

Regular Article

Cloning, Yeast Expression, and Characterization of a β -Amyrin C-28 Oxidase (CYP716A249) Involved in Triterpenoid Biosynthesis in *Polygala tenuifolia*

Fu-Sheng Zhang,^{*,a} Xuan Zhang,^{a,b} Qian-Yu Wang,^{a,b} Ya-Jie Pu,^{a,b} Chen-Hui Du,^c Xue-Mei Qin,^a Hong-Ling Tian,^d Yun-Lan Lian,^e Min-Sheng Li,^e Yu Chen,^e and Cun-Gen Ma^{*,c}

^aModern Research Center for Traditional Chinese Medicine, Shanxi University; Taiyuan, Shanxi 030006, China:

^bCollege of Chemistry and Chemical Engineering, Shanxi University; Taiyuan, Shanxi 030006, China: ^cShanxi University of Chinese Medicine; Shanxi 030619, China: ^dResearch Institute of Economics Crop Shanxi Academy of Agriculture Science; Fenyang 032200, China: and ^eShanxi Institute for Food and Drug Control; Taiyuan, Shanxi 030001, China.

Received April 13, 2020; accepted August 31, 2020

Polygala tenuifolia Willd. is a traditional Chinese herbal medicine that is widely used in treating nervous system disorders. Triterpene saponins in *P. tenuifolia* (polygala saponins) have excellent biological activity. As a precursor for the synthesis of presenegin, oleanolic acid (OA) plays an important role in the biosynthesis of polygala saponins. However, the mechanism behind the biosynthesis of polygala saponins remains to be elucidated. In this study, we found that CYP716A249 (GenBank: ASB17946) oxidized the C-28 position of β -amyrin to produce OA. Using quantitative real-time PCR, we observed that CYP716A249 had the highest expression in the roots of 2-year-old *P. tenuifolia*, which provided a basis for the selection of samples for gene cloning. To identify the function of CYP716A249, the strain R-BE-20 was constructed by expressing β -amyrin synthase in yeast. Then, CYP716A249 was co-expressed with β -amyrin synthase to construct the strain R-BPE-20 by using the lithium acetate method. Finally, we detected β -amyrin and OA by ultra-HPLC-Q Exactive hybrid quadrupole-Orbitrap high-resolution accurate mass spectrometry and GC-MS. The results of this study provide insights into the biosynthesis pathway of polygala saponins.

Key words *Polygala tenuifolia*; C-28 oxidase; CYP monooxygenase; yeast; biosynthesis; triterpene saponin

INTRODUCTION

Polygala tenuifolia Willd. is an important traditional medicine in China. It has many functions, such as calming the nerves, improving intelligence, and eliminating phlegm and swelling.¹⁾ *P. tenuifolia* is the third most frequently used prescription in treating nervous system disorders.²⁾ As one of the main active components of *P. tenuifolia*, polygala saponins have pharmacological effects, including anti-dementia, memory enhancement, and brain protection.^{3,4)} Polygala saponins are oleanane-type pentacyclic triterpenes; thus, they have a complex structure. Polygala saponins are difficult to obtain through traditional chemical methods and plant tissue culture. In recent years, the development of synthetic biology has provided a new way of acquiring polygala saponins. However, the biosynthetic pathway of polygala saponins has not been elucidated to date. Hence, discovering and identifying the functions of crucial enzymes have become the key steps to clarifying the biosynthesis pathway of polygala saponins.

A previous study found that the content of tenuifolin was higher in 2- to 3-year-old *P. tenuifolia* than in 1-year-old plants.⁵⁾ Moreover, the content of sapogenins in the head of roots is much lower than that in the roots of *P. tenuifolia*.⁶⁾ In view of the above phenomena, we investigated the key genes in the biosynthesis pathway of polygala saponins. First, we found that the expression levels of *SQS*, *SQE*, and *β -AS* were highly correlated with polygala saponins by digital gene expression (DGE) profiling.⁷⁾ We also found that the mevalonate (MVA) pathway has important functions in the biosynthesis of

triterpene saponins of *P. tenuifolia*. The precursors of *P. tenuifolia*, namely, isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP), were synthesized via the MVA pathway (Fig. 1). *β -AS* was isolated from *P. tenuifolia* and expressed in *Saccharomyces cerevisiae* to obtain β -amyrin.⁸⁾

In general, the biosynthesis of polygala saponins in plants can be divided into three steps^{9–11)} (Fig. 1). The first two steps have been identified, but oleanolic acid synthase (OAS) has not yet been elucidated in *P. tenuifolia* (Fig. 1). It is well known that CYPs play an important role in the formation of triterpenes. So far, 271 enzymes involved in triterpene biosynthesis have been found in 70 different plants.¹²⁾ Among them, clan CYP85 (members: CYP 87D, CYP88D, CYP88L, CYP708A, CYP716A, CYP716C, CYP716E, CYP716S, CYP716U, CYP716Y) have been found to be related to the structural modification of triterpenes.¹³⁾ A total of 22 CYP716 family enzymes were isolated from *Centella asiatica*, *Platycodon grandiflorum*, and *Aquilegia coerulea*. Besides, CYP716 family enzymes mainly oxidize C-2 α , C-6 β , C-12/13 α , C-16 β , and C-28 of β -amyrin after functional verification.¹⁴⁾ Hence, whether the CYP716 enzyme in *P. tenuifolia* oxidizes β -amyrin remains to be elucidated. Moreover, the discovery of this key enzyme will be a crucial step to clarify the biosynthesis pathway of polygala triterpene saponins. In our previous research, 92 P450s were found after DGE profiling.¹⁵⁾ Furthermore, we found 15 P450s that are related to polygala saponin accumulation after co-expression analysis, including 1 sequence belonging to the CYP716 family, through the analysis of full-length transcriptome data (unpublished data). Aside from identifying

*To whom correspondence should be addressed. e-mail: ample1007@sxu.edu.cn; macungen2001@163.com

