**ASIAN J NAT PROD BIOCHEM**
Volume 24, Number 1, June 2026
Pages: 8

P-ISSN: 2775-4189
E-ISSN: 2775-4197
https://doi.org/10.13057/biofar/f240105

# Time-resolved volatile profiling and endpoint microbiome characterization in spontaneous *Coffea arabica* wet fermentation

**BASTIAN NOVA[1,♥], LINDA WATI[1], RIDWAN RIZKYANTO[1], EPI SUPRI WARDI[2], PADRIANSYAH LUBIS[1]**
[1]Department of Food and Agricultural Product Technology, Faculty of Agricultural Technology, Universitas Andalas. Jl. Limau Manis, Kampus Unand Limau Manis, Padang 25163, West Sumatra, Indonesia. Tel.: +62-751-72701, Fax.: +62-751-72702, ♥email: bastiannova@ae.unand.ac.id
[2]Faculty of Pharmacy, Universitas Perintis Indonesia. Jl. Adinegoro Simp. Kalumpang Lubuk Buaya, Padang 25586, West Sumatra, Indonesia

Manuscript received: 13 January 2026. Revision accepted: 26 May 2026.

**Abstract.** *Nova B, Wati L, Rizkyanto R, Wardi ES, Lubis P. 2026. Time-resolved volatile profiling and endpoint microbiome characterization in spontaneous* Coffea arabica *wet fermentation. Asian J Nat Prod Biochem 24 (1): f240105. https://doi.org/10.13057/biofar/f240105.* Spontaneous fermentation during wet coffee processing involves complex microbial communities that influence Volatile Organic Compound (VOC) profiles and aroma characteristics. However, time-resolved descriptions of VOC dynamics alongside endpoint microbial community composition in spontaneous *Coffea arabica* wet fermentation remain limited. This study combined GC-MS volatilomics with amplicon-based community profiling to characterize chemical and microbial changes during a single 36-hour spontaneous fermentation batch. VOC profiles were measured by GC-MS at 0, 12, 24, and 36 h, while bacterial 16S rRNA and fungal ITS amplicon sequencing were performed only at 36 h to characterize the endpoint community. After contaminant curation, 32 unique volatile compounds were retained across timepoints. Temporal profiles indicated depletion of plant-derived compounds and appearance of fermentation-associated products. Eicosanoic acid decreased from 9.98% to 0.27% (Area%), and caryophyllene decreased from 6.24% to 1.07%. Linalool was detected only at 36 h (0.34%), whereas pentadecanal appeared at 12 h and persisted through 36 h. The 36-h bacterial community was dominated by *Leuconostoc* spp. (88.54% of reads; *L. falkenbergense* 47.3%, *L. pseudomesenteroides* 29.02%), while the fungal community was dominated by *Hanseniaspora uvarum* (73.0% of ITS reads). Microbiome characterization was performed only at the 36-h fermentation endpoint; no time-series microbiome sampling and no statistical association testing between microbial taxa and VOC data were conducted. Accordingly, all proposed microbe–metabolite relationships are derived solely from published literature and must be treated as hypothesis-generating rather than analytically validated findings. This work provides baseline data for future, time-series-replicated microbiome studies and controlled fermentation trials in coffee processing.

**Keywords:** 16S rRNA, amplicon sequencing, spontaneous fermentation, wet coffee processing, volatilomics

**Abbreviations:** Area%: Semi-quantitative peak area percentage, GC-MS: Gas Chromatography-Mass Spectrometry, ITS: Internal Transcribed Spacer, LAB: Lactic Acid Bacteria, ND: Not detected, VOCs: Volatile Organic Compounds

## INTRODUCTION

Coffee (*Coffea arabica* L.) is one of the world's most economically important agricultural commodities, cultivated in over 70 tropical and subtropical countries and serving the economic foundation for tens of millions of smallholder farming households (Zhang et al. 2019; Pereira et al. 2025). Global demand for quality differentiation has increased interest in postharvest processing, where fermentation is increasingly recognized as a critical determinant of aroma precursor composition and final beverage quality (Zhang et al. 2019). In wet processing, depulped beans retain a mucilage layer that is removed by microbial and enzymatic activity under aqueous conditions (Elhalis et al. 2023b). This fermentation can alter the pool of aroma precursors and Volatile Organic Compounds (VOCs) that later contribute to beverage aroma after roasting.

Spontaneous coffee fermentation is driven by indigenous microorganisms associated with the fruit, water, and processing surfaces. Because the process is not inoculated, microbial succession can vary across sites and batches (Schwan et al. 2023). The dominant microbial groups during wet coffee fermentation typically include Lactic Acid Bacteria (LAB) and non-*Saccharomyces* yeasts such as *Hanseniaspora* species (Elhalis et al. 2020; Pothakos et al. 2020). LAB contribute to pH reduction and may produce organic acids, while yeasts are considered essential for mucilage degradation and the generation of aroma-active compounds through enzymatic activity and metabolic conversion (Elhalis et al. 2023a).

