国家项目点击

DOI:10.13193/j.issn.1673-7717.2026.03.004

牛蒡4CL基因家族的克隆鉴定与生物信息学分析

邢艳萍,李春妍,安星华,高洺泽,赵容,康廷国,窦德强,许亮
(辽宁中医药大学药学院,辽宁 大连 116600)

摘要:目的 4-香豆酸:辅酶A连接酶(4-coumarate CoA ligase,4CL)是苯丙烷途径中控制木质素、木脂素等化合物生物合成的关键酶与牛蒡子显微特征石细胞发育和牛蒡苷积累密切相关,研究旨在鉴定牛蒡中的4CL基因,探索牛蒡*Arctium lappa* L.4CL酶基因家族的生物学功能。方法 研究对牛蒡基因组中预测的8个4CL基因编码序列(Coding sequence,CDS)进行克隆测序,对获得序列进行生物信息学分析,并对Al4CL3进行原核表达。结果 鉴定得到牛蒡4CL基因家族的8个成员克隆后除Al4CL5(96%)、Al4CL8存在5个变异位点之外,其他序列与预测序列一致性均为100%;预测Al4CL3、Al4CL5、Al4CL8蛋白具有2个跨膜螺旋结构;亚细胞定位于细胞质、质膜、叶绿体和内质网;基因序列结构存在一定差异,不均匀地分布于牛蒡1、2、3、9、10和14号染色体上;Al4CL1、Al4CL6、Al4CL8氨基酸序列中存在4CL氨基酸保守域;Al4CL基因启动子区域存在多个顺式作用元件,受多种激素或胁迫诱导;系统进化分析显示Al4CL分为3个亚族,共线性分析表明Al4CL基因家族成员与刺苞菜蓟之间存在更近的亲缘关系;构建重组质粒pET-28a-chAl4CL3,成功诱导表达出chAl4CL3蛋白,并完成纯化。结论 对牛蒡4CL基因家族的挖掘与鉴定,为牛蒡木脂素的生物合成及其品质的提高提供了一定的理论依据,为深入研究Al4CL及其生物学功能提供参考。

关键词:4-香豆酸:辅酶A连接酶;基因克隆;生物信息学分析;原核表达
中图分类号:R282.71 **文献标志码**:A **文章编号**:1673-7717(2026)03-0020-10

Cloning, Identification and Bioinformatics Analysis of Niubang(*Arctium lappa*) 4CL Gene Families

XING Yanping, LI Chunyan, AN Xinghua, GAO Mingze, ZHAO Rong, KANG Tingguo, DOU Deqiang, XU Liang
(College of Pharmacy, Liaoning University of Traditional Chinese Medicine, Dalian 116600, Liaoning, China)

Abstract: *Objective* 4-coumarate CoA ligase(4CL) is a key enzyme in the phenylpropane pathway controlling the biosynthesis of lignan, lignin and other compounds closely related to Niubang(*Arctium lappa*) microscopic characterization of stone cell development and Niubang(*Arctium lappa*) glycoside accumulation, the aim of this study was to identify the 4CL genes in Niubang(*Arctium lappa*) and explore the biological functions of Niubang(*Arctium lappa*) 4CL enzyme gene family's biological functions. *Methods* In this study, it cloned and sequenced 8 predicted 4CL gene coding sequences(CDS) in the burdock genome, performed bioinformatics analysis on the obtained sequences and performed prokaryotic expression of Al4CL3. *Results* The eight members of Niubang(*Arctium lappa*) 4CL gene family were identified and cloned with 100% consistency with the predicted sequences, except for Al4CL5 (96%) and Al4CL8, which had five variant sites. Al4CL3, Al4CL5 and Al4CL8 proteins had two transmembrane helical structures; subcellular localization in the cytoplasm, plasma membrane, chloroplast and endoplasmic reticulum and gene structure. There were some differences in the gene structure, unevenly distributed on chromosomes 1, 2, 3, 9, 10 and 14, and there was a 4CL amino acid conserved domain in the amino acid sequence. There were multiple cis-acting elements in the promoter region of the Al4CL gene, which was induced by multiple hormones or stresses. The phylogenetic analysis showed that Al4CL was divided into three subfamilies, and the covariance analysis showed that there was a closer relationship between members of the Al4CL family of genes and the prickly bracteate thistle. There was a closer relationship between members of the Al4CL gene family and the prickly bracteate thistle. Recombinant plasmid pET-28a-chAl4CL3 was constructed, successfully induced the ex-

基金项目:国家自然科学基金面上项目(81773852,81874338);辽宁省联合基金项目(2023-MSLH-180);辽宁省教育厅高校基本科研项目储备项目(2024-JYTCB-022)
作者简介:邢艳萍(1986-),女,辽宁北票人,高级实验师,博士,研究方向:中药资源评价与鉴定。
通讯作者:许亮(1978-),男,辽宁沈阳人,教授,博士研究生导师,研究方向:中药资源评价鉴定与分子生药。E-mail:861364054@qq.com。

pression of chAl4CL3 protein and purified. *Conclusion* The excavation and characterization of burdock 4CL gene family provides a certain theoretical basis for the biosynthesis of Niubang(*Arctium lappa*) and its quality improvement, and provides a reference for the in - depth study of Al4CL and its biological functions.

Keywords: 4 - coumarate; CoA ligase; gene cloning; bioinformatics analysis; prokaryotic expression

牛蒡子是菊科植物牛蒡 *Arctium lappa* L. 的干燥成熟果实，又称“大力子”，是我国传统中药，始载于《名医别录》，具有疏散风热、宣肺透疹、解毒利咽的功效，牛蒡子也是“抗疫三药”—金花清感颗粒的组成部分[1]。现代药理研究表明其主要活性成分牛蒡苷及牛蒡苷元等木脂素类成分具有治疗癌症、抑郁、矽肺病、白血病、降血糖以及杀虫等作用，在全世界范围内均有应用，具有重要的药用价值[2]。

4 - 香豆酸：辅酶 A 连接酶(4 - coumarate CoA ligase，4CL)作为限速酶和关键酶，有助于引导不同苯丙烷前体，在决定苯丙烷总途径特定流向木脂素、木质素、类黄酮等生物合成分支途径中发挥着关键作用[3]。4CL 酶的底物特异性被认为决定了下游反应中代谢通量的方向，4CL 酶催化肉桂酸及其羟基或甲氧基衍生物等形成相应的辅酶 A 酯，这些辅酶 A 酯作为关键中间体进入支链苯丙烷衍生物合成途径。对 4CL 基因和酶活性的深入研究不仅在植物生长发育过程中以及在不利环境中的生长中发挥着重要作用，也可为次级代谢物的生产和调节提供重要的指导[4]。4CL 基因已在许多药用植物中被研究[5-7]，在牛蒡中尚无报道。

本研究基于牛蒡基因组数据[8]，筛选并鉴定出牛蒡 4CL 基因家族的成员，通过生物信息学分析其理化性质、系统进化发育、染色体定位、基因结构和保守基序、启动子区域的顺式作用元件、共线性和亚细胞定位等，对牛蒡 4CL 基因家族成员进行克隆和原核表达研究，为后续深入研究牛蒡 4CL 基因的生物学功能提供理论依据。

