

RESEARCH

Open Access



Comparative profiling of microbial communities and volatile organic compounds in fermented wrapper, binder, and filler cigar tobaccos

Mingzhu Zhang^{1†}, Dongfeng Guo^{2*†}, Haiqing Wang¹, Guanglong Wu¹, Yaqi Shi^{1,2}, Jinlong Zhou², Tianfei Zheng^{1,2}, Eryong Zhao¹, Xiaolei Wu³ and Xingjiang Li^{1*}

Abstract

Background Economic benefits for tobacco growers are closely linked to the quality of fermented cigar tobacco leaves (CTLs). This research focused on an in-depth examination of the microbial community and flavor compounds within CTLs, specifically analyzing the wrapper, binder, and filler components of a cigar. The primary objective was to unravel the complex relationship between the microbial composition and the resultant flavor profiles, thereby providing insights that could enhance the economic value of CTLs.

Results The study revealed distinct variations in flavor chemicals and microbiota across different sections of CTLs. Prominent species identified in the fermented CTLs included *Corynebacterium*, *Pseudomonas*, *Staphylococcus*, *Aspergillus*, and *Cladosporium*. Bidirectional orthogonal partial least squares (O2PLS) analysis pinpointed five bacterial and four fungal species as key contributors to flavor compound formation. Additionally, an analysis considering Within-module and Among-module connectivity highlighted two bacterial and thirteen fungal genera as keystone species. The insights from Partial Least Squares Structural Equation Modeling (PLS-SEM) further underscored the influential role of fungal microorganisms in defining CTLs' flavor profile.

Conclusions The research findings illuminate the intricate interplay between flavor chemicals and microbes in the traditional fermentation process of CTLs.

Keywords Cigar tobacco leaf, Microbial communities, Volatile organic compounds, Yunnan region

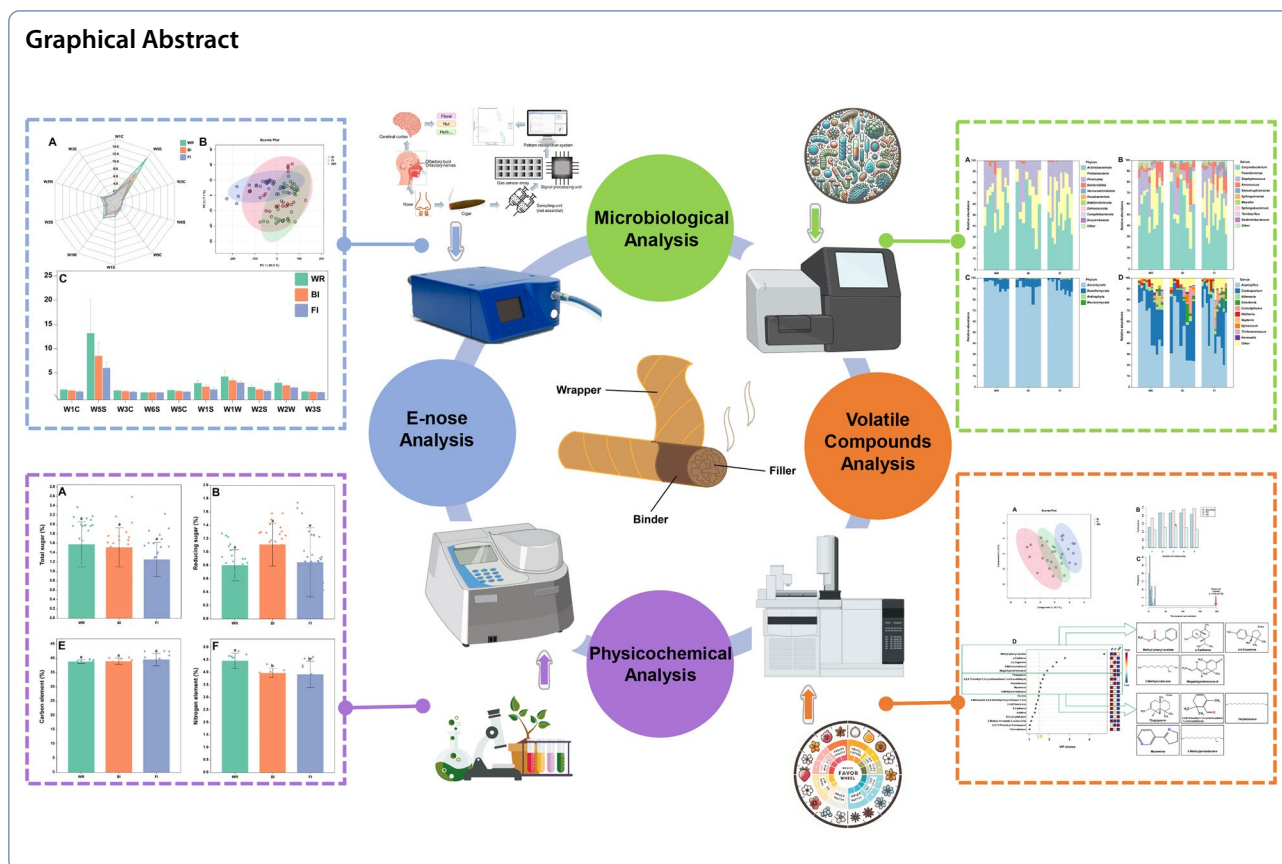
[†]Mingzhu Zhang and Dongfeng Guo contributed equally to this article and should be considered co-first authors.

*Correspondence:

Dongfeng Guo
gdf0221@163.com
Xingjiang Li

lixingjiang@hfut.edu.cn

Full list of author information is available at the end of the article



Introduction

Cigars are traditionally composed of three main parts: the filler, binder, and wrapper, each possessing unique characteristics and roles that collectively define a cigar's flavor, combustion properties, and appearance. The Filler constitutes the core of the cigar, determining its strength and primary flavors. The Binder encases the filler leaves, providing structural stability and adding layers of flavor. The Wrapper, as the cigar's outermost leaf, plays a crucial role in influencing the overall quality and smoking experience through its quality and appearance. Together, these components synergize to create the distinct flavor profile and smoking experience of cigars. Therefore, to understand the multifaceted flavor characteristics of cigars, a comprehensive analysis of the wrapper, binder, and filler is indispensable. It is widely recognized that the quality of the wrapper, binder, and filler is closely linked to the cigar type, cultivation methods, harvest maturity, and processing techniques. Particularly, the fermentation process of cigar tobacco leaves is crucial for enhancing their quality. Initiated after harvesting and processing, this phase involves complex chemical and biochemical transformations within the leaf's organic compounds, facilitated by the combined action of inorganic elements,

enzymes, and microorganisms. These transformations, catalyzed by inorganic elements such as iron (Fe) and magnesium (Mg), lead to the oxidation of organic matter in the presence of atmospheric oxygen [1]. This fermentation harmonizes the leaf's chemical components, enhancing the overall quality. Fresh, unfermented CTLs often possess green, earthy, and woody odors, producing a harsh and irritating smoke. Conversely, the fermentation process significantly reduces the greenness and earthy odors, resulting in smoother, less irritating smoke. The leaves gain elasticity and improved combustibility. This improvement in physical properties, physicochemical characteristics, aroma, and smoking quality is a direct outcome of the fermentation process [2].

Different CTLs may harbor unique microbial communities, which are essential in driving their fermentation process [3]. These microbial activities, along with enzymatic catalysis and complex chemical reactions, lead to the breakdown of proteins and starch, generating flavor compounds like acids, ketones, aldehydes, and alcohols [4, 5]. These substances, particularly organic acids, and amino acids, act as flavor components or precursors in CTLs, while aromatic substances significantly impact tobacco quality and sensory characteristics [6]. For

instance, trimethylamine has been identified as a contributor to the undesirable ammonia taste in cigars [7]. In the fermentation process, research has pinpointed benzaldehyde, noted for its almond and cherry aromas, as a key metabolite [8]. This parallels findings in fermented foods, where microbial roles in flavor compound formation are well-documented [9] but less so in cigar tobacco leaf fermentation. A deeper understanding of these microbial communities and their contributions to flavor compounds is thus imperative. This knowledge will enhance our grasp of the fermentation process and aid in achieving consistency in the quality of fermented leaves. Analytical methods like bidirectional orthogonal partial least squares (O2PLS) modeling, and analyses of within-module (Zi) and among-module (Pi) connectivity, along with PLS-SEM, are instrumental in this research. These methods have been effectively used in food fermentation and environmental microbiology. For instance, Guan et al. [10] investigated microbial succession and flavor changes during suansun fermentation, identifying crucial microbes like *Lactobacillus*, *Clostridium sensu stricto 1*, *Enterobacter*, and *Leuconostoc* using O2PLS. Zhou et al. [11] studied the dynamics of pit mud microbial communities in Baijiu fermentation, correlating them with key physicochemical factors and identifying significant OTUs related to pit mud aging through Zi and Pi analysis. Additionally, Wang et al. [12] employed PLS-SEM to quantify the direct and indirect impacts and interactions of natural environments and human activities on wetland changes. However, there has been no extensive application of O2PLS, Zi, Pi, and PLS-SEM to analyze the relationships among microorganisms, flavor compounds, and chemical components in tobacco fermentation samples.

Past research has highlighted significant regional variations in microbial communities and flavor profiles of CTLs [5]. Despite this, there has been no comprehensive analysis of the bioinformatics characteristics and quality attributes of the wrapper, binder, and filler leaves from CTLs in Yunnan's unified fermentation process. This study, therefore, zeroes in on CTLs from two emblematic production areas in Yunnan—Mang City, Dehong Prefecture (DHMS) and Hani-Yi Autonomous County of Jiangcheng, Pu'er City (PEJC). It meticulously examines the physicochemical properties, flavor metabolites, and microbial composition of various types of CTLs post-fermentation. Utilizing O2PLS, Zi, and Pi, the study identifies functional genera in different CTL types and employs PLS-SEM modeling to explore the intricate relationships driving the interplay between chemical components, microbial communities, and flavor compounds. The goal of this research is to delve into the core reasons for the distinct characteristics observed in different types of CTLs, particularly from the angles of microorganisms

and characteristic metabolites. This investigation aims to provide a theoretical framework for enhancing the standardization and precise control in the fermentation of CTLs.

Materials and methods

Sample collection

In April 2023, a total of 90 fermented CTL samples were collected from the Ganzhuang Fermentation Center in Yunnan Province, China. This collection included 30 samples from WR (Wrapper), 24 from BI (Binder), and 36 from FI (Filler). The specifics of each sample, derived from different fermentation piles, are detailed in Additional file 1: Table S1. To ensure comprehensive and representative sampling, leaves were gathered from two distinct planting locations, DHMS and PEJC. The collection spanned five strategic points across the high, middle, and bottom layers of the fermentation heaps, as depicted in Additional file 1: Fig. S1. Samples from these three strata were amalgamated to form a single biological replicate, aiming to encompass the full range of variation in the different samples. Post-collection, all samples were promptly frozen and pulverized in liquid nitrogen. They were then stored at two different temperatures, -20°C and -80°C , to preserve their integrity for subsequent chemical component analysis and DNA extraction.

Chemical components and color analyses

The color parameters of the WR, BI, and FI CTLs were quantitatively assessed using an SR-68 portable colorimeter (D65 light source, 3nh, China). This device measured the L^* value (lightness), and b^* value (yellowness), and used a \pm calibration plate. Each tobacco leaf sample underwent three measurements, with the average value representing its color. In studying the color differences among various CTLs, the overall color differences were calculated [13], with ΔL^* , Δa^* , and Δb^* representing changes in lightness, redness, and yellowness, respectively. According to perceptual thresholds, a ΔE^* greater than 3 is distinguishable by the human eye, between 1 and 3 is less discernible, and less than 1 is indistinguishable [14].

