Doctoral Dissertation

Identification of enzyme genes for triterpenoid biosynthesis and their transcriptional regulation in *Glycyrrhiza uralensis* and *Platycodon grandiflorus*

（薬用植物カンゾウおよびキキョウにおけるトリテルペノイド生合成酵素遺伝子の同定とその転写制御に関する研究）

Keita Tamura

December 2017

Cell Technology Laboratory

Department of Biotechnology, Graduate School of Engineering

Osaka University

Thesis committee

Professor Toshiya Muranaka, Ph.D.

Cell Technology Laboratory, Department of Biotechnology, Graduate School of Engineering, Osaka University

Professor Kazuhito Fujiyama, Ph.D.

Applied Microbiology Laboratory, International Center for Biotechnology, Osaka University

Professor Hajime Watanabe, Ph.D.

Bio-environmental Systems Engineering Laboratory, Department of Biotechnology, Graduate School of Engineering, Osaka University

Table of contents

List of abbreviations

aAS	α-amyrin synthase
ADH	alcohol dehydrogenase
ADS	amorphadiene synthase
ALDH	aldehyde dehydrogenase
bAS	β-amyrin synthase
bHLH	basic helix-loop-helix
CaMV	cauliflower mosaic virus
CAS	cycloartenol synthase
CDS	coding sequence
CPR	cytochrome P450 reductase
DMAPP	dimethylallyl diphosphate
EST	expressed sequence tag
ERF	ethylene response factor
FPP	farnesyl diphosphate
GA_3	gibberellin A_3
GC-MS	gas chromatography-mass spectrometry
GGPP	gereanylgeranyl diphosphate
GFP	green fluorescent protein
GPP	geranyl diphosphate
GUS	β-glucuronidase
HMM	hidden Markov model
IPP	isopentenyl diphosphate
IS	internal standard
KEGG	Kyoto Encyclopedia of Genes and Genomes
LUC	luciferase

LUS	lupeol synthase
MeJA	methyl jasmonate
MEP pathway	2-*C*-methyl-D-erythritol 4-phosphate pathway
MSD	mass selective detector
MVA pathway	mevalonate pathway
MS medium	Murashige and Skoog medium
NAA	1-naphthaleneacetic acid
NMR	nuclear magnetic resonance
ORF	open reading frame
OSC	oxidosqualene cyclase
OX	overexpression
P450	cytochrome P450 monooxygenase
PEG	polyethylene glycol
qPCR	quantitative real-time PCR
RACE	rapid amplification of cDNA ends
RNA-Seq	RNA sequencing
RPKM	reads per kilobase of exon model per million mapped reads
SA	salicylic acid
SD	standard deviation
SD medium	synthetic defined medium
SE	standard error
TF	transcription factor
TSS	transcription start site
UGT	UDP-dependent glycosyltransferase
Y1H	yeast one-hybrid
YE	yeast extract

Chapter 1

General introduction

1-1. Medicinal plants and specialized metabolites

Since ancient times, humans have used medicinal plants to treat diseases and stay healthy (Gurib-Fakim, 2006). Japanese traditional medicine, which is known as Kampo medicine, uses various medicinal plants in its prescriptions (Efferth *et al.*, 2007). Today, Kampo medicine is used in combination with modern medicine (Watanabe *et al.*, 2011).

The effects of medicinal plants are attributed to compounds found in these plants (Cragg and Newman, 2013), which also make plants a useful source for drug discovery (Rates, 2001; Cragg and Newman, 2013). The biologically active compounds found in medicinal plants are often secondary metabolites (Balandrin *et al.*, 1985). In plants, primary metabolites are distributed in virtually all plant species and are needed for fundamental cell metabolism processes, while secondary metabolites are biosynthesized from primary metabolites and have limited distributions in the plant kingdom (Balandrin *et al.*, 1985). Despite some pioneering studies on their physiological functions, such as studies of plant-insect interactions involving secondary metabolites conducted in the late 19th century, secondary metabolites were considered of "secondary importance" or "waste-products" until the 1980s (Hartmann, 2008). Since the importance of secondary metabolites for plant survival was recognized in the late 1950s (Fraenkel, 1959; Hartmann, 2008), their specialized functions in plants have been studied (Pichersky and Gang, 2000). Now, the term "specialized metabolites" is preferred to "secondary metabolites" (Gang, 2005).

The total number of metabolites produced by the plant kingdom is estimated to be range from 200,000 to 1,000,000 (Saito, 2013). Specialized metabolites can be classified into three groups: nitrogen-containing molecules (alkaloids), terpenoids, and phenolics (Schmidt *et al.*, 2007; Patra *et al.*, 2013). Of these, terpenoids are the most diverse family of plant specialized metabolites (Roberts, 2007; Hamberger and Bak, 2013). Terpenoids are biosynthesized from a five-carbon building block, isopentenyl diphosphate (IPP) and its allylic isomer dimethylallyl diphosphate (DMAPP), which are

Table 1-1. Representative triterpenoid-containing crude drugs used in 294 Kampo formulas approved by the Ministry of Health, Labour and Welfare of Japan

Japanese name[a]	English name[b]	Origin	Family	Order	Medicinal parts	Representatives of reported triterpenoids[c]	Ref[d]
甘草 (カンゾウ)	Glycyrrhiza (214)	*Glycyrrhiza uralensis*, *G. glabra*	Fabaceae	Fabales	root, stolon	1: glycyrrhizin	1
大棗 (タイソウ)	Jujube (90)	*Ziziphus jujuba*	Rhamnaceae	Rosales	fruit	1: oleanolic acid, maslinic acid; 2: ursolic acid; 3: betulinic acid, alphitolic acid; 4: jujuboside B, zizyphus saponin II	1,2
人参 (ニンジン)	Ginseng (75)	*Panax ginseng*	Araliaceae	Apiales	root	1: ginsenoside Ro; 4: ginsenoside Rg_1/Rg_3	2
柴胡 (サイコ)	Bupleurum Root (43)	*Bupleurum falcatum*	Apiaceae	Apiales	root	1: saikosaponin A/D	2
桔梗 (キキョウ)	Platycodon Root (31)	*Platycodon grandiflorus*	Campanulaceae	Asterales	root	1: platycodin D, polygalacin D	2
葛根 (カッコン)	Pueraria Root (14)	*Pueraria montana* var. *lobata*	Fabaceae	Fabales	root	1: soyasaponin I , kudzusaponin A_1/SA_3,	3
連翹 (レンギョウ)	Forsythia Fruit (14)	*Forsythia suspensa*	Oleaceae	Lamiales	fruit	1: oleanolic acid; 2: ursolic acid; 3: betulinic acid	1
木通 (モクツウ)	Akebia Stem (8)	*Akebia quinata*, *A. trifoliata*	Lardizabalaceae	Ranunculales	stem	1: akeboside Std/Ste	1
牛膝 (ゴシツ)	Achyranthes Root (7)	*Achyranthes fauriei*, *A. bidentata*	Amaranthaceae	Caryophyllales	root	1: achyranthoside A/B	4
山茱萸 (サンシュユ)	Cornus Fruit (6)	*Cornus officinalis*	Cornaceae	Cornales	accessory fruit	1: oleanolic acid; 2: ursolic acid	1
桑白皮 (ソウハクヒ)	Mulberry Bark (6)	*Morus alba*	Moraceae	Rosales	root bark	1: β-amyrin; 2: α-amyrin; 3: betulinic acid	1
酸棗仁 (サンソウニン)	Jujube Seed (5)	*Ziziphus jujuba*	Rhamnaceae	Rosales	seed	4: protojujuboside A; jujuboside B	1
山査子 (サンザシ)	Crataegus Fruit (4)	*Crataegus cuneata*, *C. pinnatifida*	Rosaceae	Rosales	accessory fruit	1: oleanolic acid, crataegolic acid (=maslinic acid); 2: ursolic acid	1
丁香/丁子 (チョウジ)	Clove (4)	*Syzygium aromaticum*	Myrtaceae	Myrtales	flower bud	1: oleanolic acid	1
遠志(オンジ)	Polygala Root (4)	*Polygala tenuifolia*	Polygalaceae	Fabales	root	1: onjisaponin F	1
*竹茹/竹茹 (チクジョ)	Bamboo Culm (4)	*Bambusa textilis*, *B. pervariabilis*, *B. beecheyana*, *B. tuldoides*, *Phyllostachys nigra* var. *henonis*, *P. bambusoides*	Poaceae	Poales	culm	3: lupeol, lupenone; 5: glutinone, *epi*-friedelanol, friedelin	1
威霊仙 (イレイセン)	Clematis Root (3)	*Clematis chinensis*, *C. mandshurica*, *C. hexapetala*	Ranunculaceae	Ranunculales	root, rhizome	1: oleanolic acid, hederagenin, clematichinenoside A/B	1
*紫菀/紫苑 (シオン)	Aster Root (3)	*Aster tataricus*	Asteraceae	Asterales	root, rhizome	1: astersaponin E, 5: *epi*-fridelanol, friedelin, shionone	1,5, 6
*烏梅(ウバイ)	Processed Mume (2)	*Prunus mume*	Rosaceae	Rosales	fruit	1: oleanolic acid	1
竹節人参 (チクセツニンジン)	Panax Japonicus Rhizome (2)	*Panax japonicus*	Araliaceae	Apiales	rhizome	1: chikusetsusaponin V (=ginsenoside Ro); 4: chikusetsusaponin III	1
*柿蒂(シテイ)	Persimmon Calyx (2)	*Diospyros kaki*	Ebenaceae	Ericales	calyx	1: oleanolic acid; 2: ursolic acid; 3: betulinic acid	1
枇杷葉 (ビワヨウ)	Loquat Leaf (2)	*Eriobotrya japonica*	Rosaceae	Rosales	leaf	1: oleanolic acid, maslinic acid; 2: ursolic acid, corosolic acid, euscaphic acid	1,2
*竹葉 (チクヨウ)	Bamboo Leaf (1)	*Phyllostachys nigra* var. *henonis*, *P. bambusoides*, *Bambusa textilis*, *B. emeiensis*	Poaceae	Poales	leaf	3: lupenone; 5: taraxerol, glutinol, glutinone, *epi*-fridelanol, friedelin	1

[a]Japanese name without asterisks indicates a crude drug listed in the Japanese Pharmacopoeia 17th Edition. Japanese name with asterisks indicates a crude drug listed in the Japanese standards for non-pharmacopoeial crude drugs 2015.

[b]The numbers of formulas in the 294 Kampo formulas approved by the Ministry of Health, Labour and Welfare of Japan (Nemoto, 2016) are shown in parentheses.

[c]Reported in at least one of the references. Triterpenoids are listed based on their skeletons (1, oleanane; 2, ursane; 3, lupane; 4, dammarane; 5, others) and in order of appearance in Figure 1-1.

[d]Reference: 1, KEGG Drug (http://www.genome.jp/kegg/drug/drug_ja.html); 2, Comprehensive Medicinal Plant

Database (http://mpdb.nibiohn.go.jp/), Research Center for Medicinal Plant Resources, National Institute of Biomedical Innovation, Japan (2013–); 3, Wong *et al.*, 2011; 4, Ida *et al.*, 1994; 5, Sawai *et al.*, 2011; 6, Nagao *et al.*, 1990.

biosynthesized via either the cytosolic mevalonate (MVA) pathway or the plastidial 2-*C*-methyl-D-erythritol 4-phosphate (MEP) pathway (Roberts, 2007). Terpenoids are further classified based on the number of C5 building blocks used (Hamberger and Bak, 2013). The condensation of IPP and DMAPP yields geranyl diphosphate (GPP, C_{10}), the precursor of monoterpenoids. The condensation of IPP and GPP yields farnesyl diphosphate (FPP, C_{15}), the precursor of sesquiterpenoids. Furthermore, the condensation of IPP and FPP yields geranylgeranyl diphosphate (GGPP, C_{20}), the precursor of diterpenoids. The condensation of two molecules of FPP yields squalene (C_{30}) the precursor of triterpenoids. The condensation of two molecules of GGPP yields phytoene (C_{40}), the precursor of tetraterpenoids (Roberts, 2007). In general, the MVA pathway involves the production of the precursors for sesquiterpenoids and triterpenoids, while the MEP pathway involves the production of the precursors for monoterpenoids, diterpenoids, and tetraterpenoids (Roberts, 2007). Some terpenoids have basic functions in plants in general and are regarded as primary metabolites. These primary terpenoids include gibberellins (diterpenoids), carotenoids (tetraterpenoids), and sterols (triterpenoids) (Roberts, 2007).

1-2. Overview of triterpenoid biosynthesis

Triterpenoids can be found in various medicinal plants, and many of these triterpenoids are bioactive, with anti-inflammatory, antiviral, cytotoxic, or antitumor activities (Dinda *et al.*, 2010). I extracted representative triterpenoids from the medicinal plants that are used in Kampo formulas approved by the Ministry of Health, Labour and Welfare of Japan (Table 1-1 and Figure 1-1). Some triterpenoids are widespread across the plant kingdom, such as oleanolic acid and ursolic acid, while others are found only in a very few species, such as glycyrrhizin. Many of these triterpenoids contain glycosyl groups. Triterpenoids containing glycosyl groups are called triterpenoid saponins, and their non-sugar parts (aglycones) are called sapogenins.

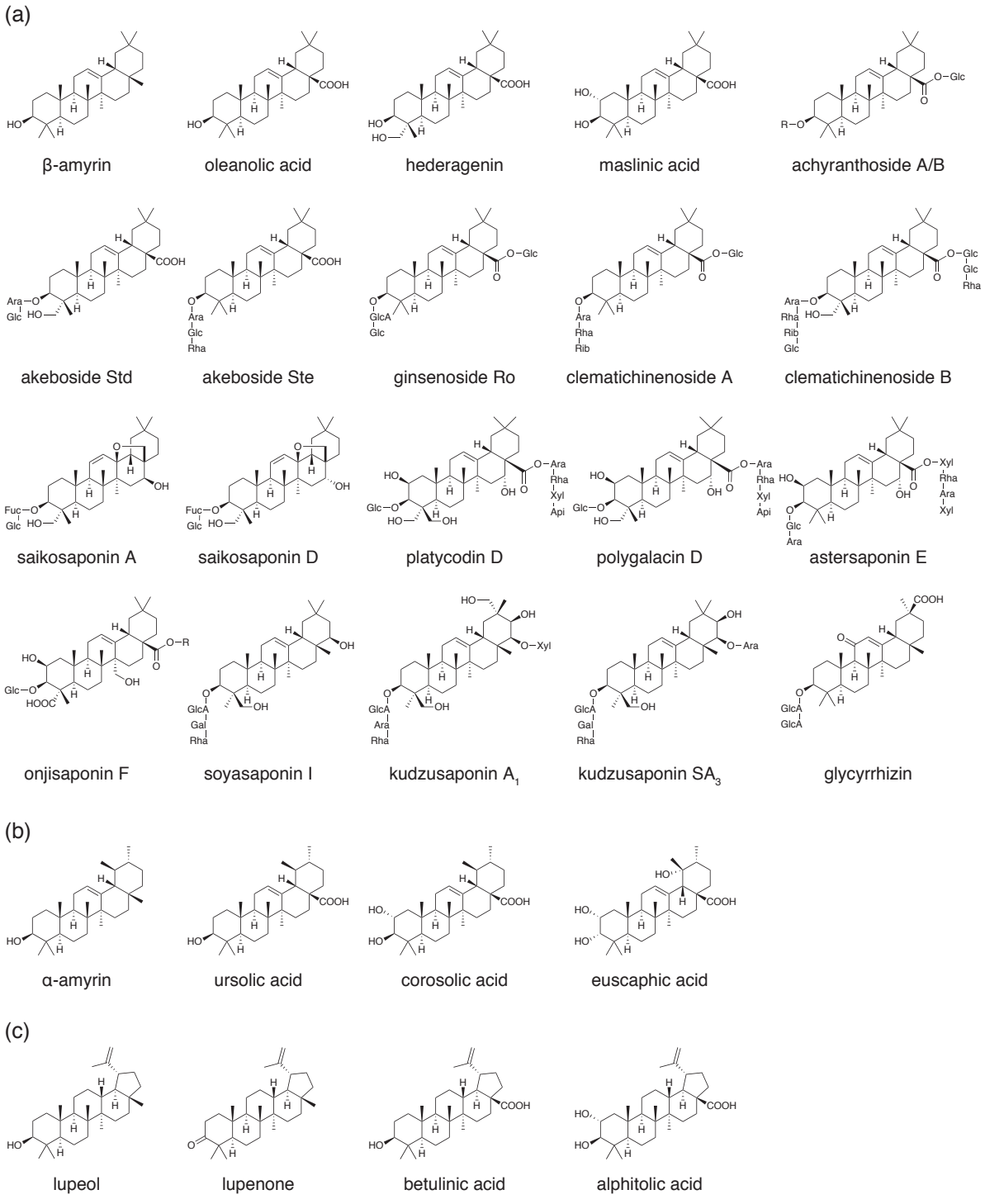


Figure 1-1. Structures of the triterpenoids shown in Table 1-1. Triterpenoids are grouped based on their skeletons: (a) Oleanane, (b) ursane, (c) lupane, (d) dammarane, and (e) others. Sugar moieties are shown as abbreviated forms. The complex achyranthoside A/B and onjisaponin F functional groups are omitted and shown as R. Abbreviations: Api, apiose; Ara, arabinose; Fuc, fucose; Gal, galactose; Glc, glucose; GlcA, gulucuronic acid; Rib, ribose; Rha, rhamnose; Xyl, xylose.

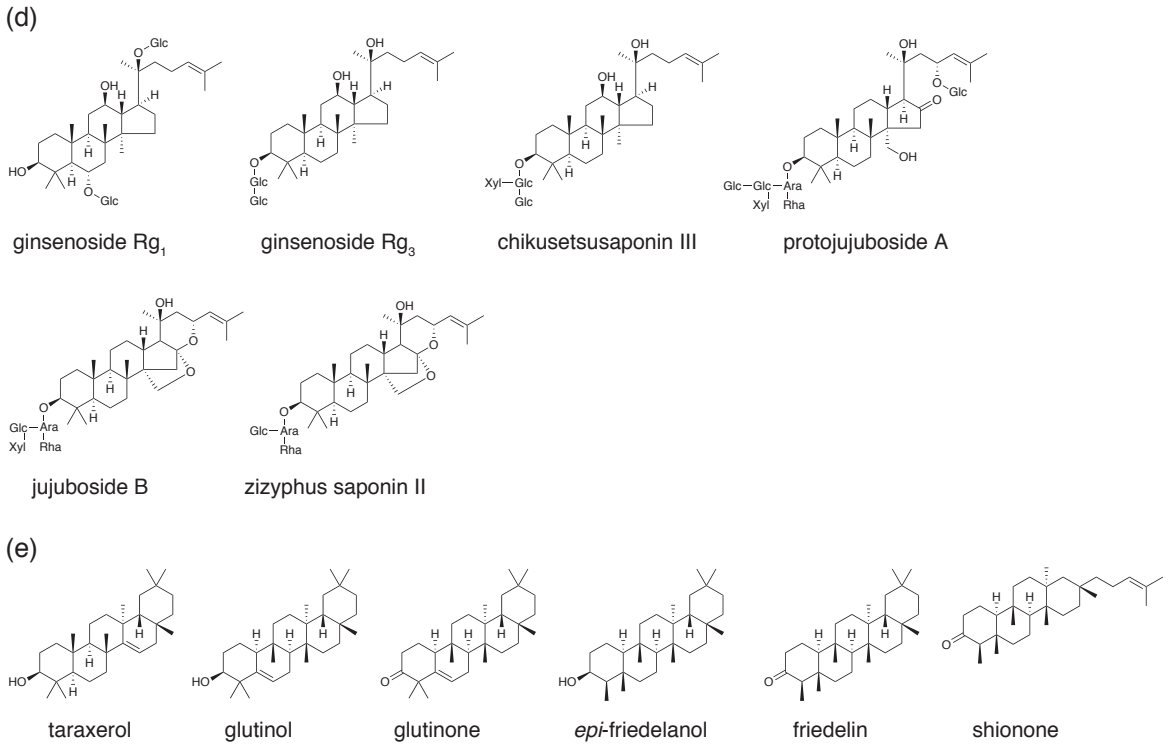


Figure 1-1. (continued)

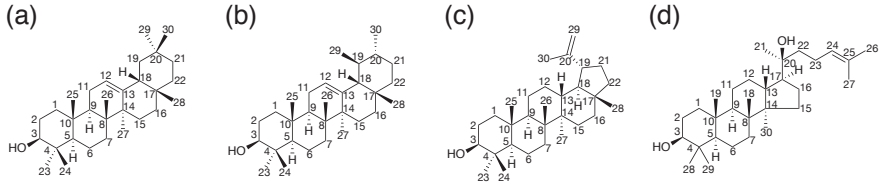


Figure 1-2. Common triterpene skeletons: (a) β-amyrin (oleanane), (b) α-amyrin (ursane), (c) lupeol (lupane), and (d) dammarenediol II (dammarane). Numbers on each triterpene skeletons indicate carbon positions.

Currently, more than 14,000 triterpenoid molecules have been reported as present in the plant kingdom (Hamberger and Bak, 2013). Triterpenoids can be classified based on their skeletons, and include dammarane, lupane, ursane, and oleanane types (Hill and Connolly, 2012) (Figure 1-2). In higher plants, triterpene skeletons originate from the cyclization of 2,3-oxidosqualene, which is biosynthesized from squalene by squalene epoxidase (Figures 1-3a,b). Oxidosqualene cyclases (OSCs) cyclize 2,3-oxidosqualene into a variety of triterpene skeletons. Most triterpenoids are tetracycles or pentacycles. The cyclization of tetra- or pentacyclic triterpenoids from 2,3-oxidosqualene involves two different intermediates: the protosteryl cation, which has a chair-boat-chair conformation, or the dammarenyl cation, which has a chair-chair-chair conformation (Xu *et al.*, 2004). Cycloartenol, a

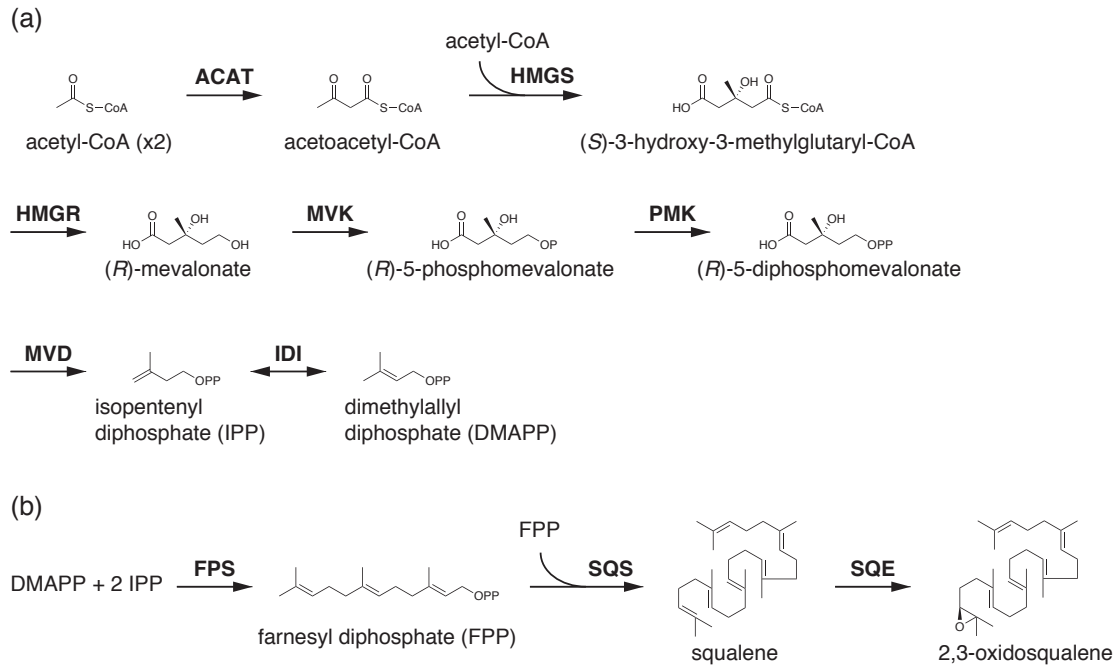


Figure 1-3. Biosynthetic pathway of 2,3-oxidosqualene.
(a) Mevalonate (MVA) pathway producing isopentenyl diphosphate (IPP) and its allylic isomer dimethylallyl diphosphate (DMAPP).
(b) Condensation of isoprene units (IPP and DMAPP) producing squalene and its epoxidation to generate 2,3-oxidosqualene.
Enzymes are indicated in bold. Abbreviations: ACAT, acetyl-CoA C-acetyltransferase; FPS, farnesyl diphosphate synthase; HMGR, hydroxymethylglutaryl-CoA reductase; HMGS, hydroxymethylglutaryl-CoA synthase; IDI, isopentenyl diphosphate isomerase; MVD, diphosphomevalonate decarboxylase; MVK, mevalonate kinase; PMK, phosphomevalonate kinase; SQE, squalene epoxidase; SQS, squalene synthase.

precursor of the primary membrane sterols of plants, such as β-sitosterol and stigmasterol, or plant hormone brassinosteroids, is biosynthesized via the protosteryl cation. In comparison, most of the diverse specialized triterpenoids are biosynthesized via the dammarenyl cation (Thimmappa *et al.*, 2014). In this dissertation, I subsequently refer to the specialized triterpenoids derived from the dammarenyl cation as "triterpenoids", based on the classification presented in a review paper (Thimmappa *et al.*, 2014). The first cloned plant OSC was *cycloartenol synthase* (*CAS1*) from *Arabidopsis thaliana* (Corey *et al.*, 1993). This discovery led to the identification of plant OSCs from various plant species using homology-based cloning (Phillips *et al.*, 2006). Today, more than 80 OSCs have been functionally characterized from various plant species (Thimmappa *et al.*, 2014).

Some triterpenoids found in plants, such as β-amyrin (oleanane), α-amyrin (ursane), and lupeol (lupane), are simply triterpene skeletons generated through the cyclization of 2,3-oxidosqualene by OSCs. However, most of the diverse triterpenoids undergo further modification reactions, such as

oxidation, glycosylation, and acetylation reactions (Thimmappa *et al.*, 2014). Recently, cytochrome P450 monooxygenases (P450s) catalyzing the site-specific oxidation of triterpene skeletons and UDP-dependent glycosyltransferases (UGTs) transferring sugar moieties to aglycones or triterpene glycosides have been characterized from various plant species (Seki *et al.*, 2015). In triterpenoid saponin biosynthesis, glycosylation often occurs at hydroxyl or carboxyl groups introduced during the oxidation of triterpene skeletons. Therefore, the identification of triterpene oxidases is the first important approach for elucidating the molecular mechanisms of the structural diversification of triterpenoids.

The first P450 found to act as a triterpene oxidase was CYP93E1, which was identified in soybean (*Glycine max*) (Shibuya *et al.*, 2006). Since this discovery, several P450s have been identified as triterpene oxidases (Table 1-2). P450s are one of the largest plant gene superfamilies and comprise approximately 1% of all protein-coding genes in plant genomes (Nelson and Werck-Reichhart, 2011). For example, *A. thaliana*, *Oryza sativa* (rice), and *G. max* have 245, 334, and 332 P450s in their respective genomes (Nelson and Werck-Reichhart, 2011). In comparison, baker's yeast (*Saccharomyces cerevisiae*) has only three P450s (Werck-Reichhart and Feyereisen, 2000), while *Homo sapiens* (human) has 57 P450 genes (Nelson, 2003). P450s are named using a nomenclature system based on homology and phylogenetic criteria (Nelson and Werck-Reichhart, 2011). P450s in the same family (e.g., within the CYP93 and CYP716 families) share at least 40% amino acid identity, while P450s in the same subfamily (e.g., within the CYP93E and CYP716A subfamilies) share at least 55% amino acid identity, and P450s in the same subfamily are numbered according to the order of naming (e.g., CYP93E1, CYP93E2, CYP93E3…) (Bak *et al.*, 2011). P450s in plants are grouped into 127 families, compared with 19 families in vertebrates and 67 families in insects, although this is less than the 399 families in fungi and 333 families in bacteria (Nelson and Werck-Reichhart, 2011). The highly diversified plant P450s have various functions, including essential functions such as the production of membrane sterols, carotenoids needed for light harvesting, or the biosynthesis and catabolism of plant hormones, and the biosynthesis of various classes of specialized metabolites, including triterpenoids (Bak *et al.*, 2011; Nelson and Werck-Reichhart, 2011). More than half of the reported P450s that function as triterpene oxidases catalyze oxidation reactions at the β-amyrin

skeleton. The CYP716A subfamily contains the best-known triterpene oxidases. They have been identified from various plant species across various families; many catalyze oxidation at the C-28 position of β-amyrin, while some have also been reported to catalyze oxidation at the C-28 positions of α-amyrin and lupeol. Other well-characterized subfamilies are the CYP72A and CYP93E subfamilies; however, unlike the CYP716A subfamily, these subfamilies catalyze oxidation reactions of the β-amyrin skeleton only in the plant family Fabaceae (Seki *et al.*, 2015).

1-3. Biotechnological applications of enzyme genes for specialized metabolites

The discovery of the enzymes involved in triterpenoid biosynthesis has made it theoretically possible to produce a variety of triterpenoids in heterologous hosts by combining the appropriate enzymes from one or more plant species. Generally, the accumulation of a pharmaceutically active compound in the plant tissues of medicinal plants requires a long time, and extraction from natural resources may lead to the extinction of valuable plant species. Recently, the price of many crude drugs used in Kampo medicine has increased (Japan Kampo Medicines Manufacturers Association, 2015). One promising approach is the production of bioactive specialized metabolites in heterologous hosts. For example, the heterologous production of taxadiene, the first key intermediate in the biosynthesis of taxol (a diterpenoid produced by *Taxus brevifolia*), was achieved in *Escherichia coli* (Ajikumar *et al.*, 2010). *S. cerevisiae* is widely used for the heterologous production of plant specialized metabolites and several biotechnology companies are now developing commercial production processes (Leavell *et al.*, 2016). One of the most successful examples is the production of artemisinic acid, a precursor of artemisinin (a sesquiterpenoid produced by *Artemisia annua*), in *S. cerevisiae* (Paddon *et al.*, 2013). In this case, the expression of *amorphadiene synthase* (*ADS*), *CYP71AV1*, and its native redox partner, *cytochrome P450 reductase* (*CPR*), was sufficient for the conversion of FPP into artemisinic acid (Ro *et al.*, 2006); however, the co-expression of a previously identified aldehyde dehydrogenase (*ALDH1*) (Teoh *et al.*, 2009) and a novel alcohol dehydrogenase (*ADH1*), both of which are highly expressed in the glandular trichome of *A. annua*, in which artemisinin is biosynthesized and accumulated, was a key factor in the high production of artemisinic acid in *S. cerevisiae* (Paddon *et al.*, 2013). Therefore, even in an

attempt to produce specialized metabolites using heterologous hosts, it is important to understand the molecular mechanisms of the biosynthesis of specialized metabolites in their native plants. Another approach is the use of the native host plant to obtain valuable specialized metabolites. Cultivation of medicinal plants or cultures of their tissues could solve the problems of limited access to natural resources; however, the production levels of specialized metabolites in plants are usually low and often require specific stimuli to induce their production (Bourgaud *et al.*, 2001). The screening of high-producing plant or cell lines and optimization of the conditions for abundant production are important steps (Bourgaud *et al.*, 2001). The identification of biosynthetic enzyme genes could provide marker genes for effective screening. Moreover, the identification of limiting enzymes makes it possible to engineer the metabolic pathways of plants (Bourgaud *et al.*, 2001).

1-4. Transcriptional regulation of the biosynthesis of specialized metabolites

In addition to the elucidation of enzymes catalyzing the transformation of plant specialized metabolites from simple precursor compounds, it is important to understand the regulatory mechanisms underlying the biosynthesis of specialized metabolites in plant tissues. Generally, plants produce specialized metabolites in specific tissues or under specific conditions (Yang *et al.*, 2012). For example, artemisinin accumulates only in the glandular trichome of *A. annua* leaves (Zhang *et al.*, 2008), while many ripening fruits, such as grapes and apples, accumulate anthocyanins during fruit ripening (Allan *et al.*, 2008). Plants have a mechanism to synthesize and accumulate specialized metabolites when needed by coordinately expressing several biosynthetic enzyme genes for the production of specific specialized metabolites and accompanying genes for the transportation or accumulation of those metabolites. This coordinated recruitment of multiple genes is usually achieved through transcription factors (TFs), which are regulatory proteins that bind to specific DNA motifs (*cis*-elements) on the promoter sequences of their target genes, and modulate the expression of their target genes (Yang *et al.*, 2012). TFs involved in regulating the biosynthesis of plant specialized metabolites have been isolated from various plant species (Yang *et al.*, 2012; Patra *et al.*, 2013) and TFs involved in flavonoid and alkaloid biosynthesis have been well characterized. Various TFs have

Table 1-2. List of P450s involved in triterpenoid biosynthesis

P450 name[a]	Species	Family	Order	Reaction (substrate)[b]	Reference	Accession no.[c]
CYP51H10	*Avena strigosa*	Poaceae	Poales	C-12,13β epoxidase; C-16β oxidase (1)	Geisler *et al*., 2013	ABG88961
CYP71A16	*Arabidopsis thaliana*	Brassicaceae	Brassicales	C-23 oxidase (6,7)	Castillo *et al*., 2013	AED94832
CYP71D353	*Lotus japonicus*	Fabaceae	Fabales	C-20, C-28 oxidase (3)	Krokida *et al*., 2013	AHB62239
CYP72A61v2	*Medicago truncatula*	Fabaceae	Fabales	C-22β oxidase (1)	Fukushima *et al*., 2013	BAL45199
CYP72A63	*Medicago truncatula*	Fabaceae	Fabales	C-30 oxidase (1)	Seki *et al*., 2011	BAL45200
CYP72A67	*Medicago truncatula*	Fabaceae	Fabales	C-2β oxidase (1)	Biazzi *et al*., 2015	ABC59075
CYP72A68v2	*Medicago truncatula*	Fabaceae	Fabales	C-23 oxidase (1)	Fukushima *et al*., 2013	BAL45204
CYP72A69	*Glycine max* cv. Enrei	Fabaceae	Fabales	C-21β oxidase (1)	Yano *et al*., 2017	BAW35009
CYP72A154	*Glycyrrhiza uralensis*	Fabaceae	Fabales	C-30 oxidase (1)	Seki *et al*., 2011	BAL45207
CYP72A566	*Glycyrrhiza uralensis*	Fabaceae	Fabales	C-22β oxidase (1)	**This study**	LC318134[d]
CYP87D16	*Maesa lanceolata*	Primulaceae	Ericales	C-16α oxidase (1)	Moses *et al*., 2015a	AHF22090
CYP88D6	*Glycyrrhiza uralensis*	Fabaceae	Fabales	C-11 oxidase (1)	Seki *et al*., 2008	BAG68929
CYP93E1	*Glycine max*	Fabaceae	Fabales	C-24 oxidase (1)	Shibuya *et al*., 2006	BAE94181
CYP93E2	*Medicago truncatula*	Fabaceae	Fabales	C-24 oxidase (1)	Fukushima *et al*., 2013	ABC59085
CYP93E3	*Glycyrrhiza uralensis*	Fabaceae	Fabales	C-24 oxidase (1)	Seki *et al*., 2008	BAG68930
CYP93E4	*Arachis hypogaea*	Fabaceae	Fabales	C-24 oxidase (1)	Moses *et al*., 2014b	AIN25416
CYP93E5	*Cicer arietinum*	Fabaceae	Fabales	C-24 oxidase (1)	Moses *et al*., 2014b	AIN25417
CYP93E6	*Glycyrrhiza glabra*	Fabaceae	Fabales	C-24 oxidase (1)	Moses *et al*., 2014b	AIN25418
CYP93E7	*Lens culinaris*	Fabaceae	Fabales	C-24 oxidase (1)	Moses *et al*., 2014b	AIN25419
CYP93E8	*Pisum sativum*	Fabaceae	Fabales	C-24 oxidase (1)	Moses *et al*., 2014b	AIN25420
CYP93E9	*Phaseolus vulgaris*	Fabaceae	Fabales	C-24 oxidase (1)	Moses *et al*., 2014b	AIN25421
CYP705A1	*Arabidopsis thaliana*	Brassicaceae	Brassicales	C-15,16 cleavarage (8)	Castillo *et al*., 2013	AEE83585
CYP708A2	*Arabidopsis thaliana*	Brassicaceae	Brassicales	C-7β oxidase (9)	Castillo *et al*., 2013	AED95604
CYP716A1	*Arabidopsis thaliana*	Brassicaceae	Brassicales	C-28 oxidase (1,2,3)	Yasumoto *et al*., 2016	AED94045
CYP716A2	*Arabidopsis thaliana*	Brassicaceae	Brassicales	C-22α oxidase (2); C-28 oxidase (1,2,3)	Yasumoto *et al*., 2016	BAU61505
CYP716A12	*Medicago truncatula*	Fabaceae	Fabales	C-28 oxidase (1,2,3)	Carelli *et al*., 2011 Fukushima *et al*., 2011	ABC59076
CYP716A14v2	*Artemisia annua*	Asteraceae	Asterales	C-3 oxidase (1,2,3)	Moses *et al*., 2015b	AHF22083
CYP716A15	*Vitis vinifera*	Vitaceae	Vitales	C-28 oxidase (1,2,3)	Fukushima *et al*., 2011	BAJ84106
CYP716A17	*Vitis vinifera*	Vitaceae	Vitales	C-28 oxidase (1)	Fukushima *et al*., 2011	BAJ84107
CYP716A44	*Solanum lycopersicum*	Solanaceae	Solanales	C-28 oxidase (1,2)	Yasumoto *et al*., 2017	AK329870[d]
CYP716A46	*Solanum lycopersicum*	Solanaceae	Solanales	C-28 oxidase (1,2)	Yasumoto *et al*., 2017	XP_004243906
CYP716A52v2	*Panax ginseng*	Araliaceae	Apiales	C-28 oxidase (1)	Han *et al*., 2013	AFO63032
CYP716A75	*Maesa lanceolata*	Primulaceae	Ericales	C-28 oxidase (1)	Moses *et al*., 2015a	AHF22088
CYP716A78	*Chenopodium quinoa*	Amaranthaceae	Caryophyllales	C-28 oxidase (1)	Fiallos-Jurado *et al*., 2016	ANY30853
CYP716A79	*Chenopodium quinoa*	Amaranthaceae	Caryophyllales	C-28 oxidase (1)	Fiallos-Jurado *et al*., 2016	ANY30854
CYP716A80	*Barbarea vulgaris* subsp. *arcuata*	Brassicaceae	Brassicales	C-28 oxidase (1)	Khakimov *et al*., 2015	ALR73782
CYP716A81	*Barbarea vulgaris* subsp. *arcuata*	Brassicaceae	Brassicales	C-28 oxidase (1)	Khakimov *et al*., 2015	ALR73781
CYP716A83	*Centella asiatica*	Apiaceae	Apiales	C-28 oxidase (1,2)	Miettinen *et al*., 2017	AOG74832
CYP716A86	*Centella asiatica*	Apiaceae	Apiales	C-28 oxidase (1)	Miettinen *et al*., 2017	AOG74831
CYP716A110	*Aquilegia coerulea*	Ranunculaceae	Ranunculales	C-28 oxidase (1)	Miettinen *et al*., 2017	AOG74847
CYP716A111	*Aquilegia coerulea*	Ranunculaceae	Ranunculales	C-16β oxidase (1)	Miettinen *et al*., 2017	APG38190
CYP716A140	*Platycodon grandiflorus*	Campanulaceae	Asterales	C-28 oxidase (1)	Miettinen *et al*., 2017	AOG74836
CYP716A140v2	*Platycodon grandiflorus*	Campanulaceae	Asterales	C-28 oxidase (1)	**This study**	BAX04007
CYP716A141	*Platycodon grandiflorus*	Campanulaceae	Asterales	C-16β oxidase (1)	**This study** Miettinen *et al*., 2017	BAX04008
CYP716A154 (CYP716AL1)	*Catharanthus roseus*	Apocynaceae	Gentianales	C-28 oxidase (1,2,3)	Huang *et al*., 2012	AEX07773
CYP716A175	*Malus domestica*	Rosaceae	Rosales	C-28 oxidase (1,2,3,5)	Andre *et al*., 2016	EB148173[d,e]
CYP716A179	*Glycyrrhiza uralensis*	Fabaceae	Fabales	C-28 oxidase (1,2,3)	**This study**	BAW34647
CYP716A180	*Betula platyphylla*	Betulaceae	Fagales	C-28 oxidase (3)	Zhou *et al*., 2016	AHL46848
CYP716A244	*Eleutherococcus senticosus*	Araliaceae	Apiales	C-28 oxidase (1)	Jo *et al*., 2017	APZ88353
CYP716A252	*Ocimum basilicum*	Lamiaceae	Lamiales	C-28 oxidase (1,2)	Misra *et al*., 2017	AFZ40057
CYP716A253	*Ocimum basilicum*	Lamiaceae	Lamiales	C-28 oxidase (1,2)	Misra *et al*., 2017	AFZ40058
CYP716C11	*Centella asiatica*	Apiaceae	Apiales	C-2α oxidase (1,2)	Miettinen *et al*., 2017	AOG74835
CYP716E26 (CYP716A43)	*Solanum lycopersicum*	Solanaceae	Solanales	C-6β oxidase (1,2)	Yasumoto *et al*., 2017	XP_004241821
CYP716E41	*Centella asiatica*	Apiaceae	Apiales	C-6β oxidase (1,2)	Miettinen *et al*., 2017	AOG74834
CYP716S1v2 (CYP716A53v2)	*Panax ginseng*	Araliaceae	Apiales	C-6α oxidase (4)	Han *et al*., 2012	AFO63031
CYP716S5	*Platycodon grandiflorus*	Campanulaceae	Asterales	C-12,13α epoxidase (1)	Miettinen *et al*., 2017	AOG74839

Table 1-2. (continued)

P450 name[a]	Species	Family	Order	Reaction (substrate)[b]	Reference	Accession no.[c]
CYP716U1 (CYP716A47)	*Panax ginseng*	Araliaceae	Apiales	C-12β oxidase (4)	Han *et al.*, 2011	AEY75212
CYP716Y1	*Bupleurum falcatum*	Apiaceae	Apiales	C-16α oxidase (1)	Moses *et al.*, 2014a	AHF45909

[a]Some P450s have been renamed recently (Miettinen *et al.*, 2017; Yasumoto *et al.*, 2017). Outdated names are shown in parentheses. P450s characterized in this study are indicated in bold.
[b]The determined biochemical activities of the P450s. Numbers in parentheses show their substrate triterpene skeletons or compounds: 1, oleanane; 2, ursane; 3, lupane; 4, dammarane; 5, germanicane; 6, marnerol; 7, marneral; 8, arabidiol; 9, thalianol.
[c]GenBank protein accession number.
[d]Accession number of nucleotide.
[e]Partial sequence.

been identified in flavonoid biosynthesis, including MYB and basic helix-loop-helix (bHLH) (Koes *et al.*, 2005; Xu *et al.*, 2015). In alkaloid biosynthesis, the main TFs that have been identified are involved in the biosynthesis of nicotine (Shoji *et al.*, 2010; Todd *et al.*, 2010; Shoji and Hashimoto, 2011) and terpenoid indole alkaloids (van der Fits and Memelink, 2000; Suttipanta *et al.*, 2011; Zhang *et al.*, 2011; Van Moerkercke *et al.*, 2015, 2016). In comparison with the TFs involved in these two classes of specialized metabolites, TFs involved in terpenoid synthesis have been less well characterized (Patra *et al.*, 2013); however, there are several reports on TFs regulating the biosynthesis of artemisinin in *A. annua* (Yu *et al.*, 2012; Lv *et al.*, 2016; Shen *et al.*, 2016). These TFs have helped our understanding of how plants act to produce specific specialized metabolites. Moreover, the identification of TFs could be a useful tool for engineering metabolic pathways in plants that suppress the pathways for undesired side products and enhance the production of valuable compounds (Grotewold, 2008). One successful example is the expression of two TFs from snapdragon (*Antirrhinum majus*): Delila (bHLH) and Rosea1 (MYB-related). These TFs interact to induce anthocyanin biosynthesis in snapdragon flowers, and they significantly increased the anthocyanin content in the fruit of transgenic tomatoes (Butelli *et al.*, 2008). Nevertheless, the TFs involved in triterpenoid biosynthesis remain largely unknown.

1-5. Scope and outline of this study

The purpose of this study is to elucidate molecular mechanisms of triterpenoid biosynthesis in medicinal plants by identifying P450 genes and the TFs regulating their expression. I focused on two important medicinal plants, *Glycyrrhiza uralensis* (Fabaceae) and *Platycodon grandiflorus* (Campanulaceae), which belong to different plant families, and performed functional characterization of the P450s and TFs involved in triterpenoid biosynthesis in these plants. First, I mined the P450s involved in triterpenoid biosynthesis in these two different medicinal plants as described in Chapters 2 and 3, respectively. I identified one CYP716A subfamily enzyme involved in the biosynthesis of oleanolic acid and betulinic acid, and one CYP72A subfamily enzyme involved in the biosynthesis of soyasaponins in *G. uralensis* (Chapter 2). The full identification of the OSCs and P450s for the major *G. uralensis* triterpenoids is also described in this chapter. A comparative analysis of the triterpenoid profiles and gene expression in field-grown intact roots and tissue-cultured stolons of *G. uralensis* suggested that the different triterpenoid profiles could be explained by the differential regulation of biosynthesis gene expression in these two samples. Since most of the P450s involved in triterpenoid biosynthesis have been identified from Fabaceae plants (legumes), it is important to analyze the triterpenoid biosynthetic mechanisms in non-Fabaceae plants for comparison. To address this question, I mined the P450s for triterpenoid biosynthesis in *P. grandiflorus* (Chapter 3). I identified two CYP716A subfamily enzymes involved in triterpenoid biosynthesis in *P. grandiflorus*, but the predicted CYP72A subfamily enzymes, which have been identified only from legumes as triterpene oxidases, did not show catalytic activity on the β-amyrin skeleton. Finally, as described in Chapter 4, I sought to identify the TFs regulating these triterpenoid biosynthetic pathways in *G. uralensis* based on the transcriptional regulation suggested in Chapter 2. From three candidate bHLH TFs selected by two different approaches using model plants, I identified a bHLH TF regulating soyasaponin biosynthesis in *G. uralensis*. The phylogenetic relationships of the identified TF for triterpenoid biosynthesis and putative bHLH TFs in *P. grandiflorus* are discussed in Chapter 5.

Chapter 2

Identification of P450s involved in triterpenoid biosynthesis in *Glycyrrhiza uralensis*

2-1. Introduction

Glycyrrhiza uralensis is a *Glycyrrhiza* (licorice, 'kanzo' in Japanese) species belonging to the Fabaceae (Figure 2-1). Roots and stolons of *G. uralensis*, and its related species, *G. glabra*, and *G. inflata*, are widely used as medicines and natural sweeteners because they contain substantial amounts of glycyrrhizin, a sweet-tasting oleanane-type pentacyclic triterpenoid saponin with various pharmacological activities (Hayashi and Sudo, 2009). In addition to glycyrrhizin, licorice produces other pentacyclic triterpenoids, including soyasaponins, betulinic acid, and oleanolic acid. Several studies have examined the distribution of these triterpenoids in tissue cultures and in different organs in licorice (Hayashi *et al.*, 1988, 1990, 1993; Kojoma *et al.*, 2010). Glycyrrhizin is accumulated mainly in thickened roots and was not detected in leaves or stems of *G. glabra* plants (Hayashi *et al.*, 1993). Soyasaponins and betulinic acid are accumulated in callus and suspension cultures of *G. glabra* instead of glycyrrhizin (Hayashi *et al.*, 1988, 1990). Similarly, tissue-cultured stolons of *G. uralensis* accumulate higher levels of oleanolic acid and betulinic acid, while glycyrrhizin production is much lower than in intact stolons (Kojoma *et al.*, 2010). Although several attempts have been made to produce glycyrrhizin in licorice tissue cultures, the production of glycyrrhizin in these cultures remains challenging (Kojoma *et al.*, 2010).

Most of the enzymes involved in the production of major triterpenoids in licorice have been characterized (Figure 2-2). Glycyrrhizin and soyasaponins are derived from β-amyrin, which is biosynthesized from 2,3-oxidosqualene by β-amyrin synthase (bAS) (Hayashi *et al.*, 2001). In glycyrrhizin biosynthesis, CYP88D6 and CYP72A154 catalyze oxidation reactions at the C-11 and C-30 positions of β-amyrin, respectively, to produce glycyrrhetinic acid, an aglycone of glycyrrhizin (Seki *et al.*, 2008, 2011). Recently, a single UDP-dependent glycosyltransferase (UGT), *Gu*UGAT, was reported to transfer two glucuronic acid moieties to the C-3 positions of glycyrrhetinic acid to produce glycyrrhizin (Xu *et al.*, 2016). In soyasaponin biosynthesis, CYP93E3 has been characterized

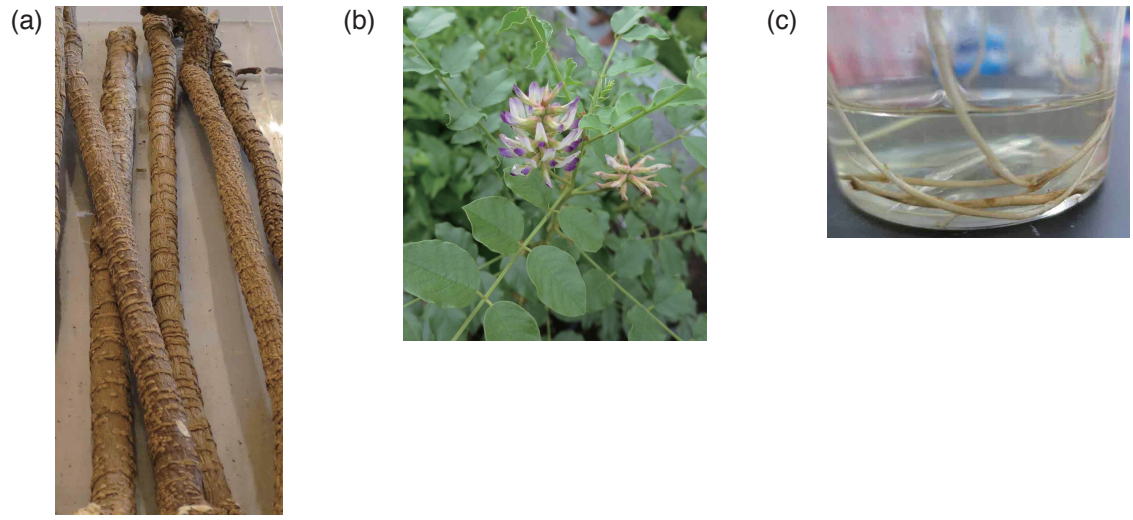


Figure 2-1. *Glycyrrhiza uralensis* plants and tissue cultures.
(a) Thickened roots exhibited at the Takeda Garden for Medicinal Plant Conservation, Kyoto, Japan (Takeda Pharmaceutical Company Limited).
(b) Aerial parts grown at the Takeda Garden for Medicinal Plant Conservation, Kyoto, in June 2016.
(c) Tissue-cultured stolons maintained in liquid culture.

as a β-amyrin C-24 hydroxylase (Seki *et al.*, 2008); however, β-amyrin C-22β hydroxylase, required for the biosynthesis of soyasapogenol B, a common aglycone of soyasaponin I and soyasaponin II, the two soyasaponins previously identified in cultured licorice cells (Hayashi *et al.*, 1990), remains unknown in licorice. Soyasapogenol B is expected to be further glycosylated by UGTs to produce soyasaponins; however, no UGTs involved in soyasaponin biosynthesis have been reported in licorice. Oleanolic acid and betulinic acid have a carboxyl group at the C-28 positions of β-amyrin and lupeol, respectively. Lupeol is a lupane-type triterpene skeleton produced by lupeol synthase (LUS) (Hayashi *et al.*, 2004). However, no enzymes catalyzing carboxylation at C-28 positions of these triterpene skeletons (triterpene C-28 oxidase) have been reported in licorice.

