P450s and UGTs: Key Players in the Structural Diversity of Triterpenoid Saponins

Hikaru Seki*, Keita Tamura and Toshiya Muranaka

Department of Biotechnology, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita Osaka, 565-0871 Japan

*Corresponding author: E-mail, hseki@bio.eng.osaka-u.ac.jp; Fax, +81-6-6879-7426.

(Received February 9, 2015; Accepted April 20, 2015)

The recent spread of next-generation sequencing techniques has facilitated transcriptome analyses of non-model plants. As a result, many of the genes encoding enzymes related to the production of specialized metabolites have been identified. Compounds derived from 2,3-oxidosqualene (the common precursor of sterols, steroids and triterpenoids), a linear compound of 30 carbon atoms produced through the mevalonate pathway, are called triterpenes. These include essential sterols, which are structural components of biomembranes; steroids such as the plant hormones, brassinolides and the toxin in potatoes, solanine; as well as the structurally diverse triterpenoids. Triterpenoids containing one or more sugar moieties attached to triterpenoid aglycones are called triterpenoid saponins. Triterpenoid saponins have been shown to have various medicinal properties, such as anti-inflammatory, anticancerogenic and antiviral effects. This review summarizes the recent progress in gene discovery and elucidates the biochemical functions of biosynthetic enzymes in triterpenoid saponin biosynthesis. Special focus is placed on key players in generating the structural diversity of triterpenoid saponins, cytochrome P450 monoxygenases (P450s) and the UDP-dependent glycosyltransferases (UGTs). Perspectives on further gene discovery and the use of biosynthetic enzymes for the microbial production of plant-derived triterpenoid saponins are also discussed.

Keywords: Biosynthesis • Cytochrome P450 monoxygenase • Oxidosqualene cyclase • Plant specialized metabolites • Triterpenoid saponin • UDP-dependent glycosyltransferase.

Abbreviations: DDMP, 2,3-dihydro-2,5-dihydroxy-6-methyl-4H-pyran-4-one; OSC, oxidosqualene cyclase; P450, cytochrome P450; UGT, UDP-dependent glycosyltransferase.

Introduction

Triterpenoid saponins are a class of structurally diverse plant specialized metabolites containing one or more sugar moieties attached to hydrophobic triterpenoid aglycones. Although the biological role of triterpenoid saponins in plants remains largely unclear, these compounds have been shown to have various medicinal properties, including anti-inflammatory (Matsui et al. 2004, Sun et al. 2010), anticancerogenic (Man et al. 2010) and antiviral (Cinatl et al. 2003, Zhao et al. 2008) effects. Like glycyrhizin in licorice and ginsenosides in ginseng, several of these compounds are major components of traditional herbal medicines described in Japanese pharmacopoeia.

The first step in the structural diversification of triterpenoid saponins is the cyclization of 2,3-oxidosqualene by a group of enzymes called oxidosqualene cyclases (OSCs). Numerous OSCs exist, each with its own cyclization mechanism, and >100 different cyclic triterpene skeletons have been found in nature (Xu et al. 2004). More than 80 OSCs have been functionally characterized from plants (Thimmappa et al. 2014). The OSC genes for synthesizing β-amyrin (a pentacyclic triterpene having an oleane skeleton) or lupeol (a pentacyclic triterpene having a lupane skeleton), which are widely distributed in the plant kingdom, have been isolated from many types of plant, including medicinal herbs [for other plant-derived OSC enzyme functions identified to date, refer to Augustin et al. (2011) and Thimmappa et al. (2014)].

In many cases, the cyclic skeleton synthesized by OSC undergoes site-specific oxidation by cytochrome P450s (P450s) to produce sapogenins (the non-sugar part of the saponin) with various structures. P450s are monoxygenases that introduce an oxygen atom into their substrates from molecular oxygen. This process is involved in the biosynthesis of various specialized metabolites in plants. As mentioned above, triterpenoid saponin biosynthesis involves the site-specific oxidation of the cyclized skeleton, primarily through the introduction of hydroxyl, carboxyl or epoxy groups. In plants, P450s form a large gene family. There are around 300 P450 genes in the genomes of flowering plants (Nelson and Werck-Reichhart 2011). The nomenclature of P450s is based on the homology of their amino acid sequences. If the amino acid sequence homology of a P450 is >50%, it is placed into a given family. If the homology is >55%, it is classified into a given subfamily.

