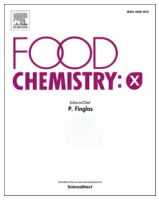
Food Chemistry: X 23 (2024) 101646

Contents lists available at ScienceDirect

Food Chemistry: X

journal homepage: www.sciencedirect.com/journal/food-chemistry-x



Identifying distinct markers in two Sorghum varieties for baijiu fermentation using untargeted metabolomics and molecular network approaches

Yulan Li [a], Yi Ma [b], Hui Zhu [b], Yin Liu [c], Shijiang Pan [d], Xi Chen [e], Tao Wu [a,*]

[a] School of Food and Biotechnology, Xihua University, No.9999 Guangchang Road, Chengdu 610039, China
[b] Sichuan Province Engineering Technology Research Center of Liquor-Making Grains, Sichuan University of Science and Engineering, Zigong 643000, China
[c] Sichuan Yibin Hengshengfu Liquor Industry Group Co., Ltd., Yibin 644100, China
[d] Sichuan Yibin Agriculture and Rural Affairs Bureau, Yibin 644100, China
[e] SCIEX Analytical Instrument Trading Co., Shanghai, 200335, China

ARTICLE INFO

Keywords:
Sorghum
Zaopei
UHPLC-Q-TOF-MS/MS
Sphingolipid
Methionine
Feature-based molecular network

ABSTRACT

The quality of strong-flavor Baijiu, a prominent Chinese liquor, is intricately tied to the choice of sorghum variety used in fermentation. However, a significant gap remains in our understanding of how glutinous and non-glutinous sorghum varieties comprehensively impact Baijiu flavor formation through fermentation metabolites. This study employed untargeted metabolomics combined with feature-based molecular networking (FBMN) to explore the unique metabolic characteristics of these two sorghum varieties during fermentation. FBMN analysis revealed 267 metabolites within both types of fermented sorghum (Zaopei) in the cellar. Further multidimensional statistical analyses highlighted sphingolipids, 2,5-diketopiperazines, and methionine derivatives as critical markers for quality control. These findings represent a significant advancement in our understanding and provide valuable insights for regulating the quality of Baijiu flavors.

1. Introduction

Baijiu, a Chinese distilled spirit comprising 52%–60% alcohol, ranks among the top six distilled liquors globally (Wei, Zou, Shen, & Yang, 2020). The dominance of strong-flavor Baijiu secures approximately 70% of the market share in China's liquor industry (He et al., 2020). Sorghum grains (*Sorghum bicolor* L. Moench) play a crucial role in the fermentation of strong-flavor Baijiu, contributing to increased alcohol yield and flavor enhancement (Liu et al., 2023). Sorghum varieties for Baijiu production are categorized into glutinous and non-glutinous types based on their amylopectin and amylose proportions (Chen et al., 2019; M.-K. Liu et al., 2023). Baijiu derived from glutinous sorghum is considered superior (Cao et al., 2020). However, the specific mechanisms underlying the superiority of glutinous sorghum remain uncertain.

Zaopei refers to the fermented sorghum grains in the cellar, where the core components of baijiu, including alcohols, aldehydes, and esters, are produced under microbial action (Hu et al., 2021). Research on non-volatile metabolites within Zaopei has been limited due to their complexity, with prior studies primarily focusing on volatile aromatic compounds (Li et al., 2020). However, gaining a comprehensive understanding of these non-volatile metabolites could significantly enhance our knowledge of the relationship between sorghum varieties and Baijiu quality.

Recently, untargeted liquid chromatography-mass spectrometry (LC-MS) metabolomics techniques have significantly advanced the field of food science (Bian et al., 2023; Han et al., 2023; Mei, Ding, & Chen, 2023; Fang, Liu, Xiao, Ma, & Huang, 2023), enabling systematic measurement of both known and unknown metabolites. For annotating known metabolites, the most commonly used approach involves searching the exact mass of precursor ions and/or tandem mass spectra against standard spectral libraries. However, due to limitations in database records, comprehensively understanding the structures of these unfamiliar compounds is challenging. To address these limitations, data-driven chemical structural mining methods such as Global Natural Products Social Molecular Networking (GNPS) have been established (Robin et al., 2021; Phelan, 2020). GNPS serves as a web-based mass spectrometry ecosystem, providing an open-access knowledge base for

* Corresponding author.
E-mail address: wutaobox@gmail.com (T. Wu).

https://doi.org/10.1016/j.fochx.2024.101646
Received 17 June 2024; Received in revised form 8 July 2024; Accepted 9 July 2024
Available online 14 July 2024

community-wide collaboration. Within GNPS, the feature-based molecular networking (FBMN) tool integrates isotopic patterns, sample retention times, and tandem mass spectrometry data using specific algorithms to conduct cluster analysis on raw data, effectively constructing a visualized molecular network. Compounds sharing structural similarities can form molecular families, significantly enhancing the discovery of unknown metabolites (Li et al., 2023).

Therefore, this research aims to broaden current knowledge by employing untargeted metabolomics techniques, specifically integrating molecular networking analysis. It focuses on a comprehensive examination of non-volatile metabolites in Zaopei derived from both glutinous and non-glutinous sorghum varieties. The principal objective is to establish a fundamental correlation between sorghum varieties and their characteristic metabolites, laying the groundwork for process control in the production of high-quality Baijiu.

2. Methods and materials

2.1. Experimental design and sample collection

The fermentation trial was conducted at Sichuan Yibin Hengshengfu Liquor Industry Group Co., Ltd. (Yibin, Sichuan, China) and involved two distinct sorghum types: glutinous sorghum (Chuanhongliang No. 9) and non-glutinous sorghum (Dongzajiao). Each sorghum variant underwent a 75-day fermentation process in separate fermentation cellars. Due to the spatial heterogeneity of the samples, sampling was conducted from the upper, middle, and lower layers of the fermentation cellars, with four replicates per layer. Specifically, the upper layer samples were taken from the exposed part of the cellar, the middle layer samples were collected 20 cm below the cellar's surface, and the lower layer samples were obtained 160 cm below the cellar's surface, resulting in a total of 12 samples per sorghum variety. These samples were promptly stored in sterile polyethylene bags at −20 °C for subsequent analysis.

2.2. Samples extraction

For each Zaopei sample, a precise measurement of one gram was taken and subsequently blended with a 10-ml mixture of solvent, which was prepared in a volumetric ratio of methanol, acetonitrile, and water at 2:2:1, respectively. Ultrasonic extraction lasted for 30 min at room temperature, followed by centrifugation at 2980*g* for 10 min to gather the supernatant. Subsequently, the supernatant was filtered through a 0.45 μm organic membrane filter and then transferred into sample vials for further use.

2.3. UHPLC-Q-TOF-MS/MS detection

The HSS T3 column (2.1*100 mm, 1.8 μm) was employed at 40 °C with a flow rate of 0.4 ml/min. Mobile phase A consisted of 0.1% formic acid (FA) with acetonitrile (95:5), while mobile phase B comprised acetonitrile (with 0.1% FA). The injection volume was 5 μl. The gradient elution proceeded as follows: 0–2 min at 100% A, 2–11 min at 50% A, 11–13.5 min at 5% A, 13.5–17.5 min at 5% A, and 17.5–17.7 min returning to 0% A.

Mass spectrometry employed an X500R QTOF in Information Dependent Acquisition (IDA) modes. Parameters were configured as follows: In positive ion mode, the cluster voltage was 80 V, collision energy was 35 ± 15 V, ion spray voltage was 5500 V. Both ion source gases were maintained at 50 psi, and the ion source temperature was set at 500 °C. Conversely, in the negative ion mode, the cluster voltage was −80 V, the collision energy ranged from −35 ± 15 V, and the ion spray voltage was −4500 V. The mass scan range was from 40 to 1200 *m/z*, and the accumulation time was 40 ms. Both positive and negative ion modes were used in the instrument.

2.4. Molecular network analysis

Utilizing the feature-based molecular networking (FBMN) approach, the analysis was performed on the GNPS platform (http://gnps.ucsd.edu) as detailed in references (W. Jia, Du, Fan, Wang, & Shi, 2022; Li et al., 2023). Raw data was converted to mgf format utilizing MSDIAL 4.6. Precursor ion mass tolerance and MS^2 fragment ion tolerance were both set at 0.02 Da. Only comparisons showing at least 6 mass-similar fragments and a similarity score above 0.7 were considered valid. The resulting feature molecular network underwent online analysis on the GNPS platform. Spectra within the network were searched against public spectral databases, including MassBank, ReSpect, and the NIST library. The generated task from the FBMN analysis is accessible through the following URLs: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=e8f62e3663ca4cf68213a87d414c0496. Additionally, for visualization, the created network was visualized using Cytoscape 3.9.1.

