

## 基于机器学习的代谢组学解析腊八蒜储藏过程中代谢物差异

王 斌, 王珍珍, 陈其彪, 戴 静, 李宪秀, 沙如意, 王宏鹏, 毛建卫, 黄 俊

### Metabolomics Analysis of Metabolite Differences during the Storage Process of Laba Garlic Based on Machine Learning

WANG Cheng, WANG Zhenzhen, CHEN Qibiao, DAI Jing, LI Xianxiu, SHA Ruyi, WANG Hongpeng, MAO Jianwei, and HUANG Jun

在线阅读 View online: <https://doi.org/10.13386/j.issn1002-0306.2022060332>

## 您可能感兴趣的其他文章

### Articles you may be interested in

#### 基于代谢组学分析工夫红茶发酵过程中代谢物的变化

Metabolomics Analysis of Changes in Metabolites During the Fermentation Process of Congou Black Tea

食品工业科技. 2021, 42(21): 8-16 <https://doi.org/10.13386/j.issn1002-0306.2021030361>

#### 基于代谢组学技术研究凤冈锌硒茶对正常大鼠机体代谢调节规律

Global Metabolic Regulation of Fenggang Zinc-Selenium Tea in Normal Rat Based on Metabolomics Technology

食品工业科技. 2020, 41(6): 99-105 <https://doi.org/10.13386/j.issn1002-0306.2020.06.017>

#### 转基因作物代谢组差异的协同性分析

Cooperative analysis of the differences in metabolome for genetically modified crops

食品工业科技. 2017(17): 135-140 <https://doi.org/10.13386/j.issn1002-0306.2017.17.026>

#### 菊花不同提取物代谢组学分析及其抗氧化活性功效物质成分筛选

Metabonomics Analysis of Different Extracts from *Chrysanthemum morifolium* and Screening of Its Antioxidant Active Components

食品工业科技. 2021, 42(16): 8-19 <https://doi.org/10.13386/j.issn1002-0306.2021010248>

#### 基于UPLC-Q-TOF-MS的植物乳杆菌P-8发酵乳代谢物轮廓分析

Metabolites Profile Analysis of Fermented Milk with *Lactobacillus plantarum* P-8 Based on Ultra-performance Liquid Chromatography-quadrupole-time of Flight Mass Spectrometry(UPLC-Q-TOF-MS)

食品工业科技. 2019, 40(11): 152-160 <https://doi.org/10.13386/j.issn1002-0306.2019.11.026>

#### 代谢组学技术在葡萄生理研究中的应用

Application of metabolomics techniques in grape physiology

食品工业科技. 2018, 39(3): 338-342,352 <https://doi.org/10.13386/j.issn1002-0306.2018.03.065>



关注微信公众号, 获得更多资讯信息

王铖, 王珍珍, 陈其彪, 等. 基于机器学习的代谢组学解析腊八蒜储藏过程中代谢物差异 [J]. 食品工业科技, 2023, 44(8): 26-34.  
doi: 10.13386/j.issn1002-0306.2022060332

WANG Cheng, WANG Zhenzhen, CHEN Qibiao, et al. Metabolomics Analysis of Metabolite Differences during the Storage Process of Laba Garlic Based on Machine Learning[J]. Science and Technology of Food Industry, 2023, 44(8): 26-34. (in Chinese with English abstract). doi: 10.13386/j.issn1002-0306.2022060332

· 未来食品 ·

# 基于机器学习的代谢组学解析腊八蒜储藏过程中代谢物差异

王 铖<sup>1</sup>, 王珍珍<sup>1</sup>, 陈其彪<sup>1</sup>, 戴 静<sup>1</sup>, 李宪秀<sup>1</sup>, 沙如意<sup>1,\*</sup>, 王宏鹏<sup>1</sup>, 毛建卫<sup>1,2</sup>, 黄 俊<sup>1</sup>  
(1. 浙江科技学院生物与化学工程学院, 浙江省农产品化学与生物加工技术重点实验室, 浙江省农业生物资源生化制造协同中心, 浙江杭州 310023;  
2. 浙江工业职业技术学院, 浙江绍兴 312000)

**摘 要:** 本研究采用气相色谱-质谱法对不同储藏时间的腊八蒜进行非靶向代谢组学研究, 利用多元统计分析, 筛选出不同储藏时间腊八蒜的显著性代谢物, 进一步利用深度学习支持向量机递归特征消除 (SVM-RFE) 算法确定其标志性代谢物, 并对其标志性代谢物的途径进行富集分析。结果表明, 基于偏最小二乘-判别分析 (PLS-DA) 为标准进行筛选, 共筛选出 57 种差异代谢物 (VIP $\geq$ 1,  $P<0.05$ , FDR $<0.05$ ), 包括酸类 (6 种)、醇类 (12 种)、胺类 (6 种)、糖苷类 (1 种)、酯类 (3 种)、醚类 (4 种)、芳香类 (13 种)、烯烃类 (2 种)、其他 (10 种)。进一步地基于 SVM-RFE 过程筛选出 6 个标志性代谢物, 其中 5-己炔-1-醇是区别大蒜和腊八蒜的关键因素, 而有机活性小分子肌醇、9-十八碳烯酰胺、硬脂酸、棕榈酸、苯甲酸是区分腊八蒜储藏的第 35 和 85 d 的关键因素。KEGG 富集分析表明, 在腊八蒜储藏过程中较为重要的代谢通路为不饱和脂肪酸生物合成途径, 脂肪酸生物合成, 角质、亚伯碱和蜡的生物合成, 抗坏血酸和醛糖酸代谢, 脂肪酸链伸长系统, 磷脂酰肌醇信号系统, 半乳糖代谢, 磷酸肌醇代谢, 脂肪酸降解。本研究解析出的腊八蒜储藏过程中 6 种标志性代谢物可以为腊八蒜的储藏和评价提供理论基础。

**关键词:** 代谢组学, 腊八蒜, 机器学习, 标志性代谢物, 支持向量机

中图分类号: S633.4; TS207.3 文献标识码: A 文章编号: 1002-0306(2023)08-0026-09

DOI: 10.13386/j.issn1002-0306.2022060332



本文网刊:

## Metabolomics Analysis of Metabolite Differences during the Storage Process of Laba Garlic Based on Machine Learning

WANG Cheng<sup>1</sup>, WANG Zhenzhen<sup>1</sup>, CHEN Qibiao<sup>1</sup>, DAI Jing<sup>1</sup>, LI Xianxiu<sup>1</sup>, SHA Ruyi<sup>1,\*</sup>, WANG Hongpeng<sup>1</sup>, MAO Jianwei<sup>1,2</sup>, HUANG Jun<sup>1</sup>

(1. School of Biological and Chemical Engineering, Zhejiang University of Science & Technology, Zhejiang Provincial Key Lab for Chem & Bio Processing Technology of Farm Product, Zhejiang Provincial Collaborative Innovation Center of Agricultural Biological Resources Biochemical Manufacturing, Hangzhou 310023, China;

