

## Expressed sequence tags from rhizomes of *Glycyrrhiza uralensis*

Hiroshi Sudo<sup>1,2,\*</sup>, Hikaru Seki<sup>3,4</sup>, Nozomu Sakurai<sup>5</sup>, Hideyuki Suzuki<sup>5</sup>,  
Daisuke Shibata<sup>5</sup>, Atsushi Toyoda<sup>6</sup>, Yasushi Totoki<sup>6</sup>, Yoshiyuki Sakaki<sup>6</sup>, Osamu Iida<sup>7</sup>,  
Toshiro Shibata<sup>8</sup>, Mareshige Kojoma<sup>9</sup>, Toshiya Muranaka<sup>3,4</sup>, Kazuki Saito<sup>2,4</sup>

<sup>1</sup> Tokiwa Phytochemical Co., Ltd., Sakura, Chiba 285-0801, Japan; <sup>2</sup> Graduate School of Pharmaceutical Sciences, Chiba University, Chiba 263-8522, Japan; <sup>3</sup> Kihara Institute for Biological Research, Yokohama City University, Yokohama, Kanagawa 244-0813, Japan; <sup>4</sup> RIKEN Plant Science Center, Yokohama, Kanagawa 230-0045, Japan; <sup>5</sup> Kazusa DNA Research Institute, Kisarazu, Chiba 292-0818, Japan; <sup>6</sup> RIKEN Genomic Sciences Center, Yokohama, Kanagawa 230-0045, Japan; <sup>7</sup> Tanegashima Division, Research Center for Medicinal Plant Resources, National Institute of Biomedical Innovation, Kumage-gun, Kagoshima 891-3604, Japan; <sup>8</sup> Hokkaido Division, Research Center for Medicinal Plant Resources, National Institute of Biomedical Innovation, Nayoro, Hokkaido 096-0065, Japan; <sup>9</sup> Faculty of Pharmaceutical Sciences, Health Sciences University of Hokkaido, Kanazawa, Hokkaido 061-0293, Japan  
\* E-mail: mesemb@gmail.com Tel: +81-43-498-0079 Fax: +81-43-498-2392

Received November 5, 2008; accepted November 25, 2008 (Edited by S. Sato)

**Abstract** *Glycyrrhiza uralensis* (Leguminosae, Fabaceae) is the most important medicinal plant used as a natural sweetener, in Chinese herbal medicines, and as a tobacco flavoring agent. Here, we report the generation of 56,857 expressed sequence tags (ESTs) derived from two full-length cDNA libraries produced from rhizomes (stolons) of two strains of *G. uralensis*, together with differential expression data obtained by Megasort™ analysis. The Megasort™ analysis was performed using two plant organs, namely, thickened roots and a rhizome-like organ, which differed in their glycyrrhizin content. After clustering of the 5′-end of ESTs, 5,542 contigs and 4,932 singletons were generated. A total of 10,474 nonredundant sequences were annotated on the basis of the TAIR database, of which 7,905 (75.5%) sequences exhibited homology to the sequences of registered genes. The gene functions of these 7,905 sequences were classified into Biological Process and Molecular Function on the basis of Gene Ontology (GO).

**Key words:** Expressed sequence tags, Fabaceae, gene ontology, *Glycyrrhiza uralensis*, glycyrrhizin, Leguminosae.

Licorice is the dried roots and rhizomes (stolons) of *Glycyrrhiza uralensis*, *G. glabra*, and *G. inflata*—leguminous plants native to northern Asia, the Middle East, and southern Europe. The use of licorice dates back thousands of years, and it is widely used as a crude drug, a natural sweetener for foods, and a flavoring agent for American tobacco. In Japan, although *G. uralensis* has been traditionally used to produce a crude drug (Shibata et al. 2003), there are no cultivars or selected strains, and the actual supply of licorice depends on imported materials from China, Russia, Afghanistan, and other countries. The major ingredients of licorice are glycyrrhizin—an oleanane-type triterpenoid saponin—and various flavonoids, particularly liquiritin, isoliquiritin, and their aglycones. To isolate the genes involved in the biosynthesis pathways of secondary metabolites of licorice, we analyzed expressed sequence tags (ESTs) derived from cDNA libraries of *G. uralensis*.

