

白及反式-肉桂酸 4-单加氧酶基因 *C4H* 的序列分析

赵永彬¹, 黄恻隐², 徐程², 钱露², 李清清², 何毅怀³, 徐德林^{2*}

¹遵义医科大学珠海校区 广东珠海

²遵义医科大学细胞生物学教研室 贵州遵义

³遵义医科大学附属医院 贵州遵义

【摘要】反式-肉桂酸 4-单加氧酶是植物中苯丙素和木质素合成的关键酶之一, 研究 *C4H* 基因的序列信息对揭示其蛋白结构及功能具有重要的指导意义。从白及转录组数据库中分离得到 *BsC4H* 基因, 运用生物信息学软件对该基因进行序列分析。分析得到的结果显示: 该基因长度为 1993 bp, 编码的蛋白质长度为 505 aa、分子量为 58.01858 kD、理论等电点为 9.20, 该基因的二级结构主要由 α 螺旋结构和无规则卷曲结构组成, 三级结构预测为单体结构。氨基酸序列与铁皮石斛 *C4H* 基因相似性最高, 达 95%, 结构域分析发现该蛋白具有高度结构保守域, 但并不包含有跨膜结构域和信号肽, 亚细胞定位分析发现该酶被定位于内质网中。该研究结果为研究白及 *BsC4H* 基因提供了基础数据, 也为研究其它物种 *C4H* 基因提供基础资料。

【关键词】白及; 反式-肉桂酸 4-单加氧酶; 序列分析; 结构预测

【基金项目】国家自然科学基金项目 (31560079、31960074); 遵义市科技局遵义医科大学科学技术联合资金招标专项 (遵市科合 HZ 字 (2020) 91 号); 贵州省科学技术基金项目 (黔科合平台人才 [2017]5733-050); 贵州省农业重大产业科学研究攻关项目 (黔科教 KY 字 [2019]016); 遵义医科大学 2019 年度学术新苗培养及创新探索专项项目 (黔科合平台人才 [2019]-027)

Bioinformatic Analysis of Trans-cinnamate 4-monooxygenase Gene *C4H* in *Bletilla striata*

Yongbin Zhao¹, Ceyin Huang², Cheng Xu², Lu Qian², Qingqing Li², Yihuai He³, Delin Xu^{2*}

¹Zunyi Medical University Zhuhai campus, Zhuhai, Guangdong

²Department of cell biology, Zunyi Medical University, Zunyi, Guizhou

³Affiliated Hospital of Zunyi Medical University, Zunyi, Guizhou

【Abstract】 Trans-cinnamic acid 4-monooxygenase is one of the key enzymes in the synthesis of phenylpropanoid and Lignin in plants. The *C4H* gene of *Bletilla striata* (*BsC4H*) was isolated from our former constructed transcriptome database and then analyzed by a series of bioinformatics software. The results showed that the length of the gene was 1993 bp, which encoding a protein with 505 aa in length which have molecular weight of 58.01858 kd and theoretical isoelectric point of 9.20. The predicted secondary structure of this protein was composed of helix structure and irregular coil structure, the tertiary structure was predicted to be monomer structure. *BsC4H* sequences of amino acid showed 95% similarity to *C4H* gene derived from *Dendrobium officinale*. The domain detection showed that *BsC4H* protein had a highly conserved domain, but no transmembrane domain and signal peptide. While the detection of subcellular localization found *BsC4H* was mainly functioned in the endoplasmic reticulum. The results of this study provided basic data for the subsequent study of *BsC4H* and developed inspiriment for of *C4H* studies of other species.

【Keywords】 *Bletilla striata*; Trans-cinnamate 4-monooxygenase; sequence analysis; structure prediction

作者简介: 赵永彬 (1998-), 男, 本科生, 研究方向为医学生物信息学。

*通讯作者: 徐德林, 男, 博士, 教授, 研究方向为中药材遗传育种。

白及 (*Bletilla striata* (Thunb.) Rchb.f.) 是我国一种常用中药, 药用部位通常是其假鳞茎, 具有收敛止血, 清热利湿, 消肿生肌之功^[1]; 在临床上广泛用于消化道出血, 烫伤、烧伤, 栓塞剂等应用^[2]。另外, 白及可以制作外用制剂, 比如膜剂、涂剂和洗剂等^[3]; 白及还是一些化妆品的原料, 还可用作观赏的花草。