Plant triterpenoids are further diversified through glycosylations at hydroxyl and/or carboxyl groups by UGTs. UGTs catalyze glycosylation using a UDP-sugar donor such as UDP-glucose, UDP-galactose, UDP-arabinose, UDP-rhamnose, UDP-xyllose or UDP-glucuronic acid. Variations in the number of sugar chains, their composition and their position on the triterpene scaffold are thought to have a large impact on intra-/extracellular transport and storage in plants, as well as biological activity, taste and bioabsorbability (Bowles et al. 2005). Similar to the P450s, UGTs also form a large gene family in plants. For example, there are 115 UGTs in the...
P450s in the Biosynthesis of Triterpenoid Saponins

P450s in soyasaponin biosynthesis

Many oleanane-type triterpenoid saponins, called soyasaponins, have been isolated and characterized in soybean seeds (Takada et al., 2013, and references therein). Soyasaponins are divided into two groups according to their aglycone components. DDMP (2,3-dihydro-2,5-dihydroxy-6-methyl-4H-pyran-4-one) saponins, found in both seed hypocotyls and cotyledons, use soyasapogenol B as a sapogenin. Group A saponins, which are highly concentrated in the seed hypocotyl, use soyasapogenol A as a sapogenin. Soyasapogenol B is biosynthesized by the hydroxylation of C-22 and C-24 in β-amyrin. Soyasapogenol A includes an additional hydroxylation at C-21. In 2006, CYP93E1, which catalyzes the hydroxylation of C-24 in β-amyrin, was the first P450 to be identified in relation to triterpenoid saponin biosynthesis (Shibuya et al. 2006). Soyasapogenol B is a sapogenin common to the Fabaceae and it is therefore expected that CYP93E1 orthologs would be common in plants of the Fabaceae. *Medicago truncatula* (Shibuya et al. 2006), *Glycyrhiza uralensis* CYP93E3 (Seki et al. 2008) and six additional CYP93E orthologs (CYP93E4–CYP93E9) (Moses et al. 2014b), which hydroxylate C-24 in β-amyrin, have been identified with a high sequence identity (>80%) to CYP93E1. Oleane-type saponins, which have a hydroxyl or carboxyl group at the C-24 position, have been discovered in various plants in addition to the Fabaceae. However, P450s of the CYP93E subfamily have yet to be found outside the Fabaceae.

CYP72A61v2 in *M. truncatula* catalyzes the hydroxylation of C-22 in 24-hydroxy-β-amyrin (Fukushima et al. 2013). Consequently, it appears that CYP93E2 and CYP72A61v2 function in a co-ordinated manner to produce soyasapogenol B in *M. truncatula*. The enzyme that catalyzes the additional hydroxylation of C-21 in the biosynthesis of soyasapogenol A has not yet been identified.

P450s in glycyrrhizin biosynthesis

Glycyrrhizin is a triterpenoid saponin derived from the underground parts of *Glycyrrhiza* plants, which are common medicinal herbs known as Chinese licorice, of the Fabaceae family. Glycyrrhizin has medicinal properties such as an anti-inflammatory activity, and is involved in the recovery of liver function. Glycyrrhizin is also 150 times sweeter than sucrose (Kitagawa 2002). Many forms of licorice and its extracts are commercially available worldwide as medicinal materials and sweetening agents (Hayashi and Sudo 2009, Kojoma et al. 2010).

Glycyrrhizin is biosynthesized by a glycosylation reaction following the production of sapogenin glycyrrhetic acid by the oxidation of C-11 and C-30 in β-amyrin. In *G. uralensis*, CYP88D6 (β-amyrin-11-oxidase; Seki et al. 2008) catalyzes the oxidation at the C-11 position of β-amyrin, thereby converting it to an intermediate product, 11-oxo-β-amyrin. CYP72A154 is a P450 that produces glycyrrhetic acid by catalyzing the oxidation of C-30 in 11-oxo-β-amyrin (Seki et al. 2011). These data show that the CYP88D and CYP72A subfamilies are involved in the biosynthesis of triterpenoid saponins. In addition, two CYP88D subfamilies of P450 from *M. truncatula* (CYP88D2 and CYP88D3), *Lotus japonicus* (CYP88D4 and CYP88D5) and *Astragalus sinicus* (CYP88D7 and CYP88D8) have been identified, although their enzyme functions have not been elucidated. Like the CYP93Es, the CYP88D subfamily P450s had been found only in members of the Fabaceae family. These results suggest that CYP88Ds evolved specifically in one group of the Fabaceae family. In addition, CYP72A63, CYP72A61v2 and CYP72A68v2 from *M. truncatula* oxidize C-30, C-22 and C-23 of the oleanane skeleton, respectively (Seki et al. 2011, Fukushima et al. 2013). The CYP72A subfamily is distributed not just among members of the Fabaceae, but across the entire plant kingdom, and an assortment of >200 different molecules has been identified. However, the biosynthetic functions of most of these have not yet been elucidated. Other than the aforementioned Fabaceae CYP72As, only the enzyme function of *Catharanthus roseus* CYP72A1 (secologanin synthase; Irmler et al. 2000) and CYP72A224 (7-deoxyloganic acid 7-hydroxylase; Salim et al. 2013, Miettinen et al. 2014), which are involved in indole alkaloid biosynthesis, has been elucidated.

