The bHLH Transcription Factors TSAR1 and TSAR2 Regulate Triterpene Saponin Biosynthesis in *Medicago truncatula*[1][OPEN]

Jan Mertens, Jacob Pollier, Robin Vanden Bossche, Irene Lopez-Vidriero, José Manuel Franco-Zorrilla, and Alain Goossens*

Department of Plant Systems Biology, VIB, B–9052 Ghent, Belgium (J.M., J.P., R.V.B., A.G.); Department of Plant Biotechnology and Bioinformatics, Ghent University, B–9052 Ghent, Belgium (J.M., J.P., R.V.B., A.G.); and Genomics Unit, Centro Nacional de Biotecnología-Consejo Superior de Investigaciones Científicas, 28049 Madrid, Spain (I.L.-V., J.M.F.-Z.)

ORCID IDs: 0000-0003-1878-9903 (J.M.); 0000-0002-1134-9238 (J.P.); 0000-0002-4593-9658 (R.V.B.); 0000-0001-9362-8662 (I.L.-V.); 0000-0001-6769-7349 (J.M.F.-Z.); 0000-0002-1599-551X (A.G.).

Plants respond to stresses by producing a broad spectrum of bioactive specialized metabolites. Hormonal elicitors, such as jasmonates, trigger a complex signaling circuit leading to the concerted activation of specific metabolic pathways. However, for many specialized metabolic pathways, the transcription factors involved remain unknown. Here, we report on two homologous jasmonate-inducible transcription factors of the basic helix-loop-helix family, TRITERPENE SAPONIN BIOSYNTHESIS ACTIVATING REGULATOR1 (TSAR1) and TSAR2, which direct triterpene saponin biosynthesis in *Medicago truncatula*. TSAR1 and TSAR2 are coregulated with and transactivate the genes encoding 3-HYDROXY-3-METHYLGLUTARYL-COENZYME A REDUCTASE1 (HMGR1) and MAKIBISHI1, the rate-limiting enzyme for triterpene biosynthesis and an E3 ubiquitin ligase that controls HMGR1 levels, respectively. Transactivation is mediated by direct binding of TSARs to the N-box in the promoter of HMGR1. In transient expression assays in tobacco (*Nicotiana tabacum*) protoplasts, TSAR1 and TSAR2 exhibit different patterns of transactivation of downstream triterpene saponin biosynthetic genes, hinting at distinct functionalities within the regulation of the pathway. Correspondingly, overexpression of *TSAR1* or *TSAR2* in *M. truncatula* hairy roots resulted in elevated transcript levels of known triterpene saponin biosynthetic genes and strongly increased the accumulation of triterpene saponins. *TSAR2* overexpression specifically boosted hemolytic saponin biosynthesis, whereas *TSAR1* overexpression primarily stimulated nonhemolytic soyasaponin biosynthesis. Both TSARs also activated all genes of the precursor mevalonate pathway but did not affect sterol biosynthetic genes, pointing to their specific role as regulators of specialized triterpene metabolism in *M. truncatula*.

Plants are frequently confronted with various sorts of biotic and abiotic stress situations. This triggers defense responses such as the production of bioactive specialized metabolites. These compounds are often family, genus, or even species specific and thereby constitute a distinct metabolic fingerprint. A specific group of defense compounds are the saponins, a structurally diverse class of amphipathic glycosides with a lipophilic triterpenoid, steroid, or steroidal alkaloid aglycone backbone, also called sapogenin, which is covalently linked to one or more hydrophilic sugar chains via a glycosidic bond (Augustin et al., 2011; Osbourn et al., 2011; Gholami et al., 2014). The structural and functional diversity of the saponins is reflected by their broad spectrum of biological activities that encompass, among others, antimicrobial, antiinsect, allelopathic, anticarcinogenic, cholesterol-lowering, antiinflammatory, and hepatoprotective activities (Avato et al., 2006; Vincken et al., 2007; Augustin et al., 2011; Pollier and Goossens, 2012; Moses et al., 2013). The model legume *Medicago truncatula* (barrel medic), a member of the Fabaceae plant family, provides a rich source of pentacyclic, oleanane-type triterpene saponins (TSs) and has been widely used to study TS biosynthesis (Gholami et al., 2014).

The TS-specific biosynthesis starts with the cyclization of 2,3-oxidosqualene (Supplemental Fig. S1). This is a precursor shared with the phytosterol synthesis route and is a condensation product of six isopentenyl pyrophosphate (IPP) units. IPP is generated through the

[1] This work was supported by the Special Research Funds from Ghent University (project no. BOF01J14813) and by the Research Foundation Flanders (predoctoral fellowship to J.M. and postdoctoral fellowship to J.P.).

* Address correspondence to alain.goossens@psb-vib.ugent.be.

The author 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.plantphysiol.org) is: Alain Goossens (alain.goossens@psb-vib.ugent.be).

J.M., J.P., and A.G. designed the research; J.M., J.P., R.V.B. and I.L.-V. performed the research; J.M., J.P., J.M.F.-Z., and A.G. analyzed the data; J.M., J.P., and A.G. wrote the article.

[OPEN] Articles can be viewed without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.15.01645

cytosolic mevalonate (MVA) pathway. The key rate-limiting enzyme of this pathway is 3-HYDROXY-3-METHYLGLUTARYL-COA REDUCTASE (HMGR), which catalyzes the formation of MVA and of which five isoforms have been characterized in *M. truncatula* (Kevei et al., 2007). The cyclization of 2,3-oxidosqualene forms the branch point between primary phytosterol and secondary TS metabolism. During primary sterol metabolism, 2,3-oxidosqualene is cyclized to cycloartenol by cycloartenol synthase (Corey et al., 1993), whereas during TS biosynthesis, 2,3-oxidosqualene is cyclized to the pentacyclic aglycone β-amyrin by β-amyrin synthase (BAS; Suzuki et al., 2002; Iturbe-Ormaetxe et al., 2003). Subsequently, the competitive action of two cytochrome P450-dependent monooxygenases (P450s) causes another branching of the TS biosynthetic pathway in *M. truncatula*, and consequently, the *M. truncatula* TSs are divided into two distinct classes: hemolytic and nonhemolytic TSs. The first committed step for the production of hemolytic TSs is carried out by CYP716A12, which performs three consecutive oxidations at position C-28 of β-amyrin to yield oleanolic acid (Carelli et al., 2011; Fukushima et al., 2011). Subsequently, additional oxidations frequently occur at C-2 and C-23 and are carried out by the P450 enzymes CYP72A67 and CYP72A68v2 (Fukushima et al., 2013; Biazzi et al., 2015). The nonhemolytic soyasaponins are distinguished by hydroxylation of β-amyrin at position C-24 catalyzed by CYP93E2 (Fukushima et al., 2013). In planta, this modification precludes oxidation at C-28 (Tava et al., 2011) and is followed by oxidation at the C-22 position by CYP72A61v2 to yield soyasapogenol B (Fukushima et al., 2013). Additional decorations of the backbone by P450s and the covalent attachment of sugar moieties by UDP-dependent glycosyltransferases (UGTs) further diversify the TS compendium in *Medicago* spp. (Achnine et al., 2005; Naoumkina et al., 2010; Gholami et al., 2014).

All *M. truncatula* organs appear to accumulate TSs, more particularly as tissue-specific mixes of tens of different TSs. Besides this constitutive accumulation, induced TS biosynthesis is often observed in the response to herbivore feeding or pathogen attack (Gholami et al., 2014). Inducible TS biosynthesis under stress conditions is mediated by concerted transcriptional activation of the TS pathway (Broeckling et al., 2005; Suzuki et al., 2005; Pollier et al., 2013a), a molecular process in which jasmonates (JAs) play a crucial role. JAs are oxylipin-derived phytohormones that mediate the reprogramming of many metabolic pathways in response to different environmental and developmental cues (Pauwels et al., 2009; De Geyter et al., 2012). Accordingly, TS production is strongly enhanced in *M. truncatula* cell suspension cultures treated exogenously with JAs (Broeckling et al., 2005; Suzuki et al., 2005). To date, little is known about the regulators involved. Posttranslational regulation of TS biosynthesis has been shown to be imposed by MAKIBISHI1 (MKB1), a RING membrane anchor-like E3 ubiquitin ligase that monitors TS production by targeting HMGR for endoplasmic reticulum-associated degradation by the 26S proteasome (Pollier et al., 2013a). However, the transcription factors (TFs) triggering the concerted transcriptional activation of TS biosynthetic genes following JA perception have remained elusive. In fact, only a few TFs specifically modulating plant terpene biosynthesis have been identified in general. The basic helix-loop-helix (bHLH) TF MYC2, also known as a primary player in the JA signaling cascade (Kazan and Manners, 2013), and its homologs have been shown to play a role in the regulation of the biosynthesis of sesquiterpenes in Arabidopsis (*Arabidopsis thaliana*), tomato (*Solanum lycopersicum*), and *Artemisia annua* (Hong et al., 2012; Ji et al., 2014; Spyropoulou et al., 2014). Very recently, two other bHLH TFs, Bl (bitter leaf) and Bt (bitter fruit), not related to MYC2, were found to regulate the accumulation of cucurbitacin triterpenes in cucumber (*Cucumis sativus*; Shang et al., 2014). Likewise, another bHLH TF not related to MYC2, BHLH IRIDOID SYNTHESIS1 (BIS1), has been shown to control the monoterpene (iridoid) branch of the monoterpene indole alkaloid (MIA) pathway in *Catharanthus roseus* (Van Moerkercke et al., 2015).

In this study, we examined transcriptomics data sets from *M. truncatula*, and through coexpression analyses, we identified two highly specialized bHLH TFs that are involved in the regulation of the different branches of TS metabolism in *M. truncatula*.

RESULTS

Coexpression Analyses Reveal Candidate Regulators for TS Metabolism in *M. truncatula*

Previously, we observed a strikingly strong coexpression of *HMGR1* and *MKB1* in *M. truncatula* roots and suspension cells under various stress conditions and/or treated with phytohormones such as JAs (Pollier et al., 2013a). TFs regulating specialized metabolite pathways are often also coexpressed with the target genes encoding the pathway enzymes (De Geyter et al., 2012). Hence, in order to identify candidate regulators of the MVA and/or TS biosynthesis pathways in *M. truncatula*, we mined the *Medicago truncatula* Gene Expression Atlas (MtGEA [http://bioinfo.noble.org/gene-atlas/]; He et al., 2009) for TF-encoding genes with expression profiles that strongly overlap with those of the *HMGR1* and *MKB1* genes in the tissues and conditions mentioned above. This allowed the compilation of a short list of six TFs that were coexpressed with *HMGR1* and *MKB1* with a Pearson's correlation coefficient higher than 0.6 (Table I; Fig. 1A; Supplemental Fig. S2). This list comprised genes encoding four bHLH proteins, one MYB protein, and one homeodomain-leucine zipper (HD-ZIP) protein. By subsequent BLAST analysis with these TF sequences against the *M. truncatula* genome, we identified a seventh TF-encoding gene, *Medtr4g066460*, a homolog of the bHLH Medtr7g080780, which was also JA inducible and followed a similar trend under some of the selected expression conditions; therefore, it was also selected for further functional analysis.

Table I. *Selected TFs that display highly overlapping expression profiles with HMGR1 or MKB1*

Coexpression was analyzed using the MtGEA tool (He et al., 2009). All conditions shown in Figure 1A were taken into account for the analyses. Pearson's correlation coefficients were calculated to measure the degree of coexpression. Probe set sequences were blasted against the *M. truncatula* genome version 4.0 to retrieve the corresponding *M. truncatula* gene identifiers (Tang et al., 2014). Medtr4g066460 is a homolog of Medtr7g080780 that displays a similar trend under some of the conditions based on visual inspection.

Probe Set	Gene Identifier	TF Family	Pearson's Correlation Coefficient	
			HMGR1	MKB1
Mtr.10397.1.S1_at	Medtr5g026500/HMGR1	–	1.0	0.7983
Mtr.43815.1.S1_at	MKB1[a]	–	0.7983	1.0
Mtr.28568.1.S1_at	Medtr7g117670	bHLH	0.8027	<0.6
Mtr.38762.1.S1_at	Medtr8g027495	bHLH	0.6495	0.7246
Mtr.51379.1.S1_at	Medtr2g038040	bHLH	0.7742	0.6379
Mtr.43316.1.S1_at	Medtr7g080780/TSAR1	bHLH	0.7527	0.646
Mtr.38413.1.S1_at	Medtr3g065440	MYB	0.6685	<0.6
Mtr.18769.1.S1_at	Medtr8g026960	HD-ZIP	0.8204	0.6084
Mtr.9397.1.S1_at	Medtr4g066460/TSAR2	bHLH	<0.6	<0.6

[a]The *MKB1* gene is not present on the *M. truncatula* genome version 4.0 (Tang et al., 2014).

Two Subclade IVa bHLH TFs Transactivate the Promoters of *HMGR1* and *MKB1*

To test whether the putative regulators are able to transactivate the promoters of TS biosynthetic genes, we launched a transient expression assay (TEA) screen in tobacco (*Nicotiana tabacum*) protoplasts (De Sutter et al., 2005; Vanden Bossche et al., 2013). To this end, we cloned the 1,000-bp region upstream of the start codon of *HMGR1* (*ProHMGR1*) and fused it to the *FIREFLY LUCIFERASE* (*fLUC*) gene to create a reporter construct.

This promoter construct was cotransformed in tobacco protoplasts with the candidate TFs driven by the cauliflower mosaic virus (CaMV) 35S promoter, revealing that two bHLH TFs strongly induced the luciferase activity by 9- and 28-fold, respectively, compared with the fLUC activity in protoplasts cotransfected with a GUS control (Fig. 1B). These two TFs, which we named TRITERPENE SAPONIN ACTIVATION REGULATOR1 (TSAR1) and TSAR2, correspond to the homologous genes *Medtr7g080780* and *Medtr4g066460*, respectively. Since the remaining five TFs did not have an effect on *ProHMGR1* transactivation comparable with that of the two TSARs (Supplemental Fig. S3), we focused on TSAR1 and TSAR2. We first examined whether they could modulate *MKB1* expression in a TEA using the 1,000-bp promoter region upstream of the *MKB1* transcriptional start site (*ProMKB1*). This region was defined as such because the *MKB1* open reading frame (ORF) is preceded by a 252-bp 5′ untranscribed region (UTR) containing an intron of 1,106 bp. Both TSAR1 and TSAR2 transactivated *ProMKB1* with strengths comparable to those of *ProHMGR1* (Fig. 1B), indicating that they represent potential general regulators of *M. truncatula* TS biosynthesis.

Members of the bHLH family possess an overall low sequence homology but are defined by their bHLH signature domain that spans about 50 amino acids (Heim et al., 2003; Toledo-Ortiz et al., 2003; Carretero-Paulet et al., 2010; Pires and Dolan, 2010). The N-terminal basic region of approximately 15 amino acids is responsible for DNA binding and specificity. The C-terminal helix-loop-helix region of approximately 45 amino acids contains two amphipathic α-helices separated by a loop region and promotes the formation of homodimeric or heterodimeric protein complexes. To determine which phylogenetic clade of the bHLH family harbors TSAR1 and TSAR2, we constructed a neighbor-joining tree based on the alignment of their bHLH domains, including all bHLH proteins from Arabidopsis clades I, III, and IV and the bHLH-type regulators of triterpene synthesis in cucumber. Both TSAR1 and TSAR2 are confined in subclade IVa of the bHLH family, as defined by Heim et al. (2003; Supplemental Figs. S4 and S5), whereas the well-known MYC2-type and the recently identified cucumber triterpene-regulating bHLH proteins belong to subclades IIIe and Ib, respectively. Notably, BIS1, the recently identified positive regulator of the iridoid branch of MIA biosynthesis in *C. roseus*, also belongs to subclade IVa of the bHLH proteins (Van Moerkercke et al., 2015) and, therefore, was included in our phylogenetic analysis for further comparison (Supplemental Figs. S4 and S5).

TSAR1 and TSAR2 Mediate *HMGR* Transactivation by Binding the N-Box in the *HMGR* Promoter

Many studied bHLH proteins, including MYC2 and the cucumber Bl and Bt, act through recognition of an E-box hexanucleotide sequence in the promoter region, which has the consensus sequence 5′-CANNTG-3′ (Carretero-Paulet et al., 2010; Kazan and Manners, 2013; Shang et al., 2014). However, it has been shown that bHLHs have different affinities for variations of this consensus sequence (Fernández-Calvo et al., 2011). As several potential E-box sequences are present in *ProHMGR1*, we pinpointed the promoter elements that

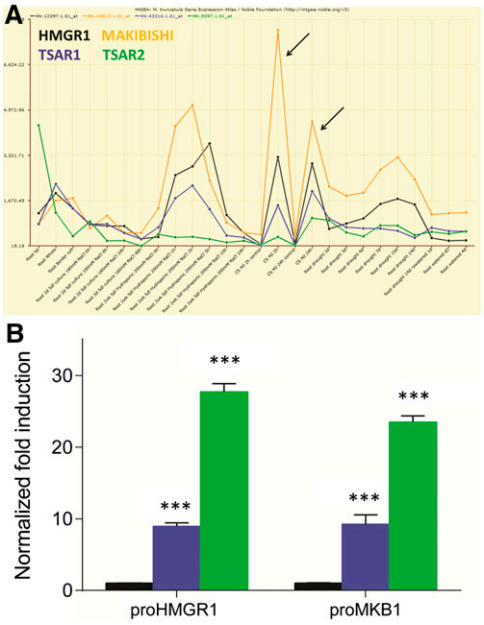


Figure 1. *TSAR1* and *TSAR2* are coexpressed with and transactivate *HMGR1* and *MKB1*. A, Coexpression profiles of *HMGR1* (black; Mtr.10397), *MKB1* (orange; Mtr.43815), *TSAR1* (blue; Mtr43316), and *TSAR2* (green; Mtr.9397) in *M. truncatula* roots under various culturing conditions, generated with the MtGEA tool (He et al., 2009). Values on the *y* axis reflect transcript levels determined by microarray analysis (He et al., 2009). Arrows depict values in methyl jasmonate-treated cell suspension cultures. B, Transactivation of *ProHMGR1* and *ProMKB1* by TSAR1 (blue) and TSAR2 (green) 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 (black). For the normalization procedure, see "Materials and Methods." *ProHMGR1* and *ProMKB1* span the 1,000-bp region upstream of the translational and transcriptional start site of the *HMGR1* and *MKB1* gene promoters, respectively. The error bars designate SE ($n = 8$). Statistical significance was determined by Student's *t* test (***, $P < 0.001$).

are essential for TSAR1- and TSAR2-mediated activation by performing protoplast assays using promoter deletion constructs. First, two promoter constructs were generated that span the regions −1 to −500 (*ProHMGR1[-1, -500]*) and −101 to −281 (*ProHMGR1[-101, -281]*) relative to the start of the ORF. In the latter construct, the first 101 bp upstream of the ORF that represent the 5′ UTR were omitted. In TEAs in tobacco protoplasts, TSAR1 and TSAR2 transactivated both promoters (Fig. 2A). Upon examining *ProHMGR1[-101, -281]*, we identified one E-box-like motif (5′-CACGAG-3′), also referred to as an N-box (Pires and Dolan, 2010), at position −246 (Supplemental Table S1). To assess the importance of this box, we replaced it with the sequence TGAATT to create a mutated version (*ProHMGR1[-101, -281]* mut). This replacement completely impeded transactivation by both TSAR1 and TSAR2, indicating that the presence of this box is necessary for TSAR1/TSAR2 activity (Fig. 2A).

