

刺五加 *P450* 基因的筛选及其表达对皂苷含量的影响

宋 鑫, 崔明晖, 张朵朵, 张 杰, 林丽梅, 邢朝斌*

华北理工大学生命科学学院, 河北 唐山 063210

摘 要: 目的 筛选和鉴定刺五加 *Eleutherococcus senticosus* 中的细胞色素 *P450* 基因 (*EsP450*), 分析其进化特征, 探究其与刺五加总皂苷含量的相关性。方法 根据转录组测序结果, 筛选得到 *EsP450* 基因, 并对其进行生物信息学分析与适应性进化分析, 采用 qRT-PCR 法检测 *EsP450* 基因的表达量, 采用分光光度计测定刺五加的总皂苷含量。结果 筛选得到了 18 条 *EsP450* 基因。刺五加 *P450* 蛋白不存在跨膜区域, 二级结构以 α -螺旋和无规则卷曲结构为主。*EsP450* 基因不存在正选择位点。各 *EsP450* 间的表达量差异显著, 部分基因表达量差值在 10 倍以上, 7 条 *EsP450* 基因的表达与皂苷含量呈正相关关系 ($P < 0.05$)。结论 鉴定出与刺五加总皂苷含量存在正相关关系的 *EsP450* 基因, *EsP450* 基因在进化中受到纯净选择。

关键词: 刺五加; 细胞色素 *P450*; 生物信息学分析; 基因表达量; qRT-PCR

中图分类号: R282.12 文献标志码: A 文章编号: 0253-2670(2021)16-5012-09

DOI: 10.7501/j.issn.0253-2670.2021.16.026

Identification of *P450* gene of *Eleutherococcus senticosus* and its effect on saponin content

SONG Xin, CUI Ming-hui, ZHANG Duo-duo, ZHANG Jie, LIN Li-mei, XING Zhao-bin

College of Life Science, North China University of Science and Technology, Tangshan 063210, China

Abstract: Objective To screen and identify *P450* gene in *Eleutherococcus senticosus* (*EsP450*), analyze its evolutionary characteristics, and explore the correlation between *P450* gene and total saponins content of *E. senticosus*. **Methods** According to the results of transcriptome sequencing, *EsP450* was screened and analyzed by bioinformatics and adaptive evolution, qRT-PCR was used to detect the expression of *EsP450* and spectrophotometer was used to determine the total saponin content of *E. senticosus*. **Results** Eighteen *EsP450* were screened. There was no transmembrane region in *EsP450*, and the secondary structure was mainly α -helix and irregular curly structure. There was no positive selection site of *EsP450*. The expression of *EsP450* was significantly different. Seven *EsP450* were identified, which were positively correlated with the total saponin content ($P < 0.05$). **Conclusion** The *EsP450* were subjected to purifying selection pressure during evolution.

Key words: *Eleutherococcus senticosus* (Rupr.et Maxim) Maxim; cytochrome *P450*; bioinformatics analysis; gene expression level; qRT-PCR

刺五加 *Eleutherococcus senticosus* (Rupr. et Maxim.) Maxim. 是我国珍贵药用植物, 具有抗疲劳、抗缺氧、抗肿瘤等许多药理作用, 与人参一样, 被广泛用作“适应原”样药物^[1], 三萜类化合物是其主要活性成分之一。三萜皂苷类化合物在生物体中均需通过依赖甲羟戊酸的类异戊二烯途径进行合成^[2], 合成过程分为前体形成、骨架构建以及后修饰 3 个阶段。

其中, 后修饰过程是决定形成何种皂苷和其产量的关键环节, 细胞色素 *P450* (cytochrome *P450* enzyme, *P450*) 即在此过程中负责三萜烯骨架的羟基化、氧化和糖基化等复杂的后修饰作用, 是三萜皂苷类化合物生物合成中的关键酶^[3]。

P450 是第一组被归类于“超级家族”的酶类, 在动植物、细菌和真菌等细胞中广泛分布^[4], 它

收稿日期: 2021-01-23

基金项目: 河北省自然科学基金生物医药联合基金项目 (H2020209302); 河北省教育厅资助科研项目 (ZD2019075); 国家自然科学基金项目 (31570683); 河北省大学生创新创业训练计划 (S202010081016); 华北理工大学大学生创新创业训练计划 (R2020044); 河北省大学生科技创新能力培育专项项目 (2021H020911)

作者简介: 宋 鑫 (2000—), 男, 研究方向为生物信息在药用植物中的应用。

*通信作者: 邢朝斌 (1975—), 男, 教授, 研究方向为分子生物学。Tel/Fax: (0315)8805590 E-mail: xingzhb@ncst.edu.cn

是自然界中最大且最古老的家族。自从 1958 年从小鼠肝脏微粒体中分离到第一个细胞色素 P450, 以及 1989 年从鳄梨 *Persea americana* Mill. 中克隆出第 1 个植物 P450 基因序列^[5]以来, P450 基因就因其多样化的功能成为了研究热点。在植物中, P450 主要参与植物体内初级和次级代谢反应, 并发挥重要的催化作用^[6]。迄今为止, 利用功能缺失突变体法、差异筛选法、抗体或筛选 cDNA 文库、同源序列法等方法已成功地分离了多个 P450 基因, 部分 P450 基因功能已经被鉴定^[7]。Vasav 等^[8]在番茄中鉴定出 233 个 P450 基因以及保守基序并进行了系统进化分析, 为番茄中 P450 基因家族的功能和进化提供具有重要价值的信息。

本研究通过刺五加转录组测序数据筛选刺五加 P450 (*EsP450*) 基因, 利用 BLAST 工具对其序列进行分析和鉴定, 同时利用生物信息学方法来进行分析, 通过 qRT-PCR 分析 *EsP450* 基因的表达量与皂苷含量间的相关性, 为进一步研究刺五加药用成分生物合成机理奠定基础。为三萜生物合成途径的进一步解析与 P450 的功能研究提供借鉴。

