Plant Molecular Biology
https://doi.org/10.1007/s11103-018-0723-z

A tandem array of UDP-glycosyltransferases from the UGT73C subfamily glycosylate sapogenins, forming a spectrum of mono- and bisdesmosidic saponins

Pernille Østerbye Erthmann[1] · Niels Agerbirk[1] · Søren Bak[1]

Received: 22 November 2017 / Accepted: 25 March 2018

Abstract

Key message **This study identifies six UGT73Cs all able to glucosylate sapogenins at positions 3 and/or 28 which demonstrates that *B. vulgaris* has a much richer arsenal of UGTs involved in saponin biosynthesis than initially anticipated.**

Abstract The wild cruciferous plant *Barbarea vulgaris* is resistant to some insects due to accumulation of two monodesmosidic triterpenoid saponins, oleanolic acid 3-*O*-β-cellobioside and hederagenin 3-*O*-β-cellobioside. Insect resistance depends on the structure of the sapogenin aglycone and the glycosylation pattern. The *B. vulgaris* saponin profile is complex with at least 49 saponin-like metabolites, derived from eight sapogenins and including up to five monosaccharide units. Two *B. vulgaris* UDP-glycosyltransferases, UGT73C11 and UGT73C13, *O*-glucosylate sapogenins at positions 3 and 28, forming mainly 3-*O*-β-D-glucosides. The aim of this study was to identify UGTs responsible for the diverse saponin oligoglycoside moieties observed in *B. vulgaris*. Twenty UGT genes from the insect resistant genotype were selected and heterologously expressed in *Nicotiana benthamiana* and/or *Escherichia coli*. The extracts were screened for their ability to glycosylate sapogenins (oleanolic acid, hederagenin), the hormone 24-epibrassinolide and sapogenin monoglucosides (hederagenin and oleanolic acid 3-*O*-β-D-glucosides). Six UGTs from the UGT73C subfamily were able to glucosylate both sapogenins and both monoglucosides at positions 3 and/or 28. Some UGTs formed bisdesmosidic saponins efficiently. At least four *UGT73C* genes were localized in a tandem array with *UGT73C11* and possibly *UGT73C13*. This organization most likely reflects duplication events followed by sub- and neofunctionalization. Indeed, signs of positive selection on several amino acid sites were identified and modelled to be localized on the UGT protein surface. This tandem array is proposed to initiate higher order bisdesmosidic glycosylation of *B. vulgaris* saponins, leading to the recently discovered saponin structural diversity, however, not directly to known cellobiosidic saponins.

Keywords UDP-glucosyltransferases · Bisdesmosidic triterpenoid saponins · Tandem repeat · *Barbarea vulgaris* · Evolution of plant chemical defense compounds

Introduction

Plants produce an array of low molecular weight defense compounds like glucosinolates, flavonoids and saponins. Glucosinolates are found mostly in the plant order Brassicales, while flavonoids are found in most plant species. In contrast, saponins are found scattered in many unrelated plant species, and this phylogenetic distribution of saponins suggests that the saponin pathways have convergently evolved several times (Augustin et al. 2011; Osbourn 2003; Vincken et al. 2007).

Saponins consist of a hydrophobic triterpenoid backbone, the sapogenin, and typically one or two hydrophilic carbohydrate moieties, with each moiety consisting of varying

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11103-018-0723-z) contains supplementary material, which is available to authorized users.

✉ Søren Bak
bak@plen.ku.dk

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

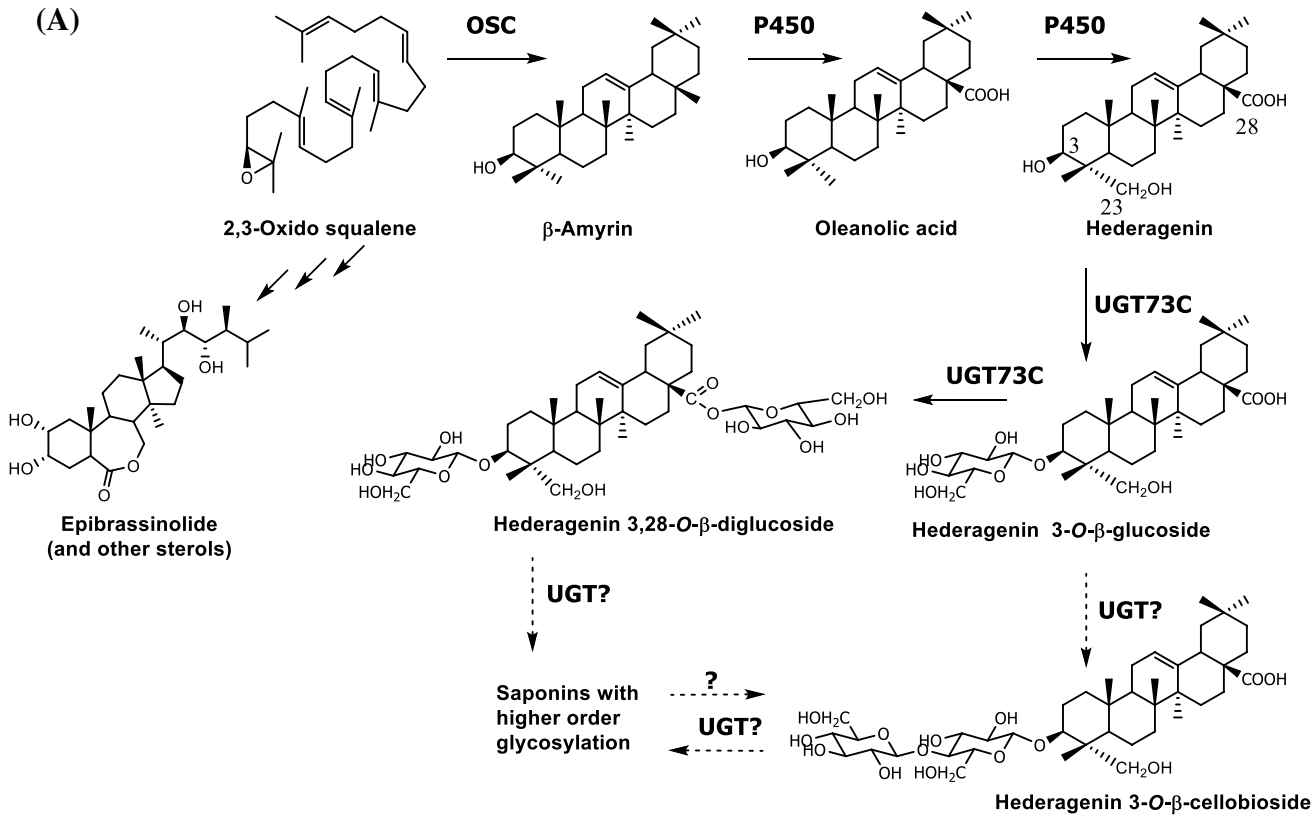
Published online: 30 March 2018

numbers of monosaccharide units (Augustin et al. 2011; Osbourn 2003; Vincken et al. 2007). The backbone can be classified as either a triterpenoid or a sterol backbone, with the former derived directly from 2,3-oxidosqualene and the latter formed via precursors from the sterol pathway, although a more detailed sub-classification based on backbone structures as suggested by Vincken et al. (2007) is rewarding. Saponins with one carbohydrate moiety are known as monodesmosidic, while those with two carbohydrate moieties are known as bisdesmosidic. Monodesmosidic saponins are reported to exert toxicity or other negative effects on a variety of organisms, with anti-fungal monodesmosidic saponins from oat (*Avena sativa*) being a well-studied example (Osbourn 1996). The molecular mechanism behind monodesmosidic saponin activity is not fully understood but is believed to involve disturbance of membrane structure due to association with membrane steroids (Augustin et al. 2011; Osbourn 2003). Additionally, saponins are well known to have hemolytic activities on red blood cells (Vo et al. 2016), and are able to permeate insect cell membranes (De Geyter et al. 2012).

The biosynthesis of saponins with a triterpenoid backbone (called saponins hereafter) generally involves three main steps (reviewed by Augustin et al. 2011; Seki et al. 2008; Thimmappa et al. 2014); first the backbone ring structures are formed by cyclization of 2,3-oxidosqualene mediated by 2,3-oxidosqualene cyclases (OSCs) (Khakimov et al. 2015; Osbourn 2003), then functional groups e.g. hydroxyl or epoxide groups are added by cytochromes P450 (P450s) to form the sapogenin (Carelli et al. 2011; Geisler et al. 2013; Khakimov et al. 2015; Kunii et al. 2012; Seki et al. 2008), and finally various saccharides are added by family 1 UDP-glycosyltransferases (UGTs) generating the saponins (Li et al. 2016; Meesapyodsuk et al. 2007; Shibuya et al. 2010) (Fig. 1a).

Our research on saponins relates to how pest-deterrent saponins in the wild cruciferous plant *Barbarea vulgaris* have evolved at the molecular level, and their function in plant-insect interactions. *Barbarea* is unique within the Brassicales order because it is the only genus known to produce saponins. *B. vulgaris* is polymorphic for saponin content (Khakimov et al. 2016) and for resistance to crucifer-adapted herbivorous insects such as a particular flea beetle species (*Phyllotreta nemorum*) and the diamondback moth (*Plutella xylostella*) (Agerbirk et al. 2001, 2003b). In Eastern and Central Europe, two ecotypes of *B. vulgaris* are found, generally known as the G-type (Glabrous type) and the P-type (Pubescent type) (Agerbirk et al. 2015; Christensen et al. 2014; Heimes et al. 2016). The P-type (Pubescent type) is the common *B. vulgaris* ecotype in Eastern Europe and is susceptible to the above mentioned herbivores. In contrast, the G-type is resistant to *P. nemorum* and the diamondback moth due to the accumulation of saponins (Kuzina et al. 2009; Nielsen et al. 2010a; Shinoda et al. 2002). Five triterpenoid saponins, all derived from β-amyrin have been structurally identified from the G-type; hederagenin 3-*O*-β-cellobioside, oleanolic acid 3-*O*-β-cellobioside, 4-epihederagenin 3-*O*-β-cellobioside, gypsogenin 3-*O*-β-cellobioside and cochalic acid 3-*O*-β-cellobioside (Fig. 1b) (Agerbirk et al. 2003a; Nielsen et al. 2010a; Shinoda et al. 2002). For brevity, the shared 3-*O*-β-cellobioside moieties are in the rest of this paper simply referred to as cellobioside moieties. Hederagenin cellobioside and oleanolic cellobioside are quantitatively dominant and seem to be major determinants of resistance due to strong post-ingestive deterrent activity (Agerbirk et al. 2003a; Kuzina et al. 2009; Nielsen et al. 2010b; Shinoda et al. 2002). In contrast to the known β-amyrin derived G-type saponins, the P-type saponins are mainly derived from lupeol (Khakimov et al. 2016). However, no P-type saponin structure has yet been elucidated with respect to the glycosidic part.