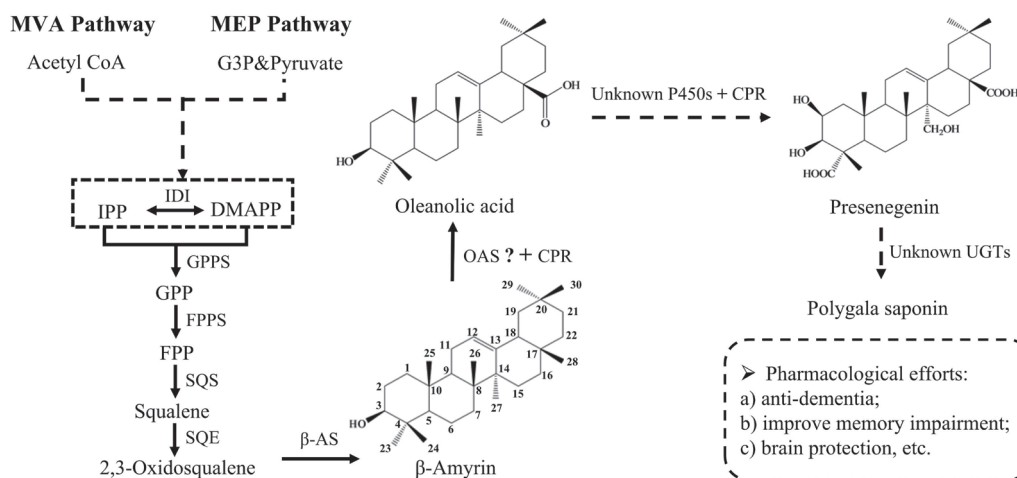


Fig. 1. Biosynthetic Pathway of Polygala Saponins

One-step catalytic reactions are indicated by solid arrows, and multi-step catalytic reactions are indicated by dashed arrows. G3P: glyceraldehyde 3-phosphate; MEP: methyl-D-erythritol phosphate; MVA: mevalonate; IPP: isopentenyl pyrophosphate; DMAPP: dimethylallyl diphosphate; IDI: isopentenyl diphosphate isomerase; GPPS: geranyl diphosphate synthase; GPP: geranyl pyrophosphate; FPP: farnesyl pyrophosphate; FPPS: farnesyl diphosphate synthase; SQS: squalene synthase; SQE: squalene epoxidase; β -AS: β -amyrin synthase; OAS: oleanolic acid synthase; CPR: CYP reductase; P450s: CYP monooxygenases; UGTs: UDP-dependent glycosyltransferases.

the enzyme, understanding its function is also important.

As one of the yeast species, *S. cerevisiae* was sequenced in 1996.¹⁶⁾ It is often used in genetic engineering due to its fast growth rate and suitability for large-scale culture. Moreover, it can provide natural precursor 2,3-oxidosqualene for the biosynthesis of triterpene saponins, which facilitates the heterologous biosynthesis of triterpene saponins by recombinant yeasts. At present, *S. cerevisiae* is often used as a cellular chassis in the heterogeneous synthesis of natural compounds.^{17–19)} In *Panax ginseng*, CYP716A47 was identified as a protopanaxadiol synthase by introducing CYP716A47-expression vector constructed using restriction enzymes into recombinant WAT21 yeasts.²⁰⁾ CYP716A52v2 and CYP716A244 were both verified in *S. cerevisiae* using the same method.^{21,22)} Based on relevant studies, we constructed strains that produce β -amyrin and OA.

In the present study, a unigene, which may be OAS, corresponded to CYP716A249. Thus, the synthetic pathway of β -amyrin was constructed in *S. cerevisiae* by introducing *Glycyrrhiza glabra* β -AS. *ERG20* (farnesyl diphosphate synthase) was overexpressed to increase precursor supply, and the synthetic pathway of OA was constructed in the strain R-BE-20 to verify the function of CYP716A249. Then, we analyzed the mRNA expression levels of CYP716A249 in different tissues at different growth stages of *P. tenuifolia*. Finally, the product of β -amyrin and OA was detected by ultra-HPLC-Q Exactive hybrid quadrupole-Orbitrap high-resolution accurate mass spectrometry (UHPLC-Q-Orbitrap HRMS) and GC-MS. The function of C-28 oxidase in *P. tenuifolia* was first predicted and verified. The results of this study may provide a foundation for the discovery of P450s in the biosynthesis pathway of polygala saponins.

MATERIALS AND METHODS

Plant Samples Fresh samples of 1- to 3-year-old cultivated *P. tenuifolia*, including roots, stems, leaves, and flowers, were collected from Xinjiang County (Shanxi Province, China). In the process of collection, three sites were selected as the three biological repeats for quantitative real-time PCR

(qRT-PCR) in each growth period. One to two plants were collected at each site. They were then wrapped with aluminum foil and immediately placed in liquid nitrogen. Then, the samples were stored at -80°C for preservation. All samples were identified as *P. tenuifolia* by Professor Xue-mei Qin and preserved in the Modern Research Center for Traditional Chinese Medicine, Shanxi University (China).

Phylogenetic Analysis The amino acid sequences of CYP716s from over 15 plant species (Table S1) were collected from NCBI (<http://www.ncbi.nlm.nih.gov>). Sequence alignment was performed by Clustal W analysis. Evolutionary distances were computed using the overall mean distance of MEGA5.0. The phylogenetic tree was built by neighbor-joining methods with bootstrap values for 1000 replicates.²³⁾

Total RNA Extraction and qRT-PCR Fresh tissues of *P. tenuifolia* and fresh strains of R-EBP-20 were collected and frozen in liquid nitrogen. The purity and concentration were measured via 1.5% agarose gel electrophoresis and by using a NanoDrop2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, U.S.A.). cDNA was amplified from the total RNA (1000ng) by qRT-PCR. The reaction system was 20 μL , including 10 μL of SYBR Premix Ex Taq II 2 $\times$ (TaKaRa Biomedical Technology (Beijing) Co., Ltd., China), 0.8 μL of Prime F (10 μM), 0.8 μL of Prime R (10 μM), 2 μL of cDNA, and 6.4 μL of ddH₂O. All reactions were carried out in 96-well plates using the Heal Force Real-Time PCR System (Heal Force CG-05, Hangzhou Jingle Scientific Instrument Co., Ltd., China). The cycling conditions were based on the manufacturer's recommendations (95 $^{\circ}\text{C}$ for 60s, followed by 40 cycles of 15s at 95 $^{\circ}\text{C}$ for DNA denaturation, 15s at 60 $^{\circ}\text{C}$ for primer annealing, and 45s at 72 $^{\circ}\text{C}$ for extension). Each plate contained three technical replicates. The gene expression was calculated using the $2^{-\Delta\Delta\text{Ct}}$ method. Primers of qRT-PCR are shown in Table S2.

Strains and Medium *S. cerevisiae* strain W303a (genotype: *MATa* {*leu2-3,112*; *ura3-1*; *trp1-1*; *can1-100*; *his3-11,15*; *ade2-1*}) was obtained from the American Type Culture Collection (<http://www.atcc.org/>) and used as the parent strain for all yeast strains. All strains were grown in defective yeast extract peptone dextrose medium, which consisted of 10g/L of

yeast extract, 10 g/L of peptone, and 10 g/L of dextrose. The strain W303a lacked leucine, uracil, tryptophan histidine, and adenine, whereas the R-BE-20 and R-BPE-20 strains lacked tryptophan. All strains used in this study are listed in Table S3. *Escherichia coli* DH5 α (TIANGEN Biochemical Technology Co., Ltd., Beijing, China) was used for transformation. The strains were cultivated at 37°C in Luria–Bertani (LB) medium with 10 g/L of peptone, 5 g/L of yeast extract, and 10 g/L of NaCl supplemented with ampicillin (100 μ g/mL). The materials required for the medium were provided by Beijing Solarbio Science & Technology Co., Ltd. (China).