1 材料与仪器

1.1 材料

刺五加样本采自河北省承德市雾灵山国家级自然保护区, 经华北理工大学邢朝斌教授鉴定为五加科植物刺五加 *E. senticosus* (Rupr. et Maxim) Maxim.。选取刺五加植株的叶片, 经清水洗净、滤纸吸干水分保存于 -80°C , 用作后续提取刺五加总 RNA 和测定总皂苷含量的样本, 根据总皂苷含量差异筛选 4 株同一产地、三年生、长势相似的刺五加样本 (A、B、C、D), 进行后续转录本测序。

1.2 仪器

核酸定量仪 (NanoDrop One, Thermo Fisher Scientific); 7900HT 型实时荧光定量 PCR 系统 (Thermo scientific); AC2-s 型无菌操作台 (ESCO 公司); 3K15 型冷冻高速离心机 (Sigma 公司); PCR 扩增仪 (ProFlex PCR System, Life technologies); 核酸电泳仪、DYY-12 型水平电泳槽 (北京市六一仪器厂); C300 型凝胶成像仪 (Azure biosystems); 2ZX-2 型旋片式真空泵 (临海市谭氏真空设备有限公司); V-5100B 型可见光光度计 (上海元析仪器有限公司)。

2 方法

2.1 *EsP450* 基因的筛选及生物信息学分析

使用 RNApure Pure 植物总 RNA 提取试剂盒提

取和纯化刺五加叶片总 RNA, 使用琼脂糖凝胶电泳法检测所提取总 RNA 的完整性, 并将其逆转录为 cDNA, 构建 cDNA 文库, 进行高通量测序。经过转录组拼接、组装与功能注释后, 获得 unigene 的注释信息^[9]。

根据 unigene 注释信息筛选出刺五加转录组测序数据中的 *EsP450*, 通过 BLAST 比对和人工鉴定, 最终得到 18 条 *EsP450* 基因。使用 ExPASy 中的 ProtParam 预测蛋白质氨基酸数量、相对分子质量、理论等电点 (PI) 等蛋白质基本性质。在 NCBI 中的 Conserved Domain Database 在线软件中分析蛋白质结构功能域。通过 TMHMM Server (v2.0) 软件对蛋白质跨膜结构进行预测。在 TargetP 1.1 Server 软件中分析蛋白质亚细胞定位, SignalP 3.0 Server 进行信号肽的预测, 使用 SOPMA 软件对蛋白质二级结构进行预测。之后使用 SWISS-MODEL 进行蛋白质三级结构的预测。最后, 在 MEGA7.0 软件中使用邻位相连法 (neighbor-joining) 构建 P450 基因的系统进化树

2.2 进化分析与正选择位点的确定

通过 Clustal X 软件进行多序列比对, 采用 MEGA7.0 软件中的临近法, bootstrapping 值设为 1000, 构建系统发育树。通过分析系统发育关系, 进行适应性进化分析, 在分子水平上度量选择压力时, 可通过估算核苷酸序列非同义替换率 (d_N) 与同义替换率 (d_S) 的比值 (ω) 来推测出其进化趋势与所经受的选择压力。当 $d_N = d_S$, 即 $\omega = 1$ 表现为中性选择; 当 $d_N < d_S$, 即 $\omega < 1$, 表明受到负选择; 当 $d_N > d_S$, 即 $\omega > 1$, 表明出现了正选择, 若 ω 显著大于 1 可被视为蛋白质发生适应性进化的证据^[10]。使用 PAML 软件中的 Comdelc 程序的位点模型对刺五加的 P450 基因在进化过程中所受到的选择压力进行分析。

位点模型假设不同位点的 ω 值不同, 但进化树上各分支 ω 值相同。本研究采用 M1a (近中性) 对 M2a (正选择), M0 (单一比值) 对 M3 (离散) 和 M7 (beta) 对 M8 (beta& ω) 3 对模型进行 LRT 检验来验证模型之间是否存在差异。其中 M0 (单一比值) 对 M3 (离散) 的比较是为了验证不同位点的 ω 值是否不同, M7 (beta) 对 M8 (beta& ω) 易出现假阳性结果, 所以选择 M1a (近中性) 对 M2a 的检测结果为可能的正选择位点^[10]。并将序列数据提交到 Datamonkey (<http://www.datamonkey.org/>)

和 MEC(<http://selecton.tau.ac.il/>)。使用 Datamonkey Web 服务器中的固定效应似然模型 (fixed effects likelihood model, FEL) 和单一似然祖先计数法 (single likelihood ancestor counting, SLAC) 对位点的选择压力进行在线分析, 对于 SLAC、FEL 方法而言, 处于 $P < 0.1$ 水平的位点即可认为是受正选择的影响^[10]。

2.3 差异表达基因的筛选

使用 RSEM 得到每个样品比对到每个基因上的 read count 数目, 以 FPKM (Fragments Per Kilobase of transcript per Million mapped reads) 来确定每个基因的表达水平。接着, 使用 EBSeq 进行差异分析, 将 $FDR < 0.01$ 且 $|\log_2(FC)| \geq 1$ 作为筛选基因表达出现显著差异的标准, Fold Change 表示 2 样品 (组) 间表达量的比值, 错误发现率

