

• 药材与资源 •

大花红景天尿苷二磷酸葡萄糖基转移酶基因组 DNA 的克隆及分析

常 凯^{1,2}, 马丽利^{1,2}, 王亚雄^{1,2}, 兰小中³, 廖志华^{1,2*}

1. 西南大学生命科学学院 重庆市甘薯工程技术研究中心, 重庆 400715

2. 西南大学 三峡库区教育部生态环境教育部重点实验室, 重庆 400715

3. 西藏农牧学院 生物技术教研室, 西藏 林芝 860000

摘要: **目的** 克隆大花红景天 *Rhodiola crenulata* 尿苷二磷酸葡萄糖基转移酶基因组 DNA, 并利用生物信息学对其编码区与启动子进行分析。**方法** 以大花红景天为样品, 采用优化的 CTAB 法提取大花红景天基因组 DNA, 并利用 hiTAIR-PCR 技术经过 3 次步步移获得大花红景天 UDPGT (RcUDPGT) 基因 DNA 序列, 并利用生物信息学对其进行分析。**结果** 克隆得到 RcUDPGT 基因的 DNA 序列 2 977 bp, 其中包含 1 499 bp 的编码区域 (包括 82 bp 的内含子序列) 和 1 500 bp 的编码区 5' 上游侧翼域序列 (包括启动子区域)。生物信息学分析证明该序列为 UDPGT 基因且与其他植物 UDPGT 同源, 具有亲水性且定位于细胞质中, 三级结构预测表明该基因编码的蛋白具有 UDPGT 功能结合位点。启动子分析表明其存在光响应、厌氧响应、热响应、低温响应、压力与防御响应和花色素苷合成响应元件等。**结论** 克隆的 RcUDPGT 基因结构完整, 具有典型的功能结合位点, 为红景天分子生物学与代谢工程研究提供参考。

关键词: 大花红景天; 红景天苷; 尿苷二磷酸葡萄糖基转移酶; 基因组; 启动子

中图分类号: R282.12 文献标志码: A 文章编号: 0253-2670(2013)06-0736-08

DOI: 10.7501/j.issn.0253-2670.2013.06.020

Genomic DNA cloning and analysis of uridine diphosphate-glucosyltransferase gene from *Rhodiola crenulata*CHANG Kai^{1,2}, MA Li-li^{1,2}, WANG Ya-xiong^{1,2}, LAN Xiao-zhong³, LIAO Zhi-hua^{1,2}

1. Chongqing Engineering and Technology Research Center for Sweetpotato, School of Life Sciences, Southwest University, Chongqing 400715, China

2. Key Laboratory of Eco-environments in Three Gorges Reservoir Region, Ministry of Education, Chongqing 400715, China

3. Department of Biotechnology, Tibet Agricultural and Animal Husbandry College, Nyingchi 860000, China

Abstract: Objective To obtain the genomic DNA sequence of uridine diphosphate (UDP)-glucosyltransferase (GT) gene from *Rhodiola crenulata* and to analyze its DNA sequence at the level of bioinformatics. **Methods** The optimized CTAB method was used to extract the genomic DNA from *R. crenulata*, and after three times genomic walking, the genomic DNA sequence of UDPGT in *R. crenulata* (RcUDPGT) was obtained by hiTAIR-PCR. The DNA sequence was analyzed by bioinformatics method. **Results** The 2 977 bp DNA sequence of RcUDPGT was obtained, which contained the 1 499 bp gene sequence (including the 82 bp intron sequence) and 1 500 bp 5' upstream flanking sequence of coding region (including promoter sequence). The bioinformatic analysis showed that the RcUDPGT was hydrophilic, located in the cytoplasm, and had high homology with UDPGTs of other plants. The tertiary structure of RcUDPGT indicated that protein had UDPGT functional sites. Promoter analysis showed that it contained *cis*-acting elements responding to light, heat, pressure, temperature, anaerobic, anthocyanins biosynthesis sequences, etc. **Conclusion** The structure of RcUDPGT gene is integral and has functional sites. The research provides the reference for the molecular biology and metabolic engineering of *R. crenulata*.

Key words: *Rhodiola crenulata* (Hook. F. et Thoms.) S. H. Fu; salidroside; uridine diphosphate-glucosyltransferase; genome; promoter

收稿日期: 2012-11-05

基金项目: 国家科技支撑计划项目 (2011BAI13B06); 重庆市科学技术委员会自然科学基金计划资助项目 (2010BB5296)

作者简介: 常 凯, 男, 四川成都人, 硕士研究生, 研究方向为药用植物分子生物学与代谢工程。Tel: (023)68367146 E-mail: ck8148@126.com

*通信作者 廖志华, 博士, 教授, 博士生导师。Tel: (023)68367146 E-mail: zhliiao@swu.edu.cn

大花红景天 *Rhodiola crenulata* (Hook. F. et Thoms.) S. H. Fu 是蔷薇目景天科 (Crassulaceae) 红景天属 *Rhodiola* L. 多年生草本植物, 具有肉质匍匐根状茎, 为珍稀药用植物, 因花、根、茎浸出液均为红色而得名, 主要生长于青藏高原海拔 4 500 m 以上的严寒、缺氧、强日照、无污染地带, 是红景天中的极品。大花红景天中红景天苷、酪醇、黄酮等成分的量是其他品种红景天的 5 倍^[1], 被誉为“雪域人参”。近年的研究表明, 红景天苷具有抗缺氧、抗辐射、抗寒冷、抗疲劳、抗病毒等多种显著功效, 还具有延缓机体衰老, 防止老年疾病的功效^[2], 是一种具有发展前景的适用于特殊地区开发的环境适应性药物。但天然药源植物野生资源数量锐减, 且红景天中红景天苷量较低 (0.5%~4.1%), 限制了工业化的生产, 红景天苷的产量远远不能满足市场的需求^[3]。因而对红景天代谢途径进行深入研究, 利用基因工程手段增加红景天中红景天苷的量成为缓解市场需求的有效途径, 尤其是对大花红景天相关工作的开展显得尤为重要。