For protein content analysis, we adhered to the Chinese National Standard GB/T 5009.5–2016, utilizing a Kjeldahl apparatus (Hanon Technologies CO., Ltd., Shandong, China). Total and reducing sugar levels were determined via ultraviolet spectroscopy (Shanghai Metash Instruments CO., Ltd., Shanghai, China). The starch content was ascertained using the iodine colorimetry method. Furthermore, an Elemental Analyzer (Flash Smart, Thermo Fisher, America) was employed to quantify chemical elements, including carbon C, hydrogen (H), sulfur (S), and nitrogen (N), in the samples.

Free amino acids (FAAs) analysis

Chen et al. [15] described an improved method for processing samples to analyze FAAs. Initially, to eliminate any interference from proteins, they were precipitated using a sulfosalicylic acid solution. This involved mixing 0.4 g of CTLs with 8 mL of 15% (w/v) sulfosalicylic acid solution. The mixture was then incubated at 4 °C for 1 h, followed by centrifugation at 7000 rpm for 15 min at the same temperature. For amino acid analysis, the supernatant was filtered through a 0.22 µm filter and then transferred to 1 mL vials for assessment using an automatic amino acid analyzer (S-433D, Sykam, Eresing, Germany). The detection of proline was carried out at a UV wavelength of 440 nm, while the general detection wavelength for amino acids was set at 570 nm. Quantification was performed using an external standard method.

E-nose analysis

The E-nose (PEN3, Airsense Analytics GmbH, Schwerin, Germany) was utilized to evaluate the aroma of tobacco leaf samples. For the analysis, 0.5 g of finely chopped and homogenized samples were placed in 20 mL headspace vials at 60 °C for preheated 10 min. Post-heating, the samples were ready for testing. The E-nose probe underwent a cleaning cycle for 90 s, followed by a reset period of 5 s, and a pre-sampling phase of 5 s. The carrier gas flow during sampling was maintained at 400 mL/min, and each sample's measurement duration was 150 s. The E-nose is equipped with ten different metal oxide sensors, each sensitive to specific compounds: W5C detects alkane aromatic compounds; W1S is tuned to short-chain alkanes; W1W identifies sulfides and terpenes; W2S recognizes alcohols, aldehydes, and ketones; W2W is for organic sulfides and aromatic components; W3S is for long-chain alkanes; W1C detects aromatic compounds; W5S is sensitive to nitrogen oxides; W3C responds to aromatic ammonia; and W6S detects hydrogen presence.

HS-SPME-GC-MS analysis

For the extraction of volatile organic compounds (VOCs) from CTLs, Headspace Solid-Phase Microextraction coupled with Gas Chromatography-Mass Spectrometry (HS-SPME-GC-MS) was employed. Following the methodology of Zheng et al. [16], with slight modifications, 0.5 g of tobacco samples were placed in a 20 mL headspace vial. To the vial, 8 mL of saturated NaCl solution and 1 µL of phenylethyl acetate (128.75 µg/µL) were added, and the mixture was then heated in a water bath at 75 °C for 20 min. VOC extraction was performed for 35 min using DVB-CAR-PDMS fibers (50/30 µm, Supelco Inc., Bellefonte, PA, USA). The extracted VOCs were analyzed using an Agilent 8890-7000D gas chromatograph-mass spectrometer system equipped with a fused

quartz capillary column (Agilent, Santa Clara, CA, USA). The separation of target compounds was conducted on an HP-5MS column (30 m×0.32 mm i.d., 0.25 µm film thickness; J&W Scientific, CA, USA) using helium as the carrier gas at a flow rate of 0.8 mL/min. The column temperature was initially set at 60 °C for 2 min, followed by a temperature ramp to 180 °C at a rate of 3 °C/min for 2 min. The temperature was further increased to 260 °C at 6 °C/min and held for 2 min. Mass spectrometry detection was performed over a range of 35–450 m/z with an ionization voltage of 70 eV.

Characterization of microbiota

Genomic DNA from all 90 samples was extracted using the HiPure Soil DNA Kits (Magen, Guangzhou, China). The DNA's quantity and quality were assessed with a NanoDrop 2000 spectrophotometer (Thermo Scientific, USA). To identify various bacterial and fungal types, universal primers were employed for amplifying the 16S rRNA and ITS regions. The primer pairs 515F (5'-GTG YCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTA CNVGGGTWCTAAT-3') were used for 16S rRNA, and ITS1F (5'-CTTGGTCATTTAGAGGAAGTAA-3') and ITS2 (5'-GCTGCGTTCTTCATCGATGC-3') for the ITS region of fungi. The PCR conditions were as follows: initial denaturation at 95 °C for 2 min, 35 cycles of denaturation at 95 °C for 30 s, annealing at 60 °C for 45 s, and extension at 72 °C for 90 s, concluding with a final extension at 72 °C for 10 min. The PCR products were mixed and purified using the Qiagen Gel Extraction Kit (Qiagen, Germany). Sequencing libraries were prepared using the Illumina TruSeq® DNA PCR-Free Sample Preparation Kit (Illumina, USA), following the manufacturer's guidelines. Post library quality assessment, sequencing was performed on the Illumina NovaSeq platform using a 250 bp paired-end run configuration.

Bioinformatics and statistical analyses

The raw data from the Illumina platform were processed using FASTP (version 0.18.0) [17] with specific criteria: 1) Reads containing unknown nucleotides (N) ≥ 10% were removed; 2) Reads with bases having a Phred quality score ≤ 20 comprising ≥ 50% of the read were removed; 3) Reads containing adapters were deleted. The resulting clean reads were then assembled with FLASH (version 1.2.11) [18], setting a minimum overlap of 10 bp and a maximum mismatch rate of 2%. Following the filtering criteria from the literature [19], low-quality tags were removed to obtain high-quality clean tags. Adhering to the tags quality control process of Qiime [20], tags were truncated and filtered based on length. UPARSE (version 9.2.64) [21] grouped clean tags into operational taxonomic units (OTUs) with a similarity threshold of ≥ 97%.

The UCHIME algorithm [22] was employed for chimera detection in tags. Post chimera filtering, the effective tags were used for OTU abundance statistics and further analyses. For fungal identification, the UNITE database was utilized, while for bacterial identification, the Silva database was employed. Qiime [20] (version 1.9.1) calculated the ACE and Shannon indices for alpha diversity in the microbial community. The Wilcoxon rank-sum test was conducted to evaluate the significance of diversity differences between groups. The Vegan package in R language was used for principal coordinates analysis (PCoA) based on (un)weighted Unifrac, Jaccard, and Bray–Curtis distances [23] (version 2.5.3). The top ten species' relative abundances were estimated using the largest abundance ranking method. LEfSe analysis determined differences in species abundance, with the LDA threshold set at 2.5 for both bacteria and fungi. OTUs were subjected to a two-condition filter [24]: only those found in over 30% of samples and with a relative abundance greater than 0.1% were retained. This filter yielded 12 bacterial and 28 fungal OTUs (Additional file 1: Table S2) for network construction. The final network comprised 246 edges ($|r| > 0.6$, $p < 0.05$) (Additional file 1: Table S3). Within-module connectivity (Z_i) and among-module connectivity (P_i) were utilized as metrics. OTUs were classified into four categories based on Z_i and P_i values: peripherals ($Z_i \leq 2.5$, $P_i \leq 0.62$), connections ($Z_i \leq 0.25$, $P_i > 0.62$), module hubs ($Z_i > 2.5$, $P_i \leq 0.62$), and network hubs ($Z_i > 2.5$, $P_i > 0.62$) [24]. Gephi (version 0.10.1) visualized the correlation network. MetaboAnalyst conducted a partial least-squares discriminant analysis (PLS-DA) on the dataset. O2PLS analysis, using OmicsPLS [25], selected species associated with flavor chemicals. The vegan package 'cca' analyzed interactions between bacteria, taste compounds, and fermentation chemical components. PLS-SEM [26, 27] explored how CTLs' chemical component characteristics mediate microbial diversity and key core species, affecting metabolic product changes. PLS-SEM, a data analysis approach, uses a latent variable to summarize observed variables and assumes linear correlations between latent variables [26]. Path coefficients and R^2 values were estimated, and fit indices like SRMR, d_{ULS} , and d_G evaluated the models. Acceptable PLS model values are $SRMR < 0.08$, $d_{ULS} < 0.95$, and $d_G < 0.95$ [28]. Models were built using SmartPLS (version 4.0.9.2.), and SPSS Statistics (version 27) conducted all statistical analyses.

Accession numbers

The raw sequencing data were deposited in the Genome Sequence Archive at the China National Center for Bio-information, under the BioProject IDs PRJCA022549 and PRJCA022550. The associated BioSample accession

numbers for these submissions are subSAM116878 and subSAM116879.

Results

Chemical components and color analysis

Additional file 1: Fig S2, utilizing proportions and concentrations of chemical components, compares the content of total sugar, reducing sugar, starch, protein, and elements C, N, H, and S in three categories of CTLs samples: WR, BI, and FI. The analysis revealed that the protein content (26.32 g/100 g) and N content (4.46%) in WR were notably higher than in BI (23.52 g/100 g and 3.98%) and FI (22.58 g/100 g and 3.92%). However, the composition of other chemical components was relatively homogeneous across the groups. Furthermore, there were no significant differences in color values L^* , a^* , b^* , and ΔE^* as shown in Additional file 1: Table S4.

FAAs and E-nose analysis

The CTLs contained seventeen FAAs, categorized based on flavor characteristics into umami, sweet, bitter, and salt-taste groups. The total concentrations of these amino acids varied from 0.30 to 148.03 mg/g. Umami amino acids were predominant, constituting about 80% of the total free amino acids. The levels of sweet and bitter amino acids were comparable, each accounting for approximately 10% and 8% of the total free amino acids, respectively (Additional file 1: Table S5). The amino acid composition of the samples was generally similar, with aspartic acid (Asp) and glutamic acid (Glu) being the most abundant, while cysteine (Cys) and tyrosine (Tyr) had lower contents (Fig. 1A). Asp and Glu, exceeding the taste thresholds of 1.0 mg/g and 0.3 mg/g, were identified as key contributors to umami flavor. The taste active value (TAV) of Asp and Glu for the three types of CTLs was calculated as 4.40, 4.28; 4.12, 2.85; 1.95, and 1.97, respectively (Additional file 1: Table S6). $TAV < 1$ compounds are typically thought to have little taste contribution in food, thus Asp and Glu contributed to the umami taste of all samples, while other amino acids did not (Fig. 1D). PLS-DA revealed no distinct differentiation in amino acid concentrations among WR, BI, and FI CTLs (Fig. 1B). However, variable importance in projection (VIP) scores ≥ 1.0 identified Glu, Ser, and Asp as the most discriminative amino acids (Fig. 1C, Additional file 1: Table S7), indicating their significant role in CTL flavor. The E-nose system, showed nearly identical radar images for WR, BI, and FI across sensors W3S, W1C, W3C, W6S, and W5C, with relatively low signal intensity. This suggests minimal production of aromatic compounds, ammonia substances, and short-chain alkanes in fermented CTLs (Fig. 2A). The W2S sensor indicated modest signal intensity but varied between samples,