2.5. Multivariate statistical analysis

MetaboAnalyst 5.0 (http://www.metaboanalyst.ca/) was used to perform Principal Component Analysis (PCA) and Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) on the UHPLC-Q-TOF-MS/MS data. Significant metabolites were identified based on variables meeting the VIP >1 criterion. Subsequently, further analysis was undertaken to explore metabolic pathways using these identified significant metabolites.

3. Result and discussion

3.1. Exploring chemical space in two sorghum groups via FBMN

The research utilized a non-targeted UHPLC-Q-TOF-MS/MS approach to comprehensively analyze all metabolites present in Zaopei. As depicted in Fig. S1, the optimized total ion chromatograms (TIC) effectively illustrate the separation of ions within 20 min for both groups of Zaopei. Notably, the TIC highlights distinct peaks corresponding to various metabolites in the two sample sets, signifying significant distribution differences between them.

The study employed FBMN technology to analyze the varied chemical composition of Zaopei samples. In the positive ion mode, a total of 24,030 nodes (compounds) were extracted, forming 22,666 molecular families (nodes ≥2) (Fig. S2). Conversely, the negative ion mode yielded 19,956 nodes, resulting in 17,893 molecular families (nodes ≥2) (Fig. S3). Nodes were represented in color-coded pie charts, illustrating the metabolite's relative abundance in both groups—red representing the glutinous sorghum group and green depicting the non-glutinous sorghum group. Larger nodes corresponded to compounds identified within the GNPS database, leading to the identification of a total of 267 compounds (Table S1). The metabolites can be broadly categorized into 16 groups, encompassing lipids and their derivatives (81), amino acids and their derivatives (44), flavonoids (30), peptides (26), organic heterocyclic compounds (18), phenylpropanoids and polyketones (16), aromatic compounds (11), organic acids and their derivatives (9), alkaloids and their derivatives (7), organic nitrogen compounds (6), organic oxygen compounds (6), other (6), organic nitrogen compounds (2), indoles and their derivatives (1), lignans, neolignans, and related compounds (1), organic oxygen compounds (1). Lipids and amino acids emerge as the primary metabolites. The numbers in parentheses denote the quantity of compounds in each category.

3.2. Visualizing chemical structural correlations

Utilizing the FBMN algorithm, compounds sharing analogous MS^2 fragmentation patterns have a tendency to cluster into molecular families. Consequently, an in-depth manual analysis was conducted to explore the structural interconnections among constituents within

sphingolipin, peptides, and flavonoid molecular families. The objective was to visually represent the chemical structural correlations existing within these molecular groups. This step serves as a foundational framework for deducing components that might not be encompassed within the GNPS database.

3.2.1. Annotation of sphingolipid molecular families

In this study, a total of 81 lipid substances were identified, representing the largest number of non-volatile metabolites in Zaopei. The lipids annotated in this study all belong to the sphingolipid category, which, to our knowledge, is the first exploration of sphingolipids in Zaopei. Sphingolipids have a highly complex structure, typically composed of three main structural parts, including a sphingoid base, a carboxylic alkyl chain, and a polar functional group (Li et al., 2017). Fig. 1A displays molecular families M1, encompassing all identified sphingolipid compounds originating from Zaopei. Fig. 1B and C portray the MS^2 spectra of sphinganine and phytosphingosine, respectively, highlighting compounds matched from the GNPS database. Here, manual analysis of representative components from distinct molecular families was performed to enhance the identification's reliability.

Compound B's precursor ion was observed at *m/z* 302.3044 $[M + H]^+$, corresponding to sphinganine (Fig. 1B). Its notable fragment ions include *m/z* 284.2951, *m/z* 266.2870, *m/z* 254.2834, and *m/z* 60.0446. Among these, *m/z* 284.2951 originates from the precursor ion losing a water molecule ($[M + H\text{-}H_2O]^+$), while *m/z* 266.2870 is derived from $[M + H\text{-}2H_2O]^+$, and *m/z* 254.2834 originates from $[M + H\text{-}H_2O\text{-}CH_2O]^+$. Additionally, *m/z* 60 results from the main chain cleavage $[M + H\text{-}C_{16}H_{33}O]^+$, a distinctive fragment ion of sphinganol (Wigger, Gulbins, Kleuser, & Schumacher, 2019; Cho, Maurer, Reynolds, & Kang, 2019). The C_2-C_3 bond cleavage in sphinganol serves as a discriminant between 1-deoxy and 3-deoxy compounds (Kováčik, Pullmannová, Pavlíková, Maixner, & Vávrová, 2020).

Compound C exhibited a precursor ion at *m/z* 318.2971 $[M + H]^+$, corresponding to phytosphingosine (Fig. 1C). Noteworthy fragment ions encompassed *m/z* 300.2865, *m/z* 282.2762, *m/z* 270.2789, and *m/z* 60.0432.

3.2.2. Annotation of peptide molecular families

2,5-Diketopiperazines (DKPs) are cyclic dipeptides formed by the condensation of two α-amino acids, discovered in various natural resources such as bacteria and fungi (Jia, Du, et al., *2022*; Goher, Abdrabo, Veerakanellore, & Elgendy, 2024).

Utilizing the FBMN platform, this study identified representative peptides within the Zaopei. Fig. 2A illustrates molecular families M2 and M3, representing the categories of cyclic peptides and linear peptides, respectively. Compound B's parent ion registers at *m/z* 211.1439 $[M + H]^+$, as depicted in Fig. 2B. A distinctive fragment ion at *m/z* 183.1482 is observed, generated by the elimination of a carbonyl group from the parent ion, thereby identifying it as cyclo(leucylprolyl). This alignment concurs with previous reports (Floris et al., 2019).

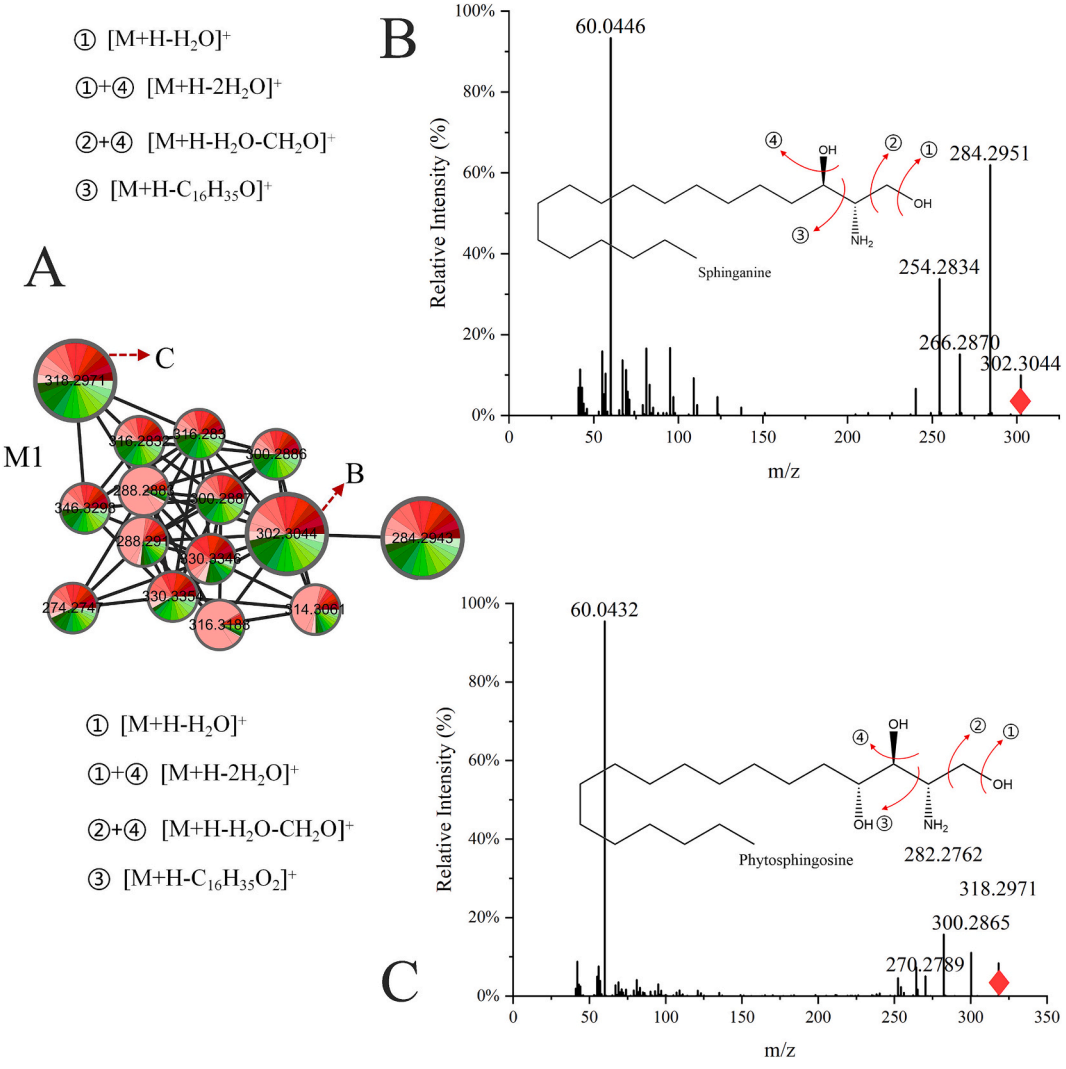


Fig. 1. Molecular families of lipids and structure deduction through MS^2 spectrometry. Molecular families of lipids (A); Sphinganine (B); Phytosphingosine (C).