尿苷二磷酸葡萄糖基转移酶 (uridine diphosphate-glucosyltransferase, UDPGT) 以尿苷二磷酸葡萄糖 (uridine diphosphoglucose, UDPG) 和酪醇为底物催化合成红景天苷。然而, 酪醇的具体代谢通路及调控方式仍不确定^[4]。关于酪醇来源主要有苯丙烷代谢途径和酪氨酸脱羧途径^[5-7], 其上游均起始于莽草酸途径^[8]。因而对大花红景天 UDPGT (RcUDPGT) 基因组 DNA 的克隆为研究上述两条途径对红景天苷生物合成影响的相关工作提供依据。本研究优化适用于西藏大花红景天基因组 DNA 提取的方法。利用 hiTAIR-PCR 方法^[9-10], 成功克隆 RcUDPGT 基因的 DNA 序列及启动子区域序列, 并进行生物信息学分析, 为大花红景天基因工程相关工作的开展提供参考。

1 材料与方法

1.1 大花红景天材料的获取

样品为采自西藏林芝的野生大花红景天 *Rhodiola crenulata* (Hook. F. et Thoms.) S. H. Fu, 由西藏农牧学院兰小中副教授鉴定。

1.2 大花红景天基因组 DNA 的提取

本实验采用经优化的 CTAB 法提取大花红景天基因组 DNA^[11]。取 3 g 大花红景天新鲜材料, 立即在液氮中研磨成粉末状。快速装入离心管, 加入 60 °C 水中预热过的 CTAB 提取液 9 mL, 混匀, 60 °C

水浴 40 min。加入 2/3 体积的氯仿-异戊醇 (24:1), 10 000 r/min 离心 5 min。取上清于 50 mL 离心管中, 加等体积的氯仿-异戊醇 (24:1), 12 000 r/min 离心 10 min, 重提 1 次。重取上清, 加 0.8 倍体积冰异丙醇, 缓缓颠倒离心管, 4 °C 静置 30 min 后, 12 000 r/min 离心 15 min。去上清后用 70% 乙醇洗 3 次, 将沉淀转到 8 mL 离心管中, 风干。加 1.5 mL TE 溶液, 加入 10 mg/mL RNaseA (Takara) 于 37 °C 酶解 1 h 后 12 000 r/min 离心 5 min 吸取上清, -20 °C 保存。

1.3 引物的设计与 DNA 扩增及序列分析

利用 hiTAIR-PCR 原理, 基于本实验室已克隆的 RcUDPGT 基因的 cDNA 序列 (GenBank 登录号: JX228125) 及其他物种 UDPGT 基因设计引物 sp1-1、sp1-2、sp1-3。sp1-1 分别与 LAD1-1、LAD1-2、LAD1-3 和 LAD1-4 进行热不对称 PCR 反应, 通常情况下, 其中至少有一种简并引物可以与特异性引物之间利用退火温度的差异进行热不对称 PCR 反应, 通过 3 次巢式 PCR 反应即可获取已知序列上游的一段序列。电泳后回收目的条带连接 T 载体 (Takara PMD19-T), 送华大基因测序后得到该段序列。该实验 1 次扩增获取的长度不能满足实验要求, 因而根据第 1 次步移获取的序列信息, 继续进行上游侧翼序列的扩增。

以大花红景天 DNA 为模板, 用表 1 中的引物进行 hiTAIR-PCR 扩增, 扩增体系为 25 μ L: 10 $\times$ ExTaq 缓冲液 2.5 μ L, dNTP (2.5 mmol) 2 μ L, 上下游引物各 1 μ L, ExTaq (5 U) 0.125 μ L (Takara), DNA 模板 2 μ L。扩增温度设置按照热不对称 PCR 原理, 每次步移应用 3 对引物分别进行 3 次扩增, 即预扩增, 第 1 次扩增和第 2 次扩增。预扩增 PCR 产物稀释 20 倍后作为下一步 PCR 的模板, 进行第 1 次扩增, 同样第 1 次扩增产物稀释 20 倍后作为第 2 次扩增的模板。hiTAIR-PCR 与其他的染色体步移技术一样, PCR 的目的产物大小存在不确定因素, 在该实验中, 为了避免小片段影响试验进度, 使用了抑制性 PCR 技术使得小片段被抑制从而增加大片段被扩增出来的可能性, 即在每组的第 1 次扩增引物的 5' 端添加一段 AC1 序列以达到抑制小片段的目的。

所测得的序列结果经 VectorNTI 拼接后再次扩增得到物理全长, 在利用 DNASTar 软件翻译为蛋白质后用生物信息学软件 (ExPASy ProtParam) 预测编码蛋白属性, TargetP 软件预测细胞定位信息。经

