RESEARCH ARTICLE

A Seed-Specific Regulator of Triterpene Saponin Biosynthesis in *Medicago truncatula*[W]

Bianca Ribeiro,[a,b,1] Elia Lacchini,[a,b,1] Keylla U. Bicalho,[a,b,c,1] Jan Mertens,[a,b] Philipp Arendt,[a,b] Robin Vanden Bossche,[a,b] Gabriela Calegario,[a,b] Lore Gryffroy,[a,b] Evi Ceulemans,[a,b] Julia Buitink,[d] Alain Goossens,[a,b,2] and Jacob Pollier[a,b,e,2]

[a]Ghent University, Department of Plant Biotechnology and Bioinformatics, 9052 Ghent, Belgium.
[b]VIB Center for Plant Systems Biology, 9052 Ghent, Belgium.
[c]Department of Organic Chemistry, Institute of Chemistry, São Paulo State University (UNESP), Araraquara, São Paulo, Brazil.
[d]IRHS-UMR1345, Université d'Angers, INRAE, Institut Agro, SFR 4207 QuaSaV, 49071, Beaucouzé, France.
[e]VIB Metabolomics Core, 9052 Ghent, Belgium.
[1]These authors contributed equally to this work.
[2]Corrresponding Authors: alain.goossens@psb-vib.ugent.be and jacob.pollier@psb-vib.ugent.be

ORCID IDs: 0000-0002-1843-0258 (B.R.); 0000-0002-1598-8950 (E.L.); 0000-0002-5165-9070 (K.U.B.); 0000-0002-8095-0748 (J.M.); 0000-0002-7883-5070 (P.A.); 0000-0002-4593-9658 (R.V.B); 0000-0002-3772-8447 (G.C.); 0000-0001-9202-2992 (L.G.); 0000-0002-3083-5768 (E.C.); 0000-0002-1457-764X (J.B.); 0000-0002-1599-551X (A.G.); 0000-0002-1134-9238 (J.P.)

Short title: Seed-specific saponin biosynthesis in *Medicago*

One-sentence summary: The discovery of a seed-specific regulator of hemolytic saponin biosynthesis in *Medicago truncatula* led to the identification of the missing P450 of the hemolytic saponin biosynthesis branch.

The authors responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) are: Alain Goossens (alain.goossens@psb-vib.ugent.be) and Jacob Pollier (jacob.pollier@psb-vib.ugent.be).

[W]Online version contains Web-only data.

ABSTRACT

Plants produce a vast array of defense compounds to protect themselves from pathogen attack or herbivore predation. A specific class of defense compounds are the saponins, bioactive glycosides with a steroidal or triterpenoid aglycone backbone. The model legume *Medicago truncatula* synthesizes two types of saponins, hemolytic saponins and non-hemolytic soyasaponins, which accumulate as specific blends in the different organs of the plant. Here we report the identification of a seed-specific transcription factor, TRITERPENE SAPONIN ACTIVATION REGULATOR 3

(TSAR3), which controls the hemolytic saponin biosynthesis in developing *M. truncatula* seeds. Co-expression analyses with TSAR3 in transcriptome datasets from developing *M. truncatula* seeds led to the identification of CYP88A13, a cytochrome P450 that catalyzes the C-16α hydroxylation of medicagenic acid towards zanhic acid, the final, yet unknown, oxidation step of the hemolytic saponin biosynthesis branch in *M. truncatula*. In addition, two uridine diphosphate glycosyltransferases, UGT73F18 and UGT73F19, which glucosylate hemolytic sapogenins at the C-3 position, were identified. The identified biosynthesis enzymes are organized in clusters of duplicated genes on the *M. truncatula* genome. This appears to be a quite common theme among saponin biosynthesis genes, especially glycosyltransferases, and may be the driving force of saponin metabolic evolution.

INTRODUCTION

As sessile organisms, plants are often confronted with various abiotic and biotic stress situations, such as herbivore and pathogen predation. As a defensive measure, plants may activate the production of bioactive specialized metabolites to combat or deter the attackers. A specific group of defense molecules are the saponins, a structurally diverse class of glycosides with a steroid, steroidal alkaloid or triterpenoid aglycone backbone that are widely distributed in the plant kingdom, but also occur in certain marine invertebrates (Osbourn et al., 2011; Claereboudt et al., 2019). With their lipophilic aglycone or sapogenin that is covalently bound to one or more hydrophilic sugar chains, saponins are amphipathic molecules that form a colloidal solution in water, and which forms a stable soap-like foam when shaken (Vincken et al., 2007). Because of these physicochemical properties, saponins are often applied as emulsifiers and foaming agents in the food and beverage industries. Furthermore, their structural and functional diversity is reflected by their broad spectrum of biological activities, making them attractive molecules for the cosmetics and pharmaceutical industries (Vincken et al., 2007; Augustin et al., 2011; Osbourn et al., 2011; Gholami et al., 2014).

The model legume *Medicago truncatula* (barrel medic), a member of the Fabaceae plant family and a close relative of the important forage crop *M. sativa* (alfalfa), accumulates a complex mixture of up to 100 different oleanane-type triterpene saponins and has been widely used to study triterpene saponin biosynthesis (Suzuki et al., 2002; Pollier et al., 2011; Tava et al., 2011; Gholami et al., 2014). The saponins produced by *Medicago* species provide protection against herbivores as a result of their bitter and astringent flavor that deters the attackers from eating the plant. Furthermore, *Medicago* saponins also serve anti-nutritive roles by reducing the palatability and digestibility in ruminants and because of their toxicity towards monogastric animals and herbivorous insects (Gholami et al., 2014; Rafińska et al., 2017). Based on the biological properties of the molecules, two types of triterpene saponins can be distinguished in *Medicago* species: the hemolytic saponins and the non-hemolytic soyasaponins. The hemolytic properties of the saponins are due to their affinity for membrane sterols and are determined by the aglycone moiety (Podolak et al., 2010; Carelli et al., 2011). The aglycones of the hemolytic saponins are oxidized at the C-28 position, which is often accompanied with different degrees of oxidation at the C-23 position. The aglycones of the non-hemolytic soyasaponins have a hydroxy group at the C-24 position, which seemingly excludes oxidation at the C-28 position (Carelli et al., 2011; Tava et al., 2011).

The biosynthetic precursor of the *M. truncatula* saponins is 2,3-oxidosqualene (Figure 1), which is synthesized via the cytosolic mevalonate pathway. The cyclization of 2,3-oxidosqualene by β-amyrin synthase leads to the production of β-amyrin and forms the branch point between the primary sterol and the specialized triterpene saponin metabolism (Figure 1) (Suzuki et al., 2002; Iturbe-Ormaetxe et al., 2003). Subsequently, the competitive action of two cytochrome P450 enzymes (P450s) causes branching of the *M. truncatula* triterpene saponin biosynthesis pathway (Figure 1). CYP716A12 catalyzes the carboxylation of β-amyrin at the C-28 position, thereby yielding oleanolic acid, the precursor of the hemolytic saponins (Carelli et al., 2011; Fukushima et al., 2011), whereas hydroxylation at the C-24 position, catalyzed by CYP93E2, leads to the production of the non-hemolytic soyasaponins (Fukushima et al., 2013). In the hemolytic branch, oleanolic acid is the substrate of two additional P450s, CYP72A67 and CYP72A68, that catalyze the hydroxylation of the oleanane backbone at the C-2 position and the three-step carboxylation at the C-23 position, respectively, thereby yielding the major *Medicago* aglycone medicagenic acid in addition to a set of reaction intermediates like hederagenin and bayogenin (Figure 1) (Fukushima et al., 2013; Biazzi et al., 2015; Tzin et al., 2019). The produced medicagenic acid can be further converted into zanhic acid by a yet unknown C-16α hydroxylase (Figure 1). Zanhic acid glycosides are major metabolites in the aerial parts of the plant, but they were reported to be absent in the roots of *M. truncatula* (Confalonieri et al., 2009; Pollier et al., 2011; Lei et al., 2019). In the non-hemolytic branch, 24-hydroxy β-amyrin is the substrate of CYP72A61 that catalyzes hydroxylation at the C-22 position, leading to the major soyasaponin aglycone soyasapogenol B (Figure 1) (Fukushima et al., 2013). Soyasapogenol B can be further oxidized at the C-21 position towards soyasapogenol A (Figure 1). The P450 enzyme catalyzing this reaction has not been identified in *M. truncatula*, however, this activity was described for CYP72A69 in *Glycine max* (Yano et al., 2017; Rehman et al., 2018; Sundaramoorthy et al., 2018), a legume closely related to *M. truncatula*. After oxidation of the backbone at the various positions, the aglycones are glycosylated with one or more sugar chains. So far, only two UDP-dependent glycosyltransferases have been described in *M. truncatula*. UGT73F3 was shown to catalyze the glucosylation of the aglycones at the C-28 position and is thus only involved in the biosynthesis of hemolytic saponins (Naoumkina et al., 2010). UGT73K1 was shown to glucosylate aglycones from both the hemolytic and the non-hemolytic branches, however, the position to which the glucosyl residues were transferred was not determined (Achnine et al., 2005).

As defense compounds, *M. truncatula* saponins accumulate constitutively as tissue-specific mixtures of tens of different metabolites. Upon pathogen attack or herbivore feeding, however, the saponin biosynthesis in *Medicago* species is further enhanced (Agrell et al., 2004; Gholami et al., 2014). The concerted transcriptional activation of the saponin biosynthesis genes is mediated by a signaling cascade in which the jasmonate (JA) phytohormones play an essential role (Broeckling et al., 2005; Suzuki et al., 2005; Pollier et al., 2013; Mertens et al., 2016b). In plants, the oxylipin-derived JAs are implicated in growth and development and in biotic and abiotic stress responses (Wasternack and Hause, 2013; Goossens et al., 2016; Wasternack and Feussner, 2018). The primary JA-mediated signaling cascade leads to the activation of MYC2, a transcription factor that induces the expression of JA-responsive genes that are either directly the biosynthesis genes or downstream transcription factors (Kazan and Manners, 2013; Chini et al., 2016; Colinas and Goossens, 2018). In *M. truncatula*, the JA-responsive transcription factors TRITERPENE SAPONIN ACTIVATION REGULATOR (TSAR) 1 and TSAR2 were shown to trigger the concerted transcriptional activation of the non-hemolytic and hemolytic saponin biosynthesis genes, respectively (Mertens et al., 2016b). However, the exact position of the TSAR transcription factors in the JA signaling cascade relative to MYC2 remains elusive.

Here we report the characterization of TSAR3, a seed-specific transcription factor that positively regulates hemolytic saponin biosynthesis in *M. truncatula*. A co-expression analysis with *TSAR3* in transcriptome datasets from developing *M. truncatula* seeds led to the identification and functional characterization of CYP88A13, a cytochrome P450 that catalyzes the final oxidation step of the hemolytic saponin biosynthesis branch in *M. truncatula*, the C-16α hydroxylation of medicagenic acid towards zanhic acid. Furthermore, an extended co-expression analysis revealed two uridine diphosphate glycosyltransferases, UGT73F18 and UGT73F19, which were shown to glucosylate hemolytic sapogenins at the C-3 position when expressed in yeast and tobacco.

RESULTS

TSAR3, a Seed-Specific Regulator of Triterpene Saponin Biosynthesis

The bHLH transcription factors TSAR1 and TSAR2 positively regulate the production of non-hemolytic and hemolytic triterpene saponins, respectively, in the model legume *M. truncatula*

(Mertens et al., 2016b). The *M. truncatula* Gene Expression Atlas (MtGEA; http://bioinfo.noble.org/gene-atlas; He et al. (2009)) revealed that a third paralogue of these TSAR transcription factors, *Medtr2g104650*, which we named *TSAR3*, is exclusively expressed during seed development, with a transient expression profile peaking mid-maturation that correlates well with the expression profile of *HMGR1* (Pearson's correlation coefficient: 0.91) and *MKB1* (Pearson's correlation coefficient: 0.87) in developing seeds (Supplemental Figure 1). Previously, we have shown that TSAR1 and TSAR2 were capable of transactivating reporter constructs in which the 1,000-bp promoter regions of *HMGR1* (*ProHMGR1*) and *MKB1* (*ProMKB1*) were fused to the *FIREFLY LUCIFERASE* (*fLUC*) gene (Mertens et al., 2016b). Transient expression assays (TEAs) in tobacco (*Nicotiana tabacum*) protoplasts revealed that TSAR3 was capable of inducing the luciferase activity of the *ProHMGR1* and *ProMKB1* reporter constructs, by 20- and 32-fold, respectively (Figures 2A and 2B), which is similar to the effect of TSAR2. Furthermore, TSAR3 was also capable of inducing the luciferase activity of a reporter construct with the 1,000-bp promoter region of *β-AMYRIN SYNTHASE* (*ProBAS*; Figure 2C), confirming TSAR3 as a potential regulator of *M. truncatula* triterpene saponin biosynthesis.

Because the TSAR1 and TSAR2 transcription factors differentially regulate the non-hemolytic and hemolytic triterpene saponin biosynthesis branches, TEAs with TSAR3 and reporter constructs with the promoter regions of P450s from both biosynthetic branches were carried out. Like TSAR2, TSAR3 was capable of inducing the luciferase activity of reporter constructs with the promoters of *CYP72A67* (*ProCYP72A67*) and *CYP72A68* (*ProCYP72A68*; Figures 2D and 2E), corresponding to P450s involved in the biosynthesis of hemolytic triterpene saponins (Fukushima et al., 2013; Biazzi et al., 2015). TSAR3 was also capable of inducing the luciferase activity of reporter constructs with the promoters of *CYP93E2* (*ProCYP93E2*) and *CYP72A61* (*ProCYP72A61*; Figures 2F and 2G), corresponding to P450s involved in the biosynthesis of non-hemolytic triterpene saponins (Fukushima et al., 2011; Fukushima et al., 2013). However, the activation of these promoters by TSAR3 was markedly lower as compared to that by TSAR1, the known regulator of the non-hemolytic saponin biosynthesis branch, indicating that the non-hemolytic branch is likely not the preferred target of TSAR3, as was also the case for TSAR2 (Mertens et al., 2016b). Taken together, these data indicate that TSAR3 is a potential regulator of the triterpene saponin biosynthesis pathway in *M. truncatula*, with a preference for the hemolytic branch.

Overexpression of *TSAR3* Boosts Hemolytic Triterpene Saponin Biosynthesis in *M. truncatula* Hairy Roots

To examine the functionality of TSAR3 *in planta*, three independent *M. truncatula* hairy root lines ectopically overexpressing *TSAR3* (TSAR3^{OE}) were generated. As control, three independent *M. truncatula* hairy root lines that express *GUS* were used. According to previously generated RNA-Seq data (Mertens et al., 2016b), the seed-specific *TSAR3* is not expressed in control *M. truncatula* hairy roots. Thus, expression of *TSAR3* in the TSAR3^{OE} lines was verified by reverse transcription PCR (RT-PCR; Figure 3A), which confirmed ectopic *TSAR3* expression in the TSAR3^{OE} lines, and absence thereof in the control hairy root lines.

Quantitative RT-PCR (qRT-PCR) analysis of triterpene saponin biosynthesis genes in the TSAR3^{OE} lines revealed increased *HMGR1* expression levels (Figure 3B), whereas no effect on *MKB1* expression was observed (Figure 3C). Furthermore, significant elevations of *BAS*, *CYP716A12*, *CYP72A67*, *CYP72A68*, and *UGT73F3* transcript levels were observed in the TSAR3^{OE} lines (Figures 3D to 3H). In contrast, transcript levels of *CYP93E2*, *CYP72A61*, and *UGT73K1* did not increase, but rather decreased in the TSAR3^{OE} lines (Figures 3I to 3K). Overall, this pattern is reminiscent of that of TSAR2^{OE} lines (Mertens et al., 2016b), and indicative of a specific transcriptional upregulation of the hemolytic triterpene saponin biosynthesis branch by overexpression of *TSAR3*.

To explore the effect of *TSAR3* expression at the metabolite level, untargeted metabolite profiling of the three independent TSAR3^{OE} and GUS control lines was carried out by LC-MS. Peak integration and alignment yielded a total of 17,922 *m*/*z* features that were log-transformed and pareto-scaled for a principal component analysis (PCA). This PCA showed that expression of *TSAR3* explained most of the variation (44.2 %) in the dataset as the first principal component (PC1) clearly separated the TSAR3^{OE} and GUS control lines (Figure 4A, Supplemental Figure 2). Within the 100 most abundant *m*/*z* features of the metabolite profiling dataset, 41 *m*/*z* features, corresponding to 37 metabolites, were over fivefold more abundant or discretely present in the TSAR3^{OE} lines, and four were over fivefold more abundant in the GUS control lines. The metabolites corresponding to these differential *m*/*z* features were identified based on their accurate mass, MS^n and MS/MS fragmentation spectra and available standards, thereby revealing that ectopic expression of *TSAR3* leads to the accumulation of hemolytic triterpene saponins in the

M. truncatula hairy roots, whereas the abundance of the major soyasaponins remained unaltered (Figures 4B to 4G, Supplemental Dataset 1).

Within the set of compounds that were unique to the TSAR3^{OE} lines, two types of saponins could be distinguished. First, several zanhic acid glycosides were shown to accumulate in the TSAR3^{OE} lines. Zanhic acid is a major aglycone in the aerial parts of *M. truncatula* (Kapusta et al., 2005a; Kapusta et al., 2005b), but it was reported to be absent in *M. truncatula* roots and hairy roots (Confalonieri et al., 2009; Pollier et al., 2011; Lei et al., 2019). In addition, also several of the major medicagenic acid glycosides normally occurring in the aerial parts of the plant, but not in roots, were observed in the TSAR3^{OE} lines. These metabolites include 3-Glc-28-Ara-Rha-Xyl-medicagenic acid and 3-GlcA-28-Ara-Rha-Xyl-medicagenic acid that were confirmed with authentic standards. Interestingly, the latter metabolite is the dominant saponin in *M. truncatula* seeds (Huhman et al., 2005). Thus, ectopic expression of *TSAR3* in *M. truncatula* hairy roots leads to transcriptional activation of the hemolytic saponin biosynthesis pathway and the accumulation of hemolytic saponins that are normally restricted to the aerial parts of the plant.

TSAR3 Regulates Hemolytic Triterpene Saponin Biosynthesis in Developing *M. truncatula* Seeds

To further validate the role of TSAR3 *in planta*, we screened a *M. truncatula Tnt1* retrotransposon insertion population (Tadege et al., 2008) for *TSAR3* loss-of-function mutants. From the *Tnt1* flanking sequence tag (FST) database (https://medicago-mutant.noble.org/mutant/index.php), three candidate mutant lines were selected: NF14212, NF15672, and NF17107. Homozygous insertion mutants were obtained for all three insertion lines, with a retrotransposon insertion in the first exon of *TSAR3* for *Tnt1* insertion line NF15672 and in the second exon for *Tnt1* insertion lines NF14212 and NF17107 (Figure 5A; Supplemental Figure 3). For each insertion line, homozygous escape lines were obtained in which both *TSAR3* alleles are intact (Supplemental Figure 3). In accordance with the *M. truncatula* Gene Expression Atlas, RT-PCR analysis confirmed that *TSAR3* expression is restricted to the developing seeds of the wild-type plants (Supplemental Figure 4).

qRT-PCR analysis with primers spanning the *Tnt1* retrotransposon revealed that transcript levels of *TSAR3* were drastically reduced or even undetectable in developing seeds of the insertion lines (Figure 5B-C). Furthermore, a drastic decrease of transcript levels of *HMGR1*, *BAS* and the hemolytic triterpene saponin biosynthesis genes *CYP716A12*, *CYP72A67*, *CYP72A68*, and

UGT73F3 was observed. Transcript levels of genes involved in soyasaponin biosynthesis, *CYP93E2*, *CYP72A61*, and *UGT73K1*, slightly increased, whereas the expression levels of *MKB1* were hardly affected (Figures 5D to 5M). In leaves of the same plants, no major effect on the transcript levels of the saponin biosynthesis genes was observed (Supplemental Figure 4). Together, these data implicate a specific transcriptional downregulation of the hemolytic triterpene saponin biosynthesis branch in the seeds of *tsar3* plants, hence supporting the role of TSAR3 as a seed-specific regulator of hemolytic triterpene saponin biosynthesis.

Untargeted metabolite profiling of developing seeds and leaves of wild-type and homozygous *tsar3* plants was carried out by LC-MS and yielded a total of 15,819 *m*/*z* features. A Principal component analysis revealed a clear separation of the samples derived from wild-type and *tsar3* seeds (Figure 6A), whereas no such separation was observed for samples derived from leaves of the same plants (Figure 6B), indicating a seed-specific effect in *tsar3* plants. In accordance with the transcriptome data, the levels of the major hemolytic saponins in the seeds were drastically reduced, whereas soyasaponin levels remained unaltered. No effects on the accumulation of hemolytic or non-hemolytic saponins was observed in the leaves (Figures 6C to 6H). To identify the metabolites that contribute most to the difference between the seeds of *tsar3* and wild-type plants, we carried out a Partial least squares discriminant analysis (PLS-DA) with the samples derived from developing seeds (Figure 6I). The 50 *m*/*z* features that contribute most to the separation of the *tsar3* and wild-type plants are listed in Supplemental Dataset 2. Identification of the corresponding metabolites revealed that most of these *m*/*z* features correspond to hemolytic saponins that are present in developing wild-type seeds, but are absent in developing *tsar3* seeds. Together, these data corroborate a specific role for TSAR3 in the regulation of the hemolytic saponin biosynthesis in developing *M. truncatula* seeds.

Selection of Candidate P450s Involved in Hemolytic Triterpene Saponin Biosynthesis

As the ectopic expression of the seed-specific *TSAR3* in *M. truncatula* hairy roots leads to the production of zanhic acid glycosides, we reasoned that this transcription factor might also regulate the yet unknown C-16α hydroxylase that catalyzes the oxidation of medicagenic acid to zanhic acid in *M. truncatula*. To identify candidate P450s that might carry out C-16α hydroxylation, two co-expression analyses were carried out on microarray datasets of developing *M. truncatula* seeds. First, we calculated the Pearson's correlation coefficient between the probeset corresponding to

TSAR3 and all 19,012 differentially expressed probesets during *M. truncatula* seed development (Verdier et al., 2013). Within the set of 88 probesets with a Pearson's correlation coefficient above 0.95, probesets corresponding to all characterized P450s of the hemolytic triterpene saponin branch were present (Supplemental Dataset 3). Remarkably, also a probeset corresponding to *CYP72A65*, encoding a P450 shown to catalyze the C-30 oxidation of oleanolic acid to queretaroic acid (Seki et al., 2011; Reed et al., 2017), was present in this dataset. Furthermore, also probesets corresponding to more upstream triterpene saponin biosynthesis genes, such as *BAS* and *MEVALONATE KINASE* were present in this set, implying a tight co-regulation of the hemolytic triterpene saponin biosynthesis genes and *TSAR3* in developing seeds. Within these highly co-regulated probesets, several probesets corresponding to P450s that might be the missing C-16α hydroxylase are present.

A second co-expression analysis was carried out on a dataset consisting of 60 transcriptomes from developing seeds that were produced under different environmental conditions (Righetti et al., 2015) (Supplemental Dataset 4). Also in this dataset, the Pearson's correlation coefficient between the probeset corresponding to *TSAR3* and all other probesets was calculated. Within the set of 121 probesets with a Pearson's correlation coefficient above 0.85, probesets corresponding to *CYP716A12* and *HMGR1* were present, however, probesets corresponding to *CYP72A67* and *CYP72A68* were absent. This implies a less strict co-regulation of the hemolytic triterpene saponin biosynthesis genes within this dataset. In both datasets, *Medtr5g014240* was the P450 that had the highest co-expression level with *TSAR3*. As such, this gene was considered our primary candidate C-16α hydroxylase. Next to *Medtr5g014240*, twelve additional P450s were present in one or both datasets. As several of these candidates corresponded to incomplete genes in the genome or genes that were not expressed in *M. truncatula* shoots, which are known to accumulate zanhic acid glycosides (Kapusta et al., 2005a; Kapusta et al., 2005b), the list of candidate genes was further reduced. Finally, from the remaining candidates we were only able to successfully clone the full-length sequences of three genes, the top-candidate *Medtr5g014240* (*CYP88A13*), as well as *Medtr5g023680* (*CYP83E45*) and *Medtr8g035780* (*CYP704G5*).

Functional Characterization of Candidate P450s in Yeast and *Nicotiana benthamiana*

In *Medicago* species, only two compounds with a hydroxylation at the C-16α position have been reported, caulophyllogenin and zanhic acid, which are oxidation products of hederagenin and

medicagenic acid, respectively (Tava et al., 2011; Gholami et al., 2014). In spite of its occurrence in *M. polymorpha*, caulophyllogenin has not been reported as a saponin aglycone in *M. truncatula* (Tava et al., 2011). Hence, we assumed medicagenic acid to be the substrate of the unknown C-16α hydroxylase in *M. truncatula*. To screen for a functional C-16α hydroxylase in yeast, we therefore first engineered a yeast strain that produces sufficiently high levels of medicagenic acid. This was achieved by co-expression of the genes encoding β-amyrin synthase from *Glycyrrhiza glabra* (GgbAS), cytochrome P450 reductase from *M. truncatula* (MtCPR1), and the C-2 (CYP72A67), C-23 (CYP72A68) and C-28 (CYP716A12) oxidases from *M. truncatula* in the starter strain JP034 (Supplemental Table 2). JP034 is a BY4742-derived yeast strain in which an additional auxotrophic marker (TRP1) and a disruption of the phosphatidic acid phosphatase-encoding *PAH1* was introduced using CRISPR/Cas9. Disruption of *PAH1* leads to a dramatic expansion of the endoplasmic reticulum, which stimulates the production of recombinant triterpene biosynthesis enzymes and ultimately results in an increased heterologous triterpenoid production (Arendt et al., 2017; Bicalho et al., 2019). Furthermore, the native lanosterol synthase (*ERG7*) promoter was replaced by a methionine-repressible promoter in strain JP034, allowing downregulation of *ERG7* by addition of methionine to the cultivation medium. This downregulation of *ERG7* leads to a higher β-amyrin production in yeast, presumably by increased 2,3-oxidosqualene availability for the expressed BAS (Kirby et al., 2008; Moses et al., 2014).