In addition to the above-mentioned saponins, a vast number of other saponins have been tentatively identified in *B. vulgaris* (Khakimov et al. 2016). These structures were not fully elucidated, rather, numbers of glycosyl groups, chemical shifts of anomeric carbons and apparent molecular masses of the aglycons were identified. In total, 27 saponin-like metabolites were partially identified in the G-type and 22 in the P-type, all with between two and five monosaccharides linked to the sapogenin moiety (Khakimov et al. 2016). It is an open question whether each structural variant is reflected in bioactivity. However, the presence of the cellobioside moiety at position 3 in G-type saponins has been confirmed to be critical for insect deterrence. In feeding assays, flea beetles were presented to either the main naturally occurring monodesmosidic saponin cellobiosides; oleanolic acid cellobioside and hederagenin cellobioside or the corresponding sapogenins; oleanolic acid and hederagenin (Nielsen et al. 2010a), and only the saponin cellobiosides exhibited deterrence, demonstrating the impact of glycosylation in position 3 for insect deterrence. Furthermore, the presence of a hydroxyl group in position 23 on the sapogenin is critical as hederagenin cellobioside was significantly more flea beetle-deterrent than oleanolic acid cellobioside (Nielsen et al. 2010a). Finally, experiments with flea beetles resistant to hederagenin cellobioside were carried out, demonstrating that a single flea beetle gene (Mendelian locus) for resistance to hederagenin cellobioside did not confer resistance to another monodesmosidic saponin—α-hederin—with the same hederagenin aglycone but containing a different disaccharide moiety (Nielsen et al. 2010a). In combination, these experiments provided evidence that specific backbone-decorating steps and specific glycosylation steps are critical to the natural pest resistance of *B. vulgaris*. Similar results, also showing saponin activity to be depending on glycoside



(B)

	R_1	R_2	R_3
Oleanolic acid cellobioside	H	CH_3	CH_3
Cochalic acid cellobioside	OH	CH_3	CH_3
4-epi-hederagenin cellobioside	H	CH_3	CH_2OH
Hederagenin cellobioside	H	CH_2OH	CH_3
Gypsogenin cellobioside	H	CHO	CH_3

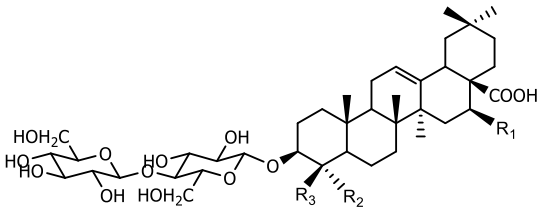


Fig. 1 Overview of saponins and their biosynthesis in *Barbarea vulgaris*. **a** Overview of known and hypothetic reactions in hederagenin-based saponin biosynthesis in *B. vulgaris*. Known reactions are indicated with fully drawn arrows. After the formation of 2,3-oxidosqualene, the saponin biosynthesis branches off from the sterol biosynthesis by cyclizing 2,3-oxidosqualene into the saponin backbone structures exemplified by β-amyrin. The backbone structures are further modified by enzymes such as cytochromes P450 monooxygenases (P450s) and UDP-glycosyltransferases (UGTs). Positions for putative glycosylation of hederagenin are marked with the corresponding carbon number. Hypothetic reactions discussed in the text are indicated with broken arrows and question marks after enzyme names. **b** Structures of the five structurally similar saponins that have been identified in *B. vulgaris* to date. All five have been isolated from the G-type of the species, and oleanolic acid cellobioside and hederagenin cellobioside are quantitatively dominant

moiety position and structure, have been reported by other authors (Osbourn 2003; Vo et al. 2016).

Until now, two UGTs involved in sapogenin glucosylation from the insect resistant G-type have been identified and characterized; BvUGT73C13 and BvUGT73C11 (Augustin et al. 2012). BvUGT73C13 glucosylates oleanolic acid and hederagenin at position 3 and 28 (Fig. 1a), whereas BvUGT73C11 is more position specific and preferentially glucosylates at position 3 with a low K_M. Surprisingly, despite the importance of glycosylation for pest resistance, the genomic localizations of BvUGT73C11 and BvUGT73C13 did not co-localize with the QTL locations (quantitative trait loci) for neither flea beetle resistance nor saponin production (Kuzina et al. 2011).

In the present study, we mined the transcriptome (Kuzina et al. 2011) for additional UGTs possibly involved in *B. vulgaris* saponin biosynthesis. Using this extended set of UGTs, we explored their genomic architecture and enzymatic activities. We characterized 23 *UGT* genes in G-type *B. vulgaris* and showed that four or perhaps six of these *(BvUGT73C21,*

BvUGT73C22, BvUGT73C23, BvUGT73C26 and possibly *BvUGT73C25, BvUGT73C27*) are organized in a tandem array on pseudomolecule 3 and may be involved in saponin biosynthesis, based on their ability to glucosylate sapogenins after transient expression in *Nicotiana benthamiana* and heterologous expression in *Escherichia coli*.

Results

Screening of *UGT* candidates for involvement in saponin biosynthesis

To identify *UGTs* involved in glycosylation of saponins in *B. vulgaris*, a *B. vulgaris* transcriptome (Kuzina et al. 2011) derived from 3- to 12-week-old G-type plants grown in a growth chamber and known to accumulate saponins was mined. As other genes in saponin biosynthesis were known to be represented in this transcriptome (Augustin et al. 2012; Khakimov et al. 2015), the missing additional UGT(s) were also expected to be represented. Nineteen full-length *UGT* candidates covering UGT families 71, 72, 73, 79 and 91 (Fig. 2) were identified. Partial hits were not pursued based on the hypothesis that good candidates will be expressed at a reasonable level and thus be full-length in the transcriptome. However, an additional subfamily 73C member (BvUGT73C27) was later included in functional analysis after discovery in the genome, because this subfamily attracted particular attention, increasing the number of screened UGTs to 20. Finally, three members of the family 74 as well as the two previously characterized BvUGT73C11 and BvUGT73C13 were included in the phylogenetic tree (Fig. 2). The genomic positions of the *UGT* candidates were also compared with genomic areas (QTLs) previously linked to saponin production or flea beetle resistance (Byrne et al. 2017; Kuzina et al. 2011), however, none of the candidates localized in these areas.

The 19 initially identified *UGTs* were screened for involvement in glucosylation of sapogenins and saponin glucosides by transient expression in *N. benthamiana*. To assess their ability to glucosylate the monoglucosidic products of BvUGT73C11 they were transiently co-expressed with three enzymes forming the expected monoglucosidic saponin precursors; BvLUP5-G, BvCYP716A81, and BvUGT73C11 (Khakimov et al. 2015). In the biosynthesis of oleanolic acid 3-*O*-β-glucoside, BvLUP5-G catalyzes the cyclization of 2,3-oxidosqualene to β-amyrin, BvCYP716A81 hydroxylates β-amyrin at C28 into oleanolic acid, and finally BvUGT73C11 glucosylates oleanolic acid into oleanolic acid 3-*O*-β-glucoside. The candidate *UGTs* were initially screened in pools of 7, however some of the UGT candidates later turned out to be caused by PCR errors, and these were neither given a systematic name nor included in the overview (Table S1 in Supplementary Material). Leaf methanol extracts were analyzed by LC-MS/MS and total ion chromatograms showed minor changes when expressing Pool #1 compared to the control (*BvLUP5-G, BvCYP716A80* and *BvUGT73C11*), whereas Pool #2, Pool #3 and Pool #4 showed highly differentiated metabolite profiles (Fig. 3).

To identify UGT pools conferring glucosylation of saponins, targeted extracted ion chromatograms were analyzed for the presence of mono- and also higher order sapogenin glycosides as it is well known that *N. benthamiana* has endogenous glucosyl transferases that may glucosylate metabolically engineered saponins (Khakimov et al. 2015). *B. vulgaris UGT* pool #4 (*BvUGT71B11, BvUGT73C21, BvUGT73C22, BvUGT73C23, BvUGT73C24, BvUGT73C25*, and *BvUGT73C26*) resulted in changes in the profile of putative saponins, while no apparent major effects on putative saponins were observed for the remaining three *B. vulgaris UGT* pools when compared to the control (Fig. 3).

In addition to the *in planta* screening in *N. benthamiana*, Transcription/Translation (TnT) assays were performed using the 23 identified *UGT* genes (Table S1 in Supplementary Material). In these assays, the ability of the UGTs to glucosylate the UGT generic substrate 2,4,6-trichlorophenol (TCP) was included as a positive control for expression of the UGTs in a functional fold (Meßner et al. 2003) (Table S1 in Supplementary Material). Five of the newly discovered *BvUGT73Cs* (*BvUGT73C21, BvUGT73C22, BvUGT73C23, BvUGT73C24* and *BvUGT73C26*), and seven of the remaining 14 *UGT* genes, (*BvUGT71B10, BvUGT71C6, BvUGT71C7, BvUGT71C8, BvUGT71D3, BvUGT72D3* and *BvUGT71B11*) were able to glucosylate TCP, meaning that we can consider them to encode proteins folded in a functionally state. Further work focused on the newly discovered BvUGT73C sub-family members (*BvUGT73C21, BvUGT73C22, BvUGT73C23, BvUGT73C25, BvUGT73C26*) since their pool lead to changes in the saponin targeted LC-MS chromatograms in the *N. benthamiana* screening (Fig. 2). This included the later discovered subfamily-member (*BvUGT73C27*). The last members of pool 4, *BvUGT71B11* and *BvUGT73C24*, were not further pursued. The 71B subfamily member was not further characterized because it was prioritized to characterize the subfamily 73C members found in the same cluster as the previously implied UGTs in the biosynthesis. Likewise, *BvUGT73C24* was not further characterized because it was orthologous to *AtUGT73C7* from *A. thaliana* (Fig. 2) and thereby assumed not being involved in saponin biosynthesis in *B. vulgaris*.

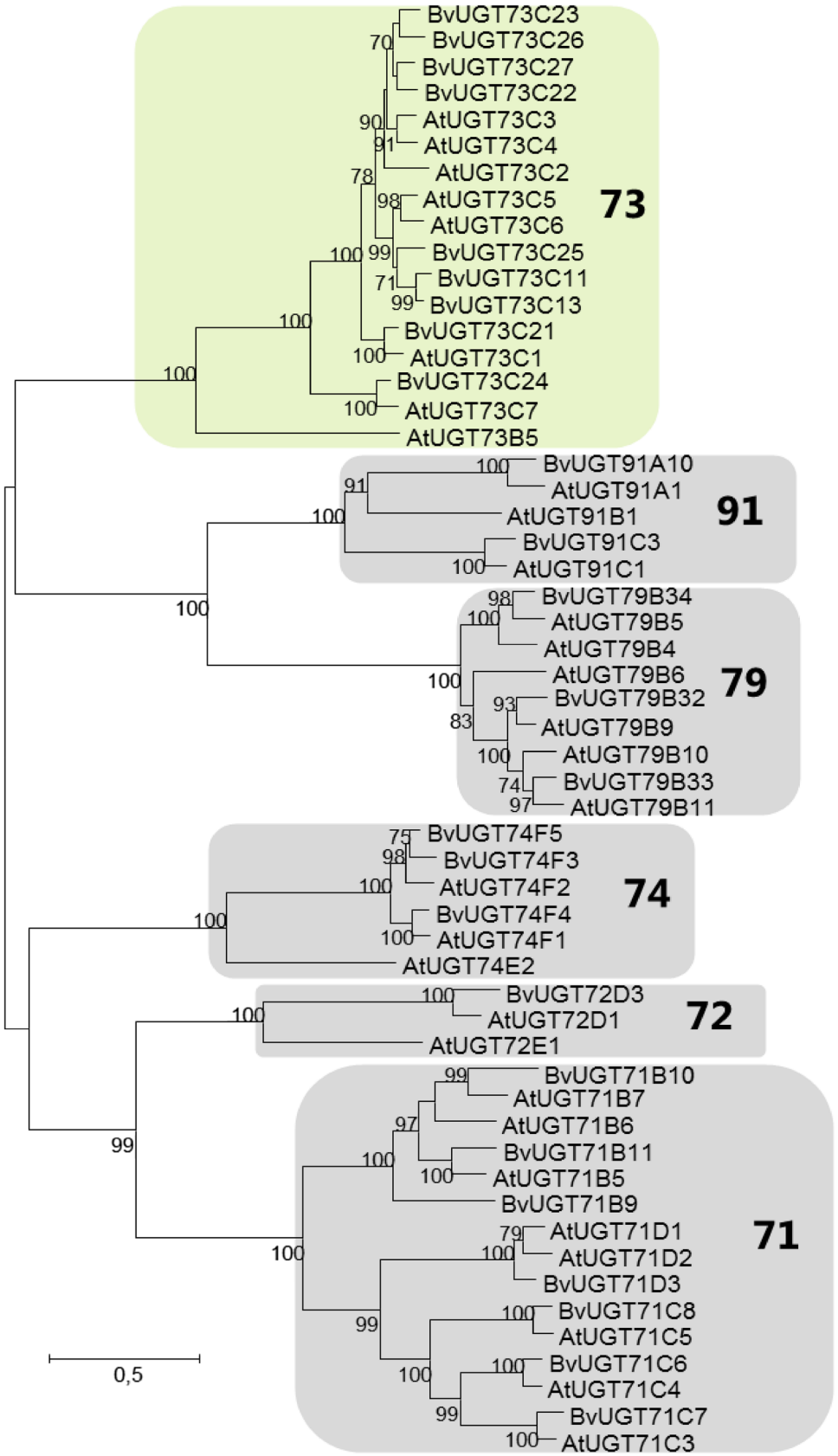


Fig. 2 Maximum Likelihood phylogenetic tree of the UGT candidates from *B. vulgaris* and homologous UGTs in *A. thaliana*. *B. vulgaris* sequences are marked with a Bv prefix, and *A. thaliana* sequences are marked with an At prefix. Each clade is framed with the UGT-family number indicated. The green box represents the UGT73-family from which some members were characterized by heterologous expression in *E. coli*. Only bootstrap values above 70% are shown in the tree. The bar represents 0.5 substitutions per site

