

Supplementary Data

***In silico* motif diversity analysis of the glycon preferentiality of plant secondary metabolic glycosyltransferases**

Ritesh Kumar, Rajender S. Sangwan, Smrati Mishra, Farzana. Sabir and Neelam S. Sangwan*

Supplementary Table 1. Detail of the fourteen *A. thaliana* GT sequences used in the study along with their E-value, P-value, and amino acid length.

S.N.	Accession No.	Description	E-value	P-value	Length
1.	NP_567953.1	UGT73B3 (UDP-glucosyl transferase 73B3); UDP-glycosyltransferase/ abscisic acid glucosyltransferase/ quercetin 3-O-glucosyltransferase/ transferase, transferring hexosyl groups [<i>A. thaliana</i>]	1.5e-139	2.97e-141	481
2.	NP_181215.1	UDP-glucuronosyl/UDP-glucosyl transferase family protein [<i>A. thaliana</i>]	4.4e-144	9.70e-146	496
3.	NP_179151.2	UGT73B4 (UDP-GLYCOSYLTRANSFERASE 73B4); UDP-glycosyltransferase/ UDP-glycosyltransferase/ quercetin 3-O-glucosyltransferase/ quercetin 7-O-glucosyltransferase/ transferase, transferring glycosyl groups [<i>A. thaliana</i>]	1e-142	1.85e-144	484
4.	NP_181216.1	UDP-glucuronosyl/UDP-glucosyl transferase family protein [<i>A. thaliana</i>]	9.7e-149	2.18e-150	496
5.	CAA17559.1	glucosyltransferase-like protein [<i>A. thaliana</i>]	6e-144	1.36e-145	478
6.	NP_567955.1	UGT73B1 (UDP-glucosyl transferase 73B1); UDP-glycosyltransferase/ abscisic acid glucosyltransferase/ quercetin 3-O-glucosyltransferase/ quercetin 7-O-glucosyltransferase [<i>A. thaliana</i>]	4.8e-138	9.04e-140	488
7.	NP_181213.1	UGT73C1 (UDP-GLUCOSYL TRANSFERASE 73C1); UDP-glycosyltransferase/ UDP-glycosyltransferase/ cis-zeatin O-beta-D-glucosyltransferase/ trans-zeatin O-beta-D-glucosyltransferase/ transferase, transferring glycosyl groups [<i>A. thaliana</i>]	2.7e-145	5.68e-147	491
8.	NP_567954.1	UGT73B2 (UDP-GLUCOSYLTRANSFERASE 73B2); UDP-glycosyltransferase/ UDP-glycosyltransferase/ flavonol 3-O-glucosyltransferase/ quercetin 7-O-glucosyltransferase [<i>A. thaliana</i>]	5.2e-141	1.13e-142	483
9.	BAC42195.1	Putative glucosyl transferase [<i>A. thaliana</i>]	3.9e-155	8.02e-157	495
10.	NP_181214.1	UGT73C2 (UDP-glucosyl transferase 73C2); UDP-glycosyltransferase/ transferase, transferring glycosyl groups / transferase, transferring hexosyl groups [<i>A. thaliana</i>]	3.7e-149	8.17e-150	496
11.	NP_181217.1	UGT73C6 (UDP-glucosyl transferase 73C6); UDP-glycosyltransferase/ UDP-glycosyltransferase/ quercetin 3-O-glucosyltransferase/ quercetin 4'-O-glucosyltransferase/ quercetin 7-O-glucosyltransferase/ transferase, transferring glycosyl groups [<i>A. thaliana</i>]	3.9e-155	8.02e-157	495
12.	AAD17393.1	Putative glucosyltransferase [<i>A. thaliana</i>]	3.9e-104	1.34e-105	460
13.	NP_181218.1	DOG1 (DON-GLUCOSYLTRANSFERASE 1); UDP-glycosyltransferase/ cis-zeatin O-beta-D-glucosyltransferase/ glucosyltransferase/ quercetin 4'-	1.8e-154	4.03e-156	495

		O-glucosyltransferase/ quercetin 7-O-glucosyltransferase/ trans-zeatin O-beta-D-glucosyltransferase/ transferase, transferring glycosyl groups [<i>A. thaliana</i>]			
14.	NP_179150.3	UGT73B5 (UDP-glucosyl transferase 73B5); UDP-glucosyltransferase/ UDP-glycosyltransferase/ quercetin 3-O-glucosyltransferase/ transferase, transferring glycosyl groups [<i>A. thaliana</i>]	1.2e-137	2.28e-139	484

Supplementary Table 2. Details of the four most diversified GT sequences in terms of number of amino acids, sugar donor specificity and plant origin used in the study along with their E-value, P-value and amino acid length.

