Electronic Supplementary Information

A Highly Selective 2”-*O*-Glycosyltransferase from *Ziziphus jujuba* and De novo Biosynthesis of Isovitexin 2”-*O*-glucoside

Ya-Qun Zhang,[‡a] Zi-Long Wang,[‡a] Zhuo Chen,[‡b,c] Zheng-Tong Jin,[a] Aobulikasimu Hasan,[a] Hai-Dong Wang,[a] Yu-Wei Sun,[b] Xue Qiao,[*a] Yong Wang[*b] and Min Ye[*a]

[a]State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Sciences, Peking University, 38 Xueyuan Road, Beijing 100191, China

[b]CAS-Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai 200032, China

[c]University of Chinese Academy of Sciences, Beijing 100039, China

Corresponding Authors

* E-mail: yemin@bjmu.edu.cn.
* E-mail: yongwang@cemps.ac.cn.
* E-mail: qiaoxue@bjmu.edu.cn.

‡ Y. Z., Z. W., and Z. C. contributed equally to this work.

Table of Contents

1. Experimental Procedures

1.1 General remarks

Compounds **1**, **4**, **5**, and **6** were previously prepared by our group using enzymatic catalysis. Isovitexin (**2**), vitexin (**3**), and all the other substrates used in this study were purchased from Shanghai Yuanye Bio-Technology Co., Ltd (Shanghai, China). The UDP-sugar donors were purchased from Guangzhou Angfei Biological Technology Co., Ltd, (Guangzhou, China). Methanol and acetonitrile (Fisher Scientific, USA) were of HPLC grade. All other chemicals and reagents were purchased from Sigma-Aldrich (St. Louis, MO, USA) and Beijing Chemical Corporation (Beijing, China) unless otherwise specified. HPLC analysis was conducted on an Agilent 1260 instrument (Agilent Technologies, Waldbronn, Germany), or a Thermo Vanquish (Thermo Scientific, Shanghai) UHPLC system. LC/MS analysis was performed on Q-Exactive Orbitrap LC/MS (Thermo Fisher, CA, USA). Chemical analysis of different parts of *Ziziphus jujuba* was conducted on the Thermo UHPLC system coupled with a Q-Exactive hybrid quadrupole-Orbitrap mass spectrometer equipped with a heated ESI source (Thermo Fisher Scientific, San Jose, CA, USA). The glycosylated products were isolated and purified by semi-preparative HPLC on an Agilent 1200 instrument (Agilent Technologies, Waldbronn, Germany). NMR spectra were recorded at 25°C on a Bruker AVANCE 600 instrument at 600 (^{1}H) and 150 (^{13}C) MHz in DMSO-d_6. Chemical shifts (δ) are given in parts per million (ppm) and coupling constants (J) are given in Hertz (Hz).

1.2 Plant materials

Fresh leaves, stems, and kernels of *Z. jujuba* were obtained from a plant collected in Shanxi Province (China), and were immediately frozen before chemical analysis. The leaves were used for DNA barcoding analysis. The kernels were stored at -80°C and were used for transcriptome sequencing and RNA extraction.

1.3 Species identification and chemical analysis of *Ziziphus jujuba*

The total genome was extracted using Plant Genomic DNA Kit (Tiangen Biotech, China). Universal primer pairs ITS2F-ITS2R and psbK-psbI were used to amplify the fragments with total genome as template. The sequencing results were uploaded and blasted in NCBI. The plant was identified as *Ziziphus jujuba* Mill. var. *spinosa* (Bunge) Hu ex H. F. Chow.

The chemical constituents of different parts of the plant were analyzed using Q-Exactive Orbitrap LC/MS. An aliquot of 10 mg sample was ground and extracted in 1 mL of 50% methanol in an ultrasonic water bath for 30 min. After filtration through a 0.22-μm membrane, the subsequent filtrate was used for UHPLC/Orbitrap-MS analysis. The samples were separated on an Acquity UPLC HSS T3 column (2.1 × 100 mm, 1.8 μm). Water containing formic acid (0.1%, *v*/*v*) (A) and methanol (B) were used as mobile phase. The gradient elution program was as follows: 0 min, 20% B; 4 min, 35% B; 7 min, 40% B; 11.5-13min, 100% B; 13.1-16min, 20% B. The flow rate was 0.3 mL min^{-1}, and the column temperature was not controlled. The DAD scanned from 210 to 400 nm. The Orbitrap parameters were set as follows: sheath gas pressure 45 arb, aux gas pressure 10 arb,

discharge voltage 4.5 kV, capillary temperature 350 °C. MS^1 resolution was set as 70,000 FWHM, AGC target 1*E6, maximum injection time 50 ms, and scan range *m/z* 100-1000. MS^2 resolution was set as 17,500 FWHM, AGC target 1*E5, maximum injection time 100 ms.

1.4 Transcriptome sequencing, data assembly, and candidate gene screening

The transcriptome of fresh kernels was sequenced and assembled by Novogene Bioinformatics Technology Co., Ltd. (Beijing, China). The raw data were uploaded in NCBI with accession No. PRJNA764572. For BLAST analysis, the gene sequences of 38 functionally characterized plant glycosyltransferases (including 20 OGTs and 18 CGTs) were used as templates (Table S1). With these query sequences, 56 open reading frames (ORFs) (including 44 OGT candidate genes and 12 CGT candidate genes) were obtained as candidate genes from the transcriptome of *Z. jujuba*.

1.5 Molecular cloning, heterologous expression, and protein purification

The total RNA from fresh semen of *Ziziphus jujuba* var. *spinosa* was extracted using the TranZol™ kit (Tiangen Biotech, China) and was reverse-transcribed to cDNA by TransScript supermix (Transgen biotech, Beijing, China) following the manufacturer's instructions. The full-length of candidate genes were amplified by PCR using the gene-specific primer pairs and TransScript KD Plus and Fast-Pfu DNA polymerases (Transgen biotech, Beijing, China). Sequences of these primers were listed in Table S2. After verification of sequences by Tsingke Biotech (Beijing, China), the amplification products of candidate genes were respectively inserted into pET28a (+) vector (Invitrogen, USA) using Quick-change method. The recombinant plasmids were transformed into *Escherichia coli* BL21 (DE3) (TransGen Biotech, China) for heterologous expression. *E. coli* cells were grown in 500 mL Luria-Bertani (LB) medium containing 50 μg/ml kanamycin at 37°C with shaking (200 rpm). After OD_{600} reached 0.4-0.6, the cells were induced by adding 0.1 mM Isopropyl β-D-thiogalactoside (IPTG), and shaking at 16°C (100rpm) for 15 h. The culture was centrifuged to collect the cells, which were resuspended in 15 mL of lysis buffer (pH 8.0, 50 mM NaH_2PO4, 30 mM NaCl, 10 mM imidazole) and ruptured by sonication on ice. The expressed protein was isolated from the cell debris by centrifugation at 12,000 rpm for 50 min at 4°C. The supernatant including crude protein was applied to a nickel-affinity chromatographic column (TransGen Biotech, China) which had been pre-equilibrated with lysis buffer. Miscellaneous proteins were eluted by elution buffer containing 30 mM imidazole, followed by elution of the purified proteins in a buffer containing 300 mM imidazole. The eluent was concentrated using Amicon Ultra-15 Ultracel-30K (Merck Millipore). The purified proteins were analysis in SDS-PAGE (Fig. S3), and stored in a storage buffer (50 mM Tris-HCl, 20% glycerol, pH 7.5) at -80°C.

1.6 Glycosyltransferase activity assay

The reaction was carried out in a 100-μL system containing 50 mM Tris-HCl (pH 8.0), 0.1 mM substrate, 0.5 mM sugar donor, and 5 μg of purified enzyme at 37 °C overnight. Then the reaction was ended with 200 μL methanol and analyzed by HPLC. For dehydration, 50 mM HCl was added

into the system after enzymatic reaction. Samples were concentrated to dryness by centrifuging in vacuum, and redissolved in 100 μL 50%methanol. An Agilent 1260 series HPLC instrument (Agilent Technologies, Waldbronn, Germany) was used. Samples were separated on an Agilent ZORBAX SB-C18 column (4.6 × 250 mm, 5 μm). The mobile phase consisted of methanol (A) and water containing 0.1% formic acid (*v*/*v*, B). The analytes were eluted using a linear gradient of 20% to 100% methanol in H_2O containing 0.1% formic acid in 20 min, and followed by 100% methanol for 5 min. The flow rate was 1 mL/min. The column temperature was 30°C. The detection wavelength was 300 nm. MS analysis was performed on the UHPLC-Q-Exactive HRMS instrument.

1.7 Preparative-scale reactions

To prepare compound **5a**, 50 mM nothofagin was dissolved in methanol (MeOH) as sugar acceptor, and 50 mM UDP-Glc was used as sugar donor. The reaction mixture contained 1 ml buffer (50 mM Tris-HCl, pH 8.0), 20 μL sugar acceptor (50 mM), 50 μL sugar donor (50 mM), and ZjOGT38. The reaction was performed at 37°C overnight and terminated by adding two volume of MeOH. The mixture was then centrifuged at 8,000 rpm for 30 min, and the supernatant was concentrated to dryness and dissolved in 1.5 mL 50%methanol. The glycosylated product **5a** was subsequently purified by reversed-phase semi-preparative HPLC and characterized as nothofagin 2''-*O*-glucoside by NMR analyses (Fig. S9-S12) with 68% yield.

1.8 Effects of reaction time, pH, temperature, and divalent metal ions

To optimize the reaction pH, the enzymatic reactions were carried out in various buffers ranged in pH values from 4.0-6.0 (citric acid-sodium citrate buffer), 6.0-8.0 (Na_2HPO_4-NaH_2PO_4 buffer), 7.0-9.0 (Tirs-HCl buffer), and 9.0-10.6 (Na_2CO_3-$NaHCO_3$ buffer). To investigate the optimal reaction temperature, the enzymatic reactions were incubated at different temperatures (4-60 °C). To test the dependence of divalent metal ions for ZjOGT38 activity, different divalent cations Mg^{2+}, Ba^{2+}, Fe^{2+}, Co^{2+}, Cu^{2+}, Zn^{2+}, Mn^{2+} and EDTA in a final concentration of 5 mM were used, individually. All enzymatic reactions were conducted for 1 h with UDP-glucose (UDP-Glc) as the sugar donor and nothofagin as the acceptor. All experiments were performed in triplicate and the mean value was used. The reactions were terminated with pre-cooled methanol and centrifuged at 12,000 rpm for 20 min for HPLC analysis. The samples were analyzed on an Agilent Zorbax SB C18 column (250 mm×4.6mm, 5 μm) protected by a Zorbax Extend-C18 guard column (4.6 ×12.5 mm, 5 μm) at a flow rate of 1 mL min^{-1}. The column temperature was 30°C and the enzymatic products were eluted with a linear gradient of 20% to 100% methanol in H_2O containing 0.1% formic acid in 20 min, followed by 100% methanol for 5 min. The conversion rates were calculated from HPLC peak areas of glycosylated products and substrates.

1.9 Kinetic studies

For kinetic studies of ZjOGT38, enzymatic assays were performed in a final volume of 50 μL containing 50 mM Tris-HCl (pH 8.0), 0.125 μg of purified ZjOGT38, 500 μM of saturated UDP-

Glc, and varying concentrations (2-40 μM) of nothofagin (**5**). The reactions were conducted at 37°C for 10 min, and were then quenched with 100 μL ice cold methanol. Samples were centrifuged at 12,000 rpm for 20 min and analyzed by reversed-phase HPLC as described above. All experiments were performed in triplicate. Michaelis-Menten plot was fitted.

