Electronic Supplementary Material (ESI) for Organic Chemistry Frontiers.

Information for

Discovery and Characterization of Four Glycosyltransferases Involved in Anthraquinone Glycoside Biosynthesis in *Rubia yunnanensis*

Shanyong Yi[#], Tongdong Kuang[#], Yuanyuan Miao, Yanqing Xu, Zhe Wang, Liao-Bin Dong, Ninghua Tan*

State Key Laboratory of Natural Medicines, Department of TCMs Pharmaceuticals, School of Traditional Chinese Pharmacy, China Pharmaceutical University, Nanjing 211198, Jiangsu, China.

*Corresponding authors. E-mail addresses: nhtan@cpu.edu.cn (N. T.).

[#]S. Y. and T. K. contributed equally to this work.

Table of Contents

1. Experimental Procedures

1.1 General remarks

6-Hydroxyalizarin (**1**), xanthopurpurin (**7**), 1-hydroxy-2-hydroxymethyl-9,10-anthraquinone (**10**), 1-hydroxy-2-hydroxymethyl-9,10-anthraquinone 2-CH_2-*O*-*β*-D-glucopyranoside (**10a**), rubiarbonol G (**42**), RA-V (**52**), rubiquinone-3-*O*-*β*-D-xylopyranosyl-(1→2)-(6'-*O*-acetyl)-*β*-D-glucopyranoside (**54**), rubiquinone-3-*O*-*β*-L-rhamnopyranosyl-(1→2)-*β*-D-glucopyranoside (**55**), and rubiquinone-3-*O*-*β*-D-glucopyranoside (**56**) were previously isolated from *R. yunnanensis*.[1] 2-Methyl-3-hydroxy-9,10-anthraquinone (**2**), 3,6-dihydroxy-xanthen-9-one (**3**), emodin (**4**), 2-hydroxy-9,10-anthraquinone (**5**), 2-amino-3-hydroxy-9,10-anthraquinone (**6**), 2,6-dihydroxy-9,10-anthraquinone (**8**), aloe-emodin (**9**), purpurin (**11**), 2,6-diamino-9,10-anthraquinone (**12**), apigenin (**13**), baicalein (**14**), kaempferol (**15**), luteolin (**16**), quercetin (**17**), myricetin (**18**), naringenin (**19**), daidzein (**20**), phloretin (**21**), butein (**22**), hematoxylin (**23**), silibinin (**24**), resveratrol (**25**), bis-demethoxycurcumin (**26**), magnolol (**27**), chlorogenic acid (**28**), ferulic acid (**29**), paeonol (**30**), 3,4-dichloroaniline (**31**), 3,4-dichlorobenzenethiol (**32**), chrysophanol (**33**), anthrarufin (**34**), 1,4-dihydroxy-9,10-anthraquinone (**35**), 2-hydroxy-1,4-naphthoquinone (**36**), 5-hydroxy-1,4-naphthalenedione (**37**), mollugin (**38**), *β*-mangostin (**39**), cyanidin chloride (**40**), senegenin (**41**), dihydroartemisinin (**43**), cyclovirobuxine D (**44**), andrographolide (**45**), camptothecin (**46**), 7-hydroxycoumarin (**47**), L-tyrosine (**48**), 4-hydroxyphenethyl alcohol (**49**), 4-hydroxybenzyl alcohol (**50**), crocetin (**51**), and 4-nitrophenyl *β*-D-glucopyranoside (**53**) were purchased from Yuanye Bio-Technology Co., Ltd (Shanghai, China). UDP-glucose, UDP-galactose, UDP-N-acetylgluccosamine and UDP-glucuronic acid were purchased from Sigma-Aldrich (St. Louis, MO, USA). Vectors, the chemically competent cells, and Trelief™ SoSoo Clonging Kit were purchased from Tsingke Biotech Co., Ltd (Beijing, China); Universal DNA Purification Kit were purchased from Vazyme Biotech Co., Ltd. (Nanjing, China). Rabbit Anti-His Tag antibody (bs-10582R) and anti-rabbit IgG/HRP antibody (bs-0295G-HRP) were purchased from Bioss Biotechnology

Co., Ltd. (Nanjing, China). SuperSignal West Pico Chemiluminescent Substrate was purchased from Pierce, Waltham, USA. Methanol and acetonitrile (Merck, Germany) were of HPLC grade. Substrate specificity and conversion rates were analyzed by HPLC on Waters ACQUITY Arc with YMC-Triart C_{18} column (250 × 4.6 mm, 5 μm). LC-MS analysis was performed on Waters Xevo-TQD MS spectrometer with Waters ACQUITY UPLC BEH C_{18} column (2.1 × 50 mm, 1.7 μm). The glycosylated products were isolated and purified by LC-3000 HPLC spectrometer (Beijing Chuangxintongheng Science & Technology Co., Ltd.) with YMC-Pack ODS-A C_{18} column (10 × 250 mm, 5 μm) and YMC-Pack ODS-A C_{18} column (20 × 250 mm, 5 μm). NMR spectra were recorded on a Bruker AVANCE III-400/500 instrument at 400/500 (^{1}H) and 100/125 (^{13}C) MHz in DMSO-d_6 and HR-ESI-MS spectra were obtained with Agilent Q-6200-TOF spectrometer.

1.2 Plant materials

The one-year-old *R. yunnanensis* materials were collected as described in our previous report.[2]

1.3 Transcriptome analysis, molecular cloning and phylogenetic analysis for screening candidate genes

1.3.1 Transcriptome analysis

Anthraquinone glycosides are one of the major natural products produced in *R. yunnanensis*. According to our previous work, interestingly, the roots produced more anthraquinone glycosides than those from the stems or leaves (Figure S1), which is extremely useful to screen out the candidate GT genes by comparing the expression levels in the transcriptome of different parts of *R. yunnanensis*. Thus, the transcriptome data of root and stem leaf (mixed stem and leaf) of annual *R. yunnanensis* were finished by our laboratory (unpublished data). A total of 187 and 312 candidate unigenes were annotated as GT genes based on the results of comparison with Nr and Swiss-prot databases, and then 57 and 111 candidate genes were further screened out from the above two groups by

significant difference expression analysis, respectively (unpublished data). Then, 31 target GT unigenes (Table S1) were selected as target genes according to the analysis results of GO molecular function enrichment and significant difference expression using the criterion of high gene expression in root or stem leaf (unpublished data). Among them, 17 GT unigenes were found to have full-length protein coding sequences (CDs) referring to prediction results of their open reading frames (ORFs) based on ORFfinder (https://www.ncbi.nlm.nih.gov/orffinder/).

1.3.2 Molecular cloning

The methods of total RNA isolation and quality control were described in our previous report.[2] First-strand cDNA was prepared with total or poly A^+ RNA according to the instructions suggested in SMARTer^{TM} RACE cDNA Amplification Kit (Clontech Laboratories, Inc., Mountain View, CA, USA). The full-length CDs of 17 GT genes were obtained with 17 full-length unigenes by regular PCR and the remaining 14 unigenes without the full-length CDs were used as the templates to design gene-specific primers (Table S2) for amplification of 5'- and/or 3'- end fragments by SMARTer^{TM} RACE cDNA Amplification Kit. All the 5'- and 3'- end fragments were sequenced to assemble full length cDNA. Among them, we used the "RyUGT3: >TRINITY_DN123143_c0_g2_i4" as the template to design RACE primers (Table S2). The amplification results were shown in Figure S2A. Then, the recovered products from line 2 and 3 were connected to the clone vector, and positive clones were selected for sequencing, so as to obtain the gene information of the 3' and 5' -ends of the target genes. Finally, the splicing was carried out with the help of DNAMAN software, and the nucleotide sequence of the splicing results was further compared. Results as shown in Figure S2B, two different splicing results were observed among the nine positive clones randomly selected. The sequence information named *UGT3A4*, *UGT3A8*, *UGT3A9* was relatively consistent with the template sequence used, while the other six sequences (named *UGT3B-1~7*) were quite different from the template sequence used. But their sequences also have many identical parts, which may be explained by the fact that there is another gene sequence with high similarity to the template sequence used. Therefore, considering that there are two *UGT* sequences similar

to transcript "RyUGT3", they are named as RyUGT3A and RyUGT3B, respectively. Thus, 15 sequences were obtained from the 14 partial unigenes by RACE with gene-specific primers designed by their nucleotide sequences. Finally, we totally obtained 32 GT genes containing the complete CDs.

1.3.3 Phylogenetic analysis

For the phylogenetic analysis, the full-length amino acid sequences of 32 GTs from *R. yunnanensis* were aligned with 122 GTs from *Arabidopsis thaliana* using ClustalW.[3,4] The resulting alignment was used to build an unrooted phylogenetic tree using the neighbor-joining method in the MEGA6.0. One thousand bootstrapped replicates were applied to estimate the confidence of each tree clade (Figure S3). The respective protein names and genebank numbers are listed in Table S3.

1.4 Heterologous expression, protein purification and enzyme catalytic activity assay

1.4.1 Heterologous expression and protein purification

Multiple sequence alignment of amino acid sequences of the 32 candidate GTs were conducted by DNAMAN. The results showed that the amino acid identities of RyUGT3A and RyUGT10, RyUGT3B and RyUGT21, RyUGT11 and RyUGT24, and RyUGT12 and RyUGT13 were 98.8%, 99.8%, 99.9%, and 99.8%, respectively, which indicated that they are orthologous genes. Therefore, the 28 GT genes (except *RyUGT10*, *RyUGT13*, *RyUGT24*, and *RyUGT21*) were inserted into pET-28a (+) vector according to the manual of Trelief™ SoSoo Clonging Kit. After verifying by sequencing, all the pET28a (+)-UGT expression constructs were transformed in *E. coli* BL21(DE3) or *E. coli* BL21(DE3) plysS. The colonies were grown at 37 ℃ in 50 mL of Luria-Bertani (LB) media containing 50 mg/L kanamycin. The cells were induced with 0.5 mM isopropyl *β*-D-thiogalactoside (IPTG) when the OD600 value was reached 0.6-0.8. After growing about 16 h at 16 ℃ with shaking, the cultures were harvested by centrifugation at 12,000 rpm, 4 ℃ for 15 min and stored at -80 ℃. Frozen cells were resuspended in 15 mL pre-cooling lysis buffer (20 mM phosphate buffer, 50 mM NaCl, pH 7.5) and disrupted by sonication in ice-water bath. The cell lysate was centrifuged at 12,000 rpm, 4 ℃ for 15 min to obtain the

supernatant (the crude enzyme solution). The supernatant was further confirmed by SDS-PAGE and Western Blotting (WB) and then purified referring to a previous study (Figure S5A and S5B).[5,6] The purified protein was further confirmed by SDS-PAGE (Figure S5C). After their concentrations were determined according to BCA protein assay kit (Beyotime Institute of Biotechnology, Shanghai, China), all recombinant enzymes were stored at –80 ℃ until used for catalytic activity assays. Finally, 16 recombinant RyUGTs were soluble in *E. coli* and the rest were either not expressed (seven RyUGTs) or formed in inclusion body (five RyUGTs) (Table S4).

1.4.2 Enzyme catalytic activity assay

A standardized procedure was conducted with different native substrates for all RyUGTs screening. The glycosylation reactions were performed in 200 μL system containing 50 mM Tris-HCl buffer (pH 7.5), 14 mM β-mercaptoethanol, 5 mM UDP-Glc, 0.5 mM substrate and 100 μl crude enzyme. After incubation at 30 ℃ for 2 h, the reactions were terminated by adding equal volume of methanol. Then the mixed samples were centrifuged at 12000 rpm for 15 min. Negative controls were carried out without adding sugar donor. HPLC and LC/MS were employed to analyze the collected supernatants. Specifically, samples were analyzed by Waters ACQUITY Arc. HPLC and LC-MS analysis were performed on Waters ACQUITY Arc and Waters Xevo-TQD MS spectrometer with YMC-Triart C_{18} column (250 × 4.6 mm, 5 μm) or Waters ACQUITY UPLC BEH C_{18} column (2.1 × 50 mm, 1.7 μm), respectively. Two different HPLC programs (method **a** and **b**) with a mobile phase of water containing 0.1% formic acid (A) and acetonitrile (B) were applied: method **a** with a flow rate of 0.8 mL/min and injection volume 20 μL for YMC-Triart C_{18} column: 0-5 min, 5%-20% B; 5-8 min, 20%-22% B; 8-17 min, 22-25% B; 17-23 min, 25-35% B； 23-25 min, 35-50% B; 25-32 min, 50-100% B; method **b** with a flow rate of 0.2 mL/min and injection volume 2 μL for BEH C_{18} column: 0-3 min, 10-100% B. The mass spectrum program was set as follows: ion source, ESI (negative mode); capillary voltage 2.5 kV; N_2 flow rate 800 L/h; dissociation temperature 500 ℃. In addition, three other sugar donors (UDP-GluA, UDP-GlcNAc, and UDP-Gal) were tested to explore the promiscuity of sugar donor when using **1** as acceptor

under the condition mentioned above (Figure S10). After confirming that the substrates are glycosylated by RyUGT3A/RyUGT12, new glycosylation reactions were performed in 200 μL system containing 50 mM Tris-HCl buffer (pH 9.0/8.0), 14 mM β-mercaptoethanol, 5 mM $MgCl_2$, 5 mM UDP-Glc, 0.5 mM substrate and 5 μg purified enzyme. The reaction conditions and sample treatment are the same as above. The conversion rates (%) were measured by HPLC peak area ($A_{product}/A_{substrate+product} \times 100\%$) and listed in Table S5.

1.5 Effects of pH, temperature, and divalent metal ions

In order to optimize the reaction temperature of RyUGT3A and RyUGT12 (Figures S6 and S7), different temperatures from 4–60 ℃ were investigated at pH 7.5. The optimal pH were determined at 30 ℃ in the range of pH 5.0-6.0 (Citric acid-sodium citrate buffer) and 7.0–11.0 (Tris-HCl buffer). To explore the metal dependence of RyUGT3A and RyUGT12, reactions were performed at 30 ℃ and pH 7.5 in the presence of different metals, Mg^{2+}, Ca^{2+}, Fe^{2+}, Zn^{2+}, Mn^{2+}, Co^{2+} and Ba^{2+}, at a final concentration of 5 mM. Experiment with 5 mM EDTA or Tris-HCl (pH 7.5, control blank) was performed as a negative control. All experiments were performed in triplicate and the mean value was used.

1.6 Kinetic studies

For determination of the kinetic parameters of the acceptor (Figure S8), reactions were performed in a final volume of 200 μL containing 50 mM Tris-HCl (pH 9.0/8.0), 50 μg of purified enzymes (RyUGT3A and RyUGT12), varying concentrations (0.65 mM, 1.3 mM, 1.95 mM, 2.6 mM, 3.25 mM and 3.9 mM) of **1**. All the kinetic experiments were incubated at 30 ℃ or 40 ℃ for 10 min and repeated in triplicate. After adding 200 μL methanol, the samples were centrifuged and analyzed as the methods described above. All reactions were carried out with UDP-Glc as a donor and **1** as an acceptor. The K_m values were calculated using the Lineweaver-Burk plot methods.

1.7 Homology modeling and site-directed mutagenesis

1.7.1 Homology modeling and molecular docking

Homology modeling was conducted in MOE v2015.1001 (Chemical Computing Group, Canada) according to the method described by Liu et al.[7] The template protein crystal structures of RyUGT3A and RyUGT12 were identified through BLAST and downloaded from RCSB Protein Data Bank (PDB ID: 5NLM and 2ACW), respectively. The Chain A of 5NLM was utilized as template for RyUGT3A and that of 2ACW for RyUGT12. The overall identity of the amino acid sequence was 40.45% and 30.94% for protein 12 and 3A, respectively. The protonation state of the proteins and the orientation of the hydrogens were optimized by LigX at pH7 and the temperature of 300 K. First, the target sequence was aligned to the template sequence, and ten independent intermediate models were built. These different homology models were the results of the permutational selection of different loop candidates and side chain rotamers. Then, the intermediate model which scored the best according to the GB/VI scoring function was chosen as the final model, and was subjected to further energy minimization using the AMBER10: EHT force field (Figures S33 and S34).

The 2D structures of UDP-Glc and substrate (**1**) were converted to 3D structures in MOE through energy minimization. Prior to docking, the force field of AMBER10: EHT and the implicit solvation model of Reaction Field (R-field) were selected. MOE-Dock was used for molecular docking. The binding sites of RyUGT3A and RyUGT12 were confirmed by superimposing their structures with the structure of the templates. The docking workflow followed the "induced fit" protocol,[8] in which the side chains of the receptor pocket were allowed to move according to ligand conformations, with a constraint on their positions. The weight used for tethering side chain atoms to their original positions was 10. Before conducting molecular docking, we have referred to the work of other researchers.[9,10] In their work, sugar groups are also stable and participate in interactions. So when we do molecular docking, UDP-Glc was firstly docked into RyUGT3A and RyUGT12 individually. All docked poses of which were ranked by London dG scoring first, then a force field refinement was carried out on the top 20 poses

followed by a rescoring of GBVI/WSA dG. The conformations with the lowest free energies of binding were selected as the best (probable) binding modes for further docking of substrate (**1**) by using the same "induced fit" protocol. Molecular graphics were generated by PyMOL (http://www.pymol.org).

1.7.2 Site-directed mutagenesis and their mutant activity assay

According to the results, site-directed mutagenesis experiment was carried out by overlap PCR method and the corresponding primers were listed in Table S6. The products of site-directed mutagenesis were transformed in *E. coli* BL21(DE3). After confirming the expression of each enzyme, they were purified in turn for analyzing the conversion yield of each enzyme to compound **1**.

1.8 One-pot reactions catalyzed by RyUGT3A and preparative-scale reactions

1.8.1 One-pot reactions

Firstly, verification test of reversible reaction of RyUGT3A was conducted with 200 μL Tris-HCl (50 mM, pH 9.0) containing 14 mM β-mercaptoethanol, 50 mM 4-nitrophenyl-β-D-glucopyranoside (**53**), 5 mM UDP, and 5 μg purified RyUGT3A. Then, the reaction was incubated at 30 ℃ for 24 h. After adding with 200 μL ice-cold methanol, the centrifuged supernatant was analyzed by HPLC and LC-MS as described above. Finally, the 200 μL reaction contained 50 mM Tris-HCl (pH 9.0), 14 mM β-mercaptoethanol, 50 mM **53**, 5 mM UDP, 0.5 mM substrate (**1)** and 5 μg purified RyUGT3A was incubated at the optimum condition for 24 h. The same treatment and detect method were used to analysis the clarified reaction mixture. The negative control reaction was without UDP, and the positive control was directly using UDP-Glc as sugar donor.

1.8.2 Preparative-scale reactions

Eleven aglycones were selected to prepare their corresponding glycosylated products for determining the glycosylation sites by scaled-up one-pot glycosylation reactions of RyUGT3A. The specific amplification system was as follows: 100 mL reaction contained 50 mM Tris-HCl (pH 9.0), 14 mM β-mercaptoethanol, 50 mM **53**, 5 mM UDP, 0.5 mM

substrate and 250 μg purified RyUGT3A. After incubation at 30 ℃ for 24 h, the reactions were terminated by adding 100 mL methanol. The mixtures were centrifuged at 12000 rpm for 25 min and then the supernatants were concentrated and dissolved in 50% methanol. All the chromatograms were obtained and detected between 190 nm and 400 nm. The glycosylated products were purified by a reverse-phase preparative HPLC on an LC 3000 instrument (Tong Heng, Beijing, China) equipped with an YMC-Pack ODS-A C_{18} column (250 mm × 10 mm, 5 μm). The purified products were characterized by LC-MS, NMR and HR-ESI-MS (Figures S12-S32).

1.9 MS, ^{1}H and ^{13}C NMR data of glycosylated products

6-Hydroxyalizarin 3-*O*-*β*-D-glucoside (1a) (12.4 mg): brownish yellow powder. ESI-MS for $C_{21}H_{19}O_{10}$ $[M-H]^-$: 431.35. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 7.40 (1H, s, H-4), 7.42 (1H, s, H-5), 7.17 (1H, d, J = 8.4 Hz, H-7), 8.06 (1H, d, J = 8.4 Hz, H-8), 5.09 (1H, d, J = 6.0 Hz, H-1′), 3.34 (2H, overlap, H-2′, 5′), 3.25 (1H, overlap, H-4′), 3.41 (1H, overlap, H-3′), 3.70 (1H, d, J = 11.8 Hz, H-6′a), 3.54 (1H, dd, J = 11.8, 5.1 Hz, H-6′b), 2.16 (3H, s, 2-CH_3). ^{13}C NMR (100 MHz, DMSO-d_6) δ: 161.3 (s, C-1), 120.8 (s, C-2), 160.6 (s, C-3), 105.5 (d, C-4), 113.1 (d, C-5), 165.0 (s, C-6), 121.8 (d, C-7), 129.6 (d, C-8), 186.2 (s, C-9), 181.8 (s, C-10), 110.6 (s, C-1a), 132.0 (s, C-4a), 135.3 (s, C-5a), 123.6 (s, C-8a), 100.4 (d, C-1′), 73.2 (d, C-2′), 77.3 (d, C-3′), 69.4 (d, C-4′), 76.3 (d, C-5′), 60.5 (t, C-6′), 8.5 (q, 2-CH_3).[11]

6-Hydroxyalizarin 6-*O*-*β*-D-glucoside (1b) (10.8 mg): brownish black powder. HR-ESI-MS calcd. for $C_{21}H_{21}O_{10}$ $[M+H]^+$: 433.1129; found: 433.1139. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 7.20 (1H, s, H-4), 7.63 (1H, s, H-5), 7.49 (1H, d, J = 7.8 Hz, H-7), 8.12 (1H, d, J = 8.6 Hz, H-8), 5.13 (1H, d, J = 6.7 Hz, H-1′), 3.22-3.71 (6H, sugar protons), 2.04 (3H, s, 2-CH_3). ^{13}C NMR (100 MHz, DMSO-d_6) δ: 163.5 (s, C-1), 117.4 (s, C-2), 162.4 (s, C-3), 108.2 (d, C-4), 113.3 (d, C-5), 161.8 (s, C-6), 121.7 (d, C-7), 128.9 (d, C-8), 185.2 (s, C-9), 181.7 (s, C-10), 107.9 (s, C-1a), 131.8 (s, C-4a), 134.9 (s, C-5a), 127.2 (s, C-8a), 100.1 (d, C-1′), 73.1 (d, C-2′), 77.3 (d, C-3′), 69.5 (d, C-4′), 76.3 (d, C-5′), 60.5 (t, C-6′), 8.1 (q, 2-CH_3).

6-Hydroxyalizarin 3,6-di-*O*-*β*-D-glucoside (1c) (9.7 mg): brownish black powder. HR-ESI-MS calcd. for $C_{27}H_{30}O_{15}Na$ $[M+Na]^+$: 617.1477; found: 617.1475. ^1H NMR (500 MHz, DMSO-d_6) δ: 7.45 (1H, s, H-4), 7.68 (1H, d, J = 2.5 Hz, H-5), 7.52 (1H, dd, J = 8.6, 2.5 Hz, H-7), 8.19 (1H, d, J = 8.6 Hz, H-8), 5.15 (1H, d, J = 6.9 Hz, H-1′), 5.12 (1H, d, J = 6.1 Hz, H-1″), 3.23-3.70 (12 H, sugar protons), 2.18 (3H, s, 2-CH_3). ^{13}C NMR (125 MHz, DMSO-d_6) δ: 161.0 (s, C-1), 120.8 (s, C-2), 161.4 (s, C-3), 105.8 (d, C-4), 113.4 (d, C-5), 162.1 (s, C-6), 121.9 (d, C-7), 129.2 (s, C-8), 186.4 (s, C-9), 181.2 (s, C-10), 110.7 (s, C-1a), 132.0 (s, C-4a), 135.0 (s, C-5a), 126.9 (s, C-8a), 100.3 (d, C-1′), 73.2 (d, C-2′), 77.3 (d, C-3′), 69.5 (d, C-4′), 75.7 (d, C-5′), 60.5 (t, C-6′), 100.0 (d, C-1″), 73.1 (d, C-2″), 77.3 (d, C-3″), 69.4 (d, C-4″), 76.3 (d, C-5″), 60.5 (t, C-6″), 8.5 (q, 2-CH_3).

