

doi: 10.3969/j.issn.0490-6756.2017.06.030

通过进化模块审视拟南芥尿苷二磷酸 葡萄糖基转移酶功能多样性

方 阳¹, 牛毓龙¹, 刘程程², 林江依³, 李旭锋¹, 杨毅¹

(1. 四川大学生命科学学院生物资源与生态环境教育部重点实验室, 成都 610064;

2. 四川大学华西口腔医院口腔疾病国家重点实验室, 成都 610041; 3. 四川大学吴玉章学院, 成都 610064)

摘要: 属于糖基转移酶(GTs)家族1的尿苷二磷酸(UDP)-糖基转移酶(UGT)在转移拟南芥中的糖基的转移和糖基化次级代谢物中起重要作用. 重建的系统发育树对这个多基因家族的进化关系和功能提供了新的见解和研究. 通过综合分析系统发育谱和重建的系统发育树, 我们系统地分析了112个拟南芥UGTs. 在我们的工作中, 我们检测到五个不同的进化保守模块, 大多数UGTs分为同一模块, 表明拟南芥UGTs在进化过程中是相当保守的. 接下来收集拟南芥UGT的底物, 并将它们映射到直向同源组, 揭示UGT在结合特异性底物中的进化关系. 通过扫描271不同物种, 还预测一些未报告的蛋白质与拟南芥UGTs密切相关.

关键词: 进化保守模块; 系统发育树; 尿苷二磷酸葡萄糖基转移酶

中图分类号: Q811.2 **文献标识码:** A **文章编号:** 0490-6756(2017)06-1317-06

An evolutionary modular view of functional diversity in *Arabidopsis* uridine diphosphate glucosyltransferase

FANG Yang¹, NIU Yu-Long¹, LIU Cheng-Cheng², LIN Jiang-Yi³, LI Xu-Feng¹, YANG Yi¹

(1. Key Lab of Bio-resources and Eco-environment of Ministry of Education, College of Life Sciences,

Sichuan University, Chengdu 610064, China; 2. State Key Laboratory of Oral Disease, West China Hospital of Stomatology, Sichuan University, Chengdu 610041, China; 3. Wu YuZhang Honors College of Sichuan University, Chengdu 610064, China)

Abstract: Uridine diphosphate (UDP)-glycosyltransferases (UGTs), belonging to the family 1 glycosyltransferases (GTs), play important roles in transferring glycosyl moieties and glycosylating secondary metabolites in *Arabidopsis thaliana*. Reconstructed phylogenomic trees can provide novel insights in studying evolutionary relationship and functions of this multigene family. By combining comprehensive phylogenetic profiles and reconstructed phylogenomic trees, we systemically analyzed 112 *Arabidopsis* UGTs. In total, we detected five distinct evolutionary conserved modules, and most UGTs were divided into the same module indicating a rather conserved evolutionary routine of *Arabidopsis* UGTs. We next manually collected the substrates of *Arabidopsis* UGTs and mapped them to orthologous groups, revealing evolutionary relationships of UGTs in binding specific substrates. By scanning 271 different species, we also predicted several unreported proteins that were closely related to *Arabidopsis* UGTs.

Keywords: Evolutionarily conserved module; Phylogenomic tree; UGTs

收稿日期: 2017-01-03

基金项目: 973计划(2015CB755702)

作者简介: 方阳(1990-),男,河南信阳人,硕士研究生,研究方向为生物信息学. E-mail: 506528950@qq.com

通讯作者: 杨毅. E-mail: yangyi528@scu.edu.cn

1 引言

糖基转移酶(GTs)是一组主要在生物体中进行糖基化反应的酶.有着广泛的核苷酸糖活性供体,例如尿苷二磷酸(UDP)-葡萄糖和 UDP-鼠李糖,GTs 能催化糖基转移到包括脂质、蛋白质和抗生素在内的多种受体分子^[1].根据其序列相似性将 GTs 分为 97 个家族其名称从 GT1 到 GT97^[2].GT1(家族 1 型)包含 UDP-葡萄糖基转移酶(UGT)其广泛分布在生物圈中,如植物、动物、真菌和细菌甚至在病毒中也存在.大多数 UGTs 都包含一个位于羧基末端共同的保守蛋白质结构域 UDPGT.在植物中该氨基酸共有序列也称为能够结合 UDP-活性糖的植物次级产物糖基转移酶 PSPG 盒子^[3].UGT 分类方法主要是:具有超过 40%的氨基酸序列同一性的 UGTs 属于一个家族,并且超过 60%的同一性的属于一个亚家族.

