The saponin bomb: a nucleolar liquid-liquid phase-separated β-glucosidase hydrolyses triterpene saponins in Medicago truncatula

Marie-Laure Erffelinck
VIB-Ghent University https://orcid.org/0000-0002-6854-3023
Jan Mertens
VIB-Ghent University
Elia Lacchini
VIB-Ghent University
Francisco Javier Molina-Hidalgo
VIB-Ghent University
Oren Tzfadia
VIB-Ghent University
Pablo Cárdenas
University of Copenhagen
Jacob Pollier
Ghent University
Aldo Tava
CREA Research Centre for Animal Production and Aquaculture https://orcid.org/0000-0003-4657-4053
Søren Bak
University of Copenhagen https://orcid.org/0000-0003-4100-115X
Alain Goossens (alain.goossens@psb.vib-ugent.be)
VIB-Ghent University

Article

Keywords: beta-glucosidase, bioactive triterpenes, jasmonate, specialized metabolism, plant defense

Posted Date: April 4th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1427048/v1

The saponin bomb: a nucleolar liquid-liquid phase-separated β-glucosidase hydrolyses triterpene saponins in *Medicago truncatula*

Marie-Laure Erffelinck[a,b,1], Jan Mertens[a,b,1], Elia Lacchini[a,b,1], Francisco Javier Molina-Hidalgo[a,b,1,2], Oren Tzfadia[a,b,3], Pablo D. Cárdenas[c], Jacob Pollier[a,b], Aldo Tava[d], Søren Bak[c] and Alain Goossens[a,b,*]

[a]Department of Plant Biotechnology and Bioinformatics, Ghent University, B-9052 Ghent, Belgium
[b]VIB Center for Plant Systems Biology, B-9052 Ghent, Belgium
[c]Department of Plant and Environmental Sciences, University of Copenhagen, DK-1871 Frederiksberg C, Denmark
[d]CREA Research Centre for Animal Production and Aquaculture, Lodi 26900, Italy

[1]These authors contributed equally
[2]Current affiliation: Department of Biochemistry and Molecular Biology, University of Córdoba, E-14014 Córdoba, Spain
[3]Current affiliation: Department of Biomedical Sciences, Faculty of Pharmaceutical, Biomedical and Veterinary Sciences, Institute of Tropical Medicine, B-2000 Antwerp, Belgium
*Corresponding author: Alain Goossens, VIB-UGent Center for Plant Systems Biology, Technologiepark 71, B-9052 Ghent, Belgium, 003293313851, alain.goossens@psb-vib.ugent.be

ORCID IDs: 0000-0002-6854-3023 (M.-L.E.); 0000-0002-8095-0748 (J.M.); 0000-0002-1598-8950 (E.L.); 0000-0003-1184-6354 (F.J.M.-H.), 0000-0002-1134-9238 (J.P.); 0000-0002-7096-8063 (O.T.), 0000-0002-4141-2410 (P.D.C.), 0000-0003-4657-4053 (A.T.), 0000-0003-4100-115X (S.B.), 0000-0002-1599-551X (A.G.)

Author Contributions: A.G. conceived and designed the research and supervised the experiments; M.-L.E., J.M., E.L., F.-J. M.-H., O.T., P.D.C., J.P. and A.T. performed the experiments;

M.-L.E., J.M., E.L., F.-J. M.-H., J.P., O.T., P.D.C., S.B. and A.G. designed the experiments and analyzed the data; M.-L.E., J.P., and A.G. wrote the article with contributions of all the authors.

Competing Interest Statement: The authors declare no competing interests.

Keywords: beta-glucosidase, bioactive triterpenes, jasmonate, specialized metabolism, plant defense

Abstract

Plants often protect themselves from their own bioactive defense metabolites by storing them in less active forms. Consequently, plants also need systems allowing correct spatiotemporal reactivation of such metabolites, for instance under pathogen attack. Here we show that the model legume *Medicago truncatula* has evolved a two-component system composed of a β-glucosidase and triterpene saponins, which are physically separated from each other in intact cells. The β-glucosidase, which is stored in the nucleolus and subjected to liquid-liquid phase separation in intact cells, is released and united with its substrates only upon tissue damage, partly mediated by the surfactant action of the saponins themselves. Subsequently, enzymatic removal of carbohydrate groups from the saponins creates a pool of metabolites with an increased broad-spectrum antimicrobial activity. The evolution of this peculiar defense system benefited from both the intrinsic condensation abilities of the enzyme and the bioactivity properties of its substrates.

Introduction

As sessile organisms, plants are unable to flee from biotic and environmental stress situations. Consequently, they arm themselves by producing a plethora of specialized metabolites that function as defense compounds or modulators of abiotic environmental adaptation. These metabolites can be classified based on their biochemical origin but also on their production patterns. As such, a distinction between so-called phytoanticipins and phytoalexins can be made, with the former accumulating constitutively in the plant or in specific plant organs or tissues and acting as a first chemical barrier upon herbivory or pathogen invasion, and the latter being biosynthesized predominantly in response to herbivore or pathogen attack[1].

To protect the plant from its own chemical arsenal, defense metabolites are often stored in a non-active, for instance glycosylated, form[1,2]. Glycosylation also increases the chemical stability and solubility of the defense compounds, allowing their long-term storage, typically in the vacuole[3]. Upon tissue damage by herbivory or pathogen attack, the glycosylated defense compounds can be swiftly activated by hydrolysis of the glycosidic bond. This reaction is catalyzed by β-glucosidases that are stored separately from their substrates in intact plant tissue, but are mixed with their substrates after tissue damage. Together, the glycosylated defense compounds and the β-glucosidases that activate them constitute a two-component system that allows an immediate chemical response in critical stress situations[1,2]. Across the plant kingdom, this two-component system has been reported for different classes of specialized metabolites, the most renowned one being the "mustard oil bomb" commonly used to describe the myrosinase-glucosinolate defense system in Brassicaceae, but similar systems have also been reported for benzoxazinoid glucosides, avenacosides, cyanogenic glucosides and strictosidine derivatives[1,2,4,5].

To prevent autotoxicity, a strict separation of the β-glucosidase and its glycoside substrate is critical. Spatial separation may occur either at the tissue level, in which enzyme and substrate are produced or stored in different tissues, or at the cellular level, in which enzyme and substrate are stored in different cellular compartments [1,2].

Triterpene saponins (TS) are a structurally diverse class of such glycosylated defense metabolites that can be found in a wide variety of plant species and taxa. The structural and functional complexity of this class of specialized metabolites is reflected by their broad range of

bioactivities, among which allelopathic, anti-insect, and anti-microbial properties[6,7]. The model legume *Medicago truncatula* (barrel medic), a member of the Fabaceae plant family and close relative of *M. sativa* (alfalfa), accumulates a complex mixture of TS and has been extensively used to study the organization of TS biosynthesis[8,9]. The first committed steps of TS biosynthesis in *M. truncatula* involve the cyclization of 2,3-oxidosqualene by β-amyrin synthase[10] and the subsequent competing actions of the cytochrome P450-dependent monooxygenases (P450) CYP716A12 and CYP93E2, which cause branching of the pathway towards the production of hemolytic and non-hemolytic TS, respectively. The three-step carboxylation of β-amyrin at the C-28 position by CYP716A12 yields oleanolic acid and represents the first step of the hemolytic TS branch[11,12]. Further oxidations of the oleanolic acid backbone are catalyzed by CYP72A67, CYP72A68, and CYP88A13 and lead to the major hemolytic saponin aglycones medicagenic acid and zanhic acid[13-16]. In the non-hemolytic branch, CYP93E2 catalyzes the hydroxylation of β-amyrin at the C-24 position, thereby excluding oxidation at position C-28[12,17]. Subsequent hydroxylation at the C-22 position by CYP72A61 yields the major non-hemolytic soyasaponin backbone soyasapogenol B[14], which can be further oxidized at the C-21 position towards soyasapogenol A. The produced saponin aglycones are subsequently glycosylated with one or more carbohydrate chains. Both hemolytic and non-hemolytic TS can be glycosylated at the C-3 position, whereas glycosylation at the C-28 carboxylic acid is restricted to the hemolytic TS[15]. In *M. truncatula*, the UDP-dependent glycosyltransferase (UGT) UGT73F3 has been shown to catalyze the glucosylation of hemolytic saponin aglycones at the C-28 position. In *UGT73F3* loss-of-function mutants, both the levels of C-28 glucosylated triterpenes and plant growth decreased, suggesting toxicity of the accumulation of saponins or sapogenins with a free C-28 carboxylic acid moiety[18]. Notably, this corresponds with the outcome of several studies on structure–activity relationships of hemolytic *Medicago spp.* TS, which reported that so-called monodesmoside TS, particularly the hemolytic TS glucosylated at the C-3 position and with a free C-28 carboxylic acid moiety, display the highest bioactivities, also in comparison with their bidesmoside version, in which the C-28 position was glycosylated. This was the case in assays for hemolysis and with plant pathogenic and dermatophytic fungi[19-23].

As defense compounds, TS accumulate constitutively in different *M. truncatula* organs as tissue-specific mixtures of tens of different metabolites[24,25]. Upon pathogen or herbivore attacks, *Medicago* TS biosynthesis is further enhanced[9,26]. Accordingly, since the transcriptional response to pathogen or herbivore attack is controlled by a signaling cascade that involves the phytohormone jasmonate (JA), treatment of *Medicago* plants or cells with JAs leads to the transcriptional activation of TS biosynthetic genes and a subsequently increased TS accumulation[10,27]. In *M. truncatula*, the JA-responsive basic helix-loop-helix (bHLH) transcription factors (TFs) TRITERPENE SAPONIN ACTIVATION REGULATOR1 (TSAR1) 1 and TSAR2 were shown to directly trigger the concerted JA-mediated transcriptional activation of the non-hemolytic and hemolytic TS biosynthesis genes, respectively[28]. The JA signaling cascade also involves posttranslational mechanisms of TS biosynthesis, involving a subset of the endoplasmic reticulum (ER)-associated degradation (ERAD) machinery, more specifically composed of the E3 ubiquitin ligase MAKIBISHI 1 (MKB1) and an interacting chaperone[27,29]. This ERAD machinery controls TS biosynthesis in *M. truncatula* by controlling the levels of the enzyme 3-HYDROXY-3-METHYLGLUTARYL-COA REDUCTASE, which is a rate-limiting enzyme for the biosynthesis of the triterpene precursor isopentenyl pyrophosphate[27].

