

## 研究报告

### A Letter

## 玳玳花咖啡酸甲基转移酶基因的电子克隆及序列分析

张强<sup>✉</sup>, 张晓伟<sup>✉</sup>, 孟月娥<sup>✉</sup>, 李艳敏<sup>✉</sup>, 王慧娟<sup>✉</sup>, 王利民<sup>✉</sup>

河南省农业科学院园艺研究所, 郑州, 450002

✉ 通讯作者: qiangzh76@163.com ✉ 作者

分子植物育种, 2011年, 第9卷, 第14篇 doi: 10.5376/mpb.cn.2011.09.0014

收稿日期: 2010年08月11日

接受日期: 2010年10月21日

发表日期: 2011年02月22日

这是一篇采用 Creative Commons Attribution License 进行授权的开放取阅论文。只要对本原作有恰当的引用, 版权所有人允许并同意第三方无条件的使用与传播。

引用格式:

张强等, 2011, 玳玳花咖啡酸甲基转移酶基因的电子克隆及序列分析, 分子植物育种 Vol.9 No.14 (doi: 10.5376/mpb.cn.2011.09.0014)

**摘要** 许多植物的次生代谢化合物的合成涉及到一个或多个羟基由甲基转移酶(O-methyltransferases, OMTs)催化的甲基化反应。本研究根据植物 OMTs 的保守结构域设计了 1 对简并引物, 利用同源克隆的方法从香花植物玳玳花的花蕾中获得的一个 OMT 片段。随后, 通过电子克隆的方法对 NCBI 的 EST 数据库进行搜索和比对获得了其 cDNA 序列全长, 并经过 RT-PCR 验证。其推断的最大 ORF 框为 1 101 bp, 编码蛋白由 366 个氨基酸组成, 分子量为 39.9 kD, 等电点为 5.76。序列分析发现从第 35 个到第 86 个氨基酸为蛋白二聚化(Dimerisation)结构域, 在许多植物甲基转移酶基因的 N 端存在, 从第 104 个到第 341 个氨基酸为甲基转移酶结构域。同源比对和分子进化树分析发现, 所获得的玳玳花 OMT 基因与植物 COMTs (caffeic acid O-methyltransferases)具有较高的相似性而被命名为 CaCOMT (GenBank 登录号: HM641694)。最后, 对 CaCOMT 的功能进行了讨论, 其可能同时在木质素生物合成和花香物质代谢中起作用。

**关键词** 玳玳花; COMT; 电子克隆; 序列分析

## In Silicon Cloning and Sequence Analysis of Caffeic Acid O-methyltransferases from *Citrus Aurantium*

Zhang Qiang<sup>✉</sup>, Zhang Xiaowei<sup>✉</sup>, Meng Yue<sup>✉</sup>, Li Yanmin<sup>✉</sup>, Wang Huijuan<sup>✉</sup>, Wang Limin<sup>✉</sup>

Institute of Horticulture, Henan Academy of Agricultural Sciences, Zhengzhou, 450002, P.R. China

✉ Corresponding author, qiangzh76@163.com; ✉ Authors

**Abstract** The biosynthesis of many plant secondary compounds involves the methylation of one or more hydroxyl groups, catalyzed by O-methyltransferases (OMTs). We obtained a plant OMTs fragment from the flower buds of *Citrus aurantium* by employing a pair of degenerate primers designed based on the conservative regions of different plant OMTs. Subsequently, its whole cDNA sequence was silicon cloned by comparing and searching in ESTs library of NCBI and confirmed by RT-PCR. Its induced the largest ORF consisted of 1 101 bp encoding a protein of 366 amino acids with MW 39.9 kD and PI 5.76. Sequence analysis indicated that the amino acids residuals at the 35th to 86th form a protein dimerisation domain which was found in N terminal of many plant OMTs protein and the amino acids residuals at the 104th to 341th form a O-methyltransferases domain. Phylogenetic analysis revealed that it share high homology to plant caffeic acid O-methyltransferases (COMTs), named as CaCOMT (GenBank Accession Number: HM641694). At last, its probable function was discussed, which could play a role in lignin biosynthesis and floral scent compounds metabolism.

**Keywords** *Citrus aurantium*; COMT; In silicon cloning; Sequence analysis

### 研究背景

许多植物次生化合物的合成涉及到一个或多个羟基的甲基化过程(Schroder et al., 2002)。植物O-甲基转移酶(O-methyltransferases, OMTs)在次生代谢中具有重要作用, 能够催化S腺苷甲硫氨酸(S-adenosylmethionine, SAM)的甲基转移到许多受体分子的羟基, 参与一些次生代谢产物的合成。植

物的OMTs基于蛋白的氨基酸序列被分为两类(Joshi and Chiang, 1998)。I类OMTs中研究最多的是咖啡酰辅酶A(CoA)-O-甲基转移酶(caffeoyl coenzyme A OMTs, CCoAOMT), 它催化木质素合成前体咖啡酰CoA的甲基化生成阿魏酰CoA (Humphreys and Chapple, 2002)。II类OMTs, 咖啡酸甲基转移酶(caffeic acid O-methyltransferases, COMTs)由于被认为参与S木

质素的合成而受到广泛的研究。虽然名为COMT基因, 咖啡酸并不是COMT酶喜欢的作用底物, 它首选的作用底物是咖啡醛(caffeoyl aldehyde)和5-羟基松柏醛(5-hydroxyconiferaldehyde) (Osakabe et al., 1999; Li et al., 2000; Dixon et al., 2001; Parvathi et al., 2001); 另外一些II类OMTs则催化黄酮、类黄酮、苯丙烷类和多酚类化合物的甲基化。目前, 已经从很多物种中克隆了植物OMTs基因, 它们的分子分析、共同特性和分类已经被综述(Ibrahim et al., 1998)。