(false discovery rate, FDR) 是通过对差异显著性 P 值进行校正得到的。

2.4 基因表达量的测定

以“2.1”项中的 cDNA 作为 qRT-PCR 模板, 以刺五加 *GAPDH* 基因为内参基因, 并对筛选得到的 9 条差异表达的 *EsP450* 基因, 设计 qRT-PCR 的特异性扩增引物 (表 1)。进行 qRT-PCR 反应, 每个样本重复 3 次。总反应体系为 10 μ L, 上、下游引物各 0.3 μ L, cDNA 模板 0.5 μ L, 50 $\times$ ROX Reference Dye 1 μ L, RNase-Free ddH₂O 2.9 μ L, 2 $\times$ Talent qPCR PreMix 5 μ L。反应条件: 预变性 95 $^{\circ}$ C, 3 min; 变性 95 $^{\circ}$ C, 5 s; 退火 55 $^{\circ}$ C, 10s; 延伸 72 $^{\circ}$ C, 15s, 共 40 个循环。基于 $2^{-\Delta\Delta C_t}$ 方法计算各样本中的 *EsP450* 基因相对表达量。

表 1 *EsP450* 基因及 *GAPDH* 基因 qRT-PCR 的引物序列

Table 1 Sequence information of qRT-PCR primers for *EsP450* gene and *GAPDH* gene

序列	引物名称	引物序列 (5'-3')	片段长度/bp
TRINITY_DN10211_c0_g2	CP1S	CGGCGGTTCTTAAATAATAGAG	101
	CP1X	CAACTTGACATGAGGTCGTTAG	
TRINITY_DN1436_c0_g1	CP2S	TAGAAGAGGCTGTCTTGGAAT	321
	CP2X	GTTCCCTTATTACAATATGACGAG	
TRINITY_DN1439_c0_g1	CP3S	ATGGGCACTTTGCTTGCTAC	109
	CP3X	TTTGAGATCGGATTCAATCAC	
TRINITY_DN3773_c0_g1	CP4S	CCATTTGGTTCAGGTCGAAG	115
	CP4X	ACATGCCATTAGGCAACTCCC	
TRINITY_DN39772_c0_g3	CP5S	AGAAGCTGTGGCTGATGTTG	155
	CP5X	TATTAGGTTTGGGTGCGTTC	
TRINITY_DN41836_c0_g1	CP6S	GTTGAGATGTTGAACCGAAGT	141
	CP6X	ACAAAATCCAGCCAAGAAAT	
TRINITY_DN4544_c0_g1	CP7S	CATAGGAACGATGGGAGTAGG	144
	CP7X	TGTCACCTGCCGAGGTGTCTG	
TRINITY_DN5264_c0_g2	CP8S	TTTACCCTGTGCGCTACTTAT	183
	CP8X	TCACCTCCAATATTCTTCG	
TRINITY_DN6616_c0_g1	CP9S	TAGACGACGATGTTCTACCCG	140
	CP9X	ATCTTTCAACCACCTTTCAGG	
GAPDH	RGS	GCAAGGACTGGAGAGGTGGA	134
	RGX	AGTGGAACCTCGGAAGGACA	

2.5 皂苷含量的测定及相关性分析

参考文献的方法^[11], 对样品进行烘干、研末、称重、乙醇溶解超声破碎提取后, 进行石油醚除色素, 正丁醇萃取, 通过香草醛-浓硫酸显色法在 543 nm 的波长下测得其吸光度后, 将其带入齐墩果酸标准曲线 ($Y=0.1856X-0.0117$, $R^2=0.9964$) 进行重复多次计算, 得到刺五加样本总皂苷含量。使

用 SPSS 18.0 软件分析 *EsP450* 基因表达量与皂苷含量间的相关性。

3 结果与分析

3.1 *P450* 基因鉴定及其蛋白理化性质

筛选转录组测序数据获得的刺五加 unigenes, 将其中标记为 *P450* 基因的序列调出, 与 NCBI 数据库进行逐一比对, 之后借助 Pfam 数据库进行功能

结构域的鉴定, 最终确定其中存在 18 条 *EsP450*, 其中 TRINITY_DN41836_c0_g1、TRINITY_DN5264_c0_g2、TRINITY_DN10211_c0_g2 和 TRINITY_DN1308_c0_g1 均缺少了编码蛋白质的核酸序列, 推测为转录组拼接过程中丢失了前端部分序列, 其余 14 条基因具有完整的开放阅读框。

EsP450 蛋白的理化性质分析结果表明 (表 2), 各

EsP450 的长度、等电点方面有很大差异。编码的氨基酸长度在 122~746 aa; 相对分子质量介于 13 629.50~83 620.79; 等电点在 5.78~9.55, 碱性蛋白占据多数, 共 12 个, 酸性蛋白共 6 个; 不稳定系数介于 31.88~64.80, 平均值为 42.22, 总体上稳定性较差; 脂溶指数在 83.09~102.21, 平均值为 95.21, 亲水系数均为负值, 因此推测 *EsP450* 蛋白为亲脂性蛋白质。

