The Plant Cell, Vol. 32: 2020–2042, June 2020, www.plantcell.org © 2020 ASPB.

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

**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 14800-900, Brazil
[d] Institut de Recherche en Horticulture et Semences-Unités Mixtes de Recherche, Université d'Angers, INRAE, Institut Agro, SFR 4207 QuaSaV, 49071 Beaucouzé, France
[e] VIB Metabolomics Core, 9052 Ghent, Belgium

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.); 0000-0001-7429-0803 (P.A.); 0000-0001-6407-8139 (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.)

**Plants produce a vast array of defense compounds to protect themselves from pathogen attack or herbivore predation. Saponins are a specific class of defense compounds comprising bioactive glycosides with a steroidal or triterpenoid aglycone backbone. The model legume *Medicago truncatula* synthesizes two types of saponins, hemolytic saponins and nonhemolytic soyasaponins, which accumulate as specific blends in different plant organs. Here, we report the identification of the seed-specific transcription factor TRITERPENE SAPONIN ACTIVATION REGULATOR3 (TSAR3), which controls hemolytic saponin biosynthesis in developing *M. truncatula* seeds. Analysis of genes that are coexpressed with *TSAR3* in transcriptome data sets from developing *M. truncatula* seeds led to the identification of CYP88A13, a cytochrome P450 that catalyzes the C-16α hydroxylation of medicagenic acid toward zanhic acid, the final 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 genes encoding the identified biosynthetic enzymes are present in clusters of duplicated genes in the *M. truncatula* genome. This appears to be a common theme among saponin biosynthesis genes, especially glycosyltransferases, and may be the driving force of the metabolic evolution of saponins.**

## INTRODUCTION

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 type 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 covalently bound to one or more hydrophilic sugar chains, saponins are amphipathic molecules that form a colloidal solution in water, which forms a stable soap-like foam when shaken (Vincken et al., 2007). Due to these physicochemical properties, saponins are often used 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 barrel medic (*Medicago truncatula*), a member of the Fabaceae plant family and a close relative of the important forage crop alfalfa (*Medicago sativa*), 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 due to their bitter, astringent flavor, which deters the attackers from eating the plant. *Medicago* saponins also serve as antinutritive substances by reducing the palatability and digestibility of plants in ruminants and due to their toxicity toward monogastric animals and herbivorous insects (Gholami et al., 2014; Rafińska et al., 2017). Two types of triterpene saponins can be distinguished in *Medicago* species based on their biological properties: the hemolytic

[1] These authors contributed equally to this work.
[2] Address correspondence to alain.goossens@psb-vib.ugent.be and jacob.pollier@psb-vib.ugent.be.
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).
www.plantcell.org/cgi/doi/10.1105/tpc.19.00609

## IN A NUTSHELL

**Background:** Plants are constantly confronted by a variety of threats, such as herbivores and pathogens, and as sessile organisms they cannot flee from their predators or search for more favorable growth conditions. As a defense mechanism, plants produce a vast arsenal of chemicals that may be toxic to their attackers or deter them from eating the plant. One class of such defense compounds are the saponins. These amphipathic glycosides act as deterrents against herbivores due to their bitter, astringent flavor and toxicity. Alfalfa (*Medicago sativa*) is a major hay crop with a high feeding value, but it synthesizes a complex saponin mixture with a certain degree of toxicity towards monogastric animals. We previously characterized two transcription factors, TSAR1 and TSAR2, which regulate saponin biosynthesis in response to stress in *M. truncatula*, a model plant and close relative of alfalfa.

**Question:** We wanted to gain further insight in how saponins are biosynthesized in *M. truncatula* and the genetic control of saponin biosynthesis that leads to the distinct saponin mixtures that accumulate in different plant tissues.

**Findings:** We identified TSAR3, a seed-specific transcription factor that determines the saponin composition in developing *M. truncatula* seeds by controlling the expression of saponin biosynthetic genes. Co-expression analysis with *TSAR3* led to the identification of CYP88A13, a cytochrome P450 that catalyzes the C-16α hydroxylation of medicagenic acid towards zanhic acid, a specific reaction in the biosynthesis of saponin backbones that only occurs in the aerial parts of the plant. In addition, we identified two glucosyltransferases, UGT73F18 and UGT73F19, which transfer a glucose moiety to the C-3 position of the saponin backbones. The genes encoding these saponin biosynthetic enzymes are present in clusters of duplicated genes in the *M. truncatula* genome, which may be a driving force of metabolic evolution in plants.

**Next steps:** Future research is needed to find out how the TSAR transcription factors target the specific set of biosynthetic genes that determine the saponin mixture in a certain tissue or environmental condition.

saponins and the nonhemolytic 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 by different degrees of oxidation at the C-23 position. The aglycones of the nonhemolytic soyasaponins have a hydroxy group at the C-24 position, which appears to exclude oxidation at the C-28 position (Carelli et al., 2011; Tava et al., 2011).

The biosynthetic precursor of *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 primary sterol and 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 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 nonhemolytic soyasaponins (Fukushima et al., 2013). In the hemolytic branch, oleanolic acid is the substrate of two additional P450s, CYP72A67 and CYP72A68, which catalyze hydroxylation of the oleanane backbone at the C-2 position and a 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 such as hederagenin and bayogenin (Figure 1; Fukushima et al., 2013; Biazzi et al., 2015; Tzin et al., 2019). 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 are absent in *M. truncatula* roots (Confalonieri et al., 2009; Pollier et al., 2011; Lei et al., 2019). In the nonhemolytic branch, 24-hydroxy β-amyrin is the substrate of CYP72A61 that catalyzes hydroxylation at the C-22 position, leading to the production of the major soyasaponin aglycone soyasapogenol B (Figure 1; Fukushima et al., 2013). Soyasapogenol B can be further oxidized at the C-21 position toward 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 soybean (*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 various positions, the aglycones are glycosylated with one or more sugar chains. To date, only two UDP-dependent glycosyltransferases have been described in *M. truncatula*. UGT73F3 catalyzes 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 glucosylates aglycones from both the hemolytic and the nonhemolytic branches; however, the position to which the glucosyl residues are transferred has not been 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, 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., 2016a). 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

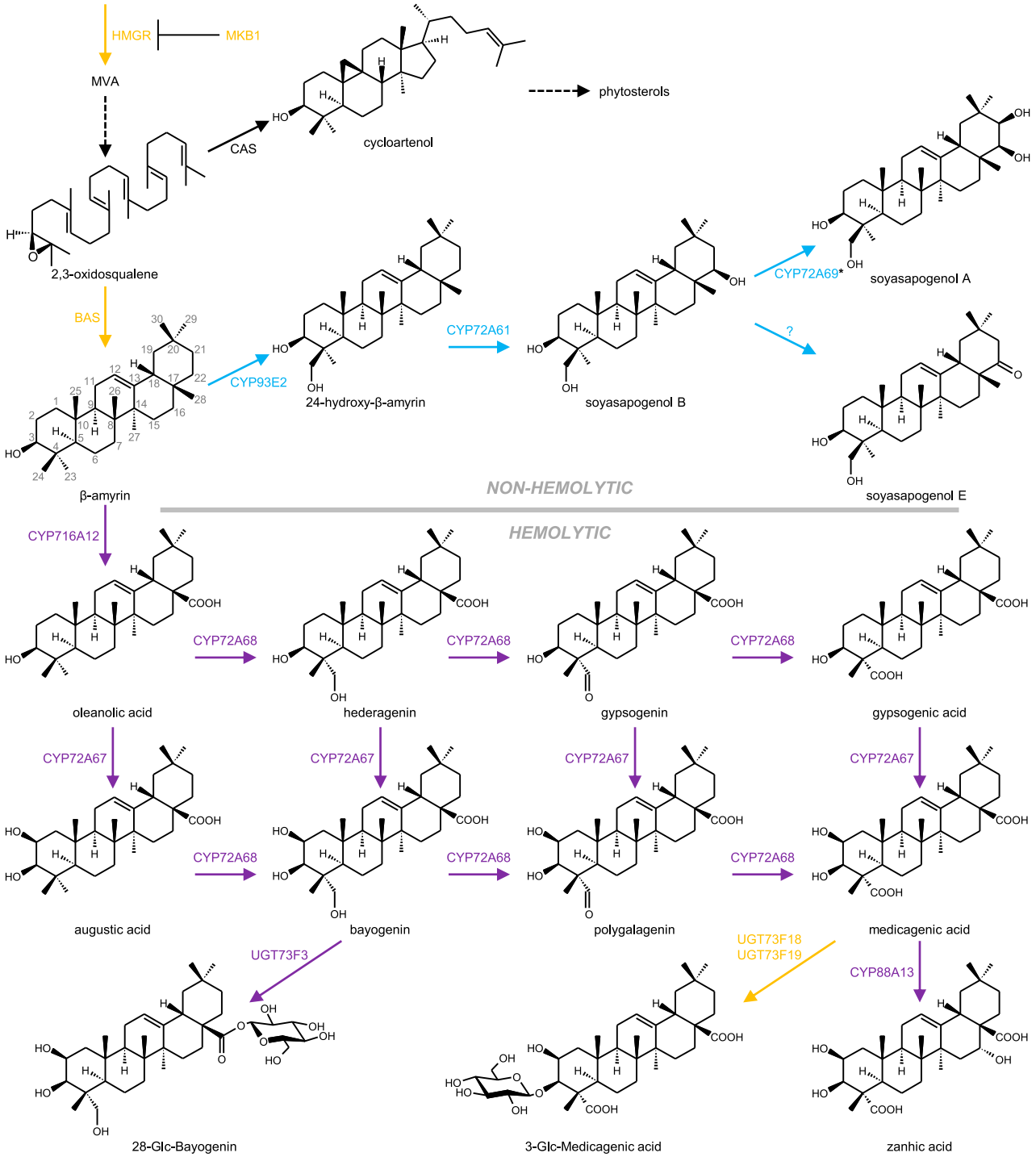



**Figure 1.** The Saponin Biosynthesis Pathway in *M. truncatula*.

Blue, purple, and orange indicate the biosynthetic steps controlled by TSAR1, TSAR2/3, and 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*.

the expression of JA-responsive genes that are either 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 REGULATOR1 (TSAR1) and TSAR2 trigger the concerted transcriptional activation of the nonhemolytic and hemolytic saponin biosynthesis genes, respectively (Mertens et al., 2016a). 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*. Analysis of genes coexpressed with *TSAR3* in transcriptome data sets 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*, i.e., the C-16α hydroxylation of medicagenic acid toward zanhic acid. Furthermore, an extended coexpression analysis revealed two uridine diphosphate glycosyltransferases, UGT73F18 and UGT73F19, which glucosylate hemolytic sapogenins at the C-3 position when expressed in yeast (*Saccharomyces cerevisiae*) and wild tobacco (*Nicotiana benthamiana*).