To identify the functional C-16α hydroxylase, the candidate P450s were expressed in the generated medicagenic acid-producing yeast strain. Compared to an empty vector control, the GC-MS chromatogram of an extract of the spent medium of the medicagenic acid-producing yeast strain expressing *CYP88A13* showed a single new peak at retention time 51.50 min. (Figure 7A). To confirm that this new peak corresponds to zanhic acid, the spent medium of the medicagenic acid-producing yeast strain expressing *CYP716Y1*, a C-16α hydroxylase from the medicinal plant *Bupleurum falcatum* (Moses et al., 2014), was analyzed by GC-MS. A peak with the same retention time and EI-MS fragmentation spectrum as the unique peak in the CYP88A13 chromatogram was visible (Figures 7A and 7B). However, in contrast to CYP88A13, several additional unique peaks were visible in the CYP716Y1 GC-MS chromatogram (Figure 7A, Supplemental Figure 5). These peaks likely correspond to other oleananes with a C-16α hydroxyl group, indicating that CYP88A13, unlike CYP716Y1, only recognizes the highly oxidized medicagenic acid as a substrate.

In addition to yeast, the candidate P450s were also expressed in combination with *GgbAS*, *CYP72A67*, *CYP72A68* and *CYP716A12* in *N. benthamiana* leaves by agroinfiltration. Methanolic extracts of leaves harvested three days after agroinfiltration were analyzed by GC-MS. Again, compared to the empty vector control or any of the other infiltrated candidate P450s, zanhic acid was only present in the GC-MS chromatograms of *N. benthamiana* leaves infiltrated with *CYP88A13* and the previously characterized *B. falcatum* C-16α hydroxylase *CYP716Y1* (Figures 7C and 7D). Taken together, these data indicate that CYP88A13 is a functional medicagenic acid C-16α hydroxylase when expressed in yeast or *N. benthamiana*.

CYP88A13 Catalyzes the Final Oxidation Step in the Hemolytic Saponin Biosynthesis Pathway

In spite of the occurrence of caulophyllogenin in *M. polymorpha* (Tava et al., 2011), the only reported sapogenin with a C-16α hydroxy group in *M. truncatula* is zanhic acid (Tava et al., 2011; Gholami et al., 2014). To investigate whether the identified C-16α hydroxylase CYP88A13 is capable of oxidizing intermediates of the triterpene saponin biosynthesis pathway other than medicagenic acid, we expressed it in yeast in combination with enzymes leading to the production of biosynthetic intermediates. First, the capacity of CYP88A13 to oxidize β-amyrin was investigated. Upon expression of *CYP88A13* in a β-amyrin-producing yeast strain, no 16α-hydroxy β-amyrin was produced, whereas expression of *CYP716Y1* led to the accumulation of 16α-hydroxy β-amyrin (Figure 8A; Supplemental Figure 6), indicating that β-amyrin is likely not a substrate of CYP88A13. Likewise, in contrast to expression of *CYP716Y1*, expression of *CYP88A13* in an oleanolic acid-producing yeast strain failed to yield echinocystic acid (Figure 8B; Supplemental Figures 7 and 8), excluding oleanolic acid as a substrate of CYP88A13. Similarly, no 2β-hydroxy echinocystic acid accumulated upon expression of *CYP88A13* in an augustic acid-producing yeast strain (Figure 8C; Supplemental Figure 9). Finally, *CYP88A13* was also expressed in a yeast strain that accumulates gypsogenic acid. Like for the positive control *CYP716Y1*, expression of *CYP88A13* in this strain led to the production of 16α-hydroxy gypsogenic acid (Figure 8D; Supplemental Figures 7 and 10), implying gypsogenic acid may serve as a substrate for CYP88A13. The presence of 16α-hydroxy gypsogenic acid has been observed in the leaves of a *M. truncatula* mutant lacking CYP72A67 activity (Biazzi et al., 2015), further supporting that gypsogenic acid may indeed serve as a substrate of the identified CYP88A13. However, as

gypsogenic acid is only present in mutants lacking CYP72A67 activity, this metabolite may not serve as a substrate in *M. truncatula*. Similar results were obtained upon infiltration of *N. benthamiana* leaves with *CYP88A13* and the enzymes leading to the various intermediates (Supplemental Figure 11).

Taken together, these data point towards the following order of oxidation of the β-amyrin backbone in the hemolytic saponin biosynthesis in *M. truncatula*. First, β-amyrin is oxidized at its C-28 position by CYP716A12 to yield oleanolic acid (Carelli et al., 2011; Fukushima et al., 2011). Next, oleanolic acid is concomitantly oxidized at the C-2 and C-23 positions by CYP72A67 and CYP72A68, respectively (Fukushima et al., 2013; Biazzi et al., 2015; Tzin et al., 2019), thereby yielding medicagenic acid in addition to reaction intermediates such as hederagenin and bayogenin. Finally, medicagenic acid is hydroxylated at its C-16α position by CYP88A13, leading to zanhic acid (Figure 1).

The Expression of *CYP88A13* Is Regulated by TSAR3

As *CYP88A13* is strictly co-regulated with *TSAR3* in developing *M. truncatula* seeds (Supplemental Datasets 3 and 4), we next investigated whether TSAR3 can regulate the expression of *CYP88A13*. First, RT-PCR analysis revealed expression of *CYP88A13* in the TSAR3^{OE} lines, but not in the GUS control lines (Figure 9A), implying that, like for the other hemolytic P450s, TSAR3 can control the expression of *CYP88A13* in *M. truncatula* hairy roots. Furthermore, transactivation assays in tobacco protoplasts with different *CYP88A13* promoter (*ProCYP88A13*) constructs were carried out. In a first assay, TSAR3 failed to induce the luciferase activity of a reporter construct containing the 1,000-bp region upstream of the start codon of *CYP88A13* (Figure 9B). A more detailed analysis of the *ProCYP88A13* revealed an N-box (5'-CACGAG-3'), a promoter element required for TSAR binding (Mertens et al., 2016b), 2,051 bp upstream of the *CYP88A13* start codon. Furthermore, three additional N-box-like sequences were located within 50 bp upstream of the first identified N-box, among which an E-box (5'CACGCG-3'), 2,096 bp upstream of the *CYP88A13* start codon. Hence, a 2,118-bp *ProCYP88A13* fragment was cloned and used for TEAs with TSAR3, revealing a three-fold induced luciferase activity compared to protoplasts co-transfected with a GUS control (Figure 9C). Furthermore, in line with its role in hemolytic saponin biosynthesis, *ProCYP88A13* was also transactivated by TSAR2. The regulator of soyasaponin biosynthesis, TSAR1, could not transactivate *ProCYP88A13* (Figure 9C). Finally,

the expression of *CYP88A13* was considerably lower in developing seeds of the three *tsar3 Tnt1* insertion lines (Figure 9D), implying transcriptional control of *CYP88A13* by TSAR3 in developing *M. truncatula* seeds. No effect on *CYP88A13* expression was observed in leaves of these insertion lines (Figure 9E), in line with the absence of *TSAR3* expression in leaves of wild-type plants (Supplemental Figure 4). Taken together, these data imply that TSAR3 stimulates expression of *CYP88A13* in developing *M. truncatula* seeds.

Functional Characterization of *CYP88A13* in *M. truncatula*

As CYP88A13 was shown to catalyze the C-16α hydroxylation of medicagenic acid in yeast and *N. benthamiana*, we postulated that ectopic expression of *CYP88A13* in *M. truncatula* hairy roots would lead to the accumulation of zanhic acid glycosides. To verify this hypothesis, we generated three *M. truncatula* hairy root lines expressing *CYP88A13* (CYP88A13^{OE}). As control, three *M. truncatula* hairy root lines expressing *GUS* were used. RT-PCR analysis confirmed ectopic expression of *CYP88A13* in the CYP88A13^{OE} but not in the GUS control hairy root lines (Figure 10A) and metabolite profiling by LC-MS confirmed that zanhic acid glycosides, such as 3-Glc-zanhic acid, were accumulating in the CYP88A13^{OE} lines (Figure 10B).

To further validate the role of CYP88A13 *in planta*, we screened the *M. truncatula Tnt1* retrotransposon insertion population (Tadege et al., 2008) for *CYP88A13* loss-of-function mutants. Like for TSAR3, three candidate mutant lines were selected from the *Tnt1* FST database: NF11894, NF17520, and NF13195. Homozygous insertion mutants were obtained for line NF17520, with a retrotransposon insertion in intron 2 of the 5652 bp *CYP88A13* gene, and line NF11894, with a *Tnt1* insertion in the first intron of the gene (Figure 10C; Supplemental Figure 12). Leaves of homozygous plants of both insertion lines were investigated by LC-MS, revealing that both lines still accumulate zanhic acid glycosides (Figure 10D). RT-PCR analysis with primers spanning the *Tnt1* insertion revealed that the *Tnt1*-containing introns get spliced out, as a result of which *CYP88A13* expression is not affected (Figure 10E).

Inspection of the *M. truncatula* genome v4.0 revealed that *CYP88A13* is part of a tandem repeat with *CYP88A14* (*Medtr5g014250*) and a third, seemingly incomplete pseudogene *Medtr5g014230* that is the result of a 1,628-bp genomic repeat that overlaps with *CYP88A13* (Figure 10F). Since CYP88A13 and CYP88A14 share 77% amino acid identity, we speculated that also *CYP88A14* may encode a functional C-16α hydroxylase. To validate this hypothesis, we

expressed *CYP88A14* in the medicagenic acid-producing yeast strain and in *N. benthamiana* leaves in combination with *GgbAS*, *CYP72A67*, *CYP72A68* and *CYP716A12*. In both systems, expression of *CYP88A14* failed to lead to the production of zanhic acid, but instead led to the production of two unknown metabolites (Supplemental Figure 13), indicating that *CYP88A14* encodes a functional protein, but is not involved in the biosynthesis of zanhic acid.

Selection of Candidate UGTs Involved in Hemolytic Triterpene Saponin Biosynthesis

Next to the production of zanhic acid glycosides, ectopic expression of *TSAR3* in *M. truncatula* hairy roots also led to the accumulation of several of the major medicagenic acid glycosides that normally only occur in the aerial parts of the plant. We thus reasoned that this transcription factor might also regulate the UDP-dependent glycosyltransferases (UGTs) responsible for glycosylation of the medicagenic acid aglycone. The UGT-encoding genes that were co-expressed with TSAR3 in the two co-expression analyses that led to the identification of CYP88A13 (Supplemental Datasets 3 and 4) were considered candidate UGTs involved in hemolytic saponin biosynthesis. This list was expanded with UGT candidates from a third co-expression analysis in which the *M. truncatula* gene expression atlas (He et al., 2009) was probed for UGTs co-expressed with *CYP716A12*. UGT-encoding genes with a Pearson's correlation coefficient above 0.6 were considered as candidates. Combined, these three co-expression analyses led to a comprehensive set of 20 candidate UGTs, of which we were able to clone ten genes for further analysis (Supplemental Table 2).

Functional characterization of candidate UGTs in yeast and *N. benthamiana*

To identify functional UGTs using *S. cerevisiae,* the UGT candidates were expressed in the yeast strain engineered for the accumulation of medicagenic acid and used for the identification of CYP88A13. Next to medicagenic acid, this yeast strain also accumulates pathway intermediates such as oleanolic acid, hederagenin, and bayogenin, thus providing a whole array of potential substrates with different degrees of oxidation of the β-amyrin backbone. As a positive control, the C-3 glucosyltransferase UGT73C11 (Augustin et al., 2012) and the C-28 glucosyltransferase UGT73F3 (Naoumkina et al., 2010) were used. LC-MS analysis of the cell extracts of the strains expressing the different UGTs revealed the production of glycosides for the strains expressing two

of the candidate UGTs and for the strains expressing the positive controls *UGT73C11* and *UGT73F3* (Figures 11A and 11B).

Expression of *UGT73F19* in the medicagenic acid-producing yeast strain led to the production of two major glycosides (Figure 11A). Based on their accurate mass, MS^n fragmentation spectra and an available standard, these metabolites were identified as 3-Glc-medicagenic acid and 3-Glc-gypsogenic acid. Both metabolites were also present in the extract of a yeast strain expressing the previously characterized C-3 glucosyltransferase *UGT73C11* (Figure 11A). Thus, like UGT73C11, UGT73F19 catalyzes the 3-*O*-glucosylation of the sapogenins medicagenic acid and gypsogenic acid when expressed in yeast. Unlike UGT73C11, however, UGT73F19 did not accept substrates with a lower degree of oxidation of the β-amyrin backbone.

Expression of *UGT73F18* in the medicagenic acid-producing yeast strain resulted in the production of a whole array of different glycosides that were absent in the control strain (Figure 11B). Based on their accurate mass, MS^n fragmentation spectra and an available standard, these metabolites were identified as 3-*O*-glucosides of medicagenic acid (1), hederagenin (3), polygalagenin (4), gypsogenin (5), augustic acid (6), and oleanolic acid (7). In addition, small amounts 3-Glc-bayogenin and 3-Glc-gypsogenic acid were co-migrating with 3-Glc-medicagenic acid and 3-Glc-hederagenin, respectively. Thus, like UGT73C11, but in contrast to UGT73F19, UGT73F18 also catalyzes the 3-*O*-glucosylation of sapogenins with a lower degree of oxidation of the triterpenoid backbone, with oleanolic acid being the least oxidized substrate that is still accepted by both UGTs upon expression in yeast.

In addition to yeast, the candidate UGTs were also expressed in combination with *GgbAS*, *CYP716A12*, *CYP72A67*, and *CYP72A68* in *N. benthamiana* leaves by agroinfiltration. Methanolic extracts of leaves harvested four days after agroinfiltration were analyzed by LC-MS and confirmed the production of 3-*O*-glucosides of medicagenic acid and gypsogenic acid upon infiltration with *UGT73F19* (Figure 11C). Remarkably, only trace amounts of 3-Glc-medicagenic acid were observed in leaves infiltrated with *UGT73C11*, whereas 28-Glc-medicagenic acid accumulated. As the accumulation of C-28 glucosides was also observed when no candidate UGT was infiltrated and as they were also observed in a previous study in which *UGT73C11* was infiltrated in *N. benthamiana* leaves (Khakimov et al., 2015), we postulated that *N. benthamiana* encodes a UGT that catalyzes the C-28 glycosylation of triterpenoid aglycones and that competes

for the triterpenoid aglycone substrates with the expressed UGTs. Upon expression of *UGT73F19*, hardly any 28-Glc-medicagenic acid was observed (Figure 11C), pointing to a likely higher affinity of UGT73F19 for the medicagenic acid substrate. Infiltration of *UGT73F18* in *N. benthamiana* leaves resulted in the accumulation of both 28-Glc-medicagenic acid and 3-Glc-medicagenic acid in addition to small amounts of 3-Glc-hederagenin and 3-Glc-polygalagenin, suggesting a slightly lower affinity of this glucosyltransferase for the sapogenin substrates as compared to UGT73F19. Taken together, the above data suggest that both *UGT73F19* and *UGT73F18* encode functional sapogenin C3 glucosyltransferases when expressed in yeast and *N. benthamiana*, with UGT73F18 having a more relaxed substrate tolerance as compared to UGT73F19.

UGT73F18 and *UGT73F19* Are Part of a Cluster of Duplicated Genes

The two newly discovered glucosyltransferases *UGT73F18* and *UGT73F19*, encoded by *MTR_120s0024* and *MTR_120s0028*, respectively, are part of a gene cluster encoding six highly homologous UGTs spread over a range of 57 kb of the *M. truncatula* genome (Figure 12A). This cluster of duplicated genes is located on sequence contig 120s of the *M. truncatula* genome release Mt3.5 (Young et al., 2011) but can unfortunately no longer be encountered in the more recent Mt4.0 genome release (Tang et al., 2014). The duplicated UGTs show a high nucleotide sequence similarity of 92% and of 88% at the amino acid level (Figure 12B). As both UGTs carry out the same reaction, *i.e.*, C-3 glucosylation of triterpenoid sapogenins, it is plausible that other members of this gene cluster also encode functional C3 glucosyltransferases.

Compared to *CYP88A13*, the degree of co-regulation between the two characterized UGTs and *TSAR3* is less pronounced. A qRT-PCR analysis of *UGT73F18* and *UGT73F19* in the TSAR3^{OE} lines revealed that also these genes seem regulated by TSAR3. However, the transcriptional upregulation is only about fourfold (Figure 12C), which is comparable to the upregulation of the C-28 glucosyltransferase UGT73F3, but less pronounced than the P450s involved in the hemolytic saponin biosynthesis (Figure 2). Transactivation assays in tobacco protoplasts, however, show a substantial activation of the *UGT73F18* (*ProUGT73F18*) and *UGT73F19* (*ProUGT73F19*) promoters by all three TSARs (Figure 12D).

DISCUSSION

To protect itself from pathogen attack or herbivore predation, *M. truncatula* accumulates organ-specific blends of triterpene saponins, of which the biosynthesis is regulated in a JA-dependent manner. Here, we report the identification of TSAR3, a seed-specific transcription factor that regulates hemolytic saponin biosynthesis in developing *M. truncatula* seeds. Extensive co-expression analyses with *TSAR3* in transcriptome datasets from developing *M. truncatula* seeds led to the identification and functional characterization of CYP88A13, which catalyzes the final oxidation step of the hemolytic saponin biosynthesis branch in *M. truncatula*, the C-16α hydroxylation of medicagenic acid towards zanhic acid. Furthermore, the co-expression analyses revealed UGT73F18 and UGT73F19, which were shown to glucosylate hemolytic sapogenins at the C-3 position when expressed in yeast and tobacco. Together with the previously characterized hemolytic saponin biosynthesis genes, these newly characterized biosynthesis genes form a TSAR3-regulated regulon in developing *M. truncatula* seeds.

Recruitment of Clade IVa bHLH Transcription Factors for Metabolic Pathway Regulation

TSAR3 belongs to the clade IVa bHLH transcription factors, which encompass several, often JA-responsive, transcription factors involved in the regulation of diverse metabolic pathways in different plant species (Goossens et al., 2017; Shoji, 2019). Like TSAR1 and TSAR2 (Mertens et al., 2016b), TSAR3 was shown to be involved in the regulation of triterpene saponin biosynthesis in *M. truncatula*. Like TSAR1, the JA-responsive TSAR ortholog GubHLH3 regulates soyasaponin biosynthesis in *Glycyrrhiza uralensis* (Tamura et al., 2018), implying conservation of the role of TSAR transcription factors in legumes. Similarly, the TSAR-Like (TSARL) transcription factors TSARL1 and TSARL2 regulate saponin biosynthesis in *Chenopodium quinoa* (quinoa), with TSARL1, like TSAR3, being exclusively expressed in the seeds where it controls the synthesis of hemolytic, bitter-tasting saponins (Jarvis et al., 2017). Belonging to the amaranth family, quinoa is not related to the legumes *M. truncatula* and *G. uralensis*, pointing to a possible conservation of the role of clade IVa bHLH transcription factors in the regulation of saponin biosynthesis beyond the legume family. However, not all dicots accumulate saponins, and in *Catharanthus roseus*, the JA-responsive clade IVa bHLH transcription factors BIS1 and BIS2 were shown to regulate monoterpenoid indole alkaloid biosynthesis (Van Moerkercke et al., 2015; Mertens et al., 2016a; Van Moerkercke et al., 2016). It thus seems more likely that the transcriptional control of the saponin biosynthesis by TSAR or TSARL transcription factors in

M. truncatula and quinoa is a case of convergent evolution. By gaining cis-regulatory elements in the promoters of the biosynthetic genes, metabolic pathways could be recruited into a regulon controlled by the clade IVa bHLH transcription factors (Shoji, 2019) that are often an integral part of the conserved JA signaling cascade.

The TSAR Transcription Factors Determine the Tissue-Specific Saponin Blend

With TSAR3, we characterized a third clade IVa bHLH transcription factor involved in the regulation of triterpene saponin biosynthesis in *M. truncatula*. Upon inspection of the *M. truncatula* genome (Tang et al., 2014), we could identify a staggering number of 30 additional clade IVa bHLH transcription factors (Supplemental Figure 14). Based on the role of the three characterized TSAR transcription factors, we speculate that the other TSAR transcription factors may also be capable of inducing (a specific branch of) the triterpene saponin biosynthesis pathway in *M. truncatula* in (specific) responses to environmental and developmental cues. As such, their specific expression could allow the plant to fine-tune the composition of its triterpene saponin blend depending on the specific needs of the moment and/or the organ.

An important question that remains is how the TSAR transcription factors differentiate between the hemolytic and the non-hemolytic saponin biosynthesis branches. Like TSAR2, TSAR3 controls the hemolytic triterpene saponin biosynthesis, yet, based on the phylogeny of its characteristic bHLH domain (Supplemental Figure 14), TSAR3 seems more closely related to TSAR1, a regulator of the non-hemolytic soyasaponin biosynthesis (Mertens et al., 2016b). Thus, the DNA-binding bHLH domain itself is seemingly not responsible for the specificity of the TSAR transcription factors. In addition to the bHLH signature domain, clade IVa transcription factors also contain an ACT-like domain that is required for the formation of homo- and heterodimers (Van Moerkercke et al., 2016). It is therefore possible that interaction with other proteins, not necessarily clade IVa bHLH transcription factors, may determine the specificity towards the specific saponin branches.

A Conserved Role for CYP88 Proteins in Triterpenoid Biosynthesis

Strictly co-regulated with *TSAR3* in developing *M. truncatula* seeds is *CYP88A13* that encodes a cytochrome P450 that catalyzes the C-16α hydroxylation of medicagenic acid towards zanhic acid. Previously, C-16α hydroxylation of triterpene backbones was also shown for cytochromes P450

belonging to the CYP716 and CYP87 families (Moses et al., 2014; Moses et al., 2015; Miettinen et al., 2017). This makes the CYP88 family the third P450 family with the inherent capacity to catalyze the C-16α hydroxylation of triterpenoids. The discovery that a CYP88 family member is responsible for this reaction suggests a conserved role of CYP88 proteins in the oxidation of triterpene backbones. Indeed, two other members of this family were reported to be involved in triterpenoid biosynthesis. *G. uralensis* CYP88D6 and *Cucumis sativus* (cucumber) CYP88L2 were shown to carry out the two-step oxidation of β-amyrin at the C-11 position to produce 11-oxo-β-amyrin, and the C-19 hydroxylation of cucurbitadienol during cucurbitacin biosynthesis, respectively (Seki et al., 2008; Shang et al., 2014). Hence, like the CYP716 family (Miettinen et al., 2017), the CYP88 family also seems to be an important P450 family in the contribution to triterpenoid diversity. However, in contrast to the CYP716 family, which seems evolved specifically towards triterpenoid biosynthesis (Miettinen et al., 2017), CYP88 family members are mostly considered *ent*-kaurenoic acid oxidases involved in gibberellin biosynthesis (Bak et al., 2011). Because of the structural similarity around the C-7 position of *ent*-kaurenoic acid and the C-16 position of medicagenic acid (Supplemental Figure 15), neofunctionalization of a duplicated ent-kaurenoic acid oxidase towards a medicagenic acid C-16α hydroxylase seems plausible. A similar recruitment of a CYP88 protein from gibberellin biosynthesis was proposed for CYP88D6 (Hamberger and Bak, 2013). Thus, unlike the CYP716 family, the CYP88 family might originally not be dedicated to triterpene biosynthesis, but its reoccurring involvement in it may rather be a consequence of the structural and physicochemical similarities between diterpenes and triterpenes.

Gene Duplications as Driving Force of Molecular Evolution

Upon inspection of the *M. truncatula* genome, we noticed that *CYP88A13* appears in a tandem duplication with *CYP88A14*. Unlike *CYP88A13*, however, *CYP88A14* seems not involved in zanhic acid biosynthesis. Yet, we were able to show that also this P450 has a capacity for triterpene oxidation, although its exact catalytic activity remains to be determined. Gene duplication followed by sub- or neofunctionalization is considered a major driving force of evolutionary novelty in plant specialized metabolism (Ober, 2010). This notion is supported here with the apparent different functionality of CYP88A14 towards triterpenes compared to its tandem repeat CYP88A13. Like *CYP88A13* and *CYP88A14*, also the characterized UGTs, *UGT73F18* and *UGT73F19*, appear in a tandem repeat, in this case a repeat of six genes encoding highly similar proteins. Although these

two UGTs catalyze the C-3 glucosylation of triterpenoid sapogenins, UGT73F18 displays a more relaxed substrate tolerance compared to UGT73F19. As such, also in this case the duplicated genes seem to have evolved slightly different functions. Because of the sequence similarity between the two characterized UGTs and the other UGTs within the tandem repeat, it is likely that at least some of the other UGT-encoding genes within this tandem repeat also code for functional C-3 glycosyltransferases.

Tandem gene duplications of UGTs involved in saponin biosynthesis seem to be the rule rather than the exception. In the cruciferous plant *Barbarea vulgaris*, a gene tandem array encoding at least six functional UGTs is involved in C-3 and C-28 glucosylation of oleanolic acid and hederagenin (Erthmann et al., 2018). Similarly, three genes encoding highly similar UGT94 family proteins and organized in a tandem repeat are involved in the biosynthesis of the natural triterpene sweetener mogroside V in the Chinese cucurbit *Siraitia grosvenorii* (Itkin et al., 2016). Furthermore, in soybean, *UGT73F2* and *UGT73F4* encode UGTs involved in triterpene saponin biosynthesis and are arranged as a tandem repeat within the *Sg-1* locus (Sayama et al., 2012). Hence, UGTs involved in saponin biosynthesis from different plant families and belonging to different UGT families all appear in tandem repeats. With the increasing availability of sequenced genomes, more and more of these tandem repeats could be encountered. Indeed, upon inspection of the *M. truncatula* genome, *UGT73F3* (*Medtr2g035020*), which encodes a C-28 glucosyltransferase, seems to be part of a tandem repeat encoding two additional UGTs. One of these UGTs, encoded by *Medtr2g035050*, appeared in our list of candidate UGTs (Supplemental Table 2), but we failed to clone it for our initial screening in yeast and *N. benthamiana*. *Medtr4g031800*, encoding UGT73K1, however, is not part of a tandem repeat. Interestingly, this gene is located right next to *CYP72A61* (*Medtr4g031820*), which encodes a P450 involved in non-hemolytic soyasaponin biosynthesis. Thereby *UGT73K1* seems to be part of a small metabolic gene cluster, a phenomenon not uncommon in plant specialized metabolism (Nützmann et al., 2016; Nützmann et al., 2018). Hence, in addition to being co-expressed, saponin biosynthesis genes appear to often occur in tandem arrays or small metabolic gene clusters. As such, the location of a candidate gene in the genome may be an additional selection criterion for candidate saponin biosynthesis genes.