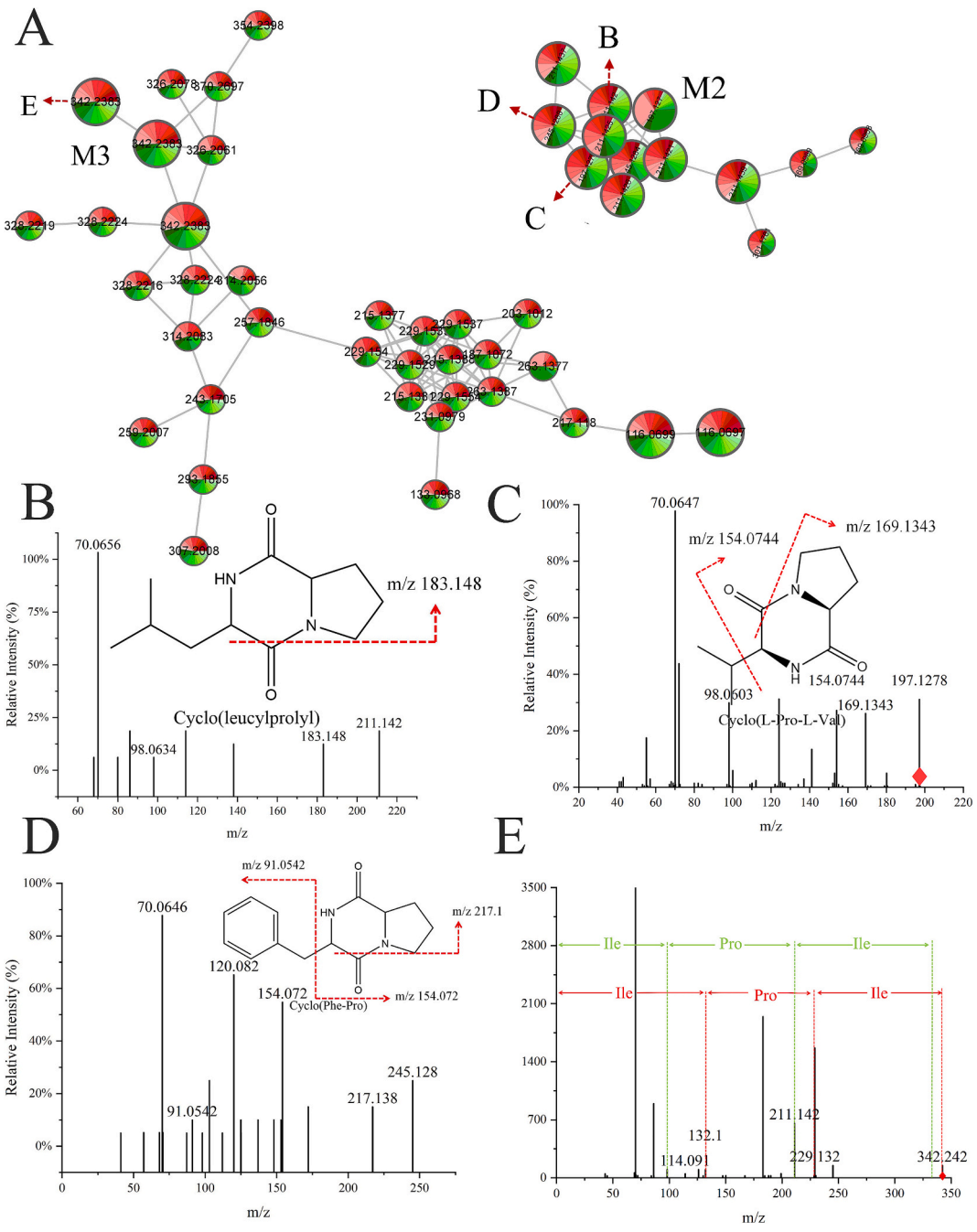


Fig. 2. Molecular families of peptides and structure deduction through MS^2 spectrometry. Molecular families of peptides (A); Cyclo(leucylprolyl) (B); Cyclo(L-Val-L-Pro) (C); Cyclo(L-Phe-D-Pro) (D); Ile-Pro-Ile (E).

Following a similar pattern, compound C's parent ion is at *m/z* 197.1278, and the characteristic ion at *m/z* 169.1343 is obtained by the loss of a CO from the parent ion. As literature reference, this component was identified as cyclo(L-Val-L-Pro) (Fig. 2C). Valine-proline derivatives are typically crucial contributors to bitterness (Amandine et al., 2022; Pawel J. Andruszkiewicz, D'Souza, Altun, Corno, & Kuhnert, 2019a).

The parent ion of compound D stands at *m/z* 245.1288, identified as cyclo(Phe-Pro) (Fig. 2D). Additionally, its fragment ion at *m/z* 217.1378 is generated through the loss of a single molecule of CO from the parent ion. Previous data also indicate that Pro-based pyrrolidone ketone derivatives share a primary product ion mass of *m/z* 70, identical to the ion mass of proline (Paweł J. Andruszkiewicz, D'Souza, Altun, Corno, & Kuhnert, 2019b). Additionally, 2,5-diketopiperazines primarily display consecutive neutral losses, such as -CO (28 Da), -OH moiety (18 Da), and a subsequent -CO (28 Da) loss, within their main mass spectrometric fragmentation pathways.

In the molecular family M3, compound E (Fig. 2E) exhibits a parent ion mass of *m/z* 342.2383 $[M + H]^+$, identified as Ile-Pro-Ile. At *m/z* 114.091, there's an indication of the potential neutral loss of C_2H_5 (−29 Da). Simultaneously, the ion at *m/z* 86.0961 suggests the presence of an Ile residue, while *m/z* 98.0607 corresponds to the proline residue. These observed fragmentation patterns align consistently with those previously reported (Chen et al., 2021).

3.2.3. *Annotation of flavonoid molecular families*

Thirty flavonoid compounds were identified from the two Zaopei groups. M4 in Fig. 3A illustrates the flavonoid molecular family, while Fig. 4B-G respectively present the MS^2 spectra of naringenin, apigenin,