Of the triterpenoids, oleanolic acid and betulinic acid are widely distributed among plant species (Jäger *et al.*, 2009), and there are many reports on the cytochrome P450 monooxygenases (P450s) involved in the biosynthesis of these compounds. The first reported triterpene C-28 oxidase was CYP716A12, which was isolated from *Medicago truncatula* (barrel medic, Fabaceae); this enzyme catalyzes a three-step oxidation reaction at the C-28 position of β-amyrin to yield oleanolic acid (Carelli *et al.*, 2011; Fukushima *et al.*, 2011). CYP716A12 also accepts α-amyrin and lupeol as substrates to produce ursolic acid and betulinic acid, respectively (Fukushima *et al.*, 2011). Since the

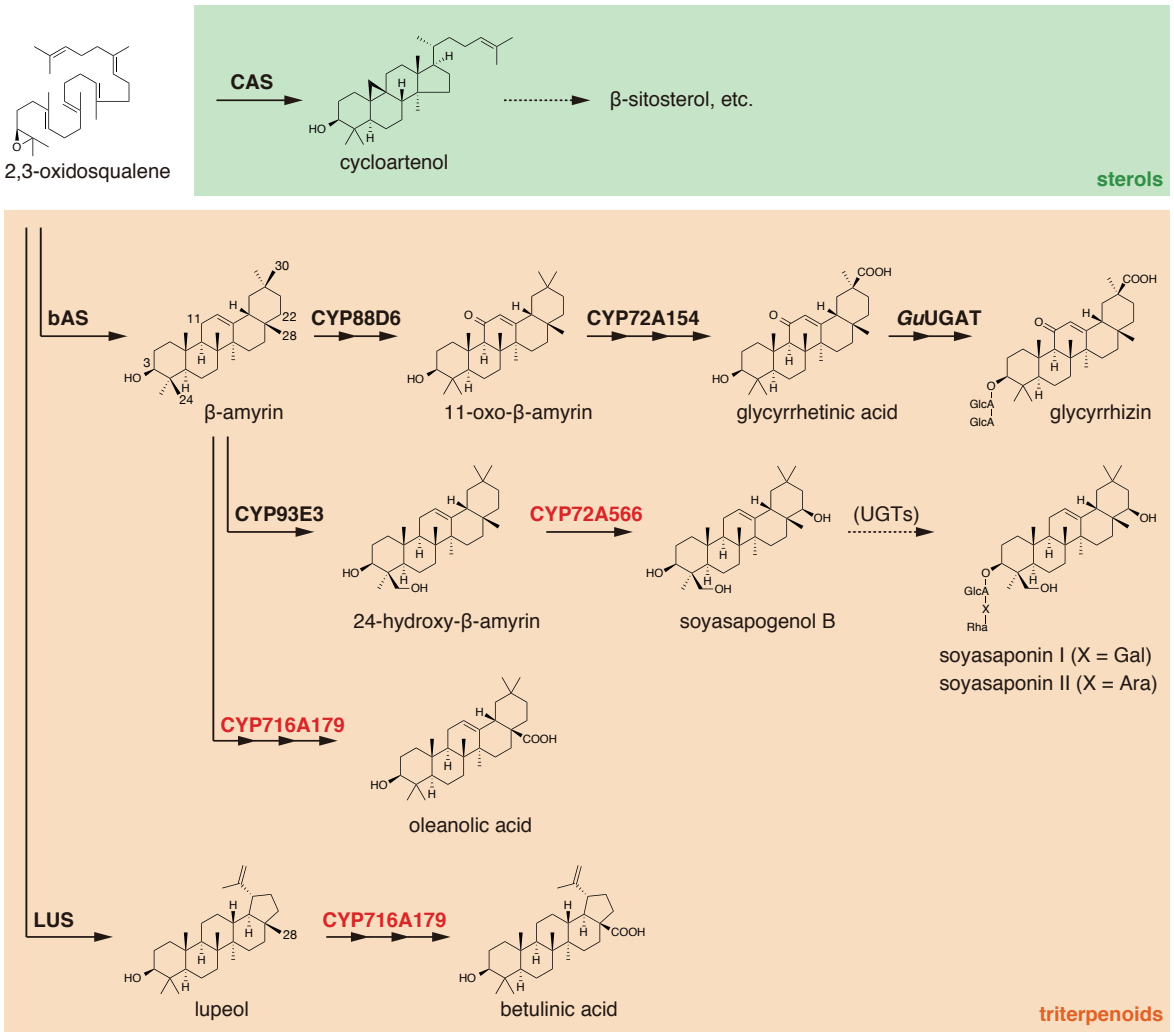


Figure 2-2. Proposed biosynthetic pathways of triterpenoids and sterols in licorice. P450 enzymes functionally characterized in this chapter are indicated in red. Arrows indicate a cyclization, single oxidation, or single glycosylation step. Numbers on β-amyrin and lupeol indicate carbon positions. Abbreviations: Ara, arabinose; bAS, β-amyrin synthase; CAS, cycloartenol synthase; Gal, galactose; GlcA, glucuronic acid; LUS, lupeol synthase; Rha, rhamnose.

discovery of CYP716A12, a number of CYP716A enzymes with similar catalytic activities have been reported in various plant species (Seki *et al.*, 2015). Although licorice and *M. truncatula* are members of the Fabaceae, no CYP716A subfamily P450 has been reported in licorice.

On the other hand, soyasaponins are widely distributed among legumes, including *Glycine max* (soybean), *M. truncatula*, and *G. glabra* (Hayashi *et al.*, 1990; Zhang and Popovich, 2009; Sawai and Saito, 2011). These are classified mainly into group A and B saponins, which possess soyasapogenol A and B as their aglycones, respectively (Rupasinghe *et al.*, 2003). In *G. glabra*, the formation of soyasaponin I and soyasaponin II has been reported (Hayashi *et al.*, 1990). Both are

group B saponins, which are glycosides of soyasapogenol B. In *M. truncatula*, co-expression of CYP93E2 and CYP72A61v2 produced soyasapogenol B in engineered yeast endogenously producing β-amyrin (Fukushima *et al.*, 2013). CYP72A155 isolated from *G. uralensis* has been shown to be phylogenetically close to CYP72A61v2; however, this enzyme did not show catalytic activity toward β-amyrin (Seki *et al.*, 2011).

In this chapter, I performed RNA sequencing (RNA-Seq) of tissue-cultured stolons of *G. uralensis*, which accumulates high levels of oleanolic acid and betulinic acid (Kojoma *et al.*, 2010), and probably soyasaponins rather than glycyrrhizin. From the obtained sequencing data, I identified a triterpene C-28 oxidase and β-amyrin C-22β hydroxylase in *G. uralensis*. I also performed a comparative analysis of the expression levels of triterpenoid biosynthetic genes and triterpenoid profiles between tissue-cultured stolons and intact roots of *G. uralensis*.

2-2. Materials and methods

2-2-1. Plant materials

Tissue-cultured stolons of *G. uralensis* (Hokkaido-iryodai strain) were maintained at 26°C with rotary shaking at 90 rpm in the dark in 100 mL of Murashige and Skoog (MS) medium (Duchefa Biochemie, Haarlem, the Netherlands) supplemented with 6% sucrose and 0.01 μM 1-naphthaleneacetic acid (NAA) as reported previously (Kojoma *et al.*, 2010). Intact roots and leaves were collected from *G. uralensis* plants (2-year-old) of a glycyrrhizin high-producing strain (308-19) harvested in June 2011 in Takeda Garden for Medicinal Plant Conservation, Kyoto, Japan (Ramilowski *et al.*, 2013).

2-2-2. Chemicals

β-Amyrin, erythrodiol, oleanolic acid, α-amyrin, uvaol, ursolic acid, lupeol, glycyrrhetinic acid (18β-glycyrrhetinic acid), echinocystic acid, and hederagenin were purchased from Extrasynthese (Genay, France). Betulin and NAA were purchased from Sigma-Aldrich (St. Louis, MO, USA). Betulinic acid and methyl-β-cyclodextrin were purchased from Tokyo Chemical Industry (Tokyo, Japan). Soyasapogenol B was purchased from Tokiwa Phytochemical (Chiba, Japan).

22α-hydroxy-α-amyrin was extracted and purified from transgenic yeast co-expressing *α-amyrin synthase* (*aAS*) and *CYP716A2* (Yasumoto *et al.*, 2016). Oleanolic aldehyde, ursolic aldehyde, betulinic aldehyde, sophoradiol, and 24-hydroxy-β-amyrin were kindly gifted by Dr. Kiyoshi Ohyama (Tokyo Institute of Technology, Tokyo, Japan).

2-2-3. RNA extraction

Total RNA was extracted using PureLink Plant RNA Reagent (Thermo Fisher Scientific, Waltham, MA, USA) from 1 g of frozen plant tissues and treated with recombinant DNase I (RNase-free) (Takara Bio, Shiga, Japan), then purified using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) following the RNA clean-up protocol.

2-2-4. RNA-Seq of tissue-cultured stolons of *G. uralensis*

A 10-μg aliquot of total RNA was used to construct a cDNA library using an Illumina TruSeq RNA Sample Prep Kit v2 (Illumina, San Diego, CA, USA) according to the manufacturer's protocol. The resulting cDNA library was sequenced using HiSeq 2000 (Illumina) with 100 bp paired-end reads in high-output mode. Total reads were assembled using Trinity (Grabherr *et al.*, 2011) after adaptor sequences and low-quality reads were removed by Trimmomatic ver. 0.33 (Bolger *et al.*, 2014). The raw RNA-Seq reads obtained in this chapter were deposited in the DDBJ Sequence Read Archive (DRA) under the accession number DRA004898.

2-2-5. Cloning of *CYP716A179* and *CYP72A566*

First-strand cDNA was synthesized from 1 μg of total RNA prepared from the tissue-cultured stolons of *G. uralensis* using the SMARTer RACE cDNA Amplification Kit (Clontech/Takara Bio). The full-length coding sequence (CDS) of *CYP716A179* was amplified from the obtained cDNA library with primers 1 and 2 (Table 2-1) and cloned into pENTR/D-TOPO (Thermo Fisher Scientific) to make an entry clone. cDNA for *CYP716A179* was transferred into a Gateway-compatible version of the pELC vector (Seki *et al.*, 2011) using the Gateway LR Clonase II Enzyme mix (Thermo Fisher Scientific) to generate pELC-CYP716A179, a construct for galactose-inducible dual expression of

Table 2-1. Primers used in this chapter

Primer no.	Sequence (5' to 3')	Comment
1	CACCATGGAGCATTTCTACATGTCC	Cloning of *CYP716A179*
2	TCAGGTATCGTGAGGATAAAGG	Cloning of *CYP716A179*
3	CACCATGGTAGAATTGTTAGGAATAACA	Cloning of *CYP72A566*
4	CTACAGTTTATGTAAAATGAGATGAGC	Cloning of *CYP72A566*
5	TCTTCGCAAACTGGCAGTGA	qPCR for *β-tubulin*
6	CGAGATGTGAGTGGGGCAAA	qPCR for *β-tubulin*
7	GGTGGTTTATCAGCGTGGGA	qPCR for *bAS*
8	TGCTCAACTACAATGTCCGCA	qPCR for *bAS*
9	CCGGCTGAGACTTTTGGTGA	qPCR for *CAS*
10	TCTCGACGATGCCCAGGATA	qPCR for *CAS*
11	GAAAGCAATACCCACAGCACAG	qPCR for *LUS*
12	GCAAATTCCCCAACACCCATAC	qPCR for *LUS*
13	GACGCCTCATTACTTCCCCAA	qPCR for *CYP88D6*
14	TTCAAGAGCTCAACGGGGTG	qPCR for *CYP88D6*
15	ATGGCGACCCTTACAAGCTC	qPCR for *CYP72A154*
16	AGATTCGTGGTGCAGCATCA	qPCR for *CYP72A154*
17	GCCGCCGATTACTTCTTCATTC	qPCR for *CYP93E3*
18	CGCGGATATTGACAAAGTGCTC	qPCR for *CYP93E3*
19	AGCGGTTCAAGTGGGAGATG	qPCR for *CYP716A179*
20	ATCGGGAGGTCATTTGCAGG	qPCR for *CYP716A179*
21	CAGGGGAGCCTAGTAACAATGAC	qPCR for *CYP72A566*
22	GCCCTGCCAAGTAAAATAGCTTC	qPCR for *CYP72A566*

The underlined sequences were added to facilitate unidirectional cloning of the product into pENTR/D-TOPO (Thermo Fisher Scientific).

cytochrome P450 reductase (*Lotus japonicus CPR1*) and *CYP716A179*. The pELC vector was generated by the cloning of full-length CDS of *LjCPR1* into the NotI and PacI sites of pESC-LEU (Agilent Technologies, Santa Clara, CA, USA) for galactose-inducible expression of *LjCPR1* under the control of the *GAL10* promoter (Seki *et al.*, 2008). The nucleotide sequence of CYP716A179 has been submitted to the DDBJ under the accession number LC157867.

Similarly, CDS of *CYP72A566* was amplified with primers 3 and 4 (Table 2-1) from the same cDNA library and cloned into pENTR/D-TOPO (Thermo Fisher Scientific) to make an entry clone. cDNA for *CYP72A566* was transferred into pYES-DEST52 (Thermo Fisher Scientific) and a Gateway compatible version of pESC-HIS (Agilent Technologies) using the Gateway LR Clonase II Enzyme mix (Thermo Fisher Scientific) to generate galactose-inducible expression of *CYP72A566* (pYES-DEST52-CYP72A566 and pESC-HIS-CYP72A566). The nucleotide sequence of CYP72A566 has been submitted to the DDBJ under the accession number LC318134.

2-2-6. *In vivo* enzyme assay of CYP716A179

Saccharomyces cerevisiae INV*Sc*1 (*MAT*a *his3Δ1 leu2 trp1-289 ura3-52/MAT*α *his3Δ1 leu2 trp1-289 ura3-52*; Thermo Fisher Scientific) harboring pYES3-ADH-aAS (constitutive expression of *Olea europaea aAS* under the control of the *ADH1* promoter), pYES3-ADH-OSC1 (constitutive expression

Table 2-2. Yeast strains generated in this chapter

Strain no.	Genotype
1	INV*Sc*1; pYES3[*ADH1*/*bAS*]; pESC-LEU[*GAL10*/*CPR1*]
2	INV*Sc*1; pYES3[*ADH1*/*bAS*]; pESC-LEU[*GAL10*/*CPR1*, *GAL1*/*CYP716A179*]
3	INV*Sc*1; pYES3[*ADH1*/*aAS*]; pESC-LEU[*GAL10*/*CPR1*]
4	INV*Sc*1; pYES3[*ADH1*/*aAS*]; pESC-LEU[*GAL10*/*CPR1*, *GAL1*/*CYP716A179*]
5	INV*Sc*1; pYES3[*ADH1*/*LUS*]; pESC-LEU[*GAL10*/*CPR1*]
6	INV*Sc*1; pYES3[*ADH1*/*LUS*]; pESC-LEU[*GAL10*/*CPR1*, *GAL1*/*CYP716A179*]
7	INV*Sc*1; pYES3[*ADH1*/*bAS*]; pESC-LEU[*GAL10*/*CPR1*, *GAL1*/*CYP93E3*]; pYES-DEST52; pESC-HIS
8	INV*Sc*1; pYES3[*ADH1*/*bAS*]; pESC-LEU[*GAL10*/*CPR1*, *GAL1*/*CYP93E3*]; pYES-DEST52[*GAL1*/*CYP72A566*]; pESC-HIS[*GAL1*/*CYP72A566*]

Selection markers in yeast are as follows: pYES3, *TRP1*; pESC-LEU, *LEU2*; pYES-DEST52, *URA3*; pESC-HIS, *HIS3*.

of *L. japonicus bAS* under the control of the *ADH1* promoter), or pYES3-ADH-LUS (constitutive expression of *G. uralensis LUS* under the control of the *ADH1* promoter) (Fukushima *et al.*, 2011) was transformed with pELC-CYP716A179 or pELC using the Frozen-EZ Yeast Transformation II Kit (Zymo Research, Irvine, CA, USA). A glycerol stock of each yeast strain was inoculated into 2 mL of appropriate synthetic defined (SD) medium containing 2% glucose, and cultured overnight at 30°C, shaking at 200 rpm. Each starter culture was transferred into 10 mL of identical medium, and cultured for an additional 24 h at 30°C, shaking at 200 rpm. The yeast cells were collected and resuspended in 10 mL of SD medium containing 2% galactose. These were cultured at 30°C for a further 2 days at 200 rpm. The yeast culture was extracted three times with ethyl acetate, concentrated *in vacuo*, and resuspended in 500 μL of methanol/chloroform (1:1, v/v). Then, 100 μL of the solution was evaporated and trimethylsilylated with 100 μL of *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide (Sigma-Aldrich) at 80°C for 20 min before gas chromatography-mass spectrometry (GC-MS) analysis. Hederagenin was added as an internal standard prior to extraction. The engineered yeast strains made for this chapter are listed (Table 2-2). Outline of *in vivo* enzyme assay of P450s is shown in Figure 2-3.

2-2-7. *In vivo* enzyme assay of CYP72A566

S. cerevisiae INV*Sc*1 pre-transformed with pYES3-ADH-OSC1 (Fukushima *et al.*, 2011) for constitutive expression of *bAS* and pELC-CYP93E3 (Seki *et al.*, 2008) for galactose-inducible dual expression of *L. japonicus CPR1* and *CYP93E3* was transformed with pYES-DEST52-CYP72A566 and pESC-HIS-CYP72A566 using the Frozen-EZ Yeast Transformation II Kit (Zymo Research). The

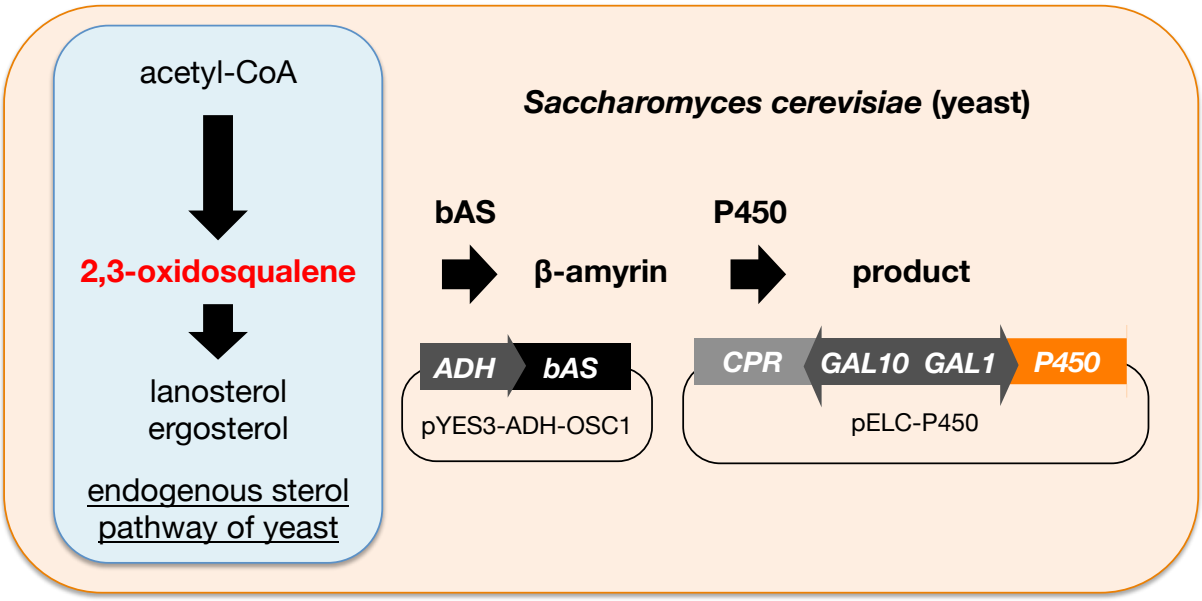


Figure 2-3. Outline of the *in vivo* enzyme assay of P450s using engineered yeast strains. Yeasts have an endogenous sterol pathway for biosynthesis of ergosterol via lanosterol for membrane structures. 2,3-Oxidosqualene, an intermediate in the endogenous sterol pathway, is converted into β-amyrin by introduced β-amyrin synthase (bAS). By expressing *cytochrome P450 reductase* (*CPR*) and P450 from galactose-inducible promoters, the enzymatic activities of the P450 can be examined.

recombinant yeast strains were inoculated into 2 mL of appropriate SD medium containing 2% glucose, and cultured overnight at 30°C, shaking at 200 rpm. Each starter culture was transferred into 10 mL of identical medium, and cultured for an additional 24 h at 30°C, shaking at 200 rpm. The yeast cells were collected and resuspended in 10 mL of SD medium containing 2% galactose and 25 mM methyl-β-cyclodextrin. These were cultured at 30°C for a further 5 days at 200 rpm. The yeast culture was extracted three times with ethyl acetate, dried *in vacuo*, and resuspended in 500 µL of methanol/chloroform (1:1, v/v). For derivatization, 100 µL of the solution was evaporated and trimethylsilylated with 100 µL of *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide (Sigma-Aldrich) at 80°C for 30 min before GC-MS analysis.

2-2-8. Analysis of sapogenins in plant tissues

Freeze-dried powder of plant tissues (40 mg) were mixed with 200 µL of internal standard (0.1 mg/mL of echinocystic acid) and extracted twice with methanol/chloroform (1:1, v/v). After the solvents were removed, 1 mL of methanol and 1 mL of 4 M hydrogen chloride were added to the residue and hydrolyzed at 80°C for 1 h. The hydrolyzed products were extracted twice with hexane/ethyl acetate

(1:1, v/v), dried *in vacuo*, and resuspended in 400 μL of methanol/chloroform (1:1, v/v). For derivatization, 100 μL of the solution was evaporated, resuspended in 50 μL of *N*,*N*-dimethyl formamide, and trimethylsilylated with 50 μL of *N*,*O*-bis(trimethylsilyl)trifluoroacetamide + trimethylchlorosilane, 99:1 (Sigma-Aldrich) at 80°C for 30 min before GC-MS analysis.

2-2-9. GC-MS analysis

GC-MS analysis was performed using a 5977A mass selective detector (MSD) (Agilent Technologies) coupled with a 7890B gas chromatograph system (Agilent Technologies) and DB-1ms (30 m × 0.25 mm internal diameter, 0.25 μm film thickness; Agilent Technologies) capillary column. The injection component and the MSD transfer line were set at 250°C and the oven temperature was programmed as follows: 80°C for 1 min, followed by a rise to 300°C at a rate of 20°C/min, and a hold at 300°C for 15 min (*in vivo* enzyme assay of CYP716A179) or 28 min (*in vivo* enzyme assay of CYP72A566 and analysis of sapogenins in plant tissues). The carrier gas was He and the flow rate was 1 mL/min. Mass spectra were recorded in the range of 50–750 *m*/*z* (*in vivo* enzyme assay of CYP716A179) or 50–850 *m*/*z* (*in vivo* enzyme assay of CYP72A566 and analysis of sapogenins in plant tissues). Peaks were identified by comparing their retention times and mass spectra with those of authentic standards.

2-2-10. Quantitative real-time PCR

First-strand cDNA was synthesized using the PrimeScript RT Master Mix (Perfect Real Time; Takara Bio) from 2.5 μg of total RNA in a 50-μL reaction. Then, 1× FastStart Essential DNA Green Master (Roche, Basel, Switzerland), 500 nM of primers, and 0.5 μL of cDNA were mixed together and the reaction volume was brought to 10 μL with PCR-grade water. Reactions were performed using a LightCycler Nano (Roche) with the following conditions: 95°C for 10 min, 45 cycles of 95°C for 10 sec, 60°C for 10 sec, and 72°C for 15 sec. Data were analyzed using LightCycler Nano Software ver. 1.1.0 (Roche). Relative transcript levels of each target gene were calculated using *β-tubulin* (Seki *et al.*, 2008) (GenBank accession number LC318135) as a reference gene. The amplification of each sample was performed three times, using primers 5–22 (Table 2-1) designed using the Primer3 website (http://bioinfo.ut.ee/primer3-0.4.0/) (Koressaar and Remm, 2007; Untergasser *et al.*, 2012).

2-2-11. Phylogenetic analysis

Full-length amino acid sequences were collected from GenBank (http://www.ncbi.nlm.nih.gov/genbank/) or DDBJ (http://www.ddbj.nig.ac.jp/). A phylogenetic tree was generated using the neighbor-joining method with 1,000 bootstrap replicates in the ClustalX 2.1 software (Larkin *et al.*, 2007) and visualized with FigTree ver. 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/) software.

2-3. Results

2-3-1. Triterpenoid profiles of intact roots and tissue-cultured stolons

I compared the triterpenoid profiles of intact roots of field-grown plants and tissue-cultured stolons of *G. uralensis*. Quantification of sapogenins in these two plant tissues was performed after removal of sugar moieties by acid hydrolysis (Table 2-3 and Figure 2-4). The content of glycyrrhetinic acid, an aglycone of glycyrrhizin, in intact roots was almost 1,000 times higher than in tissue-cultured stolons. In contrast, oleanolic acid and betulinic acid were detected only in tissue-cultured stolons. I also quantified the accumulation of soyasapogenol B, a common aglycone of soyasaponin I and soyasaponin II. This analysis revealed that accumulation of soyasapogenol B in tissue-cultured stolons was more than 20 times higher than in intact roots.

2-3-2. Identification and cloning of *CYP716A179*

To identify P450s involved in the biosynthesis of oleanolic acid and betulinic acid in *G. uralensis*, I searched for a putative ortholog of *CYP716A12* in the previous transcriptome analysis of intact roots and leaves (Ramilowski *et al.*, 2013) or intact stolons (Sudo *et al.*, 2009); however, no candidate contigs or unigenes were found in either transcriptome data set. Because oleanolic acid and betulinic acid were only detected in tissue-cultured stolons (Table 2-3), I expected that a putative ortholog of *CYP716A12* would be expressed in tissue-cultured stolons. Hence, total RNA from tissue-cultured stolons of *G. uralensis* was isolated and RNA-Seq analysis was performed. The obtained total 28,925,182 reads were assembled into 71,673 contigs (Table 2-4). The amino acid sequence of CYP716A12 was used as a query and a contig encoding protein with 86% amino acid identity to

Table 2-3. Quantification of sapogenins from intact roots and tissue-cultured stolons of *G. uralensis* by GC-MS

Sapogenin	Content of sapogenin (μg/g of dry weight)	
	Intact roots	Tissue-cultured stolons
β-amyrin	nd	37.09 ± 0.57
α-amyrin	nd	nd
lupeol	15.59 ± 0.50	nd
oleanolic acid	nd	18.08 ± 4.56
ursolic acid	nd	nd
betulinic acid	nd	30.95 ± 0.45
soyasapogenol B	9.77 ± 0.99	207.4 ± 4.5
glycyrrhetinic acid	27260 ± 1120	29.19 ± 10.73

Values are mean ± standard error (SE) of three technical replicates calculated from the peak areas of each compound and the internal standard (echinocystic acid). nd, not detected.

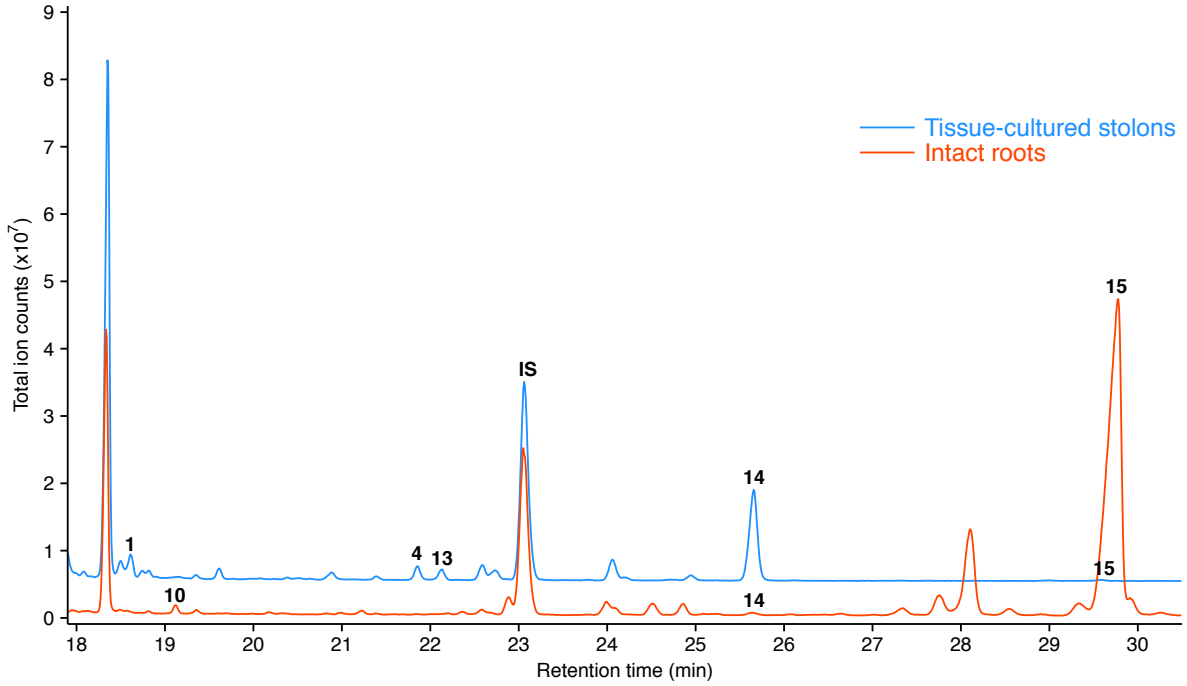


Figure 2-4. GC-MS analysis of sapogenins from tissue-cultured stolons and intact roots of *G. uralensis*. β-amyrin (**1**), oleanolic acid (**4**), betulinic acid (**13**), soyasapogenol B (**14**), and glycyrrhetinic acid (**15**) were identified in tissue-cultured stolons, whereas lupeol (**10**), soyasapogenol B (**14**), and glycyrrhetinic acid (**15**) were identified in intact roots. Echinocystic acid was used as an internal standard (IS).

Table 2-4. Summary of the RNA-Seq analysis of tissue-cultured stolons of *G. uralensis*

Total read counts	28,925,182
Number of contigs	71,673
N50 of contigs (bp)	1,639
Average length of contigs (bp)	1,001
Minimum length of contigs (bp)	224
Maximum length of contigs (bp)	13,598

CYP716A12 was found in the RNA-Seq data from tissue-cultured stolons of *G. uralensis* using blast (Altschul *et al.*, 1997). The full-length cDNA of the corresponding contig was isolated and designated as CYP716A179 according to P450 nomenclature. Alignments of amino acid sequences of CYP716A12 and CYP716A179 are shown in Figure 2-5. The phylogenetic tree of P450s involved in

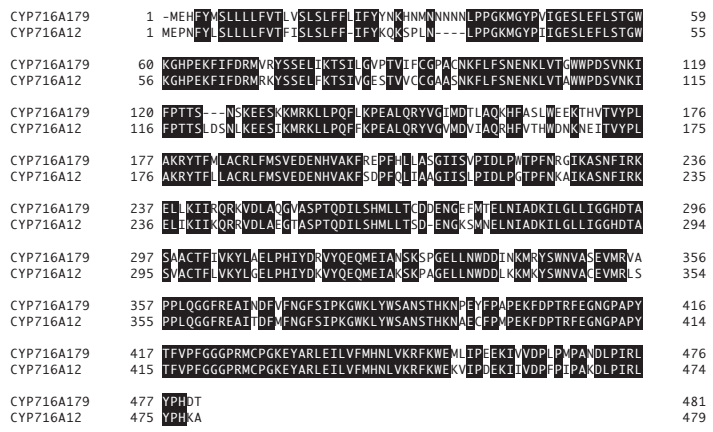


Figure 2-5. Multiple alignments of CYP716A179 and CYP716A12. Multiple alignments were generated using GENETYX-MAC ver. 18.0.3 software. Conserved residues are highlighted with a black background.

triterpenoid biosynthesis in legumes (listed in Table 2-5) indicates that CYP716A179 is an ortholog of CYP716A12 (Figure 2-6).

2-3-3. Enzyme assay of CYP716A179 in engineered yeast

To elucidate whether CYP716A179 catalyzes oxidation at the C-28 positions of β-amyrin, α-amyrin, and lupeol as reported for CYP716A12 (Fukushima *et al.*, 2011), CYP716A179 was expressed together with *CPR* in engineered yeast strains harboring *bAS*, *aAS*, or *LUS* that endogenously produce β-amyrin, α-amyrin, or lupeol, respectively. Following the culture of each yeast strain, their *in vivo* metabolites were extracted and analyzed by GC-MS.

In *bAS*/*CPR*/*CYP716A179*-expressing yeast (strain 2, Table 2-2), erythrodiol (**2**), oleanolic aldehyde (**3**), and oleanolic acid (**4**) were detected (Figure 2-7a; mass spectra in Figure 2-8). In *aAS*/*CPR*/*CYP716A179*-expressing yeast (strain 4, Table 2-2), uvaol (**6**), ursolic aldehyde (**7**), and urosolic acid (**8**) were detected (Figure 2-7b). In addition to these metabolites oxidized at the C-28 position of α-amyrin, a trace amount of 22α-hydroxy-α-amyrin (**9**) was detected. 22α-hydroxy-α-amyrin (**9**) is a major product of CYP716A2 from *Arabidopsis thaliana* when heterologously expressed in α-amyrin-producing yeast (Yasumoto *et al.*, 2016). Note that since aAS is a multiproduct OSC that produces β-amyrin as well as α-amyrin, β-amyrin-derived metabolites, erythrodiol (**2**), oleanolic aldehyde (**3**), and oleanolic acid (**4**), were also detected in strain 4. In *LUS*/*CPR*/*CYP716A179*-expressing yeast (strain 6, Table 2-2), betulin (**11**), betulinic aldehyde (**12**), and

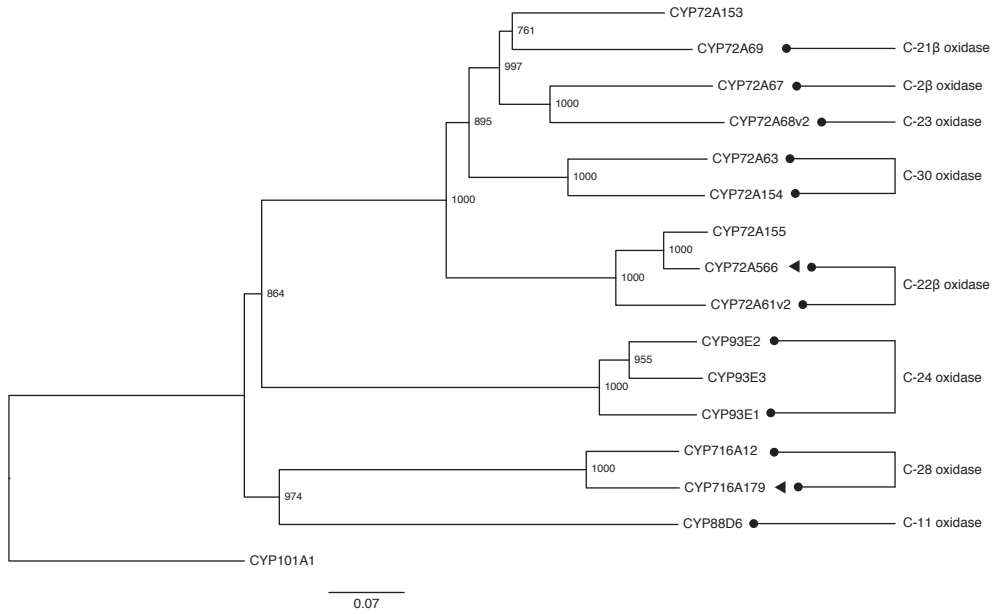


Figure 2-6. Phylogenetic tree of P450s related to oxidation of the β-amyrin skeleton in legumes. P450s characterized in this chapter are indicated with filled arrowheads. Reported functions against the β-amyrin skeleton are indicated at the right of the P450 names. Numbers indicate bootstrap values for 1,000 replicates. The scale bar shows the amino acid substitution ratio. GenBank protein accession numbers and species of previously characterized P450s in this tree are shown in Table 2-5. *Pseudomonas putida* CYP101A1 (GenBank protein accession number AAA25760) was used as an outgroup.

Table 2-5. A list of previously identified P450s included in the phylogenetic tree in Figure 2-6

P450 name	Species	GenBank protein accession no.	Reference
CYP72A61v2	*Medicago truncatula*	BAL45199	Fukushima *et al.*, 2013
CYP72A63	*Medicago truncatula*	BAL45200	Seki *et al.*, 2011
CYP72A67	*Medicago truncatula*	ABC59075	Biazzi *et al.*, 2015
CYP72A68v2	*Medicago truncatula*	BAL45204	Fukushima *et al.*, 2013
CYP72A153	*Glycyrrhiza uralensis*	BAL45206	Seki *et al.*, 2011
CYP72A154	*Glycyrrhiza uralensis*	BAL45207	Seki *et al.*, 2011
CYP72A155	*Glycyrrhiza uralensis*	BAL45208	Seki *et al.*, 2011
CYP72A69	*Glycine max* cv. Enrei	BAW35009	Yano *et al.*, 2017
CYP88D6	*Glycyrrhiza uralensis*	BAG68929	Seki *et al.*, 2008
CYP93E1	*Glycine max*	BAE94181	Shibuya *et al.*, 2006
CYP93E2	*Medicago truncatula*	ABC59085	Fukushima *et al.*, 2013
CYP93E3	*Glycyrrhiza uralensis*	BAG68930	Seki *et al.*, 2008
CYP716A12	*Medicago truncatula*	ABC59076	Carelli *et al.*, 2011; Fukushima *et al.*, 2011

betulinic acid (**13**) were detected (Figure 2-7c). These results were almost identical to the enzymatic activities of CYP716A12 (Fukushima *et al.*, 2011), except for the production of 22α-hydroxy-α-amyrin (**9**) in strain 4. The catalytic activities of CYP716A179 confirmed in this chapter are summarized in Figure 2-9.

2-3-4. Functional characterization of CYP72A566

CYP72A155 isolated from *G. uralensis* was shown to be a putative ortholog of CYP72A61v2, a

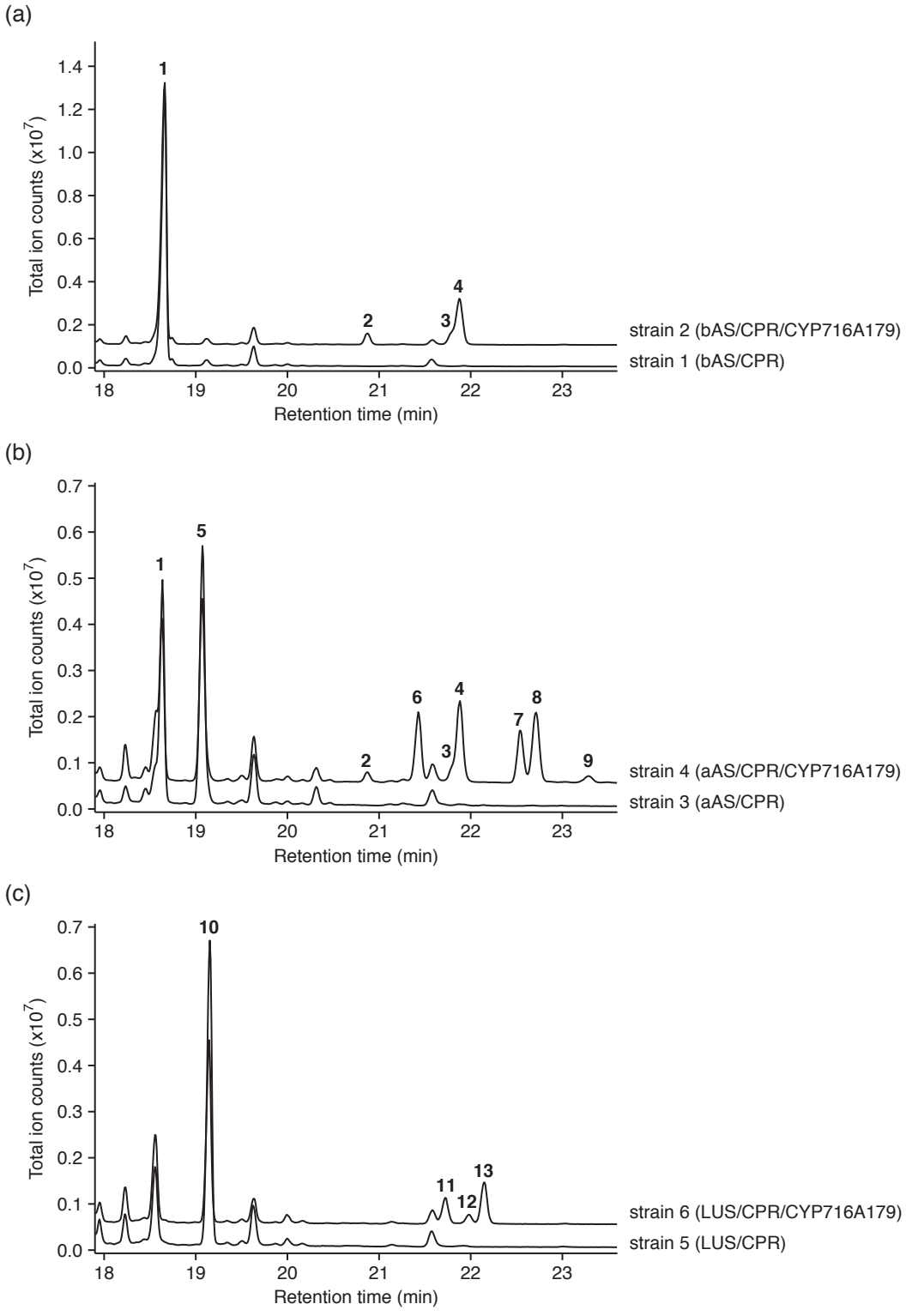


Figure 2-7. GC-MS analysis of products isolated from engineered yeast strains for *in vivo* enzyme assay of CYP716A179. *CYP716A179* was expressed in (a) β-amyrin-, (b) α-amyrin-, and (c) lupeol-producing yeasts. Numbers indicated in the chromatograms correspond to the compounds shown in Figure 2-9. Mass spectra of identified peaks are shown in Figure 2-8.

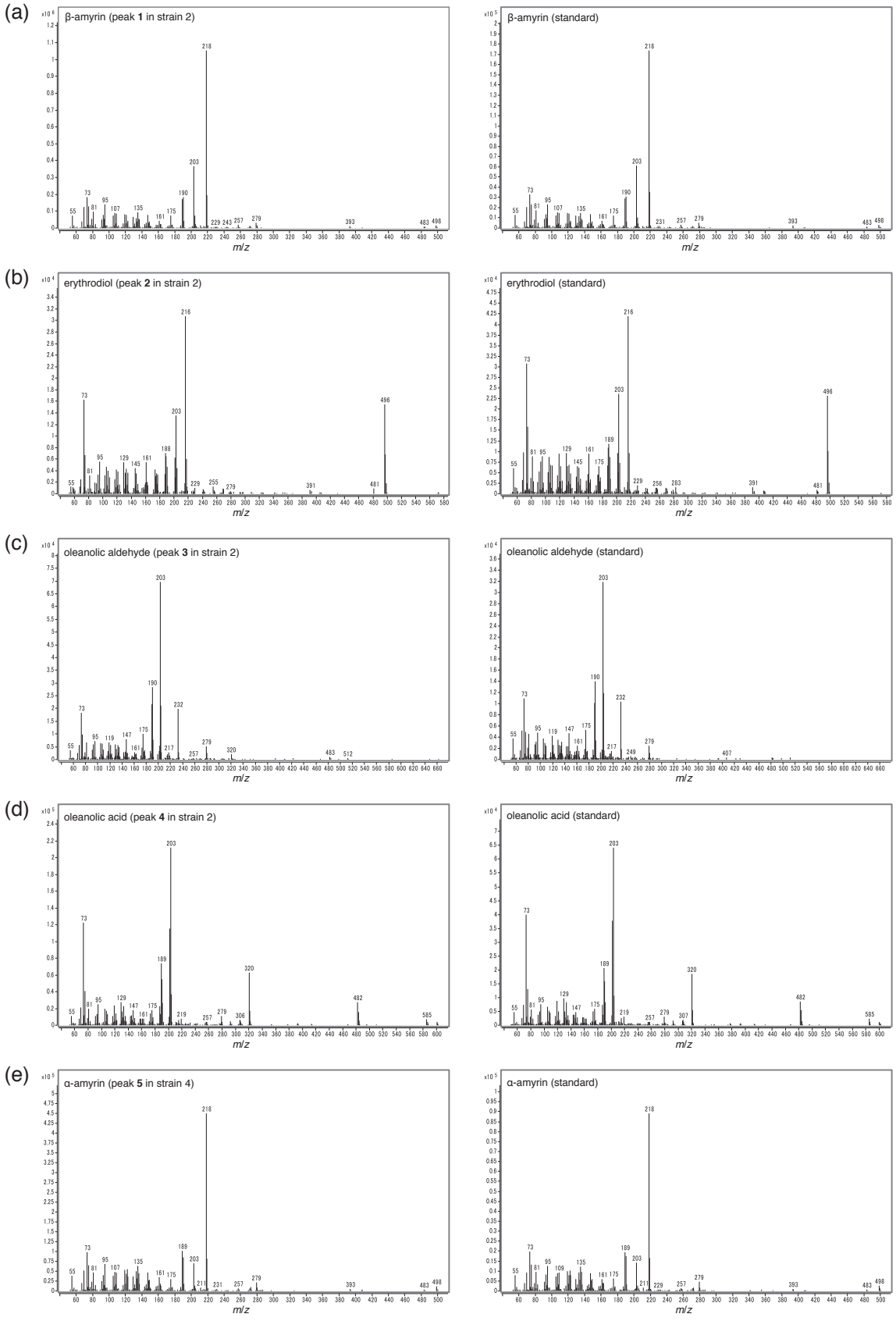


Figure 2-8. Mass spectra of peaks identified in Figures 2-7a–c and those of authentic standards: (a) β-amyrin, (b) erythrodiol, (c) oleanolic aldehyde, (d) oleanolic acid, (e) α-amyrin, (f) uvaol, (g) ursolic aldehyde, (h) ursolic acid, (i) 22α-hydroxy-α-amyrin, (j) lupeol, (k) betulin, (l) betulinic aldehyde, and (m) betulinic acid.

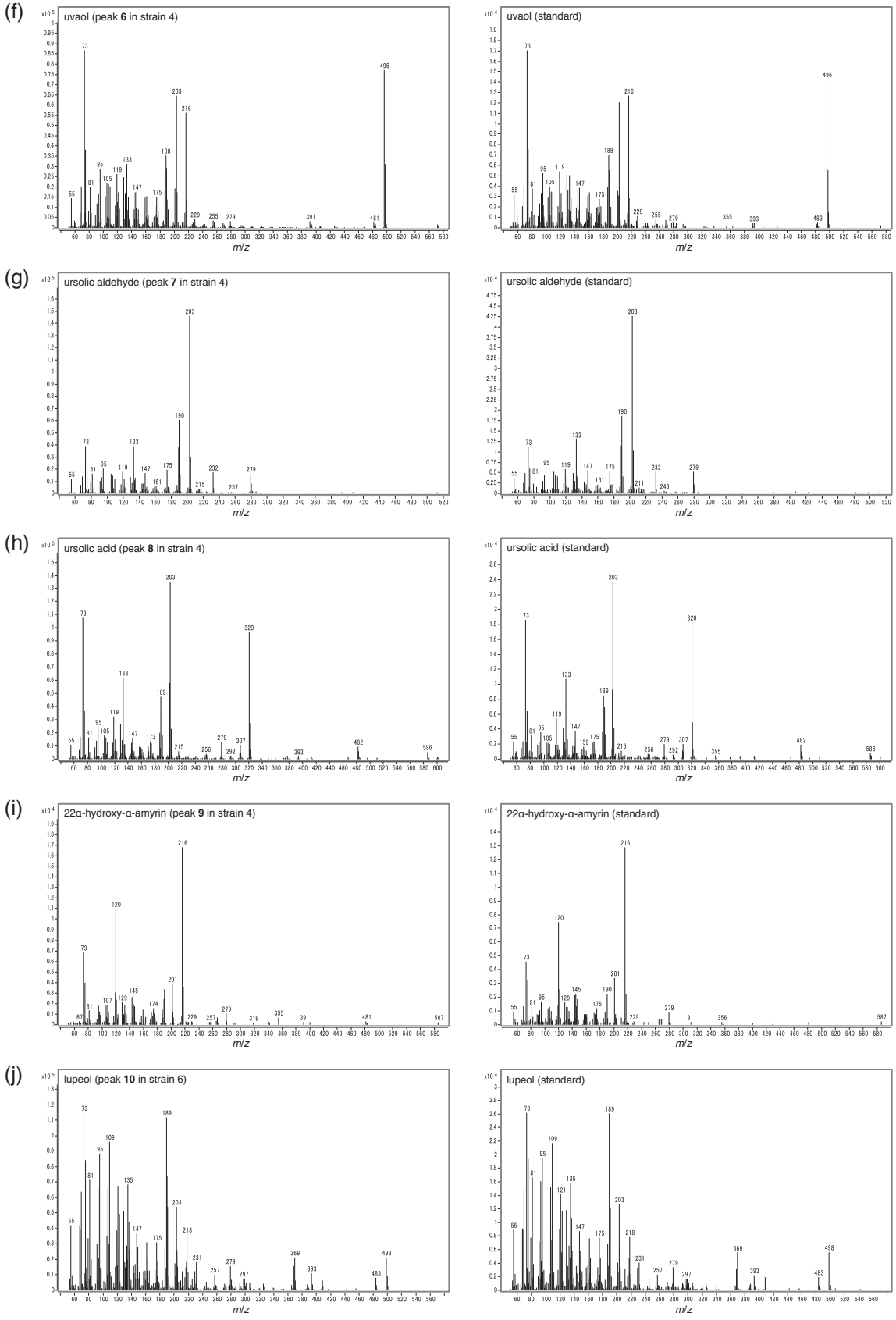


Figure 2-8. (continued)

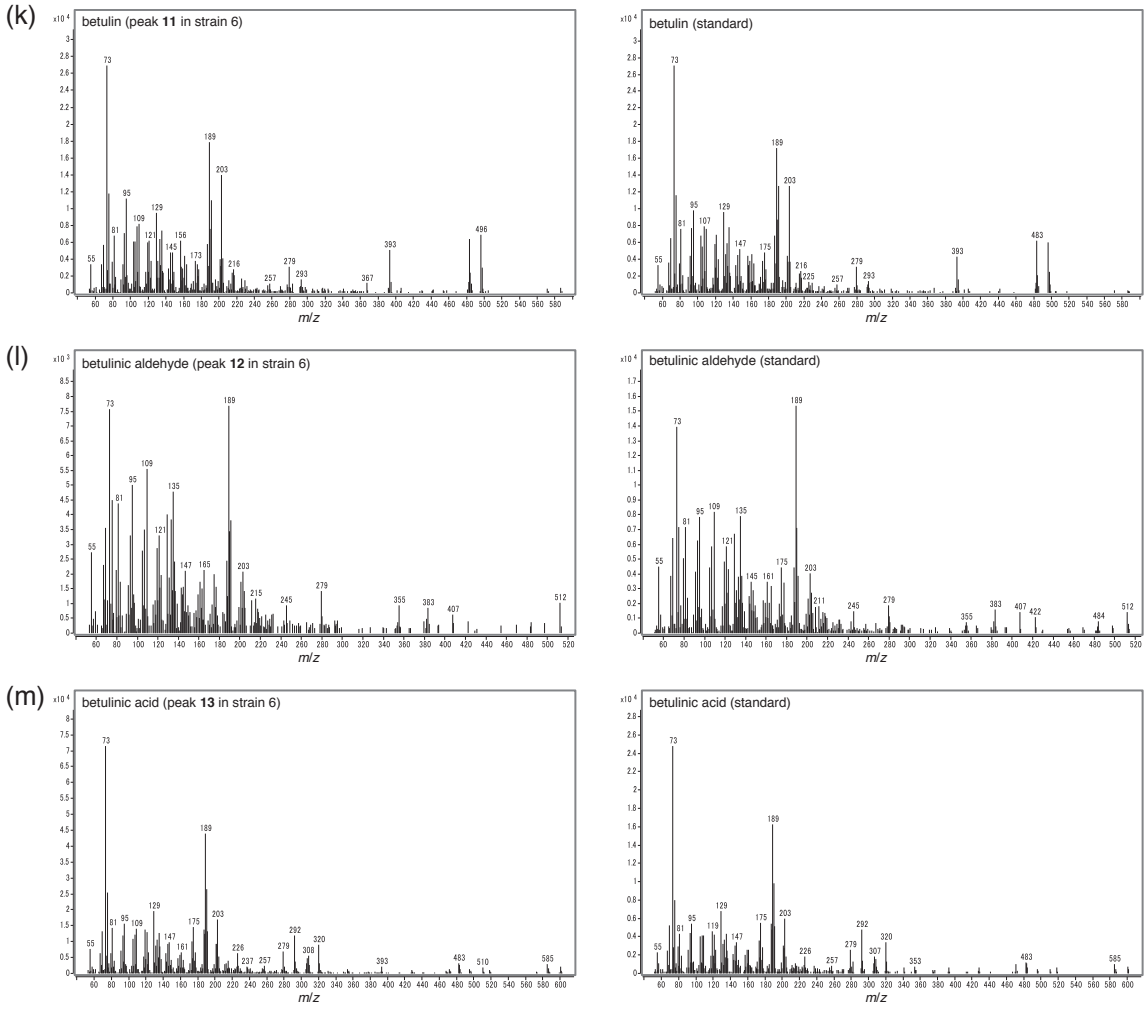


Figure 2-8. (continued)

β-amyrin C-22β hydroxylase in *M. truncatula*; however, this enzyme did not oxidize β-amyrin in the previous study (Seki *et al.*, 2011). When I searched for a contig corresponding to *CYP72A155* in the RNA-Seq data obtained in this chapter, the best hit exhibited many mismatches with *CYP72A155* (96% nucleotide identity). Therefore, I suspected that this CYP72A155-like contig was the putative β-amyrin C-22β hydroxylase gene in *G. uralensis*. To test this hypothesis, I first cloned this contig (designated CYP72A566 by the P450 nomenclature committee) and compared amino acid sequences with CYP72A61v2 and CYP72A155 (Figure 2-10; phylogenetic relationships in Figure 2-6). The amino acid sequence identity between CYP72A566 and CYP72A61v2 was 85%, slightly higher than the identity between CYP72A155 and CYP72A61v2 (82% amino acid sequence identity). Next, the enzymatic activities of CYP72A566 against β-amyrin and 24-hydroxy-β-amyrin were tested in an