2-Hydroxy-3-methyl-9,10-anthraquinone 2-*O*-*β*-D-glucoside (**2a**) (7.9 mg): bronzing powder. HR-ESI-MS calcd. for $C_{21}H_{21}O_8$ $[M+H]^+$: 401.1231; found: 401.1229. ^1H NMR (400 MHz, DMSO-d_6) δ: 7.71 (1H, s, H-1), 7.97 (1H, s, H-4), 8.13 (2H, overlap, H-5, 8), 7.88 (2H, overlap, H-6, 7), 5.13 (1H, d, J = 7.3 Hz, H-1′), 3.28-3.72 (6 H, sugar protons), 2.36 (3H, s, 3-CH_3). ^{13}C NMR (100 MHz, DMSO-d_6) δ: 110.7 (d, C-1), 160.2 (s, C-2), 134.3 (s, C-3), 129.5 (d, C-4), 126.7 (d, C-5, 8), 134.4 (d, C-6), 134.5 (d, C-7), 182.2 (s, C-9), 181.7 (s, C-10), 133.1 (s, C-1a, 8a), 127.0 (s, C-4a), 133.0 (s, C-5a), 100.5 (d, C-1′), 73.3 (d, C-2′), 77.4 (d, C-3′), 69.5 (d, C-4′), 76.4 (d, C-5′), 60.5 (t, C-6′), 16.5 (q, 3-CH_3).

Xanthen-9-one 3-*O*-*β*-D-glucoside (**3a**) (8.5 mg): yellow oil. HR-ESI-MS calcd. for $C_{19}H_{19}O_9$ $[M+H]^+$: 391.1024; found: 391.1031. ^1H NMR (400 MHz, DMSO-d_6) δ: 7.94 (1H, d, J = 8.8 Hz, H-1), 6.94 (1H, dd, J = 8.8, 2.3 Hz, H-2), 7.02 (1H, d, J = 2.3 Hz, H-4), 6.09 (1H, d, J = 2.2 Hz, H-5), 6.31 (1H, dd, J = 9.0, 2.2 Hz, H-7), 7.66 (1H, d, J = 9.0 Hz, H-8), 5.06 (1H, J = 7.3 Hz, H-1′), 3.28-3.72 (6 H, sugar protons). ^{13}C NMR (100 MHz, DMSO-d_6) δ: 126.9 (d, C-1), 112.5 (d, C-2), 161.0 (s, C-3), 103.0 (d, C-4), 102.0 (d, C-5), 177.0 (s, C-6), 119.8 (d, C-7), 126.7 (d, C-8), 172.5 (s, C-9), 116.6 (s, C-1a), 156.8 (s, C-4a), 159.5 (s, C-5a), 106.9 (s, C-8a), 100.0 (d, C-1′), 73.3 (d, C-2′), 77.2 (d, C-3′), 69.7 (d, C-4′), 76.6 (d, C-5′), 60.8 (t, C-6′).

Xanthen-9-one 3,6-di-*O*-*β*-D-glucoside (**3b**) (7.9 mg): yellow oil. ESI-MS calcd. for

$C_{25}H_{29}O_{14}$ $[M+H]^+$: 553.24. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 8.10 (2H, d, J = 8.8 Hz, H-1, 8), 7.10 (2H, dd, J = 8.8, 2.3 Hz, H-2, 7), 7.21 (2H, d, J = 2.3 Hz, H-4, 5), 5.17 (2H, J = 7.1 Hz, H-1′, 1″), 3.16-3.70 (12 H, sugar protons). ^{13}C NMR (100 MHz, DMSO-d_6) δ: 127.6 (d, C-1, 8), 114.6 (d, C-2, 7), 162.5 (s, C-3, 6), 103.1 (d, C-4, 5), 174.3 (s, C-9), 115.9 (s, C-1a, 8a), 157.3 (s, C-4a, 5a), 99.8 (d, C-1′, 1″), 73.2 (d, C-2′, 2″), 77.2 (d, C-3′, 3″), 69.6 (d, C-4′, 4″), 76.4 (d, C-5′, 5″), 60.7 (t, C-6′, 6″).[12]

Emodin 6-*O*-*β*-D-glucoside (4a) (14.6 mg): brownish black powder. ESI-MS for $C_{21}H_{19}O_{10}$ $[M-H]^-$: 431.31. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 7.10 (1H, s, H-2), 7.43 (1H, s, H-4), 7.10 (1H, s, H-5), 6.91 (1H, s, H-7), 5.11 (1H, d, J = 7.2 Hz, H-1′), 3.32 (2H, overlap, H-2′, 5′), 3.46 (1H, overlap, H-3′), 3.22 (1H, overlap, H-4′), 3.71 (1H, d, J = 10.8 Hz, H-6′a), 3.52 (1H, overlap, H-6′b), 2.38 (3H, s, 3-CH_3). ^{13}C NMR (100 MHz, DMSO-d_6) δ: 163.8 (s, C-1), 124.3 (d, C-2), 148.4 (s, C-3), 120.5 (d, C-4), 108.8 (d, C-5), 164.5 (s, C-6), 109.2 (d, C-7), 161.7 (s, C-8), 189.7 (s, C-9), 181.1 (s, C-10), 113.5 (s, C-1a), 134.8 (s, C-4a), 132.8 (s, C-5a), 110.8 (s, C-8a), 100.0 (d, C-1′), 73.1 (d, C-2′), 77.3 (d, C-3′), 69.5 (d, C-4′), 76.3 (d, C-5′), 60.6 (t, C-6′), 21.6 (q, 3-CH_3).[13]

2-Hydroxy-9,10-anthraquinone 2-*O*-*β*-D-glucoside (5a) (11.2 mg): brown powder. ESI-MS for $C_{21}H_{19}O_{10}$ $[M+HCOO]^-$: 431.17. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 7.70 (1H, s, H-1), 7.53 (1H, d, J = 8.8 Hz, H-3), 8.18 (3H, overlap, H-4, 5, 8), 7.92 (2H, d, J = 6.1 Hz, H-6, 7), 5.15 (1H, d, J = 6.3 Hz, H-1′), 3.34 (2H, overlap, H-2′, 5′), 3.46 (1H, overlap, H-3′), 3.23 (1H, overlap, H-4′), 3.71 (1H, d, J = 11.8 Hz, H-6′a), 3.52 (1H, overlap, H-6′b). ^{13}C NMR (100 MHz, DMSO-d_6) δ: 113.2 (d, C-1), 162.0 (s, C-2), 122.0 (d, C-3), 129.5 (d, C-4), 126.8 (d, C-5), 134.7 (d, C-6), 134.3 (d, C-7), 126.7 (d, C-8), 182.2 (s, C-9), 181.4 (s, C-10), 135.0 (s, C-1a), 127.3 (s, C-4a), 133.1 (s, C-5a), 133.1 (s, C-8a), 100.1 (d, C-1′), 73.2 (d, C-2′), 77.3 (d, C-3′), 69.5 (d, C-4′), 76.4 (d, C-5′), 60.5 (t, C-6′).[14]

2-Amino-3-hydroxy-9,10-anthraquinone 3-*O*-*β*-D-glucoside (6a) (8.7 mg): brownish red powder. ESI-MS for $C_{20}H_{18}NO_8$ $[M-H]^-$: 400.35. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 7.39 (1H, s, H-1), 7.70 (1H, s, H-4), 8.10 (2H, d, H-5, 8), 7.83 (2H, overlap, H-6, 7), 4.89 (1H, d, J = 5.7 Hz, H-1′), 3.39 (4H, overlap, H-2′, 3′, 4′, 5′), 3.73 (1H, d, J = 11.9 Hz, H-

6′a), 3.58 (1H, dd, J = 11.9, 4.9 Hz, H-6′b), 6.43 (2H, s, 2-NH_2). ^{13}C NMR (100 MHz, DMSO-d_6) δ: 110.2 (d, C-1), 145.2 (s, C-2), 147.7 (s, C-3), 112.7 (d, C-4), 126.3 (d, C-5, 8), 133.5 (d, C-6, 7), 182.6 (s, C-9), 180.4 (s, C-10), 129.9 (s, C-1a), 122.4 (s, C-4a), 133.7 (s, C-5a), 134.1 (s, C-8a), 101.7 (d, C-1′), 73.2 (d, C-2′), 77.3 (d, C-3′), 69.6 (d, C-4′), 75.7 (d, C-5′), 60.5 (t, C-6′).[15]

Apigenin 7-*O*-*β*-D-glucoside (13a) (13.2 mg): brown powder. ESI-MS for $C_{21}H_{19}O_{10}$ $[M-H]^-$: 431.35. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 6.85 (1H, s, H-3), 6.43 (1H, s, H-6), 6.82 (1H, s, H-8), 7.94 (2H, d, J = 8.3 Hz, H-2′, 6′), 6.94 (2H, d, J = 8.3 Hz, H-3′, 5′), 5.07 (1H, d, J = 7.2 Hz, H-1″), 3.28 (2H, overlap, H-2″, H-5″), 3.48 (2H, overlap, H-3″, H-6″b), 3.19 (1H, t, J = 8.9 Hz, H-4″), 3.72 (1H, d, J = 10.9 Hz, H-6″a). ^{13}C NMR (100 MHz, DMSO-d_6) δ: 164.3 (s, C-2), 103.1 (d, C-3), 182.0 (s, C-4), 161.6 (s, C-5), 99.6 (d, C-6), 163.0 (s, C-7), 94.8 (d, C-8), 157.0 (s, C-9), 105.4 (s, C-10), 120.9 (s, C-1′), 128.7 (d, C-2′, 6′), 116.7 (d, C-3′, 5′), 161.3 (s, C-4′), 99.9 (d, C-1″), 73.2 (d, C-2″), 77.2 (d, C-3″), 69.6 (d, C-4″), 76.5 (d, C-5″), 60.7 (t, C-6″).[16]

Baicalein 6-*O*-*β*-D-glucoside (14a) (12.1 mg): brown powder. ESI-MS for $C_{21}H_{19}O_{10}$ $[M-H]^-$: 431.31. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 6.94 (1H, s, H-3), 6.60 (1H, s, H-8), 8.06 (2H, d, J = 6.8 Hz, H-2′, 6′), 7.58 (3H, overlap, H-3′, 4′, 5′), 4.80 (1H, d, J = 7.6 Hz, H-1″), 3.33 (1H, m, H-2″), 3.16 (1H, m, H-3″), 3.20 (1H, m, H-4″), 3.25 (1H, m, H-5″), 3.64 (1H, d, J = 11.0 Hz, H-6″a), 3.48 (1H, dd, J = 11.0, 4.9 Hz, H-6″b). ^{13}C NMR (100 MHz, DMSO-d_6) δ: 163.1 (s, C-2), 104.7 (d, C-3), 182.1 (s, C-4), 152.6 (s, C-5), 128.8 (s, C-6), 158.6 (s, C-7), 94.7 (d, C-8), 153.2 (s, C-9), 103.9 (s, C-10), 130.8 (s, C-1′), 126.4 (d, C-2′, 6′), 129.2 (d, C-3′, 5′), 132.0 (d, C-4′), 104.4 (d, C-1″), 74.0 (d, C-2″), 77.3 (d, C-3″), 69.6 (d, C-4″), 76.3 (d, C-5″), 60.7 (t, C-6″).[17]

Baicalein 7-*O*-*β*-D-glucoside (14b) (10.3 mg): brown powder. ESI-MS for $C_{21}H_{19}O_{10}$ $[M-H]^-$: 431.35. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 7.00 (1H, s, H-3), 7.05 (1H, s, H-8), 8.07 (2H, d, J = 7.2 H, H-2′, 6′), 7.60 (3H, overlap, H-3′, 4′, 5′), 5.02 (1H, d, J = 7.1 Hz, H-1″), 3.19-3.76 (6H, sugar protons). ^{13}C NMR (100 MHz, DMSO-d_6) δ: 163.5 (s, C-2), 104.7 (d, C-3), 182.6 (s, C-4), 146.5 (s, C-5), 130.9 (s, C-6), 151.7 (s, C-7), 94.3 (d, C-8), 149.2 (s, C-9), 106.1 (s, C-10), 130.7 (s, C-1′), 126.4 (d, C-2′, 6′), 129.2 (d, C-3′, 5′),

132.1 (d, C-4′), 101.0 (d, C-1″), 73.2 (d, C-2″), 77.4 (d, C-3″), 69.7 (d, C-4″), 75.9 (d, C-5″), 60.7 (t, C-6″).[18]

Kaempferol 3-*O*-*β*-D-glucoside (15a) (10.5 mg): brownish black powder. ESI-MS for $C_{21}H_{19}O_{11}$ $[M-H]^-$: 447.33. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 6.21 (1H, s, H-6), 6.43 (1H, s, H-8), 8.03 (2H, d, J = 8.5 Hz, H-2′, 6′), 6.88 (2H, d, J = 8.5 Hz, H-3′, 5′), 5.46 (1H, d, J = 7.3 Hz, H-1″), 3.08-3.53 (6H, sugar protons). ^{13}C NMR (100 MHz, DMSO-d_6) δ: 156.4 (s, C-2), 133.2 (s, C-3), 177.5 (s, C-4), 161.2 (s, C-5), 98.8 (d, C-6), 164.4 (s, C-7), 93.7 (d, C-8), 156.2 (s, C-9), 103.9 (s, C-10), 120.9 (s, C-1′), 130.9 (d, C-2′, 6′), 115.1 (d, C-3′, 5′), 160.0 (s, C-4′), 100.9 (d, C-1″), 74.2 (d, C-2″), 77.5 (d, C-3″), 69.9 (d, C-4″), 76.4 (d, C-5″), 60.9 (t, C-6″).[19]

Kaempferol 7-*O*-*β*-D-glucoside (15b) (11.6 mg): brownish black powder. ESI-MS for $C_{21}H_{19}O_{11}$ $[M-H]^-$: 447.33. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 6.42 (1H, s, H-6), 6.80 (1H, s, H-8), 8.09 (2H, d, J = 8.3 Hz, H-2′, 6′), 6.95 (2H, d, J = 8.3 Hz, H-3′, 5′), 5.07 (1H, d, J = 7.4 Hz, H-1″), 3.28 (1H, overlap, H-2″), 3.47 (1H, overlap, H-3″), 3.17 (1H, overlap, H-4″), 3.28 (1H, overlap, H-5″), 3.70 (1H, d, J = 10.4 Hz, H-6″a), 3.47 (1H, overlap, H-6″b). ^{13}C NMR (100 MHz, DMSO-d_6) δ: 147.5 (s, C-2), 136.4 (s, C-3), 176.2 (s, C-4), 160.4 (s, C-5), 98.8 (d, C-6), 162.7 (s, C-7), 94.3 (d, C-8), 155.8 (s, C-9), 104.7 (s, C-10), 121.6 (s, C-1′), 129.5 (d, C-2′, 6′), 115.5 (d, C-3′, 5′), 159.4 (s, C-4′), 99.9 (d, C-1″), 73.1 (d, C-2″), 77.2 (d, C-3″), 69.6 (d, C-4″), 76.5 (d, C-5″), 60.6 (t, C-6″).[19]

Kaempferol 3,7-di-*O*-*β*-D-glucoside (15c) (12.3 mg): brownish black powder. ESI-MS for $C_{27}H_{29}O_{16}$ $[M-H]^-$: 609.34. ^{1}H NMR (500 MHz, DMSO-d_6) δ: 6.42 (1H, d, J = 2.1 Hz, H-6), 6.77 (1H, d, J = 2.1 Hz, H-8), 8.04 (2H, d, J = 8.5 Hz, H-2′, 6′), 6.89 (2H, d, J = 8.5 Hz, H-3′, 5′), 5.06 (1H, d, J = 7.5 Hz, H-1″), 5.45 (1H, d, J = 7.3 Hz, H-1‴), 3.06-3.67 (12H, sugar protons), 12.6 (1H, s, 4′-OH). ^{13}C NMR (125 MHz, DMSO-d_6) δ: 156.8 (s, C-2), 133.4 (s, C-3), 177.6 (s, C-4), 160.8 (s, C-5), 99.3 (d, C-6), 162.8 (s, C-7), 94.5 (d, C-8), 156.0 (s, C-9), 105.6 (s, C-10), 120.7 (s, C-1′), 130.9 (d, C-2′, 6′), 115.2 (d, C-3′, 5′), 160.2 (s, C-4′), 99.7 (d, C-1″), 73.1 (d, C-2″), 77.1 (d, C-3″), 69.6 (d, C-4″), 76.4 (d, C-5″, 5‴), 60.8 (t, C-6″), 100.7 (d, C-1‴), 74.2 (d, C-2‴), 77.5 (d, C-3‴), 69.9 (d, C-4‴), 60.6 (t, C-6‴).[20]

Daidzin (20a) (9.2 mg): brownish yellow powder. ESI-MS for $C_{21}H_{21}O_9$ $[M+H]^+$: 417.38. ^{1}H NMR (500 MHz, DMSO-d_6) δ: 8.38 (1H, s, H-2), 8.05 (1H, d, J = 8.8 Hz, H-5), 7.14 (1H, dd, J = 8.8, 2.3 Hz, H-6), 7.23 (1H, d, J = 2.3 Hz, H-8), 7.41 (2H, d, J = 8.2 Hz, H-2′, 6′), 6.82 (2H, d, J = 8.2 Hz, H-3′, 5′), 5.11 (1H, d, J = 6.9 Hz, H-1″), 3.18-3.74 (6H, sugar protons). ^{13}C NMR (125 MHz, DMSO-d_6) δ: 153.3 (d, C-2), 123.7 (s, C-3), 174.8 (s, C-4), 127.0 (d, C-5), 115.6 (d, C-6), 161.4 (s, C-7), 103.4 (d, C-8), 157.0 (s, C-9), 118.5 (s, C-10), 122.3 (s, C-1′), 130.1 (d, C-2′, 6′), 115.0 (d, C-3′, 5′), 157.3 (s, C-4′), 100.0 (d, C-1″), 73.2 (d, C-2″), 77.2 (d, C-3″), 69.7 (d, C-4″), 76.5 (d, C-5″), 60.7 (t, C-6″).[21]

Resveratrol 3-*O*-*β*-D-glucoside (25a) (7.5 mg): brownish yellow powder. ESI-MS for $C_{20}H_{22}O_8Na$ $[M+Na]^+$: 413.39. ^{1}H NMR (500 MHz, DMSO-d_6) δ: 6.73 (1H, s, H-2), 6.34 (1H, d, J = 2.1 Hz, H-4), 6.57 (1H, s, H-6), 6.86 (1H, d, J = 16.3 Hz, H-7), 7.03 (1H, d, J = 16.3 Hz, H-8), 7.40 (2H, d, J = 8.2 Hz, H-2′, 6′), 6.76 (2H, d, J = 8.2 Hz, H-3′, 5′), 4.80 (1H, d, J = 7.6 Hz, H-1″), 3.17-3.73 (6H, sugar protons). ^{13}C NMR (125 MHz, DMSO-d_6) δ: 139.4 (s, C-1), 102.8 (d, C-2), 158.9 (s, C-3), 104.8 (d, C-4), 158.4 (s, C-5), 107.2 (d, C-6), 125.2 (d, C-7), 128.0 (d, C-8), 128.6 (s, C-1′), 127.9 (d, C-2′, 6′), 115.5 (d, C-3′, 5′), 157.3 (s, C-4′), 100.7 (d, C-1″), 73.3 (d, C-2″), 77.1 (d, C-3″), 69.8 (d, C-4″), 76.1 (d, C-5″), 60.8 (t, C-6″).[22]

Resveratrol 3,4′-di-*O*-*β*-D-glucoside (25b) (8.7 mg): brownish yellow powder. ESI-MS for $C_{26}H_{32}O_{13}Na$ $[M+Na]^+$: 575.45. ^{1}H NMR (500 MHz, DMSO-d_6) δ: 6.76 (1H, s, H-2), 6.36 (1H, s, H-4), 6.59 (1H, s, H-6), 6.97 (1H, d, J = 16.3 Hz, H-7), 7.09 (1H, d, J = 16.3 Hz, H-8), 7.51 (2H, d, J = 8.5 Hz, H-2′, 6′), 7.02 (2H, d, J = 8.5 Hz, H-3′, 5′), 4.88 (1H, d, J = 7.4 Hz, H-1″), 4.80 (1H, d, J = 7.6 Hz, H-1‴), 3.18-3.72 (12H, sugar protons). ^{13}C NMR (125 MHz, DMSO-d_6) δ: 139.4 (s, C-1), 103.1 (d, C-2), 158.9 (s, C-3), 105.0 (d, C-4), 158.4 (s, C-5), 107.4 (d, C-6), 126.8 (d, C-7), 128.1 (d, C-8), 130.8 (s, C-1′), 127.7 (d, C-2′, 6′), 116.4 (d, C-3′, 5′), 157.1 (s, C-4′), 100.7 (d, C-1″), 73.3 (d, C-2″), 77.2 (d, C-3″), 69.8 (d, C-4″), 76.7 (d, C-5″), 60.8 (t, C-6″), 100.3 (d, C-1‴), 73.3 (d, C-2‴), 77.1 (d, C-3‴), 69.8 (d, C-4‴), 76.7 (d, C-5‴), 60.7 (t, C-6‴).[23]

1.10 Gene expression and real-time quantitative PCR (RT-qPCR) analysis

To determine the expression levels of four putative *RyUGTs* (*RyUGT3A*, *RyUGT3B*, *RyUGT11* and *RyUGT12*) involved in the anthraquinone biosynthesis in different tissues (roots, stems and leaves) of *R. yunnanensis*, RT-qPCR analysis was conducted on a StepOne™ Real-time PCR instrument (Thermo Fisher Scientific, USA) using SYBR Green PCR Master Mix (Vazyme, Nanjing, China). The mean values of three replicates was normalized with *hnRNP* gene based on our previous report.[2] RT-qPCR analysis was conducted with three biological and technical replicates and all primers used for RT-qPCR were listed in Table S7. The $2^{-\Delta\Delta Ct}$ method was used to calculate their relative expression levels.[24] The way to determine whether the designed primer pairs are specific is to refer to our previous work (Figure S37).[2] In order to better explain the specificity and applicability of the designed primers for RT-qPCR, four *RyUGTs* genes were firstly normalized for the expression levels of root and stem leaf with the same RNA templates as that applied for *de novo* transcriptome sequencing of root and stem leaf of *R. yunnanensis* (Figure S38). To explore the possible function of these four putative *RyUGT* genes in *R. yunnanensis*, one-year old *R. yunnanensis* plants were treated with 200 μM MeJA according to the method reported.[25] Then, the treated samples (roots, stems and leaves) and the untreated control samples were harvested for analysis at 1, 6, 12 and 24 h after elicitation, respectively. In addition, hairy roots of *R. yunnanensis* established by our laboratory were used to analyze the changes in this main anthraquinone aglycone (**1**) and its three corresponding anthraquinone glycosides (**54**, **55**, and **56**, Figure S39) after MeJA treatment. Specifically, 0.3 g hair roots were cultured in 100 mL ½MS liquid medium in an orbital shaker at 25 ℃ in dark. MeJA (100 μM) was supplemented to the growing hairy roots culture at day 45, and the hairy roots were harvested at day 59. The contents of **1**, **54**, **55**, and **56** were quantified via LC-MS/MS analysis with the MRM method established by our lab (unpublished). The controls of the above experiments were without MeJA treatment. The experiments were repeated three times.

1.11 Subcellular localization

The recombined pCAMBIA1302-RyUGT3A and -RyUGT12 vectors were constructed using the specific primers listed in Table S8. Specifically, the ORF of each *RyUGT* gene were inserted into pCAMBIA1302 at the 5'-terminal of the *GFP* gene under the control of the CaMV 35S promoter. The *Agrobacterium tumefaciens* strain GV3101 containing the recombined vectors was obtained based on a freeze-thaw protocol and transient transformation of tobacco (*Nicotiana benthamiana*) epidermal cells were performed according to Sparkes et al.[26,27] After 36 h infiltration with the transgenic tobacco, the infected areas were cut and then torn off the epidermis for imaging with a Zeiss LSM700 confocal laser scanning microscope (Zeiss, Germany). Meanwhile, the empty vector pCAMBIA1302 was infiltrated into tobacco leaves as control. Subcellular localization was further confirmed with three replications.

