PCP
PLANT & CELL PHYSIOLOGY

Regular Paper

Functional Characterization of Cucurbitadienol Synthase and Triterpene Glycosyltransferase Involved in Biosynthesis of Mogrosides from *Siraitia grosvenorii*

Longhai Dai, Can Liu, Yueming Zhu, Jiangsheng Zhang, Yan Men, Yan Zeng and Yuanxia Sun*

National Engineering Laboratory for Industrial Enzymes, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin, 300308, China
*Corresponding author: E-mail, syx0430@hotmail.com; Fax, +86-2284861961.
(Received November 26, 2014; Accepted March 2, 2015)

Mogrosides, the major bioactive components isolated from the fruits of *Siraitia grosvenorii*, are a family of cucurbitane-type tetracyclic triterpenoid saponins that are used worldwide as high-potency sweeteners and possess a variety of notable pharmacological activities. Mogrosides are synthesized from 2,3-oxidosqualene via a series of reactions catalyzed by cucurbitadienol synthase (CbQ), Cyt P450s (P450s) and UDP glycosyltransferases (UGTs) in vivo. However, the relevant genes have not been characterized to date. In this study, we report successful identification of *SgCbQ* and *UGT74AC1*, which were previously predicted via RNA-sequencing (RNA-seq) and digital gene expression (DGE) profile analysis of the fruits of *S. grosvenorii*. *SgCbQ* was functionally characterized by expression in the lanosterol synthase-deficient yeast strain GIL77 and was found to accumulate cucurbitadienol as the sole product. *UGT74AC1* was heterologously expressed in *Escherichia coli* as a His-tag protein and it showed specificity for mogrol by transfer of a glucose moiety to the C-3 hydroxyl to form mogroside IE by in vitro enzymatic activity assays. This study reports the identification of CbQ and glycosyltransferase from *S. grosvenorii* for the first time. The results also suggest that RNA-seq, combined with DGE profile analysis, is a promising approach for discovery of candidate genes involved in biosynthesis of triterpene saponins.

Keywords: Cucurbitadienol • Cucurbitadienol synthase • Glycosyltransferase • Mogrosides • *Siraitia grosvenorii*.

Abbreviations: CbQ, cucurbitadienol synthase; DAF, days after flowering; DGE, digital gene expression; GC-MS, gas chromatography–mass spectrometry; LC-ESI-MS, liquid chromatography–electrospray ionisation–mass spectrometry; NGS, next-generation sequencing; NMR, nuclear magnetic resonance; OSC, oxidosqualene cyclase; P450s, Cyt P450s; RNA-seq, RNA-sequencing; SQE, squalene epoxidase; SS, squalene synthase; UGT, UDP glycosyltransferases.

Introduction

Siraitia grosvenorii is a traditional Chinese medicinal plant that is principally cultivated in Guangxi Province of China (Chaturvedula and Prakash 2011). Fruits of *S. grosvenorii* have been used for thousands of years as a natural sweetener, and as folk medicine for treating lung congestion, colds, pharyngitis, constipation and acute bronchitis in the Chinese Pharmacopoeia (Prakash and Chaturvedula 2014). Mogrosides, the major active constituents of *S. grosvenorii*, are a mixture of cucurbitane-type triterpenoid glycosides that have been proven to be powerful and zero-caloric sweeteners and can hence be used as a sucrose substitute for diabetic and obese patients (Xia et al. 2008). In recent years, a considerable number of studies have also reported that mogrosides possess various remarkable pharmacological characteristics, including antioxidative, anti-inflammatory, anticancer, antiviral and antidiabetic properties (Jin and Lee 2012).

Despite the fact that there have been numerous studies on the chemical structure and pharmacological properties of mogrosides, genes involved in the biosynthetic pathway of mogrosides in *S. grosvenorii* or other plants that produce mogrosides have not been characterized. The initial step in the biosynthesis of cucurbitane-type mogrosides is cyclization of 2,3-oxidosqualene to form the triterpenoid skeleton of cucurbitadienol (**Fig. 1**), which is catalyzed by cucurbitadienol synthase (CbQ) that belongs to a family of oxidosqualene cyclases (OSCs). To date, CbQ has been functionally characterized from *Cucurbita pepo* (Shibuya et al. 2004), *Citrullus colocynthis* (Davidovich-Rikanati et al. 2015) and *Cucumis sativus* (Shang et al. 2014). Judging from the structure of mogrosides and research results for other triterpene saponins, cucurbitadienol can be converted to mogrol by a series of hydroxylation reactions by specific Cyt P450s (P450s) at positions C-11, C-24 and C-25. The C-3 and C-24 hydroxyl groups of the mogrol aglycone can be glycosylated by UDP glycosyltransferases (UGTs) to form the corresponding 3- or 24-monodesmosides and 3, 24-bidesmoside. Further glycosylation at C-2′ and C-6′ of 3-*O*-glucose and 24-*O*-glucose produces a mixture of mono-, di-, tri-, tetra-, penta- and hexaglycosides. However, the exact order of the intermediate products remains unknown.

UGT genes are a supergene family in all living organisms. In plants, UGTs catalyze the transfer of activated sugars, such as UDP-glucose, to hydrophobic aglycones during biosynthesis of a wide variety of natural products (Vogt and Jones 2000). Glycosylation helps to stabilize, detoxify, store, transport and solubilize secondary metabolites, and is often the last step in their biosynthetic pathways. Since no reliable method currently

Plant Cell Physiol. 56(6): 1172–1182 (2015) doi:10.1093/pcp/pcv043, Advance Access publication on 9 March 2015, available online at www.pcp.oxfordjournals.org

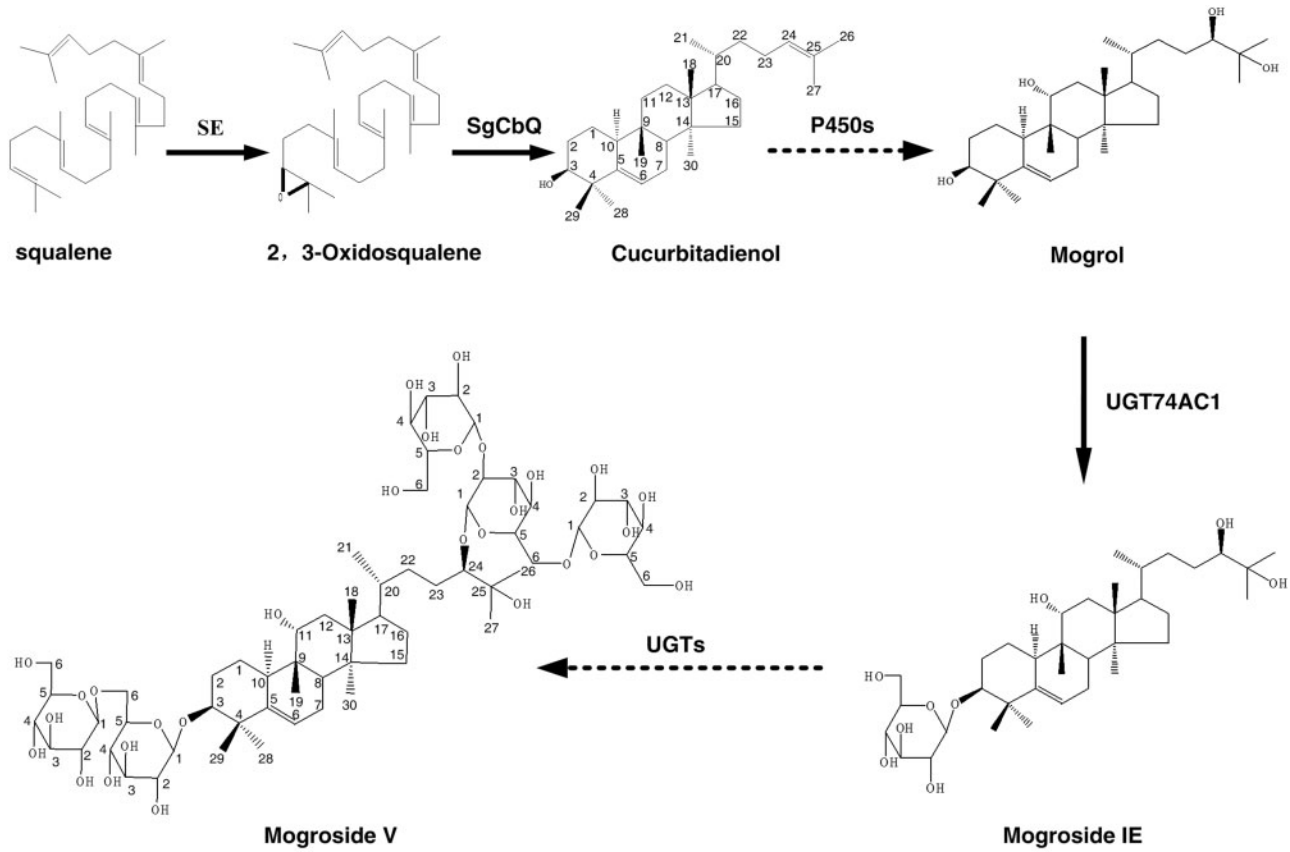

Fig. 1 Proposed biosynthetic pathway of mogrosides in *S. grosvenorii*. SE, squalene epoxidase; SgCbQ, cucurbitadienol synthase; P450s, Cyt P450s; UGTs, uridine diphosphate glycosyltransferases. Dark solid arrows represent one step; dark dotted arrows represent multiple steps.

