CellPress
PARTNER JOURNAL

Molecular Plant
Research Article

Unraveling the Triterpenoid Saponin Biosynthesis of the African Shrub *Maesa lanceolata*

Tessa Moses[1,2,3,4,8,7], Jacob Pollier[1,2,7], Ahmad Faizal[5,9], Sandra Apers[6], Luc Pieters[6], Johan M. Thevelein[3,4], Danny Geelen[5] and Alain Goossens[1,2,*]

[1]Department of Plant Systems Biology, VIB, Technologiepark 927, B-9052 Gent, Belgium

[2]Department of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 927, B-9052 Gent, Belgium

[3]Department of Molecular Microbiology, VIB, Kasteelpark Arenberg 31, B-3001 Leuven-Heverlee, Belgium

[4]Laboratory of Molecular Cell Biology, Institute of Botany and Microbiology, KU Leuven, Kasteelpark Arenberg 31, B-3001 Leuven-Heverlee, Belgium

[5]Department of Plant Production, Faculty of Bioscience Engineering, Ghent University, Coupure Links 653, 9000 Ghent, Belgium

[6]Laboratory of Pharmacognosy and Pharmaceutical Analysis, Department of Pharmaceutical Sciences, University of Antwerp, Universiteitsplein 1, B-2610 Antwerp, Belgium

[7]These authors contributed equally to the article.

[8]Present address: John Innes Centre, Norwich Research Park, Colney Lane, Norwich NR4 7UH, UK

[9]Present address: School of Life Sciences & Technology, Institut Teknologi Bandung, Jl. Ganesa No. 10, Bandung 40132, Indonesia

***Correspondence: Alain Goossens (algoo@psb.vib-ugent.be)**

http://dx.doi.org/10.1016/j.molp.2014.11.004

ABSTRACT

Maesasaponins produced by the African shrub *Maesa lanceolata* are oleanane-type saponins with diverse biological activities. Through a combination of transcript profiling of methyl jasmonate-elicited *M. lanceolata* shoot cultures, functional analysis in transgenic *M. lanceolata* plants and the heterologous hosts *Medicago truncatula* and *Saccharomyces cerevisiae*, we identified three maesaponin biosynthesis genes. These include a β-amyrin synthase and two cytochrome P450s, CYP716A75 and CYP87D16, which catalyze the C-28 and C-16α oxidations of β-amyrin, respectively.

Key words: cytochrome P450, oxidosqualene cyclase, C-28 oxidase, C-16α oxidase, triterpenoid saponin, *Maesa lanceolata*

Moses T., Pollier J., Faizal A., Apers S., Pieters L., Thevelein J.M., Geelen D., and Goossens A. (2015). Unraveling the Triterpenoid Saponin Biosynthesis of the African Shrub *Maesa lanceolata*. Mol. Plant. **8**, 122–135.

INTRODUCTION

Triterpenoid saponins are a large class of structurally diverse and biologically active specialized metabolites produced by numerous plant species. These amphipathic glycosides consist of a hydrophobic backbone or sapogenin with one or more hydrophilic sugar chains attached to it. Triterpenoid saponins display a wide range of commercial applications in the agricultural, food, cosmetic, and pharmaceutical sectors as pesticides, preservatives, surfactants, adjuvants, antimicrobial, anti-inflammatory, and anticancer agents (Augustin et al., 2011; Osbourn et al., 2011; Pollier et al., 2011c; Sparg et al., 2004; Vincken et al., 2007). Nonetheless, their applications are still limited, mainly because they occur as complex product mixtures from which isolation of a pure compound is challenging, or because of the low *in planta* production level for many of them. One way of overcoming the latter limitation is through the engineering of plants, cell cultures, and/or microorganisms to obtain an increased production rate (Moses et al., 2013). A thorough understanding of the biosynthetic pathways and regulatory mechanisms involved in target compound synthesis in the native host is a prerequisite for engineering organisms for enhanced production. However, for most triterpenoid saponins, such comprehensive information is currently lacking.

Like sterols, the triterpenoid saponins are biosynthesized from isopentenyl pyrophosphate (IPP) generated via the 3-hydroxy-3-methylglutaryl CoA reductase (HMGR)-dependent mevalonate pathway. The head-to-tail fusion of two C_5 IPP units with one C_5 dimethylallyl pyrophosphate unit results in the formation of C_{15} farnesyl pyrophosphate, two of which condense head-to-head to form C_{30} squalene, the precursor of all triterpenoids in eukaryotes. The linear squalene is subsequently epoxidized to 2,3-oxidosqualene, the last common intermediate in triterpenoid saponin and sterol biosynthesis. The two pathways diverge after 2,3-oxidosqualene, which is cyclized by specific oxidosqualene

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and IPPE, SIBS, CAS.

cyclases (OSCs) to either cycloartenol, the plant sterol precursor, or a wide variety of specific triterpenoid backbones. In nature, more than 100 distinct triterpenoid skeletons have been identified; the most common ones are β-amyrin (oleanane), α-amyrin (ursane), lupeol (lupane), and dammarenediol (dammarane) (Xu et al., 2004; Xue et al., 2012). After cyclization, the triterpenoids are decorated by multiple cytochrome P450-dependent monooxygenases (P450s) to generate a diverse array of sapogenins containing reactive functional groups. To date, 22 triterpenoid-modifying P450s that catalyze different reactions on 16 different carbon positions have been reported from different plant species (Han et al., 2013; Moses et al., 2014b; Thimmappa et al., 2014). Ultimately, the sapogenins are glycosylated by UDP-dependent glycosyltransferases (UGTs) and may be further decorated by other "tailoring" enzymes, such as acetyltransferases and malonyltransferases, to generate the biologically active saponins (Osbourn et al., 2011; Thimmappa et al., 2014).

In plants, saponins serve as defense molecules and their production is often triggered by biotic and abiotic stress. The concerted transcriptional response of the entire saponin biosynthetic pathway to different stresses is controlled by a complex signaling cascade in which the jasmonate hormones play a crucial role (Pauwels et al., 2009; De Geyter et al., 2012). As such, treatment of whole plants or plant cell cultures with jasmonates leads to the transcriptional activation of saponin biosynthetic genes, and transcript profiling of jasmonate-elicited plants or cell cultures can be used as a tool for identifying candidate saponin biosynthesis genes. The identified candidates can consequently be expressed in heterologous systems for functional characterization. Likewise, loss of function of a biosynthetic enzyme can lead to the lack of accumulation of the resulting saponins. For instance, loss of function of CYP716A12, the *Medicago truncatula* P450 that catalyzes the three-step oxidation of β-amyrin to oleanolic acid, blocks the synthesis of hemolytic saponins in the mutant called *lacking hemolytic activity* (*lha*) (Carelli et al., 2011). Therefore, a second way of characterizing gene function is by using gene knockdown and knock-out strategies for candidate biosynthesis genes in the native plant host, followed by analysis of the saponin spectrum.

Maesa lanceolata, a member of the Myrsinaceae family, is a shrub or small tree indigenous to eastern, southern, and central Africa. African traditional healers use parts and extracts of the plant for the treatment of a wide range of diseases, including infectious hepatitis, bacillary dysentery, impetigo, ozena, dermatoses, and neuropathies. Methanol extracts of *M. lanceolata* leaves are rich in triterpenoid saponins, called maesasaponins, which have been shown to possess virucidal, molluscicidal, fungistatic, and antimutagenic activities (Sindambiwe et al., 1998). The maesasaponins identified from *M. lanceolata* are of the oleanane type, derived from multiple modifications of the β-amyrin backbone. Many contain characteristic C-13,28 hemiacetal or ester bridges, and display oxidations on the C-16, C-21, and/or C-22 positions (Figure 1A and 1B) (Foubert et al., 2010; Manguro et al., 2011). The hemiacetal and ester bridges are thought to result from the reaction between a C-13 hydroxyl and C-28 aldehyde or carboxyl group, respectively (Vincken et al., 2007). The occurrence of these diverse saponins in *M. lanceolata* implies the presence of OSCs and

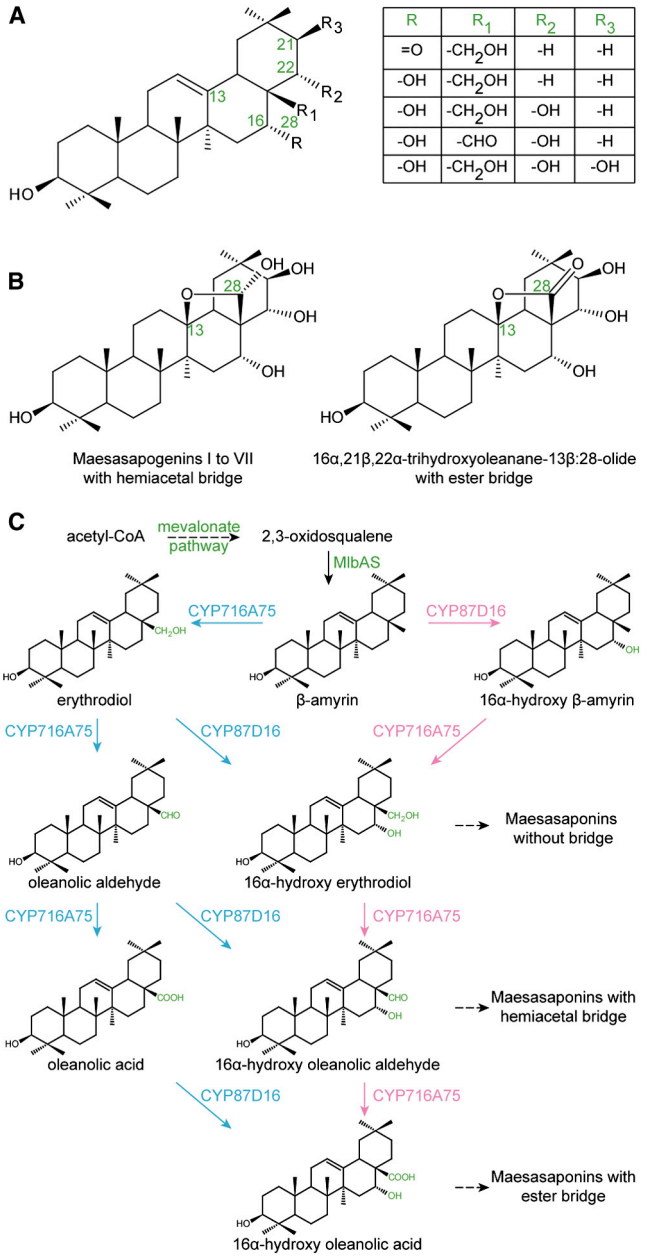

Figure 1. Chemical Structures of Maesasaponins.

(A) Maesasaponins are oleanane-type triterpenoids with oxidations on C-16, C-21, C-22, and/or C-28 positions (numbered). The possible combinations of R groups in the *M. lanceolata* triterpenoids are indicated in the adjoining table.

