FEBS Letters

FEBS PRESS science publishing by scientists

Novel triterpene oxidizing activity of *Arabidopsis thaliana* CYP716A subfamily enzymes

Shuhei Yasumoto[1], Ery O. Fukushima[1,2], Hikaru Seki[1] and Toshiya Muranaka[1]

1 Department of Biotechnology, Graduate School of Engineering, Osaka University, Suita, Osaka, Japan
2 Frontier Research Base for Global Young Researchers, Graduate School of Engineering, Osaka University, Suita, Osaka, Japan

Correspondence
T. Muranaka, Department of Biotechnology, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan
Fax: +81 6 6879 7426
Tel: +81 6 6879 7423
E-mail: muranaka@bio.eng.osaka-u.ac.jp

(Received 3 December 2015, revised 1 January 2016, accepted 15 January 2016, available online 15 February 2016)

doi:10.1002/1873-3468.12074

Edited by Ulf-Ingo Flügge

Triterpenoids have diverse chemical structures and bioactivities. Cytochrome P450 monooxygenases play a key role in their structural diversification. In higher plants, CYP716A subfamily enzymes are triterpene oxidases. In this study, *Arabidopsis thaliana* CYP716A1 and CYP716A2 were characterized by heterologously expressing them in simple triterpene-producing yeast strains. In contrast to the C-28 oxidative activity of CYP716A1 shown in several CYP716A subfamily enzymes, remarkably, CYP716A2 displayed 22α-hydroxylation activity against α-amyrin that has not been previously reported, which produces the cytotoxic triterpenoid, 22α-hydroxy-α-amyrin. Our results contribute to the enrichment of the molecular toolbox that allows for the combinatorial biosynthesis of diverse triterpenoids.

Keywords: 22α-hydroxylation; *Arabidopsis thaliana*; combinatorial biosynthesis; CYP716A; cytochrome P450 monooxygenase; triterpenoids

Highlights

- The full-length CDS of *CYP716A2*, miss-annotated on TAIR, is successfully cloned.
- Novel 22α-hydroxylation activity of CYP716A subfamily enzymes is reported.
- The chemical structure of 22α-hydroxy-α-amyrin is confirmed by NMR analysis.

Triterpenoids represent one of the largest groups of plant specialized metabolites and are composed of six C5 (isopentenyl diphosphate) units [7]. Their carbon skeletons are synthesized from the common precursor 2,3-oxidosqualene by oxidosqualene cyclases (OSCs), and these skeletons are further oxidized by cytochrome P450 monooxygenases (P450s) and glycosylated by UDP-dependent glycosyltransferases (UGTs), which generate diverse triterpenoid/triterpenoid saponin structures [17,19]. Around 80 OSCs, 30 P450s, and 10 UGTs are reportedly involved in triterpenoid biosynthesis in plants [22,25]. Among them, P450s have been described as the key players in diversifying triterpenoid scaffolds, and were recently identified in many triterpenoid-producing plant species [22]. For example, CYP88D6 and CYP72A154 from *Glycyrrhiza uralensis* were identified as components in the biosynthesis of glycyrrhetinic acid, the aglycone of glycyrrhizin [20,21]. In addition, CYP716A12 and other CYP716A subfamily enzymes have been identified as multifunc-

Abbreviations
AGI, Arabidopsis Genome Initiative; GC-MS, gas chromatography-mass spectrometry; LUS, lupeol synthase; NMR, nuclear magnetic resonance; OSCs, oxidosqualene cyclases; SREBP, sterol regulatory element-binding protein; TICs, total ion chromatograms; UGTs UDP-dependent glycosyltransferases; αAS, α-amyrin synthase; βAS, β-amyrin synthase.

tional oxidases that oxidize α-amyrin, β-amyrin, and lupeol to ursolic acid, oleanolic acid, and betulinic acid, respectively [5,8]. These discoveries have thus provided a pool of enzymes for the heterologous production of useful triterpenoids in microbes [12]. Moreover, as demonstrated in a previous study, natural, and rare triterpenoids can be generated by combinatorial expression of P450s in heterologous hosts [6]. Therefore, the discovery of P450s that exhibit novel and/or strong oxidative activities is important for the production of novel and useful compounds.

Although all 13 OSCs have been identified in the model plant *Arabidopsis thaliana* (Brassicaceae) [25], only a few triterpene skeleton oxidizing P450s have been functionally characterized [3,4]. Boutanaev *et al.* found that CYP716A1 (At5g36110) can oxidize an unknown position of tirucalla-7,24-dien-3β-ol, the main product of PEN3 (At5g36150), which is one of the 13 OSCs in *A. thaliana* [2,11]. However, the oxidative activity of CYP716A1 against known substrates for other CYP716A subfamily enzymes is unknown. In addition, the *A. thaliana* genome encodes another CYP716A subfamily, CYP716A2, whose biochemical function is entirely unknown.