exists to identify the catalytical function and substrate specificity of plant UGTs by sequence similarity analysis, the study of the functions of UGTs has lagged far behind other biosynthetic steps involved in biosynthesis of triterpene saponins (Modolo et al. 2007). To date, only a few UGTs with specificity for different triterpenoid aglycones have been functionally characterized from several plants, including *Medicago truncatula* (Achnine et al. 2005, Naoumkina et al. 2010), *Saponaria vaccaria* (Meesapyodsuk et al. 2007), *Barbarea vulgaris* (Augustin et al. 2012), *Panax ginseng* (Jung et al. 2014, Yan et al. 2014) and *Glycine max* (Shibuya et al. 2010).

In the past decade, RNA-sequencing (RNA-seq) and digital gene expression (DGE) profile analysis based on next-generation sequencing (NGS) technology have emerged as powerful tools for identification of candidate genes responsible for the biosynthesis of triterpene saponins in non-model plants (AC't Hoen et al. 2008). With improvements in sequencing depth and high accuracy of expression pattern analysis using these tools, a large number of candidate genes involved in the biosynthesis of triterpene saponins have been identified; examples of such genes include ginsenoside biosynthesis in *P. ginseng* (Subramaniyam et al. 2015), steroidal sapogenin biosynthesis in *Asparagus racemosus* (Upadhyay et al. 2014), saikosaponin biosynthesis in *Bupleurum chinense* (Sui et al. 2011) and mogroside biosynthesis in *S. grosvenorii* (Tang et al. 2011). However, a minimal number of studies have been carried out to validate the specific biosynthetic functions of these target genes.

To elucidate the molecular mechanisms of mogrosides biosynthesis in *S. grosvenorii*, we investigated a number of major candidate genes that had previously been identified via RNA-seq and DGE profile analysis. First, *SgCbQ* was expressed in the lanosterol synthase-deficient yeast strain GIL77 and the cucurbitadienol level was measured to illustrate its enzymatic function. Secondly, a number of *UGT* genes were heterologously expressed in *Escherichia coli* as His-tag proteins to evaluate their enzymatic activity towards mogrosides in vitro.

Results

Functional expression of *SgCbQ* cDNA in yeast

Cucurbitadienol is the common precursor of cucurbitane-type mogrosides. A previous DGE profile analysis had shown that the expression level of *SgCbQ* in the fruits of *S. grosvenorii* increased by 3-fold at 70 days after flowering (DAF) compared with that at 50 DAF, and 10-fold at 70 DAF compared with that at 3 DAF (Tang et al. 2011). The expression pattern of *SgCbQ* was proportional to the accumulation of mogroside V. Thus, *SgCbQ* was identified as a cucurbitadienol synthase involved in the biosynthesis of mogrosides.

The full-length cDNA of *SgCbQ* is 2,800 bp and contains an open reading frame (ORF) of 2,280 bp that encodes a protein with 759 amino acids and a predicted molecular mass of

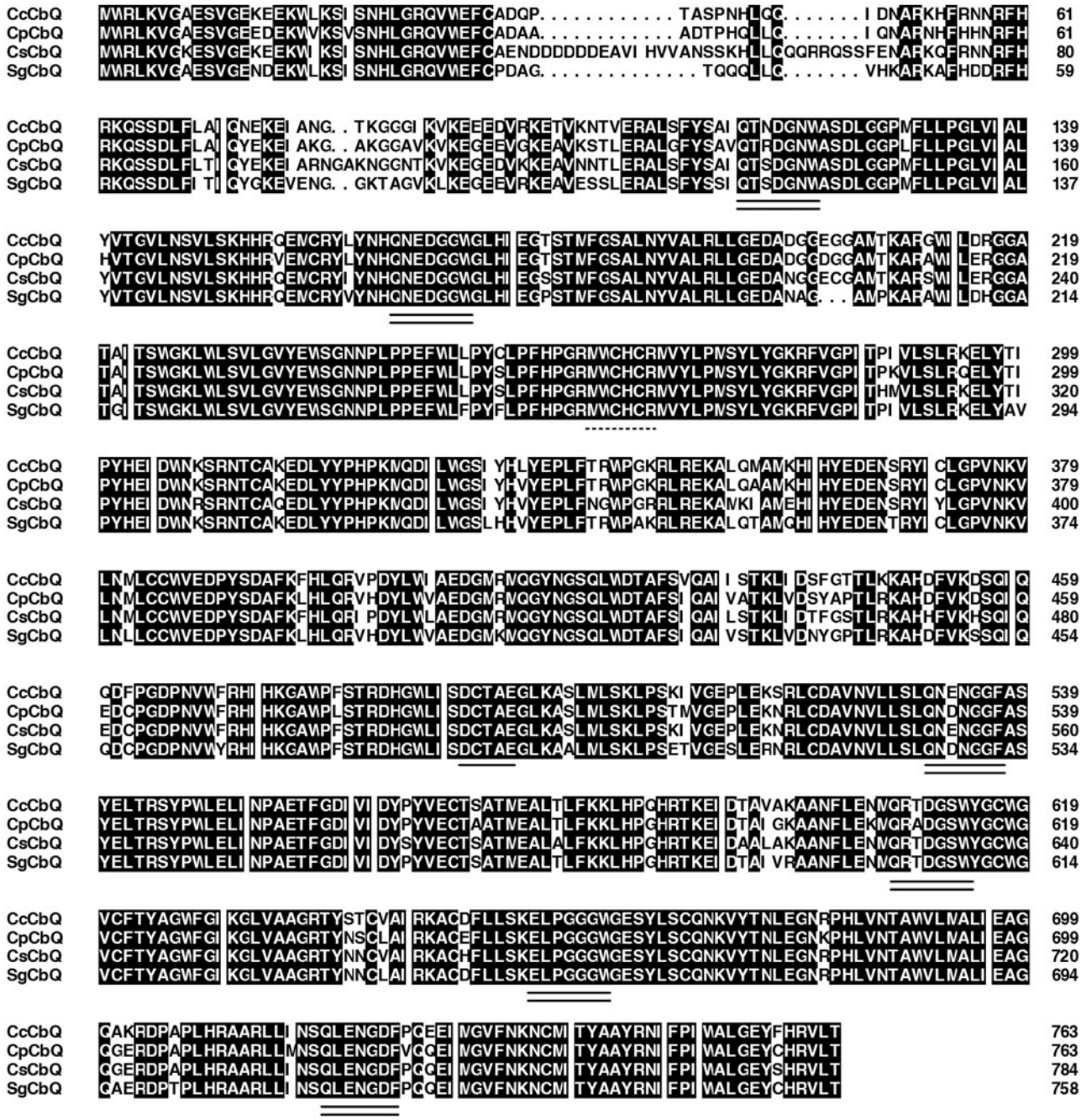

Fig. 2 Alignment of the deduced amino acid sequence of SgCbQ from *S. grosvenorii* with other cucurbitadienol synthases. GenBank accession numbers: SgCbQ (HQ128567), CpCbQ (AB116238), CsCbQ (AIT72030), CcCbQ (KM821405). Motifs are indicated as follows: DCTAE motif, single underline; QW motif, double underline; MWCHCR motif, dotted underline.

84.4 kDa. The deduced amino acid sequence of SgCbQ shares high similarity with that of cucurbitadienol synthases (89% with CpCbQ from *C. pepo*, 85% with CcCbQ from *C. colocynthis* and 84% with CsCbQ from *C. sativus*). More detailed sequence analysis (**Fig. 2**) reveals that SgCbQ contains the DCTAE motif, which is presumed to be responsible for the initiation of the cyclization reaction (Abe et al. 1993). SgCbQ also contains six repeats of QW motifs that can strengthen the structure of the protein and stabilize the carbocation intermediates during cyclization (Poralla et al. 1994). In addition, SgCbQ has a histidine residue in the MWCHCR motif, which has been presumed to be important for stabilization of the protosteryl cation intermediate (Corey et al. 1997). The protosteryl cation intermediate is catalyzed by cucurbitadienol synthase, cycloartenol synthase and lanosterol synthase, and is completely different from other triterpene synthases (Brendolise et al. 2011).

