UNIVERSITY OF COPENHAGEN

**Københavns Universitet**

# The cytochrome P450 CYP72A552 is key to production of hederageninbased saponins that mediate plant defense against herbivores

Liu, Qing; Khakimov, Bekzod; Cárdenas, Pablo D.; Cozzi, Federico; Olsen, Carl Erik; Jensen, Karen Rysbjerg; Hauser, Thure Pavlo; Bak, Søren

*Published in:*
New Phytologist (Online)

*DOI:*
10.1111/nph.15689

*Publication date:*
2019

*Document version*
Early version, also known as pre-print

*Citation for published version (APA):*
Liu, Q., Khakimov, B., Cárdenas, P. D., Cozzi, F., Olsen, C. E., Jensen, K. R., ... Bak, S. (2019). The cytochrome P450 CYP72A552 is key to production of hederageninbased saponins that mediate plant defense against herbivores. *New Phytologist (Online)*, *222*(3), 1599-1609. https://doi.org/10.1111/nph.15689

Download date: 18. nov.. 2019

# The cytochrome P450 CYP72A552 plays a key role in triterpenoid saponin mediated plant defense against herbivores

| | |
|---|---|
| Journal: | *New Phytologist* |
| Manuscript ID | NPH-MS-2018-28407 |
| Manuscript Type: | MS - Regular Manuscript |
| Date Submitted by the Author: | 29-Oct-2018 |
| Complete List of Authors: | Liu, Qing; Kobenhavns Universitet Det Natur- og Biovidenskabelige Fakultet, Plant and Environmental Sciences<br>Khakimov , Bekzod; Kobenhavns Universitet Det Natur- og Biovidenskabelige Fakultet, Plant and Environmental Sciences; Kobenhavns Universitet Det Natur- og Biovidenskabelige Fakultet, Food Science<br>Cozzi , Federico; University of Copenhagen, Department of Plant and Environmental Sciences<br>Olsen, Carl; Kobenhavns Universitet Det Natur- og Biovidenskabelige Fakultet, Food Science<br>Jensen, Karen; Kobenhavns Universitet Det Natur- og Biovidenskabelige Fakultet, Department of Plant and Environmental Sciences<br>Hauser, Thure Pavlo ; Kobenhavns Universitet Det Natur- og Biovidenskabelige Fakultet, Department of Plant and Environmental Sciences<br>Bak, Soren; Kobenhavns Universitet Det Natur- og Biovidenskabelige Fakultet, Department of Plant and Environmental Sciences |
| Key Words: | cytochrome P450, molecular evolution, triterpenoid saponin biosynthesis, plant-insect interaction, pathway elucidation |
| | |

SCHOLARONE™
Manuscripts

# The cytochrome P450 CYP72A552 plays a key role in triterpenoid saponin mediated plant defense against herbivores

Qing Liu [1], Bekzod Khakimov [1,2], Federico Cozzi [1], Carl Erik Olsen [1], Karen Rysbjerg Jensen [1], Thure Pavlo Hauser [1], and Søren Bak [1]

[1] Department of Plant and Environmental Sciences and Copenhagen Plant Science Center, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark.

[2] Department of Food Science, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark.

Authors for correspondence:

Professor Søren Bak

Tel: +45 35333346

Email: bak@plen.ku.dk

Dr. Qing Liu

Tel: +45 35331495

Email: qliu@plen.ku.dk

For Peer Review

**Summary**

Plants continuously evolve new defense compounds. One class of such compounds are triterpenoid saponins. A few species in the Barbarea genus produce saponins as the only ones in the large crucifer family. However, the molecular mechanism behind saponin biosynthesis and their role in plant defense remains unclear.

We used a combination of pathway reconstitution *in planta*, enzymatic production of saponins *in vitro*, insect feeding assays, and bioinformatics to identify a missing gene involved in saponin biosynthesis and saponin-based herbivore defense.

A tandem repeat of eight *CYP72A* cytochromes P450 co-localize with a QTL for saponin accumulation and flea beetle resistance in *B. vulgaris*. We found that CYP72A552 oxidizes oleanolic acid at position C-23 to hederagenin. *In vitro* produced hederagenin mono-glucosides reduced larval feeding by up to 90 % and caused 75 % larval mortality of the major crucifer pest diamondback moth and the tobacco hornworm. Sequence analysis indicate that *CYP72A552* evolved through gene duplication and has been under strong selection pressure.

In conclusion, CYP72A552 has evolved to play an essential role in plant defense against herbivores. Our study highlights evolution of chemical novelties by gene duplication and selection for enzyme innovations, and the importance of chemical modification in plant defense evolution.

Key words: cytochrome P450, molecular evolution, triterpenoid saponin biosynthesis, plant-insect interaction, pathway elucidation

For Peer Review

## Introduction

Plants continuously evolve new defense compounds to survive in the arms race against herbivores. Saponins are one such class of specialized compounds. They consist of a hydrophobic triterpenoid backbone with one or more hydrophilic saccharide groups attached. Some saponins are toxic or deterrent to insects, molluscs, fungi, and other microorganisms, and have probably evolved for defense against these, although their mode of action and evolution is not fully understood (De Geyter *et al.*, 2007). Research on different *Lepidoptera* (moths and butterflies) cell lines indicates that saponins can affect midgut cell membranes (De Geyter *et al.*, 2012), which may explain their toxicity. In addition to cell membrane disruption, saponins can also act as deterrents to insect feeding due to their bitter taste (De Geyter *et al.*, 2007; Augustin *et al.*, 2012).

The ability to produce saponins has evolved convergently in several different plant families, testifying their biological significance. Within the large crucifer family (Brassicaceae), only a few species in the genus *Barbarea* are known to produce saponins (Badenes-Perez *et al.*, 2013). One of the species, *B. vulgaris*, has diverged into two different "plant types" that differ in saponin profile and herbivore resistance, as well as in other defense-related traits (Kuzina *et al.*, 2011; Hauser *et al.*, 2012; Toneatto *et al.*, 2012). One type, the G-type (Glabrous), is completely resistant to some specialist insect herbivores, including the devastating agricultural pest diamondback moth (*Plutella xylostella*) and the flea beetle (*Phyllotreta nemorum*). The other type, the P-type (Pubescent), is susceptible to these insects (Agerbirk *et al.*, 2003; Kuzina *et al.*, 2011). Due to the interesting defense properties of *Barbarea*, including an unusual glucosinolate profile, it has become a model for evolution and ecology of plant defense compounds (Augustin *et al.*, 2011; Byrne *et al.*, 2017).

Insect resistance in *B. vulgaris* correlates in dose-dependent manner with four triterpenoid saponins: the cellobiosides of hederagenin, oleanolic acid, gypsogenin, and 4-epi-hederagenin (Shinoda *et al.*, 2002; Agerbirk *et al.*, 2003; Kuzina *et al.*, 2009) (Fig. 1a). These four saponins are constitutively present and can be induced to elevated levels upon insect and pathogen attack (van Mölken *et al.*, 2014). Among these saponins, hederagenin cellobioside is especially toxic or deterrent, as indicated by feeding assays with purified compounds painted on radish leaf discs (Renwick, 2002; Badenes-Perez *et al.*, 2013). In comparison, oleanolic acid cellobioside, which differs





For Peer Review

from hederagenin cellobioside only by the lack of a C-23 alcohol group, has much less effect (Nielsen *et al.*, 2010).

Knowledge of genes and enzymes involved in the biosynthesis of saponins is incomplete. According to a proposed model, the omnipresent precursor 2,3-oxidosqualene is cyclized to a limited number of core structures, which are subsequently decorated with functional groups, and finally, sugars are added (Seki *et al.*, 2015). Oxidosqualene cyclases (OSCs) catalyze the first committed step in the pathway forming the core structures, to which functional oxygenated groups are added by cytochromes P450 (P450s), and sugars by UDP- glycosyltransferases (UGTs) (Seki *et al.*, 2015).

Several studies have made significant progress in uncovering genes and enzymes involved in saponin biosynthesis and the associated flea beetle resistance in *B. vulgaris*. Two OSCs, LUP5 and LUP2, have been identified and characterized from G- and P-type plants. LUP5 is preferentially expressed in G-type (flea beetle resistant) plants and produces mainly β-amyrin, the backbone of the resistance-conferring oleane-type saponins. LUP2 is preferentially expressed in P-type (insect herbivore susceptible) plants and produces mainly lupeol, the backbone of lupeol-type saponins (Khakimov *et al.*, 2015), for which we do not know any biological function in *B. vulgaris*. Other OSCs with similar function have been reported from licorice, ginseng, lotus, *Arabidopsis thaliana,* etc. (Kushiro *et al.*, 1998; Hayashi *et al.*, 2001; Husselstein–Muller *et al.*, 2001; Sawai *et al.*, 2006). The P450s CYP716A80 and CYP716A81 from the G-type and P-type, respectively, oxidizes β-amyrin at the C-28 position to oleanolic acid, the precursor of hederagenin (Khakimov *et al.*, 2015). Several UDP-glycosyltransferases from the UGT73C subfamily in *B vulgaris* have been isolated (Augustin *et al.*, 2012; Erthmann *et al.*, 2018). Among them, UGT73C11 and its ortholog UCT73C10 have evolved to specifically catalyze 3-*O*-glucosylation of sapogenins (triterpenoid saponin backbones), including oleanolic acid and hederagenin.