2 Supplemental Tables

Table S1. Information of 31 target GT unigenes with significantly differential expressions used in this study

Gene Name	Gene ID	R1_FPKM	R2_FPKM	R3_FPKM	SL1_FPKM	SL2_FPKM	SL3_FPKM	P Value
RyUGT1	TRINITY_DN128349_c5_g1	1.07	4.56	1.77	5.68	19.93	6.44	0.00932349
RyUGT2	TRINITY_DN129196_c0_g3	16.79	1.64	55.75	1.04	13.31	1.66	4.03E-05
RyUGT3	TRINITY_DN123143_c0_g2	583.47	666.89	494.28	83.04	137.99	179.47	0.000983536
RyUGT4	TRINITY_DN127830_c1_g1	16.04	17.14	11.38	7.52	3.27	2.07	0.008577635
RyUGT5	TRINITY_DN129704_c0_g1	10.77	5.75	2.9	0.64	0.1	0.62	3.21E-06
RyUGT6	TRINITY_DN129045_c0_g3	59.97	54.75	31.95	0.68	0	0.46	6.05E-24
RyUGT8	TRINITY_DN128241_c3_g1	5.02	5.35	1.52	9.49	25.34	62.01	0.0002304
RyUGT9	TRINITY_DN127760_c1_g2	263.2	291.4	281.29	3.24	16.3	35.1	5.18E-11
RyUGT10	TRINITY_DN128133_c4_g1	175.2	142.08	173.2	48	56.89	134.7	0.010716415
RyUGT11	TRINITY_DN131159_c3_g3	322.1	230.76	335.54	22.19	47.86	42.13	2.10E-08
RyUGT12	TRINITY_DN131626_c3_g1	155.1	233.5	133.87	0.47	1.57	0.09	6.29E-34
RyUGT13	TRINITY_DN131750_c7_g2	240.6	282.67	174.86	0	0.25	0.08	2.28E-56
RyUGT14	TRINITY_DN122510_c0_g1	36.61	108.55	50.03	0.82	3.59	2.38	1.57E-17

RyUGT15	TRINITY_DN122681_c7_g3	28.81	30.6	15.91	2.44	3.08	2.16	1.09E-12
RyUGT16	TRINITY_DN129109_c3_g5	28.6	31.66	44.67	5.33	1.83	6.55	5.38E-09
RyUGT18	TRINITY_DN124382_c1_g1	227.4	237.99	151.42	2.21	1.4	1.31	4.81E-38
RyUGT19	TRINITY_DN131204_c1_g4	73.65	92.69	91.7	12.19	9.12	19.57	9.09E-09
RyUGT20	TRINITY_DN131204_c1_g1	184.1	112.86	301.63	81.04	30.66	129.44	0.005131196
RyUGT21	TRINITY_DN128133_c4_g8	397.6	450.37	309.81	3.63	16.25	18.16	2.02E-20
RyUGT22	TRINITY_DN131492_c0_g9	108.9	362.3	450.9	51.28	116.44	102.73	3.57E-05
RyUGT23	TRINITY_DN126735_c3_g3	142.3	120.11	131.19	1.62	1.6	3.03	1.70E-35
RyUGT24	TRINITY_DN130359_c1_g11	207.8	165.97	271.04	4.16	0.84	2.12	1.67E-25
RyUGT26	TRINITY_DN123452_c1_g1	28.39	23.59	20.19	2.11	9.71	3.24	4.42E-05
RyUGT27	TRINITY_DN128407_c1_g1	11.48	15.52	18.77	1.36	1.91	3.84	1.61E-07
RyUGT28	TRINITY_DN128133_c4_g9	15.51	23.6	41	9.35	15.42	10.69	0.008183448
RyUGT29	TRINITY_DN123424_c2_g7	11.67	24.62	21.36	7.8	12.17	11.53	0.009155941
RyUGT32	TRINITY_DN130075_c1_g1	172.2	192.41	94.04	0	0	0.06	4.29E-54
RyUGT33	TRINITY_DN131897_c2_g2	22.07	91.81	104.75	789.32	653.43	556.28	2.53E-07
RyUGT34	TRINITY_DN130529_c0_g5	115.3	129.75	80.54	88.69	249.93	107.91	0.767266167

RyUGT35	TRINITY_DN130201_c2_g9	2.12	2.27	0.41	36.14	23.01	22.61	3.46E-09
RyUGT36	TRINITY_DN123035_c10_g2	56.09	62.93	73.35	73.42	98.8	90.22	0.685501246

Note: Gene ID: the assembled gene ID; FPKM: the relative expression of the gene in root (R1, R2 and R3), stem and leaf samples (SL1, SL2 and SL3) of *R. yunnanensis* under three biological duplication; P Value: the lower the value, the more significant the difference of gene expression.

Table S2. Primers used for 5'-RACE and 3'-RACE

Primer name	5'→3'
RyUGT1-5'-GSP1	GATTACGCCAAGCTTGCTCCTTCCGCTTTCATTCAATCCCTC
RyUGT2-3'-GSP1	GATTACGCCAAGCTTAGGGATGAAGTTGAGGGTTTGGTGAGG
RyUGT2-5'-GSP1	GATTACGCCAAGCTTCCTCCACTCCAAAGCCTTCTCCCTCAT
RyUGT3-3'-GSP1	GATTACGCCAAGCTTGCCGTCACATCCACCATCATCACAA
RyUGT3-5'-GSP1	GATTACGCCAAGCTTGAGGCAATCTGGGCGGAGTTCTTCT
RyUGT3-3'-GSP2	GATTACGCCAAGCTTCCAAATCTTCTCCAACCTTCAAAAACGC
RyUGT3-5'-GSP2	GAGGCAATCTGGGCGGAGTTCTTCT
RyUGT10-3'-GSP1	GATTACGCCAAGCTTGACAGCGAGAAATGGTTCCCGAAGG
RyUGT10-5'-GSP1	GATTACGCCAAGCTTCCCTTTCAGAACCTCCACATTCACCCTA
RyUGT11-3'-GSP1	GATTACGCCAAGCTTCGCCTCAGGTTGCGATTCTTTCCC
RyUGT11-5'-GSP1	GATTACGCCAAGCTTTCCAGCCACAGTGCGAAACAAAGC
RyUGT12-3'-GSP1	GATTACGCCAAGCTTGCAAATCGCCTACGCACTTCAGACC
RyUGT12-5'-GSP1	GATTACGCCAAGCTTGTCAAGTCGATACTCGCAAACGTGCC
RyUGT13-3'-GSP1	GATTACGCCAAGCTTCCACCTGTTTACCCAATCGGCCCTCT
RyUGT18-5'-GSP1	GATTACGCCAAGCTTCCGAAATTGACCAAGACAACCGATGA
RyUGT19-3'-GSP1	GATTACGCCAAGCTTGGCCAAATCCGCAGCTCCATCA
RyUGT19-5'-GSP1	GATTACGCCAAGCTTGCACCGTGATCTTCTCGCTTTCGTC
RyUGT21-3'-GSP1	GATTACGCCAAGCTTTCATCCAATGGCTAGATTCCAAACCTCA
RyUGT21-5'-GSP1	GATTACGCCAAGCTTCCCTTTCCGTTTGCTTTGATTCTGTCC
RyUGT22-3'-GSP1	GATTACGCCAAGCTTAACTGGAGCCTCCGGTGACGTGTATT
RyUGT24-3'-GSP1	GATTACGCCAAGCTTGGAGACGAAAGGGCTCATCATCAACAC
RyUGT24-5'-GSP1	GATTACGCCAAGCTTGATGATGAGCCCTTTCGTCTCCCTGTAC

RyUGT26-3'-GSP1	GATTACGCCAAGCTTCCTCGTACAGGGGATAGAATTGTGGAGC
RyUGT27-5'-GSP1	GATTACGCCAAGCTTGCTCCAAACCGCAGGCTAGTTCAGTG
RyUGT28-3'-GSP1	GATTACGCCAAGCTTTGAGGGCGGAATCAGAGAAGTGATGC

Table S3. Putative GT genes cloned from *R. yunnanensis*

No.	Name/Family	Accession No.	No.	Name/Family	Accession No.
RyUGT1	UGT75M2	MT075677	RyUGT18	UGT94AM1	MT075693
RyUGT2	UGT85A94	MT075678	RyUGT19	UGT84A63	MT075694
RyUGT3A	UGT73A27	MT075679	RyUGT20	UGT84A64	MT075695
RyUGT3B	UGT73A28	MT075680	RyUGT21	UGT73A30	MT075696
RyUGT4	UGT76AB1	MT075681	RyUGT22	UGT85AK1	MT075697
RyUGT5	UGT91AC1	MT075682	RyUGT23	UGT71AS4	MT075698
RyUGT6	UGT91AD1	MT075683	RyUGT24	UGT71AS1	MT075687
RyUGT8	UGT75Y1	MT075684	RyUGT26	UGT709T1	MT075699
RyUGT9	UGT71AR1	MT075685	RyUGT27	UGT72AL3	MT075700
RyUGT10	UGT73A29	MT075686	RyUGT28	UGT71AR2	MT075701
RyUGT11	UGT71AS1	MT075687	RyUGT29	UGT79A14	MT075702
RyUGT12	UGT71AS2	MT075688	RyUGT32	UGT71AS5	MT075703
RyUGT13	UGT71AS3	MT075689	RyUGT33	UGT88A31	MT075704
RyUGT14	UGT86A19	MT075690	RyUGT34	UGT72B52	MT075705
RyUGT15	*None	MT075691	RyUGT35	UGT85K33	MT075706
RyUGT16	*None	MT075692	RyUGT36	UGT87AA1	MT075707

Note: "*" indicated that are unable to assign name since it does not align with other named glycosyltransferases, which might be another different transferase superfamily.

Table S4. Gene-specific primers, restriction sites corresponding to pET-28a(+), endonucleases used and protein expression

Primer name	Primer pairs (5'→3') (Containing homologous sequences from pET-28a(+) vectors)	Endonuclease	Protein expression or not	kDa
RyUGT1	UGT1-F: AATGGGTCGCGGATCCGAATTCATGGTGGAAAAGCAGCACC UGT1-R: CTCGAGTGCGGCCGCAAGCTTGTTATGATCCCAGATTCTGTAAAAAG	*Eco*R I/*Hin*d III	Yes	51.54
RyUGT2	UGT2-F: ACAGCAAATGGGTCGCGGATCCATGGATGCACCTGATCATCAGC UGT2-R: CTCGAGTGCGGCCGCAAGCTTGCTACTGCATGATTGCTTCTATAAACTTGTC	*Bam*H I/*Hin*d III	Yes	54.20
RyUGT3A	UGT3A-F: AATGGGTCGCGGATCCGAATTCATGGGCCGGAAGCAGCTG UGT3A-R: CTCGAGTGCGGCCGCAAGCTTGTCAATTATTTGAATGGTATGCACTCAATTCTT	*Eco*R I/*Hin*d III	Yes	54.45
RyUGT3B	UGT3B-F: ACAGCAAATGGGTCGCGGATCCATGGGGCGGCAGCAGCTG UGT3B-R: TGGTGCTCGAGTGCGGCCGCATCAACGGTACGCACTCAATTCC	*Bam*H I/*Not* I	Yes	53.51
RyUGT4	UGT4-F: AATGGGTCGCGGATCCGAATTCATGGCAAAACCCAGGGCAC UGT4-R: CTCGAGTGCGGCCGCAAGCTTGTCACGGGAATGAGCAGATAAAATCTG	*Eco*R I/*Hin*d III	inclusion body	47.63
RyUGT5	UGT5-F: AATGGGTCGCGGATCCGAATTCATGGCTACTGAAACTAAGAAGCATCA UGT5-R: TGGTGCTCGAGTGCGGCCGCATTAAACGGGACGTTTGAATTTCTCAA	*Eco*R I/*Not* I	inclusion body	51.50
RyUGT6	UGT6-F: AATGGGTCGCGGATCCGAATTCATGGAGAGCAAAACTGATCAAATCCA UGT6-R: CTCGAGTGCGGCCGCAAGCTTGCTAGCATAACTTTTTCACCCCATTGA	*Eco*R I/*Hin*d III	Yes	50.94
RyUGT8	UGT8-F: AATGGGTCGCGGATCCGAATTCATGGAAAAATGCCATTTTCTCATCGT UGT8-R: CTCGAGTGCGGCCGCAAGCTTGCTAAGTCAAAAGCAAAGAATTATTGACACA	*Eco*R I/*Hin*d III	inclusion body	52.88

RyUGT9	UGT9-F: AATGGGTCGCGGATCCGAATTCATGCTCGCAAAAGGGCACTT UGT9-R: CTCGAGTGCGGCCGCAAGCTTGCTAGTTTGAATATTCTTCAATGAAGCGC	*Eco*R I/*Hin*d III	Yes	50.69
RyUGT11	UGT11-F: AATGGGTCGCGGATCCGAATTCATGAAGAAAGCCGAGCTGGT UGT11-R: CTCGAGTGCGGCCGCAAGCTTGTCAAGGGATACTGTCGATTACATTGTCA	*Eco*R I/*Hin*d III	Yes	52.64
RyUGT12	UGT12-F: ACAGCAAATGGGTCGCGGATCCATGGGAAAAGATACGAAGAATGCAGA UGT12-R: TGGTGCTCGAGTGCGGCCGCATCATGAACTGATGTTATCCAACGCA	*Bam*H I/*Not* I	Yes	52.73
RyUGT14	UGT14-F: ACAGCAAATGGGTCGCGGATCCATGGCGGGCACCAACC UGT14-R: TGGTGCTCGAGTGCGGCCGCACTAATGGCCATTAGCCGTTATCATCC	*Bam*H I/*Not* I	Yes	54.67
RyUGT15	UGT15-F: AATGGGTCGCGGATCCGAATTCATGCGGGGTTCTCATCATCA UGT15-R: CTCGAGTGCGGCCGCAAGCTTGTTAGCCATCCTTCTGCTCCG	*Eco*R I/*Hin*d III	Yes	52.33
RyUGT16	UGT16-F: AATGGGTCGCGGATCCGAATTCATGTACTCCAAGAATAGAACTCACGG UGT16-R: CTCGAGTGCGGCCGCAAGCTTGTTAAGCTAATAATTCCTCAGATTCGCCG	*Eco*R I/*Hin*d III	No	50.57
RyUGT18	UGT18-F: AATGGGTCGCGGATCCGAATTCATGGAGAAGCTTGAAAATTCCTCT UGT18-R: CTCGAGTGCGGCCGCAAGCTTGCTAGTTTTTTTCCCTCATATCATTCCCA	*Eco*R I/*Hin*d III	Yes	47.60
RyUGT19	UGT19-F: AATGGGTCGCGGATCCGAATTCATGGTCGGCCAAATCCGC UGT19-R: TGGTGCTCGAGTGCGGCCGCACTACCCTAAAGTAACAGTTTGAATGTCA	*Eco*R I/*Not* I	Yes	52.09
RyUGT20	UGT20-F: AATGGGTCGCGGATCCGAATTCATGGTGGCCCAAACCTACGA UGT20-R: CTCGAGTGCGGCCGCAAGCTTGCTACCCTAAAGTAACAGTTTGAATGTCA	*Eco*R I/*Hin*d III	No	51.96
RyUGT22	UGT22-F: AATGGGTCGCGGATCCGAATTCATGGCGTCCTCCCCGC UGT22-R: TGGTGCTCGAGTGCGGCCGCATTATGGGGCACATGAGGTCTTT	*Eco*R I/*Not* I	Yes	52.73

RyUGT23	UGT23-F: AATGGGTCGCGGATCCGAATTCATGAAGAGAGCAGAGCTGGT UGT23-R: CTCGAGTGCGGCCGCAAGCTTGTCAAGAAGGGATGTTGTCCATAACG	*Eco*R I/*Hin*d III	inclusion body	53.33
RyUGT26	UGT26-F: ACAGCAAATGGGTCGCGGATCCATGGGTCGCCATCCTCAC UGT26-R: CTCGAGTGCGGCCGCAAGCTTGCTAATCACTGAATCTGCCGAATTTGA	*Bam*H I/*Hin*d III	Yes	53.26
RyUGT27	UGT27-F: ACAGCAAATGGGTCGCGGATCCATGACAACCTCAGCGGAGTT UGT27-R: TGGTGCTCGAGTGCGGCCGCATCAGAATAATATATTCGTCGGGTTCAGC	*Bam*H I/*Not* I	No	48.85
RyUGT28	UGT28-F: AATGGGTCGCGGATCCGAATTCATGAAAACCGAGCTGGTTTTCATCCC UGT28-R: CTCGAGTGCGGCCGCAAGCTTGTCAATCACGAATGTCTTCAATGAACCGC	*Eco*R I/*Hin*d III	No	51.26
RyUGT29	UGT29-F: ACAGCAAATGGGTCGCGGATCCATGACAACCTCAGCGGAGT UGT29-R: TGGTGCTCGAGTGCGGCCGCATCAGAATAATATATTCGTCGGGTTCAGC	*Bam*H I/*Not* I	inclusion body	52.45
RyUGT32	UGT32-F: ACAGCAAATGGGTCGCGGATCCATGGAGAAGCTCGAGCTGG UGT32-R: TGGTGCTCGAGTGCGGCCGCATCACAAGGTAGAAGATGAAGAAGAGATG	*Bam*H I/*Not* I	Yes	53.75
RyUGT33	UGT33-F: AATGGGTCGCGGATCCGAATTCATGGCGACCGTAGTCCTCTA UGT33-R: CTCGAGTGCGGCCGCAAGCTTGCTATGATATCTGTTTCCATGAATCCACC	*Eco*R I/*Hin*d III	No	51.50
RyUGT34	UGT34-F: AATGGGTCGCGGATCCGAATTCATGGCCCACGCGCCGC UGT34-R: CTCGAGTGCGGCCGCAAGCTTGTCATGCAACACCGACTTTACTCTTCC	*Eco*R I/*Hin*d III	No	51.67
RyUGT35	UGT35-F: AATGGGTCGCGGATCCGAATTCATGGGTTTTCTCCAGAATATTCCCA UGT35-R: CTCGAGTGCGGCCGCAAGCTTGTCATCCCTTCATGATGTAATGAATGAAG	*Eco*R I/*Hin*d III	Yes	54.72
RyUGT36	UGT36-F: AATGGGTCGCGGATCCGAATTCATGGCTTCATCAAATTCCATCACT UGT36-R: CTCGAGTGCGGCCGCAAGCTTGCTAATTGGCAATGTCTTTGATGAAAG	*Eco*R I/*Hin*d III	No	50.55

Table S5. LC-MS data summary of glycosylated compounds

No.	Compound name	Detected Substrate Mass (m/z)	Product Retention Time (min)		Detected Product Mass (m/z)		Product Type		Yield (%)	
			by RyUGT3A	by RyUGT12	by RyUGT3A	by RyUGT12	by RyUGT3A	by RyUGT12	by RyUGT3A	by RyUGT12
1	6-Hydroxyalizarin	269.43[M-H]$^-$	27.294	27.294	431.31[M-H]$^-$	431.31[M-H]$^-$	Mono-	Mono-	78	9
			28.111	28.111	431.31[M-H]$^-$	431.31[M-H]$^-$	Mono-	Mono-	17	88
			15.109	15.109	639.15 [M+HCOOH-H]$^-$	639.15 [M+HCOOH-H]$^-$	Di-	Di-	5	1.5
2	2-Methyl-3-hydroxy -9,10-anthraquinone	269.30[M-H]$^-$	27.273	27.273	551.33[M-H]$^-$	551.33[M-H]$^-$	Mono-	Mono-	50	41
3	3,6-Dihydroxy-xanthen-9-one	227.41[M-H]$^-$	13.277	13.277	389.46[M-H]$^-$	389.46[M-H]$^-$	Mono-	Mono-	5	46
			18.83	N.D.	551.15[M-H]$^-$	N.D.	Di-	-	94	-
4	Emodin	269.25[M-H]$^-$	28.87	28.87	431.31[M-H]$^-$	431.31[M-H]$^-$	Mono-	Mono-	98	85
5	2-Hydroxy-9,10-anthraquinone	223.26[M-H]$^-$	24.482	24.482	431.17 [M+HCOOH-H]$^-$	431.17 [M+HCOOH-H]$^-$	Mono-	Mono-	78	42

6	2-Amino-3-hydroxy-9,10-anthraquinone	238.29[M-H]$^-$	23.315	23.315	400.35[M-H]$^-$	400.35[M-H]$^-$	Mono-	Mono-	77	60
7	1,3-Dihydroxy-9,10-anthraquinone	239.09[M-H]$^-$	27.018	27.018	447.18 [M+HCOOH-H]$^-$	447.18 [M+HCOOH-H]$^-$	Mono-	Mono-	79	65
8	2,6-Dihydroxy-9,10-anthraquinone	239.24[M-H]$^-$	16.266	16.266	401.34[M-H]$^-$	401.34[M-H]$^-$	Mono-	Mono-	65	35
9	Aloe-emodin	269.30[M-H]$^-$	27.705	N.D.	431.40[M-H]$^-$	N.D.	Mono-	-	25	-
10	1-Hydroxy-2-hydroxymethyl-9,10-anthraquinone	253.28[M-H]$^-$	27.195	N.D.	415.33[M-H]$^-$	N.D.	Mono-	-	5	-
11	Purpurin	255.38[M-H]$^-$	32.4	32.4	417.35[M-H]$^-$	417.35[M-H]$^-$	Mono-	Mono-	85	84
12	2,6-Diamino-9,10-anthraquinone	239.42[M+H]$^-$	13.747	13.747	423.10[M+Na]$^+$	423.10[M+Na]$^+$	Mono-	Mono-	88	40
13	Apigenin	269.30[M-H]$^-$	20.914	20.914	431.35[M-H]$^-$	431.35[M-H]$^-$	Mono-	Mono-	95	43
			21.876	N.D.	431.35[M-H]$^-$	N.D.	Mono-	-	4	-
14	Baicalein	269.30[M-H]$^-$	24.755	24.755	431.35[M-H]$^-$	431.35[M-H]$^-$	Mono-	Mono-	90	68
			27.973	27.973	431.35[M-H]$^-$	431.35[M-H]$^-$	Mono-	Mono-	9	-

15	Kaempferol	285.32[M-H]$^-$	20.267	20.267	447.33[M-H]$^-$	447.33[M-H]$^-$	Mono-	Mono-	51	27
			20.972	20.972	447.33[M-H]$^-$	447.33[M-H]$^-$	Mono-	Mono-	2	36
			9.89	N.D.	609.34[M-H]$^-$	N.D.	Di-	-	46	-
16	Luteolin	285.27[M-H]$^-$	16.248	16.248	447.28[M-H]$^-$	447.28[M-H]$^-$	Mono-	Mono-	19	78
			21.435	N.D.	447.28[M-H]$^-$	N.D.	Mono-	-	55	-
17	Quercetin	301.25[M-H]$^-$	16.805	16.805	463.26[M-H]$^-$	463.26[M-H]$^-$	Mono-	Mono-	38	25
			15.838	15.838	625.4[M-H]$^-$	625.4[M-H]$^-$	Di-	Di-	60	72
18	Myricetin	317.37[M-H]$^-$	13.524	13.524	479.54[M-H]$^-$	479.54[M-H]$^-$	Mono-	Mono-	51	47
			18.432	18.432	479.54[M-H]$^-$	479.54[M-H]$^-$	Mono-	Mono-	28	5
			20.034	20.03	479.54[M-H]$^-$	479.54[M-H]$^-$	Mono-	Mono-	14	17
19	Naringenin	271.32[M-H]$^-$	20.944	20.944	433.38[M-H]$^-$	433.38[M-H]$^-$	Mono-	Mono-	79	59
			22.08	N.D.	433.38[M-H]$^-$	N.D.	Mono-	-	19	-
20	Daidzein	253.28[M-H]$^-$	12.941	N.D.	415.33[M-H]$^-$	N.D.	Mono-	-	98	-
21	Phloretin	273.25[M-H]$^-$	23.26	23.26	435.45[M-H]$^-$	435.45[M-H]$^-$	Mono-	Mono-	8.4	5
			25.796	25.796	435.45[M-H]$^-$	435.45[M-H]$^-$	Mono-	Mono-	3.7	47

			28.166	28.166	435.45[M-H]$^-$	435.45[M-H]$^-$	Mono-	Mono-	76	3
22	Butein	271.32[M-H]$^-$	25.663	25.663	433.38[M-H]$^-$	433.38[M-H]$^-$	Mono-	Mono-	30	91.5
			19.453	N.D.	433.38[M-H]$^-$	N.D.	Mono-	-	13	-
			13.354	N.D.	595.38[M-H]$^-$	N.D.	Di-	-	49	-
23	Hematoxylin	301.44[M-H]$^-$	7.47	N.D.	463.45[M-H]$^-$	N.D.	Mono-	-.	36	-.
24	Silibinin	481.45[M-H]$^-$	13.938	13.938	643.39[M-H]$^-$	643.39[M-H]$^-$	Mono-	Mono-	62	13
			11.5	11.5	643.39[M-H]$^-$	643.39[M-H]$^-$	Mono-	Mono-	34	9
			11.15	11.15	643.39[M-H]$^-$	643.39[M-H]$^-$	Mono-	Mono-	3	77
25	Resveratrol	227.31[M-H]$^-$	16.504	16.504	389.32[M-H]$^-$	389.32[M-H]$^-$	Mono-	Mono-	13	14
			13.293	13.293	551.33[M-H]$^-$	551.33[M-H]$^-$	Di-	Mono-	63	61
26	Bis-demethoxycurcumin	307.37[M-H]$^-$	23.731	23.731	469.38[M-H]$^-$	469.38[M-H]$^-$	Mono-	Mono-	18	52
			28.916	28.916	631.43[M-H]$^-$	631.43[M-H]$^-$	Di-	Di-	72	4
27	Magnolol	265.43[M-H]$^-$	30.203	30.203	427.12[M-H]$^-$	427.12[M-H]$^-$	Mono-	Mono-	97	30
28	Chlorogenic acid	353.43[M-H]$^-$	9.139	9.139	515.46[M-H]$^-$	515.46[M-H]$^-$	Mono-	Mono-	35	33
29	Ferulic acid	193.44[M-H]$^-$	11.692	11.692	355.36[M-H]$^-$	355.36[M-H]$^-$	Mono-	Mono-	99	99
30	Paeonol	165.50[M-H]$^-$	12.028	12.028	373.42	373.42	Mono-	Mono-	91	96

					$[M+HCOOH-H]^-$	$[M+HCOOH-H]^-$				
31	3,4-Dichloroaniline	160.23$[M-H]^-$	11.378	11.378	322.37$[M-H]^-$	322.37$[M-H]^-$	Mono-	Mono-	83	36
32	3,4-Dichlorobenzenethiol	177.06$[M-H]^-$	11.927	11.927	385.28 $[M+HCOOH-H]^-$	385.28 $[M+HCOOH-H]^-$	Mono-	Mono-	99	70

Note: N.D. indicates no glycosylation products are detected.