Phylogenetic analysis based on the Neighbor–Joining method (Tamura et al. 2013) showed that SgCbQ clustered with the other three functionally characterized cucurbitadienol synthases and formed a new subfamily that lies between the branches of cycloartenol synthase and lanosterol synthase (**Fig. 3**). The position of SgCbQ in the tree indicates

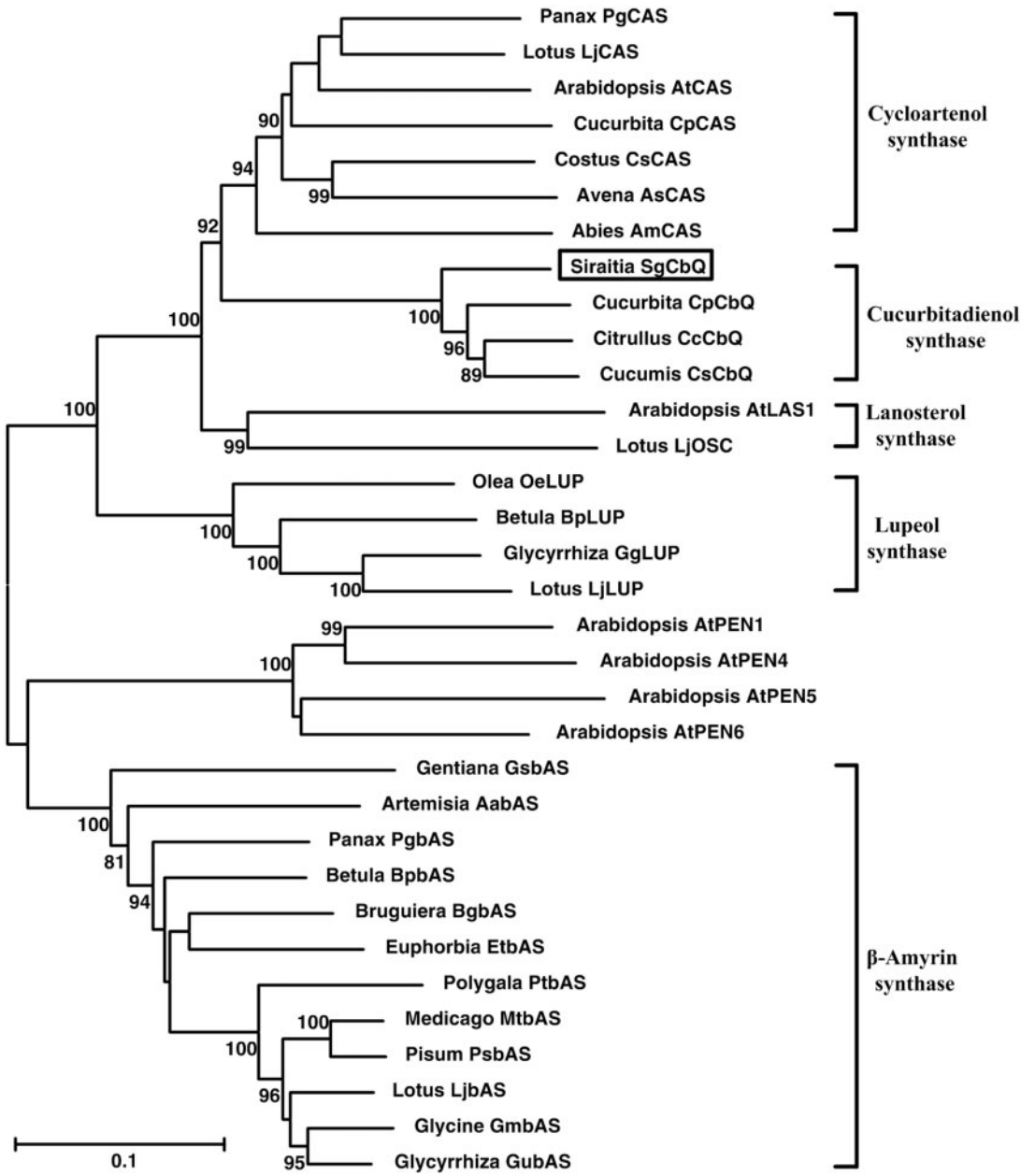

Fig. 3 Non-rooted phylogenetic tree analysis of SgCbQ from *S. grosvenorii* with previously characterized OSCs from other plants. The tree was constructed using the Neighbor–Joining method as described in Materials and Methods. The DDBJ/GenBank/EMBL accession numbers of the sequences used for analysis are as follows: AY095999 (GmbAS, *Glycine max*), GU072921 (GubAS, *Glycyrrhiza uralensis*), AB181244 (LjbAS, *Lotus japonicus*), AJ430607 (MtbAS, *Medicago truncatula*), AB034802 (PsbAS, *Pisum sativum*), EF107623 (PtbAS, *Polygala tenuifolia*), AB289585 (BgbAS, *Bruguiera gymnorhiza*), AB206468 (EtbAS, *Euphorbia tirucalli*), AB055512 (BpbAS, *Betula platyphylla*), AB009030 (PgbAS, *Panax ginseng*), EU330197 (AabAS, *Artemisia annua*), FJ790411 (GsbAS, *Gentiana streminea*), At1g78500 (AtPEN6, *Arabidopsis thaliana*), At5g42600 (AtPEN5, *Arabidopsis thaliana*), At4g15340 (AtPEN1, *Arabidopsis thaliana*), At5g48010 (AtPEN4, *Arabidopsis thaliana*), AB116228 (GgLUP, *Glycyrrhiza glabra*), AB181245 (LjLUP, *Lotus japonicus*), AB055511 (BpLUP, *Betula platyphylla*), AB025343 (OeLUP, *Olea europaea*), At3g45130 (AtLAS1, *Arabidopsis thaliana*), AB244671 (LjLAS1, *Lotus japonicus*), AB116238 (CpCbQ, *Cucurbita pepo*), HQ128567 (SgCbQ, *Siraitia grovenorii*), KM821405 (CcCbQ, *Citrullus colocynthis*), AIT72030 (CsCbQ, *Cucumis sativus*), AF216755 (AmCAS, *Abies magnifica*), AB058507 (CsCAS, *Costus speciosus*), Gu253907 (AsCAS, *Avena strigosa*), AB116237 (CpCAS, *Cucurbita pepo*), At2g07050 (AtCAS, *Arabidopsis thaliana*), AB009029 (PgCAS, *Panax ginseng*), AB181246 (LjCAS, *Lotus japonicus*).

a high degree of similarity in reaction mechanisms among CbQ, cycloartenol synthase and lanosterol synthase. Taken together, all of these data indicate that SgCbQ belongs to the triterpene synthase superfamily and may be a cucurbitadienol synthase.

To examine the biochemical function of *SgCbQ*, *SgCbQ* and *CpCbQ* (functionally characterized cucurbitadienol synthase) were codon-optimized (Grote et al. 2005), synthesized and subcloned into the yeast expression vector pYES2. *CpCbQ* was used as a positive control and as a source of authentic

