Doctoral Dissertation

Characterization of fine flavor cocoa in parent hybrids combination using metabolomic approach

Enik Nurlaili Afifah

June 2025

Graduate School of Engineering

Osaka University

List of Abbreviations

(*In alphabetical order*)

a*	Green-to-red
ANOVA	Analysis of variance
b*	Blue-to-yellow
DVB/CAR/PDMS	Divinylbenzene/Carboxen/Polydimethylsiloxane
EI	Electron Ionization
FAEE	Fatty Acid Ethyl Esters
GC-MS	Gas chromatography–mass spectrometry
HCl	Hydrochloric Acid
HS-SPME	Headspace Solid-Phase Microextraction
HSD	Tukey's Honest Significant Difference
ICCO	International Cocoa Organization
ICCRI	Indonesian Coffee and Cocoa Research Institute
ISCQF	International Standards for the Assessment of Cocoa Quality and Flavor
L*	Lightness
MSTFA	N-methyl-N-trimethylsilyl-trifluoroacetamide
OPLS-R	Orthogonal Projection to Latent Structure-Regression
PCA	Principal Component Analysis
PLS	Projection to Latent Structure
PLS-R	Projection to Latent Structure-Regression

QC	Quality Control
RIs	Retention Indices
RMSECv	Root-Mean-Square-Error of Cross-validation
RMSEE	Root-Mean-Square-Error of Estimation
RSD	Relative Standard Deviation
VIP	Variable Importance in Projection

Table of Contents

Chapter 1

General introduction

1.1 The importance of fine flavor cocoa

Fine flavor cocoa refers to high-quality chocolate produced from the *Theobroma cacao* tree, one of the most significant agricultural commodities globally due to its substantial economic (Castro-Alayo, Idrogo-V Asquez, et al., 2019) (Bagnulo et al., 2023). The beans of *Theobroma cacao* serve as the primary raw material for the chocolate (Moreno-Rojas et al., 2023). From a commercial and industrial perspective, cocoa is classified into two categories: bulk cocoa, known for its basic flavor, and fine flavor cocoa, distinguished by its outstanding aromatic qualities (Rottiers et al., 2019).

Fine flavor cocoa is highly prized for its superior quality and distinctive flavor profiles, which command a significantly higher market price compared to bulk cocoa (Escobar et al., 2021). According to the International Cocoa Organization (ICCO, 2017), fine flavor cocoa is typically priced between $5,000 and $10,000 per metric ton, while bulk cocoa ranges from $3,000 to $3,500 per metric ton. This price disparity reflects not only the differences in quality but also the increasing consumer demand for unique and premium cocoa varieties. As consumer preferences shift toward high-quality cocoa products, the demand for fine flavor cocoa is expected to grow (Tscharntke et al., 2023). To meet this rising demand and maintain its premium market value, it is crucial to enhance fine flavor cocoa production.

Fine flavor cocoa is primarily produced from *Theobroma cacao* trees of the Criollo and Trinitario varieties (Dillon et al., 2023). The cacao variety plays a significant role in determining the levels of storage polysaccharides, proteins, and polyphenols in the beans, which in turn influence the types and quantities of flavor precursors formed during fermentation, drying, and roasting. These precursors are essential for flavor development and substantially shape the final flavor profile (Kongor et al., 2016). The Criollo variety is particularly valued for its outstanding flavor quality but has notable limitations, including low vigor, poor productivity, and high susceptibility to pests and diseases (Kongor et al., 2016) (Nguyen et al., 2021). As a result, Criollo has become a rare variety, with some populations nearing extinction (Lachenaud & Motamayor, 2017), particularly in regions like Indonesia, where the original Criollo variety is no longer found. In contrast, Trinitario, a hybrid of Criollo and Forastero, combines the superior vigor and resilience of Forastero with some of the desirable flavor characteristics of Criollo. However, the flavor quality and pest and disease resistance of Trinitario trees can vary depending on the genetic contributions from both parent varieties (Colonges, Jimenez, et al., 2022) (Kongor et al., 2016). This variability underscores the importance of breeding programs aimed at enhancing consistency in flavor quality and improving disease resistance.

1.2 Breeding for the improvement of fine flavor cocoa

The development of fine flavor cocoa cultivars is vital not only to meet market demand but also to preserve the biodiversity of cacao species (Colonges, Loor Solorzano, et al., 2022). Efforts to enhance fine flavor cocoa often combine selective breeding, hybridization, and advanced techniques such as genomics and metabolomics to identify desirable traits (Bekele & Phillips-Mora, 2019). These initiatives aim to improve flavor

quality while promoting sustainability by increasing yield and resistance to pests and diseases (Colonges, Jimenez, et al., 2022).

The Indonesian Coffee and Cocoa Research Institute (ICCRI) conducted various crossing combinations to enhance resistance against pests and diseases, flavor quality, and yield. This breeding project resulted in several hybrids, which were subsequently evaluated for their resistance levels and yield (Sari, Setyawan, et al., 2022). However, the flavor quality of these parent-hybrid combinations has yet to be comprehensively assessed or characterized.

1.3 Characterization of fine flavor cocoa

Characterization of fine flavor cocoa involves evaluating and documenting its phenotypic and biochemical traits, as these are closely linked to its flavor quality. Currently, fine flavor cocoa can be characterized based on its bean appearance or fresh bean color (Kongor et al., 2016). Oliva-Cruz et al. (Oliva-Cruz et al., 2021) emphasized that the fresh color of cacao beans is a crucial characteristic for discriminating between genotypes. Specifically, cacao varieties from the Criollo and certain Trinitario groups, characterized by up to 80% white beans, were categorized as fine flavor cocoa. In contrast, the Forastero and some Trinitario varieties with up to 80% purple beans are classified as bulk cacao (Devy et al., 2019). However, the fresh bean color of hybrids can vary depending on the parentage, because fresh bean color is inherited through additive or incomplete dominance (Lachenaud & Motamayor, 2017).

Another method for distinguishing between fine flavors and bulk cocoa is sensory evaluation (Escobar et al., 2021) According to this approach, fine flavor cocoa is

characterized by outstanding and highly complex notes (Santander et al., 2021). In contrast, bulk cocoa is characterized by a basic or simple flavor without aromatic notes (Herrera-Rocha et al., 2024). However, this sensory analysis has limitations, as there is no universal standard of quality owing to the complexity of the attributes, the lack of expert or certified panelists, and the potential for inconsistent results (Krähmer et al., 2015) (Smulders et al., n.d.). This method is impractical when dealing with a large number of samples because there is a limit to the number of samples that can be evaluated by panelists (W. Fang et al., 2014). Therefore, it is necessary to employ a robust method to complement the existing approach for characterizing fine flavor cocoa.

1.4 Metabolomics as a tool for cocoa characterization

Metabolomics is a powerful and robust method for characterizing agricultural products including cocoa (Grissa et al., 2016), by revealing both volatile and non-volatile compounds that determine the flavor quality of cocoa (Herrera-Rocha et al., 2021). The analytical methods employed in metabolomics, such as gas chromatography-mass spectrometry (GC-MS), have led to a wide range of compound detection in cocoa beans and are appropriate for examining metabolite profiles in cacao (Michel et al., 2021). This method can profile different metabolite levels between hybrids and their parents (Le et al., 2023) and provide a high level of precision in identifying and characterizing specific traits associated with the complexity of plant phenotypic diversity, such as cocoa flavor quality (Vincent Colantonioa, 2022).

The metabolomic approach has been widely used to identify food quality markers by correlating chemical profiles with sensory attributes (Zhang et al., 2024) (Kim et al.,

2023). Gas chromatography-mass spectrometry (GC-MS) has been increasingly employed for chemical analyses of cacao because of its reproducibility, stability, and ease of use (Putri et al., 2022). Combined with sensory analysis, this method can significantly enhance our understanding of the flavor characteristics of different cacao genotypes. Correlating the metabolomic profiles of cocoa with sensory attributes can validate sensory perceptions and provide a scientific basis for understanding flavor characteristics by identifying the specific compounds responsible for certain attributes (L. Zhao et al., 2023) (Y. Zhao et al., 2024) (Kim et al., 2023).

Previous studies have the metabolomic approaches of Criollo and Forastero. Some of these studies have focused on dynamic postharvest processes (Castro-Alayo, Idrogo-V Asquez, et al., 2019) (Moreno-Rojas et al., 2023), while others have characterized unfermented beans (Qin et al., 2017). Most of these studies have revealed the presence of volatile components. However, both volatile and non-volatile organic compounds influence the distinct aromas and flavors of cocoa (Velásquez-Reyes et al., 2023). While prior studies have provided valuable insights, only a few have employed a comprehensive approach that integrates volatile and non-volatile metabolomic analysis, phenotypic evaluation (such as fresh bean color), and sensory evaluation to characterize fine flavor cocoa. Hence, the present study is the first to report a comprehensive characterization of fine flavor cocoa resulting from several cross combinations in a plant breeding program.

1.5 Objective and strategy

This study aimed to characterize fine flavor cacao in parent-hybrid combinations using a metabolomic approach. To achieve this objective, two strategies were employed:

(1) investigating the characteristics of fine flavor cacao in parent-hybrid combinations through metabolite profiling and fresh bean color analysis, and (2) correlating sensory properties with metabolite profiles to express the characteristics of fine flavor cacao in parent-hybrid combinations.

1.6 Thesis outline

This dissertation is organized into four chapters, each focusing on a different method to comprehensively characterize fine flavor cacao in parent-hybrid combinations. Chapter 1 provides a general introduction, highlighting the importance of fine flavor cacao, the role of cacao breeding in developing fine flavor varieties, and current methods for characterizing such cacao. It also discusses the application of metabolomic approaches in this context and identifies existing research gaps in metabolomic profiling of fine flavor cacao.

Chapter 2 focuses on characterizing the non-volatile metabolite profiles of parent-hybrid combinations through widely targeted Gas Chromatography-Mass Spectrometry (GC–MS) and bean phenotype analysis. This chapter reports the distinct bean color phenotypes and metabolite profiles that differentiate fine flavor cacao from bulk cacao, as well as the promising hybrids resulting from the crosses.

Building on the findings from Chapter 2, Chapter 3 explores the characterization of a specific hybrid that showed promising results in Chapter 2. This chapter investigates both the volatile and non-volatile metabolites of the hybrid and evaluates sensory attributes through correlation with the metabolite profiles. Notably, the correlation

between metabolite profiles and sensory attributes serves to highlight the exceptional characteristics of this particular hybrid.

Finally, Chapter 4 provides a comprehensive conclusion, summarizing the key findings and their implications. It also discusses potential directions for future research, based on the topics explored throughout the dissertation.

Chapter 2

Investigation of the characteristics of fine flavor cocoa in parent-hybrid combinations through non-volatile metabolite profiling and fresh bean color analysis

2.1 Introduction

Fine flavor cocoa is generally characterized by fresh bean color and specific sensory attributes (Kongor et al., 2016) (Escobar et al., 2021). However, these methods are not suitable for evaluating progenies or hybrids because their colors may vary depending on their parental genetics. Additionally, sensory evaluation presents significant challenges due to the large number of samples required and the need for expert panelists to assess the complexity of the attributes (Y. Fang et al., 2020). These limitations highlight the need for robust complementary characterization methods.

Metabolomics offers a powerful and reliable approach for investigating both volatile and non-volatile compounds that define the flavor profile of cocoa (Herrera-Rocha et al., 2021). Most previous metabolomics studies on cocoa have focused on differentiating the flavor attributes of Criollo and Forastero varieties under various postharvest treatments, primarily emphasizing volatile (Castro-Alayo, Idrogo-Vásquez, et al., 2019). In contrast, only a few studies have comprehensively examined non-volatile compounds to characterize fine flavor cocoa. For instance, Velazquez et al. (Velásquez-Reyes et al., 2023) analyzed a limited set of non-volatile components, focusing primarily on a few organic acids, in traditional varieties such as Criollo and Forastero, without extending the profiling to other important metabolites. These studies were also limited in

genetic scope and often did not incorporate sensory analysis. In the present study, we focus exclusively on the non-volatile metabolite profiles of fine-flavor cocoa derived from multiple hybrid combinations developed through a plant breeding program. Sensory evaluation data from a previous study are used as a benchmark to explore how non-volatile compounds contribute to flavor expression across diverse genotypes.

2.2 Materials and Method

2.2.1 Plant materials

Twelve *Theobroma cacao* clones from the Indonesian Coffee and Cocoa Research Institute (ICCRI) were utilized in this study. These clones were cultivated in Jember, East Java, Indonesia, at an altitude of 45 m above sea level, with an average rainfall of 224 mm and a mean temperature of 32°C. All clones were under the same cultivation management. The cacao trees were mature (4 years old) and had a similar height of 2–3 m. The genetic backgrounds of the samples included Trinitario varieties, specifically T1, T4, T3, T5 (parental clones), and PNT 12 (a non-parent selected for its white bean color). The Forastero group consisted of F1 (parent), F2 (a non-parent selected to represent dark purple bean color), and hybrids derived from conventional breeding through controlled crosses between selected Trinitario parental clones. These hybrids were produced by manually transferring pollen from the flower of one parent clone to the stigma of another, followed by bagging the pollinated flowers to prevent unwanted cross-pollination.

After seed development, the resulting hybrid plants, along with the parent plants, were propagated vegetatively, using techniques such as grafting or budding to ensure that each new tree maintained the exact genetic makeup of the selected hybrid. Because these

trees are propagated asexually and not grown from seed, each resulting plant is genetically identical to its source, and thus referred to as a clone.

For ease of reference in data analysis and figure presentation, all clone names were abbreviated into codes, as shown in Table 1. However, the full names are also provided to aid interpretation. The sensory profiling of these samples was used as a benchmark to characterize flavor quality via metabolomic analysis. The sensory data of these samples was used as a benchmark to characterize the flavor quality by metabolomic analysis. These sensory data were obtained in 2022 using different harvesting batches from this study. Details of the sensory results are shown in Fig. S1.

Table 1. Sample information

Clones	Genetic background	Status based on sensory analysis	Sample code
DR 2	Trinitario	Fine flavor cacao	T1
PNT 16	Trinitario		T2
SULAWESI 01	Trinitario	Bulk cacao	T3
TSH 858	Trinitario	Fine favor cacao	T4
SCA 06	Trinitario		T5
KEE 02	Forastero	Bulk cacao	F1
KW 516	Forastero	Bulk cacao	F2
ICCRI 03	Hybrid (DR 2 (Female parent) x SCA 06 (Male parent))		H1
ICCRI 09	Hybrid (TSH 858 (Female parent) x SULAWESI 01 (Male parent))		H2
KW 733	Hybrid (TSH 858 (Female parent) x SULAWESI 01 (Male parent))		H3
KW 746	Hybrid (TSH 858 (Male parent) x SULAWESI 01 (Male parent))		H4
KW 742	Hybrid (SULAWESI 01 (Male parent) X KEE 02 (Male parent))		H5

2.2.2 Sample collection

Cacao pods were manually pollinated to expedite fruit development and maintain fruit uniformity. Around six months later, in October 2023, fully ripened and healthy pods were manually harvested. The pods were then opened, and the pulp and bean exocarp were removed to examine the bean color. The remaining beans were gathered and combined for fermentation.

2.2.3 Bean color measurement

Three ripe pods per clone were collected as replicates. From each pod, three beans were chosen for fresh bean color measurement. The beans were opened, and the pulp and exocarp were manually removed. Color was measured using the L*a*b* system (L for lightness, a for green-to-red, and b for blue-to-yellow) with a calibrated colorimeter (CM-2500d, Minolta, Japan) and a Munsell color chart. Measurements were based on the dominant bean color per pod for each clone.

2.2.4 Fermentation and drying process

Each cacao sample clone was opened and combined into a 5 kg bamboo wood basket. Small-scale fermentation was carried out at an environmental temperature of 27 °C and humidity of 80% for 98 h. Following the ICCRI fermentation standard, the beans were manually turned off after 48 h to ensure homogeneous fermentation. The endpoint of the fermentation was determined using a cut test. After fermentation, the beans of each clone were spread on a round bamboo wood tray and naturally dried under sunlight for three days. Each day, the beans were manually mixed to achieve a moisture

content of 7–8%. After drying, all beans were cut and only fully fermented beans were selected for metabolomic analysis.

2.2.5 Sensory Analysis

A sensory evaluation was previously conducted in 2022 by (Sari, Murti, et al., 2022) using cocoa liquor samples prepared following the International Standards for the Assessment of Cocoa Quality and Flavor (ISCQF) protocol. This evaluation, intended solely as a benchmark, involved three professionally trained panelists. Approximately 1–2 g of cocoa liquor from each sample was melted at 48–50 °C and presented in coded cups for blind assessment. Panelists evaluated the samples in random order and scored 21 attributes based on the ISCQF criteria, using a scale from 0 (not detected) to 10 (strong intensity).

2.2.6 Sample extraction and derivatization procedures

Sample extractions were conducted by adding 5 mg of cacao powder to 1000 µl of a mixed solvent (composed of methanol, water, and chloroform with a ratio of 5:2:2 v/v/v) containing 100 µg/mL of the internal standard (Ribitol). 600 µL of the supernatants were placed into a 1.5 mL microtube, added with 300 µL ultrapure water, and vortexed to homogenize the liquid. 200 µL of supernatant was transferred into a new 1.5 mL microtube, and 200 µL of supernatant in each sample was pooled for Quality Control (QC). All samples, including the QC and blank, were sealed with hole caps and then centrifuged under vacuum conditions using a centrifugal concentrator (TAITEC, Saitama, Japan) for 2 h at room temperature to remove water. Oximization and silylation were performed as derivatization procedures in this analysis based on a previous study.

2.2.7 Gas Chromatography-Mass Spectrometry (GC-MS) analysis

A GC-MS-QP2010 Ultra system was employed for metabolomic analysis. Before analysis, the mass spectrometer underwent tuning and calibration checks. Hydrogen was used as the carrier gas in this analysis, with a linear velocity of 39.0 cm/s and a flow rate of 1.2 mL/min. The column temperature initiated at 80°C for 4 min, then increased to 330°C at a rate of 15 °C/min, and was kept steady for approximately 8 min. The interface and ion source temperatures were consistently maintained at 310 °C and 280°C, respectively. Electron ionization (EI) was generated at 1.00 kV. The spectra were recorded within the mass range of m/z 85–500 were recorded. Prior to the first sample injection, a standard mixture of alkanes (C8–C40) was injected to determine the retention time (instrumental peak identification).

2.2.8 Data processing

The raw GC-MS data were converted into Andi (AIA) file format using GC/MS solution software (Shimadzu), then into Analysis Base File (ABF) format with the ABF Converter program. Peak alignment, filtering, and tentative annotation were carried out in MS-DIAL (version 4.38) using the GL-Science DB spectral database (InertCap 5MS-NP, Fiehn RI). The database was obtained from the MS-DIAL website. Annotations were verified by cross-referencing peaks with the NIST library (NIST/EPA/NIH EI-MS Library) in GC/MS software, using a metabolite similarity threshold of >80%. Ribitol was used as the internal standard for normalizing annotated metabolites' relative intensities. Metabolites with a relative standard deviation (RSD) above 30% in QC samples were excluded from the analysis.

2.2.9 Statistical analysis

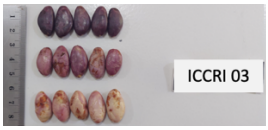
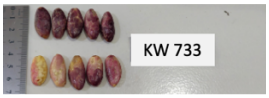
Multivariate data analysis was employed to visualize the data through principal component analysis (PCA) with auto-scaling and no transformation. Variations in the dataset were analyzed using analysis of variance (ANOVA), followed by a post-hoc Tukey's honest significant difference (HSD) test and a volcano plot. Significant differences were determined at *P-value* of ≤ 0.05. OPLS-R analysis (Orthogonal Projection to Latent Structure Regression) was also constructed to identify the correlation between metabolites corresponding to fine flavor cocoa and fresh bean color. In the PCA and OPLS-R analysis, metabolites served as explanatory variables, while fresh bean color values (L*, a*, b*) were used as response variables for OPLS-R data analysis. The b* value and percentage of bean color were not normally distributed; therefore, the data were log-transformed. From the OPLS-R analysis, VIP score (variable importance in projection) and coefficient values were obtained. A VIP value greater than one was considered to have an important contribution to the model. Model performance was evaluated by R^2 and Q^2 values, indicating the proportion of variance in the data elucidated and predicted by the model (Kim et al., 2023). This OPLS-R model was further validated using CV-ANOVA at *P-value* of ≤ 0.05. Multivariate data analysis and OPLS-R were analyzed using SIMCA-P software ver. 13 developed by Umetrics, Umea, Sweden, while ANOVA was conducted using R software (ver. 2023.12.0+369).

2.3 Result and discussion

2.3.1 Fresh bean color of cacao clones

Fresh bean color is one of the phenotypic traits used to differentiate Criollo and Forastero varieties. Criollo, known for its fine flavor, has white bean cotyledons, while Forastero, commonly used for bulk production, has purple beans (Lachenaud & Motamayor, 2017). Most previous studies have focused on describing the fresh bean and fruit colors of these two varieties (Oliva-Cruz et al., 2021). This study is the first to explore fresh bean color in parent-hybrid combinations of cacao clones. To provide a more detailed analysis of fine flavor cocoa in both parents and hybrids, fresh bean color was evaluated using L*a*b* values, and the percentage of different bean colors per pod was measured (Table 2, Fig. S2).