Plasmid and Strain Construction PrimeSTAR HS Polymerase and T4 DNA Ligase were purchased from TaKaRa Biomedical Technology Co., Ltd. Restriction enzymes were purchased from New England Biolabs (Beijing), Ltd. (China). DNA gel purification and plasmid extraction kits were purchased from TIANGEN Biotech (Beijing) Co., Ltd. (China). Amino acids, yeast nitrogen base medium, D-sorbitol, and dextrose were supplied by Beijing Solarbio Science & Technology Co., Ltd. (China). All primers were synthesized by GenScript Biotechnology Co., Ltd. (Nanjing, China).

R-BE-20: β -AS was PCR-amplified from the cDNA of *G. glabra* L. Promoter TYrosyl-tRNA Synthetase (TYS1p), promoter Triose-phosphate DeHydrogenase (TDH3p), terminator TYrosyl-tRNA Synthetase (TYS1t), and terminator CYtochrome C (CYC1t) were amplified from strain W303a DNA.

R-BPE-20: OAS was PCR-amplified from the cDNA of *P. tenuifolia*. CPR (CYP reductase) was commissioned by GenScript Biotechnology Co., Ltd. for codon optimization and fully synthesized. Promoter Alcohol DeHydrogenase (ADH1p), promoter ALAnyI-tRNA synthetase (ALA1p), terminator Alcohol DeHydrogenase (ADH1t), and terminator ALAnyI-tRNA synthetase (ALA1t) were amplified from *S. cerevisiae* W303a DNA. All genetic elements were linked in turn by overlap extension PCR (OE-PCR) to obtain gene expression cassettes with restriction sites at both ends.²⁴ These plasmids and gene expression cassettes were digested with two restriction endonucleases (the primers used are shown in Table S4). Then, validated plasmids were transformed into yeast by the lithium acetate method.

Metabolite Squalene, Ergosterol, and β -Amyrin Extraction for GC-MS Sample preparation: The cells were collect-

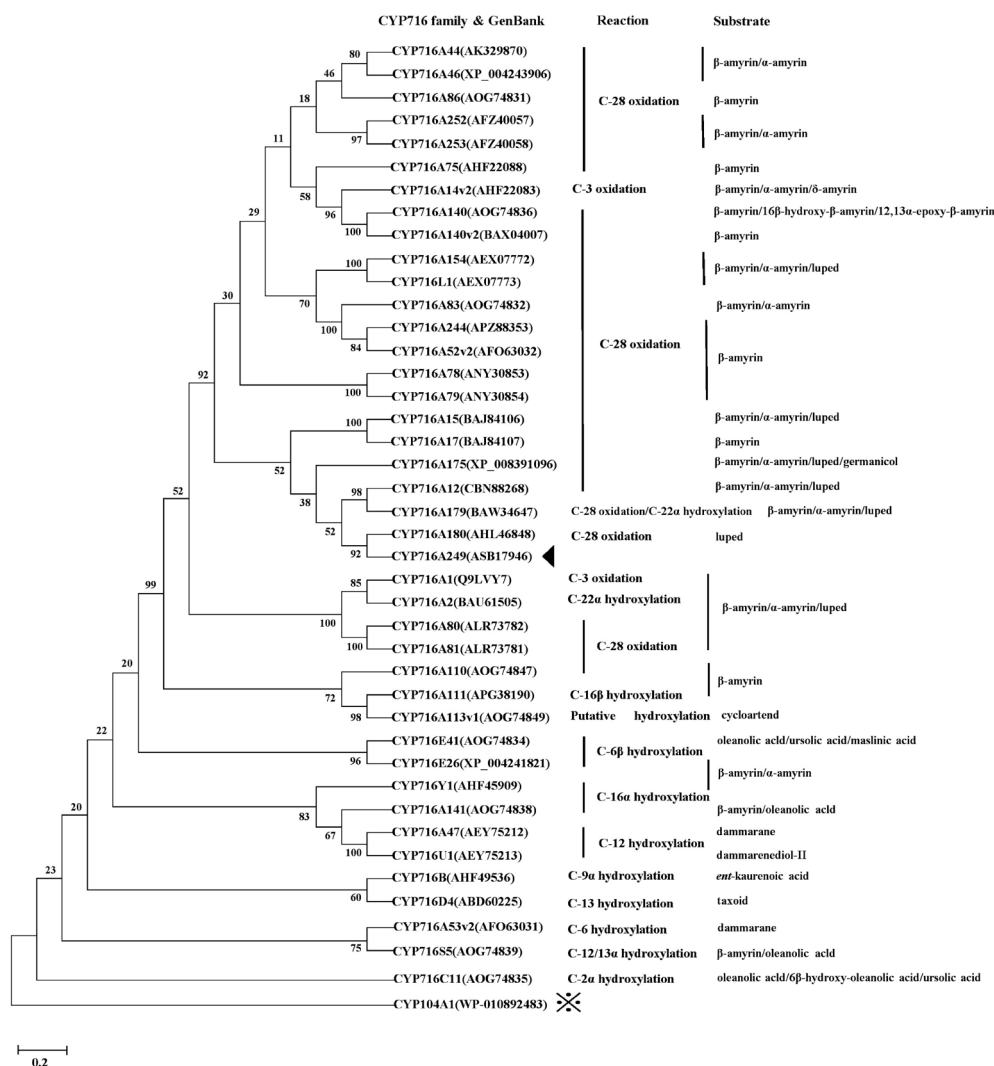


Fig. 2. Phylogenetic Analysis of CYP716s

Phylogenetic Tree of the deduced amino acid sequences including CYP716s and CYP104A1. The Neighbor–Joining bootstrapped phylogenetic tree was built using MEGA5.0 software. Bootstrap analysis values with 1000 replicates are shown at the nodal branches. The indicated scale represents 0.1 amino acid substitutions per site. Each amino acid sequence was provided with the corresponding reaction and substrate. * represents *Agrobacterium tumefaciens* CYP104A1, which was included as the out group. ► represents *P. tenuifolia* CYP716A249.

