

研究报告 Research Report

杜仲 *EuFLC1* 基因克隆与生物信息学分析

高源隆^{1,2}✉, 董旋^{1,3}✉, 赵德刚^{1,2,3}✉

1 贵州大学生命科学学院山地植物资源保护与种质创新省部共建教育部重点实验室, 贵阳, 550025

2 贵州大学农业生物工程研究院贵州省农业生物工程重点实验室, 贵阳, 550025

3 贵州大学绿色农药与农业生物工程国家重点实验室培育基地, 贵阳, 550025

✉ 通讯作者, dgzhao@gzu.edu.cn ✉ 作者

计算分子生物学, 2014 年, 第 3 卷, 第 4 篇 doi: 10.5376/cmb.cn.2014.03.0004

本文首次发表在《基因组学与应用生物学》(2014 年第 33 卷第 4 期 836-844 页)上。现依据版权所有人授权的许可协议, 采用 Creative Commons Attribution License 协议对其进行授权, 再次发表与传播。只要对原作有恰当的引用, 版权所有人允许并同意第三方无条件的使用与传播。

建议最佳引用格式:

引用格式(中文):

高源隆等, 2014, 杜仲 *EuFLC1* 基因克隆与生物信息学分析, 计算分子生物学(online), 3(4): 1-10 (doi: 10.5376/cmb.cn.2014.03.0004)

引用格式(英文):

Li et al., 2014, Molecular Cloning and Bioinformatics Analysis of Gene *EuFLC1* in *Eucommia ulmoides* Olive, Computational Molecular Biology (online), 3(4): 1-10 (doi: 10.5376/cmb.cn.2014.03.0004)

摘要 以杜仲(*Eucommia ulmoides* Olive)雄花花芽为材料, 依靠实验室构建的杜仲转录组库, 采用 cDNA 末端快速扩增法克隆了 3' 端包含完整开放阅读框的 745 bp cDNA 序列。经 3' RACE 产物和转录组库中 5' 端序列拼接, 获得全长为 872 bp 的杜仲 FLC 的 cDNA 序列, 编码 86 个氨基酸, 命名为 *EuFLC1*。生物信息学分析显示: *EuFLC1* 蛋白分子量约为 10.005 kD, 理论等电点为 10.47, 为亲水性蛋白; 存在 3 个可能的磷酸化位点、1 个可能的核定位信号和 1 个可能的核输出信号; 不存在跨膜螺旋和信号肽; 蛋白二级结构中包含 2 个 α 螺旋和 4 个 β 折叠, 无规卷曲连接 α 螺旋及 β 折叠; 蛋白三级结构同源建模结果表明, *EuFLC1* 蛋白包含 2 个互相垂直的 α 螺旋, α 螺旋间有 2 个处于同一平面的 β 折叠, 蛋白的 N 端和 C 端为无规卷曲并伸向整体结构的一侧; 是含有 MEF2_like 型 MADS-box 结构域的 MADS-Box 基因。Blastp 结果中, 与 *EuFLC1* 同源性较高的序列有: 葡萄的 flowering locus C、小粒种咖啡的 MADS-box protein FLC subfamily、巨峰葡萄的 flowering locus C-like MADS-box protein, 因此认为 *EuFLC1* 是杜仲 FLC 基因。系统发育树表明杜仲与同属唇形类植物的小粒咖啡聚为一支, 与 APG III 分类系统的分类结果一致。该基因为首次从第三纪孑遗植物杜仲中克隆到的 MADS-Box 基因, 为研究 MADS-Box 基因的演化的提供了一个重要材料, 为研究杜仲开花时间的分子调控机制奠定了基础。

关键词 杜仲; *EuFLC1*; 克隆; 生物信息学分析

Molecular Cloning and Bioinformatics Analysis of Gene *EuFLC1* in *Eucommia ulmoides* Olive

Gao Yuanlong^{1,2}✉, Dong Xuan^{1,3}✉, Zhao Degang^{1,2,3}✉

1 Key Laboratory of Plant Resources Conservation and Germplasm Innovation in Mountainous Region, Ministry of Education, College of Life Sciences, Guizhou University, Guiyang, 550025

2 Guizhou Key Lab of Agro-bioengineering, Institute of Agro-bioengineering, Guizhou University, Guiyang, 550025

3 State Key Laboratory Breeding Base of Green Pesticide and Agricultural Bioengineering, Guizhou University, Guiyang, 550025

✉ Corresponding author, dgzhao@gzu.edu.cn ✉ Authors

收稿日期: 2014 年 09 月 08 日

接受日期: 2014 年 09 月 22 日

发表日期: 2014 年 10 月 08 日

基金项目: 本研究由国家 863 计划“特色植物功能基因组学研究与应用”子课题“特色林木功能基因组学研究与应用”子课题“杜仲功能基因组研究与应用”(2013AA102605-05)资助。

Abstract Based on the transcriptome library constructed from *Eucommia ulmoides* Olive by our laboratory, the gene *FLC* of *Eucommia ulmoides* was cloned by rapid amplification of cDNA ends (RACE) by using male flower buds of *Eucommia ulmoides* as experiment material and named as *EuFLC1*. The gene encodes a

protein composed of 86 amino acids. The predicted molecular mass and the isoelectric point (PI) of EuFLC1 are 10.005 kD and 10.47, respectively. Protein sequence analysis revealed the presence of three phosphorylation sites, one nuclear localization signal (NLS) and one nuclear export signal (NES), and the absence of signal peptide sequence and transmembrane helix structure. The protein secondary structure of EuFLC1 has two α -helixes and four β -strands. The result of protein structure homology modeling shows that EuFLC1 has two mutually perpendicular α -helixes and two coplanar β -strands between the α -helixes, and the N-terminal and C-terminal of EuFLC1 are random coils extending toward the same side of the overall structure. With EuFLC1 containing a MEF2_like sequence, *EuFLC1* belongs to the MADS-Box gene. According to Blastp analysis, EuFLC1 is more homologous to flowering locus C (*Vitis vinifera*), MADS-box protein FLC subfamily (*Coffea arabica*) and flowering locus C-like MADS-box protein (*Vitis labrusca* × *Vitis vinifera*). Phylogenetic tree shows that *Eucommia ulmoides* and *Coffea arabica* are in same branch, which is concordant with APG III classification results that two trees are all members of the lamiales. *EuFLC1* is a MADS-Box gene that was cloned for the first time from the tertiary relict plant *Eucommia ulmoides*, which may provide an important material for studying the evolution of MADS-Box gene and lay the foundation for the study of the molecular mechanism of *EuFLC1* regulating flowering time in *Eucommia ulmoides*.