表 2 刺五加 *P450* 蛋白的基本理化特性

Table 2 Basic physicochemical properties of *EsP450* protein

基因 ID	转录本长度/bp	氨基酸数目	相对分子质量	等电点	亲水系数	不稳定系数	脂溶指数	总原子数
TRINITY_DN10211_c0_g2	2008	338	38 985.59	8.84	-0.186	37.32	95.74	5541
TRINITY_DN104_c0_g1	1692	502	56 419.29	7.63	-0.149	42.87	102.21	8040
TRINITY_DN1051_c3_g1	1741	482	54 033.98	6.29	-0.119	48.23	83.09	7627
TRINITY_DN11618_c0_g1	2666	746	83 620.79	6.36	-0.161	32.84	99.03	11 909
TRINITY_DN1308_c0_g1	2627	415	47 296.05	8.56	-0.191	45.78	98.96	6737
TRINITY_DN1436_c0_g1	2289	509	57 175.37	8.68	-0.127	36.69	99.61	8136
TRINITY_DN1439_c0_g1	2142	522	58 815.45	8.89	-0.042	34.80	99.87	8353
TRINITY_DN2136_c1_g1	2875	527	59 673.34	9.08	-0.081	40.17	96.43	8459
TRINITY_DN3437_c0_g1	1727	493	56 014.30	7.67	-0.126	31.88	93.57	7938
TRINITY_DN3773_c0_g1	1977	504	57 683.50	6.91	-0.020	43.28	101.35	8200
TRINITY_DN39772_c0_g3	1860	471	53 531.57	9.55	-0.174	48.00	87.79	7612
TRINITY_DN41836_c0_g1	934	122	13 629.50	5.78	-0.312	64.80	81.64	1908
TRINITY_DN4544_c0_g1	1764	502	56 420.19	8.71	-0.176	45.16	99.64	8006
TRINITY_DN4853_c0_g1	2186	510	57 403.82	6.50	-0.094	37.64	99.43	8165
TRINITY_DN5264_c0_g2	1419	346	38 917.06	6.34	-0.055	35.72	96.10	5528
TRINITY_DN5386_c0_g1	2298	525	59 269.64	8.94	-0.088	37.66	99.14	8410
TRINITY_DN6616_c0_g1	1556	368	41 882.46	8.91	-0.016	55.54	93.80	5942
TRINITY_DN8899_c0_g1	1882	483	54 278.26	9.08	-0.085	41.63	86.38	7710

3.2 *P450* 蛋白二级结构及亚细胞定位分析

在刺五加 *P450* 蛋白二级结构 (表 3) 中, α -螺旋 (23.77%~54.35%) > 无规则卷曲 (31.79%~49.18%) > 延伸链 (9.24%~18.03%) > β -转角 (3.24%~9.02%)。因此推测二级结构中起关键作用的是 α -螺旋。信号肽预测结果表明, *P450* 蛋白中含有信号肽的可能性较小, 推测为胞内蛋白。亚细胞定位结果分析, 1 个位于细胞核, 1 个位于线粒体, 2 个位于质膜, 1 个位于叶绿体类囊体膜, 5 个位于过氧化物酶体, 8 个位于内质网膜, 推测 *P450* 蛋白在不同的位置实现氧化功能。

3.3 保守结构域分析

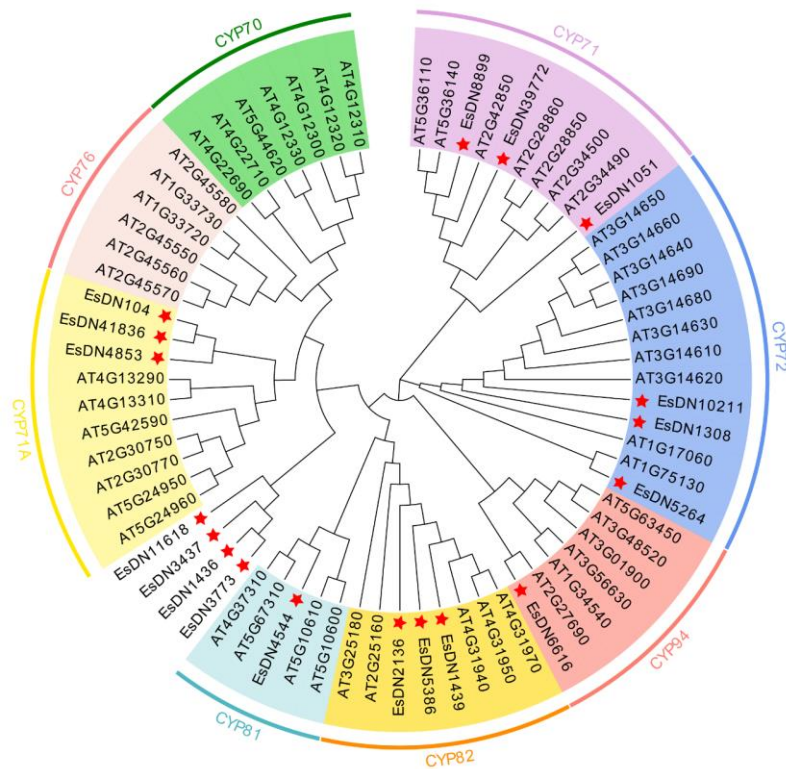
为了进一步探索 *EsP450* 的功能作用和进化关系, 使用筛选出的 18 条 *EsP450* 以及来自拟南芥的 52 条 *P450* 的氨基酸序列, 利用 MEGA7.0 软件构建了 *P450* 的 ML 系统发育树 (图 1)。如图 1 所示, *EsP450* 分别

属于不同家族。其中, CYP72 家族与 CYP82 家族各有 3 条; CYP81 家族与 CYP94 家族各有 1 条为 TRINITY_DN4544_c0_g1 和 TRINITY_DN6616_c0_g1; 而刺五加 *P450* 基因并未分布在 CYP70 家族和 CYP76 家族中; 而 4 条 (TRINITY_DN11618_c0_g1、TRINITY_DN3437_c0_g1、TRINITY_DN1436_c0_g1、TRINITY_DN3773_c0_g1) 序列与代表序列的同源性过小, 因此无法归属; 其余序列全部分布在多基因簇 CYP71 家族中。根据刺五加 *P450* 核苷酸序列, 使用邻位相连法再次进行了系统发育进化树的构建 (图 2), 未鉴定出所在家族的 4 条基因, 经过 BLAST 比对与基因结构分析后, 确认 TRINITY_DN3437_c0_g1、TRINITY_DN1436_c0_g1 和 TRINITY_DN3773_c0_g1 属于 CYP71 家族, TRINITY_DN11618_c0_g1, 除了 dnak 保守结构域外, 还包含了 CYP71 家族保守基序, 因此将其归属为 CYP71 家族。