Table 2. Quantitative data and visual appearance of fresh bean color

No.	Clones code	L*	a*	b*	Color	Color code**	Visual appearance
1	T1	73.28 a	5.13 de	25.75 a	White	75Y8/2	
2	T2	79.29 a	3.05 e	22.33 a	White	75Y8/2	
3	T3	29.89 c	7.01 cde	0.88 b	Dark purple	5RP3/2	
4	T4	43.80 bc	13.26 ab	2.91 b	Purple	5RP44	
5	T5	32.26 c	8.76 bcde	1.10 b	Dark purple	5RP3/2	
6	F1	31.48 c	7.45 bcde	1.07 b	Dark purple	5RP3/2	
7	F2	33.48 c	8.33 bcde	2.26 b	Dark purple	5RP3/2	
8	H1	36.04 c	10.48 bcd	3.35 b	Purple	5RP44	
9	H2	29.92 c	5.46 de	1.32 b	Dark purple	5RP3/2	
10	H2	51.80 b	17.05 a	2.95 b	Light purple	5RP66	
11	H3	40.31 bc	9.69 bcd	4.87 b	Light purple	5RP66	
12	H4	35.18 c	11.80 abc	2.63 b	Dark purple	5RP3/2	

Remarks: Different letters in the same column indicate statistically significant differences (Tukey's HSD, p-value < 0.05). L*: Lightness; a*: Green to red; b*: Blue to yellow. **) Codes are based on Munsell color chart.

Trinitario clones, T1 (identified as fine flavor based on sensory analysis in Fig. S1) and T2 had a high proportion of white beans per pod, exceeding 88% (Fig. S2). These clones also exhibited a lighter color. In contrast, the Forastero clones F1 and F2, used for bulk cocoa, produced dark purple beans, over 80% per pod. White beans are typically found in the rare Criollo variety and some Trinitario types, which are known for their mild nutty and aromatic notes. On the other hand, Forastero's purple beans contribute to a stronger, more astringent flavor (Collin et al., 2023).

Trinitario clones T3, T5, and T4 had significantly lower lightness. T3 and T5 had dark purple beans in 100% and 79.89% of pods, respectively, while T4 had over 79% purple beans (Fig. S2). These results suggest that these three clones resemble the Forastero type in fresh bean color. However, sensory analysis (Fig. S1) showed that T4 had more aromatic attributes, while T3 had a stronger astringency. This indicates that fresh bean color alone may not be the main criterion for characterizing fine flavor cocoa (Lachenaud & Motamayor, 2017).

Hybrid progenies displayed varying percentages of fresh bean color per pod (Fig. S2) and had significantly darker beans compared to T1 and T2 (Table 2). The H1 hybrid, derived from a cross between T1 (Trinitario white bean) and T5 (Trinitario dark purple bean), produced beans that were 16.67% dark purple, 75% purple, and 8.33% light purple. Previous research suggests that fresh bean color inheritance in cacao follows an additive or incomplete dominance pattern. In an additive pattern, each allele contributes a

measurable effect to the phenotype, resulting in an intermediate expression. In incomplete dominance, the heterozygous phenotype is also intermediate but does not fully resemble either parent, showing a blending of traits. This implies that bean color is influenced by multiple genes, and the resulting phenotype reflects the combined effects of parental alleles (Lachenaud & Motamayor, 2017). Similar trends were observed in hybrids H3, H4, and H2, which resulted from crosses between T4 (Trinitario light purple bean) and T3 (Trinitario dark purple bean). Additionally, the cross between T3 (Trinitario, dark purple bean) and F1 (Forastero, dark purple bean) produced the H5 hybrid, which was predominantly dark purple. Based on these findings, fresh bean color alone is not a sufficient criterion for classifying fine flavor cocoa, specifically in the hybrid progenies. Therefore, this study also analyzed all cacao clones using metabolomic data.

2.3.2 Metabolite profiles of all cacao clones

The metabolome profiling of cacao clones (Fig. 1) highlights their overall characteristics, including parent-hybrid combinations and representative samples: T2 (Trinitario, white bean) and F2 (Forastero, dark purple bean). The PCA score plot (Fig. 1A) shows clear clustering based on variety and bean color. Trinitario dark purple beans (T3, T5) separate from Trinitario white beans (T1, T2) along PC1 (22.4% variance), while Forastero clones (F1, F2) cluster on the negative side. Notably, Trinitario white beans were clustered in the center between the Forastero and Trinitario dark purple beans. Another remarkable result is that the hybrids could not be classified according to their bean appearance.

This study performed separate Principal Component Analysis (PCA) for the T1 (Trinitario white beans) and Forastero varieties (F1 and F2), as shown in Fig. 1C and 1D, to differentiate between fine flavor and bulk cocoa, used as a basis for analyzing parent-hybrid combinations. Sensory results from 2022 (Fig. S1) classified T1 as fine flavor cocoa due to its stronger aromatic notes, such as floral, nutty, and fruity. In contrast, the Forastero varieties (F1 and F2) were identified as bulk cocoa, characterized by strong bitterness and astringency. The PCA score plot (Fig. 1C) clearly separates fine flavor cocoa (Trinitario white beans) from bulk cocoa (Forastero dark purple beans) along PC1, with 41% of the variance. This metabolomic result aligns with the sensory analysis, providing a clear distinction between fine flavor and bulk cocoa.

PCA loading plot illustrated in Fig. 1D and Table S2, show shows fine flavor cocoa has higher caffeine and organic acids, supported by volcano plot analysis (Fig. 1E). which showed that 26 metabolites had notable differences, as determined by a *t-test* with significant *p-values* below 0.05 (Fig. 1E). Among these significant metabolites, five compounds, including caffeine and organic acids such as malic acid, fumaric acid, citric acid, and tartaric acid, were found to be twice as high in fine flavor cocoa. This finding aligns with prior research reporting that Criollo, the finest cacao types are characterized by lower theobromine and higher caffeine contents (Zapata-Alvarez et al., 2024). Caffeine plays a role in the taste and aroma enhancement in cocoa. While organic acids are essential for balancing the flavor (Luna et al., 2002) (Davrieux et al., n.d.). In contrast, Forastero cocoa variety contains higher levels of theobromine but lower caffeine, consistent with its bitter, astringent profile. This result aligns with those of previous studies, which reported that the Forastero variety tends to have higher theobromine and

lower caffeine content (Guzmán Penella et al., 2023). While previous studies have identified caffeine as a key metabolite in Criollo, one of the finest cacao varieties (Zapata-Alvarez et al., 2024) (Velásquez-Reyes et al., 2023)., this study provides new insight by showing that Trinitario white beans—also considered fine-flavor cocoa. Additionally, this study identified specific organic acids that may also play a critical role in defining the fine-flavor characteristics. Thereby, expanding the current understanding of flavor-related metabolites in fine flavor cocoa

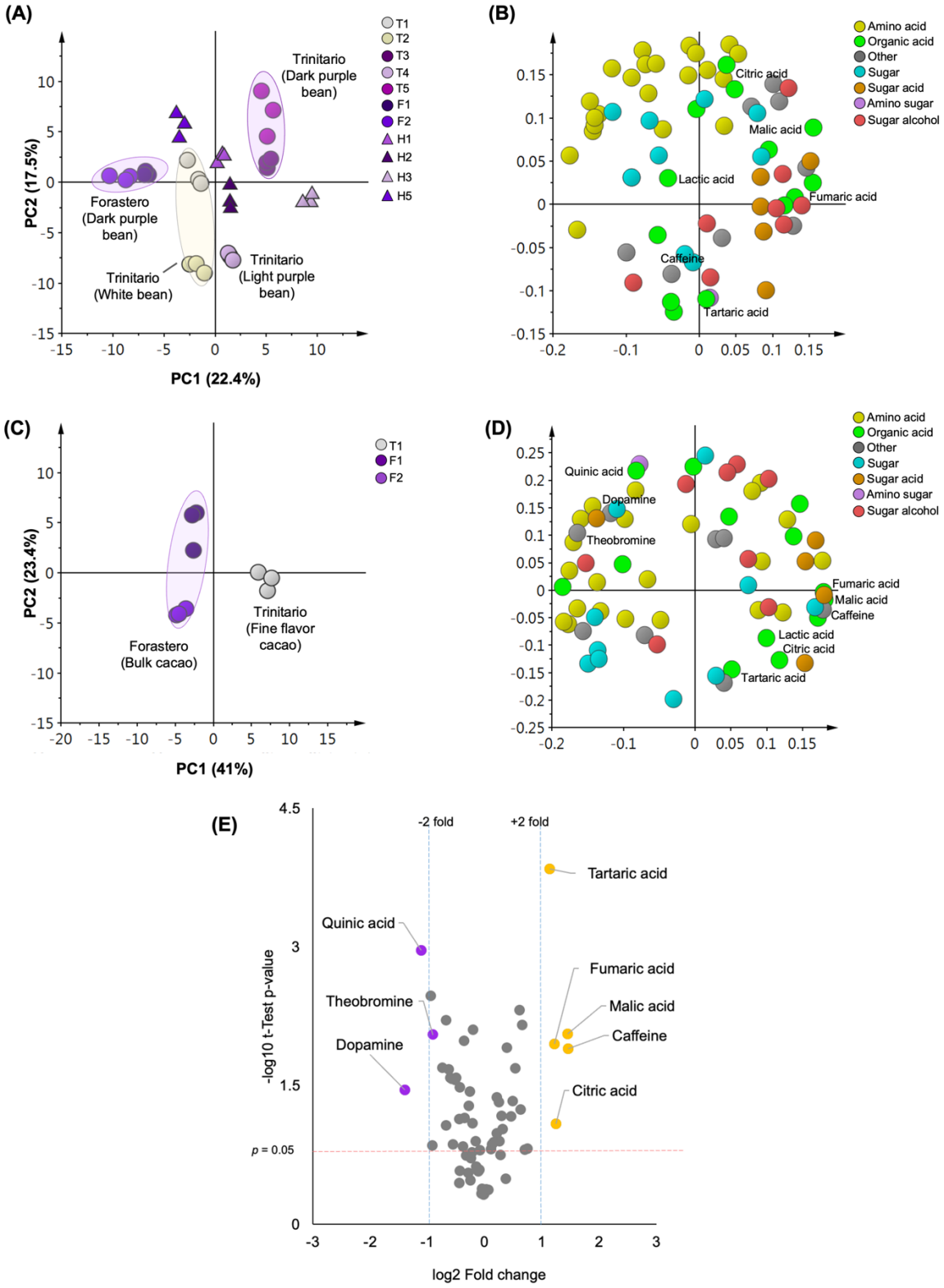


Fig. 1. The PCA score plot (A) and loading plot (B) were generated from the GC-MS-based metabolomic analysis of all samples, with auto-scaling and no transformation, using n=3 replicates. The score plot (C)

and loading score plot (D) illustrate the PCA comparison between fine flavor and bulk cacao. Plot (E) is a volcano plot of fine flavor and bulk cacao. In the score plot, circles represent cacao clones from Forastero, Trinitario with white beans, Trinitario with dark purple beans, and Trinitario with light purple beans. Triangles indicate hybrids (Trinitario × Trinitario and Trinitario × Forastero crosses). Dots in the loading plot denote metabolites associated with the observed separation, with different colors representing various metabolite classes. Highlighted metabolite names indicate influential compounds that contributed to sample clustering. In the volcano plot, yellow dots represent significantly increased metabolites (p-value < 0.05) with at least a 2-fold change, while purple dots indicate significantly decreased metabolites (p-value < 0.05) with at least a 2-fold change. Grey dots represent non-significant metabolites (p-value > 0.05) with less than a 2-fold change.

2.3.3 Metabolite profiles of parent-hybrid combination

This study performed separate principal component analyses (PCA) (Fig. 2) to gain a more comprehensive understanding of the characteristics of parent-hybrid combinations. The PCA score plot (Fig. 2A) revealed three distinct clusters in the T1 × T5 cross, which resulted in the hybrid H1. PC1, which explained 50.7% of the variance, differentiated the female and male parents, with the hybrid clustering alongside its female parent on the negative side of PC1. The PCA loading plot (Fig. 2B) identified caffeine as a key metabolite on the negative side of PC1, indicating its accumulation in both the female parent and the hybrid. A bar graph analysis (Fig. 3) further confirmed significantly higher levels of caffeine, malic acid, fumaric acid, and tartaric acid in the hybrid and its female parent compared to the male parent and Forastero varieties. These results suggest that this parent-hybrid combination has the potential to produce hybrids with higher levels of key metabolites associated with fine flavor cocoa. Furthermore, this finding underscores the influence of female parents on cocoa's flavor quality. While a previous

study demonstrated the impact of pollen donors on cocoa flavor through sensory analysis(Sukha et al., 2008), this study provides evidence from a metabolomic perspective.

The other hybrids, H2, H4, and H3, were derived from a cross between T4 (a Trinitario genotype with light purple beans) as the female parent and T3 (a Trinitario genotype with dark purple beans) as the male parent. Unfortunately, the beans of H4 were not sufficient for the fermentation process; thus, this hybrid was excluded from metabolomic analysis. PCA (Fig. 2C) revealed four distinct clusters, with PC1 (44.7% variance) clearly separating female and male parents. Notably, the H2 hybrid clustered with its female parent, indicating that this hybrid exhibited similar metabolite profiles to that of its female parent as a fine flavor cocoa. PCA loadings (Fig. 2D) further showed that H2 and its female parent had higher levels of caffeine, lactic acid, and amino acids, with key metabolite levels comparable to those of the female parent and T1, both classified as fine-flavor cocoa. Based on these metabolite profiles, H2 appears to exhibit superior flavor characteristics compared to its sibling hybrid H3, though not as pronounced as those of H1 and T1. Prior study highlighted that the flavor quality of cocoa beans is influenced by the genetic composition of their parents (Seguine, 2009). The inherent heterozygosity of Trinitario contributes to the substantial variability observed in the metabolome profiles of its hybrids (Davrieux et al., n.d.).

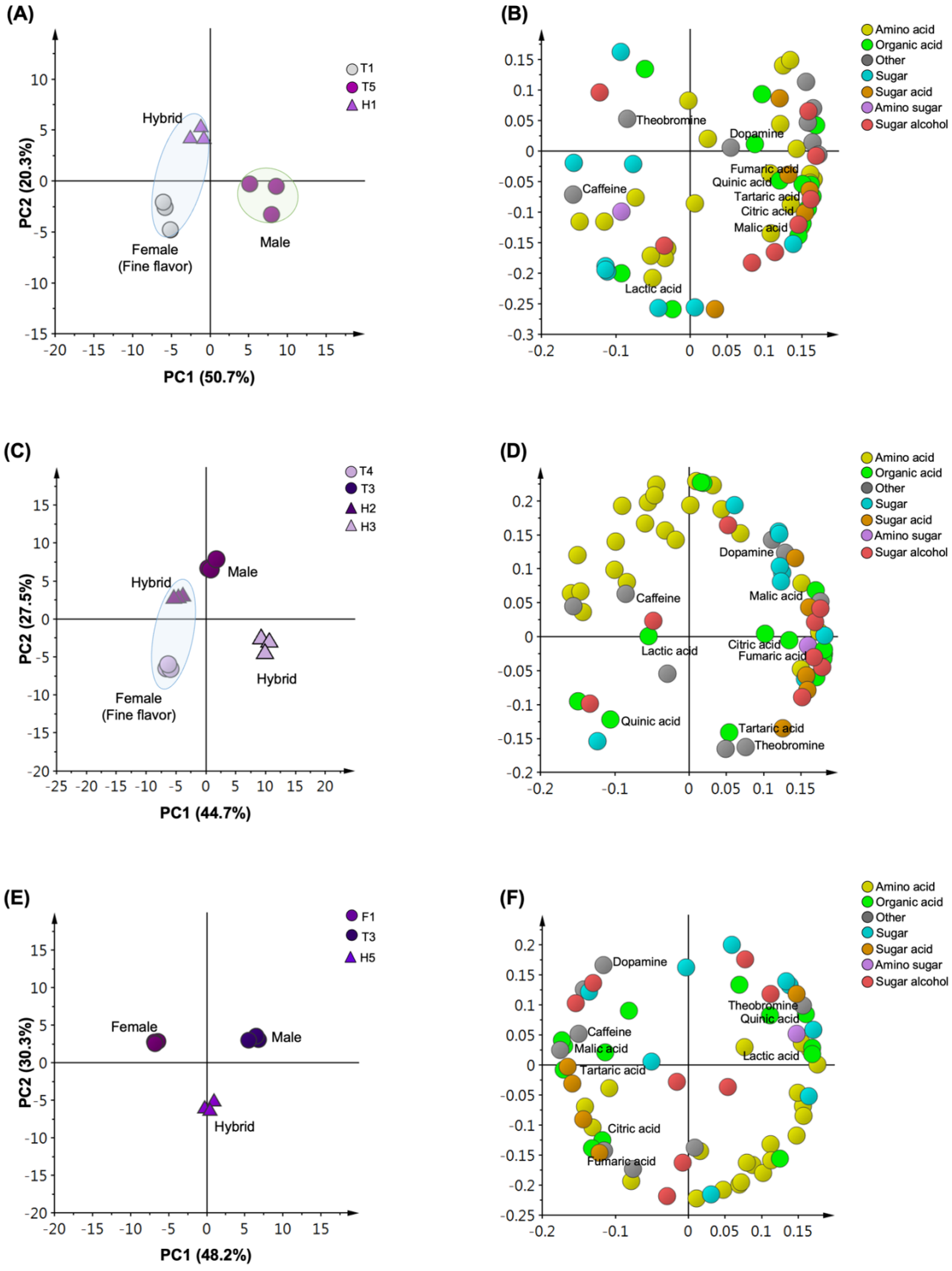


Fig. 2. The PCA score plots (A, C, E) and loading plots (B, D, F) were generated from GC-MS-based metabolomic analysis of the parent-hybrid combinations: (T1 × T5), (T4 × T3), and (T3 × F1). In the score

plots, circles represent the parent cacao clones, and triangles represent the hybrids. Dots in the loading plots indicate metabolites linked to the separation. The names of the highlighted metabolites are the influential compounds that clustered the samples.

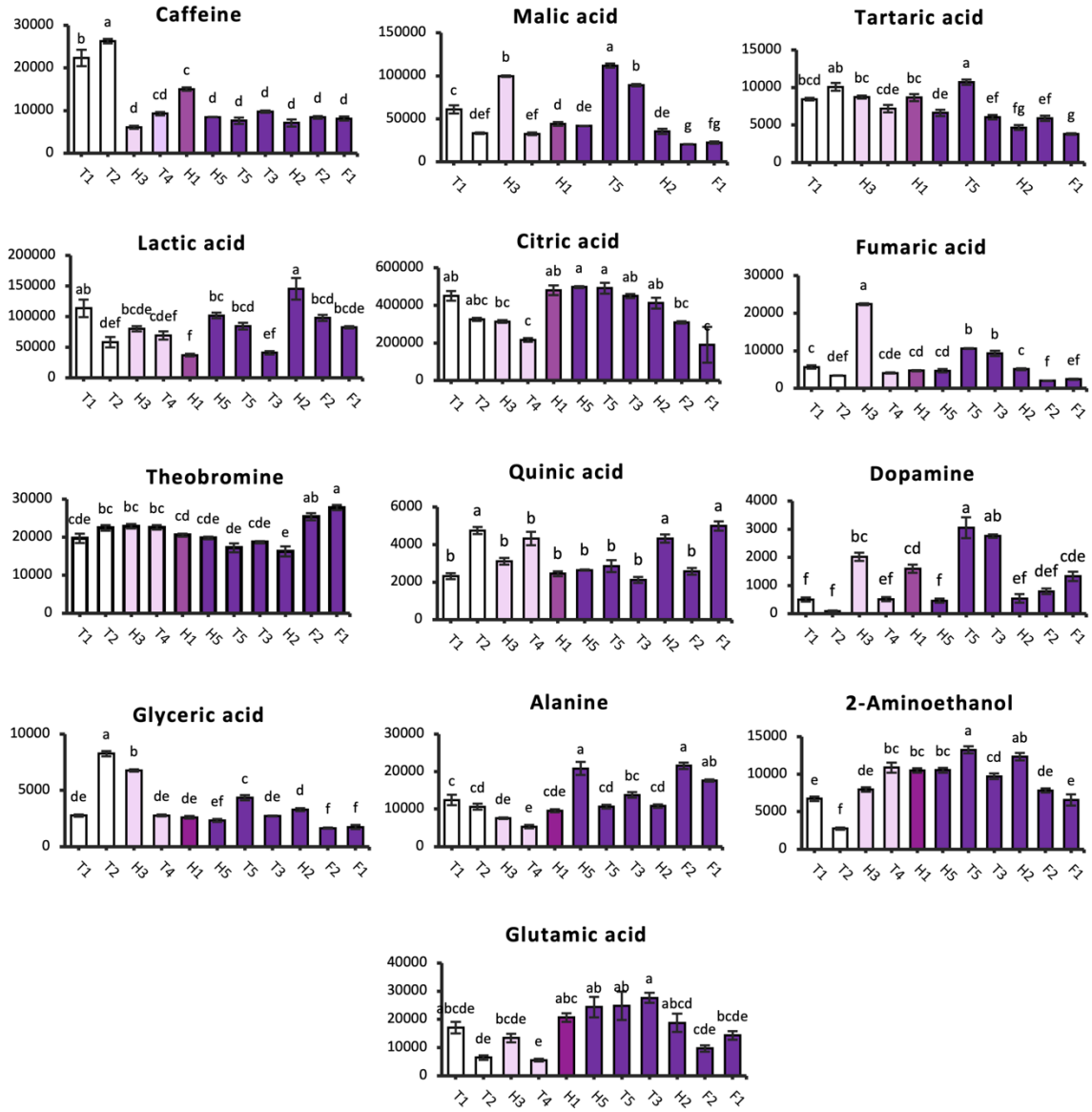


Fig. 3. The bar graphs show the relative intensity of the top higher VIP metabolites highlighted in the loading plot. Some of these graphs illustrate the relative intensity of the top three VIP metabolites listed in Tables S4-S8. The vertical axis represents the relative intensity, while the horizontal axis indicates the cacao clones with the codes. Bars labeled with the same letters do not show a statistically significant

difference according to Tukey's HSD test, with a p-value < 0.05. The colors in the bar graphs represent the fresh bean color of each sample.