玳玳花(*Citrus aurantium*), 又名玳玳橘、回青橙、回春橙, 是芸香科柑桔属常绿灌木, 分布于我国南部各地, 多见于扬州、苏州。玳玳花春末夏初盛开, 其花香浓郁, 沁人心脾, 为我国南方常见的香花植物。花蕾含挥发油, 油中主要含柠檬烯(Limonene)、芳樟醇(Linalool)、牻牛儿醇(Geraniol)、香茅醇(Citronellol)、缬草酸(Valeric acid)等。迄今为止, 还未见到有玳玳花中OMTs基因被克隆的报道。随着分子生物学技术和基因组科学的发展, Genbank数据库中的表达序列标签(expressed sequence tags, ESTs)信息日渐丰富, 电子克隆(*in silico* cloning)正是基于数据库中大量已克隆测序的ESTs信息于近年来发展起来的一门快速克隆基因的新技术, 其技术核心是利用生物信息学技术组装延伸ESTs序列, 获得基因的部分乃至全长cDNA序列进一步利用RT-PCR的方法进行克隆分析、验证。在本研究中, 我们在利用同源克隆的方法从玳玳花的花瓣中获得了一个植物OMTs片段的基础上, 利用电子克隆的方法推断出了序列全长, 并经过RT-PCR和测序验证, 同源比对发现与COMT基因具有较高的相似性而被命名为CaCOMT (GenBank登录号: HM641694), 最后, 对它进行了序列分析和功能预测。

## 1 结果与分析

### 1.1 玳玳花COMT序列全长的扩增

以玳玳花花瓣的cDNA为模板, 用OMT5和OMT3的引物组合进行PCR扩增, 对250 bp左右的亮带切胶回收(图1A), 连接到pGEM-T载体, 测序, 获得玳玳花OMT序列片段。对NCBI的blastn结果表明, 获得的OMT序列片段, 与COMT基因序列有较高的同源性; 对NCBI的EST数据库进行电子克隆最终获得玳玳花COMT基因序列全长, 用全长引物COMTW5和COMTW3进行RT-PCR验证, 发现扩增得到的序

列大小在1 000 bp左右, 与预期大小基本相符(图1B)。测序结果表明, 玳玳花COMT基因的ORF框为1 101 bp, 与电子克隆推断的序列仅存在3个碱基的差别, 但推断的氨基酸编码序列完全相同, 命名为CaCOMT。

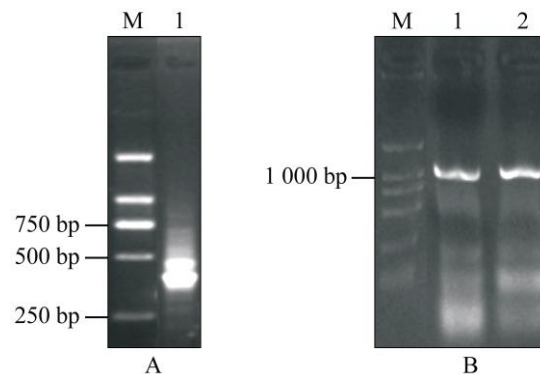


图1 玳玳花OMT基因的克隆

注: A: 玳玳花OMT基因片段的扩增; M: DL2000 Marker; 1: OMT基因片段; B: 玳玳花OMT基因全长的扩增; M: DL2000 Marker; 1, 2: OMT基因全长

Figure 1 The cloning of OMT gene from *Citrus aurantium*

Note: A: Amplification of OMT fragment from *Citrus aurantium*; M: DL2000 Marker; 1: OMT fragment; B: Amplification of the cDNA full length of OMT from *Citrus aurantium*; M: DL2000 Marker; 1, 2: the cDNA full length of OMT