2. Zhejiang Industry Polytechnic College, Shaoxing 312000, China)

**Abstract:** In this study, non-targeted metabolomic research was conducted using gas chromatography-mass spectrometry (GC-MS) on Laba garlic at different storage times to explore the signature metabolite changes during storage. Multivariate statistical analysis was used to screen out the significant metabolites of Laba garlic at different storage times. Further, the deep learning support vector machine recursive feature elimination (SVM-RFE) algorithm was used to determine its

收稿日期: 2022-07-01

基金项目: 江苏省科技厅富民强县项目 (XZ-SZ202137); 浙江省基础公益研究计划项目 (LGN22C200034)。

作者简介: 王铖 (1994-), 男, 硕士研究生, 研究方向: 生物医学大数据, E-mail: 834278309@qq.com。

\* 通信作者: 沙如意 (1982-), 男, 博士, 副教授, 研究方向: 食品生物技术, E-mail: kevinsha\_0204@163.com。

landmark metabolites, and enrichment analysis of the pathways of its landmark metabolites was carried out. A total of 57 different metabolites were screened out ( $VIP \geq 1$ ,  $P < 0.05$ ,  $FDR < 0.05$ ) based on partial least squares-discriminant analysis (PLS-DA) as a criterion for screening, including acids (6), alcohols (12), amines (6), glycosides (1), esters (3), ethers (4), aromatics (13), olefins (2), and others (10). Further, 6 landmark metabolites were screened based on the SVM-RFE process, among which 5-hexyn-1-ol was the key factor to distinguish garlic and Laba garlic, while the organic active small molecule including inositol, 9-octadecenamide, stearic acid, palmitic acid and benzoic acid were the key factors to distinguish the 35th and 85th days of Laba garlic storage. KEGG enrichment analysis indicated that the more important metabolic pathways during storage were biosynthesis of unsaturated fatty acid, fatty acid biosynthesis, cutin, suberine and wax biosynthesis, ascorbate aldarate metabolism, fatty acid elongation, phosphatidylinositol signaling system, galactose metabolism, inositol phosphate metabolism and fatty acid degradation. The six landmark metabolites in the storage process of Laba garlic analyzed in this study would provide a theoretical basis for the storage and evaluation of Laba garlic.

**Key words:** metabolomics; Laba garlic; machine learning; landmark metabolites; support vector machine (SVM)

我国地域辽阔,大蒜品种资源丰富,据联合国粮食及农业组织(FAO)数据显示我国大蒜种植面积和产量均居世界第一,截止到 2021 年,中国大蒜种植面积为 1013 万亩,大蒜产量为 2162.5 万吨。相比于其它蔬菜,大蒜被推荐为人们日常消耗食物中多酚类物质的重要来源之一<sup>[1]</sup>。目前我国大蒜以原料和初级产品加工为主。将产量丰富的新鲜大蒜进行深加工,一方面可以解决大蒜在贮藏过程中因呼吸作用或腐败变质而造成的浪费,另一方面可以提高大蒜产品的附加值,优化我国大蒜产业结构,提高我国大蒜出口竞争能力。所以,研究大蒜加工过程,对于进一步开发我国大蒜深加工产品以及品质保证均具有重要意义。

腊八蒜是众多大蒜加工产品之一,是中国的传统美食,日久岁深。因其颜色碧绿、酸甜爽口颇受我国北方地区人民喜爱,但腊八蒜绿色素不稳定,随着储藏时间的延长,其颜色会逐渐变黄。多年来国内外学者十分注重对腊八蒜的护色研究,通过物理法<sup>[2]</sup>、化学法<sup>[3]</sup>等延长腊八蒜维持绿色状态的时间,目前,市场上的腊八蒜的保质期一般为 1~3 个月。

代谢组学是一种以高通量检测技术来检测生物样品中小分子代谢物的方法,可用于探索代谢物在不同时间的变化情况,因此被广泛应用于食品的代谢物组分变化研究中,如辣椒、黄酒等,但在腊八蒜方面的应用研究还较少,现有的应用主要集中在腊八蒜的品质评价<sup>[4]</sup>,关于腊八蒜与白蒜之间、不同储藏时间的腊八蒜之间差异代谢物等的研究还未见报道。

机器学习作为一种分析高通量数据的新手段,是统计学、人工智能和计算机科学的交叉学科,随着方法的日益成熟,越来越多研究者将其开发应用于生命科学领域<sup>[5-6]</sup>。本文以腊八蒜为研究对象,采用机器学习结合代谢组学分析,研究腊八蒜与白蒜之间以及不同储藏时间点的腊八蒜样品之间的代谢物差异,并结合多元统计分析和深度学习 SVM-RFE 算法,获得不同样品的标志性代谢物,分析相关的代谢通路,以期为腊八蒜储藏过程中标志性代谢产物的筛选及其生产工艺品质控制提供理论基础。

## 1 材料与方法

### 1.1 材料与仪器

优质保鲜蒜米 购自龙腾蒜业,挑选样品大小均匀,表面无破损,直径在 5.5 cm 左右且收获期一致,将样品于实验室冰箱(4 °C)保存;白米醋 杭州市食品酿造有限公司;甲醇 色谱纯,德国默克公司;正构烷烃( $C_7 \sim C_{30}$ )(纯度  $\geq 98\%$ )、N-甲基-N-(三甲基硅烷)三氟乙酰胺(MSTFA)(纯度  $\geq 99\%$ )、甲氧胺盐酸盐(纯度  $\geq 99\%$ )、吡啶(色谱纯) 上海阿拉丁生化科技股份有限公司;核糖醇 纯度  $\geq 98\%$ ,上海源叶生物科技有限公司。

FA2004 电子天平 上海舜宇横平科学仪器有限公司;Free Zone 2.5 L 冷冻干燥机 美国 Labconco 有限公司;Allegra X-12R 冷冻离心机 美国贝克曼库尔特有限公司;JT-DCY-12Y 水浴氮吹仪 杭州聚同电子有限公司;CHA-2 恒温气浴箱 苏州威尔实验用品有限公司;Agilent 7890-5975C 气相色谱串联质谱仪 安捷伦科技有限公司。

### 1.2 实验方法

1.2.1 样品准备 取 5 kg 保鲜蒜米,利用去离子水清洗三次,放入大 10 L 玻璃罐中,加入 4 kg 白米醋浸泡,于 25 °C 下浸泡 15 d,然后置于 4 °C 冰箱储藏,分别于储藏的第 35 d(绿色)、第 85 d(黄色)取样,每个时间点取 5 个重复样本,以白蒜为空白对照,进行各项品质指标的测定。

1.2.2 色泽测定 将腊八蒜样品置于液氮中冷冻 15 min,于 -80 °C 保存 24 h 后真空冷冻干燥。将冻干的样品粉碎,过 80 目筛,置于干燥器内保存。采用色差仪对样品的色差进行分析,记录样品的  $L^*$ 、 $a^*$ 、 $b^*$  值,每个样品重复三次并取平均值。

