

## 基于修饰组和探针分子的重要途径代谢物筛选和注释新方法

李在芳<sup>1,2#</sup>, 郑福建<sup>1,2#</sup>, 夏悦怡<sup>1,2</sup>, 张秀琼<sup>1,2</sup>, 王鑫欣<sup>1,2</sup>,  
赵春霞<sup>1,2</sup>, 赵欣捷<sup>1,2</sup>, 路鑫<sup>1,2\*</sup>, 许国旺<sup>1,2</sup>

(1. 中国科学院大连化学物理研究所, 中国科学院分离分析化学重点实验室,  
辽宁省代谢组学重点实验室, 辽宁 大连 116023; 2. 中国科学院大学, 北京 100049)

**摘要:**植物次生代谢物在抵御生物/非生物胁迫、生物间互作以及信息传递等方面发挥重要作用,次生代谢途径解析对植物分子育种、天然产物合成等方面具有重要意义。液相色谱-高分辨串联质谱(LC-HRMS/MS)为次生代谢物鉴定及途径表征提供了技术手段。非靶向LC-HRMS/MS方法可获得丰富的质谱信号,包括一级质谱和二级质谱(MS, MS/MS),但受质谱数据库规模以及次生代谢物复杂性的制约,次生代谢物注释十分困难。该研究以玉米叶片中苯丙烷途径代谢物为例,发展用于非靶向代谢组数据中重要途径代谢物的高效筛选和注释新方法。首先,利用公共代谢途径数据库及文献获取参与苯丙烷代谢途径的61种修饰反应类型,进而从非靶向实验数据中筛选出修饰代谢组。其次,获取开源串联质谱数据中的苯丙烷类化合物作为探针分子,构建探针分子质谱数据库。将探针分子与修饰代谢组共建分子网络,锁定目标途径代谢物并注释结构。该方法在正、负离子模式下分别筛选出玉米叶片中392个和417个苯丙烷途径候选代谢物,去冗余后共注释出129个代谢物,涉及苯丙烷代谢的主要分支途径,如黄酮途径的8个类黄酮、19个氧苷类黄酮和32个碳苷类黄酮,31个羧基肉桂酸途径代谢物以及22个木脂素途径代谢物;其中26个在PubChem和SciFinder数据库中未见收录。该研究利用探针分子结合修饰组可快速锁定途径代谢物,且有助于快速、准确的网络传播注释,可显著提高目标途径代谢物筛选与注释效率,为植物次生代谢途径的深入解析提供分析手段。

**关键词:**液相色谱-高分辨串联质谱;修饰代谢组;次生代谢物;探针分子;注释

**中图分类号:** O658 **文献标识码:** A **文章编号:** 1000-8713(2022)09-0788-09

## A novel method for efficient screening and annotation of important pathway-associated metabolites based on the modified metabolome and probe molecules

LI Zaifang<sup>1,2#</sup>, ZHENG Fujian<sup>1,2#</sup>, XIA Yueyi<sup>1,2</sup>, ZHANG Xiuqiong<sup>1,2</sup>, WANG Xinxin<sup>1,2</sup>,  
ZHAO Chunxia<sup>1,2</sup>, ZHAO Xinjie<sup>1,2</sup>, LU Xin<sup>1,2\*</sup>, XU Guowang<sup>1,2</sup>

(1. CAS Key Laboratory of Separation Science for Analytical Chemistry, Liaoning Province Key Laboratory of Metabolomics, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China)

**Abstract:** Plants produce a wide variety of secondary metabolites in the process of evolution. Secondary metabolites have highly diverse structures due to the modification of the basic skeletons of metabolites. They are required for interaction with the environment and are produced in response to abiotic/biotic stress. Characterization of secondary metabolic pathways is significant to plant molecular breeding and natural product biosynthesis. The liquid chromatography-

收稿日期: 2022-03-17

# 共同第一作者.

\* 通讯联系人. Tel: (0411) 84379532, E-mail: luxin001@dicp.ac.cn.

基金项目: 国家自然科学基金项目(21874132, 22074145); 中国科学院大连化学物理研究所科研创新基金项目(DICP I202004).

**Foundation item:** National Natural Science Foundation of China (Nos. 21874132, 22074145); Innovation Program of Science and Research from Dalian Institute of Chemical Physics, Chinese Academy of Sciences (No. DICP I202004).

high resolution tandem mass spectrometry (LC-HRMS/MS) is one of the major techniques for untargeted metabolomics study. The LC-HRMS/MS method could detect tens of thousands of metabolic features and provide abundant structural information. It has been widely used in the discovery and characterization of the secondary metabolome. However, due to the largely diverse structure and limited records in the mass spectral library, the annotation of the secondary metabolome is very difficult.

To address the analytical challenges associated with the vast structural diversity and the large numbers of secondary metabolites, particularly those previously unknown structural metabolites, a novel method for the efficient characterization of pathway-associated metabolites was developed. Modification reactions and MS/MS spectral information were collected using the metabolic pathways database and mass spectral library. Screening and annotation of metabolites involved in phenylpropanoid metabolism in maize leaves were used as an example. First, a database of modified groups was established via pathway-associated modifications from open access metabolic pathway database and literature. Here, pathway databases included the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Plant Metabolic Pathways (PlantCyc). A total of 61 modification types were enrolled, including 10 generic and 51 pathway-specific modifications. Modified metabolomes were filtered from untargeted LC-HRMS/MS metabolomics data. Next, MS/MS spectra of the pathway-associated compounds (probe molecules) were collected in the Global Natural Products Social Molecular Networking (GNPS) MS/MS spectral library. The MS/MS of compounds assigned to chemical classes of phenylpropanoids were kept. An MS/MS spectral database of the probe molecules was constructed. It included 2 677 spectra of 1 542 phenylpropanoid compounds in the positive mode and 814 spectra of 661 phenylpropanoid compounds in the negative mode. Then, an MS/MS molecular network was generated by modified metabolome and probe molecules. The clusters comprising both probe molecules and modified metabolites were kept. To explore more previously unknown structural metabolites, the clusters with one more pathway-specific modified metabolite were retained even though they didn't contain any probe molecule. A total of 392 and 417 phenylpropanoid pathway-related metabolic metabolites were obtained in positive and negative ion modes, respectively. The pathway-associated metabolites were annotated based on the propagation of the molecular network. For the metabolites within the co-cluster, annotations were performed using the probe molecules as the initial seed. The modification group's substructure information was used for network propagation annotation. For the clusters containing only pathway-specific modified metabolites, the annotation is similar to the above process if identified nodes were present within the cluster. Otherwise, de novo annotation was manually executed based on substructure information. Finally, 129 unique metabolites were annotated after integration and removal of redundancy. Ten annotated metabolites were validated using commercially available or synthesized reference compounds. The other annotation results were validated using predicted chemical classes, *in silico* MS/MS, and predicted retention time. They are mainly involved in the downstream branch of

引用本文:李在芳,郑福建,夏悦怡,张秀琼,王鑫欣,赵春霞,赵欣捷,路鑫,许国旺. 基于修饰组和探针分子的重要途径代谢物筛选和注释新方法. 色谱,2022,40(9):788-796.