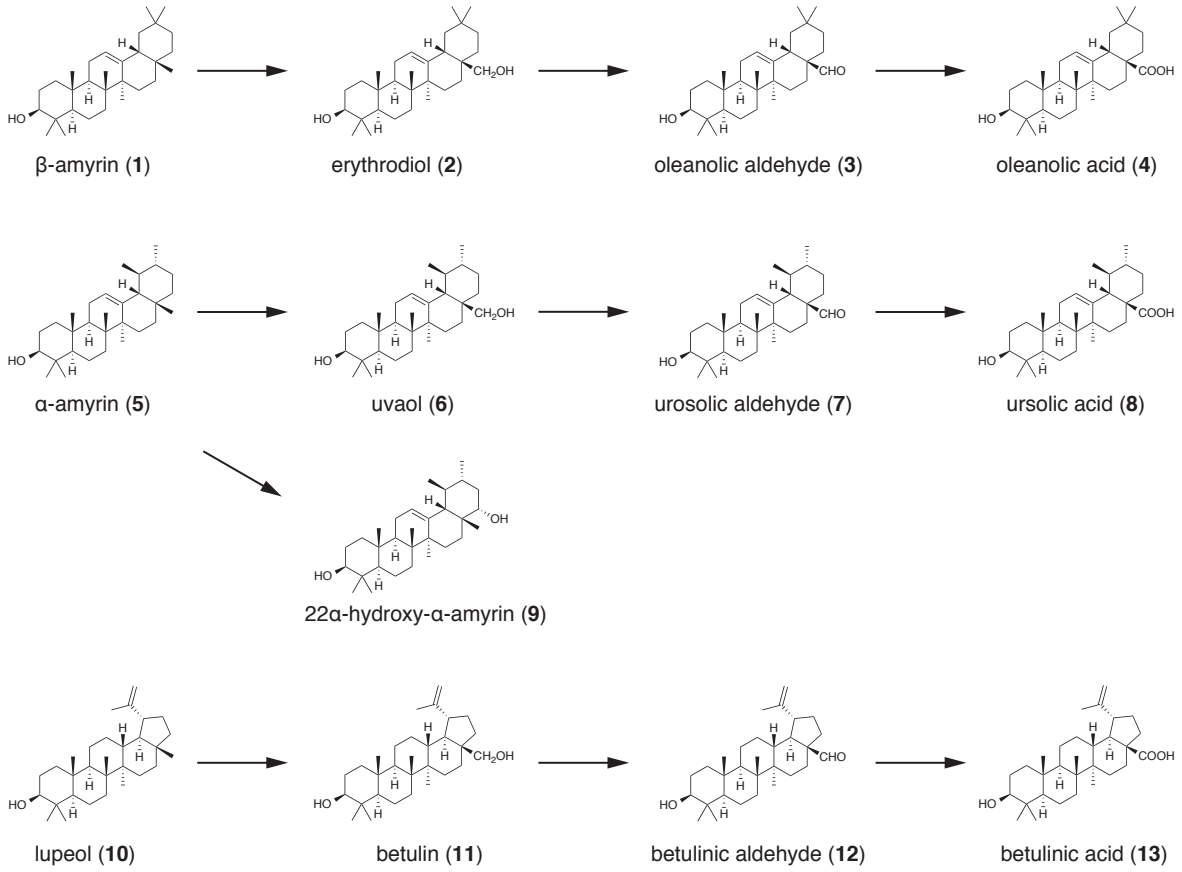


Figure 2-9. Reactions catalyzed by CYP716A179 in the heterologous expression of *CYP716A179* in engineered yeast strains. Arrows represents single oxidation reactions catalyzed by CYP716A179.

```
CYP72A566       1 MVELLGITSSMNLAPTTIAISTVIGSVLIWWVWSALNVVWLRPKSIERRLREQGLQGNSY   60
CYP72A61v2      1 MENLGTLLSSLNLQPTKVAITTVISSVLIWWIWSALDWIWFTPKRIEKRLKQQGLKGNSY   60
CYP72A155       1 MGESLGITSSMNLAPTTIAISTVIGSVLMWWVWSALNVVWLRPKSIERRLREQGLQGNSY   60

CYP72A566      61 RPFVGDIRDMVKMIKEAKAKFMDPYSNNIAPRVLPYVVHTVAKYGKNSFMWLGPKPRVFI  120
CYP72A61v2     61 RLMVGDIRDMVKMIKEAKSKPMDPYSNDIAPRVLPFVVHTIAKYGKSSFMWLGPRPRVFI  120
CYP72A155      61 RPLVGDIRDMIKMIKEAKARYMDPYSDIIAPRVLPYVVHTVAKYGKNSFMWLGPNPRVFI  120

CYP72A566     121 MDPDKIKEMTTKVYEFQKPETSPLFKLLASGFANYDGDKWAKHRKIVSPAFNVEKMKMLI  180
CYP72A61v2    121 MEPDKIKEMTTKVYDFQKPDTSPLFKLLASGFANYDGDKWAKHRKIVSPAFNVEKMKMLI  180
CYP72A155     121 MDPDKIKEMTTKVYEFQKPETSPLFKLLASGFANYDGDKWAKHRKIVSPAFNVEKLKMLI  180

CYP72A566     181 PIFSQSCDEMVKKLESVVSTSNGTCELDVWPFVQNVSSDVLARAGFGSSYGEGKRVFELQ  240
CYP72A61v2    181 PIFCHSCDDMVNKLDKVVSSSNGECELDIWPFVQNVSSDVLARAGFGSSFEEGKRVFQLQ  240
CYP72A155     181 PIFSQSCDEMVNKLESVVSASNGTCELDIWPFCQNVSRDVLARAGFGSSYEEGKRVFELQ  240

CYP72A566     241 REMLTLTMTLFRFAFIPGYRFLPTYTNRRMKAIDKEIRTSLMVIINRRLKAMKAGEPSNN  300
CYP72A61v2    241 KEMISLTMTLFKFAFIPGYRFLPTYTNRRMKAIDLEIRTSLMKIINRRLKAIKAGEPTNN  300
CYP72A155     241 KEMLRLTVTLFRFAFILGYRFLPTYTNRRMKAIDKEVRTSLMAIINRRLKAMKEGEPTNN  300

CYP72A566     301 DLLGHLLESNYKESEKKISGGGMSLREVVEEVKLFYLAGQEANAELLVWTLLLLSKHPEW  360
CYP72A61v2    301 DLLGHLLESNYKESEKGNGGGMSLRDVVDEVKLFYLAGQEANAELLVWTLLLLAKNPEW  360
CYP72A155     301 DLLGLLLESNCK-SEKNISGGGMSLREVVEEVKLFYLAGQEANAELLVWTLLLLSKHPKW  359

CYP72A566     361 QAKAREEVFQVFGHGKPDYDKIGQLKIVSMILQESLRLYPPVVMLSRYLRKDAKLGDLTI  420
CYP72A61v2    361 QAKAREESFQVFGNENPDYEKIGQLKIVSMILQESLRLYPPVIMLSRFLRKDTKLGDLTL  420
CYP72A155     360 QAKAREEVFQVLGHGKPDYDKIGQLKIVSMILQESLRLYSPAVVISRYLRKDAKLGDLTL  419

CYP72A566     421 PAGVELIVPVSMMHQEKEFWGDDAGEFNPERFSEGVSKATNGKVGYLPFGWGPRLCIGQN  480
CYP72A61v2    421 PAGVELIVPVSMMHQEKEFWGDDAGDFKPERFSEGVSKATNGKVSYLPFGWGPRLCIGQN  480
CYP72A155     420 PAGVELIIPVSMMHQEKEFWGDDAGEFNPERFSEGVSKATNGKVGYLPFGWGPRLCIGQN  479

CYP72A566     481 FGLLEAKIAVAMILQRFSLELSPSYSHAPSFIITLQPERGAHLILHKL              528
CYP72A61v2    481 FGLLEAKIAVSMILRQFSLEFSPSYTHAPSFIITLQPEFGAHLILHKL              528
CYP72A155     480 FGLLEAKIAVAMILQRFSLELSPSYSHAPSFIITLQPEFGAPLILHKL              527
```

Figure 2-10. Multiple alignments of CYP72A566, CYP72A61v2, and CYP72A155. Multiple alignments were generated using GENETYX-MAC ver. 18.0.3 software. The fully conserved residues are highlighted with a black background, and residues conserved in at least two of the three P450 proteins are highlighted with a gray background.

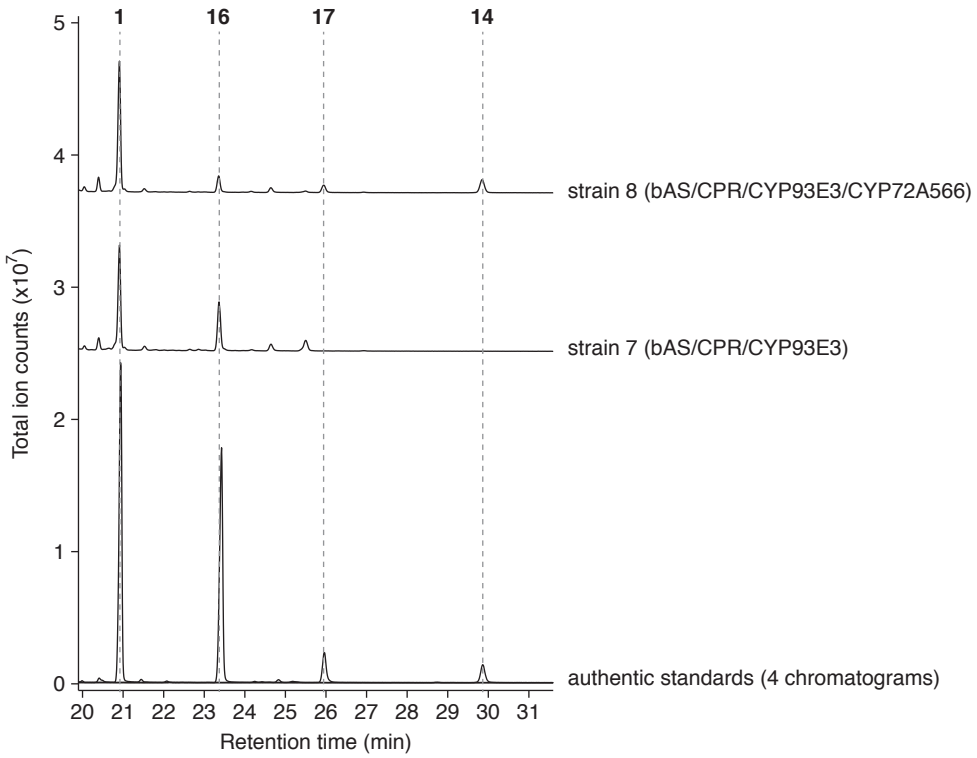


Figure 2-11. GC-MS analysis of products isolated from engineered yeast strains for *in vivo* enzyme assay of CYP72A566. β-Amyrin (**1**), 24-hydroxy-β-amyrin (**16**), sophoradiol (**17**), and soyasapogenol B (**14**) were identified through comparison with authentic standards. To avoid overlapping chromatograms, the baselines for strains 7 and 8 were moved 2.5×10^7 and 3.7×10^7 total ion counts, respectively. Mass spectra of identified peaks are shown in Figure 2-12.

engineered yeast strain endogeneously producing β-amyrin and 24-hydroxy-β-amyrin (strain 7, Table 2-2). In strain 7, production of β-amyrin (**1**) and 24-hydroxy-β-amyrin (**16**) was observed (Figure 2-11; mass spectra in Figure 2-12). When CYP72A566 was introduced into this strain to generate strain 8 (Table 2-2), new peaks corresponding to sophoradiol (22β-hydroxy-β-amyrin) (**17**) and soyasapogenol B (**14**) were confirmed (Figure 2-11). Therefore, CYP93E3 and CYP72A566 are two P450s involved in the biosynthesis of soyasapogenol B from β-amyrin.

2-3-5. Transcript levels of triterpenoid biosynthetic genes in intact roots and tissue-cultured stolons

To address the differences in triterpenoid biosynthesis between intact roots and tissue-cultured stolons, transcript levels of triterpenoid biosynthetic genes including *CYP716A179* and *CYP72A566* were examined by quantitative real-time PCR (qPCR) analysis (Figure 2-13). As predicted from the RNA-Seq analysis, the transcript level of *CYP716A179* in tissue-cultured stolons was about 500 times higher than in intact roots. Similarly, the transcript level of *LUS* in tissue-cultured stolons was about

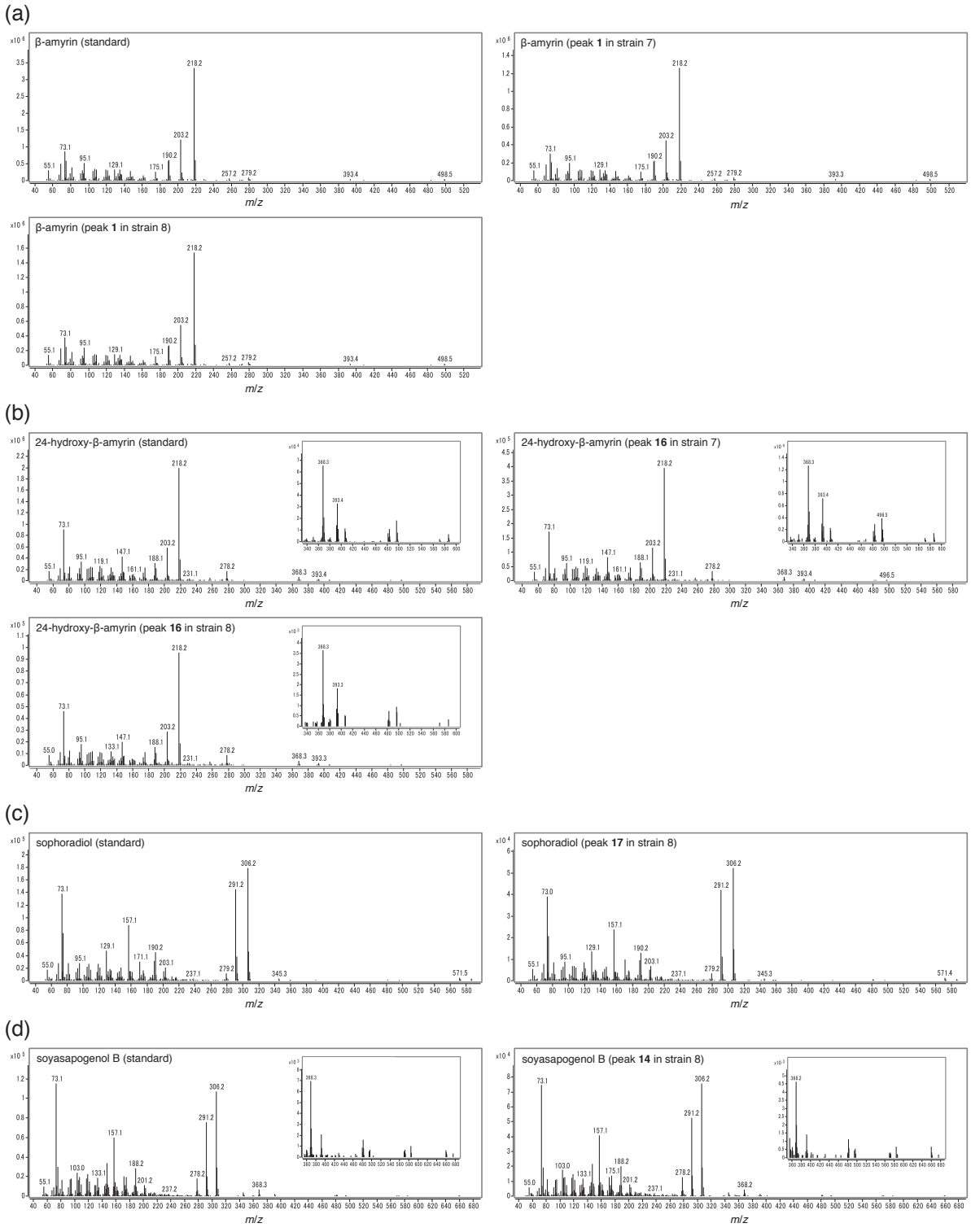


Figure 2-12. Mass spectra of the peaks identified in Figure 2-11 and those of authentic standards: (a) β-amyrin, (b) 24-hydroxy-β-amyrin, (c) sophoradiol, and (d) soyasapogenol B. Insets show enlarged spectra.

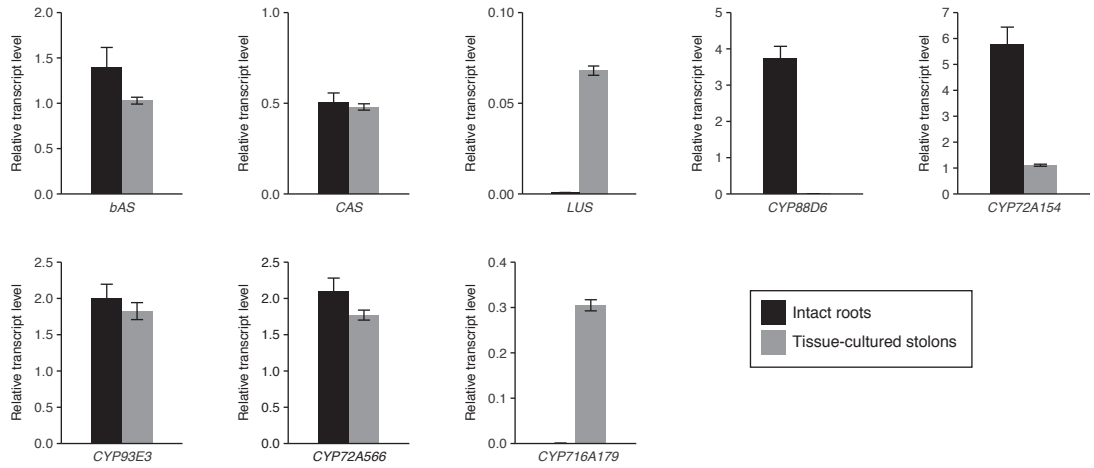


Figure 2-13. Relative transcript levels of OSCs and P450s in intact roots and tissue-cultured stolons. The transcript levels of each gene were normalized to *β-tubulin*. Error bars indicate the SE of three technical replicates.

90 times higher than in intact roots. In contrast, transcript levels of *CYP88D6* and *CYP72A154*, two key P450 genes for glycyrrhizin biosynthesis, were much higher in intact roots; particularly, the transcript level of *CYP88D6* in intact roots was about 700 times higher than that in tissue-cultured stolons. The transcript levels of *bAS*, *CYP93E3*, *CYP72A566*, and *cycloartenol synthase* (*CAS*) were comparable between these two samples.

2-4. Discussion

P450s are key elements in the biosynthesis of various plant triterpenoids. In *G. uralensis*, three P450s involved in the biosynthesis of glycyrrhizin (CYP88D6 and CYP72A154) and soyasaponins (CYP93E3) have been reported thus far (Seki *et al.*, 2008, 2011). In this chapter, I identified two additional P450s, CYP716A179 and CYP72A566, which are involved in the triterpenoid biosynthesis in licorice.

CYP716A179 shares 86% amino acid identity with CYP716A12 from *M. truncatula* (Figure 2-5). The CYP716A subfamily is the most abundant of the P450s involved in triterpenoid biosynthesis (Seki *et al.*, 2015). Most of the CYP716A subfamily enzymes currently identified catalyze oxidation at the C-28 position of pentacyclic triterpene skeletons, such as β-amyrin, α-amyrin,

and/or lupeol. As predicted from the phylogenetic relationships between CYP716A179 and CYP716A12 (Figure 2-6), CYP716A179 showed almost identical enzymatic activities to CYP716A12 (Figure 2-9).

Identification of *CYP716A179* was accomplished by RNA-Seq analysis of tissue-cultured stolons of *G. uralensis*. The initial attempt to identify genes similar to *CYP716A12* in previous transcriptome or expressed sequence tag (EST) analyses of intact roots, leaves, or stolons of *G. uralensis* (Sudo *et al.*, 2009; Ramilowski *et al.*, 2013) was unsuccessful due to extremely low transcript levels of *CYP716A179* in intact plants (Figure 2-13). I also note that no CYP716A subfamily gene was annotated in previous EST analyses of vegetative organs (roots, stems, and leaves) of *G. uralensis* performed by another research group (Li *et al.*, 2010).

CYP72A566 was identified as the corresponding contig of *CYP72A155* by RNA-Seq of tissue-cultured stolons of *G. uralensis*. The amino acid identity of P450s encoded by these two genes was 92% (Figure 2-10). A comparison of amino acid sequences with CYP72A61v2, a previously identified C-22β hydroxylase of the β-amyrin skeleton in *M. truncatula* (Fukushima *et al.*, 2013), revealed that CYP72A61v2 shares greater similarity with CYP72A566 (amino acid sequence identity = 85%) than with CYP72A155 (amino acid sequence identity = 82%). The amino acid residues that CYP72A61v2 shares with CYP72A566 but not with CYP72A155 potentially identify the residues necessary for the catalytic activity of CYP72A566. As the C-30 oxidation of the β-amyrin skeleton is catalyzed by orthologous CYP72A subfamily enzymes in *G. uralensis* and *M. truncatula*, the present study shows that C-22β oxidation of the β-amyrin skeleton is also catalyzed by orthologous CYP72A subfamily enzymes in these two legume species (Figure 2-6).

The triterpenoid profiles of intact roots and tissue-cultured stolons of *G. uralensis* indicate that glycyrrhizin is the most prominent triterpenoid in intact roots, while tissue-cultured stolons accumulate relatively high levels of soyasaponins, betulinic acid, and oleanolic acid, with low glycyrrhizin production (Table 2-3). The identification of CYP716A179 and CYP72A566 in *G. uralensis* explains the observed difference in triterpenoid profiles in terms of gene expression. A comparative analysis of the expression levels of triterpenoid biosynthetic genes between intact roots and tissue-cultured stolons showed that glycyrrhizin biosynthetic genes (*CYP88D6* and *CYP72A154*)

are highly expressed in intact roots (Figure 2-13). Furthermore, *CYP716A179* was highly expressed in tissue-cultured stolons, corresponding to the observed triterpenoid profiles of these two samples. Soyasaponin biosynthetic genes were expressed at almost identical levels between the two samples, suggesting that the higher level of accumulation of soyasaponins in tissue-cultured stolons is due to the lack of expression of glycyrrhizin biosynthetic genes. The results presented in this chapter suggest that the triterpenoid biosynthetic genes in each biosynthetic pathway are differentially regulated between intact roots and tissue-cultured stolons. Identification of potential regulators, including the transcription factors regulating each triterpenoid biosynthetic pathway, is required to understand the regulatory mechanisms underlying triterpenoid biosynthesis in different plant tissues.

Chapter 3

Identification of P450s involved in triterpenoid biosynthesis in *Platycodon grandiflorus*

3-1. Introduction

Platycodon grandiflorus (balloon flower, 'kikyo' in Japanese), the sole species of the genus *Platycodon*, belongs to the family Campanulaceae. As one of the seven flowers of autumn in Japan, *P. grandiflorus* flowers (Figure 3-1a) have been popular for a long time. On the other hand, *P. grandiflorus* roots (Figure 3-1b) are considered one of the most important crude drugs in Kampo medicine (Fukumura *et al.*, 2010); they are used to treat multiple conditions, including cough, abscess, excessive phlegm, and sore throat (Li *et al.*, 2015). Triterpenoid saponins are the most important bioactive components found in platycodon roots. To date, more than 70 different triterpenoid saponins have been isolated from platycodon roots (Nyakudya *et al.*, 2014; Zhang *et al.*, 2015). A major saponin that has sugar moieties attached to the C-3 and C-28 of platycodigenin (sapogenin) is platycodin D (Zhang *et al.*, 2015) (Figure 3-1c). Although chemical structures of platycodin D and platycodigenin were characterized more than 30 years ago (Kubota and Kitatani, 1969; Tada *et al.*, 1975), the biosynthetic enzymes that produce them have not been identified.

As mentioned in Chapter 1, recent studies have revealed that cytochrome P450 monooxygenases (P450s) and UDP-dependent glycosyltransferases (UGTs) play important roles in the biosynthesis of plant triterpenoid saponins (Seki *et al.*, 2015). From the chemical structure of platycodin D, P450s introducing a carboxyl group at C-28 and hydroxyl groups at C-2β, C-16α, C-23, and C-24 of β-amyrin skeleton are predicted to form platycodigenin, which is subsequently glycosylated to produce platycodin D (Figure 3-1c).

Several P450s catalyzing oxidation reactions of the abovementioned carbon positions of β-amyrin or β-amyrin-derived compounds have been reported in other plant species. The CYP716A subfamily enzymes carboxylate β-amyrin at C-28 to generate oleanolic acid. These enzymes have been isolated from a variety of plant species, including *Arabidopsis thaliana* and *Barbarea vulgaris* (Brassicaceae), *Medicago truncatula* (Fabaceae), *Vitis vinifera* (Vitaceae), *Catharanthus roseus*

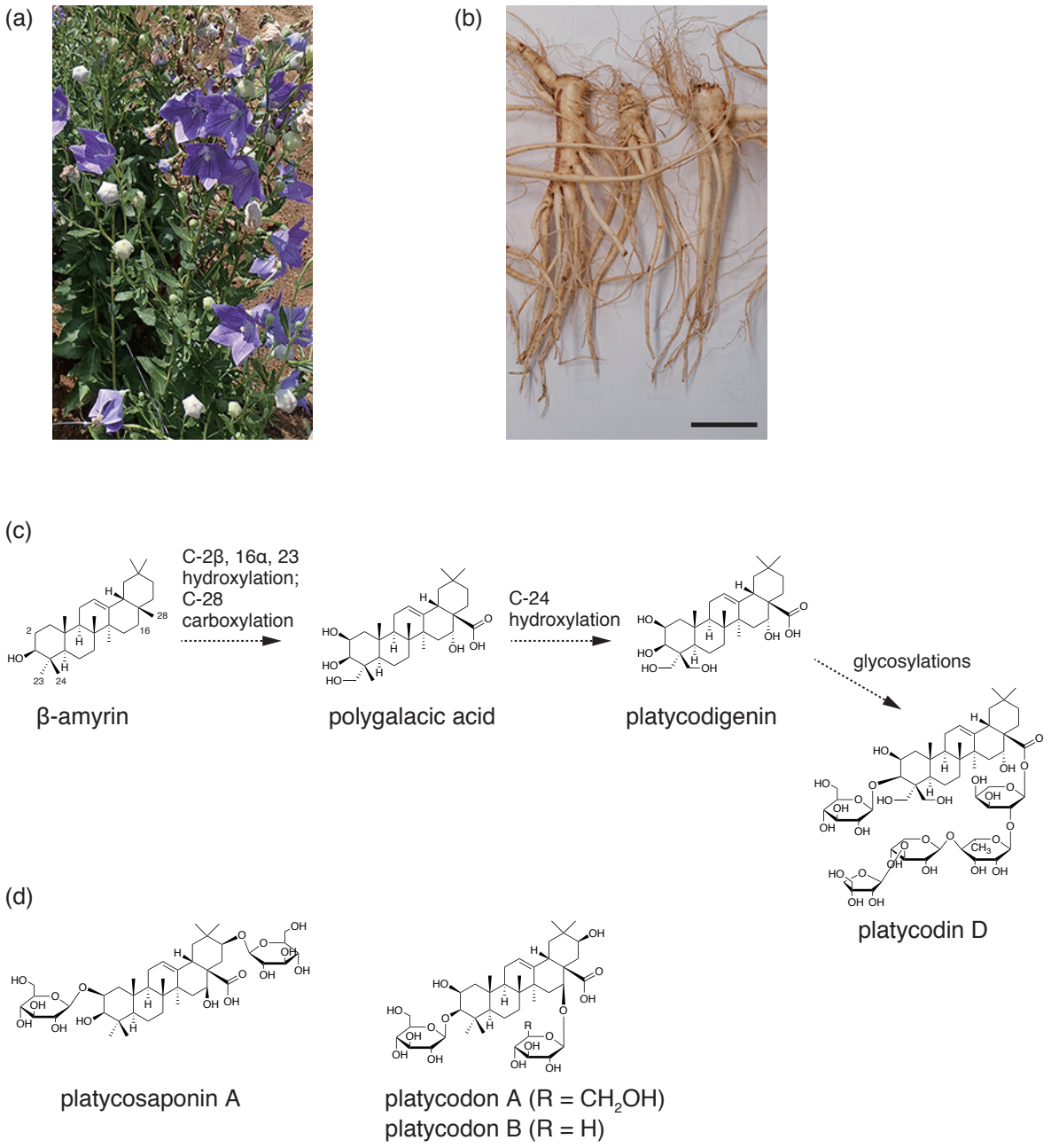


Figure 3-1. *Platycodon grandiflorus* plants and their triterpenoid saponins.
(a) Four-year-old whole plants (aerial parts) of *P. grandiflorus* cultivated at National Institutes of Biomedical Innovation, Health and Nutrition (NIBIOHN), Tsukuba, Japan. Photo by Dr. Noriaki Kawano (NIBIOHN).
(b) Four-year-old roots of *P. grandiflorus* cultivated at NIBIOHN, Tsukuba, Japan. The scale bar indicates 5 cm. Photo by Dr. Kayo Yoshimatsu (NIBIOHN).
(c) Simplified reaction scheme showing the formation of platycodin D from β-amyrin. Platycodin D is the major triterpenoid saponin in *P. grandiflorus*. Platycodigenin is the sapogenin of platycodin D, and polygalacic acid is a possible biosynthetic intermediate between β-amyrin and platycodigenin. Numbers indicated on the β-amyrin structure are carbon positions.
(d) Triterpenoid saponins derived from sapogenins containing a C-16β hydroxyl group.

(Apocynaceae), and *Panax ginseng* (Araliaceae) (Carelli *et al.*, 2011; Fukushima *et al.*, 2011; Huang *et al.*, 2012; Han *et al.*, 2013; Khakimov *et al.*, 2015; Yasumoto *et al.*, 2016). CYP716Y1 and CYP87D16 hydroxylate β-amyrin at C-16α in *Bupleurum falcatum* (Apiaceae) and *Maesa lanceolata* (Primulaceae), respectively (Moses *et al.*, 2014a, 2015a). In legumes, which include *Glycine max*, *M. truncatula*, and *Glycyrrhiza uralensis* (Fabaceae), the hydroxylation of β-amyrin at C-24 depends on the CYP93E enzyme subfamily (Shibuya *et al.*, 2006; Seki *et al.*, 2008; Fukushima *et al.*, 2013). In *M. truncatula*, the hydroxylation of oleanolic acid is catalyzed by both CYP72A67 (C-2β) and CYP72A68v2 (C-23) (Fukushima *et al.*, 2013; Biazzi *et al.*, 2015). Recent studies suggested that CYP716 family enzymes are commonly involved in triterpene oxidation in eudicots, but CYP72A and CYP93E subfamily enzymes have only been reported catalyzing oxidation of triterpenes in the legumes. Therefore, characterization of the P450s involved in triterpenoid biosynthesis in non-Fabaceae plants is important to understand the evolutionary origins of triterpenoid biosynthetic genes in eudicots.

Since roots are the medicinal parts of *P. grandiflorus* containing triterpenoid saponins, putative P450 genes for platycodigenin biosynthesis are expected to be specifically expressed in roots. In this chapter, I performed RNA sequencing (RNA-Seq) analysis of total RNA isolated from the roots, leaves, and petals of *P. grandiflorus* to identify P450 genes for platycodigenin biosynthesis in this non-Fabaceae medicinal plant, *P. grandiflorus*. Six P450 genes, classified as belonging to the CYP716A, CYP716D and CYP72A subfamilies, were found to be highly expressed in roots. The catalytic activities of these P450s were analyzed using engineered yeast that produces β-amyrin endogenously. The P450s CYP716A140v2 and CYP716A141 were able to catalyze C-28 carboxylation and C-16β hydroxylation of β-amyrin, respectively.

3-2. Materials and methods

3-2-1. Plant materials

P. grandiflorus (Plant Material No. 0522-79TS in NIBIOHN) used in this chapter was grown under natural conditions (sown on March 2012) at the Tsukuba Division, Research Center for Medicinal

Plant Resources, National Institutes of Biomedical Innovation, Health and Nutrition (NIBIOHN, Tsukuba, Japan). Roots, leaves, and petals used for RNA-Seq analysis were harvested in August 2013 and frozen immediately in liquid nitrogen. The roots harvested in August 2013, and leaves and petals harvested in July 2016 were used for sapogenin analysis in plant tissues.

3-2-2. Chemicals

β-Amyrin, erythrodiol, oleanolic acid, and glycyrrhetinic acid (18β-glycyrrhetinic acid) were purchased from Extrasynthese. Platycodigenin and polygalacic acid were purchased from Quality Phytochemicals (East Brunswick, NJ, USA). Methyl-β-cyclodextrin was purchased from Tokyo Chemical Industry. Oleanolic aldehyde was kindly gifted by Dr. Kiyoshi Ohyama (Tokyo Institute of Technology).

3-2-3. RNA extraction

RNA extraction was performed as described in Chapter 2 from frozen plant tissues (3 g of roots, 0.5 g of leaves, and 1 g of petals).

3-2-4. Library construction, Illumina sequencing, and *de novo* sequence assembly

Library construction and Illumina sequencing were performed as described in Chapter 2. The sequence reads were assembled using CLC Genomics Workbench software ver. 5.5.2 (CLC Bio, Aarhus, Denmark). A minimum contig length of 300 bp and the 'perform scaffolding' function provided assembled contigs after the removal of adaptor sequences and low quality reads. Reads were assembled into 40,420 unigenes after removing redundant sequences using the TIGR Gene Indices clustering tools (TGICL) (Pertea *et al.*, 2003). The raw RNA-Seq reads obtained in this chapter have been submitted to the DDBJ Sequence Read Archive (DRA) under the accession number DRA005429.

3-2-5. Functional annotation of the assembled reads

A blastx program- (Altschul *et al.*, 1997) based homology search for 40,420 unigenes was performed

Table 3-1. Primers used in this chapter

Primer no.	Sequence (5' to 3')	Comment
1	CACCATGGAGTTATTGTATGTCTCTCTC	Cloning of *CYP716A140v2*
2	TTAAGCTTTATGTGGATAGAGGC	Cloning of *CYP716A140v2*
3	CACCATGGATTCCCTCTTCATCATCA	Cloning of *CYP716A141*
4	TCATGCCTTGTGAGGAATGA	Cloning of *CYP716A141*
5	CACCATGATATTGATGATGATGATAGTACTC	Cloning of *CYP716D58*
6	CTACTTCACAGAGAGTTGGCA	Cloning of *CYP716D58*
7	CTTAGTGGATTCACCCCCTTAGGCTTCTG	5'-RACE PCR, Unigene2726, first
8	GATTTGGCTTCCCTGACCATCGCCGCC	5'-RACE PCR, Unigene2726, nested
9	CCACGATATTCACCATTGGCTTCGGCCC	5'-RACE PCR, Unigene2754, first
10	GCTTTGGCTTCCTTGGACATGGTCGCC	5'-RACE PCR, Unigene2754, nested
11	CCCTGGGAGGTATAATGATTGTGATGCC	5'-RACE PCR, Unigene3928, first
12	GGCCACACATCTACCTCACAAGGCCTC	5'-RACE PCR, Unigene3928, nested
13	CTTTGCAGGACAGGAGACGACTTCGAAC	3'-RACE PCR, Unigene2726, first
14	CCCGCCAGCTGTTCTACTTAGGCGG	3'-RACE PCR, Unigene2726, nested
15	CTACTTTGCAGGACAGGAGACGACCGC	3'-RACE PCR, Unigene2754, first
16	CCCGCCAGCTGCAGCAGTAATGGCACGG	3'-RACE PCR, Unigene2754, nested
17	CTTAGCTGGCCAAGAAACCACCTCAG	3'-RACE PCR, Unigene3928, first
18	GGCAAGCACGAGCGAGAGAAGAAGTTC	3'-RACE PCR, Unigene3928, nested
19	CACCATGGCCGCCAGTTGTTATGTG	Cloning of *CYP72A554*
20	CTAAAGTTTCTGCAAAATTAAGTGAGC	Cloning of *CYP72A554*
21	CACCATGGCTGTCAATTGTGCGGC	Cloning of *CYP72A555*
22	CTAAAGCCTTTGAAGAACTAAGTGAG	Cloning of *CYP72A555*
23	CACCATGGAGATGGAGATGAAAAGCTT	Cloning of *CYP72A556*
24	TTAAAGTTTGTTTAAAATTAAGTGAGCACC	Cloning of *CYP72A556*
25	TTGACTGAGCGTGGTTATTCTTTC	qPCR for *actin*
26	TCTTCGCAGTTTCTAATTCCTGTTC	qPCR for *actin*
27	TCCCTCTCTCACTCCACTTCC	qPCR for *CYP716A140v2*
28	AACTCGATGCTCTCCCCAAC	qPCR for *CYP716A140v2*
29	GTGTTGCTCCCACCGTTTCT	qPCR for *CYP716A141*
30	ACCAAAGGGCAAACCTCAAC	qPCR for *CYP716A141*

The underlined sequences were added to facilitate unidirectional cloning of the product into pENTR/D-TOPO (Thermo Fisher Scientific).

against the NCBI-nr protein database (http://www.ncbi.nlm.nih.gov/) with a cutoff E-value < 10^{-5}, and the maximum number of permitted hits fixed at 20. For each sample, CLC Genomics Workbench ver. 5.5.2 (CLC Bio) aligned the reads to obtain reliable reads per kilobase of exon model per million mapped reads (RPKM) values. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway mapping was performed using the KEGG Automatic Annotation Server (KAAS; http://www.genome.jp/kegg/kaas/) (Moriya *et al.*, 2007) and the bi-directional best hit (BBH) method against datasets of "ath, aly, crb, brp, bna, cit, tcc, gmx, adu, aip, fve, csv, pop, vvi, sly, osa" with the default threshold. To obtain a set of putative P450 genes, protein sequences of the open reading frames (ORFs) for each unigene obtained by TransDecoder ver. 2.0.1 (https://transdecoder.github.io) were searched against the hidden Markov model (HMM) of the P450 domain (PF00067) downloaded from Pfam 30.0 (Finn *et al.*, 2016) by HMMER ver. 3.1b2 (http://hmmer.org) with a cutoff E-value < 10^{-5}.

3-2-6. Cluster analysis

Hierarchical clustering was performed using Cluster 3.0 software (de Hoon *et al.*, 2004) and the

following parameters: similarity metric, correlation (centered); clustering method, average linkage. The clustering results were illustrated using Java TreeView ver. 1.1.6r2 software (Saldanha, 2004).

3-2-7. Cloning of candidate P450 genes

First-strand cDNA was synthesized from 1 μg of total RNA prepared from the roots of *P. grandiflorus* using the SMARTer RACE cDNA Amplification Kit (Clontech/Takara Bio) according to the manufacturer's instructions. Fragments containing the full-length coding sequence (CDS) of *CYP716A140v2*, *CYP716A141*, or *CYP716D58* were amplified by PCR using primers 1–6 (Table 3-1). To obtain the missing 5' sequences of Unigene2726, Unigene2754, and Unigene3928, 5'-rapid amplification of cDNA ends (RACE) PCR was performed using the SMARTer RACE cDNA Amplification Kit (Clontech/Takara Bio) according to the manufacturer's instructions, using primers 7–12 (Table 3-1). To obtain the missing 3' sequences of Ungiene2726, Unigene2754, and Unigene3928, 3'-RACE PCR was performed using primers 13–18 (Table 3-1). Using sequence information obtained by RACE PCR, fragments containing the full-length CDS of *CYP72A554*, *CYP72A555*, or *CYP72A556* were amplified by PCR using primers 19–24 (Table 3-1). The nucleotide sequences isolated in this chapter have been submitted to in DDBJ under the accession numbers LC209199 (CYP716A140v2), LC209200 (CYP716A141), LC209201 (CYP716D58), LC209202 (CYP72A554), LC209203 (CYP72A555), and LC209204 (CYP72A556), respectively.

3-2-8. Generation of engineered yeast strains

Each P450 cDNA was transferred into a Gateway-compatible version of the pELC vector (Seki *et al.*, 2011) using the Gateway LR Clonase II Enzyme mix (Thermo Fisher Scientific) to generate a construct for galactose-inducible dual expression of *cytochrome P450 reductase* (*Lotus japonicus CPR1*) and the P450. In addition, each P450 cDNA was transferred into pYES-DEST52 (Thermo Fisher Scientific) and a Gateway-compatible version of pESC-HIS (Agilent Technologies) using the Gateway LR Clonase II Enzyme mix (Thermo Fisher Scientific) to generate galactose-inducible expression of each P450. These constructs were transformed into *Saccharomyces cerevisiae* INV*Sc*1 (*MAT*a *his3*Δ*1 leu2 trp1-289 ura3-52*/*MAT*α *his3*Δ*1 leu2 trp1-289 ura3-52*; Thermo Fisher Scientific)

Table 3-2. Yeast strains generated in this chapter

Strain no.	Genotype
1	INV*Sc*1; pYES3[*ADH1*/*bAS*]; pESC-LEU[*GAL10*/*CPR1*]
2	INV*Sc*1; pYES3[*ADH1*/*bAS*]; pESC-LEU[*GAL10*/*CPR1*, *GAL1*/*CYP716A140v2*]
3	INV*Sc*1; pYES3[*ADH1*/*bAS*]; pESC-LEU[*GAL10*/*CPR1*, *GAL1*/*CYP716A141*]
4	INV*Sc*1; pYES3[*ADH1*/*bAS*]; pESC-LEU[*GAL10*/*CPR1*, *GAL1*/*CYP716D58*]
5	INV*Sc*1; pYES3[*ADH1*/*bAS*]; pESC-LEU[*GAL10*/*CPR1*, *GAL1*/*CYP716D58*]; pYES-DEST52[*GAL1*/*CYP716D58*]; pESC-HIS[*GAL1*/*CYP716D58*]
6	INV*Sc*1; pYES3[*ADH1*/*bAS*]; pESC-LEU[*GAL10*/*CPR1*, *GAL1*/*CYP72A554*]; pYES-DEST52[*GAL1*/*CYP72A554*]; pESC-HIS[*GAL1*/*CYP72A554*]
7	INV*Sc*1; pYES3[*ADH1*/*bAS*]; pESC-LEU[*GAL10*/*CPR1*, *GAL1*/*CYP72A555*]; pYES-DEST52[*GAL1*/*CYP72A555*]; pESC-HIS[*GAL1*/*CYP72A555*]
8	INV*Sc*1; pYES3[*ADH1*/*bAS*]; pESC-LEU[*GAL10*/*CPR1*, *GAL1*/*CYP72A556*]; pYES-DEST52[*GAL1*/*CYP72A556*]; pESC-HIS[*GAL1*/*CYP72A556*]
9	INV*Sc*1; pYES3[*ADH1*/*bAS*]; pESC-LEU[*GAL10*/*CPR1*, *GAL1*/*CYP716A141*]; pYES-DEST52[*GAL1*/*CYP716A141*]; pESC-HIS[*GAL1*/*CYP716A141*]
10	INV*Sc*1; pYES3[*ADH1*/*bAS*]; pESC-LEU[*GAL10*/*CPR1*, *GAL1*/*CYP716A141*]; pYES-DEST52[*GAL1*/*CYP716A140v2*]
11	INV*Sc*1; pYES3[*ADH1*/*bAS*]; pESC-LEU[*GAL10*/*CPR1*, *GAL1*/*CYP716A140v2*]; pYES-DEST52[*GAL1*/*CYP716D58*]
12	INV*Sc*1; pYES3[*ADH1*/*bAS*]; pESC-LEU[*GAL10*/*CPR1*, *GAL1*/*CYP716A140v2*]; pYES-DEST52[*GAL1*/*CYP72A554*]
13	INV*Sc*1; pYES3[*ADH1*/*bAS*]; pESC-LEU[*GAL10*/*CPR1*, *GAL1*/*CYP716A140v2*]; pYES-DEST52[*GAL1*/*CYP72A555*]
14	INV*Sc*1; pYES3[*ADH1*/*bAS*]; pESC-LEU[*GAL10*/*CPR1*, *GAL1*/*CYP716A140v2*]; pYES-DEST52[*GAL1*/*CYP72A556*]

Selection markers in yeast are as follows: pYES3, *TRP1*; pESC-LEU, *LEU2*; pYES-DEST52, *URA3*; pESC-HIS, *HIS3*.

harboring pYES3-ADH-OSC1 (bAS, β-amyrin synthase) (Fukushima *et al.*, 2011) using the Frozen-EZ Yeast Transformation II Kit (Zymo Research). The engineered yeast strains made for this chapter are listed (Table 3-2).

3-2-9. *In vivo* P450 enzyme assays

A glycerol stock of each yeast strain was inoculated into 2 mL of appropriate synthetic defined (SD) medium containing 2% glucose, and cultured overnight at 30°C, shaking at 200 rpm. Each starter culture was transferred into 10 mL of identical medium, and cultured for an additional 24 h at 30°C, shaking at 200 rpm. The yeast cells were collected and resuspended in 10 mL of SD medium containing 2% galactose and 25 mM methyl-β-cyclodextrin. These were cultured at 30°C for a further 2 days at 200 rpm. The yeast cultures were extracted three times with 6 mL of ethyl acetate. After the liquid had evaporated, the remaining residue was dissolved in 1 mL of ethyl acetate. The sample was transferred into a Sep-pak Silica 6 cc Vac Cartridge (Waters, Milford, MA, USA) and eluted with 10 mL of ethyl acetate and 10 mL of chloroform/methanol (1:1, v/v). The eluate was allowed to evaporate and the residue was dissolved in 500 μL of chloroform/methanol (1:1, v/v). A total of 100 μL of this solution was allowed to evaporate and trimethylsilylated with 100 μL of *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide (Sigma-Aldrich) at 80°C for 20 min before gas chromatography-mass spectrometry (GC-MS) analysis.

3-2-10. Analysis of sapogenins in plant tissues

Extraction of metabolites, acid hydrolysis, and sample preparation for GC-MS analysis were performed as described in Chapter 2 from freeze-dried powder of plant tissues (40 mg), but the internal standard was 0.1 mg/mL glycyrrhetinic acid and 150 μL of the solution was derivatized.

3-2-11. GC-MS analysis

GC-MS analysis was performed as described in Chapter 2, but the hold time at 300°C was 21 min (extracts from yeasts) or 45 min (extracts from plants), and the *m*/*z* range was 50–750 (extracts from yeasts) or 50–1,050 (extracts from plants).

3-2-12. Identification of compound 5

To purify compound **5**, yeast strain 9 (Table 3-2) was cultured for 96 h in 6 L (24 × 250 mL) of SD medium containing 2% galactose but without tryptophane, uracil, leucine, or histidine. Yeast cells from 6 L of culture were collected and lysed for 2 h using a mixture of 40% (w/v) potassium hydroxide and methanol (1:4, v/v) at 80°C. The lysed cells were extracted three times with hexane. Compound **5** was purified from this organic extract by column chromatography using Silica Gel 60 N (spherical, neutral) (Kanto Chemical, Tokyo, Japan). The extract was passed through the column twice and 19.5 mg of compound **5** was obtained. The structure of compound **5** was determined by nuclear magnetic resonance (NMR) analysis and confirmed using the NMR data reported previously (Quijano *et al.*, 1998).

Maniladiol (**5**): ^{1}H NMR ($CDCl_3$, 400 MHz): δ 0.79 (s, 3H), 0.80 (s, 3H), 0.89 (s, 3H), 0.91 (s, 3H), 0.94 (s, 3H), 0.99 (s, 3H), 1.00 (s, 3H), 1.22 (s, 3H), 3.22 (dd, 1H, J = 11.2, 4.8 Hz), 4.20 (dd, 1H, J = 11.9, 5.5 Hz), 5.24 (t, J = 3.7 Hz). ^{13}C NMR ($CDCl_3$, 100 MHz): δ 15.5, 15.6, 16.8, 18.3, 21.5, 23.5, 24.0, 27.1, 27.2, 28.1, 30.6, 30.9, 32.7, 33.3, 34.2, 35.6, 36.9, 37.3, 38.6, 38.8, 39.9, 43.8, 46.5, 46.8, 49.1, 55.2, 66.0, 79.0, 122.3, 143.5.

3-2-13. NMR analysis

The NMR spectra were recorded on a JNM-ECS400 system (JEOL, Tokyo, Japan) in deuterated chloroform ($CDCl_3$). Tetramethylsilane was added as an internal reference.

3-2-14. Quantitative real-time PCR

Quantitative real-time PCR (qPCR) was performed as described in Chapter 2, but first-strand cDNA was synthesized from 0.5 μg of total RNA in a 10-μL reaction. Relative transcript levels of each target gene were calculated using *actin* (GenBank accession number JF781303) as a reference gene. The amplification of each sample was performed three times, using primers 25–30 (Table 3-1) designed using the Primer3 website (http://bioinfo.ut.ee/primer3-0.4.0/) (Koressaar and Remm, 2007; Untergasser *et al.*, 2012).

3-2-15. Phylogenetic analysis

Phylogenetic analysis was performed as described in Chapter 2.

3-3. Results

3-3-1. RNA-Seq analysis of *P. grandiflorus* and functional annotation of unigenes

Total RNA extracted from *P. grandiflorus* roots, leaves, and petals was sequenced using the Illumina HiSeq 2000 platform. The obtained total 49,247,989 reads were assembled into 41,168 contigs (Table 3-3). After removing the redundant sequences, 40,420 unigenes were identified and 22,251 of these (55.0%) were annotated in the NCBI non-redundant (NCBI-nr) protein database (http://www.ncbi.nlm.nih.gov/). Genes potentially involved in triterpenoid backbone biosynthesis were retrieved from annotations in the KEGG database (Kanehisa and Goto, 2000; Kanehisa *et al.*, 2016) using the KEGG Automatic Annotation Server (KAAS; http://www.genome.jp/kegg/kaas/) (Moriya *et al.*, 2007). The RPKM values for unigenes related to the mevalonate pathway are shown in Table 3-4. To obtain a set of putative P450 genes, I searched protein sequences from the possible ORFs of each unigene using the HMM of P450 domains obtained from PF00067 of the Pfam database and HMMER ver. 3.1b2. I

Table 3-3. Summary of the RNA-Seq analysis of *P. grandiflorus*

Number of total reads used in the assembly	49,247,989
Number of contigs	41,168
N50 of contigs (bp)	1,475
Average length of contigs (bp)	1,024
Minimum length of contigs (bp)	272
Maximum length of contigs (bp)	14,520
Number of unigenes	40,420

Table 3-4. Unigenes related to the mevalonate pathway

Unigene	KO assignment[a]	EC number	Definition	RPKM		
				Root	Leaf	Petal
Unigene1971	K00626	EC 2.3.1.9	acetyl-CoA C-acetyltransferase (ACAT)	47.92	21.71	13.03
Unigene4759	K00626	EC 2.3.1.9	acetyl-CoA C-acetyltransferase (ACAT)	21.54	40.57	36.38
Unigene5769	K01641	EC 2.3.3.10	hydroxymethylglutaryl-CoA synthase (HMGS)	36.69	36.96	25.05
Unigene15440	K00021	EC 1.1.1.34	hydroxymethylglutaryl-CoA reductase (HMGR)	2.04	7.54	112.69
Unigene17091	K00021	EC 1.1.1.34	hydroxymethylglutaryl-CoA reductase (HMGR)	1.56	0.91	12.66
Unigene17052	K00869	EC 2.7.1.36	mevalonate kinase (MVK)	7.69	5.90	11.48
Unigene1343	K00938	EC 2.7.4.2	phosphomevalonate kinase (PMK)	15.81	10.75	12.36
Unigene12762	K01597	EC 4.1.1.33	diphosphomevalonate decarboxylase (MVD)	6.53	11.76	16.13
Unigene4254	K01823	EC 5.3.3.2	Isopentenyl diphosphate isomerase (IDI)	63.75	15.71	69.45
Unigene8008	K00787	EC 2.5.1.1 EC 2.5.1.10	farnesyl diphosphate synthase (FPS)	19.35	31.83	13.25
Unigene7047	K00801	EC 2.5.1.21	squalene synthase (SQS)	55.03	58.91	49.51
Unigene9322	K00511	EC 1.14.14.17	squalene epoxidase (SQE)	14.83	0.06	10.77
Unigene10255	K00511	EC 1.14.14.17	squalene epoxidase (SQE)	22.41	41.84	369.17
Unigene12837	K00511	EC 1.14.14.17	squalene epoxidase (SQE)	15.37	22.19	92.97
Unigene2844	K15813	EC 5.4.99.39	beta-amyrin synthase (bAS)	20.27	8.44	8.91
Unigene20943	K15813	EC 5.4.99.39	beta-amyrin synthase (bAS)	0.65	16.86	29.99
Unigene31367	K15813	EC 5.4.99.39	beta-amyrin synthase (bAS)	3.70	74.27	67.25

[a]KO, KEGG orthology.

annotated 154 unigenes as putative P450 genes.