1 材料

牛蒡子采收地点为辽宁中医药大学大连校区药草园。经液氮速冻研磨后，保存于 - 80 ℃冰箱备用。

2 方法

2.1 牛蒡 4CL 基因的挖掘及鉴定

从 TAIR 网站(https://www.arabidopsis.org/index.jsp)下载拟南芥 *Arabidopsis thaliana* 13 条 4CL 蛋白序列，作为参考序列。从 Pfam 数据库下载 4CL 基因家族保守结构域隐马尔可夫模型 HMM 文件(PF00501)[6]。从牛蒡数据库(http://210.22.121.250:41352/)下载牛蒡基因组数据[8]。首先，利用拟南芥的 4CL 蛋白序列和牛蒡基因组蛋白序列使用 TBtools Ⅱ[9]中的 TBtools Blast Compare Two Seqs 和 Simple HMM Search 得到候选基因集 1(51 条基因)和候选基因集 2(55 条基因)。将 2 个基因集中的序列合并去重后，结合前期课题组筛选出与牛蒡苷积累相关基因[10]，筛选出 4CL 家族成员命名为 Al4CL1 ~ Al4CL8(插页Ⅵ图 1)。

2.2 Al4CL 基因克隆

牛蒡种子总 RNA 的提取方法参考上海生工 Trizol 提取试剂盒说明书操作，反转录成 cDNA 参照上海生工 M - MuLV Reverse Transcriptase，RNase H - 反转录试剂盒说明书操作，并保存在 - 20 ℃冰箱备用。在牛蒡数据库[8]下载得到 8 个 Al4CL 家族成员 CDS(编码序列，coding sequences)，利用 Primer 5 软件设计各家族成员的扩增引物(表 1)，以 cDNA 作为模板，利用 LA Taq DNA 聚合酶反应试剂盒进行 PCR。扩增体系(25 μL)：2 × GC Buffer I 12.5 μL，F (10 μmol/L) 0.5 μL，R (10 μmol/L) 0.5 μL，dNTP(10 mmol/L) 0.2 μL，ddH_2O 10.1 μL，cDNA 1 μL，Taq 酶(5 U/μL)0.2 μL。扩增程序：95 ℃ 预变性 3 min，94 ℃变性 30 s，58 ℃退火 30 s，72 ℃延伸 2 min，共 33 个循环；72 ℃延伸 7 min，4 ℃保存。PCR 产物 1% 琼脂糖凝胶电泳观察，将目的条带参照上海生工柱式 DNA 胶回收试剂盒说明书进行回收。将回收产物连接到 pMD18 - T 载体上，转入感受态细胞(SK2301)，涂布在预先用 20 μL 100 mmol/L 异丙基 - β - D - 硫代半乳糖苷(Isopropylβ - D - thiogalactoside，IPTG)和 100 μL 20 mg/mL 5 - 溴 - 4 - 氯 - 3 - 吲哚基 - β - 吡咯 - D - 半乳糖苷(X - gal)涂布的氨苄青霉素平板中，倒置培养过夜。菌液 PCR 鉴定后，挑选阳性克隆送至上海生物工程技术服务有限公司测序。

表 1 本实验所使用的 PCR 引物

基因名称	引物序列(5’ ~3’)
Al4CL1	F:ATGGAGCCGGAAAACGAGAT; R:TTAATTTGGAACGCCGGCTGCA
Al4CL2	F:ATGGAGCCGGAAAACGAGAT; R:TTAATTTGGAACGCCGGCTGCA
Al4CL3	F:ATGGCTACATTCTCCAAAACCCA; R:GGGTTAAAGTCTTGAATATAATTTCTCC
Al4CL4	F:ATGGCGGATCTCTCTCCCTAC; R:AATCAGAGTTTGGATTGAAGTTTCT
Al4CL5	F:ATGGCCAACACACAAAAAGACA; R:TCATTGAGATTTGGTAATCTTGTCC
Al4CL6	F:ATGGCGGATCTCTCTCCCTAC; R:AATCAGAGTTTGGATTGAAGTTTCT
Al4CL7	F:ATGGAGCCAACGTCGGAGACC; R:TCAATTCGGAACACCAGCTGCT
Al4CL8	F:ATGAGTAATCAACCACAAGTCAAAAAC; R:TTACAATCTAGAGGTTGCAAGTTGTATGA

2.3 Al4CL 蛋白理化性质及 Al4CL 基因染色体定位分析

通过上海纽普生物科技有限公司在线工具(https://www.novopro.cn/tools/translate.html)将 DNA 序列翻译为氨基酸序列。利用 Expasy 在线网站(https://www.expasy.org)预测 Al4CL 蛋白的相对分子质量、等电点和蛋白质的亲疏水性等理化性质。利用亚细胞定位 WoLF PSORT(https://wolfpsort.hgc.jp)预测 Al4CL 蛋白亚细胞定位，信号肽预测网站 SignalP - 4.1(https://services.heAlthtech.dtu.dk/services/SignAlP - 4.1/)分析信号肽，跨膜结构分析(https://services.

heAlthtech. dtu. dk/services/TMHMM - 2. 0/)网站分析跨膜结构。根据牛蒡基因组注释文件,获取 Al4CL1 ~ Al4CL8 基因在染色体的位置信息,使用 TBtools Ⅱ[9] 中 Gene Location Visualize from GTF/GFF 程序进行染色体位置映射,One Step McscanX 程序分析牛蒡基因组内 Al4CL 基因家族成员间复制关系,并通过 Advance Circos 模块对复制关系进行展示。

2.4 Al4CL 基因系统发育关系分析

从 TAIR 网站下载拟南芥 *Arabidopsis thaliana* 的 4CL 蛋白序列,从 NCBI(https://www. ncbi. nlm. nih. gov)下载杨树 *Populus przewalskii*、大豆 *Glycine max*、向日葵 *Helianthus annuus*、刺苞菜蓟 *Cynara cardunculus*、莴苣 *Lactuca sativa*、小蓬草 *Erigeron canadensis*、水稻 *Oryza sativa* 的 4CL 蛋白序列,利用 MAGA 11. 0 软件对牛蒡与其他物种的 4CL 蛋白进行系统发育关系分析,采用邻接法(neighbor - joining, NJ)构建进化树(boostrap = 1000)。随后,用 Evolview(https://www. evolgenius. info/evolview - v2/#login)对系统发育树进行可视化。

2.5 Al4CL 基因保守基序、基因结构及蛋白结构域分析

使用 MEME(https://meme - suite. org/meme/)在线网站分析 Al4CL 家族氨基酸保守基序,将 motif 查找数量定为 13。通过 TBtools Ⅱ[9] 将保守基序结果进行可视化。通过 NCBI - CDD(https://www. ncbi. nlm. nih. gov/cdd/term =)网站分析 Al4CL 家族 Domain,通过 TBtools Ⅱ[9] 中 Gene Structure View 程序将蛋白结构域等可视化。

2.6 Al4CL 基因顺式作用元件预测

利用 TBtools Ⅱ[9] 中 Gtf/Gff3 Sequences Extract 获取 Al4CL 家族成员上游 2000 bp 序列作为假定的启动子区域,利用 PlantCARE(https://bioinformatics. psb. ugent. be/webtools/plantcare/html/)进行启动子顺式作用元件的预测,利用 TBtools Ⅱ[9] 中 Simple Biosequence Viewer 作图,展示启动子区域的顺式作用元件。

2.7 Al4CL 基因编码氨基酸序列比对分析

利用 MEGA 11. 0 软件将克隆得到的 Al4CL 与拟南芥 At4CL1(AT1G51680. 1)、拟南芥 At4CL2(AT3G21240. 1)、拟南芥 At4CL3(AT1G65060. 1)、拟南芥 At4CL4(AT3G21230. 1)的氨基酸序列进行比对,分析功能域,使用 iTOL(https://itol. embl. de)对多序列比对进行可视化。

2.8 Al4CL 基因共线性分析

利用 TBtools Ⅱ[9] 中 One step MCScan X 和 Multiple Synteny Plot 程序分析牛蒡、拟南芥和刺苞菜蓟 4CL 基因间的共线性。