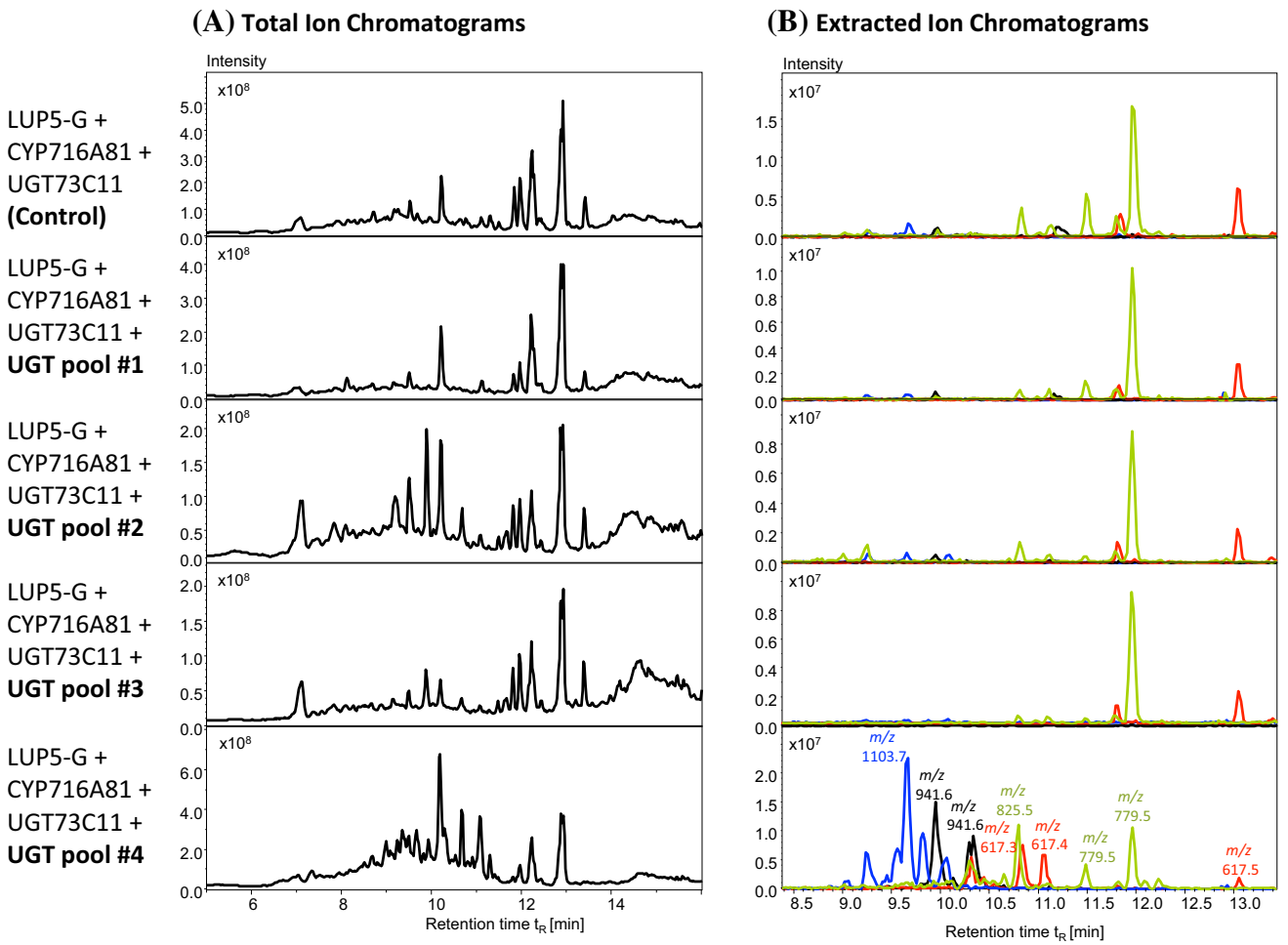


Fig. 3 LC-MS analysis of metabolites in *N. benthamiana* leaves after transient expression of UGT candidates divided into four pools. Genes for formation of oleanolic acid 3-glucoside (*BvLUP5-G, BvCYP716A81* and *BvUGT73C11*) were co-expressed in *N. benthamiana* alone or with each of four UGT pools, and leaf extracts were analysed by LC-MS. **a** Total ion chromatograms representing the mock control (upper lane) and the UGT pools (the lower four lanes). **b** Extracted ion chromatograms using *m/z* values (±0.5) to represent saponins with various numbers of hexose residues: oleanolic acid with one hexose residue, *m/z* 617.4 $[M–H]^-$ (red); with two hexose residues, *m/z* 779.5 $[M–H]^-$ and *m/z* 825.5 [M+formiate]$^-$ (green); with three hexose residues, *m/z* 941.5 $[M–H]^-$ (black); with four hexose residues, *m/z* 1103.6 $[M–H]^-$ (blue). Putative saponins peaks are indicated with *m/z* values of the observed $[M–H]^-$ and/or [M+formiate]$^-$ ions. Pool #1: BvUGT71B9, BvUGT71B10, BvUGT71C6. Pool #2: BvUGT71C7, BvUGT71C8, BvUGT71D3, BvUGT72D3, BvUGT91A10, BvUGT79B32. Pool #3: BvUGT79B33, BvUGT79B34, BvUGT91C3. Pool #4: BvUGT71B11, BvUGT73C21, BvUGT73C22, BvUGT73C23, BvUGT73C24, BvUGT73C25, BvUGT73C26

Mono- and bisdesmosidic glucosides formed from sapogenins and saponins by BvUGT73Cs

To characterize the product profiles of the newly identified *BvUGT73C*s, each were expressed in *E. coli* and crude protein extracts were incubated with UDP-D-glucose and 10 µM of the different sapogenin substrates. In this section, we identify the individual products, while product profiles of individual UGT enzymes are concluded in the next section. As UDP-D-glucose was used, all products were D-glucosides. Two products were identified as oleanolic acid 3-*O*-β-glucoside (t_R 26.6 min, *m/z* value 1237 $[2M–H]^-$, minor MS ion and major MS^2 fragment at *m/z* 618 $[M–H]^-$) and hederagenin 3-*O*-β-glucoside (t_R 22.8 min, *m/z* value 1268 $[2M–H]^-$ and 680 [M+formiate]$^-$, major MS^2 fragment of $[2M–H]^-$ at *m/z* 634 $[M–H]^-$) by comparison with authentic standards (Figs. 4, 5, S1 in Supplementary Material). A total of five additional products were observed, provisionally named a–e, all tentatively identified from substrate structures by *m/z* values, susceptibility to alkaline hydrolysis and fragmentation in MS^2 (Figs. 4, 5). While the 3-glucosides were resistant to alkaline hydrolysis (cleaving ester bonded glycosides but not ordinary *O*-glycosides), four of the five unknowns were susceptible to alkaline hydrolysis while product d was at so low level that the results were inconclusive (Fig. S2 in Supplementary Material).

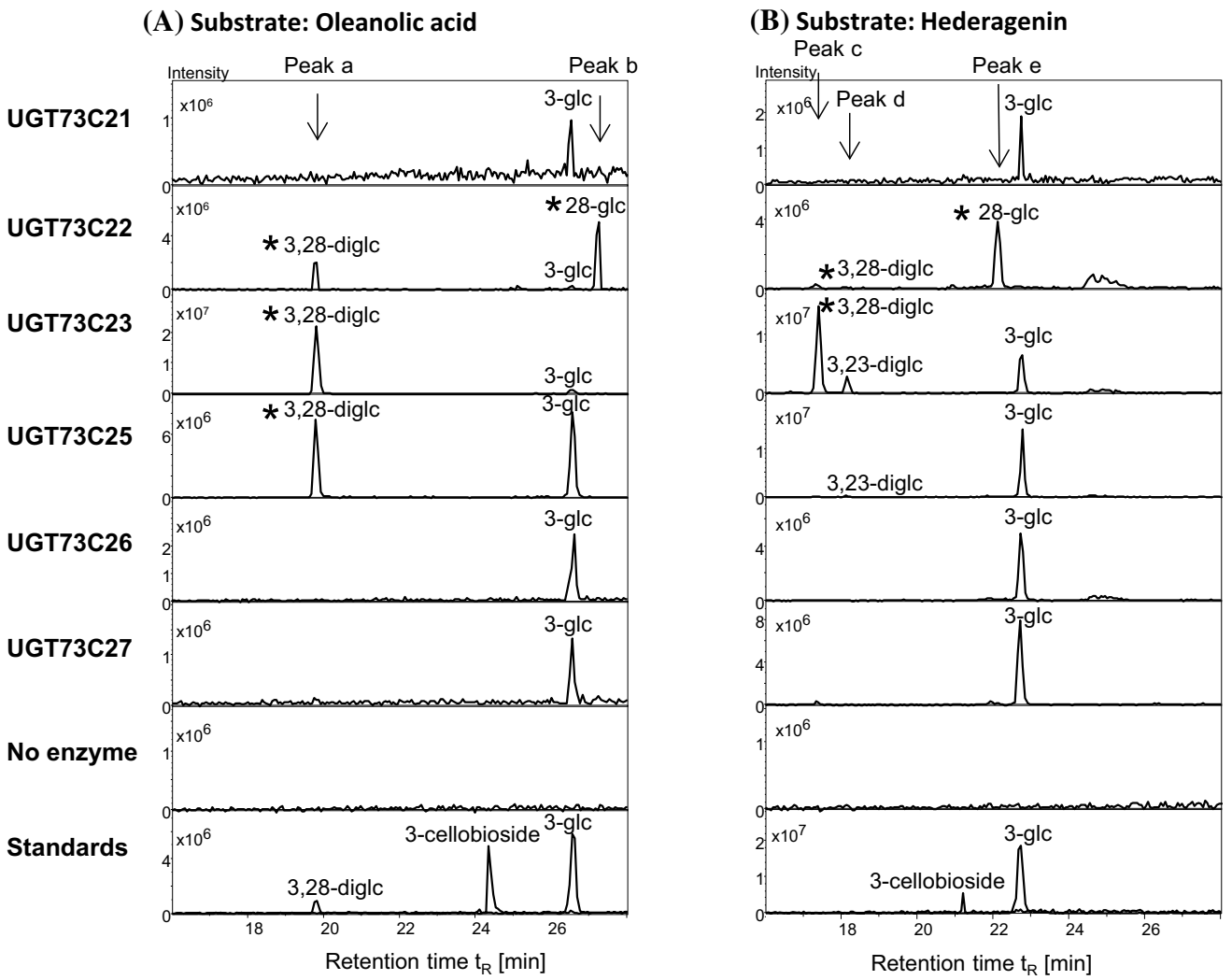


Fig. 4 LC-MS analysis of products of *B. vulgaris* UGTs expressed in *E.coli* and assayed with oleanolic acid or hederagenin. The extracted ion chromatograms are based on *m/z*-values ± 0.5 of 634 $[M_D$-anhydroglc-H$]^-$; 680 $[M_M$+formiate$]^-$; 842 $[M_D$+formiate$]^-$; 1268 $[2M_D-H]^-$ for determination of mono- (M_M) and diglucosides (M_D) of hederagenin and 618 $[M_D-$anhydroglc$-H]^-$; 664 $[M_M$+formiate$]^-$; 780 $[M_D-H]^-$; 826 $[M_D$+formiate$]^-$; 1237 $[2M_D-H]^-$ for determination of mono- (M_M) and diglucosides (M_D) of oleanolic acid. Visual inspection of raw data did not reveal other putative saponin products. Peaks marked with asterisks were susceptible to alkaline hydrolysis suggesting ester-like linkage. **a** Chromatograms with oleanolic acid as substrate. **b** Chromatograms with hederagenin as substrate

From incubation with oleanolic acid or oleanolic acid 3-*O*-β-glucoside as substrates, oleanolic acid 3-*O*-β-glucoside and two additional products (a, b) were observed (Figs. 4, 5). As the same UGT enzyme (BvUGT73C22) formed both oleanolic acid 3-*O*-β-glucoside and products a and b, we suggest from a shared reaction mechanism argument that the glucosides of products a and b are also β-linked. Product b (t_R 27.3 min) corresponds to a monoglucoside of oleanolic acid according to its *m/z* value (664, $[M+$formiate$]^-$) and MS^2 fragments of 618 ($[M-H]^-$) and 455 ($[M$-anhydroglucose-H$]^-$) (Fig. S1 in Supplementary Material). As the product was not the 3-linked monoglucoside available as standard, it had to be a glucosyl ester at position 28, a position well-known to be glucosylated in *B. vulgaris* saponins (Augustin et al. 2012; Khakimov et al. 2016). Susceptibility to alkaline hydrolysis agreed with this deduction. Accordingly, we suggest that product b is oleanolic acid 28-*O*-β-glucoside. Product a (t_R 19.8 min) corresponds to a bisdesmosidic diglucoside according to its *m/z* value (826, $[M+$formiate$]^-$), lack of loss of a diglucoside moiety in MS^2 (in contrast to the case for the authentic oleanolic acid cellobioside standard), and a major MS^2 fragment corresponding to loss of anhydroglucose (*m/z* value of 618 $[M-$anhydroGlc$-H]^-$) (Fig. S1 in Supplementary Material). We propose that product a is oleanolic acid 3,28-*O*-β-diglucoside, as positions 3 and 28 are the only positions at