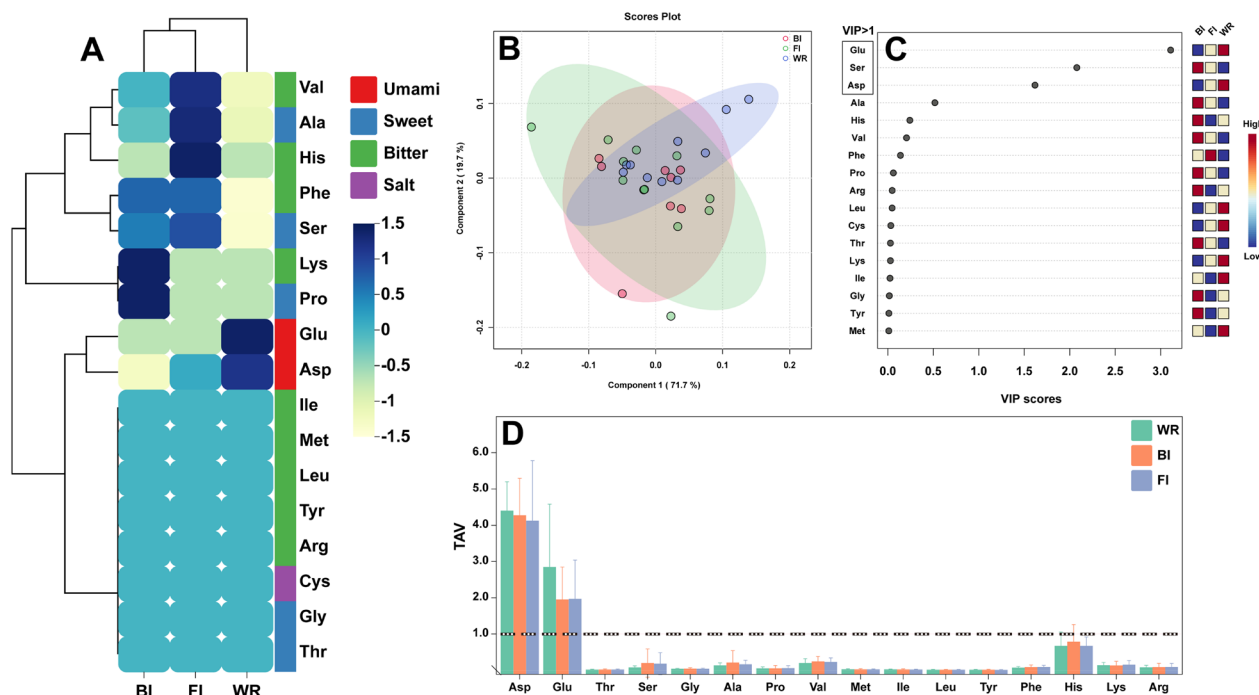


Fig. 1 The free amino acids in CTLs from different types. **A** Heatmap and hierarchical clustering of amino acids content in CTLs. The horizontal axis represents the group name, and the vertical axis represents the amino acid name. The color gradient within each color block indicates the abundance variation of the respective amino acids in the sample. Bitter, Salt, Sweet, and Umami represent different tastes. **B** PLS-DA analysis for amino acid content in CTLs. **C** The VIP value associated with PLS-DA analysis for amino acids content in CTLs from WR, BI, and FI. **D** The TAV analysis for amino acids content in CTLs

implying differences in alcohols, aldehydes, ketones, ethers, and other compounds throughout different types. W2W, W1W, and W1S sensors showed slightly higher signals, but without significant difference, suggesting the presence of certain organic sulfides in CTLs. The W5S sensor's signal strength varied significantly for WR, BI, and FI, highlighting nitrogen oxides as differentiators of CTL types (Fig. 2C). The PCA diagram reveals that the cumulative variance contribution rate of PC1 (89.5%) and PC2 (7.1%) amounts to 96.6% (Fig. 2B). This suggests that the primary components effectively represent the overall flavor distribution of each sample. A partial overlap between the WR and FI samples on the PCA plot implies a similarity in their flavor profiles. Additionally, the close positioning of the BI sample to the WR and FI indicates a high degree of flavor resemblance among them. This similarity is attributed to the processing practices of binder CTLs; in production, some BI leaves are often selected from the WR and FI batches. While the E-nose system proficiently captures the overall fragrance characteristics of the samples, it falls short in pinpointing specific changes in individual taste chemicals before and after treatment. To address this, further analysis using GC-MS is necessary for a more detailed understanding of the flavor compounds.

Qualitative analysis of VOCs by HS-SPME-GC-MS

In the analysis of VOCs in three different types of CTL samples, a total of 90 compounds were identified. A detailed list of these compounds, along with their CAS numbers, names, and VIP values, is provided in Additional file 1: Table S8. To discern any significant differences in the volatile profiles among the various CTL types, PLS-DA was conducted. The resulting PLS-DA model (Fig. 3A) exhibited a good fit with $R^2=0.73$ and predictive accuracy of $Q^2=0.44$, indicating that the VOCs could be effectively classified into three distinct categories (Fig. 3B). This classification was reliable for both the training and testing data sets. Additionally, a permutation test confirmed that the models were not overfitted (Fig. 3C). Each compound was assigned a VIP score, which quantifies its contribution to the separation of the groups. A higher VIP score implies a more significant role in distinguishing between groups. As shown in Fig. 3D, ten VOCs—Methyl phenyl acetate, γ -Cadinene, (+)-Cuparene, 3-Methylundecane, Megastigmatrienone-A, Thujopsene, 2,6,6-Trimethyl-1,3-cyclohexadiene-1-carboxaldehyde, Heptadecane, Myosmine, and 3-Methylpentadecane—had VIP scores greater than 1.5. These compounds were considered to be substantially differentiating factors among the three CTL groups. The majority

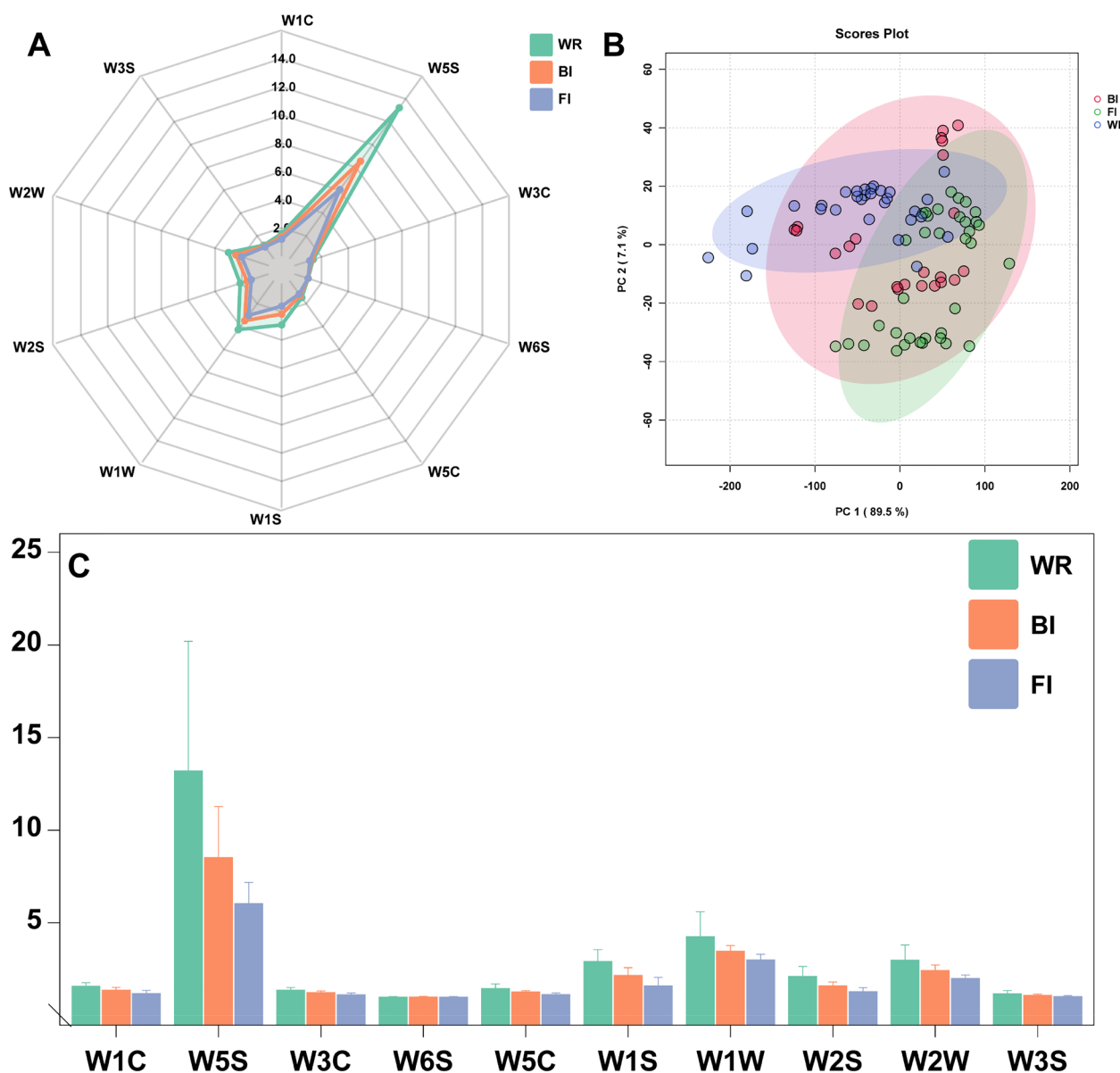


Fig. 2 **A** E-nose radar of different types of CTLs. **B** The PCA plots of sensory evaluation of CTLs in different types. **C** Changes of E-nose W1C, W5S, W3C, W6S, W5C, W1S, W1W, W2S, W2W, W3S sensor signal response of CTLs in different types

of these substances were predominantly enriched in BI and FI, while their concentration in WR was comparatively lower.

Microbiota diversity and microbial community composition analysis

To assess the diversity of microbial communities in different types of CTLs, the ACE and Shannon indices were employed. For bacterial communities, WR exhibited a higher diversity compared to BI and FI, although the difference was not statistically significant ($p > 0.05$)

(Additional file 1: Fig. S3A, B). Conversely, the fungal community in BI showed greater diversity than in WR and FI, but again, the difference was not statistically significant ($p > 0.05$) (Additional file 1: Fig. S3E, F). In terms of bacterial OTUs, the WR, BI, and FI samples contained 37, 32, and 34 OTUs, respectively. Among these, 22 OTUs were common to all three types, while the unique OTUs for WR, BI, and FI were 13, 6, and 6, respectively (Additional file 1: Fig. S3C). Fungal analysis identified 177, 206, and 175 OTUs in WR, BI, and FI samples, respectively. There were 101 OTUs shared across the