The volatile organic compounds generated during coffee fermentation constitute a chemically diverse set of natural products whose biosynthetic origins are linked to specific microbial enzymatic activities and metabolic pathways. Monoterpene alcohols such as linalool and geraniol are produced principally by β-glucosidase-mediated hydrolysis of glycosidically bound aroma precursors present in coffee mucilage, releasing free terpene aglycones that contribute floral and citrus character to the beverage (Zhang et al. 2021; Fan et al. 2022). Aliphatic aldehydes can be generated through microbial lipid oxidation via β oxidation or α-oxidation pathways, or through amino acid catabolism under fermentative conditions via α-keto acid decarboxylation reactions (Cassimiro et al. 2023). Esters, which impart fruity aroma attributes, are produced through esterification reactions catalyzed by esterases and alcohol acyltransferases expressed by non-Saccharomyces yeasts and certain LAB (Gao et al. 2022a). Higher alcohols can be further produced from amino acid degradation through the Ehrlich pathway, catalyzed by specific decarboxylases and alcohol dehydrogenases in fermentative yeasts. Understanding these biochemical pathways is essential for interpreting how microbial community composition shapes the natural product VOC profile of fermented coffee, and for developing targeted strategies to modulate aroma quality in processed beans.

Recent studies have examined microbial community composition in coffee fermentation using amplicon sequencing (Holguín-Sterling et al. 2023; Góngora et al. 2024; Todhanakasem et al. 2024) and have profiled VOC changes using GC-MS and related techniques (Galarza and Figueroa 2022; Silva et al. 2025). Integrated multi-omics approaches have also begun to link fermentation microbiota with metabolite and flavor outcomes in Arabica wet processing, including studies from Ecuador (Zhang et al. 2019) and Africa (Pereira et al. 2025). However, most studies either focus on microbial composition without detailed temporal VOC data or measure VOCs without concurrent microbiome characterization. The specific gap in the existing literature is the lack of integrated datasets that combine time-resolved VOC profiling with endpoint microbial community characterization from within the same spontaneous fermentation batch, particularly for wet *C. arabica* fermentation in smallholder processing environments. Such baseline data are necessary to identify biochemically plausible microbe-metabolite associations and to formulate testable hypotheses for future mechanistic studies.

This study addressed three Research Questions: (RQ1) How do Volatile Organic Compound (VOC) profiles evolve temporally across 0, 12, 24, and 36 h of spontaneous wet fermentation of West Sumatran, Indonesia *C. arabica*? (RQ2) What bacterial and fungal taxa dominate the fermentation community at the 36-h endpoint under the conditions tested? (RQ3) Which literature-supported microbe metabolite associations are plausible given the observed community composition and temporal VOC dynamics, and can therefore serve as testable hypotheses for future time-series experiments? We hypothesize that (H1) plant-derived substrate compounds will decrease while fermentation-associated metabolites will accumulate over the 36-h period; and (H2) the endpoint community will be dominated by lactic acid bacteria and non-Saccharomyces yeasts previously documented in spontaneous coffee fermentation, and that these taxa will be associated with specific VOC classes via literature-informed pathways. These hypotheses are evaluated in a descriptive, single-batch framework intended to guide future replicated, mechanistic studies.

## MATERIALS AND METHODS

### Study site and raw material

Ripe *C. arabica* cherries (Sigarar Utang, Kartika, Andung Sari) were obtained from a smallholder farm in West Sumatra, Indonesia, at an altitude of approximately 1,500-1,600 m a.s.l.; mean ambient temperature 16-18°C; humid tropical climate, during the 2025 harvest season. Cherries were hand-picked at commercial maturity (predominantly red stage) and transported to the processing facility within 4 hours of harvest. The processing facility was a covered open-air structure typical of smallholder operations in the region, with no temperature or humidity control.

### Fermentation set up

Approximately 2 kg of depulped coffee beans with mucilage were placed in a plastic fermentation vessel (5 L capacity) and soaked in 1.5 L of tap water obtained from the local municipal supply. The vessel was loosely covered to limit oxygen exposure while allowing gas exchange. Fermentation was conducted at ambient temperature (16-18°C) without agitation. Liquid fermentation samples (approximately 50 mL) were collected at 0, 12, 24, and 36 h for VOC analysis. Liquid fermentation was chosen for sampling as it represents the aqueous phase where microbial metabolic products accumulate and can be analyzed with minimal matrix interference. At 36 h, an additional sample of liquid fermentation was collected for microbial community analysis.

### GC-MS headspace volatile profiling

Static headspace GC-MS analyzed volatile organic compounds. For headspace sampling, liquid was transferred to a headspace vial sealed with a polytetrafluoroethylene (PTFE)-lined silicone cap. Samples were equilibrated at 60.0°C for 3.0 min before automated headspace injection. The GC-2030 was coupled to a GC-MS-TQ8050 NX operated in full-scan mode (m/z 45-500) with a solvent cut of 2.50 min. The injector was set to 250°C and operated in splitless mode with a 1.00 min sampling time. The oven temperature program was 60°C (2.0 min), ramp 5°C/min to 100°C (0 min), ramp 10°C/min to 200°C (5.0 min), and ramp 15°C/min to 250°C (2.0 min), followed by a 3.0 min equilibration. Carrier gas flow was pressure-controlled at 131.1 kPa with a column flow of 2.14 mL/min, a linear velocity of 53.4 cm/s, and a purge flow of 3.0 mL/min. MS data were acquired from 3.00 to 30.33 min with an event time of 0.300 s and a scan speed of 1666. Compounds were identified by comparison with the NIST mass spectral library, and only hits with a similarity index ≥85% were retained for reporting.

### Microbial community profiling

DNA was extracted from 2 mL of 36-h liquid fermentation using a Quick-DNA Magbead Plus Kit (D4082; Zymo Research, Irvine, CA, USA) according to the manufacturer's instructions. Bacterial communities were characterized by amplification of the full-length 16S rRNA gene using universal primers 27F and 1492R. Fungal communities were characterized by amplification of the ITS region using primers ITS1 and ITS4. PCR products were purified and sequenced on an Oxford Nanopore GridION platform.