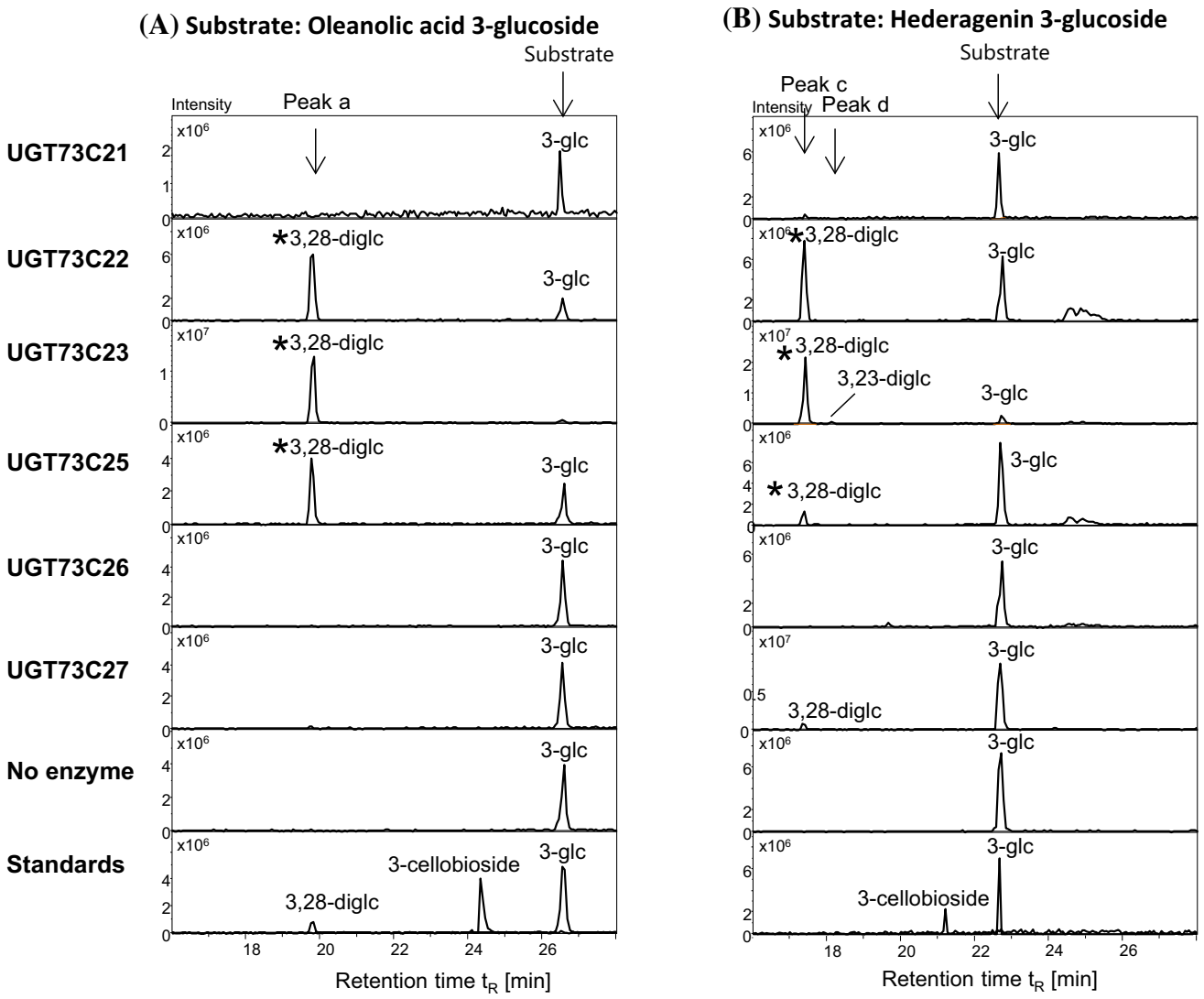


Fig. 5 LC-MS analysis of products of *B. vulgaris* UGTs expressed in *E.coli* and assayed with oleanolic acid 3-*O*-β-glucoside or hederagenin 3-*O*-β-glucoside. Visual inspection of raw data did not reveal other putative saponin products. Peaks marked with asterisks were susceptible to alkaline hydrolysis suggesting ester-like linkage. The extracted ion chromatograms are based on *m/z*-values ± 0.5 of 634 $[M_D-anhydroglc-H]^-$; 680 $[M_M+formiate]^-$; 842 $[M_D+formiate]^-$; 1268 $[2M_D-H]^-$ for determination of mono- (M_M) and diglucosides (M_D) of hederagenin and 618 $[M_D-anhydroglc-H]^-$; 664 $[M_M+formiate]^-$; 826 $[M_D+formiate]^-$; 1237 $[2M_D-H]^-$ for determination of mono- (M_M) and diglucosides (M_D) of oleanolic acid. **a** Oleanolic acid 3-*O*-β-D-glucoside as substrate (indicated with arrow). **b** Hederagenin 3-*O*-β-D-glucoside as substrate (indicated with arrow)

which *O*-glycosylation can occur and as product b was also formed from the 3-glucoside as substrate (hence, product b cannot be a diglucoside at position 28). The 3,28-diglucoside is also a logical product from enzymes forming the deduced 3- and 28-monoglucosides. Susceptibility to alkaline hydrolysis agreed with this deduction. Finally product b co-eluted with an in-house standard (Augustin et al. 2012).

From incubation with hederagenin and hederagenin 3-*O*-β-glucoside as substrates, hederagenin 3-*O*-β-glucoside and three additional products were observed (c, d, e) (Figs. 4, 5). In this case, three potential sites for glucosylation were present in the substrate. According to *m/z* values, product e is a monoglucoside while products c and d are diglucosides. From a shared UGT reaction mechanism argument, we again assume that all glucosides produced are β-linked. Product e (t_R 22.1 min) corresponds to a monoglucoside according to its *m/z* value (680, $[M+formiate]^-$) and a major fragment corresponding to loss of anhydroglucose (*m/z* value of 472, $[M-anhydroGlc-H]^-$) (Fig. S1 in Supplementary Material). Theoretically, for the e product, it can either be glucosylated at position 23 or 28, but as BvUGT73C22 mainly forms the 28-glucoside (product b) from oleanolic acid and also formed product e from hederagenin, a shared substrate recognition argument would suggest product e to be the ester

at position 28. Susceptibility to alkaline hydrolysis agreed with this deduction. Indeed, MS^2 spectra of products b and e were similar except for the mass difference from the extra oxygen of hederagenin. Accordingly, we conclude product e to be hederagenin 28-*O*-β-glucoside. The products with t_R 17.4 min (c) and 18.2 min (d) both correspond to bidesmosidic diglucosides according to their *m*/*z* values (842, $[M+formiate]^-$ in both cases) (Fig. S1 in Supplementary Material). Additionally, both products also have a major fragment corresponding to loss of anhydroglucose (*m*/*z* value of 634 $[M-anhydroGlc-H]^-$) (Fig. S1 in Supplementary Material). Product c was no longer detectable after alkaline hydrolysis. This suggests one of the glucose moieties is at position 28. As product c was formed when hederagenin 3-*O*-β-glucoside was used as substrate for BvUGT73C22, we conclude product c to be hederagenin 3,28-*O*-β-diglucoside. Finally, product d also corresponds to a bisdesmosidic diglucosylation. Like product c, product d is formed from both hederagenin and hederagenin 3-*O*-β-glucoside, so the glucoside at position 3 is definitely β-linked. We suggest the second position for glucosylation at product d is at position 23 because the 3,28-*O*-β-diglucoside is product c. In conclusion, product d is hederagenin 3,23-*O*-β-diglucoside.

Products with three or more glycosides were searched for in the chromatograms but not found. As all products are concluded to be *O*-D-glucosides in β-linkage, only the glucoside position is mentioned hereafter, e.g. as hederagenin 3-glucoside for the 3-*O*-β-D-glucoside.

BvUGT73 family members accept both hederagenin and oleanolic acid as substrates but with different product profiles

Extracts of *E. coli* expressing six of the newly discovered BvUGT73Cs individually, were characterized for their ability to accept sapogenins and their corresponding 3-glucosides as substrates. All six UGTs catalyzed glucosylation of both oleanolic acid and hederagenin (Fig. 4). Two of them (BvUGT73C21 and BvUGT73C26) exclusively glucosylated at position 3, forming oleanolic acid 3-glucoside and hederagenin 3-glucoside, and thus appear to be functionally equivalent to BvUGT73C11, the UGT previously reported to catalyze this specific reaction (Augustin et al. 2012).

The remaining four UGTs (BvUGT73C22, BvUGT73C23, BvUGT73C25, BvUGT73C27) also catalyze additional glucosylation at position 28 to various extents, both with sapogenins and their corresponding 3-glucosides as substrates. BvUGT73C27 mainly formed 3-glucosides of the tested sapogenins. Hence it represents an intermediate product profile between the above mentioned that exclusively form 3-glucosides and three UGTs discussed below. However, when 3-glucosides of both oleanolic acid and hederagenin were used as substrates of BvUGT73C27, only low levels of the 3,28-diglucosides were formed. This reveals a modest ability to glucosylate at position 28. A second UGT (BvUGT73C22) accepts both sapogenins and both 3-glucosides as substrates but mainly glucosylates at position 28. BvUGT73C22 also produces small amounts of the 3-glucosides from both sapogenins, as well as the corresponding bisdesmosidic diglucosides observed when the 3-glucosides were used as substrates. Hence, BvUGT73C22mainly glucosylates both sapogenins in position 28 but also in position 3, forming bisdesmosidic diglucosides. A third UGT (BvUGT73C23) seemed to be more processive and/or less regioselective than the remaining characterized UGT73Cs, as bisdesmosidic diglucoside products were mainly observed from sapogenins. Furthermore, BvUGT73C23 catalyzed formation of moderate amounts of the 3,23-diglucoside when hederagenin was used as substrate, indicating reduced regioselectivity. Product peaks were moderately more intense with this BvUGT73C23, but relative magnitudes of kinetic constants cannot be concluded from this observation, as expression levels of *UGT* genes are likely to vary in the experiment. A fourth UGT (BvUGT73C25) also catalyzes glucosylation in both positions 3 and 28. However, the ratio of the corresponding products was different according to the substrate; when oleanolic acid was used as substrate similar amounts of 3- and 28-glucosylation were observed, however, when hederagenin was used as substrate the 3-glucoside product dominated. Hence, of the BvUGT73Cs analyzed, BvUGT73C25 is the only one to exhibit different reactivity with oleanolic acid and hederagenin. Apparently, the presence of a hydroxyl group at position 23 in hederagenin stimulates glucosylation at the nearby position 3 or disturbs glucosylation at the distant position 28 by BvUGT73C25. However, the characterization of this enzyme (BvUGT73C25) was unintentionally carried out with a sequence elongated form. This was due to a deletion and frame shift near the end of the cloned gene, resulting in a 29 amino acid extension at the C-terminal as detailed in materials and methods. Hence, it cannot be rules out that this extension might have influenced the folding of the enzyme and the product profile.

The currently characterized saponins from *B. vulgaris* all have a cellobiosidic sugar moiety attached to position 3 of the sapogenin. No product peak matched the sapogenin cellobioside authentic standards (Figs. 4, 5). Hence, the six BvUGT73Cs do not, at our conditions, convert the 3-glucosides of hederagenin and oleanolic acid into their corresponding cellobiosides, although this route would be a straight forward pathway for their formation (Fig. 1a).

