of chemical reactions, including the Maillard reaction. This reaction leads to the synthesis of pyrazines from amino acids and reducing sugars (Arnoldi et al., 1988).

Different families of volatile compounds such as alcohols, esters, and ketones represent fresh fruit aromas. During alcoholic fermentation, yeasts synthesize short-chain alcohols (e.g., propan-1-ol; butan-1-ol, 2-methyl-propan-1-ol, pentan-1-ol, hexan-1-ol, heptan-1-ol, octan-1-ol) from simple sugars. These alcohols can also be synthesised from a pyruvate molecule, present in all organisms (Sun et al., 2015). Aldehydes are formed through the oxidation of alcohol. This is the case, for example, in the Lilac flower (*Syringa vulgaris* L.) where the "lilac alcohols" are oxidised to give "lilac aldehydes". The esters are the result of the esterification reaction that is a reversible chemical reaction. Esterification allows the synthesis of an ester molecule and a water molecule from an alcohol molecule and an acid molecule (Aranda et al., 2008).

The objectives of this study are to decipher the genetic, genomic, and biochemical bases of the fruity aromas of the Nacional cocoa variety. A GWAS (Genome-Wide Association Study) study was carried out to discover the areas of the genome responsible for the fruity flavour. It involved phenotyping data related to volatile compounds, potentially linked to the fruity taste, as well as sensory analysis data. To refine the determinants of the fruity aroma of the modern Nacional variety, candidate genes potentially involved in the biosynthesis pathways of fruity compounds were sought.

## 2. Material and methods

### 2.1. Vegetal material

The plant material used for these experiments were composed of a collection of 152 cocoa trees from Ecuador conserved in the Pichilingue experimental station of the "Instituto Nacional de Investigaciones Agropecuarias" (INIAP) and the "Coleccion de Cacao de Aroma Tenguel" (CCAT) of Tenguel. This population represents the Nacional variety currently grown in Ecuador and has been described by Loor, 2007 (Loor S., 2007).

### 2.2. Fermentation processes

Micro-fermentations of cocoa beans were carried out in a wooden box in the most homogeneous way possible with a homogeneous cocoa mass. The process lasted 4 days with two turns at 24 and 72 h after the beginning of the fermentation. Each clone sample (152) was placed in a protective laundry bag and micro-fermented in a cocoa mass. After fermentation, the samples were put in a dry place. They were considered dried when their moisture content was below 8%.

Actual roasting conditions were 120 °C × 22 min. Each bean sample was adjusted from this basis roast using the validated ISCQF protocol (ISCQF, 2020) moisture and bean size (bean weight) from that source. Times measured from -2C of set point.

### 2.3. Sensorial analysis

One hundred and forty-six individuals were characterised by sensory analyses based on blind tastings carried out on three repetitions per sample by Edward Seguine. Samples were prepared and then evaluated using the protocols in the ISCQF standards (ISCQF, 2020). The flavor evaluation was also done using the protocols of these standards (ISCQF, 2020). In all cases, flavor evaluation sessions were conducted using defined attribute Cocoa of Excellence Reference Liquors (chocolate liquors, dried fermented roasted crushed beans) whose attributes were defined previously by the Technical Committee of the Cocoa of Excellence program. As far as the tasting by one person, all samples were maintained blind using the four digit codes and were evaluated in random order for all three blind evaluations. Thirteen fruity notes were judged with a score ranging from zero (no fruity notes detected) to ten (intense note detected) according to ISCQF protocol (ISCQF, 2020) (table A.1).

### 2.4. Volatile compound analysis by GC-MS

The analysis of volatile compounds was carried out on dried fermented beans (UR) and roasted beans (R). The volatile compounds of cocoa samples were extracted using the technique of solid-phase microextraction in the headspace (SPME-HS). GC/MS analysis were conducted according to the condition described by Assi-Clair et al. (2019) (Assi-Clair et al., 2019).

### 2.5. Statistical analysis

PCA analysis and visualization were made with "mixOmics" R package. Calculation of correlation was made with "agricolae" R package and visualization of correlation matrix with "corrplot" R package.

### 2.6. DNA extraction protocol

DNA extraction was conducted according to Risterucci et al. (2000) protocol (Risterucci et al., 2000).

### 2.7. Genotyping by SSR

This population was genotyped using SSR markers by Loor S. (2007) (Loor S., 2007).

### 2.8. Genotyping by sequencing (GBS)

DNA samples were genotyped by sequencing (GBS) using DArTseq (Diversity Arrays Technology Sequencing) technology (Kilian et al., 2012). Reads were aligned with the V2 sequence of the Criollo genome (Argout et al., 2017). Markers with unknown locations were discarded for analysis.

**Table 1**

List of biochemical compounds related to fruity traits and used for the GWAS analysis of unroasted (UR) and roasted (R) beans.

UR	R	Biochemical traits	Compound family	Path	Aroma
X	X	Acetic acid	Acid	FA	Vinegar (Ferreira et al., 1997; Mahajan et al., 2004)
X	X	Pentanoic acid	Acid	FA	Unpleasant, acre, cheese, sweat (Genovese et al., 2007; Jezussek et al., 2002; Kalua et al., 2007; Mahajan et al., 2004)
	X	Hexanoic acid	Acid	FA	Fatty acid, cheese, sweat, rancid, sweet and vinegar (Colahan-Sederstrom and Peterson, 2005; Ferreira et al., 1997; Genovese et al., 2007; Jezussek et al., 2002; Mahajan et al., 2004; Perestrello et al., 2006)
X	X	Octanoic acid	Acid	FA	Fatty acid, rancid, cheese, waxy (Colahan-Sederstrom and Peterson, 2005; Ferreira et al., 1997; Genovese et al., 2007; Karagül-Yüceer et al., 2003; Perestrello et al., 2006)
X	X	Nonanoic acid	Acid	FA	Acid (Colahan-Sederstrom and Peterson, 2005)
X	X	Ethanol	Alcohol	SS	Alcohol (Kalua et al., 2007)
	X	Pentan-1-ol	Alcohol	FA/ SS	Fruity (Kalua et al., 2007)
X	X	Pentan-2-ol	Alcohol	FA/ SS	
	X	Hexan-2-ol	Alcohol	SS	Green grass (Guichard et al., 2003)
X	X	Heptan-2-ol	Alcohol	SS	
X		Nonan-2-ol	Alcohol	SS	Cucumber (Arn and Acree, 1998)
X		2-methylbutan-1-ol	Alcohol	SS	
X	X	3-methylbutan-1-ol	Alcohol	SS	Plastic, Fuel oil, whiskey characteristic pungent (Guichard et al., 2003)
X		2-methylbutan-2-ol	Alcohol	SS	
	X	2-methyl-but-3-en-2-ol	Alcohol	SS	
	X	2-ethylhexan-1-ol	Alcohol	SS	
	X	Butan-2,3-diol	Alcohol	FA/ SS	Fruit, onions (Arn and Acree, 1998)
	X	Butane-1,3-diol	Alcohol	FA/ SS	
	X	Cis-hept-4-en-1-ol	Alcohol	SS	
X	X	Benzyl alcohol	Alcohol	L-phe	Fruity (Ito et al., 2002)
X	X	Acetaldehyde	Aldehyde	SS	Acre, ether (Arn and Acree, 1998; Wang et al., 2014)
	X	2-methylpropanal	Aldehyde	SS	Acre, malted, green (Arn and Acree, 1998)
X	X	2-methylbutanal	Aldehyde	SS	Malted (Fickert and Schieberle, 1998)
X	X	3-methylbutanal	Aldehyde	SS	Stimulant, malted (Kumazawa and Masuda, 2002)
X		Nonanal	Aldehyde	Other	Orange-like, floral, soapy (Karagül-Yüceer et al., 2003; Kumazawa and Masuda, 2002; Mahajan et al., 2004)
	X	2-phenylbut-2-enal	Aldehyde	Other	
	X	5-methyl-2-phenylhex-2-enal	Aldehyde	SS	Cacao (Garg et al., 2018)
X	X	Benzaldehyde	Aldehyde	L-phe	Bitter, cherry, almond, fruity (Perestrello et al., 2006; Pham et al., 2008; Wang et al., 2014)
X	X	Methyl acetate	Ester	SS	Ester, green (Garg et al., 2018)
	X	Methyl octanoate	Ester	FA	Fruity, orange, wax, wine (Garg et al., 2018)
X	X	Ethyl acetate	Ester	FA/ SS	Fruity, solvent, apple, sweet (Ferreira et al., 1997; Genovese et al., 2007; Karagül-Yüceer et al., 2003)
X	X	Ethyl propanoate	Ester	FA/ SS	Fruity, strong, strawberry (Kalua et al., 2007)
	X	Ethyl butanoate	Ester	FA	Strawberry, fruity, sweet, kiwi, pineapple (Ferreira et al., 1997; Genovese et al., 2007; Larsen and Poll, 1992; Wang et al., 2014)
	X	Ethyl hexanoate	Ester	Other	Strawberry, green apple, fruity and floral (Ferreira et al., 1997; Genovese et al., 2007; Larsen and Poll, 1992; Wang et al., 2014)
X		Ethyl 3-hydroxyhexanoate	Ester	SS	Strawberry (Genovese et al., 2007)
X	X	Ethyl octanoate	Ester	FA/ SS	Sweet, fruity, fresh, pineapples (Genovese et al., 2007; Perestrello et al., 2006)
	X	Ethyl dodecanoate	Ester	FA	Floral, fruity, leaf (Garg et al., 2018)
X		Ethyl hexadecanoate	Ester	SS	
X		Ethyl benzoate	Ester	L-phe	Ripe fruit, fruity, violets, candy (Ferreira et al., 1997)
X		Ethyl 2-methylpropanoate	Ester	FA	Fruity, sweet, rubber (Arn and Acree, 1998; Schieberle et al., 1990)
	X	Ethyl 2-methylbutanoate	Ester	FA	Green apple, fruity, red fruit (Baek et al., 1997; Genovese et al., 2007)
X		Butyl benzoate	Ester	FA	
	X	Propyl acetate	Ester	Other	Celery, floral, pear, red fruit (Garg et al., 2018)
X		Hexyl acetate	Ester	Other	Fruity, pear, sweet and floral (Guichard et al., 2003; Wang et al., 2014)
X		2-methylpropyl acetate	Ester	FA	Strawberry (Aznar et al., 2001)
X	X	1-methylbutyle acetate	Ester	SS	
X	X	3-methylbutyle acetate	Ester	FA/ SS	Banana (Genovese et al., 2007; Guichard et al., 2003)
X		Butan-2,3-diyl diacetate	Ester	SS	
X		Meso butan-2,3-di-yl diacetate	Ester	SS	
X	X	3-methyl-but-2-enyl acetate	Ester	SS	Jasmin, banana, fresh, ripe, heliotrope, sweet, fruity, balsam, lavender (Garg et al., 2018)
X	X	1-methylhexyle acetate	Ester	SS	
	X	1-phenylethyl acetate	Ester	L-phe	Fruity (Garg et al., 2018)
X	X	Furfural	Furan	Other	Incense, fruity, flowery, roasted, sweet and almond (Colahan-Sederstrom and Peterson, 2005; Ferreira et al., 1997; Wang et al., 2014)
	X	Acetone	Ketone	?	Acre (Garg et al., 2018)
X		Butan-2-one	Ketone	SS	Fruity, ethereal (Kalua et al., 2007)
X	X	3-hydroxy-butan-2-one	Ketone		Butter, cream (Arn and Acree, 1998)