Glycyrrhizin is accumulated predominantly in underground organs (rhizomes and thickened roots) of several *Glycyrrhiza* species. Our extensive chemical

analyses of aboveground organs (leaves, stems, and flowers), however, failed to detect any traces of glycyrrhizin. These observations strongly suggest that rhizome and root tissue are the sites of glycyrrhizin biosynthesis for both accumulation and storage. Two full-length cDNA libraries were constructed independently from the rhizomes of 6- to 7-year-old field-grown *G. uralensis* plants, each cultivated in different areas in Japan. The rhizomes of “Hokkaido-iryodai” strain cultivated in Nayoro, Hokkaido prefecture, were harvested in May, while those of “Kanzo-yashiki” strain cultivated in Tsukuba, Ibaraki prefecture, were harvested in October. The construction of each cDNA library from these two sources was carried out from mRNA following the vector-capping protocol (Kato et al. 2005) of Hitachi Instruments Service Co., Ltd. (Tokyo, Japan) by using the vector pGCAPzf3 (Tsugane et al. 2005) or the oligo-capping protocol (Maruyama et al. 1994; Suzuki et al. 1997) of Toyobo Co., Ltd. (Osaka, Japan) by using the vector pCMVFL3 (Invitrogen, Carlsbad, CA, USA). The average insertion length of the

“Hokkaido-iryodai” cDNA library and the “Kanzoyashiki” cDNA library was 1.5 and 1.2 kb, respectively. Approximately 30,000 or 26,000 sequences were isolated from each cDNA library. Sequencing of the 5'-end of each clone was performed on an automatic DNA sequencer (ABI Prism 3700 or ABI Prism 3730XL, Applied Biosystems, Foster City, USA) by using a BigDye Terminator Cycle Sequencing kit (Applied Biosystems).

The genes involved in secondary metabolism are usually expressed at a low level compared to those involved in primary metabolism. To isolate rare transcripts, Megasort™ analysis was performed at Takara Bio, Inc. (Otsu, Shiga) as described earlier (Brenner et al. 2000). Two samples differing in glycyrrhizin content, namely, thickened roots cultured in a pot for six months and a rhizome/stolon-like organ cultured *in vitro* (Kojoma et al. 2005), were used as mRNA sources. The glycyrrhizin content of the thickened roots and the rhizome-like organ was 0.24% of dry weight and trace, respectively. After competitive hybridization, the beads corresponding to high- and/or low-expressed genes were sorted with a cell sorter. A total of 400 beads corresponding to upregulated genes and 200 beads corresponding to downregulated genes were obtained, and the 5'-end of each clone was sequenced. Finally, we added 454 5'-end sequences (272 upregulated genes and 182 downregulated genes) obtained by the Megasort™ analysis to the ESTs obtained from two cDNA libraries described above. The complete set of ESTs has been submitted to DNA Data Bank of Japan (DDBJ).

The 56,857 sequences derived from the two cDNA libraries and the Megasort™ analysis were assembled into independent clusters by the following process. Vector-derived and ambiguous sequences were eliminated using a combination of the PHRED program (Ewing et al. 1998) and Cross\_match software (<http://www.phrap.org>). Clustering of the resulting EST dataset was performed using the CAP3 sequence assembly program (Huang et al. 1999). Finally, 10,474 unique sequences longer than 200 nt were obtained and compared using the BlastX program with an E-value threshold of  $1 \times 10^{-5}$  against the TAIR8 database provided by The Arabidopsis Information Resource (TAIR, <http://www.arabidopsis.org>). Of the 10,474 unique sequences, 7,905 (75.5%) sequences showed homology to those of registered genes in the TAIR8 database, and the remaining 2,569 sequences had no significant matches. To obtain an overview of gene functions within the ESTs, the annotated 7,905 sequences were categorized according to Gene Ontology (GO) on the basis of AGI codes and TAIR GO slim provided by TAIR. The latest file was downloaded from the ftp site [ftp://ftp.arabidopsis.org/home/tair/Ontologies/Gene\\_Ontology/ATH\\_GO\\_](ftp://ftp.arabidopsis.org/home/tair/Ontologies/Gene_Ontology/ATH_GO_)

GOSLIM.txt on September 2, 2008. The annotation counts and relative frequencies of GO functional categories are summarized as pie charts of Biological Process and Molecular Function (Figure 1). With regard to Biological Process and Molecular Function, the relative frequencies of annotation numbers of each functional category were compared with those of *Arabidopsis thaliana*. No significant difference was observed in the relative frequencies of GO hits between the two plant species (Figure 2). Recently, the genome sequences and chromosomal structure of *Lotus japonicus*, a leguminous plant, has been revealed (Sato et al. 2008). The genome information of the model legume will provide important knowledge for the gene discovery and genetic improvement of *Glycyrrhiza* species.