In the T3 × F1 parent-hybrid combination (Fig. 2E), three groups emerged. The female parent clustered separately on the negative side of PC1 (48.2% variance), while hybrid H5 was distinct from both parents along PC2 (30.3% variance), indicating intermediate flavor characteristics. The highlighted metabolites presented in Fig. 3 revealed that caffeine levels in the hybrid and its parent were comparable to those in the Forastero variety. Additionally, organic acids such as citric acid, tartaric acid, and quinic acid showed no significant differences between the hybrid and its female parent, indicating a flavor profile similar to that of its female parent, which is classified as bulk cocoa. Notably, this parent-hybrid combination originates from Trinitario and Forastero varieties. Consequently, this crossbreeding may result in a hybrid with lower flavor quality than H1 and H2.

2.3.4 Correlation between metabolite profiles and fresh bean color of cacao

Fine flavor cocoa has traditionally been characterized by cotyledon appearance or fresh bean color. Criollo varieties, known for finest cocoa type, have white beans due to the absence of anthocyanins, while bulk Forastero cocoa has dark purple beans rich in anthocyanins (Kongor et al., 2016). Although fermentation can alter bean color, it could potentially influence the flavor quality of cacao. This study explores the correlation between fresh bean color and cacao metabolome profiles, as previous research has mainly focused on volatiles without quantitatively linking fresh bean color to metabolites.

Using OPLS-R, with fresh bean color as the response variable, the analysis obtained a good prediction model ($R^2 > 0.6$, $Q^2 > 0.5$) validated by CV-ANOVA ($P < 0.05$). Since no clear trend was observed in a* (red-green) or lightness percentage per pod, the analysis focused on lightness (L*), b* (yellow-blue), and the percentage of white and dark purple beans per pod. Detailed parameters are provided in Table S3.

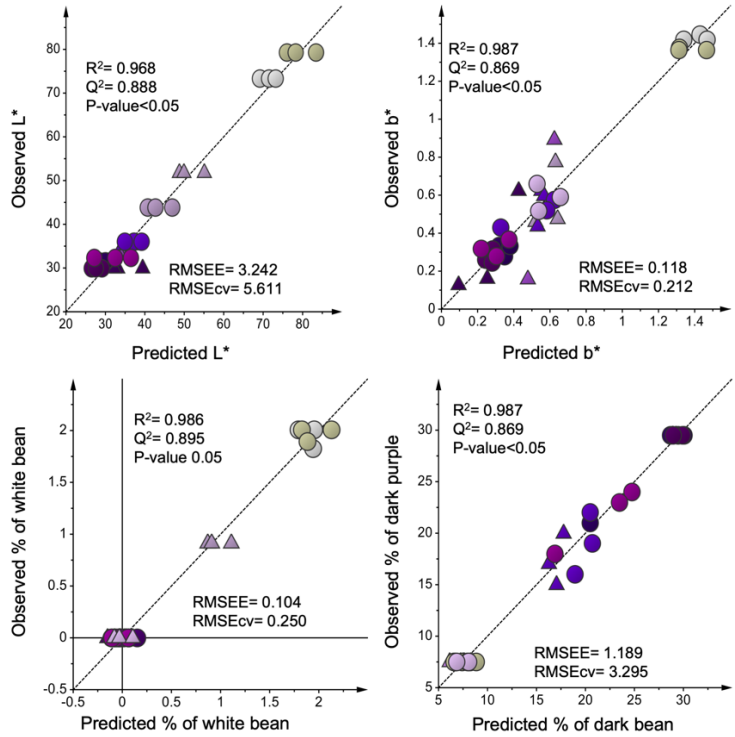


Fig. 4. Correlation models of L*a*b* value, the percentage of white bean per pod, and the percentage of dark purple bean per pod obtained from Orthogonal Projection to Latent Square Regression (OPLS-R) analysis show the correlation between metabolites and fresh bean color

The VIP score greater than 1 indicates a significant contribution to the model and was used to identify metabolites highly correlated with the response variable. The top ten metabolites (Table S4-S8), including caffeine, tartaric acid, and glyceric acid, exhibited strong positive correlations with L* (lightness), b* (blue to yellow), and the percentage

of white beans per pod. A previous study found that caffeine content is higher in Criollo (white beans) than in Forastero (purple beans) and contributes to flavor enhancement (Brunetto et al., 2007). Tartaric and glyceric acids are important for balancing flavors (Kim et al., 2023). The percentage of dark purple beans correlates with alanine, 2-aminoethanol, and glutamic acid. Sugars play a critical role in flavor development through Maillard reactions with amino acids (Aprotosoaie et al., 2016). This correlation between fresh bean color and metabolites suggests that certain candidate metabolites for fine flavor cacao are associated to white beans. In contrast, darker bean colors do not appear to correlate with metabolites typically associated with Forastero. Interestingly, hybrids or mixed genotypes may exhibit darker beans while maintaining metabolite profiles with traits from Criollo.

1.4 Conclusion

This chapter provides a comprehensive characterization of fine flavor cocoa in parent-hybrid combinations resulting from a plant breeding program, utilizing a widely targeted Gas Chromatography-Mass Spectrometry (GC-MS) metabolomics approach alongside bean phenotype analysis. The non-volatile profiles of fine flavor cacao were characterized by high levels of caffeine and organic acids, including malic acid, fumaric acid, citric acid, and tartaric acid. This study is the first to report non-volatile compounds, particularly organic acids, as key metabolites for differentiating fine flavor cocoa from bulk cocoa. Each crossbreed exhibited distinct flavor profiles, with the H1- and H2-hybrids identified as promising candidates. Further research should focus on exploring the volatile and sensory profiles of these selected hybrids to fully uncover their complex flavor characteristics.

Chapter 3

Correlation between sensory attributes and metabolomic profiles of cocoa liquor to express the characteristics of fine flavor cocoa in parent hybrid combination

3.1 Introduction

The previous chapter suggested that the flavors attributes of fine flavor cacao are also influenced by volatile compounds (Moreno-Rojas et al., 2023), suggesting that it is necessary to profile volatile compounds in these parent-hybrid combinations. A previous study (Herrera-Rocha et al., 2021) emphasized that the flavor characteristics of cocoa are determined by both volatile and non-volatile organic compounds. Furthermore, fine flavor cocoa is characterized by outstanding and highly complex notes (Santander et al., 2021). In contrast, bulk cocoa is characterized by a basic or simple flavor without aromatic notes (Herrera-Rocha et al., 2024).

Recent findings in the previous chapter revealed that hybrid progeny derived from parent-hybrid combination crosses exhibit promising flavor potential. Building on these insights, sensory and volatile compound analysis are essential to express the detailed flavor characteristics of the selected cacao hybrid. Accordingly, this chapter provides a comprehensive analysis of the flavor components in selected hybrids by correlating their metabolite profiles, encompassing both volatile and non-volatile compounds with sensory attributes.

Gas chromatography-mass spectrometry (GC-MS) has been increasingly employed for chemical analyses of cacao because of its reproducibility, stability, and ease of (Putri et al., 2022). Combined with sensory analysis, this method can significantly enhance our understanding of the flavor characteristics of different cacao genotypes. Correlating the metabolomic profiles of cocoa with sensory attributes can validate sensory perceptions and provide a scientific basis for understanding flavor characteristics by identifying the specific compounds responsible for certain attributes (L. Zhao et al., 2023) (Y. Zhao et al., 2024) (Kim et al., 2023).

3.2 Materials and method

3.2.1 Cacao sample

Cacao genotypes with a code F2 (Forastero variety), T3 (Trinitario variety, as male parent), and H2 (a hybrid from T4 and T3) were used in this chapter. These clones were grown in Jember, East Java Province, Indonesia, at 45 meters above sea level, with an average temperature of 28 °C and 300 mm of rainfall in 2024. These clones were managed under standard practices by the Indonesian Coffee and Cocoa Research Institute (ICCRI) and reached a height of approximately 3 meters at 5 years of maturity. Cacao pods were harvested during the rainy season in April 2024, shelled, and placed in fermentation boxes for the fermentation process (Fig. S4).

3.2.2 Post-harvest processing

All cacao samples underwent the same postharvest processing. After pod breaking, the cacao was placed in fermentation boxes (Fig. S4) with a 10 kg capacity per clone.

Fermentation occurred at 28 °C and 85% humidity for 114 h. Following the ICCRI standard, the beans were manually turned after 48 hours for even fermentation, with the endpoint determined by a cutoff test. After fermentation, the beans were spread on bamboo trays (Fig. S4) and sun-dried for four days, with daily mixing to achieve a moisture content of 7–8%. After drying, the beans underwent fermentation index and cut test analysis.

The procedure for roasting and producing cocoa liquor was based on the ISCQF (International Standards for the Assessment of Cocoa Quality and Flavor) protocol(Gutiérrez, n.d.) and the ICCRI standard, as follows: After drying, the fermented beans of each genotype were placed separately on perforated metal trays. The samples were roasted in an oven at 120 °C for 30 min, following a preheating period of 15 min at the same temperature. The actual roasting temperature and time for each clone were adjusted based on the moisture content of the beans and their size (weight of 100 grains). The timing was measured from 2°C below the set point. After roasting, the beans were cooled immediately to stop roasting. To prepare the cocoa liquor, 500 g of roasted beans was winnowed to separate the nibs from their shells (using a basic winnower, Cocoa TownTM, Alpharetta, USA). The shelled cocoa nibs were ground using a melanger until smooth cocoa liquor was obtained. All samples were then packed, vacuum-sealed, and stored in a fridge at -30 °C until further analysis.

3.2.3 Degree of fermentation index and cut test analysis

Dried cacao beans from each clone were collected (100 grains per sample) and used for the cutting test and fermentation index analysis. For the cut test analysis, beans were

manually cut lengthwise into two parts to expose the cotyledons. Each sample was placed on a white background, analyzed, and classified based on the color and texture of the exposed surface, as shown in Fig. S14.

The fermentation index was measured according to a previous study by Kongor et al. (Edem Kongor et al., 2013) using the Gourieva and Tsernetivinov methods. Five dried beans were ground using a mortar and pestle, and then approximately 0.5 grams of the crushed beans were extracted in 50 ml of a mixed solution (composed of methanol and HCl in a 97:3 v/v ratio). The samples was then left to homogenize in a refrigerator (8 °C) for approximately 20 h. The solution was filtered using Whatman paper (No.1) and the filtered solution was analyzed using a Shimadzu UV-1601 UV-vis spectrophotometer in the wavelength range of 400–700 nm. The fermentation index was calculated based on the ratio of the absorbance values at 460 nm and 530 nm.

3.2.4 Sensory evaluation

The cocoa liquor of each clone was tested by six trained panelists from the Indonesian Coffee and Cocoa Research Institute (ICCRI) board members, including four females and two males (30–50 years old). Cocoa samples for sensory analysis were prepared according to (The International Standards for the Assessment of Cocoa Quality and Flavor (ISCQF) protocol(Gutiérrez, n.d.). Approximately 1–2 gr of cocoa liquor from each clone were put in a cup then heated at 48–50 °C until cocoa paste was melted. Then, a melt cocoa liquor was served to the panelists. All samples were maintained blindly using alphabet codes and evaluated in random order for all blind evaluations. Twenty-one attributes according to the ISCQF protocol (cocoa, acidity, astringency, fresh fruit,

browned fruit, floral, vegetal, woody, spicy, nutty, sweet, browned roast, dusty, meaty, putrid, smoky, moldy, other off-flavors, global quality, overall flavor, and uniqueness) were scored ranging from zero to ten with the following criteria: 0 (no attribute detected), 1 (just a trace), 2 (low intensity), 3-5 (clearly characterized), 6-8 (dominant in the sample), 9-10 (Strong intensity).

3.2.5 Sample preparation and GC-MS non-volatile analysis

The GC-MS non-volatile analysis was performed based on a previous chapter. The cocoa liquor (5 g) was crushed in into a powder. Approximately 5 mg of each sample was prepared in 2 mL microtubes, with three replicates. Each sample was then mixed with a solvent composed of methanol (Wako Chemical, Osaka, Japan), chloroform (Kishida Chemical, Osaka, Japan), and ultrapure water (Wako Chemical) in a ratio of 2.5:1:1 v/v/v, containing 0.1 mg/mL ribitol as an internal standard. 600 μL of the supernatant was transferred to a 1.5 mL tube, and 300 μL of ultrapure water was added. 200 μL of the aqueous phase from each sample was pooled to create quality control (QC) samples, which were also transferred to a new 1.5 mL tube and sealed with a holed cap. The samples were then evaporated for 2 h at room temperature using a centrifuge concentrator. Oximization and silylation were performed for derivatization. Finally, 100 μL of each sample was transferred to a vial for GC-MS analysis.

A GC-MS QP2010 Ultra system (Shimadzu, Kyoto, Japan) was used for the analysis. Hydrogen was used as the carrier gas, with a linear velocity of 39.0 cm/s and a flow rate of 1.2 mL/min was used in this analysis. The column temperature was initially set at 80 °C for 4 min, then increased by 15 °C per minute to 330 °C, and maintained at

330 °C for 8 min. Ions were generated using the electron ionization (EI) method with a filament bias voltage of 70.0 V. EI mass spectra were recorded over a mass range of 85–500 m/z with an event time of 0.15 seconds. At the beginning of the analysis, a standard alkane mixture (C10–C40) was injected to determine retention indices (RIs) for tentative identification.

3.2.6 Volatile compounds analysis by HS-SPME arrow GC-MS

The volatile components of cocoa liquor from different cacao genotypes were extracted using the HS-SPME arrow and analyzed by GC-MS, following previous studies by Velásquez-Reyes et al(Velásquez-Reyes et al., 2023) with modification. A solid form of cocoa liquor was broken into pieces, and approximately 2, 5 g of each sample was weighed using three replicates. QC sample was also prepared by weighing 2,5 g from each sample and mixed to obtain a QC pool sample (n=3). All samples were then placed in a 20 mL screw vial and closed with a magnetic screw cap. The SPME procedure was conducted automatically using a multifunctional autosampler (AOC-6000 by Shimadzu) fitted with a DVB/CAR/PDMS (Divinylbenzene/Carboxen/Polydimethylsiloxane) fiber (20 mm × o.d. 1.1 mm, df = 100 μm; Shimadzu).

A GC-MS (GCMS-TQ8050 NX; Shimadzu) was utilized in this analysis. The analyte desorbed from the SPME fibers was injected at a split ratio of 25:1 (v/v). A helium carrier gas with a linear velocity of 40.7 cm/s. The column oven temperature was initially set at 40 °C for 5 min, then gradually increased to 250 °C at a rate of 3 °C/min, and maintained at 250 °C for 45 min. The MS was operated in scan mode (m/z 24–350). Both the ion source and interface temperatures were set at 250 °C. Prior to the analysis, a

standard of fatty acid ethyl esters (FAEEs, C4–C24) was injected to calculate the RIs for tentative identification.

3.2.7 Raw data processing

Raw GC-MS data were converted to AIA format using GC-MS Solution (Shimadzu) and then to .abf format via ABF Converter. Data processing in MS-DIAL 4.4 included baseline correction, peak filtering, alignment, noise reduction, and annotation using the GL-Science DB spectral database. Tentative annotations were verified against the NIST library. Metabolites with similarity scores above 80% were included in the analysis and normalized to an internal standard. HS-SPME arrow GC/MS intensities were adjusted using LOWESS, and metabolites with RSD >30% in QC samples were excluded.

3.2.8 Statistical analysis

All the data obtained from this study including metabolite, sensory attribute, fermentation index, and cut test analysis data were analyzed using ANOVA (analysis of variance) to observe the significant variation within the sample, then followed by a post-hoc Tukey's honest significant difference (HSD) test at $\alpha = 5\%$. Analysis of variance, biplot analysis, and Pearson's correlation were performed using the R software version 4.3.2 (2023-10-31). Principal component analysis (PCA) and partial least squares (PLS) regression analyses were performed using SIMCA-P+ version 13 (Umetricts, Umea, Sweden). PCA was used to visualize metabolite data within the cacao genotypes from dried fermented beans, roasted beans, and cocoa liquor. PLS regression analysis was conducted to determine the relationships between metabolite profiles as explanatory variables (x) and sensory attributes as response variables (y) (Kim et al., 2023) for

different cacao genotypes. The PLS regression model was then validated by cross-validation using the R^2 (a coefficient of determination which represents the proportion of the variance in the dependent variable) and Q^2 (model's predictive ability) parameters obtained from a random test with n = 200.

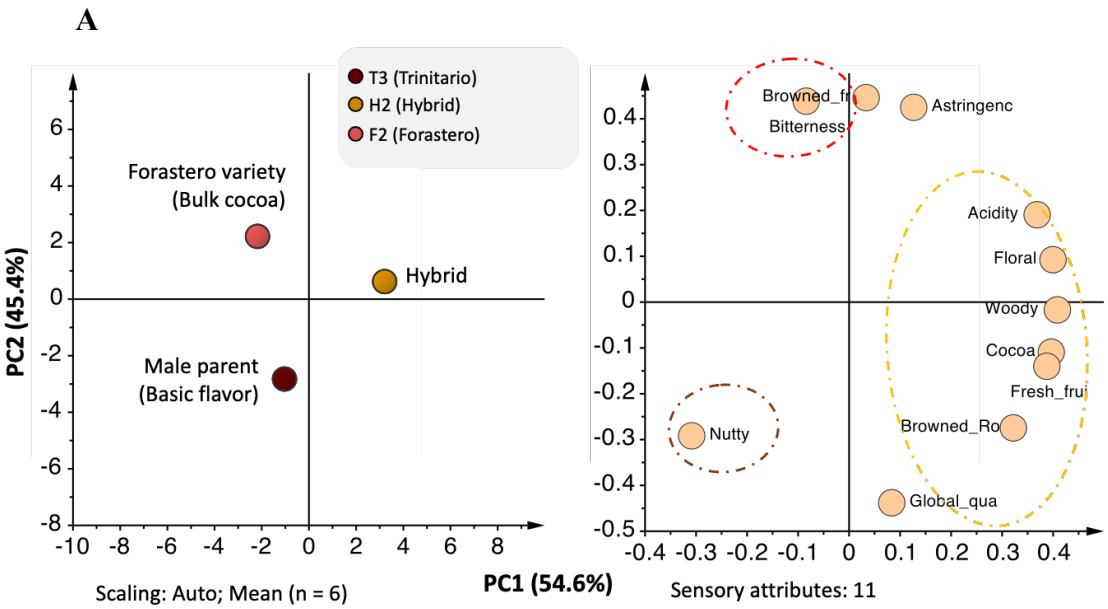
3.3 Result and discussion

3.3.1 Sensory characteristics

The sensory evaluation of the three cacao genotypes was analyzed using a PCA biplot (Fig. 5A). The first two principal components, PC1 and PC2, explained 54.6% and 45.4% of the variance, respectively, clearly differentiating the hybrid H2, its male parent (T3), and Forastero (F2) as bulk cocoa. The ISCQF protocol categorizes cacao's sensory attributes into core, complementary, and off-flavors. In this study, 21 attributes were assessed, with 11 identified in the genotypes: bitterness, astringency, acidity, cocoa (core), browned roast, browned fruit, floral, woody, fresh fruit, nutty, and global quality. No off-flavors were found, indicating optimal fermentation, confirmed by the fermentation index and cut test (Fig. S14).

The H2 hybrid exhibited outstanding sensory notes, including floral, woody, cocoa, fresh fruit, browned roast, and global quality. The male parent, T3, had a nutty flavor but fewer distinct attributes, while the Forastero variety F2 was dominated by bitterness, astringency, and browned fruit. Notably, H2's flavor profile differed significantly from its male parent and F2 as bulk cocoa. Previous studies have qualitatively reported that breeding cacao between the Trinitario and Forastero varieties can yield promising results, enhancing both resistance and flavor quality (Colonges, Jimenez, et al., 2022). However,

this study quantitatively revealed that the flavor quality of the H2 hybrid was more complex than that of its male parent and Forastero variety.



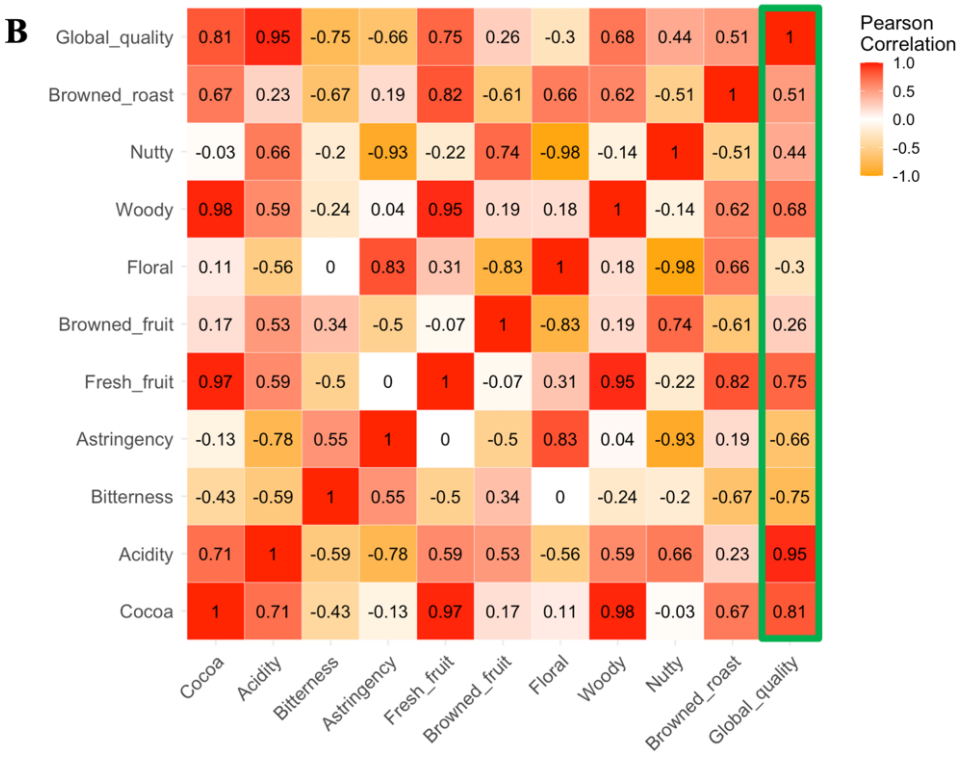


Fig. 5. Biplot analysis showing the sensory characteristics of cacao genotypes (A). Heatmap depicting the relationships between sensory attributes in cacao genotype based on Pearson's correlations (B).

