Investigation of glucosyltransferases essential for biosynthesis of the antifungal triterpene glycoside avenacin in oats

Anastasia Lily Orme

A thesis submitted to the University of East Anglia for the degree of Doctor of Philosophy

©November 2017

John Innes Centre
Norwich Research Park
Norwich
Norfolk
United Kingdom

©This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there from must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Abstract

Triterpene glycosides are a large, diverse class of amphipathic plant-derived natural products that have high potential as medicinal, agricultural and industrial products. Currently their use is limited as they are typically produced in small amounts in plants and they are difficult to extract or chemically synthesise.

In this thesis, two glucosyltransferases from oat (*Avena strigosa*) required for the biosynthesis of the trisaccharide sugar chain of the antifungal triterpene glycoside, avenacin A-1, were identified. A candidate UDP-glucose dependent glycosyltransferase, AsUGT91, was identified by mining an oat root transcriptome database, phylogenetic analysis and expression profiling. *In vitro* assays and transient expression in *Nicotiana benthamiana* confirmed triterpene 3-*O*-arabinoside β-1,2-glucosyltransferase activity.

The triterpene 3-*O*-arabinoside β-1,4-glucosyltransferase, AsTG, was identified by physical linkage to the avenacin genes, which are clustered in the oat genome. AsTG is a vacuolar transglucosidase in glycosyl hydrolase family 1 (GH1) and is the first of this class of enzyme to be involved in triterpene biosynthesis.

Subsequent analysis of oat mutants revealed that *AsUGT91* and *AsTG* correspond to loci required for avenacin glycosylation. Both sets of mutants have root developmental defects, are deficient in avenacin production and show increased susceptibility to the take-all fungal pathogen, *Gaeumannomyces graminis* var. *tritici*.

This work increases the range of glycoside glycosyltransferase activities available to create novel triterpene glycosides by synthetic biology; contributes towards engineering resistance in other crop species to the agriculturally important root disease, 'take-all'; and expands the knowledge of triterpene glycoside biosynthesis to include an unusual class of plant specialised metabolite glycosyltransferases.

Acknowledgements

I would like to express my deepest thanks to my supervisor, Prof Anne Osbourn, for giving me the opportunity to work in her fantastic group, and for her excellent guidance, feedback and help in writing this thesis. I would also like to thank her for her incredible support, particularly in the last months of this PhD.

I am grateful to my secondary supervisor Prof Rob Field, for his helpful advice and input on this project over the years, and Dr Thomas Louveau for his invaluable guidance, patience and advice, including in the writing of this project; and for sharing his vast knowledge of working with UGTs and triterpene glycosides.

I am grateful to the many members of the Osbourn group, past and present, for making the lab such an enjoyable place to work. In particular, I thank Dr James Reed for his constant encouragement, assistance and friendship throughout this PhD. I am additionally grateful to Dr Aymeric Leveau for his tireless help in troubleshooting my experiments; Dr Michael Stephenson for extensive advice and assistance in my attempts to purify my compounds; and Rachel Melton for her considerable help and training throughout my project. I would particularly like to thank Thomas, Hannah and James for their support throughout the writing process, and in these past few weeks.

I would also like to thank Dr Ingo Appelhagen for his generous assistance with the localisation assays; and Dr D. M. Jones (congratulations Marc!) for helping me learn to code, and for sharing his scripts and advice for the RNA-seq co-expression analysis. And also for his habit of leaving flammable items in the oven.

I am very grateful for the support of Graduate Studies, and in particular Prof Mike Merrick for his assistance and advice; and Prof Mark Buttner and Dr Jake Malone for their understanding, supervision and support at the start of this PhD. I am grateful to the BBSRC for funding and excellent training, and the JIC, for teaching me about plants and being a wonderful place to carry out research. In extension I would like to acknowledge the amazing Media Kitchen staff, Horticultural Services, and especially Metabolite Services: thank you to Paul and Lionel for their patience, training and friendship.

I would like to thank my friends and housemates (in particular Zena, Laura, Hannah, Nathan, Peter, Rachel and Jemima); Natasha for her understanding, fun and LaTeX advice; and Jake and Jack for welcoming me into their home, and breaking up my writing with their antics. I would also like to thank Lynn James for her beautiful emails.

I would like to thank my partner, Guy, for his constant support throughout everything that I decide to do; and whose interest in my project and enthusiasm to talk through chemistry-related aspects over four years have been incredible. Thank you for everything – and for all of the coffee.

Finally, I thank my family for their humour and strength: my brother, for his inspiring curiosity and thought-provoking questions; my sister-in-law for her support from afar; my sister for her wisdom, (and coding help) and wonderful support; my mother for her advice, positivity and love; and my Dad, for always picking up the phone to talk through anything, and for his incredible strength, perspective and love.

List of common abbreviations

AsbAS1	*Avena strigosa* β-amyrin synthase
cM	centiMorgan
EDTA	Ethylenediaminetetraacetic acid
EpHβA	12,13β-epoxy,16β-hydroxy-β-amyrin
ER	Endoplasmic Reticulum
GADPH	Glyceraldehyde-3-phosphate dehydrogenase
GFP	Green Fluorescent Protein
GGT	Glycoside-specific glycosyltransferase
GT	Glycosyltransferase
HMGR	3-hydroxy,3-methylglutaryl-CoA reductase
HPLC	High performance liquid chromatography
HPLC-CAD	High performance liquid chromatography charged aerosol detection
IPTG	Isopropyl β-D-1-thiogalactoside
IS	Internal Standard
LB	Luria-Bertani
LC-MS	Liquid Chromatography Mass Spectrometry
m/z	Mass-to-charge ion ratio
MEP	2-C-methyl-D-erythritol 4-phosphate
MS^2	Mass spectrometry/mass spectrometry
MVA	Mevalonate
NMR	Nuclear Magnetic Resonance
OD	Optical density
OSC	Oxidosqualene cyclase
P450	Cytochrome P450
PAGE	Polyacrylamide gel electrophoresis
PSPG	Plant Secondary Product Glycosyltransferase
RFP	Red Fluorescent Protein
sad	saponin-deficient
SDS	Sodium Dodecyl Sulphate
TG	Transglycosidase
TIC	Total Ion Chromatogram
TLC	Thin Layer Chromatography
Tris	Tris(hydroxymethyl)aminomethane
UDP	Uridine diphosphate
UGT	UDP-dependent glycosyltranferases

Contents

List of Figures

List of Tables

Chapter 1

General introduction

1.1 Triterpene glycosides (saponins)

Plants make thousands of specialised metabolites that have a diverse role in their survival in the environment. These metabolites fall into three main groups: phenylpropanoids, alkaloids and terpenoids. Triterpene glycosides are a large and varied class of terpenoids that are often associated with plant defence mechanisms due to the prevalence of antimicrobial, allelopathic, molluscicidal, insecticidal and cytotoxic properties (Sparg et al., 2004; Augustin et al., 2011; Osbourn et al., 2011; Moses et al., 2014). Although most triterpene diversity is found in plants, particularly in dicots, some examples are also found in microbes and other eukaryotes such as starfish and sea cucumbers.

Triterpene glycosides have a wide range of different properties with many potential applications in medicine, industry and agriculture (Osbourn et al., 2011; Moses et al., 2014). They are formed of a hydrophobic triterpene backbone attached to hydrophilic saccharide chains. The amphipathic nature of triterpene glycosides confers foaming and emulsifying properties, and together with the related steroidal glycosides these compounds are classed as 'saponins'. They have been used traditionally as soaps and more recently as beverage additives (Güçlü-Üstündağ and Mazza, 2007). The amphipathic properties of triterpene glycosides can also confer the ability to permeabilise eukaryotic membranes, causing cell lysis. Effective cell lysis depends on various aspects of the molecule, such as the type of triterpene backbone, the triterpene oxidations, and differences in the types of sugar and linkages within the saccharide chains. Not all triterpene glycosides have this ability (Augustin et al., 2011; Osbourn et al., 2011; Moses et al., 2014).

Triterpene glycosides can activate the mammalian immune system and are used or are being developed as vaccine adjuvants (Sun et al., 2009). The triterpene glycoside glycyrrhizic acid from liquorice (*Glycyrrhiza*) is proposed to have wide-ranging properties such as anti-inflammatory, antiviral, antioxidative, anticancer, immunomodulatory activities (Augustin et al., 2011; Osbourn et al., 2011) and the triterpene mogroside glycosides from the Chinese curcubit *Siraitia grosvenorii* are being developed as low-calorie sweeteners (Itkin et al., 2016).

It is difficult to study the properties of triterpene glycosides and their potential uses due to the lack of availability of pure compounds. Triterpene glycosides are difficult to isolate from natural sources due to their presence in low amounts in composite mixtures, and their chemical complexity impedes chemical or partial synthesis. However, recent advances in the knowledge of triterpene glycoside biosynthesis now enable the production of multiple triterpene glycosides in heterologous hosts in sufficient amounts to evaluate their properties systematically, opening up opportunities to scale-up production for applications in medicine and industry. This approach can be combined with semi-synthetic chemistry to create novel compounds with enhanced properties such as reduced toxicity (Augustin et al., 2011; Osbourn et al., 2011). In addition, knowledge of the biosynthetic pathways can be used to engineer new traits into crops, or to modify undesirable traits that are associated with triterpene glycosides (Heng et al., 2006; Osbourn et al., 2011).

1.2 Triterpene glycoside biosynthesis

Terpenes are formed from the condensation of five-carbon building blocks, 3-isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). In plants, these two compounds can be formed in the plastids via the methyl-erythritol-phosphate (MEP) pathway or in the cytosol via the mevalonate (MVA) pathway (Vranova et al., 2013). Most classes of terpenes (monoterpenes, diterpenes and carotenoids) are synthesized via the MEP pathway, however triterpenes are synthesized by the MVA pathway, and are closely related to sterols of primary metabolism (Figure 1.1).

Both triterpenes and sterols are formed by the cyclisation of the linear precursor 2,3-oxidosqualene by oxidosqualene cyclases (OSCs) (Figure 1.2). Triterpenes and sterols are distinguished by the conformation of cyclisation (Thimmappa et al., 2014). The triterpene aglycone backbones are stereospecifically oxidised by cytochrome P450-dependent mono-oxygenases (P450s), increasing the polarity of the scaffold and

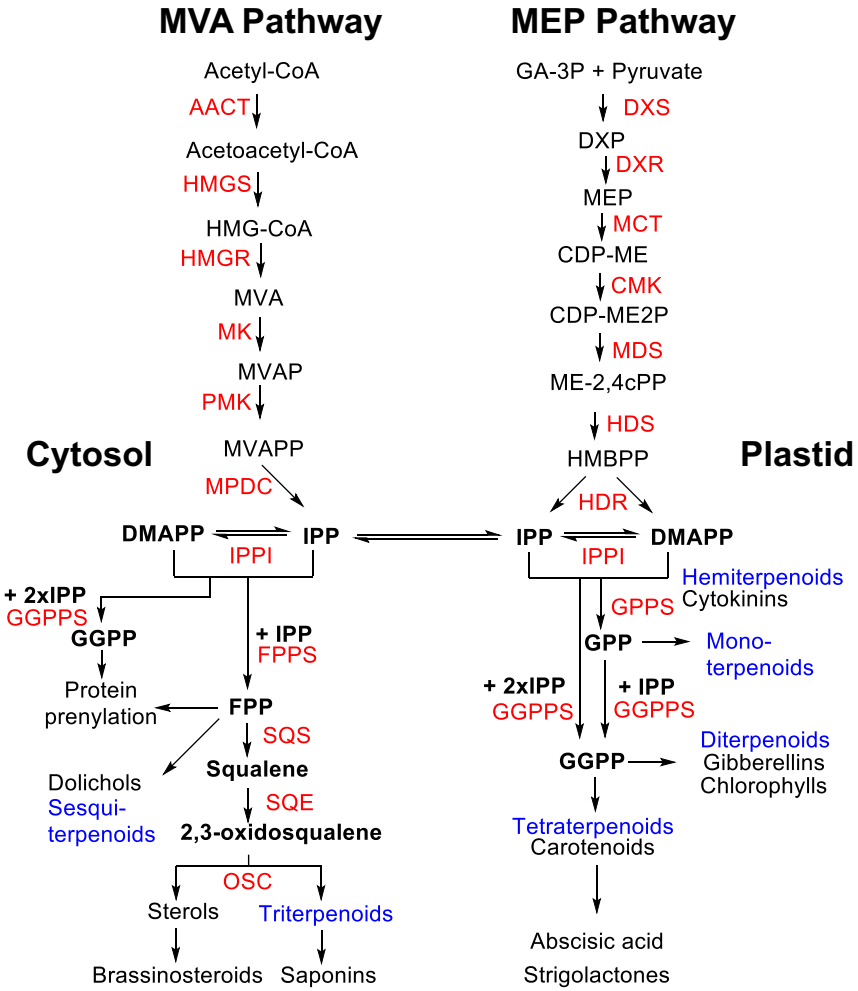


Figure 1.1: The MVA and MEP pathways.
Trafficking of IPP and DMAPP occurs between compartments (double- headed arrows). Compound abbreviations: MVA Pathway: CoA, Coenzyme A; HMG, 3- hydroxy-3-methylglutaryl-CoA; MVA, mevalonate; MVAP, Mevalonate-5-phosphate; MVAPP, Mevalonate-5-diphosphate. MEP pathway: GA-3P, D-glyceraldehyde-3-phosphate; DXP, 1-deoxy-D-xylulose 5-phosphate; MEP, 2-C-Methyl-D-erythritol 4-phosphate; CDP-ME, 4- (Cytidine 5′-diphospho)-2-C-methyl-D-erythritol; CDP-ME2P, 2-Phospho-4-(cytidine 5′- diphospho)-2-C-methyl-D-erythritol; ME-2,4-cPP, 2-C-Methyl-D-erythritol ,2,4- cyclodiphosphate; HMBPP, 4-Hydroxy-3-methylbut-2-enyl-diphosphate; Downstream: IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate; GPP, geranyl diphosphate; FPP, farnesyl diphosphate; GGPP, geranylgeranyl diphosphate. Enzyme abbreviations: MVA Pathway: AACT, Acetyl-CoA C-acetyltransferase; HMGS, HMG-CoA synthase; HMGR, HMG-CoA reductase; MK, MVA kinase; PMK, Phospho-MVA kinase; MPDC, Diphospho-MVA kinase. MEP pathway: DXS, DXP synthase; DXR, DXP reductoisomerase; MCT, MEP cytidylyltransferase; CMK. CDP-ME kinase; MDS, ME-2,4cPP synthase; HDS, HMBPP synthase; HDR, HMBPP reductase; Downstream: IPPI, IPP isomerase; GPPS, GPP synthase; FPPS, FPP synthase; GGPPS, GGPP synthase; SQS, Squalene synthase; SQE, Squalene epoxidase; OSC, Oxidosqualene cyclase. Figure and figure legend adapted from Reed (2016).

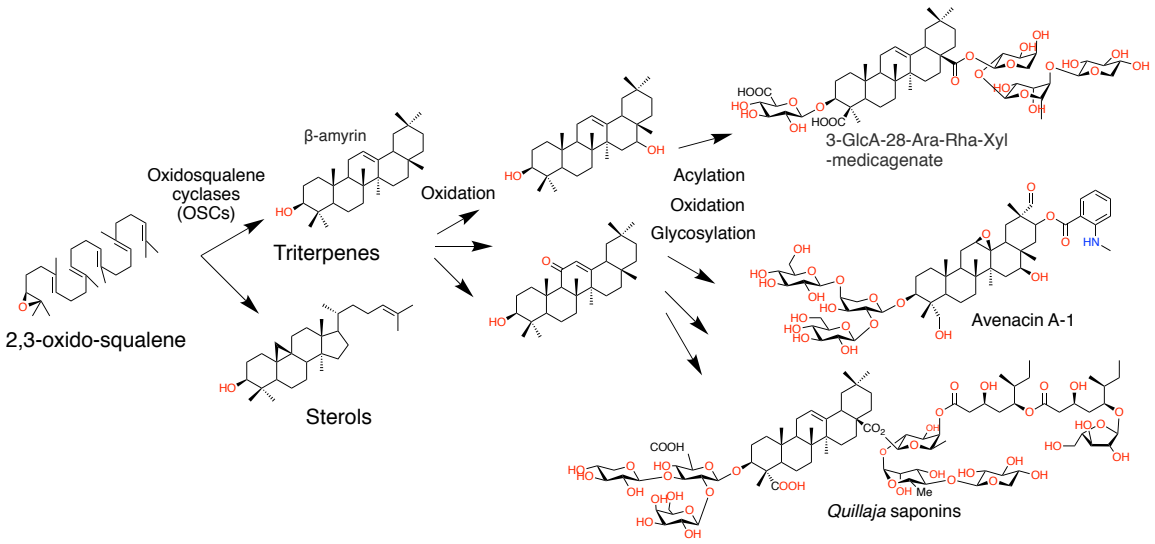


Figure 1.2: Triterpene glycoside biosynthesis and diversification.
Figure adapted from Reed (2016) and Osbourn (2011).

creating functional groups for further modification. The saccharide side chains are then added by glycosyltransferases, although triterpene glycosides may also be modified by acyltranferases, malonyltranferases and methyltransferases.

The addition of sugar molecules contributes drastically to the immense diversity of triterpene glycosides. Sugar modifications can be added to -OH, -COOH, -NH_2, -SH and C-C groups, and can be monodesmosidic (only one position of the aglycone is modified), bidesmosidic (two saccharide chains) and tridesmosidic in rare cases. Multiple sugars can be added (typically 2-5 units) and the saccharide chains may be branched, and include acyl groups. Glucose, galactose, glucoronic acid, rhamnose, xylose and arabinose are common sugars in the structures of plant natural products, but fucose, quinovose, ribose and apiose may also be incorporated (Bowles et al., 2006; Vincken et al., 2007).

Few triterpene glycosyltransferases and even fewer triterpene glycoside glycosyltransferases have been described, which would be required to open up the diversity of triterpenes for metabolic engineering.

1.3 Plant triterpene glycoside glycosyltransferases

Glycosyltransferases (GTs) catalyze the formation of glycosidic bonds between activated sugar molecules and many varied acceptors (Lairson et al., 2008). GTs are clas-

sified according to amino acid similarities and fall into over 95 different families in the CAZy online database (http://www.cazy.org/) (Coutinho et al., 2003; Campbell et al., 1997; Lombard et al., 2014). Although primary sequences vary widely between GTs, nucleotide sugar-dependent GTs have highly conserved tertiary structures that mainly fall into two main folds, GT-A and GT-B, suggesting a common evolutionary origin within these groups (Coutinho et al., 2003). These groups can be sub-divided into clans based on the glycosyltransfer mechanism, which either occurs with inversion or retention of the anomeric configuration of the sugar donor (Coutinho et al., 2003).

Plant glycosyltransferases that are involved in the biosynthesis of specialised metabolites use uridine diphosphate (UDP) sugars as activated sugar donors and are classed in glycosyltransferase family 1 (GT1)[1]. GT1 enzymes have a GT-B fold which consists of two flexibly linked $\beta/\alpha/\beta$ Rossmann domains with the catalytic site located in a cleft between the N- and the C-terminal domains (Lairson et al., 2008). The sugar acceptor and donor pockets are located in the N-and C-terminal domains, respectively, although both domains contribute important interactions in the binding of both substrates (Osmani et al., 2009).

These enzymes catalyse the direct S_N2-like displacement of the activated sugar leaving group, inverting the anomeric stereochemistry with respect to the donor sugar (Coutinho et al., 2003; Shao et al., 2005; Lairson et al., 2008). The majority of plant Family 1 GTs have a conserved histidine in the N-terminal domain at approximately residue 22 which acts as a base catalyst to deprotonate the nucleophile of the acceptor (Lairson et al., 2008; Wang, 2009). A conserved aspartate positioned around the 121^{st} residue forms a hydrogen bond with the histidine, potentially forming an acceptor-His-Asp triad to stabilise the interaction (Wang, 2009). The conserved histidine and aspartate are occasionally replaced with cysteine and arginine/serine/asparagine respectively, and may not required for the mechanism of action of all plant GT1 enzymes (Noguchi et al., 2007; Wang, 2009).

GTs that use UDP-sugars are called UDP-dependent glycosyltranferases (UGTs) and contain a conserved motif (UDPGT, PROSITE accession number PS00375) which is associated with binding of the UDP-sugar substrate. The motif is used to catalogue UGTs from animals, yeasts, plants and bacteria into superfamilies containing 40% or more amino acid identity; UGT families 71-100 are reserved for plants (Mackenzie et al., 1997).

[1] A few flavanoid glycosyltransferases are members of glycosyl hydrolase family 1 (Matsuba et al., 2010; Miyahara et al., 2012; Luang et al., 2013; Nishizaki et al., 2013; Miyahara et al., 2014).