表3 刺五加 P450 蛋白二级结构及亚细胞定位

Table 3 Secondary structure and subcellular localization of *EsP450* protein

基因 ID	α -螺旋		β -转角		无规则卷曲		延伸链		信号肽	亚细胞定位
	数量/个	占比/%	数量/个	占比/%	数量/个	占比/%	数量/个	占比/%		
TRINITY_DN10211_c0_g2	168	49.70	17	5.03	110	32.54	43	12.72	0.001 3	过氧化物酶体
TRINITY_DN104_c0_g1	235	46.81	29	5.78	171	34.06	67	13.35	0.476 6	质膜
TRINITY_DN1051_c3_g1	197	40.87	20	4.15	199	41.29	66	13.69	0.004 4	内质网膜
TRINITY_DN11618_c0_g1	318	42.63	61	8.18	258	34.58	109	14.61	0.002 6	线粒体
TRINITY_DN1308_c0_g1	212	51.08	19	4.58	137	33.01	47	11.33	0.000 4	过氧化物酶体
TRINITY_DN1436_c0_g1	244	47.94	21	4.13	183	35.95	61	11.98	0.555 2	内质网膜
TRINITY_DN1439_c0_g1	240	45.94	19	3.64	187	35.82	76	14.56	0.005 2	内质网膜
TRINITY_DN2136_c1_g1	235	44.59	25	4.74	196	37.19	71	13.47	0.002 7	过氧化物酶体
TRINITY_DN3437_c0_g1	256	51.93	24	4.87	155	31.44	58	11.76	0.384 6	质膜
TRINITY_DN3773_c0_g1	262	51.98	20	3.79	161	31.94	61	12.10	0.133 0	内质网膜
TRINITY_DN39772_c0_g3	222	47.13	27	5.73	165	35.03	57	12.10	0.135 9	内质网膜
TRINITY_DN41836_c0_g1	29	23.77	11	9.02	60	49.18	22	18.03	0.008 9	内质网膜
TRINITY_DN4544_c0_g1	235	46.81	22	4.38	175	34.86	70	13.94	0.162 3	内质网膜
TRINITY_DN4853_c0_g1	250	49.02	30	5.88	167	32.75	63	12.35	0.112 5	叶绿体类囊体膜
TRINITY_DN5264_c0_g2	178	51.45	14	4.05	113	32.66	41	11.85	0.001 1	过氧化物酶体
TRINITY_DN5386_c0_g1	241	45.90	17	3.24	198	37.71	69	13.14	0.143 7	过氧化物酶体
TRINITY_DN6616_c0_g1	200	54.36	17	4.62	117	31.79	34	9.24	0.000 9	细胞核
TRINITY_DN8899_c0_g1	225	46.58	21	4.35	169	34.99	68	14.08	0.185 0	内质网膜



在进行保守结构域分析时发现(图2), *EsP450* 基因的 motif 中存在 *P450* 典型的保守结构域。这些典型的保守结构域主要包括 PERF 结构域(PXRX)、K 螺旋(EXXR)和 C 末端的血红素结合域(FXXGXXXCXG)。血红素结合域存在铁元素形成硫醇盐离子键的绝对保守的半胱氨酸残基, 该结构域也是 *P450* 蛋白的特征结构, 其中, 血红素结合域存在于 motif 1, PERF 结构域存在于 motif 7, K 螺旋存在于 motif 5。在全部序列中均较测到了 motif 1 与 motif 7, 除 TRINITY_DN41836_c0_g1 外,

其他序列还共同拥有 motif 3。但 K 螺旋所在的 motif 5 只存在于分支 II 全部序列、分支 I 除 TRINITY_DN41836_c0_g1 外全部序列、以及 TRINITY_DN1051_c3_g1、TRINITY_DN6616_c0_g1 中, 在进行多序列比对确认后, 发现 K 螺旋确实存在于除 TRINITY_DN41836_c0_g1 外全部序列, 而 TRINITY_DN41836_c0_g1 转录本序列为 934 bp, 但其编码形成的多肽链仅有 122 aa, 因此推测为在转录组拼接时出现了丢失, 从而证实了筛选得到的 *P450* 真实性。

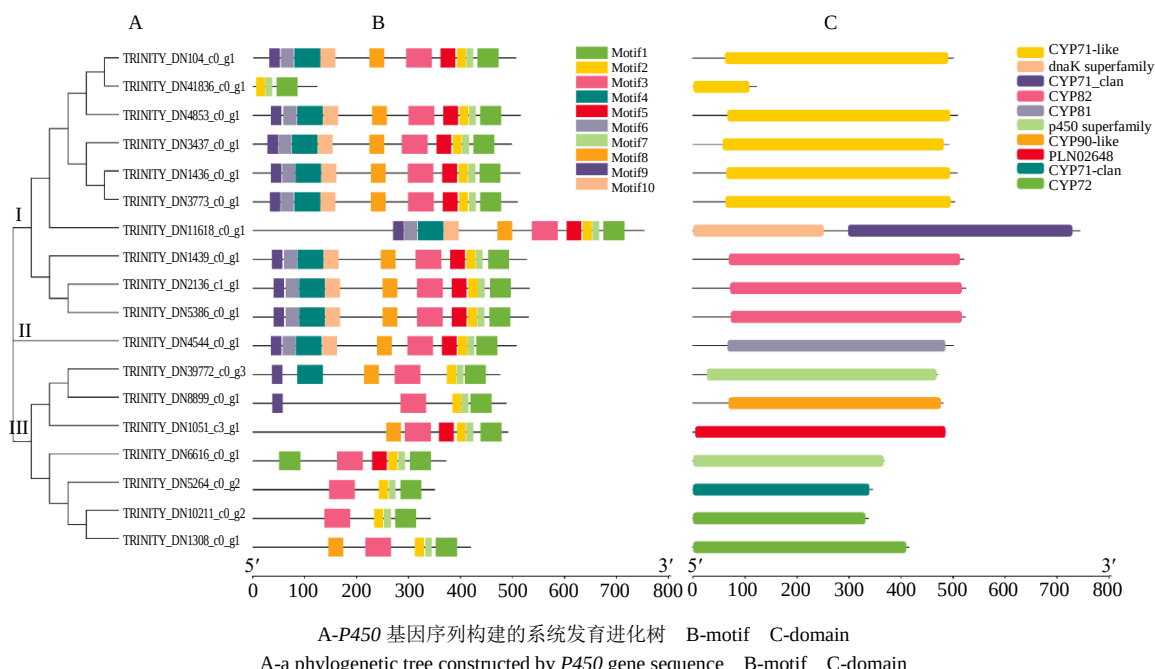


图2 刺五加 *P450* 系统发育进化树 (NJ 法) 及蛋白结构分析
Fig. 2 Phylogenetic tree and protein structure analysis of *EsP450*