### 1.2 CaCOMT序列分析

利用DNASTar软件的EditSeq程序对玳玳花CaCOMT氨基酸组成、分子质量、等电点分析发现, CaCOMT蛋白分子量为39 985.27 Daltons, 等电点为5.76, 由366个氨基酸组成, 其中, 正电荷氨基酸(K, R) 33个, 负电荷氨基酸(D, E) 41个, 疏水氨基酸(A, I, L, F, W, V) 138个, 极性氨基酸(N, C, Q, S, T, Y) 87个。在NCBI的CDD (Conserved Domain Database)数据库(网址为<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>)对CaCOMT蛋白序列进行结构特征分析发现(图2), 从35到86个氨基酸为蛋白二聚化(Dimerisation)结构域, 在许多植物甲基转移酶基因的N端存在, 从104到341个氨基酸为甲基转移酶结构域, 能够利用S腺苷甲硫氨酸作为甲基的供体。利用DNAMAN6.0和MEGA4.0软件对植物OMTs进行多重序列比对并利用NJ法构建系统进化树(图3), 结果表明, CaCOMT同CbCOMT、CbIEMT、NtOMT1、BpCOMT、EgCOMT、RcOMT2、MsCOMT和AtCOMT聚为一类, 与RcOMT2、MsCOMT和

```

1  ATGGGTTCAACCAGTTCAGAACTCAAATAAGTCCAGCCCAAGGCTCGGATGAAGAGGCAAACCTCTTGGCC
   M G S T S S E T Q I S P A Q G S D E E A N L L A M
76  ATGCAATTAACCAGTGCCTCAGTCTTGCCTATGGTTCTCAAATCAGCCATTGAGCTTGATCTTTTAGAGATCATA
   Q L T S A S V L P M V L K S A I E L D L L E I I A
151 GCTAAAGCTGGGCCCAGTGTTCATGTCTCCAAAAGACATAGCTTCTCAGCTGCCACAAAGAACCCAGATGCC
   K A G P D A F M S P K D I A S Q L P T K N P D A H
226 CATATCGTGTGATCGTATATTGCGCCTTCTGGCGAGCTATTCAGTCCTTAATTGCTCTTTGCGCAATCTCCCC
   I V L D R I L R L L A S Y S V L N C S L R N L P D
301 GACGGCAAAGTTGAGAGGCTTTATGGCCTTGCCCGTTTGTAAATCCTCACTAAAAATGAAGATGGTGTACAG
   G K V E R L Y G L A P V C K F L T K N E D G V T L
376 CTTTCCGATCTTTGTCTCATGAACCAAGACAAGGTTCTCATGGAGAGCTGGTACTACTTAAAAGATGCAGTGCTT
   S D L C L M N Q D K V L M E S W Y Y L K D A V L E
451 GAAGGTGGCATTCCATTTAACAAGGCCTATGGGATGAATGCATTGATTACCACGGCAAAGATCTAAGATTCAAC
   G G I P F N K A Y G M N A F D Y H G K D L R F N K
526 AAGATTTTCAACAATGGAATGTCTTCTATTCTACCATTACCATGAAGAAAATTCTTGAAAATTACAAAGGGTTT
   I F N N G M S S H S T I T M K K I L E N Y K G F E
601 GAAGGCCTCAACTCAGTTGTGACGTTGGTGGTGGAAATTGGAGCCACACTTAACATGATTATCTCCAAGTATCCA
   G L N S V V D V G G G I G A T L N M I I S K Y P S
676 TCGATTAAGGCATTAACCTTGTGATTTGCCACATGTTATTCCAGGATGCTCCAGCTTTTCTGGTGTGCGAGCATGT
   I K G I N F D L P H V I Q D A P A F P G V E H V G
751 GGGGGAGACATGTTTGTAGTGTCCAAAGGGAGATGCCATTTTTATCAAGTGGATATGTCATGATTGGAGTGAT
   G D M F V S V P K G D A I F I K W I C H D W S D E
826 GAGCACTGCGTGAAATTTCTGAAGAACTGCTATGAAGCACTCCAGTAAATGGGAAAATGTCATTGTTGCTGAATCT
   H C V K F L K N C Y E A L P V N G K V I V A E S I
901 ATCCTCCCAGTAACCCCGGACACAAGCCTTGCATCCAAAGTAGTCATCCATGTCGACTGCATCATGTTGGCTCAT
   L P V T P D T S L A S K V V I H V D C I M L A H N
976 AACCCGGGTGGCAAAGAGAGGACTGAACAAGAGTTCAGAGCATTGGCTAAGGCTGCTGGATTCCAAGTTTCCAA
   P G G K E R T E Q E F R A L A K A A G F Q G F Q V
1051 GTTGTGAGCTCTGCTTTTAATACTTACATTATGGAATTTCTCAAGAGTGCTTGA
   V S S A F N T Y I M E F L K S A *

```

图2 玳玳花*CaCOMT*基因序列及推断的氨基酸

注: 加框的氨基酸为蛋白二聚化结构域, 下划线部分为甲基转移酶结构域

Figure 2 Nucleotide and deduced amino acid sequences of cDNA full length of *CaCOMT* from *Citrus aurantium*

Note: Boxed region indicate protein dimerisation domain; Underlined sequences indicate O-methyltransferases domain

*AtCOMT*的亲缘关系最近, 氨基酸之间相似性分别为79.7%, 77.3%, 76.0%。多个*OMT*基因编码氨基酸序列多重比对结果表明(图4), 玳玳花*COMT*具有植物*OMTs*的5个保守结构域。

## 2讨论

早期研究认为*COMT*甲基化主要发生在羟基肉桂酸水平。*COMT*催化咖啡酸和5-羟基阿魏酸(5-hydroxyferulic acid)甲基化分别生成阿魏酸(ferulic acid)和芥子酸(sinapic acid), 芥子酸进一步参与S木质素的合成(Guo et al., 2001)。近几年研究表明,*COMT*催化的反应除了在肉桂酸水平进行外, 还可在醛和醇的水平上发生(Osakabe et al., 1999; Li et al., 2000; Dixon et al., 2001; Parvathi et al., 2001)。故现在认为*COMT*在木质素生物合成中催化咖啡酸、5-羟基松柏醛和5羟基松柏醇甲基化分别生成

阿魏酸、芥子醛和芥子醇; *CCoAOMT*催化咖啡酰CoA甲基化生成阿魏酰CoA。除此之外, *COMT*还参与了芳香化合物的合成, 月季*RcOMT2*与*COMT*氨基酸具有高度的相似性, 它除了可以利用咖啡酸作为底物外, 还对芳香物质1,3,5-三甲基苯(1,3,5-trimethoxybenzene)合成的3个前体都具有一定程度的活性(Wu et al., 2003)。花香物质香草素(vanillin)合成的最后一步酶促反应被认为是3,4-二羟基苯甲醛的甲基化反应(Pak et al., 2004)。烟草的I类*COMT*对底物咖啡酸和3,4-二羟基苯甲醛的活性效率相当, 而II类*COMT*对咖啡酸没有活性, 只对3,4-二羟基苯甲醛具有活性(Maury et al., 1999)。3,4-二羟基苯甲醛虽然不是香草兰(*Vanilla planifolia*)*COMT*的首选底物, 但也能够催化3,4-二羟基苯甲醛的甲基化反应(Pak et al., 2004)。又如, 仙女扇

