

光皮桦中 4-香豆酸辅酶 A 连接酶基因 *BI4CL* 的克隆和表达分析

程龙军*, 童再康, 黄华宏, 楼雄珍

浙江林学院林业与生物技术学院, 浙江临安 311300

摘要: 以从光皮桦茎叶组织提取的 mRNA 为模板, 根据其他已克隆到的阔叶类树种中 4-香豆酸辅酶 A 连接酶(4CL)基因的同源序列设计兼并引物, 进行 RT-PCR 扩增, 获得部分基因片段, 然后结合 5', 3' RACE 方法从光皮桦中扩增出 1 个 4CL 基因的全长 cDNA 序列, 命名为 *BI4CL*。该基因 cDNA 全长为 1 983 bp (GenBank 登录号 FJ410448), 具有完整的开放阅读框架 (69~1 697 bp), 编码蛋白为 542 个氨基酸, 包含一个 AMP 结合功能域和一个含有 12 个氨基酸的功能基序。与其他植物中的 4CL 进行同源性比对的结果显示, *BI4CL* 蛋白与东北白桦的同源性最高, 达到了 98%。该基因在光皮桦的根和茎中表达量较高, 而在花和叶中的表达量低。

关键词: 光皮桦; 4-香豆酸辅酶 A 连接酶; 基因克隆; 表达

Cloning and Expression Analysis of 4-coumarate: CoA Ligase Gene *BI4CL* in *Betula luminifera* H. Winkl.

CHENG Long-Jun*, TONG Zai-Kang, HUANG Hua-Hong, LOU Xiong-Zhen

School of Forestry and Biotechnology, Zhejiang Forestry College, Lin'an, Zhejiang 311300, China

Abstract: A 4-coumarate: CoA ligase gene (*BI4CL*) was amplified by RT-PCR and 5', 3' RACE methods, when mRNA from stem and leaf of *Betula luminifera* were used as a template, and the primers were designed according to the homologous sequences of 4CL from broadleaf species. Sequence analysis showed that *BI4CL* cDNA was 1 983 bp and contained a single open reading frame (69–1 697 bp) encoding a protein of 542 amino acids (GenBank No. FJ410448). The protein had a motif of 12 amino acids and a functional AMP-binding domain. The results of homology comparison showed that *BI4CL* protein shared a 98% homology with *Betula platyphylla*. And the expression of *BI4CL* gene were high in roots and stems of *Betula luminifera*, and low in leaves and flowers.

Key words: *Betula luminifera*; 4-coumarate: CoA ligase; gene clone; expression

木质素合成过程中, 4-香豆酸辅酶 A 连接酶(4-coumarate: CoA ligase, 4CL)是一个关键酶。它通过苯丙烷途径在木质素合成早期产生单体木质素的前体 ρ - 香豆酰辅酶 A、咖啡酰辅酶 A、阿魏酰辅酶 A 和芥子酰辅酶 A 等, 进而促进各种木质素单体的合成。同时, 4CL 催化形成的 ρ - 香豆酰辅酶 A 参与植物次生代谢物黄酮、二苯乙烯类等前体的合成。鉴于 4CL 基因在植物木质素合成和次生代谢中都起作用, 其在杨树、松树以及各种模式植物中都已有了一定程度的研究(Whetten 和 Sederoff 1995; Wagner 等 2009)。在烟草中, 4CL 基因的沉默大幅降低木质纤维主要构成成分的 S 型木质素的含量(Kajita 等 1997); 而在拟南芥中, 沉默 4CL 则主要导致构成微管组分的 G 型木质素含量的降低(Lee 等 1997)。同样, 4CL 基因的超表达能够提高植物中木质素的含量, 在由于挤压、弯曲而形成的具高

木质素、低纤维素含量, 称为“压缩木”的组织中, 4CL 的表达量大大上调, 说明木质素含量的增加与 4CL 基因的表达水平密切相关(Zhang 和 Chiang 1997)。