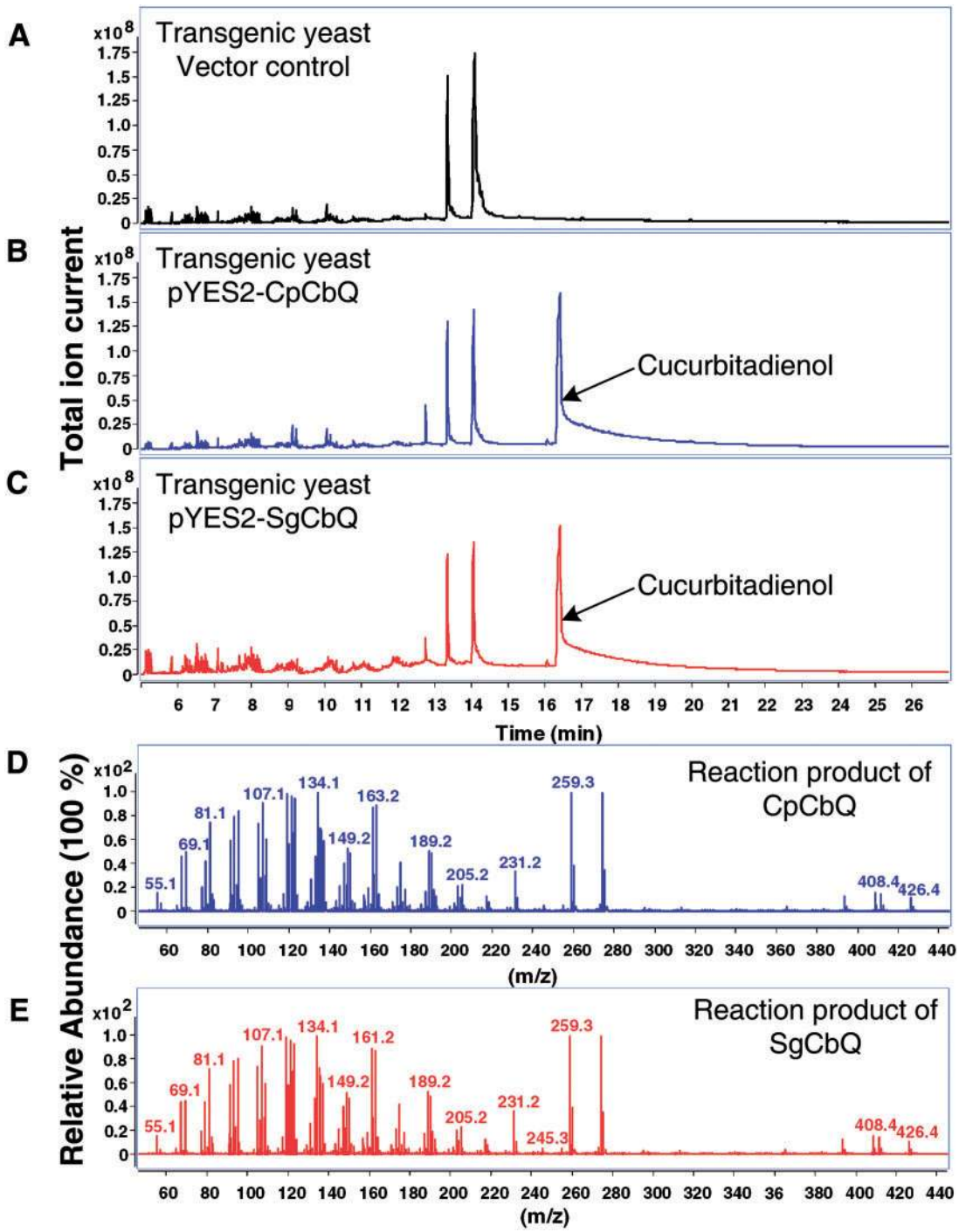

Fig. 4 GC-MS chromatograms of yeast cell extracts. (A) Products from the yeast cells expressing the empty vector pYES2. (B) Products from the yeast cells expressing pYES2-CpCbQ. (C) Products from the yeast cells expressing pYES2-SgCbQ. (D) The mass spectrum for the compound eluting at 16.42 min from the yeast cells expressing pYES2-CpCbQ. (E) The mass spectrum for the compound eluting at 16.42 min from the yeast cells expressing pYES2-SgCbQ.

cucurbitadienol to serve as a standard. The resulting recombinant plasmids combined with the empty vector pYES2, used as a negative control, were transformed into the lanosterol synthase-deficient yeast strain GIL77, which accumulates high endogenous concentrations of 2,3-oxidosqualene that can serve as a substrate for heterologously expressed OSC genes (Wang et al. 2010). After induction with galactose, the cells were collected, saponified and extracted with *n*-hexane. Gas chromatography–mass spectrometry (GC-MS) analysis of the cell extracts showed that all three transformants contained two common compounds, which were identified as squalene and ergosterol. Extracts from yeast transformants expressing *SgCbQ* or *CpCbQ* produced only one new peak with the same retention time of 16.42 min, whereas a signal was not detected in the negative control yeast harboring the empty vector at the retention time of 16.42 min (**Fig. 4**). Furthermore, mass

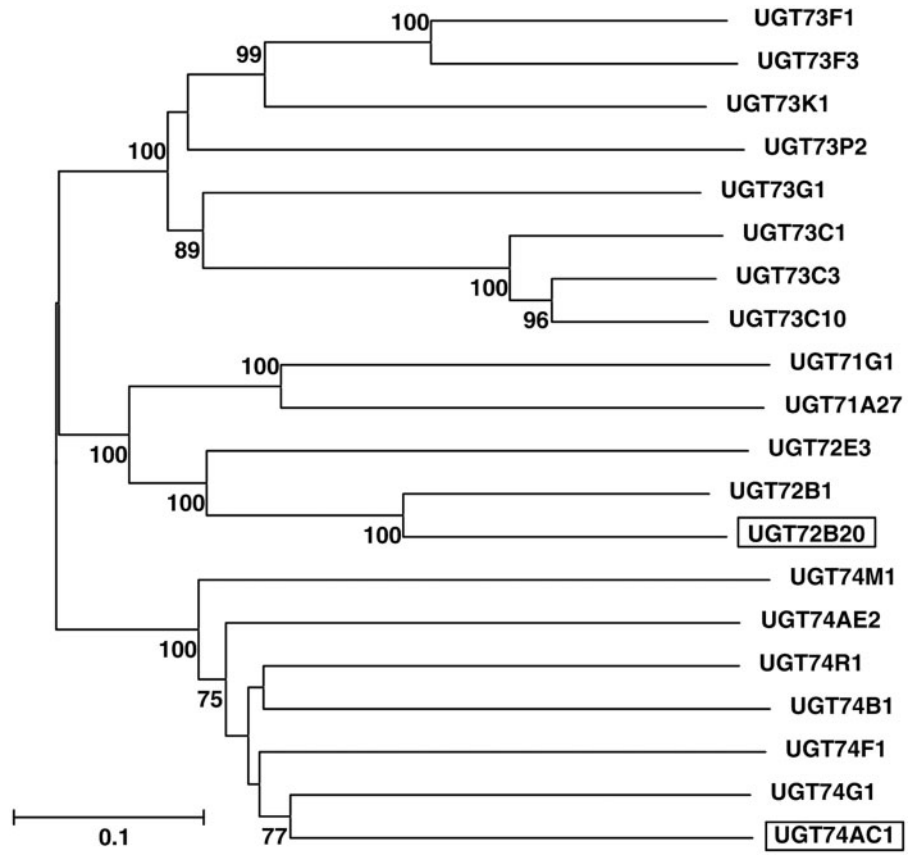

Fig. 5 Non-rooted phylogenetic tree analysis of UGT74AC1 and UGT72B20 from *S. grosvenorii* with previously characterized UGTs from other plants using the Neighbor–Joining method. The tree was constructed as described in the Materials and Methods. Accession numbers, references and substrate acceptors of the UGTs used for phylogenetic analysis are given in **Supplementary Table S2**.

spectrometry (MS) fragmentation patterns of *SgCbQ* and *CpCbQ* at 16.42 min (*m/z*; relative intensity: 119 [100], 134 [99], 259 [75], 274 [71], 121 [70], 95 [58], 107 [55], 207 [52], 161 [51], 81 [44], 136 [36], 69 [36], 149 [28], 175 [23], 408 [16], 393 [14], 231 [13], 55 [6], 411 $[M\text{-}CH_3]^+$ [4], 426 $[M]^+$ [3]) were identical (**Fig. 4**). These results clearly indicate that SgCbQ is a cucurbitadienol synthase that produces cucurbitadienol as the sole product.

Functional characterization of UGT candidates

In the previous RNA-seq and DGE profile analyses, six *UGT* genes with similar expression patterns to the squalene epoxidase gene (*SE*) and *SgCbQ* were identified as candidate genes involved in mogrosides biosynthesis (Tang et al. 2011). Specific primers (**Supplementary Table S1**) were designed for amplification of these *UGT* genes, and the candidate *UGT* genes thus obtained were expressed in *E. coli* Rosetta-gami (DE3) cells as N-terminal His_6-tag fusion proteins. Only two *UGT* genes, named *UGT74AC1* and *UGT72B20* by the UGT nomenclature committee (Mackenzie et al. 1997), were expressed solubly and could be purified by nickel-affinity chromatography. Coomassie staining of SDS–polyacrylamide gel analysis showed that the resulting molecular weight of the purified recombinant UGT74AC1 was approximately 52 kDa, which is consistent with its predicted size (455 amino acids); no expression was detected in the control harboring empty pET-21a. Similar results were obtained with the recombinant UGT72B20 (**Supplementary Fig. S1**).