Library preparation was performed using the Nanopore Ligation sequencing amplicons Native Barcoding Kit 96 V14 (SQK-NBD114.96), and the final library was sequenced on the Nanopore GridIon platform. Nanopore sequencing was operated by MinKNOW software version 25.03.7. Basecalling was performed using Dorado version 7.8.3 with a high-accuracy model. For bioinformatic analysis, adapters and primers were trimmed using Dorado version 7.8.3. Reads were filtered for quality ($Q \geq 10$) and length (500-1600 bp for 16S; 150-600 bp for ITS) using NanoFilt, and quality statistics were visualized using NanoPlot (De Coster et al. 2018). Filtered reads were taxonomically classified using the Centrifuge classifier (Kim et al. 2016), with bacterial and fungal reference indices built from the NCBI 16S and ITS TargetedLoci databases. Downstream composition analysis and visualizations were performed using Pavian (Breitwieser and Salzberg 2020), Krona Tools (Ondov et al. 2011), and RStudio using R version 4.3. Diversity indices (Shannon and Simpson) and richness estimators (observed richness, Chao1) were calculated using the vegan package (Oksanen et al. 2022) in R.

### Data analysis and interpretation

VOC temporal dynamics were described based on changes in Area% across the four timepoints. A literature-informed hypothesis framework (presented as Table 4 in the Discussion section) links observed microbial taxa to putative metabolic capabilities reported in prior studies. This table does not represent functional inference from the amplicon data collected in this study. It maps literature-reported enzymatic activities and pathways to the observed community composition to propose testable hypotheses for future work. No KEGG module assignment, enzyme assays, or correlation analyses were performed.

## RESULTS AND DISCUSSION

### Temporal VOC dynamics

GC-MS analysis identified 32 unique volatile compounds across the four sampling time points after contaminant exclusion. A summary of key compounds exhibiting notable temporal changes (Table 1)

Temporal profiling indicated depletion of several plant-derived compounds and the appearance of fermentation-associated metabolites. Eicosanoic acid (arachidic acid) decreased from 9.98% at 0 h to 0.27% at 36 h. Caryophyllene declined from 6.24% at 0 h to 1.07% at 36 h. Conversely, linalool was not detected at 0, 12, or 24 h but was detected at 36 h (0.34%). Pentadecanal was not detected at 0 h, appeared at 12 h (0.78%), and remained detectable at 24 h (0.70%) and 36 h (0.39%).

### Bacterial community composition at 36 h

Full-length 16S rRNA amplicon sequencing of the 36 h fermentation sample yielded 139,075 quality-filtered reads and identified the top 10 bacterial species (Figure 1) in the reported species-level profile (Table 2). The community was dominated by *Leuconostoc falkenbergense* (47.3%; 70,718 reads), followed by *Leuconostoc pseudomesenteroides* (29.02%; 43,386 reads) and *Leuconostoc mesenteroides* (12.22%; 18,267 reads). *Leuconostoc* accounted for 88.54% of reads at the genus level, while *Weissella* (2.00%) and *Klebsiella* (1.41%) were present at lower relative abundances; other genera collectively accounted for 7.96%. Diversity indices indicated moderate evenness with dominance by a small number of taxa (Shannon index 1.69; Simpson index 0.69). Richness estimators suggested additional low-abundance taxa beyond the dominant species-level profile (observed richness 417 taxa; Chao1 976.62); however, these estimates encompass all Centrifuge-classified reads, including low-confidence taxonomic assignments, and are likely to overestimate true community richness due to classification artifacts inherent to reference-based amplicon classification. These values should therefore be interpreted as upper-bound estimates of community diversity rather than precise ecological richness measurements. Minor LAB (*Weissella cibaria* 1.62%; *Weissella confusa* 0.38%) and low-abundance Enterobacteriaceae (*Klebsiella pasteurii* 1.13%; *Enterobacter asburiae* 0.14%) were detected in the dataset. These results indicate that LAB, particularly Leuconostoc, dominated the bacterial community at the fermentation endpoint under the conditions tested. Given the known limitations of species-level classification from Nanopore amplicon data classified by Centrifuge (see Materials and Methods), genus-level assignments are considered more reliable, and conclusions regarding community composition are drawn primarily at the genus level throughout this study.

### Fungal community composition (36-hour timepoint)

ITS amplicon sequencing of the 36 h fermentation sample yielded 130,450 quality-filtered reads and identified the top 10 fungal species (Figure 2) in the reported species-level profile (Table 3). The community was dominated by *Hanseniaspora uvarum* (73.0%; 101,459 reads), followed by *Hanseniaspora pseudoguilliermondii* (6.7%; 9,318 reads) and *Saturnispora gosingensis* (4.89%; 6,794 reads); *Starmerella davenportii* accounted for 4.86% (6,748 reads). At the genus level, *Hanseniaspora* comprised 80.05% of reads, while *Starmerella* accounted for 6.80% and *Kurtzmaniella* for 0.84%; remaining genera were each <1%. Diversity indices indicated low evenness (Shannon index 1.38; Simpson index 0.46), consistent with dominance by a few taxa. Richness estimators indicated the presence of additional low-abundance taxa (observed richness 861 taxa; Chao1 1779.86). As with the bacterial dataset, these estimates likely overestimate true community richness due to spurious assignments and the inclusion of low confidence classifications inherent to Centrifuge-based taxonomic profiling of amplicon data. Overall, the fungal community at 36 h was dominated by *Hanseniaspora* spp., with several other yeasts present at lower relative abundance. Species-level designations are reported for completeness but should be interpreted cautiously; genus-level assignments are considered more reliable for interpretation.

