

# 木质素生物合成酶 4CL 基因的遗传进化分析

黄胜雄, 胡尚连, 孙霞, 曹颖, 卢学琴, 蒋瑶

(西南科技大学 生命科学与工程学院, 四川 绵阳 621010)

**摘要** 【目的】4CL 是木质素生物合成途径中的关键合成酶, 通过对 4CL 基因的遗传进化分析, 为其研究和利用提供理论依据。【方法】利用生物信息学手段, 对 4CL 基因完整 cDNA 和编码氨基酸序列进行数据挖掘, 利用 Mega 软件、ExPASy 的 Prosite 数据库、NCBI 的 Conserved Domains 数据库, 进行 4CL 基因核酸和蛋白质水平上的遗传进化分析。【结果】构建了 4CL 基因的系统发生树, 揭示出植物中的 4CL 基因聚类与植物分类大体一致, 主要以基因家族的形式存在, 但基因家族中部分基因存在着结构和功能的分化, 4CL 基因中存在着 GC 含量偏高和偏低两大类基因。4CL 基因编码的氨基酸序列保守区高度相似, 但是在单、双子叶两大类植物之间, 其氨基酸平均含量存在着较大差异。【结论】在植物的进化过程中, 虽然 4CL 基因较保守, 但部分 4CL 基因发生了分化, 体现在基因的结构和功能方面, 在 4CL 基因的研究中必须充分考虑到这种分化。

**关键词** 木质素; 4CL 基因; 遗传进化分析

**中图分类号** Q343.1<sup>+</sup>2

**文献标识码** A

**文章编号** 1671-9387(2008)10-0199-08

## Genetic and evolutionary analysis of lignin biosynthase 4CL gene

HUANG Sheng-xiong, HU Shang-lian, SUN Xia, CAO Ying, LU Xue-qin, JIANG Yao

(Life Science and Engineering College, Southwest University of Science and Technology, Mianyang, Sichuan 621010, China)

**Abstract:** 【Objective】4CL (4-coumarate:CoA ligase) is an important biosynthase in the lignin biosynthesis. Through genetic and evolutionary analysis of 4CL gene, some theoretic foundation could be found for researches of 4CL gene. 【Method】With the bioinformational tools, the complete sequences of 4CL cDNA and corresponding amino acid have been found from the gene database. With the Mega, Prosite database of ExPASy, Conserved Domains database of NCBI, the genetic and evolutionary analysis was respectively carried out in nucleotides and proteins level. 【Result】The Phylogenetic tree of 4CL genes was constructed. The sorting of 4CL genes tree and the sorting of plants were mainly identical. There were 4CL family genes in the 4CL genes, in which the divergence of the construction and function were found. Two kinds of 4CL genes, which respectively contain high percents and low percents of GC contents, were found. The conserved amino acid blocks of individual 4CL gene were highly similar. But there were differences in the average amino acid contents between the monocots and dicots. 【Conclusion】In the process of plants evolution, some 4CL genes diverged, although most of 4CL genes were reserved. The divergence of structure and the function of 4CL gene were included. The divergence should be fully considered in the 4CL gene research.

**Key words:** lignin; 4CL gene; genetic and evolutionary analysis

\* [收稿日期] 2007-11-05

[基金项目] 四川省应用基础研究基金项目(05JY029-101); 四川省教育厅资助项目(2003031022); 四川省青年基金项目(041100); 西南科技大学重点项目(2006031030)

[作者简介] 黄胜雄(1983-), 男, 四川眉山人, 在读硕士, 主要从事植物遗传与品种改良研究。

[通讯作者] 胡尚连(1966-), 女, 河北秦皇岛人, 教授, 博士, 主要从事植物生理与生物技术研究。E-mail: hushanglian@126.com

木质素是地球上含量非常丰富的一类有机物质,仅次于纤维素的含量,其具有丰富的生物学功能,在抵御植物病害袭击、抗击外来侵袭、维持植物正常生长等方面发挥着巨大的作用<sup>[1]</sup>。

木质素在植物体内的生物合成途径目前尚不完全清楚,但普遍认为大致可分为以下 3 个主要阶段:①植物光合作用的同化产物通过莽草酸途径形成苯丙氨酸、酪氨酸;②生成 3 种木质素主要单体的苯丙氨酸途径;③ 3 种单体分别聚合形成 3 种不同的木质素<sup>[2-3]</sup>。其中第②阶段——苯丙氨酸途径包含了大量的酶类和生化反应,目前是木质素研究的重点。

4-香豆酸辅酶 A 连接酶(4-coumarate:CoA ligase, 4CL)是苯丙氨酸途径中的关键酶。在苯丙氨酸途径中,4CL 是联系木质素前体和各个分支反应的纽带,分别催化 p-香豆酸、咖啡酸、阿魏酸和芥子酸生成相应的 CoA 酯,促进木质素单体的合成<sup>[2-4]</sup>。