To assess whether TSAR1 and TSAR2 bind directly to the N-box, a yeast one-hybrid (Y1H) assay was carried out. To this end, a yeast strain was generated that contains a synthetic promoter construct in which the *ProHMGR1* N-box was repeated in triplicate (*3xCACGAG*$_{[HMGR1]}$). Yeast growth on selective medium was observed for yeast expressing *TSAR2* but, unexpectedly, not for yeast expressing *TSAR1* (Fig. 2B). The latter could be due to impaired folding and/or functionality of TSAR1 in yeast.

To further investigate the DNA-binding properties of TSAR1 and TSAR2, we used a protein-binding microarray, an in vitro system that allows screening for 11-mer nucleotide sequences targeted by TFs (Godoy et al., 2011). To this end, both TSAR TFs were fused with a maltose-binding protein (MBP) allowing detection by anti-MBP antibodies. Enrichment scores represent binding affinities per 8-mer motifs (Fig. 2C). In this assay, both TSAR1 and TSAR2 exhibit strongest affinity for the G-box (CACGTG), similar to Arabidopsis MYC2 (Fig. 2C). TSAR1, TSAR2, and MYC2 also show similar affinities for the N-box CACGAG but lower affinities compared with the G-box. Unlike MYC2, however, TSAR1 and TSAR2 show no to low affinity for the G-box-like motifs CATGTG and AACGTG. Conversely, as compared with MYC2, TSAR1 and TSAR2 show higher affinity for the box variant CACGCG. Together, our data demonstrate that the TSAR proteins preferably bind to the motif 5′-CACGHG-3′, in which H may be T, A, and C, and therefore support that this element is necessary and sufficient for the TSAR proteins to exert their activity.

TSAR1 and TSAR2 Transactivate TS Biosynthesis Gene Promoters

To assess the regulatory range of the TSARs, we cloned the promoter sequences of the genes encoding two nonhemolytic TS P450s (*ProCYP93E2* and *ProCYP72A61v2*), two hemolytic TS P450 promoters (*ProCYP716A12* and *ProCYP72A67*), as well as two UGT promoters (*ProUGT73K1* and *ProUGT73F3*) to make reporter constructs containing the 1,000-bp regions upstream of the start codon of all mentioned genes to carry out transactivation assays in tobacco protoplasts.

TSAR1 was able to transactivate the promoters of the nonhemolytic TS P450s, *CYP93E2* and *CYP72A61v2*, by 39- and 14-fold, respectively, whereas TSAR2 could only transactivate these constructs by 8- and 3-fold, respectively (Fig. 3A). We identified two N-boxes, at positions −252 and −210, within the promoter sequence of *CYP93E2* (*ProCYP93E2*; Supplemental Table S1). To assess the necessity of these elements for the transactivation of *ProCYP93E2*, we generated a reporter construct containing a promoter fragment spanning the region from −160 to −300 that encompasses both motifs. In parallel, we created a second construct, spanning the same promoter region but in which the N-boxes at −252 and −210 were mutated. As expected, this small promoter fragment was sufficient to mediate TSAR1 transactivation, whereas this was completely abolished in the mutant version (Fig. 3B), further supporting that the presence of an N-box is necessary and sufficient to enable TSAR1 activity.

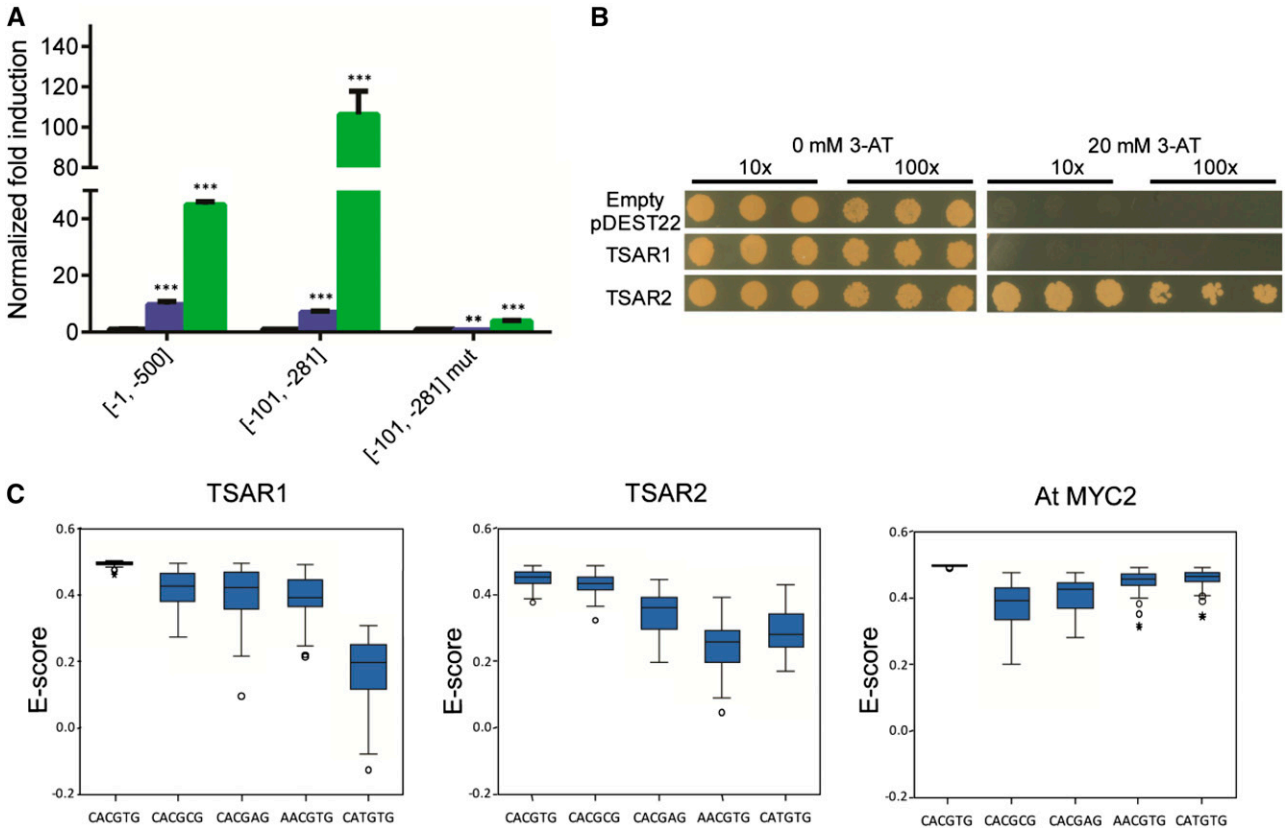


Figure 2. TSARs interact with the N-box in the *HMGR* promoter. A, Transactivation of *HMGR1* promoter fragments by TSAR1 (blue) and TSAR2 (green). Promoter fragments comprise the indicated regions relative to the start codon (in brackets). In the third fragment (mut), the N-box (CACGAG motif) was substituted by TGAATT. Values on the *y* axis are normalized fold changes relative to protoplasts cotransfected with the reporter constructs and a pCaMV35S:GUS control plasmid (in black). The error bars designate SE ($n = 8$). Statistical significance was determined by Student's *t* test (**, $P < 0.01$ and ***, $P < 0.001$). B, Y1H analysis of the binding of TSAR1 and TSAR2 to a *3xCACGAG*$_{[HMGR1]}$ promoter element. The *TSAR1* and *TSAR2* ORFs fused to GAL4AD or the empty vector control (pDEST22) were expressed in reporter strains harboring the *HIS3* gene under the control of a synthetic promoter element consisting of a triple repeat of the CACGAG element with 10 flanking nucleotides as found in the *HMGR1* promoter (*3xCACGAG*$_{[HMGR1]}$). Transformed yeast cultures dropped in serial dilutions (10- and 100-fold) were grown for 6 d on selective medium (minus His and plus 3-amino-1,2,4-triazole [3-AT]). C, Identification of TSAR-binding motifs in vitro by protein-binding microarray analysis. Shown are box plots of enrichment (E) values from G-box and N-box variants. The line in each box indicates the median (quartile 50%). Boxes indicate the quartiles from 25% to 75% of the distribution. Bars represent the quartiles from 1% to 25% and from 75% to 100%. Dots represent outliers. For comparison, we included Arabidopsis MYC2 (Godoy et al., 2011).

Neither TSAR1 nor TSAR2 transactivated the 1,000-bp promoters of the hemolytic TS P450s by more than 1.5-fold (Fig. 3A). This was a puzzling observation, but we reasoned that by spanning only 1,000 bp of the promoter region, we could have missed important elements for transactivation. Indeed, *ProCYP72A67* contained N-box sequences within the 1,500-bp region of the start codon (Supplemental Table S1). The corresponding 1,500-bp promoter fragment could successfully be cloned and used for TEAs. Using this reporter construct, clear transactivation by TSAR2 (9-fold) could be observed (Fig. 3B), further pointing to the importance of the N-box sequence also for TSAR2 activity. Remarkably, this long *ProCYP72A67* reporter construct was hardly activated by TSAR1 (only 2-fold; Fig. 3B). Taking into account the fact that TSAR1 appeared more efficient in transactivating nonhemolytic TS P450 gene promoters, this may point to specificities for TSAR1 and TSAR2 in the distinct TS pathway branches. No N-box-like sequences could be detected in the available sequence upstream of the cloned *ProCYP716A12*, and unfortunately, we did not manage to clone larger promoter fragments of *CYP716A12* either, based on the available *M. truncatula* genome version 4.0. Hence, transactivation of *ProCYP716A12* was not further assessed.

Finally, TSAR1 strongly induced luciferase activity using *ProUGT73K1*, compared with TSAR2 (i.e. 31- versus 5-fold), whereas *ProUGT73F3* was strongly transactivated by both TSAR1 and TSAR2 (i.e. by 24- and 48-fold; Fig. 3A). Together, these findings indicate that, at least in TEAs in tobacco protoplasts, both TSAR1 and TSAR2 encompass the whole TS pathway,

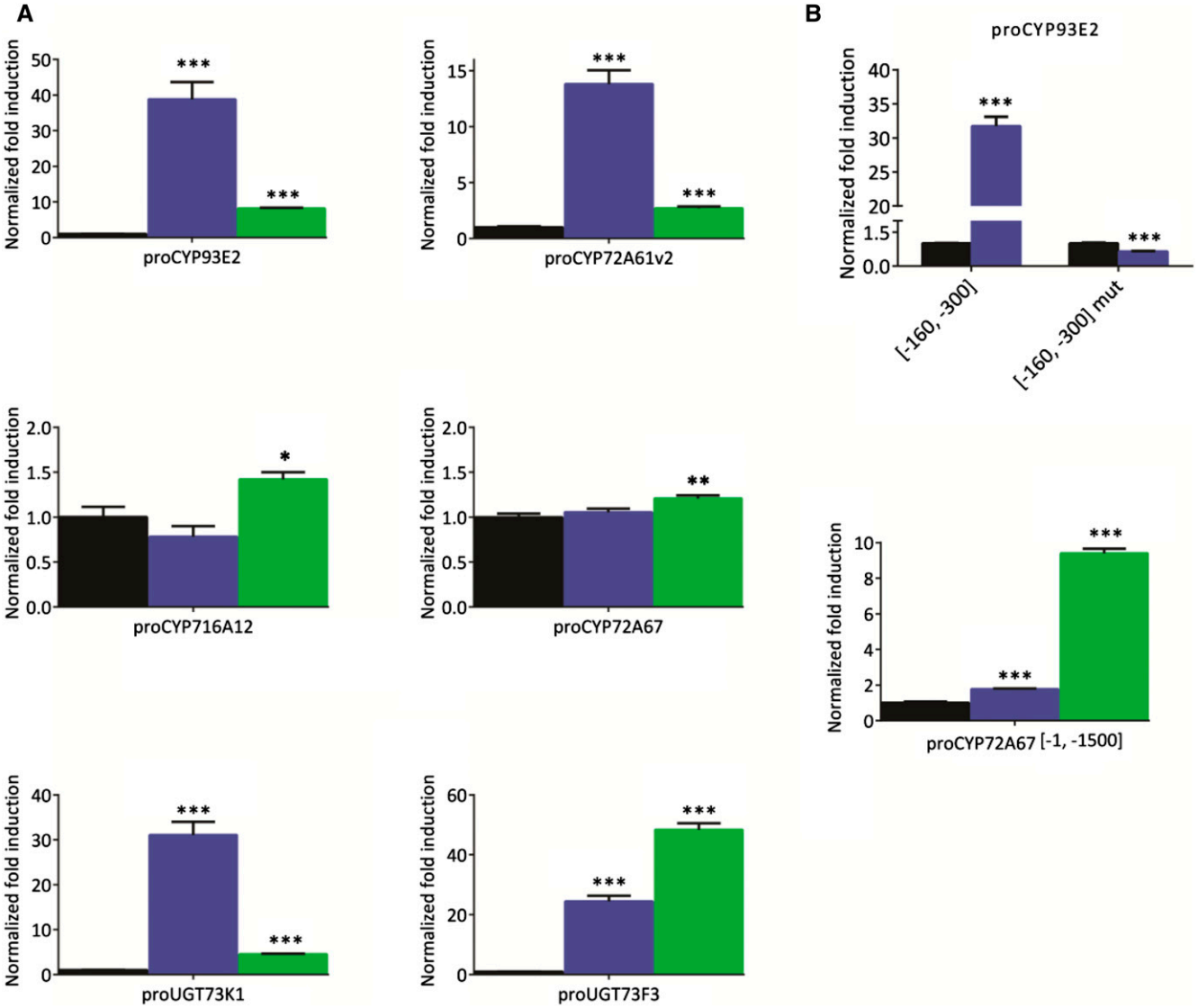


Figure 3. TSARs transactivate TS biosynthesis gene promoters. A, Transactivation of *ProCYP93E2*, *ProCYP72A61v2*, *ProCYP716A12*, *ProCYP72A67*, *ProUGT73K1*, and *ProUGT73F3* by TSAR1 (blue) and TSAR2 (green) in transfected tobacco protoplasts. Values on the *y* axis are normalized fold changes relative to protoplasts cotransfected with the reporter constructs and a pCaMV35S:GUS control plasmid (black). The error bars designate SE ($n = 4$). Statistical significance was determined by Student's *t* test (*, $P < 0.05$; **, $P < 0.01$; and ***, $P < 0.001$). All promoters comprise the 1,000-bp region upstream of their respective translational start sites. B, TSAR1 and TSAR2 depend on the N-boxes in *ProCYP93E2* and *ProCYP72A67[-1, -1500]*, respectively. Promoter fragments comprise the indicated regions relative to the start codon (in brackets). In the mutant (mut) version of *ProCYP93E2*, the two N-boxes (Supplemental Table S1) were substituted with TGAATT and CTATTA. Values on the *y* axis are normalized fold changes relative to protoplasts cotransfected with the reporter constructs and a pCaMV35S:GUS control plasmid (in black). The error bars designate SE ($n = 4$). Statistical significance was determined by Student's *t* test (***, $P < 0.001$).

albeit possibly with distinct specificities for the two branches.

Overexpression of *TSAR1* Increases the Biosynthesis of Soyasaponins in *M. truncatula* Hairy Roots

To assess the role and specificity of TSAR1 in planta, we generated three independent stably transformed *M. truncatula* hairy root lines overexpressing *TSAR1* ($TSAR1^{OE}$). For controls, we made three independent lines expressing the *GUS* gene. Quantitative reverse transcription-PCR (qPCR) analysis confirmed overexpression of *TSAR1* by approximately 3- to 5-fold (Fig. 4A). All three $TSAR1^{OE}$ lines exhibited significant but modest increases in *HMGR1* and *MKB1* transcripts, between 1.5- and 2-fold, and significant elevations of *BAS*, *CYP93E2*, *UGT73F3*, and *UGT73K1* transcript levels,

ranging from 3- to 8-fold (Fig. 4A). Finally, in two of the three TSAR1^{OE} lines, we observed modest increases in *TSAR2* and *CYP716A12* transcript levels, by approximately 1.75- and 3-fold, respectively (Fig. 4A). Overall, this indicates that *TSAR1* overexpression has a pathway-encompassing effect on the expression of TS genes.

To investigate the effect of *TSAR1* overexpression on the *M. truncatula* metabolome, we performed untargeted metabolite profiling of hairy root extracts by liquid chromatography-mass spectrometry (LC-MS). Five technical replicates of three TSAR1^{OE} and three control lines were profiled, yielding a total of 2,813 mass-to-charge ratio (m/z) peaks. To identify the peaks that are different between the control and TSAR1^{OE} lines, a partial least-squares discriminant analysis (PLS-DA) model that separates the TSAR1^{OE} and control hairy roots was generated (Fig. 4B). This PLS-DA model was subsequently used to generate an S-plot of the

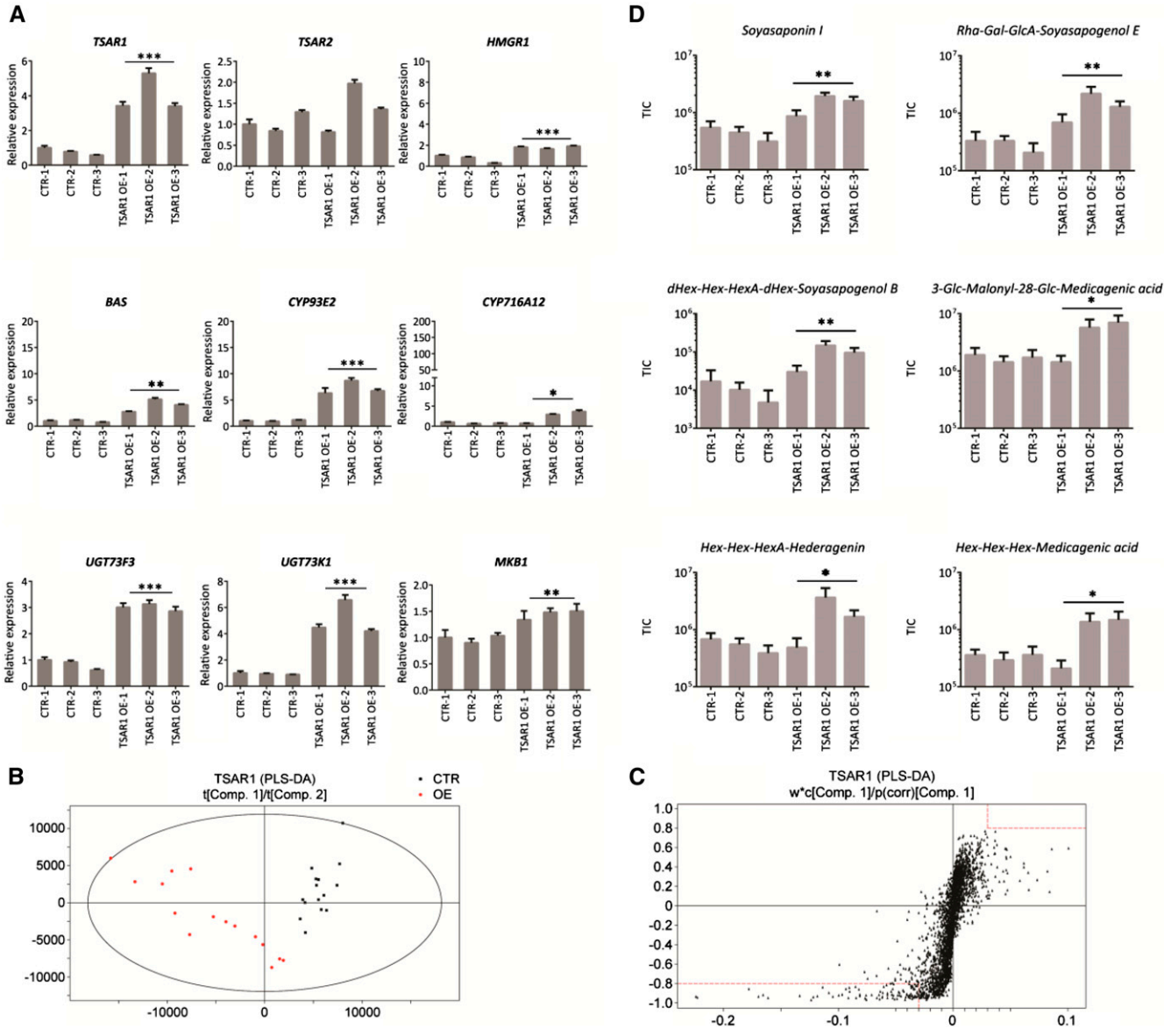


Figure 4. Overexpresssion of *TSAR1* boosts nonhemolytic TS biosynthesis in *M. truncatula* hairy roots. A, qPCR analysis of TS biosynthetic genes in three independent control (CTR) and three independent TSAR1^{OE} *M. truncatula* hairy root lines. Control lines were transformed with a *pCaMV35S:GUS* construct. Expression ratios were plotted relative to the normalized CTR-1. The error bars designate SE ($n = 3$). B, PLS-DA of samples from TSAR1^{OE} (red) and CTR (black) roots. C, S-plot for correlation [p(corr)] and covariance (w*c) derived from PLS-DA. Metabolites in the bottom left and top right quadrants (marked by dotted red lines) are significantly higher and lower, respectively, in abundance in the TSAR1^{OE} samples. D, Average total ion current (TIC) of peaks corresponding to TS. The error bars designate SE ($n = 5$). Statistical significance was calculated by Student's t test (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$).

correlation and covariance of all *m/z* peaks. Peaks with an absolute covariance value above 0.03 and an absolute correlation value above 0.8 were considered significantly different. As such, only peaks that were higher in the TSAR1^{OE} roots contributed to the observed differences (Fig. 4C). The metabolites corresponding to these peaks were elucidated based on their MS^n spectra, thereby revealing that *TSAR1* overexpression leads to higher levels of specific TSs (Fig. 4D; Supplemental Table S2). In particular, we observed significant elevations, by 3- to 9-fold relative to control lines, of nonhemolytic soyasaponins, such as soyasaponin I, Rha-Gal-GlcA-soyasapogenol E, and dHex-Hex-HexA-dHex-soyasapogenol B (Fig. 4D). Taken together, these data indicate that *TSAR1* overexpression has a strong effect on the accumulation of nonhemolytic TSs and support its role as a regulator of nonhemolytic TS biosynthesis.