Plant UGTs fall into three groups: two minor clades of sterol and lipid glycosyltransferases (UGT80 and UGT81, respectively) that appear more related to non-plant UGTs and are classed in GT family 28; and a large monophyletic group in GT family 1 that contains a characteristic version of the UGT motif, the Plant Secondary Product Glycosyltransferase (PSPG) motif (Figure 1.3) (Hughes and Hughes, 1994; Paquette et al., 2003). Plant GT1 UGTs can be further classified into 16 phylogenetic groups, from group A to group P (Ross et al., 2001; Caputi et al., 2012).

These GT1 UGTs have diverse roles in the modification of small lipophilic molecules. For example, glycosylation can stabilise volatile or labile compounds such as glucosinolates or pigmented anthocyanins that would otherwise convert to colourless products at physiological pH (Rask et al., 2000; Jones and Vogt, 2001; Zhao et al., 2014). Glycosylation takes place at nucleophilic sites on compounds that may interact deleteriously with other molecules in plant cells; sugar modifications can reduce toxicity by blocking the interaction site of toxic compounds such as xenobiotics or toxic metabolic intermediates (Jones and Vogt, 2001). The change in polarity and solubility affects the free diffusion of lipophilic compounds across lipid membranes and allows their storage in specific subcellular compartments (Bowles and Lim, 2010). In addition, the glycosylation of lipophilic scaffolds affects the binding interactions of these molecules with various biological consequences. For instance, molecules that may be inherently stable can be recognised by catabolic enzymes depending on their glycosylation state (Jones and Vogt, 2001) and glycosylation is used as a rapid and transient signal between active and inactive forms, such as in the homeostasis of plant hormones (Bowles and Lim, 2010).

1.3.1 Predicting UGT activity based on primary structure features

The GT1 family is expanded in vascular plants and can account for around 0.5% of all protein-coding genes (Caputi et al., 2012). This typically equates to 100-200 possible UGTs in a plant genome, and it can be difficult to infer which UGT is involved in a specific biosynthetic pathway.

UGTs are relatively specific in terms of the sugar donors that they can use, with preferential activity with one sugar donor where multiple activities are reported. The 44-amino acid PSPG motif in the C-terminal domain lines one side of the sugar donor pocket, where highly conserved residues interact with the UDP-sugar donor or contribute to intramolecular interactions (Osmani et al., 2009).

The crystallised structures of plant UGTs (which all have a preference for UDP-

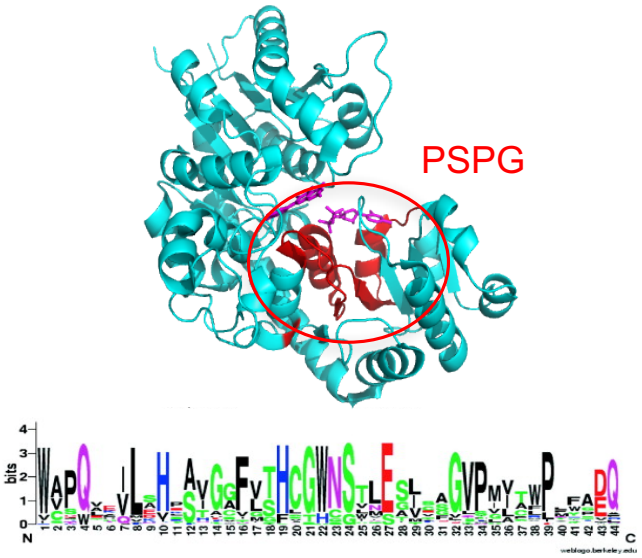


Figure 1.3: A crystal structure of a UGT (PDB:2C9Z) with the PSPG motif highlighted in red. WebLogo of the plant specific PSPG motif (Osmani et al., 2009). Letter size is proportional to the degree of amino acid conservation.

glucose) show ten of the PSPG motif residues to interact directly with a UDP-glucose substrate (Osmani et al., 2009). Seven of these ten direct interactions involve the invariant UDP part of the UDP-glucose, whilst three form hydrogen-bonds with hydroxyl groups of the glucose moiety (Osmani et al., 2009). These residues are a highly conserved tryptophan (W) in position 22 of the PSPG motif, and the last two positions (43 and 44), which are conserved as aspartic acid/glutamic acid/serine (D/E/S) and glutamine/glutamic acid/histidine/asparagine (Q/E/H/N) respectively.

The direct interactions between residues of the PSPG motif to the sugar hydroxyl groups would be altered with different sugar donors. Whilst the residues involved could be expected to vary with the sugar preference of the UGT, this does not appear to occur (Osmani et al., 2009). For example, residue 43 (aspartic acid in the crystallised structures) interacts with the 4-OH of UDP-glucose. The 4-OH of D-galactose and L-arabinose are axial on the sugar ring and point in a different direction compared to the equatorial 4-OH of D-glucose. However it is predicted that the aspartic acid can still form an interaction with these sugars, and this residue is conserved in arabinosyltransferases and galactosyltransferases (Kubo et al., 2004).

The final PSPG residue, (Q) is often replaced with a histidine (H) in UGTs that transfer D-galactose or L-arabinose and may be indicative of these activities (Kubo et al., 2004; Louveau et al, manuscript in preparation). However a histidine at this po-

sition is not necessary or sufficient for galactosyltransferase activity. Studies of the *Aralia chordata* galactosyltransferase showed that mutation of the histidine (H) to glutamine (Q) conferred glucosyltransferase activity; however the reverse mutation tends to simply reduce activities with UDP-glucose, and UDP-galactose if present (Kubo et al., 2004). *Solanum tuberosum* SGT and *Vitis vinefera GT1* have a glutamine (Q) in the 44th position and exhibit both glucosyl- and galactosyl-transferase activity (Offen et al., 2006; Kohara et al., 2007).

Other residues and regions, including within the N-terminus, have been implicated in sugar donor preference (Osmani et al., 2008; Dai et al., 2017). Highly homologous UGTs that differ in only a few amino acids in their protein sequences can have different sugar donor specificities. For example, the sugar donor specificity of the highly similar UGT73F2 and UGT73F4 from *Glycine max* is UDP-glucose and UDP-xylose respectively, and proposed to depend on one amino acid difference (Gly-138 and Ser-138) (Sayama et al., 2012). Similarly, UGT89A2 in different *Arabidopsis thaliana* accessions have an amino acid difference (Ile-153 and Ser-153) in the same region that confers specificity for UDP-xylose or activity towards both UDP-xylose and UDP-glucose (Chen and Li, 2017). Sugar donor preference appears to depend on the interaction of several unknown residues and as yet it is unclear which combinations of residues confers specificity (Offen et al., 2006; Osmani et al., 2009).

The sugar acceptor pocket is mostly formed from the N-terminal domain, although the C-terminal domain contributes to its formation (Osmani et al., 2009). It is composed of several helices and loops that are highly varied in sequence and length and interactions with acceptors are mainly hydrophobic in nature (Osmani et al., 2009; Wang, 2009).

In general, acceptor specificity appears to rely on the ability to correctly position the acceptor near the catalytic base and the 1C of the sugar donor by the position of variable loops and the size of the acceptor pocket (Osmani et al., 2009; Wang, 2009). A promiscuous UGT from *Medicago truncatula*, UGT71G1, that can glycosylate triterpenes and all five hydroxyls of quercetin was shown to have a large, more open pocket, and mutations to alter the shape of the pocket altered quercetin glycosylation regiospecificity (He et al., 2006).

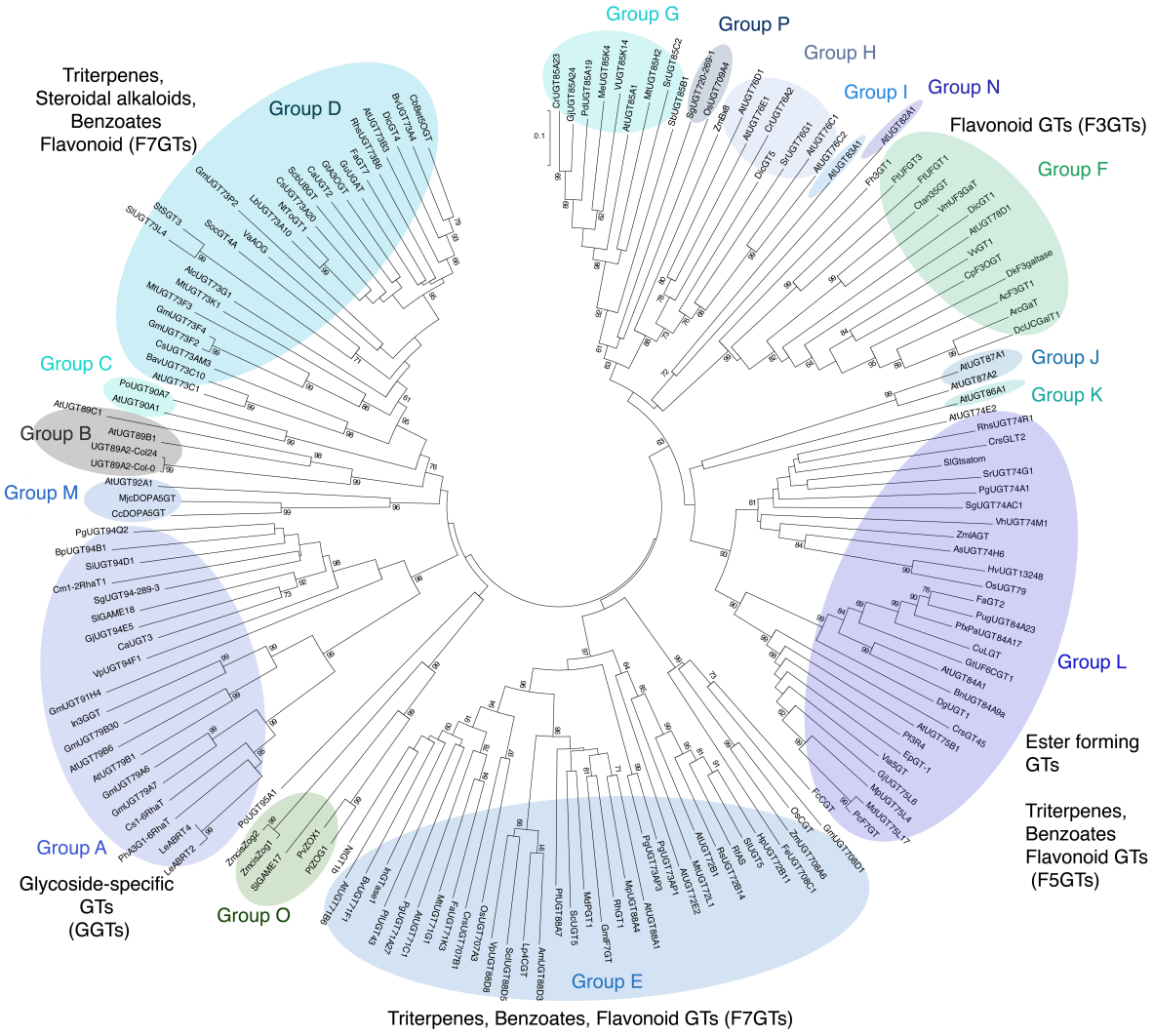


Figure 1.4: Phylogenetic tree of plant UGT sequences. The UGT phylogenetic groups (Groups A-N) are labelled as described in (Ross et al., 2001; Caputi et al., 2012; Osmani et al., 2009). Protein sequences of UGTs from different plant species (listed in Table S3) were aligned using MAFFT (https://mafft.cbrc.jp/alignment/software/) and the evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987) and analysed with 1000 bootstrap replicates which are shown at the nodes. The branches with <60% support were not marked. The unrooted tree was constructed using MEGA7 (Kumar et al., 2016).

Attempts to infer acceptor specificity based on UGT phylogeny are generally difficult due to the lack of any characterised members within many UGT phylogenetic groups (such as groups I, J or K), the high levels of convergent evolution and the promiscuity shown by UGTs in *in vitro* assays (Bowles et al., 2006; Osmani et al., 2009; Augustin et al., 2011; Caputi et al., 2012). For example, members of the same phylogenetic group can recognise acceptors belonging to different compound classes, such as flavonoids, terpenoids and benzoates (Caputi et al., 2008; Lim et al., 2004, 2002). Multiple members of three different phylogenetic groups (D, E and F) were all able to recognize the same acceptor substrate, quercetin, and glucosylate it at the 3-*O* position in an *in vitro* assay (Lim et al., 2004). Enzymes in groups D and E are particularly promiscuous; UGT73AE1 (Group D) from *Carthamus tinctorius* was found to be a highly active trifunctional (creating *O*-, *S*- and *N*-glycosides) and reversible UGT (Xie et al., 2014). RhGT1 from Group E sequentially glucosylates anthocyanidin at the 5-*O*, then the 3-*O* position *in planta* (Ogata et al., 2005). The promiscuity of a subset of UGTs may be involved in the protective response to abiotic or biotic stress *in planta* by glycosylating xenobiotics or toxic metabolic intermediates (Brazier-Hicks et al., 2017). Not all UGTs are promiscuous enzymes; some may be intrinsically specific or they may be spatially separated from potential substrates or pre- or post-transcriptionally regulated to prevent deleterious cross-talk impacting processes in primary or secondary metabolism (Jones and Vogt, 2001; Lim et al., 2004; Bowles et al., 2006; Osmani et al., 2009).

Some general rules may be implied by phylogeny (Figure 1.4). Some GTs may glycosylate varied acceptors at structurally similar acceptor sites, with related enzymes sharing the same regiospecificity (Li et al., 2001; Lim et al., 2003; Frydman et al., 2013). Group L enzymes are distinctive in their ability to glycosylate the carboxyl group of various small molecules to form glucose ester bonds (Lim et al., 2001; Caputi et al., 2008) and group G enzymes from *A. thaliana* were associated with the glycosylation of primary hydroxyl groups of linear or cyclic substrates (Caputi et al., 2008). While not unique to this group, all of the characterised enzymes from Group A (UGT79, UGT91 and UGT94) are glycoside-specific glycosyltransferases (GGTs), forming β-1,2- or β-1,6-linkages, strongly implicating these enzymes in the extension of sugar chains.

Not many UGTs have been reported as triterpene glycosyltransferases (Table 1.1). The currently characterised triterpene UGTs fall into groups A, D, E, L and P, although it may emerge that other UGT groups also contain triterpene UGTs as more enzymes are characterised.

Table 1.1: Characterised triterpene GT Family 1 UGTs

Enzyme name	UGT family	Group	Activity	Plant species	Accession Number	Reference
GmUGT91H4	UGT91	A	Triterpene 3-*O*-galactoside 1,2-rhamnosyltransferase	*Glycine max*	BAI99585	Shibuya et al. (2010)
PgUGT94Q2	UGT94	A	Triterpene 3-*O*-glucoside 1,2-glucosyltransferase	*Panax ginseng*	AGR44632	Jung et al. (2014)
SgUGT94-289-3	UGT94	A	Triterpene 24-*O*-glucoside/3-*O*-glucoside 1,2-/1,6-glucosyltransferase	*Siraitia grosvenorii*		Itkin et al. (2016)
SgUGT94-289-2	UGT94	A	Triterpene 24-*O*-glucoside 1,2-/1,6-glucosyltransferase	*Siraitia grosvenorii*		Itkin et al. (2016)
SgUGT94-289-1	UGT94	A	Triterpene 24-*O*-glucoside/3-*O*-glucoside 1,6-glucosyltransferase	*Siraitia grosvenorii*		Itkin et al. (2016)
BvUGT73C10	UGT73	D	Triterpene 3-*O*-glucosyltransferase	*Barbarea vulgaris*	AFN26666	Augustin et al. (2012)
BvUGT73C11	UGT73	D	Triterpene 3-*O*-glucosyltransferase	*Barbarea vulgaris*	AFN26667	Augustin et al. (2012)
BvUGT73C12	UGT73	D	Triterpene 3-*O*-glucosyltransferase	*Barbarea vulgaris*	AFN26668	Augustin et al. (2012)
BvUGT73C13	UGT73	D	Triterpene 3-*O*-glucosyltransferase	*Barbarea vulgaris*	AFN26669	Augustin et al. (2012)
CsUGT73AM3	UGT73	D	Triterpene 3-*O* glucosyltranferase	*Cucumis sativus*	KGN59015	Zhong et al. (2017)
GmUGT73F2	UGT73	D	Triterpene 22-*O*-arabinoside 1,3-glucosyltransferase	*Glycine max*	BAM29362	Sayama et al. (2012)
GmUGT73F4	UGT73	D	Triterpene 22-*O*-arabinoside 1,3-xylosyltransferase	*Glycine max*	BAM29363	Sayama et al. (2012)
GmUGT73P2	UGT73	D	Triterpene 3-*O*-glucoronide 1,2-galactosyltransferase	*Glycine max*	BAI99584	Shibuya et al. (2010)
GuUGAT	UGT73	D	Triterpene 3-*O*-glucoronosyltransferase/ Triterpene 3-*O*-glucuronide 1,2-glucuronosyltransferase	*Glycyrhiza uralensis*	ANJ03631	Xu et al. (2016a)
MtUGT73K1	UGT73	D	Triterpene glucosyltransferase	*Medicago truncatula*	AAW56091	Achnine et al. (2005)
MtUGT73F3	UGT73	D	Triterpene 28-*O*-glucosyltransferase	*Medicago truncatula*	ACT34898	Naoumkina et al. (2010)
MtUGT71G1	UGT71	E	Triterpene glucosyltransferase	*Medicago truncatula*	AAW56092	Achnine et al. (2005)
PgUGT71A27	UGT71	E	Triterpene 20-*O*-glucosyltransferase	*Panax ginseng*	AIZ00429	Wei et al. (2015)
UGTPg100	UGT71	E	Triterpene 6-*O*-glucosyltransferase	*Panax ginseng*	AKQ76388	Wei et al. (2015)
UGTPg101	UGT71	E	Triterpene 20-*O*/6-*O*-glucosyltransferase	*Panax ginseng*	AKQ76389	Wei et al. (2015)
VhUGT74M1	UGT74	L	Triterpene 28-*O* carboxylic acid glucosyltransferase	*Vaccaria hispanica*	ABK76266	Meesapyodsuk et al. (2007)
PgUGT74AE2	UGT74	L	Triterpene 3-*O*-glucosyltransferase	*Panax ginseng*	AGR44631	Jung et al. (2014)
SgUGT74AC1	UGT74	L	Triterpene 3-*O*-glucosyltransferase	*Siraitia grosvenorii*	AEM42999	Dai et al. (2015)
SgUGT720-269-1	UGT720	P	Triterpene 24-*O*/3-*O*-glucosyltransferase	*Siraitia grosvenorii*		Itkin et al. (2016)

1.3.2 Triterpene glycoside biosynthetic genes are typically co-regulated with other pathway genes

Triterpene glycosides are often accumulated in specific tissues, at different developmental times or in response to abiotic or biotic stress (Augustin et al., 2011; Moses et al., 2014; Thimmappa et al., 2014). The metabolome and transcriptome of plants have been shown to overlap (Matsuda et al., 2010) and most triterpene OSCs have been identified using expression-based strategies (Thimmappa et al., 2014).

Co-expression with key enzymes such as the OSC is also a highly successful method to identify triterpene biosynthetic enzymes (Thimmappa et al., 2014). For example, the first triterpenoid UGTs, UGT71G1 and UGT73K1 were identifed with expressed sequence tag (EST) libraries to find UGTs co-expressed with the OSC gene in *Medicago truncatula* (Achnine et al., 2005).

1.3.3 Genes encoding specialised metabolic pathways are often physically clustered in plant genomes

Genes for primary metabolic pathways in plants are generally widely dispersed in plant genomes. However, many "operon-like" gene clusters in secondary metabolism, where nonhomologous genes involved in the same pathway are located together but transcribed separately, have been discovered (Field and Osbourn, 2008; Nutzmann and Osbourn, 2014).

These clusters appear to have formed *de novo* and not by horizontal gene transfer from bacteria (Chu et al., 2011). Clusters may be formed by the duplication of gene pairs and neofunctionalisation, followed by rearrangements and recruitment of further genes (Chu et al., 2011; Field et al., 2011; Field and Osbourn, 2012). Plant genomes are dynamic, with stochastic polyploidisation, duplications and invasion by transposable elements (Chu, 2013). The maize DIMBOA pathway and the avenacin pathway are subtelomeric which have high levels of chromosome recombination (Chu et al., 2011; Field and Osbourn, 2012). The *A. thaliana* triterpene marnerol and thalianol clusters are not sub-telomeric, but are in dynamic regions of chromosomes rich in transposable elements that have arisen since the last whole-genome duplication event (Field et al., 2011).

Clustering may facilitate the co-regulation of pathway genes at the level of chromatin (Chu et al., 2011; Field and Osbourn, 2012). Since the clustered pathways make, or are likely to make, compounds that confer a selective advantage, there is likely to be strong selection for maintenance of biosynthetic gene clusters. Furthermore, dis-

ruption of clustered pathways by mutation can result in the accumulation of toxic pathway intermediates, as seen for late avenacin pathway mutants in oats (Mylona et al., 2008; Chu et al., 2011).