In this study, CYP716A1 and CYP716A2 were heterologously expressed in simple triterpene-producing yeasts to determine their function. Both CYP716A1 and CYP716A2 exhibited oxidative activities against α-amyrin, β-amyrin, and lupeol. However, the profiles of each oxidized product differed. Interestingly, CYP716A2 displayed 22α-hydroxylation activity against α-amyrin that has not been previously reported. This study shows that gene discovery in plants not known to produce useful compounds can also be a source of novel enzymes that could be used for combinatorial biosynthesis.

Materials and methods

Plasmid construction

The full-length coding sequences of CYP716A1 and CYP716A2 were amplified using Prime STAR Max DNA Polymerase (TaKaRa Bio, Shiga, Japan) with primer combinations 1403/1404 and 1405/1406, respectively (primer sequences are listed in Table S1). SuperScript Pre-Made cDNA Library (*Arabidopsis thaliana* third flower-stage seedlings cDNA library; Invitrogen, Carlsbad, CA, USA, 11474-012) was used as the PCR template. Each amplicon was cloned into the entry vector pENTR/D-TOPO (Thermo Fisher Scientific, Waltham, MA, USA). The CYP716A2 fragment had two nucleotides deleted at the 3′ end due to incomplete primer synthesis. However, cloning of this fragment into the pENTR-D-TOPO vector created an artificial in-frame stop codon at the same position. This clone was used for further experiments.

To construct Gateway-compatible yeast expression vectors pELC-GW and pESC-HIS-GW, a Gateway cassette (attL1-ccdB-Cmr-attL2) was inserted into multi-cloning site two of pELC [20] and pESC-HIS (Agilent Technologies, Santa Clara, MA, USA), respectively (H. Seki *et al.* unpublished data).

The cDNA of CYP716A2 was transferred into pELC-GW, pYES-DEST52 (Thermo Fisher Scientific), and pESC-HIS-GW vectors using GW LR Clonase II Enzyme Mix (Thermo Fisher Scientific) to create pELC-CYP716A2, pYES-DEST52-CYP716A2, and pESC-HIS-CYP716A2, respectively. The same strategy was applied for the CYP716A1 yeast expression vectors (pELC-CYP716A1, pYES-DEST52-CYP716A1, and pESC-HIS-CYP716A1).

Yeast transformation and culture conditions

Saccharomyces cerevisiae INV*Sc*1 (Thermo Fisher Scientific) harboring pYES-ADH-aAS, pYES3-ADH-bAS or pYES3-ADH-LUS [5] was transformed with the P450 expression vector(s) using a Frozen-EZ Yeast Transformation II Kit (Zymo Research, Irvine, CA, USA), and transformants were selected on appropriate SD plates. Each of the transformed yeast strains are listed in Table S2 and were stored in 25% glycerol at −80 °C. Each glycerol stock was inoculated into 2 mL of the appropriate SD medium containing 2% glucose and cultured at 30 °C overnight at 200 r.p.m. The cultures were transferred into 10 mL of the same medium and subsequently cultured at 30 °C for an additional 24 h with shaking at 200 r.p.m. Yeast cells were collected and resuspended in 10 mL SD medium containing 2% galactose to induce CYP716A expression and cultured at 30 °C for 2 days at 200 r.p.m. Yeast cultures were then stored at −80 °C until extraction.

Extraction and GC-MS analysis

Yeast cultures were extracted thrice with 6 mL ethyl acetate. The extracts were evaporated and the remaining residues dissolved in 1 mL ethyl acetate. The sample was then transferred into a Sep-pak Silica 6 cc Vac Cartridge (Waters, Milford, MA, USA) and eluted with 10 mL ethyl acetate and 10 mL chloroform : methanol (1 : 1). The eluate was then evaporated and dissolved in 500 μL chloroform : methanol, and samples were stored at −30 °C until use. Fifty microliters of the sample and standard (α-amyrin, uvaol, ursolic acid, β-amyrin, erythrodiol, oleanolic acid, lupeol, betulin, and betulinic acid, 10 p.p.m. in methanol) solutions were transferred into vial inserts and evaporated. Finally, the sample was derivatized with 100 μL *N*-methyl-*N*-(trimethylsilyl) trifluoroacetamide for 20 min at 80 °C.