(continued on next page)

Table 1 (continued)

UR	R	Biochemical traits	Compound family	Path	Aroma
				FA/ SS	
X	X	Pentan-2-one	Ketone	FA/ SS	Ether, fruity (Arn and Acree, 1998)
	X	Pentan-2,3-dione	Ketone	FA/ SS	Cream, butter (Arn and Acree, 1998)
X		Hexan-2-one	Ketone	FA/ SS	Ether (Arn and Acree, 1998)
X	X	Heptan-2-one	Ketone	FA/ SS	Sweet, fruity (Kalua et al., 2007)
	X	Octan-2-one	Ketone	FA/ SS	Green, mouldy (Kalua et al., 2007)
	X	Nonan-2-one	Ketone	FA/ SS	Warm milk, syrup, green (Arn and Acree, 1998)
X	X	Acetophenone	Ketone	L-phe	Acacia honey, floral and fruity (Genovese et al., 2007; Wang et al., 2014)
X	X	Gamma-butyrolactone	Lactone	FA?	Caramel, fat, sweet, creamy, oily (Garg et al., 2018)
	X	Methylpyrazine	Pyrazine	P	Popcorn (Arn and Acree, 1998)
	X	2,3-dimethylpyrazine	Pyrazine	P	Hazelnut, cooked (Mahajan et al., 2004; Wang et al., 2014)
	X	2,6-dimethylpyrazine	Pyrazine	P	Cooked meat, cooked rice, hazelnut (Mahajan et al., 2004; Pham et al., 2008; Wang et al., 2014)
	X	2,3,5-trimethylpyrazine	Pyrazine	P	Hazelnut, roasted (Mahajan et al., 2004; Wang et al., 2014)
X	X	2,3,5,6-tetramethylpyrazine	Pyrazine	P	Cooked (Wang et al., 2014)
	X	2,3-dimethyl-5-ethylpyrazine	Pyrazine	P	
	X	2-ethyl-5-methylpyrazine	Pyrazine	P	Fruity, green, sweet (Arn and Acree, 1998; Garg et al., 2018)
	X	2-ethyl-6-methylpyrazine	Pyrazine	P	Green, hazelnut, grilled (Garg et al., 2018)
	X	Trimethylpyridine	Pyridine	P	
X	X	Linalool	Terpene	M	Floral, lemon peel, orange blossom (Ferreira et al., 1997; Genovese et al., 2007)
X		Linalool cis pyranic oxide	Terpene	M	Fruity, citrus, green (Arn and Acree, 1998; Ito et al., 2002)
X		Linalool trans furanic oxide	Terpene	M	Citrus, leafy, floral (Arn and Acree, 1998; Ito et al., 2002)

UR: Unroasted beans; R: Roasted beans; Path.: Pathway, FA: Fatty acid degradation pathway; SS: Sugar degradation pathway; M: Monoterpene biosynthesis; P: Pyrazine production; L-phe: L-phenylalanine degradation pathway.

## 2.9. Association mapping

We performed a GWAS analysis with SNP markers associated with biochemical (146 accessions x 5195 markers) and sensory (144 accessions x 5195 markers) traits using TASSEL 219 v5.

For all the traits, we used a mixed model (MLM) and a fixed-effect model (GLM) the detailed information was described in Colonges et al. (2021).

The threshold was determined using the Bonferonni correction formula as proposed by Gao et al. (2008) with the effective number of independent tests (Meff) used as the denominator and calculated by SimpleM R package (Gao et al., 2010). Meff was 2796, which corresponds to a P-value of approximately 1.79e-05. The significance of all markers was plotted using Manhattan plots with the R QQman package.

We performed an analysis with SSR markers associated with biochemical (180 accessions x 180 markers) and sensory (197 accessions x 180 markers) traits using TASSEL v3. We used a fixed-effect model (GLM) with a structure matrix; the option of 500 permutations was chosen the detailed information was described in Colonges et al. (2021). The threshold was determined using the Bonferonni correction corresponding to a p-value about 2.78e-04.

The choice of markers for the genotyping data and the determination of the confidence interval was made using the same method as described by (Colonges et al., 2021).

## 3. Results

### 3.1. Characterisation of the traits studied

#### 3.1.1. Sensory trait analysis

Thirteen fruity notes and notes associated with pyrazine notes (roasted degree, cocoa, browned flavour, Caramel browned sugar) were evaluated during the sensory analysis carried out on the cocoa liquors.

All of these thirteen sensory traits were used to carry out this study (table A.1).

The results of a PCA for fruity notes (sensory analysis traits) showed a continuous variation within the population (Fig. 1A). The aromatic notes of brown flavour, fruity browned dried fruit, and nutty mainly define axis 1. Aromatic notes fruity acidity, Fruity-Citrus, and Fruity-Dark tree fruit mainly define axis 2.

Analyses of correlations between the different sensory traits related to the fruity taste did not show strong negative correlations. Three strong positive correlations were detected; a correlation of 0.8–1 between the hazelnut note and browned flavour characteristics, two correlations of 0.6–0.8 between the browned dried fruit note and the nutty note on the one hand and the browned flavour note, on the other hand (Fig. 2A).

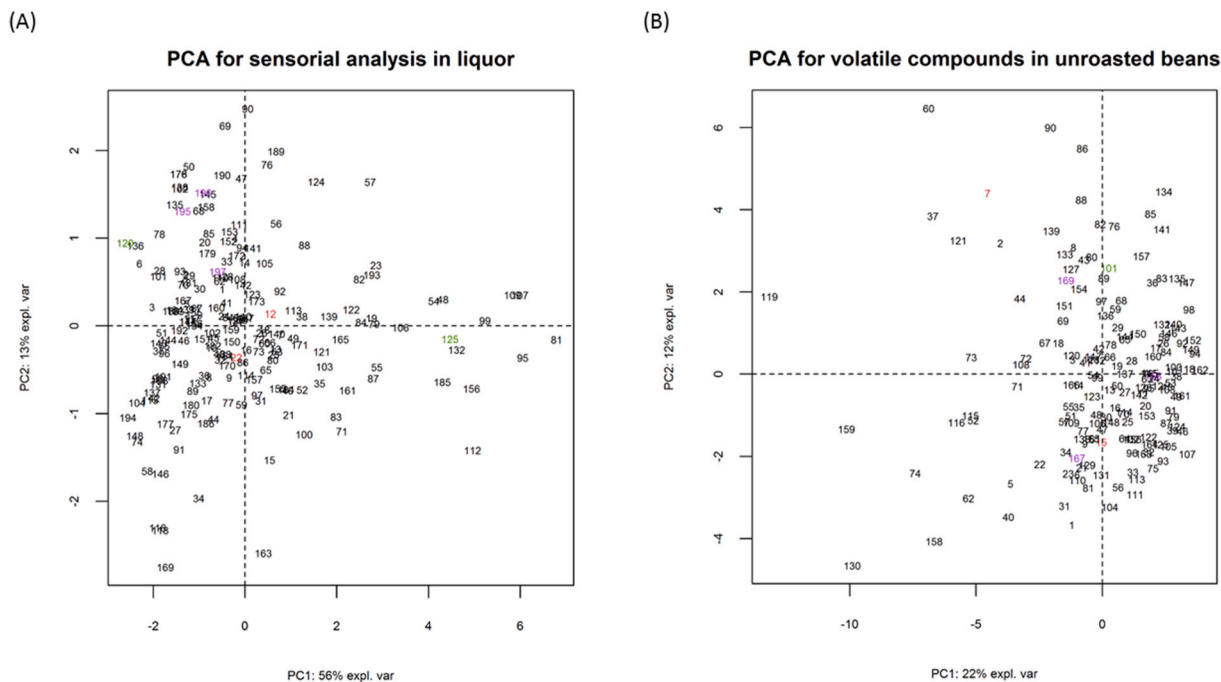
#### 3.1.2. Biochemical traits analysis

The identification of volatile compounds was carried out in cocoa beans before and after roasting. A total of one hundred and sixty-one volatile compounds were identified. Thirty-five of them are known to have fresh or dried fruity notes (Table 1). These compounds are mainly synthesised through five biosynthesis pathways: proteins degradation pathways (required for the production of pyrazines via the Maillard reaction), the fatty acid degradation pathway, the sugar degradation pathway, the L-phenylalanine degradation pathway, and the monoterpene biosynthesis pathway.