**Table 1.** Temporal changes of representative volatile compounds during spontaneous fermentation of *C. arabica* (0-36 h)

| **Metabolite** | **Category** | **t0 (%)** | **t12 (%)** | **t24(%)** | **t36 (%)** | **Biological role** |
|---|---|---|---|---|---|---|
| Linalool | Monoterpenoid alcohol | ND | ND | ND | 0.34 | Key volatile, fruity aroma |
| Pentadecanal | Aliphatic aldehyde | 0.0 | 0.78 | 0.70 | 0.39 | LAB lipid oxidation |
| Caryophyllene | Sesquiterpene | 6.24 | ND | 1.89 | 1.07 | Substrate conversion |
| Eicosanoic acid | Long-chain fatty acid | 9.98 | ND | 0.49 | 0.27 | Advanced degradation |
| Benzetriol | Polyphenolic compound | ND | 0.99 | 0.62 | 0.25 | Polyphenol metabolism |
| Tetradecanal | Aliphatic aldehyde | ND | 0.11 | ND | 0.09 | Minor aldehyde product |

**Table 2.** Bacterial community composition at 36 h based on full-length 16S rRNA amplicon sequencing

| **Species** | **Read count** | **Relative abundance (%)** | **Putative characteristics (literature-based)** |
|---|---|---|---|
| *Leuconostoc falkenbergense* | 70,718 | 47.3 | Heterofermentative LAB; acid production |
| *Leuconostoc pseudomesenteroides* | 43,386 | 29.02 | Heterofermentative LAB; acid and $CO_2$ production |
| *Leuconostoc mesenteroides* | 18,267 | 12.22 | Heterofermentative LAB; dextran production |
| *Weissella cibaria* | 2,429 | 1.62 | Heterofermentative LAB; exopolysaccharide producer |
| *Klebsiella pasteurii* | 1,685 | 1.13 | Facultative anaerobe; opportunistic environmental bacterium; potential for citrate fermentation and nitrogen metabolism |
| *Gluconobacter cerinus* | 1,167 | 0.78 | Acetic acid bacterium; incomplete oxidation of sugars and sugar alcohols; produces acetic acid and other acids. |
| *Weissella confusa* | 575 | 0.38 | Heterofermentative LAB; EPS production; acidification and flavor modulation |
| *Klebsiella grimontii* | 417 | 0.28 | Facultative anaerobe; environmental/commensal species; fermentative metabolism with acid production |
| *Pantoea vagans* | 226 | 0.15 | Facultative plant-associated bacterium; potential phytase and sugar metabolism; limited fermentative role |
| *Enterobacter asburiae* | 205 | 0.14 | Facultative anaerobe; versatile carbohydrate metabolism; opportunistic environmental microbe |

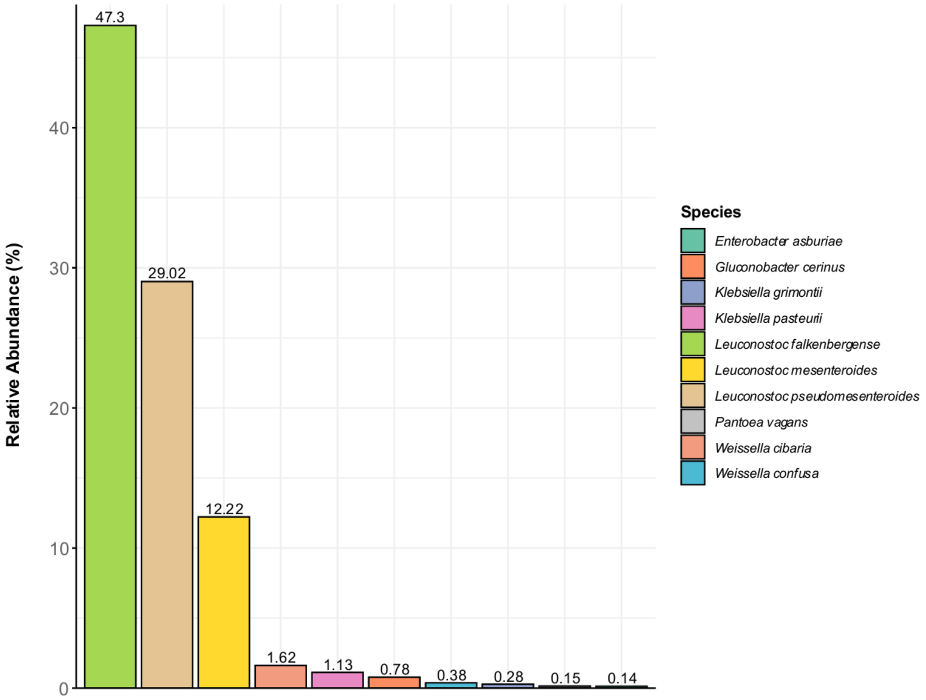

**Figure 1.** Ten bacterial species with the highest proportion of relative abundance toward total reads. The following plot presents the top 10 relative abundances at the species level