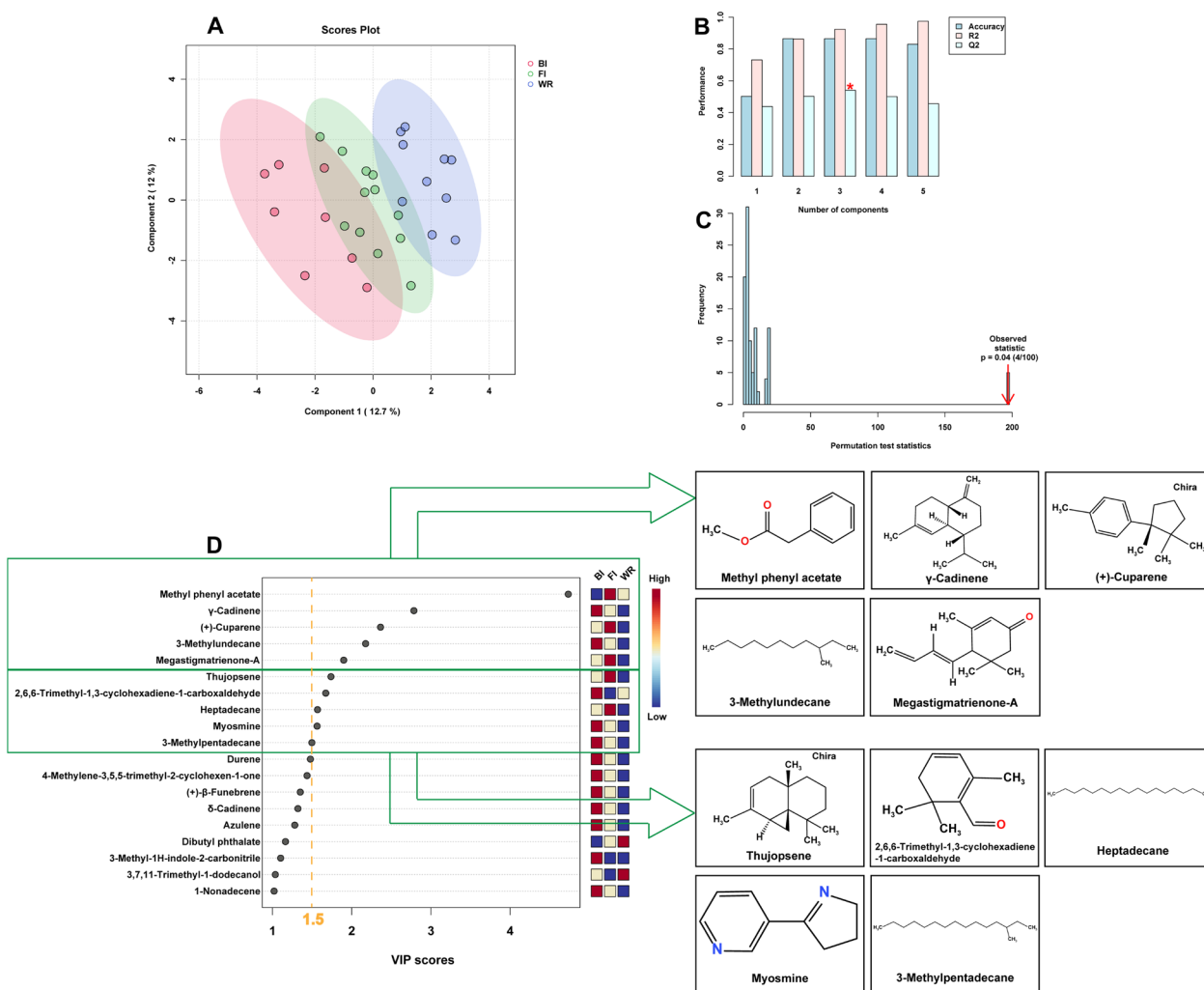


Fig. 3 **A** PLS-DA scores plot the volatile components in the CTLs from three different types. **B** Cross-validation; **C** Permutation test. **D** VIP scores plot of the volatile metabolites in the CTLs with VIP > 1.0 compounds in three different types

three types, with 34 unique to WR, 51 to BI, and 34 to FI (Additional file 1: Fig. S3G). However, PCoA based on Aitchison dissimilarity revealed that the bacterial and fungal community structures of WR, BI, and FI were highly similar and did not exhibit any significant separation characteristics (Additional file 1: Fig. S3D, H). This suggests a high degree of overlap in the microbial communities among the three types of CTLs.

In the bacterial community of CTLs, *Actinobacteriota* emerged as the most abundant phylum, representing over 45% of the total (Fig. 4A). At the genus level, *Corynebacterium* (54.30% in WR, 54.68% in BI, and 45.55% in FI) and *Pseudomonas* (18.30% in WR, 17.28% in BI, and 25.71% in FI) were the most prevalent across all groups (Fig. 4B). In the fungal community, the phyla *Ascomycota* and *Basidiomycota* dominated, collectively accounting for over 98% of the total abundance (Fig. 4C). *Aspergillus*

was the leading genus in all samples, comprising 45%–58% of the total. Notably, *Aspergillus* and *Cladosporium* together accounted for 57.50%, 44.91%, 57.91%, and 26.33%, 34.27%, and 22.47%, respectively, in the different CTL types (Fig. 4D). This analysis highlights the diversity and distinctiveness of the microbial communities present in the different types of CTLs.

Identification of core microbiota genera

LEfSe analysis was used to identify distinct microbiota in the three CTL types. At the genus level, *Sphingobacterium*, *Sediminibacterium*, *Flavisolibacter*, and *Parabacteroides* were identified in WR; *Massilia* and *Bacteroides* in BI; and *Pseudomonas*, *Mesorhizobium*, *Luteibacter*, *Bacillus*, and *Porphyromonas* in FI (Fig. 5A, B, Additional file 1: Table S9). For fungi, biomarkers at the genus level included *Alternaria* and *Candida* in WR,

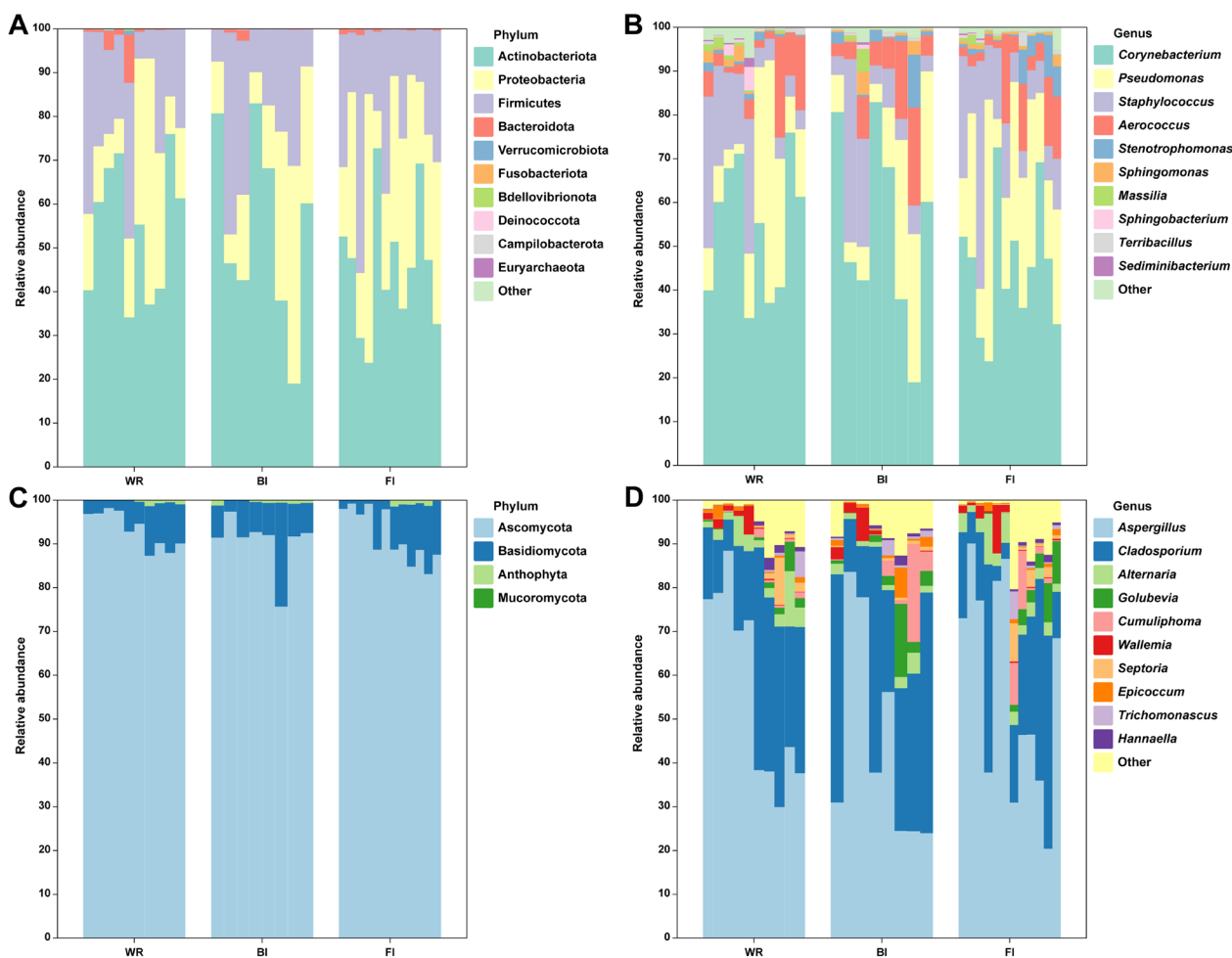


Fig. 4 Analysis of microbial community composition in CTLs from different types. The relative abundance of bacterial taxon at **A** phylum and **C** genus level, respectively. The relative abundance of a fungal taxon at **B** phylum and **D** genus level, respectively

and *Trichomonascus*, *Penicillium*, and *Eichleriella* in BI (Fig. 5C, D). The flavor substances were considerably influenced by starch as well as the combination of other chemical components, as demonstrated in Fig. 6 and suggested by the canonical correspondence analysis (CCA) (Fig. 6A). We discovered that starch and H were important variables in bacterial community makeup (Fig. 6B). While starch, total sugar, and H were important factors determining the makeup of fungal communities (Fig. 6C). OTU000044 (*Stenotrophomonas*) and OTU000037 (*Terribacillus*) were positively correlated with starch, whereas OTU000019 (*Staphylococcus*) was negatively correlated with starch. Additionally, OTU000019 showed a positive relationship with H and OTU000037 showed a negative relationship with H (Fig. 6B). In contrast, many OTUs were positively correlated with starch and total sugar coupled with OTU000001 (*Aspergillus*), OTU000021 (*Aspergillus*) and OTU000023 (*Wallemia*) were negatively correlated with starch and total sugar. Besides,

OTU000001, OTU000021, and OTU000023 showed a positive relationship with H, and other OTUs were negatively correlated with H. Starch and total sugar exhibited an opposite correlation trend with H. (Fig. 6C). We built an O2PLS model to study the link between bacteria and flavor metabolites. We discovered five bacterial OTUs that significantly influenced the flavor profile, namely OTU000037, OTU000164 (*Brevundimonas*), OTU000007 (*Corynebacterium*), OTU000044, and OTU000080 (*Massilia*) (Fig. 6E), which were consistent with the findings from the genus-level analysis (Fig. 4B). Myosmine, 3,7,11-Trimethyl-1-dodecanol, Heptadecane, 2,6,10,14-Tetramethylpentadecane-Norphytane, and Durene were the flavor compounds most impacted by the bacterial microbiota (Fig. 6D). The analysis of relative abundance further demonstrated that *Corynebacterium* held a relatively high dominant position in all the samples (Fig. 6F). Additionally, we discovered that the fungal genera OTU000052 (*Golubevia*), OTU000105