(B) Maesasapogenins with hemiacetal (left) and ester (right) bridge between C-13 and C-28.

(C) A simplified overview of *M. lanceolata* triterpenoid saponin biosynthesis. MlbAS is a β-amyrin synthase, CYP87D16 is a C-16α hydroxylase, and CYP716A75 is a C-28 oxidase involved in maesasaponin biosynthesis. Dashed arrows indicate multiple steps. Two alternate pathways for the synthesis of maesasaponins are proposed with the enzymes and arrows highlighted in pink and blue. Oxidations resulting from the activity of the P450s are highlighted in green.

P450s in the plant, which cyclize 2,3-oxidosqualene to β-amyrin, which is further oxidized at its C-13, C-16, C-21, C-22, and/or C-28 positions.

Several β-amyrin synthases (bASs) have been functionally characterized using *Saccharomyces cerevisiae*, including bAS from *Artemisia annua* (Kirby et al., 2008), *M. truncatula* (Suzuki et al., 2002), *Glycyrrhiza glabra* (Hayashi et al., 2001), and *Saponaria vaccaria* (Meesapyodsuk et al., 2007). Similarly, P450s catalyzing C-28 oxidation (Carelli et al., 2011; Fukushima et al., 2011; Huang et al., 2012; Han et al., 2013), C-16α hydroxylation (Moses et al., 2014b), C-12,13β-epoxidation together with C-16β hydroxylation (Kunii et al., 2012; Geisler et al., 2013), and C-22 hydroxylation (Seki et al., 2011) of triterpenoid backbones have been characterized using yeast strains.

In this study, we performed cDNA-amplified fragment length polymorphism (cDNA-AFLP)-based transcript profiling on methyl jasmonate (MeJA)-treated *M. lanceolata* shoot cultures to identify candidate maesasaponin biosynthesis genes. The functional activity of the candidate genes was screened for by heterologous expression in a sterol-engineered *S. cerevisiae* strain. This led to the identification of one OSC that was characterized as a bAS of which the role in maesasaponin biosynthesis was confirmed by overexpression and silencing strategies in transgenic *M. lanceolata* plants. Furthermore, we identified two P450s that, together with the characterized bAS, catalyze three enzymatic steps involved in maesasaponin biosynthesis (Figure 1C).

RESULTS

Transcript Profiling of *M. lanceolata* Shoots Reveals Candidate Saponin Biosynthesis Genes

Jasmonate inducibility of maesasaponin production in *M. lanceolata* has not yet been reported. Maesasaponins are produced in the leaves of *M. lanceolata* plants from indigenous collection sites. To assess MeJA inducibility of maesasaponin synthesis, *M. lanceolata* axenic shoot cultures were generated from this material, cultivated in liquid cultures, and elicited with MeJA. No unambiguous effect of MeJA (or any other phytohormone) on maesasaponin accumulation in these explants could be observed, but the addition of MeJA (but none of the other phytohormones) led to excessive foaming, which might be indicative of either MeJA-induced secretion, or increased production of maesasaponins (Faizal, 2013). Considering this, and taking into account the conserved transcriptional response of entire saponin biosynthetic pathways across the plant kingdom to the jasmonate hormone (De Geyter et al., 2012), we therefore carried out transcriptome analysis of MeJA-elicited *M. lanceolata* explants.

cDNA-AFLP-based transcript profiling was performed on a time-course sample set of MeJA-treated *M. lanceolata* axenic shoot cultures. The expression of 13 558 transcript tags was monitored, and 733 MeJA-responsive tags were isolated (hereafter called ML tags), reamplified, and sequenced. Good-quality sequences were obtained for 545 (74.4%) ML tags; no unique sequence could be attributed unambiguously to the remaining 188 (25.6%) tags. Of the 545 unique ML tags, 312 (57.2%) were tentatively annotated following BLAST analysis. An average linkage hierarchical clustering analysis showed that after MeJA treatment, the selected ML tags were either transcriptionally induced (clusters I–V, Supplemental Figure 1) or repressed (cluster VI, Supplemental Figure 1). Based on their MeJA response time, the induced gene tags were divided into five subclusters. Genes in cluster V responded within 2 h of MeJA treatment, whereas genes in cluster IV reached maximum expression 4–8 h after MeJA treatment. Cluster II consists of genes that were induced 2 h after MeJA treatment, maintaining a high expression level thereafter. In clusters I and III, genes were activated 24 h after MeJA treatment.

ML tags corresponding to enzymes known to be involved in triterpenoid biosynthesis, such as acetoacetyl-CoA thiolase, HMGR, and squalene epoxidase, belong to cluster III and were transcriptionally activated 24 h post MeJA elicitation (Figure 2A). No tag corresponding to bAS, the first committed enzyme for saponin biosynthesis in *M. lanceolata*, was encountered in our transcript profiling analysis, whereas 16 ML tags corresponded to P450s, which represent the second set of enzymes involved in saponin biosynthesis. Of these 16 ML tags, three (ML449, ML265, and ML177) were transcriptionally repressed after MeJA treatment and were therefore not considered as candidates for saponin biosynthesis. For 8 of the remaining 13 tags, full-length (FL) coding sequences were obtained using an integrated PCR colony hybridization approach (Pollier et al., 2011a). Two tags (ML041 and ML104) with a nearly identical expression pattern corresponded to the same gene and, hence, seven candidate P450s were retained, all of which were transcriptionally activated within 24 h of MeJA treatment (Figure 2A). The sequences obtained were submitted to GenBank and the P450 naming committee, and an overview of the names and accession numbers is given in Supplemental Table 1.

Identification and Characterization of MlbAS, the First Committed Enzyme for Maesasaponin Biosynthesis

As no ML tag corresponding to bAS was encountered in the cDNA-AFLP transcript profiling analysis, we designed degenerate oligonucleotide primers based on characterized *bAS* genes from different plants, and amplified a gene tag corresponding to bAS using an *M. lanceolata* cDNA template. The sequence obtained was then used to screen for the FL open reading frame of the *M. lanceolata* β-amyrin synthase (*MlbAS*) in a cDNA library (Pollier et al., 2011a). The FL *MlbAS* encodes a putative protein of 760 amino acids with conserved SDCTAE and MWCYCR motifs, which are essential for substrate binding and β-amyrin product specificity, respectively. In addition, the MlbAS contains six QW repeats, which are imperative for the polycyclization reaction, and a key conserved Lys residue at position 449, which is conserved in all specific bASs (Supplemental Figure 2) (Poralla, 1994; Kushiro et al., 2000).

To check its functionality, we heterologously expressed the FL *MlbAS* gene in the sterol-engineered yeast strain TM1 (Supplemental Table 2) (Moses et al., 2014b). Using gas chromatography–mass spectrometry (GC–MS) analysis, organic extracts of cultures of the yeast strain TM4 harboring the plasmid pESC-URA[*GAL10/tHMG1*; *GAL1/MlbAS*] and the control strain TM5 harboring the plasmid pESC-URA [*GAL10/tHMG1*] were compared. A single, low-abundant peak

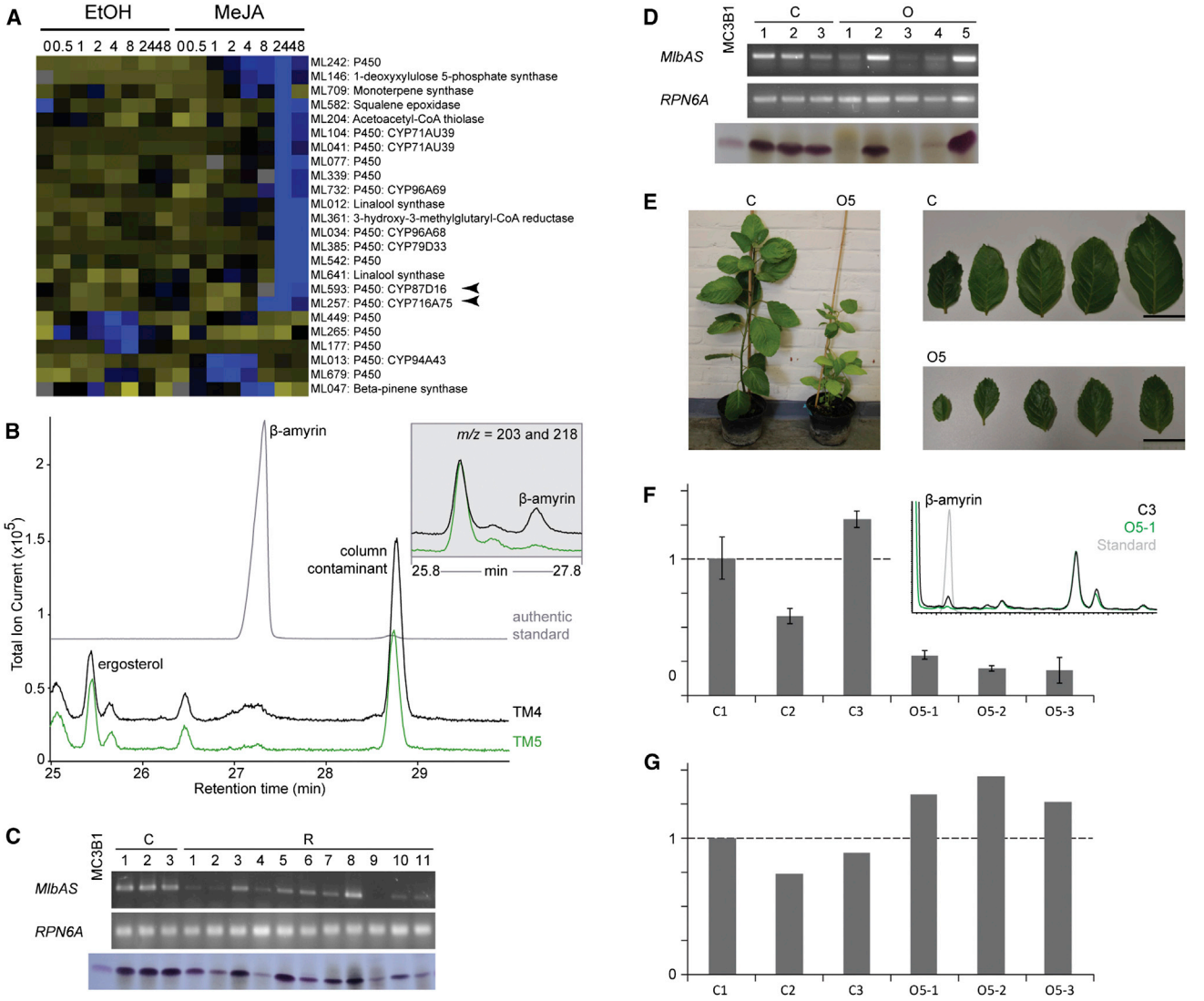

Figure 2. cDNA-AFLP-Based Transcript Profiling of MeJA-Induced *M. lanceolata* Shoots and Functional Characterization of MlbAS.