Here, we expand our knowledge of the organization of TS metabolism in *M. truncatula*, by demonstrating that the root-specific, nucleolar-localized β-glucosidase G1 is controlled by JA and the bHLH TF TSAR2 and specifically hydrolyses hemolytic TS by hydrolyzing C-28 TS conjugates. Because of the spatial separation between the enzyme and its substrates, this only occurs upon tissue damage, thereby concomitantly ensuring self-protection and acting according to a “saponin bomb” model.

Results

MORPH Analysis Reveals the β-Glucosidase-Encoding Gene *Medtr4g015460.* To identify new players in TS biosynthesis and homeostasis in *M. truncatula*, an algorithm called MORPH (MOdule-guided Ranking of candidate PatHway genes) was used[30]. MORPH leverages multiple expression datasets and clustering thereof to prioritize candidate genes involved in certain biological pathways. For this analysis, a set of characterized *M. truncatula* TS biosynthesis genes

(Table S1) was entered as target pathway and two sets of *M. truncatula* expression profiles were used. The first dataset consisted of gene expression values generated by RNA-Seq of *M. truncatula* hairy roots overexpressing *TSAR1* (*TSAR1*^{OE}) or *TSAR2* (*TSAR2*^{OE})[28], which was expanded with expression data of *M. truncatula* hairy roots treated or not with 100 µM of methyl jasmonate (MeJA) for two or 24 h[31]. The second dataset contained all publicly available expression profiles of the *M. truncatula* Gene Expression Atlas[32]. Using machine learning techniques, MORPH selects the best combination of data and analysis methods and outputs a ranking of candidate genes predicted to partake the target pathway (Fig. 1; Table S2). Functional annotation for the candidate genes was retrieved from the PLAZA comparative genomics platform[33]. An area under the curve of the self-ranked genes (AUSR) score of 0.84, representing a high accuracy of ranking, suggests that it is highly probable that the novel candidate genes belong to the given pathway. The candidate list could be subdivided into several groups with predicted functions such as acyltransferases, chaperones, P450s, ion channels, transporters, MVA pathway enzymes, and sugar metabolism enzymes (Fig. 1; Table S2). One gene, *Medtr4g015460*, with a z-score of 2.96, and annotated as a member of the sugar metabolism group, more particularly as a putative β-glucosidase (or glycosyl hydrolase family 1 protein, GH1), caught our attention as a potential candidate TS glycosyl hydrolase.

TSAR2 Controls the Expression of the MeJA-Responsive G1-Encoding *Medtr4g015460.* The β-glucosidase encoded by *Medtr4g015460* was previously identified as G1, one of four MeJA-inducible *M. truncatula* β-glucosidases in a transcriptome analysis of *M. truncatula* cell cultures treated with different elicitors[34]. Among these four β-glucosidases, the expression of *Medtr4g015460* was boosted the strongest by MeJA treatment, but the physiological target of G1 remained hitherto unknown[34]. Recently, G1 was labeled as *MtBGLU18* in a comprehensive identification of *GH1* genes in *M. truncatula*[35]. In agreement with the observations made byNaoumkina, et al.[34], RNA-Seq analysis revealed that *Medtr4g015460* transcript levels were increased 14- and 57-fold in *M. truncatula* hairy root lines treated with 100 µM of MeJA for two or 24 h, respectively (Fig. S1A). Furthermore, *Medtr4g015460* transcript levels were increased in *TSAR2*^{OE} lines, but not in *TSAR1*^{OE} lines (Fig. S1B), implying that the expression of *Medtr4g015460*

is controlled by TSAR2. To further support this, additional independent *TSAR1*^{OE}, *TSAR2*^{OE}, and control *M. truncatula* hairy root lines overexpressing β-glucuronidase (GUS) were made. Quantitative reverse transcription-PCR (qRT-PCR) analysis confirmed elevated expression levels of *Medtr4g015460* in the *TSAR2*^{OE} lines, whereas in the *TSAR1*^{OE} lines, *Medtr4g015460* transcripts were not significantly induced (Fig. S1C-E). Overall, this indicates that TSAR2 positively affects the expression of *Medtr4g015460* in *M. truncatula* hairy roots. Corroborating this, when probing the *M. truncatula* gene expression atlas[32], correlated expression between *Medtr4g015460* and *TSAR2* was observed in roots and suspension cells when exposed to various stresses and MeJA treatment (Pearson's correlation coefficient = 0.83; Fig. S2). Furthermore, like *TSAR2*, *Medtr4g015460* is specifically expressed in subterranean tissues (Fig. S2). Since *TSAR2* overexpression only affects the expression of genes associated with the production of hemolytic saponins[28], it was hypothesized that the β-glucosidase G1 might specifically act on hemolytic saponins in roots.

Phylogenetic Analysis. To investigate the evolutionary relationship between G1 and other β-glucosidases that function to bioactivate inert defense compounds, we carried out a phylogenetic analysis (Fig. S3). To this selection, we added the three other identified GH1 β-glucosidases from *M. truncatula*[34], and a number of β-glucosidases that activate different inert defense compounds from diverse species (see the legend of Fig. S3 for details). As was reported before[2], β-glucosidases from monocots and dicots are reliably separated from each other, indicating an independent evolutionary development. It was postulated that isoflavonoid glucoside β-glucosidases in legumes would have evolved from cyanogenic β-glucosidases. G2, which most likely deglycosylates isoflavonoids[34], resides indeed in the same clade as an isoflavone conjugate-hydrolyzing β-glucosidase from *Glycine max* and many cyanogenic glucoside glucosidases. Notably, G1 also confines to this clade. The targets of G3 and G4 are yet unknown but both of them are closely related with a cyanogenic glucoside (CG)-hydrolyzing β-glucosidase. This suggests that both isoflavonoid- as well as our candidate TS-hydrolyzing β-glucosidases share a common origin with CG-hydrolyzing β-glucosidases. Furthermore, G1 is evolutionary not related to the *Avena sativa* (oat) β-glucosidases, which target the avenacosides, steroidal saponins that are derived from phytosterols[36,37] and can be considered as biochemically related to TS.

G1 Acts on Hemolytic Saponins by Removing the Glucosyl Residue at the C-28 Position. In a previous study, recombinant G1 produced in *Pichia pastoris* was tested with flavonoids and isoflavonoids as potential substrates[34]. However, very high K_m and low K_{cat}/K_m values indicated that G1 may have a different, albeit unknown target. To determine the biochemical activity of G1, we produced the protein in *Escherichia coli* and incubated it with a methanolic extract of *M. truncatula* hairy roots. As a control, heat-denatured G1 was used. LC-MS analysis revealed that several metabolites disappeared and appeared in the extract treated with G1 (Fig. 2). Based on their accurate mass and MS^n fragmentation spectra, two of these metabolites were identified as 3-Glc-28-Glc-medicagenic acid and 3-Glc-malonyl-28-Glc-medicagenic acid. The identity of the former metabolite was confirmed with an authentic standard. In addition to these metabolites, several other peaks corresponding to triterpene saponins disappeared in the chromatogram of the G1-treated hairy roots extract (Fig. 2B). However, the MS^n spectra of these metabolites did not allow us to determine the position of the carbohydrates on the oleanane backbones. In addition to the peaks that disappeared, several peaks appeared more abundant in the G1-treated hairy roots extract (Fig. 2C). Two of these metabolites were identified as 3-Glc-medicagenic acid and 3-Glc-malonyl-medicagenic acid. The amounts of soyasaponin I, the most abundant triterpene saponin in *M. truncatula* hairy roots, remained unaltered (Fig. 2C).

In the previous study, recombinant G1 showed no activity on the isoflavonoids ononin and genistin[34]. We did not detect these metabolites in our *M. truncatula* hairy roots extract, however, malonyl-ononin was detected (Fig. 2). In agreement with the previous study, this metabolite was found not to be targeted by G1.Naoumkina, et al.[34] also reported activity of G1 on the flavone apigenin-7-*O*-glucoside. Again, we were not able to detect this metabolite in our *M. truncatula* hairy roots extract. Nonetheless, several other flavonoids, eluting between 12 and 22 min, were detected, but were unaltered by G1 (Fig. 2). Taken together, these data suggest that G1 specifically targets hemolytic TS, where it catalyzes the hydrolysis of the ester bond between the C-28 carboxyl group and the glucose moiety at that position. To further substantiate this, we incubated G1 with several TS standards. Glucose esterified at the C-28 position of 3-Glc-28-Glc-medicagenic acid and 3-Glc-Ara-28-Glc-hederagenin was completely removed, leading to the

monodesmosidic 3-Glc-medicagenic acid and 3-Glc-Ara-hederagenin, respectively (Fig. 3A-B). Conversely, 3-Glc-medicagenic acid itself and another monoglucosylated TS, 3-Glc-echinocystic acid, were not affected by G1 (Fig. 3C-D).

G1 is Localized in the Nucleolus. Previously, a nuclear localization was shown for an eGFP-tagged version of G1 in leaves and cell suspension cultures of *M. truncatula* and in tobacco leaves. Removal of the 108 N-terminal amino acids resulted in a cytoplasmic localization, supporting a nuclear targeting role for this N-terminal extension[34]. *In silico* analysis of the G1 sequence with the Localizer tool[38] corroborated these observations and revealed the nuclear localization signal (NLS) at the N-terminus (PPPKRKR between amino acid positions 8 and 15, Fig. S4). However, additional analysis of the G1 sequence using the nucleolar localization sequence detector[39] suggested that an extended version of this NLS, comprising the 24 N-terminal amino acids, could also correspond to a nucleolar localization sequence (Fig. S4). We therefore decided to further assess subcellular localization of G1 specifically in *M. truncatula* roots, where *G1* is normally expressed. We generated both N- and C-terminal translational fusions of G1 with the Venus fluorescent protein that we first produced in *E. coli* and used for activity assays. Both N- and C-terminal fusion proteins hydrolyzed the ester bond between the glucose moiety and the C-28 carboxyl group of 3-Glc-28-Glc-medicagenic acid, thereby yielding 3-Glc-medicagenic acid (Fig. S5A-B) and confirming functionality of the fusion proteins. As both translational fusions lead to functional proteins, transgenic *M. truncatula* hairy root lines expressing the N- and C-terminal Venus fusions were generated. In accordance with previous observations byNaoumkina, et al.[34], G1 indeed appeared to be localized in the nucleus (Fig. 4 and Fig. S5C). However, confocal microscopy analysis of the *M. truncatula* hairy root lines expressing the Venus fusions treated with the nuclear DNA stain 4',6-diamidino-2-phenylindole (DAPI) clearly revealed that within the nucleus, the protein was specifically confined to the nucleolus, corroborating the *in silico* analysis. MeJA treatment did not influence the localization of the fusion protein (Fig. 4 and Fig. S5C), indicative of a strict physical separation of the vacuolar TS and the nuclear G1 protein in intact root cells.