Where genome sequence data are available, looking for physically linked genes is a useful strategy to find genes involved in biosynthetic pathways. This technique has been successful in locating genes in multiple biosynthetic pathways such as the triterpene marneral (Field et al., 2011) and in finding cytochrome P450s (P450s) involved in diterpene production (King et al., 2014). There are however cases where biosynthetic pathways are split into two stable clusters, such as the steroidal glycoalkaloids in tomato (*Solanum lycopersicum*) and potato (*Solanum tuberosum*) (Itkin et al., 2013) and the DIBOA pathway in wheat (*Triticum aestivum*) and rye (*Secale cereale*) (Nomura et al., 2002, 2003).

1.4 Avenacin A-1 is an antifungal triterpene glycoside made by oats

Avenacins are antifungal triterpene glycosides synthesised in the roots of oat (*Avena*) species and the closely related *Arrhenatherum elatius* (Turner, 1953; Crombie and Crombie, 1986). They are preformed phytoprotectants that confer resistance against the soil-borne fungus *Gaeumannomyces graminis* var. *tritici*, which causes the agriculturally important root-infecting disease "take-all" in crops such as wheat or barley (Papadopoulou et al., 1999).

Symptoms of take-all include stunting, nutrient deficiency and root rot (Kwak and Weller, 2013). In 2006 it was estimated that 50% of UK wheat crops suffered from the disease with average yield losses of 5-20% (Home-Grown Cereals Association, 2006). There is no effective method to control take-all, but available strategies, none of which are robust, involve the use of fungicides; the establishment of "take-all decline" which results in reduced disease severity due to the build up of antagonistic rhizosphere microorganisms over several years of continuous wheat cropping; and the use of break crops with attempts to reduce carry over in the soil (Home-Grown Cereals Association, 2006; Kwak and Weller, 2013). As wheat and other susceptible cereals are genetically isolated from oats, avenacin production cannot be introduced using traditional breeding techniques and genetic engineering is the most promising approach to develop take-all resistance in wheat or barley (Osbourn et al., 1994; Louveau, 2013).

There are four types of avenacin in oat roots, of which the most abundant and

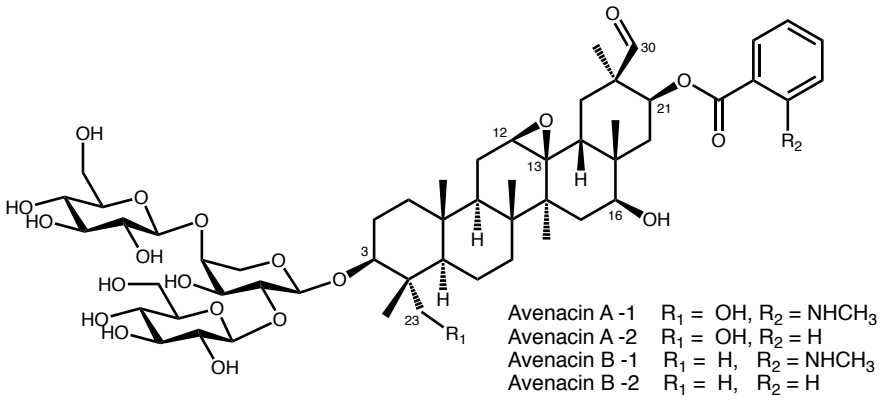


Figure 1.5: Structure of avenacins

fungitoxic is avenacin A-1 (Figure 1.5) (Crombie and Crombie, 1986). Avenacins have a β-amyrin backbone with a C-12,C-13 epoxide and oxidations at the C-16 and C-30 positions, and an additional oxidation at the C-23 position in the case of avenacins A-1 and A-2. They have a trisaccharide sugar chain of two branching D-glucose molecules with β-1,4- and β-1,2 linkages to a 3-*O*-α-L-arabinose. The avenacin scaffold is acylated at the C-21 position with benzoic acid (avenacin A-2 and B-2) or *N*-methyl-anthranilic acid (avenacin A-1 and B-1).

Multiple elements of the avenacin A-1 structure are required for its antifungal activity. The *N*-methylanthranilate ester avenacins A-1 and B-1 are more fungicidal than the benzoate esters (Crombie and Crombie, 1986) and the C-12,C-13 epoxide group has been shown to be critical for antifungal activity against *G. graminis* var. *tritici* (Geisler et al., 2013). An avenacin-resistant variety of the take-all causing fungus, *G. graminis* var. *avenae*, is able to infect oats by secreting a glycoside hydrolase enzyme, avenacinase, which deglucosylates avenacin A-1 to less toxic forms by removing one or more sugars (Turner, 1961; Crombie et al., 1986b; Osbourn et al., 1991; Bowyer et al., 1995). Avenacin A-1 is thought to permeabilise fungal membranes by forming complexes with sterols, leading to pore formation. The complete sugar chain is known to be crucial for this process (Armah et al., 1999).

The *N*-methyl-anthranilic acid ester of avenacin A-1 and B-1 confers a strong blue fluorescence to these molecules under UV light. Avenacins are accumulated in the epidermal cell layers of oat roots which can be seen to fluoresce strongly under UV illumination (Figure 1.6) (Osbourn et al., 1994; Qi et al., 2006).

As the major fluorescent compound in oat roots, reduction in levels of avenacin A-1 is clearly visible as reduced fluorescence in the roots of young oats (Osbourn et al.,

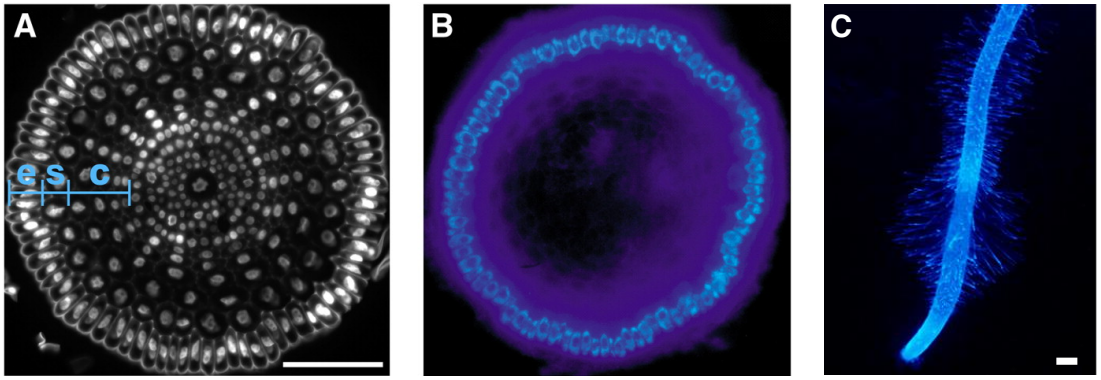


Figure 1.6: Avenacin A-1 is specifically accumulated in the epidermal cells of *A. strigosa* oat root tips. **(A)** Cross section of an oat root tip with nuclei stained with 4′,6-diamidino-2-phenylindole (DAPI) showing the different cell types (e, epidermis; s, subepidermis; c, cortex). Bar = 50 µm. **(B)** Fluorescence of avenacin A-1 in the root tip epidermal cells. **(C)** Fluorescence of avenacin A-1 in the root tip. Adapted from Wegel et al, 2009 and Kemen et al, 2014

1994). This property was used in a screen to identify genetic loci involved in the avenacin A-1 biosynthetic pathway. Sodium azide mutants of the diploid oat species *Avena strigosa* (accession S75) were screened for reduced root fluorescence, indicating avenacin deficiency (Papadopoulou et al., 1999). This identified ten distinct saponin-deficient (*sad*) mutants which all showed reduced resistance to the take-all pathogen *G. graminis* var. *tritici* (Papadopoulou et al., 1999). These loci were found to be tightly genetically clustered, with the exception of *Sad4* which is unlinked, and *Sad3* which is less tightly linked (genetic distance 3.6 cM) (Papadopoulou et al., 1999; Qi et al., 2004).

Many advances have been made in characterising the genes of the avenacin A-1 biosynthetic pathway (Figure 1.7). The screening of an *A. strigosa* root tip-specific cDNA library for oxidosqualene cyclase genes identified the β-amyrin synthase, AsbAS1, which corresponded to the *Sad1* complementation group (Haralampidis et al., 2001). This enzyme cyclises 2,3-oxidosqualene to β-amyrin, and is the first committed step in the avenacin A-1 biosynthetic pathway. Sequencing of the avenacin cluster region identified the gene encoding AsCyp51H10 (SAD2), a cytochrome P450 that both hydroxylates the β-amyrin scaffold at the C-16 position and adds the C-12,C-13 epoxide (Qi et al., 2006; Geisler et al., 2013). Phylogenetic analyses suggest that both of these enzymes have been co-opted from primary metabolism (from sterol biosynthesis) by duplication and rapid divergence (Qi et al., 2004; Geisler et al., 2013).

Expanding the sequenced region surrounding the avenacin cluster by sequencing of a bacterial artificial chromosome (BAC) contig identified genes encoding a further

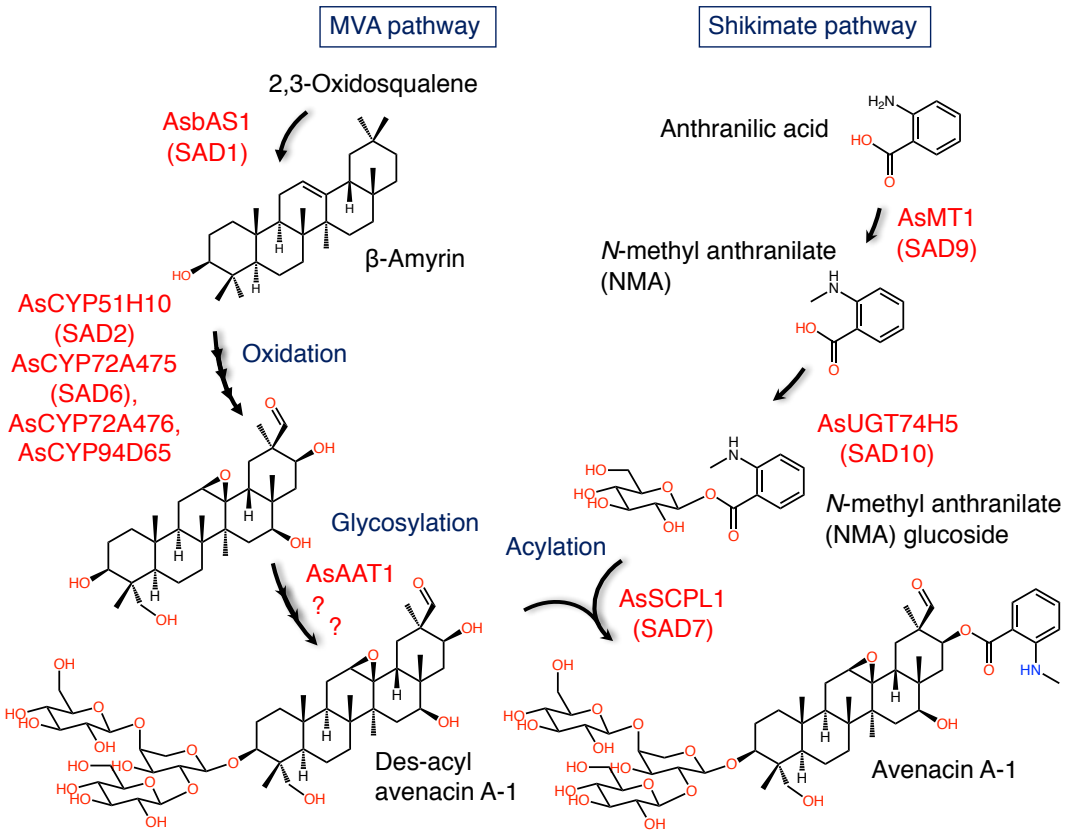


Figure 1.7: The avenacin biosynthetic pathway.
β-Amyrin synthase AsbAS1 (SAD1) cyclises 2,3-oxidosqualene to form β-amyrin, which is oxidised by four cytochromes P450, AsCYP51H10 (SAD2), AsCYP72A475 (SAD6), AsCYP72A476 and AsCYP94D65 (Haralampidis et al., 2001; Qi et al., 2006, Reed, 2016; Leveau et al., manuscript in preparation). The oxidised product is then glycosylated at the C-3 carbon position by an arabinosyltransferase, AsAAT1 (Louveau et al., manuscript in preparation) and two D-glucose molecules are added at the L-arabinose 2-*O* and 4-*O* positions by unknown glucosyltransferases.
Anthranilic acid generated by the shikimate pathway is methylated by the methyltransferase AsMT1 (SAD9) and glucosylated in the cytoplasm by AsUGT74H5 (SAD10) to give *N*-methyl anthranilate (NMA) glucoside (Owatworakit et al., 2013; Mugford et al., 2013). NMA glucoside acts as the activated acyl donor for the serine carboxypeptidase-like acyltransferase AsSCPL1 (SAD7), which acylates des-acyl avenacin A-1 at the C-21 carbon position to form avenacin A-1 (Mugford et al., 2009).

three enzymes in the pathway involved in acylation at the C-21 position. The serine carboxypeptidase-like acyltransferase, AsSCPL1 (SAD7), uses *O*-glucose esters as activated acyl donors (Mugford et al., 2009). The activated acyl donor used in the formation of avenacin A-1 and B-1, *N*-methyl anthranoloyl-*O*-glucose, is produced from the shikimate pathway intermediate anthranilate by the methyltransferase AsMT1 (SAD9) and a Group L UGT AsUGT74H5 (SAD10) (Owatworakit et al., 2013; Mugford et al., 2013). The three cytochrome P450s that catalyse the oxidations at the C21,

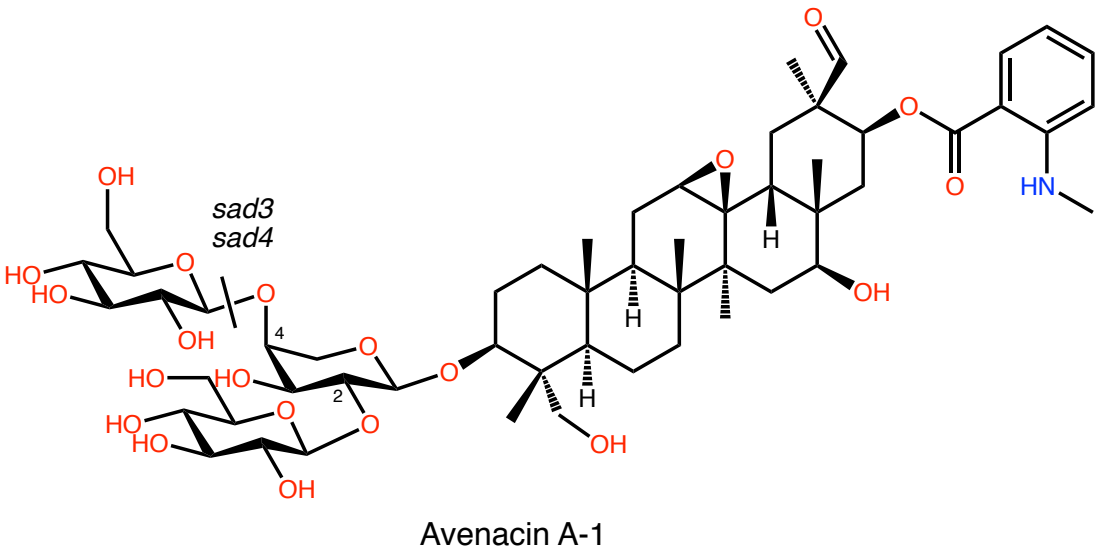


Figure 1.8: *sad3* and *sad4* mutants accumulate avenacin A-1 lacking the β-1,4-linked glucose (Papadopoulou et al., 1999; Mylona et al., 2008).

C23 and the C30 positions (AsCYP72A475 (SAD6), AsCYP94D65 and AsCYP72A476, respectively) and the arabinosyltransferase that catalyses the first step of the trisaccharide sugar chain, AsAAT1 (AsUGT99D1), were also found to be genetically linked to the characterised avenacin cluster genes (Reed, 2016; Leveau et al., manuscript in preparation; Louveau et al., manuscript in preparation). These nine genes are all co-expressed specifically in root tips, showing tight transcriptional regulation of the avenacin biosynthetic cluster (Haralampidis et al., 2001; Qi et al., 2006; Mugford et al., 2009; Wegel et al., 2009; Mugford et al., 2013; Owatworakit et al., 2013; Reed, 2016; Leveau et al., manuscript in preparation; Louveau et al., manuscript in preparation).

The final steps of the avenacin A-1 biosynthetic pathway, the addition of the two branching D-glucose molecules of the saccharide chain, are as yet unknown. Two of the *Sad* loci identified, *Sad3* and *Sad4*, have been implicated in the glycosylation of avenacin A-1. The roots of *sad3* and *sad4* mutants accumulate an avenacin intermediate lacking the β-1,4-D-glucose, mono-deglucosyl avenacin A-1 (Figure 1.8) (Papadopoulou et al., 1999; Mylona et al., 2008). Unique amongst the *sad* mutants, these both appear to have developmental defects, with short roots, membrane trafficking defects and root hair deficiency. This was attributed to the accumulation of mono-deglucosyl avenacin A-1 as a toxic intermediate, as double mutants for the *Sad1* locus, which encodes the first committed step in avenacin biosynthesis (*sad1/sad1 sad3/sad3* and *sad1/sad1 sad4/sad4*), had normal root morphology albeit with reduced root fluo-

rescence (Mylona et al., 2008).

The *sad4* mutant is only partially compromised in avenacin A-1 glycosylation, as *sad4* roots also accumulate avenacin A-1 (Mylona et al., 2008; Papadopoulou et al., 1999). *sad4* mutants are additionally defective in the glycosylation of leaf-specific steroidal glycosides, avenacosides A and B. These mutants accumulate reduced amounts of avenacoside B but have normal levels of avenacoside A (Papadopoulou et al., 1999). Avenacoside B has an additional D-glucose compared to avenacoside A, suggesting that *Sad4* has a more general role in specialised metabolite glucosylation such as transport and is not dedicated to the avenacin biosynthetic pathway (Mylona et al., 2008; Papadopoulou et al., 1999). Supporting this, the *Sad4* locus is unlinked to the avenacin cluster (Qi et al., 2004). *Sad3* may be more directly involved in the biosynthesis of avenacin A-1, as *sad3* mutants are unable to make any fully glycosylated avenacins and the locus is linked to the avenacin cluster (Papadopoulou et al., 1999; Qi et al., 2004).

1.5 Thesis Aims

The aim of this thesis is to identify the two glucosyltransferase enzymes of *A. strigosa* that are involved in the biosynthesis of avenacin A-1, an antifungal triterpene glycoside that confers resistance in oats to the agriculturally important crop root disease, 'take-all'.

The characterisation of these enzymes would contribute to the knowledge of plant triterpene glycosyltransferases, of which few have been identified. In addition, this would complete the characterisation of the avenacin biosynthetic pathway and allow its genetic manipulation into susceptible crop species.

Chapter 2

General materials and methods

Specific materials and methods are detailed within each chapter.

2.1 Biological materials

2.1.1 Bacterial strains

The bacterial strains used are listed in Table 2.1.

Table 2.1: Bacterial strains

Strain	Antibiotic selection
E. coli,DH5α (Invitrogen)	-
E. coli,BL21 (Invitrogen)	-
E. coli,BL21 Rosetta (Invitrogen)	Chloramphenicol (35 µg/ml)
A. tumefaciens GV3101	Gentamycin (25 µg/ml), rifampicin (50 µg/ml)
A. tumefaciens LBA4404	Streptomycin (100 µg/ml), rifampicin (50 µg/ml)

2.1.2 Fungal strains

Fungal strains used were *Gaeumannomyces graminis* var. *tritici* strain T5 and *G. graminis* var. *avenae* strain A3 (Bryan et al., 1999).

2.1.3 Plant material

Avena sativa seeds for avenacin purification were kindly provided by Rachel Melton. All other oat plants used in this study are *Avena strigosa* accession S75 (from the Institute of Grasslands and Environmental Research, Aberystwyth, Wales, United Kingdom). Saponin-deficient mutant lines of *A. strigosa* S75 were derived by sodium azide mutagenesis (Papadopoulou et al., 1999).

Nicotiana benthamiana plants were grown under greenhouse conditions (25°C with 16 hour lighting). Seedlings were sown in F1 compost (Levington) and grown for two weeks before transfer to individual cells in trays containing F2 compost (Levington). Plants were grown for an additional 3-4 weeks before infiltration.

2.2 Primers

Primers were designed using Primer3plus (Rozen and Skaletsky, 2000). Primers used are listed in Section 2.11 and in individual materials and methods sections of chapters.

2.3 Sequencing

Sequencing was carried out by Eurofins (http://www.eurofinsgenomics.eu/) and GATC Biotech (https://www.gatc-biotech.com/en/index.html).