1.10 Molecular modeling and molecular docking

The protein model of ZjOGT38 was calculated based on Alphafold2. The model structure displays conserved GT-B fold with two β/α/β Rossmann domains. UDP-Glc was simulated to ZjOGT38 according to GgCGT/UDP-Glc complex structure (PDB ID: 6L5P). Then we used software AutoDock 4.2 in Auto-Dock-Tools (ADT) from the MGLTools (Molecular Graphics Laboratory tools) for molecular docking. We constructed a series of complex structures of protein, UDP-Glc, and substrates (the axis of grid box is x=-32.873, y=-21.953, and z=-8.527). Then we performed Auto-Dock by Lamarckian Genetic Algorithm with default parameters for 2,500,000 evaluations in 100 cycles, and the other parameters followed the default information [1]. The model according to binding energy and reasonable conformation in sugar acceptor binding pocket was analyzed.

1.11 Analysis of De novo biosynthesis products

The product was extracted by an equivalent volume of ethyl acetate for three times. The extract was combined and evaporated to dryness. Then 1 mL methanol was added to dissolve products from around 10 mL fermentation culture. Samples were centrifuged at 15,000 rpm for 20 min, and the supernatant was injected into UHPLC for analysis. The Thermo Vanquish UHPLC system consisted of a quaternary solvent manager, an autosampler with a cooling system, a column heater, and a DAD detector. An Acquity UPLC HSS T3 column (1.8 mm, 2.1×150mm) equipped with a VanGuard pre-column was used for sample separation. The organic mobile phase (B) was methanol, and the aqueous phase (A) was water containing 0.1% (*v*/*v*) of formic acid. A linear gradient elution program was used as follows: 0-8 min, 20% B; 8-10 min, 20%-35%B; 10-18 min, 35%B; 18-19 min, 35%-20%B. The flow rate was 0.3 mL/min, and the column temperature was 45°C. The DAD detector scanned from 190 to 400 nm (extracted at 275 nm). The sample tray temperature was maintained at 15°C. A 2-μL aliquot of sample was injected for analysis.

2. Supplementary Tables

Table S1. Detailed information of reported flavonoid sugar-sugar OGTs (1-20) and CGTs (21-42) in the phylogenetic analysis

No.	Enzyme	Accession number	Plant species	Reported activity	Reference
1	AtUGT79B1	Q9LVW3	*Arabidopsis thaliana*	Anthocyanidin 3-*O*-glucoside [1,2]-xylosyltransferase	[2]
2	F3GGT1	FG404013	*Actinidia chinensis*	Cyanidin 3-*O*-galactoside [1,2]-xylosyltransferase	[3]
3	In3GGT	Q53UH4	*Ipomoea nil*	Anthocyanidin 3-*O*-glucoside [1,2]-glucosyltransferase	[4]
4	BpUGAT	BAD77944	*Bellis perennis*	Anthocyanidin 3-*O*-glucoside [1,2]-glucuronosyltransferase	[5]
5	Ph3RT	CAA81057	*Petunia hybrida*	Anthocyanidin 3-*O*-glucoside [1,6]-rhamnosyltransferase	[6]
6	LeABRT2	BAU68118	*Lobelia erinus*	Delphinidin 3-*O*-glucoside [1,6]-rhamnosyltransferase	[7]
7	PdUGT85A19	ABV68925	*Prunus dulcis*	Cyanohydrin glucoside [1,6]-glucosyltransferase	[8]
8	CsUGT707B1	CCG85331	*Crocus sativus*	Flavonol 3-*O*-glucoside [1,2]-glucosyltransferase	[9]
9	AtUGT79B6	Q9FN26	*Arabidopsis thaliana*	Flavonol 3-*O*-galactoside [1,2]-glucosyltransferase	[10]
10	UGT79B30	BAR88078	*Glycine max*	Flavonol 3-*O*-glucoside/galactoside [1,2]-glucosyltransferase	[11]
11	UGT79B31	BBE29003	*Petunia hybrida*	Flavonol 3-*O*-glucoside/galactoside [1,2]-glucosyltransferase	[12]
12	MaF3GRT	ALD83609	*Morus alba*	Flavonol 3-*O*-glucoside [1,6]-rhamnosyltransferase	[13]
13	GmUGT79A6	BAN91401	*Glycine max*	Flavonol 3-*O*-glucoside/galactoside [1,6]-rhamnosyltransferase	[14]
14	GmUGT79A7	BAV56172	*Glycine max*	Flavonol 3-*O*-glucoside/galactoside [1,6]-glucosyltransferase	[15]
15	CaUGT3	BAH80312	*Catharanthus roseus*	Quercetin 3-*O*-glucoside [1,6]-glucosyltransferase	[16]
16	VpUGT94F1	BAI44133	*Veronica persica*	Flavonoid 3-*O*-glucoside [1,2]-glucosyltransferase	[17]
17	Cm1-2RhaT1	AAL06646	*Citrus maxima*	Flavonoid 7-*O*-glucoside [1,2]-rhamnosyltransferase	[18]
18	Cs1-6RhaT	ABA18631	*Citrus sinensis*	Flavonoid 7-*O*/3-*O*-glucoside [1,6]-rhamnosyltransferase	[18]
19	UGT91L1	NP_001347041	*Zea mays*	Iso-orientin [1,2]-rhamnosyltransferase	[19]
20	TcOGT4	MK947398	*Trollius chinensis*	Orientin [1,2]-galactosyltransferase	[20]
21	GtUF6CGT	AB985754	*Gentiana triflora*	Flavone 6-*C*-glycosyltransferase	[21]
22	PlUGT43	A0A172J2G3	*Pueraria lobata*	Isoflavone 8-*C*-glycosyltransferase	[22]
23	TcCGT1	MK644229	*Trollius chinensis*	Flavone 8-*C*-glycosyltransferase	[23]

24	WjGT1	LC465149	*Eutrema japonicum*	Flavone 6-*C*-glycosyltransferase	[24]
25	AbCGT	MN747045	*Aloe barbadensis*	2-Hydroxyflavanone *C*-glycosyltransferase	[25]
26	CuCGT	LC131334	*Citrus unshiu*	2-Hydroxyflavanone *C*-glycosyltransferase	[26]
27	DcaCGT	MT452646	*Dendrobium catenatum*	2-Hydroxyflavanone *C*-glycosyltransferase	[27]
28	FcCGT	LC131333	*Fortunella crassifolia*	2-Hydroxyflavanone *C*-glycosyltransferase	[26]
29	FeCGTa	AB909375	*Fagopyrum esculentum*	2-Hydroxyflavanone *C*-glycosyltransferase	[28]
30	FeCGTb	AB909376	*Fagopyrum esculentum*	2-Hydroxyflavanone *C*-glycosyltransferase	[28]
31	GgCGT	MH998596	*Glycyrrhiza glabra*	2-Hydroxyflavanone *C*-glycosyltransferase	[29]
32	MiCGTa	KT200208	*Mangifera indica*	Benzophenone *C*-glycosyltransferase	[30]
33	MiCGTb	A0A140GC03	*Mangifera indica*	Benzophenone di-*C*-glycosyltransferase	[31]
34	NnCGT1	LOC104598527	*Nelumbo nucifera*	2-Hydroxyflavanone *C*-glycosyltransferase	[32]
35	NnCGT2	LOC104603347	*Nelumbo nucifera*	2-Hydroxyflavanone *C*-glycosyltransferase	[32]
36	OsCGT	C3W7B0	*Oryza sativa*	2-Hydroxyflavanone *C*-glycosyltransferase	[33]
37	UGT708D1	LC003312	*Glycine max*	2-Hydroxyflavanone *C*-glycosyltransferase	[34]
38	ZmCGT	A0A096SRM5	*Zea mays*	2-Hydroxyflavanone *C*-glycosyltransferase	[35]
39	SbCGTa	MK894443	*Scutellaria baicalensis*	2-Hydroxyflavanone *C*-glycosyltransferase	[36]
40	SbCGTb	MK894444	*Scutellaria baicalensis*	2-Hydroxyflavanone *C*-glycosyltransferase	[36]
41	PhCGT1	MK616588	*Phyllostachys heterocycla*	2-Hydroxyflavanone *C*-glycosyltransferase	[37]
42	PhCGT2	MK616589	*Phyllostachys heterocycla*	2-Hydroxyflavanone *C*-glycosyltransferase	[37]

Table S2. PCR primers used in this study

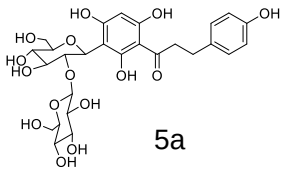
PCR primers to amplify genes from cDNA

Primer name	Sequence (5'-3')
ZjCGT1-F	GGACAGCAAATGGGTCGCCGGATGTTAAACCGCAAAGAAAATTC
ZjCGT1-R	GTCGACGGAGCTCGAATTCGGATTTTTCTGCTGCTTTAGGAATTC
ZjCGT2-F	GGACAGCAAATGGGTCGCCGGATGTCAGGTGACCTGAAGTTAG
ZjCGT2-R	GTCGACGGAGCTCGAATTCGGAAAGGTCTTCCAAGTCTCG
ZjOGT38-F	GGACAGCAAATGGGTCGCCGGATGGCCAATTTCCACATAGCC
ZjOGT38-R	GTCGACGGAGCTCGAATTCGGATTTTGGTTTACAAGCTTATTC

PCR primers to construct pCZ combined vectors

Name	Primers	Sequence (5'→3')
pCZ88	pCZ88-F	CCTAAAGCAGCAGAATAATTAAACTAGAAATAATTTTGTTTAACT TTAAGAAGGAGATATACATATGGCCAATTTCCACATAGC
	pCZ88-R	ATCTCAGTGGTGGTGGTGGTGGTGCTCGAGTGCGGCCGCAAGC TTGTTTAATTTTGGTTTACAAGCTTATTCAGA
pCZ89	pCZ89-F	ATCTTGCAGTGTTGGACTAGTAAACTAGAAATAATTTTGTTTAAC TTTAAGAAGGAGATATACATATGGCCAATTTCCACATAGC
	pCZ89-R	GTGGTGGTGGTGGTGCTCGAGTAAGCTTGCGGGATCCGCCGCTT AATTTTGGTTTACAAGCTTATTCAGATC

Table S3. ^{1}H and ^{13}C NMR data (600 MHz, δ in ppm and J in Hz) of 5a [29]



5a

Position	5a	
	δ_C	$\delta_{H,\ mult(Hz)}$
1	132.6	
2	129.1	7.01 d (J = 8.4 Hz)
3	115.1	6.66 d (J = 8.4 Hz)
4	155.4	
5	115.1	6.66 d (J = 8.4 Hz)
6	129.1	7.01 d (J = 8.4 Hz)
C=O	204.5	
α	45.6	3.20 m
β	29.7	2.76 m
1'	105.3	
2'	164.4	
3'	103.3	
4'	165.4	
5'	94.5	5.93 m
6'	161.7	
1"	71.8	4.56 m
2"	81.2	4.26 m
3"	74.8	2.87 t (J = 8.1 Hz)
4"	76.5	3.06 m
5"	81.4	3.11 m
6"	61.4	3.65 m
1"'	105.4	4.14 d (J = 7.4 Hz)
2"'	70.5	3.11 m
3"'	76.4	2.66 m
4"'	69.4	3.05 m
5"'	78.6	3.38 m
6"'	61.4	3.37 m