G1 is a Highly Efficient and Fast β-Glucosidase. By kinetic analysis using the glucose oxidase-peroxidase (GOD-POD) assay we assigned a K_m value of 154 µM, a K_{cat} value of 4.05 s^{-1} and a very high K_{cat}/K_m ratio of 26299 $s^{-1}M^{-1}$ for G1 with the hemolytic TS 3-Glc-28-Glc-medicagenic acid (Fig. S6).Naoumkina, et al.[34] reported K_{cat}/K_m ratios of 174.4 and 2528.4 $s^{-1}M^{-1}$ with the artificial substrate p-nitrophenyl glucoside (pNPG) and the flavone apigenin-7-O-glucoside, respectively. However, *in vitro* activity assays for G1 performed here, with a methanolic metabolite extract of *M. truncatula* hairy roots, showed no activity on flavonoids or on non-hemolytic TSs. Moreover, a G1 preference for hemolytic saponins, represented by the tremendous K_{cat}/K_m ratio for 3-Glc-28-Glc-medicagenic acid, correlates with a possible physiological role of TS bioactivation upon biotic attack. Accordingly, the pH optimum of G1, pH 6.0 (Fig. S6), is substantially lower than the pH of the nucleolus (typically pH 7.2) where it resides. We hypothesize that cellular disruption caused, for instance by herbivore or pathogen attack, results in a massive overall cellular pH drop, mainly due to vacuolar rupturing ($pH_{vacuole}$ 5.5), setting the stage for G1 to hydrolyze released hemolytic TSs.

G1 Catalyzes Hydrolysis of Hemolytic Saponins Following Tissue Damage. Hence, the most eminent questions that remained to be resolved were (i) whether G1 can get in touch with its TS substrates *in planta* upon cellular damage, (ii) if so, whether G1 then catalyzes hemolytic TS hydrolysis *in planta*, and (iii) whether G1 is the only *M. truncatula* β-glucosidase doing so *in planta*. To allow addressing all of these questions at once, we needed a loss-of-function *M. truncatula G1* line, which then could be compared with the wild-type. Therefore, we first mined the *M. truncatula* Tnt1 retrotransposon insertion population[40] for *G1* loss-of-function mutants. From the Tnt1 flanking sequence tag (FST) database (https://medicago-mutant.dasnr.okstate.edu/mutant/database.php), three candidate mutant lines, NF7894, NF4760, and NF3037were selected, but unfortunately, none of them turned out to be a genuine mutant for *G1*. Hence, to allow investigating the *in planta* role of G1, we generated three independent *g1* knock-out (*g1*-KO) hairy root lines using CRISPR/Cas9-mediated genome editing (Fig. S7).

Having the appropriate *M. truncatula* genotypes at hand, we next designed a robust and reproducible experimental set-up, which could serve as a proxy for tissue damage by attackers, and yet enable us to obtain sufficient material for phenotyping. A very gentle and short 'mashing' of *M. truncatula* hairy roots was found to be appropriate to generate damaged *M. truncatula* tissue for metabolite profiling. Indeed, when we specifically profiled 3-Glc-28-Glc-medicagenic acid (the G1 substrate) and 3-Glc-medicagenic acid (the G1 product) in control lines, we observed a concomitant decrease and increase in the G1 substrate and product after mashing, respectively (Fig. 5), indicating that hydrolysis at the C28 position of the saponin takes in place *in planta* upon tissue damage.

Given the proven suitability of this experimental design, we proceeded with the profiling of the *g1*-KO lines, as well as of the G1-Venus lines, which can be considered as *G1* overexpression lines. Notably, the shift in the G1 substrate/product pattern upon mashing was totally abolished in the *g1*-KO and slightly strengthened in the G1-Venus lines (Fig. 5). We could also observe an increase and decrease in 3-Glc-28-Glc-medicagenic acid and 3-Glc-medicagenic acid, respectively, already in the unmashed *g1*-KO, as compared to intact control lines. This may point to some endogenously occurring G1-mediated TS hydrolysis, possibly either because of endogenous cellular turnover or of our general experimental handlings. Further, we assume that 3-Glc-medicagenic acid may also be detectable at lower levels in intact roots because it is also an intermediate towards the biosynthesis of multiple glycosylated medicagenic acid-derived TS. Taken together, our data confirm that G1 catalyzes the hydrolysis of hemolytic TS following tissue damage, and importantly, seems the sole *M. truncatula* β-glucosidase responsible for such hydrolytic activity.

Nucleolar Storage of G1 Involves Liquid-Liquid Phase separation. In view of all of the above, the nucleolar localization intrigued us. Indeed, the nucleolus is considered as a membrane-less organelle sequestered from the nucleus by liquid droplet formation through a liquid-liquid phase separation (LLPS)[41]. LLPS can be broken by emulsifiers, including surface-active agents. Given that TS are such agents, we therefore reasoned that tissue damage would possibly not only unite substrate and enzyme, but also release LLPS-wise condensed G1. First, we assessed whether G1

is subject to LLPS. To this end, we treated *G1-Venus* expressing *M. truncatula* roots with 1,6-hexanediol, the most widely used tool to probe LLPS in cells. Indeed, 1,6-hexanediol treatment triggered partial release of G1-Venus from the nucleolus, more precisely towards the whole nucleus (Fig. 6), indicating G1 is stored in LLPS condensates. Plausibly, the NLS still safeguards the nuclear localization of G1 in the presence of 1,6-hexanediol. Next, we wanted to assess whether TS could also break the G1 LLPS condensates. Unfortunately, TS mixes from *M. truncatula* roots were not readily available, hence we used a commercially available mix of TS extracted from *Quillaja saponaria* (soap bark tree), to serve as a proxy for the endogenous *M. truncatula* TS. Notably, the *Q. saponaria* TS caused a similar partial release from the nucleolus as observed with 1,6-hexanediol (Fig. 6). This suggests that close proximity of endogenous *M. truncatula* TS could indeed release G1 from the nucleolus, in case such TS would be released themselves from the vacuole, their presumed storage site, upon tissue damage. To assess the latter hypothesis, we also scored G1 localization following mashing of the roots as described above. Though confocal analysis of cells in damaged roots was less robust, also the mashing treatment seemed to promote partial G1 release from the nucleolus (Fig. 6), supporting our model in which vacuolar release of endogenous TS triggers G1 decondensation, leading to the hydrolysis of *M. truncatula* hemolytic TS.

Discussion

Triterpenes are ubiquitous metabolites in the plant kingdom, with a wide range of structures that are produced in an organ-, species- and/or taxa-specific manner. Accordingly, triterpenes display an enormous range of bioactivities, much of which find applications in our daily lives[6,7]. Exemplary are legumes such as *Medicago spp.*, which accumulate large amounts of TS, typically in organ-specific blends of tens of different TS[8,9,25]. Several reports pointed towards the bioactivity of the hemolytic *Medicago* TS, particularly the medicagenic acid glycosides, as strong and broad-spectrum antimicrobial compounds[25,42]. Furthermore, when comparing activities of so-called monodesmosidic and bidesmosidic compounds within that subclass, typically the former, *i.e.* those with a free C-28 carboxylic acid moiety and only glucosylated at the C3 position, present the highest anti-fungal activity[19,21-23]. Yet, such monodesmosidic TS are not the most abundant TS

in *M. truncatula* tissues. Indeed, it was previously reported that *M. truncatula* roots store considerable amounts of bidesmosidic 3-Glc-28-Glc-medicagenic acid and only low amounts of 3-Glc-medicagenic acid[24]. Likewise, 3-Glc-28-Glc-medicagenic acid has been shown to be the second most abundant TS in *M. truncatula* hairy roots and, additionally, more C-28 glucose-conjugated TS have been identified in hairy roots[43]. Also aerial organs harbor high levels of bidesmosidic zanhic acid and medicagenic acid glycosides, of which the majority is conjugated with an Ara-Rha-Xyl carbohydrate chain at C-28[24,25,44,45]. Zanhic acid glycosides are not produced in *M. truncatula* roots, principally because *CYP88A13*, the gene encoding the cytochrome P450 that catalyzes the C-16a hydroxylation of medicagenic acid toward zanhic acid, is not expressed in *M. truncatula* roots[15]. Unlike medicagenic acid glycosides, zanhic acid glycosides do not possess a broad and strong anti-microbial activity and are generally considered as antinutritional and deterrent agents against herbivores in leaves instead[25,42].

Taken together, available TS profiling data indicate that the hemolytic TS with strongest antimicrobial activity are stored in a form with submaximal potency in *M. truncatula* organs. As such, these TS conform well to the definition of so-called phytoanticipins, which are plant specialized metabolites that act in the chemical defense against pathogens or herbivores, and that are produced constitutively and stored in specific organs or organelles, eventually in an inert form, to alleviate autotoxicity to the plant. From the above, it appears that also *M. truncatula* may have mounted such a double protection barrier for the medicagenic acid glycosides, namely storage in the vacuole in a less active bidesmosidic form. The need for such a strong barrier seems justified, because it is perhaps substantiated by the fact that the loss-of-function plants for *UGT73F3*, the gene encoding the UGT that catalyzes the glucosylation of hemolytic saponin aglycones at the C-28 position, show aberrant or retarded root growth[18]. Importantly, such phytoanticipin-based defense systems often involve a second, activating component, for instance a β-glucosidase. Indeed, in adverse conditions, for instance herbivore or pathogen attack, β-glucosidases, which are typically compartmentally separated from their phytoanticipin substrates, bioactivate them by removing glucosyl moieties, rendering them more harmful towards the attackers. Such a binary system has been reported for several specialized metabolite classes[2], but to date not for legume TS.

