New Phytologist

Identification of oxidosqualene cyclases from the medicinal legume tree *Bauhinia forficata*: a step toward discovering preponderant α-amyrin-producing activity

Pisanee Srisawat[1], Ery Odette Fukushima[1,2], Shuhei Yasumoto[1], Jekson Robertlee[1,3], Hideyuki Suzuki[4], Hikaru Seki[1] and Toshiya Muranaka[1]

[1]Department of Biotechnology, Graduate School of Engineering, Osaka University, Suita 565-0871, Japan; [2]Universidad Regional Amazónica IKIAM, Tena 150150, Ecuador; [3]Department of Frontier Research, Kazusa DNA Research Institute, Kisarazu 292-0818, Japan; [4]Department of Research & Development, Kazusa DNA Research Institute, Kisarazu 292-0818, Japan

Authors for correspondence:
Ery Odette Fukushima
Tel: +593 6 2999160 ext. 304
Email: ery.fukushima@ikiam.edu.ec

Toshiya Muranaka
Tel: +81 6 6879 7423
Email: muranaka@bio.eng.osaka-u.ac.jp

Received: *2 May 2019*
Accepted: *15 June 2019*

New Phytologist (2019) **224:** 352–366
doi: 10.1111/nph.16013

Key words: *Bauhinia forficata*, non-model legume, oxidosqualene cyclase, triterpenoid, α-amyrin.

Summary

- Triterpenoids are widely distributed among plants of the legume family. However, most studies have focused on triterpenoids and their biosynthetic enzymes in model legumes.
- We evaluated the triterpenoid aglycones profile of the medicinal legume tree *Bauhinia forficata* by gas chromatography–mass spectrometry. Through transcriptome analyses, homology-based cloning, and heterologous expression, we discovered four oxidosqualene cyclases (OSCs) which are responsible for the diversity of triterpenols in *B. forficata*. We also investigated the effects of the unique motif TLCYCR on α-amyrin synthase activity.
- *B. forficata* highly accumulated α-amyrin. We discovered an OSC with a preponderant α-amyrin-producing activity, which accounted for at least 95% of the total triterpenols. We also discovered three other functional OSCs (BfOSC1, BfOSC2, and BfOSC4) that produce β-amyrin, germanicol, and cycloartenol. Furthermore, by replacing the unique motif TLCYCR from BfOSC3 with the MWCYCR motif, we altered the function of BfOSC3 such that it no longer produced α-amyrin.
- Our results provide new insights into OSC cyclization, which is responsible for the diversity of triterpenoid metabolites in *B. forficata*, a non-model legume plant.

Introduction

Triterpenoids are one of the largest and most structurally diverse classes of natural specialized metabolites, and they have a variety of biological and ecological properties. Over 20 000 triterpenoids exist in nature, and they are widely used in medicine, industry and agriculture (Francis *et al.*, 2002; Xu *et al.*, 2004; González-Coloma *et al.*, 2011; Forestier *et al.*, 2019). Nonetheless, their low abundance and structural complexity limit the commercial applications of these compounds (Sawai & Saito, 2011; Moses *et al.*, 2013b). Triterpenoids can also be produced in a heterologous host by synthetic pathway manipulation (Moses *et al.*, 2013a; Reed *et al.*, 2017; Yu *et al.*, 2018), which requires identification of the enzymes involved in triterpenoid biosynthesis.

Triterpenoids and sterols are derived from the common linear precursor 2,3-oxidosqualene and are produced through two distinct pathways: the plastidial 2-C-methyl-d-erythritol 4-phosphate (MEP) and cytosolic mevalonate (MVA) pathways (Vranová *et al.*, 2013). The first committed step in triterpenoid biosynthesis is the cyclization of 2,3-oxidosqualene to various triterpenol scaffolds by oxidosqualene cyclase (OSC; Xu *et al.*, 2004). This is a complex series of reactions that leads to the formation of polycyclic molecules. This cyclization stage is the branch-off point of the sterol and triterpenoid biosynthetic pathways; the branch taken depends on the type of OSC (Xue *et al.*, 2012; Thimmappa *et al.*, 2014). The cyclization of 2,3-oxidosqualene in the chair-boat-chair conformation forms the protosteryl cation intermediate (Fig. 1), which is a sterol precursor in plants (Suzuki *et al.*, 2006). By contrast, the chair-chair-chair conformation produces the dammarenyl carbocation intermediate, which is modified to generate diverse triterpenol scaffolds (Fig. 1). Most triterpenols are further modified by oxidation and glycosylation to form triterpenoid glycosides, also known as saponins (Augustin *et al.*, 2011; Moses *et al.*, 2014; Thimmappa *et al.*, 2014; Seki *et al.*, 2015).

The ursane-, oleanane-, and lupane-type triterpenoids, derived from α-amyrin, β-amyrin, and lupeol, respectively, are the most widely distributed pentacyclic triterpenoids in the plant kingdom (Moses *et al.*, 2014). They are created via the dammarenyl carbocation intermediate by a rearrangement cascade of methyl and proton shifts (Phillips *et al.*, 2006). OSCs that catalyze the formation of β-amyrin or lupeol as a single product have been reported (Hayashi *et al.*, 2001; Zhang *et al.*, 2003; Kajikawa *et al.*, 2005; Sawai *et al.*, 2006), but there has been no report of α-amyrin

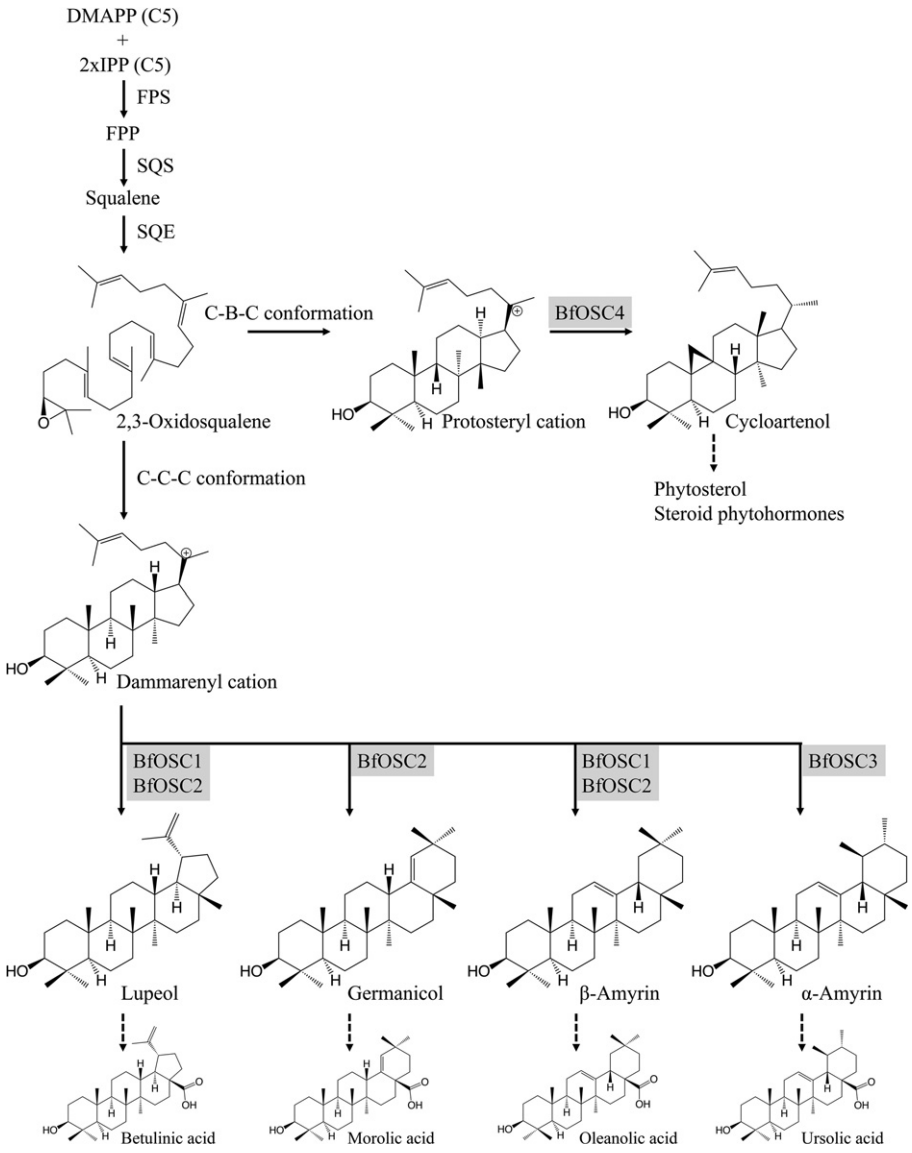

Fig. 1 Proposed sterol and triterpenoid biosynthetic pathways in *Bauhinia forficata*. Triterpenoids and sterols are synthesized from the common precursor 2,3-oxidosqualene, produced via the mevalonate pathway. The oxidosqualene cyclases (OSCs) reported in this study (BfOSCs) are shown in gray. DMAPP, dimethylallyl pyrophosphate; IPP, isopentenyl pyrophosphate; FPP, farnesyl pyrophosphate; C-C-C, chair-chair-chair; C-B-C, chair-boat-chair; FPS, farnesyl pyrophosphate synthase; SQS, squalene synthase; SQE, squalene epoxidase.

formed as a single product. All multifunctional OSCs reportedly involved in α-amyrin production generate a combination of other triterpenols, including β-amyrin and lupeol, in various proportions (Brendolise *et al.*, 2011; Huang *et al.*, 2012). Many OSCs from various plant species, including those of the Fabaceae family, have been cloned and characterized. Triterpenoid saponins are produced by plants of the Fabaceae (Leguminosae) family (Dixon & Sumner, 2003). However, most studies have focused on economically important and/or model legumes such as *Glycine max* and *Medicago truncatula*, which are of the Papilionoideae subfamily. Moreover, there is limited information on triterpenoids and their biosynthetic enzymes in legume plants other than the Papilionoideae clade.

Bauhinia is the largest genus in the Cercidoideae subfamily, which is within the clade farthest from the Papilionoideae subfamily in the Leguminosae family phylogenetic tree (LPWG, 2017). This genus consists of *c.* 300 species, which are widely distributed in most tropical countries (Sinou *et al.*, 2009). Their aerial parts have been used in folk medicine, particularly for treating diabetes, infection, and inflammation (Cechinel-Filho, 2009). *Bauhinia forficata* Link, commonly known as the Brazilian orchid tree or 'cow's foot' (Supporting Information Fig. S1), is a medicinal legume tree native to South America. The biological properties of *B. forficata* are among the most studied within the *Bauhinia* genus, due to the widespread use of this species in Brazilian folk medicine (de Sousa *et al.*, 2004; Cechinel-Filho, 2009; da Cunha *et al.*, 2010). Pre-clinical and clinical studies have suggested that the *B. forficata* leaf is a source of pharmacological flavonoids, which have been investigated as hypoglycemic agents for the treatment of diabetes, and as a diuretic agent for treating kidney and urinary disorders (de Sousa *et al.*, 2004; De Souza *et al.*, 2017). Although many studies have examined the properties of pharmacological flavonoids in *B. forficata*, triterpenoids from this species have not been studied. To date, there have been few reports of triterpenoids from plants in the *Bauhinia* genus; these have included studies of lupeol and

glutinol from *B. sirindhorniae* (Athikomkulchai *et al.*, 2003) and β-amyrin and oleanolic acid 3-O-caffeate from *B. variegata* (Rao *et al.*, 2008).