## RESULTS

### TSAR3 Is a Seed-Specific Regulator of Triterpene Saponin Biosynthesis

The basic helix-loop-helix (bHLH) transcription factors TSAR1 and TSAR2 positively regulate the production of nonhemolytic and hemolytic triterpene saponins, respectively, in the model legume *M. truncatula* (Mertens et al., 2016a). Analysis of the *M. truncatula* Gene Expression Atlas (MtGEA; https://mtgea.noble.org/v3/; He et al. [2009]) revealed that a third paralog 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 the 3-hydroxy-3-methylglutaryl-CoA reductase gene *HMGR1* (Pearson's correlation coefficient, 0.91) and *MAKIBISHI1* (*MKB1*; Pearson's correlation coefficient: 0.87) in developing seeds (Supplemental Figure 1). We previously showed that TSAR1 and TSAR2 were capable of transactivating reporter constructs harboring the 1000-bp promoter regions of *HMGR1* (*ProHMGR1*) and *MKB1* (*ProMKB1*) fused to the *FIREFLY LUCIFERASE* (*fLUC*) gene (Mertens et al., 2016a). Transient expression assays in tobacco (*Nicotiana tabacum*) protoplasts

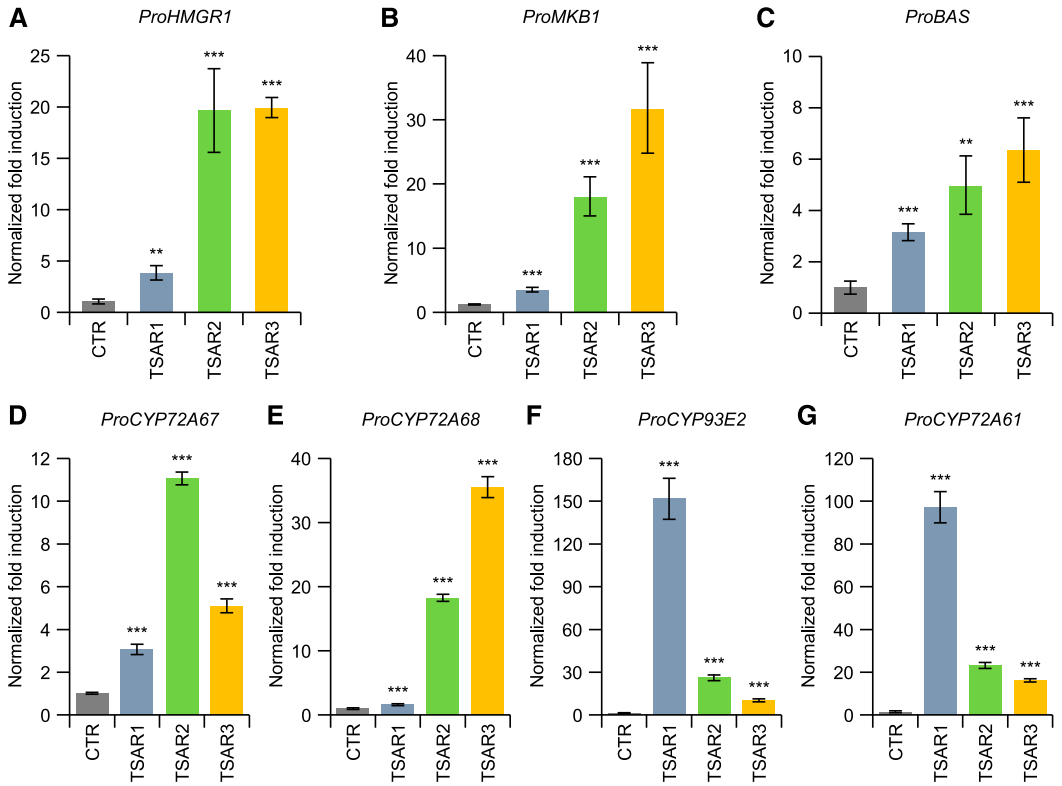



**Figure 2.** Transactivation of Triterpene Saponin Biosynthesis Gene Promoters by TSAR Transcription Factors in Transfected Tobacco Protoplasts.

**(A)** to **(G)** 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 cotransfected with the reporter constructs and a *pCaMV35S:GUS* control plasmid (CTR). The error bars represent the SE ($n = 8$ biological replicates of independently transfected protoplasts). Statistical significance was determined by a two-tailed Student's *t* test (**P < 0.01 and ***P < 0.001).

revealed that TSAR3 induced 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. TSAR3 also induced the luciferase activity of a reporter construct containing the 1000-bp promoter region of *β-AMYRIN SYNTHASE* (*ProBAS*; Figure 2C), suggesting that TSAR3 regulates triterpene saponin biosynthesis in *M. truncatula*.

Because the transcription factors TSAR1 and TSAR2 differentially regulate the nonhemolytic and hemolytic triterpene saponin biosynthesis branches, we performed transient expression assays with TSAR3 and reporter constructs containing the promoter regions of P450 genes from both biosynthetic branches. Like TSAR2, TSAR3 induced the luciferase activity of reporter constructs containing 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 also induced the luciferase activity of reporter constructs containing the promoters of *CYP93E2* (*ProCYP93E2*) and *CYP72A61* (*ProCYP72A61*; Figures 2F and 2G), corresponding to P450s involved in the biosynthesis of nonhemolytic triterpene saponins (Fukushima et al., 2011, 2013). However, the activation of these promoters by TSAR3 was markedly lower compared to that by TSAR1, the known regulator of the nonhemolytic saponin biosynthesis branch, indicating that the nonhemolytic branch is likely not the preferred target of TSAR3, as was also the case for TSAR2 (Mertens et al., 2016a). 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, we generated three independent *M. truncatula* hairy root lines ectopically overexpressing *TSAR3* (TSAR3$^{OE}$). We used three independent *M. truncatula* hairy root lines expressing *β-glucuronidase* (*GUS*) as a control. According to RNA sequencing data previously generated by Mertens et al. (2016a), the seed-specific *TSAR3* is not expressed in control *M. truncatula* hairy roots. We verified the expression of *TSAR3* in the TSAR3$^{OE}$ lines by RT-PCR (Figure 3A), which confirmed that *TSAR3* was ectopically expressed in the TSAR3$^{OE}$ lines but not in the control hairy root lines.

RT-qPCR analysis of triterpene saponin biosynthesis genes revealed increased expression of *HMGR1* in the TSAR3$^{OE}$ lines (Figure 3B), whereas no effect on *MKB1* expression was observed (Figure 3C). Furthermore, we detected significantly increased transcript levels of the triterpene saponin biosynthesis genes *BAS*, *CYP716A12*, *CYP72A67*, *CYP72A68*, and *UGT73F3* in the TSAR3$^{OE}$ lines compared to the control (Figures 3D to 3H). By contrast, the transcript levels of *CYP93E2*, *CYP72A61*, and *UGT73K1* decreased in the TSAR3$^{OE}$ lines versus the control (Figures 3I to 3K). Overall, this pattern, which is reminiscent of that in TSAR2$^{OE}$ lines (Mertens et al., 2016a), indicates that the hemolytic triterpene saponin biosynthesis branch genes are transcriptionally upregulated in response to the overexpression of *TSAR3*.

To explore the effect of *TSAR3* expression at the metabolite level, we performed untargeted metabolite profiling of the three independent TSAR3$^{OE}$ and GUS control lines by liquid chromatography-mass spectrometry (LC-MS). Peak integration and alignment yielded a total of 17,922 *m/z* features, which were log-transformed and pareto-scaled for principal component analysis (PCA). This PCA showed that the expression of *TSAR3* explained most of the variation (44.2%) in the data set, as the first principal component 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 data set, 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. We identified the metabolites corresponding to these differential *m/z* features based on their accurate mass, $MS^n$, and MS/MS fragmentation spectra and available standards. This analysis revealed that the ectopic expression of *TSAR3* led to the accumulation of hemolytic triterpene saponins in *M. truncatula* hairy roots, whereas the abundance of the major soyasaponins remained unaltered (Figures 4B to 4G; Supplemental Data Set 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 accumulated in the TSAR3$^{OE}$ lines. Zanhic acid is a major aglycone found in the aerial parts of *M. truncatula* plants (Kapusta et al., 2005a, 2005b) that is absent from *M. truncatula* roots and hairy roots (Confalonieri et al., 2009; Pollier et al., 2011; Lei et al., 2019). In addition, several of the major medicagenic acid glycosides normally present in the aerial parts of the plant, but not in roots, were detected 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, which 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 the 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 database (https://medicago-mutant.noble.org/mutant/index.php), three candidate mutant lines were selected: NF14212, NF15672, and NF17107. We obtained homozygous insertion mutants 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, we obtained homozygous escape lines in which both *TSAR3* alleles were 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 wild-type plants (Supplemental Figure 4).

RT-qPCR analysis with primers spanning the *Tnt1* retrotransposon revealed that the transcript levels of *TSAR3* were drastically reduced or even undetectable in developing seeds of the insertion lines (Figures 5B and 5C). Furthermore, we detected a drastic decrease in transcript levels of *HMGR1*, *BAS*, and the hemolytic triterpene saponin biosynthesis genes *CYP716A12*, *CYP72A67*, *CYP72A68*, and *UGT73F3*. The transcript levels of *CYP93E2*, *CYP72A61*, and *UGT73K1*, which are involved in soyasaponin biosynthesis, appeared to be slightly higher in these lines than the wild type, whereas the expression levels of *MKB1* were barely 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 point to the specific transcriptional downregulation of the hemolytic triterpene saponin biosynthesis branch in the seeds of *tsar3* plants, supporting the role of TSAR3 as a seed-specific regulator of hemolytic triterpene saponin biosynthesis.

We performed untargeted metabolite profiling of developing seeds and leaves of wild-type and homozygous *tsar3* plants by LC-MS, which yielded a total of 15,819 *m/z* features. A PCA 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), pointing to 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 nonhemolytic saponins were observed in the leaves of these plants (Figures 6C to 6H). To identify the metabolites that contribute most to the difference between the

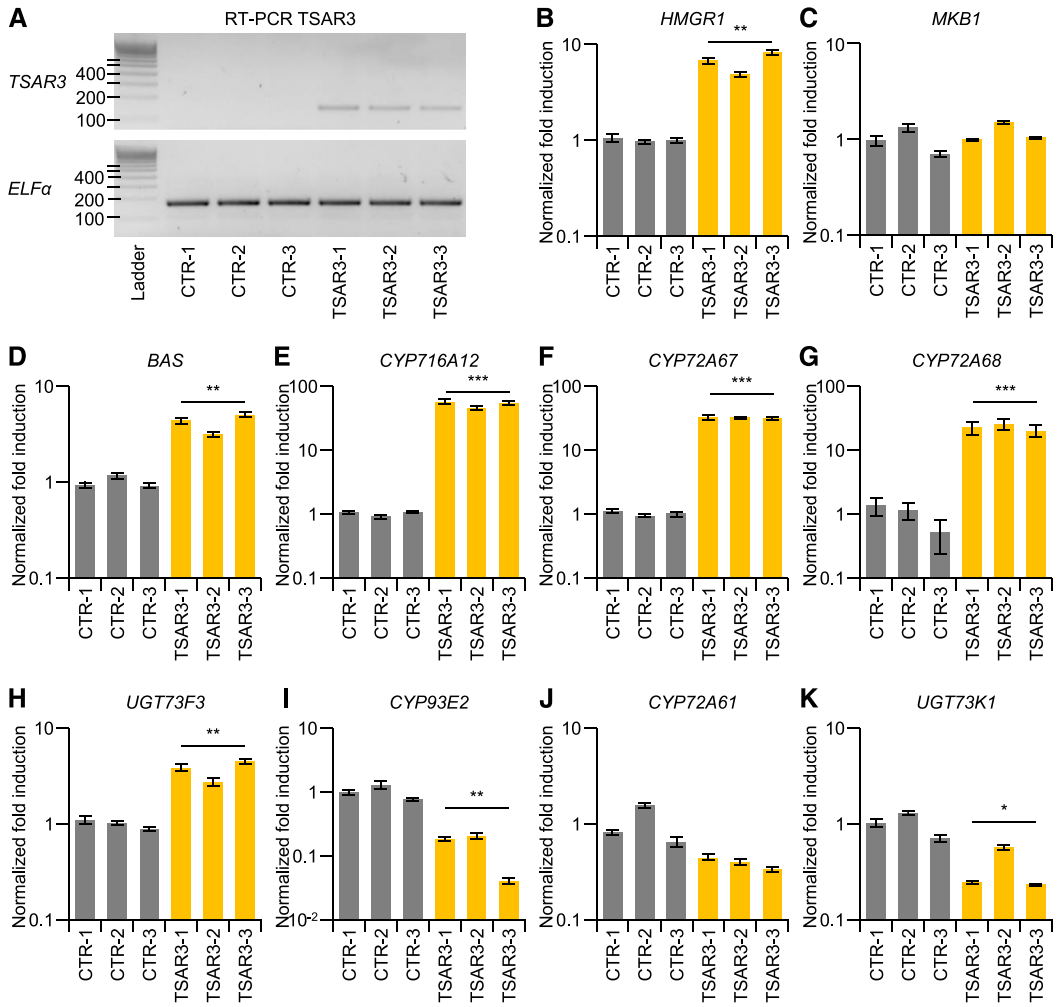



**Figure 3.** Ectopic Expression of *TSAR3* in *M. truncatula* Hairy Roots Affects the Expression of Triterpene Saponin Biosynthesis Genes.