In the qPCR analysis, we also observed an increase in *CYP716A12* expression in two of the three TSAR1^{OE} lines. CYP716A12 is the P450 that oxidizes β-amyrin at the C-28 position, thereby directing TS biosynthesis toward the hemolytic TS (Carelli et al., 2011). Hence, we probed the effect of TSAR1^{OE} on the accumulation of hemolytic TSs, and in accordance with the qPCR analysis, we observed an increase of certain hemolytic TSs only in the two TSAR1^{OE} lines with a higher expression of *CYP716A12* (Fig. 4D; Supplemental Table S2).

Overexpression of *TSAR2* Boosts the Biosynthesis of Hemolytic TSs in *M. truncatula* Hairy Roots

To evaluate TSAR2 function, we generated three independent *M. truncatula TSAR2* overexpression hairy root lines (TSAR2^{OE}). The extent of *TSAR2* overexpression was variable across the three independent lines, ranging from 5- to 23-fold, but effectuated in all cases a clear rise in *HMGR1*, *BAS*, and *UGT73F3* transcript levels, ranging from 2- to 18-fold (Fig. 5A). Like TSAR1^{OE} roots, TSAR2^{OE} roots exhibited a significant but modest increase in *MKB1* transcript levels, between 1.5- and 2-fold (Fig. 5A). In contrast to the TSAR1^{OE} lines, however, the transcript levels of *CYP93E2* and *UGT73K1* were not altered. Furthermore, the transcript levels of *CYP716A12* increased spectacularly, with more than 150-fold in the strongest TSAR2^{OE} line, pointing to a specificity of TSAR2 in the regulation of the hemolytic TSs.

To substantiate this, we investigated the changes on the metabolome of TSAR2^{OE} lines by LC-MS, which yielded a total of 2,993 *m/z* peaks. As for the TSAR1^{OE} lines, a PLS-DA model separating the TSAR2^{OE} and control lines was generated (Fig. 5B) and used to create an S-plot and depict the significantly different peaks (Fig. 5C). Analogous to the TSAR1^{OE} lines, only peaks that were higher in the TSAR2^{OE} roots contributed to the observed differences from the control roots. Identification of the corresponding metabolites revealed that hemolytic TSs overaccumulated in the TSAR2^{OE} lines (Fig. 5D; Supplemental Table S3). This was exemplified by over 10-fold increases of 3-Glc-malonyl-28-Glc-medicagenic acid, Hex-Hex-HexA-hederagenin, and Hex-Hex-Hex-medicagenic acid (Fig. 5D).

The qPCR analysis of the TSAR2^{OE} lines indicated a specific effect of TSAR2 on the hemolytic TSs, as the expression of *CYP93E2* remained unaltered (Fig. 5A). Accordingly, no soyasaponins were labeled as accumulating significantly differently in the TSAR2^{OE} lines (Supplemental Table S3). To ascertain this, we verified the accumulation pattern of the soyasaponins that were most altered in the TSAR1^{OE} lines (Fig. 4D; Supplemental Table S3). This confirmed that the effect of *TSAR2* overexpression is indeed restricted to the hemolytic TSs (Fig. 5D) and supports its role as a regulator of hemolytic TS biosynthesis. The specificity of TSAR2 in the regulation of hemolytic TSs could also explain the observed small increases in the production of hemolytic TSs in two of the three TSAR1^{OE} lines, which were correlated with increased *TSAR2* expression levels. In this regard, it is noteworthy that none of the TSAR1^{OE} lines actually showed this increased expression of *TSAR2* and *CYP716A12* nor any increase in the accumulation of hemolytic TSs early after the generation of the root lines (Supplemental Fig. S6). Only following the repeated subculturing of the TSAR1^{OE} lines, necessary for the upscaling for the metabolome and transcriptome profiling (see below), two of the three TSAR1^{OE} lines started to show modest increases in the expression of these two genes and, accordingly, in the accumulation of hemolytic TSs. Therefore, we consider that TSAR1 and TSAR2 primarily and specifically activate nonhemolytic and hemolytic TS biosynthesis, respectively. Long-term culturing of TSAR1^{OE} lines may eventually cause feedback leading to the induction of *TSAR2* expression and consequent hemolytic TS biosynthesis. Such feedback may be caused by TS pathway intermediates and/or end products and may be reminiscent of the feedback repression on the TS pathway genes in the *MKB1*^{KD} line that overaccumulates monoglycosylated TS (Pollier et al., 2013a).

TSAR1 Knockdown Results in Decreased TS Gene Expression

Examining the *TSAR* expression levels by mining MtGEA and in-house-generated RNA sequencing (RNA-Seq) data (Table II; Supplemental Tables S4 and S5), we noticed that *TSAR1* exhibited severalfold higher expression levels than *TSAR2* in (hairy) roots under standard (i.e. nonstressed) culturing conditions. Therefore, assessing the physiological role in planta of the TSAR TFs through a gene-silencing approach seemed most practical for *TSAR1*. Indeed, we successfully managed to create three independent *TSAR1* knockdown lines (TSAR1^{KD}), with less than 25% of the wild-type *TSAR1* transcript levels remaining (Fig. 6A). In contrast, we did not manage to generate *TSAR2* knockdown lines, despite repeated transformation rounds; hence, we focused further on the TSAR1^{KD} lines. Across

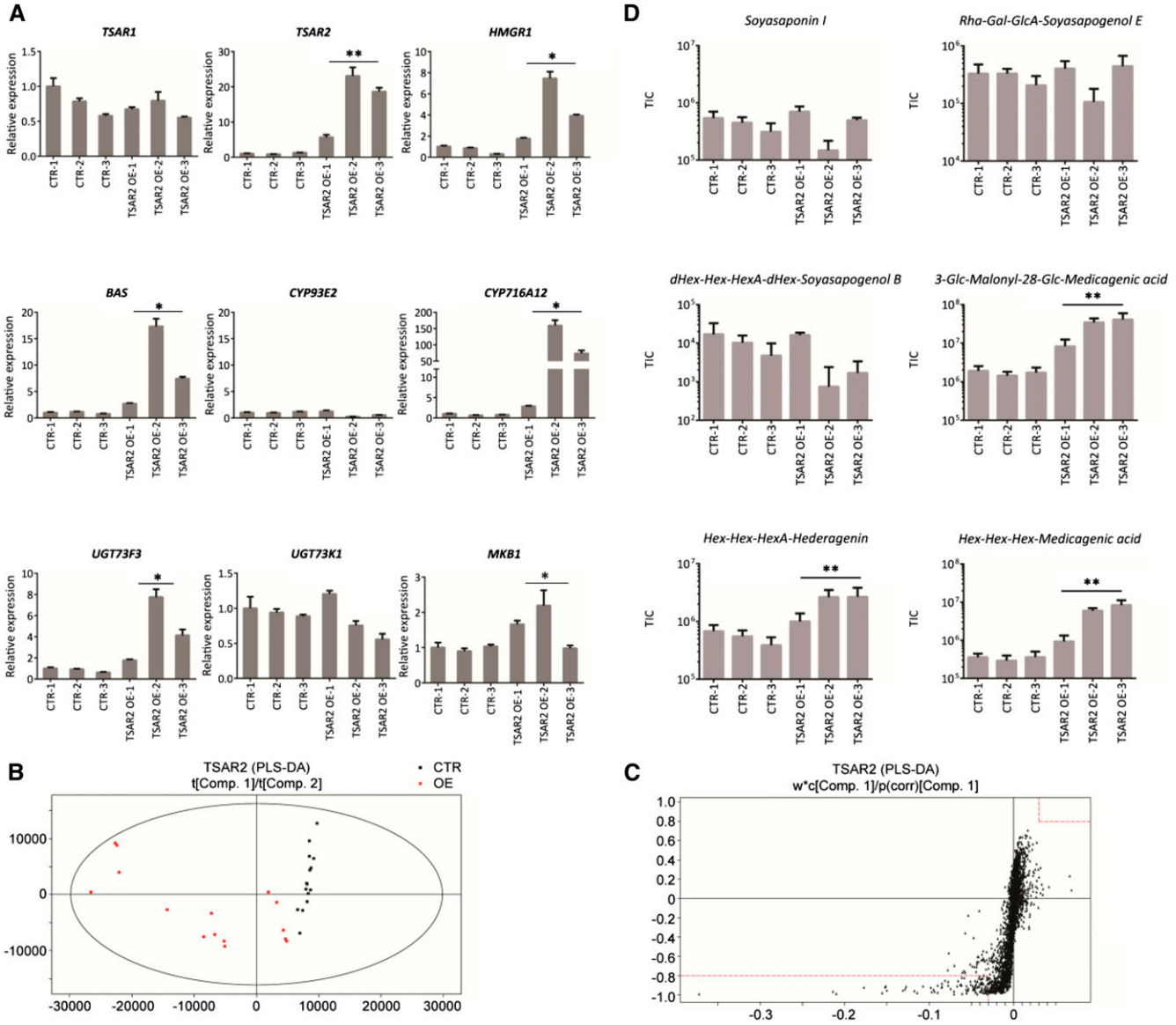


Figure 5. Overexpresssion of *TSAR2* boosts hemolytic TS biosynthesis in *M. truncatula* hairy roots. A, qPCR analysis of TS biosynthetic genes in three independent control (CTR) and three independent TSAR2^{OE} *M. truncatula* hairy root lines. Expression ratios were plotted relative to the normalized CTR-1. The error bars designate SE ($n = 3$). B, PLS-DA of samples from TSAR2^{OE} (red) and CTR (black) roots. C, S-plot for correlation [p(corr)] and covariance (w*c) derived from PLS-DA. Metabolites in the bottom left and top right quadrants (marked by dotted red lines) are significantly higher and lower, respectively, in abundance in the TSAR2^{OE} samples. D, Average total ion current (TIC) of peaks corresponding to TS. The error bars designate SE ($n = 5$). Statistical significance was calculated by Student's *t* test (*, $P < 0.05$ and **, $P < 0.01$).

those three TSAR1^{KD} lines, we observed a significant decrease in *MKB1*, *HMGR1*, *CYP93E2*, *UGT73F3*, and *UGT73K1* transcript levels, varying between 20% and 70% of the control transcript levels (Fig. 6A). No decrease in *BAS* and *CYP716A12* expression levels was apparent in the TSAR1^{KD} lines. Together, these findings support the importance of TSAR1 for expression of the genes involved in the biosynthesis of nonhemolytic TSs.

As for the *TSAR* overexpression lines, we conducted metabolite profiling of the TSAR1^{KD} root lines by LC-MS (Fig. 6, B–D). Using the same analytical criteria, we could not detect significant changes in TS accumulation (Fig. 6D); hence, the decrease in TS synthesis transcripts did not lead to significantly decreased TS metabolite levels. The PLS-DA did not indicate increased abundance of any metabolite but did reveal a decreased abundance of 137 m/z peaks (Fig. 6C), although the fold decrease was modest (less than 1.5-fold). The mass data and fragmentation patterns from the Fourier transform-mass spectrometry analysis did not allow the

Table II. *Fragments per kilobase of exon per million fragments mapped (FPKM) values of triterpene genes in transformed M. truncatula hairy roots*

Log base 2 values of fold changes in boldface designate genes that show significant differential expression. AVG, Average of the three independent control and *TSAR*-overexpressing lines.

Gene Identifier	Name	Control	TSAR1^{OE}		TSAR2^{OE}	
		AVG FPKM	AVG FPKM	Log_2 (Fold Change)	AVG FPKM	Log_2 (Fold Change)
Medtr7g080780	TSAR1	27.84	143.41	**2.30**	27.06	−0.29
Medtr4g066460	TSAR2	4.89	7.80	0.64	89.70	**4.21**
Medtr5g026500	HMGR1	6.69	27.02	**2.11**	56.89	**3.23**
Medtr4g005190	BAS	6.80	32.15	**2.32**	67.44	**3.44**
Medtr7g056103	CYP93E2	67.24	571.25	**3.13**	52.73	−0.29
Medtr8g100135	CYP716A12	5.41	17.18	**1.63**	580.16	**6.75**
Medtr2g035020	UGT73F3	17.06	94.34	**2.59**	143.81	**3.23**
Medtr4g031800	UGT73K1	54.46	359.68	**2.84**	49.75	0.04
Medtr2g023680	CYP72A67	2.21	6.20	**1.48**	197.26	**6.50**
Medtr2g055470	CYP72A68v2	3.58	9.37	**1.44**	193.89	**5.81**
Medtr4g031820	CYP72A61v2	32.30	178.60	**2.50**	34.70	0.17
Medtr5g098310	Acetyl-CoA acetyltransferase	37.10	44.87	0.37	68.61	**1.03**
Medtr5g011040	Hydroxymethylglutaryl-CoA synthase	42.71	72.86	**0.79**	133.29	**1.68**
Medtr7g113660	MVA kinase	14.04	16.99	0.32	21.86	**0.71**
Medtr3g091190	Phosphomevalonate kinase	8.51	12.86	**0.63**	34.19	**2.08**
Medtr1g112230	MVA diphosphate decarboxylase	45.39	64.45	**0.61**	102.97	**1.33**
Medtr7g080060	Isopentenyl diphosphate Δ-isomerase	104.87	222.62	**1.14**	268.16	**1.44**
Medtr2g027300	Farnesyl pyrophosphate synthase	50.93	87.97	**0.81**	212.82	**2.12**
Medtr4g071520	Squalene synthase	62.26	107.75	**0.85**	276.92	**2.25**
Medtr1g017270	Squalene epoxidase	71.87	123.42	**0.87**	406.71	**2.63**
Medtr5g008810	Cycloartenol synthase (CAS)	36.45	36.84	0.09	31.27	−0.10
Medtr3g032530	C-24 methyltransferase (C24MT)	91.87	104.74	0.22	89.97	0.05
Medtr8g006450	C-14 demethylase (CYP51G1)	45.32	46.65	0.01	55.40	0.31
Medtr1g061240	C-14 reductase (Fackel)	8.80	9.04	0.00	9.05	0.03
Medtr6g084920	C-8,7 isomerase (Hydra1)	18.40	19.39	0.03	17.57	−0.06
Medtr3g114780	C-24 methyltransferase (CVP1)	88.83	92.57	0.17	90.98	0.18
Medtr2g019640	C-22 desaturase (CYP710A15)	19.26	19.68	0.13	18.35	0.07
Medtr5g070090	UGT71G1	5.08	3.93	−0.38	4.25	−0.24

identification or tentative annotation of the corresponding metabolites.

RNA-Seq Analyses Confirm and Reveal Downstream Targets of TSAR1 and TSAR2

To assess the specificity and range of TSAR1 and TSAR2, we performed a genome-wide transcript profiling study by RNA-Seq. To this end, the three independent *M. truncatula* TSAR1^{OE}, TSAR2^{OE}, and control root lines were subjected to RNA-Seq using the Illumina HiSeq 2500 platform. A total of 353,231,400 single-end reads of 50 bp were obtained (Supplemental Table S4) and mapped on the *M. truncatula* genome version 4.0 (Tang et al., 2014). Genes with significant differential expression levels between the control lines and the TSAR1^{OE} or TSAR2^{OE} lines were selected using the Cuffdiff algorithm (Trapnell et al., 2010; Table II; Supplemental Table S5). Respectively, 859 and 845 genes showed differential expression levels in TSAR1^{OE} and TSAR2^{OE} roots. Between these two gene pools, there was an overlap of 356 genes (Supplemental Fig. S7; Supplemental Table S5).

First, the strong overexpression of *TSAR1* and *TSAR2* in their respective OE lines was confirmed. Among the differentially expressed genes, we next checked for the known or suggested TS biosynthesis pathway genes (Supplemental Fig. S1). In accordance with our qPCR analysis, we found that *TSAR1* overexpression strongly enhanced the expression of *BAS*, *CYP93E2*, *UGT73F3*, and *UGT73K1*, while *TSAR2* overexpression instigated increments of *BAS*, *CYP716A12*, and *UGT73F3* transcripts (Fig. 7; Supplemental Table S5). In addition, the increased expression of *CYP716A12* and *TSAR2* in two of the three TSAR1^{OE} lines was confirmed. In these and in the TSAR2^{OE} lines, this was accompanied by increased transcript levels of the *CYP72A68v2* and *CYP72A67* genes. CYP72A68v2 catalyzes the C-23 oxidation of the TS backbone in the synthesis of hemolytic TSs. *CYP72A67* was recently shown to hydroxylate C-2 in the hemolytic TS biosynthesis branch (Fukushima et al., 2013; Biazzi et al., 2015). In TSAR1^{OE} lines, we observed a strong increase in *CYP72A61v2* transcripts encoding the P450 that catalyzes the C-22 oxidation of the TS backbone in the synthesis of soyasaponins. These data further corroborate the specific effects of the TSAR1 and TSAR2 TFs on the soyasaponin and hemolytic branches of the TS biosynthesis pathway, respectively.