Keywords *Eucommia ulmoides* Olive; *EuFLC1*; Molecular cloning; Bioinformatics analysis

开花是植物生长发育过程中最重要的生理过程,这一过程涉及发育方式的转换、花器官的发生和发育,以及在外界环境条件的作用下,生理生化信号的产生、传递和相互作用等(雍伟东等, 2000, 科学通报, 45(5): 455-466)。不同植物开花有 4 条不同的途径,分别是: 依赖光周期的开花途径、自主开花途径、依赖春化的开花途径和赤霉素途径(He and Amasino, 2005)。FLC (FLOWERING LOCUS C) 属于 MADS-Box 基因,它编码的 MADS-Box 转录因子对拟南芥成花具有抑制作用(Michaels and Amasino, 2001)。自主开花途径和依赖春化的开花途径都通过负调控 FLC 的表达,起始植物开花的生理过程(He and Amasino, 2005)。FLC 基因处在这 2 条途径的交汇处,可见 FLC 基因在植物开花过程中有重要的地位。杜仲(*Eucommia ulmoides* Oliver)是我国传统的名贵中药材,有二千多年的药用历史(杜红岩, 2003),属于杜仲科(Eucommiaceae)杜仲属(*Eucommia*),是杜仲科中仅有的一个种,为第三纪孑遗植物。杜仲又是极具潜力的温带胶源树种,杜仲胶属于反式-聚异戊二烯,具有传统顺式-聚异戊二烯橡胶所不具有的优良特性(张继川等, 2011)。杜仲研究具有较高的经济价值,而且在研究被子植物系统演化以及中国植物区系的起源等诸多方面有重要的科学意义(杨峻山等, 1997)。杜仲的 MADS-Box 基因在 NCBI 上还未见报道,本文拟通过 RACE (rapid amplification of cDNA ends) 技术克隆杜仲的 FLC 基因的 cDNA 全长序列,旨在为杜仲的繁育和研究被子植物系统演化中 MADS-Box 基因的演化提供理论基础。

1 结果与分析

1.1 杜仲目的基因克隆

提取的杜仲雄花总 RNA 质量满足 RACE 试剂盒对 RNA 的要求,电泳检测可见清晰的 28S RNA 和 18S RNA 条带,且 28S 的宽度与亮度约为 18S 的两倍(图 1A)。反转录后,通过 3'-RACE 扩增获得目标条带,大约为 760 bp (图 1B)。将目标条带连入 T 载体并转入大肠杆菌,挑取阳性菌落,摇菌提取质粒。双酶切位点在空 T 载体上间隔约 100 bp,插入片段约 760 bp。提取阳性菌株的质粒,双酶切后较小的片段大约为 860 bp,符合预期大小(图 1C)。

1.2 碱基序列分析

将测序得到的目的基因 3'端序列和转录组库中的序列拼接,通过重叠区域确定目的基因的 cDNA 序列为 872 bp,开放阅读框位置在 127~387 bp 共 261 bp,编码 86 个氨基酸(图 2)。基因命名为 *EuFLC1*。

EuFLC1 基因其开放阅读框长度不到 cDNA 长度的 1/3,只编码 86 个氨基酸的小蛋白,这同类基因中较为少见。但 NCBI 中也有报道类似的序列,其开放阅读框占 cDNA 长度的比例较小,编码的蛋白较小(表 1)。

1.3 *EuFLC1* 蛋白中氨基酸性质分析

利用 Protean 软件分析 *EuFLC1* 的 86 个氨基酸,出现频率最高的 4 种氨基酸分别是: Arg (13.95%)、Val (10.47%)、Lys (9.30%)和 Ser (8.14%)。带电荷

氨基酸 36 个(41.86%), 占总分子量的 48.99%。酸性氨基酸 10 个(11.63%), 占总分子量的 12.20%; 碱性氨基酸 20 个(23.26%), 占总分子量的 28.98%。疏水性氨基酸有 27 个(31.40%), 占总分子量的 29.38%。极性氨基酸有 17 个(19.77%), 占总分子量的 17.79%。

1.4 *EuFLC1* 蛋白性质预测

用 ProtScale 软件和 ProtParam 在线软件分析, *EuFLC1* 蛋白分子量约为 10.005 kD, 理论PI为 10.47, GRAVY 为-0.553, 说明其为亲水性蛋白(图 3A)。在细胞内的 pH 条件下, 该蛋白整体带正电, 利于结合到带负电的核酸上。NetPhos 对 *EuFLC1* 的预测结果显示目标蛋白可能的磷酸化位点是 15、22、36 位的丝氨酸(图 3B), 得分分别是: 0.932、0.996、0.905。将 *EuFLC1* 的 N 端甲硫氨酸去除, 保留的氨基酸序列为 *EuFLC1* 前体蛋白。用 TMHMM 在线分析软件预测 *EuFLC1* 前体蛋白的跨膜区(图 3C)。预测结果显示, *EuFLC1* 不存在跨膜螺旋, 其不属于跨膜蛋白。

1.5 *EuFLC1* 蛋白定位信号预测

使用 SignalP 在线预测分析 *EuFLC1* 前体蛋白的信号肽, 结果显示(图 4A): 自然分裂位点分值(C-score)的峰值 0.113 在第 52 位, 组合分裂位点分值(Y-score)的峰值 0.111 也在第 52 位, 而信号肽分值(S-score)在 N 端无明显峰值, 综合分析认为目的蛋白中不存在信号肽。用 NLStradamus 对 *EuFLC1* 蛋白氨基酸序列进行分析, 目的蛋白存在可能的核定位信号(nuclear localization signal, NLS), 序列为 5K VEIKRIENKSKRQVAFSKRRNGLMKK31 (图