**(A)** RT-PCR analysis showing *TSAR3* expression in three independent transgenic TSAR3$^{OE}$ hairy root lines compared to three control (CTR) lines. The RT-qPCR reference gene *ELFα* was used as a control. The Eurogentec SmartLadder SF was used as a size marker (Ladder).

**(B)** to **(K)** RT-qPCR 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 represent the SE ($n = 3$ technical replicates). Statistical significance was determined by a two-tailed Student's *t* test (*P < 0.05, **P < 0.01, and ***P < 0.001).

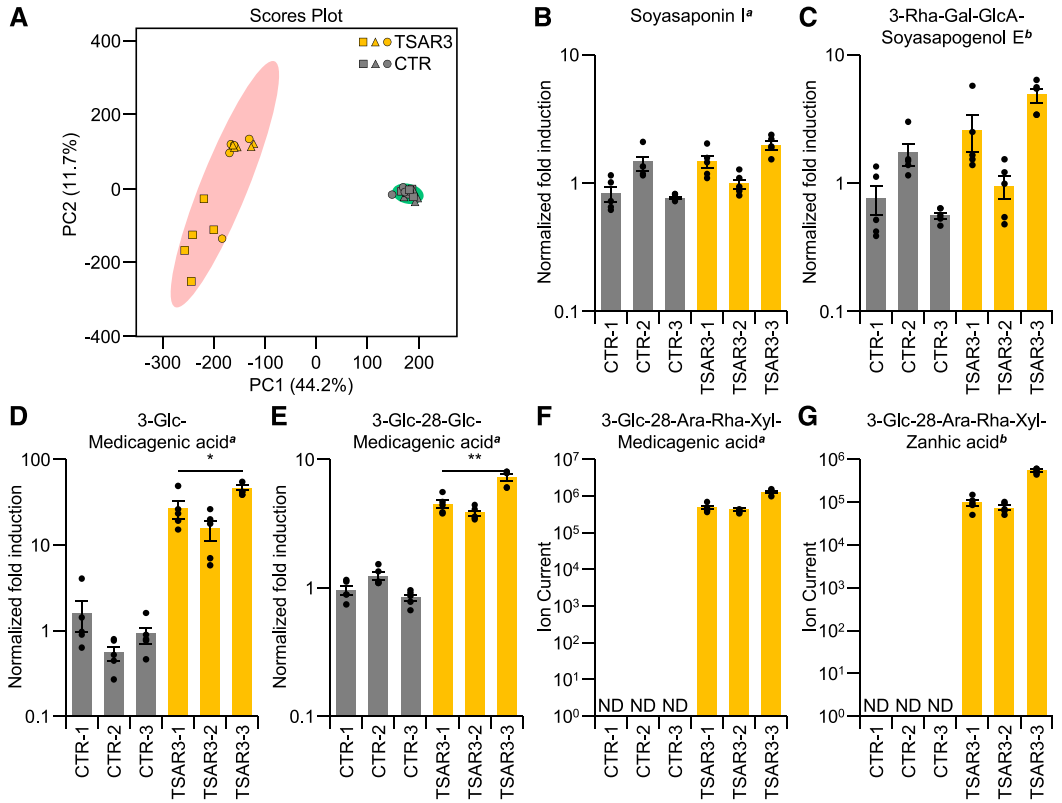



**Figure 4.** Ectopic Expression of *TSAR3* Leads to Enhanced Accumulation of Hemolytic Triterpene Saponins in *M. truncatula* Hairy Roots.

**(A)** Principal component analysis of samples from TSAR3$^{OE}$ and control (CTR) hairy roots. Different symbols indicate the different transgenic lines.
**(B)** to **(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)** and **(G)** Average total ion current of the peaks corresponding to two hemolytic triterpene saponins. The mean and SE ($n = 5$ technical replicates in which the roots were grown and analyzed independently) 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 Data Set 1) according to Pollier et al. (2011).

seeds of *tsar3* and wild-type plants, we performed a partial least squares discriminant analysis (PLS-DA) of the samples derived from developing seeds (Figure 6I). The 50 *m/z* features that contribute most to the separation of *tsar3* and wild-type plants are listed in Supplemental Data Set 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 absent in developing *tsar3* seeds. Together, these data confirm the notion that TSAR3 plays a specific role in regulating 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 led 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, we performed two coexpression analyses using microarray data sets from developing *M. truncatula* seeds. First, we calculated the Pearson's correlation coefficient between the probe set corresponding to *TSAR3* and all 19,012 differentially expressed probe sets during *M. truncatula* seed development (Verdier et al., 2013). Within the 88 probe sets with a Pearson's correlation coefficient above 0.95, probe sets corresponding to all characterized P450s of the hemolytic triterpene saponin branch were present (Supplemental Data Set 3). Remarkably, a probe set 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 also present in this data set. Furthermore, probe sets corresponding to more upstream triterpene saponin biosynthesis genes, such as *BAS* and *MEVALONATE KINASE*, were also present in this set, implying that the hemolytic triterpene saponin biosynthesis genes and *TSAR3* are tightly coregulated in developing seeds. Within these highly coregulated probe sets, we identified several probe sets corresponding to P450s that might be the missing C-16α hydroxylase.

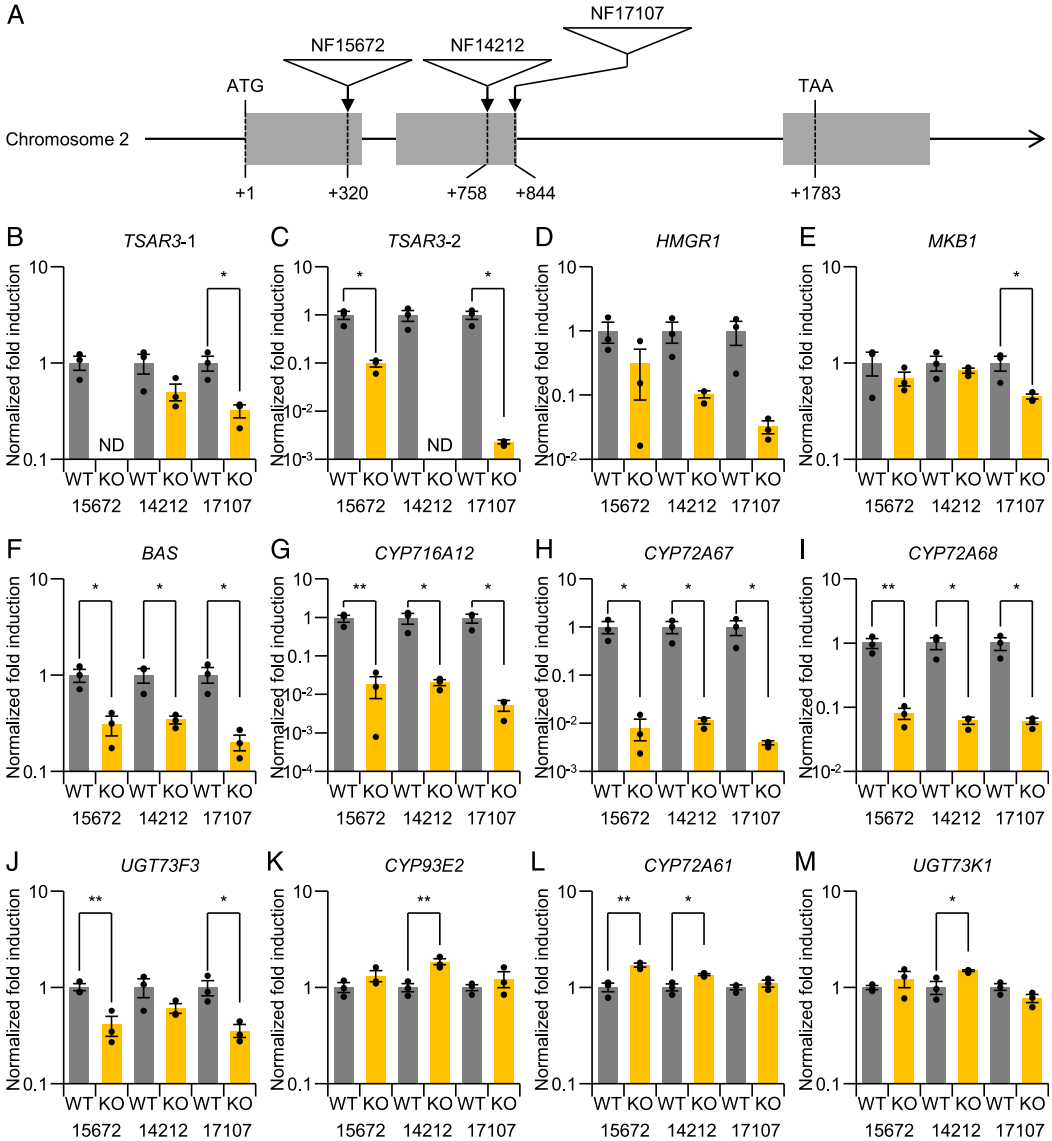



**Figure 5.** Hemolytic Saponin Biosynthesis Is Affected in the Seeds of *tsar3* Plants.

**(A)** Schematic representation of *TSAR3*. The locations of the *Tnt1* retrotransposon in the insertion lines are indicated by triangles. Exons are represented by gray boxes.

**(B)** to **(M)** RT-qPCR analysis of triterpene saponin biosynthesis genes in the 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 (wild-type [WT]) plants. The mean and SE ($n$ = 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 and **P < 0.01). TSAR3-1 primers span the first *Tnt1* insertion (NF15672); TSAR3-2 primers span the *Tnt1* insertion in lines NF14212 and NF17107. ND, Not detected.

