

咖啡 NMT 基因家族鉴定及不同时期 WGCNA 表达网络分析

张妍冰¹, 王曦奥¹, 黄丽芳¹, 吉训志¹, 方晓峰¹, 闫林^{1,2}, 胡丽松^{1,2},
廖振阳^{1,2*}

1. 中国热带农业科学院香料饮料研究所/农业农村部香辛饮料作物遗传资源利用重点实验室/咖啡海南省工程研究中心/热带香料饮料国际联合研究中心, 海南万宁 571533; 2. 中国热带农业科学院三亚研究院/热带作物生物育种全国重点实验室, 海南三亚 572024

摘要: 咖啡作为全球范围内重要的经济作物之一, 广泛种植于热带和亚热带地区, 是众多国家和地区的经济收入来源。咖啡因(1,3,7-三甲基黄嘌呤)是咖啡豆的主要次生代谢产物, 不仅赋予咖啡独特的苦涩风味和提神功效, 还发挥多重生物学功能。其中, N-甲基转移酶(NMT)基因家族在咖啡因生物合成途径中发挥关键作用。因此, 系统鉴定咖啡 NMT 基因家族成员并解析其在不同发育时期的表达模式, 对于探究咖啡因代谢、优化咖啡品质及抗逆育种具有重要意义。本研究在异源四倍体小粒种咖啡中共鉴定到 27 个 NMT 家族基因, 其中 C 亚基因组 14 个基因, E 亚基因组 13 个基因, 分布在 10 条染色体上。基因结构显示, 成员之间基序组成较为相似, 但内含子排列和数量有差异。顺式作用元件分析发现, 小粒种咖啡 *CaNMT* 基因启动子区富含多种顺式作用元件, 涉及光应答、胁迫响应和激素信号转导, 并包含 MYB、MYC 和 W-box 等转录因子结合位点。进一步通过转录组分析发现小粒种咖啡果实发育过程中基因呈现动态表达, 果实变色期差异表达基因数目最多(14 703 个), 而果实膨大期数量最少(4419 个)。小粒种咖啡 NMT 家族基因在果实不同发育时期及不同组织中也表现出明显的动态与特异性。通过加权基因共表达网络分析(WGCNA)构建 12 个共表达模块, ivory、coral1 和 darkolivegreen 模块分别富集环境应激与蛋白折叠、能量代谢、核酸代谢相关通路。核心基因网络阐明了 *CaNMT* 基因在咖啡因代谢、应激响应及发育调控中的关键作用。本研究系统鉴定了异源四倍体小粒种咖啡 NMT 家族基因, 并研究了该家族在不同果实发育时期的表达模式, 阐明了不同基因的功能, 为咖啡育种提供理论依据。

关键词: 咖啡; NMT 基因家族; WGCNA; 果实发育

中图分类号: S571.2 **文献标志码:** A

Genome-wide Identification of the NMT Gene Family and Stage-Specific WGCNA in *Coffea*

ZHANG Yanbing¹, WANG Xi'ao¹, HUANG Lifang¹, JI Xunzhi¹, FANG Xiaofeng¹, YAN Lin^{1,2},
HU Lisong^{1,2}, LIAO Zhenyang^{1,2*}

1. Spice and Beverage Research Institute, Chinese Academy of Tropical Agricultural Sciences / Key Laboratory of Genetic Resources Utilization of Spice and Beverage Crops, Ministry of Agriculture and Rural Affairs / Hainan Research Center for Coffee / International Joint Research Center for Tropical Spice Beverages, Wanning, Hainan 571533, China; 2. Sanya Research Institute, Chinese Academy of Tropical Agricultural Sciences / National Key Laboratory of Tropical Crop Biological Breeding, Sanya, Hainan 572024, China

Abstract: Coffee, one of the most important economic crops worldwide, is widely cultivated in tropical and subtropical regions and serves as a major source of income for many countries and regions. Caffeine (1,3,7-trimethylxanthine) is the main metabolite in coffee beans and related products, which not only imparts the characteristic bitter taste and stimu-

收稿日期 2025-09-16; **接受日期** 2025-10-06

基金项目 中央级公益性科研院所基本科研业务费专项 (No. 1630142024017); 国家热带植物种质资源库 (No. NTPGRC2025-014); 国家木薯产业技术体系专项 (No. CARS-11-HNYL)。

作者简介 张妍冰 (1999—), 女, 博士研究生, 研究方向: 热带植物基因组学。*通信作者 (Corresponding author): 廖振阳 (LIAO Zhenyang), E-mail: zhenyangliao@catas.cn。

lates effect of coffee but also exerts multiple biological functions. The N-methyltransferase (NMT) gene family plays a key role in the caffeine biosynthetic pathway. Therefore, systematic identification of coffee NMT family members and analysis of the expression patterns at different developmental stages are of great importance for investigating caffeine metabolism, improving coffee quality, and breeding stress-resistant varieties. In this study, a total of 27 NMT genes were identified in allotetraploid *Coffea arabica*, including 14 genes from the C subgenome and 13 from the E subgenome, distributed across 10 chromosomes. Gene structure analysis revealed that family members shared similar motif compositions but differed in intron arrangement and number. *Cis*-acting element analysis showed that the promoters of *C. arabica* NMT genes were enriched with various regulatory elements related to light response, stress response and hormone signaling, and contained transcription factor binding sites such as MYB, MYC and W-box. Transcriptome analysis further demonstrated that NMT genes exhibited dynamic expression during fruit development, with the ripening stage showing the highest number of differentially expressed genes (14 703), while the expansion stage showed the lowest (4419). Moreover, *C. arabica* NMT genes displayed clear temporal and tissue-specific expression patterns. Weighted gene co-expression network analysis (WGCNA) identified 12 co-expression modules, among which the ivory, coral1 and darkolivegreen modules were enriched in pathways related to environmental stress and protein folding, energy metabolism and nucleic acid metabolism, respectively. Core gene network analysis revealed the critical roles of NMT genes in caffeine metabolism, stress response, and developmental regulation. Taken together, this study systematically identified the NMT gene family in allotetraploid *C. arabica*, characterized the expression patterns during different stages of fruit development, and elucidated the potential functional differentiation, thereby providing a theoretical basis for coffee breeding.

Keywords: *Coffea*; NMT gene family; WGCNA; fruit development

DOI: 10.3969/j.issn.1000-2561.2025.12.003

咖啡 (*Coffea*) 是一种在全球范围内广泛种植的热带经济作物, 其种子烘焙制成的饮品深受消费者喜爱^[1]。咖啡产业链涵盖了种植、加工、贸易、烘焙和零售等多个环节^[2], 在全球贸易中占据重要地位, 是许多热带国家的重要出口创汇产品之一^[3]。咖啡风味的形成涉及多种化学成分, 包括挥发性化合物 (如醛类、酮类和酯类) 和非挥发性化合物 (如绿原酸、咖啡因和多糖), 这些成分在咖啡豆的种植、加工、烘焙和冲泡过程中发生复杂的化学变化, 最终决定了咖啡的香气和口感^[4-5]。咖啡因 (1,3,7-三甲基黄嘌呤) 是咖啡豆及相关制品中的主要代谢产物, 不仅赋予咖啡独特的苦涩风味和提神功效, 还在抗氧化、心血管健康和代谢调节等方面具有重要作用^[6-7]。近年来, 中粒种咖啡 (*Coffea canephora*) 参考基因组^[8]与多套高质量的小粒种咖啡 (*C. arabica*) 基因组^[9-10]相继发表, 为咖啡风味解析和品质改良提供了研究基础。小粒种咖啡为异源四倍体 ($2n=4x=44$), 源自 *C. eugenioides* 与 *C. canephora* 的天然杂交^[9], 已有研究表明 2 个亚基因组在咖啡因、萜类及脂肪酸合成等关键代谢基因簇中呈现局部表达偏倚^[11-12], 这为等位及同源基因的互作网络解析提供了思路。

NMT 基因家族属于甲基转移酶超家族, 负责催化 S-腺苷甲硫氨酸 (SAM) 依赖的甲基化反应,

在咖啡因生物合成中发挥核心作用。NMT 在咖啡属植物中的序列保守性较高^[13], 该家族主要包括黄嘌呤甲基转移酶 (XMT)、7-甲基黄嘌呤甲基转移酶 (MXMT) 和 3,7-二甲基黄嘌呤甲基转移酶 (DXMT), 分别催化咖啡因合成途径中的 3 个关键步骤: 从黄苷到 7-甲基黄苷、7-甲基黄嘌呤到可可碱, 以及可可碱到咖啡因。已有研究克隆出小粒种咖啡中 *CaXMT1*、*CaMXMT* 和 *CaDXMT1/CCS1* 基因并进行功能鉴定, 明确了其底物专一性和催化顺序^[14-15]。通过结构生物学分析, 进一步解析中粒种咖啡中 XMT 和 DXMT 的 SAM 依赖甲基转移机制及其底物选择的关键决定因子^[16]。在基因组分布方面, 已有研究发现小粒种咖啡中控制咖啡因生物合成的 NMT 基因主要位于 1 号和 9 号染色体, 呈基因聚簇分布特性^[8]。DXMT 基因在咖啡因合成最后一步中起关键作用, 其在 Chr1 上的结构变异与咖啡因含量种间差异相关。此外, Chr9c 和 Chr9e 上 XMT 和 MXMT 基因形成功能性基因簇, 序列保守性高。这些特征为研究 NMT 基因家族的功能分化及咖啡因代谢调控提供了重要基础。在表达模式方面, NMT 基因在幼叶、花芽和未成熟果实中表达丰度较高, 与体内甲基黄嘌呤和咖啡因的积累趋势一致。*C. canephora* 中 DXMT 基因在绿色果实中的表达水平最高, 而在 *C. eugenioides* 中, DXMT 基因仅在绿色果实阶段