LI Zaifang, ZHENG Fujian, XIA Yueyi, ZHANG Xiuqiong, WANG Xinxin, ZHAO Chunxia, ZHAO Xinjie, LU Xin, XU Guowang. A novel method for efficient screening and annotation of important pathway-associated metabolites based on the modified metabolome and probe molecules. Chinese Journal of Chromatography, 2022, 40(9): 788-796.

phenylpropanoid pathways, including the flavonoid pathway (8 flavonoids, 19 flavonoid *O*-glycosides, 32 flavonoid *C*-glycosides), the hydroxycinnamic acid pathway (31 hydroxycinnamic acids and derivatives), and the lignan pathway (22 neo-lignans/lignan/lignan glycosides). All the annotated structures were searched against the PubChem and SciFinder databases. Among them, 26 metabolites were previously unreported in both the databases. In this study, the pathway-associated metabolites could be quickly discovered and annotated by the integration of probe molecules and modified metabolome. It provides a method for the in-depth study of the phenylpropanoid pathway.

**Key words:** liquid chromatography-high resolution tandem mass spectrometry (LC-HRMS/MS); modified metabolome; secondary metabolites; probe molecule; annotation

植物在进化进程中,为适应复杂的生存环境及满足发育需要,产生出种类繁多、数量巨大、结构非常丰富的次生代谢物,在抵御生物/非生物胁迫、生物间互作以及信息传递等方面发挥重要作用<sup>[1]</sup>,次生代谢途径解析对植物分子育种及天然产物生物合成等具有重要意义。苯丙烷代谢是植物最重要的次生代谢合成途径之一,与植物应激诱导有着密切关系<sup>[2,3]</sup>。苯丙烷代谢有多个分支途径,如黄酮途径、木质素途径、木脂素途径、羟基肉桂酸酰胺途径等;骨架结构经多种后修饰,如羟基化、糖基化、乙酰化、异戊二烯化、硫酸化和甲基化等,产生了复杂多样的苯丙烷途径代谢物。

基于液相色谱-高分辨串联质谱(LC-HRMS/MS)的代谢组学分析技术为次生代谢物鉴定及途径阐释提供了手段<sup>[4-7]</sup>。如Wen等<sup>[8]</sup>基于非靶向LC-HRMS/MS代谢组学技术鉴别了玉米中29个类黄酮,通过整合基因组学、转录组学信息,揭示了玉米黄酮生物合成的遗传学基础。牟红梅等<sup>[9]</sup>基于超高效液相色谱-串联质谱的非靶向代谢组学分析研究了成熟期茄梨和红茄梨果皮代谢物差异,发现差异代谢物涉及黄酮代谢、氨基酸代谢、苯丙烷代谢以及苯丙烷分支途径黄酮代谢等。胡永丹等<sup>[10]</sup>基于超高效液相色谱-高分辨质谱联用技术分析茶树花化学成分,采用氮规则、质量亏损和特征子离子筛选目标化学成分,从7个茶树花样本中共鉴定出137个化合物。非靶向代谢组学采集了非常丰富的质谱信号,包括一级质谱和二级质谱(MS & MS/MS),代谢组鉴定多采用质谱数据库搜索方式<sup>[11]</sup>。尽管全球天然产物社会分子网络(Global Natural Products Social Molecular Networking, GNPS)数据库已收录免费共享的83 314张MS/MS谱图,但与复杂多样的植物代谢组,特别是次生代谢组相比,质谱

库的覆盖范围仍不足,搜库鉴定能力有限<sup>[12]</sup>。利用非靶向代谢组实验数据,构建基于质谱相似性(MS/MS similarity)的分子网络(molecular networking, MN)<sup>[13,14]</sup>,可人工推断注释数据库无法鉴定的结构类似物<sup>[15]</sup>。如Clements等<sup>[16]</sup>采用代谢组学结合分子网络的方法,推测了一种新型开环赛氏菌缩肽(serratamolide)类似物的结构。但该方法不能有效筛选目标途径相关分子簇,且方法注释能力主要依赖分子簇内可搜库注释的种子节点。植物次生代谢物结构存在大量修饰基团,在MS/MS中常以中性丢失形式体现<sup>[17]</sup>。植物代谢途径数据库包含大量代谢反应信息,如植物代谢途径(Plant Metabolic Pathways, PlantCyc, <https://www.plantcyc.org/>)<sup>[18]</sup>收录了5 234个反应,京都基因与基因组百科全书(Kyoto Encyclopedia of Genes and Genomes, KEGG, <https://www.genome.jp/kegg/>)收录了11 744个生化反应。将代谢途径中的修饰反应用于非靶向数据中代谢特征的筛选,可有助于提高途径代谢物的发现效率<sup>[19]</sup>。

为此,本研究针对次生代谢物种类多、结构复杂,且存在大量未知结构代谢物的分析挑战,以玉米叶片苯丙烷途径代谢物为例,利用代谢途径的修饰反应以及现有质谱数据库,发展用于非靶向LC-HRMS/MS代谢组数据中重要途径代谢物筛选和注释的新方法,为次生代谢物的注释和途径解析提供新思路。

## 1 实验部分

### 1.1 仪器、试剂与材料

ACQUITY UPLC超高效液相色谱系统购自美国Waters公司。Triple TOF 5600+飞行时间质谱(TOF-MS)购自美国AB SCIEX公司。超纯水由购