In this study, we solved this elusive gap by reporting on the discovery and characterization of the β-glucosidase G1 in *M. truncatula*. G1 is a GH1 family member that specifically hydrolyzes the ester bond between the C-28 carboxyl group and the corresponding glucose moiety at that position of the hemolytic TS. In contrast to β-glucosidases that hydrolyze glycosidic bonds, only a few β-glucosidases that catalyze glucosyl ester hydrolysis have been reported to date. These include the *Oryza sativa* (rice) and *Hordeum vulgare* (barley) β-glucosidases that hydrolyze gibberellin and abscisic acid glucosyl esters, respectively[46,47]. GH1 enzymes contain two essential glutamate (Glu) residues in the conserved TFNEP and I/VTENG motifs, which serve essential roles in the two-step enzymatic hydrolysis of the glycosidic bond. In the initial glycosylation step, the Glu of the I/VTENG motif serves as a nucleophile and performs a nucleophilic attack on the anomeric carbon, leading to a glucose enzyme intermediate. In the subsequent deglycosylation step, the Glu from the TFNEP motif serves as a base catalyst for the activation of a water molecule that acts as a nucleophile and breaks the glycosidic bond, thereby releasing the glucose molecule[2,46,47]. As both TFNEP and I/VTENG motifs are also conserved in G1 (Fig. S4), a similar reaction mechanism for glucosyl ester hydrolysis by G1 can be assumed.

Notably, expression of *G1* is root specific, as is that of *UGT73F3*[18]. This pattern seems to correlate also with the root-specific accumulation of medicagenic acid conjugates, with Glc conjugated to both the C-3 and C-28 positions, whereas aerial organs primarily bear glycosides with GlcA and an Ara-Rha-Xyl carbohydrate chain at the C-3 and C-28 positions, respectively[24,44,45]. Whether G1 may also catalyze hydrolysis of such TS, remains to be determined. Remarkably also, the expression of *G1* is inducible by MeJA and is controlled by TSAR2, the transcriptional regulator of hemolytic TS biosynthesis in *M. truncatula* roots[28]. As such, G1 forms clearly an integral part of the transcriptional regulon that monitors the stress-mediated elicitation of this defense element in *M. truncatula*.

The binary systems that depend on β-glucosidases typically involve physical separation of the two components. This is also the case for G1, which localizes to the nucleolus, and is thereby well disconnected from its substrates, which accumulate in the vacuole. This separation is not disrupted by mere elicitation of TS biosynthesis, for instance by treatment with JA, despite that the JA signaling cascade increases the production of both G1 and the TS biosynthesis enzymes,

and thus the G1 substrates. Consequently, G1 can only hydrolyze hemolytic TS following cell disruption, caused for instance by chewing insects or protruding pathogens, which we mimicked by a mashing experiment in this study. Therefore, we dub the *M. truncatula* two-component system composed of G1 and the hemolytic TS as the saponin bomb, in analogy to the renowned mustard oil and cyanide bombs, respectively composed of glucosinolates and myrosinases found in Brassicaceae species and cyanogenic glucosides and their β-glucosidases found in thousands of different plant species across the plant kingdom[2]. However, to our knowledge, a nucleolar localization of the β-glucosidase in such a two-component system has not been reported to date.

The closest similar subcellular localization of a β-glucosidase was that of strictosidine β-D-glucosidase (SGD) in *Catharanthus roseus* (Madagascar periwinkle), which hydrolyses the glucose moiety of strictosidine. The resulting unstable aglycone is rapidly converted into a highly reactive dialdehyde, from which numerous bioactive monoterpene indole alkaloids can be derived, but which, because of its toxicity, also could be a potential herbivore deterrent or defense chemical against necrotrophic pathogens itself[4]. The SGD sequence was found to contain an NLS. Accordingly, SGD was shown to accumulate as highly stable supramolecular aggregates within the nucleus of *C. roseus* cells, but not in the nucleolus[4]. Notably, the tertiary structure of plant β-glucosidases belonging to the GH1 family is highly conserved, and oligomerization of β-glucosidases, as well as interaction with aggregating factors and other binding proteins has being reported[2]. This may explain the occurrence of the SGD aggregates in the nucleus of *C. roseus* cells. *M. truncatula* G1 also contains a 7-amino acid NLS in its N-terminus, which, in an extended form of 24 amino acids, may also correspond to a nucleolar targeting signal. We corroborated the latter by confocal microscopy analysis of transformed *M. truncatula* roots expressing Venus-tagged G1. The nucleolar localization of G1 is particularly intriguing within a two-component system that involves TS. Indeed, the nucleolus is considered as a membrane-less organelle sequestered from the nucleus by liquid droplet formation through LLPS[41]. LLPS can be broken by emulsifiers, including surface-active agents, such as the TS. Accordingly, we could demonstrate that TS promote the release of G1 from the LLPS condensates. The fact that we could only observe a partial G1 decondensation and that G1 still remained in the nucleus following TS treatment could be because the TS treatment in our experiments may not have been efficient enough or its effect

only transient. It may however also be due to the NLS and/or other sequence stretches in the G1 structure that endorse the reported intrinsic capacity of the GH1-family β-glucosidases to oligomerize and/or aggregate[2]. LLPS or, accordingly, the formation of so-called liquid droplets, allows condensation of cellular material into membrane-less organelles to tune biochemical reactions, improve cellular fitness during stress, and enable diverse dynamic roles of the nucleolus, among others[41,48]. Very recently, LLPS events have been shown to play important roles in chloroplast cargo sorting, photoregulation, circadian clock, thermosensing and immunity programs in plants as well[49-51]. Here, we report on yet another plant LLPS case, in which a classical two-component defense system composed of a β-glucosidase and its bioactive substrates evolved to benefit from both the intrinsic condensation abilities of the enzyme and the bioactivity (surfactant) properties of its substrates.

Methods

MORPH Analysis. MORPH tests multiple combinations of expression datasets and gene clusters, identifies the best combination using cross-validation, and ranks genes in terms of strength of the evidence that they belong to the target pathway. Hence, MORPH builds upon co-expression analysis and finds the best combination of gene expression data and network information to assess the candidates of a specific pathway. We define a pathway's learning configuration as a combination of gene expression dataset and clustering method. The choice of expression data and a specific clustering algorithm (possibly employing a network) affects the results of the analysis. For each dataset, six clustering options are considered: no clustering, cluster identification via connectivity kernels (CLICK), Markov clustering (MCL) with protein-protein interaction (PPI), or metabolic dependencies (MD) networks, and enzyme/non-enzyme partition[30].

Generation and Cultivation of Transgenic *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 and cultivation of transgenic hairy roots were carried out as previously described[43].

Hairy roots were cultivated for 21 days in liquid medium, harvested and flash-frozen in liquid nitrogen to provide suitable and sufficient material for RNA or metabolite extraction. For the metabolite profiling of mashed roots, hairy roots were further propagated in an optimized *in vitro* upscaling system. Growing cultures were transferred to 1-L, wide-mouth bottles with membrane caps (Duran Group GmbH, Meinz, Germany) containing 100 mL of Murashige and Skoog (MS) medium. Inoculated bottles were placed on an orbital shaker (130 rpm) at 25°C in darkness. The volume of the medium was doubled weekly by adding MS medium until a final culturing volume of 800 mL. One month after inoculation of the bottles, the hairy roots were harvested, and either gently mashed using a pestle in a mortar or not (mock-treated), frozen in liquid nitrogen and stored at -70°C.

RNA-Seq Analysis. RNA-Seq analysis of control, *TSAR1*^{OE}, and *TSAR2*^{OE} lines was reported before[28]. RNA-Seq analysis of control hairy root lines treated with 100 µM of MeJA or an equivalent amount of ethanol for two or 24 h was carried out as described[28,31]. The reported RNA-Seq read data are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) under accession numbers E-MTAB-3532 (control, *TSAR1*^{OE}, and *TSAR2*^{OE} lines) and E-MTAB-8697 (control hairy roots treated with MeJA for two or 24 h).

qRT-PCR Analysis

Hairy roots grown for 21 days in liquid medium were harvested by flash freezing in liquid nitrogen. Harvested hairy roots were ground under liquid nitrogen with mortar and pestle and used for total RNA extraction and first-strand cDNA synthesis with the RNeasy Mini Kit (Qiagen) and the iScript cDNA Synthesis Kit (Bio-Rad), respectively. qRT-PCR was carried out with a LightCycler 480 system (Roche), SYBR Green Master Mix (Thermo Fisher Scientific) and qRT-PCR primers (Table S3) were designed using Beacon Designer 4 (Premier Biosoft International). For reference genes, the *M. truncatula 40S ribosomal protein S8* (*40S*) and *translation elongation factor 1α* (*ELF1α*) were used. Reactions were done in triplicate and qBase was used for relative quantification with multiple reference genes[52].

DNA Constructs. Using the primer pair COMBI3938-COMBI3939, the 1,521-bp open reading frame of G1 (GenBank accession number EU078901;[34]) was PCR-amplified from cDNA of *TSAR2*^{OE} lines[28] and Gateway recombined into pDONR221. Constructs encoding N- and C-terminal fusions of G1 with the VENUS fluorescent protein[53] were generated using overlap extension PCR and Gateway recombined into pDONR221. Sequence-verified entry clones were Gateway recombined with the destination vectors pK7WG2D or pK7WG2[54] for expression in hairy roots, and with pDEST17[55] for recombinant protein production. All primers used for cloning are reported in Table S3.

CRISPR/Cas9 Gene Editing in *M. truncatula* Hairy Roots. Targeted genome modifications in *M. truncatula* hairy roots with CRISPR/Cas9 were obtained according to a method developed for soybean[56]. Guide RNA (gRNA) targets with the $GN_{19-20}GG$ motif were identified using the CRISPR-P 2.0 tool (http://crispr.hzau.edu.cn/CRISPR2/) for genome editing in plants[57]. For each gene, three high-scoring gRNAs were selected within the first half of the coding sequence. For gRNA cloning, forward and reverse primers (Table S3) were designed with overlapping tails that encode the target sequence and used to amplify the entire pUC gRNA Shuttle vector (AddGene plasmid #47024) that contains the *M. truncatula U6.6* promoter fused to the gRNA scaffold, flanked by I-*PpoI* restriction sites. The resulting PCR products were used for an In-Fusion® cloning reaction (Clontech) according to the manufacturer's instructions. The resulting plasmids were Sanger-sequenced with the M13-reverse primer to confirm the sequence of the gRNA constructs. Sequence-verified gRNA constructs were cut from the pUC gRNA Shuttle vector using the I-*PpoI* restriction enzyme and ligated into the p201G:Cas9 vector (AddGene plasmid #59178) that contains *CAS9*, the *M. truncatula* U6.6 polymerase III promoter for efficient transcription in *M. truncatula,* and *GFP* for selection of hairy roots. The obtained plasmids were sequence-verified using the MtU6 primers (Table S3) and correct clones were transferred to the *A. rhizogenes* strain LBA 9402/12 for hairy root transformation. After transformation, GFP-expressing hairy roots were selected, propagated on solid medium without antibiotics and finally cultured in liquid MS medium for ca. 3 weeks. Upon harvest, the hairy roots were frozen and ground in liquid nitrogen. Genomic DNA was extracted as described[15] and used to PCR-amplify the gRNA target sites using

Q5® High-Fidelity DNA Polymerase (New England Biolabs) and the primers given in Table S3. The obtained PCR fragments were run on a 2% agarose gel, purified using the GeneJET Gel Extraction Kit (Thermo Scientific) and Sanger-sequenced with the same primers used for PCR amplification. The resulting sequences were analyzed using the TIDE website tool (http://tide.nki.nl), which quantifies the efficiency of genome editing and identifies the predominant types of insertions and deletions in the DNA[58].