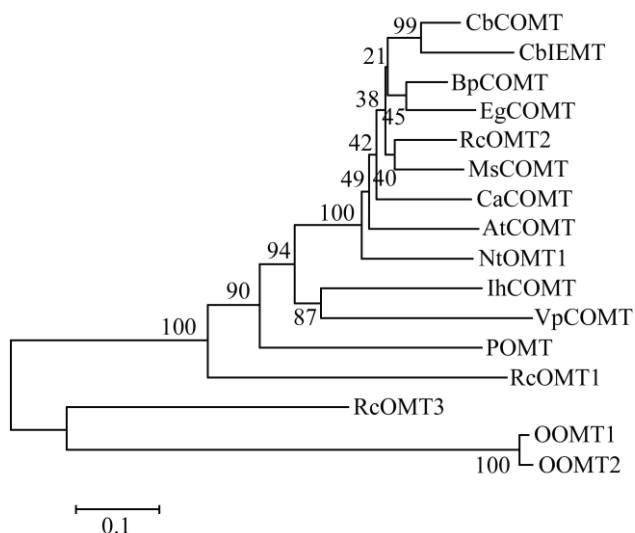


图3 根据植物OMTs蛋白构建的系统进化树

注: CaCOMT: 玳玳花; AtCOMT: 拟南芥; BpCOMT: 欧洲白桦; EgCOMT: 桉树; RcOMT1: 月季; RcOMT2: 月季; RcOMT3: 月季; OOMT1: 玫瑰; OOMT2: 玫瑰; POMT: 月季; MsCOMT: 苜蓿; IhCOMT: 鸢尾; VpCOMT: 香草兰; CbCOMT: 仙女扇; CbIEMT: 仙女扇; NtOMT1: 烟草

Figure 3 Molecular phylogenetic trees based on the deduced amino acid sequences of plant OMTs

note: CaCOMT: *Citrus aurantium*; AtCOMT: *Arabidopsis thaliana*, ATU70424; BpCOMT: *Betula pendula*, FJ667539; EgCOMT: *Eucalyptus gunnii*, AF168777; RcOMT1: *Rosa chinensis*, AB086103; RcOMT2: *Rosa chinensis*, AB086104; RcOMT3: *Rosa chinensis*, AB086105; OOMT1: *Rosa hybrida*, AF502433; OOMT2: *Rosa hybrida*, AF502434; POMT: *Rosa chinensis*, AB121046; MsCOMT: *Medicago sativa*, M63853; IhCOMT: *Iris hollandica*, AB183825; VpCOMT: *Vanilla planifolia*, AY555144; CbCOMT: *Clarkia breweri*, AF006009; CbIEMT: *Clarkia breweri*, U86760; NtOMT1: *Nicotiana tabacum*, X74452

(*Clarkia breweri*) COMT和IEMT ((iso) eugenol O-methyltransferase)的氨基酸相似性达到了83%, 但两者具有完全不同的底物选择特性, IEMT使丁香酚和异丁香酚的甲基化生成芳香物质甲基丁香酚(methyleugenol)和异甲基丁香酚(isomethyleugenol) (Wang et al., 1997; Wang and Pichersky, 1998)。定点突变分析(Wang and Pichersky, 1999)和分子模型研究(Zubieta et al., 2002)揭示仙女扇COMT和IEMT几个底物绑定残基的不同决定了底物选择特性的不同。通过序列分析和同源比对, 由于玳玳花COMT与上述基因间都有较高的同源性, 因此, 我们推测玳玳花COMT可能

同时在木质素生物合成和花香物质代谢中起作用。

### 3 材料和方法

#### 3.1 植物材料

玳玳花花瓣, 取自苏州。

#### 3.2 总RNA分离及cDNA第一链和第二链的合成

开花期取其未开放的花蕾。

RNA的分离按照北京天根生化科技有限公司RNAplant植物总RNA提取试剂的说明书进行。RNA沉淀室温干燥5~10 min, 用DEPC水溶解, -75°C保存备用。cDNA第一链和第二链的合成参照SMART<sup>TM</sup> PCR cDNA Synthesis Kit User Manual方法进行。

#### 3.3 玳玳花OMT序列片段的扩增

根据已知植物的OMTs基因信息, 设计简并引物OMT5; 5'-ATGT(A/C/T)GG(A/T/C/G)GG(A/T/C/G)GACATGTTT-3'和OMT3: 5'-GT(C/T)CTCTC(T/C)TTNCCHCC-3', 引物由上海生工生物工程技术有限公司合成。以玳玳花花瓣的cDNA为模板, 用简并引物OMT5和OMT3进行甲基转移酶的PCR扩增。PCR反应体系: cDNA模板0.5 μL, 10×PCR缓冲溶液5 μL, 25 mmol/L的Mg<sup>2+</sup> 4 μL, 特异引物(20 μmol/L) 1 μL, 锚定引物(20 μmol/L) 1 μL, 10 mmol/L dNTPs 1 μL, Taq酶(5 U/μL) 0.5 μL, 加水到总体积50 μL。PCR程序: 94°C预变性3 min; 94°C变性30s, 50°C退火30s, 72°C延伸1 min, 30个循环; 72°C延伸7 min, 4°C保存。PCR产物用AXYGEN公司的DNA凝胶回收试剂盒回收, 回收的产物插入pGEM-T easy载体, 转化大肠杆菌, 菌落PCR筛选阳性克隆。碱裂解法抽提质粒, 用EcoR I进行单酶切鉴定和PCR鉴定后, 重组质粒送上海博尚生物公司测序。