Table S6. Primers used in site-directed mutagenesis

Primer name	5'→3'
RyUGT3A-GLY17A-F1	CACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCTAGCATGACTGGTGGACA
RyUGT3A-GLY17A-tbF	CTTCTTCTTCCCCATGTTAGCTCACGCCCACATGCTCCCAATCCTGGATATGGC
RyUGT3A-GLY17A-tbR	GCCATATCCAGGATTGGGAGCATGTGGGCGTGAGCTAACATGGGGAAGAAGAAG
RyUGT3A-GLY17A-R1	GCCGGATCTCAGTGGTGGTGGTGGTGGTGCTCGAGTGCGGCCGCAAGCTTTCAATTATTTGAATGG
RyUGT3A-HIP18A-F1	CACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCTAGCATGACTGGTGGACA
RyUGT3A-HIP18A-tbF	TCTTCCCCATGTTAGCTCACGGCGCCATGCTCCCAATCCTGGATATGGCA
RyUGT3A-HIP18A-tbR	TGCCATATCCAGGATTGGGAGCATGGCGCCGTGAGCTAACATGGGGAAGA
RyUGT3A-HIP18A-R1	GCCGGATCTCAGTGGTGGTGGTGGTGGTGCTCGAGTGCGGCCGCAAGCTTTCAATTATTTGAATGG
RyUGT3A-ARG192A-F1	CAGGTTTCGTTATACGAGGCTAAGAAGCTCGAAACAGAGACGGAGTTTAGCAA
RyUGT3A-ARG192A-R1	TTTCGAGCTTCTTAGCCTCGTATAACGAAACCTGGCCTCTCGTCA
RyUGT3A-SER291A-F1	GTGTACGTCTGTTTCGGAGCTATGGCGAACTTCACACGCGATCAGTTGCACGAG
RyUGT3A-SER291A-R1	TGAAGTTCGCCATAGCTCCGAAACAGACGTACACCACGGAGTTTTGAGGTTT
RyUGT3A-SER375A-F1	CGCATTGCGGGTGGAATGCTACGCTGGAGAGTGTTTCTTCCGGGGTCCCGATGATA
RyUGT3A-SER375A-R1	GAAACACTCTCCAGCGTAGCATTCCACCCGCAATGCGTCACGAATGCT

RyUGT3A-GLU378A-F1	GGAATTCGACGCTGGCTAGTGTTTCTTCCGGGGTCCCGATGATA
RyUGT3A-GLU378A-R1	ACCCCGGAAGAAACACTAGCCAGCGTCGAATTCCACCCGCAATGC
RyUGT3A-GLU394A-F1	CACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCTAGCATGACTGGTGGACA
RyUGT3A-GLU394A-tbF	TGATAACATGGCCGATGTTTGCGGCGCAGTTTTTGAATGAGAAATTGCT
RyUGT3A-GLU394A-tbR	AGCAATTTCTCATTCAAAAACTGCGCCGCAAACATCGGCCATGTTATCA
RyUGT3A-GLU394A-R1	GCCGGATCTCAGTGGTGGTGGTGGTGGTGCTCGAGTGCGGCCGCAAGCTTTCAATTATTTGAATGG
RyUGT3A-GLN395A-F1	CACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCTAGCATGACTGGTGGACA
RyUGT3A-GLN395A-tbF	ATAACATGGCCGATGTTTGCGGAGGCGTTTTTGAATGAGAAATTGCTTA
RyUGT3A-GLN395A-tbR	TAAGCAATTTCTCATTCAAAAACGCCTCCGCAAACATCGGCCATGTTAT
RyUGT3A-GLN395A-R1	GCCGGATCTCAGTGGTGGTGGTGGTGGTGCTCGAGTGCGGCCGCAAGCTTTCAATTATTTGAATGG
RyUGT12-GLY19A-F1	ATGGGAAAAGATACGAAGAATGCAGAGCTGGTTTTCATTCCCACTCCCGGAGCCGCCCACTTAACATCCACCATAGAA
RyUGT12-HIP20A-F1	ATGGGAAAAGATACGAAGAATGCAGAGCTGGTTTTCATTCCCACTCCCGGAGCCGGCGCCTTAACATCCACCATAGAAGTA
RyUGT12-SER367A-F1	GAAAGATGCCGCCGGCTTCCTCCCGGGGGACTACGACAACCTGGACG
RyUGT12-SER367A-tbF	GGTTCGTCTCGCACTGCGGCTGGGCTTCTACGCTGGAGAGCATCTGGT
RyUGT12-SER367A-tbR	ACCAGATGCTCTCCAGCGTAGAAGCCCAGCCGCAGTGCGAGACGAACC
RyUGT12-SER367A-R1	ATTAATTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTT

RyUGT12-GLU370AF1	GAAAGATGCCGCCGGCTTCCTCCCGGGGGACTACGACAACCTGGACG
RyUGT12-GLU370A-tbF	CTGCGGCTGGAATTCTACGCTGGCGAGCATCTGGTTCGGCGTTCCGA
RyUGT12-GLU370A-tbR	TCGGAACGCCGAACCAGATGCTCGCCAGCGTAGAATTCCAGCCGCAG
RyUGT12-GLU370A-R1	ATTAATTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTT
RyUGT12-TYR384A-F1	GAAAGATGCCGCCGGCTTCCTCCCGGGGGACTACGACAACCTGGACG
RyUGT12-TYR384A-tbF	TTCGTCTCGCACTGCGGCTGGAATGCTACGCTGGAGAGCATCTGGTTCGGC
RyUGT12-TYR384A-tbR	GCCGAACCAGATGCTCTCCAGCGTAGCATTCCAGCCGCAGTGCGAGACGAA
RyUGT12-TYR384A-R1	ATTAATTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTT
RyUGT12-GLU386A-F1	GAAAGATGCCGCCGGCTTCCTCCCGGGGGACTACGACAACCTGGACG
RyUGT12-GLU386A-tbF	AGCGTCGTGGCCGCAGTACGCGGCGCAGCAGACGAACGCGTTCTTTC
RyUGT12-GLU386A-tbR	GAAAGAACGCGTTCGTCTGCTGCGCCGCGTACTGCGGCCACGACGCT
RyUGT12-GLU386A-R1	ATTAATTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTT
RyUGT12-GLN387A-F1	GAAAGATGCCGCCGGCTTCCTCCCGGGGGACTACGACAACCTGGACG
RyUGT12-GLN387A-tbF	AGCGTCGTGGCCGCAGTACGCGGAGGCTCAGACGAACGCGTTCTTTCTGGTGA
RyUGT12-GLN387A-tbR	TCACCAGAAAGAACGCGTTCGTCTGAGCCTCCGCGTACTGCGGCCACGACGCT
RyUGT12-GLN387A-R1	ATTAATTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTT

RyUGT12-SER411A-F1	TTGATTATGTGAAGGCCGTCGGAACTGAGAGCACTGAGATTGTGAGCGC
RyUGT12-SER411A-R1	CTCAGTTCCGACGGCCTTCACATAATCAATCTTGATCTCCACCGCCA

Table S7. Primers used for RT-qPCR normalization

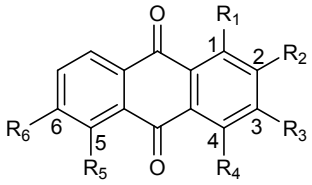
Gene name	Primer sequence (5′→3′)	Amplicon length (bp*)	Primers Tm* (℃)	E* (%)	R^2*
RyUGT3A	For: GACTCCGAACCGTTTATTATCCC Rev: CCTCTTTCTGCCTTATCTTCAACC	287	61.2/61.9	98.02	0.999
RyUGT3B	For: AAGCAAACGGAAAGGGACTAA Rev: CCACCCCAATCTTCAAAACG	210	58.7/60.2	90.27	0.994
RyUGT11	For: TTCCTCCGTTGTTTTCCTTTG Rev: TTTCGTAGTCTTCGGGGAGC	158	59.5/59.4	98.79	0.998
RyUGT12	For: AAATCGCCTACGCACTTCAG Rev: GCCGCTGTCCGAGTAAGAA	148	58.2/58.7	92.03	0.996

Note: * bp, Tm, E and R^2 indicate base pair, melting temperature, PCR efficiency, correlation coefficient, respectively. Specificity of primer pairs for RT-qPCR amplification was showed in Figure S37A and the melt curves with single peaks produced for all amplicons (Figure S37B). Standard curves of four target genes were shown in Figure S37C.

Table S8. Primers used for constructing pCAMBIA1302-RyUGT3A and -RyUGT12 vectors

Gene name	Primer sequence (5′→3′)
RyUGT3A	For:TTGGAGAGAACACGGGGGACTCTTGACCATGGTAGATCTATGGGCCGGAAGCAGCTGCACGT Rev:ACTCCAGTGAAAAGTTCTTCTCCTTTACTAGTATTATTTGAATGGTATGCACTCAA
RyUGT12	For:TTGGAGAGAACACGGGGGACTCTTGACCATGGTAGATCTATGGGAAAAGATACGAAGAATGCAGA Rev:ACTCCAGTGAAAAGTTCTTCTCCTTTACTAGTTGAACTGATGTTATCCAACGCAGT

3. Supplementary Figures



Order	Name	R_1	R_2	R_3	R_4	R_5	R_6
1	2-Hydroxymethyl-AQ	H	CH_2OH	H	H	H	H
2	2-Methoxycarbonyl-AQ	H	$COOCH_3$	H	H	H	H
3	2-Carbaldehyde-AQ	H	CHO	H	H	H	H
4	Alizarin	OH	OH	H	H	H	H
5	Xanthopurpurin	OH	H	OH	H	H	H
6	3-Hydroxy-2-hydroxymethyl-AQ	H	CH_2OH	OH	H	H	H
7	1,3-Dihydroxy-2-methyl-AQ (Rubiadin)	OH	CH_3	OH	H	H	H
8	Soranjidiol	OH	CH_3	H	H	H	OH
9	Anthragallol	OH	OH	OH	H	H	H
10	Anthragallol-2,3-dimethyl ether	OH	OCH_3	OCH_3	H	H	H
11	Anthragallol-3-methyl ether	OH	OH	OCH_3	H	H	H
12	1,4-Dihydroxy-2-methyl-AQ	OH	CH_3	H	OH	H	H
13	Pseudopurpurin	OH	OH	COOH	OH	H	H
14	1,3-Dihydroxy-6-methoxy-2-methyl-AQ	OH	CH_3	OH	H	H	OCH_3
15	1,3,6-Trihydroxy-2-methyl-AQ (6-Hydroxyrubiadin)	OH	CH_3	OH	H	H	OH
16	Rubianthraquinone	OCH_3	CH_3	OH	H	H	OH
17	Lucidin-3-*O*-*β*-D-glucopyranoside	OH	CH_2OH	OGlc	H	H	H
18	Rubiquinone-3-*O*-*β*-D-glucopyranoside	OH	CH_3	OGlc	H	H	OH

19	Rubiquinone-3-O-*β*-L-rhamnopyranosyl-(1→2)-*β*-D-glucopyranoside	OH	CH_3	-OGlc (2→1)Rha	H	H	OH
20	Rubiquinone-3-*O*-*β*-L-rhamnopyranosyl-(1→2)-(3'-*O*-acetyl)-*β*-D-glucopyranoside	OH	CH_3	(3'-OAc)-OGlc (2→1)Rha	H	H	OH
21	Rubiquinone-3-(6'-*O*-acetyl)-*O*-*β*-D-glucopyranoside	OH	CH_3	(6'-OAc)-OGlc	H	H	OH
22	Rubiquinone-3-*O*-*β*-D-xylopyranosyl-(1→2)-(6'-*O*-acetyl)-*β*-D-glucopyranoside	OH	CH_3	(6'-OAc)-OGlc (2→1)Xyl	H	H	OH
23	Purpurin-2-*O*-*β*-D-xylopyranosyl-(1→6)-*β*-D-glucopyranoside	OH	OGlc (6→1)Xyl	H	H	OH	H

AQ = 9,10-anthraquinone

Figure S1. Chemical structures of anthraquinones with high contents isolated from *R. yunnanensis.* Among them, the glycosylation sites of all anthraquinone glycosides occur in their *β*-hydroxyl groups.

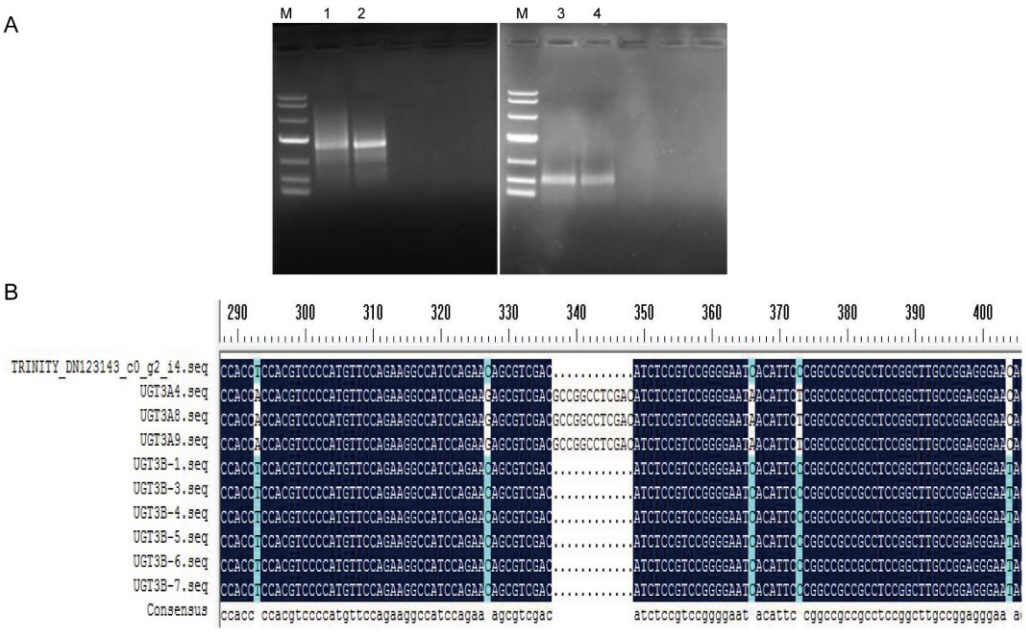


Figure S2. Results of 3'-RACE, 5'-RACE and their assemble and alignment of 'TRINITY_DN123143_c0_g2_i4'. (A) The results of 3'-RACE are displayed on the left and the right one is the 5'-RACE results. M: DL10000 Marker; 1-2: 3'-RACE by UGT3-3'-GSP1 and UGT3-3'-GSP2; 3-4: 5'-RACE by UGT3-5'-GSP1 and UGT3-5'-GSP2. (B) The assemble and alignment results of 3'-RACE and 5'-RACE of 'RyUGT3: TRINITY_DN123143_c0_g2_i4'.

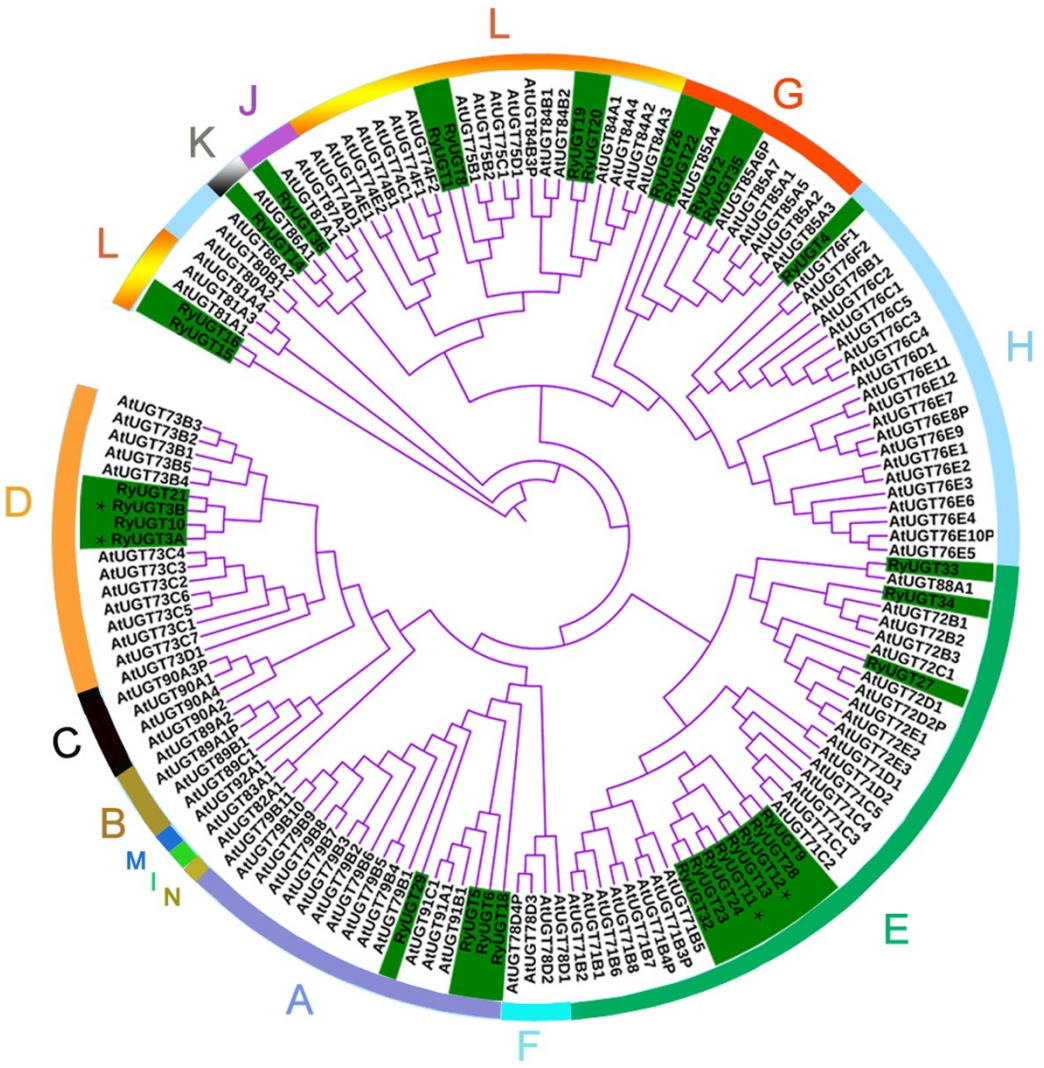


Figure S3. Phylogenetic analysis of 32 GTs from *R. yunnanensis* and 122 GTs from *A. thaliana.* The ClustalW program was employed to analyze amino acid sequences and the bootstrap consensus tree was constructed using MEGA 6.0 software with the neighbor-joining method and 1000 bootstrap replicates. RyUGTs from *R. yunnanensis* are highlighted in green and those from *A. thaliana* are in black. The four RyUGTs whose functions were characterized in this study were marked with stars. The 14 subgroups are indicated in different colors, and the 32 GTs were clustered into at least 12 subfamilies (UGT71, 72, 73, 75, 76, 79, 84, 85, 86, 87, 88, and 91) or 8 groups (A, D, E, G, H, J, K, and L).

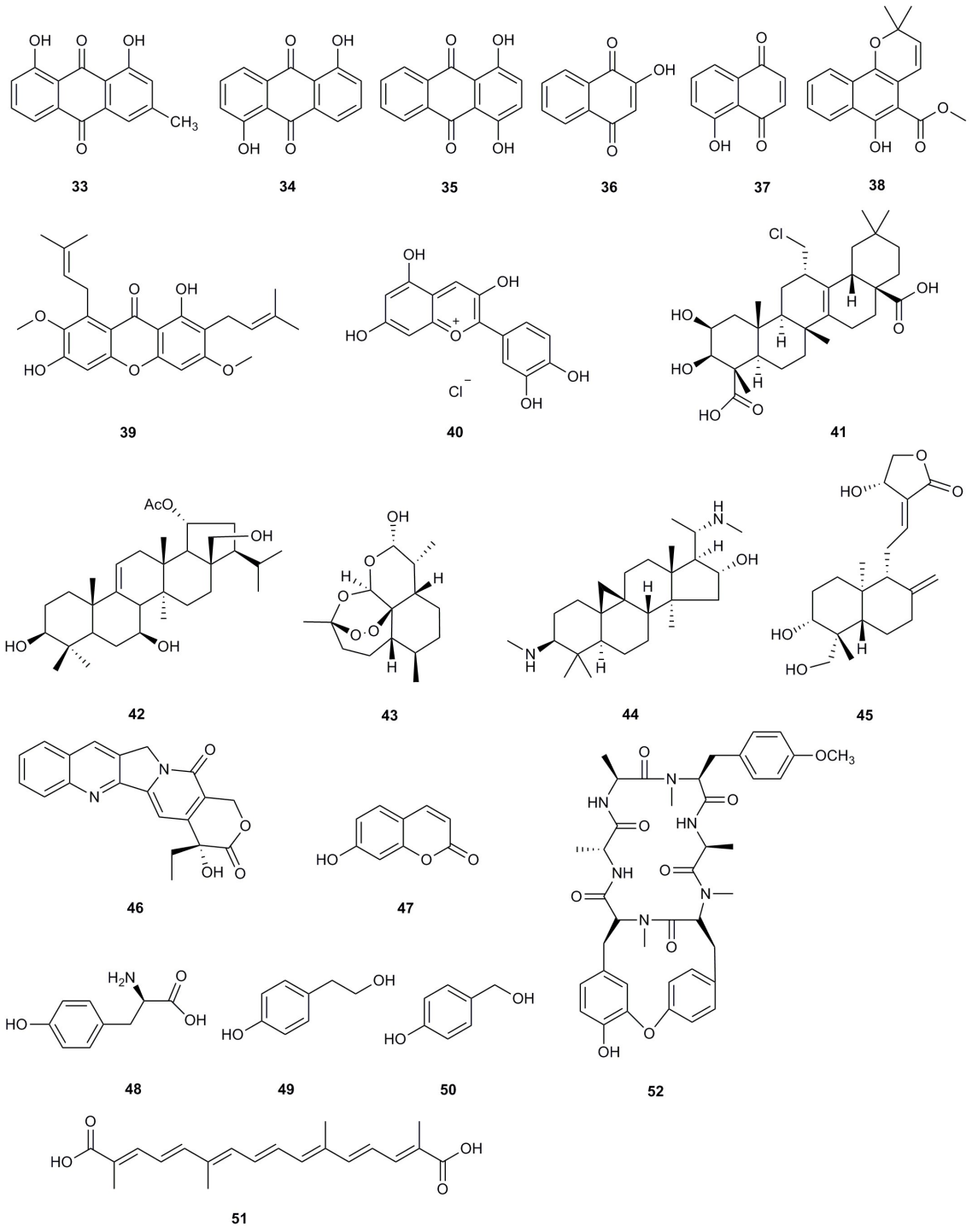


Figure S4. Compounds **33**−**52** are not transglucosylized by RyUGT3A and RyUGT12.