表1 hiTAIR-PCR 扩增引物
Table 1 Primers of hiTAIR-PCR

引物名称	引物序列 (5'→3')	退火温度 / °C
sp1-1	gtatttaacgaatttcgaacatctc	58
sp1-2	ACGATGGACTCCAGTCCGGCCgatttagaattgagatgttgcg	60
sp1-3	cattcaggaagaccagacac	60
sp2-1	gcttcgaaactgcctaaataggcct	60
sp2-2	ACGATGGACTCCAGTCCGGCCggcctcagcactttcaaaattctgcc	62
sp2-3	ctgcccttccaaagctcgtctg	63
sp3-1	acgtggtgttctagctgagactag	60
sp3-2	ACGATGGACTCCAGTCCGGCCgctgagactagaagaccttagca	58
sp3-3	ggaagtatccgaaagtaaacacgc	58
f-full	ctaattacaattgtgttcctatg	50
r-full	gtatttaacgaatttcgaacat	50
LAD1-1	ACGATGGACTCCAGAGCGGCCGC (G/C/A) N (G/C/A) NNNGGAA	60
LAD1-2	ACGATGGACTCCAGAGCGGCCGC (G/C/T) N (G/C/T) NNNGGTT	60
LAD1-3	ACGATGGACTCCAGAGCGGCCGC (G/C/A) (G/C/A) N (G/C/A) NNNCCAA	60
LAD1-4	ACGATGGACTCCAGAGCGGCCGC (G/C/T) (G/A/T) N (G/C/T) NNCGGT	60
AC1	ACGATGGACTCCAGAG	50

NCBI 上的 BLASTp 进行在线序列比对, 分析同源性。应用 CLUSTALX 和 MEGA5.0 软件基于 Neighbor-Joining 的原理构建进化树。在 SOPMA 网站上使用 MLRC 软件对 RcUDPGT 进行二级结构预测, Swiss-Model 网站上采用同源建模进行三级结构预测。用在线预测软件 BDGP 和 Plant CARE 软件对启动子进行分析。

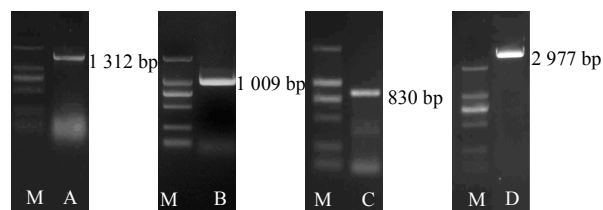
2 结果与分析

2.1 RcUDPGT 基因组 DNA 的获得

以大花红景天基因组 DNA 为模板, 经过 3 次步移得到 3 条目标片段, 分别为 1 312 bp (图 1-A)、1 009 bp (图 1-B) 和 830 bp (图 1-C), 拼接后经扩增得到 2 977 bp 的物理全长 (图 1-D)。并在 GenBank 上登录, 登录号为: JX228126。经过与 RcUDPGT 基因的 cDNA 比对分析, 得到 RcUDPGT 的 DNA 序列 1 477 bp, 得到编码框上游 (包含启动子区域) 序列 1 500 bp (图 2)。

2.2 RcUDPGT 基因编码区 DNA 的生物信息学分析

与本实验室已克隆的 RcUDPGT 基因 cDNA (GenBank 登录号: JX228125) 比对后发现在编码区域的 DNA 序列中有且只有一条含 82 bp 内含子序列, 且符合内含子左右边界规则。外显子比对结果显示该序列与来源于高山红景天的蛋白序列一致性



A-第1次扩增条带 (1 312 bp) B-第2次扩增条带 (1 009 bp)
C-第3次扩增条带 (830 bp) D-全长扩增条带 (2 977 bp)
M-Marker

A-the first amplification band (1 312 bp) B-the second amplification band (1 009 bp) C-the third amplification band (830 bp)
D-full-length amplification band (2977 bp) M-Marker

图1 RcUDPGT 基因 PCR 扩增条带电泳图

Fig. 1 Electrophoresis of amplification bands of RcUDPGT

达到 76%、与蓖麻 UDPGT 蛋白的一致性达到了 50%, 表明已成功克隆并得到 RcUDPGT 基因的 DNA 全长。预测该编码蛋白的相对分子质量为 5.093×10^4 , 等电点预测为 4.82。总原子数 4 950 个, 半衰期在哺乳动物中 30 h, 酵母中大于 20 h, 大肠杆菌中大于 10 h, 不稳定系数为 39.01, 脂肪系数 105.48, 总平均亲水性为 -0.048, 表明该蛋白为弱亲水性。TargetP 软件预测表明, RcUDPGT 无转运肽。