To verify TSAR specificity for TS biosynthesis and to examine whether they could also stimulate precursor pathways in an encompassing way, we looked at whether genes encoding enzymes involved in the MVA

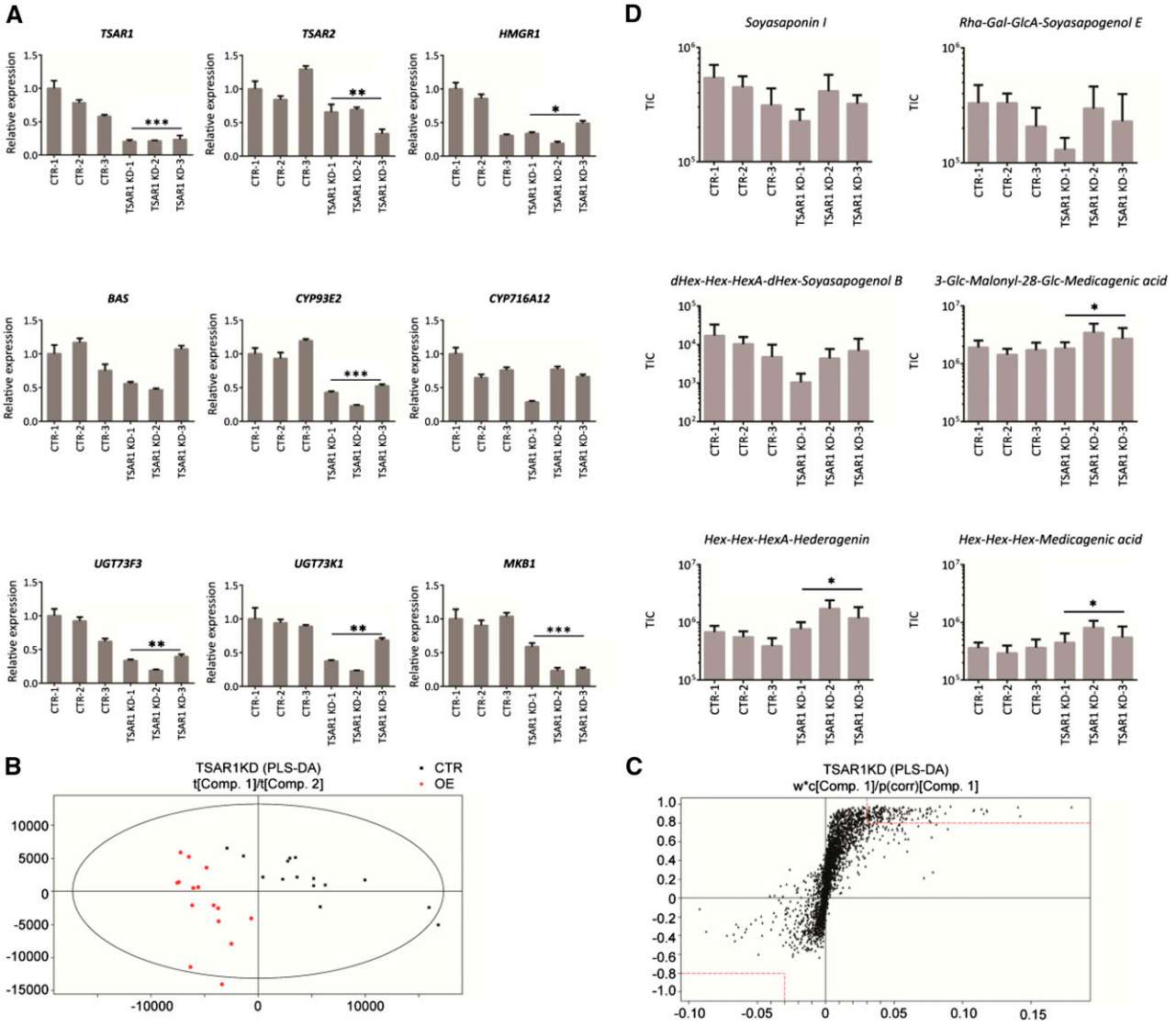


Figure 6. Knockdown of *TSAR1* lowers TS biosynthesis gene expression. A, qPCR analysis of TS biosynthetic genes in three independent control (CTR) and TSAR1^{KD} *M. truncatula* hairy root lines. Expression ratios were plotted relative to the normalized CTR-1. The error bars designate SE ($n = 3$). B, PLS-DA of samples from TSAR1^{KD} (red) and CTR (black) roots. C, S-plot for correlation [p(corr)] and covariance (w*c) derived from PLS-DA. Metabolites in the bottom left and top right quadrants (marked by dotted red lines) are significantly higher and lower in abundance in the TSAR1^{KD} samples, respectively. D, Average total ion current (TIC) of peaks corresponding to TS. The error bars designate SE ($n = 5$). Statistical significance was calculated by Student's *t* test (*, $P < 0.05$; **, $P < 0.01$; and ***, $P < 0.001$).

and sterol pathways were among the differentially expressed genes in our analysis. We retrieved the putative MVA and sterol pathway synthesis genes from the *Medicago truncatula* Pathway Database 2.0 (http://mediccyc.noble.org/; Urbanczyk-Wochniak and Sumner, 2007). When more than one homolog existed, we identified the closest homolog of the corresponding pathway genes from Arabidopsis. Nearly all MVA pathway genes were significantly up-regulated in both the TSAR1^{OE} and TSAR2^{OE} roots, albeit usually less pronounced than the TS-specific genes (Fig. 7; Table II). The strongest up-regulation was observed for *HMGR1*, the key rate-limiting enzyme of the pathway. None of the sterol synthesis genes were affected by TSAR1 or TSAR2 (Table II).

A few genes annotated as encoding enzymes involved in flavonoid biosynthesis, such as isoflavone synthases and chalcone synthases, appeared differentially expressed, but the FPKM values were not consistent for the three lines per construct (Supplemental Table S5). No other genes potentially involved in the synthesis of phenolic compounds were induced in any of the TSAR^{OE} lines (Supplemental Table S5). Hence, no concerted effect of *TSAR* overexpression on flavonoid biosynthetic gene expression was observed, further underscoring the specific activity of TSAR1 and TSAR2 for TS biosynthesis

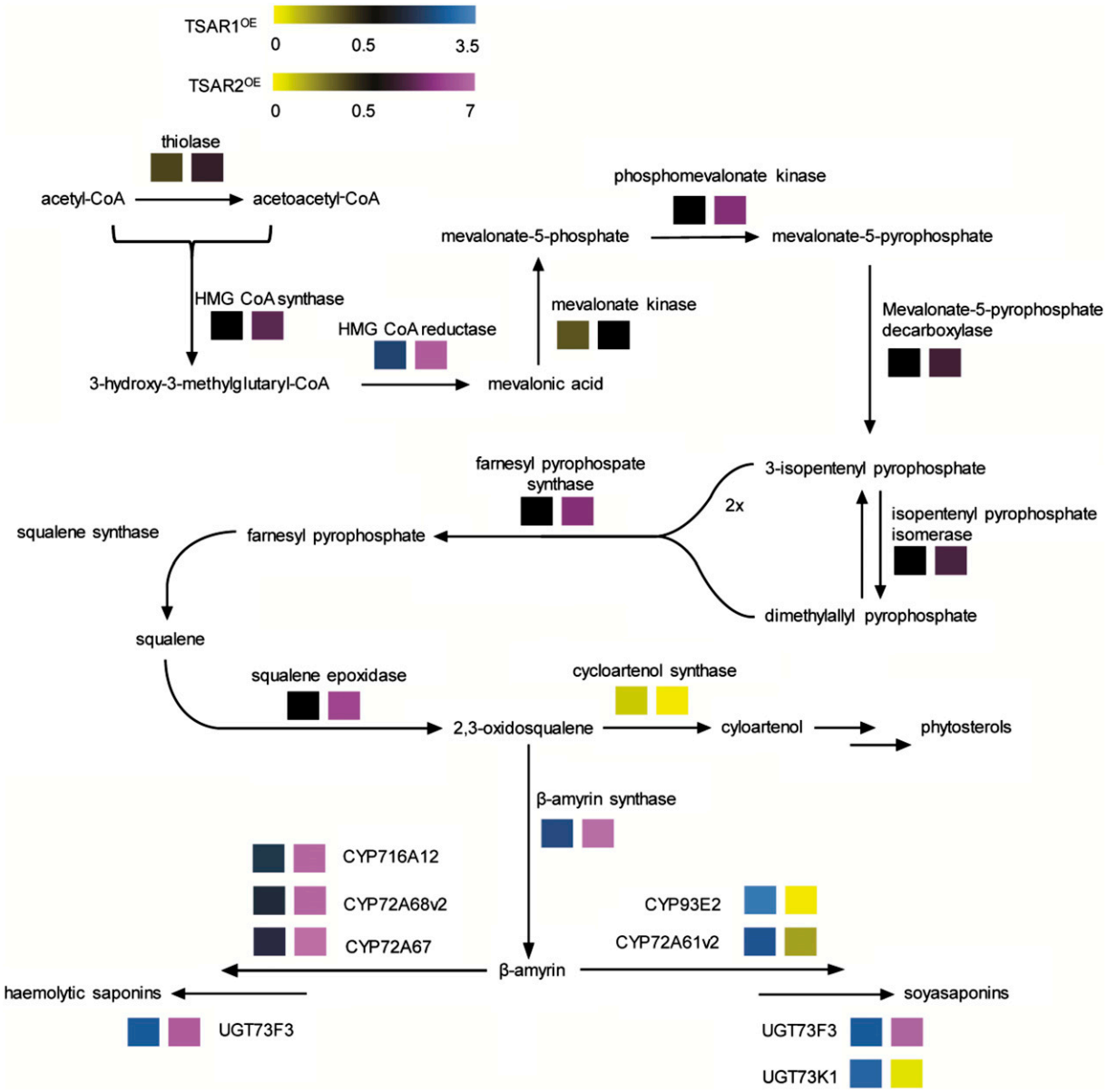


Figure 7. *TSAR1* and *TSAR2* overexpression specifically modulates the TS biosynthesis pathway in *M. truncatula*. Log base 2 values of the fold up-regulation of biosynthetic genes in TSAR1^{OE} and TSAR2^{OE} roots relative to the control roots were calculated using Cufflinks (Trapnell et al., 2010) and are depicted by color codes (left and right for TSAR1 and TSAR2, respectively).

within *M. truncatula* specialized metabolism. Notably, overexpression of *TSAR1*, but not of *TSAR2*, also stimulates the expression of genes encoding 9*S*-lipoxygenases, leginsulins (also annotated as albumins), and Kunitz-type protease inhibitors (Supplemental Table S5), suggesting that additional roles, beyond the regulation of specialized metabolism, may exist for TSAR1.

DISCUSSION

Upon predation, plants produce a plethora of specialized metabolites to safeguard their integrity and survival. The phytohormone JA plays a pivotal role therein. TFs at the top of the JA signaling hierarchy, such as MYC2, have been studied extensively in several plant species. However, less is known about the downstream TFs that boost the flow through specific biosynthesis pipelines in plant specialized metabolism.

Members of Clade IVa of the bHLH Family Activate Different Branches of TS Synthesis in *M. truncatula*

Limited data are available about TFs implicated in the activation of specialized terpene metabolite production. In cotton (*Gossypium arboreum*), a WRKY TF

has been shown to directly bind the promoter of a sesquiterpene synthase-encoding gene essential for the biosynthesis of sesquiterpene phytoalexins (Xu et al., 2004). In *A. annua*, two JA-responsive AP2 family TFs and a MYC-type bHLH TF have been identified as activators of the expression of genes encoding the sesquiterpene synthase and P450 essential for the production of the antimalaria compound artemisinin (Yu et al., 2012; Ji et al., 2014). In Arabidopsis, it has been observed that the bHLH TF MYC2 directly binds sesquiterpene synthase gene promoters, resulting in an elevated release of volatile sesquiterpenes (Hong et al., 2012). Likewise, in tomato, a MYC-type bHLH and a WRKY TF have been described to bind the promoter of a sesquiterpene synthase gene (Spyropoulou et al., 2014). Finally, the bHLH TFs Bl and Bt from cucumber and BIS1 from *C. roseus* were recently identified as activators of the production of cucurbitane-type triterpenes and iridoid-type monoterpenes, respectively (Shang et al., 2014; Van Moerkercke et al., 2015). Here, we identified two other bHLH TFs, TSAR1 and TSAR2, that activate TS biosynthesis in the model legume *M. truncatula*.

Notably, TSAR1 and TSAR2 have distinct preferences in steering fluxes to the two branches of the TS biosynthesis pathway, implying the existence of distinct control mechanisms for the different classes of TSs that accumulate in *M. truncatula* and that may have distinct biological functions or activities. TSAR1 primarily drives the expression of *BAS* and all genes encoding the known nonhemolytic or soyasaponin-specific P450s, such as *CYP93E2* and *CYP72A61v2*, as well as the UGTs *UGT73K1* and *UGT73F3*. Upon overexpression in hairy roots, this led to a clear rise in nonhemolytic TS accumulation. Conversely, TSAR2 drives the expression of the *BAS*, *CYP716A12*, *CYP72A68v2*, and *UGT73F3* genes, which upon overexpression in hairy roots led to major increases in hemolytic TS levels. Nonhemolytic TS metabolism remained unaffected in TSAR2^{OE} lines. Notably, *TSAR2* overexpression did not mediate changes in *UGT73K1* levels, suggesting that despite their ability to glucosylate sapogenin skeletons of both TS types in vitro (Achnine et al., 2005), they might only glucosylate nonhemolytic sapogenins in planta. *UGT71G1* was not induced by either *TSAR1* or *TSAR2* overexpression, suggesting that the corresponding UGT might not be implicated in TS biosynthesis and that its activity might be restricted to the glucosylation of flavonoids, for which the recombinant UGT71G1 possesses high catalytic activity (Achnine et al., 2005). Consequently, we believe that further mining of the RNA-Seq data from *TSAR1* and *TSAR2* overexpression lines will lead to the discovery of the missing enzymes in *M. truncatula* TS biosynthesis.

Despite the observation that also the expression of the MVA precursor pathway genes is controlled by both TSAR TFs, their specific commitment to the TS pathway was demonstrated by the lack of impact on the sterol biosynthesis pathway. In this regard, the TSAR TFs have a similar pathway-encompassing effect to the MKB1 E3 ubiquitin ligase (Pollier et al., 2013a), and it is striking that the TSARs also drive *MKB1* expression. However, contrary to *M. truncatula* hairy root lines with a loss of function of MKB1, in which the overaccumulation of monoglycosylated TSs is associated with perturbed root development and integrity (Pollier et al., 2013a), the pronounced increases in the accumulation of TSs did not result in any phenotypical change in any of the *TSAR* overexpression lines (Supplemental Fig. S8). Likewise, although we observed significantly lower *MKB1* transcript levels in the TSAR1^{KD} roots, this did not result in any phenotypical change, and the TS levels remained unaltered. Considering that also mutants defective in *CYP716A12* and *UGT73F3* suffer from growth retardation (Naoumkina et al., 2010; Carelli et al., 2011), it seems important for normal plant growth and development that the flux through the TS pathway proceeds to the multihydroxylated and multiglycosylated end products. Strict and concerted regulation of TS biosynthetic genes by the TSARs may provide the necessary safety mechanism to achieve this.

TSAR1 and TSAR2 Are Integrated in a JA Signaling Cascade That Steers Defense Responses

The bHLHs constitute a major family of TFs and are widely spread across the three eukaryotic kingdoms (Heim et al., 2003; Toledo-Ortiz et al., 2003; Carretero-Paulet et al., 2010; Pires and Dolan, 2010). Notably, TSAR1 and TSAR2 belong to subclade IVa of the bHLH family, distinguishing them from the MYC2-type TFs that reside in subclade IIIe and the cucumber Bl and Bt TFs that sort in subclade Ib (Heim et al., 2003). Besides divergent bHLH domains, clade IVa bHLHs constitute an ORF that is about half the size of MYC-type TFs. In addition, MYC-type TFs carry a defined JAZ interaction domain that is lacking in clade IVa bHLHs. The JAZ interaction domain is responsible for the interaction of MYCs with Jasmonate ZIM Domain (JAZ) repressor proteins that thereby block the activity of the MYCs and the transactivation of their target genes. Upon JA perception, JAZ proteins are targeted for degradation by the 26S proteasome (Chini et al., 2007; Thines et al., 2007; Fernández-Calvo et al., 2011; Pauwels and Goossens, 2011). Like the *MYC* genes, *TSAR1* and *TSAR2* also are JA inducible. Hence, they seem to constitute an integrated element in the JA signaling cascade. The determination of their exact positions in this cascade, relative to the primary signaling module composed of COI1-JAZ-MYC-NINJA (Cuéllar Pérez and Goossens, 2013; Wasternack and Hause, 2013), will be the subject of further study.

JAs trigger a signaling cascade that mediates broad-scale defense responses resulting in a strictly regulated production of various defense products in response to various biotic and abiotic stresses (Wasternack and Hause, 2013). Our data here suggest that TSAR2

specifically drives hemolytic TS production, since only those specialized metabolites were induced in the TSAR2^{OE} lines and RNA-Seq analysis did not readily indicate any concerted modulation of other defense or developmental processes. TSAR1, in contrast, clearly also stimulates the production of at least two other types of defense molecules besides soyasaponin TS, namely leginsulins and Kunitz-type protease inhibitors (Supplemental Table S5). Likewise, TSAR1^{OE} roots, but not TSAR2^{OE} roots, showed a strong elevation of several linoleate 9*S*-lipoxygenase transcripts (Supplemental Table S5). Mining of public expression data in MtGEA indicated that most of the above-mentioned genes are JA inducible, as are the TS biosynthesis genes. Leginsulins (also annotated as albumins) are Cys-rich peptides that are prevalent within the plant family Fabaceae (Louis et al., 2004). These 35- to 40-kD peptides contain three disulfide bridges that render them highly resistant to degradation during digestion (Le Gall et al., 2005). In addition, they exhibit antiinsect and hormonal functions in plants and disturb blood Glc levels in mice (Watanabe et al., 1994; Dun et al., 2007). Kunitz-type protease inhibitors are widely distributed across the plant kingdom (Oliva et al., 2011). Frequently, they consist of one polypeptide chain that harbors two disulfide bridges. Among others, they inhibit proteases that activate digestive enzymes, leading to the perturbation of digestion, an impaired uptake of amino acids, and reduced growth, which is considered to be particularly important as a defense against insect feeding (Howe and Jander, 2008; Oliva et al., 2010). Finally, 9-lipoxygenases are enzymes at the base of the committed biosynthesis of specific oxylipins that act in defense against microbial pathogens (Vicente et al., 2012).

Conserved Subclade IVa bHLH TFs Regulate Plant Terpene Biosynthesis

In Arabidopsis, it has been shown that homologous TFs can instigate different effects on JA-responsive genes and cooperate with other TFs, many of which are still unknown, to set off the full complement of JA responses (Fernández-Calvo et al., 2011; De Geyter et al., 2012).

In plants, it is generally assumed that IPP used for the generation of triterpenes and sesquiterpenes is derived from the MVA pathway, whereas the synthesis of other terpenes consumes IPP from the plastidial 2-*C*-methyl-D-erythritol 4-phosphate (MEP) pathway (Moses et al., 2013). Overexpression of the *C. roseus* bHLH TF *BIS1* was recently shown to specifically mediate the elevated expression of the biosynthesis genes required to yield the MEP pathway-dependent monoterpene (iridoid) loganic acid in cell suspension cultures and hairy roots (Van Moerkercke et al., 2015). Interestingly, *C. roseus* is also a source of nonglycosylated pentacyclic triterpenes, including oleanolic acid, but overexpression of *BIS1* does not affect the expression of MVA or triterpene pathway genes (Van Moerkercke et al., 2015). BIS1, like TSAR1 and TSAR2, belongs to subclade IVa of bHLH proteins (Supplemental Figs. S4 and S5). Remarkably, however, in *M. truncatula*, the TSARs drive MVA and TS pathway gene expression without altering MEP pathway gene expression (Table I; Supplemental Table S5). In *C. roseus*, the genes required for MIA production are induced by JAs, but this is not case for the MVA and TS pathway genes (Van Moerkercke et al., 2013). This differs from the situation in *M. truncatula*, where MVA pathway and TS synthesis genes are JA responsive (Broeckling et al., 2005; Suzuki et al., 2005; Pollier et al., 2013a).