经过 20 多年的研究,拟南芥<sup>[5]</sup>、紫草<sup>[6]</sup>、大豆<sup>[7-8]</sup>、杨树<sup>[9-10]</sup>等植物中 4CL 基因及其家族基因,已经得到了克隆和表达,同时开展了 4CL 反应机制、及涉及催化反应的主要氨基酸保守区域等方面的研究<sup>[11-12]</sup>。然而,目前大多数研究注重于分子生

物和生物化学方面,虽然在 NCBI、EMBL、DBJ 的数据库中存在着大量的 4CL 基因 EST、全长 cDNA 序列,但对 4CL 基因缺乏系统的生物信息学分析和研究报道。本研究利用生物信息学手段,对已经克隆得到的 4CL 基因完整 cDNA 和编码氨基酸序列进行数据挖掘,在核酸和蛋白质水平上进行遗传进化分析,试图揭示单子叶植物和双子叶植物中,4CL 基因的保守性与分化程度,以及进化过程和基因功能之间存在的联系,旨在为 4CL 基因的研究和利用提供一定的理论依据。

## 1 材料与方法

### 1.1 材料

通过生物信息学方法,在美国国立生物技术信息中心(www.ncbi.nlm.nih.gov)的核酸和蛋白质数据库中,对已经克隆得到的 4CL 基因完整的 cDNA 序列和编码氨基酸序列进行数据搜索。

收集得到的 4CL 基因完整 cDNA 和编码氨基酸序列各 50 条,其中双子叶植物 38 条,单子叶植物 10 条,裸子植物 2 条,单子叶植物全部为禾本科植物(表 1)。

表 1 4CL 基因的完整 cDNA 和编码氨基酸序列

Table 1 Complete cDNA and the corresponding amino acid sequences of 4CL gene

植物 Plants	物种 Organism	文章序列编号 No. in paper	数据库完整 cDNA 序列编号 Nucleotide No. in database	数据库编码氨基酸 序列编号 Protein No. in database
双子叶植物 Dicot	拟南芥 <i>Arabidopsis thaliana</i>	At4CL1	At1g51680	NP175579
		At4CL2	At3g21240	NP188761
		At4CL3	At1g65060	NP176686
		At4CL4	At3g21230	NP188760
	欧芹 <i>Petroselinum crispum</i>	Pc4CL1	X13324	CAA31696
		Pc4CL2	X13325	CAA31697
	黄芩 <i>Scutellaria baicalensis</i>	Sb4CL1	AB166767	BAD90936
		Sb4CL2	AB166768	BAD90937
	土豆 <i>Solanum tuberosum</i>	St4CL1	M62755	AAA33842
		St4CL2	AF150686	AAD40664
	丹参 <i>Salvia miltiorrhiza</i>	Sm4CL1	AY237163	AAP68990
		Sm4CL2	AY237164	AAP68991
	复盆子 <i>Rubus idaeus</i>	Ri4CL1	AF239685	AAF91308
		Ri4CL2	AF239686	AAF91309
		Ri4CL3	AF239687	AAF91310
		大豆 <i>Glycine max</i>	Gm4CL1	AF279267
Gm4CL2	AF002259		AAC97600	
Gm4CL3	AF002258		AAC97599	
Gm4CL4	X69955		CAC36095	
紫草 <i>Lithospermum erythrorhizon</i>	Le4CL1	D49366	BAA08365	
	Le4CL2	D49367	BAA08366	
烟草 <i>Nicotiana tabacum</i>	Nt4CL1	D43773	BAA07828	
	Nt4CL2	U50845	AAB18637	
	Nt4CL3	U50846	AAB18638	

续表 1 Continued table 1

植物 Plant	物种 Organism	文章序列编号 No. in paper	数据库完整 cDNA 序列编号 Nucleotide No. in database	数据库编码氨基酸 序列编号 Protein No. in database
双子叶植物 Dicot	毛白杨 <i>Populus tomentosa</i>	Pto4CL1	AF314180	AAL56850
		Pto4CL2	AY043494	AAL02144
		Pto4CL3	AY043495	AAL02145
		Pto4CL4	DQ76679	AAAY84731
	白杨 <i>Populus tremuloides</i>	Ptr4CL1	AF041049	AAC24503
		Ptr4CL2	AF041050	AAC24504
	杂交杨 <i>Populus trichocarpa</i> × <i>Populus deltoids</i>	Ptri4CL1	AF008183	AAC39365
		Ptri4CL2	AF008184	AAC39366
		Ptri4CL3	AF283552	AAK58908
		Ptri4CL4	AF283553	AAK58909
	藿香 <i>Agastache rugosa</i>	Ar4CL	AY587891	AAT02218
	紫穗槐 <i>Amorpha fruticosa</i>	Af4CL	AF435968	AAL35216
	辣椒 <i>Capsicum annuum</i>	Ca4CL	AF212317	AAG43823
赤桉 <i>Eucalyptus camaldulensis</i>	Ee4CL	DQ147001	AAZ79469	
单子叶植物 Monocot	水稻 <i>Oryza sativa</i>	Os4CL1	NM001052604	NP001046069
		Os4CL2	NM001054354	NP001047819
		Os4CL3	NM001055527	NP001048992
		Os4CL4	NM001064787	NP001058252
		Os4CL5	NM001067888	NP001061353
		Os4CL6	NM001068470	NP001061935
	黑麦草 <i>Lolium perenne</i>	Lp4CL1	AF052221	AAF37732
		Lp4CL2	AF052222	AAF37733
		Lp4CL3	AF052223	AAF37734
	玉米 <i>Zea mays</i>	Zm4CL	AY566301	AAS67644
裸子植物 Gymnosperm	火炬松 <i>Pinus taeda</i>	Pt4CL1	U12012	AAA92668
		Pt4CL2	U12013	AAA92669