2.4 DNA electrophoresis

Agarose gels were prepared by the addition of the appropriate amount of agarose (Sigma) to TAE buffer (Formedium) with approximately 0.01% ethidium bromide (1 mg/ml solution). A 2-log DNA ladder (NEB) or 1 kb DNA ladder (NEB) were included. Gels were run at constant 110 V for approximately 20 minutes and visualised under UV.

2.5 SDS-PAGE gel electrophoresis

Protein electrophoresis was carried out using the mini-cell system (XCell SureLock Mini-Cell Electrophoresis System, Invitrogen). Protein samples were prepared in 1 x LDS loading buffer (NuPAGE LDS Sample Buffer, Invitrogen) and heated for 10 minutes at 95°C. SDS-PAGE gels (NuPAGE 4-12% Bis-Tris, Invitrogen) were loaded with 10 µl of samples and 5 µl of a protein ladder (SeeBlue Pre-Stained Standard, Invitrogen). Gels were run at 200 volts for 30-50 minutes and stained with Coomassie solution (InstantBlue, Expedeon) for 20-30 minutes.

2.6 Plasmids

Table 2.2: Plasmids

Name	Supplier/Origin	Use	Selection
pDONR207	Invitrogen	Gateway Entry	Gentamycin (25 µg/ml)
pH9-GW	O'Maille group, Gateway compatible version of pET28	Gateway Destination vector; Expression in *E. coli* with 9x His N-terminal tag	Kanamycin (50 µg/ml)
pEAQ-*HT*-DEST1	Lomonossoff lab, JIC	Gateway Destination vector; Transient expression in *N. benthamiana*	Kanamycin (50 µg/ml)
pMDC45	(Curtis and Grossniklaus, 2003)	Gateway Destination vector; Transient expression in *N. benthamiana* with N-terminal GFP tag	Kanamycin (50 µg/ml)
pMDC83	(Curtis and Grossniklaus, 2003)	Gateway Destination vector; Transient expression in *N. benthamiana* with C-terminal GFP and 6-His tag	Kanamycin (50 µg/ml)
pB7RWG2	(Karimi et al., 2002)	Gateway Destination vector; Transient expression in *N. benthamiana* with C-terminal RFP	Spectinamycin (100 µg/ml) (Streptomycin (100 µg/ml) for *E. coli*)
35S:mRFP	(Moglia et al., 2014)	35S-driven mRFP control vector derived from pB7RWG2.0	Spectinamycin (100 µg/ml) (Streptomycin (100 µg/ml) for *E. coli*)
ER-rk CD3-959	(Nelson et al., 2007)	ER mCherry marker	Kanamycin (50 µg/ml)
G-rk CD3-967	(Nelson et al., 2007)	Golgi mCherry marker	Kanamycin (50 µg/ml)
pEAQ-*HT*-DEST1-AsUGT91	This study		Kanamycin (50 µg/ml)
pEAQ-*HT*-DEST1-AsTG	This study		Kanamycin (50 µg/ml)
pEAQ-*HT*-DEST1-NOSIG-AsTG	This study		Kanamycin (50 µg/ml)
35S:GFP	This study	35S-driven GFP control vector derived from pMDC83	Kanamycin (50 µg/ml)
pMDC45-AsUGT91	This study		Kanamycin (50 µg/ml)
pMDC83-AsUGT91	This study		Kanamycin (50 µg/ml)
pMDC83-AsTG	This study		Kanamycin (50 µg/ml)
pB7RWG2-SIG-ONLY	This study		Spectinamycin (100 µg/ml) (Streptomycin (100 µg/ml) for *E. coli*)
pB7RWG2-AsTG	This study		Spectinamycin (100 µg/ml) (Streptomycin (100 µg/ml) for *E. coli*)
pB7RWG2-NOSIG-AsTG	This study		Spectinamycin (100 µg/ml) (Streptomycin (100 µg/ml) for *E. coli*)

2.7 Growth of *A. strigosa* seedlings

Manually-dehusked seeds were surface-sterilised in 0.5% sodium hypochlorite solution for 5 minutes, rinsed thoroughly (10 x volume) with sterile water and dried under sterile conditions. Seeds were put onto moist sterile filter paper or distilled water agar in Petri dishes with the radicle facing the bottom of the plate. Petri dishes were sealed (with PARAFILM® for seeds on filter paper or with Micropore^{TM} tape for seeds on water agar) and kept in the dark for two days at 4°C . Petri dishes were incubated at a 45°angle for 3-7 days at 22°C in a growth cabinet (SANYO, Versatile Environmental Test Chamber) with a light-dark cycle of 16 hours of light and 8 hours of dark.

2.8 Oat genomic DNA purification

Frozen oat leaf tissue (1.5 cm^3) from five-day-old seedlings (Section 2.7) was ground in liquid nitrogen with an autoclaved pestle and mortar. Ground tissue was resuspended in 1.2 ml of extraction buffer (0.2 M Tris-HCl pH 8, 250 mM NaCl, 25 mM ethylenediaminetetraacetic acid (EDTA), 1% sodium dodecyl sulfate (SDS, Sigma-Aldrich)) and centrifuged at 13 000 x g for 5 minutes. The supernatant (4 x 375 µl) was removed to fresh tubes. An equal volume of isopropanol was added and mixed by inversion, then tubes were centrifuged at 13 000 x g for 10 minutes. The supernatant was removed and the pellets were washed with 600 µl of 70% ethanol and centrifuged at 13 000 x g for 10 minutes. The supernatant was removed and the pellets were dried at 35°C for 20 minutes in a vacuum dryer. Pellets were resuspended in 200 µl water and stored at -20 °C. Aliquots (2 µl) of the above gDNA preparations were used in Gateway PCR reactions (Section 2.11).

2.9 RNA extraction from oat seedlings and cDNA synthesis

After three days of growth (Section 2.7) one set of *A. strigosa* S75 seedlings was used to harvest whole roots and young leaves and a second set was used to harvest root tips (last 0.5 mm) and elongation zones (~0.5 mm section between the root tip and the start of the root hairs). Tissues were harvested with a sterile razor blade on a glass plate over dry ice, with the harvested tissues placed immediately into 1.5 ml Eppendorf tubes in dry ice. Harvested tissues were stored at -80 °C.

Frozen oat tissues were ground in liquid nitrogen with an autoclaved mortar and pestle. RNA extraction and clean-up was performed using an RNeasy Plant Mini

kit (Qiagen) according to the manufacturer's instructions. DNA contaminants were removed by digestion with DNase treatment (Promega) and the treated samples were cleaned up using the RNeasy Plant Mini kit (Qiagen) protocol. RNA concentration was determined using a NanoDrop 8000 spectrophotometer (Thermo Scientific) and the the A_{260}/A_{280} ratio verified. RNA integrity was checked by separation on a 1% agarose gel at 110 V for 15 minutes. RNA samples were stored at -80 °C before cDNA synthesis.

RNA (5 µg) for each tissue was used for cDNA synthesis was using the SuperScript-III ®First-Strand Synthesis System (Invitrogen) using the supplied oligo(dT) primers. cDNA libraries were stored at -20 °C.

2.10 RT-PCR profiling

cDNA libraries from *A. strigosa* tissues (Section 2.9) were diluted to 100 ng/µl. Fragment amplication of constitutively expressed glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) was used to validate whether cDNA concentrations were normalised across tissues. PCR mixes were prepared as Table 2.3 and RT-PCR cycling conditions were as Table 2.4. PCR products (10 µl) were analysed by DNA electrophoresis (Section 2.4) on a 1.5% agarose gel at 110 V for 17 minutes.

Table 2.3: PCR mix components for RT-PCR

Component	Volume	Final concentration
cDNA from *A. strigosa* tissue (100 ng/µl)	1 µl	5 ng/µl
5 x Green GoTaq®Flexi buffer	4 µl	1x
$MgCl_2$ solution (25 mM)	2 µl	2.5 mM
dNTP mix (Promega) (10 mM each dATP, dCTP, dGTP, dTTP)	0.5 µl	0.25 mM
primers (10 µM each upstream and downstream)	1 µl	0.5 µM
GoTaq®G2 Flexi DNA polymerase (5 u/µl)	0.2 µl	1 u
sterile water	11.3 µl	

Table 2.4: PCR conditions for RT-PCR

Step	Duration	Temperature	Cycles
Denaturation	2 min	95 °C	x 1
Denaturation	30 sec	95 °C	
Annealing	30 sec	variable*	x variable*
Elongation	45 sec	72 °C	
Elongation	5 min	72 °C	x 1

*see Section 3.3.4 for annealing temperatures and number of cycles used.

2.11 Gateway cloning

2.11.1 Two-step Gateway cloning

For two-step Gateway cloning, an initial amplification with gene-specific primers containing 12 nucleotides of the attB sites was used (Tables 2.6 and 2.7) followed by a second amplification using attB adapters (Tables 2.5, 2.8 and 2.9) to result in PCR products flanked by the full attB sites. PCR products were analysed by gel electrophoresis (section 2.4) and the PCR fragment length estimated by comparison with a 1 kb DNA ladder (New England Biolabs). PCR products were purified with a PCR Cleanup kit (Qiagen) for further cloning (Section 2.11.3).

2.11.2 One-step Gateway cloning

For one-step Gateway cloning, the full attB adapter sites are included in the gene-specific primers, requiring only one amplification step (Tables 2.10 and 2.11). PCR products were analysed by gel electrophoresis (section 2.4) and the PCR fragment length estimated by comparison with a 1 kb DNA ladder (New England Biolabs). PCR products were purified with a PCR Cleanup kit (Qiagen) for further cloning (Section 2.11.3).

Table 2.5: General Gateway primers

Name	Sequence
attB1F-1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTA
attB2R-1	GGGGACCACTTTGTACAAGAAAGCTGGGTA
attLF-1	TCGCGTTAACGCTAGCATGGATCTC
attLR-2	GTAACATCAGAGATTTTGAGACAC

Table 2.6: PCR mix components for Gateway PCR step 1

Component	Volume
DNA template	1-2 µl
5 x iProof GC buffer	4 µl
dNTP mix (Promega) (10 mM each dATP, dCTP, dGTP, dTTP)	0.4 µl
gene-specific primers (10 µM each upstream and downstream)	1 µl
iProof (Promega) DNA polymerase (2 u/µl)	0.2 µl
sterile water	to 20 µl

Table 2.7: PCR conditions for Gateway PCR step 1

Step	Duration	Temperature	Cycles
Denaturation	2 min	98 °C	x 1
Denaturation	10 sec	98 °C	
Annealing	30 sec	60 °C	x 18
Elongation	1 min 20 sec	72 °C	

Table 2.8: PCR mix components for Gateway PCR step 2

Component	Volume
Gateway PCR step 1	10 µl
5 x iProof GC buffer	8 µl
dNTP mix (Promega) (10 mM each dATP, dCTP, dGTP, dTTP)	0.9 µl
attB1F-1 primer (10 µM)	2 µl
attB2R-1 primer (10 µM)	2 µl
iProof (Promega) DNA polymerase (2 u/µl)	0.4 µl
sterile water	26.7 µl

Table 2.9: PCR conditions for Gateway PCR step 2

Step	Duration	Temperature	Cycles
Denaturation	2 min	98 °C	x 1
Denaturation	10 sec	98 °C	
Annealing	30 sec	50 °C	x 25
Elongation	1 min 20 sec	72 °C	
Elongation	5 min	72 °C	x 1

2.11.3 Gateway entry clones

pDONR207 (100 ng) and purified PCR products (50 ng) were added to 1 µl Gateway®BP Clonase™ II enzyme mix (Invitrogen) and made up to 5 µl total volume with Tris-EDTA (10 mM Tris-HCl pH 8; 1 mM EDTA) buffer. Reactions were incubated at room temperature for 2 hours and 2.5 µl of the reaction mixture was used to transform 10 µl chemically competant *E. coli* as detailed in Section 2.14. Colony

Table 2.10: PCR mix components for one-step Gateway PCR

Component	Volume
plasmid template (5 ng/µl)	1 µl
5 x iProof GC buffer	10 µl
dNTP mix (Promega) (10 mM each dATP, dCTP, dGTP, dTTP)	1 µl
upstream primer (10 µM)	2.5 µl
downstream primer (10 µM)	2.5 µl
iProof (Promega) DNA polymerase (2 u/µl)	0.2 µl
sterile water	32.8 µl

Table 2.11: PCR conditions for one-step Gateway PCR

Step	Duration	Temperature	Cycles
Denaturation	2 min	98 °C	x 1
Denaturation	10 sec	98 °C	
Annealing	30 sec	65 °C	x 30
Elongation	1 min 20 sec	72 °C	
Elongation	5 min	72 °C	x 1

PCR (section 2.12) with attL primers (Table 2.5) was used to screen colonies for inserts of the correct size. Plasmids from positive colonies were extracted (Section 2.13) and sent for sequencing (Section 2.3) with attLF-1/R-2 primers (Table 2.5) and custom primers as required.

2.11.4 Gateway expression clones

The pDONR207-derived entry clone (100 ng) (Section 2.11.3) and the destination vector (100 ng) were added to to 1 µl Gateway®LR Clonase™ II enzyme mix (Invitrogen) and made up to 5 µl total volume with Tris-EDTA (10 mM Tris-HCl pH 8; 1 mM EDTA) buffer. Reactions were incubated at room temperature for 2 hours and 2.5 µl of the reaction mixture was used to transform 10 µl chemically competant *E. coli* as detailed in Section 2.14. Colony PCR (section 2.12) with attB primers (Table 2.5) was used to screen colonies for inserts of the correct size. Plasmids from positive colonies were extracted (Section 2.13). On occasion, for validation of correct cloning, plasmids were sent for sequencing (Section 2.3) with attBF-1/R-2 primers (Table 2.5) or custom primers as required.

2.12 Colony PCR

Single colonies were picked with a sterile toothpick and diluted into 30 µl of sterile water, and heated at 95°C for 5-10 minutes. Cooled cell samples were centrifuged briefly, and 2 µL were used as the DNA template in PCR reactions (Tables 2.12 and 2.13). Colonies were additionally streaked onto fresh LB agar plates with the appropriate antibiotic and incubated at 37°C overnight for subsequent cloning steps (Section 2.13).

Table 2.12: PCR mix components for colony PCR

Component	Volume	Final concentration
template	2 µl	-
5 x Green GoTaq®Flexi buffer	4 µl	1x
$MgCl_2$ solution (25 mM)	2 µl	2.5 mM
dNTP mix (Promega) (10 mM each dATP, dCTP, dGTP, dTTP)	0.5 µl	0.25 mM
upstream primer (10 µM)	1 µl	0.5 µM
downstream primer (10 µM)	1 µl	0.5 µM
GoTaq®G2 Flexi DNA polymerase (5 u/µl)	0.2 µl	1 u
sterile water	9.3 µl	

Table 2.13: PCR conditions for colony PCR

Step	Duration	Temperature	Cycles
Denaturation	2 min	95 °C	x 1
Denaturation	30 sec	95 °C	
Annealing	30 sec	50 °C	x 30
Elongation	1min/kb	72 °C	
Elongation	5 min	72 °C	x 1

2.13 Plasmid extraction and purification

Positive colonies from colony PCR (Section 2.12) were cultured overnight at 37°C and 200 rpm in 10 ml LB media with the appropriate antibiotic (Table 2.2). Plasmids were extracted using QIAprep Spin Miniprep Kit (Qiagen) following supplier instructions. DNA concentrations were quantified by Nanodrop (NanoDrop 8000 spectrophotometer, Thermo Scientific).

2.14 Transformation of chemically-competent *E. coli*

Chemically competent *E.coli* were stored at -80°C. Frozen aliquots of *E.coli* were allowed to thaw on ice, and incubated with DNA plasmids for 30 minutes on ice before heat shock at 42°C in a water bath for 45 seconds. Cells were placed on ice for 2 minutes before adding 90 µl of LB media. Cells were incubated for 1 hour at 37°C at 200 rpm and 50 µl of the cells were plated onto LB agar plates with antibiotic selection (Table 2.2) and incubated at 37°C overnight.

2.15 Transformation of *A. tumefaciens* strain LBA4404

Aliquots (50 µl) of chemically competent *A. tumefaciens* strain LBA4404 stocks (kindly provided by Dr James Reed) were thawed on ice. Cells were incubated for 30 minutes on ice with 1-200 ng of the pEAQ-*HT*-DEST1 destination vector containing the gene of interest. The tubes were transferred to liquid nitrogen for 30 seconds before removal and thawing at room temperature. LB (1 ml) was added and the cells were incubated at 28°C at 200 rpm for 3-4 hours. Cells were plated on LB agar plates with rifampicin (50 µg/ml), kanamycin (50 µg/ml) and streptomycin (100 µg/ml) and incubated at 28°C for 3 days until colonies could be seen. Single colonies were picked to make glycerol stocks (Section 2.17).

2.16 Transformation of *A. tumefaciens* strain GV3101

Aliquots (30 µl) of chemically competent *A. tumefaciens* strain GV3101 stocks (kindly provided by Dr Aymeric Leveau) were thawed on ice. Cells were incubated on ice for 30 minutes with 0.5 µl of plasmid DNA and transferred to a chilled electroporation cuvette. Cells were electroporated at 2.2 kV and placed back on ice briefly before adding 600 µl of LB media. Cells were incubated at 28°C and 200 rpm for 1

hour before 100 µl l was plated on LB agar plates containing gentamycin (25 µg/ml), rifampicin (50 µg/ml) and plasmid-specific antibiotics (Table 2.2). Plates were incubated at 28 °C for 1-2 days until colonies could be seen. Single colonies were picked to make glycerol stocks (Section 2.17).

2.17 Preparation of glycerol stocks

Single bacterial colonies were used to inoculate 10 ml LB with antibiotic selection and were grown overnight at 200 rpm (37°C for *E. coli* and 28°C for *A. tumefaciens* strains). Cultures were mixed with an equal volume of sterile 40% glycerol and stored at -80°C.

2.18 Cell culture and protein production for *in vitro* assays

Transformed *E. coli* Rosetta cells were grown overnight in 10 ml LB with appropriate antibiotics at 37°C and 200 rpm. The overnight culture (1 ml) was used to inoculate 100 ml LB with antibiotics and grown at 37°C, 200 rpm until the cultures reached an OD_{600} between 0.5-0.6. The cultures were acclimatised for 30 minutes at 16°C at 200 rpm, then 0.05 mM isopropyl-β-D-thiogalactopyranoside (IPTG) (Sigma-Aldrich) was added. Cells were grown for 4-5 hours (for AsTG) or overnight (for UGTs) and harvested by centrifugation at 3220 x g for 10 minutes. The supernatant was discarded and the cell pellets stored at -80°C.

The frozen cell pellets were allowed to thaw on ice and were resuspended in 6 ml of chilled sonication buffer (300 mM NaCl, 50 mM Tris-HCl pH 7.8, 20 mM imidazole, 5% glycerol, cOmplete^{TM} EDTA-free protease inhibitor cocktail (Roche) (1 tablet per 50 ml sonication buffer), 0.1% Tween 20 (Sigma- Aldrich)).

Resuspended cells were sonicated with a benchtop ultrasonic disintegrator (Soniprep 150 plus, MSE) in ice water for 5 x 10 seconds (amplitude = 7.0) with 20 seconds rest. Cell lysates were centrifuged at 12 000 x g for 20 minutes at 4°C. Supernatants were incubated with 150 pre-equilibrated µl Ni-NTA Agarose beads (Roche) with agitation at 4°C for 1 hour. Beads were transferred to 1.5 ml Eppendorf tubes and washed 3 times with 500 µl filtered Buffer A (300mM NaCl, 50mM Tris-HCl pH 7.8, 20mM imidazole, 5% glycerol).

Proteins were eluted with 3 x 200 µl Buffer B (300mM NaCl, 50mM Tris-HCl pH 8, 500 mM imidazole, 5% glycerol). Protein elution fractions were combined, and the buffer exchanged by adding 2 x 2 ml 50 mM Tris-HCl pH 7.5 and concentrating

in Amicon®Ultra-4 Centrifugal Filter Units with Ultracel-10 membranes (Merck) at 3220 x g, with a final volume of approximately 250 µl.

Protein concentration was estimated using a Bradford assay (Bio-Rad Protein Assay Kit I, with bovine γ-globulin standard) as per manufacturer's instructions and the protein purity assessed by SDS-PAGE (section 2.5).

Enzyme assays were carried out as detailed in Material and Methods sections of Chapters 4 and 5 (Sections 4.5.5 and 5.5.5).

The analysis of reaction products was carried out by reverse phase HPLC using a 50x2.1mm 2.6 µ Kinetex XB-C18 column (Phenomenex) with a column oven temperature of 30°C. Detection was by UV/Vis absorbance (Shimadzu SPD-M20A), collecting spectra from 200-600nm. Electrospray MS data (Shimadzu LC-2020 dual source MS) were collected by electrospray in positive mode and negative mode from *m/z* 50-1500. The gradient was run at 0.3 ml/min with 100% water as Buffer A, 100% acetonitrile as Buffer B and was as follows: 25% Buffer B from 0-0.6 minutes; 25-80% Buffer B from 0.6-7 minutes; 80-100% Buffer B from 7-7.2 minutes; a linear gradient between 7.2-8 minutes; 100 to 25% Buffer B from 8-8.1 minutes, and held at 25% Buffer B until 10 minutes.