In the *B. vulgaris* genome, two quantitative trait loci (QTL) for the four oleanolic acid-based saponins were found to co-localize with a QTL for flea beetles resistance (Khakimov *et al.*, 2015; Byrne *et al.*, 2017). The first QTL also co-localizes with LUP5, the other QTL is syntenic to a region on *A. thaliana* chromosome 3 containing eight highly similar *CYP72As* (Kuzina *et al.*, 2011). Several members of the *CYP72A*





For Peer Review

subfamily from *Glycyrrhiza* and other legumes are involved in oleanolic acid derived saponin biosynthesis, by oxidizing at C2, C21, C22 and C30 positions (Seki *et al.*, 2011; Fukushima *et al.*, 2013; Biazzi *et al.*, 2015). These findings, and their location in a QTL for flea beetle resistance, suggest that CYP72As from *B. vulgaris* may have an important role in the biosynthesis of hederagenin (C-23 oxidized oleanolic acid) and derived saponins, and thereby for defense against insect herbivores.

In this study, we identified *CYP72A552* from a QTL region involved in flea beetle resistance, and show that it plays a key role in activating toxicity and/or deterrence against the devastating agricultural pest diamondback moth (*Plutella xylostella*) and the tobacco specialist tobacco hornworm (*Manduca sexta*). This key gene in the saponin biosynthesis evolved through gene duplication, most likely at some time after the origin of the *Barbarea* genus, and is still under selection.

## Materials and Methods

### Gene mining and cloning of CYP72A candidate genes from *b. vulgaris*

Putative P450 candidates within the QTLs for flea beetles resistance were searched using the CLC Workbench (QIAGEN). Eight CYP72A candidate genes were identified in the QTL region, for which PCR primer pairs (Table S3) were designed either from genome DNA or cDNA. The full-length sequence of all eight candidate genes was verified with Sanger sequencing and deposited into the NCBI database with accession numbers from MH252567 to MH252574. Phylogenies were analyzed as described in Supplementary Materials and Methods.

### Transient expression in *Nicotiana benthamiana*

For transient expression in tobacco (*N. benthamiana*), the selected genes were amplified by PCR from *B. vulgaris* cDNA and cloned into the pEAQ-HT-DEST expression vector (Sainsbury *et al.*, 2009), using USER technology (Nour-Eldin *et al.*, 2010). *Agrobacterium tumefaciens* infiltration (agro-infiltration) for transient expression in tobacco leaves was performed as described by Khakimov *et al,*(2015). After infiltration, plants were grown for another four-and-a-half days before harvesting for analysis. All constructs were verified by sequencing.

For Peer Review

**Expression in yeast**

For expression analysis in yeast (*S. cerevisiae*), INVSc1 (ThermoFisher Scientific), *CYP72A552* was cloned into the pYeDP60 vector using USER cloning technology (Nour-Eldin *et al.*, 2010). pYES3-ADH-OSC1 (Seki *et al.*, 2008) and pELC-CYP716A80 was co-transformed with pYeDP60-CYP72A552 constructs, using Frozen-EX Yeast Transformation II™ (Zymo Research). Yeast cells were cultured in synthetic complete medium without uracil, tryptophan or leucine medium containing 2% glucose at 30°C for 24 hours. The cells were then collected and resuspended in synthetic complete medium containing 2% galactose, and cultured at 30 °C for 2 days.

**Yeast *in vitro* microsome assay**

Yeast microsome assays were performed as described by Liu *et al.* (2014). 100 µl isolated microsomal, 1 µl substrate (50µM), 100 µl NADPH, 20 µl potassium buffer (1 M, pH7.5), and 278 µl water were mixed and incubated for 1 h at 25 °C with shaking (200 rpm), and the mixture was extracted with 1.5 ml ethyl acetate two times. The combined ethyl acetate extracts were dried with ScanVac prior to GC-MS analysis.

**Metabolite analysis by GC-MS and LC-MS**

Tobacco leaf discs were ground to fine powder under liquid $N_2$ with mortar and pestle. 100 mg of powder was used for either GC-MS or LC-MS analysis (Supplementary Materials and Methods). LC-MS data were processed using MetAlign version 4.0 (www.metAlign.nl) as described in Yang *et al.,* (2011) with modifications. MetAlign allowed baseline correction, noise elimination and subsequent spectral data alignment (De Vos *et al.*, 2007). The processing parameters were set to analyze from scan numbers 31-3827 at the retention time of 0.3-32.0 min. The output from MetAlign (signal intensities of each variable) was subjected to an ANOVA to identify metabolites that differed between C1 and C2 with the cutoff value of 2-fold change. Metabolite identification at level 1 and 2, according to the Metabolomics Standards Initiatives (http://www.metabolomics-msi.org/), were assisted by molecular formula, generated from an accurate mass spectrometry data (± 5 ppm), MS/MS fragmentation pattern, and the use of authentic standards.

For Peer Review

GC-MS data were processed using DataAnalysis software (version 4.3, Bruker, Germany). Intensities of saponins aglycones, including β-amyrin, oleanolic acid, hederagenin, gypsogenin, and gypsogenic acid, were extracted using characteristic *m/z* ions including 129, 203, 218, 320, and 471.

**Non-choice insect feeding assays**

Tobacco hornworm (*Manduca sexta*) and diamondback moth (*Plutella xylostella*) eggs were stored in insect cages in a greenhouse at 26 °C- 16 h light : 24°C- 8 h darkness, and in a growth room 20 °C- 16 h light : 20°C- 8 h darkness, respectively, until the larvae hatched. Freshly hatched neonates were placed on tobacco or broccoli (*Brassica oleracea*) overnight for acclimation. First instar larvae of tobacco hornworm and 3$^{rd}$ instar larvae of diamondback moth were starved 4 to 5 hours before used for feeding assays. Non-choice assays were performed as described by Kumar *et al.* (2012) and Augustin *et al.* (2012) with modifications. Briefly, leaf discs (1.57 cm$^2$) from (1) tobacco agro-infiltrated with saponin biosynthetic genes or (2) tobacco and broccoli applied with pure monoglucosides were placed into 24-well plates (15.6 mm well diameter), one leaf disc per well, with three layers of filter paper (Whatman, UK) beneath. 60 µL of water was added on the filter papers to secure humidity. One larva was placed on each leaf disc and fed for 24 to 52 hours. 20 to 30 replicates (larvae) were used for each treatment. Consumed leaf area was measured with a Portable Area Meter (LI-COR Inc., USA) after 24 hours and larvae survivorship was assessed after 52 hours.

***In vitro* synthesis and NMR analysis**

Monoglucosides of 3-*O*-oleanolic acid, 3-*O*-hederagenin, and 3-*O*-gypsogenic acid were synthesized by incubating the corresponding aglycon and UDP-glucose with purified UGT73C11 enzymes, as described by Augustin *et al*. (2012). The structures of the *in vitro* produced monoglucosides were verified with either in-house standards or NMR (Supporting Information Materials and Methods).

**Sequence evolution analysis**

Nucleotide sequences of CYP72As from *B. vulgaris* and *A. thaliana* were aligned on the amino acid level based on codons using MUSCLE and used to construct a





For Peer Review

maximum-likelihood bootstrapped phylogenetic tree, using the JTT matrix-based model (Jones *et al.*, 1992). All the *A. thaliana* P450 sequences were obtained from the website for *Arabidopsis* cytochrome P450, cytochrome b5, P450 reductase, beta-glucosidase, and glycosyltransferases (http://www.p450.kvl.dk/) (Paquette *et al.*, 2000). The tree with the highest log-likelihood (-6728.4268) was selected. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log-likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.1865)). The tree was drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 16 amino acid sequences. The coding data was translated assuming a Standard genetic code table. There were a total of 522 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar *et al.*, 2016). Nonsynonymous to synonymous substitution rate ratios (Yang, 2007) were calculated for codon-based nucleotide alignments with the program 'codeml' from the PAML package and PAMLx (Xu & Yang, 2013). All site-models (NSsites) were tested with model 0 and branch models 1:b (Hall, 2011).