(A) Subcluster of the *M. lanceolata* transcriptome with all gene tags corresponding to P450s and terpenoid biosynthesis genes. Treatments and time points (in h) are indicated at the top. Blue and yellow boxes reflect transcriptional activation and repression relative to the average expression level, respectively. Gray boxes correspond to missing time points.

(B–E) MlbAS is a functional β-amyrin synthase. **(B)** Overlay of GC chromatograms from spent medium of yeast strain TM4 expressing *MlbAS* (black), control strain TM5 (green), and β-amyrin standard (gray). Inset shows the β-amyrin peak corresponding to the select ion monitoring for *m/z* 203 and 218, which represent the most abundant ions resulting from electron impact ionization of β-amyrin. **(C)** Expression analyzed by RT–PCR in leaves of wild-type control (C) and *MlbAS* hpRNAi transgenic (R) plants, for *MlbAS* (top) and control RPN6A (middle) genes, and their corresponding maesasaponin content (bottom) analyzed by TLC. **(D)** Expression analyzed by RT-PCR in leaves of wild-type control (C) and *MlbAS* overexpression transgenic (O) plants, for *MlbAS* (top) and control RPN6A (middle) genes, and their corresponding maesasaponin content (bottom). MC3B1 is an HPLC-purified maesasaponin mixture (Apers et al., 1998) and a 0.1% solution (w/v) was used as reference. **(E)** Phenotypic comparison between C and O (saponin overproducing line) plants of *M. lanceolata*. Full plants (left), comparison of leaf sizes between C (right, top) and O (right, bottom) plants. The size bar indicates 5 cm.

(F) Quantification of β-amyrin in control lines (C) and the transgenic *MlbAS* overexpressing line 5 (O). The amount of β-amyrin relative to the level in control line 1 is plotted. Error bars show ±SEM for $n = 3$. The inset shows the total ion chromatogram of control line 3 (C3, black), *MlbAS* overexpressing line 5 clone 1 (O5-1, green), and a β-amyrin standard (gray).

(G) Saponin quantification in control lines (C) and the transgenic *MlbAS* overexpressing line 5 (O). The amount of maesasaponins relative to the level in control line 1 is plotted.

at 27.2 min was observed to be unique to the GC chromatogram of TM4 (Figure 2B). Its corresponding electron ionization–mass spectrometry (EI–MS) pattern was similar to that of a β-amyrin standard, but consisted of only the most abundant *m/z* fragment ions 189, 203, and 218. A clear β-amyrin peak was observed in TM4 with select ion monitoring of *m/z* ions 203 and 218 (inset, Figure 2B). Although β-amyrin was detected in the yeast strain TM4, the 2,3-oxidosqualene cyclization efficiency of MlbAS in yeast was observed to be very low.

MlbAS Overexpression Leads to Growth and Developmental Defects in *M. lanceolata* Plants

To confirm the functionality of the MlbAS identified and its role in maesasaponin biosynthesis, we analyzed the effect of *MlbAS* overexpression and silencing *in planta*. Transgenic *M. lanceolata* plants constitutively expressing a hairpin RNA-mediated interference (hpRNAi) construct against *MlbAS* were generated by *Agrobacterium tumefaciens*-mediated transformation, and transcript levels were determined by RT–PCR. The maesasaponin accumulation in independent hpRNAi lines, with varying degrees of *MlbAS* suppression, was determined by qualitative thin-layer chromatography (TLC) analysis. We observed a clear correlation between reduced transcript levels and reduced maesasaponin accumulation *in planta* (Figure 2C), suggesting that *MlbAS* is indeed a key player involved in the biosynthesis of maesasaponins.

In addition to gene silencing, we analyzed the effect of *MlbAS* overexpression on the accumulation of maesasaponins in transgenic *M. lanceolata* plants. Analysis of transcript levels by RT–PCR revealed that *MlbAS* overexpression resulted in either higher or lower gene expression, the latter presumably resulting from cosuppression. Also in these transgenic plants, the accumulation of maesasaponins correlated with the elevated or reduced *MlbAS* transcript levels (Figure 2D). Furthermore, the *M. lanceolata* plants with elevated *MlbAS* transcript levels showed growth and morphologic defects when grown under greenhouse conditions. When compared with a control plant, the *MlbAS*-overexpressing plants exhibited a dwarf phenotype with small yellowish leaves (Figure 2E), presumably resulting from the altered *in planta* flux balances among triterpenoid-derived sterol, hormone, chlorophyll, and/or saponin biosynthesis. Similar phenotypic effects have been reported for several other plants for which saponin biosynthesis was altered *in planta* to result in the accumulation of biosynthetic saponin intermediates that do not normally accumulate in the plant, suggesting a role for saponins and their intermediates in plant growth and developmental processes (Moses et al., 2014a).

To correlate the phenotypic effects with saponins and/or saponin biosynthetic intermediates, we first assessed the accumulation of β-amyrin using GC–MS in control plants and the *MlbAS*-overexpressing plant O5, which has the most severely altered growth and developmental phenotype. Notably, β-amyrin levels were lower in the leaves of O5 plants when compared with control plants (Figure 2F). Since the *M. lanceolata* transgenic lines have functional downstream enzymes for the subsequent modification of β-amyrin to maesasaponins, we reasoned that the decreased accumulation of β-amyrin in the O5 line could be a result of its further modification *in planta*, which would then result in altered accumulation of maesasapogenins and/or maesasaponins, as suggested by the TLC analysis (Figure 2D). To further assess this, we screened the GC chromatograms of O5 plants for the accumulation of potential maesasaponin intermediates erythrodiol, oleanolic acid, and echinocystic acid, and performed liquid chromatography–mass spectrometry (LC–MS) analysis on the control and O5 lines to determine levels of maesasaponins in these lines. None of the maesasapogenins screened could be detected in the GC chromatograms, but higher accumulation of maesasaponins could indeed be observed in the O5 line (Figure 2G). The higher accumulation of maesasaponins in O5 plants suggests an increased flux through the pathway resulting from the overexpression of *MlbAS*, and links the phenotypic effects to either the decreased accumulation of β-amyrin or the increased accumulation of maesasaponins.

Yeast-Based *In Vivo* Screening of Candidate P450s Identifies Two Functional Enzymes

To identify functional enzymes capable of modifying the oleanane triterpenoid β-amyrin, we generated seven yeast strains, heterologously expressing a candidate P450 and the *Arabidopsis thaliana* P450 reductase (*AtATR1*), from the β-amyrin-producing strain TM3 expressing the *bAS* gene from *G. glabra* (Moses et al., 2014b), since the *MlbAS*-expressing strain TM4 showed only very low accumulation of β-amyrin. Cell extracts of 72-h-old cultures of these seven yeast strains were compared with a control strain by GC–MS analysis. The GC chromatograms of two candidate P450s, i.e. CYP716A75 (tag ML257) and CYP87D16 (tag ML593), showed unique peaks that did not occur in the control strain, suggesting a functional activity in the modification of β-amyrin *in vivo* (Figure 3A).

The P450s CYP716A75 and CYP87D16 represent proteins of 487 and 475 amino acids, respectively. A phylogenetic analysis with all the functionally characterized triterpenoid-modifying P450s clearly grouped CYP716A75 with the functionally diverse CYP716A subfamily, whereas CYP87D16 did not group with any particular P450 family (Figure 3B and Supplemental Figure 3).

CYP716A75 Is a C-28 Oxidase

To identify the enzymatic activity of CYP716A75, yeast strain TM20 that heterologously produces the protein was cultured in the presence of methyl-β-cyclodextrin (MβCD) to sequester triterpenoids from yeast cells to the growth medium (Moses et al., 2014b). The GC chromatogram of the organic extract obtained from its spent culture medium was compared with that of the control strain TM27, not expressing a P450. Three peaks were unique to strain TM20 and two of these matched authentic standards of erythrodiol and oleanolic acid. From the EI–MS pattern, the third peak was deduced to be oleanolic aldehyde, but its identity could not be confirmed due to the lack of an authentic standard (Figure 4A and 4B). A semiquantitation of the peak intensities revealed the accumulation of erythrodiol:oleanolic aldehyde:oleanolic acid to be in the ratio 8.7:1.1:0.2. This observation suggests a higher efficiency of CYP716A75 to catalyze the first oxidation from β-amyrin to erythrodiol, and a considerably lower efficiency to continue the subsequent oxidations to oleanolic acid via oleanolic aldehyde, when heterologously expressed in yeast. Furthermore, the products formed by CYP716A75 were identical to those detected in the strain TM17 (Figure 4A) expressing *CYP716A12*, the *M. truncatula* P450 that catalyzes the three-step oxidation of β-amyrin to oleanolic acid (Carelli et al., 2011). CYP716A12 displays an erythrodiol:oleanolic aldehyde:oleanolic acid product ratio of 6.5:2.4:1.1, further underscoring the relatively low efficiency of CYP716A75 to continue the subsequent oxidations to oleanolic acid in yeast. Nonetheless, our data collectively confirm the C-28 oxidase activity of CYP716A75 (Figure 4C). In addition, in the

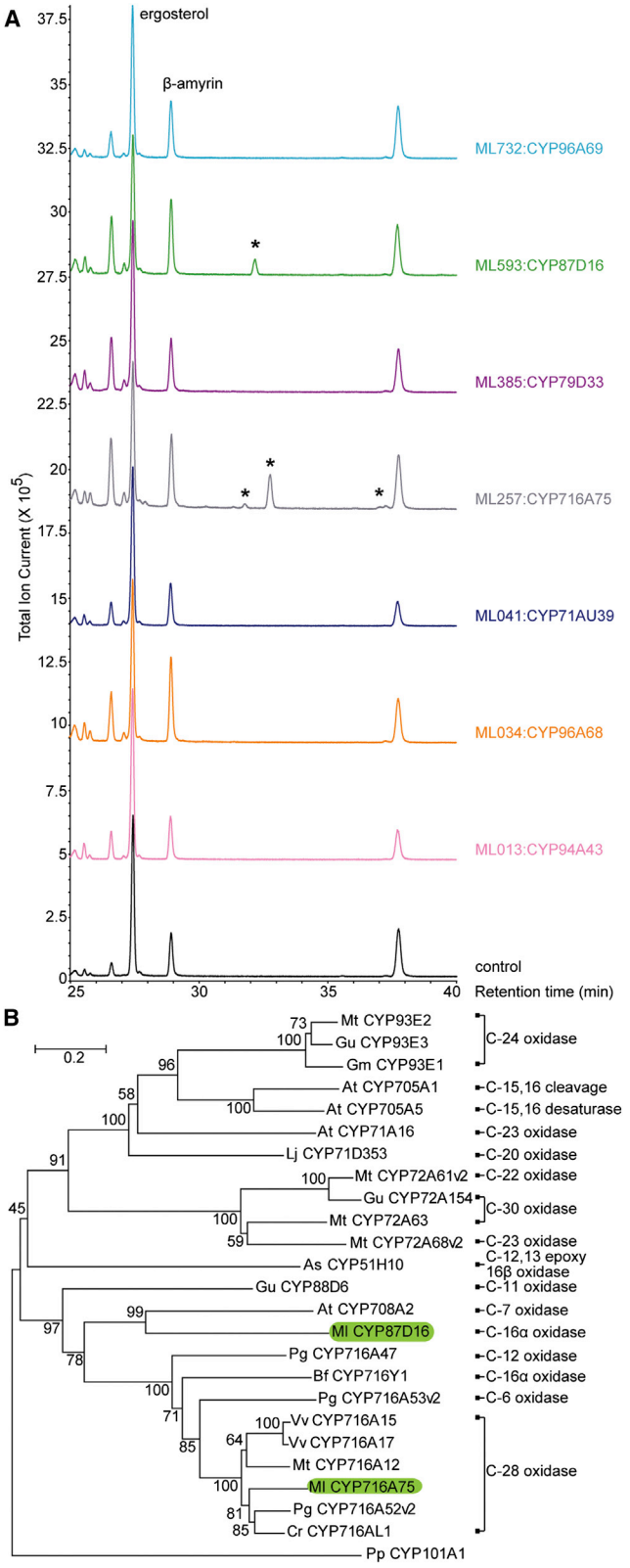

Figure 3. Two of the Seven *M. lanceolata* P450s Accept β-Amyrin as a Substrate When Heterologously Expressed in Yeast.

