



**Supplementary Information, Figure S2.** Unrooted tree dendrogram comparison of the amino acid sequences of glycosyltransferases.

Peptide sequences of these UDP-glycosyltransferase encoding genes and reference sequences with UDP-glycosyltransferase activity were aligned using MAFFT [1]. The output of MAFFT result was used to create phylogenetic tree by MEGA 5.2 using Neighbor-Joining method, with Poisson correction as the model [2]. The bootstrap confidence values were obtained based on 1000 replicates. Functionally characterized plant glycosyltransferases were specially colored. The triterpene glucosyltransferases were labeled with green; the flavonoid-3-*O*-UGTs were labeled with orange; the sterol UGTs were labeled with blue; the anthocyanin 5-*O*-UGTs were labeled with brown, the 16 cloned *Panax ginseng* glycosyltransferases cloned from *P. ginseng* were labeled with red and the UGTPg1 was specially labeled with an asterisk, the flavonol 3-*O*-UGTs were labeled with purple and other functionally characterized glycosyltransferases were labeled with black. In particular, as shown in the databases of CAZy and/or UniProtKB/Swiss-Prot (**Supplementary Information, Table S3**), one UGT from *Saponaria vaccaria* (*Vaccaria hispanica*) [3] and three from the model legume *Medicago truncatula* [4] are testified to glycosylate the C-3, C-23 or C-28 hydroxyl of pentacyclic triterpene. UGTPgs (9, 8, and 1) clustered within two of these

UGTs are the best candidates for screening of tetracyclic triterpene specific glycosyltransferases.

### References

- 1 Katoh K, Misawa K, Kuma Ki, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002; **30**:3059-3066.
- 2 Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987; **4**:406-425.
- 3 Meesapyodsuk D, Balsevich J, Reed DW, Covello PS. Saponin biosynthesis in *Saponaria vaccaria*. cDNAs encoding  $\beta$ -amyrin synthase and a triterpene carboxylic acid glucosyltransferase. *Plant Physiol* 2007; **143**:959-969.
- 4 Achnine L, Huhman DV, Farag MA, Sumner LW, Blount JW, Dixon RA. Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume *Medicago truncatula*. *Plant J* 2005; **41**:875-887.