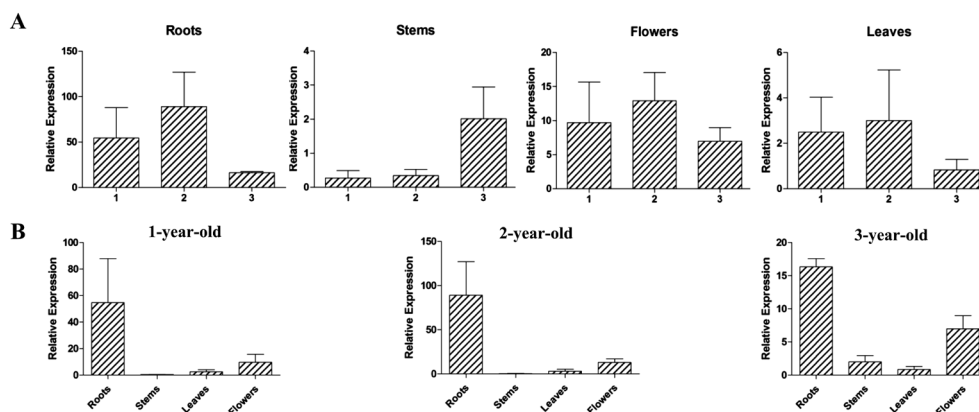


Fig. 3. Expression of *CYP716A249* in *P. tenuifolia* (A/B: Expression Levels of *CYP716A249* at Different Growth Periods in Roots, Stems, Flowers, and Leaves)

Note: the error bar indicates the S.D. calculated from the technical replicates.

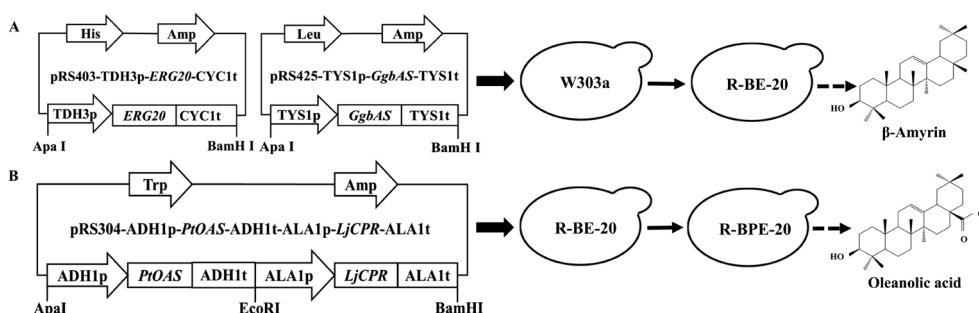


Fig. 4. Strains in the Study (A: R-BE-20; B: R-BPE-20; $\Rightarrow$ Represents LiAc Transformation; $\rightarrow$ Represents Successful Construction of New Strain; $\dashrightarrow$ Represents Shaking Fermentation)

ed by centrifuging yeast culture (40 mL) for 5 min at 5000 rpm and were resuspended from yeast culture media with equal volume of 20% (w/v) KOH and 50% (v/v) ethanol (Damao Chemical Reagent Factory, China) by boiling for 10 min. When the temperature dropped to room temperature, 10 mL of hexane (Damao Chemical Reagent Factory) was added to extract twice. Then, the hexane phase was merged and blown dry with nitrogen. The extracts were trimethylsilylated with pyridine and *N,O*-bis(trimethylsilyl)-trifluoroacetamide (vol: 1:2) at 85 °C for 1 h.

Chromatographic conditions: GC-MS analysis was performed using the GC model 6890 and MS model 5975B (Agilent Technologies, Inc., China). The aliquot (2 μ L) was injected (splitless mode) into a ZB-5 capillary column (Phenomenex, 30.0 m $\times$ 250 $\times$ 0.25 μ m). The carrier gas was He with a flow rate of 1 mL/min. The injector temperature was 280 °C, and the oven temperature was held at 80 °C for 1 min. Then, the temperature was ramped to 280 °C at 20 °C/min and held at 280 °C for 45 min. Finally, the temperature was ramped to 320 °C at 20 °C/min and held at 320 °C for 1 min. The temperatures of the MS transfer line, MS ion source, and quadrupole were 250, 230, and 150 °C, respectively.

Mass conditions: For the identification of metabolites, full electron ionization-MS spectra were generated by scanning the *m/z* range of 60–800 with a solvent delay of 2.0 min.^{25,26)}

Metabolite OA Extraction for UHPLC-Q-Orbitrap HRMS Sample preparation: All yeast cells were collected and ground into a powder with liquid nitrogen. The powder was extracted with methanol/acetone at a ratio of 1:1 (Damao

Chemical Reagent Factory) for 30 min, and the process was repeated twice. The supernatants were combined and filtered through 0.22 μ m microporous membrane.

Chromatographic conditions: The chromatography was performed on a Thermo Fisher U3000 UHPLC system (Thermo Fisher Scientific). For chromatographic separation, a Kinetex 2.6U C18 column (2.1 $\times$ 100 mm) was used. The mobile phase consisted of 0.1% formic acid and 10% methanol (A) and acetonitrile (B) (Damao Chemical Reagent Factory), and the chromatography was performed as follows: A:B = 15:85 for 15 min. The solvent flow rate was 0.2 mL/min, and the column temperature was 30 °C.

Mass conditions: Mass spectrometry analysis was performed on a Thermo Scientific™ Q Exactive hybrid quadrupole-Orbitrap high-resolution accurate mass spectrometer (Thermo Fisher Scientific) equipped with heated electrospray ionization source in both positive and negative ion modes. The MS conditions were as follows: ionspray voltage was 3.5 kV (+) and 2.5 kV (–), auxiliary volume was 10, capillary temperature was 320 °C, and mass resolution was 70000. All spectra were obtained at an *m/z* range of 100–1000.^{27,28)}

RESULTS

Phylogenetic Analysis of CYP716s According to the comparison between the DGE data of different phenological phases²¹⁾ and previous transcriptome data,^{19,29)} we found that a unigene may be OAS in the P450s after further analyzing the expression of P450s and transcriptome data of *P. tenuifolia*.

After sequence alignment and functional annotation on NCBI, the unigene was found to correspond to *CYP716A249* of *P. tenuifolia*.

