

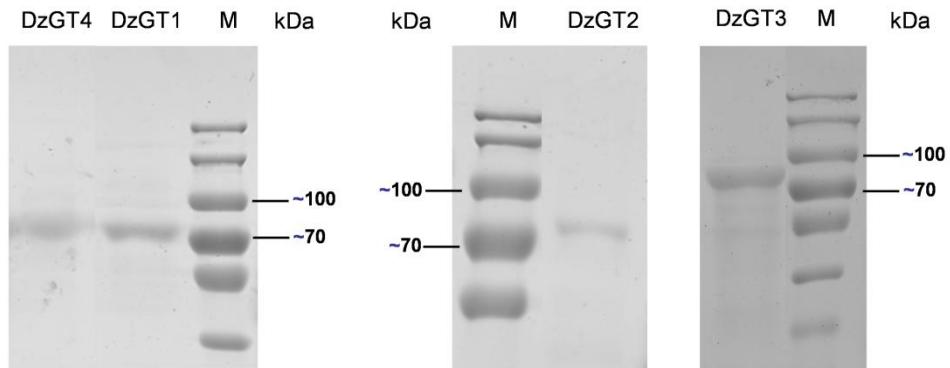
## ***Supplementary Material***

### **1 Supplementary Figures and Tables**

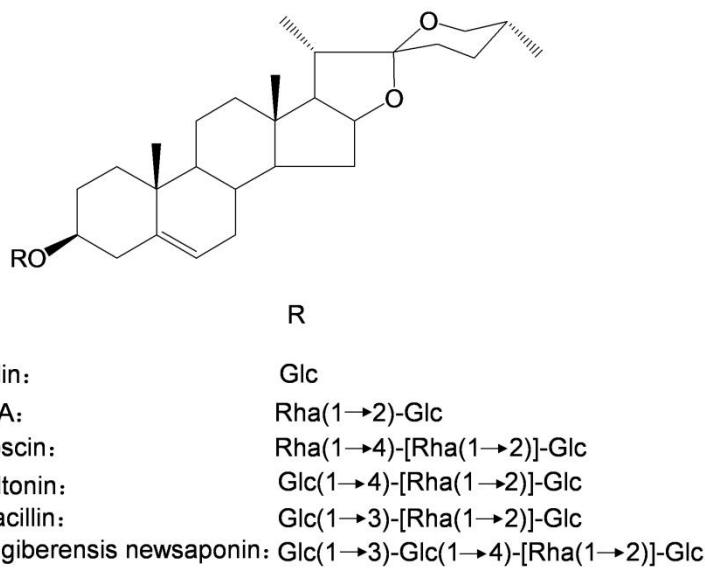
#### **1.1 Supplementary Figures**

DzGT1	MATHAGHEPRPLRVFFIFPFATGHMIPMVDIARLFAARGVADTVLVTPSNAALISRTIDD	60
DzGT3	-MGTETKTQVPLRFFFPLMSPGHIMPMIDVAKLFARGATTIITTPANEFPARPAIRD	59
DzGT2	--MTSKAKDNHLHILFFFPMAGGHIIPTLDLAKAFATHGVTTILTPSYSPLIQRTINH	57
DzGT4	-MGSNTPPHNLHILFFFPMAPGHAIPMLDMITKLFALHGVRTTFITTPANAPLVQPTINR	59
	*:::;*:;:: ** *;*:; ** :*. ;*:;**: : :;	
DzGT1	AAA---SGLPIRTLLYPPSAEV-GLPPGAENIASVPLADAPKVD-AASLLTAGDHDR1	114
DzGT3	ANSSSPPLPHPINLIFLPFPPSVHHLLPAGHENFSSITSDDDFNKLFTALFALDTPFGHL	119
DzGT2	YS----GQLPIHLSLISFPTTIS-GLPSGCDSVTSLSQHERRIIFLCSLSSLQPFKRA	111
DzGT4	ANSANP-TAPPKLTLPFFPSAEV-GLPSGCENAQSLESMDRSNFYKAVMMLRQPFDQV	117
	**. : **;: ** * . *. : . : : .	
DzGT1	LRLHHPDAVVSDFTHFAWTTSIARDLIRPRTFHAILGLFPVCVLGSLIRNLPHLSVTDD-V	173
DzGT3	LQSHQPDCLVSDSVFTWSAAISADLNIPRLIFHGPGINFTLCIHDNLNLDV----YGQA	174
DzGT2	LTELRPDCIITDMFFPWTSAVARELQIPRLVFQSESEFFSLCMTDSLRLYKPHLTISDDID	171
DzGT4	LKELHPDAIVTDSFLPWTTDAASEFGIPRLVFHGTSFALCAHWNVDEYNPHE---RPE	173
	* : ;***.::: * : *;: : : ***; *:. . * : * ..	
DzGT1	HPIVVPDLPHPVQMFRPELPDFLRSENTAIQAMSALNEAEGSLGVVNSFAEIESAYA	233
DzGT3	TTFTIQDIPHDIHMNRWEVPNIFTL-P---MWLQMLRDAEKDSYGVVVNSFYELEPDYA	229
DzGT2	ESFILGPLPHPIKMSKSQILDPTK1-TPEFSELFSLMGESEAESYGVVVNSFLEPEYA	230
DzGT4	ESFTIPGLPHRIHMLKSQVPIPKT-NPSFTDVMRAVKESDAKSYGAVVNSFYELEPEYY	232
	* : .: **.::: * : : : : : : : * . * . * . * . * .	