METHODS

DNA Constructs

The sequences corresponding to the full-length open reading frames of all genes used were retrieved from the *M. truncatula* genome versions 3.5 (Young et al., 2011) or 4.0 (Tang et al., 2014) and were cloned using Gateway technology (Invitrogen). The gene sequences were PCR amplified (for primers, see Supplemental Dataset 5) and recombined into the donor vector pDONR221 or pDONR207. Sequence-verified entry clones for the candidate P450s and UGTs were Gateway recombined into the pEAQ-*HT*-DEST1 vector (Sainsbury et al., 2009) and sequence-verified. The resulting constructs were used for transformation of *A. tumefaciens* C58C1 RifR (PMP90) for transient expression in *N. benthamiana* leaves. For expression in yeast, entry clones of the P450s and UGTs were Gateway recombined into the high-copy number yeast destination vector pAG424GAL-ccdB (Addgene plasmid 14151; (Alberti et al., 2007)). For expression in hairy roots, the binary overexpression vector pK7WG2D (Karimi et al., 2002) was used as destination vector.

For protoplast assays, the sequence-verified *TSAR3* entry clone was Gateway recombined into the p2GW7 vector (Vanden Bossche et al., 2013). The promoter regions of *BAS*, *CYP72A68*, *CYP88A13*, *UGT73F18*, and *UGT73F19* were obtained from the *M. truncatula* genome version 4.0 (Tang et al., 2014), PCR-amplified from *M. truncatula* genomic DNA (ecotype Jemalong J5) and recombined into the donor vector pDONR221. Sequence-verified entry clones were recombined into the pGWL7 plasmid (Vanden Bossche et al., 2013) for protoplast assays. The generation of the effector plasmids for *TSAR1* and *TSAR2*, and the reporter plasmids for *ProHMGR1*, *ProMKB1*, *ProCYP72A67*, *ProCYP93E2*, *ProCYP72A61*, *ProUGT73F3*, and *ProUGT73K1*, was reported previously (Mertens et al., 2016b).

Transient Expression Assays in Tobacco Protoplasts

Transient expression assays in tobacco (*Nicotiana tabacum* 'Bright Yellow-2') protoplasts were carried out as previously described (Vanden Bossche et al., 2013). Briefly, protoplasts were transfected with a reporter, an effector, and a normalizer plasmid. The reporter plasmid consisted of a fusion between the promoter fragment of interest and the *fLUC* gene. The effector plasmid contained the selected transcription factor driven by the CaMV35S promoter. The normalizer plasmid contained the *RENILLA LUCIFERASE* (*rLUC*) gene under control of the CaMV35S

promoter. Protoplasts were incubated overnight and lysed. fLUC and rLUC readouts were collected using the Dual-Luciferase Reporter Assay System (Promega). Each assay was carried out in eight biological repeats. Promoter activities were normalized by dividing the fLUC values by the corresponding rLUC values. The average of the normalized fLUC values was calculated and set out relative to the control fLUC values (*i.e.* protoplasts transfected with an effector plasmid carrying a *GUS* gene).

Generation and Cultivation of *M. truncatula* Hairy Roots

M. truncatula seed sterilization (ecotype Jemalong J5), *Agrobacterium rhizogenes*-mediated transformation of *M. truncatula* seedlings (strain LBA 9402/12), cultivation of hairy roots and metabolite extractions for LC-MS analysis were done as previously described (Pollier et al., 2011).

qRT-PCR Analysis

Hairy roots grown for 21 days in liquid medium as described (Pollier et al., 2011) were harvested by flash freezing in liquid nitrogen. Harvested hairy roots were ground under liquid nitrogen with mortar and pestle and used for total RNA extraction and first-strand cDNA synthesis with the RNeasy Mini Kit (Qiagen) and the iScript cDNA Synthesis Kit (Bio-Rad), respectively. qRT-PCR was carried out with a LightCycler 480 system (Roche), SYBR Green Master Mix (Thermo Fisher Scientific) and qRT-PCR primers (Supplemental Dataset 5) were designed using primer3 version 0.4.0 (Thornton and Basu, 2011; Untergasser et al., 2012). For reference genes, the *M. truncatula 40S ribosomal protein S8* (*40S*) and *translation elongation factor 1α* (*ELF1α*) were used. Reactions were done in triplicate and qBase (Hellemans et al., 2007) was used for relative quantification with multiple reference genes.

Co-expression Analyses

To select candidate P450s, two co-expression analyses were carried out on microarray datasets of developing *M. truncatula* seeds. The first dataset consisted of 19,012 differentially expressed probesets during *M. truncatula* seed development (Verdier et al., 2013). Genes co-expressed with *TSAR3* were identified by calculating the Pearson's correlation coefficient between the probeset corresponding to *TSAR3* and all other probesets in this dataset. The second dataset contained five time series of developing seeds. Four series correspond to the A17 genotype for which seeds were

harvested at different stages from plants that were subjected to various environmental conditions from flowering onwards (Righetti et al., 2015). An additional developmental time series was obtained from seeds of the DZA315.16 genotype (GEO submission GSE137174). Like for the first dataset, genes co-expressed with *TSAR3* were identified by calculating the Pearson's correlation coefficient between the probeset corresponding to *TSAR3* and all other probesets in this dataset. Finally, for the selection of candidate UGTs, a third co-expression analysis was carried in which the *M. truncatula* gene expression atlas (He et al., 2009) was probed for UGTs co-expressed with *CYP716A12*. To this end, the expression values of all probesets in all 274 experiments included in the *M. truncatula* gene expression atlas were considered, and genes with a Pearson's correlation coefficient above 0.65 were considered co-expressed with *TSAR3*.

Yeast Engineering

Yeast strain PA147 was created from strain PA059 by CRISPR-mediated knockout of *PAH1* as described previously (Arendt et al., 2017). All yeast strains used in this study are listed in Supplemental Table 1. Subsequently, the *ERG7* promoter of strain PA147 was replaced by the methionine-repressible *MET3* promoter. This was achieved by PCR amplification of two overlapping fragments, a and b (Supplemental Figure 16), using primer pairs 1162/1311 and 1312/1161, respectively, with genomic DNA of the similarly engineered strain TM1 (Moses et al., 2014) as PCR template. The amplified fragments were ligated in the cloning vector pJET1.2/blunt (Thermo Fisher Scientific) and sequence verified. The overlapping fragments a and b were subsequently amplified from the sequence-verified pJET1.2 clones and strain PA147 was co-transformed with the resulting PCR fragments. Transformants were selected on YPD (Clontech) plates supplemented with 200 µg/mL G-418 disulfate (Duchefa). Promoter replacement in the resulting colonies (strain JP034) was confirmed by PCR using primer pair 3428/1161.

Production of medicagenic acid was achieved by co-expression of the genes encoding β-amyrin synthase from *Glycyrrhiza glabra* (GgbAS), cytochrome P450 reductase from *M. truncatula* (MtCPR1), and the C-2 (CYP72A67), C-23 (CYP72A68) and C-28 (CYP716A12) oxidases from *M. truncatula* in the starter strain JP034. As only three auxotrophic markers were available in this strain to express these five genes, two expression plasmids were generated that each express two genes as a self-cleaving polyprotein in which the two enzymes are linked via a viral 2A peptide. The first plasmid, pAG425GAL[MTR1-T2A-CYP716A12] was obtained by

Gateway recombination of the MTR1-T2A-CYP716A12 entry clone (Arendt et al., 2017) into the high-copy number yeast destination vector pAG424GAL-ccdB (Addgene plasmid 14153; (Alberti et al., 2007)). The second plasmid, pESC-URA-tHMG1-DEST[CYP72A68-T2A-GgbAS], was obtained by Gateway recombination of a CYP72A68-T2A-GgbAS entry clone into the high-copy number yeast destination vector pESC-URA-tHMG1-DEST (Fiallos-Jurado et al., 2016). To generate this entry clone, *CYP72A68* and *GgBAS* were first PCR amplified from plasmid DNA using primer pairs 2088/3716 and 3715/1297, respectively. The fragments were joined by overlap extension PCR using primer pair 2088/1297 and Gateway recombined into the donor vector pDONR221. Finally, *CYP72A67* was expressed from the high-copy number yeast destination vector pAG423GAL-ccdB (Addgene plasmid 14149; (Alberti et al., 2007)).

Yeast Cultivations for Triterpenoid Production

Yeast transformations were carried out using the lithium acetate/single-stranded carrier DNA/polyethylene glycol method (Gietz and Woods, 2002) and transformed cells were selected on SD medium supplemented with appropriate dropout supplements (Clontech). For evalulation of the P450s, yeast cells were cultivated in the presence of methyl-β-cyclodextrin (MβCD) as described (Moses et al., 2014). For GC-MS analysis, 1 mL of the spent medium was extracted thrice with 0.5 volumes of ethyl acetate. The organic extracts were pooled, evaporated to dryness under vacuum and the resulting residue was used for GC-MS analysis. For evalulation of the UGTs, yeast cells were cultivated without MβCD. For LC-MS analysis, yeast cells from 10 mL of cultivation medium were harvested in a 2-mL Eppendorf tube by centrifugation. Yeast cells were broken by flash freezing the cells in liquid nitrogen, adding metal balls (one ø5-mm and two ø3-mm balls) and milling using a RETCH ball mill (two times for 30 s at 30 Hz). Broken cells were extracted with 1 mL of methanol for 30 min at room temperature. After centrifugation, the methanol phase was collected and evaporated to dryness under vacuum. The resulting residue was used for LC-MS analysis.

Agroinfiltration of *N. benthamiana* Leaves

A. tumefaciens C58C1 RifR (pMP90) strains were cultivated for two days in a shaking incubator (150 rpm) at 28°C in 5 mL of yeast extract broth (YEB) medium, supplemented with 20 μg/mL gentamycin, 100 μg/mL rifampicin and 25 μg/mL kanamycin. After incubation, 500 μL of the pre-

cultures were used to inoculate 9.5 mL of YEB medium supplemented with antibiotics, 10 mM of MES (pH 5.7) and 20 μM of acetosyringone. After overnight incubation, bacteria for transient co-expression were mixed, collected via centrifugation, and resuspended in 5 mL of infiltration buffer (100 μM acetosyringone, 10 mM $MgCl_2$, and 10 mM MES, pH 5.7). The amount of bacteria harvested for each construct was adjusted to a final OD_{600} of 1.0 after resuspension in the infiltration buffer. *A. tumefaciens* containing *M. truncatula tHMGR1* (Pollier et al., 2013) was infiltrated together with the strains containing the triterpene biosynthesis genes to enhance triterpene yield (Reed and Osbourn, 2018). After a 3-h incubation at room temperature, the bacterial mixtures were infiltrated to the abaxial side of fully expanded leaves of 3- to 4-week-old *N. benthamiana* plants grown at 21°C in a 16-h/8-h light/dark regime. The infiltrated plants were incubated under normal growth conditions for four days prior to metabolite analysis. Four days after agroinfiltration, *N. benthamiana* leaves were harvested and ground to a fine powder in liquid nitrogen. For each sample, 100 mg of leaf material was extracted with 1 mL of methanol. The resulting organic extract was evaporated to dryness under vacuum and the residue was used for GC-MS or LC-MS analysis.

GC-MS Analysis

For GC-MS analysis, the residue obtained from metabolite extraction of yeast media or *N. benthamiana* leaves was trimethylsilylated using 10 μL of pyridine and 50 μL of *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide. GC-MS analysis of yeast samples was carried out using a GC model 6890 and MS model 5973 (Agilent, Santa Clara, United States). One μL of the sample was injected in splitless mode with the injector port set to 280°C. Separation was achieved with a VF-5ms column (30 m x 0.25 mm, 0.25 μm; Varian CP9013; Agilent) with helium carrier gas at a constant flow of 1 mL/min. The oven was held at 80°C for 1 min post-injection, ramped to 280°C at 20°C/min, held at 280°C for 45 min, ramped to 320°C at 20°C/min, held at 320°C for 1 min, and finally cooled to 80°C at 50°C/min at the end of the run. The MS transfer line, ion source and quadrupole temperatures were set to 250°C, 230°C, and 150°C, respectively. For metabolite identification, full EI-MS spectra were recorded between *m/*z 60-800 with a solvent delay of 7.8 min. GC-MS analysis of *N. benthamiana* leaf extracts was carried out using a 7890B GC system equipped with a 7693A Automatic Liquid Sampler and a 7250 Accurate-Mass Quadrupole Time-of-Flight MS system (Agilent). One μL of the sample was injected in splitless mode with the

injector port set to 280°C. Separation was achieved with a VF-5ms column (30 m x 0.25 mm, 0.25 µm; Varian CP9013; Agilent) with helium carrier gas at a constant flow of 1.2 mL/min. The oven was held at 80°C for 1 min post-injection, ramped to 280°C at 20°C/min, held at 280°C for 60 min, ramped to 320°C at 20°C/min, held at 320°C for 5 min, and finally cooled to 80°C at 50°C/min at the end of the run. The MSD transfer line was set to 280°C and the electron ionization energy was 70 eV. Full EI-MS spectra were recorded between *m*/*z* 50-800 at a resolution of >25,000 and with a solvent delay of 20.0 min.

LC-MS Analysis

The residues obtained from metabolite extraction of yeast cells, *N. benthamiana* leaves or *M. truncatula* hairy roots were dissolved in 800 µL of H_2O/cyclohexane (1:1, v/v). Samples were centrifuged (10 min at 20,800 x g), and 200 µL of the aqueous phase was retained for LC-MS analysis.

For LC-MS, 10 µL of the sample was injected into an Acquity UPLC BEH C18 column (2.1 x 150 mm, 1.7 µm) mounted on either a Waters Acquity UPLC system coupled to a SYNAPT HDMS Q-TOF or a Thermo instrument equipped with an LTQ FT Ultra for accurate mass measurements. Ionization was achieved via an electrospray ionization source. The following gradient was run using acidified (0.1% (v/v) formic acid) solvents A (water/acetonitrile, 99:1, v/v) and B (acetonitrile/water; 99:1, v/v) at a flow rate of 350 µL/min: time 0 min, 5% B; 30 min, 50% B; 33 min, 100% B. The mass spectrometer was set to negative ionization mode with the following parameter values: capillary temperature 150°C, sheath gas 25 (arbitrary units), aux. gas 3 (arbitrary units) and spray voltage 4.5 kV. For analyte identification using the LTQ FT Ultra, full MS spectra were interchanged with a dependent MS^2 scan event in which the most abundant ion in the previous full MS scan was fragmented, two dependent MS^3 scan events in which the two most abundant daughter ions were fragmented and a dependent MS^4 scan event in which the most abundant daughter ion of the first MS^3 scan event was fragmented. The collision energy was set to 35%.

Peak areas for untargeted metabolite profiling experiments were determined using the Progenesis QI software (Waters) and the principal component analysis and partial least squares discriminant analysis was performed using log-transformed, pareto-scaled mass spectrometry data with standard settings of the MetaboAnalyst 4.0 software package (Chong et al., 2018).

Analytical Standards

Analytical standards for GC-MS analysis (β-amyrin, erythrodiol, oleanolic acid, hederagenin, medicagenic acid, quillaic acid, and caulophyllogenin) were purchased from Extrasynthese. Saponin standards for LC-MS analysis (Supplemental Datasets 1 and 2) were obtained from Dr. Aldo Tava (CRA-FLC, Italy) and Prof. Wiesław Oleszek (IUNG, Poland) and were described previously (Pollier et al., 2013).

Genotyping of *Tnt1* Insertion Mutants

Tnt1 insertion mutants of *Medtr2g104650* (*TSAR3*) and *Medtr5g014240* (*CYP88A13*) were screened for according to a method modified from Cheng et al. (2011). Briefly, NF14212, NF15672, NF17107, NF11894, NF17520, and NF13195 seeds received from the Noble Research Institute were germinated and grown in soil. Plants were grown in the greenhouse at 21°C in a 16-h/8-h light/dark regime. After 4 weeks, two leaves were cut from each plant, frozen in liquid nitrogen and ground using a RETCH ball mill. DNA was extracted as reported (Cheng et al., 2011) and used as PCR template. PCR was carried out with gene-specific primers and primers within the *Tnt1* retrotransposon (Supplemental Figures 3 and 12, Supplemental Dataset 5). The obtained PCR fragments were sequenced to determine the exact location of the *Tnt1* insertion. R108 wild-type plants and homozygous and hemizygous insertion mutants were allowed to self-pollinate for seed production. Plants grown from the resulting seeds were re-genotyped and used for transcript and metabolite profiling. For the *TSAR3 Tnt1* insertion lines, seeds were harvested approximately 24 days after pollination (24DAP, Supplemental Figure 17) and leaves were harvested at the same time.

Phylogenetic Analysis

An alignment of the bHLH domains according to Heim et al. (2003) was used as input file (Supplemental Dataset 6) to generate a Maximum Likelihood tree using MEGA5 software (Tamura et al., 2011). The Jones, Taylor and Thornton model was applied to determine evolutionary distances. Bootstrap analysis was carried out with 1,000 replicates. A machine-readable tree file is provided as Supplemental Dataset 7.

Accession Numbers

Sequence data from this article can be found in the GenBank database under accession numbers MN200610-MN200629.

Supplemental Data

The following materials are available in the online version of this article:

Supplemental Figure 1. Expression of *TSAR3* according to the *Medicago truncatula* gene expression atlas.

Supplemental Figure 2. Partial least squares discriminant analysis (PLS-DA) of the samples from the $TSAR3^{OE}$ and control (CTR) hairy roots.

Supplemental Figure 3. Genotyping of the *TSAR3 Tnt1* insertion lines.

Supplemental Figure 4. Hemolytic saponin biosynthesis is not affected in leaves of *tsar3* plants.

Supplemental Figure 5. EI-MS spectra of additional metabolites identified in Figure 7A.

Supplemental Figure 6. EI-MS spectra and reaction scheme of the metabolites identified in Figure 8A.

Supplemental Figure 7. Reaction schemes of the metabolites identified in Figures 8B and 8D.

Supplemental Figure 8. EI-MS spectra of the metabolites identified in Figure 8B.

Supplemental Figure 9. EI-MS spectra and reaction scheme of the metabolites identified in Figure 8C.

Supplemental Figure 10. EI-MS spectra of the metabolites identified in Figure 8D.

Supplemental Figure 11. Investigation of the substrate tolerance of CYP88A13 in *N. benthamiana*.

Supplemental Figure 12. Genotyping of the *Tnt1* insertion lines.

Supplemental Figure 13. Analysis of *CYP88A14* in yeast and *N. benthamiana*.

Supplemental Figure 14. Phylogenetic analysis of the identified *M. truncatula* and *Arabidopsis thaliana* clade IVa proteins.

Supplemental Figure 15. Comparison of the enzymatic reactions catalyzed by CYP88A13 and *ent*-kaurenoic acid oxidase (KAO) belonging to the CYP88A subfamily.

Supplemental Figure 16. Promoter replacement in strain PA147, leading to strain JP034.

Supplemental Figure 17. Development of *M. truncatula* seeds.

Supplemental Table 1. Yeast strains used in this study.

Supplemental Table 2. Candidate UGTs identified through the different co-expression analyses.

Supplemental Dataset 1. Differential metabolites between the TSAR3^{OE} and GUS control lines (separate Excel file).

Supplemental Dataset 2. Differential metabolites in developing seeds of *tsar3* and wild-type plants (separate Excel file).

Supplemental Dataset 3. Co-expression analysis in publicly available microarray datasets from developing *M. truncatula* seeds (separate Excel file).

Supplemental Dataset 4. Co-expression analysis in a set of 60 microarray datasets from developing *M. truncatula* seeds (separate Excel file).

Supplemental Dataset 5. List of oligonucleotides used in this study (separate Excel file).

Supplemental Dataset 6. Sequence alignment of the bHLH domains of the identified *M. truncatula* and *A. thaliana* clade IVa proteins.

Supplemental Dataset 7. Machine-readable tree file in Newick format in support of Supplemental Figure 14.

ACKNOWLEDGMENTS

We thank Prof. David Nelson for naming the P450s, Dr. Michael Court and the UGT nomenclature committee for naming the UGTs, Steven Vandersyppe and Geert Goeminne (VIB Metabolomics Core) for help with the metabolite profiling and LC-MS data processing, Dr. Aldo Tava (CRA-FLC, Italy) and Prof. Wiesław Oleszek (IUNG, Poland) for the *Medicago* saponin standards, and Annick Bleys for help with preparing the manuscript. This work was supported by the Research Foundation-Flanders for predoctoral fellowships to E.C. and J.M. and for a postdoctoral fellowship to J.P., the Program Ciências Sem Fronteiras for a predoctoral fellowship to B.R. (Grant 201135/2014-0) and a SWE fellowship to G.C. (Grant 202776/2014-0), the FAPESP - São Paulo Research Foundation (Process number 2017/10911-1) for a postdoctoral fellowship to K.U.B., the VIB International PhD Fellowship Program for a predoctoral fellowship to P.A., the European Union Horizon 2020 program for a research project grant to A.G. under grant agreement Endoscape-825730, and the Concerted Research Actions from the Ghent University for a research project grant to A.G. (BOF18/GOA/013).

AUTHOR CONTRIBUTIONS

A.G., and J.P. designed experiments. J.B. contributed the expression data. B.R., E.L., K.U.B., J.M., P.A., R.V.B, G.C., L.G., E.C., and J.P. performed experiments. J.P. analyzed the data and wrote the manuscript. A.G. complemented the writing.

REFERENCES

Achnine, L., Huhman, D.V., Farag, M.A., Sumner, L.W., Blount, J.W., and Dixon, R.A. (2005). Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume *Medicago truncatula*. Plant Journal **41,** 875-887.

Agrell, J., Anderson, P., Oleszek, W., Stochmal, A., and Agrell, C. (2004). Combined effects of elevated CO_2 and herbivore damage on alfalfa and cotton. Journal of Chemical Ecology **30,** 2309-2324.

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.

Arendt, P., Miettinen, K., Pollier, J., De Rycke, R., Callewaert, N., and Goossens, A. (2017). An endoplasmic reticulum-engineered yeast platform for overproduction of triterpenoids. Metabolic engineering **40,** 165-175.

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

Augustin, J.M., Drok, S., Shinoda, T., Sanmiya, K., Nielsen, J.K., Khakimov, B., Olsen, C.E., Hansen, E.H., Kuzina, V., Ekstrøm, C.T., Hauser, T.P., and Bak, S. (2012). UDP-glycosyltransferases

from the UGT73C subfamily in *Barbarea vulgaris* catalyze sapogenin 3-*O*-glucosylation in saponin-mediated insect resistance. Plant Physiology **160,** 1881-1895.

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

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

Bicalho, K.U., Santoni, M.M., Arendt, P., Zanelli, C.F., Furlan, M., Goossens, A., and Pollier, J. (2019). CYP712K4 catalyzes the C-29 oxidation of friedelin in the *Maytenus ilicifolia* quinone methide triterpenoid biosynthesis pathway. Plant & Cell Physiology **60,** 2510-2522.

Broeckling, C.D., Huhman, D.V., Farag, M.A., Smith, J.T., May, G.D., Mendes, P., Dixon, R.A., and Sumner, L.W. (2005). Metabolic profiling of *Medicago truncatula* cell cultures reveals the effects of biotic and abiotic elicitors on metabolism. Journal of Experimental Botany **56,** 323-336.

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

Cheng, X., Wen, J., Tadege, M., Ratet, P., and Mysore, K.S. (2011). Reverse genetics in *Medicago truncatula* using *Tnt1* insertion mutants. Methods in Molecular Biology **678,** 179-190.

Chini, A., Gimenez-Ibanez, S., Goossens, A., and Solano, R. (2016). Redundancy and specificity in jasmonate signalling. Current Opinion in Plant Biology **33,** 147-156.

Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., Wishart, D.S., and Xia, J. (2018). MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic Acids Research **46,** W486-W494.

Claereboudt, E.J.S., Caulier, G., Decroo, C., Colson, E., Gerbaux, P., Claereboudt, M.R., Schaller, H., Flammang, P., Deleu, M., and Eeckhaut, I. (2019). Triterpenoids in echinoderms: fundamental differences in diversity and biosynthetic pathways. Marine Drugs **17,** 352.

Colinas, M., and Goossens, A. (2018). Combinatorial transcriptional control of plant specialized metabolism. Trends in Plant Science **23,** 324-336.

Confalonieri, M., Cammareri, M., Biazzi, E., Pecchia, P., Fevereiro, M.P.S., Balestrazzi, A., Tava, A., and Conicella, C. (2009). Enhanced triterpene saponin biosynthesis and root nodulation in transgenic barrel medic (*Medicago truncatula* Gaertn.) expressing a novel β-amyrin synthase (*AsOXA1*) gene. Plant Biotechnology Journal **7,** 172-182.

Erthmann, P.Ø., Agerbirk, N., and Bak, S. (2018). A tandem array of UDP-glycosyltransferases from the UGT73C subfamily glycosylate sapogenins, forming a spectrum of mono- and bisdesmosidic saponins. Plant Molecular Biology **97,** 37-55.

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

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 Physiology **54,** 740-749.

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 Physiology **52,** 2050-2061.

Gholami, A., De Geyter, N., Pollier, J., Goormachtig, S., and Goossens, A. (2014). Natural product biosynthesis in *Medicago* species. Natural Product Reports **31,** 356-380.

Gietz, R.D., and Woods, R.A. (2002). Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. Methods in Enzymology **350,** 87-96.

Goossens, J., Mertens, J., and Goossens, A. (2017). Role and functioning of bHLH transcription factors in jasmonate signalling. Journal of Experimental Botany **68,** 1333-1347.