反式-肉桂酸 4-单加氧酶, 又称肉桂酸 4-羟化酶 (Cinnamate 4-hydroxylase, C4H) 是苯丙烷代谢途径中的第二个关键酶, 它参与调节一些代谢途径的下游产物, 如在木质素、黄酮类和芳香类等化合物的调节^[4], 它属于细胞色素 P450 单加氧酶 (P450s) 中的一种。细胞色素 P450 单加氧酶广泛存在于各种微生物中, 细胞色素 P450 单加氧酶具有催化活性混杂性的特点, 所以可催化羟化反应, 环氧化反应和脱烷基反应等氧化反应^[5]。*C4H* 基因在苯丙烷代谢中有重要作用, 能催化反式肉桂酸生成对香豆酸^[6]; 对香豆酸是肉桂酸的一个酚类衍生物, 具有抗氧化作用, 它在植物中能够起着染色和保护植物免受环境胁迫的作用^[7]。*C4H* 基因的表达量可以影响许多条代谢途径的进展方向, 它不仅是苯丙烷代谢途径的重要酶之一, 而且在电子传递反应中也起着重要作用^[8]。

目前, 关于白及 *C4H* 基因 (*BsC4H*) 的研究报道很少。因此, 本文章通过生物信息分析对该基因的结构、理化性质和功能等特征进行探究, 为以后研究 *BsC4H* 基因以及其它植物 *C4H* 基因提供基础资料和参考。

1 材料与方法

1.1 材料

白及材料来自于遵义正安地区, 种植在本实验室的种质资源圃, 选取生长状态良好的嫩叶进行 DNA 提取。

1.2 方法

(1) 序列来源

BsC4H 基因的核苷酸序列来源于本课题组前期构建并注解的白及核苷酸数据库^[9]。

(2) 理化性质的分析

应用 ProtScale 软件对反式-肉桂酸 4-单加氧酶进行了亲疏水性分析, 运用在线程序 Protparam tool 对 *BsC4H* 基因的氨基酸序列进行了氨基酸含量、分

子量和等电点的分析。

(3) 序列特征分析

应用在线 NCBI 的 ORF Finder

(<https://www.ncbi.nlm.nih.gov/orffinder>) 检测 *BsC4H* 基因的阅读框和氨基酸序列, 用 NCBI 的 blastp 进行在线的同源序列检测, 使用 DNAMAN 软件进行多物种氨基酸序列比对, 然后用 MEGA5.10 软件的构建系统进化树。

(4) 结构域和二级、三级结构的分析

运用在线软件 InterProScan 和 GenBank 对基因进行了功能域分析, 应用 SPOMA 和 SWISS-MODEL 进行了 *BsC4H* 蛋白的二级结构与三级结构预测。

(5) 功能分析

应用在线软件 WORFPSOPRT 对 *BsC4H* 基因的氨基酸序列进行了亚细胞定位, 用 Signalp5.0 软件进行信号肽分析, 用 TMHMM Server v.2.0 对该蛋白进行跨膜结构域分析。

2 结果与分析

2.1 理化性质分析

运用 ExPASy 网站的 Protparam tool 程序分析白及 *BsC4H* 蛋白序列分子量为 58.01858kD、理论等电点为 9.20, 原子组成 C2640H4173N719O718S17, 带的负电荷残基为 60 个, 带的正电荷残基为 69 个, 结果表明该蛋白为碱性蛋白。通过 ProtScale 程序对该蛋白进行了亲疏水性分析对比 (如图 1), 发现该蛋白为亲水性蛋白质, 其亲水峰值为 -2.772, 疏水峰值为 2.867。

2.2 氨基酸序列的分析

从白及转录组数据库中获得了 *BsC4H* 基因的核苷酸序列全长为 1993bp, 用 ORF Finder 进行开放阅读框检索发现该基因共编码了 505 个氨基酸残基。通过软件 DNAMAN 进行进行了多物种 *C4H* 基因的氨基酸序列比对, 发现 *BsC4H* 的氨基酸序列与洋葱 (*Allium cepa*)、铁皮石斛 (*Dendrobium officinale*)、油棕 (*Elaeis guineensis*)、深圳拟兰 (*Apostasia shenzhenica*)、小兰屿蝴蝶兰 (*Phalaenopsis equestris*) 等植物 *C4H* 基因的的氨基酸序列具有同源性 (如图 2), 从结果中可看出该基因具有高度的保守性。根据同源性对比发现 *BsC4H* 基因的氨基酸序列与铁皮石斛氨基酸 (NCBI 登录号: XP_020701874.1) 的同源性相似性最高, 达 95%。