2.9 Al4CL3 的原核表达分析

参照李胜男的研究[11],选取与木脂素和牛蒡苷生物合成高度相关的 Al4CL3 的序列,截取高度保守的膜外肽区,经密码子优化后进行亚克隆,命名为 chAl4CL3,合成 pCold Ⅱ - chAl4CL3 质粒,经过 BamH Ⅰ、Sal Ⅰ酶切连接到 pET - 28a(+)原核表达载体上,命名为 pET - 28a - chAl4CL3,将重组质粒 pET - 28a - chAl4CL3 进一步转化到大肠杆菌感受态细胞 Rosetta(DE3)培养,诱导表达,收集菌体,纯化蛋白,最后通过 SDS - PAGE 以及 Western - Blot 检测重组蛋白。

3 结果

3.1 牛蒡 4CL 基因克隆

PCR 产物经回收、克隆和菌落 PCR 鉴定,获得 8 条特异性较强,序列大小在 1500 ~ 2000 bp 的基因片段(插页Ⅵ图 2A)。对其阳性克隆进行测序的结果显示,PCR 所得开放阅读框(Open Reading Frame, ORF)的长度为 1596 ~ 1764 bp,与预期目的基因片段大小相符,利用牛蒡数据库(http://210. 22. 121. 250:41352/)Blast 程序进行比对,同源性除 Al4CL5(NCBI 登录号:PP781509. 1)为 96%(插页Ⅵ图 2B),Al4CL8 有 5 个变异位点,分别在 263、807、955、1102、1487 存在 C - T、G - A、G - A、G - A、C - T 变异,其他序列与预测序列一致性均为 100%。

3.2 Al4CL 蛋白理化性质分析

Al4CL1 ~ Al4CL8 基因序列的阅读框长度为 1596 ~ 1764 bp,编码氨基酸数目介于 531 ~ 587,最大的为 Al4CL4,最小的为 Al4CL6;编码蛋白质分子量介于 58. 45 ~ 63. 70 kD,平均分子量为 60. 74 kD,理论等电点 pI 介于 5. 33 ~ 8. 85,平均值为 6. 95,不稳定系数介于 32. 55 ~ 46. 24(大于 40 为不稳定),Al4CL1、Al4CL2、Al4CL7 为疏水性蛋白质,Al4CL3、Al4CL4、Al4CL5、Al4CL6、Al4CL8 为亲水性蛋白质,脂肪系数介于 87. 84 ~ 101. 43。Al4CL3、Al4CL6、Al4CL8 蛋白含有 2 个跨膜螺旋结构。预测结果表明 Al4CL2 ~ Al4CL8 蛋白可能含有信号肽,可能为分泌蛋白。亚细胞定位可能定位于细胞质、质膜、叶绿体和内质网。

3.3 Al4CL 基因染色体定位分析

Al4CL 家族成员染色体定位分析表明,Al4CL1、Al4CL2 位于 Hic_asm_1 染色体,Al4CL4 位于 Hic_asm_2 染色体,Al4CL5、Al4CL8 位于 Hic_asm_3 染色体,Al4CL6 位于 Hic_asm_9 染色体,Al4CL3 位于 Hic_asm_10 染色体,Al4CL7 位于

表 2 Al4CL 理化性质、跨膜结构域、信号肽和亚细胞定位预测

基因	基因号	阅读框 ORF/(bp)	氨基酸数目/(aa)	分子量/(kD)	等电点	不稳定系数	总平均亲水性	脂肪酸系数	跨膜结构域	亚细胞定位	信号肽
Al4CL1	evm. TU. 000019F. 176	1656	551	59. 79	8. 85	46. 24	0. 022	91. 98	0	内质网	无
Al4CL2	evm. TU. 000050F. 503	1623	540	59. 02	5. 33	34. 69	0. 052	99. 67	0	叶绿体	有
Al4CL3	evm. TU. 000137F. 68	1713	570	62. 79	8. 58	39. 15	-0. 050	87. 84	2	质膜	有
Al4CL4	evm. TU. 000201F. 2	1764	587	63. 55	6. 16	35. 93	-0. 006	94. 47	0	细胞质	有
Al4CL5	evm. TU. 000117F. 44	1638	545	59. 86	5. 90	36. 40	-0. 093	98. 37	0	细胞质	有
Al4CL6	evm. TU. 000039F. 388	1596	531	58. 45	8. 43	42. 11	-0. 046	97. 70	0	质膜	有
Al4CL7	evm. TU. 000054F. 293	1620	539	58. 77	5. 49	32. 55	0. 098	101. 43	0	叶绿体	有
Al4CL8	evm. TU. 000058F. 165	1755	584	63. 70	6. 85	36. 07	-0. 108	89. 95	2	质膜	有

Hic_asm_14 染色体(插页Ⅶ图 3A)。

基因复制分析结果显示,Al4CL 家族存在两对串联复制基因,Al4CL2 和 Al4CL7 互为串联复制基因,Al4CL6 和 Hic_asm_8 染色体上的其他基因互为串联复制基因(插页Ⅶ图 3B)。

3.4 Al4CL 基因结构、保守基序及蛋白结构域分析

对 Al4CL 家族成员进行蛋白保守基序研究,鉴定得到 13 个保守基序(插页Ⅶ图 4)。除 Al4CL1 缺失 motif 8 外,这些序列均包含 motif 1 ~ motif 10。除此以外,motif 11 和 motif 13 仅在 Al4CL2、Al4CL4、Al4CL7 存在,motif 12 仅在 Al4CL1 和 Al4CL3 存在。Al4CL 家族的基因外显子数目集中在 5、6 个,非翻译区(Untranslated Region,UTR)为 0 ~ 2 个,分布存在一定差异。Al4CL 基因家族成员中共有 3 个保守结构域,其中 Al4CL1、Al4CL6、Al4CL8 具有 4CL 结构域,Al4CL2、Al4CL4、Al4CL7 具有 PLN02246 结构域,Al4CL3、Al4CL5 具有 AFD_class_ I superfamily 结构域,所有序列的结构域均为单个结构域。

3.5 Al4CL 基因启动子顺式作用元件分析

通过 PlantCARE 在线网站分析 Al4CL 家族基因启动子上的顺式作用元件及其响应激素和非生物胁迫的功能,参考 LIU 等[12]的研究,将预测的顺式作用元件归为发育响应(development responsive)、胁迫响应(stress responsive)、激素响应(hormone responsive)和光响应(light responsive)4 种类型。大多数的 Al4CL 启动子区域都具有这 4 种类别的顺式作用元件,但是这些序列包含的顺式作用元件的种类差异比较明显(插页Ⅷ图 5),预测到的启动子顺式作用元件分析结果提示 Al4CL 在牛蒡生长发育、抵御环境胁迫中可能参与具体的生理过程,也预示这些序列在牛蒡体内不同的生理功能。

3.6 Al4CL 基因编码氨基酸多序列比对与系统发育关系分析

氨基酸序列比对分析显示(插页Ⅷ图 6A),Al4CL2、Al4CL4、Al4CL7 蛋白包含保守 Box Ⅰ(SSGTTGLPKGV)和 Box Ⅱ(GEICIRG)的结构域,Box Ⅰ作为腺苷酸甲基化酶结合功能域,直接参与到催化反应的核心过程,Box Ⅱ在 4CL 蛋白家族中展现出显著的序列保守性,尽管它并不直接参与到催化反应的主体环节,但缺失可能导致蛋白质构象发生不可逆改变,从而严重影响其催化活性。而其他不同成员在这两处的序列发生变异,推测其功能可能也存在差异。除已知的 4CL 蛋白保守序列 Box Ⅰ、Box Ⅱ,牛蒡 4CL 蛋白还存在 3 段保守序列即 GWLHTG 为腺苷 - 1 - 磷酸结合部位、FIVDRLKELKYKGFQVAPAE 为辅酶 A 结合部位和 SGKILRKDL 为 4CL 活性部位和辅酶 A 结合部位[13],Al4CL2、Al4CL7 具有这 3 个保守序列。其他家族成员也都存在与之较为相似的保守序列,这些保守基序和序列对于 Al4CL 蛋白催化活性或催化特异性至关重要。