3-3-2. Selection of candidate P450s

To identify candidate P450s involved in triterpenoid saponin biosynthesis, hierarchical clustering of the 154 putative P450 genes was performed based on the RPKM values of each sample (Figure 3-2). As platycodon roots are a known source of triterpenoid saponins (Nyakudya *et al.*, 2014), I focused on a cluster of 59 unigenes with high expression in roots. To select the best candidate unigenes, I performed blastx searches using the 32 triterpenoid biosynthesis P450s identified to date (Table 3-5). Based on the best hits from these blastx searches against triterpenoid biosynthetic P450s, I selected three unigenes encoding proteins with more than 50% identity to CYP716A15, CYP716A17, or CYP716A154 (Unigene3622, Unigene5427, and Unigene5914), and three unigenes encoding proteins with more than 50% identity to CYP72A67 (Unigene2726, Unigene2754, and Unigene3928). No unigenes encoded proteins with more than 50% identity to the CYP72A68v2, CYP716Y1, CYP87D16, or CYP93E enzymes. The full-length cDNA sequences of these unigenes were determined and

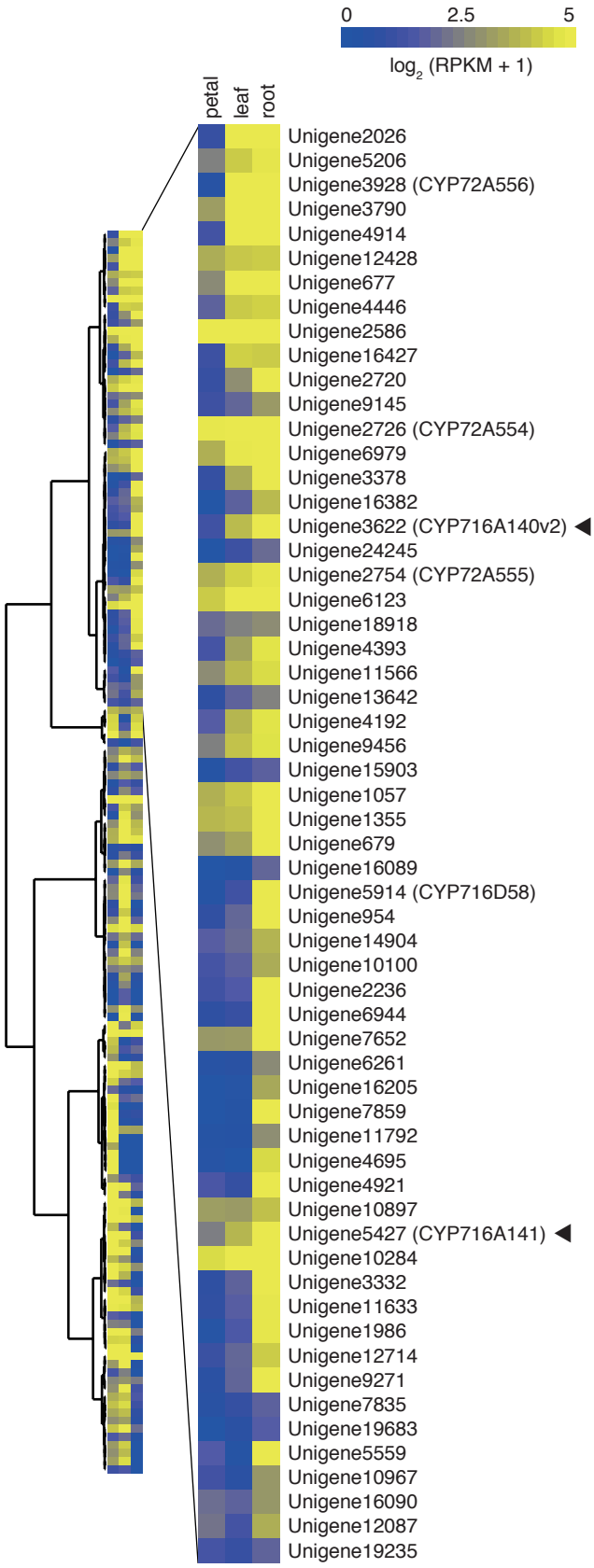


Figure 3-2. Hierarchical clustering of unigenes annotated as putative P450s. The filled arrowheads indicate the P450s defined functionally in this chapter.

Table 3-5. A list of 59 unigenes encoding putative P450s that were highly expressed in roots

Unigene	RPKM			Best hit in the 32 P450s related to triterpenoid biosynthesis[b]			
	Root	Leaf	Petal	P450 name[a]	Amino acid identities (%)	E-value	Score (bit)
Unigene2026	89.03	40.56	0.83	CYP93E2	26.72	2.00E-21	85.5
Unigene5206	27.70	18.38	4.88	CYP88D6	31.58	8.00E-64	210
Unigene3928	85.74	43.54	0.30	CYP72A67	55.87	8.00E-115	337
Unigene3790	110.52	80.05	8.51	CYP72A154	54.25	1.00E-55	180
Unigene4914	62.17	41.22	1.43	CYP93E2	26.92	6.00E-20	82.0
Unigene12428	18.50	17.33	11.40	CYP72A154	57.84	2.00E-37	127
Unigene677	76.55	73.89	5.96	CYP72A68v2	35.84	1.00E-24	92.4
Unigene4446	19.90	18.17	2.56	CYP72A154	44.36	2.00E-115	345
Unigene2586	434.46	458.30	105.43	CYP93E2	44.13	2.00E-57	189
Unigene16427	18.43	19.68	1.05	CYP71D353	38.61	2.00E-19	82.4
Unigene2720	47.79	6.39	0.86	CYP72A67	30.57	4.00E-84	267
Unigene9145	7.99	2.88	1.11	CYP72A61v2	40.82	2.00E-115	349
Unigene2726	128.42	60.12	29.50	CYP72A67	54.07	3.00E-169	481
Unigene6979	74.11	28.88	11.75	CYP88D6	29.22	6.00E-40	140
Unigene3378	112.86	10.91	0.80	CYP93E2	28.15	1.00E-21	86.3
Unigene16382	14.19	2.50	0.00	CYP71D353	39.87	1.00E-119	360
Unigene3622	83.67	14.50	1.30	CYP716A154 (CYP716AL1)	76.40	0.00E+00	733
Unigene24245	3.17	1.08	0.00	CYP71D353	44.91	2.00E-72	226
Unigene2754	27.50	20.59	12.04	CYP72A67	55.13	2.00E-166	474
Unigene6123	51.07	34.79	18.33	CYP72A154	43.77	4.00E-91	280
Unigene18918	6.20	4.89	3.19	CYP72A67	24.10	1.00E-21	89.4
Unigene4393	26.20	9.42	1.55	CYP71D353	34.65	8.00E-71	238
Unigene11566	22.94	13.89	6.22	CYP93E2	31.82	1.00E-83	268
Unigene13642	4.98	2.61	0.65	CYP93E3	42.61	9.00E-33	117
Unigene4192	25.75	13.06	2.24	CYP705A1	34.41	1.00E-76	248
Unigene9456	24.71	16.00	4.82	No hit	-	-	-
Unigene15903	2.46	1.43	0.20	CYP71D353	30.69	2.00E-28	105
Unigene1057	241.69	17.53	12.03	CYP71D353	37.89	2.00E-101	315
Unigene1355	46.96	15.49	13.25	CYP51H10	48.53	4.00E-159	465
Unigene679	115.79	9.95	6.59	CYP705A1	25.17	1.00E-08	47.4
Unigene16089	2.88	0.21	0.00	No hit	-	-	-
Unigene5914	60.04	1.27	0.18	CYP716A17	54.42	6.00E-180	516
Unigene954	218.96	2.92	0.68	CYP72A63	31.42	3.00E-89	282
Unigene14904	12.81	3.19	2.26	CYP71A16	44.92	2.00E-127	379
Unigene10100	11.18	2.48	1.55	CYP88D6	40.66	3.00E-120	369
Unigene2236	51.48	2.03	1.30	CYP716S1v2 (CYP716A53v2)	28.12	2.00E-09	47.0
Unigene6944	33.86	0.95	0.57	CYP71D353	51.70	8.00E-172	492
Unigene7652	72.98	8.16	7.91	CYP716S1v2 (CYP716A53v2)	24.66	2.00E-06	41.6
Unigene6261	5.90	0.33	0.27	CYP72A68v2	38.43	3.00E-119	360
Unigene16205	10.44	0.07	0.00	CYP72A68v2	35.60	1.00E-99	309
Unigene7859	40.32	0.19	0.00	CYP93E2	31.80	3.00E-67	223
Unigene11792	6.38	0.26	0.18	CYP716A52v2	39.90	4.00E-115	343
Unigene4695	22.30	0.00	0.12	CYP87D16	32.67	2.00E-75	240
Unigene4921	245.19	0.77	1.75	CYP716A15	33.18	3.00E-76	246
Unigene10897	15.68	8.08	8.80	CYP705A1	25.38	5.00E-45	160
Unigene5427	75.84	13.02	4.60	CYP716A15	54.44	7.00E-175	502
Unigene10284	155.06	45.56	22.94	CYP93E3	33.92	1.00E-87	274
Unigene3332	44.37	2.62	0.55	CYP93E3	40.60	2.00E-114	346
Unigene11633	27.94	2.28	0.69	CYP705A1	24.80	2.00E-46	164
Unigene1986	28.59	1.88	0.07	CYP93E3	27.66	6.00E-07	42.4
Unigene12714	18.63	2.90	0.99	CYP93E2	34.63	7.00E-95	297
Unigene9271	37.63	2.76	0.48	CYP72A67	34.86	6.00E-89	283
Unigene7835	2.52	0.78	0.31	CYP88D6	26.52	5.00E-14	61.2
Unigene19683	2.19	0.37	0.00	CYP72A67	38.26	1.00E-22	84.3
Unigene5559	35.22	0.22	2.08	CYP71D353	41.00	2.00E-118	356
Unigene10967	7.70	0.36	1.40	CYP71D353	24.05	5.00E-20	84.3
Unigene16090	7.67	2.58	3.38	CYP72A61v2	35.45	1.00E-21	88.2
Unigene12087	11.48	1.46	3.77	CYP71D353	33.55	2.00E-81	261
Unigene19235	2.64	0.83	1.52	CYP87D16	42.57	1.00E-37	129

[a]Outdated names are shown in parentheses.

[b]The 32 P450s used as queries are as follows: CYP51H10, CYP71A16, CYP71D353, CYP72A61v2, CYP72A63, CYP72A67, CYP72A68v2, CYP72A154, CYP87D16, CYP88D6, CYP93E1, CYP93E2, CYP93E3, CYP705A1, CYP708A2, CYP716A1, CYP716A2, CYP716A14v2, CYP716A12, CYP716A15, CYP716A17, CYP716U1 (CYP716A47), CYP716A52v2, CYP716S1v2 (CYP716A53v2), CYP716A75, CYP716A78, CYP716A79, CYP716A80, CYP716A81, CYP716A179, CYP716A154 (CYP716AL1), and CYP716Y1.

the corresponding proteins were named by the P450 nomenclature committee: CYP716A140v2 (Unigene3622), CYP716A141 (Unigene5427), CYP716D58 (Unigene5914), CYP72A554 (Unigene2726), CYP72A555 (Unigene2754), and CYP72A556 (Unigene3928). The phylogenetic relationships among the characterized P450s involved in triterpenoid biosynthesis in plants were analyzed (Figure 3-3).

3-3-3. Enzymatic activities of CYP716A140v2 and CYP716A141

As platycodigenin has a β-amyrin skeleton, each of the six candidate P450s was first expressed together with a *CPR* in a strain of engineered yeast that had been pretransformed with *bAS* and produced β-amyrin endogenously (strains 2–8, Table 3-2). Only CYP716A140v2 and CYP716A141 were found to oxidize the β-amyrin. In the *bAS*/*CPR*/*CYP716A140v2*-expressing yeast (strain 2), erythrodiol (**2**), oleanolic aldehyde (**3**), and oleanolic acid (**4**) were detected by comparing these with authentic standards (Figures 3-4a,b; mass spectra in Figure 3-5). These compounds are oxidation products of β-amyrin (**1**) modified at C-28. Erythrodiol (**2**) and oleanolic aldehyde (**3**) are reaction intermediates occurring between β-amyrin (**1**) and oleanolic acid (**4**). Therefore, CYP716A140v2 was found to be a β-amyrin C-28 oxidase in *P. grandiflorus*. In the *bAS*/*CPR*/*CYP716A141*-expressing yeast (strain 3), an unknown compound corresponding to peak **5** was detected as a major product (Figure 3-4a). To increase the yield and identify compound **5**, a yeast strain with two additional *CYP716A141*-expression vectors (strain 9, Table 3-2) was generated, and obtained 19.5 mg of purified compound **5** from 6 L of culture. NMR analysis showed that compound **5** was a C-16β hydroxylated β-amyrin, called maniladiol (olean-12-ene-3β,16β-diol) (Figure 3-4c). In addition to maniladiol (**5**), erythrodiol (**2**) and oleanolic acid (**4**) were also detected as minor compounds in strain 3 (Figure 3-4a), suggesting that CYP716A141 is capable of catalyzing oxidation at both C-16β and C-28. Strain 3 also produced an unknown minor compound corresponding to peak **6**. In the *bAS*/*CPR*/*CYP716A140v2*/*CYP716A141*-expressing yeast (strain 10, Table 3-2), maniladiol (**5**), erythrodiol (**2**), oleanolic aldehyde (**3**), oleanolic acid (**4**), and the compound corresponding to peak **6** were detected (Figure 3-4a). Compared with the *bAS*/*CPR*/*CYP716A141*-expressing yeast (strain 3), the relative amount of peak **6** increased when *CYP716A140v2* and *CYP716A141* were expressed together (strain 10),

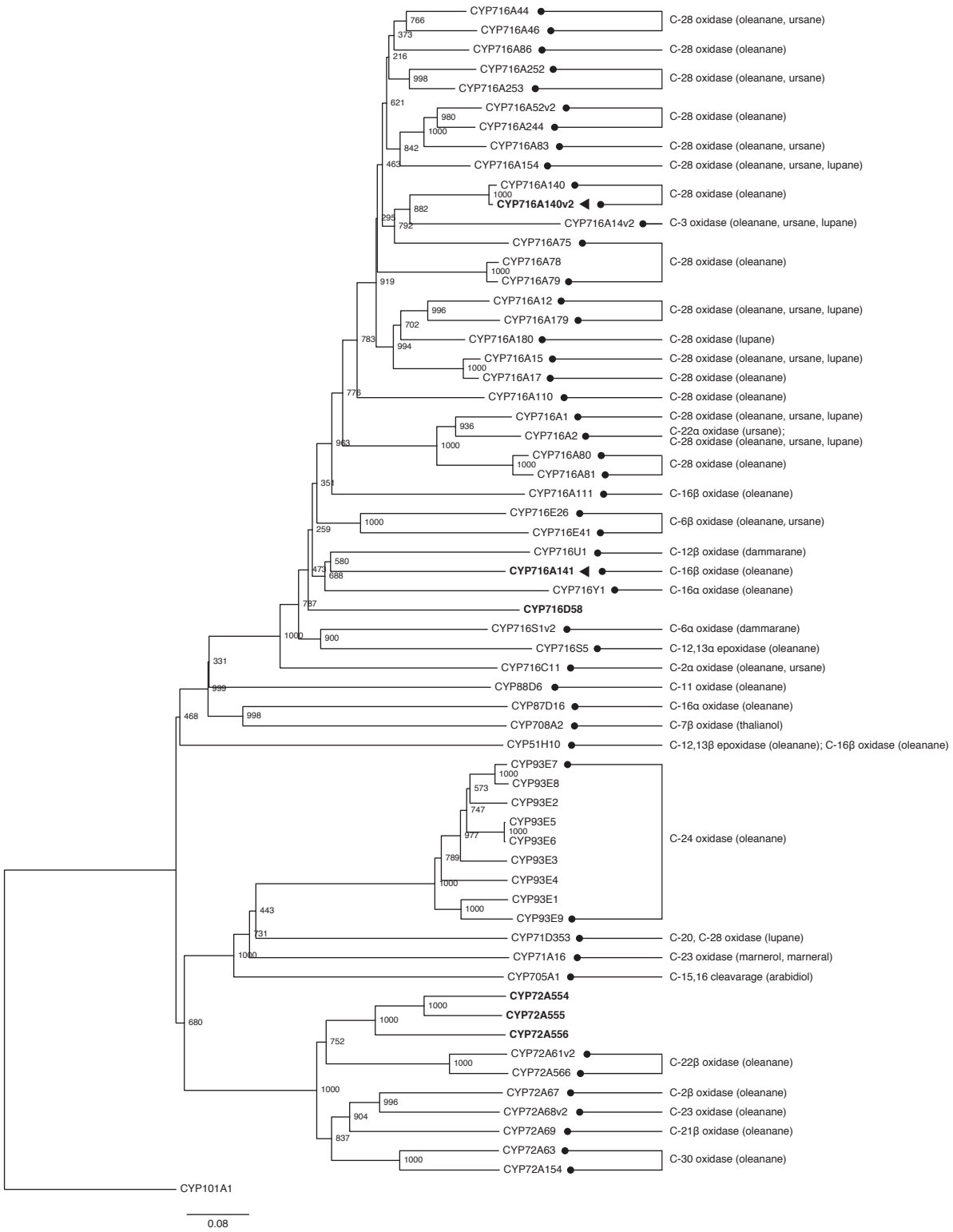


Figure 3-3. Phylogenetic tree of characterized triterpenoid biosynthesis P450s and the *P. grandiflorus* P450s isolated in this chapter. Numbers indicate bootstrap values for 1,000 replicates. The scale bar shows the amino acid substitution ratio. P450s isolated in this chapter are indicated in bold. The filled arrowheads indicate the P450s defined functionally in this chapter. The determined biochemical activities of the P450s are indicated on the right, with their substrate triterpene skeletons or compounds shown in parentheses. GenBank protein accession numbers and species of characterized P450s in this tree are shown in Table 1-2. *Pseudomonas putida* CYP101A1 (GenBank protein accession number AAA25760) was used as an outgroup.

suggesting that peak **6** was the compound carboxylated at C-28 and hydroxylated at C-16β. In addition, the mass spectrum of the peak **6** was similar to that of echinocystic acid (3β,16α-dihydroxy-olean-12-en-28-oic acid) (Moses *et al.*, 2014a) (Figure 3-5). However, since CYP716A141 catalyzed C-16 hydroxylation only in the β-configuration, the compound corresponding to peak **6** is likely to be cochalic acid (3β,16β-dihydroxy-olean-12-en-28-oic acid) (Figure 3-4c).

Among the six candidate P450s cloned in this chapter, four (CYP716D58, CYP72A554, CYP72A555, and CYP72A556) had no catalytic activity on β-amyrin (strains 4–8). Nevertheless, while β-amyrin is not a substrate for the P450s CYP72A67 and CYP72A68v2, these enzymes showed activity on oleanolic acid (Fukushima *et al.*, 2013; Biazzi *et al.*, 2015). Therefore, to assess the potential activity of four candidate P450s on oleanolic acid, I generated yeast strains expressing *CYP716A140v2* and each of the four P450s together (strains 11–14, Table 3-2). However, none of these candidates showed enzymatic activity on oleanolic acid.

3-3-4. Transcript levels of *CYP716A140v2* and *CYP716A141*, and triterpenoid profiles in plant tissues

The relative transcript levels of *CYP716A140v2* and *CYP716A141* in three different tissues of *P. grandiflorus* were determined by qPCR (Figure 3-6a). As predicted by RPKM values obtained during the RNA-Seq analysis, both P450s were highly expressed in roots compared with leaves and petals. *CYP716A140v2* transcripts in roots were more than nine times higher than in leaves, and more than 250 times higher than in petals. In addition, *CYP716A141* transcripts in roots were more than nine times higher than in leaves, and more than 25 times higher than in petals, suggesting that their catalysis occurs predominantly in roots. I also analyzed triterpenoid profiles in these tissues of *P. grandiflorus* (Figure 3-6b). Platycodigenin (**7**) and polygalacic acid (**8**) were detected mainly in roots, but only trace amounts were detected in leaves and petals. On the other hand, β-amyrin (**1**) was detected in leaves and petals at higher levels, but was barely detected in roots. These results suggest that β-amyrin is converted efficiently to sapogenins by P450s highly expressed in roots.

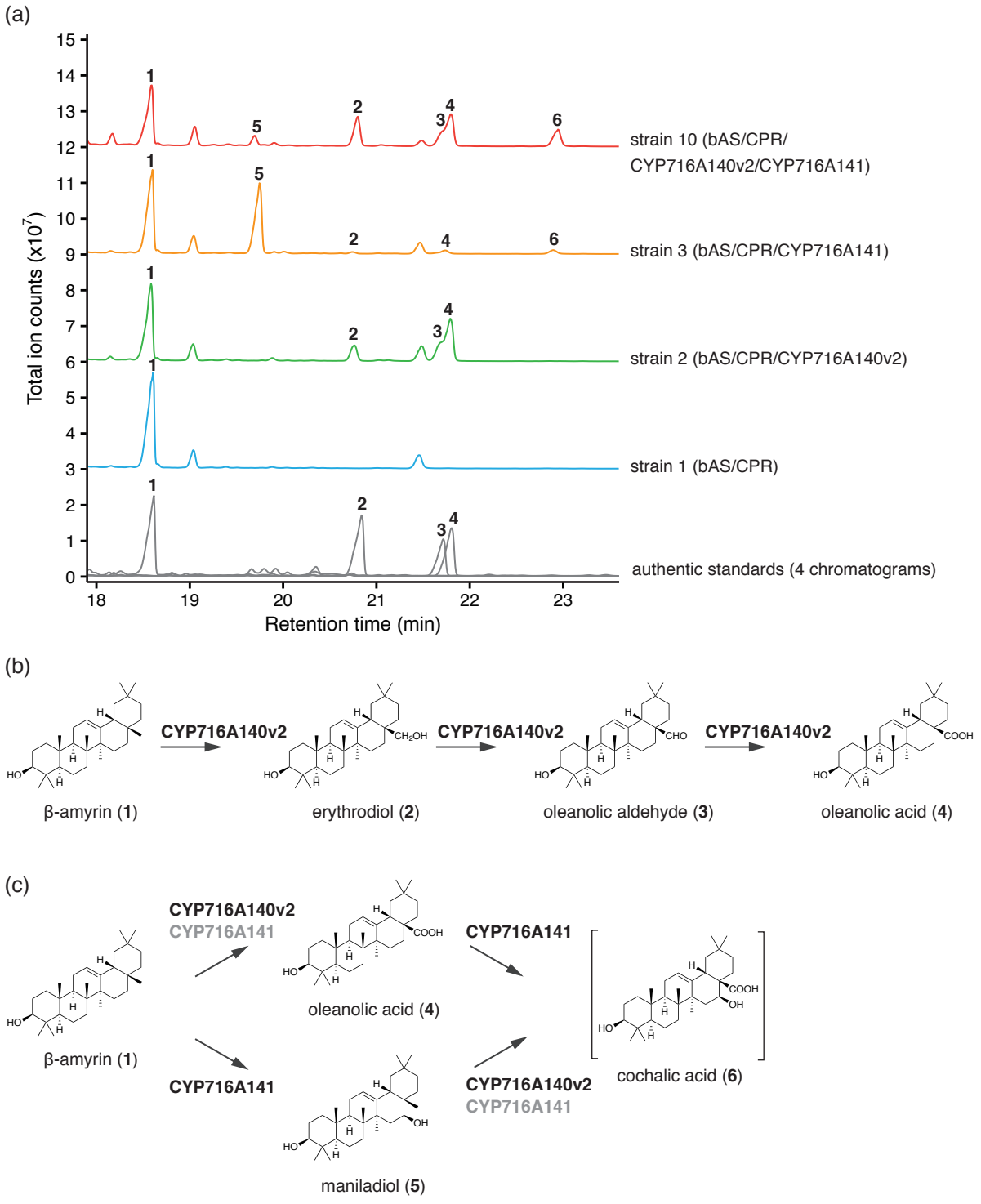


Figure 3-4. Enzyme assays of CYP716A140v2 and CYP716A141 in engineered yeast strains endogenously producing β-amyrin.
(a) GC-MS analysis of the products of engineered yeast expressing *bAS*, *CPR*, and candidate P450(s). To avoid overlapping chromatograms, the baselines for strains 1, 2, 3, and 10 were moved 3×10^7, 6×10^7, 9×10^7, and 1.2×10^8 total ion counts, respectively.
(b) Reactions catalyzed by CYP716A140v2 in β-amyrin-producing engineered yeast. Arrows represent single oxidation reactions catalyzed by CYP716A140v2.
(c) Proposed reactions catalyzed by CYP716A140v2 and CYP716A141 in β-amyrin-producing engineered yeast. The minor reaction catalyzed by CYP716A141 is marked in gray.

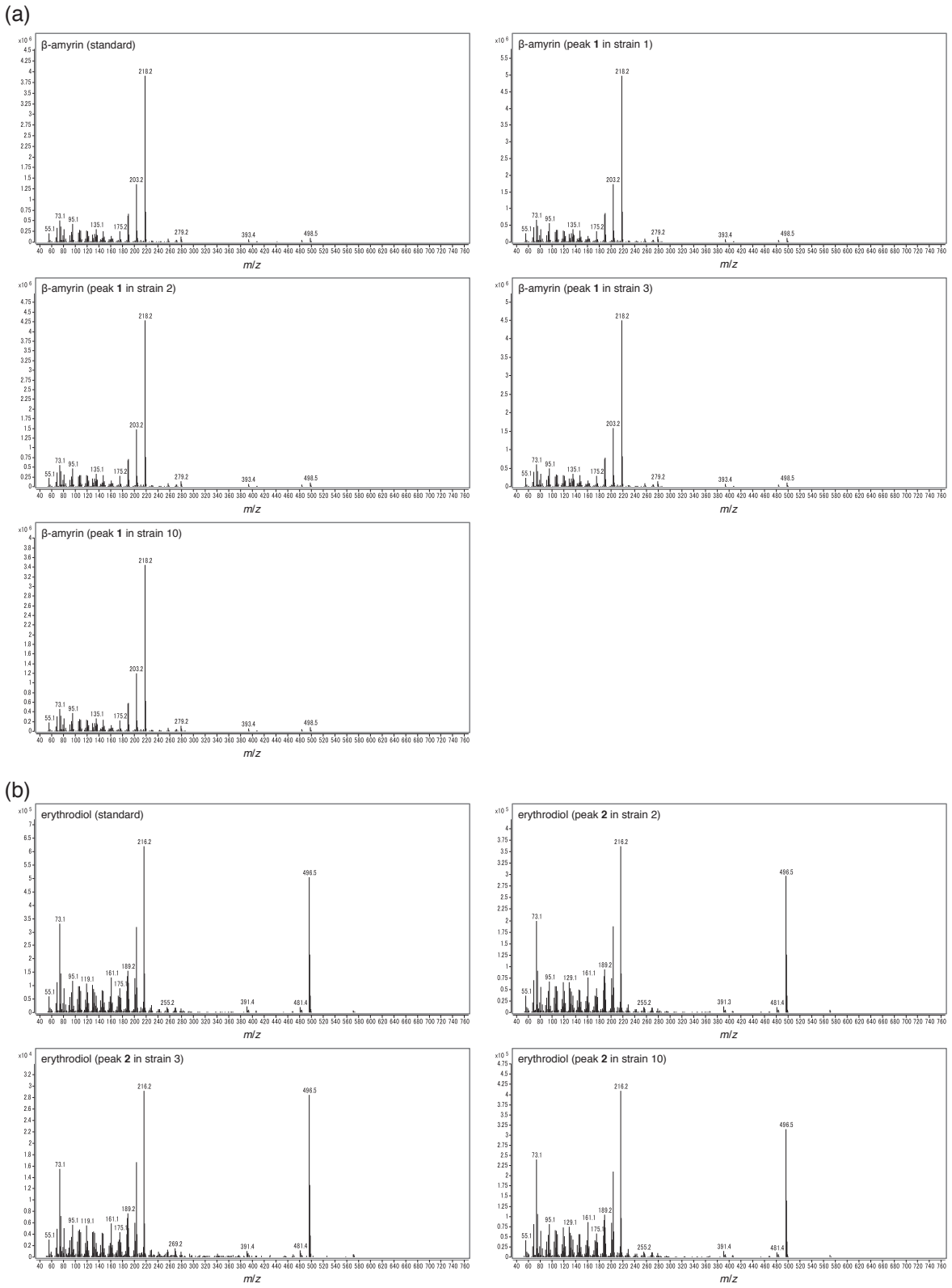
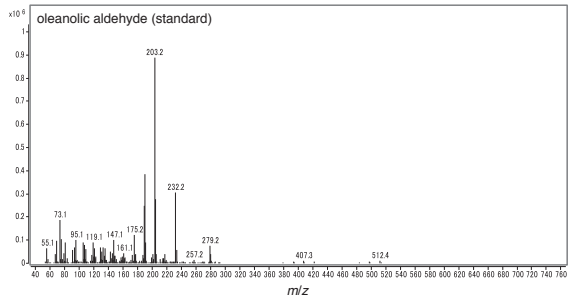
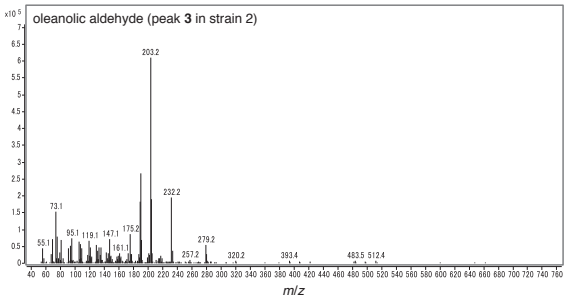
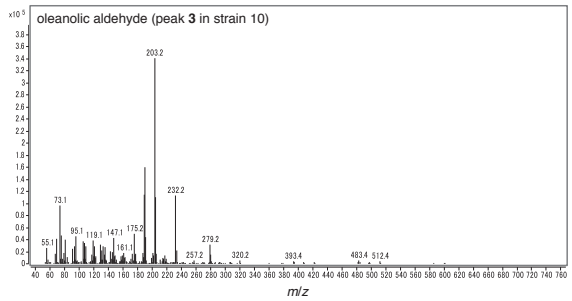


Figure 3-5. Mass spectra of peaks identified in Figure 3-4a and those of authentic standards: (a) β-amyrin, (b) erythrodiol, (c) oleanolic aldehyde, (d) oleanolic acid, (e) maniladiol, and (f) putative cochalic acid.

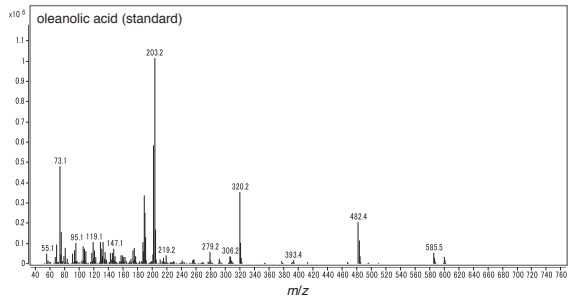
(c)

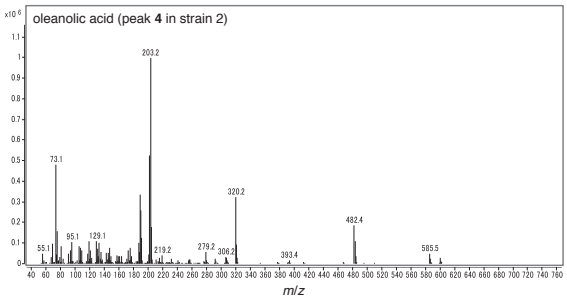


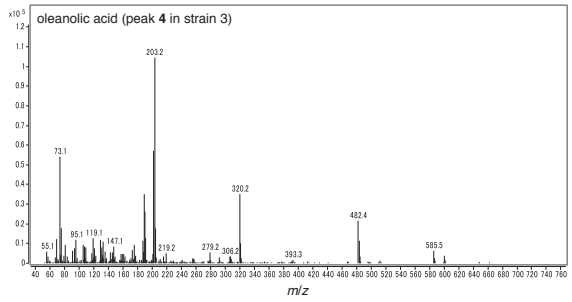


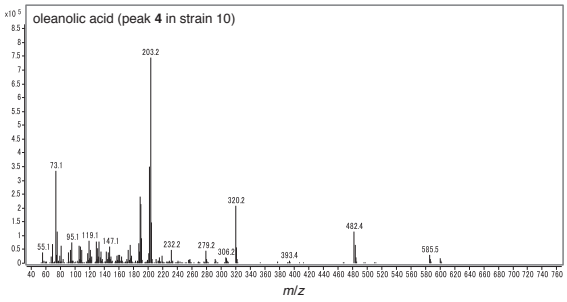


(d)

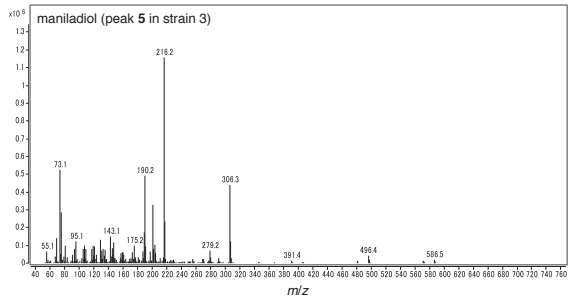








(e)



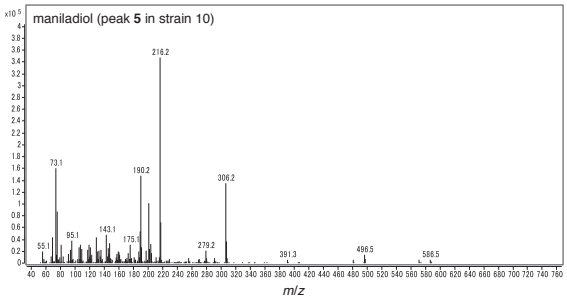


Figure 3-5. (continued)

(f)

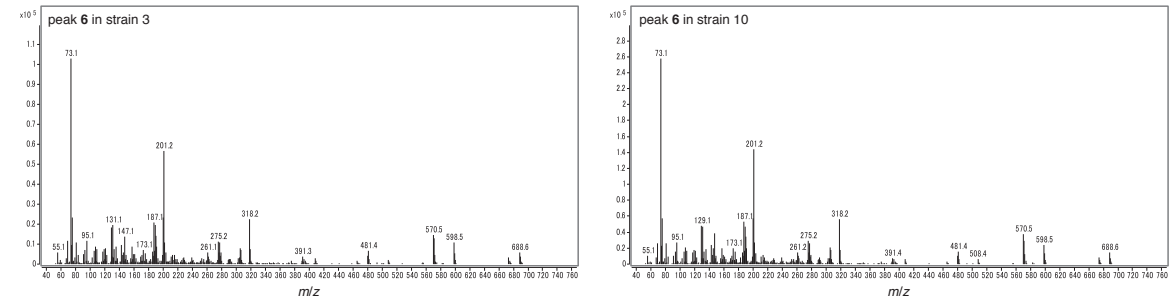


Figure 3-5. (continued)

3-4. Discussion

The results presented here suggested that two P450s, CYP716A140v2 and CYP716A141, function as triterpene oxidases in the medicinal plant, *P. grandiflorus*, the roots of which contain a variety of triterpenoid saponins. CYP716A140v2 is highly expressed in roots, and it catalyzes a three-step oxidation reaction at C-28 of β-amyrin to produce oleanolic acid. These catalytic activities are the same as those of CYP716A subfamily enzymes isolated from other plant species (Carelli *et al.*, 2011; Fukushima *et al.*, 2011; Huang *et al.*, 2012; Han *et al.*, 2013; Khakimov *et al.*, 2015; Yasumoto *et al.*, 2016), indicating that C-28 multiple oxidases are highly conserved in land plants.

CYP716A141 is another P450 belonging to the CYP716A subfamily that is also highly expressed in roots. Heterologous expression of CYP716A141 in β-amyrin-producing yeast indicated that its main activity was hydroxylation of β-amyrin at C-16β. Interestingly, CYP716A141 also had weak activity catalyzing the three-step oxidation at C-28. Although most of the triterpenoid saponins reported in *P. grandiflorus*, including platycodin D, have a C-16α hydroxyl group (Zhang *et al.*, 2015), C-16β hydroxylated triterpenoid saponins, such as platycosaponin A (Fukumura *et al.*, 2010), platycodon A, and platycodon B (Ma *et al.*, 2013), have also been reported (Figure 3-1d). Therefore, CYP716A141 is likely to be involved in the biosynthesis of these triterpenoid saponins. *B. falcatum* produces both C-16α and C-16β hydroxylated bioactive triterpenoid saponins (saikosaponin D and saikosaponin A, respectively) (Chen *et al.*, 2013). However, only the C-16α oxidase has been isolated from this plant (Moses *et al.*, 2014a). In monocots, *Avena strigosa* CYP51H10 is a multifunctional P450 that simultaneously catalyzes the C-12,13β epoxidation and C-16β hydroxylation of β-amyrin in

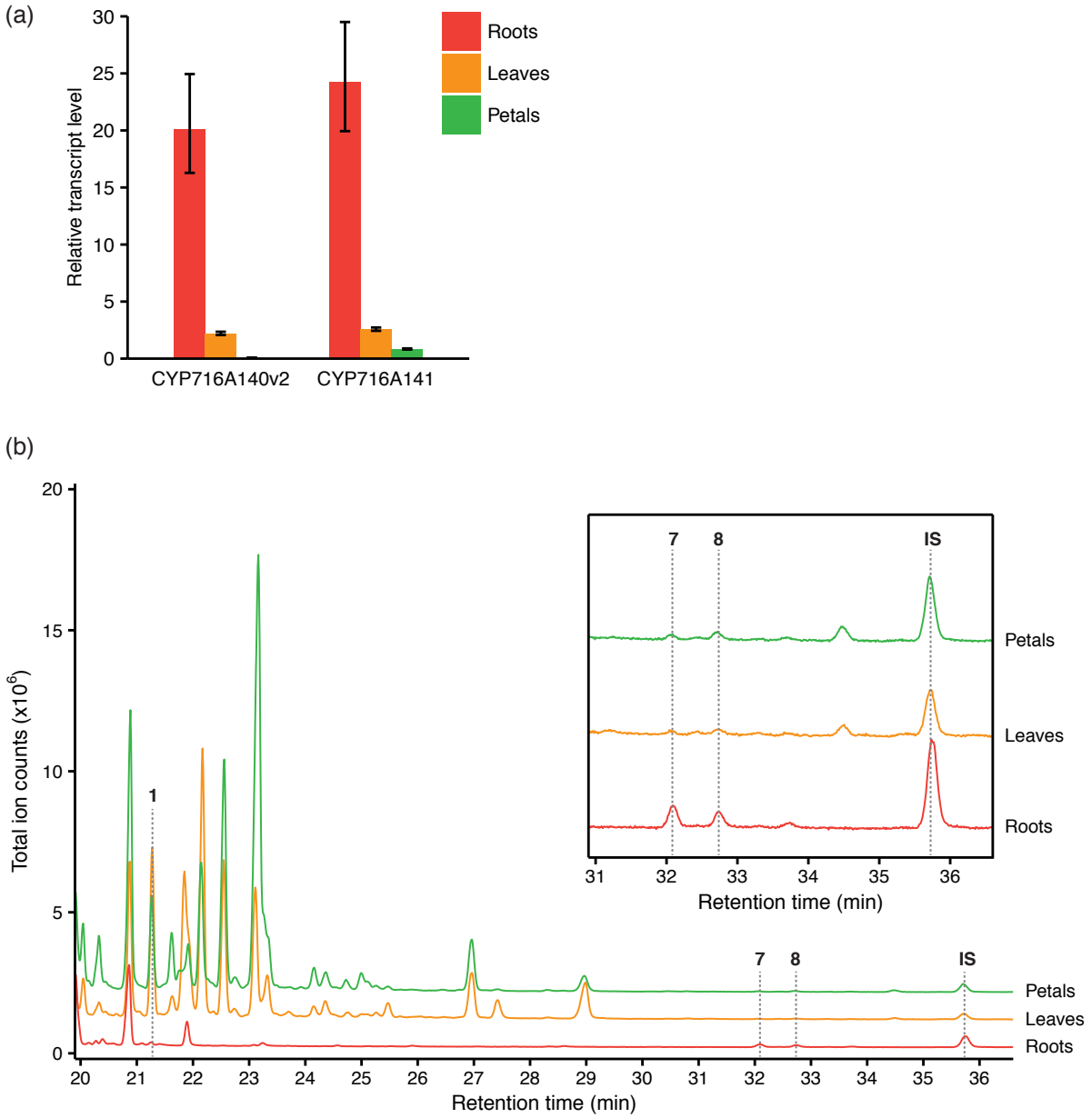


Figure 3-6. Relative transcript levels of *CYP716A140v2* and *CYP716A141*, and triterpenoid profiles in roots, leaves, and petals of *P. grandiflorus*.
(a) Quantitative real-time PCR analysis of *CYP716A140v2* and *CYP716A141* in roots, leaves, and petals. The transcript level of each gene was normalized relative to that of *actin*. Error bars indicate the standard error (SE) of three technical replicates.
(b) GC-MS analysis of the acid-hydrolyzed extracts from roots, leaves, and petals of *P. grandiflorus*. Enlarged chromatograms between 31 and 36.6 min are shown in the inset. Among the products produced in the engineered yeast strains (compounds **1**–**6**, Figure 3-4), only β-amyrin (**1**) was detected. In addition to β-amyrin, platycodigenin (**7**) and polygalacic acid (**8**) were identified. To avoid overlapping chromatograms, the baselines for leaves and petals were moved 1×10^6 and 2×10^6 total ion counts, respectively. Glycyrrhetinic acid was added as an internal standard (IS).

avenacin biosynthesis (Geisler *et al.*, 2013). To my knowledge, however, this is the first report of a P450 that preferentially catalyzes C-16β hydroxylation of β-amyrin.

As described above, the C-16 hydroxyl group of most of the triterpenoid saponins in *P. grandiflorus* is in the α-configuration. However, the P450s responsible for C-16α hydroxylation of β-amyrin in *P. grandiflorus* remain unknown. Previous studies demonstrated that C-16α hydroxylation of β-amyrin is catalyzed by CYP716Y1 in *B. falcatum* and by CYP87D16 in *M. lanceolata* (Moses *et al.*, 2014a, 2015a). Of all the unigene sequences obtained from this transcriptome analysis, Unigene5427 (CYP716A141) encoded the protein with the highest alignment score to CYP716Y1 (amino acid sequence identity = 48.6%) (Table 3-6). In addition, I failed to find unigenes encoding proteins likely to be part of the CYP87D subfamily (Table 3-6). These results suggest that P450 subfamilies other than CYP716Y and CYP87D probably catalyze the C-16α hydroxylation of β-amyrin in *P. grandiflorus*.

In *M. truncatula*, CYP72A67 and CYP72A68v2 hydroxylate the C-2β and C-23 of the β-amyrin skeleton, respectively (Fukushima *et al.*, 2013; Biazzi *et al.*, 2015). Therefore, I cloned three CYP72A subfamily genes that were highly expressed in roots, and analyzed their catalytic activities. However, none of these enzymes had catalytic activity for β-amyrin or oleanolic acid substrates. In this chapter, I selected candidate P450s for their high expression in roots, and so P450s with lower expression in roots but relatively high similarity to CYP72A67 or CYP72A68v2 might have catalytic activity on β-amyrin or related compounds (Table 3-6). Another possibility is that the P450s responsible for hydroxylation at C-2β and C-23 in *P. grandiflorus* are not members of the CYP72A subfamily. CYP93E subfamily members isolated from nine different legume species catalyze hydroxylation at the C-24 of β-amyrin (Shibuya *et al.*, 2006; Seki *et al.*, 2008; Fukushima *et al.*, 2013; Moses *et al.*, 2014b); however, I failed to find a unigene structurally similar to the CYP93E subfamily P450s (Table 3-6), suggesting that P450s belonging to a class different from the CYP93E subfamily may catalyze C-24 hydroxylation in platycodigenin biosynthesis. Interestingly, two P450s belonging to different families, CYP716Y1 in *B. falcatum* (Apiaceae) and CYP87D16 in *M. lanceolata* (Primulaceae), catalyze the same reaction: C-16α hydroxylation of β-amyrin. CYP716Y1 and CYP87D16 are the first examples of two structurally unrelated plant P450s that evolved independently

Table 3-6. The first 20 unigene hits from a tblastn search using CYP716Y1, CYP87D16, CYP72A67, CYP72A68v2, or CYP93E1 as a query sequence

Query	Subject	Amino acid identities (%)	E Value	Score (Bits)	RPKM		
					Root	Leaf	Petal
CYP716Y1	Unigene5427	48.62	7.00E-168	491	75.84	13.02	4.60
	Unigene3622	49.55	5.00E-161	473	83.67	14.50	1.30
	Unigene5914	40.62	5.00E-128	390	60.04	1.27	0.18
	Unigene33496	40.29	3.00E-120	367	0.00	29.27	0.56
	Unigene33576	39.67	2.00E-119	366	0.00	35.83	33.80
	Unigene11792	34.92	2.00E-93	295	6.38	0.26	0.18
	Unigene33696	37.14	1.00E-87	277	0.00	1.58	3.36
	Unigene4921	34.31	1.00E-70	239	245.19	0.77	1.75
	Unigene28297	51.31	2.00E-68	221	0.25	3.31	1.84
	Unigene17820	32.33	2.00E-63	218	6.73	26.77	0.10
	Unigene164	26.95	2.00E-48	181	6.75	21.95	43.29
	Unigene16193	30.52	6.00E-48	173	8.81	17.96	1.55
	Unigene4695	29.32	1.00E-45	167	22.30	0.00	0.12
	Unigene35540	36.04	2.00E-41	149	0.00	1.76	2.28
	Unigene27222	25.13	2.00E-35	137	0.71	0.10	7.34
	Unigene5206	23.64	4.00E-35	137	27.70	18.38	4.88
	Unigene10100	22.12	9.00E-27	114	11.18	2.48	1.55
	Unigene6979	25.32	3.00E-26	109	74.11	28.88	11.75
	Unigene37898	55.13	1.00E-24	99.4	0.00	0.54	8.14
	Unigene21694	25.42	2.00E-21	97.8	3.45	2.41	37.04
CYP87D16	Unigene164	32.82	4.00E-73	253	6.75	21.95	43.29
	Unigene4695	32.67	2.00E-72	240	22.30	0.00	0.12
	Unigene17820	29.61	2.00E-52	188	6.73	26.77	0.10
	Unigene4921	28.54	2.00E-52	188	245.19	0.77	1.75
	Unigene5914	28.31	4.00E-48	176	60.04	1.27	0.18
	Unigene33576	25.05	4.00E-47	173	0.00	35.83	33.80
	Unigene5206	27.40	9.00E-47	171	27.70	18.38	4.88
	Unigene33496	24.78	8.00E-45	166	0.00	29.27	0.56
	Unigene5427	25.46	6.00E-42	158	75.84	13.02	4.60
	Unigene3622	25.98	3.00E-39	150	83.67	14.50	1.30
	Unigene27222	23.20	5.00E-39	147	0.71	0.10	7.34
	Unigene11792	24.94	9.00E-38	145	6.38	0.26	0.18
	Unigene16193	28.14	5.00E-37	142	8.81	17.96	1.55
	Unigene19235	42.57	3.00E-34	129	2.64	0.83	1.52
	Unigene33696	24.01	7.00E-30	120	0.00	1.58	3.36
	Unigene28297	33.73	4.00E-27	108	0.25	3.31	1.84
	Unigene10100	30.15	5.00E-21	96.7	11.18	2.48	1.55
	Unigene14133	37.69	3.00E-19	90.5	15.60	20.35	65.20
	Unigene2402	37.25	3.00E-17	82.4	46.35	28.10	81.19
	Unigene24107	28.94	3.00E-16	81.3	0.16	36.82	150.71
CYP72A67	Unigene2726	54.07	1.00E-159	465	128.42	60.12	29.50
	Unigene2754	55.13	8.00E-156	455	27.50	20.59	12.04
	Unigene3486	56.57	3.00E-119	356	47.70	110.07	77.91
	Unigene6261	40.08	7.00E-113	352	5.90	0.33	0.27
	Unigene3928	55.87	9.00E-112	337	85.74	43.54	0.30
	Unigene9145	37.92	1.00E-110	344	7.99	2.88	1.11
	Unigene7391	55.28	2.00E-109	331	11.38	18.66	11.01
	Unigene4446	39.32	2.00E-100	313	19.90	18.17	2.56
	Unigene16205	33.84	5.00E-97	310	10.44	0.07	0.00
	Unigene9271	34.86	3.00E-88	289	37.63	2.76	0.48
	Unigene6123	41.54	7.00E-86	274	51.07	34.79	18.33
	Unigene11339	34.89	6.00E-85	277	0.08	0.00	24.07
	Unigene2725	50.58	1.00E-81	258	1.22	0.16	17.09
	Unigene954	31.39	5.00E-81	268	218.96	2.92	0.68
	Unigene2720	32.28	1.00E-78	261	47.79	6.39	0.86
	Unigene12311	38.48	1.00E-73	240	2.06	4.71	0.59
	Unigene82	57.58	2.00E-72	242	4.96	128.75	3.54
	Unigene27128	51.27	1.00E-62	206	0.23	31.89	7.97
	Unigene7262	39.00	3.00E-61	209	0.76	0.53	71.26
	Unigene22507	57.24	1.00E-53	180	0.34	17.83	0.98
CYP72A68v2	Unigene2726	53.35	2.00E-164	477	128.42	60.12	29.50
	Unigene2754	53.44	2.00E-158	461	27.50	20.59	12.04
	Unigene3486	54.88	3.00E-116	349	47.70	110.07	77.91
	Unigene6261	38.63	4.00E-116	360	5.90	0.33	0.27
	Unigene9145	37.45	1.00E-110	344	7.99	2.88	1.11
	Unigene3928	54.45	1.00E-108	329	85.74	43.54	0.30
	Unigene7391	54.23	5.00E-106	322	11.38	18.66	11.01
	Unigene16205	34.89	1.00E-102	324	10.44	0.07	0.00
	Unigene4446	39.12	3.00E-101	315	19.90	18.17	2.56
	Unigene6123	41.05	4.00E-85	272	51.07	34.79	18.33
	Unigene954	30.17	5.00E-85	278	218.96	2.92	0.68
	Unigene9271	32.40	6.00E-84	278	37.63	2.76	0.48

Table 3-6. (continued)

Query	Subject	Amino acid identities (%)	E Value	Score (Bits)	RPKM		
					Root	Leaf	Petal
CYP72A68v2	Unigene82	54.22	5.00E-80	263	4.96	128.75	3.54
	Unigene2725	49.43	8.00E-80	254	1.22	0.16	17.09
	Unigene2720	28.99	1.00E-79	263	47.79	6.39	0.86
	Unigene11339	30.56	3.00E-77	257	0.08	0.00	24.07
	Unigene12311	36.44	2.00E-68	226	2.06	4.71	0.59
	Unigene27128	50.48	9.00E-67	218	0.23	31.89	7.97
	Unigene7262	41.74	3.00E-61	209	0.76	0.53	71.26
	Unigene22507	57.93	1.00E-53	180	0.34	17.83	0.98
CYP93E1	Unigene24107	43.48	1.00E-137	416	0.16	36.82	150.71
	Unigene3332	40.53	6.00E-116	357	44.37	2.62	0.55
	Unigene39822	35.32	1.00E-100	317	0.00	0.00	10.04
	Unigene15394	37.04	2.00E-100	319	12.79	47.54	55.29
	Unigene25782	35.80	1.00E-98	314	0.36	4.02	129.13
	Unigene26455	35.56	7.00E-97	309	0.00	0.00	87.02
	Unigene16382	33.47	2.00E-96	307	14.19	2.50	0.00
	Unigene13210	35.42	1.00E-95	304	1.24	36.56	97.14
	Unigene21615	37.47	2.00E-95	309	2.63	113.98	5.77
	Unigene14904	37.21	2.00E-92	296	12.81	3.19	2.26
	Unigene5559	34.90	2.00E-92	296	35.22	0.22	2.08
	Unigene6944	33.94	5.00E-90	289	33.86	0.95	0.57
	Unigene9193	36.40	2.00E-89	290	2.11	3.24	19.71
	Unigene29345	32.75	7.00E-88	286	0.07	41.50	0.00
	Unigene5015	32.73	8.00E-87	281	87.56	7.65	58.38
	Unigene10284	32.64	4.00E-86	278	155.06	45.56	22.94
	Unigene9476	34.12	7.00E-85	273	0.11	5.47	0.13
	Unigene9559	34.33	4.00E-84	280	16.21	31.27	11.80
	Unigene2099	33.06	4.00E-83	280	12.51	8.46	13.25
	Unigene33688	32.70	5.00E-83	273	0.00	16.42	0.27

with the same properties in different plant lineages (Moses *et al.*, 2015a). The identification of C-2β, C-23, and C-24 oxidases in *P. grandiflorus* may lead to the discovery of other examples of convergent evolution in triterpene oxidases in different plant species.

I focused on P450s that are highly expressed in the roots of *P. grandiflorus* and on similarities to triterpene oxidase families, and identified two triterpene oxidases. CYP716A140v2 catalyzes the three-step oxidation at C-28 of β-amyrin, while CYP716A141 mainly catalyzes the hydroxylation at C-16β of β-amyrin. Interestingly, CYP716A141 also has negligible C-28 oxidation activity. This may be a biochemical trace of the ancestral enzyme activity, suggesting that this specialized C-16β oxidase is derived from a more highly conserved C-28 oxidase. Recently, the C-16β hydroxylation of β-amyrin by CYP716A111 was shown in *Aquilegia coerulea* (Ranunculaceae) (Miettinen *et al.*, 2017); however, CYP716A111 and CYP716A141 cluster in separate groups (Figure 3-3). C-16β hydroxylated oleanene-type triterpenoids have been reported from *Endopappus macrocarpus* (Asteraceae) (Boutaghane *et al.*, 2013), *Calendula officinalis* (Asteraceae) (Yoshikawa *et al.*, 2001), *Silphium asteriscus* (Asteraceae) (Masullo *et al.*, 2014), and *Phyllanthus hirsutus* (Phyllanthaceae) (Thang *et al.*, 2017), as well as *B. falcatum*. Further identification of C-16β

hydroxylases from these plants may provide important insight into the evolution of C-16β hydroxylases in eudicots.