Table S4. Plasmids and strains used in this study

Designation	Relevant characteristics	Source/references
Plasmids		
pYH55	pCDF-T7-4CL-T7-PAL-T7-CHS-T7-CHI	[38]
pCZ201	pET-Duet-T7-2B1-tr29 Syn_ZmCYP93G5-T7-AtCPR2	[37]
pCZ88	pET-28a-T7-ZjCGT1-ZjOGT38	This work
pCZ89	pET-28a-T7- PhUGT708A43-ZjOGT38	This work
Strains		
E. coli BL21(DE3)		NEB
sCZ112	BL21(DE3): pYH55+pCZ201	[37]
sCZ121	BL21(DE3): pYH55+pCZ201+pCZ88	This work
sCZ122	BL21(DE3): pYH55+pCZ201+pCZ89	This work

3. Supplementary Figures

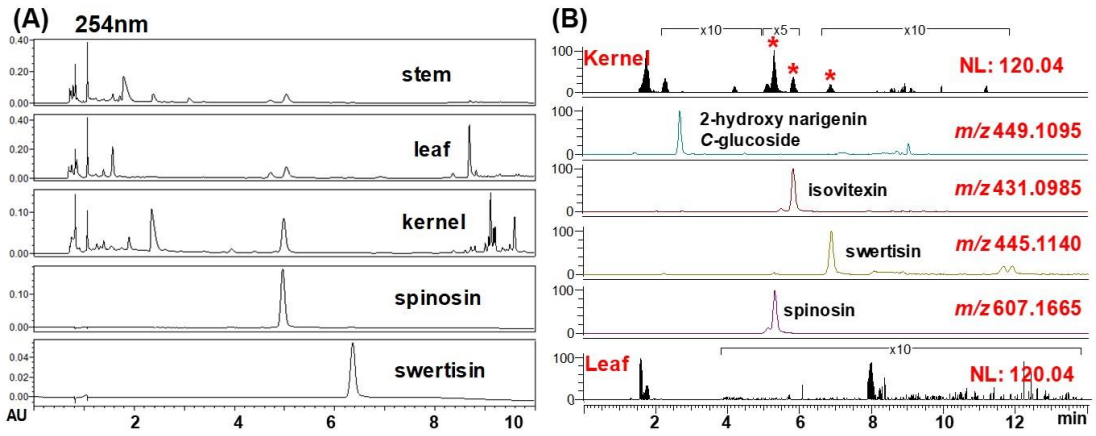


Fig. S1. Chemical analysis of different parts of *Ziziphus jujuba*. (A) The UPLC chromatograms of stem, leaf, kernel, and the reference standards. (B) Extracted ion chromatograms. NL, neutral loss.

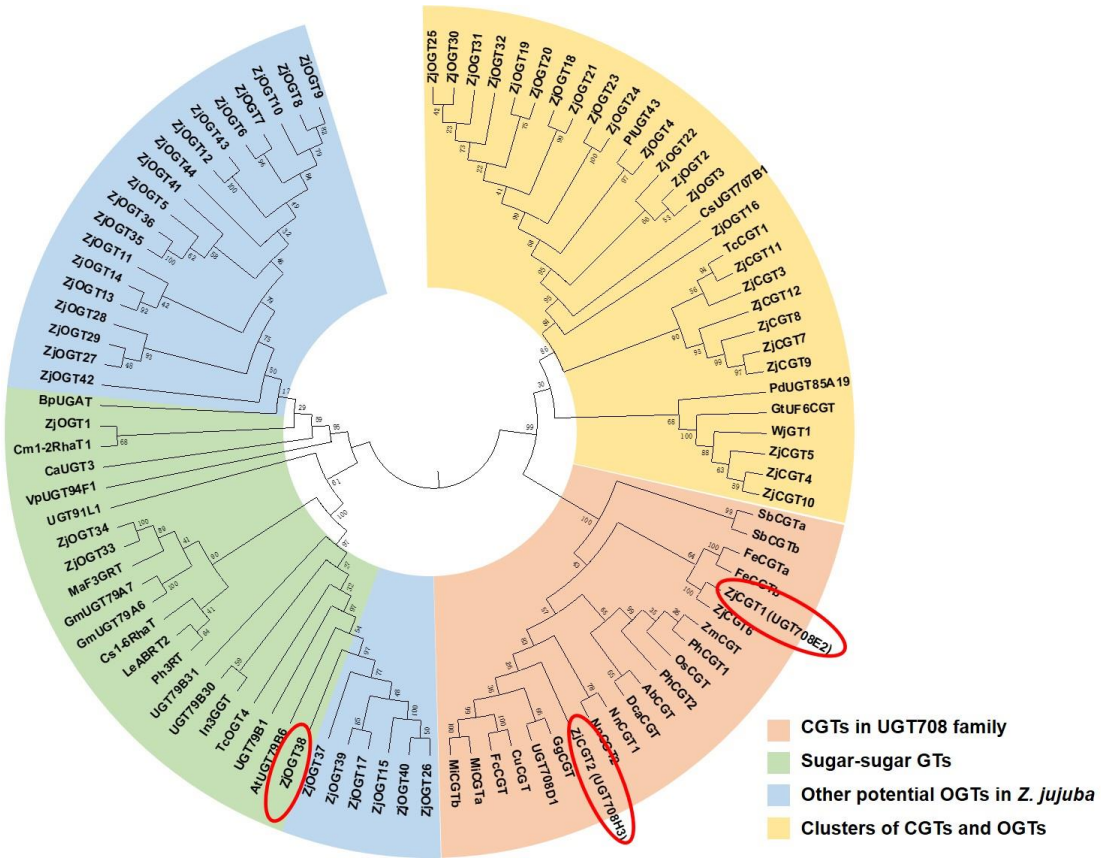


Fig. S2. The phylogenetic relationships between reported OGTs, CGTs and candidate genes from *Z. jujuba.* The protein sequences were aligned using ClustalW. The neighbour-joining phylogenetic tree was drawn using MEGA 6. The bootstrap value was 1000, and the branch lengths represent the relative genetic distances. The abbreviations of the protein sequences and their accession numbers are listed in Table S1.

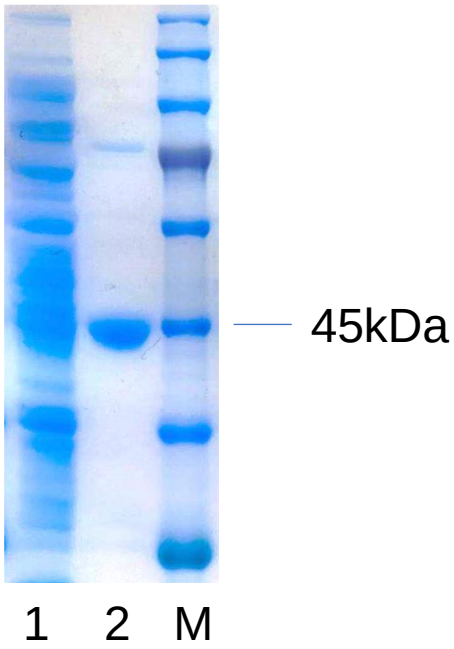


Fig. S3. SDS-PAGE of the His-tagged ZjOGT38 purified by affinity chromatography. Lane 1: Crude enzymes. Lane 2: Purified ZjOGT38. Lane M: Protein marker.

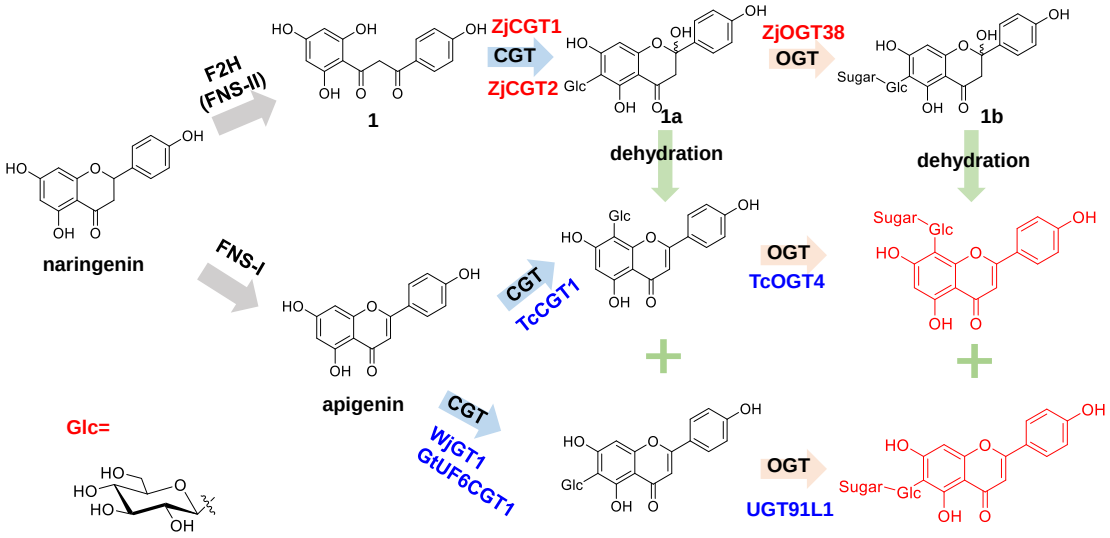


Fig. S4. Proposed biosynthetic pathway of flavonoid *O*-glycosyl-*C*-glycosides based on previous reports. FNS: flavone synthase; F2H: flavanone 2-hydroxylase. Information for the enzymes and their accession numbers are listed in Table S1.

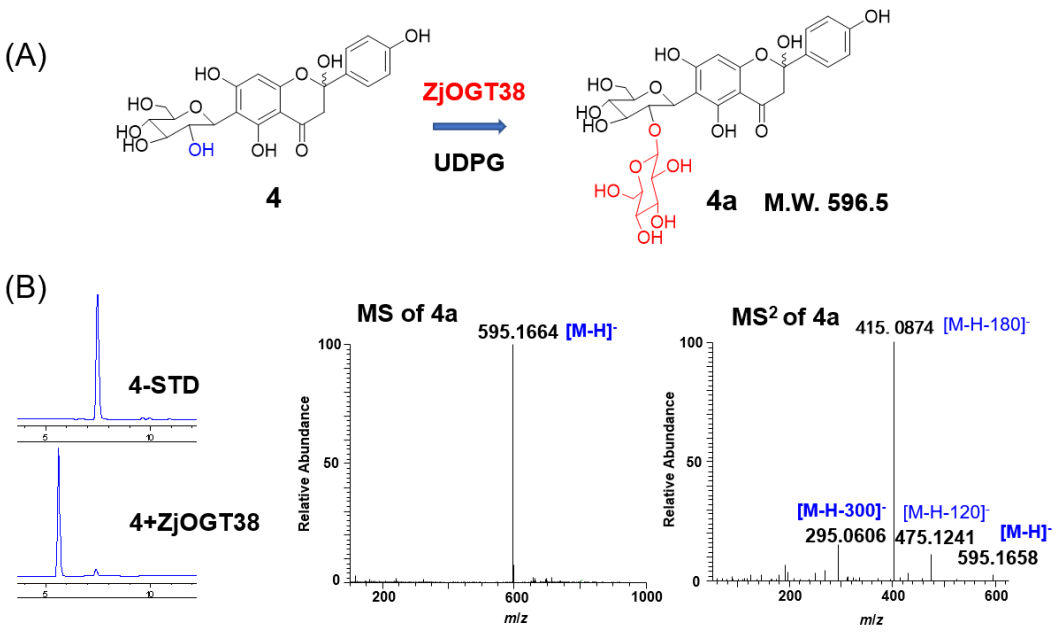


Fig. S5. HPLC and MS analyses of ZjOGT38 catalytic reaction mixtures for compound **4** (*C*-glycosyl 2-hydroxypinocembrin) with UDP-Glc. (A) The enzymatic reaction. (B) HPLC chromatograms and (-)-ESI-MS spectra for product **4a**. UV chromatogram recorded at 300nm. The analytical conditions are given under the experimental procedures.