## 1.2 方法

1.2.1 4CL 基因完整 cDNA 序列的分析 (1)4CL 基因系统发生树的构建。采用 Mega3.1<sup>[13]</sup> 软件内置的 ClustalW<sup>[14]</sup> 程序进行核酸序列多重比对(比对采用程序默认的 IUB 记分矩阵),采用 Mega3.1 对 ClustalW 产生的多重比对结果构建系统发生树。参数设置:采取最大简约法(Maximum Parsimony, MP)构建系统树,采用随机逐步比较的方式搜索最佳系统树,对生成的系统发生树进行 Bootstrap 校正,最终生成系统发生树。

(2)4CL 基因 GC 含量的分析。采用 Mega3.1 软件进行分析。

1.2.2 4CL 基因编码氨基酸序列的分析 采用 ExpASy 的 Prosite 数据库(www.expasy.org/prosite/),NCBI 的 Conserved Domains 数据库<sup>[15]</sup>(www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi)和 Mega3.1 软件,对 4CL 基因编码氨基酸序列进行氨基酸保守区分析;采用 Mega3.1 软件,对 4CL 基因编码氨基酸序列的氨基酸平均含量进行分析。

## 2 结果与分析

### 2.1 4CL 基因的完整 cDNA 序列

利用 Mega3.1 软件,构建了 4CL 基因的系统发生树,见图 1。

2.1.1 4CL 基因的进化 由图 1 可知,系统发生树中的 4CL 基因主要聚为 2 类:第 1 类包括拟南芥、大豆、烟草、杨树等大部分双子叶植物的 4CL 基因(以○标记);第 2 类包括水稻、玉米、黑麦草这几种单子叶植物和裸子植物的 4CL 基因,另外还包括来自双子叶植物中 7 种植物的 8 条 4CL cDNA 序列,如拟南芥 At4CL3、藿香 Ar4CL、大豆 Gm4CL3 和 Gm4CL4、紫草 Le4CL2、白杨 Ptr4CL2、杂交杨 Ptri4CL4、复盆子 Ri4CL1(以△标记)。

由图 1 还可知,被子植物在进化过程中发生了单子叶植物和双子叶植物的分化,但是在第 2 类 4CL 基因中,双子叶植物拟南芥、藿香、大豆、紫草、白杨、杂交杨和复盆子 6 个物种的 4CL 基因与单子叶植物的 4CL 基因聚在一类,而其余双子叶植物的 4CL 基因全部聚在第 1 类中。表明在单子叶植物和

双子叶植物分化之前,这些 4CL 基因在植物中已经存在,而且在植物中的进化时间超前于单子叶和双子叶植物的分化时间。在分类地位上,虽然拟南芥、藿香、大豆、紫草、白杨、杂交杨和复盆子 6 种植物与单子叶植物差异较大,但其 4CL 基因的功能却非常相似,在 4CL 系统发生树中,它们的 4CL 基因聚在同一类中。这种基因的聚类 and 植物分类存在冲突的

现象,在植物中已经被广泛发现。有研究表明,基因的倍增和重组、水平的基因转移等都是这种差异存在的原因<sup>[16-17]</sup>。

系统发生树(图 1)中,4CL 基因聚为 2 类,这与植物中单子叶、双子叶植物的分类大体一致。表明植物和 4CL 基因的进化基本是一致的,4CL 基因和植物的进化过程密切相关。