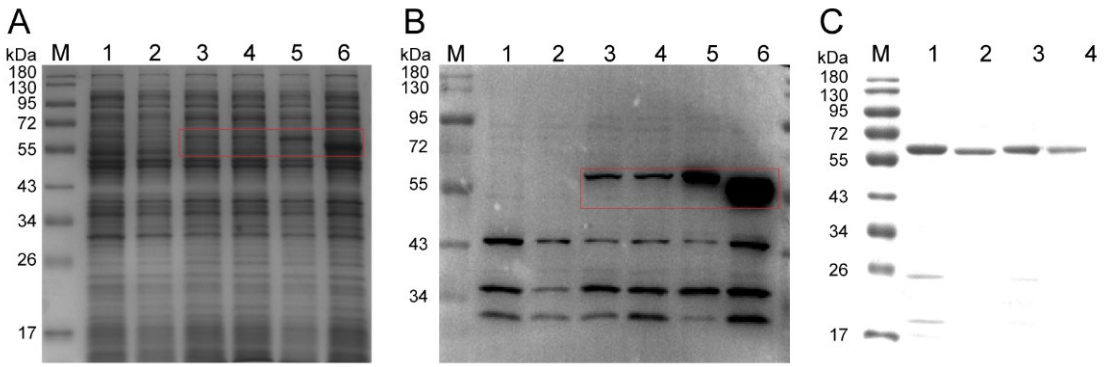


Figure S5. SDS-PAGE and Western Blotting detection of soluble expressions of the four His-tagged RyUGTs. (A and B): M indicated the protein Markers; Lane 1-6: crude protein extract; among them, Lane 1 and 2 indicated *E. coli* BL21(DE3) and the strain containing empty vector, respectively. Lane 3-6 indicated the strains containing recombinant plasmid pET-28a(+)-RyUGT11, pET-28a(+)-RyUGT12, pET-28a(+)-RyUGT3A and pET-28a(+)-RyUGT3B, respectively. (C): SDS-PAGE of the four His-tagged RyUGTs purified by Ni-NTA affinity chromatography. Lane 1-4 indicated the purified recombinant protein RyUGT3A, RyUGT3B, RyUGT11 and RyUGT12, respectively.

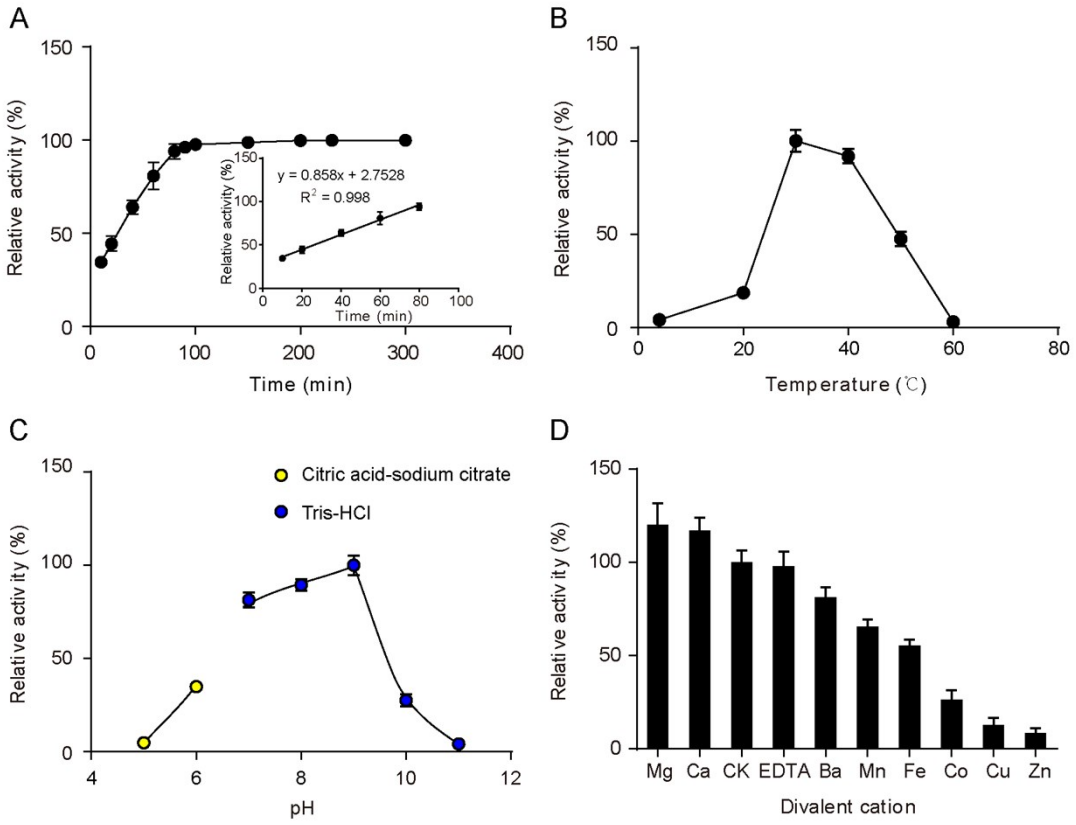


Figure S6. Effects of reaction time (A), temperature (B), reaction buffer (C), and divalent metal ions (D) on the activities of RyUGT3A. 6-Hydroxyrubiadin (**1**) was used as the acceptor and UDP-Glc was used as the sugar donor. An optimized reaction time of 60 min was used for (B), (C) and (D). Experiment with 5 mM EDTA or Tris-HCl (pH 7.5, control blank (CK)) was performed as a negative control in (D). RyUGT3A exhibited its maximum activity at pH 9.0 (50 mM Tris-HCl buffer) and 30 ℃.

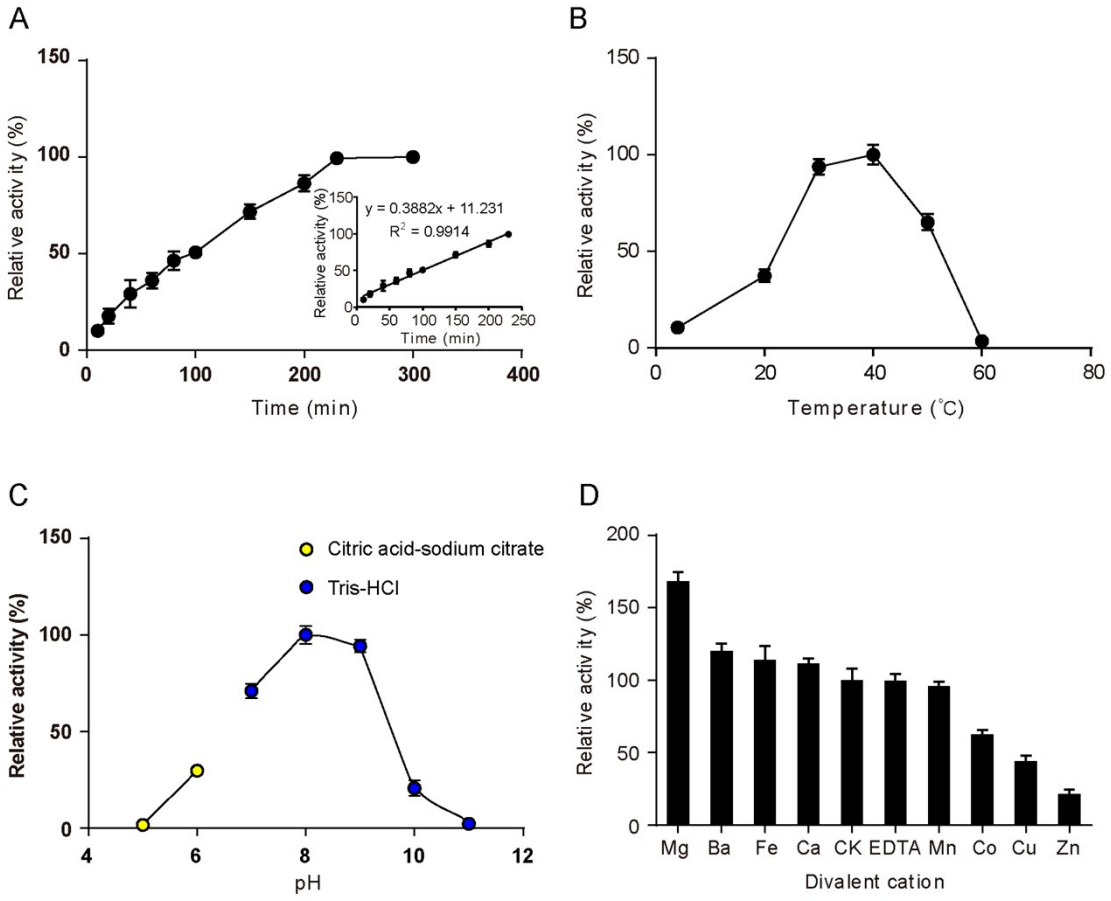


Figure S7. Effects of reaction time (A), temperature (B), reaction buffer (C), and divalent metal ions (D) on the activities of RyUGT12. 6-Hydroxyrubiadin (**1**) was used as the acceptor and UDP-Glc was used as the sugar donor. An optimized reaction time of 4 h was used for (B), (C) and (D). Experiment with 5 mM EDTA or Tris-HCl (pH 7.5, control blank (CK)) was performed as a negative control in (D). RyUGT3A exhibited its maximum activity at pH 8.0 (50 mM Tris-HCl buffer) and 40 ℃.

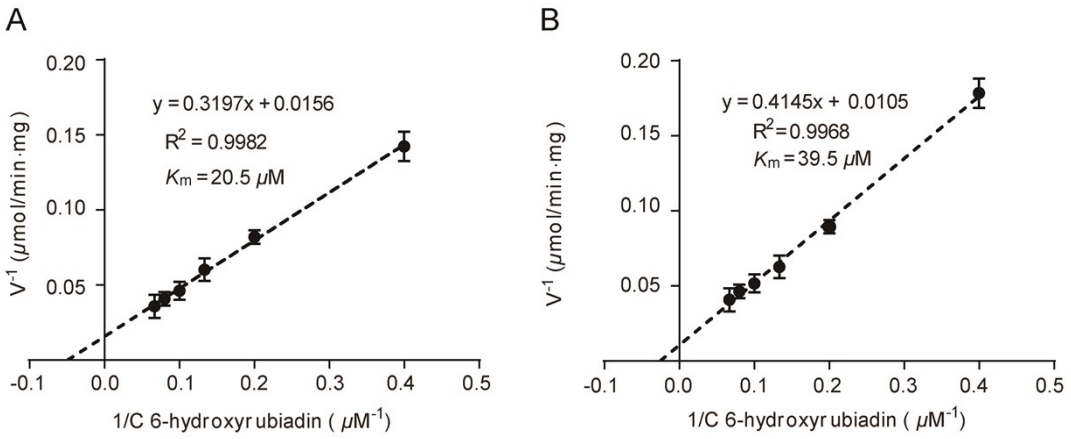


Figure S8. Determination of kinetic parameters for purified RyUGT3A (A) /RyUGT12 (B). The apparent K_m value was determined using 6-hydroxyrubiadin (**1**) as the acceptor and UDP-Glc as the sugar donor at 30 ℃/40 ℃ and pH 9.0/8.0 for 10 min. The apparent K_m values are calculated as 20.5 μM and 39.5 μM, respectively.

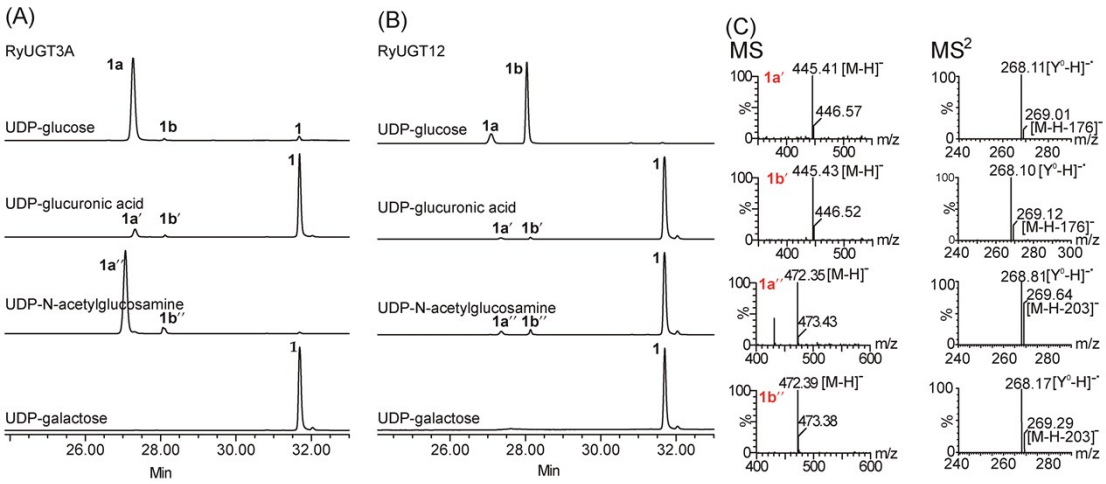


Figure S9. Assessment of UDP-sugar donor catalyzed by RyUGT3A and RyUGT12. (A and B) RyUGT3A and RyUGT12 were incubated with 6-hydroxyrubiadin (**1**) as an acceptor and UDP-glucose, UDP-glucuronic acid, UDP-N-acetylglucosamine and UDP-galactose as sugar donors, respectively; (C) LC-MS/MS analysis of glycosylated products (**1a′**, **1b′**, **1a″**, and **1b″**) in the negative ion mode.

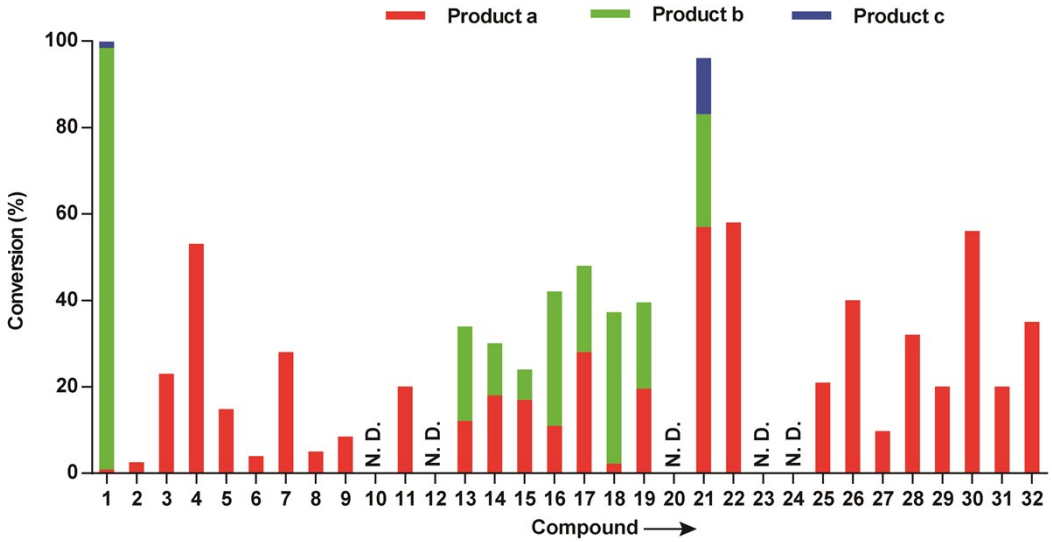


Figure S10. Percent yields of glucosylated products catalyzed by RyUGT11. Compounds (**1-32**) are listed based on the structural scaffolds with numbering corresponding to the structures listed in Fig. 2B. The colors in the bar graphs (Product **a**, **b**, and **c**) represent different ratios of various glycosylated products. N.D. indicates no products are detected.

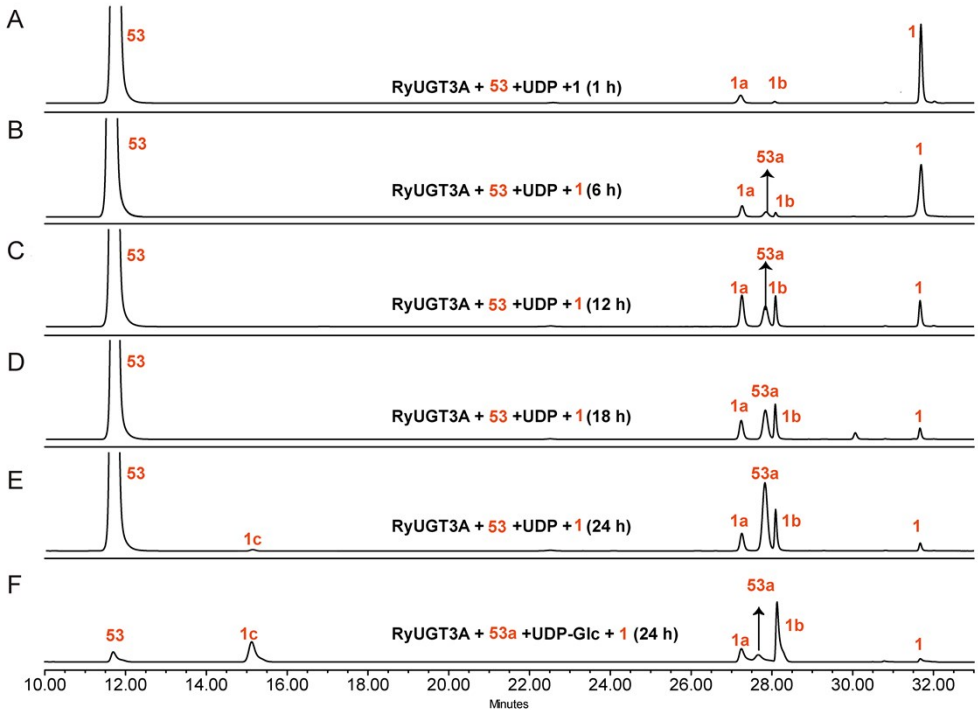


Figure S11. HPLC analysis of time course of the reactions in one-pot method. (A-E) The reaction at 30 ℃ for 24 h with 200 μL Tris-HCl buffer solution (50 mM, pH 9.0) containing 14 mM β-mercaptoethanol, 5 mM UDP, 50 mM 4-nitrophenyl-β-D-glucopyranoside (**53**), 0.5 mM 6-hydroxyrubiadin (**1**) and 5 μg purified RyUGT3A. 1, 6, 12, 18, and 24 h represent the different reaction time points. (F) The reaction at 30 ℃ for 24 h with 200 μL Tris-HCl buffer solution (50 mM, pH 9.0) containing 14 mM β-mercaptoethanol, 0.5mM substrate **1**, 0.5 mM aglycone **53a,** 5 mM UDP-Glc and 5 μg purified RyUGT3A.

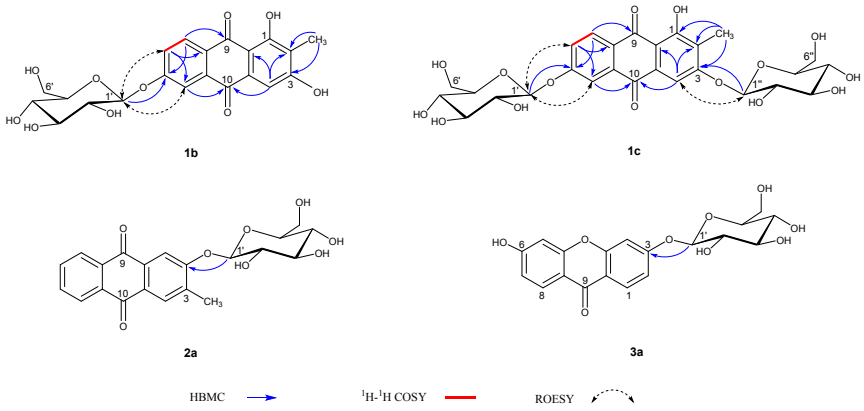


Figure S12. Key 2D NMR correlations of four new compounds.

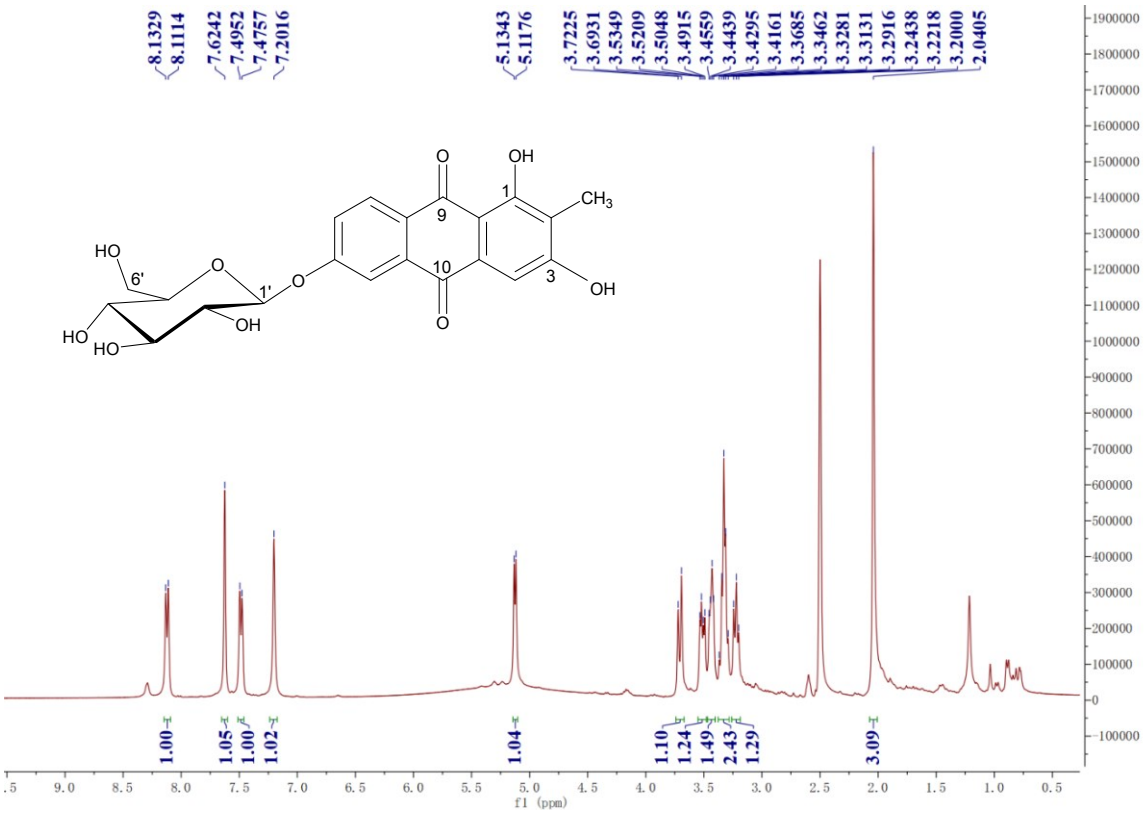


Figure S13. ^{1}H NMR spectrum (400 MHz) of compound **1b** in DMSO-d_6.

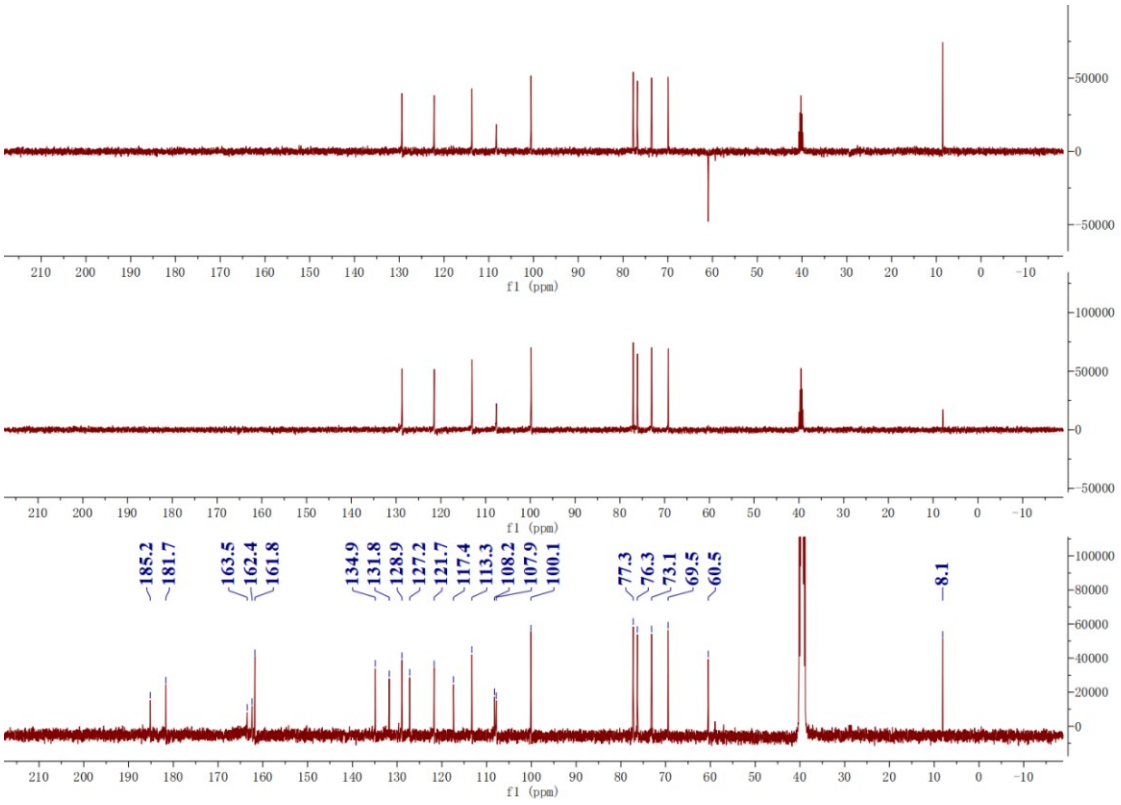


Figure S14. ^{13}C NMR spectrum (100 MHz) of compound **1b** in DMSO-d_6.