CYP716Y1 in saikosaponin biosynthesis

*Bupleurum* roots (Apiaceae family) are commonly used in traditional Asian medicines for reducing fever and relieving irritability. Saikosaponins are recognized as the most important pharmacological constituents in *Bupleurum* root extract and, to date, >120 oleanane- and ursane-type saponins have been isolated from *Bupleurum* *species* (Ashour and Wink 2011). The biosynthesis of two major saikosaponins, saikosaponin A and saikosaponin D, includes hydroxylation at C-16 and C-23 of the oleanane skeleton and the formation of an ether linkage between C-13 and C-28. Based on these reactions, CYP716Y1, which catalyzes the hydroxylation of C-16 of β-amyrin, has been identified in *Bupleurum falcatum* (Moses et al. 2014a). Saikosaponins D and A have a 16α-hydroxyl and 16β-hydroxyl...
Fig. 1 Modification of triterpene scaffolds catalyzed by characterized P450s. Arrows represent the conversion catalyzed by biochemically characterized P450s. Functional groups introduced by P450s are highlighted in red.
group, respectively. Note that because CYP716Y1 specifically catalyzes hydroxylation only in the α-configuration, it has been hypothesized that an unknown enzyme catalyzes 16β-hydroxylation.

Like some CYP716A P450s (described later), CYP716Y1 exhibits the same degree of activity towards β- and α-amarin (ursane skeleton), both pentacyclic skeletons. The enzyme that catalyzes the hydroxylation of the C-23 position remains unidentified. However, since the discovery that CYP72A68v2 of *M. truncatula* catalyzes the hydroxylation of C-23 in β-amarin-based scaffolds (oleanolic acid) to produce gypsogenic acid, the participation of a P450 belonging to the same subfamily is possible in *Bupleurum*.

**P450s in maesasaponin biosynthesis**

*Maesa lanceolata* is a shrub or small tree of the Myrsinaceae family indigenous to East, South and Central Africa. African
traditional healers use parts and plant extracts to treat a wide range of diseases, including infectious hepatitis, bacillary dysentery, impetigo, zona, dermatoses and neuropathies (Moses et al. 2015a). Maesasaponins are derived from multiple modifications of oleane-type backbones. Many contain characteristic C-13,28 hemiacetal or ester bridges, and display oxidations at the C-16, C-21 and/or C-22 positions. CYP87D16, which catalyzes the C-16x hydroxylation of α-amyrin, was recently identified based on these reactions (Moses et al. 2015a). As mentioned above, another P450, CYP716Y1 from *B. falcatum*, is an active α-amyrin C-16x hydroxylase (Moses et al. 2014a). CYP87D16 has 27% homology with CYP716Y1 and therefore belongs to a different P450 family, although they are functional homologs. At present, this is the only such example in which P450s from two different families exhibit the same biochemical function.

The same research group identified CYP716A75 from *M. lanceolata* as a β-amyrin-28-oxidase that produces oleanolic acid (Moses et al. 2015a). Note that CYP716A12 of *M. truncatula* and both CYP716A15 and CYP716A17 of *Vitis vinifera* produce oleanolic acid through an identical β-amyrin-28-oxidase activity (Carelli et al. 2011, Fukushima et al. 2011). In addition to β-amyrin, CYP716A12 and CYP716A15 catalyze the oxidation of C-28 in α-amyrin and lupeol, producing ursolic acid or betulinic acid, respectively (Fukushima et al. 2011). A similar activity has been reported for *C. roseus* CYP716A1 (Huang et al. 2012).

There are >20 subfamilies of CYP716 among the dicots. The specific functions of these enzymes are hypothesized to vary according to the plant’s speciation. Other than the aforementioned CYP716s, which are involved in the oxidation of triterpene, only *Ginkgo biloba* CYP716B (Zhang et al. 2014) and *Stevia rebaudiana* CYP716D (Brandle and Richman 2008, Mizutani and Ohta 2010) catalyze the hydroxylation of diterpenoids, taxoids and ent-kaurenoic acid, respectively.

**P450s in avenacin biosynthesis**

Avenacins, triterpenoid saponins that accumulate in the roots of oats (*Avena* sp., Poaceae family), are thought to be important determinants of disease resistance. The biosynthesis of avenacin A-1 includes the hydroxylation of C-16, C-21 and C-23, the oxidation of C-30 and the C-12,13 epoxidation in the oleane skeleton. These reactions indicated that CYP51H10 from *Avena strigosa* is a multifunctional enzyme that catalyzes both the hydroxylation of C-16 and the C-12,13 epoxidation (Geisler et al. 2013). While P450s capable of catalyzing both hydroxylation and epoxidation have been identified in microorganisms (Anzai et al. 2008, Crešnar and Petrović 2011), to the best of our knowledge CYP51H10 is, at present, the only such example in plants. The CYP51H1 subfamily is found only in monocots, and nine members exist in *Oryza sativa* (CYP51H1–CYP51H9). However, to date, enzyme function has only been elucidated for CYP51H10.