显著表达,且受同源特异性重复 DNA 调控^[17]。利用 RNAi 技术抑制 *CaMXMT* 基因可同步降低可可碱和咖啡因含量,验证了可可碱作为咖啡因合成通路关键中间体的功能节点^[18]。*NMT* 基因家族不仅在咖啡因生物合成中至关重要,还在植物代谢调控和适应性进化中发挥广泛作用。例如,拟南芥中 *NMT* 基因功能缺失会导致植物生长和繁殖能力受损^[19]。

加权基因共表达网络分析 (weighted gene co-expression network analysis, WGCNA) 是一种系统生物学方法,通过构建基因表达的共表达网络,挖掘基因间的协同表达关系及其与表型性状的关联^[20]。近年来, WGCNA 广泛应用于植物基因功能研究,如水稻抗逆基因网络^[21]、玉米籽粒发育调控^[22]等。然而,在咖啡研究中, WGCNA 的应用尚处于起步阶段,尤其针对 *NMT* 基因家族在咖啡不同发育时期的共表达网络分析仍显不足。利用 WGCNA 方法,可系统解析 *NMT* 基因与其他基因间的共表达关系,挖掘关键调控模块及候选基因,为揭示咖啡因代谢的分子机制提供新视角。

近年来,咖啡基因组学研究在高通量测序技术和生物信息学工具的推动下取得了重要突破,为解析咖啡因代谢的分子机制提供了坚实基础。本研究旨在通过全基因组分析鉴定咖啡 *NMT* 基因家族成员,结合 RNA-seq 数据,系统解析其不同发育时期的表达模式,并利用 WGCNA 方法构建 *NMT* 基因的共表达网络,挖掘关键调控模块及候选基因。研究结果将为咖啡因代谢的深入研究提供数据支持,并为咖啡品质改良及抗逆育种提供新的理论基础。

1 材料与方法

1.1 材料

选取生长于海南省兴隆热带植物园咖啡种质资源圃中长势良好的小粒种咖啡多年生植株为试验材料,分别采集其根、茎、叶以及 5 个不同发育时期的果实样品 (Fruit-1 为坐果期; Fruit-2 为幼果期; Fruit-3 为果实膨大期; Fruit-4 为果实变色期; Fruit-5 为果实成熟期),取样结束后,立即将咖啡种子与果肉分离,所有组织迅速置于液氮中速冻,并保存于 -80°C ,后续用于测序分析。

1.2 方法

1.2.1 小粒种咖啡 *NMT* 基因家族的鉴定 从 NCBI 数据库 (<https://www.ncbi.nlm.nih.gov/>) 中

下载中粒种咖啡 *NMT* 基因的 CDS 序列、蛋白序列,基于小粒种咖啡基因组数据,利用 BLASTN^[23] 比对工具,获取与已知 *NMT* 基因相似的序列,获得小粒种的 *NMT* 候选基因, *E*-value 设置为 $1\text{e-}5$ 。使用 SMART^[24] 工具进行保守结构域验证,删除不具备完整结构和氨基酸长度小于 100 aa 的序列,经去重复后最终获得 *CaNMT* 基因。

1.2.2 *CaNMT* 基因理化性质和亚细胞定位分析 采用 ProtParam (<http://web.expasy.org/protparam>) 推测 *CaNMT* 蛋白的长度、等电点和理论分子量。使用 Softberry Home Page (<http://linux1.softberry.com/berry.phtml>) 网站对 *CaNMT* 基因进行亚细胞定位预测。

1.2.3 咖啡 *NMT* 基因家族系统进化树和染色体定位分析 利用 MEGA^[25] 软件中的 MUSCLE 模块,对小粒种咖啡 *NMT* 基因家族成员的全长氨基酸序列进行多序列比对。随后,使用 MEGA 软件 Find Best DNA/Protein Models 功能评估并确定最优的进化模型,其中缺失数据处理方式选择 partial deletion。通过邻接法 (neighbor-joining, NJ) 构建系统进化树,并对树的分支可靠性进行自展检验 (bootstrap),重复次数设定为 1000 次。此外,从小粒种咖啡基因组注释文件中获取 *NMT* 基因家族成员的染色体定位信息,利用 TBtools^[26] 软件绘制染色体定位分布图。

1.2.4 基因家族的基因结构和保守结构域分析 采用 GSDS (<http://gsds.cbi.pku.edu.cn/>) 分析 *CaNMT* 基因结构;采用 MEME (<http://meme-suite.org/>) 分析其蛋白序列的保守结构域,最大结构域数目设置为 10。利用 bedtools^[27] 工具提取 *CaNMT* 基因上游 1500 bp 序列,提交至 PlantCare (<https://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) 进行顺式作用元件预测,对预测结果进行汇总、筛选和简化后通过 TBtools 软件对预测的顺式作用元件进行可视化分析。

1.2.5 转录组分析及 WGCNA 分析 转录组原始数据经 fastp 软件过滤后,使用 HISAT^[28] 工具进行比对,并利用 featureCount^[29] 软件获取 reads 数,经过矫正后获取基因表达 TPM 值。基于 TPM 表达矩阵使用 R 软件 WGCNA 包进行共表达网络分析,其中以中位数绝对偏差 (MAD) 过滤低变异基因,用 goodSamplesGenes 剔除异常基因后进行样本层次聚类质控,用 pickSoftThreshold 确定软阈值 β ,构建邻接矩阵并转为拓扑重叠矩阵 (TOM),

然后在 TOM 距离上层次聚类，采用动态剪切识别模块（minModuleSize=50，deepSplit=2），并在特征基因（ME）间距离<0.33 合并相近模块。对每个模块计算 ME，并与表型矩阵做 Pearson 相关分析，同时计算每个基因的模块归属度（MM/kME）和对性状的基因显著性（GS）。结果通过样本树+性状热图、模块-性状相关热图及 TOM 子网络热图等进行可视化，并将各模块的 TOM 子网（阈值为 0.02）导出为 Cytoscape 边/节点文件。提取 *CaNMT* 基因所在模块的基因集后，使用 TBtools 软件进行 KEGG 富集分析。

2 结果与分析

2.1 *CaNMT* 基因鉴定及理化性质分析

基于小粒种基因组，共筛选出 27 个 *CaNMT*

基因。理化性质分析结果显示，小粒种咖啡 NMT 基因家族成员的蛋白质特性具有多样性（表 1）。*CaNMT* 蛋白的氨基酸长度范围在 107~416 aa 之间，其中，*CaNMT3c* 编码的蛋白长度最短，仅为 107 aa，而 *CaNMT16e* 编码的蛋白长度最长，达到 416 aa。这种长度的差异可能与基因的功能分化及其在咖啡因代谢途径中的具体作用相关。预测的分子质量范围为 11957.56 Da（*CaNMT3c*）~46992.74 Da（*CaNMT16e*），反映了 *CaNMT* 蛋白在结构复杂性上的差异。理论等电点（pI）分布在 5.02（*CaNMT24e*）~9.03（*CaNMT18e*）之间，表明 NMT 蛋白的电荷性质差异较大。其中，*CaNMT1c*（pI 8.11）、*CaNMT7c*（pI 8.22）、*CaNMT22e*（pI 8.24）、*CaNMT5c*（pI 8.60）以及 *CaNMT18e*（pI 9.03）为碱性蛋白（pI>7），而其