UDPGT 氨基酸序列同源性分析结果表明 RcUDPGT 与其他植物的 UDPGT 的同源性介于

1 CTAATTACAATTTGTTGTTTCCTATGTGGCATGGGTTTTGGCAAAACGACTTCAGAGTGT
61 CAAGCAGAAGTCAAGGAGCTGTCAACAAAAGATTGGAACACGGATGCTGCGCAAACTG
121 AAGAACGGAAGTTTCCAAAATTTGGAATTTCCAAATATCGCAGATTGGAGTCGGTTGAAGA
181 TCGGAGTGGTTGTCTAAAAGATAGAAGGAAAATATTCTGGAAAACATTATCAAATATCCA
241 CCAATATTTTCGAGTAACCAATATTTCAAGGATGATACTCAGAAGGTTATCAACAAGAA
301 GATTTGGAAGCTTGGAGAATATTTATGAAGATATTCCTAAATATTTGAGATAGAGTATCT
361 CAACCGAATAGGAAATCTGCCTTGCACATATAAATACAAGGCTAGGGTTTCCCTATTCACT
421 AGTTCGTCTAGAGGAGAAGTTATACACGATACACACTCAAGCCGAGGAGGAGGAGGAGAC
481 GACGGCTGGATAGTTTTCTACACCGGAAAACTAATCGGCGCTGAGGCAGTCTGTTCAGTG
541 TGCTAATTAGACATGTCCGTGGAACAAACCGCTGCACACACCACCTGACGCTACGCAT
601 TCGGGAGAACCGCGCTATCTTCGGGAGAAGTCACTGAGCTTCGGGAGAAGCCGTCTGGAC
661 TTCGGGAGAGGTCCTAGCATAGGTTCGGGAAGGAGTCTCGGACCTACCCACAGGGAATCGA
721 GGAGGAGCCGAGATCGCAAAATCGAGAAGAAAAGTGTAGAGGCTAGACAGTTACAAACGG
781 GAGTTGTGTAAGTGTAAAGACCATCTATACCTCTAAACTTTTTTGATAAGATTTGCTACCT
841 TTTGGGCGTGGGAAGCCCCCATAGTAGGCTGGTGTGCTGAACTGGGAAATATATCGTGTG
901 TGCTTGCCTGTTTACTTTTCGGATACCTTCTGCTAAGGCTTCTAGTCTCAGCTAGAACAC
961 CACGTGGATCTAGTCTTAAACAGACAACATGATTTCAAGAACATTTTGACACAAAATCAC
1 021 GGTCAACCAATGACCTGAGTCATGACGGCTAGGAAGGGCCGAAGGAGTTCTGGTCTCTCT
1 081 GAATTTAAAGTTTATTTTCTATAATTTAATGTTGTTTTGTTTTGTTTTGTTTTGAT
1 141 TGGTATAATTTAATATTTGTTAGACTAGACCTTTTTTTTTTTTTTTTTTTTTTGATAAGAG
1 201 ACTAGACCCATTTGGTTTCTTTAAAAAATATATAATTCAGACTAAAATATGAGATTT
1 261 TTCATTTTCTTGGAAGAAGTCAAAAAAATTTGCCATTTGCGTAAGCAGTGACATGATCAT
1 321 CTACTCGTCATGAATGATCAACCTCTCAAACTTCATTGTACTTCTTCATCTTCAATTAA
1 381 TTTTGAAAACTTTAAACACTCAACAAACCATCAACTGCATGTGTGGGGACTCATCACACT
1 441 TCACGCTCAGATCTATTTAAAGAACTCAACCAGTAACCATCTCCTAAGAAAACATAACCA
1 501 ATGGCGACGAAGAAAACTCACATTTCTTATCTCCCATACCCAAGTCAAGGCCACATCAAT
1 561 CCGATGCTCCAGTTCGCCAAACGCTCGCTCCAAGTCAGACAACTTCATCTTCACTTTG
1 621 CTCCTCCCCACCTCTCAAACTAAGTCCATAACACCGCGCATTTGGCTCCATAAACATCCAG
1 681 CCTATATCAGACGGAGCTTTGGAAGGGCAGAATTTTGAAAGTGCTGAGGCCTATTTTAGG
1 741 CAGTTTCGAAGCGTAGTACCTGGTAGCTTGGACGAATTGATTCGTCGCGAATTTGATCGA
1 801 GGCGATGAATTTAGCCTAAAGTGATTATTTATGACTCTTTCTTACCTTGGGCGCTGGAC
1 861 GTCGCTCATGGGAATGGATTGGATGGCGCGCTTTTTTCACTCAGTTTGTCTGTGTAGC
1 921 TCTCTTTACTTTCTTTACAAGGAAGGGAATTTGAGCTCGTCGGAGCTACTTTATGGTTG
1 981 CCGAGACTGGAACACGCGACTTGCTTTCGTTTATACGGGATAAAGAGAGCTCGTCGTCG
2 041 TGTCTGTTGGAGATGCTTGTGGATCAGTTTTCGAATTTGGACAAGGCTGAGTATGTTTTT
2 101 TTCAACACTTTTGACAACTTGAGTCTCAAGTAAGTTCATTTATTTCAAATTTGAGCTTT
2 161 GTAATTTCTAATTTAAGCTATTTAGATTCTAAAGCCTTGCTCAAAAATGGCAGATGGTGA
2 221 ATGGATGTCAAGCCAATGGCCGTTTAAACCGTTGGACCAACGATTCCATCGATGTATTT
2 281 GGATAAATATATCGAAGATGATACAAGCTATGGTCTCAACTTGTTCAAGCCAAATCGGCA
2 341 GACATGTCAGGACTGGCTAAGCGAAAAGAAAATCAGGTTCCGGTGATCTACGTATCATTTGG
2 401 GAGCATGGCCATTTTGGAGCAGGATCAAGTGTGTGAAATAGCTCAATGCCTTGAGAATCT
2 461 TCAAACTCATTTTATATGGTGGTCCGAGAAACCGAAATGACGAAATGGCCACTGAATT
2 521 TGTGGAATGGAATTTAAGGTGAGGCTTAGGCCTGGTGGTGACTTGGTGTAACCAAGTTAGA
2 581 TGTTTTGGCACATGAGGCAGTGGGCTGCTTCGTGACGCATTTGTTGGGTGGAACCTGACGTT
2 641 GGAGGCGTTGTGTTTCGGAGTGCCGATGGTGGGGATGCCGAATGGTCTGACCAGCCGAC
2 701 GAATGCTAAGTTTGTGGAGGATGTTTGGGGAGTTGGGGTGAGGACAATGGAAGATGAGAA
2 761 TGGGATTGTGACTGGTTTGGAGTTGGAGAAAATGTGTGAGGGCAGTTTGGAAAAAGGTGA
2 821 AAAGGGTGATGAGATTAGAAGAAATGCTGAGAAGATGAAGGATTGGGCTTTAAAGGCTAT
2 881 TGAAATTGATGGCAGCTCTGATTGTAATATTTGTCAAGTTTGTGTCTGGTCTTACCTGAAT
2 941 GATTTAGAATTGAGATGTTGCGAAATTCGTTAAATAC