C. roseus and *M. truncatula* belong to the Apocynaceae and Fabaceae, respectively. Both are dicot plant families, but they are representative of the two different clades within that group, the asterids and rosids, respectively. Hence, our results show that clade IVa bHLH TFs in two distantly related dicot species exert a similar function in regulating terpene biosynthesis but are capable of acting on different classes of terpenes depending on the species. Moreover, they act not only in the species-specific specialized metabolite pathway branches but also in the respective primary precursor pathways (i.e. the MEP pathway for MIA synthesis in *C. roseus* and the MVA pathway for TS biosynthesis in *M. truncatula*). Like nearly all genes encoding the enzymes involved in all the above-mentioned pathways, the genes encoding clade IVa bHLH TFs are JA responsive themselves. This indicates that they are essential elements in the universal capacity of JA to elicit specific specialized terpene pathways across the plant kingdom.

MATERIALS AND METHODS

DNA Constructs

Sequences of the full-length ORFs of all TFs were retrieved from the *Medicago truncatula* genome version 4.0 (Tang et al., 2014) and were cloned using Gateway technology (Invitrogen). Full-length coding sequences were PCR amplified (for primers, see Supplemental Table S6) and recombined in the donor vector pDONR221. Following sequence verification, the entry clones were recombined with the destination vector p2GW7 for protoplast assays (Vanden Bossche et al., 2013). For overexpression in *M. truncatula* hairy roots, TSAR1 and TSAR2 entry clones were recombined with the binary overexpression vector pK7WG2D (Karimi et al., 2002). A construct to silence *TSAR1* through hairpin RNA interference was generated by amplifying a 231-bp fragment from the 3′ UTR from the *TSAR1* mRNA. After insertion into pDONR221, the construct was recombined into the binary vector pK7GWIWG2D(II) (Karimi et al., 2002).

The promoter regions of *HMGR1*, *CYP93E2*, *CYP716A12*, *CYP72A61v2*, *CYP72A67*, *UGT73K1*, and *UGT73F3* were determined using the *M. truncatula* genome version 4.0 (Tang et al., 2014). This version of the genome did not contain *MKB1*, but we were able to trace *MKB1* and its promoter in another version of the *M. truncatula* genome, namely Mt20120830-LIPM (Roux et al., 2014). The 1,000-bp regions upstream of the respective translational start sites were PCR amplified (Supplemental Table S1), except for *MKB1*, for which 1,000 bp upstream of the transcriptional start was amplified (Supplemental Table S1). In addition, a longer fragment of 1,500 bp, upstream of the ORF of *CYP72A67*, was generated (Supplemental Table S1). Likewise, we amplified shorter fragments of the *HMGR1* promoter. A promoter fragment in which the hexanucleotide CACGAG was substituted by TGAATT was generated by overlap extension PCR. A shorter fragment was also generated for *ProCYP93E2*. A fragment, of which the two N-boxes were substituted with 5′-TGAATT-3′ and 5′-CTATTA-3′, was constructed by overlap extension PCR. All promoter sequences were successively recombined into pDONR221, and sequence-verified entry clones were recombined with the pGWL7 plasmid to generate promoter:fLUC reporter constructs (Vanden Bossche et al., 2013). A Y1H bait fragment

was generated by overlap extension PCR. Three identical (CACGAG) motifs with their 10 flanking nucleotides from the *HMGR1* promoter were fused using two linker sequences (Supplemental Fig. S9). This construct was cloned in the reporter plasmid pMW#2 (Deplancke et al., 2006).

Transient Expression Assays in Tobacco Protoplasts

Transient expression assays in tobacco (*Nicotiana tabacum* 'Bright Yellow-2') protoplasts were carried out as described previously (De Sutter et al., 2005; Vanden Bossche et al., 2013). Briefly, protoplasts were transfected with a reporter, an effector, and a normalizer plasmid. The reporter plasmid consists of a fusion between the promoter fragment of interest and the *fLUC* gene. The effector plasmid contains the selected TF driven by the CaMV 35S promoter. The normalizer plasmid, harboring the *RENILLA LUCIFERASE* (*rLUC*), is under the control of the CaMV 35S promoter. Protoplasts were incubated overnight and lysed. fLUC and rLUC readouts were collected using the Dual-Luciferase Reporter Assay System (Promega). Each assay incorporated eight or four biological repeats. Promoter activities were normalized by dividing the fLUC values by the corresponding rLUC values. The average of the normalized fLUC values was calculated and set out relative to the control fLUC values (i.e. protoplasts transfected with an effector plasmid carrying a *GUS* gene).

Phylogenetic Analysis

The bHLH domain amino acid sequences of Arabidopsis (*Arabidopsis thaliana*), cucumber (*Cucumis sativus*), and *M. truncatula* were defined as by Heim et al. (2003). All members of Arabidopsis from subclades I, III, and IV were used for the assembly. We incorporated *M. truncatula* TSAR1 and TSAR2, *Catharanthus roseus* BIS1, and cucumber Bl and Bt. All sequences were aligned with the ClustalW tool of BioEdit7. Gaps, which are prevalent in the loop region, were dealt with by complete deletion of the corresponding sites. A neighbor-joining tree was constructed in MEGA5 using the Jones, Taylor, and Thornton amino acid substitution model (Jones et al., 1992; Tamura et al., 2011). A bootstrap analysis was carried out with 1,000 replicates, and an unrooted tree was generated.

Protein-Binding Microarrays

N-terminal fusions of TSAR1 and TSAR2 with a His-MBP tag were generated by cloning into the plasmid pDEST-HisMBP (Nallamsetty et al., 2005). The plasmids were introduced into *Escherichia coli* One Shot BL21 Star (DE3) cells (Thermo Scientific). Protein expression, purification, and DNA binding in the protein-binding microarray were carried out according to Godoy et al. (2011).

Generation of *M. truncatula* Hairy Roots

Sterilization of *M. truncatula* seeds (ecotype Jemalong J5), transformation of seedlings by *Agrobacterium rhizogenes* (strain LBA 9402/12), and the subsequent generation of hairy roots were carried out as described previously (Pollier et al., 2011). Hairy roots were cultivated for 21 d in liquid medium to provide proper amounts to be used for RNA and metabolite extraction.

qPCR

Frozen roots of three independent transgenic lines were ground under liquid nitrogen. The material was used to prepare total RNA and first-strand complementary DNA with the RNeasy Mini Kit (Qiagen) and the iScript cDNA Synthesis Kit (Bio-Rad), respectively, according to each manufacturer's instructions. qPCR primers for *TSAR1* and *TSAR2* were designed using Beacon Designer 4 (Premier Biosoft International). The *M. truncatula* 40S ribosomal protein S8 and translation elongation factor1α were used as reference genes. All qRT primers used are listed in Supplemental Table S6. The qPCR was carried out with a LightCycler 480 (Roche) and the LightCycler 480 SYBR Green I Master Kit (Roche) according to the manufacturer's guidelines. Three replicates were made for each reaction. Relative expression levels using multiple reference genes were calculated using qBase (Hellemans et al., 2007).

Y1H Assays

The Y1H reporter strain was made as described (Deplancke et al., 2006). *TSAR1* and *TSAR2* full-length ORFs were cloned into pDEST22 (Invitrogen), thereby creating a fusion with the GAL4 activation domain. Empty pDEST22 served as a negative control. The yeast reporter strain was transformed with the pDEST22 preys followed by the assessment of growth on synthetic defined -His plates with and without 20 mM 3-amino-1,2,4-triazole after an incubation period of 4 d at 30°C.

RNA-Seq Analysis

Total RNA of three independent transformant lines per construct was submitted to GATC Biotech (http://www.gatc-biotech.com/) for Illumina HiSeq 2500 RNA sequencing (50 bp, single-end read). As described (Pollier et al., 2013b) and using default parameters, the raw RNA-Seq reads were quality trimmed and mapped on the *M. truncatula* genome version 4.0 (Tang et al., 2014) with TopHat version 2.0.6. Uniquely mapped reads were counted, and FPKM values were determined with Cufflinks version 2.2.1. (Trapnell et al., 2010; Kim et al., 2013). Differential expression analyses were performed using Cuffdiff (Trapnell et al., 2010).

Metabolite Profiling of Transformed *M. truncatula* Hairy Roots

Upon harvest, *M. truncatula* hairy roots were rinsed with purified water, frozen, and ground in liquid nitrogen. A total of 400 mg of the ground material was used for metabolite extractions as described previously (Pollier et al., 2011). Liquid chromatography-electrospray ionization-mass spectrometry analysis was performed using an Acquity UPLC BEH C18 column (150 $\times$ 2.1 mm, 1.7 μm; Waters) mounted on an Acquity UPLC system (Waters). The liquid chromatography system was coupled to an LTQ XL linear ion-trap mass spectrometer (Thermo Electron) or an LTQ FT Ultra (Thermo Electron) via an electrospray ionization source operated in the negative mode. The following gradient was run using acidified (0.1% formic acid) water:acetonitrile (99:1, v/v; solvent A) and acetonitrile:water (99:1, v/v; solvent B): 0 min, 5% B; 30 min, 55% B; and 35 min, 100% B. The injection volume was 10 μL, the flow rate was 300 μL min^{-1}, and the column temperature was 40°C. Negative ionization was obtained with a capillary temperature of 150°C, sheath gas of 25 (arbitrary units), auxiliary gas of 3 (arbitrary units), and a spray voltage of 4.5 kV. Full mass spectrometry spectra between *m/z* 120 and 1,400 were recorded. For identification, full mass spectrometry spectra were interchanged with a dependent MS^2 scan event, in which the most abundant ion in the previous full mass spectrometry scan was fragmented, and two dependent MS^3 scan events, in which the two most abundant daughter ions were fragmented. The collision energy was set at 35%. All samples were analyzed on the LTQ XL linear ion-trap system, and for identification, representative samples were reanalyzed on the LTQ FT Ultra system. The resulting chromatograms were integrated and aligned using the Progenesis QI software (Waters). The PLS-DA was performed with the SIMCA-P 11 software package (Umetrics) with Pareto-scaled mass spectrometry data. Peaks with an absolute covariance value above 0.03 and an absolute correlation value above 0.8 were considered significantly different.

The GenBank/EMBL/DNA Data Bank of Japan accession numbers for TSAR1 and TSAR2 are KM409647 and KR349466, respectively. The raw RNA-Seq read data reported here are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-3532.

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Schematic overview of the TS and sterol biosynthesis pathways in *M. truncatula*.

Supplemental Figure S2. Coexpression analysis of *HMGR1*, *MKB1*, and TF genes.

Supplemental Figure S3. Relative transactivation of *ProHMGR1* by the seven selected TF candidates for involvement in TS metabolism.

Supplemental Figure S4. Alignment of bHLH domains.

Supplemental Figure S5. Phylogenetic analysis of TSAR1 and TSAR2.

Supplemental Figure S6. Overexpresssion of *TSAR1* primarily activates nonhemolytic TS biosynthesis in *M. truncatula* hairy roots.

Supplemental Figure S7. Number of differentially expressed genes in *TSAR*-overexpressing *M. truncatula* hairy roots.

Supplemental Figure S8. TSAR1^{OE}, TSAR2^{OE}, and TSAR1^{KD} roots exhibit no phenotypical changes compared with control roots.

Supplemental Figure S9. Nucleotide sequence of *3xCACGAG*$_{[HMGR1]}$.

Supplemental Table S1. TS promoter sequences.

Supplemental Table S2. Differential peaks identified by LC-MS analysis in TSAR1^{OE} *M. truncatula* hairy roots.

Supplemental Table S3. Differential peaks identified by LC-MS analysis in TSAR2^{OE} *M. truncatula* hairy roots.

Supplemental Table S4. Number of sequenced reads in the RNA-Seq analysis of *TSAR*-overexpressing *M. truncatula* hairy roots.

Supplemental Table S5. FPKM values of all differentially expressed genes in *TSAR*-overexpressing *M. truncatula* hairy roots.

Supplemental Table S6. Primers used.

ACKNOWLEDGMENTS

We thank Jennifer Fiallos-Jurado for technical assistance, Frederik Coppens for support with the RNA-Seq analysis, Geert Goeminne for aid with the data processing of the metabolite profiling, and Annick Bleys for help with preparing the article.

Received October 22, 2015; accepted November 19, 2015; published November 23, 2015.

LITERATURE CITED

Achnine L, Huhman DV, Farag MA, Sumner LW, Blount JW, Dixon RA (2005) Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume *Medicago truncatula*. Plant J **41:** 875–887

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

Avato P, Bucci R, Tava A, Vitali C, Rosato A, Bialy Z, Jurzysta M (2006) Antimicrobial activity of saponins from *Medicago* sp.: structure-activity relationship. Phytother Res **20:** 454–457

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

Broeckling CD, Huhman DV, Farag MA, Smith JT, May GD, Mendes P, Dixon RA, Sumner LW (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, Biazzi E, Panara F, Tava A, Scaramelli L, Porceddu A, Graham N, Odoardi M, Piano E, Arcioni S, et al (2011) *Medicago truncatula* CYP716A12 is a multifunctional oxidase involved in the biosynthesis of hemolytic saponins. Plant Cell **23:** 3070–3081

Carretero-Paulet L, Galstyan A, Roig-Villanova I, Martínez-García JF, Bilbao-Castro JR, Robertson DL (2010) Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in Arabidopsis, poplar, rice, moss, and algae. Plant Physiol **153:** 1398–1412

Chini A, Fonseca S, Fernández G, Adie B, Chico JM, Lorenzo O, García-Casado G, López-Vidriero I, Lozano FM, Ponce MR, et al (2007) The JAZ family of repressors is the missing link in jasmonate signalling. Nature **448:** 666–671

Corey EJ, Matsuda SPT, Bartel B (1993) Isolation of an *Arabidopsis thaliana* gene encoding cycloartenol synthase by functional expression in a yeast mutant lacking lanosterol synthase by the use of a chromatographic screen. Proc Natl Acad Sci USA **90:** 11628–11632

Cuéllar Pérez AC, Goossens A (2013) Jasmonate signalling: a copycat of auxin signalling? Plant Cell Environ **36:** 2071–2084

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

Deplancke B, Vermeirssen V, Arda HE, Martinez NJ, Walhout AJM (2006) Gateway-compatible yeast one-hybrid screens. Cold Spring Harb Protoc **2006:** pdb.prot4590

De Sutter V, Vanderhaeghen R, Tilleman S, Lammertyn F, Vanhoutte I, Karimi M, Inzé D, Goossens A, Hilson P (2005) Exploration of jasmonate signalling via automated and standardized transient expression assays in tobacco cells. Plant J **44:** 1065–1076

Dun XP, Wang JH, Chen L, Lu J, Li FF, Zhao YY, Cederlund E, Bryzgalova G, Efendic S, Jörnvall H, et al (2007) Activity of the plant peptide aglycin in mammalian systems. FEBS J **274:** 751–759

Fernández-Calvo P, Chini A, Fernández-Barbero G, Chico JM, Gimenez-Ibanez S, Geerinck J, Eeckhout D, Schweizer F, Godoy M, Franco-Zorrilla JM, et al (2011) The *Arabidopsis* bHLH transcription factors MYC3 and MYC4 are targets of JAZ repressors and act additively with MYC2 in the activation of jasmonate responses. Plant Cell **23:** 701–715

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

Fukushima EO, Seki H, Sawai S, Suzuki M, Ohyama K, Saito K, 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, Goossens A (2014) Natural product biosynthesis in *Medicago* species. Nat Prod Rep **31:** 356–380

Godoy M, Franco-Zorrilla JM, Pérez-Pérez J, Oliveros JC, Lorenzo O, Solano R (2011) Improved protein-binding microarrays for the identification of DNA-binding specificities of transcription factors. Plant J **66:** 700–711

He J, Benedito VA, Wang M, Murray JD, Zhao PX, Tang Y, Udvardi MK (2009) The *Medicago truncatula* gene expression atlas web server. BMC Bioinformatics **10:** 441

Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B, Bailey PC (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, Vandesompele J (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol **8:** R19

Hong GJ, Xue XY, Mao YB, Wang LJ, Chen XY (2012) *Arabidopsis* MYC2 interacts with DELLA proteins in regulating sesquiterpene synthase gene expression. Plant Cell **24:** 2635–2648

Howe GA, Jander G (2008) Plant immunity to insect herbivores. Annu Rev Plant Biol **59:** 41–66

Iturbe-Ormaetxe I, Haralampidis K, Papadopoulou K, Osbourn AE (2003) Molecular cloning and characterization of triterpene synthases from *Medicago truncatula* and *Lotus japonicus*. Plant Mol Biol **51:** 731–743

Ji Y, Xiao J, Shen Y, Ma D, Li Z, Pu G, Li X, Huang L, Liu B, Ye H, et al (2014) Cloning and characterization of AabHLH1, a bHLH transcription factor that positively regulates artemisinin biosynthesis in *Artemisia annua*. Plant Cell Physiol **55:** 1592–1604

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci **8:** 275–282

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

Kazan K, Manners JM (2013) MYC2: the master in action. Mol Plant **6:** 686–703

Kevei Z, Lougnon G, Mergaert P, Horváth GV, Kereszt A, Jayaraman D, Zaman N, Marcel F, Regulski K, Kiss GB, et al (2007) 3-Hydroxy-3-methylglutaryl coenzyme A reductase 1 interacts with NORK and is crucial for nodulation in *Medicago truncatula*. Plant Cell **19:** 3974–3989

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol **14:** R36

Le Gall M, Quillien L, Guéguen J, Rogniaux H, Sève B (2005) Identification of dietary and endogenous ileal protein losses in pigs by immunoblotting and mass spectrometry. J Nutr **135:** 1215–1222

Louis S, Delobel B, Gressent F, Rahioui I, Quillien L, Vallier A, Rahbé Y (2004) Molecular and biological screening for insect-toxic seed albumins from four legume species. Plant Sci **167:** 705–714

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

Nallamsetty S, Austin BP, Penrose KJ, Waugh DS (2005) Gateway vectors for the production of combinatorially-tagged His6-MBP fusion proteins in the cytoplasm and periplasm of *Escherichia coli*. Protein Sci **14:** 2964–2971

Naoumkina MA, Modolo LV, Huhman DV, Urbanczyk-Wochniak E, Tang Y, Sumner LW, Dixon RA (2010) Genomic and coexpression analyses predict multiple genes involved in triterpene saponin biosynthesis in *Medicago truncatula*. Plant Cell **22:** 850–866

Oliva ML, Ferreira RdaS, Ferreira JG, de Paula CA, Salas CE, Sampaio MU (2011) Structural and functional properties of Kunitz proteinase inhibitors from Leguminosae: a mini review. Curr Protein Pept Sci **12:** 348–357

Oliva MLV, Silva MCC, Sallai RC, Brito MV, Sampaio MU (2010) A novel subclassification for Kunitz proteinase inhibitors from leguminous seeds. Biochimie **92:** 1667–1673

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

Pauwels L, Goossens A (2011) The JAZ proteins: a crucial interface in the jasmonate signaling cascade. Plant Cell **23:** 3089–3100

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

Pires N, Dolan L (2010) Origin and diversification of basic-helix-loop-helix proteins in plants. Mol Biol Evol **27:** 862–874

Pollier J, Goossens A (2012) Oleanolic acid. Phytochemistry **77:** 10–15

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

Pollier J, Moses T, González-Guzmán M, De Geyter N, Lippens S, Vanden Bossche R, Marhavý P, Kremer A, Morreel K, Guérin CJ, et al (2013a) The protein quality control system manages plant defence compound synthesis. Nature **504:** 148–152

Pollier J, Rombauts S, Goossens A (2013b) Analysis of RNA-Seq data with TopHat and Cufflinks for genome-wide expression analysis of jasmonate-treated plants and plant cultures. Methods Mol Biol **1011:** 305–315