#### 3.4 玳玳花COMT序列全长的电子克隆及RT-PCR验证

在同源克隆获得玳玳花OMT序列片段的基础上, 利用blastn搜索NCBI的EST数据库, 下载检索到的高相似性的EST序列, 下载的EST进行第一次拼接, 得到一个叠连群(contig); 将此contig在NCBI的EST数据库上重新搜索匹配序列, 并下载匹配序列; 再将匹配序列进行拼接, 重复搜索、拼接, 直到不能再延伸为止, 最后获得1个玳玳花COMT基因的cDNA序列。设计引物COMTW5: 5'-ACAAAA

CbCOMT : MGSTGNAETQLTPT-----HVSDEEANLFAMQLASASVLEMLVLAETLELDVLEETMAKSI PHGSGAYTSPABTAAQLPPTNEDAP-VMLDRVLRLLASYS : 93  
 CbIEMT : MGSTGNAEIQIIPPT-----HSSDEEANLFAMQLASAAVLEMLVLAETLELDVLEETMAKSVVPP--SGYISPBATAAQLPPTNEDAP-VMLDRVLRLLASYS : 91  
 RcOMT2 : MGSTG--ETQMTPT-----QVSDEEANLFAMQLASASVLEMLVLAETLELDVLEETMAKAGPG--AFLSPENDTASOLPPTNEDAP-VMLDRMLRLLASYS : 88  
 MsCOMT : MGSTG--ETQITPT-----HISDEEANLFAMQLASASVLEMLVLAETLELDVLEETMAKAGPG--AQISPIETASOLPPTNEDAP-VMLDRMLRLLACYI : 88  
 BpCOMT : MGSTG--ETQMTPT-----QVSDEEANLFAMQLASASVLEMLVLAETLELDVLEETMAKAGPG--AYLSPBETASOLPPTNEDAP-VMLDRMLRLLASYS : 88  
 EgCOMT : MGSTG--ETQMTPT-----QVSDEEANLFAMQLASASVLEMLVLAETLELDVLEETMAKAGPG--AFLSPGVEAAQLPPTNEDAP-VMLDRMLRLLASYS : 89  
 CaCOMT : MGSTSS--ETQISPA-----QGSDEEANLFAMQLASASVLEMLVLAETLELDVLEETMAKAGPD--AFMSPKDASOLPPTNEDAP-VMLDRMLRLLASYS : 89  
 AtCOMT : MGSTA--ETQLTPTV-----QVTDEEALFAMQLASASVLEMLVLAETLELDVLEETMAKNGSP--MSPTBASKLPPTNEDAP-VMLDRMLRLLSAYS : 86  
 NtOMT1 : MGSTS--ESQSNL-----THTDEEALFAMQLASASVLEMLVLAETLELDVLEETMAKAGPG--AAISPSBAAQLPPTNEDAP-VMLDRMLRLLASYS : 88  
 IhCOMT : -MDSTPVIQIPPTFY-----PFSDEEALYAMQLRSFVLEMLVLAETLELDVLEETMAKAGPG--AYLSPBETASOLPPTNEDAP-VMLDRMLRLLASYS : 89  
 VpCOMT : MMTWVEHQQQNGS-----KDVDEEACMYAMQLRSFVLEMLVLAETLELDVLEETMAKAGPD--SYLAEEDAAARLGNSSNEDAP-VMLDRMLRLLSAYS : 94  
 POMT : MRSTSPQTLTDFPMSPH--QOESEELGKQATRLANVVLEMLVLAETLELDVLEETMAKAGDG--ESLSPSDTAAQLPPTNEDAP-VMLDRMLRLLASYS : 90  
 RcOMT1 : MASPLEIEIFGPCSIPHVDAKRKEEBSSEPHAVQLMLSSVLEMLVLAETLELDVLEETMAKAGTD--AKLSALDAAKIGKKNHAP-VMLDRMLRLLTAHS : 96  
 OOMT1 : -MERLNSFRHLNQKWSNGEHSNELLHAQAHIWNHIFSEFINSMSLSNSATLGGIPDITIN-KHG----FMTLSBETASALPIHPTKS--HSVYRLMRLLVHSG : 92  
 OOMT2 : -MERLNSFKHLNQKWSNGEHSNELLHAQAHIWNHIFSEFINSMSLSNSATLGGIPDITIN-KHG----FMTLSBETASALPIHPTKS--HSVYRLMRLLVHSG : 91  
 RcOMT3 : -MEAVT-----ELDEASLRGQADVWKYMLRFADSMALNSAVLEIRIPDITIN-SHDH--ALTLSBETCSFDSASLSPDITCLSRIMRLLVRRN : 82