Goossens, J., Fernández-Calvo, P., Schweizer, F., and Goossens, A. (2016). Jasmonates: signal transduction components and their roles in environmental stress responses. Plant molecular biology **91,** 673-689.

Hamberger, B., and Bak, S. (2013). Plant P450s as versatile drivers for evolution of species-specific chemical diversity. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences **368,** 20120426.

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

Heim, M.A., Jakoby, M., Werber, M., Martin, C., Weisshaar, B., and Bailey, P.C. (2003). The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. Molecular Biology and Evolution **20,** 735-747.

Hellemans, J., Mortier, G., De Paepe, A., Speleman, F., and Vandesompele, J. (2007). qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biology **8,** R19.

Huhman, D.V., Berhow, M.A., and Sumner, L.W. (2005). Quantification of saponins in aerial and subterranean tissues of *Medicago truncatula*. Journal of Agricultural and Food Chemistry **53,** 1914-1920.

Itkin, M., Davidovich-Rikanati, R., Cohen, S., Portnoy, V., Doron-Faigenboim, A., Oren, E., Freilich, S., Tzuri, G., Baranes, N., Shen, S., Petreikov, M., Sertchook, R., Ben-Dor, S., Gottlieb, H., Hernandez, A., Nelson, D.R., Paris, H.S., Tadmor, Y., Burger, Y., Lewinsohn, E., Katzir, N., and Schaffer, A. (2016). The biosynthetic pathway of the nonsugar, high-intensity sweetener mogroside V from *Siraitia grosvenorii*. Proceedings of the National Academy of Sciences of the United States of America **113,** E7619-E7628.

Iturbe-Ormaetxe, I., Haralampidis, K., Papadopoulou, K., and Osbourn, A.E. (2003). Molecular cloning and characterization of triterpene synthases from *Medicago truncatula* and *Lotus japonicus*. Plant Molecular Biology **51,** 731-743.

Jarvis, D.E., Ho, Y.S., Lightfoot, D.J., Schmöckel, S.M., Li, B., Borm, T.J.A., Ohyanagi, H., Mineta, K., Michell, C.T., Saber, N., Kharbatia, N.M., Rupper, R.R., Sharp, A.R., Dally, N., Boughton, B.A., Woo, Y.H., Gao, G., Schijlen, E.G.W.M., Guo, X., Momin, A.A., Negrão, S., Al-Babili, S., Gehring, C., Roessner, U., Jung, C., Murphy, K., Arold, S.T., Gojobori, T., van der Linden, C.G., van Loo, E.N., Jellen, E.N., Maughan, P.J., and Tester, M. (2017). The genome of *Chenopodium quinoa*. Nature **542,** 307-312.

Kapusta, I., Janda, B., Stochmal, A., and Oleszek, W. (2005a). Determination of saponins in aerial parts of barrel medic (*Medicago truncatula*) by liquid chromatography-electrospray ionization/mass spectrometry. Journal of Agricultural and Food Chemistry **53,** 7654-7660.

Kapusta, I., Stochmal, A., Perrone, A., Piacente, S., Pizza, C., and Oleszek, W. (2005b). Triterpene saponins from barrel medic (*Medicago truncatula*) aerial parts. Journal of Agricultural and Food Chemistry **53,** 2164-2170.

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

Kazan, K., and Manners, J.M. (2013). MYC2: the master in action. Molecular Plant **6,** 686-703.

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

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

Lei, Z., Watson, B.S., Huhman, D., Yang, D.S., and Sumner, L.W. (2019). Large-scale profiling of saponins in different ecotypes of *Medicago truncatula*. Frontiers in Plant Science **10,** 850.

Mertens, J., Van Moerkercke, A., Vanden Bossche, R., Pollier, J., and Goossens, A. (2016a). Clade IVa basic helix-loop-helix transcription factors form part of a conserved jasmonate signaling circuit for the regulation of bioactive plant terpenoid biosynthesis. Plant & Cell Physiology **57,** 2564-2575.

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

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

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*. Molecular Plant **8,** 122-135.

Moses, T., Pollier, J., Almagro, L., Buyst, D., Van Montagu, M., Pedreño, M.A., Martins, J.C., Thevelein, J.M., and Goossens, A. (2014). Combinatorial biosynthesis of sapogenins and saponins in *Saccharomyces cerevisiae* using a C-16α hydroxylase from *Bupleurum falcatum*. Proceedings of the National Academy of Sciences of the United States of America **111,** 1634-1639.

Naoumkina, M.A., Modolo, L.V., Huhman, D.V., Urbanczyk-Wochniak, E., Tang, Y., Sumner, L.W., and Dixon, R.A. (2010). Genomic and coexpression analyses predict multiple genes involved in triterpene saponin biosynthesis in *Medicago truncatula*. Plant Cell **22,** 850-866.

Nützmann, H.-W., Huang, A., and Osbourn, A. (2016). Plant metabolic clusters - from genetics to genomics. New Phytologist **211,** 771-789.

Nützmann, H.-W., Scazzocchio, C., and Osbourn, A. (2018). Metabolic gene clusters in eukaryotes. Annual Review of Genetics **52,** 159-183.

Ober, D. (2010). Gene duplications and the time thereafter – examples from plant secondary metabolism. Plant Biology **12,** 570-577.

Osbourn, A., Goss, R.J.M., and Field, R.A. (2011). The saponins – polar isoprenoids with important and diverse biological activities. Natural Product Reports **28,** 1261-1268.

Podolak, I., Galanty, A., and Sobolewska, D. (2010). Saponins as cytotoxic agents: a review. Phytochemistry Reviews **9,** 425-474.

Pollier, J., Morreel, K., Geelen, D., and 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.

Pollier, J., Moses, T., González-Guzmán, M., De Geyter, N., Lippens, S., Vanden Bossche, R., Marhavý, P., Kremer, A., Morreel, K., Guérin, C.J., Tava, A., Oleszek, W., Thevelein, J.M., Campos, N., Goormachtig, S., and Goossens, A. (2013). The protein quality control system manages plant defence compound synthesis. Nature **504,** 148-152.

Rafińska, K., Pomastowski, P., Wrona, O., Górecki, R., and Buszewski, B. (2017). *Medicago sativa* as a source of secondary metabolites for agriculture and pharmaceutical industry. Phytochemistry Letters **20,** 520-539.

Reed, J., and Osbourn, A. (2018). Engineering terpenoid production through transient expression in *Nicotiana benthamiana*. Plant cell reports **37,** 1431-1441.

Reed, J., Stephenson, M.J., Miettinen, K., Brouwer, B., Leveau, A., Brett, P., Goss, R.J.M., Goossens, A., O'Connell, M.A., and 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.

Rehman, H.M., Nawaz, M.A., Shah, Z.H., Yang, S.H., and Chung, G. (2018). Functional characterization of naturally occurring wild soybean mutant (*sg-5*) lacking astringent saponins using whole genome sequencing approach. Plant Science **267,** 148-156.

Righetti, K., Vu, J.L., Pelletier, S., Vu, B.L., Glaab, E., Lalanne, D., Pasha, A., Patel, R.V., Provart, N.J., Verdier, J., Leprince, O., and Buitink, J. (2015). Inference of longevity-related genes from a robust coexpression network of seed maturation identifies regulators linking seed storability to biotic defense-related pathways. Plant Cell **27,** 2692-2708.

Sainsbury, F., Thuenemann, E.C., and Lomonossoff, G.P. (2009). pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. Plant Biotechnology Journal **7,** 682-693.

Sayama, T., Ono, E., Takagi, K., Takada, Y., Horikawa, M., Nakamoto, Y., Hirose, A., Sasama, H., Ohashi, M., Hasegawa, H., Terakawa, T., Kikuchi, A., Kato, S., Tatsuzaki, N., Tsukamoto, C., and Ishimoto, M. (2012). The *Sg-1* glycosyltransferase locus regulates structural diversity of triterpenoid saponins of soybean. Plant Cell **24,** 2123-2128.

Seki, H., Ohyama, K., Sawai, S., Mizutani, M., Ohnishi, T., Sudo, H., Akashi, T., Aoki, T., Saito, K., and Muranaka, T. (2008). Licorice β-amyrin 11-oxidase, a cytochrome P450 with a key role in the biosynthesis of the triterpene sweetener glycyrrhizin. Proceedings of the National Academy of Sciences of the United States of America **105,** 14204-14209.

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

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

Shoji, T. (2019). The recruitment model of metabolic evolution: jasmonate-responsive transcription factors and a conceptual model for the evolution of metabolic pathways. Frontiers in Plant Science **10,** 560.

Sundaramoorthy, J., Park, G.T., Mukaiyama, K., Tsukamoto, C., Chang, J.H., Lee, J.-D., Kim, J.H., Seo, H.S., and Song, J.T. (2018). Molecular elucidation of a new allelic variation at the *Sg-5* gene associated with the absence of group A saponins in wild soybean. PLoS ONE **13,** e0192150.

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

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

Tadege, M., Wen, J., He, J., Tu, H., Kwak, Y., Eschstruth, A., Cayrel, A., Endre, G., Zhao, P.X., Chabaud, M., Ratet, P., and Mysore, K.S. (2008). Large-scale insertional mutagenesis using the *Tnt1* retrotransposon in the model legume *Medicago truncatula*. Plant Journal **54,** 335-347.

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. Molecular Biology and Evolution **28,** 2731-2739.

Tamura, K., Yoshida, K., Hiraoka, Y., Sakaguchi, D., Chikugo, A., Mochida, K., Kojoma, M., Mitsuda, N., Saito, K., Muranaka, T., and Seki, H. (2018). The basic helix–loop–helix transcription factor GubHLH3 positively regulates soyasaponin biosynthetic genes in *Glycyrrhiza uralensis*. Plant & Cell Physiology **59,** 783-796.

Tang, H., Krishnakumar, V., Bidwell, S., Rosen, B., Chan, A., Zhou, S., Gentzbittel, L., Childs, K.L., Yandell, M., Gundlach, H., Mayer, K.F.X., Schwartz, D.C., and Town, C.D. (2014). An improved genome release (version Mt4.0) for the model legume *Medicago truncatula*. BMC Genomics **15,** 312.

Tava, A., Scotti, C., and Avato, P. (2011). Biosynthesis of saponins in the genus *Medicago*. Phytochemistry Reviews **10,** 459-469.

Thornton, B., and Basu, C. (2011). Real-time PCR (qPCR) primer design using free online software. Biochemistry and Molecular Biology Education **39,** 145-154.

Tzin, V., Snyder, J.H., Yang, D.S., Huhman, D.V., Watson, B.S., Allen, S.N., Tang, Y., Miettinen, K., Arendt, P., Pollier, J., Goossens, A., and Sumner, L.W. (2019). Integrated metabolomics identifes CYP72A67 and CYP72A68 oxidases in the biosynthesis of *Medicago truncatula* oleanate sapogenins. Metabolomics **15,** 85.

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

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

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

Vanden Bossche, R., Demedts, B., Vanderhaeghen, R., and Goossens, A. (2013). Transient expression assays in tobacco protoplasts. Methods in Molecular Biology **1011,** 227-239.

Verdier, J., Lalanne, D., Pelletier, S., Torres-Jerez, I., Righetti, K., Bandyopadhyay, K., Leprince, O., Chatelain, E., Vu, B.L., Gouzy, J., Gamas, P., Udvardi, M.K., and Buitink, J. (2013). A regulatory network-based approach dissects late maturation processes related to the acquisition of desiccation tolerance and longevity of *Medicago truncatula* seeds. Plant Physiology **163,** 757-774.

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

Wasternack, C., and Hause, B. (2013). Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in *Annals of Botany*. Annals of Botany **111,** 1021-1058.

Wasternack, C., and Feussner, I. (2018). The oxylipin pathways: biochemistry and function. Annual Review of Plant Biology **69,** 363-386.

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

Young, N.D., Debellé, F., Oldroyd, G.E.D., Geurts, R., Cannon, S.B., Udvardi, M.K., Benedito, V.A., Mayer, K.F.X., Gouzy, J., Schoof, H., Van de Peer, Y., Proost, S., Cook, D.R., Meyers, B.C., Spannagl, M., Cheung, F., De Mita, S., Krishnakumar, V., Gundlach, H., Zhou, S., Mudge, J., Bharti, A.K., Murray, J.D., Naoumkina, M.A., Rosen, B., Silverstein, K.A.T., Tang, H., Rombauts, S., Zhao, P.X., Zhou, P., Barbe, V., Bardou, P., Bechner, M., Bellec, A., Berger, A., Bergès, H., Bidwell, S., Bisseling, T., Choisne, N., Couloux, A., Denny, R., Deshpande, S., Dai, X., Doyle, J.J., Dudez, A.-M., Farmer, A.D., Fouteau, S., Franken, C., Gibelin, C., Gish, J., Goldstein, S., González, A.J., Green, P.J., Hallab, A., Hartog, M., Hua, A., Humphray, S.J., Jeong, D.-H., Jing, Y., Jöcker, A., Kenton, S.M., Kim, D.-J., Klee, K., Lai, H., Lang, C., Lin, S., Macmil, S.L., Magdelenat, G., Matthews, L., McCorrison, J., Monaghan, E.L., Mun, J.-H., Najar, F.Z., Nicholson, C., Noirot, C., O'Bleness, M., Paule, C.R., Poulain, J., Prion, F., Qin, B.F., Qu, C., Retzel, E.F., Riddle, C., Sallet, E., Samain, S., Samson, N., Sanders, I., Saurat, O., Scarpelli, C., Schiex, T., Segurens, B., Severin, A.J., Sherrier, D.J., Shi, R., Sims, S., Singer, S.R., Sinharoy, S., Sterck, L., Viollet, A., Wang, B.-B., Wang, K., Wang, M., Wang, X., Warfsmann, J., Weissenbach, J., White, D.D., White, J.D., Wiley, G.B., Wincker, P., Xing, Y., Yang, L., Yao, Z., Ying, F., Zhai, J., Zhou, L., Zuber, A., Dénarié, J., Dixon, R.A., May, G.D.,

Schwartz, D.C., Rogers, J., Quétier, F., Town, C.D., and Roe, B.A. (2011). The *Medicago* genome provides insight into the evolution of rhizobial symbioses. Nature **480,** 520-524.

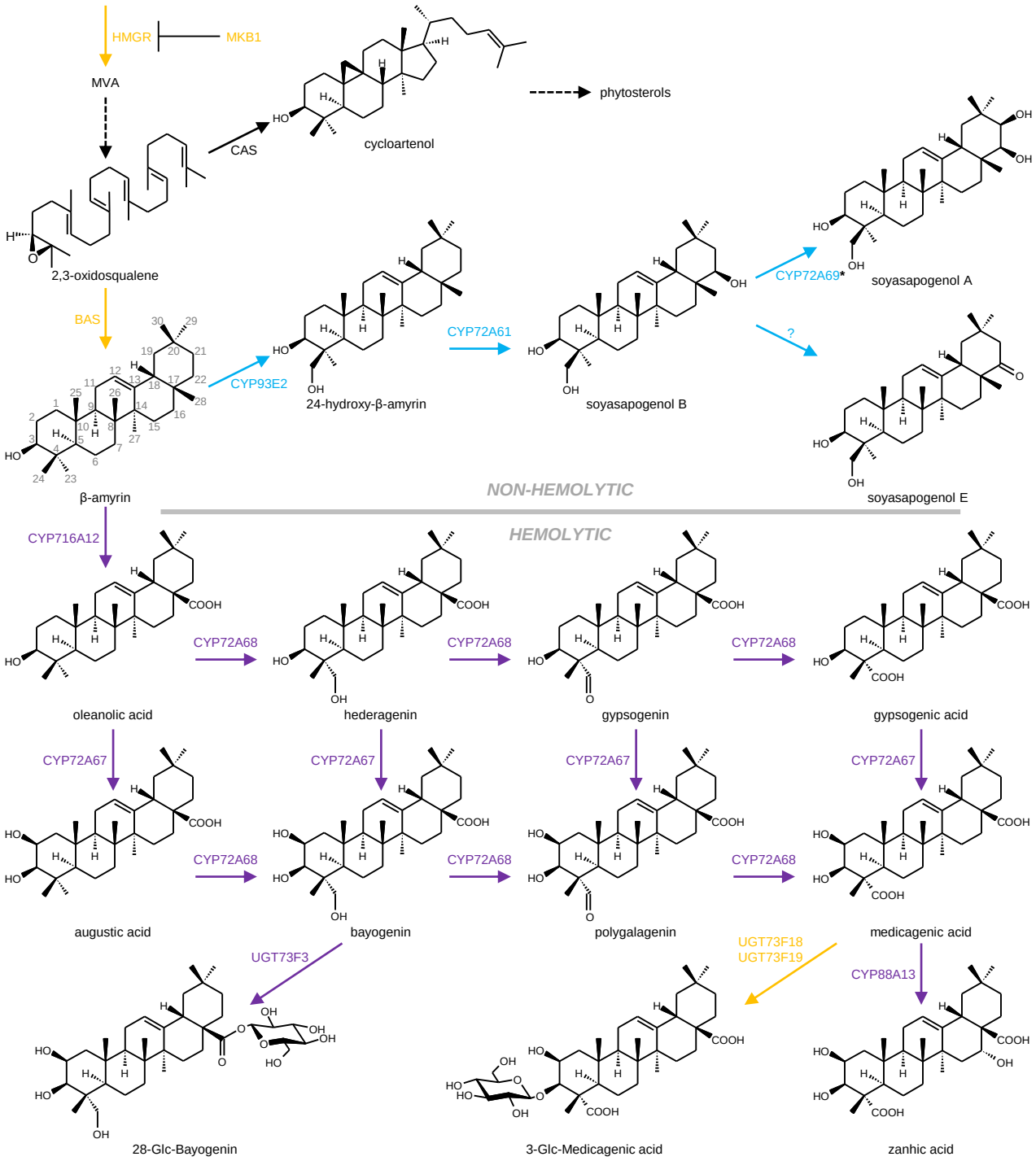


Figure 1. The saponin biosynthesis pathway in *M. truncatula*. Blue, purple, and orange colors indicate the biosynthetic steps controlled by TSAR1, TSAR2/3 or TSAR1/2/3, respectively. Dashed arrows indicate multiple steps. HMGR, 3-hydroxy-3-methylglutaryl-CoA reductase; BAS, β-amyrin synthase; CAS, cycloartenol synthase; MVA, mevalonate. *****CYP72A69 was characterized in *Glycine max*.

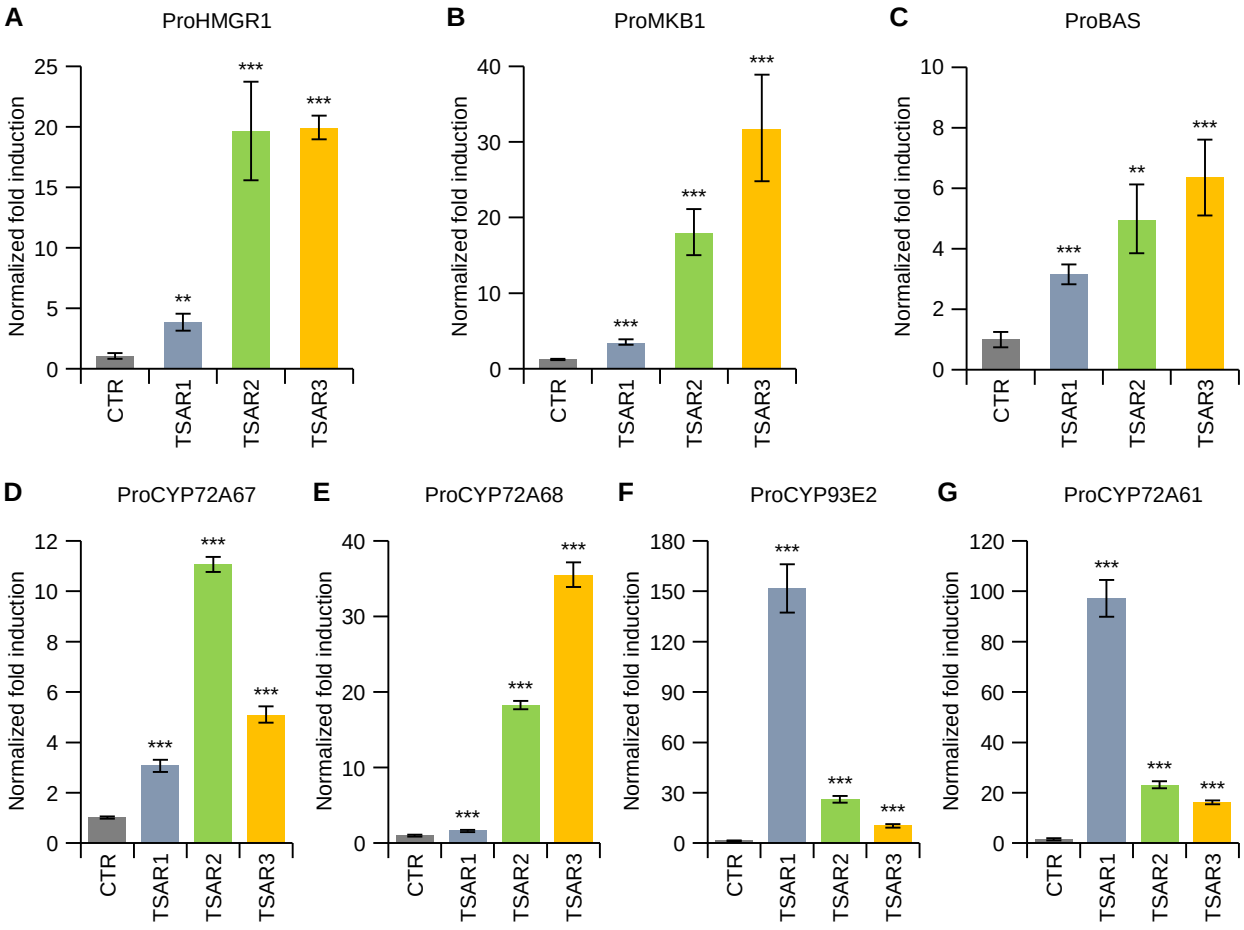


Figure 2. Transactivation of triterpene saponin biosynthesis gene promoters by TSAR transcription factors in transfected tobacco protoplasts. Transactivation of *ProHMGR1* (**A**), *ProMKB1* (**B**), *ProBAS* (**C**), *ProCYP72A67* (**D**), *ProCYP72A68* (**E**), *ProCYP93E2* (**F**), and *ProCYP72A61* (**G**) by TSAR1, TSAR2, and TSAR3 in transfected tobacco protoplasts. Values on the y-axis are normalized fold changes relative to protoplasts co-transfected with the reporter constructs and a *pCaMV35S:GUS* control plasmid (CTR). The error bars designate the standard error (n = 8 biological replicates). Statistical significance was determined by a two-tailed Student's *t*-test (*, $P < 0.05$; **, $P < 0.01$; and ***, $P < 0.001$).

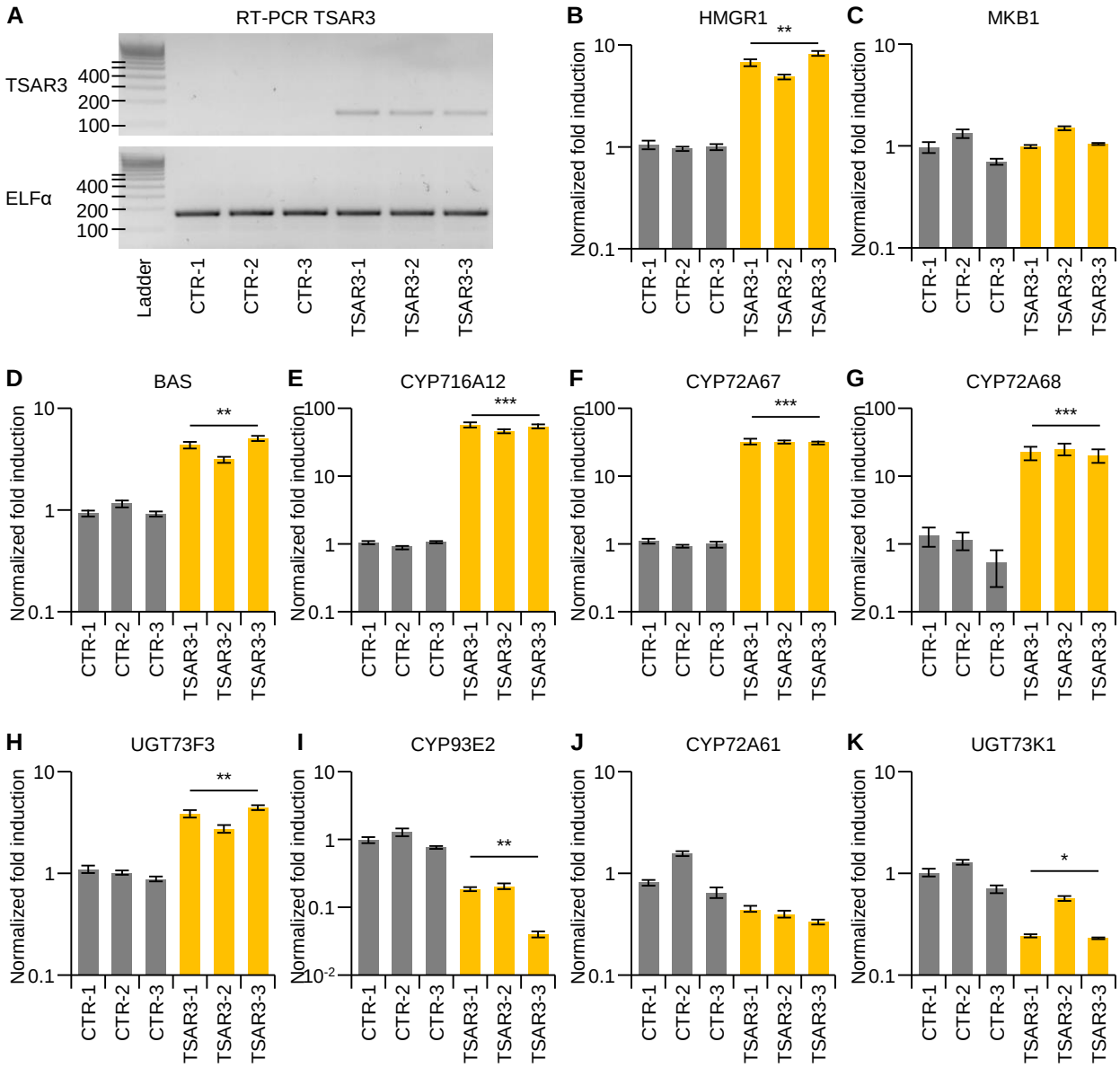


Figure 3. Ectopic expression of *TSAR3* in *M. truncatula* hairy roots affects triterpene saponin biosynthesis gene expression. (**A**) RT-PCR analysis showing *TSAR3* expression in three independent transgenic TSAR3^{OE} hairy root lines compared to three control (CTR) lines. The qRT-PCR reference gene *ELFα* was used as a control. The Eurogentec SmartLadder SF was used as a size marker (Ladder). (**B-K**) qRT-PCR analysis of triterpene saponin biosynthesis genes in three independent control and TSAR3^{OE} *M. truncatula* hairy root lines. Values on the y-axis are normalized fold changes relative to the average of the three control lines. The error bars designate the standard error (n = 3 technical replicates). Statistical significance was determined by a two-tailed Student's *t*-test (*, $P < 0.05$; **, $P < 0.01$; and ***, $P < 0.001$).