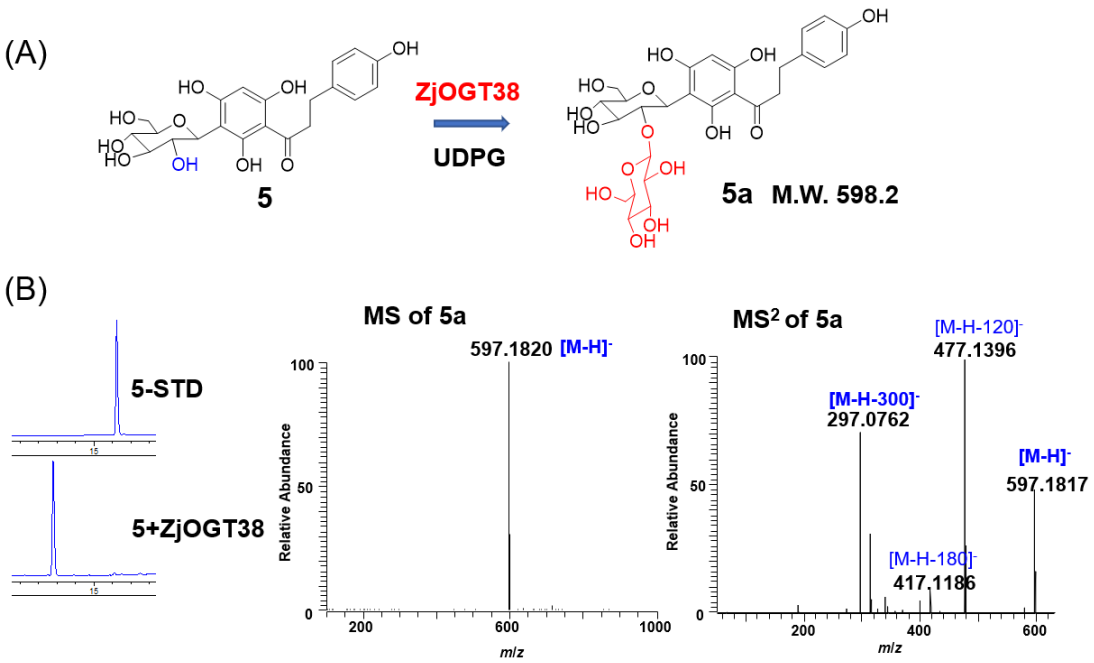


Fig. S6. HPLC and MS analyses of ZjOGT38 catalytic reaction mixtures for compound **5** (nothofagin) with UDP-Glc. (A) The enzymatic reaction. (B) HPLC chromatograms and (-)-ESI-MS spectra for product **5a**. UV chromatogram recorded at 300nm. The analytical conditions are given under the experimental procedures.

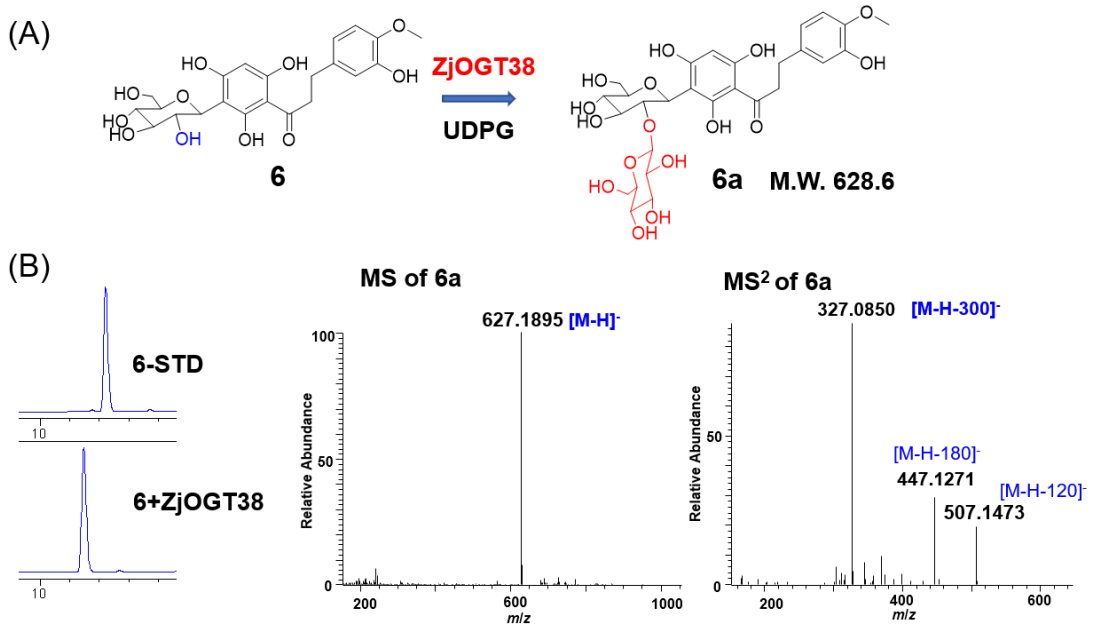


Fig. S7. HPLC and MS analyses of ZjOGT38 catalytic reaction mixtures for compound **6** (*C*-glycosyl hesperetin dihydrochalcone) with UDP-Glc. (A) The enzymatic reaction. (B) HPLC chromatograms and (-)-ESI-MS spectra for product **6a**. UV chromatogram recorded at 300nm. The analytical conditions are given under the experimental procedures.

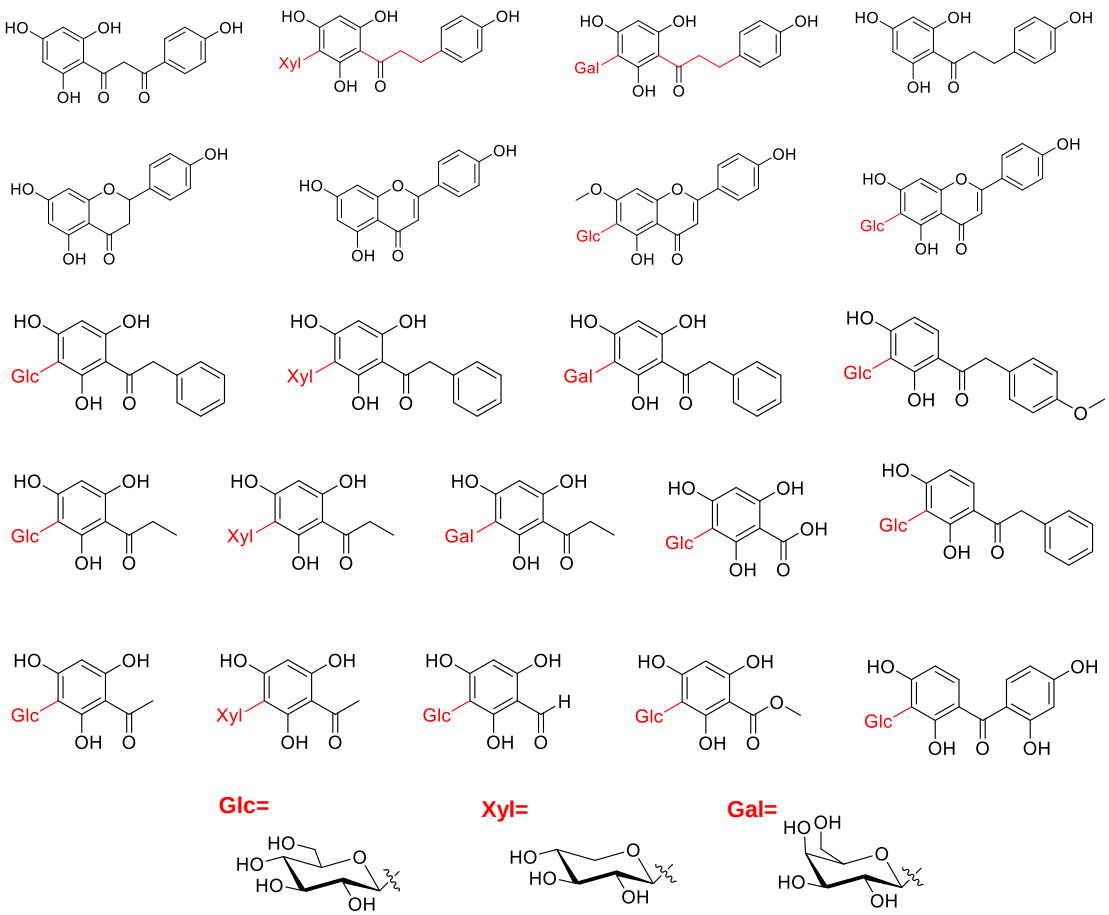


Fig. S8. Substrates that cannot be recognized by ZjOGT38.

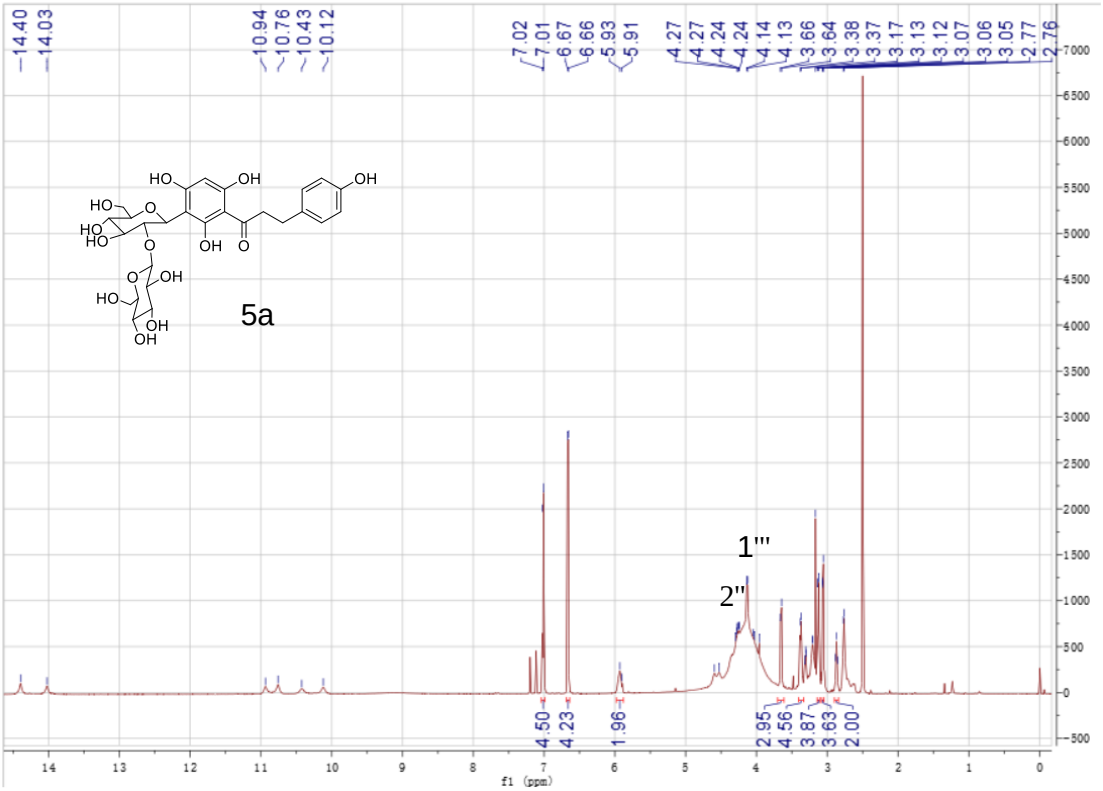


Fig. S9. ^{1}H NMR (600 MHz, DMSO-d_6) spectrum of **5a**.