(A) Overlay of GC chromatograms of extracts of yeast strains expressing the seven candidate P450s, compared with that of a control strain. *CYP87D16* and *CYP716A75* encode functional P450s when expressed in yeast and correspond to gene tags ML593 and ML257, respectively. The unique peaks in each chromatogram are indicated by an asterisk.

(B) Phylogenetic analysis of all hitherto characterized triterpenoid-modifying P450s. The enzymatic activity of the P450s is indicated on the right. The percentage of replicate trees that clustered together in the bootstrap test is indicated on the branches. The scale bar represents the number of amino acid substitutions per site. The bacterial CYP101A1 was included as an outgroup. CYP87D16 and CYP716A75 characterized in this study are highlighted in green. Aa, *Artemisia annua*; As, *Avena strigosa*; At, *Arabidopsis thaliana*; Bf, *Bupleurum falcatum*; Cr, *Catharanthus roseus*; Gm, *Glycine max*; Gu, *Glycyrrhiza uralensis*; Lj, *Lotus japonicus*; Ml, *Maesa lanceolata*; Mt, *Medicago truncatula*; Pg, *Panax ginseng*; Pp, *Pseudomonas putida*; Vv, *Vitis vinifera*.

phylogenetic analysis, CYP716A75 clustered together with the other CYP716A subfamily of C-28 oxidases that have been characterized to be involved in saponin biosynthesis (Figure 3B).

CYP716A75 Complements the *M. truncatula lha-1* Mutant

Since CYP716A75 from *M. lanceolata* is a functional C-28 oxidase, we determined its ability to functionally complement the *lha-1* mutant, a *cyp716a12* mutant of *M. truncatula* that lacks the C-28 oxidase activity and therefore does not accumulate oleanolic acid-derived hemolytic saponins (Carelli et al., 2011). We generated five different types of transgenic *M. truncatula* hairy roots for this complementation assay. Three types of hairy roots were generated from the loss-of-function *lha-1* mutant line expressing the *M. lanceolata CYP716A75*, a functional copy of the *M. truncatula CYP716A12*, or a control *GUS* construct. Two additional types of control hairy roots were generated by expressing the control *GUS* construct in hairy roots derived from the E25-08 wild-type background and the E113 escape line (Carelli et al., 2011). The saponin compositions of the transgenic hairy roots were analyzed by LC–MS. Similar to *lha-1* mutant plants and contrary to hairy roots derived from the E25-08 and E113 control lines, the *lha-1* control hairy roots did not accumulate hemolytic saponins, whereas the accumulation of nonhemolytic soyasaponins was comparable between hairy roots derived from the three *M. truncatula* backgrounds. The expression of a functional copy of the *M. truncatula CYP716A12* in hairy roots derived from the *lha-1* mutant restored the C-28 oxidase activity in the mutant and resumed the accumulation of hemolytic saponins, confirming the utility of this screening approach (Figure 5A). It should be noted, however, that most, if not all, accumulating hemolytic saponins were less abundant in the *CYP716A12*-expressing *lha-1* hairy roots in comparison with hairy roots derived from the E25-08 and E113 control lines (Figure 5B–5E). Analysis of the LC chromatograms of the *lha-1* hairy roots heterologously expressing the *M. lanceolata CYP716A75* also showed the accumulation of hemolytic saponins, confirming restored C-28 oxidase activity and, hence, complementation of the loss of function of the native CYP716A12 in *M. truncatula* (Figure 5A). However, when compared with *lha-1* hairy roots expressing *CYP716A12*, the *lha-1* hairy roots expressing *CYP716A75* accumulated even lower levels of the hemolytic saponins (Figure 5B–5E), indicating lower C-28 carboxylation activity of *M. lanceolata* CYP716A75 compared with *M. truncatula* CYP716A12.

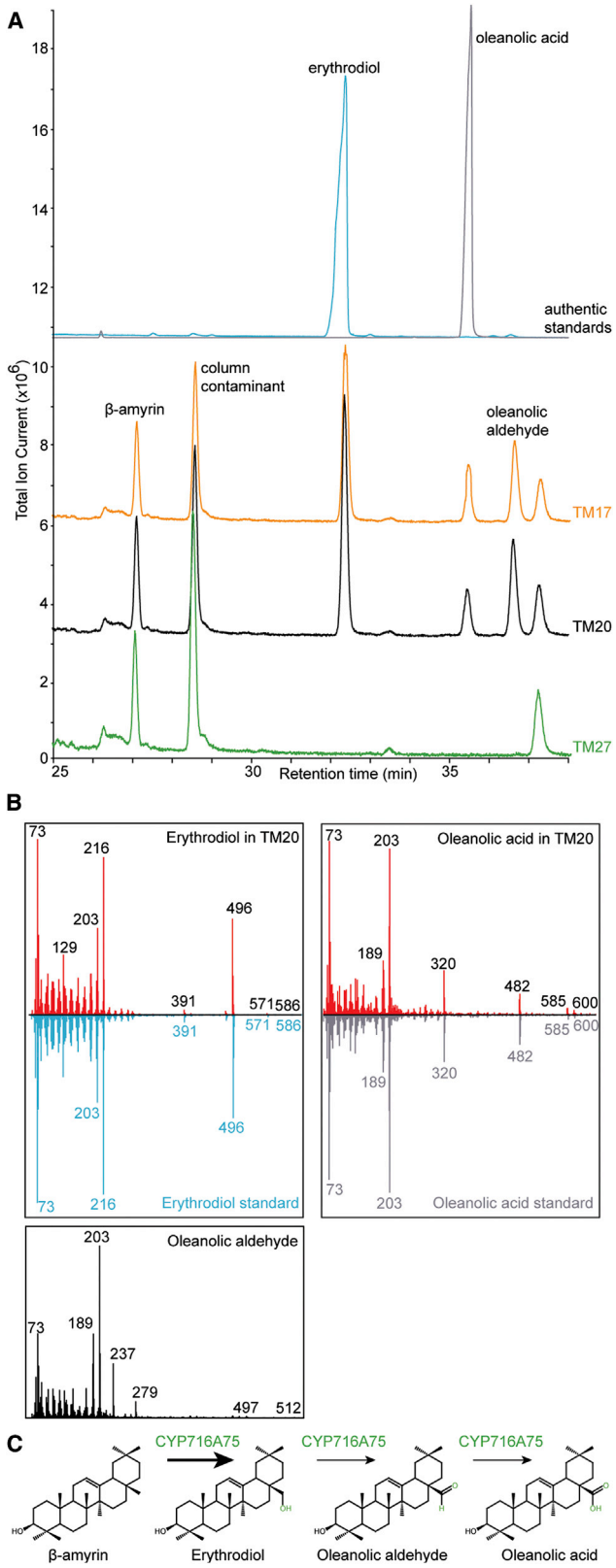

Figure 4. Characterization of CYP716A75 as a C-28 Oxidase.

(A) Overlay of GC chromatograms from spent medium of yeast strain TM20 expressing *CYP716A75* (black), TM17 expressing *CYP716A12* (orange), and control strain TM27 (green), erythrodiol standard (blue), and oleanolic acid standard (gray).

(B) EI-MS spectrum of trimethylsilylated erythrodiol, oleanolic aldehyde, and oleanolic acid eluting at 32.4 min, 36.8 min, and 35.5 min, respectively, from strain TM20.

(C) Structures of the substrate, intermediates, and product involved in the three-step oxidation reaction from β-amyrin to oleanolic acid, via erythrodiol and oleanolic aldehyde, catalyzed by CYP716A75.

Alternatively, variation in transgene copy number, level of transcript expression, and folding and stability of the resulting protein cannot be ruled out as possible experimental factors contributing to the lower accumulation.

In addition to lower levels of natural *Medicago* saponins, *lha-1* hairy roots expressing *CYP716A75* accumulated high levels of certain saponins, which were nearly absent in *lha-1* hairy roots expressing *CYP716A12* or hairy roots derived from the E25-08 and E113 control lines (Figure 5C). Based on the accurate masses and the MS^n spectra generated, these compounds were tentatively identified as saponins with aglycones having incomplete oxidation (alcohol or aldehyde) at position C-28 (Supplemental Figure 4).

CYP87D16 Is a C-16α Oxidase

To functionally characterize CYP87D16, yeast strain TM21, harboring the plasmids pAG415[*GAL1/AtATR1*] and pAG423 [*GAL1/CYP87D16*] in the parent strain TM3, and the control strain TM27, were cultured in the presence of MβCD. Organic extracts of the spent medium obtained from the strains were then analyzed by GC–MS, revealing a unique peak in the GC chromatogram of strain TM21 (Figure 6A). The parent ion at *m/z* 586 in the EI–MS pattern of this peak suggested the presence of an extra hydroxyl group on β-amyrin (Figure 6B). Furthermore, the fragment ions at *m/z* 201, 216, and 306 resulting from a retro-Diels–Alder fragmentation of the oleanane backbone indicated that the additional hydroxyl group was located either on the D or E ring of β-amyrin. The retention time and EI–MS pattern of this peak were identical to those of 16α-hydroxy β-amyrin (Figure 6C and 6D), which was produced in strain TM9 as a result of the hydroxylation of β-amyrin by CYP716Y1, a P450 we previously characterized (Moses et al., 2014b). Therefore, we conclude that the *M. lanceolata CYP87D16* encodes a C-16α oxidase involved in maesasaponin biosynthesis.