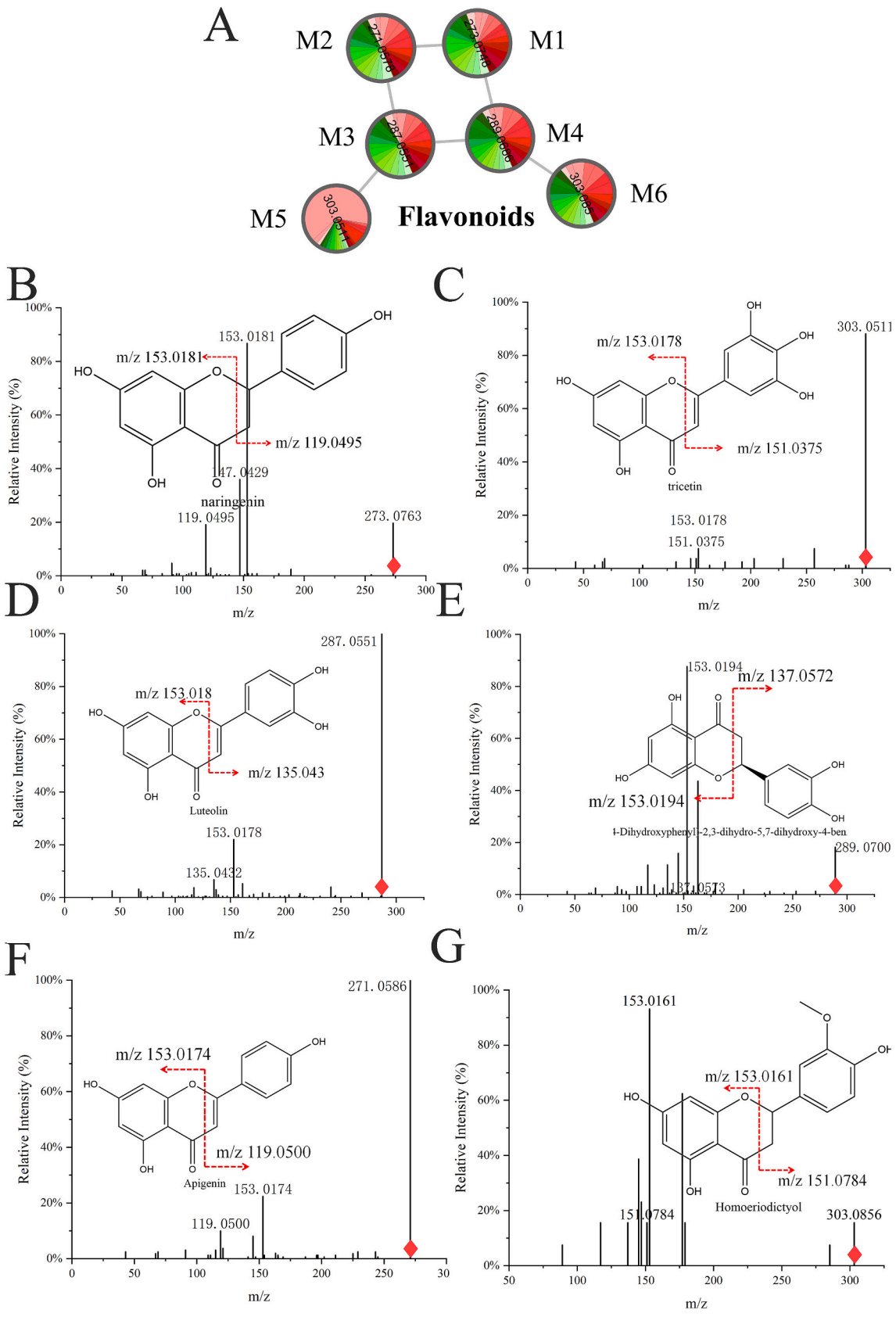


Fig. 3. Fig. 4: Flavonoid molecular families and structure deduction via MS^2 spectrometry. Molecular families of flavonoids (A); Naringenin (B); Apigenin (C); Luteolin (D); (*S*)-2-(3,4-Dihydroxyphenyl)-2,3-dihydro-5,7-dihydroxy-4-benzopyrone (E); Tricetin (F); Homoeriodictyol (G).

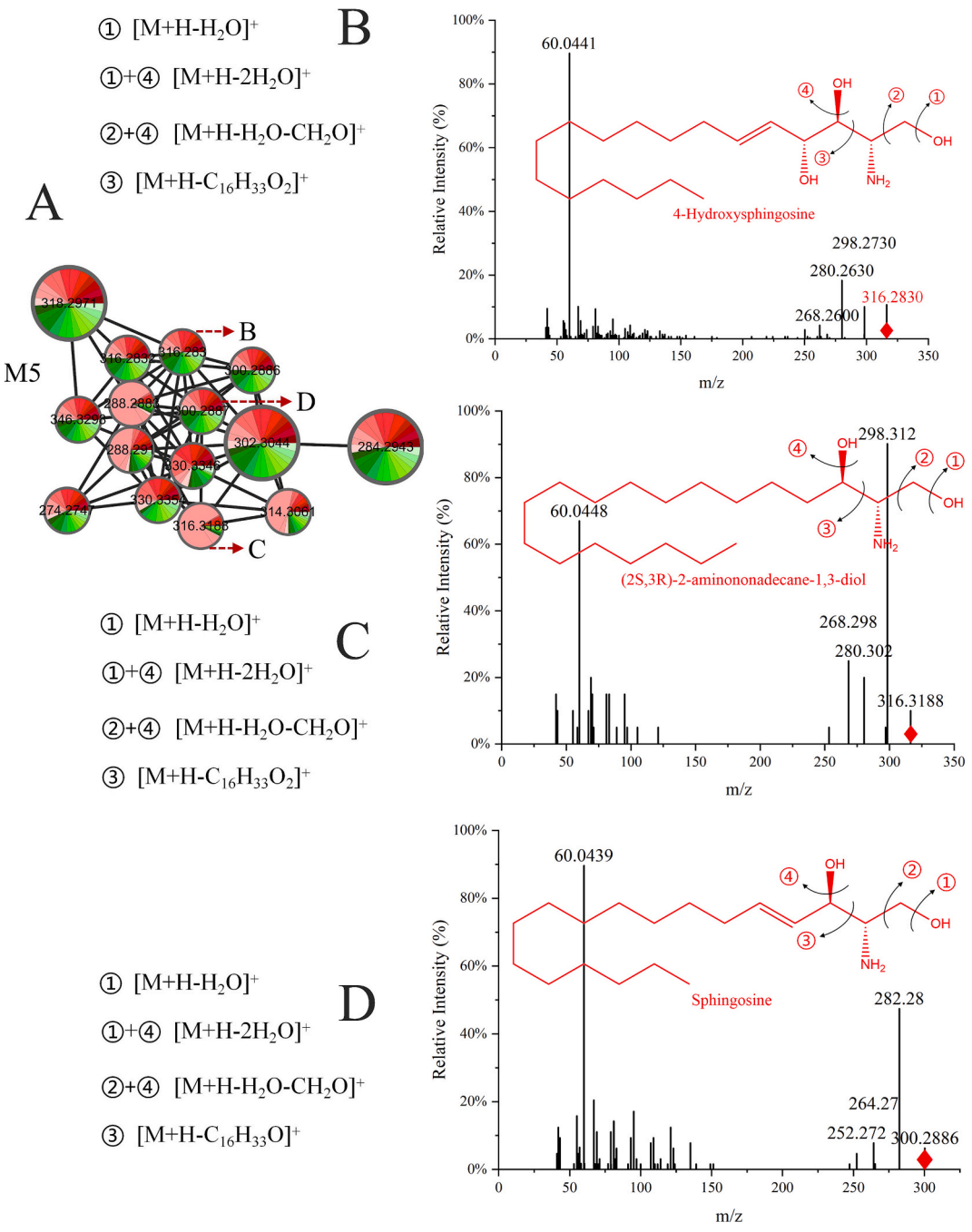


Fig. 4. Inferred lipids not present in the database (A); 4-Hydroxysphingosine (B); (2*S*,3*R*)-2-Aminononadecane-1,3-diol (C); Sphingosine (D).

luteolin, (*S*)-2-(3,4-Dihydroxyphenyl)-2,3-dihydro-5,7-dihydroxy-4-benzopyrone, tricetin, and homoeriodictyol.

Compound B (Fig. 3B) exhibits a parent ion at *m/z* 273.0763, identified as naringenin. Characteristic fragments at *m/z* 153.0181 $[M + H\text{-}C_7H_4O_4]^+$ and *m/z* 119.0495 $[M + H\text{-}C_8H_8O]^+$ originated from the Retro-Diels-Alder (RDA) reaction, confirming its identity as naringenin, consistent with previous reports (Li, Mei, Zhou, Salman Farid, et al., 2023).

Compound E (Fig. 3E) exhibits a parent ion at *m/z* 289.0700 with a molecular formula of $C_{15}H_{12}O_6$. Its MS^2 fragments at *m/z* 153.0194 and *m/z* 137.0572 were derived from the RDA reaction, preliminarily identifying it as (*S*)-2-(3,4-Dihydroxyphenyl)-2,3-dihydro-5,7-dihydroxy-4-benzopyrone (Kang, Price, Ashton, Tapsell, & Johnson, 2016).

Compound F exhibits a parent ion at *m/z* 271.0586, identified as apigenin (Fig. 3F). With a 2 Da difference from node B's parent ion mass and shared fragments at *m/z* 153.0174 and *m/z* 119.0500, this compound was identified as apigenin (Abu-Reidah, Arráez-Román, Warad, Fernández-Gutiérrez, & Segura-Carretero, 2017; Li, Mei, Zhou, Farid, et al., 2023).

Similarly, compound D was identified as luteolin (Fig. 3D), in line with earlier studies (Śliwka-Kaszyńska, Anusiewicz, & Skurski, 2022).

Additionally, compounds C and G are isomeric, with parent ions at *m/z* 303.0511 Da and *m/z* 303.0856, identified as tricetin or homoeriodictyol, respectively. Flavonoids, owing to their differing connecting groups, showcase variable polarity. The lipophilicity parameter (ClogP) calculation aids in estimating isomeric retention times within a reverse-phase chromatography column, facilitating their differentiation. Typically, higher ClogP values correlate with longer retention times (Liu

et al., 2022; Y. Zhang et al., 2023). Tricetin exhibits a slightly lower ClogP value of 1.12 compared to homoeriodictyol, which demonstrates a value of 1.5. Utilizing the difference in retention times, compounds E (RT = 7.89 min) and F (RT = 9.72 min) were tentatively identified as tricetin and homoeriodictyol, respectively (Wu & Tian, 2019).

3.3. Inference of sphingolipid not present in the database

FBMN offers an advantage by utilizing known structural components as 'seeds' to deduce the structure of unidentified lipids within the same molecular family. This is achieved through analogous mass spectrometric fragmentation patterns. Notably, the structures of compounds B, C, and D were inferred by analyzing the mass spectrometric fragmentation patterns of the sphingolipid family. It's important to note that these structures are not available in the GNPS database (Fig. 4).