进化分析表明(插页Ⅷ图 6B),8 个 Al4CL 成员聚为 3 个分支,Al4CL2、Al4CL7 与拟南芥 At4CL1、At4CL2、At4CL4 和杨树 Pto4CL3 等参与木脂素、木质素生物合成的 4CL 聚为一支,为 Group Ⅰ亚家族;Al4CL4 与拟南芥 At4CL3 及杨树 Pto4CL2 等参与类黄酮生物合成的蛋白序列聚为一支,为 Group Ⅱ亚家族;而 Al4CL1、Al4CL3、Al4CL5、Al4CL6、Al4CL8 与拟南芥等亲缘关系较近,为 Group Ⅲ,属于 4CL - like 基因。

3.7 Al4CL 基因共线性分析

Al4CL 与刺苞菜蓟和拟南芥基因序列均具有同源共线性关系(插页Ⅸ图 7),与同科植物刺苞菜蓟之间有 10 个同源基因对,与拟南芥之间仅有 3 个同源基因对。

3.8 Al4CL3 原核表达

(1)含有 chAl4CL3 基因的原核表达质粒的构建与鉴定。

用 BamH Ⅰ、Sal Ⅰ分别双酶切质粒 pCold Ⅱ - chAl4CL3 与原核表达载体 pET - 28a(+),将目的基因 chAl4CL3 连接到表达载体得到重组质粒 pET - 28a - chAl4CL3(图 8A、8B),进行酶切鉴定,可见约 5000 bp 的载体片段和约 750 bp 的目的片段(chAl4CL3)(图 8C),与理论值一致,表明质粒构建成功。将质粒 pET - 28a - chAl4CL3 转化到 Rosetta(DE3)感受态细胞中。

(2)chAl4CL3 蛋白的原核表达、鉴定和纯化。

Al4CL3 蛋白膜外结构域片段 chAl4CL3 蛋白的分子质量约为 31 kD,将含有裂解缓冲液的细胞破碎物离心后,取菌体沉淀(即包涵体)与上清蛋白样品 chAl4CL3 进行融合蛋白表达小试 SDS - PAGE(插页Ⅸ图 9A、9B)。最终纯化蛋白进行 SDS - PAGE 及 Western - blot 分析。结果表明,目的蛋白主要以包涵体形式表达,在 31 kD 左右有相应的目的条带,与预期一致,表明目的蛋白成功表达。纯化后进行 SDS - PAGE 及 Western - blot 分析(插页Ⅸ图 9C、9D),有相对单一的目的条带,表明目的蛋白已被纯化成功。

4 讨论

木脂素和木质素是植物中由苯丙烷途径产生的一类重要的次生代谢物,4CL 基因在维管束植物中普遍存在,处于从一般苯丙烷代谢到几个主要分支途径如黄酮、香豆素、木质素、木脂素等的分化点,可催化肉桂酸及其羟基或甲氧基衍生物生成相应的辅酶 A 酯,这些中间体进入支链苯丙烷衍生物合成途径,生成最终的化合物单体,该基因也是苯丙烷代谢途径产物多样化的一个关键基因[14-19]。本研究对牛蒡全基因组中鉴定出 8 个 4CL 家族成员进行克隆鉴定、生物信息学和原核表达分析,为牛蒡子药材质量的形成分子机制研究提供基础。

Al4CL1 ~ Al4CL8 的扩增、测序结果与预测序列一致性达到 95% 以上。生物信息学分析显示,Al4CL 基因家族氨基酸保守基序研究发现 Al4CL2、Al4CL7 共含有 5 段保守序列:SSGTTGLPKGV (Box Ⅰ)、GEICIRG (Box Ⅱ)、GWLHTG、FIVDRLKELKYKGFQVAPAE、SGKILRKDL,并且这 5 条序列分别存在 motif 5(21 ~ 31)、motif 2(27 ~ 33)、motif 1(4 ~ 9)、motif 1(21 ~ 41)、motif 3(33 ~ 41)中。基序为基因中保守的部分,基序的改变可能对基因的生物学功能具有重要影响,这也与物种进化相一致[14]。基于牛蒡、拟南芥、水稻和刺苞菜蓟等 8 种植物的 4CL 基因家族的系统发育分析,8 个 Al4CL 成员分布在 3 个亚家族。4CL 基因家族的不同类别成员在植物次生代谢中承担特定且分工明确的角色,Group Ⅰ亚家族 4CL 主要参与植物中木脂素和木质素化合物的生物合成,Group Ⅱ亚家族 4CL 主要参与并调控类黄酮化合物的生物合成,Ⅲ类 4CL 为4CL - Like,具体作用尚未确定,Al4CL2 和 Al4CL7 作为 Group Ⅰ亚家

