African Journal of Biotechnology Vol. 10(21), pp. 4339-4346, 23 May, 2011
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB10.2015
ISSN 1684–5315 © 2011 Academic Journals

Full Length Research Paper

Bioinformatic data mining on UDP-glucose:flavonoid 7-O-glucosyltransferase (*UBGAT*) genes and their encoding proteins in two plants of genus *Scutellaria*

Wei Lei[1, 2], Jidong Zhang[1,2], Ruixia Yao[1] and Aimin Qiao[1]*

[1]College of Horticulture and Landscape Architecture, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China.
[2]Key Laboratory of Eco-environments in three Gorges Reservoir Region of Ministry of Education, College of Life Science, Southwest University, Chongqing 400715, China.

Accepted 1 April, 2011

The GenBank sequences of *UBGAT* (UDP-glucose:flavonoid 7-O-glucosyltransferase) genes and their encoding proteins from *Scutellaria baicalensis* and *Scutellaria laeteviolacea*, were predicted and analyzed by some bioinformatics methods and tools such as expasy, vector NTI, National Center for Biotechnology Information (NCBI) etc. The results obtained are as follows: both *Sbubgat* and *Slubgat* genes contained complete open reading frame (ORF); the formula of SbUBGAT and SbUBGAT were $C_{2193}H_{3462}N_{590}O_{638}S_{11}$ and $C_{2263}H_{3562}N_{596}O_{664}S_{13}$, respectively and both proteins were found in the cytoplasm, without transmembrane topological structure, and were hydrophic proteins related to intermediary metabolism and amino acid synthesis; the secondary structure of UBGATs constitute mainly α-helix and random coil and the tertiary structure were modeling. In brief, this study lay theoretical foundation for metabolic mechanism and genetic regulation researches of flavonoids biosynthesis.

Key words: *Scutellaria baicalensis*, *Scutellaria laeteviolacea*, flavonoids, bioinformatics, UDP-glucose:flavonoid 7-O-glucosyltransferase.

INTRODUCTION

Glycosylation is the enzymatic process that attaches glycans to proteins, lipids or other organic molecules. It produces one of the fundamental biopolymers along with DNA, RNA, and proteins and is a form of co-translational and post-translational modification (Ajit, 2009). Hence, as far as the plant secondary metabolites is concerned, glycosylation was considered as a key step at the end of their biosynthetic pathway and this reaction is catalyzed by a uridine diphosphate (UDP)-glucose: glucosyl-transferase to form the stable water-soluble compound transported to the vacuole (Reay and Conn, 1974; Poulton, 1990; Sun and Hrazdina, 1991; Stapleton et al., 1992). Flavonoid, as a widely existed secondary metabolite, play an important role in plant adaptive responses to physiological function and environment challenges such as flower coloration, UV protection and disease defense (Stefan and Axel, 2005). The glycosides of the favonoids are extensively distributed in the form of the glycosides of glucose, galactose and xylose. The glucose is one of the most common sugar residue and the glucosyl moiety is transferred from UDP-glucose to the favonoid acceptor by the enzyme UDP-glucose:favonoid glucosyltransferase (Masao et al., 2000 Noguchi et al., 2009).

Scutellaria plants are the perennial herb distributed mainly in Asia and its roots are widely used in treating

*Corresponding author. E-mail: qiaoam@126.com or thdmast@gmail.com. Tel: +86-20-89003155. Fax: +86-20-89003155.

Abbreviations: *UBGAT*, Uridine diphosphate-glucose:flavonoid 7-O-glucosyltransferase*;* **CDS,** coding regions; **UTR,** untranslated region; **CDD,** conserved domain database; **NJ,** neighbor-joining; **ME,** minimum evolution.

Table 1. Molecular structure and physicochemical properties.

index	SbUBGAT	SlUBGAT
Formula	$C_{2193}H_{3462}N_{590}O_{638}S_{11}$	$C_{2263}H_{3562}N_{596}O_{664}S_{13}$
Amino acid residues	441	455
Molecular weight (kDa)	48.6538	50.1595
PI	5.99	5.38
Molar extinction coefficient	39670	37930
Estimated half-life	30h	30h
Instability index	40.99 (unstable)	35.77(stable)
Aliphatic index	94.65	94.70
Negatively charged residues	54	58
Positively charged residues	48	46
Hydrophobic amino acids (%)	39.00	39.12
GRAVY	-0.088	-0.022

inflammatory and bacterial diseases as the staple Chinese herbal medicine. Generally, *Scutellaria baicalensis* and *Scutellaria laeteviolacea* are both the traditional medicinal varieties. Recent pharmacological studies revealed that, their main active ingredients are flavornoids, including baicalin, baicalein, wogonoside and wogonin, which are synthesized with the help of many enzymes involved in the flavonoid biosynthetic pathway. Thereinto, UDP-glucose: flavonoid 7-O-glucosyltransferase (*UBGAT*) is one of the most important enzymes, which catalyzes the transformation of glucuronic acid from UDP-glucuronic acid to 7-OH of baicalein (Masao et al., 2000; Noguchi et al., 2009). Some studies showed that, SbUBGAT and SlUBGAT (SlUGT88D5) played the crucial role in the officinal composition accumulation in *S. baicalensis* and *S. laeteviolacea*, respectively (Wang et al., 2006; Noguchi et al., 2009). However, little information is available about the enzyme in *Scutellaria* so far.