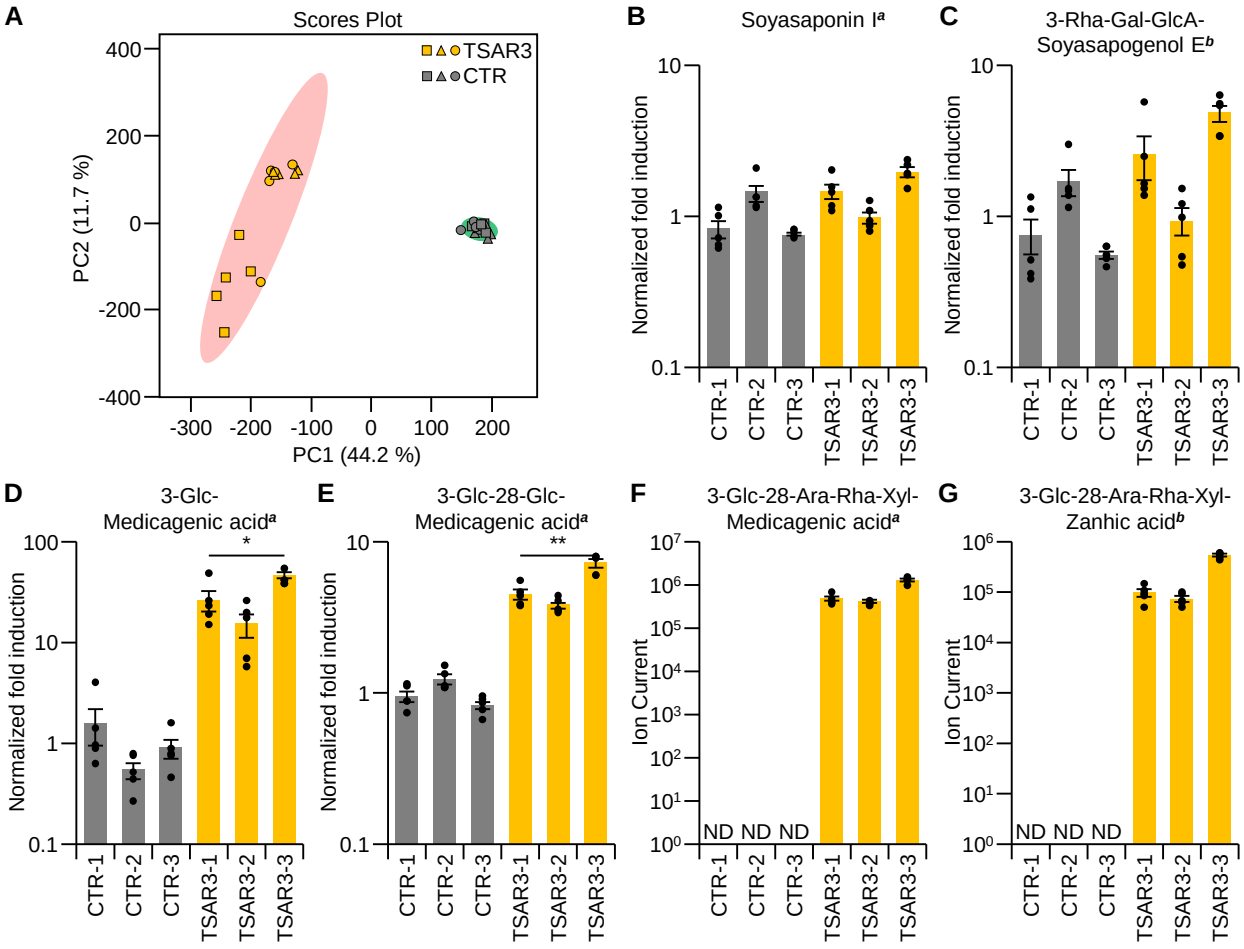


Figure 4. Ectopic expression of *TSAR3* leads to enhanced accumulation of hemolytic triterpene saponins in *M. truncatula* hairy roots. (**A)** Principal component analysis of samples from TSAR3^{OE} and control (CTR) hairy roots. Different symbols indicate the different transgenic lines. (**B-E**) Relative accumulation of triterpene saponins in three independent control and TSAR3^{OE} *M. truncatula* hairy root lines. Values on the y-axis are normalized fold changes relative to the average of the three control lines. (**F-G**) Average total ion current of the peaks corresponding to two hemolytic triterpene saponins. The mean and standard error (n = 5 technical replicates) are shown and dot plots (black dots) are overlaid. Statistical significance was determined by a two-tailed Student's *t*-test (*, $P < 0.05$; **, $P < 0.01$; and ***, $P < 0.001$). ND, not detected; metabolites were (*[a]*) identified using authentic standards or (*[b]*) predicted based on their accurate mass and MS/MS fragmentation spectra (Supplemental Dataset 1) according to Pollier et al. (2011).

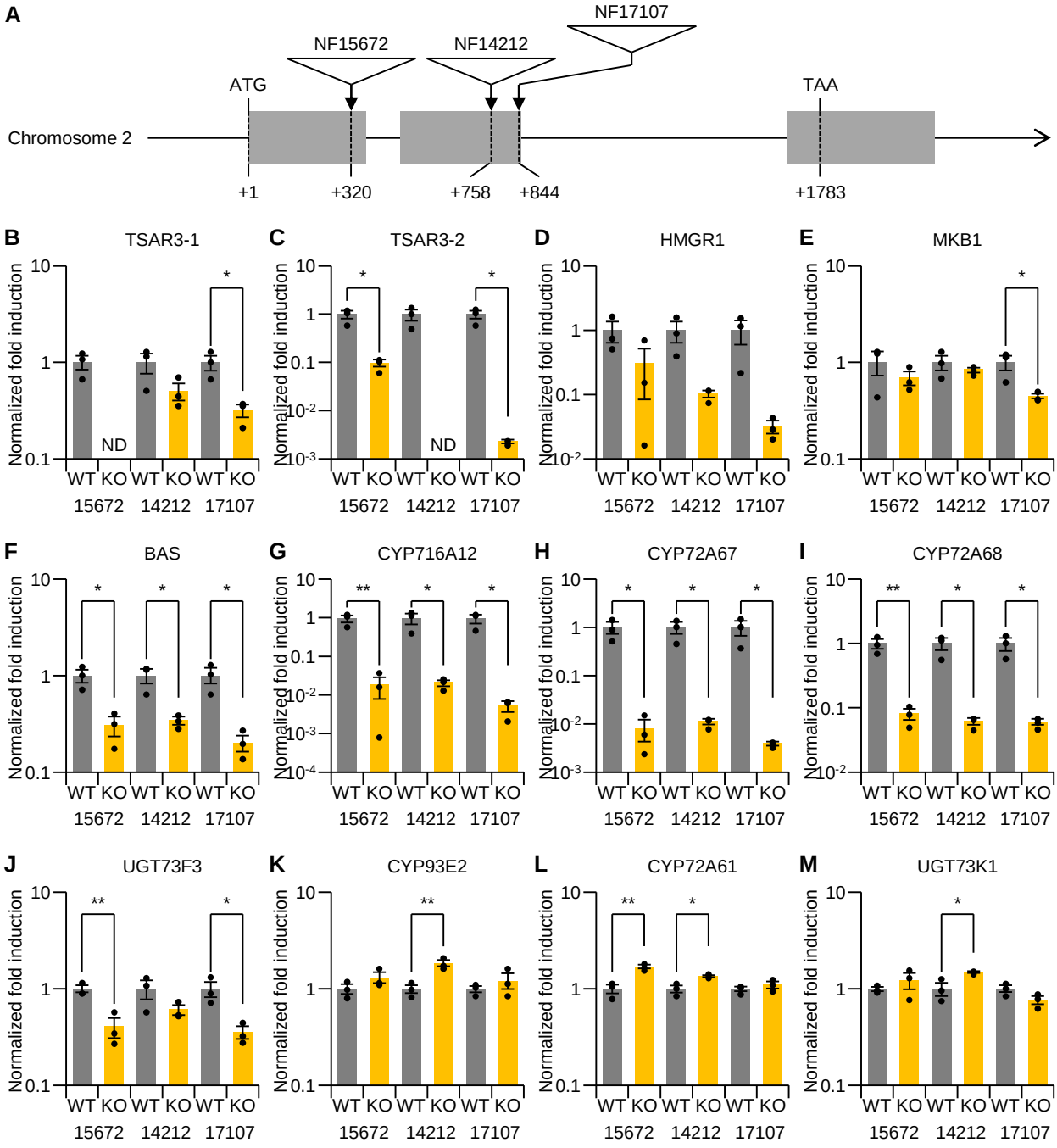


Figure 5. Hemolytic saponin biosynthesis is affected in seeds of *tsar3* plants. (**A**) Schematic representation of *TSAR3*. The locations of the *Tnt1* retrotransposon in the insertion lines are indicated with triangles. Exons are represented by gray boxes. (**B-M**) qRT-PCR analysis of triterpene saponin biosynthesis genes in seeds of three independent *tsar3 Tnt1* lines. Values on the y-axis are average normalized fold changes of three plants per insertion line (KO), relative to the average of three individual control (WT) plants. The mean and standard error (n = 3 biological replicates) are shown and dot plots (black dots) are overlaid. Statistical significance was determined by a two-tailed Student's *t*-test (*, $P < 0.05$; **, $P < 0.01$; and ***, $P < 0.001$). TSAR3-1 primers span the first *Tnt1* insertion (NF15672), TSAR3-2 primers span the *Tnt1* insertion in lines NF14212 and NF17107. ND, not detected.

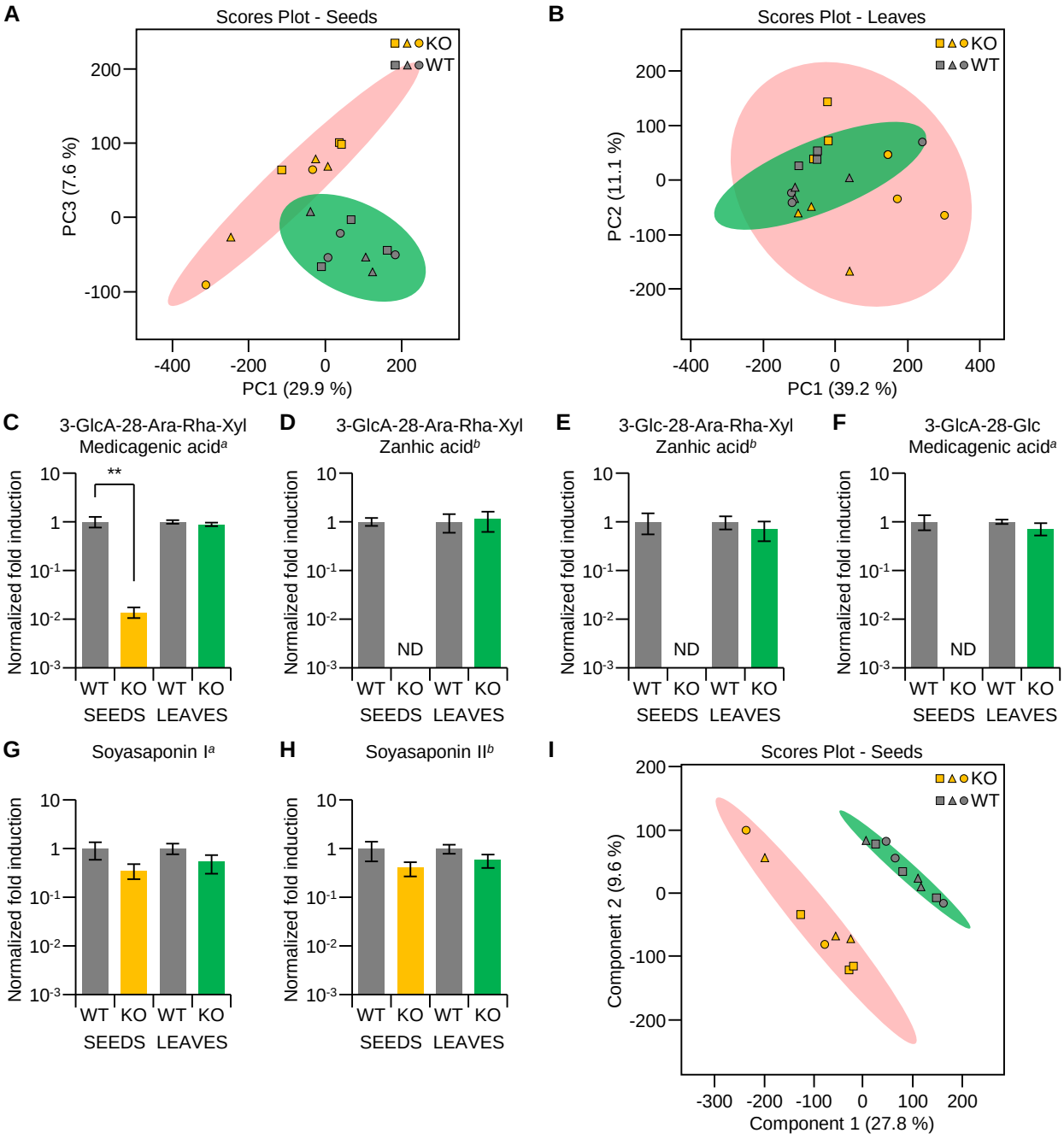


Figure 6. Reduced accumulation of hemolytic triterpene saponins in developing seeds of *tsar3 Tnt1* insertion mutants. (**A-B**) Principal component analysis of samples from *tsar3* (KO) and control (CTR) seeds (**A**) and leaves (**B**). Different symbols indicate the different transgenic lines. (**C-H**) Relative accumulation of triterpene saponins in developing seeds and leaves of *tsar3* (KO) and wild-type (WT) plants. Values on the y-axis are normalized fold changes relative to the average of the WT plants. The mean and standard error (n = 8 or 9 biological replicates) are shown. Statistical significance was determined by a two-tailed Student's *t*-test (**, $P < 0.01$). ND, not detected; metabolites were ([a]) identified using authentic standards or ([b]) predicted based on their accurate mass and MS/MS fragmentation spectra (Supplemental Dataset 2). (**I**) Partial least squares discriminant analysis (PLS-DA) of the samples from developing seeds of the *tsar3* (KO) and wild-type (WT) plants. Different symbols indicate the different transgenic lines.

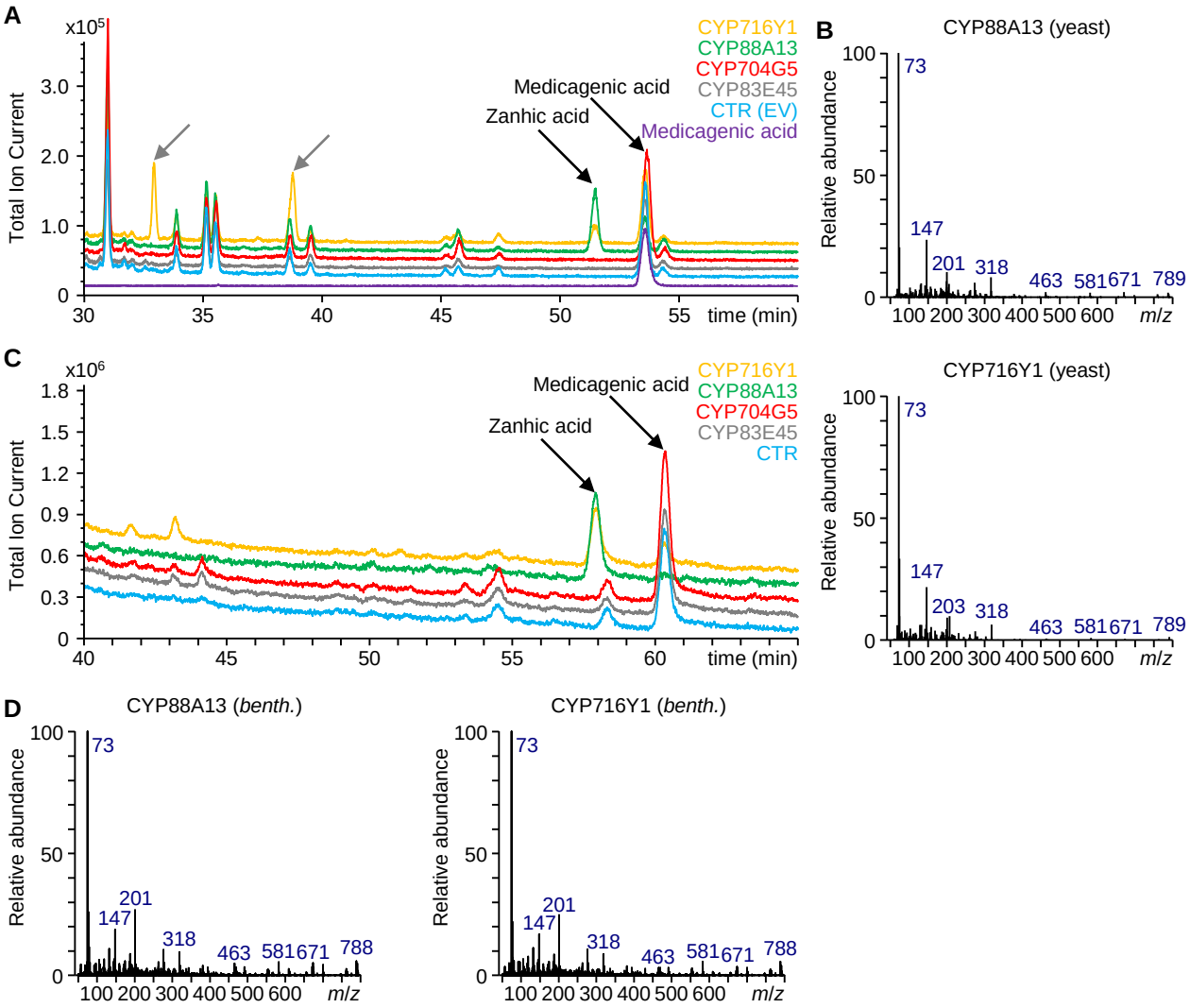


Figure 7. CYP88A13 catalyzes the C-16α hydroxylation of medicagenic acid in yeast and *N. benthamiana*. (**A**) Overlay of GC-MS chromatograms of a medicagenic acid-producing yeast strain expressing the candidate P450s, *CYP716Y1* or an empty vector (EV) control (CTR). Zanhic acid accumulates only in the strain expressing the characterized C-16α hydroxylase *CYP716Y1* and *CYP88A13*. Next to zanhic acid, several other C-16α hydroxylated triterpenoids (indicated with gray arrows; Supplemental Figure 5) accumulate in the yeast strain expressing *CYP716Y1*. (**B**) Comparison of the EI-MS spectra of trimethylsilylated zanhic acid produced by yeast expressing the characterized C-16α hydroxylase *CYP88A13* (top) or *CYP716Y1* (bottom). **(C)** Overlay of GC-qTOF-MS chromatograms of medicagenic acid-producing *N. benthamiana* leaves infiltrated with the candidate P450s or *CYP716Y1*. Zanhic acid accumulates only in the leaves infiltrated with the characterized C-16α hydroxylase *CYP716Y1* and *CYP88A13*. **(D)** Comparison of the EI-MS spectra of trimethylsilylated zanhic acid produced by *N. benthamiana* leaves infiltrated with *CYP88A13* (left) or *CYP716Y1* (right).

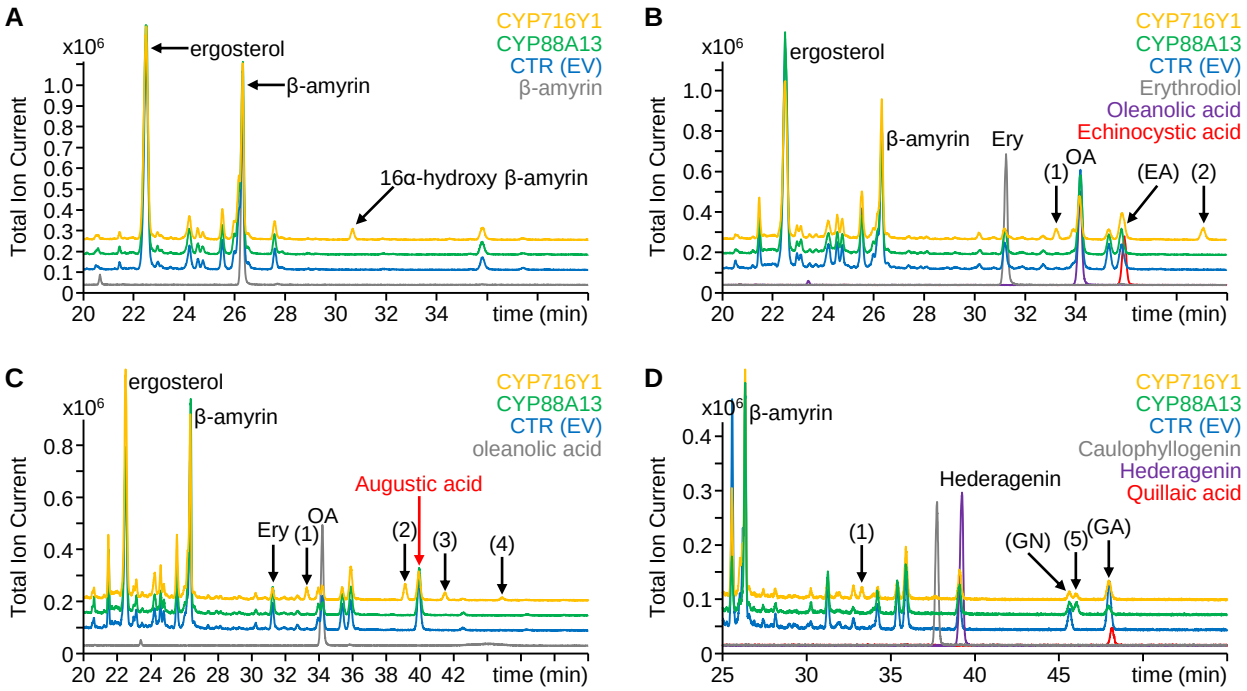


Figure 8. Investigation of the substrate tolerance of CYP88A13 in yeast. (**A**) Overlay of GC-MS chromatograms of a β-amyrin standard and extracts of the spent medium of a β-amyrin-producing yeast strain expressing *CYP88A13*, *CYP716Y1* or an empty vector (EV) control (CTR). 16α-hydroxy β-amyrin accumulates only in the β-amyrin-producing yeast strain expressing *CYP716Y1*. (**B**) Overlay of GC-MS chromatograms of authentic erythrodiol (Ery), oleanolic acid (OA) and echinocystic acid (EA) standards and extracts of the spent medium of an oleanolic acid-producing yeast strain expressing *CYP88A13*, *CYP716Y1* or an empty vector control. Echinocystic acid, which elutes close to a metabolite (tris(2,4-di-tert-butylphenyl) phosphate; Supplemental Figure 8) present in all samples, primulagenin A (1), and 16α-hydroxy oleanolic aldehyde (2) accumulate only in the oleanolic acid-producing yeast strain expressing *CYP716Y1*. (**C**) Overlay of GC-MS chromatograms of an oleanolic acid standard and extracts of the spent medium of an augustic acid-producing yeast strain expressing *CYP88A13*, *CYP716Y1* or an empty vector control. 2β-hydroxy echinocystic acid (3) and 2-oxo echinocystic acid (4) accumulate only in the strain expressing *CYP716Y1*. (**D**) Overlay of GC-MS chromatograms of caulophyllogenin, hederagenin, and quillaic acid standards and extracts of the spent medium of a gypsogenic acid (GA)-producing yeast strain expressing *CYP88A13*, *CYP716Y1* or an empty vector control. 16α-hydroxy gypsogenic acid (5) accumulates in the strains expressing *CYP716Y1* and *CYP88A13*. GN, gypsogenin. The reaction schemes and EI-MS spectra of the identified metabolites and standards are presented in Supplemental Figures 6 to 10.