The parent ion of compound B registers at *m/z* 316.2830 $[M + H]^+$, displaying characteristic ions at *m/z* 298.273 and *m/z* 280.263 (Fig. 4B). The specific fragmentation pattern observed in the sphingolipid family (section 3.2) suggests that the parent ion undergoes consecutive water molecule losses, generating fragments at *m/z* 298.273 and *m/z* 280.263. Additionally, the 2 Da variance from phytosphingosine indicates a potential presence of a double bond in Compound B, tentatively identified as 4-hydroxysphingosine.

Compound C displays a parent ion at *m/z* 316.3188 $[M + H]^+$,

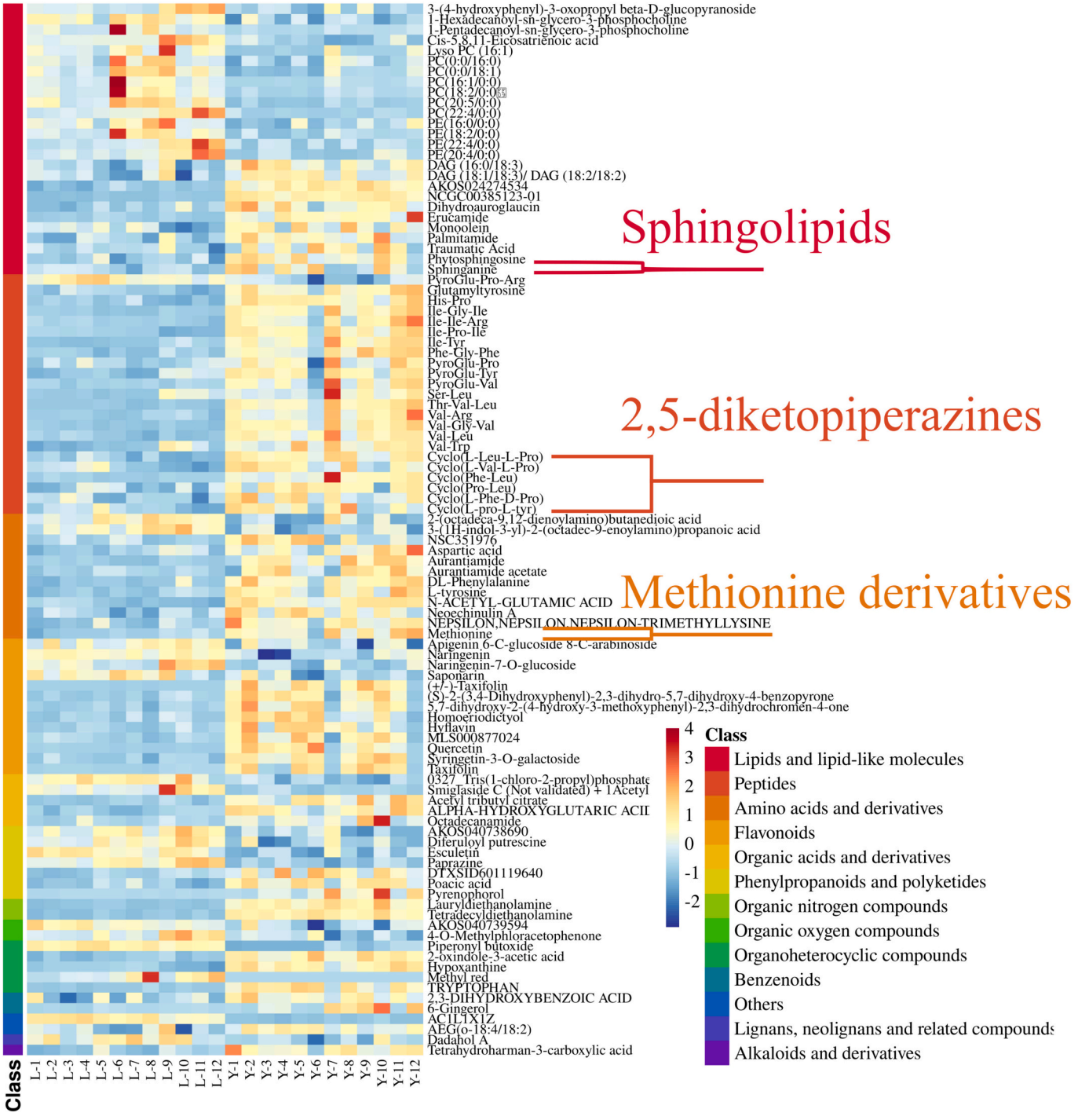


Fig. 5. Visualization of differential non-volatile compounds detected in Zaopei from two sorghum varieties (VIP > 1).

featuring characteristic ions at *m/z* 298.312 and *m/z* 280.302 (Fig. 4C). These ions stem from the fragmentation involving the loss of one and two water molecules from the parent ion, respectively. Furthermore, a 14 Da deviation from sphinganine at m/z suggests the possible presence of an additional -CH_2 group in this compound. By consulting prior literature (Li et al., 2017; Steshin & Ishikawa, 2021), it is tentatively identified as (2*S*,3*R*)-2-aminononadecane-1,3-diol.

Compound D has a parent ion at *m/z* 300.2886 $[M + H]^+$, with characteristic fragment ions at *m/z* 282.28, *m/z* 264.27, *m/z* 252.272, and *m/z* 60.0439 (Fig. 4D). Its fragmentation pattern resembles that of phytosphingosine. Hence, based on previous literature, it is deduced to be sphingosine (Li et al., 2017).

3.4. Multivariate statistical analysis

Principal Component Analysis (PCA) is an unsupervised data analysis method aimed at reducing dimensions by generating a set of new composite variables(Du et al., 2020). The PCA analysis results (Fig. S4A) clearly demonstrate the model's ability to effectively separate various samples into two major clusters, highlighting its excellent discriminative capability. To pinpoint the variables contributing to the distinction between the two groups, an additional OPLS-DA analysis model was utilized (Fig. S4B). The model showcases R^2Y and Q^2 values of 0.976 and 0.948, respectively. Employing a permutation test of 1000 iterations to assess the predictive ability and robustness of the OPLS-DA model resulted in a *p*-value below 0.001, indicating the model's substantial capability to differentiate between the sample groups and its excellent fitting and predictive capacity (R. Zhang, Zhu, & Jia, 2022).

The study employed a VIP > 1 criterion to identify 101 differential metabolites visualized in a heatmap (Fig. 5), delineating the influence of various metabolites within the FGs of two sorghum varieties. These metabolites were sorted into 14 categories: Lipids and lipid-like molecules (26)、Peptides (23)、Flavonoids (13)、Amino acids and derivatives (12)、Phenylpropanoids and polyketides (7)、Organic acids and derivatives (5)、Organoheterocyclic compounds (5)、Benzenoids (2)、Organic nitrogen compounds (2)、Organic oxygen compounds (2)、Others (2)、Alkaloids and derivatives (1)、Lignans, neolignans and related compounds (1). The numbers in parentheses represent the count of compounds in each group. This thorough classification provides an intricate breakdown of diverse metabolites, highlighting their distribution across various chemical groups in the FGs of the two sorghum varieties.

This study reveals significant differences in 26 lipid and lipid-derived compounds during the fermentation of two sorghum varieties. Of these, 15 lipid and lipid-derived compounds displayed higher levels in the Zaopei of the glutinous sorghum compared to the non-glutinous variety. Notably, PC (phosphatidylcholine) species like PC (0:0/16:0), PC (0:0/18:1), PC (16:1/0:0), PC (18:2/0:0), PC (20:5/0:0), PC (22:4/0:0), and PE (phosphatidylethanolamine) species such as PE (16:0/0:0), PE (18:2/0:0), PE (20:4/0:0), and PE (22:4/0:0) were notably elevated. Conversely, 11 lipid and lipid-derived compounds showed increased levels in the non-glutinous sorghum group compared to the glutinous group, particularly phytosphingosine and sphinganine.Cyclo(L-Phe-D-Pro), Cyclo(L-Val-L-Pro), and Cyclo(L-pro-L-tyr) comprise a notable subset within the second major category of distinct metabolites.

Specifically, the levels of sphingolipids, 2,5-diketopiperazines, and methionine derivatives in glutinous sorghum varieties are 1.21–1.71 times higher than those in non-glutinous sorghum varieties. Understanding these variations could provide valuable insights into the unique flavor profiles of Baijiu produced from different sorghum varieties.