黑底斜体表示启动子转录起始区域，黑底部分表示内含子序列，白底部分为外显子，转录起始点、起始密码子、终止密码子和内含子边界均用粗体表示
Promoter sequences were showed with black background and italic, intron sequences were showed with black background, exon sequences were showed with white background, the transcription start point, the start codon, termination codon, and intron boundaries were showed in bold font

图 2 RcUDPGT 基因组 DNA 序列

Fig. 2 Genomic DNA sequence of RcUDPGT

30%~75%，同源性较高（图 3）。从整体上看，保守区主要集中在中后部催化区域，而变异区集中在两端非催化区域，尤其以羧基端为甚，羧基末端变异区的氨基酸差别很大，并且长短不一。植物中的 UDPGT 含有相似的结构域，所有的 UDPGTC-末端都含有一个 44 个氨基酸的特征性区域，称为

PSPG 盒，被认为是 UDPGT 的标志性结合区域，在 RcUDPGT 氨基酸序列中同样具有 PSPG 盒^[12]。上述分析结果均表明本研究所获得的 DNA 序列确为 UDPGT 基因。

将 RcUDPGT 与来源于其他植物、动物、细菌的 UDPGT 进行进化树分析，结果表明，来源于植

物、动物、细菌的 UDPGTs 在发育进化树上聚为 3 支 (图 4), 并且在进化树上的距离与其所属类别的亲缘关系相符。

对 RcUDPGT 进行二级结构预测表明: RcUDPGT 包含 36.67% α -螺旋、14.67% β -片层和 48.66% 随意卷曲。对 RcUDPGT 蛋白进行预测建模 (图 5), 对比发现, RcUDPGT 与 2C1Z (MMDB ID: 86870)

的相似较高。RcUDPGT 包含两个结构域, 特别是在 C-端结构域的第 6 个 β -片层和随后的螺旋有助于 Trp 327 形成疏水性的平台, 随后 Gln 330 虽然没有直接相互作用于核苷酸糖, 但参与维持 Gln 353 在正确的方向结合核糖羟基^[13]。RcUDPGT 的蛋白质三维结构的预测能更好地了解活性蛋白发挥功能时的立体化学性质和活性位点的部位。

		301		360
<i>Rhodiola crenulata</i>	(258)	L	SERKSGSVIYVSFGSMAILEQDQVCEIAQCLENLQTHFIWVVR-----ETEMTKLPTE	
<i>R. sachalinensis</i>	(259)	L	CERRASSVIYVSFGSMAILKQEQLIEEIAKCLLENLQTRFIWVVR-----ETEMAKLPSE	
<i>Arabidopsis thaliana</i>	(283)	L	DSKPNSSVIYVSFGSAFFKNEQLFEIAAGLEASGTSFIWVVR---KTKDDREEWLPPEG	
<i>Brassica napus</i>	(274)	L	DSRPESSVYIISFGTVAYLKQEQEMEMAHGVLKTGLSFLWVIRLPLPLDKLETHVLPQE	
<i>Capsicum annuum</i>	(276)	L	DAKSPSSVYIISFGSVILKKEQAEIAYGLLNSGVNFWVIRPPTKLQNFDSLLEPSE	
<i>Carica papaya</i>	(281)	L	DSKQPSSVIYVSFGSLASLSEQMTLARGLQMSCDHFLWVVR-----DLEKLKPES	
<i>Isatis tinctoria</i>	(275)	L	DTKTDSSVIYISFGTLAVLSTROLVELCMALIQRPELWVITDKTYRIKEDGEETEEE	
Consensus	(301)	L	DSK SSVIYVSFGSVAILEKQEL EIA GL NS FLWVVR D E LP E	
		361		420
<i>Rhodiola crenulata</i>	(312)	F	VEWNLRSGLGLVVTWCNQLDVLAEAVGCFVTHCGWNSILEALCFGVPMVGMNPWS	
<i>R. sachalinensis</i>	(313)	F	VEWNLSSGLGLVVTWCNQLDILAHETVGCFTVTHCGWNSILEALCLGVPMVGMNPWS	
<i>Arabidopsis thaliana</i>	(340)	F	EEERVKGKGLMIIRGWAPQVLIDHQAATGGFVTHCGWNSLLEGVAAGLPMVTWPVGAE	
<i>Brassica napus</i>	(334)	L	KEASGKGLGKIVENCPEQVLHSSVACFVTHCGWNSILEALTSGVFPVCFPQWGD	
<i>Capsicum annuum</i>	(336)	F	LEKAG-D---RAKIVQWCPQEVLSHPVACFVTHCGWNSILEALSSGMPVLAFFPQWGD	
<i>Carica papaya</i>	(335)	F	KEETSDK---GLVVSWSPELVLAHKSVMGCFMTHCGWNSILEALSLGVPMVAMPQWTD	
<i>Isatis tinctoria</i>	(335)	S	IRSFREELDEIGMVVSQDQFSLKRSIGCYMTHCGWNSILESLVAGVPMVAMPQWTD	
Consensus	(361)	F	VE LGLVVSWCPL VLAH SVGCFVTHCGWNSILEAL AGVPMVAFPOWSD	