3.4 选择压力分析

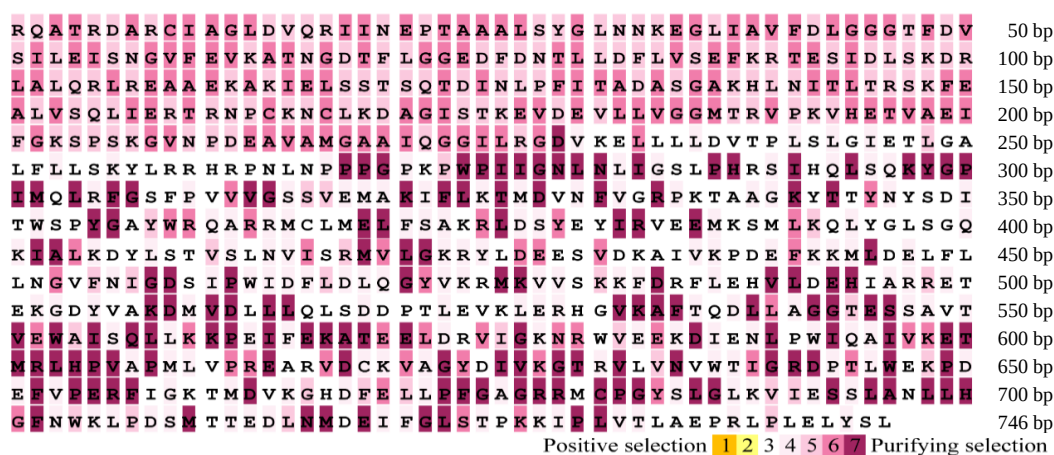
利用 PLAM 软件中的 Comedlc 程序检测 *EsP450* 家族中的每个位点的选择压力(表4)。单比率模型 M0 的参数 $np=35$, 似然值 $\ln L=-4\ 587.954\ 733$; 离散模型 M3 的参数 $np=39$, 似然值 $\ln L=-4\ 519.625\ 208$, 两者之间的 LRT 检验 $P<0.001$ 。备择假设模型 M3 成立, M3 优于 M0, 说明各个位点存在选择压压力的差异。M3 的 ω_2 小于 1, 说明不存在正选择位点。模型 M1a 与 M2a 的 $2\Delta\ln L=0$, $P=1$, 说明备择假设 M2a 不成立。在模型 M7 与 M8 比较中, $df=2$, $2\Delta\ln L=0.198\ 608$, 得到 $P=0.905\ 07$, 说明模型 M8 不成立基于 Datamonkey 检测选择压力: 分别以 SLAC 模型与 IFEL 模型进行正选择位点的鉴定。SLAC 模型在 $P<0.1$ 水平下检测到 6 个(2E、34K、64K、125R、223L、334E) 正选择位点, 14 个负选

择位点; 在 $P<0.05$ 水平下, 检测到 5 个(160K、254E、308A、356H、437K) 负选择位点, 未检测到正选择位点; 在 $P<0.01$ 水平下未检测到选择位点。在 IFEL 模型中, 当 $P<0.1$ 时检测到 3 个(13L、26K、45G) 正选择位点, 19 个负选择位点; 在 $P<0.01$ 水平下检测到 2 个(197F、310D) 正选择位点, 6 个(33K、171V、256K、343L、353L、380E) 负选择位点。

将刺五加 18 条 *P450* 基因序列上传至在线服务器 MEC 内, 以 Clustal W 为对比方法, 于 TRINITY_DN11618_c0_g1 一级结构上标注选择压力(图3)。绝大部分的位点被标注为紫色, 其中深紫色位点有 150 个, 占总位点的 20.1%。不存在橙色标记与黄色标记。表明了纯净选择在 *P450* 基因家族的进化过程中占主导地位。

表 4 刺五加 *P450* 基因基于 PAML 软件的适应性分析Table 4 Likelihood ratio test of selective on *EsP450* using site models in PAML

模型	参数	$\Delta\ln L$	参数估计值	正选择位点
M0	35	-4 587.954 733	$\omega=0.101\ 23$	不允许
M1a	36	-4 582.942 956	$P_0=0.972\ 22$, $\omega_0=0.100\ 45$ $P_1=0.027\ 78$, $\omega_1=1.000\ 00$	不允许
M2a	38	-4 582.942 956	$P_0=0.972\ 22$, $\omega_0=0.100\ 45$ $P_1=0.001\ 09$, $\omega_1=1.000\ 00$ $P_2=0.026\ 69$, $\omega_2=1.000\ 00$	不允许
M3	39	-4 519.625 208	$P_0=0.029\ 58$, $\omega_0=0.000\ 00$ $P_1=0.277\ 47$, $\omega_1=0.342\ 8$ $P_2=0.692\ 95$, $\omega_2=0.144\ 44$	不允许
M7	36	-4 518.932 888	$P=1.546\ 24$, $q=12.089\ 79$	允许
M8	38	-4 518.833 584	$P_0=0.999\ 99$, $P=1.546\ 25$ $q=12.089\ 89$, $P_1=0.000\ 01$ $\omega=1.000\ 0$	无