**P450s in ginsenoside biosynthesis**

Ginseng (*Panax ginseng*, of the family Araliaceae) is one of the most popular medicinal herbs and contains a series of pharmacologically active triterpenoid saponins, ginsenosides, in its roots. *Panax ginseng* produces two different types of triterpenoid saponins: dammarane-type saponins, such as ginsenoside Rb1 and Rg2, derived from dammaranediol-II, and ginsenoside Rg3, which is an oleane-type saponin derived from α-amyrin. In the biosynthesis of dammarane-type saponins, CYP716A47, which produces protopanaxadiol (the sapogenin of ginsenoside Rb1) through the hydroxylation of C-12 in dammaranediol-II, and CYP716A53v2, which hydroxylates the C-6 of protopanaxadiol to produce protopanaxatriol (the sapogenin of ginsenoside Rg3), have been identified (Han et al. 2011, Han et al. 2012). The same research group identified CYP716A52v2 (Han et al. 2013) as a β-amyrin-28-oxidase that produces oleanolic acid as the sapogenin of ginsenoside Ro. This activity is identical to the reported activities of close homologs from other plant species, as described above.

**P450s related to the biosynthesis of other triterpenoids**

CYP71D353 is a multifunctional enzyme found in *L. japonicus* that catalyzes both the hydroxylation of C-20 and the oxidation of C-28 in dihydro-lupeol to produce 20-hydroxy-betulnic acid via 20-hydroxy-lupeol (Krokida et al. 2013). We have previously identified *M. truncatula* CYP716A12 and *V. vinifera* CYP716A15 as P450s that catalyze the oxidation of C-28 of both β-amyrin and lupeol. CYP71D353 differs from these in showing no activity towards β-amyrin. There are several CYP71D subfamily P450s that are related to mono- and sesquiterpenoid biosynthesis in plants of the Lamiaceae and Solanaceae family, respectively (Weitzel and Simonsen 2013). CYP71D353 is the first P450 found to be involved in triterpenoid biosynthesis in the CYP71D subfamily. It is still unknown whether 20-hydroxy-betulnic acid is the final product in this pathway or whether it is an intermediate (sapogenin) to subsequent glycosylation.

*Aster terrastris*, a member of the Asteraceae family, is an extensively studied species for its ability to produce the anti-malarial sesquiterpenoid, artemisinin. In addition to producing a variety of sesquiterpenoids, *A. annua* also accumulates triterpenoids in the leaf surface wax. CYP716A14v2 catalyzes oxidation of the hydroxyl group at C-3 of the pentacyclic triterpene α-amyrin, β-amyrin and δ-amyrin to produce triterpenoids with a C-3 carbonyl group (Moses et al. 2015b).

In nature, several pentacyclic triterpenes with a carbonyl group at the C-3 position have been identified. Shionone, the major triterpenoid of the roots of the medicinal plant *Aster tataricus*, is a tetracyclic triterpene containing a C-5 carbonyl group: shionone is cyclized from 2,3-oxidosqualene in a single enzymatic reaction catalyzed by the OSC, shionone synthase (Sawai et al. 2011). Likewise, friedelin, a pentacyclic triterpene with a carbonyl group at the C-3 position, is cyclized from 2,3-oxidosqualene by friedelin synthase (Wang et al. 2010). In contrast, the carbonyl groups at the C-3 positions of triterpenoids found in leaf surface wax of *A. annua* are generated via an alternative biosynthetic pathway: a two-step enzymatic process catalyzed by OSC and CYP716A14v2.

The enzyme functions of three *Arabidopsis* P450s have been characterized. Castillo et al. (2013) showed that CYP708A2 oxidizes a tricyclic triterpenoid, thalianol, to 7β-hydroxylthalianol,
CYP705A1 cleaves the side chain of arabidiole to give a C19 methyl ketone, and CYP71A16 hydroxylates an allylic methyl group of marneral/marnerol to yield 23-hydroxymarneral/23-hydroxymarnerol (Castillo et al. 2013). Another P450, CYP705A5, appears to catalyze the desaturation of the product of CYP708A2, 7β-hydroxythalianol (Field and Osbourn 2008). The biochemical function of CYP705A5 remains uncharacterized.

**UGTs Responsible for Decorating Triterpene Scaffolds**

**UGTs related to soyasaponin biosynthesis in soybean**

Soyasaponins are divided into two groups according to their aglycone components: DDMP saponins that use soyasapogenol B as a sapogenin, and Group A saponins that use soyasapogenol A as a sapogenin. The primary DDMP saponin, soyasaponin βg, contains one sugar chain comprising a glucuronic acid, a galactose and a rhamnose attached at the C-3 hydroxyl group of soyasapogenol B. UGT73P2 transfers a galactose moiety to the glucuronic acid already bound to the C-3 of soyasapogenol B as a second sugar, and UGT91H4 transfers a rhamnose moiety to the galactose as the third sugar (Shibuya et al. 2010).