**Table 3.** Fungal community composition at 36 h based on ITS region amplicon sequencing

| Species | Read count | Relative abundance (%) | Putative characteristics (literature-based) |
|---|---|---|---|
| *Hanseniaspora uvarum* | 101,459 | 73.0 | Non-*Saccharomyces* yeast; β-glucosidase activity |
| *Hanseniaspora pseudoguilliermondii* | 9,318 | 6.7 | Non-*Saccharomyces* yeast; β-glucosidase activity |
| *Saturnispora gosingensis* | 6,794 | 4.89 | Non-*Saccharomyces* yeast; osmophilic tendency; moderate ethanol tolerance; potential ester production |
| *Starmerella davenportii* | 6,748 | 4.86 | Fructophilic yeast; high glycerol production; contributes to acidity and aroma complexity. |
| *Starmerella stellata* | 2,039 | 1.47 | Fructophilic yeast; high glycerol and organic acid production; low ethanol yield |
| *Pichia kluyveri* | 1,191 | 0.86 | Non-*Saccharomyces* yeast; strong ester producer (e.g., acetate esters); β-glucosidase activity |
| *Kurtzmaniella quercitrusa* | 1,170 | 0.84 | Environmental yeast; limited fermentative capacity; possible role in early-stage fermentation ecology |
| *Starmerella khaoyaiensis* | 656 | 0.47 | Fructophilic yeast; osmophilic; glycerol-producing; adapted to high-sugar substrates |
| *Meyerozyma caribbica* | 591 | 0.43 | Non-*Saccharomyces* yeast; biocontrol potential; β-glucosidase and other hydrolytic enzymes |
| *Hanseniaspora lachancei* | 484 | 0.35 | Non-*Saccharomyces* yeast; ester-producing; β-glucosidase activity; early fermentation colonizer |

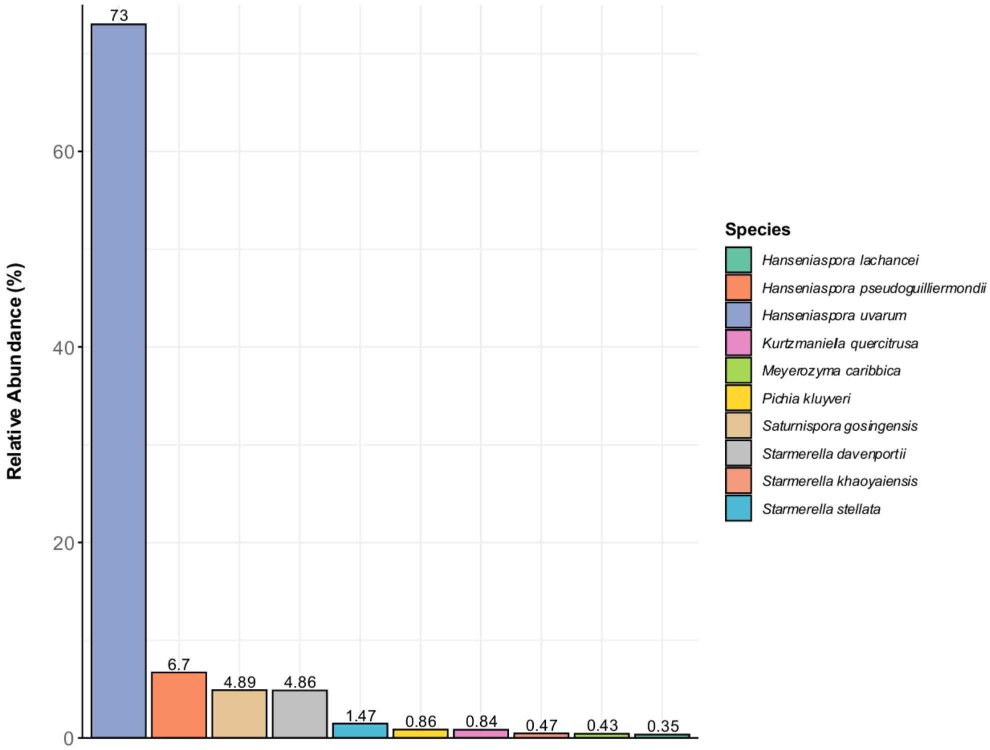

**Figure 2.** Ten fungal species with the highest proportion of relative abundance toward total reads. The following plot presents the top 10 relative abundances at the species level

## Discussion

This study provides a time-resolved description of volatile compound dynamics during a single spontaneous *C. arabica* wet fermentation batch, contextualized by endpoint bacterial and fungal community composition. The observed depletion of plant-derived compounds, such as eicosanoic acid and caryophyllene, alongside the appearance of fermentation-associated metabolites such as linalool and pentadecanal, is consistent with patterns reported in prior studies of spontaneous coffee fermentation (Galarza and Figueroa 2022; Silva et al. 2025).

### *VOC temporal patterns and putative microbial associations*

VOC data are reported as semi-quantitative Area% values, reflecting relative peak areas; absolute concentrations cannot be determined without external standards. Additionally, microbiome data were collected only at the 36-h endpoint, precluding time-series correlation analysis between specific taxa and specific VOC changes.

The observed temporal patterns of depletion of plant-derived compounds, such as eicosanoic acid and caryophyllene, alongside the appearance of fermentation-associated metabolites, including linalool and pentadecanal, are consistent with microbial transformation of substrate-derived compounds reported in prior studies of spontaneous coffee fermentation (Zhang et al. 2019; Galarza and Figueroa 2022; Silva et al. 2025). Specifically, this transformation is driven by a network of microbial enzymatic mechanisms: β-glucosidase-mediated hydrolysis releases volatile terpene aglycones (linalool) from glycosidically bound plant precursors. In contrast, microbial lipid oxidation (via β-oxidation) and esterase activities degrade long-chain fatty acids (eicosanoic acid) into shorter-chain aliphatic aldehydes and complex esters. These biochemical patterns support the hypothesis that substrate catabolism by the dominant LAB and yeast communities drives progressive VOC remodeling over the fermentation period.