Gas chromatography-mass spectrometry (GC-MS) analysis was performed using a 5977A GC-MS system (Agilent technologies) with a DB-1 ms (30 m × 0.250 mm, 0.25 μm film thickness; Agilent technologies) or DB-1ht (30 m × 0.250 mm, 0.10 μm film thickness; Agilent technologies) capillary column. The injection component, MSD transfer line and ion source temperatures were 250 °C, 250 °C, and 230 °C, respectively. The oven program used was as follows: 80 °C for 1 min, followed by an increase to 320 °C at a rate of 20 °C·min^{-1} and a hold at 320 °C for 28 min. The carrier gas was He, and the flow rate was 1.0 mL·min^{-1}. The injection volume was 1 μL, and the mass scan range was 50–750. The gain value of the mass detector was 1.0. Peaks were identified by comparing the retention times and mass fragmentation patterns with those of the standards.

Chemical structure determination

The yeast strain SY26 was cultured in ~ 7 L (~ 200 mL × 36) SD medium containing 2% galactose and extracted thrice with ~ 4 L (~ 120 mL × 36) ethyl acetate. The extracts were evaporated and the remaining residue subjected to column chromatography on silica (60N 40–50 μm) in a hexane : ethyl acetate gradient (5 : 1–3 : 1). After running through the column three times for purification, 3.6 mg of compound **4** was obtained.

Nuclear magnetic resonance (NMR) spectra (^{1}H, ^{13}C, HMBC, HSQC, DEPT135, DQF-COSY, HSQC-TOCSY, and NOESY) of purified compound **4** were reported on a JMN-ECS400 system (400 MHz; JEOL, Tokyo, Japan) in deuterated chloroform. Tetramethylsilane was added as an internal standard.

Phylogenetic tree analysis

Full-length protein sequences were collected from GenBank and aligned using CLUSTALW [26]. The unrooted Neighbor-joining tree was generated using MEGA6.06 software [23] (Bootstrap method, 10 000 replications).

Results

Isolation of CYP716A1 and CYP716A2

According to the TAIR website (https://www.arabidopsis.org/index.jsp), *CYP716A2* (*At5g36140*) encodes a polypeptide of 318 amino acid residues (AED94048.1), which shows high similarity to the N-terminal region of CYP716A1 (AED94045.1). However, this 318 amino acid polypeptide does not contain the cytochrome P450 cysteine heme-iron ligand signature, which is conserved among P450 superfamily proteins. Next to *CYP716A2*, a *cytochrome P450 superfamily protein* (*At5g36130*) is located over 2 kbp away from *At5g36140* in the *A. thaliana* genome. This gene encodes a protein of 140 amino acid residues (AED94047.1) and contains the cytochrome P450 cysteine heme-iron ligand signature. In addition, it shows high similarity to the C-terminal region of CYP716A1. On the basis of these observations, we deduced that one P450 protein, CYP716A2, is encoded by *At5g36130* and *At5g36140*. To isolate the full-length cDNA of *CYP716A1* (*At5g36110*) and *CYP716A2* (*At5g36140* and *At5g36130*), DNA fragments from the *A. thaliana* cDNA library were amplified with primers complementary to the 5′ and 3′ sequences of *CYP716A1* (*At5g36110*) for *CYP716A1*, and the 5′ sequence of *CYP716A2* (*At5g36140*) and the 3′ sequence of *cytochrome P450 family protein* (*At5g36130*) for *CYP716A2*. DNA fragments were cloned into pENTR-D-TOPO vectors and sequenced. The *CYP716A1* fragment completely matched the expected sequence, and the combined product of *At5g36140* and *At5g36130* was successfully amplified. Putative full-length CYP716A1 (NM_123002.1) and CYP716A2 (LC106013) cDNAs encoded polypeptides of 477 and 473 amino acid residues, respectively, and showed 83.7% homology with each other. Alignment of the deduced amino acid sequences with previously characterized CYP716A subfamily amyrin oxidases indicated that the cDNA sequences of CYP716A1 and CYP716A2 encoded complete CYP genes (Fig. S1). Unfortunately, the annotation of CYP716A1 in TAIR completely matched CYP716A2 in the Cytochrome P450 Homepage (accessed 29 January, 2016) [16], and CYP716A2 partially matched CYP716A1. To avoid a misunderstanding, we decided to use CYP716A1 and

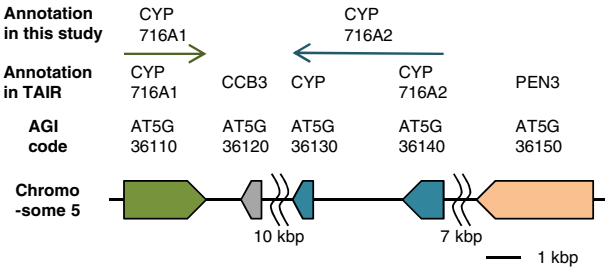

Fig. 1. Structure of the genomic region surrounding *CYP716As*. The genomic structure of chromosome 5 surrounding CYP716As is shown. Annotation and Arabidopsis Genome Initiative (AGI) code of each gene are cited from the TAIR homepage (https://www.arabidopsis.org/about/tairpubs.jsp, accessed October 7, 2015). CYP716A1 and CYP716A2 are indicated by green and blue arrows, respectively. CCB3, cofactor assembly complex C; CYP, cytochrome P450 superfamily protein; PEN3, pentacyclic triterpene synthase 3 (tirucalla-7,24-dien-3β-ol synthase).