Phylogenetic analysis based on the Neighbor–Joining method (Tamura et al. 2013) of UGT74AC1 and UGT72B20 from other biochemically characterized plant UGTs (**Supplementary Table S2**) showed that UGT74AC1 and UGT72B20 clustered together with UGT74 and UGT72 family members (**Fig. 5**), respectively. In the UGT74 family, UGT74M1 from *S. vaccaria* has been shown to glycosylate various oleanane triterpenes. Recently, PgUGT74AE2, another UGT74 family member from *P. ginseng*, was reported to glycosylate dammarene-type protopanaxadiol and compound K through *O*-linkage at position C3 to form ginsenoside Rh_2 and ginsenoside F_2 (Jung et al. 2014). UGT72B20 clustered into the UGT72 family and was closely related to the UGT71 family. In the UGT71 family, UGT71G1 from *M. truncatula* showed specificity for triperpene medicagenic acid, and UGT71A27 from *P. ginseng* was found to glycosylate protopanaxadiol to produce ginsenoside compound K. These findings suggest that UGT74AC1 and UGT72B20 have roles in triterpene saponin biosynthesis, considering that *S. grosvenorii* accumulates considerable amounts of mogrosides.

Glycosyltransferase activity of purified UGT74AC1 and UGT72B20 was tested using prepared mogrol, mogroside IE, mogroside IA and other purchased mogroside standards (**Supplementary Table S3**) as acceptors and UDP-glucose as

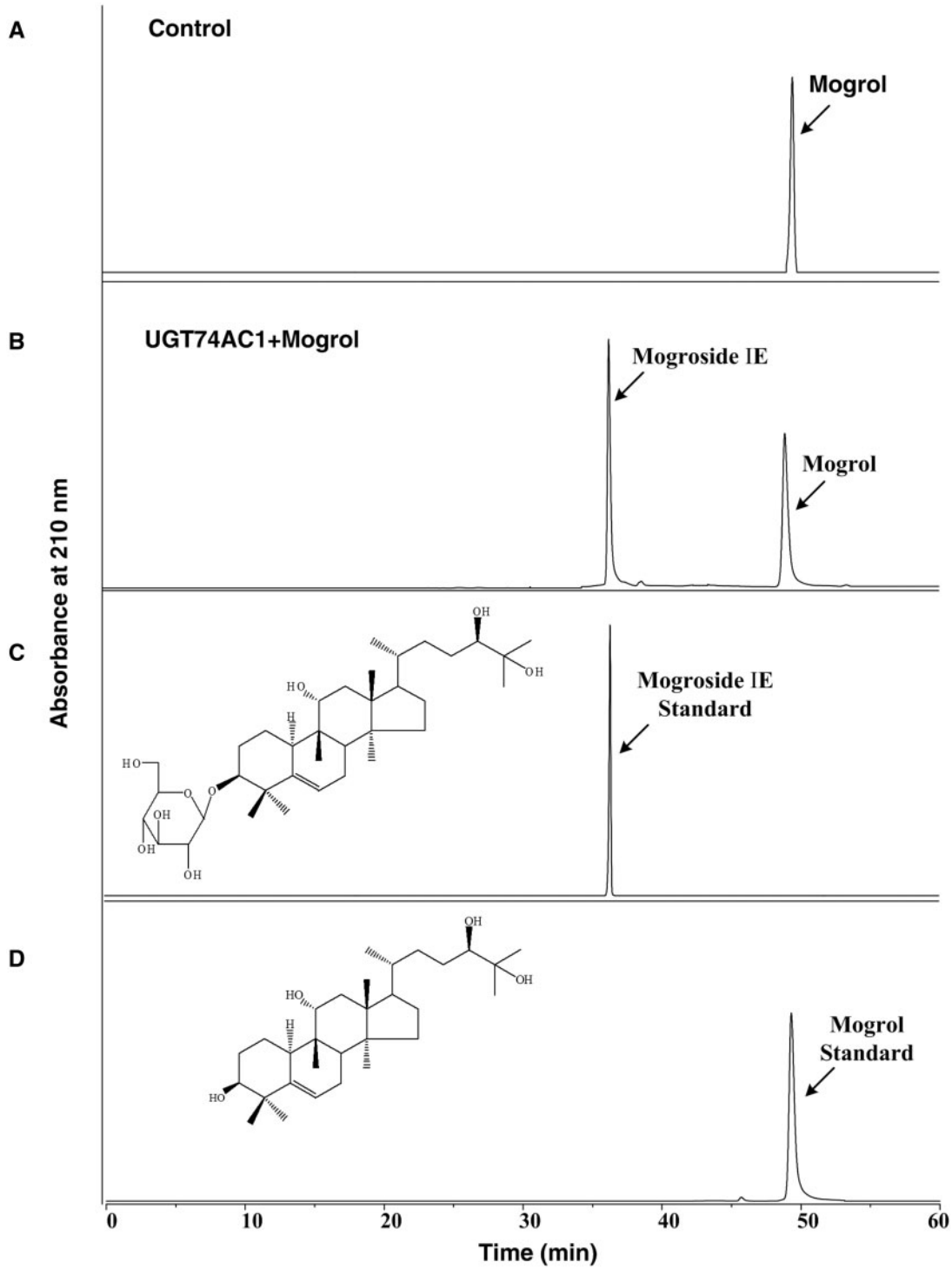

Fig. 6 HPLC analysis of the products by UGT74AC1. (A) The LC chromatogram of the reaction products from mogrol with total proteins with the empty vector. (B) The LC chromatogram of the reaction products from mogrol with purified UGT74AC1. (C) The LC chromatogram of the mogroside IE standard. (D) The LC chromatogram of the mogrol standard.

sugar donor. Reaction products were analyzed by HPLC and the results showed that UGT74AC1 using mogrol as an acceptor produced a new peak at 36.10 min with the same retention time as that of authentic mogroside IE, whereas new products were not observed in the vector control at the retention time of 36.10 min (**Fig. 6**). Liquid chromatography–electrospray ionization–mass spectrometry (LC-ESI-MS) in the negative ion mode was used to identify further the newly formed product. The MS fragmentation pattern of the product was identical to that of authentic mogroside IE with the same *m/z* ions for 637 $[M-H]^-$, 683 $[M-H+HCOOH]^-$ and 751 $[M-H+TFA]^-$ (**Fig. 7**). Products were not detected when UGT72B20 was incubated with mogrol or mogroside standards (data not shown), which may be attributed to the narrow range of tested mogroside standards.

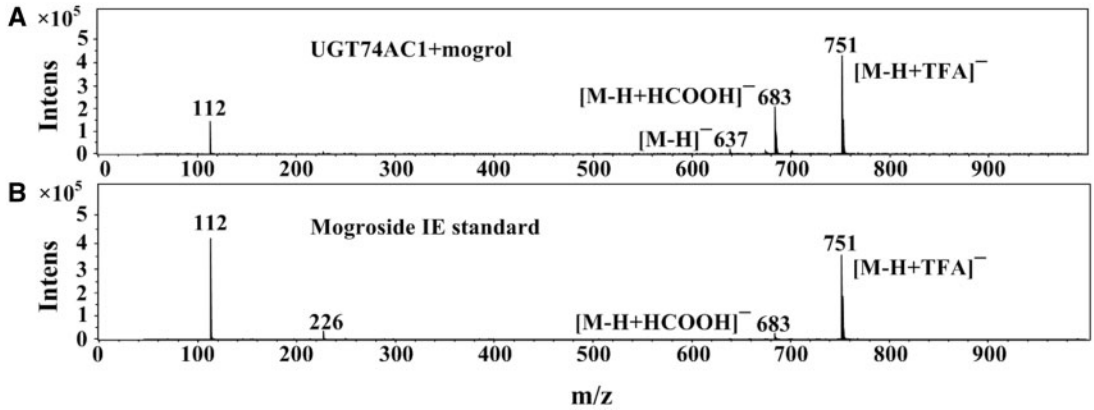

Fig. 7 MS spectra of the reaction products of UGT74AC1 with mogrol. (A) The mass spectrum for a peak (36.10 min) detected in the reaction products from UGT74AC1 and mogrol. (B) The MS spectrum of the mogroside IE standard.

Table 1 Kinetic parameters of UGT74AC1 toward mogrol, quercetin and naringenin

Substrate	K_m (μM)	k_{cat} (S^{-1})	K^{cat}/K_m ($S^{-1} M^{-1}$)
Mogrol	41.4 ± 4.2	$4.11 \times 10^{-4} \pm 0.55 \times 10^{-4}$	9.94 ± 0.75
Quercetin	58.2 ± 1.4	0.41 ± 0.06	$6.88 \times 10^3 \pm 0.55 \times 10^3$
Naringenin	54.7 ± 1.8	0.45 ± 0.04	$8.23 \times 10^3 \pm 0.41 \times 10^3$

Many plant glycosyltransferases are thought to be promiscuous. The potential activity of UGT74AC1 against a range of acceptors, including different types of triterpenoids, phenolic compounds and flavonoids (**Supplementary Table S4**), was tested using UDP-glucose as sugar donor. Of these substrates, only quercetin and naringenin were glycosylated by UGT74AC1, with the molecular masses of the reaction products increased by 162 or 324 Da, respectively, as determined by LC-ESI-MS analysis (**Supplementary Figs. S2, S3**).