The decrease in eicosanoic acid from 9.98% to 0.27% over 36 h is consistent with microbial lipid metabolism as a biochemically plausible hypothesis. However, no direct evidence of this process was obtained in this study. Biochemically, long-chain fatty acids such as eicosanoic acid can be degraded via β-oxidation pathways or esterified by microbial lipases and esterases to form shorter-chain fatty acid derivatives, thereby depleting the parent compound. LAB, including Leuconostoc species, have been reported to possess lipase and esterase activities in fermented food systems (Jun et al. 2024; Vargas-Luna et al. 2025), providing a literature-based mechanistic rationale for this proposed transformation. Direct demonstration of lipase or esterase activity in the present system was not performed, and this proposed pathway remains a testable hypothesis for future work.

The appearance of linalool exclusively at 36 h (0.34%) is noteworthy. Linalool is a monoterpene alcohol that can arise through enzymatic hydrolysis of glycosidic precursors or biotransformation of other terpenes. *Hanseniaspora uvarum*, which dominated the fungal community at 36 h (73.0%), has been reported to possess β-glucosidase activity capable of releasing volatile terpenes from glycosylated precursors (Zhang et al. 2021; Fan et al. 2022). This temporal alignment is biochemically plausible under the β-glucosidase mechanism described above, but does not constitute evidence of causality; the microbiome data were collected only at the 36-h endpoint and cannot establish whether *Hanseniaspora* was present and metabolically active at earlier timepoints when linalool was not yet detected. Time series microbiome sampling concurrent with VOC profiling would be required to determine whether increasing *Hanseniaspora* relative abundance precedes or coincides with linalool accumulation, thereby testing this hypothesis.

Pentadecanal appeared at 12 h and persisted through 36 h. Aldehydes can be produced through lipid oxidation pathways or amino acid catabolism. Both LAB and yeasts are capable of aldehyde production under fermentative conditions (Elhalis et al. 2021; Cassimiro et al. 2023), but the specific enzymatic mechanisms and responsible taxa in this batch remain uncharacterized.

### *Microbial community composition and functional hypotheses*

This study analyzed a single spontaneous fermentation batch without biological replication, and the observed community and VOC profiles may not be representative of other batches, seasons, or farms. Species-level assignments from Nanopore sequencing with Centrifuge classification are subject to known taxonomic resolution limitations and should be interpreted at the genus level with caution.

The dominance of *Leuconostoc* species (88.54% of bacterial reads) at 36 h is consistent with reports from other spontaneous LAB-driven fermentation systems, including cocoa, wine, and kimchi (Franco et al. 2021; Hamdaoui et al. 2022; Kim et al. 2024). *Leuconostoc* species are heterofermentative LAB that produce lactic acid, acetic acid, ethanol, and $CO_2$ from hexose sugars. They are often early colonizers in spontaneous fermentations due to their acid tolerance and metabolic versatility (Mudoor Sooresh et al. 2023). The high relative abundance of *L. falkenbergense* and *L. pseudomesenteroides* suggests these species may be well-adapted to the coffee fermentation niche. Still, their specific functional contributions to VOC production in this system have not been experimentally validated.

The dominance of *H. uvarum* (73.0% of fungal reads) aligns with observations from other spontaneous coffee fermentation studies (Elhalis et al. 2020; Pregolini et al. 2021; Martins et al. 2025). *Hanseniaspora uvarum* is a non-*Saccharomyces* yeast frequently associated with the early stages of fruit fermentation. While lactic acid bacteria primarily drive acidification, fungi (particularly yeasts) are generally considered more dominant in directly influencing coffee aroma complexity. Specifically, *H. uvarum* has been reported to produce aromatic compounds, including higher alcohols, esters, and volatile terpenes, and to exhibit β-glucosidase activity that can release bound aroma precursors (Lee and Park 2020; Gao et al. 2022b). However, direct enzyme assays or targeted metabolite quantification were not performed in this study, and the specific contributions of *H. uvarum* to the observed VOC profile remain hypothetical.

Table 4 presents a conceptual framework linking the observed microbial taxa to putative metabolic pathways reported in the literature. This table is intended to generate testable hypotheses for future work and does not represent functional inference from the amplicon data collected in this study. No statistical association testing was performed; the proposed microbe–metabolite relationships remain hypothetical and are inferred exclusively from published literature.

**Table 4.** Identified metabolic pathways during coffee fermentation. Five distinct metabolic routes were inferred from literature-informed pathway analysis linking detected organisms with detected metabolites

| Taxon | Putative pathway (Literature-based) | Potential products (Based on literature) | Measured in this study | Reference |
|---|---|---|---|---|
| *Leuconostoc* spp. | Heterofermentative glycolysis | Lactic acid, acetic acid, ethanol, $CO_2$ | Not measured | (Hamdaoui et al. 2022) |
| *Leuconostoc* spp. | Lipid metabolism | Short-chain fatty acids, aldehydes | Hexanoic acid, pentadecanal | (Jun et al. 2024) |
| *Hanseniaspora uvarum* | β-Glucosidase activity | Monoterpene alcohols (linalool, geraniol) | Linalool | (Zhang et al. 2021; Fan et al. 2022) |
| *Hanseniaspora uvarum* | Ester biosynthesis | Ethyl esters, acetate esters | Not measured | (Gao et al. 2022a) |
| *Lactococcus lactis* | Homofermentative glycolysis | Lactic acid | Not measured | (Franco et al. 2021) |
| *Pichia kudriavzevii* | Ester formation | Aromatic esters | Not measured | (Martins et al. 2025) |