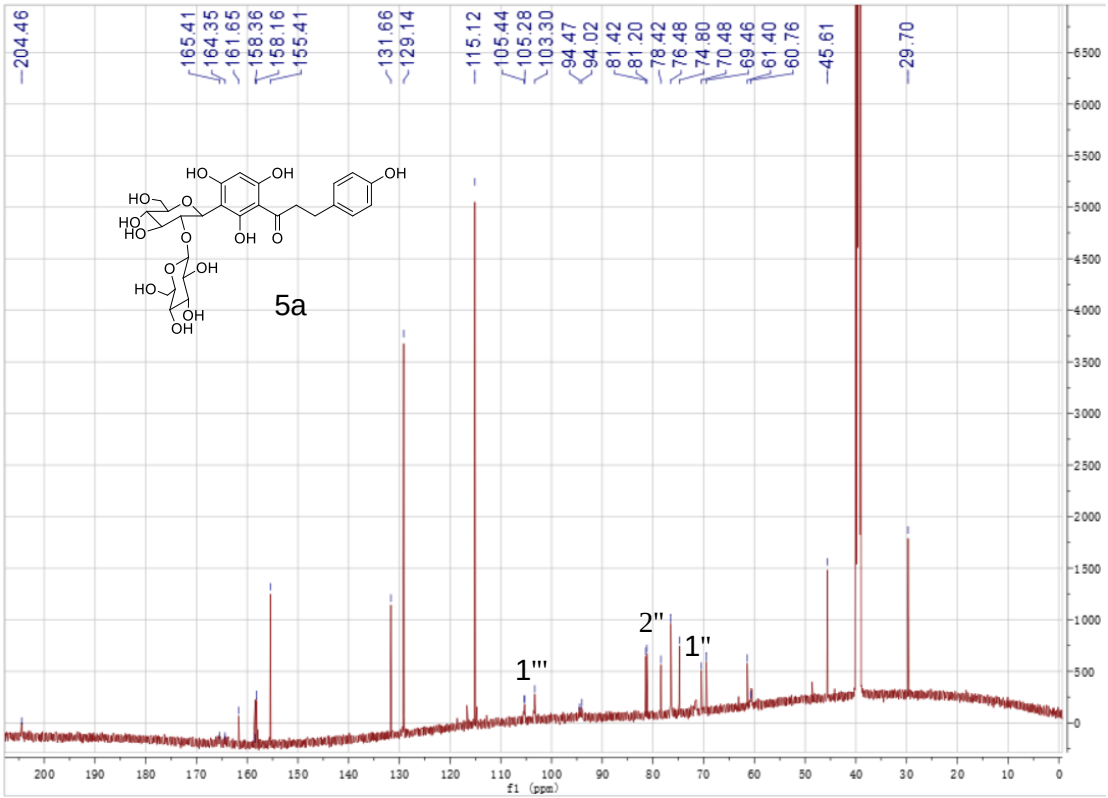


Fig. S10. ^{13}C NMR (150 MHz, DMSO-d_6) spectrum of **5a**.

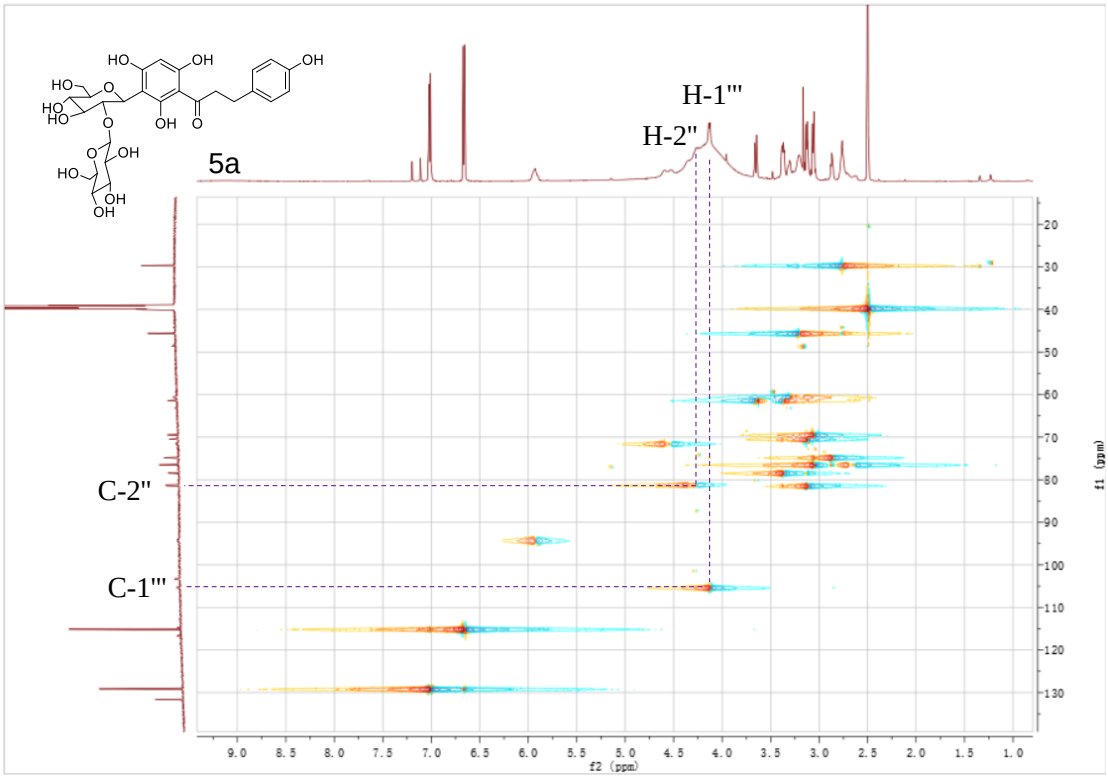


Fig. S11. HSQC (600 MHz, DMSO-d_6) spectrum of **5a**.

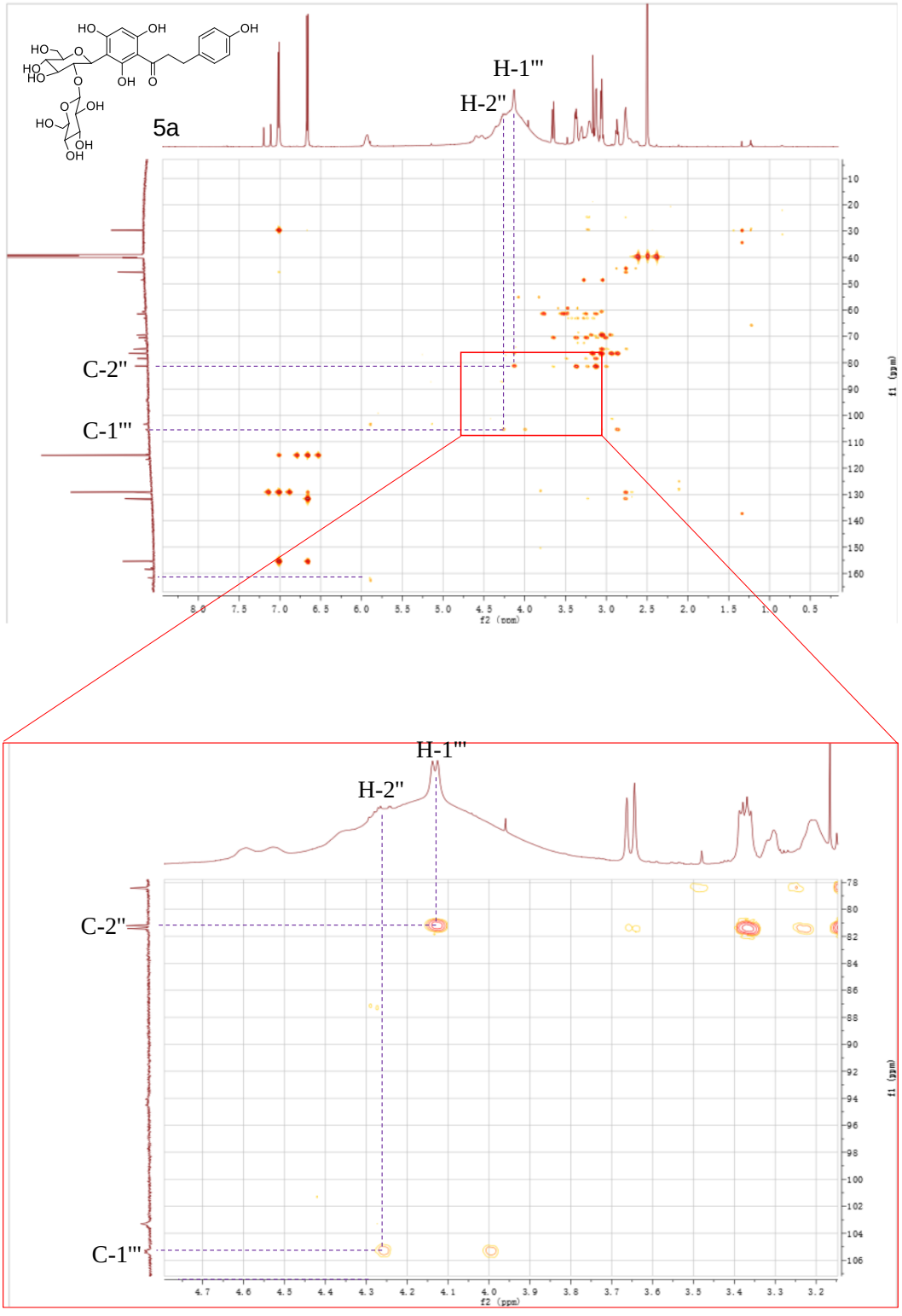


Fig. S12. HMBC (600 MHz, DMSO-d_6) spectrum of **5a**.