光皮桦又名亮叶桦, 作为我国的一种材质优良、在南方广泛分布的珍贵用材树种, 对其材质形成相关基因的研究很有必要, 因此克隆和分析光皮桦中木质素合成途径中的基因 4CL 对研究其材质形成过程中的某些机制及其材质改良可能有一些参考价值。本文用其他植物中已克隆的 4CL 全长 cDNA 序列, 通过序列同源比对, 根据保守性序列设计兼并引物, 对从光皮桦茎叶组织中提取的

收稿 2009-09-21 修定 2009-12-22

资助 浙江省重大科技专项(2008C02004-1)。

* 通讯作者(E-mail: ljcheng@zju.edu.cn; Tel: 0571-63742885)。

mRNA 进行 RT-PCR, 获得基因序列片段后, 再用 RACE-PCR 技术得到 *4CL* 的 cDNA 全长序列, 并对不同物种之间该基因编码蛋白进行了同源性与进化分析。此外, 对该基因在光皮桦不同器官中的表达情况也作了初步研究。

材料与方 法

光皮桦(*Betula luminifera* H. Winkl.)材料取自本学院光皮桦良种繁育中心苗圃, 取三年树龄的光皮桦茎叶作为 mRNA 提取材料。

克隆用大肠杆菌 *Escherichia coli* 5H α 为本实验室保存; T4 DNA 连接酶、Taq DNA 聚合酶、克隆载体 pMD-T 为 TaKaRa 公司产品; RT-PCR 试剂盒为 Invitrogen 公司产品; 引物由南京金斯特生物工程技术有限公司合成; RACE 试剂盒为 Clone-Tech 产品。

光皮桦茎叶组织中 mRNA 提取采用改进的 CTAB 法(候夫云等 2008), 取 150 mg 幼嫩光皮桦茎叶组织, 于液氮中研磨后用药匙将粉末加入到 65 °C 水浴的 700 mL CTAB (2%) 提取液中, 摇动混匀。然后 65 °C 水浴 20 min 后冷却至室温。加等体积 (700 mL) 的氯仿 / 异戊醇 (24:1), 混匀, 4 °C 下以 10 000 \times g 离心 10 min。小心吸取上清液, 重复上述步骤, 直至中间层无白色絮状沉淀。吸取上清液, 加 1/4 体积的 10 mol·L⁻¹ LiCl, 混合后于 4 °C 沉淀过夜, 然后以 10 000 \times g 离心 20 min。弃上清液, 用 500 μ L RNA 溶解液 SSTE 溶解沉淀, 加等体积氯仿 / 异戊醇混匀 (24:1), 4 °C 下以 10 000 \times g 离心 10 min。离心吸取上清液, 加 2 倍体积的无水乙醇, 于 -70 °C 沉淀至少 30 min, 或 -20 °C 下沉淀 2 h。4 °C 下以 12 000 \times g 离心 20 min。去上清, 用 75% 的乙醇洗 2 次, 于 4 °C 下以 12 000 \times g 离心 5 min, 小心弃去上清液并吹干沉淀。用 40 μ L DEPC 处理的水溶解沉淀, 得到 RNA 样品。

进行引物合成时, 首先从 GenBank 数据库中获得 3 种阔叶树 *4CL* 基因的编码序列: 欧洲山杨 (*Populus tremula* L.), 登录号 AF283552; 东北白桦 (*Betula platyphylla* Suk.), 登录号 AY792353; 桉树 (*Eucalyptus camaldulensis* Dehnh.), 登录号 DQ147001。用 ClawstW 软件做多序列比对, 根据序列保守性, 设计基因保守区的 RT-PCR 引物 4CL-F 和 4CL-R, 序

列为: 4CL-F: 5' AAAGGDGTSATGYTVACDCAC 3'; 4CL-R: 5' CCAHCCWTCTTRTCKATBGT 3'。

cDNA 的合成按反转录合成及 PCR 扩增均参照试剂盒说明书步骤进行。获得的片段以 T 载体法克隆 PCR 产物, 阳性克隆送南京金斯特生物工程技术有限公司测序。测序结果经 BLAST 检索序列数据库, 确定所克隆序列为光皮桦 *4CL* 基因编码序列。