## Results

### QTL map and genome-based cloning of *BvCYP72A*s

A quantitative trait locus (QTL) for oleanolic acid-derived saponins was previously found to co-localize with a QTL for flea beetles resistance in *B. vulgaris* (Kuzina et al., 2009). Through fine mapping of this QTL region, a tandem repeat of six *CYP72A*s was identified (Fig. 1b) that is syntenic to a region on *A. thaliana* chromosome 3 containing a similar tandem repeat of eight *CYP72As*. The remaining two *CYP72As* in the QTL were placed in *B. vulgaris* contig 4433 and contig 1898. Nucleotide-based sequence alignment of the *CYP72A*s showed that four of the *BvCYP72A* candidate genes share high nucleotide sequence identity (from 87% to 93%) which may partially explain why the remaining two *CYP72As* were not assembled in the same position as the rest, as assembling algorithms do not facilitate assembly of such highly repetitive regions





For Peer Review

(Byrne *et al.*, 2017) (Fig. S1). Based on their phylogenetic clustering with *CYP72A*s, three of them were officially named *BvCYP72A7*, *BvCYP72A8*, and *BvCYP72A9,* in agreement with the P450 nomenclature system (http://drnelson.uthsc.edu/cytochromeP450.html). These three CYP72As are considered orthologues to *A. thaliana CYP72A7, CYP72A8, and CYP72A9* and accordingly have received the same names as their *A. thaliana* orthologues. Of the remaining *BvCYP72A*s, one was identified as a pseudogene, named *CYP72A561*p, and the remaining were named *CYP72A552*, *CYP72A553*, *CYP72A562,* and *CYP72A563*, respectively.

**CYP72A552 oxidizes oleanolic acid to hederagenin**

To determine which of the *BvCYP72A* candidates is involved in the biosynthesis of oleanolic acid-derived saponins, each of the seven candidate genes were transiently expressed in tobacco (*Nicotiana benthamiana*) leaves together with two previously characterized upstream genes in the saponin biosynthesis pathway from *B. vulgaris*: *BvLUP5* and *BvCYP716A80* to provide substrates for the CYP72As. We have previously shown that when these two genes are transiently co-expressed in tobacco leaves, oleanolic acid is the major product (Khakimov *et al.*, 2015). When *CYP72A552* was co-expressed with *BvLUP5* and *BvCYP716A80*, two major peaks and one minor could be identified (Fig. 2a); this was not found for any of the other *CYP72A* gene combinations. The two major peaks were identified as hederagenin and gypsogenic acid, and the minor peak as gypsogenin, based on authentic standards. Hederagenin, gypsogenin and gypsogenic acid are the corresponding alcohol, aldehyde and carboxylic acid derivative of oleanolic acid when oxidized at the C-23 position, demonstrating that CYP72A552 oxidizes at this position.

To further verify its catalytic function, *CYP72A552* was expressed in yeast cells to enable *in vitro* microsome enzyme assays. When oleanolic acid was administrated as the substrate to yeast microsomes harboring CYP72A552, only hederagenin was detected (Fig. 2b). We also expressed *BvCYP72A552* in yeast cells together with *LjOSC1* from *Lotus japonicus* (which produces mainly β-amyrin), and *BvCYP716A80*. This resulted in four new peaks (Fig. 2c): hederagenin, gypsogenin, gypsogenic acid,

and 23-hydroxy erythrodiol, the latter putatively identified from the mass spectrum and fragmentation pattern. To determine if the gypsogenin and gypsogenic acid identified from the yeast and tobacco were converted by CYP72A552, hederagenin was administrated as substrate to the microsomes harboring CYP72A552. Hederagenin was in that case not further oxidized, demonstrating that CYP72A552 only oxidizes oleanolic acid to hederagenin (Fig. S2). Thus, CYP72A552 oxidizes oleanolic acid to hederagenin, which is further oxidized to gypsogenin and gypsogenic acid probably by endogenous enzymes in both yeast cells and tobacco plants. In conclusion, of the eight CYP72As in the QTL for flea beetle resistance, only CYP72A552 oxidizes oleanolic acid at the C-23 position to hederagenin (Fig. 2d).

**Hederagenin-based saponins is a major deterrent to both diamondback moth and tobacco hornworm**

The sapogenins (saponin aglycons) from *B. vulgaris* have previously been shown not to affect the tested crucifer herbivores, but only to become a deterrent in their 3-*O*-glucosylated form(s) (Nielsen *et al.*, 2010; Augustin *et al.*, 2012). To determine the antifeedant activity of C-23 oxidized saponins, the monoglucoside of oleanolic acid (OA), hederagenin (OA alcohol) and gypsogenic acid (OA acid) were synthesized *in vitro* (yielding 18.0, 15.6, and 13.0 mg of each, respectively) and tested for antifeedant activity against the crucifer specialist herbivore diamondback moth (*P. xylostella*), after being purified and structurally verified with in-house standards or by NMR (Fig. 3a, Fig. S3-5, Table S1) as previously described (Augustin *et al.*, 2012). The monoglucosides were applied to broccoli (*Brassica oleracea*) leaf discs at concentrations of 25 and 100 nmol $cm^{-2}$, and presented to diamondback moth (*P. xylostella*) larvae in non-choice assays. These concentrations where chosen to reflect natural concentrations of hederagenin cellobioside in fresh leaves of *B. vulgaris*, which is approximately 94 nmol $cm^{-2}$. A dose of 100 nmol $cm^{-2}$ hederagenin 3-*O*-monoglucoside reduced larval feeding by 90% compared to controls, which is a four-fold stronger feeding reduction than oleanolic acid 3-*O*-monoglucoside (Fig. 3b). At the dose of 25 nmol $cm^{-2}$, only hederagenin 3-*O*-monoglucoside reduced larval feeding, which is in agreement with the expected dose dependency. Even more strikingly, hederagenin 3-*O*-monoglucoside caused up to 75% larval mortality, which is 2-fold

higher than that of oleanolic acid 3-*O*-monoglucoside or gypsogenic acid 3-*O*-monoglucoside. In conclusion, hederagenin 3-*O*-monoglucoside is a much stronger defense-component against diamondback moth than oleanolic acid 3-*O*-monoglucoside or gypsogenic acid 3-*O*-monoglucoside, and substantiates the role of C-23 hydroxylation in saponin mediated defense

To develop a fast and efficient test system to evaluate the antifeedant activity of transiently expressed defense compounds in tobacco, effects of the saponins were also tested against larvae of tobacco hornworm (*Manduca sexta*), an important Solanaceous 'specialist' herbivore. As done above for broccoli leaf disks, tobacco leaf discs were coated with the three *in vitro* synthesized monoglucosides at the same concentrations as above. At the dose of 100 nmol $cm^{-2}$, hederagenin 3-*O*-monoglucoside reduced larval feeding by 88% compared to controls, which is a seven-fold stronger feeding reduction than oleanolic acid 3-*O*-monoglucoside. Further, hederagenin 3-*O*-monoglucoside caused 75% larval mortality compared to less than 5% for the other two monoglucosides (Fig. 3). At the dose of 25 nmol $cm^{-2}$, none of the monoglucosides affected consumption, confirming the dose dependency of saponins for defense. In conclusion, hederagenin 3-*O*-monoglucoside is a major deterrent or toxin against both diamond back moth and tobacco hornworm. Thus, the system consisting of tobacco leaf discs and tobacco hornworm larvae can be used for structure-activity relationship studies of saponin-mediated plant defense.

**BvCYP72A552 mediates herbivore defense**

Genetic modification, mutants and overexpression lines are not available in *B. vulgaris*, and we therefore analysed further the role of *CYP72A552* in plant defense against insects through transient expression in tobacco leaves. The saponin biosynthesis pathway was first transiently reconstructed in tobacco leaves, which were subsequently presented to the tobacco specialist herbivore, tobacco hornworm (*Manduca sexta*), for bioassays. Two gene combinations, differing with respect to inclusion of *CYP72A552*, were expressed in leaves using agro-infiltration: *LUP5+CYP716A80+UGT73C11* (C1) and *LUP5+CYP716A80+CYP72A552+UGT73C11* (C2) (Fig. 4a). To verify that the pathway was properly reconstituted, the major monoglycosides, as well as sapogenins, were identified in leaf extracts and quantified by LC-MS and GC-MS, respectively (Fig.

316 4b). When CYP72552 (C2) was co-expressed, hederagenin 3-*O*-monoglucoside
317 accumulated to 36.6 nmol $g^{-1}$ FW while concentrations of oleanolic acid 3-*O*-
318 monoglucoside decreased by 55% compared to C1 to 82.1 nmol $g^{-1}$ FW . Gypsogenic
319 acid 3-*O*-monoglucoside accumulated to 109 nmol/g FW in C2, while no gypsogenic
320 acid 3-*O*-monoglucoside was detected for C1. As for saponin aglycons, the
321 concentration of oleanolic acid decreased by 38% in C2 compared to C1, whereas
322 hederagenin accumulated to 5.8 nmol/g FW in C2 compared to trace amount in C1.
323 These results suggest that both gene combinations performed as expected in tobacco
324 plants and that hederagenin 3-*O*-monoglucoside only accumulated when *CYP72A552*
325 was co-expressed. The results also confirm the presence of enzyme systems in tobacco
326 that metabolize hederagenin further to gypsogenic acid (Fig. 2a), leading to an
327 accumulation of the non-toxic gypsogenic acid 3-*O*-monoglucoside at the expense of
328 toxic hederagenin 3-*O*-monoglucoside.