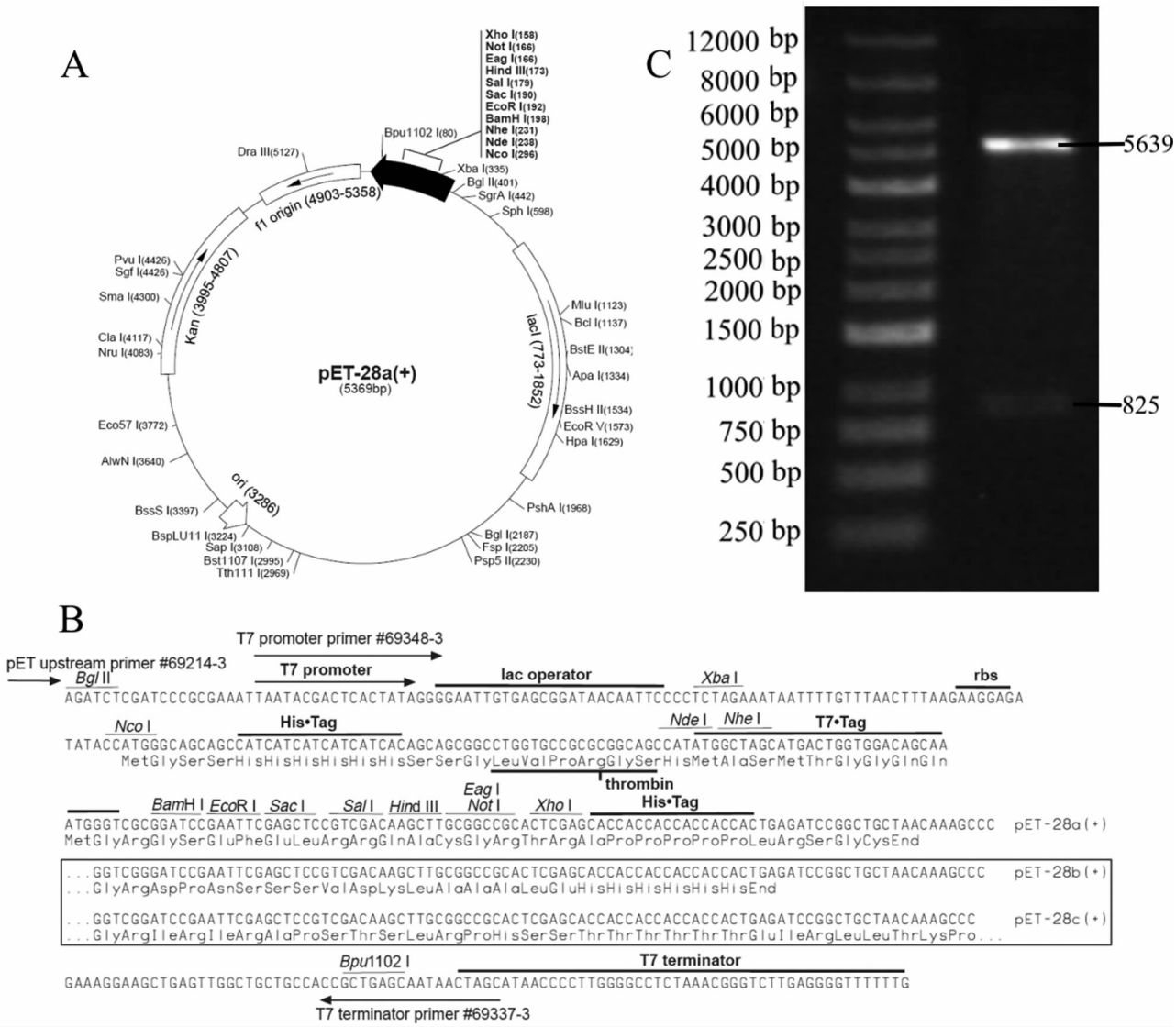


图8 载体信息(A、B)及重组质粒酶切检测(C)

族成员,其潜在功能可能与木脂素、木质素等生物合成有关,而Al4CL4作为Group Ⅱ亚家族成员,其潜在功能可能与类黄酮类等生物合成有关。牛蒡与刺苞菜蓟等菊科植物遗传距离相对较近,与水稻的遗传距离较远,推测原因可能为4CL主要以基因家族的形式存在,其功能与植物的进化过程密切相关[20]。

Al4CL启动子区域包括发育响应(development responsive)、压力响应(stress responsive)、激素响应(hormone responsive)和光响应(light responsive)4种类型,光响应元件在4CL基因家族成员的启动子区域是最普遍的,共有81个。此外,激素响应元件的数量有61个,压力响应元件的数量有41个,发育响应元件的数量有15个。激素响应元件中包含了赤霉素响应元件(8个)、水杨酸响应元件(5个)、脱落酸响应元件(19个)、茉莉酸甲酯响应元件(24个)等。发育响应元件中包含了参与玉米醇溶蛋白代谢调控元件(4个)、参与分生组织表达的调控元件(2个)、参与胚乳表达的顺式调控元件(4个)、昼夜节律调控元件(2个)等。压力响应元件中包含了厌氧诱导所必需的顺式调控元件(31个)、参与低温响应调控元件(3个)、参与干旱响应调控元件(1个)。启动子区域大量的顺式作用元件,对4CL基因的转录和表达具有显著影响。SUN S C等[21]发现过度表达Gh4CL7的拟南芥对干旱处理更具耐受性,可能是由于木质素含量的增加,进而积极应对干旱压力,LI Y等[22]发现在低温胁迫下,河北梨花中与类黄酮相关途径中的两个4CL基因表达上调。YAO P等[23]采用茉莉酸甲酯(methyl jasmonate,MeJA)处理苦荞Fagopyrum tataricum,发现Ft4CL基因表达水平显著上调,从而促进芦丁的积累。这些结果表明,Al4CL的表达与激素以及外界环境胁迫等有关,进而催化并合成具有抵抗逆境功能的类黄酮、木脂素及木质素等多种次生代谢物质。

原核表达分析是验证基因能否体外表达的重要方法[24-26]。YANG Y等[10]研究表明Al4CL3与牛蒡子牛蒡苷生物合成密切相关,所以本研究对该基因进行原核表达分析,验证其是否可以正常表达。在密码子优化的前提下,成功转入原核表达载体pET-28a(+)中,并表达出重组蛋白31 kD,获得纯化蛋白,为后续酶活力研究和分子功能验证提供基础。本研

究探明了 Al4CL 家族的基因序列和基因结构特性，为进一步研究牛蒡果实发育过程中该基因的作用和功能等提供基础数据。

5 结论

本研究基于牛蒡基因组数据，克隆 8 条 Al4CL 基因，通过生物信息学分析表明 Al4CL 家族成员的理化性质、结构、保守区域等具有一定相似性，同时也具有很大的特异性。Al4CL 家族成员不均匀地分布于 1、2、3、9、10 和 14 号染色体上。氨基酸序列中存在 4CL 氨基酸保守域，Al4CL 启动子区域存在多个顺式作用元件，受多种激素或胁迫诱导，共线性分析表明 Al4CL 家族成员与刺苞菜蓟之间存在更近的亲缘关系，进化树分析发现 Al4CL 家族可分为 3 个亚家族，同一亚家族的 4CL 成员基因结构、蛋白保守基序较为相似，其中 Al4CL2、Al4CL7 属于 4CL Group Ⅰ可能具有调控木脂素、木质素合成的功能。成功构建了 chAl4CL3 在大肠杆菌内的原核表达体系，并获得了纯化蛋白，提示该蛋白可以用于后续研究。为研究牛蒡子木脂素的生物合成及其品质的提高提供一定的理论依据，为深入研究 Al4CL 及其生物学功能提供参考。

参考文献

[1] 国家药典委员会. 中华人民共和国药典[S]. 北京：中国医药科技出版社，2020：73.

[2] BOKELMANN J M. Burdock (*Arctium lappa*) [J]. Medicinal Herbs in Primary Care, 2022(2):253 - 256.

[3] 李逸，徐欢欢，张宇辰，等. 大葱 4 - 香豆酸辅酶 A 连接酶基因 *Af4CL* 的克隆及表达分析[J]. 农业生物技术学报，2023，30(2):273.

[4] LAVHALE S G, KALUNKE R M, GIRI A P. Structural, functional and evolutionary diversity of 4 - coumarate - CoA ligase in plants [J]. Planta, 2018, 245(5):1063 - 1078.

[5] 张苛苛，谭宇萍，徐欢欢，等. 菘蓝 4 - 香豆酸：辅酶 A 连接酶全基因家族的鉴定及表达分析[J]. 中国中药杂志，2024，49(2)：361 - 369.

[6] 郑飞雄，陈俊宇，江林琪，等. 掌叶覆盆子 4CL 基因家族鉴定及表达分析[J]. 农业生物技术学报，2024，32(2):311 - 321.

[7] 周如月. 灵芝 4 - 香豆酸辅酶 A 连接酶基因(4CL)的功能分析[D]. 济南：山东农业大学，2024.

[8] SONG Y, YANG Y, XU L, et al. The burdock database: a multi - omic database for *Arctium lappa*, a food and medicinal plant[J]. BMC Plant Biol, 2023, 23: 86.

[9] CHEN C, WU Y. TBtools - II: A "one for all, all for one" bioinformatics platform for biological big - data mining[J]. Molecular Plant, 2023, 16: 1733 - 1742.

[10] YANG Y, LI S, XING Y, et al. The first high - quality chromosomal genome assembly of a medicinal and edible plant *Arctium lappa* [J]. Molecular Ecology Resources, 2022, 22(4): 1493 - 1507.

[11] 李胜男. 不同生长时期牛蒡果实中牛蒡苷积累规律及相关代谢基因分析[D]. 沈阳：辽宁中医药大学，2022.