DzGT1	DCYYKFDNMRTWFVGPVALADGGEKASRGAAADAKAVANRELCMRWLKDQRERSVYYACF	293
DzGT3	ELYRRKPGRAWVYGPVSLCSQRVGDD--DDA--SKDHYTECLKWLDSKKPGSVVYCF	286
DzGT2	KHYREFIGRAWVYGPVSLCSQRVGDD--DDA--SKDHYTECLKWLDSKKPGSVVYCF	285
DzGT4	DHYRNVMGRRAWVYGPVSLCNEEVVDSL--AGRGEAVIDQDECLNWLDNKOPNSVLYCF	290
	. * . . *;:;***: * . : . : . * . * . * . * .	
DzGT1	GSWCHFSAAQLRELALGLEAAGHPFLWVRDDE--EWMPEGFEELGKGLVVRGWAPQ	351
DzGT3	GSLCVFSVDQVEEIAGLEISGHCFWVVRDKGPST---PEGFEERVKGRGVVVRGWAPQ	343
DzGT2	GSMFSMPSEQNTELAIGLEACEHPIIWVIRMKGHDHEE-KKVRVLEGGKGVINGWAPQ	344
DzGT4	GSVGRFTSQAQLHEIALGLESSRQPFIWVVKLFDDEREWLPEYEKRVEGRGLIIRGWAPQ	350
	** : *;: *;: * . * . : *;: *** . : *;: * . *	
DzGT1	VAVLGHDAVGGFVTHCGWNSVLEGVSSGLPMVTWPLSTEQFINEKLVVEVLGTAVEREG	411
DzGT3	VVLLGHVAVGGFVNHCWGWSVLEGISAGLPMVTWPLMHEQFINEKLVNEVLGIGVRARRE	403
DzGT2	LVLINHPAVGAFITHCGWNSTMESICAGVPMITWPLFGDQFYNERFIIDVVLKVGVEVGK	404
DzGT4	ILYLINHEAVGGFLTHCGWNSALEAVCAGVPMVTMPMLAEQFFNERLLVDVLKIGVGIGVK	410
	: *.* . * . * . * . * . : * . * . * . * .	
DzGT1	--FRSTEEGEKEVVKAGDVAAAEVKMGGGEEGEKKRVTAREFGEKAKAAVKGG	464
DzGT3	--G-----DQPTVVIIGREEVAGAVRRLMDGGDEAEVRRGKAREYARMARTAIEDGG	452
DzGT2	KFVVD-----EEDMELVKGEDVKKAVERMMGDGDEKESMRRLMELGEMARRAVEEGG	457
DzGT4	EYGTGQKFGMTENVASALVCGEDIEKAVAMLMNAGEEAEEMRRRARELRESAKMALEKGG	470
	: : : : : * . * . * . * . * . * . * . * . * .	
DzGT1	SSYEGLSRLIGEIRDWPKKINGCDAATA---	492
DzGT3	SSRVDLNRLMDELTRQRGRVVGEEKVETSSN	483
DzGT2	SSYVDLKNLNLINELCEVCSEHDMLNSKQ---	484
DzGT4	SSSLDICNLIRELTEKQRPVPSLY-----	494
	** . : . * : * : :	