We performed a second coexpression analysis of a data set consisting of 60 transcriptomes from developing seeds that were produced under different environmental conditions (Supplemental Data Set 4; Righetti et al., 2015). We also calculated the Pearson's correlation coefficient between the probe set corresponding to *TSAR3* and all other probe sets in this data set. Within the 121 probe sets with a Pearson's correlation coefficient above 0.85, probe sets corresponding to *CYP716A12* and *HMGR1* were present, but probe sets corresponding to *CYP72A67* and *CYP72A68* were absent, implying less strict coregulation of the hemolytic triterpene saponin biosynthesis genes within this data set. In both data sets, *Medtr5g014240* was the P450 gene that had the highest coexpression level with *TSAR3*. We therefore considered this gene to be our primary candidate C-16α hydroxylase gene. In addition to *Medtr5g014240*, 12 other P450s were present in one or both data sets. As several of these candidates corresponded to incomplete genes in the genome or genes that were not expressed in *M.*

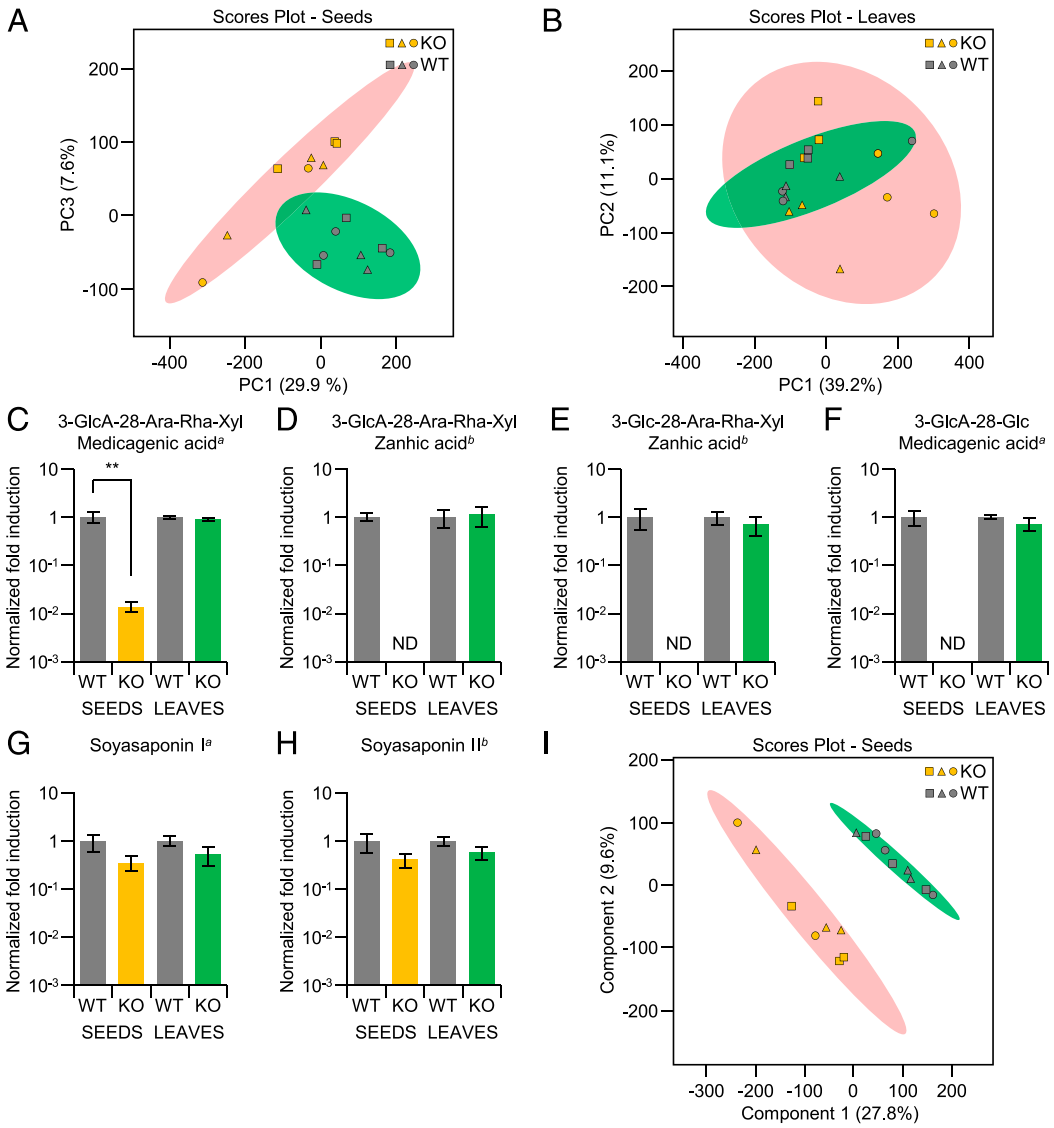



**Figure 6.** Hemolytic Triterpene Saponin Accumulation Is Reduced in Developing Seeds of *tsar3 Tnt1* Insertion Mutants.

**(A)** and **(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)** to **(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 SE ($n = 8$ 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 Data Set 2).

**(I)** PLS-DA of the samples from developing seeds of *tsar3* (KO) and wild-type (WT) plants. Different symbols indicate the different transgenic lines.

*truncatula* shoots, which are known to accumulate zanhic acid glycosides (Kapusta et al., 2005a, 2005b), the list of candidate genes was further reduced. Finally, of the remaining candidate genes, 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 *N. benthamiana*

In *Medicago* species, only two compounds with 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). Despite its occurrence in *Medicago polymorpha*, caulophyllogenin has not been reported as a saponin aglycone in *M. truncatula* (Tava et al., 2011). Hence, we assumed that medicagenic acid is the substrate of the unknown C-16α hydroxylase in *M. truncatula*. To screen for a functional C-16α hydroxylase in yeast, we engineered a yeast strain that produces sufficiently high levels of medicagenic acid. This was achieved by coexpressing the genes encoding β-amyrin synthase from licorice (*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 1). JP034 is a BY4742-derived yeast strain in which an additional auxotrophic marker (TRP1) and a disruption of the phosphatidic acid phosphatase-encoding gene *PAH1* were 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 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 *ERG7* to be downregulated by the addition of methionine to the cultivation medium. This downregulation of *ERG7* leads to higher β-amyrin production in yeast, presumably by increasing the availability of 2,3-oxidosqualene for the expressed BAS (Kirby et al., 2008; Moses et al., 2014).

To identify the functional C-16α hydroxylase, we expressed the candidate P450s in the medicagenic acid-producing yeast strain. Compared to the empty vector control, the gas chromatography-mass spectrometry (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, we analyzed 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), by GC-MS. A peak with the same retention time and electron ionization (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 hydroxy group, indicating that CYP88A13, unlike CYP716Y1, only recognizes the highly oxidized medicagenic acid as a substrate.

In addition to yeast, we also expressed the candidate P450s in combination with *GgbAS*, *CYP72A67*, *CYP72A68*, and *CYP716A12* in wild tobacco (*N. benthamiana*) leaves by agroinfiltration. We analyzed methanolic extracts of leaves harvested 3 d after agroinfiltration by GC-MS. Again, compared to the empty vector control and to 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 gene *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

Despite 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, we investigated the capacity of CYP88A13 to oxidize β-amyrin. Upon expression of *CYP88A13* in a β-amyrin-producing yeast strain, no 16α-hydroxy β-amyrin was produced, whereas the 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 the expression of *CYP716Y1*, the 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, we expressed *CYP88A13* in a yeast strain that accumulates gypsogenic acid. Like the positive control *CYP716Y1*, the expression of *CYP88A13* in this strain led to the production of 16α-hydroxy gypsogenic acid (Figure 8D; Supplemental Figures 7 and 10), suggesting that gypsogenic acid serves as a substrate for CYP88A13. The presence of 16α-hydroxy gypsogenic acid was observed in the leaves of a *M. truncatula* mutant lacking CYP72A67 activity (Biazzi et al., 2015), further supporting the notion that gypsogenic acid indeed serves as a substrate of 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 genes encoding enzymes leading to the production of various intermediates (Supplemental Figure 11).

Taken together, these data point toward the following order of oxidation of the β-amyrin backbone in the hemolytic saponin biosynthesis pathway 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 the formation of zanhic acid (Figure 1).

### The Expression of *CYP88A13* Is Regulated by TSAR3

As *CYP88A13* is strictly coregulated with *TSAR3* in developing *M. truncatula* seeds (Supplemental Data Sets 3 and 4), we investigated whether TSAR3 regulates the expression of *CYP88A13*. First, RT-PCR analysis revealed *CYP88A13*

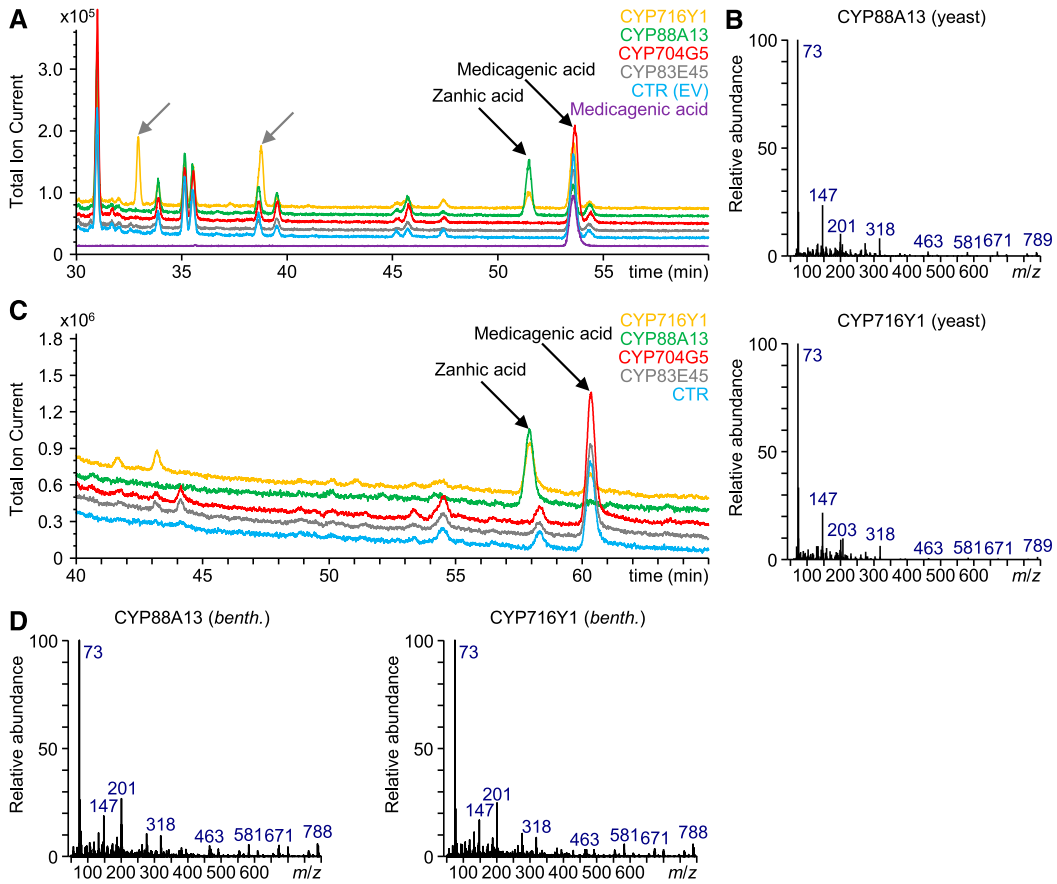



**Figure 7.** CYP88A13 Catalyzes the C-16α Hydroxylation of Medicagenic Acid in Yeast and *N. benthamiana*.

**(A)** Overlay of GC-MS chromatograms of a medicagenic acid-producing yeast strain expressing the candidate P450 genes, *CYP716Y1*, or an empty vector (EV) control (CTR). Zanhic acid accumulates only in the strain expressing the characterized C-16α hydroxylase gene *CYP716Y1* and *CYP88A13*. In addition to zanhic acid, several other C-16α hydroxylated triterpenoids (indicated by 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 gene *CYP88A13* (top) or *CYP716Y1* (bottom).