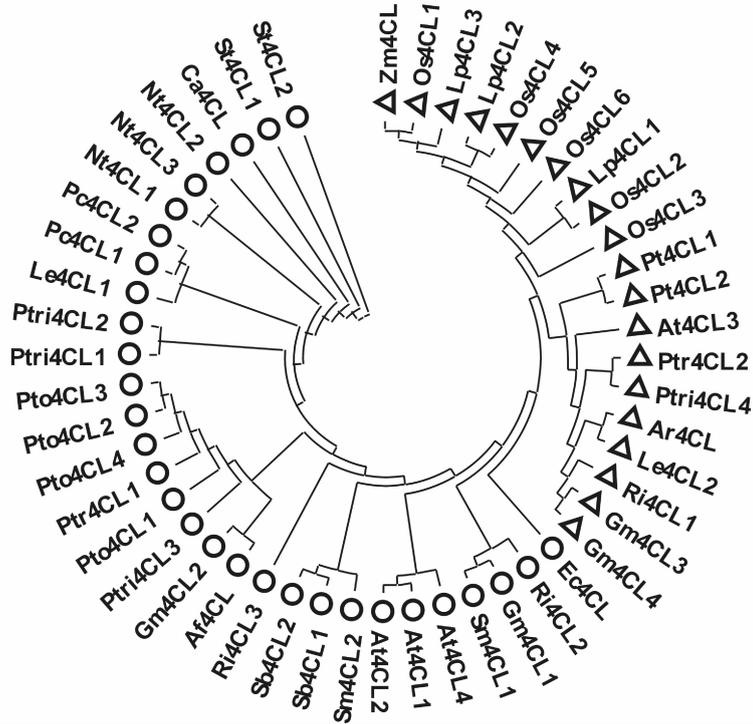


图 1 4CL 基因的系统发生树

Fig. 1 Phylogenetic tree of 4CL gene

2.1.2 家族基因的存在 在系统发生树的单子叶和双子叶植物 4CL 基因序列中,发现大部分 4CL 基因是以基因家族的形式存在的,如拟南芥、水稻、大豆、烟草等(图 1、表 1),表明 4CL 基因在这些植物的进化过程中,随着植物基因组的倍增产生了基因的多个拷贝,形成了并系同源的家族基因。Wendel<sup>[18]</sup>发现,在开花植物中,全基因组倍增是进化过程中普遍存在的重要现象。Yu 等<sup>[19]</sup>、Guyot 等<sup>[20]</sup>也发现,水稻基因组在单、双子叶植物分化之前,曾经发生过基因组倍增。Blanc 等<sup>[21]</sup>利用最新的数据库和同义替换率分布检测方法,在拟南芥、玉米、大豆、土豆等 9 种植物中,发现了全基因组倍增的证据。在以上发生过基因组倍增的植物中,均存在 4CL 家族基因。

此外,这些新 4CL 并系同源家族基因的出现,与植物进化过程中一些新功能的产生有着密切的联系。Ehltling 等<sup>[22]</sup>研究表明,在拟南芥中的 4CL 基因中,At4CL1、At4CL2、At4CL4 在木质化程度较高

的组织细胞中表达,与木质素单体的合成密切相关;其中 At4CL4 的表达水平在拟南芥中较低,且在特定条件下表达,可能作用于另外更加专一的底物;相反,At4CL3 却激活 p-香豆酸作为查耳酮合成酶的底物,最后进入到黄酮合成途径。Lindermayr 等<sup>[8]</sup>发现,在大豆中克隆的 Gm4CL1、Gm4CL2、Gm4CL3 和 Gm4CL4 4 条 4CL cDNA 序列,前 3 种 4CL 基因在结构和功能上均不同,仅 Gm4CL4 与 Gm4CL3 高度相似,Gm4CL1 和 Gm4CL2 参与植物生长和发育(包括木质化),而 Gm4CL3 和 Gm4CL4 对环境刺激发生反应。

另外,从杨树<sup>[9]</sup>、松树<sup>[23]</sup>等克隆得到的 4CL 基因也存在基因结构、功能方面的差异,且杨树<sup>[9,24]</sup>、松树<sup>[23]</sup>中的 4CL 基因不能以芥子酸为作用底物。

2.1.3 4CL 基因的 GC 含量 对得到的 50 条 4CL 基因完整 cDNA 序列进行 GC 含量分析(表 2)发现,单子叶植物 4CL 基因的 GC 含量远远高于双子

叶植物,其中黑麦草 Lp4CL1 的 GC 含量达 69.2%;水稻的 Os4CL3 GC 含量最低,为 59.4%;平均为 65.41%。而双子叶植物中,GC 含量最高为赤桉的 Ec4CL,达 61.2%;辣椒的 Ca4CL 最低,只有 41.7%;平均为 47.55%。

Carels 等<sup>[25]</sup>研究发现,在水稻、玉米、大麦 3 种禾本科植物和拟南芥、大豆、烟草等 6 种双子叶植物中,均存在高 GC 含量和低 GC 含量两类基因。4CL 基因 GC 含量在两大类植物中的明显差异,可能与两大类植物的进化过程和生存环境的差异有一定联系。