CYP716A2 in place of At5g36110 and At5g36130 + At5g36140, respectively (Fig. 1).

Triterpene oxidizing activities of CYP716A1 and CYP716A2

To date, many CYP716 family enzymes, such as triterpene oxidases, CYP716A12 and other CYP716A subfamily enzymes have been reported to be C-28 oxidases that oxidize α-amyrin, β-amyrin and lupeol into ursolic acid, oleanolic acid, and betulinic acid, respectively [5,8]. Other CYP716 family enzymes such as CYP716Y1 from *Bupleurum falcatum* and CYP716A14v2 from *Artemisia annua* oxidize the C-16α and C-3 positions of the oleanane backbone, respectively [12,13]. As members of the CYP716 family, we expected that CYP716A1 and CYP716A2 would oxidize triterpene skeletons, such as α-amyrin, β-amyrin, and lupeol. First, we introduced one P450 expression vector into α-amyrin-, β-amyrin-, or lupeol-producing yeasts to confirm their oxidative activities against each triterpene backbone. However, only small peaks of oxidized products were detected by GC-MS analysis, probably due to the low oxidation activities of CYP716A enzymes or their low expression levels in yeast (Fig. S2). Therefore, to demonstrate the oxidative activities of CYP716A1 and CYP716A2, we increased the expression levels of each CYP716A gene in yeast by generating yeast strains harboring three P450 expression vectors (Fig. 2, Table S2).

In both αAS/CYP716A1 and αAS/CYP716A2 co-expressing yeast strains, uvaol (**2**) was detected, and its successive oxidation to ursolic acid (**3**) was confirmed in CYP716A1-expressing yeast, but not in CYP716A2-expressing yeast (Fig. 2A). The major product in αAS/CYP716A2 co-expressing yeast, compound **4**, was predicted to correspond to a mono-hydroxylated product of α-amyrin by MS analysis. A small amount of compound **4** was also detected in αAS/CYP716A1 co-expressing yeast. The mass spectra of compound **4** showed $m/z = 216$ as the main signal, which indicates the presence of a hydroxyl group on the D or E rings of the α-amyrin skeleton (Fig. S3A). Based on NMR analysis, compound **4** was identified as 22α-hydroxy-α-amyrin (Fig. S3). The chemical shift (^{13}C) matched well with a previous report [18] (Table S3).

In βAS/CYP716A1 co-expressing yeast, erythrodiol (**7**) and oleanolic acid (**8**) were detected as products of sequential oxidation at the C-28 position of β-amyrin (**6**) (Fig. 2B). In contrast, oleanolic acid was not detected in βAS/CYP716A2 co-expressing yeast. However, erythrodiol and two unknown peaks, **9** and **10**, presumably corresponding to hydroxylated β-amyrin compounds, were detected (Fig. 2B, Fig. S3B). Since the ion fragmentation pattern of peak **9** matched well with that of 16α-hydroxy-β-amyrin [12], we concluded that compound **9** was 16-hydroxy-β-amyrin. Since the mass spectra of compound **10** showed a $m/z = 216$ as the main signal, which indicates the presence of a hydroxyl group on the D or E ring of the β-amyrin skeleton (Fig. S3B), and CYP716A2 exhibited 22α-hydroxylation activity against α-amyrin, we hypothesized that CYP716A2 also exhibits 22α-hydroxylation activity against β-amyrin. Therefore, peak **10** was predicted to be 22α-hydroxy-β-amyrin.

Betulin (**12**) was detected in both LUS/CYP716A1 and LUS/CYP716A2 co-expressing yeast (Fig. 2C). However, betulinic acid (**13**) was not detected in either strain. In addition, two unknown peaks, **14** and **15**, were detected in LUS/CYP716A2 co-expressing yeast. Peak **15** was also observed in LUS/CYP716A1 co-expressing yeast as a minor product, and its retention time (18.12–18.13 min) was similar to that of betulinic acid (**13**) (18.11 min) (Fig. 2C); however, their mass fragmentation patterns were completely different (Fig. S3C). Thus, compounds **14** and **15** could not be identified in this work.