In Chapters 2 and 3, two medicinal plants belonging to different plant families were analyzed for the mining of P450s involved in triterpenoid biosynthesis. CYP716A subfamily enzymes catalyzed the oxidation of β-amyrin in both plants, while CYP72A subfamily enzymes catalyzed the oxidation of β-amyrin only in *G. uralensis* (Fabaceae). These results further confirm that the evolution of CYP716A subfamily enzymes for triterpenoid biosynthesis is a shared character among eudicots, while CYP72A subfamily enzymes are specifically involved in triterpenoid biosynthesis in legumes.

Chapter 4

Identification of a transcription factor regulating triterpenoid biosynthesis in *Glycyrrhiza uralensis*

4-1. Introduction

As mentioned in Chapter 2, *Glycyrrhiza uralensis* (licorice) is an important medicinal plant that produces various triterpenoids including glycyrrhizin, soyasaponins, betulinic acid, and oleanolic acid. Oxidosqualene cyclases (OSCs) and cytochrome P450 monooxygenases (P450s) involved in the biosynthesis of major triterpenoids in licorice have been elucidated in previous studies and the results presented in Chapter 2 (Figure 2-2). The accumulation patterns of these triterpenoids in licorice plants or in tissue cultures differ (Hayashi *et al.*, 1988, 1993; Kojoma *et al.*, 2010). Moreover, as shown in Chapter 2, intact roots and tissue-cultured stolons of *G. uralensis* accumulated different triterpenoids, and the expression patterns of these biosynthetic genes were largely correlated with metabolite accumulation.

These observations imply that certain mechanisms coordinately regulate triterpenoid biosynthetic genes in licorice. Generally, biosynthesis of plant specialized metabolites is temporally and spatially regulated. This can often be achieved by regulating biosynthetic genes at the transcriptional level. Key elements of transcriptional regulation are transcription factors (TFs) that bind to promoter regions of target genes or form complexes with other DNA-binding proteins to modulate gene expression (Yang *et al.*, 2012). There are several reports on TFs involved in specialized metabolites derived from sterol precursors (biosynthesized via protosteryl cation), such as basic helix-loop-helix (bHLH) TFs regulating cucurbitacin biosynthesis in the cucumber (Shang *et al.*, 2014), ethylene response factor (ERF) TFs regulating steroidal glycoalkaloid biosynthesis in the potato and the tomato (Cárdenas *et al.*, 2016; Thagun *et al.*, 2016), and a WRKY TF regulating the biosynthesis of withanolides in *Withania somnifera*, which are thought to be derived from primary sterols (Singh *et al.*, 2017). However, reports on TFs regulating triterpenoid biosynthesis are very limited. Recently, two bHLH TFs, TSAR1 and TSAR2 from the model legume *Medicago truncatula*,

were shown to regulate the biosynthesis of non-hemolytic triterpenoid saponins (soyasaponins; possess a hydroxyl group at the C-24 position of β-amyrin) and hemolytic triterpenoid saponins (possess a carboxyl group at the C-28 position of β-amyrin), respectively (Mertens *et al.*, 2016). In licorice, biosynthetic genes for major triterpenoids have been elucidated, and these genes are differentially expressed (Seki *et al.*, 2008, 2011; Ramilowski *et al.*, 2013; Chapter 2). However, no TFs involved in the regulation of triterpenoid biosynthesis have yet been reported in licorice.

The results presented in Chapter 2 suggested that genes involved in each triterpenoid pathway are transcriptionally co-regulated for the effective conversion of precursors into the final products by their cognate TFs. In this chapter, I sought to identify the TFs that regulate each triterpenoid pathway in *G. uralensis*. As a result, I identified a bHLH TF, GubHLH3, which positively regulates soyasaponin biosynthesis. Overexpression of *GubHLH3* in transgenic hairy roots of *G. uralensis* enhanced the expression levels of *β-amyrin synthase* (*bAS*), *CYP93E3*, and *CYP72A566*, suggesting that GubHLH3 coordinately regulates the soyasaponin biosynthetic genes. Moreover, accumulation of soyasapogenol B and its biosynthetic intermediate, sophoradiol (22β-hydroxy-β-amyrin), were enhanced in transgenic hairy roots overexpressing *GubHLH3*. The regulatory mechanisms of triterpenoid biosynthesis in legumes are compared and discussed.

4-2. Materials and methods

4-2-1. Plant materials

Tissue-cultured stolons of *G. uralensis* (Hokkaido-iryodai strain) were maintained as described in Chapter 2. Leaves of *G. uralensis* strain 308-19 (Mochida *et al.*, 2017) were used for genomic DNA extraction. Roots of *G. uralensis* strain 308-19 (2-year-old) harvested in June 2011 (Ramilowski *et al.*, 2013) were used for RNA extraction. *G. uralensis* (Hokkaido-iryodai strain) seeds were collected in 2007. Suspension cultures of tobacco BY-2 cells (*Nicotiana tabacum* L. cultivar Bright Yellow 2; a kind gift from Dr. Ryo Misaki, Osaka University, Suita, Japan) were maintained at 26°C with rotary shaking at 125 rpm in 100 mL of Murashige and Skoog plant salt mixture (Wako Pure Chemical Industries, Osaka, Japan) supplemented with 3% sucrose, 0.2 g/L potassium dihydrogenphosphate,

Table 4-1. Primers used in this chapter

Primer no.	Sequence (5' to 3')	Comment
1	CACCATGGAAGACGCCAAAAACAT	Construction of pAM-PAT-LUC
2	TTACACGGCGATCTTTCCGC	Construction of pAM-PAT-LUC
3	CACAGGGCTTGACCTGGAAGCGGTTGT	5'-RACE PCR, *bAS*, first; Genome walking
4	GACGAGCGGCATCAACCTGAGCTCTCT	5'-RACE PCR, *bAS*, nested; Genome walking
5	GATGGATTGCAAAATTACCAACCGTGAG	Genome walking
6	GTGTCTTGAGATTGATCCGGTGCTGAAC	Genome walking
7	CACCACCCCGGTCTAGACAAAAAATACT	Cloning of *bASpro*; Genome walking
8	TTCCGCTATCTTCAGCCTCCACAT	Cloning of *bASpro*; Genome walking
9	CATGCTGGACAATTCTTCCAACGAATTG	5'-RACE PCR, *CYP88D6*, first; Genome walking
10	ATTGCTAGCGCCTTGTGACCCACGATT	5'-RACE PCR, *CYP88D6*, nested; Genome walking
11	CTGTTATGCCAAATAGGGCCTAGATAG	Genome walking
12	GATTTGTTTCCGCAATCACCGCCATAGA	Genome walking
13	CACCCTGTAGTATATAACAAAAGAGAAATG	Cloning of *CYP88D6pro*; Genome walking
14	CATTTTTCTGAATCTGTTTCGTTGTTC	Genome walking
15	AGCGTCTTCCCACTAAGAAGCTCTGTC	5'-RACE PCR, *CYP93E3*, first; Genome walking
16	GAGGACGCAACGATGACATGTTGCGAA	5'-RACE PCR, *CYP93E3*, nested; Genome walking
17	CTGTGCTGAGGGTTTCAATCATGGGAG	Genome walking
18	TGTCACGGTGTGTGCGTCTTCTCCCAA	Genome walking
19	GGTCGAGCTTTAGGAGCCGAGGTGTAC	Genome walking
20	TCGTGGCAGGTCTCTTCCTCGTGATGC	Genome walking
21	CACCTTAATTGAAAGCTAAATTTCCAAGTTC	Cloning of *CYP93E3pro*; Genome walking
22	CATGGTTGAGTTGAGGGAAAACAG	Genome walking
23	GGACATGCAAACCCAATGTACTTCCAT	Cloning of *CYP88D6pro*
24	GTAGCCTTGGATGTCAAGCAT	Cloning of *CYP93E3pro*; *CYP93E3pro*, deletion 3
25	CACCACATATGCTGGCCTCCCCACTACAC	Cloning of *CASpro*
26	CTCCGCAATCTTGAGCTTCCACAT	Cloning of *CASpro*
27	CACCAAAATACTGCTTGCTTATGCGTTCTATTCATC	Cloning of *LUSpro*
28	TCCTCCTTCTCCTATCTTCAGCTTCCACAT	Cloning of *LUSpro*
29	CACCGTGATATCCGCGGGCCATGGATATT	Cloning of *CYP72A154pro*
30	CCCTGGTGTGGAAGATGCATCCAT	Cloning of *CYP72A154pro*
31	CACCTTGATGCTCATGTCATGTGTGAGG	Cloning of *CYP716A179pro*
32	AAGGGACATGTAGAAATGCTCCAT	Cloning of *CYP716A179pro*
33	GCTGTGTTTGTGAGTGAAGCGGTT	5'-RACE PCR, *CAS*, first
34	TCCAAGCTTAGGATCGAACTCCCACACC	5'-RACE PCR, *CAS*, nested
35	GATGGGCAAGGGAGAAGATTCGTGGTG	5'-RACE PCR, *CYP72A154*, first
36	GCCTCAGCCACAGCGTGTTCACCA	5'-RACE PCR, *CYP72A154*, nested
37	TGAGTTCCGAGGAGTACCGGACCATCCG	5'-RACE PCR, *CYP716A179*, first
38	GGGTAACCCATCTTTCCCGGTGGCAGG	5'-RACE PCR, *CYP716A179*, nested
39	GTTTTTGGTGAATTCCCGGCGTAGCCT	5'-RACE PCR, *LUS*, first
40	GGAGTTCCTGCATTTGGATCGAACTCCC	5'-RACE PCR, *LUS*, nested
41	atagaaaagttgACCCCGGTCTAGACAAAAAA	*bASpro-#*1
42	gtacaaacttggAAAATTAATTCAATCATTTTTTTCTGAACAGA	*bASpro-#*1
43	atagaaaagttgTCAGAAAAAAATGATTGAATTAATTTTGTTGT	*bASpro-#2*
44	gtacaaacttggTAAATCTTTTTTATTACATATGCTGACAACT	*bASpro-#2*
45	atagaaaagttgCTCTTCAATGTGTAAAGGCATCTAT	*bASpro-#*3
46	gtacaaacttggGATTTGGAGAGTGACTCGTGAAA	*bASpro-#*3
47	atagaaaagttgCATAAACAATAAATAAGAATGCTATAAGTAGGG	*bASpro-#*4
48	gtacaaacttggCTTCCTTCCTCTTCTGATTCTTCTT	*bASpro-#*4
49	atagaaaagttgCTGTAGTATATAACAAAAGAGAAATGTT	*CYP88D6pro-#*1
50	gtacaaacttggTACTATTTTGATAAATAAATAAAAGTGTGC	*CYP88D6pro-#*1
51	atagaaaagttgGGGGCTTTTTTATCCCATTAGC	*CYP88D6pro-#2*
52	gtacaaacttggAGCTGAGATGCAACACTAAGAA	*CYP88D6pro-#2*
53	atagaaaagttgGTGGACTAAAACAGTGAACTAAATG	*CYP88D6pro-#*3
54	gtacaaacttggTAGAGAGGCTTTTTTACCCCAT	*CYP88D6pro-#*3
55	atagaaaagttgTGATAAATGAAAAAAAAAGGTGTTATTTTGA	*CYP88D6pro-#*4
56	gtacaaacttggTTTTCTGAATCTGTTTCGTTGTTC	*CYP88D6pro-#*4
57	atagaaaagttgTTAATTGAAAGCTAAATTTCCAAGTTCA	*CYP93E3pro-#*1
58	gtacaaacttggCTTTGAAAGAAAACCCTCTTATAAGTTGAT	*CYP93E3pro-#*1
59	atagaaaagttgTTTTATCATGATAAAAAGCAAGTGAAA	*CYP93E3pro-#2*
60	gtacaaacttggGCATATACTATTTTCATAATTAAATTTTCTCA	*CYP93E3pro-#2*
61	atagaaaagttgGTAGATATTTGAGAAAATTTAATTATGAAAAT	*CYP93E3pro-#*3
62	gtacaaacttggCTTTTAATATTGGTGGCGGACA	*CYP93E3pro-#*3
63	atagaaaagttgTGATCCTTATAAGAGCTGTCCG	*CYP93E3pro-#*4
64	gtacaaacttggGGTTGAGTTGAGGGAAAACAGA	*CYP93E3pro-#*4
65	ggggacaactttgtatagaaaagttg	*att*B4 adapter
66	ggggactgctttttgtacaaacttgg	*att*B1r adapter
67	CCTTCACCTATAAGAGGGTTTTCTTT	*CYP93E3pro*, deletion 1
68	TCTTATAGGTGAAGGGGGCGGCCGC	*CYP93E3pro*, deletion 1
69	CCTTCACCGTTTGGAAGAGTTTATGT	*CYP93E3pro*, deletion 2
70	TCCAAACGGTGAAGGGGGCGGCCGC	*CYP93E3pro*, deletion 2
71	CACCCCTAGAATGATCCTTATAAGA	*CYP93E3pro*, deletion 3
72	CGGAAtgccaaTCACTTTGTTTTCAATAA	*CYP93E3pro*, mE1
73	GTGAttggcaTTCCGCGTCCACTGTCAC	*CYP93E3pro*, mE1
74	GCAGtgaaaaGAACACGGTACCTTGTAT	*CYP93E3pro*, mN1
75	TGTTCttttcaCTGCATATATTATATAAT	*CYP93E3pro*, mN1
76	GTCCtgaattGTTCGCAAAAATTAGTAA	*CYP93E3pro*, mN2
77	CGAACaattcaGGACAATATTGTATTATG	*CYP93E3pro*, mN2

Table 4-1. (continued)

Primer no.	Sequence (5' to 3')	Comment
78	CACCATGACAGATTACCGGTCACCACCAAC	Cloning of *GubHLH1*
79	TCGTGCATCGCCAACTTTGGAGGA	Cloning of *GubHLH1*
80	CACCATGAATCTGTGGAGCGATGA	Cloning of *GubHLH2*
81	GGGAACATCCCCTACTTTAGAC	Cloning of *GubHLH2*
82	CACCATGGAGGAAATGAACACACCAATGA	Cloning of *GubHLH3*
83	TGACATAGACTTCAATGTTGCCATGCG	Cloning of *GubHLH3*
84	CACCATGGAGGAAATGAACACACCAATAA	Dual expression of *GubHLH3*(-*SRDX*) and *GFP*
85	GGCGCGCCGAATTCCTTATC	Dual expression of *GubHLH3*(-*SRDX*) and *GFP*
86	TCTTCGCAAACTGGCAGTGA	qPCR for *β-tubulin*
87	CGAGATGTGAGTGGGGCAAA	qPCR for *β-tubulin*
88	GGTGGTTTATCAGCGTGGGA	qPCR for *bAS*
89	TGCTCAACTACAATGTCCGCA	qPCR for *bAS*
90	CCGGCTGAGACTTTTGGTGA	qPCR for *CAS*
91	TCTCGACGATGCCCAGGATA	qPCR for *CAS*
92	GAAAGCAATACCCACAGCACAG	qPCR for *LUS*
93	GCAAATTCCCCAACACCCATAC	qPCR for *LUS*
94	ATGGACGAAAATTGGAGGACGA	qPCR for *CYP88D6*
95	CTGGTTGCTGTACTTTCATGGC	qPCR for *CYP88D6*
96	GACGCCTCATTACTTCCCCAA	qPCR for *CYP88D6* (Figure 4-15 only)
97	TTCAAGAGCTCAACGGGGTG	qPCR for *CYP88D6* (Figure 4-15 only)
98	ATGGCGACCCTTACAAGCTC	qPCR for *CYP72A154*
99	AGATTCGTGGTGCAGCATCA	qPCR for *CYP72A154*
100	CCAAGCACTCTACAAACTCTCCA	qPCR for *CYP93E3*
101	GGATTTGCTTAGCCATTTCTGCA	qPCR for *CYP93E3*
102	GCCGCCGATTACTTCTTCATTC	qPCR for *CYP93E3* (Figure 4-15 only)
103	CGCGGATATTGACAAAGTGCTC	qPCR for *CYP93E3* (Figure 4-15 only)
104	CAGGGGAGCCTAGTAACAATGAC	qPCR for *CYP72A566*
105	GCCCTGCCAAGTAAAATAGCTTC	qPCR for *CYP72A566*
106	AGCGGTTCAAGTGGGAGATG	qPCR for *CYP716A179*
107	ATCGGGAGGTCATTTGCAGG	qPCR for *CYP716A179*
108	TTGAGCTCGCTTCCACTGAG	qPCR for *GubHLH1*
109	TCGCTCTGGTCGGTTGTAAC	qPCR for *GubHLH1*
110	CTAATCGAGGGAGCAAGAGAGA	qPCR for *GubHLH2*
111	CTTCGTCGCCTTTGTAGTAACC	qPCR for *GubHLH2*
112	TGCTGATGACGAAGTCGAAG	qPCR for *GubHLH3*
113	ACTGGTCTCATCGGTGAAGG	qPCR for *GubHLH3*

The underlined sequences were added to facilitate unidirectional cloning of the product into pENTR/D-TOPO (Thermo Fisher Scientific).

1 mg/L thiamine hydrochloride, 0.1 g/L *myo*-inositol, and 0.2 mg/L 2,4-dichlorophenoxyacetic acid, and subcultured every 7 days.

4-2-2. Chemicals

β-Amyrin, α-amyrin, lupeol, oleanolic acid, ursolic acid, glycyrrhetinic acid (18β-glycyrrhetinic acid), echinocystic acid were purchased from Extrasynthese. Soyasapogenol B was purchased from Tokiwa Phytochemical. Sophoradiol, 24-hydroxy-β-amyrin, 11-oxo-β-amyrin, 30-hydroxy-β-amyrin, and 11-deoxoglycyrrhetinic acid were kindly gifted by Dr. Kiyoshi Ohyama (Tokyo Institute of Technology). Betulinic acid and gibberellin A_3 (GA_3) were purchased from Tokyo Chemical Industry. Methyl jasmonate (MeJA), 1-naphthaleneacetic acid (NAA), and salicylic acid (SA) were purchased from Sigma-Aldrich.

4-2-3. Vector construction

pGEM-GW-GUS-NOS

A GUS reporter gene-NOS terminator fusion fragment was excised from pBI121 (Clontech/Takara Bio) with SmaI and EcoRI, and ligated into the SmaI and EcoRI sites of pGEM-7Zf(+) (Promega, Madison, WI, USA) to obtain pGEM-GUS-NOS. Then, Reading Frame Cassette A equipped with Gateway Vector Conversion System (Thermo Fisher Scientific) was inserted into SmaI-digested pGEM-GUS-NOS to obtain pGEM-GW-GUS-NOS.

pAM-PAT-LUC

luc+ (modified firefly luciferase) was amplified from pGL3-basic vector (Promega) with primers 1 and 2 (Table 4-1) and cloned into pENTR/D-TOPO (Thermo Fisher Scientific). Then, *luc*+ was transferred into the pAM-PAT-GW destination vector (a generous gift from Dr. Bekir Ülker, Max Planck Institute for Plant Breeding Research, Cologne, Germany) using the Gateway LR Clonase II Enzyme mix (Thermo Fisher Scientific) to obtain pAM-PAT-LUC.

4-2-4. Isolation of genomic DNA and total RNA from plant tissues

Genomic DNA was isolated using Nucleon Phytopure Genomic DNA Extraction Kits (GE Healthcare, Little Chalfont, UK) according to the manufacturer's instructions from frozen plant tissues (1 g of leaves). Total RNA was extracted as described in Chapter 2 from frozen plant tissues (1 g of roots, 0.1–0.5 g of tissue-cultured stolons treated with plant hormons or elicitors, and 0.07–0.3 g of hairy roots).

4-2-5. Cloning of promoter regions of *bAS*, *CYP88D6*, and *CYP93E3* by PCR-based genome walking

Cloning of the promoter regions of *bAS*, *CYP88D6*, and *CYP93E3* was conducted by a PCR-based genome walking method using a GenomeWalker Universal Kit (Clontech/Takara Bio). Genomic DNA (2.5 µg) isolated from tissue-cultured stolons of *G. uralensis* (Hokkaido-iryodai strain) was digested with a single blunt-cutting restriction endonuclease, DraI, EcoRV, PvuII, StuI, or SspI. Digested samples were used in separate ligation reactions with the GenomeWalker adapters provided with the

kit. This process created five libraries for use in several rounds of PCR-based genome walking. A primary PCR amplification of each library using a gene-specific primer (GSP1) and an adapter sequence-specific primer (AP1, provided in the kit) was followed by a secondary reaction in which a nested gene-specific primer (GSP2) and a nested adapter primer (AP2, provided in the kit) were used to amplify the diluted primary PCR reaction product.

The GSPs used in *bAS* promoter cloning were 3 (GSP1) and 4 (GSP2) (Table 4-1) for the first round of genome walking. An approximately 0.7-kb DNA fragment amplified from adapter-ligated DraI-digested DNA was cloned and sequenced. Based on the sequence data obtained, new GSPs, 5 and 6 (Table 4-1), were designed for a second round of walking. An approximately 1.5-kb DNA fragment amplified from adapter-ligated EcoRV-digested DNA was cloned and sequenced. Based on the sequence data obtained, primers 7 and 8 (Table 4-1) were designed to amplify a 1.8-kb promoter fragment of *bAS*.

To clone the *CYP88D6* gene promoter fragment, primers 9 and 10 (Table 4-1) were used for the first round of genome walking. An approximately 1.4-kb DNA fragment amplified from adapter-ligated DraI-digested DNA was cloned and sequenced. Based on the sequence data obtained, new GSPs, 11 and 12 (Table 4-1), were designed for a second round of walking. A 1.4-kb DNA fragment amplified from adapter-ligated PvuII-digested DNA was cloned and sequenced. Based on the sequence data obtained, primers 13 and 14 (Table 4-1) were designed to amplify a 1.8-kb promoter fragment of *CYP88D6*.

To clone the *CYP93E3* gene promoter fragment, primers 15 and 16 (Table 4-1) were used for the first round of genome walking. An approximately 0.8-kb fragment amplified from adapter-ligated DraI-digested DNA was cloned and sequenced. Based on the sequence data obtained, new GSPs, 17 and 18 (Table 4-1), were designed for a second round of walking. A 0.9-kb DNA fragment amplified from adapter-ligated EcoRV-digested DNA was cloned and sequenced. Based on the sequence data obtained, primers 19 and 20 (Table 4-1) were designed for a third round of genome walking. A 1.0-kb fragment amplified from adapter-ligated PvuII-digested DNA was cloned and sequenced. Based on the sequence data obtained, primers 21 and 22 (Table 4-1) were designed to amplify a 2-kb promoter fragment of *CYP93E3*.

4-2-6. Isolation of promoter regions

Based on the promoter sequences obtained by PCR-based genome walking, I isolated promoter regions of *bAS*, *CYP88D6*, and *CYP93E3* from the genomic DNA of strain 308-19 with primers 7, 8, 13, 21, 23, and 24 (Table 4-1) by PCR. The promoter regions of *cycloartenol synthase* (*CAS*), *lupeol synthase* (*LUS*), *CYP72A154*, and *CYP716A179* were isolated from genomic DNA of strain 308-19 by PCR with primers 25–32 (Table 4-1) designed from the genomic information of strain 308-19 (Mochida *et al.*, 2017). All of the amplified promoter regions were cloned into pENTR/D-TOPO (Thermo Fisher Scientific) to make entry clones.

4-2-7. Determination of the transcriptional start sites

The transcriptional start sites were determined by 5′-rapid amplification of cDNA ends (RACE) PCR method using the SMARTer RACE cDNA Amplification Kit (Clontech/Takara Bio) with primers 3, 4, 9, 10, 15, 16, and 33–40 (Table 4-1). First-strand cDNA was synthesized from 1 μg of total RNA according to the manufacturer's instructions. Total RNA obtained from the roots of strain 308-19 was used to determine the transcriptional start sites of *bAS*, *CAS*, *CYP88D6*, *CYP72A154*, and *CYP93E3*, and total RNA obtained from tissue-cultured stolons of the Hokkaido-iryodai strain was used for *LUS* and *CYP716A179*.

4-2-8. High-throughput yeast one-hybrid screening of *Arabidopsis* TF library

Fragments (approx. 500 bp) of promoter regions of *bAS*, *CYP88D6*, and *CYP93E3* were amplified by primers 41–64 and adapter primers 65 and 66 (Table 4-1) and cloned into pDONRGm-P4P1R (Oshima *et al.*, 2011) using Gateway BP Clonase II Enzyme mix (Thermo Fisher Scientific) to make entry clones containing *att*L4 and *att*R1 sites, then transferred into R4L1pDEST_HISi (Mitsuda *et al.*, 2010; Figure 4-1a) using Gateway LR Clonase II Enzyme mix (Thermo Fisher Scientific). The obtained bait constructs were integrated into the *URA3* locus of yeast strain YM4271 by homologous recombination to make yeast one-hybrid (Y1H) yeast strains. Y1H screening was performed as previously described (Oda-Yamamizo *et al.*, 2016) with some modifications. Briefly, each Y1H strain harboring a bait construct was first transformed with pDEST_GAD424 (Mitsuda *et al.*, 2010;

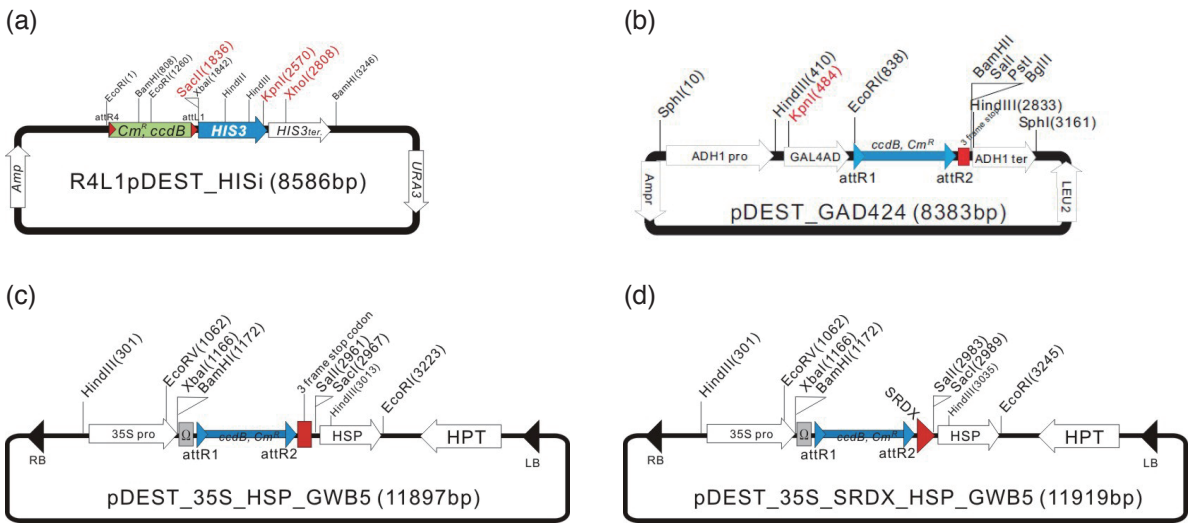


Figure 4-1. Maps of the vectors used in yeast one-hybrid assays and overexpression of transcription factors or their chimeric repressors: (a) R4L1pDEST_HISi, (b) pDEST_GAD424, (c) pDEST_35S_HSP_GWB5, and (d) pDEST_35S_SRDX_HSP_GWB5. These vectors and maps were kindly gifted by Dr. Nobutaka Mitsuda, National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan. The meanings of key components in the maps are as follows: 35S pro, cauliflower mosaic virus (CaMV) *35S* promoter; GAL4AD, GAL4 activation domain; *HIS3*, yeast *HIS3* gene downstream of the minimal promoter of the *HIS3* locus containing TATA boxes; HSP, heat shock protein (HSP) terminator; LB, left border; RB, right border; SRDX, modified plant-specific EAR-motif repression domain (LDLDLELRLGFA; Hiratsu *et al.*, 2003); Ω, the translational enhancer sequence from tobacco mosaic virus.

Figure 4-1b) empty vector, and then spotted on the appropriate synthetic defined (SD) medium containing 0–100 mM 3-amino-1,2,4-triazole (3-AT) to determine the opitimal concentration of 3-AT to suppress background *HIS3* expression. Next, using SD medium containing the optimal concentration of 3-AT, each Y1H strain was transformed with the *Arabidopsis* TF library (cloned into pDEST_GAD424) prepared in 96-well plates, which covers 1,428 *Arabidopsis* TFs with 376 mini-pools (the library was prepared by the Plant Gene Regulation Research Group, Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology). Yeast cell growth was then observed for 1–2 weeks. To narrow down the TFs that interacted with the bait, Y1H strains were transformed individually with the TFs contained in the positive wells and observed. An outline of the Y1H screening procedure is shown in Figure 4-2.

4-2-9. Construction of reporter constructs for transient co-transfection assays

To produce promoter:*GUS* reporter constructs, each promoter region was transferred to pGEM-GW-GUS-NOS using Gateway LR Clonase II Enzyme mix (Thermo Fisher Scientific). To generate

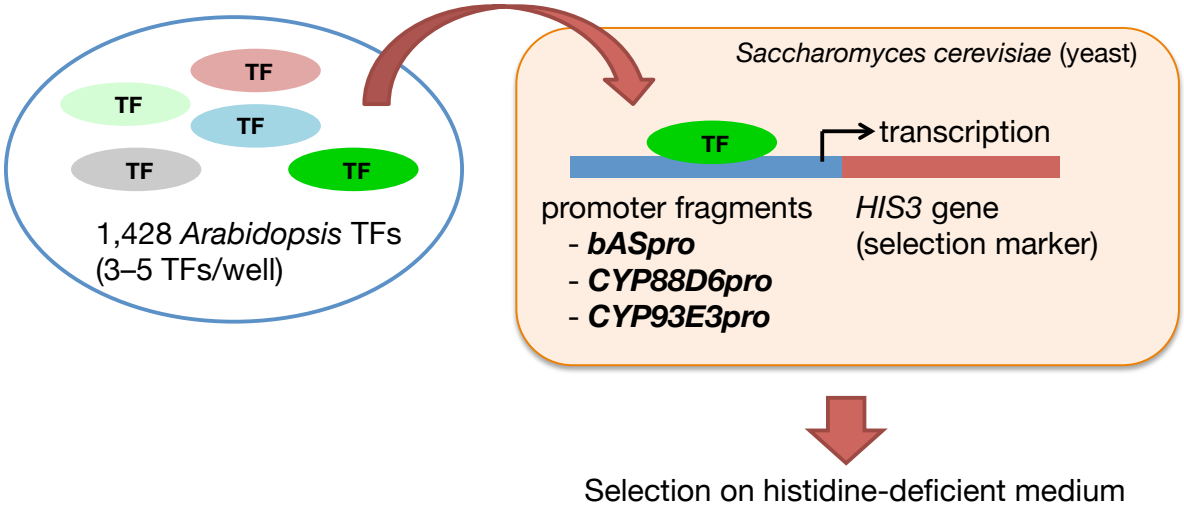


Figure 4-2. Outline of high-throughput yeast one-hybrid (Y1H) assays. Expression of the *HIS3* gene was driven by a promoter fragment of interest; this construct (bait) was integrated into the genome of yeast strain YM4271. A library of 1,428 *Arabidopsis* transcription factors (TFs) was covered by 376 mini-pools (three to five TFs/well); these TFs were transformed into yeast cells pre-transformed with the bait construct. After selection on histidine-deficient medium, TF binding to the integrated promoter sequences was determined by growth of the yeast cells on the selection medium.

deletion constructs for the *CYP93E3* promoter, primers 67–70 (Table 4-1) were used to obtain deletions 1 and 2 by inverse PCR, and primers 71 and 24 (Table 4-1) were used to obtain deletion 3 by cloning into pENTR/D-TOPO (Thermo Fisher Scientific). To introduce mutation at the E-box (E1) and N-box (N1 and N2), inverse PCR was performed using primers 72–77 (Table 4-1). The obtained deletion and mutation constructs were transferred to pGEM-GW-GUS-NOS using Gateway LR Clonase II Enzyme mix (Thermo Fisher Scientific).

4-2-10. Cloning of TF genes and construction of effector constructs for transient co-transfection assays

First-strand cDNA obtained from total RNA isolated from the roots of strain 308-19 was used as a PCR template. The full-length coding sequences (CDSs) without stop codons of *GubHLH1*, *GubHLH2*, and *GubHLH3* were amplified by PCR with primers 78–83 (Table 4-1) and cloned into pENTR/D-TOPO (Thermo Fisher Scientific) to make entry clones. Each CDS was transferred into pDEST_35S_HSP_GWB5 (Fujiwara *et al.*, 2014; Figure 4-1c) using Gateway LR Clonase II Enzyme mix (Thermo Fisher Scientific) to make GubHLH1–3/pDEST_35S_HSP_GWB5 constructs, and used as effector constructs for transient co-transfection assays. The nucleotide sequences isolated in this

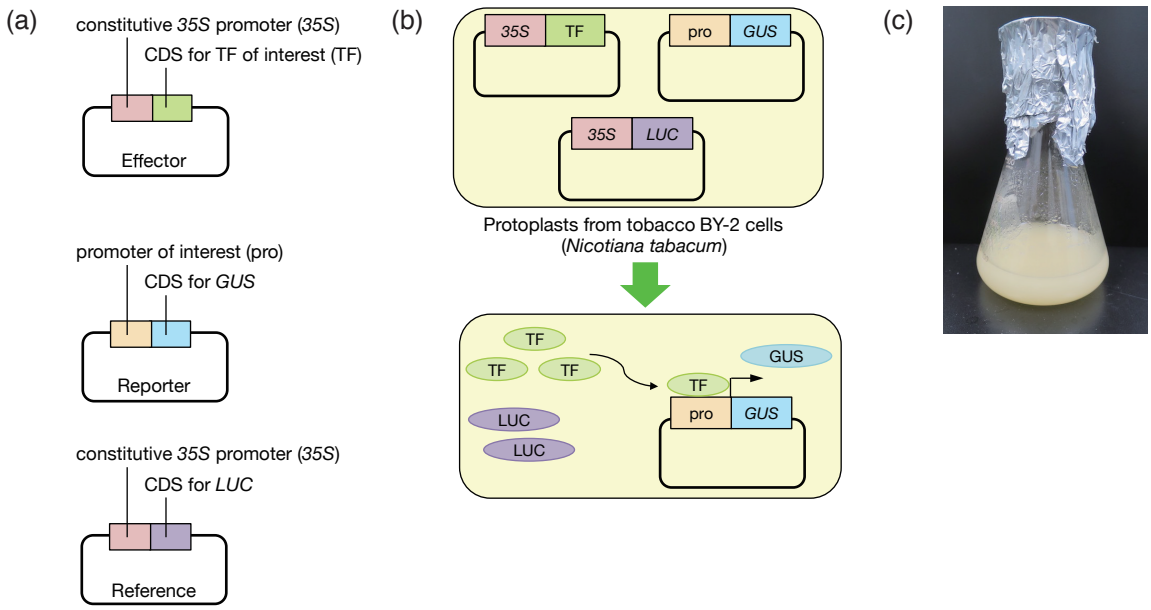


Figure 4-3. Outline of the transient co-transfection assays using protoplasts isolated from tobacco BY-2 cells.
(a) The three constructs used in the transient co-transfection assays. A reference construct (*35S*:*LUC*) was used as an internal control to normalize the transfection efficiency.
(b) Illustrated mechanisms of observed transactivation. The three constructs were introduced into protoplasts isolated from tobacco BY-2 cells by polyethylene glycol (PEG)-mediated transient transformation. A TF of interest was expressed from the *35S* promoter. TF binding to and transactivation of a promoter of interest leads to *GUS* expression, which is determined by the level of GUS activity measured.
(c) Tobacco BY-2 cells at 7 days after the previous subculture.

chapter have been submitted to the DDBJ under the accession numbers LC318131 (GubHLH1), LC318132 (GubHLH2), and LC318133 (GubHLH3), respectively.

4-2-11. Transient co-transfection assays

Transient co-transfection assays were performed using polyethylene glycol (PEG)-mediated transient transformation of protoplasts isolated from tobacco BY-2 cells as previously described (Sakamoto *et al.*, 2012), with some modifications. Briefly, 12.5 mL of BY-2 cells were collected in a 50-mL tube 7 days after subculturing and incubated with 20 mL of digestion buffer (Table 4-2) for 2 h at 25°C with reciprocal shaking at 80 rpm in the dark. After digestion, the protoplasts were filtered through a 70-µm nylon mesh cell strainer (Corning, Corning, NY, USA), washed three times with W5 buffer (Table 4-2), and then incubated in W5 buffer at 4°C for 30 min in the dark. After incubation, the protoplasts were collected and resuspended in MMg solution (Table 4-2) to prepare a suspension (3–5 × 10^5 cells/mL). For co-transfection assays, 35 µL of protoplast suspension, 45 µL of the PEG solution

Table 4-2. Buffers and reagents used in transient co-transfection assays

Buffer/Reagent	Composition	Reference
Digestion buffer	1% (w/v) cellulase "onozuka" R-10 (Yakulto, Tokyo, Japan), 0.1% (w/v) pectolyase Y-23 (Kyowa Kasei, Osaka, Japan), 0.1% (w/v) bovine serum albumin, 400 mM mannitol, 20 mM MES (pH 5.7), 20 mM potassium chloride, 10 mM calcium chloride, 5 mM 2-mercaptoethanol	Sakamoto *et al*., 2012
W5 buffer	150 mM sodium chloride, 125 mM calcium chloride, 5 mM potassium chloride, 2 mM MES (pH 5.7)	Sakamoto *et al*., 2012
MMg solution	400 mM mannitol, 15 mM magnesium chloride, 4 mM MES (pH 5.7)	Sakamoto *et al*., 2012
PEG solution	40% (w/v) PEG 4000 (Fluka/Sigma-Aldrich), 200 mM mannitol, 100 mM calcium chloride	Sakamoto *et al*., 2012
MUG substrate mix for GUS assay	10 mM Tris-HCl (pH 8.0), 1 mM 4-methylumbelliferyl-β-D-glucuronide (MUG; Sigma-Aldrich), 2 mM magnesium chloride	Yoo *et al*., 2007

(Table 4-2), and 10 μL of plasmid solution containing 1 μg of effector construct, 1 μg of reporter construct, and 0.5 μg of *35S*:*LUC* internal control (pAM-PAT-LUC) were mixed in a 96-well plate (Thermo Fisher Scientific) by vortexing at 900 rpm for 15 sec. After incubation at room temperature for 10 min in the dark, the mixture was suspended with 200 μL of W5 buffer, then centrifuged at 100 × *g* for 5 min. After washing three times with the W5 buffer, 200 μL of W5 buffer was added and the protoplasts were incubated at 25°C for 24 h in the dark. The incubated protoplasts were vortexed at 900 rpm for 15 sec, then 20 μL of Luciferase Cell Culture Lysis 5X Reagent (Promega) was added. The mixture was vortexed for another 15 sec and centrifuged at 100 × *g* for 1 min. The supernatant was used for luciferase (LUC) and β-glucuronidase (GUS) reporter assays. For LUC assays, 50 μL of Luciferase Assay Reagent (Promega) was added to 10 μL of supernatant, and LUC activities were determined by luminescence measurement using a Wallac 1420 multilabel counter (PerkinElmer, Waltham, MA, USA). For GUS assays, 100 μL of MUG substrate mix for GUS assay (Table 4-2) was added to 10 μL of the supernatant or Luciferase Cell Culture Lysis Reagent diluted with the W5 buffer as a blank. After incubation at 37°C for 90 min, GUS activities were measured with a Wallac 1420 multilabel counter (PerkinElmer) by excitation at 355 nm and measuring emission at 460 nm. The average value of the blanks (four wells) was subtracted from the measured GUS activities of the samples to calculate the GUS/LUC activities. An outline of the transient co-transfection assays is shown in Figure 4-3.

4-2-12. Generation of transgenic hairy root lines

Constructs for dual expression of *GubHLH3*(-*SRDX*) and green fluorescent protein (*GFP*) were used to generate transgenic hairy root lines. To construct the plasmid for dual expression of *GubHLH3* and

GFP, a fragment containing *GubHLH3* and *HSP* terminator of GubHLH3/pDEST_35S_HSP_GWB5 was amplified with primers 84 and 85 (Table 4-1) and cloned into pENTR/D-TOPO (Thermo Fisher Scientific). The cassette was then transferred into pBI-OX-GW (GFP selection) (Inplanta Innovations, Yokohama, Japan) to make *GubHLH3* and *GFP* dual expression plasmid. To construct the plasmid for dual expression of *GubHLH3-SRDX* and *GFP*, the CDS of GubHLH3 was transferred into pDEST_35S_SRDX_HSP_GWB5 (Fujiwara *et al.*, 2014; Figure 4-1d) using Gateway LR Clonase II Enzyme mix (Thermo Fisher Scientific) to make GubHLH3/pDEST_35S_SRDX_HSP_GWB5. Then, a fragment containing *GubHLH3-SRDX* (a fusion of *GubHLH3* and *SRDX* chimeric repressor domain) and *HSP* terminator was amplified with primers 84 and 85 (Table 4-1) and transferred into pBI-OX-GW (GFP selection) (Inplanta Innovations) accordingly. pBI-OX-GW (GFP selection) lacking a Gateway cassette was used as an empty vector control. These three constructs were introduced into *Agrobacterium rhizogenes* strain ATCC15834 by electroporation.

G. uralensis (Hokkaido-iryodai strain) seeds were scratched to stimulate germination, and placed on 1/2 Murashige and Skoog (MS) medium (Duchefa Biochemie) supplemented with 1% sucrose at 22°C with a 16-h light/8-h dark cycle. These temperature and light conditions were kept constant throughout hairy root culture. After 6–8 days, roots were removed and the sectioned surface was coated with pellets of *A. rhizogenes*. The infected seedlings were placed on the same medium and co-cultured for 2 days. Then the seedlings were transferred into 1/2 MS medium supplemented with 1% sucrose and 125 mg/L cefotaxime to eliminate *A. rhizogenes*. The seedlings were transferred into fresh medium twice before isolation of hairy roots. One month after infection, hairy roots strongly expressing GFP were isolated from the seedlings and placed on 1/2 McCown woody plant medium (Duchefa Biochemie) supplemented with 1% sucrose, 0.01 μM GA_3, and 125 mg/L cefotaxime. One day after isolation, hairy roots were harvested, frozen in liquid nitrogen, and used for RNA extraction. The rest of the hairy roots were subcultured several times on the same medium to remove *A. rhizogenes* completely, and then transferred to the same medium lacking cefotaxime. Hairy roots cultured for one month after subculture were used for metabolite analysis.

4-2-13. Analysis of sapogenins in hairy roots

Extraction of metabolites, acid hydrolysis, and sample preparation for gas chromatography-mass spectrometry (GC-MS) analysis were performed as described in Chapter 2 from 40 mg of freeze-dried powder of hairy roots. GC-MS analysis was performed as described in Chapter 2, but the hold time at 300°C was 38 min and *m*/*z* range was 50–850. Peaks were identified by comparing their retention times and mass spectra with those of authentic standards.

4-2-14. Treatment of tissue-cultured stolons with plant hormones or elicitors

Tissue-cultured stolons were cultured for more than 2 weeks in MS medium supplemented with 6% sucrose before treatment. MeJA and SA were diluted with ethanol to make 100-mM stock solutions, then 100 μL was added to 100 mL medium (final concentration, 100 μM). Yeast extract (YE) (BD Biosciences, Franklin Lakes, NJ, USA) was dissolved in water to make 10% (w/v) stock solution, and 1 mL was added to 100 mL culture medium (final concentration, 0.1%). For mock treatment in the time-course qPCR analysis after MeJA treatment, 100 μL of ethanol was added to 100 mL medium. Untreated tissue-cultured stolons were used as the 0-h time point in each series of experiments.

4-2-15. Quantitative real-time PCR

Quantitative real-time PCR (qPCR) was performed as described in Chapter 2. Relative transcript levels of each target gene were calculated using *β-tubulin* (Seki *et al.*, 2008) (GenBank accession number LC318135) as a reference gene. The amplification of each sample was performed three times, using primers 86–113 (Table 4-1) designed using the Primer3 website (http://bioinfo.ut.ee/primer3-0.4.0/) (Koressaar and Remm, 2007; Untergasser *et al.*, 2012).

4-2-16. Phylogenetic analysis

Full-length amino acid sequences of previously characterized bHLH TFs were collected from GenBank (http://www.ncbi.nlm.nih.gov/genbank/). To collect putative bHLH TFs in the previous transcriptome analysis of *G. uralensis* (Ramilowski *et al.*, 2013), the predicted protein sequences were searched against a hidden Markov model (HMM) of HLH (PF00010) obtained from the Pfam database

(Finn *et al.*, 2016) as previously described (Mochida *et al.*, 2009). Phylogenetic trees were generated as described in Chapter 2.

4-2-17. Statistical analysis

Dunnett's tests were conducted using the multcomp package in R (Hothorn *et al.*, 2008).

4-3. Results

4-3-1. Mining of candidate TFs

I performed two different approaches using model plant species to identify candidate TFs regulating glycyrrhizin or soyasaponin biosynthetic pathways in *G. uralensis*. The first approach used the publicly available gene co-expression analysis tool for *M. truncatula* (https://mtgea.noble.org/v3/) (Benedito *et al.*, 2008; He *et al.*, 2009) to find TFs co-expressed with soyasaponin biosynthetic genes. Although *M. truncatula* does not produce glycyrrhizin, soyasaponin I (soyasapogenol B glycoside) has been identified in this plant (Huhman *et al.*, 2005; Carelli *et al.*, 2011). In addition to bAS, CYP93E2 and CYP72A61v2 have been characterized as P450s for the production of soyasapogenol B from β-amyrin (Fukushima *et al.*, 2013). In co-expression analysis with a correlation threshold of 0.6, no TFs were identified when *bAS* (Probeset ID: Mtr.32384.1.S1_s_at) was used as a query. However, when *CYP93E2* (Mtr.8618.1.S1_at) or *CYP72A61v2* (Mtr.43117.1.S1_at) were used as queries, a sequence annotated as a bHLH TF (Mtr.1885.1.S1_at) was identified as a unique TF by both searches. According to the database, Mtr.1885.1.S1_at corresponds to Medtr0246s0020.1 in the Medicago v4.0 annotation.

