DOI:10.13275/j.cnki.lykxyj.2021.06.002

沙棘 UGT 基因家族的全基因组鉴定与 表达分析

吕中睿¹,刘 宏¹,张国昀¹,于立洋¹,罗红梅²,何彩云^{1*}

(1.国家林业和草原局林木培育重点实验室,中国林业科学研究院林业研究所,北京 100091;2.中国林业科学研究院沙漠林业实验中心,内蒙古 磴口 015200)

摘要:[目的]本研究通过系统检测沙棘 UGT 基因家族成员,探讨其结构特征及潜在功能,为解析沙棘类黄酮 糖苷生物合成机制及其积累模式奠定基础。[方法]利用 BLASTP 和 hmmsearch 搜索沙棘基因组,并通过 Pfam、CDD 和 SMART 数据库验证保守结构域。使用 Prot-Param、MUSCLE、MAGA7.0、MEME、MC-ScanX 等工具分析蛋白理化性质、系统发育、蛋白基序和基因结构及基因复制事件。利用转录组数据和实时荧 光定量 PCR 分析沙棘 UGT 基因在果实发育过程中的表达模式。[结果]本研究从沙棘基因组中鉴定出 89 个沙 棘 UGT 基因,全部包含 UGT 基因家族保守结构域 PSPG box。通过分析发现:沙棘 UGT 蛋白长度为 266~533 氨基酸,平均分子量 50.00 KDa,平均等电点 5.89。根据系统发育关系,可将 89 个沙棘 UGT 分为 16 个组,其中,A 组包含最多的沙棘 UGT 基因家族成员。除 7 号染色体外,UGT 基因共分布在 11 条沙棘染 色体上。研究发现,串联重复是导致沙棘 UGT 基因家族扩张的主要复制事件。转录组分析和实时荧光定量 PCR 表明,大部分基因具有广泛的果实发育阶段特异性表达。[结论]本研究提供了沙棘 UGT 基因家族的完 整信息,为进一步研究沙棘基因家族成员的生物学功能提供了数据参考和理论依据。 关键词:沙棘;UDP 糖基转移酶;基因表达;

中图分类号: S718.46 文献标志码: A

文章编号:1001-1498(2021)06-0009-11

沙棘 (*Hippophae rhamnoides* L.) 属胡颓子科多 年生落叶灌木、小乔木或乔木^[1],原产于俄罗斯、 中国和北欧^[2]。沙棘营养极高,富含维生素、类胡 萝卜素、脂类、甾醇和类黄酮^[3]。类黄酮是植物最 主要的次生代谢物之一^[4],在沙棘叶和果实中大量 存在,具有降血糖、降血脂、抗衰老、抗氧化等多 种生理活性^[5],在食药保健领域受到广泛关注。类 黄酮在沙棘中通常以糖基衍生物的形式出现^[6], Teleszko等研究发现,黄酮醇糖苷是沙棘中最丰富 的酚类化合物^[7]。然而,类黄酮在沙棘中的糖基化 作用机理仍不清楚。

糖基化修饰是类黄酮生物合成的关键修饰之 一,这种修饰促进类黄酮的溶解性、稳定性和

修回日期: 2021-09-18

收稿日期: 2021-03-03

生物活性,以防御和适应环境变化^[8]。植物次生代 谢物的糖基化是由 UDP 糖基转移酶 (UGT, UDPglycosyltransferase) 催化的^[9],可以催化糖基加到底 物的特定位置或特定区域。植物中 UGT 基因长度 约为 1 000~1 500 bp,UGT 基因在植物中保守性 较强,尤其在终止密码子附近有一段编码 44 个氨 基酸的极强保守序列,称为 PSPG box^[10],可以作 为挑选 UGT 基因的依据。作为模式植物,拟南芥 UGT 家族最早被研究,Li等发现,拟南芥中共有 107 个成员,根据序列同源性被划分为 14 个系统 发育组,命名为 A-N^[11]。随后,在毛果杨、玉米、 葡萄、苹果和茶等植物中陆续发现了 O、P、Q 和 R 组^[12-13]。近期,Wilson 等分析了 65 个全序列 的植物基因组,应用严格的标准来选择候选的 UGTs,并进行系统发育分析,重建了被子植物原 有的 18 个系统发育组(A-R)和 OG^[14]。在高等植 物的进化过程中,A、D、E、G和L这5个组群 扩展较快,E组扩展最快,不同物种中E组中的基 因占 UGT 家族的 20%~25%^[15]。

迄今,多个物种中的数百个 UGT 基因已经被 克隆出来,并对其功能进行了表征。如 Lim 等以 槲皮素为底物,对拟南芥中 91 个糖基转移酶进行 了鉴定,其中,29 个能够催化相关的糖基化反 应^[16]。Trapero 等对番红花中糖基转移酶功能验证 发现,UGT707B1 可以催化山奈酚、槲皮素生成相 应的糖苷衍生物^[17]。然而,与植物基因组中 UGT 基因庞大的数量相比,功能被验证的特征蛋 白的数量仍然相对较低^[15]。