Discussion

The full-length coding sequences of *CYP716A1* and *CYP716A2* were cloned from the *A. thaliana* cDNA library. Although AT5G36130 and AT5G36140 were annotated as separate P450s on the TAIR website, our results indicate that AT5G36130 and AT5G36140 are transcribed as a single mRNA to encode a single protein, CYP716A2 (Fig. 1). To confirm their oxidative activities against α-amyrin, β-amyrin, and lupeol, we expressed CYP716A1 and CYP716A2 in triterpene skeleton-producing yeast strains. The results showed that CYP716A1 and CYP716A2 catalyzed the C-28 hydroxylation of α-amyrin, β-amyrin and lupeol, producing uvaol, erythrodiol and betulin, respectively (Fig. 2). However, only CYP716A1 showed carboxylative activity for the C-28 position of α-amyrin and β-amyrin. In addition, CYP716A2 catalyzed the C-22α hydroxylation of α-amyrin and β-amyrin, a catalytic activity that has not been previously reported. The main product of αAS and CYP716A2 co-expressing yeast, 22α-hydroxy-α-amyrin, was previously isolated from *Nardophyllum bryoides* as a cytotoxic compound against an LM3 (murine lung adenocarcinoma) cell line [18]. Interestingly, while the presence of 22α-hydroxy-α-amyrin in *A. thaliana* is unknown, our discovery allows us to produce this bioactive compound by using a heterologous yeast expression system. In a manner that differs from previously reported

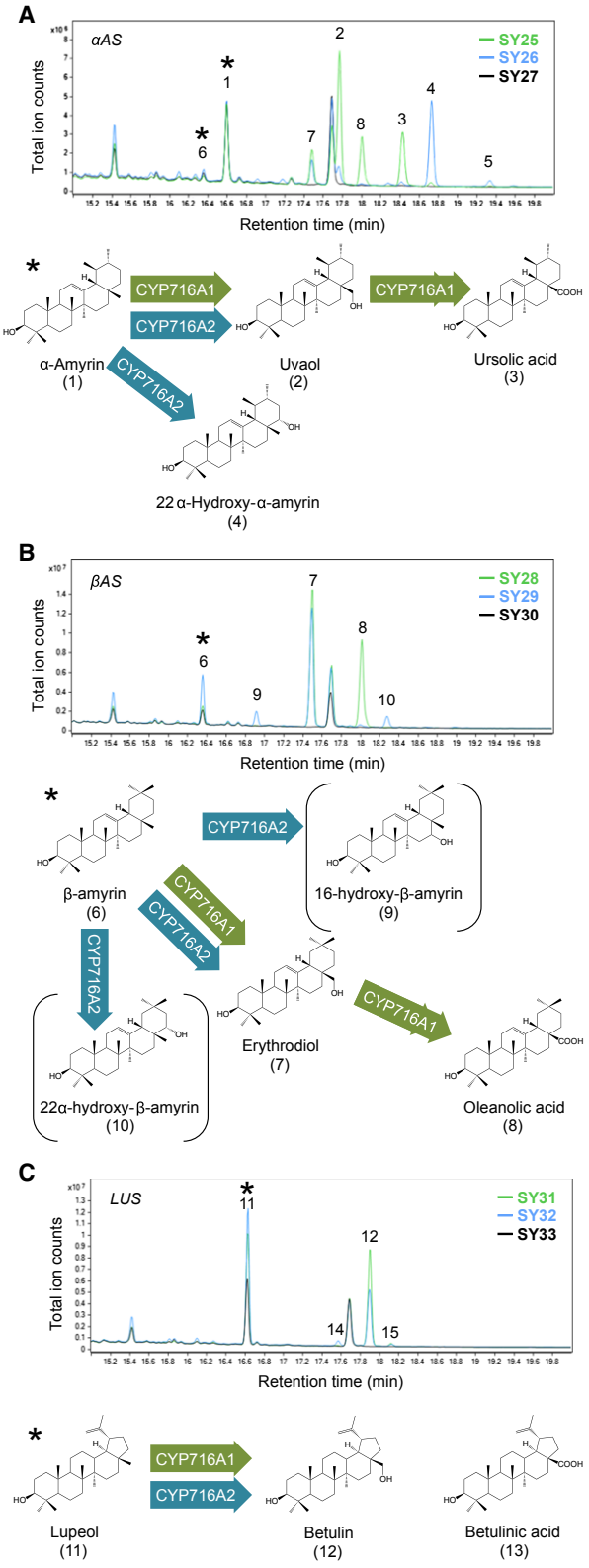

Fig. 2. *In vivo* activity of CYP716A1 and CYP716A2 toward triterpene skeletons. Total ion chromatograms (TICs) of yeast extracts and oxidative reactions catalyzed by CYP716A1 and CYP716A2. Yeast strains harboring one oxidosqualene cyclase expression vector and three P450 expression vectors (pELC-CYP716A, pYES-DEST52-CYP716A, and pESC-HIS-CYP716A) were used (see Table S2). CYP716A1 and CYP716A2 were co-expressed in yeast with α-amyrin synthase (αAS) (A), β-amyrin synthase (βAS) (B), and lupeol synthase (LUS) (C). TICs corresponding to CYP716A1-expressing yeast (SY25, SY28, and SY31), CYP716A2-expressing yeast (SY26, SY29, and SY32) and an empty vector-harboring yeast strain (SY27, SY30, and SY33) are indicated in green, blue, and black, respectively. Peaks corresponding to the reaction products of CYP716As are numbered. Oxidation reactions catalyzed by each CYP716A are indicated by arrows. Substrate peaks are marked with asterisks in each chromatogram. The predicted chemical structures of the product are shown in brackets. The ion fragmentation patterns of each product are shown in Fig. S3.