黄色代表受到正选择的位点，白色代表受到中性选择，红色代表受到负选择的位点，颜色越深代表在统计学上的显著性越强

Yellow: positive selection, white: neutral selection, purple: purifying selection, The deeper the color, the more the statistical significance (color figure online)

图 3 刺五加 *P450* 基因 MEC 模型分析结果Fig. 3 MEC model analysis of *EsP450*

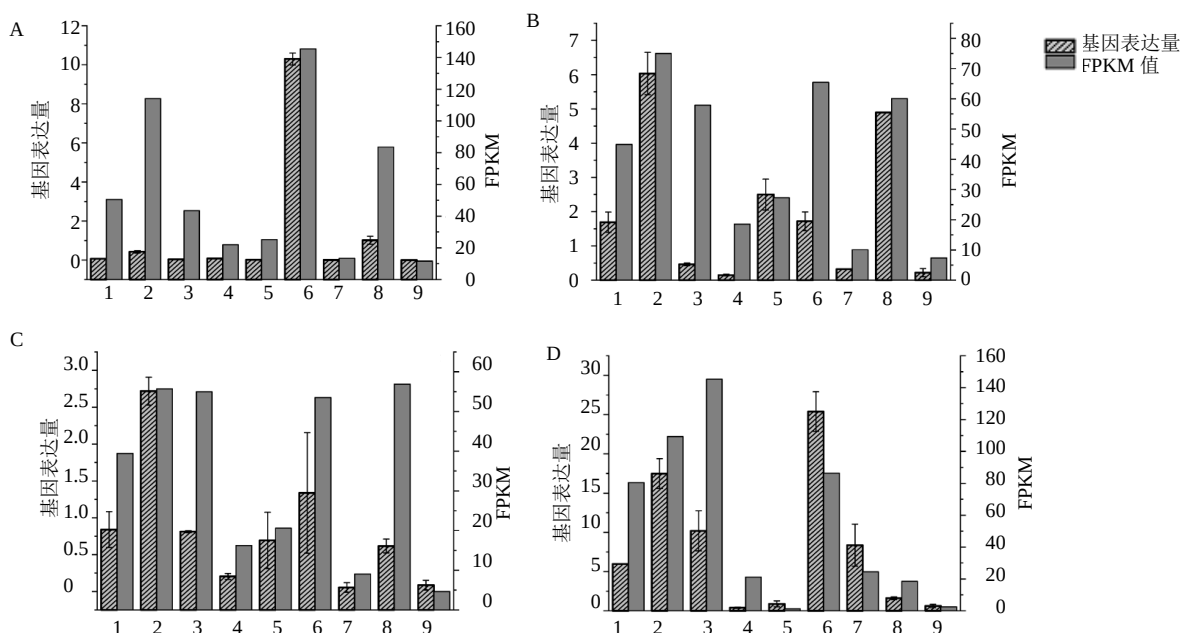
3.5 基因相对表达量与皂苷含量

从 18 条 *EsP450* 基因中筛选得到 9 条差异表达基因，对它们进行 qRT-PCR 验证的结果显示，qRT-PCR 得到的 *EsP450* 在各个样品中的表达量与转录组结果基本一致 ($P < 0.05$)。TRINITY_DN41836_c0_g1 在样本 D 中表达量最高，是样本 C 中表达量的 19.5 倍；TRINITY_DN6616_c0_g1 在各样本中表达量均较低，其中在样本 A 中未检测到其表达。4 个样本中总皂苷的含量从高到低依次为 D (47.2 g/kg) > C (37.6 g/kg) > B (32.2 g/kg) > A (26.1 g/kg)。使用 SPSS 软件对 *EsP450* 基因表达量与总皂苷含量进行相关性分析。结果

显示，在 0.01 级别检测到了 6 条基因与总皂苷含量相关性显著；在 0.05 级别检测到了 TRINITY_DN41836_c0_g1 与总皂苷含量相关性显著；TRINITY_DN39772_c0_g3 和 TRINITY_DN5264_c0_g2 与总皂苷含量不存在相关性。见图 4。

4 讨论

三萜皂苷是许多药用植物的主要活性成分之一，*P450* 基因的表达水平影响了三萜皂苷在药用植物中合成和积累量的多寡^[11]。本研究从刺五加转录组数据中挖掘出了 *EsP450* 基因家族成员，对其进行了系统的生物信息学分析。研究结果表明，CYP450 蛋白二级结构以 α -螺旋和无规则卷曲结构为主，延



A~D 分别代表 4 株刺五加样本 1-TRINITY_DN10211_c0_g2 2-TRINITY_DN1436_c0_g1 3-TRINITY_DN1439_c0_g1 4-TRINITY_DN3773_c0_g1 5-TRINITY_DN39772_c0_g3 6-TRINITY_DN41836_c0_g1 7-TRINITY_DN4544_c0_g1 8-TRINITY_DN5264_c0_g2 9-TRINITY_DN6616_c0_g1
A—D represent four samples respectively 1-TRINITY_DN10211_c0_g2 2-TRINITY_DN1436_c0_g1 3-TRINITY_DN1439_c0_g1 4-TRINITY_DN3773_c0_g1 5-TRINITY_DN39772_c0_g3 6-TRINITY_DN41836_c0_g1 7-TRINITY_DN4544_c0_g1 8-TRINITY_DN5264_c0_g2 9-TRINITY_DN6616_c0_g1