表 2 4CL 基因中的 GC 含量  
Table 2 GC content of the 4CL gene

植物 Plant	4CL 基因 4CL gene	GC 含量/% GC content	4CL 基因 4CL gene	GC 含量/% GC content	4CL 基因 4CL gene	GC 含量/% GC content
单子叶植物 Monocot	Os4CL1	64.5	Os4CL2	68.9	Os4CL3	59.4
	Os4CL4	63.0	Os4CL5	64.8	Os4CL6	67.0
	Lp4CL1	69.2	Lp4CL2	64.3	Lp4CL3	66.1
	Zm4CL	66.9				
双子叶植物 Dicot	At4CL1	48.0	At4CL2	48.4	At4CL3	45.9
	At4CL4	47.6	Ec4CL	61.2	Gm4CL1	49.9
	Gm4CL2	49.6	Gm4CL3	52.3	Gm4CL4	52.6
	Nt4CL1	42.1	Nt4CL2	42.8	Nt4CL3	42.2
	Pto4CL1	45.4	Pto4CL2	44.9	Pto4CL3	44.9
	Pto4CL4	44.9	Sm4CL1	54.2	Sm4CL2	54.9
	Ri4CL1	52.5	Ri4CL2	52.1	Ri4CL3	50.7
	Af4CL	46.1	Ar4CL	55.6	Ca4CL	41.7
	Le4CL1	42.0	Le4CL2	45.9	Ptr4CL1	44.8
	Pc4CL1	43.1	Pc4CL2	43.4	Ptr4CL2	48.0
	Ptri4CL1	44.2	Ptri4CL2	44.8	Ptri4CL3	44.0
	Ptri4CL4	47.4	Sb4CL1	52.1	Sb4CL2	52.2
	St4CL1	42.3	St4CL2	42.3		
裸子植物 Gymnosperm	Pt4CL1	52.9	Pt4CL2	53.2		

## 2.2 4CL 基因编码的氨基酸序列

2.2.1 4CL 基因编码的氨基酸序列保守区 利用 Mega3.1 内置的 ClustalW 程序,对所有的氨基酸序列进行多重比对。氨基酸多重比对分析结果表明,从蛋白质的 C 端到 N 端,依次发现了以下 6 个氨基酸保守区:① SSGTTGLPKG V;② QGYGMTE;③ GEICIRG;④ GWLHTGD;⑤ VDRLKELIK;⑥ PK-SPSGKILR。

图 2 和图 3 列出了 SSGTTGLPKG V 和 GEICIRG 的氨基酸序列多重比对结果。因为比对结果较大,故图 2 和图 3 的比对结果中未一一列出其他氨基酸序列,仅将氨基酸序列保守区存在的差异位点全部列出。SSGTTGLPKG V 是几乎所有 4CL 中的绝对氨基酸保守区,在 50 条 4CL 多重比对的氨基酸序列中,只有 5 条氨基酸序列存在差异(图 2)。

在 ATP 和  $Mg^{2+}$  存在的条件下,4CL 催化底物形成相应的 CoA 酯,同时形成 AMP。保守区① (SSGTTGLPKG V)被认为是 4CL 催化反应中保守的 AMP 结合功能域<sup>[26]</sup>。同时,在 ExPASy 的 Pro-site 数据库中发现,SSGTTGLPKG V 的功能注释为 putative AMP-binding domain signature (推定的

AMP 结合域)。SSGTTGLPKG V 同时也是包括 4CL、荧光素酶、乙酰 CoA 合成酶、长链脂酰基 CoA 合成酶和多肽合成酶在内的腺苷酸形成酶超家族的保守区域<sup>[27]</sup>。SSGTTGLPKG V 已经成为腺苷酸形成酶超家族的标志之一<sup>[11,27]</sup>。

在几乎所有的 4CL 中,保守区③ (GEICIRG)也是一个氨基酸绝对保守区。由比对结果(图 3)发现,50 条氨基酸序列中只有 7 条存在差异。在 GEICIRG 保守区域中,甘氨酸 Gly、谷氨酸 Glu、半胱氨酸 Cys 是最为保守的氨基酸位点。

以前的研究认为,4CL GEICIRG 中的半胱氨酸 Cys 残基直接参与酶催化反应<sup>[28]</sup>。但是 Stuitable 等<sup>[11]</sup>将拟南芥 At4CL2 的 GEICIRG 保守区半胱氨酸 Cys 突变成丙氨酸 Ala,突变的拟南芥 At4CL2 酶活性降低到野生型拟南芥的 45%。接近 50% 的酶活存在,证明了半胱氨酸 Cys 并未直接参与酶的反应。由此推测,酶活性的降低可能是半胱氨酸 Cys 与稳定酶活性空间结构构象有关。At4CL2 中另一保守位点谷氨酸 Glu 残基的突变试验也得到了类似的结果。Stuitable 等<sup>[11]</sup>认为,虽然 GEICIRG 在 4CL 氨基酸序列中十分保守,但并不像保守区①

(SSGTTGLPKG<sup>V</sup>)直接参与反应,但是这个保守区十分重要,缺失 GEICIRG 保守区的 4CL 完全丧失了活性。

在 NCBI 的 Conserved Domains<sup>[15]</sup> 数据库中,

Os4CL1	SSGTTGLPKG <sup>V</sup>
At4CL1	SSGTTGLPKG <sup>V</sup>
Zm4CL	SSGTTGLPKG <sup>V</sup>
Ec4CL	SSGTTGLPKG <sup>V</sup>
Lp4CL3	SSGTTGLPKG <sup>V</sup>
At4CL4	SSGTTGLPKG <sup>V</sup>
Os4CL3	SSGTTGDSK <sup>G</sup> V
Sb4CL1	SSGTTGLPKA <sup>L</sup>
Sb4CL2	SSGTTGLPKA <sup>V</sup>
Gm4CL1	SSGTTGLPKG <sup>V</sup>
Ar4CL	SSGTTGLPKG <sup>V</sup>