图 3 来自不同植物的 UDPGT 氨基酸序列多重比对

Fig. 3 Multi-comparison on amino acid sequences of UDPGTs from different plants

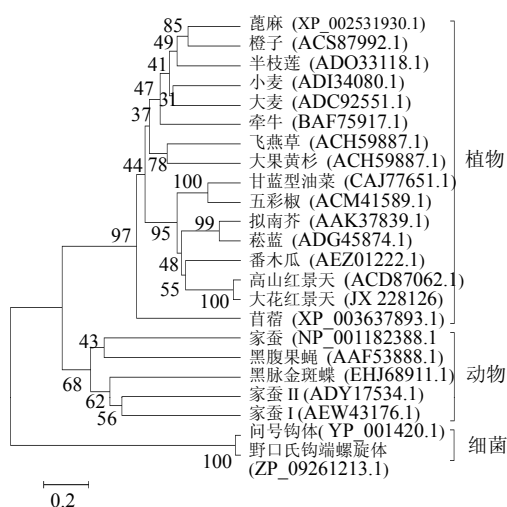
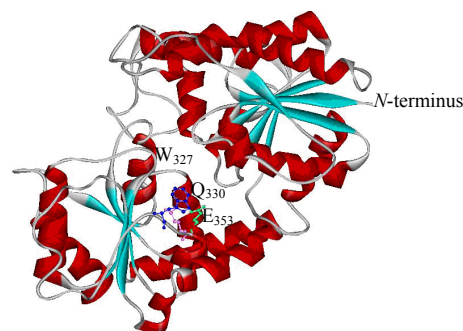


图 4 基于 MEGA 5.0 软件平台用 Neighbor-Joining 方法构建的 UDPGT 分子进化树

Fig. 4 Phylogenetic tree of UDPGT from different organisms constructed by Neighbor-Joining method on MEGA 5.0



α -螺旋用黑色螺旋状表示, β -片层用灰色板状表示, 随机卷曲用浅灰色绳状表示, 活性位点用不同颜色小球标注

The α -helix, the sheet, and the random coil were showed in black (helix-shaped), gray (plate-shaped), and light gray (rope-shaped), active sites were shown with balls of different colors, respectively

图 5 基于序列同源建模的 RcUDPGT 三维立体结构

Fig. 5 3D structure of RcUDPGT established by homology-based modeling

2.3 RcUDPGT 基因启动子区域序列的生物信息学分析

本研究得到起始密码子上游 1 500 bp 的序列, 将

得到的该段调控区序列分析预测, 在 1 354~1 404 bp 处存在可能的基础启动子区域分值为 0.97, 其序列为: ggaagaagtcaaaaaattgccatttgcgtaagcagtgacAtgatcatct。转录起始点在第 1 313 bp 处的 A^[14] (图 6)。RcUDPGT 基因核心启动子元件的预测发现该基因启动子含有 22 处核心启动子元件 TATA-box 和 43 处启动子近侧调节元件 CAAT-box^[15]。可以推断, 该序列含有基础

启动子区域。经过进一步的结构分析后, 在 RcUDPGT 基因启动子序列中分别发现了发育调控及环境响应等重要顺式作用元件。其中, ABA 响应元件与 SA 响应元件的 Matrix score 均达到 9, 表明在 RcUDPGT 基因的启动子中极有可能存在上述两个作用元件的位点; 同时还具有环境影响元件如光响应、温度响应元件等。表 2 显示作用元件的 Matrix score 值均大于 5。

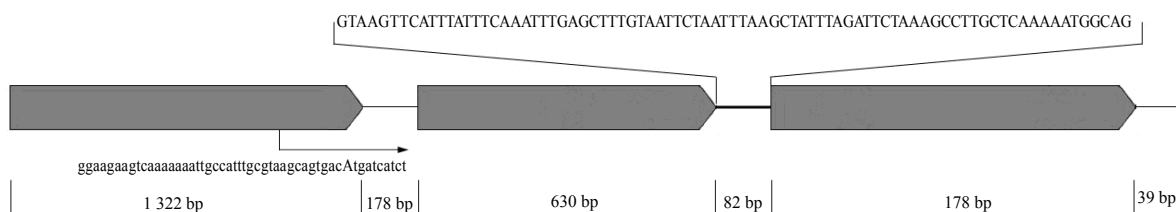


图 6 RcUDPGT 基因的 DNA 序列模式图

Fig. 6 DNA sequence mode of RcUDPGT gene

表 2 RcUDPGT 基因 5'端调控区序列的顺式作用元件预测

Table 2 Prediction of cis-acting elements in regulatory sequence in 5' terminal of RcUDPGT gene