我们的研究主要是对模式植物拟南芥的 UGTs 进行的.为了系统分析拟南芥 UGTs 的功能,我们重建了一个具有 270 个真核生物的系统发育树和扫描 27396 拟南芥蛋白质来构建同源矩阵.通过使用聚类方法,我们将 112 个拟南芥 UGTs 分为 5 个不同的进化保守模块(ECM).为了研究 UGTs 与其底物的关系,我们查询了已经发表的关于 UGTs 的文献收集 11 种类型的底物然后分析 UGTs 及其相应底物的进化过程.在我们的研究中还预测了一些与拟南芥 UGTs 在功能上密切相关的新蛋白质.

2 材料与方法

2.1 材料

我们从拟南芥 P450,细胞色素 b5,NADPH-细胞色素 P450 还原酶和家族 1 糖基转移酶(P450)的网络资源中收集了 112 个拟南芥 UGTs^[3,4].为了进一步分析 UGTs 的底物,其中去掉了 10 种假基因(UGT85A6P,UGT90A3P,UGT71B4P,UGT72D2P,UGT89A1P,UGT76E10P,UGT84B3P,UGT76E8P,UGT90A3P 和 UGT71B3P).

2.2. 方法

2.2.1 构建系统发育谱 我们从 KEGG 数据库中获得了 268 个真核生物的全基因组蛋白质序列和注释,其中有两个物种(*Sphaeroformaarctica* 和 *Capsasporaowczaraki*)来自 Broad Institute 的多

细胞测序项目^[5].27,396 个拟南芥蛋白质序列与所选物种进行序列比对的软件用的是 BLASTP^[6].我们设置了 BLASTP E-value 为 0.001(E-value<0.001)来构建同源矩阵,其中 1 表示在相应生物体中发现的拟南芥蛋白质的同源性,0 则表示未发现.

2.2.2 鉴定进化保守模块 首先,使用 Ciccarelli 等人提供的方法重建了真核物种的系统发育树^[7].我们选取的核糖体蛋白包括(S2,S4,S5,S8,S11,S12,L3,L5,L11,L9E / L22,L10E,L15)和氨酰基 tRNA 合成酶(丝氨酰-和亮氨酰-tRNA 合成酶)为标记蛋白来构建物种系统发育树.从 KEGG 下载每个物种中的这些蛋白质的氨基酸序列并用 NCBI 基因注释进行人工的核对.然后使用 MUSCLE 软件分别对真核物种进行了多重序列比对^[8].系统树通过 iTOL 进行了可视化^[9].通过结合系统发育谱和系统发育树使用 CLIME 软件来探索 112 UGT(使用默认参数)的进化保守模块(ECM)^[10].

2.2.3 拟南芥 UGTs 间的系统发育关系及底物收集方法 重建了拟南芥 112 UGTs 系统发育树,使用的是 ETE 软件^[11].为了研究拟南芥 UGT 的各种功能,我们在 PubMed 数据库和谷歌学术中使用 112 个 UGTs 名字和 *Arabidopsis thaliana* 名称作为搜索关键词查询了已发表的文献.一共收集了 102 篇相关文章,其中筛选了主要讨论了关于糖基化中的糖苷配基的文献共 27 篇.最后获得了包括植物激素,类黄酮和异生素等在内的糖苷配基约 20 种,占拟南芥全部 UGTs 的 31%.