In this study, we examine the triterpenoid aglycone profile and the OSCs involved in triterpenoid scaffold diversification in the non-model legume *B. forficata*. We first conducted transcriptome analyses of leaf samples to identify the genes encoding OSCs. Next, we conducted *in vivo* functional analyses in heterologous hosts and discovered an OSC that predominantly produces α-amyrin. We also discovered three other functional OSCs, which mainly produce β-amyrin, germanicol, and cycloartenol. Finally, we performed OSC motif analyses and site-directed mutagenesis to identify the amino acids involved in α-amyrin production. Our results provide new insights into the OSC cyclization responsible for the diversity of triterpenoid metabolites in *B. forficata*. Our findings will facilitate the production of beneficial triterpenoids.

Materials and Methods

Plant materials

Bauhinia forficata Link seeds, obtained from the Desert Legume Program of the University of Arizona, were grown in standard soil in a glasshouse under controlled conditions (20–25°C, 60% humidity, 140–300 $\mu mol\ m^{-2}\ s^{-1}$ light intensity, 16 h : 8 h, light : dark cycle) for 10 months. The flowers, leaves, stems and roots were harvested and immediately soaked in liquid nitrogen for triterpenoid analyses and RNA extraction.

Metabolite extraction from *B. forficata* tissues

B. forficata samples (flowers, leaves, stems and roots) were freeze-dried and ground to a fine powder. The powder (1 g) was reflux-extracted with methanol for 4 h at 70°C. The methanol extracts were evaporated to dryness. Equal volumes of methanol and 4 N hydrochloric acid were added to the dried methanol extract and incubated for 2 h at 80°C, followed by hexane extraction. The organic layer was collected, dried and dissolved in methanol.

Total RNA isolation and cDNA synthesis

Frozen samples were ground in liquid nitrogen into a fine powder using a mortar and pestle. Total RNA was isolated using TRIzol® Reagent (Invitrogen), following the manufacturer's protocol. Residual DNA was removed using recombinant DNase I (TaKaRa Bio, Shiga, Japan), followed by phenol/chloroform extraction. PrimeScript RT Master Mix (TaKaRa Bio) was used to synthesize complementary DNA (cDNA) from total RNA.

Illumina sequencing, *de novo* assembly and library construction

Total RNA (10 μg) was used to construct a cDNA library using the NEBNext Ultra RNA Library Prep Kit according to the manufacturer's protocol (New England Biolabs, Tokyo, Japan). The resulting cDNA library was sequenced using the HiSeq 1500 instrument (Illumina, San Diego, CA, USA) with 100 base pair (bp) paired-end reads in rapid-run mode. The sequence reads were assembled using CLC Genomics Workbench software (v.10.0.1; CLC Bio, Tokyo, Japan) with a minimum contig length of 300 bp and the 'perform scaffolding' function. The raw RNA-Seq reads obtained were submitted to the DNA Data Bank of Japan (DDBJ) Sequence Read Archive under accession no. DRA008057.

Functional annotation of assembled sequences

To obtain a set of putative *OSC*s, KEGG pathway mapping was performed using the KEGG Automatic Annotation Server (https://www.genome.jp/kegg/kaas/) with the bi-directional best hit method against a dataset comprising 'ath, aly, brp, bna, cit, gmx, mtr, lja, adu, aip, pop, vvi, sly and osa', with the default threshold setting.

Cloning of *OSC*s

OSC candidates were identified from an in-house *B. forficata* transcriptome database by Blast searching based on known OSCs from legumes and primers specific for legume OSCs. Homology-based primers designed to target the conserved region of legume OSCs (Table S1) were used to amplify the core fragment of *OSC*s from *B. forficata*. The fragments obtained by homology-based PCR and from the in-house transcriptome database were extended using 3′and 5′rapid amplification of cDNA ends (RACE) PCR. Total RNA obtained from *B. forficata* leaves was used to synthesize first-strand cDNA with a SMARTer RACE cDNA Amplification Kit (Clontech, Shiga, Japan), according to the manufacturer's instructions; the cDNA served as a template for RACE-PCR using gene-specific primers (Table S1) to amplify the open reading frame (ORF) end fragment. The full-length ORFs were amplified by PCR using 5′- and 3′- end gene-specific primers (Table S1), cloned into either pDONR-221 or pENTR-D-TOPO (Invitrogen) and confirmed by sequencing.

Phylogenetic analyses

The OSC protein sequences were retrieved from GenBank and aligned with the four sequences obtained in this study using the Muscle algorithm (Edgar, 2004). A neighbor-joining tree was constructed using Mega7 software (Kumar *et al.*, 2016) with default parameters based on 1000 bootstrap replications.

Generation of *Saccharomyces cerevisiae* SY100

The sequence upstream of the *ERG7* promoter (−1000 to −500) and of the *MET3* promoter (−500 to −1), as well as the partial *ERG7*-coding sequence (+1 to +500), were amplified from *S. cerevisiae* BY4742 genomic DNA using the primer sets pY425/pY426, pY427/pY428, and pY429/pY430, respectively (Table S1). Next, *pERG7-pMET3* and *pMET3*-ERG7 were joined by PCR. The joined amplicons were cloned into the pJET1.2 Blunt Cloning Vector (Thermo Scientific, Waltham,

MA, USA), and pJET-pERG7-pMET3 and pJET-pMET3-ERG7 were digested with *Sma*I/*Bsa*I and *Bsa*I/*Eco*RI, respectively. The DNA fragments were purified and ligated into the *Sma*I/*Eco*RI site of pAUR135 (TaKaRa Bio), resulting in pAUR135-MET3-ERG7. Next, pAUR135-MET3-ERG7 was linearized by *Xba*I digestion and introduced into *S. cerevisiae* BY4742, which was cultured on synthetic defined (SD) medium (Difco, Frankin Lakes, NJ, USA) with methionine dropout supplement (Clontech) agar containing glucose and aureobasidin A. The transformants with the DNA fragment at the *ERG7* locus were streaked on SD medium with methionine dropout supplement (Clontech) agar containing galactose to remove the vector backbone, which contained an aureobasidin A resistance gene. We selected yeast clones with replacement of the *MET3* promoter at the *ERG7* locus by genomic PCR using the primers pY431/pY432; the resulting strain was named SY100.

Functional analyses of BfOSCs

*BfOSC*s and mutant *OSC*s were ligated into a yeast expression vector (pYES-DEST52) by LR reaction (Invitrogen), and transformed into *S. cerevisiae* SY100. The resulting strains (PS0–PS21, Table S2) were cultured in 5 ml SD medium (Difco) containing glucose with uracil dropout (Clontech) for 36 h at 30°C, with agitation. To induce expression, the cell pellet was resuspended into 5 ml SD medium containing 2% galactose, 1 mM methionine, and uracil dropout supplement for 48 h at 30°C, with agitation. The cell pellet was harvested and extracted using ethyl acetate. The experiment was performed in triplicate.

Transient expression in *Nicotiana benthamiana* leaves

The ORF of BfOSC3 and mutant *Arabidopsis thaliana* HMGR1 catalytic domain (AtHMGR1cd-S577A; Robertlee *et al.*, 2017) were cloned into the pYS_015 binary vector (Methods S1) by LR reaction (Invitrogen). The RNA-silencing repressor p19 of tomato bushy stunt virus (NP_062901.1) was cloned into pRI101-AN-GW (Muangphrom *et al.*, 2016) by LR reaction (Invitrogen) and transformed into *Agrobacterium tumefaciens* strain GV3101 (pMP90). Freshly grown *A. tumefaciens* cells were resuspended in infiltration medium (10 mM 2-(N-morpholino) ethanesulfonic acid, 10 mM $MgSO_4$, and 200 µM acetosyringone; pH 5.6) to an optical density at 600 nm (OD_{600}) of 0.5. Equal volumes of *A. tumefaciens* harboring *AtHMGR1cd-S577A*, *BfOSC3*, and *p19* were combined and infiltrated into the leaves of 5-wk-old *N. benthamiana*. The leaves of five plants were harvested for triterpenoid analyses 5 d post-infiltration. Powdered freeze-dried samples (100 mg) were extracted with 1 ml chloroform : methanol (1 : 1 v/v). The mixtures were briefly vortexed, disrupted using a water-bath sonicator for 1 h at room temperature, and centrifuged at 8000 ***g*** for 10 min. The supernatants were collected, dried, and dissolved in methanol.

Gas chromatography–mass spectrometry analyses

Samples (100 µl) were concentrated under vacuum and derivatized with 50 µl N-methyl-N-(trimethyl-silyl) trifluoroacetamide (Sigma-Aldrich) at 80°C for 30 min. The samples were run on an HP-5MS capillary column (30 m × 0.25 mm inner diameter, 0.25 µm film thickness; J&W Scientific, Santa Clara, CA, USA) or DB-1MS capillary column (30 m × 0.25 mm inner diameter, 0.10 µm film thickness; J&W Scientific) in a gas chromatograph (7890B; Agilent Technologies, Santa Clara, CA, USA) connected to a mass spectrometer (5977A; Agilent Technologies). The injection temperature was 250°C with splitless injection, and the interface temperature was 280°C. The column temperature program was 80–265°C at a rate of 10°C min^{-1} and a hold at 265°C for 30.5 min; or 80–300°C at a rate of 20°C min^{-1} and a hold at 300°C for 35 min. The carrier gas was helium at a flow rate of 1.0 ml min^{-1}. The compounds were identified by comparison with β-amyrin, α-amyrin, lupeol, cycloartenol, erythrodiol, uvaol, betulin, oleanolic acid, ursolic acid, soyasapogenol B (Extrasynthese, Lyon, France), germanicol and morolic acid (Wuhan ChemFaces Biochemical, Wuhan, China) authentic standards.

PCR-based site-directed mutagenesis

Entry clones of *BfOSC1*, *BfOSC2*, and *BfOSC3* were used as templates for PCR-based site-directed mutagenesis (Table S1) to substitute the target amino-acid sequence. Mutant *OSC*s were transferred to the expression vectors and confirmed by sequencing.

Results

α-Amyrin and germanicol are the major triterpenes in *B. forficata*

The leaves of *B. forficata* harbored a greater diversity of triterpenoids than its flower, stem or root (Tables 1, S3). The major triterpene detected was α-amyrin; germanicol, β-amyrin and lupeol were also detected (Figs 2, S2). Interestingly, *B. forficata* triterpenoids accumulated as triterpenols such as α-amyrin and germanicol (Fig. S2), which were unlike the triterpenoid profile of model legumes. Model legumes such as *Medicago truncatula* (barrel medic) and *Glycine max* (soybean) tend to accumulate oxidized triterpenoids derived from the β-amyrin scaffold such as soyasapogenol B (a β-amyrin derivative with hydroxyl groups at C-22 and C-24), erythrodiol (hydroxyl group at C-28), and oleanolic acid (carboxyl group at C-28) (Pollier *et al.*, 2011; Krishnamurthy *et al.*, 2014). Lupeol and its derivatives, such as

Table 1 The major pentacyclic triterpenoid scaffolds detected in *Bauhinia forficata* tissues.