UGT73Cs glucosylates 24-epibrassinolide at a high substrate concentration

AtUGT73C5 and AtUGT73C6 from *Arabidopsis thaliana* cluster with BvUGT73Cs in phylogenetic trees (Fig. 2, Table S5 in Supplementary material), and have been shown

to glucosylate the steroidal plant hormone 24-epibrassinolide (Husar et al. 2011; Poppenberger et al. 2005). As the dedicated BvUGT73C11 forming sapogenin 3-glucosides has retained a weak side-activity for 24-epibrassinolide, we tested the additional BvUGT73C family members at the same high 24-epibrassinolide concentration (100 µM) as done by previous authors (Augustin et al. 2012). This high substrate concentration was used in order to reveal even low affinity reactivity, and was ten times higher than the sapogenins substrate concentrations used above. Although product levels seemed low, all six BvUGT73Cs were able to glucosylate 24-epibrassinolide into four different products (See Fig. S3 in Supplementary Material for more information). All four products have an apparent molecular ion with an *m/z* value of 688 $[M+formiate]^-$ corresponding to formiate adducts of 24-epibrassinolide monoglucosides. The major fragment (MS^2) of this ion has an *m/z* value of 642 $[M-H]^-$ as expected for the formiate adduct parent ion (Fig. S3 in Supplementary Material). Hence, all products seem to be monoglucosides. Diglucosides were searched for but could not be detected. As the UGTs glucosylate sapogenins in β-linkage, the products would represent β-glucosylation at each of the four hydroxyl groups of 24-epibrassinolide. Accordingly, the six BvUGT73Cs accepts 24-epibrassinolide as substrate at a high concentration, with ability to individually glucosylate each of the four available sites of the molecule. However, the relative amounts of the four monoglucosidic products differed significantly between the UGTs (Fig S3 in Supplementary Material). Chromatograms from incubations of *E. coli* extracts expressing three *BvUGT73Cs* (*BvUGT73C22, BvUGT73C23*, and *BvUGT73C25*) and assayed with 24-epibrassinolide were dominated by one product, with medium levels of two other products and traces of the fourth product. In contrast, three other *BvUGT73Cs* (*BvUGT73C21, BvUGT73C26*, and *BvUGT73C27*) yielded similar amounts of two products and lower amounts of two other, indicating less regioselectivity. Surprisingly, the latter were the most regiospecific towards oleanolic acid and hederagenin as substrates (Fig. 4).

BvUGT73C are organized in a tandem repeat

The four of the *UGTs* from subfamily 73 (*BvUGT73C21, BvUGT73C22, BvUGT73C23, BvUGT73C26*) are organized in a tandem repeat along with the previously identified *BvUGT73C11* on pseudomolecule 3 (Fig. 6c). The remaining two *BvUGT73s* (*BvUGT73C25* and *BvUGT73C27*) seemed initially to be located elsewhere in the genome (pseudomolecule 7 and Contig 7281, respectively, with the previously identified *BvUGT73C13* also on contig 7281). However, based on synteny with *A. thaliana* and the fact there is an unresolved region covering app. 13,000 nucleotides between *BvUGT73C11* and *BvUGT73C23*, it is possible that *BvUGT73C25, BvUGT73C27* and *BvUGT73C13* also are localized on pseudomolecule 3, despite the different or unresolved localization in the published genome. The order of the UGTs 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. 6a). This was also the case for a BvUGT pseudogene in the tandem repeat that was highly similar to AtUGT73C2. The pseudogene contained a nucleotide deletion causing a frameshift mutation resulting in a premature stop codon. Accordingly, the pseudogene encodes a 216 amino acid protein showing 91.3% sequence similarity to AtUGT73C2 and was therefore not included in the molecular phylogeny. The uncertainties regarding localization are due to high number of *UGT73C* paralogs combined with a high degree of sequence identity between UGTs as well as other genes from the saponin biosynthesis. This is hampering assembly and resulting in the genomic regions covering these genes being undissolved leading to many saponin genes being located on unanchored contigs (Byrne et al. 2017).

Residues on the surface of the BvUGT73Cs are under positive selection

To analyze for selection pressure on the BvUGT73C sequences, the ratio (ω) of non-synonymous (dN) to synonomous (dS) substitutions were determined. The average ω for all tested models are between 0.29 and 0.40 (data not shown) suggesting purifying selection ($\omega < 1$). The models with best fit to the data are model 3 and model 8—both significant with p values < 0.05. Due to similar log-likelihoods none of the two models are significantly better than the other. Model 3 has three site-classes and model 8 has 12 site-classes, however, in both models only one site-class has a $\omega > 1$ (Table 1). The two ω-values > 1 have p values of 0.036 and 0.071, hence 3.6–7.1% of the codons in the BvUGT73Cs are under positive selection. From a Naïve Empirical Bayes analysis with a strict threshold above 95%, model 3 identified 11 possible sites under positive selection, and model 8 identified two possible sites under positive selection, which are consistent with model 3 (Table S2 in Supplementary Material).

In order to deduce topological positions of the amino acid residues under positive selection, a homology model of AtUGT73C1 was created using the known crystal structure of the 3-*O*-flavonoid glucosyltransferase MtUGT78G1 (3hbf) from *Medicago sativa* as template. This template show high sequence identity to AtUGT73C1 and is co-crystalized with a substrate and donor, hence it was regarded the optimal template. AtUGT73C1 was chosen as it has a defined substrate and the position numbers in table S2 in Supplementary Material corresponds to this. The model was visualized

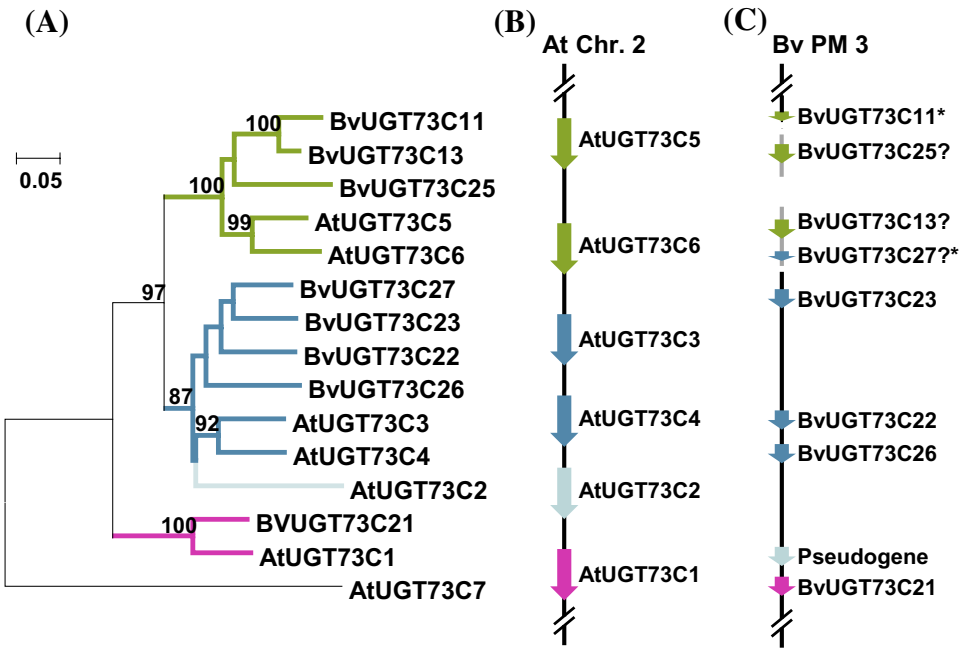


Fig. 6 Phylogenetic analysis and syntenic localization of UGT73Cs from *B. vulgaris* and *A. thaliana*. *B. vulgaris* sequences and genes are marked with a Bv prefix, and *A. thaliana* sequences and genes are marked with a At prefix. Phylogenetic clades are marked with colors (green, blue and magenta). **a** Phylogenetic analysis. Only bootstrap values above 70% are shown in the tree. The bar represents 0.05 substitutions per site. **b** Selection of *A. thaliana* chromosome 2, which covers six UGTs from the UGT73C subfamily. **c** Selection of *B. vulgaris* pseudomolecule 3, which covers at least five UGTs and a pseudogene from the UGT73C subfamily. The broken line denotes an undissolved region of pseudomolecule 3. Genes (*BvUGT73C25, BVUGT73C13, BvUGT73C27*) suggested to be anchored to this undissolved region on pseudomolecule 3 are denoted by an question mark (?) while asterisks (*) denote non full-length genes

Table 1 Nonsynonymous/synonymous substitution rates (ω) for the UGT73Cs sequences

Best fit models	Average ω	Classes of ω with corresponding shares of sites fitted to each ω
Model 3: (discrete)	0.3383	p: 0.49253 0.43660 0.07087 ω: 0.05152 0.46514 1.54976
Model 8: (β and ω)	0.3413	p: 0.09639 0.09639 0.09639 0.09639 0.09639 0.09639 0.09639 0.09639 0.09639 0.09639 0.03609 ω: 0.00213 0.01813 0.04927 0.09572 0.15822 0.23808 0.33757 0.46063 0.61516 0.82369 1.98185

p; share of sites fitted to each ω

in Pymol and the corresponding sites with signs of positive selection in UGT73Cs were identified (cyan) (Fig. 7). Codons 1–6 and 495–499 from the substitutions calculations were not included in the homology model, however no sites suggested to be under positive selection are situated in that range. The codons suggested to be under positive selection by model 2 and model 8 all localize on the solvent-facing surface of the enzyme, near the substrate and UDP-glucose entrance (Fig. 7).

The branches in the phylogenetic tree (Fig. 6a) were also tested for positive selection, however no significant evidence for positive selection on branches were found (data not shown).

Discussion

Previous research has elucidated core biosynthetic steps in the formation of sapogenin 3-glucosides that are considered intermediates of the main cellobiosidic saponins known to accumulate in *B. vulgaris* (Augustin et al. 2012; Khakimov et al. 2015). As the cellobioside moiety of the saponin end-products consists of two glucosides connected by a (1 → 4) β-glycosidic bond, the envisioned missing enzyme was expected to be a UGT (Thimmappa et al. 2014). Using several strategies aimed at identifying UGTs converting sapogenin 3-glucosides to sapogenin

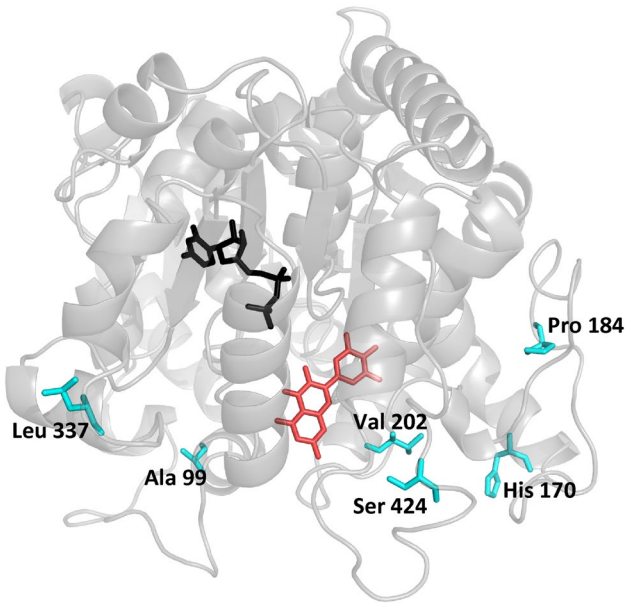


Fig. 7 Model of AtUGT73C1 with the flavonoid 3-*O*-glucosyltransferase UGT78G1 (3hbf.1.A) as template complexed with UDP and myricetin. The structure of the model is shown in grey, the donor UDP-glucose is represented by UDP shown in black and the acceptor myricetin (corresponding to the sapogenin substrates of UGT73Cs) is shown in red. Amino acid residues corresponding to sites with signs of positive selection in *A. thaliana* and *B. vulgaris* UGT73Cs are shown in cyan, and marked with the three-letter code for the amino acid and number in the alignment used for positive selection

3-cellobiosides, we screened 20 *B. vulgaris* UGTs and found six of them to catalyze glucosylation of the sapogenin and saponin 3-glucoside precursors of the major pest-deterrent saponins. These BvUGTs belonged to the UGT73C subfamily; a UGT subfamily that had previously been shown to be involved in the first glucosylation steps in saponin biosynthesis in *B. vulgaris*. Four of the six *BvUGT73C*s were organized in a tandem repeat, possibly including the remaining two *BvUGT73C*s (Fig. 6). However, our screening strategy did not identify an enzyme catalyzing the expected last step of the biosynthesis, from the 3-glucosides of oleanolic acid or hederagenin to the corresponding 3-cellobiosides. In contrast, the identified BvUGT73Cs formed monoglucosides or bisdesmosidic saponins.

At least three questions arise from our results: (I) What is the functional and evolutionary background of gene organization and functional diversity in the BvUGT73C subfamily? (II) How are multiple specific glycosylation patterns in specialized *B. vulgaris* metabolites controlled with an available set of UGTs with low catalytic selectivity? (III) What is the reason for the repeatedly observed tendency to form bisdesmosidic saponins in this species, in contrast to the known monodesmosidic nature of the major pest-deterrent saponins?