Protein Expression and Purification. pDEST17 carrying the full-length sequence of *G1* was transformed into One Shot® BL21 Star™ competent cells. An overnight culture at 37°C was added to 1.5 L LB medium, which was grown at 37°C until bacterial growth had reached its mid-exponential phase. Protein expression was induced by isopropyl β-D-1-thiogalactopyranoside (IPTG) and the cells were grown for 16 h at 18°C. Cells were centrifuged and the pellet was dissolved in lysis buffer containing 50 mM Tris, 5 mM imidazole, 500 mM NaCl, 1 cOmplete™ ULTRA Tablet, EDTA-free Protease Inhibitor Cocktail (Sigma-Aldrich) and 0.1% triton (pH 8). The sample was sonicated, followed by centrifugation for 20 min at maximum speed. The supernatant was filtered using 0,22-µm syringe filters. G1 was purified by its N-terminally fused HIS_6-tag using immobilized metal ion chromatography (IMAC). A Bio-Spin® chromatography column (Bio-Rad) was filled with Ni-Nta Superflow resin (QIAGEN) and washed with water and washing buffer (50 mM Tris (pH 8), 250 mM NaCl, 5 mM imidazole). Next, the sample was applied, the column was washed with buffer containing 60 mM imidazole and finally G1 was eluted using buffer containing 400 mM imidazole. The sample was dialyzed in Tris buffer (20 mM Tris, 200 mM NaCl, pH 7.5) using Spectra/Por®Dialysis Membrane. The presence of protein was verified by SDS-PAGE and western blot and final protein concentration was calculated using the Bradford protein assay (Bio-Rad). Aliquots were diluted with 20% glycerol and stored at -70°C. The whole procedure yielded about 10 mg of protein from 3 L of *E. coli* culture. Likewise, recombinant N- and C-terminal VENUS fusions of G1 were made. The total yield of VENUS-G1 and G1-VENUS was 7.54 and 1.73 mg, respectively, from 1.5 L of *E. coli* culture.

Glucosidase Activity Assays. The hydrolyzing activity of G1 was assessed on hairy root extracts prepared as described[43] using pure saponin standards generated previously: 3-Glc-medicagenic acid, 3-Glc-28-Glc-medicagenic acid, and 3-Glc-Ara-28-Glc-hederagenin. 3-Glc-echinocystic acid was obtained from Extrasynthese. A mixture of 50 µg G1 in 150 µL hairy root extract (dissolved in 20 mM Tris, 200 mM NaCl, pH 7.5) was prepared. To prepare the samples containing pure saponins, 50 µg of G1 was added to a mixture containing 20 µM saponin, 20 mM Tris and 200 mM NaCl (pH 7.5). The samples were incubated at 25°C for 2 h, after which the samples were boiled for 5 min. As a control, samples were incubated with G1 that was boiled preceding the incubation period. Salt was removed using C18 solid phase extraction columns (Thermo Scientific™, HyperSep™ C18 Cartridges) according to the following sequence of steps: washing with 3 mL of methanol, 3 mL of water, application of the sample, washing with 3 mL of water and finally elution with 1 mL of methanol. Next, the eluate was evaporated to dryness under vacuum and finally resuspended in 200 µL of water for LC-MS analysis.

LC-MS Analysis. All LC-MS analysis of samples generated with recombinant G1 protein was carried out as described[15]. Analysis of *M. truncatula* hairy roots samples was carried out as follows. Samples were subjected to Ultra Performance Liquid Chromatography High Resolution Mass Spectrometry (UPLC-HRMS) at the VIB Metabolomics Core Ghent (VIB-MCG). 10 µl was injected on a Waters Acquity UHPLC device connected to a Vion HDMS Q-TOF mass spectrometer (Waters, Manchester, UK). Chromatographic separation was carried out on an ACQUITY UPLC BEH C18 (150 × 2.1 mm, 1.7 µm) column (Waters, USA), column temperature was maintained at 40°C. A gradient of two buffers was used for separation: buffer A (99:1:0.1 water:acetonitrile:formic acid, pH 3) and buffer B (99:1:0.1 acetonitrile:water:formic acid, pH 3), as follows: 99% A for 0 min decreased to 50% A in 30 min, decreased to 30% from 30 to 35 min, and decreased to 0% from 35 to 37 min. The flow rate was set to 0.35 mL min^{-1}. Electrospray Ionization (ESI) was applied, LockSpray ion source was operated in negative ionization mode under the following specific conditions: capillary voltage, 2.5 kV; reference capillary voltage, 3 kV; source temperature, 120°C; desolvation gas temperature, 550°C; desolvation gas flow, 800 L h^{-1}; and cone gas flow, 50 L h^{-1}. The collision energy for full MS scan was set at 6 eV for low energy settings, for high energy

settings (HDMSe) it was ramped from 20 to 70 eV. Mass range was set from 50 to 1500 Da, scan time was set at 0.1s. Nitrogen (greater than 99.5%) was employed as desolvation and cone gas. Leucine-enkephalin (100 pg µL−1 solubilized in water:acetonitrile 1:1 [v/v], with 0.1% formic acid) was used for the lock mass calibration, with scanning every 2 min at a scan time of 0.1 s. Profile data was recorded through Unifi Workstation v2.0 (Waters). Data processing was done with Progenesis QI v2.4 (Waters).

DAPI Staining and Confocal Imaging. Pieces of *M. truncatula* hairy roots were fixed using 4% paraformaldehyde (Sigma-Aldrich) in phosphate-buffered saline (PBS, pH 7.4) for 45 min under vacuum, and subsequently washed with MilliQ water for 10 min. The samples were stained with 2.5-µg/ml 2-(4-amidinophenyl)-6-indolecarbamidine dihydrochloride (DAPI; Sigma-Aldrich) in PBS (pH 7.4) for 10 min, followed by washing with MilliQ water for 15 min. The root pieces were mounted in MilliQ water on a microscope slide for confocal imaging. The samples were imaged using a Zeiss LSM 5 Exciter confocal laser scanning microscope equipped with a C-Apochromat 63x/1.20 water corrected objective. Figures were prepared using ImageJ for image visualization.

For 1,6-hexanediol and *Q. saponaria* saponin mix treatment, *M. truncatula* hairy roots overexpressing *G1-Venus* were dipped in solutions with either 3% (w/v) 1,6-hexanediol (Sigma-Aldrich, 211-074-0) or 0.5% (w/v) Saponin from Quillaja Bark (Sigma-Aldrich, 232-462-6). For G1 localization following mashing, control (intact) hairy roots were immerged in sterile purified water whereas mashed hairy roots were squeezed between glass slides.

All treatments were performed for 15 min prior confocal imaging. Confocal microscopy was performed using a Zeiss LSM710 laser scanner microscope with Plan-Apochromat 20x/0.8 M27; a 488-nm Argon laser was adopted for excitation of the G1-Venus fusion protein. Z-sections were taken every 2 µm. Images were processed generating maximum intensity projections and adding scale bars using ImageJ software.

ACKNOWLEDGEMENTS. We thank Rebecca De Clercq for technical assistance, Ellie Himschoot for help with the confocal analysis, Frederik Coppens and Lieven Sterck for support with the RNA-Seq analysis, Steven Vandersyppe and Geert Goeminne (VIB Metabolomics Core Ghent) for the

metabolite profiling, and Annick Bleys for help with preparing the manuscript. This research was supported by the Research Foundation Flanders by a research project grant to A.G. (G004515N), a predoctoral fellowship to J.M., and a postdoctoral fellowship to J.P., the Special Research Fund from Ghent University to A.G. (project O1J14813), and the European Union's Horizon 2020 research and innovation program under Grant Agreements No 760331 (Newcotiana) and No 825730 (Endoscape) to A.G.

References

1. Pentzold, S., Zagrobelny, M., Rook, F. & Bak, S. How insects overcome two-component plant chemical defence: plant β-glucosidases as the main target for herbivore adaptation. *Biol. Rev.* **89**, 531-551 (2014).
2. Morant, A.V. *et al.* β-Glucosidases as detonators of plant chemical defense. *Phytochemistry* **69**, 1795-1813 (2008).
3. Louveau, T. & Osbourn, A. The sweet side of plant-specialized metabolism. *Cold Spring Harb. Perspect. Biol.* **11**, a034744 (2019).
4. Guirimand, G. *et al.* Strictosidine activation in Apocynaceae: towards a "nuclear time bomb"? *BMC Plant Biol.* **10**, 182 (2010).
5. Hannemann, L. *et al.* A promiscuous beta-glucosidase is involved in benzoxazinoid deglycosylation in *Lamium galeobdolon*. *Phytochemistry* **156**, 224-233 (2018).
6. Cárdenas, P.D., Almeida, A. & Bak, S. Evolution of structural diversity of triterpenoids. *Front. Plant Sci.* **10**, 1523 (2019).
7. Thimmappa, R., Geisler, K., Louveau, T., O'Maille, P. & Osbourn, A. Triterpene biosynthesis in plants. *Annu. Rev. Plant Biol.* **65**, 225-257 (2014).
8. Liu, C., Ha, C.M. & Dixon, R.A. Functional genomics in the study of metabolic pathways in *Medicago truncatula*: an overview. *Methods Mol. Biol.* **1822**, 315-337 (2018).
9. Gholami, A., De Geyter, N., Pollier, J., Goormachtig, S. & Goossens, A. Natural product biosynthesis in *Medicago* species. *Nat. Prod. Rep.* **31**, 356-380 (2014).
10. Suzuki, H., Achnine, L., Xu, R., Matsuda, S.P.T. & Dixon, R.A. A genomics approach to the early stages of triterpene saponin biosynthesis in *Medicago truncatula*. *Plant J.* **32**, 1033-1048 (2002).
11. Carelli, M. *et al. Medicago truncatula* CYP716A12 is a multifunctional oxidase involved in the biosynthesis of hemolytic saponins. *Plant Cell* **23**, 3070-3081 (2011).
12. Fukushima, E.O. *et al.* CYP716A subfamily members are multifunctional oxidases in triterpenoid biosynthesis. *Plant Cell Physiol.* **52**, 2050-2061 (2011).
13. Biazzi, E. *et al.* CYP72A67 catalyzes a key oxidative step in *Medicago truncatula* hemolytic saponin biosynthesis. *Mol. Plant* **8**, 1493-1506 (2015).
14. Fukushima, E.O. *et al.* Combinatorial biosynthesis of legume natural and rare triterpenoids in engineered yeast. *Plant Cell Physiol.* **54**, 740-749 (2013).
15. Ribeiro, B. *et al.* A seed-specific regulator of triterpene saponin biosynthesis in *Medicago truncatula*. *Plant Cell* **32**, 2020-2042 (2020).
16. Tzin, V. *et al.* Integrated metabolomics identifes CYP72A67 and CYP72A68 oxidases in the biosynthesis of *Medicago truncatula* oleanate sapogenins. *Metabolomics* **15**, 85 (2019).
17. Confalonieri, M. *et al.* CRISPR/Cas9-mediated targeted mutagenesis of CYP93E2 modulates the triterpene saponin biosynthesis in *Medicago truncatula*. *Front. Plant Sci.* **12**, 690231 (2021).

