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

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 (K. Colonges), 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
Published by Elsevier Masson SAS. This is an open access article under the CC BY license

https://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; Ziegleder, 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 (ET-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 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. The aroma of the modern 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; Ziegleder, 1990). The aroma of the

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” (INAP) 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 replications per sample by Edward Seguine. Samples were prepared and then evaluated using the protocols in the ISQCF 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.
List of biochemical compounds related to fruity traits and used for the GWAS analysis of unroasted (UR) and roasted (R) beans.

| Biochemical traits          | Compound family | Path | Aroma                                      |
|-----------------------------|----------------|------|--------------------------------------------|
| Acetic acid                 | Acid           | FA   | Vinegar (Ferreira et al., 1997; Mahajan et al., 2004) |
| Pentanoic acid              | Acid           | FA   | Unpleasant, acre, cheese, sweat (Genovese et al., 2007; Jezunek et al., 2002; Kalua et al., 2007; Mahajan et al., 2004) |
| 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; Jezunek et al., 2002; Mahajan et al., 2004; Perestrello et al., 2006) |
| 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) |
| Nonanoic acid               | Acid           | FA   | Acid (Colahan-Sederstrom and Peterson, 2005) |
| Ethanol                     | Alcohol        | SS   | Alcohol (Kalua et al., 2007) |
| Pentan-1-ol                 | Alcohol        | FA   | Fruity (Kalua et al., 2007) |
| 2-methyl-2,6-pentane alcohol| Alcohol        | SS   | Green grass (Guichard et al., 2003) |
| 2-methylhexan-2-ol           | Alcohol        | SS   | Cucumber (Arn and Acree, 1998) |
| 3-methylbutan-1-ol           | Alcohol        | SS   | Plastic, fuel oil, whiskey characteristic pungent (Guichard et al., 2003) |
| 2-methylbutan-2-ol           | Alcohol        | SS   | Fruity, solvent, apple, sweet (Ferreira et al., 1997; Genovese et al., 2007; Karagül-Yüceer et al., 2003) |
| 2-ethylhexan-1-ol            | Alcohol        | SS   | Stimulant, malted (Kumazawa and Masuda, 2002) |
| Butan-2,3-diol               | Alcohol        | FA   | Fruit, onions (Arn and Acree, 1998) |
| Butane-1,3-diol              | Alcohol        | SS   | |
| Cis-hept-4-ene-1-ol          | Alcohol        | SS   | |
| Benzy alcohol               | Alcohol        | L-phe| Fruity (Ito et al., 2002) |
| Acetalddehyde               | Aldehyde       | SS   | Acre, other (Arn and Acree, 1998; Wang et al., 2014) |
| 2-methylpropanal            | Aldehyde       | SS   | Acre, malted, green (Arn and Acree, 1999) |
| 2-methylbutanal             | Aldehyde       | SS   | Malted (Fickert and Schieberle, 1998) |
| 3-methylbutanal             | Aldehyde       | SS   | Stimulant, malted (Kumazawa and Masuda, 2002) |
| Nonanal                     | Aldehyde       | Other| Orange-like, floral, soapy (Karagül-Yüceer et al., 2003; Kumazawa and Masuda, 2002; Mahajan et al., 2004) |
| 2-phenylbut-2-enal           | Aldehyde       | SS   | |
| 5-methyl-2-phenylhex-2-enal  | Aldehyde       | SS   | Cacao (Garg et al., 2018) |
| Benzaldehyde                | Aldehyde       | L-phe| Bitter, cherry, almond, fruity (Perestrello et al., 2006; Pham et al., 2008; Wang et al., 2014) |
| Methyl acetate              | Ester          | SS   | Fruity, orange, wax, wine (Garg et al., 2018) |
| Ethyl acetate               | Ester          | FA   | Fruity, solvent, apple, sweet (Ferreira et al., 1997; Genovese et al., 2007; Karagül-Yüceer et al., 2003) |
| Ethyl propanoate            | Ester          | FA   | Fruity, strong, strawberry (Kalua et al., 2007) |
| 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) |
| 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) |
| Ethyl 3-hydroxyhexanoate     | Ester          | SS   | Strawberry (Genovese et al., 2007) |
| Ethyl octanoate             | Ester          | FA   | Sweet, fruity, fresh, pineapples (Genovese et al., 2007; Perestrello et al., 2006) |
| Ethyl 2-hexadecanoate       | Ester          | SS   | Floral, fruity, leaf (Garg et al., 2018) |
| Ethyl benzoate              | Ester          | SS   | |
| Ethyl 2-methylpropanal      | Ester          | FA   | Fruity, sweet, rubber (Arn and Acree, 1998; Schieberle et al., 1990) |
| Ethyl 2-methylbutanoate      | Ester          | FA   | Green apple, fruity, red fruit (Baek et al., 1997; Genovese et al., 2007) |
| Butyl benzoate              | Ester          | FA   | |
| Propyl acetate              | Ester          | Other| Celery, floral, pear, red fruit (Garg et al., 2018) |
| Hexyl acetate               | Ester          | Other| Fruity, pear, sweet and floral (Guichard et al., 2003; Wang et al., 2014) |
| 2-methylpropyl acetate      | Ester          | FA   | Strawberry (Anzar et al., 2001) |
| 1-methylbutyl acetate       | Ester          | SS   | Banana (Genovese et al., 2007; Guichard et al., 2003) |
| 3-methylbutyl acetate       | Ester          | FA   | Banana (Genovese et al., 2007; Guichard et al., 2003) |
| Butan-2,3-diy diacetate     | Ester          | SS   | |
| Meso butan-2,3-di-y diacetate| Ester          | SS   | |
| 3-methyl-but-2-enyl acetate | Ester          | SS   | Jasmin, banana, fresh, ripe, heliotrope, sweet, fruity, balsam, lavender (Garg et al., 2018) |
| 1-methylhexylene acetate    | Ester          | SS   | Fruity (Garg et al., 2018) |
| 1-phenylethyl acetate       | Ester          | L-phe| Fruity (Garg et al., 2018) |
| Furfural                    | Furan          | Other| Incense, fruity, flowery, roasted, sweet and almond (Colahan-Sederstrom and Peterson, 2005; Ferreira et al., 1997; Wang et al., 2014) |
| Acetone                     | Ketone         | ?    | Acre (Garg et al., 2018) |
| Butan-2-one                 | Ketone         | SS   | Fruity, ethereal (Kalua et al., 2007) |
| 3-hydroxybutan-2-one        | Ketone         | ?    | Butter, cream (Arn and Acree, 1998) |