Compounds	Flower	Leaves	Stem	Root
β-amyrin (1)	++	++	+	+
Germanicol (2)	++	++	++	Trace
α-amyrin (3)	++	+++	+++	Trace
Lupeol (4)	++	++	++	Trace

Compound concentrations indicated by plus signs are as follows: (+), 30–50 µg g^{-1} DW; (++), 100–200 µg g^{-1} DW; (+++), 300–500 µg g^{-1} DW; and (trace), < 10 µg g^{-1} DW.

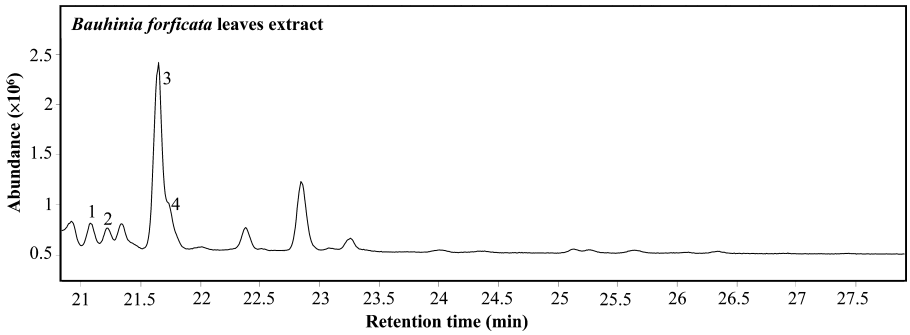

Fig. 2 The major pentacyclic triterpenoid scaffolds detected in *Bauhinia forficata* extracts. The total ion chromatogram of *B. forficata* leaves extracted on the DB-1MS column is shown. β-amyrin (1), germanicol (2), α-amyrin (3), and lupeol (4). The total ion chromatogram retention times in comparison with those of authentic standards are shown in Supporting Information Fig. S2.

betulin (a lupeol derivative with a hydroxyl group at C-28) and betulinic acid (carboxyl group at C-28), accumulate in the economically important legume *Glycyrrhiza uralensis* (licorice; Hayashi *et al.*, 2005; Tamura *et al.*, 2017). In the Leguminosae phylogenetic tree, *B. forficata* is in the same clade as the Cercidoideae subfamily, which is the farthest clade from the model legumes (Papilionoideae subfamily; LPWG, 2017). All of the model and economically important legumes such as barrel medic, licorice, and soybean are in the Papilionoideae subfamily, in which soyasapogenol B is a common metabolite (Dixon & Sumner, 2003; Pollier *et al.*, 2011; Krishnamurthy *et al.*, 2014). By contrast, we did not detect soyasapogenol B in *B. forficata* extracts but noted accumulation of germanicol and morolic acid (Figs 2, S2; Table S3). To the best of our knowledge, these compounds have not been detected in any model legume. Interestingly, morolic acid has been detected in heartwood extract of the legume tree *Mora excelsa* (Barton & Brooks, 1950) of the subfamily Caesalpinioideae. Despite the dearth of information on triterpenoid accumulation and biosynthesis in legume trees (other than the Papilionoideae subfamily), we speculated that these processes in legume trees might differ from those in members of the Papilionoideae subfamily.

Isolation of four candidate *OSC*s based on *B. forficata* transcriptomic analyses

From the triterpenoid accumulation pattern in *B. forficata*, we assumed that the *OSC*s responsible for triterpenoid scaffold cyclization are mainly expressed in leaves. Because neither genomic nor transcriptomic data for *B. forficata* are available, we performed transcriptomic analyses of *B. forficata* leaves. We obtained 40 407 566 total reads and assembled them into 45 443 contigs (Table S4). Next, we performed BLAST analyses to annotate 29 217 of the contigs using the NCBI non-redundant protein database. To identify genes potentially involved in triterpenoid biosynthesis, we performed KEGG pathway mapping using the KEGG automatic annotation server, resulting in annotation of 10 contigs as putative *OSC*s (Table S5). Among them, the full-length sequence of one was annotated as a putative cycloartenol synthase; the other nine comprised 300–800 nucleotides (nt). To isolate *OSC* genes expressed in leaf tissue, among the 10 contigs annotated as putative *OSC*s, we selected contig00009800, contig00011765 and contig00018000 for cloning due to their length and different annotations. Three pairs of gene-specific primers were designed to amplify the core fragments of these three candidate *OSC*s. In addition, a set of homology-based primers designed based on the conserved sequence of legume β-amyrin synthases were employed to amplify the core segment of the candidate β-amyrin synthase from *B. forficata*. The PCR products obtained were resolved by gel electrophoresis, resulting in one major band of the expected size (1400 nt). Next, the core fragments of the candidate *OSC*s were extended by RACE-PCR. We identified four full-length sequences of putative *OSC*s from *B. forficata*, which were named *BfOSC1* (DDBJ accession no. LC464978), *BfOSC2* (DDBJ accession no. LC464979), *BfOSC3* (DDBJ accession no. LC464980), and *BfOSC4* (DDBJ accession no. LC464981). The deduced amino-acid sequences of these four *OSC*s showed 52–84% similarity (Table S6). All four *B. forficata* OSCs possessed the DCTAE motif, which is involved in substrate binding (Ito *et al.*, 2013b), and four QW motifs characteristic of the OSC superfamily (Fig. 3a). The QW motifs may be involved in stabilizing carbocation during cyclization (Poralla, 1994; Kushiro *et al.*, 1998). We constructed a phylogenetic tree of *B. forficata* OSCs and characterized OSCs from other plant species. BfOSC1 clustered within the clade of previously characterized β-amyrin synthases; BfOSC2 and BfOSC3 were in the same clade with previously characterized multifunctional OSCs; and BfOSC4 clustered with previously characterized cycloartenol synthases (Fig. 4).

BfOSC1–BfOSC4 are responsible for the triterpenol diversity in *B. forficata*

To characterize the putative OSCs from *B. forficata*, we used a yeast *in vivo* expression system. We engineered *S. cerevisiae* BY4742 by replacing the native promoter of *ERG7* (*lanosterol synthase*) with a methionine-repressible promoter to create the SY100 yeast strain. The full-length ORFs were heterologously expressed from the high-copy yeast expression vector pYES-DEST52 in *S. cerevisiae* SY100 to generate strains PS1–PS4. We extracted and identified triterpenols by gas chromatography–mass spectrometry. The chromatogram of the PS0 strain, SY100, harboring the pYES-DEST52 empty vector, was used as a background control. We observed an increase in endogenous yeast

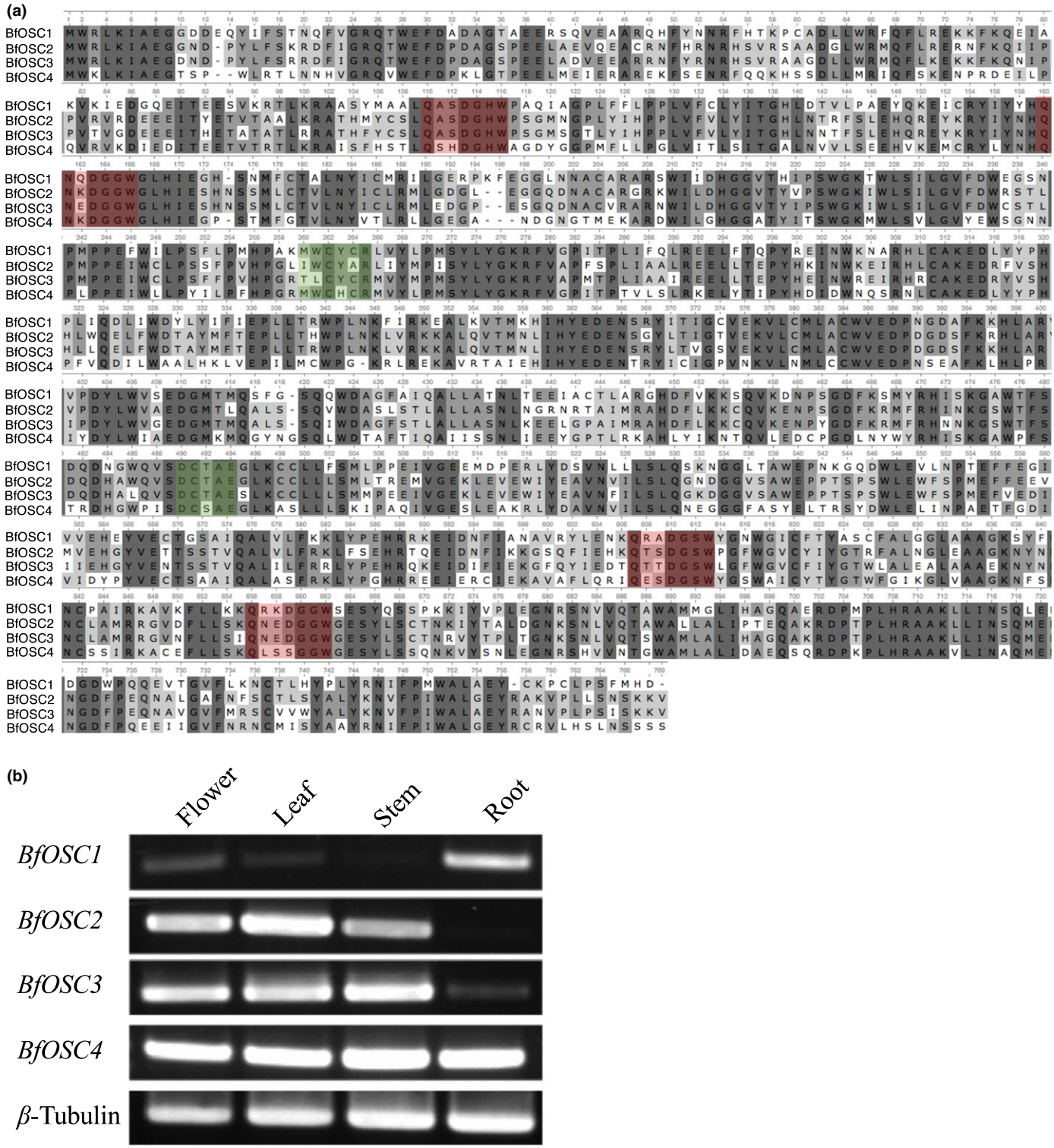

Fig. 3 Deduced amino-acid sequence alignment and gene expression patterns of four OSCs from *Bauhinia forficata* (BfOSCs). (a) The conserved DCTAE motif and (M/T)(W/L)CY(C/A)R motif sequences are highlighted in green. Four QW motifs (QXXXGXW) are highlighted in red. Light gray to dark gray indicates the increasing magnitude of sequence similarity. (b) Gene expression analysis of *BfOSCs* in different tissues. *BfOSC2* and *BfOSC3* were expressed in aerial parts, and *BfOSC1* in the root. *BfOSC4*, which generates cycloartenol, was expressed equally in all tissues tested.