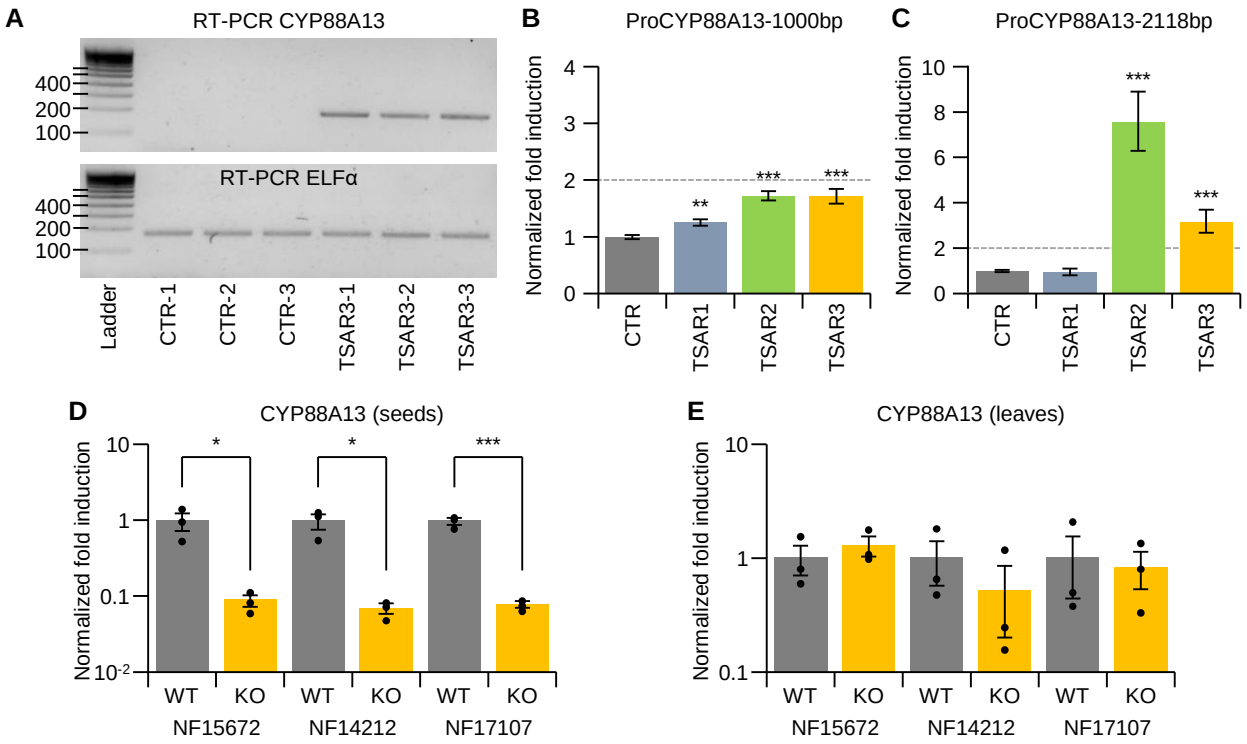


Figure 9. The expression of *CYP88A13* is regulated by TSAR3. (**A**) RT-PCR analysis showing *CYP88A13* expression in three independent transgenic TSAR3^{OE} hairy root lines, but not in the GUS control (CTR) lines (top). The qRT-PCR reference gene *ELFα* was used as a control (bottom). (**B-C**) Transactivation of the 1,000-bp (**B**) and 2,118-bp (**C**) *ProCYP88A13* reporter constructs by TSAR1, TSAR2, and TSAR3 in transfected tobacco protoplasts. Values on the y-axis are normalized fold changes relative to protoplasts co-transfected with the reporter constructs and a *pCaMV35S:GUS* control plasmid (CTR). The gray, dashed line indicates 2-fold induction compared to the control and serves as cut-off. The error bars designate the standard error (n = 8). Statistical significance was determined by a two-tailed Student's *t*-test (*, $P < 0.05$; **, $P < 0.01$; and ***, $P < 0.001$). (**D-E**) qRT-PCR analysis of *CYP88A13* in developing seeds (**D**) or leaves (**E**) of *tsar3 Tnt1* lines. For each *Tnt1* insertion line, three individual homozygous escape (WT) and homozygous knock-out (KO) plants were tested. Values on the y-axis are normalized fold changes relative to the average of three individual control plants. The mean and standard error (n = 3 biological replicates) are shown and dot plots (black dots) are overlaid. Statistical significance was determined by a two-tailed Student's *t*-test (*, $P < 0.05$; **, $P < 0.01$; and ***, $P < 0.001$).

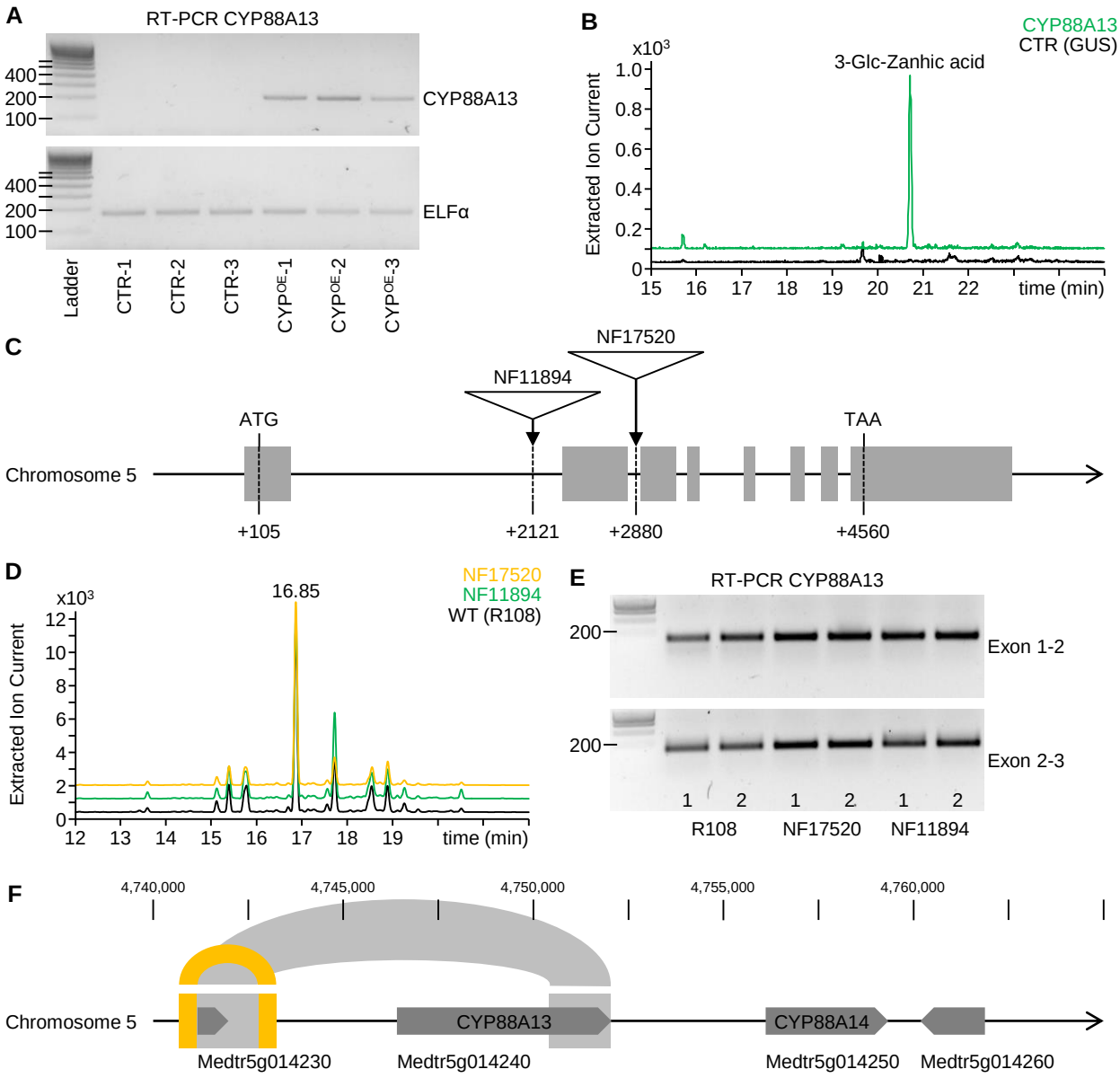


Figure 10. Functional characterization of *CYP88A13* in *M. truncatula*. (**A**) RT-PCR analysis showing *CYP88A13* expression in three independent transgenic CYP88A13^{OE} (CYP^{OE}) hairy root lines, but not in the GUS control (CTR) lines (top). The qRT-PCR reference gene *ELFα* was used as a control (bottom). (**B**) Overlay of the LC-MS chromatograms (extracted ion: 679.37) of an extract of a control hairy root line (black) and an extract of a hairy root line expressing *CYP88A13* (green). 3-Glc-zanhic acid accumulates only in the line expressing *CYP88A13*. (**C**) Schematic representation of *CYP88A13*. The locations of the *Tnt1* retrotransposon in the insertion lines are indicated with triangles. Exons are represented by gray boxes. (**D**). Overlay of the LC-MS chromatograms (extracted ion: 1103.49) of an extract of leaves of wild-type (R108) and homozygous NF17520 and NF11894 *Tnt1* insertion lines. 3-GlcA-28-Ara-Rha-Xyl-Zanhic acid (retention time 16.85 min) still accumulates in the homozygous *Tnt1* insertion lines. (**E**). RT-PCR analysis of *CYP88A13* in two wild-type (R108) and two homozygous NF17520 and NF11894 *Tnt1* insertion lines. Primers spanning the *Tnt1* insertion of NF17520 (top) and NF11894 (bottom) were used for amplification, indicating the *Tnt1* insertion is spliced out during RNA processing. (**F**) Organization of the *CYP88A13-CYP88A14* tandem duplication on chromosome 5 of the *M. truncatula* genome. Genomic repeats are indicated in orange (461 bp) and light gray (1628 bp) boxes, respectively. Genes are indicated with dark gray arrows.

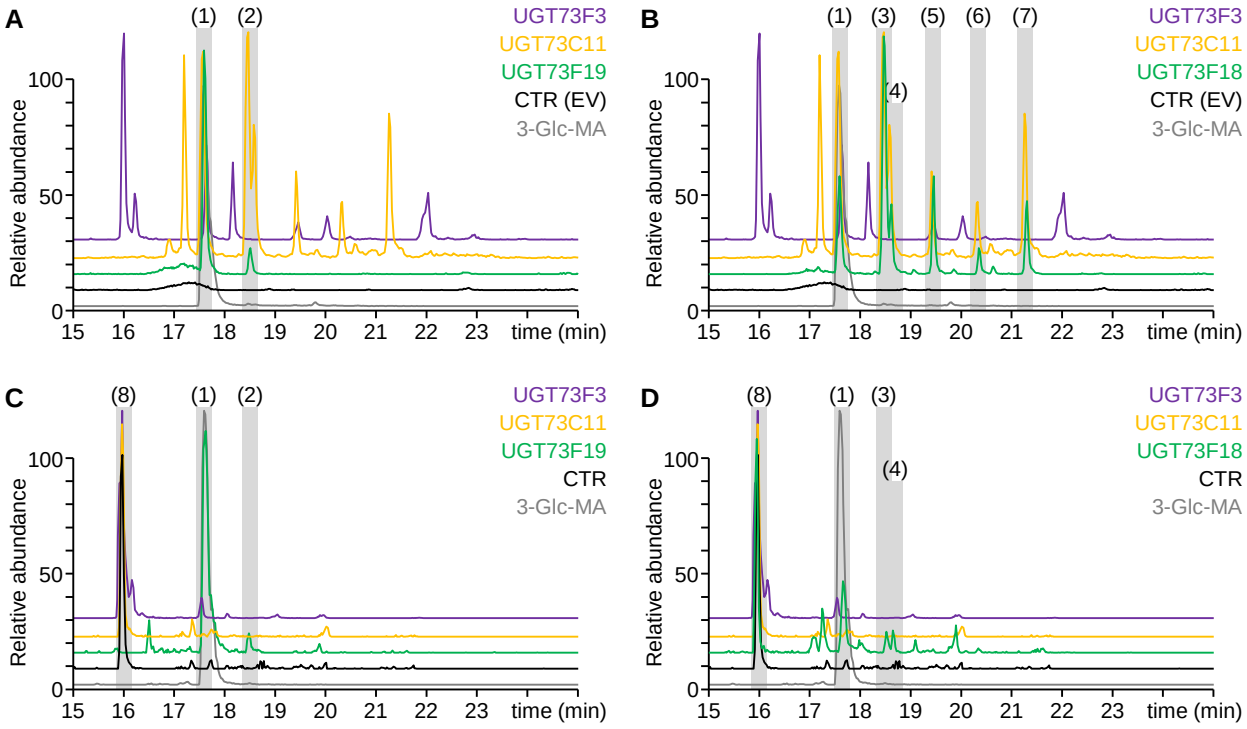
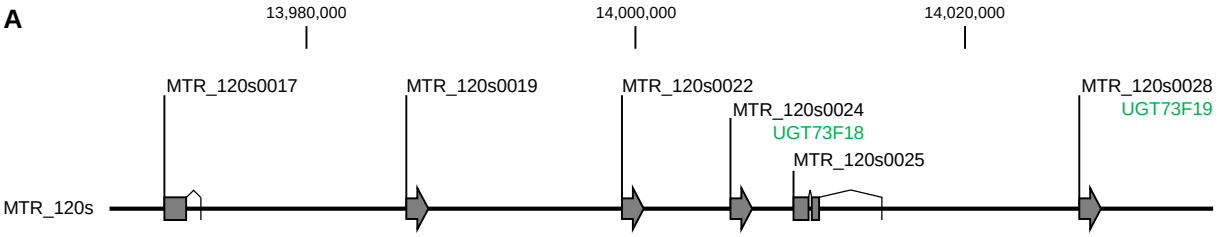


Figure 11. Functional characterization of candidate UGTs in yeast and *N. benthamiana*. (**A**) Overlay of LC-MS chromatograms of a 3-Glc-medicagenic acid (3-Glc-MA) standard and extracts of a medicagenic acid-producing yeast strain expressing the candidate glycosyltransferase *UGT73F19*, the positive controls *UGT73C11* and *UGT73F3*, or an empty vector (EV) control (CTR). (**B**) Overlay of LC-MS chromatograms of a 3-Glc-medicagenic acid standard and extracts of a medicagenic acid-producing yeast strain expressing the candidate glycosyltransferase *UGT73F18*, the positive controls *UGT73C11* and *UGT73F3*, or an empty vector control. (**C**) Overlay of LC-MS chromatograms of a 3-Glc-medicagenic acid standard and extracts of *N. benthamiana* leaves co-infiltrated with *UGT73F19*, the positive controls *UGT73C11* and *UGT73F3*, or a leaf not co-infiltrated with a UGT. (**D**) Overlay of LC-MS chromatograms of a 3-Glc-medicagenic acid standard and extracts of *N. benthamiana* leaves co-infiltrated with *UGT73F18*, the positive controls *UGT73C11* and *UGT73F3*, or a control leaf not co-infiltrated with a UGT. (1), 3-Glc-medicagenic acid; (2), 3-Glc-gypsogenic acid; (3), 3-Glc-hederagenin; (4), 3-Glc-polygalagenin; (5), 3-Glc-gypsogenin; (6), 3-Glc-augustic acid; (7), 3-Glc-oleanolic acid; (8) 28-Glc-medicagenic acid.



B

	MTR_120s0017		MTR_120s0019		MTR_120s0022		MTR_120s0024		MTR_120s0025		MTR_120s0028	
	NT	AA	NT	AA	NT	AA	NT	AA	NT	AA	NT	AA
MTR_120s0017	100	100	87	81	86	80	86	79	73	65	86	82
MTR_120s0019			100	100	93	89	95	92	75	68	93	90
MTR_120s0022					100	100	91	87	74	67	93	91
MTR_120s0024							100	100	75	67	92	88
MTR_120s0025									100	100	74	68
MTR_120s0028											100	100

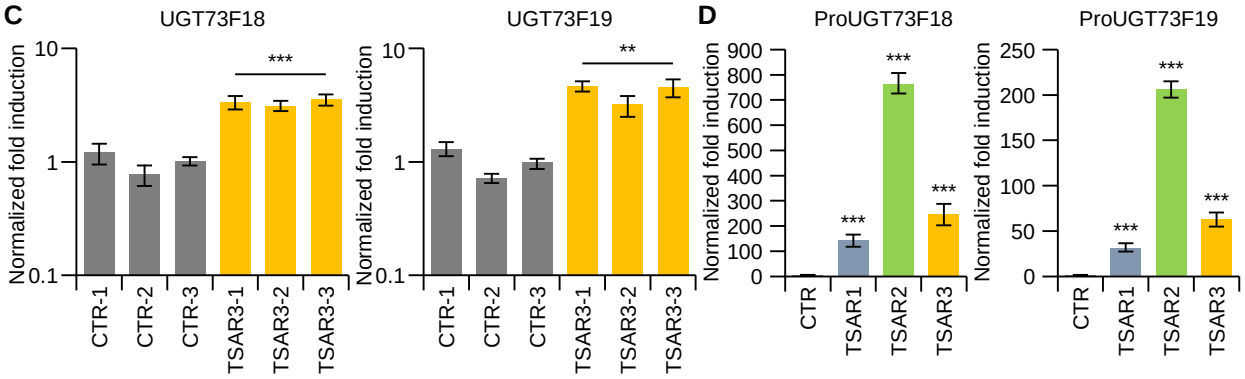


Figure 12. The two identified glucosyltransferases are part of a cluster of six duplicated UGTs. (**A**) Organization of the UGT gene cluster on sequence contig 120s of the *M. truncatula* genome release Mt3.5. (**B**) Sequence identities of the six UGTs on sequence contig 120s at the nucleotide (NT) and amino acid (AA) level. (**C**) qRT-PCR analysis of *UGT73F18* and *UGT73F19* in three independent control (CTR) and TSAR3^{OE} *M. truncatula* hairy root lines. Values on the y-axis are normalized fold changes relative to the average of the three control lines. The error bars designate the standard error (n = 3 technical replicates). Statistical significance was determined by a two-tailed Student's *t*-test (**, $P < 0.01$; and ***, $P < 0.001$). (**D**) Transactivation of the 2000-bp *ProUGT73F18* and *ProUGT73F19* reporter constructs by TSAR1, TSAR2, and TSAR3 in transfected tobacco protoplasts. Values on the y-axis are normalized fold changes relative to protoplasts co-transfected with the reporter constructs and a *pCaMV35S:GUS* control plasmid (CTR). The error bars designate the standard error (n = 8). Statistical significance was determined by a two-tailed Student's *t*-test (***, $P < 0.001$).

A Seed-Specific Regulator of Triterpene Saponin Biosynthesis in *Medicago truncatula*

Bianca Ribeiro, Elia Lacchini, Keylla U. Bicalho, Jan Mertens, Philipp Arendt, Robin Vanden Bossche, Gabriela Calegario, Lore Gryffroy, Evi Ceulemans, Julia Buitink, Alain Goossens, and Jacob Pollier

The following supplemental materials are available for this article:

SUPPLEMENTAL FIGURES:

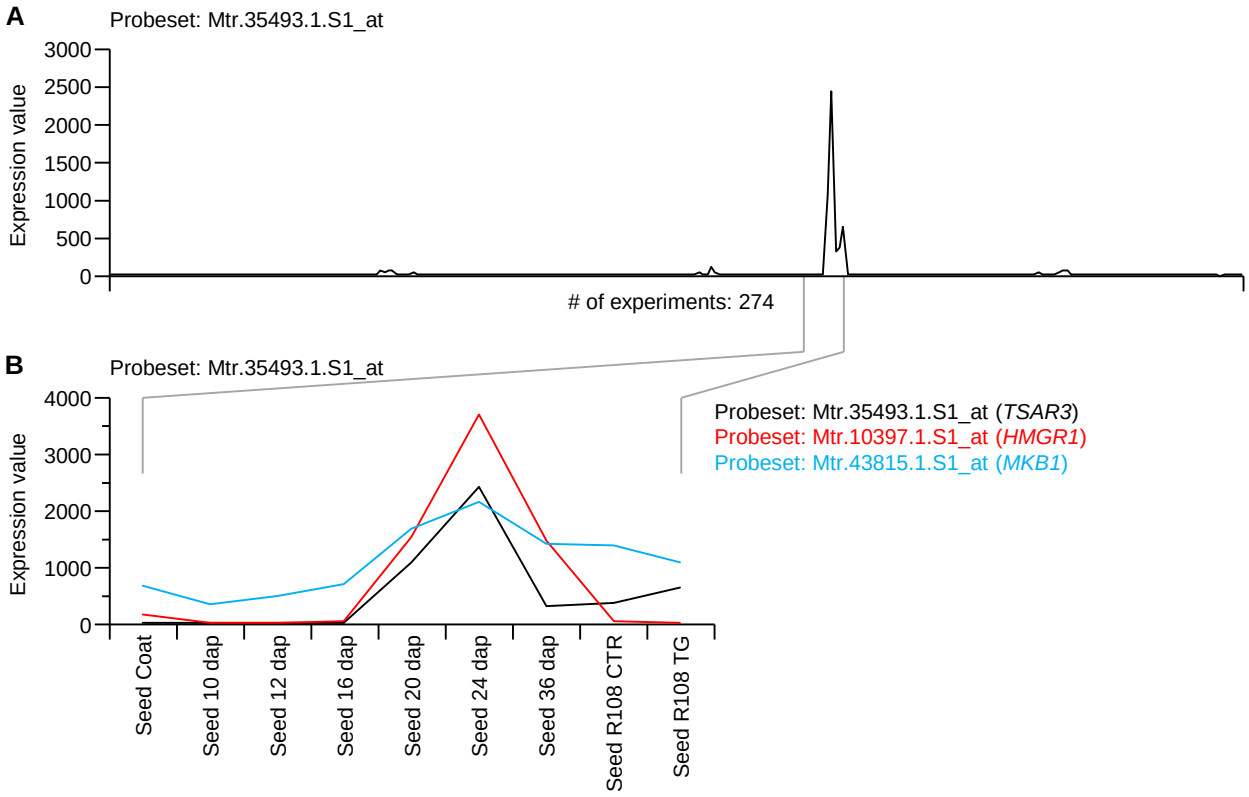
Supplemental Figure 1. Expression of *TSAR3* according to the *Medicago truncatula* gene expression atlas.
Supplemental Figure 2. Partial least squares discriminant analysis (PLS-DA) of the samples from the TSAR3^{OE} and control (CTR) hairy roots.
Supplemental Figure 3. Genotyping of the *TSAR3 Tnt1* insertion lines.
Supplemental Figure 4. Hemolytic saponin biosynthesis is not affected in leaves of *tsar3* plants.
Supplemental Figure 5. EI-MS spectra of additional metabolites identified in Figure 7A.
Supplemental Figure 6. EI-MS spectra and reaction scheme of the metabolites identified in Figure 8A.
Supplemental Figure 7. Reaction schemes of the metabolites identified in Figures 8B and 8D.
Supplemental Figure 8. EI-MS spectra of the metabolites identified in Figure 8B.
Supplemental Figure 9. EI-MS spectra and reaction scheme of the metabolites identified in Figure 8C.
Supplemental Figure 10. EI-MS spectra of the metabolites identified in Figure 8D.
Supplemental Figure 11. Investigation of the substrate tolerance of CYP88A13 in *N. benthamiana*.
Supplemental Figure 12. Genotyping of the *CYP88A13 Tnt1* insertion lines.
Supplemental Figure 13. Analysis of CYP88A14 in yeast and *N. benthamiana*.
Supplemental Figure 14. Phylogenetic analysis of the identified *M. truncatula* and *Arabidopsis thaliana* clade IVa proteins.
Supplemental Figure 15. Comparison of the enzymatic reactions catalyzed by CYP88A13 and ent-kaurenoic acid oxidase (KAO) belonging to the CYP88A subfamily.
Supplemental Figure 16. Promoter replacement in strain PA147, leading to strain JP034.
Supplemental Figure 17. Development of *M. truncatula* seeds.

SUPPLEMENTAL TABLES:

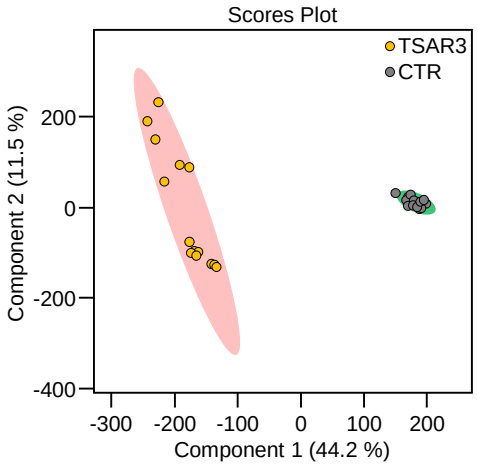
Supplemental Table 1. Yeast strains used in this study.
Supplemental Table 2. Candidate UGTs identified through the different co-expression analyses.

SUPPLEMENTAL DATASETS:

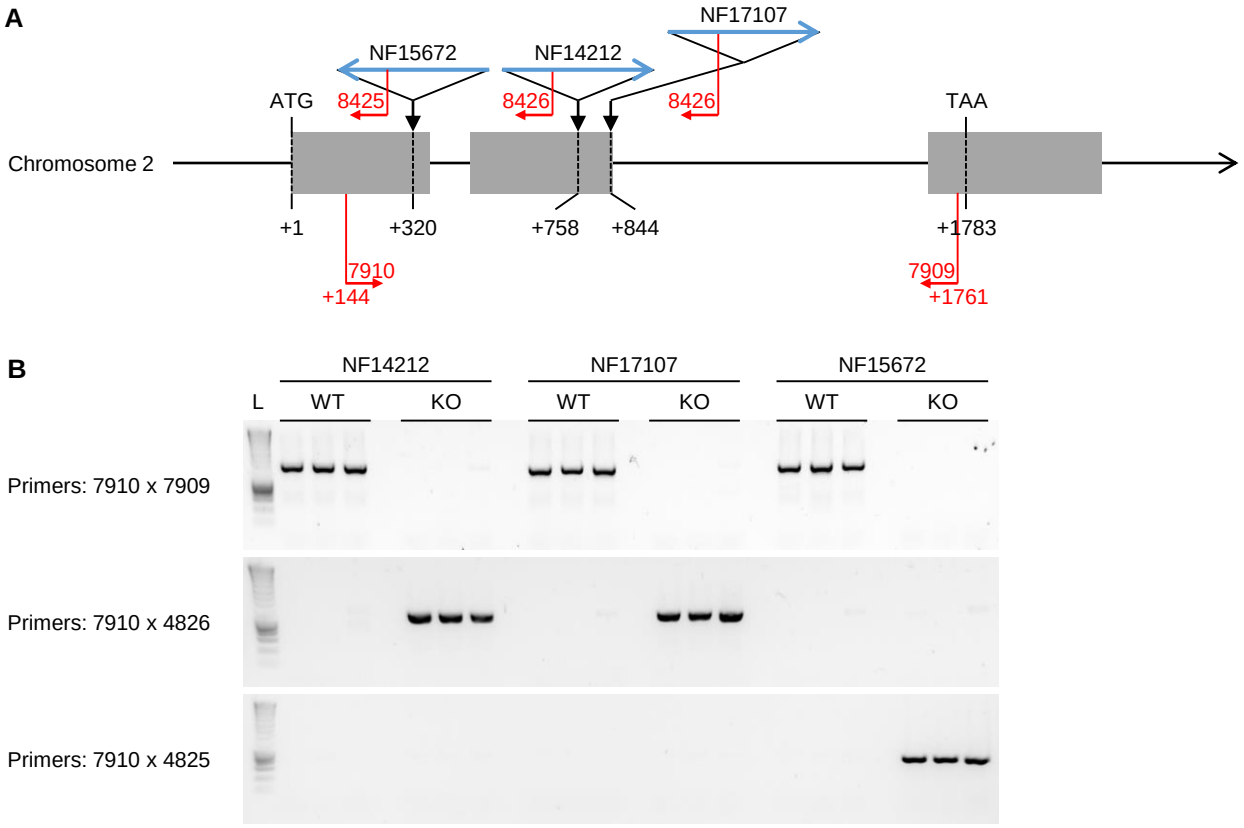
Supplemental Dataset 1. Differential metabolites between the TSAR3^{OE} and GUS control lines (separate Excel file).
Supplemental Dataset 2. Differential metabolites in developing seeds of tsar3 and wild-type plants (separate Excel file).
Supplemental Dataset 3. Co-expression analysis in publicly available microarray datasets from developing *M. truncatula* seeds (separate Excel file).
Supplemental Dataset 4. Co-expression analysis in a set of 60 microarray datasets from developing *M. truncatula* seeds (separate Excel file).
Supplemental Dataset 5. List of oligonucleotides used in this study (separate Excel file).
Supplemental Dataset 6. Sequence alignment of the bHLH domains of the identified *M. truncatula* and *A. thaliana* clade IVa proteins.
Supplemental Dataset 7. Machine-readable tree file in Newick format in support of Supplemental Figure 14.