Genomic organization and evolution

Both *A. thaliana* and *B. vulgaris* have UGT73s organized as tandem repeats in their genomes. There are seven UGT73Cs in the *A. thaliana* genome, of which six are localized on in a tandem array on chromosome 2. Similarly, until now nine full-length *UGT73C* genes and one *UGT73C* pseudogene have been identified in *B. vulgaris*. A phylogenetic analysis indicates duplications of *UGT73Cs* in both lineages after the divergence of *A. thaliana* and *B. vulgaris* (Fig. 2). Duplications of genes in plant specialized metabolism drives chemical diversity; subsequent changes of the substrate and product specificity of newly duplicated genes may result in new metabolites (Hamberger and Bak 2013; Ober 2010). *BvUGT73C11* is an example of a duplicated gene that evolved from a more promiscuous UGT73C5-like ancestor into an efficient and specific C3 sapogenin glucosyl transferase, that most likely facilitate that *B. vulgaris*, as the only know crucifer, has evolved to produce saponins from a sapogenin predisposition (Augustin et al. 2012).

The *AtUGT73C1-6*s are organized in a tandem repeat on chromosome 2 and *A. thaliana* and *B. vulgaris* have been found to show high synteny (Kuzina et al. 2011). Accordingly, we suggest a number of *BvUGT73Cs* to be organized in a tandem array on pseudomolecule 3. Despite huge sequencing efforts (Byrne et al. 2017), one-third of the genome region covering the *BvUGT73Cs* is still undissolved (indicated by broken lines in Fig. 6c). A major problem is high sequence identity within the *BvUGT73C*s, and current sequencing technologies and assemble algorithms do not facilitate assembly of such highly repetitive regions (Byrne et al. 2017). Consequently, this region may include genes that we have not been able to identify and analyze, which might have limited our screening strategy. To resolve the unknown region in pseudomolecule 3, and other unknown regions, additional sequencing of very long genomic fragments using other next generation sequencing technologies than those already used (Byrne et al. 2017) should be considered. Although it can be concluded that the diversity and expansion of the UGT73C family in *B. vulgaris* has evolved as a tandem repeat, in case of the four genes mapped with certainty (Fig. 6c), better sequence data and methods would be desirable for more detailed studies of possible evolutionary events in the genus.

Interestingly, the amino acid residues corresponding to the codons in the *UGT73C* subfamily that showed signs of positive selection, were modeled to be on the solvent-exposed surface of the enzyme, rather than in the interior. We tentatively suggest some of these positions to be involved in substrate entry to the catalytic center and/or be involved in interaction with other enzymes in the saponin pathway. Such interactions could result in protein–protein interactions that

serve to facilitate efficient channeling of specific substrates to the active site of the UGT.

Difference in product profiles in vitro reflects the structural diversity in planta

From the in vitro assay product profiles of the six BvUGT73Cs and those previously reported (Augustin et al. 2012), we concluded that the product profiles of the individually expressed enzymes were diverse but partly overlapping (Figs. 4, 5). The enzymology of sapogenin glycosylation is still in its infancy, and information in the literature of UGT selectivity is sparse. Three UGTs, BvUGT73C21, BvUGT73C26 and BvUGT73C27 glucosylate at position 3 only, like the previously identified BvUGT73C11 (Augustin et al. 2012). Glycosylation at position 3 is common for saponins from many other plant species (Vincken et al. 2007). To our current knowledge, besides the BvUGT73C members discussed here and elsewhere (Augustin et al. 2012), only UGT74AE2 from *Panax ginseng* is known to catalyze glucosylation at position 3 of a sapogenin (protopanaxadiol) (Jung et al. 2014). Three UGTs, BvUGT73C22, BvUGT73C23 and BvUGT73C25 were able to glycosylate at position 28, like the previously identified BvUGT73C13 (Augustin et al. 2012), although they have a very different product profile. Enzymes with the ability to glucosylate triterpenoid backbones in position 28 are also known in other plants: UGT73F3 from *Medicago truncatula* gluosylates hederagenin at position 28, and UGT74M1 from *Saponaria vaccaria* gluosylates gypsogenic acid at position 28 (Meesapyodsuk et al. 2007; Naoumkina et al. 2010). These two UGTs belong to two different UGT families, so their similar selectivity was proposed to be a result of convergent evolution (Seki et al. 2015). Two other UGTs, BvUGT73C23 and BvUGT73C25, are apparently able to glucosylate hederagenin at position 23, but only as a minor reaction compared to glucosylation at positions 3 and 28. To our knowledge, no UGT glycosylating preferentially at position 23 has been reported previously. The ability of BvUGT73C22, BvUGT73C23 and BvUGT73C25 to form bisdesmosidic saponins may be crucial for the immense saponin diversity in the species.

The apparent high product specificity of some of the BvUGTs, specifically BvUGT73C21 and BvUGT73C26, was challenged when incubating with the more structurally distant substrate 24-epibrassinolide (at 100 µM) since this led to the production of minor peaks of apparent glucoside products. In addition, trichlorophenol was included as a generic UGT substrate (Augustin et al. 2012), and found to be glucosylated by the BvUGT73Cs, albeit at high substrate concentration (500 µM). However, even the very product specific BvUGT73C11 catalyzing the first glycosylation step in *B. vulgaris* saponin biosynthesis showed these side reactions, possibly reflecting an ancient role of the UGT73C subfamily in *A. thaliana* 24-epibrassinolide metabolism (Chung and Choe 2013). Future research should consider whether some of the BvUGT73Cs have retained a role in 24-epibrassinolide turnover. However, as all possible substrates cannot be tested and compared in vitro, gene knockout or similar experiments or specific gene expression studies would be needed to conclude in vivo product profiles of the UGTs in *B. vulgaris*. Gene expression studies have been reported for *B. vulgaris* (Wei et al. 2013), but due to the high sequence identity between the BvUGT73Cs, distinction of the paralogous *BvUGT73Cs* transcripts is far from simple.

This and previous reports (Augustin et al. 2012; Liu et al. 2016) show that UGTs are involved in forming the diverse spectrum of many kinds of specialized metabolites in *B. vulgaris*, and a series of metabolite structure elucidation studies have shown that bond types in oligosaccharide moieties of flavonoids, glucosinolates and saponins differ (Agerbirk et al. 2015; Dalby-Brown et al. 2011; Khakimov et al. 2016; Nielsen et al. 2010b). In flavonoids, oligosaccharide moieties were elucidated and found to include β-glucosylation at position 2 and 6 of the innermost glucoside (Dalby-Brown et al. 2011). These positions differ from position 4-glucosylation in the cellobiose moiety of saponins, meaning that *B. vulgaris* has the capacity to form at least three types of connections between glucoses in specialized metabolite oligoglycoside moieties. In glucosinolates, however, the single thioglucose is not further glycosylated but occasionally acylated (Agerbirk and Olsen 2011), meaning that highly specific biosynthetic machinery governs position, presence or absence of higher order glycosylation in *B. vulgaris* specialized metabolites. Stringent regulation must take place in the known seasonal fluctuations and complex inductions of glucosinolates (Agerbirk et al. 2001; van Mölken et al. 2014), flavonoids (Dalby-Brown et al. 2011) and saponins (Agerbirk et al. 2001, 2003a; van Mölken et al. 2014) with contrasting glycoside moieties.

Biochemical functions of bisdesmosidic saponins

We did not find any UGT that catalyzed the formation of sapogenin cellobiosides, which are the dominant saponins in *B. vulgaris* and involved in insect resistance *in planta*. Our screening was based on the hypothesis that the corresponding monoglucosides were substrates of the cellobioside-forming enzyme. Although we did not characterize all potentially responsible UGTs, this study tentatively suggests that cellobiosides in *B. vulgaris* are not formed directly from 3-monoglucosides in a single step catalyzed by a UGT. Hence, the hypothesized biosynthetic pathway may need to be reevaluated. It is possible that a bisdesmosidic saponin, or a saponin exclusively glycosylated at position 28, or another conjugate is an intermediate in the biosynthesis of the cellobiosides. Thus, the UGTs

catalyzing glucosylation in position 28 could also play a role in the biosynthesis of the cellobiosides, assuming an indirect pathway (Fig. 1a).

An alternative hypothesis is biochemical transfer of the entire cellobiosyl group to the sapogenin. Such assembly of a complex substituent before conjugation has experimental support in case of a complex acyl group added to an oat-saponin (Owatworakit et al. 2013). A mechanism of this kind could be referred to as "cellobiosylation". However, we are not aware of a similar documented mechanism for transfer of complex oligosaccharides to sapogenins. Analysis of gene expression during herbivore attack (Wei et al. 2013; Zhang et al. 2015), further analysis of UGT product profiles, and seasonal transitions may shed light on these questions in the future.

Levels of insect resistance and cellobioside saponins concentrations in *B. vulgaris* are known to be seasonally regulated, with a decline during the fall (Agerbirk et al. 2003a). However, in that study higher order glycosylated forms (Khakimov et al. 2016) were not analyzed since their existence was not known at the time of the study. The function of this decline could be to avoid plant membrane damage in the cold season, where the rosette leaves are retained and metabolically active, as bisdesmosidic saponins may function as inactive or less active storage forms of saponins (Moses et al. 2014; Osbourn 1996). One possible function of the UGTs catalyzing glucosylation in position 28 could be to form inactive conjugates that could still have colligative effects reducing e.g. frost damage, and could possibly be turned over to pest-deterrent monodesmosidic saponins in the following spring.

If bisdesmosidic saponins are indeed biosynthetic intermediates of the cellobiosides, these saponins should be included as substrates in future enzyme assays aimed at reconstituting this biosynthesis. In the present study, we used the previously characterized BvUGT73C11s to produce the 3-monoglucoside substrates. The UGTs reported here are well suited for preparing sapogenin 28-glucosides and sapogenin 3,28-diglucosides for use as substrates. In this way, increasing the enzyme tool box of specific UGTs will enable us to prepare increasingly complex saponin substrates for elucidating the complex biosynthesis of saponins in pest resistant *B. vulgaris*. In addition to the complexities of defense-regulation, saponins may also serve regulatory roles in plants [reviewed by Moses et al. (2014)], so the complex glucosylation machinery and metabolite diversity could also play a role in such processes.

Conclusion

This study clearly demonstrates that *B. vulgaris* has a much richer arsenal of UGTs involved in saponin biosynthesis than initially anticipated. While the tendency to form bisdesmosidic saponins is in agreement with the recent study by Khakimov et al. (2016), the apparent lack of an UGT to convert monoglucosides to their corresponding cellobiosides challenges previous hypotheses on the biosynthesis of pest-deterrent sapogenin cellobiosides. The moderate substrate specificity and product selectivity of these enzymes agrees with the highly diverse mixture of saponins in the species, but does not explain the characteristic glycoside moieties of known *B. vulgaris* saponins responsible for insect resistance, and other classes of specialized metabolites. The possibility of UGTs taking part in biosynthetic multi-enzyme complexes (metabolons) is a tempting hypothesis to explain apparent substrate specificity and product selectivity in the biosynthesis, supported by indications of positive selection for amino acid residues on the UGT73C protein surfaces. Further work on elucidation of saponin biosynthesis in *B. vulgaris* should take into account additional biosynthetic routes than previously anticipated, including bisdesmosidic intermediates (Fig. 1a), a much higher number of UGTs involved than previously recognized in the pathway, and the possibility of controlling UGT product profiles by metabolon formation. Additional biosynthetic routes could include transfer of the cellobioside moiety in a single step, or formation of the cellobioside moiety as part of a bisdesmosidic saponin, followed by hydrolysis to a monodesmosidic sapogenin cellobioside (Fig. 1a).

Materials and methods

Reference saponins and sapogenins

Hederagenin 3-*O*-β-D-glucoside (hederagenin 3-glucoside) and oleanolic acid 3-*O*-β-D-glucoside (oleanolic acid 3-glucoside) preparations were the previously reported (Augustin et al. 2012). Oleanolic acid 3-*O*-β-cellobioside was of the originally reported preparation (Agerbirk et al. 2003a). Hederagenin 3-*O*-β-cellobioside was a kind gift from Dr. Francisco R. Badenes-Perez (Badenes-Perez et al. 2010, 2014) and the identity and purity was controlled by ^{1}H-NMR analysis (data not shown). Proposed oleanolic acid 3,28-O-β-D-diglucoside, formed from UDP-D-glucose and oleanolic acid using UGT73C13 (Augustin et al. 2012) as catalyst, was a kind gift of Dr. Qing Liu. Hederagenin and oleanolic acid were those used and confirmed by NMR in (Augustin et al. 2012), and 24-epibrassinolide was from Sigma (E1641).