3 结果

3.1 重建拟南芥 UGTs 的系统发育谱

根据早期研究,如图 1 所示,除了 UGT80 和 UGT81 外,112 个拟南芥 UGT 被分为 14 组,同一组的 UGTs 家族聚集在同一分支中^[12].有趣的是组 E(UGT71,UGT72 和 UGT88)和组 L(UGT74,UGT75 和 UGT84)等这些 UGTs 家族明显的聚集到相同的组从而显示了其间的紧密关系.除了这些已知的由多个 UGTs 家族组成的组之外,一些其它组如组 H 和组 G 被置于同一单一分支中,表明拟南芥中的该多基因家族可能从不同有限的祖先 UGT 基因进化而来的.通过 Paquette 等人的研究进一步证实了,一个额外的组是 G 组(UGT85)位于 K 组(UGT86)和 J 组(UGT87)的

不同分支中^[3]. 此外, Yonekura 等人通过比较来自各种植物的 UGTs, 将 14 组拟南芥 UGTs 分成了 11 个直系同源组(OG)中^[13]. 如图 1 所示, 我们重建的系统发育关系除了组 M(UGT92)接近 OG1

(UGT73, UGT89 和 UGT90)外其他组的聚类都与 OG 分类非常一致, 而 UGT92 被定义为包含了 UGT79 和 UGT91 的 OG 8 的成员之一.