本研究基于沙棘基因组信息,对 UGT 基因家 族进行了鉴定和分析,共鉴定到 89 个沙棘 UGT 基 因成员,划分为 16 个系统发育分组。本研究对沙 棘 UGT 基因家族的蛋白理化性质、亚细胞定位、 染色体分布、基因结构和基因复制进行了预测分 析。在此基础上,分析了 UGT 基因在沙棘果实不 同发育时期的表达模式,并通过实时荧光定量 PCR 进行验证,对日后解析沙棘类黄酮糖苷生物 合成机制及其积累模式奠定了基础。

1 材料与方法

1.1 沙棘 UGT 基因家族的鉴定

113 条拟南芥 UGT 氨基酸序列下载自拟南芥 基因组网站(https://www.arabidopsis.org/),UGT 基因家族保守结构域隐马尔科夫模型 HMM 文件 (PF00201,UDPGT.HMM)下载自 Pfam(http:// pfam.xfam.org/)。首先以拟南芥 UGT 氨基酸序列 作为 query 序列,使用 BLASTP 程序搜索沙棘基因 组蛋白数据库(未发表),evalue=1 e⁻¹⁵,构建沙棘 候选 UGT 数据集 1。通过 HMM 文件对沙棘基因 组蛋白数据库进行 hmmsearch 搜索,evalue=1 e⁻²⁰, 提取结果文件中比对一致的序列通过 hmmbuild 程序构建沙棘 UGT 保守结构域隐马尔科夫模型, 并再次进行 hmmsearch,构建沙棘候选 UGT 数据 集 2。合并 2 个数据集,提交至 CDD、Pfam 和 SMART 数据库验证保守结构域,然后手动删除氨 基酸序列小于 250 aa 和 PSPG box 不完整的序列。

1.2 UGT 基因家族理化性质和亚细胞定位分析

利用 Expasy server 的 ProtParam 工具(https:// web.expasy.org/protparam/)计算沙棘中各 UGT 蛋 白的分子量、氨基酸长度和等电点。使用 DeepLoc (http://www.cbs.dtu.dk/services/DeepLoc/)预测沙 棘 UGT 蛋白的亚细胞定位。

1.3 系统发育分析

通过 MUSCLE 对沙棘 UGT 蛋白序列进行多重 序列比对(http://www.ebi.ac.uk/Tools/msa/muscle/), 删除 gap 区域。利用 MEGA 7.0 软件,基于比对后 的 UGT 蛋白序列,采用 neighbor-joining 法,设置 bootstrap 值为 1000,构建系统发育树^[18]。

1.4 保守序列和基因结构分析

通过 GSDS 在线工具 (v2.0 http://gsds.cbi.pku. edu.cn/), 输入沙棘基因注释 GFF 文件, 将沙棘 UGT 的编码序列与其对应的基因组序列进行比 较, 展示沙棘 UGT 的外显子内含子信息。为了比 较沙棘 UGT 的差异,本研究利用 MEME 在线工具 对沙棘 UGT 蛋白的保守基序进行分析,参数设置 为: site distribution: zero or one occurrence (of a contributing motif site) per sequence, maximum number of motifs: 10, and optimum motif width \geq 6 and \leq 60。

1.5 染色体定位和基因复制分析

通过自建脚本,从沙棘基因组注释文件中提取 沙棘 UGT 位置信息。使用 MCScanX 软件分析基 因加倍事件。染色体定位和基因加倍信息通过 Circos 软件绘图展示。

1.6 表达模式分析

2个沙棘亚种(中国沙棘,"FN";蒙古沙棘, "XY")不同果实发育阶段的转录组数据下载自沙 棘基因组数据库,使用每百万映射 reads 的千碱 基片段(FPKM)来估计表达水平。利用 TBtools 软件对数据进行标准化和聚类,并绘制表达量 热图^[19]。

实时荧光定量 PCR 分析所用样品为中国林业 科学研究院沙漠林业实验中心种植的蒙古沙棘花 后 21、63、91 d 果实,每批样品设置 3 个生物学 重复,采样后迅速使用液氮速冻,并置于-80℃ 备 用。总 RNA 的提取采用天根公司 RNAprep Pure 多 糖多酚植物总 RNA 提取试剂盒,参照使用说明书 的方法进行提取。反转录试剂盒为 TAKARA 公司 的 PrimeScriptTM 1st Strand cDNA Synthesis Kit,并 按照使用说明进行。用 primer Premier 5.0 软件对选 定的 9 个 HrUGTs 进行特异性引物设计,引物信息 见表 1。实时荧光定量 PCR 反应体系按照 TAKARA 公司 TB Green[®] Premix Ex Taq TM II 试剂盒使用说 明书配置, PCR 反应程序为: 95℃ 30 s 预变性, 95℃ 5 s, 60℃ 30 s, 40 个循环扩增。使用 2^{-ΔΔCT} 法计算 HrUGT 基因的相对表达水平^[20],使用 Origin 8.0 软件作图。

	表 1 实时荧光定量 PCR 引物信息表
Table 1	The primer sequence for quantitative real-time PCR (RT-qPCR)