The second approach was comprehensive screening of an *Arabidopsis* TF library through high-throughput Y1H assays (Mitsuda *et al.*, 2010) against the promoter sequences of *bAS*, *CYP88D6*, and *CYP93E3* (Table 4-3). Promoter sequences of approximately 2 kb for *bAS*, *CYP88D6*, and *CYP93E3* isolated from *G. uralensis* (strain 308-19) were divided into approximately 500-bp segments (Figure 4-4), and used as bait. Of 1,428 *Arabidopsis* TFs, 32 TFs showed positive interactions with at least one fragment of the *bAS*, *CYP88D6*, or *CYP93E3* promoters (Table 4-4). Since a bHLH TF was

Table 4-3. Promoter sequences isolated in this chapter

bASpro
ACCCCGGTCTAGACAAAAAATACTTGTTTTTACATAATTAATCTAATGATGTTATTTTACAATAAAATTAGAAATAAAATAAATTAGAAAAATGACGTTGTATAG TATCAAGAAAAATGACGTTGTATAGAGTATCAACCTCGTTTGACAACCTCTTGTATTGAATATCATATTGCGGATTTTACCAAATAAAAATAGAAATTATCATAA TTTAATAAAAATTATATAAAATTATTTTATTTAACAATAAGAAATGTGCTAAAAATGCATTGCAAAATATATTATTAACAATTTCGTTTGTAAATTTTGCTCAGC CAGATGGTCGCCTACTCTTATTTGTAATCCTGCTCGTTTGTTGACCACGCTTCCCTCGTGTTTCATTTTCTTTCTTAAAAAAAAATTGTCCCCTGGTTTCCACGC TATACTCGCTGCAAATATAATAATGAAAGTAATATTGTTAAATTGCATTCTGTTCAGAAAAAAATGATTGAATTAATTTTGTTGTATACATATGTCTTTTTTTTT TATCATACATATATATTGTATATCACATTTTGTATATTTTAAAATAAATAGATTTGCAAACATTTTTTTTTTGCTAAAAATAAAAGAAAAAAAATAAACGATGAC TCTTTACCAGATGAAAATGAATAGAAAAAACACTTGATATTGTTATAATAAAGAGGTGAGAATGATGGCAATTCCATGGAAAAACACTTACAAGATGCATATGAA GCAATCATCGTAACTTCCAAAACATAATAATAAATTAATATTTATGCCAATACACTAAAGAAAACACTTGCAATGTTTCAAAAATAGGATTTGGAATAAGTAAAG TAGAGTAATTATAGAGTCACAAAAAATTGTAAACTCTTCAATGTGTAAAGGCATCTATTTATATAAAATTCTAATAATCTAAATGCTTGAGTTATCTTATTTAGT TGTCAGCATATGTAATAAAAAAGATTTAGTTATCATTTTTGAACATCCACTCCACTTCCAACATCCATCAAATATATTTGTATTTTTCTATCTCTTTATTTTTTT ATTACATTACTCATCTATAACTTATATCTCTTTTATTTTTTTCTCTCTCAATGTGTTAAATGACATTATGGTATCCATCTTTTTTTTTTGGGTACACATTTTGGT AGCCATCTAAGTTTCTCCTAATTAAAGATTAGCTGATTATTACGGTCATGATAATAAATAAAACGCTGCAATATTAGTTTGTCGGAATTGAAAATGTAAGAAAAT GATTTAGGTGGGTTTGACAAGTAATGTCAAACAAGTAACAACATAAACAATAAATAAGAATGCTATAAGTAGGGGCAGCACGAATGGCTCTTTCACGAGTCACTC TCCAAATCAATTGAGATTCAACAGTGTTGGTGTTGAGCGTCACCTGCTCTGCTCTGCTCTGCTTGCTTTAAAACTCGTTTAATTTGTGGAGACTAGCTGTGTGAT TCAGGTAAGCTTATAGTTCAAAAGTTCAGCACCGGATCAATCTCAAGACACACAAATTAATCTCACGGTTGGTAATTTTGCAATCCATCTTCCTTTTCTTTTGCC TGAAATATTATTATGTTAGTAAGTAGTATATTAGAAAGAGAACTCTGTTAAGGTTGGCTGGTTAGTTACGAGCTGCAAATAATGATATATATATTTACCGCGTGC GTGCGCAGGATGGAGTGGAGGAGGTAGCTAGCACTTTCAATTGATTCAAGCATTAAGGCGGGGATTGTAATAATAATAATAGAATAGAATACTAAGAAGAAGAAT CAGAAGAGGAAGGAAGATGTGGAGGCTGAAGATAGCGGAA
CASpro
ACATATGCTGGCCTCCCCACTACACATTGCAAAATGTATACAAAGTTTACCAAGAAGAGTTCAAATCAATGCCTCACCGTTCATATTGGCTAGAATACAATGAAA TTAAATTGTGCCACAATCCAGCAATACAGAGAGATCCAAATGGTCGTCCACAGTCAATGAGAATTCGTACAGAAATAGATCAAATAGAAAGGAATCGTCATCCCA AGAGATGTGGACTGTGTCAAAATGAAGGGCATAACAAGAAAAACTGCTCTTACCGTTCATGAGCTGAGACGCAACACTAAGAAAATTTCTATCATTTGCATTTAG TTCACTGTTTTATCATTTTTGTTTGTATCACTTTTGTATTTGTATTTTTAAATTTTAAATCAATTTGAGCTATTTTTTATAATTATATTATTAATTAGTTGACAA ATCGGGTTAATAAATATTTAACAATTACCATTAATATTTAATTCATTAAAAAATACTTAAATAAATATTATATTCAAAAGTGTGTTATTTTGGTAAATAAAAAAA ATATGTGTTATTTTGGTAAATAATAAAAAAAAAGTGTGTTAATGGGGAAAAAAAAGCCGTGAATGTCTATTTACCTAAAAGATTTCATAAAATGTGCCCGAGTTGT CAATTTTATCTTTTATAAACTTTTATTTGAAATGAAGAACACATGACAAGAAGAATAGAAATTTCCAAAAGATGAAGTGTGAAGCCCGGGTGCGGTCTGGATCGG GTAAAAGATAATGAGTAAATAGTCAAAATAGTCCAGGAAATTGTAAGGCGCTCTCAAATTAGCCCTTAAAATTAACGAAAACACAAAATAATCTCTTATTTTGTA AACCGTTATTCAGTTTAGTCTCTCCGTTAACCTCAGTTACTTGATGCCGTGAGTGAGTCTGACGTGTTTAGTCCCTCAAGTAATTGAGGTTAACGGAGGGACTAA ACCGAGTAACGGTTTACAAAATAAAGGGCTATTTTGAGTTTTCATTAATTTCAGGGGTTAATTTGAGAGTGTCCTACAATTTCAGGGCCTTTTGACTATTCACTC AAAAGATAATCCACAAACTAAAGCGATTTTCTCCCTACACACACTTAACATTTTTGACACTAAAAAGACCAAAGTGCCCTTTCCATAAAAACCCTCCTCAACACT CAGCCCCCAACACAGCTCCGTCGTTCCACCGCCTTGAGACGCCGCCACCGCTCCCACATGTCGTTGCACCACGAACCAACGCTAGGGAGACCATTGCAGCAAACA ACATTTGTCGTCGCAATCGTTGGTAAGCATTAATGTTTTTGTAATTTTTTTTTTTGCTGTTCTGCATTGTTCTGGAACCAATAATTCCTCAATTGAAATATTTTCA GTTACAAAATTAGTACTAAAAAACACAATTACGAAATTAATACTTTTGGAAGTGAAAATCTCTATTCCGGAATTAACACTTTAGGAGGTTAGTGAAGTCATGTTT ATCACTGTGTTACAGAACTAATTGTTTCGGAAATCAAATTTTATGTTCCGGAGGTAATTAGTATGAAAGTCAAAAGTTTGGTTCTGAAACTATTAATTTCGGTAT TTTATGTGTCTTGCACGAACTAATACTTATTGGGATTATTAATTTCAGAATTCAAAATTTCTATTTCGAAGTTAATATTTTCATAAAGTCAGGAAACGGAAAAAA AATAACTCGTGATTTACCTATGGTGGCGAGCTCGGCGGCAATAGGGGTCTATTTGGGGGTTGGAGATGGTTAGGGGTTGCTTGGATGGTTGAAGAGGAGGAAGAG GGGGAGGGGGAGGGAGATGGAAGAGAGCAGAAGGAGAAGAGAAACAAAAAAAAAAAAACGAAGAAGAAAGAGGTGGGGAGTGGGGCGCAATGGTGGGAGGATTTTA AGGGTAGGGGTAATTTGGTCTTTTCACTGACCATTTTTTTTTTTTTAGTGTGTGTCGGGAGAAAATCTCGAAACCAAATCCACGCACACACACATTCCGTAACGT GTTTTGGCTGTTTGGCCCAGCCAGCCAGCCAGCCATTAGTCCTCAGTCATCATTTGGCTTTGCGAGCTCCGTCCCCTCCGTTCGCAGTAGCTGAGCTGTTG TTGAAAATTAAATGTAATTAAATGCAATGCTCGAATTCAACAATTTCATTTTCCATTCCAGAACCTACACGCTATCTTCTAGTACTACTCCTCCAAAGCAATAAC CATCCTTCATTCCCATTTCTCCACTCTTTCTCTTTGATTGGTCAGTCAGGCCTCTTGCAAATCACAGTAGTATATCACACAGACTCACACAGGTTGCATGCTACT CATACTCACGCTTAAACGCTTGTTTCATTTCATCCATCCATCATCTCCATCTCCCAATTGTTGCTGAGTTTTATTTTATATTTGTTTATTTATTTTTGCACC CGAAACTCACTCGCATTGCAATTCCTTGCTTGCTTGTTTTTGTTAGAGGAAAATGTGGAAGCTCAAGATTGCGGAG
LUSpro
AAAATACTGCTTGCTTATGCGTTCTATTCATCTTTCTTTGTTCTTGGAAATTTTTAAGTTGCAGAGTTATTTTTCTTTTGGTTTTATTTCTGATTAATAAAAAAA GTATTTACTTACGTATTTCATTATTCATTTATCTTTGTGCCTAAAAATTTTCAAGTTGTAATAATATTTTCTTTTGATTTTGTTTCTGATCAATGAAATAGTATT ATAGGTTTCAATCGTCTTTCTTTATTCGTGTAAATTTTTAGGTTGCAAAATTACTTTTTTTAATTTTATTTCTGATCAATAAAATTGTGTTTACTTATAGGTTTT ATTCATATTTCTTTGTTCTAGTAACTTTTCAAGTAGCAAAATTTTCATTCTTTTGGTTTTATTTTCGATCAGTAAAATAATATTTGTTTACGAGTTCTATTCATC TTTATTTATTCTTGTAATTTTTCAGATTATAAATTTAGAGTATATGTTATTTATCAATTTATTACTAAATAAACATTTCATATCATATGTAAAATCAGTCAGAAA TACGGGTCAACCGACTAGTTGTAATAAAAAAGACATAGCACGAATCATATCCTACAAAACTATCACCAATCAAATACTTAATATACAAGTTAATGCTATAATCAC GATTAACAGCCACAATTAAATTAAAAAAACAATATTTCTTACTGTTCATCTTTTCCCTTTTAATAAACTATCAACCACCATATATCTTTAATTCAATTTTAAAAA CACCATCTTTAATTTATTTTTTAACAAGCTGGGTCGCTGTCCATATGTTATATTTTTCCCTCACGAAACTTCCTCACAACCACTCTTTAAGTGCATAATCGATCG ATACGTTCGTTAAGCTTTCATGTTCGTGTTTCTGTGTCCCAACAAGTGATAGTTGTTACTCTAGTAATGCGTAATTCTCACAAAATCTACGTGTCAGGTTTTAAT TAGTTACACTTACATTGCATCATCCTATTTGTGCGAATTTTAGAGTATCCGTCCTAAAATAAATGTTCATTTTGGCTTTTTTACATAAATTAAAAAATACAATTA AATAAAGGAGTGAAAATAACAAATTTATCAATTTATCCTTGCTAATTACTTATTGTTCTTGGTTAAATGAGTCATCATCACTATATTTATTAAGGGTAAAAATAA AATAATTTTATTTATAATTCATTGAAATCCTAAAGTGAACAATTATTTTGGGACACATTTTTTTGCTAAATGGACATTTATTTTGAGATGGAGGAAGTACATGAT CTTCAAAATGTGGTTAGGGTACATATCAAAAAAAATGTGGTCAGTATAAAAGAAAGTACGTAGATATTTTTGGACACAGTCACACAGGACATTATGTACTATACT AGAAATTCACAGTTTTCGTGCAAAAAATTTCTGTAGATTTTTTTCTGCAAGAATTTTACATTTTTTAATGTAAATCTCAACTTTAATCCAATTATTGTAATTAGAT GCTATCTTTTCTCACAATTAAATACGAGTTCTCATGATGATACACTCCCCCCCCCCACCCACACACACACACACACATATTTATACAGAAAAGGGATCAAATTACA CTAATGTAATACTTGGAGTTACATATCTAATAATCTTTCATCGAATAACTATATTATAAAGCGTCGTGATAACTACATGTGACATTTAGAATTATATTTAAAATG CTAAAACAATATATATTCAAGATACTTGCTATTAATCAATAGGAAAAAGATGCATAGACACCCAAGGGGTATATACACCTAGAGAGATAAAAAGAGAAATGAGAG ATAAGTGATAGATATGATAGGTGATATGATGTGACAAGAAGAGAAATATTGAAAGAGATAAGAGATGTTAAATGGGTGTAAAAAATTATTGGGTGTTCAAATATC ATCACCTTAATCAATATCATCGATCGCAGACGTCCTTCTTGGGCCATATAACAGTATCTCCCTTCGCGGCCGGTGTGCATGTGAGTAGTATATATACACACATCC CTCATCAGCCAAGAGGTGTCTAGCCCTTTGTTTCCTTTAATATATATATTGGAGGCATTTCTTTCCTCAACACTACTACTTAAGCATCTCGATTTCCGTTTAATT TTTCATCTTGTGAGTAATATATATACCATCAAACATAATTATATCTGTTTTTTTTTATCTGTTGTGAGTTGTGATTATTAATGGGATGATATTGATTTTAATAAT

ATTTTTACTGCATGTGTTGGTCCTAAACTCTAATAAAGAAGAAGAAGAAGAGTACTACTAGGCAAGGCAATTAAAGCGAGAAAGAAAAAAGATATAAAACATGTG
GAAGCTGAAGATAGGAGAAGGAGGA

CYP88D6pro

CTGTAGTATATAACAAAAGAGAAATGTTATATCATCCTACTACTTTATAAAGTAAGTATTTAATAAAAGAATAATCTAAATCCTCCTATCTTTTCTTCCTACAAA
TCCTCCTAATTACTAAATAGAGAAAGTATTAAATATTTTTTATGTTTTCAATATATTAGTTATATTCTCTTTTTTACTTTATAAAGTTGTAAGATAATATGTAGA
AGGATTTGTAGGAAGATATAACATTTCTCTTAATAAAAAATTGAAGCTTAAAAACATAAAAAATATTTAATATTTTTTTATATAATAATTAAGAAAATTTGTTGA
AATAAAAAGTAAAAAGATTTAGGATTATTAAAAAAAAAATGAAAAATGGGATTGCATTGATTGCACACGTACTTGCAGCTCACTTATTATGGCAAAAGACTTGCTA
CTAGAAAAAAGGTGGAATCGGATTCTCTACGGGGCTTTTTTATCCCATTAGCACACTTTTATTTATTTATCAAAATAGTACACTTTTGAATATAATATTTATTTA
AGTATTTTTTAATAAATTAAATATTAATAGTTGGTTTAAGATTATAATTAAGTTATTAATTAAATTAATATTTAATTGTTAAATATTTATTAACTCGATTTGTCA
ACTAATTAATAATATAATTATAGAAAATAGCTCAAATTGATTCAAAATTTAAAAATATAAATATAAAAGTGATATAAACAGAAATGATAAAACAGTGAACTAAAT
GCAAATGATAGAAATTTTCTCAGTGCTGCGTCTCAGCTCCTGAACGGTAAGGGCAGTTTTTCTTGCTATACCCTTCATTTCGATACAGTCCACATCTCTTGGGAT
GACGACTCTTTTCCATTTGATCCATTTCTGTACGAGTTCTTGTCGACTGTGGACTAAAACAGTGAACTAAATGCGAATGATTGAAATTTTCTTAGTGTTGCATCT
CAGCTCCTGAACGGTAAGGGCAGTTTTTCTTGCTATGTCCTTTATTTCGACACAGTCCACATCTCTTGGGATGACGACTCCTTTCCATTCTATCATTTGCATTTT
GTTCACTATTTTATCATTTATGTTTGTATCACTTTTGTATTTGTATTTTTAAATTTTGAATCAATTTGAGTTATTTTCTATAATTATATTATTAATTAGTTGACA
AATTAGGTTAATAAATATTTAACAATTAAATATTAATTTAATTAACAACTTAATCACAATCTTAAACTAACTATTAATATTTAATTCATTAAAAAAATATTTAAA
AAAATATTATATTCAAAAGTGTGCTATTTTCGTAAATAAATAAAAAGGTGTTATTTTGATAAATGAAAAAAAAAGGTGTTATTTTGATAAAAAAAAAAGTATGCT
AATGGGGTAAAAAAGCCTCTCTATGGCGGTGATTGCGGAAACAAATCAACTTTTTCAGTTGACAGGACAGTATTAACTGTAGATTTTTTTTTTGGTGAGTCTATT
ATAACTGCAGATTTTACACCTTATGGTTTATTCAAGATCTATCTAGGTCCTATTTGGCATAACAGTTTATTCTACAATTTATGGCATAAATTGTATAAATTATGT
TTATAATTAAAGAAAAAATATGGTGAAATTTATTTCATATAAGCTATTTTGATAAGCTATTATGGGTAGCTTATAAAAAATAAACCCAAAACAGCTTATGAATCA
GTCATAAGCTGTGTTCATAAGCTCTGCATAAGTCAATCCAAACATGCCCTAATCTTTCTCATTTACATCATGAGCAATCAATCAACTCTGTATATAAAGAGAGGG
ATCCGATGAACAACGAAACAGATTCAGAAAAATGGAAGTACATTGGGTTTGCATGTCC

CYP72A154pro

GTGATATCCGCGGGCCATGGATATTAAAAAATTAAAAAAAAATCTATAAAAATTCTATTAAAAGAAATTGTTTTCACACAATATTGATCATAATGAAGATTATAT
TAATAAGACTAGTTATACCAATCATTATAATTTTAACTCAATTGTTATATCAATAATTGTTCAATTAAATAAAAACAAATACTACAAAATAAAAAATGAGTGT
GAGAATTAATACATGACATAATAATAATACATATTTTGGAATCACTTTGTTCAATAGTAAAACAAAGCAAGACAACAACAAAAATGAATGCATCACACTCTAATA
CTTGTCAAATTGTGCATTGGTGTTTTAGATTTTTAGGTATATATATAAATAAAATGATATTTTTTATAATTTTACTAACGGATACGGGTACCTCACTACCTGCTC
CCGTCCCCATTCATTGGCGGAATGGTGGGTACCTCAAATACCTGTTTAACTTACCCGTACCTGCCCTCGCACCAGTAATGGGTTTTATCCTTGGGTACCCGTGGG
GCACGGGTTTTTTTGCCATCCTTATATATTTATTTATTTATCTTTTGAAAAATGTTTTTATGGATACCCCACATGTCATTCTATTTGAGAACTTATGAAAATAAG
TCTAAAATAGTTTATAAACTAGTAATAAAATATTTTTACTTTCTCTCCCAAACACTATTAGAAGTGCTTATGTCAATATATAAGCTTTACATAAGCCAATCCAAA
TAAGCTCCTAGTATCAGGGTCGACCCTGAGATTTTGGGGCCTTTTGCGAACTTAAAAACTGATGCATTTATAATCATTTGAATATTGTTCTCTTAATATTTTTTG
CTACAAAATCATAAATAATCAATTCACAGCCAAAGTTTTTTTTAAAAATGATTTTCGAATCAAAGTAAGACAGACTATTAAATCTATCTTGTGATATAGTAGATC
GCAAATAAGATTTTAATAATTTCAATTTAGAAAATATTCTCTCAGTACATGCAATAATAATAGTGATAATTAATATTATTCTATAAACTACGCATGCATTAGAAA
AAGAAATAAGACTTTAAAAATTATTCTATAGTAAATTGTGTACAAAAAAATTATTCTATAGTAAATAAAATTTTACATATTGGTATCCTCTAAAGATGGAGTCTT
GTGCAGTCGCTCCTTTGCACATGTTTAGAACCACCACTACCTTTTTTTTTGTTGATACAGAGAGCCACCACTACCTAGTATAAGATACGAATAGAAGAAAAATAG
TAATAAATATAATAGATGGATAATTACTAATGTGACCAAAAGAAAAAGATAGAACAAAAAAGAAGTGTGGTATGAATATTAAAAGTCGAGTATGTTCAAAAAAGT
AAATGGTTAATTTATTTATATTTTAAAATTGAGTAATCAGTTACTTTAATATCTGAGAATGTCAAGCGATGACACTTTGGCTTATGACAGATACGAAATTATAAA
AACGTCCCCAAGTGTTGGCCGTGAGTCAATTTGATCCCTTAAAGTAATTATTTACTATTATTTTGGTCCCTAAAAGTGATTACTTAATTGTTATTTTGGTCTTAA
AATGATTATTTTCCTGATGACAAATTCATATATGTCTGTCAAAAACCAAAGTGACACTTTTTGATACTCTCGAGAACCAAAATAATTATTTACTCTTTTAAAATT
GGTTTGTGCACTTTCCACGGGAGAAGGTGCCGCATAACTACAGCTGGCAAATGGGCCTACCCGGCGGGTTTCTGGCTGGGCCAAAAAAGCCCAGATTAAAACGAG
CCCTAATTTAGTTGCCCAAGCCCAATCCGCATAGTGTCAGGCTACACTGAACTCGGCTATTACTTTGTTATTTTTGGGCCGGCCCCCATTGGGCTCAGGCCAAAA
ATATCCAAATTAAAAATTGACTCAAATTTATATGCCAAAACCAGATATTTTTCGGGCAGACTCGGAATGACCCGGCGAACTAAGACCATTTTGCCAGCTCTACGC
ATAACCCACTATAAAATTCCCACCTTGGCGTGGAAAAGAAACGAAACAAATCAAGGGGATCTATACTATACCATATCTTTCTATAGTTCTAGACAGAGAAAAAGA
AAAAATGGATGCATCTTCCACACCAGGG

CYP93E3pro

TTAATTGAAAGCTAAATTTCCAAGTTCAATGATATAAATGTGTATGATTTATGAGATTATCCAAAGACTTTAAGGTGCATTTGATACACACTGAAGTATACAGGT
TTACAAGACAAATAGAACAACATAAAATAAATGCTTAAGTCATAACACAATTCTTTGTGATGTATATTGTTTGATGGATACTAGACAACACAAATAAAATATTGT
AAAATTTATTACAATACTCTTTGTATCGTGCAATATTTGTCCCACACTAAACAAATCGTAATATCATTTTATTATTCTATATATTATAATATAATATATATTATTA
CTATATTATTACATAAAATATCAATATTAAATATTTTTTATATTTAATATAAAATATTATACACACACACACACATTATTATGTCAGTAGTAAATATAGTTTCAT
AATATATTTAATATTAAATATATTTTATTTATATAAAATAAAAAATAAATATTTATGTTCAATATATTTAAATATTTTATATATGTTAAATATAATAAAATACTA
AATATTTAATATAAATATTTATGTTAAGTATATATAAAATATAAAATATTTTACTTATATTAAATCTTTAATATTTAAATAATAATAAAAAAATTAATATTTAAT
ATCTTTTATCATGATAAAAAGCAAGTGAAAATATCAACTTATAAGAGGGTTTTCTTTCAAAGTTATTCATCTGTTTTTAATTTATTAACTTAAATATATACAAAT
TATTTTCAAGAGTTAATCACCTTTCAATTAAAAAAACTTTAGAAAAAGAAACAAAAAACTATATAATTTTTTTTTTAAAACATTGTGTGAAAGCAGGATAACTTA
TGAGCCATAAATTATTAAATCAAATGCGTTATATCTCACGTGAATTTTGCAAGGGATACTGTCATGTTTACTTTGCTATTTCTTGGCATCACGAGGAAGAGACCT
GCCACGAGGGGCGTAAAATAAGTACACCTCGGCTCCTAAAGCTCGACCTATCCCCTAAAATGCGTTTGGAAGAGTTTATGTACATTTTATTTAGATTTATTTTAT
GACATAAGTATTTATGTAGATATTTGAGAAAATTTAATTATGAAAATAGTATATGCATTTGTTTTTAGTTTATTTTAATAAAATCTTCAAGATAGCTTGAAGTTT
AAACAGCTTAATCTTATTTTCTCTTTGATTGTAGAAATAGGTTATTCATAAAATCTATATTTCTCTCTCAAAAATTAAATATAAATTACACTTTAATTTATTTTA
ATATAATCAAGCTTTTTATGTATATTTTCTCTCTCTATTTTGATCATCACTTTCGTAACCATATAGTTTAAAACTCATACTATCAACTATAAAACATGTAAAAGA
TAAAATAAAAAGATTGGGACAAAGTTCTAAAATCGATAAAAAAATATATTCAAAATTTAATTTTATCTTTTAAAATTATAAAATCTATAATAAAATATACAAATT
TAGTTAATTATCCGCTATTTTGAAAGTTGTAAATGAAAAAAAAAATTATTCCTAGAATGATCCTTATAAGAGCTGTCCGCCACCAATATTAAAAGATTATTTATA
TCCAATAGACTATATCAAAATCAAAATAAGAAGAAAAATGCAACTCAACTTCCTTCAAAAGTGTGAGTGTACTTAAATTACCATTTATAGTATTGATGAATATAT
GAATTAATGGAAGCCAATTACTAAATGTGTATTTGGGAGAAGACGCACACACCGTGACAGTGGACGCGGAACACATGTCACTTTGTTTTCAATAATTATATAATA
TATGCAGCACGCGGAACACGGTACCTTGTATAACATAATACAATATTGTCCCACGAGGTTCGCAAAAATTAGTAACTTAAAATGGCTAGAATCTCAACGTAGAGC
AAAATGCATTGCTCATCATAACTTGAAAAAGGCCATCCCTTATATTTATGTCCCTCTCCCATGATTGAAACCCTCAGCACAGGTTGAGTTGTGTCAAGTTTGGCA
CATATATTGTTCTGTTTTCCCTCAACTCAACCATGCTTGACATCCAAGGCTAC

CYP716A179pro

TTGATGCTCATGTCATGTGTGAGGGCAAGGGAGTCTAGTTTTAAACAACTAATTTTATTTTAATTTTTTAAAGTAAACGGTTATTTACGTTGAGATCTACTTTTC
GCGAATAGTACTTTTCGTCGCTTAGAAATCTTGTTAAAGAGATGTTTTAACTCGGCATAAATTTTGGACCGCCATATTGGTACTTAGTAGAATGTATTTGTCTAG

CATTTATGTTGACATCAAGTTCAGCGGTATGGTTCGGCTGTATTATTCAATTCCCATCGTGTTGTCACAGAAATGAGAAAAATACGGTAGACAGAACCATTGTAG
TGACTTGTGTACATGTGAGTATTTTCCGGATCAGTAAACTTAATTCTACGAAGAGCGCTACGAACAAATTCTCTAACACATTATTTTCAATACACCATTTTTAAT
TGGTCGAAATACATGTGGATAACATTAAATCATGTGGGTTCCTGTCTCATTTAGTGGGACTCATTCCTAAAGTGGTGAGGTCTACATGCATTTCGACCAACCAAG
GAGGTGTATTGGAAAGAGCGTTGGTGTACTCCTCTGGTTCTATTTGTACTTGTGTTTCTTGGTGTGTAGGGTGGCATATGTGGTACTGGTTACAAAATGACTAAT
TTGTCTCATAGCCCGTCTGAAGGGTTGAAGTAGTGGAAATACCTTAGGATCAATGCTATTGGGGACAAATGCCTTAACTATGGTGGTTGCACACTGGTGGATCGG
ATCGGGTTTAGAGGCAAAAAAATGGATCCAAATTGAATTGAGATTTTAAAAGACTATTTTGTCATAAAAAAATCTTGTAGATTTTTTAGATTTTAAAAGACTACG
TAGGAATGTAATAATTTATCAGACTTTTTTTAAAAGTCTTGTGGATTTTAAATAACTTTTATATCATAAGAATTTATAAAATTTGAAAGGATTTTATGGATTTAT
AAAGGACCTCAGGGTTTTTTAGCAAACTTTCAAAGTTATAAGAGTTAATAAAAATGAAAATAAAGAAATTGGTTGTGTCTTATCTCCTCCTAAGTAATAATTCTC
CACATAAGTCATACTAGAATTGGTAATACTGATGGCAACATTGAAGGTGAATATTCGTCTGCTGGCTCAAAATTATAGTTTTCATTAACTGATAATACCGAAGGT
GAAGATTCGTCTGCTGGCTCAATTAAAAATTCATCAGAACTACATTTTTTGCGAAGAAAGTTATGTAATGCTGCTATGCTAACACAAATTCTTCTTGTGTTAAGC
ATACTATATAATATAATATATATACATATAAATGAAATAGTTTTTCGGTCTATAGTAATAATAAAATTATAAGAATAAATGAGAAACTAGTAAAAAAATACTTTT
TATTTATCAATAATAAAATAAAAAAAATAATTTTATATTCTAAAATTATAGAAATAAAAAAATGAATGAAACATGCTAAGAAATGTTACTATATTTATCAAAAAA
ATTTTTTTAAAAATTATAGCCTGAAAATTTATAGAAATAAAAAATATGAATAAAACCTGTTAATTAGAAAACTATTATATTTATAAAAATAAAAAAGAGGGTAAC
GAATGAAAAAATAATTGTAATTTGAATAGCGTTTGGGGAGATAAAGTATATAAAGAATGTACACACACATATTTACAGTATTGTCATCTAAGTATATATAACGCA
TAAAGAATGCAGGAAGAAAACAATCTGGAAAAGGAAAAAAGAAGAACAACAAAGGAGAATTACTTGGTGAGTCTGTCACTGCTGCAATGGCAGAAAAAAAGTCTG
AGAGAATAGGAAAAATATTGAGGGTTCGGGTGAGATTGTTGGATAGGTCTGATCTTTATAGTTTACACTAAAAAATCTAACGAAATCCACCCTGCAAAAATAATC
CTTTAAAATCTGTATATTTTTTAATATCAGGAGACTTTTTTTAAGTTGTAAAAAATCTCAACTGAGTACCACCGAATTTCATTACTCTCTTTAAAAAATCATAAT
AGTCCTAATTAAATACCATTAGATTTCATTGCTCTCTTTAAAAAGTTATAATAGTATCCCTCATTTATTAGTCCAATTTGAATAGCACCAGAGTATTATATATAA
GGTTAAATATATTTTTTATCTCTTATTTTTAATAATTTTTATTTTTAATCTCTCATATTTAAATCATAACTTTTTAATTCTATTTTTAGAAATATACCACGTTTA
ATCCATATGACTGAAATAGCTAGCTAACAGGTCTACACCCTTATATATATTACTCCCTAGTCAATTTGACGTACAGTAATTGATAATGGCACAAACATTAATGAA
AGAAATTGTTTTTTGTCTTCAATATTTCTCTAGCTTGCGCACTGCGCAGCCCTGTGCCTATAAATTGAGGCACAAAACACACGCCAACTGCCACTGCACCAATGC
CAGTTAGCTAAGTTAAATTGATTAGTATAAGAAAGTACTTTGCGATTCTTCTTAACAAATTTGCGCACTCACAAGAAAAGCGTGATCGATAATATGGAGCATTTC
TACATGTCCCTT

Transcription start sites (TSS) identified in this chapter are indicated in red. E-box (CANNTG) motifs are indicated by a gray background, and N-box (CACG(A/C)G) motifs are indicated by a pale blue background. Sequences after the start codon are underlined in black. Intron sequences are underlined in red.

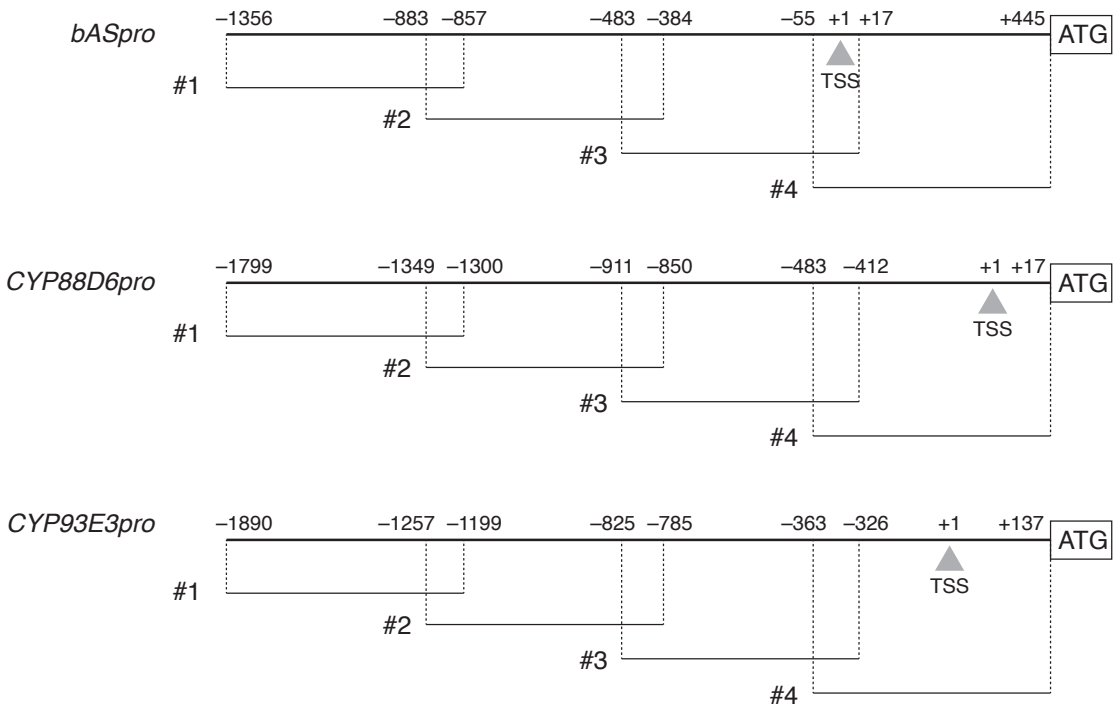


Figure 4-4. Schematic illustration of promoter fragments of *bAS*, *CYP88D6*, and *CYP93E3* used in high-throughput yeast one-hybrid (Y1H) screening of an *Arabidopsis* transcription factor (TF) library. Abbreviation: TSS, transcription start site.

Table 4-4. Observed interactions (indicated by circles) between *Arabidopsis* transcription factors (TFs) and promoter fragments of *bAS*, *CYP88D6*, or *CYP93E3* in high-throughput yeast one-hybrid (Y1H) screening of an *Arabidopsis* TF library

Locus	Promoter fragment												Common name	TF family
	bASpro				*CYP88D6pro*				*CYP93E3pro*					
	#1	#2	#3	#4	#1	#2	#3	#4	#1	#2	#3	#4		
AT1G04880									●					ARID
AT1G20910	●										●			ARID
AT1G76110											●			ARID
AT1G14490	●				●		●		●		●		AHL28	AT-hook
AT2G42940					●		●		●				TEK, AHL16	AT-hook
AT5G49700	●				●	●	●		●		●		AHL17	AT-hook
AT1G20900	●				●		●		●		●		ORE7, AHL27, ESC	AT-hook
AT1G76500	●				●		●		●		●		SOB3, AHL29	AT-hook
AT4G35390	●				●		●		●		●		AGF1, AHL25	AT-hook
AT3G04570	●				●		●		●		●		AHL19	AT-hook
AT3G55560	●				●		●		●		●		AHL15, AGF2	AT-hook
AT4G12050	●				●		●		●		●		AHL26	AT-hook
AT4G14465							●						AHL20	AT-hook
AT4G22810	●				●		●		●				AHL24	AT-hook
AT1G63470					●								AHL5	AT-hook
AT1G63480					●								AHL12	AT-hook
AT2G45850	●				●		●		●				AHL9	AT-hook
AT3G61310							●						AHL11	AT-hook
AT4G17950					●								AHL13	AT-hook
AT3G49610						●								B3
AT2G31460						●								B3
AT1G32640	●						●						JIN1, JAI1, ZBF1, RD22BP1, ATMYC2	bHLH
AT4G17880	●												bHLH004, MYC4	bHLH
AT2G22750	●												bHLH018	bHLH
AT2G21230					●						●		AtbZIP30	bZIP
AT4G38900					●						●		AtbZIP29	bZIP
AT3G02830											●		PNT1, ZFN1	C3HZnF
AT3G12680											●		HUA1	C3HZnF
AT2G37590					●								AtDof2. 4, ATDOF2.4	DOF
AT1G51700					●								ADOF1, AtDof1. 7	DOF
AT1G30210				●									ATTCP24, TCP24	TCP
AT1G14440							●						FTM2, ATHB31, ZHD4	ZF_HD

identified in the first approach, I focused on the three bHLH TFs: AtMYC2 (At1g32640), AtMYC4 (At4g17880), and AtbHLH018 (At2g22750) (Table 4-4). Among the 1,428 *Arabidopsis* TFs included in the Y1H assays, AtbHLH018 had the highest alignment score with Medtr0246s0020.1 (46% amino acid sequence identity), the *M. truncatula* TF identified in the first approach (Table 4-5).

Putative orthologs of Medtr0246s0020.1 and the three *Arabidopsis* bHLH TFs in *G. uralensis* were searched using blast (Altschul *et al.*, 1997) in the transcriptome database of *G. uralensis* (http://ngs-data-archive.psc.riken.jp/Gur/blast.pl) (Ramilowski *et al.*, 2013) (Table 4-6). Among the *G. uralensis* unigenes in the database, Unigene29516 had the highest alignment score with Medtr0246s0020.1 (65% amino acid sequence identity), and a relatively high alignment score with AtbHLH018 (42% amino acid sequence identity). Unigene 21689 had the highest alignment score

Table 4-5. The first five hits from a blastp search using Medtr0246s0020.1 as a query sequence in the *Arabidopsis* TAIR10 proteins

Query	Subject	Amino acid identities (%)	E Value	Score (Bits)
Medtr0246s0020.1	At4g37850.1[a]	34	4e-42	169
	At2g22750.2	46	4e-39	159
	At2g22750.1	46	7e-38	154
	At2g22760.1	43	3e-36	149
	At2g22770.1	39	2e-30	130

The blast search was performed at https://www.arabidopsis.org/Blast/index.jsp.
[a]At4g37850 was not included in the 1,428 *Arabidopsis* TFs used in the Y1H assays.

Table 4-6. The first five hits from blastp searches for each query sequence in the previous transcriptome analysis of *G. uralensis*

Query	Subject[a]	Amino acid identities (%)	E Value	Score (Bits)	Complete /partial CDS in the DB[b]
Medtr0246s0020.1	**Unigene29516**	65	5e-96	347	complete
	Unigene11765	51	2e-67	252	partial
	Unigene21970	51	6e-48	187	partial
	Unigene16153	40	1e-44	176	complete
	Unigene1680	41	9e-31	130	partial
AtMYC2 (At1g32640)	**Unigene21689**	56	0.0	634	complete
	Unigene723	54	0.0	630	complete
	Unigene29454	31	9e-45	177	partial
	Unigene37824	31	1e-44	176	partial
	Unigene17643	45	5e-43	171	partial
AtMYC4 (At4g17880)	**Unigene21689**	53	1e-168	588	complete
	Unigene723	51	1e-167	585	complete
	Unigene17643	32	1e-65	246	partial
	Unigene471	30	9e-52	200	complete
	Unigene705	31	3e-50	195	partial
AtbHLH018 (At2g22750)	Unigene11765	42	3e-54	207	partial
	Unigene21970	53	1e-49	192	partial
	Unigene29516	42	5e-49	190	complete
	Unigene16153	41	2e-37	151	complete
	Unigene1680	48	5e-34	140	partial

[a]Unigenes cloned in this chapter are indicated in bold.
[b]DB, database.

with AtMYC2 and AtMYC4 (56 and 53% amino acid sequence identities, respectively), and Unigene723 also had a comparably high alignment score with AtMYC2 and AtMYC4 (54 and 51% amino acid sequence identities, respectively). I cloned these three unigenes from the cDNA library obtained from roots of *G. uralensis* as candidate TFs, and named them GubHLH1 (Unigene21689), GubHLH2 (Unigene723), and GubHLH3 (Unigene29516). The alignments of these three candidate TFs are shown in Figure 4-5. The phylogenetic relationships of GubHLH1–3 and other putative bHLH TFs in the transcriptome database of *G. uralensis* with characterized bHLH TFs regulating the biosynthesis of specialized metabolites (listed in Table 4-7) were analyzed (Figure 4-6).

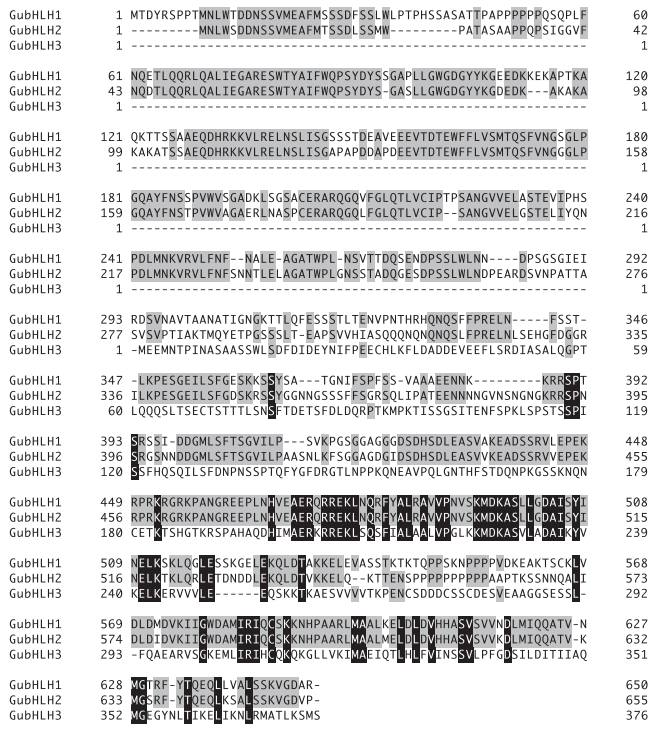


Figure 4-5. Multiple alignments of GubHLH1, GubHLH2, and GubHLH3. Multiple alignments were generated using GENETYX-MAC ver. 18.0.3 software. The fully conserved residues are highlighted with a black background, and residues conserved in at least two of the three TF proteins are highlighted with a gray background.

Table 4-7. A list of previously identified bHLH TFs included in the phylogenetic tree in Figure 4-6

TF name	Species	GenBank protein accession no.	Reference
AaMYC2	*Artemisia annua*	AKO62850	Shen *et al*., 2016
AtEGL3	*Arabidopsis thaliana*	AEE34124	Zhang *et al*., 2003
AtGL3	*Arabidopsis thaliana*	ANM71157	Zhang *et al*., 2003
AtMYC2	*Arabidopsis thaliana*	AEE31513	Schweizer *et al*., 2013
AtMYC3	*Arabidopsis thaliana*	AED95422	Schweizer *et al*., 2013
AtMYC4	*Arabidopsis thaliana*	AEE83960	Schweizer *et al*., 2013
AtTT8	*Arabidopsis thaliana*	ABE66051	Nesi *et al*., 2000
CjbHLH1	*Coptis japonica*	BAJ40865	Yamada *et al*., 2011
CrBIS1	*Catharanthus roseus*	AJE29376	Van Moerkercke *et al*., 2015
CrBIS2	*Catharanthus roseus*	AJE29375	Van Moerkercke *et al*., 2016
CrMYC2	*Catharanthus roseus*	AAQ14332	Zhang *et al*., 2011
Csa5g156220 (Bl)	*Cucumis sativus*	AIT72026	Shang *et al*., 2014
Csa5g157230 (Bt)	*Cucumis sativus*	AIT72027	Shang *et al*., 2014
MtTSAR1	*Medicago truncatula*	AJE29377	Mertens *et al*., 2016
MtTSAR2	*Medicago truncatula*	ALH07115	Mertens *et al*., 2016
NbbHLH1	*Nicotiana benthamiana*	ADH04262	Todd *et al*., 2010
NbbHLH2	*Nicotiana benthamiana*	ADH04263	Todd *et al*., 2010
NtMYC2a	*Nicotiana tabacum*	ADH04269	Shoji and Hashimoto, 2011
NtMYC2b	*Nicotiana tabacum*	ADH04270	Shoji and Hashimoto, 2011

4-3-2. Transient co-transfection assays of candidate TFs

To find potential transactivation of triterpenoid or sterol biosynthetic genes of *G. uralensis* by these candidate TFs, I performed transient co-transfection assays with these candidate TFs against promoter sequences of three OSC and four P450 genes in *G. uralensis* (Figure 4-7). Promoter regions of each

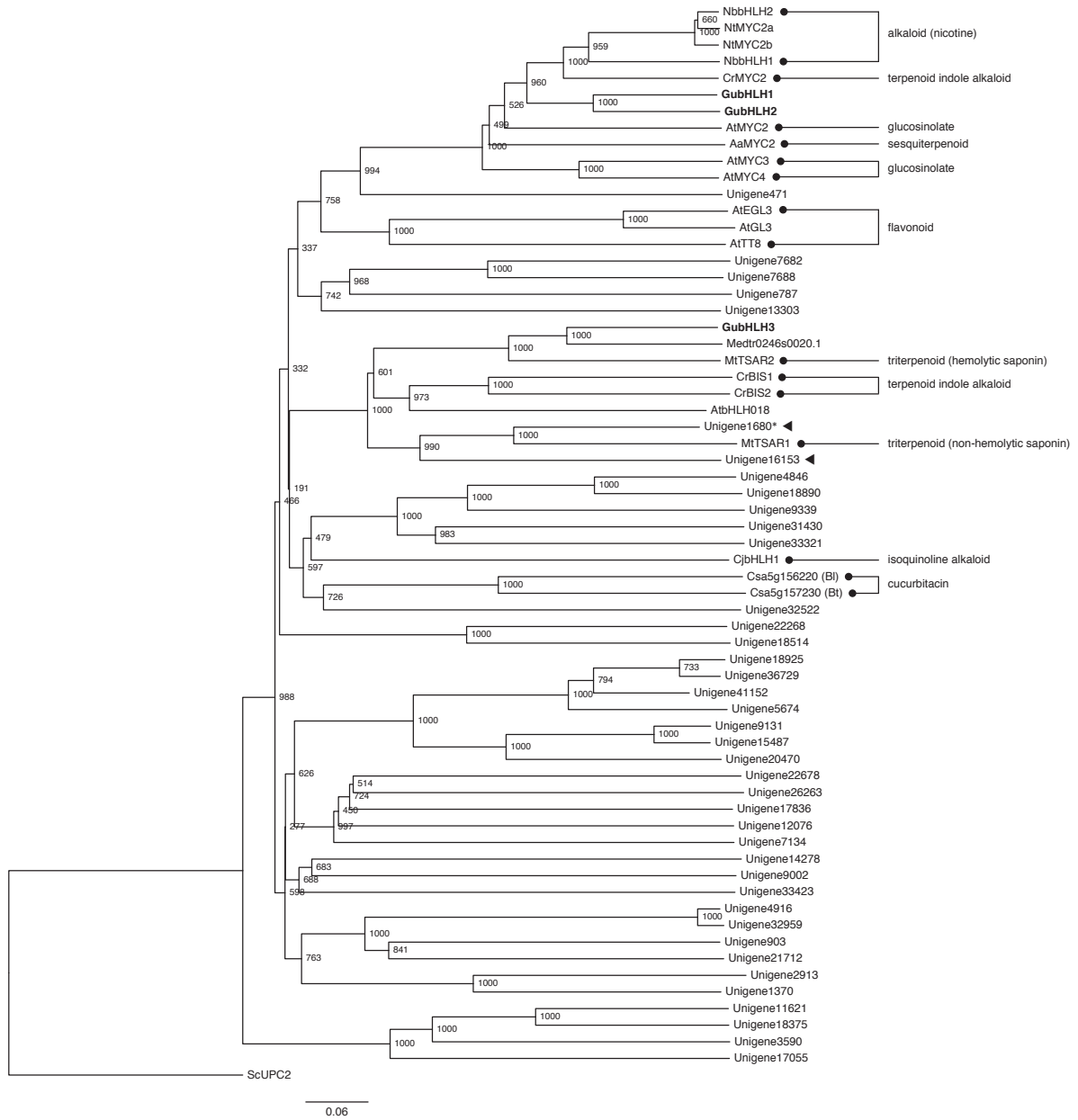


Figure 4-6. Phylogenetic tree of representative bHLH TFs involved in the biosynthesis of specialized metabolites and putative bHLH TFs in *G. uralensis*. Putative bHLH TFs in *G. uralensis* (indicated by unigene numbers) were identified in the previous transcriptome analysis of *G. uralensis* (Ramilowski *et al.*, 2013). A phylogenetic tree was constructed based on full-length protein sequences except for Unigene1680 (indicated by an asterisk), for which only a partial CDS was predicted. Reported classification of metabolites regulated by each TF is indicated at the right of TF names. GubHLH1–3 are indicated in bold, and Unigene1680 and Unigene16153 are indicated by filled arrowheads. For reference, Medtr0246s0020.1 and AtbHLH018 are also included in this phylogenetic tree. Numbers indicate bootstrap values for 1,000 replicates. The scale bar shows the amino acid substitution ratio. GenBank protein accession numbers and species of previously characterized TFs in this tree are indicated in Table 4-7. *Saccharomyces cerevisiae* UPC2 (GenBank protein accession number EDN60549) was used as an outgroup.

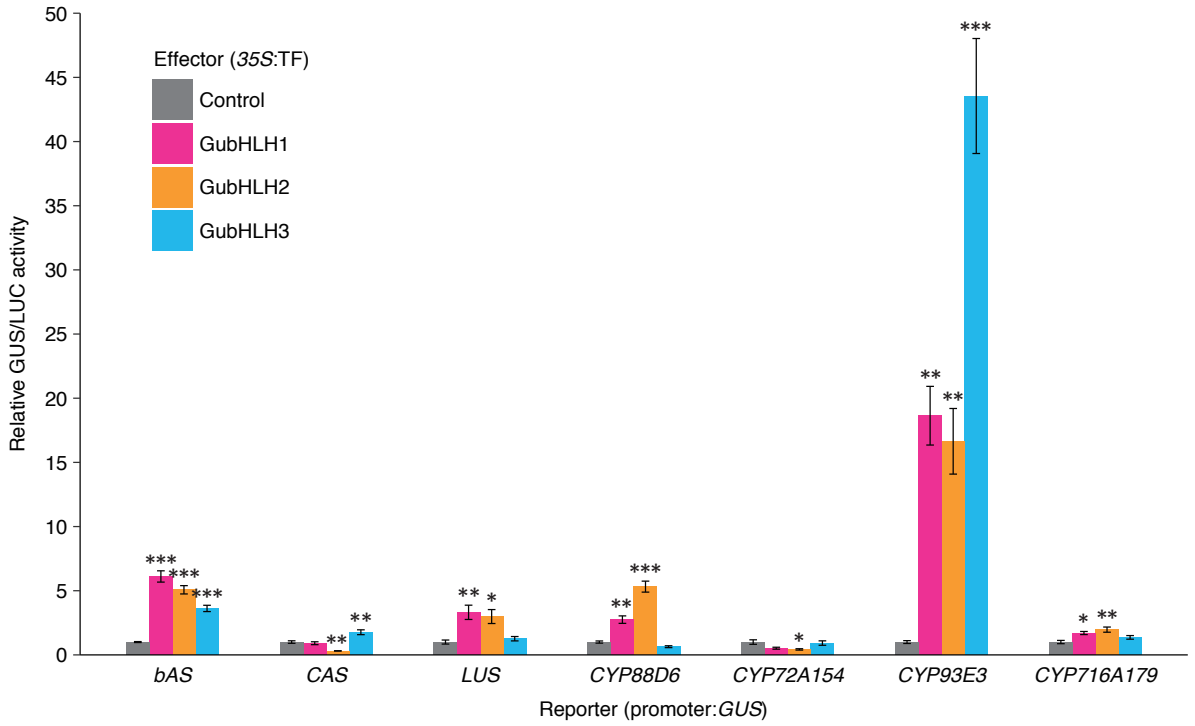


Figure 4-7. Transient co-transfection assays of promoters of triterpenoid biosynthetic genes of *G. uralensis* and GuHLH1–3. Error bars indicate the standard error (SE) of four transfections. Asterisks indicate significant differences between the control and each effector determined by Dunnett's test (**P* < 0.05, ***P* < 0.01, ****P* < 0.001).

gene (1.8–2.5 kb; Table 4-3) were fused with a *GUS* reporter gene and transiently introduced into protoplasts isolated from tobacco BY-2 cells, together with an effector construct (*35S*:TF) and an internal control (*35S*:*LUC*). GubHLH1 and GubHLH2 significantly transactivated *bAS*, *LUS*, *CYP88D6*, *CYP93E3*, and *CYP716A179* promoters compared with a control vector. Transactivation of the *CYP93E3* promoter by GubHLH1 and GubHLH2 was 19-fold and 17-fold, respectively. GubHLH3 also significantly transactivated *bAS* (3.6-fold) and *CYP93E3* (44-fold) promoters, but did not transactivate *LUS*, *CYP88D6*, *CYP72A154,* or *CYP716A179* promoters. *CAS* promoter was also significantly transactivated by GubHLH3 (1.8-fold), however, the fold change was much lower than the transactivation of the *CYP93E3* promoter. Therefore, transactivation by GubHLH3 was highly selective for promoters of soyasaponin biosynthetic genes. Among all combinations of promoters and TFs tested in this chapter, the highest transactivation was observed between GubHLH3 and the *CYP93E3* promoter (44-fold).

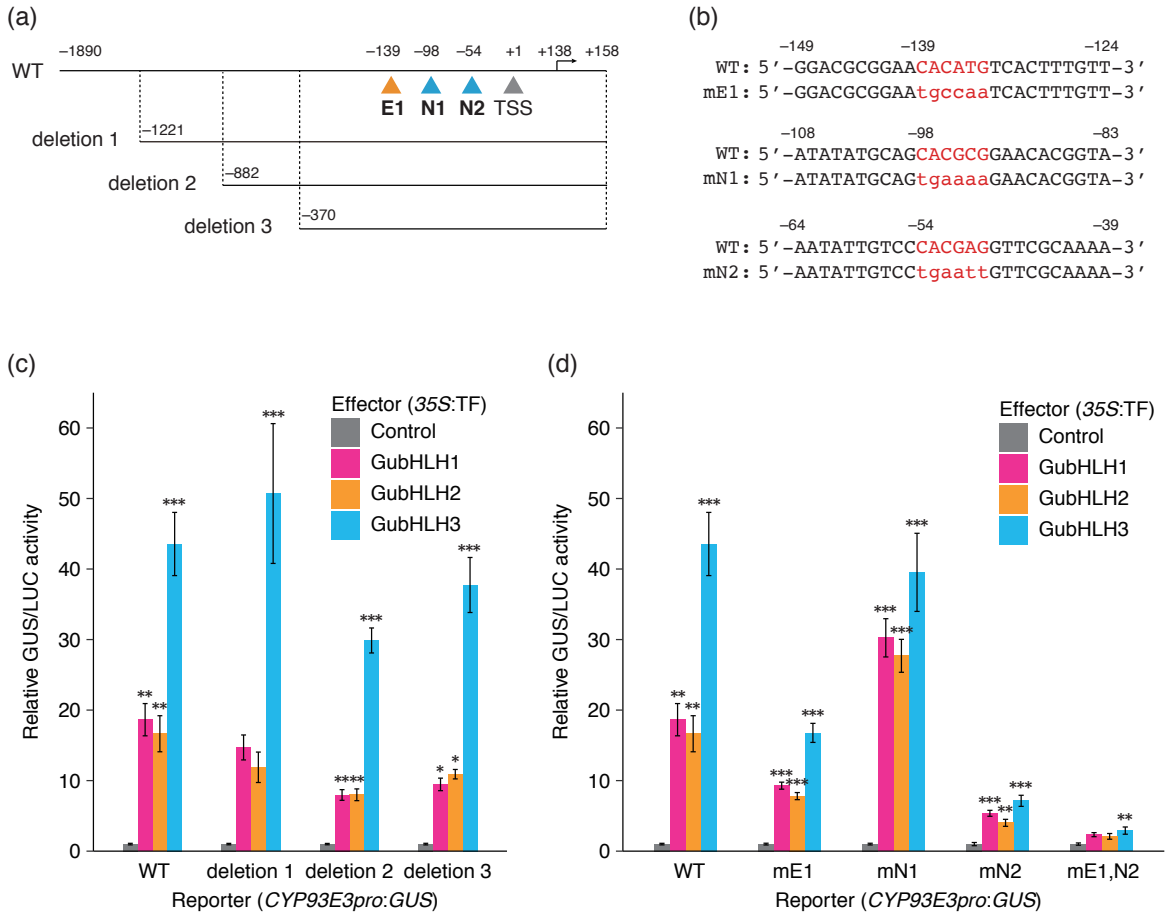


Figure 4-8. Determination of binding sites of GubHLH1–3 on *CYP93E3* promoter.
(a) Deletion constructs of *CYP93E3* promoter and the location of E-box (E1) and N-box (N1 and N2) motifs. Numbers indicate the position relative to the transcriptional start site (TSS; +1). Sequence between +138 and +158 is the coding region of *CYP93E3*.
(b) Mutations at E1, N1, and N2 introduced in this chapter.
(c) Transient co-transfection assays of deletion series of *CYP93E3* promoter and GubHLH1–3.
(d) Transient co-transfection assays of mutated *CYP93E3* promoter and GubHLH1–3.
Error bars indicate the SE of four transfections. Asterisks indicate significant differences between the control and each effector determined by Dunnett's test (**P* < 0.05, ***P* < 0.01, ****P* < 0.001). The values for WT of the *CYP93E3* promoter are the same as those described in Figure 4-7.

4-3-3. Identification of GubHLH3 binding sites on the *CYP93E3* promoter

The transient co-transfection assays indicated that GubHLH3 preferentially transactivates the *CYP93E3* promoter. To narrow down the binding sites of GubHLH3 on the *CYP93E3* promoter, I first generated deletion constructs of the *CYP93E3* promoter and performed transient co-transfection assays (Figures 4-8a,c). GubHLH3 transactivated all three deletion constructs by more than 30-fold (Figure 4-8c). Therefore, putative binding sites were predicted to be downstream of position −370 (relative to the transcription start site, TSS) on the *CYP93E3* promoter (Figure 4-8a).