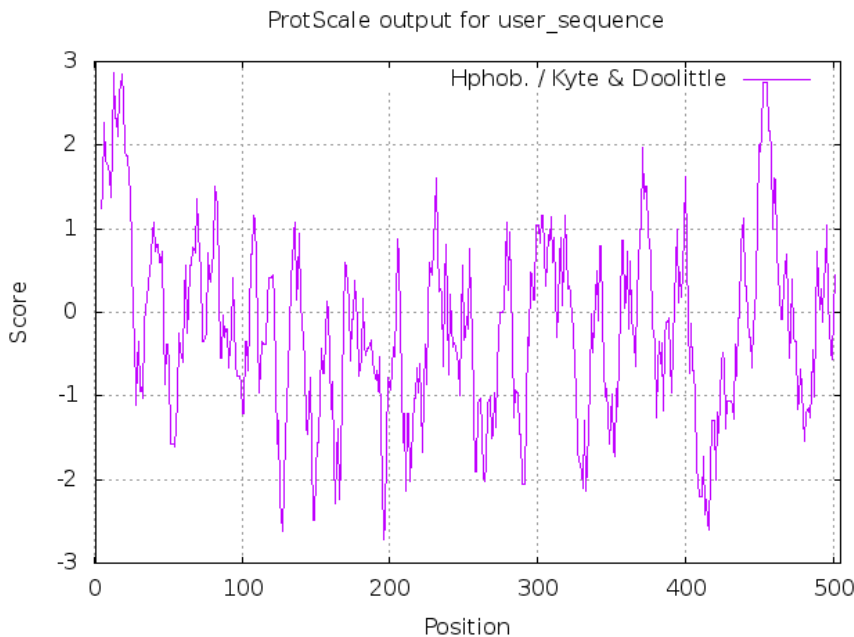


图 1 BsC4H 蛋白亲疏水性分析图

AAS48416.1	E LLL T IS F I	IVI KLR KFK	H ANF K DVF	DL D	101
AGN70893.1	D VLV A LG A V	VVV KLS KLK	H ASL K DIL	DH E	101
PKA58596.1	D LLV A LG A V	IVV KFR KLY	H AAF K DTL	DL E	101
XP_020572760.1	D VLV A LG A V	IVV KLS KLK	H ASI K EIF	EL E	101
XP_020685658.1	D LVL A LG A V	IVF TLR RFK	Q AAL K DIF	DL E	101
XP_020701874.1	D VLV A LG A V	IVV KLS KLK	H ASL K DIL	DH E	101
XP_010921158.1	D VLL A LG A V	IAV KIR RFK	H TAL R DIF	EL E	101
BsC4H	D VLV V LG A V	IVV KLS KLK	H ATL K DIL	DL E	101
Consensus	m l e k l f a t l a s g k	l p p g p l p v p i f g n w l q v g d d l n r n l	a k f g	l l r m g q r n l v v v s s p	a r v l h t q g v e f g s r
AAS48416.1		M N Y Y D GR	KKDAKAA I L K	N	202
AGN70893.1		V Q Y Y D AS	RKSPKSA V I R	D	202
PKA58596.1		M Q Y Y D AR	RRNARAA I I R	E	202
XP_020572760.1		M Q Y Y D AR	RQNPKAT I I R	E	202
XP_020685658.1		M Q Y Y D AR	RANPKSA I I R	E	202
XP_020701874.1		M Q Y Y D AS	RKNPKSA V I R	D	202
XP_010921158.1		M Q N H E RQ	KANPKAA I L R	D	202
BsC4H		M Q Y Y D AR	RQNPKSA V I R	E	202
Consensus	tr n v f d i f t g k g d m v f t v y g e h w r k	r r i m t v p f f t n k v v q r g w e e a v v e d v	t e g v r r l q l m m y n n m y r i m f d r r f e s e	d p l	
AAS48416.1	L K G	G KEV ER IQL D LE	SST PTDNAG K D K N	303	
AGN70893.1	M R N	G KDE DV LNL N VQ	SST PTDNAG K D K N	303	
PKA58596.1	V R G	G NEV ER LKL D VE	SSP PTDNAG K D K N	303	
XP_020572760.1	M R N	G KEV DV LRL D VQ	SST PTDNAG K D K N	303	
XP_020685658.1	L R G	G KEV DM ISI S VE	GNT PTDITG K D K N	303	
XP_020701874.1	M R N	G KDV DV LNL N VQ	SST PTDNAG K D K N	303	
XP_010921158.1	N K G	A QEV DR LQL D IQ	AST GASNME K E R N	303	
BsC4H	M R N	G KEV DV FSL N VQ	SST PTDNITG R D K S	303	
Consensus	f k l a l n e r s r l a q s f e y n y g d f i p i l r p f l r	y l k i c k r f k y f e r k k l k	l c a i d h i l a e k g e i e d n v l y i v e n i		
AAS48416.1		I HPKI LQH T P TQ THR I	H G YD	404	
AGN70893.1		I HPEV LRD T R NQ THK I	H G YE	404	
PKA58596.1		I HPEV LRA S P VP THR V	H G YD	404	
XP_020572760.1		I NPEV LRE T P NQ THK I	N G YE	404	
XP_020685658.1		L HQDV VRA S P VQ IPR V	H G FD	404	
XP_020701874.1		I HPEV LRE T R NQ THK I	H G YE	404	
XP_010921158.1		I HPEI LRH T P VQ TQK I	H R CD	404	
BsC4H		I HPEV LRE T P TP THK I	H G YE	404	
Consensus	n v a a i e t t l s w i e g a e l v n q r k	e l d v l g g i t e p d l p y l q a v k e t l r l r m a i p l l v p h m n l	d a k l g i p a e s k i l n v a w l a		
AAS48416.1	AH D EEF	EA YI	G I D M MDKI T	N STI A V	504
AGN70893.1	AQ N DQF	ES FI	G L E L MEKL S	K YTV A E	504
PKA58596.1	AH N EQY	EA FI	S I E L KEKL T	N YTI A V	504
XP_020572760.1	SH N EQF	EF FI	G L E L MEKL A	K YNV A V	504
XP_020685658.1	AH N DQF	EA YI	G L E L MERI T	K STV A S	504
XP_020701874.1	AQ N DQF	ES FI	G L E L MEKL S	K YTV A E	504
XP_010921158.1	DN N EEF	EA YL	G I E L QSKI T	K STI C V	504
BsC4H	AH N EQF	DF FI	G L E L MEKL A	K YTV A A	504
Consensus	n m p w k p r p e r f l e e k v e a n g n d f r	p f g v g r r s c p g i i l a l p i l i t	g r l v q n f l p p p g	d t e k g g q f s l h i l h v k p r	

