



吴亚楠,李贺,苏倩,等.沙棘 PAL 家族基因的生物信息学分析[J].黑龙江农业科学,2019(4):15-17.

沙棘 PAL 家族基因的生物信息学分析

吴亚楠,李贺,苏倩,王振铭,刘景,时一平
(大连民族大学 环境与资源学院,辽宁 大连 116600)

摘要:苯丙氨酸解氨酶(phenylalanine ammonia lyase, PAL)是黄酮合成代谢途径中的关键酶,对植物中黄酮类物质的积累有重要的作用。为研究 PAL 家族基因及其编码蛋白的结构特征,本研究以前期获得的沙棘转录组测序数据为基础,通过功能注释与分析获得了 3 个 PAL 家族成员基因,分别命名为 *HrPAL1*、*HrPAL2* 和 *HrPAL3*。生物信息学分析结果表明:沙棘 PAL 家族基因均含有完整的开放阅读框,编码以 α -螺旋和无规卷曲为主要二级结构的亲水蛋白,其中 *HrPAL1* 是具有信号肽的分泌性蛋白。3 种 *HrPAL* 蛋白在三级结构上有不同程度的相似性。

关键词:沙棘;生物信息学;黄酮合成;关键酶基因

沙棘是我国重要的生态经济植物,分布广泛,资源丰富。沙棘含有丰富的生物活性物质,广泛应用于食品、医药、化妆品和保健品等领域,具有很高的经济价值^[1-2]。作为主要活性成分的沙棘黄酮药理活性广泛,是一种十分宝贵的药用资源^[3]。沙棘果实、叶片及全株均含有黄酮,尤其是叶片中含有大量黄酮^[4]。PAL(苯丙氨酸解氨酶)家族是黄酮合成代谢途径中的关键酶,是进入苯丙烷代谢途径的第一个酶,直接影响黄酮代谢的信号通路的调控^[5]。因此了解沙棘 PAL 家族基因的序列及其编码蛋白质的结构与性质对于沙棘黄酮的有效开发具有重要的现实意义。

本研究以前期获得的沙棘转录组数据为基础,筛选 PAL 家族基因成员并进行生物信息学分析,对沙棘 PAL 家族基因及编码蛋白的结构进行预测,旨在为沙棘黄酮合成代谢的调控机理研究提供理论依据。

1 材料与方法

1.1 沙棘 PAL 蛋白相关基因的筛选

对前期的沙棘转录组测序数据进行一系列分析,包括基因表达与注释、GO 分析、KEGG 代谢通路分析等,筛选差异表达基因,基于差异表达基因进行 GO 功能显著性富集分析和 Pathway 显著性富集分析,筛选出蛋白质合成途径中的沙棘

PAL 家族基因序列。

1.2 生物信息学分析

利用多种生物信息软件和在线工具对目的基因序列及其编码蛋白的结构进行分析。登录 NCBI 网站,利用在线工具 ORF finder 确定沙棘 PAL 家族核酸序列的开放阅读框,使用 DNASTAR 软件将核苷酸序列转换成氨基酸序列。利用 ExPASy 网站分析编码蛋白的理化性质。使用 SignalP^[6]和 TMHMM 2.0 Server^[7]对蛋白的信号肽序列和跨膜区进行预测。利用 SOPMA 和 Phyre2^[8]预测沙棘 PAL 蛋白的二级结构并对其三级结构进行建模。

2 结果与分析

2.1 *HrPAL* 家族基因的注释及生物信息学分析

对转录组测序数据进行注释与分析,筛选出沙棘 PAL 家族基因中的 3 个成员,分别命名为 *HrPAL1*、*HrPAL2* 和 *HrPAL3*。利用 NCBI 网站上的 ORF finder 工具,分析这 3 个成员的核苷酸序列,结果表明,3 种 *HrPAL* 基因均具有完整的开放阅读框(open reading frame, ORF),长度分别为 1 341、1 305 和 2 160 bp,编码 446、434 和 719 个氨基酸(表 1),这表明 3 种 *HrPAL* 基因均有编码完整基因的独立性。