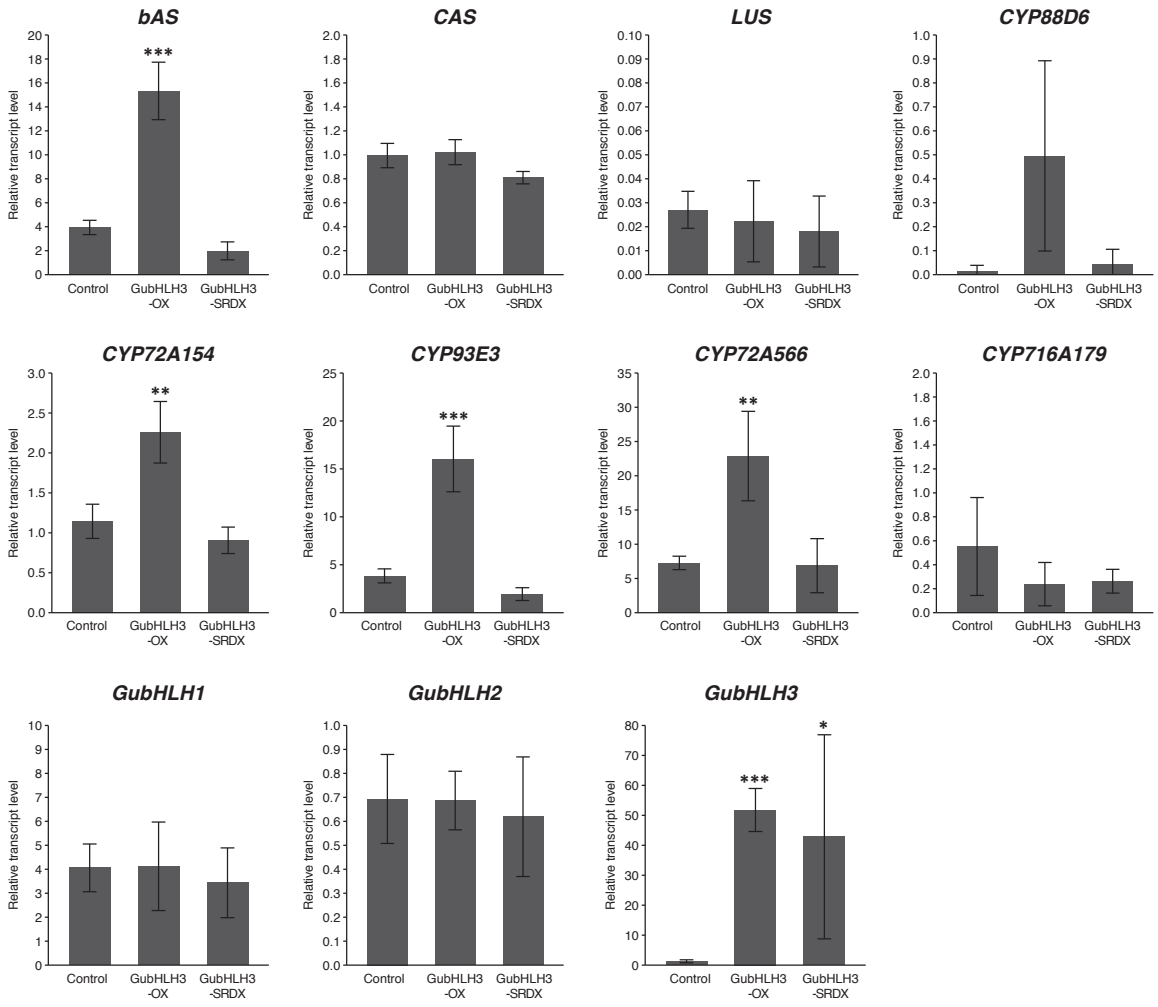


Figure 4-9. Quantitative real-time PCR (qPCR) analysis of triterpenoid biosynthetic genes and *GubHLH1–3* in transgenic hairy roots. Error bars indicate the standard deviation (SD) of four (Control), five (GubHLH3-OX), and two (GubHLH3-SRDX) independent hairy root lines. Asterisks indicate significant differences between control lines and GubHLH3-OX or GubHLH3-SRDX lines determined by Dunnett's test (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$).

bHLH TFs are known to bind to the E-box motif (CANNTG) (Toledo-Ortiz *et al.*, 2003). Moreover, GubHLH3 is similar to TSAR2 in *M. truncatula* (53% amino acid identity) and the N-box motif (CACGCG or CACGAG) (Pires and Dolan, 2010) is reportedly recognized by TSAR2 (Mertens *et al.*, 2016). Therefore, I focused on E-box and N-box motifs on the shortest promoter fragment (deletion 3) of *CYP93E3* and introduced mutations into these motifs (Figures 4-8a,b). In transient co-transfection assays using these mutated promoter fragments and *35S*:*GubHLH3*, mutations at E1 and N2 reduced transactivation by GubHLH3 to 18-fold and 8-fold, respectively, but mutation at N1 showed almost the same level of transactivation as the wild-type promoter (44-fold) (Figure 4-8d). When both E1 and N2 motifs were mutated simultaneously, the level of transactivation dropped to

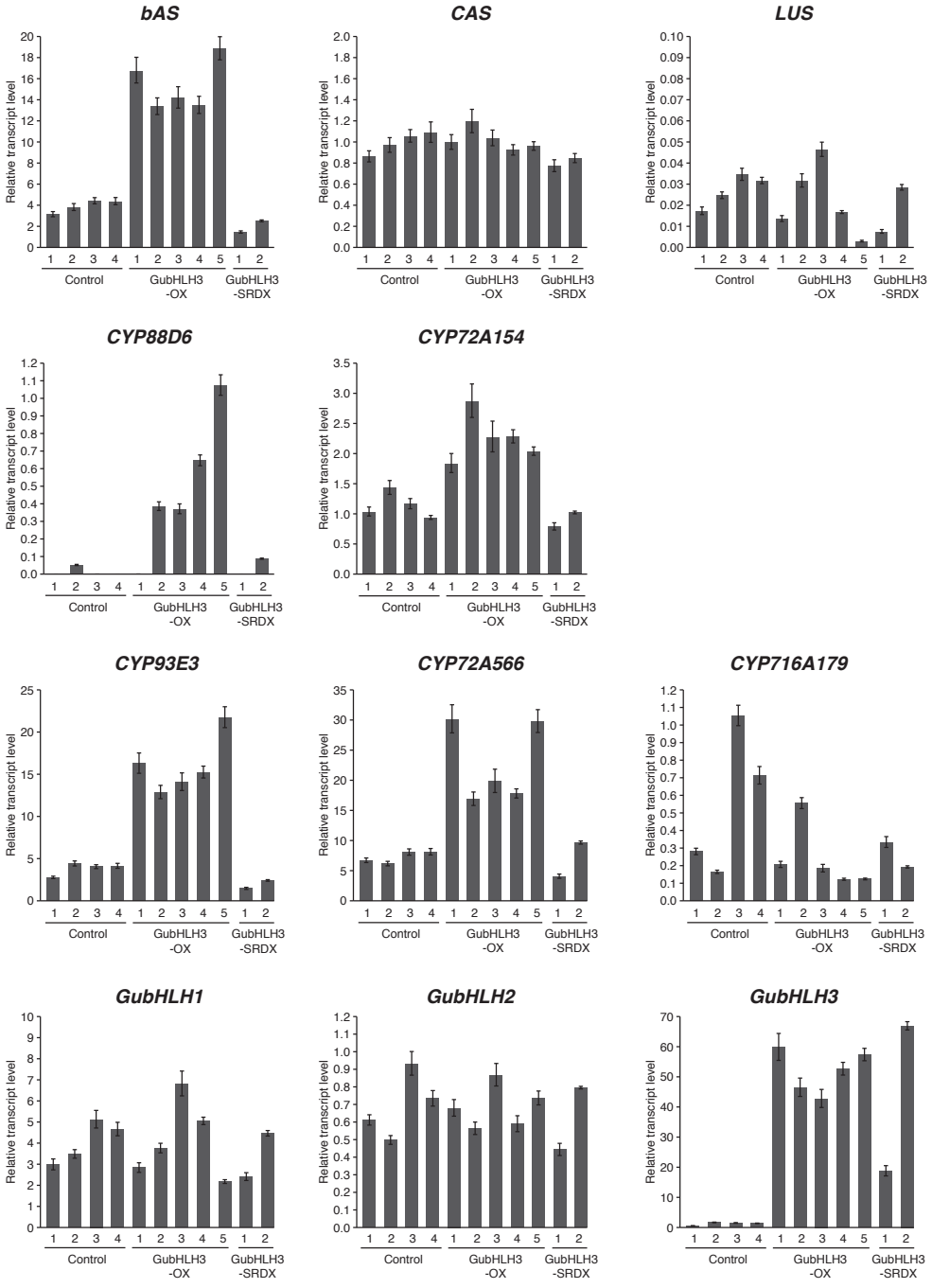


Figure 4-10. qPCR analysis of triterpenoid biosynthetic genes and *GubHLH1–3* in individual lines of transgenic hairy roots. Error bars indicate SE of three technical replicates.

3.4-fold. Therefore, E-box (E1) and N-box (N2) motifs are predicted to be involved in the binding of GubHLH3 to the *CYP93E3* promoter. Similar results were obtained when GubHLH1 or GubHLH2 were used as effector constructs (Figures 4-8c,d), suggesting that the E-box (E1) and the N-box (N2) are necessary for the binding of these TFs to the *CYP93E3* promoter.

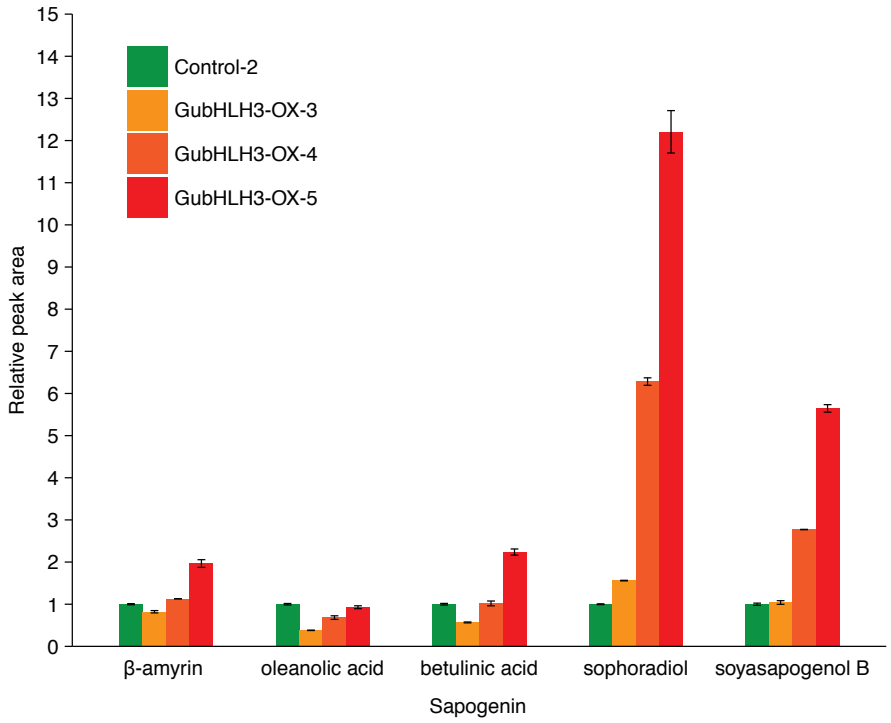


Figure 4-11. Relative quantification of sapogenins accumulated in transgenic hairy root lines. Analysis was performed for one control line (Control-2) and three GubHLH3-OX lines (GubHLH3-OX-3, 4, and 5). Line numbers correspond to qPCR analysis of individual lines shown in Figure 4-10. The peak area of each compound was normalized to an internal standard (echinocystic acid) and values are shown relative to the value of Control-2. Error bars indicate SE of two technical replicates. In addition to the indicated compounds, a trace amount of lupeol was identified in the three GubHLH3-OX lines.

4-3-4. Overexpression of *GubHLH3* enhances the expression of soyasaponin biosynthetic genes

To functionally characterize GubHLH3 in *G. uralensis*, I generated transgenic hairy roots overexpressing *GubHLH3* under the cauliflower mosaic virus (CaMV) *35S* promoter. I obtained four empty vector control lines (Control) and five *GubHLH3* overexpression (GubHLH3-OX) lines, and analyzed the transcript levels of triterpenoid or sterol biosynthetic genes and *GubHLH1–3* by qPCR (Figure 4-9). In GubHLH3-OX lines, the transcript levels of *bAS*, *CYP93E3*, and *CYP72A566*, the three genes required for the biosynthesis of soyasapogenol B from 2,3-oxidosqualene, were all significantly enhanced compared with control lines. The relative transcript levels of *bAS*, *CYP93E3*, and *CYP72A566* in GubHLH3-OX lines were 3.9, 4.2, and 3.1 times higher than in control lines, respectively. In addition to these soyasaponin biosynthetic genes, *CYP72A154* was also significantly upregulated to approximately twice the level of the control lines. The transcript level of *CYP88D6* in GubHLH3-OX lines was more than 30 times higher than in the control lines; however, the transcript

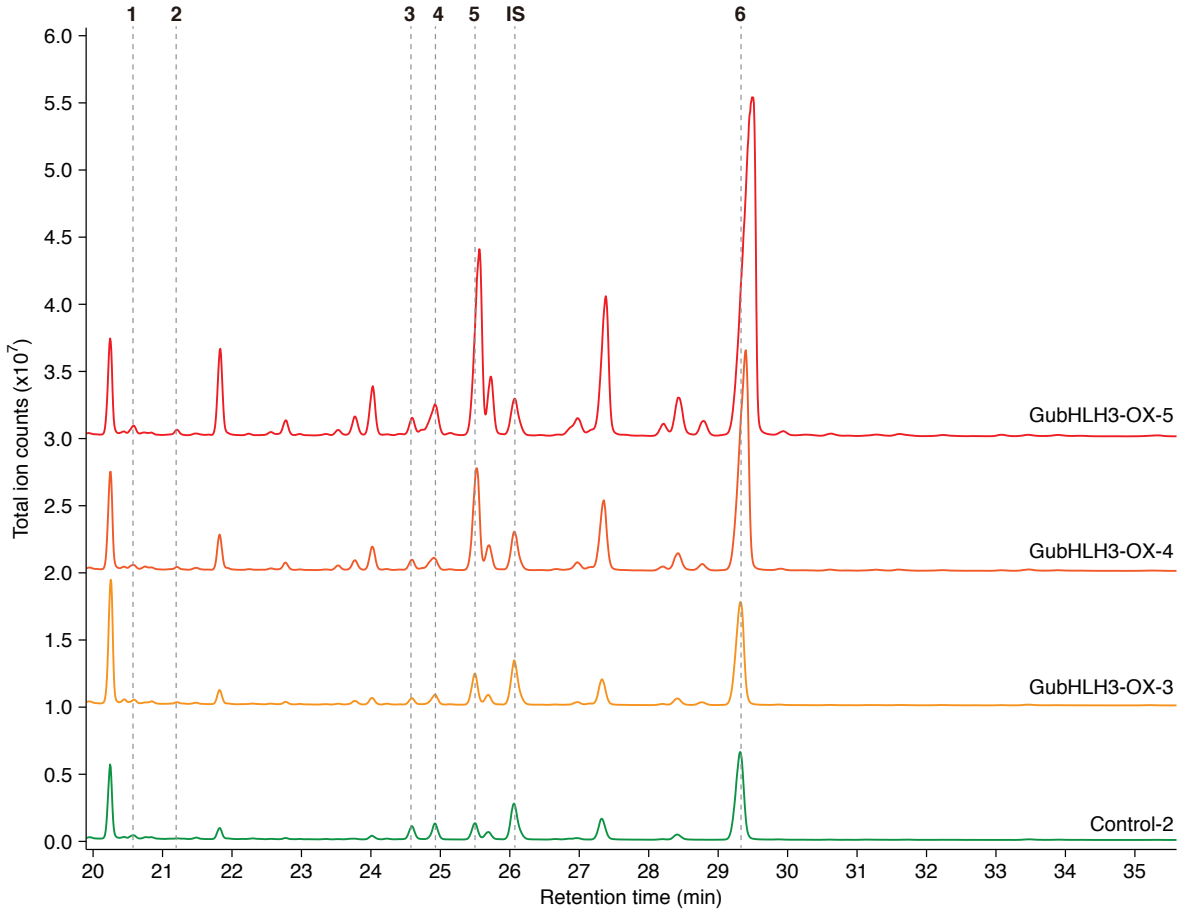


Figure 4-12. Gas chromatography-mass spectrometry (GC-MS) analysis of sapogenins isolated from transgenic hairy roots. Analysis was performed for one control line (Control-2) and three GubHLH3-OX lines (GubHLH3-OX-3, 4, and 5). Line numbers correspond to qPCR analysis of individual lines shown in Figure 4-10. β-Amyrin (**1**), oleanolic acid (**3**), betulinic acid (**4**), sophoradiol (**5**), and soyasapogenol B (**6**) were identified in all hairy root lines, and lupeol (**2**) was identified in the three GubHLH3-OX lines. Echinocystic acid was added as an internal standard (IS). 24-Hydroxy-β-amyrin, α-amyrin, ursolic acid, 11-oxo-β-amyrin, 30-hydroxy-β-amyrin, 11-deoxoglycyrrhetinic acid, and glycyrrhetinic acid were not detected from any of the chromatograms. To avoid overlapping chromatograms, the baselines for GubHLH3-OX-3, GubHLH3-OX-4, and GubHLH3-OX-5 were moved 1.0×10^7, 2.0×10^7, and 3.0×10^7 total ion counts, respectively.

Table 4-8. Quantification of sapogenins from transgenic hairy root lines

Sapogenin	Content of sapogenin (μg/g of dry weight)			
	Control-2	GubHLH3-OX-3	GubHLH3-OX-4	GubHLH3-OX-5
β-amyrin	24.95 ± 0.36	20.74 ± 0.57	28.27 ± 0.19	49.51 ± 2.12
lupeol	nd	9.48 ± 2.25	18.99 ± 1.08	37.70 ± 0.65
oleanolic acid	105.8 ± 1.73	40.71 ± 0.72	72.50 ± 4.87	99.18 ± 3.25
betulinic acid	264.3 ± 4.83	152.0 ± 3.90	270.4 ± 16.45	597.3 ± 17.31
sophoradiol	318.6 ± 2.72	503.4 ± 5.35	2009 ± 18.28	3928 ± 147.1
soyasapogenol B	1470 ± 36.65	1548 ± 54.60	4089 ± 13.82	8379 ± 101.7

Values are mean ± SE of two technical replicates calculated from the peak areas of each compound and the internal standard (echinocystic acid). nd, not detected.

levels of *CYP88D6* and of *GubHLH3* did not seem to be correlated with each other in individual hairy root lines (Figure 4-10). Therefore, it is unclear whether GubHLH3 upregulates the expression of *CYP88D6*. The transcript levels of the two remaining triterpenoid biosynthetic genes, *LUS* and *CYP716A179*, and the sterol biosynthetic gene *CAS* were not enhanced in GubHLH3-OX lines compared with control lines (Figure 4-9). I also generated two lines expressing a chimeric repressor of GubHLH3 (GubHLH3-SRDX), and analyzed the transcript levels of triterpenoid or sterol biosynthetic genes and *GubHLH1–3* (Figure 4-9). The SRDX domain is a modified plant-specific EAR-motif repression domain consisting of 12 amino acid residues (LDLDLELRLGFA; Hiratsu *et al.*, 2003). The fusion of a TF to the SRDX domain converts transcriptional activators into dominant repressors, allowing a loss-of-function analysis of the TF (Hiratsu *et al.*, 2003; Mitsuda and Ohme-Takagi, 2009). The transcript levels of *bAS* and *CYP93E3* in GubHLH3-SRDX lines were slightly lower than those of control lines; however, the *CYP72A566* level was nearly the same as the control lines (Figure 4-9).

To elucidate whether the enhanced expression of soyasaponin biosynthetic genes in GubHLH3-OX hairy root lines affects metabolite accumulation, I compared the relative quantities of sapogenins accumulated in the control line and GubHLH3-OX hairy root lines (Figure 4-11). Compared with the control line, the relative amounts of soyasapogenol B and sophoradiol, a possible intermediate between β-amyrin and soyasapogenol B, were more than doubled in two of the three GubHLH3-OX lines. Especially, GubHLH3-OX-5 accumulated 5.6 times more soyasapogenol B and 12 times more sophoradiol than the control line. The quantity of soyasapogenol B in GubHLH3-OX-5 was estimated to be 8.4 mg per 1 g of dry weight (Table 4-8). On the other hand, the amount of oleanolic acid in the GubHLH3-OX lines was less of the amount in the control line, suggesting that oleanolic acid biosynthesis competes with soyasaponin biosynthesis. I did not detect glycyrrhetinic acid or possible intermediates between β-amyrin and glycyrrhetinic acid, even in the GubHLH3-OX lines (Figure 4-12). Therefore, the observed increase in the expression level of *CYP88D6* in several GubHLH3-OX lines (Figure 4-10) is not likely to contribute to glycyrrhizin biosynthesis.

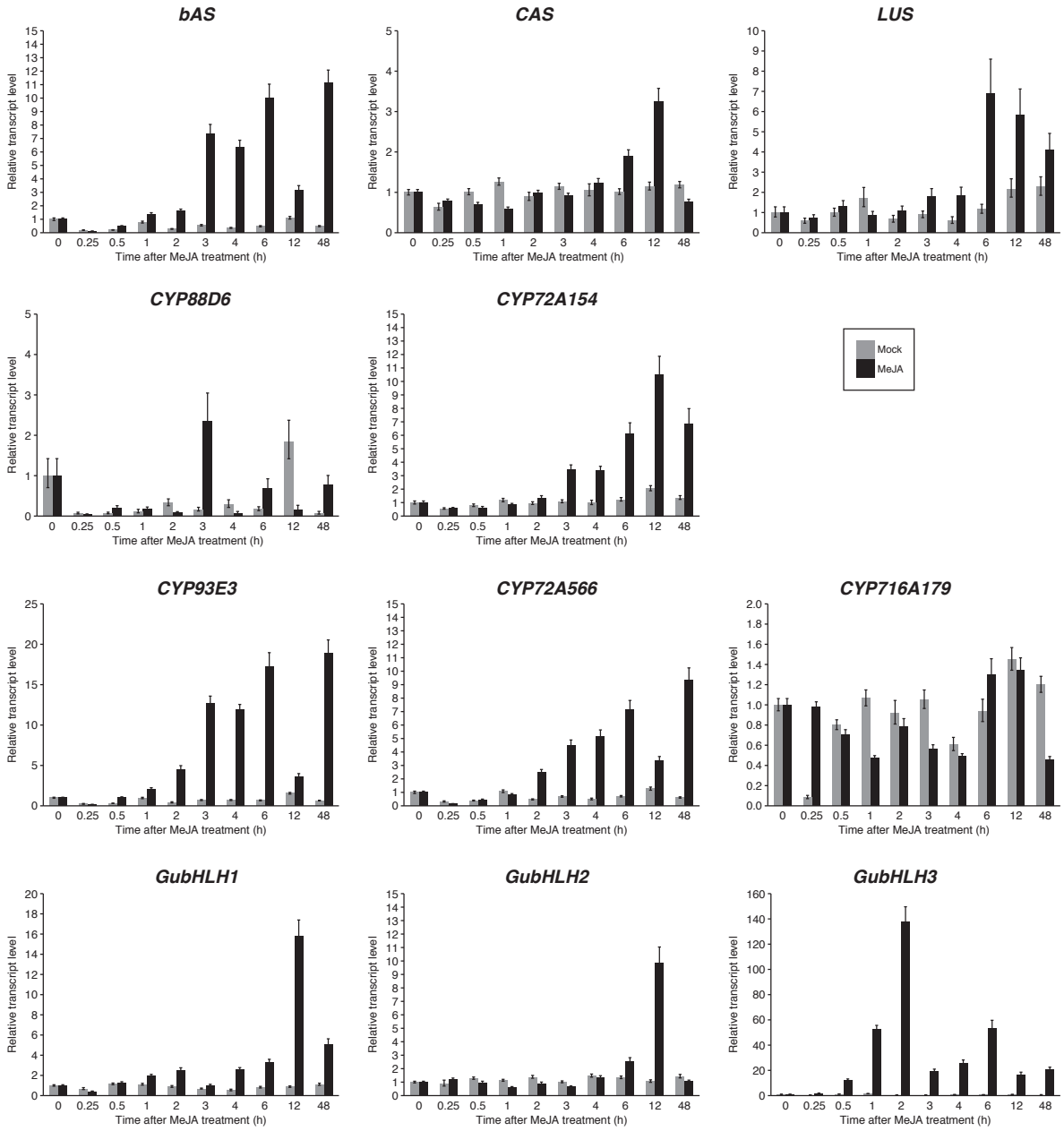


Figure 4-13. Time-course qPCR analysis of triterpenoid biosynthetic genes and *GubHLH1–3* after methyl jasmonate (MeJA) treatment. Tissue-cultured stolons were treated with 0.1% ethanol (mock) or 100 μM MeJA. Relative transcript levels at 0 h were set equal to 1. Error bars indicate SE of three technical replicates.

4-3-5. Expression of soyasaponin biosynthetic genes and *GubHLH3* is induced by methyl jasmonate treatment

The biosynthesis of soyasaponins in cultured cells of *G. glabra* was upregulated following the application of exogenous MeJA (Hayashi *et al.*, 2003). To analyze the effects of exogenously applied MeJA on the expression of triterpenoid or sterol biosynthetic genes and *GubHLH1–3*, I treated tissue-cultured stolons of *G. uralensis* with MeJA and analyzed the time-course transcriptional

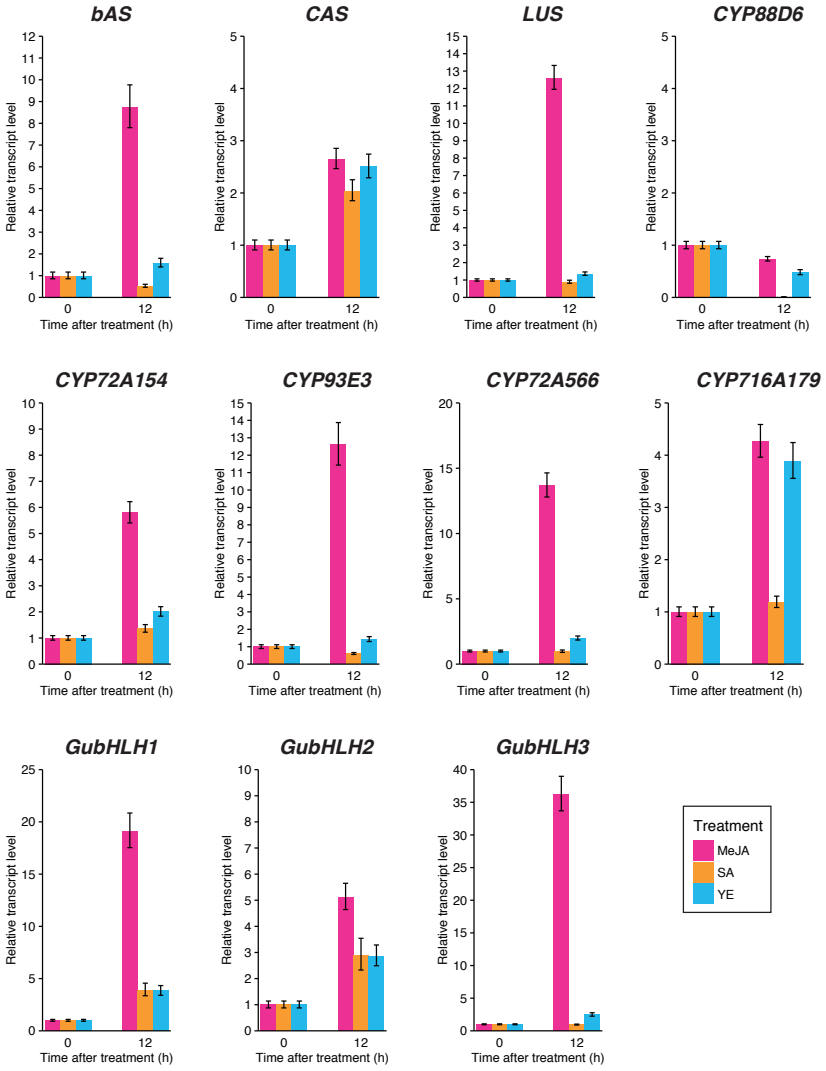


Figure 4-14. qPCR analysis of triterpenoid biosynthetic genes and *GubHLH1–3* in tissue-cultured stolons treated with methyl jasmonate (MeJA), salicylic acid (SA), or yeast extract (YE). Relative transcript levels at 0 h were set equal to 1. Error bars indicate SE of three technical replicates.

response of these genes by qPCR (Figure 4-13). The three soyasaponin biosynthetic genes, *bAS*, *CYP93E3*, and *CYP72A566*, were all upregulated by MeJA, and the three exhibited similar responses with each other. They were strongly upregulated 3 h after MeJA treatment, then their expression levels dropped 12 h after MeJA treatment. Other triterpenoid biosynthetic genes and *CAS* showed different responses from those of the soyasaponin biosynthetic genes, suggesting that *bAS*, *CYP93E3*, and *CYP72A566* are coordinately regulated under MeJA treatment. The expression of *GubHLH3* was induced 0.5−1 h after MeJA treatment, peaking 2 h after treatment (138-fold). This earlier response of *GubHLH3* to MeJA than the responses of the soyasaponin biosynthetic genes suggests that GubHLH3

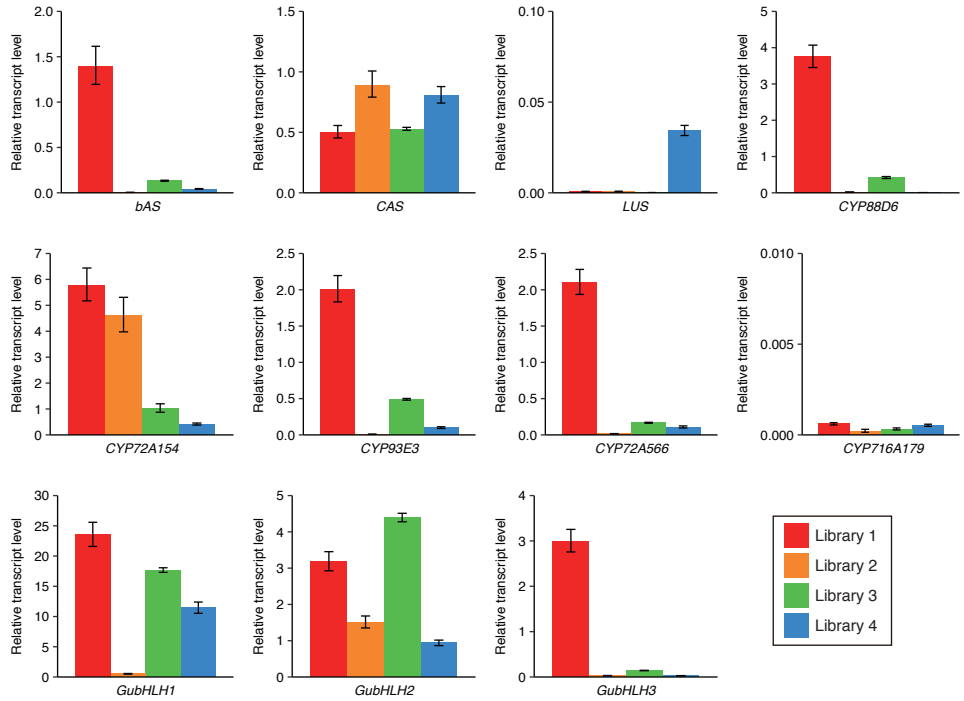


Figure 4-15. qPCR analysis of triterpenoid biosynthetic genes and *GubHLH1–3* in intact *G. uralensis* plants. Library 1, roots of strain 308-19 (glycyrrhizin [GL] high-producing) harvested in summer; Library 2, roots of strain 308-19 harvested in winter; Library 3, roots of strain 87-458 (GL low-producing) harvested in summer; Library 4, leaves of strain 308-19 harvested in summer. The RNA samples were the same as the samples used in the previous transcriptome analysis (Ramilowski *et al.*, 2013). Data on *bAS*, *CAS*, *LUS*, *CYP88D6*, *CYP72A154*, *CYP93E3*, *CYP72A566*, and *CYP716A179* in Library 1 (intact roots) have already described in Chapter 2. Error bars indicate SE of three technical replicates.

stimulates the expression of downstream soyasaponin biosynthetic genes in response to MeJA treatment. On the other hand, *GubHLH1* and *GubHLH2* were upregulated 12 h after MeJA treatment (16-fold and 9.9-fold, respectively). This is much later than the upregulation of soyasaponin biosynthetic genes; moreover, their fold change was much lower than that of *GubHLH3*. To analyze the effects of other elicitors, I applied YE or SA to tissue-cultured *G. uralensis* stolons, and analyzed the transcriptional responses (Figure 4-14). In previous studies, YE was shown to suppress the production of soyasaponins in cultured cells of *G. glabra* (Hayashi *et al.*, 2005), and SA was shown to have little effect on the expression of *bAS* in *M. truncatula* (Suzuki *et al.*, 2005). Therefore, these two elicitors were not expected to upregulate the soyasaponin biosynthetic genes. The results supported this prediction, as these two elicitors failed to upregulate the expression of the soyasaponin biosynthetic genes or *GubHLH3* (Figure 4-14). From these results, I suggest that GubHLH3 controls MeJA-responsive expression of soyasaponin biosynthetic genes in *G. uralensis*.

4-3-6. Expression of triterpenoid biosynthetic genes and *GubHLH1–3* in intact plants

To compare the expression patterns of *GubHLH3* and soyasaponin biosynthetic genes in field-grown intact *G. uralensis* plants, I performed qPCR analysis using RNA samples from the previously reported transcriptome analysis of *G. uralensis* (Figure 4-15). This transcriptome analysis was performed in four libraries to examine the differences between organs (roots or leaves), seasons (summer or winter), and strains (glycyrrhizin high-producing or low-producing strains), which were predicted to affect the expression of glycyrrhizin biosynthetic genes (Ramilowski *et al.*, 2013). The expression levels of the three soyasaponin biosynthetic genes were particularly high in the roots of the glycyrrhizin high-producing strain harvested in summer (Figure 4-15). This expression pattern is very similar to that of *GubHLH3*, further supporting the conclusion that soyasaponin biosynthetic genes are under the control of GubHLH3. In contrast, *GubHLH1* and *GubHLH2* were expressed at substantial levels in the other samples, suggesting that these two TFs are not responsible for the regulation of soyasaponin biosynthetic genes.

4-4. Discussion

Plants produce various classes of specialized metabolites, including triterpenoids. Although recent studies have revealed various enzyme genes involved in plant triterpenoid biosynthesis, their regulatory mechanisms remain largely unknown. Here, I identified GubHLH3, the bHLH TF regulating the expression of *bAS*, and the two P450 genes required for the biosynthesis of soyasaponins, in an important medicinal plant, *G. uralensis*.

The regulation of soyasaponin biosynthetic genes by GubHLH3 in *G. uralensis* is supported by a series of experimental results. First, in transient co-transfection assays of promoter:*GUS* and *35S*:TF constructs, GubHLH3 significantly transactivated *bAS* and *CYP93E3* promoters, but not the promoters of other triterpenoid biosynthetic genes (*CYP88D6* and *CYP72A154* for glycyrrhizin biosynthesis, *CYP716A179* for oleanolic acid and betulinic acid biosynthesis, and *LUS* for betulinic acid biosynthesis) (Figure 4-7). Second, in GubHLH3-OX hairy root lines of *G. uralensis*, the transcript levels of soyasaponin biosynthetic genes (*bAS*, *CYP93E3*, and *CYP72A566*)

were clearly enhanced in all five GubHLH3-OX lines compared with control lines (Figures 4-9 and 4-10). Metabolite analysis of representative transgenic hairy root lines showed higher levels of soyasapogenol B and sophoradiol in GubHLH3-OX lines compared with the control line (Figure 4-11). Third, expression of *GubHLH3* was strongly upregulated by the plant hormone MeJA, and the expression levels of *bAS*, *CYP93E3*, and *CYP72A566* were also strongly upregulated by MeJA, with similar responses among the three (Figure 4-13). Furthermore, the expression patterns of soyasaponin biosynthetic genes and *GubHLH3* in intact *G. uralensis* plants (Figure 4-15) were similar.

CYP88D6 is a key enzyme in the production of glycyrrhizin in *G. uralensis* (Seki *et al.*, 2008). In the qPCR analysis of GubHLH3-OX hairy root lines, four of the five lines showed higher transcription levels of *CYP88D6* compared with the control lines. However, the transcription levels of *GubHLH3* and *CYP88D6* did not correlate with each other (Figure 4-10). Moreover, the expression pattern of *CYP88D6* after MeJA treatment was completely different from that of *GubHLH3* (Figure 4-13), and *CYP88D6* and *GubHLH3* also showed different responses to SA or YE (Figure 4-14). Therefore, it is not reasonable to suggest that GubHLH3 directly regulates *CYP88D6* expression. There is a possibility that *GubHLH3* overexpression indirectly affects the expression of *CYP88D6*. The other two bHLH TFs cloned in this chapter, GubHLH1 and GubHLH2, showed significant transactivation against the *CYP88D6* promoter (Figure 4-7). However, their responses to MeJA, SA, and YE differed markedly from those of *GubHLH3* (Figures 4-13 and 4-14), and the expression patterns in intact plants were also quite different (Figure 4-15). I did not detect glycyrrhetinic acid or possible intermediates between β-amyrin and glycyrrhetinic acid in any of the three GubHLH3-OX lines in the metabolite analysis (Figure 4-12). Further screening, such as a comprehensive gene co-expression analysis of licorice samples, is necessary to find the regulators of glycyrrhizin biosynthetic genes including *CYP88D6*.

Both licorice and *M. truncatula* produce soyasaponins, which have soyasapogenol B as their aglycone (Hayashi *et al.*, 1990; Huhman *et al.*, 2005; Carelli *et al.*, 2011). The three enzymes required for the biosynthesis of soyasapogenol B from 2,3-oxidosqualene in licorice (bAS, CYP93E3, and CYP72A566) are the corresponding orthologs of those in *M. truncatula* (bAS, CYP93E2, and CYP72A61v2) (Figure 2-6). In *M. truncatula*, bHLH TF TSAR1 is reported to boost soyasaponin

(non-hemolytic saponin) biosynthesis and activate the expression of relevant enzyme genes, *bAS*, *CYP93E2*, and *CYP72A61v2*, whereas TSAR2 is reported to boost hemolytic saponin biosynthesis, and activate the expression of relevant enzyme genes, *bAS*, *CYP716A12* and *CYP72A68v2* (Mertens *et al.*, 2016). Interestingly, phylogenetic analysis indicated that GubHLH3 is phylogenetically closer to TSAR2 (53% amino acid sequence identity) than TSAR1 (43% amino acid sequence identity) (Figure 4-6). However, GubHLH3 activated the expression of *bAS*, *CYP93E3*, and *CYP72A566*, the three corresponding orthologous genes of which were activated by TSAR1 in *M. truncatula*. Therefore, although biosynthetic enzymes for the production of soyasaponins are shared between *M. truncatula* and *G. uralensis*, the regulatory mechanism of soyasaponin biosynthesis is not conserved in these two species. By conducting phylogenetic analysis of putative bHLH TFs encoded by unigenes identified in the previous *G. uralensis* transcriptome analysis (Ramilowski *et al.*, 2013), I found one putative ortholog of TSAR1 (Unigene1680) and another phylogenetically close unigene (Unigene16153) (Figure 4-6). Further studies on the comparative analysis of these TFs as well as GubHLH3 may provide a better understanding of the evolutionary mechanisms of the regulation of triterpenoid biosynthesis in legumes.

Chapter 5

General conclusion

In this study, I focused on two medicinal plants, *Glycyrrhiza uralensis* and *Platycodon grandiflorus*, and identified several P450 enzymes and transcription factors (TFs) involved in triterpenoid biosynthesis. Both plants are used as major crude drugs in Kampo medicine and contain triterpenoid saponins as their major bioactive constituents. Glycyrrhizin in *G. uralensis* and platycodin D in *P. grandiflorus* are oleanane-type triterpenoid saponins. These triterpenoids are commonly derived from β-amyrin and subsequently oxidized at several positions; ultimately, they are glycosylated to form each triterpenoid saponin. *G. uralensis* belongs to the Fabaceae family (order Fabales) while *P. grandiflorus* belongs to the Campanulaceae family (order Asterales). According to the recently updated Angiosperm Phylogeny Group classification (APG IV), branching of these two orders is estimated to have occurred at a relatively early stage during the evolution of eudicots (Figure 5-1). This study elucidated more detailed mechanisms of multiple triterpenoid biosynthetic pathways within one species (*G. uralensis*) and evolutionary aspects of triterpenoid biosynthetic machinery between phylogenetically distinct plant species.

In Chapter 2, I identified two P450s involved in triterpenoid biosynthesis in *G. uralensis*. Although the functions of CYP716A179 and CYP72A566 could be predicted from previously identified putative orthologs in the model legume *Medicago truncatula*, this discovery completed the identification of P450s involved in the biosynthesis of major triterpenoids found in licorice plants. Moreover, the results presented in Chapter 2 show that the functions of CYP716A and CYP72A subfamily enzymes in *G. uralensis* and *M. truncatula*, as well as previously identified CYP93E subfamily enzymes, are highly conserved.

In Chapter 3, I identified two P450s in *P. grandiflorus* catalyzing oxidation reactions of β-amyrin. Platycodigenin, an aglycone of platycodin D, is predicted to be synthesized by oxidation reactions at C-2β, C-16α, C-23, C-24, and C-28 positions of β-amyrin. CYP716A140v2 showed C-28 oxidase activity, as predicted from its amino acid sequence similarity with previously identified C-28

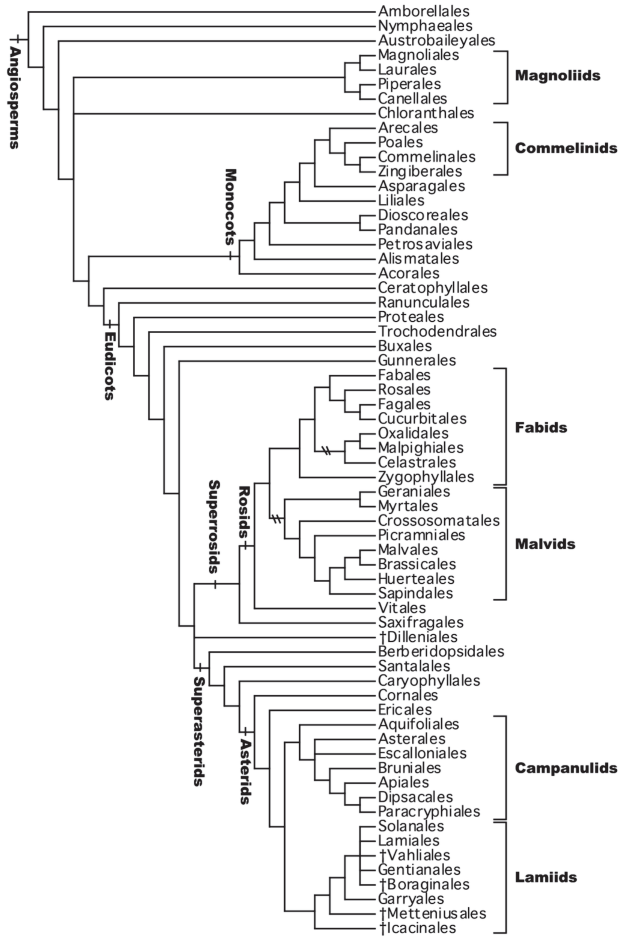


Figure 5-1. Interrelationships of the APG IV orders and some families supported by jackknife/bootstrap percentages >50 or Bayesian posterior probabilities >0.95 in large-scale analyses of angiosperms. The alternative placements representing incongruence between nuclear/mitochondrial and plastid results for the Celastrales/Oxalidales/Malpighiales (COM) clade are indicated by slash marks (\\). †Orders newly recognized in APG. This tree is reproduced from APG IV, 2016 (Figure 1 in the original paper), with permission from Oxford University Press. According to the original paper (APG IV, 2016), these relationships are supported by several studies, including an analysis of 17 genes in nuclear/plastid/mitochondrial genomes from 640 species (Soltis *et al.*, 2011), an analysis of 78 protein-coding genes in plastid genomes from 360 species (Ruhfel *et al.*, 2014), and an analysis of 73 protein-coding genes in plastid genomes from 99 species in lamiids (Stull *et al.*, 2015).

oxidases and *G. uralensis* CYP716A179 identified in this study, while CYP716A141 showed unique C-16β oxidase activity. The CYP72A subfamily enzymes isolated in Chapter 3 did not catalyze oxidation reactions at β-amyrin or oleanolic acid. Currently, CYP72A subfamily enzymes with triterpene oxidase activity, including CYP72A566 (isolated in Chapter 2), have only been isolated from legumes. Therefore, the triterpene oxidase within the CYP72A subfamily may have developed exclusively in legumes. Similarly, CYP93E3 subfamily enzymes catalyzing the oxidation of triterpene skeletons have only been isolated from legumes, and similar P450s have not been found in *P.*

grandiflorus. These results suggest that the P450s responsible for oxidation reactions at C-2β, C-23, and C-24 positions of β-amyrin evolved convergently in *P. grandiflorus*. Many medicinal plants contain triterpenoids possessing C-2β, C-23, or C-24 hydroxyl groups within their aglycone (Table 1-1 and Figure 1-1). The exploration of such medicinal plants, which belong outside the Fabaceae, may provide more information on how each plant species evolved P450s that catalyze the abovementioned oxidation reactions in the β-amyrin skeleton.

The characterization of regulator proteins (including TFs), not just biosynthetic enzymes, is also important for understanding triterpenoid biosynthesis. In Chapter 4, I identified a basic helix-loop-helix (bHLH) TF that regulates soyasaponin biosynthesis in *G. uralensis*. According to a previous study, the biosynthesis of hemolytic triterpenoid saponins (possess a carboxyl group at the C-28 position of β-amyrin) and non-hemolytic saponins (possess a hydroxyl group at the C-24 position of β-amyrin) are regulated by different bHLH TFs in *M. truncatula* (Mertens *et al.*, 2016). GubHLH3, identified in Chapter 4, selectively activated soyasaponin biosynthesis, and the overexpression of *GubHLH3* in transgenic hairy roots reduced the oleanolic acid content (Figure 4-11). Therefore, it is hypothesized that hemolytic saponin biosynthesis and non-hemolytic saponin biosynthesis are differentially regulated in legumes by different TFs. Interestingly, platycodigenin has both the characteristic functional groups of hemolytic saponin and non-hemolytic saponin. Therefore, it is of interest to determine whether putative TFs that regulate platycodin D biosynthesis resemble one of the functionally characterized bHLH TFs that regulate triterpenoid biosynthesis in legumes, or whether it is completely different from previously characterized TFs. I performed a phylogenetic analysis of putative bHLH TFs in the *P. grandiflorus* transcriptome and generated a phylogenetic tree together with previously identified bHLH TFs involved in the regulation of specialized metabolites and with GubHLH1–3 (Figure 5-2). I found two putative bHLHs (Unigene4737 and Unigene12591) with relatively similar structures to GubHLH3, TSAR1, and TSAR2. Functional analyses of these putative bHLH TFs may provide new insight into the functions of bHLH TFs involved in triterpenoid biosynthesis.

The results presented in this study could be useful in several ways. One possible application is the metabolic engineering of *G. uralensis*. Glycyrrhizin is a unique triterpenoid saponin produced in

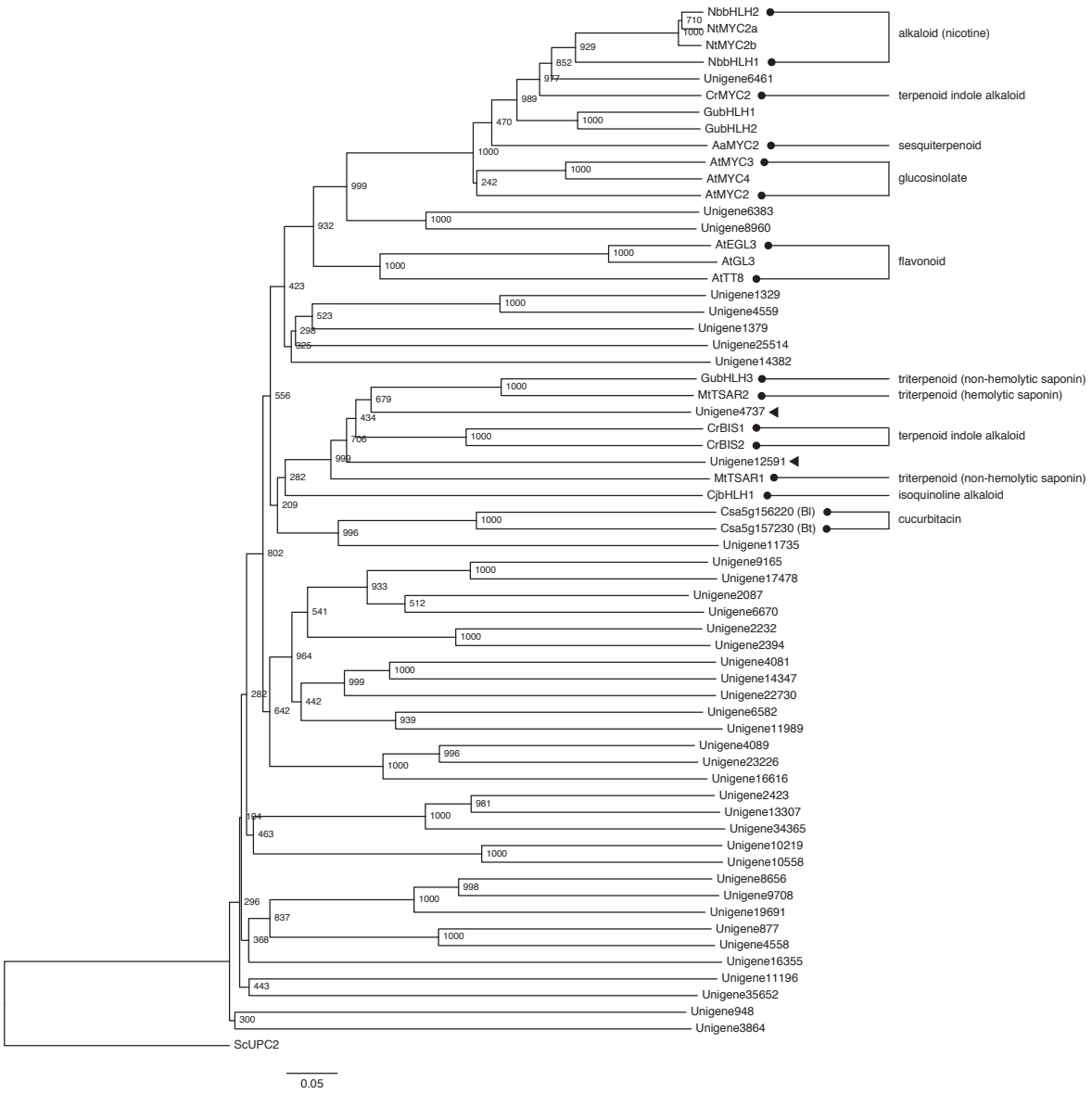


Figure 5-2. Phylogenetic tree of representative bHLH TFs involved in the biosynthesis of specialized metabolites and putative bHLH TFs in *P. grandiflorus*. Putative bHLH TFs in *P. grandiflorus* (indicated by unigene numbers) were identified from the predicted protein sequences obtained in Chapter 3 by scanning against a hidden Markov model (HMM) of HLH (PF00100). A phylogenetic tree was constructed based on full-length protein sequences. Unigene4737 and Unigene12591 are indicated by filled arrowheads. Numbers indicate bootstrap values for 1,000 replicates. The scale bar shows the amino acid substitution ratio. GenBank protein accession numbers and species of previously characterized TFs in this tree are indicated in Table 4-7. *Saccharomyces cerevisiae* UPC2 (GenBank protein accession number EDN60549) was used as an outgroup. RPKM values of putative bHLH TFs in *P. grandiflorus* are indicated in Table 5-1.

a limited number of *Glycyrrhiza* (licorice) species, including *G. uralensis*. In Kampo medicine, licorice is indispensable; more than 70% of Kampo formulas contain licorice (Table 1-1). *G. uralensis* is mainly used in Kampo medicine, and Japan imports the plant from China; however, the price of imported licorice from China has increased dramatically (Hayashi and Sudo, 2009). Plant tissue

Table 5-1. RPKM values of putative bHLH TFs in *P. grandiflorus*

Unigene	RPKM		
	Root	Leaf	Petal
Unigene6461	92.90	35.00	19.69
Unigene6383	50.34	7.13	15.98
Unigene8960	50.80	23.58	31.94
Unigene1329	37.08	43.04	3.68
Unigene4559	23.34	57.05	118.41
Unigene1379	83.86	8.51	8.82
Unigene25514	0.04	1.10	16.16
Unigene14382	7.54	0.65	0.35
Unigene4737	8.62	1.98	9.17
Unigene12591	16.47	5.58	8.63
Unigene11735	2.88	1.12	0.85
Unigene9165	20.98	1.82	5.51
Unigene17478	4.84	5.50	1.31
Unigene2087	68.15	33.87	43.41
Unigene6670	35.05	27.94	17.02
Unigene2232	74.07	29.55	8.03
Unigene2394	77.73	39.58	33.16
Unigene4081	17.21	21.34	28.06
Unigene14347	11.14	6.52	7.68
Unigene22730	1.19	24.02	4.15
Unigene6582	4.01	59.09	63.42
Unigene11989	33.02	28.62	17.19
Unigene4089	7.24	10.19	5.89
Unigene23226	0.61	10.72	10.64
Unigene16616	2.50	2.85	1.30
Unigene2423	62.01	1.21	0.06
Unigene13307	22.54	3.36	1.40
Unigene34365	0.00	6.59	3.29
Unigene10219	10.42	20.23	6.62
Unigene10558	3.74	23.11	6.26
Unigene8656	35.67	22.87	23.06
Unigene9708	20.56	26.89	23.73
Unigene19691	5.50	15.04	14.63
Unigene877	164.48	77.59	109.04
Unigene4558	44.98	55.31	39.85
Unigene16355	10.42	6.78	8.74
Unigene11196	10.67	22.21	10.70
Unigene35652	0.00	1.94	1.53
Unigene948	73.02	0.89	0.36
Unigene3864	72.93	64.10	30.58

Unigenes are listed in order of appearance (from top to bottom) in Figure 5-2. Unigene4737 and Unigene12591 are indicated in bold.

culture is an alternative to obtaining licorice plants; however, as shown in Chapter 2, the content of glycyrrhizin in tissue-cultured stolons is extremely low compared with field-grown intact roots (Table 2-3). According to the Japanese Pharmacopoeia 17th Edition, the content of glycyrrhizin in dried crude licorice drugs should be at least 2.0%. The results presented in Chapters 2 and 4 suggest strategies to increase the glycyrrhizin content in tissue cultures of *G. uralensis* (Figure 5-3). The two identified P450s, CYP716A179 and CYP72A566, and the previously identified CYP93E3 are possible targets for gene knockdown or silencing to suppress oleanolic acid, betulinic acid, and soyasaponin biosynthesis to direct the metabolic flux into glycyrrhizin biosynthesis. The application of GubHLH3 for the increased production of glycyrrhizin may be difficult because it activates the expression of *bAS* in addition to *CYP93E3* and *CYP72A566*; however, I showed the possibility of the metabolic

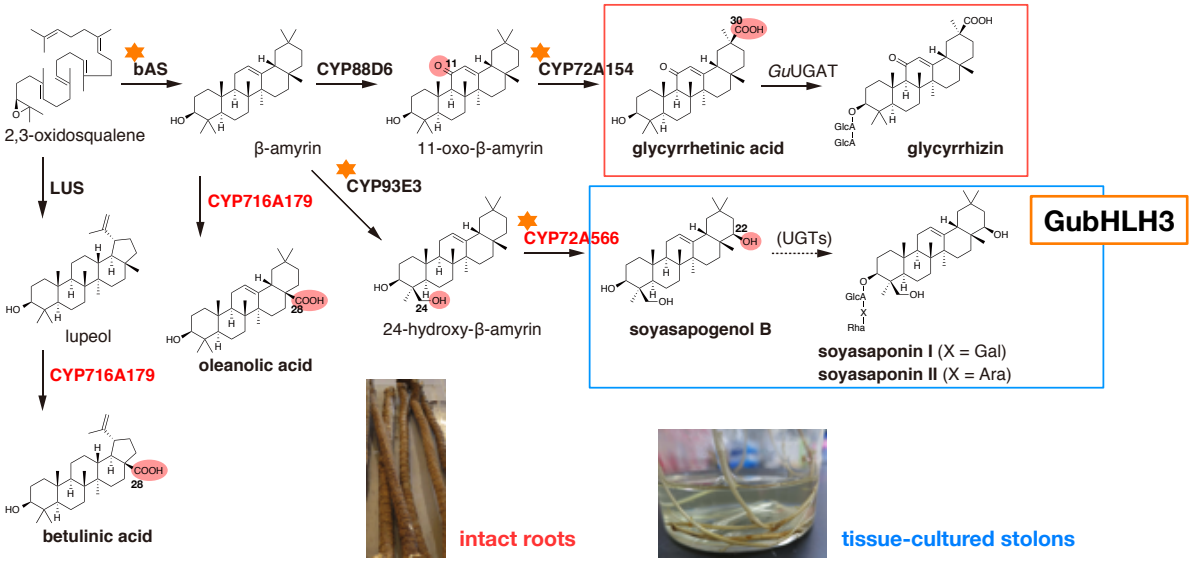


Figure 5-3. Biosynthetic pathways for major triterpenoids in *G. uralensis* and a summary of the results presented in Chapters 2 and 4. Multiple reaction steps catalyzed by CYP88D6, CYP72A154, CYP716A179, and *Gu*UGAT (Figure 2-2) are shown as a single arrow. Major triterpenoids in intact roots are indicated by a red square; those in tissue-cultured stolons are indicated by blue squares. P450s identified in this study are indicated in red. Upregulated genes due to the overexpression of *GubHLH3* are indicated by orange stars.

engineering of *G. uralensis* using TFs. Among glycyrrhizin biosynthetic genes, the expression level of *bAS* was comparable between intact roots and tissue-cultured stolons; however, that of *CYP88D6* and *CYP72A154* was suppressed in tissue-cultured stolons (Figure 2-13). Specifically, the expression of *CYP88D6* in tissue-cultured stolons was strictly suppressed. Therefore, it is possible that putative TF(s) highly expressed in tissue-cultured stolons, not only putative TF(s) that activate glycyrrhizin biosynthesis, suppress the expression of *CYP88D6* and *CYP72A154*. The future identification of such TFs in combination with metabolic engineering through the P450s identified in this study will make the production of glycyrrhizin in tissue cultures of *G. uralensis* possible.