329 When tobacco leaves with the reconstituted saponin biosynthesis pathway were
330 presented to tobacco hornworm larvae, leaf consumption was significantly reduced for
331 C2 compared to control (15% reduction, $p < 0.01$; Fig. 4c). In contrast, consumption
332 was not reduced for C1. This suggests that *CYP72A552* causes the antifeedant activity
333 of the transiently produced saponins. The relatively low antifeedant activity in the
334 transient expression system relative to when saponins were painted on leaves (above)
335 reflects the low concentration of hederagenin 3-*O*-monoglucoside (36.6 nmol $g^{-1}$) that
336 accumulate in the tobacco leaves. In comparison, *B. vulgaris* leaves contain
337 approximately ~ 1.400 nmol $g^{-1}$.

338 To further determine if the antifeedant effect was directly caused by saponins that
339 accumulated when *CYP72A552* was co-infiltrated or if other metabolites might be
340 involved, a non-targeted metabolite analysis comparing C2 to C1 was performed using
341 LC-MS. Analysis of Variance (ANOVA) revealed that 44 mass signals, out of 8491,
342 were significantly upregulated in C2 compared to C1. The 44 mass signals correspond
343 to 15 metabolites, which were further tentatively identified as saponins containing
344 sapogenins with one or two hexoses, including hederagenin 3-*O*-monoglucoside as
345 verified with an authentic standard (Table S2). This result further substantiates that the





For Peer Review

deterrent effect of saponins against tobacco hornworm in tobacco leaves is an effect of CYP72A552.

**BvCYP72A552 evolved through gene duplication and selection**

A maximum likelihood analysis of CYP72A amino acid sequences from *B. vulgaris* and *A. thaliana* was performed to reconstruct their evolutionary history. Fig. 5a (blue marking) indicates that the three orthologous pairs, CYP72A7, CYP72A8, and CYP72A9, evolved before the two species diverged. Subsequently, an ancestral *CYP72A* duplicated and diverged independently in the two evolutionary lineages leading to *A. thaliana* and *B. vulgaris* (Fig. 5a: red marking). Non-synonymous to synonymous substitution ratios ($\omega = dN/dS$) were calculated to indicate whether the CYP72As evolved under different selection pressures. PAML analysis showed that the $\omega$ values of *CYP72A552* and *CYP72A562* are much higher than for the other *CYP72A*s (Fig. 5a), indicating they have been subject to stronger positive selection, especially in comparison with those evolved before the split of *A. thaliana* and *B. vulgaris*.

The order of the *CYP72As* in the tandem array corresponds well to the topology of the phylogenetic tree, and the order is in agreement with the syntenic tandem array in *A. thaliana* (Fig. 5b). Each of the three orthologous pairs of *B. vulgaris* and *A. thaliana* *CYP72A*s, (*CYP72A7*, *CYP72A8*, *CYP72A9*) are placed at one end of the tandem array in the same (but inverted) order, whereas the three non-orthologous genes (*BvCYP72A561P*, *BvCYP72A562*, and *BvCYP72A552*) are placed in the other end of the tandem array. The bootstrapped tree and $\omega$ values indicate that *BvCYP72A552* evolved through recent gene duplication. In summary, our results suggest that evolution of the saponin-based plant defense in the *Barbarea* genus was mediated by gene duplications followed by specializing selection.

## Discussion

Combined, *B. vulgaris* G- and P-type plants accumulate up to 49 different saponins structures, as verified by LC-MS-NMR (Khakimov *et al.*, 2016). Some of these saponin structures are more efficient in mediating resistance to insect herbivores than others (Agerbirk *et al.*, 2003; Kuzina *et al.*, 2009; Augustin *et al.*, 2012). Four oleanolic acid-derived saponins from *Barbarea vulgaris* are known to be constitutively produced and





For Peer Review

accumulate upon flea beetle attack (Kuzina *et al.*, 2011; Toneatto *et al.*, 2012), and they have been shown to correlate with flea beetles resistance in a QTL region containing a tandem array of eight *CYP72A*s (Shinoda *et al.*, 2002; Kuzina *et al.*, 2011; Byrne *et al.*, 2017). Here we demonstrate that only one of them, *CYP72A552,* oxidizes oleanolic acid to its corresponding alcohol, hederagenin, only, and that the monoglucoside of hederagenin is a major deterrent to the crucifer specialist herbivore diamondback moth and the solanaceous facultative 'specialist' tobacco hornworm. Transient expression of *CYP72A552* in tobacco leaves, together with other known genes in the *B. vulgaris* saponin pathway, led to accumulation of hederagenin-derived saponins and significantly reduced feeding by tobacco hornworm. Our results indicate that *BvCYP72A552* evolved through gene duplication after the split between the *A. thaliana* and *B. vulgaris* lineages and is under stronger positive selection than the *CYP72A*s before the split. Collectively, *CYP72A552* amplifies the deterrent (or toxic) effect of the oleanolic-based saponins and thus plays an essential role in mediating insect herbivore defense in *B. vulgaris*.

In previous studies by Fukushima *et al.* (2013), and Kim *et al.* (2018), *CYP72A68v2* from *Medicago truncatula* and *CYP714E19* from *Centella asiatica*, respectively, were reported to catalyze up to three consecutive hydroxylations of oleanolic acid at the C-23 position to first hederagenin, then gypsogenin and last gypsogenic acid by one P450. In this study, we show that when *CYP72A552* was co-expressed with upstream biosynthesis genes in tobacco leaves or yeast cells hederagenin, gypsogenin, and gypsogenic acid could be detected. However, our *in vitro* enzyme assay, using isolated yeast microsomes harboring CYP72A552, showed that CYP72A552 exclusively oxidizes oleanolic acid to hederagenin. Thus, we conclude that CYP72A552 oxidizes oleanolic acid at the C-23 position only to hederagenin, and that yeast and tobacco enzymes can further oxidize hederagenin to gypsogenic acid. In agreement with this, Han *et al.* (2017) found that hederagenin was the only product of yeast microsomes harboring *Kalopanax septemlobus* CYP72A397 when administrated with oleanolic acid.

*B. vulgaris,* and a few other closely related species, are the only crucifers known to produce saponins (Badenes-Perez *et al.*, 2013). None of the eight syntenic CYP72As located on *A. thaliana* chromosome 3 could oxidize oleanolic acid to hederagenin or





For Peer Review

other compounds (data not shown). Likewise, we have previously shown that *B. vulgaris* has evolved a UDP- glycosyltransferase, UGT73C11, that specifically catalyzes 3-*O*-glucosylation of sapogenins (Augustin *et al.*, 2012), which does not have a functional homolog in *A. thaliana*. Thus, at least two of the genes involved in the *Barbarea* saponin biosynthesis seem to have evolved and specialized subsequent to their divergence from the *A. thaliana* lineage. In both cases, this seems to have been mediated by gene duplication(s) of biosynthetic genes, initially shared with a common ancestor, followed by positive selection for novel function in the *Barbarea* lineage. In contrast to *CYP72A552*, neither *UGT73C11, CYP716A*80 or *CYP716A*81 are found close to the two QTLs for flea beetle resistance in *B. vulgaris*. This is to be expected, as all saponins are glucosylated at C-3 and most of them carboxylated at C-28, and not all *Barbarea* saponins contribute to resistance. In contrast, only some *B. vulgaris* saponins are hydroxylated at C-23 and among them are the defense-conferring hederagenin-based saponins. We have previously shown that CYP716s and OSCs are promiscuous enzymes (Khakimov *et al.*, 2015), both with respect to substrate and product specificity, and this may explain how *B. vulgaris* with a limited number of genes in the pathway may accumulate up to 49 different saponin structures (Khakimov *et al.*, 2016).

Our study provides important new insight into the structure-activity relationships of saponins, by showing that oxidative modification of the triterpenoid sapogenin may strongly increase or decrease bioactivity. A previous study has shown that a carboxyl group attached to the C-28 of oleanolic acid is crucial for antitumor and hemolytic activity (Carelli *et al.*, 2011; Sui *et al.*, 2017). Here, we showed that adding a hydroxyl group at C-23 significantly decreased feeding and survival of diamondback moth and tobacco hornworm. These results substantiates previous results with the crucifier specialist flea beetle that originally was used to identify the QTLs (Nielsen *et al.*, 2010; Augustin *et al.*, 2012; Christensen *et al.*, 2018).