引物名称	序列(5'-3')	引物名称	序列(5'-3')								
Primer name	Primer sequence	Primer name	Primer sequence								
		01.11.04475									
QHrUG10002F	GCCTAAGCCTCGTATCCTCG	QLa11g0447F	TATICCCTAGCGCAGGCAT								
QHrUGT0002R	CCCATCGTCGTAACCATCAG	QLa11g0447R	TGGATTCAGATAAGGAGACGGT								
QLa2g0900F	AGTTCATTGCTCAAAAGGGTCA	QLa11g2592F	GAGATGCTTATTGATCGGCTTG								
QLa2g0900R	CGTTCTCTGGCAAATCCATAGTT	QLa11g2592R	TAGTCTATGGGTTCCTCACCTTTAT								
QLa2g3104F	GATTCAACCTTCCCAACTCCC	QLa12g1442F	ACCCTTCATTGCTCAGAGTCTC								
QLa2g3104R	CACGCCGTTAAGCCTAGCA	QLa12g1442R	GGATTTTGTCGGTGTTTTCG								
QLa9g0469F	CAACCAAATCATCACCCCCT	QLa12g2361F	CACACCAATGAACATCAATCGT								
QLa9g0469R	GAACAGCTATAAAAGCCGTGC	QLa12g2361R	AACTTTAGGTAGTGGGAAAGACACA								
QLa10g1923F	AACTTCTTCTCCAGTCCGCCAT	18SF	AAACCTTACCAGCCCTTGAC								
QLa10g1923R	ACTTCATGCCCAATCGACGAG	18SR	CGCTCGTTATAGGACTTGACC								

2 结果分析

2.1 沙棘 UGT 基因家族成员鉴定

对利用 BLASTP 和 hmmsearch 两种方法搜索 沙棘基因组蛋白数据库获得的 110个候选沙棘 UGT 基因成员,经过验证保守结构域和手动筛 选,共鉴定出 89个沙棘 UGT 基因。蛋白理化性质 分析结果(表 2)表明:沙棘 UGT 家族各成员蛋 白质长度为 266~533 aa,平均长度 462 aa,蛋白 理论分子量平均值为 52.00 KDa,平均等电点 5.89。 82 个沙棘 UGT 家族成员定位于细胞质,6个成员 定位于线粒体,1个成员定位于质体。

2.2 系统发育分析

基于蛋白同源序列的相似性进行功能预测是基因功能研究的重要手段,本研究以沙棘和拟南芥、 玉米、山柳兰等植物 UGT 蛋白序列为基础,构建 了系统发育树。图 1 表明: 89 个沙棘 UGT 可被聚 类为 16 个先前鉴定的类群^[13],沙棘 UGT 在 O 组 和 Q 组 均没有分布,大部分沙棘 UGT 聚集在 E(8)、G(8)、D(11)、L(16)和A(17)组。 多序列比对分析表明: 89 个沙棘 UGT 的 C 端序列 均存在 PSPG box,并在 1(W)、4(Q)、8(L)、 10(H)、12(S/A)、14(G)、16(F)、19-24 (HCGWNS)、27(E)、32-34(GVP)、39
(P)、43(D/E)、44(Q)位点高度保守。

2.3 蛋白基序和基因结构分析

为了进一步确定沙棘 UGT 家族的保守结构域 特征,利用在线工具 MEME 创建了 10 个基序,并 从 1 到 10 列出(图 2)。基序 1 和基序 3 为 UGT 家 族保守结构域 PSPG box。La4g1035、La5g0208、 La11g1107、La5g1327、La4g1118、La10g1561 和 La10g1574 由于 1 或 2 个氨基酸的插入并没有匹配 到基序 3,在后续的分析中发现,这些基因除 La5g0208 外均未发现表达或表达量极低。A 组和 R 组成员均未发现基序 9 的存在,这一基序中 3 个 氨基酸(GSS)之前被认为在单糖基转移酶中高度 保守^[21]。

内含子外显子结构的多样性通常在基因家族的 进化中发挥关键作用,并为支持系统发育类群提供 了额外的证据^[22]。为了进一步了解基因结构,对沙 棘 UGT 的内含子外显子结构进行了分析。在本研 究鉴定的 89 个 UGT 基因中,45 个 UGT 基因含有 内含子 (50.6%),其中,40 个 UGT 基因有 1 个内 含子,5 个 UGT 有 2 个内含子。G 组、P 组和 F 组成员大多具有较长的内含子插入。M 组、B 组