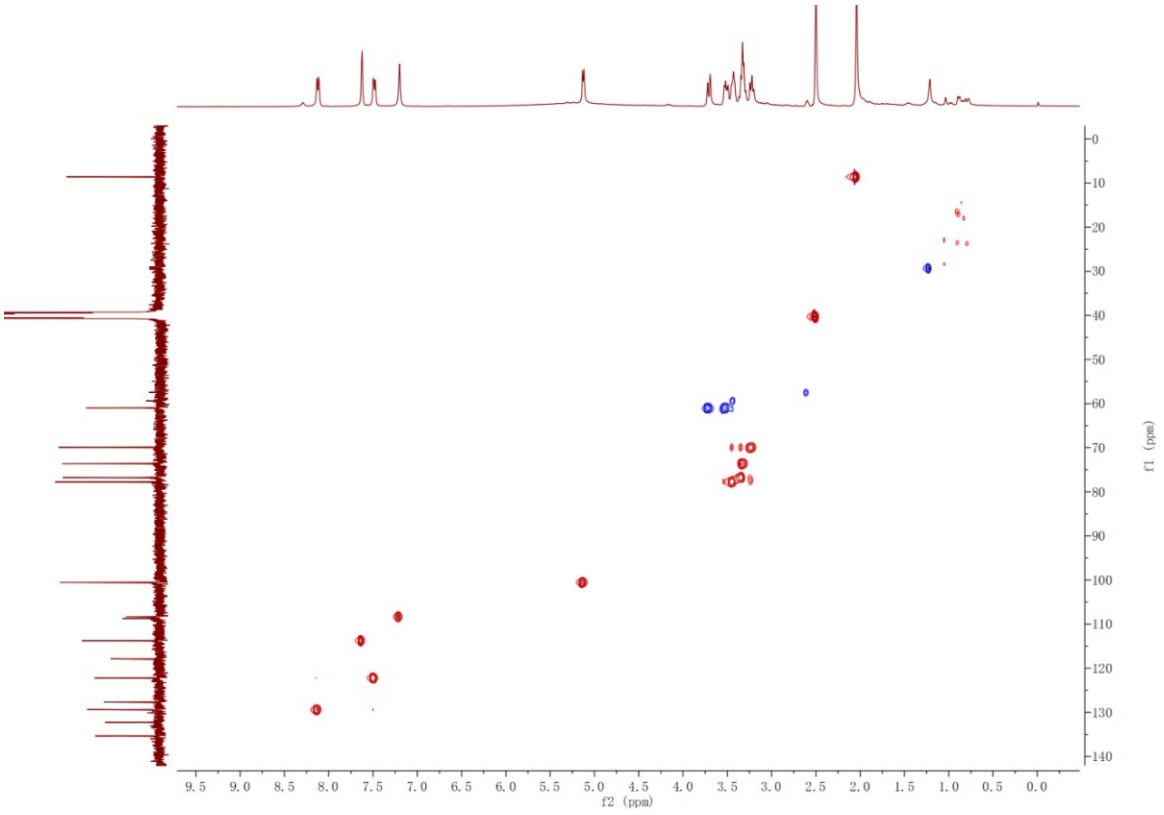


Figure S15. HSQC spectrum of compound **1b** in DMSO-d_6.

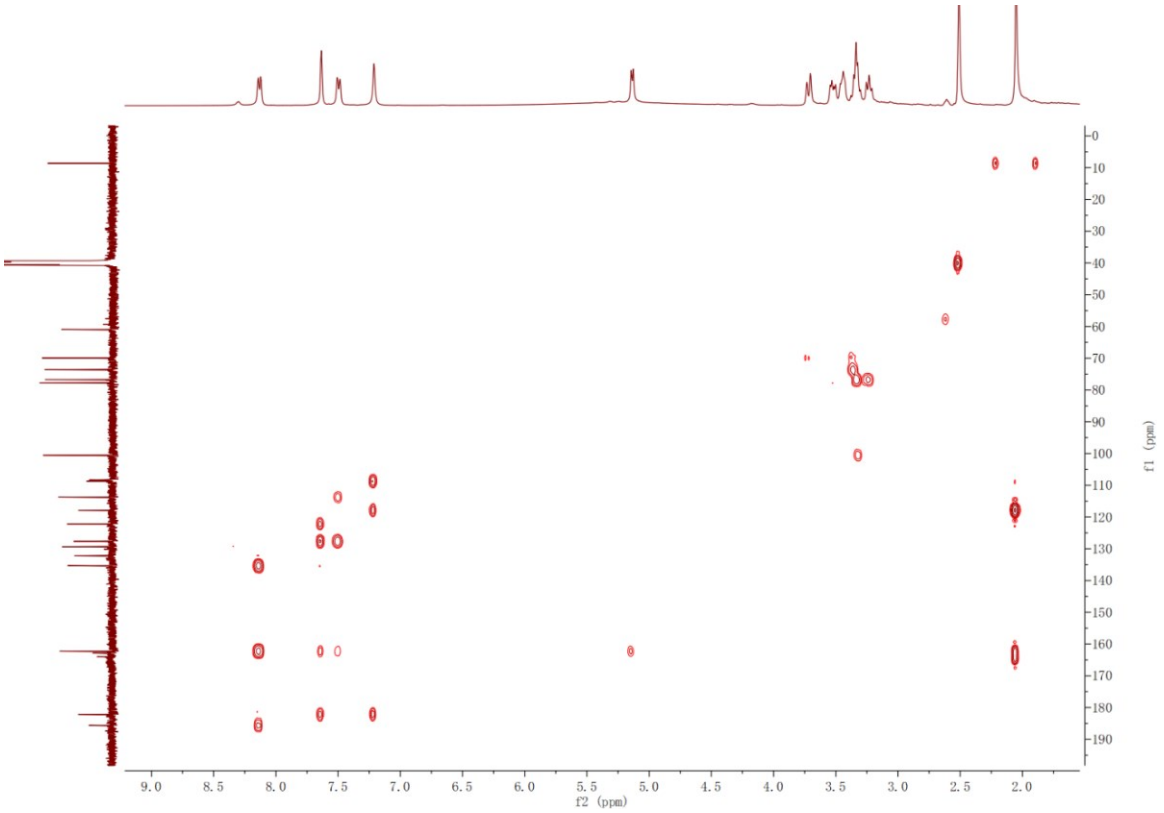


Figure S16. HMBC spectrum of compound **1b** in DMSO-d_6.

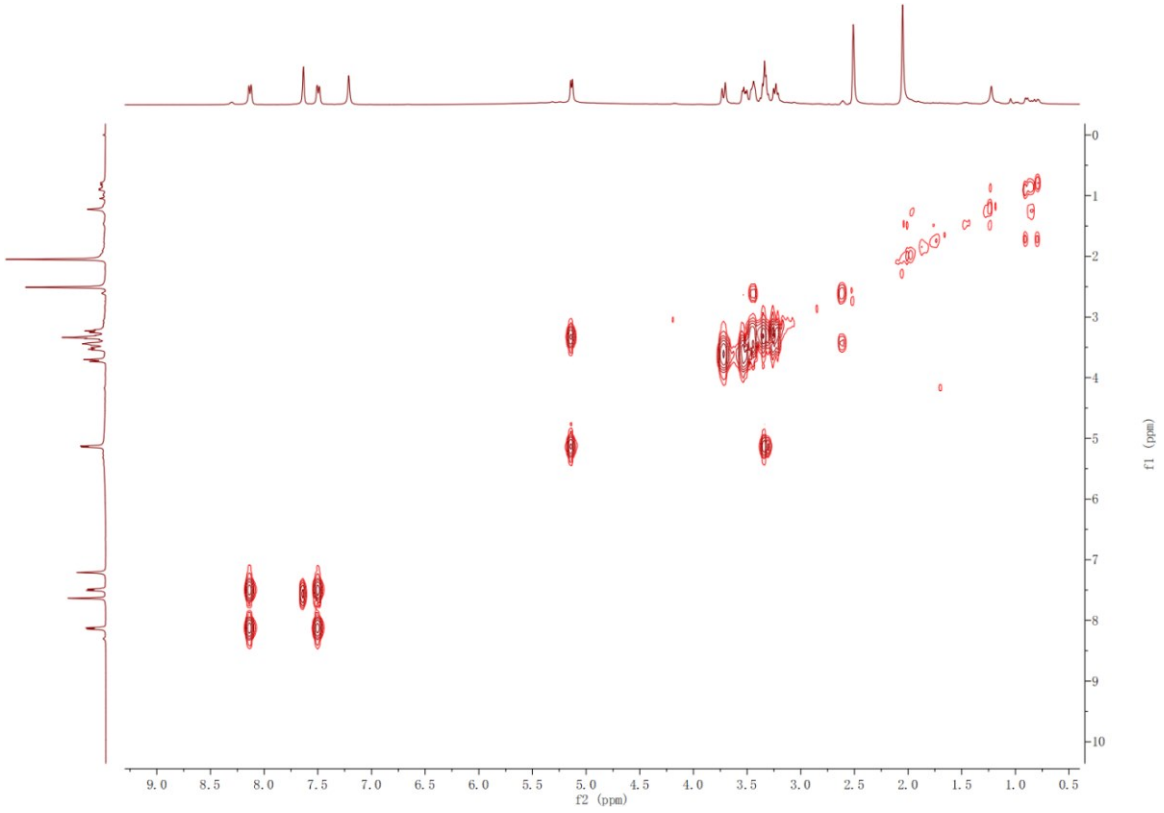


Figure S17. ^{1}H-^{1}H COSY spectrum of compound **1b** in DMSO-d_6.

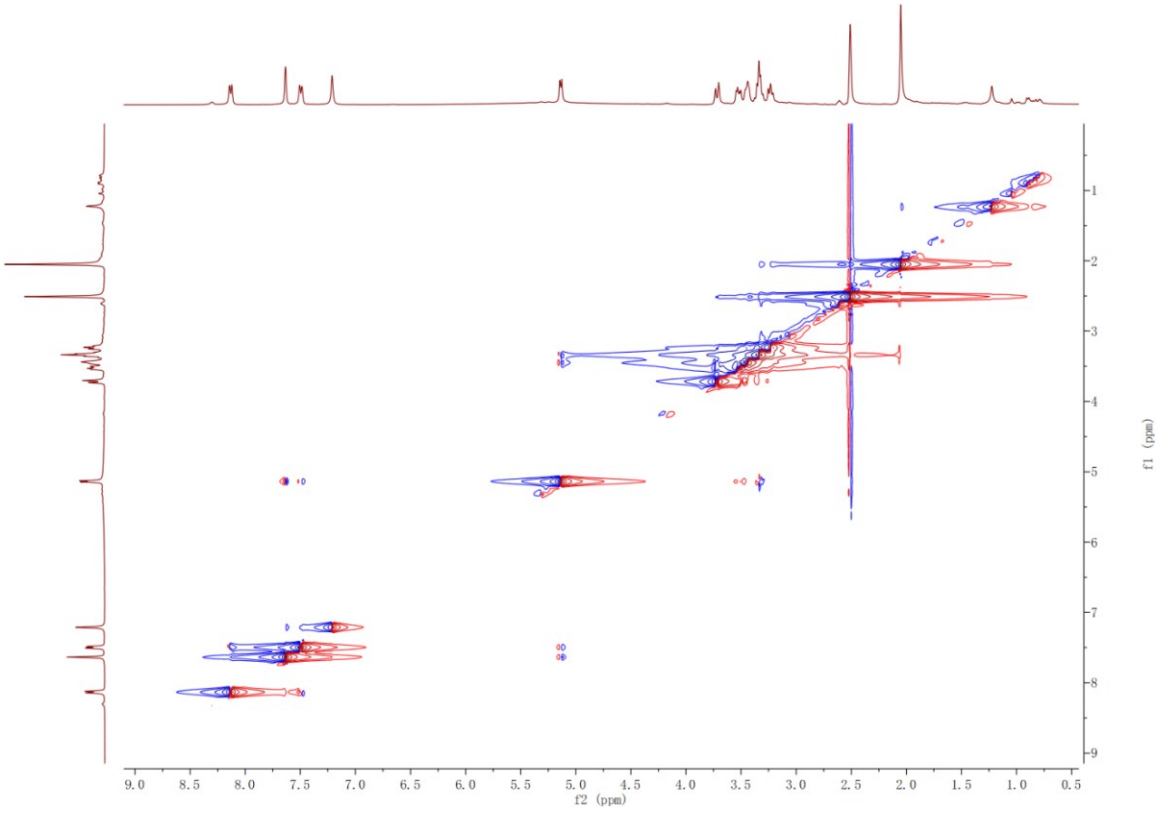


Figure S18. ROESY spectrum of compound **1b** in DMSO-d_6.

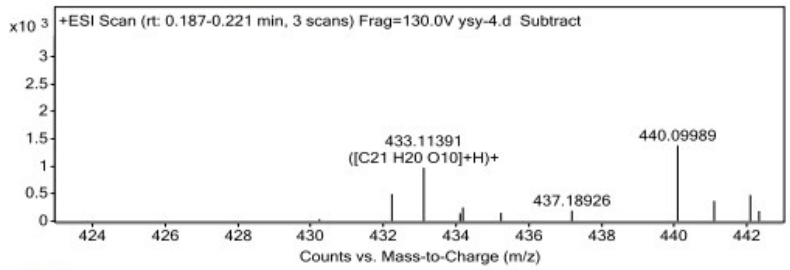
Qualitative Analysis Report

Data Filename	ysy-4.d	**Sample Name**	ysy-4
Sample Type	Sample	**Position**	Vial 64
Instrument Name	Instrument 1	**User Name**	
Acq Method	method-POSI.m	**Acquired Time**	7/12/2019 9:27:26 AM (UTC+08:00)
IRM Calibration Status	Success	**DA Method**	1.m
Comment			

Sample Group		**Info.**	
Stream Name	LC 1	**Acquisition Time (Local)**	7/12/2019 9:27:26 AM (UTC+08:00)
Acquisition SW Version	6200 series TOF/6500 series Q-TOF B.06.01 (B6157)	**TOF Driver Version**	6.00.01
TOF Firmware Version	17.643		

Spectra

Collision Energy 0 **Ionization Mode** ESI



Peak List

m/z	z	Abund
63.99844	1	10094.81
79.02177	1	268605.16
79.04096		17580.39
85.05938	1	40191.91
279.15905	1	19623.09

Formula Calculator Element Limits

Element	Min	Max
C	11	31
H	10	30
O	9	11

Formula Calculator Results

Formula	Best	Mass	Tgt Mass	Diff (ppm)	Ion Species	CalculatedMz
C21 H20 O10	TRUE	432.10638	432.10565	1.69	C21 H21 O10	433.11292

--- End Of Report ---

Figure S19. HR-ESI-MS spectrum of compound **1b**.

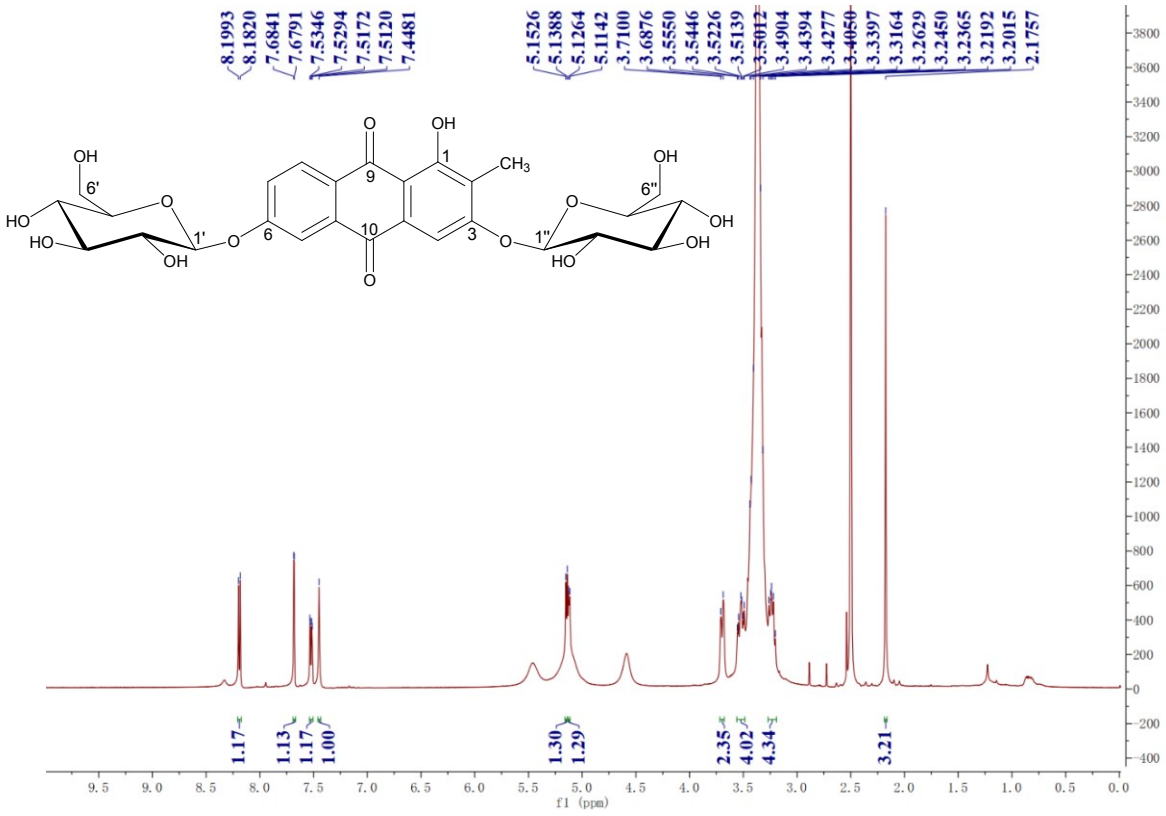


Figure S20. ^{1}H NMR spectrum (500 MHz) of compound **1c** in DMSO-d_6.

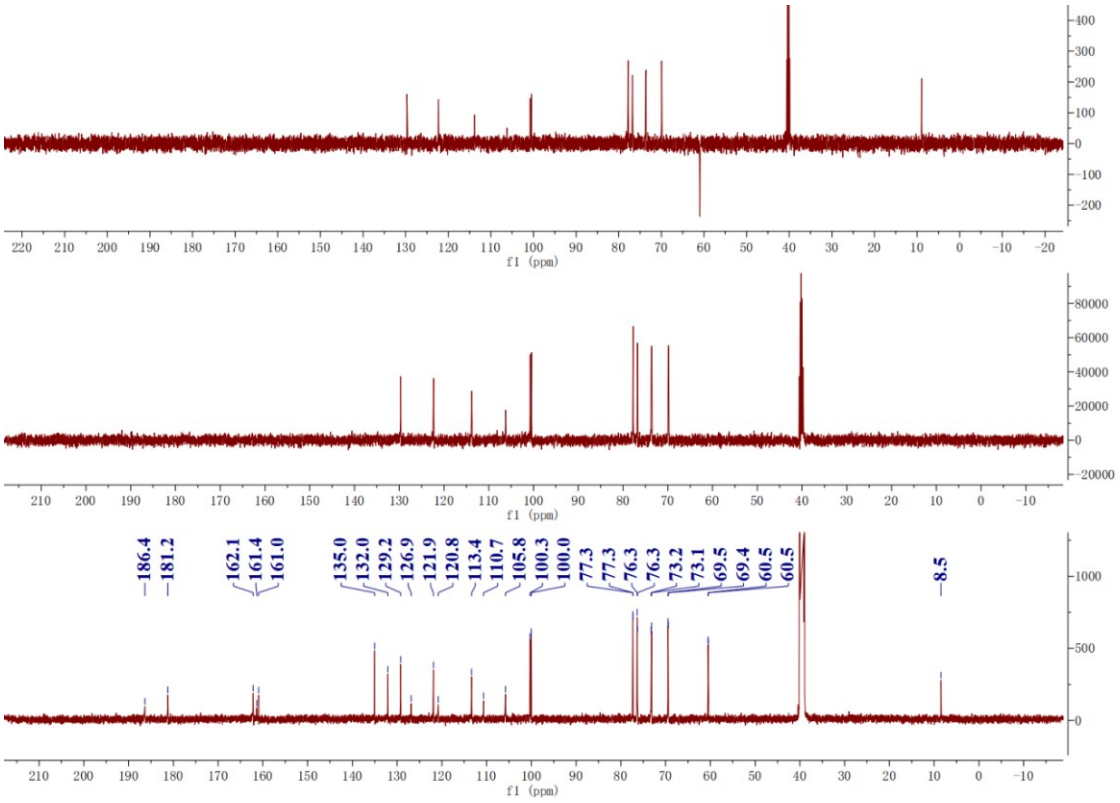


Figure S21. ^{13}C NMR spectrum (125 MHz) of compound **1c** in DMSO-d_6.

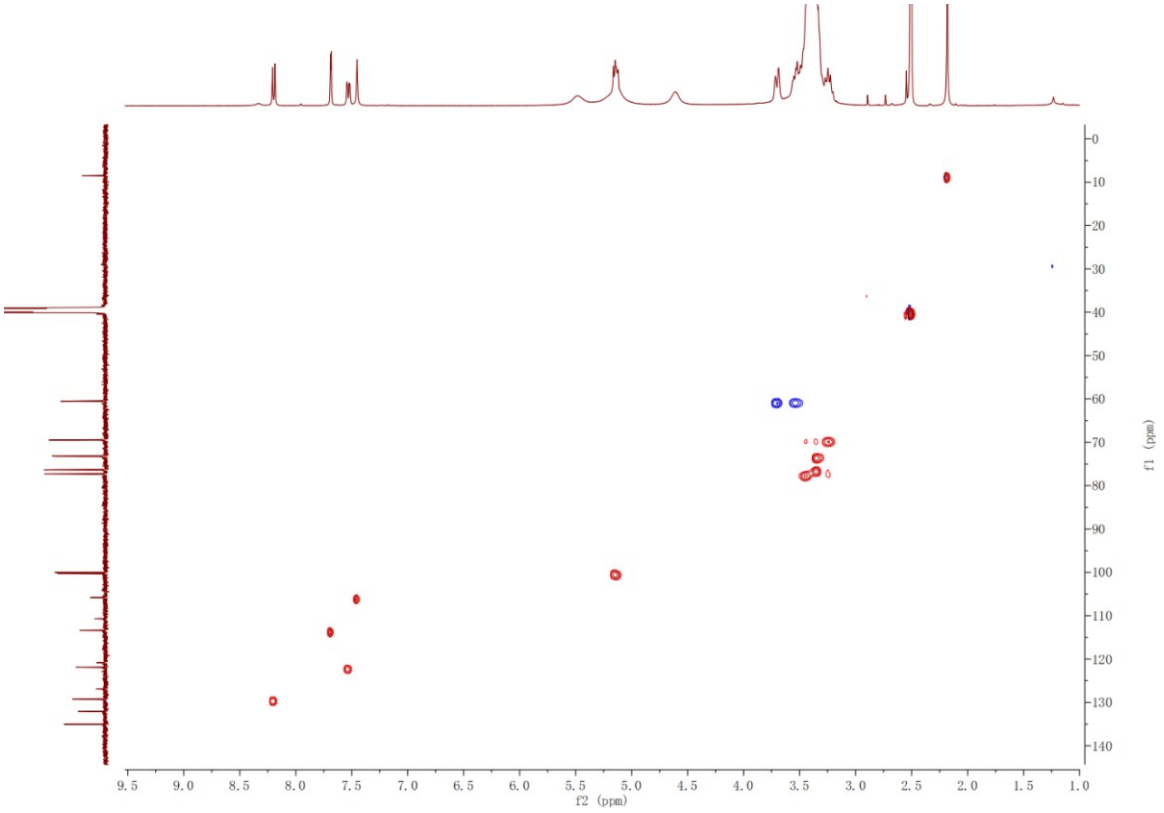


Figure S22. HSQC spectrum of compound **1c** in DMSO-d_6.