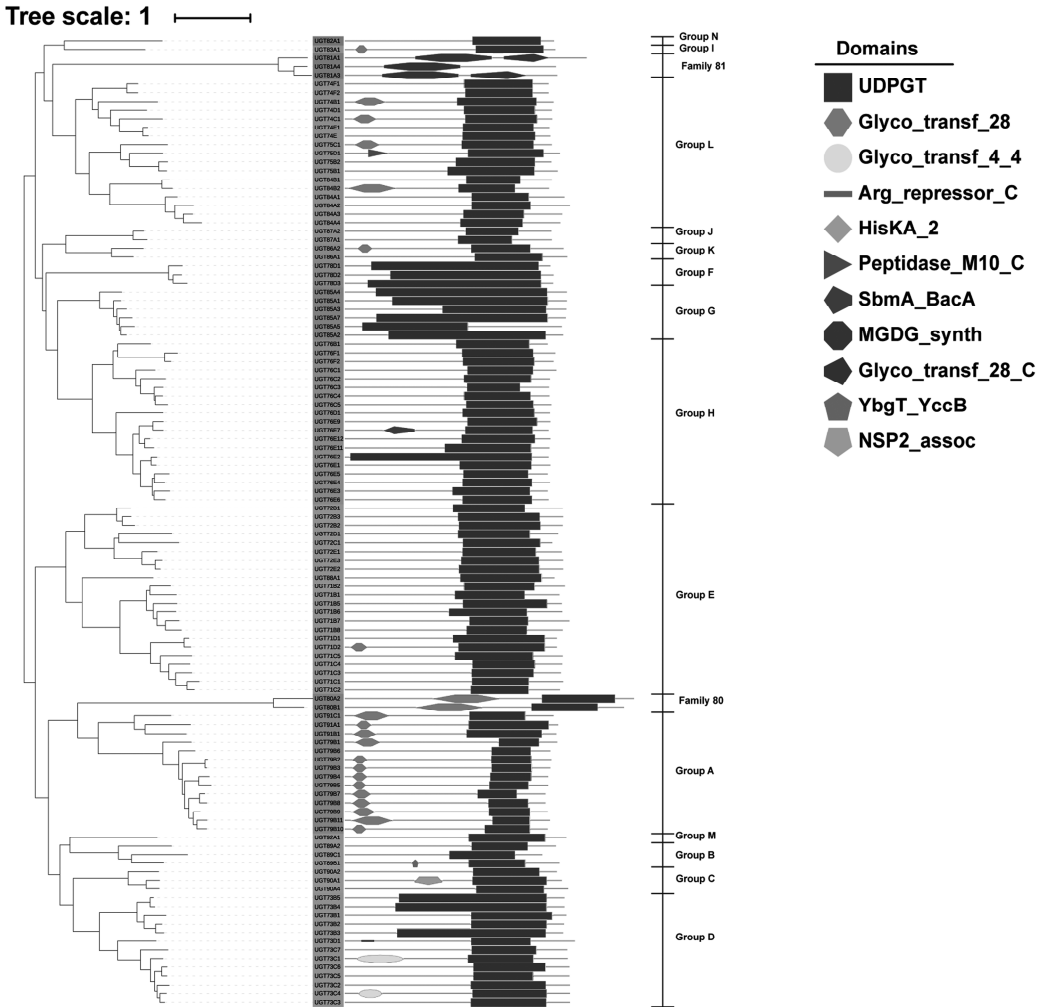


图 1 具有 112 个拟南芥 UGTs 结构域的系统发育树
Fig. 1 Phylogenetic tree with domain of 112 *Arabidopsis* UGTs

3.2 拟南芥 UGTs 的多进化模块

我们将 112 个拟南芥 UGT 分为 5 个 ECM (ECM1 至 ECM5). 如图 2 所示, 在 5 种 ECM 中, 因为它们在真核生物中具有相似的同源分布所以 83.9%(94/112)UGTs 划分到了 ECM1, 其中在大多数植物和动物中检测到同源 UGTs, 并且在真菌生物中具有类似的获得和丢失 UGTs, 表明大多数拟南芥 UGTs 可以从相同的祖先基因进化而来. 其中 ECM2 包含了几几乎所有 UGT79 的成员 (UGT79B10 除外) 和一个来自于 UGT91 家族的 UGT91A1. ECM1 和 ECM2 之间的显著差异是系统发育普中真菌范围同源 UGTs 存在大量缺失的

模式. UGT80 (UGT80A2 和 UGT80A1) 和 UGT81 (UGT81A1, UGT81A3 和 UGT81A4) 分别聚类到 ECM3 和 ECM4 进化保守模块中. 有趣的是, 这两种较小的 ECM 具有明显不同的系统发育普分布模型, ECM3 中的同源 UGTs 广泛存在于动物、植物和真菌中, 而 ECM4 中的 UGT 则主要集中在植物中. 它与图 1 所示的序列比对分析一致, UGT81 家族 (ECM4) 缺乏典型的 UDPGT 结构域, 并且 UGT80 家族的成员在氨基末端保持相同长度的结构域“Glyco_transf_28”. 值得注意的是, ECM5 是一个只包括了 UGT72B1 的单独的模块.



图 2 112 个拟南芥 UGTs 的 5 个进化保守模块

Fig. 2 Five evolutionarily conserved module of 112 *Arabidopsis* UGTs

3.3 UGT91A1 进化上更接近 UGT79 家族

以前的研究根据序列相似性表明 UGT91A1 是属于 UGT91 家族^[2]. 在图 2 中可以清楚的看到, 系统发育分析显示 UGT91A1 与真菌和动物中 UGT91 家族的其余成员具有不同的获得和缺失模型, 但是它却与 UGT79 家族具有很高相似性. 在进化的观点上表明了, UGT91A1 与 UGT79 家族可能来自于同一个祖先.

3.4 UGT72B1 具有的独特进化史

在 UGT72 家族中, UGT72E1 对松柏醛和芥子醛的体外实验证实具有高度的特异性^[14]. UGT72B1 在偶联持久性污染物 3,4-二氯苯胺 (DCA) 和 2,4,5-三氯苯酚 (TCP) 方面具有活性, 其在拟南芥中氯苯胺的代谢中发挥着重要的作用^[15]. 在图 2 中, 我们发现在系统发育谱中 UGT72B1 在 270 个真核生物中的获得和缺失模式不同于 UGT72 家族的其他成员. 虽然通过氨基酸序列比对将 UGT72B1 分类到 UGT72B 家族中, 但是 UGT72B1 的同源蛋白质仅在植物物种中发现, 表明 UGT72B1 具有特异性生物学功能.