图 2 4CL 基因编码氨基酸序列保守区

SSGTTGLPKG<sup>V</sup> 的多重比对结果

阴影部分为差异位点

Fig. 2 Multiple alignment result of SSGTTGLPKG<sup>V</sup>

The different sites were shaded

发现保守区①②③④⑤⑥一起构成了 CaiC 氨基酸保守区,主要功能注释为 Acyl-CoA synthetases (AMP-forming),行使酰基 CoA 合成酶功能,产生 AMP(图 4)。

Os4CL1	GEICIRG
At4CL1	GEICIRG
Zm4CL	GEICIRG
Ec4CL	GEICIRG
Lp4CL3	GEICV <sup>R</sup> RG
At4CL4	GEICV <sup>R</sup> RG
Os4CL3	GEICV <sup>R</sup> RG
Ri4CL1	GEICV <sup>R</sup> RG
Gm4CL1	GEICIRG
Sm4CL1	GEICIRG
Ca4CL	GENCIRG

图 3 4CL 基因编码氨基酸序列保守区

GEICIRG 的多重比对结果

阴影部分为差异位点

Fig. 3 Multiple alignment result of GEICIRG

The different sites were shaded

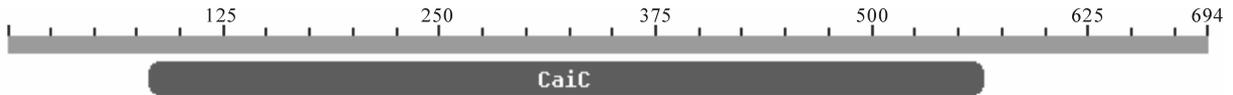


图 4 4CL 的 Caic 氨基酸序列保守区

Fig. 4 Caic amino acid conserved blocks in 4CLs

综上分析,保守区①(SSGTTGLPKG<sup>V</sup>)是单、双子叶两大类植物 4CL 中最为重要的氨基酸序列保守区,在其他保守区的辅助作用下共同行使功能。

2.2.2 4CL 基因编码氨基酸序列的氨基酸平均含量 在 50 条 4CL 基因编码的氨基酸序列中,丙氨酸 Ala(9.22%)、亮氨酸 Leu(9.40%)、缬氨酸 Val(8.63%)是平均含量最高的氨基酸,平均含量偏低

的为半胱氨酸 Cys(1.80%)、组氨酸 His(2.14%)、甲硫氨酸 Met(2.36%)、酪氨酸 Tyr(2.60%),最低的为色氨酸 Trp(0.21%)。

在组成蛋白质的 20 种氨基酸中,单子叶植物和双子叶植物 4CL 中有 10 种氨基酸平均含量存在着较大差异(表 3),其他 10 种氨基酸差异在 0.0%~0.3%。

表 3 单子叶和双子叶植物 4CL 中 11 种氨基酸的平均含量

Table 3 Average contents of 11 kinds of amino acid between monocots and dicots 4CLs

项目 Item	氨基酸 Amino Acid										
	丙氨酸 Ala	精氨酸 Arg	缬氨酸 Val	甘氨酸 Gly	赖氨酸 Lys	异亮 氨酸 Ile	天冬 酰胺 Asn	丝氨酸 Ser	谷氨 酰胺 Gln	酪氨酸 Tyr	色氨酸 Trp
所有植物 All plants	9.22	3.77	8.63	6.85	6.33	7.03	3.24	6.25	3.04	2.60	0.21
双子叶植物 Dicots	8.41	3.41	8.30	6.59	6.71	7.30	3.56	6.53	3.18	2.76	0.20
单子叶植物 Monocots	12.12	5.18	9.70	7.88	4.91	5.87	2.19	5.25	2.38	1.97	0.28
单、双子叶植物间差值 AA Difference	-3.71	-1.77	-1.4	-1.29	1.8	1.43	0.37	1.28	0.8	0.79	-0.08

注:单、双子叶两大类植物之间氨基酸平均含量差值,是利用双子叶植物中的氨基酸平均含量减去单子叶植物中的氨基酸平均含量所得到的值。

Note: The AA differences indicate the differences of the average amino acids contents between monocots and dicots.

由表 3 可知,单子叶植物和双子叶植物中平均含量较高(>5.8%)的氨基酸基本一致。单子叶植物中为:丙氨酸 Ala、缬氨酸 Val、甘氨酸 Gly、异亮氨酸 Ile,这些全部为中性氨基酸,其中丙氨酸 Ala

平均含量最大,为 12.12%;双子叶植物中为:丙氨酸 Ala、缬氨酸 Val、甘氨酸 Gly、异亮氨酸 Ile、赖氨酸 Lys、丝氨酸 Ser,其中赖氨酸 Lys 为碱性氨基酸,丝氨酸 Ser 为含羟基氨基酸。