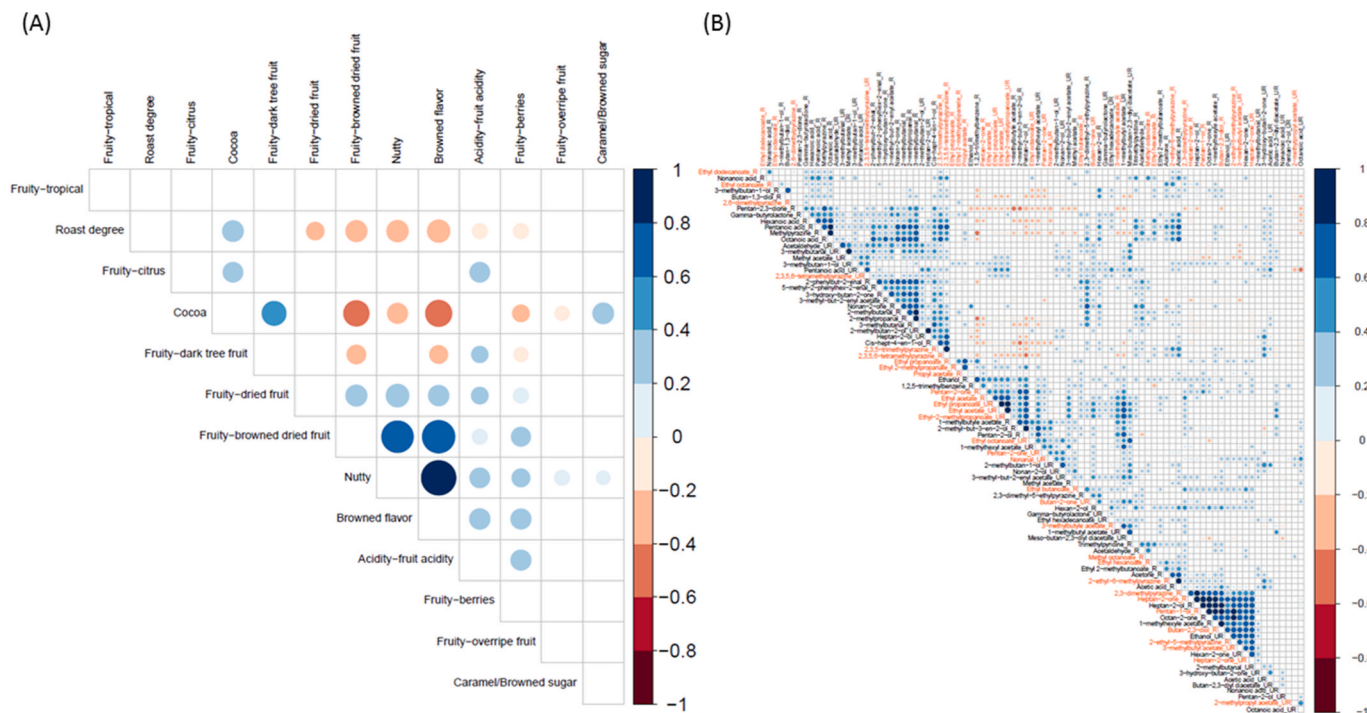
The result of a PCA for volatile compounds in unroasted beans involved in fruity notes shows a continuous variation within the population (Fig. 1B). The compounds 1-methylbutyl acetate, meso-butan-2,3-diyl acetate, and ethyl acetate mainly define axis 1. The compounds 3-methylbutyl acetate, heptan-2-one, and ethanol mainly define axis 2.

Correlation analyses were carried out between the different volatile compounds related to fruity notes. Twenty-seven strong positive correlations (greater than or equal to 0.8) were detected between several





**Fig. 1.** PCA related to fruity traits. (colour should be used for any figures in print) (A) PCA of sensorial analysis results related to fruity notes detected in liquor. (B) PCA of biochemical compounds detected related to fruity notes detected in unroasted beans. Black numbers represent the population of modern Nacional variety. Green numbers represent the two individuals closest to Amelonado’s ancestor. Red numbers represent the two individuals closest to the Criollo ancestor. Purple numbers represent SNA604 (Nacional ancestor) and two individuals closest to the Nacional ancestor. (e.g., See also Table A.1).



**Fig. 2.** Significant correlation matrix of sensorial analysis. (colour should be used for any figures in print)(A) Significant correlation matrix between the sensorial profiles determined in cocoa liquor. (B) Significant correlation matrix between the biochemical compounds measured in unroasted (UR) and roasted beans (R). The correlations were calculated by the Pearson method. The white boxes represent no significant correlations. The colour of the circles corresponds to Pearson’s correlation coefficient. The areas of circles correspond to a p-value of correlation coefficients. The p-value threshold for a significant correlation is 0.05. 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. (e.g., See also Figure A.1 and Table A.1).

volatile compounds from unroasted or roasted beans. Sixteen strong positive correlations were between compounds involved in the fatty acid and sugar degradations. Only one strong positive correlation was between the two pyrazines. Four strong positive correlations were between compounds involved in fatty acid and sugar degradation pathway and pyrazines. Seven strong negative correlations were also detected (Fig. 2B) of which five were detected between compounds involved in fatty acid and sugar degradation. The two other negative correlations are between pyrazine and pentan-2-ol (R) or 1, 2, 5-trimethylbenzene (R). No strong correlation was detected between biochemical compounds and sensory analysis data (see figure A.1).

### 3.2. Genome wide association study (GWAS)

The genetic diversity of the population and its structure has been described by Loor et al. (2009) (Loor S. et al., 2009) using SSR markers and by Colonges et al. (2021) (Colonges et al., 2021) using SNP markers. The population has a high rate of heterozygosity.

All detected association areas are available in table A.2.

#### 3.2.1. Identification of significant associations for sensorial traits

Out of all the associations, only 22 relate to fruity sensory data. With thirteen sensory perceptions related to fruity traits, associations could be detected for only three of them: "Fruity-Dark Tree Fruit", "Fruity-Dried fruit" and "Fruity-Berries". Only one association could be detected for three of the four characters (Table 2). Nineteen areas of associations could be detected for the note "Fruity-Dark Tree Fruit". These associations were detected on chromosomes 1, 2, 4, 7, 8, and 10. The strongest association detected for the note "Fruity-Dark Tree Fruit" is also the strongest association for all fruit notes and explains 24% of the variation in this trait.

#### 3.2.2. Identification of significant associations for biochemical traits

The GWAS analysis revealed 480 areas of association for biochemical compounds related to the fruity taste. All the associations found are reported in table A.2.

For all the compounds for which significant associations have been detected, three major biosynthesis pathways seem to emerge. The protein degradation pathway (required for the production of pyrazines via the Maillard reaction), the sugar degradation pathway, and the fatty acid degradation pathway. The two degradation pathways (sugars and fatty acids) are often linked (Fig. 3). They will therefore be presented jointly.

The areas of significant associations have been mapped to visualize their locations and co-locations. Two maps have been produced. A map showing the associations of traits related to pyrazine production (figure A.2). A second map showing the areas of associations related to the

compounds known to be involved in the degradation pathway of sugars or fatty acids (figure A.3). These biosynthesis pathways often have common compounds. Some results are different according to the methods (GLM and MLM) or according to the sorting of markers (MAF5 or G7) or even between the type of markers (SNP or SSR).

Repeatable results between methods appear to be the most conclusive.

#### 3.2.3. Significant associations were identified for the biochemical compounds involved in the pyrazine production pathway

Of the 74 volatile compounds related to the fruity notes, eight pyrazines were identified by GC-MS in the roasted beans and only one in the unroasted beans (2, 3, 4, 5-tetramethylpyrazine). Two hundred and sixty-eight association zones have been detected in relation to pyrazines (Table 2). Association zones were detected on chromosomes 1, 2, 3, 9, and 10. The most significant association zone for 2, 3, 5-trimethylpyrazine (R) co-locates with the most significant association zone detected for methylpyrazine (R). Six of the most significant association zones for these compounds are located outside the calculated haplotypic blocks.

The most significant association found for pyrazines is the association zone associated with 2-ethyl-5-methylpyrazine (R). The variation of this trait is explained at 49% by the genetic variation located in this association zone.

The map showing the significant associations for the traits linked to the pyrazine production pathway (figure A.2) shows interesting results because, in a large number of associations represented, co-locations can be observed between the associations detected for different pyrazines. For example, on chromosome 1 between 2, 3-dimethyl-5-ethyl pyrazine (R) and methylpyrazine (R) or between 2, 3, 5, 6-tetramethylpyrazine (R); 2, 3, 5-trimethylpyrazine (R) and 2, 3-dimethyl-5-ethylpyrazine (R).

#### 3.2.4. Co-locations between biochemical compounds and sensorial traits

Six co-locations between pyrazine-related associations and those related to sensory notes could be observed. Four are present on chromosome 1: two co-locations between the note "Fruity- Dark Tree fruit", methylpyrazine (R) and 2,3-dimethyl-5-ethylpyrazine (R); one between the note "Fruity- Dark Tree fruit" and methylpyrazine (R); one between the note "Fruity- Dark Tree fruit" and 2,3-dimethyl-5-ethylpyrazine (R). A co-location is present on chromosome 4 between the note "Fruity-Dark Tree fruit", 2,3,5,6-tetramethylpyrazine (R), and 2,3,5-trimethylpyrazine (R). A co-location is present on chromosome 10 between the note "Fruity- Dark Tree fruit" and methylpyrazine (R). Only 2,3,5,6-tetramethylpyrazine (R) and 2, 3, 5-trimethylpyrazine (R) are known to have fruity notes. It is, therefore, possible that other compounds explain the "Fruity- Dark Tree fruit" note.