3.5 预测与拟南芥 UGTs 相关联的蛋白质

为了预测与拟南芥 UGTs 相关的潜在蛋白质, 我们分别将 21 个 UGT 家族导入 CLIME 软件(使用默认的参数). 取对数似然比 (LRR) 阈值大于 20 的结果. 例如 AT5G24750 和 AT3G57260. 蛋白质 AT5G24750 由 UGT80 家族 (LLR = 44.8) 预测得到的, 并且被注释为推测的 UGT. 这暗示了, AT5G24750 与 UGT80 家族在生物学上可能具有非常相似的功能, 或是 UGT80 成员参与生物学反应通路中的一个成员. 早期研究通过用 UGT80A2 和 UGT80B1 进行序列相似性搜索, AT5G24750 被预

测为 UGT80 的第三个家族成员, 并且根据标准 UGT 命名法进一步命名为 UGT713B1^[16], 这与我们的结果是相一致的. 尽管 AT5G24750 作为 UDP-葡萄糖: 固醇葡萄糖基转移酶的作用尚未在体外得到支持^[17], 但它可能识别不是类固醇的其他特定的底物, 仍然需要进一步的研究.

3.6 通过结合的底物推断 UGTs 的进化过程

为了解释拟南芥不同的 UGTs 具有多样的功能, 我们使用 *Arabidopsis thaliana* 和 112 个 UGTs 名字等作为关键字在 PubMed 和谷歌学术中查询相关的文献资料. 我们搜索 102 篇相关的文献报道和通过人工筛选出 57 篇包括 37 种拟南芥 UGTs 的文献, 其中占总拟南芥 UGTs (112 UGTs) 的 31%. 这些已经报道的文献研究一共包括 11 种底物: 包括植物激素类、油菜素内脂、细胞分裂素、脱落酸 (ABA)、次级代谢产物等. 这些底物在糖基反应中起着非常重要的作用, 例如, 脱落酸 (ABA), UGT71C5 在拟南芥 ABA 稳态中起主要作用, UGT71C5 的过表达可以降低 ABA 表达量和增加 ABA-GE 的表达量^[18].

为了探索拟南芥 UGTs 在进化中的关系, 我们用底物聚类来分析 UGTs 的在进化中的关系. 如图 3A 所示, 通过对 40 个 UGTs 和 11 个底物的聚类分析显示了 UGTs 具有相同的催化底物是聚集在一起的. 聚类结果还表明了, UGT71、UGT72、UGT84 家族是与木质素前体聚类到一起的. UGT73、UGT78、UGT79、UGT89 和 UGT75 家族与黄酮底物聚类在一起. UGT71、74、89、75 和 72 家族与 SA 聚类在一起的. UGT71 和 UGT84 家族与 ABA 聚类在一起, UGT75 和 UGT76 与抗花青素聚类在一起, UGT72、UGT84、GT75 与异生素聚类在一起.