In the present study, we completed the bioinformatic analysis of *UBGAT* genes and their encoding protein from *S. baicalensis* and *S. laeteviolacea*. The results will give theoretical foundation for molecular mechanism and genetic regulation researches of flavonoid biosynthesis and the development and utilization of their correlative medicinal resource plants in the future.

MATERIALS AND METHODS

The sequences with the complete coding regions (CDS) of *Sbubgat* and *Slubgat* genes and their corresponding amino acid sequences were obtained from GenBank (SbUBGAT: Accession no. AB042277; SlUBGAT: Accession no. BAG31946).

Comparative bioinformatic analyses of target sequences were performed online at the websites (http://www.ncbi.nlm.nih.gov and http://www.expasy.org). Molecular structures and physicochemical properties were obtained by Protparam tools. Multiple alignment analysis, based on the full-length amino acid sequences, was performed with vector NTI suite 8 using default parameters (Lei et al., 2009). The subcellular location was predicted by TargetP 1.1 Server (Olof et al., 2000; Kristin and Siegfried 2004). The trans-membrane helices and hydro-phobicity in protein were analyzed by ProtFun 2.2 Server (Jensen et al., 2002; Jensen at al., 2003), TMHMM Server v.2.0 (Ikeda et al., 2002) and ProtScale (Kyte and Doolittle, 1982), respectively. The motifs and structural domains in protein were searched by PrositeScan (Combet et al., 2000) and CDD (Marchler et al., 2004). Target proteins and their related sequences from other species were aligned with ClustalX (Thompson et al., 1997) and subsequent-

tly, a phylogenetic tree was constructed by neighbor-joining method with 1000 replicates and another tree was recon-structed by maximum-likelihood method with 1000 replicates and reliability of each node was established by bootstrap methods using MEGA4.1 software, respectively (Saito et al., 1987; Kumar et al., 2001). The secondary structures of two UBGAT proteins were predicted by GOR4 online tool (Combet et al., 2000). And the homology-based three-dimensional (3D) structural modeling of UBGAT proteins was accomplished by Swiss-modeling (Guex and Peitsch, 1997; Schwede et al., 2003; Arnold et al., 2006) and accelrys viewerlite 4.2 was used for 3D structure editing.

RESULTS AND DISCUSSION

Analysis of molecular structure and physicochemical properties

Structures and properties of nucleotide acid and corresponding amino acid sequences of *Sbubgat* and *Slubgat* genes, designated as *Subgats*, were obtained by vector NTI suite 8 software and Protparam online server (Table 1). *Sbubgat* gene contained complete ORF and 5' untranslated region (UTR) but lacking 3' UTR and there was only ORF region in *Slubgat* sequence. Some physicochemical indices of *Subgats* encoding protein SbUBGAT and SlUBGAT, incorporately designated as SUBGATs, were also computed and the results were characterized in Table 1, including the formula, molecular weight, isoelectric point (PI), molar extinction coefficient, estimated half-life, instability index, aliphatic index, grand average of hydropathicity (GRAVY) and total number of negatively and positively charged residues. Hence, it was pertinent to determine that they are a group gene with significant functional association and close genetic relation.