4B)。用 NetNES 对 *EuFLC1* 蛋白氨基酸序列进行分析(图 4C), 结果显示目的蛋白的中部存在一个可能的核输出信号(nuclear export signal, NES), 序列为 35LSVLC DV41。

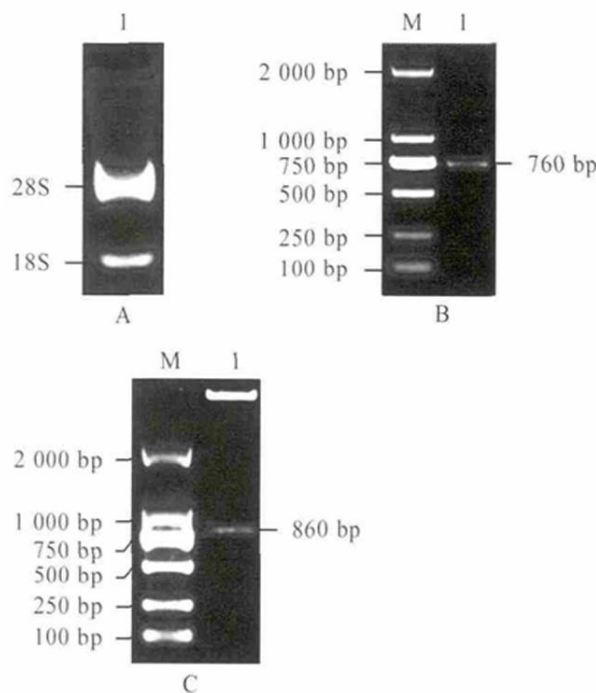


图1 电泳检测

注: M: DNA Marker; I: 样品; A: 总 RNA 电泳检测; B: 3'-RACE 扩增片段; C: 重组质粒双酶切鉴定

Figure 1 Electrophoresis

Note: M: DNA Marker; I: Sample ; A: Electrophoresis of total RNA; B: The cDNA fragment amplified by 3'-RACE; C: Restriction digestion of recombinant plasmids

```

1  CGTCCGTCATCAATCAATTTCTTGATCGAGGATTACAATTTTGCAATAATTTCTTTGAGCAAATGATATAAAAAGTTAGAGAGAGAATGAG
91  CGATTGCGAGGAGTTAGGGCTTGCGAGAGAGACAGGATGGGGCGAAGGAAGGTGGAGATAAAGCGAATCGAGAAACAAGAGCAAGAGACAG
      M G R R K V E I K R I E N K S K R Q 18
181 GTTGCCTTCTCAAAGAGACGAAATGGATTGATGAAGAAAGCTCGAGAGCTGTCCGTTCTTTGCGACGTCGACGTCGCGGTCGTCATTTTC
    V A F S K R R N G L M K K A R E L S V L C D V D V A V V I F 48
271 TCCACCCGTGGGAAGCTCTACGACTTCTCCGGCGGCACAGATCTCGAATGATTGAGAGACACGTTGATCTCGTTCATTGCTTCGGTGAG
    S T R G K L Y D F S G G H R S R M I Q R H V D L V H C F G E 78
361 ATCTCGAATGATTTAGAGACGCGTTGATCTCGTTCATACTTATCTACACTGGACAACGAAGTTATAATTAATTTAGCCTGCAAAAGGCTG
    I S N D L E T R * 86
451 TTCACGCATCATCTCAGATTCTCAGTATACGTTTCCACTTGGTTTAGGAAGCGGTTGCCGCAACTTTTTCCCCCCTGTTTTGTGCTTTT
541 GTTGCAACCGAGACCGGCTTATTGCTCGTGGTTCATAAAGTCAGTCTCGAGAAAGATCAGTGCCTAGTATATTATTATTTAAATGA
631 CTGGTGTTCATTGAAAATAATTGATTGATCCCGATGGATGTAGTTCAGTTCCTTCTTTCTTTCGTTGGGATATGTAGTTCCTTCTTTTCG
721 TGGAAGGTGCCGATGGATATGTAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC
811 CAATGAAGATGATGATTGAAAATTACTTGGTTGTAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAA
    
```

图2 *EuFLC1* 的cDNA 序列及相应的氨基酸序列

Figure 2 Nucleotide sequences of the *EuFLC1* cDNA and its deduced amino acid sequences

表1 Blastp 结果中与EuFLC1 长度相近的蛋白序列

Table 1 Protein sequences with similar length of EuFLC1 in blastp results

蛋白描述	登录号	蛋白长度(aa)	cDNA 长度(bp)
Protein definition	Accession No.	Protein length (aa)	cDNA length (bp)
预测蛋白 大叶杨 Predicted protein (<i>Populus trichocarpa</i>)	XP_002303785	90	845
MADS affecting flowering 4 变异体IV 拟南芥 MADS affecting flowering 4 variant IV (<i>Arabidopsis thaliana</i>)	AAO65318	66	761
预测蛋白 大叶杨 Predicted protein (<i>Populus trichocarpa</i>)	XP_002299316	81	>246
MADS affecting flowering 4 变异体III 拟南芥 MADS affecting flowering 4 variant III (<i>Arabidopsis thaliana</i>)	AAO65317	63	854
假定蛋白PHAVU_003G039400g 菜豆 Hypothetical protein PHAVU_003G039400g (<i>Phaseolus vulgaris</i>) MADS-box 蛋白SEP2B	ESW25477	86	>261
稷斗菜 MADS-box protein SEP2B (<i>Aquilegia coerulea</i>) MADS box 转录因子	AFX72878	85	>258
苜蓿 MADS box transcription factor (<i>Medicago truncatula</i>)	XP_003614241	76	>231

1.6 EuFLC1 蛋白二级结构分析

PSIPRED 在线对 EuFLC1 分析的结果显示, 目的蛋白二级结构中包含 2 个 α 螺旋(α -helix)和 4 个 β 折叠(β -strand), 无规卷曲(random coil)连接 α 螺旋及 β 折叠(图 5)。

1.7 EuFLC1 蛋白三级结构分析

SWISS-MODEL 以肌细胞特异性增强因子 2B(myocyte-specificenhancerfactor2B)为模板, 对 EuFLC1 蛋白三级结构进行同源建模, EuFLC1 蛋白包含 2 个互相垂直的 α 螺旋, α 螺旋间有 2 个处于