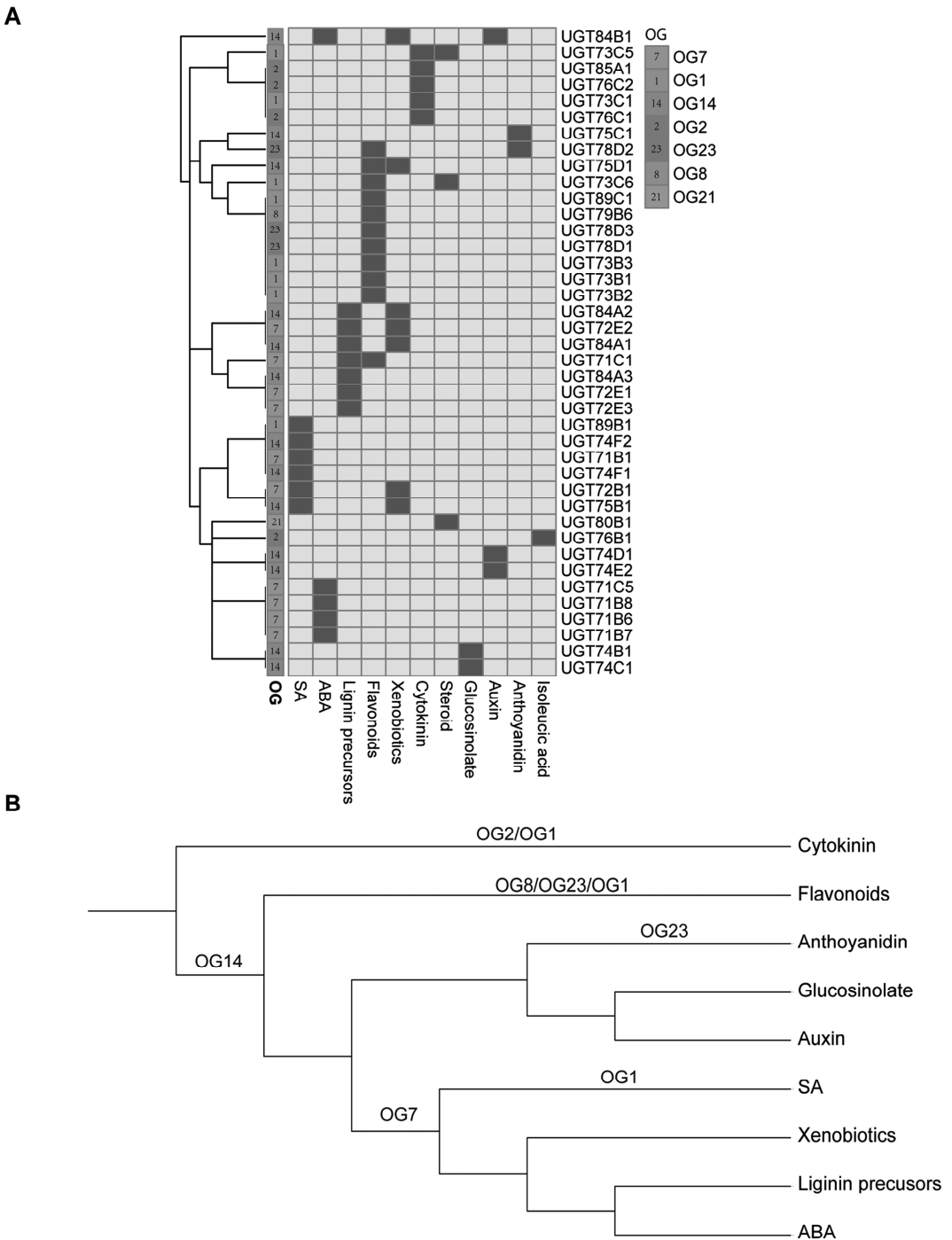


图 3 拟南芥 UGTs 进化分析
Fig. 3 Evolutionary Analysis of *Arabidopsis* UGTs

4 讨论

拟南芥 UGTs 家族是个多基因家族, 并且未知功能的 UGTs 仍然非常多的. 在拟南芥植物中, 已经鉴定功能的 UGTs 的比例是小于 20% 的^[19]. 与以前报道的拟南芥 UGTs 的系统发育分析相比较^[3], 我们对拟南芥 UGTs 进行了更为全面的系统分析, 从而增加了人们对 UGTs 进化的更进一

步理解. 如图 3B 所示, 将我们的聚类结果与 24 个同源组(OG)进行了比较^[13], 发现 OG14(UGT74、UGT75、UGT84)中的 UGTs 可以识别除细胞分裂素外的几乎所有底物, 表明 OG14 比其他 OG 更保守并且更接近祖先状态. OG14 的多样功能可能来自基因的多拷贝^[19]. 此外, 我们发现 OG7(UGT71、UGT72 和 UGT81)能够接受四种不同的底物(木质素前体, SA, ABA, TCP), 表明 OG7