CYP716As, which catalyze the three step oxidation of lupeol into betulinic acid [5,8], both CYP716A1 and CYP716A2 catalyzed only the mono-hydroxylation of lupeol to produce betulin (Fig. 2C). Betulin has shown some activity, including antitumor [14] and sterol regulatory element-binding protein (SREBP) inhibition [24]. CYP716A1 and CYP716A2 could thus contribute to the production of betulin or its derivatives in engineered yeast using combinatorial biosynthesis. Therefore, this study indicates that the genes required for combinatorial biosynthesis of useful triterpenoids could be identified not only in medicinal/herbal plants, but also in plants that are not known to produce beneficial triterpenoids, such as *Arabidopsis*.

According to the phylogenetic analysis, both CYP716A1 and CYP716A2 are clustered together with *Barbarea vulgaris* (Brassicaceae) CYP716A80 and CYP716A81 (Fig. 3), which oxidize β-amyrin to produce erythrodiol, oleanolic acid, and two unknown oxidized compounds [10]. Since the ion fragmentation patterns of the two unknown compounds differ from those of the products of βAS/CYP716A2 (22α-hydroxy-β-amyrin and 16-hydroxy-β-amyrin), we assume that the CYP716As from *B. vulgaris* and *A. thaliana* produce different compounds. These results suggest that the ancestors of CYP716A2, CYP716A80, and CYP716A81 obtained novel activities during evolution. Functional analysis of other CYP716A subfamily proteins in other Brassicaceae plants will help reveal the functional evolution of CYP716A activities. CYP72A61v2 isolated from *Medicago truncatula* (Fabaceae) is a 22β-hydroxylase that oxidizes 24-hydroxy-β-amyrin to produce soyasapogenol B [6]. It modifies the oleanane skeleton at the

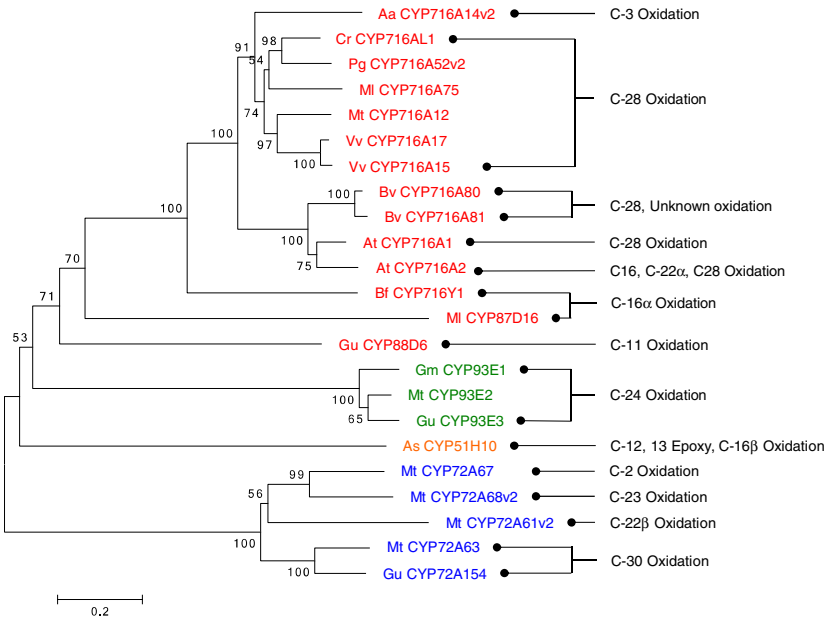

Fig. 3. Phylogenetic analysis of P450s that oxidize the oleanane skeleton. The rooted neighbor-joining tree was generated using CLUSTALW and MEGA6 software. The oxidation positions on α- or β-amyrin are indicated. Aa, *A. annua*; As, *A. strigosa*; At, *A. thaliana*; Bf, *B. falcatum*; Bv, *B. vulgaris*; Cr, *C. roseus*; Gm, *G. max*; Gu, *G. uralensis*; Ml, *M. lanceolata*; Mt, *M. truncatula*; Pg, *P. ginseng*; Vv, *V. vinifera*. P450s belonging to the CYP85, CYP51, CYP72, and CYP71 clans are represented by red, orange, blue, and green letters, respectively.