同一平面 β 折叠, 蛋白的 N 端和 C 端为无规卷曲并伸向整体结构的一侧(图 6)。

1.8 EuFLC1 蛋白功能预测

经 NCBI 的 Conserved Domain Database 的分析, EuFLC1 具有 MADS-Box 家族亚科中 MADS_MEF2_like 的特征结构域(图 7), 表明 EuFLC1 是 MADS-Box 基因。MADS-box 基因分为 I 和 II 两种类型(A1 varez-Buyllaetal., 2000), 根据 EuFLC1 包含 MEF2_like 类型的 MADS-box 区结构域, 将其归为 II 型。

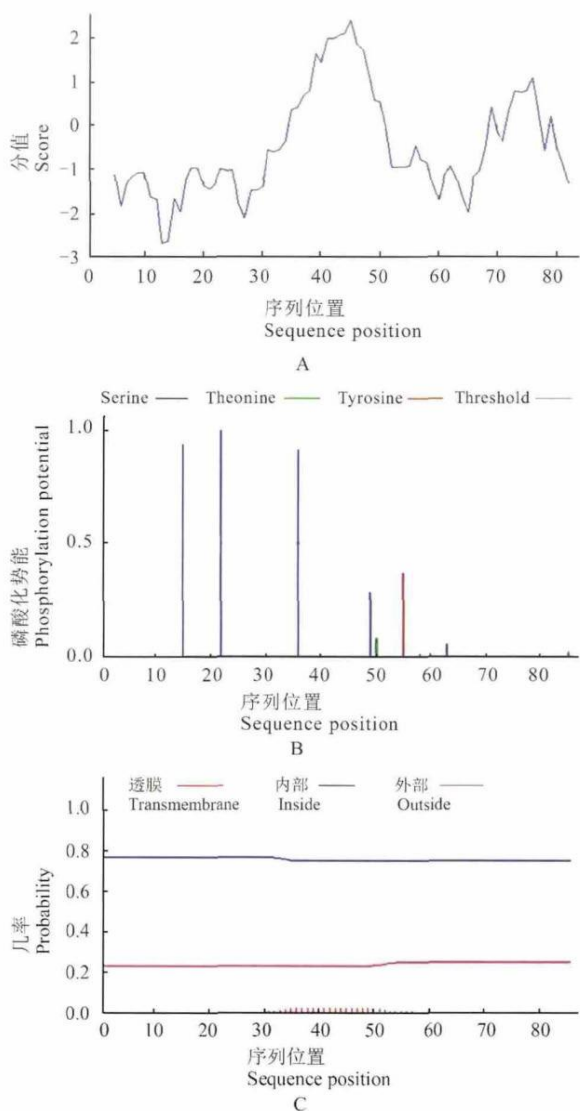


图3 *EuFLC1* 蛋白质性质预测

注: A: ProtScale 对亲水性的预测; B: NetPhos 对磷酸化位点的预测; C: TMHMM 对跨膜螺旋的预测

Figure 3 Predictions of properties for *EuFLC1* protein

Note: A: Predicted hydrophilicity by ProtScale; B: Predicted phosphorylation sites by NetPhos; C: Predicted transmembrane helices by TMHMM

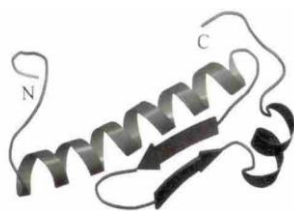


图6 SWISS-MODEL 对*EuFLC1* 蛋白三级结构预测

Figure 6 Tertiary structure prediction of *EuFLC1* by SWISS-MODEL

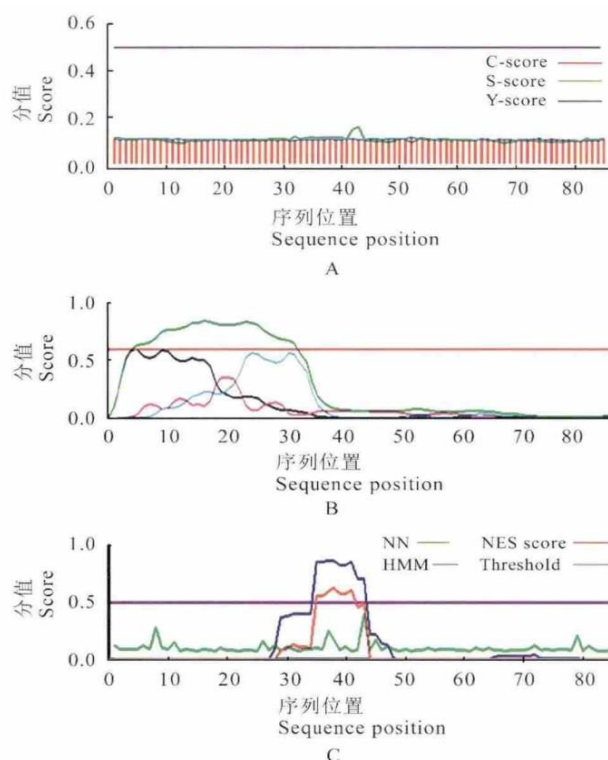


图4 *EuFLC1* 蛋白定位信号预测

注: A: SignalP 对信号肽的预测; B: NLStradamus 对NLS的预测; C: NetNES 对NES的预测

Figure 4 Predictions of orientation signal for *EuFLC1* protein

Note: A: Predicted signal peptides by SignalP; B: Predicted NLS by NLStradamus; C: Predicted NES by NetNES