[12] LIU X, GONG Q, ZHAO C, et al. Genome - wide analysis of cytochrome P450 genes in *Citrus clementina* and characterization of a CYP gene encoding flavonoid 3' - hydroxylase[J]. Hortic Res, 2023, 10(2):c283.

[13] 王星淇，孙雪丽，车婧如，等. 两个非洲菊 4CL 基因的克隆及表达[J]. 应用与环境生物学报，2020，26(2):272 - 279.

[14] LI S, CHIU T Y, JIN X, et al. Integrating genomic and multiomic data for *Angelica sinensisprovides* insights into the evolution and biosynthesis of pharmaceutically bioactive compounds[J]. Communications Biology, 2023, 6(1):1198.

[15] WANG W, LI Y, CAI C, et al. Auxin response factors fine - tune lignin biosynthesis in response to mechanical bending in bamboo [J]. New Phytologist, 2024, 241:1161 - 1176.

[16] WANG X, ZHANG J, LIU Z, et al. Comparative transcriptome analysis of three chrysanthemums provides insights into flavonoid and terpenoid biosynthesis[J]. Journal of Plant Biology, 2021, 64(5): 1 - 13.

[17] SHAN C, ZHAO L, SHI Y, et al. Transcriptome analyses reveal the expression profile of genes related to lignan biosynthesis in *Anthriscus sylvestris* L. Hoffm. Gen[J]. Physiology and Molecular Biology of Plants, 2022, 28(2): 333 - 346.

[18] ZHANG L, CHEN J, ZHOU X, et al. Dynamic metabolic and transcriptomic profiling of methyl jasmonate - treated hairy roots reveals synthetic characters and regulators of lignan biosynthesis in *Isatis indigotica* Fort[J]. Plant biotechnology journal, 2016, 14(12): 2217 - 2227.

[19] SHEN S, TONG Y, LUO Y, et al. Biosynthesis, total synthesis, and pharmacological activities of aryltetralin - type lignan podophyllotoxin and its derivatives[J]. Natural product reports, 2022, 39(9): 1856 - 1875.

[20] CUKOVICA D, EHLTING J, ZIFFLE J A V, et al. Structure and evolution of 4 - coumarate: coenzyme A ligase (4CL) gene families [J]. Biological Chemistry, 2001, 382(4):645 - 654.

[21] SUN S C, XIONG X P, ZHANG X L, et al. Characterization of the *Gh4CL* gene family reveals a role of *Gh4CL7* in drought tolerance [J]. BMC plant biology, 2020, 20: 1 - 15.

[22] LI Y, ZHANG J, WANG S, et al. Integrative transcriptomic and metabolomic analyses reveal the flavonoid biosynthesis of *Pyrus hopeiensis* flowers under cold stress[J]. Horticultural Plant Journal, 2023, 9(3): 395 - 413.

[23] YAO P, HUANG Y, DONG Q, et al. FtMYB6, a light - induced SG7 R2R3 - MYB transcription factor, promotes flavonol biosynthesis in tartary buckwheat (*Fagopyrum tataricum*)[J]. Journal of Agricultural and Food Chemistry, 2020, 68(47): 13685 - 13696.

[24] WEI H, XU C, MOVAHEDI A, et al. Characterization, expression profiling, and functional analyses of a 4CL - like gene of *Populus trichocarpa*[J]. Processes, 2019, 7(1): 45.

[25] LI M, GUO L, WANG Y, et al. Molecular and biochemical characterization of two 4 - coumarate: CoA ligase genes in tea plant (*Camellia sinensis*)[J]. Plant molecular biology, 2022, 109(4): 579 - 593.

[26] ZHOU X, CAO J, LIU X M, et al. Cloning and functional analysis of *Gb4CL1* and *Gb4CL2* from *Ginkgo biloba*[J]. The Plant Genome, 2024(24): e20440.

牛蒡 4CL 基因家族的克隆鉴定与生物信息学分析

(正文见 20 – 25 页)

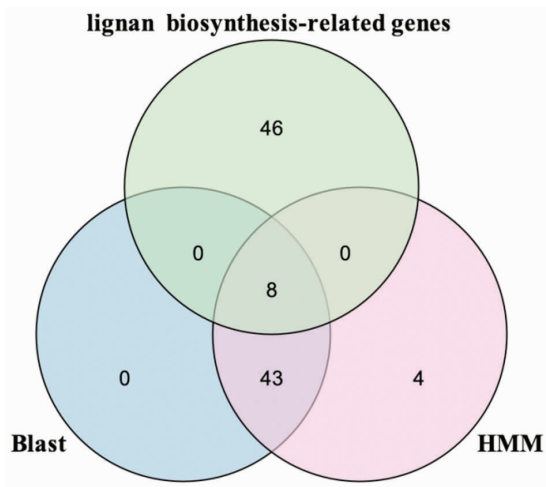


图 1 牛蒡 4CL 基因的筛选



图 2 Al4CL 基因的 ORF 的 PCR(A)及 Al4CL5 的同源性比对(B)

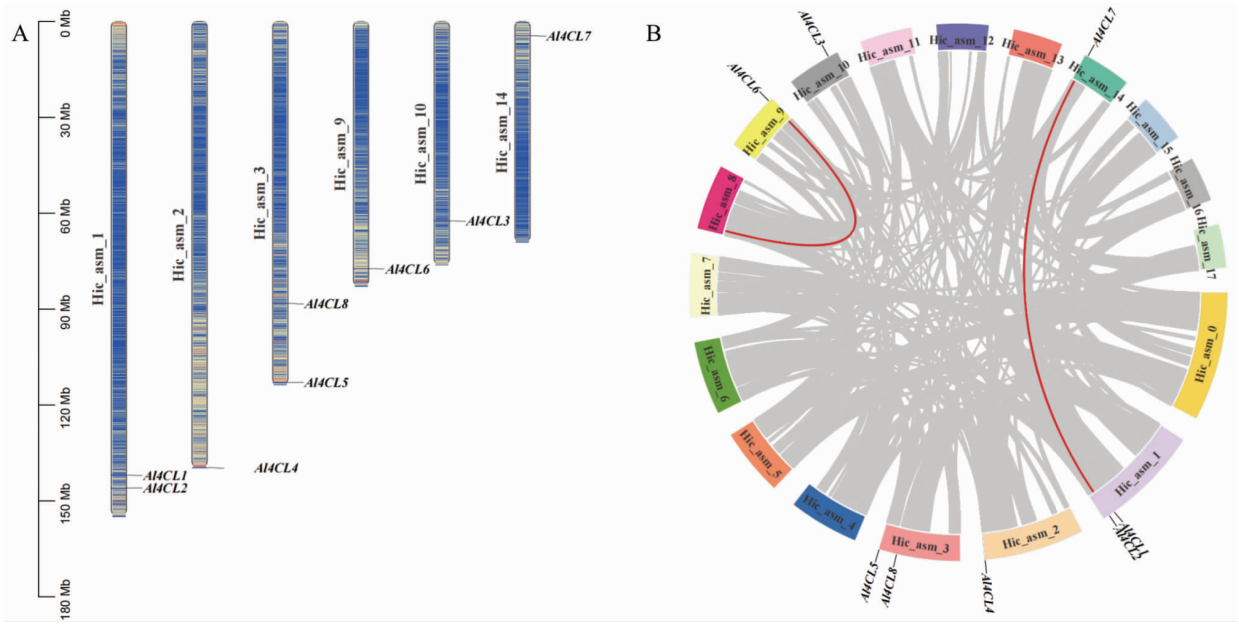


图 3 At4CL 染色体定位(A)及复制事件(B)