**(C)** Overlay of GC-qTOF-MS chromatograms of medicagenic acid-producing *N. benthamiana* leaves infiltrated with the candidate P450 genes or *CYP716Y1*. Zanhic acid accumulates only in leaves infiltrated with the characterized C-16α hydroxylase gene *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).

expression in the TSAR3$^{OE}$ lines, but not in the GUS control lines (Figure 9A), implying that, like the other hemolytic P450s, TSAR3 can control the expression of *CYP88A13* in *M. truncatula* hairy roots. We then performed transactivation assays in tobacco protoplasts with different *CYP88A13* promoter (*ProCYP88A13*) constructs. In the first assay, TSAR3 failed to induce the luciferase activity of a reporter construct containing the 1000-bp region upstream of the start codon of *CYP88A13* (Figure 9B). A more detailed analysis of *ProCYP88A13* revealed an N-box (5′-CACGAG-3′) located 2051 bp upstream of the *CYP88A13* start codon; this promoter element is required for TSAR binding (Mertens et al., 2016a). Three additional N-box-like sequences are located within 50 bp upstream of the first identified N-box, including an E-box (5′-CACGCG-3′) located 2096 bp upstream of the *CYP88A13* start codon. Hence, we cloned a 2118-bp *ProCYP88A13* fragment and used it for transient expression assays with TSAR3, revealing threefold induced luciferase activity compared to protoplasts cotransfected with the GUS control (Figure 9C). Furthermore, in line with its role in hemolytic saponin biosynthesis, *ProCYP88A13* was also transactivated by TSAR2. TSAR1, a regulator of soyasaponin biosynthesis, did not transactivate *ProCYP88A13* (Figure 9C). Finally, the expression of *CYP88A13* was considerably reduced in developing seeds of all three *tsar3 Tnt1* insertion lines (Figure 9D), implying that TSAR3 controls *CYP88A13* transcription in developing *M. truncatula* seeds. No effect on *CYP88A13* expression was observed in the leaves of these insertion lines (Figure 9E), which is 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 the expression of *CYP88A13* in developing *M. truncatula* seeds.

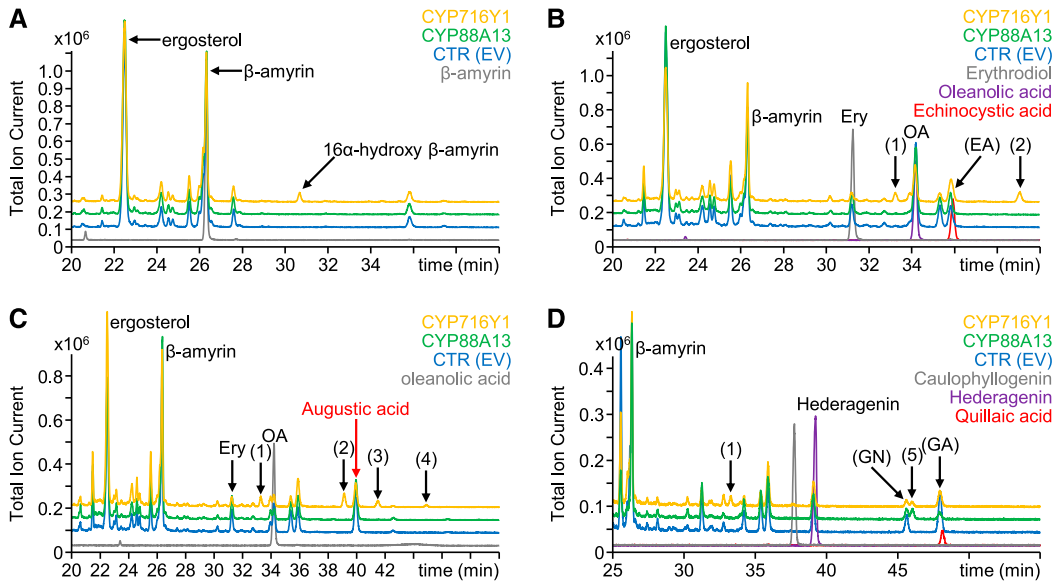



**Figure 8.** Investigation of the Substrate Tolerance of CYP88A13 in Yeast.

**(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.

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

As CYP88A13 catalyzed the C-16α hydroxylation of medicagenic acid in yeast and *N. benthamiana*, we reasoned that the 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$^{OE}$). Three *M. truncatula* hairy root lines expressing *GUS* were used as a control. RT-PCR analysis confirmed the ectopic expression of *CYP88A13* in CYP88A13$^{OE}$, but not in the GUS control hairy root lines (Figure 10A). In addition, metabolite profiling by LC-MS confirmed that zanhic acid glycosides such as 3-Glc-zanhic acid accumulated in the CYP88A13$^{OE}$ lines (Figure 10B).

To further validate the role of CYP88A13 in planta, we screened a *M. truncatula Tnt1* retrotransposon insertion population (Tadege et al., 2008) for *CYP88A13* loss-of-function mutants. As done for TSAR3, three candidate mutant lines were selected from the *Tnt1* flanking sequence tag 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). We investigated the leaves of homozygous plants of both *Tnt1* insertion lines by LC-MS, revealing that both lines still accumulated zanhic acid glycosides (Figure 10D). RT-PCR analysis with primers spanning the *Tnt1* insertion revealed that the *Tnt1*-containing introns were spliced out, which did not affect *CYP88A13* expression (Figure 10E).

Inspection of the *M. truncatula* genome v4.0 revealed that *CYP88A13* forms part of a tandem repeat with *CYP88A14* (*Medtr5g014250*) and a third, seemingly incomplete pseudogene, *Medtr5g014230*, resulting from a 1628-bp genomic repeat that overlaps with *CYP88A13* (Figure 10F). Since CYP88A13 and CYP88A14 share 77% amino acid identity, we speculated that also *CYP88A14* might 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, the expression of *CYP88A14* failed to lead to the production of zanhic acid, instead leading to the

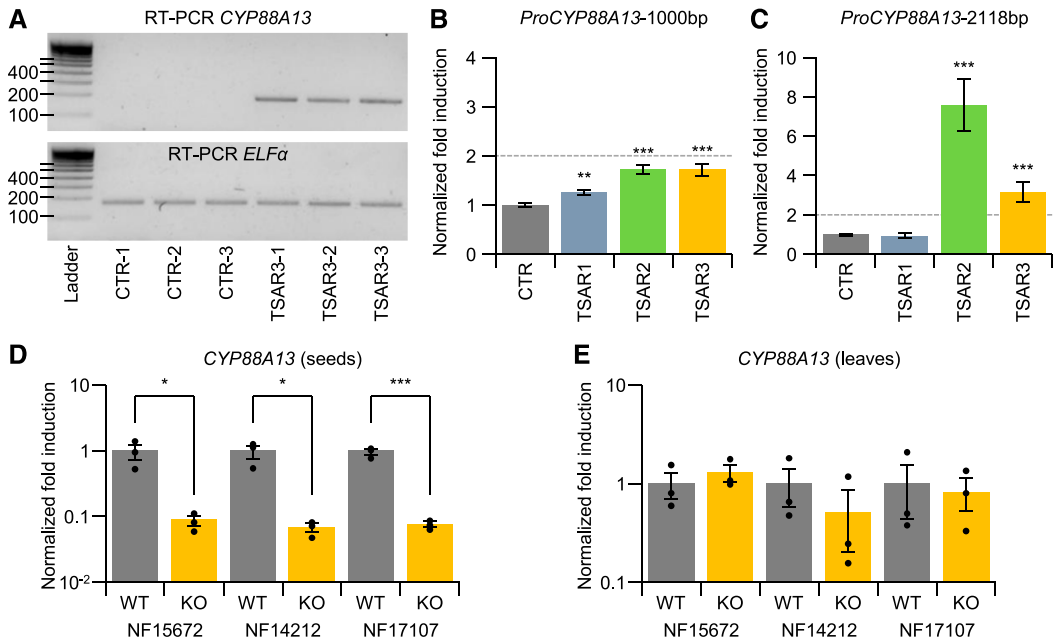



**Figure 9.** The Expression of *CYP88A13* Is Regulated by TSAR3.

**(A)** RT-PCR analysis showing that *CYP88A13* is expressed in three independent transgenic TSAR3$^{OE}$ hairy root lines, but not in the GUS control (CTR) lines (top). The RT-qPCR reference gene *ELFα* was used as a control (bottom).

**(B)** and **(C)** Transactivation of the 1000-bp **(B)** and 2118-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 cotransfected with the reporter constructs and a *pCaMV35S:GUS* control plasmid (CTR). The gray dashed line indicates twofold induction compared to the control and serves as the cut-off. The error bars represent the SE ($n = 8$). Statistical significance was determined by a two-tailed Student's *t* test (*$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$).

**(D)** and **(E)** RT-qPCR analysis of *CYP88A13* in developing seeds **(D)** or leaves **(E)** of the *tsar3 Tnt1* lines. For each *Tnt1* insertion line, three individual homozygous escape (wild-type [WT]) and homozygous knockout (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 SE ($n = 3$ 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 ***$P < 0.001$).

production of two unknown metabolites (Supplemental Figure 13). These results indicate that *CYP88A14* encodes a functional protein that is not involved in zanhic acid biosynthesis.

### Selection of Candidate UGTs Involved in Hemolytic Triterpene Saponin Biosynthesis

In addition to the production of zanhic acid glycosides, the 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 coexpressed with *TSAR3* in the two coexpression analyses that led to the identification of CYP88A13 (Supplemental Data Sets 3 and 4) were considered to be candidate UGT genes involved in hemolytic saponin biosynthesis. This list was expanded with UGT candidates from a third coexpression analysis in which the *M. truncatula* gene expression atlas (He et al., 2009) was probed for UGTs coexpressed with *CYP716A12*. UGT-encoding genes with a Pearson's correlation coefficient above 0.6 were considered to be candidate UGT genes involved in hemolytic saponin biosynthesis. Combined, these three coexpression analyses led to the generation of a comprehensive set of 20 candidate UGTs, 10 of which we were able to clone for further analysis (Supplemental Table 2).

### Functional Characterization of Candidate UGTs in Yeast and *N. benthamiana*

To identify functional UGTs using *S. cerevisiae*, we expressed the UGT candidates in the yeast strain engineered for medicagenic acid accumulation and used for the identification of CYP88A13. In addition 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 cell extracts expressing the different UGTs revealed the production of glycosides in the strains expressing two of the candidate UGTs and in the strains expressing the positive controls *UGT73C11* and *UGT73F3* (Figures 11A and 11B).