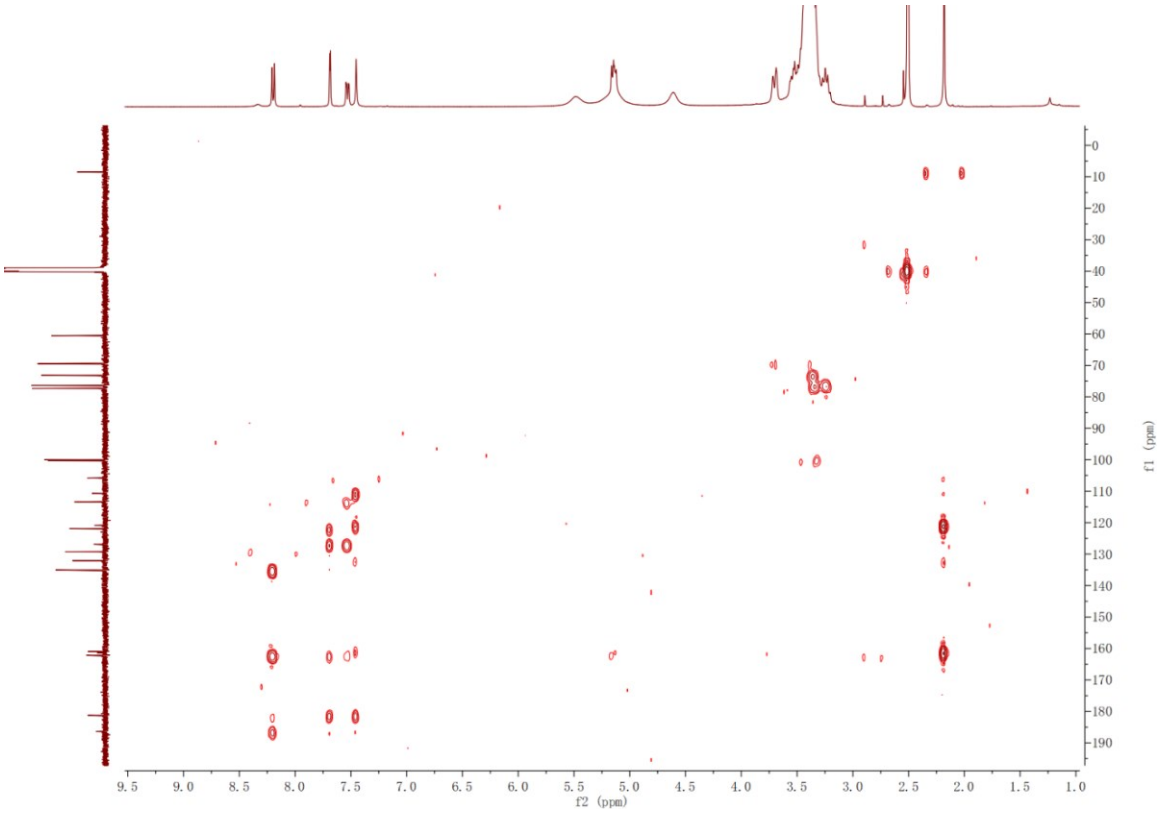


Figure S23. HMBC spectrum of compound **1c** in DMSO-d_6.

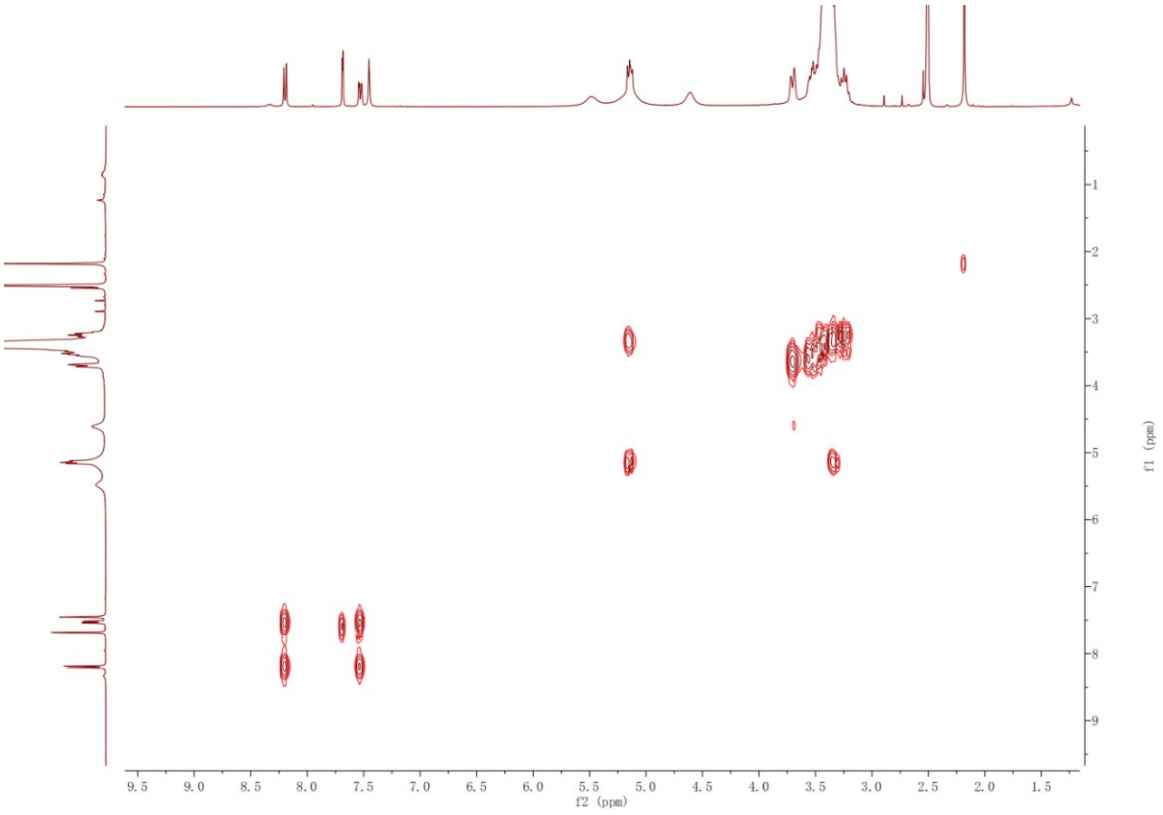


Figure S24. ^{1}H-^{1}H COSY spectrum of compound **1c** in DMSO-d_6.

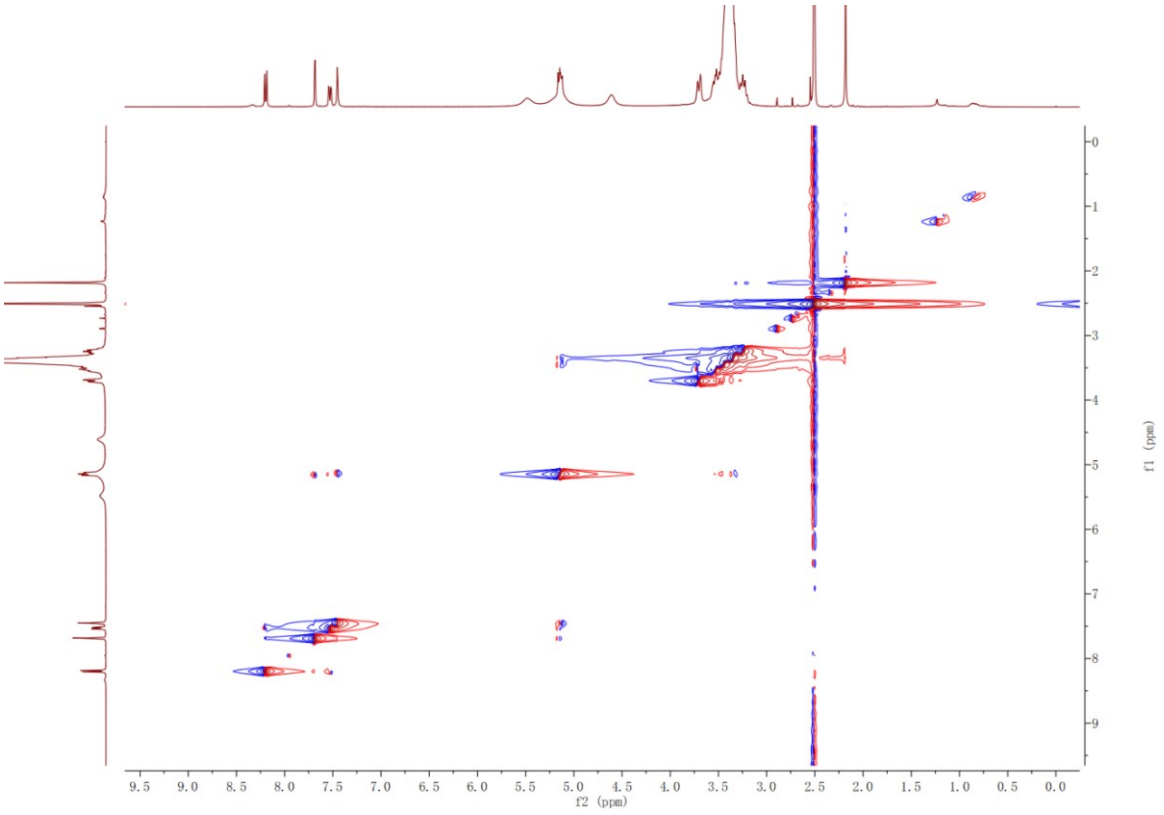


Figure S25. ROESY spectrum of compound **1c** in DMSO-d_6.

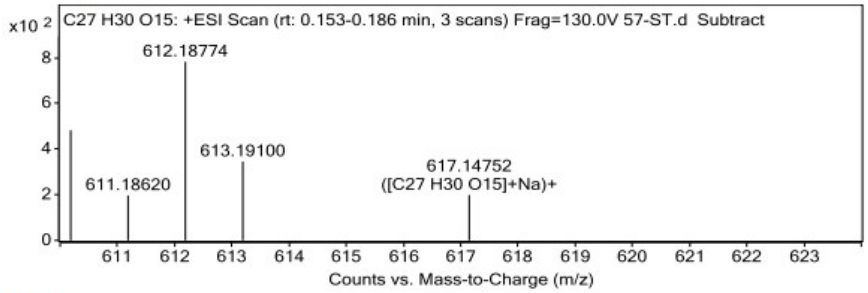
Qualitative Analysis Report

Data Filename	57-ST.d	**Sample Name**	57-ST
Sample Type	Sample	**Position**	Vial 63
Instrument Name	Instrument 1	**User Name**	
Acq Method	method-POSI.m	**Acquired Time**	7/12/2019 9:24:56 AM (UTC+08:00)
IRM Calibration Status	Success	**DA Method**	1.m
Comment			

Sample Group		**Info.**	
Stream Name	LC 1	**Acquisition Time (Local)**	7/12/2019 9:24:56 AM (UTC+08:00)
Acquisition SW Version	6200 series TOF/6500 series Q-TOF B.06.01 (B6157)	**TOF Driver Version**	6.00.01
TOF Firmware Version	17.643		

Spectra

Collision Energy 0 **Ionization Mode** ESI



Peak List

m/z	z	Abund
63.99835	1	17960.08
79.02165	1	517832.22
81.01773	1	17529.82
85.05931	1	67318.02
279.159	1	32213.05

Formula Calculator Element Limits

Element	Min	Max
C	17	37
H	20	40
O	14	16

Formula Calculator Results

Formula	Best	Mass	Tgt Mass	Diff (ppm)	Ion Species	CalculatedMz
C27 H30 O15	TRUE	594.1583	594.15847	-0.29	C27 H30 Na O15	617.14769

Figure S26. HR-ESI-MS spectrum of compound **1c**.

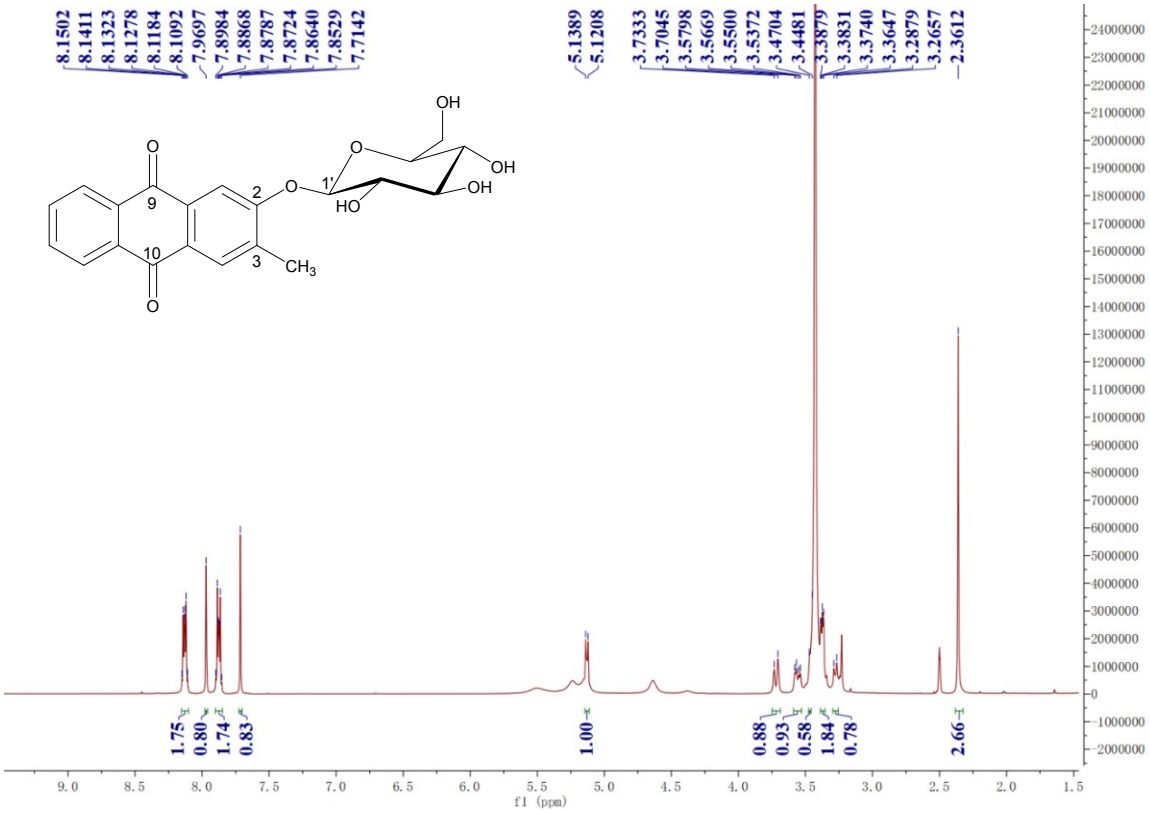


Figure S27. ^{1}H NMR spectrum (400 MHz) of compound **2a** in DMSO-d_6.

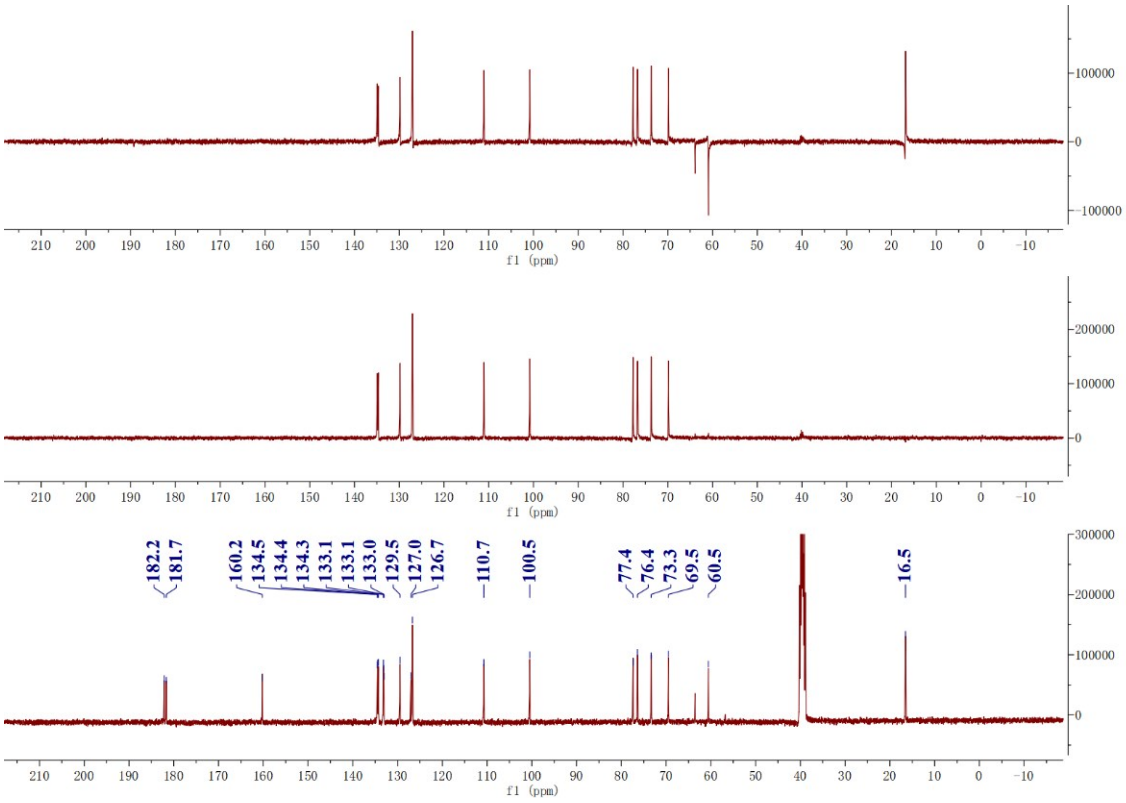


Figure S28. ^{13}C NMR spectrum (100 MHz) of compound **2a** in DMSO-d_6.

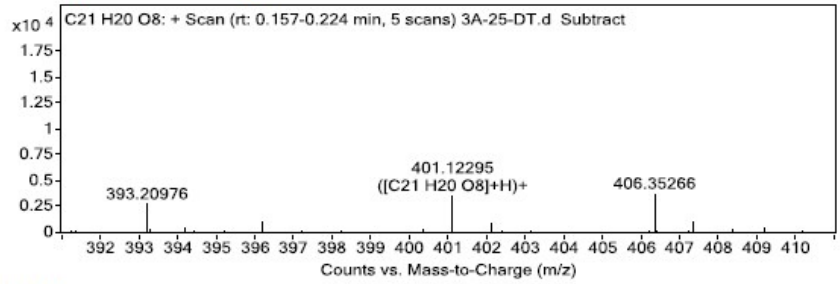
Qualitative Analysis Report

Data Filename	3A-25-DT.d	**Sample Name**	3A-25-DT
Sample Type	Sample	**Position**	Vial 3
Instrument Name	Instrument 1	**User Name**	
Acq Method	method-POSI.m	**Acquired Time**	10/31/2019 9:08:10 AM (UTC+08:00)
IRM Calibration Status	Success	**DA Method**	1.m
Comment			

Sample Group		**Info.**	
Stream Name	LC 1	**Acquisition Time (Local)**	10/31/2019 9:08:10 AM (UTC+08:00)
Acquisition SW Version	6200 series TOF/6500 series Q-TOF B.06.01 (B6157)	**TOF Driver Version**	6.00.01
TOF Firmware Version	17.643		

Spectra

Collision Energy 0 **Ionization Mode** ESI



Peak List

m/z	z	Abund
79.02274	1	2143430.25
79.1073		222520.59
101.00347	1	223199.25
157.03527	1	1517113.25
179.01715	1	183329.39

Formula Calculator Element Limits

Element	Min	Max
C	11	31
H	10	30
O	7	9

Formula Calculator Results

Formula	Best	Mass	Tgt Mass	Diff (ppm)	Ion Species	CalculatedMz
C21 H20 O8	TRUE	400.11599	400.11582	-0.43	C21 H21 O8	401.12309

--- End Of Report ---

Figure S29. HR-ESI-MS spectrum of compound **2a**.

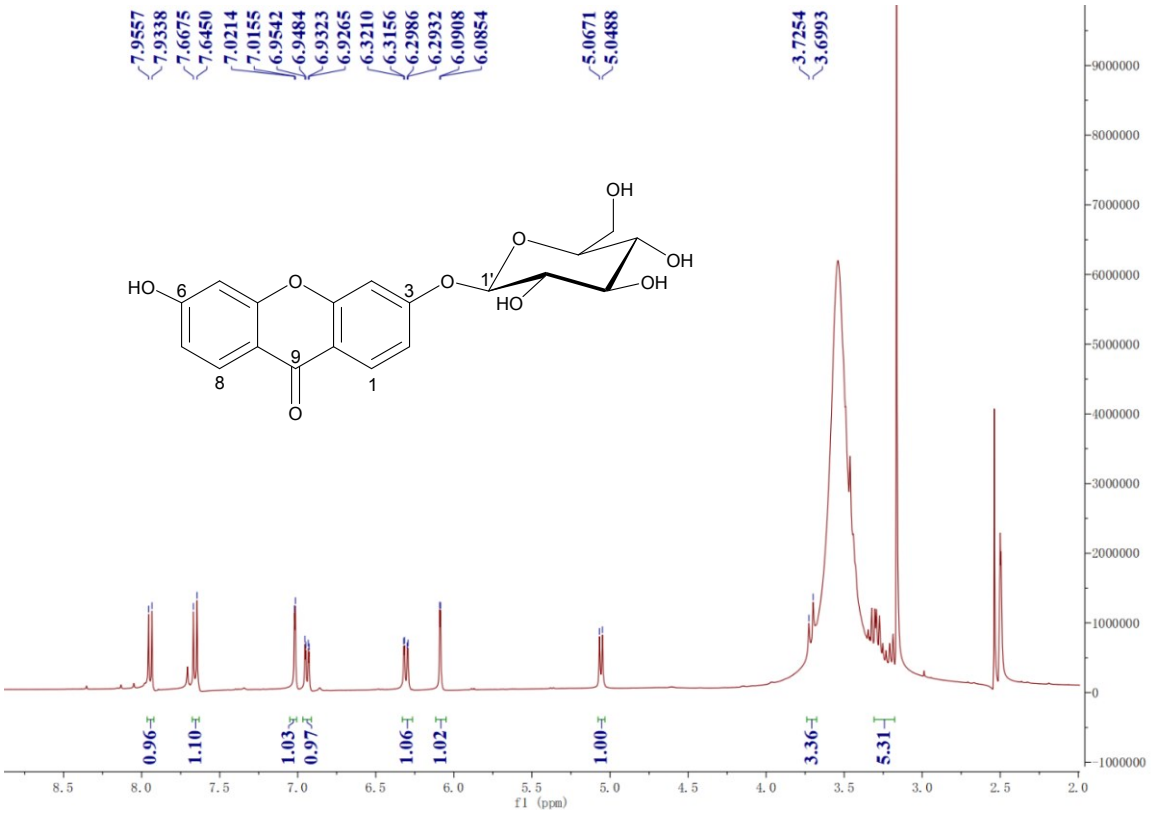


Figure S30. ^{1}H NMR spectrum (400 MHz) of compound **3a** in DMSO-d_6.