由表 3 还可知,在 10 种平均含量存在较大差异的氨基酸中,单子叶植物有 4 种氨基酸的平均含量超出双子叶植物,其中 3 种为中性氨基酸(丙氨酸 Ala、缬氨酸 Val、甘氨酸 Gly),1 种为碱性氨基酸(精氨酸 Arg)。另外 6 种氨基酸在双子叶植物中的平均含量超出单子叶植物,其中有 1 种中性氨基酸(异亮氨酸 Ile)、2 种酸性氨基酸(天冬酰胺 Asn、谷氨酰胺 Gln)、1 种碱性氨基酸(赖氨酸 Lys)、1 种含羟基氨基酸(丝氨酸 Ser)和 1 种芳香氨基酸(酪氨酸 Tyr)。

以上分析表明,与单子叶植物相比,不论是平均含量较高的氨基酸种类,还是平均含量超出单子叶植物的氨基酸种类,双子叶植物 4CL 中的氨基酸种类都要广泛得多。从 4CL 基因的系统发生树(图 1)可以看出,双子叶植物的 4CL 基因聚在第 1 和第 2 类中,单子叶植物仅聚在第 2 类中。也许,正是以上氨基酸组成的这些差异,在一定程度上决定了 4CL 基因功能的分化。单、双子叶植物中 4CL 氨基酸组成的差异,可能是单子叶和双子叶植物的分化及 4CL 基因进化的证据之一。

另外,平均含量最低的芳香氨基酸(色氨酸 Trp)(0.21%)在两类植物之间差异很小,表明在两大类植物中,Trp 相对稳定;其含量虽然很低,但是其作用却不可忽视。

### 3 讨论

4CL 作为木质素生物合成的关键酶,对其基因的研究已经开展了 20 多年。因为各个研究的侧重点不同,得到的 4CL 基因完整序列相对较少。同时,以前的研究主要集中在分子生物学和生物化学两大方面,4CL 基因的克隆、功能鉴定、转基因植株的获得等是分子生物学方面的主要内容,生物化学方面则对 4CL 的反应机制、氨基酸序列保守区等方面进行了研究。本研究利用前人研究得到的 4CL 基因完整 cDNA 序列和氨基酸序列保守区的成果,结合生物信息学手段进行了系统的生物信息学分析,得到了相对满意的结果。

本研究利用收集得到的 50 条 4CL 基因完整 cDNA 序列,构建了 4CL 基因的系统发生树。虽然得出了基因聚类和植物分类大体一致的结论,但是基因的聚类相对比较粗略,主要是 4CL 基因完整序列相对较少。随着研究的深入,不断增加的 4CL 基因完整序列将会构建出更加精细的系统发生树。

本研究利用 NCBI 的 Conserved Domains 和

ExPASy 的 Prosite 两大氨基酸序列保守区数据库进行了 4CL 基因编码氨基酸序列的分析。这 2 个数据库包括了大量的基因编码氨基酸序列,其功能主要是提供氨基酸序列保守区的检索和功能注释,差别主要在检索采用的不同计算机算法。这两大数据库对功能基因蛋白产物的研究很有裨益。

本研究采用 Mega 软件作为主要的分析工具。随着生物学和计算机科学的发展,生物信息学方面的软件工具大量出现,在进化分析方面并非只有 Mega 软件可以利用,还包括 Paup、Phylip、DNA-MAN 等优秀的软件。正是这些应用软件的出现,大大促进了生物信息学方面的研究。

### 4 结论

利用收集得到的 50 条 4CL 基因完整 cDNA 序列,构建了 4CL 基因的系统发生树,揭示出 4CL 基因的保守性和分化的事实。在植物进化过程中,随着植物基因组的倍增,一些植物中的 4CL 基因产生了新的拷贝,同时发生了基因结构和功能上的变化,发生了 4CL 基因的分化。在单、双子叶两大类植物之间,4CL 基因的 GC 含量存在着明显差异。4CL 基因编码氨基酸序列分析表明,不同植物中的氨基酸序列保守区高度相似,差异较小,但是单、双子叶两大类植物之间,4CL 的氨基酸含量存在较大差异。这些可能与植物进化及 4CL 基因分化具有密切关系。

### [参考文献]

- [1] Baucher M, Monties B, Van Montagu M, et al. Biosynthesis and genetic engineering of lignin [J]. *Critical Reviews in Plant Sciences*, 1998, 17(2): 125-197.
- [2] Humphreys J M, Chapple C. Rewriting the lignin road map [J]. *Curr Opin Plant Biol*, 2002, 5(3): 224-229.
- [3] Boerjan W, Ralph J, Baucher M. Lignin biosynthesis [J]. *Annu Rev Plant Biol*, 2003, 54: 519-546.
- [4] Lee D, Meyer K, Chapple C, et al. Antisense suppression of 4-coumarate:coenzyme A ligase activity in Arabidopsis leads to altered lignin subunit composition [J]. *Plant Cell*, 1997, 9: 1985-1998.
- [5] Lee D, Ellard M, Wanner L A, et al. The Arabidopsis thaliana 4-coumarate:CoA ligase (4CL) gene: stress and developmentally regulated expression and nucleotide sequence of its cDNA [J]. *Plant Mol Biol*, 1995, 28(5): 871-884.
- [6] Yazaki K, Ogawa A, Tabata M. Isolation and characterization of two cDNAs encoding 4-coumarate:CoA ligase in Lithospermum cell cultures [J]. *Plant Cell Physiol*, 1995, 36(7): 1319-1329.