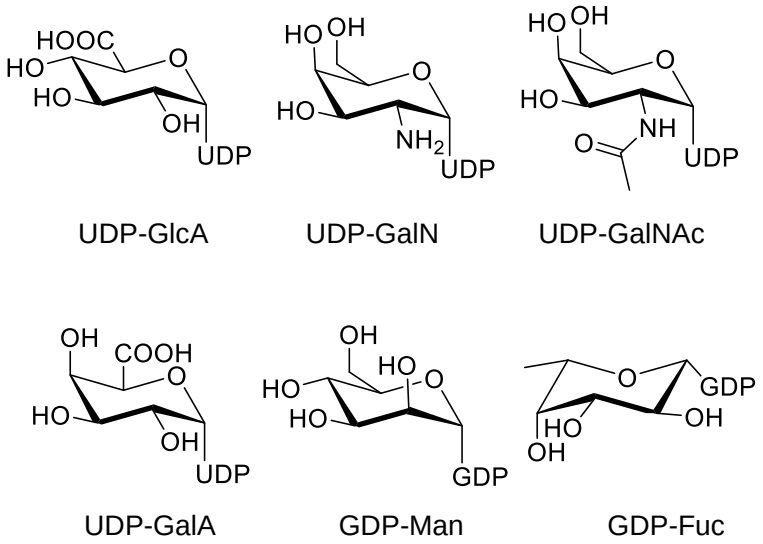


Fig. S13. Sugar donors that cannot be recognized by ZjOGT38.

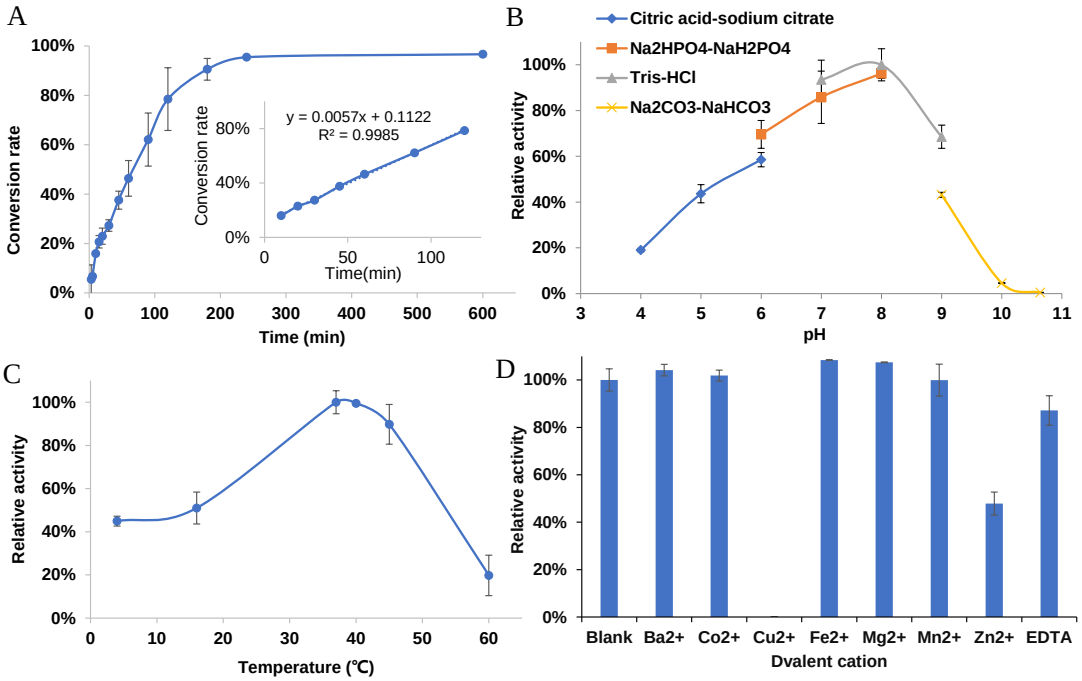
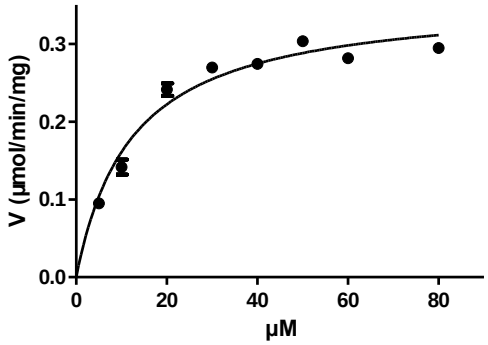


Fig. S14. Effects of reaction time (A), reaction buffer (B), temperature (C), and divalent metal ions (D) on glycosylation activity of ZjOGT38. Nothofagin was used as the sugar acceptor and UDP-Glc was used as the sugar donor. An optimized reaction time of 60 min was used.



K_m=12.34±1.479 μM

V_{max}=0.3593±0.01213 μM min^{-1} mg^{-1}

k_{cat}=0.3054 s^{-1}

k_{cat}/K_m=0.0247 $\mu M^{-1}s^{-1}$

Fig. S15. Determination of kinetic parameters for recombinant ZjOGT38. The apparent K_m value was determined using nothofagin as the acceptor and UDP-Glc as the sugar donor at 37ºC and pH 8.0 for 10 min.

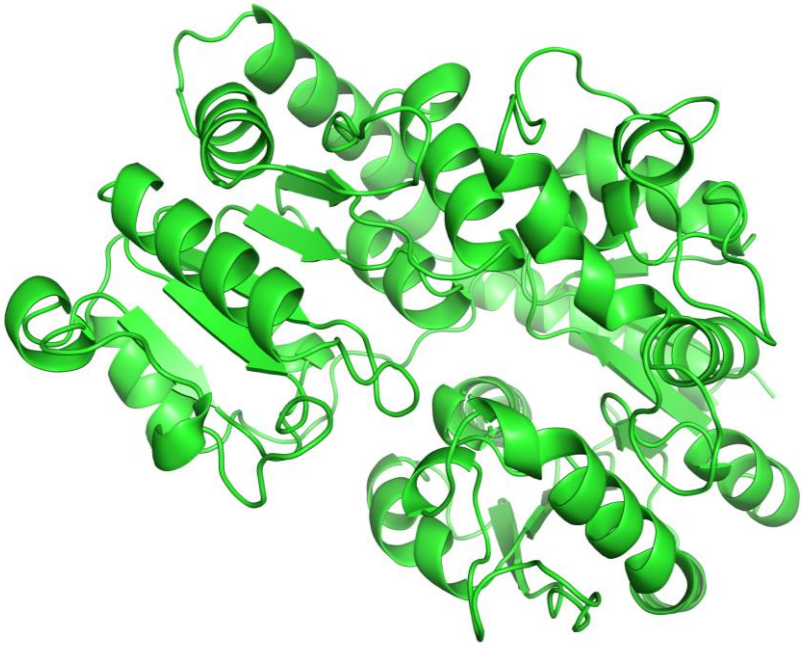


Fig. S16. ZjOGT38 protein model calculated by Alphafold2.

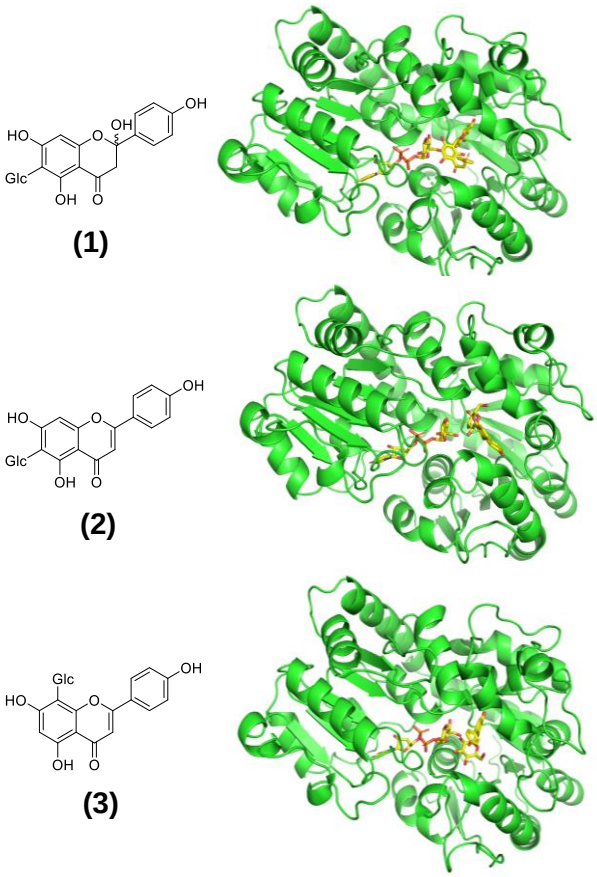
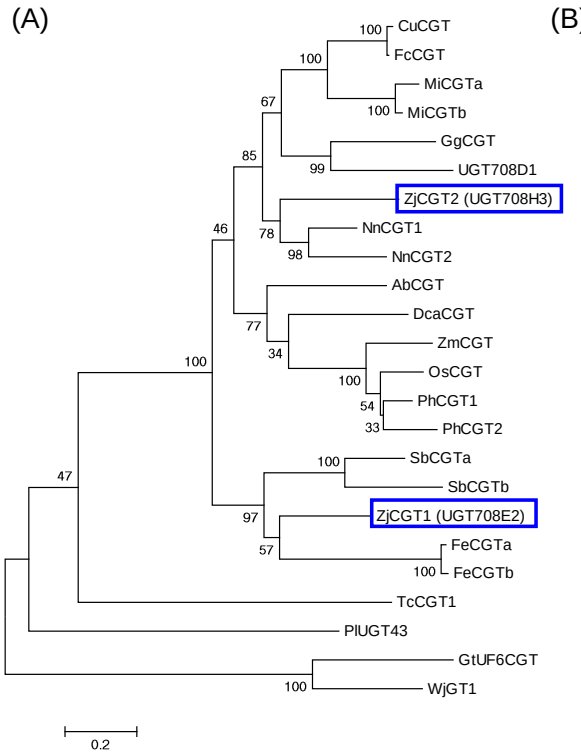


Fig. S17. Molecular docking results of ZjOGT38/UDP-Glc/**1**, ZjOGT38/UDP-Glc/**2** and ZjOGT38/UDP-Glc/**3**.



(B)

CuCGT	84	A	T	D	P	F	F	L	R	W	E	95
FcCGT	84	A	T	D	P	F	F	L	R	W	E	95
MiCGTa	84	T	T	D	P	F	F	L	R	W	E	95
MiCGTb	84	S	T	D	P	F	F	L	R	W	E	95
GgCGT	86	S	T	D	P	F	F	L	Q	F	A	97
UGT708D1	82	T	I	D	P	F	F	L	Q	F	E	93
NnCGT1	87	S	K	D	P	F	F	L	Q	F	E	98
NnCGT2	80	S	N	D	P	F	C	L	Q	F	E	91
AbCGT	84	S	P	D	P	F	F	Q	Q	W	E	95
DcaCGT	79	S	S	D	P	F	F	L	Q	M	E	90
ZmCGT	88	G	A	D	P	F	F	L	R	F	E	99
OsCGT	86	G	A	D	P	F	F	L	R	F	E	97
PhCGT1	86	G	A	D	P	F	F	L	R	F	E	97
PhCGT2	91	D	A	D	P	F	Y	V	R	Y	E	102
SbCGTa	85	N	D	D	P	F	F	I	Q	M	E	96
SbCGTb	85	N	D	D	P	F	F	I	Q	R	E	96
FeCGTa	93	N	V	D	P	F	F	L	R	Y	K	104
FeCGTb	93	N	V	D	P	F	F	L	R	Y	K	104
GuCGTa	86	S	T	D	P	F	F	L	Q	F	A	97
GuCGTb	87	T	N	D	P	F	W	L	H	F	E	98
ZjCGT1	83	T	D	D	P	F	F	L	Q	E	E	94
ZjCGT2	82	T	E	D	P	F	Y	Y	H	F	E	93

Fig. S18. Phylogenetic analysis. (A) A phylogenetic tree of plant CGTs; (B) the DPF conserved region of plant CGTs in the UGT708 family.