The actual mode of action of saponins is largely unknown and we can therefore only speculate why the C-23 hydroxy group is so crucial for biological activity. Preliminary analysis of structural changes through molecular energy minimization modelling indicates that the glucose unit at the 3-*O* position of oleanolic acid monoglucoside is twisted ~ 90 degrees relative to the plane of the sapogenin backbone when a hydroxyl





For Peer Review

group is added at position C-23, leading to formation of hederagenin (Fig. S6). However, when a carboxyl group is introduced instead, the glucose unit is further twisted to a total ~180 degrees bringing the glucose more or less back to being in the same plane as the sapogenin backbone. Thus, the presence of the hydroxyl group at C-23 appears to be crucial for how the C-3 glucose is oriented towards the triterpenoid backbone, and this may be important for how saponin exert their toxicity at the molecular level. Saponins are generally believed to disrupt membranes by forming complexes with membrane sterols or to interact with steroidal receptors (Augustin *et al.*, 2011). Possibly, the C-23 hydroxylation by CYP72A552 induces a rotation in the saponin molecule changing its physiochemical properties, and this could possibly trigger membrane disruption, affinity to steroidal receptors, or decreases efficiency of saponin detoxifying enzymes in larval guts.

In this study, we introduced a fast and powerful system for determining structure-activity-relationships of transiently expressed biosynthetic genes in tobacco leaves for saponin production and subsequent feeding assays with insects. This compliments traditional feeding studies, where the metabolites are applied to leaf surfaces, by being more physiologically realistic as the tested compounds are produced and stored within leaves. The transient system takes less than one-week from introduction of the genes to insect bioassays can be done, as compared to months needed for stably transformed plants. Furthermore, the transient expression system allows for production of compounds that may otherwise be toxic or lethal to the plant if expressed systemically.

An unexpected drawback of the transient tobacco system is that hederagenin was further metabolized by endogenous tobacco enzymes, hampering the accumulation of hederagenin-3-*O*-monoglucoside. Thus, hederagenin 3-*O*-monoglucoside accumulated to 36.6 nmol $g^{-1}$ in the tobacco leaves (corresponding to 2.5 nmol $cm^{-2}$), whereas natural concentrations in flea beetle resistant *B. vulgaris* plants is ~ 1.400 nmol $g^{-1}$ (corresponding to ~ 96 nmol $cm^{-2}$) and concentrations used in our leaf disk assays were 100 nmol $cm^{-2}$. The relatively modest decrease in feeding on transiently saponin-expressing tobacco (15% reduction), in comparison to the feeding assays with saponin-painted leaf disks at the high concentration (90% reduction), is most likely due to the much lower dose of saponins.





For Peer Review

In summary, our study provides a key link between *CYP72A552* and saponin mediated plant defense against insects and reveals that its function has evolved through gene duplication and selection for novel function. Our study further highlights evolution of chemical novelties by gene duplication, recruitment of enzymes into new functions, and the importance of chemical modification in plant defense evolution.

**Acknowledgments**

We thank Professors Ian Baldwin and Marcel Dicke for providing *Manduca sexta* and *Plutella xylostella* material, respectively, Professor Giovanni Battista Appendino, Alberto Minassi and Federica Pollastro for providing saponin aglycon standards, Assistant Professor Henrik Hjarvard de Fine Licht for helping with sequence evolution analysis and Professor David Nelson for naming CYP72As according to the P450 nomenclature. This work was supported by Marie Skłodowska-Curie European Fellowship (H2020-MSCA-IF-2014, grant no. 655559 to SSAR), the Danish Council for Independent Research, Technology and Production Sciences (grant no. 1335-00151), the European Community's Seventh Framework Programme (FP7/2007-2013; grant no. 613692 to TriForC).

**Author contributions**

Q.L., T.P.H., and S.B. designed research; Q.L., B.K., and K.R.M. performed research; F.C. and C.E.O. contributed new reagents or analytic tools; Q.L. and B.K. analyzed data; Q.L., T.P.H., and S.B. wrote the paper.

The authors declare no conflict of interest.

## References


**Agerbirk N, Olsen CE, Bibby BM, Frandsen HO, Brown LD, Nielsen JK, Renwick JAA. 2003.** A saponin correlated with variable resistance of *Barbarea vulgaris* to the diamondback moth *Plutella xylostella*. *Journal of chemical ecology* **29**(6): 1417-1433.

**Aharoni A, Giri AP, Deuerlein S, Griepink F, de Kogel W-J, Verstappen FW, Verhoeven HA, Jongsma MA, Schwab W, Bouwmeester HJ. 2003.** Terpenoid metabolism in wild-type and transgenic *Arabidopsis* plants. *The Plant Cell* **15**(12): 2866-2884.

**Augustin JM, Drok S, Shinoda T, Sanmiya K, Nielsen JK, Khakimov B, Olsen CE, Hansen EH, Kuzina V, Ekstrøm CT. 2012.** UDP-glycosyltransferases from the UGT73C subfamily in *Barbarea vulgaris* catalyze sapogenin 3-O-glucosylation in saponin-mediated insect resistance. *Plant Physiology* **160**(4): 1881-1895.

**Augustin JM, Kuzina V, Andersen SB, Bak S. 2011.** Molecular activities, biosynthesis and evolution of triterpenoid saponins. *Phytochemistry* **72**(6): 435-457.

**Badenes-Perez FR, Reichelt M, Gershenzon J, Heckel DG. 2013.** Using plant chemistry and insect preference to study the potential of *Barbarea* (Brassicaceae) as a dead-end trap crop for diamondback moth (Lepidoptera: Plutellidae). *Phytochemistry*.

**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. *Molecular Plant* **8**(10): 1493-1506.

**Byrne SL, Erthmann PØ, Agerbirk N, Bak S, Hauser TP, Nagy I, Paina C, Asp T. 2017.** The genome sequence of *Barbarea vulgaris* facilitates the study of ecological biochemistry. *Scientific Reports* **7**.

**Carelli M, Biazzi E, Panara F, Tava A, Scaramelli L, Porceddu A, Graham N, Odoardi M, Piano E, Arcioni S. 2011.** *Medicago truncatula* CYP716A12 is a multifunctional oxidase involved in the biosynthesis of hemolytic saponins. *The Plant Cell Online* **23**(8): 3070-3081.

**Christensen S, Enge S, Munk K, Müller C, Kiaer L, Agerbirk N, Heimes C, Hauser T. 2018.** Different herbivore responses to two co-occurring chemotypes of the wild crucifer Barbarea vulgaris. *Arthropod-Plant Interactions*.

**De Geyter E, Lambert E, Geelen D, Smagghe G. 2007.** Novel advances with plant saponins as natural insecticides to control pest insects. *Pest Technology* **1**(2): 96-105.

**De Geyter E, Swevers L, Soin T, Geelen D, Smagghe G. 2012.** Saponins do not affect the ecdysteroid receptor complex but cause membrane permeation in insect culture cell lines. *Journal of insect physiology* **58**(1): 18-23.

**De Vos RC, Moco S, Lommen A, Keurentjes JJ, Bino RJ, Hall RD. 2007.** Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nature protocols* **2**(4): 778.

**Erthmann PØ, Agerbirk N, Bak S. 2018.** A tandem array of UDP-glycosyltransferases from the UGT73C subfamily glycosylate sapogenins,


For Peer Review

forming a spectrum of mono-and bisdesmosidic saponins. *Plant molecular biology*: 1-19.

**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 and Cell Physiology* **54**(5): 740-749.

**Hall BG. 2011.** *Phylogenetic trees made easy: a how-to manual*: Oxford university press.

**Han JY, Chun J-H, O SA, Park S-B, Hwang H-S, Lee H, Choi YE. 2017.** Transcriptomic analysis of *Kalopanax septemlobus* and characterization of KsBAS, CYP716A94, and CYP72A397 genes involved in hederagenin saponin biosynthesis. *Plant and Cell Physiology*: pcx188-pcx188.

**Hauser TP, Toneatto F, Nielsen JK. 2012.** Genetic and geographic structure of an insect resistant and a susceptible type of *Barbarea vulgaris* in western Europe. *Evolutionary Ecology* **26**(3): 611-624.

**Hayashi H, Huang P, Kirakosyan A, Inoue K, Hiraoka N, Ikeshiro Y, Kushiro T, Shibuya M, Ebizuka Y. 2001.** Cloning and characterization of a cDNA encoding beta-amyrin synthase involved in glycyrrhizin and soyasaponin biosyntheses in licorice. *Biological & pharmaceutical bulletin* **24**(8): 912-916.

**Husselstein–Muller T, Schaller H, Benveniste P. 2001.** Molecular cloning and expression in yeast of 2, 3–oxidosqualene–triterpenoid cyclases from *Arabidopsis thaliana*. *Plant Molecular Biology* **45**(1): 75-92.

**Jones DT, Taylor WR, Thornton JM. 1992.** The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* **8**(3): 275-282.

**Khakimov B, Kuzina V, Erthmann PØ, Fukushima EO, Augustin JM, Olsen CE, Scholtalbers J, Volpin H, Andersen SB, Hauser TP. 2015.** Identification and genome organization of saponin pathway genes from a wild crucifer, and their use for transient production of saponins in *Nicotiana benthamiana*. *The Plant Journal* **84**(3): 478-490.