metabolites when heterologous OSCs were expressed in strains PS1–PS4 (Fig. S3). The chromatogram of PS1 (expressing *BfOSC1*) had a major peak corresponding to β-amyrin together with a smaller peak corresponding to α-amyrin and lupeol (Figs 5, S4). The proportions of triterpenols in PS1 suggested that *BfOSC1* encodes a multifunctional OSC that catalyzes the production of β-amyrin : α-amyrin : lupeol at a 94 : 4 : 2 ratio. The chromatogram of PS2 (expressing *BfOSC2*) had a peak

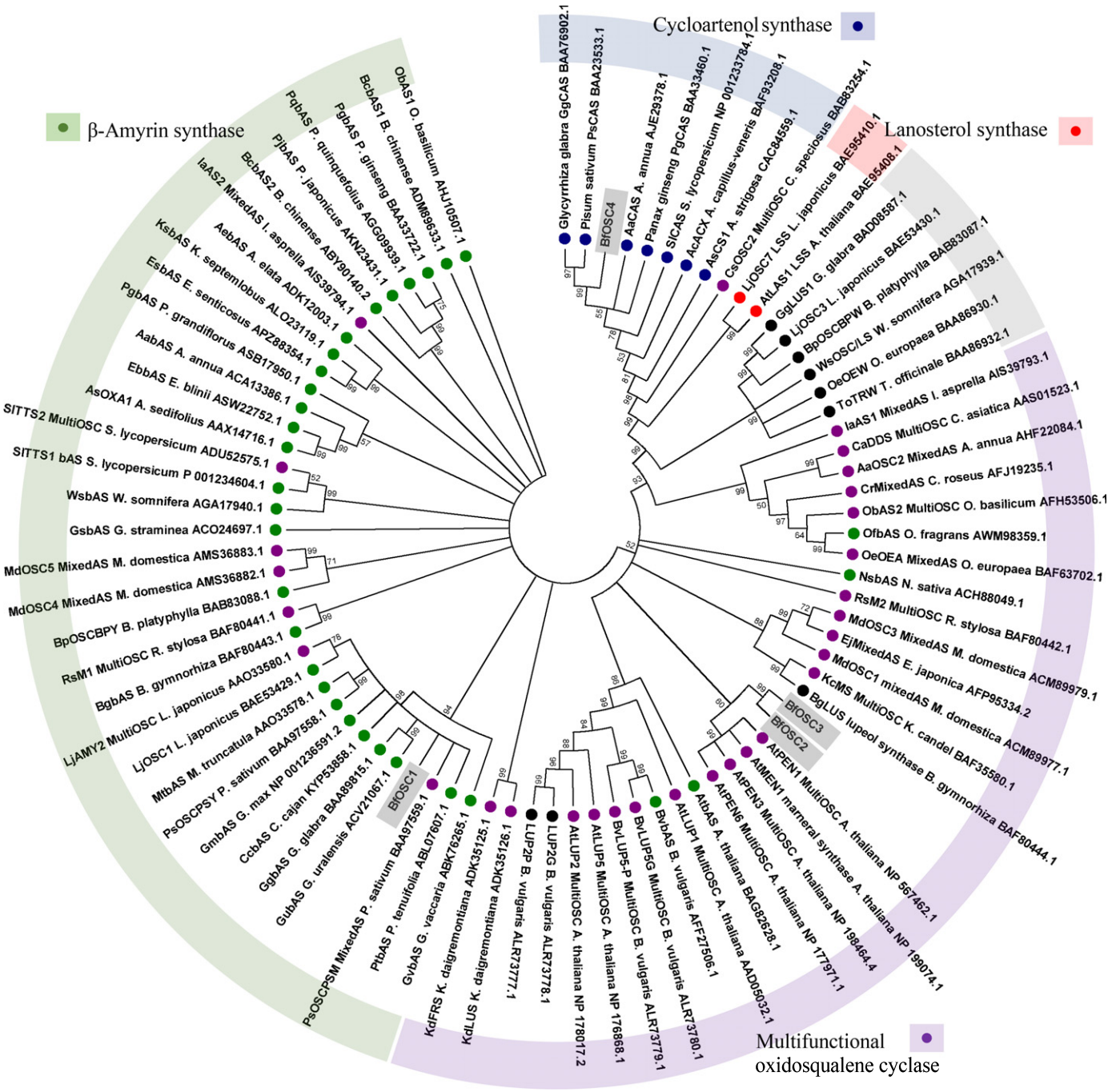

Fig. 4 Phylogenetic tree of plant oxidosqualene cyclases (OSCs), including BfOSCs. β-Amyrin synthases, green; lupeol synthases, black; lanosterol synthases, red; cycloartenol synthases, blue; multifunctional OSCs, purple. BfOSC1, BfOSC2, BfOSC3, and BfOSC4 are highlighted in gray. Amino acid sequences were retrieved from GenBank and aligned using MUSCLE. GenBank accession nos. are indicated after the protein name in the phylogenetic tree. The phylogenetic tree was constructed by MEGA7 using the neighbor-joining method with 1000 bootstrap replicates.

corresponding to germanicol together with peaks corresponding to β-amyrin and lupeol (Figs 5, S4). The product proportions indicated that *BfOSC2* encodes a multifunctional OSC that catalyzes the production of germanicol : β-amyrin : lupeol at a 90 : 9 : 1 ratio. The chromatogram of PS3 (expressing *BfOSC3*) had a peak corresponding to α-amyrin; this was replicated in the extracted ion chromatogram (EIC; $m/z = 218$), suggesting that *BfOSC3* encodes an α-amyrin synthase (Figs 5, S4).

According to our findings regarding the yeast expression system, BfOSC3 appears to produce α-amyrin specifically, but in low concentrations (Fig. 5). To confirm the function of BfOSC3, we expressed *BfOSC3* in an *Escherichia coli in vivo* expression system, but did not detect triterpenols (data not shown). Therefore, it was unclear why BfOSC3 has low productivity in a yeast expression system and no activity in an *E. coli* expression system but α-amyrin accumulates to a higher concentration in

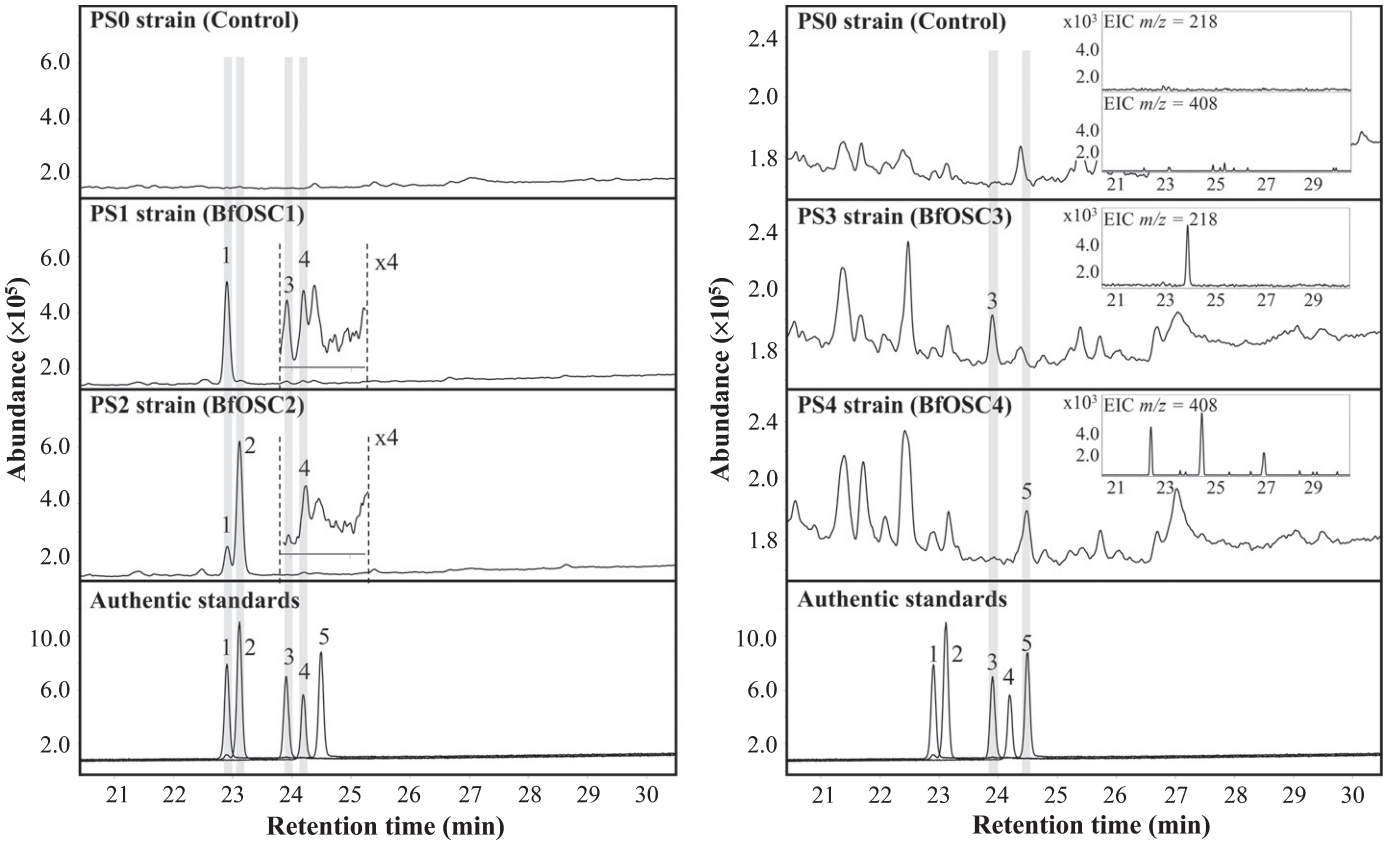

Fig. 5 Total ion chromatograms of yeast strains expressing BfOSCs. β-Amyrin (1), germanicol (2), α-amyrin (3), and cycloartenol (5) were detected as major products in yeasts expressing BfOSC1 (PS1), BfOSC2 (PS2), BfOSC3 (PS3), and BfOSC4 (PS4), respectively. Lupeol (4) was detected as a minor product in PS1 and PS2. The triterpenol products were identified by comparing their retention times and mass fragmentation patterns with the corresponding authentic standards (Supporting Information Fig. S4). The yeast SY100 (pYES-DEST52 empty vector, PS0) was used as the background control. The lower intensity peaks in the chromatograms of PS1 and PS2 are shown magnified fourfold. The EICs of PS3 ($m/z = 218$) and PS4 ($m/z = 408$) are shown. Data are representative of at least three biological replicates.