In addition to the C-3 substituents, the hydroxyl group at the C-22 position in Group A saponins is also bound to a sugar chain. Two UGTs participating in the formation of the sugar chain at C-22 have been identified (Sayama et al. 2012). UGT73F4 and UGT73F2 transfer xylose and glucose, respectively, to the arabinose moiety already bound to C-22.

However, the UGT that catalyzes the transfer of glucuronic acid to the hydroxyl group at C-3, or the transfer of an arabinose to the hydroxyl group at C-22, remains unidentified.

**UGTs related to saponin biosynthesis in M. truncatula**

The model legume *M. truncatula* synthesizes >30 different triterpenoid sapogenins from at least five oleanane-type sapogenins: soyasapogenols B and E, medicagenic acid, hederagenin and bayogenin. Two UGTs involved in these syntheses have been functionally characterized, although the precise position(s) of glycosylation by these enzymes has not been determined (Achnine et al. 2005). UGT73K1 is specific for hederagenin and soyasapogenols B and E, while UGT71G1 is specific for medicagenic acid. In addition, in vitro studies, the glycosylation of certain isoflavones and the flavonol quercetin with UGT71G1 was more efficient than the glycosylation of triterpenoid sapogenins. However, treatment with methyl jasmonate induced UGT71G1 transcripts and an accumulation of triterpenoid sapogenins but not flavonoid/isorflavonoid glycosides in cell cultures. These data suggest UGT71G1 is involved in reactions with triterpenoid sapogenins but not in planta (iso)flavonoid biosynthesis.

Another UGT from *M. truncatula*, UGT73F3, is active against multiple sapogenins and was confirmed to glucosylate hederagenin at the C-28 position in vitro (Naoumkina et al. 2010). Furthermore, loss of UGT73F3 function caused a reduction in levels of C-28 glycosylated triterpenoids with a concurrent decrease in plant growth. These results confirmed the in vivo function of UGT73F3 in saponin biosynthesis and suggest that the accumulation of non-glycosylated sapogenins might be toxic.

**UGTs in vaccaroside B biosynthesis**

*Saponaria vaccaria*, a member of the Caryophyllaceae family, is an annual herb widely distributed in Asia, Europe and other parts of the world. The seeds of this plant are used in traditional Chinese medicines prescribed for amenorrhea, breast infections and stimulation of lactation. The most common sapogenins in this family are quillaic acid, gypsogenic acid and gypsojenin, which have hydroxyl and carboxyl groups at the C-3 and C-28 positions in β-amyrin, respectively (Meesapyodsuk et al. 2007). UGT74M1 from *S. vaccaria* catalyzes the formation of an ester bond between glucose and the carboxyl group of C-28 in gypsogenic acid, which is a putative intermediate in the biosynthesis of vaccaroside B (Meesapyodsuk et al. 2007). While UGT74M1 belongs to a different UGT family from the aforementioned UGT73F3, their enzymatic functions are similar. It has been hypothesized that this functional similarity is due to convergent evolution.

**UGTs related to triterpenoid saponin biosynthesis in Barbarea vulgaris**

*Barbarea vulgaris*, a member of the Brassicaceae, is the only crucifer known to produce triterpenoid saponins. Four UGTs (UGT73C10–UGT73C13) known to catalyze 3-O-glucosylation of the oleanane sapogenins oleanolic acid and hederagenin were identified in *B. vulgaris* (Augustin et al. 2012). Of these, UGT73C10 and UGT73C11 were specific for a hydroxyl group at the C-3 position. In comparison, the regiospecificities of UGT73C12 and UGT73C13 were relatively low; these enzymes catalyze the formation of an ester bond between glucose and the carboxyl group at the C-28 position of oleanolic acid and hederagenin. These four UGTs also catalyze the transfer of glucose to C-3 and/or C-28 of betulinic acid, which is a lupane-type sapogenin.