图 4 刺五加 *P450* 基因表达水平
Fig. 4 *P450* gene expression level of *E. senticosus*

伸链和 β -转角相对较少,推测 α -螺旋和无规则卷曲在其蛋白二级结构中发挥着重要的作用。细胞定位分析结果显示不同的 CYP450 定位在不同的细胞器中实现其氧化功能,这与前人研究结果一致^[12]。基因家族分析的结果显示,*EsP450* 所在家族与已知和三萜结构修饰有关的 *P450* 分布家族基本吻合^[13]。在刺五加 *P450* 中存在的最大类型是 CYP71。CYP71 也是植物中最大的 *P450* 基因簇,这个基因簇基本上包含超过半数 *P450* 基因,功能非常丰富^[14],此家族中的 CYP716A47 和 CYP716A53v2 已被证实参与人参皂苷生物合成过程中 2 步连续的氧化反应^[15],而刺五加与人参同属五加科,是人参的近源物种,在刺五加 CYP71 家族中很可能存在参与三萜皂苷合成途径反应的 CYP450。

氨基酸序列的适应性进化能够为研究酶的活性位点和功能提供有意义的信息。在植物众多的 *P450* 中,部分为三萜皂苷生物合成通路中的关键酶,这些酶也是造成三萜皂苷元多样性的重要原因之一。利用分子适应性进化原理进行功能位点的筛选可以为植物 *P450* 的活性位点提供有价值的参考。本研究

通过 PAML、MEC 模型、Datamonkey 共 3 种方式对 *EsP450* 进行了分析,结果表明 *EsP450* 不存在正选择位点,由此推测 *EsP450* 基因以负选择为主导,它们受到强烈的负选择作用的约束,在长期进化的过程中高度保守,对维持基因的重要功能起到了保护作用。比如,维持蛋白质结构的位点一般都受负选择,因为它们一旦发生突变,就会被纯化处理而消失^[16]。

本研究所用的不同刺五加样本皂苷含量存在显著差异,并且 9 条差异表达 *EsP450* 基因中,有 7 条基因的表达量与刺五加总皂苷含量成正相关关系 ($P < 0.05$)。吴鹏等^[3]研究结果也表明,*EsP450* 基因表达量与三萜皂苷含量间确实存在正相关关系,这代表参与刺五加三萜皂苷合成途径的 *EsP450* 很可能存在于本研究筛选得到的 7 条差异基因中。目前预测 *P450* 的功能仍具有一定难度,今后可以对这 7 条基因进行深入挖掘,开展 *P450* 基因克隆及功能研究,为进一步解析三萜皂苷合成途径提供借鉴。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] Huang L, Zhao H, Huang B, et al. *Acanthopanax senticosus*: Review of botany, chemistry and pharmacology [J]. *Pharmazie*, 2011, 66(2): 83-97.
- [2] Choi D W, Jung J, Ha Y I, et al. Analysis of transcripts in methyl jasmonate-treated ginseng hairy roots to identify genes involved in the biosynthesis of ginsenosides and other secondary metabolites [J]. *Plant Cell Reports*, 2005, 23(8): 557-566.
- [3] 吴鹏, 谷俊涛, 修乐山, 等. 刺五加 P450 基因时空表达差异及与皂苷含量的相关性分析 [J]. 河北农业大学学报, 2014, 37(3): 29-33.
- [4] Nelson D R, Kamataki T, Waxman D J, et al. The P450 superfamily: Update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature [J]. *DNA Cell Biology*, 1993, 12(1): 1-51.
- [5] Bozak K R, Yu H, Sirevag R, et al. Sequence analysis of ripening-related cytochrome P-450 cDNAs from avocado fruit [J]. *Proceed Nat Academy Sci*, 1990, 87(10): 3904-3908.
- [6] 张福生, 孔冉冉, 陈彤垚, 等. P450s 介导远志皂苷等齐墩果烷型植物三萜生物合成的研究进展 [J]. 药学报, 2019, 54(6): 1000-1009.
- [7] Vasav A P, Barvkar V T. Phylogenomic analysis of cytochrome P450 multigene family and their differential expression analysis in *Solanum lycopersicum* L. suggested tissue specific promoters [J]. *Bmc Genomics*, 2019, 20(1): 116.
- [8] 董栩, 许燕, 李月亭, 等. 药用植物中细胞色素 P450 基因的研究进展 [J]. 云南中医药杂志, 2016, 37(3): 75-78.
- [9] 宋菊, 国红玉, 李志栋, 等. 刺五加转录组和差异性表达分析 [J]. 中草药, 2016, 47(22): 4049-4053.
- [10] 王卓, 国红玉, 张妍彤, 等. 五加科药用植物鲨烯合酶的适应性进化分析 [J]. 中草药, 2019, 50(14): 3413-3419.
- [11] 邢朝斌, 龙月红, 劳凤云, 等. 刺五加鲨烯合酶基因的表达及其对皂苷含量的影响 [J]. 经济林研究, 2013, 31(1): 25-29.
- [12] 赵志新, 鹿鹏鹏, 王通. 植物三萜皂苷代谢中细胞色素 P450 的生物信息学分析 [J]. 西南农业学报, 2019, 32(10): 2304-2312.
- [13] 朱灵英, 郭娟, 张爱丽, 等. 参与植物三萜生物合成的细胞色素 P450 酶研究进展 [J]. 中草药, 2019, 50(22): 5597-5610.
- [14] Nelson D, Werck-Reichhart D. A P450-centric view of plant evolution [J]. *Plant J Cell Mol Biol*, 2011, 66(1): 194-211.
- [15] Gwak Y S, Han J Y, Choi Y E. Production of ginsenoside aglycone (protopanaxatriol) and male sterility of transgenic tobacco co-overexpressing three *Panax ginseng* genes: *PgDDS*, *CYP716A47*, and *CYP716A53v2* [J]. *J Ginseng Res*, 2019, 43(2): 261-271.
- [16] 陈晓霞, 苏应娟, 王艇. 细鳞苔科 psbA 基因的适应性进化分析 [J]. 西北植物学报, 2010, 30(8): 1534-1544.

[责任编辑 时圣明]