Identification of gene candidates

A selection of UGT sequences involved in multiple glycosylation of metabolites were used as query in a tBLASTn against a *B. vulgaris* transcriptomic database (Kuzina et al. 2011) and a *B. vulgaris* genomic database [early version

of Byrne et al. (2017)] using the CLC main workbench and standard settings: *UGT71A1, UGT71A15, UGT71A16, UGT73E1, UGT74A1, UGT74B1, UGT74C1, UGT74D1, UGT74E1, UGT74F1, UGT74G1, UGT74H1, UGT71H5, UGT74J1, UGT74M1, UGT74N1, UGT74R1, UGT79A1, UGT79A2, UGT79B1, UGT79B2. UGT79B5, UGT79B15, UGT79B16, UGT79C1, UGT91A1, UGT91A2, UGT91B, UGT91C1, UGT91D1, UGT91E1, UGT91F1, UGT91F2, UGT91G1, UGT91G7, UGT91H4, UGT91H6, UGT82C2, UGT710E5, UGT705A4, UGT709A10, UGT99C4, UGT74H7, UGT98B4, UGT85B2, UGT74H5, UGT75E3, UGT88C, UGT85F13, UGT93B9, UGT703A5, UGT93B8, UGT75E2, UGT74H6, UGT710F3, UGT73C1, UGT73C2, UGT73C3, UGT73C4, UGT73C5, UGT73C6, UGT73C7, UGT73C9, UGT73C10, UGT73C11, UGT73C12, UGT73C13*.

Cloning of UGTs and transient expression in *Nicotiana benthamiana*

Primers 1–46 (Table S3 in Supplementary Material) were designed to amplify the identified full length candidate genes via PCR (98 °C 30 s, 35 cycles of 98 °C 10 s, 55 °C 30 s, 72 °C 90 s, 72 °C 10 min) with the X7 polymerase (Nørholm 2010). The template DNA was from *B. vulgaris* G-type genomic DNA extracted as in (Byrne et al. 2017). Except for *BvUGT74F3, BvUGT74F4* and *BvUGT74F5* all the *UGT*s were intron less. The cloned genes were assigned GeneIDs and a UGT name according to the UGT nomenclature (Mackenzie et al. 2005) (http://prime.vetmed.wsu.edu/resources/udp-glucuronsyltransferase-homepage) (Tables S1, S4 in Supplementary Material). Unintentionally, elongated versions of *BvUGT73C25* and *BvUGT91A10*, respectively, were used in this study. *BvUGT91A10* had a premature start codon 10 codons upstream the true start codon, and *BvUGT73C25* had a deletion of nucleotide no 1477 which caused a change of reading frame in the last four amino acids as well as a C- terminal sequence elongation of additional 29 amino acid residues (Tabel S4 in Supplementary Material). When *BvUGT73C25* and *BvUGT91A10* are mentioned it is the elongated versions that are analyzed. Constructs were cloned using Gateway cloning BP and LR kits from Invitrogen (BP clonase ® Enzymes Mix, and LR clonase ® Enzymes Mix) according to the manufacturers protocol. pDONR207 (Invitrogen) and pEAQ-HT-DEST1 (Sainsbury et al. 2009) were used as donor and destination vectors, respectively. UGT expression constructs were transformed into *Agrobacterium tumefaciens* strain AGL1 by electroporation, and positive clones were used for transient expression in *N. benthamiana*.

The *UGT* candidates were divided into four pools and each pool was screened in combination with known genes from the saponin biosynthesis; *BvLUP5-G, BvCYP716A81* and *BvUGT73C11* (Khakimov et al. 2015). Transient expression of the *UGT* candidates in *N. benthamina* leaves was performed according to (Khakimov et al. 2015), and metabolites were extracted with 80% aq. MeOH.

Saponin and general metabolite analysis

Metabolite extracts in 80% aq. MeOH were analyzed for saponins and general polar metabolites by LC-MS/MS. The instrument used was an Agilent 1100 Series HPLC (Agilent Technologies), equipped with a Gemini NX column (35 °C; 2.0 × 150 mm, 3.5 mm; Phenomenex) and coupled to a Bruker HCT Ultra ion-trap mass spectrometer (Bruker Daltonics). Mobile phases were eluent A, water with 0.1% (v/v) formic acid and 10 µM NaCl, and eluent B, 80% acetonitrile with 0.1% (v/v) formic acid. A short gradient was used for screening of *N. benthamiana* extracts, while a longer gradient was used for identifying specific saponin products from UGT product profile assays. The short gradient program was as follows: 0–1 min, isocratic 17% B; 1–12.5 min, linear gradient 17–80% B; 12.5–12.6 min, linear gradient 80–98% B; 12.6–16 min, isocratic 98% B; 16–18 min, isocratic 17% B. The flow rate varied during the program as follows: 0–8 min, 2 ml/min; 8–12.5 min, 0.3 ml/min; 12.5–18 min, 0.45 ml/min.The longer gradient was the same as previously published (Augustin et al. 2012). The detector was operated in negative electrospray-ionization mode, and MS^2 spectra were acquired in automatic mode using the SmartFrag function with the instrument default settings. The obtained data were analyzed using Bruker Daltronics Dataanalysis 4.0, including total ion chromatograms to reveal polar metabolites in general and extracted ion chromatograms targeted at relevant saponin oligoglycoside *m/z* values (± 0.5 to allow for the inherent moderate accuracy of the ion trap detector).

Activity assays of UGTs using the TnT system

To analyze for the UGTs ability to glucosylate 50 µM hederagenin 3-monoglucoside the Transcription/Translation system (TnT) from Promega was used. Primers 55–88 (Table S3 in Supplementary Material) were designed to add TnT T7 sites to the UGTs via PCR (98 °C 30 s, 35 cycles of 98 °C 10 s, 62 °C 30 s, 72 °C 90 s, final extension 72 °C 10 min) with the X7 polymerase (Nørholm 2010). The DNS templates used were the plasmids from the *N. benthamiana* expressions plasmids. PCR products were purified using the Qiagen MinElute gel extraction kit. Transcription/translation of the UGTs were performed in total amounts of 10µl: 40 ng purified PCR product, Nuclease-Free water, 0,5 µCi fresh L-[35S] methionine (PerkinElmer), 55 µM Methionine and 8 µl TnT T7 PCR Quick Master Mix were mixed and incubated 90 min at 30 °C. The formed proteins were separated by SDS-PAGE and visualized by incubating the dried gels for 4 days on phosphorimager

screens, following scanning with a STORM 860 molecular imager (Molecular Dynamics). The assays were performed in total volumes of 50 µl:10 µl transcription/translation reaction (produced using only non-radiolabeled methionine), 2 µCi [Glucose-^{14}C (U)]-UDP glucose (PerkinElmer), 0.5 mM 2,4,6 trichlorophenol (TCP), 5 mM $MgCl_2$, 1 mM $CaCl_2$ and 25 mM TAPS-HCl buffer (pH 8.6). The reaction mix was incubated 30 °C for 1 h, and the reaction stopped by addition of 150 µl 100% MeOH. The reaction was centrifuged 10 min at 12,000×*g*, filtered by centrifugation for 2 min 3000×*g* using 96-well filter plates (polyvinylidene difluoride; 0,45 µm). The samples were concentrated using a speed vacuum concentrator 30 °C 1 h. The formed metabolites were resuspended in 15 µl 100% MeOH and separated by thin layer chromatography [silica gel 60 F254 plates (Merck), mobile phase: dichloromethane:MeOH:H_2O (90:9.5:0.5)] and visualized by incubating the dried plates for 3 days on phosphorimager screens, following scanning with a STORM 860 molecular imager (Molecular Dynamics).

Cloning of *UGTs*, expression in *E. coli*, and analysis of saponin product profile

The most promising UGTs from transient expression in *N. benthamiana* identified by LC-MS/MS analysis were re-amplified as above using primers listed in Tabel S3 in Supplementary Material, cloned into pDONR 207 and then into the destination vector pJAM1786 (Lou et al. 2007; Takos et al. 2011). The expression constructs were transformed into Xjb DE3 autolysis TM *E. coli* cells (Zymo Rescearch) according to Augustin et al. (2012) and used for *E. coli* expression.

Activity assays for identifying UGT products of sapogenins were performed in a total volume of 50 µl containing (final concentrations) 750 µM UDP-glucose, 10 or 100 µM substrate, 5 mM $MgCl_2$, 1 mM $CaCl_2$ 1 mM dithiothreitol (DTT) and 25 mM TAPS-HCl pH 8.6 as described elsewhere. The following substrates were used: 10 µM hederagenin, 10 µM hederagenin 3-*O*-β-D-glucoside, 10 µM oleanolic acid, 10 µM oleanolic acid 3-*O*-β-D-glucoside, and 100 µM 24-epibrassinolide (Sigma) all dissolved in DMSO. The assays were incubated for 1 h at 30 °C and stopped by the addition of 150 µl MeOH. Samples were filtered (polyvinylidene difluoride; 0.45 µm) and solvent evaporated by vacuum concentrator. Samples were dissolved in 50 µl 80% MeOH prior to analysis by LC-MS/MS (long gradient) as described above.

Testing for ester-linked glucosides by alkaline hydrolysis

Aqueous NaOH was used as a reagent for specific hydrolysis of ester but not general glycoside bonds. The protocol was modified from (Augustin et al. 2012): Products from UGTs expressed in *E. coli* and incubated with UGT substrates were dried under vacuum at 42 °C. 0.5 ml 2 M aq. NaOH was added, followed by brief vortex mixing (30 s), followed by agitation at room temperature for 1.5 h at 1400 rpm. Then, 100 µl of 12 M aq. HCl was added to ensure low pH keeping carboxylic acids uncharged. The resulting solution was extracted twice with a total of 3 ml of diethyl ether, in order to remove saponins from the high ionic strength aqueous phase. The diethyl ether fractions were washed with 1/2 volume of H_2O and vortexed for 30 s, followed by centrifugation at 4000×*g* for 5 min. The ether fraction was then transferred to a new tube, completely dried under nitrogen flow and re-solubilized in 50 µl of 85% aq. MeOH before analysis by LC-MS/MS as described above. Peaks still detected after alkaline hydrolysis were considered not to contain ester bonds (glucoside at position 28), while major peaks not observed after alkaline hydrolysis were considered to contain ester bonds. Due to less than full recovery, one minor peak (the proposed 3,23-diglucoside of hederagenin) could not be unequivocally assigned to linkage type in this way.

Sequence analysis and phylogenetic trees

Calculation of sequence identity were performed at the webserver Clustal omega (EMBL EBI) using standard settings (McWilliam et al. 2013) (Table S5 in Supplementary Material).

Prior to the phylogenetic analysis alignments of the sequences were performed using the MUSCLE alignment algorithm (Edgar 2004) in MEGA version 6.6 (Tamura et al. 2013) (Fig. S4, S5, Table S6, S7 all in Supplementary Material).

The phylogenetic analysis of aligned sequences used in Fig. 2 was based on the Maximum Likelihood method using on the JTT matrix-based model in MEGA version 6.6 (Tamura et al. 2013). The tree with the highest log likelihood (-22361.9316 for Fig. 2 and −5618.6034 for Fig. 6a) is shown. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining and BioNJ method to a matrix of pairwise distances estimated using a JTT model. A discrete Gamma distribution was used to model evolutionary rate differences among sites [5 categories (+G, parameter = 2.0744 for Fig. 2 and 0.8622 for Fig. 6a)]. The rate variation model allowed for Fig. 2 some sites to be evolutionarily invariable ([+*I*], 3.1685% sites). The tree was drawn to scale, with branch lengths measured in the number of substitutions per site. All positions with less than 95% site coverage were eliminated. There were a total of 387 positions for Fig. 2 and 485 positions for Fig. 6a in the final dataset.

Analysis for sites and branches under selection pressure

To analyze for selection pressure on the *UGT73C* sequences, the ratio (ω) of non-synonymous to synonymous substitutions (dN and dS, respectively) was calculated for a codon-based nucleotide alignment using the program Codeml from the PAML X user interface (Xu and Yang 2013; Yang 2000). In Codeml, model = 0 was selected (which uses a single ω among branches), and then different codon substitutions models (NSsite models M0–M8) (which account for heterogeneous ω ratios among codons) were tested (Yang and Bielawski 2000; Yang et al. 2000). The log likelihoods of the models were compared, and to test their significance log likelihood ratio tests were carried out. Where probable positive selection was found, a naïve empirical Bayes analysis identifying positively selected sites with a $p > 95\%$ and $p > 99\%$ was made using Codeml.