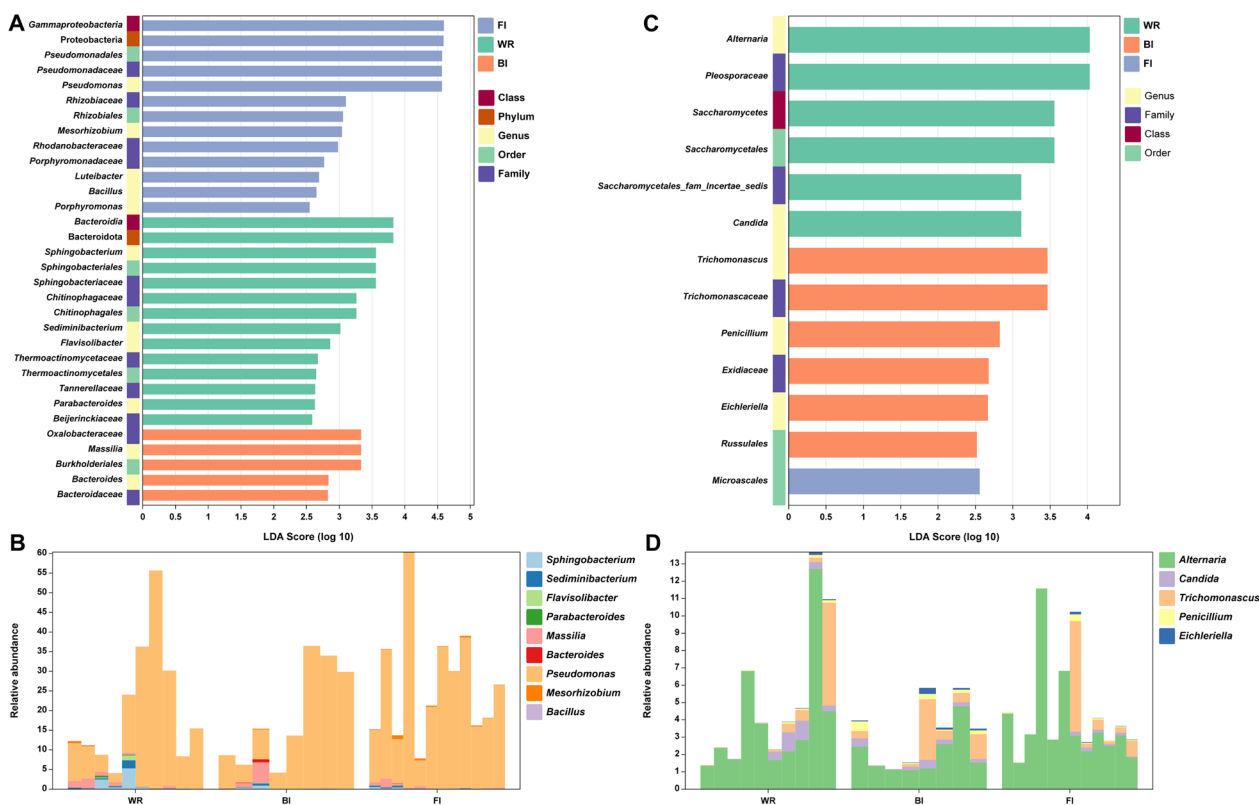


Fig. 5 Lefse Analysis of microbial community differences in different CTLs. **A** The bacterial biomarker of different CTLs. **B** Relative abundance at the genus level of bacteria screened in WR, BI, and FI. **C** The fungal biomarker of different CTLs. **D** Relative abundance of fungal genus levels screened in WR, BI, and FI

(*Bulleromyces*), OTU000021, OTU000041 (*Aspergillus*) and OTU000117 (*Candida*) were highly linked to the flavor components (Fig. 6H), particularly Tridecane, Thujopene, (+)-Cuparene, Cedrol and Octadecane (Fig. 6G). Notably, the dominant taxon *Aspergillus* occupied a dominant position in samples of different types of CTLs (Fig. 6I).

In fermented CTLs, the network of prokaryotic and eukaryotic communities was divided into four separate modules (Fig. 7A): Module 1, Module 2, Module 3, and Module 4. Module 1 included 13 OTUs, Module 2 had 9 OTUs, Module 3 included 9 OTUs, and Module 4 had 2 OTUs (Additional file 1: Table S10). It was discovered by computing the Zi and Pi that OTU000027 (Module1, *Cumuliphoma*), OTU000043 (Module1, *Golubevia*), OTU000016 (Module1, *Septoria*), OTU000054 (Module1, *Nicotiana*), OTU000073 (Module1, *Hannaella*), OTU000099 (Module1, *Sarocladium*), OTU000112 (Module1, *Hannaella*), OTU000001 (Module2, *Aspergillus*), OTU000014 (Module2, *Aspergillus*), OTU000023 (Module2, *Wallemia*), OTU000071 (Module2, *Trichomonascus*), OTU000053 (Module2, *Aspergillus*), OTU000019 (Module3, *Staphylococcus*), OTU000080 (Module3, *Massilia*),

OTU000097 (Module3, *Lepista*), OTU000117 (Module3, *Candida*), OTU000098 (Module3, *Phaeosphaeria*) and OTU000111 (Module3, *Schizophyllum*) functioned as network connections (Fig. 7B, C). These microbial genera were believed to play an important role in prokaryotic community interactions and evolution. Notably, among the 15 identified microbial genera, 13 belong to fungi, with only 2 being bacterial genera. This showed that fungi may be more important in the microbial network than bacteria (Fig. 7D–F). The number and composition of the various modules differed significantly. When it came to abundance, the connectors in Module 1 displayed a low abundance in some CTL samples of WR, BI, and FI, interestingly, all these samples originated from the same planting area (DHMS). Module 2 exhibited a low abundance in certain CTL samples of WR, BI, and FI, and these samples all came from another planting area (PEJC). In contrast, the total abundance of Module 2 was found to be high across all samples. In Module 3, bacterial genera were the dominant microorganisms in CTLs, but the relative abundance of fungal microorganisms was lower. Overall, *Staphylococcus* and *Massilia* were consistently prevalent among the bacterial taxa (Fig. 7F). Notably, the relative abundance

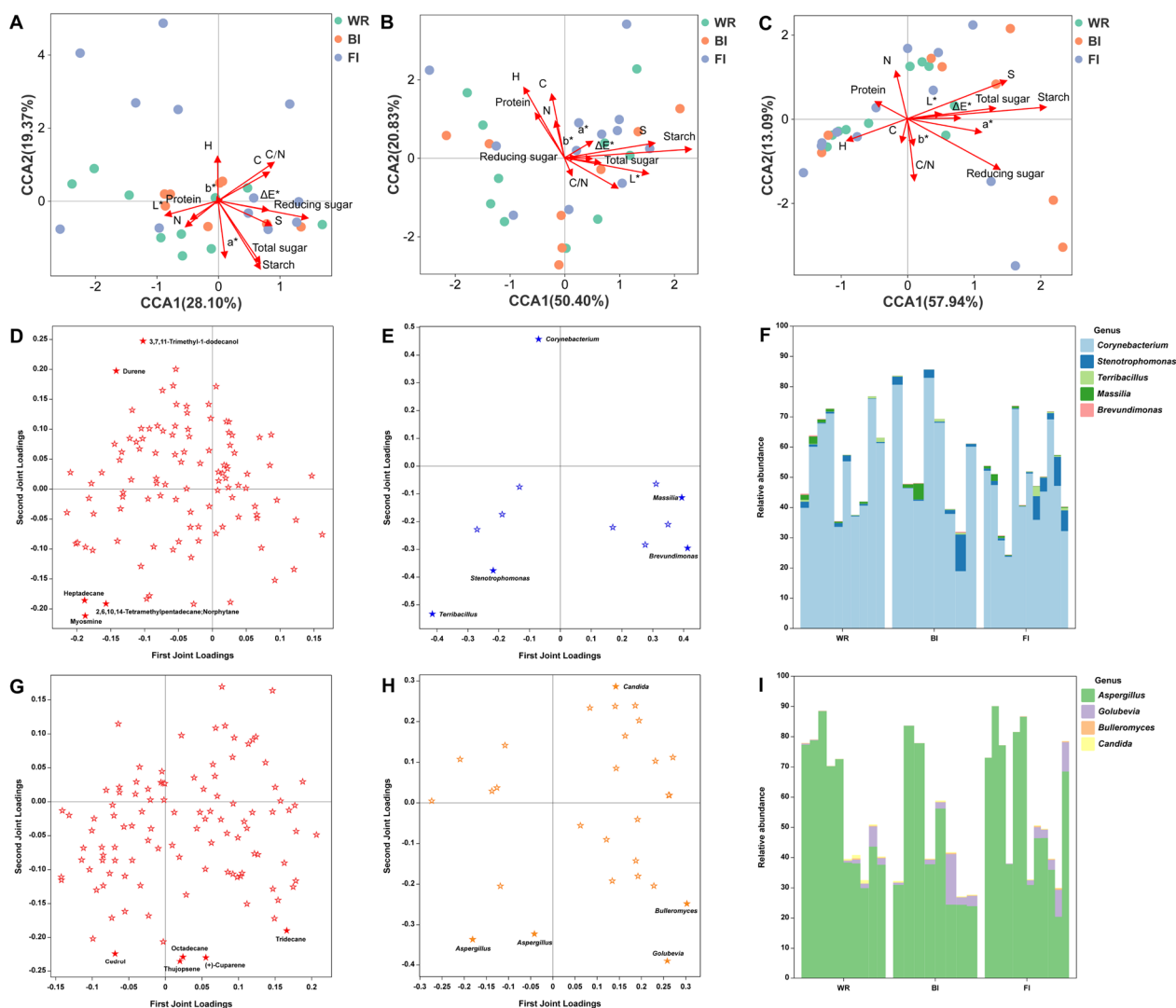


Fig. 6 Correlation analysis of physicochemical profiles, microbial communities, and flavor compounds in CTLs from different types. **A** CCA analysis between flavor compounds and physicochemical profiles. **B** CCA analysis between bacterial OTU and chemical components. **C** CCA analysis between fungal OTU and chemical components. **D** O2PLS analysis between bacteria OTU and flavor compounds. **E** O2PLS analysis between bacteria OTU and flavor compounds. **F** O2PLS screening of bacterial genus abundance highly correlated with metabolome. **G** O2PLS analysis between fungi OTU and flavor compounds. **H** O2PLS analysis between fungi OTU and flavor compounds. **I** O2PLS screening of fungal genus abundance highly correlated with metabolome

of *Staphylococcus* was generally higher in WR, BI, and FI samples from DHMS compared to those from PEJ/C. In WR, BI, and FI, the relative abundance of *Staphylococcus* was highest in FI (15.21%), followed by WR (14.51%) and BI (12.36%). Conversely, *Massilia* showed an opposite pattern, with its relative abundance being highest in BI (0.82%), followed by WR (0.72%), and lowest in FI (0.44%). Among the fungal genera (Fig. 7D, E), all connectors in Module 1 and Module 2 were composed entirely of fungal genera, with *Aspergillus*, *Golubevia*, *Cumuliphoma*, *Wallemia*, *Septoria*, *Trichomonascus*, and *Hannaella* dominating in all samples,

consistent with the dominant genera identified in Fig. 4D. Therefore, the fungal genera in Module 1 and Module 2 may bear greater responsibility for the stability and characteristic changes of the microbial interaction network and are likely key core species of significance in the fermentation of CTLs.