自美国 Billerica 公司的 Milli-Q 系统纯化制备。乙腈(HPLC 级)和甲醇(HPLC 级)购自德国 Merck 公司。甲酸(纯度 98%)和碳酸氢铵购自中国 J & K Scientific 有限公司。

## 1.2 样品预处理

玉米叶片代谢组提取:准确称取玉米新鲜叶片组织冻干粉 50 mg,置于 1.5 mL 离心管中。加入 1.0 mL 80% (v/v) 甲醇水提取剂,涡旋提取 5 min,在 4 °C 条件下,以 21 500 g 转速离心 10 min。取 700  $\mu$ L 上清液,放入真空离心浓缩仪冻干。冻干样品加入 100  $\mu$ L 80% (v/v) 甲醇水复溶,复溶液在 4 °C 条件下,以 21 500 g 转速离心 10 min。取上清液用于仪器分析。

## 1.3 非靶向代谢组学分析

正离子模式下的液相色谱条件 Waters ACQUITY BEH C<sub>18</sub> 色谱柱(100 mm×2.1 mm, 1.7  $\mu$ m);柱温:50 °C;进样器温度:4 °C;流动相 A:0.1% (v/v) 甲酸水溶液;流动相 B:0.1% (v/v) 甲酸乙腈溶液;流速:0.35 mL/min;进样量:5  $\mu$ L。洗脱梯度:0~1.0 min, 5% B; 1.0~24.0 min, 5% B~100% B; 24.0~28.0 min, 100% B; 28.0~28.1 min, 100% B~5% B; 28.1~30.0 min, 5% B。

负离子模式下的液相色谱条件 Waters ACQUITY HSS T<sub>3</sub> 色谱柱(100 mm×2.1 mm, 1.8  $\mu$ m);柱温:50 °C;进样器温度:4 °C;流动相 A:6.5 mmol/L 碳酸氢铵水溶液;流动相 B:含 6.5 mmol/L 碳酸氢铵的 95% (v/v) 甲醇水溶液;流速:0.35 mL/min;进样量:5  $\mu$ L。洗脱梯度:0~1.0 min, 2% B; 1.0~18.0 min, 2% B~100% B; 18.0~22.0 min, 100% B; 22.0~22.1 min, 100% B~2% B; 22.1~25.0 min, 2% B。

质谱条件 离子源:电喷雾电离(ESI)源,采用正、负离子模式检测;扫描方式:一级全扫描质量范围  $m/z$  50~1 250;二级数据依赖型扫描(Top 15)质量范围  $m/z$  50~1 250;碰撞能量:15、30 和 45 eV;离子源温度:500 °C;电喷雾电压:4 500 V;气帘气压强:0.241 MPa;雾化气(gas 1)压强:0.345 MPa;加热气(gas 2)压强:0.345 MPa。

## 1.4 数据处理

### 1.4.1 原始数据预处理

采用 MarkerView 1.2.1 软件处理质谱原始数据,获得包含代谢特征保留时间、质荷比和峰强度的质谱峰列表。采用 ProteoWizard 3.0.10240 软件将质

谱原始数据文件转化为 .mgf 格式的二级质谱文件。

### 1.4.2 修饰代谢组筛选

首先,从代谢途径数据库收集修饰反应,将修饰类型的名称、分子式和精确相对分子质量整理至 .csv 格式的文件中,生成修饰类型列表。将非靶向代谢组数据的质谱峰列表、二级质谱文件,以及修饰类型列表作为输入文件,运行自编程序包 Modified-MetMRM。设置质谱峰列表与二级质谱文件的匹配参数为质量精度  $15\times 10^{-6}$ ,保留时间窗口 12 s。基于修饰类型列表获取非靶向代谢组数据中修饰代谢组的匹配参数为质量窗口  $15\times 10^{-6}$ ,子离子最低绝对强度为 100。离子融合参数为质量窗口  $15\times 10^{-6}$ ,保留时间窗口 12 s。

### 1.4.3 探针分子数据库构建

下载 GNPS (<https://gnps.ucsd.edu/>) 中的 3 个质谱数据集(GNPS Library, NIH Natural Products Library Round 1 和 NIH Natural Products Library Round 2)。采用自编代码收集 3 个数据库化合物的 SMILES,并在 ChemDes 平台 (<http://www.scbdd.com/convert/convert/>) 将其转化为 InChIKey。利用 ClassyFire (<https://cfb.fiehnlab.ucdavis.edu/>) 工具获得化合物的化学分类,保留其中苯丙烷类化合物,获取化合物的相关信息,包括采集 MS/MS 的仪器类型、SMILES 和二级质谱图等。

### 1.4.4 分子网络构建

分子网络由开源工具 GNPS 平台完成。建网参数:不少于 6 个子离子匹配,MS/MS 相似度阈值为 0.7;母离子质量精度阈值:0.01 Da;子离子质量精度阈值:0.02 Da;单个连通网络节点数最大值:500;单个节点最大相邻节点数:50。分子网络可视化由软件 Cytoscape 3.8.0 实现。

### 1.4.5 代谢物定性

采用基于 SMRT 数据集构建的 GNN-RT 模型<sup>[20]</sup>,通过标准品迁移学习预测本实验色谱条件下的保留时间;使用开源工具 CFM-ID 4.4.3 (<https://hub.docker.com/r/wishartlab/cfmid>) 预测代谢物的虚拟(*in silico*)二级质谱;二级质谱相似度计算采用谱嫡算法<sup>[21]</sup>。代谢物化学类别预测采用 SIRI-US 4<sup>[22]</sup>中的 CANOPUS<sup>[12]</sup>完成,使用默认参数。