B. forficata (Fig. 2, Table 1). We hypothesized that microbial expression systems might not be suitable for *BfOSC3* expression. Therefore, we used the *N. benthamiana* transient expression system to evaluate the function of BfOSC3. We co-expressed BfOSC3 with AtHMGR1cd-S577A, which lacks the inactivating phosphorylation site and has greater activity *in vitro* than AtHMGR1cd (Robertlee *et al.*, 2017, 2018). The HMGR1 catalytic domain with the S577A mutation enhances the production of isoprenoids (latex) in dandelion (Pütter *et al.*, 2017). Moreover, co-expression of the *HMGR* catalytic domain with triterpenoid biosynthetic genes increases triterpenoid production in an *N. benthamiana* transient expression system (Reed *et al.*, 2017). Therefore, we applied these strategies to our plant transient expression system. *BfOSC3* and *AtHMGR1cd-S577A* were cloned downstream of the CaMV 35S promoter in a binary vector and transformed individually into *A. tumefaciens*. We co-infiltrated *A. tumefaciens*, harboring the constructs into *N. benthamiana* leaves, which were harvested 5 d later and processed for metabolite extraction. Leaves transfected with *A. tumefaciens* harboring *p19* and *AtHMGRcd_S577A* were used as the background control. We observed a single major peak in the gas chromatogram of *N. benthamiana* leaves transiently expressing *BfOSC3* but not in the background control (Fig. 6). This peak was identified as α-amyrin by comparison of its mass fragmentation pattern with that of the α-amyrin authentic standard (Fig. S5). We also detected five minor peaks corresponding to β-amyrin, lupeol, and three unidentified compounds (Figs 6, S5). The minor peak at 26.0 min was putatively identified as taraxasterol, based on the literature (Wang *et al.*, 2011). Catalysis by BfOSC3 resulted in a total triterpenol profile comprising ≥95% α-amyrin. We cannot definitively conclude that BfOSC3 is a single-product α-amyrin synthase, but we also cannot exclude the possibility that by-products might be derived from the *N. benthamiana* background. The data suggest that plant expression systems are suitable for *BfOSC3*, and that BfOSC3 is the main contributor to α-amyrin production in *B. forficata*.

In the chromatogram of yeast expressing *BfOSC4* (PS4), cycloartenol was detected (Fig. 5), indicating that BfOSC4 has cycloartenol synthase activity. The EIC ($m/z = 408$) contained unidentified minor peaks (Fig. 5), suggesting that BfOSC4 is a multifunctional OSC or that its characterization was hindered by native yeast sterol biosynthesis enzymes. To confirm the function of BfOSC4, we performed an *in vivo* enzymatic assay in *E. coli*. We assembled the mevalonate pathway from *S. cerevisiae* and introduced squalene synthase (SQS) and squalene epoxidase (SQE) from *Arabidopsis thaliana* to supply 2,3-oxidosqualene as

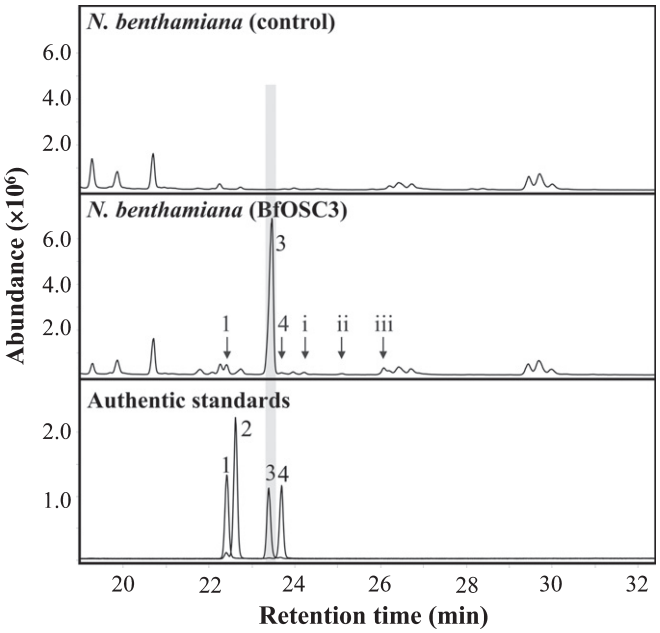

Fig. 6 Total ion chromatogram of *Nicotiana benthamiana* leaves transiently expressing *BfOSC3*. *N. benthamiana* transiently expressing *p19* and AtHMGRcd_S577A was used as the background control. α-Amyrin (3) was detected as the major product in the leaf extract. Five minor peaks were also observed; these corresponded to β-amyrin (1), lupeol (4), putative taraxasterol (iii), and two unidentified compounds (i, ii). The mass fragmentation patterns of α-amyrin and the three unidentified (peaks i–iii) are shown in Supporting Information Fig. S5. Data are representative of five biological replicates.

an OSC precursor (Methods S2). *BfOSC4* was heterologously expressed from the *E. coli* expression vector, pDEST17. Gas chromatography–mass spectrometry analyses showed a single peak of cycloartenol in both the total ion and extracted ion (EIC *m/z* = 408) chromatograms (Fig. S6), confirming that *BfOSC4* encodes a cycloartenol synthase. Therefore, the minor peaks detected in PS4 may be derived from the conversion of cycloartenol by yeast enzyme(s).

Reverse transcription PCR (RT-PCR) was conducted to assess *OSC* expression in *B. forficata* leaf, root, stem, and flower. *BfOSC4*, which encodes cycloartenol synthase, was expressed equally in all tested tissues (Fig. 3b). *BfOSC2* and *BfOSC3* encode multifunctional OSCs that catalyze the production of mainly germanicol and α-amyrin, respectively; both were highly expressed in the aerial parts (Fig. 3b). *BfOSC1*, which encodes a multifunctional OSC that catalyzes the production of mainly β-amyrin, was highly expressed in the root (Fig. 3b). Thus, we speculate that BfOSC2 plays a role in producing β-amyrin in flower, leaf, and stem. These results are in agreement with triterpenol accumulation in *B. forficata* tissues (Table 1), suggesting a role for these four BfOSCs in triterpenol diversification.

Residues T257 and L258 are important for BfOSC3 α-amyrin synthase activity and selectivity

We further analyzed the MXCYXR motif sequence, which is a determinant of the products generated by OSCs (Kushiro *et al.*, 2000), in BfOSCs and previously characterized OSCs. Amino acid sequence alignment of the MXCYXR motif (Table 2) showed that the tryptophan residue of MWCYCR is conserved among β-amyrin synthases, while the leucine residue is conserved among lupeol synthases (Kushiro *et al.*, 2000). In the multifunctional OSCs that catalyze the production of α-amyrin, we found variation in the methionine and tryptophan residues of MXCYXR (Table 2). Interestingly, BfOSC3, which preponderantly produces α-amyrin, has a TLCYCR (257–262) sequence. Multifunctional OSCs that mainly produce α-amyrin from *Barbarea vulgaris* (BvLUP5P) also exhibit this motif sequence (Khakimov *et al.*, 2015). Therefore, we hypothesized that the T257 and L258 residues in the TLCYCR motif play roles in the α-amyrin synthase activity of BfOSC3. By site-directed mutagenesis we introduced T257M and L258W mutations individually and also created the double mutation. The *BfOSC3* mutants were heterologously expressed in SY100 to generate the yeast strains PS5–PS7. The gas chromatogram of PS5 (expressing BfOSC3_T257M) contained α-amyrin and had the same product profile as BfOSC3 (Fig. 7a). The L258 residue of the MXCYXR motif mediates the lupeol specificity of OSCs (Kushiro *et al.*, 2000; Ito *et al.*, 2017). Unexpectedly, the gas chromatogram of PS6 and PS7 (expressing BfOSC3_L258W and BfOSC3_T257M/L258W) lacked triterpenol peaks, as did the EIC (*m/z* = 218; Fig. 7a), suggesting that L258 is important for the catalytic activity of BfOSC3. However, we cannot exclude the possibility that loss of activity in the BfOSC3 mutants was caused by protein instability in the yeast expression system. Therefore, we determined the activity of the *BfOSC3* mutants in the *N. benthamiana* transient expression system. As expected, there was no peak corresponding to pentacyclic triterpenols in the gas chromatogram of *N. benthamiana* leaves transiently expressing *BfOSC3*_L258W and *BfOSC3*_T257M/L258W (Fig. 7b). However, the peak corresponding to tetracyclic triterpenols was of greater area (Figs 7b, S5). The peak at 23.9 min was identified as cycloartenol by comparison with the authentic cycloartenol standard, and one minor peak at 28.0 min was identified as a putative damarenediol-II by comparison of its mass fragmentation pattern with a prior report (Salmon *et al.*, 2016). These results suggest the importance of both residues for pentacyclic cyclization by BfOSC3. Similar results have been reported previously for SAD1 (β-amyrin synthase from *Avena strigosa*), which transformed enzyme activity (from mainly producing pentacyclic products to mainly producing tetracyclic products) via a single mutation at S728F (Salmon *et al.*, 2016).

The TLCYCR motif affects the products of BfOSC1 and BfOSC2

Next, we investigated the effects of the threonine and leucine residues in the MXCYXR motif of other BfOSCs. We replaced the MXCYXR motif of BfOSC1 and BfOSC2 with a TLCYCR motif to generate the yeast strains PS8–PS17. The gas chromatogram of PS8–PS10 (*BfOSC1* mutants) showed a trace amount of lupeol; β-amyrin and α-amyrin were not detected (Fig. 8a), suggesting suppression of catalytic activity. These

Table 2 Amino acid sequence alignment of the M(T)W(L)CYC(T/A)R motif of BfOSCs and other plant oxidosqualene cyclases (OSCs).

Enzyme	Plant/Enzyme name	Accession no.	Motif	M(T)	W(L)	CYC	(T/A)	R	
β-AS	*G. glabra*/GgβAS	BAA89815	256	M	W	C	Y	C	R	261
	A. annua/AaβAS	ACA13386	256	M	W	C	Y	C	R	261
	B. platyphylla/BpBPY	BAB83088	256	M	W	C	Y	C	R	261
	P. sativum/PsPSY	BAA97558	256	M	W	C	Y	C	R	261
	L. japonicus/LjOSC1	BAE53429	256	M	W	C	Y	C	R	261
	B. gymnorhiza/BgβAS	BAF80443	256	M	W	C	Y	C	R	261
	M. truncatula/MtβAS	AAO33578	256	M	W	C	Y	C	R	261
	N. sativa/NsβAS	ACH88049	260	M	W	C	Y	C	R	265
LUS	*L. japonicus*/LjOSC3	BAE53430	254	M	L	C	Y	C	R	259
	O. europaea/OeOEW	BAA86930	255	M	L	C	Y	C	R	260
	T. officinale/ToTRW	BAA86932	257	M	L	C	Y	C	R	262
	G. glabra/GgLUS1	BAD08587	254	M	L	C	Y	C	R	259
	B. platyphylla/BpBPW	BAB83087	254	M	L	C	Y	C	R	259
	B. gymnorhiza/BgLUS	BAF80444	256	M	L	C	Y	C	R	261
MFS	*A. annua*/AaOSC2	AHF22084	256	M	W	C	Y	C	R	261
	A. thaliana/AtLUP2	NP178018	257	T	L	C	Y	T	R	262
	B. vulgaris/BvLUP5P	ALR73780	257	T	L	C	Y	C	R	262
	O. basilicum/ObAS2	AFH53506	260	M	W	C	Y	C	R	264
	P. sativum/PsPSM	BAA97559	256	M	L	C	Y	C	R	261
	M. domestica/MdOSC1	ACM89977	256	M	F	C	Y	C	R	261
	E. japonica/EjAS	AFP95334	256	M	F	C	Y	C	R	261
	O. europaea/OeOEA	BAF63702	257	M	W	C	Y	C	R	262
	C. roseus/CrAS	AFJ19235	257	M	W	C	Y	C	R	262
	I. asprella/MixedAS1	AIS39739	256	M	W	C	Y	C	R	261
BfOSC	*B. forficata*/BfOSC1	LC464978	259	M	W	C	Y	C	R	264
	B. forficata/BfOSC2	LC464979	257	I	W	C	Y	A	R	262
	B. forficata/BfOSC3	LC464980	257	T	L	C	Y	C	R	262
	B. forficata/BfOSC4	LC464981	254	M	W	C	H	C	R	259

β-AS, β-amyrin synthase; LUS, lupeol synthase; MFS, multifunctional OSC that produces α-amyrin. Shading light gray to dark gray indicates the increasing magnitude of sequence similarity.

results indicate the importance of M259 and W260 for the catalytic activity and β-amyrin product specificity of BfOSC1.