根据 RT-PCR 片段的序列测定结果, 合成 4CL 和 4CL 特异引物 GSP1 和 GSP2, 分别用于克隆 cDNA 的 5' 和 3' 端的序列, 序列为: 4CL-GSP1: 5' GCTCATCGTCGTCGTCGAATTAAGC 3'; 4CL-GSP2: 5' GGAAGCCACGGCAAGCACCATAGACA 3'。

5' RACE 和 3' RACE 具体步骤均按 Clone-Tech Racer Kit (Clone-Tech) 的说明书进行, PCR 产物回收后经 T 载体克隆、测序, 并与已克隆的 *4CL* 基因编码区序列拼接, 从而获得全长 cDNA 序列。然后, 以拼接得到的 *4CL* 全长 cDNA 设计引物, 再用 PCR 扩增, 凝胶电泳分析、测序, 验证所获基因全长的准确性。*4CL* 全长 cDNA 设计引物序列: 4CL-F': 5' ATCTCACCAACAAGACCAACATC 3'; 4CL-R': 5' ATAATTGAGGAAAGTCGTGGCTGC 3'。

分析氨基酸序列及其基本性质时, 用序列在线处理工具包网站 <http://www.bioinformatics.org/sms/> 上的 show translation 软件分析推测光皮桦 *4CL* 基因编码蛋白质 4CL 的氨基酸组成以及分子量。并利用推测的蛋白质序列在 http://au.expasy.org/tools/pi_tool.html 上用 Compute pI/MW 工具分析其等电点 (pI); 在 <http://www.bioinfo.tsinghua.edu.cn/SubLoc/> 进行蛋白质的亚细胞定位分析。

分析 4CL 蛋白质的结构特征时, 联网到 <http://www.expasy.ch/cgi-bin/prosite/PSScan.cgi> 上, 对 4CL 蛋白质序列进行重要功能基序 (motif) 的搜寻, 并用 NCBI 上的蛋白质保守功能域推测工具 Conserved Domain Search 在 <http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi> 上分析 4CL 的功能域。

分析光皮桦中 4CL 与其他植物中同源蛋白序列的同源性时, 在 NCBI 上用其 Blast (<http://www.ncbi.nlm.nih.gov/blast>) 程序分别将光皮桦的 4CL 蛋白质序列与其他植物的 4CL 蛋白质序列进行同源比对。并同时用 Clustal 软件对不同植物的 4CL 蛋

白序列与光皮桦4CL蛋白序列进行多重序列比对,比较它们之间的同源性。并用Phylip软件构建这些不同植物的4CL基因的进化树,分析他们之间的进化关系。

对光皮桦中4CL进行不同组织的表达分析时,分别用上述方法提取光皮桦叶、茎、花、根组织中的RNA,反转录成cDNA。以4CL的cDNA为模板设计半定量、定量PCR引物,进行该基因的组织表达特异性分析。其中半定量PCR扩增产物用1.2%的琼脂糖凝胶进行电泳分析。定量PCR采用Applied Biosystems (ABI)公司生产的7500型荧光定量PCR进行分析。半定量RT引物4CL-F: 5' CTTGCTGCTCACCCACCCAAACAT 3', 4CL-R: 5' GAGGAAAGTCGTGGCTGCAACGTC 3'; 定量RT-PCR引物4CL-F: 5' TCGTCGGTAAAGATGCTCAAG 3', 4CL-R: 5' AAAACTGGACCTGCCTCTGTC 3'。