In conclusion, this study provides a time-resolved description of Volatile Organic Compound (VOC) dynamics during a single 36-hour spontaneous *C. arabica* wet fermentation batch conducted under West Sumatran smallholder conditions. Among the major findings and contributions of this work: (i) GC-MS profiling across four timepoints identified 36 unique metabolites, revealing a progressive depletion of plant derived compounds notably eicosanoic acid (9.98% to 0.27%) and caryophyllene (6.24% to 1.07%) and the de novo appearance of fermentation-associated metabolites including linalool (0.34% at 36 h) and pentadecanal (first detected at 12 h); (ii) full-length 16S rRNA Nanopore sequencing of the 36-h endpoint community identified *Leuconostoc* spp. as the dominant bacterial genus (88.54% of reads), with *L. falkenbergense* as the most abundant species (47.3%) a community composition not previously reported from Indonesian Arabica fermentation and ITS sequencing identified *H. uvarum* as the dominant fungal species (73.0%); (iii) the temporal co-occurrence of linalool detection with *H. uvarum* dominance at 36 h represents a novel, testable hypothesis linking β-glucosidase mediated terpene release to this yeast's enzymatic activity in this system. These findings contribute baseline data on the fermentation microbiome and volatilome of West Sumatran *C. arabica* and position this system for future mechanistic studies. Functional validation through enzyme assays, targeted metabolite quantification with authentic standards, and controlled inoculation experiments remains a priority for future work, as does biological replication and time-series microbiome sampling to formally test the microbe metabolite associations proposed here.

## ACKNOWLEDGEMENTS

Universitas Andalas, Indonesia, funded this study under the Penelitian Dosen Pemula (PDP) scheme (Associate Lecturer Research Scheme), Batch I, Research Contract No. 33/UN16.19/PT.01.03/PDP/2025, Fiscal Year 2025. The authors also wish to express their gratitude to Koperasi Produsen Serba Usaha (KPSU) Solok Radjo, Indonesia, for providing the coffee samples and for their technical assistance during the field data collection.

## REFERENCES

Breitwieser FP, Salzberg SL. 2020. Pavian: interactive analysis of metagenomics data for microbiome studies and forensic investigations. Bioinformatics 36 (4): 1303-1304. https://doi.org/10.1093/bioinformatics/btz715.

Cassimiro DM, Batista NN, Coutrim MX, Ávila e Silva S, Dias DR, Schwan RF. 2023. Wet fermentation of *Coffea canephora* by lactic acid bacteria and yeasts using the self-induced anaerobic fermentation (SIAF) method enhances the coffee quality. Food Microbiol 110: 104161. https://doi.org/10.1016/j.fm.2022.104161.

De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: Visualizing and processing long-read sequencing data. Bioinformatics 34 (15): 2666–2669. https://doi.org/10.1093/bioinformatics/bty149.

Elhalis H, Cox J, Frank D, Zhao J. 2021. Microbiological and biochemical performances of six yeast species as potential starter cultures for wet fermentation of coffee beans. LWT 137: 110430. https://doi.org/10.1016/j.lwt.2020.110430.

Elhalis H, Cox J, Zhao J. 2020. Ecological diversity, evolution and metabolism of microbial communities in the wet fermentation of Australian coffee beans. Intl J Food Microbiol 321: 108544. https://doi.org/10.1016/j.ijfoodmicro.2020.108544.

Elhalis H, Cox J, Zhao J. 2023a. Coffee fermentation: Expedition from traditional to controlled process and perspectives for industrialization. Appl Food Res 3 (1): 100253. https://doi.org/10.1016/j.afres.2022.100253.

Elhalis H, Cox J, Zhao J. 2023b. Yeasts are essential for mucilage degradation of coffee beans during wet fermentation. Yeast 40 (9): 425-436. https://doi.org/10.1002/yea.3888.

Fan T, Xu Y, Li Y, Shi Y, Xu F, Zhang M, Zhou Y, Du L. 2022. Localization, purification, and characterization of a novel β-glucosidase from *Hanseniaspora uvarum* Yun268. J Food Sci 87 (3): 886-894. https://doi.org/10.1111/1750-3841.16068.

Franco W, Benavides S, Valencia P, Ramírez C, Urtubia A. 2021. Native yeasts and lactic acid bacteria isolated from spontaneous fermentation of seven grape cultivars from the Maule Region (Chile). Foods 10 (8): 1737. https://doi.org/10.3390/foods10081737.

Galarza G, Figueroa JG. 2022. Volatile compound characterization of coffee (*Coffea arabica*) processed at different fermentation times using SPME–GC–MS. Molecules 27 (6): 2004. https://doi.org/10.3390/molecules27062004.

Gao P, Peng S, Luo J, Wang Q, Liu J, Feng Y. 2022a. Indigenous non-*Saccharomyces* yeasts with β-glucosidase activity in sequential fermentation with *Saccharomyces cerevisiae*: A strategy to improve the volatile composition and sensory characteristics of wines. Front Microbiol 13: 845837. https://doi.org/10.3389/fmicb.2022.845837.

Gao P, Sam FE, Liu J, Zhang N, Chen Y, Luo J, Feng Y. 2022b. Enzymatic characterization of purified β-glucosidase from non-*Saccharomyces* yeasts and application on chardonnay aging. Foods 11 (6): 852. https://doi.org/10.3390/foods11060852.

Góngora CE, Tapias VA, Figueroa JG, Oliveros CE. 2024. Metataxonomic identification of microorganisms during the coffee fermentation process in Colombian farms (Cesar Department). Foods 13 (6): 839. https://doi.org/10.3390/foods13060839.