Combined Expression of *CYP87D16* and *CYP716A75* in Yeast

To generate intermediates of the maesasaponin biosynthesis pathway in yeast, we generated yeast strain TM104 from the parent strain TM3. Next to the *A. thaliana* P450 reductase (*AtATR1*), yeast strain TM104 was engineered to express both *CYP87D16* and *CYP716A75* from a single high-copy number plasmid to produce a self-processing polyprotein in which the two P450s are linked via a 2A oligopeptide (de Felipe et al., 2006). Comparison of GC chromatograms of organic extracts obtained from the culture medium of this strain with those of strains TM20 and TM21 revealed the presence of three peaks that are unique to the strain TM104 (Figure 7A). The retention time and EI–MS pattern of the second peak (comigrating with another peak present in all extracts) matched those of an authentic echinocystic acid standard (Figure 7B), confirming the

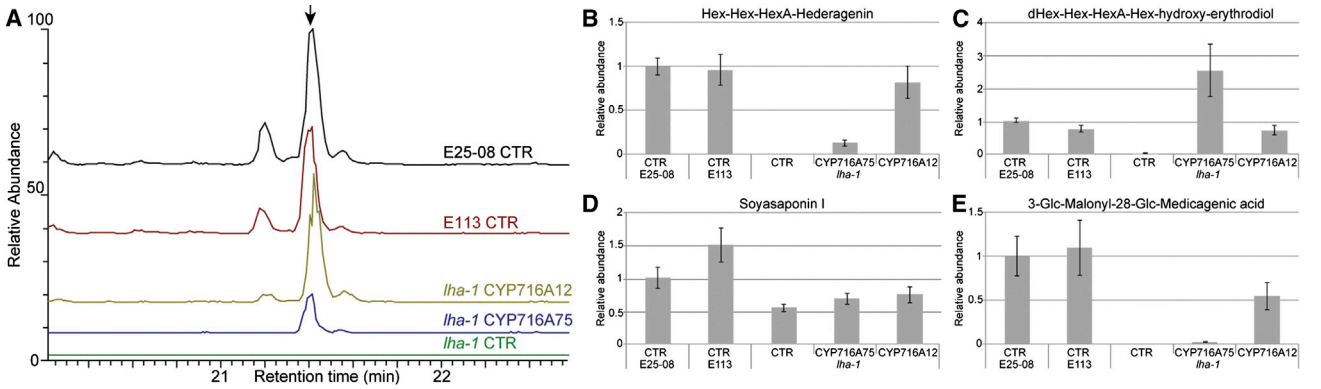

Figure 5. Heterologous Expression of *CYP716A75* Leads to Functional Complementation of the *M. truncatula lha-1* Mutant.
(A) Overlay of LC chromatograms from E25-08 (wild-type), E113 (escape), and *lha-1* (loss-of-function mutant) lines expressing a control (CTR) construct, and *lha-1* mutants expressing *CYP716A12* or *CYP716A75*. The Hex-Hex-HexA-Hederagenin peak at 21.4 min (indicated with an arrow) was detected by scanning a mass range of 971–972 Da.
(B–E) Relative abundance of Hex-Hex-HexA-Hederagenin **(B)**; a tentatively annotated (Supplemental Figure 4) saponin dHex-Hex-HexA-Hex-hydroxy erythrodiol with higher abundance in *lha-1* mutants expressing *CYP716A75* **(C)**; and two of the most abundant saponins in *M. truncatula* hairy roots (Pollier et al., 2011b), non-hemolytic Soyasaponin I **(D)**, and hemolytic 3-Glc-Malonyl-28-Glc-Medicagenic acid **(E)** in E25-08, E113, *lha-1* control lines, and *lha-1* lines expressing *CYP716A75* or *CYP716A12*. dHex, a pentose; Glc, glucose; Hex, a hexose; HexA, an uronic acid.

C-16α oxidation by CYP87D16 and the C-28 oxidation by CYP716A75. For the first and third unique peaks, no authentic standards were available. However, based on the enzymatic activities of the expressed P450s, it is expected that these peaks correspond to 16α-hydroxy erythrodiol and 16α-hydroxy oleanolic aldehyde. Interpretation of the EI–MS fragmentation patterns of these peaks (Figure 7C) corroborated this hypothesis. The parent ion at *m/z* 675 in the EI–MS pattern of the first unique peak in the organic extract of the strain TM104 corresponds with the presence of three hydroxyl groups on the β-amyrin backbone, as is the case for 16α-hydroxy erythrodiol, whereas the parent ion at *m/z* 601 in the EI–MS pattern of the third unique peak would be the expected mass for 16α-hydroxy oleanolic aldehyde. In addition, the most abundant fragment ions in both peaks corresponding to the loss of trimethylsilylated moieties and resulting from retro-Diels–Alder cleavage of ring C in the backbones (Figure 7C) are in agreement with the expected fragments of 16α-hydroxy erythrodiol and 16α-hydroxy oleanolic aldehyde. Hence, the combined expression of *CYP87D16* and *CYP716A75* in yeast leads to the accumulation of three intermediates involved in maesasaponin biosynthesis. The first, 16α-hydroxy erythrodiol, is a precursor of maesasaponins without a C-13,28 bridge, whereas 16α-hydroxy oleanolic aldehyde and echinocystic acid are precursors of maesasaponins with a hemiacetal or ester bridge, respectively (Figure 1C).

DISCUSSION

Despite the broad applications of triterpenoid saponins, their industrial utilization is limited by the low availability of pure compounds from plant sources for pharmaceutical and cosmetic applications, and the restricted knowledge of their biosynthetic pathways, which prevents generation of heterologous production systems. Maesasaponins constitute a wide spectrum of structurally unique triterpenoids with promising pharmaceutical activities, but no corresponding biosynthesis genes have been identified. In this study, we characterized three enzymes involved in the synthesis of maesasaponins, which are inherent to the African shrub *M. lanceolata*.

Functional Activity of MlbAS in *S. cerevisiae* and *M. lanceolata*

The *MlbAS* gene was cloned through a degenerate PCR approach and functionally characterized using our sterol-modified *S. cerevisiae* strain TM1. However, very low β-amyrin accumulation was observed in the resulting yeast, indicating poor exogenous activity of MlbAS in this heterologous host. Previously, we reported accumulation of 36 mg/l and 19 mg/l β-amyrin in strains derived from TM1 when the *bAS* gene from *G. glabra* and *M. truncatula bAS*, respectively, were expressed (Moses et al., 2014b), suggesting differences in the functional activities of bAS enzymes obtained from different plants. These differences in enzyme activities could be attributed to several factors, including the lack of upstream elements for the optimal expression of plant genes in yeast, nonoptimized codon usage, protein mislocalization, protein misfolding, and/or protein instability. Nonetheless, the overexpression of *MlbAS* in *M. lanceolata* resulted in increased accumulation of maesasaponins, and its silencing or cosuppression resulted in decreased maesasaponin levels in the transgenic plants, suggesting that this gene indeed codes for a functional enzyme and plays an important role *in planta* for the production of maesasaponins.

Functional Characterization of CYP87D16 Suggests Independent Evolvement of 16α Oxidase Activity in the Plant Kingdom

Plant P450s are conventionally classified into two main classes: the A type and non-A type. The A-type P450s belong to a single clade and are involved in the synthesis of plant-specific specialized metabolites, spanning alkaloids, terpenoids, and phenolics, whereas the non-A-type P450s are distributed over six clades

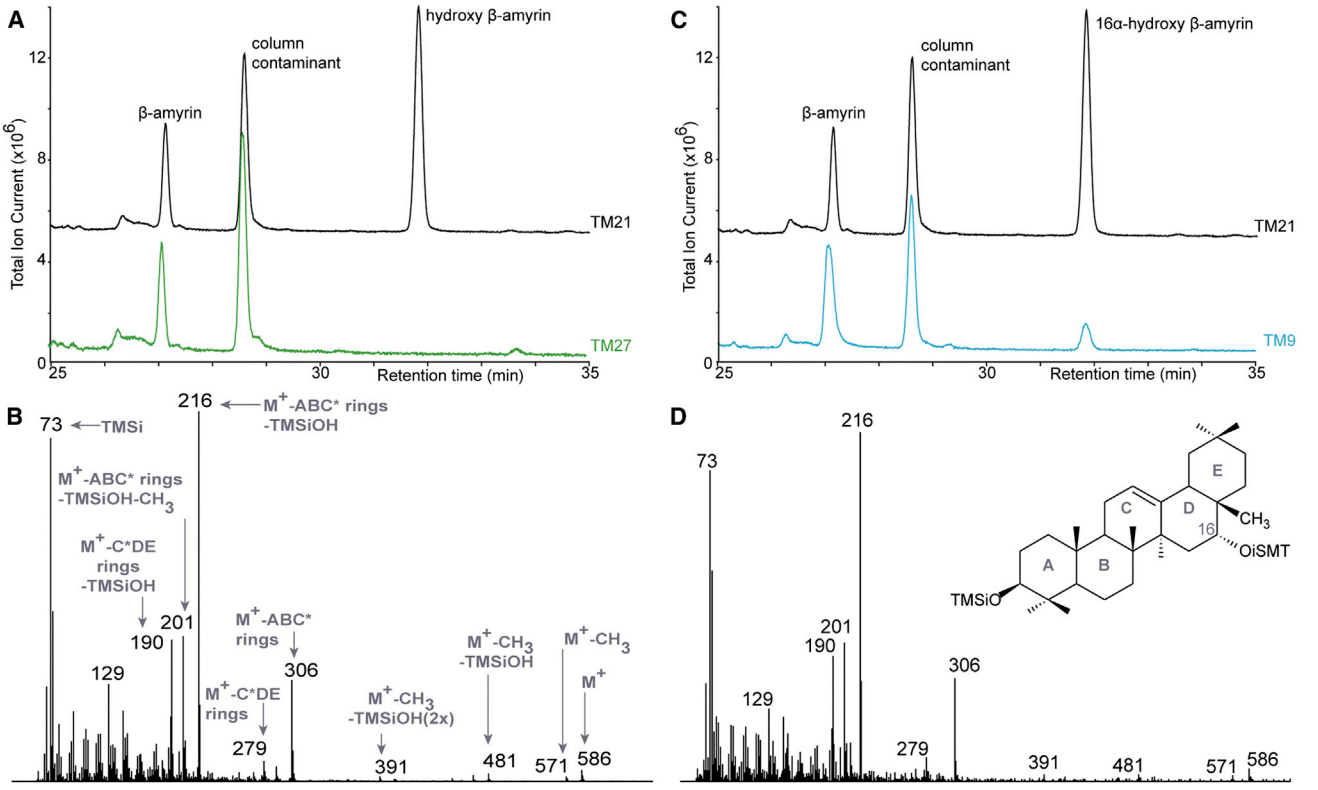

Figure 6. Characterization of CYP87D16 as a C-16α Oxidase.

(A) Overlay of GC chromatograms from spent medium of yeast strain TM21 expressing *CYP87D16* (black) and control strain TM27 (green).