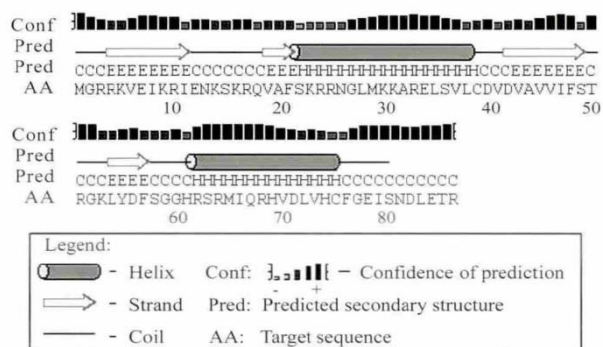


图5 PSIPRED 对*EuFLC1* 蛋白二级结构预测

Figure 5 Secondary structure prediction of *EuFLC1* by PSIPRED

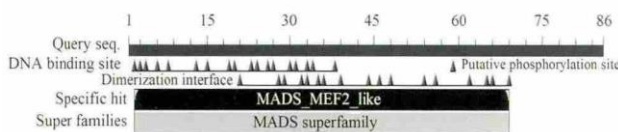


图7 *EuFLC1* 氨基酸序列的保守结构域

Figure 7 Conserved domains on deduced *EuFLC1* amino acid sequence

Protein blast 结果中,与 EuFLC1 同源性较高的序列为:葡萄(*Vitis vinifera*)的 flowering locus C (NP_001268057)、小粒种咖啡(*Coffea arabica*)的 MADS-box protein FLCsubfamily(ADU56823)、巨峰葡萄(*Vitis labrusca* × *Vitis vinifera*)的 flowering locus C-like MADS-box protein (ABR68644) (图 8)。据此判断 *EuFLC1* 基因行使类似 *FLC* 基因的功能,调控开花时间。

1.9 EuFLC1 蛋白氨基酸序列进化树分析

分析 NCBI 中已报道的所有 FLC 蛋白序列,选取已报道的每个种属中具有代表性的、氨基酸序列完整的一个序列作为数据源,以此构建系统发育树(图 9)。结果显示,在被子植物中 FLC 最初进化为 2 个分支,一支主要是草本植物,另一支主要为木本植物。在木本植物的分支中进一步分为 2 个分支,其中一支包括杜仲和小粒咖啡。这说明二者的亲缘关系较近,同时在系统进化上较为原始。在 APGIII 分类系统 (Bremer et al., 2009) 中,杜仲属于绞木目 (Garryales), 小粒咖啡属于龙胆目 (Gentianales), 同属唇形类植物 (lamiids), 这与 EuFLC1 的系统发育树的结果一致。山萮菜 (*Eutrema wasabi*)、芥菜 (*Capsella rubella*)、白芥 (*Sinapis alba*)、盐芥 (*Eutrema halophilum*)、拟南芥 (*Arabidopsis thaliana*)、白菜 (*Brassica rapa*)、油菜 (*Brassica napus*)、甘蓝 (*Brassica oleracea*)、萝卜 (*Raphanus*

sativus) 都是十字花目 (Brassicales) 植物,在发育树中聚在一个分支上。大豆 (*Glycine max*) 是豆类植物 (fabids), 独立一支,与十字花目植物分支并列。可可 (*Theobroma cacao*)、枳 (*Citrus trifoliata*)、胡桃 (*Juglans regia*)、山核桃 (*Carya cathayensis*)、娑罗双 (*Shorea beccariana*)、葡萄、白桦 (*Betula platyphylla*) 分属于豆类植物 (fabids)、锦葵类植物 (malvids) 和葡萄目 (Vitales) 的植物,聚在一个大的分支上,与唇形类植物分支并列。系统发育树中,除杜仲和小粒咖啡是菊类植物 (asteroids), 其它均为蔷薇类植物 (rosids)。

2 讨论

杜仲是第三纪孑遗植物,与它同时期的大部分植物都已灭绝(杨峻山等, 1997)。杜仲胶属于反式-聚异戊二烯,目前仅已知我国特有的杜仲(张继川等, 2011)、东南亚热带雨林里的古塔波树 (*Palaquium gutta*) (Bamba et al., 2001) 和南美热带雨林里的巴拉塔树 (*Mimusops balata*) (Lovering, 1970) 等少数植物能够合成反式-1,4-聚异戊二烯,因此国际上杜仲胶又被称作古塔波胶或者巴拉塔胶。已知约 2 500 种植物会产生橡胶 (Mooibroek and Cornish, 2000), 但它们合成的是顺式-聚异戊二烯,说明杜仲具有特殊的胶合成基因。一些杜仲次生代谢相关基因已被克隆并进行了研究(周明兵等, 2003; 赵丹等, 2012; 刘攀峰等, 2013), 基因序列相似性都没有超过 87%。

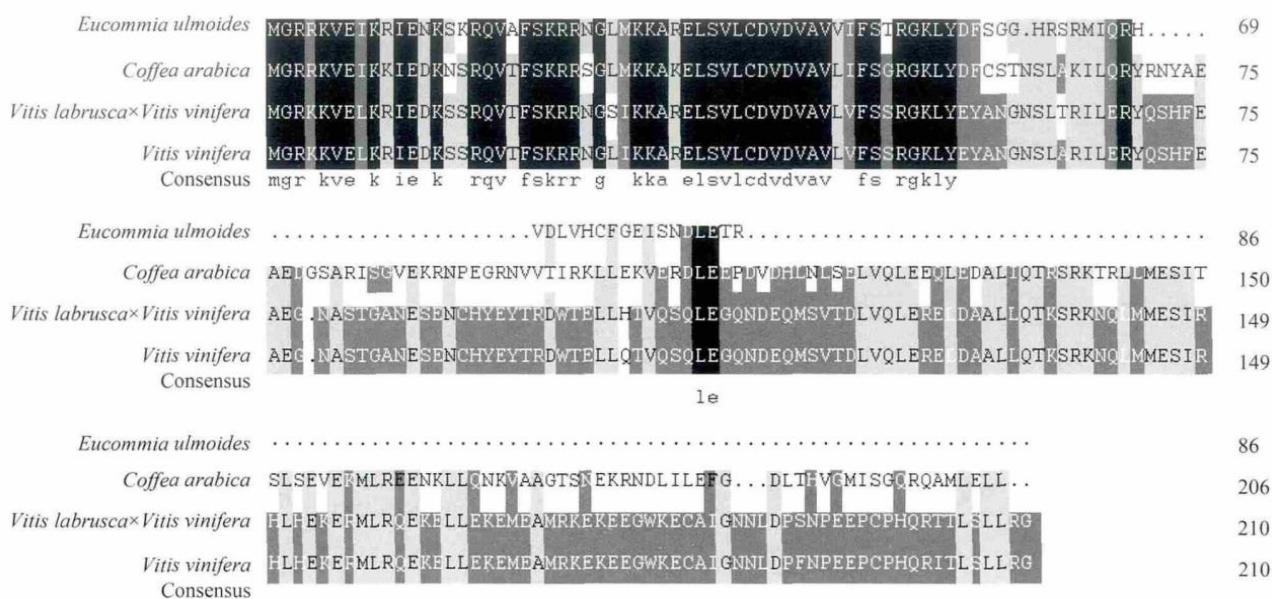


图8 EuFLC1 氨基酸序列与同源序列的多重比对

Figure 8 Multiple sequence alignment in deduced EuFLC1 amino acid sequence and homologous sequences