表 1 *CaNMT* 蛋白理化性质
Tab. 1 Physical and chemical properties of *CaNMT* proteins

基因名称 Gene name	基因 ID Gene ID	氨基酸数目 Number of amino acid/aa	分子量 Molecular weight/Da	等电点 Theoretical pI	不稳定指数 Instability index	脂肪族指数 Aliphatic index	平均亲水性 Grand average of hydropathicity	定位 Location
<i>CaNMT1c</i>	Chr1c006347	252	28 575.75	8.11	53.86	75.44	-0.297	细胞质
<i>CaNMT2c</i>	Chr1c006370	384	43 276.26	5.38	54.07	87.60	-0.246	细胞质
<i>CaNMT3c</i>	Chr1c006565	107	11 957.56	5.51	37.61	88.60	-0.277	细胞质
<i>CaNMT4c</i>	Chr1c006566	192	21 514.33	5.62	54.53	89.38	-0.152	叶绿体
<i>CaNMT5c</i>	Chr1c006585	130	14 554.80	8.60	30.79	87.85	-0.188	细胞质
<i>CaNMT6c</i>	Chr2c011551	297	33 539.09	5.44	46.44	89.33	-0.115	细胞质
<i>CaNMT7c</i>	Chr5c026745	207	23 498.15	8.22	51.26	95.17	0.004	细胞质
<i>CaNMT8c</i>	Chr9c039876	385	43 270.65	5.69	44.13	90.42	-0.168	细胞质
<i>CaNMT9c</i>	Chr9c039878	392	44 341.90	5.27	39.32	96.96	-0.101	细胞质
<i>CaNMT10c</i>	Chr9c039881	378	42 747.12	5.40	42.85	93.65	-0.170	细胞质
<i>CaNMT11c</i>	Chr9c039884	343	38 601.50	5.53	42.34	98.05	-0.133	细胞质
<i>CaNMT12c</i>	Chr9c039886	394	44 735.03	5.41	48.20	86.40	-0.201	细胞质
<i>CaNMT13c</i>	Chr11c004549	360	39 684.31	5.76	43.37	86.42	-0.114	细胞质
<i>CaNMT14c</i>	Chr11c004550	145	15 946.17	5.89	47.23	96.90	0.027	细胞质
<i>CaNMT15e</i>	Chr1e048197	384	43 163.23	5.48	52.17	89.64	-0.249	细胞质
<i>CaNMT16e</i>	Chr1e048199	416	46 992.74	5.34	53.06	91.39	-0.161	细胞质
<i>CaNMT17e</i>	Chr1e048201	389	44 039.35	5.40	50.37	90.98	-0.211	细胞质
<i>CaNMT18e</i>	Chr1e048416	152	17 205.88	9.03	32.39	91.78	-0.189	细胞质
<i>CaNMT19e</i>	Chr1e048556	227	25 413.36	5.95	41.39	95.33	-0.002	叶绿体
<i>CaNMT20e</i>	Chr2e053130	207	23 134.31	5.23	42.97	94.64	-0.196	细胞质
<i>CaNMT21e</i>	Chr4e065015	367	41 987.88	6.25	49.22	78.58	-0.344	细胞质
<i>CaNMT22e</i>	Chr5e067541	197	22 161.58	8.24	53.75	95.08	0.030	细胞质
<i>CaNMT23e</i>	Chr9e081520	385	43 270.56	5.41	43.42	93.71	-0.161	细胞质
<i>CaNMT24e</i>	Chr9e081521	318	36 062.20	5.02	43.27	94.37	-0.190	细胞质
<i>CaNMT25e</i>	Chr9e081529	384	43 405.72	5.58	40.77	92.45	-0.180	细胞质
<i>CaNMT26e</i>	Chr9e081533	385	43 530.06	5.70	39.64	97.48	-0.119	细胞质
<i>CaNMT27e</i>	Chr9e081536	241	27 255.03	5.31	39.39	92.28	-0.305	细胞质

余蛋白均为酸性蛋白 ($pI < 7$)。这种酸碱性差异可能影响蛋白质在细胞环境中的稳定性和相互作用特性。

不稳定指数的分析显示, NMT 蛋白的不稳定指数范围为 30.79 (CaNMT5c)~54.53 (CaNMT4c), 其中大部分蛋白的不稳定指数高于 40, 表明多数 NMT 蛋白在体外环境中可能具有较低稳定性, 需依赖特定的细胞环境或伴侣蛋白以维持其功能活性。脂肪族指数介于 75.44 (CaNMT1c)~98.05 (CaNMT11c) 之间, 较高的脂肪族指数提示 NMT 蛋白具有较多的疏水性氨基酸, 这可能与其在细胞内的折叠和膜相关功能有关。平均亲水性值在 -0.344 (CaNMT21e)~0.030 (CaNMT22e) 之间, 均偏向负值, 表明 NMT 蛋白总体呈现亲水性, 有利于其在水性细胞环境中发挥功能, 仅少数蛋白如 CaNMT22e (0.03) 和 CaNMT14c (0.027) 接近中性。

亚细胞定位预测结果显示, NMT 基因家族成员的蛋白产物主要分布于细胞质, 少数分布于叶绿体, 反映了其在代谢过程或蛋白质修饰中的潜在功能。具体而言, 在分析的 27 个成员中, 2 个蛋白 (CaNMT4c、CaNMT19e) 定位于叶绿体, 其余定位于细胞质。这种分布模式表明 NMT 基因家族的蛋白产物主要在细胞质中执行功能, 可

能与 N-甲基转移酶相关的代谢途径或蛋白质修饰过程密切相关。叶绿体定位的蛋白 (CaNMT4c 和 CaNMT19e) 可能参与光合作用相关的次生代谢产物的合成或调控, 提示 NMT 基因家族在不同亚细胞环境中具有多样化的生物学功能。

2.2 NMT 基因染色体定位及系统进化分析

27 个 *CaNMT* 基因分布在咖啡的 10 条染色体上 (图 1), 整体呈现不均衡的分布格局。其中, *NMT* 基因在 C 与 E 亚基因组的 Chr1 与 Chr9 上均显示出明显的富集趋势, Chr1c、Chr1e、Chr9c 和 Chr9e 各含 5 个基因; 而在 Chr2 (Chr2c 和 Chr2e) 和 Chr5 (Chr5c 和 Chr5e) 上则仅各有 1 个基因。这种分布模式可能与基因复制事件或染色体重排相关。此外, Chr11c 上分布有 2 个基因, 而 Chr4e 仅检测到 1 个基因。此外, 虽然 Chr1、Chr2、Chr5 和 Chr9 在 C 与 E 亚基因组间均表现出一致的基因分布模式, 但 C 亚基因组的 Chr11c 包含 2 个 *NMT* 基因 (CaNMT13c 和 CaNMT14c), 而 E 亚基因组的 Chr4e 仅含 1 个基因 (CaNMT21e), 显示出亚基因组间在基因分布模式上的不对称性。

为了明确 NMT 基因家族的系统进化关系, 对茶 (*Camellia*)、可可 (*Theobroma*) 和咖啡 (*Coffea*) 61 个 NMT 蛋白全长序列进行了系统进化树构建, 其中 (图 2)。分析结果显示, 咖啡 *NMT* 基因在进

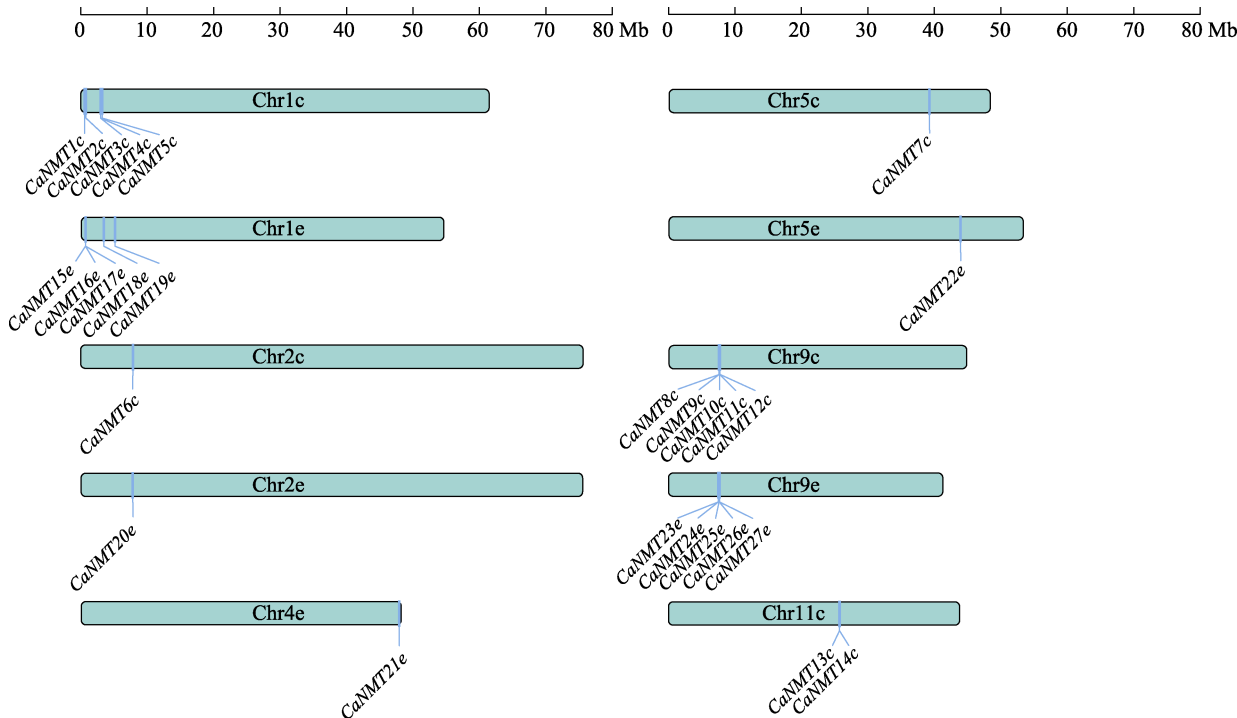
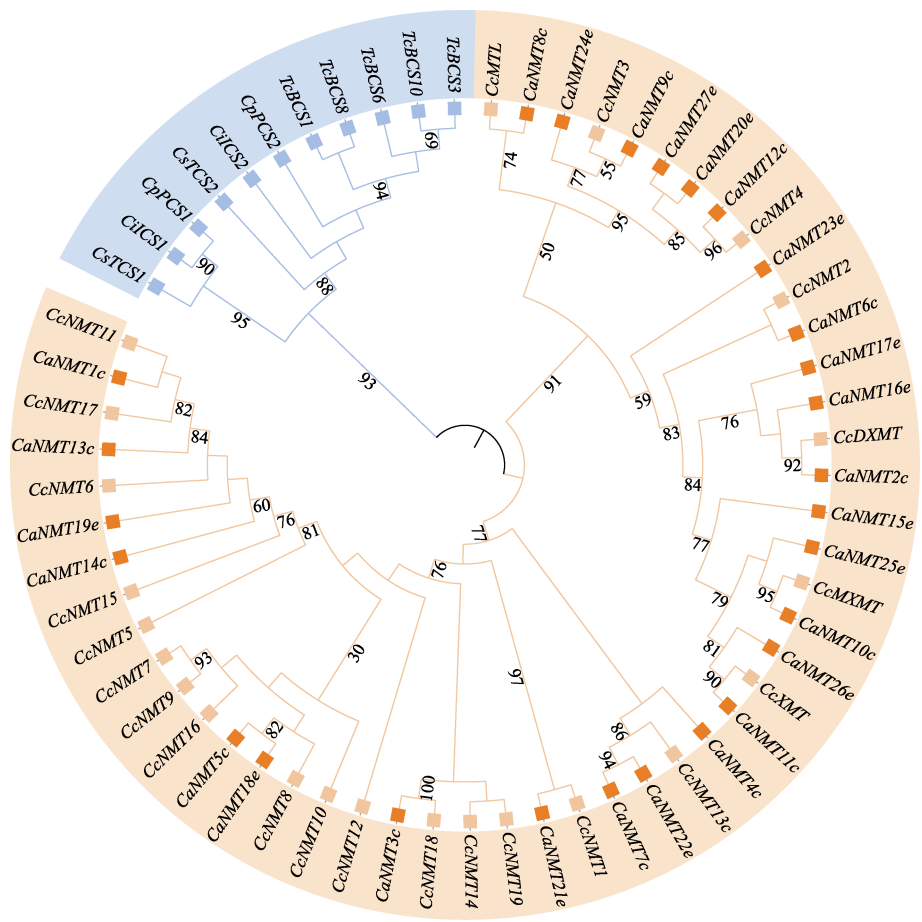