**Supplementary Figure 1.** The amino acid sequence alignment of DzGT1-DzGT4. The putative plant secondary product glycosyltransferase (PSPG) box is highlighted by red colour.



**Supplementary Figure 2.** SDS-PAGE electrophoresis analysis of the purified recombinant DzGTs.



R

Trillin:

Glc

PSA:

Rha(1→2)-Glc

Dioscin:

Rha(1→4)-[Rha(1→2)-Glc]

Deltonin:

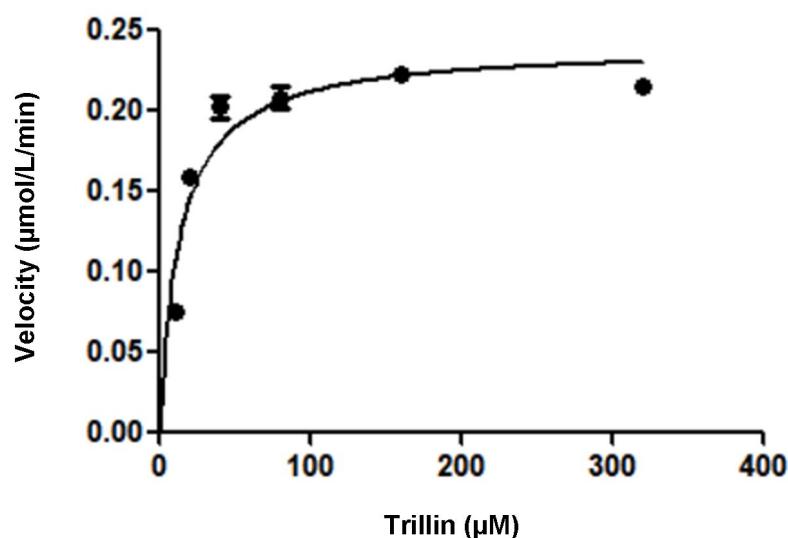
Glc(1→4)-[Rha(1→2)-Glc]

Gracillin:

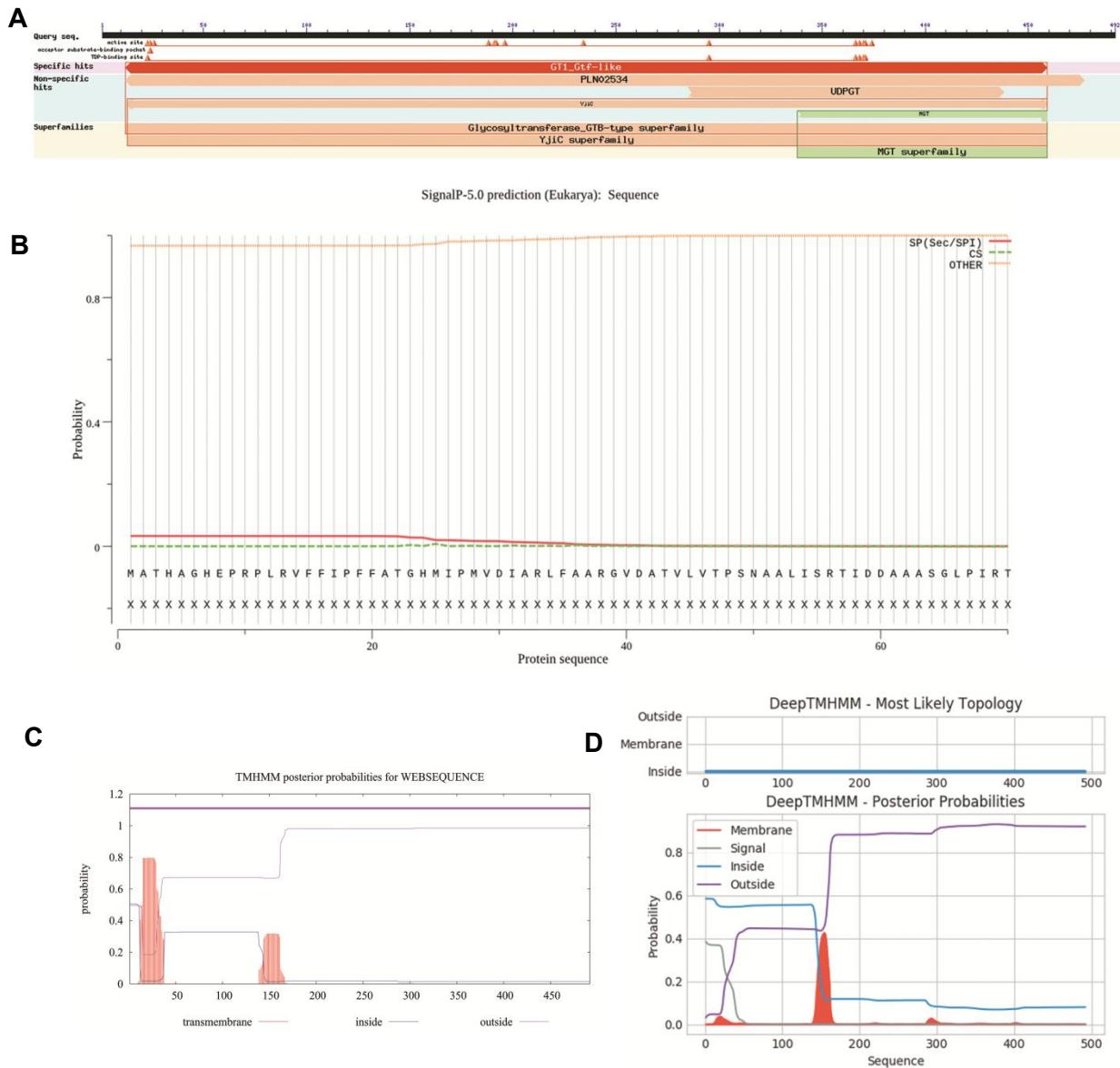
Glc(1→3)-[Rha(1→2)-Glc]

Zingiberensis newsaponin: Glc(1→3)-Glc(1→4)-[Rha(1→2)-Glc]

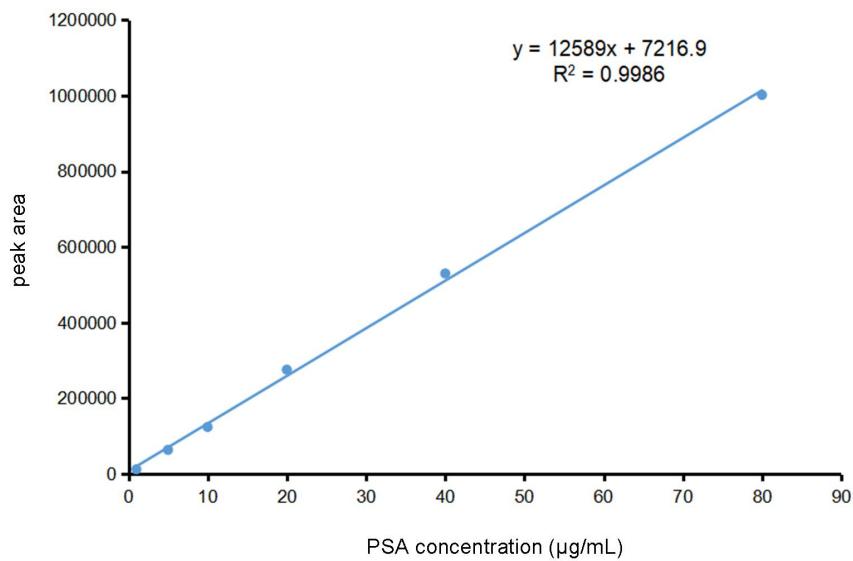
**Supplementary Figure 3.** Structure of the substrates used for the enzyme assays of DzGT1.