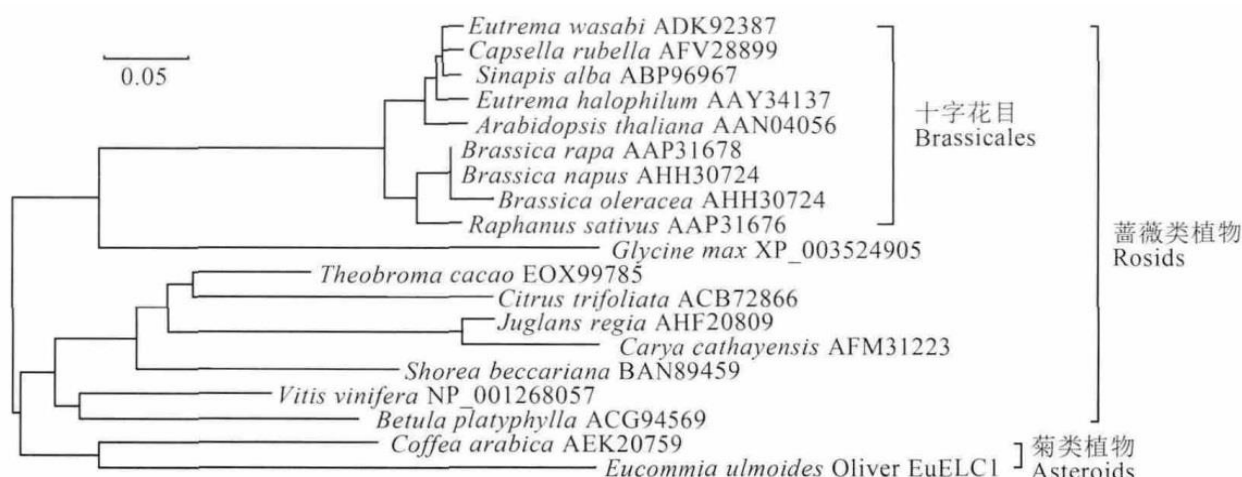


图9 *EuFLC1* 蛋白和NCBI 中报道的主要FLC 蛋白氨基酸序列的系统发育树

Figure 9 Phylogenetic tree of amino acid sequences of *EuFLC1* protein and other FLC protein from NCBI

MADS-Box 基因是植物中重要的基因家族之一，但对杜仲的此类基因克隆和研究还未见报道。本研究克隆的 *EuFLC1* 基因和 NCBI 上已报道的 *FLC* 基因有较大的差异，表现在开放阅读框的长度较短、占 mRNA 的比重较小。花发育相关的 MADS-Box 基因包含约 180 bp 的 MADS-box 区以外，还含有 1 个约 90 bp 的 I-box 区，1 个约 210 bp 的 K-box 区和 1 个比较多变的 C 末端 (Purugganan et al., 1995)。这类 MADS-Box 基因被定义为 II 型，其 MADS-box 区为 MEF2_{like} 型 (Alvarez-Buylla et al., 2000)。根据 MADS-box 区进行分类，含有 MEF2_{like} 型 MADS-box 区的 *EuFLC1* 是花发育中的 II 型 MADS-Box 基因。植物的 II 型 MADS-Box 基因又被分为 MIKCC 型和 MIKC* 型，但二者均包含 K-box (Henscheletal., 2002)。

EuFLC1 不包含 K-box 和比较多变的 C 末端，不属于这二者之一，而独立成为一类，为植物 II 型 MADS-Box 基因中的第三类。动物中的 II 型 MADS-Box 基因不包含 K-box 和比较多变的 C 末端，只有 MEF2 型一种，*EuFLC1* 结构与其相同。Theissen 等 (1996) 认为，在大约 10 亿年前真核生物的最后 一个祖先可能已经拥有至少一个真正的 MADS-Box 基因，MIKC 型基因的最后 一个祖先可能是 MEF2 型基因。推测动物中和杜仲中 MEF2 型基因保留了原始特征，现在大多数植物的 MEF2 型基因为进化的结果。这表明杜仲基因处在较原始的状态，同时也在一定程度上支持了 Theissen 等的观点。处于原始地位的杜仲为单性植株，也在一定程

度上支持了被子植物的祖先可能具有单性的生殖结构的观点 (Donoghue and Doyle, 2000)。

EuFLC1 蛋白有 86 个氨基酸，但在其 Blastp 比对结果中，完整的蛋白序列有近 90% 包含 163~254 个氨基酸。出现这种情况的原因可能是杜仲是第三纪子遗植物，基因较为原始。在 Blastp 比对的结果中，拟南芥中的 MADS affecting flowering 4 variant III 和 IV 是 2 个突变基因的产物，过表达每一个都会影响开花，功能类似 *FLC*，起到抑制开花的作用 (Ratcliffe et al., 2003)。这说明具有较短的氨基酸序列的类 *FLC* 蛋白质一样可以起到 *FLC* 的功能。*EuFLC1* 与这 2 个蛋白具有一定的同源性和相似的大小，又和其它 *FLC* 有较高同源性，因此 *EuFLC1* 可能在杜仲中就发挥着 *FLC* 的功能。