1.2.3 样本前处理 称取粉碎后的样品约 50 mg 至 EP 管中,加入 1 mL 去离子水(含 25  $\mu$ g 核糖醇),漩涡混合 5 s,600 W 超声处理 15 min,13000 r/min 离心 15 min,分离得到上清液和残渣。保留上清液,向残渣中加入 1 mL 的甲醇进行同样的提取处理。合并两次上清液,进行衍生化(胺化)反应。每个待测样

本取等体积混合后制成 QC 样本,用以校正混合样品分析结果产生偏差以及由于仪器自身原因造成的失误。取 800  $\mu\text{L}$  的上清液样品,用氮吹仪吹干。加入 100  $\mu\text{L}$  甲氧胺盐试剂(每 1 mL 吡啶溶解 15 mg 甲氧胺盐酸盐),轻轻混匀后,放置 37  $^{\circ}\text{C}$  恒温箱胺化 2 h;然后加入 100  $\mu\text{L}$  MSTFA(含有 1% TMCS, V/V), 37  $^{\circ}\text{C}$  继续胺化 30 min 后过 0.22  $\mu\text{m}$  微孔滤膜,上机检测,每个时间点 5 组样品,每个样品做 3 次个重复。

**1.2.4 GC-MS 分析** 色谱条件:色谱柱为 DB-5MS 毛细管柱(30 m $\times$ 0.25 mm, 0.25  $\mu\text{m}$ ),升温程序为起始温度 80  $^{\circ}\text{C}$ ,保持 2 min,以 5  $^{\circ}\text{C}/\text{min}$  的速度升温至 300  $^{\circ}\text{C}$  并保持 12 min;载气为氦气;进样模式为不分流;进样量为 1  $\mu\text{L}$ 。质谱条件:电子轰击离子源;电子能量 70 eV;传输线温度 280  $^{\circ}\text{C}$ ;离子源温度 240  $^{\circ}\text{C}$ ;母离子  $m/z$  285;激活电压 1.2 kV;质量扫描范围  $m/z$  50~650;采用标准烷烃混合物( $\text{C}_7\sim\text{C}_{30}$ )进行 GC-MS 质量控制和保留指数计算<sup>[7]</sup>。采用 NIST14 数据库(对荷质比  $m/z$  进行比较)进行定性分析,最终得到代谢物鉴定结果。

### 1.3 数据处理

主成分分析(Principal Component Analysis, PCA)、偏最小二乘分析(Partial Least Squares Discriminant Analysis, PLS-DA)、聚类热图分析、代谢通路分析和可视化使用的是 Metaboanalyst 5.0(<http://www.metaboanalyst.ca>), SVM-RFE 算法使用 R 统计软件(<http://www.r-project.org>)。

首先采用主成分分析(PCA)、偏最小二乘-判别分析(PLS-DA)等多元统计方法对腊八蒜样品进行差异分析,利用模型的 VIP(Variable Importance in the Projection)值、 $t$  检验( $t$ -test)的  $P(P$ -value)值以及方差分析的 FDR(False Discovery Rate)值来研究比较空白对照、第 35 和第 85 d 的组间代谢物差异,以  $P<0.05$  且  $\text{FDR}<0.05$  和  $\text{VIP}\geq 1$  作为截止标准,鉴定差异代谢物。第 0、50、100 d 腊八蒜之间标志性代谢物的筛选使用 R 统计软件包(<http://www.r-project.org>)中的 Limma 软件包进行。使用 Limma 的线性模型进行背景校正和归一化处理。使用 caret 包中的 RFE 函数进行标志性代谢物选择,并进行 5 倍交叉验证<sup>[8]</sup>。SVM 分类器使用带有 5 倍交叉验证的 R 包 e1071 构建。 $P<0.05$  被认为表明差异有统计学意义<sup>[9]</sup>。代谢通路分析使用基于京都基因和基因组百科全书(Kyoto Encyclopedia of Genes and Genomes, KEGG)代谢途径库的 MetaboAnalyst 5.0 对所有具有统计学意义的代谢物进行。

## 2 结果与分析

### 2.1 储藏过程中腊八蒜色差变化

色泽是衡量腊八蒜品质的重要物理指标,腊八蒜储藏过程中的色差分析见表 1,其中  $L^*$  值表示样品的明度, $a^*$  值代表红(+)/绿(-)色度, $b^*$  值显示黄(+)/蓝

(-)色度。由表 1 可知,在储藏过程中  $L^*$  值依次减小; $a^*$  值则先减小后增大,而  $b^*$  值依次增大,说明腊八蒜在储藏过程中颜色逐渐由黄绿色变为黄色,储藏第 35 d 的样品可以维持腊八蒜的色泽,之后,样品的色差变化小,储藏至第 85 d 的样品失去绿色。

表 1 腊八蒜储藏过程中  $L^*$ 、 $a^*$ 、 $b^*$  值的变化  
Table 1 Change of  $L^*$ ,  $a^*$  and  $b^*$  values of Laba garlic during the storage process

贮藏时间(d)	$L^*$	$a^*$	$b^*$
空白对照	94.63	0.34	14.69
35	79.85	-3.26	34.78
85	74.90	0.44	40.02

### 2.2 腊八蒜不同储藏时间点样品数据采集分析

将质控(quality control, QC)样本的总离子流图进行谱图重叠比较,如图 1a 所示, QC 样本的基线稳定,各色谱图峰的相应强度和保留时间基本一致,说明仪器数据采集稳定性较好,在整个实验过程中仪器误差引起的变异较小。由图 1b 可知不同储藏时间点的腊八蒜总离子流图有明显的不同,说明三组间的代谢物组成有明显的差异。

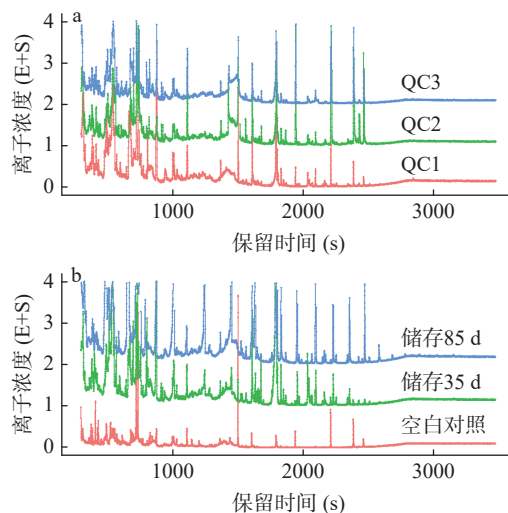


图 1 腊八蒜不同储藏时间点及质控样品的总离子流图

Fig.1 Total ion current diagram of Laba garlic at different storage time points and quality control samples