顺式作用元件	功能名称	基序名称	核心序列	数量
发育调控元件	胚乳表达调控	Skn-1 motif	GTCAT	4
		GCN4-motif	TGTGTCA	2
激素响应	生物钟调控	Circadian	CANNNNATC	1
	生长素响应	TGA	AACGAC	1
	赤霉素响应	P-box	CCTTTTG	1
	MeJA 响应	GARE-motif	AAACAGA	1
		CGTCA-motif	CGTCA	3
	ABA 响应	TGACG-motif	TGACG	3
		ABRE	CCTACGTGGC	1
环境响应元件	SA 响应		CACGTG	1
	光响应	TCA-element	GAGAAGAATA	1
		ACE	AAAACGTTTA	1
		Box4	ATTAAT	1
		Box1	TTTCAAA	1
		G-Box	CACGTG	3
		SP1	CCGCCC	1
		厌氧响应	TGGTTT	2
		热响应	AAAAAATTTT	1
		低温响应	CCGAAA	1
		压力与防御响应	A/GTTTTCTC/TC/AC/A	2
		花色素苷合成响应	C/TAACGTG	2
			CGGTCA	1

3 讨论

红景天苷是红景天类植物最为重要的具有很高药用价值的天然产物, 其生物合成途径的分子生物学研究引起了科学家广泛的关注^[16]。这些研究大多集中在高山红景天 *R. sachalinensis*、狭叶红景天 *R. kirilowii*、喜马红景天 *R. himalensis*、帕里红景天 *R. phariensis* 等材料中, 但对于被《中国药典》2010 年版收录的大花红景天 *R. crenulata* 这一红景天植物中

最为优良的种质资源的分子生物学研究处于空白^[17]; 而且在高山红景天等样品的研究都集中于 cDNA 水平, 缺乏对于相关基因的基因组水平的研究。

糖基转移酶 (glycosyltransferases, GT) 几乎存在于所有生物中, 催化糖基转移反应, 是最重要的生物转化反应之一, 直接参与二糖、单糖苷、低聚糖等生物合成。UDP-葡萄糖是植物中最常见的糖基供体参与 UDPGT 的反应, 而 UDPGT 又与红景天

苷的生物合成直接相关。因此,对 RcUDPGT 基因的克隆工作显得尤为重要。

植物 UDPGT 催化的反应底物非常广泛,如次生代谢产物、植物激素等,在植物生长发育、合成、代谢、储存等方面具有重要作用^[18-19]。通常大部分的 UDPGT 为可溶性蛋白定位于细胞质中,目前未见 UDPGT 蛋白含有明确跨膜区的报道^[20-21]。本研究所克隆基因编码的蛋白经疏水性分析和 TargetP 预测不存在质体转运肽,同样是亲水性蛋白且定位于细胞质中。GT 的空间结构具有两种类型,一是含有两个紧密相连的 $\beta/\alpha/\beta$ 类 Rossmann 折叠区域的 GT-A 型;二是含有两个正对的 $\beta/\alpha/\beta$ 类 Rossmann 折叠区域的 GT-B 型。采用同源建模的方式对 RcUDPGT 蛋白进行预测建模,结合上述特征分析 RcUDPGT 空间结构为 GT-B 型^[22-23]。

通过本实验室相关研究发现 RcUDPGT 基因在大花红景天不同组织中的表达量具有显著性差异,因而克隆 RcUDPGT 基因组 DNA 及启动子这一工作有助于研究其调控机制和差异性表达机制,同时也为红景天苷生物合成途径的研究提供参考。在本实验中,对该基因的启动子区域进行生物信息学分析表明,不仅存在植物本身应有的发育元件和激素响应元件,还存在光响应、厌氧响应、热响应、低温响应、压力与防御响应,这些都可能与其已在遗传上适应了高寒多变的恶劣环境有关。

在本研究中发现 RcUDPGT 启动子序列中存在 MYB 的结合序列,MYB 类转录因子家族相关基因的调节活性是自然界中植物着色模式多变的主要原因,MYB 家族与类黄酮和花色苷生物合成途径调控紧密相关^[24]。该转录因子通过识别、结合结构基因启动子中顺式作用元件,激活花色苷生物合成途径一个或多个结构基因表达,从而有效启动花色苷生物合成^[25]。花色苷是植物中重要的次生代谢产物,其稳定性和在水中的溶解度均需要糖基化步骤的参与,这些花色苷合成响应元件极可能与大花红景天花、根、茎之浸出液均为红色有着紧密联系。因而在 RcUDPGT 的启动子中发现 MYB 的结合序列是对 UDPGT 基因与花色苷合成相关性研究的又一佐证^[26]。

利用基因工程遗传改良红景天,开展红景天苷代谢工程研究,进而利用细胞工程生产红景天苷具有很大优势,也是本课题组的最终目标。借鉴前人的研究思路,深入剖析红景天苷合成代谢的关键酶

的作用及调控机制,加以基因工程手段干预后可能会显著增加红景天苷的量^[27]。利用基因工程结合细胞工程是生产红景天苷的理想途径,红景天苷生物合成的分子生物学研究和细胞工程研究在这方面将起到决定作用,为最终实现红景天苷商业化生产奠定理论基础。