Roux B, Rodde N, Jardinaud MF, Timmers T, Sauviac L, Cottret L, Carrère S, Sallet E, Courcelle E, Moreau S, et al (2014) An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. Plant J **77:** 817–837

Shang Y, Ma Y, Zhou Y, Zhang H, Duan L, Chen H, Zeng J, Zhou Q, Wang S, Gu W, et al (2014) Biosynthesis, regulation, and domestication of bitterness in cucumber. Science **346:** 1084–1088

Spyropoulou EA, Haring MA, Schuurink RC (2014) RNA sequencing on *Solanum lycopersicum* trichomes identifies transcription factors that activate terpene synthase promoters. BMC Genomics **15:** 402

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

Suzuki H, Reddy MSS, Naoumkina M, Aziz N, May GD, Huhman DV, Sumner LW, Blount JW, Mendes P, Dixon RA (2005) Methyl jasmonate and yeast elicitor induce differential transcriptional and metabolic reprogramming in cell suspension cultures of the model legume *Medicago truncatula*. Planta **220:** 696–707

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

Tang H, Krishnakumar V, Bidwell S, Rosen B, Chan A, Zhou S, Gentzbittel L, Childs KL, Yandell M, Gundlach 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, Avato P (2011) Biosynthesis of saponins in the genus *Medicago*. Phytochem Rev **10:** 459–469

Thines B, Katsir L, Melotto M, Niu Y, Mandaokar A, Liu G, Nomura K, He SY, Howe GA, Browse J (2007) JAZ repressor proteins are targets of the SCF^{COI1} complex during jasmonate signalling. Nature **448:** 661–665

Toledo-Ortiz G, Huq E, Quail PH (2003) The *Arabidopsis* basic/helix-loop-helix transcription factor family. Plant Cell **15:** 1749–1770

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol **28:** 511–515

Urbanczyk-Wochniak E, Sumner LW (2007) MedicCyc: a biochemical pathway database for *Medicago truncatula*. Bioinformatics **23:** 1418–1423

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

Van Moerkercke A, Fabris M, Pollier J, Baart GJ, Rombauts S, Hasnain G, Rischer H, Memelink J, Oksman-Caldentey KM, Goossens A (2013) CathaCyc, a metabolic pathway database built from *Catharanthus roseus* RNA-Seq data. Plant Cell Physiol **54:** 673–685

Van Moerkercke A, Steensma P, Schweizer F, Pollier J, Gariboldi I, Payne R, Vanden Bossche R, Miettinen K, Espoz J, Purnama PC, 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

Vicente J, Cascón T, Vicedo B, García-Agustín P, Hamberg M, Castresana C (2012) Role of 9-lipoxygenase and α-dioxygenase oxylipin pathways as modulators of local and systemic defense. Mol Plant **5:** 914–928

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

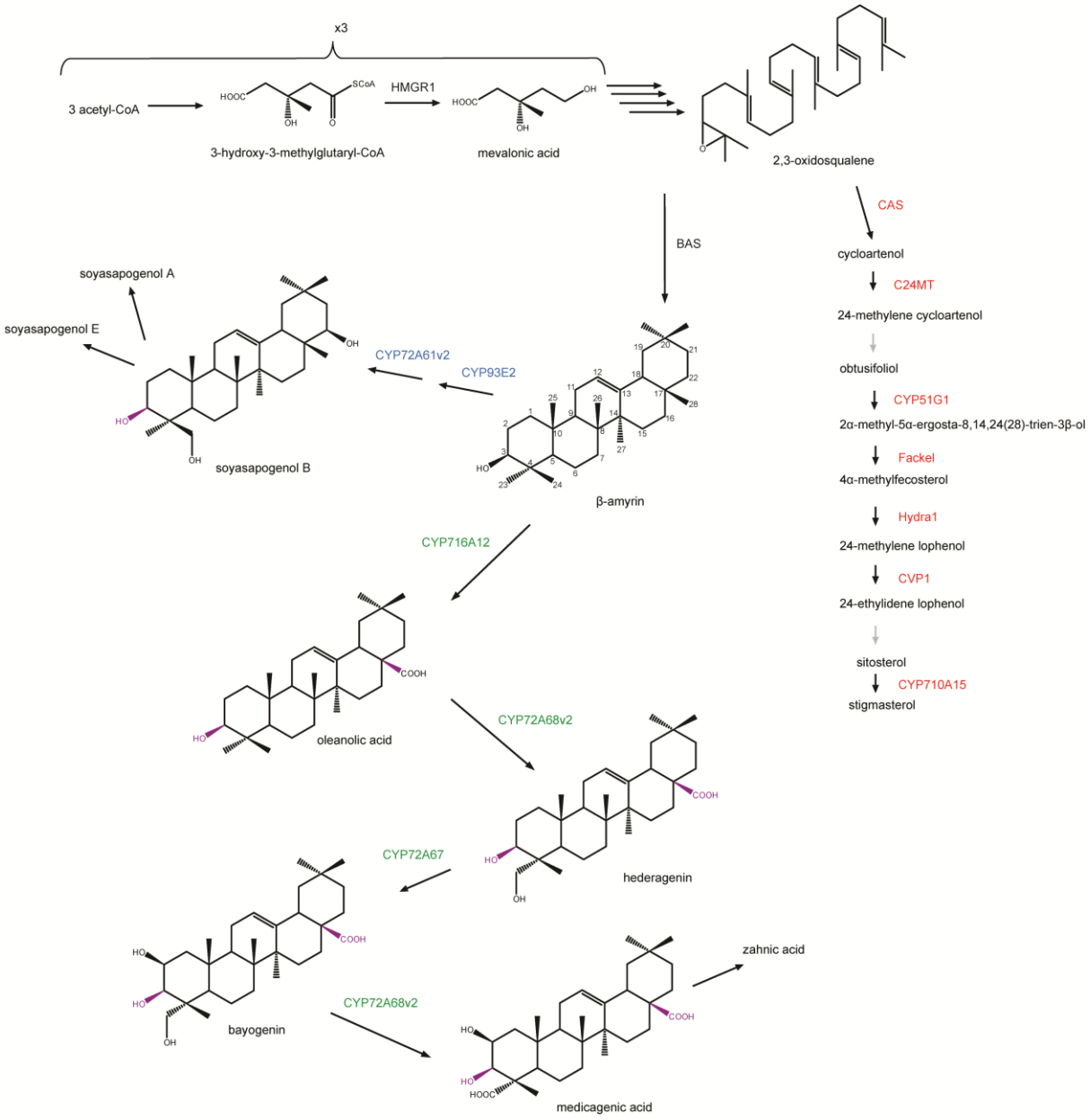
Wasternack C, Hause B (2013) Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in *Annals of Botany*. Ann Bot (Lond) **111:** 1021–1058

Watanabe Y, Barbashov SF, Komatsu S, Hemmings AM, Miyagi M, Tsunasawa S, Hirano H (1994) A peptide that stimulates phosphorylation of the plant insulin-binding protein: isolation, primary structure and cDNA cloning. Eur J Biochem **224:** 167–172

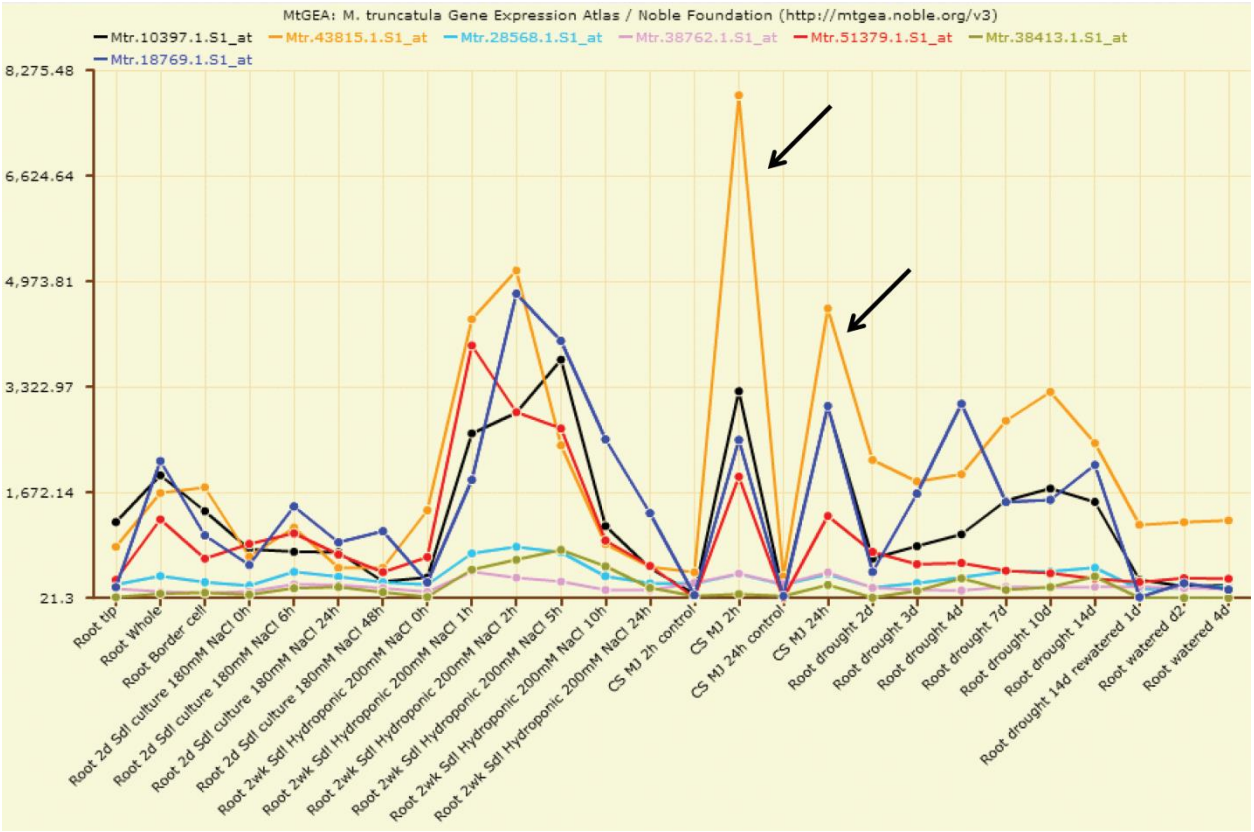
Xu YH, Wang JW, Wang S, Wang JY, Chen XY (2004) Characterization of GaWRKY1, a cotton transcription factor that regulates the sesquiterpene synthase gene (+)-δ-cadinene synthase-A. Plant Physiol **135:** 507–515

Yu ZX, Li JX, Yang CQ, Hu WL, Wang LJ, Chen XY (2012) The jasmonate-responsive AP2/ERF transcription factors AaERF1 and AaERF2 positively regulate artemisinin biosynthesis in *Artemisia annua* L. Mol Plant **5:** 353–365

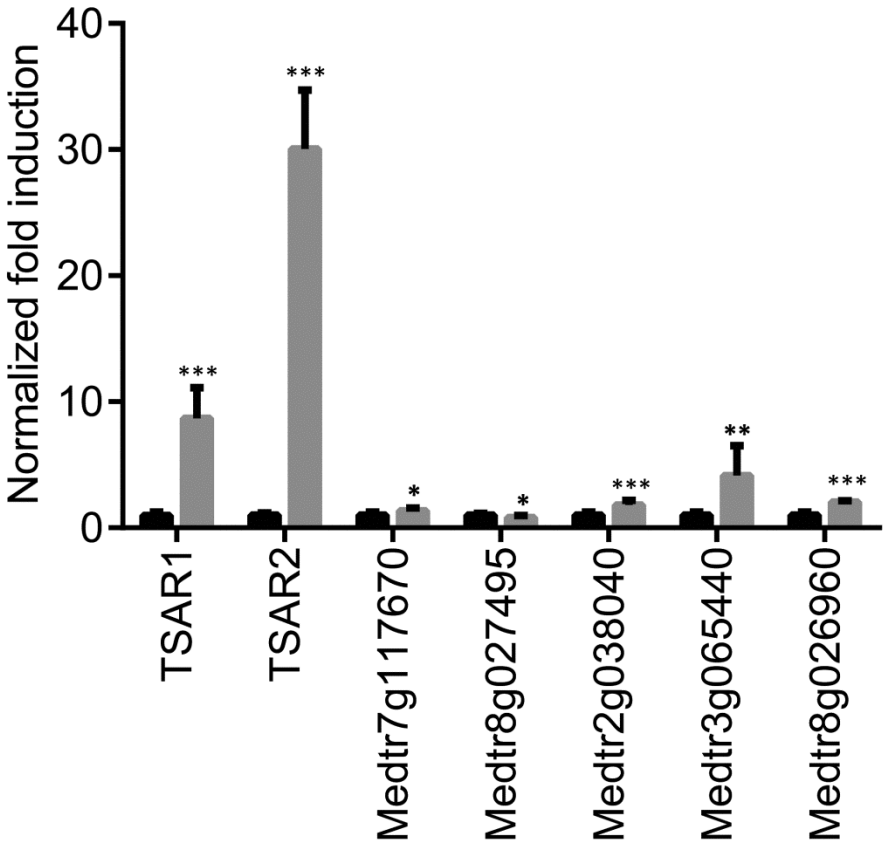
Supplemental Data



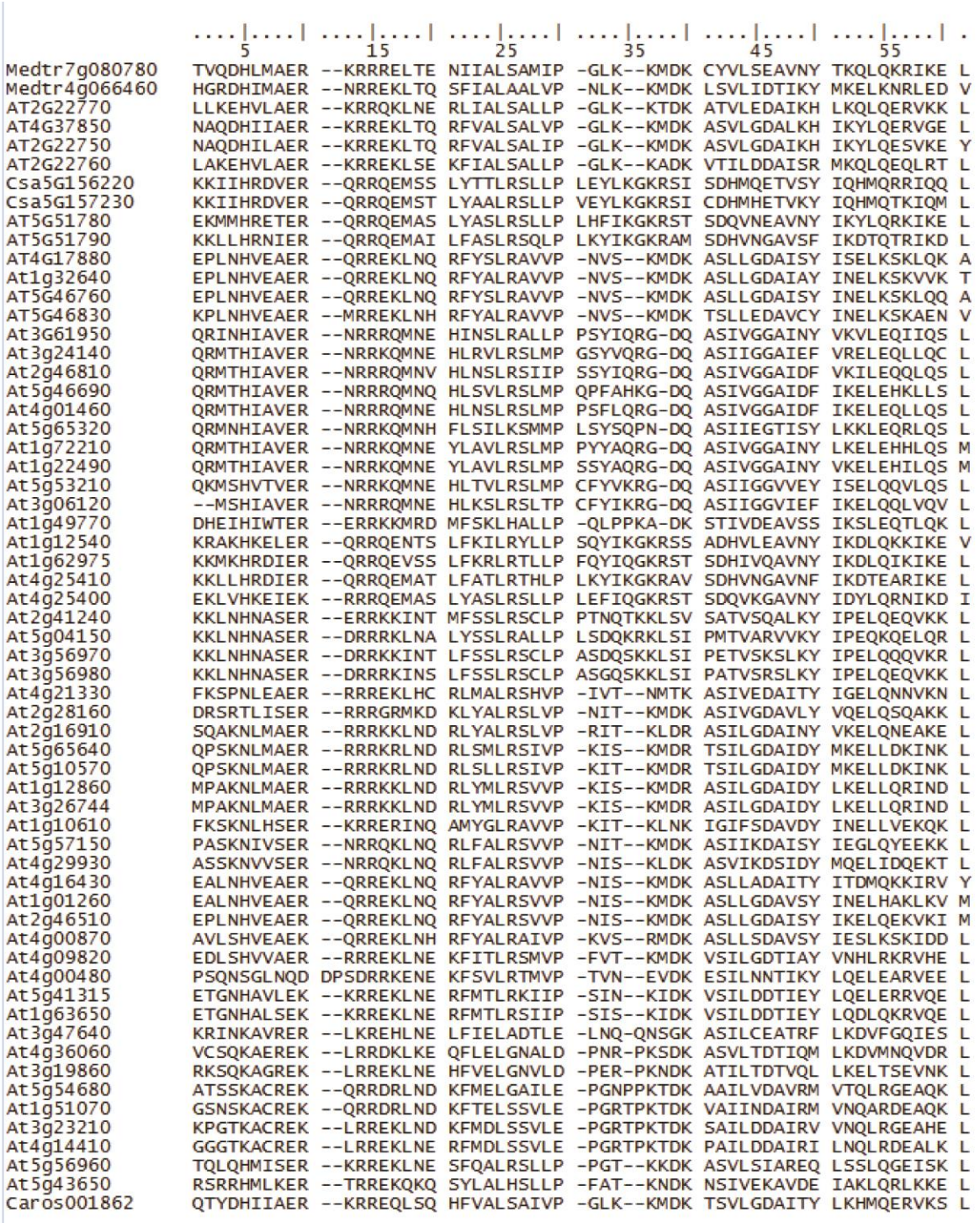
Supplemental Figure S1. Schematic overview of the triterpene saponin and sterol biosynthesis pathways in *Medicago truncatula*. Known cytochrome P450-dependent mono-oxygenases acting in the non-haemolytic and haemolytic triterpene saponin synthesis pathways are designated in blue and green respectively. Enzymes of the sterol biosynthesis pathway are depicted in red. Gray arrows signify multiple enzymatic steps. Substituents in purple are locations where glycosylation occurs. HMGR, 3-hydroxy-3-methylglutaryl-CoA reductase; BAS, β-amyrin synthase; CAS, cycloartenol synthase; C24MT, C-24 methyltransferase; CYP51G1, C-14 demethylase; Fackel, C-14 reductase; Hydra1, C8,7-isomerase; CVP1, C-24 methyltransferase; CYP710A15, C-22 desaturase.



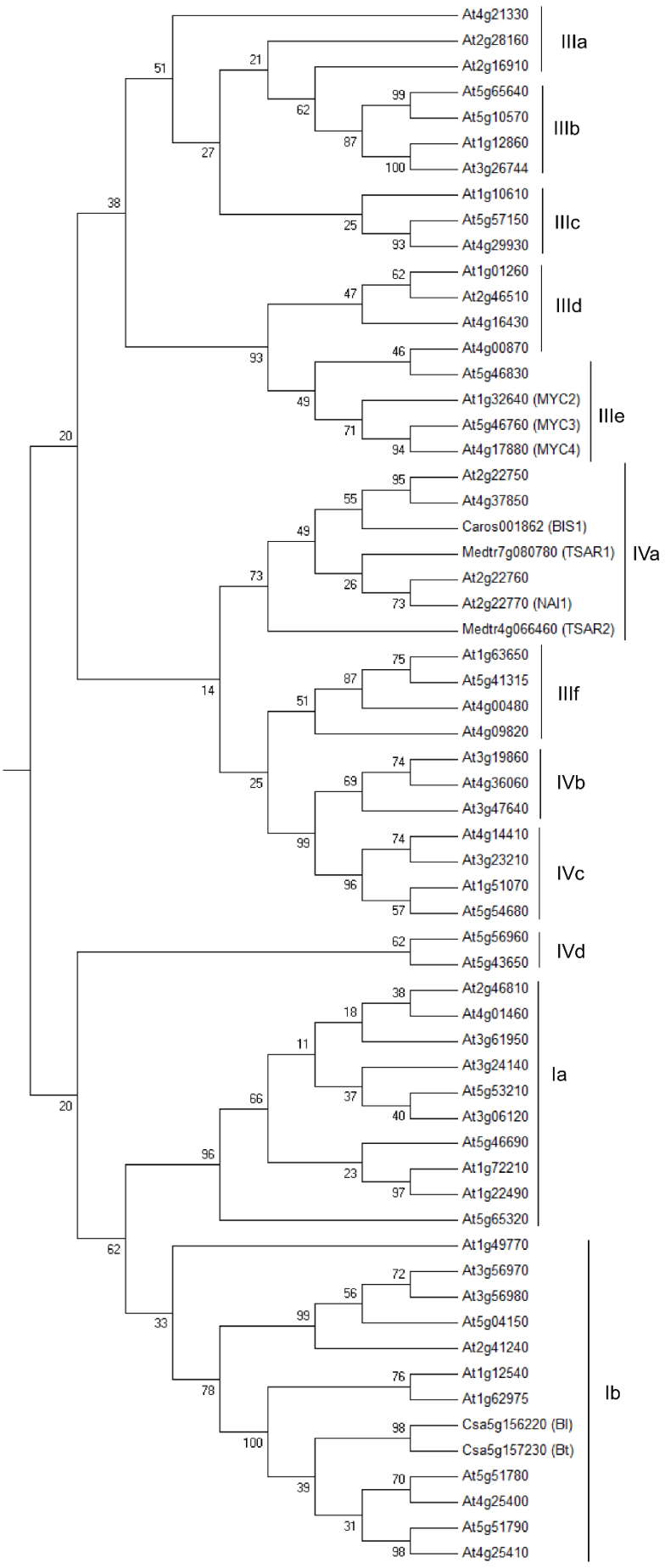
Supplemental Figure S2. Coexpression analysis of *HMGR1*, *MKB1* and TF genes. Coexpression profiles of *HMGR1* (black; Mtr.10397), *MKB1* (orange; Mtr.43815), *Medtr2g038040* (red; Mtr.51379), *Medtr7g117670* (light blue; Mtr.28568), *Medtr8g02749* (pink; Mtr.38762), *Medtr3g065440* (army green; Mtr.38413) and *Medtr8g026960* (blue: Mtr.18769) in *M. truncatula* roots under various culturing conditions, generated with the MtGEA tool (He et al., 2009). Values in the y-axis reflect transcript levels determined by microarray analysis (He et al., 2009). Arrows depict values in methyl jasmonate (MJ)-treated cell suspension cultures.