2.19 Transient expression in *N. benthamiana*

Strains of *Agrobacterium tumefaciens* LBA4404 containing pEAQ-*HT*-DEST1 plasmids expressing HMGR, AsbAS1, AsCYP51H10 or AsCYP72A475 were provided by Dr James Reed. An *A. tumefaciens* LBA4404 strain containing an AsAAT1-expressing pEAQ-*HT*-DEST1 construct was provided by Dr Thomas Louveau. Strains of *Agrobacterium tumefaciens* GV3101 containing pEAQ-*HT*-DEST1 plasmids expressing HMGR, AsbAS1, AsCYP51H10, AsCYP72A475, AsAAT1 were provided by Dr Aymeric Leveau.

Transient expression in *N. benthamiana* was carried out based on the method as described in Sainsbury et al. (2012). Transformed *A. tumefaciens* glycerol stocks were streaked out onto fresh LB agar plates with antibiotics (Tables 2.1 and 2.2) and grown overnight at 28°C. Colonies were picked and grown in 10 ml or 50 ml LB with antibiotics (Tables 2.1 and 2.2) at 28°C and 200 rpm for 15-18 hours. Cultures were pelleted by centrifugation at 3220 x g for 12 minutes and the supernatant discarded. The cells were resuspended in 10 ml (or 20 ml for 50 ml cultures) of MMA solution (10 mM MES-KOH, 10 mM $MgCl_2$, 150 µM acetosyringone) and incubated at room temperature in the dark for one hour. The concentration of the cell cultures was determined

by measuring the optical density at 600 nm of culture dilutions in a spectrophotometer. For metabolite analysis experiments, a strain containing green fluorescent protein (GFP) in pEAQ-*HT*-DEST1 (Lomonossoff group; Reed, 2016) was used as a control to maintain the same concentration of bacteria in all infiltration combinations. For subcellular localisation assays, the pEAQ-*HT*-DEST1-GFP strain was substituted with MMA solution, and combinations were co-infiltrated with a strain containing the suppressor of silencing P19 (in plasmid pEAQ-*HT*-DEST1-AsUGT91). Culture densities were adjusted with MMA solution to a final OD_{600} of n x c, where n = the number of bacterial strains to be combined and c is between 0.1 and 0.2, adjusted so that the final OD_{600} did not exceed 1. Diluted cell suspensions of individual strains were combined in equal ratios. The bacterial suspension mixes were hand-infiltrated into the underside of 5-week-old *N. benthamiana* leaves using a needle-less syringe. Plants were grown under the same greenhouse conditions following infiltration.

2.20 Analysis of metabolites from *N. benthamiana* leaves

Five days after infiltration (Section 2.19), *N. benthamiana* leaves were harvested, frozen at -80°C, and freeze-dried. Dried leaves were stored at -80°C. Leaf samples (10 mg) were measured into 2 ml screw-cap tubes. Two tungsten beads (3mm) were added, and leaf tissues were disrupted at 1000 rpm for 2 x 30 seconds in a Geno/Grinder SPEX Sample Prep 2010. After brief centrifugation, 550 µl of 80% MeOH with 20 µM digitoxin standard (Merck) was added. Samples were agitated at 1400 rpm at 18 °C for 20 minutes, then centrifuged at 20 000 x g at 4°C for 2 minutes. 400 µl of the supernatent was removed to a fresh pre-chilled 1.5 ml Eppendorf tube on ice, and samples were partitioned twice with hexane. Aqueous fractions were dried in a Genevac EZ-2 Elite centrifugal evaporator at a maximum temperature of 30°C and stored at -80°C.

For high-performance liquid chromatography, samples were resuspended in 75 µl methanol and filtered through Corning®Costar®Spin-X®centrifuge tube filters (Sigma-Aldrich). The filtrate (50 µl) was combined with 50 µl 50% MeOH and 10 µl was analysed by reverse phase HPLC using a 50x2.1mm 2.6 µ Kinetex XB-C18 column (Phenomenex) and detection by CAD or CAD-MS (as below).

For HPLC-CAD-MS: The column oven temperature was set at 30°C and detection was by charged aerosol detector (CAD, Corona Ultra RS from Dionex), as well as electrospray MS (Shimadzu LC-2020 dual source MS) collected in positive mode and negative mode from *m/z* 50 -1500. The gradient was run at 0.3 ml/min with 100%

water as Buffer A, 100% acetonitrile as Buffer B and was as follows: 15% Buffer B from 0-1.5 minutes; 15-60% Buffer B from 1.5-26 minutes; 60-100% Buffer B from 26-26.5 minutes; a linear gradient between 26.5-28.5 minutes; 100 to 15% Buffer B from 28.5-29 minutes, and held at 15% Buffer B until 30 minutes.

For HPLC-CAD: The column oven temperature was set at 25°C and detection was by charged aerosol detector (CAD, Corona Ultra RS from Dionex). The gradient was run at 0.3 ml/min with 100% water as Buffer A, 100% acetonitrile as Buffer B and was as follows: 10% Buffer B from 0-1.5 minutes; 10-50% Buffer B from 1.5-21 minutes; 50-95% Buffer B from 21-21.5 minutes; a linear gradient between 21.5-23.5 minutes; 95 to 10% Buffer B from 23.5-24 minutes, and held at 10% Buffer B until 25 minutes.

Chapter 3

Searching for *A. strigosa* glucosyltransferase candidates

Avenacin A-1 is an antifungal triterpene glycoside that protects oat species from the important wheat root pathogen, *Gaeumannomyces graminis* var. *tritici* (Papadopoulou et al., 1999). The biosynthesis of the avenacin A-1 aglycone involves multiple steps including several oxidations and the addition of an acyl group, all of which have been characterised in the diploid oat species *Avena strigosa* accession number S75 (Haralampidis et al., 2001; Qi et al., 2006; Mugford et al., 2009, 2013; Owatworakit et al., 2013; Reed, 2016; Leveau et al., manuscript in preparation; Louveau et al., manuscript in preparation).

The sugar chain of avenacin A-1 is a trisaccharide of an L-arabinose linked in the α-configuration to the 3-*O* of the avenacin backbone with two branching β-1,4- and β-1,2-D-glucose molecules. The arabinosyltransferase that initiates the avenacin saccharide chain has been recently characterised as AsAAT1 (UGT99D1) (Louveau et al., manuscript in preparation), however the steps involved in adding the two branching D-glucoses of the avenacin sugar chain have not been elucidated.

All previously characterised plant enzymes that glycosylate triterpene scaffolds are uridine diphosphate sugar-dependent glycosyltransferases (UGTs) belonging to glycosyltransferase family 1, an enzyme family that glycosylates a vast range of plant specialised metabolites (Chapter 1, Table 1.1). There are likely to be over 100 UGTs in the *A. strigosa* genome (Louveau, 2013) making it difficult to identify the missing glucosyltransferases of the avenacin A-1 biosynthetic pathway. As discussed in Chapter 1, candidates can be prioritised based on homology to characterised UGTs with similar activities to the activity in question; co-expression with other genes of the biosynthetic

pathway; and genetic linkage to characterised biosynthetic pathway genes.

Characterised glycoside-specific UGTs (GGTs) that extend sugar chains of plant specialised metabolites are listed in Table 3.1. The majority of these GGTs are specific in the sugar-sugar linkage that they create, and do not glycosylate sugars at multiple positions (with the exception of SgUGT94-289-3 and SgUGT94-289-2 from *Siraitia grosvenorii* which have been described as catalysing both 1,2- and 1,6-linkages) (Table 3.1). The *A. strigosa Sad3* and *Sad4* loci are both required for the addition of the β-1,4-linked D-glucose, although it is not known if these correspond UGTs (Papadopoulou et al., 1999; Mylona et al., 2008). The addition of the β-1,2-linked D-glucose must involve another UGT, and it is therefore anticipated that characterising the glucosylation steps of avenacin A-1 biosynthesis will involve two different UGTs.

Nine triterpene GGTs have previously been characterised (highlighted in blue in Table 3.1) (Shibuya et al., 2010; Sayama et al., 2012; Jung et al., 2014; Itkin et al., 2016; Xu et al., 2016a). Five of these GGTs (GmUGT91H4 from *Glycine max*, PgUGT94Q2 from *Panax ginseng* and SgUGT94-289-1, SgUGT94-289-2 and SgUGT94-289-3 from *Siraitia grosvenorii*) are in plant GT1 phylogenetic Group A. Group A UGTs appear to be dedicated GGTs and catalyse the formation of β-1,2 and β-1,6 sugar linkages. The other four triterpene GGTs (GmUGT73F2, GmUGT73F4, and GmUGT73P2 from *Glycine max* and GuUGAT from *Glycyrrhiza uralensis*) are in Group D and form β-1,2 and β-1,3 sugar linkages. GGTs that form β-1,4-sugar linkages are not well characterised, with only one example, SlGAME17 from *Solanum lycopersicum*, which is a steroidal alkaloid 3-*O*-galactoside β-1,4-D-glucosyltransferase in Group O (Itkin et al., 2013).

Whilst the link between UGT phylogeny and activity is generally weak (Osmani et al., 2009), Groups A, D and O are good candidate phylogenetic groups to find potential triterpene GGTs that form β-1,2 and β-1,4-linkages. All of the characterised avenacin pathway genes are clustered in the genome and co-regulated. Sequencing of the *A. strigosa* S75 genome is currently underway (Bin Han, Chinese Academy of Sciences) and will elucidate the genomic region surrounding the avenacin cluster. However, this information was not available at the time of this work.

Table 3.1: Characterised GT Family 1 UGTs with glycoside-specific glycosyltransferase activity

Enzyme name	UGT family	Group	Activity	Plant species	Publication
AtUGT79B1	UGT79	A	Anthocyanidin 3-*O*-glucoside 1,2-xylosyltransferase	*Arabidopsis thaliana*	Yonekura-Sakakibara et al. (2012)
AtUGT79B6	UGT79	A	Flavonol 3-*O*-galactoside 1,2-glucosyltransferase	*Arabidopsis thaliana*	Yonekura-Sakakibara et al. (2014)
BpUGT94B1	UGT94	A	Anthocyanidin 3-*O*-glucoside 1,2-glucuronosyltransferase	*Bellis perennis*	Sawada et al. (2005)
CaUGT3	UGT94	A	Flavonol 3-*O*-glucoside 1,6-glucosyltransferase (processive)	*Catharanthus roseus*	Masada et al. (2009)
Cm1-2RhaT1	UGT94	A	Flavonol 7-*O*-glucoside 1,2-rhamnosyltransferase	*Citrus maxima*	Frydman et al. (2013)
Cs1-6RhaT	UGT91	A	Flavonol 7-O/3-O glucoside 1,6-rhamnosyltransferase	*Citrus sinensis*	Frydman et al. (2013)
GjUGT94E5	UGT94	A	Apocarotenoid glucoside 1,6-glucosyltransferase	*Gardenia jasminoides*	Nagatoshi et al. (2012)
GmUGT79A6	UGT79	A	Flavonol 3-*O*-glucoside/galactoside 1,6-rhamnosyltransferase	*Glycine max*	Rojas Rodas et al. (2014)
GmUGT79A7	UGT79	A	Flavonol 3-*O*-glucoside/galactoside 1,6-glucosyltransferase	*Glycine max*	Rojas Rodas et al. (2016)
GmUGT79B30	UGT79	A	Flavonol 3-*O*-glucoside/galactoside 1,2-glucosyltransferase	*Glycine max*	Di et al. (2015)
GmUGT91H4	**UGT91**	**A**	**Triterpene 3-*O*-galactoside 1,2-rhamnosyltransferase**	*Glycine max*	Shibuya et al. (2010)
LeABRT2	UGT79	A	Flavonol 3-*O*-glucoside 1,6-rhamnosyltransferase	*Lobelia erinus*	Hsu et al. (2017)
LeABRT4	UGT79	A	Flavonol 3-*O*-glucoside 1,6-rhamnosyltransferase	*Lobelia erinus*	Hsu et al. (2017)
PgUGT94Q2	**UGT94**	**A**	**Triterpene 3-*O*-glucoside 1,2-glucosyltransferase**	*Panax ginseng*	Jung et al. (2014)
PhA3G1-6RhaT	UGT79	A	Anthocyanidin 3-*O*-glucoside 1,6-rhamnosyltransferase	*Petunia hybrida*	Brugliera et al. (1994)
In3GGT	UGT91	A	Anthocyanidin 3-*O*-glucoside 1,2-glucosyltransferase	*Ipomoea nil*	Morita et al. (2005)
SiUGT94D1	UGT94	A	Lignan 2'-*O*-glucoside 1,6-glucosyltransferase	*Sesamum indicum*	Noguchi et al. (2008)
SgUGT94-289-3	**UGT94**	**A**	**Triterpene 24-*O*-glucoside/3-*O*-glucoside 1,2-/1,6-glucosyltransferase**	*Siraitia grosvenorii*	Itkin et al. (2016)
SgUGT94-289-2	**UGT94**	**A**	**Triterpene 24-*O*-glucoside 1,2-/1,6-glucosyltransferase**	*Siraitia grosvenorii*	Itkin et al. (2016)
SgUGT94-289-1	**UGT94**	**A**	**Triterpene 24-*O*-glucoside/3-*O*-glucoside 1,6-glucosyltransferase**	*Siraitia grosvenorii*	Itkin et al. (2016)
SlGAME18	UGT94	A	Steroidal alkaloid 3-*O*-glucoside 1,2-glucosyltransferase	*Solanum lycopersicum*	Itkin et al. (2013)
VpUGT94F1	UGT94	A	Flavonol 3-*O*-glucoside 1,2-glucosyltransferase	*Veronica persica*	Ono et al. (2010)
GmUGT73F2	**UGT73**	**D**	**Triterpene 22-*O*-arabinoside 1,3-glucosyltransferase**	*Glycine max*	Sayama et al. (2012)
GmUGT73F4	**UGT73**	**D**	**Triterpene 22-*O*-arabinoside 1,3-xylosyltransferase**	*Glycine max*	Sayama et al. (2012)
GmUGT73P2	**UGT73**	**D**	**Triterpene 3-*O*-glucuronide 1,2-galactosyltransferase**	*Glycine max*	Shibuya et al. (2010)
GuUGAT	**UGT73**	**D**	**Triterpene 3-*O*glucoronosyltransferase/ Triterpene 3-*O*-glucuronide 1,2-glucuronosyltransferase**	*Glycyrrhiza uralensis*	Xu et al. (2016a)
SlUGT73L4	UGT73	D	Steroidal alkaloid 3-*O*-glucoside 1,3-xylosyltransferase	*Solanum lycopersicum*	Itkin et al. (2013)
StSGT3	UGT73	D	Steroidal alkaloid 3-*O*-glucoside/galactoside 1,2-rhamnosyltransferase	*Solanum tuberosum*	McCue et al. (2007)
CsUGT707B1	UGT707	E	Flavonol 3-*O*-glucoside 1,2-glucosyltransferase	*Crocus sativus*	Trapero et al. (2012)
PdUGT85A19	UGT85	G	Cyanohydrin glucoside 1,6-glucosyltransferase	*Prunus dulcis*	Franks et al. (2008)
SrUGT76G1	UGT76	H	Diterpene 13-*O*-glucoside 1,3-glucosyltransferase	*Stevia rebaudiana*	Richman et al. (2005)
SlGAME17	UGT93	O	Steroidal alkaloid 3-*O*-galactoside 1,4-glucosyltransferase	*Solanum lycopersicum*	Itkin et al. (2013)

Triterpene glycoside-specific UGTs are indicated in blue.

3.1 Aims

The aim of the work presented in this chapter was to identify and prioritise candidate *A. strigosa* S75 triterpene GGTs based on primary sequence identity to UGT groups (Groups A, D and O) and co-expression with avenacin biosynthetic genes.

3.2 Results and Discussion

Two independent transcriptome datasets were available as sources of candidate avenacin GGTs:

1. An *A. strigosa* S75 root tip (terminal 0.5 cm) transcriptome library generated using 454 titanium chemistry, in which all transcripts of the previously characterised avenacin biosynthetic genes are represented (Kemen et al., 2014; Reed, 2016; Louveau, 2013).

2. An RNA-seq dataset of six *A. strigosa* S75 tissues (whole root, root tip (terminal 0.5 mm), leaf, panicle, shoot and spikelet) generated in a collaboration with Bin Han, Chinese Academy of Sciences.

3.2.1 Mining the 454 transcriptome database for candidate UGTs

Previously, preliminary mining of the *A. strigosa* S75 root tip 454 transcriptome resource identified 110 UGT-like sequences of which 53 corresponded to full-length UGTs (Louveau, 2013). Recently, a draft version (70% coverage) of the *A. strigosa* S75 genome became available as part of the ongoing *A. strigosa* S75 sequencing project in the group of Bin Han (CAS). Although the draft genome sequence was not of high enough assembly quality to enable the avenacin gene cluster to be extended, the sequence information allowed the validation of contig sequences and the retrieval of full-length sequences where these were not available in the 454 transcriptome database.

In this work, the 454 transcriptome database was searched for UGT sequences by a BLAST (tBLASTn) search using full-length sequences of representative UGTs from every plant UGT subfamily (A-N) (Materials and Methods, Table 3.5). Consistent with the previous analysis, this identified 110 UGT-like sequences (Supplementary Table S1). These sequences were then verified by a BLAST (BLASTn) search against the draft version of the *A. strigosa* S75 genome to resolve chimeric sequences and incomplete transcripts. Where full-length sequences were not available from the 454 database, full-length sequences were predicted using the genescan tool (http://genes.mit.edu/GENSCAN.html, (Burge and Karlin, 1997)). This resulted in 93 unique UGT-like sequences which include the previously characterised avenacin UGTs, AsAAT1 (the avenacin arabinosyltransferase) (Louveau et al., manuscript in preparation), and SAD10 (AsUGT74H5) and its related homologue AsUGT74H7 (which are required for the generation of the acyl glucose donor used by the avenacin

acyltransferase SAD7) (Mugford et al., 2009; Owatworakit et al., 2013; Mugford et al., 2013) (Table 3.2).

These UGTs were named following the same nomenclature as in Louveau (2013). UGTs identified in the 454 sequencing project are named by the contig number and preceded by "AsGT" for *A. strigosa* S75 glycosyltransferases (Louveau, 2013). Twenty-six *A. strigosa* enzymes were identified in an earlier expressed sequence tag (EST) database and these enzymes are named by the accession number from the original EST collection (Haralampidis et al., 2001). The resulting UGT names and their corresponding contig numbers can be found in Supplementary Table S2.

These UGTs were assigned to families by a neighbor-joining phylogenetic analysis of the protein sequences along with functionally characterised UGT protein sequences from other plant species. (Figure 3.1). This revealed five UGTs in Group A, seventeen in Group D and seven in Group O (Table 3.2).

Table 3.2: UGTs from the 454 transcriptome dataset assigned to plant GH1 phylogenetic groups.

UGT	Group	UGT	Group	UGT	Group
AsGT01194	A	AsGT06751	E	AsGT4h2	H
AsGT01332	A	AsGT08700	E	AsGT03883	I
AsGT18280	A	AsGT10772	E	AsGT11637	I
AsGT19358	A	AsGT16496	E	AsGT18535	I
AsGT18279	A	AsGT17930	E	AsGT23002	I
AsGT00243	B	AsGT01315	E	AsGT24951	I
AsGT22027	B	AsGT28947A	E	AsGT17328	I
AsGT00733	C	AsGT28947B	E	AsGT26962	J
AsGT01461	C	AsGT20n10	E	AsGT10326	K
AsGT02436	D	AsGT01092	E	AsGT15a11	L
AsGT16f23	D	AsGT05602	E	AsGT01989	L
AsGT12842	D	AsGT06492	E	AsGT02699	L
AsGT14h21	D	AsGT10189	E	SAD10	L
AsAAT1	D	AsGT25n16	E	AsGT07784	L
AsGT000892-1	D	AsGT17673	E	AsGT01599	L
AsGT000892-2	D	AsGt21862	E	AsGT03158	L
AsGT27f7	D	AsGT23340	E	AsGT10759	L
AsGT23781	D	AsGT23586	E	AsGT28b19	L
AsGT24i2	D	AsGT23818	E	AsGT17576	L
AsGT10433	D	AsGT27586	E	AsGT11140	L
AsGT11099	D	AsGT00260	G	AsGT15275	L
AsGT23141	D	AsGT02132	G	AsGT1a15	L
AsGT14h20	D	AsGT05740	G	AsGT01341	O
AsGT11i11	D	AsGT8i4	G	AsGT05827	O
AsGT3i21	D	AsGT16h6	G	AsGT18257	O
AsGT03295	D	AsGT18035	G	AsGT27a12	O
AsGT16525	E	AsGT26167	G	AsGT22388	O
AsGT22538	E	AsGT10811	G	AsGT24248	O
AsGT27009	E	AsGT01670	H	AsGT24a3	O
AsGT28561	E	AsGT04347	H	AsGT06218	P
AsGT00931	E	AsGT04598	H	AsGT12o13	P

Groups A, D and O are highlighted.