The Neighbor-Joining bootstrapped phylogenetic tree of *CYP716A249* and 40 other *CYP716s* (Table S1), whose functions were already known, was constructed. The enzymes of the *CYP716* family have different functions, such as C-3-oxidase, C-28-oxidase, C-16 α / β -hydroxylase, C-13-hydroxylase, and C-12-hydroxylase (Fig. 2). Among them, 24 enzymes in the

CYP716A family play the role of C-28 oxidase. *CYP716A249* of *P. tenuifolia* was clustered into the group of *CYP716* family enzymes with another three P450s, whose functions are C-28 oxidase (*Betula platyphylla* *CYP716A180*, *Glycyrrhiza uralensis* *CYP716A179*, *Medicago truncatula* *CYP716A12*). *CYP716A249* was far away from C-3-oxidase (*Artemisia annua* *CYP716A14v2*), C-6 β -hydroxylase (*Solanum lycopersicum* *CYP716E26*, *C. asiatica* *CYP716E26*), and C-2 α -hydroxylase (*P. ginseng* *CYP716A53v2*). Thus, *CYP716A249*

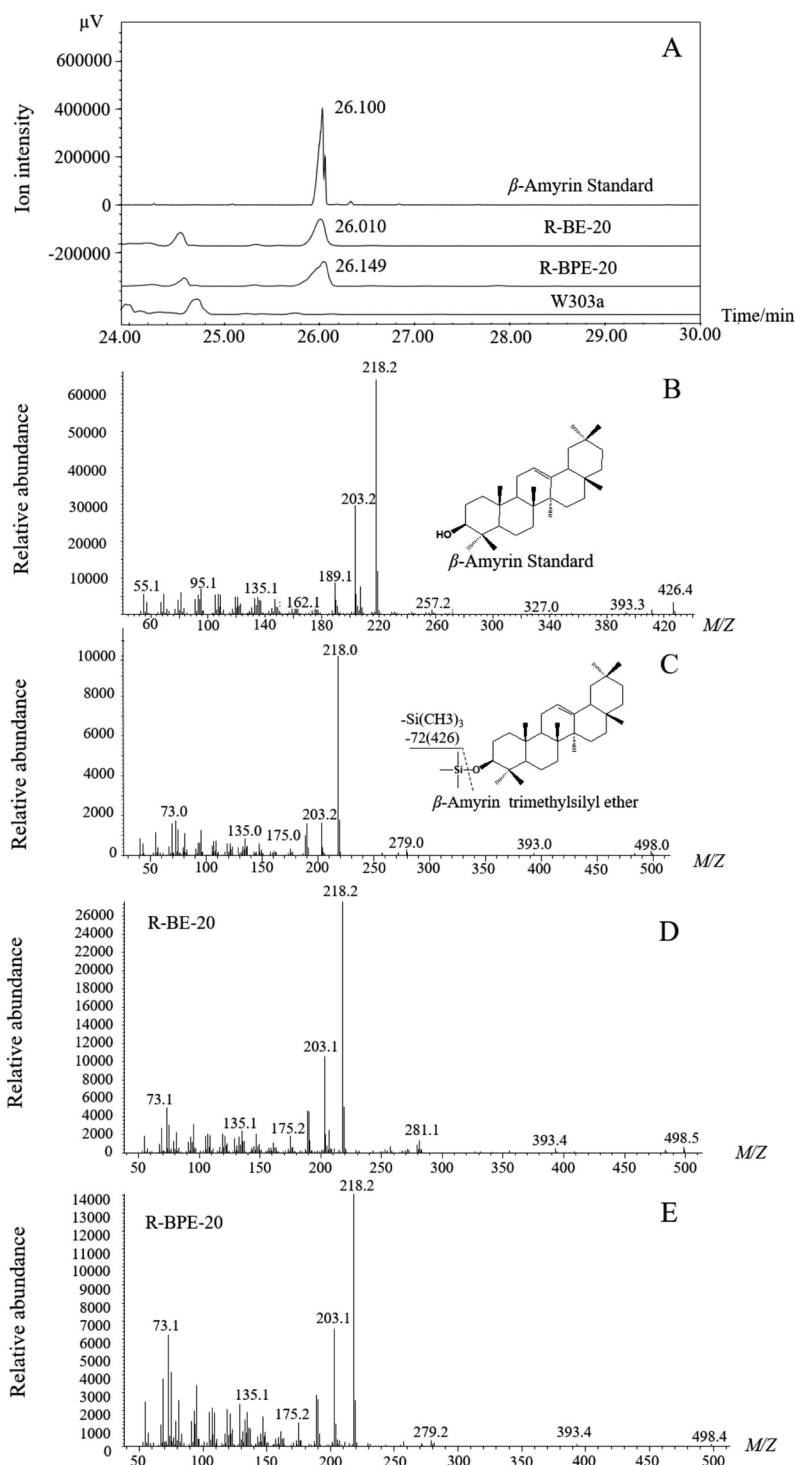


Fig. 5. GC-MS Analysis of β -Amyrin Production in R-BE-20 and R-BPE-20

A: Total ion chromatogram of β -amyrin standard, R-BE-20, and R-BPE-20; B: Mass spectrometry spectra of the β -amyrin standard at 26.100 min; C: Mass spectrometry spectra of β -amyrin trimethylsilyl ether; D: Mass spectrometry spectra of R-BE-20 at 26.010 min; E: Mass spectrometry spectra of R-BPE-20 at 26.149 min.

was suspected to oxidize the C-28 position of β -amyrin to form OA in the biosynthesis pathway of polygala saponins.

Space-Time Expression of *CYP716A249* in *P. tenuifolia* RNA was extracted from 1–3-year-old *P. tenuifolia* with different tissues (roots, stems, leaves, and flowers). The extraction quality was detected by agarose gel electrophoresis (Fig. S1). The 28S and 18S bands were clear and bright, indicating that RNA had great integrity and no degradation. Using the Nanodrop2000 spectrophotometer, the result showed that the OD_{260}/OD_{280} of RNA was approximately 1.90, which indicated no protein and small salt impurity pollution. Thus, the RNA was of great quality and could be used for subsequent analysis.

The mRNA expression levels of *CYP716A249* in different tissues of *P. tenuifolia* were detected by qRT-PCR with *Cdc42* as the reference gene. From the fusion curves, *Cdc42* and *CYP716A249* were single peaks (Fig. S2), indicating that primers of qRT-PCR did not show obvious non-specific amplification and all results were reliable. These results showed that the expression trends of *CYP716A249* in the roots, leaves, and flowers were the same. In other words, the expression of *CYP716A249* in 2-year-old *P. tenuifolia* was increased, but the expression of *CYP716A249* in 3-year-old *P. tenuifolia* was decreased. The expression of *CYP716A249* in stems increased with the growth period (Fig. 3A). The expression trend of *CYP716A249* in 1-year-old *P. tenuifolia* was the same as that in 2-year-old *P. tenuifolia*, which was the highest in roots and the lowest in stems. The expressions of *CYP716A249* in all tissues were ordered as follows: roots > flowers > leaves > stems. *CYP716A249* of 3-year-old *P. tenuifolia* was highly expressed in roots but the least expressed in leaves. The expression was ordered as follows: roots > flowers > stems > leaves (Fig. 3B).