## 2 结果与讨论

### 2.1 方法框架

方法总体框架如图 1 所示,途径代谢物的筛选

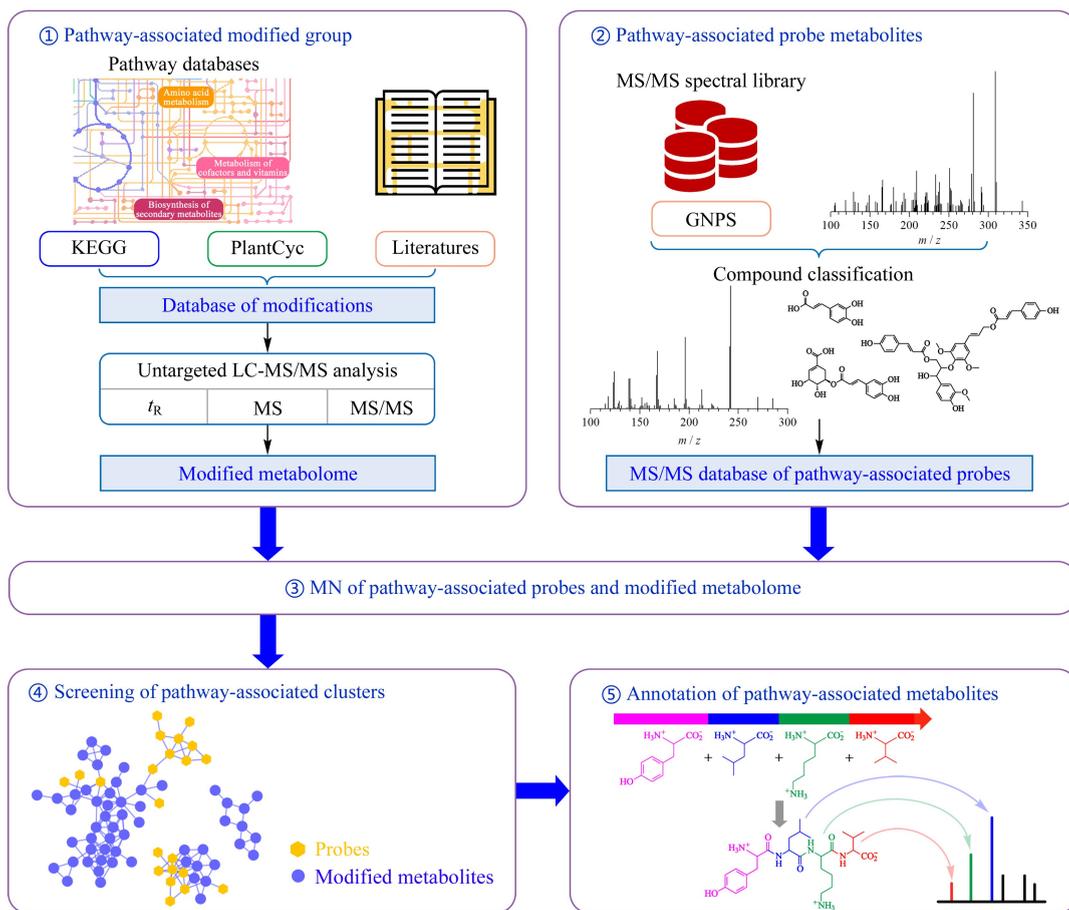


图 1 重要途径代谢物的筛选与注释流程

Fig. 1 Workflow of screening and annotation of pathway-associated metabolites

KEGG; the Kyoto Encyclopedia of Genes and Genomes; MN: molecular networking; GNPS: Global Natural Products Social Molecular Networking.

与注释步骤主要包括:1) 基于公共代谢途径数据库及文献报道,收集参与途径的修饰反应,构建修饰基团质谱数据库,进而从非靶向实验数据中筛选修饰代谢组(含有修饰基团的代谢物);2) 从开源质谱数据库中收集目标途径代谢物及其类似物作为探针分子,构建探针分子串联质谱数据库;3) 将探针分子与修饰代谢组共建分子网络;4) 筛选目标途径代谢物分子簇;5) 基于探针分子结构、修饰基团、子结构信息等注释代谢物。

## 2.2 修饰代谢组筛选

利用开源代谢途径知识库,如:KEGG、PlantCyc 以及文献报道<sup>[23-25]</sup>,收集参与苯丙烷途径的代谢反应,提取修饰基团。表 1 给出了收集到的 61 种修饰反应类型,包括 11 种通用修饰类型(甲基化、甲氧基化、羟基化、甲氨基化、乙酰化、羧基化、硫酸化、丙二酰化、戊糖基化、脱氧己糖基化和己糖基化)和 50 种途径特异性修饰类型(13 种胺结合、5

种羟基肉桂酰结合、13 种酸结合、18 种醇结合和 1 种异戊二烯化修饰)。

从玉米叶片正、负离子模式非靶向代谢组学数据中筛选到 60 种修饰类型,对应 1 385 个和 1 412 个代谢物。其中,正、负离子模式下分别有 32.27% (447/1 385) 和 26.06% (368/1 412) 的代谢物含有两种及以上修饰基团,通用修饰类型占比 32.31% (正离子模式) 和 34.85% (负离子模式);正离子模式下,特异性修饰类型主要是胺结合(7.76%)、羟基肉桂酰结合(8.09%)、酸结合(13.91%) 和醇结合(36.28%) 修饰;异戊二烯化修饰占比较低,仅为 1.66%。负离子模式与正离子模式类似,上述特异性修饰及异戊二烯化修饰的占比分别为 3.94%、13.62%、19.21%、28.12% 和 0.26%。

## 2.3 探针分子串联质谱数据库构建

基于 ClassyFire<sup>[26]</sup> 化学分类从 GNPS 数据库中包含天然产物较多的 3 个质谱数据集(GNPSLibrary、

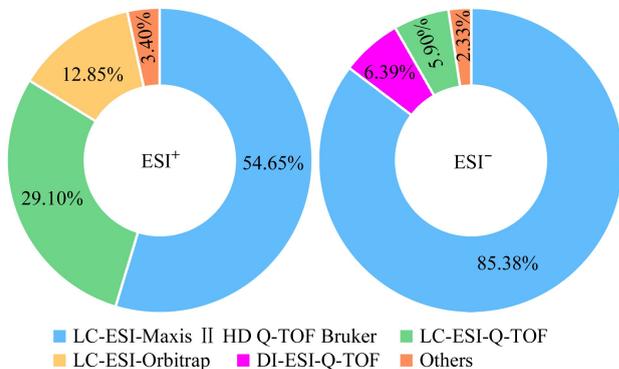
表 1 参与苯丙烷代谢的后修饰反应  
Table 1 Modifications involved in phenylpropanoids biosynthesis