18. Naoumkina, M.A. *et al.* Genomic and coexpression analyses predict multiple genes involved in triterpene saponin biosynthesis in *Medicago truncatula*. *Plant Cell* **22**, 850-866 (2010).
19. Abbruscato, P. *et al.* Triterpenoid glycosides from *Medicago sativa* as antifungal agents against *Pyricularia oryzae*. *J. Agric. Food Chem.* **62**, 11030-11036 (2014).
20. Oleszek, W. Structural specificity of alfalfa (*Medicago sativa*) saponin haemolysis and its impact on two haemolysis based quantification methods. *J. Sci. Food Agric.* **53**, 477-485 (1990).
21. Saniewska, A., Jarecka, A., Bialy, Z. & Jurzysta, M. Antifungal activity of saponins from *Medicago arabica* L. shoots against some pathogens. *Allelopathy Journal* **16**, 105-112 (2005).
22. Jarecka, A., Saniewska, A., Biały, Z. & Jurzysta, M. The effect of *Medicago arabica*, *M. hybrida* and *M. sativa* saponins on the growth and development of *Fusarium oxysporum* Schlecht f. sp. *tulipae* Apt. *Acta Agrobot.* **61**, 147-155 (2008).
23. Houghton, P., Patel, N., Jurzysta, M., Bialy, Z. & Cheung, C. Antidermatophyte activity of Medicago extracts and contained saponins and their structure-activity relationships. *Phytother. Res.* **20**, 1061-1066 (2006).
24. Huhman, D.V., Berhow, M.A. & Sumner, L.W. Quantification of saponins in aerial and subterranean tissues of *Medicago truncatula*. *J. Agric. Food Chem.* **53**, 1914-1920 (2005).
25. Lei, Z., Watson, B.S., Huhman, D., Yang, D.S. & Sumner, L.W. Large-scale profiling of saponins in different ecotypes of *Medicago truncatula*. *Front. Plant Sci.* **10**, 850 (2019).
26. Agrelli, J., Oleszek, W., Stochmal, A., Olsen, M. & Anderson, P. Herbivore-induced responses in alfalfa (*Medicago sativa*). *J. Chem. Ecol.* **29**, 303-320 (2003).
27. Pollier, J. *et al.* The protein quality control system manages plant defence compound synthesis. *Nature* **504**, 148-152 (2013).
28. Mertens, J. *et al.* The bHLH transcription factors TSAR1 and TSAR2 regulate triterpene saponin biosynthesis in *Medicago truncatula*. *Plant Physiol.* **170**, 194-210 (2016).
29. Erffelinck, M.-L. *et al.* The heat shock protein 40-type chaperone MASH supports the endoplasmic reticulum-associated degradation E3 ubiquitin ligase MAKIBISHI1 in *Medicago truncatula*. *Front. Plant Sci.* **12**, 639625 (2021).
30. Tzfadia, O., Amar, D., Bradbury, L.M., Wurtzel, E.T. & Shamir, R. The MORPH algorithm: ranking candidate genes for membership in *Arabidopsis* and tomato pathways. *Plant Cell* **24**, 4389-4406 (2012).
31. Ribeiro, B. *et al.* ER-anchored transcription factors bZIP17 and bZIP60 regulate triterpene saponin biosynthesis in *Medicago truncatula*. *bioRxiv* **2020.01.17.910802**(2020).
32. He, J. *et al.* The *Medicago truncatula* gene expression atlas web server. *BMC Bioinformatics* **10**, 441 (2009).
33. Van Bel, M. *et al.* PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Res.* **46**, D1190-D1196 (2018).
34. Naoumkina, M. *et al.* Different mechanisms for phytoalexin induction by pathogen and wound signals in *Medicago truncatula*. *Proc. Natl. Acad. Sci. USA* **104**, 17909-17915 (2007).
35. Yang, J. *et al.* Comprehensive identification and characterization of abiotic stress and hormone responsive glycosyl hydrolase family 1 genes in *Medicago truncatula*. *Plant Physiol. Biochem.* **158**, 21-33 (2021).
36. Gus-Mayer, S., Brunner, H., Schneider-Poetsch, H.A.W. & Rüdiger, W. Avenacosidase from oat: purification, sequence analysis and biochemical characterization of a new member of the BGA family of β-glucosidases. *Plant Mol. Biol.* **26**, 909-921 (1994).
37. Kim, Y.-W., Kang, K.-S., Kim, S.-Y. & Kim, I.-S. Formation of fibrillar multimers of oat β-glucosidase isoenzymes is mediated by the As-Glu1 monomer. *J. Mol. Biol.* **303**, 831-842 (2000).
38. Sperschneider, J. *et al.* LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. *Sci. Rep.* **7**, 44598 (2017).

39. Scott, M.S., Troshin, P.V. & Barton, G.J. NoD: a nucleolar localization sequence detector for eukaryotic and viral proteins. *BMC Bioinformatics* **12**, 317 (2011).
40. Tadege, M. *et al.* Large-scale insertional mutagenesis using the *Tnt1* retrotransposon in the model legume *Medicago truncatula*. *Plant J.* **54**, 335-347 (2008).
41. Yoneda, M., Nakagawa, T., Hattori, N. & Ito, T. The nucleolus from a liquid droplet perspective. *J. Biochem.* **170**, 153-162 (2021).
42. Avato, P. *et al.* Antimicrobial activity of saponins from *Medicago* sp.: structure-activity relationship. *Phytother. Res.* **20**, 454-457 (2006).
43. Pollier, J., Morreel, K., Geelen, D. & Goossens, A. 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 (2011).
44. Kapusta, I., Janda, B., Stochmal, A. & Oleszek, W. Determination of saponins in aerial parts of barrel medic (*Medicago truncatula*) by liquid chromatography-electrospray ionization/mass spectrometry. *J. Agric. Food Chem.* **53**, 7654-7660 (2005).
45. Kapusta, I. *et al.* Triterpene saponins from barrel medic (*Medicago truncatula*) aerial parts. *J. Agric. Food Chem.* **53**, 2164-2170 (2005).
46. Hua, Y., Sansenya, S., Saetang, C., Wakuta, S. & Ketudat Cairns, J.R. Enzymatic and structural characterization of hydrolysis of gibberellin A4 glucosyl ester by a rice β-D-glucosidase. *Arch. Biochem. Biophys.* **537**, 39-48 (2013).
47. Dietz, K.-J., Sauter, A., Wichert, K., Messdaghi, D. & Hartung, W. Extracellular β-glucosidase activity in barley involved in the hydrolysis of ABA glucose conjugate in leaves. *J. Exp. Bot.* **51**, 937-944 (2000).
48. Kim, J., Lee, H., Lee, H.G. & Seo, P.J. Get closer and make hotspots: liquid–liquid phase separation in plants. *EMBO Rep.* **22**, e51656 (2021).
49. Wang, X. *et al.* A photoregulatory mechanism of the circadian clock in *Arabidopsis*. *Nat. Plants* **7**, 1397-1408 (2021).
50. Huang, S., Zhu, S., Kumar, P. & MacMicking, J.D. A phase-separated nuclear GBPL circuit controls immunity in plants. *Nature* **594**, 424-429 (2021).
51. Ouyang, M. *et al.* Liquid-liquid phase transition drives intra-chloroplast cargo sorting. *Cell* **180**, 1144-1159 (2020).
52. Hellemans, J., Mortier, G., De Paepe, A., Speleman, F. & Vandesompele, J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* **8**, R19 (2007).
53. Nagai, T. *et al.* A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* **20**, 87-90 (2002).
54. Karimi, M., Inzé, D. & Depicker, A. GATEWAY^{TM} vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* **7**, 193-195 (2002).
55. Nallamsetty, S., Austin, B.P., Penrose, K.J. & Waugh, D.S. Gateway vectors for the production of combinatorially-tagged His_6-MBP fusion proteins in the cytoplasm and periplasm of *Escherichia coli*. *Protein Sci.* **14**, 2964-2971 (2005).
56. Jacobs, T.B., LaFayette, P.R., Schmitz, R.J. & Parrott, W.A. Targeted genome modifications in soybean with CRISPR/Cas9. *BMC Biotechnol.* **15**, 16 (2015).
57. Liu, H. *et al.* CRISPR-P 2.0: an improved CRISPR-Cas9 tool for genome editing in plants. *Mol. Plant* **10**, 530-532 (2017).
58. Brinkman, E.K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, e168 (2014).