(continued on next page)
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 dependent tests (Meff) used as the denominator and calculated by

All of these thirteen sensory traits were used to carry out this study (Table 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. 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 A.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 monoterpenic biosynthesis pathway.

The result of a PCA for volatile compounds in unroasted 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 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.
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).

Repeateable 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 methlypyrazine (R). Six of the most significant association zones for these compounds are located outside the calculated haploacypic 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 methlypyrazine (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 "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

| 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).
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), monoterpenic biosynthesis and L-phenylalanine degradation pathways are involved in the synthesis of fruity notes. Two volatile compounds with fruity notes belonging to the monoterpenic 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 could 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.

3.3. Candidate genes 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.

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    | 14                    | /             | 2-methylbutanal (R) |
| 1          | 4 522    | 23                    | /             | 5-methyl-2-phenylhexan-2-enal (R) |
| 1          | 6 281    | 26                    | /             | pentanoic acid (R) |
| 1          | 6 855    | 91                    | /             | 3-methylbutan-1-ol (R) acetone (R) |
| 2          | 8 389    | NA                    | /             | Heptan-2-ol R |
| 2          | 11 233   | NA                    | /             | 3-Methyl butyl acetate-UR |
| 2          | 25 201   | 65                    | /             | 3-hydroxy-butan-2-one (UR) |
| 2          | 5 956    | 24                    | /             | 2-methylbutan-1-ol (UR) |
| 2          | 5 607    | 373                   | /             | ethyl benzoate (UR) ethyl butanoate (R) |
| 2          | 5 607    | 523                   | /             | hexan-2-one (UR) pentan-2,3-dione (R) |
| 2          | 5 607    | 833                   | /             | 2-methylbutan-1-ol (UR) |
| 2          | 5 607    | 377                   | /             | meso-2,3-butan-diyl diacetate (UR) |
| 2          | 5 607    | 398                   | /             | heptan-2-one (UR) |
| 2          | 5 607    | 204                   | /             | hydroxy-butan-2-one (UR) |
| 2          | 5 607    | 204                   | /             | nonan-2-ol (UR) |
| 2          | 5 607    | 398                   | /             | 2-methylbutan-2-ol (UR) |
| 2          | 5 607    | 398                   | /             | nonan-2-ol (UR) |
| 2          | 5 607    | 398                   | /             | meso-2,3-butan-diyl diacetate (UR) |
| 2          | 5 607    | 398                   | /             | methyl acetate (UR) |
| 7          | 10 459   | 31                    | /             | hexyl acetate (UR) |
| 7          | 1 099    | 5                     | /             | pentanoic acid (UR) 2-ethylhexan-1-ol (R) |
| 9          | 3 755    | 54                    | /             | 2-methylpropanal (R) methyl acetate (R) nonanal (UR) |
| 9          | 3 755    | 266                   | /             | 2-methylbutan-2-ol (UR) 2-phenylbut-2-enal (R) 3-methylbutanal (R) octanoic acid (R) and ethyl butanoate (R) |
| 9          | 3 755    | 289                   | /             | 2-methylbutan-2-ol (UR) 2-phenylbut-2-enal (R) 3-methylbutanal (R) octanoic acid (R) and ethyl butanoate (R) |
| 1          | 3 688    | 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          | 3 688    | 6                     | Dark Tree fruit | 2-phenylbut-2-enal (R) 3-methylbutanal (R) and ethyl butanoate (R) |
| 1          | 3 688    | 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 |