(B) EI-MS spectrum of trimethylsilylated 16α-hydroxy β-amyrin eluting at 31.8 min from strain TM21. The fragment ions corresponding to the loss of specific groups from the parent ion (M^+) during electron impact ionization are indicated.

(C) Overlay of GC chromatograms from spent medium of yeast strain TM21 (black) expressing *CYP87D16*, and strain TM9 (blue) expressing *CYP716Y1*, a functionally characterized C-16α oxidase.

(D) EI-MS spectrum of trimethylsilylated 16α-hydroxy β-amyrin (structure) eluting at 31.8 min from strain TM9. The rings and C-16 position are indicated in the structure. TMSiO/OiSMT, trimethylsilylated oxygen.

representing enzymes involved in metabolic housekeeping functions, such as the synthesis of sterols, fatty acids, hormones, and other signaling molecules (Schuler and Werck-Reichhart, 2003). The CYP87 family belongs to the non-A-type multifamily CYP85 clan, of which most are involved in hormone metabolism (Bak et al., 2011). During evolution, the CYP87 family first appeared in monocots, but it is also present in many dicots, including Ericales, the order to which the large family of Myrsinaceae belongs. The CYP87 family shares a common ancestor with two other P450 families belonging to the CYP85 clan, namely CYP702 and CYP708, both of which are involved in triterpenoid biosynthesis (Nelson and Werck-Reichhart, 2011). Extensive P450 phylogenetic analyses have shown no indication of gene duplication in the CYP87 family, suggesting their involvement in an essential cellular function. More than 40 CYP87 family genes and pseudogenes belonging to the A, B, C, and D subfamilies have been identified in various plants including *A. thaliana*, *Oryza sativa* (rice), *Glycine max* (soybean), *M. truncatula*, *Populus trichocarpa* (poplar), *Vitis vinifera* (grape), and *Carica papaya* (papaya) (Nelson et al., 2008; Guttikonda et al., 2010). Of these, however, only one, CYP87A3, has been partially characterized to be implicated in auxin responsiveness of rice coleoptiles (Chaban et al., 2003).

We functionally characterized CYP87D16 from *M. lanceolata*, and, contrary to the general assumption that the CYP85 clan members are involved in hormone biosynthesis, found that it catalyzes the C-16α oxidation of oleanane-type triterpenoid backbones during triterpenoid saponin biosynthesis. In the field of saponin biosynthesis, another P450, CYP716Y1 from *Bupleurum falcatum* (order Apiales), was recently characterized to also have C-16α oxidase activity (Moses et al., 2014b). This enzyme belongs to the CYP716 family, which evolved with terrestrial plants and is, like the CYP87 family, a member of the CYP85 clan (Bak et al., 2011). All members of the CYP716 family identified to date are involved in triterpenoid saponin metabolism with a total of nine characterized enzymes, all of which catalyze various oxidations on different carbon positions of the oleanane-, ursane-, lupane-, and dammarane-type triterpenoid backbones. This list also includes *M. lanceolata* CYP716A75, identified here as a C-28 oxidase, which is identical to the reported activity of its close homologs from other plant species such as *Catharanthus roseus*, *M. truncatula*, *Panax ginseng*, and *V. vinifera* (Carelli et al., 2011; Fukushima et al., 2011; Han et al., 2013; Huang et al., 2012).

New P450s are named based on amino acid sequence identities, with members of a family and subfamily sharing ≥40%

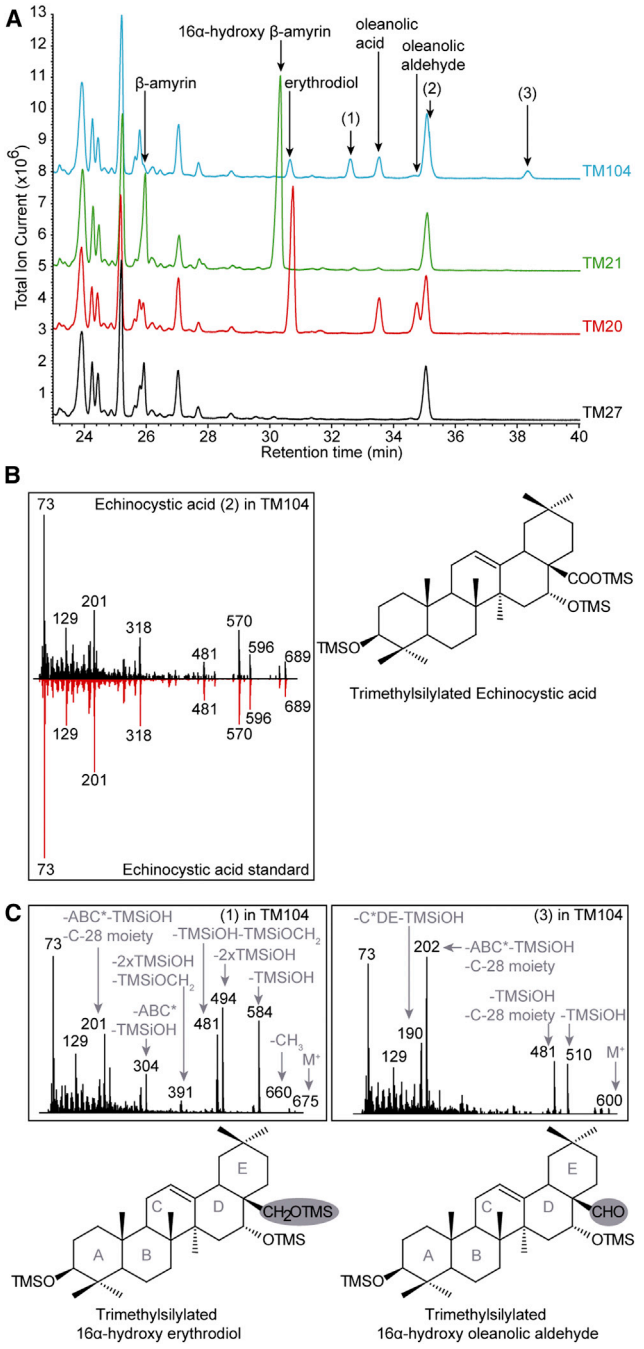

Figure 7. *In Vivo* Synthesis of Maesasaponin Biosynthesis Intermediates in Yeast.

(A) Overlay of GC chromatograms from spent medium of control yeast strain TM27 (black) and strains expressing *CYP716A75* alone (TM20, red), *CYP87D16* alone (TM21, green), or *CYP716A75* and *CYP87D16* together (TM104, blue). Three peaks unique to strain TM104 are indicated.

(B) EI-MS spectrum of trimethylsilylated echinocystic acid produced by strain TM104 (black) compared with the standard (red).

(C) EI-MS spectra of the peaks (1) and (3) that are unique to strain TM104. The predicted fragment loss corresponding to the most abundant *m/z* ions are given in gray. ABC*, mass corresponding to ring A, ring B, and partial ring C resulting from retro-Diels–Alder fragmentation; C*DE, mass corresponding to partial ring C, ring D, and ring E; CH_3, methyl group; M^+, trimethylsilylated parent ion; $TMSiOCH_2$, trimethylsilylated methanol; TMSiOH, trimethylsilanol. The C-28 moiety on the structure is highlighted in gray.

and ≥55% identity, respectively (Nelson et al., 2008). The CYP87D16 protein shows 27% homology to the CYP716Y1 protein, and therefore belongs to a different P450 family, although they are functional homologs of each other. To the best of our knowledge, no other examples of P450s from two different gene families exhibiting the same biochemical function have been reported to date. Previous phylogenetic analyses have not indicated the presence of common ancestry between the CYP87 and CYP716 families. Our current findings, however, suggest a possible, yet unexplored, evolutionary branching within the CYP85 clan of land plants, and suggest that C-16α oxidase activity could have evolved independently and distinctly in different plant species or lineages.

Functionality of Other Candidate P450s

Through heterologous gene expression in yeast, we show the functional activity of two of the seven candidate P450s identified through our cDNA-AFLP analysis. Despite the lack of activity on the β-amyrin backbone, the remaining five candidate P450s might be capable of modifying intermediates generated as a result of the catalytic action of CYP87D16 and/or CYP716A75. Recent combinatorial synthesis efforts in yeast have demonstrated that some P450s show sequential oxidation specificity, and can accept and oxidize only the reaction products of other P450s. For instance, CYP72A61v2 and CYP72A68v2 from *M. truncatula* do not oxidize β-amyrin, but catalyze the C-22 and C-23 oxidations of 24-hydroxy β-amyrin and oleanolic acid to generate soyasapogenol B and gypsogenic acid, respectively (Fukushima et al., 2013). Hence, it would be of interest to determine the possible functionality of the other candidate P450s in combination with CYP87D16 and CYP716A75, in particular considering the similar expression profiles of these gene tags after MeJA elicitation in *M. lanceolata* shoot cultures. The African shrub *M. lanceolata* accumulates saponins containing characteristic C-13,28 bridges, like the Asian plant *B. falcatum*, although the bridges occurring in saikosaponins are predominantly of the ether type, in contrast to the hemiacetal and ester bridges that occur in maesasaponins (Vincken et al., 2007). Perhaps some of these P450s can open the synthetic biology gate to these bridges.

METHODS

Chemicals

β-Amyrin, erythrodiol, oleanolic acid, and echinocystic acid were purchased from Extrasynthese. Hexane, pyridine, and *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide were purchased from Sigma-Aldrich. Methyl-β-cyclodextrin was purchased from CAVASOL.

Cultivation and Elicitation of *M. lanceolata* Plants

M. lanceolata axenic shoot cultures were generated and maintained as described previously (Faizal et al., 2011). For elicitation, each pot of shoot cultures was sprayed with 2 ml of deionized water containing 0.05% (v/v) Tween-20 in combination with 500 μM MeJA (dissolved in ethanol) or an equivalent amount of ethanol as control. For transcript profiling, samples were collected 0, 0.5, 1, 2, 4, 8, 24, and 48 h after elicitor or mock treatments. For each sample, three different plants from three different pots were pooled.

Transcript Profiling

The cDNA-AFLP based transcript profiling was performed with all 128 possible BstYI+1/MseI+2 primer combinations as described previously (Vuylsteke et al., 2007). For accurate quantification of band intensities, the gel images were analyzed with the AFLP-Quantar Pro software (KeyGene). The raw expression data obtained were processed and differentially expressed gene tags were selected for further analysis as described by Moses et al. (2014b). Cluster analysis, sequencing, and BLAST analysis were performed according to Rischer et al. (2006).

Phylogenetic Analysis and Alignment of *bAS* Sequences

The protein sequences were retrieved from GenBank and aligned with ClustalW. The phylogenetic tree was generated with MEGA 5.10 software (Tamura et al., 2011), according to the Neighbor-Joining method and bootstrapping with 1000 replicates. The evolutionary distances were computed with the Poisson correction method, and all positions containing gaps and missing data were eliminated from the dataset (complete deletion option).