Construction of Plasmid and Yeast Strains R-BE-20: To obtain β -amyrin as the substrate of OA in yeast, β -AS (*GgbAS*, GenBank: AB037203.1) of *G. glabra* as the exogenous gene and *ERG20* as the endogenous regulatory gene were selected to conduct overexpression plasmids. Then, the expression plasmids of pRS403-TDH3p-*ERG20*-CYC1t and pRS425-TYS1p-*GgbAS*-TYS1t were constructed (Figs. S3A, C). All gene elements in these two plasmids were amplified by OE-PCR and verified by agarose gel electrophoresis (Figs. S3B, D). Finally, we transferred them into strain W303a to obtain strain R-BE-20 (Fig. 4A).

R-BPE-20: CPR is the key limiting enzyme in the redox process, which delivers the electrons of reduced nicotinamide adenine dinucleotide phosphate (NADPH) to P450s in order to make them produce catalytic activity. Therefore, *LjCPR* (*Lotus japonicus*, GenBank: AB433810) and *CYP716A249* were used for the plasmid of pRS304-ADH1p-*PtOAS*-ADH1t-*ALA1p*-*LjCPR*-*ALA1t* (Fig. S2E). These gene elements were amplified and verified by the same methods (Fig. S3F). Thereafter, the strain R-BPE-20 was constructed successfully (Fig. 4B).

In addition, qRT-PCR was used to detect whether the plasmid of β -AS and OAS were transferred into yeast successfully. As can be seen from Figs. S4 and S5, β -AS and OAS were both successfully expressed in the strain R-BPE-20.

Detection of β -Amyrin by GC-MS GC-MS was used to detect β -amyrin of strains R-BE-20 and R-BPE-20. The total ion chromatogram revealed that strain R-BE-20 has a clear peak at a retention time of 26.010 min, and strain R-BPE-20 has a peak at a retention time of 26.149 min, but strain W303a was not detected (Fig. 5A). Then, two signals at 26.010 and

26.149 min were further analyzed using GC-MS. The mass spectrometry fragmentation patterns of the two signals at 26.010 and 26.149 min were identical to β -amyrin trimethylsilyl ether (Figs. 5D, E), indicating that β -amyrin trimethylsilyl ether was detected in strain R-BE-20 and R-BPE-20. Trimethylsilyl ($m/z = 72$) was added on the structural formula of β -amyrin ($m/z = 426$) after silylation, and then β -amyrin became β -amyrin trimethylsilyl ether ($m/z = 498.0$) (Figs. 5B, C). Subsequently, we could conclude that β -amyrin was produced in strain R-BE-20 and R-BPE-20, but it was transformed into

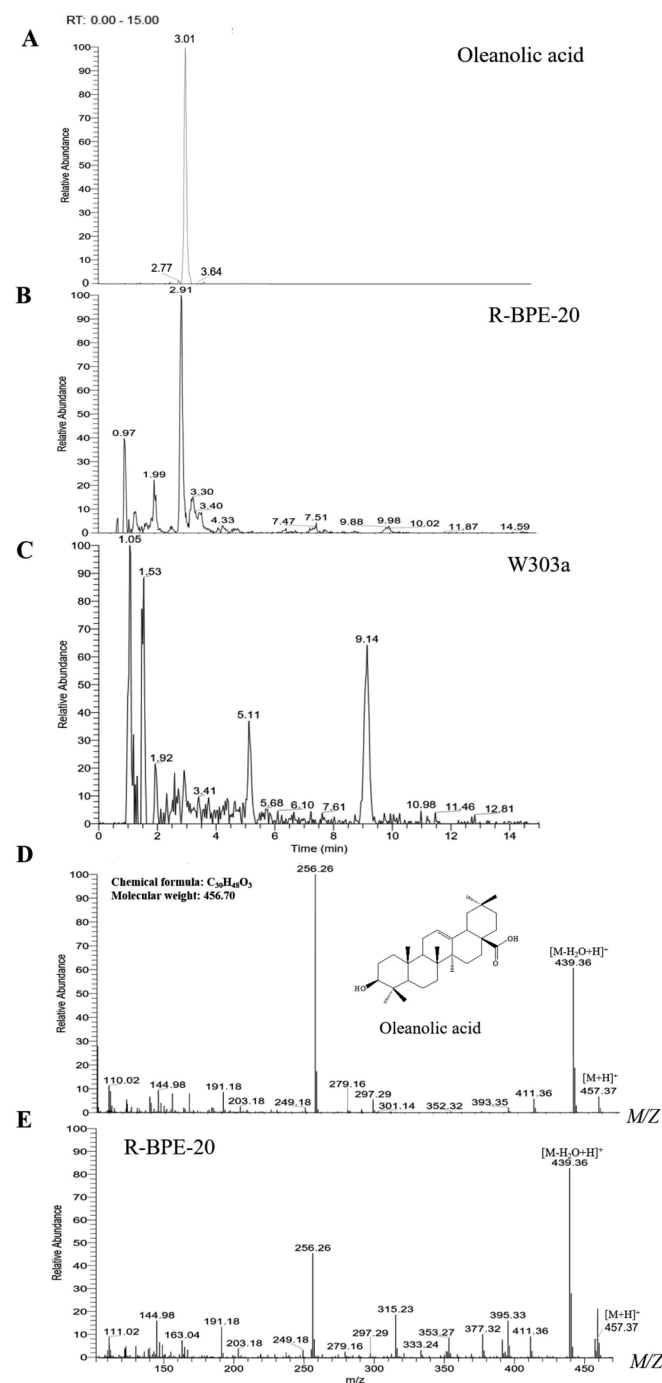


Fig. 6. Extracted Ion Chromatogram and Mass Spectra by UHPLC-Q-Orbitrap HRMS in Positive Ion Mode

A: Extracted ion chromatogram of OA; B: Extracted ion chromatogram of R-BPE-20; C: Extracted ion chromatogram of W303a; D: Mass spectra of OA; E: Mass spectra of R-BPE-20.

β -amyrin trimethylsilyl ether due to a long derivatization time.

Detection of OA by UHPLC-Q-Orbitrap HRMS OA's variance in *S. cerevisiae* could be analyzed by the UHPLC system, and accurate mass measurement was provided through Q-Orbitrap HRMS. By comparing the peak of OA in the positive and negative ion mode, we found that the peak of OA was better in the positive ion mode. In the positive ion mode, the retention time of the standard was 3.01 min, and that of the strain R-BPE-20 was 2.91 min (Figs. 6A, B). We found that the retention time of strain R-BPE-20 was almost identical to that of the OA standard. Then, the corresponding mass spectra of strain R-BPE-20 were compared with those of the OA standard. The results showed that the fragment ion diagrams of strain R-BPE-20 were consistent with the fragment information of OA standard, with $m/z = 439.36 [M-H_2O + H]^+$, $457.37 [M + H]^+$. Thus, we could conclude that OA was produced in the strain R-BPE-20.