Glycyrrhizin, a major bioactive component of licorice, belongs to the family of triterpenoid saponin. Three genes, namely, two genes encoding squalene synthase homologs (Hayashi et al. 1999) and one gene encoding  $\beta$ -amyrin synthase (Hayashi et al. 2001) involved in the early stage of glycyrrhizin biosynthesis, were isolated from *G. glabra*. However, no genes involved in the subsequent steps of the conversion of  $\beta$ -amyrin to glycyrrhizin have been isolated. Recently, a new cytochrome P450 monooxygenase (P450) gene, namely, *CYP88D6*, encoding a  $\beta$ -amyrin 11-oxidase involved in the late stage of glycyrrhizin biosynthesis, was isolated by transcript profiling-based selection from our ESTs (Seki et al. 2008). Fifty-four P450s were found in the ESTs by keyword search of the annotations in the TAIR8

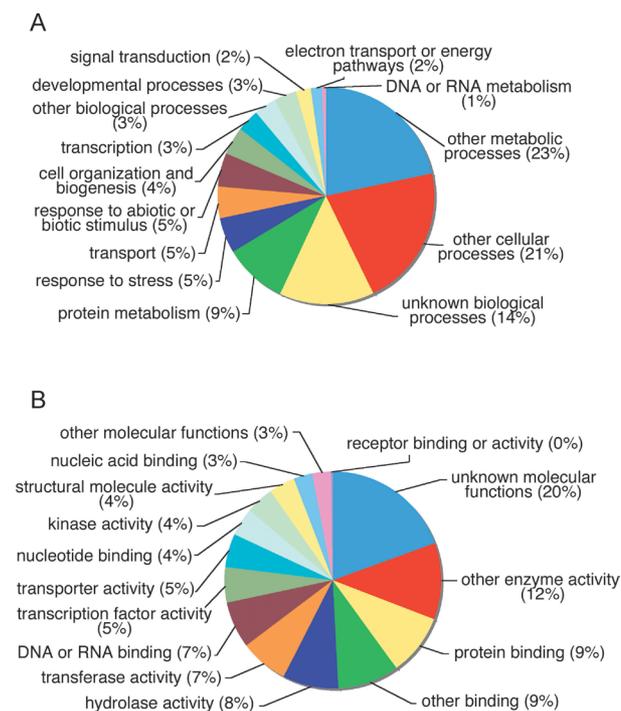


Figure 1. Gene ontology classification of the *Glycyrrhiza uralensis* ESTs. (A) Biological process, (B) Molecular function.