The kinetic parameters of purified UGT74AC1 for mogrol, quercetin and naringenin were determined (**Table 1**). The K_m values of UGT74AC1 for mogrol (41.4 μM), quercetin (58.2 μM) and naringenin (54.7 μM) were comparable. However, the k_{cat} value of UGT74AC1 for mogrol ($4.11 \times 10^{-4} S^{-1}$) was quite low compared with quercetin ($0.41 S^{-1}$) and naringenin ($0.45 S^{-1}$), resulting in low catalytic efficiency ($9.94 S^{-1} M^{-1}$ for mogrol compared with $6.88 \times 10^3 S^{-1} M^{-1}$ for quercetin and $8.23 \times 10^3 S^{-1} M^{-1}$ for naringenin).

Discussion

Triterpene saponins are a large group of natural compounds that exhibit great structural diversity in a considerable number of plant species. The first step in biosynthesis of triterpenoids is the cyclization of the common precursor 2,3-oxidosqualene, which can generate >100 different skeleton types. In this study, SgCbQ was confirmed as a novel cucurbitadienol synthase with cucurbitadienol as its sole product. Successful characterization of *SgCbQ* not only confirmed its correct identification based on the DGE profile of *S. grosvenorii*, but also indicated that further DGE profile analysis of candidate P450 genes and *UGT* genes with *SgCbQ* as the reference gene was credible. We used degenerate primers designed based on the highly conserved regions of known OSCs (Shibuya et al. 2004) for further cloning of other CbQ genes from *S. grosvenorii*. However, all attempts were unsuccessful (data not shown). Although functional characterization of all OSCs from *S. grosvenorii* is not possible, *SgCbQ* is the most probable gene involved in biosynthesis of mogrosides in *S. grosvenorii* based on the gene expression pattern with the accumulation of mogroside V and catalytic products of *SgCbQ*. To date, four cucurbitadienol synthases (SgCbQ, CpCbQ, CsCbQ and CcCbQ) have been functionally characterized from different plant species. However, key amino acid residues responsible for formation of cucurbitadienol have not been identified. Therefore, further studies based on homologous modeling and site-directed mutagenesis should be carried out.

Glycosylation is the last step in the biosynthesis of mogrosides and is a critical determinant of their taste. However, genes involved in this step have not been reported to date. In this study, UGT74AC1 was confirmed to glycosylate mogrol at the C-3 hydroxyl group to produce mogroside IE. This is the first reported UGT that shows specificity for cucurbitane-type triterpenoid. Similar to other UGTs, UGT74AC1 showed strong regioselectivity and stereospecificity with mogrol. UGT74AC1 can only glycosylate mogrol at the C-3 hydroxyl, whereas the C-24 or C-11 hydroxyls of mogrol cannot be catalyzed. Furthermore, mogroside IA, which also contains a C-3 hydroxyl group, cannot serve as an acceptor for UGT74AC1, indicating that UGT74AC1 can distinguish the C-24 hydroxyl group of the acceptor.

In general, UGTs exhibited substrate regiospecificity rather than absolute specificity in vitro. Besides mogrol, UGT74AC1 also showed activity with quercetin and naringenin. Kinetic data suggested that UGT74AC1 was most efficient with naringenin, exhibiting the highest k_{cat} value and k_{cat}/K_m ratio. Similarly, *Medicago* triterpene glycosyltransferase UGT71G1 also showed higher catalytic efficiency with quercetin and genistein compared with hederagenin. We could not expect

UGT74AC1 to be a flavonoid glycosyltransferase in vivo considering that naringenin and naringenin glycosides have never been detected in the fruits of *S. grosvenorii* and that at least four UGTs active with triterpenes can also glycosylate flavonoids (Achnine et al. 2005, Naoumkina et al. 2010, Augustin et al. 2012). Typically, plant UGTs belong to the GT-B fold that comprises two Rossmann-like domains. The crystal structures of several plant UGTs, including multifunctional triterpene/flavonoid UGT71G1 (Shao et al. 2005), multifunctional (iso)flavonoid UGT85H2 (Li et al. 2007) and UGT78G1 from *M. truncatula* (Modolo et al. 2009), anthocyanidin 3-*O*-glucosyltransferase Ct3GT-A from *Clitoria ternatea* (Hiromoto et al. 2013) and flavonoid 3-*O* -glycosyltransferase VvGT1 from *Vitis vinifera* (Offen et al. 2006) have been determined. The conserved His20 of VvGT1 and His22 of UGT71G1 in the acceptor-binding pocket were identified as a general base for enzyme activity, while Asp119 in VvGT1 and Asp121 in UGT71G1 formed a hydrogen bond with the histidine to help deprotonation of acceptor molecules. Sequence alignment analysis of UGT74AC1 with these five UGTs showed that the two key amino acid residues in the acceptor-binding pocket were relatively conserved (**Supplementary Fig. S4**). Crystallization of UGT74AC1 will be required for better understanding of the detailed interactions between catalytical function and substrate specificity considering the low sequence similarity among plant UGTs.

The glycosyltransferase superfamily is a large and diverse group of enzymes. For example, 120 and 165 UGT genes were reported in the model plant *Arabidopsis thaliana* (Paquette et al. 2003) and *M. truncatula* (Achnine et al. 2005), respectively. Identification of UGTs responsible for the synthesis of triterpene saponins is difficult because of limited predictability between biochemical function and primary sequence similarity among UGTs. However, the expression patterns of the squalene synthase gene (*SS*), *SE* and *OSC* genes with *UGT* genes are often clustered tightly, which can be used for screening of candidate *UGT* genes involved in triterpene saponin biosynthesis. In a previous study, UGT73F3 was identified and found to show specificity for multiple sapogenins by cluster analysis of *UGT* genes with similar expression patterns to that of the β-amyrin synthase gene in *M. truncatula* (Naoumkina et al. 2010). Using a similar method, UGT71G1 and UGT73K1 were screened from > 300 expressed sequence tags (ESTs) and found to show activity toward medicagenic acid/hederagenin and soyasapongenols B and E/hederagenin, respectively (Achnine et al. 2005). Thus, RNA-seq based on NGS, combined with DGE profile analysis with *SS*, *SE* or *OSC* genes as reference genes may provide a basic method for identification of candidate *UGT* genes in triterpene saponin biosynthesis.

To date, mogroside production has depended completely on extraction from fruits of *S. grosvenorii*, and the production of mogrosides by organic synthesis is not practically possible. *SgCbQ* and *UGT74AC1* identified in this study may potentially allow improvement in mogroside production by genetic transformation of *S. grosvenorii*. The pathway from cucurbitadienol to mogrol is thought to involve a series of hydroxylations by the P450s, which is critical for mogrosides production. Functional characterization of candidate P450s responsible for oxidation of cucurbitadienol will be of particular interest.

Materials and Methods

Plant materials and chemicals

Fruits of *S. grosvenorii* were kindly provided by Professor Qinhong Wang at the3 Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences. UDP-glucose was purchased from Sigma-Aldrich. Oleanolic acid, betulinic acid, ursolic acid, protopanoxadiol, protopanoxatriol, ferulic acid, *p*-coumaric acid, caffeic acid, quercetin and naringenin were purchased from J&K Scientific Ltd. Chemical standard mogrosides (mogroside IIE, mogroside III, mogroside IV, siamenoside I and mogroside V) were purchased from Chengdu Biopurify Phytochemicals. Mogroside extract (containing 55% mogroside V) was purchased from Guilin Layn Natural Ingredients. All other chemicals were of commercial reagent grade.

Preparation of mogrol, mogroside IE and mogroside IA

For preparation of mogrol, mogroside IE and mogroside IA, 300 mg of mogroside V purified from mogroside extract was dissolved in 100 ml of sodium acetate buffer (0.1 mol l^{-1}, pH 5.0) adding 10 ml of crude pectinase from *Aspergillus niger* (Sigma, P2611). Enzymolysis was performed at 50°C for 48 h, and reaction products were extracted three times using an equal volume of ethyl acetate (Kim et al. 2006). The organic extracts were concentrated and evaporated under reduced pressure distillation, then resuspended in appropriate volumes of methanol and purified using an Agilent 1200 preparative HPLC system coupled with a reverse-phase Ultimate C18 column (21.2 × 250 mm, 5 µm particle, Welch) to obtain pure mogrol, mogroside IE and mogroside IA.