表 2 沙棘 UGT 基因家族成员信息 Table 2 The information of HrUGTs

基因ID Gene ID	蛋白长度/aa Protein length	分子量/KDa Molecular weight	等电点 pI	亚细胞定位 Subcellular localization	基因ID Gene ID	蛋白长度/aa Protein length	分子量/KDa Molecular weight	等电点 pI	亚细胞定位 Subcellular localization
HrUGT0001	485	54.97	5.44	细胞质	La6g1086	503	56.83	6.25	细胞质
HrUGT0002	483	54.45	4.88	细胞质	La6g1126	490	55.62	6.33	细胞质
HrUGT0003	486	54.90	6.00	细胞质	La8g0188	457	51.01	5.83	细胞质
HrUGT0004	485	55.08	5.54	细胞质	La8g0189	266	29.74	6.27	细胞质
HrUGT0005	478	53.38	5.38	细胞质	La8g0492	495	55.55	6.25	细胞质
La1g0563	462	52.15	5.26	细胞质	La9g0184	415	46.42	5.63	细胞质
Lalg1021	496	56.08	5.52	细胞质	La9g0185	478	53.30	5.61	细胞质
Lalg1077	515	57.92	6.77	细胞质	La9g0469	482	53.46	6.04	细胞质
La1g1078	472	52.65	7.19	线粒体	La10g1046	485	54.78	5.30	细胞质
La1g2297	468	51.73	6.76	细胞质	La10g1047	463	52.58	5.32	线粒体
La1g2301	473	53.62	5.45	细胞质	La10g1561	489	56.06	6.53	细胞质
La1g2837	470	52.24	5.57	细胞质	La10g1574	388	44.45	6.28	细胞质
La2g0136	288	32.71	5.54	细胞质	La10g1923	461	51.48	5.86	细胞质
La2g0150	490	55.53	5.82	细胞质	La10g1932	468	52.89	5.39	细胞质
La2g0151	488	55.59	6.01	细胞质	La10g2527	486	55.10	5.49	细胞质
La2g0165	461	52.49	5.76	细胞质	La10g2528	285	32.18	5.69	线粒体
La2g0900	467	53.27	5.59	细胞质	La10g2530	288	32.38	4.90	细胞质
La2g1189	495	56.01	6.81	细胞质	La10g2531	440	49.93	5.50	细胞质
La2g2279	488	55.39	6.13	细胞质	La10g2632	465	51.88	6.12	细胞质
La2g2282	499	55.73	5.27	细胞质	La11g0447	490	54.83	5.30	细胞质
La2g3104	453	50.55	5.17	细胞质	La11g0570	469	52.88	5.75	细胞质
La3g0020	462	52.27	5.73	细胞质	La11g1107	484	53.88	5.99	细胞质
La3g0035	456	51.57	5.72	细胞质	La11g1417	481	53.85	6.80	细胞质
La3g0196	456	51.58	6.21	线粒体	La11g1418	533	60.00	5.13	细胞质
La3g0199	466	52.30	6.62	细胞质	La11g1941	463	52.33	5.73	质体
La3g0203	466	52.23	6.84	细胞质	La11g2409	510	58.41	7.64	细胞质
La3g0694	380	42.57	5.90	细胞质	La11g2588	453	50.38	5.68	细胞质
La3g0945	470	51.47	6.28	细胞质	La11g2591	458	51.90	5.78	细胞质
La3g1197	481	53.90	5.27	细胞质	La11g2592	461	52.31	6.22	细胞质
La3g1210	480	54.52	6.53	细胞质	La11g2624	453	50.62	5.52	细胞质
La4g1118	475	53.47	6.42	细胞质	La11g2653	450	50.58	6.03	细胞质
La4g1305	459	51.77	5.94	细胞质	La11g2654	452	50.88	7.99	细胞质
La4g2072	475	52.35	6.15	细胞质	La12g0379	490	54.82	5.25	细胞质
La5g0208	470	52.75	5.94	细胞质	La12g0383	490	54.82	5.25	细胞质
La5g0668	532	58.53	5.69	细胞质	La12g0737	491	55.35	5.77	细胞质
La5g0841	464	51.91	5.20	细胞质	La12g1195	483	53.20	6.10	细胞质
La5g0951	476	53.06	6.08	细胞质	La12g1196	474	52.91	6.33	细胞质
La5g1081	454	50.96	5.24	细胞质	La12g1442	489	55.27	5.14	细胞质
La5g1082	461	51.85	5.31	细胞质	La12g1628	488	54.56	6.65	细胞质
La5g1269	478	53.21	6.34	细胞质	La12g1841	446	49.84	5.23	细胞质
La5g1327	482	54.31	6.04	细胞质	La12g2361	292	32.80	6.76	线粒体
La5g1328	472	53.10	5.46	线粒体	La12g3948	481	54.07	5.52	细胞质
La6g1080	487	54.65	7.10	细胞质	La12g4351	444	50.55	5.91	细胞质
La6g1081	492	55.53	5.40	细胞质	La12g4479	444	50.55	5.83	细胞质
La6g1082	489	55.19	5.39	细胞质				_	

注/Notes: HrUGT0001 (Sph_Contig02792G000010)、HrUGT0002 (Sph_Contig03881_ERROPOS16800000_G000350)、 HrUGT0003 (Sph_Contig03890_ERROPOS700000_G000100)、HrUGT0004 (Sph_Contig03890_ERROPOS700000_G000150)、 HrUGT0005 (Sph_Contig03932_ERROPOS2100000_G000520)。