CbCOMT : VVTCSLRELP-DGKVERLRLGLAPVCKPFTKNEDE--VSLAPFLCMMNOQKVLMSWYHLKDAVLDGG-IFFNKAYGNTAREYHGTGIFRNKVFNRMSDHS : 189  
 CbIEMT : VVYTLRELP-DGKVERLRLGLAPVCKPFTKNEDE--VSLAPFLCMMNOQKVLMSWYHLKDAVLDGG-IFFNKAYGNTAREYHGTGIFRNKVFNRMSDHS : 187  
 RcOMT2 : ILYYSRLTLP-DGKVERLRLGLAPVCKPFTKNEDE--VSLAPFLCMMNOQKVLMSWYHLKDAVLDGG-IFFNKAYGNTAREYHGTGIFRNKVFNRMSDHS : 184  
 MsCOMT : ILLCSVRTQP-DGKVERLRLGLATVAKYLVKNEDE--VSIISANNMNOQKVLMSWYHLKDAVLDGG-IFFNKAYGNTAREYHGTGIFRNKVFNRMSDHS : 184  
 BpCOMT : VLYTSLRLLP-DGKVERLRLGLAPVCKPFTKNEDE--VSIISANNMNOQKVLMSWYHLKDAVLDGG-IFFNKAYGNTAREYHGTGIFRNKVFNRMSDHS : 184  
 EgCOMT : VLTCTLRNLP-DGKVERLRLGLAPVCKPFTKNEDE--VSIISANNMNOQKVLMSWYHLKDAVLDGG-IFFNKAYGNTAREYHGTGIFRNKVFNRMSDHS : 185  
 CaCOMT : VINCSSLRNLDP-DGKVERLRLGLAPVCKPFTKNEDE--VSLSDPCMMNOQKVLMSWYHLKDAVLDGG-IFFNKAYGNTAREYHGTGIFRNKVFNRMSDHS : 185  
 AtCOMT : VILYCSNRKLS-DGKVERLRLGLAPVCKPFTKNEDE--VSIISANNMNOQKVLMSWYHLKDAVLDGG-IFFNKAYGNTAREYHGTGIFRNKVFNRMSDHS : 182  
 NtOMT1 : VLNCTLRLLP-DSVSRRLRLGLAPVCKPFTKNEDE--VSVAPFLMNOQKVLMSWYHLKDAVLDGG-IFFNKAYGNTAREYHGTGIFRNKVFNRMSDHS : 184  
 IhCOMT : VVSCRVDAAAGDGKPRRRNGAERIVKYLTKNEDE--VSMMAATMNOQKVLMSWYHLKDAVLDGG-IFFNKAYGNTAREYHGTGIFRNKVFNRMSDHS : 186  
 VpCOMT : IILNF-TDVTVDGEGRTVRSYGAHVCKYLTPTQDC--VSMAPFLMNTQKVLMSWYHLKDAVLDGG-IFFNKAYGNTAREYHGTGIFRNKVFNRMSDHS : 186  
 POMT : ILLCSARTGS-DGQVERLRLSAGPICKPFTKNEDE--VSLSDPCMMNOQKVLMSWYHLKDAVLDGG-IFFNKAYGNTAREYHGTGIFRNKVFNRMSDHS : 192  
 RcOMT1 : ILLCSAVVTGQ-----RLNLSYAVSHVFTSEDE--ASLSYVMSASIOAVVMSWYHLKDAVLDGG-IFFNKAYGNTAREYHGTGIFRNKVFNRMSDHS : 187  
 OOMT1 : FFAKKKLS-KTDEE--GATLTASQLLKLHPL--SLTPLYLTAMLEPLVLTENNYLSTWFWQNDDEPFDPTAHGTFEDYGNHQPSTAHLENDAMASDA : 185  
 OOMT2 : FFAKKKLS-KTDEE--GATLTASQLLKLHPL--SLTPLYLTAMLEPLVLTENNYLSTWFWQNDDEPFDPTAHGTFEDYGNHQPSTAHLENDAMASDA : 184  
 RcOMT3 : FETAQPDGGDSGET--LGGTTHSSRWLHSEF--FLAPVIMETLSTLMSFWHCFSCQYKQETPCABEKRAHGRDITQFSEKPEFNRLENDEMSCTA : 176

CbCOMT : TITMKKILEMNT--FEGLETVIVDVGGGTCAVNMHIVSYRPSIKGINFDLPHVTEDAEYTPGVVHVGGMDFVS--VEKGDALFMKNI CHDWSDEHCLRLK : 286  
 CbIEMT : TITMKKILEMNT--FEGLETVIVDVGGGTCAVNMHIVSYRPSIKGINFDLPHVTEDAEYTPGVVHVGGMDFVS--VEKGDALFMKNI CHDWSDEHCLRLK : 284  
 RcOMT2 : TITMKKILETYK--FEGLETVIVDVGGGTCAVNMHIVSYRPSIKGINFDLPHVTEDAEYTPGVVHVGGMDFVS--VEKGDALFMKNI CHDWSDEHCLRLK : 281  
 MsCOMT : TITMKKILETYK--FEGLETVIVDVGGGTCAVNMHIVSYRPSIKGINFDLPHVTEDAEYTPGVVHVGGMDFVS--VEKGDALFMKNI CHDWSDEHCLRLK : 281  
 BpCOMT : TITMKKILETYK--FEGLETVIVDVGGGTCAVNMHIVSYRPSIKGINFDLPHVTEDAEYTPGVVHVGGMDFVS--VEKGDALFMKNI CHDWSDEHCLRLK : 281  
 EgCOMT : TITMKKILETYK--FEGLETVIVDVGGGTCAVNMHIVSYRPSIKGINFDLPHVTEDAEYTPGVVHVGGMDFVS--VEKGDALFMKNI CHDWSDEHCLRLK : 282  
 CaCOMT : TITMKKILETYK--FEGLETVIVDVGGGTCAVNMHIVSYRPSIKGINFDLPHVTEDAEYTPGVVHVGGMDFVS--VEKGDALFMKNI CHDWSDEHCLRLK : 282  
 AtCOMT : TITMKKILETYK--FEGLETVIVDVGGGTCAVNMHIVSYRPSIKGINFDLPHVTEDAEYTPGVVHVGGMDFVS--VEKGDALFMKNI CHDWSDEHCLRLK : 279  
 NtOMT1 : TITMKKILETYK--FEGLETVIVDVGGGTCAVNMHIVSYRPSIKGINFDLPHVTEDAEYTPGVVHVGGMDFVS--VEKGDALFMKNI CHDWSDEHCLRLK : 281  
 IhCOMT : AFTMKKILETYK--FEGLETVIVDVGGGTCAVNMHIVSYRPSIKGINFDLPHVTEDAEYTPGVVHVGGMDFVS--VEKGDALFMKNI CHDWSDEHCLRLK : 283  
 VpCOMT : IITMKKILETYK--FEGLETVIVDVGGGTCAVNMHIVSYRPSIKGINFDLPHVTEDAEYTPGVVHVGGMDFVS--VEKGDALFMKNI CHDWSDEHCLRLK : 283  
 POMT : TITMKKILETYK--FEGLETVIVDVGGGTCAVNMHIVSYRPSIKGINFDLPHVTEDAEYTPGVVHVGGMDFVS--VEKGDALFMKNI CHDWSDEHCLRLK : 289  
 RcOMT1 : TITMKKILETYK--FEGLETVIVDVGGGTCAVNMHIVSYRPSIKGINFDLPHVTEDAEYTPGVVHVGGMDFVS--VEKGDALFMKNI CHDWSDEHCLRLK : 284  
 OOMT1 : RLVTSVIINDCKVDFEGLESLVDVGGGTCTLAKAHADAFPHIECTVLDLPHVADLQGSKNLKYTGDMFEA--VEPADTVLKRILHDDWSDEHCLRLK : 283  
 OOMT2 : RLVTSVIINDCKVDFEGLESLVDVGGGTCTLAKAHADAFPHIECTVLDLPHVADLQGSKNLKYTGDMFEA--VEPADTVLKRILHDDWSDEHCLRLK : 282  
 RcOMT3 : KTRRRAMLGKKGDFDNVATLVDVGGGTCSAVABIVKLYPSIKGINFDLPHVVAIAEYHGVSHVGGDMFDGNTENTDAIEMKNIHDDWSDEHCLRLK : 276