The chromatogram of PS11 (expressing *BfOSC2*_I257T) was identical to that of the BfOSC2 wild-type (PS2; Fig. 8b). The chromatogram of PS13 (expressing *BfOSC2*_A261C) showed germanicol but not β-amyrin or lupeol (Fig. 8b). Interestingly, the chromatogram of PS15 (expressing *BfOSC2*_I257T/A261C) had germanicol as a major product together with β-amyrin and lupeol, similar to the BfOSC2 wild-type (PS2). These results are unexpected because the I257T/A261C (PS15) double mutation restored the BfOSC2 product profile to that of wild-type. Therefore, the I257 and A261 residues are important for the catalytic activity, but not the product selectivity, of BfOSC2. In the yeast strains expressing *BfOSC2*_W258L (PS12), *BfOSC2*_I257T/W258L (PS14), and *BfOSC2*_W258L/A261C (PS16), lupeol was a major product and germanicol was a minor product (Fig. 8b). These results suggest that these residues are essential for product selectivity, and that W258 is required for the germanicol synthase activity of BfOSC2. Moreover, mutation of A261 of *BfOSC2* suppressed triterpenol production (compare PS13 to PS16). By contrast, mutation of I257 enhanced triterpenol production (compare PS12 to PS14, and PS13 to PS15). Therefore, A261 may be important for the catalytic activity of BfOSC2. In the yeast strain expressing *BfOSC2*_I257T/W258L/A261C (PS17), germanicol and lupeol were detected in equal concentrations, while β-amyrin was not detected (Fig. 8b). Therefore, these three sites are determinants of the product profile and the germanicol synthase activity of BfOSC2.

Discussion

We cloned and characterized the OSCs responsible for the diversity of triterpenoids in *B. forficata* and investigated the role of the TLCYCR motif in α-amyrin synthase activity. α-Amyrin accumulated in most of the *B. forficata* tissues tested, unlike all of the other model legumes. Therefore, we hypothesized that *B. forficata* possesses a unique OSC that produces α-amyrin. Transcriptome analyses and homology-based cloning resulted in identification of four full-length *OSC*s.

An *in vivo* assay in *S. cerevisiae* confirmed that BfOSC1 catalyzes the conversion of 2,3-oxidosqualene to β-amyrin, α-amyrin, and lupeol at a 94 : 4 : 2 ratio. In addition, BfOSC2 catalyzes the conversion of 2,3-oxidosqualene to germanicol, β-amyrin, and lupeol at a 90 : 9 : 1 ratio. To the best of our knowledge, an OSC in a legume plant that mainly catalyzes the production of germanicol has not been reported to date. Therefore, BfOSC2 is the first legume enzyme reported to produce mainly germanicol. Multifunctional OSCs that produce germanicol have been reported in *Rhizophora styloza* (mangrove; RsM1, produces germanicol : β-amyrin : lupeol at a 63 : 33 : 4 ratio) (Basyuni

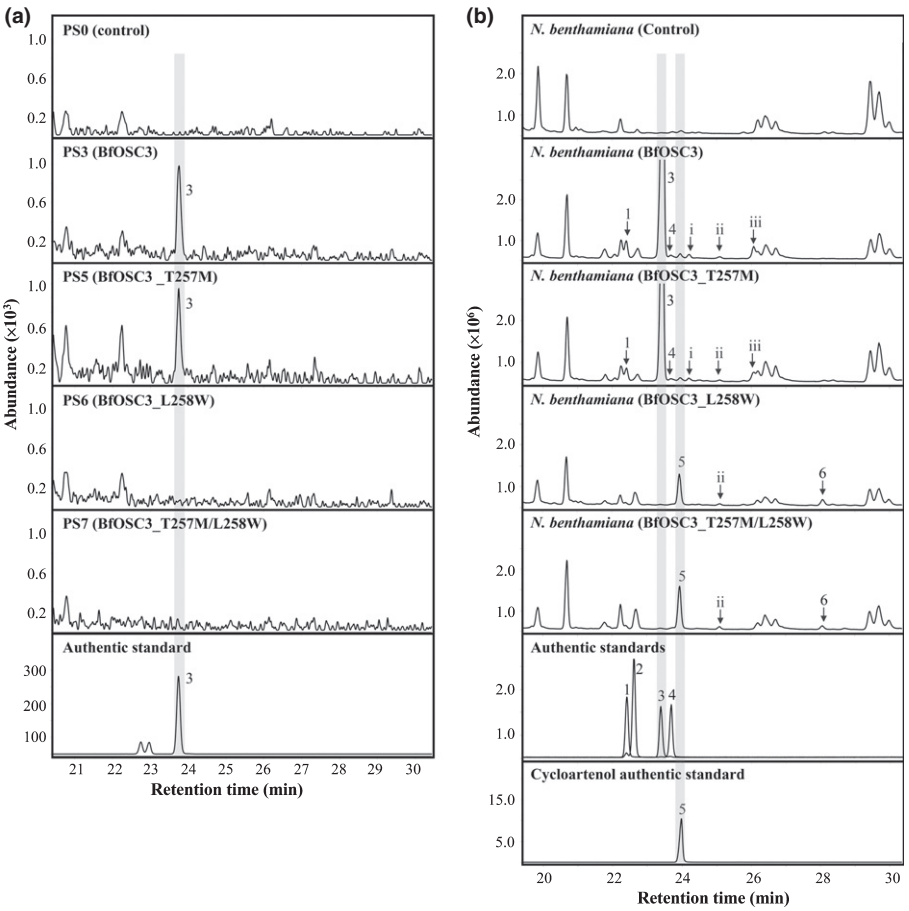

Fig. 7 Gas chromatograms of yeasts expressing BfOSC3 and its TLCYCR-motif mutants. (a) EIC ($m/z = 218$) of yeast expressing the wild-type (PS3) and mutants of BfOSC3 (PS5–PS7). In BfOSC3-expressing yeast (PS3), α-amyrin (3) was detected as a single product. In yeast expressing BfOSC3_T257M (PS5), α-amyrin (3) was detected similarly to PS3. In yeast expressing BfOSC3_L258W (PS6) and BfOSC3_T257M/L258W (PS7), no peak corresponding to triterpenol was detected. The SY100 yeast strain (pYES-DEST52 empty vector, PS0) was used as the background control. Data are representative of at least three biological replicates. (b) α-Amyrin (3) was detected in the leaf extract of *Nicotiana benthamiana* transiently expressing BfOSC3 and BfOSC3_T257M but not BfOSC3_L258W and BfOSC3_T257M/L258W. The peaks corresponding to cycloartenol (5) and two minor peaks, putative damarenediol-II (6) and an unidentified compound (ii), were of increased area in leaves transiently expressing either BfOSC3_L258W or BfOSC3_T257M/L258W. Mass fragmentation patterns are shown in Supporting Information Fig. S5. Leaves transiently expressing p19 and AtHMGRcd_S577A were used as background controls. Data are representative of five biological replicates.

et al., 2007) and *Malus domestica* (apple; MdOSC4, produces germanicol : β-amyrin : lupeol at an 82 : 14 : 4 ratio) (Andre *et al.*, 2016). Although BfOSC3 catalyzed the conversion of 2,3-oxidosqualene to only α-amyrin, we were puzzled by the inconsistency between the accumulation of α-amyrin in *B. forficata* and the low productivity of BfOSC3 in yeast. An attempt to characterize BfOSC3 using an *in vivo E. coli* expression system failed to detect triterpenol products. However, using an *N. benthamiana* transient expression system, we found that the product profile of BfOSC3 comprised ≥95% α-amyrin. Therefore, the catalytic activity of BfOSC3 is suitable for plant expression systems. Codon optimization might enhance the α-amyrin–generating activity of BfOSC3 in yeast or other microbial expression systems.

Several OSCs that produce β-amyrin or lupeol as a single product have been reported (Hayashi *et al.*, 2001; Zhang *et al.*, 2003; Kajikawa *et al.*, 2005); however, no OSC that produces α-amyrin as a single product has been found (Andre *et al.*, 2016; Yu *et al.*, 2018), and may not exist in nature (Saimaru *et al.*, 2007). To date, only MdOSC1 from *M. domestica*, CrAS from *Catharanthus roseus*, and IaAS1 from *Ilax asprella* have a product profile comprising ≥80% α-amyrin (Brendolise *et al.*, 2011; Huang *et al.*, 2012; Zheng *et al.*, 2015). The biosynthesis of α-amyrin and β-amyrin diverges after formation of the oleanyl cation (Ito *et al.*, 2013a). β-Amyrin is formed after the hydride arrangement from C-18 to C-12, whereas the methyl group at the C-29 position shifts to C-19 and is subsequently subjected to hydride arrangement, resulting in the generation of α-amyrin (Fig. S7; Ito *et al.*, 2013a). Homology modeling, domain swapping and site-directed mutagenesis using yeast expression systems have been conducted to assess the function and cyclization mechanism of OSCs (Kushiro *et al.*, 1999, 2000; Segura *et al.*, 2003; Chang *et al.*, 2013; Ito *et al.*, 2013a, 2014, 2017). However, the branch point and the amino acids essential for β-amyrin and α-amyrin biosynthesis are unknown.

To identify the amino acids essential for the α-amyrin synthase activity of BfOSC3, we performed site-directed mutagenesis of the MXCYXR motif, a determinant of the product selectivity of OSCs (Kushiro *et al.*, 2000). We found that L258 residue on the TLCYCR motif was crucial for BfOSC3 α-amyrin synthase activity. We further performed homology-based structural modeling and molecular docking analyses (Methods S3) to investigate the cyclization mechanism. In our model, the TLCYCR motif was located near the catalytic pocket, but we could not observe its direct interaction with α-amyrin (Fig. S8). Therefore, we could not assess the cyclization mechanism, most likely because the model was based on the human lanosterol synthase (HsLAS, 1W6K), which only has 34% similarity to BfOSC3.