Association among chemical components, core microbiota genera, and VOCs in CTLs

Following LefSe analysis, O2PLS screening, and the calculation of Z_i and P_i values, 8 bacterial genera and 9

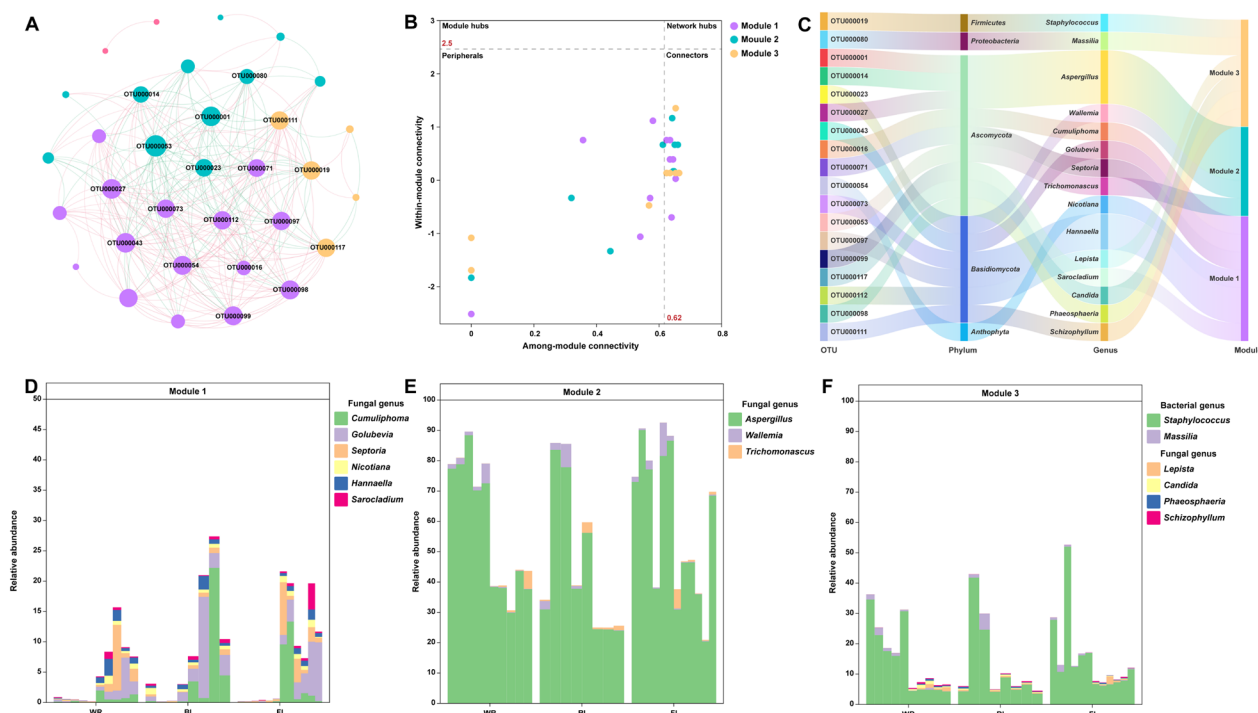


Fig. 7 **A** Network observed between the 33 OTUs (Spearman correlation coefficient $|r| > 0.6$, $p < 0.05$). **B** Distribution of OTUs based on Z_i and P_i . **C** Taxonomic annotation levels (phylum level and genus level) of OTUs in the Connectors region and the modules they belong to. **D** Composition of bacterial and fungal communities in Module 1 in the Connectors region at the genus level. **E** Composition of bacterial and fungal communities in Module 2 in the Connectors region at the genus level. **F** Composition of bacterial and fungal communities in Module 3 in the Connectors region at the genus level

fungal genera were identified as dominant marker core genera. These were then correlated with the chemical components' properties using the Mantel test (Additional file 1: Table S11). The heat map demonstrated that the protein and starch were the most interconnected chemical components (Fig. 8A, B). Additionally, the Mantel test revealed more highly significant physico-chemical influences on fungal microorganisms (Fig. 8B). Surprisingly, N, C, C/N, and reducing sugar did not affect any of the microbial strains. To further clarify the impact of chemical components on microbial diversity and genus composition, PLS-SEM was performed. This analysis investigated the direct and indirect correlations between chemical components, fungal diversity, fungal genus, and flavor metabolites (Fig. 8C, D). After multiple iterations of the algorithm, it was determined that there were no significant path relationships between N, C, C/N, reducing sugar, and total sugar with bacterial diversity and bacterial genus. Therefore, bacterial diversity and bacterial genus were excluded from the analysis to enhance the model's rationality. The main parameters used in PLS-SEM indicated that the final model was reasonable, with predictions closely aligning with actual results (SRMR=0.052, 0.061; d_{ULS} =0.150, 0.393; and

d_G =0.599, 0.337). Starch exerted the strongest positive and significant effect ($p < 0.05$) on both fungal diversity and fungal genus. Protein had a significant negative impact on fungal diversity ($p < 0.05$) but a positive, albeit non-significant, effect on the fungal genus. The Ace, Chao, Shannon, and Sobs indices could explain most of the variation in fungal diversity, with R^2 values being 0.979, 0.988, 0.913, and 0.990, respectively. Among the selected 9 fungal genera, *Aspergillus*, *Candida*, and *Hannaella* could explain most of the variation in the fungal genus, with R^2 values of -0.851, 0.855, and 0.922, respectively. Additionally, fungal diversity and fungal genus had a direct impact on flavor metabolites. Fungal diversity showed a significant positive effect on alkenes and n-heterocyclic carbenes, while fungal genus had a significant positive effect on alcohols and n-heterocyclic carbenes. Therefore, overall, starch is the most influential chemical component. It positively affects the abundance and diversity of fungal microbial genera, which in turn directly impacts flavor metabolites. Thus, in practical production, altering the starch content in cigar fermentation CTLs might be a strategy to regulate microbial community changes and consequently alter the flavor quality characteristics of CTLs.

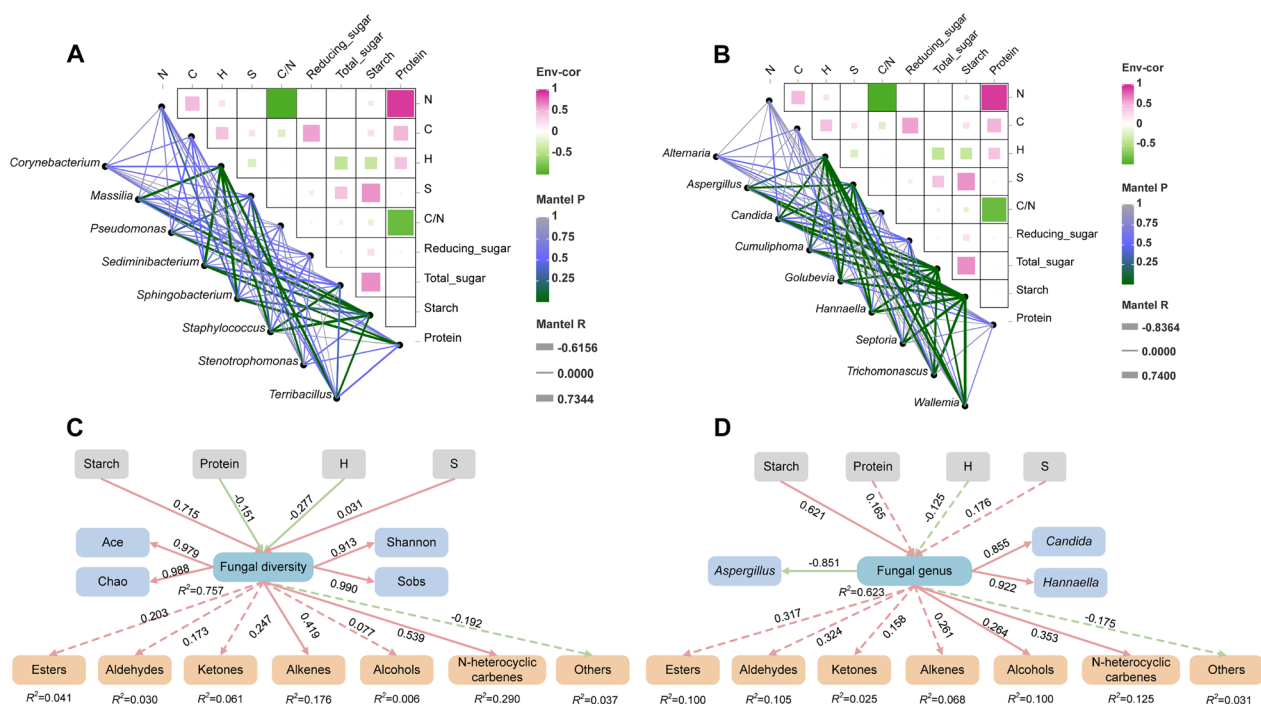


Fig. 8 **A** Mantel test between bacterial communities and chemical components. **B** Mantel test between fungal communities and chemical components. (Bacteria and fungi as core microorganisms after screening). **C** PLS-SEM in chemical components, microorganisms, and metabolites based on the results of fungal diversity. **D** PLS-SEM in chemical components, microorganism, and metabolites based on the results of key microorganism genus of fungi. Specific values for the Mantel test were given in the supplementary material (Additional file 1: Table S11), and the results of bacterial diversity and genus could not construct a valid model due to the low number of significant chemical components

Discussion

Chemical components impact on CTLs quality

The quality of CTLs generally includes aspects like appearance, chemical components, and internal quality [7]. The internal quality mainly refers to the harmony in the content and proportions of various chemical components within the leaves, with a direct correlation existing between the leaf quality and the content and proportion of these chemical constituents. The results of this study indicate that the average content of conventional chemical components in the samples of WR, BI, and FI is within the appropriate range for each component, an overall harmonious chemical composition (Additional file 1: Fig. S2). These findings suggest that despite some variations among WR, BI, and FI post-uniform fermentation, the overall degree of differentiation is minimal, indicating a trend toward homogenization. This further underscores the pivotal role of fermentation in standardizing the quality of CTLs. Nitrogen compounds and carbohydrates are among the most crucial conventional chemical components in WR, BI, and FI, and their content significantly impacts the overall quality of the leaves. Generally, the pyrolysis products of carbohydrates and nitrogen compounds while burning CTLs have opposite effects on the taste experience. The pyrolysis products of

carbohydrates are acidic, while those of nitrogen compounds (especially alkaloids) are alkaline. The harmonious proportional relationship between these two types of compounds produces a desirable flavor when smoked [2, 29].

From the average content of the tested components, the N contents in the WR were significantly higher than in the BI and FI. There were no significant differences in other chemical components between WR, BI, and FI. Moreover, based on the selection criteria of VIP > 1 and TAV > 1, Glu and Asp are important taste-flavor amino acids distinguishing WR, BI, and FI (Fig. 1D). Tobacco leaf-free amino acid concentration is closely connected to leaf quality. They are precursors for protein and nicotine synthesis and participate in enzymatic and non-enzymatic browning reactions with reducing sugars (or carbonyl compounds) during tobacco processing, fermentation, and even burning. These reactions produce various heterocyclic compounds with roasted or popcorn-like aromas, such as pyrans, pyrazines, pyrroles, and pyridines. Some amino acids, like phenylalanine, can also decompose into aromatic compounds like benzyl alcohol and phenylethanol [30]. In CTLs, Glu provides a unique umami taste, crucial for balancing and enriching the overall flavor characteristics of cigars. Asp

significantly affects the sweetness and ash cohesiveness of CTLs. Studies have shown a positive correlation between ASP and irritancy and off-flavor scores in cigarettes [31]. The impact of FAAs on tobacco quality is not the effect of a single amino acid but the result of interactions among various FAAs.