Modified type	Neutral loss	Generic or specific	Modified type	Neutral loss	Generic or specific
Methylation	CH <sub>2</sub>	generic	Phenylacetyl coupling	C <sub>8</sub> H <sub>6</sub> O	specific
Methoxylation	CH <sub>2</sub> O	generic	Hydroxybenzoate	C <sub>7</sub> H <sub>4</sub> O <sub>2</sub>	specific
Hydroxylation	H <sub>2</sub> O	generic	<i>p</i> -Hydroxybenzoylation	C <sub>7</sub> H <sub>6</sub> O <sub>2</sub>	specific
Methyl ammonia	CH <sub>3</sub> NH <sub>2</sub>	generic	Tartarate	C <sub>4</sub> H <sub>4</sub> O <sub>5</sub>	specific
Acetylation	CH <sub>2</sub> CO	generic	Dihydroxybenzoic acid coupling	C <sub>7</sub> H <sub>4</sub> O <sub>3</sub>	specific
Carboxylation	COO	generic	Hydroxyadipic acid coupling	C <sub>6</sub> H <sub>8</sub> O <sub>4</sub>	specific
Sulfation	SO <sub>3</sub>	generic	Vanillate	C <sub>8</sub> H <sub>6</sub> O <sub>3</sub>	specific
Malonyl	C <sub>3</sub> H <sub>2</sub> O <sub>3</sub>	generic	3-Dehydroshikimic acid coupling	C <sub>7</sub> H <sub>6</sub> O <sub>4</sub>	specific
Pentosylation	C <sub>5</sub> H <sub>8</sub> O <sub>4</sub>	generic	Shikimic acid coupling	C <sub>7</sub> H <sub>8</sub> O <sub>4</sub>	specific
Deoxyhexosylation	C <sub>6</sub> H <sub>10</sub> O <sub>4</sub>	generic	Quinic acid coupling	C <sub>7</sub> H <sub>10</sub> O <sub>5</sub>	specific
Hexosylation	C <sub>6</sub> H <sub>10</sub> O <sub>5</sub>	generic	Syringate	C <sub>9</sub> H <sub>8</sub> O <sub>4</sub>	specific
Putrescine	C <sub>4</sub> H <sub>12</sub> N <sub>2</sub>	specific	Glycerol	C <sub>3</sub> H <sub>6</sub> O <sub>2</sub>	specific
Cadaverine	C <sub>5</sub> H <sub>14</sub> N <sub>2</sub>	specific	Quinol	C <sub>6</sub> H <sub>4</sub> O	specific
Agmatine	C <sub>5</sub> H <sub>14</sub> N <sub>4</sub>	specific	Hydroxybenzyl alcohol	C <sub>7</sub> H <sub>6</sub> O	specific
Tyramine	C <sub>8</sub> H <sub>11</sub> NO	specific	Hydroxyquinol	C <sub>6</sub> H <sub>4</sub> O <sub>2</sub>	specific
Spermidine	C <sub>7</sub> H <sub>19</sub> N <sub>3</sub>	specific	Vanillyl alcohol	C <sub>8</sub> H <sub>8</sub> O <sub>2</sub>	specific
Octopamine/dopamine	C <sub>8</sub> H <sub>11</sub> NO <sub>2</sub>	specific	Coumaryl alcohol	C <sub>9</sub> H <sub>8</sub> O	specific
Tryptamine	C <sub>10</sub> H <sub>12</sub> N <sub>2</sub>	specific	Caffeyl alcohol	C <sub>9</sub> H <sub>8</sub> O <sub>2</sub>	specific
3-Methoxytyramine	C <sub>9</sub> H <sub>13</sub> NO <sub>2</sub>	specific	Coniferyl alcohol	C <sub>10</sub> H <sub>10</sub> O <sub>2</sub>	specific
Noradrenaline	C <sub>8</sub> H <sub>11</sub> NO <sub>3</sub>	specific	5-OH-Feruloyl alcohol	C <sub>10</sub> H <sub>10</sub> O <sub>3</sub>	specific
Serotonin	C <sub>10</sub> H <sub>12</sub> N <sub>2</sub> O	specific	Sinapyl alcohol	C <sub>11</sub> H <sub>12</sub> O <sub>3</sub>	specific
3'-Methoxyoctopamine	C <sub>9</sub> H <sub>13</sub> NO <sub>3</sub>	specific	Non-condensed vanillyl alcohol	C <sub>8</sub> H <sub>10</sub> O <sub>3</sub>	specific
5-Methoxytryptamine	C <sub>11</sub> H <sub>14</sub> N <sub>2</sub> O	specific	Non-condensed coumaryl alcohol	C <sub>9</sub> H <sub>10</sub> O <sub>2</sub>	specific
Spermine	C <sub>10</sub> H <sub>26</sub> N <sub>4</sub>	specific	Non-condensed caffeyl alcohol	C <sub>9</sub> H <sub>10</sub> O <sub>3</sub>	specific
Coumaryl	C <sub>9</sub> H <sub>6</sub> O <sub>2</sub>	specific	Non-condensed coniferyl alcohol	C <sub>10</sub> H <sub>12</sub> O <sub>3</sub>	specific
Caffeoyl	C <sub>9</sub> H <sub>6</sub> O <sub>3</sub>	specific	Non-condensed 5-OH-feruloyl alcohol	C <sub>10</sub> H <sub>12</sub> O <sub>4</sub>	specific
Feruloyl	C <sub>10</sub> H <sub>8</sub> O <sub>3</sub>	specific	Non-condensed sinapyl alcohol	C <sub>11</sub> H <sub>14</sub> O <sub>4</sub>	specific
5-OH-Feruloyl	C <sub>10</sub> H <sub>8</sub> O <sub>4</sub>	specific	Dimethoxyquinol	C <sub>8</sub> H <sub>8</sub> O <sub>3</sub>	specific
Sinapyl	C <sub>11</sub> H <sub>10</sub> O <sub>4</sub>	specific	Syringyl alcohol	C <sub>9</sub> H <sub>10</sub> O <sub>3</sub>	specific
Malate	C <sub>4</sub> H <sub>4</sub> O <sub>4</sub>	specific	Isoprenylation	C <sub>5</sub> H <sub>8</sub>	specific
Glyceric acid coupling	C <sub>3</sub> H <sub>4</sub> O <sub>3</sub>	specific			

NIH Natural Products Library Round 1 和 NIH Natural Products Library Round 2) 中共收集到 1 542 个苯丙烷类化合物的正离子模式二级谱图 2 677 张和 661 个苯丙烷类化合物的负离子模式二级谱图 814 张, 建立探针分子串联质谱数据库。