DISCUSSION

β -Amyrin is not only an important pentacyclic triterpenoid but also an important precursor in OA biosynthesis.³⁰⁾ In this study, phylogenetic analysis was used to predict the C-28 oxidative activity of CYP716A249, and the final product OA was obtained by co-expressing β -AS and CYP716A249 in yeast. We determined that CYP716A249 catalyzed β -amyrin into OA.

Previous studies reported that CYP716s evolved specifically toward triterpenoid biosynthesis by kingdom-wide phylogenetic analysis among more than 400 CYP716s from over 200 plants.^{14,31)} Most CYP716s have been characterized as C-28 oxidases, which oxidize α -amyrin and β -amyrin into triterpenoids ursolic acid and OA.³²⁾ Recently, CYP716C, CYP716E, CYP716S, and CYP716Y subfamily members were found to catalyze the oxidation of other triterpenoids from different families (higher eudicots), including *C. asiatica* CYP716E41 (C6 β -hydroxylation),³²⁾ *C. asiatica* CYP716C11 (C2 α /6 β -hydroxylation),¹²⁾ and *Bupleurum falcatum* CYP716Y1 (C16 α -hydroxylation).³³⁾

A total of 42 P450s can mediate the catalytic reaction of β -amyrin, belonging to 26 different species. They can catalyze C-3, C-6, C-11, C-12/13, C-16, C-22, C-24, C-28, and C-30 on the skeleton of β -amyrin and then generate various intermediate products such as OA and β -amyrin derivatives.^{12,34)} Among them, 21 P450s can directly oxidize β -amyrin to OA, which belong to C-28 oxidase. For instance, CYP716A12 has been proved to be the three-step oxidation at the C-28 position of β -amyrin in the triterpene saponin biosynthesis of *M. truncatula*.³⁴⁾

At present, genome or transcriptome heterologous reconstruction is the main method to analyze the biosynthesis pathway of a natural plant.³⁵⁾ Transcriptome sequencing technology and bioinformatics have been widely used in recent years. In 2018, Kim's group in Korea uploaded 49 P450s sequences of *P. tenuifolia* to the GenBank database.³⁰⁾ To obtain more full-length P450 sequences, our group also completed Iso-Seq high-throughput transcriptome sequencing (PacBio platform, 20G clean data) with root, stem, and leaf samples (mixed) of cultivated *P. tenuifolia* at different growth periods (1, 2, and 3 years). We found that transcriptome data have a great significance for further screening candidate P450s, and phylogenetic analysis can be used to predict the functions of candidate genes. However, CYP716A249 was only verified by yeast in

this study. Functional verification of genes mainly includes *Escherichia coli*, yeast *in vivo/vitro*, hairy root, and genetic transformation system. Thus, the candidate genes could be verified from multiple levels in the future.

Studies on the biosynthesis pathway of polygala saponins are few, and the candidate genes of P450s related to oxidation state formation at C-2, C-23, and C-27 sites of presenegenin are still unknown. Accordingly, the first task for us is to explore the relevant candidate genes. Although transcriptome sequencing can provide more full-length genes, how to effectively screen potential functional candidate genes is still a major problem that needs to be solved.

Acknowledgments This work was supported by Shanxi Institute for Food and Drug Control and the Large Instrument Center of Shanxi University, funded by the Shanxi Province Key Projects for Key Research and Development Program (Grant number 201603D3111003), the Nature Science Young Foundation of Shanxi Province (Grant No. 201901D111039), the Nature Science Foundation of Shanxi Province and the Key Laboratory of Effective Substances Research and Utilization in TCM of Shanxi Province (Grant No. 201605D111004).

Conflict of Interest The authors declare no conflict of interest.

Supplementary Materials The online version of this article contains supplementary materials.

REFERENCES

- 1) Chinese Pharmacopoeia Commission. *Pharmacopoeia of the People's Republic of China* (10th ed.). China Medical Science Press, Beijing, Vol. 1, p. 156 (2015).
- 2) Lin ZH, Gu J, Xiu J, Mi TY, Dong J, Tiwari JK. Traditional Chinese medicine for senile dementia. *Evid. Based Complement. Alternat. Med.*, **2012**, 692621 (2012).
- 3) Mo W, Lin ZH. Study development on cognitive benefits of radix polygalae. *Wood Sci. Technol.*, **14**, 1913–1916 (2012).
- 4) Li CJ, Yang JZ, Yu SS, Chen NH, Xue W, Hu JF, Zhang DM. Triterpenoid saponins with neuroprotective effects from the roots of *Polygala tenuifolia* Willd. *Planta Med.*, **74**, 133–141 (2008).
- 5) Li J, Dong XB, Jiang Y, Dong TX, Tu PF, Zhan HQ. HPLC determination of total saponins in Radix Polygalae. *Chin. J. Pharm. Anal.*, **27**, 1329–1332 (2007).
- 6) Zhang FS, Chen TY, Wang DD, Yan Y, Tian HL, Qin XM, Ma CG. Research progress in correlation between commodity specification level and quality of *Polygala tenuifolia*. *Chin. Tradit. Herbal Drugs*, **48**, 2538–2547 (2017).
- 7) Zhang FS, Li XW, Li ZY, Xu XS, Peng B, Qin XM, Du GH. UPLC/Q-TOF MS-based metabolomics and qRT-PCR in enzyme gene screening with key role in triterpenoid saponin biosynthesis of *Polygala tenuifolia*. *PLOS ONE*, **9**, e105765 (2014).
- 8) Jin ML, Lee DY, Um Y, Lee JH, Park CG, Jetter R, Kim OT. Isolation and characterization of an oxidosqualene cyclase gene encoding a β -amyrin synthase involved in *Polygala tenuifolia* Willd. saponin biosynthesis. *Plant Cell Rep.*, **33**, 511–519 (2014).
- 9) Xu XS, Zhang FS, Qin XM. Research advances on triterpenoid saponins biosynthesis and its key enzymes. *Wood Sci. Technol.*, **11**, 2440–2448 (2014).
- 10) Chen L, Wu YS. Biosynthesis pathway and related enzymes of triterpenoid saponins. *Drugs & Clinic*, **19**, 156–161 (2004).
- 11) Zhao YS, Wan DG, Chen X, Li ZL, Tian HL. Advances in studies