杜仲的雌雄花均无花被等复杂结构，这表明杜仲在花的进化上也表现的较为原始。*AP1* 基因是花器官形态特征基因，其在花发育的 ABC 模型、ABCD 模型和 ABCDE 模型中均属于 A 类基因，是萼片和花瓣正常发育所必需的。Gustafson-Brown 等 (1994) 发现在 *AP1* 突变体中，处在第一轮萼片转变为叶片状结构，第二轮花器官多数不发育，而第三轮的雄蕊和第四轮的心皮则发育正常。遗传学上的研究也显示 *AP1* 在决定花的分生组织特性和第一轮、第二轮花器官的正常发育中起作用 (Haughn et al., 1995)。杜仲无花被，且花芽的结构类似 *AP1* 突变体，因此认为杜仲无花被的现象可能与 *AP1* 有关。如果杜仲中无 *AP1* 基因或类似功能的基因，则其自然不会产生花被结构。如果杜仲

API 存在,则分为 2 种情况。第一种情况是杜仲 *API* 的序列与现在 *API* 有较大差异,功能有所不同。第二种情况是杜仲 *API* 正常,但调控受到影响。*FLC* 基因通过 *FT* 基因间接调控 *API* (He and Amasino, 2005)。K-box 区是发生二聚体化的结构基础(Krizek and Meyerowitz, 1996), C 末端在蛋白复合体的形成和转录激活中起重要作用 (Egea-Cortines et al., 1999), EuFLC1 蛋白无这两结构,可能会对杜仲 *API* 的间接调控产生影响。

3 材料与amp;方法

3.1 材料

3.1.1 主要材料

取杜仲雄花花芽置液氮中备用。大肠杆菌 (*Escherichia coli*)菌株 DH5 α 由本实验室保存。

3.1.2 主要试剂

OMEG A 公司的 Plant RNA Kit、Plasmid Mini Kit、Gel Extraction Kit; Clontech 公司的 SMARTer RACE cDNA Amplification Kit; 北京全式金生物技术有限公司的 pEASY-T1 Cloning Kit; TaKaRa 公司的 DL2000 DNA Maker、Competent Cell Preparation Kit、限制性内切酶 *Xba* I 和 *Kpn* I。

3.2 方法

3.2.1 总 RNA 提取

将 RNA 提取需要使用器皿根据材质分别用 0.1% 的 DEPC 溶液 37 $^{\circ}$ C 浸泡 12 h 或 200 $^{\circ}$ C 干热处理 4 h。液氮研磨杜仲雄花样品后,按照 Plant RNA Kit 说明书提供的实验步骤提取杜仲雄花总 RNA。将总 RNA 保存于 -80 $^{\circ}$ C 冰箱中,并取 5 μ L 样品用 1% 的琼脂糖凝胶电泳检测 RNA 的品质

3.2.2 引物设计

在本实验室构建的杜仲转录组库(另文发表)中筛选与 *FLC* 基因有同源性的基因片段。Blastn 比对结果显示,已知基因片段的 5'端包含 MADS-Box 基因完整的 MADS-box 区,且起始密码子的位置与已报道的同源序列相同。因此确定目标片段包含开放阅读框 5'端的完整序列,只需克隆目的基因的 3'端。根据基因片段的序列和 RACE cDNA Amplification Kit 的要求设计 3'-RACE 引物,通过 PrimerSelect 软件对引物及引物对进行质量检测,

由上海英俊生物技术有限公司合成。GSP: 5'-ATGGGGCGAAGGAAGGTGGAG-3'

3.2.3 逆转录和 3'-RACE

逆转录和 3'-RACE 均使用 SMARTer RACE cDNA Amplification Kit 完成。按照说明书提供的实验步骤反转录杜仲雄花总 RNA。用 GSP 和试剂盒提供的通用引物 UPM 作为双侧引物,按照说明书提供的实验步骤进行 3'-RACE。将 RACE-PCR 产物保存于 -20 $^{\circ}$ C 冰箱中,并取 5 μ L 样品用 1% 的琼脂糖凝胶电泳检测 RACE-PCR 扩增的结果。

3.2.4 连接 T 载体与测序

使用 Gel Extraction Kit 回收目标片段;把目标片段连接到 pEASY-T1 Cloning Vector 上;将重组 T 载体用热激法导入用 Competent Cell Preparation Kit 制备的大肠杆菌 DH5 α 感受态细胞中。挑取阳性大肠杆菌菌落液体培养,用 Plasmid Mini Kit 提取质粒 DNA,限制性内切酶双酶切鉴定,送北京诺赛基因组研究中心有限公司测序。

3.2.5 生物信息学分析

使用 DNAMAN 进行序列分析、序列拼接、开放阅读框查找、氨基酸翻译、多重比对,用 Protean 软件分析氨基酸序列,用 ProtScale 和 ProtParam 在线软件分析分子量、理论等电点(isoelectric point, PI)、亲水性指数(grand average of hydropathicity, GRAVY),用 NetPhos 2.0 Server 在线软件分析磷酸化位点,用 TMHMM Server v.2.0 在线分析软件预测跨膜区,用 SignalP 4.1 Server 在线预测信号肽,用 NLStradamus 在线预测核定位信号(nuclear localization signal, NLS),用 NetNES 1.1 Server 在线预测核输出信号(nuclear export signal, NES),用 PSIPRED 在线对分析蛋白二级结构,用 SWISS-MODEL 在线对蛋白三级结构进行同源建模,用 NCBI 的 Conserved Domain Database 的分析保守结构域,用 Protein blast (Program: BLASTP 2.2.29+)对蛋白进行同源性比对,由 MEGA 6 软件根据相邻连接方法(Neighbor Jointing, NJ)建立系统发生树。

作者贡献

高源隆是本研究的实验设计和实验研究的执行人,并完成数据分析,论文初稿的写作;董旋参

与实验设计和实验研究; 赵德刚教授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

本研究由国家 863 计划“特色植物功能基因组学研究与应用”子项目“特色林木功能基因组学研究与应用”子课题“杜仲功能基因组学研究与应用”(2013AA102605-05)资助, 在此表示感谢!