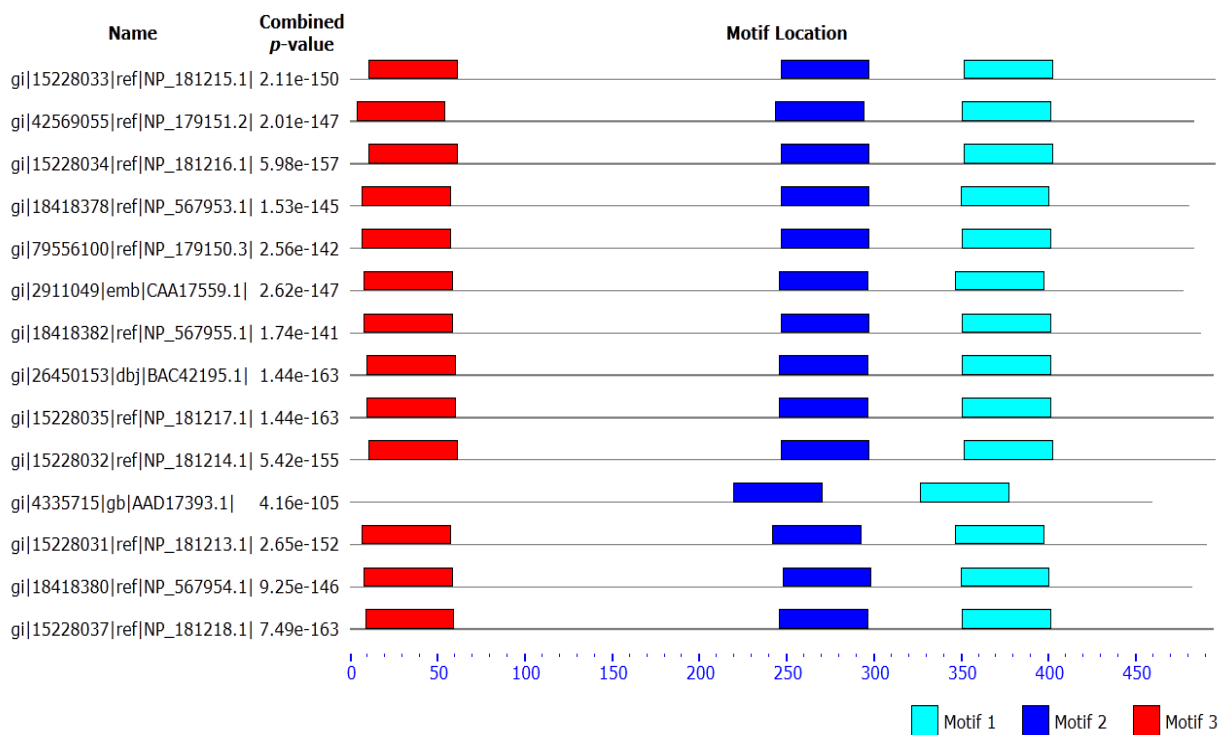
S.N.	Accession No.	Description	E-value	P-value	Length
1	XP_001765134.1	Predicted protein [<i>P. patens</i> subsp. <i>patens</i>]	2.40E-83	3.69E-52	265
2	NP_001148090.1	Cytokinin-O-glucosyltransferase3[<i>Z. mays</i>]	7.90E-119	4.70E-51	525
3	ACB47884.1	UDP-glucuronosyl/UDP-glucosyl transferase protein [<i>T. aestivum</i>]	1.80E-156	1.80E-60	496
4	ACR43490.1	UDP-glucosyl transferase [<i>S. cereale</i>]	4.60E-156	4.42E-61	496

Supplementary Table 3. Detail of GT sequences from 40 diverse plants used in the study with their E-value, P-value and amino acid length.