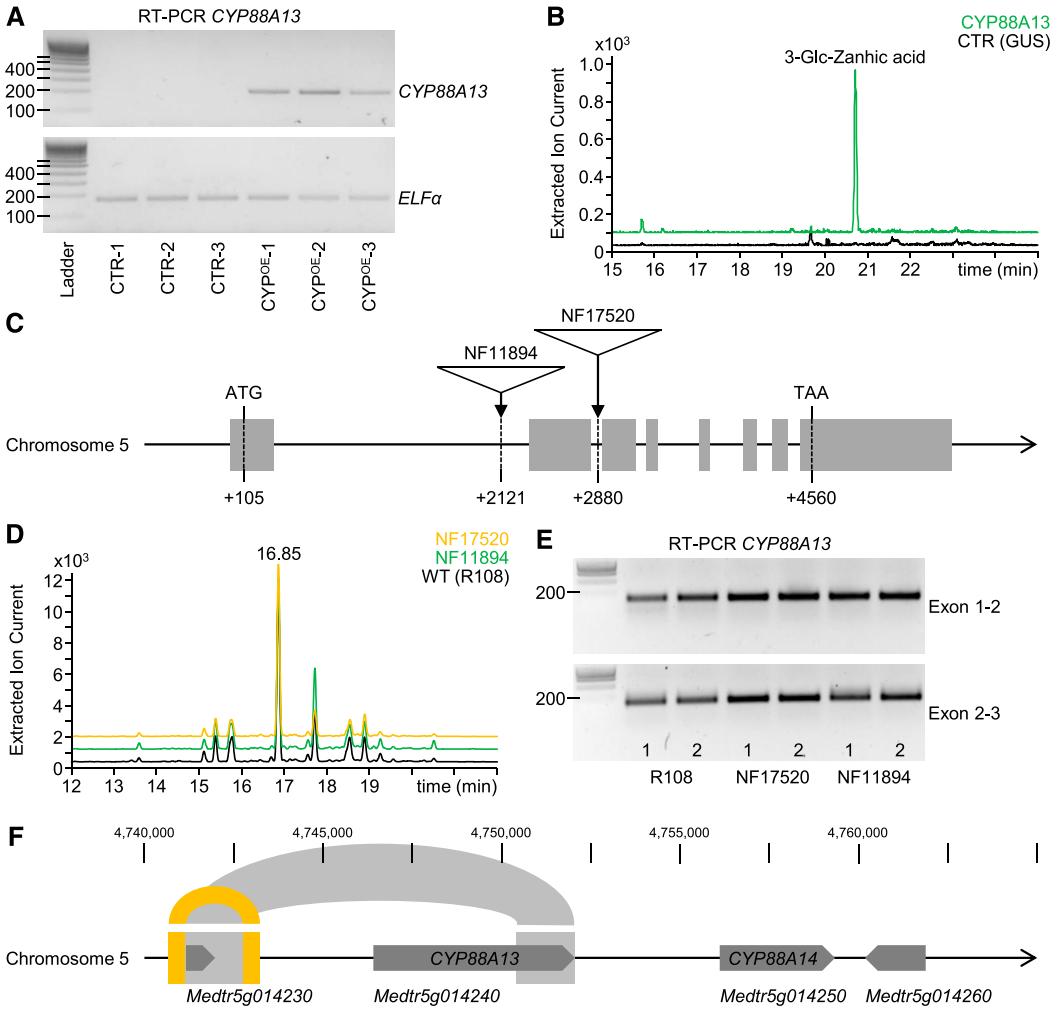



**Figure 10.** Functional Characterization of *CYP88A13* in *M. truncatula*.

**(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 RT-qPCR 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 by 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 by orange (461 bp) and by light gray (1628 bp) boxes, respectively. Genes are indicated by dark gray arrows.

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 gene *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),

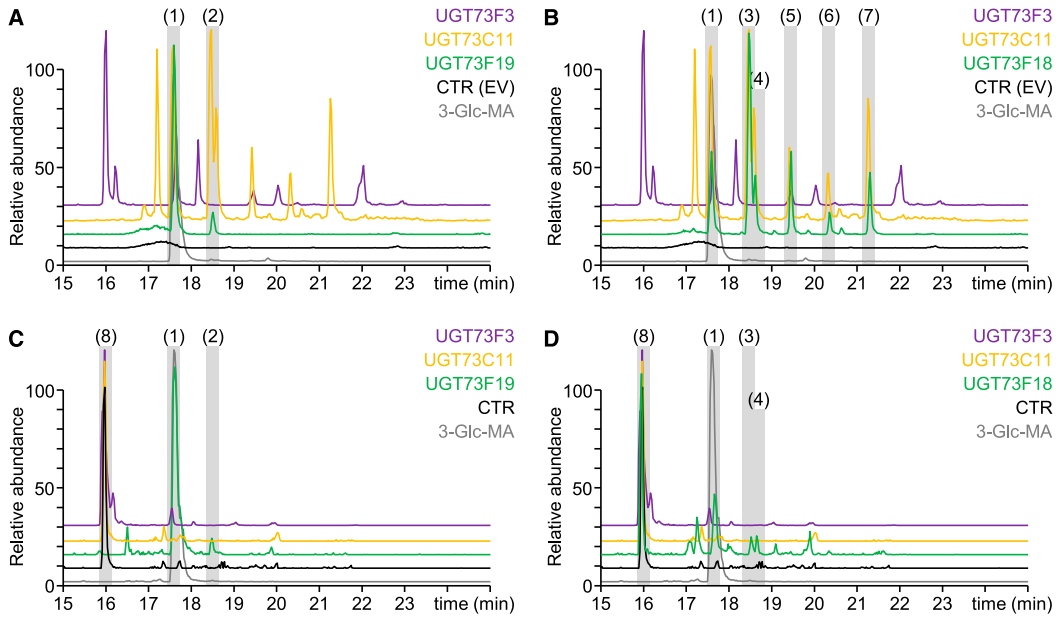



**Figure 11.** Functional Characterization of Candidate UGTs in Yeast and *N. benthamiana*.

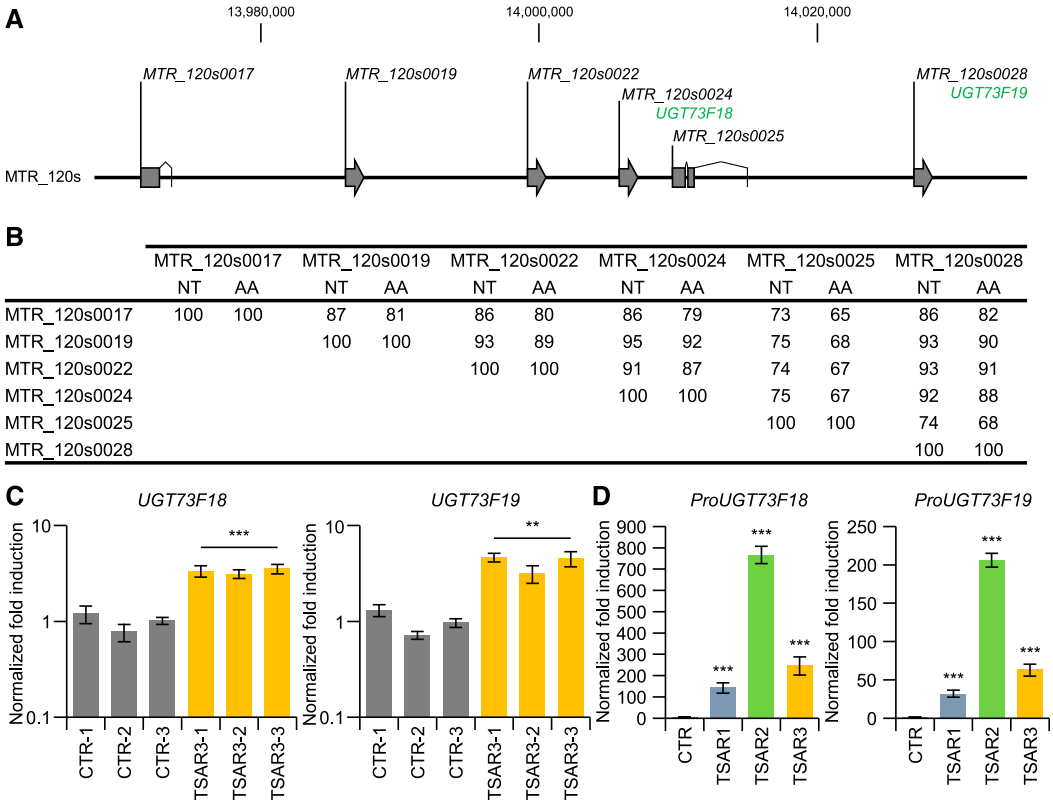
**(A)** Overlay of LC-MS chromatograms of a 3-Glc-medicagenic acid (3-Glc-MA) standard and extracts of a medicagenic-acid-producing yeast strain expressing the candidate glycosyltransferase gene *UGT73F19*, the positive controls *UGT73C11* and *UGT73F3*, or an empty vector (EV) control (CTR).
**(B)** 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 gene *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 coinfiltrated with *UGT73F19*, the positive controls *UGT73C11* and *UGT73F3*, or a leaf not coinfiltrated with a UGT gene.
**(D)** Overlay of LC-MS chromatograms of a 3-Glc-medicagenic acid standard and extracts of *N. benthamiana* leaves coinfiltrated with *UGT73F18*, the positive controls *UGT73C11* and *UGT73F3*, or a control leaf not coinfiltrated with a UGT gene. (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.

polygalagenin (4), gypsogenin (5), augustic acid (6), and oleanolic acid (7). In addition, small amounts 3-Glc-bayogenin and 3-Glc-gypsogenic acid comigrated 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, we also expressed the candidate UGTs in combination with *GgbAS*, *CYP716A12*, *CYP72A67*, and *CYP72A68* in *N. benthamiana* leaves by agroinfiltration. We analyzed methanolic extracts of leaves harvested 4 d after agroinfiltration 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 in these leaves. The accumulation of C-28 glucosides was also observed when no candidate UGT gene was infiltrated and they were also observed in a previous study in which *UGT73C11* was infiltrated into *N. benthamiana* leaves (Khakimov et al., 2015). Hence, we postulated that *N. benthamiana* contains a UGT that catalyzes the C-28 glycosylation of triterpenoid aglycones that competes for the triterpenoid aglycone substrates with the expressed UGTs. Upon expression of *UGT73F19*, almost no 28-Glc-medicagenic acid was detected (Figure 11C), suggesting that UGT73F19 has a higher affinity 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 that this glucosyltransferase has a slightly lower affinity for the sapogenin substrates compared to UGT73F19. Taken together, these results 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 compared to UGT73F19.

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

The glucosyltransferase genes *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 in 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 is unfortunately no longer present in the more recent Mt4.0 genome release (Tang et al., 2014).




| | 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)** RT-qPCR 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 represent the SE ($n = 3$ 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 cotransfected with the reporter constructs and a *pCaMV35S:GUS* control plasmid (CTR). The error bars represent the SE ($n = 8$). Statistical significance was determined by a two-tailed Student's *t* test (***P < 0.001).

The duplicated UGTs show a high sequence similarity of 92% at the nucleotide level and 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 coregulation between the two characterized UGTs and *TSAR3* is less pronounced. RT-qPCR analysis of *UGT73F18* and *UGT73F19* in the TSAR3$^{OE}$ lines revealed that these genes also appear to be regulated by TSAR3. However, their transcriptional upregulation was only approximately fourfold (Figure 12C), which is comparable to the upregulation of the C-28 glucosyltransferase gene *UGT73F3* but less pronounced than the upregulation of P450 genes involved in hemolytic saponin biosynthesis (Figure 2). Transactivation assays in tobacco protoplasts, however, revealed the substantial activation of the *UGT73F18* (*ProUGT73F18*) and *UGT73F19* (*ProUGT73F19*) promoters by all three TSARs (Figure 12D).