结果与讨论

1 光皮桦4CL基因的克隆和分析

用根据杨树、东北白桦以及桉树4CL的序列保守区域设计的兼并引物4CL-F和4CL-R,经过RT-PCR扩增,获得一个约510 bp的产物,连接到T载体上经转化DH5 α 后,挑取阳性克隆测序。测序结果用NCBI的BLAST程序检索数据库后,观察到该序列与杨树、东北白桦以及桉树的同源性都达到了80%以上,尤其是与白桦的序列同源程度最高,达到95%。确定该片段为光皮桦的4CL基因片段,然后用该序列设计5'端RACE特异引物4CL-GSP1和3'端RACE特异引物4CL-GSP2,分别用于基因5'端和3'端序列的扩增。经5'和3'RACE分别获得约为1.7 kb和750 bp的PCR产物,克隆测序的结果表明,5'和3'端序列与已获得的光皮桦4CL编码序列有重叠,为4CL基因mRNA的5'和3'端序列,3个序列经软件拼接,得到光皮桦4CL基因完整的cDNA序列,共1983 bp,GenBank登录号为FJ410448。该基因全长扩增(图1)和测序的结果也验证了所获基因的准确性。

2 光皮桦4CL基因编码蛋白质的分析

分析光皮桦4CL基因的全长cDNA序列表明,它的ORF有1629 bp,从cDNA全长的69位到1697位,编码542个氨基酸(图2),预测其分子量

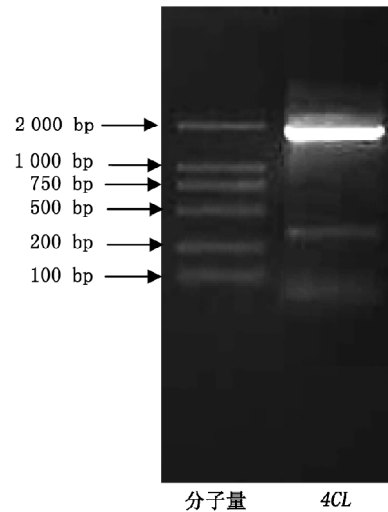


图1 光皮桦4CL全长cDNA RT-PCR产物的琼脂糖电泳图谱

Fig.1 Agrose gel electropherogram of RT-PCR product of 4CL full-length cDNA in *B. luminifera*

为59.05 kDa,等电点为5.848。蛋白质亚细胞定位预测认定其为细胞质蛋白(reliability index=3; expected accury=84%)。这与其他模式植物中经研究认定的4CL亚细胞定位结果是相吻合的。而且它不是分泌蛋白,不具有信号肽结构。此外,在进行保守结构功能域分析的结果显示,该蛋白质具有一个AMP结合结构功能域,该结构域为4CL蛋白的一个特征性结构域,负责催化木质素合成途径中4-香豆酸和辅酶A的连接,形成4-香豆酸辅酶A。另外,进行序列重要基序的搜寻,发现光皮桦4CL序列中第186~197个氨基酸为4CL的共有基序结构,它由12个氨基酸(LPYSSTGTLPK)组成。