图 1 *CaNMT* 基因在染色体上的分布
Fig. 1 Locations of *CaNMT* genes on chromosomes



Cs、Cp 代表茶；Ci、Tc 代表可可；Cc 代表中粒种咖啡；Ca 代表小粒种咖啡。
Cs, Cp represent *Camellia*; Ci, Tc represent *Theobroma*; Cc represents *C. canephora*; Ca represents *C. arabica*.

图 2 4 个物种的 NMT 基因家族系统进化分析

Fig. 2 Phylogenetic analysis of NMT gene family in four species

化树中形成 1 个独立的聚类分支，表明其在进化过程中具有较高的特异性。茶和可可的 *NMT* 基因在进化树中聚类较近，显示出较高的序列相似性。此外，进化树中还观察到相比于中粒种咖啡，小粒种咖啡中 *NMT* 基因出现收缩现象，多个 *CcNMT* 基因比对为同一个 *CaNMT* 基因，如图 2 所示，*CcNMT14*、*CcNMT18*、*CcNMT19* 基因均与 *CaNMT3c* 同源，反映了基因家族在进化过程中的收敛或功能冗余。

2.3 *CaNMT* 基因结构与保守结构域分析

为进一步了解 *CaNMT* 基因系统进化与其编码蛋白中保守基序及其基因结构之间的关系，对 *NMT* 编码蛋白保守基序以及外显子与内含子的结构进行了分析。基因结构分析表明，*CaNMT* 基因的外显子（Exon）-内含子（Intron）组成呈现多样性，外显子数量在 1~6 个之间变化，例如 *CaNMT14c* 和 *CaNMT18e* 基因外显子较少，而 *CaNMT1c* 基因外显子数量较多（图 3），显示出基

因结构上的异质性。保守基序分析结果进一步表明，*NMT* 家族成员均包含 1 个核心保守基序，属于 *Methyltransf_7* 家族或其超家族（*Methyltransf_7* superfamily）（图 3），反映了 *NMT* 蛋白在功能上的保守性及潜在的多样化作用。

2.4 顺式作用元件分析

CaNMT 基因含有大量的顺式作用元件，主要参与光应答、胁迫响应和激素响应。在光响应调控方面，*CaNMT* 基因启动子包含 16 种不同的光响应元件，这些元件广泛分布于 27 个 *CaNMT* 基因的启动子区域，包括 AE-box、TCT-motif、ACE、GT1-motif、I-box、ATCT-motif、Box 4、C-box、G-box、GA-motif、GATA-motif、AT1-motif、MRE、Sp1 和 chs-CMA2a 等（图 4）。这些元件可能介导 *CaNMT* 基因对光周期、光质及光照强度变化的响应。例如，*CaNMT4c* 和 *CaNMT15e* 等基因启动子中富含 G-box 和 I-box 元件，可能增强对光周期和光质变化的敏感性，而 *CaNMT1c* 和 *CaNMT20e*

温和厌氧条件的潜在响应能力。这些元件的存在表明 NMT 基因可能通过调控咖啡因含量在植物抗逆应答中发挥独特作用, 尤其是面对干旱和低温等环境胁迫时, 咖啡因作为次生代谢产物的抗逆特性可能与 *CaNMT* 基因的表达调控密切相关。

在植物激素信号转导方面, *CaNMT* 基因启动子包含多种激素响应元件, 比如生长素 (IAA) 响应元件 (TGA-element、AuxRR-core)、水杨酸 (SA) 响应元件 (TCA-element)、赤霉素 (GA) 响应元件 (P-box、TATC-box)、茉莉酸甲酯 (MeJA) 响应元件 (CGTCA-motif、TGACG-motif) 和脱落酸 (ABA) 响应元件 (ABRE)。例如, *CaNMT26e* 和 *CaNMT18e* 中 TGA-element 和 AuxRR-core 的富集可能增强对 IAA 信号的敏感性, 而 *CaNMT10c* 和 *CaNMT21e* 中 CGTCA-motif 的分布提示其对 MeJA 介导的防御反应的调控潜力。

此外, 启动子区域还包含类黄酮生物合成相关元件 (MBSI, 如 *CaNMT9c*)、玉米醇溶蛋白代谢相关元件 (O2-site, 如 *CaNMT3c*)、昼夜节律调控元件 (circadian, 如 *CaNMT19e*)、细胞分化相关元件 (HD-Zip 1, 如 *CaNMT1c*)、胚乳特异性表达元件 (GCN4_motif, 如 *CaNMT6c*) 以及分生组织表达相关元件 (CAT-box, 如 *CaNMT2c*), 这些元件参与特定生物学过程, 可能与 *CaNMT* 基因在咖啡果实发育和种子成熟中的功能分化相关。

同时, *CaNMT* 基因启动子区域还包含多种转录因子结合位点, 包括 MYB、MYC 和 W-box 等。这些元件是相应转录因子识别并结合的特异性 DNA 序列, 在调控基因表达中发挥关键作用。其中, W-box 作为 WRKY 转录因子的典型结合基序, 广泛参与植物胁迫响应和免疫防御; 而 MYB 和 MYC 结合位点则与干旱、高盐、低温胁迫及多种植物激素信号通路的调控密切相关。

2.5 *CaNMT* 基因在不同组织中表达分析

对小粒种果实不同发育时期的转录组数据进行分析发现, 果实发育 Fruit-1、Fruit-2 阶段有 4797 个差异表达基因, Fruit-2、Fruit-3 阶段有 4419 个, Fruit-3、Fruit-4 阶段高达 14703 个, 而 Fruit-4、Fruit-5 阶段则为 6339 个。其中, Fruit-2、Fruit-3 差异表达基因数量最少, Fruit-3、Fruit-4 则最多 (图 5)。进一步比较发现, Fruit-1、Fruit-2 时期特有 1065 个差异表达基因, Fruit-2、Fruit-3 时期特有 755 个差异表达基因, Fruit-3、Fruit-4 时期

特有 7662 个差异表达基因, 而 Fruit-4、Fruit-5 时期特有 2062 个差异表达基因。与各阶段间特有的基因相比, 共享的差异表达基因数量相对较少, 这表明特异性表达基因在果实发育中具有更重要的作用。对不同发育时期特有的差异表达基因进行 KEGG 富集, 发现 Fruit-1、Fruit-2 阶段的基因主要参与跨膜转运、脂肪酸分解、昼夜节律以及过氧化物酶体等过程。Fruit-2、Fruit-3 阶段, 基因主要参与 DNA 修复与重组、蛋白质加工等过程, 同时涉及部分氨基酸和脂类代谢。在 Fruit-3、Fruit-4 阶段, 基因主要富集在光合作用、核苷酸代谢、能量代谢, 以及泛素系统和 RNA 降解等关键通路。Fruit-4、Fruit-5 阶段的基因主要参与糖酵解、氨基酸与脂类代谢、核糖体生成以及细胞骨架构建等核心活动 (图 5)。

对咖啡不同组织 (根、茎、叶、果实) 及果实不同发育时期 (Fruit-1~Fruit-5) 的转录组数据分析, *CaNMT* 基因表现出动态的表达模式, 并具有组织特异性 (图 6)。*CaNMT2c* 在果实早期发育阶段 (Fruit-1) 表现出极高的表达量, 在 Fruit-2 和 Fruit-3 阶段仍保持较高水平, 但在 Fruit-4 和 Fruit-5 阶段逐渐下降, 表明其可能在果实早期发育中具有重要作用。类似地, *CaNMT11c* 在果实发育早期至中期 (Fruit-1~Fruit-4) 均维持较高表达量, 仅在 Fruit-5 略有下降, 显示其在果实成熟过程中具有持续的功能活性。相比之下, *CaNMT16e* 在 Fruit-1 阶段表达较高, 但在后续阶段迅速下降 (Fruit-3 后接近 0), 表明其功能主要集中在果实发育的初期。而 *CaNMT23e* 则在 Fruit-4 阶段达到表达峰值, 远高于 Fruit-1, 表明其可能在果实后期发育中发挥重要作用。