2.2 *HrPAL* 家族蛋白的理化特性

表 2 显示了 3 种 *HrPAL* 蛋白的主要理化性质,其分子式相似,不稳定指数均为 II 级,表现为稳定蛋白,疏水性为负值,表明全部为亲水性蛋白。预测 3 种 *HrPAL* 基因所编码蛋白的分子量分别为 49.55、47.12 和 78.81 kD,理论等电点为 5.72、6.91 和 6.05。

收稿日期:2018-11-02

基金项目:2017 年度大连民族大学大学生创新训练项目(201712026550)。

第一作者简介:吴亚楠(1997-),女,在读学士,从事生物技术研究。E-mail:wuyan_1997@163.com。

通讯作者:李贺(1975-),女,硕士,副教授,从事植物遗传育种研究。E-mail:lihe@dlnu.edu.cn。

表 1 *HrPAL* 家族基因核酸序列分析
Table 1 Analysis of nucleic acid sequence of *HrPAL* gene family

基因 Gene	标签 Label	链 Strand	阅读框 Reading frame	长度 Length (nt aa)
<i>HrPAL1</i>	ORF1	+	2	1341 446
<i>HrPAL2</i>	ORF9	-	2	1305 434
<i>HrPAL3</i>	ORF13	-	2	2160 719

2.3 *PAL* 家族蛋白信号肽预测

应用 SignalP. 4. 1 对 *HrPAL* 家族蛋白进行

表 2 *HrPAL* 家族基因编码蛋白的理化特性分析

Table 2 Analysis of physicochemical properties of *HrPAL* gene family

基因 Gene	分子式 Formula	氨基酸 数 AAs	相对分子质量/kD Molecular weight	等电点 PI	疏水性 Hydrophobicity	不稳定指数 Instability index
<i>HrPAL1</i>	C ₂₁₉₈ H ₃₅₀₅ N ₅₉₃ O ₆₆₉ S ₁₉	446	49. 55	5. 72	-0. 254	39. 86(Ⅱ)
<i>HrPAL2</i>	C ₂₀₇₃ H ₃₃₄₇ N ₅₈₇ O ₆₂₈ S ₁₈	434	47. 12	6. 91	-0. 137	31. 60(Ⅱ)
<i>HrPAL3</i>	C ₃₄₇₁ H ₅₅₅₁ N ₉₆₃ O ₁₀₆₅ S ₃₁	719	78. 81	6. 05	-0. 234	33. 04(Ⅱ)

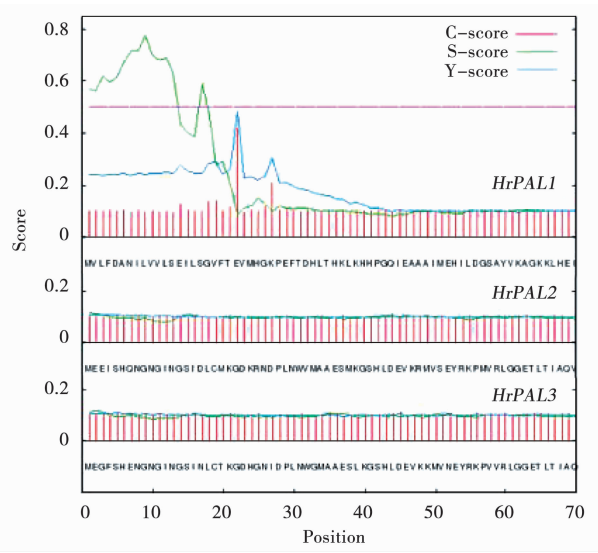


图 1 *HrPAL* 蛋白的信号肽预测

Fig. 1 Signal peptide prediction of *HrPAL* protein

2.4 *HrPAL* 家族蛋白的结构预测

应用在线工具 SOPMA 预测 *HrPAL* 家族蛋白的二级结构,结果如表 3 所示,3 种 *HrPAL* 蛋白的二级结构十分相似,均由 α -螺旋、 β -折叠和无规卷曲三种结构元件组成,其中 α -螺旋所占比例最高(49. 4%~57. 6%),其次为无规卷曲(40. 4%~42. 3%), β -折叠所占比例最少(<10%)。根据蛋白结构类型的分类方法,3