3 不同物种间4CL基因的同源性分析

将光皮桦4CL基因与另外几个已知4CL基因编码蛋白序列的阔叶树种欧洲山杨(*Populartremula*, AF283552)、东北白桦(*Betula platyphylla*, AY792353)、桉树(*Eucalyptus camaldulensis*, DQ147001)和几个双子叶植物树莓(*Rubus idaeus*, AF239687)、大豆(*Glycine max*, FJ770469)、藿香(*Agastache rugosa*, AY587891)、烟草(*Nicotiana tabacum*, U50845)以及用作模式植物的拟南芥(*Arabidopsis thaliana*, NM_104046)进行同源比较的结果表明,在核酸水平上光皮桦4CL与阔叶类树种有很高的一致性。其与东北白桦最高为98%,与

```

1 aatatctcaccacaagaccacaacatcttccccatctctttgtgcttaattctcgtcga
1 M A I Q A K E Q E F I F R S K L P D
61 cagtccttATGGCTATCCAAGCAAAGAGCAAGAATTCATTTCCGGTCCAAACTCCGG
19 I Y I P K H L P I H S Y C F E N I S K V
121 ACATTTACATTCAAAACATCTCCCCATCCACTCCTATTGCTTCGAAAACATAICCAAGG
39 G S R P C L I N G L T G K V Y T Y Y D V
181 TCGGCTCGCGCCCGTGTGTTGATCAATGGCTTGACCGGCAAGGTATACACGTATTATGATG
59 E L T A R K V A S G L S K L G I Q K G D
241 TTGAGCTCACCGCGTAAGGTCGCTTCGGCCTTAGCAAGCTCGGCATTAGAAAGGGAG
79 V V M L L L P N S P E F A F V F L C A S
301 ATGTTGATGCTTTTGCTACCCAATCGCCGGAATTCGCCTTCGTATTCCCTGGCGCTT
99 Y L G A M T T A A N P F C T A G E V S K
361 CTTATTTAGGGGCCATGACCACGGCTGCCAACCCCTTCTGCACAGCCGGGGAGGTCTCAA
119 Q A K S A N A K I V V I Q A C Y Y D R V
421 AGCAAGCGAAATCGGCCAACGCAAAAATCGTTGTCACCCAAGCTTGTACTACGACAGAG
139 K D Y A N E N G V K I I C I D S P P E D
481 TGAAGACTACGCAATGAAAACGGTGTAAAGATTATCTGCATCGACTCGCCCCAGAAAG
159 C L H F S E L T K A D E N D V A E V D I
541 ACTGCTTGCAATTTCTCGGAGCTAAGGCGGACGAGACGCTCGCCGAGGTGGATA
179 S P D D V V A L P Y S S G T T G L P K G
601 TTAGCCCCGACGACGTCGTGGCGCTGCCGTACTCTTCGGGGACGACTGGGTGCCAAAAG
199 V M L T H K G L V T S V A Q Q V D G E N
661 GGGTGAIGTTAACGCACAAAGGGCTGGTACTAGCGTGGCTCAACAAGTGGACGGAGAAA
219 P N L Y Y H S E D V I L C V L P L F H I
721 ACCCAAACCTCTATTATCACAGCGAGGACGTGATCTGTGCGTTTTGCGGCTGTTCACA
239 Y S L N S V F L C G L R A G A S I L I L
781 TTTACTCTCAACTCGGTCTTCTTTGCGGCCTGCGAGCGGGTGTCTCGATTCTGATAT