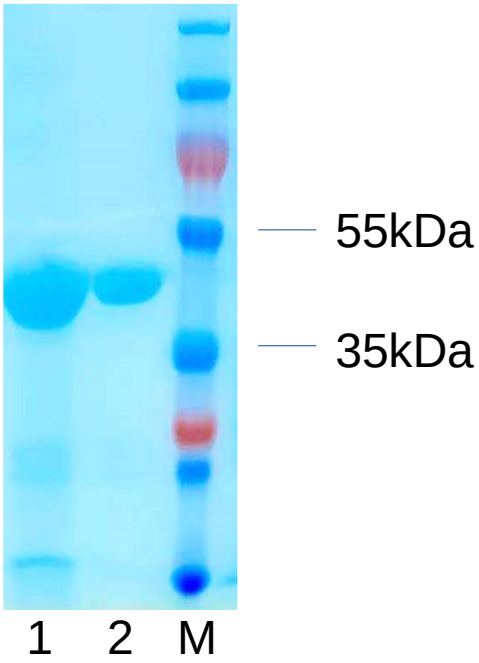


Fig. S19. SDS-PAGE of the His-tagged ZjCGT1 and ZjCGT2 purified by affinity chromatography. Lane 1: Purified ZjCGT1. Lane 2: Purified ZjCGT2. Lane M: Protein marker.

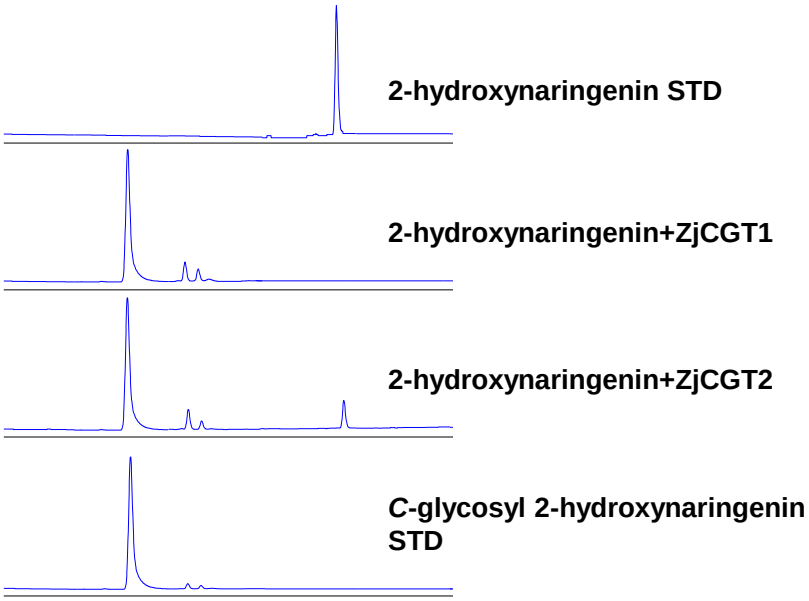


Fig. S20. The UV chromatograms of ZjCGTs reactions recorded at 300 nm. ZjCGT1 and ZjCGT2 exhibited *C*-glycosylation activity towards 2-hydroxynaringenin, and the product was identified as *C*-glycosyl 2-hydroxynaringenin (**1**) by comparing with a reference standard. The analytical conditions are given under experimental procedures.

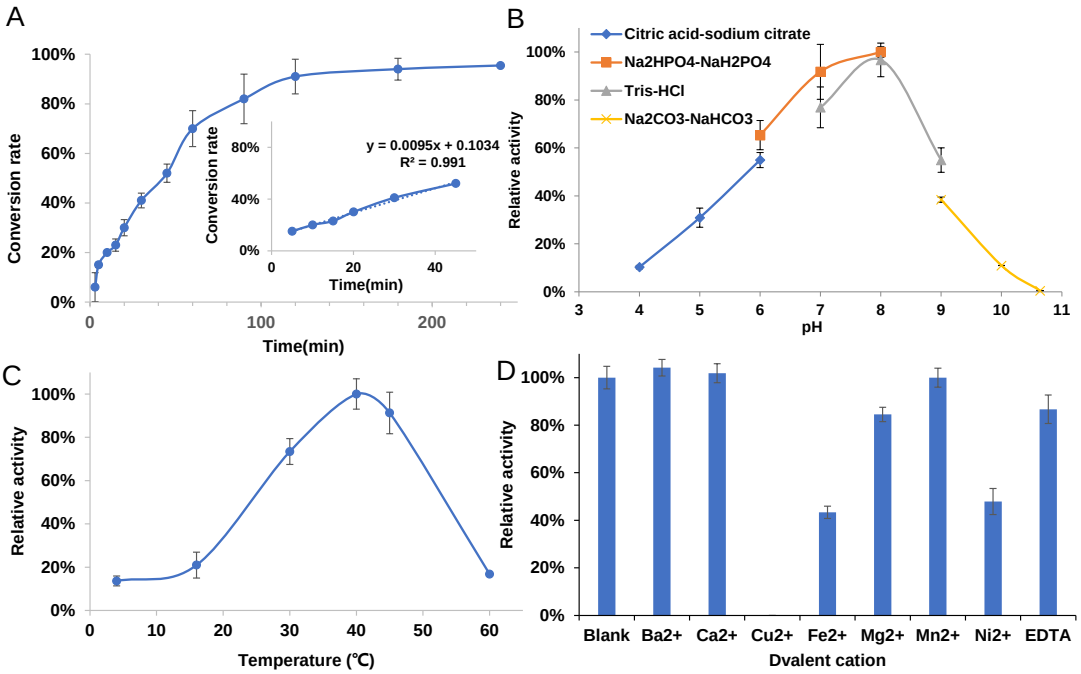


Fig. S21. Effects of reaction time (A), reaction buffer (B), temperature (C), and divalent metal ions (D) on glycosylation activity of ZjCGT1. Phloretin was used as the acceptor and UDP-Glc was used as the sugar donor. An optimized reaction time of 90 min was used. ZjCGT1 exhibited its maximum activity at pH 8.0 (50 mM PBS buffer) and 40ºC, and was independent of divalent cations.

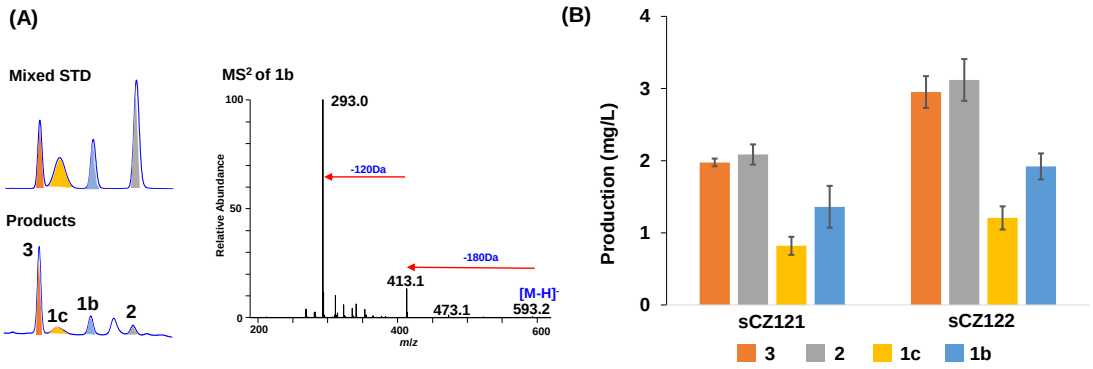


Fig. S22. Chemical analysis of the de novo biosynthetic products. (A) HPLC chromatogram of the fermentation mixture, and mass spectrum of product **1b**. (B) Production of four glycosides in sCZ121 (pCZ88 as GT module), and sCZ122 (pCZ89 as GT module) after fermentation for 96 h.

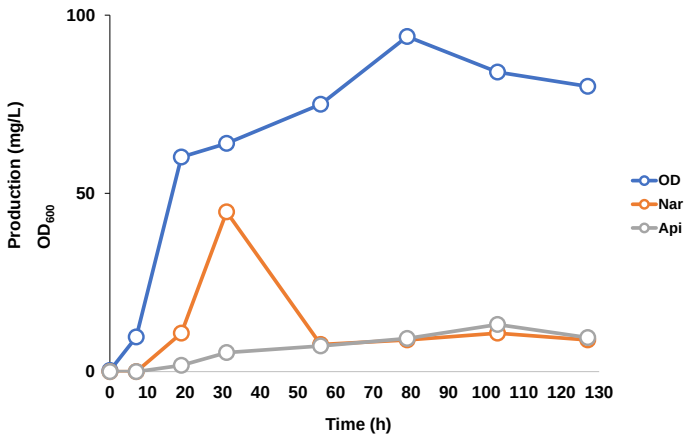


Fig. S23. Large-scale production of apigenin (Api) and naringenin (Nar) in a 5-L bioreactor. OD, optical density.

References

1. Trott, O.; Olson, A. J., Software news and update AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2012**, *31*, 455-461.
2. Yonekura Sakakibara, K.; Fukushima, A.; Nakabayashi, R.; Hanada, K.; Matsuda, F.; Sugawara, S.; Inoue, E.; Kuromori, T.; Ito, T.; Shinozaki, K.; Wangwattana, B.; Yamazaki, M.; Saito, K., Two glycosyltransferases involved in anthocyanin modification delineated by transcriptome independent component analysis in *Arabidopsis thaliana*. *Plant J.* **2012**, *69*, 154-167.
3. Montefiori, M.; Espley, R. V.; Stevenson, D.; Cooney, J.; Datson, P. M.; Saiz, A.; Atkinson, R. G.; Hellens, R. P.; Allan, A. C., Identification and characterization of F3GT1 and F3GGT1, two glycosyltransferases responsible for anthocyanin biosynthesis in red-fleshed kiwifruit (*Actinidia chinensis*). *Plant J.* **2011**, *65*, 106-118.
4. Morita, Y.; Hoshino, A.; Kikuchi, Y.; Okuhara, H.; Ono, E.; Tanaka, Y.; Fukui, Y.; Saito, N.; Nitasaka, E.; Noguchi, H.; Iida, S., 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. *Plant J.* **2005**, *42*, 353-363.
5. Sawada, S.; Suzuki, H.; Ichimaida, F.; Yamaguchi, M.; Iwashita, T.; Fukui, Y.; Hemmi, H.; Nishino, T.; Nakayama, T., UDP-glucuronic Acid: Anthocyanin Glucuronosyltransferase from Red Daisy (*Bellis perennis*) Flowers. *J. Biol. Chem.* **2005**, *280*, 899-906.
6. Brugliera, F.; Holton, T. A.; Stevenson, T. W.; Farcy, E.; Lu, C. Y.; Cornish, E. C., Isolation and characterization of a cDNA clone corresponding to the Rt locus of *Petunia hybrida*. *Plant J.* **1994**, *5*, 81-92.
7. Hsu, Y.; Tagami, T.; Matsunaga, K.; Okuyama, M.; Suzuki, T.; Noda, N.; Suzuki, M.; Shimura, H., Functional characterization of UDP-rhamnose-dependent rhamnosyltransferase involved in anthocyanin modification, a key enzyme determining blue coloration in *Lobelia erinus*. *Plant J.* **2017**, *89*, 325-337.
8. Franks, T. K.; Yadollahi, A.; Wirthensohn, M. G.; Guerin, J. R.; Kaiser, B. N.; Sedgley, M.; Ford, C. M., A seed coat cyanohydrin glucosyltransferase is associated with bitterness in almond (*Prunus dulcis*) kernels. *Funct. Plant Biol.* **2008**, *35*, 236.
9. Trapero, A.; Ahrazem, O.; Rubio-Moraga, A.; Jimeno, M. L.; Gómez, M. D.; Gómez-Gómez, L., Characterization of a glucosyltransferase enzyme involved in the formation of kaempferol and quercetin sophorosides in *Crocus sativus*. *Plant Physiol.* **2012**, *159*, 1335-1354.
10. Yonekura Sakakibara, K.; Nakabayashi, R.; Sugawara, S.; Tohge, T.; Ito, T.; Koyanagi, M.; Kitajima, M.; Takayama, H.; Saito, K., A flavonoid 3−*O*−glucoside: 2″−*O*−glucosyltransferase responsible for terminal modification of pollen−specific flavonols in *Arabidopsis thaliana*. *Plant J.* **2014**, *79*, 769-782.
11. Di, S.; Yan, F.; Rodas, F. R.; Rodriguez, T. O.; Murai, Y.; Iwashina, T.; Sugawara, S.; Mori, T.; Nakabayashi, R.; Yonekura-Sakakibara, K.; Saito, K.; Takahashi, R., Linkage mapping, molecular cloning and functional analysis of soybean gene Fg3 encoding flavonol 3-*O*-glucoside/galactoside (1 → 2) glucosyltransferase. *BMC Plant Biol.* **2015**, *15*, 126.
12. Knoch, E.; Sugawara, S.; Mori, T.; Nakabayashi, R.; Saito, K.; Yonekura-Sakakibara, K., UGT79B31 is responsible for the final modification step of pollen-specific flavonoid biosynthesis in *Petunia hybrida*. *Planta* **2018**, *247*, 779-790.