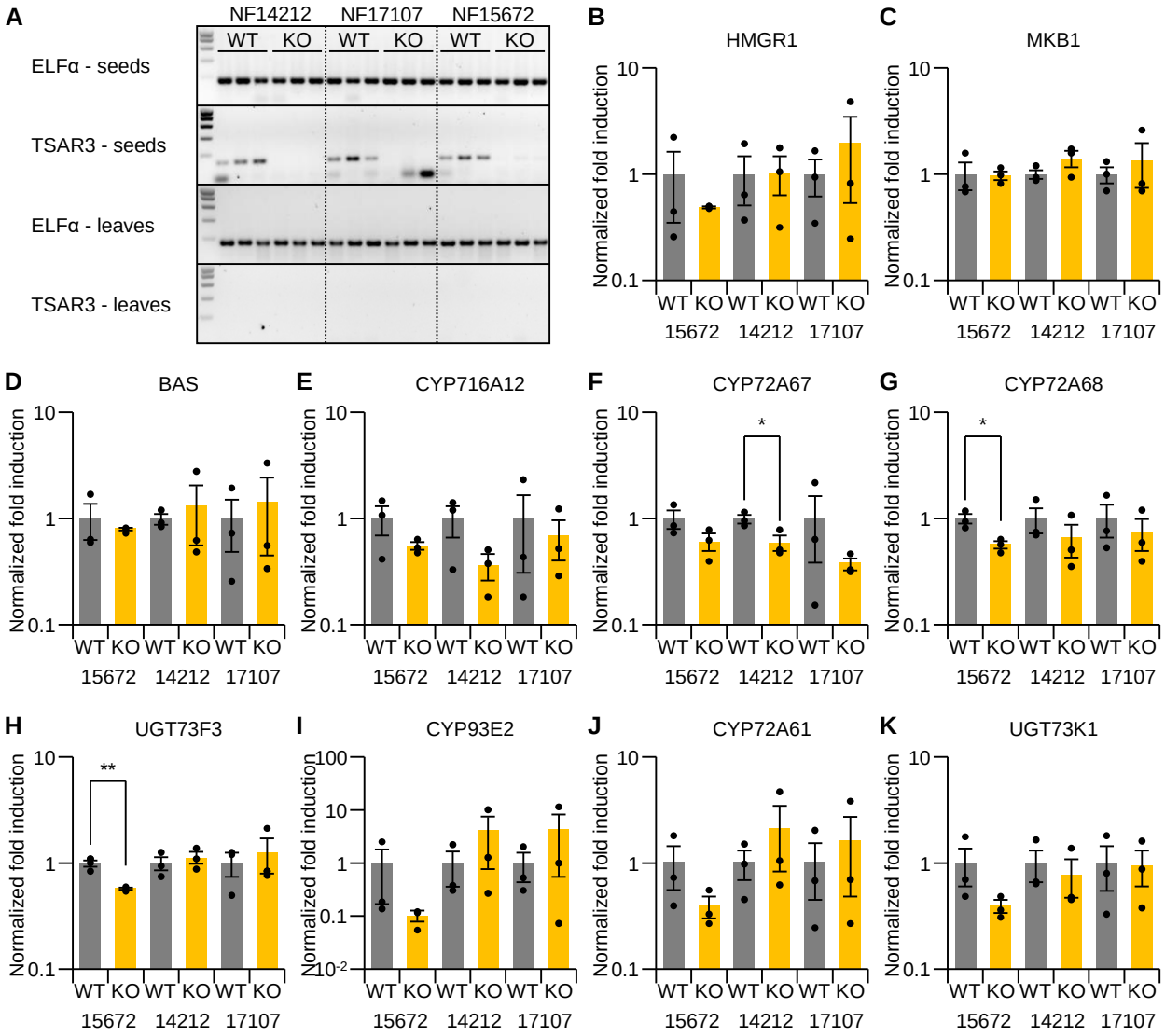
Supplemental Figure 1. Expression of *TSAR3* according to the *Medicago truncatula* gene expression atlas (supports a statement in the first results section). **(A)** Expression of *TSAR3* (represented by probeset Mtr.35493.1.S1_at) in all 274 experiments included in the *Medicago truncatula* gene expression atlas. **(B)** Expression of *TSAR3* (black) in developing seeds overlaid with the expression profile of HMGR1 (red, probeset Mtr.10397.1.S1_at) and MKB1 (blue, probeset Mtr.43815.1.S1_at). Pearson's correlation coefficients of 0.91 and 0.87 between *TSAR3* and *HMGR1*, and *TSAR3* and *MKB1*, respectively, indicate a certain degree of co-expression.



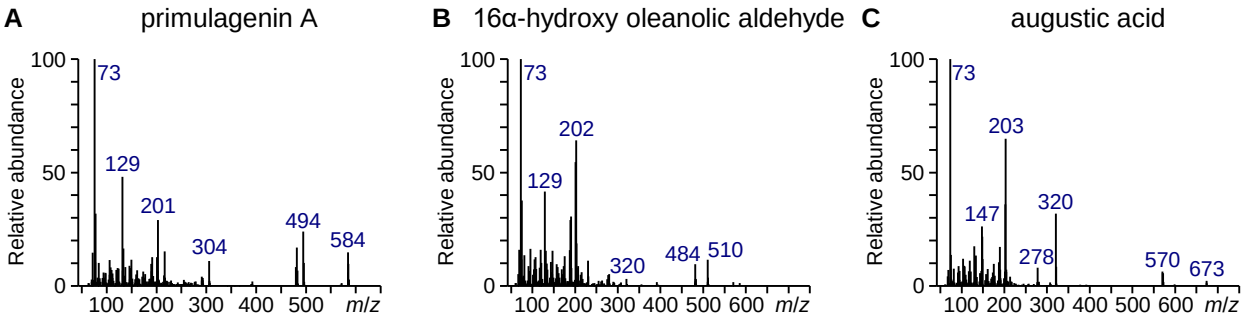
Supplemental Figure 2. Partial least squares discriminant analysis (PLS-DA) of the samples from the TSAR3^{OE} and control (CTR) hairy roots (Supports Figure 4A). A clear separation of the CTR and TSAR3^{OE} samples is observed. Each *m*/*z* feature within this PLS-DA model has a Variable Importance in Projection (VIP) score, which estimates the importance of the *m*/*z* feature in the PLS-DA model. Interestingly, the metabolite with the highest VIP score was 3-GlcA-28-Ara-Rha-Xyl-Zanhic acid. Furthermore, many of the saponins identified in Supplemental Dataset 1 have a high VIP score (>1) and can be considered important features in the PLS-DA model.



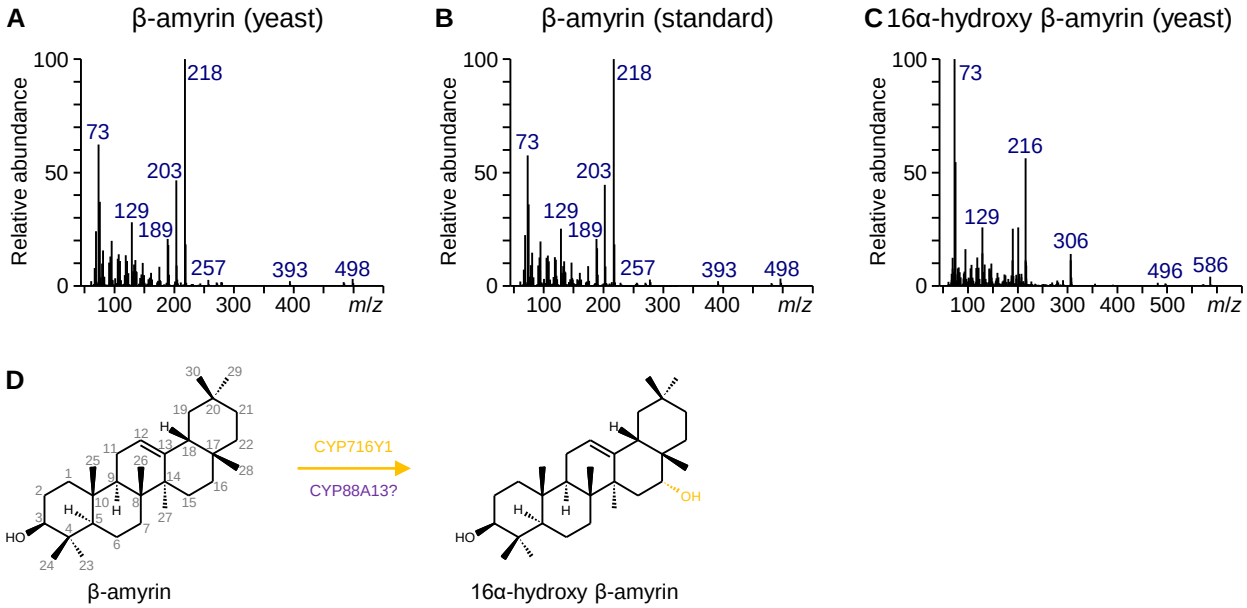
Supplemental Figure 3. Genotyping of the *TSAR3 Tnt1* insertion lines (Supports Figure 5). (**A**) Schematic representation of the *TSAR3* locus on chromosome 2. The locations of the *Tnt1* retrotransposon in the insertion lines are indicated with triangles and the arrow indicates the orientation (5'-3') of the *Tnt1* insertion. Exons are represented by gray boxes. Genotyping primers are indicated with small red arrows. (**B**) Genotyping PCRs of three individual homozygous escape plants (WT) and three individual homozygous *Tnt1* insertion plants (KO) for each of the three *Tnt1* insertion lines. The Eurogentec SmartLadder was used as a size marker (L). Primer sequences can be found in Supplemental Data Set 5.



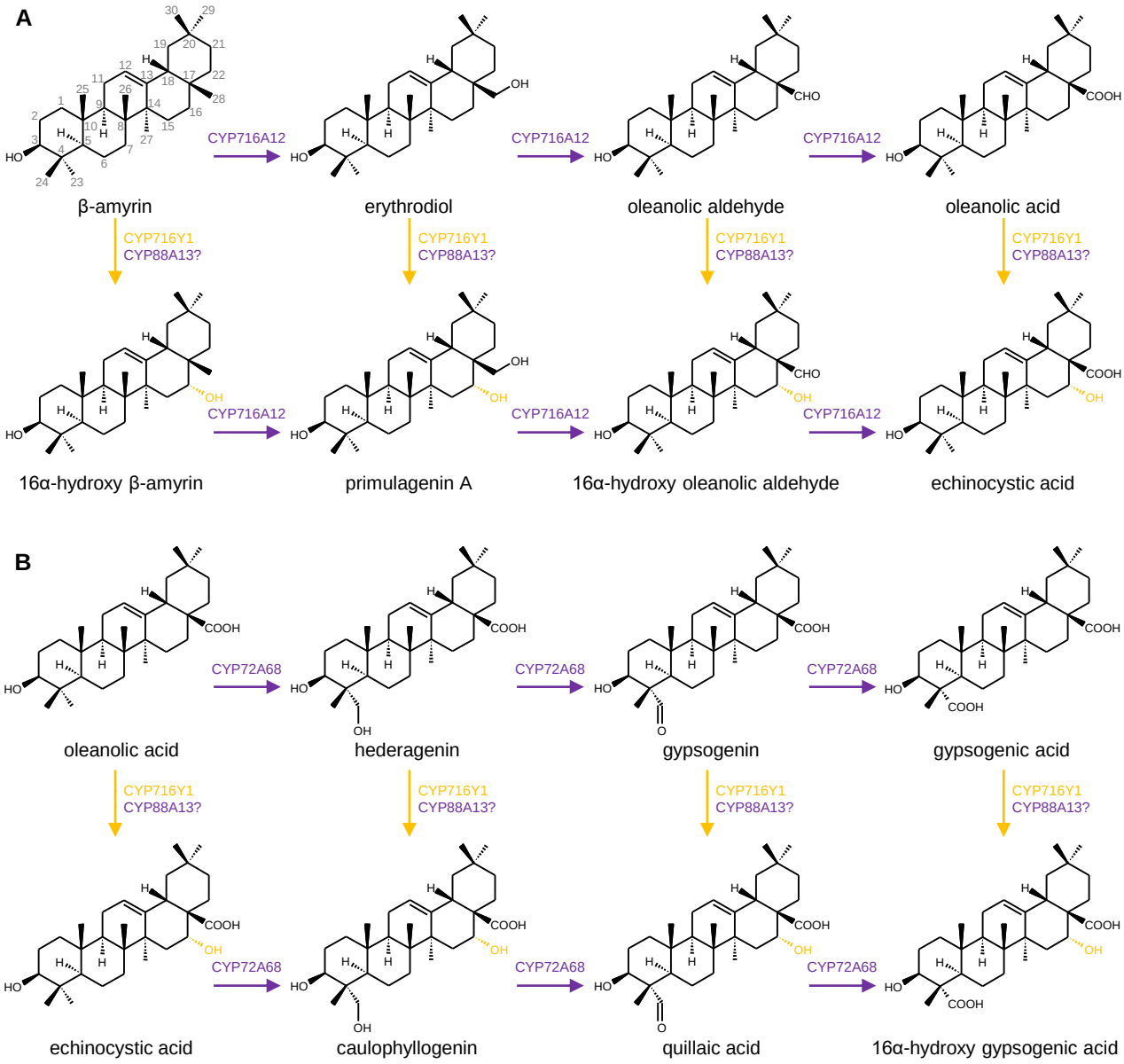
Supplemental Figure 4. Hemolytic saponin biosynthesis is not affected in leaves of *tsar3* plants (Supports Figure 5). (**A**) RT-PCR showing no expression of *TSAR3* in seeds of *tsar3 Tnt1* plants and in leaves of WT and *tsar3 Tnt1* plants. For each *Tnt1* insertion line, three individual homozygous escape (WT) and homozygous knock-out (KO) plants were tested. (**B-K**) qRT-PCR analysis of triterpene saponin biosynthesis genes in leaves of three independent *tsar3 Tnt1* insertion lines. Values on the y-axis are normalized fold changes relative to the average of three individual control plants. The mean and standard error (n = 3 biological replicates) are shown and dot plots (black dots) are overlaid. Statistical significance was determined by a Student's *t*-test (*, $P < 0.05$; **, $P < 0.01$; and ***, $P < 0.001$).



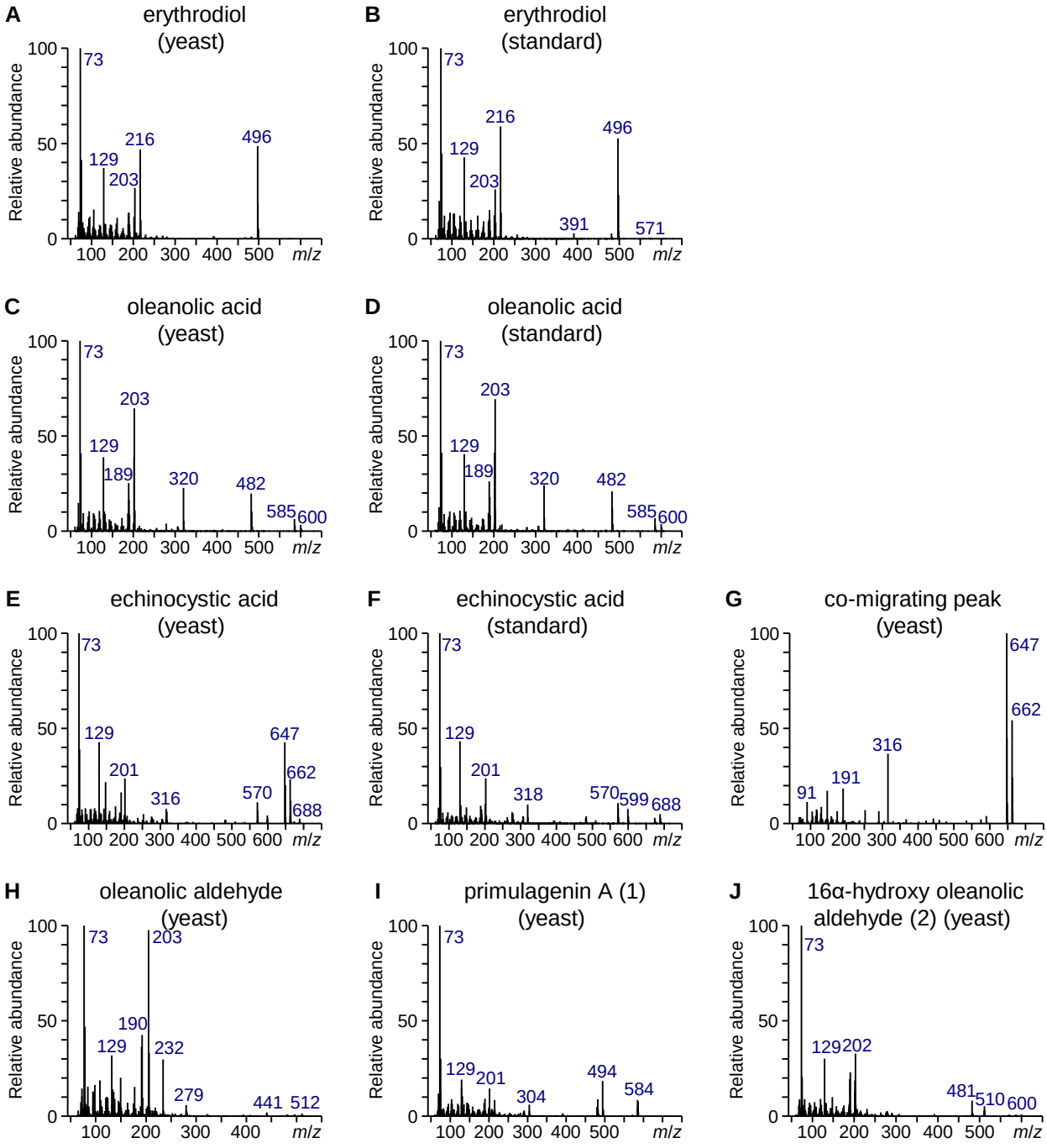
Supplemental Figure 5. EI-MS spectra of additional metabolites identified in Figure 7A. Primulagenin A (**A**) and 16α-hydroxy oleanolic aldehyde (**B**) are unique to the CYP716Y1 sample and elute at 32.93 and 38.78 min, respectively. Augustic acid (**C**) elutes at 38.67 min and partially co-elutes with 16α-hydroxy oleanolic aldehyde.



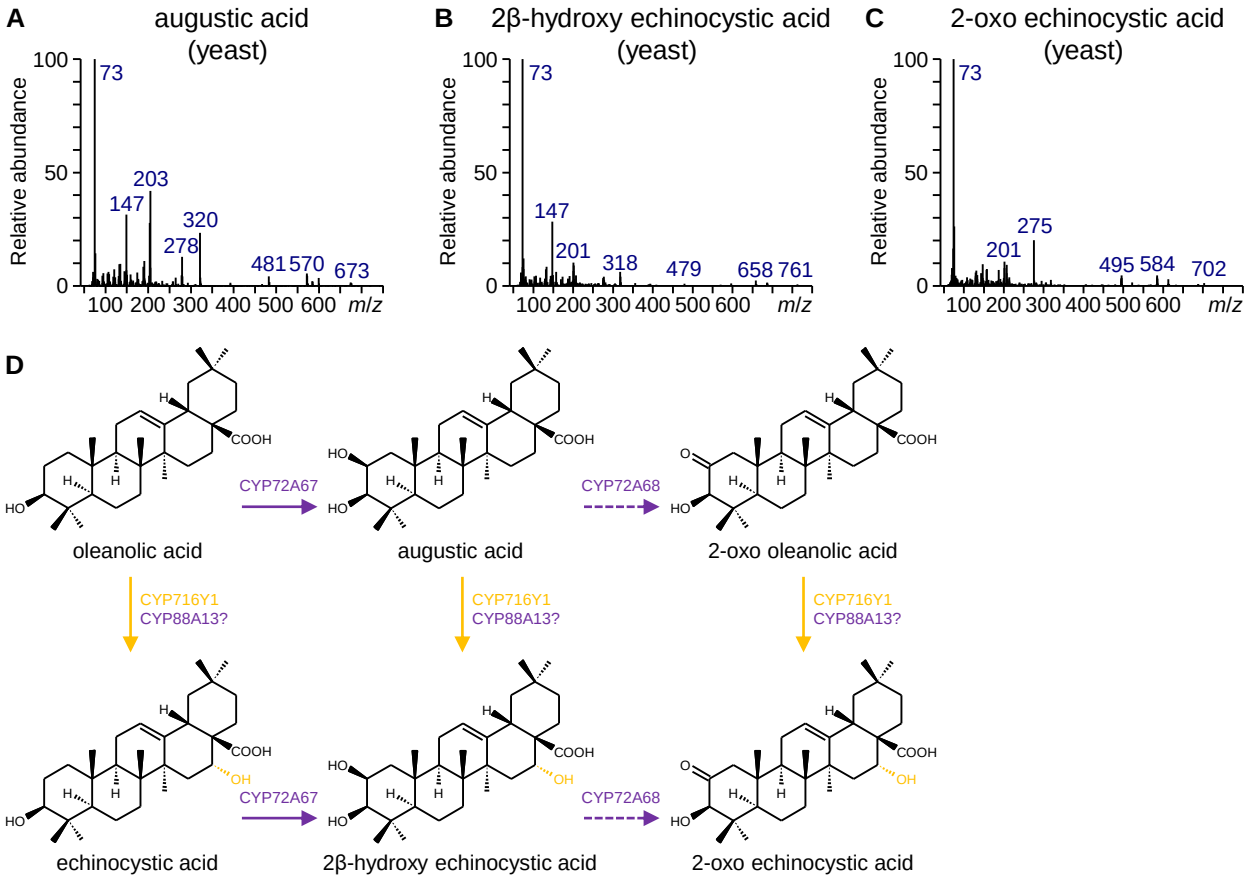
Supplemental Figure 6. EI-MS spectra and reaction scheme of the metabolites identified in Figure 8A.



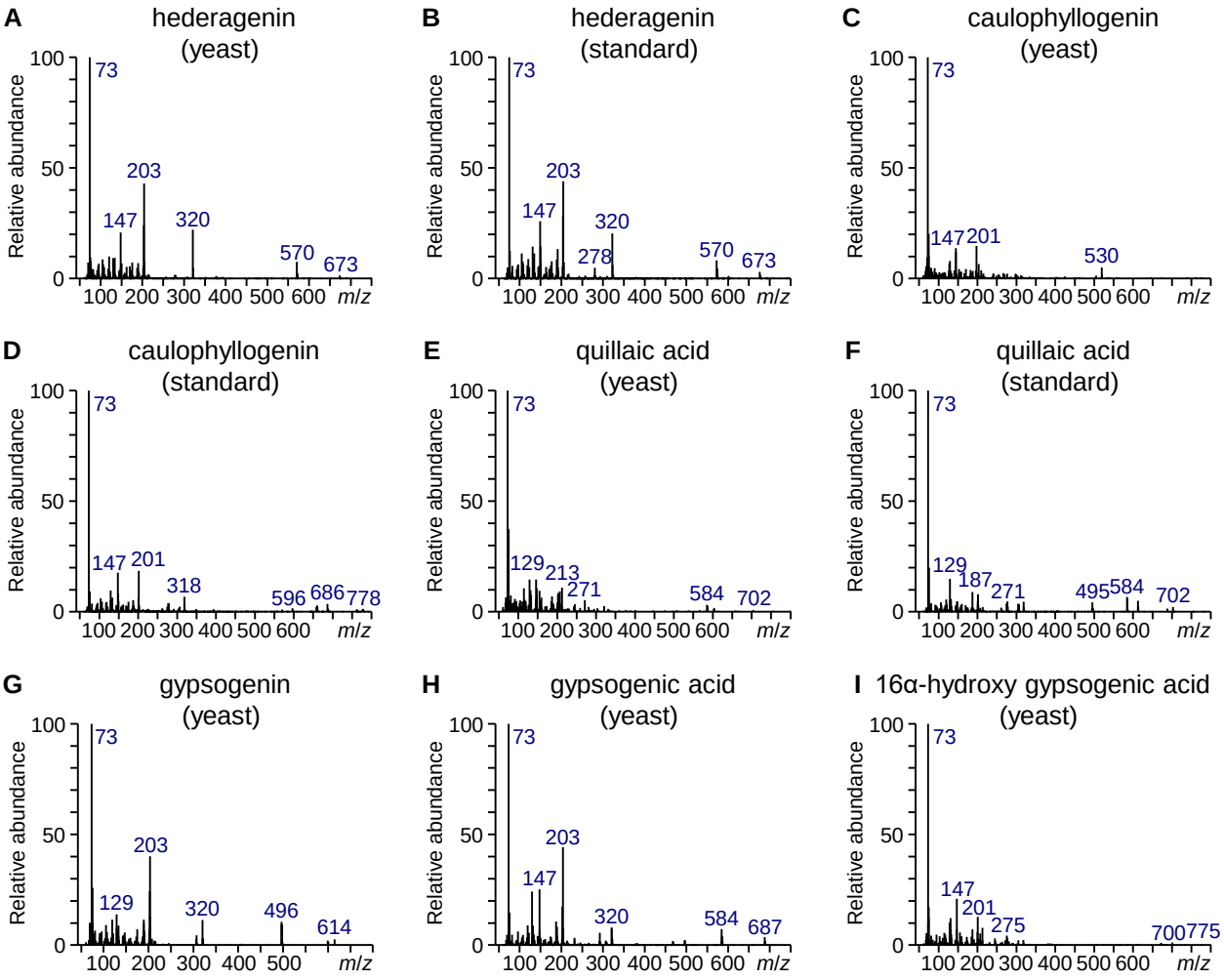
Supplemental Figure 7. Reaction schemes leading to the metabolites identified in Figures 8B (**A**) and 8D (**B**).