图 2 给出了探针分子二级质谱采集条件统计, 其中来自飞行时间质谱的二级谱图占主要的比重, 正、负离子模式占比分别为 83. 75% 和 97. 67% (见图 2a)。对探针分子的化学类别进行统计, 分属 36 个亚类, 其中类黄酮、异黄酮、香豆素及其衍生物、肉桂酸及

a. The distributions of instrument types of MS/MS spectra



b. Metabolite classes

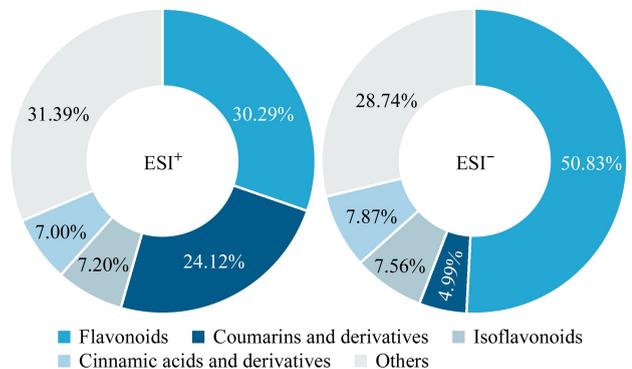


图 2 探针分子串联质谱数据库收录统计

Fig. 2 Statistics of the probe molecule MS/MS database

其衍生物占比较大,正、负离子模式占比分别为 68.61% 和 71.26% (见图 2b)。

### 2.4 途径代谢物筛选

将探针分子与玉米叶片中筛选出的修饰代谢组共建分子网络,选取其中苯丙烷途径代谢物所在的分子簇。分子簇筛选条件为:簇内同时含有探针分子与修饰代谢物,或虽不含探针分子但含有一个及以上途径特异性修饰的代谢物。满足筛选条件的分子簇中共含有 392 (正离子模式) 和 417 个 (负离子模式) 修饰代谢物。图 3a 所示为负离子模式筛选出的分子簇,簇内的探针分子可快速提示簇内代谢物所属的途径信息。如图 3b 所示,簇内 8 个探针分子 (黄色) 均为木脂素类化合物,提示该分子簇为木脂素及其结构类似物。仅含途径特异性修饰基团的代谢物分子簇如图 3c,簇内含有木脂素途径特异性醇结合的修饰基团,如松柏醇 (non-condensed coniferyl alcohol) 和芥子醇 (non-condensed sinapyl alcohol) 等,该分子簇也被快速识别为木脂素途径代谢物。

### 2.5 途径代谢物注释

对筛选出的途径代谢物进行结构注释,对于含有探针分子的分子簇,以探针分子为初始种子节点,结合修饰基团信息进行网络传播注释。以图 3b 红色虚框中紫色节点 ( $m/z$  697.2286) 的注释为例,其相邻黄色节点为探针分子 9''-O-Z-p-香豆酰-9'''-O-E-p-香豆酰-(7S,8S)-愈创木酰甘油 8-O-4'-松柏酰 (9''-O-Z-p-coumaroyl-9'''-O-E-p-coumaroyl-(7S,8S)-guaiacylglycerol 8-O-4'-coniferyl ether,  $m/z$  667.2180); 它们之间存在  $\Delta m/z$  为 30.0105 的质量差,推断其为探针分子甲氧基化修饰的产物。此外,该节点只有香豆酰 (coumaroyl) 一种修饰 (中性丢失),说明甲氧基化修饰未发生在香豆酰部分。进一步从该节点的二级谱图可知,香豆酰 (146.0357) 以中性形式丢失后,产生子离子  $m/z$