Flavor compound variations in different types of CTLs

The composition and concentration of aromatic chemicals influence the sensory qualities of CTLs. Currently, E-nose are used to detect odors or flavors by mimicking human olfaction. They are widely applied in various fields including food, beverages, environmental monitoring, and medical diagnostics. In the analysis of CTLs, E-nose can rapidly and non-destructively evaluate the aroma characteristics of different tobacco products. Through E-nose analysis, it was found that WR, BI, and FI samples showed higher response values at sensor W5S, indicating that nitrogen oxides have a significant presence in CTLs. Moreover, the signal intensity in WR was notably higher than in BI and FI (Fig. 2C), suggesting distinct differences in the aromatic profile among these components of WR, BI, and FI. Although E-nose provides rapid aroma analysis, they cannot offer detailed structural information about compounds and have relatively lower sensitivity and specificity. Therefore, we further employed HS-SPME-GC-MS to measure the VOCs in WR, BI, and FI. A total of 90 VOCs were identified, among which 10 VOCs had a VIP > 1.5. These 10 VOCs were identified as key VOCs critical in differentiating the flavor characteristics of WR, BI, and FI (Fig. 3D). The phenylalanine breakdown products, such as benzyl alcohol, phenyl ethyl alcohol, benzaldehyde, and phenylacetaldehyde, had a floral odor of a rose, a strong almond odor, and a honey-like odor [32]. Phenylacetaldehyde can be further converted into phenylacetic acid, which when reacted with methanol in an esterification process, forms phenylacetic acid methyl ester. Phenylacetic acid methyl ester significantly contributes to the aroma of CTLs, providing a sweet floral fragrance that enriches the overall flavor profile of cigars, adding a specific floral and fruity note [33]. γ -Cadinene, a type of monoterpenoid, and (+)-Cuparene, a sesquiterpenoid, are produced in plants through the terpenoid synthesis pathway involving isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) [34]. These compounds are characterized by their distinctive woody and citrus scents, respectively [35]. Megastigmatrienone-A, a naturally occurring substance in CTLs, is an essential component of tobacco aroma and plays a primary role in fragrance. It significantly enhances the tobacco scent, improves the taste experience, harmonizes the smoke, and reduces irritation. Megastigmatrienone-A is a degradation product

of carotenoids, formed from the breakdown of lutein, and contributes significantly to the enhancement of tobacco aroma, modulation of flavor, and removal of off-flavors [36]. Myosmine is a natural alkaline compound found in tobacco and other plants, belonging to the nicotine class of compounds. It, along with nicotine, nornicotine, anatabine, and anabasine, influences the overall chemical and sensory properties of CTLs [37]. According to the results of PLS-DA and VIP analysis, after fermentation, aromatic components in WR, BI, and FI can be somewhat distinguished. Methyl phenyl acetate (VIP = 4.73), γ -Cadinene (VIP = 2.78), and (+)-Cuparene (VIP = 2.36) are key VOCs that differentiate WR, BI, and FI (Additional file 1: Table S8). CTLs are known to contain a range of aromatic compounds, and the degradation of these aromatics may result in the production of new aromatic compounds. Notably, in contrast to flue-cured tobacco, which mostly comprises neutral flavor compounds, cigars are distinguished by the predominance of alkaline flavor substances. As a result, additional research is required to thoroughly evaluate the properties of alkaline aromatic components in WR, BI, and FI.

Impact of chemical components and core microbial genera on flavor characteristics

It is noteworthy that the top 10 bacterial and fungal genera in terms of relative abundance are common across WR, BI, and FI, indicating that these dominant genera are better adapted to the fermentation environment and occupy a significant position in WR, BI, and FI (Fig. 4B–D). Previous research has shown that during the fermentation process of CTLs, it is primarily the bacterial communities that influence the generation of flavor compounds [5]. However, whether bacteria or fungi play a more significant role in shaping the distinct flavor characteristics of WR, BI, and FI has not yet been reported. As a consequence, OTUs from bacteria and fungi in WR, BI, and FI that appeared in more than 30% of the samples and had a relative abundance larger than 0.1% were chosen for this investigation. These OTUs were chosen as the primary microbiota and were designated as independent variables (x), while 90 flavor compounds were designated as dependent variables (y). The O2PLS model was then used to investigate the relationships between bacterial and fungal communities and flavor chemicals in a variety of samples. We identified 5 bacterial and 4 fungal genera that were significantly correlated with flavor compounds (Fig. 6E–H). These microbial genera had been detected in CTLs before, but their functions in fermented CTLs had not been previously established.

Moreover, it has been claimed that these genera are important in the synthesis of flavor components in tobacco. As an example, *Sphingobacterium* can produce

microbial lignin oxidase and is known for lignin degradation [38]. In a study examining the impact of flue-curing techniques on the dynamic changes in microbial diversity of CTLs, Hu et al. found that the trends in *Sediminibacterium* were significantly correlated with changes in flue-curing processes [39]. *Flavisolibacter* is a dominant genus of bacteria in the rhizosphere soil of tobacco plants, while *Alternaria* is a very common microbial genus in tobacco, although it may cause brown spot disease in the crop [40]. *Candida* plays an important role in the fermentation of cigars, increasing the content of chlorophyll degradation products and carotenoid degradation products, enhancing the roasted, nutty, cocoa, and honey flavors of cigars, and improving the flavor components of CTLs [41]. *Massilia* and *Bacteroides* are also ubiquitous in CTLs, with *Massilia* identified as a primary predictive taxon in the bacterial composition of CTLs after re-drying in different regions [42]. Furthermore, fungi such as *Aspergillus*, *Penicillium*, *Fusarium*, *Cladosporium*, and *Trichomonascus* have been identified as indicator microorganisms for microbial community succession during the pile fermentation of CTLs [43]. *Pseudomonas*, as a main genus of bacteria engaged in nicotine breakdown, could sustain high nicotine concentrations and use nicotine as the sole carbon and nitrogen source for growth [44]. *Bacillus species* can synthesize tiny aromatic compounds by digesting big molecules like carotene. After 2 days of co-culture of *Bacillus Amylolytic* and *Bacillus Kochi*, the levels of the most important reaction products and terpene metabolites were reported to be greater than those of other samples, which boosted the scent and softness of the samples [45]. Studies have indicated that WR, BI, and FI exhibit similar dominant microbial compositions during the pile fermentation process. Through their interactions, key microorganisms such as *Staphylococcus*, *Corynebacterium*, *Aerococcus*, and *Aspergillus* have the most influence on the microbial community structure and distinctive bacteria in CTLs. These interactions influence the transformation of volatile flavor compounds [46]. As a result, in CTLs, bacteria, and fungi work together to degrade carbohydrates and nitrogen molecules, as well as to chemosynthesis VOCs throughout the fermentation process.

Microorganisms often live in intricate communities rather than in isolation, forming tight relationships with one another [47]. Microbial co-occurrence networks are commonly utilized to investigate microbial community connections. Key species can be found by evaluating the topological properties of species within these networks [48]. Co-occurrence networks were used in this investigation to analyze potential connections between microorganisms in WR, BI, and FI. Significant connections between bacterial and fungal nodes were found in these

networks ($p < 0.05$, $|r| > 0.6$, Additional file 1: Table S3), demonstrating that microorganisms work together to adapt to the phyllosphere and sustain community structure [49]. Nodes in the network were separated into four modules and classified as peripherals, connectors, module hubs, and network hubs based on Zi and Pi values. Within microbial co-occurrence networks, nodes categorized as connectors, module hubs, and network hubs are considered keystone nodes that contribute to the sustainability and stability of the ecosystem and are important for the bacterial community assembly and function; the removal of this keystone may have a significant impact on the community structure [50, 51]. In all samples, a total of 18 nodes (54.55%) were identified as connectors. Among these module hubs, there were 13 fungal genera and 2 bacterial genera, with high-abundance dominant genera such as *Aspergillus*, *Golubevia*, *Septoria*, *Cumuliphoma*, *Wallemia*, *Trichomonascus*, and *Hannaella* occupying central positions in the modules (Fig. 7). This suggests that fungi may have a stronger role than bacteria in shaping the microbial populations of WR, BI, and FI. The diversity of fungal species is critical for shaping the ecological and functional stability of WR, BI, and FIs' microbial communities. This emphasizes the significance of taking into account the diversity and dominant species when researching microbial communities in CTLs. Observations similar to these have been made in the fermentation ecosystems of Chinese strong-flavored Baijiu [11].

Structural Equation Modeling (SEM) is a quantitative research tool for dealing with multifactor causal interactions that is based on statistical analysis techniques [52]. By connecting empirical data and theoretical analysis, the relationships between multiple causes and outcomes can be established utilizing integrated path analysis, factor analysis, regression analysis, and analysis of variance. SEM is employed in the estimation of latent variables and the construction of a complex variable prediction model [53]. This method, in addition to the results of a normal multivariate statistical analysis, provides a better knowledge of the elements' direct and indirect relationships [54, 55]. Considering the complex interactions between microbial communities and chemical components that determine the flavor quality of WR, BI, and FI, this study employed an SEM model to investigate the composite factors affecting microbial communities and the main genera impacting VOCs. The diversity of fungi and fungal genera such as *Aspergillus*, *Candida*, and *Hannaella* significantly affect nitrogenous compounds in CTLs. Nitrogenous compounds, like nicotine and other alkaloids, are produced through various metabolic pathways in CTLs and play a crucial role in regulating their flavor, aroma, and burning characteristics. Starch was identified

as a key chemical component significantly influencing fungal microbial diversity and genera ($p < 0.05$, $R^2 > 0.6$) and is an important factor affecting the flavor quality of WR, BI, and FI (Fig. 8C, D). Starch, a primary metabolite in tobacco, was converted into water-soluble carbohydrates and ultimately into aromatic compounds, according to Banozic et al. [56]. During the baking process, the Maillard reaction between amino acids released by protein hydrolysis and sugar formed by starch hydrolysis is the primary source of an aroma precursor in flue-cured tobacco.

Conclusion

In summary, our findings extensively evaluated the traditional chemical components, flavor compounds, and microbial community architectures of WR, BI, and FI and investigated their interrelationships. The findings conclude that the conventional chemical components in WR, BI, and FI show a trend of homogenization, with Asp and Glu being the main taste-contributing amino acids in CTLs. Compared to fungal microbiota, the bacterial community structures in WR, BI, and FI are more similar. Additionally, five bacterial genera and four fungal genera are significantly related to flavor compounds, with seven fungal genera identified as functional microorganisms that may play an essential role in maintaining the sustainability and stability of the tobacco phyllosphere ecosystem. The starch of WR, BI, and FI significantly positively influences fungal microbes such as *Aspergillus*, *Candida*, and *Hannaella*, thereby indirectly affecting the formation of nitrogenous flavor compounds. In subsequent practical production, it may be possible to regulate the changes in the fungal microbial communities by altering the starch content in fermented cigar tobacco leaves, thereby modifying the flavor quality characteristics of WR, BI, and FI. These findings improve our understanding of the properties of WR, BI, and FI and give a sound theoretical underpinning for increasing CTL quality.

Abbreviations

CTLs	Cigar Tobacco Leaves
O2PLS	Bidirectional orthogonal partial least squares
PLS-SEM	Partial Least Squares Structural Equation Modeling
Zi	Within-module
Pi	Among-module
DHMSYunnan	Mang City, Dehong Prefecture
PEJC	Hani-Yi Autonomous County of Jiangcheng, Pu'er City
VOCs	Volatile Organic Compounds
HS-SPME-GC-MS	Headspace Solid-Phase Microextraction coupled with Gas Chromatography-Mass Spectrometry
OTUs	Operational Taxonomic Units
PCoA	principal coordinates analysis
TAV	Taste Active Value
PLS-DA	Partial Least-Squares Discriminant Analysis
CCA	Canonical Correspondence Analysis
VIP	Variable Importance in Projection
PCA	Principal Component Analysis

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40538-024-00582-0>.

Additional file 1. Additional files and tables.

Author contributions

MZ: investigation, writing. DG: writing review and editing. HW: resources. GW: investigation. XW and YS: software. JZ and TZ: investigation and formal analysis. EZ and XL: conceptualization, methodology, and writing-review and editing. All authors reviewed the manuscript.

Funding

This research was funded by the State-Owned Enterprises Cooperative External Program, a grant number W2023JSKF0083.

Availability of data materials

The data described in this study can be obtained from the corresponding author upon request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

This research has been confirmed for publication in the journal.

Competing interests

The authors have no competing interests.

Author details

¹Key Laboratory for Agricultural Products Processing, School of Food and Biological Engineering, Hefei University of Technology, Danxia Road 485#, Zip Code 230601, Hefei, Anhui, China. ²China Tobacco Anhui Industrial Co., Ltd, Huangshan Road 606#, Zip Code 230088, Hefei, Anhui, China. ³College of Engineering, Peking University, Beijing, China.