Hamdaoui N, Nasri N, Zoghlami W, Ayadi MA. 2022. Technological aptitude and sensitivity of lactic acid bacteria *Leuconostoc* isolated from raw milk of cows: From step-by-step experimental procedure to the results. Indones J Sci Technol 8 (2): 157-170. https://doi.org/10.17509/ijost.v8i2.53730.

Holguín-Sterling L, Zapata-Zapata AD, Durango-Restrepo DL. 2023. Physical–chemical and metataxonomic characterization of the microbial communities present during the fermentation of three varieties of coffee from Colombia and their sensory qualities. Agriculture 13 (10): 1980. https://doi.org/10.3390/agriculture13101980.

Jun BG, Kim AR, Lee YG, Kim DW, Park DH, Park MK. 2024. Metabolomic comparison of guava (*Psidium guajava* L.) leaf extracts fermented by *Limosilactobacillus fermentum* and *Lactiplantibacillus plantarum* and their antioxidant and antiglycation activities. Nutrients 16 (6): 841. https://doi.org/10.3390/nu16060841.

Kim D, Song L, Breitwieser FP, Salzberg SL. 2016. Centrifuge: Rapid and sensitive classification of metagenomic sequences. Genome Res 26 (12): 1721-1729. https://doi.org/10.1101/gr.210641.116.

Kim S, Jung MJ, Song HS, Whon TW, Roh SW, Park DS. 2024. *Leuconostoc aquikimchii* sp. nov., a lactic acid bacterium isolated from cabbage watery kimchi. J Microbiol 62 (12): 1089-1097. https://doi.org/10.1007/s12275-024-00188-z.

Lee SB, Park HD. 2020. Isolation and investigation of potential non-*Saccharomyces* yeasts to improve the volatile terpene compounds in Korean muscat bailey a wine. Microorganisms 8 (10): 1552. https://doi.org/10.3390/microorganisms8101552.

Martins PMM, Lopes MG, Duarte WF, Dias DR, Schwan RF. 2025. Selection and characterization of non-*Saccharomyces* yeast strains for potential use in Arabica and Conilon coffee fermentations. J Food Sci 90 (7): e70431. https://doi.org/10.1111/1750-3841.70431.

Mudoor Sooresh M, Willing BP, Bourrie BCT. 2023. Opportunities and challenges of understanding community assembly in spontaneous food fermentation. Foods 12 (3): 673. https://doi.org/10.3390/foods12030673.

Oksanen J, Simpson G, Blanchet F et al. 2022. Vegan: Community Ecology Package. R Package Version 2.6-4. https://doi.org/10.32614/CRAN.package.vegan.

Ondov BD, Bergman NH, Phillippy AM. 2011. Interactive metagenomic visualization in a web browser. BMC Bioinf 12: 385. https://doi.org/10.1186/1471-2105-12-385.

Pereira GVM, Vale AS, Ribeiro-Barros AI, Rodrigues LRS, Mirção GMFB, Camilo B, Tapaça IPE, Sampaio VM, Brar SK, Soccol CR. 2025. Integrated microbial–metabolomic analysis reveals how fermentation contributes to the unique flavor of African Arabica coffee. Food Chem Mol Sci 12: 100344. https://doi.org/10.1016/j.fochms.2025.100344.

Pothakos V, Snauwaert I, De Vos P, Huys G, Devlieghere F. 2020. Temporal shotgun metagenomics of an Ecuadorian coffee fermentation process highlights the predominance of lactic acid bacteria. Curr Res Biotechnol 2: 1-15. https://doi.org/10.1016/j.crbiot.2020.02.001.

Pregolini VB, Savoi S, Pisano MB, Costantini A. 2021. Influence of environmental microbiota on the activity and metabolism of starter cultures used in coffee beans fermentation. Fermentation 7 (4): 278. https://doi.org/10.3390/fermentation7040278.

Schwan RF, Silva CF, Lopez-Palestino A, Pereira GVM. 2023. The essential role of spontaneous and starter yeasts in cocoa and coffee fermentation. FEMS Yeast Res 23: foad019. https://doi.org/10.1093/femsyr/foad019.

Silva ME, Melo PS, Alencar SM, Lemes GF, Pereira GVM. 2025. Impact of fermentation time on the bioactive and volatile composition of coffee: Insights for producers and researchers. Food Chem 490: 145067. https://doi.org/10.1016/j.foodchem.2025.145067.

Todhanakasem T, Saenchai C, Krasaesakul N, Areesirisuk A, Thanonkeo P, Ratanapongleka K. 2024. The relationship between microbial communities in coffee fermentation and aroma with metabolite attributes of finished products. Foods 13 (15): 2332. https://doi.org/10.3390/foods13152332.

Vargas-Luna C, Melo D, Opazo MC, Romero J. 2025. Screening and selection of native lactic acid bacteria isolated from Chilean grapes. Foods 14 (1): 143. https://doi.org/10.3390/foods14010143.

Zhang P, Barlow S, Krstic M, Herderich M, Fuentes S, Howell K. 2021. Beta-glucosidase activity of wine yeasts and its impacts on wine volatiles and phenolics: A mini-review. Food Microbiol 100: 103859. https://doi.org/10.1016/j.fm.2021.103859.

Zhang SJ, De Bruyn F, Pothakos V, Torres J, Falconi C, Moccand C, Weckx S, De Vuyst L. 2019. Following coffee production from cherries to cup: Microbiological and metabolomic analysis of wet processing of *Coffea arabica*. Appl Environ Microbiol 85 (6): e02635-18. https://doi.org/10.1128/AEM.02635-18.