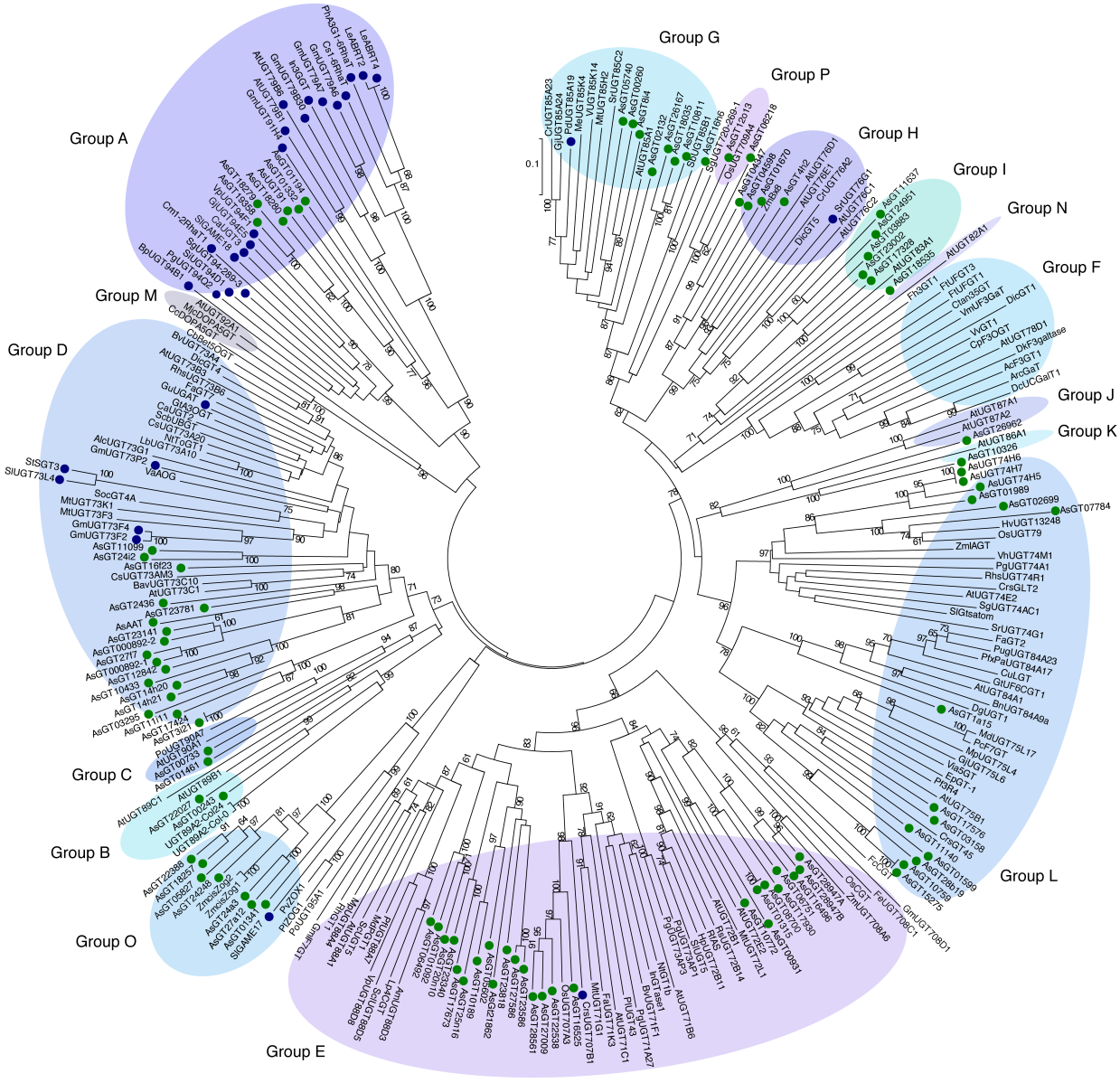


Figure 3.1: Phylogenetic analysis of *A. strigosa* UGT sequences from the 454 transcriptome dataset.
Phylogenetic analysis of *A. strigosa* UGT candidates (green circles) (listed in Supplementary section 6) with characterised UGTs from other plant species (listed in Table S3). Functionally characterised glycoside glycosyltransferases (GGTs) are indicated (blue circles).
The UGT phylogenetic groups (Groups A-N) are labelled as described in (Ross et al., 2001; Caputi et al., 2012). Protein sequences were aligned using MAFFT (https://mafft.cbrc.jp/alignment/software/) and the evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987) and analysed with 1000 bootstrap replicates which are shown at the nodes. The branches with <60% support were not marked. The unrooted tree was constructed using MEGA7 (Kumar et al., 2016).

3.2.2 Mining RNA-seq data for candidate UGTs

An RNA-seq database of six *A. strigosa* S75 tissues (whole root, root tip (terminal 0.5 mm), leaf, panicle, shoot and spikelet) was created as part of a collaboration with the group of Bin Han (CAS).

To identify possible UGT candidates from this transcriptomic dataset, the gene expression profiles were clustered using a self-organising map (SOM) (Wehrens and Buydens, 2007; Kohonen, 2013). Self-organising maps cluster datasets using an unsupervised and therefore unbiased method and have successfully identifying genes involved in plant specialised metabolic pathways (Payne et al., 2017).

The SOM output can be represented on a two-dimensional hexagonal grid of map nodes, where genes with the same or similar expression patterns are mapped to the same node or a neighbouring node. The characterised avenacin biosynthetic pathway genes (coloured circles) can be seen to cluster together in two nodes in a SOM of the *A. strigosa* S75 RNA-seq dataset (Figure 3.2).

The random presentation of data during the training of a SOM can lead to different mapping results if the training is repeated (Wehrens and Buydens, 2007). Therefore, the SOM training was repeated 100 times, and genes that mapped to the same node as any of the avenacin biosynthetic genes were extracted from each SOM run. This resulted in a list of 1090 contigs, 918 of which corresponded to annotated genes in the draft oat genome sequence (Supplementary Table S4). Within this list were 45 genes that were annotated as UGTs (Table 3.3).

The majority of these UGTs correspond to genes identified in the screen of the 454 transcriptome database, including two in Group A, eleven in Group D and four in Group O (Table 3.3). There were eight additional genes in this analysis, marked as N/A in Table 3.3. These genes may not have been represented in the 454 transcriptome database due to their low expression levels, as they had reads per kilobase of exon model per million reads (RPKM) values of less than 12. The additional genes in Groups A and D were not considered as good candidates for avenacin glucosylation as they had RPKM values below 5. An exception was AS01_006890_0018824 in Group A (named AsUGT91), which had a RPKM value in the root tip of 22.5.

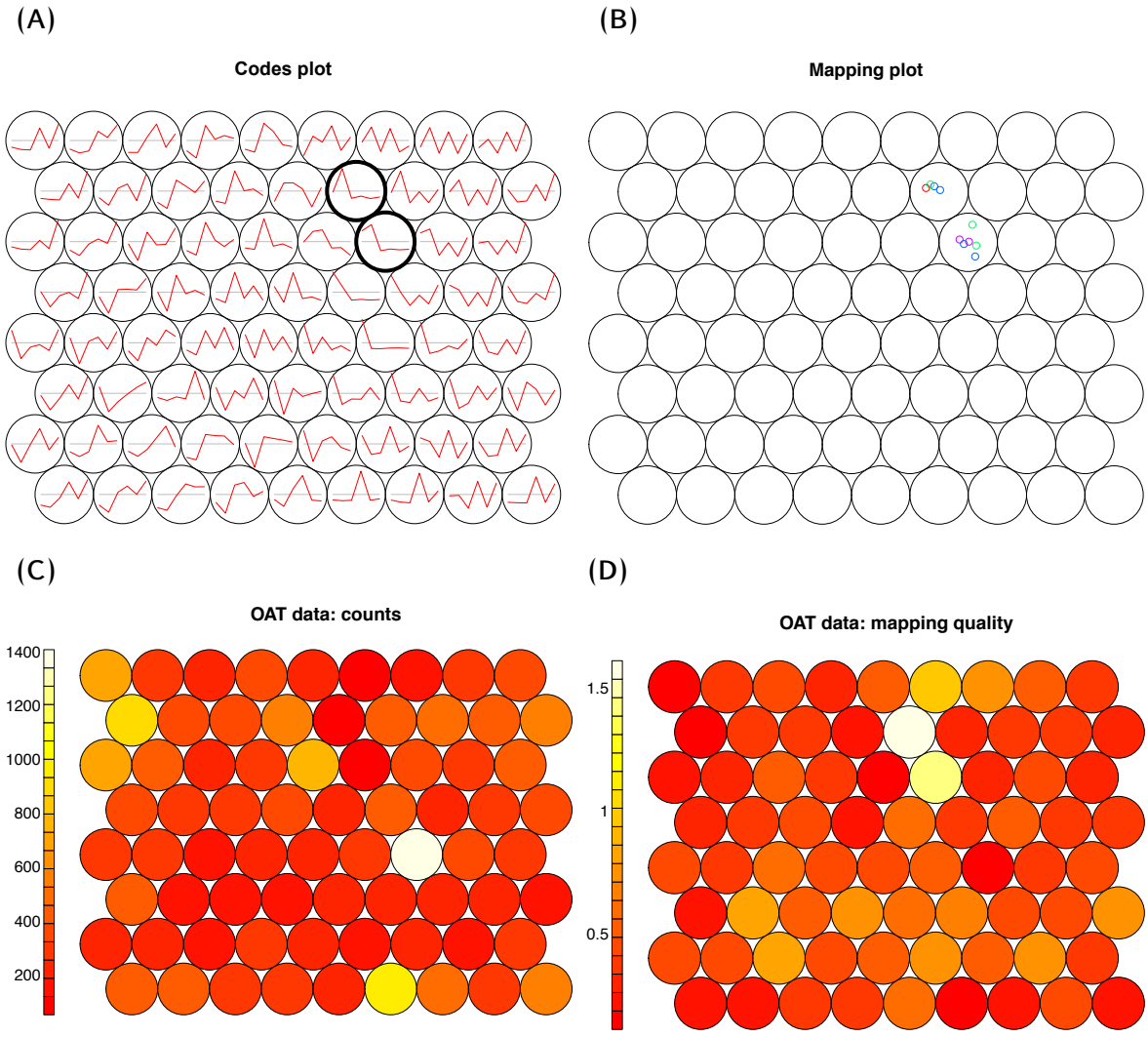


Figure 3.2: Self-organising map of the *A. strigosa* RNA-seq dataset. The characterised avenacin biosynthetic pathway genes cluster together in two good quality nodes of the SOM. **(A)** The expression profile in the six tissues (root, root tip, leaf, panicle, shoot and spikelet) associated with the node is plotted within each node. Nodes containing avenacin biosynthetic genes are highlighted in bold. **(B)** The characterised avenacin biosynthetic pathway genes are indicated as follows: AsbAS1, red circle; AsCYP51H10, AsCYP94D65, AsCYP72A475, AsCYP72A476, blue circles; AsUGT99D1, AsUGT74H5, AsUGT74H7, green circles; AsSCPL1, AsMT1, magenta circles. **(C)** The colour of the units indicates the number of genes that are mapped to each node. The genes are reasonably spread out over the grid. **(D)** The unit colour shows the mean distance of gene expression profiles, mapped to a particular unit, to the expression profile of that unit. Small distances across the map indicates good mapping (Wehrens and Buydens, 2007).

Table 3.3: UDP glycosyltransferases identified in the self-organising map analysis

Present in avenacin cluster nodes	Gene	Group	UGT from 454 transcriptome	Maximum RPKM value
100%	AS01_002796_0048708	A	N/A	3.1
20%	AS01_002796_0056524	A	AsGT18280	20.9
100%	AS01_003718_0016409	A	AsGT01332	19.6
100%	AS01_006890_0018824	A	N/A	22.5
100%	AS01_000040_0343206	D	AsGT14h20	47.5
100%	AS01_000040_0580230	D	AsGT14h21	17.8
100%	AS01_000232_0183840	D	AsGT02436	13.7
100%	AS01_001384_0023180	D	N/A	3.1
100%	AS01_002760_0003285	D	AsGT23781	33.5
100%	AS01_003018_0175304	D	AsGT3i21	5.1
100%	AS01_003827_0329694	D	AsAAT	61.5
100%	AS01_008603_0020929	D	AsGT11i11	62.9
100%	AS01_009157_0340895	D	AsGT24i2	25.8
100%	AS01_011433_0183300	D	AsGT12842	18.3
100%	AS01_015816_0099017	D	AsGT16f23	92.0
100%	AS01_018433_0019113	D	AsGT27f7	29.6
20%	AS01_001240_0004470	E	AsGT01092	12.4
100%	AS01_009625_0484622	E	AsGT25n16	115.4
89%	AS01_009997_0231261	E	AsGT27586	74.1
100%	AS01_010493_0015704	E	AsGT20n10	12.9
94%	AS01_003251_0510695	G	AsGT00260	21.9
48%	AS01_003866_0515441	G	AsGT18035	4.4
100%	AS01_012751_0126886	G	AsGT16h6	108.2
100%	AS01_009625_0051477	H	AsGT4h2	10.8
100%	AS01_002110_0512100	I	AsGT11637	4.7
38%	AS01_004898_0191830	I	AsGT18535	20.2
100%	AS01_006545_0031520	I	N/A	11.3
100%	AS01_014560_0017746	I	AsGT24951	14.5
100%	AS01_017784_0074517	I	AsGT17328	17.7
100%	AS01_002774_0633711	J	N/A	5.2
39%	AS01_000525_0200999	L	AsGT1a15	22.0
100%	AS01_001959_0203229	L	AsGT17576	6.5
100%	AS01_003827_0182130	L	AsUGT75H5	75.0
100%	AS01_003827_0587827	L	AsUGT74H7	93.6
100%	AS01_003944_0076156	L	AsGT10759	9.7
100%	AS01_005173_0292209	L	N/A	2.6
100%	AS01_007006_0066560	L	N/A	2.4
100%	AS01_007207_0097636	L	AsGT01989	112.1
100%	AS01_009143_0044829	L	N/A	5.0
57%	AS01_015483_0023201	L	AsGT28b19	4.7
88%	AS01_000207_0282310	O	AsGT05827	12.1
100%	AS01_000815_0222774	O	AsGT24a3	11.8
20%	AS01_002123_0323801	O	AsGT24248	5.3
100%	AS01_009625_0058710	O	AsGT27a12	7.0
4%	AS01_011449_0125658	P	AsGT12o13	12.8

N/A, not present in 454 transcriptome database.

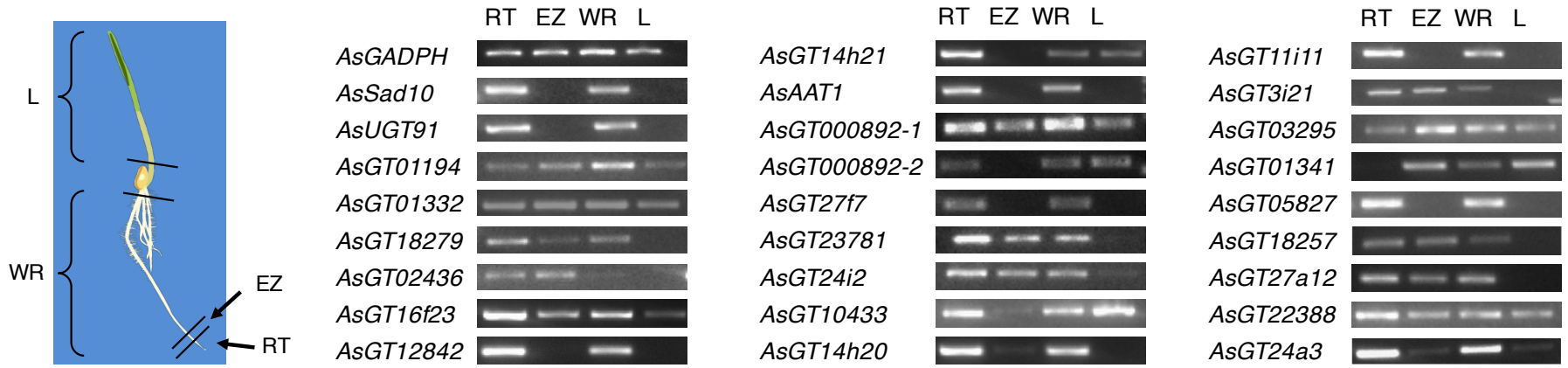


Figure 3.3: Analysis of the expression profile of candidate UGTs.
RT-PCR expression profile of candidate UGTs in Groups A, D and O. The profile of the previously characterised AsUGT74H5 (*Sad10*) gene and the housekeeping oat glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) gene were included as positive controls. RNA was extracted from 3-day-old *Avena strigosa* plants. The oat tissues used are illustrated: root tips (RT), elongation zones (EZ), whole roots (WR) and young leaves (L).

3.2.3 Validation of candidate UGT expression profiles by RT-PCR

The expression profiles of UGTs in Group A, D and O identified in the 454 transcriptome database and the RNA-seq dataset were validated by RT-PCR (Figure 3.3).

This showed seven genes (*AsUGT91*, *AsGT12842*, *AsGT27f7*, *AsGT14h20*, *AsGT11i11*, *AsGT05827* and *AsGT24a3*) to have similar root tip-specific expression patterns to the avenacin biosynthetic pathway UGTs, *Sad10* and *AsAAT1*, in the tissues tested (Figure 3.3). These genes were all identified as being co-regulated with the avenacin biosynthetic pathway genes in the SOM analysis (clustering with avenacin biosynthetic genes in more than 85% of SOM runs), suggesting these genes as candidates for avenacin glucosyltransferase activity.

3.2.4 Protein sequence features of candidate UGTs in Groups A, D and O

An alignment of full-length protein sequences of UGT candidates in Groups A, D and O shows that the two residues proposed to form a His-Asp-acceptor catalytic triad (Wang, 2009) are conserved in all sequences except for AsGT02436, which has a glycine instead of the aspartate residue (data not shown). This does not rule out that all of the candidate proteins are potentially active, as the promiscuous flavonoid glucosyltransferase UGT73G1 from onion (*Allium cepa*) also has a glycine at this position in multiple sequence alignments (Kramer et al., 2003).

Analysis of the Plant Secondary Product Glycosyltransferase (PSPG) motif of the candidates shows that the Group D enzyme, AsGT12842, has a histidine at the 44th PSPG position (Figure 3.4). The final PSPG residue (Q) is often replaced with a histidine (H) in UGTs that transfer D-galactose or L-arabinose, such as the avenacin arabinosyltransferase AsAAT1 (Figure 3.4) (Kubo et al., 2004; Louveau et al., manuscript in preparation). As AsGT12842 may not have glucosyltransferase activity, this enzyme was considered to be a weaker candidate for avenacin glucosyltransferase activity.