是 OG14 的后代。OG14 作为祖先,具有催化更多底物的能力,而由于基因复制的结果,OG7 能够催化的底物仅是 OG14 的子集。如图 3B 所示,我们还预测了 OG14 的一些其他的子代,例如 OG23 (UGT78)和 OG8(UGT79、UGT91 和 UGT92)可以催化两种相似的底物花青素和黄酮类。

参考文献:

- [1] Lairson L L, Henrissat B, Davies G J, *et al.* Glycosyltransferases: structures, functions, and mechanisms [J]. *Annu Rev Biochem*, 2008, 77:21.
- [2] Lombard V, Golaconda Ramulu H, Drula E, *et al.* The carbohydrate-active enzymes database (CAZy) in 2013 [J]. *Nucleic Acids Res*, 2014, 42: D490.
- [3] Paquette S, Moller B L, Bak S. On the origin of family 1 plant glycosyltransferases [J]. *Phytochemistry*, 2003, 62: 399.
- [4] Paquette S M, Jensen K, Bak S. A web-based resource for the Arabidopsis P450, cytochromes b5, NADPH-cytochrome P450 reductases, and family 1 glycosyltransferases (<http://www.P450.kvl.dk>) [J]. *Phytochemistry*, 2009, 70: 1940.
- [5] Ruiz-Trillo I, Burger G, Holland P W, *et al.* The origins of multicellularity: a multi-taxon genome initiative [J]. *Trends Genet : TIG*, 2007, 23: 113.
- [6] Camacho C, Coulouris G, Avagyan V, *et al.* BLAST+: architecture and applications [J]. *BMC bioinformatics*, 2009, 10:21.
- [7] Ciccarelli F D, Doerks T, Von Mering C, *et al.* Toward automatic reconstruction of a highly resolved tree of life [J]. *Science*, 2006, 311: 1283.
- [8] Edgar R C. MUSCLE: multiple sequence alignment with high accuracy and high throughput [J]. *Nucleic Acids Res*, 2004, 32: 1792.
- [9] Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees[J]. *Nucleic Acids Res*, 2016, 44: W242.
- [10] Li Y, Calvo S E, Gutman R, *et al.* Expansion of biological pathways based on evolutionary inference [J]. *Cell*, 2014, 158: 213.
- [11] Huerta-Cepas J, Serra F, Bork P. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data [J]. *Mol Biol Evol*, 2016, 33: 1635.
- [12] Ross J, Li Y, Lim E, *et al.* Higher plant glycosyltransferases [J]. *Genome Biol*, 2001, 2: REVIEWS3004.
- [13] Yonekura-Sakakibara K, Hanada K. An evolutionary view of functional diversity in family 1 glycosyltransferases [J]. *Plant J*, 2011, 66: 182.
- [14] Lim E K, Jackson R G, Bowles D J. Identification and characterisation of Arabidopsis glycosyltransferases capable of glucosylating coniferyl aldehyde and sinapyl aldehyde [J]. *Febs Lett*, 2005, 579: 2802.
- [15] Brazier-Hicks M, Edwards R. Functional importance of the family 1 glucosyltransferase UGT72B1 in the metabolism of xenobiotics in Arabidopsis thaliana [J]. *Plant J*, 2005, 42: 556.
- [16] Mackenzie P I, Owens I S, Burchell B, *et al.* The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence [J]. *Pharmacogenetics*, 1997, 7: 255.
- [17] Stucky D F, Arpin J C, Schrick K. Functional diversification of two UGT80 enzymes required for steryl glucoside synthesis in Arabidopsis [J]. *J Exp Bot*, 2015, 66: 189.
- [18] 岑咏一, 刘震, 李旭锋, 等. UGT75B1, UGT71B6, UGT71C5 酶活性及酶反应动力学分析[J]. *四川大学学报: 自然科学版*, 2016, 53: 169.
- [19] Hudson W H, Kossmann B R, De Vera I M, *et al.* Distal substitutions drive divergent DNA specificity among paralogous transcription factors through subdivision of conformational space [J]. *P Natl Acad Sci USA*, 2016, 113: 326.