注:利用 MEGA 7.0 程序构建了 89 个来自沙棘的 UGTs(●), 19 个拟南芥 UGTs、4 个玉米 UGTs 和 1 个山柳兰 UGT(★) 的 neighborjoining 树。每个节点计算 1 000 次 bootstrap 值并以百分比标注在分支上。上述 UGTs 被划分为 18 个不同的分组,并用不同颜色标出。 Notes: A total of 89 UGTs from sea buckthorn (●), 19 UGTs from *Arabidopsis*, 4 UGTs from maize and 1 UGT from mouse-ear hawkweed (★) were used to

Notes: A total of 89 UG1s from sea buckthorn (\bullet), 19 UG1s from *Arabidopsis*, 4 UG1s from maize and 1 UG1 from mouse-ear hawkweed (\star) were used to construct the neighbor-joining tree using MEGA 7.0 program. The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (1 000 replicates) are shown next to the branches. The UGTs were classified into 18 groups, A~R, the color of each group was different.

图 1 沙棘、拟南芥、玉米和山柳兰 UGT 系统发育树

Fig. 1 Phylogenetic tree of UGT proteins of sea buckthorn, Arabidopsis, maize and mouse-ear hawkweed

和R组成员均不含内含子。

2.4 染色体定位和基因复制分析

在鉴定出的 89 个沙棘 UGT 中,84 个 UGT 被 定位于沙棘染色体上。图 3 表明:在12 条沙棘染 色体中,只有11 条沙棘染色体包含 UGT 基因。 11 号染色体包含最多的共13 个 UGT 家族成员, 而7 号染色体中没有 UGT 基因存在。12 号染色体 含有12 个 UGT 基因,10 号染色体包含11 个 UGT 基因,4 号、8 号和9 号染色体均只含3 个 UGT 基 因。沙棘 UGT 基因在染色体上的这种不平衡分 布,说明沙棘在进化过程中存在遗传变异。

为了揭示沙棘 UGT 基因家族的扩展和进化机制,对沙棘基因组中潜在的基因复制事件进行了分析。本研究利用 MCScanX 软件基于氨基酸序列同源性在沙棘全基因组内进行了比对,发现 UGT 基因家族成员中存在 12 个串联重复基因簇和 11 个共线基因对(图3),这一结果表明,串联重复是导致沙棘 UGT 基因家族扩张的主要复制事件。本研究计算了复制基因间的 Ka 和 Ks 值,其比值均小

于1, 说明 UGT 基因在进化过程中受到纯化选择。

2.5 沙棘 UGT 基因在果实不同发育阶段的表达模式

本研究利用两个沙棘亚种果实 3 个发育阶段的 转录组数据,来进一步了解沙棘 UGT 基因的表达 模式,结果发现:沙棘各 UGT 在种间和时间上的 表达表现出显著差异(图4)。La10g1046,La10g2527 和 La3g0035 只在中国沙棘果实中微量表达,在 蒙古沙棘果实中不表达。La2g0165,La3g0199 和 La1g2297 则表现相反;而 La9g0469 在蒙古沙棘果 实中高表达,在中国沙棘中不表达。La5g0668 在两个亚种不同发育时期均高表达。La11g2592、 HrUGT0002、La12g1442 等基因在两个亚种果实中 表达较高且随着果实发育表达量逐渐升高。大多 数 UGT 基因主要在果实发育的前期或中期表达量 较高,而在果实发育后期表达量降低。

从沙棘 UGT 基因所处的系统发育分组看, A 组中, La5g0208 和 La12g2361 两个基因表达水 平相对较高,且均随果实发育表达量逐渐降低。相 似的,La5g0951 只在果实发育初期表达,而在果





实发育的中到后期均不表达。C组中,Lal1g1107 在两个沙棘亚种果实中均不表达,而Lal2g1442在 两个沙棘中表达量相对较高且主要在果实发育的中 后期表达。D组和E组均包含较多的沙棘UGT基 因家族成员,但两组基因的表达模式却有着巨大差 异。在D组中,除La2g1189外,其他10个基因 在中国沙棘中均不表达,这些基因在蒙古沙棘果实 中的表达水平也相对较低甚至不表达。而在E组 中,除La9g0469在中国沙棘果实中不表达,其余 基因在两个沙棘亚种果实中均有一定程度的表达。 La11g0447和La11g0570在两个沙棘亚种果实中表 达量相对较高,Lal1g0447 在蒙古沙棘中表达量随 着果实发育先升高后降低,而在中国沙棘中表现出 相反的趋势;Lal1g0570 在中国沙棘中随着果实发 育表达量逐渐降低,而在蒙古沙棘果实中的表达水 平先小幅升高,在果实成熟时下降到较低水平。 F组成员在沙棘果实中除前期有少量表达外,其余 时期表达水平较低或不表达。G组中,Lal2g0737 表达量整体较高,在两个沙棘亚种果实中均表现为 随着果实发育表达量先升高后降低;La2g2279 在 中国沙棘果实中有着较高的表达水平,且随着果实 发育表达量逐渐升高,而在蒙古沙棘果实发育末期



注: 片段重复基因对由一条红线连接, 蓝线连接的基因对来自于全基因组加倍, 星号表示串联重复序列

Notes: The segmental duplication genes were linked by a red line. The genes connected by the blue lines were derived from the whole genome duplication. Asterisks indicated tandem duplication genes.