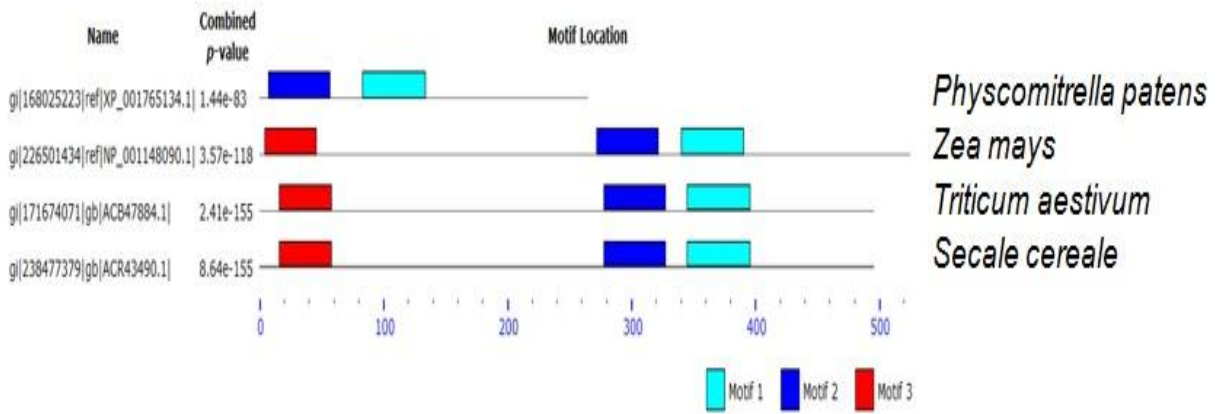
S.N.	Accession No.	Description	E-value	P-value	Length
1	ABB85236.1	Glucosyltransferase [<i>G. max</i>]	9.90E-106	1.62E-106	476
2	BAC78438.1	Isoflavonoid glucosyltransferase [<i>G. echinata</i>]	3.10E-113	5.62E-114	482
3	ACT34898.1	GT3, UDP-glucosyltransferase family protein [<i>M. truncatula</i>]	2.40E-103	3.62E-104	497
4	BAA36410.1	UDP-glucose:flavonoid glycosyltransferase [<i>V. mungo</i>]	7.70E-99	1.00E-99	477
5	XP_002518722.1	UDP-glucosyltransferase, putative [<i>R. communis</i>]	6.70E-121	1.45E-121	461
6	XP_002298737.1	Predicted protein, UDP-glucosyltransferase family protein [<i>P. trichocarpa</i>]	7.40E-116	1.54E-116	491
7	ACJ72160.1	UGT3 [<i>P. montana</i> var. <i>lobata</i>]	1.50E-135	4.77E-136	475
8	CAB88666.1	Putative UDP-glycose [<i>C. arietinum</i>]	2.80E-96	3.48E-97	438
9	ABB92749.1	UDP-glucose glucosyltransferase [<i>F. x ananassa</i>]	2.30E-123	5.86E-124	487
10	XP_002456026.1	Hypothetical protein SORBIDRAFT_03g029080 UDP-glucosyltransferase family protein [<i>S. bicolor</i>]	2.10E-103	3.11E-104	491
11	BAF75890.1	Tetrahydroxychalcone glucosyltransferase [<i>D. caryophyllus</i>]	8.00E-124	1.19E-124	483
12	BAG71127.1	Glucosyltransferase [<i>P. americana</i>]	1.90E-133	5.65E-134	485
13	BAG80549.1	UDP-glucose:glucosyltransferase [<i>L. barbarum</i>]	5.60E-108	9.25E-109	485
14	ACD03255.1	UDP-glycosyltransferase UGT703A5 [<i>A. strigosa</i>]	7.70E-104	1.18E-104	502
15	AAK28303.1	Phenylpropanoid:glucosyltransferase 1 [<i>N. tabacum</i>]	3.10E-136	1.05E-136	476
16	CAN65903.1	Hypothetical protein, GT1_Gtf_like [<i>V. vinifera</i>]	1.20E-125	2.63E-126	482
17	BAG31950.1	UGT73A9 [<i>Antirrhinum majus</i>]	1.30E-127	3.25E-128	481
18	BAF49308.1	Putative glycosyltransferase [<i>E. grandiflorum</i>]	1.40E-123	3.48E-124	482
19	BAA83484.1	UDP-glucose: flavonoid 7-O-glucosyltransferase [<i>S. baicalensis</i>]	6.50E-130	1.83E-130	476
20	AAS55083.1	UDP-glucose glucosyltransferase [<i>R. sachalinensis</i>]	4.30E-131	1.27E-131	480
21	ACO44747.1	UDP-glycosyltransferase [<i>W. somnifera</i>]	1.10E-131	3.36E-132	470
22	BAD29722.1	UDP-glucose glucosyltransferase [<i>C. roseus</i>]	2.60E-123	6.72E-124	487
23	BAG31952.1	UGT73A13 [<i>P. frutescens</i>]	2.50E-123	6.06E-124	479
24	BAI63589.1	UDP-glucose glucosyltransferase [<i>L. japonicus</i>]	2.40E-122	5.24E-123	491
25	NP_181215.1	UDP-glucuronosyl/UDP-glucosyl transferase family protein [<i>A. thaliana</i>]	2.80E-99	3.83e_100	496
26	AAS94329.1	UDP-glucose:flavonoid-O-glucosyltransferase [<i>B. vulgaris</i>]	5.40E-129	1.61E-129	476
27	CAA59450.1	UDP-glucosyltransferase family protein [<i>S. lycopersicum</i>]	1.60E-126	4.39E-127	466