C-22 position, which is the same as CYP716A2, but has different stereo-specificity. As CYP716A and CYP72A belong to different clans (CYP85 clan and CYP72 clan, respectively) [1] (Fig. 3), these P450s likely independently obtained similar activities during evolution.

CYP716A1, *CYP716A2*, and *PEN3* make up a triterpene gene cluster in the *A. thaliana* genome. Boutanaev *et al.* [2] co-expressed PEN3 and CYP716A1 in yeast and *N. benthamiana,* and detected putative hydroxylated tirucalla-7,24-dien-3β-ol by GC-MS analysis. In this work, we confirmed the oxidative activities of CYP716A1 and CYP716A2 against α-amyrin, β-amyrin, and lupeol. Although the presence of α-amyrin and β-amyrin has been reported in *Arabidopsis* (in wild-type WS and Ler [9]), the presence of α-amyrin, β-amyrin, lupeol, and tirucalla-7,24-dien-3β-ol oxidized products in the yeast expression system was not reported. Therefore, the *in planta* functions of CYP716A1 and CYP716A2 remain unknown. To reveal the roles of CYP716A1 and CYP716A2 *in planta*, profiling the expression patterns and metabolites in wild-type plants and over-expressing and knockout lines will be required. Because CYP716A1 and CYP716A2 share some activity (C-28 hydroxylation), the generation of *cyp716a1 cyp716a2* double knockout plants will be necessary to reveal their *in planta* functions. Individual T-DNA-inserted single knockout plants are available; however, both P450 genes are located near each other on chromosome 5 (Fig. 1), making it difficult to create double knockout plants by crossing single knockout plants. In a previous study, *Arabidopsis cyp71a12 cyp71a13* double knockout plants were generated using genome-editing tools. This finding supports that artificial nucleases, such as transcription activator-like effector nuclease (TALEN), will be useful in the generation of multigene knockout lines of linked genes [15]. The generation of double knockout plants using genome-editing tools and analysis of these plants are ongoing, and will be the subject of future research efforts in our lab.

Acknowledgements

We thank Dr. Satoru Sawai for insightful and helpful discussion. We thank Dr. Anchaya Mongkolchaiyaphruek (Kasetsart University) and Keiko Fukamoto for their technical assistance. This study was partially supported by a Grant-in-Aid for JSPS (The Japan Society for the Promotion of Science) Fellows awarded to SY, a Grant-in-Aid for Scientific Research (B) from JSPS awarded to TM, the Frontier Research Base for Young Researchers, Osaka University from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT) awarded to EOF, and a research grant from the Nagase Science and Technology Foundation awarded to HS.

Author contributions

EOF, HS, and TM conceived and supervised the study; SY designed and performed experiments; SY and EOF analyzed data; SY wrote the manuscript; EOF, HS, and TM made manuscript revisions.

References

1 Bak S, Beisson F, Bishop G, Hamberger B, Höfer R, Paquette S and Werck-Reichhart D (2011) Cytochromes P450. *Arabidopsis Book* **9**, e0144.

2 Boutanaev AM, Moses T, Zi J, Nelson DR, Mugford ST, Peters RJ and Osbourn A (2015) Investigation of terpene diversification across multiple sequenced plant genomes. *Proc Natl Acad Sci USA* **112**, E81–E88.

3 Castillo DA, Kolesnikova MD and Matsuda SPT (2013) An effective strategy for exploring unknown metabolic pathways by genome mining. *J Am Chem Soc* **135**, 5885–5894.

4 Field B and Osbourn AE (2008) Metabolic diversification – independent assembly of operon-like gene clusters in different plants. *Science* **320**, 543–547.

5 Fukushima EO, Seki H, Ohyama K, Ono E, Umemoto N, Mizutani M, Saito K and Muranaka T (2011) CYP716A subfamily members are multifunctional oxidases in triterpenoid biosynthesis. *Plant Cell Physiol* **52**, 2050–2061.

6 Fukushima EO, Seki H, Sawai S, Suzuki M, Ohyama K, Saito K and Muranaka T (2013) Combinatorial biosynthesis of legume natural and rare triterpenoids in engineered yeast. *Plant Cell Physiol* **54**, 740–749.

7 Hamberger B and Bak S (2013) Plant P450s as versatile drivers for evolution of species-specific chemical diversity. *Philos Trans R Soc Lond B Biol Sci* **368**, 20120426.

8 Huang L, Li J, Ye H, Li C, Wang H, Liu B and Zhang Y (2012) Molecular characterization of the pentacyclic triterpenoid biosynthetic pathway in *Catharanthus roseus*. *Planta* **236**, 1571–1581.