**Supplementary Figure 4.** Nonlinear regression analysis of DzGT1 with trillin using the GraphPad Prism 5.01 software. The best-fit regression equations is  $V = 0.24[S]/(13.05 + [S])$ . V, enzyme velocity ( $\mu\text{mol/L/min}$ ); [S], substrate (trillin) concentration.



**Supplementary Figure 5.** Bioinformatic analysis of DzGT1. **(A)** Putative conserved domains of DzGT1 detected by the NCBI conserved domains database. **(B)** Signal peptide prediction of DzGT1 with SignalP 5.0. **(C, D)** Transmembrane topology prediction of DzGT1 with TMHMM2 and DeepTMHMM.



**Supplementary Figure 6.** The calibration curve of PSA. The linear fit equation is  $y = 12589x + 7216.9$  and  $R^2 = 0.9986$ .

## 1.2 Supplementary Tables

**Supplementary Table 1.** Information of the glycosyltransferases used in the phylogenetic analysis

Protein	GenBank accession Number	Plant	Function
UGT80A1	CAB06081	<i>Avena sativa</i>	Sterol 3- <i>O</i> -glucosyltransferase
UGT80B1	OAP13743	<i>Arabidopsis thaliana</i>	Sterol 3- <i>O</i> -glucosyltransferase
GhSGT1	AHX00584	<i>Gossypium hirsutum</i>	Sterol 3- <i>O</i> -glucosyltransferase
SGTL1	ABC96116	<i>Withania somnifera</i>	Sterol 3- <i>O</i> -glucosyltransferase
Dz3GT1	AVI57699	<i>Dioscorea zingiberensis</i>	Sterol 3- <i>O</i> -glucosyltransferase
Dz3GT2	AVI57700	<i>D. zingiberensis</i>	Sterol 3- <i>O</i> -glucosyltransferase
GmSGT3	NP_001240857	<i>Glycine max</i>	soyasaponin III rhamnosyltransferase
StSGT3	NP_001274799	<i>Solanum tuberosum</i>	$\beta$ -solanine/ $\beta$ -chaconine rhamnosyltransferase

**Supplementary Table 2.** Primers used for the gene amplifications in this study

Name	Primers 5' to 3' sequence
For pMD18-T /pGEX-2T	
<i>DzGT1</i> -F	GATCTGGTTCCCGCGTGGATCCATGCCACCCATGCC
<i>DzGT1</i> -R	TCAGTCAGTCACGATGAATTCTCAAGCGGTAGCAGCATC
<i>DzGT2</i> -F	GATCTGGTTCCCGCGTGGATCCATGACTCAAAAGCCAAAGA
<i>DzGT2</i> -R	TCAGTCAGTCACGATGAATTCTTATTGCTTGGACAGATTCA
<i>DzGT3</i> -F	GATCTGGTTCCCGCGTGGATCCATGGGGACTGAAACAAAAAC
<i>DzGT3</i> -R	TCAGTCAGTCACGATGAATT CCTAGTTGAAGAGATTCTACTTT
<i>DzGT4</i> -F	GATCTGGTTCCCGCGTGGATCCATGGGATCCAACAGTACTACT
<i>DzGT4</i> -R	TCAGTCAGTCACGATGAATT CCTAGTATAAACTAGGTACTGGTCT
For real-time PCR	
<i>Actin</i> -F	TGCTGGATTCTGGTGATGGT
<i>Actin</i> -R	CCCGTTCTGCGGTAGTAGTG
<i>DzGT1</i> -F	TTCCTCCGCTCCGAAAACA
<i>DzGT1</i> -R	CGAACCACGTCCGCATATTG