|------------|----------|-----------------------|---------------|-------------------|
| 1          | 3 688    | 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          | 1 099    | 5                     | /             | pentanoic acid (UR) 2-ethylhexan-1-ol (R) |
| 1          | 3 755    | 54                    | /             | 2-methylpropanal (R) methyl acetate (R) nonanal (UR) |
| 1          | 3 755    | 266                   | /             | 2-methylbutan-2-ol (UR) 2-phenylbut-2-enal (R) 3-methylbutanal (R) octanoic acid (R) and ethyl butanoate (R) |
| 1          | 3 755    | 289                   | /             | 2-methylbutan-2-ol (UR) 2-phenylbut-2-enal (R) 3-methylbutanal (R) octanoic acid (R) and ethyl butanoate (R) |
| 1          | 3 688    | 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          | 3 688    | 6                     | Dark Tree fruit | 2-phenylbut-2-enal (R) 3-methylbutanal (R) and ethyl butanoate (R) |
| 1          | 3 688    | 7                     | Dark Tree fruit | 2-phenylbut-2-enal (R) 3-methylbutanal (R) octanoic 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 (Czeke 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 acids, Sugars degradations | Pyrazine degradation |
| 17              | Protease function | Proteins | Pyrazine production |
| 16              | Peptidase function | Proteins | Pyrazine production |
| 11              | Glutathione S-transferase | Proteins | Pyrazine production |
| 5               | Oxidase/reductase or hydrolase function | Proteins, Fatty acids, Sugars degradations | Pyrazine production |
| 4               | Proteinase function | Proteins | Pyrazine production |
| 3               | Aminotransferase function | Proteins | Pyrazine production |
| 3               | Amino acid decarboxylase function | Proteins | Pyrazine production |
| 1               | Delta-1-pyrroline-5-carboxylate synthase | Proteins | Pyrazine production |
| 1               | Anthranilate synthase | Proteins | Pyrazine production |
| 1               | Bifunctional aspartokinase/homoserine dehydrogenase | Proteins | Pyrazine production |
| 1               | Lactoylglutathione lyase | Proteins | Pyrazine production |
| 1               | Peptidyl-prolyl cis-trans isomerase | Proteins | Pyrazine production |
| 1               | Protein S-acetyltransferase 16 | Proteins | Pyrazine production |
| 1               | Peptide-N4-(N-acetyl-beta-glycosaminy1) asparagine amidase | Proteins | Pyrazine production |
| 1               | Endoglucanase | Sugars | Pyrazine production |
| 1               | Xyloglucan endotransglycosylase/hydrolase | Sugars | Pyrazine production |
| 1               | Inositol oxygenase 1 | Sugars | Pyrazine production |
| 4               | Carboxylesterase | Fatty acid | Pyrazine production |
| 1               | Isopenylcysteine alpha-carbonyl methyltransferase | Fatty acid | Pyrazine production |
| 1               | Phospholipase | Fatty acid | Pyrazine production |
| 1               | Branched-chain-amino-acid aminotransferase | Amino acid | Pyrazine production |
| 1               | N-acetyltransferase | Amino acid | Pyrazine production |
| 1               | Peptidyl prolyl cis-trans isomerase | Amino acid | Pyrazine production |
| 1               | Arabinosyltransferase | Amino acid | Pyrazine production |
| 2               | Methyltransferase | Amino acid | Pyrazine production |
| 1               | Malonate CoA ligase | Carboxylic acids degradations | Pyrazine production |
| 1               | Methylmalonate semialdehyde dehydrogenase | Carboxylic acids degradations | Pyrazine production |
| 2               | Malate dehydrogenase | Carboxylic acids degradations | Pyrazine production |
| 30              | Alpha-beta hydrolase | Fatty acid and sugars degradations | Pyrazine production |
| 8               | Hydrolase | Fatty acid and sugars degradations | Pyrazine production |
| 1               | 3-ketoacyl-CoA thiolase | Fatty acid and sugars degradations | Pyrazine production |
| 1               | Plastidial glycolate/glycerate translocator | Fatty acid and sugars degradations | Pyrazine production |
| 1               | Enoyl CoA hydratase | Fatty acid and sugars degradations | Pyrazine production |