信号肽预测和分析,如图 1 所示,S 值表示信号肽氨基酸数,C 值表示剪切位点,Y 为 S 和 C 的综合值。结果表明,*HrPAL1* 蛋白具有 1 个明显的信号肽,为分泌性蛋白;*HrPAL2* 和 *HrPAL3* 蛋白的 C 值和 Y 值均小于 1,S 值无明显变化,且无明显的信号肽特征,表明这两种蛋白均不含信号肽,属于非分泌蛋白(图 1)。

利用 TMHMM 2. 0 Server 对 *HrPAL* 蛋白的跨膜区进行预测,结果表明,*HrPAL* 蛋白均无跨膜结构域,非跨膜蛋白。预测结果与疏水性分析结果一致,表明 3 个蛋白均为亲水蛋白。

种 *HrPAL* 蛋白均属于 mixed 蛋白,二级结构相似性较高。*HrPAL* 家族蛋白无规卷曲结构所含比例非常高,这种结构常受侧链影响从而构成蛋白质的活性部位^[9]。

表 3 *HrPAL* 蛋白的二级结构预测

Table 3 Secondary structure prediction of *HrPAL* protein

蛋白 Protein	α -螺旋 α -helix	β -折叠 β -sheet	无规卷曲 Random coil
<i>HrPAL1</i>	367(57. 6%)	21(2. 0%)	234(40. 4%)
<i>HrPAL2</i>	253(49. 4%)	57(9. 2%)	255(41. 4%)
<i>HrPAL3</i>	334(53. 1%)	35(4. 6%)	267(42. 3%)

3 种 *HrPAL* 蛋白的三维结构建模如图 2 所示,可见 *HrPAL* 家族的蛋白分子结构和二级结

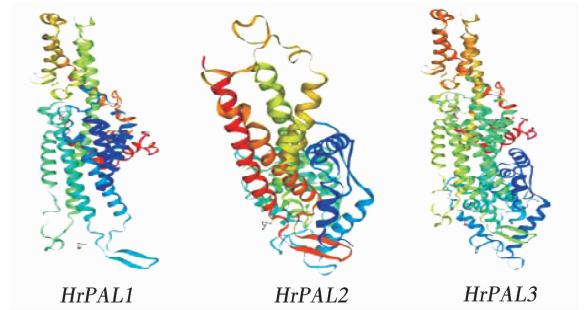


图 2 *HrPAL* 蛋白的三级结构预测

Fig. 2 Predicted tertiary structures of *HrPAL* protein

构的预测结果相一致, α -螺旋数量最多,其次为无规卷曲,二者所占比例非常高,而 β -折叠所占比例极低。另外,*HrPAL* 蛋白的三维结构之间具有不同程度的相似性。

3 结论与讨论

植物黄酮是一种有效的活性氧清除剂和抗氧化剂,具有较强的清除自由基的能力,因而具有抗衰老、增强免疫、调节血糖、抑制肿瘤生长和抗心血管疾病等多种功能,药理活性非常广泛^[3]。植物 *PAL* 基因为多拷贝多家族基因,近年来,越来越多的 *PAL* 基因在不同物种中被发掘^[10-11]。

沙棘 *HrPAL* 家族编码蛋白均为亲水性蛋白,这与孙宇蛟等^[6]对葫芦科植物 *PAL* 家族基因的研究结果相符。二级和三级结构分析表明,*HrPAL* 家族基因编码蛋白的二级结构均由 α -螺旋、 β -折叠和无规卷曲三种结构组成,其中 α -螺旋和 β -折叠所占比例较高,这与李义红等^[12]对梨苯丙氨酸解氨酶基因的生物信息学研究结果相符。本研究依托生物信息学手段,对沙棘的 *PAL* 家族基因及其蛋白质结构进行了初步的分析和预测,结果可为沙棘 *PAL* 基因进一步的挖掘、黄酮合成代谢的调控及品种改良提供新的理论依据。