For structure analysis, mogrol and mogroside IE were dissolved in pyridine-d5 with 1% tetramethylsilane (TMS) as internal standard. ^{1}H- and ^{13}C-nuclear magnetic resonance (NMR) spectra were obtained using a Bruker DMX-600 NMR spectrometer operating at 600 MHz (**Supplementary Table S5**). Considering that all the products were hydrolyzed from mogroside V, the minor compound, whose level was too low for NMR analysis, eluted at 38.50 min (analyzed by HPLC with the same conditions as mogroside IE), exhibited the same ions for 637 $[M-H]^-$, 683 $[M-H+HCOOH]^-$ and 751 $[M-H+HCOOH]^-$, and so was speculated to be mogroside IA (mogrol glycosylated at the C-24 hydroxyl group).

Isolation of RNA and cloning of candidate *UGT* genes

Total RNA was isolated from the fruits of *S. grosvenorii* using the Trizol reagent according to the manufacturer's instructions (Invitrogen). The RNA sample was treated with DNase I (Fermentas) at 37°C for 30 min, then 3 µg of the DNase I-treated total RNA was used as template for the synthesis of first-strand cDNA using the Quantscript RT Kit (Tiangen) in a 20 µl reaction with oligo$(dT)_{15}$ primer following the manufacturer's recommended protocols. The coding regions of the six *UGT* genes were PCR amplified using Q5 High-Fidelity DNA Polymerase (NEB) with specific primers (**Supplementary Table S1**) and ligated into the pEASY-Blunt Simple Cloning Vector (TransGen Biotech) for sequencing. The correctly sequenced plasmids were digested with the corresponding restriction enzymes, and then the resulting DNA fragments were recycled and cloned into the restriction sites of the pET21a expression vector (Novagen).

Functional characterization of *SgCbQ* in yeast

SgCbQ (GenBank accession No. HQ128567) and *CpCbQ* (GenBank accession No. AB116238) were codon-optimized for synthesis and cloned into the *Kpn*I/*Xho*I sites of the pYES2 yeast expression vector under the control of the GAL1 promoter to construct pYES2-SgCbQ and pYES2-CpCbQ. The two recombinant plasmids and the empty vector pYES2 were transformed into *Saccharomyces cerevisiae* GIL77 (*gal2 hem3-6 erg7 ura3-167*) by the lithium acetate method (Daniel Gietz and Woods 2002). Yeast expression experiments

were carried out as described before (Kushiro et al. 1998). After galactose induction, cells were collected, refluxed with 5 ml of 20% KOH/50% ethanol and extracted three times with the same volume of *n*-hexane. The combined extracts were dried under reduced pressure distillation and dissolved in 500 µl of *n*-hexane for GC-MS analysis as described by Dai et al. (2014).

Heterologous expression and purification of UGTs in *E. coli*

Escherichia coli Rosetta-gami (DE3), which is optimized for the expression of eukaryotic genes, was used as host for protein expression. Transformants were cultured with Luria–Bertani (LB) medium plus ampicillin (100 µg ml^{-1}), kanamycin (15 µg ml^{-1}), chloramphenicol (34 µg ml^{-1}) and tetracycline (12.5 µg ml^{-1}L) at 37°C until the OD_{600} reached 0.6–0.8. The expression was initiated after addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM, after which, the cells were further cultured for 18–20 h at 16°C.

For protein purification, cell pellets containing the recombinant fusion protein were harvested by centrifugation at 4,000 × *g* for 10 min and resuspended in lysis buffer (50 mM Tris–HCl, 150 mM NaCl, 25 mM imidazole, pH 7.2). After disruption by sonication, the homogenate was centrifuged at 10,000 × *g* for 30 min at 4°C. The supernatant soluble protein was purified using an Ni-NTA agarose affinity column (GE Healthcare) according to the manufacturer's protocols. The purity and molecular mass of the UGTs were determined by SDS–PAGE. Protein concentration was measured by the Bradford method with bovine serum albumin (BSA) as standard.

In vitro enzyme assays of UGTs

Triterpene UGT activity assay (500 µl) was conducted with 50 µg of purified protein, 50 mM Tris–HCl (pH 7.2), 5 mM $MgCl_2$, 1 mM UDP-glucose and 200 µM of different triterpene substrates. The reaction mixture was incubated at 30°C for 4 h, then samples were extracted three times with the same volume of ethyl acetate. Then, the ethyl acetate phase was collected and evaporated under a stream of nitrogen gas, and the residue was re-diluted in 200 µl of methanol for LC-ESI-MS analysis.

To test the activity of UGTs towards flavonoids and phenolic compounds, the assay mixture (200 µl) contained 10 µg of purified recombinant protein, 1 mM of UDP-glucose in 50 mM Tris–HCl (pH 7.2), 5 mM $MgCl_2$ and 200 µM of variant flavonoids and phenolic compounds. The reactions were carried out at 30°C for 30 min and stopped by adding an equal volume of methanol. After centrifugation at 10,000 × *g* for 10 min, the reaction mixture was filtered with a 0.22 µm filter prior to analysis by LC-MS.

Kinetic studies for mogrol were carried out in 400 µl volumes with 100 µg of purified recombinant protein, 50 mM Tris–HCl (pH 7.2), 5 mM $MgCl_2$, 1 mM UDP-glucose and mogrol (6.25–120 µM). Mixtures were incubated at 30°C for 1 h and stopped by adding an equal volume of ethyl acetate. All subsequent steps were conducted as described above. Kinetic studies for quercetin and naringenin were performed in enzyme assays with 3 µg of purified recombinant protein, quercetin and naringenin concentrations of 10–200 µM and a reaction time of 10 min. Kinetic parameters were determined from Lineweaver–Burk plots from triplicate experiments. The k_{cat} values were calculated using the predicted molecular mass of 51,274 g mol^{-1} for UGT74AC1.

LC-ESI-MS analysis

An Agilent 1260 HPLC system coupled to a Bruker-micrOTOF-II mass spectrometer with an electrospray ionization (ESI) probe was used for LC-ESI-MS analysis. Glycosylated products of triterpene were subjected to a reverse-phase Ultimate C18 column (4.6 × 250 mm, 5 µm particle, Welch, Shanghai, China) with a flow rate of 1 ml min^{-1} and UV wavelength of 210 nm. The column was eluted with solvent A (H_2O) and solvent B (100% CH_3CN) using the following gradients: 0–20 min, 23% acetonitrile; 20–40 min, 23–50%; 40–60 min, 50% B. ESI parameters were optimized as follows: the scan range was 100–1,000 *m/z* in negative ion mode, spray voltage was 4,500 V, capillary temperature was 400°C, dry gas was 6 ml min^{-1}, dry temperature was 180°C and nebulizer pressure was 1 bar.

For analysis of products formed by glycosylation of flavonoids, the same column was eluted with a linear gradient of 20–85% acetonitrile in H_2O (all containing 0.1% formic acid) at 1.0 ml min^{-1} over 25 min and detected with a wavelength of 254 nm. ESI parameters were set as described above but the spectrometer was operated in positive ion mode.

Phylogenetic tree construction

Sequence alignments analyses of SgCbQ (**Suppplementary Fig. S5**) with OSCs, and of UGT74AC1 with UGTs (**Suppplementary Fig. S6**) were carried out with ClustalW using the respective amino acid sequences. The non-rooted phylogenetic tree was constructed by the Neighbor–Joining method using the MEGA6 program (Tamura et al. 2013) with the following parameters: bootstrap method (1,000 replicates), p-distance model, uniform rates and complete deletion. In the phylogenetic tree, the scale bar indicates 0.1 amino acid substitution per site, and numbers at the nodes indicate the percentage of bootstrap values.

Supplementary data

Supplementary data are available at PCP online.

Funding

This work was supported by the National High Technology Research and Development Program of China [No. 2013AA102102]; the Chinese Academy of Sciences [Key Programs KSZD-EW-Z-019]; Tianjin Enterprise Postdoctoral Innovation Project Special Foundation.

Acknowledgments

The authors are very grateful to Professor Reinhard Jetter (University of British Columbia) for his generous gift of the yeast strain GIL77. We appreciate Dr. Caixia Dong (Tianjin University of Traditional Chinese Medicine) for her kind assistance with NMR analysis. We also thank Yi Cai and Yonghong Yao (Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences) for LC-MS and GC-MS analysis.

Disclosures

The authors have no conflicts of interest to declare.

References

Abe, I., Rohmer, M. and Prestwich, G.D. (1993) Enzymatic cyclization of squalene and oxidosqualene to sterols and triterpenes. *Chem.l Rev.* 93: 2189–2206.

AC't Hoen, P., Ariyurek, Y., Thygesen, H.H., Vreugdenhil, E., Vossen, R.H., de Menezes, R.X. et al. (2008) Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. *Nucleic Acids Res.* 36: e141–e141.

Achnine, L., Huhman, D.V., Farag, M.A., Sumner, L.W., Blount, J.W. and Dixon, R.A. (2005) Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume *Medicago truncatula*. *Plant J.* 41: 875–887.