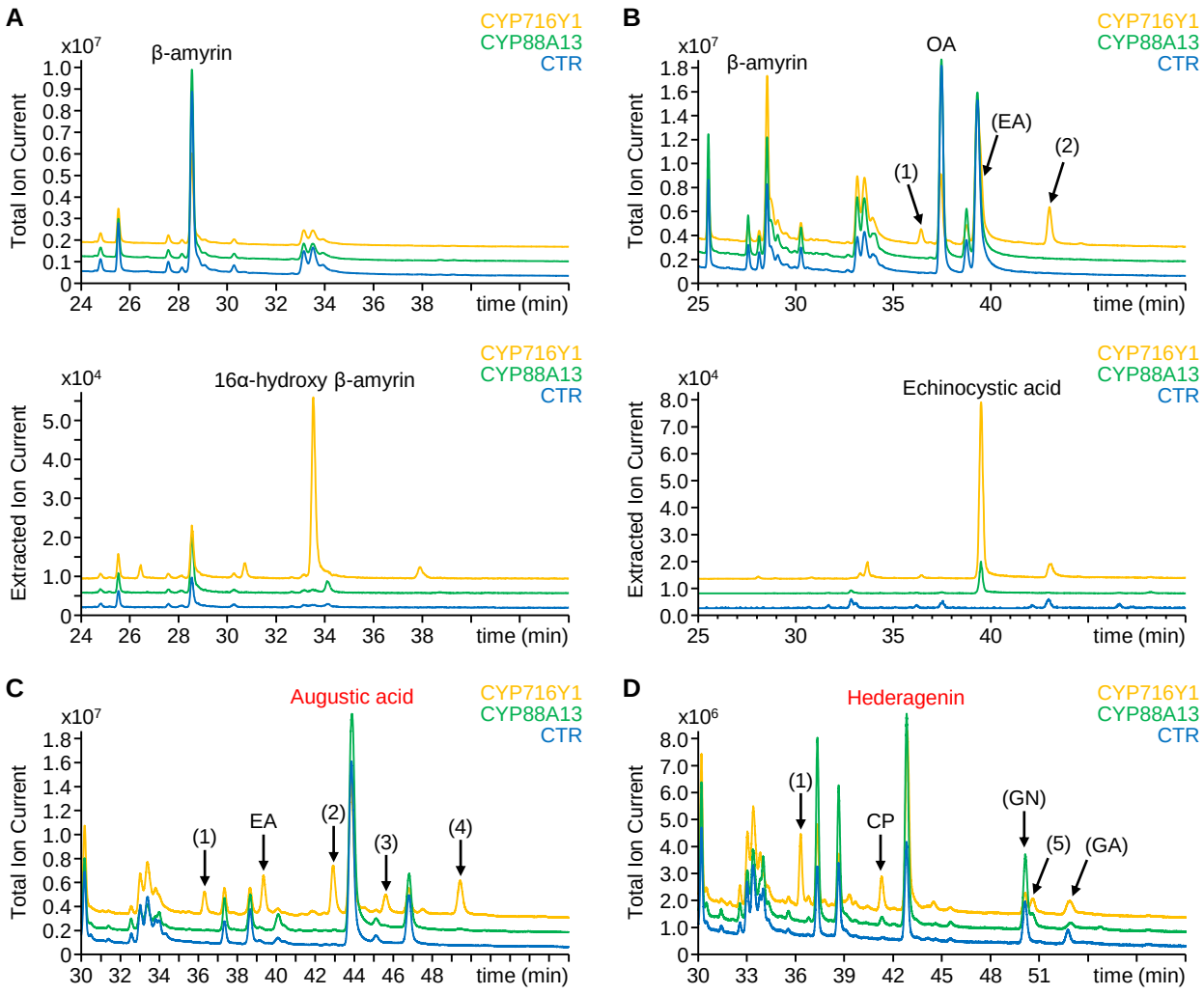
Supplemental Figure 8. EI-MS spectra of the metabolites identified in Figure 8B. A search in the NIST 17 mass spectral library suggests the co-migrating metabolite (**G**) corresponds to tris(2,4-di-tert-butylphenyl) phosphate (Match Factor = 73.8).



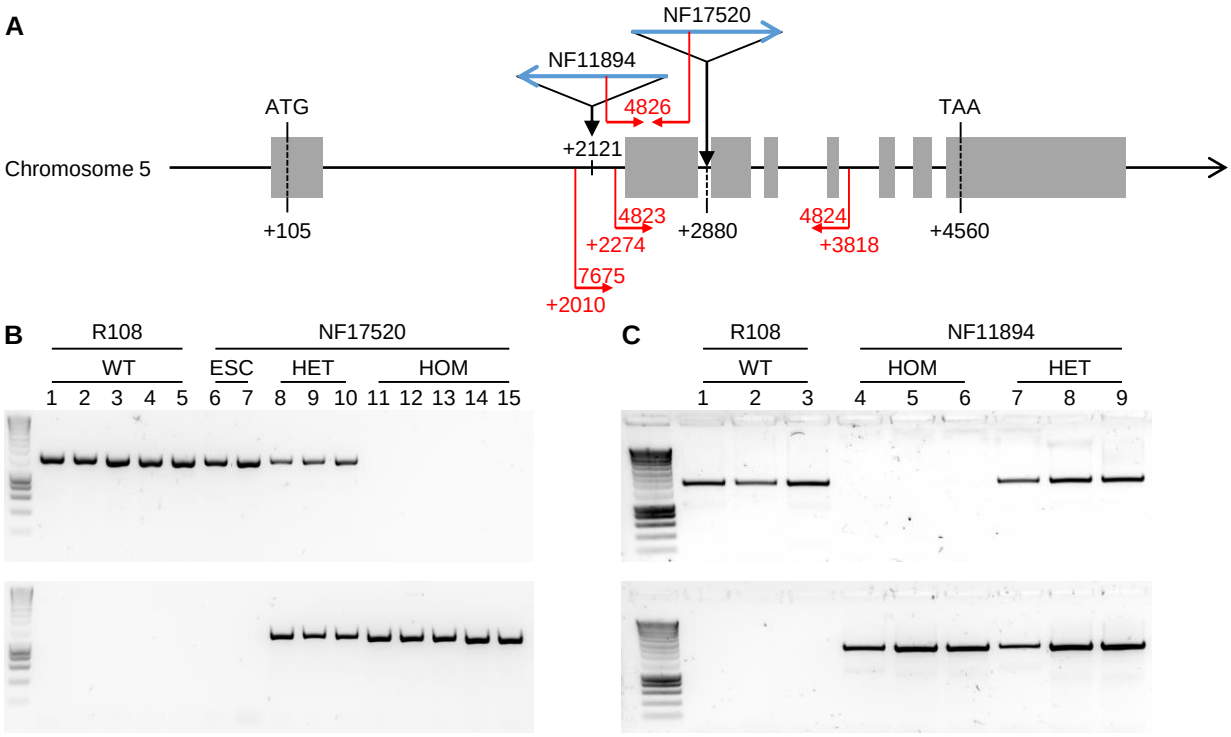
Supplemental Figure 9. EI-MS spectra and reaction scheme of the metabolites identified in Figure 8C.



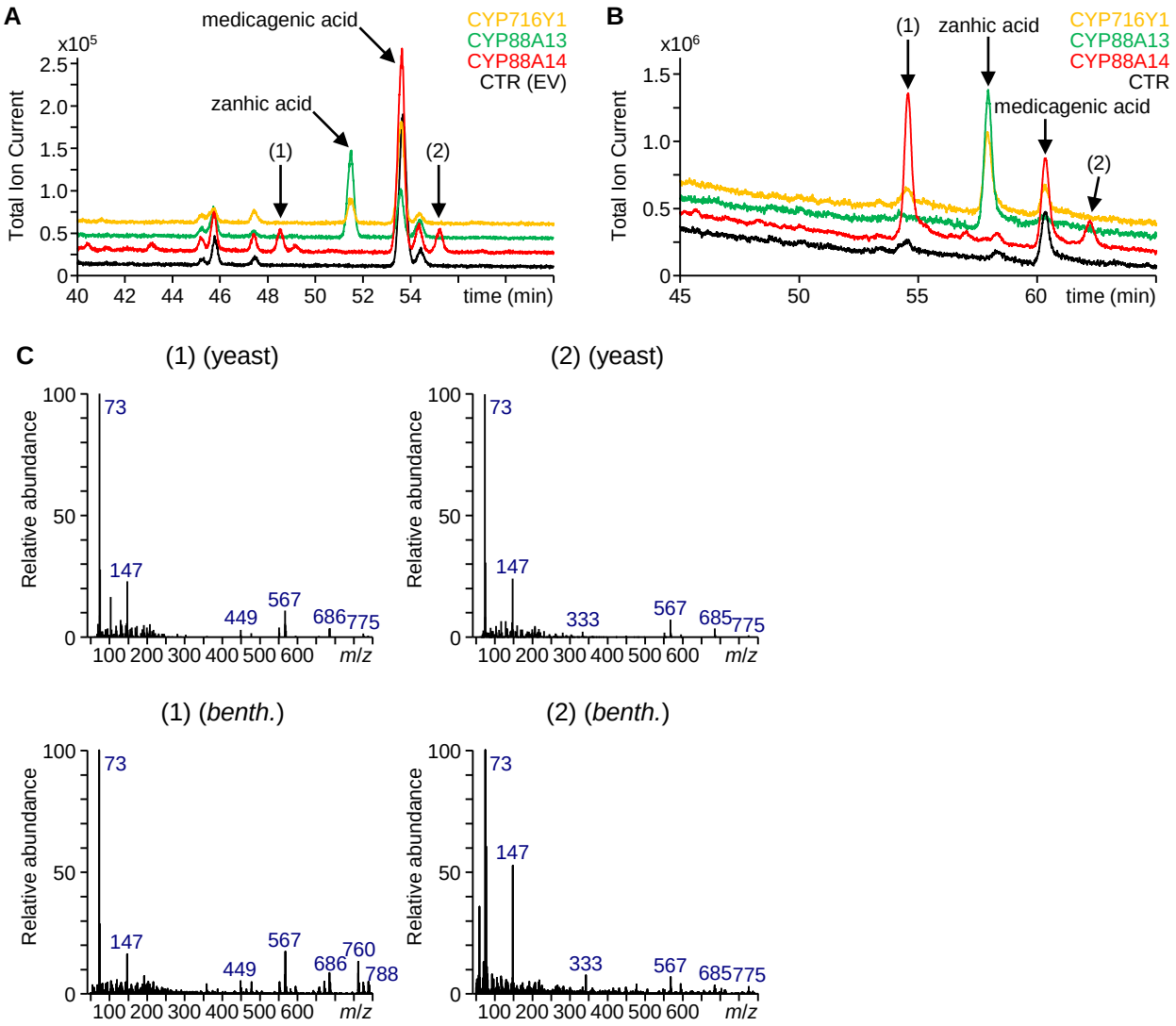
Supplemental Figure 10. EI-MS spectra of the metabolites identified in Figure 8D.



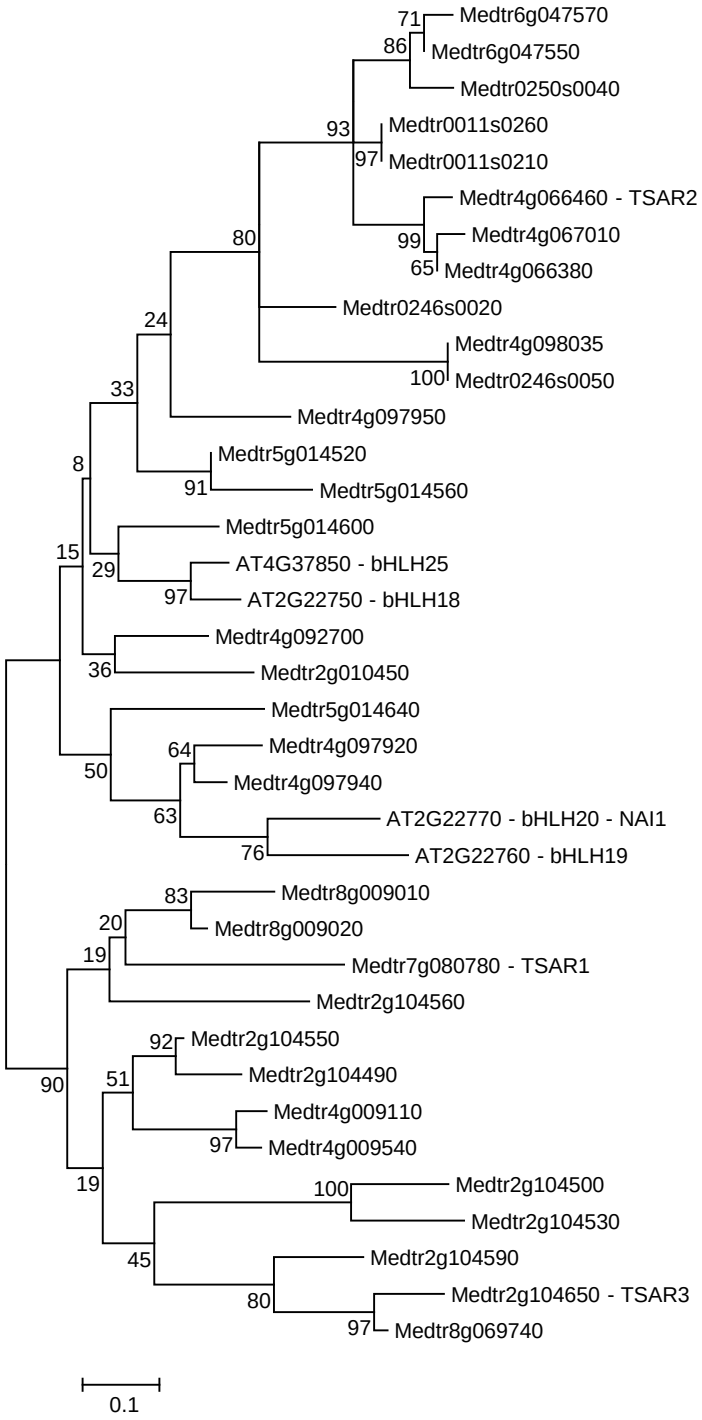
Supplemental Figure 11. Investigation of the substrate tolerance of CYP88A13 in *N. benthamiana* (Supports Figure 8). (**A**) Overlay of the total ion current (top) and extracted ion current (216.19 ± 0.1 Da, bottom) GC-MS chromatograms of β-amyrin-producing *N. benthamiana* leaves expressing *CYP88A13*, *CYP716Y1* or an infiltration control (CTR). 16α-hydroxy β-amyrin accumulates only in leaves infiltrated with *CYP716Y1*. (**B**) Overlay of the total ion current (top) and extracted ion current (570.43 ± 0.1 Da, bottom) GC-MS chromatograms of oleanolic acid-producing *N. benthamiana* leaves expressing *CYP88A13*, *CYP716Y1* or an infiltration control (CTR). Primulagenin A (1) and 16α-hydroxy oleanolic aldehyde (2) accumulate only in the leaves expressing *CYP716Y1*. Echinocystic acid accumulates mainly in the leaves expressing *CYP716Y1*, however, trace amounts were also observed in leaves expressing *CYP88A13*. (**C**) Overlay of the total ion current GC-MS chromatograms of augustic acid-producing *N. benthamiana* leaves expressing *CYP88A13*, *CYP716Y1* or an infiltration control (CTR). 2β-hydroxy echinocystic acid (3) and 2-oxo echinocystic acid (4) accumulate mainly in leaves expressing *CYP716Y1*, however, traces of both metabolites were also detected in leaves expressing *CYP88A13*. (**D**) Overlay of the total ion current GC-MS chromatograms of gypsogenic acid (GA)-producing *N. benthamiana* leaves expressing *CYP88A13*, *CYP716Y1* or an infiltration control (CTR). Caulophyllogenin (CP) and 16α-hydroxy gypsogenic acid (5) accumulate in the leaves expressing both *CYP716Y1* and *CYP88A13*. GN, gypsogenin.



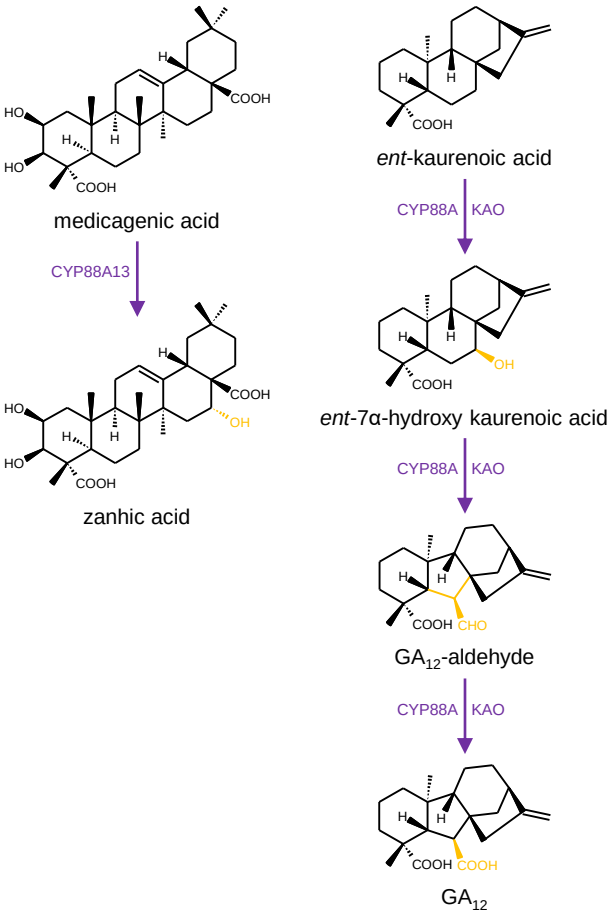
Supplemental Figure 12. Genotyping of the *Tnt1* insertion lines (Supports Figure 10). (**A**) Schematic representation of the *CYP88A13* locus on chromosome 5. The locations of the *Tnt1* retrotransposon in the insertion lines are indicated with triangles and the arrow indicates the orientation (5'-3') of the *Tnt1* insertion. Exons are represented by gray boxes. Genotyping primers for NF17520 and NF11894 are indicated with small red arrows; primer sequences can be found in Supplemental Data Set 5. (**B**) Genotyping PCR of five R108 wild-type (WT) plants and ten NF17520 plants, consisting of two escape (ESC) wild-type, three heterozygous (HET), and five homozygous (HOM) plants. The top gel corresponds to a PCR with primers 4823 and 4824; the lower gel to a PCR with primers 4823 and 4826. (**C**) Genotyping PCR of three R108 wild-type (WT) plants and six NF11894 plants, of which three heterozygous (HET) and three homozygous (HOM) plants. The top gel corresponds to a PCR with primers 7675 and 4824; the lower gel to a PCR with primers 4826 and 4824.



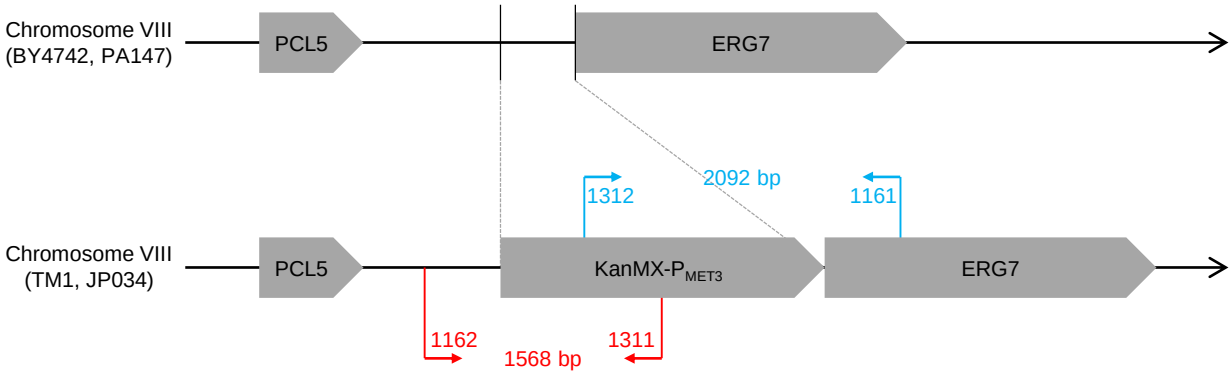
Supplemental Figure 13. Analysis of CYP88A14 activity in yeast and *N. benthamiana*. (Supports Figure 10). (**A**) Overlay of GC-MS chromatograms of extracts of the spent medium of a medicagenic acid-producing yeast strain expressing *CYP716Y1* (orange), *CYP88A13* (green), *CYP88A14* (red) or an empty vector control (black). No zanhic acid accumulates in the strain expressing *CYP88A14*, however, two unknown metabolites (numbered) accumulate in this strain. (**B**) Overlay of GC-MS chromatograms of extracts of medicagenic acid-producing *N. benthamiana* leaves expressing *CYP716Y1* (orange), *CYP88A13* (green), *CYP88A14* (red) or mock infiltration (black). No zanhic acid was produced in leaves infiltrated with *CYP88A14*, however, two unknown metabolites (numbered and indicated with red arrows) accumulate in these leaves. (**C**) EI-MS spectra of the unknown metabolites produced in *CYP88A14*-expressing yeast cells or *N. benthamiana* leaves.



Supplemental Figure 14. Phylogenetic analysis of the identified *M. truncatula* and *Arabidopsis thaliana* clade IVa proteins (Supports a statement in the Discussion section). An alignment of the bHLH domains according to Heim et al. (2003) in which gaps were omitted (complete deletion option) was used as input file to generate a Maximum Likelihood tree using MEGA5 software (Tamura et al., 2011). The Jones, Taylor and Thornton model was applied to determine evolutionary distances. Bootstrap analysis was carried out with 1,000 replicates and bootstrap values are indicated as percentages on the nodes of each clade. The tree branch length indicates the number of substitutions per amino acid site; scale bar: 0.1 substitutions per site.



Supplemental Figure 15. Comparison of the enzymatic reactions catalyzed by CYP88A13 and ent-kaurenoic acid oxidase (KAO) belonging to the CYP88A subfamily (Supports a statement in the Discussion section).



Supplemental Figure 16. Promoter replacement in strain PA147, leading to strain JP034 (Supports a statement in the Methods section). A schematic representation of the genome around the *ERG7* gene in the BY4742-derived strain PA147 (top) and the promoter-replaced strains TM1 (Moses et al., 2014) and JP034 (bottom). Two DNA fragments amplified from strain TM1 were introduced in strain PA147, leading to promoter replacement via homologous recombination. Primer sequences can be found in Supplemental Data Set 5.

Supplemental Figure 17. Development of *M. truncatula* seeds (Supports a statement in the Methods section). (**A**) Early seed development (1 to 14 DAP) with seeds attached to the plant. (**B**) Later stages of the *M. truncatula* seed development. (**C**) Pictures of a seedpod and seeds harvested 24 DAP for the metabolite and transcript profiling of the *tsar3* lines. The size bar indicates 5 mm.

Supplemental Table 1. Yeast strains used in this study.

Strain	Genotype	Reference
BY4742	*MATα*; *his3Δ1*; *leu2Δ0*; *ura3Δ0*; *lys2Δ0*	EUROSCARF
TM1	BY4742; P_{ERG7}::P_{MET3}-*ERG7*	(Moses et al., 2014)[a]
PA059	BY4742; *trp1*	(Miettinen et al., 2017)[b]
PA147	PA059; *pah1*	This study
JP034	PA147; P_{ERG7}::P_{MET3}-*ERG7*	This study

[a]Moses, T. *et al*. (2014). Proc Natl Acad Sci U S A 111 (4): 1634-1639.
[b]Miettinen, K. *et al*. (2017). Nat Commun 8, 14153.

Supplemental Table 2. Candidate UGTs identified through the different co-expression analyses.

Number	Gene ID	UGT ID
Candidate UGTs that were successfully cloned		
1	Medtr2g008225	UGT91H6
2	Medtr4g123553	UGT91H12
3	Medtr5g058320	UGT85H8
4	Medtr5g067170	UGT73F20
5	Medtr6g013990	UGT85H10
6	Medtr6g042310	UGT91H13
7	Medtr7g012120	UGT73F21
8	Medtr7g071040	UGT73P17
9	MTR_120s0024	UGT73F18
10	MTR_120s0028	UGT73F19
Candidate UGTs that we failed to clone		
11	Medtr2g035050	-
12	Medtr4g117970	-
13	Medtr4g117980	-
14	Medtr4g485630	-
15	Medtr5g015260	-
16	Medtr7g070740	-
17	Medtr7g070910	-
18	Medtr7g070740	-
19	Medtr8g032950	-
20	MTR_120s0025	-