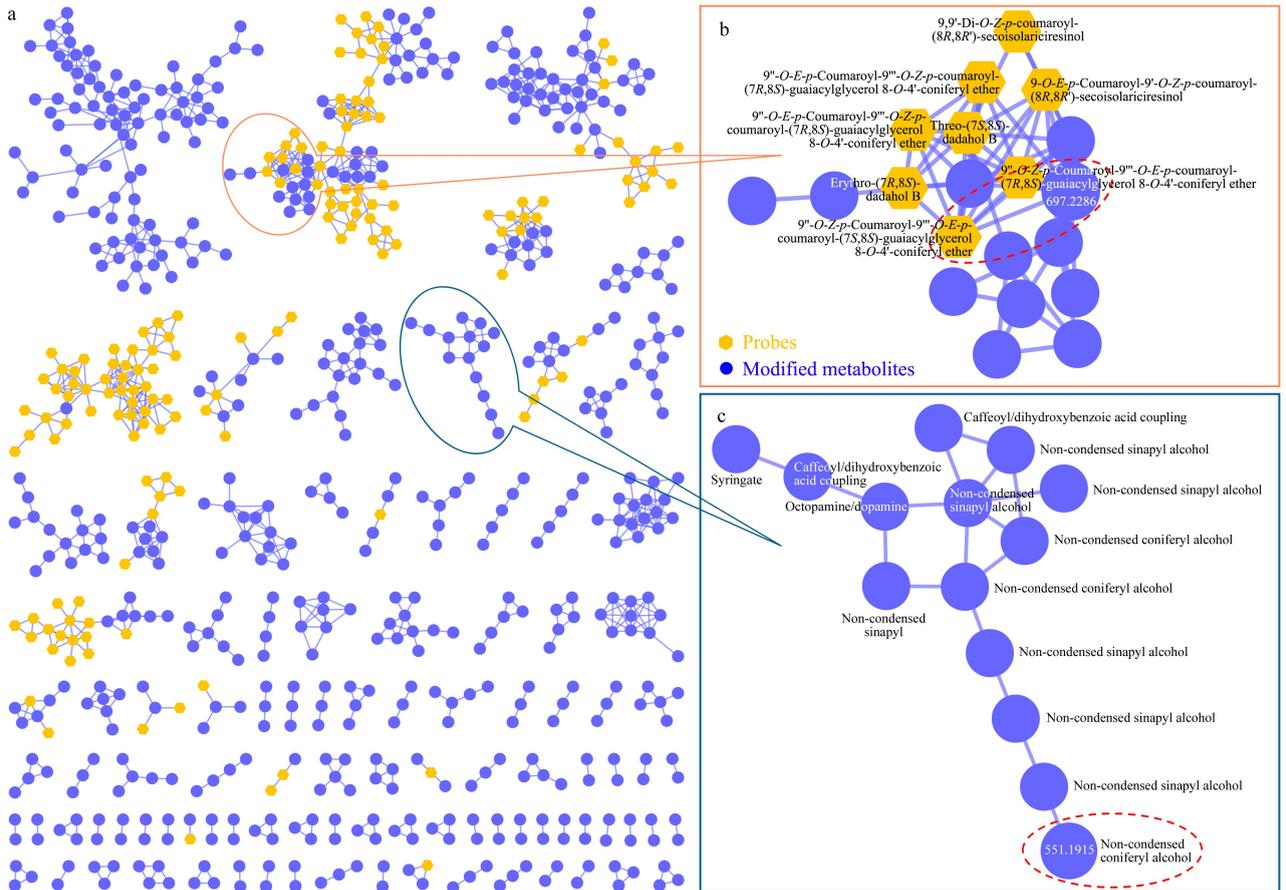


图 3 基于探针-修饰组分子网络筛选出的苯丙烷途径代谢物

Fig. 3 Phenylpropanoids pathway-associated metabolites screened by probes-modified metabolome molecular network  
a. phenylpropanoids pathway-associated clusters in negative ion mode; b. co-cluster consisting of both probe molecules and modified metabolites; c. modified metabolites with one or more pathway-specific modification types.

551.1929,它与子离子  $m/z$  341.1012 之间可能存在  $m/z$  210.0917 的中性丢失,从表1可知归属为非缩合芥子醇(non-condensed sinapyl alcohol),基于该子结构推断甲氧基化修饰仅能发生在阿魏醇(feruloyl alcohol)上。对其二级质谱碎片离子进行子结构注释(见图4a),并将碎片信息进行化合物结构拼接,将得到的结构搜索 PubChem 数据库,确定为稀有木脂素 A(dadahol A)。

对只含有途径特异性修饰代谢物的分子簇,如簇内有可被搜库鉴定的节点,则以注释节点为初始种子,注释过程与含有探针分子的分子簇类似。对于无相邻注释节点的修饰代谢物采用从头注释方式,如图3c所示簇内代谢物仅有修饰基团提供的子结构信息。以簇末端代谢物( $m/z$  551.1915)的结构解析为例说明如下:该代谢物含有非缩合松柏醇(non-condensed coniferyl alcohol)子结构,经中性丢失( $m/z$  180.0786)产生  $m/z$  371.1112 的子离子;另外还有香豆酸(coumaric acid)特征离子( $m/z$  163.0377),它与子离子  $m/z$  371.1112 之间可能存在  $m/z$  208.0735 的中性丢失;该中性丢失与修饰基团库中的非缩合芥子醇( $m/z$  210.0892)相差 2.0157,推断可能是非缩合芥子醇开环失去 2 个 H;将得到的子结构非缩合松柏醇、香豆酸和非缩合芥子醇进行拼接,符合二级谱图注释结果的仅存在一种合理结构(见图4b)。该结构经 PubChem

和 SciFinder 数据库搜索,均未见收录。采用上述注释方法,将正、负离子均注释出的同一代谢物,根据其结构去冗余后共初步注释出 129 个苯丙烷途径代谢物,其中 89 个在 PubChem 和 SciFinder 数据库中已有收录,26 个为数据库未报道的“未知结构”化合物。采用注释代谢物的 MS/MS 对其结构进一步验证,其中 68 个注释代谢物的化合物类别可被准确预测为苯丙烷代谢物。对其中 115 个有确切结构的代谢物(其余 14 个为同分异构体)预测其保留时间,其中 102 个代谢物的预测保留时间相对误差小于 30%。采用 CFM-ID 工具预测了 115 个代谢物的二级谱图,并计算了与实验二级谱图的相似性;其中,具有较高相似性( $>0.5$ )的代谢物有 28 个,较相似( $0.4\sim 0.5$ )的代谢物有 30 个,有一定相似性( $<0.4$ )的有 57 个。此外,对其中 10 个有标准品的代谢物进行了验证,验证结果显示注释结构正确。

从注释结果可知,129 个苯丙烷途径代谢物涉及苯丙烷下游主要分支途径产生的次生代谢物(见图5),如黄酮分支途径的 8 个类黄酮、19 个氧苷类黄酮、32 个碳苷类黄酮,羟基肉桂酸途径的 31 个羟基肉桂酸酰胺及其衍生物,以及木脂素合成途径的 22 个(新)木脂素/木脂素苷等。其中 4 个氧苷类黄酮、4 个碳苷类黄酮、6 个羟基肉桂酸酰胺及其衍生物和 11 个木脂素结构在 PubChem 和 SciFinder 数据库中均未被收录(见图5括号内红色数字)。