To further clarify the sensory characteristics of these cacao clones, a Pearson's correlation heatmap was generated (Fig. 5B). The results revealed that the global quality of cocoa liquor from three cacao clones was positively correlated with cocoa flavor (0.81), acidity (0.95), fresh fruit (0.75), and woody characteristics (0.68). "Global quality" represents the overall cocoa quality. These results indicate that cocoa flavors, acidity, fresh fruit, and woody notes play a role in determining overall quality. Notably, H2 hybrid demonstrated a higher global quality, suggesting that this hybrid may have potential as a high-flavor cocoa preferred by chocolate consumers. Previous research has shown that cocoa samples with higher astringency levels generally exhibit lower flavor and overall acceptability (Palma-Morales et al., 2024).

3.3.2 Metabolite profiles of cocoa liquor

A total of 70 non-volatile metabolites, including amino acids, organic acids, and sugars, were detected (Table S9). HS-SPME Arrow GC/MS analysis identified 66 volatile compounds, such as acids, alcohols, esters, and ketones (Table S10). PCA of the non-volatile metabolites (Fig. 6A) revealed clear differences among cacao clones, with 50.8% of the variance distinguishing the H2 hybrid, T3, and F2, aligning with the sensory evaluation results. The loading plot (Fig. 6B) showed a consistent trend with findings from the first chapter, indicating that the H2 hybrid contained higher levels of certain organic acids and sugars.

The cocoa liquor of H2 hybrid, having complex sensory attributes and show promising based on the previous chapter, consistently exhibited significantly higher levels of organic acids (citric, malic, and 2-hydroxyglutaric acid) and sugars (glucono 1,5-lactone, trehalose, sorbitol, inositol, psicose, ribulose, and galactinol) compared to Forastero (F2, Bulk cocoa) and its male parent (T3) (Fig. 7). Organic acids, such as citric and malic acid, contribute to cocoa's acidity and are key to flavor balance. Previous studies have mainly focused on caffeine and theobromine for differentiating Criollo and Forastero (Velásquez-Reyes et al., 2023) (Zapata-Alvarez et al., 2024), whereas this study notably highlights the abundant non-volatile organic acids and sugars in a hybrid with outstanding sensory properties.

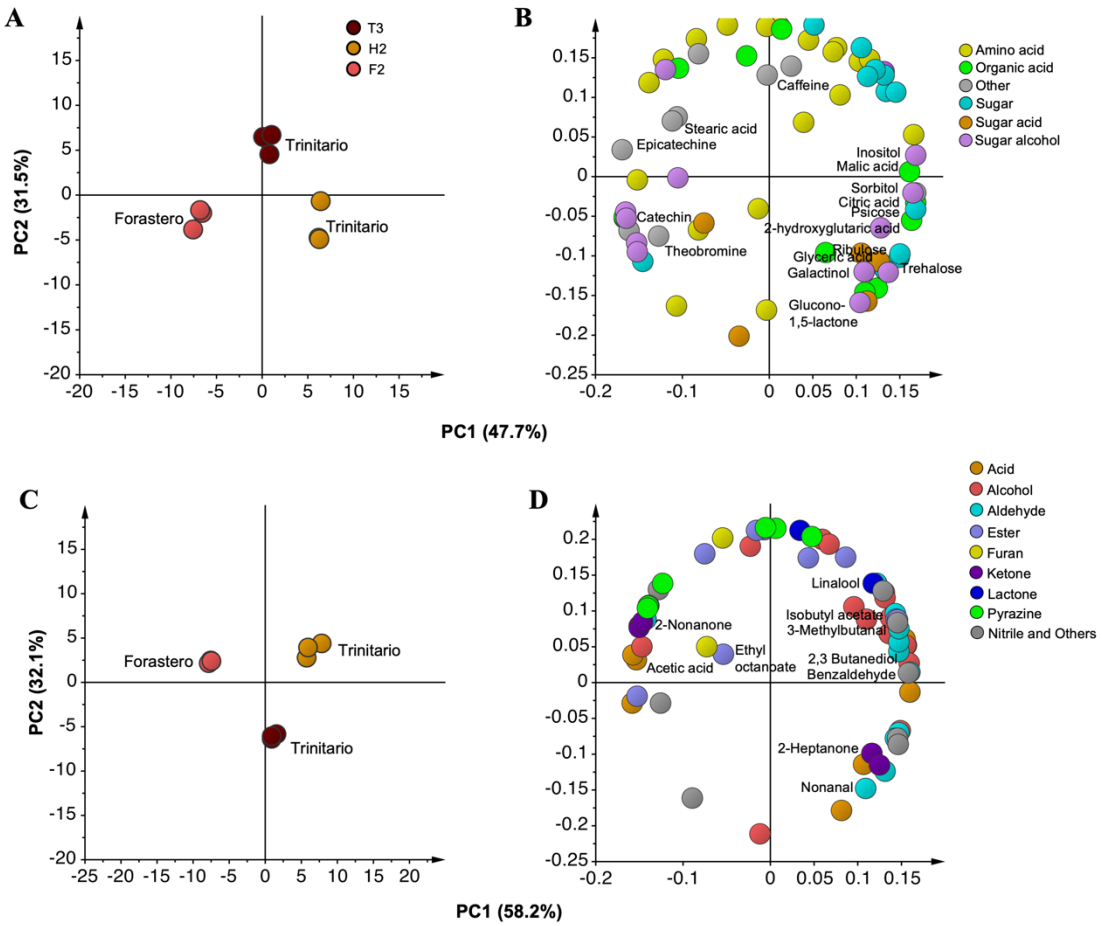


Fig. 6. The PCA score plots (A) and loading plots (B) for non-volatile metabolites, along with the score plot (C) and loading plot (D) for volatile metabolites, were generated from GC-MS-based metabolomic analysis using auto-scaling and no transformation, with three replicates. In the score plots, circles represent cacao clones, whereas in the loading plots, circle indicate metabolites contributing to the observed separation. Different colors correspond to various metabolite classes, and highlighted metabolite names denote the key compounds driving sample clustering.

T3, the male parent, had notably higher levels of gentiobiose, isoleucine, and methionine. Previous studies have reported that sugars and amino acids serve as precursors for flavor development, which may contribute to the subtle differences between T3 and Forastero. In contrast, F2, a Forastero variety, exhibited significantly higher levels of alkaloids and polyphenols, including theobromine, epicatechin, catechin, and stearic acid, compounds typically abundant in Forastero and inversely related to cocoa flavor quality. The correlation between these key compounds and the sensory attributes is further discussed in the correlation analysis.

The PCA of volatile components (Fig. 6C) consistently demonstrated a clear distinction among three cacao clones. Based on the loading score (Fig. 6D) and bar graph (Fig. 8), H2 hybrid was notably rich in volatile alcohols, aldehydes, pyrazines, and esters, including 2,3-butanediol, linalool, isobutyl acetate, 3-methylbutanal, 2-ethyl-3,5-dimethylpyrazine, and benzaldehyde. In contrast, male parent, T3 exhibited higher concentrations of 2-heptanone and nonanal, while Forastero F2 was predominantly characterized by acetic acid, ethyl octanoate, and 2-nonanone. Previous studies have identified benzaldehyde and 3-methylbutanal as key markers of Criollo-type cocoa liquor (considered the finest cocoa), whereas Forastero is known for its higher acetic acid and ethyl octanoate content. This study notably highlights these key volatile compounds in a

hybrid derived from cross breeding. The relationship between these genotypes' sensory profiles and volatile components is further explored in the correlation analysis.

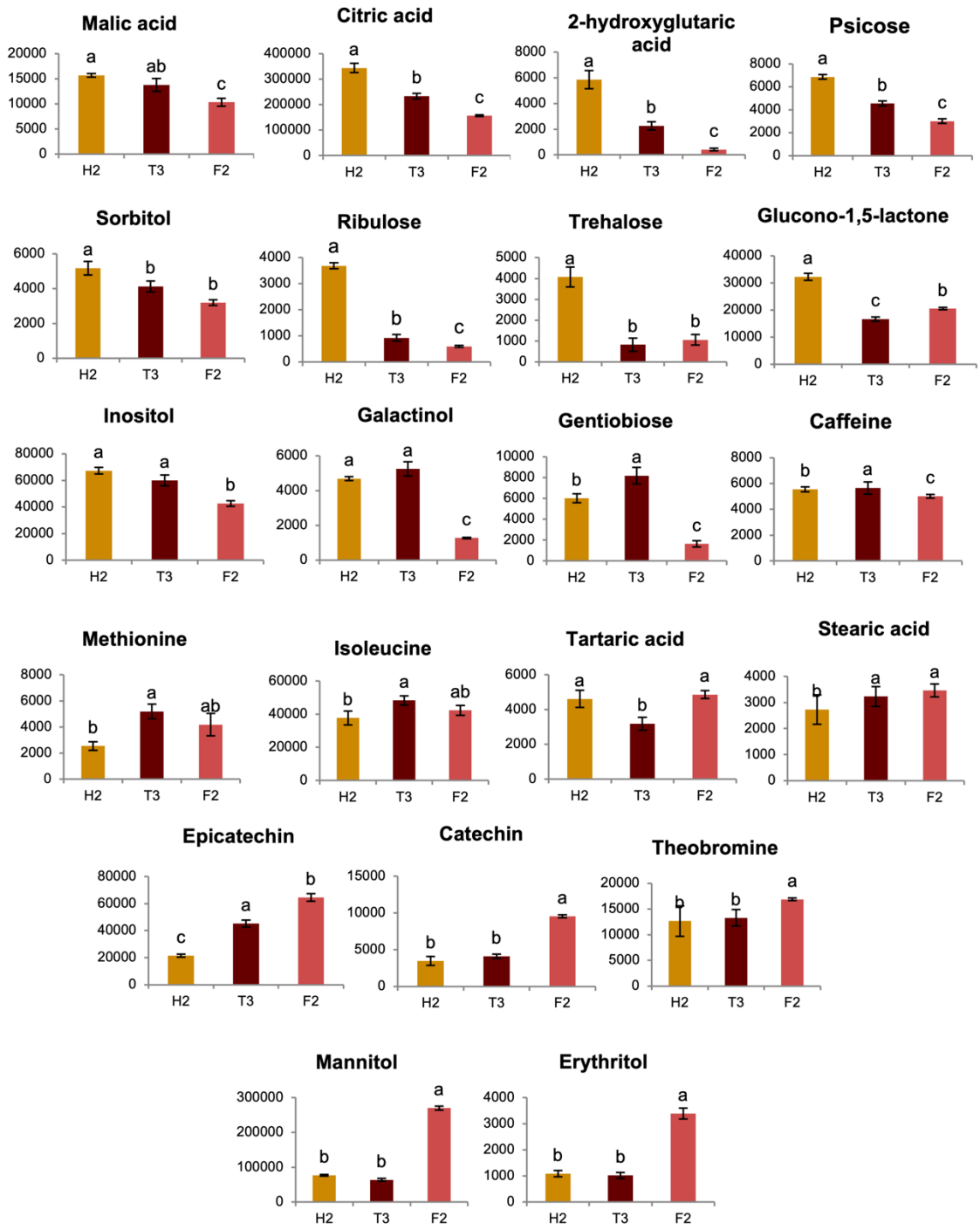


Fig. 7. The bar graphs display the relative intensity of the top VIP non-volatile metabolites highlighted in the loading plot, which are highly correlated with the sensory attributes shown in the PLS plot. The vertical axis represents the relative intensity, whereas the horizontal axis indicates the cacao clones. Bars labeled with the same letters do not show a statistically significant difference according to Tukey's HSD test, with a *p-value* < 0.05.

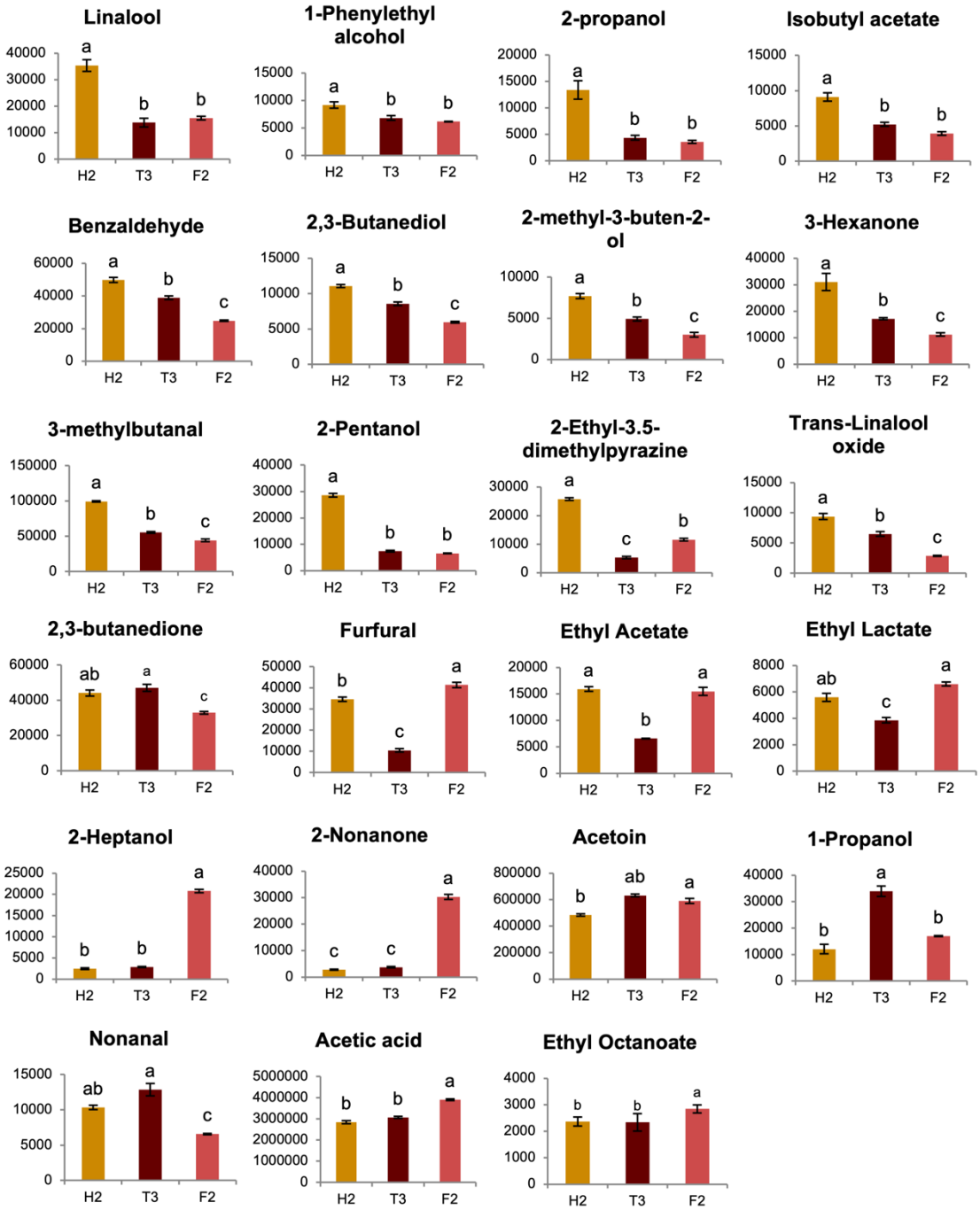


Fig. 8. The bar graphs represent the relative intensity of the top VIP volatile components highlighted in the loading plot, which are highly correlated with the sensory attributes shown in the PLS plot. The vertical axis represents the relative intensity, whereas the horizontal axis indicates the cacao clones. Bars labeled with the same letters do not show a statistically significant difference according to Tukey's HSD test, with a *p-value* < 0.05.

3.3.3 The correlation between sensory attributes and metabolite profiles

The relationship between sensory properties and metabolite profiles of cocoa liquors from different genotypes was analyzed using PLS regression (Fig. 9A, 9B). As illustrated in Fig. 9A, the model evaluates how non-volatile metabolites (x, n=70) contribute to sensory attributes (y, n=11). Since non-volatile metabolites play a key role in shaping the balance of taste and aroma in cocoa, this model incorporates all detected sensory profiles to investigate their correlation with flavor components in cocoa liquor from three cacao clones: F2, T3, and H2 hybrid. A previous study categorized sensory notes such as floral, fruity, nutty, cocoa, browned roast, browned fruit, and woody as aroma components. Therefore, a second PLS model (Fig. 9B) was developed to explore the relationship between volatile compounds (x, n=66) and sensory aroma attributes (y, n=7), focusing on the aroma characteristics of cocoa liquor.

The biplot (Fig. 9) shows H2 hybrid distinctly separated from Forastero (F2) owing to its markedly different sensory profile and metabolite composition. A 200-permutation test validated the PLS model by ensuring that Q^2 and R^2 values from permuted datasets were lower than those from the actual dataset. The results (Fig. S6) confirm strong correlations between sensory attributes and the volatile and non-volatile metabolite profiles of cocoa liquor.

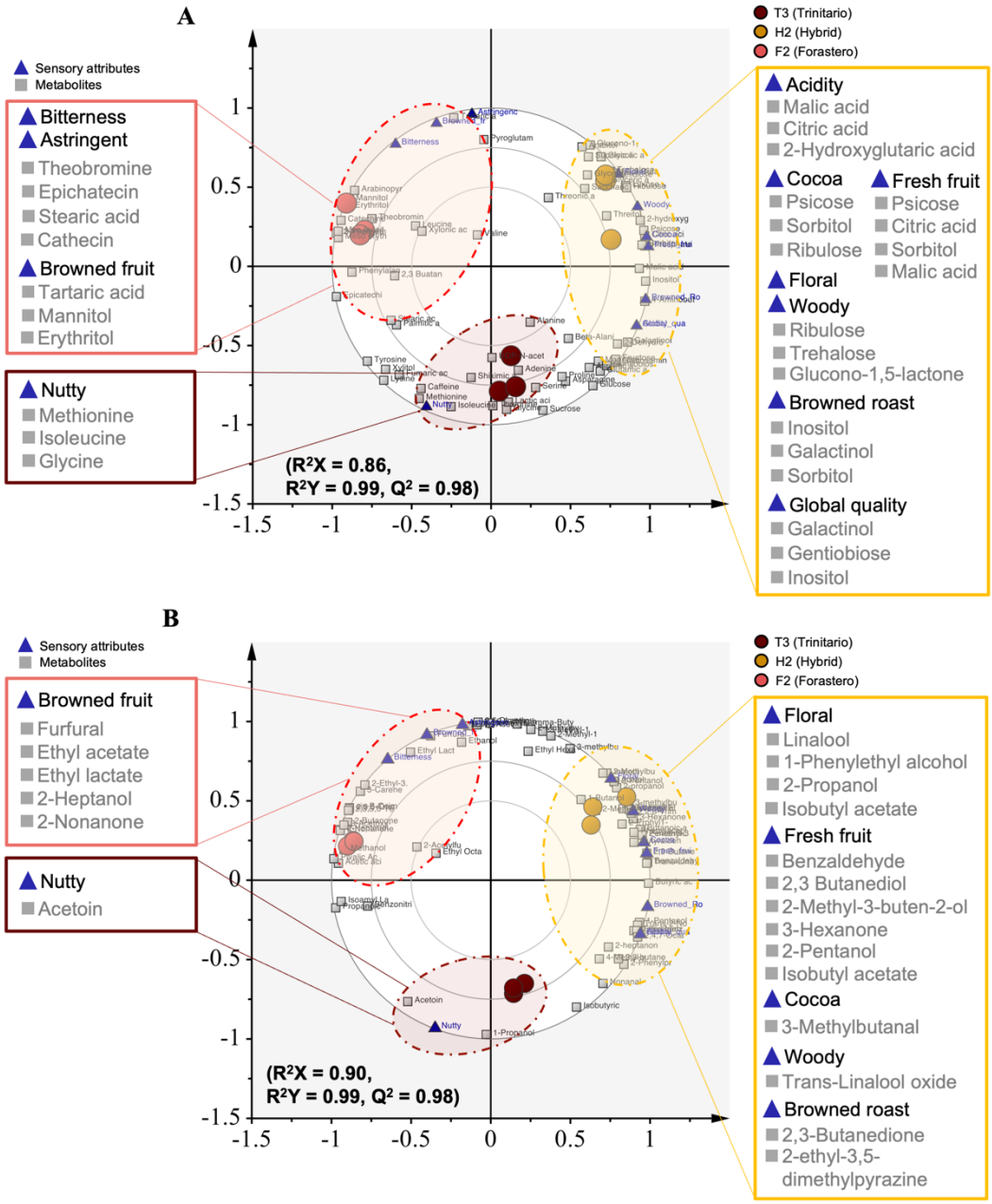


Fig. 9. PLS regression plot illustrating the relationships between sensory attributes and metabolite profiles of cacao genotypes: non-volatile compounds (A) and volatile compounds (B).

H2 exhibited distinctive sensory attributes, including floral, woody, cocoa, acidity, fresh fruit, browned roast, and global quality (Fig. 9A, Table S11). These attributes correlated with nonvolatile compounds such as organic acids, sugars, and sugar

derivatives. Floral and woody notes were linked to ribulose, trehalose, and glucono-1,5-lactone, while acidic attributes were associated with malic acid, citric acid, and 2-hydroxyglutaric acid. Fresh fruit notes correlated with psicose, citric acid, sorbitol, inositol, and malic acid, whereas cocoa notes were linked to psicose, sorbitol, and ribulose. Browned roast attributes were associated with inositol, galactinol, and sorbitol, while global quality correlated with galactinol, gentiobiose, and inositol. These findings highlight H2 hybrids exhibiting rich flavor profile. This result align with previous study which reported that the combination of certain organic acids and sugars enhanced the balance between taste and flavor in cocoa (Holm et al., 1993).

T3, male parent, characterized by a nutty flavor, which linked to methionine, isoleucine, and glycine. A study reported that glycine contributes to sweetness, whereas isoleucine and methionine contribute to bitterness (Bachmanov et al., 2016). However, no studies have yet reported an association between these compounds and the nutty flavor profile, although non-volatile compounds are known to influence the unique taste and aroma of cocoa (Castro-Alayo, Idrogo-Vásquez, et al., 2019).