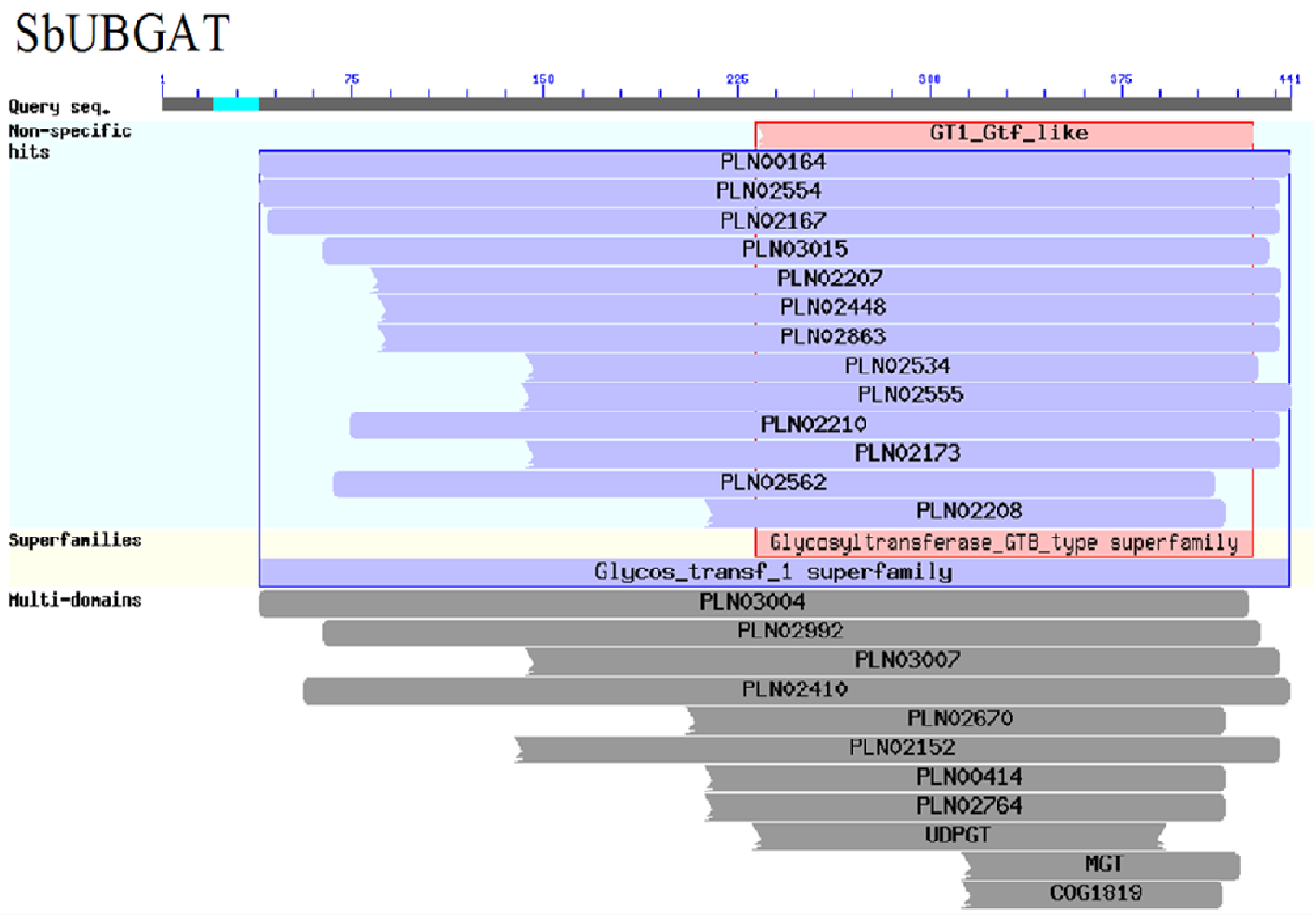

Figure 1. Conserved domains analysis of SbUBGAT and SlUBGAT proteins.

Subcellular localization and functions analysis

TMHMM v2.0 server online predicted that, there was no transmembrane topological structure in SUBGATs proteins, implying that they may exert catalytic function directly in cytosol without transportation. With the help of TargetP 1.1 tool, the subcellular localization and trans-membrane helix prediction suggested that, SUBGATs were both localized in the cytoplasmic matrix without transit peptide, indicating that the enzyme drove directly the substrates glycosylation reaction and the flavonoid compounds biosynthesis, which just corresponded with Geza Hrazdina's research report that, flavonoids were synthesized in the cytoplasmic matrix (Hrazdina, 1992).

ProFun 2.2 server online analysis indicated that the cellular function of both SUBGATs may have some relation to central intermediary metabolism and this laid some evidence to the subcellular localization prediction of SUBGATs, because central intermediary metabolite occurred usually within the cell.

Furthermore, protein may contain some motifs, whose corresponding amino acid regions carry out specific biochemical functions and meanwhile, include respective genetic evolutional information. The tool PrositeScan recognized the presence of an N-glycosylation site in each SUBGATs protein, playing the important role in recognizing and binding between the enzymes and their own substrate. And then the tool conserved domain database (CDD) recognized the present of conserved domains in these proteins, as shown in Figure 3, a series of functional domains were detected, such as catalytic residues, substrate binding pocket and active site, etc. (Figure 1).

Multiple alignment and molecular phylogram analysis

The amino acid sequences of SbUBGAT, SlUBGAT and UBGATs from other nine species of plants were aligned according to the algorithm of the Vector NTI Suite 8. The