注: a: 腊八蒜质控样品谱图; b: 不同储藏时间的腊八蒜样品谱图。

### 2.3 PCA 和 PLS-DA 分析

在无监督 PCA 得分图中,每个实心圆代表一个样本,对代谢物进行主成分分析可以总体反映样本组间和组内差异,观察样本间的总体分布趋势,判断可能存在的变异点<sup>[10]</sup>,如图 2 所示,得分图的所有样本都在 95% 置信椭圆内,这说明在被分析的样本中不存在异常值。而且从图形中可以看出,储藏第 35 和第 85 d 的样品在 PC1 上有聚类,和空白对照组有明显的分离,初步认为大蒜原料和腊八蒜之间的代谢物有显著差异,同时储藏第 35 和第 85 d 的样品在 PC2 上有显著分离,则说明储藏过程中部分代谢物变

化对腊八蒜的品质有很大的影响<sup>[11]</sup>。而  $R^2X=0.29$ , 说明两个主成分的贡献率为 29%, 拟合性不是很好,  $Q^2=0.193$ , 与  $R^2X$  相差较小, 说明拟合方程的稳定性较好。

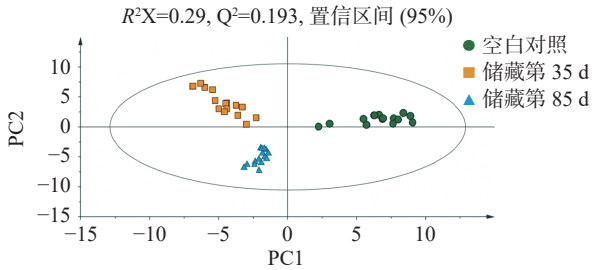


图 2 主成分分析(PCA) (n=15)

Fig.2 Principal component analysis (PCA) drawing Laba garlic samples (n=15)

PLS-DA 是在 PCA 的基础上, 多因变量对多自变量的回归建模方法, 该有监督判别统计方法可弥补 PCA 无法剔除差异变量的不足, 在产品质量预报、毒物分析、食品分析、环境化学等方面有着广泛的应用<sup>[12]</sup>。在 PLS-DA 模型中,  $R^2Y$  和  $Q^2$  的值常用于判断其稳定性和观察原始数据的可靠性, 一般成功的模型两者的值都接近 1<sup>[13]</sup>。如图 3 所示,  $R^2Y=0.963$ ,  $Q^2=0.936$ , 表明模型能很好地解释样本间的差异。为防止模型过拟合, 采用 200 次响应的置换检验(Response Permutation Testing, RPT 检验)对模型进行验证, 其中  $R^2$ 、 $Q^2$  为回归直线与 y 轴的截距值,  $R^2$  表示模型能够解释的方差总和,  $Q^2$  表示模型的预测能力; 使用 RPT 检验时, 一般要求  $Q^2$  小于零。由图 4 可知,  $R^2=0.569$ ,  $Q^2=-0.257$ , 回归线呈向上的趋势,  $Q^2$  小于零, 没有出现过拟合的现象, 模型可靠, 可

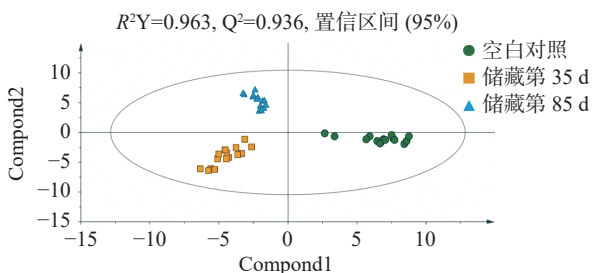


图 3 偏最小二乘-判别分析(PLS-DA) (n=15)

Fig.3 Partial least squares analysis (PLS-DA) was used to draw Laba garlic samples (n=15)

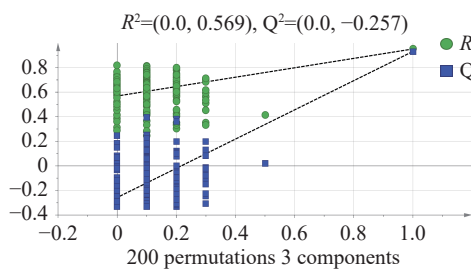


图 4 PLS-DA 模型置换检验

Fig.4 Permutation test plots of PLS-DA model

用于后续的分析。

### 2.4 差异代谢物的筛选

VIP 值通常用来反映 PLS-DA 模型变量的重要性, VIP 值越大表示该变量越重要<sup>[12]</sup>。根据 PLS-DA 模型进行差异代谢物的筛选, 共筛选出 57 种差异代谢物 ( $VIP \geq 1, P < 0.05, FDR < 0.05$ ), 如表 2 所示, 差异代谢物的热图如图 5 所示。

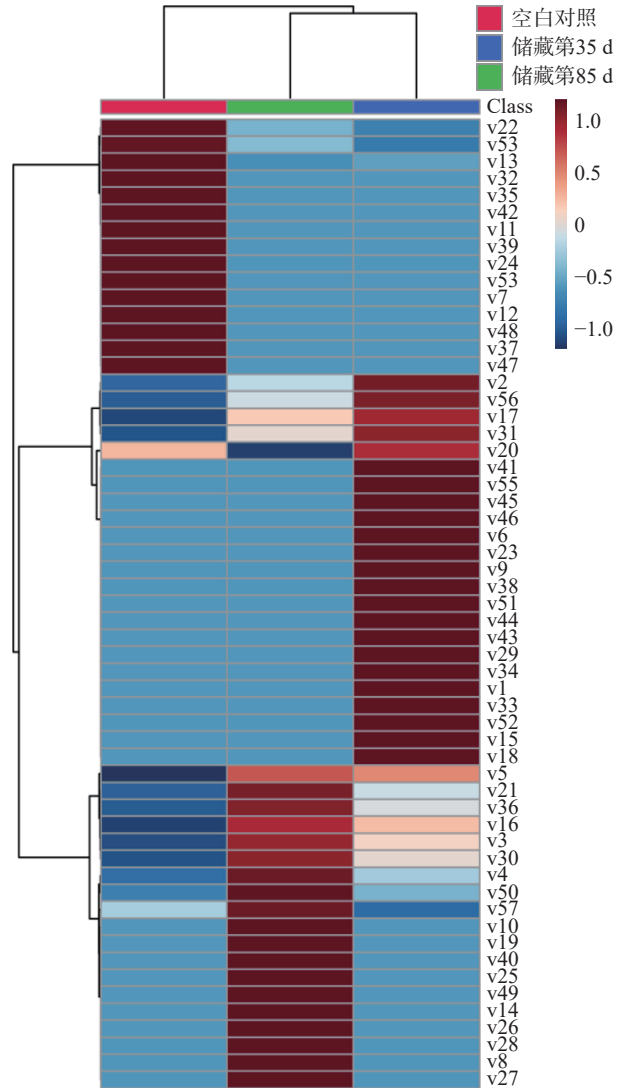


图 5 不同样品组的差异代谢物热图

Fig.5 Heat map of different metabolites in different sample groups

通过表 2 可知, 腊八蒜在储藏过程中的差异代谢物主要包括酸类(6 种)、醇类(12 种)、胺类(6 种)、糖苷类(1 种)、酯类(3 种)、醚类(4 种)、芳香类物(13 种)、烯炔类(2 种)、其他(10 种), 共 57 种。利用热图聚类分析可将储藏第 35 和第 85 d 的样品组聚为一类, 验证了 2.3 中 PCA 的结论。