CaNMT 基因在果实 (Fruit-5)、叶、根和茎等不同组织中展现出组织特异性。*CaNMT2c* 在茎表达量最高, 其次是叶和根, 而果实中较低; *CaNMT18e* 在茎表达高于叶和果实, 表明其在茎部和叶片代谢中可能具有主要作用。*CaNMT10c* 和 *CaNMT26e* 在叶表达量最高, Fruit-5 较低, 反映其在叶片光合作用或代谢中的重要性。*CaNMT1c* 在根表达量最高, 叶次之, Fruit-5 极低, 表明其可能与根部营养吸收相关。

2.6 咖啡不同时期 WGCNA 表达网络分析

通过转录组分析共获得 6 6023 个基因的表达谱数据, 过滤低表达量的基因, 获得高表达基因用于基因共表达网络构建。通过对权重值采用动态剪

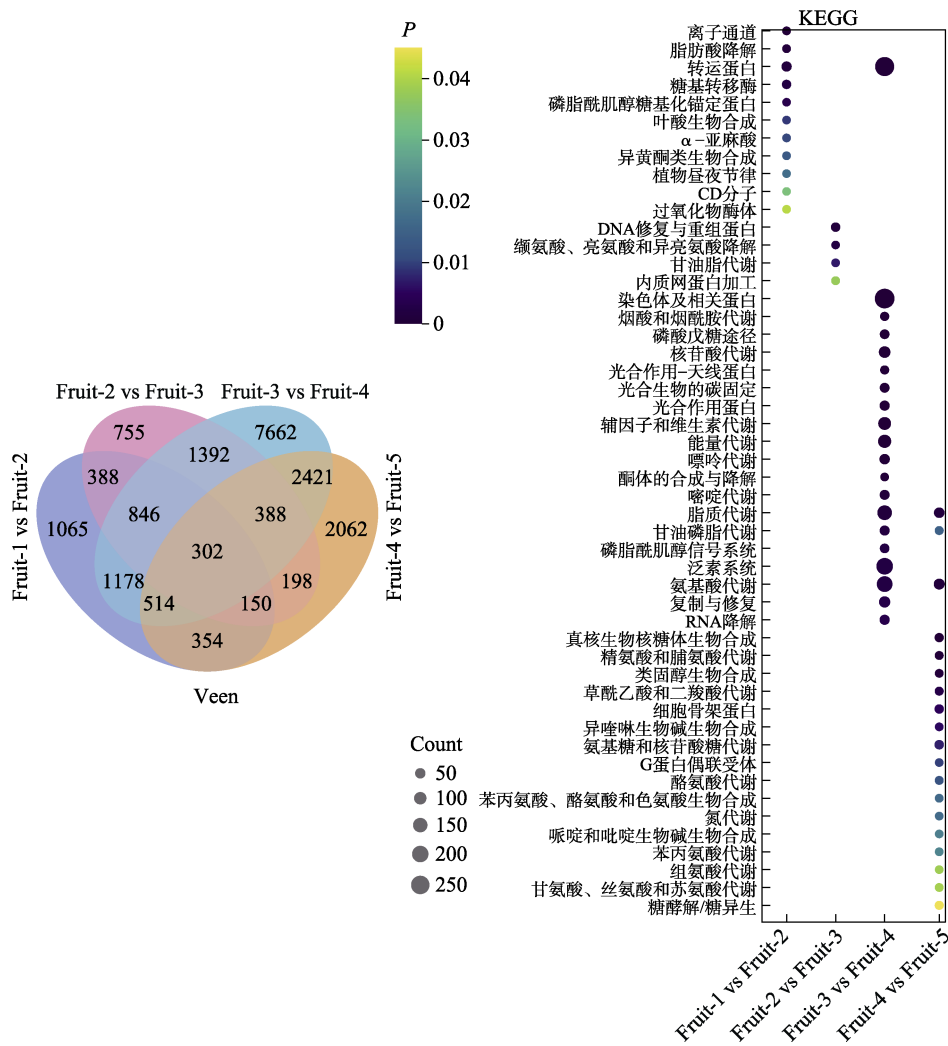


图 5 果实不同发育时期差异表达基因 Venn 图及其特有差异表达基因 KEGG 富集分析

Fig. 5 Venn diagram of differentially expressed genes at different fruit developmental stages and KEGG enrichment analysis of stage-specific DEGs

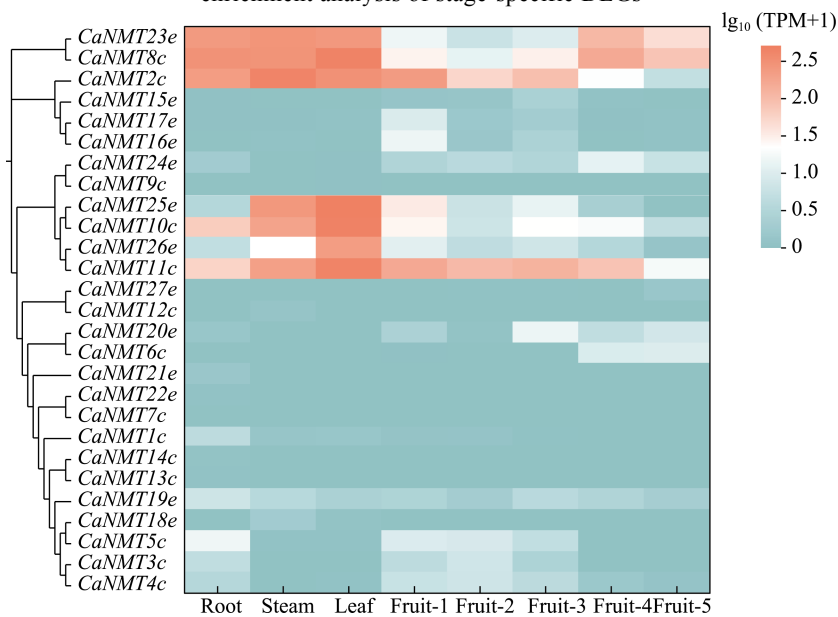


图 6 小粒种咖啡根、茎、叶和果实不同发育阶段 *CaNMT* 基因表达

Fig. 6 Expressions of *CaNMT* genes in the roots, leaves, stems and fruits at different developmental stages of *C. arabica*

切树法合并表达相似的模块，获得 12 个共表达模块，不同的颜色代表不同的模块。每个模块包含的基因数目不同，darkseagreen 4 模块包含基因数量最多，为 9193 个基因，grey 模块包含基因数目最少，为 279 个基因，平均每个模块包含 3159 个基因（图 7）。

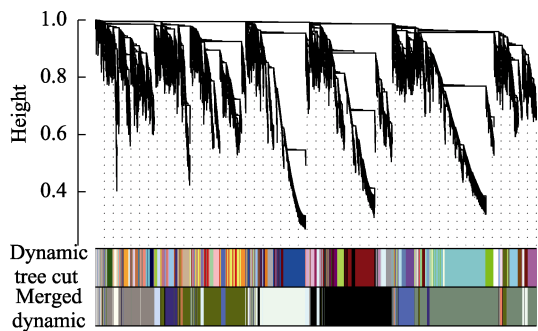


图 7 共表达模块与层次聚类树

Fig. 7 Co-expression module and hierarchical clustering tree

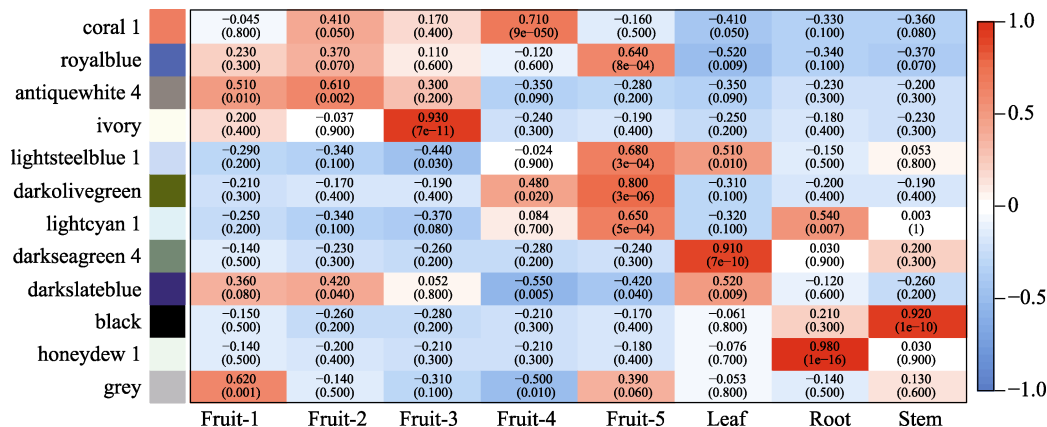
按照 $|r|>0.70$ 且 $P<0.005$ 的标准，6 个模块均与组织存在高度相关性，比如 honeydew 1 ($r=0.98$, $P<1e-16$) 与根高度相关, black($r=0.92$, $P<1e-10$) 与茎高度相关, darkseagreen 4 ($r=0.91$, $P<7e-10$) 与叶高度相关（图 8），ivory ($r=0.93$, $P<7e-11$) 与果实高度相关。为筛选与咖啡发育相关基因，本研究对与咖啡发育显著关联的模块基因进行富集分析。