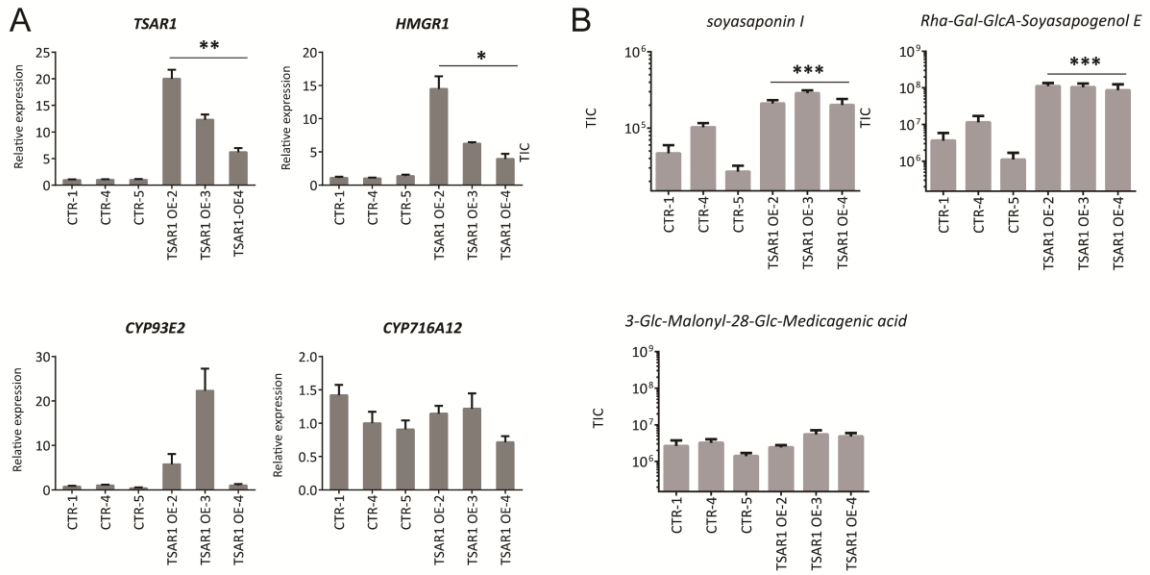
Supplemental Figure S3. Relative transactivation of *ProHMGR1* by the seven selected TF candidates (in gray) for involvement in TS metabolism. Values in the y-axis are normalized fold-changes relative to protoplasts cotransfected with the reporter constructs and a pCaMV35S:GUS (GUS) control plasmid (in black). The error bars designate SE of the mean (n=8). Statistical significance was determined by a Student's *t*-test (*P<0.05, **P<0.01, ***P<0.001).



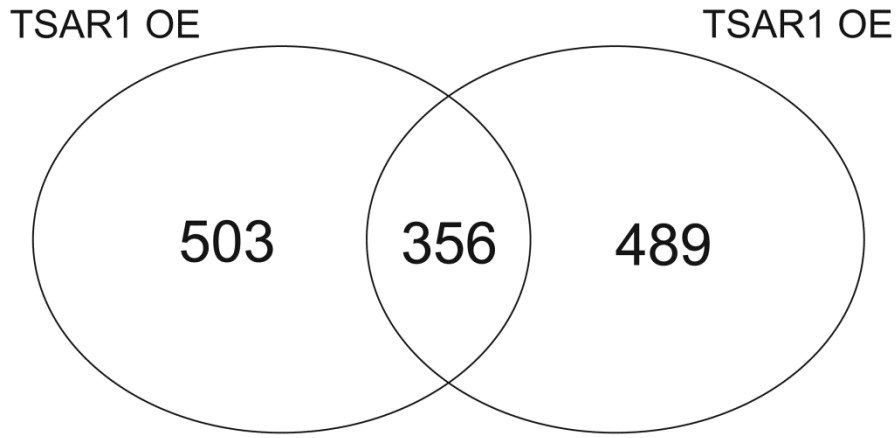
Supplemental Figure S4. Alignment of bHLH domains. The bHLH domains of *M. truncatula* TSAR1 and TSAR2, *C. roseus* BIS1, cucumber Bl and Bt and all *A. thaliana* bHLH proteins from clades I, III and IV, were aligned using the ClustalW tool of BioEdit7.


At4g21330
At2g28160
At2g16910
IIIa
At5g65640
At5g10570
At1g12860
At3g26744
IIIb
At1g10610
At5g57150
At4g29930
IIIc
At1g01260
At2g46510
At4g16430
IIId
At4g00870
At5g46830
At1g32640 (MYC2)
At5g46760 (MYC3)
At4g17880 (MYC4)
IIIe
At2g22750
At4g37850
Caros001862 (BIS1)
Medtr7g080780 (TSAR1)
At2g22760
At2g22770 (NAI1)
Medtr4g066460 (TSAR2)
IVa
At1g63650
At5g41315
At4g00480
At4g09820
IIIf
At3g19860
At4g36060
At3g47640
IVb
At4g14410
At3g23210
At1g51070
At5g54680
IVc
At5g56960
At5g43650
IVd
At2g46810
At4g01460
At3g61950
At3g24140
At5g53210
At3g06120
At5g46690
At1g72210
At1g22490
At5g65320
Ia
At1g49770
At3g56970
At3g56980
At5g04150
At2g41240
At1g12540
At1g62975
Csa5g156220 (BI)
Csa5g157230 (Bt)
At5g51780
At4g25400
At5g51790
At4g25410
Ib

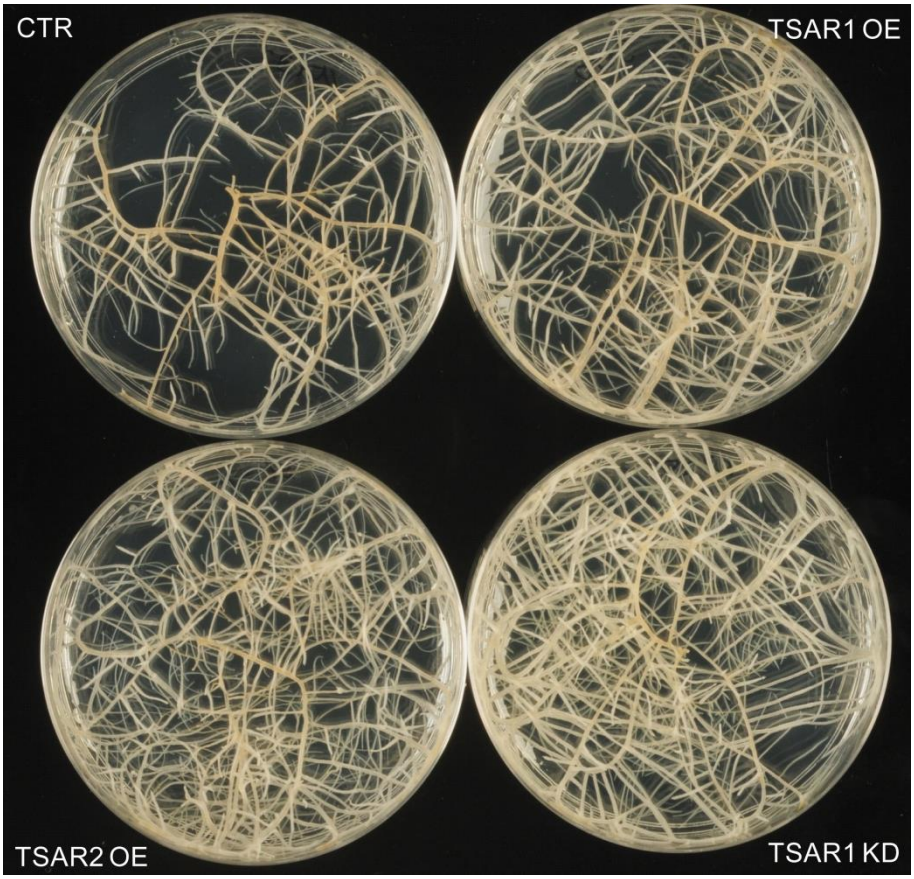
Supplemental Figure S5. Phylogenetic analysis of TSAR1 and TSAR2. A neighbor-joining tree was constructed in Mega5 (Tamura et al., 2011). Positions containing gaps were eliminated. Evolutionary distances were calculated using the Jones, Taylor, and Thorton (JTT) model (Jones et al., 1992). Bootstrap analyses were performed with 1,000 bootstrap replicates. No outgroup was used. Subclades were defined as in Heim et al. (2003). Subclades are designated on the right.



Supplemental Figure S6. Overexpresssion of *TSAR1* primarily activates non-haemolytic TS biosynthesis in *M. truncatula* hairy roots. TSAR1^{OE} lines did not show increased expression of *TSAR2* and *CYP716A12* (A), nor any increase in the accumulation of haemolytic TSs (B), early after the generation of the root lines. A, qRT-PCR analysis of TS biosynthetic genes in three independent control (CTR) and three independent TSAR1^{OE} *M. truncatula* hairy root lines. Control lines were transformed with a *pCaMV35S:GUS* (GUS) construct. Expression ratios were plotted relatively to the normalized CTR-1. The error bars designate SE of the mean (n=3). B, Average total ion current (TIC) of peaks corresponding to TS. The error bars designate SE of the mean (n=3). Statistical significance was determined by a Student's *t*-test (*P<0.05, **P<0.01, ***P<0.001).



Supplemental Figure S7. Number of differentially expressed genes in *TSAR*-overexpressing *M. truncatula* hairy roots.



Supplemental Figure S8. TSAR1^{OE}, TSAR2^{OE} and TSAR1^{KD} roots exhibit no phenotypical changes compared to control roots. *M. truncatula* hairy roots grown on solid medium are depicted.

CCTCGTTGTTCACGAGTTCCTTACGAGATAATCCTGGCCCACCTCGTTGTTCACGAGTTCCT
TACGATGCTAACATGCTGTGCCTCGTTGTTCACGAGTTCCTTACGA

Supplemental Figure S9. Nucleotide sequence of *3xCACGAG*$_{[HMGR1]}$. N-boxes (CACGAG) are marked in yellow. Two independent linker sequences are indicated in green and purple, respectively.

Supplemental Table S1. TS Promoter sequences. N- (CACGAG) and G-boxes (CACGTG) are bold underlined.

ProHMGR1

ATTGGTAGGATCAATTGTTGATTGTCTTGGTTTATGTAAAATGTTAAAATGTGACAATTAAAAAGAAACGGAGGT
AGTACTTCTATCATATTTAATGCTATTACTTCTAAATTCGTTTTCAATTATTTATGTTTGAATCCTCCGAGTGTC
AATTTTTGTTGATTTGTTTATAGTGGGATTGGTCCTTTGGATTAGTCGGTTCTTGAATCAAATACTAAATTTTCA
AACAAAAATTTAGATTTTACTTTTCTATAACCATTTTATTATCATAGAAAAAAAAATTAAATCTCCGAAAAATTA
TCATGAACAAGTTTTCATGAGGGACAATTTAGAAAATGCAATTATGGCTAAACTATGTTTCAAGTCCCTTAAATT
GTTCATTTTGTTCACTTAAATTACAAACATTGAGTTATTCAAATAAAAATTACAAACATTGAGTAAGAAAACAAA
ATTAAACACGTTTTTGACTTTAGATTCGATTCATATTAAACACATTTGAGTCAAAAACAAAATTAAACACGTTTT
TTTTGCAGCACAAGTGATGTATAAAACTAATTAGTTTTTATTTCAGCAACAAAAAATGTACTTATAGTTTGATTA
GTTTGTGCAAATAAGAAATATATATAGTATTTTAGAAACGTCAAGTAAAGGCTTCTTCTAGTGATGTGTAATGTT
GGCATCTCTCCTCCATAAAATATGTTATCTTGGAGTAAAAGTTAGGTACTGATTACAATATTCTTCCCTCGTTGT
T**CACGAG**TTCCTTACGATCTAGCTATCAAGCTACTGTTACTTGACTTCTTAAAATATTGGCATATACCTAATTCC
TCACTCACTCCCTGAGTACTCTTTATCTATAAATAACATTCAAAATAAATCTTGGCATTCATTTTCAAGTTAAGT
TTCAATCATCCCCTAACCTGAACTCCTTTTTCTCCTTTCTTTTTCCATTTTCGGATTTCAAATCCTTCACTTAAT
AACCAAGAGAAAGAGAGAGAGAGAG

ProCYP93E2

AAAATAATATGTCATTGTAGACATGTTTTCTTTTTGCCTACTACCATAAGAATCTTTTCCTAGTAGGCGTCCAAA
ATATTTAGATTTAAATATTATAATTATTATTTTAGATTTAAATGTATTTATTATTACTTTTAATCGAATAGTTTA
GAGGCGAGAAGAAACACACAATAAATATTAAGAAGTGGAAAATTCAAAATTAGAACCCGAGATCAAAAACTTGCG
GTTTTCAACCCGGGATCATACTAACAAGTGTCATAAATATTTTGTTTAAAGAATACTCTCTGGTAACATTTATAA
GCAAAAAAAAACTTTTTAGGTACATTGAATAACTAATATATATGACATATAAATGTGACCAGATACATTGATTATT
CAATGAATTTAAAATGCTGATTTTTACTTACAAATTTGATCGGAGGGAGTACATAAAAAGAAAATTTATATTAGA
AATATAAATCAAACATACGTCGAAGTCATATATACATCATTTTTTATGCAAAAGTTGCAATTCCTAACAATGTTC
TCACGATATTTGTTAGCCTTTCTCAATTAAAAATGTATTTTTAAAATAATTTGCGAGTGTGTAAATTAACATTGG
GATTATACAAGTATTCTCAACCATTTATTCATCAAGACAAAACAAAATCTTAACAATTTACCATAAATAAACATA
AAAAACACGTATCAATCACACACCGTTAGCTGACAAACAACGTCAGCAATCATGAAATTATTTTATGTAC**CACGA**
GGGACACATAAGCAACATTCTTATTATAATATTCGAC**CACGAG**GATCGCAAAAATTAATAACTTAACTTCACTCA
ACAGTACCATTCAAAATCATACTAGACCAAGTCAAATTTGTTCTTCTATTTATGTTCCATCATAGTTCAAGAATT
TAGATTCATCTTAAGCATAGACGTTGAGTTGTGTTTGGTATTTTGCATTGACAACATTGTCCTTTTTTGTCACAC
CAATTACCAAATCAAGTACTCAACC

ProCYP72a61v2

TCGATTTTTTATTTTATTTTTTCAATTGCGATTTTGCGCGTTTCGGTCGAAAAAATTAAAAAAAAAAATCGAGACT
TATTCGAAATGTTACGAAACTTTTCAAATCATCTATATGTATTTTTATAGAGGTGTCTGGAAAGTTTCGTAGAAT
TCTGAATAACTTTCGATTATTTTGATTTTTCTACTGAAACGCATAAAATTATAATAAAAAGCAGCAAAGTCAACG

GCAGTTAACTGCCTCTGGAGGAGCCAGCGGCAGTTAACTCATGACGCATATACCAGAGGCACTTAACTACCGCTG
GTGGTAATTTGATCATTTATTTGTGTCAATAGGAAATATTCTATGTTAAAAGCAATTTTCCACCACCAATGCTAG
CATGAGGCCGCGAGACAGAGAAAAGAGACCCAGAGTTAATATGGGTCCTATGATCACTATCAAACGATTAACCAA
ACAACAAGCCCAAAAAAGATGCAACCACCTAACAATTTCACTCTTACATGGCAACTTTAACGACGTATAAAACGA
GCTACCCACTCTAGAAAAGGTACACGCTTTGGTTTTGAAGAAGAACCTGAAAGAACCTTGATCATGCCCAATTTT
CTAATGACTAATATATACAGTTTTTTTTATTTTTTTTTATTGACAACTATTTAATTGGGAAATAGTTTATTTAGCT
GACCGACCTGATACTACGCTAAACGCCAGTAACAAATGACATCATACTTTTGTGTTTTTTACTTTTGTGGTGATG
AATACTTTTAGTATTTTTTCCTTACAAAAAATTTGCACCGTTCAAGTGTATTTCTTTTCCTTATACAAACCTATA
TCATGGTATCATATCATATATTATATTATTACTGCGTGGCT**CTCGTG**GGTCGCT**CACGTG**GTATTATATAATCAT
TATATAATCATAAAAATAATATTTGTTTATAAGCTATCGTAGCATTATATTGTAACAACGATGAGTTTCCTTTGA
GCACTATCACACACACACATAAACA

ProCYP716A12

AGCTAATCATTATATACCTGACCAACATCATCATTGTTCTTTCACCAAAACAAAAGCAAAAAAAAAACAAAGAAAA
ATAGATTCTCAAATTTGCAAAGAAGACTTGACCAAGCAGTAAAAACCAATTTAAAGACGATAAGAAACACAGTGA
CTTATACACAATTGACCGATGTTAATTTGAAACAAAGGAATACGGAAGAAAAATCTCACTAAAACACTTATTTAT
TTTAACTTTCAATGGATTTAATATAATAGAGAGAAGTGATAGAGGAGTTCCAATTGTCGAACACTTTTTTTTTTT
TTATGAGGATGGAAAAATGGAGTTTGCCACAACAAGGTTTGAATTAGGATACTCTTGGTGTTGCTATACTTAAAA
TTGTACACTTTTTGCATTAAGATCTAGTAACTTTTTTGAAGGTCGGTGGTCGGACCATTTTTGTTTCTTCAAAAA
AAAAAACTTTTTTAAAGGTCTCGACTTTACTTTTTTCATGCTTCTAGCTAAGCCGTTAGTTTCATCCTTCAAAAC
TCTAACATATATCATGCTCAAATAATGATAAGCTAGAACATATTTATAATCAGCTAAAATACATAATGACATTGT
TTGACAAATAAATAATCAATATTGCATTAAACTTTAAAAATAACAACTATATAAAAATGCGACAATTAAAATGGA
AACGAGGGAGTGATATCTCTCTTTTTTTCATACGATTTATCAAAAGAGTTATCGAAGAATAATTTTAAAAATTTA
GCTTCTTAATTCTTCCACCGTTATATAGGTTTACAAAACACATGACTAAAATACTTTCTTTAATCAATTGCCGTA
TTGTGACTTGTCATATTGTATTGTTTTATAACCATGAAATTGAACACCCCCCTAACATTATGCCTATAAATTGCG
GCATCCAAAACATGGACACGCAGTCCAACTGCCATTGCACCATTGTAAGCTAATAAAGAGAATTTTATTATAAAC
AAGTTTTGTTAGTCACAAAAAAAGC