**UGTs in ginsenoside biosynthesis**

Three UGTs that participate in the biosynthesis of dammarane-type saponins have been identified in *P. ginseng*. Of these, UGT71A27 catalyzes the glucosylation of the hydroxyl group at the C-20 position in dammarenediol-II to produce compound K (Yan et al. 2014). UGT74AE2 catalyzes the glucosylation of the C-3 hydroxyl group of protopanaxadiol and compound K to produce ginsenoside Rh2 and ginsenoside F2, respectively. UGT94Q2 transfers glucose as a second sugar onto the glucose moiety already bound to C-3 of ginsenoside Rh2 and ginsenoside F2 to produce ginsenoside Rg3 and ginsenoside Rd, respectively (Jung et al. 2014). However, the UGTs participating in the formation of the sugar chain of ginsenoside Ro, which is an oleanane-type saponin derived from β-amyrin, remain unidentified.
Perspective
The recent availability of inexpensive and high-throughput sequencing technologies has enabled single laboratories to produce sequence libraries and information on gene expression profiles from species of interest. At the same time, the development of several large-scale, consortium-based transcriptome analyses has led to the accumulation of massive volumes of sequence data in many non-model plant species. The PhytoMetaSyn project (www.phytometasyn.ca) provides the transcriptome data of 75 non-model plants that produce compounds of interest for biotechnological applications, and the web-based BLAST server allows public access to assembled transcriptome databases for all 75 plant species (Facchini et al. 2012, Xiao et al. 2013). The 1KP project aims to obtain transcriptome data from 1,000 plant species with exemplars from all of the major lineages of green plants (Matasci et al. 2014). The Medicinal Plant Genomics Resource project provides transcriptome and metabolome analyses of 14 medicinal plant species. It is expected that the discovery of new enzyme genes in the biosynthesis of triterpenoid saponins, as well as other types of specialized metabolites, will be further accelerated by candidate gene mining from these data sets. In addition to transcriptome mining, genome mining for potential biosynthetic gene clusters is another standard strategy for gene discovery in plant-specialized metabolic pathways. Clusters of non-homologous genes involved in the biosynthesis of specialized metabolites from several chemical classes have been found in numerous plant species including O. sativa (diterpene phytalexin), Papaver somniferum (isoquinoline alkaloids) and Solanum tuberosum (steroidal glycoalkaloids). These clusters are described in more detail in Boycheva et al. (2014). Regarding triterpenoids, (partial) gene clusters comprising an OSC gene and functionally associated P450(s) have been identified for the biosynthesis of avenacin (Avena sativa), desaturated thalian-diol (A. thaliana) and 20-hydroxy-betulinic acid (L. japonicus).

Remarkable progress has been made in engineering the production of various plant-specialized metabolites in microbial hosts, such as Escherichia coli and yeast, through the introduction of biosynthetic genes. Engineered Saccharomyces cerevisiae strains optimized to accumulate 2,3-oxidosqualene, the precursor of triterpenoids and sterols, has been used as the host for the expression of triterpenoid saponin biosynthetic genes, OSC, P450 genes and UGT genes. Recently, Yan et al. (2014) succeeded in producing a bioactive ginsenoside (compound K), which has received approval from the China Food and Drug Administration to commence clinical trials (CDEL20130379) for arthritis prevention and treatment (Yan et al. 2014). This was accomplished by introducing genes for ginsenoside biosynthetic enzymes, dammarenediol-II synthase (an OSC that produces dammarenediol-II from 2,3-oxidosqualene), CYP716A47 (produces protopanaxadiol via the hydroxylation of C-12 in dammarenediol-II) and UGT71A27 (glucosylation of the C-20 hydroxyl group). Another example by Moses et al. (2014a) describes the production of 3-O-Glc-echinocystic acid by the combined expression of genes for Glycyrrhiza glabra β-amyrin synthase, M. truncatula CYP716A12 (C-28 oxidation), B. falcatus CYP716Y1 (C-16 hydroxylation) and B. vulgaris UGT73C11 (glucosylation of the C3-hydroxyl group). Thus, the combination of various OSCs, P450s and UGTs isolated from various plant species in a yeast cell allows biosyntheses of rare natural or even non-natural triterpenoid saponins.

It is expected that triterpenoid pathway engineering in recombinant yeasts will develop into a process amenable to the commercial production of useful triterpenoids. This strategy will also generate more highly functional triterpenoids through close evaluations of biological activity and the construction of triterpenoid libraries.

Funding
This work was supported by the Ministry of Education, Culture, Sports, Science, and Technology of Japan [Grants-in-Aid for Scientific Research (Nos. 25108715 and 26450123 to H.S.); the Yamada Science Foundation [funding to H.S.]; the Yoshida Scholarship Foundation [Doctor 21 scholarship to K.T.].

Disclosures
The authors have no conflicts of interest to declare.