- [7] Uhlmann A, Ebel J. Molecular cloning and expression of 4-coumarate:coenzyme A ligase, an enzyme involved in the resistance response of soybean (*Glycine max* L.) against pathogen attack [J]. *Plant Physiol*, 1993, 102(4): 1147-1156.
- [8] Lindermayr C, Mollers B, Fliegmann J, et al. Divergent members of a soybean (*Glycine max* L.) 4-coumarate:coenzyme A ligase gene family [J]. *Eur J Biochem*, 2002, 269(4): 1304-1315.
- [9] Hu W J, Kawaoka A, Tsai C J, et al. Compartmentalized expression of two structurally and functionally distinct 4-coumarate:CoA ligase genes in aspen (*Populus tremuloides*) [J]. *Proc Natl Acad Sci USA*, 1998, 95(9): 5407-5412.
- [10] Cukovic D, Ehlting J, Van Ziffle J A, et al. Structure and evolution of 4-coumarate:coenzyme A ligase (4CL) gene families [J]. *Biol Chem*, 2001, 382(4): 645-654.
- [11] Stuible H P, Buttner D, Ehlting J. Mutational analysis of 4-coumarate:CoA ligase identifies functionally important amino acids and verifies its close relationship to other adenylate-forming enzymes [J]. *FEBS Letters*, 2000, 467: 117-122.
- [12] Stuible H P, Kombrink E. Identification of the substrate specificity-conferring amino acid residues of 4-coumarate:coenzyme A ligase allows the rational design of mutant enzymes with new catalytic properties [J]. *J Biol Chem*, 2001, 276: 26893-26897.
- [13] Kumar S, Tamura K, Nei M. MEGA3; Integrated software for molecular evolutionary genetics analysis and sequence alignment [J]. *Briefings in Bioinformatics*, 2004, 5: 150-163.
- [14] Thompson J D, Higgins D G, Gibson T J. CLUSTAL W; improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice [J]. *Nucleic Acids Research*, 1994, 22: 4673-4680.
- [15] Marchler-Bauer A, Bryant S H. CD-Search; protein domain annotations on the fly [J]. *Nucleic Acids Res*, 2004, 32(W): 327-331.
- [16] Doyle J J. Trees within trees; genes and species, molecules and morphology [J]. *Syst Biol*, 1997, 46: 537-553.
- [17] Page R D M, Charleston M A. From gene to organismal phylogeny; reconciled trees and the gene tree/species tree problem [J]. *Mol Phylogenet Evol*, 1997, 7: 231-240.
- [18] Wendel J F. Genome evolution in polyploids [J]. *Plant Mol Biol*, 2000, 42: 225-249.
- [19] Yu J, Wang J, Lin W, et al. The genome of *Oryza sativa*; A history of duplication [J]. *PLoS Biol*, 2005, 3(2): 266-281.
- [20] Guyot R, Keller B. Ancestral genome duplication in rice [J]. *Genome*, 2004, 47(3): 610-614.
- [21] Blanc G, Wolfe K H. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes [J]. *Plant Cell*, 2004, 16: 1667-1678.
- [22] Ehlting J, Buttner D, Wang Q, et al. Three 4-coumarate:coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms [J]. *Plant J*, 1999, 19: 9-20.
- [23] Voo K S, Whetten R W, O'Malley D M, et al. 4-coumarate:coenzyme a ligase from loblolly pine xylem; Isolation, characterization, and complementary DNA cloning [J]. *Plant Physiol*, 1995, 108(1): 85-97.
- [24] 范丙友, 胡诗宇, 陆海, 等. 毛白杨 4-香豆酸:辅酶 A 连接酶可溶性原核表达及活性检测 [J]. *北京林业大学学报*, 2006, 28(2): 1-8.  
Fan B Y, Hu S Y, Lu H, et al. Soluble prokaryotic expression of 4-coumarate:coenzyme A ligase from *Populus tomentosa* and enzyme activity of recombinant 4CL1 [J]. *Journal of Beijing Forestry University*, 2006, 28(2): 1-8. (in Chinese)
- [25] Carels N, Bernardi G. Two classes of genes in plants [J]. *Genetics*, 2000, 154: 1819-1825.
- [26] Challis G L, Ravel J, Townsend C A. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains [J]. *Chem Biol*, 2000, 7: 211-224.
- [27] Fulda M, Heinz E, Wolter F P. The fadD gene of *Escherichia coli* K12 is located close to rnd at 39.6 min of the chromosomal map and is a new member of the AMP-binding protein family [J]. *Mol Gen Genet*, 1994, 242(3): 241-249.
- [28] Becker-Andre M, Schulze-Lefert P, Hahlbrock K. Structural comparison, modes of expression, and putative cis-acting elements of the two 4-coumarate:CoA ligase genes in potato [J]. *J Biol Chem*, 1991, 266: 8551-8559.