## DISCUSSION

To protect themselves from pathogen attack or herbivore predation, *M. truncatula* plants accumulate organ-specific blends of triterpene saponins whose 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 coexpression analyses with *TSAR3* in transcriptome data sets 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*, i.e., the C-16α hydroxylation of medicagenic acid toward zanhic acid. Furthermore, the coexpression 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 biosynthesis genes form a TSAR3-regulated regulon in developing *M. truncatula* seeds.

### Recruitment of Clade IVa bHLH Transcription Factors to Regulate Metabolic Pathways

TSAR3 is a clade IVa bHLH transcription factor. These often JA-responsive transcription factors are involved in regulating diverse metabolic pathways in different plant species (Goossens et al., 2017; Shoji, 2019). Like TSAR1 and TSAR2 (Mertens et al., 2016a), we demonstrated that TSAR3 is also involved in regulating triterpene saponin biosynthesis in *M. truncatula*. Like TSAR1, the JA-responsive TSAR ortholog GubHLH3 regulates soyasaponin biosynthesis in Chinese licorice (*Glycyrrhiza uralensis*; Tamura et al., 2018), implying that TSAR transcription factors play conserved roles in legumes. Similarly, the TSAR-Like (TSARL) transcription factors TSARL1 and TSARL2 regulate saponin biosynthesis in quinoa (*Chenopodium quinoa*), with TSARL1, like TSAR3, being exclusively expressed in the seeds where it controls the biosynthesis of hemolytic, bitter-tasting saponins (Jarvis et al., 2017). Quinoa, a member of the amaranth family, is not related to the legumes *M. truncatula* and *G. uralensis*, pointing to the possible conservation of the role of clade IVa bHLH transcription factors in regulating saponin biosynthesis beyond the legume family. However, not all dicots accumulate saponins, and in Madagascar periwinkle (*Catharanthus roseus*), the JA-responsive clade IVa bHLH transcription factors BIS1 and BIS2 regulate monoterpenoid indole alkaloid biosynthesis (Van Moerkercke et al., 2015; Mertens et al., 2016b; Van Moerkercke et al., 2016). It thus seems more likely that the transcriptional control of saponin biosynthesis by TSAR or TSARL transcription factors in *M. truncatula* and quinoa represents a case of convergent evolution. By gaining cis-regulatory elements in the promoters of 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.

### TSAR Transcription Factors Determine the Tissue-Specific Blend of Saponins

TSAR3 is the third clade IVa bHLH transcription factor shown to be involved in regulating triterpene saponin biosynthesis in *M. truncatula*. Upon inspection of the *M. truncatula* genome (Tang et al., 2014), we identified a staggering number (30) of 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 might also be capable of inducing (a specific branch of) the triterpene saponin biosynthesis pathway in *M. truncatula* during (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 nonhemolytic saponin biosynthesis branches. Like TSAR2, TSAR3 controls hemolytic triterpene saponin biosynthesis, yet based on the phylogeny of its characteristic bHLH domain (Supplemental Figure 14), TSAR3 appears to be more closely related to TSAR1, a regulator of non-hemolytic soyasaponin biosynthesis (Mertens et al., 2016a). Thus, it appears that the DNA-binding bHLH domain itself is 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 homo- and heterodimer formation (Van Moerkercke et al., 2016). It is therefore possible that interactions with other proteins, not necessarily clade IVa bHLH transcription factors, determine the specificity of these transcription factors toward the specific saponin branches.

### CYP88 Proteins Play a Conserved Role in Triterpenoid Biosynthesis

*CYP88A13*, which is strictly coregulated with *TSAR3* in developing *M. truncatula* seeds, encodes a cytochrome P450 that catalyzes the C-16$\alpha$ hydroxylation of medicagenic acid toward zanhic acid. C-16$\alpha$ hydroxylation of triterpene backbones has also been demonstrated for cytochrome P450s belonging to the CYP716 and CYP87 families (Moses et al., 2014, 2015; Miettinen et al., 2017), making the CYP88 family the third P450 family shown to have the inherent capacity to catalyze the C-16$\alpha$ hydroxylation of triterpenoids. The discovery that a CYP88 family member is responsible for this reaction suggests that CYP88 proteins play a conserved role 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 cucumber (*Cucumis sativus*) CYP88L2 carry out the two-step oxidation of $\beta$-amyrin at the C-11 position to produce 11-oxo-$\beta$-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), it appears that the CYP88 family is also an important P450 family that contributes to triterpenoid diversity. However, in contrast to the CYP716 family, whose role appears to have evolved specifically toward triterpenoid biosynthesis (Miettinen et al., 2017), CYP88 family members are primarily considered to be *ent*-kaurenoic acid oxidases involved in gibberellin biosynthesis (Bak et al., 2011). Based on the structural similarity around the C-7 position of *ent*-kaurenoic acid and the C-16 position of medicagenic acid (Supplemental Figure 15), it appears that neofunctionalization of a duplicated ent-kaurenoic acid oxidase toward a medicagenic acid C-16$\alpha$ hydroxylase might have occurred. A similar recruitment of a CYP88 protein that functioned in gibberellin biosynthesis was proposed for CYP88D6 (Hamberger and Bak, 2013). Thus, unlike the CYP716 family, the CYP88 family might not have originally been dedicated to triterpene biosynthesis, but its reoccurring involvement in this process might instead be a consequence of the structural and physicochemical similarities between diterpenes and triterpenes.

### Gene Duplications as a 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, it appears that *CYP88A14* is not involved in zanhic acid biosynthesis. Yet, we were able to show that this P450 also has a capacity for triterpene oxidation, although

its exact catalytic activity remains to be determined. Gene duplication followed by sub- or neofunctionalization is considered to be a major driving force of evolutionary novelty in plant specialized metabolism (Ober, 2010). This notion is supported by the apparently different functionality of CYP88A14 toward triterpenes compared to CYP88A13, which is encoded by its tandem repeat gene. Like *CYP88A13* and *CYP88A14*, the UGT genes *UGT73F18* and *UGT73F19* also 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, these duplicated genes also appear 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 encode functional C-3 glycosyltransferases.

Tandem gene duplications of UGTs involved in saponin biosynthesis appear 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 that are 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* glycosyltransferase 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, appears 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*. However, *Medtr4g031800*, encoding UGT73K1, is not part of a tandem repeat. Interestingly, this gene is located right next to *CYP72A61* (*Medtr4g031820*), which encodes a P450 involved in nonhemolytic soyasaponin biosynthesis. Therefore, *UGT73K1* appears to be part of a small metabolic gene cluster, a phenomenon not uncommon in plant specialized metabolism (Nützmann et al., 2016, 2018). Hence, in addition to being coexpressed, saponin biosynthesis genes often appear to occur in tandem arrays or small metabolic gene clusters. As such, the location of a candidate gene in the genome might serve as 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 *Medicago truncatula* genome version 3.5 (Young et al., 2011) or 4.0 (Tang et al., 2014) and cloned using Gateway technology (Invitrogen). The gene sequences were PCR amplified (for primers, see Supplemental Data Set 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 *Agrobacterium tumefaciens* C58C1 RifR (PMP90) for transient expression in *Nicotiana benthamiana* leaves. For expression in yeast (*Saccharomyces cerevisiae*), 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 the 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 by Mertens et al. (2016a).

### Transient Expression Assays in Tobacco Protoplasts

Transient expression assays in tobacco (*Nicotiana tabacum* 'Bright Yellow-2') protoplasts were performed as previously described by Vanden Bossche et al. (2013). In brief, 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 the 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 performed in eight biological repeats of independently transfected protoplasts. Promoter activity was normalized by dividing the fLUC value by the corresponding rLUC value. The average of the normalized fLUC values was calculated and expressed 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 performed as previously described by Pollier et al. (2011).

### RT-qPCR Analysis

Hairy roots grown for 21 d in liquid medium as described (Pollier et al., 2011) were harvested by flash freezing in liquid nitrogen. The samples were ground in liquid nitrogen with a mortar and pestle and used for total RNA extraction and first-strand cDNA synthesis with an RNeasy mini kit (Qiagen) and an iScript cDNA synthesis kit (Bio-Rad), respectively. RT-qPCR was performed with a LightCycler 480 system (Roche), SYBR green master mix (Thermo Fisher Scientific), and RT-qPCR primers (Supplemental Data Set 5) designed using primer3 version 0.4.0 (Thornton and Basu, 2011;

Untergasser et al., 2012). RT-qPCR was performed in a 5-μL of reaction mixture in 384-well reaction plates. After initial denaturation for 10 min at 95°C, the PCR consisted of 45 cycles of denaturation (95°C for 10 s), annealing (60°C for 15 s), and extension (72°C for 15 s). For the 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.

### Coexpression Analyses

To select candidate P450s, two coexpression analyses were performed on microarray data sets of developing *M. truncatula* seeds. The first data set consisted of 19,012 differentially expressed probe sets during *M. truncatula* seed development (Verdier et al., 2013). Genes coexpressed with *TSAR3* were identified by calculating the Pearson's correlation coefficient between the probe set corresponding to *TSAR3* and all other probe sets in the data set. The second data set 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 the first data set, genes coexpressed with *TSAR3* were identified by calculating the Pearson's correlation coefficient between the probe set corresponding to *TSAR3* and all other probe sets in this data set. Finally, to select candidate UGTs, a third coexpression analysis was performed in which the *M. truncatula* gene expression atlas (He et al., 2009) was probed for UGTs coexpressed with *CYP716A12*. To this end, the expression values of all probe sets 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 to be coexpressed 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 from the similarly engineered strain TM1 (Moses et al., 2014) as the 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 cotransformed with the resulting PCR fragments. Transformants were selected on Yeast Peptone Dextrose (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.

The production of medicagenic acid was achieved by coexpressing 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 Cultivation for Triterpenoid Production

Yeast transformations were performed using the lithium acetate/single-stranded carrier DNA/polyethylene glycol method (Gietz and Woods, 2002), and transformed cells were selected on Synthetic Defined medium supplemented with the appropriate dropout supplements (Clontech). To evaluate the P450s, yeast cells were cultivated in the presence of methyl-β-cyclodextrin as described by Moses et al. (2014). For GC-MS analysis, 1 mL of the spent medium was extracted three times with 0.5 volumes of ethyl acetate. The organic extracts were pooled, evaporated to dryness under a vacuum, and the resulting residue was used for GC-MS analysis. To evaluate the UGTs, yeast cells were cultivated without methyl-β-cyclodextrin. For LC-MS analysis, yeast cells from 10 mL of cultivation medium were harvested in a 2-mL Eppendorf tube by centrifugation. The 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 a vacuum. The resulting residue was used for LC-MS analysis.

