

油茶 BAHD 基因家族全基因组鉴定与组织特异性表达分析

摘要:【目的】本研究在全基因组水平上检测油茶 BAHD 基因家族成员,挖掘鉴定油茶 BAHD 酰基转移酶家族基因,系统分析候选基因在油茶植株不同组织部位的特异性表达模式,为解析油茶皂苷类化合物生物合成机制奠定基础。【方法】通过 Pfam 转移酶家族蛋白质结构模型 (PF02458),以及 BAHD 家族的保守基序 HXXXD 和 DFGWG 来鉴定油茶 BAHD 家族基因,并结合油茶植株不同部位转录组数据及实时荧光定量 PCR 筛选 BAHD 家族中与萜类化合物酰基化修饰相关的关键候选基因。【结果】从油茶基因组中全面鉴定出 158 个 BAHD 基因,所有的油茶 BAHD 基因随机分布在油茶的 15 条染色体上,在进化过程中产生 14 对重复基因。系统发育分析表明,这些油茶 BAHD 基因可划分为 6 个分支。BAHD 的蛋白基序和基因结构在各组中高度保守,反映了它们的功能保守性。共线性分析表明,大部分 BAHD 基因可能来自纯化选择的串联重复,这为了解 BAHD 基因家族成员之间可能存在的功能差异提供了新的思路。此外,在 BAHD 启动子上发现了大量的激素和应激应答元件。利用 qRT-PCR 验证 BAHD 基因中参与萜类化合物修饰的核心成员,发现 *Co10g1264.11*、*Co04g0659.15*、*Co06g0230.08* 基因在叶、根和茎中的表达量较高,而 *Co10g1264.11* 基因在果中的表达量最高。【结论】油茶不同组织中 BAHD 家族基因的组织特异性表达,体现了其功能的多样性,了解 BAHD 家族基因的表达模式有助于推测基因的功能。通过 qRT-PCR 分析发现,油茶的根、叶及种仁部分 BAHD 基因具有显著表达差异。其中 *Co10g1264.11*、*Co04g0659.15*、*Co06g0230.08* 基因在叶、根和茎中的表达量较高;*Co10g1264.11* 基因在种仁的表达量最高,而其他基因在种仁中表达量较低,这些成员可作为后续功能验证研究的核心候选基因。本研究挖掘了油茶皂苷 BAHD 酰基化修饰的关键候选酶基因,阐明油茶 BAHD 家族的特征,为进一步解析油茶皂苷生物合成途径奠定了研究基础。

关键词: 油茶; 酰基转移酶; BAHD 家族; 基因挖掘

Whole Genome Identification and Tissue-Specific Expression Analysis of the BAHD Gene Family in *Camellia oleifera*

Abstract: 【Objective】 To identify and characterize the BAHD acyltransferases of *Camellia oleifera*, the BAHD gene family members were detected at the whole genome level. We systematically analyze the specific expression patterns of candidate genes in different tissues, to obtain key acyltransferase genes for elucidating the biosynthesis mechanism of camellia saponins. 【Methods】 The BAHD gene family was identified using the Pfam acyltransferase family protein structure model (PF02458) and the conserved motifs HXXXD and DFGWG of the BAHD family. Transcriptome data and real-time fluorescence quantitative PCR from different tissues of the *C. oleifera* were used to screen for candidate genes related to the acylation modification of terpenoids within the BAHD family. 【Results】 A comprehensive identification of 158 BAHD genes was performed in the *Camellia* genome. All of the *Camellia* BAHD genes are randomly distributed across its 15 chromosomes, and 14 pairs of duplicated genes have been generated during evolution. Phylogenetic analysis indicates that these *Camellia* BAHD genes can be divided into 6 branches. The protein motifs and gene structures of BAHD are highly conserved within each group, reflecting their functional conservation. Synteny analysis suggests that most BAHD genes likely originate from tandem duplications under purifying selection, providing new insights into potential functional differences among BAHD gene family members. Additionally, numerous hormone and stress response elements were found in the BAHD promoters. Using qRT-PCR to verify the core members of the BAHD gene involved in the modification of terpenoid compounds, it was found that the *Co10g1264.11*, *Co04g0659.15*, and *Co06g0230.08*

genes had higher expression levels in leaves, roots, and stems, while the *Co10g1264.11* gene had the highest expression level in fruits. **【Conclusion】** The tissue-specific expression of the BAHD family genes in different tissues of *C.oleifera* reflects their functional diversity. Understanding the expression patterns of the BAHD family genes helps in predicting their functions. qRT-PCR analysis revealed significant differences in the expression of certain BAHD genes in the roots, leaves, and seeds of *C.oleifera*. Among them, the genes *Co10g1264.11*, *Co04g0659.15*, and *Co06g0230.08* showed higher expression levels in the leaves, roots, and stems. The gene *Co10g1264.11* had the highest expression level in seeds, while other genes exhibited lower expression levels in seeds. These members can serve as core candidate genes for subsequent functional validation studies. This study identified key candidate enzyme genes for the acylation modification within the BAHD family of camellia saponin, clarified the characteristics of *C.oleifera* BAHD gene family, and laid a research foundation for further elucidation of the camellia saponin biosynthesis pathway.

Keywords: *Camellia oleifera*; acyltransferase; BAHD family; gene mining