图 3 沙棘 UGT 基因的染色体分布和基因重复 Fig. 3 Chromosomal distribution and gene duplications of the HrUGTs

表达量下降到较低水平。L组中,La10g1081和 La10g1082均在果实发育中期表达量较高,且蒙古 沙棘高于中国沙棘;HrUGT0002随着果实发育表 达量逐渐增加,且与发育初期相比HrUGT0002在 蒙古沙棘果实成熟期的表达水平提高了16.7倍, 而在中国沙棘中达到了25倍。J组和R组中的沙 棘UGT 基因在果实发育的各个时期均有相对较高 的表达水平,H组、I组、K组、M组和N组中各 成员在沙棘果实中表达量均相对较低。

本研究对部分表达差异较大的沙棘 UGT 利用 实时荧光定量 PCR 进行验证,结果(图 5)表明: 在蒙古沙棘中 HrUGT0002、La2g0900、La9g0469 和 La11g2592 均随果实成熟表达量逐渐上升,而 La2g3104、La10g1923、La12g2361 总体呈下降趋势,La11g0447 和 La11g0570 基因则在果实发育的中期表达较高。总体来看,实时荧光定量 PCR 结果与转录组结果基本一致。

3 讨论

为了从功能上对沙棘 UGT 进行鉴定,通过系 统发育分析将鉴定到的 89 个沙棘 UGT 基因聚类 为 16 个组。沙棘中的 UGT 基因约占沙棘全基因组 基因总数的 0.29%,低于桃(0.6%)^[23]和拟南芥 (0.44%)^[11],高于石斛(0.28%)^[24]和玉米(0.23%)^[12]



注: 沙棘 UGT 基因的表达水平以 log₂ (FPKM+1)标准化

Note: Expression levels of the HrUGTs were shown as the log₂ (FPKM+1), transformed FPKM values obtained from the RNA-Seq data.

图 4 沙棘 UGT 基因在两个亚种不同发育时期的表达模式

Fig. 4 Expression profiles of HrUGTs in various developmental stages of two sea buckthorn subspecies

的 UGT 基因占比。A 组、L 组、D 组、G 组和 E 组被认为是高等植物进化过程中进化最快的分 组^[15],在沙棘中这些分组包含了最多的 UGT 基因 家族成员,这一结果与 Ren 等^[24]和 Cui 等^[13]的研 究高度一致。A 组中的多数 UGT 被鉴定为能够催 化类黄酮糖苷再次糖基化的糖基转移酶[25-27].本研 究发现,沙棘 UGT 家族 A 组成员均不含单糖基转 移酶中高度保守的 C 端 GSS 基序,这一结构特征 也暗示着沙棘 UGT 家族 A 组成员可能和多糖基类 黄酮糖苷的生物合成存在重要联系。O组和O组 在沙棘中未发现有成员存在,这两个分组最早在玉 米中鉴定出来[12]并被认为可能与细胞分裂素的糖 基化有关。La12g1195、La11g1196 和 La5g0668 被 划分为 UGT95 亚家族,这一亚家族在山柳兰中首 先被鉴定出来,能够催化木犀草素和槲皮素的 3'-OH 基团和山萘酚的 7-OH 基团糖基化^[28]。在石

榴^[29] 和茶树^[13] 中均发现了 UGT95 亚家族成员的存在, Cui 等将其划为 R 组^[13],在本研究中延续了这一分组的划分。

在鉴定到的 89 沙棘 UGT 基因中,有 84 个基 因被定位到染色体上。这些基因在染色体上通常成 簇存在且表现出较高的序列相似性,这一特征与石 斛和棉花表现一致^[24,30]。本研究基于序列相似性和 基因间距鉴定出 12 个串联重复基因簇和 11 个共线 基因对,证明串联重复是导致沙棘 UGT 基因家族 扩张的主要复制事件。内含子的位置、丢失和获得 可以作为了解基因家族在系统发育类群内进化的重 要指标。超过一半 (50.6%)的沙棘 UGT 有内含子 插入,低于玉米 (60%)^[12] 和拟南芥 (58%)^[11] 的内含 子数量。利用 MEME 在线工具来搜索 UGT 蛋白之 间共享的保守基序,共发现了 10 个不同的保守基 序,其中,在所有鉴定的 UGT 中都发现了编码



Note: The fruits at three developmental stages (F1, F2, F3) were analyzed by RT-qPCR.

图 5 沙棘 UGT 基因在果实不同发育时期的实时荧光定量 PCR 分析

Fig. 5 Expression analysis of selected HrUGTsin various developmental stages using RT-qPCR.