The Group A enzyme AsUGT91 has an insertion in the PSPG motif (Figure 3.4). However, the active arabinosyltransferase AsAAT1 also has an insertion of two amino acids in the same region, suggesting that AsUGT91 may still be active.

```
AsGT01341     WAPQLEILAHGATAAFVSHCGWNSLLEGLGHGKPILAWPMH-C--DQ
AsGT27a12     WAPQLEILAHRATAAFMSHCGWNSTMESLSHGKPILAWPMH-C--DQ
AsGT24a3      WAPQLEILAHGATAAFMSHCGWNSTVESLSHGKPILAWPMH-C--DQ
AsGT05827     WAPQLEILAHGATAAFMSHCGWNSTMESLSHGKPILAWPMH-S--DQ
AsGT24248     WAPQLEILAHGATAAFMSHCGWNSTLESLSHGKPMLAWPMH-S--DQ
AsGT18257     WAPQLEILAHGATAAFMSHCGWNSTMESLSHGKPILAWPMH-S--DQ
AsGT22388     WAPQLEILAHGATAAFMSHCGWNSTMESLSHGKPILAWPMH-S--DQ
AsGT01194     WVPQVRFLAHASVGGFLTHAGWNSVAEGLAQGVRLVLLPLV-F--DQ
AsGT01332     WVPQVRVLAHGAVGAFLTHCGWGSTVESFRFGHALVMLPFV-A--DQ
AsUGT91       WVPQVSILAHGAVAAFLTHCGWSSTIEGLRFGRPLVMLPIATG--DQ
AsGT18280     WVPQISVLAHGAVAAFLTHCGWNSTIEGLLFGHPLIMLPIF-G--DQ
AsGT19358     WVPQMSILAHAAVGGFLTHCGRNSLIESLLFGHPLVMLPIF-G--DQ
AsGT18279     WLPQMSILAHAAVGGFLTHCGRNSLIEGLLFGHPLVMLPIF-G--DQ
AsGT03295     WAPQSEILSLSSIGAFLTHCGSSSLLEAAAAGVPMLTWPLV-F--DQ
AsGT11i11     WAPQISILGHHAAGAFVTQCGWNSVLETVAAAVPMLTWPLA-F--EQ
AsGT3i21      WAPQTAILGHPAVGAFVTHCGWNSVLETVAAGVPVLTWPMV-F--EQ
AsGT24i2      WAPQMLILNHRALGGFVTHCGWNSTLESVSAGVPMVTWPRF-A--DQ
AsGT11099     WAPQVLVLNHPSIGGFVTHCGWNSVLEAVSAGVPMVTWPRC-G--DQ
AsGT02436     WAPQVAILAHPAVGGFMTHCGWGSTLESVAAGVPMLTWPLF-A--EQ
AsGT12842     WAPQLAILSHRAVGGFLTHCGWNSLVESVTLGVPLLTWPHF-A--DH
AsGT10433     WAPQVAILSHPAVGGFVTHCGWNSLLESITHGVAVVTWPKF-A--DQ
AsGT14h21     WAPQLAILSHRAVGGFVTHCGWNSLLESVAHGVPVVTWPHS-G--DQ
AsGT14h20     WVSQLAILSHRAIGGFVTHCGWNSLLESVAHGVPVVTWPHF-G--DQ
AsGT000892-1  WAPQVTILSHPAVGGFLTHCGWNGTLEALSIGVPALTWPNI-A--DQ
AsGT27f7      WAPQVTILSHPAVGGFLTHCGWNGTLETLSLGVPTLTWPTI-A--DQ
AsGT000892-2  WAPQVTILSHPAVGGFLTHCGWNATLEAISHGVPALTWPSF-A--DQ
AsGT23141     WAPQVTILSHPAVGGFLTHCGWNATLETISHGVPALTWPNF-A--DQ
AsGT16f23     WAPQMMILWHQAIGGFMTHCGWNSTVEGICAGVPMITWPHF-A--EQ
AsAAT1        WVPQLLILSHAAVGCLFTHSGWNSVMEAITAGKPAVTWPRL-IGSDH
AsGT23781     WAPQALILSHRAAGAFVTHCGWNSTLEAVAAGLPVVTWPHF-T--DQ
```

Figure 3.4: Alignment of the Plant Secondary Product Glycosyltransferase (PSPG) motif of UGT candidates in groups A, D and O.
AsUGT91 and AsAAT1 have amino acid insertions in the PSPG motif (bold underline). AsAAT1 and AsGT12842 have a histidine as the final residue of the PSPG motif (highlighted in green).

3.2.5 Candidate UGTs for avenacin glucosyltransferase activity

Six UGT candidates for triterpene-3-*O*-α-L-arabinoside β-1,2- and 1,4-D-glucosyltransferase activity were prioritised for functional characterisation:

1. AsUGT91
2. AsGT27f7
3. AsGT14h20
4. AsGT11i11
5. AsGT05827
6. AsGT24a3

These candidates were chosen based on their root tip-specific expression profiles as assessed by RT-PCR; co-expression with avenacin cluster genes in the SOM analysis (clustered with avenacin biosynthetic genes in more than 85% of SOM runs); expression levels (RPKM values higher than ten); and primary structure features consistent with glucosyltranferase activity (summarised in Table 3.4).

Enzyme	Group	Gene	RT-PCR	SOM	Maximum RPKM value	PSPG
AsUGT91	A	AS01_006890_0018824	Root tip	100%	22.5	1 insertion (1 residue)
N/A	A	AS01_002796_0048708	n.d.	100%	3.1	
AsGT01194	A	AS01_012363_0009496	Root and leaf	0%	19.1	
AsGT01332	A	AS01_003718_0016409	Root and leaf	100%	19.6	
AsGT18280	A	AS01_002796_0056524	n.d.	20%	20.9	
AsGT19358	A	AS01_013835_0105701	n.d.	0%	5.2	
AsGT18279	A	AS01_002740_0084538	Root	0%	33.2	
AsGT02436	D	AS01_000232_0183840	Root	100%	13.7	
AsGT16f23	D	AS01_015816_0099017	Root and leaf	100%	92	
AsGT12842	D	AS01_011433_0183300	Root tip	100%	18.3	Possible arabinosyl/ galactosyltransferase
AsGT14h21	D	AS01_000040_0580230	Root and leaf	100%	17.8	
AsGT000892-1	D	AS01_000892_0475555	Root and leaf	0%	39.5	
AsGT000892-2	D	AS01_000892_0649070	Root and leaf	0%	32.8	
AsGT27f7	D	AS01_018433_0019113	Root tip	100%	29.6	
AsGT23781	D	AS01_002760_0003285	Root	100%	33.5	
AsGT24i2	D	AS01_009157_0340895	Root and leaf	100%	25.8	
AsGT10433	D	AS01_001764_0065978	Root and leaf	0%	74.9	
AsGT11099	D	AS01_003427_0107181	n.d.	0%	0	
AsGT23141	D	AS01_004255_0482187	n.d.	0%	9.5	
AsGT14h20	D	AS01_000040_0343206	Root tip	100%	47.5	
N/A	D	AS01_001384_0023180	n.d.	100%	3.1	
AsGT11i11	D	AS01_008603_0020929	Root tip	100%	62.9	
AsGT3i21	D	AS01_003018_0175304	Root	100%	5.1	
AsGT03295	D	AS01_001569_0429119	Root and leaf	0%	8.9	
AsGT01341	O	AS01_000815_0292318	Root and leaf	0%	46.1	
AsGT05827	O	AS01_000207_0282310	Root tip	88%	12.1	
AsGT18257	O	AS01_000815_0141040	Root	0%	1.4	
AsGT27a12	O	AS01_009625_0058710	Root	100%	7	
AsGT22388	O	AS01_000815_0187839	Root and leaf	0%	3.6	
AsGT24248	O	AS01_002123_0323801	n.d.	20%	5.3	
AsGT24a3	O	AS01_000815_0222774	Root tip	100%	11.8	

Table 3.4: Summary of shortlisted candidate UGTs in Groups A, D and O for avenacin glucosyltransferase activity.
Shortlisted candidates from UGT phylogenetic groups A, D and O are highlighted. Candidates were shortlisted based on: a similar RT-PCR expression profile to the avenacin biosynthetic genes which are root tip-specific; clustering with avenacin biosynthetic genes in a self-organising map (SOM) co-expression analysis; expression levels; and primary structural features. RT-PCR expression patterns are described as follows: Root tip, highest expression in root tip; Root: root-specific; Root and leaf, expressed in both roots and leaves; n.d.: not determined. Genes that were clustered less than 85% with the avenacin biosynthetic genes in the SOM analysis were not considered likely candidates. The maximum reads per kilobase of exon model per million reads (RPKM) for each gene in the RNA-seq dataset is indicated. Candidates with maximum RPKM values lower than 10 were not considered likely candidates.

3.3 Materials and Methods for Chapter 3

3.3.1 Mining the *A. strigosa* root tip 454 transcriptome database

Table 3.5: Table of UGTs used as queries in tBLASTn searches to mine a 454-based transcriptomic dataset from *A. strigosa* root tips.

UGT group	Enzyme	GenBank Protein accession	Organism
A	GjUGT94E5	BAM28984	*Gardenia jasminoides*
A	GmUGT91H4	BAI99585	*Glycine max*
A	AtUGT91B1	Q9LSM0	*Arabidopsis thaliana*
B	AtUGT89B1	NP_177529	*Arabidopsis thaliana*
C	PoUGT90A7	ACB56926	*Pilosella officinarum*
D	AtUGT73B3	AAM47999	*Arabidopsis thaliana*
E	AtUGT72B1	Q9M156	*Arabidopsis thaliana*
E	AtUGT88A1	AEE75831	*Arabidopsis thaliana*
E	MtUGT71G1	AAW56092	*Medicago truncatula*
F	AtUGT78D1	Q9S9P6	*Arabidopsis thaliana*
G	AtUGT85A1	AAF18537	*Arabidopsis thaliana*
H	AtUGT76B1	NP_187742	*Arabidopsis thaliana*
I	AtUGT83A1	Q9SGA8	*Arabidopsis thaliana*
J	AtUGT87A1	O64732	*Arabidopsis thaliana*
K	AtUGT86A1	Q9SJL0	*Arabidopsis thaliana*
L	AtUGT84A1	Q5XF20	*Arabidopsis thaliana*
L	McUGT75L4	ABL85474	*Maclura pomifera*
N	AtUGT82A1	Q9LHJ2	*Arabidopsis thaliana*
O	ZmcisZOG1	AAK53551	*Zea mays*
O	ZmcisZOG2	AAL92460	*Zea mays*
P	ZmUGT_P	DAA40852	*Zea mays*

UGTs in Table 3.5 were used as queries to search the *A. strigosa* root tip 454 transcriptome database with the tBLASTn search tool with a cutoff value of e^{-10}.

3.3.2 Phylogenetic analysis

Protein sequences were aligned using MAFFT (https://mafft.cbrc.jp/alignment/software/) and the evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and are in the units of the number of amino acid substitutions per site. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016).

AsUGT91 from the *A. strigosa* RNA-seq dataset was additionally included in the analysis with the sequences derived from the 454 transcriptome mining. Protein sequences from other plant species were included in the analysis as listed in Table S3.

3.3.3 Generation of self-organising maps

The Kohonen analysis was carried out based on the method implemented in Jones et al. (2017) and Payne et al. (2017). The data were filtered by removing contigs that did not have RPKM (reads per kilobase of transcript per million mapped reads) expression values above two in any tissue. The filtered data were normalised to have a mean expression level of zero and unit variance across conditions. Self-organising maps were implemented and visualised in R (R Core Team, 2017) using the Kohonen package (Wehrens and Buydens, 2007).

To avoid boundary effects a toroidal map was used so that every node had the same number of neighbours. The shape of the SOM was chosen such that the ratio between the two edge lengths was the same as the two largest eigenvalues of the data (Kohonen, 2013; Payne et al., 2017; Jones et al., 2017), and the size (72 nodes) chosen to represents at least 85% of the variance of the data as in Jones (2017). This was calculated as follows:

$$\frac{\sum_{c=1}^{S} N_c(x_c - \bar{x})^2}{\sum_{g=1}^{N}(x_g - \bar{x})^2}$$

Where N is the total number of genes; x_g is the expression vector for gene g; $\bar{x}$ is the global mean of all expression vectors; x_c is the expression vector for a SOM node c; N_c is the total number of genes assigned to SOM node c and S is the total number of SOM nodes.

3.3.4 RT-PCR profiling

RT-PCR was carried out as in Materials and Methods section 2.10 with primers, annealing temperature and number of cycles as in Table 3.6.

Table 3.6: Chapter 3 RT-PCR primers

Name	Sequence	Annealing Temperature	PCR Cycles
GAPDH-RT-PCR-F	GGTGGTCATTTCAGCCCCTA	58 °C	27
GAPDH-RT-PCR-R	CTCCCACCTCTCCAGTCCTT	58 °C	27
SAD10FWD	GAGGGAGGAGTTGGAGAGGT	58 °C	30
SAD10REV	GGGCCACAGATCGATCCATT	58 °C	30
Frt-18279	GAAGCCTCACTGGATCTTCG	58 °C	30
Rrt-18279	ATCGTGCACCTCTCTTTCGT	58 °C	30
Futr-GT01332	GGACTACGACGATGGGTCAT	58 °C	30
Rutr-GT01332	AGCAACCAGGCTCACAACTC	58 °C	30
Futr-01194	GTGTTCGGAGATGACGAGGT	58 °C	30
Rutr-01194	TCAAAACATTTTGCCTGCAC	58 °C	30
Futr-05827	CGAGCTCCTCTGCAACTACC	58 °C	30
Rutr-05827	GACATGCATTTGGCAAGAAA	58 °C	30
Futr-24a3	GTCCATGGGAGAAGCACAGT	58 °C	30
Rutr-24a3	ACAGACCCACAGCGATCTTT	58 °C	30
Futr-27a12	AATGCCATTCAGCGAGTCAT	58 °C	30
Rutr-27a12	TGGAATGTAAACTGGGAGAACA	58 °C	30
Futr-18257	CGGAATTTCTGTGCAAGTACC	58 °C	30
Rutr-18257	TCTCTGGTGATCGGATTTTT	58 °C	30
Fp-01341	TAAACCCACTGCTCGACCTC	58 °C	30
Rp-01341	CCCTCGGTCTGCTTAGTGAA	58 °C	30
Futr-02436	AGCTGATCGTGGACGTTCTT	58 °C	30
Rutr-02436	GCTCATTCGTTAACATGAATCAA	58 °C	30
Fp-11i11bis	GATGATGATGATGCAGTGGTGGAG	58 °C	30
Rp-11i11bis	TGCTGCTAGATATTGGCGGC	58 °C	30
Frp-14h20	CTGTGGTGCGTGAACACATT	58 °C	30
Rrp-14h20	TCCGATCTTGGGTAGTCTGC	58 °C	30
Fep-14h21	AGGCTGCCCCTTGAAATAGT	58 °C	30
Rep-14h21	ACGTGTCCTTGGTCATCTCC	58 °C	30
Futr-24i2	TGTGATACGCGGTGAGGTAA	58 °C	30
Rutr-24i2	AAAGCGAGCGAGGTACAAAA	58 °C	30
Futr-AsGT000892-1	ACTCCGATCTGACGGACTTG	58 °C	30
Rutr-AsGT000892-1	CACGTACGGACTGCATCATC	58 °C	30
Futr-AsGT000892-2	AGGGCGTTCAGGTGGAGAG	58 °C	30
Rutr-AsGT000892-2	TGCCTTTGTTGTTTCTGAGC	58 °C	30
Frp27f7	TCCAGGAAGAGAAGCTTGGA	58 °C	30
Rrp-27f7	TTGTATCGCTGCTCTCGTTG	58 °C	30
Futr-UGT91like	AGGAGAGAGGGGTGGGACTA	58 °C	35
Rutr-UGT91bis	GGAACCATATTGAAAAATCGCTTA	58 °C	35
Frp-16f23	CATCAATGGATGAGGCACAG	58 °C	35
Rrp-16f23	TTTCACTCCAACCTCCAACC	58 °C	35
Futr-22388	GTACCTCAAGGTCGGCTTCC	58 °C	35
Rutr-22388	GTGCATCGGAGGATCTTAGC	58 °C	35
Fp-12842	CAGCTGCATCATCAACGACT	58 °C	35
Rp-12842	ACACAGGGGTGGTCAAGAAG	58 °C	35
Futr-10433	CCAAAGAGTACGGCCAGAGA	58 °C	35
Rutr-10433	CCACGAACAAGCGATTTCAT	58 °C	35
Frt-GT15351	CGAGCACAACGTCCACGAGT	58 °C	35
Rrt-GT15351	TTCGCCTCTACAGGTGGTGG	58 °C	35
Futr-03295	AAGGTGGGTCCTCACACAGT	58 °C	35
Rutr-03295	GTGAGATTGGAGGCATAGGG	58 °C	35
Frt-GT23781	GGTTGAGGCCGCTGTGAG	60 °C	35
Rrt-GT23781	ATGTCACCTCCACCGGTTCC	60 °C	35
Fp-3i21	CAGCCGAAAGCACCATCCC	60 °C	35
Rp-3i21	TGGAGTAGTCCACCTCCTGCTTC	60 °C	35

Chapter 4

Identification of a new avenacin UGT

4.1 General introduction

In Chapter 3, UGT candidates for the missing glucosylation steps in the biosynthesis of avenacin A-1 were identified by mining two *A. strigosa* transcriptome databases and prioritised for further investigation by phylogenetic analysis and expression profiling.

The availability of complete plant genome sequences has allowed rapid advances in the functional characterisation of UDP-dependent glycosyltransferases (UGTs) (Vogt and Jones, 2000; Bowles and Lim, 2010; Tiwari et al., 2016). Historically, attempts to purify UGTs directly from plant tissues were complicated by the low natural levels of UGTs and the difficulty of separating them from other UGTs present (Vogt and Jones, 2000). Based on genomic information, UGTs can be cloned and expressed in hosts such as *E. coli* for *in vitro* enzymatic assays (Vogt and Jones, 2000; Bowles and Lim, 2010), and entire metabolic pathways can be reconstituted by expressing pathway genes in heterologous hosts such as yeast or tobacco (Thimmappa et al., 2014).

Nonetheless, the lack of available substrates is a significant hurdle to evaluate the activity of these enzymes, as physiological substrates are generally structurally complicated molecules that are difficult to purify or to generate by synthetic chemistry. In addition, many UGTs are highly promiscuous *in vitro*, and so results obtained in *in vitro* assays may not be a true reflection of their physiological role (Bowles et al., 2005). Their specificity *in planta* might be influenced by compartmentalisation, protein-protein interactions, metabolite channelling and the presence of co-factors that are missing *in vitro* (Bowles et al., 2006). Promiscuous activity may also occur in

more physiologically relevant hosts such as *N. benthamiana* (Thimmappa et al., 2014).

Whilst this promiscuity can be a useful tool in synthetic biology to create new-to-nature molecules, confirmation of the physiological role of UGTs should ideally be carried out in the plant of origin, by gene knock-out, RNA-interference and/or overexpression (Bowles et al., 2005; Thimmappa et al., 2014).

4.1.1 Generation of substrates to screen glycoside-specific glycosyltranferases of the avenacin pathway

An *in vitro* assay is an attractive option for preliminary screening of the UGT candidates shortlisted in Chapter 3. However, this approach requires the availability of suitable substrates.

The glucosylation events under study here involve the addition of two D-glucose molecules onto an L-arabinose that is linked to the avenacin triterpene scaffold at the C-3 position (Figure 4.1). The current model of the avenacin biosynthetic pathway is shown in Figure 4.1. The early steps that involve the formation and oxidation of the triterpene backbone are catalysed by the oxidosqualene cyclase AsbAS1 (SAD1) and the P450s, AsCYP51H10 (SAD2), AsCYP72A475 (SAD6), AsCYP72A476 and AsCYP94D65 (Haralampidis et al., 2001; Qi et al., 2006; Reed, 2016; Leveau et al., manuscript in preparation). These steps are likely to be associated with the cytosolic face of the endomembrane system (Wegel et al., 2009; Reed, 2016).

The next step is predicted to be the addition of an α-L-arabinose at the C-3 position by the avenacin arabinosyltransferase, AsAAT1, in the cytosol (Louveau et al., manuscript in preparation).

The triterpene scaffold is acylated at the C-21 position by the action of three avenacin pathway enzymes, AsMT1 (SAD9), AsUGT74H5 (SAD10) and AsSCPL1 (SAD7) (Mugford et al., 2009; Owatworakit et al., 2013; Mugford et al., 2013). The methyl transferase AsMT1 methylates anthranilate from the shikimate pathway, which is then glucosylated in the cytosol by the UGT AsUGT74H5 to form *N*-methyl anthraniloyl glucose. This product is transported (by an unknown mechanism) into the vacuole and acts as the the acyl donor substrate for the vacuolar acyltransferase AsSCPL1 (Mugford et al., 2013).