Received: 17 January 2024 Accepted: 8 April 2024

Published online: 08 May 2024

References

- Tao J, Chen Q, Chen S, Lu P, Chen Y, Jin J, et al. Metagenomic insight into the microbial degradation of organic compounds in fermented plant leaves. *Environ Res.* 2022;214: 113902.
- Zhang Q, Kong G, Zhao G, Liu J, Jin H, Li Z, et al. Microbial and enzymatic changes in cigar tobacco leaves during air-curing and fermentation. *Appl Microbiol Biotechnol.* 2023;107:5789–801.
- Ning Y, Zhang L-Y, Mai J, Su J-E, Cai J-Y, Chen Y, et al. Tobacco microbial screening and application in improving the quality of tobacco in different physical states. *Bioresour Bioprocess.* 2023;10:32.
- Dai J, Dong A, Xiong G, Liu Y, Hossain MS, Liu S, et al. Production of highly active extracellular amylase and cellulase from *Bacillus subtilis* ZIM3 and a recombinant strain with a potential application in tobacco fermentation. *Front in Microbiol.* 2020;11:1539.
- Zheng T, Zhang Q, Li P, Wu X, Liu Y, Yang Z, et al. Analysis of microbial community, volatile flavor compounds, and flavor of cigar tobacco leaves from different regions. *Front Microbiol.* 2022;13: 907270.
- Hu W, Cai W, Zheng Z, Liu Y, Luo C, Xue F, et al. Study on the chemical compositions and microbial communities of cigar tobacco leaves fermented with exogenous additive. *Sci Rep.* 2022;12:19182.
- Shi Y, Pan Y, Du F, Zhao Z, Li Z, Wang Y, et al. Identification and discrimination of characteristic aroma components of different cigar leaves based on static headspace/gas chromatography-Ion mobility spectrometry

- combined with relative odor activity value and multivariate statistical analysis. *J Instrum Anal.* 2023;42:674–83.
8. Fan J, Kong G, Yao H, Wu Y, Zhao G, Li F, et al. Widely targeted metabolomic analysis reveals that volatile metabolites in cigar tobacco leaves dynamically change during fermentation. *Biochem Biophys Rep.* 2023;35:101532.
 9. An F, Wu J, Feng Y, Pan G, Ma Y, Jiang J, et al. A systematic review on the flavor of soy-based fermented foods: Core fermentation microbiome, multisensory flavor substances, key enzymes, and metabolic pathways. *Compr Rev Food Sci F.* 2023;22:2773–801.
 10. Guan Q, Huang T, Peng F, Huang J, Liu Z, Peng Z, et al. The microbial succession and their correlation with the dynamics of flavor compounds involved in the natural fermentation of suansun, a traditional Chinese fermented bamboo shoots. *Food Res Int.* 2022;157: 111216.
 11. Zhou H, Xu S, Xu B, Jiang C, Zhao E, Xu Q, et al. Effect of *Caproicbacterium lactatifermentans* inoculation on the microbial succession and flavor formation of pit mud used in Chinese Baijiu fermentation. *Food Res Int.* 2024;175: 113730.
 12. Wang C, Ma L, Zhang Y, Chen N, Wang W. Spatiotemporal dynamics of wetlands and their driving factors based on PLS-SEM: A case study in Wuhan. *Sci Total Environ.* 2022;806: 151310.
 13. Choe J-H, Kim H-Y, Lee J-M, Kim Y-J, Kim C-J. Quality of frankfurter-type sausages with added pig skin and wheat fiber mixture as fat replacers. *Meat Sci.* 2013;93:849–54.
 14. Gu Z, Jin Z, Schwarz P, Rao J, Chen B. Unraveling the role of germination days on the aroma variations of roasted barley malts via gas chromatography-mass spectrometry based untargeted and targeted flavoromics. *Food Chem.* 2023;426: 136563.
 15. Chen Z, Feng Y, Cui C, Zhao H, Zhao M. Effects of *koji*-making with mixed strains on physicochemical and sensory properties of Chinese-type soy sauce. *J Sci Food Agric.* 2015;95:2145–54.
 16. Zheng T, Zhang Q, Wu Q, Li D, Wu X, Li P, et al. Effects of inoculation with *Acinetobacter* on fermentation of cigar tobacco leaves. *Front Microbiol.* 2022;13: 911791.
 17. Guo M, Wu F, Hao G, Qi Q, Li R, Li N, et al. *Bacillus subtilis* improves immunity and disease resistance in rabbits. *Front Immunol.* 2017;8:354.
 18. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics.* 2011;27:2957–63.
 19. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, et al. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat Methods.* 2013;10:57–9.
 20. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods.* 2010;7:335–6.
 21. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods.* 2013;10:996–8.
 22. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics.* 2011;27:2194–200.
 23. Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci.* 2003;14:927–30.
 24. Shi Y, Delgado-Baquerizo M, Li Y, Yang Y, Zhu Y-G, Peñuelas J, et al. Abundance of kinless hubs within soil microbial networks are associated with high functional potential in agricultural ecosystems. *Environ Int.* 2020;142: 105869.
 25. Bouhaddani SE, Uh H-W, Jongbloed G, Hayward C, Klarić L, Kielbasa SM, et al. Integrating omics datasets with the OmicsPLS package. *BMC Bioinformatics.* 2018;19:371.
 26. Sugrue D. Introducing article numbering to computational statistics and data analysis. *Comput Stat Data An.* 2020;142: 106857.
 27. Wagg C, Bender SF, Widmer F, Van Der Heijden MGA. Soil biodiversity and soil community composition determine ecosystem multifunctionality. *Proc Natl Acad Sci USA.* 2014;111:5266–70.
 28. Rönkkö M, Evermann J. A critical examination of common beliefs about partial least squares path modeling. *Organ Res Methods.* 2013;16:425–48.
 29. Macris BJ. Production and characterization of cellulase and β -glucosidase from a mutant of *Alternaria alternata*. *Appl Environ Microbiol.* 1984;47:560–5.
 30. Zhao T, Shi H, Ji X, Gu S, Wu C, Hou L, et al. Difference in composition and content of free amino acids among various tobacco types. *Acta Tabacaria Sinica.* 2011;17:13–7.
 31. Shi H, Liu G, Yang H, Ji X. Tobacco aromatology. China Agriculture Press. 2011.
 32. Wu H, Huang W, Chen Z, Chen Z, Shi J, Kong Q, et al. GC–MS-based metabolomic study reveals dynamic changes of chemical compositions during black tea processing. *Food Res Int.* 2019;120:330–8.
 33. Wahlberg I, Karlsson K, Austin DJ, Junker N, Roeraade J, Enzell CR, et al. Effects of flue-curing and ageing on the volatile, neutral and acidic constituents of Virginia tobacco. *Phytochemistry.* 1977;16:1217–31.
 34. Zachariah TJ, Leela NK. Volatiles from herbs and spices. Woodhead Publishing. 2006; 177–218.
 35. Ivanova T, Popova V, Atanasova T, Nenov N, Omar A. Chemical composition of aromatic products from tobacco (*N. tabacum L.*) 1. Virginia Flue-Cured. 2009.
 36. Qi D, Zhou Y, Wang J, Fei T, Wu D, Lu J. Determination of volatiles in flue-cured tobacco by gas chromatography–mass spectrometry (GC–MS) with chemometrics. *Anal Lett.* 2022;55:1398–411.
 37. Lisko JG, Stanfill SB, Duncan BW, Watson CH. Application of GC–MS/MS for the analysis of tobacco alkaloids in cigarette filler and various tobacco species. *Anal Chem.* 2013;85:3380–4.
 38. Rashid GMM, Taylor CR, Liu Y, Zhang X, Rea D, Fülöp V, et al. Identification of manganese superoxide dismutase from *Sphingobacterium sp. T2* as a novel bacterial enzyme for lignin oxidation. *ACS Chem Biol.* 2015;10:2286–94.
 39. Hu B, Gu K, Gong J, Zhang K, Chen D, He X, et al. The effect of flue-curing procedure on the dynamic change of microbial diversity of tobaccos. *Sci Rep.* 2021;11:5354.
 40. Chen Y-H, Lu M-H, Guo D-S, Zhai Y-Y, Miao D, Yue J-Y, et al. Antifungal effect of magnolol and honokiol from *magnolia officinalis* on *Alternaria alternata* causing tobacco brown spot. *Molecules.* 2019;24:2140.
 41. Jia Y, Zhou W, Yang Z, Zhou Q, Wang Y, Liu Y, et al. A critical assessment of the *Candida strains* isolated from cigar tobacco leaves. *Front Bioeng Biotechnol.* 2023;11:1201957.
 42. Zhang Y, Xu Q, Yang M, Yang Y, Fu J, Miao C, et al. Analysis of differences in tobacco leaf microbial communities after redrying in Chinese provinces and from abroad. *AMB Expr.* 2023;13:80.
 43. Zhang G, Zhao L, Li W, Yao H, Lu C, Zhao G, et al. Changes in physicochemical properties and microbial community succession during leaf stacking fermentation. *AMB Expr.* 2023;13:132.
 44. Li H, Li X, Duan Y, Zhang K-Q, Yang J. Biotransformation of nicotine by microorganism: the case of *Pseudomonas spp.* *Appl Microbiol Biotechnol.* 2010;86:11–7.
 45. Huang S, Liu D, Chen M, Xi G, Yang P, Jia C, et al. Effects of *Bacillus subtilis* subsp. on the microbial community and aroma components of flue-cured tobacco leaves based on metagenome analysis. *Arch Microbiol.* 2022;204:726.
 46. Wu Q, Peng Z, Pan Y, Liu L, Li L, Zhang J, et al. Interaction analysis of tobacco leaf microbial community structure and volatiles flavor compounds during cigar stacking fermentation. *Front Microbiol.* 2023;14:1168122.
 47. Hughes AR, Inouye BD, Johnson MTJ, Underwood N, Vellend M. Ecological consequences of genetic diversity. *Ecol Lett.* 2008;11:609–23.
 48. Proulx S, Promislow D, Phillips P. Network thinking in ecology and evolution. *Trends in Ecol Evol.* 2005;20:345–53.
 49. Jiang C, Liu Y, Li H, Zhu S, Sun X, Wu K, et al. The characterization of microbial communities and associations in karst tiankeng. *Front Microbiol.* 2022;13:1002198.
 50. Steele JA, Countway PD, Xia L, Vigil PD, Beman JM, Kim DY, et al. Marine bacterial, archaeal and protistan association networks reveal ecological linkages. *ISME J.* 2011;5:1414–25.
 51. Li Q, Li Y, Luo Y, Zhang Y, Chen Y, Lin H, et al. Shifts in diversity and function of the bacterial community during the manufacture of Fu brick tea. *Food Microbiol.* 2019;80:70–6.
 52. Hayes AF, Montoya AK, Rockwood NJ. The analysis of mechanisms and their contingencies: process versus structural equation modeling. *Australas Mark.* 2017;25:76–81.
 53. Cepeda-Carrion G, Cegarra-Navarro J-G, Cillo V. Tips to use partial least squares structural equation modelling (PLS-SEM) in knowledge management. *JKM.* 2019;23:67–89.
 54. Ren L, Li J, Li C, Dang P. Can ecotourism contribute to ecosystem? Evidence from local residents' ecological behaviors. *Sci Total Environ.* 2021;757: 143814.

55. Schweiger EW, Grace JB, Cooper D, Bobowski B, Britten M. Using structural equation modeling to link human activities to wetland ecological integrity. *Ecosphere*. 2016;7: e01548.
56. Banožić M, Jokić S, Ačkar Đ, Blažić M, Šubarić D. Carbohydrates-key players in tobacco aroma formation and quality determination. *Molecules*. 2020;25:1734.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.