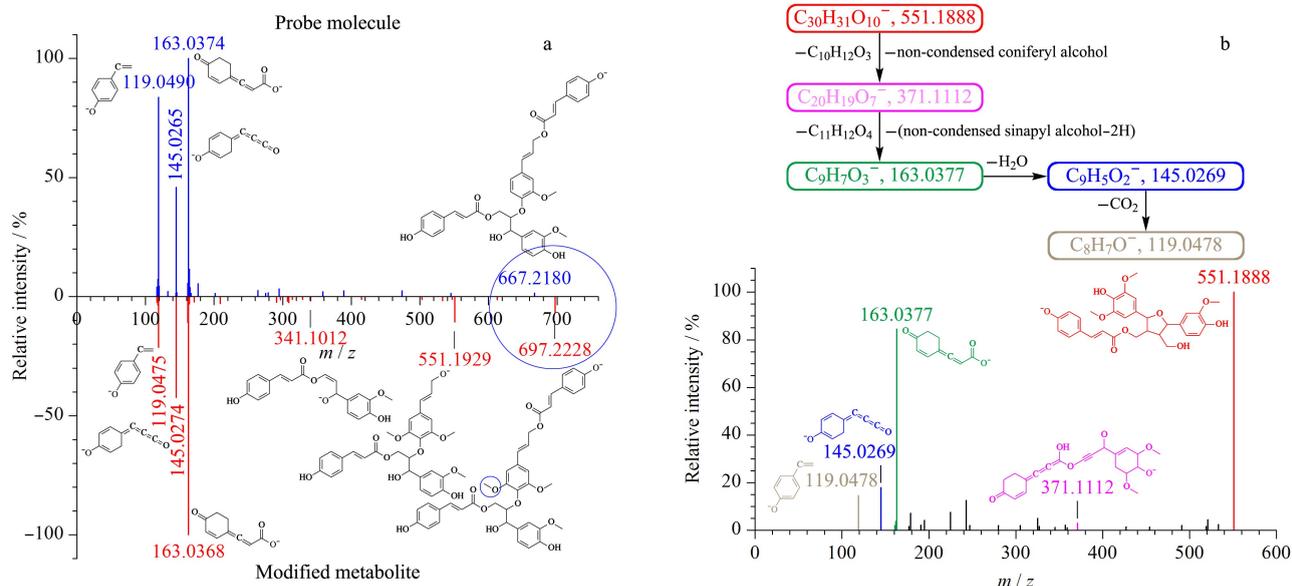


图4 (a)基于已知相邻节点的代谢物传播注释和(b)代谢物从头注释

Fig. 4 (a) Annotation of metabolites with known neighbor-nodes and (b) *de novo* annotation of metabolites

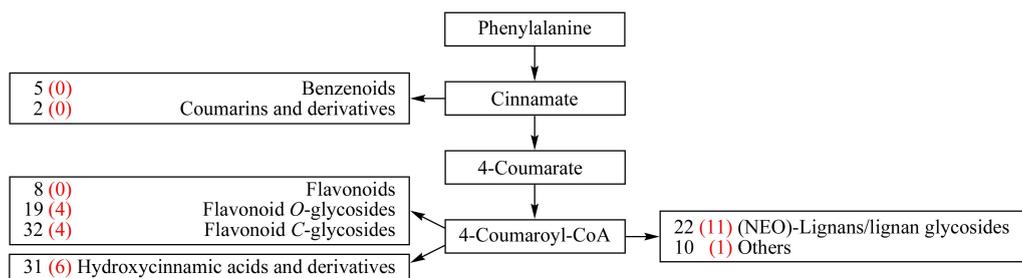


图 5 注释代谢物的分支途径分布

Fig. 5 Distributions of annotated metabolites in the downstream branches of the phenylpropanoid pathway

Red numbers in brackets; previously unreported in both PubChem and SciFinder databases.

### 3 结论

本研究以苯丙烷途径代谢物的筛选和注释为例,发展了一种基于修饰组和探针分子的重要途径代谢物高效筛选和注释方法。仅从一种玉米叶片组织中就注释出了 129 个苯丙烷途径代谢物,其中有 26 个未被 PubChem 和 SciFinder 数据库收录,显示了方法在发现与注释途径代谢物特别是未在数据库收录的“未知结构”代谢物方面的能力。鉴于目前途径数据库提供的修饰反应尚不全面,以及开源质谱数据库因谱图质量、仪器采集条件差异等原因,方法注释能力还未能充分发挥。随着数据库信息的不断积累和完善,以及基于机器学习的子结构预测方法不断成熟,未来将在重要代谢途径挖掘和利用方面发挥更大的作用。

### 参考文献:

- [1] Isah T. *Biol Res*, 2019, 52: 39
- [2] Dixon R A, Paiva N L. *Plant Cell*, 1995, 7(7): 1085
- [3] Deng Y X, Lu S F. *Crit Rev Plant Sci*, 2017, 36(4): 257
- [4] Deng M, Zhang X H, Luo J Y, et al. *Plant J*, 2020, 103(5): 1710
- [5] Tohge T, Scossa F, Wendenburg R, et al. *Mol Plant*, 2020, 13(7): 1027
- [6] Chen J, Hu X, Shi T T, et al. *Plant Biotechnol J*, 2020, 18(8): 1722
- [7] Perez D S L, Alseekh S, Naake T, et al. *Curr Protoc Plant Biol*, 2019, 4(4): e20100
- [8] Jin M, Zhang X H, Zhao M C, et al. *BMC Plant Biol*, 2017, 17(1): 17
- [9] Mu H M, Ci Z J, Ai S J, et al. *Chinese Journal of Chromatography*, 2021, 39(11): 1203

牟红梅, 慈志娟, 艾沙江, 等. 色谱, 2021, 39(11): 1203

- [10] Hu Y D, Huang S C, Ren D B, et al. *Chinese Journal of Chromatography*, 2022, 40(3): 242
- 胡永丹, 黄斯晨, 任达兵, 等. 色谱, 2022, 40(3): 242
- [11] Boonchaisri S, Rochfort S, Stevenson T, et al. *Metabolomics*, 2021, 17(3): 28
- [12] Dührkop K, Nothias L F, Fleischauer M, et al. *Nat Biotechnol*, 2021, 39(4): 462
- [13] Nothias L F, Petras D, Schmid R, et al. *Nat Methods*, 2020, 17(9): 905
- [14] Sedio B E, Boya P C A, Rojas E J C. *Appl Plant Sci*, 2018, 6(3): e1033
- [15] Xu S, Wang J J, Wei Y, et al. *J Agric Food Chem*, 2019, 67(43): 12084
- [16] Clements T, Rautenbach M, Ndlovu T, et al. *Front Chem*, 2021, 9: 633870
- [17] Dai W D, Yin P Y, Zeng Z D, et al. *Anal Chem*, 2014, 86(18): 9146
- [18] Hawkins C, Ginzburg D, Zhao K M, et al. *J Integr Plant Biol*, 2021, 63(11): 1888
- [19] Yang B, Li H, Ruan Q F, et al. *J Pharm Biomed Anal*, 2018, 157: 44
- [20] Yang Q, Ji H C, Lu H M, et al. *Anal Chem*, 2021, 93(4): 2200
- [21] Li Y Y, Kind T, Folz J, et al. *Nat Methods*, 2021, 18(12): 1524
- [22] Dührkop K, Fleischauer M, Ludwig M, et al. *Nat Methods*, 2019, 16(4): 299
- [23] Morreel K, Saeys Y, Dima O, et al. *Plant Cell*, 2014, 26(3): 929
- [24] Wang S C, Alseekh S, Fernie A R, et al. *Mol Plant*, 2019, 12(7): 899
- [25] Li Z F, Zhao C X, Zhao X J, et al. *Anal Chem*, 2018, 90(24): 14321
- [26] Djoumbou F Y, Eisner R, Knox C, et al. *J Cheminformatics*, 2016, 8(1): 61