**Table 2**

Most significant association detected for each sensory fruity note and pyrazines.

CH	Position of the association peak	N° hap. bloc	Fruity note and pyrazines detected	p-value of the strongest association	Explanation rate of the trait of the strongest association	Associations detected
1	2 430 002 bp	4	Fruity-Dark Tree Fruit	6,33E-09	24%	1/19
2	32 194 678 bp	61	Fruity-Dried fruit	3,19E-06	16%	1
7	4 613 061bp	19	Fruity-Berries	1,18E-06	15%	1
1	7,081,063	26	2,3,5-trimethylpyrazine_R	2.15E-08	25%	17
1	7,081,063	26	Methylpyrazine_R	2.57E-08	25%	38
2	8,389,914	NA	2,3-dimethylpyrazine_R	1.99E-11	29%	37
2	11,231,613	NA	2-ethyl-5-methylpyrazine_R	7.80E-16	49%	94
3	31,136,388	NA	2,6-dimethylpyrazine_R	4.67E-07	25%	6
4	23,199,282	39	2,3,5,6-tetramethylpyrazine_R	5.19E-09	24%	18
4	30,166,293	70	2,3-dimethyl-5-ethylpyrazine_R	1.35E-08	55%	50
9	29,324,700	NA	2,3,5,6-tetramethylpyrazine_UR	9.14E-07	18%	2
10	3,974,481	19	2-ethyl-6-methylpyrazine_R	4.45E-07	21%	6

CH: chromosome; N°: number; hap.: haplotypic; R: roasted beans; UR: Unroasted beans (e.g., See also Figure A.2).

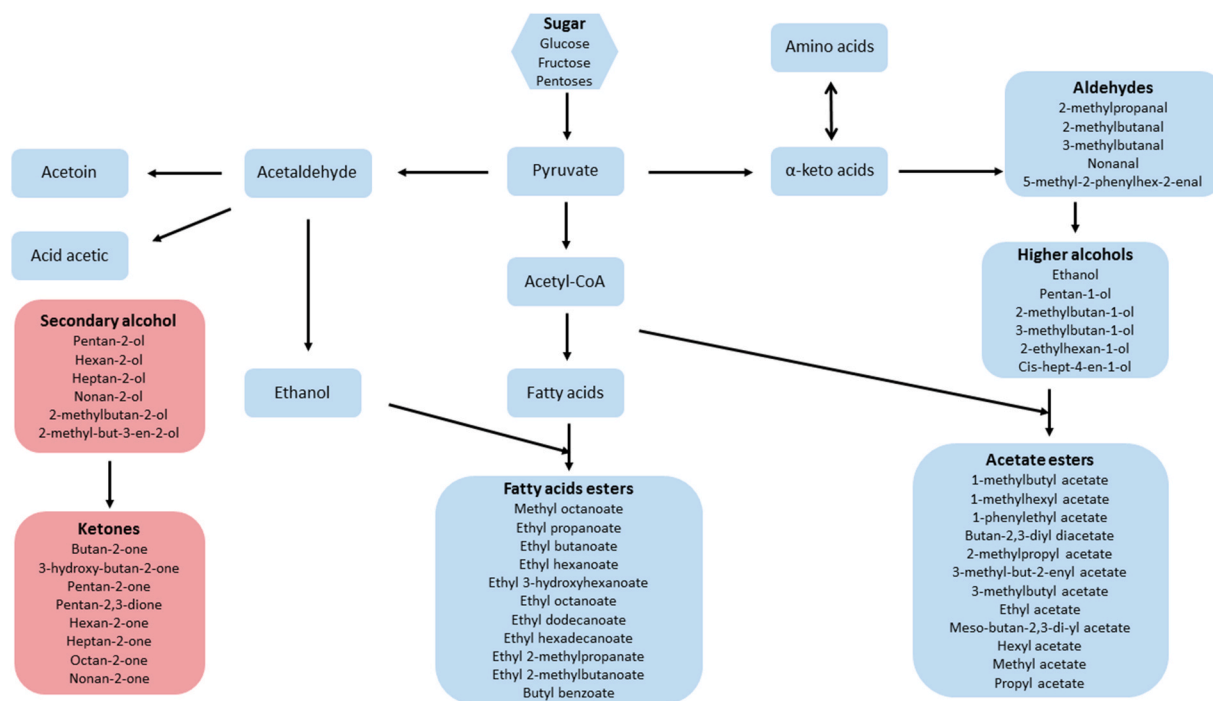


Fig. 3. Scheme representation of fatty acid and sugar degradations adapted to Dzialo et al. (2017) (Dzialo et al., 2017). (colour should be used for any figures in print).

### 3.2.5. Significant associations were identified for the biochemical compounds involved in the degradation of fatty acid and sugar pathway

Four hundred and eighty associations were detected in connection with the compounds involved in the degradation pathways of sugars and fatty acids (table A.3). Some of the most significant associations are located in the same haplotypic block but with a different position of the peak of the associations (Table 3). Some peaks of the most significant associations co-locate (Table 3).

Some associations outside the calculated haplotypic blocks present co-localizations, this is the case for the most significant associations detected for heptan-2-ol (R), heptan-2-one (R), octan-2-one (R); the most significant associations detected for 2 and 3-Methyl-butyl acetate (UR), 1-methylhexyl acetate (R), ethanol (UR) and pentan-1-ol (R).

### 3.2.6. Co-locations between biochemical compounds and sensorial traits

Nine co-locations between associations linked to fruity notes and associations linked to volatile compounds were detected (Table 3) of which seven are between the fruit note “Dark tree fruit”, one with the dried fruit note and one with the berries note.

### 3.2.7. Involvement of monoterpenes and L-phenylalanine pathway compounds in the fruity notes of cocoa

As shown by (Colonges et al., 2021), monoterpene biosynthesis and L-phenylalanine degradation pathways are involved in the synthesis of fruity notes. Two volatile compounds with fruity notes belonging to the monoterpene biosynthesis have been identified in this population of cocoa: the linalool cis pyranic oxide, known for its citrus note, and the linalool Trans furanic oxide known for its floral and citrus notes. Respectively, twenty-seven and twenty-nine association area was detected related to these two compounds.

Five volatile compounds known for fruity notes belonging to the L-phenylalanine degradation pathway have been identified: benzaldehyde, benzyl alcohol, 1-phenylethyl acetate, ethyl benzoate, and acetophenone. Two associations' zones were identified for benzaldehyde (UR) and seventy-two for benzaldehyde (R), known for its cherry and bitter almond notes. Forty-two associations' zones were identified related to benzyl alcohol (UR), seventy-three for 1-phenylethyl acetate

(R), two for ethyl benzoate (UR), thirty-six for acetophenone (UR), and forty for acetophenone (R), known for their fruity notes.

### 3.3. Candidate genes are potentially involved in the formation of the dried fruity aroma

Across the pyrazine association zones, 100 candidate genes were identified (table A.4). 94 genes are identified as being involved in the synthesis of precursors of the Maillard reaction (amino acids and reducing sugars). These genes are therefore involved in either protein degradation or sugar degradation (figure A.3).

Of the 100 genes, 30 could be involved in the degradation of proteins, sugar, and fatty acid (Table 4), including 25 genes coding for the Alpha/Beta hydrolase family of proteins known to have various functions, including peptidase function (Holmquist, 2000; Mindrebo et al., 2016). In addition, five genes coding for enzymes with oxidase/reductase or hydrolase functions.

Of the 100 genes, 61 are involved in protein degradation (Table 4). These genes could therefore be at the origin of the synthesis of amino acids, necessary for the Maillard reaction, by degrading existing proteins.

Among the 100 genes, three are involved in the degradation of sugars, and six are involved in fatty acid degradation (Table 4). These genes could be at the origin of the synthesis of sugars also necessary for the Maillard reaction.

### 3.4. Candidate genes potentially involved in the formation of the fresh fruity aroma

#### 3.4.1. Fatty acid and sugar degradation pathways

In the set of association zones linked to volatile compounds (acids, alcohols, aldehydes, esters, ketones) from the degradation pathways of sugars and fatty acids, 227 candidate genes were identified (table A.4).

Of the 227 genes identified, 30 encode an "alpha-beta hydrolase" with a probable lipase function (Holmquist, 2000; Mindrebo et al., 2016) and 8 for enzymes with hydrolase functions. As their name indicates, the enzymes encoded by these genes would participate in lipids



**Table 3**

Co-localisations between compounds involved in the fatty acid and sugar degradations pathways or/and in the L-phenylalanine degradation pathway.