Another possible application is the heterologous production of triterpenoids containing a C-16β hydroxyl group using the CYP716A141 identified from *P. grandiflorus* in Chapter 3. Although platycodigenin has a hydroxyl group at the C-16α position of β-amyrin, this identified C-16β hydroxylase can be used for the heterologous production of bioactive triterpenoids containing the C-16β hydroxyl group of the β-amyrin skeleton, including saikosaponin A, which is produced in another important medicinal plant used in Kampo medicine, *Bupleurum falcatum*.

The number of characterized genes involved in triterpenoid biosynthesis has increased

dramatically. However, although various structures of triterpenoids have been discovered from a wide range of medicinal plants belonging to various families, gene discovery of triterpenoid biosynthesis is limited within certain plant families. As predicted from the results of Chapter 3, even in the same oxidation reaction, there is a possibility that P450s belonging to different families catalyze identical reactions in different plant species. Therefore, unlike oxidosqualene cyclases (OSCs) and triterpene C-28 oxidases, the isolation of various enzymes for the modification of triterpene skeletons based on sequence similarities with characterized enzymes may be difficult in cases of plants belonging to different families. The triterpenoids reported in medicinal plants often have complicated structures with highly oxidized aglycones and multiple sugar moieties (Figure 1-1). Hence, enzyme assays of complicated triterpenoids may be difficult, owing to the selection of appropriate substrates and their availability. The recent discovery of the WRKY TF, which regulates withanolide biosynthesis in *Withania somnifera*, was achieved only via the partial identification of biosynthetic genes (Singh *et al.*, 2017). TFs can coordinately regulate several genes involved in the biosynthetic pathways of specialized metabolites (Grotewold, 2008). Therefore, the identification of a key TF gene controlling the desired biosynthetic pathway could lead to the discovery of novel enzyme genes, and could also provide useful molecular tools to enhance the production of desired triterpenoids in plants more efficiently. Moreover, TFs often determine the tissue-specific biosynthesis of specialized metabolites. Although the triterpenoid profiles of each medicinal plant are different, many medicinal plants, including *G. uralensis* and *P. grandiflorus*, accumulate bioactive triterpenoids in roots (Table 1-1). Some MYB TFs play conserved roles in the tissue-specific biosynthesis of proantocyanidines and anthocyanin biosynthesis (Yoshida *et al.*, 2015). The characterization of TFs involved in triterpenoid biosynthesis from a wide range of medicinal plant species may uncover the conserved regulator of root-specific triterpenoid biosynthesis in medicinal plants.

According to the PlaBi database (http://www.plabipd.de/), the genomes of a number of medicinal plants were recently sequenced, including *Morus notabilis* (He *et al.*, 2013), *Ziziphus jujuba* (Liu *et al.*, 2014), *G. uralensis* (Mochida *et al.*, 2017), and *Panax notoginseng* (Chen *et al.*, 2017). Genomic analyses provide useful information on regulatory elements, including promoter sequences, as well as biosynthetic gene clusters, as shown previously in several metabolic pathways (Nützmann

and Osbourn, 2014). The comparison of *G. uralensis* and *P. grandiflorus* performed in this study implies that even in species producing structually similar triterpenoids, the triterpenoids are synthesized by uniquely evolved molecular mechanisms. A comprehensive understanding of the enzymes and regulatory mechanisms of valuable triterpenoids found in medicinal plants, with the aid of genomic sequencing of additional plant species, will help address questions on the evolutionary origins of specialized triterpenoids and provide solutions for the sustainable use of medicinal plants.

References

Ajikumar, P.K., Xiao, W.-H., Tyo, K.E.J., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T.H., Pfeifer, B. and Stephanopoulos, G. (2010) Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli*. *Science*, **330**, 70–74.

Allan, A.C., Hellens, R.P. and Laing, W.A. (2008) MYB transcription factors that colour our fruit. *Trends Plant Sci.*, **13**, 99–102.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

APG IV (2016) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.*, **181**, 1–20.

Andre, C.M., Legay, S., Deleruelle, A., Nieuwenhuizen, N., Punter, M., Brendolise, C., Cooney, J.M., Lateur, M., Hausman, J.-F., Larondelle, Y. and Laing, W.A. (2016) Multifunctional oxidosqualene cyclases and cytochrome P450 involved in the biosynthesis of apple fruit triterpenic acids. *New Phytol.*, **211**, 1279–1294.

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

Balandrin, M.F., Klocke, J.A., Wurtele, E.S. and Bollinger, W.H. (1985) Natural plant chemicals: sources of industrial and medicinal materials. *Science*, **228**, 1154–1160.

Benedito, V.A., Torres-Jerez, I., Murray, J.D., Andriankaja, A., Allen, S., Kakar, K., Wandrey, M., Verdier, J., Zuber, H., Ott, T., Moreau, S., Niebel, A., Frickey, T., Weiller, G., He, J., Dai, X., Zhao, P.X., Tang, Y. and Udvardi, M.K. (2008) A gene expression atlas of the model legume *Medicago truncatula*. *Plant J.*, **55**, 504–513.

Biazzi, E., Carelli, M., Tava, A., Abbruscato, P., Losini, I., Avato, P., Scotti, C. and Calderini, O. (2015) CYP72A67 catalyzes a key oxidative step in *Medicago truncatula* hemolytic saponin biosynthesis. *Mol. Plant*, **8**, 1493–1506.

Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Bourgaud, F., Gravot, A., Milesi, S. and Gontier, E. (2001) Production of plant secondary metabolites: a historical perspective. *Plant Sci.*, **161**, 839–851.

Boutaghane, N., Voutquenne-Nazabadioko, L., Simon, A., Harakat, D., Benlabed, K. and Kabouche, Z. (2013) A new triterpenic diester from the aerial parts of *Chrysanthemum macrocarpum*. *Phytochem. Lett.*, **6**, 519–525.

Butelli, E., Titta, L., Giorgio, M., Mock, H.-P., Matros, A., Peterek, S., Schijlen, E.G.W.M., Hall, R.D., Bovy, A.G., Luo, J. and Martin, C. (2008) Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nat. Biotechnol.*, **26**, 1301–1308.

Cárdenas, P.D., Sonawane, P.D., Pollier, J., Vanden Bossche, R., Dewangan, V., Weithorn, E.,

Tal, L., Meir, S., Rogachev, I., Malitsky, S., Giri, A.P., Goossens, A., Burdman, S. and Aharoni, A. (2016) GAME9 regulates the biosynthesis of steroidal alkaloids and upstream isoprenoids in the plant mevalonate pathway. *Nat. Commun.*, **7**, 10654.

Carelli, M., Biazzi, E., Panara, F., Tava, A., Scaramelli, L., Porceddu, A., Graham, N., Odoardi, M., Piano, E., Arcioni, S., May, S., Scotti, C. and Calderini, O. (2011) *Medicago truncatula* CYP716A12 is a multifunctional oxidase involved in the biosynthesis of hemolytic saponins. *Plant Cell*, **23**, 3070–3081.

Castillo, D.A., Kolesnikova, M.D. and Matsuda, S.P.T. (2013) An effective strategy for exploring unknown metabolic pathways by genome mining. *J. Am. Chem. Soc.*, **135**, 5885–5894.

Chen, M.F., Huang, C.C., Liu, P.S., Chen, C.H. and Shiu, L.Y. (2013) Saikosaponin a and saikosaponin d inhibit proliferation and migratory activity of rat HSC-T6 cells. *J. Med. Food*, **16**, 793–800.

Chen, W., Kui, L., Zhang, G., Zhu, S., Zhang, J., Wang, X., Yang, M., Huang, H., Liu, Y., Wang, Y., Li, Y., Zeng, L., Wang, W., He, X., Dong, Y. and Yang, S. (2017) Whole-genome sequencing and analysis of the Chinese herbal plant *Panax notoginseng*. *Mol. Plant*, **10**, 899–902.

Corey, E.J., Matsuda, S.P.T. and Bartel, B. (1993) Isolation of an *Arabidopsis thaliana* gene encoding cycloartenol synthase by functional expression in a yeast mutant lacking lanosterol synthase by the use of a chromatographic screen. *Proc. Natl Acad. Sci. USA*, **90**, 11628–11632.

Cragg, G.M. and Newman, D.J. (2013) Natural products: a continuing source of novel drug leads. *Biochim. Biophys. Acta*, **1830**, 3670–3695.

de Hoon, M.J.L., Imoto, S., Nolan, J. and Miyano, S. (2004) Open source clustering software. *Bioinformatics*, **20**, 1453–1454.

Dinda, B., Debnath, S., Mohanta, B.C. and Harigaya, Y. (2010) Naturally occurring triterpenoid saponins. *Chem. Biodivers.*, **7**, 2327–2580.

Efferth, T., Miyachi, H. and Bartsch, H. (2007) Pharmacogenomics of a traditional Japanese herbal medicine (Kampo) for cancer therapy. *Cancer Genomics Proteomics*, **4**, 81–91.

Fiallos-Jurado, J., Pollier, J., Moses, T., Arendt, P., Barriga-Medina, N., Morillo, E., Arahana, V., de Lourdes Torres, M., Goossens, A. and Leon-Reyes, A. (2016) Saponin determination, expression analysis and functional characterization of saponin biosynthetic genes in *Chenopodium quinoa* leaves. *Plant Sci.*, **250**, 188–197.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J. and Bateman, A. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.*, **44**, D279–D285.

Fraenkel, G.S. (1959) The raison d'être of secondary plant substances. *Science*, **129**, 1466–1470.

Fujiwara, S., Sakamoto, S., Kigoshi, K., Suzuki, K. and Ohme-Takagi, M. (2014) VP16 fusion induces the multiple-knockout phenotype of redundant transcriptional repressors partly by Med25-independent mechanisms in *Arabidopsis*. *FEBS Lett.*, **588**, 3665–3672.

Fukumura, M., Iwasaki, D., Hirai, Y., Hori, Y., Toriizuka, K., Kenny, P.T.M., Kuchino, Y. and Ida, Y. (2010) Eight new oleanane-type triterpenoid saponins from platycodon root.

Heterocycles, **81**, 2793–2806.

Fukushima, E.O., 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.

Fukushima, E.O., 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.

Gang, D.R. (2005) Evolution of flavors and scents. *Annu. Rev. Plant Biol.*, **56**, 301–325.

Geisler, K., Hughes, R.K., Sainsbury, F., Lomonossoff, G.P., Rejzek, M., Fairhurst, S., Olsen, C.-E., Motawia, M.S., Melton, R.E., Hemmings, A.M., Bak, S. and Osbourn, A. (2013) Biochemical analysis of a multifunctional cytochrome P450 (CYP51) enzyme required for synthesis of antimicrobial triterpenes in plants. *Proc. Natl Acad. Sci. USA*, **110**, E3360–E3367.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. and Regev, A. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.*, **29**, 644–652.

Grotewold, E. (2008) Transcription factors for predictive plant metabolic engineering: are we there yet? *Curr. Opin. Biotechnol.*, **19**, 138–144.

Gurib-Fakim, A. (2006) Medicinal plants: traditions of yesterday and drugs of tomorrow. *Mol. Aspects Med.*, **27**, 1–93.

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.

Han, J.-Y., Hwang, H.-S., Choi, S.-W., Kim, H.-J. and Choi, Y.-E. (2012) Cytochrome P450 CYP716A53v2 catalyzes the formation of protopanaxatriol from protopanaxadiol during ginsenoside biosynthesis in *Panax ginseng*. *Plant Cell Physiol.*, **53**, 1535–1545.

Han, J.-Y., Kim, H.-J., Kwon, Y.-S. and Choi, Y.-E. (2011) The cyt P450 enzyme CYP716A47 catalyzes the formation of protopanaxadiol from dammarenediol-II during ginsenoside biosynthesis in *Panax ginseng*. *Plant Cell Physiol.*, **52**, 2062–2073.

Han, J.-Y., Kim, M.-J., Ban, Y.-W., Hwang, H.-S. and Choi, Y.-E. (2013) The involvement of β-amyrin 28-oxidase (CYP716A52v2) in oleanane-type ginsenoside biosynthesis in *Panax ginseng*. *Plant Cell Physiol.*, **54**, 2034–2046.

Hartmann, T. (2008) The lost origin of chemical ecology in the late 19th century. *Proc. Natl Acad. Sci. USA*, **105**, 4541–4546.

Hayashi, H., Fukui, H. and Tabata, M. (1988) Examination of triterpenoids produced by callus and cell suspension cultures of *Glycyrrhiza glabra*. *Plant Cell Rep.*, **7**, 508–511.

Hayashi, H., Fukui, H. and Tabata, M. (1993) Distribution pattern of saponins in different organs of *Glycyrrhiza glabra*. *Planta Med.*, **59**, 351–353.

Hayashi, H., Hiraoka, N. and Ikeshiro, Y. (2005) Differential regulation of soyasaponin and betulinic acid production by yeast extract in cultured licorice cells. *Plant Biotechnol.*, **22**, 241–244.

Hayashi, H., Huang, P. and Inoue, K. (2003) Up-regulation of soyasaponin biosynthesis by methyl jasmonate in cultured cells of *Glycyrrhiza glabra*. *Plant Cell Physiol.*, **44**, 404–411.

Hayashi, H., Huang, P., Kirakosyan, A., Inoue, K., Hiraoka, N., Ikeshiro, Y., Kushiro, T., Shibuya, M. and Ebizuka, Y. (2001) Cloning and characterization of a cDNA encoding β-amyrin synthase involved in glycyrrhizin and soyasaponin biosyntheses in licorice. *Biol. Pharm. Bull.*, **24**, 912–916.

Hayashi, H., Huang, P., Takada, S., Obinata, M., Inoue, K., Shibuya, M. and Ebizuka, Y. (2004) Differential expression of three oxidosqualene cyclase mRNAs in *Glycyrrhiza glabra*. *Biol. Pharm. Bull.*, **27**, 1086–1092.

Hayashi, H., Sakai, T., Fukui, H. and Tabata, M. (1990) Formation of soyasaponins in licorice cell suspension cultures. *Phytochemistry*, **29**, 3127–3129.

Hayashi, H. and Sudo, H. (2009) Economic importance of licorice. *Plant Biotechnol.*, **26**, 101–104.

He, J., Benedito, V.A., Wang, M., Murray, J.D., Zhao, P.X., Tang, Y. and Udvardi, M.K. (2009) The *Medicago truncatula* gene expression atlas web server. *BMC Bioinformatics*, **10**, 441.

He, N., Zhang, C., Qi, X., Zhao, S., Tao, Y., Yang, G., Lee, T.-H., Wang, X., Cai, Q., Li, D., Lu, M., Liao, S., Luo, G., He, R., Tan, X., Xu, Y., Li, T., Zhao, A., Jia, L., Fu, Q., Zeng, Q., Gao, C., Ma, B., Liang, J., Wang, X., Shang, J., Song, P., Wu, H., Fan, L., Wang, Q., Shuai, Q., Zhu, J., Wei, C., Zhu-Salzman, K., Jin, D., Wang, J., Liu, T., Yu, M., Tang, C., Wang, Z., Dai, F., Chen, J., Liu, Y., Zhao, S., Lin, T., Zhang, S., Wang, J., Wang, J., Yang, H., Yang, G., Wang, J., Paterson, A.H., Xia, Q., Ji, D. and Xiang, Z. (2013) Draft genome sequence of the mulberry tree *Morus notabilis*. *Nat. Commun.*, **4**, 2445.

Hill, R.A. and Connolly, J.D. (2012) Triterpenoids. *Nat. Prod. Rep.*, **29**, 780–818.

Hiratsu, K., Matsui, K., Koyama, T. and Ohme-Takagi, M. (2003) Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in *Arabidopsis*. *Plant J.*, **34**, 733–739.

Hothorn, T., Bretz, F. and Westfall, P. (2008) Simultaneous inference in general parametric models. *Biom. J.*, **50**, 346–363.

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.

Huhman, D.V., Berhow, M.A. and Sumner, L.W. (2005) Quantification of saponins in aerial and subterranean tissues of *Medicago truncatula*. *J. Agric. Food Chem.*, **53**, 1914–1920.

Ida, Y., Satoh, Y., Katoh, M., Katsumata, M., Nagasao, M., Yamaguchi, K., Kamei, H. and Shoji, J. (1994) Achyranthosides A and B, novel cytotoxic saponins from *Achyranthes fauriei* root. *Tetrahedron Lett.*, **35**, 6887–6890.

Jäger, S., Trojan, H., Kopp, T., Laszczyk, M.N. and Scheffler, A. (2009) Pentacyclic triterpene distribution in various plants—rich sources for a new group of multi-potent plant extracts. *Molecules*, **14**, 2016–2031.

Japan Kampo Medicines Manufacturers Association (2015) 第 3 回中国産原料生薬の価格指数調査. Retreved November 21, 2017, from http://www.nikkankyo.org/aboutus/investigation/kakaku-chousa/kakaku-chousa_003.html

Jo, H.-J., Han, J.Y., Hwang, H.-S. and Choi, Y.E. (2017) β-Amyrin synthase (EsBAS) and

β-amyrin 28-oxidase (CYP716A244) in oleanane-type triterpene saponin biosynthesis in *Eleutherococcus senticosus*. *Phytochemistry*, **135**, 53–63.

Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M. (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.*, **44**, D457–D462.

Khakimov, B., Kuzina, V., Erthmann, P.Ø., Fukushima, E.O., Augustin, J.M., Olsen, C.E., Scholtalbers, J., Volpin, H., Andersen, S.B., Hauser, T.P., Muranaka, T. and Bak, S. (2015) Identification and genome organization of saponin pathway genes from a wild crucifer, and their use for transient production of saponins in *Nicotiana benthamiana*. *Plant J.*, **84**, 478–490.

Koes, R., Verweij, W. and Quattrocchio, F. (2005) Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. *Trends Plant Sci.*, **10**, 236–242.

Kojoma, M., Ohyama, K., Seki, H., Hiraoka, Y., Asazu, S.N., Sawa, S., Sekizaki, H., Yoshida, S. and Muranaka, T. (2010) *In vitro* proliferation and triterpenoid characteristics of licorice (*Glycyrrhiza uralensis* Fischer, Leguminosae) stolons. *Plant Biotechnol.*, **27**, 59–66.

Koressaar, T. and Remm, M. (2007) Enhancements and modifications of primer design program Primer3. *Bioinformatics*, **23**, 1289–1291.

Krokida, A., Delis, C., Geisler, K., Garagounis, C., Tsikou, D., Peña-Rodríguez, L.M., Katsarou, D., Field, B., Osbourn, A.E. and Papadopoulou, K.K. (2013) A metabolic gene cluster in *Lotus japonicus* discloses novel enzyme functions and products in triterpene biosynthesis. *New Phytol.*, **200**, 675–690.

Kubota, T. and Kitatani, H. (1969) The structure of platycodigenin, a 4,4-di(hydroxymethyl)-triterpene. *J. Chem. Soc. D*, 190–191.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins, D.G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.

Leavell, M.D., McPhee, D.J. and Paddon, C.J. (2016) Developing fermentative terpenoid production for commercial usage. *Curr. Opin. Biotechnol.*, **37**, 114–119.

Li, W., Liu, Y., Wang, Z., Han, Y., Tian, Y.-H., Zhang, G.-S., Sun, Y.-S. and Wang, Y.-P. (2015) Platycodin D isolated from the aerial parts of *Platycodon grandiflorum* protects alcohol-induced liver injury in mice. *Food Funct.*, **6**, 1418–1427.

Li, Y., Luo, H.-M., Sun, C., Song, J.-Y., Sun, Y.-Z., Wu, Q., Wang, N., Yao, H., Steinmetz, A. and Chen, S.-L. (2010) EST analysis reveals putative genes involved in glycyrrhizin biosynthesis. *BMC Genomics*, **11**, 268.

Liu, M.-J., Zhao, J., Cai, Q.-L., Liu, G.-C., Wang, J.-R., Zhao, Z.-H., Liu, P., Dai, L., Yan, G., Wang, W.-J., Li, X.-S., Chen, Y., Sun, Y.-D., Liu, Z.-G., Lin, M.-J., Xiao, J., Chen, Y.-Y., Li, X.-F., Wu, B., Ma, Y., Jian, J.-B., Yang, W., Yuan, Z., Sun, X.-C., Wei, Y.-L., Yu, L.-L., Zhang, C., Liao, S.-G., He, R.-J., Guang, X.-M., Wang, Z., Zhang, Y.-Y. and Luo, L.-H. (2014) The complex jujube genome provides insights into fruit tree biology. *Nat. Commun.*, **5**, 5315.

Lv, Z., Wang, S., Zhang, F., Chen, L., Hao, X., Pan, Q., Fu, X., Li, L., Sun, X. and Tang, K.

(2016) Overexpression of a novel NAC domain-containing transcription factor gene (*AaNAC1*) enhances the content of artemisinin and increases tolerance to drought and *Botrytis cinerea* in *Artemisia annua*. *Plant Cell Physiol.*, **57**, 1961–1971.

Ma, G., Guo, W., Zhao, L., Zheng, Q., Sun, Z., Wei, J., Yang, J. and Xu, X. (2013) Two new triterpenoid saponins from the root of *Platycodon grandiflorum*. *Chem. Pharm. Bull.*, **61**, 101–104.

Masullo, M., Calabria, L., Gallotta, D., Pizza, C. and Piacente, S. (2014) Saponins with highly hydroxylated oleanane-type aglycones from *Silphium asteriscus* L. *Phytochemistry*, **97**, 70–80.

Mertens, J., Pollier, J., Vanden Bossche, R., Lopez-Vidriero, I., Franco-Zorrilla, J.M. and Goossens, A. (2016) The bHLH transcription factors TSAR1 and TSAR2 regulate triterpene saponin biosynthesis in *Medicago truncatula*. *Plant Physiol.*, **170**, 194–210.

Miettinen, K., Pollier, J., Buyst, D., Arendt, P., Csuk, R., Sommerwerk, S., Moses, T., Mertens, J., Sonawane, P.D., Pauwels, L., Aharoni, A., Martins, J., Nelson, D.R. and Goossens, A. (2017) The ancient CYP716 family is a major contributor to the diversification of eudicot triterpenoid biosynthesis. *Nat. Commun.*, **8**, 14153.

Misra, R.C., Sharma, S., Sandeep, Garg, A., Chanotiya, C.S. and Ghosh, S. (2017) Two CYP716A subfamily cytochrome P450 monooxygenases of sweet basil play similar but nonredundant roles in ursane- and oleanane-type pentacyclic triterpene biosynthesis. *New Phytol.*, **214**, 706–720.

Mitsuda, N., Ikeda, M., Takada, S., Takiguchi, Y., Kondou, Y., Yoshizumi, T., Fujita, M., Shinozaki, K., Matsui, M. and Ohme-Takagi, M. (2010) Efficient yeast one-/two-hybrid screening using a library composed only of transcription factors in *Arabidopsis thaliana*. *Plant Cell Physiol.*, **51**, 2145–2151.

Mitsuda, N. and Ohme-Takagi, M. (2009) Functional analysis of transcription factors in Arabidopsis. *Plant Cell Physiol.*, **50**, 1232–1248.

Mochida, K., Sakurai, T., Seki, H., Yoshida, T., Takahagi, K., Sawai, S., Uchiyama, H., Muranaka, T. and Saito, K. (2017) Draft genome assembly and annotation of *Glycyrrhiza uralensis*, a medicinal legume. *Plant J.*, **89**, 181–194.

Mochida, K., Yoshida, T., Sakurai, T., Yamaguchi-Shinozaki, K., Shinozaki, K. and Tran, L.-S.P. (2009) *In silico* analysis of transcription factor repertoire and prediction of stress responsive transcription factors in soybean. *DNA Res.*, **16**, 353–369.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C. and Kanehisa, M. (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.*, **35**, W182–W185.

Moses, T., Pollier, J., Almagro, L., Buyst, D., Van Montagu, M., Pedreño, M.A., Martins, J.C., Thevelein, J.M. and Goossens, A. (2014a) 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.

Moses, T., Pollier, J., Faizal, A., Apers, S., Pieters, L., Thevelein, J.M., Geelen, D. and Goossens, A. (2015a) Unraveling the triterpenoid saponin biosynthesis of the African shrub *Maesa lanceolata*. *Mol. Plant*, **8**, 122–135.

Moses, T., Pollier, J., Shen, Q., Soetaert, S., Reed, J., Erffelinck, M.-L., Van Nieuwerburgh,

F.C.W., Vanden Bossche, R., Osbourn, A., Thevelein, J.M., Deforce, D., Tang, K. and Goossens, A. (2015b) OSC2 and CYP716A14v2 catalyze the biosynthesis of triterpenoids for the cuticle of aerial organs of *Artemisia annua*. *Plant Cell*, **27**, 286–301.

Moses, T., Thevelein, J.M., Goossens, A. and Pollier, J. (2014b) Comparative analysis of CYP93E proteins for improved microbial synthesis of plant triterpenoids. *Phytochemistry*, **108**, 47–56.

Nagao, T., Okabe, H. and Yamauchi, T. (1990) Studies on the constituents of *Aster tataricus* L. f. III. Structures of aster saponins E and F isolated from the root. *Chem. Pharm. Bull.*, **38**, 783–785.

Nelson, D. and Werck-Reichhart, D. (2011) A P450-centric view of plant evolution. *Plant J.*, **66**, 194–211.

Nelson, D.R. (2003) Comparison of P450s from human and fugu: 420 million years of vertebrate P450 evolution. *Arch. Biochem. Biophys.*, **409**, 18–24.

Nemoto, Y. (2016) *漢方 294 処方生薬解説 その基礎から運用まで* Tokyo: Jiho Inc.

Nesi, N., Debeaujon, I., Jond, C., Pelletier, G., Caboche, M. and Lepiniec, L. (2000) The *TT8* gene encodes a basic helix-loop-helix domain protein required for expression of *DFR* and *BAN* genes in Arabidopsis siliques. *Plant Cell*, **12**, 1863–1878.

Nützmann, H.-W. and Osbourn, A. (2014) Gene clustering in plant specialized metabolism. *Curr. Opin. Biotechnol.*, **26**, 91–99.

Nyakudya, E., Jeong, J.H., Lee, N.K. and Jeong, Y.-S. (2014) Platycosides from the roots of *Platycodon grandiflorum* and their health benefits. *Prev. Nutr. Food. Sci.*, **19**, 59–68.

Oda-Yamamizo, C., Mitsuda, N., Sakamoto, S., Ogawa, D., Ohme-Takagi, M. and Ohmiya, A. (2016) The NAC transcription factor ANAC046 is a positive regulator of chlorophyll degradation and senescence in Arabidopsis leaves. *Sci. Rep.*, **6**, 23609.

Oshima, Y., Mitsuda, N., Nakata, M., Nakagawa, T., Nagaya, S., Kato, K. and Ohme-Takagi, M. (2011) Novel vector systems to accelerate functional analysis of transcription factors using chimeric repressor gene-silencing technology (CRES-T). *Plant Biotechnol.*, **28**, 201–210.

Paddon, C.J., Westfall, P.J., Pitera, D.J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M.D., Tai, A., Main, A., Eng, D., Polichuk, D.R., Teoh, K.H., Reed, D.W., Treynor, T., Lenihan, J., Fleck, M., Bajad, S., Dang, G., Dengrove, D., Diola, D., Dorin, G., Ellens, K.W., Fickes, S., Galazzo, J., Gaucher, S.P., Geistlinger, T., Henry, R., Hepp, M., Horning, T., Iqbal, T., Jiang, H., Kizer, L., Lieu, B., Melis, D., Moss, N., Regentin, R., Secrest, S., Tsuruta, H., Vazquez, R., Westblade, L.F., Xu, L., Yu, M., Zhang, Y., Zhao, L., Lievense, J., Covello, P.S., Keasling, J.D., Reiling, K.K., Renninger, N.S. and Newman, J.D. (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, **496**, 528–532.

Patra, B., Schluttenhofer, C., Wu, Y., Pattanaik, S. and Yuan, L. (2013) Transcriptional regulation of secondary metabolite biosynthesis in plants. *Biochim. Biophys. Acta*, **1829**, 1236–1247.

Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., Lee, Y., White, J., Cheung, F., Parvizi, B., Tsai, J. and Quackenbush, J. (2003) TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics*, **19**, 651–652.

Phillips, D.R., Rasbery, J.M., Bartel, B. and Matsuda, S.P.T. (2006) Biosynthetic diversity in plant

triterpene cyclization. *Curr. Opin. Plant Biol.*, **9**, 305–314.

Pichersky, E. and Gang, D.R. (2000) Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends Plant Sci.*, **5**, 439–445.

Pires, N. and Dolan, L. (2010) Origin and diversification of basic-helix-loop-helix proteins in plants. *Mol. Biol. Evol.*, **27**, 862–874.

Quijano, L., Rios, T., Fronczek, F.R. and Fischer, N.H. (1998) The molecular structure of maniladiol from *Baccharis salicina*. *Phytochemistry*, **49**, 2065–2068.

Ramilowski, J.A., Sawai, S., Seki, H., Mochida, K., Yoshida, T., Sakurai, T., Muranaka, T., Saito, K. and Daub, C.O. (2013) *Glycyrrhiza uralensis* transcriptome landscape and study of phytochemicals. *Plant Cell Physiol.*, **54**, 697–710.

Rates, S.M.K. (2001) Plants as source of drugs. *Toxicon*, **39**, 603–613.

Ro, D.-K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S., Kirby, J., Chang, M.C.Y., Withers, S.T., Shiba, Y., Sarpong, R. and Keasling, J.D. (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, **440**, 940–943.

Roberts, S.C. (2007) Production and engineering of terpenoids in plant cell culture. *Nat. Chem. Biol.*, **3**, 387–395.

Ruhfel, B.R., Gitzendanner, M.A., Soltis, P.S., Soltis, D.E. and Burleigh, J.G. (2014) From algae to angiosperms—inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evol. Biol.*, **14**, 23.

Rupasinghe, H.P.V., Jackson, C.-J.C., Poysa, V., Di Berardo, C., Bewley, J.D. and Jenkinson, J. (2003) Soyasapogenol A and B distribution in soybean (*Glycine max* L. Merr.) in relation to seed physiology, genetic variability, and growing location. *J. Agric. Food Chem.*, **51**, 5888–5894.

Saito, K. (2013) Phytochemical genomics—a new trend. *Curr. Opin. Plant Biol.*, **16**, 373–380.

Sakamoto, S., Fujikawa, Y., Tanaka, N. and Esaka, M. (2012) Molecular cloning and characterization of L-galactose-1-phosphate phosphatase from tobacco (*Nicotiana tabacum*). *Biosci. Biotechnol. Biochem.*, **76**, 1155–1162.

Saldanha, A.J. (2004) Java Treeview—extensible visualization of microarray data. *Bioinformatics*, **20**, 3246–3248.

Sawai, S. and Saito, K. (2011) Triterpenoid biosynthesis and engineering in plants. *Front. Plant Sci.*, **2**, 25.

Sawai, S., Uchiyama, H., Mizuno, S., Aoki, T., Akashi, T., Ayabe, S. and Takahashi, T. (2011) Molecular characterization of an oxidosqualene cyclase that yields shionone, a unique tetracyclic triterpene ketone of *Aster tataricus*. *FEBS Lett.*, **585**, 1031–1036.

Schmidt, B.M., Ribnicky, D.M., Lipsky, P.E. and Raskin, I. (2007) Revisiting the ancient concept of botanical therapeutics. *Nat. Chem. Biol.*, **3**, 360–366.

Schweizer, F., Fernández-Calvo, P., Zander, M., Diez-Diaz, M., Fonseca, S., Glauser, G., Lewsey, M.G., Ecker, J.R., Solano, R. and Reymond, P. (2013) *Arabidopsis* basic helix-loop-helix transcription factors MYC2, MYC3, and MYC4 regulate glucosinolate biosynthesis, insect performance, and feeding behavior. *Plant Cell*, **25**, 3117–3132.

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.

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

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.

Shang, Y., Ma, Y., Zhou, Y., Zhang, H., Duan, L., Chen, H., Zeng, J., Zhou, Q., Wang, S., Gu, W., Liu, M., Ren, J., Gu, X., Zhang, S., Wang, Y., Yasukawa, K., Bouwmeester, H.J., Qi, X., Zhang, Z., Lucas, W.J. and Huang, S. (2014) Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science*, **346**, 1084–1088.

Shen, Q., Lu, X., Yan, T., Fu, X., Lv, Z., Zhang, F., Pan, Q., Wang, G., Sun, X. and Tang, K. (2016) The jasmonate-responsive AaMYC2 transcription factor positively regulates artemisinin biosynthesis in *Artemisia annua*. *New Phytol.*, **210**, 1269–1281.

Shibuya, M., Hoshino, M., Katsube, Y., Hayashi, H., Kushiro, T. and Ebizuka, Y. (2006) Identification of β-amyrin and sophoradiol 24-hydroxylase by expressed sequence tag mining and functional expression assay. *FEBS J.*, **273**, 948–959.

Shoji, T. and Hashimoto, T. (2011) Tobacco MYC2 regulates jasmonate-inducible nicotine biosynthesis genes directly and by way of the *NIC2*-locus *ERF* genes. *Plant Cell Physiol.*, **52**, 1117–1130.

Shoji, T., Kajikawa, M. and Hashimoto, T. (2010) Clustered transcription factor genes regulate nicotine biosynthesis in tobacco. *Plant Cell*, **22**, 3390–3409.

Singh, A.K., Kumar, S.R., Dwivedi, V., Rai, A., Pal, S., Shasany, A.K. and Nagegowda, D.A. (2017) A WRKY transcription factor from *Withania somnifera* regulates triterpenoid withanolide accumulation and biotic stress tolerance through modulation of phytosterol and defense pathways. *New Phytol.*, **215**, 1115–1131.

Soltis, D.E., Smith, S.A., Cellinese, N., Wurdack, K.J., Tank, D.C., Brockington, S.F., Refulio-Rodriguez, N.F., Walker, J.B., Moore, M.J., Carlsward, B.S., Bell, C.D., Latvis, M., Crawley, S., Black, C., Diouf, D., Xi, Z., Rushworth, C.A., Gitzendanner, M.A., Sytsma, K.J., Qiu, Y.-L., Hilu, K.W., Davis, C.C., Sanderson, M.J., Beaman, R.S., Olmstead, R.G., Judd, W.S., Donoghue, M.J. and Soltis, P.S. (2011) Angiosperm phylogeny: 17 genes, 640 taxa. *Am. J. Bot.*, **98**, 704–730.

Stull, G.W., Duno de Stefano, R., Soltis, D.E. and Soltis, P.S. (2015) Resolving basal lamiid phylogeny and the circumscription of Icacinaceae with a plastome-scale data set. *Am. J. Bot.*, **102**, 1794–1813.

Sudo, H., Seki, H., Sakurai, N., Suzuki, H., Shibata, D., Toyoda, A., Totoki, Y., Sakaki, Y., Iida, O., Shibata, T., Kojoma, M., Muranaka, T. and Saito, K. (2009) Expressed sequence tags from rhizomes of *Glycyrrhiza uralensis*. *Plant Biotechnol.*, **26**, 105–107.

Suttipanta, N., Pattanaik, S., Kulshrestha, M., Patra, B., Singh, S.K. and Yuan, L. (2011) The transcription factor CrWRKY1 positively regulates the terpenoid indole alkaloid biosynthesis

in *Catharanthus roseus*. *Plant Physiol.*, **157**, 2081–2093.

Suzuki, H., Reddy, M.S.S., Naoumkina, M., Aziz, N., May, G.D., Huhman, D.V., Sumner, L.W., Blount, J.W., Mendes, P. and Dixon, R.A. (2005) Methyl jasmonate and yeast elicitor induce differential transcriptional and metabolic re-programming in cell suspension cultures of the model legume *Medicago truncatula*. *Planta*, **220**, 696–707.

Tada, A., Kaneiwa, Y., Shoji, J. and Shibata, S. (1975) Studies on the saponins of the root of *Platycodon grandiflorum* A. De Candolle. I. Isolation and the structure of platycodin-D. *Chem. Pharm. Bull.*, **23**, 2965–2972.

Teoh, K.H., Polichuk, D.R., Reed, D.W. and Covello, P.S. (2009) Molecular cloning of an aldehyde dehydrogenase implicated in artemisinin biosynthesis in *Artemisia annua*. *Botany*, **87**, 635–642.

Thagun, C., Imanishi, S., Kudo, T., Nakabayashi, R., Ohyama, K., Mori, T., Kawamoto, K., Nakamura, Y., Katayama, M., Nonaka, S., Matsukura, C., Yano, K., Ezura, H., Saito, K., Hashimoto, T. and Shoji, T. (2016) Jasmonate-responsive ERF transcription factors regulate steroidal glycoalkaloid biosynthesis in tomato. *Plant Cell Physiol.*, **57**, 961–975.

Thang, N.V., Thu, V.K., Nhiem, N.X., Dung, D.T., Quang, T.H., Tai, B.H., Anh, H.L.T., Yen, P.H., Ngan, N.T.T., Hoang, N.H. and Kiem, P.V. (2017) Oleanane-type saponins from *Glochidion hirsutum* and their cytotoxic activities. *Chem. Biodivers.*, **14**, e1600445.

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

Todd, A.T., Liu, E., Polvi, S.L., Pammett, R.T. and Page, J.E. (2010) A functional genomics screen identifies diverse transcription factors that regulate alkaloid biosynthesis in *Nicotiana benthamiana*. *Plant J.*, **62**, 589–600.

Toledo-Ortiz, G., Huq, E. and Quail, P.H. (2003) The Arabidopsis basic/helix-loop-helix transcription factor family. *Plant Cell*, **15**, 1749–1770.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. and Rozen, S.G. (2012) Primer3—new capabilities and interfaces. *Nucleic Acids Res.*, **40**, e115.

van der Fits, L. and Memelink, J. (2000) ORCA3, a jasmonate-responsive transcriptional regulator of plant primary and secondary metabolism. *Science*, **289**, 295–297.

Van Moerkercke, A., Steensma, P., Gariboldi, I., Espoz, J., Purnama, P.C., Schweizer, F., Miettinen, K., Vanden Bossche, R., De Clercq, R., Memelink, J. and Goossens, A. (2016) The basic helix-loop-helix transcription factor BIS2 is essential for monoterpenoid indole alkaloid production in the medicinal plant *Catharanthus roseus*. *Plant J.*, **88**, 3–12.

Van Moerkercke, A., Steensma, P., Schweizer, F., Pollier, J., Gariboldi, I., Payne, R., Vanden Bossche, R., Miettinen, K., Espoz, J., Purnama, P.C., Kellner, F., Seppänen-Laakso, T., O'Connor, S.E., Rischer, H., Memelink, J. and Goossens, A. (2015) The bHLH transcription factor BIS1 controls the iridoid branch of the monoterpenoid indole alkaloid pathway in *Catharanthus roseus*. *Proc. Natl Acad. Sci. USA*, **112**, 8130–8135.

Watanabe, K., Matsuura, K., Gao, P., Hottenbacher, L., Tokunaga, H., Nishimura, K., Imazu, Y., Reissenweber, H. and Witt, C.M. (2011) Traditional Japanese Kampo medicine: clinical research between modernity and traditional medicine—the state of research and methodological suggestions for the future. *Evid. Based Complement. Alternat. Med.*, **2011**,

1–19.

Werck-Reichhart, D. and Feyereisen, R. (2000) Cytochromes P450: a success story. *Genome Biol.*, **1**, reviews3003.

Wong, K.H., Li, G.Q., Li, K.M., Razmovski-Naumovski, V. and Chan, K. (2011) Kudzu root: traditional uses and potential medicinal benefits in diabetes and cardiovascular diseases. *J. Ethnopharmacol.*, **134**, 584–607.

Xu, G., Cai, W., Gao, W. and Liu, C. (2016) A novel glucuronosyltransferase has an unprecedented ability to catalyse continuous two-step glucuronosylation of glycyrrhetinic acid to yield glycyrrhizin. *New Phytol.*, **212**, 123–135.

Xu, R., Fazio, G.C. and Matsuda, S.P.T. (2004) On the origins of triterpenoid skeletal diversity. *Phytochemistry*, **65**, 261–291.

Xu, W., Dubos, C. and Lepiniec, L. (2015) Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends Plant Sci.*, **20**, 176–185.

Yamada, Y., Kokabu, Y., Chaki, K., Yoshimoto, T., Ohgaki, M., Yoshida, S., Kato, N., Koyama, T. and Sato, F. (2011) Isoquinoline alkaloid biosynthesis is regulated by a unique bHLH-type transcription factor in *Coptis japonica*. *Plant Cell Physiol.*, **52**, 1131–1141.

Yang, C.-Q., Fang, X., Wu, X.-M., Mao, Y.-B., Wang, L.-J. and Chen, X.-Y. (2012) Transcriptional regulation of plant secondary metabolism. *J. Integr. Plant Biol.*, **54**, 703–712.

Yano, R., Takagi, K., Takada, Y., Mukaiyama, K., Tsukamoto, C., Sayama, T., Kaga, A., Anai, T., Sawai, S., Ohyama, K., Saito, K. and Ishimoto, M. (2017) Metabolic switching of astringent and beneficial triterpenoid saponins in soybean is achieved by a loss-of-function mutation in cytochrome P450 72A69. *Plant J.*, **89**, 527–539.

Yasumoto, S., Fukushima, E.O., Seki, H. and Muranaka, T. (2016) Novel triterpene oxidizing activity of *Arabidopsis thaliana* CYP716A subfamily enzymes. *FEBS Lett.*, **590**, 533–540.

Yasumoto, S., Seki, H., Shimizu, Y., Fukushima, E.O. and Muranaka, T. (2017) Functional characterization of CYP716 family P450 enzymes in triterpenoid biosynthesis in tomato. *Front. Plant Sci.*, **8**, 21.

Yoo, S.-D., Cho, Y.-H. and Sheen, J. (2007) *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.*, **2**, 1565–1572.

Yoshida, K., Ma, D. and Constabel, C.P. (2015) The MYB182 protein down-regulates proanthocyanidin and anthocyanin biosynthesis in poplar by repressing both structural and regulatory flavonoid genes. *Plant Physiol.*, **167**, 693–710.

Yoshikawa, M., Murakami, T., Kishi, A., Kageura, T. and Matsuda, H. (2001) Medicinal flowers. III. Marigold. (1): hypoglycemic, gastric emptying inhibitory, and gastroprotective principles and new oleanane-type triterpene oligoglycosides, calendasaponins A, B, C, and D, from Egyptian *Calendula officinalis*. *Chem. Pharm. Bull.*, **49**, 863–870.

Yu, Z.-X., Li, J.-X., Yang, C.-Q., Hu, W.-L., Wang, L.-J. and Chen, X.-Y. (2012) The jasmonate-responsive AP2/ERF transcription factors AaERF1 and AaERF2 positively regulate artemisinin biosynthesis in *Artemisia annua* L. *Mol. Plant*, **5**, 353–365.

Zhang, F., Gonzalez, A., Zhao, M., Payne, C.T. and Lloyd, A. (2003) A network of redundant bHLH proteins functions in all TTG1-dependent pathways of *Arabidopsis*. *Development*, **130**, 4859–4869.

Zhang, H., Hedhili, S., Montiel, G., Zhang, Y., Chatel, G., Pré, M., Gantet, P. and Memelink, J. (2011) The basic helix-loop-helix transcription factor CrMYC2 controls the jasmonate-responsive expression of the *ORCA* genes that regulate alkaloid biosynthesis in *Catharanthus roseus*. *Plant J.*, **67**, 61–71.

Zhang, L., Wang, Y., Yang, D., Zhang, C., Zhang, N., Li, M. and Liu, Y. (2015) *Platycodon grandiflorus*—an ethnopharmacological, phytochemical and pharmacological review. *J. Ethnopharmacol.*, **164**, 147–161.

Zhang, W. and Popovich, D.G. (2009) Chemical and biological characterization of oleanane triterpenoids from soy. *Molecules*, **14**, 2959–2975.

Zhang, Y., Teoh, K.H., Reed, D.W., Maes, L., Goossens, A., Olson, D.J.H., Ross, A.R.S. and Covello, P.S. (2008) The molecular cloning of artemisinic aldehyde Δ11(13) reductase and its role in glandular trichome-dependent biosynthesis of artemisinin in *Artemisia annua*. *J. Biol. Chem.*, **283**, 21501–21508.

Zhou, C., Li, J., Li, C. and Zhang, Y. (2016) Improvement of betulinic acid biosynthesis in yeast employing multiple strategies. *BMC Biotechnol.*, **16**, 59.

List of publications

This dissertation is based on the following publications:

1. **Tamura, K., Seki, H., Suzuki, H., Kojoma, M., Saito, K. and Muranaka, T.** (2017) CYP716A179 functions as a triterpene C-28 oxidase in tissue-cultured stolons of *Glycyrrhiza uralensis*. *Plant Cell Rep.*, **36**, 437–445.
2. **Tamura, K., Teranishi, Y., Ueda, S., Suzuki, H., Kawano, N., Yoshimatsu, K., Saito, K., Kawahara, N., Muranaka, T. and Seki, H.** (2017) Cytochrome P450 monooxygenase CYP716A141 is a unique β-amyrin C-16β oxidase involved in triterpenoid saponin biosynthesis in *Platycodon grandiflorus*. *Plant Cell Physiol.*, **58**, 874–884.
3. **Tamura, K., Yoshida, K., Hiraoka, Y., Sakaguchi, D., Chikugo, A., Mochida, K., Kojoma, M., Mitsuda, N., Saito, K., Muranaka, T. and Seki, H.** The basic helix-loop-helix transcription factor GubHLH3 positively regulates soyasaponin biosynthetic genes in *Glycyrrhiza uralensis*. (in preparation)

Reviews:

1. **關光, 田村啓太, 村中俊哉** (2014) トリテルペノイド生合成における P450 の多様な機能, 薬用植物・生薬の最前線 ～国内栽培技術から品質評価, 製品開発まで～, シーエムシー出版 (東京), 132–139. (in Japanese)
2. **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.
3. **關光, 田村啓太, 村中俊哉** (2016) 植物の非糖質系天然甘味成分の生合成, *Foods Food Ingredients J. Jpn.* **221,** 225–231. (in Japanese)
4. **Seki, H., Tamura, K. and Muranaka, T.** (2017) Plant-derived isoprenoid sweeteners: recent progress in biosynthetic gene discovery and perspectives on microbial production. *Biosci. Biotechnol. Biochem.* (in press)

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Prof. Dr. Toshiya Muranaka, for his guidance, support, and encouragement during my PhD study, and for providing me with many opportunities to meet and establish collaborations with other researchers. I would like to thank Prof. Dr. Kazuhito Fujiyama and Prof. Dr. Hajime Watanabe for their critical reading of my dissertation and for providing useful suggestions.

I would like to thank Dr. Hikaru Seki for his helpful advice and support during my PhD study. I would also like to thank Dr. Ery Odette Fukushima, Dr. Satoru Sawai, Dr. Yuhta Nomura, Dr. Takatoshi Wakabayashi, Dr. Shuhei Yasumoto, and Dr. Keiko Kobayashi for their useful advice and suggestions. I would like to acknowledge Ms. Yasuko Hiraoka, Mr. Koki Yoshida, Mr. Yuga Teranishi, Mr. Daiki Sakaguchi, Mr. Shinya Ueda, and Ms. Ayaka Chikugo for their considerable contributions to several experiments. I would like to thank Dr. Kyoko Inoue for technical support with NMR analysis and Ms. Keiko Fukamoto for technical assistance.

I would like to deeply acknowledge many collaborators who supported this study. I would like to thank Dr. Nobutaka Mitsuda for the opportunity to perform high-throughput screening of transcription factors at the National Institute of Advanced Industrial Science and Technology (AIST), for providing several experimental materials, and for helpful suggestions on my manuscript. I would like to thank Ms. Fumie Tobe for technical assistance on Y1H experiments at the AIST and Dr. Shingo Sakamoto for technical advice on transient co-transfection assays using protoplasts. I would like to thank Dr. Hideyuki Suzuki for the opportunity to learn bioinformatics at the Kazusa DNA Research Institute and for assistance with RNA-Seq analysis. I would like to thank Dr. Kazuto Mannen for his kind support at the institute. I would like to thank Dr. Mareshige Kojoma for providing tissue-cultured stolons and seeds of *G. uralensis*, and for technical advice for establishing hairy root lines. I would like to thank Dr. Noriaki Kawano and Dr. Kayo Yoshimatsu for providing plant materials and photos of *P. grandiflorus*. I would like to thank Dr. Keiich Mochida for mining transcription factors from transcriptome data and Dr. Ryosuke Sano for technical advice with the *de*

novo assembly tools of RNA-Seq analysis. I would like to thank Dr. Kiyoshi Ohyama for providing authentic standard compounds and Dr. Ryo Misaki for providing tobacco BY-2 cells. I would like to thank Prof. Dr. Kazuki Saito and Dr. Nobuo Kawahara for encouragement, and the Takeda Garden for Medicinal Plant Conservation, Kyoto for providing *G. uralensis* plant materials.

I would like to deeply acknowledge the Yoshida Scholarship Foundation for financial and hearty support during my PhD study at Osaka University. They also provided great opportunities to establish good friendships with students studying at other universities.

I would like to express my thanks to all members of the Cell Technology Laboratory, both past and present, for their kind support and encouragement. I would also like to thank my friends for their encouragement.

Finally, I would like to thank my family, especially my parents, who have always supported and encouraged me during my PhD study.

December 2017