Figures and Tables

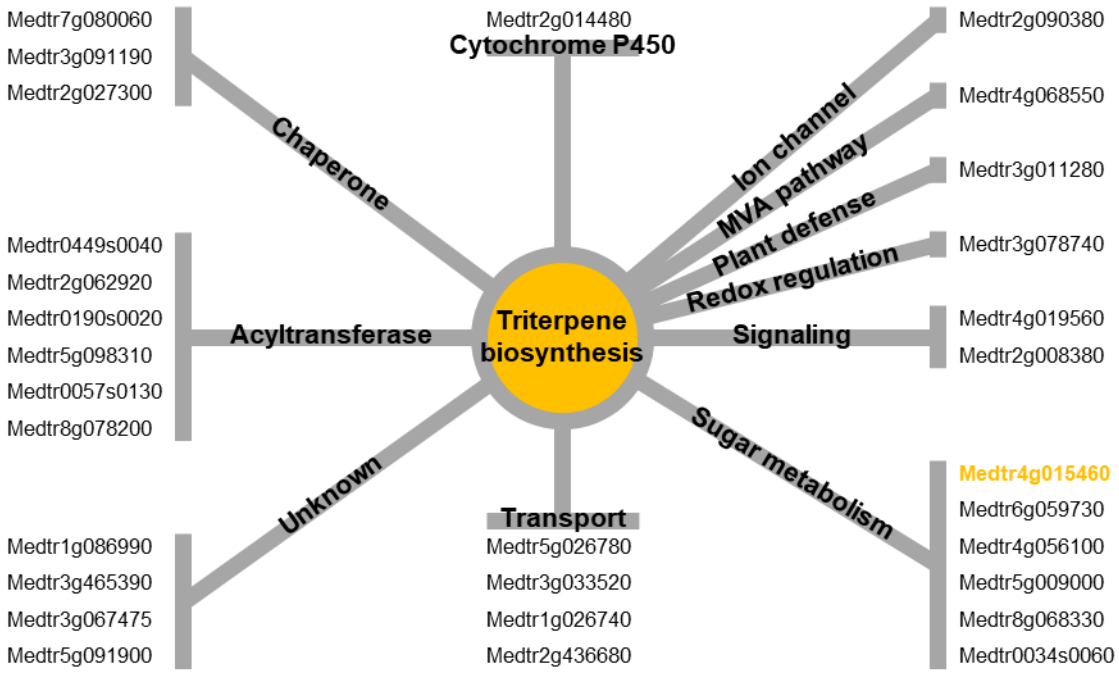

Figure 1. Candidate triterpene saponin biosynthesis genes as predicted by MORPH[30]. The top candidate β-glucosidase, *Medtr4g015460*, is indicated in yellow.

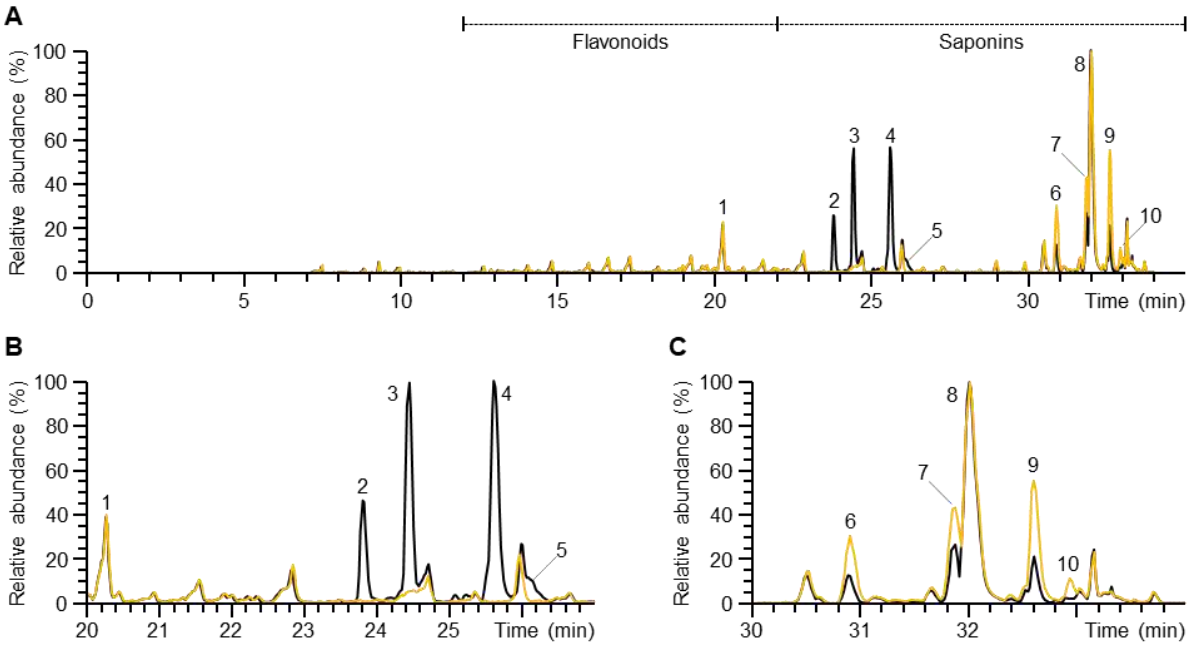

Figure 2. Determination of the biochemical activity of G1 on *M. truncatula* hairy roots extracts. A, LC-MS chromatograms comparing methanolic hairy roots extracts incubated with boiled G1 (black) and G1 (orange). Flavonoids elute between 12 and 22 min, whereas triterpene saponins elute between 22 and 35 min. B, Zoomed LC-MS chromatogram between 20 and 27 min, showing the substrates of G1 disappearing. C, Zoomed LC-MS chromatogram between 30 and 34 min, showing the higher abundance of the products of G1. 1: malonyl-ononin; 2: Hex-Hex-Hex-medicagenic acid; 3: 3-Glc-28-Glc-medicagenic acid; 4: 3-Glc-malonyl-28-Glc-medicagenic acid; 5: Hex-Hex-HexA-hederagenin; 6: Hex-Hex-medicagenic acid; 7: 3-Glc-medicagenic acid; 8: soyasaponin I; 9: 3-Glc-malonyl-medicagenic acid; 10: Hex-HexA-hederagenin.

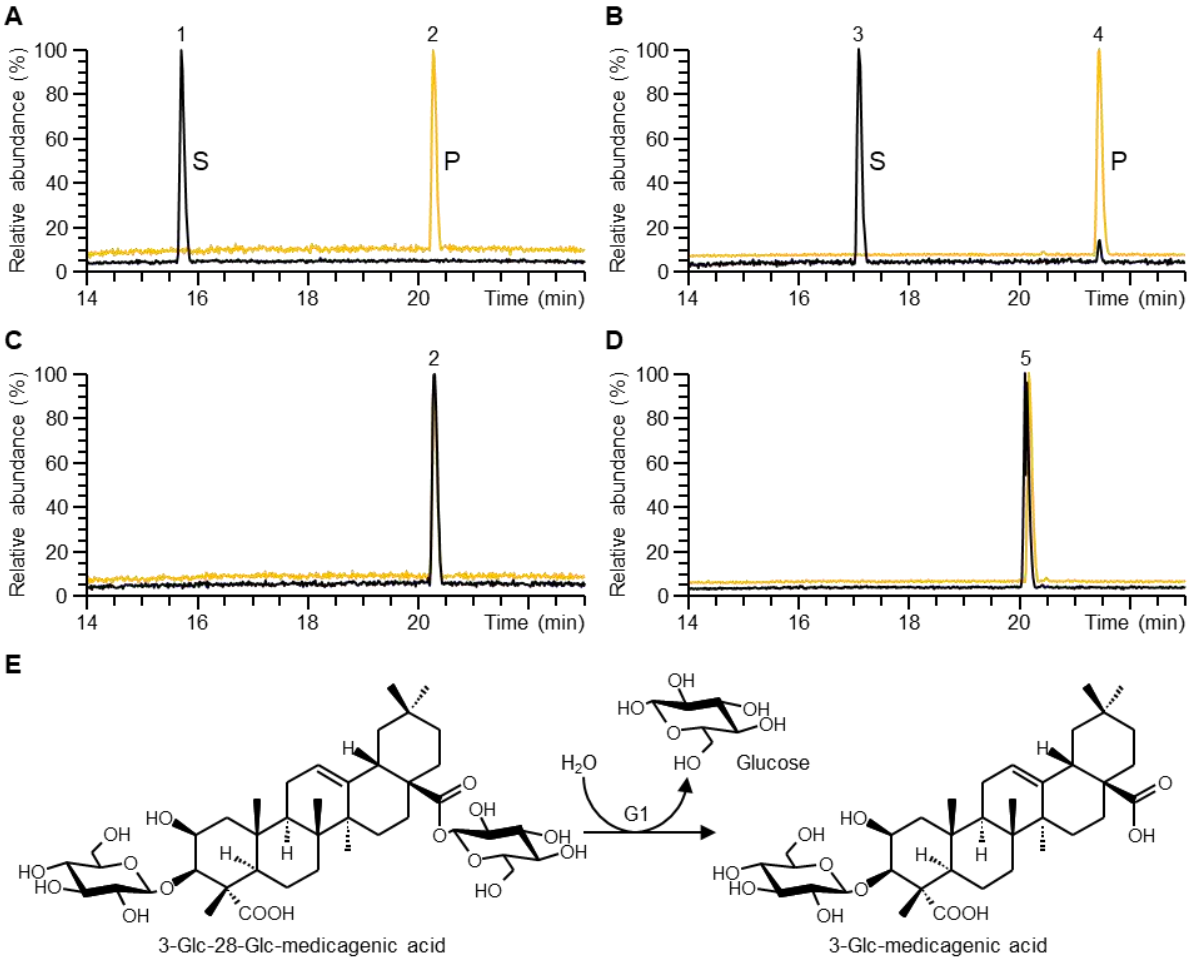

Figure 3. *In vitro* activity of G1 with triterpene saponin standards. A, Incubation of G1 with 3-Glc-28-Glc-medicagenic acid. B, Incubation of G1 with 3-Glc-Ara-28-Glc-hederagenin. C, Incubation of G1 with 3-Glc-medicagenic acid. D, Incubation of G1 with 3-Glc-echinocystic acid. The standards were incubated with boiled G1 (black) or native G1 (orange). E, Hydrolysis of the ester bond by G1 removes the glucose moiety from the C-28 carboxyl group of hemolytic triterpene saponins. 1, 3-Glc-28-Glc-medicagenic acid; 2, 3-Glc-medicagenic acid; 3, 3-Glc-Ara-28-Glc-hederagenin; 4, 3-Glc-Ara-hederagenin; 5, 3-Glc-echinocystic acid; S, substrate; P, product.