**Khakimov B, Tseng LH, Godejohann M, Bak S, Engelsen SB. 2016.** Screening for Triterpenoid Saponins in Plants Using Hyphenated Analytical Platforms. *Molecules* **21**(12): 1614.

**Kim OT, Um Y, Jin ML, Kim JU, Hegebarth D, Busta L, Racovita RC, Jetter R. 2018.** A Novel Multifunctional C-23 Oxidase, CYP714E19, is Involved in Asiaticoside Biosynthesis. *Plant and Cell Physiology* **59**(6): 1200-1213.

**Kumar P, Pandit SS, Baldwin IT. 2012.** Tobacco rattle virus vector: a rapid and transient means of silencing *Manduca sexta* genes by plant mediated RNA interference. *PloS one* **7**(2): e31347.

**Kumar S, Stecher G, Tamura K. 2016.** MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution* **33**(7): 1870-1874.

**Kushiro T, Shibuya M, Ebizuka Y. 1998.** β-Amyrin synthase: cloning of oxidosqualene cyclase that catalyzes the formation of the most popular triterpene among higher plants. *European journal of biochemistry* **256**(1): 238-244.

**Kuzina V, Ekstrøm CT, Andersen SB, Nielsen JK, Olsen CE, Bak S. 2009.** Identification of defense compounds in *Barbarea vulgaris* against the herbivore *Phyllotreta nemorum* by an ecometabolomic approach. *Plant physiology* **151**(4): 1977-1990.

**Kuzina V, Nielsen JK, Augustin JM, Torp AM, Bak S, Andersen SB. 2011.** *Barbarea vulgaris* linkage map and quantitative trait loci for saponins, glucosinolates, hairiness and resistance to the herbivore *Phyllotreta nemorum*. *Phytochemistry* **72**(2–3): 188-198.

**Liu Q, Manzano D, Tanić N, Pesic M, Bankovic J, Pateraki I, Ricard L, Ferrer A, de Vos R, de Krol Sv. 2014.** Elucidation and *in planta* reconstitution of the parthenolide biosynthetic pathway. *Metabolic Engineering*.

**Ma C, Wang H, Lu X, Wang H, Xu G, Liu B. 2009.** Terpenoid metabolic profiling analysis of transgenic *Artemisia annua* L. by comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry. *Metabolomics* **5**(4): 497-506.

**Nielsen JK, Nagao T, Okabe H, Shinoda T. 2010.** Resistance in the plant, *Barbarea vulgaris*, and counter-adaptations in flea beetles mediated by saponins. *Journal of chemical ecology* **36**(3): 277-285.

**Nour-Eldin H, Geu-Flores F, Halkier B 2010.** USER Cloning and USER Fusion: The Ideal Cloning Techniques for Small and Big Laboratories. In: Fett-Neto AG ed. *Plant Secondary Metabolism Engineering*: Humana Press, 185-200.

**Paquette SM, Bak S, Feyereisen R. 2000.** Intron–exon organization and phylogeny in a large superfamily, the paralogous cytochrome P450 genes of *Arabidopsis thaliana*. *DNA and cell biology* **19**(5): 307-317.

**Renwick JAA. 2002.** The chemical world of crucivores: lures, treats and traps. *Entomologia experimentalis et applicata* **104**(1): 35-42.

**Sainsbury F, Thuenemann EC, Lomonossoff GP. 2009.** pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant biotechnology journal* **7**(7): 682-693.

**Sawai S, Shindo T, Sato S, Kaneko T, Tabata S, Ayabe S-i, Aoki T. 2006.** Functional and structural analysis of genes encoding oxidosqualene cyclases of *Lotus japonicus*. *Plant Science* **170**(2): 247-257.

**Seki H, Ohyama K, Sawai S, Mizutani M, Ohnishi T, Sudo H, Akashi T, Aoki T, Saito K, Muranaka T. 2008.** Licorice β-amyrin 11-oxidase, a cytochrome P450 with a key role in the biosynthesis of the triterpene sweetener glycyrrhizin. *Proceedings of the National Academy of Sciences* **105**(37): 14204-14209.

**Seki H, Sawai S, Ohyama K, Mizutani M, Ohnishi T, Sudo H, Fukushima EO, Akashi T, Aoki T, Saito K. 2011.** Triterpene functional genomics in licorice for identification of CYP72A154 involved in the biosynthesis of glycyrrhizin. *The Plant Cell Online* **23**(11): 4112-4123.

**Seki H, Tamura K, Muranaka T. 2015.** P450s and UGTs: key players in the structural diversity of triterpenoid saponins. *Plant Cell Physiol*.

**Shinoda T, Nagao T, Nakayama M, Serizawa H, Koshioka M, Okabe H, Kawai A. 2002.** Identification of a Triterpenoid Saponin from a Crucifer, *Barbarea vulgaris*, as a Feeding Deterrent to the Diamondback Moth, *Plutella xylostella*. *Journal of Chemical Ecology* **28**(3): 587-599.

**Sui J-J, Zhou W-H, Liu D-Y, Li M-Q, Sun J-S. 2017.** Highly efficient synthesis of bioactive oleanane-type saponins. *Carbohydrate Research*.

**Toneatto F, Hauser TP, Nielsen JK, Ørgaard M. 2012.** Genetic diversity and similarity in the *Barbarea vulgaris* complex (Brassicaceae). *Nordic Journal of Botany* **30**(4): 506-512.

**van Mölken T, Kuzina V, Munk KR, Olsen CE, Sundelin T, van Dam NM, Hauser TP. 2014.** Consequences of combined herbivore feeding and pathogen infection for fitness of *Barbarea vulgaris* plants. *Oecologia* **175**(2): 589-600.

**Xu B, Yang Z. 2013.** PAMLX: a graphical user interface for PAML. *Molecular biology and evolution* **30**(12): 2723-2724.

**Xu H, Moghe GD, Wiegert-Rininger K, Schilmiller AL, Barry CS, Last RL, Pichersky E. 2018.** Coexpression Analysis Identifies Two Oxidoreductases Involved in the Biosynthesis of the Monoterpene Acid Moiety of Natural Pyrethrin Insecticides in *Tanacetum cinerariifolium*. *Plant Physiology* **176**(1): 524-537.

**Yang T, Stoopen G, Yalpani N, Vervoort J, de Vos R, Voster A, Verstappen FWA, Bouwmeester HJ, Jongsma MA. 2011.** Metabolic engineering of geranic acid in maize to achieve fungal resistance is compromised by novel glycosylation patterns. *Metabolic Engineering* **13**(4): 414-425.

**Yang Z. 2007.** PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution* **24**(8): 1586-1591.

For Peer Review

**Supporting Information**

Additional supporting information may be found in the online version of this article.

Fig. S1. Amino acid sequences alignment of the seven CYP72As from *B. vulgaris* possible involved in saponin biosynthesis and the eight syntenic *A. thaliana* CYP72s.

Fig. S2. GC-MS profile of yeast (*S. cerevisiae*) microsome expressing *CYP72A552* administrated with hederagenin (HE).

Fig. S3. Large-scale production of monoglucosides of oleanolic acid (OA), hederagenin (HE), and gypsogenic acid (GA).

Fig. S4. Purification of the monoglucosides of oleanolic acid (OA), hederagenin (HE), and gypsogenic acid (GA).

Fig. S5. Structure of 3-*O*-monoglucoside gypsogenic acid elucidated by NMR.

Fig. S6. 3D structure of 3-*O*-oleanolic acid mono-glucoside (a), 3-*O*-hederagenin mono-glucoside (b), and 3-*O*-gypsogenic acid mono-glucoside (c).

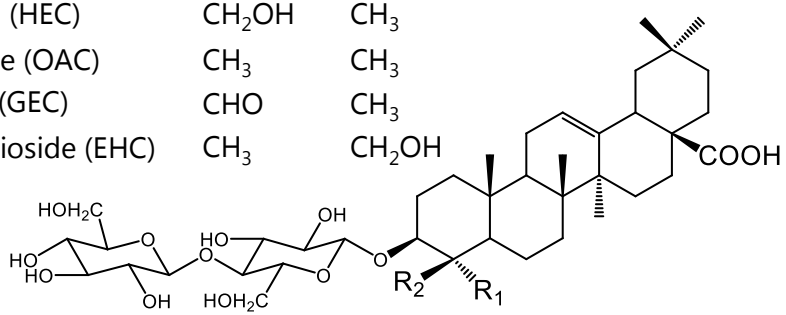
Table S1. 1D 1H and HSQC based 13C NMR data of 3-*O*-monoglucoside gypsogenic acid.

Table S2. Tentative identification of metabolites significantly changed between plants transiently expressing the two saponin biosynthetic gene combinations.