3.5. Enrichment analysis and KEGG pathway analysis

To unveil potential differential metabolic pathways and gain deeper insights into how these sorghum varieties influence metabolite production mechanisms, we conducted KEGG pathway analysis using yeast strains as the reference species. In Fig. 6A, each bubble represents a specific metabolic pathway, where larger bubbles signify a higher impact factor. Meanwhile, deeper colors indicate smaller *p*-values, suggesting a more significant enrichment of that pathway. The identified key metabolites predominantly participate in 16 metabolic pathways: Aminoacyl-tRNA biosynthesis, Sphingolipid metabolism, Arginine biosynthesis, Phenylalanine, tyrosine, and tryptophan biosynthesis, Ubiquinone and other terpenoid-quinone biosynthesis, Glycine, serine, and threonine metabolism, Monobactam biosynthesis, Cysteine, and methionine metabolism, Cyanoamino acid metabolism, Beta-alanine metabolism, Tyrosine metabolism, Lysine biosynthesis, Thiamine metabolism, Alanine, aspartate, and glutamate metabolism, Tryptophan metabolism, and Purine metabolism.

4. Discussion

Glutinous sorghum is traditionally believed to produce a richer flavor profile in Baijiu compared to non-glutinous varieties. Therefore, we hypothesized that glutinous sorghums generate a distinct array of non-volatile metabolites during fermentation, ultimately significantly influencing the flavor of Baijiu. Establishing the connection between these metabolites and Baijiu aroma is crucial for elucidating the mechanisms affecting Baijiu quality. Due to the complexity of these metabolites, there is limited scientific understanding of their roles. This study employs a data-driven computational method to annotate these “dark matter” metabolites in Zaopei for the first time, providing deeper insights into the biochemical processes at play.

4.1. FBMN in metabolite annotation

The implementation of FBMN in our study marks a significant advancement over conventional local database search methodologies, offering distinct advantages in metabolite analysis of Zaopei. Firstly, FBMN provides a holistic perspective on the metabolite landscape through molecular family analysis, categorizing metabolites based on shared structural characteristics (S. Chen et al., 2021). This approach simplifies the complex metabolic profile and facilitates the recognition of metabolic patterns.

Secondly, FBMN allows for the identification of unknown metabolites using a single known metabolite from a reaction pair. This is a significant departure from traditional methods, which often require extensive databases or multiple known standards for identification. By focusing on shared features in the MS^2 spectra, FBMN significantly expands our reach into uncharted chemical territories, enhancing our ability to explore and understand novel or rare metabolites(Li, Mei, Zhou, Salman Farid, et al., 2023).

The adoption of FBMN has refined our analytical process and broadened the scope of our chemical exploration. It has enabled us to delve deeper into the intricacies of Zaopei's metabolite profile, providing a robust foundation for future studies in the realm of fermented food chemistry.

4.2. Roles of key differential metabolites: Sphingolipids, DKPs, and amino acid derivatives

Among the myriad metabolites identified, our investigation specifically targets three pivotal differential metabolites in the Zaopei of Baijiu: sphingolipids, 2,5-diketopiperazines (DKPs), and amino acid derivatives.

4.2.1. Sphingolipids

Sphingolipids, a unique class of lipids similar to phospholipids but without a glycerol backbone, initiate their biosynthesis with sphingosine as the cornerstone molecule (Fig. 6B). Characterized by amide linkages and extended fatty acid chains (C20 to C26), sphingolipids exhibit distinctive properties that set them apart from other lipid classes

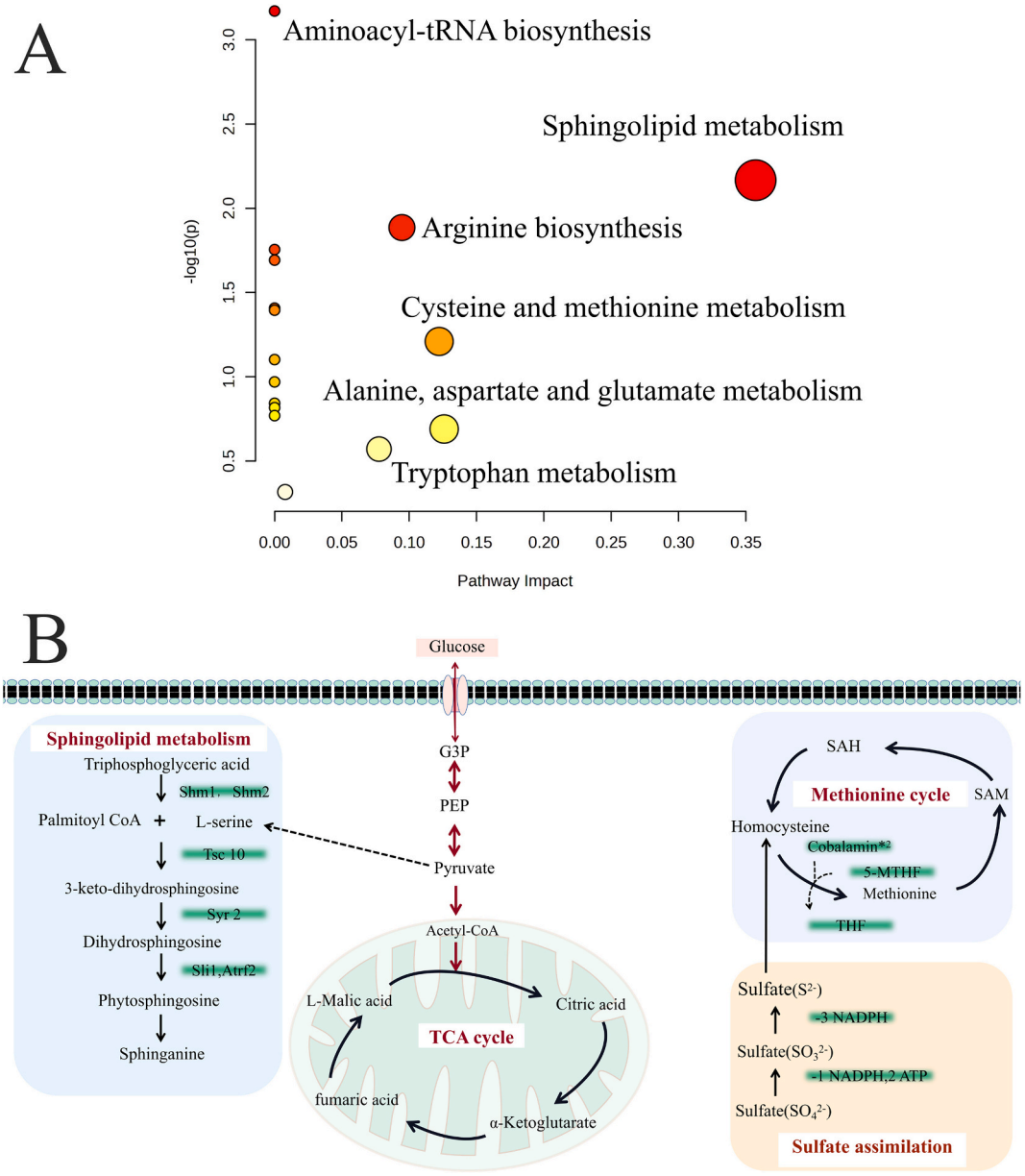


Fig. 6. KEGG Pathway analysis depicting key metabolic pathways (A); Metabolic pathway schematic in Zaopei (B).

(Hannun & Obeid, 2018). Research has highlighted the significant impact of glycosyl sphingomyelins on yeast fermentation traits, particularly concerning cell membrane fluidity (Ferdouse et al., 2018). An increase in membrane fluidity correlates with reduced generation of ethyl esters, crucial for Baijiu's aroma profile (Saerens et al., 2008). The presence of a 9-methyl base in the sphingoid backbone is suggested to suppress the emergence of volatile flavors, indicating a complex regulatory role in flavor development.

4.2.2. 2,5-DKPs

The synthesis of 2,5-diketopiperazines (DKPs) is catalyzed by non-ribosomal peptide synthetase (NRPS), a complex and highly specialized enzyme system (Fig. 6B)(Borgman, Lopez, & Lane, 2019). NRPSs facilitate the formation of DKPs through a series of coordinated steps, each mediated by distinct enzyme domains. DKPs exhibit unique chemical properties, notably their ability to engage in hydrogen bonding due to the presence of two hydrogen bond donor and two hydrogen bond acceptor sites. This structural feature enables DKPs to interact with a variety of biological targets, influencing both their biological activity and sensory characteristics.

In the context of Baijiu, DKPs are known to contribute to the beverage's complex flavor profile, often imparting bitter, metallic, or astringent taste notes. These sensory attributes are critical in shaping the overall flavor perception of Baijiu, highlighting the importance of DKPs in flavor chemistry.

The content of kafirin, a major storage protein in sorghum, plays a pivotal role in the biosynthesis of DKPs. Factors such as sorghum genotype and environmental conditions can significantly influence the levels of kafirin, and consequently, the levels of DKPs in sorghum-derived Zaopei (Abdelbost et al., 2023). Genotypic variations can lead to differences in protein composition and expression, while environmental factors such as soil composition, climate, and agricultural practices can affect protein stability and availability. These variations directly impact the fermentation process, as the breakdown or release of kafirin during fermentation serves as a precursor for DKP formation.