酸类代谢物在储藏过程中总相对含量不断升高。腊八蒜在储藏过程中是浸泡在酸性环境中的, 而且腊八蒜口感偏酸, 因此酸类化合物在其中发挥着重要作用。从图 5 中可以看出, 随着储藏时间的延长, VIP 值较大的饱和脂肪酸(硬脂酸(v3)、棕榈酸

表2 不同储藏时间腊八蒜中差异代谢物筛选结果

Table 2 Screening results of different metabolites in Laba garlic with different storage time

编号	化合物名称	VIP值	P值	FDR值
v3	硬脂酸	1.83	2.20E-07	1.59E-06
v4	棕榈酸	1.73	8.84E-06	4.75E-05
v5	苯甲酸	1.71	1.19E-33	5.76E-32
v51	草酸	1.06	2.05E-05	8.48E-05
v29	苯甲酸衍生物	1.18	1.90E-05	8.09E-05
v12	二亚油酸	1.48	7.16E-15	2.08E-13
	酸类6种			
v2	肌醇	1.91	2.69E-12	4.88E-11
v10	丁四醇	1.49	2.82E-11	3.72E-10
v15	甘油	1.40	2.46E-11	3.57E-10
v24	5-己炔-1-醇	1.23	4.69E-08	4.25E-07
v28	硫二甘醇	1.19	1.69E-05	7.45E-05
v31	双甘油	1.17	1.28E-06	8.07E-06
v52	三异丙基硅烷醇	1.05	0.000582	0.001383
v53	别嘌呤醇	1.03	7.80E-05	0.00029
v16	蔗糖醇	1.40	1.18E-11	1.90E-10
v27	赤藓糖醇	1.20	8.78E-05	0.000301
v43	阿糖醇	1.09	0.00011	0.000355
v56	木糖醇	1.01	1.44E-05	6.52E-05
	醇类12种			
v1	9-十八碳烯酰胺	1.97	2.60E-46	3.77E-44
v6	棕榈酰胺	1.67	2.25E-12	4.67E-11
v9	硬脂酸酰胺	1.52	2.50E-08	2.58E-07
v18	4-乙基苯甲酰胺	1.33	7.31E-08	6.24E-07
v19	丙酰乙基苯丙胺	1.32	1.37E-06	8.26E-06
v41	三炔丙基胺	1.10	0.000143	0.000426
	胺类6种			
v55	阿拉伯糖苷	1.03	0.000775	0.001703
	糖苷类1种			
v33	棕榈酸甲酯	1.14	8.94E-05	0.000301
v36	棕榈酸甘油酯	1.13	0.001184	0.002345
v57	氯代磷酸二乙酯	1.00	0.007507	0.010269
	酯类3种			
v20	4-茴香醚	1.32	2.68E-05	0.000108
v22	儿茶酚次甲醚	1.24	0.000541	0.001329
v30	2-羟乙基丁基硫醚	1.17	4.80E-06	2.68E-05
v54	邻苯二甲醚	1.03	0.004743	0.006678
	醚类4种			
v8	噻吩	1.52	3.05E-08	2.94E-07
v11	苯腈	1.49	3.34E-19	1.21E-17
v14	2-丁基噻唑啉	1.43	1.59E-10	1.93E-09
v21	溴胺苯	1.25	2.39E-06	1.39E-05
v23	3-苯基-4-异恶唑-5-酮	1.24	1.04E-07	8.40E-07
v25	1-氢咪唑并[1,2-b]吡唑, 2,3-二氢	1.21	8.77E-05	0.000301
v32	2-卞吡啶	1.15	1.01E-06	6.62E-06
v42	2,2-乙烯双(1,3-二噻烷)	1.10	1.03E-05	5.14E-05
v44	四氢噻喃-4-酮	1.09	0.000504	0.001281
v46	6-氨基-2-甲基噻啉	1.08	0.000484	0.001275
v47	3-氨基二苯并呋喃	1.07	1.16E-05	5.58E-05
v48	2-氨基二苯并呋喃	1.07	1.30E-05	6.06E-05
v49	2-(乙胺基)-1,3,4-噻二唑	1.07	0.000144	0.000426
	芳香类14种			
v7	3-乙烯基-1,2-二硫环己-4-烯	1.66	9.19E-35	6.66E-33
v37	3-乙烯基-1,2-二硫环己-5-烯	1.12	1.24E-07	9.43E-07

续表 2

编号	化合物名称	VIP值	P值	FDR值
	烯烃类2种			
v13	5-甲基-2-糠醛	1.47	2.05E-08	2.28E-07
v17	丁烷	1.38	1.79E-12	4.34E-11
v26	戊烷	1.20	8.87E-05	0.000301
v34	醋酸甲脒	1.13	0.000521	0.001303
v35	2-乙烯基-1,3-二噻烷	1.13	9.68E-07	6.62E-06
v38	交叉三环化合物	1.11	0.000474	0.001272
v39	4-羟基苯乙腈	1.11	9.59E-06	4.97E-05
v40	丙腈	1.10	0.000161	0.000466
v45	乙二醇毒素	1.08	0.002991	0.005046
v50	三十一碳烷	1.06	0.006328	0.008738
	其他10种			

注: 化合物名称对应编号顺序全文同。

(v4)、苯甲酸(v5)的相对含量不断增加, 而不饱和脂肪酸(亚油酸(v12))是仅在空白组中出现, 草酸(v51)的相对含量则是先增加后减小。苯甲酸是植物的次级代谢产物, 在植物应对生物或非生物胁迫时也起着重要的作用<sup>[14]</sup>; 将大蒜置于酸性环境中后, 样品中的苯甲酸的相对含量会逐渐增高。

醇类作为广泛分布在植物中的生物标记物之一, 在腊八蒜的储藏过程中也发挥着重要作用<sup>[15]</sup>。由图 6 可知, 醇类化合物的相对含量在储藏第 35 d 时达到最高, 第 85 d 次之, 空白组最少。醇类中 VIP 值最高的是肌醇(v2, VIP=1.9), 其余的醇类物质 VIP 值都小于 1.5。肌醇是化学上非常稳定的极性小分子, 属于 B 族维生素, 是微生物的生长因子, 它和它的衍生物参与磷脂酸酰肌醇信号通路, 作为信号分子和关键代谢产物行使双重功能, 在植物的生长和胁迫应答过程中起协调作用, 并且具有一定的抗氧化作用<sup>[16]</sup>, 其合成需要消耗糖酵解的中心代谢物葡萄糖-6-磷酸。刘建<sup>[4]</sup>研究发现腊八蒜加工后期(腊八蒜由绿变黄), 糖醇含量明显下降而蒜醋液中糖醇含量升高, 本研究肌醇在储藏的第 85 d 样品中的相对含量显著低于储藏的第 35 d 的样品, 可能也与肌醇的逐