GO 富集分析结果显示，在生物学过程（biological process）方面，ivory 模块主要富集于与环境应激响应和蛋白质稳态维持相关的过

程，包括响应高温（GO:0009408）、响应高光强度（GO:0009644）、蛋白质折叠（GO:0006457）等；coral1 模块主要富集于与能量代谢相关的过程，包括碳水化合物磷酸化（GO:0046835）、丙酮酸生物合成过程（GO:0042866）、糖酵解过程（GO:0006096）等；darkolivegreen 模块主要富集于核酸代谢过程，包括核酸代谢过程（GO:0090304）、RNA 代谢过程（GO:0016070）、基因表达（GO:0010467）、mRNA 代谢过程（GO:0016071）和 RNA 加工（GO:0006396）。在分子功能（molecular function）方面，ivory 模块主要富集于半乳糖基转移酶活性（GO:0008378）；coral1 模块主要富集于碳水化合物激酶活性（GO:0019200）；darkolivegreen 模块主要富集于核酸内切酶活性（GO:0004519）。在细胞组分（cellular component）方面，ivory 模块主要富集于脂滴（GO:0005811）和内质网（GO:0005783）；darkolivegreen 模块主要富集于细胞内解剖结构（GO:0005622）和核蛋白复合物（GO:0140513）。coral1 模块主要富集于细胞质（GO:0005829）（图 9）。

KEGG 富集分析结果显示，ivory 模块主要富集在脂质代谢、半乳糖代谢和糖酵解/糖异生等通路；coral1 模块主要富集在果糖和甘露糖代谢、蛋白质磷酸酶及其相关蛋白、半胱氨酸和蛋氨酸代谢等通路；darkolivegreen 模块主要富集在维生素 B₆ 代谢、谷胱甘肽代谢、植物昼夜节律等通路（图 9）。

选取 ivory、coral1 和 darkolivegreen 模块中选取连通度排名前 10 的基因作为核心基因，并以权重值



括号中的数值为 P 值。
The values in parentheses are P values.

图 8 基因共表达网络模块与性状的关联热图

Fig. 8 Association analysis of gene co-expression network modules with traits

>0.20 为标准, 构建核心基因及其关联基因的互作网络图。在 ivory 模块中, 筛选到 NMT 基因家族成员 *CaNMT15e* 和 *CaNMT20e*; 在 darkolivegreen 模块中, 筛选到 NMT 基因家族成员 *CaNMT24e* (图 10)。

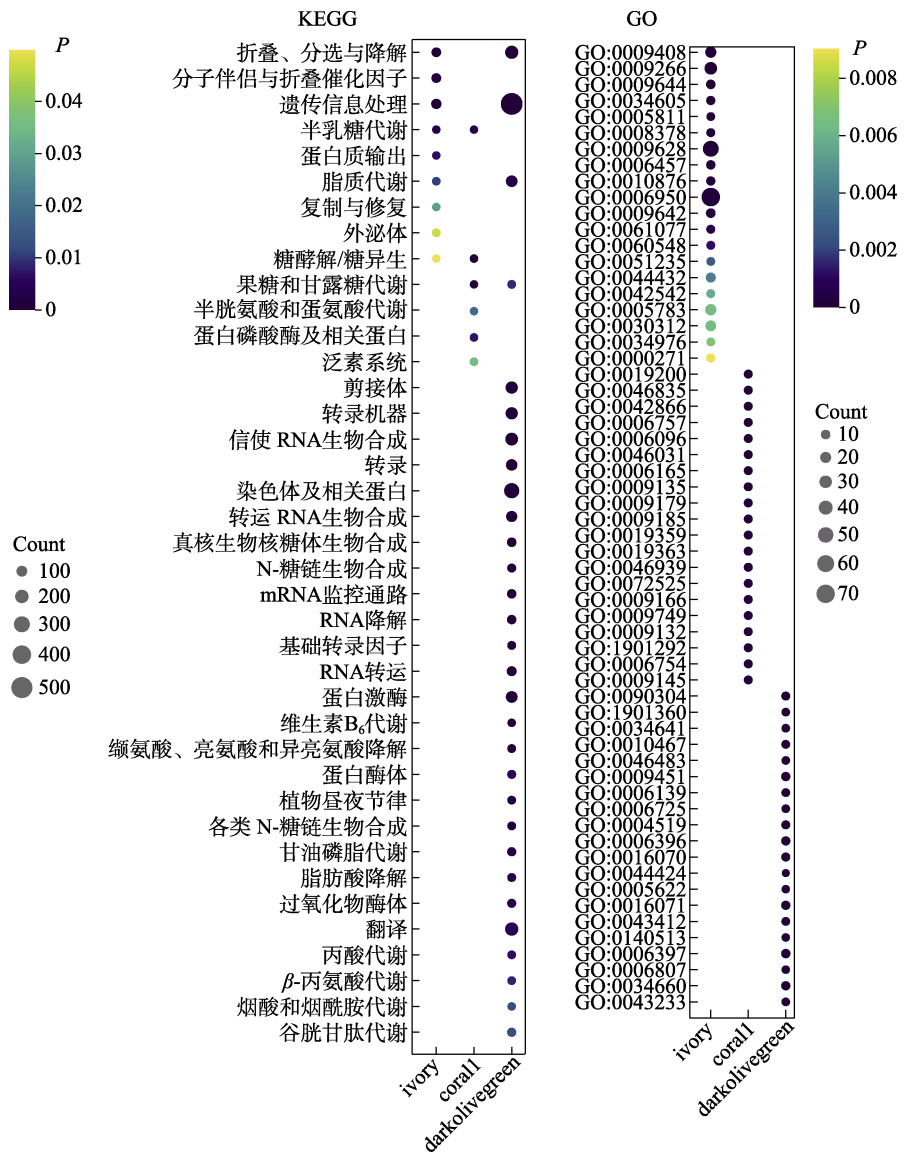


图 9 目标模块内基因 KEGG 及 GO 富集分析
Fig. 9 KEGG and GO enrichment analysis of genes in target module

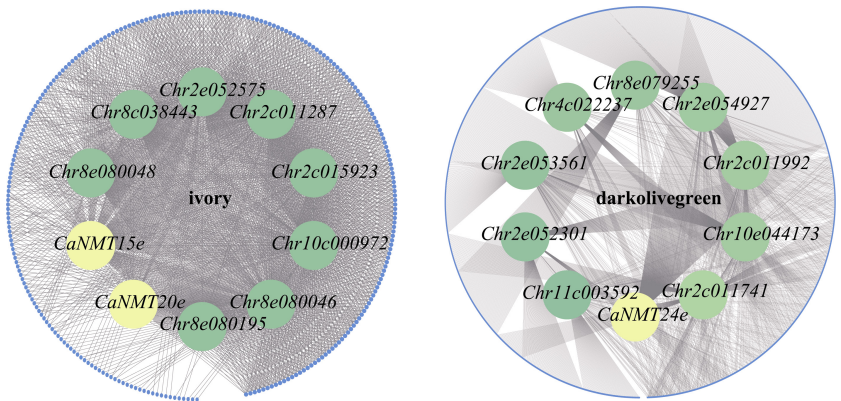


图 10 ivory 和 darkolivegreen 模块内的基因共表达网络及其核心基因
Fig. 10 Gene co-expression network and hub genes in ivory and darkolivegreen modules

3 讨 论

本研究系统鉴定了小粒种咖啡 NMT 基因家族成员,解析了其在不同组织及果实不同发育时期的特异表达模式,并借助加权基因共表达网络分析(WGCNA)在多组织和果实发育阶段构建了共表达网络,进而在模块水平上分析了与 NMT 家族协同调控的核心基因,为后续研究提供了候选基因和潜在调控靶点。

基于基因组学分析,本研究鉴定到 27 个小粒种咖啡 NMT 家族基因,观察到其在 Chr1 和 Chr9 染色体上的基因聚簇分布,以及亚基因组间相似的分布格局,这符合局部复制与染色体重排共同塑造基因家族结构的机制^[30]。从结构域层面分析,家族成员普遍携带 Methyltransf_7 保守结构域,该域定义了底物识别与 SAM 依赖的转甲基化共同化学框架,为在相同催化框架上演化出位点特异性的 N7/N3/N1 甲基化提供了基础。已有研究基于 GFP 融合蛋白的原位观察,发现咖啡 3 类 NMT (XMT/MXMT/DXMT) 主要定位于细胞质,并可形成二聚体^[16,31],这与本研究的亚细胞定位预测结果基本一致,表明细胞质是咖啡因生物合成的主要场所。值得注意的是,本研究预测到 CaNMT4c 和 CaNMT19e 蛋白定位于叶绿体,这暗示 NMT 基因家族可能演化出了新的功能。叶绿体是多种次生代谢物(如部分萜类、生物碱)合成的场所^[32],这 2 个 NMT 蛋白可能参与了与光合作用偶联的特定甲基化修饰过程,或催化了某种未知底物的合成,这为理解咖啡因之外的 NMT 功能提供了新线索。此外,理化性质分析揭示了 CaNMT 蛋白在氨基酸长度、等电点和稳定性上的显著多样性,这种多样性可能反映了不同成员在底物亲和力、酶活调控以及与其他蛋白相互作用模式上的分化,是基因家族成员功能细化的体现。此外,本研究在 CaNMT 基因启动子上检测到 G-box、ACE、I-box 等典型光响应元件,以及 ABRE、TCA-element、MeJA 和 MBS 等激素/胁迫响应元件。已有研究表明, G-box 作为 bZIP/bHLH (如 HY5、MYC 家族) 转录因子的核心结合位点,调控大量次生代谢基因的表达^[33];同时,JA/SA 信号途径可直接调控咖啡或茶中咖啡因相关 NMT 基因的转录水平^[34-35]。这些发现表明环境因子和激素信号在调控咖啡 NMT 基因表达中具有重要作用。