图 2 8 个物种的 *C4H* 基因的氨基酸序列比对。同源性: 黑色 100%, 红色 75%, 蓝色 25%, 白色 0%。

2.3 系统发育树分析

BsC4H 基因的氨基酸序列与多个物种 *C4H* 基因氨基酸序列通过 MEGA 5.10 软件构建了系统进化树进行分析。分析结果如图 3 所示, 由其结果可看出白及 TC4M 基因与其它物种 TC4M 基因, 如与洋葱 (AAS48416.1)、油棕 (XP_010921158.1) 等表现出较远的亲缘关系, 而与处于同一分支的小兰屿蝴蝶兰 (XP_020572760.1) 亲缘关系最近。

2.4 保守结构域和二级结构、三级结构的预测分析

使用在线软件 InterProScan 预测 *BsC4H* 蛋白的保守结构域 (如图 4A), 共有 2 个 C4H 家族, 1 个同源家族, 1 个保守位点。2 个 C4H 家族其中一个位于 34-495 位 (IPR001128) 属于细胞色素 P450 酶含有血红素的单加氧酶的一个超家族; 另外一个位于 62-470 位 (IPR002421) 细胞色素 P450 中 e 类细胞色素 p450 蛋白, 属于序列集群 i 组, 是成员中最丰富的。属于细胞色素 P450 酶同源家族位于 23-503 位 (IPR036396)。位于 440-449 位 (PS00086) 的半胱氨酸血红素铁结合位点, 此处存在半胱氨酸残基是与第 5 位血红素铁结合。运用 Genbank 对该氨基酸进行保守结构域分析 (如图 4B), 该蛋白有细胞 P450 酶结构域, 这结果与 InterProScan 分析的结果一致。

运用在线软件 SPOMA 对 *BsC4H* 蛋白的二级结

构的预测分析 (如图 5A)。发现其二级结构中含有 49.70% 的 α -螺旋结构, 32.67% 的无规则结构, 12.28% 的线状结构, 5.35% 的 β 折叠结构。使用 SWISS-MODEL 构建了该蛋白的三级结构, 发现其三级结构为单体结构 (如图 5B), 进一步运用 QMEAN4 软件分析得到表 1, 由表 1 可知该蛋白的三级结构更稳定。

表 1 *BsC4H* 蛋白三级结构的 QMEAN4 检测值

计分功能(Scoring function)	Z 分数(Z-score)
C β 相互作用能(C beta interaction energy)	-3.18
所有原子成对能(All-atom pairwise energy)	-1.61
融化剂能(Solvation energy)	0.59
扭转角能(Torsion angle energy)	-3.82
Qmean 评分(Qmean score)	-4.11

2.5 信号肽、跨膜结构域、亚细胞定位预测分析

使用在线软件 SignalP 5.0 预测 *BsC4H* 蛋白的信号肽 (如图 6A), 其剪切位点在发生在 19-20 位氨基酸之间, 其 mean S-score 值为 0.1644 < 0.5, 由此推断该蛋白不是分泌蛋白。通过在线软件 TMHMM Server v.2.0 分析 (如图 6B), 由此推断 *BsC4H* 蛋白无跨膜结构域。用在线工具 WOLFP SOPRT 亚细胞定位分析预测 *BsC4H* 蛋白定位在内质网中。