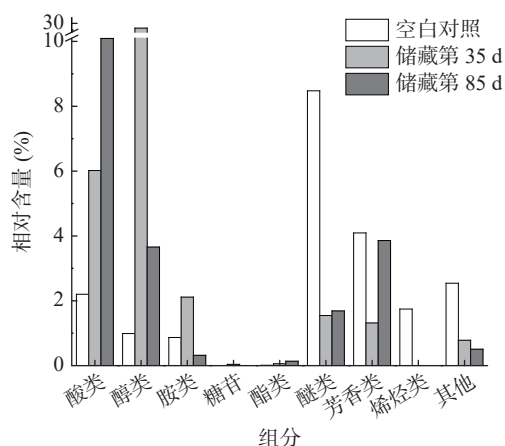


图 6 不同储藏时间样品中各类化合物累积相对含量变化  
Fig.6 Cumulative relative contents of various compounds in Laba garlic at different storage times

渐溶出有关。除此之外, 5-己炔-1-醇(v24)等不饱和醇只出现在空白组, 炔醇类化合物药理作用广泛, 具有抗肿瘤、抗炎等作用<sup>[17]</sup>。甘油(v10)、苏糖醇(v16)、赤藓糖醇(v27)在储藏的第 85 d 的相对含量最高, 阿糖醇(v43)、木糖醇(v56)在储藏的第 35 d 最高。

胺类化合物中多为酰胺化合物, 其相对含量整体呈现先上升后下降的趋势。VIP 值高于 1.5 的胺类化合物有: 9-十八碳烯酰胺(v1)、棕榈酰胺(v6)和硬脂酸酰胺(v9), 但上述 3 种物质均只出现在储藏第 35 d 的样品中。9-十八碳烯酰胺是一种玛咖酰胺类物质, 玛咖酰胺具有抗疲劳、抗抑郁、促进生育、保护神经、抗骨质疏松等作用<sup>[18]</sup>。

糖苷类和脂类化合物的相对含量较低, 醚类、芳香类、烯烃类以及其他类化合物的相对含量变化趋势相同。研究发现, VIP 大于 1.5 的代谢物中的 3-乙炔基-1,2-二硫环己-4-烯(v7)具有抗氧化活性<sup>[1]</sup>。

## 2.5 标志性代谢物的筛选

SVM-RFE 过程(后向特征选择过程)是通过不断的迭代, 每次删除  $w^2$  值最小的特征, 最终得到一个属性的排序, 然后通过交叉验证的方法确定最好的特征子集。与 PLS-DA(Filter 方法)相比, SVM-RFE(Wrapper 方法)分类正确率的结果一般会更好, 而且特征选择数量上会更少, 但在时间方面 SVM-RFE 也更耗时<sup>[19-21]</sup>。在 PLS-DA 筛选出的差异代谢物的基础上进行 SVM-RFE 标志性代谢物的筛选既提高了分类预测的正确率, 又缩短了时间, 提高了效率<sup>[22]</sup>。

由图 7 可知, 基于 SVM-RFE 算法, 将 PLS-DA 模型下的 57 个差异代谢物在 SVM 分类器中进一步拟合, 得出了均方根误差最小的 6 个代谢物, 分别为: 9-十八碳烯酰胺(v1)、肌醇(v2)、硬脂酸(v3)、棕榈酸(v4)、苯甲酸(v5)、5-己炔-1-醇(v24), 进入 SVM 分类器, 其误差为 0, 预测正确率为 100%。观察图 5 可知, 5-己炔-1-醇在空白组中的表达极其显著, 9-十八碳烯酰胺、肌醇在储藏的第 35 d 的样品中相对含量最高, 硬脂酸、棕榈酸、苯甲酸在储藏的第 85 d 的

样品中相对含量最高,6个代谢物分别标志了三个不同时间段。

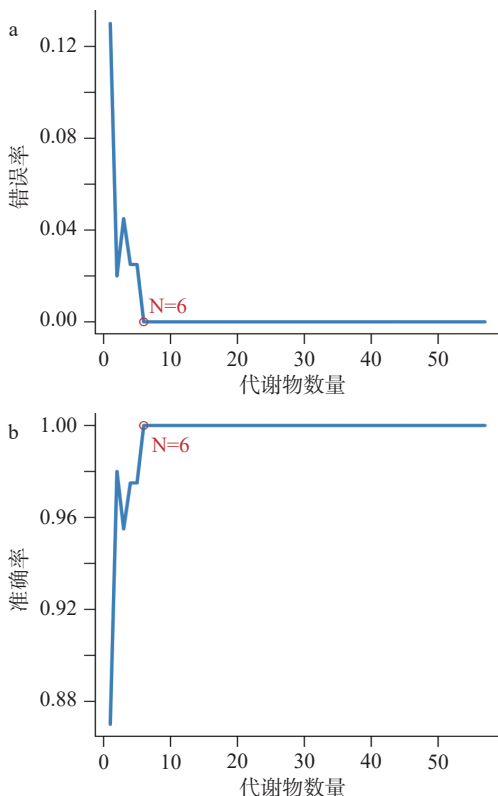


图7 通过 SVM-RFE 算法得出的特征代谢物选择图

Fig.7 A plot of characteristic metabolites selection by recursive feature elimination

注: a 代表模型错误率; b 代表模型准确率。

### 2.6 标志性代谢物的通路分析

KEGG(Kyoto Encyclopedia of Genes and Genomes)即京都基因及基因组百科全书,是汇集了基因组、生物化学产物、分子互作网络等信息的在线数据库<sup>[23]</sup>。通过 KEGG 对 6 种标志性代谢物进行途径富集分析<sup>[24]</sup>,揭示了空白组、储藏的第 35 和 85 d 的腊八蒜样品中标志性代谢物的代谢途径。从图 8 中可以看出,一共有 9 条代谢通路: 不饱和脂肪酸生物合成途径(Biosynthesis of unsaturated fatty acids)、脂肪酸生物合成(Fatty acid biosynthesis)、角质、亚伯碱和蜡的生物合成(Cutin, suberine and wax biosynthesis)、抗坏血酸和醛糖酸代谢(Ascorbate and aldarate metabolism)、脂肪酸链伸长系统(Fatty acid elongation)、磷脂酰肌醇信号系统(Phosphatidylinositol signaling system)、半乳糖代谢(Galactose metabolism)、磷酸肌醇代谢(Inositol phosphate metabolism)、脂肪酸降解(Fatty acid degradation)。其中 *P* 值小于 0.05 的代谢通路有不饱和脂肪酸生物合成途径、脂肪酸生物合成、角质、亚伯碱和蜡的生物合成途径。9 条代谢通路中与硬脂酸代谢相关的通路有 5 条,其中有 2 条同时与棕榈酸相关,另外 4 条通路与肌醇相关。