在表达模式方面,本研究观察到 CaNMT 基因在果实转色期表现出显著的差异表达峰值,这与

先前研究中咖啡果皮/种子发育过程中色素与挥发性化合物积累的转变以及碳代谢重分配的观点一致^[36-37]。相关研究显示,JA/SA 处理可上调咖啡因合成途径中 NMT 基因的转录水平并提升咖啡因含量^[38],且已鉴定的转录因子可直接激活关键合成酶基因的启动子^[39],表明在果实发育关键节点,转录调控网络驱动了代谢通量的重新分配;然而,关于咖啡因是否作为内源信号分子及其通量调控的动态机制,仍需通过酶活性测定与同位素示踪等进一步验证。除了在果实中的动态表达,本研究还探究了 CaNMT 基因在根、茎、叶等器官中的组织特异性表达模式,这表明咖啡因在植物整体防御体系中的多重角色。这种组织特异性的表达模式,体现了植物在不同部位精准调配防御资源以应对环境胁迫的精细策略,也说明 NMT 基因家族的调控网络远比我们之前认为的更为复杂。

进一步结合 WGCNA 分析结果,本研究发现 CaNMT15e 和 CaNMT20e 这 2 个基因在与果实发育呈高度正相关的 ivory 模块中表现出高连通度。根据富集分析,ivory 模块中的基因主要参与糖酵解/糖异生、半乳糖代谢和脂质代谢等核心能量代谢通路。咖啡因作为嘌呤生物碱,其生物合成是一个能量密集型过程,需要消耗大量的 ATP^[40]。糖酵解等中心碳代谢途径正是细胞内 ATP 的主要来源,为合成反应提供必需的能量。因此,ivory 模块中能量代谢通路的活跃,支持了细胞在果实成熟过程中维持高水平的 ATP 供应,这对于合成咖啡因至关重要。此外,所有 NMT 酶催化的甲基化反应均依赖 S-腺苷甲硫氨酸(SAM)作为甲基供体,而 SAM 的再生循环本身也高度依赖 ATP 供应。ivory 模块中能量代谢通路的活跃,不仅为 ATP 供应提供保障,还通过维持细胞内 SAM 的高水平,确保 CaNMT 酶能够高效催化甲基化反应。另一方面,CaNMT24e 在与 RNA 代谢和基因表达相关的 darkolivegreen 模块中处于核心地位,这表明咖啡因的合成调控不仅仅局限于代谢层面,还涉及转录及转录后等多个层次的精密调控,体现了其调控网络的复杂性。

已有研究通过 RNAi 技术抑制咖啡中 MXMT 和 XMT 等基因,成功获得了咖啡因含量显著降低的植株^[18,41]。此外,在异源表达系统中同时表达 XMT、MXMT 和 DXMT 3 种酶,能够表现出抗虫效果^[42-43]。结合本研究的 WGCNA 模块分析和

顺式元件解析, 未来可利用 CRISPR 技术精准调控 NMT 基因家族核心成员或其启动子中的关键元件, 从而实现降低咖啡因含量的同时保留咖啡风味和抗逆性的育种目标。

参考文献

- [1] EMADI R C, KAMANGAR F. Coffee's impact on health and well-being[J]. *Nutrients*, 2025, 17(15): 2558.
- [2] BARRETO PEIXOTO J A, SILVA J F, OLIVEIRA M, ALVES R C. Sustainability issues along the coffee chain: from the field to the cup[J]. *Comprehensive Reviews in Food Science and Food Safety*, 2023, 22(1): 287-332.
- [3] 闫林, 黄丽芳, 王晓阳, 孙燕, 林兴军, 董云萍, 龙宇宙. 咖啡种质资源遗传多样性的 ISSR 分析[J]. *南方农业学报*, 2019, 50(3): 491-499.
YAN L, HUANG L F, WANG X Y, SUN Y, LIN X J, DONG Y P, LONG Y Z. Genetic diversity of coffee germplasms by ISSR analysis[J]. *Journal of Southern Agriculture*, 2019, 50(3): 491-499. (in Chinese)
- [4] MAHMUD M M C, SHELLIE R A, KEAST R. Unravelling the relationship between aroma compounds and consumer acceptance: coffee as an example[J]. *Comprehensive Reviews in Food Science and Food Safety*, 2020, 19(5): 2380-2420.
- [5] VEZZULLI F, ROCCHETTI G, LAMBRI M, LUCINI L. Metabolomics combined with sensory analysis reveals the impact of different extraction methods on coffee beverages from *Coffea arabica* and *Coffea canephora* var. Robusta[J]. *Foods*, 2022, 11(6): 807.
- [6] FIANI B, ZHU L, MUSCH B L, BRICENO S, ANDEL R, SADEQ N, ANSARI A Z. The neurophysiology of caffeine as a central nervous system stimulant and the resultant effects on cognitive function[J]. *Cureus*, 2021, 13(5): e15032.
- [7] ZHAO E H, ERGUL B, ZHAO W. Caffeine's antioxidant potency optically sensed with double-stranded DNA-encased single-walled carbon nanotubes[J]. *The Journal of Physical Chemistry B*, 2015, 119(10): 4068-4075.
- [8] DENOEUDE F, CARRETERO-PAULET L, DEREPPER A, DROC G, GUYOT R, PIETRELLA M, ZHENG C, ALBERTI A, ANTHONY F, APREA G, AURY J M, BENTO P, BERNARD M, BOCS S, CAMPA C, CENCI A, COMBES M C, CROUZILLAT D, DA SILVA C, DADDIEGO L, DE BELLIS F, DUSSERT S, GARSMEUR O, GAYRAUD T, GUIGNON V, JAHN K, JAMILLOUX V, JOET T, LABADIE K, LAN T, LECLERCQ J, LEPELLEY M, LEROY T, LI L T, LIBRADO P, LOPEZ L, MUÑOZ A, NOEL B, PALLAVICINI A, PERROTTA G, PONCET V, POT D, PRIYONO, RIGOREAU M, ROUARD M, ROZAS J, TRANCHANT-DUBREUIL C, VANBUREN R, ZHANG Q, ANDRADE A C, ARGOUT X, BERTRAND B, DE KOCHKO A, GRAZIOSI G, HENRY R J, JAYARAMA, MING R, NAGAI C, ROUNSLEY S, SANKOFF D, GIULIANO G, ALBERT V A, WINCKER P, LASHERMES P. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis[J]. *Science*, 2014, 345(6201): 1181-1184.
- [9] SALOJARVI J, RAMBANI A, YU Z, GUYOT R, STRICKLER S, LEPELLEY M, WANG C, RAJARAMAN S, RASTAS P, ZHENG C, MUNOZ D S, MEIDANIS J, PASCHOAL A R, BAWIN Y, KRABBENHOFT T J, WANG Z Q, FLECK S J, AUSSEL R, BELLANGER L, CHARPAGNE A, FOURNIER C, KASSAM M, LEFEBVRE G, METAIRON S, MOINE D, RIGOREAU M, STOLTE J, HAMON P, COUTURON E, TRANCHANT-DUBREUIL C, MUKHERJEE M, LAN T, ENGELHARDT J, STADLER P, CORREIA DE LEMOS S M, SUZUKI S I, SUMIRAT U, WAI C M, DAUCHOT N, OROZCO-ARIAS S, GARAVITO A, KIWUKA C, MUSOLI P, NALUKENGE A, GUICHOUX E, REINOUT H, SMIT M, CARRETERO-PAULET L, FILHO O G, BRAGHINI M T, PADILHA L, SERA G H, RUTTINK T, HENRY R, MARRACCINI P, VAN DE PEER Y, ANDRADE A, DOMINGUES D, GIULIANO G, MUELLER L, PEREIRA L F, PLAISANCE S, PONCET V, ROMBAUTS S, SANKOFF D, ALBERT V A, CROUZILLAT D, DE KOCHKO A, DESCOMBES P. The genome and population genomics of allopolyploid *Coffea arabica* reveal the diversification history of modern coffee cultivars[J]. *Nature Genetics*, 2024, 56(4): 721-731.
- [10] MEDRANO J F, CANTU D, MINIO A, DREISCHER C, GIBBONS T, CHIN J, CHEN S, VAN DEYNZE A, HULSE-KEMP A M. *De novo* whole-genome assembly and annotation of *Coffea arabica* var. Geisha, a high-quality coffee variety from the primary origin of coffee[J]. *G3 Genes Genomes Genetics*, 2025, 15(1): jkae262.
- [11] COMBES M C, JOET T, STAVRINIDES A K, LASHERMES P. New cup out of old coffee: contribution of parental gene expression legacy to phenotypic novelty in coffee beans of the allopolyploid *Coffea arabica* L.[J]. *Annals of Botany*, 2023, 131(1): 157-170.
- [12] 张振伟, 董相书, 杨婧, 杞美军, 王步天, 施学东, 黄文龙, 李学俊, 葛宇. 小粒咖啡 CaFAB2 基因家族的鉴定及其在种子脂肪酸合成中的潜在作用[J]. *农业生物技术学报*, 2025, 33(1): 79-91.
ZHANG Z W, DONG X S, YANG J, QI M J, WANG B T, SHI X D, HUANG W L, LI X J, GE Y. Identification of the