ProCYP72A67

CTGTTTGAAGTTCAAATTAAGTTTTAATTACAATTTATTTGGGAATATAAGTTTTAGTACATAAATTGGTCTTTT
TATTTTGAAAACTAAAAAAGTTAGGATAAATCTGTGTTGATAAAAGTGATAGTGCAATTTTTAAGAAGTAGAGCG
AGAATTCCATTAGT**CACGAG**ATA**CACGAG**TAATGACCCTCGAGATG**CACGAG**TACATTACCCTCGAGATG**CACGA**
GTATTATAAACATTTATGACTTTTTTTTTTTGGGGAACACTAATATTTTTTGTAAGGCATGCCCACACATGACGG
TAGGAACCAGAAGACGTGCTGCATTAATGTAGGAATGAGTGTAGTACGTTGTTGTGATAATCCTACATTATGTAG
GAAACATCTTACACTCTCCCTCCCGTCGTTATTATAAACAAAATTTTGCATTTTAAATTCATTCATTAATATGGT
AATTTTTACAAAAATATTATGGGAAGACACATACTCTCCGGTTATTAATATAAGCAAAAAACTACATTTTAGATT
CATTTATTTAATGATGTATGTGATCTATAATAAAGACCATGTACATTATTAAATGAATGAATCTAAAATATAGAT
TTTTGTTTATAATGGTGACCGGAGGGATTCTTTCATACCATTCACTATATTTATACGGAGTACCATTTTTTATAG
GCTATAAAGAATGATGATTTGGAAGCACCACCCTTCTTTAATTTTAATAGATTCCACTTAATAAATTATATTAAT
AATATTGTTCAAAAAAAAATTATATTAATAATAATATGAACATTTAGTAAATATGATAACTCACATGTTTTTATA
TTGACATGACAATATATTAGCATATGTATTTATCCTATATGGTACTCCCTCCGTCTCAAAATATAAGTAAATTTT

ATTTTTTAGGTTCATTCAATTAATAATATATATGGTTCATAATATAAACCACATACGTCATTAATTGAATGAACC
TAAAAAATTAAATTTGCTTATATTTTGAAACGGATGGAGTATGCTAATTGAATTTATAGTAACATTAACATGCCG
CAAAATCAACAATGTGTACATGTGACGCTCATTTAAATATTCATAATGAAGAGATAATGCTTTTAATATTGTTGT
TGACCATCACATATACAAACTTAAGCGGTAGCCAAATACATTTGTCAATAATATCAAGTTTGTTTTAATAGAAAT
CAAACCACTATTATTATTGCCCACACTTTCTCTTTCCCTTTCTGTTTCTCTTAGCAGAGATCAACCAACTAGATT
CCAAAACACCAGTAAAAAGGTAAGAAAAACCCACGTCCACACCAATGCTCCCACAAACAATAATGAATGATTCTT
ACTTGCCCACACCTACCAAATTCTATAAAGCAATTAGAGAAATGAAGCATCATTCACTGGCCATAAATTCTCTAT
TCTCTACTTAATTTGGTTTCATTGTTGCATTCTCTATTCTCTATCTACTACTTTTTATCACCTACAAGCAGAGAA

ProUGT73K1

TTTTATCTATGGAGTGACTTTATGAAGACATCCACCAAATTATCTTTCGATGCAACTTTCTCAGTAATAATTTTC
TTAAGTCTGATCTACAACAAACCATATCATACATGGTGCTTCCTACGACACTTGCATATGGTGTACCCTCCATCT
TCTCATCCTCAAATTGAGGATAATGCATAGTCAAAAGTTTTGTGTGATGATCCAAAGGAATCTCAGTGGTCATAG
CACTACTCTTTATGAGTTGTCCCACCACTTTCTTTAGGTAAACATATTGAGATAAGAAAAATTCATCCTTGTCCT
AGTTTTTCCGGTATTTTTTTCAAGTATACTCTTAGCAACAACAAAAAATTTCATCTCAAGCTAGGGGACTAAGTT
GTCCCTTGTATAAAAAACGAGACACTAATTTGGTATTATCTCTTGTCAGGAACTAATTTGTGTGATGAAACTCAG
ATACACTAAATGTTTCATTATACGTAGTCTTCGTGAGCATAGCTCAGTTGACAAGGACAATACATAATATCGCAA
GGTTTGAGTTCAAATCTCAGACACAAAAAAAAAAATGTAAAAAATTATATGTACGAATTTGTGCTTCAATATATT
AAATATTGTTAAACTTTATCATCTGAATTGTACTTATAAAAATGTCCACTCTAGGATCAAATATAGAAAATAATA
AGATTCAATCATTTACAAAAAATATTTTAAAAATTATTCACGCAGTT**CACGAG**TACAGTTTTTATCAATATAAAC
AATGAACAATAAACTAATAAAATATGCAGCCCGTGTATCTCGCTATTGCAAGTATATATATTTTAATGAATCATT
TCGTTTTCTTTGGACCGTCTTTAATTTAGATAAATAATAGTGACAAATAATACATTCATAATTCATCAATCAACC
CCTTTTATAAACACTTCTAAGTTGTAACAGATTTAGAACACAGAGCACTAAACCAAGAAGAAGAAAAAAGAGAAG
TAGAAGAATCACATAAGCTAAAAAA

ProUGT73F3

TTATCGTAATATATGATCAATATTCTCTATTATGAGTTACAAAAAAAAGAAAAAAAATCAATTTTATTTCGCCAA
CACGTTTTGTAAATAAGATTTAGTTATTAAAGTTATAAAGACTAGCGAAAAGACGGGTACT**CACGTG**CCCGTCTT
TCCGCTCTTTTTATTGCGCTAATGTTTGAAAATTATGAATGATGTTTTTTTTATGAAATTTTGATTTTGATTAAA
ACTAAAAAATATATAAATGTTTGTTTCTTTCAACTTATAAAAATCAACTAACTATGATTATCTATTTCAACAACA
CCCACAATAATACAATATAACAAAAAATATAATCTAAATTCCTTTTTAATGACAGCAACAACAATTTTATTATAG
AAAAAGTTGTTTAAGGATATAATTGGAATGATGAAAAAGTAAGTATATTCATATCGTGTTTGTTACATTTGTTTT
AATATTCAAATTTATTCTTATCTAATTGTTACATGTATTATTTGTATTAAAAAATTTTGGAAAGAAAAATATAAC
CAAAAAATACTGAAAGGGATAGTTCTCTTTATATATAGTATAGATAAAAAATATTTCTTTTTCAAAATAAAAAAC
TCATTTAATTATTAATATAAAATCATATGTATCTGATACATTTAAATTTTCGAGTGTAGCATGTAGTTGTCATTT
GGGAATGCACAACAAATTGTATCTGATACAATTTTTTCTTTTGAAGGATGAAGTTACAATCTTTAATTTAGTATT
AATAGTCGGCATTCAAAACTAAGATGGAAATCTCGCGGAAAAGATTCCTCTC**CTCGTG**GGA**CACGAG**CACATAGA
TCATAGATAAACTAATAACAAAAACAACATATAATATTAAAGAAGTAAATTATTTATGTTAAACCAAAGTATGGT
TGAAATTGAATATATAGAGATGCACTCACAGTGACAAACACAAACATAGAAACGTGAATAGAATAGAATAGAATA
GAGAGAATAATTTGATCCATAAGCA

ProMKB1
TTTACGGGAGGGATTGCATGACTAAAATAGTAGTGTGAATAAAAATAGGTGCAAACTTAATTTTTTATTGAATTA CATTTAATTTAATATAAGATAAATACAATCAAACATCGATCGTCAATAGTTTTGAATTGAGTCTTTTAGTGTTAA GCCTTGTGCTCAATTTTTTTAGCCTATCTTTGTATCGTATCGATGTATTGAGCGTATTGAGCAGGACTTTTAGTT GAGTTGTATCAAAGCATACATCTTAACAAAAACAATATAATACACATATATCTAAATTGATTGATTGCCCATGTC TTTTACTCAGATCATTGATTAAATACGCGATCTTTAAATAAAAATCAAATCAACATAGTAATAACATGAGAATTA AAAATTAATATTTAACAAAGACAATAAACAACACGATTTCCTCAGGATCATATAAGGCAAAGTAACCTTCCCTTG AATTTAATAAGAATGCGAACAACACGGTTTCGCACTCAAGGAATGGAAGCGCAGACAATAGTCAACCTAGACAGT CATGTCATTTGTATATTTTGCGAACGAGTTAAATATTTTGTGCGGGAAAAGAAAACATACACGGTTAAAAAACTA TTTCGATATTTTGTTGATAGTCTTTATGAATCGATTTGATTTACTCGATTATGCGCGAGCAACGTCATTACGTTT ATTCAATCATGGATGTATATTTATTGTGCAGATCGAACAAATGCTTAATATTATTTTTAAATATCTATTATTATT TTGTTTAAAATTTTGATCCTTATTCATTTTATACCCAATACTATTTGATAAAAACGCATGGGTTTAATACTAGTA TATCACCATTGAATATGCTCCTAGAGGTTTCATGGTCGGCATGCTTTTTAAAACATTGATGGATGATTATGATGT TCTAGATGGTGCACGCTTAATTCGTGGTCTCCAACATATATTTTATTTATAAGAAACTACTATTAAAATAAATAT ACTACTATATTTTTTAAGCATAAAA

Supplemental Table S2. Differential peaks identified by LC-MS analysis in TSAR1^{OE} *M. truncatula* hairy roots.
See separate excel file.

Supplemental Table S3. Differential peaks identified by LC-MS analysis in TSAR2^{OE} *M. truncatula* hairy roots.
See separate excel file.

Supplemental Table S4. Number of sequenced reads in the RNA-Seq analysis of *TSAR*-overexpressing *M. truncatula* hairy roots.

Samples	# Sequenced reads
CTR-1	43,159,000
CTR-2	34,485,300
CTR-3	39,390,600
TSAR1^{OE}-1	38,831,500
TSAR1^{OE}-2	40,822,800
TSAR1^{OE}-3	35,581,400
TSAR2^{OE}-1	41,075,700
TSAR2^{OE}-2	40,288,500
TSAR2^{OE}-3	39,596,600

Supplemental Table S5. FPKM values of all differentially expressed genes in *TSAR*-overexpressing *M. truncatula* hairy roots.
See separate excel file with 3 tabs.

Supplemental Table S6. Primers Used.

Primer name	Primer sequence (5'-3')
Cloning	
Genes	
attB1_TSAR1_FW	5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGAGGATTCACTGGAAAATTTG-3′
attB2_TSAR1_RV	5′-GGGGACCACTTTGTACAAGAAAGCTGGGTATCMCAAGTAACTCTCATGAG-3′
attB1_TSAR2_FW	5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGAGGAAATCAACAAC-3′
attB2_TSAR2_RV	5′-GGGGACCACTTTGTACAAGAAAGCTGGGTATTATGATGACGTAAACTTCAATAATG-3′
attB1_Medtr7g117670_FW	5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGGATCAATAAGGCGCATG-3′
attB2_Medtr7g117670_RV	5′-GGGGACCACTTTGTACAAGAAAGCTGGGTATCMTATGGACAAAGTTTGGGTTTG-3′
attB1_Medtr8g027495_FW	5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGAAGATTGAAGTGGGTTTAG-3′
attB2_Medtr8g027495_RV	5′-GGGGACCACTTTGTACAAGAAAGCTGGGTATCMTAAGGAGTTGGTTTCTTTGG-3′
attB1_Medtr2g038040_FW	5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGAGAACATTGGTGATGAGTAC-3′
attB2_Medtr2g038040_RV	5′-GGGGACCACTTTGTACAAGAAAGCTGGGTATCMGATGCTCATAGGGCTTAGAG-3′
attB1_Medtr3g065440_FW	5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGGAAGACAACCTTGTTG-3′
attB2_Medtr3g065440_RV	5′-GGGGACCACTTTGTACAAGAAAGCTGGGTATCMAAATAATCCATAGGCCC-3′
attB1_Medtr8g026960_FW	5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGATGTTTGAATTAGAAAAC-3′
attB2_Medtr8g026960_RV	5′-GGGGACCACTTTGTACAAGAAAGCTGGGTATCMAGACCAAAAGTCCCAC-3′
attB1_Medtr7g116890_FW	5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGGAAATAGCGATGAAG-3′
attB2_Medtr7g116890_RV	5′-GGGGACCACTTTGTACAAGAAAGCTGGGTATCMACCAGCAGCCACAGC-3′
attB1_Medtr7g091390_FW	5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCTTCGTCTTCCTCC-3′
attB2_Medtr7g091390_RV	5′-GGGGACCACTTTGTACAAGAAAGCTGGGTATCMCTCCTGTGGCTGAAACv
attB1_Medtr4g107650_FW	5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCGGGTGGGAGAGTTTTTTC-3′
attB2_Medtr4g107650_RV	5′-GGGGACCACTTTGTACAAGAAAGCTGGGTATCMATAAGACCAAGGCCAAAG-3′
Promoters	
attB1_HMGR1 -1000_FW	5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATTGGTAGGATCAATTGTTG-3′
attB2_HMGR1 -1_RV	5′-GGGGACCACTTTGTACAAGAAAGCTGGGTACTCTCTCTCTCTCTTTCTCTTGG-3′

attB1_HMGR1 -500_FW	5’-GGGGACAAGTTTGTACAAAAAAGCAGGCTTATCAAAAACAAAATTAAACACG-3’
attB1_HMGR1 -281_FW	5’-GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGGTACTGATTACAATATTCTTCC-3’
attB2_HMGR1 -101_RV	5’-GGGGACCACTTTGTACAAGAAAGCTGGGTACTTAACTTGAAAATGAATGCC-3’
attB1_HMGR1 -281_mut_FW	5’-GGTACTGATTACAATATTCTTCCCTCGTTGTTTGAATTTTCCTTACGATCTAGCTATC-3’
attB1_MKB1 -1000_FW	5’-GGGGACAAGTTTGTACAAAAAAGCAGGCTTATTTACGGGAGGGATTGCATGAC-3’
attB2_MKB1 -1_RV	5’-GGGGACCACTTTGTACAAGAAAGCTGGGTATTTTATGCTTAAAAAATATAG-3’
attB1_CYP93E2 -1000_FW	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTAAAAATAATATGTCATTGTAGACATG-3'
attB2_CYP93E2 -1_RV	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTAGGTTGAGTACTTGATTTGG-3'
attB1_CYP93E2 -300_FW	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTATTAGCTGACAAACAACGTCAG-3'
attB2_CYP93E2 -160_RV	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTATTTGAATGGTACTGTTGAGTG-3'
CYP93E2 -210_mut_FW	5'-ATAAGCAACATTCTTATTATAATATTCGACCTATTAGATCGCAAAAATTAATAAC-3'
CYP93E2 -252_mut_RV	5'-TATTATAATAAGAATGTTGCTTATGTGTCCAATTCAGTACATAAAATAATTTC-3'
attB1_CYP72A61v2 -1000_FW	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATAAATCGAGACTTTGTGGAAATG-3'
attB2_CYP72A61v2 -1_RV	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTATGTTTATGTGTGTGTGTGATAG-3'
attB1_CYP716A12 -1000_FW	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTAAGCTAATCATTATATACCTGACC-3'
attB2_CYP716A12 -1_RV	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTAGCTTTTTTTGTGACTAACAAAAC-3'
attB1_CYP72A67 -1000_FW	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTATAAGCAAAAAACTACATTTTAGATTC-3'
attB2_CYP72A67 -1_RV	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTATTCTCTGCTTGTAGGTGATAAAAAG-3'
attB1_UGT73K1 -1000 Fw	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTATTTTATCTATGGAGTGACTTTATG-3'
attB2_UGT73K1 -1_RV	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTATTTTTTAGCTTATGTGATTCTTC-3'
attB1_UGT73F3 -1000 Fw	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTATTATCGTAATATATGATCAATATTC-3'
attB2_UGT73F3 -1_RV	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTATGCTTATGGATCAAATTATTCTC-3'
attB1_CYP72A67 -1500_FW	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTACTGTTTGAAGTTCAAATTAAG-3'
RNAi construct	
attB1_TSAR1_3'UTR_FW	5’-GGGGACAAGTTTGTACAAAAAAGCAGGCTTATCCGTTCTAGTGCATCAAAATAAAG-3’
attB2_TSAR1_3'UTR_RV	5’-GGGGACCACTTTGTACAAGAAAGCTGGGTAACAAAATCAAGTTCTTACTAGC-3’
Y1H promoter construct	
attB4_HMGR1_L1_FW	5’-GGGGACAACTTTGTATAGAAAAGTTGAACCTCGTTGTTCACGAGTTCCTTACGAGATAATCCTGGCCCA-3’

L1_HMGR1_L2_FW	5′-GATAATCCTGGCCCACCTCGTTGTTCACGAGTTCCTTACGATGCTAACATGCTGTG-3′
L1_HMGR1_L2_RV	5′-CACAGCATGTTAGCATCGTAAGGAACTCGTGAACAACGAGGTGGGCCAGGATTATC-3′
attB1R_HMGR1_L2_RV	5′-GGGGACTGCTTTTTTGTACAAACTTGGTCGTAAGGAACTCGTGAACAACGAGGCACAGCATGTTAGCA-3′
qRT-PCR	
TSAR1_FW	5′-TGTGGTGATTATGAGAATGTTGA-3′
TSAR1_RV	5′-AAGGTATGTGGCTGGAGAA-3′
TSAR2_FW	5′-TCAGTTTCAAGTTCCATCTT-3′
TSAR2_FW	5′-AATCGGTTGGAGACAATG-3′
40S_FW	5′-GCCATTGTCGAATTTGATGCTG-3′
40S_RV	5′-TTTTCCTACCAACTTCAAAACACCG-3′
ELFα_FW	5′-ACTGTGCAGTAGTACTTGGTG-3′
ELFα_RV	5′-AAGCTAGGAGGTATTGACAAG-3′
HMGR1_FW	5′-CAGGATTCACAGTCACAACAAC-3′
HMGR1_RV	5′-GTAGACGAAGGAAGCGATGAG-3′
BAS_FW	5′-AGAGGGAAAGAATGAGCCATAC-3′
BAS_RV	5′-CTACCTGCTTCTGGATCATACTC-3′
CYP93E2_FW	5′-ATTGGTGAACTTCTTGGTG-3′
CYP93E2_RV	5′-TCCTTCTTCCTATCACTACC-3′
CYP716A12_FW	5′-AAGGGACAGCATCACCAACAC-3′
CYP716A12_RV	5′-CGCCGAGATATTTGACAAGGAAAG-3′
UGT73K1_FW	5′-CGGATTCTTAACGCATTGTG-3′
UGT73K1_RV	5′-CTCACCACTGTCTTCTTAGC-3′
UGT73F3_FW	5′-CAGCAAGTAACAGTCATCAC-3′
UGT73F3_RV	5′-AGACATAGATTCAATACCTTCAG-3′
MKB1_FW	5′-CTGTGGTCACCTGTATTG-3′
MKB1_RV	5′-CCTGTAGTTTATTGGATTCG-3′