Chromosome	Position	N° of Haplotypic bloc	Sensory trait	Biochemical Trait
1	4 498 932 4 522 166	14	/	2-methylbutanal (R)
1	6 281 788 6 448 063 6 448 063	23	/	5-methyl-2-phenylhex-2-enal (R) hexan-2-ol (R) 2-phenylbut-2-enal (R) 3-methylbutanal-R
1	6 855 567 7 081 063	26	/	pentatonic acid (R) octanoic acid (R)
1	36 447 062 36 447 686	91	/	3-methylbutan-1-ol (R) acetone (R)
2	8 389 914	NA	/	Heptan-2-ol-R Heptan-2-one-R Octan-2-one-R
2	11 231 613	NA	/	3-Methyl-butyl acetate-UR 1-methylhexyl acetate-R Ethanol-UR Pentan-1-ol-R
5	27 513 744	45	/	1-methylhexyl acetate (UR) ethyl benzoate (UR) ethyl butanoate (R)
5	38 411 373	87	/	hexan-2-one (UR) pentan-2,3-dione (R)
6	25 201 654 25 355 204	65	/	3-hydroxy-butan-2-one (UR) nonan-2-ol (UR)
7	5 596 523 5 607 833 5 607 833 5 607 833	24	/	2-methylbutan-1-ol (UR) meso-2,3-butan-di-yl diacetate (UR) methyl acetate (UR)
7	10 459 413	31	/	hexyl acetate (UR)
9	1 099 704	5	/	pentanoic acid (UR) 2-ethylhexan-1-ol (R)
9	37 554 266 37 557 289	54	/	2-methylpropanal (R) methyl acetate (R) nonanal (UR)
1	Start 2 368 915	4	Dark Tree fruit	2-methylbutan-2-ol (UR), 2-phenylbut-2-enal (R), 3-methylbutanal (R), octanoic acid (R) and ethyl butanoate (R)
1	End 2 445 782 Start 2 694 498	6	Dark Tree fruit	2-phenylbut-2-enal (R), 3-methylbutanal (R) and ethyl butanoate (R)
1	End 2 793 273 Start 3 083 032	7	Dark Tree fruit	2-phenylbut-2-enal (R), 3-methylbutanal (R), octanoic acid (R),

**Table 3 (continued)**

Chromosome	Position	N° of Haplotypic bloc	Sensory trait	Biochemical Trait
				pentanoic acid (R), ethyl butanoate (R), 3-hydroxy-butan-2-one (R) and gamma-butyrolactone (R)
1	End 3 398169 Start 3 721 132	9	Dark Tree fruit	2-phenylbut-2-enal (R), 3-methylbutanal (R), 1-methylhexyl acetate (R), propyl acetate (R), octanoic acid (R), pentanoic acid (R), ethyl butanoate (R), gamma-butyrolactone (R), heptan-2-ol (UR), heptan-2-one (R) and pentan-1-ol (R)
1	End 4 022 340 Start 4 114 978	10	Dark Tree fruit	octanoic acid (R), pentanoic acid (R), 2-phenylbut-2-enal (R), 3-methylbutanal (R) and ethyl dodecanoate (R)
2	End 4 134 570 Start 31 957 983 End 32 194 678	61	Dried fruit	Methyl acetate (UR) and pentan-2-ol (UR)
4	Start 22 748 126	39	Dark Tree fruit	2 and 3-methylbutyl acetate (UR), 2-methylbutan-2-ol (UR), 2-phenylbut-2-enal (R), 3-hydroxy-butan-2-one (UR), pentanoic acid (R), butan-2,3-diol (R), heptan-2-ol (R), heptan-2-one (UR), octan-2-one (R), pentan-1-ol (R) and pentan-2,3-dione (R)
7	End 23 333 424 4 613 061	19	Berries	Ethyl octanoate (R)
10	6 167 221	NA	Dark tree fruit	3-methylbutanal (R) and pentanoic acid (R)

N°: number, R: roasted beans; UR: Unroasted beans (e.g., See also [Figure A.3](#)).

or sugars degradations.

Of the 227 genes identified, 121 have functions for fatty acid degradation ([Table 4](#)), 11 genes were identified as being involved in the fatty acid synthesis, 47 involved in sugars degradation, six involved in amino acid degradation, four involved in carboxylic acids degradation. The degradation of sugars, amino acids, and carboxylic acids are necessary steps for the synthesis of compounds such as acids, alcohols, or esters.

#### 3.4.2. Monoterpene biosynthesis pathway

For the linalool cis pyranic oxide four genes coding for "Probable terpene synthase 9" was identified (three on chromosome 7 and one on chromosome 10) ([table A.4](#)). This gene is known to be responsible for the transformation of the geranyl diphosphate into linalool, the precursor of linalool cis-pyranic oxide ([Cseke et al., 1998](#)). For the linalool Trans furanic oxide six candidate's genes were identified on chromosome 5: five coding for *Cytochrome P450 89A2*" and one for *"Cytochrome P450 89A9"*. At least, two studies have shown the implication of *Cytochrome P450* for the transformation of linalool into 6,7-epoxylinalool

**Table 4**  
Synthesis of gene functions found in association zones linked to fruity traits.

Number of genes	Gene function	General function	Pathway
25	Alpha/Beta hydrolase family	Proteins, Fatty acid, Sugars degradations	Pyrazine production
17	Protease function	Proteins degradation	Pyrazine production
16	Peptidase function	Proteins degradation	Pyrazine production
11	Glutathione S-transferase	Proteins degradation	Pyrazine production
5	Oxidase/reductase or hydrolase function	Proteins, Fatty acid, Sugars degradations	Pyrazine production
4	Proteinase function	Proteins degradation	Pyrazine production
3	Aminotransferase function	Proteins degradation	Pyrazine production
3	Amino acid decarboxylase function	Proteins degradation	Pyrazine production
1	Delta-1-pyrroline-5-carboxylate	Proteins degradation	Pyrazine production
1	Anthranilate synthase	Proteins degradation	Pyrazine production
1	Bifunctional aspartokinase/homoserine dehydrogenase	Proteins degradation	Pyrazine production
1	Lactoylglutathione lyase	Proteins degradation	Pyrazine production
1	Peptidyl-prolyl cis-trans isomerase	Proteins degradation	Pyrazine production
1	Protein S-acyltransferase 16	Proteins degradation	Pyrazine production
1	Peptide-N4-(N-acetyl-beta-glucosaminyl) asparagine amidase	Proteins degradation	Pyrazine production
1	Endoglucanase	Sugars degradation	Pyrazine production
1	Xyloglucan endotransglucosylase/hydrolase	Sugars degradation	Pyrazine production
1	Inositol oxygenase 1	Sugars degradation	Pyrazine production
4	Carboxylesterase	Fatty acid degradation	Pyrazine production
1	Isoprenylcysteine alpha-carbonyl methylesterase	Fatty acid degradation	Pyrazine production
1	Phospholipase	Fatty acid degradation	Pyrazine production
1	Branched-chain-amino-acid aminotransferase	Amino acid degradation	FA/SS
1	N-acetyltransferase	Amino acid degradation	FA/SS
1	Peptidyl prolyl cis-trans isomerase	Amino acid degradation	FA/SS
1	Arabinosyl transferase	Amino acid degradation	FA/SS
2	Methyltransferase	Amino acid degradation	FA/SS
1	Malonate CoA ligase	Carboxylic acids degradation	FA/SS
1	Methylmalonate semialdehyde dehydrogenase	Carboxylic acids degradation	FA/SS
2	Malate dehydrogenase	Carboxylic acids degradation	FA/SS
30	Alpha-beta hydrolase	Fatty acid and sugars degradation	FA/SS
8	Hydrolase	Fatty acid and sugars degradations	FA/SS
1	3-ketoacyl-coA thiolase	Fatty acid degradation	FA/SS
1	Plastidial glycolate/glycerate translocator	Fatty acid degradation	FA/SS
1	Enoyl CoA hydratase	Fatty acid degradation	FA/SS

**Table 4 (continued)**

Number of genes	Gene function	General function	Pathway
1	Non-specific lipid transfer	Fatty acid degradation	FA/SS
1	Malonyl CoA decarboxylase enzyme	Fatty acid degradation	FA/SS
4	Desaturase	Fatty acid degradation	FA/SS
4	Linoleate lipoxygenase	Fatty acid degradation	FA/SS
6	Dehydrogenase, dehydratase or oxygenase	Fatty acid degradation	FA/SS
9	Oxidation-reduction of fatty acids	Fatty acid degradation	FA/SS
9	Esterase	Fatty acid degradation	FA/SS
19	Lipase or phospholipase functions	Fatty acid degradation	FA/SS
21	Acyltransferase or transferase	Fatty acid degradation	FA/SS
44	GDSL esterase/lipase enzymes	Fatty acid degradation	FA/SS
1	Acetyl-coenzyme A synthetase	Fatty acid synthesis	FA/SS
1	Lipoyl synthase	Fatty acid synthesis	FA/SS
1	Lipid-A-disaccharide synthase	Fatty acid synthesis	FA/SS
1	Long-chain acyl-CoA synthetase 1	Fatty acid synthesis	FA/SS
2	3-ketoacyl-CoA synthase	Fatty acid synthesis	FA/SS
5	3-oxoacyl-[acyl-carrier-protein] synthase	Fatty acid synthesis	FA/SS
1	Amylase	Sugars degradation	FA/SS
1	Fructofuranosidase	Sugars degradation	FA/SS
1	Simple sugar dehydrogenase	Sugars degradation	FA/SS
1	Glucose-6-phosphate-1-epimerase	Sugars degradation	FA/SS
1	Polygalacturonase	Sugars degradation	FA/SS
2	Phosphatase	Sugars degradation	FA/SS
6	Xylodase	Sugars degradation	FA/SS
7	Glucosidase	Sugars degradation	FA/SS
8	Transferase function on sugar molecules	Sugars degradation	FA/SS
8	Galactosidase	Sugars degradation	FA/SS
11	Glycosyltransferase	Sugars degradation	FA/SS

FAS: Fatty acid and sugar degradation pathway (e.g., See also [Figure A.2](#) and [Figure A.3](#)).

([Chen et al., 2010](#); [Meesters et al., 2007](#)).