References
ACHNINE, L., HUHMAN, D.V., FARAG, M.A., SUMNER, L.W., BLOUNT, J.W. AND DIXON, R.A. (2005) Genomes-based selection and functional characterization of triterpene glycosyltransferases from the model legume Medicago truncatula. Plant J. 41: 875–887.
ANZAI, Y., LI, S., CHAULAGAIN, M.R., KINOSHITA, K., KATO, F., MONTGOMERY, J., ET AL. (2008) Functional analysis of MyccI and MyccG, cytochrome P450 enzymes involved in biosynthesis of mycinamicin macrolide antibiotics. Chem. Biol. 15: 950–959.
ASHOUR, M.L. AND WINK, M. (2011) Genus Bupleurum: a review of its phytochemistry, pharmacology and modes of action. J. Pharm. Pharmacol. 63: 305–321.
AUGUSTIN, J.M., DROK, S., SHINODA, T., SANMIYA, K., NIelsen, J.K., KHAKIMOV, B., ET AL. (2012) UDP-glycosyltransferases from the UGT73C subfamily in Barbarea vulgaris catalyze sapogenin 3-O-glucosylation in saponin-mediated insect resistance. Plant Physiol. 160: 1881–1895.
AUGUSTIN, J.M., KUZINA, V., ANDERSEN, S.B. AND BAK, S. (2011) Molecular activities, biosynthesis and evolution of triterpenoid saponins. Phytochemistry 72: 435–457.
BOWLES, D., ISAYENKOVA, J., LIM, E-K. AND POPPENBERGER, B. (2005) Glycosyltransferases: managers of small molecules.Curr. Opin. Plant Biol. 8: 254–263.
BOYCHEVA, S., DAVIES, L., WOLFENDER, J.-L. AND FITZPATRICK, T.B. (2014) The rise of operon-like gene clusters in plants. Trends Plant Sci. 19: 447–459.
BRANDLE, J. AND RICHMAN, A. (2008) Composition and Methods for Producing Steviol and Steviol Glycosides. European Patent No. EP1897951
CARELLI, M., BIAZZI, E., PANARA, F., TAVA, A., SCARAMELLI, L., PORCEDDU, A., ET AL. (2011) Medicagrace 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. Amer. Chem. Soc. 135, 5885–5894.
Cinatl, J., Morgenstern, B., Bauer, G., Chandra, P., Rabenau, H. and Doerr, H.W. (2003) Glycyrrhizin, an active component of liquorice roots, and replication of SARS-associated coronavirus. *Lancet* 361: 2045–2046.

Crešían, B. and Petrič, S. (2011) Cytochrome P450 enzymes in the fungal kingdom. *Biochim. Biophys. Acta* 1814: 29–35.

Facchini, P.J., Bohlmann, J., Covello, P.S., De Luca, V., Mahadevan, R., Page, J.E., et al. (2012) Synthetic biosystems for the production of high-value plant metabolites. *Trends Biotechnol.* 30: 127–131.

Field, B. and Osbourn, A.E. (2008) Metabolic diversification—indigenous assembly of opeon-gen-like clusters in different plants. *Science* 320: 543–547.

Fukushima, E.O., Seki, H., Ohyama, K., Ono, E., Umemoto, N., Mizutani, M., et al. (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., et al. (2013) Combinatorial biosynthesis of legume natural and rare triterpenoids in engineered yeast. *Plant Cell Physiol.* 54: 740–749.

Geisler, K., Hughes, R.K., Sainsbury, F., Lomonossoff, G.P., Rejzek, M., Han, J.Y., Hwang, H.S., Choi, S.W., Kim, H.J. and Choi, Y.E. (2011) The cyt P450 enzyme CYP716A47 catalyzes the formation of protopanaxadiol from dammar-enediol-II during ginsenoside biosynthesis in *Panax ginseng*. *Plant Cell Physiol.* 53: 1535–1545.

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

Huang, L., Li, J., Ye, H., Li, C., Wang, H., Liu, B., et al. (2012) Molecular characterization of the pentacyclic triterpenoid biosynthetic pathway in *Catharanthus roseus*. *Plant Physiol.* 156: 1571–1581.

Kojoma, M., Ohyama, K., Seki, H., Hiraoka, Y., Aazu, S.N., Sawa, S., et al. (2010) In vitro proliferation and triterpenoid characteristics of licorice ingredient in Chinese medicine. *FEBS J.* 273: 948–959.

Kitagawa, I. (2002) Licorice root. A natural sweetener and an important ingredient in Chinese medicine. *Pure Appl. Chem.* 74: 1189–1198.

Kojima, M., Ohyama, K., Seki, H., Hiraoka, Y., Aazu, S.N., Sawa, S., et al. (2010) *in vitro* proliferation and triterpenoid characteristics of licorice (*Glycyrrhiza uralensis* Fischer, *Leguminosae*) stolons. *Plant Biotechnol.* 27: 59–66.

Krokida, A., Delis, C., Geisler, K., Garagounis, C., Tsikou, D., Peña-Rodríguez, L.M., et al. (2013) A metabolic gene cluster in *Lotus japonicus* discloses novel enzyme functions and products in triterpenoid biosynthesis. *New Phytol.* 200: 675–690.

Man, S., Gao, W., Zhang, Y., Huang, L. and Liu, C. (2010) Chemical study and medical application of saponins as anti-cancer agents. *Fitoterapia* 81: 703–714.

Matasci, N., Hung, L.H., Yan, Z., Carpenter, E.J., Wickert, N.J., Mirarab, S., et al. (2014) Data access for the 1,000 Plants (1Kp) project. *GigaScience* 3: 17.

Matsui, S., Matsumoto, H., Sonoda, Y., Ando, K., Aizu-Yokota, E., Sato, T., et al. (2004) Glycyrrhizin and related compounds down-regulate production of inflammatory chemokines IL-8 and eotaxin 1 in a human lung fibroblast cell line. *Int. Immunopharmacol.* 4: 1633–1644.