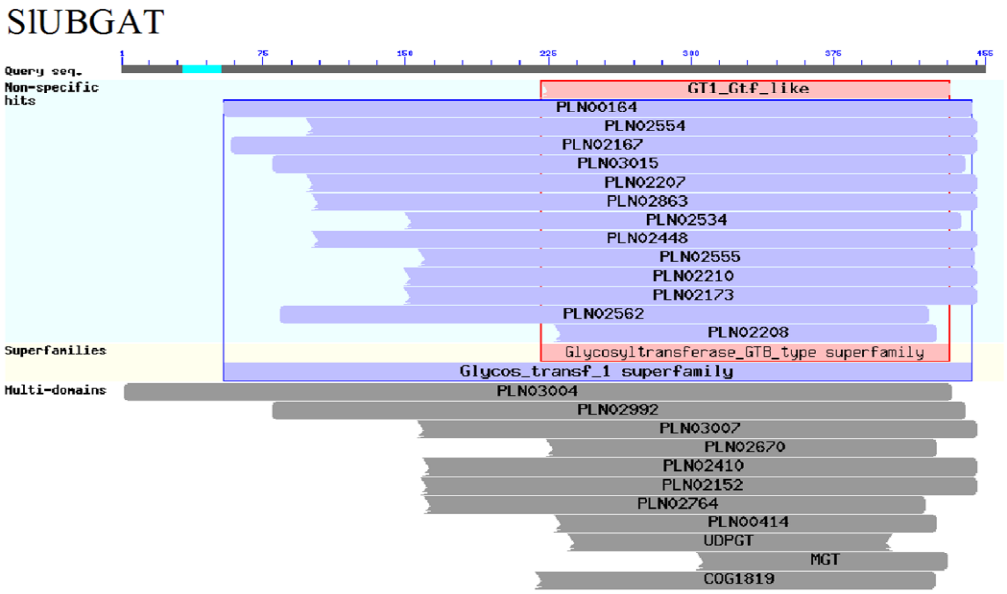

Figure 1. Contd.

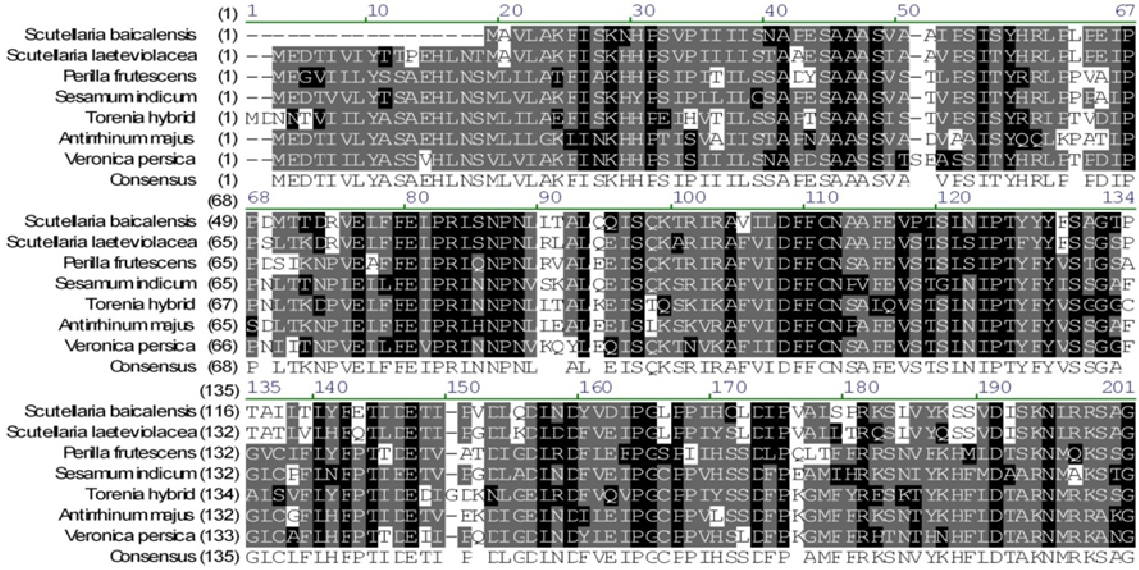

Figure 2. The multiple alignment of amino acid sequences of SbUBGAT, SlUBGAT and other plant UBGATs. The identical sites are shown in white letters and black background; the conservative sites are shown in white letters and gray background; other sites were all shown in black letters and white background. The database accession numbers of the sequences used for the phylogenic analysis: *Scutellaria baicalensis* (GenBank accession no.: AB042277), *Scutellaria laeteviolacea* (GenBank accession no.: BAG31946), *Sesamum indicum* (GenBank accession no.: BAG31947), *Veronica persica* (GenBank accession no.: BAH47552), *Perilla frutescens* (GenBank accession no.: BAG31948), *Torenia hybrid* (GenBank accession no.: BAH14961), *Antirrhinum majus* (GenBank accession no.: BAG31945).

result showed that, these UBGATs had relatively high similarity throughout entire coding region and some highly conserved residues, especially N-glycosylation site (425 to 428) and an N-myristoylation site, were possibly catalytic domains (Figure 2).