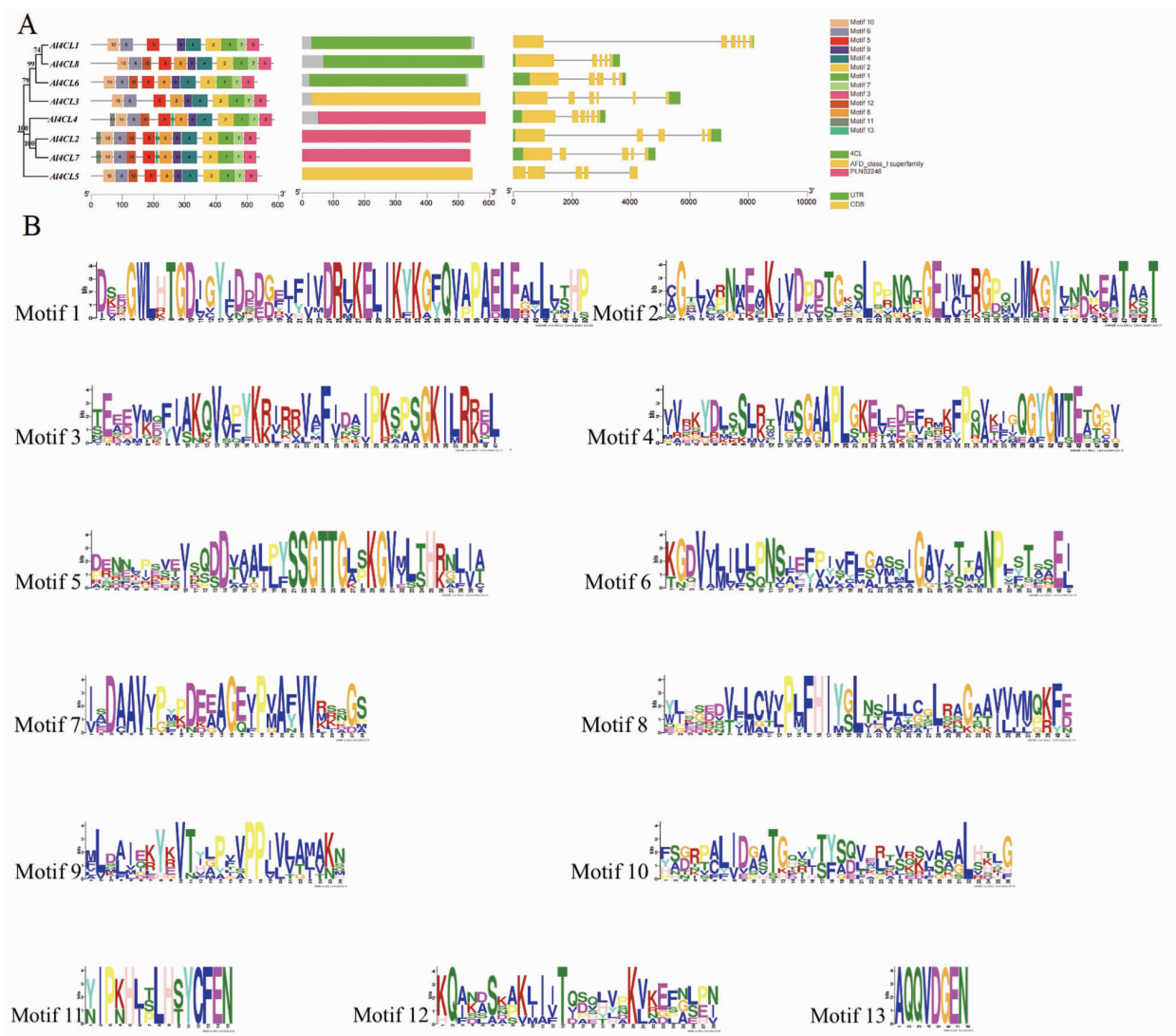


图 4 At4CL 结构与保守结构域分布分析(A)及基序标识(B)