The Forastero variety, F2 is characterized by dominant bitterness and astringency, linked to alkaloids and polyphenols such as theobromine, epicatechin, catechin, and stearic acid. These compounds known contributing to bitterness and astringency taste. Despite its intense taste, F2 also exhibited a slight browned fruit note, associated with sugar derivatives such as tartaric acid, mannitol, and erythritol. Although the chemical composition of non-volatile compounds provides valuable insights into flavor, it may not fully explain all the flavor components of cocoa liquor from different genotypes.

Therefore, in this study, we explored the correlation between sensory attributes and volatile components.

The correlation between sensory profiles and volatile compounds (Fig. 9B, Table S12) identified the top VIP metabolites that clearly differentiated H2 from its male parent (T3) and Forastero variety. H2 had significantly higher levels of linalool, 1-phenylethyl alcohol, 2-propanol, and isobutyl acetate, linked to a flowery aroma, as well as benzaldehyde, 2,3-butanediol, 2-methyl-3-buten-2-ol, 3-hexanone, 2-pentanol, and isobutyl acetate, contributing to fruity notes. Additionally, 3-methylbutanal was associated with a cocoa/malty aroma, while trans-linalool oxide, 2,3-butanedione, and 2-ethyl-3,5-dimethylpyrazine contributed to woody and browned roast notes. These findings align with previous studies on profiling volatile compounds in cocoa , but no prior research has quantitatively linked these volatile compounds to sensory attributes across different genotypes. This study establishes these correlations, underscoring H2's distinctive sensory profile and potential as high-quality cocoa.

T3 consistently grouped with nutty sensory profiles, linked to volatile compounds like acetoin. However, previous studies associate acetoin with a creamy aroma(Velásquez-Reyes et al., 2023)(Colonges, Jimenez, et al., 2022). Nutty, almond-like aromas are instead influenced by phenylacetaldehyde and pyrazines (e.g., 2-ethyl-3,5-dimethylpyrazine, 2,5-dimethylpyrazine, and 2,6-dimethylpyrazine). Since this study did not assess creamy aromas, future sensory analyses should include buttery flavors to clarify this correlation.

F2, a Forastero genotype, differed from H2 hybrid and T3, which were characterized by bitterness and astringency. However, it exhibited a browned fruit aroma

associated with furfural, ethyl acetate, ethyl lactate, 2-heptanol, and 2-nonanone compounds. This study did not include undesired flavors, such as the sour vinegar aroma, in the sensory analysis; therefore, the correlation analysis did not highlight that acetic acid was the predominant aroma in this Forastero genotype.

3.4 Conclusion

The correlation between the sensory attributes and metabolite profiles, both volatile and non-volatile compounds of selected cocoa hybrid was confirmed using PLS regression analysis. The correlation notably highlights the distinctive qualities of H2, expressing its potential as fine-flavor cocoa. This hybrid exhibits complex sensory attributes linked to some organic acids, sugars, and volatile compounds, including 3-methylbutanal, 2,3-butanediol, benzaldehyde, linalool, trans-linalool oxide, isobutyl acetate, 3-methyl-1-butanol, 2-methyl-3-buten-2-ol, and 1-phenylethyl alcohol.

Chapter 4

Conclusion

Cacao beans resulting from crossbreeding have been comprehensively characterized using a metabolomics approach. The initial chapter successfully investigates the characteristics of fine flavor cocoa through non-volatile metabolite profiling, using widely targeted Gas Chromatography-Mass Spectrometry (GC-MS) and bean phenotype analysis, which led to the identification of five key compounds differentiating fine flavor cocoa. This study notably identified organic acids such as malic acid, fumaric acid, citric acid, and tartaric acid as key metabolites distinguishing fine flavor cocoa from bulk cocoa, with the H1 hybrid (ICCRI 03) and H2 (ICCRI 09 clones) identified as promising candidates.

To provide a more detailed characterization of the selected hybrids, the correlation between sensory properties and metabolite profiles both non-volatile and volatile compounds was investigated using a combined metabolomics-based approach and sensory analysis. This study revealed that both non-volatile and volatile compounds (e.g., organic acids, certain sugars, 3-methylbutanal, 2,3-butanediol, benzaldehyde, linalool, trans-linalool oxide, and isobutyl acetate) are associated with outstanding sensory profiles, including floral, fruity, woody, cocoa, and browned roast notes, highlighting the potential of the H2 (ICCRI 09) hybrid as a fine flavor cocoa.

The results of this study provide a scientific basis for authenticating the high flavor quality of cocoa. Furthermore, the comprehensive information on cocoa's flavor characteristics will be valuable for future fine flavor cacao selection. Future research

should explore a broader range of cacao hybrids and investigate flavor changes across the post-harvest process to further enhance flavor characterization.

Acknowledgement

I would like to express my deepest gratitude to my supervisors, **Prof. Eiichiro Fukusaki, Assoc. Prof. Sastia Prama Putri, and Assoc. Prof. Shuichi Shimma** for their invaluable guidance, support, and encouragement throughout this study. Their expertise and insights have been instrumental in shaping this research.

I am also sincerely grateful to our collaborators, **Dr. Indah Anita Sari, Hendy Firmanto, M.Sc., Fitratin, Abdul Malik, and Dr. Agung Wahyu Susilo** from the **Indonesian Coffee and Cocoa Research Institute**, for their assistance and support during the field research. Their contributions have been crucial to the success of this study.

I would also like to extend my sincere appreciation to **Prof. Honda Kousuke** and **Prof. Aoki Wataru** for their valuable comments and suggestions, which have significantly contributed to the improvement of my dissertation.

A heartfelt appreciation goes to the **staff and members of the Fukusaki Laboratory** for their unwavering support, thoughtful discussions, and encouragement. Their kindness and support have made this journey all the more meaningful.

Lastly, my deepest gratitude goes to my husband, **Aris Setyawan**, for his incredible support, unwavering patience, and love. His dedication, especially in caring for our beloved daughter, **Keinara Arini Setyawan**, during my studies, has been my greatest source of strength. I could not have done this without him.

Thank you all sincerely and wholeheartedly.

References

Aprotosoaie, A. C., Luca, S. V., & Miron, A. (2016). Flavor Chemistry of Cocoa and Cocoa Products-An Overview. *Comprehensive Reviews in Food Science and Food Safety*, *15*(1), 73–91. https://doi.org/10.1111/1541-4337.12180

Bachmanov, A. A., Bosak, N. P., Glendinning, J. I., Inoue, M., Li, X., Manita, S., McCaughey, S. A., Murata, Y., Reed, D. R., Tordoff, M. G., & Beauchamp, G. K. (2016). Genetics of amino acid taste and appetite. In *Advances in nutrition (Bethesda, Md.)*, *7*(4), 806–822. https://doi.org/10.3945/an.115.011270

Bagnulo, E., Scavarda, C., Bortolini, C., Cordero, C., Bicchi, C., & Liberto, E. (2023). Cocoa quality: Chemical relationship of cocoa beans and liquors in origin identitation. *Food Research International*, *172*. https://doi.org/10.1016/j.foodres.2023.113199

Bekele, F., & Phillips-Mora, W. (2019). Cacao (Theobroma cacao L.) breeding. In *Advances in Plant Breeding Strategies: Industrial and Food Crops*, *6*, 409–487). Springer International Publishing. https://doi.org/10.1007/978-3-030-23265-8_12

Brunetto, M. del R., Gutiérrez, L., Delgado, Y., Gallignani, M., Zambrano, A., Gómez, Á., Ramos, G., & Romero, C. (2007). Determination of theobromine, theophylline and caffeine in cocoa samples by a high-performance liquid chromatographic method with on-line sample cleanup in a switching-column system. *Food Chemistry*, *100*(2), 459–467. https://doi.org/10.1016/j.foodchem.2005.10.007

Castro-Alayo, E. M., Idrogo-V Asquez, G., Ul Siche, R., & Cardenas-Toro, F. P. (2019). Formation of aromatic compounds precursors during fermentation of Criollo and Forastero cocoa. *Forastero Cocoa. Heliyon*, *5*, 1157. https://doi.org/10.1016/j.heliyon.2019

Castro-Alayo, E. M., Idrogo-Vásquez, G., Siche, R., & Cardenas-Toro, F. P. (2019). Formation of aromatic compounds precursors during fermentation of Criollo

and Forastero cocoa. In *Heliyon*, *5*(1). Elsevier Ltd. https://doi.org/10.1016/j.heliyon.2019.e01157

Collin, S., Fisette, T., Pinto, A., Souza, J., & Rogez, H. (2023). Discriminating aroma compounds in five cocoa bean genotypes from two Brazilian States: White Kerosene-like Catongo, Red Whisky-like FL89 (Bahia), Forasteros IMC67, PA121 and P7 (Pará). *Molecules*, *28*(4). https://doi.org/10.3390/molecules28041548

Colonges, K., Jimenez, J. C., Saltos, A., Seguine, E., Loor Solorzano, R. G., Fouet, O., Argout, X., Assemat, S., Davrieux, F., Cros, E., Lanaud, C., & Boulanger, R. (2022). 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*, *171*, 213–225. https://doi.org/10.1016/j.plaphy.2021.11.006

Colonges, K., Loor Solorzano, R. G., Jimenez, J. C., Lahon, M. C., Seguine, E., Calderon, D., Subia, C., Sotomayor, I., Fernández, F., Lebrun, M., Fouet, O., Rhoné, B., Argout, X., Costet, P., Lanaud, C., & Boulanger, R. (2022). Variability and genetic determinants of cocoa aromas in trees native to South Ecuadorian Amazonia. *Plants People Planet*, *4*(6), 618–637. https://doi.org/10.1002/ppp3.10268

Davrieux, F., Assemat, Sukha, Portillo, Boulanger, Bastianelli, & Cros E. (n.d.). Genotype characterization of cocoa into genetic groups through caffeine and theobromine content predicted by near infra red spectroscopy.

Devy, L., Susilo, A. W., Wachjar, A., & Sobir. (2019). Metabolite profiling of Indonesian cacao using Gas Chromatography-Mass Spectrometry. *IOP Conference Series: Earth and Environmental Science*, *347*(1). https://doi.org/10.1088/1755-1315/347/1/012071

Dillon, N. L., Zhang, D., Nauheimer, L., Toramo, E., Nagalevu, P., Melteras, M. V., Wallez, S., Finau, K., Nakidakida, S., Lepou, P., & Diczbalis, Y. (2023). Understanding the cocoa genetic resources in the Pacific to assist producers to

supply the growing craft market. *New Zealand Journal of Crop and Horticultural Science*. https://doi.org/10.1080/01140671.2023.2278788

Edem Kongor, J., Afoakwa, E. O., Takrama, J., Budu, A., Kongor, J. E., Felix Takrama, J., Simpson Budu, A., & Mensah-Brown, H. (2013). Effects of fermentation and drying on the fermentation index and cut test of pulp pre-conditioned Ghanaian cocoa (Theobroma cacao) beans. In *Journal of Food Science and Engineering*, *3*. https://www.researchgate.net/publication/270272357

Escobar, S., Santander, M., Zuluaga, M., Chacón, I., Rodríguez, J., & Vaillant, F. (2021). Fine cocoa beans production: Tracking aroma precursors through a comprehensive analysis of flavor attributes formation. *Food Chemistry*, *365*. https://doi.org/10.1016/j.foodchem.2021.130627

Fang, W., Meinhardt, L. W., Mischke, S., Bellato, C. M., Motilal, L., & Zhang, D. (2014). Accurate determination of genetic identity for a single cacao bean, using molecular markers with a nanofluidic system, ensures cocoa authentication. *Journal of Agricultural and Food Chemistry*, *62*(2), 481–487. https://doi.org/10.1021/jf404402v

Fang, Y., Li, R., Chu, Z., Zhu, K., Gu, F., & Zhang, Y. (2020). Chemical and flavor profile changes of cocoa beans (Theobroma cacao L.) during primary fermentation. *Food Science and Nutrition*, *8*(8), 4121–4133. https://doi.org/10.1002/fsn3.1701

Grissa, D., Pétéra, M., Brandolini, M., Napoli, A., Comte, B., & Pujos-Guillot, E. (2016). Feature selection methods for early predictive biomarker discovery using untargeted metabolomic data. *Frontiers in Molecular Biosciences*, *3*(7). https://doi.org/10.3389/fmolb.2016.00030

Gutiérrez, D. (n.d.). Compilers and Editors Brigitte Laliberté, Dolores Alvarado, Nadia Villaseñor (Bioversity International) and Sara Fusi (RB-ELLI).

Guzmán Penella, S., Boulanger, R., Maraval, I., Kopp, G., Corno, M., Fontez, B., & Fontana, A. (2023). Link between flavor perception and volatile compound composition of dark chocolates derived from trinitario cocoa beans from Dominican Republic. *Molecules*, *28*(9). https://doi.org/10.3390/molecules28093805

Herrera-Rocha, F., Cala, M. P., Aguirre Mejía, J. L., Rodríguez-López, C. M., Chica, M. J., Olarte, H. H., Fernández-Niño, M., & Gonzalez Barrios, A. F. (2021). Dissecting fine-flavor cocoa bean fermentation through metabolomics analysis to break down the current metabolic paradigm. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-01427-8

Herrera-Rocha, F., León-Inga, A. M., Aguirre Mejía, J. L., Rodríguez-López, C. M., Chica, M. J., Wessjohann, L. A., González Barrios, A. F., Cala, M. P., & Fernández-Niño, M. (2024). Bioactive and flavor compounds in cocoa liquor and their traceability over the major steps of cocoa post-harvesting processes. *Food Chemistry*, *435*. https://doi.org/10.1016/j.foodchem.2023.137529

Holm, C. S., Aston, J. W., & Douglas, K. (1993). The effects of the organic acids in cocoa on the flavour of chocolate. *Journal of the Science of Food and Agriculture*, *61*(1), 65–71. https://doi.org/10.1002/jsfa.2740610111

ICCO. (2017, September 2). *What is Fine Flavour Cocoa?*

Kim, K., Chun, I. J., Suh, J. H., & Sung, J. (2023). Relationships between sensory properties and metabolomic profiles of different apple cultivars. *Food Chemistry: X*, *18*. https://doi.org/10.1016/j.fochx.2023.100641

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

Krähmer, A., Engel, A., Kadow, D., Ali, N., Umaharan, P., Kroh, L. W., & Schulz, H. (2015). Fast and neat - Determination of biochemical quality parameters in cocoa using near infrared spectroscopy. *Food Chemistry*, *181*, 152–159. https://doi.org/10.1016/j.foodchem.2015.02.084

Lachenaud, P., & Motamayor, J. C. (2017). The Criollo cacao tree (Theobroma cacao L.): a review. *Genetic Resources and Crop Evolution*, *64*(8), 1807–1820. https://doi.org/10.1007/s10722-017-0563-8

Le, Q. T. N., Sugi, N., Yamaguchi, M., Hirayama, T., Kobayashi, M., Suzuki, Y., Kusano, M., & Shiba, H. (2023). Morphological and metabolomics profiling of intraspecific Arabidopsis hybrids in relation to biomass heterosis. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-36618-y

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

Michel, S., Baraka, L. F., Ibañez, A. J., & Mansurova, M. (2021). Mass spectrometry-based flavor monitoring of peruvian chocolate fabrication process. *Metabolites*, *11*(2), 1–16. https://doi.org/10.3390/metabo11020071

Moreno-Rojas, J. M., Yadira Erazo Solorzano, C., Tuárez García, D. A., Pereira-Caro, G., Ordóñez Díaz, J. L., Muñoz-Redondo, J. M., & Rodríguez-Solana, R. (2023). Impact of the pre-drying process on the volatile profile of on-farm processed Ecuadorian bulk and fine-flavour cocoa varieties. *Food Research International*, *169*. https://doi.org/10.1016/j.foodres.2023.112938

Nguyen, V. T., Tran, A. X., & Le, V. A. T. (2021). Microencapsulation of phenolic-enriched extract from cocoa pod husk (Theobroma cacao L.). *Powder Technology*, *386*, 136–143. https://doi.org/10.1016/j.powtec.2021.03.033

Oliva-Cruz, M., Goñas, M., García, L. M., Rabanal-Oyarse, R., Alvarado-Chuqui, C., Escobedo-Ocampo, P., & Maicelo-Quintana, J. L. (2021). Phenotypic

Characterization of Fine-Aroma Cocoa from Northeastern Peru. *International Journal of Agronomy*, *2021*. https://doi.org/10.1155/2021/2909909

Palma-Morales, M., Rune, C. J. B., Castilla-Ortega, E., Giacalone, D., & Rodríguez-Pérez, C. (2024). Factors affecting consumer perception and acceptability of chocolate beverages. *LWT*, *201*. https://doi.org/10.1016/j.lwt.2024.116257

Putri, S. P., Ikram, M. M. M., Sato, A., Dahlan, H. A., Rahmawati, D., Ohto, Y., & Fukusaki, E. (2022). Application of gas chromatography-mass spectrometry-based metabolomics in food science and technology. In *Journal of Bioscience and Bioengineering* (Vol. 133, Issue 5, pp. 425–435). Elsevier B.V. https://doi.org/10.1016/j.jbiosc.2022.01.011

Qin, X. W., Lai, J. X., Tan, L. H., Hao, C. Y., Li, F. P., He, S. Z., & Song, Y. H. (2017). Characterization of volatile compounds in Criollo, Forastero, and Trinitario cocoa seeds (Theobroma cacao L.) in China. *International Journal of Food Properties*, *20*(10), 2261–2275. https://doi.org/10.1080/10942912.2016.1236270

Rottiers, H., Tzompa Sosa, D. A., De Winne, A., Ruales, J., De Clippeleer, J., De Leersnyder, I., De Wever, J., Everaert, H., Messens, K., & Dewettinck, K. (2019). Dynamics of volatile compounds and flavor precursors during spontaneous fermentation of fine flavor Trinitario cocoa beans. *European Food Research and Technology*, *245*(9), 1917–1937. https://doi.org/10.1007/s00217-019-03307-y

Santander, M., Vaillant, F., Sinuco, D., Rodríguez, J., & Escobar, S. (2021). Enhancement of fine flavour cocoa attributes under a controlled postharvest process. *Food Research International*, *143*. https://doi.org/10.1016/j.foodres.2021.110236

Sari, I. A., Murti, R. H., Misnawi, Putra, E. T. S., & Susilo, A. W. (2022). Sensory profiles of cocoa genotypes in Indonesia. *Biodiversitas*, *23*(2), 648–654. https://doi.org/10.13057/biodiv/d230205

Sari, I. A., Setyawan, B., Wahyu Susilo, A., Fitri Isnaini, N., Paputpungan, S., Nur'aini, F., & Solecha Ruseani, N. (2022). Flush characteristics of several cocoa genotypes different in resistant to vascular streak dieback. *Pelita Perkebunan (a Coffee and Cocoa Research Journal)*, *38*(2), 120–127. https://doi.org/10.22302/iccri.jur.pelitaperkebunan.v38i2.518

Seguine, E. S. (2009). Evidence for the Effect of the Cocoa Bean Flavour Environment during Fermentation on the Final Flavour Profile of Cocoa Liquor and Chocolate. https://www.researchgate.net/publication/228754936

Smulders, M. J. M., Esselink, D., Amores, F., Ramos, G., Sukha, D. A., Butler, D. R., Vosman, B., & Van Loo, E. N. (n.d.). Identification of Cocoa (Theobroma cacao L.) Varieties with Different Quality Attributes and Parentage Analysis of Their Beans. www.ICCO.org

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

Tscharntke, T., Ocampo-Ariza, C., Vansynghel, J., Ivañez-Ballesteros, B., Aycart, P., Rodriguez, L., Ramirez, M., Steffan-Dewenter, I., Maas, B., & Thomas, E. (2023). Socio-ecological benefits of fine-flavor cacao in its center of origin. *Conservation Letters*, *16*(1). https://doi.org/10.1111/conl.12936

Velásquez-Reyes, D., Rodríguez-Campos, J., Avendaño-Arrazate, C., Gschaedler, A., Alcázar-Valle, M., & Lugo-Cervantes, E. (2023). Forastero and Criollo cocoa beans, differences on the profile of volatile and non-volatile compounds in the process from fermentation to liquor. *Heliyon*, *9*(4). https://doi.org/10.1016/j.heliyon.2023.e15129

Vincent Colantonioa, 1, Luis Felipe V. Ferra~oa, 1, Denise M. Tiemana, Nikolay Bliznyukb, c, d, Charles Simse, Harry J. Kleea, 2, Patricio Munoza, 2, and

Marcio F. R. Resende Jr. (2022). Metabolomic selection for enhanced fruit flavor. *PNAS*, *119*, 1–11.

Zapata-Alvarez, A., Bedoya-Vergara, C., Porras-Barrientos, L. D., Rojas-Mora, J. M., Rodríguez-Cabal, H. A., Gil-Garzon, M. A., Martinez-Alvarez, O. L., Ocampo-Arango, C. M., Ardila-Castañeda, M. P., & Monsalve-F, Z. I. (2024). Molecular, biochemical, and sensorial characterization of cocoa (Theobroma cacao L.) beans: A methodological pathway for the identification of new regional materials with outstanding profiles. *Heliyon*, *10*(3). https://doi.org/10.1016/j.heliyon.2024.e24544

Zhang, W., Bai, B., Du, H., Hao, Q., Zhang, L., Chen, Z., Mao, J., Zhu, C., Yan, M., Qin, H., & Abd El-Aty, A. M. (2024). Co-expression of metabolites and sensory attributes through weighted correlation network analysis to explore flavor-contributing factors in various Pyrus spp. Cultivars. *Food Chemistry: X*, *21*. https://doi.org/10.1016/j.fochx.2024.101189

Zhao, L., Shang, S., Tian, Y., Gao, Y., Song, Z., Peng, L., Li, Z., & Wang, B. (2023). Integrative analysis of sensory evaluation and non-targeted metabolomics to unravel tobacco leaf metabolites associated with sensory quality of heated tobacco. *Frontiers in Plant Science*, *14*. https://doi.org/10.3389/fpls.2023.1123100

Zhao, Y., Zhang, Y., Yang, H., Xu, Z., Li, Z., Zhang, Z., Zhang, W., & Deng, J. (2024). A comparative metabolomics analysis of phytochemcials and antioxidant activity between broccoli floret and by-products (leaves and stalks). *Food Chemistry*, *443*. https://doi.org/10.1016/j.foodchem.2024.138517

List of publications

Original paper

Afifah, E.N., Sari, I.A., Susilo, A.W., Malik, A., Fukusaki, E., & Putri, S.P. (2024). Characterization of fine-flavor cocoa in parent-hybrid combinations using metabolomics approach. *Food Chemistry: X*, 24, 101832. https://doi.org/10.1016/j.fochx.2024.101832.