After the multiple alignments by the software ClustalX, two phylogenetic trees were constructed in parallel with the neighbor-joining (NJ) method and minimum evolution (ME) method, respectively (Figure 3). It is most intriguing that three clusters were composed at the evolutionary

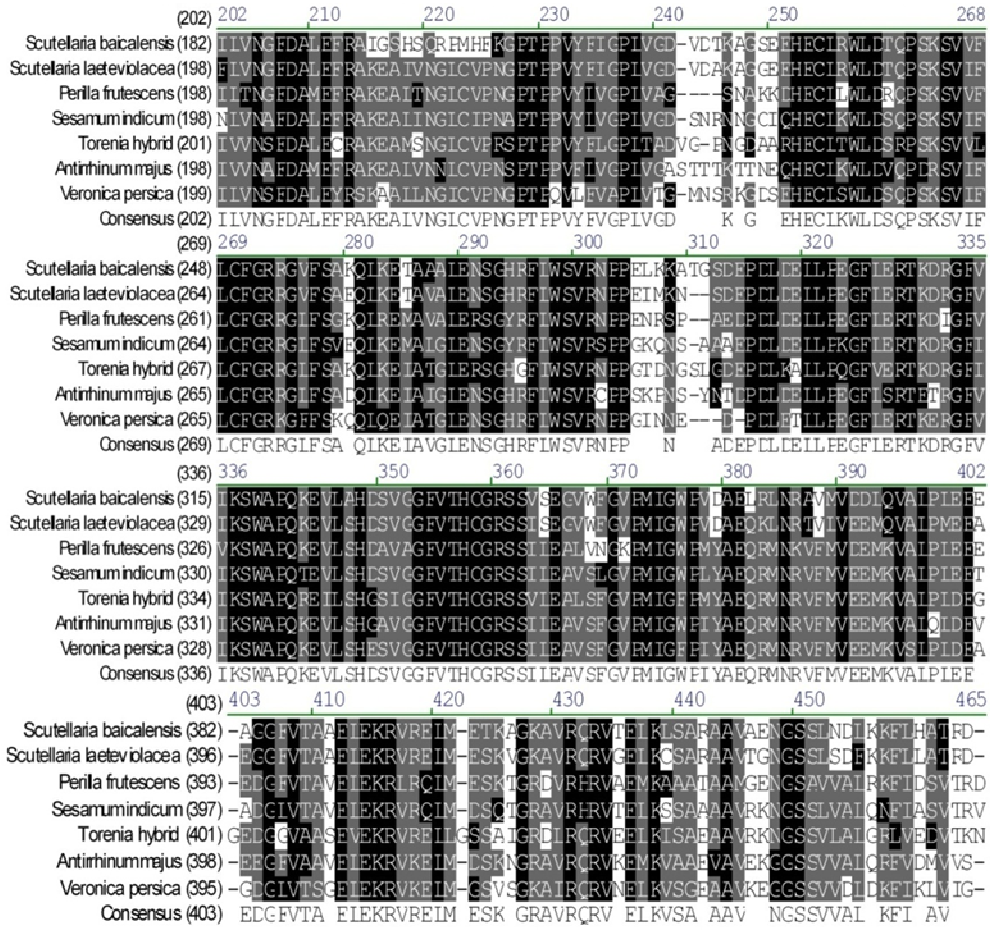

Figure 2. Contd.

trees and meanwhile, SbUBGAT had the closest relation with SlUBGAT, because both of them belonged to *Scutellaria* family. In addition, genetic distance was very small between SUBGATs and medicinal plants as well as ornamental plants, which contain abundant flavonoid compounds. To sum up, the data suggested that, *UBGAT* genes from the various plants had somewhat different expression to meet the diverse requirements of flavonoids function and furthermore, *UBGAT* genes of different plants might have originated from the same ancestor.

Prediction of secondary and tertiary structures

The secondary structure of two SUBGATs proteins were predicted by GOR4 online tool. Random coil, α-helix and extended strand shared 50.34, 35.37 and 14.29% in SbUBGAT, respectively. And there was similar composition proportion in SlUBGAT. It was extrapolated that, the first one was the main constitution and extended strand intersperse in the whole protein. The very high coil structures are attributed to abundance in flexible glycine and hydrophobic praline amino acids, which can effectively create links in polypeptide chains and disrupting ordered secondary structure (Figure 4).

The homology-based 3D structural models of the target SUBGATs was constructed by Swiss-modeling based on *Medicago truncatula* UBGAT crystal structure and displayed by accelrys viewerlite (Figure 5). Taken together, there were strong evidences that, UBGAT proteins from *S. baicalensis* and *S. laeteviolacea* have very typical molecular structure and biological function of glycosylation enzyme.

In conclusion, as the rate-limiting enzyme involved in flavonoids biosynthetic pathway, UDP-glucose:flavonoid 7-O-glucosyltransferase is always paid much attention. In this study, molecular structures and biochemical functions of UBGAT from genus *Scutellaria* plants, including *S. baicalensis* and *S. laeteviolacea*, were predicted and analyzed by some bioinformatics tools. This study will throw important theoretical references for enzymology properties of UBGAT protein and molecular mechanism of flavonoids biosynthesis and become a successful application of computer skills and information technology on analysis of ometabolic engineering of traditional Chinese medicines.