- on biosynthesis and regulation of pentacyclic triterpenoid saponin. *Chin. Tradit. Herbal Drugs*, **40**, 327–330 (2009).
- 12) Miettinen K, Iñigo S, Kreft L, Pollier J, De Bo C, Botzki A, Coppen F, Bak S, Goossens A. TheTriForC database: a comprehensive up-to-date resource of plant triterpene biosynthesis. *Nucleic Acids Res.*, **46** (D1), D586–D594 (2018).
 - 13) Sumit G. Triterpene structural diversification by plant cytochrome P450 enzymes. *Front. Plant Sci.*, **8**, 1886 (2017).
 - 14) Miettinen K, Pollier J, Buyst D, Arendt P, Csuk R, Sommerwerk S, Moses T, Mertens J, Sonawane PD, Pauwels L, Aharoni A, Martins J, Nelson DR, Goossens A. The ancient CYP716 family is a major contributor to the diversification of eudicot triterpenoid biosynthesis. *Nat. Commun.*, **8**, 14153 (2017).
 - 15) Xu XS, Xue Y, Zhang FS, Qin XM, Peng B, Tian HL, Zeng ZP. Analysis of digital gene expression profiles of the cultivated *Polygala tenuifolia* in different phenological phases. *Yao Hsueh Hsueh Pao*, **51**, 1165–1174 (2016).
 - 16) Engel SR, Dietrich FS, Fisk DG, Binkley G, Balakrishnan R, Costanzo MC, Dwight SS, Hitz BC, Karra K, Nash RS, Weng S, Wong ED, Lloyd P, Skrzypek MS, Miyasato SR, Simison M, Cherry JM. The reference genome sequence of *Saccharomyces cerevisiae*: then and now. *G3* (Bethesda), **4**, 389–398 (2014).
 - 17) Sun MX, Zhou JW. Research advances of heterologous synthesis of monoterpenes in *Saccharomyces cerevisiae*. *Industrial Microbiol.*, **46**, 54–58 (2016).
 - 18) Zhou YJ, Gao W, Rong Q, Jin G, Chu H, Liu W, Yang W, Zhu Z, Li G, Zhu G, Huang L, Zhao ZK. Modular pathway engineering of diterpenoid synthases and the mevalonic acid pathway for miltiradiene production. *J. Am. Chem. Soc.*, **134**, 3234–3241 (2012).
 - 19) Xie WP, Liu M, Lv XM, Lu WQ, Gu JL, Yu HW. Construction of a controllable beta-carotene biosynthetic pathway by decentralized assembly strategy in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.*, **111**, 125–133 (2014).
 - 20) Han JY, Kim HJ, Kwon YS, Choi YE. The Cyt P450 enzyme CYP716A47 catalyzes the formation of protopanaxadiol from damarenediol-II during ginsenoside biosynthesis in *Panax ginseng*. *Plant Cell Physiol.*, **52**, 2062–2073 (2011).
 - 21) Han JY, Kim MJ, Ban YW, Hwang HS, Choi YE. The involvement of β -amyrin 28-oxidase (CYP716A52v2) in oleanane-type ginsenoside biosynthesis in *Panax ginseng*. *Plant Cell Physiol.*, **54**, 2034–2046 (2013).
 - 22) Jo HJ, Han JY, Hwang HS, Choi YE. β -Amyrin synthase (EsBAS) and β -amyrin 28-oxidase (CYP716A244) in oleanane-type triterpene saponin biosynthesis in *Eleutherococcus senticosus*. *Phytochemistry*, **135**, 53–63 (2017).
 - 23) Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.*, **28**, 2731–2739 (2011).
 - 24) Horton RM, Hunt HD, Ho SN, Pullen JK, Pease LR. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene*, **77**, 61–68 (1989).
 - 25) Kirby J, Romanini DW, Paradise EM, Keasling JD. Engineering triterpene production in *Saccharomyces cerevisiae*– β -amyrin synthase from *Artemisia annua*. *FEBS J.*, **275**, 1852–1859 (2008).
 - 26) Zhang GL, Cao Q, Liu JZ, Liu BY, Li J, Li C. Refactoring β -amyrin synthesis in *Saccharomyces cerevisiae*. *AIChE J.*, **61**, 3172–3179 (2015).
 - 27) Wang D, Wang BB, Liu Y, Shi MY, Wang DG, Huang LQ, Dai ZB, Zhang XL. Optimization of synthetic pathway and fermentation process of yeast cell factories for production of oleanolic acid. *Chung Kuo Chung Yao Tsa Chih*, **39**, 2640–2645 (2014).
 - 28) Wang BB, Shi MY, Wang D, Xu JY, Liu Y, Yang HJ, Dai ZB, Zhang XL. Production of carotene by metabolically engineered *Saccharomyces cerevisiae*. *Chin. J. Biotechnol.*, **30**, 1204–1216 (2014).
 - 29) Tian HL, Xu XS, Zhang FS, Wang QY, Guo SH, Qin XM, Du GH. Analysis of *Polygala tenuifolia* transcriptome and description of secondary metabolite biosynthetic pathways by illumina sequencing. *Int. J. Genomics*, **2015**, 782635 (2015).
 - 30) Zhang FS, Kong RR, Chen TY, Wang QY, Qin XM, Du CH, Ma CG. Advance in biosynthesis with catalysis by P450s of plant-derived oleanane type triterpenoids such as Polygala saponins. *Yao Hsueh Hsueh Pao*, **6**, 1000–1009 (2019).
 - 31) Ghosh S. Triterpene structural diversification by plant cytochrome P450 enzymes. *Front. Plant Sci.*, **8**, 1886 (2017).
 - 32) Yasumoto S, Seki H, Shimizu Y, Fukushima EO, Muranaka T. Functional characterization of CYP716 family P450 enzymes in triterpenoid biosynthesis in Tomato. *Front. Plant Sci.*, **8**, 21 (2017).
 - 33) Moses T, Pollier J, Almagro L, Buyst D, Van Montagu M, Pedreño MA, Martins JC, Thevelein JM, Goossens A. Combinatorial biosynthesis of sapogenins and saponins in *Saccharomyces cerevisiae* using a C-16 α hydroxylase from *Bupleurum falcatum*. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 1634–1639 (2014).
 - 34) Carelli M, Biazzi E, Panara F, Tava A, Scaramelli L, Porceddu A, Graham N, Odoardi M, Piano E, Arcioni S, May S, Scotti C, Calderini O. *Medicago truncatula* CYP716A12 is a multifunctional oxidase involved in the biosynthesis of hemolytic saponins. *Plant Cell*, **23**, 3070–3081 (2011).
 - 35) Wang D, Dai ZB, Zhang XL. Production of natural products from plants by yeast synthetic cells. *Acta Microbiol. Sin.*, **311**, 195–208 (2016).