参考文献:

- [1] Bal M L, Meda V, Naik N S, et al. Sea buckthorn: A potential source of valuable nutrients for nutraceuticals and cosmeceuticals[J]. Food Research International, 2011, 44: 1718-1727.
- [2] Chen C, Xu X M, Chen Y, et al. Identification, quantification

and antioxidant activity of acylated flavonol glycosides from sea buckthorn (*Hippophae rhamnoides* ssp. *sinensis*) [J]. Food Chemistry, 2013, 141: 1573-1579.

- [3] 张祚, 冉丽霞, 万方琼, 等. 沙棘叶总黄酮的提取法与药理作用研究进展[J]. 中国临床药理学杂志, 2018, 34(9): 1122-1124.
- [4] 秦莉, 程文杰, 王军扬, 等. 四种沙棘枝叶总黄酮含量的比较研究[J]. 家畜生态学报, 2013, 43(2): 45-48.
- [5] 孙宇蛟, 陈欣, 崔浩楠, 等. 葫芦科 *PAL* 基因家族生物信息学分析及甜瓜 *PAL4* 基因克隆[J]. 分子植物育种, 2018, 16(5): 4910-4920.
- [6] Petersen T N, Brunak S, von Heijne G, et al. SignalP 4.0: Discriminating signal peptides from transmembrane regions[J]. Nature Methods, 2011, 8(10): 785-786.
- [7] Moller S, Croning M D, Apweiler R. Evaluation of methods for the prediction of membrane spanning regions[J]. Bioinformatics, 2001, 17(7): 646-653.
- [8] Kelley L A, Sternberg M J. Protein structure prediction on the web: A case study using the Phyre server[J]. Nature Protocols, 2009, 4(3): 363-371.
- [9] 马帅, 王勤, 李晓琴. α/β 类蛋白质折叠类型的分类方法研究[J]. 生物信息学, 2014, 12(2): 123-132.
- [10] Huang J, Gu M, Lai Z, et al. Functional analysis of the *Arabidopsis* *PAL* gene family in plant growth, development, and response to environmental stress[J]. Plant Physiology, 2010, 153(4): 1526-1538.
- [11] Shang Q M, Li L, Dong C J. Multiple tandem duplication of the phenylalanine ammonia-lyase genes in *Cucumis sativus* L. [J]. Planta, 2012, 236(4): 1093-1105.
- [12] 李义红, 于凤鸣, 张立彬. 梨苯丙氨酸解氨酶基因的生物信息学分析[J]. 安徽农业科学, 2012, 40(32): 15569-15574, 15580.

Bioinformatics Analysis of *PAL* Gene Family in Sea Buckthorn

WU Ya-nan, LI He, SU Qian, WANG Zhen-ming, LIU Jing, SHI Yi-ping

(College of Environment and Resources, Dalian Minzu University, Dalian 116600, China)

Abstract: Phenylalanine ammonia lyase(PAL) is the key enzyme of flavonoid synthesis metabolism. It plays an important role in accumulation of flavonoids in plants. We studied the nucleotide sequences and structural characteristics of encoded proteins of *PAL* family genes. According to our previous RNA Sequencing data of sea buckthorn and subsequent analysis of functional annotation, a total of three genes were characterized to encode for the *PAL* family, named as *HrPAT1*, *HrPAT2* and *HrPAT3*, respectively. The results of bioinformatics analysis showed that all the identified *HrPAL* genes contain complete open-reading frame and encode hydrophilic proteins with secondary structure mainly including α -helix and irregular crimp. *HrPAL1* is a secretory protein with one signal peptide. Among three *HrPAL* proteins, there is different degrees of similarity in tertiary structure.

Keywords: sea buckthorn; bioinformatics; flavonoid synthesis; key enzyme