Table 4 (continued)

| Number of genes | Gene function | General function | Pathway |
|-----------------|---------------|------------------|---------|
| 1               | Non-specific lipid transfer | Pyrazine production | 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               | 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               | Xylose | 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).

Ethylene is also a phytohormone that identifies in Arabidopsis Thaliana as affecting cell death in response to the ethylene dependent systemic resistance pathway. 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).

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 Zielinski, 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

Table 5

| N°  | Gene function | Role in the plant defences | Type of associated compounds |
|-----|---------------|---------------------------|-----------------------------|
| 1   | Indole-3-pyruvate monooxygenase | Auxin biosynthesis | FA/SS |
| 12  | xyloglucan endotransglycosylase/hydrolase | Cell walls biogenesis and organization | FA/SS |
| 2   | Chitinase 2 | Cell walls catabolism process and defense response to fungus | FA/SS |
| 1   | Leucine-rich repeat receptor-like ser/thre protein kinase | Jasmonic acid and ethylene dependent systemic resistance | FA/SS |
| 1   | Transport proteins that can interact with LRPs | 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 isofrom 1 (synthesis of TOPLESS prot) | Auxin biosynthesis | Pyrazines |
| 3   | Indole-3-glycerol phosphate synthase | Auxin biosynthesis | Pyrazines |
| 1   | Indole-3-pyruvate monooxygenase | 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 organization | Pyrazines |
| 1   | 1-aminocyclop propane-1-carboxylic acid 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.
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 co-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 determination 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://tropgene.cirad.fr/tropgene/JSJ/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 Gaston 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 Service), 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

Abbot, C.A., 2006. Production of antioxidans, 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_16.

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/s10521-007-9318-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, 720. https://doi.org/10.1186/s12864-017-4120-9.

Argout, X., Salse, J., Aury, J.-M., Guillinat, M.J., Droc, G., Gouzy, J., Allegre, M., Chaparro, C., Legavre, T., Maximova, S.N., Abrous, M., Murat, F., Fouet, O., Poulin, J., Ruiz, M., Roget, Y., Roudier-Goud, M., Barbona-Neto, J.F., Sabot, F., Kudrna, D., Ammiraju, J.S.S., Schuster, S.C., Carlon, J.E., Sallet, E., Schiex, T., Dievart, A., Kramer, M., Gelley, I., Shi, Z., Bérard, A., Viot, C., Boccazzi, M., Risterucci, A.M., Guignon, V., Sabau, X., Astell, M.J., Ma, Z., Zhang, Y., Brown, S., Bourge, M., Goulie, W., Song, X., Clement, D., Rivallan, R., Tabi, M., Akaza, J.M., Pitallat, 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., Winkler, P., Bocq, S., Lanaud, C., 2011. The genome of Thesmophora 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/10.1007/s10562-007-9880-9.

Arnold, 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 raw cocoa and sensory attributes of chocolate produced thereof. Eur. Food Res. Technol. 245, 1459-1471. https://doi.org/10.1007/s00217-018-3181-x.
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/s00217-014-2279-2.

Ziegleder, G., 1990. Linalool contents as characteristic of some flavor grade cocoas. Z. Für Lebensm.-Unters. Forsch. 191, 306–309. https://doi.org/10.1007/BF01268432.

Zhou, Z., Zhi, T., Han, C., Peng, Z., Wang, R., Tong, J., Zhu, Q., Ren, C., 2020. Cell death resulted from loss of fumarylacetooacetate 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.