参考文献

- [1] 张祖荣,廖志华. 红景天苷的生物合成途径及生物技术研究进展 [J]. 中草药, 2010, 41(9): 1571-1574.
- [2] 李伟,黄勤妮. 红景天属植物的研究及应用 [J]. 首都师范大学学报: 自然科学版, 2003, 24(1): 55-59.
- [3] 祝顺琴,刘万宏,戴传云. 红景天苷生物合成的细胞与分子生物学研究进展 [J]. 食品科学, 2012, 33(3): 267-271.
- [4] 马兰青,柳春梅,于寒松,等. 红景天甙生物合成途径: 酪醇合成的起始反应及其糖基化 [J]. 生物工程学报, 2012, 28(3): 282-294.
- [5] Rippert P, Matringe M. Molecular and biochemical characterization of an *Arabidopsis thaliana* arogonate dehydrogenase with two highly similar and active protein domains [J]. *Plant Mol Biol*, 2002, 48(4): 361-368.
- [6] Landtag J, Baumert A, Degenkolb T, et al. Accumulation of tyrosol glucoside in transgenic potato plants expressing a parsley tyrosine decarboxylases [J]. *Phytochemistry*, 2002, 60(7): 683-689.
- [7] Ellis B E. Production of hydroxyphenylethanol glycosides in suspension cultures of *Syringa vulgaris* [J]. *Phytochemistry*, 1983, 22(9): 1941-1943.
- [8] Yao K, De Luca V, Brisson N. Creation of a metabolic sink for tryptophan alters the phenylpropanoid pathway and the susceptibility of potato to *Phytophthora infestans* [J]. *Plant Cell*, 1995, 7(11): 1787-1799.
- [9] Liu Y G, Chen Y L. High-efficiency thermal asymmetric interlaced PCR for amplification of unknown flanking sequences [J]. *Biotechniques*, 2007, 43: 649-656.
- [10] Liu Y G, Whittier R F. Thermal asymmetric interlaced PCR: Automatable amplification and sequencing of insert end fragments from PI and YAC clones for chromosome walking [J]. *Genomics*, 1995, 25: 674-681.
- [11] 宋国立,崔荣霞,王坤波,等. 改良 CTAB 法快速提取棉花 DNA [J]. 棉花学报, 1998, 10(5): 273-275.
- [12] Hughes J, Hughes M. Multiple secondary plant product UDP-glucose glucosyltransferase genes expressed in cassava (*Manihotes culenta* Crantz) cotyledons [J]. *DNA Seq*, 1994, 5(1): 41-49.
- [13] Offen W, Martinez-Fleites C, Yang M, et al. Structure of a flavonoid glucosyltransferase reveals the basis for plant

- natural product modification [J]. *EMBO J*, 2006, 25(6): 1396-1405.
- [14] Reese M G R. Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome [J]. *Comput Chem*, 2001, 26(1): 51-56.
- [15] Lescot M, Dhais P, Thijs G, *et al.* A database of plant *cis*-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences [J]. *Nucleic Acids Res*, 2002, 30(1): 325-327.
- [16] 明海泉, 夏光成, 张瑞钧, 等. 红景天研究进展 [J]. 中草药, 1988, 19(5): 37-42.
- [17] 中国药典 [S]. 一部. 2010.
- [18] Vogt T, Jones P. Glycosyltransferases in plant natural product synthesis: Characterization of a supergene family [J]. *Trends Plant Sci*, 2000, 5(9): 380-386.
- [19] Jones P, Vogt T. Glycosyltransferases in secondary plant metabolism: Tranquilizers and stimulant controllers [J]. *Planta*, 2001, 213(2): 164-174.
- [20] Li Y, Baldauf S, Lim E K, *et al.* Phylogenetic analysis of the UDP-glycosyltransferase multigene family of *Arabidopsis thaliana* [J]. *J Biol Chem*, 2001, 276(6): 4338-4343.
- [21] Jones P, Vogt T. Glycosyltransferases in secondary plant metabolism: Tranquilizers and stimulant controllers [J]. *Planta*, 2001, 213(2): 164-174.
- [22] Bourne Y, Henrissat B. Glycoside hydrolases and glycosyltransferases: Families and functional modules [J]. *Curr Opin Struct Biol*, 2001, 11(5): 593-600.
- [23] Lairson L L, Henrissat B, Davies G J, *et al.* Glycosyltransferases: Structures, functions, and mechanism [J]. *Annu Rev Biochem*, 2008, 77: 521-557.
- [24] Wang K L, Bolitho K, Grafton K, *et al.* An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in Rosaceae [J]. *BMC Plant Biol*, 2010, 10: 10-50.
- [25] Ramsay N A, Glover B J. MYB-bHLH-WD40 protein complex and the evolution of cellular diversity [J]. *Trends Plant Sci*, 2005, 10(2): 63-70.
- [26] Ogata J, Kanno Y, Itoh Y, *et al.* Plant biochemistry: Anthocyanin biosynthesis in roses [J]. *Nature*, 2005, 435: 757-758.
- [27] Yamada T, Matsuda F, Kasai K, *et al.* Mutation of a rice gene encoding a phenylalanine biosynthetic enzyme results in accumulation of phenylalanine and tryptophan [J]. *Plant Cell*, 2008, 20(5): 1316-1329.