Given their derivation from kafirin and their significant impact on flavor, DKPs present a promising avenue for quality control in Baijiu production. Monitoring the presence and levels of DKPs in Zaopei can

serve as a reliable indicator for the selection and screening of optimal sorghum varieties tailored for Baijiu production. By establishing a correlation between DKP profiles and Baijiu quality, producers can refine their fermentation processes to enhance flavor consistency and quality. This approach not only improves the sensory attributes of Baijiu but also provides a scientific basis for breeding and selecting sorghum varieties with desirable fermentation properties.

4.2.3. Amino Acid Derivatives

The differential metabolite analysis within the Zaopei of glutinous and non-glutinous sorghum has unveiled significant disparities in the levels of amino acids and their derivatives, with methionine being a standout. Methionine's influence on flavor is multifaceted. Its oxidation via dicarbonyl compounds initiates the Strecker degradation, a pivotal pathway for generating aroma-active aldehydes such as benzaldehyde and methyl butanal (Yaylayan, 2003). These compounds are known for their distinct floral, honey-like, malty, and chocolaty notes, which are highly desirable in the olfactory landscape of Baijiu (Zhu et al., 2020). Moreover, the Strecker degradation intersects with the Maillard reaction, a complex series of non-enzymatic browning reactions fundamental in food flavor development. This interaction is particularly crucial for the formation of pyrazine compounds, which contribute earthy, nutty, and toasty aromas to Baijiu (Song et al., 2020).

The synthesis of methionine in yeast is intricately linked to the sulfate assimilation pathway. This process demands considerable energy investment, reflecting the biological significance of methionine. The conversion of sulfate to sulfite and then to sulfide involves multiple enzymatic reactions, culminating in the production of homocysteine, which is the direct precursor to methionine (Lauinger & Kaiser, 2021). The methylation of homocysteine, facilitated by the folate cycle, is a critical step in methionine biosynthesis, highlighting the interplay between metabolic pathways and the final flavor compounds present in Baijiu.

The observed differences in methionine levels between glutinous and non-glutinous sorghum Zaopei may be attributed to variations in the efficiency of the sulfate assimilation pathway or the activity of enzymes involved in methionine synthesis. Understanding these metabolic nuances is essential for manipulating the flavor development process in Baijiu production, potentially leading to the creation of spirits with tailored aroma profiles.

5. Conclusion

This study presents several new findings and advances our understanding of the biochemical processes in Baijiu production. By employing the data-driven computational method Feature-Based Molecular Networking (FBMN), we successfully annotated 267 previously unknown metabolites, most of which were identified in fermented sorghum for the first time. Through FBMN, we identified key metabolites that play significant roles in shaping Baijiu's flavor profile. The differential presence of sphingolipids, 2,5-diketopiperazines (DKPs), and amino acid derivatives like methionine provides new insights into the flavor-enhancing properties of glutinous sorghum.

Given the complexity of interactions among metabolites, microbes, and enzymes, further investigation is both necessary and timely. Particularly, exploring how sorghum-derived sphingolipids impact yeast during fermentation is a compelling area for future research. This investigation will not only deepen our understanding of the role of sphingolipids in shaping Baijiu flavors but also shed light on the broader mechanisms of flavor development in fermented products.

These findings pave the way for more targeted quality control measures in Baijiu production, ultimately leading to improved flavor consistency and quality. Understanding the intricate metabolic pathways and their impact on flavor development not only enhances Baijiu production but also contributes broadly to the field of flavor chemistry in fermented foods.

CRediT authorship contribution statement

Yulan Li: Writing – original draft, Validation. **Yi Ma:** Validation. **Hui Zhu:** Writing – review & editing. **Yin Liu:** Methodology. **Shijiang Pan:** Supervision. **Xi Chen:** Methodology. **Tao Wu:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgements

We are grateful for the financial support from Sichuan Engineering Technolog Research Centerfor Liquor-Making Grains (2023-02,2022-04). This study was funded by the Science and Technology Research Plan Program of Chongqing Municipal Education Commission (KJQN202001605), the Research Project of Chongqing University of Education (KY202121C), China.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fochx.2024.101646.

References

Abdelbost, L., Morel, M.-H., Nascimento, T. P. D., Cameron, L.-C., Bonicel, J., Larraz, M. F. S., & Mameri, H. (2023). Sorghum grain germination as a route to improve kafirin digestibility: Biochemical and label free proteomics insights. *Food Chemistry, 424*, Article 136407.

Abu-Reidah, I. M., Arráez-Román, D., Warad, I., Fernández-Gutiérrez, A., & Segura-Carretero, A. (2017). UHPLC/MS2-based approach for the comprehensive metabolite profiling of bean (Vicia faba L.) by-products: A promising source of bioactive constituents. *Food Research International, 93*, 87–96.

Amandine, A., Amandine, A., Bettina, C., Bettina, C., Lisa, U., Lisa, U., Irene, C., & Irene, C. (2022). *Use of molecular networking to identify 2,5-diketopiperazines in chocolates as potential markers of bean variety*. Heliyon.

Andruszkiewicz, P. J., D'Souza, R. N., Altun, I., Corno, M., & Kuhnert, N. (2019a). Thermally-induced formation of taste-active 2,5-diketopiperazines from short-chain peptide precursors in cocoa. *Food Research International, 121*(JUL.), 217–228.

Andruszkiewicz, P. J., D'Souza, R. N., Altun, I., Corno, M., & Kuhnert, N. (2019b). Thermally-induced formation of taste-active 2,5-diketopiperazines from short-chain peptide precursors in cocoa. *Food Research International, 121*, 217–228.

Bian, X., Zhang, Y., Li, N., Shi, M., Chen, X., Zhang, H. L., … Wu, J. L. (2023). Ultrasensitive quantification of trace amines based on N-phosphorylation labeling chip 2D LC-QQQ/MS. *Journal of Pharmaceutical Analysis, 13*(3), 315–322.

Borgman, P., Lopez, R. D., & Lane, A. L. (2019). The expanding spectrum of diketopiperazine natural product biosynthetic pathways containing cyclodipeptide synthases. *Organic & Biomolecular Chemistry, 17*(9), 2305–2314.

Cao, S., Wang, L., Wu, Q., Ni, D., Xu, Y., & Lin, L. (2020). *Regional aroma characteristics of sorghum for Chinese liquor production, 126*(3), 306–315.

Chen, B.-R., Wang, C.-Y., Wang, P., Zhu, Z.-X., Xu, N., Shi, G.-S., Yu, M., Wang, N., Li, J.-H., Hou, J.-M., Li, S.-J., Zhou, Y.-F., Gao, S.-J., Lu, X.-C., & Huang, R.-D. (2019). Genome-wide association study for starch content and constitution in sorghum (Sorghum bicolor (L.) Moench). *Journal of integrative. Agriculture, 18*(11), 2446–2456.

Chen, S., Huang, G., Liao, W., Gong, S., Xiao, J., Bai, J., … Wu, J.-L. (2021). Discovery of the bioactive peptides secreted by Bifidobacterium using integrated MCX coupled with LC–MS and feature-based molecular networking. *Food Chemistry, 347*, Article 129008.

Cho, H. E., Maurer, B. J., Reynolds, C. P., & Kang, M. H. (2019). Hydrophilic interaction liquid chromatography-tandem mass spectrometric approach for simultaneous determination of safingol and D-erythro-sphinganine in human plasma. *Journal of Chromatography Analytical Technologies in the Biomedical and Life Sciences, 1112*, 16–23.

Du, H., Lv, H., Xu, Z., Zhao, S., Huang, T., Manyande, A., & Xiong, S. (2020). The mechanism for improving the flesh quality of grass carp (Ctenopharyngodon idella) following the micro-flowing water treatment using a UPLC-QTOF/MS based metabolomics method. *Food Chemistry, 327*, Article 126777.