9 Jenks MA, Tuttle HA, Eigenbrode SD and Feldmann KA (1995) Leaf epicuticular waxes of the eceriferum mutants in Arabidopsis. *Plant Physiol* **108**, 369–377.

10 Khakimov B, Kuzina V, Erthmann PØ, Fukushima EO, Augustin JM, Olsen CE, Scholtalbers J, Volpin H, Andersen SB, Hauser TP *et al.* (2015) Identification and genome organization of saponin pathway genes from a wild crucifer and their use for transient production of saponins in tobacco. *Plant J* **84**, 478–490.

11 Morlacchi P, Wilson WK, Xiong Q, Bhaduri A, Sttivend D, Kolesnikova MD and Matsuda SPT (2009) Product profile of PEN3: the last unexamined oxidosqualene cyclase in Arabidopsis thaliana. *Org Lett* **11**, 2627–2630.

12 Moses T, Pollier J, Almagro L, Buyst D, Van Montagu M, Pedreño MA, Martins JC, Thevelein JM and Goossens A (2014) Combinatorial biosynthesis of sapogenins and saponins in *Saccharomyces cerevisiae* using a C-16α hydroxylase from *Bupleurum falcatum*. *Proc Natl Acad Sci USA* **111**, 1634–1639.

13 Moses T, Pollier J, Shen Q, Soetaert S, Reed J, Erffelinck ML, Van Nieuwerburgh FCW, Bossche RV, Osbourn A, Thevelein JM *et al.* (2015) OSC2 and CYP716A14v2 catalyze the biosynthesis of triterpenoids for the cuticle of aerial organs of *Artemisia annua*. *Plant Cell* **27**, 286–301.

14 Mullauer FB, Kessler JH and Medema JP (2009) Betulin is a potent anti-tumor agent that is enhanced by cholesterol. *PLoS One* **4**, e1.

15 Müller TM, Böttcher C, Morbitzer R, Götz CC, Lehmann J, Lahaye T and Glawischnig E (2015) Transcription activator-like effector nuclease-mediated generation and metabolic analysis of camalexin-deficient cyp71a12 cyp71a13 double knockout lines. *Plant Physiol* **168**, 849–858.

16 Nelson DR (2009) The cytochrome p450 homepage. *Hum Genomics* **4**, 59.

17 Osbourn A, Goss RJM and Field RA (2011) The saponins – polar isoprenoids with important and diverse biological activities. *Nat Prod Rep* **28**, 1261–1268.

18 Sánchez M, Mazzuca M, Veloso MJ, Fernández LR, Siless G, Puricelli L and Palermo JA (2010) Cytotoxic terpenoids from *Nardophyllum bryoides*. *Phytochemistry* **71**, 1395–1399.

19 Sawai S and Saito K (2011) Triterpenoid biosynthesis and engineering in plants. *Front Plant Sci* **2**, doi: 10.3389/fpls.2011.00025.

20 Seki H, Ohyama K, Sawai S, Mizutani M, Ohnishi T, Sudo H, Akashi T, Aoki T, Saito K and Muranaka T (2008) Licorice β-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.

21 Seki H, Sawai S, Ohyama K, Mizutani M, Ohnishi T, Sudo H, Fukushima EO, Akashi T, Aoki T, Saito K *et al.* (2011) Triterpene functional genomics in licorice for identification of CYP72A154 involved in the biosynthesis of glycyrrhizin. *Plant Cell* **23**, 4112–4123.

22 Seki H, Tamura K and Muranaka T (2015) P450s and UGTs: key players in the structural diversity of triterpenoid saponins. *Plant Cell Physiol* **56**, 1463–1471.

23 Tamura K, Stecher G, Peterson D, Filipski A and Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol and Evol*, **30**, 2725–2729.

24 Tang J-J, Li JG, Qi W, Qiu WW, Li PS, Li BL and Song BL (2011) Inhibition of SREBP by a small molecule, betulin, improves hyperlipidemia and insulin resistance and reduces atherosclerotic plaques. *Cell Metab* **13**, 44–56.

25 Thimmappa R, Geisler K, Louveau T, O'Maille P and Osbourn A (2014) Triterpene biosynthesis in plants. *Annu Rev Plant Biol* **65**, 225–257.

26 Thompson JD, Higgins DG and Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site:
Fig. S1. Multiple alignment of selected CYP716A subfamily proteins.
Fig. S2. *In vivo* activity of CYP716A1 and CYP716A2 toward triterpene skeletons using yeast strains harboring one oxidosqualene cyclase expression vector and one P450 expression vector.
Fig. S3. Ion fragmentation patterns.
Fig. S4. Structural identification of unknown hydroxy-α-amyrin by NMR.
Table S1. Primers used in this study.
Table S2. Yeast strains used in this study.
Table S3. ^{13}C NMR spectra data of compound **4**.