UGT 结构域的基序 1。这些基序在组间有着显著差异,特别是 R 组和 A 组均不含在其他分组中普遍存在的基序 9。这些特定的基序可能会导致沙棘 UGTs 功能的分化。

了解基因的时空表达模式有助于推测基因的功能。在蒙古沙棘中,48个UGT基因在果实发育过程中表达(FPKM>1),在中国沙棘中这一数字为51。R组3个成员表达量在两个亚种果实发育时期均较高。除La9g0469外,E组成员在两个亚种果实中均有不同程度的表达。La9g0469在中国沙棘中不表达,而在蒙古沙棘中高表达,且随着果实发育表达量逐渐上升,这种特异性的表达可能对两个亚种果实中代谢物组成造成一定影响。

4 结论

本研究在沙棘全基因组范围内鉴定获得 89 条 含有 UGT 保守结构域的 HrUGTs 蛋白序列,并划 分为 16 个系统发育分组。同一分组内沙棘 UGT 具 有相似的蛋白基序和基因结构,但在组间存在着巨 大差异。沙棘 UGT 家族在进化过程中受到纯化选 择。沙棘 UGT 基因家族成员在两个沙棘亚种和果 实不同发育阶段的表达模式具有显著差异。沙棘 UGT 基因家族的表达模式和生物信息学分析将为 进一步鉴定沙棘类黄酮糖基转移酶功能和催化机理 奠定基础。

参考文献:

- [1] 宋 彬, 胡安鸿, 田永芝, 等. 沙棘PEPCK基因的克隆及表达研 究[J]. 西北植物学报, 2017, 37 (10): 1934-1940.
- Fatima T, Kesari V, Watt I, *et al.* Metabolite profiling and expression analysis of flavonoid, vitamin C and tocopherol biosynthesis genes in the antioxidant-rich sea buckthorn (*Hippophae rhamnoides* L.)[J]. Phytochemistry, 2015, 118(9): 181-191.
- [3] Fang R, Veitch N C, Kite G C, *et al.* Enhanced profiling of flavonol glycosides in the fruits of sea buckthorn (*Hippophae rhamnoides*)[J].
 J Agric Food Chem, 2013, 61(16): 3868-3875.
- [4] Yonekura K, Saito K. Function, structure, and evolution of flavonoid

glycosyltransferases in plants[J]. Recent Advances in Polyphenol Research, 2014, 4: 61-82.

- [5] Yogendra K M S, Tirpude R J, Maheshwari D T, et al. Antioxidant and antimicrobial properties of phenolic rich fraction of Seabuckthorn (*Hippophae rhamnoides* L.) leaves in vitro[J]. Food Chem, 2013, 141(4): 3443-3450.
- [6] Rosch D, Krumbein A, Mugge C, et al. Structural investigations of flavonol glycosides from sea buckthorn (*Hippophae rhamnoides*) pomace by NMR spectroscopy and HPLC-ESI-MS(n)[J]. J Agric Food Chem, 2004, 52(13): 4039-4046.
- [7] Teleszko M, Wojdylo A, Rudzinska M, et al. Analysis of Lipophilic and Hydrophilic Bioactive Compounds Content in Sea Buckthorn (*Hippophae rhamnoides* L.) Berries[J]. J Agric Food Chem, 2015, 63(16): 4120-4129.
- [8] Vogt T, and Jones P. Glycosyltransferases in plant natural product synthesis: characterization of a supergene family[J]. Trends Plant Sci, 2000, 5(9): 380-386.
- [9] Bowles D, Lim E K, Poppenberger B, et al. Glycosyltransferases of lipophilic small molecules[J]. Annu Rev Plant Biol, 2006, 57: 567-597.
- [10] Gachon C M, Langlois-Meurinne M, Saindrenan P. Plant secondary metabolism glycosyltransferases: the emerging functional analysis[J]. Trends Plant Sci, 2005, 10(11): 542-549.
- [11] Li Y, Baldauf S, Lim E K, et al. Phylogenetic analysis of the UDPglycosyltransferase multigene family of Arabidopsis thaliana[J]. J Biol Chem, 2001, 276(6): 4338-4343.
- [12] Li Y, Li P, Wang Y, et al. Genomewide identification and phylogenetic analysis of Family-1 UDP glycosyltransferases in maize (Zea mays)[J]. Planta, 2014, 239(6): 1265-1279.
- [13] Cui L, Yao S, Dai X, et al. Identification of UDP-glycosyltransferases involved in the biosynthesis of astringent taste compounds in tea (*Camellia sinensis*)[J]. J Exp Bot, 2016, 67(8): 2285-2297.
- [14] Wilson A E, Tian L. Phylogenomic analysis of UDP-dependent glycosyltransferases provides insights into the evolutionary landscape of glycosylation in plant metabolism[J]. Plant J, 2019, 100(6): 1273-1288.
- [15] Caputi L, Malnoy M, Goremykin V, et al. A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land[J]. Plant J, 2012, 69(6): 1030-1042.
- [16] Lim E K, Ashford D A, Hou B, et al. Arabidopsis glycosyltransferases as biocatalysts in fermentation for regioselective synthesis of diverse quercetin glucosides[J]. Biotechnol Bioeng, 2004, 87(5): 623-631.
- [17] Trapero A, Ahrazem O, Rubio-Moraga A, et al. Characterization of a glucosyltransferase enzyme involved in the formation of kaempferol and quercetin sophorosides in *Crocus sativus* [J]. Plant Physiol, 2012,

159(4): 1335-1354.