28	CAB56231.1	Betanidin-5-O-glucosyltransferase [<i>D. bellidiformis</i>]	2.30E-119	5.20E-120	489
29	ACB47884.1	UDP-glucuronosyl/UDP-glucosyl transferase protein [<i>T. aestivum</i>]	3.90E-108	6.51E-109	496
30	ACR43490.1	UDP-glucosyl transferase [<i>S. cereale</i>]	4.10E-104	6.58E-105	496
31	EAY74811.1	Hypothetical protein OsI_02703, UDP-glucosyltransferase family protein [<i>O. sativa</i> Indica Group]	1.10E-104	1.64E-105	494
32	AAR06917.1	UDP-glycosyltransferase 73E1 [<i>S. rebaudiana</i>]	3.70E-102	6.04E-103	495
33	ACS87992.1	UDP-glycosyltransferase family 1 protein [<i>C. sinensis</i>]	2.40E-111	4.61E-112	504
34	BAI65914.1	UDP-sugar:glycosyltransferase [<i>F. x intermedia</i>]	7.30E-84	6.61E-85	486
35	BAI65915.1	UDP-sugar:glycosyltransferase [<i>A. sylvestris</i>]	2.20E-83	2.01E-84	485
36	ABR17572.1	Unknown, Glycosyltransferase_GTB_type [<i>P. sitchensis</i>]	4.60E-85	4.38E-86	498
37	XP_001765134.1	Predicted protein Glycosyltransferase_GTB_type [<i>P. patens</i> subsp. patens]	2.10E-66	1.39e_67	265
38	AAP88405.1	Flavonoid glucosyl-transferase [<i>A. cepa</i>]	3.90E-93	4.43E-94	487
39	NP_001148090.1	Cytokinin-O-glucosyltransferase 3 [<i>Z. mays</i>]	3.70E-89	4.10E-90	525
40	BAF96582.1	Lignan glucosyltransferase [<i>S. indicum</i>]	3.20E-81	2.80E-82	476

Supplementary Figures



Supplementary Fig 1. Combined block diagram for all the motifs, constructed using set I sequences of 14 *A. thaliana* sequences, showing its occurrence and location. One putative glycosyltransferase of *A. thaliana* did not exhibit motif 3.



Supplementary Fig 2. Combined block diagram for all the motifs, constructed from set II of 4 most diversified sequences, showing its occurrence and location. Glycosyl transferase (265 amino acids) of *P. patens subsp. Patens* (XP_001765134.1) did not exhibit motif 3, and considerable shift in both the motif towards N-terminal was noticed.