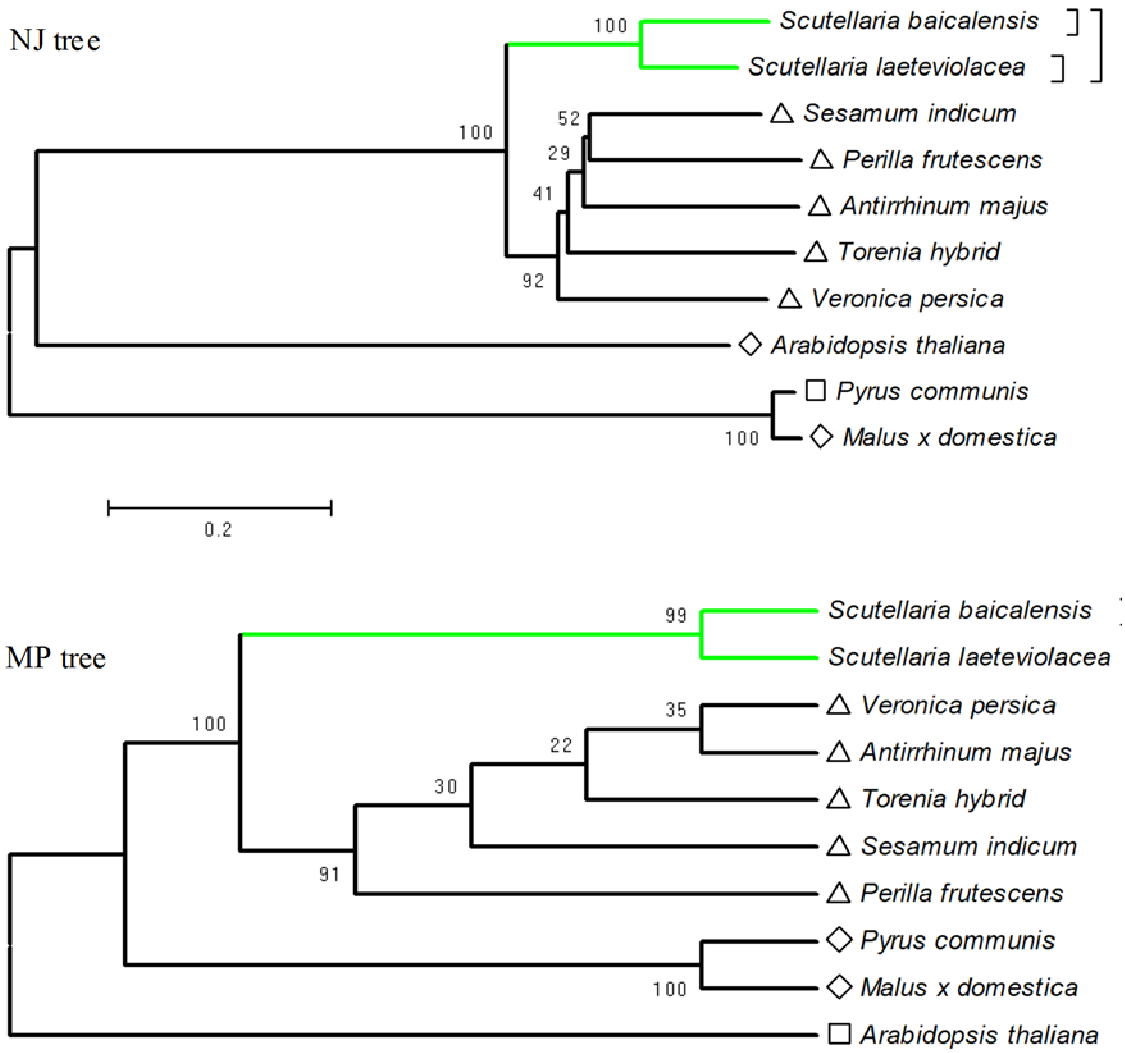

Figure 3. Phylogenetic tree analyses of SbUBGAT, SlUBGAT and UBGATs from other plants. The database accession numbers of the sequences used for the phylogenic analysis: *S. baicalensis* (GenBank accession no.: AB042277), *S. laeteviolacea* (GenBank accession no.: BAG31946), *S. indicum* (GenBank accession no.: BAG31947), *V. persica* (GenBank accession no.: BAH47552), *P. frutescens* (GenBank accession no.: BAG31948), *T. hybrid* (GenBank accession no.: BAH14961), *A. majus* (GenBank accession no.: BAG31945).