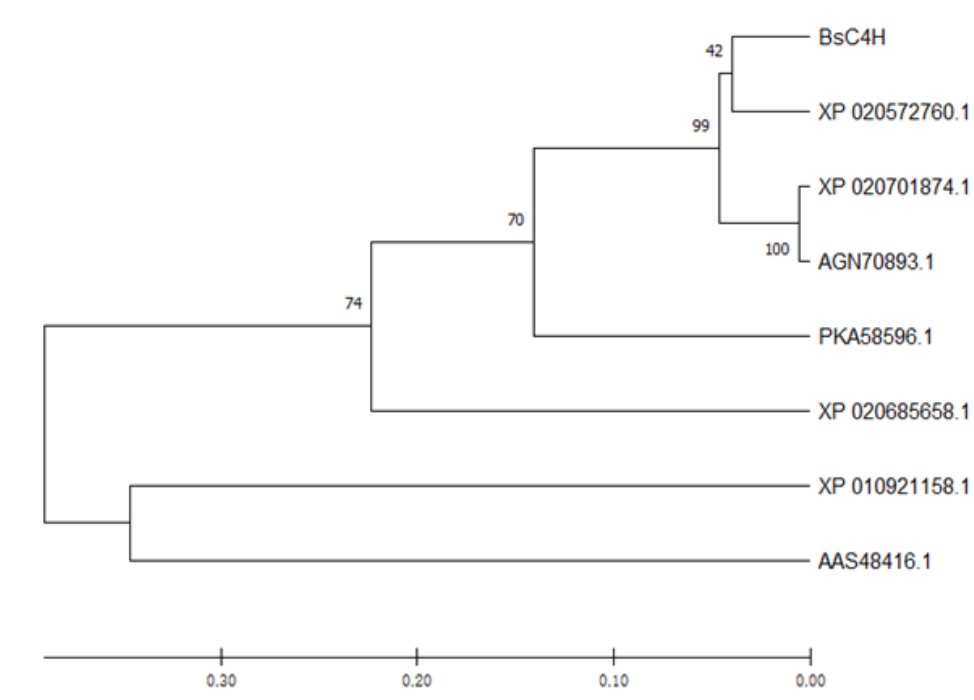


图 3 氨基酸序列进化树分析

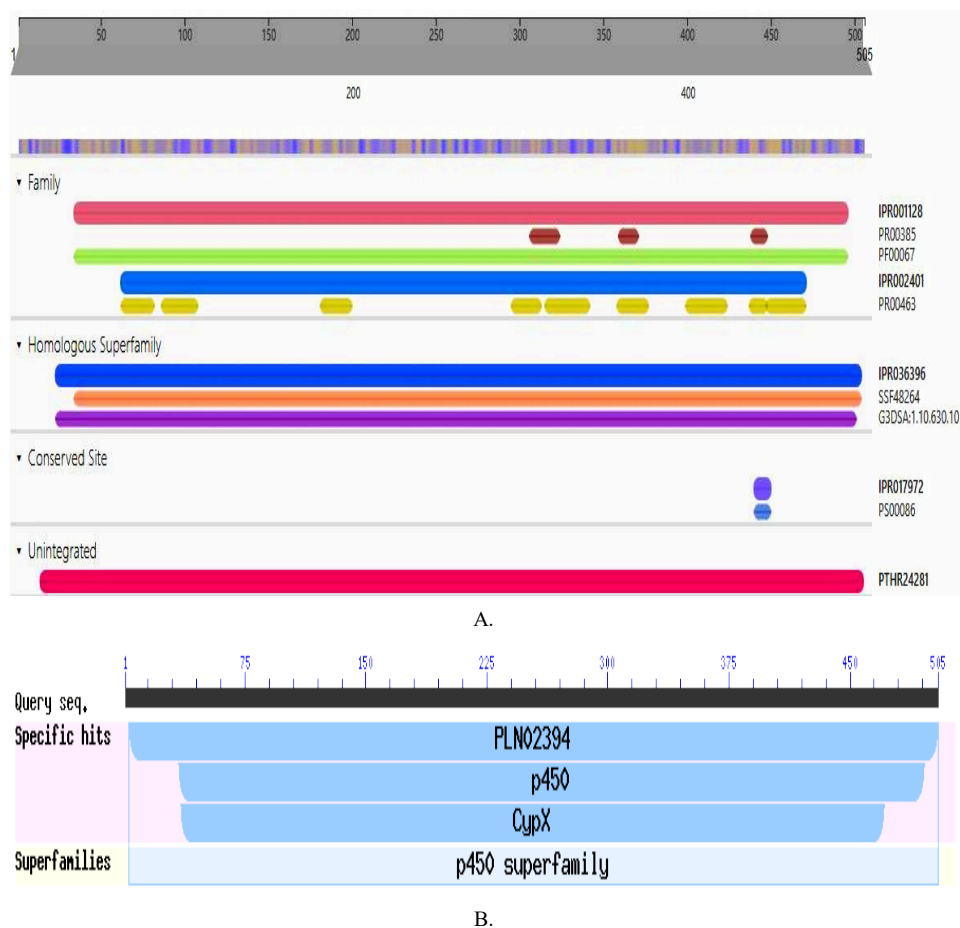


图 4 蛋白序列的保守结构域分析。A. InterProScan 分析结果, B. GeneBank 的 Conserved domains 分析结果。