### 3.4.3. L-phenylalanine degradation pathway

In total, twenty-nine candidate genes were identified in these associations' zones involved in their biosynthesis. Any candidate gene was detected in associations linked to ethyl benzoate (UR). Forty-six candidate genes were identified in the areas of association with the other compounds. They were chosen of their predictive functions. They indicate possible involvement in the degradation pathway of L-phenylalanine ([table A.4](#)).

### 3.5. Candidates genes potentially involved in the general plant defence identified in association areas linked with biochemical compounds

As suggested earlier (Sabau et al., 2006), the detection of fermentative micro-organisms seems to induce defensive responses that may be responsible for the synthesis of certain aromatic compounds.

In the association zones linked to compounds involved in fatty acid and sugar degradation, twenty-three genes are involved in the general defences of the plant (Table 5). Jasmonic acid and salicylic acid are two phytohormones with an important role in the defence mechanisms.

In the association zones linked to pyrazines compounds, sixteen genes coding for enzymes with functions involved in general plant defences were identifying (Table 5). The functioning of the proteasome plays an essential role in the activation of the jasmonic acid response pathway (Song et al., 2014). The fumaryl-acetoacetate hydrolase was identified in *Arabidopsis thaliana* as affecting cell death in response to jasmonic acid (Zhou et al., 2020). Ethylene is also a phytohormone that plays a role in the response to biotic stresses (Song et al., 2014).

**Table 5**

Synthesis of genes found in association zones and involved in the general plant defences.

N° of genes	Gene function	Role in the plant defences	Type of associated compounds
1	Indole-3-pyruvate monooxygenase	Auxin biosynthesis	FA/SS
12	xyloglucan endotransglucosylase/hydrolase	Cell walls biogenesis and organisation	FA/SS
2	Chitinase 2	Cell walls catabolic process and defense response to fungus	FA/SS
1	Leucine-rich repeat receptor-like ser/thr protein kinase	Jasmonic acid and ethylene dependent systemic resistance	FA/SS
1	Transport proteins that can interact with LRRs	Jasmonic acid and ethylene dependent systemic resistance	FA/SS
2	Jasmonate O-methyltransferase	Jasmonic acid biosynthesis	FA/SS
1	Jasmonate-zim-domain protein	Jasmonic acid biosynthesis	FA/SS
1	12-oxophytodienoate reductase 2	Jasmonic acid biosynthesis	FA/SS
2	Salicylic acid-binding protein 2	Jasmonic acid biosynthesis	FA/SS
1	C-terminal LisH motif isoform 1 (synthesis TOPLESS prot)	Auxin biosynthesis	Pyrazines
3	Indole-3-glycerol phosphate synthase	Auxin biosynthesis	Pyrazines
1	Indole-3-pyruvate monooxygenase YUCCA6	Auxin biosynthesis	Pyrazines
4	Indole-3-acetaldehyde oxidase	Auxin biosynthesis	Pyrazines
1	Fumaryl-acetoacetate hydrolase	Cell death in response to jasmonic acid	Pyrazines
1	xyloglucan endotransglucosylase/hydrolase	Cell walls biogenesis and organisation	Pyrazines
1	1-aminocyclopropane-1-carboxylate synthase	Ethylene biosynthesis	Pyrazines
1	Proteasome subunit alpha type-7	Functioning of the proteasome	Pyrazines
1	Proteasome subunit alpha type-1-B	Functioning of the proteasome	Pyrazines
1	26S proteasome regulatory subunit	Functioning of the proteasome	Pyrazines
1	Gibberellin 3-beta-dioxygenase 1	Gibberellin signalling pathway	Pyrazines

N°: number; FA/SS: Fatty acid or sugar degradation compounds.

## 4. Discussion

It is the first time that integrative analysis of genetic, biochemical and sensorial fruity traits was conducted, leading to the identification of a few numbers of metabolomic pathways involved in fruity traits.

### 4.1. Insights into the genetic architecture of fruity aromas in cocoa

A large number of associations could be detected in relation to the different analysed traits related to fruity taste. Out of the thirteen sensory descriptors relating to fruity notes, associations were detected for only three of them. This is far less than what was detected for floral notes. (Colonges et al., 2021), detected associations for eleven of the sixteen floral descriptors obtained by sensory analysis. These results show that there is greater variability in the presence of floral notes than fruity notes in this population of modern Nacional.

The statistical power of GWAS may be somewhat limited due to the population size (152 clones). These clones come from genetic diversity conservation orchards and therefore do not have an optimal design for association genetics analyses. In order to improve future analyses, it would be interesting to plan several years of harvesting and analysis to confirm the results obtained. In addition, meta-analyses with other populations could also provide more statistical power.

However, a greater number of associations with volatile compounds known to have a fruity taste were detected. Possibly, these compounds are not in sufficient concentration to be statistically detected in the sensory analysis. Each compound would thus have a weak effect on the fruity note and in this case, they are not detectable by the GWAS method. In addition, some volatile compounds with a fruity taste can be synthesised by the micro-organisms present during fermentation (Schwan and Wheals, 2004). This is the case, for example, of esters, which are largely synthesised by fermentative yeasts (Ho et al., 2014; Soles et al., 1982). In our case, it is possible that these microorganisms played a role in the production of volatile aromatic compounds, but they are not the only ones responsible, since a large number of associations related to volatile compounds known to have a fruity taste have been detected. This means that the cocoa tree is also involved in the synthesis of these compounds in associations.

Of the 74 volatile compounds related to the fruity notes, association zones could not be detected for 13 of them. These compounds belong to different chemical families (two acids, three alcohols, one aldehyde, five esters, and two ketones). As these different compounds do not appear to be genetically related to the cocoa trees in this population, the micro-organisms present during fermentation likely synthesised them (Abbas, 2006; Ho et al., 2014; Soles et al., 1982).

### 4.2. Metabolomic pathways involved in fruity traits

Based on the previously presented results, five biosynthesis pathways appear to be involved in the synthesis of fruity compounds in cocoa. Two have been identified as being involved in the synthesis of floral compounds (Colonges et al., 2021): the monoterpene biosynthesis pathway and the L-phenylalanine degradation pathway. The three other biosynthesis pathways appear to be the pyrazine production pathway, the sugar degradation pathway and the fatty acid degradation pathway.

The pyrazine production seems to be achieved by the Maillard reaction during roasting. Indeed, ten of the eleven pyrazines were detected in cocoa beans after roasting. The eleventh pyrazine, tetramethylpyrazine, was certainly synthesised during drying also through the Maillard reaction (Starowicz and Zieliński, 2019). For this reaction to take place, free amino acids as well as reducing sugars are needed. This is why genes coding for proteinase or sugar degradation enzymes were found in the pyrazine association zones. As was observed for floral notes, co-locations between pyrazines known to have a fruity note and other pyrazines with no known notes suggest that a gene coding for a key enzyme involved in their biosynthesis could be responsible for the presence or absence of the

fruity note. This is the case, for example, of the co-locations between methylpyrazine, known to have a nutty note, and 2,3-dimethyl-5-ethylpyrazine, known to have burnt notes, on chromosomes 1, 4, and 6. It is possible that in these areas of the genome, the same enzyme is at the origin of their precursors. In the case where the enzyme is active, the precursors for pyrazine A would be synthesised and in the case where the enzyme is inactive, the precursors for pyrazine B would be present. If pyrazine A is the one bringing a dry fruit note then when the enzyme is activated, a dry fruit note is present; otherwise it is absent.

The pathways of sugar degradation and fatty acid degradation are linked and have several biochemical compounds in common. A large number of areas of the association have been detected for these many traits. With the large number of results and the multiple possibilities of compound synthesis, it is difficult to establish a hypothetical biosynthetic pathway in cocoa from these results. Furthermore, some intermediate compounds may be synthesised by the microorganisms present during fermentation. The biosynthesis of these compounds could be a synergy between the enzymatic actions of the microorganisms and the cocoa beans. Further studies on the presence of these volatile compounds with or without some micro-organisms (yeast or bacteria) could identify more precisely the biosynthetic mechanisms. However, the localisations observed between certain compounds suggest enzymatic actions initiated in the cocoa beans. This is the case, for example, on chromosome 1 where a co-location between the association linked to pentan-1-ol (R) and pentanoic acid (R) is observed in the haplotypic block number 18 (5 356 899 pb–5 820 933 pb). In this case, a cocoa enzyme could be responsible for the oxidation of the alcohol pentan-1-ol leading to the carboxylic acid: pentanoic acid. A candidate gene "Acyl-coenzyme A oxidase 2, peroxisomal" encoding an enzyme with an oxidative function was found at position 5 520 690 on chromosome 1 next to the haplotypic block of the associations (figure A.3).

The fruity aroma of Nacional could also include compounds belonging to the L-phenylalanine degradation pathway such as benzaldehyde, benzyl alcohol and 1-phenylethyl acetate, known to have fruity flavours. Associations have been detected for these three compounds and exposed in a previous study (Colonges et al., 2021). These compounds are also precursors of compounds known to have floral tastes. The enzymatic activity allowing the degradation of benzaldehyde or benzyl alcohol is one of the keys to the increased presence of compounds with fruity or floral tastes.

## 5. Conclusions and perspectives

The study of the determinism of aromas is complex and involves several disciplines. No biosynthesis pathway produces only one of the compounds with the same type of taste. It is the balance between all the compounds that allow the synthesis of an aromatic profile and that of Nacional cocoa is complex. The presence of an aromatic compound is not necessarily synonymous with the perception of its aromatic note. Identification of the volatile compounds with a perceived fruity note could be carried out thanks to a gas chromatography analysis coupled with olfactometry. With this approach, it would be possible to identify the molecules whose fruity notes are perceived and thus select the candidate genes favourable alleles associated with these compounds, providing new tools for marker-assisted selection.