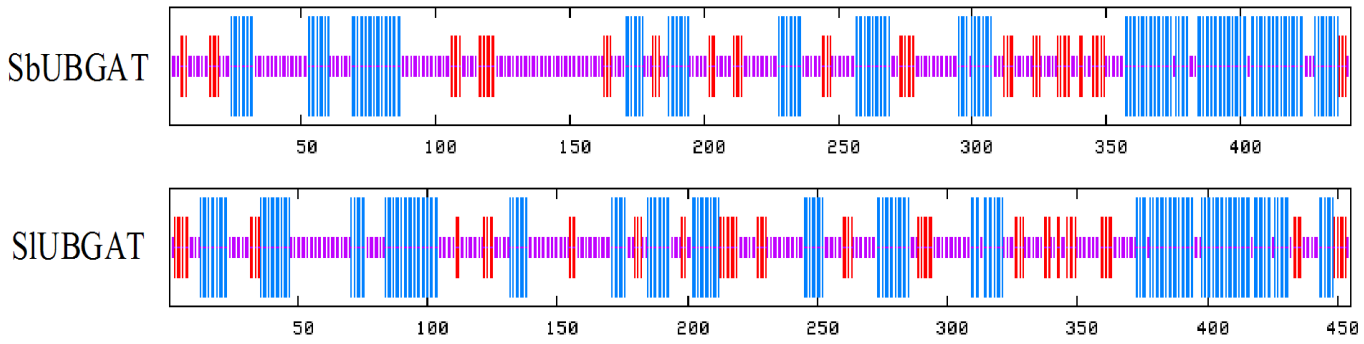

Figure 4. The secondary structure model of SbUBGAT and SlUBGAT. The α-helix and extended strand were indicated as and , respectively. Random coil was indicated as .

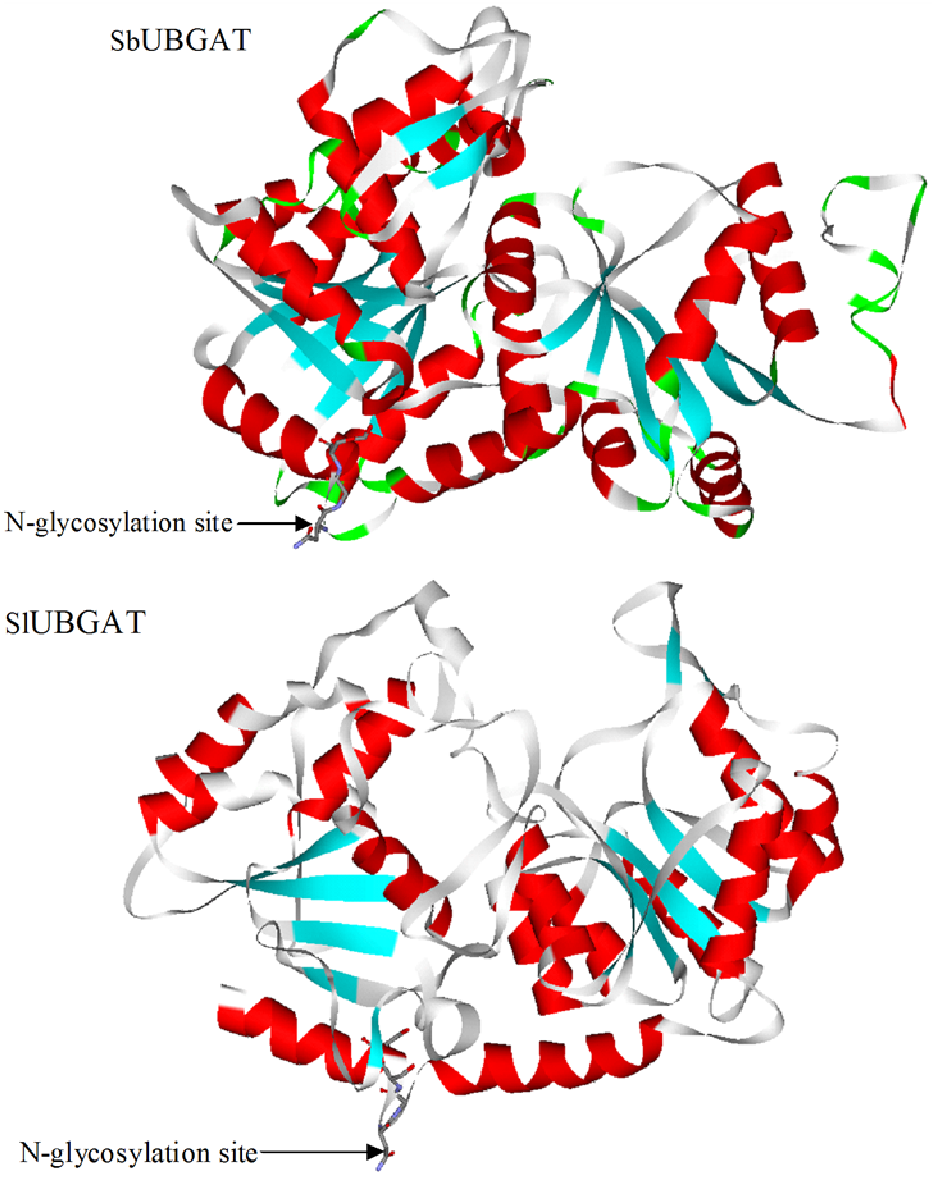

Figure 5. The 3D structural model of SbUBGAT and SlUBGAT. The α-helix and extended strand were indicated in red and blue, respectively. Random coil was indicated in silver.