### Agroinfiltration of *N. benthamiana* Leaves

*A. tumefaciens* C58C1 RifR (pMP90) cells were cultivated for 2 d in a shaking incubator (150 rpm) at 28°C in 5 mL of yeast extract broth medium supplemented with 20 μg/mL gentamicin, 100 μg/mL rifampicin, and 25 μg/mL kanamycin. After incubation, 500 μL of the precultures were used to inoculate 9.5 mL of yeast extract broth medium supplemented with antibiotics, 10 mM of MES (pH 5.7), and 20 μM of acetosyringone. After overnight incubation, bacteria for transient coexpression 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 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 into the abaxial sides of fully expanded leaves of 3- to 4-week-old *N. benthamiana* plants grown in the greenhouse (21°C, 60% humidity, 16-h/8-h light/dark regime, 40 to 60 μmol/$m^2$s photosynthetic photon flux). The infiltrated plants were incubated under normal growth conditions for 4 d 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 a 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 medium 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 performed using a GC model 6890 and MS model 5973 (Agilent). One microliter 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 × 0.25 mm, 0.25 μm; Varian CP9013; Agilent) with helium carrier gas at a constant flow rate of 1 mL/min. The oven was held at 80°C for 1 min postinjection, 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 cooled to 80°C at 50°C/min at the end of the run. The Mass Selective Detector 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 performed 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 microliter 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 × 0.25 mm, 0.25 μm; Varian CP9013; Agilent) with helium carrier gas at a constant flow rate of 1.2 mL/min. The oven was held at 80°C for 1 min postinjection, 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 cooled to 80°C at 50°C/min at the end of the run. The Mass Selective Detector 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 to 800 at a resolution of >25,000 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 water/cyclohexane (1:1, v/v). The samples were centrifuged (10 min at 20,800*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 × 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 the untargeted metabolite profiling experiments were determined using Progenesis QI software (Waters). PCA and PLS-DA were performed using log-transformed, pareto-scaled MS 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 Data Sets 1 and 2) were obtained from Aldo Tava (CRA-FLC, Italy) and Prof. Wiesław Oleszek (Institute of Soil Science and Plant Cultivation, Poland) and were previously described by Pollier et al. (2013).

### Genotyping of *Tnt1* Insertion Mutants

*Tnt1* insertion mutants of *Medtr2g104650* (*TSAR3*) and *Medtr5g014240* (*CYP88A13*) were screened using a method modified from Cheng et al. (2011). In brief, NF14212, NF15672, NF17107, NF11894, NF17520, and NF13195 seeds from the Noble Research Institute were germinated and grown in soil. Plants were grown in the greenhouse (21°C, 60% humidity, 16-h/8-h light/dark regime, 40 to 60 μmol/$m^2$s photosynthetic photon flux). 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 by Cheng et al. (2011) and used as a PCR template. PCR was performed with gene-specific primers and primers within the *Tnt1* retrotransposon (Supplemental Figures 3 and 12; Supplemental Data Set 5). The 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 regenotyped and used for transcript and metabolite profiling. For the *TSAR3 Tnt1* insertion lines, seeds were harvested ~24 d after pollination (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 Data Set 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 performed with 1000 replicates. A machine-readable tree file is provided as Supplemental Data Set 7.

### Accession Numbers

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

### Supplemental Data

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

**Supplemental Figure 2.** PLS-DA of samples from $TSAR3^{OE}$ and control hairy roots

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

**Supplemental Figure 4.** Hemolytic saponin biosynthesis is not affected in the 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 Figure 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 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 the generation of strain JP034

**Supplemental Figure 17.** *M. truncatula* seed development

**Supplemental Table 1.** Yeast strains used in this study

**Supplemental Table 2.** Candidate UGTs identified through the different coexpression analyses

**Supplemental Data Set 1.** Differential metabolites between $TSAR3^{OE}$ and GUS control lines

**Supplemental Data Set 2.** Differential metabolites in developing seeds of *tsar3* and wild-type plants

**Supplemental Data Set 3.** Coexpression analysis of publicly available microarray data sets from developing *M. truncatula* seeds

**Supplemental Data Set 4.** Coexpression analysis of a set of 60 microarray data sets from developing *M. truncatula* seeds

**Supplemental Data Set 5.** List of oligonucleotides used in this study

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

**Supplemental Data Set 7.** Machine-readable tree file in Newick format in support of Supplemental Figure 14

## ACKNOWLEDGMENTS

We thank David Nelson (Committee on Standardized Cytochrome P450 Nomenclature) for naming the P450s, Michael Court (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, Aldo Tava (CRA-FLC, Italy) and Wiesław Oleszek (Institute of Soil Science and Plant Cultivation, Poland) for the *Medicago* saponin standards, and Annick Bleys (VIB-Ghent University) for help with preparing the article. This work was supported by the Research Foundation Flanders (predoctoral fellowships to E.C. and J.M. and a postdoctoral fellowship to J.P.), the Program Ciências Sem Fronteiras (predoctoral fellowship grant 201135/2014-0 to B.R. and a Society of Women Engineers fellowship grant 202776/2014-0 to G.C.), the FAPESP - São Paulo Research Foundation (postdoctoral fellowship grant 2017/10911-1 to K.U.B.), the VIB (International PhD Fellowship Program predoctoral fellowship to P.A.), the European Union Horizon 2020 (research project grants Endoscape-825730 and Eucleg-727312 to A.G.), and Ghent University (Concerted Research Actions project grant BOF18/GOA/013 to A.G.).

## 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 article. A.G. helped with the writing.

Received August 7, 2019; revised March 27, 2020; accepted April 10, 2020; published April 17, 2020.

## 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 J. **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. J. Chem. Ecol. **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. Metab. Eng. **40:** 165–175.

**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., and Bak, S.** (2012). UDP-glycosyltransferases from the UGT73C subfamily in *Barbarea vulgaris* catalyze sapogenin 3-*O*-glucosylation in saponin-mediated insect resistance. Plant Physiol. **160:** 1881–1895.

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

**Bak, S., Beisson, F., Bishop, G., Hamberger, B., Höfer, R., Paquette, S., and Werck-Reichhart, D.** (2011). Cytochromes p450. 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. Mol. 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 Physiol. **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. J. Exp. Bot. **56:** 323–336.

**Carelli, M., et al.** (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 Mol. Biol. **678:** 179–190.

**Chini, A., Gimenez-Ibanez, S., Goossens, A., and Solano, R.** (2016). Redundancy and specificity in jasmonate signalling. Curr. Opin. Plant Biol. **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 Res. **46** (W1): 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. Mar. Drugs **17:** E352.

**Colinas, M., and Goossens, A.** (2018). Combinatorial transcriptional control of plant specialized metabolism. Trends Plant Sci. **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 Biotechnol. J. **7:** 172–182.

**Erthmann, P., 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 Mol. Biol. **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 Sci. **250:** 188–197.

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

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

**Gholami, A., De Geyter, N., Pollier, J., Goormachtig, S., and Goossens, A.** (2014). Natural product biosynthesis in *Medicago* species. Nat. Prod. Rep. **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 Enzymol. **350:** 87–96.

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

**Goossens, J., Mertens, J., and Goossens, A.** (2017). Role and functioning of bHLH transcription factors in jasmonate signalling. J. Exp. Bot. **68:** 1333–1347.

**Hamberger, B., and Bak, S.** (2013). Plant P450s as versatile drivers for evolution of species-specific chemical diversity. Philos. Trans. R. Soc. Lond. B Biol. Sci. **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. Mol. Biol. Evol. **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 Biol. **8:** R19.

**Huhman, D.V., Berhow, M.A., and Sumner, L.W.** (2005). Quantification of saponins in aerial and subterranean tissues of *Medicago truncatula*. J. Agric. Food Chem. **53:** 1914–1920.

**Itkin, M., et al.** (2016). The biosynthetic pathway of the nonsugar, high-intensity sweetener mogroside V from *Siraitia grosvenorii*. Proc. Natl. Acad. Sci. USA **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 Mol. Biol. **51:** 731–743.

**Jarvis, D.E., et al.** (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. J. Agric. Food Chem. **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. J. Agric. Food Chem. **53:** 2164–2170.

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

**Kazan, K., and Manners, J.M.** (2013). MYC2: The master in action. Mol. Plant **6:** 686–703.

**Khakimov, B., Kuzina, V., Erthmann, P., 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 J. **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 J. **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*. Front Plant Sci **10:** 850.

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

**Mertens, J., Van Moerkercke, A., Vanden Bossche, R., Pollier, J., and Goossens, A.** (2016b). 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 Physiol. **57:** 2564–2575.

**Miettinen, K., et al.** (2017). The ancient CYP716 family is a major contributor to the diversification of eudicot triterpenoid biosynthesis. Nat. Commun. **8:** 14153.

**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*. Proc. Natl. Acad. Sci. USA **111:** 1634–1639.

**Moses, T., Pollier, J., Faizal, A., Apers, S., Pieters, L., Thevelein, J.M., Geelen, D., and Goossens, A.** (2015). Unraveling the triterpenoid saponin biosynthesis of the African shrub *Maesa lanceolata*. Mol. Plant **8:** 122–135.

**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 Phytol. **211:** 771–789.

**Nützmann, H.-W., Scazzocchio, C., and Osbourn, A.** (2018). Metabolic gene clusters in eukaryotes. Annu. Rev. Genet. **52:** 159–183.

**Ober, D.** (2010). Gene duplications and the time thereafter: Examples from plant secondary metabolism. Plant Biol (Stuttg) **12:** 570–577.

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

**Podolak, I., Galanty, A., and Sobolewska, D.** (2010). Saponins as cytotoxic agents: A review. Phytochem. Rev. **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. J. Nat. Prod. **74:** 1462–1476.

**Pollier, J., et al.** (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. Phytochem. Lett. **20:** 520–539.

**Reed, J., and Osbourn, A.** (2018). Engineering terpenoid production through transient expression in *Nicotiana benthamiana*. Plant Cell Rep. **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. Metab. Eng. **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 Sci. **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 Biotechnol. J. **7:** 682–693.

**Sayama, T., et al.** (2012). The *Sg-1* glycosyltransferase locus regulates structural diversity of triterpenoid saponins of soybean. Plant Cell **24:** 2123–2138.

**Seki, H., Ohyama, K., Sawai, S., Mizutani, M., Ohnishi, T., Sudo, H., 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. Proc. Natl. Acad. Sci. USA **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., et al.** (2014). Plant science. 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. Front Plant Sci **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 J. **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 J. **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. Mol. Biol. Evol. **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 Physiol. **59:** 778–791.

**Tang, H., et al.** (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*. Phytochem. Rev. **10:** 459–469.

**Thornton, B., and Basu, C.** (2011). Real-time PCR (qPCR) primer design using free online software. Biochem. Mol. Biol. Educ. **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 identifies 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 Res. **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 J. **88:** 3–12.

**Van Moerkercke, A., et al.** (2015). The bHLH transcription factor BIS1 controls the iridoid branch of the monoterpenoid indole alkaloid pathway in *Catharanthus roseus*. Proc. Natl. Acad. Sci. USA **112:** 8130–8135.

**Vanden Bossche, R., Demedts, B., Vanderhaeghen, R., and Goossens, A.** (2013). Transient expression assays in tobacco protoplasts. Methods Mol. Biol. **1011:** 227–239.

**Verdier, J., et al.** (2013). A regulatory network-based approach dissects late maturation processes related to the acquisition of desiccation tolerance and longevity of *Medicago truncatula* seeds. Plant Physiol. **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 Feussner, I.** (2018). The oxylipin pathways: Biochemistry and function. Annu. Rev. Plant Biol. **69:** 363–386.

**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*. Ann. Bot. **111:** 1021–1058.

**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 J. **89:** 527–539.

**Young, N.D., et al.** (2011). The *Medicago* genome provides insight into the evolution of rhizobial symbioses. Nature **480:** 520–524.