The bAS protein sequences were retrieved from GenBank and aligned using default settings of Clustal X2 to generate a clustal consensus sequence and visualize conserved motifs (Larkin et al., 2007).

Full-Length Cloning and Generation of Yeast Expression Vectors

The FL coding regions of *CYP94A43*, *CYP96A68*, *CYP71AU39*, *CYP716A75*, *CYP79D33*, *CYP87D16*, and *CYP96A69* were screened for in an *M. lanceolata* Uncut Nanoquantity cDNA library (custom-made by Invitrogen) as reported previously (Pollier et al., 2011a), using the primer pairs P1+P2, P3+P4, P5+P6, P7+P8, P9+P10, P11+P12, and P13+P14, respectively (Supplemental Table 3).

Degenerate oligonucleotide primers were designed based on highly conserved amino acid regions of known plant *bAS* genes, and used to amplify a fragment of the *MlbAS* gene. First, a PCR was performed on *M. lanceolata* cDNA using the primers P15 and P16, corresponding to the amino acid motifs DGGWGLH and PLHRAAKL, respectively. The resulting amplicon was purified and used as a template for a nested PCR with the primers P17 and P18, corresponding to the amino acid motifs FLPMHPAKMW and AWEPAGAQE, respectively. The resulting PCR fragment was purified once again and cloned into the pGEM-T easy vector (Promega) for sequencing. The sequence obtained was used to design the primers P19 and P20, which were used to generate a DNA probe to screen for the FL coding sequence of *MlbAS* in the *M. lanceolata* Uncut Nanoquantity cDNA library (Pollier et al., 2011a).

The FL *MlbAS*, *CYP94A43*, *CYP96A68*, *CYP71AU39*, *CYP716A75*, *CYP79D33*, *CYP87D16*, and *CYP96A69* were PCR amplified for Gateway cloning into pDONR221 using the primer pairs P21+P22, P23+P24, P25+P26, P27+P28, P29+P30, P31+P32, P33+P34, and P35+P36, respectively. The *MlbAS* was excised from its entry vector using XhoI and KpnI and inserted into MCS2 of the plasmid pESC-URA[*GAL10/tHMG1*] to create the expression vector pESC-URA[*GAL10/tHMG1*; *GAL1/MlbAS*]. The entry clones of the candidate P450s were Gateway recombined into the high-copy number expression vector pAG423GAL-ccdB (Addgene plasmid 14149; Alberti et al., 2007). The *A. thaliana* cytochrome P450 reductase *AtATR1* (At4g24520) was cloned into the low-copy number yeast expression vector pAG415GAL-ccdB (Addgene plasmid 14145, Alberti et al. [2007]) as described previously (Moses et al., 2014b).

The self-processing polyprotein of *CYP87D16* and *CYP716A75* was generated by the fusion of two amplicons. *CYP87D16* was amplified without a stop codon using primers P33+P37 to add a 3′-overhang of the partial T2A sequence. *CYP716A75* was amplified with primers P38+P30 to add a 5′-overhang of the remaining T2A sequence, creating an overlap of 22 bp in the T2A sequence between the two amplicons. The resulting amplicons were fused by PCR using the primers P33+P30, thereby generating a fragment for Gateway recombination. The resulting entry clone was further recombined into pAG423GAL-ccdB.

Generation of Plant Expression Vectors, Transgenic Plants, and RT–PCR Analysis

For the *M. truncatula* complementation assay, entry clones of *CYP716A75* and *CYP716A12* were Gateway recombined into the binary vector pK7WG2D (Karimi et al., 2002). For control hairy roots, the entry clone pEN-L1-S-L2, containing the *GUS* gene (Karimi et al., 2007), was Gateway recombined with pK7WG2D. The resulting clones were transformed into *Agrobacterium rhizogenes* and used to transform *M. truncatula* E25-08, E113 and *lha-1* seedlings as reported previously (Pollier et al., 2011b).

The 388-bp-long *MlbAS* hpRNAi construct was PCR amplified using primers P39 and P40 and cloned into the binary vector pK7GWIWG2D(II) (Karimi et al., 2002) by Gateway recombination. The resulting hpRNAi construct and the expression vector of *MlbAS* in pK7WG2D were transformed into *A. tumefaciens* LBA4404 and used for transforming *M. lanceolata* leaves, as reported previously (Faizal and Geelen, 2012).

Total RNA was extracted from 3-month-old *M. lanceolata* leaves with Concert Plant RNA Reagent (Invitrogen) and cDNA prepared with Super-Script II Reverse Transcriptase (Invitrogen). The *M. lanceolata* homolog of the *A. thaliana* 26S proteasome non-ATPase regulatory subunit 11 (RPN6A, At1g29150) was used as a control gene. Primer pairs P41+P42 and P43+P44 were used for RT–PCR analysis of *MlbAS* and *MlRPN6A*, respectively.

Culturing of Yeast Cells, Metabolite Extraction, and GC–MS Analysis

For *in vivo* screening of functional P450s, yeast precultures were grown in synthetic defined (SD) medium containing glucose (Duchefa, Clontech) and appropriate drop-out (DO) supplements (Duchefa, Clontech) at 30°C and 250 rpm for 18–20 h. The cultures were induced by inoculating to a starting optical density of 0.25 in SD Gal/Raf medium containing galactose and raffinose as the carbon source (Duchefa, Clontech) and appropriate DO supplements. The cultures were incubated at 30°C and 250 rpm for 24 h before addition of methionine to a concentration of 1.5 mM, and cultured further for 48 h. Yeast cells from 1 ml of the 72-h culture were lysed by boiling in equal volumes of 40% potassium hydroxide and 50% ethanol. The hydrophobic metabolites were extracted with hexane and 1 μl of the extract was analyzed by GC-MS.

For functional characterization of CYP87D16 and CYP716A75, yeast strains were cultured as above but with the addition of MβCD to a final concentration of 5 mM to the induction medium at 24 and 48 h after inoculation. Metabolite extractions were carried out on 72-h-old cultures using hexane. The organic phases resulting from a triple hexane extraction were pooled, evaporated to dryness, and trimethylsilylated using pyridine and *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide for GC–MS analysis.

For the synthesis of maesasaponin pathway intermediates, yeast strains were cultured in the presence of MβCD as described above. Metabolite extractions were carried out on 72-h-old cultures using a double hexane extraction, followed by an extraction using the more polar ethyl acetate. The resulting organic phases were pooled, evaporated to dryness, and trimethylsilylated using pyridine and *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide for GC–MS analysis.

GC–MS analysis was performed on a GC model 6890 and MS model 5973 (Agilent) in splitless mode. A VF-5ms capillary column (Varian CP9013, Agilent) was operated at a constant helium flow of 1 ml/min. The oven was held at 80°C for 1 min after injection, ramped to 280°C at a rate

of 20°C/min, held at 280°C for 45 min, ramped to 320°C at a rate of 20°C/min, held at 320°C for 1 min, and finally cooled to 80°C at a rate of 50°C/min at the end of the run. The injector was set to 280°C, the MS transfer line to 250°C, the MS ion source to 230°C, and the quadrupole to 150°C throughout. A full EI–MS spectrum was generated for each sample by scanning the *m/z* range of 60–800 with a solvent delay of 7.8 min. For select ion monitoring of β-amyrin, the *m/z* ions 218 and 203 were screened. For relative quantification, the peak areas were calculated using the default settings of AMDIS software (version 2.6, NIST, USA).

Metabolite Profiling of *M. truncatula* Hairy Roots

M. truncatula hairy roots were generated and maintained as described previously (Pollier et al., 2011b). Three technical repeats of five independent transgenic lines per construct were grown for 21 days in liquid medium. The roots were harvested by removing the medium, rinsed with purified water, frozen and ground in liquid nitrogen. Metabolites from 400 mg of ground hairy roots were extracted as described by Pollier et al. (2011b).

For reversed-phase LC–MS analysis, an Acquity UPLC BEH C18 column (150 × 2.1 mm, 1.7 μm; Waters) was coupled to an LTQ XL Linear Ion Trap MS via an electrospray ionization source. The following gradient was operated: 0 min, 5% B; 30 min, 55% B; 35 min, 100% B, using solvent A water:acetonitrile (99:1) and solvent B acetonitrile:water (99:1), both acidified with 0.1% formic acid. The loop size, flow, and column temperature were 25 μl, 300 μl/min, and 80°C, respectively. For negative ionization, the capillary temperature, sheath gas, auxiliary gas, and spray voltage were set to 300°C, 30 (arbitrary units), 5 (arbitrary units), and 5.0 kV, respectively. A full ion-trap MS spectrum was monitored between *m/z* 120 and 1400.

The resulting chromatograms were integrated and aligned with the XCMS package (Smith et al., 2006) in R version 2.6.1, with the following parameter values: xcmsSet(fwhm = 10, max = 300, snthresh = 2, mzdiff = 0.5), group(bw = 10, max = 300, mzwid = 0.5), retcor(method = loess, family = symmetric). A second grouping was done with the same parameter values.

For compound identification, the accurate masses of the eluting compounds were determined, and MS^n spectra were generated by reanalyzing representative samples by LC Fourier transform ion cyclotron resonance MS as described by Pollier et al. (2011b).

TLC Analysis of Maesasaponins

Metabolites from 50 mg of dry powdered plant material were extracted twice in 50% methanol by sonication for 1 h each. The extracts were centrifuged for 10 min at 18 000 *g*, and the resulting supernatants pooled and evaporated to dryness. The dry material was resuspended in 80% methanol and 10 μl was spotted on normal-phase silica gel 60 plates with fluorescence indicator (F_{254}) (Merck, Germany) for TLC analysis. The mobile phase corresponded to the upper phase of a mixture of *n*-butanol:acetic acid:water (40:10:50). The TLC plates were developed by spraying with *p*-anisaldehyde reagent and heating at 100°C for 10 min.

GC–MS Analysis of *M. lanceolata* Transgenic Lines

Metabolites from 10 mg of dry powdered leaf material were extracted with 1 ml of 100% methanol for 1 h. The extracts were centrifuged for 10 min at 20 800 *g* and the resulting supernatant was evaporated to dryness. For quantitative analysis, 30 μg/ml cholesterol was added as an internal standard to each sample. The residue was dissolved in 0.5 ml of water and extracted twice with hexane (0.5 ml per extraction), followed by an extraction with 0.5 ml of the more polar solvent ethyl acetate. The organic fractions were pooled, evaporated to dryness, and trimethylsilylated using pyridine and *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide for GC–MS analysis. GC-MS was performed as described above. The area under the curve of peaks corresponding to β-amyrin and cholesterol were calculated using the default settings of AMDIS software (version 2.6, NIST, USA) and the area for β-amyrin normalized against the area of the internal standard.