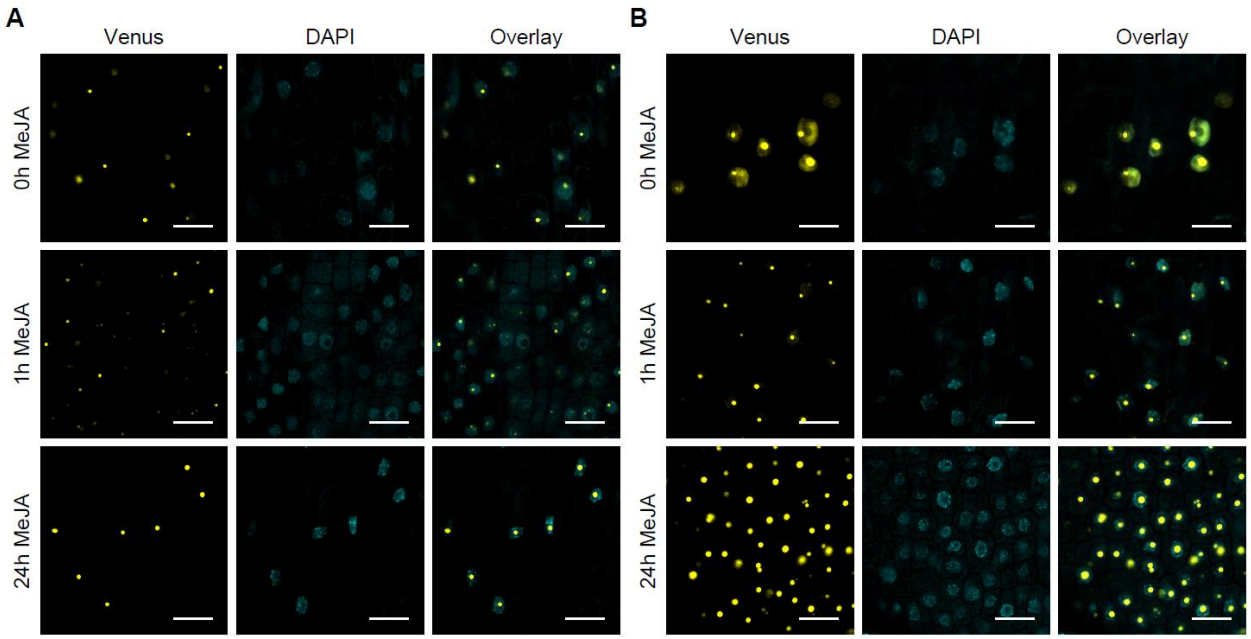

Figure 4. Nucleolar localization of G1. A-B, Confocal microscopy analysis of transgenic *M. truncatula* hairy roots expressing G1 tagged with Venus to its N-terminus (Venus-G1, A) or C-terminus (G1-Venus, B). DAPI staining of the nuclei revealed that the fusion proteins mainly localize within the nucleoli within the nucleus. MeJA treatment (100 µM of MeJA for 0 h, 1 h, or 24 h) does not affect the localization of the fusion protein. Scale bars, 20 µm.

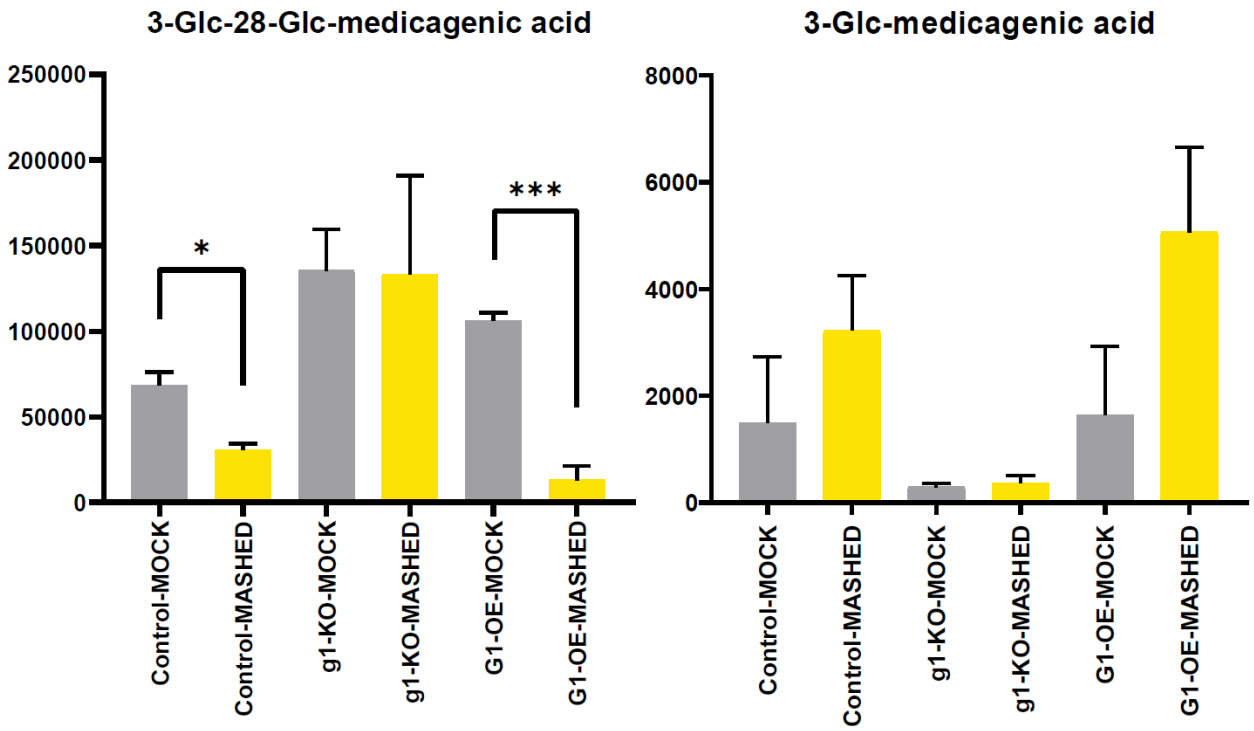

Figure 5. Accumulation of hydrolyzed hemolytic saponins in mashed *M. truncatula* hairy roots. Profiling of 3-Glc-28-Glc-medicagenic acid (left) and 3-Glc-medicagenic acid (right) in intact (mock) and mashed hairy roots of wild type (control), G1 knock-out (g1-KO) and G1-Venus overexpressing (G1-OE) lines. Values on the y-axis are normalized average peak values. The error bars designate the standard error (n = 5 biological replicates). Statistical significance was determined by a Student's *t*-test (*, $P < 0.05$; ***, $P < 0.001$).

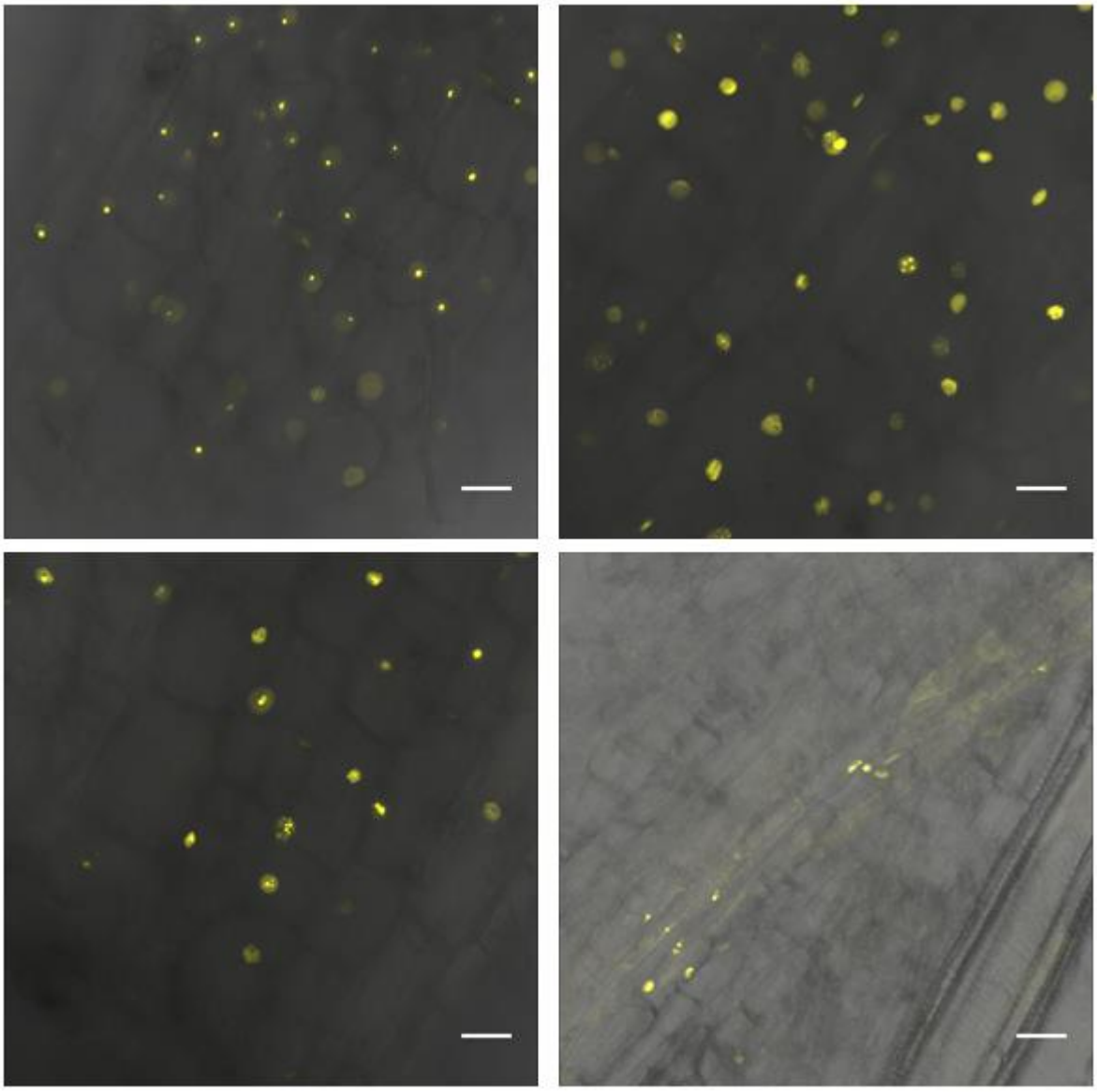

Figure 6. G1 is stored in the nucleolus via liquid-liquid phase separation. Confocal images were taken 15 min after mock (top left), 3% (w/v) 1,6-hexanediol (top right), 0.5% (w/v) Quillaja bark saponin mix (bottom left) or mashing (bottom right) treatment. Compared to mock, the G1-Venus fusion protein signal spread within the nucleus when surfactants were added or roots damaged. Scale bars, 20 µm.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- G1NatComSI.pdf