- CaFAB2 gene family in *Coffea arabica* and its potential role in seed fatty acid synthesis[J]. Journal of Agricultural Biotechnology, 2025, 33(1): 79-91. (in Chinese)
- [13] RAHARIMALALA N, ROMBAUTS S, MCCARTHY A, GARAVITO A, OROZCO-ARIAS S, BELLANGER L, MORALES-CORREA A Y, FROGER S, MICHAUX S, BERRY V, METAIRON S, FOURNIER C, LEPELLEY M, MUELLER L, COUTURON E, HAMON P, RAKOTOMALALA J J, DESCOMBES P, GUYOT R, CROUZILLAT D. The absence of the caffeine synthase gene is involved in the naturally decaffeinated status of *Coffea humblotiana*, a wild species from Comoro archipelago[J]. Scientific Reports, 2021, 11(1): 8119.
- [14] UEFUJI H, OGITA S, YAMAGUCHI Y, KOIZUMI N, SANO H. Molecular cloning and functional characterization of three distinct N-methyltransferases involved in the caffeine biosynthetic pathway in coffee plants[J]. Plant Physiology, 2003, 132(1): 372-380.
- [15] MIZUNO K, OKUDA A, KATO M, YONEYAMA N, TANAKA H, ASHIHARA H, FUJIMURA T. Isolation of a new dual-functional caffeine synthase gene encoding an enzyme for the conversion of 7-methylxanthine to caffeine from coffee (*Coffea arabica* L.)[J]. FEBS Letters, 2003, 534(1/3): 75-81.
- [16] MCCARTHY A A, MCCARTHY J G. The structure of two N-methyltransferases from the caffeine biosynthetic pathway[J]. Plant Physiology, 2007, 144(2): 879-889.
- [17] SCALABRIN S, MAGRIS G, LIVA M, VITULO N, VIDOTTO M, SCAGLIONE D, DEL TERRA L, RUOSI M R, NAVARINI L, PELLEGRINO G, BERNY MIER Y T J C, TONIUTTI L, SUGGI LIVERANI F, CERUTTI M, DI GASPERO G, MORGANTE M. A chromosome-scale assembly reveals chromosomal aberrations and exchanges generating genetic diversity in *Coffea arabica* germplasm[J]. Nature Communications, 2024, 15(1): 463.
- [18] OGITA S, UEFUJI H, MORIMOTO M, SANO H. Application of RNAi to confirm theobromine as the major intermediate for caffeine biosynthesis in coffee plants with potential for construction of decaffeinated varieties[J]. Plant Molecular Biology, 2004, 54(6): 931-941.
- [19] CHEN W, SALARI H, TAYLOR M C, JOST R, BERKOWITZ O, BARROW R, QIU D, BRANCO R, MASLE J. NMT1 and NMT3 N-methyltransferase activity is critical to lipid homeostasis, morphogenesis, and reproduction[J]. Plant Physiology, 2018, 177(4): 1605-1628.
- [20] ZHANG B, HORVATH S. A general framework for weighted gene co-expression network analysis[J]. Statistical Applications in Genetics and Molecular Biology, 2005, 4: 17.
- [21] YIN M, ZHENG Z Z, ZHANG Y, WANG S W, ZUO L Y, LEI Y, ZHAO Y Q, ZHAO X Q, FU B Y, SHI Y Y, XU J L, WANG W S. Identification of key genes and pathways for anaerobic germination tolerance in rice using weighted gene co-expression network analysis (WGCNA) in association with quantitative trait locus (QTL) mapping[J]. Rice, 2024, 17: 37.
- [22] WANG Y Z, NIE L H, MA J, ZHOU B, HAN X H, CHENG J L, LU X M, FAN Z F, LI Y L, CAO Y Y. Transcriptomic variations and network hubs controlling seed size and weight during maize seed development[J]. Frontiers in Plant Science, 2022, 13: 828923.
- [23] CAMACHO C, COULOURIS G, AVAGYAN V, MA N, PAPADOPOULOS J, BEALER K, MADDEN T L. BLAST+: architecture and applications[J]. BMC Bioinformatics, 2009, 10: 421.
- [24] LETUNIC I, KHEDKAR S, BORK P. SMART: recent updates, new developments and status in 2020[J]. Nucleic Acids Research, 2021, 49(D1): D458-D460.
- [25] KUMAR S, STECHER G, LI M, KNYAZ C, TAMURA K, MEKA X. Molecular evolutionary genetics analysis across computing platforms[J]. Molecular Biology and Evolution, 2018, 35(6): 1547-1549.
- [26] CHEN C, CHEN H, ZHANG Y, THOMAS H R, FRANK M H, HE Y, XIA R. TBtools: an integrative toolkit developed for interactive analyses of big biological data[J]. Molecular Plant, 2020, 13(8): 1194-1202.
- [27] QUINLAN A R, HALL I M. BEDTools: a flexible suite of utilities for comparing genomic features[J]. Bioinformatics, 2010, 26(6): 841-842.
- [28] KIM D, LANGMEAD B, SALZBERG S L. HISAT: a fast spliced aligner with low memory requirements[J]. Nature Methods, 2015, 12(4): 357-360.
- [29] LIAO Y, SMYTH G K, SHI W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features[J]. Bioinformatics, 2014, 30(7): 923-930.
- [30] PANCHY N, LEHTI-SHIU M, SHIU S H. Evolution of gene duplication in plants[J]. Plant Physiology, 2016, 171(4): 2294-2316.
- [31] KODAMA Y, SHINYA T, SANO H. Dimerization of N-methyltransferases involved in caffeine biosynthesis[J]. Biochimie, 2008, 90(3): 547-551.
- [32] FU P P, XIA Q, LIN G, CHOU M W. Pyrrolizidine alkaloids—genotoxicity, metabolism enzymes, metabolic activation, and mechanisms[J]. Drug Metabolism Reviews, 2004, 36(1): 1-55.

- [33] EZER D, SHEPHERD S J K, BRESTOVITSKY A, DICKINSON P, CORTIJO S, CHAROENSAWAN V, BOX M S, BISWAS S, JAEGER K E, WIGGE P A. The G-box transcriptional regulatory code in *Arabidopsis*[J]. *Plant Physiology*, 2017, 175(2): 628-640.
- [34] YE Z L, ZHENG Z Q, WANG J S, ZHONG Q Y, MU X Y, YUAN Z F, HUANG Y, GHAZALA Z, ZUO H, LI Y L, LIANG X Y, QI Y J, SI X Y, ZHANG Y R, XIA E H, ZHAO J, LI P H. JA regulates caffeine biosynthesis in tea leaf for resistance against fungal infection and antagonises with GA to balance the defence-growth trade-off via CsDELLA-JAZ-MYC2-MYB184-TCS1 module[J/OL]. *Plant Biotechnology Journal*, 1-26(2025-09-03)[2025-09-06]. <https://doi.org/10.1111/pbi.70355>.
- [35] KUMAR A, NAIK G K, GIRIDHAR P. Dataset on exogenous application of salicylic acid and methyljasmonate and the accumulation of caffeine in young leaf tissues and catabolically inactive endosperms[J]. *Data in Brief*, 2017, 13: 22-27.
- [36] CHENG B, FURTADO A, HENRY R J. The coffee bean transcriptome explains the accumulation of the major bean components through ripening[J]. *Scientific Reports*, 2018, 8(1): 11414.
- [37] PERROIS C, STRICKLER S R, MATHIEU G, LEPELLEY M, BEDON L, MICHAUX S, HUSSON J, MUELLER L, PRIVAT I. Differential regulation of caffeine metabolism in *Coffea arabica* (Arabica) and *Coffea canephora* (Robusta)[J]. *Planta*, 2015, 241(1): 179-191.
- [38] KUMAR A, GIRIDHAR P. Salicylic acid and methyljasmonate restore the transcription of caffeine biosynthetic N-methyltransferases from a transcription inhibition noticed during late endosperm maturation in coffee[J]. *Plant Gene*, 2015, 4: 38-44.
- [39] MA W, KANG X, LIU P, SHE K, ZHANG Y, LIN X, LI B, CHEN Z. The NAC-like transcription factor CsNAC7 positively regulates the caffeine biosynthesis-related gene *yhNMT1* in *Camellia sinensis*[J]. *Horticulture Research*, 2022, 9: uhab046.
- [40] ASHIHARA H, SANO H, CROZIER A. Caffeine and related purine alkaloids: biosynthesis, catabolism, function and genetic engineering[J]. *Phytochemistry*, 2008, 69(4): 841-856.
- [41] OGITA S, UEFUJI H, YAMAGUCHI Y, KOIZUMI N, SANO H. Producing decaffeinated coffee plants[J]. *Nature*, 2003, 423(6942): 823.
- [42] UEFUJI H, TATSUMI Y, MORIMOTO M, KAOTHIEN-NAKAYAMA P, OGITA S, SANO H. Caffeine production in tobacco plants by simultaneous expression of three coffee N-methyltransferases and its potential as a pest repellent[J]. *Plant Molecular Biology*, 2005, 59(2): 221-227.
- [43] KIM Y S, SANO H. Pathogen resistance of transgenic tobacco plants producing caffeine[J]. *Phytochemistry*, 2008, 69(4): 882-888.