I II III

CbCOMT : NQYAALE---EHGKVI VASCIIPLSLDPSSLATKGVVHIDVIMLAHNPGGKERTKEKEPEALAKGAGGQGFVACCAFNTYVMEFLKTA : 370  
 CbIEMT : NQYAALE---DHGKVI VASCIIPLSLDPSSLATKGVVHIDVIMLAHNPGGKERTKEKEPEALAKGAGGQGFVACCAFNTYVMEFLKTA : 368  
 RcOMT2 : NQYAALE---DNGKVI VASCIIPVADTSLATKGVVHIDVIMLAHNPGGKERTQGEPEALAKGAGGQGFVACCAFNTYVMEFLKTA : 365  
 MsCOMT : NQYAALE---DNGKVI VASCIIPVADTSLATKGVVHIDVIMLAHNPGGKERTQGEPEALAKGAGGQGFVACCAFNTYVMEFLKTA : 365  
 BpCOMT : NQYDALE---NNGKVI VASCIIPVADTSLATKGVVHIDVIMLAHNPGGKERTKEKEPEALAKGAGGQGFVACCAFNTYVMEFLKTA : 365  
 EgCOMT : NQYDALE---NNGKVI VASCIIPVADTSLATKGVVHIDVIMLAHNPGGKERTQGEPEALAKGAGGQGFVACCAFNTYVMEFLKTA : 366  
 CaCOMT : NQYDALE---VNGKVI VASCIIPVADTSLATKGVVHIDVIMLAHNPGGKERTQGEPEALAKGAGGQGFVACCAFNTYVMEFLKTA : 366  
 AtCOMT : NQYDALE---EDGKVI VASCIIPVADTSLATKGVVHIDVIMLAHNPGGKERTKEKEPEALAKGAGGQGFVACCAFNTYVMEFLKTA : 363  
 NtOMT1 : NQYDALE---ANGKVI VASCIIPVADTSLATKGVVHIDVIMLAHNPGGKERTKEKEPEALAKGAGGQGFVACCAFNTYVMEFLKTA : 364  
 IhCOMT : NCKKALE---EDGKVI VASCIIPVADTSLATKGVVHIDVIMLAHNPGGKERTKAEPEALAKGAGGQGFVACCAFNTYVMEFLKTA : 365  
 VpCOMT : NCAKSLA---DKGKVI VASCIIPVADTSLATKGVVHIDVIMLAHNPGGKERTKAEPEALAKGAGGQGFVACCAFNTYVMEFLKTA : 365  
 POMT : NCKKSLA---SSGKVI VASCIIPVADTSLATKGVVHIDVIMLAHNPGGKERTKAEPEALAKGAGGQGFVACCAFNTYVMEFLKTA : 371  
 RcOMT1 : NNYTATL---DNGKVI VASCIIPVADTSLATKGVVHIDVIMLAHNPGGKERTQGEPEALAKGAGGQGFVACCAFNTYVMEFLKTA : 366  
 OOMT1 : RSRVATIGKKGKVIIDMMENKQGDSEIETQDFDMLMMLALVGG--KERNEKEMAKLFTDAGSDYKITPILGLRSLIEVYV-- : 367  
 OOMT2 : RSRVATIGKKGKVIIDMMENKQGDSEIETQDFDMLMMLALVGG--KERNEKEMAKLFTDAGSDYKITPILGLRSLIEVYV-- : 366  
 RcOMT3 : NCRKAI---ERSGKVI VASCIIPVADTSLATKGVVHIDVIMLAHNPGGKERTKEKEPEALAKGAGGQGFVACCAFNTYVMEFLKTA : 359