Afifah, E.N., Sari, I.A., Susilo, A.W., Firmanto, H., Malik, A., Fukusaki, E., & Putri, S.P. (2025). Correlation between sensory attributes and metabolomic profiles of cocoa liquor from different cacao genotypes. Food Chemistry: X, 28, 102498. https://doi.org/10.1016/j.fochx.2025.102498.

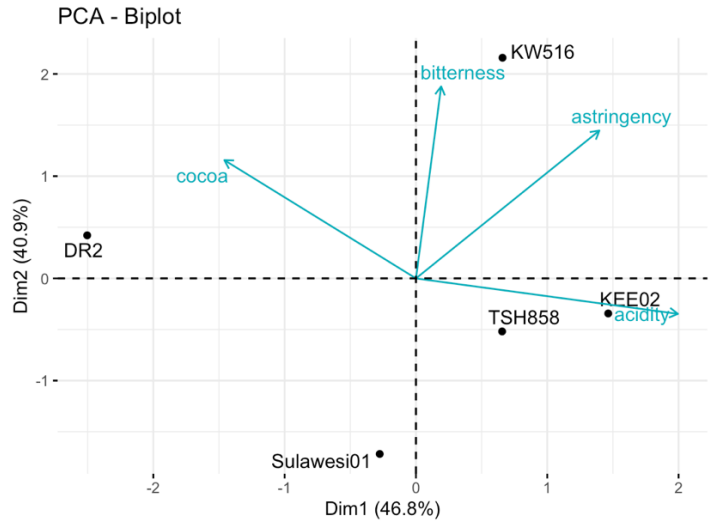
Presentations:

1. **Afifah, E.N.,** Sari, I.A., Susilo, A.W., Malik, A., Fukusaki, E., & Putri, S.P. Characterization of fine-flavor cocoa in parent-hybrid combinations using metabolomics approach. *(Poster presentation) 20th Annual Conference of the Metabolomic Society. 2024.*

2. **Afifah, E.N.,** Sari, I.A., Susilo, A.W., Malik, A., Fukusaki, E., & Putri, S.P. Correlation between Sensory Attributes and Metabolomic Profiles of Cocoa Liquor from Different Cacao Genotypes. *(Online presentation) 76th Annual Society of Biotechnology Japan Meeting. 2024.*

Supplementary

(A)



(B)

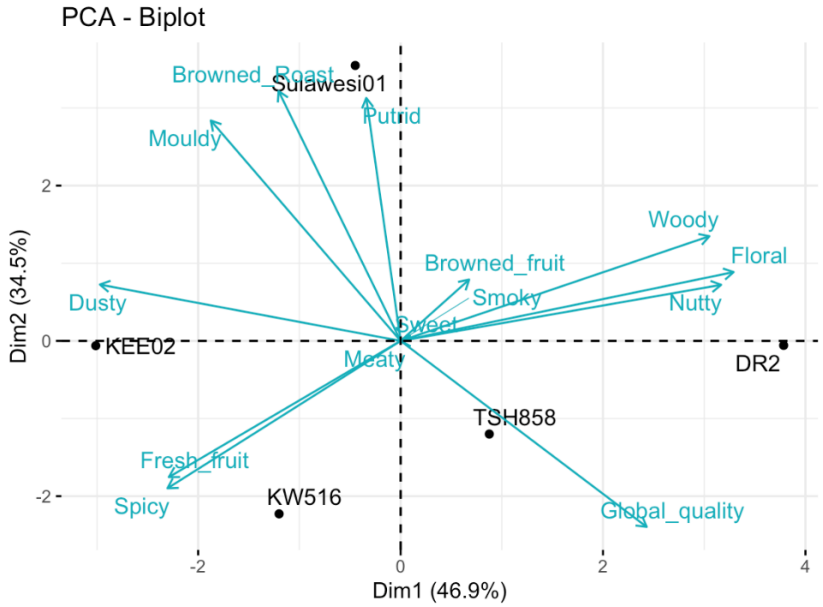


Figure S1. Biplot analysis of sensory data from five cacao clones using 2021 sample batch, showing (A) taste and (B) flavor attributes. The sensory analysis was conducted

by The Indonesian Coffee and Cocoa Research Institute (ICCRI) and Guittard Chocolate Company. 3 replicates (3 certified panelists). Data were transformed using $\sqrt{X} + 1$.

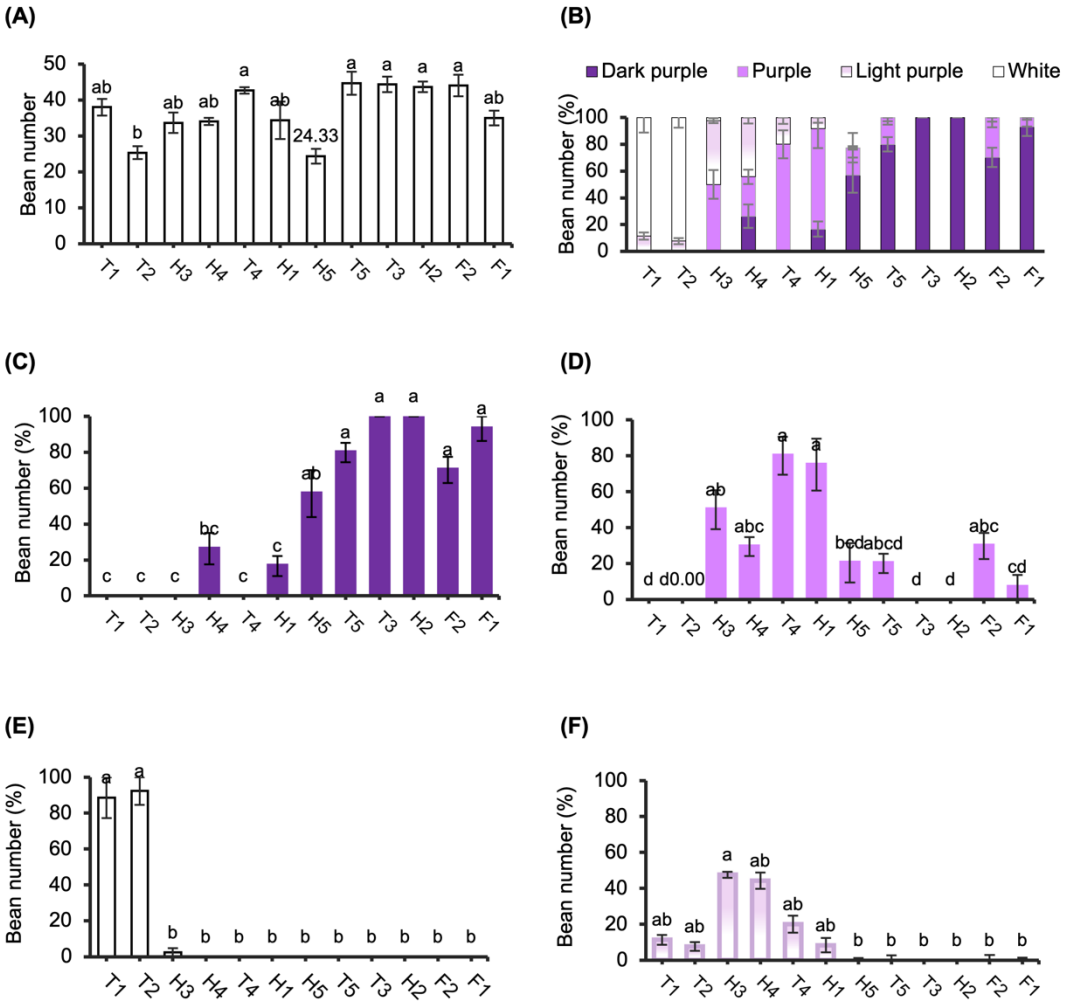


Figure S2. Bar chart of bean number (A), the total percentage of fresh bean color per pod (B), the percentage of dark purple bean per pod (C), the percentage of purple bean per pod (D), the percentage of white bean per pod (E), and the percentage of light purple bean per pod (F).

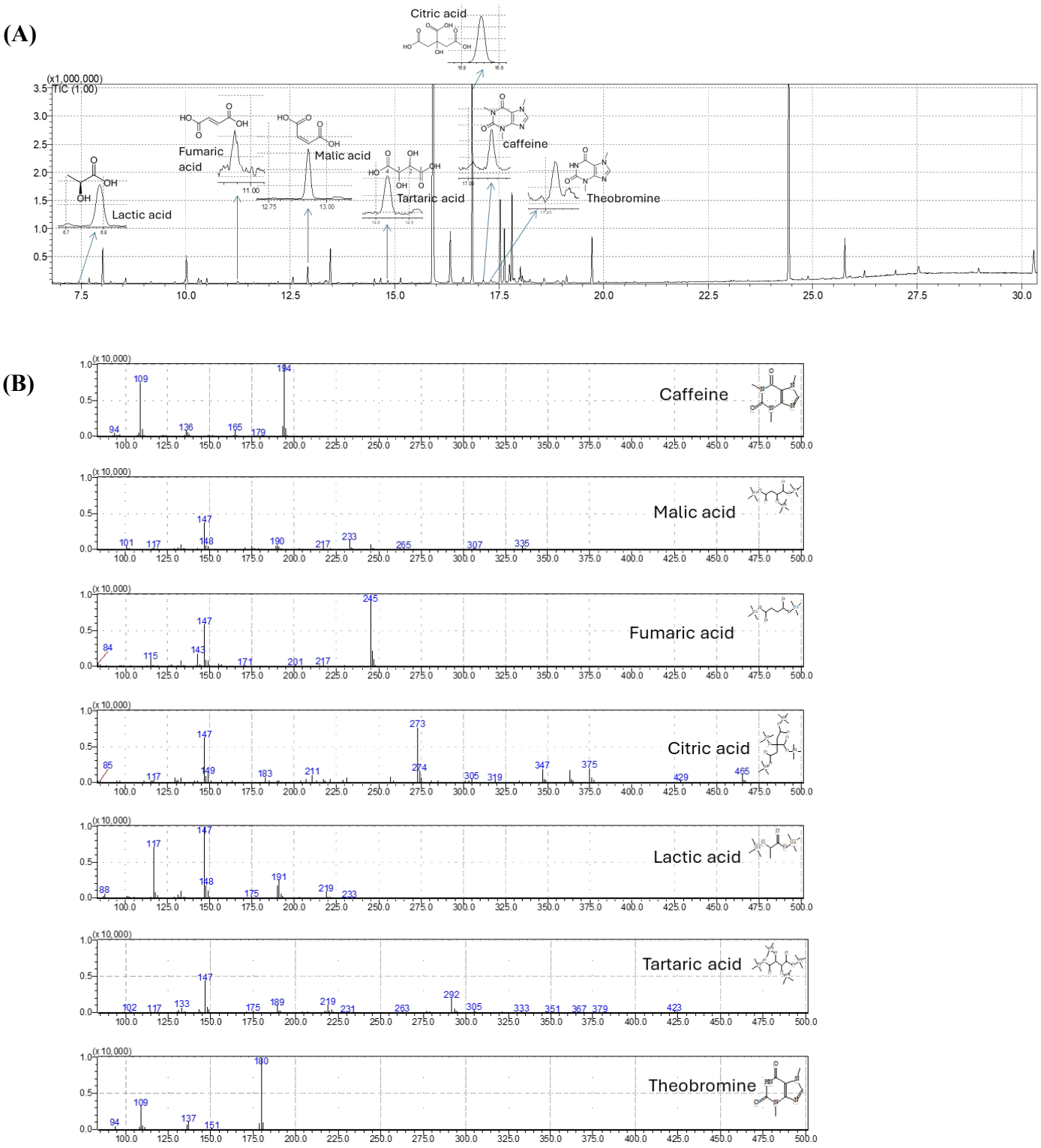


Figure S3. Total Ion Chromatogram (TIC) (A) and Mass spectrum (B) obtained from Gas Chromatography-Mass Spectrometry (GC-MS) analysis.

Table S1. Loading Scores from PCA of all samples based on PC1 and PC2

No.	Metabolite name	PC1	Metabolite name	PC2
1	Malic acid	0.15537	4-Aminobutyric acid	0.185122
2	Fumaric acid	0.15528	Asparagine	0.184114
3	Glucarate	0.151526	Glycine	0.178483
4	Phosphate	0.146305	Glutamine	0.174896
5	Mannitol	0.139852	Glutamic acid	0.174656
6	3-Hydroxy-3-methylglutarate	0.130539	Tryptophan	0.162934
7	Propyleneglycol	0.128417	Serine	0.162106
8	Inositol	0.122103	Isocitric acid	0.161904
9	Malonic acid	0.11657	Lysine	0.157533
10	2,3-Butanediol	0.115287	Histidine	0.155478
11	Sorbitol	0.114206	Pyroglutamic acid	0.149318
12	Dopamine	0.109148	Threonine	0.14669
13	Glycerol	0.105147	Aspartic acid	0.145712
14	Serotonin	0.101497	Serotonin	0.139462
15	2-Isopropylmalic acid	0.095145	Inositol	0.134823
16	Glyceric acid	0.090495	Citric acid	0.133392
17	Gluconic acid	0.088062	L-Threonine	0.128657
18	Arabionose	0.084783	Ribose	0.12226
19	Xylonic acid	0.083961	Tyrosine	0.119156
20	Tartaric acid	0.083788	Dopamine	0.118979
21	Sucrose	0.079299	2-Aminoethanol	0.113194
22	2-Aminoethanol	0.070912	Succinic acid	0.110367
23	Glutamic acid	0.053159	Galactose	0.107338
24	Citric acid	0.048078	Sucrose	0.105519
25	4-Aminobutyric acid	0.040781	Alanine	0.104656
26	Isocitric acid	0.037436	Valine	0.100502
27	L-Aspartic acid	0.033814	Glucose	0.097415
28	Aspartic acid	0.033121	Phenylalanine	0.091688
29	Catechine	0.030414	L-Aspartic acid	0.091172
30	Meso erythritol	0.014913	Malic acid	0.088849
31	Glucosamine	0.013875	Proline	0.086687
32	Histidine	0.011113	Isoleucine	0.08523
33	Xylitol	0.010095	2-Isopropylmalic acid	0.063221
34	Threonic acid	0.00908	2-Dehydro gluconate_2	0.056689
35	Ribose	0.006816	Leucine	0.056392
36	Succinic acid	-0.00428	Arabionose	0.055203
37	Glutamine	-0.00508	Glucarate	0.050283

38	D-Glucopyranoside	-0.00838	Phosphate	0.04235
39	Asparagine	-0.01327	Sorbitol	0.035244
40	Pyroglutamic acid	-0.01521	Xylonic acid	0.032081
41	Mannose	-0.01973	Fructose	0.031311
42	Quinic acid	-0.03447	Sorbose	0.031132
43	Caffeine	-0.03835	Lactic acid	0.030775
44	Shikimic acid	-0.03871	Fumaric acid	0.025225
45	Lactic acid	-0.04306	3-Hydroxy-3-methylglutarate	0.008084
46	Proline	-0.04986	Malonic acid	-0.00083
47	2-Dehydro gluconate_2	-0.0553	Mannitol	-0.00089
48	Phthalic acid	-0.05709	Tartaric acid	-0.00267
49	Tryptophan	-0.05959	Glycerol	-0.00456
50	Glucose	-0.06771	Xylitol	-0.02181
51	L-Threonine	-0.06994	2,3-Butanediol	-0.02358
52	Serine	-0.07368	Propyleneglycol	-0.02452
53	Glycine	-0.07777	L-Tyrosine	-0.02963
54	Erythritol	-0.09048	Gluconic acid	-0.0311
55	Threonine	-0.09217	Phthalic acid	-0.0354
56	Fructose	-0.09236	Catechine	-0.03852
57	Sorbose	-0.09281	Theobromine	-0.05528
58	Theobromine	-0.09967	Mannose	-0.05789
59	Galactose	-0.11896	D-Glucopyranoside	-0.0673
60	Lysine	-0.11963	Caffeine	-0.08018
61	Alanine	-0.13754	Meso erythritol	-0.08454
62	Valine	-0.14283	Erythritol	-0.09089
63	Phenylalanine	-0.14295	Glyceric acid	-0.09956
64	Isoleucine	-0.14564	Glucosamine	-0.10778
65	Tyrosine	-0.14975	Threonic acid	-0.10924
66	L-Tyrosine	-0.16723	Shikimic acid	-0.11309
67	Leucine	-0.17735	Quinic acid	-0.12399

Table S2. Loading Scores from PCA of fine flavor cacao and bulk cacao based on PC1 and PC2

No.	**Metabolite name**	**PC1**	**Metabolite name**	**PC2**
1	Glutamine	0.137782	D-Glucopyranoside	0.188099
2	Malic acid	0.137375	Glucosamine	0.181372

3	Fumaric acid	0.136991	Quinic acid	0.174786
4	Glyceric acid	0.135164	Inositol	0.172201
5	Caffeine	0.134886	Malonic acid	0.171421
6	Glucarate	0.132729	Meso erythritol	0.155916
7	Isocitric acid	0.128854	Mannitol	0.153711
8	Sucrose	0.128271	Sorbitol	0.14892
9	2-Isopropylmalic acid	0.116442	Aspartic acid	0.1442
10	Xylonic acid	0.113633	Histidine	0.137248
11	Tartaric acid	0.112232	Tryptophan	0.131404
12	Shikimic acid	0.110881	Pyroglutamic acid	0.129561
13	Glutamic acid	0.101532	Mannose	0.121269
14	L-Threonine	0.0882458	Dopamine	0.11935
15	Sorbitol	0.0853805	Lysine	0.113351
16	Citric acid	0.0846837	Gluconic acid	0.110355
17	Glycerol	0.0787341	2-Isopropylmalic acid	0.109541
18	Lactic acid	0.075984	Phenylalanine	0.103547
19	4-Aminobutyric acid	0.0750791	Threonic acid	0.10128
20	Aspartic acid	0.0747173	Proline	0.094672
21	Pyroglutamic acid	0.0686278	Theobromine	0.08918
22	Threonine	0.0605013	Glutamic acid	0.085218
23	Xylitol	0.0598639	Glycine	0.077136
24	2-Dehydro gluconate_2	0.0578793	Serotonin	0.0753
25	Inositol	0.0552025	Catechine	0.074042
26	Meso erythritol	0.0403522	Shikimic acid	0.06299
27	Threonic acid	0.039443	Glucarate	0.058602
28	3-Hydroxy-3-methylglutarate	0.0306709	2,3-Butanediol	0.049459
29	Serotonin	0.0276665	Xylitol	0.046842
30	Catechine	0.0258121	Phthalic acid	0.040723
31	Propyleneglycol	0.0235857	Tyrosine	0.037961
32	D-Glucopyranoside	0.0211861	Glutamine	0.030649
33	Ribose	0.0110804	Xylonic acid	0.028666
34	Malonic acid	0.00821319	4-Aminobutyric acid	0.026695
35	Mannitol	-0.0010474	Isoleucine	0.015776

36	Proline	-0.0060683	Succinic acid	0.015489
37	Glucose	-0.029082	Serine	0.014041
38	Asparagine	-0.0418704	2-Dehydro gluconate_2	-0.00022
39	Isoleucine	-0.0426153	Glyceric acid	-0.01637
40	Erythritol	-0.0459604	Valine	-0.01645
41	Glucosamine	-0.0503821	Fumaric acid	-0.01665
42	Quinic acid	-0.0535739	Leucine	-0.01691
43	Histidine	-0.0542296	Threonine	-0.02097
44	2-Aminoethanol	-0.0583417	L-Threonine	-0.02607
45	L-Aspartic acid	-0.0685118	Malic acid	-0.02898
46	Phenylalanine	-0.0727132	L-Tyrosine	-0.03178
47	Mannose	-0.0775188	Galactose	-0.03413
48	Phthalic acid	-0.0784151	Alanine	-0.03633
49	Dopamine	-0.0836833	L-Aspartic acid	-0.0364
50	Serine	-0.0991687	Glycerol	-0.03987
51	Gluconic acid	-0.0992096	Sucrose	-0.04106
52	Valine	-0.100625	Caffeine	-0.04388
53	Tryptophan	-0.103035	Asparagine	-0.0471
54	Galactose	-0.105187	Phosphate	-0.05046
55	Fructose	-0.106496	Isocitric acid	-0.05253
56	Sorbose	-0.107083	2-Aminoethanol	-0.05863
57	2,3-Butanediol	-0.109902	Erythritol	-0.07555
58	Lysine	-0.116703	Fructose	-0.08033
59	Arabionose	-0.118199	Lactic acid	-0.08516
60	Theobromine	-0.12135	Sorbose	-0.09021
61	Phosphate	-0.12236	Arabionose	-0.09464
62	Leucine	-0.122965	Citric acid	-0.10821
63	Glycine	-0.125135	Tartaric acid	-0.11474
64	Tyrosine	-0.131789	Ribose	-0.11847
65	Alanine	-0.135574	3-Hydroxy-3-methylglutarate	-0.1212
66	Succinic acid	-0.139647	Propyleneglycol	-0.13237
67	L-Tyrosine	-0.140283	Glucose	-0.15721

Table S3. Assessment of the correlation models construdted by OLPS-R analysis

No.	Attributes	N	Latent variable	R^2X (cum)	R^2Y (cum)	Q^2 (cum)	*p-value*
1	L*	33	1+3+0	0.596	0.968	0.888	1.23e-09
2	a*	33	1+3+0	0.593	0.958	0.883	2.09e-09
3	b*	33	1+3+0	0.642	0.931	0.726	2.81e-08
4	Percentage of white bean per pod	33	1+3+0	0.726	0.986	0.895	2.11e-09
5	Percentage of dark purple bean per pod	33	1+3+0	0.668	0.987	0.869	2.356e-08

N: number of sample used to construct OPLS-R analysis

Table S4. List of the top ten metabolites with VIP values greater than 1 from the OPLS-R analysis for the L* value

No.	**Metabolite name**	**VIP score**	**Coefficient**
1	Caffeine*	2.45835	0.849526
2	Tartaric acid*	2.15996	0.545985
3	Threonic acid	1.89783	0.514354
4	Glyceric acid*	1.86079	0.400589
5	Glucosamine	1.84477	0.322079
6	Meso erythritol	1.78584	0.884982
7	Tryptophan	1.75178	0.363086
8	3,4-Dihydroxybenzoate	1.71536	0.42061
9	Sucrose	1.63998	0.323999
10	Glycine	1.63732	0.155801

(*) highlighted in the bar chart

Table S5. List of the top ten metabolites with VIP values greater than 1 from the OPLS-R analysis for the a* value.