### Data availability

The SNPs used in this study are available at <https://tropgenedb.cirad.fr/tropgene/JSP/interface.jsp?module=COCOA> in a file named *Cocoa\_Nacional\_aroma.zip* containing: *GBS\_Nac\_G7.xls* and *GBS\_Nac\_MAF5.xls*.

### Author contributions

**Kelly Colonges:** Formal analysis, Investigation, Visualization and

Writing - original draft. **Juan-Carlos Jimenez:** Data curation, Methodology, Investigation. **Alejandra Saltos:** Data curation, Methodology, Investigation. **Edward Seguine:** Data curation, Methodology. **Rey Gastón Loor Solorzano:** Conceptualization, Funding acquisition. **Olivier Fouet:** Data curation, Methodology. **Xavier Argout,** Methodology. **Sophie Assemat:** Supervision. **Fabrice Davrieux:** Supervision. **Emile Cros:** Conceptualization, Funding acquisition. **Claire Lanaud:** Project administration, Supervision, Conceptualization, Funding acquisition, Validation and Writing - review & editing. **Renaud Boulanger:** Project administration, Supervision, Validation and Writing - review & editing.

### Funding

This work was supported by the United States Department of State (U.S. Foreign Ministry), the U.S. Embassy, Quito and the U.S. Department of Agriculture (USDA-ARS) [ 58-4001-2-F128] and the MUSE Amazcacao project [ANR-16-IDEX-0006].

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgement

We thank the USDA and the I-Site MUSE 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.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.plaphy.2021.11.006>.

### References

- Abbas, C.A., 2006. Production of antioxidants, aromas, colours, flavours, and vitamins by yeasts. In: *Yeasts in Food and Beverages*. Springer, Berlin, Heidelberg, pp. 285–334. [https://doi.org/10.1007/978-3-540-28398-0\\_10](https://doi.org/10.1007/978-3-540-28398-0_10).
- Aranda, D.A.G., Santos, R.T.P., Tapanes, N.C.O., Ramos, A.L.D., Antunes, O.A.C., 2008. Acid-catalyzed homogeneous esterification reaction for biodiesel production from palm fatty acids. *Catal. Lett.* 122, 20–25. <https://doi.org/10.1007/s10562-007-9318-z>.
- 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 Genom.* 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., Kudrna, D., Ammiraju, J.S.S., Schuster, S.C., Carlson, J.E., Sallet, E., Schiex, T., Dievart, A., Kramer, M., Gelle, L., Shi, Z., Bérard, A., Viot, C., Boccara, M., Risterucci, A.M., Guignon, V., Sabau, X., Axtell, M.J., Ma, Z., Zhang, Y., Brown, S., Bourge, M., Golsner, W., Song, X., Clement, D., Rivallan, R., Tahiri, M., Akaza, J.M., Pitollat, B., Gramacho, K., D'Hont, A., Brunel, D., Infante, D., Kebe, I., Costet, P., Wing, R., McCombie, W.R., Guiderdoni, E., Quetier, F., Panaud, O., Wincker, P., Bocs, S., Lanaud, C., 2011. The genome of *Theobroma cacao*. *Nat. Genet.* 43, 101. <https://doi.org/10.1038/ng.736>.
- Arn, H., Acree, T.E., 1998. Flavornet: a database of aroma compounds based on odor potency in natural products. *Dev. Food Sci., Food Flavors: Formation, Analysis and Packaging Influences* 40, 27. [https://doi.org/10.1016/S0167-4501\(98\)80029-0](https://doi.org/10.1016/S0167-4501(98)80029-0).
- Arnoldi, A., Arnoldi, C., Baldi, O., Griffini, A., 1988. Flavor components in the Maillard reaction of different amino acids with fructose in cocoa butter-water. *Qualitative and quantitative analysis of pyrazines*. *J. Agric. Food Chem.* 36, 988–992.
- 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. *Eur. Food Res. Technol.* 245, 1459–1471. <https://doi.org/10.1007/s00217-018-3181-6>.