IV V

图4 不同植物OMTs基因编码氨基酸序列多重比对结果  
 注: 下划线部分为OMT基因编码氨基酸序列保守区; I: LVDGGGxG; II: GINFDLPHV; III: EHVGGDMF; IV: NGKVI; V: GKGERT  
 Figure 4 Multiple alignment result of different plant O-methyltransferase  
 Note: The underlined parts indicate conservative regions of different plant OMTs encoded protein; I: LVDGGGxG; II: GINFDLPHV; III: EHVGGDMF; IV: NGKVI; V: GKGERT



TGGGTTCAACCAGTTC-3'和COMTW3: 5'-TCCA GTAGAACAAAAGACTCAAGC-3', 以玳玳花花瓣的cDNA为模板, 进行RT-PCR验证。获得的PCR产物用AXYGEN公司的DNA凝胶回收试剂盒回收, 克隆至pGEM-T easy载体, 转化大肠杆菌, 菌落PCR筛选阳性克隆送上海博尚生物公司测序。

### 3.5 序列分析

用DNASTar软件EditSeq程序对玳玳花CaCOMT进行氨基酸组成、分子质量、等电点分析, 并通过NCBI的CDD (Conserved Domain Database)数据库(网址为<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>)预测CaCOMT蛋白的保守结构域; 利用DNAMAN6.0和MEGA4.0软件对CaCOMT进行序列分析并与数据库中植物OMTs进行同源序列比对、聚类分析, NJ法构建系统进化树。

### 作者贡献

张强是本研究的实验设计和实验研究的执行人, 并和张晓伟完成数据分析, 论文初稿的写作; 李艳敏和王慧娟及王利民参与实验设计, 试验结果分析; 孟月娥是项目的构思者及负责人, 指导实验设计, 数据分析, 论文写作与修改。全体作者都阅读并同意最终的文本。

### 致谢

本研究由河南省基础与前沿技术研究计划项目(082300430100)资助。

### 参考文献

- Dixon R.A., Chen F., Guo D., and Parvathi K., 2001, The biosynthesis of monolignols: a “metabolic grid”, or independent pathways to guaiacyl and syringyl units, *Phytochemistry*, 57: 1069-1084
- Guo D.J., Chen F., Inoue K., Blount J.W., and Dixon R.A., 2001, Downregulation of caffeic acid 3-O-methyltransferase and caffeoyl CoA 3-O-methyltransferase in transgenic alfalfa: impacts on lignin structure and implications for the biosynthesis of G and S lignin, *Plant Cell*, 13(1): 73-88
- Humphreys J.M. and Chapple C., 2002, Rewriting the lignin roadmap, *Current Opinion in Plant Biology*, 5: 224-229
- Ibrahim R.K., Bruneau A., and Bantignies B., 1998, Plant O-methyltransferases: molecular analysis, common signature and classification, *Plant Mol. Biol.*, 36: 1-10
- Joshi C.P., and Chiang V.L., 1998, Conserved sequence motifs in

- plant S-adenosyl-L-methionine-dependent methyltransferases, *Plant Mol. Biol.*, 37: 663-674
- Li L., Popko J.L., Umezawa T., and Chiang V.L., 2000, 5-Hydroxyconiferyl aldehyde modulates enzymatic methylation for syringyl monolignol formation, a new view of monolignol biosynthesis in angiosperms, *J. Biol. Chem.*, 275: 6537-6545
- Osakabe K., Tsao C.C., Li L., Popko J.L., Umezawa T., and Carraway D.T., 1999, Coniferyl aldehyde 5-hydroxylation and methylation direct syringyl lignin biosynthesis in angiosperms, *Proc. Natl. Acad. Sci. USA*, 96: 8955-8960
- Pak F.E., Gropper S., Dai W.D., Havkin-Frenkel D., and Belanger F.C., 2004, Characterization of a multifunctional methyltransferase from the orchid *Vanilla planifolia*, *Plant Cell Rep.*, 22: 959-966
- Parvathi K., Chen F., Guo D., Blount J.W., and Dixon R.A., 2001, Substrate preferences of O-methyltransferases in alfalfa suggest new pathways for 3-O-methylation of monolignols, *Plant J.*, 25: 193-202
- Schroder G., Wehinger E. and Schroder J., 2002, Predicting the substrates of cloned plant O-methyltransferases, *Phytochemistry*, 59: 1-8
- Wang J., and Pichersky E., 1999, Identification of specific residues involved in substrate discrimination in two plant O-methyltransferases, *Arch. Biochem. Biophys.*, 368: 172-180
- Wang J., and Pichersky E., 1998, Characterization of S-adenosyl-L-methionine: (iso) eugenol O-methyltransferase involved in floral scent production in *Clarkia breweri*, *Arch. Biochem. Biophys.*, 349: 153-160
- Wang J., Dudareva N., Bhakta S., Raguso R.A., and Pichersky E., 1997, Floral scent production in *Clarkia breweri* (Onagraceae) II. Localization and developmental modulation of the enzyme S-adenosyl-L-methionine: (iso) eugenol O-methyltransferase and phenylpropanoid emission, *Plant Physiol.*, 114:213-221
- Wu S., Watanabe N., Mita S., Ueda Y., Shibuya M., and Ebizuka Y., 2003, Two O-methyltransferases isolated from flower petals of *Rosa chinensis* var. *spontanea* involved in scent biosynthesis, *J. Biosci. Bioeng.*, 96: 119-128
- Zubieta C., Kota P., Ferrer J-L., Dixon R.A., and Noel J.P., 2002, Structural basis for the modulation of lignin monomer methylation by caffeic acid/5-hydroxyferulic acid 3/5-O-methyltransferase, *Plant Cell*, 14: 1265-1277