LC–MS Analysis of *M. lanceolata* Transgenic Lines

Metabolites from 10 mg of dry powdered leaf material were extracted with 1 ml of 90% methanol for 1 h. The extracts were centrifuged for 10 min at 20 800 *g* and the resulting supernatant was evaporated to dryness. The residue was dissolved in 600 μl of water/cyclohexane (2:1, v/v) and centrifuged for 10 min at 20 800 *g*. Two hundred microliters of the aqueous phase was retained for analysis. LC–MS analysis was performed as described for *M. truncatula* hairy roots. For saponin quantification, the peak areas corresponding to the 15 characterized maesasaponins of the MC3B1 mixture (Theunis et al., 2007) were integrated using the default settings of Xcalibur software (Thermo Fisher Scientific, USA). The areas obtained were added up and plotted relative to the sum of the peak areas of control line 1.

ACCESSION NUMBERS

The sequence data of *MlbAS*, *CYP71AU39*, *CYP79D33*, *CYP87D16*, *CYP94A43*, *CYP96A68*, *CYP96A69*, and *CYP716A75* have been deposited in GenBank with accession numbers KF425519, KF318732, KF318734, KF318735, KF318730, KF318731, KF318736, and KF318733, respectively.

SUPPLEMENTAL INFORMATION

Supplemental Information is available at *Molecular Plant Online*.

FUNDING

This research has received funding from the Agency for Innovation by Science and Technology in Flanders ("Strategisch Basisonderzoek" Combiplan project SBO040093) and the Research Foundation-Flanders through project G.0014.08. T.M. is indebted to the VIB International PhD Fellowship Program for a predoctoral fellowship. J.P. is a postdoctoral fellow of the Research Foundation Flanders (FWO). A.F. is supported by the Directorate General of Higher Education, Ministry of National Education, Republic of Indonesia, for a predoctoral fellowship.

ACKNOWLEDGMENTS

We thank Carla Scotti for the *Medicago* mutant seeds, David Nelson for naming the P450s, Robin Vanden Bossche for technical assistance, and Annick Bleys for help with preparing the manuscript. No conflict of interest declared.

Received: June 10, 2014
Accepted: September 27, 2014
Published: October 8, 2014

REFERENCES

Alberti, S., Gitler, A.D., and Lindquist, S. (2007). A suite of Gateway cloning vectors for high-throughput genetic analysis in Saccharomyces cerevisiae. Yeast **24**:913–919.

Apers, S., Foriers, A., Sindambiwe, J.B., Vlietinck, A., and Pieters, L. (1998). Separation of a triterpenoid saponin mixture from Maesa lanceolata: semipreparative reversed-phase wide pore high performance liquid chromatography with temperature control. J. Pharm. Biomed. Anal. **18**:737–743.

Augustin, J.M., Kuzina, V., Andersen, S.B., and Bak, S. (2011). Molecular activities, biosynthesis and evolution of triterpenoid saponins. Phytochemistry **72**:435–457.

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

Carelli, M., Biazzi, E., Panara, F., Tava, A., Scaramelli, L., Porceddu, A., Graham, N., Odoardi, M., Piano, E., Arcioni, S., et al. (2011). Medicago truncatula CYP716A12 is a multifunctional oxidase

involved in the biosynthesis of hemolytic saponins. Plant Cell **23**:3070–3081.

Chaban, C., Waller, F., Furuya, M., and Nick, P. (2003). Auxin responsiveness of a novel cytochrome p450 in rice coleoptiles. Plant Physiol. **133**:2000–2009.

de Felipe, P., Luke, G.A., Hughes, L.E., Gani, D., Halpin, C., and Ryan, M.D. (2006). E unum pluribus: multiple proteins from a self-processing polyprotein. Trends Biotechnol. **24**:68–75.

De Geyter, N., Gholami, A., Goormachtig, S., and Goossens, A. (2012). Transcriptional machineries in jasmonate-elicited plant secondary metabolism. Trends Plant Sci. **17**:349–359.

Faizal, A. (2013). Modulation of saponin production in medicinal Maesa spp, PhD thesis (Belgium: Ghent University). Available online at http://lib.ugent.be/catalog/pug01:3191661.

Faizal, A., and Geelen, D. (2012). Agroinfiltration of intact leaves as a method for the transient and stable transformation of saponin producing Maesa lanceolata. Plant Cell Rep. **31**:1517–1526.

Faizal, A., Lambert, E., Foubert, K., Apers, S., and Geelen, D. (2011). In vitro propagation of four saponin producing Maesa species. Plant Cell Tiss. Org. Cult. **106**:215–223.

Foubert, K., Cuyckens, F., Vleeschouwer, K., Theunis, M., Vlietinck, A., Pieters, L., and Apers, S. (2010). Rapid quantification of 14 saponins of Maesa lanceolata by UPLC-MS/MS. Talanta **81**:1258–1263.

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

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

Geisler, K., Hughes, R.K., Sainsbury, F., Lomonossoff, G.P., Rejzek, M., Fairhurst, S., Olsen, C.E., Motawia, M.S., Melton, R.E., Hemmings, A.M., et al. (2013). Biochemical analysis of a multifunctional cytochrome P450 (CYP51) enzyme required for synthesis of antimicrobial triterpenes in plants. Proc. Natl. Acad. Sci. U S A **110**:E3360–3367.

Guttikonda, S.K., Trupti, J., Bisht, N.C., Chen, H., An, Y.Q., Pandey, S., Xu, D., and Yu, O. (2010). Whole genome co-expression analysis of soybean cytochrome P450 genes identifies nodulation-specific P450 monooxygenases. BMC Plant Biol. **10**:243.

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

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

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

Karimi, M., Bleys, A., Vanderhaeghen, R., and Hilson, P. (2007). Building blocks for plant gene assembly. Plant Physiol. **145**:1183–1191.

Karimi, M., Inzé, D., and Depicker, A. (2002). GATEWAY vectors for Agrobacterium-mediated plant transformation. Trends Plant Sci. **7**:193–195.

Kirby, J., Romanini, D.W., Paradise, E.M., and Keasling, J.D. (2008). Engineering triterpene production in Saccharomyces cerevisiae-beta-amyrin synthase from Artemisia annua. FEBS J. **275**:1852–1859.

Kunii, M., Kitahama, Y., Fukushima, E.O., Seki, H., Muranaka, T., Yoshida, Y., and Aoyama, Y. (2012). β-Amyrin oxidation by oat CYP51H10 expressed heterologously in yeast cells: the first example of CYP51-dependent metabolism other than the 14-demethylation of sterol precursors. Biol. Pharm. Bull. **35**:801–804.

Kushiro, T., Shibuya, M., Masuda, K., and Ebizuka, Y. (2000). Mutational studies on triterpene synthases: engineering lupeol synthase into β-amyrin synthase. J. Am. Chem. Soc. **122**:6816–6824.

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

Manguro, L.O.A., Midiwo, J.O., Tietze, L.F., and Hao, P. (2011). Triterpene saponins of Maesa lanceolata leaves. ARKIVOC **ii**:172–198.

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

Moses, T., Papadopoulou, K.K., and Osbourn, A. (2014a). Metabolic and functional diversity of saponins, biosynthetic intermediates and semi-synthetic derivatives. Crit. Rev. Biochem. Mol. Biol. **49**:439–462.

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

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

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

Nelson, D.R., Ming, R., Alam, M., and Schuler, M.A. (2008). Comparison of cytochrome P450 genes from six plant genomes. Trop. Plant Biol. **1**:216–235.

Osbourn, A., Goss, R.J.M., and Field, R.A. (2011). The saponins – polar isoprenoids with important and diverse biological activities. Nat. Prod. Rep. **28**:1261–1268.

Pauwels, L., Inzé, D., and Goossens, A. (2009). Jasmonate-inducible gene: what does it mean? Trends Plant Sci. **14**:87–91.

Pollier, J., González-Guzmán, M., Ardiles-Diaz, W., Geelen, D., and Goossens, A. (2011a). An integrated PCR colony hybridization approach to screen cDNA libraries for full-length coding sequences. PLoS ONE **6**:e24978.

Pollier, J., Morreel, K., Geelen, D., and Goossens, A. (2011b). Metabolite profiling of triterpene saponins in Medicago truncatula hairy roots by liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry. J. Nat. Prod. **74**:1462–1476.

Pollier, J., Moses, T., and Goossens, A. (2011c). Combinatorial biosynthesis in plants: a (p)review on its potential and future exploitation. Nat. Prod. Rep. **28**:1897–1916.

Poralla, K. (1994). The possible role of a repetitive amino acid motif in evolution of triterpenoid cyclases. Bioorg. Med. Chem. Lett. **4**:285–290.

Rischer, H., Orešič, M., Seppänen-Laakso, T., Katajamaa, M., Lammertyn, F., Ardiles-Diaz, W., Van Montagu, M.C., Inzé, D., Oksman-Caldentey, K.M., and Goossens, A. (2006). Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in

Catharanthus roseus cells. Proc. Natl. Acad. Sci. U S A **103**:5614–5619.

Schuler, M.A., and Werck-Reichhart, D. (2003). Functional genomics of P450s. Annu. Rev. Plant Biol. **54**:629–667.

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

Sindambiwe, J.B., Calomme, M., Geerts, S., Pieters, L., Vlietinck, A.J., and Vanden Berghe, D.A. (1998). Evaluation of biological activities of triterpenoid saponins from Maesa lanceolata. J. Nat. Prod. **61**:585–590.

Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R., and Siuzdak, G. (2006). XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal. Chem. **78**:779–787.

Sparg, S.G., Light, M.E., and van Staden, J. (2004). Biological activities and distribution of plant saponins. J. Ethnopharmacol. **94**:219–243.

Suzuki, H., Achnine, L., Xu, R., Matsuda, S.P., and Dixon, R.A. (2002). A genomics approach to the early stages of triterpene saponin biosynthesis in Medicago truncatula. Plant J. **32**:1033–1048.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28**:2731–2739.

Theunis, M.H., Foubert, K., Pollier, J., Gonzalez-Guzman, M., Goossens, A., Vlietinck, A.J., Pieters, L.A., and Apers, S. (2007). Determination of saponins in Maesa lanceolata by LC-UV: development and validation. Phytochemistry **68**:2825–2830.

Thimmappa, R., Geisler, K., Louveau, T., O'Maille, P., and Osbourn, A. (2014). Triterpene biosynthesis in plants. Annu. Rev. Plant Biol. **65**:16.1–16.33.

Vincken, J.P., Heng, L., de Groot, A., and Gruppen, H. (2007). Saponins, classification and occurrence in the plant kingdom. Phytochemistry **68**:275–297.

Vuylsteke, M., Peleman, J.D., and van Eijk, M.J. (2007). AFLP-based transcript profiling (cDNA-AFLP) for genome-wide expression analysis. Nat. Protoc. **2**:1399–1413.

Xu, R., Fazio, G.C., and Matsuda, S.P. (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., and Qi, X. (2012). Divergent evolution of oxidosqualene cyclases in plants. New Phytol. **193**:1022–1038.