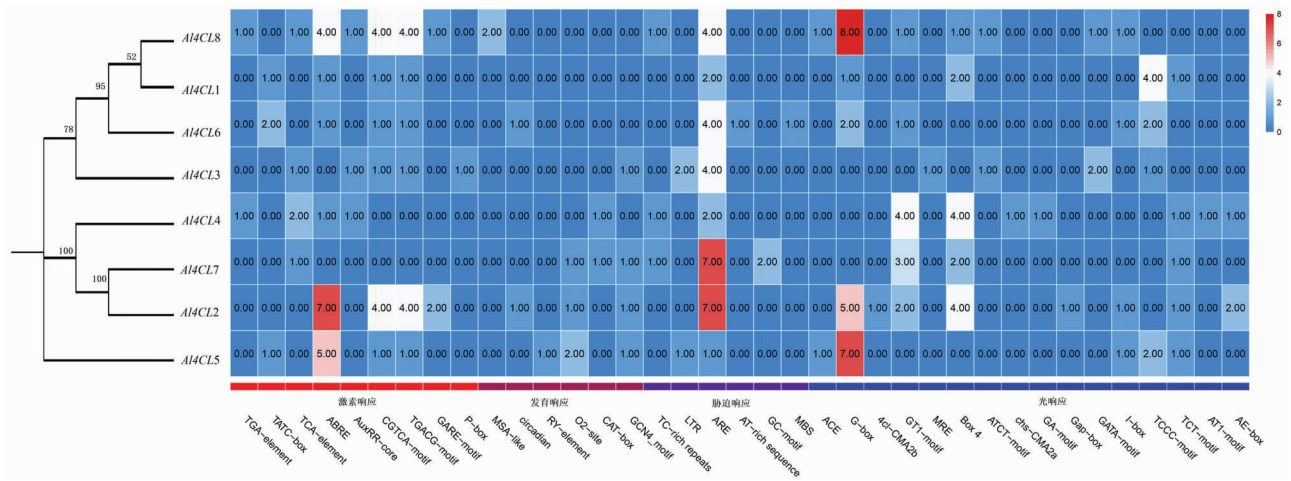


图 5　Al4CL 启动子顺式作用元件预测

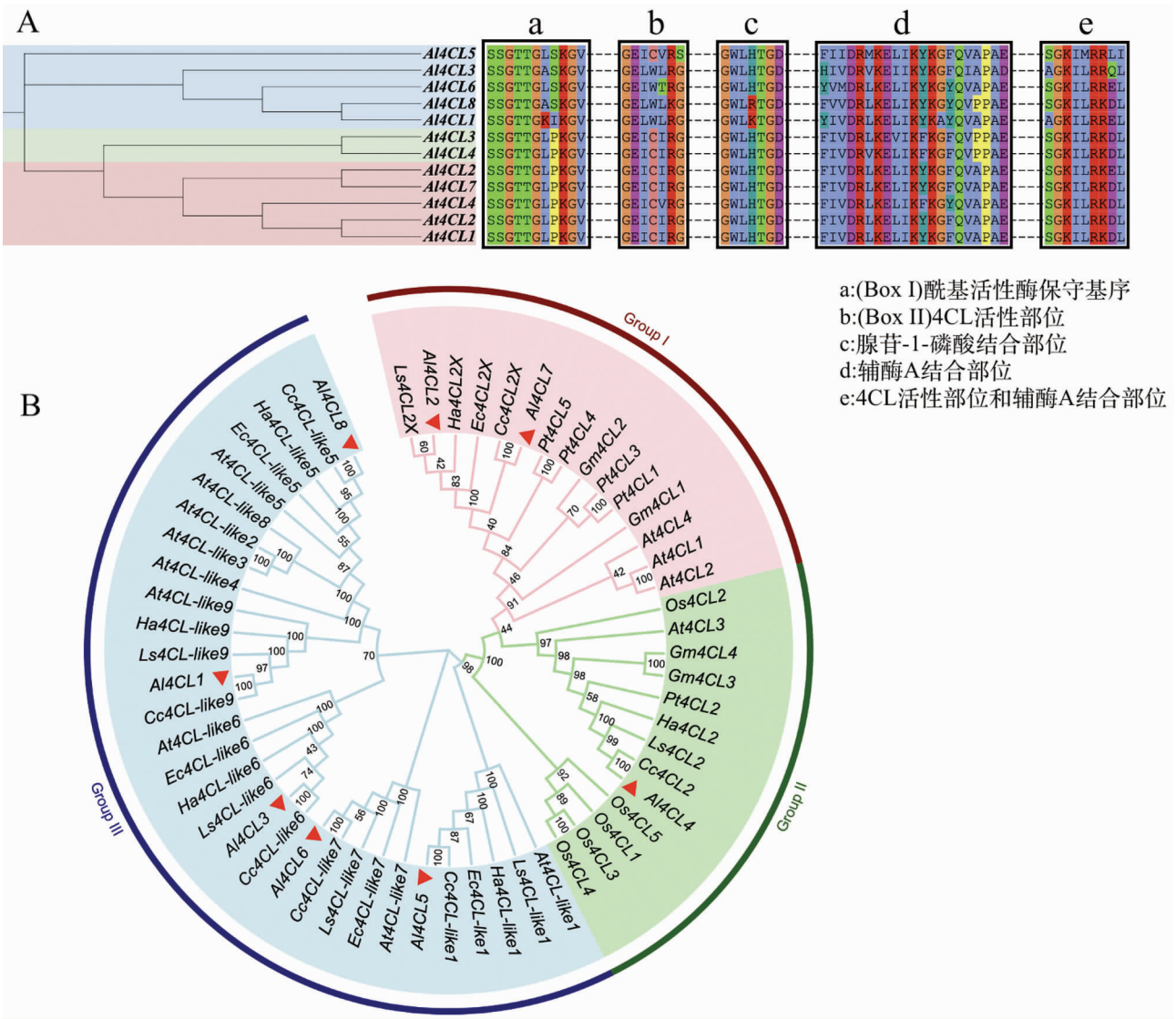


图 6　Al4CL 氨基酸序列比对(A)及 Al4CL 家族的系统进化分析(B)

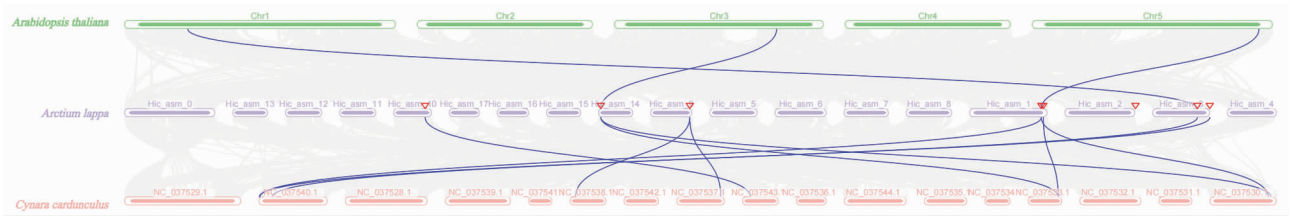
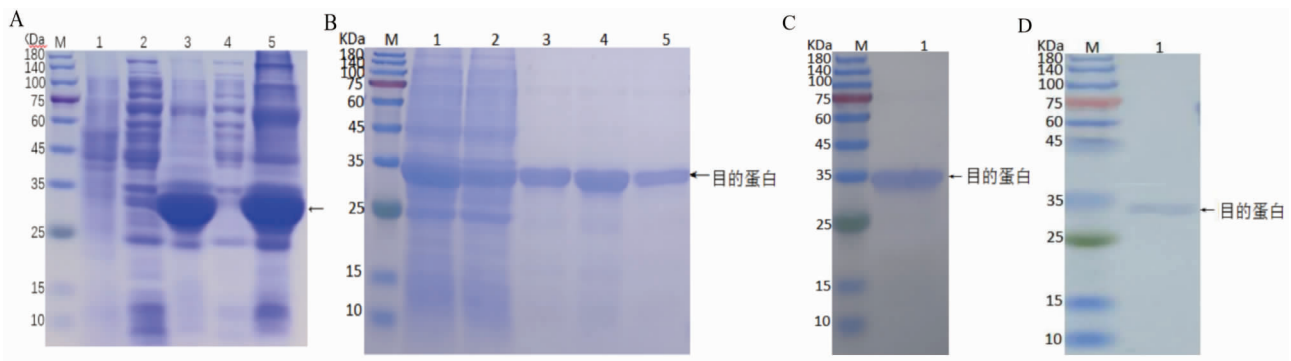


图 7　牛蒡与刺苞菜蓟、拟南芥 4CL 基因的共线性分析



注：A. 融合蛋白表达小试 SDS－PAGE 分析图，M 为 Protein Marker；1 为诱导前总蛋白；2 为 20 ℃上清；3 为 20 ℃沉淀；4 为 37 ℃上清；5 为 37 ℃沉淀；B. 融合蛋白镍琼脂糖亲和层析纯化 SDS－PAGE 分析图，M 为 Protein marker；1 为上样；2 为流出；3 为 20 mmol/L Imidazole 洗脱组分；4 为 50 mmol/L Imidazole 洗脱组分；5 为 500 mmol/L Imidazole 洗脱组分；C. 最终纯化蛋白 SDS－PAGE 分析图，M 为 Protein marker；1 为融合目的蛋白；D. 最终纯化蛋白 Western Blot 分析图，M 为 Protein marker；1 为融合目的蛋白。

图 9　chAl4CL 蛋白的原核表达结果

小儿定喘颗粒治疗呼吸道合胞病毒肺炎（痰热闭肺证）多中心随机对照临床研究

（正文见 1－6 页）

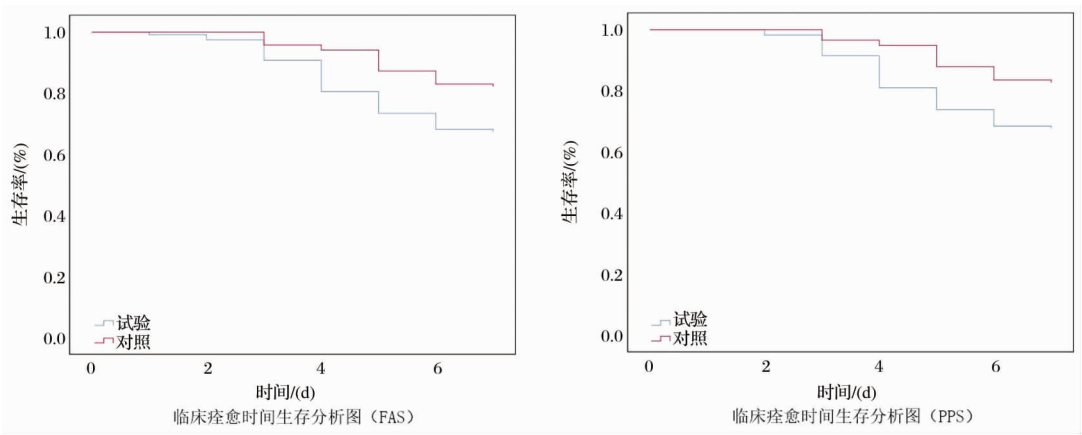


图 1　两组患儿临床痊愈时间生存分析图