- Aznar, M., López, R., Cacho, J.F., Ferreira, V., 2001. Identification and quantification of impact odorants of aged red wines from rioja. GC–Olfactometry, quantitative GC–MS, and odor evaluation of HPLC fractions. *J. Agric. Food Chem.* 49, 2924–2929. <https://doi.org/10.1021/jf001372u>.
- Baek, H.H., Cadwallader, K.R., Marroquin, E., Silva, J.L., 1997. Identification of predominant aroma compounds in muscadine grape juice. *J. Food Sci.* 62, 249–252. <https://doi.org/10.1111/j.1365-2621.1997.tb03978.x>.
- Bayer, C., Kubitzki, K., 2003. Malvaceae. In: *The Families and Genera of Vascular Plants. Dicotyledons: Malvales, Capparales and Non-betain Caryophyllales*. Springer, Berlin, pp. 225–311.
- Beckett, S.T., Fowler, M., Ziegler, G.R., 2017. *Beckett's Industrial Chocolate Manufacture and Use*, fifth ed. Wiley-Blackwell.
- Boza, E.J., Motamayor, J.C., Amores, F.M., Cedeño-Amador, S., Tondo, C.L., Livingstone, D.S., Schnell, R.J., Gutiérrez, O.A., 2014. Genetic characterization of the cacao cultivar CCN 51: its impact and significance on global cacao improvement and production. *J. Am. Soc. Hortic. Sci.* 139, 219–229. <https://doi.org/10.21273/JASHS.139.2.219>.
- Cheesman, E., 1944. Notes on the nomenclature, classification and possible relationships of cacao populations. *Trop. Agric.* 21, 144–159.
- Chen, X., Yauk, Y.-K., Nieuwenhuizen, N.J., Matic, A.J., Wang, M.Y., Perez, R.L., Atkinson, R.G., Beuning, L.L., 2010. Characterisation of an (S)-linalool synthase from kiwifruit (*Actinidia arguta*) that catalyses the first committed step in the production of floral lilac compounds. *Funct. Plant Biol.* 37, 232–243. <https://doi.org/10.1071/FP09179>.
- Colahan-Sederstrom, P.M., Peterson, D.G., 2005. Inhibition of key aroma compound generated during ultrahigh-temperature processing of bovine milk via epicatechin addition. *J. Agric. Food Chem.* 53, 398–402. <https://doi.org/10.1021/jf0487248>.
- Colonges, K., Jimenez, J.C., Saltos, A., Seguíne, E., Solorzano, R.L., Fouet, O., Argout, X., Assemat, S., Davrieux, F., Cros, E., Boulanger, R., Lanaud, C., 2021. Two main biosynthesis pathways involved in the synthesis of the floral aroma of the Nacional cocoa variety. *Front. Plant Sci., Crop and Product Physiology online*. <https://doi.org/10.3389/fpls.2021.681979>, 2064–2064.
- Cseke, L., Dudareva, N., Pichersky, E., 1998. Structure and evolution of linalool synthase. *Mol. Biol. Evol.* 15, 1491–1498. <https://doi.org/10.1093/oxfordjournals.molbev.a025876>.
- Dzalian, M.C., Park, R., Steensels, J., Lievens, B., Verstrepen, K.J., 2017. Physiology, ecology and industrial applications of aroma formation in yeast. *FEMS Microbiol. Rev.* 41, S95–S128. <https://doi.org/10.1093/femsre/fux031>.
- Ferreira, V., López, R., Escudero, A., Cacho, J.F., 1997. The aroma of Grenache red wine: hierarchy and nature of its main odorants. *J. Sci. Food Agric.* 77, 259–267. [https://doi.org/10.1002/\(SICI\)1097-0010\(199707\)77:2<259::AID-JSFA36>3.0.CO;2-Q](https://doi.org/10.1002/(SICI)1097-0010(199707)77:2<259::AID-JSFA36>3.0.CO;2-Q), 199806.
- Fickert, B., Schieberle, P., 1998. Identification of the key odorants in barley malt (caramalt) using GC/MS techniques and odour dilution analyses. *Food Nahr* 42, 371–375. [https://doi.org/10.1002/\(SICI\)1521-3803\(199812\)42:06<371::AID-FOOD371>3.0.CO;2-V](https://doi.org/10.1002/(SICI)1521-3803(199812)42:06<371::AID-FOOD371>3.0.CO;2-V).
- 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. *Genet. Epidemiol.* 34. <https://doi.org/10.1002/gepi.20430>, 100–100.
- Gao, X., Starmer, J., Martin, E.R., 2008. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet. Epidemiol.* 32, 361–369. <https://doi.org/10.1002/gepi.20310>. In press.
- Garg, N., Sethupathy, A., Tuwani, R., Nk, R., Dokania, S., Iyer, A., Gupta, A., Agrawal, S., Singh, N., Shukla, S., Kathuria, K., Badhwar, R., Kanji, R., Jain, A., Kaur, A., Nagpal, R., Bagler, G., 2018. FlavorDB: a database of flavor molecules. *Nucleic Acids Res.* 46, D1210–D1216. <https://doi.org/10.1093/nar/gkx957>.
- Genovese, A., Gambuti, A., Piombino, P., Moio, L., 2007. Sensory properties and aroma compounds of sweet Fiano wine. *Food Chem.* 103, 1228–1236. <https://doi.org/10.1016/j.foodchem.2006.10.027>.
- Guichard, H., Lemesle, S., Ledauphin, J., Barillier, D., Picoche, B., 2003. Chemical and sensorial aroma characterization of freshly distilled calvados. 1. evaluation of quality and defects on the basis of key odorants by olfactometry and sensory analysis. *J. Agric. Food Chem.* 51, 424–432. <https://doi.org/10.1021/jf020372m>.
- Ho, V.T.T., Zhao, J., Fleet, G., 2014. Yeasts are essential for cocoa bean fermentation. *Int. J. Food Microbiol.* 174, 72–87. <https://doi.org/10.1016/j.ijfoodmicro.2013.12.014>.
- Holmquist, M., 2000. Alpha beta-hydrolase fold enzymes structures, functions and mechanisms. *Curr. Protein Pept. Sci.* 1, 209–235. <https://doi.org/10.1021/1389203003381405>.
- International Cocoa Organization, 2017. *Fine or Flavour Cocoa*.
- ISCF, 2020. *First Draft of the Protocol for Cocoa Liqueur Sensory Evaluation: Part of the International Standards for the Assessment of Cocoa Quality and Flavour (ISCF)*. Compiled by Bioversity International, in Collaboration with the Members of the ISCF Working Group.
- Ito, Y., Sugimoto, A., Kakuda, T., Kubota, K., 2002. Identification of potent odorants in Chinese jasmine green tea scented with flowers of *Jasminum sambac*. *J. Agric. Food Chem.* 50, 4878–4884. <https://doi.org/10.1021/jf020282h>.
- Jezussek, M., Juliano, B.O., Schieberle, P., 2002. Comparison of key aroma compounds in cooked brown rice varieties based on aroma extract dilution analyses. *J. Agric. Food Chem.* 50, 1101–1105. <https://doi.org/10.1021/jf0108720>.
- Kalua, C.M., Allen, M.S., Bedgood, D.R., Bishop, A.G., Prenzler, P.D., Robards, K., 2007. Olive oil volatile compounds, flavour development and quality: a critical review. *Food Chem.* 100, 273–286. <https://doi.org/10.1016/j.foodchem.2005.09.059>.
- Karagül-Yüceer, Y., Vlahovich, K.N., Drake, M., Cadwallader, K.R., 2003. Characteristic aroma components of rennet casein. *J. Agric. Food Chem.* 51, 6797–6801. <https://doi.org/10.1021/jf0345806>.
- 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: Pompanon, F., Bonin, A. (Eds.), *Data Production and Analysis in Population Genomics: Methods and Protocols, Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp. 67–89. [https://doi.org/10.1007/978-1-61779-870-2\\_5](https://doi.org/10.1007/978-1-61779-870-2_5).
- Kumazawa, K., Masuda, H., 2002. Identification of potent odorants in different green tea varieties using flavor dilution technique. *J. Agric. Food Chem.* 50, 5660–5663. <https://doi.org/10.1021/jf020498j>.
- Lachenaud, P., Motamayor, J.C., 2017. The Criollo cacao tree (*Theobroma cacao* L.): a review. *Genet. Resour. Crop Evol.* 64, 1807–1820. <https://doi.org/10.1007/s10722-017-0563-8>.
- Larsen, M., Poll, L., 1992. Odour thresholds of some important aroma compounds in strawberries. *Z. Für Lebensm.-Unters. Forsch.* 195, 120–123. <https://doi.org/10.1007/BF01201770>.
- Loor, S.R.G., 2007. Contribution à l'étude de la domestication de la variété de cacaoyer Nacional d'Equateur : recherche de la variété native et de ses ancêtres sauvages. *INIAP Archivo Historico*.
- 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 Genet. Genomes* 5, 421–433. <https://doi.org/10.1007/s11295-008-0196-3>.
- Luna, F., Cruzillat, D., Cirou, L., Bucheli, P., 2002. Chemical composition and flavor of Ecuadorian cocoa liquor. *J. Agric. Food Chem.* 50, 3527–3532.
- Mahajan, S.S., Goddik, L., Qian, M.C., 2004. Aroma compounds in sweet whey powder. *J. Dairy Sci.* 87, 4057–4063. [https://doi.org/10.3168/jds.S0022-0302\(04\)73547-X](https://doi.org/10.3168/jds.S0022-0302(04)73547-X).
- Meesters, R.J.W., Duisken, M., Hollender, J., 2007. Study on the cytochrome P450-mediated oxidative metabolism of the terpene alcohol linalool: indication of biological epoxidation. *Xenobiotica Fate Foreign Compd. Biol. Syst.* 37, 604–617. <https://doi.org/10.1080/00498250701393191>.
- Mindrebo, J.T., Nartey, C.M., Seto, Y., Burkart, M.D., Noel, J.P., 2016. Unveiling the functional diversity of the Alpha-Beta hydrolase fold in plants. *Curr. Opin. Struct. Biol.* 41, 233–246. <https://doi.org/10.1016/j.sbi.2016.08.005>.
- Motamayor, J.C., Lachenaud, P., Mota, J.W., da, S. e, 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, e3311. <https://doi.org/10.1371/journal.pone.0003311>.
- Motamayor, J.C., Mockaitis, K., Schmutz, J., Haiminen, N., III, D.L., Cornejo, O., Findley, S.D., Zheng, P., Utro, F., Royraert, S., Sasaki, C., Jenkins, J., Podicheti, R., Zhao, M., Scheffler, B.E., Stack, J.C., Feltus, F.A., Mustiga, G.M., Amores, F., Phillips, W., Marelli, J.P., May, G.D., Shapiro, H., Ma, J., Bustamante, C.D., Schnell, R.J., Main, D., Gilbert, D., Parida, L., 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 Biol.* 14, r53. <https://doi.org/10.1186/gb-2013-14-6-r53>.
- Perestrelo, R., Fernandes, A., Albuquerque, F.F., Marques, J.C., Câmara, J.S., 2006. Analytical characterization of the aroma of Tinta Negra Mole red wine: identification of the main odorants compounds. *Anal. Chim. Acta, Papers presented at the 4th SYMPOSIUM IN VINO ANALYTICA SCIENTIA* 563, 154–164. <https://doi.org/10.1016/j.aca.2005.10.023>.
- Pham, A.J., Schilling, M.W., Yoon, Y., Kamadia, V.V., Marshall, D.L., 2008. Characterization of fish sauce aroma-impact compounds using GC-MS, SPME-OsmegCO, and Stevens' power Law Exponents. *J. Food Sci.* 73, C268–C274. <https://doi.org/10.1111/j.1750-3841.2008.00709.x>.
- Risterucci, A.M., Grivet, L., N'Goran, J.A.K., Pieretti, I., Flament, M.H., Lanaud, C., 2000. A high-density linkage map of *Theobroma cacao* L. *Theor. Appl. Genet.* 101, 948–955. <https://doi.org/10.1007/s001220051566>.
- Rottiers, H., Tzompa Sosa, D.A., Lemarcq, V., De Winne, A., De Wever, J., Everaert, H., Bonilla Jaime, J.A., Dewettinck, K., Messens, K., 2019. A multipronged flavor comparison of Ecuadorian CCN51 and Nacional cocoa cultivars. *Eur. Food Res. Technol.* 245, 2459–2478. <https://doi.org/10.1007/s00217-019-03364-3>.
- 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. In: *15th International Cocoa Research Conference : Cocoa Productivity, Quality, Profitability, Human Health and the Environment*. Presented at the 15th International Cocoa Research Conference. Cocoa Producers' Alliance.
- Schieberle, P., Ofner, S., Grosch, W., 1990. Evaluation of potent odorants in Cucumbers (*Cucumis sativus*) and Muskmelons (*Cucumis melo*) by aroma extract dilution analysis. *J. Food Sci.* 55, 193–195. <https://doi.org/10.1111/j.1365-2621.1990.tb06050.x>.
- Schwan, R.F., Wheals, A.E., 2004. The microbiology of cocoa fermentation and its role in chocolate quality. *Crit. Rev. Food Sci. Nutr.* 44, 205–221. <https://doi.org/10.1080/10408690490464104>.
- Soles, R.M., Ough, C.S., Kunkel, R.E., 1982. Ester concentration differences in wine fermented by various species and strains of yeasts. *Am. J. Enol. Vitic.* 33, 94–98.
- Song, S., Qi, T., Wasternack, C., Xie, D., 2014. Jasmonate signaling and crosstalk with gibberellin and ethylene. *Curr. Opin. Plant Biol.* 15, C1. <https://doi.org/10.1016/j.pbi.2014.07.005>.
- Starowicz, M., Zieliński, H., 2019. How Maillard reaction influences sensorial properties (color, flavor and texture) of food products? *Food Rev. Int.* 35, 707–725. <https://doi.org/10.1080/87559129.2019.1600538>.
- Sukha, D.A., Butler, D.R., Umaharan, P., Boulte, 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. *Eur. Food Res. Technol.* 226, 405–413.



- Sun, X., Shen, X., Jain, R., Lin, Y., Wang, Jian, Sun, J., Wang, Jia, Yan, Y., Yuan, Q., 2015. Synthesis of chemicals by metabolic engineering of microbes. *Chem. Soc. Rev.* 44, 3760–3785. <https://doi.org/10.1039/C5CS00159E>.
- Wang, X., Fan, W., Xu, Y., 2014. Comparison on aroma compounds in Chinese soy sauce and strong aroma type liquors by gas chromatography–olfactometry, chemical quantitative and odor activity values analysis. *Eur. Food Res. Technol.* 239, 813–825. <https://doi.org/10.1007/s00217-014-2275-z>.
- Zhou, Z., Zhi, T., Han, C., Peng, Z., Wang, R., Tong, J., Zhu, Q., Ren, C., 2020. Cell death resulted from loss of fumarylacetoacetate hydrolase in Arabidopsis is related to phytohormone jasmonate but not salicylic acid. *Sci. Rep.* 10, 13714. <https://doi.org/10.1038/s41598-020-70567-0>.
- Ziegler, G., 1990. Linalool contents as characteristic of some flavor grade cocoas. *Z. Für Lebensm.-Unters. Forsch.* 191, 306–309. <https://doi.org/10.1007/BF01202432>.