Our findings provide insights into triterpenoid metabolites in the non-model legume *B. forficata*. In addition, using a suitable heterologous expression platform, we discovered four OSCs and identified the amino acid residues involved in the α-amyrin

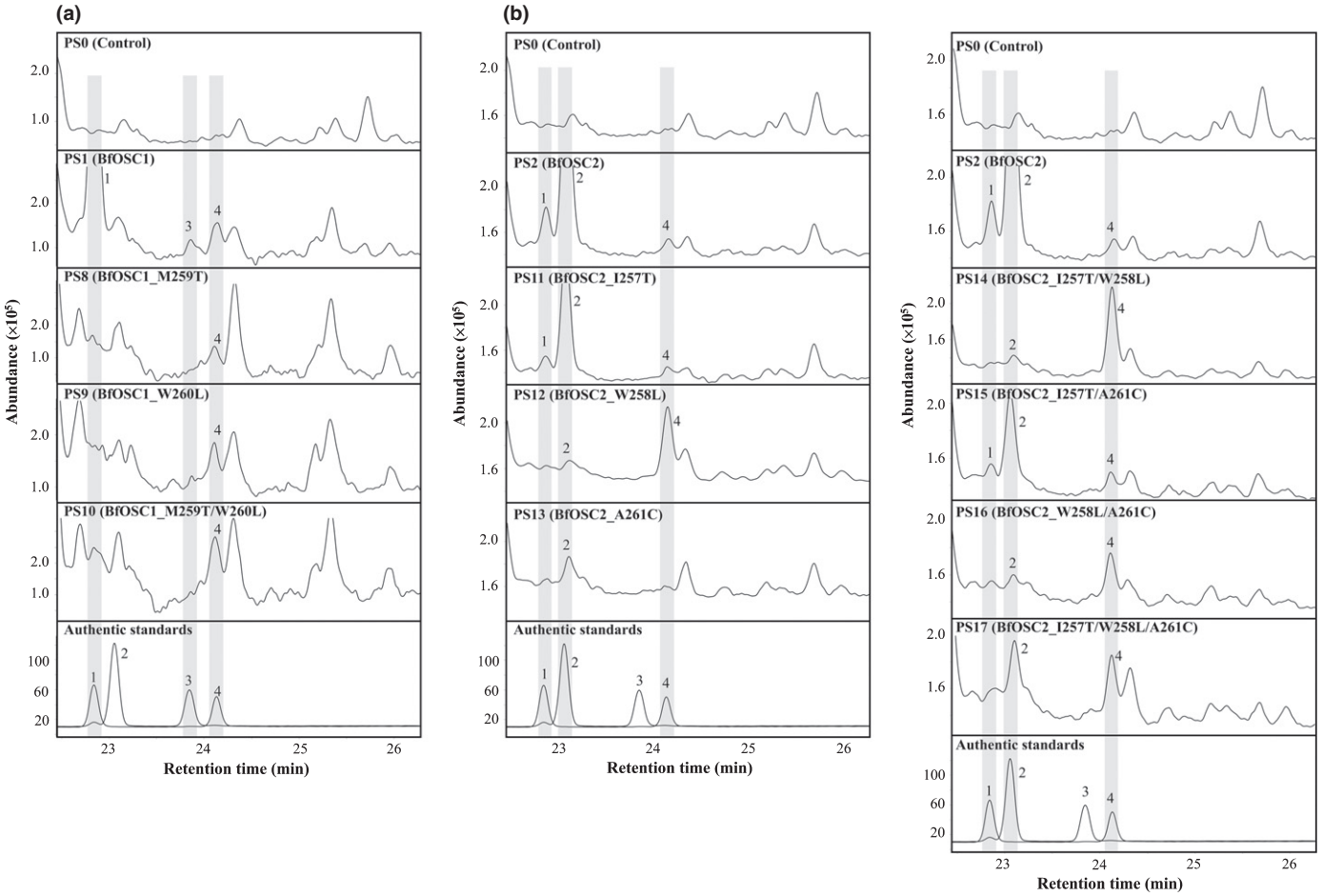

Fig. 8 Total ion chromatograms of heterologous host expressing BfOSC1, BfOSC2, and mutants thereof in the MXCYXR motif. (a) In the chromatogram of BfOSC1-expressing yeast (PS1), β-amyrin (1) was detected as the major product, and α-amyrin (3) and lupeol (4) were detected in trace amounts. In the chromatograms of yeasts expressing the BfOSC1 mutants (PS8–PS10), BfOSC1_M259T (PS8), BfOSC1_W260L (PS9), and BfOSC1_M259T/W260L (PS10), β-amyrin (1) and α-amyrin (3) were not detected, and lupeol (4) was detected at a low concentration. Data are representative of at least three biological replicates. (b) In the chromatogram of BfOSC2-expressing yeast (PS2), germanicol (2) was detected as the major product, and β-amyrin (1) and lupeol (4) as minor products. The product profiles of the yeasts expressing BfOSC2_I257T (PS11) and BfOSC2_I257T/A261C (PS15) were identical to that of PS2. In the chromatogram of yeast expressing BfOSC2_W258L (PS12), lupeol (4) was detected as the major product and germanicol (2) as a minor product. In the chromatogram of yeast expressing BfOSC2_A261C (PS13), a trace amount of germanicol (2) was detected, whereas β-amyrin (1) and lupeol (4) were not detected. In the chromatograms of yeasts expressing BfOSC2_I257T/W258L (PS14) and BfOSC2_W258L/A261C (PS16), lupeol (4) was detected as the major product and germanicol (2) as a minor product. In the chromatogram of yeast expressing BfOSC2_I257T/W258L/A261C (PS17), germanicol (2) and lupeol (4) were detected at equal concentrations and β-amyrin (1) was not detected. Data are representative of at least three biological replicates.

synthase activity of a newly discovered OSC (BfOSC3). Although the mechanism of α-amyrin formation remains unclear, our results will facilitate the engineering of OSCs with the desired product profile. In addition, an understanding of the product selectivity of OSCs will enable the design of efficient OSCs for commercial applications.

Acknowledgements

We thank Dr Matthew B. Johnson (Desert Legume Program, The University of Arizona) for *B. forficata* seeds. We thank the Leaf Tobacco Research Center, Japan Tobacco Inc., for *N. bentamiana* seeds. We thank Mr. Jezer Sauceda Galves for germinating the *B. forficata* seeds. We also thank Mr Jutapat Romsuk and Mr Much Zaenal Fanani for helpful discussion. This study was financially supported in part by the Grant-in-Aid for Scientific Research no. JP15H04485 awarded to TM and JP17H05442 awarded to HS; the Frontier Research Base for Global Young Researchers, Osaka University, supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) to EOF; and the Monbukagakusho Scholarship from MEXT awarded to PS.

Author contributions

PS, EOF, SY, JR, HSeki and TM designed the experiments. PS and SY conducted the experiments. HSuzuki processed the RNA-seq data. PS and EOF analyzed the results. TM, HSeki and

EOF conceived and supervised the study. PS wrote the manuscript. EOF, HSeki, SY, JR, HSuzuki and TM made manuscript revisions.

ORCID

Ery Odette Fukushima https://orcid.org/0000-0003-4099-4292
Toshiya Muranaka https://orcid.org/0000-0003-1058-2473

References

Andre CM, Legay S, Deleruelle A, Nieuwenhuizen N, Punter M, Brendolise C, Cooney JM, Lateur M, Hausman JF, Larondelle Y *et al.* 2016. Multifunctional oxidosqualene cyclases and cytochrome P450 involved in the biosynthesis of apple fruit triterpenic acids. *New Phytologist* **211**: 1279–1294.

Athikomkulchai S, Ruangrungsi N, Sekine T, Sumino M, Igarashi K, Ikegami F. 2003. Chemical constituents of *Bauhinia sirindhorniae*. *Natural Medicines* **57**: 150–153.

Augustin JM, Kuzina V, Andersen SB, Bak S. 2011. Molecular activities, biosynthesis and evolution of triterpenoid saponins. *Phytochemistry* **72**: 435–457.

Barton DHR, Brooks CJW. 1950. Morolic acid, a triterpenoid sapogenin. *Journal of the American Chemical Society* **72**: 3314.

Basyuni M, Oku H, Tsujimoto E, Kinjo K, Baba S, Takara K. 2007. Triterpene synthases from the Okinawan mangrove tribe, Rhizophoraceae. *FEBS Journal* **274**: 5028–5042.

Brendolise C, Yauk YK, Eberhard ED, Wang M, Chagne D, Andre C, Greenwood DR, Beuning LL. 2011. An unusual plant triterpene synthase with predominant α-amyrin- producing activity identified by characterizing oxidosqualene cyclases from *Malus* × *domestica*. *FEBS Journal* **278**: 2485–2499.

Cechinel-Filho V. 2009. Chemical composition and biological potential of plant from the genus *Bauhinia*. *Phytotherapy Research* **23**: 1347–1354.

Chang CH, Wen HY, Shie WS, Lu CT, Li ME, Liu YT, Li WH, Wu TK. 2013. Protein engineering of oxidosqualene-lanosterol cyclase into triterpene monocyclase. *Organic and Biomolecular Chemistry* **11**: 4214–4219.

da Cunha AM, Menon S, Menon R, Couta AG, Bürger C, Biavatti MW. 2010. Hypoglycemic activity of dried extracts of *Bauhinia forficata* Link. *Phytomedicine* **17**: 37–41.

De Souza P, Mota L, Boeing T, Somensi LB, Cecconi C, Zanchett C, Campos A, De Medeiros C, Krueger A, Bastos JK *et al.* 2017. Influence of prostanoids in the diuretic and natriuretic effects of extracts and kaempferitrin from *Bauhinia forficata* Link Leaves in Rats. *Phytotherapy Research* **31**: 1521–1528.

Dixon RA, Sumner LW. 2003. Legume natural products: understanding and manipulating complex pathways for human and animal health. *Plant Physiology* **131**: 878–885.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792–1797.

Forestier E, Romero-segura C, Pateraki I, Centeno E, Compagnon V, Preiss M, Berna A, Boronat A, Bach TJ, Darnet S *et al.* 2019. Distinct triterpene synthases in the laticifers of *Euphorbia lathyris*. *Scientific Reports* **9**: 4840.

Francis G, Kerem Z, Makkar HPS, Becker K. 2002. The biological action of saponins in animal systems: a review. *British Journal of Nutrition* **88**: 587–605.

González-Coloma A, López-Balboa C, Santana O, Reina M, Fraga BM. 2011. Triterpene-based plant defenses. *Phytochemistry Reviews* **10**: 245–260.

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

Hayashi H, Huang P, Kirakosyan A, Inoue K, Hiraoka N, Ikeshiro Y, Kushiro T, Shibuya M, Ebizuka Y. 2001. Cloning and characterization of a cDNA encoding β-Amyrin synthase involved in glycyrrhizin and soyasaponin biosyntheses in licorice. *Biological and Pharmaceutical Bulletin* **24**: 912–916.

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

Ito R, Hashimoto I, Masukawa Y, Hoshino T. 2013a. Effect of cation–π interactions and steric bulk on the catalytic action of oxidosqualene cyclase: a case study of Phe728 of β-amyrin synthase from *Euphorbia tirucalli* L. *Chemistry – A European Journal* **19**: 17150–17158.

Ito R, Masukawa Y, Hoshino T. 2013b. Purification, kinetics, inhibitors and CD for recombinant β-amyrin synthase from *Euphorbia tirucalli* L and functional analysis of the DCTA motif, which is highly conserved among oxidosqualene cyclases. *FEBS Journal* **280**: 1267–1280.