259 P K F E I V S L L Q L I Q K H K V Y V M
841 TGCCCAAGTTTGAGATTGTTTGTGTTGTCAGCTGATACAGAAACATAAGGTTACTGTCA
279 F I V P P I V L A I T K F P D L D K Y D
901 TGCCCAATGTCCCGCGATTGTGCTGGCCATAACTAAGTTCACAGATCTCGATAAGTACG
299 L S S V K M L K S G G A P L G K E I E E
961 ACCTGTGTCGGTAAAGATGCTCAAGTCCGGCGGGCGCGCTGGGAAGGAGATTGAAG
319 T V K A K F P N A L F G Q G Y G M T E A
1021 AAAGTGTAAAGCCAAAGTTTCCCAATGCTCTATTGGTCAAGGATACGGAATGACAGAGG
339 G P V L A M C L A F A K E P M Q V K S G
1081 CAGTCCAGTTTTGCAATGTGCTTGGCATTGCCCCAAGAACCAATGAGTGAATCAG
359 A C G T V V R N A E M K I V D P E T G A
1141 GAGCATGTGTTGTTGAGGAATGCAGAAATGAAAATGTGACCCCGAAACCGGTG
379 S L A R N Q P G E I C I R G D Q I M K G
1201 CCTCTTTGGCCGAAACCAACCGGGAGAGATATGCATCAGAGGAGACCAGATCATGAAAG
388 Y I N D P E A T A S T I D K E G W L H T
1261 GTTATATTAATGATCCGGAAGCCACGGCAAGCACCATAGACAAAGAAGGTTGGTTACATA
419 G D I G L I D D D D E L F I V D R L K E
1321 CCGGCGATATTGGCTTAATTGACGACGATGATGAGCTCTTCATCGTTGACCGTTTGAAGG
439 L I K Y K G F Q V A P A E L E A L L T
1381 AGCTGATCAAATACAAAGGTTCCAGGTAGCCCCCGCGAGCTTGAGGCCTTGCTGCTCA
459 H P N I S D A A V V P M K D D L A G E V
1441 CCCACCAACATCTCTGATCGCCGCTGTCCTCAATGAAGGACGATCTAGCTGAGGAGG
479 F V A F V A R S N G S Q V T E D E I K Q
1501 TCCTGTGGCATTGTGCGGAGATCAAACGGGTCTCAAGTCAACAGAGGATGAAATTAAGC
499 F V S K Q V V F Y K R I S R V F F I D V
1561 AGTTTCTCTAAACAGCTGCTGTTTTATAAAAAGATAAGCCGACTATTTTTCATCAGATC
519 V P K S P S G K I L R K E L R A K L A A
1621 TGGTTCCCAAGTCACCATCGGGGAAGATATTACGCAAGGAGTTGAGGGCAAACTGGCAG
539 G F S N *
1681 CTGGCTTTTCAAATGAatctacagatattgcaaggcttgaagaagaatgaaaagctatt
1741 aatggacgttgcagccacgactttcctcaattatttggatatttggaaagaaagcctgt
1801 ataaaggatttgagttgcaattcccaactacaacaccttttatgtaatcgcaaatctt
1861 tctttttctttaccgtgtttgcatttgagacaatggagccaaatcctacaatatggct
1921 tgttggaaacttatatcaattataagtttttttaaaaaaaaaaaaaaaaaaaaaaaaaaaa
1981 aaa