Fang, X., Liu, Y., Xiao, J., Ma, C., & Huang, Y. (2023). GC–MS and LC-MS/MS metabolomics revealed dynamic changes of volatile and non-volatile compounds during withering process of black tea. *Food Chemistry, 410*, 135396.
Ferdouse, J., Yamamoto, Y., Taguchi, S., Yoshizaki, Y., Takamine, K., & Kitagaki, H. (2018). Glycosylceramide modifies the flavor and metabolic characteristics of sake yeast. *PeerJ, 6*, Article e4768.
Floris, P., McGillicuddy, N., Morrissey, B., Albrecht, S., Kaisermayer, C., Hawe, D., Riordan, L., Lindeberg, A., Forestell, S., & Bones, J. (2019). A LC–MS/MS platform for the identification of productivity markers in industrial mammalian cell culture media. *Process Biochemistry, 86*, 136–143.
Goher, S. S., Abdrabo, W. S., Veerakanellore, G. B., & Elgendy, B. (2024). 2,5-Diketopiperazines (DKPs): promising scaffolds for anticancer agents. *Current Pharmaceutical Design, 30*(8), 597–623.
Han, J., Gong, S., Bian, X., Qian, Y., Wang, G., Li, N., & Wu, J.-L. (2023). Polarity-regulated derivatization-assisted LC-MS method for amino-containing metabolites profiling in gastric cancer. *Journal of Pharmaceutical Analysis, 13*(11), 1353–1364.
Hannun, Y. A., & Obeid, L. M. (2018). Sphingolipids and their metabolism in physiology and disease. *Nature Reviews. Molecular Cell Biology, 19*(3), 175–191.
He, Y., Liu, Z., Qian, M., Yu, X., Xu, Y., & Chen, S. (2020). Unraveling the chemosensory characteristics of strong-aroma type baijiu from different regions using comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry and descriptive sensory analysis. *Food Chemistry, 331*, Article 127335.
Hu, X., Tian, R., Wang, K., Cao, Z., Yan, P., Li, F., Li, X., Li, S., & He, P. (2021). The prokaryotic community, physicochemical properties and flavors dynamics and their correlations in fermented grains for Chinese strong-flavor baijiu production. *Food Research International, 148*, Article 110626.
Jia, W., Du, A., Fan, Z., Wang, Y., & Shi, L. (2022). Effects of short-chain peptides on the flavor profile of baijiu by the density functional theory: Peptidomics, Sensomics, flavor reconstitution, and sensory evaluation. *Journal of Agricultural and Food Chemistry, 70*(30), 9547–9556.
Kang, J., Price, W. E., Ashton, J., Tapsell, L. C., & Johnson, S. (2016). Identification and characterization of phenolic compounds in hydromethanolic extracts of sorghum wholegrains by LC-ESI-MSn. *Food Chemistry, 211*, 215–226.
Kováčik, A., Pullmannová, P., Pavlíková, L., Maixner, J., & Vávrová, K. (2020). Behavior of 1-deoxy-, 3-deoxy- and N-methyl-ceramides in skin barrier lipid models. *Scientific Reports, 10*(1), 3832.
Lauinger, L., & Kaiser, P. (2021). Sensing and signaling of methionine metabolism. *Metabolites, 11*(2).
Li, K., Chen, Y., Liu, T., Deng, M., Xu, Z., Fu, G., … Zheng, F. (2020). Analysis of spatial distribution of bacterial community associated with accumulation of volatile compounds in Jiupei during the brewing of special-flavor liquor. *LWT, 130*, 109620.
Li, W., Mei, S., Zhou, H., Salman Farid, M., Hu, T., & Wu, T. (2023). Metabolite fingerprinting of the ripening process in Pixian douban using a feature-based molecular network and metabolomics analysis. *Food Chemistry, 418*, 135940.
Li, W., Mei, S., Zhou, H., Salman Farid, M., Hu, T., & Wu, T. (2023). Metabolite fingerprinting of the ripening process in Pixian douban using a feature-based molecular network and metabolomics analysis. *Food Chemistry, 418*, Article 135940.
Li, Y., Lou, Y., Mu, T., Ke, A., Ran, Z., Xu, J., Chen, J., Zhou, C., Yan, X., Xu, Q., & Tan, Y. (2017). Sphingolipids in marine microalgae: Development and application of a mass spectrometric method for global structural characterization of ceramides and glycosphingolipids in three major phyla. *Analytica Chimica Acta, 986*, 82–94.
Liu, K.-X., Li, N., Yin, Y.-H., Zhong, Z.-J., Li, P., Liu, L.-F., & Xin, G.-Z. (2022). An in-house database-driven untargeted identification strategy for deep profiling of chemicalome in Chinese medicinal formula. *Journal of Chromatography A, 1666*, Article 462862.
Liu, M.-K., Tang, Y.-M., Liu, C.-Y., Tian, X.-H., Zhang, J.-W., Fan, X.-L., Jiang, K.-F., Ni, X.-L., & Zhang, X.-Y. (2023). Variation in microbiological heterogeneity in Chinese strong-flavor baijiu fermentation for four representative varieties of sorghum. *International Journal of Food Microbiology, 397*, Article 110212.
Robin, S., Robin, S., Daniel, P., Louis-Félix, N., Mingxun, W., Mingxun, W., Allegra, T. A., Annika, J., Hiroshi, T., Hiroshi, T., Hiroshi, T., Johannes, R., Mar, G.-A., Mar, G.-A., Kai, D., Ansgar, K., Tomáš, P., Tomáš, P., Tomáš, P., Zdeněk, K., Alan, K. J., Andrés Mauricio, C.-R., Kelly, W., Mélissa, N.-E., Alexander, A. A., Anelize, B., Anelize, B., Anelize, B., Andrea Georgina Albarracín, O., Carlismari, O. G., Carlismari, O. G., Fernando, V., Irina, K., Julia, M. G., Emily, C. G., Yannick, H., Svetlana, A. K., Matthew, A. P., Morgan, P., Richard, T., Audrey Le, G., Gajender, A., Helena Mannochio, R., Helena Mannochio, R., Birgit, A., Florian, H., Heiko, H., Hui, Z., Manuela, R., Kimberly, A. P., Lihini, I. A., Sebastian, B., Kerry, L. M., Hans-Ulrich, H., Uwe, K., & Pieter, C. D. (2021). Ion identity molecular networking for mass spectrometry-based metabolomics in the GNPS environment. (Nature Communications).
Mei, S., Ding, J., & Chen, X. (2023). Identification of differential volatile and non-volatile compounds in coffee leaves prepared from different tea processing steps using HS-SPME/GC–MS and HPLC-Orbitrap-MS/MS and investigation of the binding mechanism of key phytochemicals with olfactory and taste receptors using molecular docking. *Food Research International, 168*, 112760.
Phelan, V. V. (2020). Feature-based molecular networking for metabolite annotation. *Methods of Molecular Biology, 2104*, 227–243.
Saerens, S. M., Delvaux, F., Verstrepen, K. J., Van Dijck, P., Thevelein, J. M., & Delvaux, F. R. (2008). Parameters affecting ethyl ester production by Saccharomyces cerevisiae during fermentation. *Applied and Environmental Microbiology, 74*(2), 454–461.
Śliwka-Kaszyńska, M., Anusiewicz, I., & Skurski, P. (2022). The mechanism of a retro-Diels-Alder fragmentation of Luteolin: Theoretical studies supported by electrospray ionization tandem mass spectrometry results. *Molecules, 27*(3).
Song, X., Jing, S., Zhu, L., Ma, C., Song, T., Wu, J., Zhao, Q., Zheng, F., Zhao, M., & Chen, F. (2020). Untargeted and targeted metabolomics strategy for the classification of strong aroma-type baijiu (liquor) according to geographical origin using comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry. *Food Chemistry, 314*, Article 126098.
Steshin, M., & Ishikawa, T. (2021). Liquid chromatography-tandem mass spectrometry with a new separation mode for rapid profiling of the Z/E isomers of plant glucosylceramides. *Journal of Chromatography B, 1178*, Article 122807.
Wei, Y., Zou, W., Shen, C. H., & Yang, J. G. (2020). Basic flavor types and component characteristics of Chinese traditional liquors: A review. *Journal of Food Science, 85*(12), 4096–4107.
Wigger, D., Gulbins, E., Kleuser, B., & Schumacher, F. (2019). Monitoring the sphingolipid de novo synthesis by stable-isotope labeling and liquid chromatography-mass spectrometry. *Frontiers in Cell and Developmental Biology, 7*, 210.
Wu, S., & Tian, L. (2019). A new flavone glucoside together with known ellagitannins and flavones with anti-diabetic and anti-obesity activities from the flowers of pomegranate (Punica granatum). *Natural Product Research, 33*(2), 252–257.
Yaylayan, V. A. (2003). Recent advances in the chemistry of Strecker degradation and Amadori rearrangement: Implications to aroma and color formation. *Food Science and Technology Research, 9*(1), 1–6.
Zhang, R., Zhu, Z., & Jia, W. (2022). Molecular mechanism associated with the use of magnetic fermentation in modulating the dietary lipid composition and nutritional quality of goat milk. *Food Chemistry, 366*, Article 130554.
Zhang, Y., Xu, D., Xing, X., Yang, H., Gao, W., & Li, P. (2023). The chemistry and activity-orientedcharacterization of isoflavones difference between roots of Pueraria lobata and P. Thomsonii guided by feature-based molecular networking. *Food Chemistry, 422*, Article 136198.
Zhu, L., Wang, X., Song, X., Zheng, F., Li, H., Chen, F., Zhang, Y., & Zhang, F. (2020). Evolution of the key odorants and aroma profiles in traditional Laowuzeng baijiu during its one-year ageing. *Food Chemistry, 310*, Article 125898.