To test if any branches of the tree (Fig. 6a) are under positive selection we used a branch model (model = 1b in Codeml), which assumes an independent ω ratio for each branch and compared this log likelihood to the log likelihood of the model 0 (same ω for all branches) (Hall 2004).

Using Swiss model (Bienert et al. 2017), the crystal structure of UGT78G1 (3hbf) (Modolo et al. 2009) was used as template and the model was visualized in Pymol (Schrödinger LLC).The model was included in the alignment used for calculation of positive selection, and the amino acid residues corresponding to the sites where positive selection were identified and visualized.

Acknowledgements We thank Dr. Francisco R. Badenes-Pérez for the generous gift of hederagenin 3-*O*-β-cellobioside, Dr. Carl Erik Olsen for NMR-spectroscopy of this saponin, Dr. Elizabeth Heather Jakobsen Neilson for assistance in LC-MS analysis, Dr. Mika Zagrobelny for assistance in phylogenetic analysis and analysis for positive selection, Dr. Qing Liu for discussions and sharing a standard, Dr. Bekzod Khakimov for discussions and sharing a protocol for alkaline hydrolysis, Mette Sørensen for thorough reading and constructive ideas during writing, Michael Court and the UGT nomenclature committee for systematic UGT names, and three anonymous reviewers for constructive comments and suggestions. This work was funded by the Danish Council for Independent Research, Technology and Production Sciences (Grant No. 1335-00151) and Department of Plant and Environmental Sciences, University of Copenhagen (PhD stipend to PØE).

Author contributions Conceived research: PØE, SB. Planned experiments: All. Carried out research and analyzed data: PØE supervised by NA and SB. Wrote draft: PØE. Wrote paper: All.

References

Agerbirk N, Olsen CE (2011) Isoferuloyl derivatives of five seed glucosinolates in the crucifer genus. Barbarea Phytochem 72:610–623

Agerbirk N, Olsen CE, Nielsen JK (2001) Seasonal variation in leaf glucosinolates and insect resistance in two types of *Barbarea vulgaris* ssp. arcuata. Phytochemistry 58:91–100

Agerbirk N, Olsen CE, Bibby BM, Frandsen HO, Brown LD, Nielsen JK, Renwick JAA (2003a) A saponin correlated with variable resistance of *Barbarea vulgaris* to the diamondback moth *Plutella xylostella*. J Chem Ecol 29:1417–1433

Agerbirk N, Ørgaard M, Nielsen JK (2003b) Glucosinolates, flea beetle resistance, and leaf pubescence as taxonomic characters in the genus *Barbarea* (Brassicaceae). Phytochemistry 63:69–80

Agerbirk N, Olsen CE, Heimes C, Christensen S, Bak S, Hauser TP (2015) Multiple hydroxyphenethyl glucosinolate isomers and their tandem mass spectrometric distinction in a geographically structured polymorphism in the crucifer *Barbarea vulgaris*. Phytochemistry 115:130–142

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

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

Badenes-Perez FR, Reichelt M, Heckel DG (2010) Can sulfur fertilisation improve the effectiveness of trap crops for diamondback moth, *Plutella xylostella* (L.)(Lepidoptera: Plutellidae)? Pest Manag Sci 66:832–838

Badenes-Perez FR, Reichelt M, Gershenzon J, Heckel DG (2014) 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 98:137–144

Bienert S, Waterhouse A, de Beer TA, Tauriello G, Studer G, Bordoli L, Schwede T (2017) The SWISS-MODEL Repository—new features and functionality. Nucleic Acids Res 45:D313-D319

Byrne SL, Erthmann P, 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. Sci Rep 7:p 40728

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. Plant Cell 23:3070–3081

Christensen S, Heimes C, Agerbirk N, Kuzina V, Olsen CE, Hauser TP (2014) Different geographical distributions of two chemotypes of *Barbarea vulgaris* that differ in resistance to insects and a pathogen. J Chem Ecol 40:491–501

Chung Y, Choe S (2013) The regulation of brassinosteroid biosynthesis in Arabidopsis. Crit Rev Plant Sci 32:396–410

Dalby-Brown L, Olsen CE, Nielsen JK, Agerbirk N (2011) Polymorphism for novel tetraglycosylated flavonols in an eco-model crucifer *Barbarea vulgaris*. J Agric Food Chem 59:6947–6956

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. J Insect Physiol 58:18–23

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797

Geisler K, Hughes RK, Sainsbury F, Lomonossoff GP, Rejzek M, Fairhurst S, Olsen C-E, Motawia MS, Melton RE, Hemmings AM (2013) Biochemical analysis of a multifunctional cytochrome P450 (CYP51) enzyme required for synthesis of antimicrobial triterpenes in plants. Proc Natl Acad Sci 110:E3360–E3367

Hall BG (2004) Phylogenetic trees made easy: a how-to manual, vol 547. Sinauer Associates Sunderland, Sunderland

Hamberger B, Bak S (2013) Plant P450s as versatile drivers for evolution of species-specific chemical diversity. Philos Trans R Soc B 368:20120426

Heimes C, Agerbirk N, Sørensen H, van Mölken T, Hauser TP (2016) Ecotypic differentiation of two sympatric chemotypes of *Barbarea vulgaris* (Brassicaceae) with different biotic resistances. Plant Ecol 217:1055–1068

Husar S, Berthiller F, Fujioka S, Rozhon W, Khan M, Kalaivanan F, Elias L, Higgins GS, Li Y, Schuhmacher R (2011) Overexpression of the UGT73C6 alters brassinosteroid glucoside formation in *Arabidopsis thaliana*. BMC Plant Biol 11:1

Jung S-C, Kim W, Park SC, Jeong J, Park MK, Lim S, Lee Y, Im W-T, Lee JH, Choi G (2014) Two ginseng UDP-glycosyltransferases synthesize ginsenoside Rg3 and Rd. Plant Cell Physiol 55:2177–2188

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

Khakimov B, Tseng LH, Godejohann M, Bak S, Engelsen SB (2016) Screening for triterpenoid saponins in plants using hyphenated analytical platforms. Molecules 21:1614

Kunii M, Kitahama Y, Fukushima EO, Seki H, Muranaka T, Yoshida Y, Aoyama Y (2012) β-Amyrin oxidation by oat CYP51H10 expressed heterologously in yeast cells: the first example of CYP51-dependent metabolism other than the 14-demethylation of sterol precursors. Biol Pharma Bull 35:801–804

Kuzina V, Ekstrom 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 Physiol 151:1977–1990

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 72:188–198

Li J, Liu S, Wang J, Li J, Liu D, Li J, Gao W (2016) Fungal elicitors enhance ginsenosides biosynthesis, expression of functional genes as well as signal molecules accumulation in adventitious roots of Panax ginseng CA Mey. J Biotechnol 239:106–114

Liu T, Zhang X, Yang H, Agerbirk N, Qiu Y, Wang H, Shen D, Song J, Li X (2016) Aromatic glucosinolate biosynthesis pathway in *Barbarea vulgaris* and its response to *Plutella xylostella* infestation. Front Plant Sci 7:83

Lou X-M, Yao Q-H, Zhang Z, Peng R-H, Xiong A-S, Wang H-K (2007) Expression of the human hepatitis B virus large surface antigen gene in transgenic tomato plants. Clin Vaccine Immunol 14:464–469

Mackenzie PI, Bock KW, Burchell B, Guillemette C, Ikushiro S-i, Iyanagi T, Miners JO, Owens IS, Nebert DW (2005) Nomenclature update for the mammalian UDP glycosyltransferase (UGT) gene superfamily. Pharmacogenet Genomics 15:677–685

McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R (2013) Analysis tool web services from the EMBL-EBI. Nucleic Acids Res 41:W597-W600

Meesapyodsuk D, Balsevich J, Reed DW, Covello PS (2007) Saponin biosynthesis in *Saponaria vaccaria*. cDNAs encoding β-amyrin synthase and a triterpene carboxylic acid glucosyltransferase. Plant Physiol 143:959–969

Meßner B, Thulke O, Schäffner AR (2003) Arabidopsis glucosyltransferases with activities toward both endogenous and xenobiotic substrates. Planta 217:138–146

Modolo LV, Li L, Pan H, Blount JW, Dixon RA, Wang X (2009) Crystal structures of glycosyltransferase UGT78G1 reveal the molecular basis for glycosylation and deglycosylation of (iso) flavonoids. J Mol Biol 392:1292–1302

Moses T, Papadopoulou KK, Osbourn A (2014) Metabolic and functional diversity of saponins, biosynthetic intermediates and semi-synthetic derivatives. Crit Rev Biochem Mol Biol 49:439–462

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

Nielsen JK, Nagao T, Okabe H, Shinoda T (2010a) Resistance in the plant, Barbarea vulgaris, and counter-adaptations in flea beetles mediated by saponins. J Chem Ecol 36:277–285

Nielsen NJ, Nielsen J, Staerk D (2010b) New resistance-correlated saponins from the insect-resistant crucifer *Barbarea vulgaris*. J Agric Food Chem 58:5509–5514

Nørholm MH (2010) A mutant Pfu DNA polymerase designed for advanced uracil-excision DNA engineering. BMC Biotechnol 10:1

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

Osbourn A (1996) Saponins and plant defence—a soap story. Trends Plant Sci 1:4–9

Osbourn AE (2003) Saponins in cereals. Phytochemistry 62:1–4

Owatworakit A, Townsend B, Louveau T, Jenner H, Rejzek M, Hughes RK, Saalbach G, Qi X, Bakht S, Roy AD (2013) Glycosyltransferases from oat (Avena) implicated in the acylation of avenacins. J Biol Chem 288:3696–3704

Poppenberger B, Fujioka S, Soeno K, George GL, Vaistij FE, Hiranuma S, Seto H, Takatsuto S, Adam G, Yoshida S (2005) The UGT73C5 of *Arabidopsis thaliana* glucosylates brassinosteroids. Proc Natl Acad Sci USA 102:15253–15258

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

Schrödinger (LLC) The PyMOL Molecular Graphics System version 1.3. https://www.pymol.org/

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. Proceed Natl Acad Sci 105:14204–14209

Seki H, Tamura K, Muranaka T (2015) P450s and UGTs: key players in the structural diversity of triterpenoid saponins. Plant Cell Physiol 56:1463–1471

Shibuya M, Nishimura K, Yasuyama N, Ebizuka Y (2010) Identification and characterization of glycosyltransferases involved in the biosynthesis of soyasaponin I in Glycine max. FEBS Lett 584:2258–2264

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*. J Chem Ecol 28:587–599

Takos AM, Knudsen C, Lai D, Kannangara R, Mikkelsen L, Motawia MS, Olsen CE, Sato S, Tabata S, Jørgensen K (2011) Genomic clustering of cyanogenic glucoside biosynthetic genes aids their identification in *Lotus japonicus* and suggests the repeated evolution of this chemical defence pathway. Plant J 68:273–286

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729

Thimmappa R, Geisler K, Louveau T, O'Maille P, Osbourn A (2014) Triterpene biosynthesis in plants. Annu Rev Plant Biol 65:225–257

van Mölken T, Heimes C, Hauser TP, Sundelin T (2014) Phylogeny of an *Albugo* sp. infecting *Barbarea vulgaris* in Denmark and its frequency of symptom development in natural populations of two evolutionary divergent plant types. Fungal Biol 118:340–347

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

Vo NNQ, Fukushima EO, Muranaka T (2016) Structure and hemolytic activity relationships of triterpenoid saponins and sapogenins. J Nat Med:1–9

Wei X, Zhang X, Shen D, Wang H, Wu Q, Lu P, Qiu Y, Song J, Zhang Y, Li X (2013) Transcriptome analysis of *Barbarea vulgaris* infested with diamondback moth (*Plutella xylostella*) larvae. PLoS ONE 8:e64481

Xu B, Yang Z (2013) PAMLX: a graphical user interface for PAML. Mol Biol Evol 30:2723–2724

Yang Z (2000) Phylogenetic analysis by maximum likelihood (PAML). Version. https://abacus.gene.ucl.ac.uk/software/paml.html

Yang Z, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. Trends Ecol Evol 15:496–503

Yang Z, Nielsen R, Goldman N, Pedersen A-MK (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449

Zhang X, Liu T, Wei X, Qiu Y, Song J, Wang H, Shen D, Agerbirk N, Li X (2015) Expression patterns, molecular markers and genetic diversity of insect-susceptible and resistant *Barbarea* genotypes by comparative transcriptome analysis. BMC Genomics 16:486