```

图2 光皮桦4CL基因的cDNA序列及由其推测的氨基酸序列

Fig.2 cDNA sequence of 4CL gene and its deduced protein in *B. luminifera*

用*号标出的为终止密码子, 方框内为功能基序的氨基酸序列。

藿香的最低为51%, 其余均在70%左右。氨基酸序列的同源性也是与藿香最低为57%, 与东北白桦的同源性最高同样是98% (图3)。从Phylip软件构建的4CL基因系统发育树可以看出, 同属桦木科的东北白桦和光皮桦亲缘关系最近, 二者又与茄科的烟

草和桃金娘科的桉树亲缘关系较近, 它们可归为一类; 欧洲山杨和藿香的亲缘关系也比较近。从进化树上看, 蔷薇科的树莓和豆科的大豆4CL蛋白分化较早, 光皮桦4CL蛋白与杨树的4CL蛋白亲缘关系比较远, 而与烟草和桉树的亲缘关系比较近(图4)。

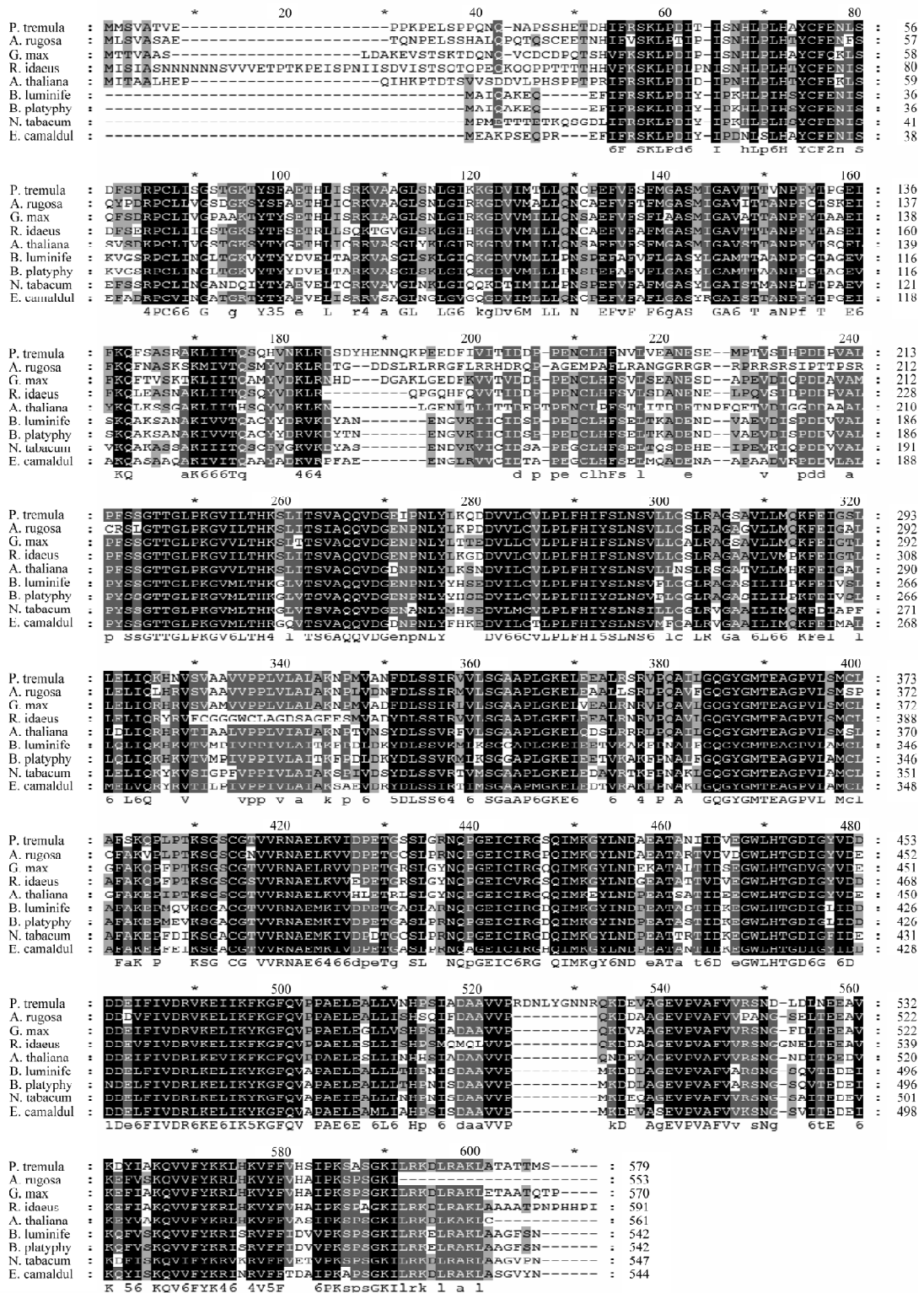


图3 光皮桦4CL蛋白序列及其他植物同源蛋白序列的比较

Fig.3 Comparison of 4CL in *B. luminifera* and its homologous proteins from other plants

欧洲山杨(*P. tremula*, AF283552)、东北白桦(*B. platyphylla*, AY792353)、桉树(*E. camaldulensis*, DQ147001)、树莓(*R. idaeus*, AF239687)、大豆(*G. max*, FJ770469)、藿香(*A. rugosa*, AY587891)、烟草(*N. tabacum*, U50845)、拟南芥(*A. thaliana*, NM_104046)。