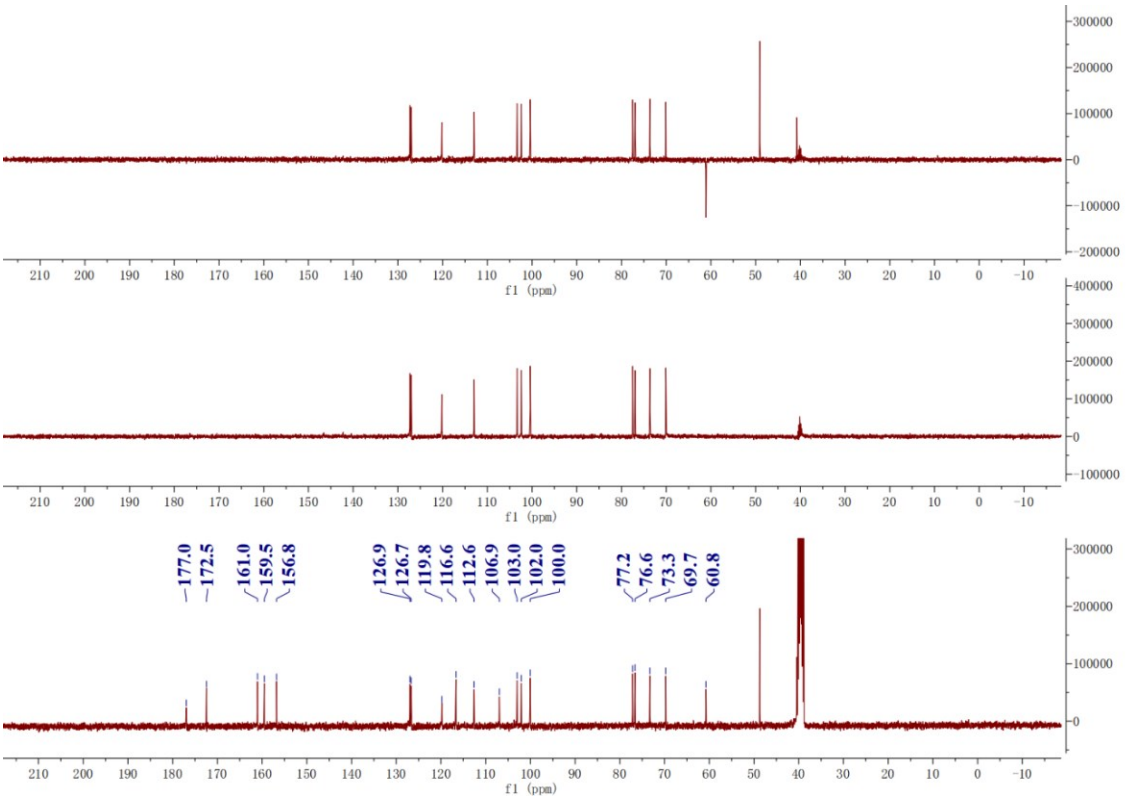


Figure S31. ^{13}C NMR spectrum (100 MHz) of compound **3a** in DMSO-d_6.

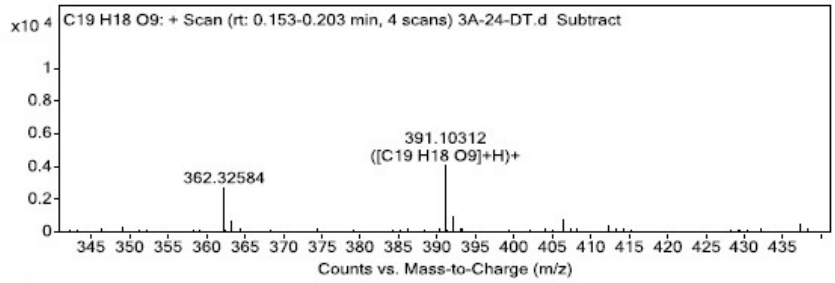
Qualitative Analysis Report

Data Filename	3A-24-DT.d	**Sample Name**	3A-24-DT
Sample Type	Sample	**Position**	Vial 1
Instrument Name	Instrument 1	**User Name**	
Acq Method	method-POSI.m	**Acquired Time**	10/31/2019 9:03:06 AM (UTC+08:00)
IRM Calibration Status	Success	**DA Method**	1.m
Comment			

Sample Group		**Info.**	
Stream Name	LC 1	**Acquisition Time (Local)**	10/31/2019 9:03:06 AM (UTC+08:00)
Acquisition SW Version	6200 series TOF/6500 series Q-TOF B.06.01 (B6157)	**TOF Driver Version**	6.00.01
TOF Firmware Version	17.643		

Spectra

Collision Energy 0 **Ionization Mode** ESI



Peak List

m/z	z	Abund
79.02184	1	501523.19
96.04812		50202.74
102.12793	1	93869.98
105.0425	1	103574.69
157.03499	1	107829.84

Formula Calculator Element Limits

Element	Min	Max
C	9	29
H	8	28
O	8	10

Formula Calculator Results

Formula	Best	Mass	Tgt Mass	Diff (ppm)	Ion Species	CalculatedMz
C19 H18 O9	TRUE	390.09548	390.09508	-1.02	C19 H19 O9	391.10236

--- End Of Report ---

Figure S32. HR-ESI-MS spectrum of compound **3a**.

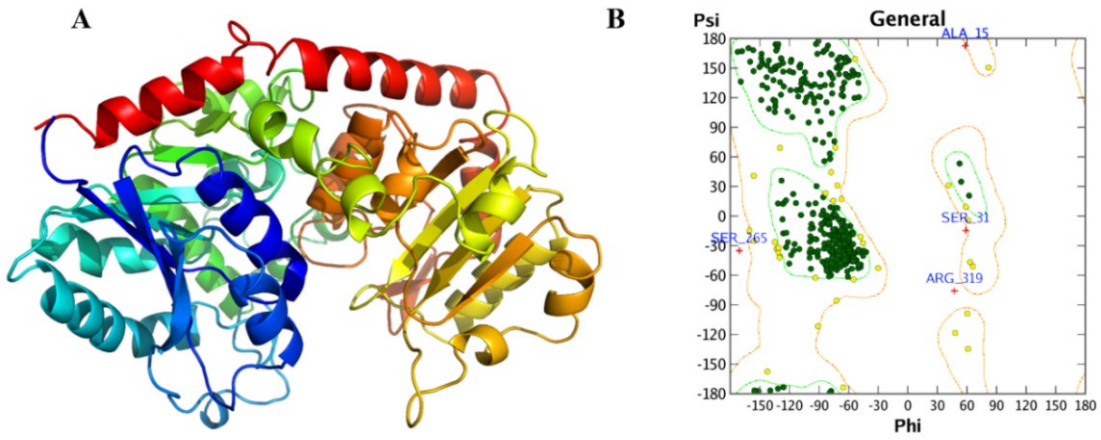


Figure S33. (A) Homology model of RyUGT3A. (B) Ramachandran plot for RyUGT3A. Dark green dots represent the residues in favored regions; yellow dots represent the residues in allowed regions; red cross represents the residues in irrational regions.

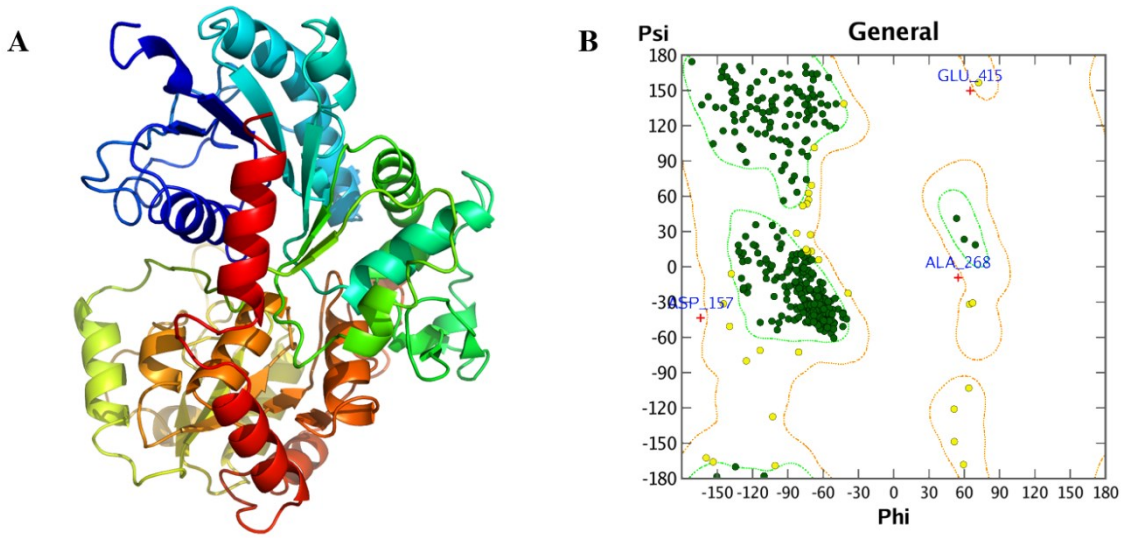


Figure S34. (A) Homology model of RyUGT12. (B) Ramachandran plot for RyUGT12. Dark green dots represent the residues in favored regions; yellow dots represent the residues in allowed regions; red cross represents the residues in irrational regions.

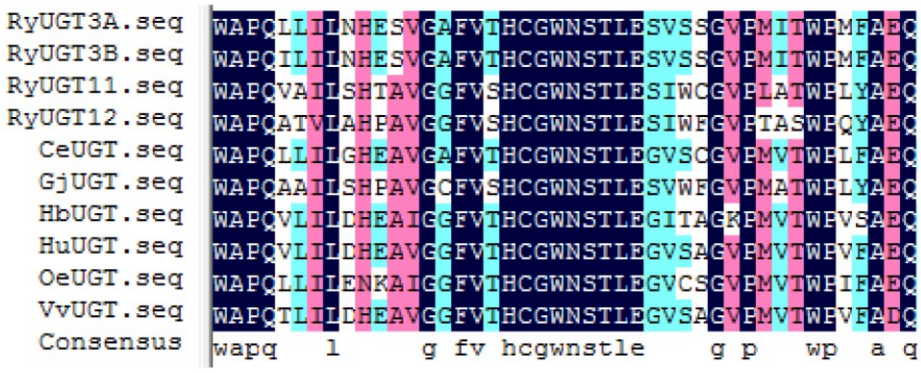


Figure S35. Multiple sequence alignment of PSPG motifs of four RyUGTs. Identical residues are highlighted with a blue background and similar residues with a pink or cyan background. Gene names, GenBank accession numbers and plant species are as follows: CeUGT, XP_027161059.1, from *Coffea eugenioides*; GjUGT, BAK55746.1, from *Gardenia jasminoides*; HbUGT, XP_021684532.1, from *Hevea brasiliensis*; HuUGT, XP_021298085.1, from *Herrania umbratica*; OeUGT, XP_022868325.1, from *Olea europaea*; VvUGT, CAN65903.1, from *Vitis vinifera.*

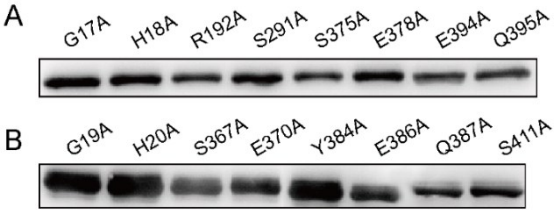


Figure S36. Western blotting analysis of mutants of RyUGT3A (A) and RyUGT12 (B).

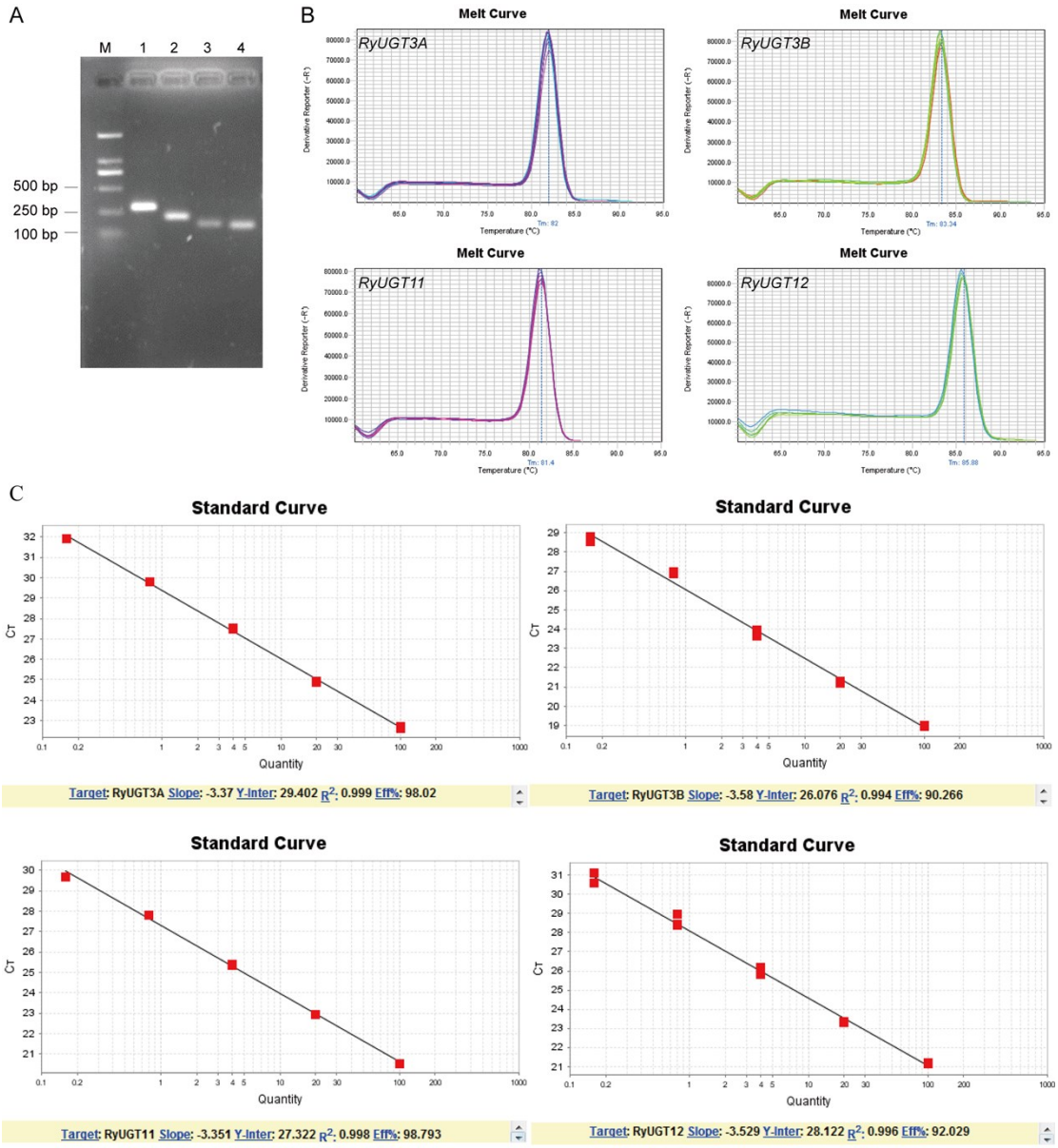


Figure S37. Specificity of primer pairs for RT-qPCR amplification. (A) The 2% agarose gel electrophoresis showed the expected size of a single band for each candidate reference gene, M represents the DNA size marker, lane 1-lane 4: *RyUGT3A*, *RyUGT3B*, *RyUGT11*, *RyUGT12*. (B) Melt curves with single peaks produced for these four amplicons. (C) Standard curves of four target genes directly generated by StepOne™ Real-time PCR system.

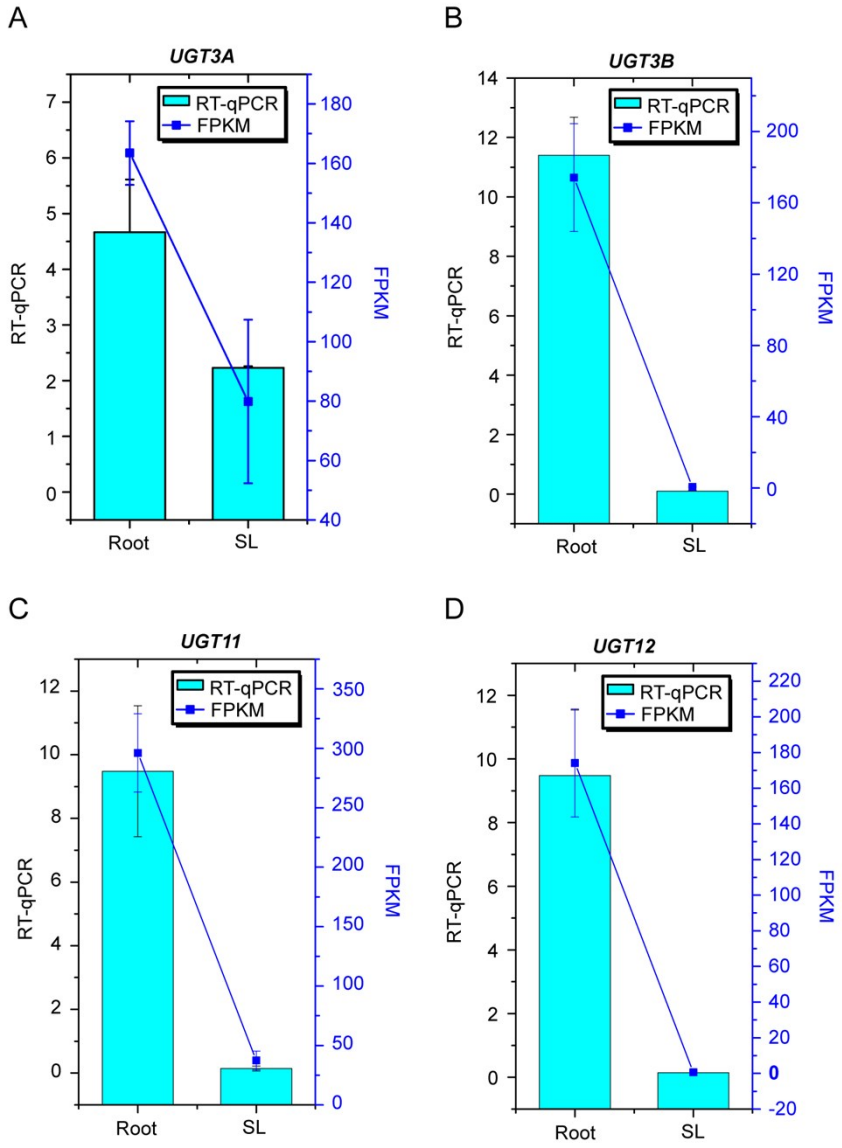


Figure S38. Quantitative comparison of relative expression levels of four RyUGTs in root and stem leaf (SL) of *R. yunnanensis* by RT-qPCR and FPKM. Columns indicated relative expression levels of four *RyUGT* genes calculated by RT-qPCR (left y-axis) and *hnRNP* was used as the reference gene ($2^{-\Delta\Delta Ct}$ method); Lines indicating the relative expression levels were obtained by FPKM method (right y-axis). Error bars indicate the standard deviation of mean values of three replications.

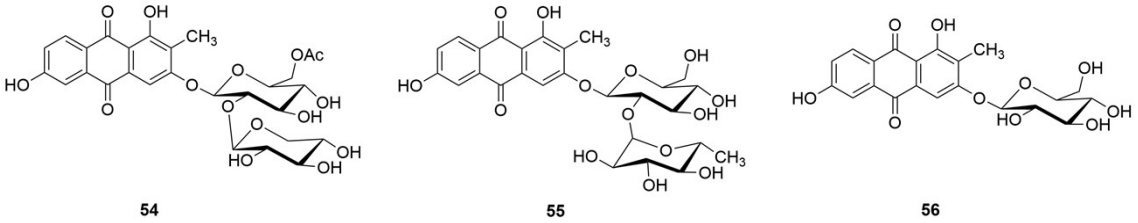


Figure S39. Chemical structures of anthraquinone glycosides **54**, **55**, and **56** isolated from *R. yunnanensis*.

References

(1) Fan, J. T.; Kuang, B.; Zeng, G. Z.; Zhao, S. M.; Ji, C. J.; Zhang, Y. M.; Tan, N. H. *J Nat Prod*. **2011**, *74*, 2069-2080.

(2) Yi, S. Y.; Lin, Q. W.; Zhang, X. J.; Wang, J.; Miao, Y. Y.; Tan, N. H. *BioMed Res Int*. **2020**, *2020*.

(3) Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S. *Mol Biol Evo*. **2013**, *30*, 2725-2729.

(4) Ross, J.; Li, Y.; Lim, E.; Bowles, D. J. *Genome Biol*. **2001**, *2*, reviews3004.

(5) Guo, D.; Wang, Y. R.; Wang, J.; Song, L. H.; Wang, Z.; Mao, B. Y.; Tan, N. H. *Molecules*. **2019**, *24*, 1-14.

(6) Feng, J.; Zhang, P.; Cui, Y. L.; Li, K.; Qiao, X.; Zhang, Y. T.; Li, S. M.; Cox, R. J.; Wu, B.; Ye, M.; Yin, W. B. *Adv Synth Catal*. **2017**, *6*, 995-1006.

(7) Liu, M. S.; Zheng, N.; Li, D. M.; Zheng, H. L.; Zhang, L. L.; Ge, H.; Liu, W. D. *Med Mycol*. **2016**, *54*, 400-408.

(8) Yin, Q. G.; Shen, G. A.; Chang, Z. Z.; Tang, Y. H.; Gao, H. W.; Pang, Y. Z. *J Exp Bot*. **2017**, *68*, 597-612.

(9) He, J. B.; Zhao, P.; Hu, Z. M.; Liu, S.; Kuang, Y.; Zhang, M.;. Li, B.; Yun, C. H.; Qiao, X.; Ye, M. *Angew Chem Int Ed Engl.* **2019**, *58*, 11513-11520.

(10) Zhang, M.; Li, F. D.; Li, K.; Wang, Z. L.; Wang, Y. X.; He, J. B.; Su, H. F.; Zhang, Z. Y.; Chi, C. B.; Shi, X. M.; Yun, C. H.; Zhang, Z. Y.; Liu, Z. M.; Zhang, L. R.; Yang, D. H.; Ma, M.; Qiao, X.; Ye, M. *J Am Chem Soc.* **2020**, *142*, 3506-3512.

(11) Itokawa, H.; Qiao, A.; Takeya, K. *Phytochemistry*. **1989**, *28*, 3465-3468.

(12) Correia-da-Silva, M.; Sousa, E.; Duarte, B.; Marques, F.; Carvalho, F.; Cunha-Ribeiro, L. M.; Pinto, M. M. *J Med Chem.* **2011**, *54*, 5373-5384.

(13) Zhang, W.; Ye, M.; Zhan, J. X.; Chen, Y. J.; Guo, D. *Biotechnol Lett.* **2004**, *26*, 127-131.

(14) Cudlín, J.; Blumauerová, M.; Steinbeová, N.; Matějů, J.; Zalabák, V. *Folia Microbiolog.* **1976**, *21*, 54-57.

(15) Nguyen, T. T. H.; Pandey, R. P.; Parajuli, P.; Han, J. M.; Jung, H. J.; Park, Y. I.; Sohng, J. K. *Molecules*. **2018**, *23*, E2171.

(16) Masayoshi, K.; Hiroshi, S.; Tohrv, E.; Heihachiro, T.; Itiro, Y. *Chem Pharm Bull*. **1986**, *34*, 3097-3101.

(17) Liu, Q. A.; Dixon, R. J.; Mabry, T. *Phytochemistry*. **1993**, *34*, 167-170.

(18) Liu, Q.; Markhama, K. R.; Paré, P. W.; Dixon, R. A.; Mabry, T. J. *Phytochemistry*. **1993**, *32*, 925-928.

(19) He, C. N.; Wang, C. L.; Guo, S. X.; Yang, J. S.; Xiao, P. G. *J Integr Plant Biol.* **2006**, *48*, 359-363.

(20) Wu, H.; Dushenkov, S.; Ho, C. T.; Sang, S. M. *Food Chem*. **2009**, *115*, 592-595.

(21) Kinjo, J. E.; Furusawa, J. I.; Baba, J.; Takeshita, T.; Yamasaki, M.; Nohara, T. *Chem Pharm Bull.* **1987**, *35*, 4846-4850.

(22) Hanawa, F.; Tahara, S.; Mizijtani, J. *Phytochemistry*. **1992**, *31*, 3005-3007.

(23) Kanchanapoom, T.; Suga, K.; Kasai, R.; Yamasaki, K.; Kamel, M. S.; Mohamed, M. H. *Chem Pharm Bull.* **2002**, *50*, 863-865.

(24) Pfaffl, M. W. *Nucleic Acids Res*. **2001**. *29*, e45.

(25) Zhao, Y. C.; Wang, N. N.; Zeng, Z. X.; Xu, S.; Huang, C. L.; Wang, W.; Liu, T. T.; Luo, J.; Kong, L. Y. *Front Plant Sci*. **2016**, *7*, 722.

(26) Hofgen, R.; Willmitzer, L. S. *Nucleic Acids Res*. **1988**, *16*, 9877.

(27) Sparkes, I. A.; Runions, J.; Kearns, A.; Hawes, C. *Nat Protoc*. **2006**, *1*, 2019-2025.