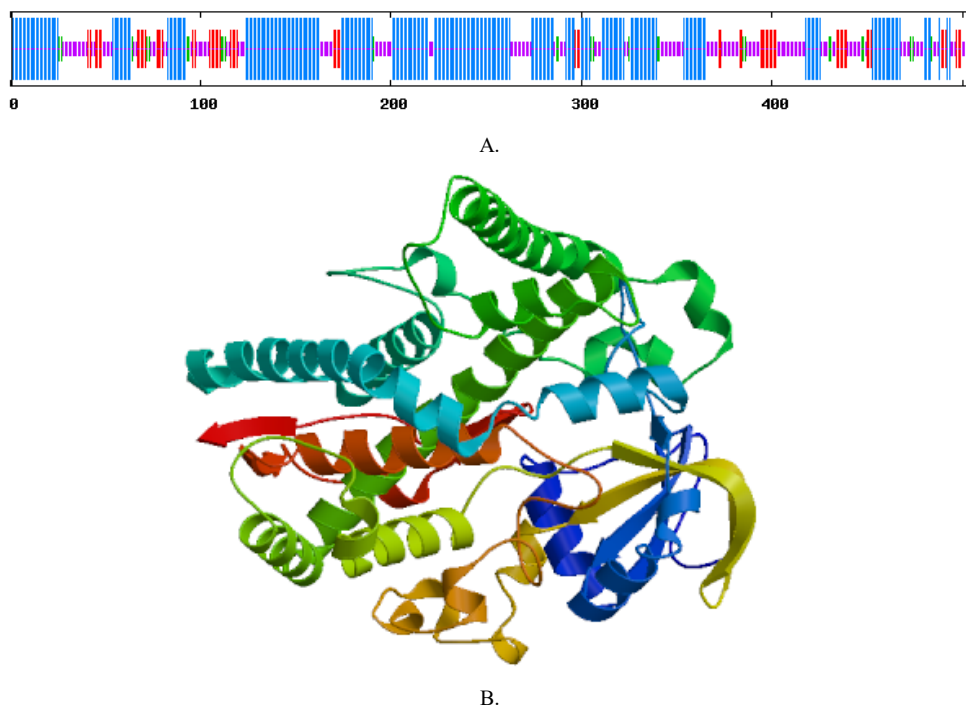


图 5 BsC4H 蛋白的二级和三级结构分析。A. 蓝色是 α -螺旋结构, 红色是线状结构, 绿色是 β 折叠结构, 橙色是无规则结构。B. SWISS-MODEL 预测的三级结构。

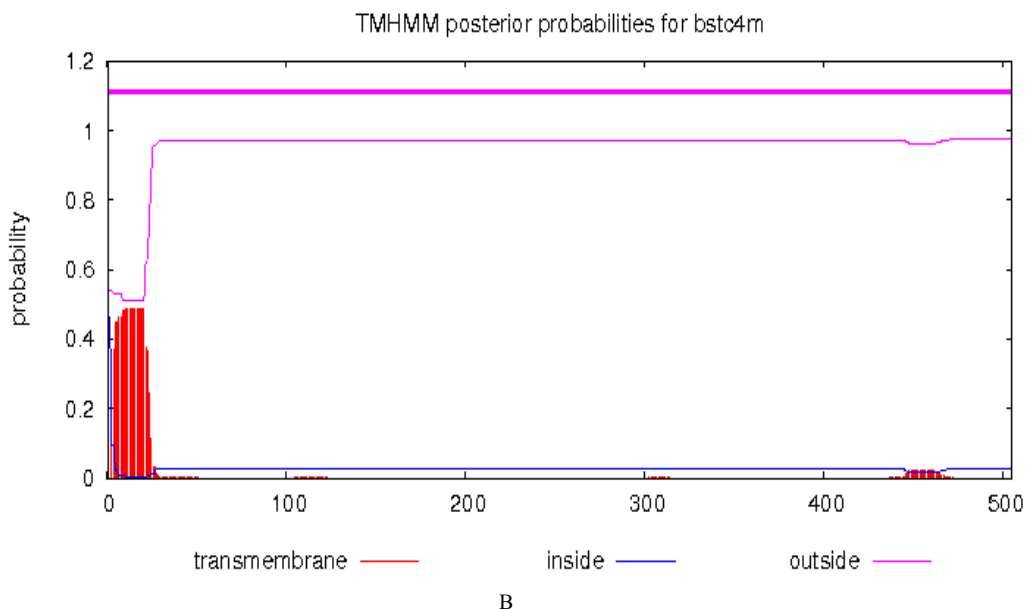
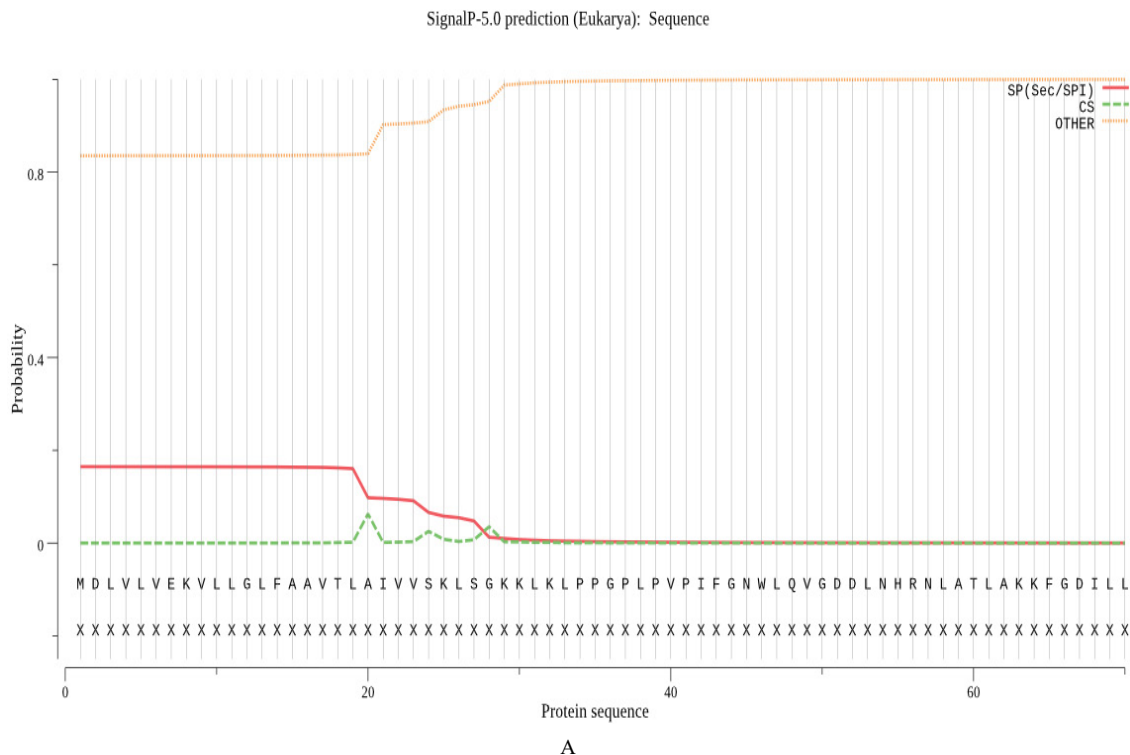