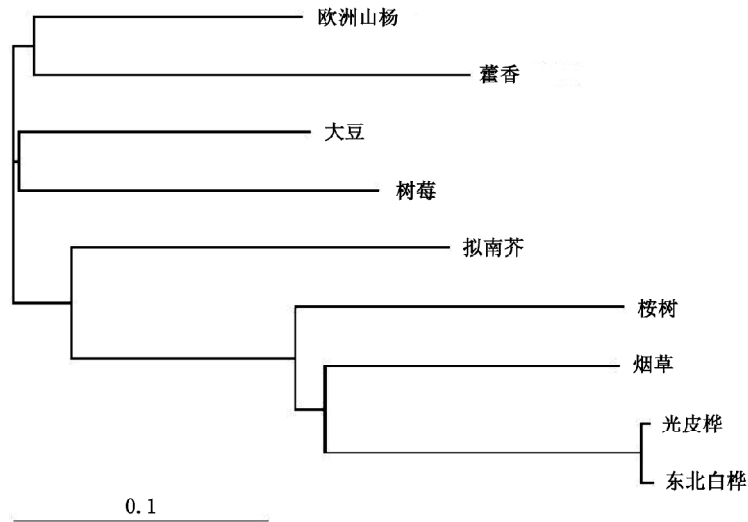


图4 植物4CL蛋白的系统发育树

Fig.4 Phylogenetic tree of 4CL proteins in different plants

4 光皮桉不同器官中 4CL 基因的表达分析

在半定量RT-PCR分析的结果表明,光皮桉中的 4CL 基因在叶、茎、花、根中都有表达,以

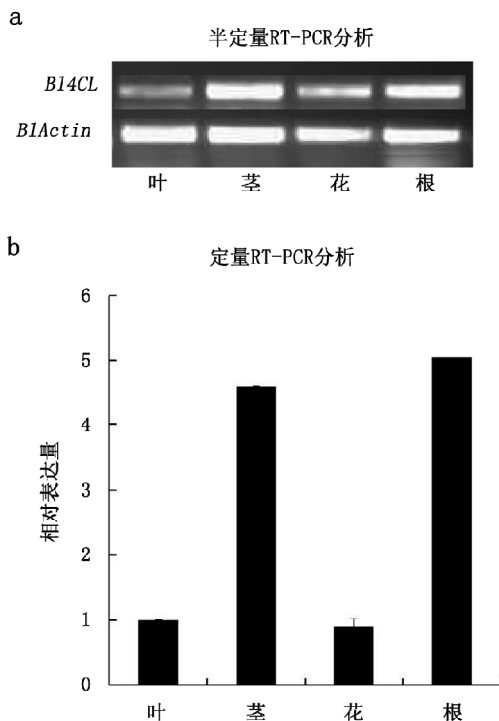


图5 光皮桉 4CL 基因在叶、茎、花、根中的表达分析

Fig.5 Analysis of 4CL gene expression in leaf, stem, flower and root in *B. luminifera*

根和茎的表达量为高,而花和叶中的表达量比较低(图 5-a)。定量 RT-PCR 与半定量 RT-PCR 的结果是一致的,以叶中 4CL 基因的表达量为基准,该基因在花中的表达量是叶中的0.89倍,而根和茎中的表达量分别是叶中的 5.04 倍和 4.59 倍(图 5-b)。这种表达模式同样与 4CL 作为木质素合成途径中关键酶的功能相一致。

参考文献

- 侯夫云, 赵兵, 赵宝杰, 王庆美, 李爱贤, 张海燕, 董顺旭, 张立明 (2008). 利用改良的 CTAB-LiCl 法提取甘薯块根 RNA. 山东农业科学, 9: 90~92
- Kajita S, Hishiyama S, Tomimura Y, Katayama Y, Omori S (1997). Structural characterization of modified lignin in transgenic tobacco plants in which the activity of 4-coumarate: coenzyme A ligase is depressed. *Plant Physiol*, 114: 871~879
- Lee D, Meyer K, Chapple C, Douglas JC (1997). Antisense suppression of 4-coumarate: coenzyme A ligase activity in *Arabidopsis* leads to altered lignin subunit composition. *Plant Cell*, 9: 1985~1998
- Wagner A, Donaldson L, Kim H, Phillips L, Flint H, Steward D, Torr K, Koch G, Schmitt U et al (2009). Suppression of 4-coumarate-CoA ligase in the coniferous gymnosperm *Pinus radiata*. *Plant Physiol*, 149: 370~383
- Whetten R, Sederoff R (1995). Lignin biosynthesis. *Plant Cell*, 7: 1001~1013
- Zhang XH, Chiang VL (1997). Molecular cloning of 4-coumarate: coenzyme A ligase in loblolly pine and the roles of this enzyme in the biosynthesis of lignin in compression wood. *Plant Physiol*, 113: 65~74