- [18] Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets[J]. Mol Biol Evol, 2016, 33(7): 1870-1874.
- [19] Chen C, Chen H, Zhang Y, et al. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data[J]. Mol Plant, 2020, 13(8): 1194-1202.
- [20] Livak K J, Schmittgen T D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method[J]. Methods, 2001, 25(4): 402-408.
- [21] Huang F C, Giri A, Daniilidis M, et al. Structural and Functional Analysis of UGT92G6 Suggests an Evolutionary Link Between Mono- and Disaccharide Glycoside-Forming Transferases[J]. Plant Cell Physiol, 2018, 59(4): 857-870.
- [22] Zhu Y X, Yang L, Liu N, *et al.* Genome-wide identification, structure characterization, and expression pattern profiling of aquaporin gene family in cucumber[J]. BMC Plant Biol, 2019, 19(1): 345.
- [23] Wu B, Gao L, Gao J, et al. Genome-Wide Identification, Expression Patterns, and Functional Analysis of UDP Glycosyltransferase Family in Peach (*Prunus persica* L. Batsch)[J]. Front Plant Sci, 2017, 8: 389.
- [24] Ren Z, Ji X, Jiao Z, et al. Functional analysis of a novel C-glycosyltransferase in the orchid *Dendrobium catenatum*[J]. Hortic Res, 2020, 7(1): 111.
- [25] Cheng J, Wei G, Zhou H, et al. Unraveling the mechanism underlying the glycosylation and methylation of anthocyanins in peach[J]. Plant Physiol, 2014, 166(2): 1044-1058.
- [26] Montefiori M, Espley R V, Stevenson D, et al. Identification and characterisation of F3GT1 and F3GGT1, two glycosyltransferases responsible for anthocyanin biosynthesis in red-fleshed kiwifruit (Actinidia chinensis)[J]. Plant J, 2011, 65(1): 106-118.
- [27] Morita Y, Hoshino A, Kikuchi Y, et al. Japanese morning glory dusky mutants displaying reddish-brown or purplish-gray flowers are deficient in a novel glycosylation enzyme for anthocyanin biosynthesis, UDP-glucose: anthocyanidin 3-O-glucoside-2"-O-glucosyltransferase, due to 4-bp insertions in the gene[J]. Plant J, 2005, 42(3): 353-363.
- [28] Witte S, Moco S, Vervoort J, et al. Recombinant expression and functional characterisation of regiospecific flavonoid glucosyltransferases from *Hieracium pilosella* L[J]. Planta, 2009, 229(5): 1135-1146.
- [29] Wilson A E, Wu S, Tian L. PgUGT95B2 preferentially metabolizes flavones/flavonols and has evolved independently from flavone/flavonol UGTs identified in *Arabidopsis thaliana* [J]. Phytochemistry, 2019, 157: 184-193.
- [30] Huang J, Pang C, Fan S, *et al.* Genome-wide analysis of the family 1 glycosyltransferases in cotton[J]. Mol Genet Genomics, 2015, 290(5): 1805-1818.

Genome-wide Identification, Characterization, and Expression Analysis of UGT Gene Family Members in Sea Buckthorn (*Hippophae rhamnoides* L.)

LYU Zhong-rui¹, LIU Hong¹, ZHANG Guo-yun¹, YU Li-yang¹, LUO Hong-mei², HE Cai-yun¹

(1. Key Laboratory of Tree Breeding and Cultivation of National Forestry and Grassland Administration, Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, China; 2. Experimental Center of Desert Forestry, Chinese Academy of Forestry, Dengkou 015200, Inner Mongolia, China)

Abstract: [Objective] To study the characteristics and potential functions of sea buckthorn (*Hippophae rhamnoides*) and to analyze the biosynthesis mechanism and accumulation pattern of flavonoid glycosides by identifying the UGT gene family members in sea buckthorn. [Method] BLASTP and hmmsearch were used to identify the members of the HrUGT gene family based on sea buckthorn genome database. The protein physical and chemical properties, phylogenesis, protein motif and gene structure and gene duplication were analyzed by using Prot-Param, MUSCLE, MAGA7.0, MEME and MCScanX. [Result] 89 HrUGTs containing the plant secondary product glycosyltransferase motif (PSPG) were identified from the sea buckthorn genome. The length of sea buckthorn UGT proteins ranged from 266 to 533 amino acids, the average molecular weight was 50.00 KDa, and the average isoelectric point was 5.89. According to the phylogenetic relationship, the 89 HrUGTs could be divided into 16 major groups. 84 HrUGTs were distributed on 11 chromosomes except chromosome 7. Tandem duplication was a predominant duplication event which caused the expansion of HrUGT genes. Transcriptomic data and RT-qPCR analysis indicated that most of UGT genes had a wide range of fruit development stage expression characteristics. [Conclusion] The complete information of the HrUGT gene family is obtained, which will benefit the study on the biological functions of HrUGTs. Keywords: sea buckthorn; *Hippophae rhamnoides*; UDP-glycosyltransferase; gene expression

(责任编辑:张研)