Tao 等<sup>[25]</sup>研究表明,腊八蒜的绿变是由醋对大蒜细胞的损伤所触发的,即在醋中浸泡会损伤大蒜的细胞膜,使得大蒜细胞内的蒜氨酸和蒜氨酸酶得以接触,在短时间内反应产生大蒜素,大蒜素和一些色素

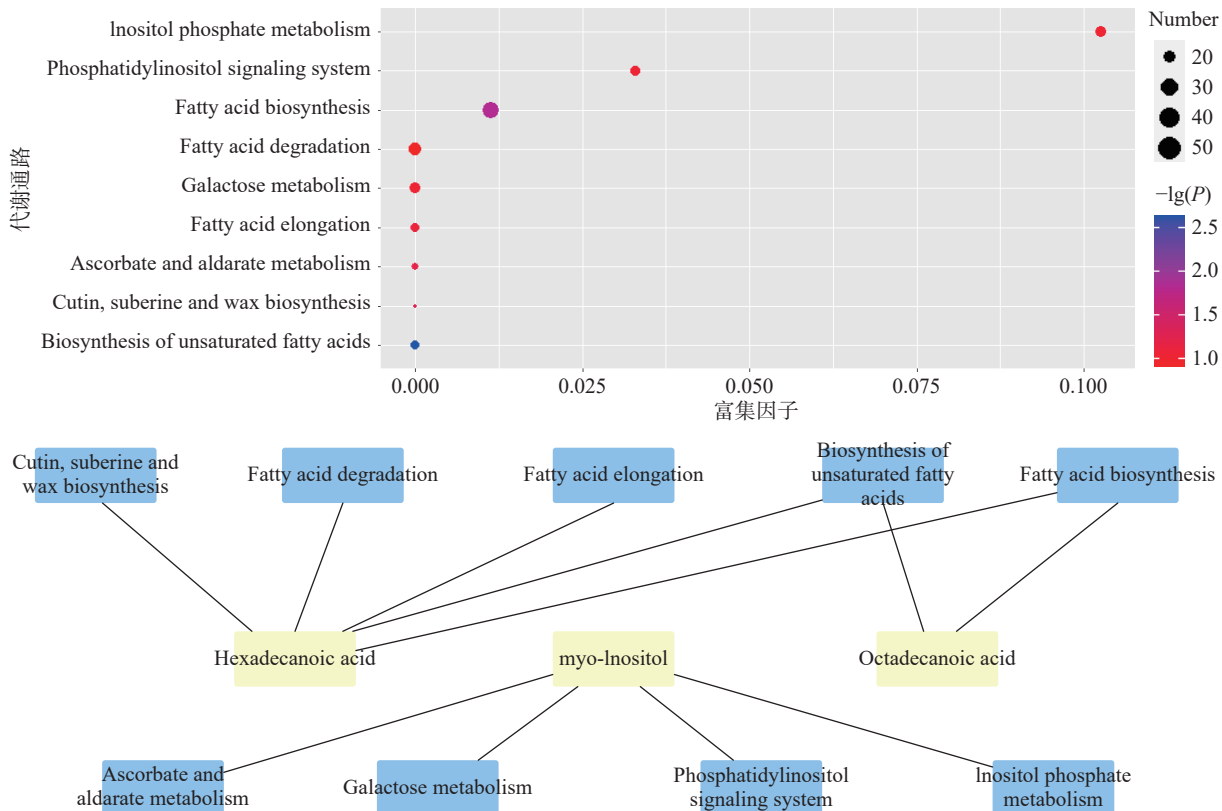


图8 代谢通路图

Fig.8 Metabolic pathway diagram



前体化合物继续反应,并引发一系列的酶促反应和非酶促反应,最终生成色素类物质,使腊八蒜表现为绿色。而程碧君<sup>[26]</sup>发现蒜氨酸通过 P53/AMPK/mTOR 信号通路激活自噬从而改善 1,3-DCP 诱导的脂质积累。同时蒋绪顺<sup>[27]</sup>等发现棕榈酸也可以诱导足细胞内的自噬增加和脂质积累。在 9 条代谢通路中的不饱和脂肪酸生物合成途径和脂肪酸生物合成、脂肪酸链伸长系统以及脂肪酸降解均与脂质积累相关,推测硬脂酸和棕榈酸等脂肪酸对绿变过程中有一定的影响。

### 3 结论

综上所述,基于 GC-MS 的非靶向代谢组学研究显示,在相同储藏条件下,不同储藏时期的腊八蒜具有明显的代谢物特征。经筛选得到的腊八蒜差异代谢物 57 种,为腊八蒜的生产和研发提供了重要的科学信息。此外,从 57 种差异代谢物种筛选出来的 6 种标志性代谢物表明挥发性物质(5-己炔-1-醇)是区别大蒜和腊八蒜的关键因素,而有机活性小分子肌醇、9-十八碳烯酰胺、硬脂酸、棕榈酸、苯甲酸是区分储藏的第 35 和 85 d 的关键因素,储藏至第 35 d 的腊八蒜呈黄绿色,第 85 d 时完全失去绿色。因此在储藏腊八蒜的过程中可以通过以上代谢成分的变化来进行储藏质量的评价,并用于指导腊八蒜的加工和生产。

### 参考文献

- [1] 刘平香. 基于代谢组学的大蒜生长贮藏过程中特征成分变化研究[D]. 北京: 中国农业科学院, 2020. [LIU P X. Study on changes of characteristic components in garlic during growth and storage based on metabolomics [D]. Beijing: Chinese Academy of Agricultural Sciences, 2020.]
- [2] 白冰, 纪淑娟, 王东梅, 等. 腊八蒜绿色素影响因素及护绿方法研究[J]. 食品工业科技, 2011, 32(2): 129-132. [BAI B, JI S J, WANG D M, et al. Methods and influence factors in green protecting of "laba" garlic[J]. Science and Technology of Food Industry, 2011, 32(2): 129-132.]
- [3] 王丹, 胡小松, 赵广华. 吡咯基丙氨酸的合成表征及对腊八蒜的增色作用[J]. 食品科学, 2011, 32(15): 1-4. [WANG D, HU X S, ZHAO G H. Synthesis and characterization of pyrrolyl alanine and its hyperchromic effect on Laba garlic pigments[J]. Food Science, 2011, 32(15): 1-4.]
- [4] 刘建. 基于代谢组学的大蒜品质评价研究[D]. 合肥: 合肥工业大学, 2018. [LIU J. Study on garlic quality evaluation based on metabolomics [D]. Hefei: Hefei University of Technology, 2018.]
- [5] SOLTANINEJAD M, YANG G, LAMBROU T, et al. Supervised learning based multimodal MRI brain tumour segmentation using texture features from supervoxels[J]. *Computer Methods and Programs in Biomedicine*, 2018, 157: 69-84.
- [6] DONG J F, XUE Q, CHEN T, et al. Machine learning approach to predict acute kidney injury after liver surgery[J]. *World Journal of Clinical Cases*, 2021, 9(36): 11255-11264.
- [7] LEITE V, REIS M R, PINTO F G. Untargeted metabolomics

reveals metabolic changes linked to bulb purpling in garlic (*Allium sativum* L.)[J]. *ACS Food Science and Technology*, 2021, 1(2): 242-248.