ACKNOWLEDGEMENT

We are grateful to the reviewers for their valuable suggestions and comments on the manuscript.

REFERENCES

Ajit V (2009). Essentials of glycobiology (2nd ed.). New York: Cold Spring Harbor Laboratories Press.
Arnold K, Bordoli L, Kopp J, Schwede T (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling. Bioinformatics, 22: 195- 201.
Combet C, Blanchet C, Geourjon C, Deléage G (2000). NPS@: Network protein sequence Analysis. TIBS, 25: 147- 150.
Guex N, Peitsch MC (1997). SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis, 18: 2714- 2723.
Hrazdina G (1992). Compartmentation in aromatic metabolism, In: Stafford HA, Ibrahim RK, eds. Recent advances in phytochemistry. New York, Plenum Press.
Ikeda M, Arai M, Lao DM (2002). Transmembrane topology prediction methods: A reassessment and improvement by a consensus method using a dataser of experimentally characterized transmembrance topologies. Silico Biol. 2: 19- 33.
Jensen JL, Gupta R, Blom N, Devos D, Tamames J, Kesmir C, Nielsen H, Staerfeldt H, Rapacki K, Workman C (2002). Ab initio prediction of human orphan protein function from post-translational modifications and localization features. J. Mol. Biol. 319: 1257- 1265.
Jensen JL, Stærfeldt HH, Brunak S (2003). Prediction of human protein function according to Gene Ontology categories. Bioinformatics. 19: 635- 642.
Kristin E, Siegfried H (2004). InvB is required for type III-dependent secretion of sopA in *Salmonella enterica* serovar typhimurium. J.

Bacteriol. 186: 1215-1219.
Kumar SK., Tmamura K, Jakobseni IB, Nei M (2001). MEGA2: Molecular evolutionary genetics analysis software. Bioinformatics, 17: 1244-1245.
Kyte J, Doolittle RF (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157: 105-132.
Lei W, Sun M, Luo KM, Shui XR, Sun YM, Tang SH (2009). Compute simulations to characterize chalcone synthase from *Scutellaria baicalensis* Georgi. Mol. Biol. 43(6): 1012-1017.
Marchler BA, Bryant SH (2004). CD-Search: protein domain annotations on the fly. Nucleic Acids Res. 32: w327-w331.
Masao H, Ryo K, Hideyuki S, Takafumi Y (2000). Cloning and expression of UDP-glucose:favonoid 7-O-glucosyltransferase from hairy root cultures of *Scutellaria baicalensis*. Planta, 210: 1006- 1213.
Noguchi A, Kunikane S, Homma H, Liu WH, Sekiya T, Hosoya M, Kwon S, Ohiwa S, Katsuragi H, Nishino T, Takahashi S, Hamada H, Nakayama T (2009). Identification of an inducible glucosyltransferase from *Phytolacca americana* L. cells that are capable of glucosylating capsaicin. Plant Biotechnol. 26: 285-292.
Olof E, Henrik N, Søren B. Gunnar H (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. 300: 1005-1016.
Poulton JE (1990). Cyanogenesis in plants. Plant Physiol. 94: 401-405.
Reay PF, Conn EE (1974). The purification and properties of a uridine diphosphate glucose: aldehyde cyanohydrin β-glucosyl-transferase from sorghum seedlings. J. Biol. Chem. 249: 5826- 5830.
Saito N, Nei M (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4: 406-425.
Schwede T, Kopp J, Guex N, and Peitsch MC (2003). SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res. 31: 3381-3385.
Stapleton A, Allen PV, Tao HP, Belknap WR, Friedman M (1992). Partial amino acid sequence of potato solanidine UDP-glucose glucosyltransferase purified by new anion-exchange and size exclusion media. Prot. Express Purific. 3: 85-92.
Stefan M, Axel M (2005). Flavones and flavone synthases. Phytochemistry, 66: 2399-2407.
Sun Y, Hrazdina G (1991). Isolation and characterization of a UDP glucose: favonol O^3-glucosyltransferase from illuminated red cabbage (*Brassica oleracea* cv red danish) seedlings. Plant Physiol. 95: 570-576.
Thompson JD, Gibson TJ, Plewniaki F, Jeanmougin F, Higgins DG (1997). The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25: 4876-4882.
Wang W, Chen JK, Zhou TS (2006). CHS and UBGAT expression and baicalin accumulation in the roots of *Scutellaria baicalensis* Georgi during cultivation seasons. J. Fudan University (Natural Science). 45(5): 674- 678.