图 6 BsC4H 蛋白的亚细胞定位分析。A. 信号肽预测, B. 跨膜结构域分析。

3 讨论

近年来, 植物 *C4H* 基因的克隆和表达分析在甘薯^[8]、山葡萄^[10]、莲雾果实^[11]和芍药^[12]等植物中已有许多的文献报道, 而在白及中未有报道。通过对白及 *BsC4H* 基因核苷酸和氨基酸序列分析, 与其它物种相比较, 如图 2 所示 *BsC4H* 基因与其它物种具有非常高的覆盖率, 所以 *BsC4H* 基因具有高度的保

守性。通过 *BsC4H* 蛋白进化树分析, 发现 *BsC4H* 与铁皮石斛处于同一分支上, 亲缘关系非常接近, 由此可以推断 *BsC4H* 与铁皮石斛的功能与作用机制具有更高的相似性。

对 *BsC4H* 蛋白的结构分析主要是细胞色素 P450 酶结构域, 还有个半胱氨酸血红素铁结合位点, 通过 GenBank 能证明它的保守结构域。在 *BsC4H* 结

构预测中主要是 α -螺旋结构与无规则卷曲结构元件组成, 蛋白二级结构中 α -螺旋结构最经典, 在抗菌的效应研究中也发现与 α -螺旋结构有关, 而与其抑制剂的活性位点没有关系^[13], 而它的三级结构为单体结构。

Emmanuel Chekwube Ossai^[14]在亲水性材料上关于亲水性有所研究; 而在其它领域中认为亲水性材料与免疫调节有关, 亲水性材料的免疫调节作用与其水合特性及其在复合介质中抵抗非特异性结合的能力密切相关^[15]; 根据一些研究表明自组装的生物过程与亲水性的相互作用是有关的^[16], 这些功能是否能够体现在更多的亲水性蛋白上, 需要进一步的深层次研究。

植物中的黄酮类化合物主要是在内质网上生成的, 由一些与黄酮类物质合成有关的调控酶作用其中, 以多复合酶体的形式在苯丙烷中参与电子信号的传递并进行黄酮类物质的合成^[17]。通过亚细胞定位发现白及 *BsC4H* 蛋白定位于内质网上, 其催化反式肉桂酸生成对香豆酸, 是影响苯丙素代谢能力的关键因素^[18]。结合 *BsC4H* 信号肽与跨膜结构分析, 其不具有信号肽及跨膜结构。

本研究通过对白及 *BsC4H* 基因进行生物信息学分析探究该基因的结构与功能, 以进一步对研究白及的其他作用机制和功能提供理论基础, 也对后续研究其它植物的 *BsC4H* 基因提供基础依据和资料。