Ito R, Masukawa Y, Nakada C, Amari K, Nakano C, Hoshino T. 2014. β-Amyrin synthase from *Euphorbia tirucalli*. Steric bulk, not the π-electrons of Phe, at position 474 has a key role in affording the correct folding of the substrate to complete the normal polycyclization cascade. *Organic and Biomolecular Chemistry* **12**: 3836–3846.

Ito R, Nakada C, Hoshino T. 2017. β-Amyrin synthase from *Euphorbia tirucalli* L. functional analyses of the highly conserved aromatic residues Phe413, Tyr259 and Trp257 disclose the importance of the appropriate steric bulk, and cation–π and CH–π interactions for the efficient catalytic action of the polyolefin cyclization cascade. *Organic and Biomolecular Chemistry* **15**: 177–188.

Kajikawa M, Yamato KT, Fukuzawa H, Sakai Y, Uchida H, Ohyama K. 2005. Cloning and characterization of a cDNA encoding β-amyrin synthase from petroleum plant *Euphorbia tirucalli* L. *Phytochemistry* **66**: 1759–1766.

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

Krishnamurthy P, Tsukamoto C, Takahashi Y, Hongo Y, Singh RJ, Lee JD, Chung G. 2014. Comparison of saponin composition and content in wild soybean (*Glycine soja* Sieb. and Zucc.) before and after germination. *Bioscience, Biotechnology and Biochemistry* **78**: 1988–1996.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33**: 1870–1874.

Kushiro T, Shibuya M, Ebizuka Y. 1998. β-Amyrin synthase. Cloning of oxidosqualene cyclase that catalyzes the formation of the most popular triterpene among higher plants. *European Journal of Biochemistry* **256**: 238–244.

Kushiro T, Shibuya M, Ebizuka Y. 1999. Chimeric triterpene synthase. A possible model for multifunctional triterpene synthase. *Journal of the American Chemical Society* **121**: 1208–1216.

Kushiro T, Shibuya M, Masuda K, Ebizuka Y. 2000. Mutational studies on triterpene synthases: engineering lupeol synthase into β-amyrin synthase. *Journal of the American Chemical Society* **122**: 6816–6824.

LPWG. 2017. A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny. *Taxon* **66**: 44–77.

Moses T, Papadopoulou KK, Osbourn A. 2014. Metabolic and functional diversity of saponins, biosynthetic intermediates and semi-synthetic derivatives. *Critical Reviews in Biochemistry and Molecular Biology* **49**: 439–462.

Moses T, Pollier J, Almagro L, Buyst D, Van Montagu M, Pedreño MA, Martins JC, Thevelein JM, Goossens A. 2013a. Combinatorial biosynthesis of sapogenins and saponins in *Saccharomyces cerevisiae* using a C-16α hydroxylase from *Bupleurum falcatum*. *Proceedings of the National Academy of Sciences, USA* **111**: 1634–1639.

Moses T, Pollier J, Thevelein JM, Goossens A. 2013b. Bioengineering of plant (tri)terpenoids: from metabolic engineering of plants to synthetic biology *in vivo* and *in vitro*. *New Phytologist* **200**: 27–43.

Muangphrom P, Seki H, Suzuki M, Komori A, Nishiwaki M, Mikawa R, Fukushima EO, Muranaka T. 2016. Functional analysis of amorpha-4,11-diene synthase (ADS) homologs from non-artemisinin-producing artemisia species: the discovery of novel koidzumiol and (+)-α-bisabolol synthases. *Plant and Cell Physiology* **57**: 1678–1688.

Phillips DR, Rasbery JM, Bartel B, Matsuda SPT. 2006. Biosynthetic diversity in plant triterpene cyclization. *Current Opinion in Plant Biology* **9**: 305–314.

Pollier J, Morreel K, Geelen D, Goossens A. 2011. Metabolite profiling of triterpene saponins in *Medicago truncatula* hairy roots by liquid chromatography fourier transform ion cyclotron resonance mass spectrometry. *Journal of Natural Products* **74**: 1462–1476.

Poralla K. 1994. The possible role of a repetitive amino acid motif in evolution of triterpenoid cyclases. *Bioorganic and Medicinal Chemistry Letters* **4**: 285–290.

Pütter KM, Van Deenen N, Unland K, Prüfer D, Gronover CS. 2017. Isoprenoid biosynthesis in dandelion latex is enhanced by the overexpression of three key enzymes involved in mevalonate pathway. *BMC Plant Biology* **17**: 1–13.

Rao YK, Fang SH, Tzeng YM. 2008. Antiinflammatory activities of flavonoids and a triterpene caffeate isolated from *Bauhinia variegata*. *Phytotherapy Research* **22**: 957–962.

Reed J, Stephenson MJ, Miettinen K, Brouwer B, Leveau A, Brett P, Goss RJM, Goossens A, O'Connell MA, Osbourn A. 2017. A translational synthetic biology platform for rapid access to gram-scale quantities of novel drug-like molecules. *Metabolic Engineering* **42**: 185–193.

Robertlee J, Kobayashi K, Suzuki M, Muranaka T. 2017. AKIN10, a representative *Arabidopsis* SNF1-related protein kinase 1 (SnRK1), phosphorylates and downregulates plant HMG-CoA reductase. *FEBS Letters* **591**: 1159–1166.

Robertlee J, Kobayashi K, Tang J, Suzuki M, Muranaka T. 2018. Evidence that the *Arabidopsis thaliana* 3-hydroxy-3-methylglutaryl-CoA reductase 1 is phosphorylated at Ser577 *in planta*. *Plant Biotechnology* **35**: 1–7.

Saimaru H, Orihara Y, Tansakul P, Kang YH, Shibuya M, Ebizuka Y. 2007. Production of triterpene acids by cell-suspension cultures of *Olea europaea*. *Chemical and Pharmaceutical Bulletin* **55**: 784–788.

Salmon M, Thimmappa RB, Minto RE, Melton RE, Hughes RK, O'Maille PE, Hemmings AM, Osbourn A. 2016. A conserved amino acid residue critical for product and substrate specificity in plant triterpene synthases. *Proceedings of the National Academy of Sciences* **113**: E4407–E4414.

Sawai S, Saito K. 2011. Triterpenoid biosynthesis and engineering in plants. *Frontiers in Plant Science* **2**: 1–8.

Sawai S, Shindo T, Sato S, Kaneko T, Tabata S, Ayabe S, Aoki T. 2006. Functional and structural analysis of genes encoding oxidosqualene cyclases of *Lotus japonicus*. *Plant Science* **170**: 247–257.

Segura MJR, Jackson BE, Matsuda SPT. 2003. Mutagenesis approaches to deduce structure–function relationships in terpene synthases. *Natural Product Reports* **20**: 304–317.

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

Sinou C, Forest F, Lewis GP, Bruneau A. 2009. The genus *Bauhinia* s.l. (Leguminosae): a phylogeny based on the plastid *trn*L–*trn* F region. *Botany-Botanique* **87**: 947–960.

de Sousa E, Zanatta L, Seifriz I, Creczynski-Pasa TB, Pizzolatti MG, Szpoganicz B, Silva FRMB. 2004. Hypoglycemic effect and antioxidant potential of kaempferol-3,7-O-(α)-dirhamnoside from *Bauhinia forficata* leaves. *Journal of Natural Products* **67**: 829–832.

Suzuki M, Xiang T, Ohyama K, Seki H, Saito K, Muranaka T, Hayashi H, Katsube Y, Kushiro T, Shibuya M *et al.* 2006. Lanosterol synthase in dicotyledonous plants. *Plant and Cell Physiology* **47**: 565–571.

Tamura K, Seki H, Suzuki H, Kojoma M, Saito K, Muranaka T. 2017. CYP716A179 functions as a triterpene C-28 oxidase in tissue-cultured stolons of *Glycyrrhiza uralensis*. *Plant Cell Reports* **36**: 437–445.

Thimmappa R, Geisler K, Louveau T, O'Maille P, Osbourn A. 2014. Triterpene biosynthesis in plants. *Annual Review of Plant Biology* **65**: 225–257.

Vranová E, Coman D, Gruissem W. 2013. Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annual Review of Plant Biology* **64**: 665–700.

Wang Z, Guhling O, Yao R, Li F, Yeats TH, Rose JKC, Jetter R. 2011. Two oxidosqualene cyclases responsible for biosynthesis of tomato fruit cuticular triterpenoids. *Plant Physiology* **155**: 540–552.

Xu R, Fazio GC, Matsuda SPT. 2004. On the origins of triterpenoid skeletal diversity. *Phytochemistry* **65**: 261–291.

Xue Z, Duan L, Liu D, Guo J, Ge S, Dicks J, ÓMáille P, Osbourn A, Qi X. 2012. Divergent evolution of oxidosqualene cyclases in plants. *New Phytologist* **193**: 1022–1038.

Yu Y, Chang P, Yu H, Ren H, Hong D, Li Z, Wang Y, Song H, Huo Y, Li C. 2018. Productive amyrin synthases for efficient α-amyrin synthesis in engineered *Saccharomyces cerevisiae*. *ACS Synthetic Biology* **7**: 2391–2402.

Zhang H, Shibuya M, Yokota S, Ebizuka Y. 2003. Oxidosqualene cyclases from cell suspension cultures of *Betula platyphylla* var. *japonica*: molecular evolution of oxidosqualene cyclases in higher plants. *Biological and Pharmaceutical Bulletin* **26**: 642–650.

Zheng X, Luo X, Ye G, Chen Y, Ji X, Wen L, Xu Y, Xu H, Zhan R, Chen W. 2015. Characterisation of two oxidosqualene cyclases responsible for triterpenoid biosynthesis in *Ilex asprella*. *International Journal of Molecular Sciences* **16**: 3564–3578.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Photographs of a *B. forficata* flower and leaf.

Fig. S2 Total ion chromatogram of the major pentacyclic triterpenoids detected in *B. forficata*.

Fig. S3 Total ion chromatogram (zoomed-in view) of yeast strains expressing BfOSCs (PS1–PS4).

Fig. S4 Mass fragmentation patterns of the major products of yeast strains expressing BfOSCs (PS1–PS4).

Fig. S5 Mass fragmentation patterns of the products detected in *N. benthamiana* transiently expressing BfOSC3.

Fig. S6 Gas chromatography–mass spectrometry analyses of *E. coli* expressing BfOSC4.

Fig. S7 Proposed biosynthetic pathway of the triterpenols detected in this study.

Fig. S8 Homology-based structural modeling and molecular-docking analyses of BfOSC3 with α-amyrin.

Methods S1 Construction of the pYS_015 binary vector.

Methods S2 Construction of the *E. coli* expression system.

Methods S3 Homology-modeling and molecular docking.

Table S1 List of primers used in this study.

Table S2 List of yeast strains used in this study.

Table S3 Pentacyclic triterpenoids detected in *B. forficata* extracts after acid hydrolysis.

Table S4 Summary of RNA sequencing analysis of *B. forficata* leaves.

Table S5 Summary of gene contigs related to the mevalonate biosynthesis pathway.

Table S6 Amino acid sequence identity: comparison among four OSCs cloned from *B. forficata*.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.