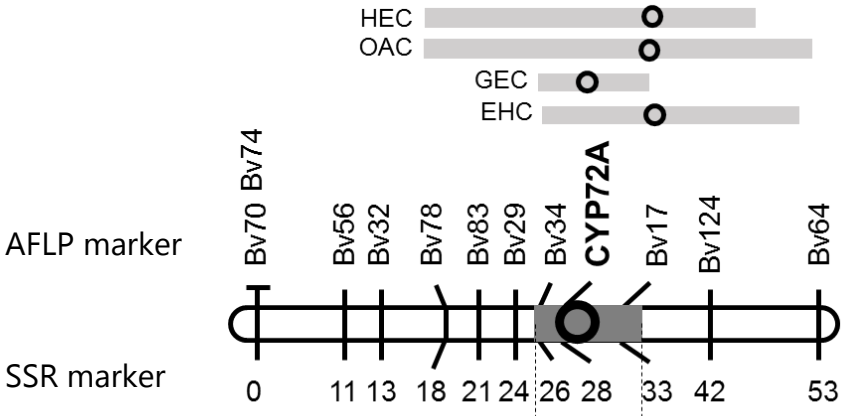
For Peer Review

a

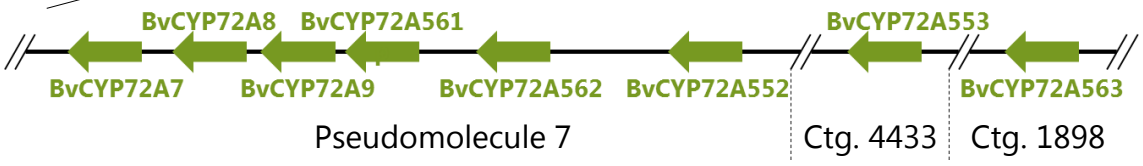
| | $R_1$ | $R_2$ |
|---|---|---|
| hederagenin cellobioside (HEC) | $CH_2OH$ | $CH_3$ |
| oleanolic acid cellobioside (OAC) | $CH_3$ | $CH_3$ |
| gypsogenin cellobioside (GEC) | CHO | $CH_3$ |
| 4-epi hederagenin cellobioside (EHC) | $CH_3$ | $CH_2OH$ |