- [8] XIE H, ZHAO E, ZHANG Y. Identification of featured metabolism-related genes in patients with acute myocardial infarction[J]. *Disease Markers*, 2020, 2020(1-3): 1-9.
- [9] 崔保伟, 秦广利, 朱俊朋. 腊八蒜绿变过程中的品质变化研究[J]. 中国调味品, 2021, 46(9): 81-85. [CUI B W, QIN G L, ZHU J P. Study on the quality change of Laba garlic in the greening process[J]. *China Condiment*, 2021, 46(9): 81-85.]
- [10] 高旭东. 腊八蒜的化学成分、生物活性及相关机理研究[D]. 天津: 天津大学, 2020. [GAO X D. Studies on chemical constituents, biological activities and related mechanisms of Laba garlic [D]. Tianjin: Tianjin University, 2020.]
- [11] 高歌, 鹿雪莉, 刘海华, 等. 基于 GC-MS-O 香气成分分析和多元统计分析的柚子品种鉴别[J]. 中国食品学报, 2020, 20(5): 283-292. [GAO G, PANG X L, LIU H H, et al. Identification of grapefruit varieties based on GC-MS-O aroma composition analysis and multivariate statistical analysis[J]. *Journal of Chinese Institute of Food Science and Technology*, 2020, 20(5): 283-292.]
- [12] 冯玉超. 基于 GC-MS 技术的黑龙江省主产区水稻 (*Oryza sativa* L.) 代谢组学分析[D]. 大庆: 黑龙江八一农垦大学, 2019. [FENG Y C. Analysis of metabonomics of rice (*Oryza sativa* L.) in the main production area of Heilongjiang Province based on GC-MS technology [D]. Daqing: Heilongjiang Bayi Agricultural University, 2019.]
- [13] 陈贤双, 李科, 李震宇, 等. 基于 SPME-GC-MS 结合多元统计的不同产地黄芪挥发性成分差异分析[J]. 药学学报, 2020, 55(5): 979-986. [CHEN X S, LI K, LI Z Y, et al. Analysis of the volatile components of *Astragali radix* from different habitats by SPME-GC-MS combined with multivariate statistical analysis[J]. *Acta Pharmaceutica Sinica*, 2020, 55(5): 979-986.]
- [14] 张跃举. 苯甲酸激励的小球藻抗倒酯合成与细胞代谢调节[D]. 沈阳: 东北师范大学, 2021. [ZHANG Y J. Benzoic acid stimulates the synthesis of chlorella antideposide and the regulation of cell metabolism [D]. Shenyang: Northeast Normal University, 2021.]
- [15] 段中华, 全小龙, 乔有明, 等. 植物与土壤中醇类化合物对高寒草甸退化指示作用初探[J]. 草业科学, 2018, 35(5): 969-977. [DUAN Z H, QUAN X L, QIAO Y M, et al. Alcohols in plants and soil as indicators of alpine meadow degradation[J]. *Pratacultural Science*, 2018, 35(5): 969-977.]
- [16] 乔腾生. 肌醇提高单针藻 *Monoraphidium* sp. QLY-1 在糖蜜废醪液中油脂积累的过程及机制[D]. 昆明: 昆明理工大学, 2021. [QIAO T S. The process and mechanism of inositol increasing the oil accumulation of *Monoraphidium* sp. QLY-1 in molasses waste mash [D]. Kunming: Kunming University of Technology, 2021.]
- [17] 徐念智, 李兵. 中药中多炔类化学成分药理研究进展[J]. 中医药导报, 2021, 27(9): 137-141, 163. [XUN Z, LI B. Research progress on pharmacology of polyacetylenes in traditional Chinese medicine[J]. *Guiding Journal of Traditional Chinese Medicine and Pharmacology*, 2021, 27(9): 137-141, 163.]
- [18] 付玲, 陈芮. 玛咖酰胺及玛咖烯的研究进展[J]. 化工管理,

- 2019(1): 55–56. [ FU L, CHEN R. Research progress of macamides and macalene[J]. *Chemical Enterprise Management*, 2019(1): 55–56. ]
- [ 19 ] 严康. 基于支持向量机的特征选择算法研究[D]. 大连: 大连理工大学, 2010. [ YAN K. Research on feature selection algorithm based on support vector machine [D]. Dalian: Dalian University of Technology, 2010. ]
- [ 20 ] LIU Z, MI M, LI X, et al. A lncRNA prognostic signature associated with immune infiltration and tumour mutation burden in breast cancer[J]. *Journal of Cellular and Molecular Medicine*, 2020, 24(21): 12444–12456.
- [ 21 ] SANKAR, MAHADEVAN, SIRISH, et al. Analysis of metabolomic data using support vector machines[J]. *Analytical Chemistry*, 2008, 80(19): 7562–7570.
- [ 22 ] 赵燕妮, 张坤, 许牡丹, 等. 基于 GC-MS 代谢组学技术的不同品种猕猴桃果实化学成分差异性研究[J]. *陕西科技大学学报*, 2022, 40(1): 45–50. [ ZHAO Y N, ZHANG K, XU M D, et al. Study on the chemical constituents of different kiwifruit cultivars based on GC-MS metabolomics[J]. *Journal of Shaanxi University of Science & Technology*, 2022, 40(1): 45–50. ]
- [ 23 ] KANEHISA M, GOTO S, FURUMICHI M, et al. KEGG for representation and analysis of molecular networks involving diseases and drugs[J]. *Nucleic Acids Research*, 2010, 38: 355–360.
- [ 24 ] KHATRI P, SIROTA M, BUTTE A J, et al. Ten years of pathway analysis: Current approaches and outstanding challenges [J]. *Plos Computational Biology*, 2012, 8(2): e1002375.
- [ 25 ] TAO D, ZHOU B, ZHANG L, et al. 'Laba' garlic processed by dense phase carbon dioxide: The relation between green colour generation and cellular structure, alliin consumption and alliinase activity[J]. *Journal of the Science of Food & Agriculture*, 2016, 96(9): 2969–2975.
- [ 26 ] 程碧君. 蒜氨酸通过激活自噬改善 1,3-二氯-2-丙醇诱导的 HepG2 细胞脂质积聚的机制研究[D]. 长春: 吉林大学, 2019. [ CHENG B J. Study on the mechanism of allicin improving lipid accumulation in HepG2 cells induced by 1,3-dichloro-2-propanol by activating autophagy [D]. Changchun: Jilin University, 2019. ]
- [ 27 ] 蒋绪顺. 自噬在棕榈酸诱导的足细胞凋亡中的作用及机制[D]. 重庆: 重庆医科大学, 2017. [ JIANG X S. The role and mechanism of autophagy in palmitic acid-induced podocyte apoptosis [D]. Chongqing: Chongqing Medical University, 2017. ]