No.	**Metabolite name**	**VIP score**	**Coefficient**
1	2,3-Butanediol	2.2803	0.655385
2	Arabionose	2.05525	0.561594
3	Mannitol	1.97764	0.543135
4	Fumaric acid*	1.95712	0.43896
5	Propyleneglycol	1.92796	0.575483

6	Caffeine*	1.91107	1.00015
7	Sorbitol	1.66574	0.546568
8	Glucarate	1.64976	0.344166
9	Phosphate	1.60755	0.155797
10	Gluconic acid	1.60738	0.922878

(*) highlighted in the bar chart

Table S6. List of the top ten metabolites with VIP values greater than 1 from the OPLS-R analysis for the b* value.

No.	Metabolite name	VIP score	Coefficient
1	Caffeine*	3.19924	0.519175
2	2-Aminoethanol	2.32127	0.367004
3	Glycolic acid	1.94488	0.897934
4	Tartaric acid*	1.90562	0.325048
5	Meso erythritol	1.7532	1.26066
6	3,4-Dihydroxybenzoate	1.6162	0.297407
7	Glucosamine	1.55688	0.363233
8	Arabionose	1.54366	0.430822
9	Sucrose	1.53048	0.294988
10	Glyceric acid*	1.50946	0.285054

(*) highlighted in the bar chart

Table S7. List of the top ten metabolites with VIP values greater than 1 from the OPLS-R analysis for the percentage of white bean.

No.	Metabolite name	VIP score	Coefficient
1	Caffeine*	2.99024	0.461239
2	2-Aminoethanol	2.3161	0.261443
3	Glyceric acid*	2.06979	0.359425
4	Meso erythritol	1.94546	1.14169
5	3,4-Dihydroxybenzoate	1.78471	0.272794
6	Glucosamine	1.77913	0.310551
7	Mannose	1.72103	0.377974
8	Glyceric acid	1.64576	0.346883
9	Sucrose	1.52606	0.196998
10	Tartaric acid*	1.45017	0.441418

(*) highlighted in the bar chart

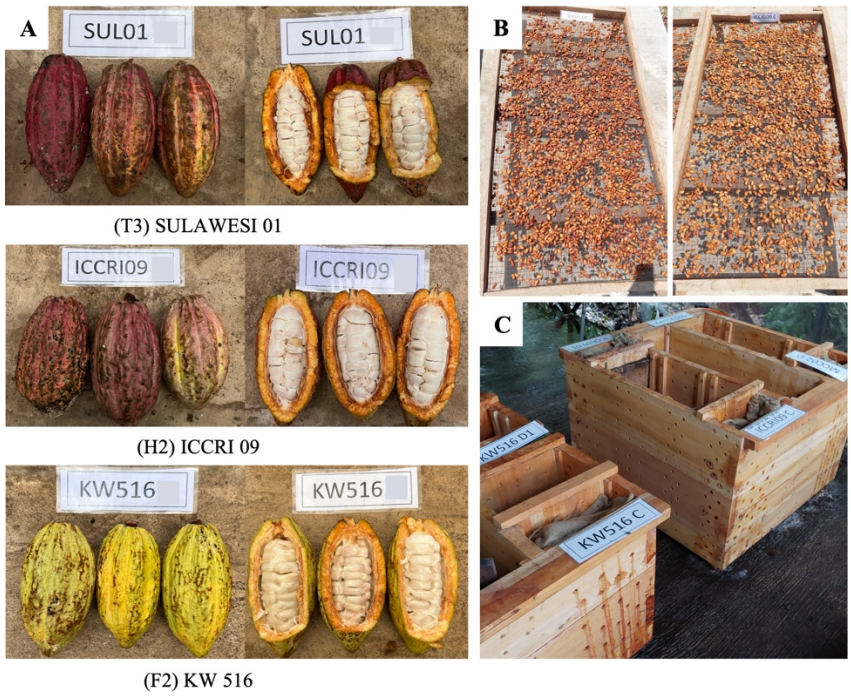
Table S8. List of the top ten metabolites with VIP values greater than 1 from the OPLS-R analysis for the percentage of dark purple bean.

No.	Metabolite name	VIP score	Coefficient
1	Alanine*	1.90728	0.877365
2	Lysine	1.81866	0.339679
3	Glycine	1.74785	0.455654
4	Asparagine	1.68066	0.575365
5	Serine*	1.63671	0.528898
6	Tryptophan	1.58596	0.826745
7	Oxalate	1.55473	0.758194
8	Glyceric acid	1.50946	0.285054
9	Glutamic acid*	1.49886	0.389418
10	2-Aminoethanol	1.25147	1.43097

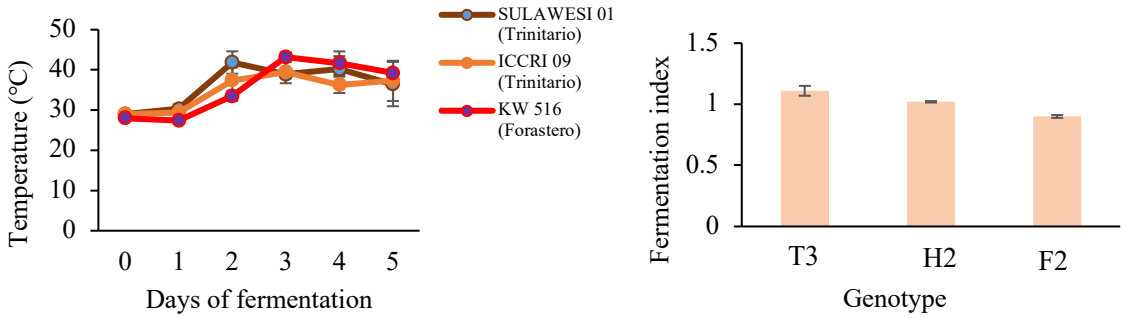
(*) highlighted in the bar chart

Figure S4. Cacao pod of each genotype (A), fermentation (B), and drying condition (C).

(A)



(B)



(C)

Figure S5. Bar chart of fermentation index (A), temperature during fermentation process (B), and cut test analysis (C).

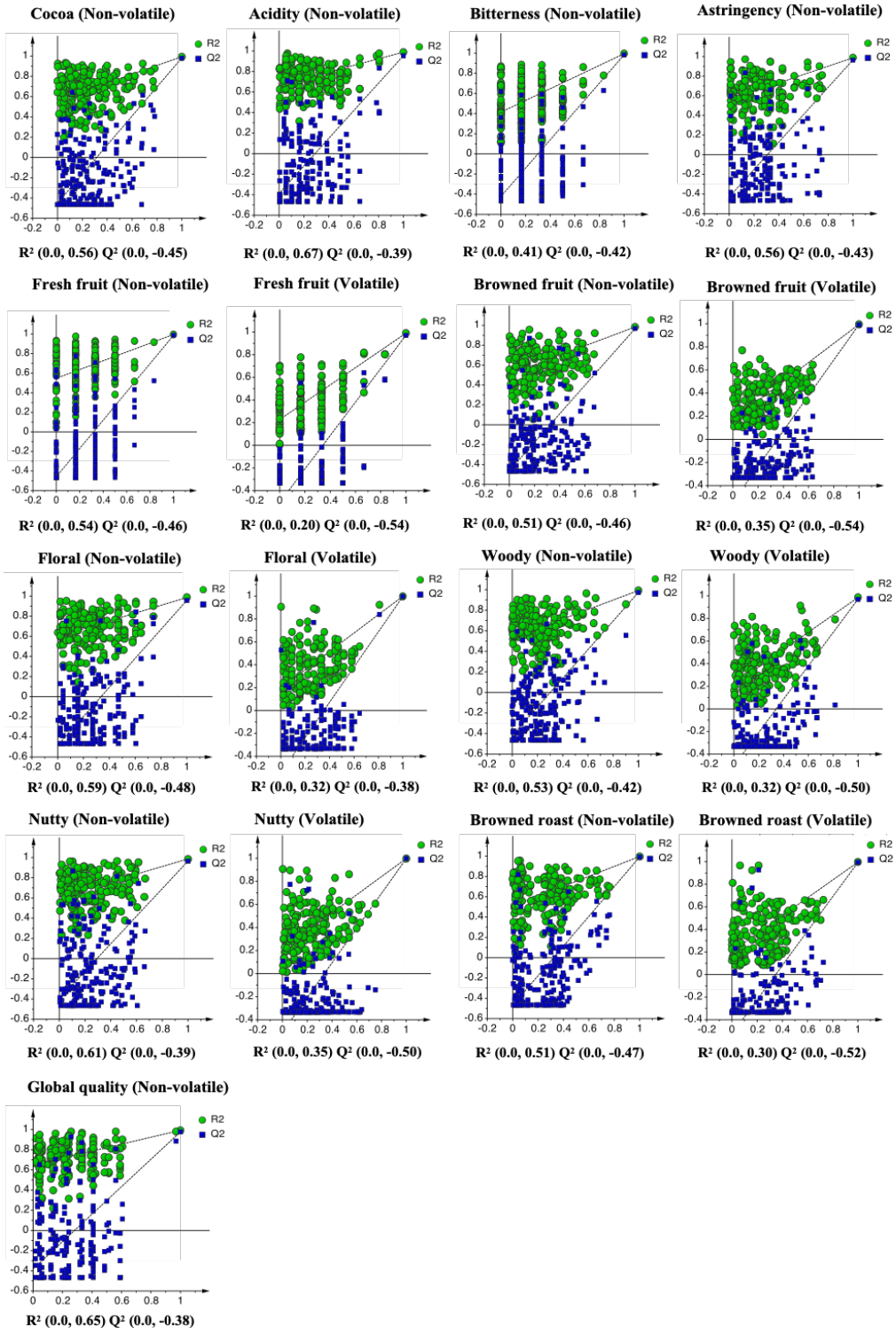


Figure S6. The result of validation by permutation test analysis (n = 200) of each sensory attribute

Table S9. Loading scores from PCA of non-volatile metabolites based on PC1 and PC2

No.	Metabolite name	PC1	Metabolite name	PC2
1	Citric acid	0.168575	Sucrose	0.191519
2	Phosphate	0.168399	Isoleucine	0.191107
3	Inositol	0.168199	Glycine	0.18977
4	Psicose	0.168169	Threonine	0.189528
5	4-Aminobutyric acid	0.166054	Lactic acid	0.186352
6	Sorbitol	0.165538	Methionine	0.174038
7	2-hydroxyglutaric acid	0.163646	Serine	0.172321
8	Malic acid	0.161204	Asparagine	0.163095
9	Ribulose	0.15017	Glucose	0.162609
10	Lyxose	0.149419	Proline	0.157569
11	Galactinol	0.14528	Caffeine	0.154841
12	1,2,3-Butanetriol	0.136614	Shikimic acid	0.152317
13	2-Dehydro gluconate	0.134174	Lysine	0.147825
14	Trehalose	0.133211	Glutamic acid	0.147743
15	Melibiose	0.133001	Aspartic acid	0.145415
16	Glucosamine	0.132336	Adenine	0.139841
17	Fructose	0.131605	Fumaric acid	0.136513
18	Threitol	0.127764	Xylitol	0.135274
19	Glyceric acid	0.126631	Gentiobiose	0.13504
20	Glycolic acid	0.124458	Glucosamine	0.132605
21	Gentiobiose	0.121511	Fructose	0.128594
22	Glutamic acid	0.115479	UDP-N-acetylglucosamine	0.12851
23	Mannose	0.112784	Mannose	0.126227
24	Glucono-1,5-lactone	0.112596	Tyrosine	0.118673
25	Succinic acid	0.109885	2-Dehydro gluconate	0.107226
26	Glycerol	0.108968	Galactinol	0.106898
27	Glucose	0.106133	Beta-Alanine	0.103239
28	Saccharic acid	0.105803	Palmitic acid	0.075388
29	Arabitol	0.104048	Stearic acid	0.070386
30	Aspartic acid	0.103154	Alanine	0.068851
31	Beta-Alanine	0.081336	4-Aminobutyric acid	0.052619
32	Asparagine	0.077276	Epicatechin	0.033851
33	Proline	0.073822	Inositol	0.026943
34	Threonic acid	0.0649	Malic acid	0.006623
35	Sucrose	0.051151	2,3 Buatanediol	-0.00181
36	Serine	0.04474	Phenylalanine	-0.00404
37	Alanine	0.039811	Phosphate	-0.02041

38	Adenine	0.025605	Sorbitol	-0.02082
39	Lactic acid	0.013894	Citric acid	-0.03231
40	Glycine	0.011432	Valine	-0.04034
41	UDP-N-acetylglucosamine	-0.00205	Psicose	-0.04145
42	Pyroglutamic acid	-0.00324	Meso erythritol	-0.04467
43	Threonine	-0.00351	2-Aminoethanol	-0.05079
44	Valine	-0.01275	Myo-Inositol	-0.05267
45	Shikimic acid	-0.0262	3-Phenyllactic acid	-0.05267
46	Tartaric acid	-0.03505	2-hydroxyglutaric acid	-0.05493
47	Isoleucine	-0.04864	Xylonic acid	-0.05891
48	Xylonic acid	-0.07489	Threitol	-0.06434
49	Leucine	-0.08121	Leucine	-0.06722
50	Caffeine	-0.08128	Catechine	-0.0677
51	Methionine	-0.08358	Theobromine	-0.07506
52	Fumaric acid	-0.1043	Erythritol	-0.08399
53	2,3 Buatanediol	-0.1055	Mannitol	-0.09451
54	Palmitic acid	-0.10591	Threonic acid	-0.09569
55	L-Alanine	-0.10704	Saccharic acid	-0.09624
56	Stearic acid	-0.11118	Ribulose	-0.09716
57	Xylitol	-0.1191	Lyxose	-0.1018
58	Lysine	-0.12174	Arabinopyranose	-0.10691
59	Theobromine	-0.12769	Glyceric acid	-0.10738
60	Tyrosine	-0.13873	Melibiose	-0.11737
61	Arabinopyranose	-0.14546	Trehalose	-0.1182
62	Phenylalanine	-0.15178	Glycerol	-0.1198
63	Mannitol	-0.1521	1,2,3-Butanetriol	-0.1208
64	Erythritol	-0.15267	Glycolic acid	-0.14113
65	Catechine	-0.16137	Succinic acid	-0.14612
66	Myo-Inositol	-0.16474	Glucono-1,5-lactone	-0.15701
67	Meso erythritol	-0.1655	Arabitol	-0.15874
68	3-Phenyllactic acid	-0.16665	L-Alanine	-0.16302
69	2-Aminoethanol	-0.16678	Pyroglutamic acid	-0.16859
70	Epicatechin	-0.16935	Tartaric acid	-0.20133

Table S10. Loading scores from PCA of volatile compounds based on PC1 and PC2

No.	Metabolite name	PC1	Metabolite name	PC2
1	Benzaldehyde	0.159444	2,6-Dimethylpyrazine	0.216446
2	Butyric acid	0.159356	2,5-dimethylpyrazine	0.215002
3	2,3-Butanediol	0.158885	3-Methoxybutyl Acetate	0.213148
4	Trans-Linalool oxide	0.158371	Gamma-Butyrolactone	0.212129
5	Pentanoic acid	0.155409	Ethyl Acetate	0.212124
6	2-methyl-3-buten-2-ol	0.154958	2-Methylpyrazine	0.203628
7	Butanoic acid	0.154058	Furfural	0.201488
8	1-Pentanol	0.148373	3-Methyl-1-butanol	0.199941
9	Phenylacetaldehyde	0.147243	2-Methyl-1-butanol	0.193756
10	3-Hexanone	0.147209	Ethanol	0.190166
11	Trans-2-Nonenal	0.146877	Ethyl Lactate	0.18
12	Butyraldehyde	0.146782	3-methylbutyl acetate	0.174676
13	2,4,7-Octanetrione	0.146457	Ethyl Hexanoate	0.174413
14	2,2,6-Trimethyloctane	0.145705	2-Methylbutyraldehyde	0.138915
15	Ethyl benzene	0.145333	Linalool	0.138839
16	isobutyl acetate	0.145296	2-Ethyl-3.5-dimethylpyrazine	0.13812
17	Isoamyl Ether	0.143921	3-Carene	0.129572
18	3-methylbutanal	0.143157	Acetol	0.128173
19	Valeraldehyde	0.142897	2-Pentanol	0.126328
20	1-Phenylethyl alcohol	0.137259	2-propanol	0.118388
21	2-Prphyl1-Pentanol	0.135693	cis-B-Ocimene	0.108232
22	2-propanol	0.131202	2,3,5-Trimethylpyrazine	0.106983
23	2-Phenylpropionaldehyde	0.130938	1-Butanol	0.105547
24	2-Pentanol	0.130307	2,3,5,6-Tetramethylpyrazine	0.103948
25	Acetol	0.129135	3-methylbutanal	0.095987
26	2,3-butanedione	0.125188	1-Phenylethyl alcohol	0.090383
27	2-Methylbutyraldehyde	0.121775	2-Methyl-1-propanol	0.089001
28	Linalool	0.118093	2-Butanone	0.088397
29	2-heptanone	0.116118	Isoamyl Ether	0.088128
30	2-Methyl-1-propanol	0.109547	isobutyl acetate	0.087767
31	Nonanal	0.108794	3-Acetoxy-2-butanone	0.084766
32	4-Methyl-valeric acid	0.106639	2,2,6-Trimethyloctane	0.0833
33	1-Butanol	0.095333	2-Heptanol	0.078883
34	3-methylbutyl acetate	0.085876	2-Nonanone	0.076891
35	Isobutyric Acid	0.081141	3-Hexanone	0.075108
36	2-Methyl-1-butanol	0.06728	2-Prphyl1-Pentanol	0.068765
37	3-Methyl-1-butanol	0.059235	Butanoic acid	0.060226

38	2-Methylpyrazine	0.047502	Phenylacetaldehyde	0.056679
39	Ethyl Hexanoate	0.04358	Pentanoic acid	0.052379
40	Gamma-Butyrolactone	0.03367	2-methyl-3-buten-2-ol	0.052363
41	2,5-dimethylpyrazine	0.005902	Methanol	0.050354
42	2,6-Dimethylpyrazine	-0.00551	2-Acetylfuran	0.049934
43	3-Methoxybutyl Acetate	-0.00719	Butyraldehyde	0.0435
44	1-Propanol	-0.01155	Ethyl Octanoate	0.039734
45	Ethyl Acetate	-0.01584	Pivalic Acid	0.038951
46	Ethanol	-0.02282	Acetic acid	0.032253
47	Ethyl Octanoate	-0.05413	2,3-Butanediol	0.026993
48	Furfural	-0.05481	Benzaldehyde	0.014803
49	2-Acetylfuran	-0.07272	Trans-Linalool oxide	0.013811
50	Ethyl Lactate	-0.07543	Butyric acid	-0.01342
51	Acetoin	-0.08921	Isoamyl Lactate	-0.01935
52	2-Ethyl-3.5-dimethylpyrazine	-0.12346	Benzonitrile	-0.02831
53	Benzonitrile	-0.12612	Propanoic acid	-0.02857
54	3-Carene	-0.1278	1-Pentanol	-0.06692
55	cis-B-Ocimene	-0.13938	Trans-2-Nonenal	-0.07033
56	2,3,5-Trimethylpyrazine	-0.13994	Ethyl benzene	-0.0775
57	2,3,5,6-Tetramethylpyrazine	-0.14068	Valeraldehyde	-0.07765
58	2-Butanone	-0.14225	2,4,7-Octanetrione	-0.08668
59	3-Acetoxy-2-butanone	-0.14621	2-heptanone	-0.09898
60	Methanol	-0.14687	4-Methyl-valeric acid	-0.11414
61	2-Heptanol	-0.14995	2,3-butanedione	-0.1156
62	2-Nonanone	-0.15016	2-Phenylpropionaldehyde	-0.12359
63	Isoamyl Lactate	-0.15259	Nonanal	-0.14812
64	Acetic acid	-0.15357	Acetoin	-0.16097
65	Pivalic Acid	-0.15739	Isobutyric Acid	-0.17823
66	Propanoic acid	-0.15852	1-Propanol	-0.21098

Table S11. List of the top VIP non-volatile metabolites highlighted in the correlation biplot (**Figure 9**)

Response variable: Cocoa			
No.	**Metabolite name**	**VIP score**	**Coefficient**
1	Psicose	1.442	0.135
2	Sorbitol	1.389	0.287
3	Ribulose	1.357	0.312
4	Inositol	1.349	0.235