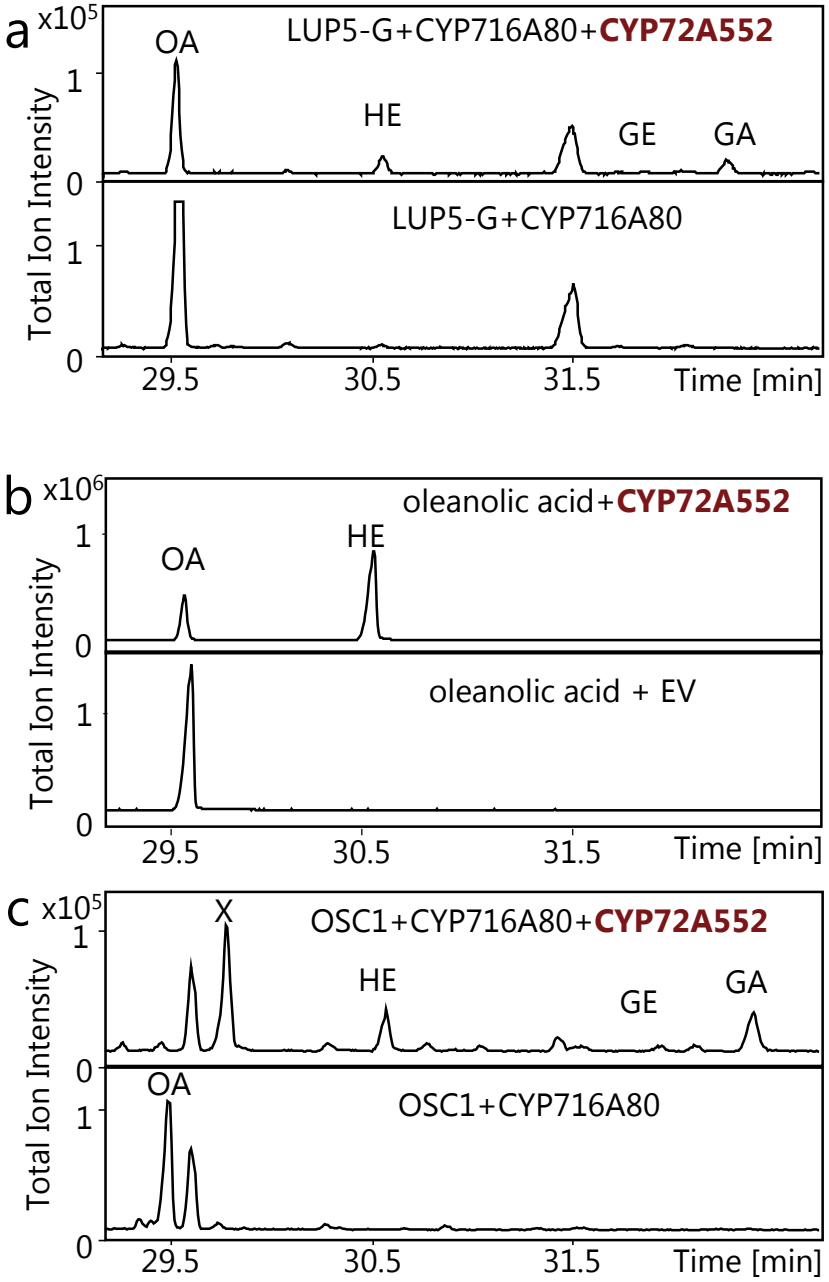


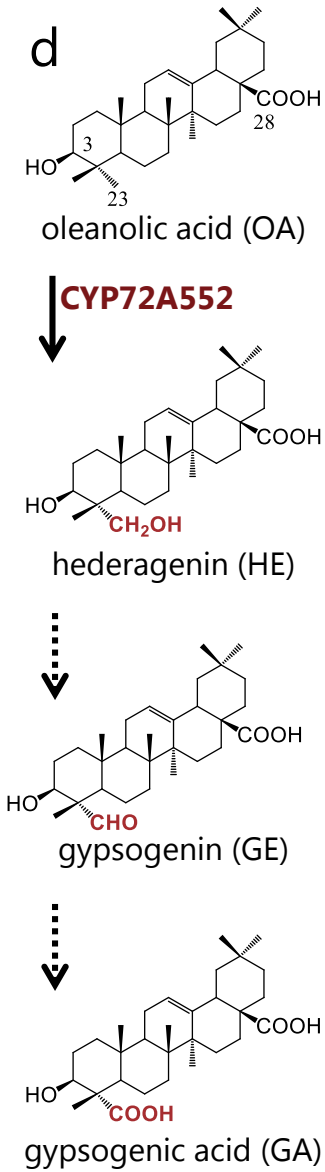

b

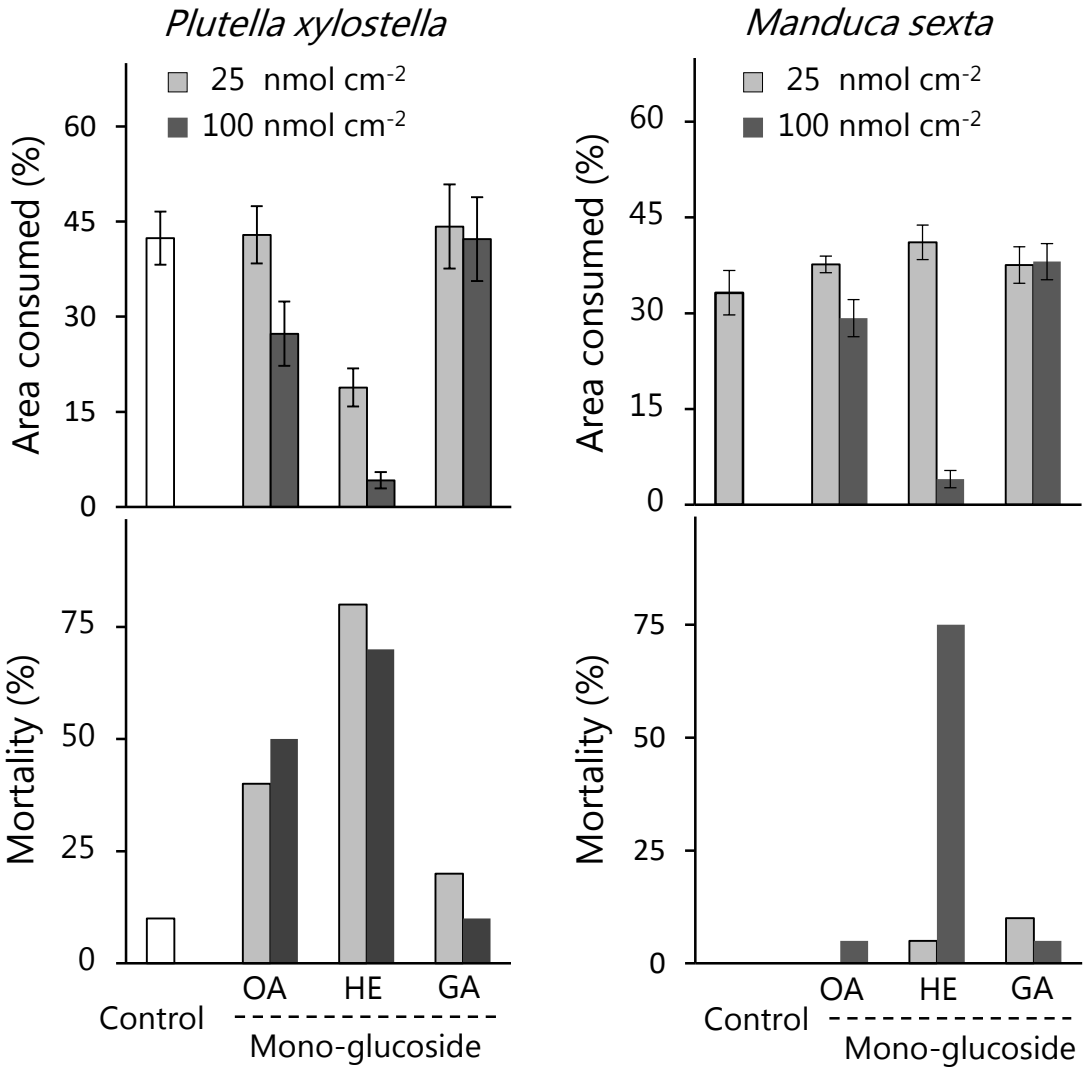



c



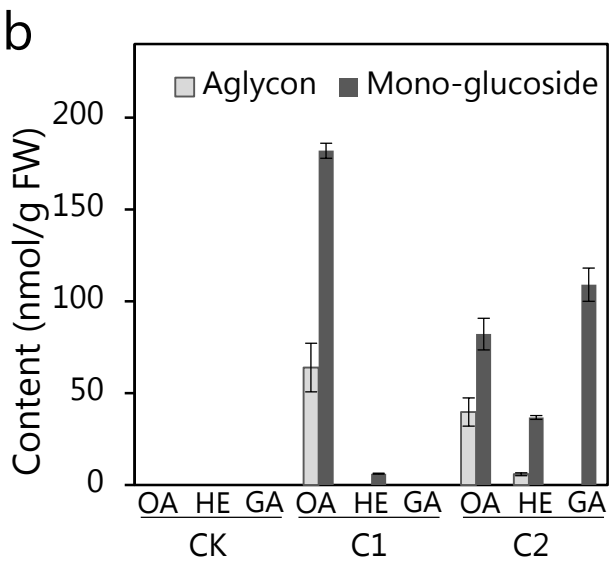

a
x10^5
LUP5-G+CYP716A80+CYP72A552
OA
HE
GE
GA
LUP5-G+CYP716A80
Total Ion Intensity
Time [min]
b
x10^6
oleanolic acid+CYP72A552
oleanolic acid + EV
c
x10^5
X
OSC1+CYP716A80+CYP72A552
OSC1+CYP716A80
29.5
30.5
31.5

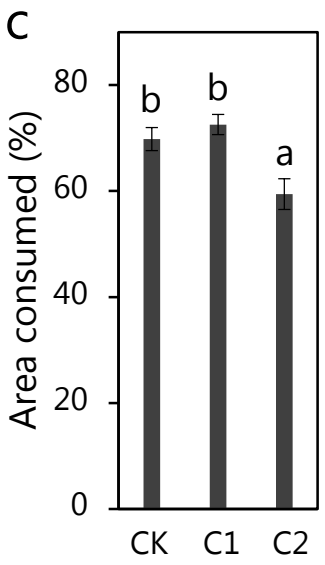

d
COOH
28
HO
3
23
oleanolic acid (OA)
CYP72A552
CH2OH
hederagenin (HE)
CHO
gypsogenin (GE)
gypsogenic acid (GA)

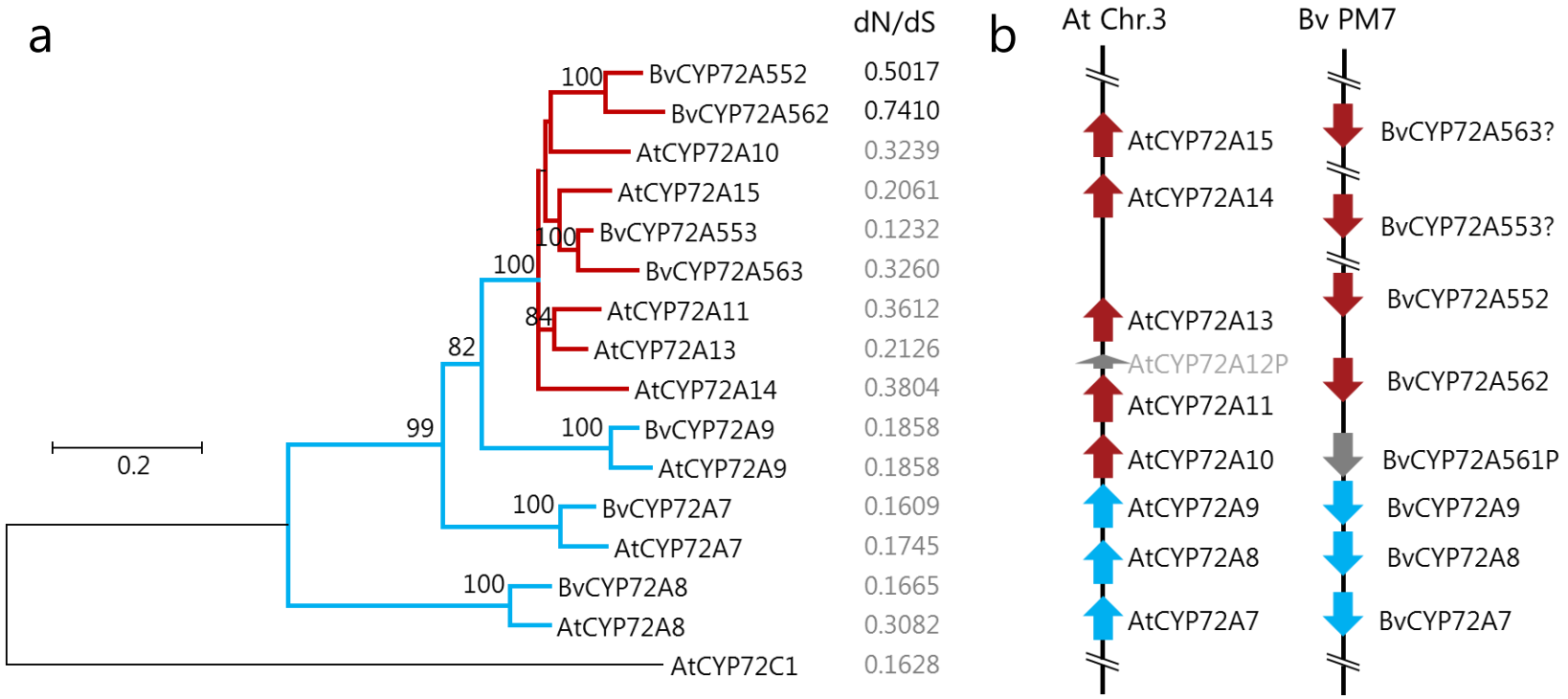

Plutella xylostella
25 nmol cm-2
100 nmol cm-2
Area consumed (%)
60
45
30
15
0
Mortality (%)
75
50
25
0
Control
OA
HE
GA
Mono-glucoside
Manduca sexta
25 nmol cm-2
100 nmol cm-2
Area consumed (%)
60
45
30
15
0
Mortality (%)
75
50
25
0
Control
OA
HE
GA
Mono-glucoside

a

| | Gene combination |
|---|---|
| C1 | LUP5+CYP716A80+UGT73C11 |
| C2 | LUP5+CYP716A80+**CYP72A552**+UGT73C11 |
| Control (CK) | Empty vector |




a
dN/dS
BvCYP72A552
BvCYP72A562
AtCYP72A10
AtCYP72A15
BvCYP72A553
BvCYP72A563
AtCYP72A11
AtCYP72A13
AtCYP72A14
BvCYP72A9
AtCYP72A9
BvCYP72A7
AtCYP72A7
BvCYP72A8
AtCYP72A8
AtCYP72C1
100
100
100
84
82
99
100
100
100
0.2
0.5017
0.7410
0.3239
0.2061
0.1232
0.3260
0.3612
0.2126
0.3804
0.1858
0.1858
0.1609
0.1745
0.1665
0.3082
0.1628
b
At Chr.3
Bv PM7
AtCYP72A15
AtCYP72A14
AtCYP72A13
AtCYP72A12P
AtCYP72A11
AtCYP72A10
AtCYP72A9
AtCYP72A8
AtCYP72A7
BvCYP72A563?
BvCYP72A553?
BvCYP72A552
BvCYP72A562
BvCYP72A561P
BvCYP72A9
BvCYP72A8
BvCYP72A7