Augustin, J.M., Drok, S., Shinoda, T., Sanmiya, K., Nielsen, J.K., Khakimov, B. et al. (2012) UDP-glycosyltransferases from the UGT73C subfamily in *Barbarea vulgaris*catalyze sapogenin 3-O-glucosylation in saponin-mediated insect resistance. *Plant Physiol.* 160: 1881–1895.

Brendolise, C., Yauk, Y.K., Eberhard, E.D., Wang, M., Chagne, D., Andre, C. et al. (2011) An unusual plant triterpene synthase with predominant

α-amyrin-producing activity identified by characterizing oxidosqualene cyclases from *Malus* × *domestica*. *FEBS J.* 278: 2485–2499.
Chaturvedula, V.S.P. and Prakash, I. (2011) Cucurbitane glycosides from *Siraitia grosvenorii*. *J. Carbohydr. Chem.* 30: 16–26.
Corey, E.J., Cheng, H.M., Baker, C.H., Matsuda, S.P.T., Li, D. and Song, X.L. (1997) Studies on the substrate binding segments and catalytic action of lanosterol synthase. Affinity labeling with carbocations derived from mechanism-based analogs of 2,3-oxidosqualene and site-directed mutagenesis probes. *J. Amer. Chem. Soc.* 119: 1289–1296.
Dai, Z., Wang, B., Liu, Y., Shi, M., Wang, D., Zhang, X. et al. (2014) Producing aglycons of ginsenosides in bakers' yeast. *Sci. Rep.* 4: 3698.
Daniel Gietz, R. and Woods, R.A. (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.* 350: 87–96.
Davidovich-Rikanati, R., Shalev, L., Baranes, N., Meir, A., Itkin, M., Cohen, S. et al. (2015) Recombinant yeast as a functional tool for understanding bitterness and cucurbitacin biosynthesis in watermelon (*Citrullus*spp.). *Yeast* 32: 103–114.
Grote, A., Hiller, K., Scheer, M., Münch, R., Nörtemann, B., Hempel, D.C. et al. (2005) JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.* 33: W526–W531.
Hiromoto, T., Honjo, E., Tamada, T., Noda, N., Kazuma, K., Suzuki, M. et al. (2013) Crystal structure of UDP-glucose: anthocyanidin 3-O-glucosyltransferase from *Clitoria ternatea*. *J. Synchrotron Radiat.* 20: 894–898.
Jin, J.-S. and Lee, J.-H. (2012) Phytochemical and pharmacological aspects of *Siraitia grosvenorii*luo han kuo. *Orient. Pharm. Exp. Med.* 12: 233–239.
Jung, S.-C., Kim, W., Park, S.C., Jeong, J., Park, M.K., Lim, S. et al. (2014) Two ginseng UDP-glycosyltransferases synthesize ginsenoside Rg3 and Rd. *Plant Cell Physiol.* 55: 2177–2188.
Kim, B.H., Lee, S.Y., Cho, H.J., You, S.N., Kim, Y.J., Park, Y.M. et al. (2006) Biotransformation of Korean *Panax ginseng*by Pectinex. *Biol. Pharm. Bull.* 29: 2472–2478.
Kushiro, T., Shibuya, M. and Ebizuka, Y. (1998) β-Amyrin synthase. *Eur. J. Biochem.* 256: 238–244.
Li, L., Modolo, L.V., Escamilla-Trevino, L.L., Achnine, L., Dixon, R.A. and Wang, X. (2007) Crystal structure of *Medicago truncatula*UGT85H2—insights into the structural basis of a multifunctional (iso) flavonoid glycosyltransferase. *J. Mol. Biol.* 370: 951–963.
Mackenzie, P.I., Owens, I.S., Burchell, B., Bock, K., Bairoch, A., Belanger, A. et al. (1997) The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. *Pharmacogenet. Genomics* 7: 255–269.
Meesapyodsuk, D., Balsevich, J., Reed, D.W. and Covello, P.S. (2007) Saponin biosynthesis in *Saponaria vaccaria*. cDNAs encoding beta-amyrin synthase and a triterpene carboxylic acid glucosyltransferase. *Plant Physiol.* 143: 959–969.
Modolo, L.V., Blount, J.W., Achnine, L., Naoumkina, M.A., Wang, X. and Dixon, R.A. (2007) A functional genomics approach to (iso) flavonoid glycosylation in the model legume *Medicago truncatula*. *Plant Mol. Biol.* 64: 499–518.
Modolo, L.V., Li, L., Pan, H., Blount, J.W., Dixon, R.A. and Wang, X. (2009) Crystal structures of glycosyltransferase UGT78G1 reveal the molecular basis for glycosylation and deglycosylation of (iso) flavonoids. *J. Mol. Biol.* 392: 1292–1302.
Naoumkina, M.A., Modolo, L.V., Huhman, D.V., Urbanczyk-Wochniak, E., Tang, Y., Sumner, L.W. et al. (2010) Genomic and coexpression analyses predict multiple genes involved in triterpene saponin biosynthesis in *Medicago truncatula*. *Plant Cell* 22: 850–866.
Offen, W., Martinez-Fleites, C., Yang, M., Kiat-Lim, E., Davis, B.G., Tarling, C.A. et al. (2006) Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. *EMBO J.*I 25: 1396–1405.
Paquette, S., Møller, B.L. and Bak, S. (2003) On the origin of family 1 plant glycosyltransferases. *Phytochemistry* 62: 399–413.
Poralla, K., Hewelt, A., Prestwich, G.D., Abe, I., Reipen, I. and Sprenger, G. (1994) A specific amino acid repeat in squalene and oxidosqualene cyclases. *Trends Biochem. Sci.* 19: 157–158.
Prakash, I. and Chaturvedula, V.S.P. (2014) Additional new minor cucurbitane glycosides from *Siraitia grosvenorii*. *Molecules* 19: 3669–3680.
Shang, Y., Ma, Y., Zhou, Y., Zhang, H., Duan, L., Chen, H. et al. (2014) Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science* 346: 1084–1088.
Shao, H., He, X., Achnine, L., Blount, J.W., Dixon, R.A. and Wang, X. (2005) Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from *Medicago truncatula*. *Plant Cell* 17: 3141–3154.
Shibuya, M., Adachi, S. and Ebizuka, Y. (2004) Cucurbitadienol synthase, the first committed enzyme for cucurbitacin biosynthesis, is a distinct enzyme from cycloartenol synthase for phytosterol biosynthesis. *Tetrahedron* 60: 6995–7003.
Shibuya, M., Nishimura, K., Yasuyama, N. and Ebizuka, Y. (2010) Identification and characterization of glycosyltransferases involved in the biosynthesis of soyasaponin I in *Glycine max*. *FEBS Lett.* 584: 2258–2264.
Subramaniyam, S., Mathiyalagan, R., Natarajan, S., Kim, Y.-J., Jang, M.-g., Park, J.-H. et al. (2015) Transcript expression profiling for adventitious roots of *Panax ginseng*Meyer. *Gene* 546: 89–96.
Sui, C., Zhang, J., Wei, J., Chen, S., Li, Y., Xu, J. et al. (2011) Transcriptome analysis of *Bupleurum chinense*focusing on genes involved in the biosynthesis of saikosaponins. *BMC Genomics* 12: 539.
Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30: 2725–2729.
Tang, Q., Ma, X., Mo, C., Wilson, I.W., Song, C., Zhao, H. et al. (2011) An efficient approach to finding *Siraitia grosvenorii*triterpene biosynthetic genes by RNA-seq and digital gene expression analysis. *BMC Genomics* 12: 343.
Upadhyay, S., Phukan, U.J., Mishra, S. and Shukla, R.K. (2014) De novo leaf and root transcriptome analysis identified novel genes involved in steroidal sapogenin biosynthesis in *Asparagus racemosus*. *BMC Genomics* 15: 746.
Vogt, T. and Jones, P. (2000) Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. *Trends Plant Sci.* 5: 380–386.
Wang, Z., Yeats, T., Han, H. and Jetter, R. (2010) Cloning and characterization of oxidosqualene cyclases from *Kalanchoe daigremontiana*enzymes catalyzing up to 10 rearrangement steps yielding friedelin and other triterpenoids. *J. Biol. Chem.* 285: 29703–29712.
Xia, Y., Rivero-Huguet, M.E., Hughes, B.H. and Marshall, W.D. (2008) Isolation of the sweet components from *Siraitia grosvenorii*. *Food Chem.* 107: 1022–1028.
Yan, X., Fan, Y., Wei, W., Wang, P., Liu, Q., Wei, Y. et al. (2014) Production of bioactive ginsenoside compound K in metabolically engineered yeast. *Cell Res.* 24: 770–773.