5	Lyxose	1.340	0.382
Response variable: Acidity			
No.	**Metabolite name**	**VIP score**	**Coefficient**
1	Malic acid	1.325	0.366
2	Citric acid	1.200	0.191
3	2-hydroxyglutaric acid	1.124	0.204
4	Glutamic acid	1.300	0.207
5	Aspartic acid	1.138	0.444
Response variable: Astringency			
No.	**Metabolite name**	**VIP score**	**Coefficient**
1	Tartaric acid	1.709	0.345
2	Pyroglutamic acid	1.526	0.524
3	Lactic acid	1.487	0.735
4	Tartaric acid	1.709	0.345
Response variable: Bitterness			
No.	**Metabolite name**	**VIP score**	**Coefficient**
1	Catechine	1.330	0.308
2	Glycine	1.285	0.814
3	Theobromine	1.029	0.394
4	Epicatechin	1.028	0.654
Response variable: Fresh fruit			
No.	**Metabolite name**	**VIP score**	**Coefficient**
1	Psicose	1.431	0.153
2	Citric acid	1.429	0.174
3	Sorbitol	1.386	0.282
4	Inositol	1.372	0.215
5	Malic acid	1.359	0.245
Response variable: Browned fruit			
No.	**Metabolite name**	**VIP score**	**Coefficient**
1	Tartaric acid	1.616	0.452
2	Mannitol	1.291	0.458
3	Erythritol	1.251	0.502
4	Isoleucine	1.177	0.887
5	Caffeine	1.118	1.026
Response variable: Floral			
No.	**Metabolite name**	**VIP score**	**Coefficient**
1	Ribulose	1.497	0.237
2	Trehalose	1.479	0.512
3	Glycolic acid	1.477	0.450
4	Glucono-1,5-lactone	1.460	0.410

5	Lyxose	1.452	0.296
Response variable: Woody			
No.	**Metabolite name**	**VIP score**	**Coefficient**
1	Ribulose	1.451	0.286
2	Psicose	1.441	0.110
3	Lyxose	1.421	0.356
4	Trehalose	1.374	0.493
5	Glycolic acid	1.347	0.484
Response variable: Nutty			
No.	**Metabolite name**	**VIP score**	**Coefficient**
1	Methionine	1.501	0.360
2	Isoleucine	1.410	0.850
3	Glycine	1.278	0.784
4	Sucrose	1.262	0.487
5	Pyroglutamic acid	1.239	0.764
Response variable: Browned Roast			
No.	**Metabolite name**	**VIP score**	**Coefficient**
1	Inositol	1.416	0.191
2	Galactinol	1.389	0.219
3	Sorbitol	1.305	0.266
4	Glucose	1.175	0.298
5	Glutamic acid	1.174	0.267
Response variable: Global Quality			
No.	**Metabolite name**	**VIP score**	**Coefficient**
1	Galactinol	1.468	0.191
2	Gentiobiose	1.400	0.189
3	Inositol	1.383	0.204
4	Glucose	1.337	0.224
5	Malic acid	1.325	0.366

Table S12. List of the top VIP volatile metabolites highlighted in the correlation biplot (Figure 9)

Response variable: Fresh fruit			
No.	**Metabolite name**	**VIP score**	**Coefficient**
1	2,3-Butanediol	1.311	0.101
2	Benzaldehyde	1.304	0.118
3	2-methyl-3-buten-2-ol	1.290	0.181
4	3-methylbutanal	1.249	0.279

5	isobutyl acetate	1.230	0.277
6	3-Hexanone	1.226	0.316
7	2-Pentanol	1.166	0.353
Response variable: Browned fruit			
No.	**Metabolite name**	**VIP score**	**Coefficient**
1	Furfural	1.670	0.274
2	Ethyl Lactate	1.601	0.277
3	Ethyl Acetate	1.594	0.475
4	2-Heptanol	1.131	0.506
5	2-Nonanone	1.116	0.521
Response variable: Floral			
No.	**Metabolite name**	**VIP score**	**Coefficient**
1	Linalool	1.406	0.420
2	2-propanol	1.370	0.407
3	isobutyl acetate	1.360	0.243
4	1-Phenylethyl alcohol	1.287	0.407
Response variable: Woody			
No.	**Metabolite name**	**VIP score**	**Coefficient**
1	Ribulose	1.450	0.286
2	Trehalose	1.374	0.493
3	Glucono-1,5-lactone	1.296	0.480
Response variable: Nutty			
No.	**Metabolite name**	**VIP score**	**Coefficient**
1	Acetoin	1.554	0.366
2	2,5-dimethylpyrazine	1.549	0.409
	2-Methylbutyraldehyde	1.491	0.204
Response variable: Browned Roast			
No.	**Metabolite name**	**VIP score**	**Coefficient**
1	2-Ethyl-3.5-dimethylpyrazine	1.174	0.232
2	2,3-butanedione	1.160	0.249
Response variable: Cocoa			
No.	**Metabolite name**	**VIP score**	**Coefficient**
1	3-Methylbutanal	1.283	0.269
2	Phenylacetaldehyde	1.250	0.188

Table S13. All volatile metabolites detected in this study and the aroma descriptor

Metabolites	Group	H2	T3	F2	Aroma description
4-Methyl-valeric acid	Acid	27474.22	30701.69	20176.28	

Acetic acid	Acid	2836865.00	3066434.33	3896622.33	Sour, vinegar
Butyric acid	Acid	891739.40	773195.07	514922.60	
Isobutyric Acid	Acid	750288.97	926363.63	638143.13	
Pivalic Acid	Acid	26492.54	45377.54	109705.13	
Propanoic acid	Acid	141895.70	4341779.33	8458767.00	Acidic, pungent, cheese
Pentanoic acid	Acid	174487.27	113795.07	72065.59	
Butanoic acid	Acid	455453.77	278576.10	174276.90	Acetic, cheese, butter
1-Butanol	Alcohol	2109.79	835.11	853.20	
1-Pentanol	Alcohol	6512.45	6278.89	3217.36	Fruity, green
2-Heptanol	Alcohol	2465.40	2864.66	20801.27	Floral, earthy, fruity, citrus
2-Methyl-1-propanol	Alcohol	3836.12	3080.08	2985.43	Wine
2-Methyl-3-buten-2-ol	Alcohol	7695.92	4928.49	3015.27	Herbal, earth, oily
2-Pentanol	Alcohol	28600.52	7434.69	6596.78	Green, fruity, sweet, fusel oil
2-propanol	Alcohol	13371.46	4319.32	3562.14	
2-Prphyl-1-Pentanol	Alcohol	2430.73	1888.68	1649.83	
3-Methyl-1-butanol	Alcohol	9155.58	3171.07	6206.91	Fruity
2-Methyl-1-butanol	Alcohol	3332.54	1306.41	2234.99	Fruity
Ethanol	Alcohol	13605.12	10336.21	13835.87	Undesirable note
Methanol	Alcohol	3795.04	4318.01	7490.61	
1-Phenylethyl alcohol	Alcohol	9136.40	6805.04	6147.89	Honey, flowery
2,3-Butanediol	Alcohol	11073.48	8559.69	5934.05	Fruity, creamy
1-Propanol	Alcohol	12076.86	33940.21	16917.02	Pungent, sweet, candy
2-Butanone	Aldehyde	6002.41	5859.16	10514.38	
2-Methylbutyraldehyde	Aldehyde	40968.83	20463.19	21502.55	
2-Phenylpropionaldehyde	Aldehyde	12596.72	15768.97	2226.83	
3-methylbutanal	Aldehyde	99225.65	55473.40	44285.47	Malty, chocolate
Benzaldehyde	Aldehyde	49795.71	38869.34	24813.59	Fruity, almond
Butyraldehyde	Aldehyde	48286.21	30808.61	17499.33	
Nonanal	Aldehyde	10360.02	12866.85	6574.69	Citrus
Phenylacetaldehyde	Aldehyde	52218.04	39713.27	32272.91	Honey, nutty
Trans-2-Nonenal	Aldehyde	7168.45	7078.00	3850.42	
Valeraldehyde	Aldehyde	8481.40	8473.82	4535.76	
3-Hexanone	Aldehyde	31031.93	17180.09	11204.75	Wine, fruity
3-Methoxybutyl Acetate	Ester	4793.40	2397.68	4585.74	
3-methylbutyl acetate	Ester	3342.83	1818.25	2322.23	Fruity
Ethyl Acetate	Ester	15917.04	6565.02	15482.69	Fruity, sweet
Ethyl Lactate	Ester	5587.65	3855.73	6597.11	Fruity
Ethyl Octanoate	Ester	2574.55	2539.28	2849.37	Fruity, apricot

Isoamyl Ether	Ester	7280.34	3929.83	2868.78	
Isoamyl Lactate	Ester	2777.68	7662.88	13103.81	
isobutyl acetate	Ester	9113.04	5215.37	3925.92	Fruity, floral
Ethyl Hexanoate	Ester	1264.17	914.35	1112.10	Fruity, green, floral
Ethyl acetate	Ester	185075.73	615666.20	1474126.33	Fruity, sweet
2-Acetylfuran	Furans	206.05	120.73	1397.09	
Furfural	Furans	34553.79	10444.74	41343.97	Fruity, flowery, roasted, almond
2-heptanone	Ketone	9020.01	9695.58	6084.89	Fruity
2-Nonanone	Ketone	2796.13	3742.86	30285.89	Fruity, sweet
2,3-butanedione	Ketone	44177.18	47108.68	32938.49	Buttery, creamy
3-Acetoxy-2-butanone	Ketone	17121.53	17113.53	29369.83	
Gamma-Butyrolactone	Lactone	45358.14	13835.52	34595.18	Caramel, sweet, creamy
Linalool	Lactones	35404.76	13832.96	15490.94	Rose, floral, green
Benzonitrile	Nitriles	484.94	571.60	655.00	
3-Carene	Others	1513.08	655.24	3873.36	
Acetoin	Others	483727.53	632059.07	590896.07	Buttery, creamy
Acetol	Others	48631.63	26657.56	26014.91	
cis-B-Ocimene	Others	1703.02	883.12	6717.77	Floral, herbal
Ethyl benzene	Others	9008.91	8864.91	4533.66	Fruity
Trans-Linalool oxide	Others	9370.89	6470.68	2854.29	Floral, honey, woody, lemon peel
2,4,7-Octanetrione	Others	23198.68	23374.28	10463.43	
2-Ethyl-3.5-dimethylpyrazine	Pyrazine	11553.88	5301.21	25750.18	Caramel, roasted, nutty
2-Methylpyrazine	Pyrazine	6843.38	2542.18	5000.96	Nutty, caramel, roasted
2,3,5-Trimethylpyrazine	Pyrazine	53601.82	34586.69	178249.40	Nutty, cocoa
2,3,5,6-Tetramethylpyrazine	Pyrazine	147438.80	99182.86	504097.50	Nutty, cocoa
2,5-dimethylpyrazine	Pyrazine	7476.07	3166.78	6682.65	Caramel, roasted, nutty
2,6-Dimethylpyrazine	Pyrazine	12923.46	5886.77	12238.68	Caramel, roasted, nutty

Note: The flavor descriptor was based on the previous studies (Yang et al., 2024; Velásquez-Reyes et al., 2023; Akoa et al., 2023; Colonges et al., 2022; Bastos et al., 2019)

List of Annotated Metabolites Obtained from GC/MS analysis

No.	Metabolite name	RT[a] (min)	RI[b]	Quant mass (m/z)	Similarity (%)	Library
1	2,3 Buatanediol	4.189	1047.43	117.0782	92	NIST 20s
2	Lactic acid	4.438	1066.08	147.0718	99.9	In-house
3	Glycolic acid	4.629	1080.47	147.0697	96.4	In-house

4	L-Alanine	5.02	1109.24	147.05	90	NIST 20s
5	Alanine	5.025	1109.56	116.0974	96.4	In-house
6	Isoleucine	6.027	1180.29	86.0423	83.6	In-house
7	Valine	6.661	1225.34	144.1231	98	In-house
8	Serine	7.217	1265.06	116.0766	93.3	In-house
9	2-Aminoethanol	7.376	1276.4	174.0949	98	In-house
10	Leucine	7.458	1282.24	158.109	93.4	In-house
11	Phosphate	7.498	1285.13	299.0936	99.1	In-house
12	Glycerol	7.538	1287.95	147.0782	99.2	In-house
13	Threonine	7.749	1303.15	117.0919	86.7	In-house
14	Proline	7.778	1305.32	142.1192	93.7	In-house
15	1,2,3-Butanetriol	7.807	1307.55	117.0727	90	NIST 20s
16	Glycine	7.928	1316.6	174.0987	99.3	In-house
17	Succinic acid	7.973	1319.9	147.0697	96.7	In-house
18	Glyceric acid	8.315	1345.51	147.075	94.2	In-house
19	Beta-Alanine	9.58	1442.16	174.0965	81.2	In-house
20	Malic acid	10.368	1504.4	147.0705	97.8	In-house
21	Threitol	10.61	1524.53	147.0667	90.7	In-house
22	Methionine	10.679	1530.26	176.0689	84.5	In-house
23	Pyroglutamic acid	10.701	1532.09	156.0766	95.4	In-house
24	Meso erythritol	10.706	1532.5	147.0714	97.1	In-house
25	Aspartic acid	10.753	1536.43	232.0872	98	In-house
26	4-Aminobutyric acid	10.821	1542.06	174.0962	99.5	In-house
27	Threonic acid	11.351	1586.01	147.072	88.6	In-house
28	2-hydroxyglutaric acid	11.402	1590.27	129.096	83.5	In-house
29	3-Phenyllactic acid	11.479	1596.69	193.0736	93	In-house
30	Glutamic acid	11.921	1635.25	246.1103	98.2	In-house
31	Phenylalanine	11.968	1639.34	218.0667	99	In-house
32	Tartaric acid	12.265	1665.33	147.0676	97.9	In-house
33	Asparagine	12.514	1687.13	116.0875	95.9	In-house
34	Lyxose	12.591	1693.86	103.0125	87.1	In-house
35	Ribulose	12.761	1709.25	147.0673	92.1	In-house
36	Xylitol	13.113	1741.68	217.066	87.1	In-house
37	Arabitol	13.248	1754.12	217.0818	99.1	In-house
38	UDP-N-acetylglucosamine	13.698	1795.53	129.1036	80.5	In-house
39	Xylonic acid	13.753	1800.61	147.0659	89.9	In-house
40	Shikimic acid	14.084	1832.64	204.0603	81.8	In-house
41	Citric acid	14.228	1846.59	273.1042	97.8	In-house

42	Caffeine	14.272	1850.8	194.0395	93.4	In-house
43	Theobromine	14.454	1868.4	180.041	98	NIST 20s
44	Adenine	14.559	1878.53	264.1051	87.5	In-house
45	Psicose	14.737	1895.76	103.0344	87.3	In-house
46	Fructose	14.898	1911.77	103.0375	99.6	In-house
47	Mannose	15.041	1926.21	160.0739	90.2	In-house
48	2-Dehydro gluconate	15.101	1932.37	103.025	96.1	In-house
49	Glucose	15.16	1938.31	205.0759	99.1	In-house
50	Lysine	15.196	1941.99	174.0968	96.7	In-house
51	Glucosamine	15.296	1952.03	203.0682	87.7	In-house
52	Tyrosine	15.358	1958.34	218.0692	95.7	In-house
53	Fumaric acid	15.386	1961.11	144.1141	95	NIST 20s
54	Mannitol	15.49	1971.67	205.0757	99.3	In-house
55	Sorbitol	15.565	1979.23	205.0705	93.3	In-house
56	Erythritol	15.879	2011.48	205.0841	89	NIST 20s
57	Glucono-1,5-lactone	16.199	2045.43	147.0734	98.5	In-house
58	Palmitic acid	16.231	2048.75	313.275	94.3	In-house
59	Saccharic acid	16.337	2060.02	147.0774	86.7	In-house
60	Myo-Inositol	16.429	2069.74	217.0797	93	NIST 20s
61	Inositol	17.002	2131.71	217.0782	99.1	In-house
62	Stearic acid	18.025	2246.54	117.0349	90.1	In-house
63	Sucrose	21.729	2710.71	361.1819	98.2	In-house
64	Trehalose	22.493	2817.11	361.1758	96.2	In-house
65	Epicatechin	23.071	2900.11	368.1757	98.2	In-house
66	Catechine	23.234	2924.12	368.1676	85	NIST 20s
67	Gentiobiose	23.27	2929.31	204.05	87.2	In-house
68	Melibiose	23.48	2960.21	204.066	74.4	In-house
69	Galactinol	24.22	3072.19	204.0576	82.6	In-house
70	Arabinopyranose	24.936	3184.11	204.0621	95	NIST 20s

Remarks: [a] (RT) Retention time in minutes, [b] (RI) Retention Indices, calculated by standard alkane mixture (C10-C40).

List of Annotated Metabolites Obtained from HS-SPME Arrow GC/MS analysis

No.	Metabolite name	RT[a] (min)	RI[b]	Quant mass (m/z)	Similarity (%)	Library
1	Butyraldehyde	4.623	176.31	43.08	84.5	In-house
2	Ethyl Acetate	6.087	220.56	43.04445	83.6	In-house

3	Methanol	6.323	229.41	31.05	80.5	In-house
4	2-Butanone	6.434	233.54	43.03636	80.5	In-house
5	2-Methylbutyraldehyde	6.732	244.7	41.06429	86.9	In-house
6	3-methylbutanal	6.861	249.54	44.05	86.2	In-house
7	2-propanol	6.926	251.96	45.05	79.3	In-house
8	Ethanol	7.307	266.21	45.05	85.4	In-house
9	Valeraldehyde	8.841	323.57	44.05	81.9	In-house
10	2,3-Butanedione	8.879	325	43.05	86.1	In-house
11	2,4,7-Octanetrione	9.028	330.58	43.065	84	NIST 20s
12	isobutyl acetate	9.978	366.1	43.05	82.6	In-house
13	2-methyl-3-buten-2-ol	10.905	400.45	43.05	91.9	In-house
14	3-Hexanone	11.157	405.82	71.1	89	NIST 20s
15	Isoamyl Ether	11.963	423.03	71.1	81.9	In-house
16	2-Methyl-1-propanol	13.033	445.86	43.06667	94.3	In-house
17	2-Pentanol	14.284	472.55	45.05455	98.4	In-house
18	3-methylbutyl acetate	14.597	479.24	43.05	98.5	In-house
19	1-Butanol	15.459	497.62	56.1	95.2	In-house
20	Ethyl benzene	15.58	500.2	91.07917	88.2	In-house
21	3-Carene	16.386	517.39	93.1	81.6	In-house
22	2-heptanone	17.434	539.77	43.05	92.1	In-house
23	DL-2-Methyl-1-butanol	18.323	558.74	57.1	90.9	In-house
24	3-Methyl-1-butanol	18.352	559.34	55.0625	95	In-house
25	cis-B-Ocimene	19.764	589.49	93.09167	90	In-house
26	Ethyl Hexanoate	19.893	592.24	98.06667	82	NIST 20s
27	1-Pentanol	20.392	602.78	42.05833	92	In-house
28	2-Methylpyrazine	21.555	626.8	94.075	92.1	In-house
29	Acetoin	22.414	644.55	45.05	92.4	In-house
30	Acetol	23.3	662.85	43.04667	94.9	In-house
31	2-Heptanol	23.551	668.02	45.05	87.9	In-house
32	2,5-dimethylpyrazine	24.393	685.42	108.1	97.2	In-house
33	3-Methoxybutyl Acetate	25.052	699.02	59.03333	76.7	In-house
34	2,6-Dimethylpyrazine	25.274	703.61	108.1	86.6	In-house
35	Ethyl Lactate	25.78	714.05	45.05	84.5	In-house
36	3-Acetoxy-2-butanone	26.94	738	43.04615	82	In-house
37	2-Nonanone	27.035	739.97	58.05	86.3	In-house
38	Nonanal	27.272	744.86	57.06818	85	In-house
39	2,3,5-Trimethylpyrazine	27.845	756.69	42.05	89.7	In-house
40	Ethyl Octanoate	28.947	779.46	88.05833	82.6	In-house
41	Acetic acid	29.954	800.26	43.05	78.5	In-house

42	2-Ethyl-3.5-dimethylpyrazine	30.36	809.18	135.1	96.6	In-house
43	Trans-Linalool oxide	30.623	814.98	59.06	81.3	In-house
44	Furfural	30.926	821.63	95.05	79.1	In-house
45	2,3,5,6-Tetramethylpyrazine	30.939	821.93	54.05	85.2	In-house
46	2-Prophyl-1-Pentanol	31.297	829.79	57.1	87.5	In-house
47	2-Acetylfuran	32.778	862.37	95.01667	72.1	In-house
48	Benzaldehyde	33.531	878.92	106.05	83.8	In-house
49	2,3-Butanediol	33.537	879.05	50.05	98	NIST 20s
50	Propanoic acid	33.618	880.83	45.075	91	NIST 20s
51	Linalool	33.783	884.47	71.0625	74.4	In-house
52	Trans-2-Nonenal	34.208	893.81	56.07273	81.1	In-house
53	Ethyl acetate	34.604	902.53	43.05	85	NIST 20s
54	Isobutyric Acid	34.836	907.62	43.1	73.9	In-house
55	Pivalic Acid	35.15	914.52	57.1	70.4	In-house
56	1-Propanol	35.719	927.04	43.05	81	NIST 20s
57	Isoamyl Lactate	36.354	940.99	43.07222	81.5	In-house
58	Benzonitrile	36.825	951.36	103.05	74	In-house
59	Gamma-Butyrolactone	38.131	980.07	42.06818	81.1	In-house
60	Phenylacetaldehyde	38.474	987.62	91.0625	71.9	In-house
61	2-Phenylpropionaldehyde	38.792	994.6	105.05	72.5	In-house
62	Butyric acid	38.939	997.84	60.02	89.5	In-house
63	Pentanoic acid	38.962	998.36	29.05	88	NIST 20s
64	Butanoic acid	38.967	998.46	74.05	90	NIST 20s
65	4-Methyl-valeric acid	45.6	1159.74	60.025	84.3	In-house

Remarks: [a] (RT) Retention time in minutes, [b] (RI) Retention Indices, calculated by standard FAEE (Fatty Acid Ethyl Esters, C4–C24)