参考文献

- [1] 万大群, 赵仁全, 刘海, 等. 白及的成分、药理作用和临床应用研究进展[J]. 中国药业, 2017,26 (2):93-96.
- [2] 沈金好. 中药白及药理作用及临床应用进展[J]. 养生保健指南, 2018, (47):279.
- [3] 董永成, 王启斌, 高蕾, 等. 中药白及外用制剂研究进展[J]. 中国药师, 2019, 22(1):133-136.
- [4] 车玉红, 杨波, 郭春苗, 等. 槭寄生 *C4H* 基因的克隆、序列分析及表达[J]. 经济林研究, 2020, 38(2):1-8.
- [5] 李敏奇, 谢凌志, 陈永正, 等. 高通量筛选策略在细胞色素 P450 单加氧酶定向进化中的应用[J]. 合成化学, 2020, 28(4):371-384.
- [6] 黄利娜, 吴光斌, 匡凤元, 等. 莲雾果实 *C4H* 基因的克隆及在 NO 处理下的表达分析[J]. 集美大学学报(自然科学版), 2020, 25(2):105-112.
- [7] Showkat Ahmad Bhawani, Nursuhailah Binti Suhaili, Rachel Marcella Roland, et al. Template Assisted Synthesis of Molecularly Imprinted Polymer for the Extraction of p-Coumaric Acid [J]. Asian Journal of Chemistry, 2020, 32(9):2342-2346.
- [8] 张华玲, 刘绪, 杨春贤, 等. 甘薯肉桂酸-4-羟化酶基因克隆及序列分析[J]. 广西植物, 2018,38 (4):501-508.
- [9] Delin Xu, Hongbo Chen, Murat Aci, et al. De Novo Assembly, Characterization and Development of EST-SSRs from *Bletilla striata* Transcriptomes Profiled throughout the Whole Growing Period. Plos ONE, 2018,13(10): e0205954.
- [10] 陈蒙, 刘海峰. 山葡萄 *C4H* 基因的克隆表达及遗传转化分析[J]. 西南大学学报(自然科学版), 2019, 41(10):11-21.
- [11] 黄利娜, 吴光斌, 匡凤元, 等. 莲雾果实 *C4H* 基因的克隆及在 NO 处理下的表达分析[J]. 集美大学学报(自然科学版), 2020, 25(2):105-112.
- [12] 陈德伟, 王昕, 李红叶, 等. 芍药 *C4H* 基因的克隆及其在钙处理下的表达特性分析[J]. 分子植物育种, 2018, 16(9):2795-2801.
- [13] 崔晓东, 王婷芬, 杜晶晶, 等. 苦荞 α -螺旋发夹抗菌肽的分子改造抑制植物真菌活性[J]. 中国生物化学与分子生物学报, 2019, 35(2):196-204.
- [14] Emmanuel Chekwube Ossai, Takashi Kuroiwa, Kaname Horikoshi, et al. Lipid Vesicle Preparation Using W/O/W Multiple Emulsions via Solvent Evaporation: The Effect of Emulsifiers on the Entrapment Yield of Hydrophilic Materials [J]. Journal of the American Oil Chemists' Society, 2019, 96(12):1-12.
- [15] Li B, Xie J, Yuan Z, et al. Mitigation of Inflammatory Immune Responses with Hydrophilic Nanoparticles. Angew Chem Int Ed Engl. 2018, 57(17):4527-4531.
- [16] Rhys NH, Soper AK, Dougan L. Hydrophilic Association in a Dilute Glutamine Solution Persists Independent of Increasing Temperature. J Phys Chem B. 2015,119 (51):15644-15651.
- [17] 刘宽亮, 赵志常, 高爱平, 等. 芒果 *C4H* 基因的克隆及其表达分析[J]. 江苏农业科学, 2017, 45(14):8-12.
- [18] Kong JQ, Lu D, Wang ZB. Molecular cloning and yeast expression of cinnamate 4-hydroxylase from *Ornithogalum saundersiae* baker. Molecules. 2014, 19(2):1608-1621.

收稿日期: 2021 年 4 月 26 日

出刊日期: 2021 年 6 月 30 日

引用本文: 赵永彬, 黄恻隐, 徐程, 钱露, 李清清, 何毅怀, 徐德林, 白及反式-肉桂酸 4-单加氧酶基因 *C4H* 的序列分析[J]. 细胞与分子生物学研究, 2021, 1(1): 4-11.

DOI: 10.12208/j.ijcmbr.20210002

检索信息: 中国知网 (CNKI Scholar)、万方数据 (WANFANG DATA) 等数据库收录期刊

版权声明: ©2021 作者与开放获取期刊研究中心 (OAJRC) 所有。本文章按照知识共享署名许可条款发表。 <http://creativecommons.org/licenses/by/4.0/>



OPEN ACCESS