Meesapopdulsak, D., Balsevich, J., Reed, D.W. and Covello, P.S. (2007) Saponin biosynthesis in *Saponaria vaccaria* cDNAs encoding β-amyrin synthase and a triterpene carboxylic acid glycosyltransferase. *Plant Physiol.* 143: 959–969.

Miettinen, K., Dong, L., Navrot, N., Schneider, T., Burlat, V., Pollier, J., et al. (2014) The seco-iridoid pathway from *Catharanthus roseus*. *Nat. Commun.* 5: 3606.

Mizutani, M. and Ohto, D. (2010) Diversification of P450 genes during land plant evolution. *Annu. Rev. Plant Biol.* 61: 291–315.

Moses, T., Pollier, J., Almagno, L., Buyst, D., Van Montagu, M., Pedreño, M.A., et al. (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., Apera, S., Pieters, L., Thevelein, J.M., et al. (2015a) Unraveling the triterpenoid saponin biosynthesis of the African shrub *Maesa lanceolata*. *Plant Sci.* 122: 125–135.

Moses, T., Pollier, J., Shen, Q., Soetaert, S., Reed, J., Erffelink, M.-L., et al. (2015b) OSC2 and CYP716A14v2 catalyze the biosynthesis of triterpenoids for the cuticle of aerial organs of *Artemisia annua*. *Plant Cell* 27: 286–310.

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

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

Salm, V., Yu, F., Altarejos, J. and De Luca, V. (2013) Virus-induced gene silencing identifies *Catharanthus roseus* 7-deoxyoligonic acid-7-hydroxylase, a step in iridoid and monoterpenoid indole alkaloid biosynthesis. *Plant J.* 76: 754–765.

Sayama, T., Ono, E., Takagi, K., Takada, Y., Horikawa, M., Nakamoto, Y., et al. (2012) The Cp-g1 glycosyltransferase locus regulates structural diversity of triterpenoid saponins of soybean. *Plant Cell* 24: 2123–2138.

Seki, H., Ohyama, K., Sawai, S., Mizutani, M., Ohnishi, T., Sudo, H., et al. (2008) Licorice β-amyrin 11-oxidase, a cytochrome P450 with a key role in the biosynthesis of the triterpene sweeter glycyrrhizin. *Proc. Natl Acad. Sci. USA* 105: 14204–14209.

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

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

Shibuya, M., Nishimura, K., Yasuyama, N. and Ebizuka, Y. (2010) Identification and characterization of glycosyltransferases involved in the biosynthesis of soyasaponin I in *Glycine max*. *FEBS Lett.* 584: 2258–2264.

Sun, S.X., Li, Y.M., Fang, W.R., Cheng, P., Liu, L. and Li, F. (2010) Effect and mechanism of AR-6 in experimental rheumatoid arthritis. *Clin. Exp. Med.* 10: 113–121.

Takada, Y., Sasama, H., Sayama, T., Kikuchi, A., Kato, S., Ishimoto, M., et al. (2013) Genetic and chemical analysis of a key biosynthetic step for soyasapogenol A, an aglycone of group A saponins that influence soy milk flavor. *Theor. Appl. Genet.* 126: 721–731.

Thimmappa, R., Geisler, K., Louveau, T., O’Maille, P. and Osbourn, A. (2014) Triterpenes biosynthesis in plants. *Annu. Rev. Plant Biol.* 65: 225–257.
Wang, Z., Yeats, T., Han, H. and Jetter, R. (2010) Cloning and characterization of oxidosqualene cyclases from Kalanchoe daigremontiana: enzymes catalyzing up to 10 rearrangement steps yielding friedelin and other triterpenoids. J. Biol. Chem. 285: 29703–29712.

Weitzel, C. and Simonsen, H.T. (2013) Cytochrome P450-enzymes involved in the biosynthesis of mono- and sesquiterpenes. Phytochem. Rev. 1–18.

Xiao, M., Zhang, Y., Chen, X., Lee, E.-J., Barber, C.J.S., Chakrabarty, R., et al. (2013) Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. J. Biotechnol. 166: 122–134.

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

Yan, X., Fan, Y., Wei, W., Wang, P., Liu, Q., Wei, Y., et al. (2014) Production of bioactive ginsenoside compound K in metabolically engineered yeast. Cell Res. 24: 770–773.

Yonekura-Sakakibara, K. and Hanada, K. (2011) An evolutionary view of functional diversity in family 1 glycosyltransferases. Plant J. 66: 182–193.

Zhang, N., Han, Z., Sun, G., Hoffman, A., Wilson, I.W., Yang, Y., et al. (2014) Molecular cloning and characterization of a cytochrome P450 taxoid 9z-hydroxylase in Ginkgo biloba cells. Biochem. Biophys. Res. Commun. 443: 938–943.

Zhao, Y.L., Cai, G.M., Hong, X., Shan, L.M. and Xiao, X.H. (2008) Anti-hepatitis B virus activities of triterpenoid saponins compound from Potentilla anserine L. Phytomedicine 15: 253–258.