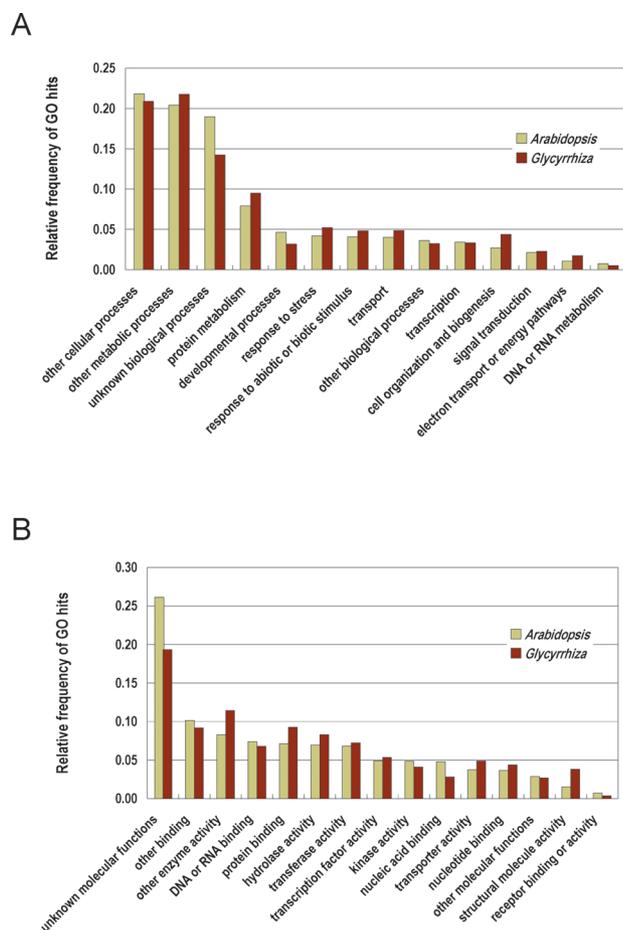


Figure 2. Comparison of GO classification frequencies between the *Glycyrrhiza uralensis* ESTs and *Arabidopsis thaliana* genes. (A) Biological process, (B) Molecular function.

database. The other P450s involved in the oxidation of oleanane skeleton of glycyrrhizin will be searched from our ESTs. The study of microarray data based on the ESTs will reveal the transcriptome profiles, leading to the elucidation of control mechanisms of the secondary metabolism of *G. uralensis*.

### Acknowledgements

This study was supported in part by a grant from the New Energy and Industrial Technology Development Organization (NEDO), Japan, and by a Grant-in-Aid (No. 18780082) for Young Scientists

(B) from the Ministry of Education, Culture, Sports, Science and Technology.

### References

- Brenner S, Williams SR, Vermaas EH, Storck T, Moon K, McCollum C, Mao JI, Luo S, Kirchner JJ, Eletr S, DuBridgde RB, Burcham T, Albrecht G (2000) *in vitro* cloning of complex mixtures of DNA on microbeads: Physical separation of differentially expressed cDNAs. *Proc Natl Acad Sci USA* 97: 1665–1670
- Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8: 175–185
- Hayashi H, Hirota A, Hiraoka N, Ikeshiro Y (1999) Molecular cloning and characterization of two cDNAs for *Glycyrrhiza glabra* squalene synthase. *Biol Pharm Bull* 22: 947–950
- Hayashi H, Huang P, Kirakosyan A, Inoue K, Hiraoka N, Ikeshiro Y, Kushiro T, Shibuya M, Ebizuka Y (2001) Cloning and characterization of a cDNA encoding  $\beta$ -amyrin synthase involved in glycyrrhizin and soyasaponin biosynthesis of *Glycyrrhiza glabra*. *Biol Pharm Bull* 24: 912–916
- Huang X. and Madan A. (1999) CAP3: A DNA sequence assembly program. *Genome Res* 9: 868–877
- Kojoma M, Muranaka T, Yoshida S (2005) Method for culturing tissue of *Glycyrrhiza* plant. *Patent* JP2005137291
- Maruyama K, Sugano S (1994) Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. *Gene* 138: 171–174
- Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, Sasamoto S, Watanabe A, Ono A, Kawashima K, Fujishiro T, Katoh M, Kohara M, Kishida Y, Minami C, Nakayama S, Nakazaki N, Shimizu Y, Shinpo S, Takahashi C, Wada T, Yamada M, Ohmido N, Hayashi M, Fukui K, Baba T, Nakamichi T, Mori H, Tabata S (2008) Genome structure of the legume, *Lotus japonicus*. *DNA Res* 15: 227–239
- Seki H, Ohyama K, Sawai S, Mizutani M, Ohnishi T, Sudo H, Akashi T, Aoki T, Saito K, Muranaka T (2008) Licorice  $\beta$ -amyrin 11-oxidase, a cytochrome P450 with a key role in the biosynthesis of the triterpene sweetener glycyrrhizin. *Proc Natl Acad Sci USA* 105: 14204–14209
- Shibata S, Kondo K, Terabayashi S (2003) Identification of the licorice root stored in Shosoin based on the sequences of internal transcribed spacer (ITS) on nrDNA and the chemotaxonomic consideration. *Proc Jpn Acad, Ser B*, 79: 176–180
- Suzuki Y, Yoshitomo K, Maruyama K, Suyama A, Sugano S (1997) Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library *Gene* 200: 149–156