13. Li, D.; Chen, G.; Ma, B.; Zhong, C.; He, N., Metabolic profiling and transcriptome analysis of mulberry leaves provide insights into flavonoid biosynthesis. *J. Agric. Food Chem.* **2020**, *68*, 1494-1504.
14. Rojas Rodas, F.; Rodriguez, T. O.; Murai, Y.; Iwashina, T.; Sugawara, S.; Suzuki, M.; Nakabayashi, R.; Yonekura-Sakakibara, K.; Saito, K.; Kitajima, J.; Toda, K.; Takahashi, R., Linkage mapping, molecular cloning and functional analysis of soybean gene Fg2 encoding flavonol 3-*O*-glucoside (1→6) rhamnosyltransferase. *Plant Mol. Biol.* **2014**, *84*, 287-300.
15. Rojas Rodas, F.; Di, S.; Murai, Y.; Iwashina, T.; Sugawara, S.; Mori, T.; Nakabayashi, R.; Yonekura-Sakakibara, K.; Saito, K.; Takahashi, R., Cloning and characterization of soybean gene Fg1 encoding flavonol 3-*O*-glucoside/galactoside (1→6) glucosyltransferase. *Plant Mol. Biol.* **2016**, *92*, 445-456.
16. Masada, S.; Terasaka, K.; Oguchi, Y.; Okazaki, S.; Mizushima, T.; Mizukami, H., Functional and Structural Characterization of a Flavonoid Glucoside 1,6-Glucosyltransferase from *Catharanthus roseus*. *Plant Cell Physiol.* **2009**, *50*, 1401-1415.
17. Ono, E.; Ruike, M.; Iwashita, T.; Nomoto, K.; Fukui, Y., Co-pigmentation and flavonoid glycosyltransferases in blue *Veronica persica* flowers. *Phytochemistry* **2010**, *71*, 726-735.
18. Frydman, A.; Liberman, R.; Huhman, D. V.; Carmeli-Weissberg, M.; Sapir-Mir, M.; Ophir, R.; W. Sumner, L.; Eyal, Y., The molecular and enzymatic basis of bitter/non-bitter flavor of citrus fruit: evolution of branch-forming rhamnosyltransferases under domestication. *Plant J.* **2013**, *73*, 166-178.
19. Casas, M. I.; Falcone-Ferreyra, M. L.; Jiang, N.; Mejía-Guerra, M. K.; Rodríguez, E.; Wilson, T.; Engelmeier, J.; Casati, P.; Grotewold, E., Identification and characterization of maizesalmon silks genes involved in insecticidal maysin Biosynthesis. *The Plant Cell* **2016**, *28*, 1297-1309.
20. Liu, S.; Zhang, M.; Bao, Y. O. J.; Chen, K.; Xu, L. L.; Su, H. F.; Kuang, Y.; Wang, Z. L.; Qiao, X.; Ye, M., Characterization of a highly selective 2″-*O*-galactosyltransferase from *Trollius chinensis* and structure-guided engineering for improving UDP-glucose selectivity. *Org. Lett.* **2021**, *23*, 9020-9024.
21. Sasaki, N.; Nishizaki, Y.; Yamada, E.; Tatsuzawa, F.; Nakatsuka, T.; Takahashi, H.; Nishihara, M., Identification of the glucosyltransferase that mediates direct flavone *C*-glucosylation in *Gentiana triflora*. *FEBS Lett.* **2015**, *589*, 182-187.
22. Wang, X.; Li, C.; Zhou, C.; Li, J.; Zhang, Y., Molecular characterization of the *C*-glucosylation for puerarin biosynthesis in *Pueraria lobata*. *Plant J.* **2017**, *90*, 535-546.
23. He, J. B.; Zhao, P.; Hu, Z. M.; Liu, S.; Kuang, Y.; Zhang, M.; Li, B.; Yun, C. H.; Qiao, X.; Ye, M., Molecular and structural characterization of a promiscuous *C*-glycosyltransferase from *Trollius chinensis*. *Angew. Chem., Int. Ed.* **2019,** *58*, 11513-11520.
24. Mashima, K.; Hatano, M.; Suzuki, H.; Shimosaka, M.; Taguchi, G., Identification and characterization of apigenin 6-*C*-glucosyltransferase involved in biosynthesis of isosaponarin in Wasabi (*Eutrema japonicum*). *Plant Cell Physiol.* **2019**, *60*, 2733-2743.
25. Xie, K.; Zhang, X.; Sui, S.; Ye, F.; Dai, J., Exploring and applying the substrate promiscuity of a *C*-glycosyltransferase in the chemo-enzymatic synthesis of bioactive *C*-glycosides. *Nat. Commun.* **2020**, *11*, 5162.
26. Ito, T.; Fujimoto, S.; Suito, F.; Shimosaka, M.; Taguchi, G., *C*-Glycosyltransferases catalyzing the formation of di-*C*-glucosyl flavonoids in citrus plants. *Plant J.* **2017**, *91*, 187-198.
27. Ren, Z.; Ji, X.; Jiao, Z.; Luo, Y.; Zhang, G.; Tao, S.; Lei, Z.; Zhang, J.; Wang, Y.; Liu, Z.; Wei, G., Functional analysis of a novel *C*-glycosyltransferase in the orchid *Dendrobium catenatum*. *Hortic.*

Res. **2020**, *7*, 111.

28. Nagatomo, Y.; Usui, S.; Ito, T.; Kato, A.; Shimosaka, M.; Taguchi, G., Purification, molecular cloning and functional characterization of flavonoid *C*-glucosyltransferases from *Fagopyrum esculentum* M. (buckwheat) cotyledon. *Plant J.* **2014**, *80*, 437-448.
29. Zhang, M.; Li, F.; Li, K.; Wang, Z.; Wang, Y.; He, J.; Su, H.; Zhang, Z.; Chi, C.; Shi, X.; Yun, C.; Zhang, Z.; Liu, Z.; Zhang, L.; Yang, D.; Ma, M.; Qiao, X.; Ye, M., Functional characterization and structural basis of an efficient di-*C*-glycosyltransferase from *Glycyrrhiza glabra*. *J. Am. Chem. Soc.* **2020**, *142*, 3506-3512.
30. Chen, D.; Chen, R.; Wang, R.; Li, J.; Xie, K.; Bian, C.; Sun, L.; Zhang, X.; Liu, J.; Yang, L.; Ye, F.; Yu, X.; Dai, J., Probing the catalytic promiscuity of a regio- and stereospecific *C*-glycosyltransferase from *Mangifera indica*. *Angew. Chem., Int. Ed.* **2015**, *54*, 12678-12682.
31. Chen, D.; Fan, S.; Chen, R.; Xie, K.; Yin, S.; Sun, L.; Liu, J.; Yang, L.; Kong, J.; Yang, Z.; Dai, J., Probing and engineering key residues for bis-*C*-glycosylation and promiscuity of a *C*-glycosyltransferase. *ACS Catal.* **2018**, *8*, 4917-4927.
32. Feng, C. Y.; Li, S. S.; Taguchi, G.; Wu, Q.; Yin, D. D.; Gu, Z. Y.; Wu, J.; Xu, W. Z.; Liu, C.; Wang, L. S., Enzymatic basis for stepwise *C*−glycosylation in the formation of flavonoid di−*C*−glycosides in sacred lotus (*Nelumbo nucifera* Gaertn.). *Plant J.* **2021**, *106*, 351-365.
33. Brazier-Hicks, M.; Evans, K. M.; Gershater, M. C.; Puschmann, H.; Steel, P. G.; Edwards, R., The *C*-glycosylation of flavonoids in cereals. *J. Biol. Chem.* **2009**, *284*, 17926-17934.
34. Hirade, Y.; Kotoku, N.; Terasaka, K.; Saijo-Hamano, Y.; Fukumoto, A.; Mizukami, H., Identification and functional analysis of 2-hydroxyflavanone *C*-glucosyltransferase in soybean (*Glycine max*). *FEBS Lett.* **2015**, *589*, 1778-1786.
35. Falcone Ferreyra, M. L.; Rodriguez, E.; Casas, M. I.; Labadie, G.; Grotewold, E.; Casati, P., Identification of a bifunctional maize *C*- and *O*-glucosyltransferase. *J. Biol. Chem.* **2013**, *288*, 31678-31688.
36. Wang, Z.; Gao, H.; Wang, S.; Zhang, M.; Chen, K.; Zhang, Y.; Wang, H.; Han, B.; Xu, L.; Song, T.; Yun, C.; Qiao, X.; Ye, M., Dissection of the general two-step di-*C*-glycosylation pathway for the biosynthesis of (iso)schaftosides in higher plants. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 30816-30823.
37. Sun, Y.; Chen, Z.; Yang, J.; Mutanda, I.; Li, S.; Zhang, Q.; Zhang, Y.; Zhang, Y.; Wang, Y., Pathway-specific enzymes from bamboo and crop leaves biosynthesize anti-nociceptive *C*-glycosylated flavones. *Commun. Biol.* **2020**, *3*, 110.
38. Li, J.; Tian, C.; Xia, Y.; Mutanda, I.; Wang, K.; Wang, Y., Production of plant-specific flavones baicalein and scutellarein in an engineered *E. coli* from available phenylalanine and tyrosine. *Metab. Eng.* **2019**, *52*, 124-133.