参考文献

- Alvarez-Buylla E.R., Pelaz S., Liljegren S.J., Gold S.E., Burgeff C., Ditta G.S., De Pouplana L.R., Martínez-Castilla L., and Yanofsky M.F., 2000, An ancestral MADS-box gene duplication occurred before the divergence of plants and animals, *Proc. Natl. Acad. Sci. USA*, 97(10): 5328-5333
- Bamba T., Fukusaki E., Kajivama S., Ute K., Kitayama T., and Kobayashi A., 2001, The occurrence of geometric polyprenol isomers in the rubber-producing plant, *Eucommia ulmoides* Oliver, *Lipids*, 36(7): 727-732
- Bremer B., Bremer K., Chase M., Fay M., Reveal J., Soltis D., Soltis P., and Stevens P., 2009, An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG III, *Botanical Journal of the Linnean Society*, 161(2): 105-121
- Donoghue M.J., and Doyle J.A., 2000, Seed plant phylogeny: Demise of the anthophyte hypothesis? *Current Biology*, 10(3): R106-R109
- Du H.Y., 2003, The progress in research of the active component from *Eucommia ulmoides* and its pharmacology, *Jingjilin Yanjiu (Economic Forest Researches)*, 21(2): 58-61 (杜红岩, 2003, 杜仲活性成分与药理研究的新进展, *经济林研究*, 21(2): 58-61)
- Egea-Cortines M., Saedler H., and Sommer H., 1999, Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in *Antirrhinum majus*, *The EMBO Journal*, 18(19): 5370-5379
- Gustafson-Brown C., Savidge B., and Yanofsky M.F., 1994, Regulation of the arabidopsis floral homeotic gene *APETALA1*, *Cell*, 76(1): 131-143
- Haughn G.W., Schultz E.A., and Martinez-Zapater J.M., 1995, The regulation of flowering in *Arabidopsis thaliana*: Meristems, morphogenesis, and mutants, *Canadian Journal of Botany*, 73(7): 959-981
- He Y., and Amasino R.M., 2005, Role of chromatin modification in flowering-time control, *Trends in Plant Science*, 10(1): 30-35
- Henschel K., Kofuji R., Hasebe M., Saedler H., Münster T., and Theißen G., 2002, Two ancient classes of MIKC-type MADS-box genes are present in the moss *Physcomitrella patens*, *Molecular Biology and Evolution*, 19(6): 801-814
- Krizek B.A., and Meyerowitz E.M., 1996, Mapping the protein regions responsible for the functional specificities of the *Arabidopsis* MADS domain organ-identity proteins, *Proc. Natl. Acad. Sci. USA*, 93(9): 4063-4070
- Liu P.F., Du H.Y., Wuyun T.N., Du L.Y., and Sun Z.Q., 2013, Cloning and sequence analysis of 1-hydroxy-2-methyl-2-e-butenyl-4-diphosphate reductase gene cDNA from *Eucommia ulmoides*, *Linye Kexue Yanjiu (Forest Research)*, 26(4): 447-453 (刘攀峰, 杜红岩, 乌云塔娜, 杜兰英, 孙志强, 2013, 杜仲 *HDR* 基因全长 cDNA 克隆与序列分析, *林业科学研究*, 26(4): 447-453)
- Lovering E.G., 1970, Transcrystallinity and x-ray diffraction in trans-1, 4-polyisoprene, *Journal of Polymer Science Part A-2: Polymer Physics*, 8(10): 1697-1701
- Michaels S.D., and Amasino R.M., 2001, Loss of *FLOWERING LOCUS C* activity eliminates the late-flowering phenotype of *FRIGIDA* and autonomous pathway mutations but not responsiveness to vernalization, *The Plant Cell Online*, 13(4): 935-941
- Mooibroek H., and Cornish K., 2000, Alternative sources of natural rubber, *Applied Microbiology and Biotechnology*, 53(4): 355-365
- Purugganan M.D., Rounsley S.D., Schmidt R.J., and Yanofsky M.F., 1995, Molecular evolution of flower development Diversification of the plant MADS-box regulatory gene family, *Genetics*, 140(1): 345-356
- Ratcliffe O.J., Kumimoto R.W., Wong B.J., and Riechmann J.L., 2003, Analysis of the *Arabidopsis* MADS *AFFECTING FLOWERING* gene family: *MAF2* prevents vernalization by short periods of cold, *The Plant Cell Online*, 15(5): 1159-1169
- Theissen G., Kim J.T., and Saedler H., 1996, Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the

- morpho- logical evolution of eukaryotes, *Journal of Molecular Evolu- tion*, 43(5): 484-516
- Yang J.S., Zhang Y.M., and Jiang S.H., 1997, Recent situation and prospects of the study of *Eucommia ulmoides*, *Ziran Ziyuan Xuebao (Journal of Natural Resources)*, 12(1): 60-67 (杨峻山, 张聿梅, 姜声虎, 1997, 杜仲研究的现状与展望, *自然资源学报*, 12(1): 60-67)
- Zhang J.C., Xue Z.H., Yan R.F., and Fang S.B., 2011, Natural polymer material-recent studies on *Eucommia ulmoides* gum, *Gaofenzi Xuebao (Acta Polymerica Sinica)*, (10): 1105-1117 (张继川, 薛兆弘, 严瑞芳, 方世璧, 2011, 天然高分子材料-杜仲胶的研究进展, *高分子学报*, (10): 1105-1117)
- Zhao D., Li X.Y., Chen J., and Zhao D.G., 2012, Cloning and se- quence analysis of the full-length cDNA of cinnamyl alcohol dehydrogenase gene from *Eucommia ulmoides* Olive, *Shandi Nongye Shengwu Xuebao (Journal of Mountain Agriculture and Biology)*, 31(4): 283-287 (赵丹, 李晓毓, 陈建, 赵德刚, 2012, 杜仲肉桂醇脱氢酶基因全长 cDNA 克隆及序列分析, *山地农业生物学报*, 31(4): 283-287)
- Zhou M.B., Xiao Y.H., Zhu D.X., Pei Y., and Zhao D.G., 2003, Molecular cloning and sequence analysis of *EuFPS* involved into rubber biosynthesis in *Eucommia ulmoides* Olive, *Fenzi Zhiwu Yuzhong (Molecular Plant Breeding)*, 1(1): 66-71 (周 明兵, 肖月华, 朱冬雪, 裴炎, 赵德刚, 2003, 杜仲胶合成相关基因 *EuFPS* 的克隆及序列分析, *分子植物育种*, 1(1): 66-71)