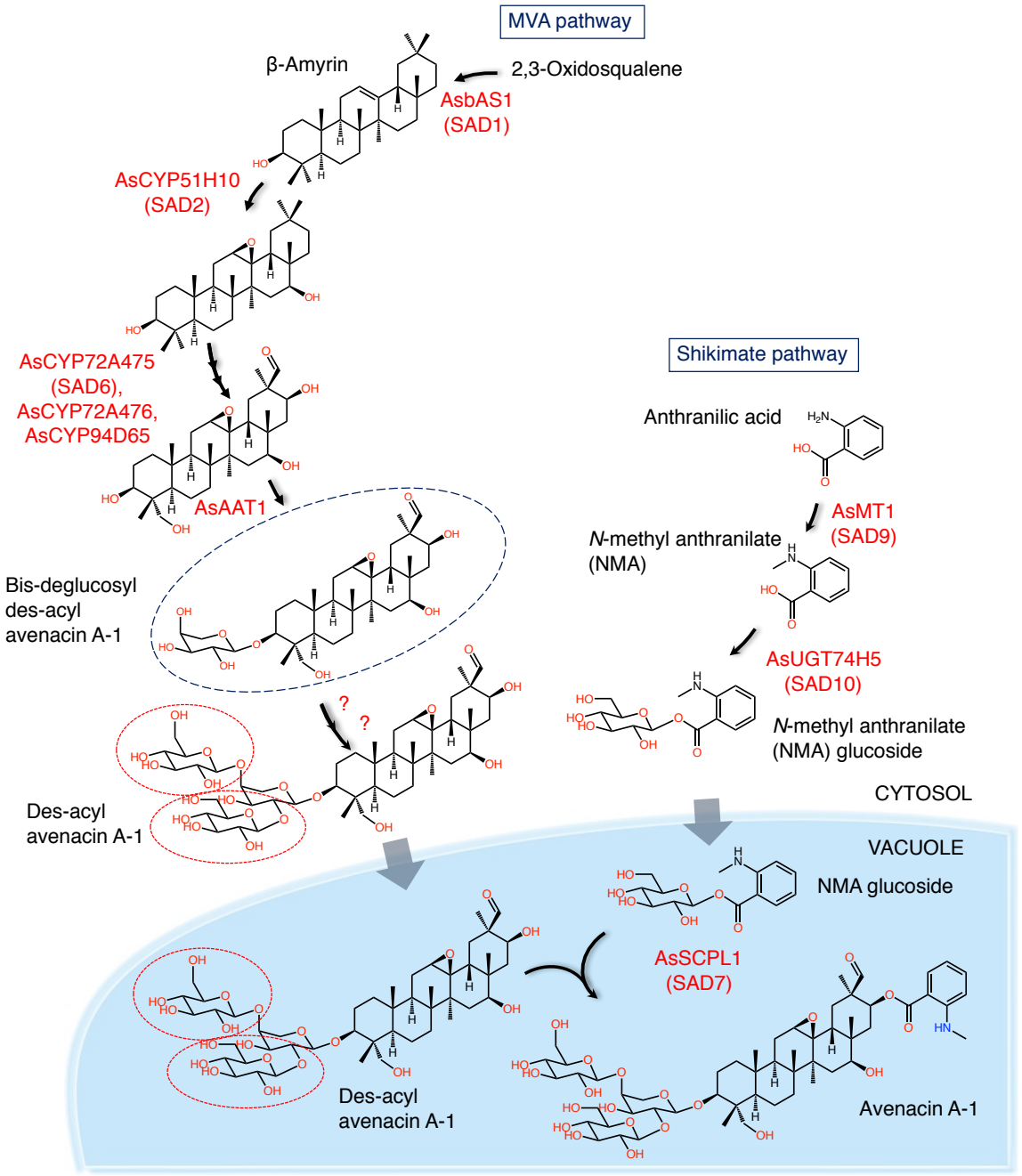


Figure 4.1: The predicted substrate for the avenacin α-L-arabinoside D-glucosyltransferase enzyme is bis-deglucosyl-des-acyl-avenacin A-1. β-Amyrin is formed by the cyclisation of 2,3-oxidosqualene by the β-amyrin synthase AsbAS1 (SAD1) and modified by cytochrome P450 AsCYP51H10 (SAD2) by both epoxidation and hydroxylation (Haralampidis et al., 2001; Qi et al., 2006). This product is then further oxidised by three cytochromes P450, AsCYP72A475 (SAD6), AsCYP72A476 and AsCYP94D65 and glycosylated in the cytosol at the C-3 carbon position by an arabinosyltransferase, AsAAT1, to form bis-deglucosyl des-acyl avenacin A-1 (blue dashed circle) (Reed, 2016; Leveau et al., manuscript in preparation; Louveau et al., manuscript in preparation). Two D-glucose molecules are added at the L-arabinose 2-*O* and 4-*O* positions (red dashed circles) by unknown glucosyltransferases to form des-acyl avenacin A-1.
Anthranilic acid generated by the shikimate pathway is methylated by the methyltransferase AsMT1 (SAD9) and glucosylated in the cytoplasm by AsUGT74H5 (SAD10) to give *N*-methyl anthranilate (NMA) glucoside (Owatworakit et al., 2013; Mugford et al., 2013). NMA glucoside is transported to the vacuole where it acts as the activated acyl donor for the serine carboxypeptidase-like acyltransferase AsSCPL1 (SAD7), which acylates avenacin A-1 at the C-21 carbon position (Mugford et al., 2009). Grey arrows, unknown transport mechanisms.

As plant UGTs are generally assumed to be cytosolic (Jones and Vogt, 2001; Caputi et al., 2012) as has been shown for the avenacin arabinosyltransferase, AsAAT1 (Louveau et al., manuscript in preparation), the branching glucosylation steps are proposed to occur before the triterpene scaffold is transported (by an unknown mechanism) into the vacuole for acylation. The predicted substrate for the first branching glucosylation step is therefore the non-acylated bis-deglucosyl des-acyl avenacin A-1 (ringed with blue dash line in Figure 4.1).

The roots of *A. strigosa sad3* and *sad4* mutants, which are both deficient in avenacin A-1 glucosylation (Papadopoulou et al., 1999; Mylona et al., 2008) accumulate a disaccharide avenacin intermediate that lacks the β-1,4-D-glucose, named mono-deglucosyl avenacin A-1. This suggests that the glucose molecules are not added by a single multifunctional UGT, but that another enzyme is involved in the addition of the β-1,2-D-glucose. In addition it suggests that during the biosynthesis of avenacin A-1, the β-1,2-D-glucose is either added first or can be added in the absence of the β-1,4-linked D-glucose.

Therefore it is likely that a bis-deglucosyl substrate lacking both branching D-glucose moieties will be a suitable *in vitro* substrate for the β-1,2-glucosyltransferase. Depending on the order of glucose addition, either bis- or mono-deglucosyl (lacking the β-1,4-D-glucose) substrates are likely to be relevant for the β-1,4-glucosyltransferase.

4.2 Aims

The aims of this work are to:

1. Generate suitable substrates for the identification of UGTs that are able to add a D-glucose to the L-arabinose of the avenacin triterpene scaffold.

2. Determine which of the candidate UGTs identified in Chapter 3 have triterpene-*O*-arabinoside glucosyltransferase activity when heterologously expressed in *E. coli* and *N. benthamiana.*

3. Investigate the role of positive UGT candidates in avenacin biosynthesis in *A. strigosa*.

4.3 Results and Discussion

4.3.1 Generation of substrates for *in vitro* evaluation of avenacin UGT candidates

The complex glycosylated triterpenes bis-deglucosyl des-acyl avenacin A-1 and mono-deglucosyl (lacking the β-1,4-D-glucose) des-acyl avenacin A-1 are not commercially available. Achieving the regio-specificity of these substrates through chemical synthesis from available precursors such as β-amyrin would be extremely challenging. In addition, these intermediates are not detectable or accumulate as very minor components in wild type oat roots (Crombie et al., 1986a) and so cannot be purified directly from oat plants.

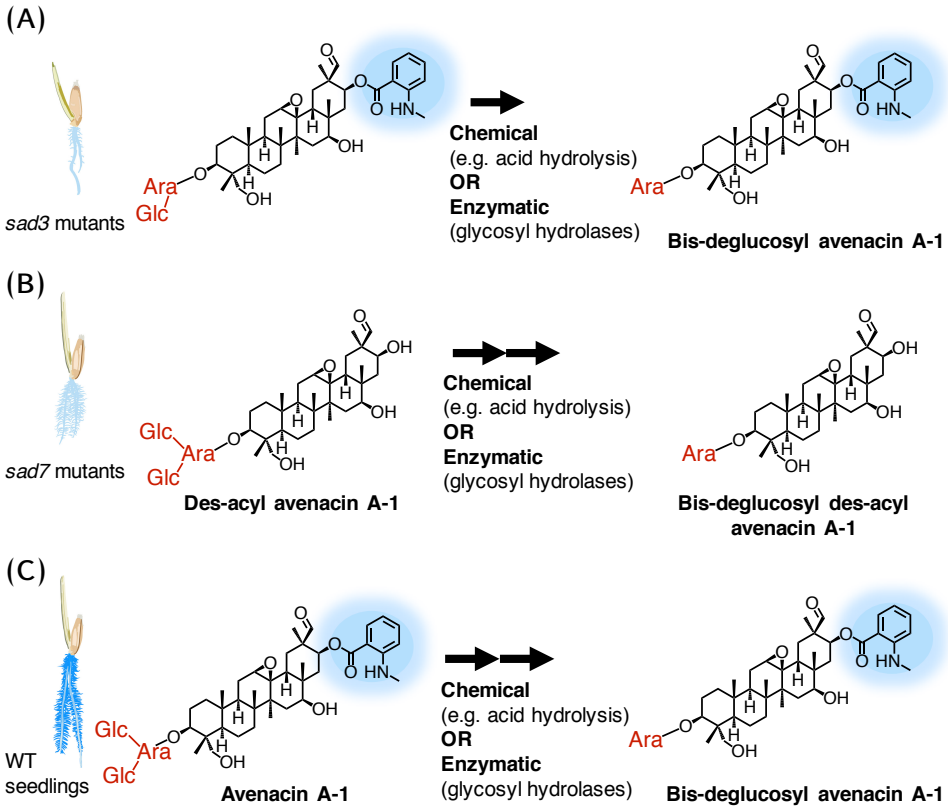


Figure 4.2: Routes to generate substrates by deglucosylation of glycosylated compounds. The roots of *sad3* and *sad7* mutants of *A. strigosa* accumulate mono-deglucosyl avenacin A-1 and des-acyl avenacin A-1, respectively (Papadopoulou et al., 1999; Mylona et al., 2008; Mugford et al., 2009). Wild-type (WT) oat seedlings accumulate avenacin A-1 (Begley et al., 1986). These compounds could be extracted and deglucosylated chemically or enzymatically (black arrows) to obtain (**A**) and (**C**), bis-deglucosyl avenacin A-1, a substrate with the UV-active (blue highlight) acyl group; or (**B**), bis-deglucosyl-des-acyl avenacin A-1, the proposed substrate for the first avenacin-3-*O*-L-arabinoside D-glucosyltransferase. Glc = D-glucose; Ara = L-arabinose.

An alternative approach could be to extract mono-deglycosyl avenacin A-1 from *sad3* mutants which accumulate this compound (Papadopoulou et al., 1999; Mylona et al., 2008), and to additionally obtain bis-deglycosyl avenacin A-1 by removing the β-1,2-D-glucose either chemically or enzymatically (Figure 4.2A). These substrates would be acylated with the UV-active *N*-methyl anthranilate group (highlighted in blue in Figure 4.2A), which is predicted to be added at a later step in avenacin biosynthesis. Therefore des-acyl avenacin A-1 could potentially be extracted from the roots of *sad7* mutants, which are unable to acylate avenacin (Papadopoulou et al., 1999; Mugford et al., 2009), and subsequently deglucosylated to obtain the predicted substrates (Figure 4.2B). However, the small size of the *sad3* mutant roots and the availability of oat mutant seeds of *sad3* and *sad7* makes it unfeasible to extract sufficient quantities of these compounds for subsequent de-glucosylation and purification of products.

Avenacins accumulate at 0.22-1.0 mg/g dry weight of roots in several analysed oat species (Crombie and Crombie, 1986), and therefore avenacin A-1 can be purified in milligram quantities from oat roots of the readily available wild type hexaploid *Avena sativa*. Although de-glucosylated products derived from avenacin A-1 will also have the UV-active acyl group (Figure 4.2C), they will have many chemical features of the proposed substrates. Therefore avenacin extraction from *A. sativa* was the chosen strategy to obtain substrates.

For avenacin A-1 extraction, *A. sativa* oat seedlings were grown hydroponically and the roots harvested and freeze-dried as described in Materials and Methods section 4.5.1. A crude root extract was obtained by concentration and precipitation of an 80% methanol extract of the freeze-dried oat roots (34 g) (Materials and Methods section 4.5.1). Avenacin A-1 is present in oat roots along with the other three avenacins, avenacins A-2, B-1 and B-2 (Begley et al., 1986), and other compounds including the fluorescent compounds scopoletin and scopoletin glycoside (Goodwin and Pollock, 1954). LC-mass spectrometry analysis showed that the crude root extract contained a large amount of avenacin A-1, along with other molecules such as avenacin B-1 (Figure S1).

Avenacins have been shown to separate well by reverse-phase chromatography (Begley et al., 1986). Therefore the oat root extract was separated on a C-18 flash chromatography column as described in Materials and Methods Section 4.5.1 to result in fractions (such as fraction 14) predominantly containing avenacin A-1, with a small amount of avenacin B-1 (Figure 4.3). To reduce yield losses from further purification steps, deglycosylation was carried out at this stage.

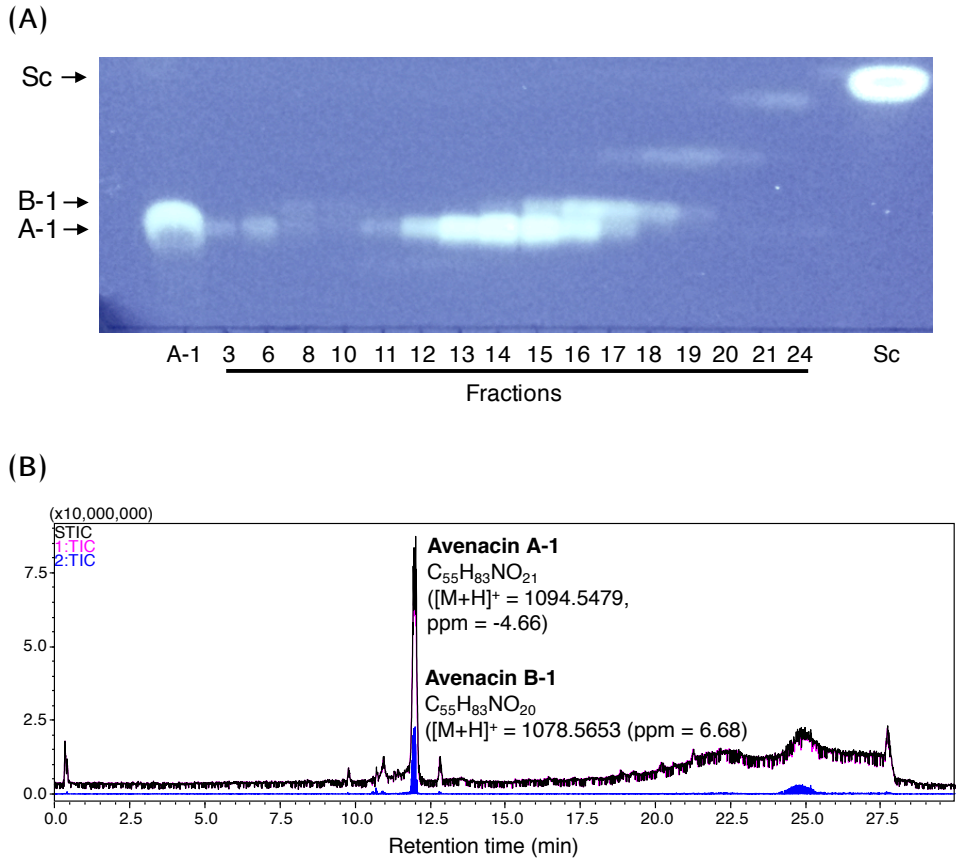


Figure 4.3: Generation of oat root extract fractions enriched for avenacin A-1.
An extract enriched for avenacin A-1 was obtained by separating a crude oat root methanolic extract on a C-18 flash chromatography column (see Materials and Methods section 4.5.1).
(A) Analytical thin layer chromatography of the C-18 flash chromatography fractions. Fractions were loaded on a silica TLC plate, developed with chloroform/methanol/water (13:6:1) and visualised under UV light at 365 nm. A-1 = avenacin A-1 standard; Sc = scopoletin standard.
(B) Total ion chromatogram (TIC) of C-18 flash chromatography fraction 14 analysed by HPLC-IT-ToF in positive mode shows that fraction 14 mainly consists of peaks corresponding to avenacin A-1 ($[M+H]^+$ = 1094.5479) and avenacin B-1 ($[M+H]^+$ = 1078.5653). The difference in observed and expected mass values in parts per million is indicated (ppm).

Development of methods for avenacin A-1 deglucosylation

Although acidic hydrolysis in mild conditions may allow the removal of the terminal D-glucose molecules (Ikeda et al., 1998), this method is likely to degrade other elements of the avenacin scaffold such as the epoxide moiety (Crombie et al., 1984b, 1986a; Geisler et al., 2013).

Avenacin A-1 can be enzymatically de-glucosylated by secreted glycosyl hydrolases (GHs) from different fungal species. A variety of root-infecting fungi secrete enzymes that hydrolyse avenacin A-1 to less toxic forms (Carter et al., 1999; Bowyer et al., 1995). These enzymes may enable them to infect and cause disease on oat roots as has been shown for *Gaeumannomyces graminis* var. *avenae*, which produces an enzyme known as avenacinase (Carter et al., 1999; Turner, 1961; Osbourn et al., 1991; Bowyer et al., 1995). Avenacinase first removes the β-1,4-D-glucose to give mono-deglucosyl avenacin A-1 (Osbourn et al., 1991; Papadopoulou et al., 1999; Mylona et al., 2008). With further enzymatic digestion, the β-1,2-D-glucose is also removed to give bis-deglucosyl avenacin A-1. Therefore this enzyme was used to generate both of these two substrates.

To obtain a preparation of avenacinase, a culture filtrate of *G. graminis* var. *avenae* strain A3 (Bryan et al., 1999) was concentrated with ammonium sulphate and dialysed as in Materials and Methods section 4.5.2. A preliminary assay was carried out to investigate suitable conditions to generate both bis- and mono-deglucosyl avenacin A-1 from a large-scale deglucosylation reaction of avenacin A-1, based on the method of Osbourn et al. (1991). Reactions were set up in a total volume of 100 µl, with two concentrations of the avenacin A-1 substrate (0.25 mg/ml and 1 mg/ml of the dried oat root extract fractions enriched for avenacin A-1) and 3 or 9 µl of the ammonium sulphate concentrated avenacinase preparation (3% and 9% v/v of the total reaction volume).

The reactions were incubated at 37°C overnight and analysed by separation on a silica TLC plate (Figure 4.4). At the higher substrate concentration of 1 mg/ml, there was still a substantial amount of fully-glucosylated avenacin A-1 and the major product in this reaction was mono-deglucosyl avenacin A-1.

At the lower substrate concentration of 0.25 mg/ml, most of the starting avenacin A-1 substrate had been deglucosylated and a higher proportion of mono-deglucosyl avenacin A-1 was converted to bis-deglucosyl avenacin A-1 (Figure 4.4). There was not an appreciable advantage in using the higher concentration (9%) of the avenacinase preparation, and therefore the condition with 3% v/v of the avenacinase preparation

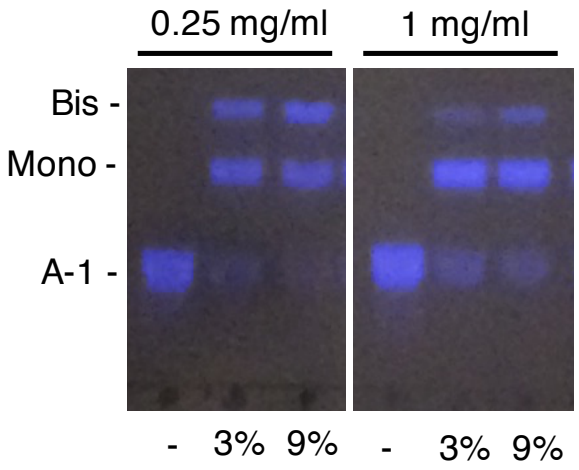


Figure 4.4: Assay to test reaction conditions for the enzymatic deglucosylation of avenacin A-1 Analytical thin layer chromatography of the small scale enzymatic reactions with 0.25 mg/ml (left) or 1 mg/ml (right) of the oat root extract enriched for avenacin A-1 (A-1) as substrate, with 0 (-), 3 (3%) or 9 (9%) µl of the ammonium sulphate concentrated avenacinase preparation. Reactions containing avenacinase resulted in the generation of mono-deglucosyl avenacin A-1 (Mono). Bis-deglucosyl avenacin A-1 (Bis) was more clearly accumulated at the substrate concentration of 0.25 mg/ml. No large difference was seen between 3 or 9 % of the avenacinase preparation. Experiments were carried out in a total volume of 100 µl with 100 mM sodium acetate buffer pH 5.0 at 37°C overnight. 10% and 5% of the 0.25 mg/ml and 1 mg/ml reaction volumes respectively were loaded onto a silica TLC plate. The plate was developed with chloroform/methanol/water (13:6:1) and visualised under UV light at 365 nm.

in the total reaction volume and 0.25 mg/ml of the avenacin A-1 extract was used in the scaled-up reaction to produce both bis- and mono-deglucosyl avenacin A-1. The de-glucosylated products were purified by preparative thin-layer chromatography as in Materials and Methods section 4.5.3.

This bis-deglucosyl avenacin A-1 product was analysed by HPLC-IT-ToF mass spectrometry (Figure 4.5). The molecular formula of bis-deglucosyl avenacin A-1 is $C_{43}H_{63}NO_{11}$ and the expected $[M+H]^+$ is 770.4474. The observed signal in the HPLC-IT-TOF analysis was 770.4414; which has an error of -6.0mDa, -7.79ppm (Figure 4.5C). As bis-deglucosyl avenacin A-1 may be the substrate for the 1,2-glucosyltransferase or the 1,4-glucosyltransferase, this was used as the substrate for the initial enzymatic assay.

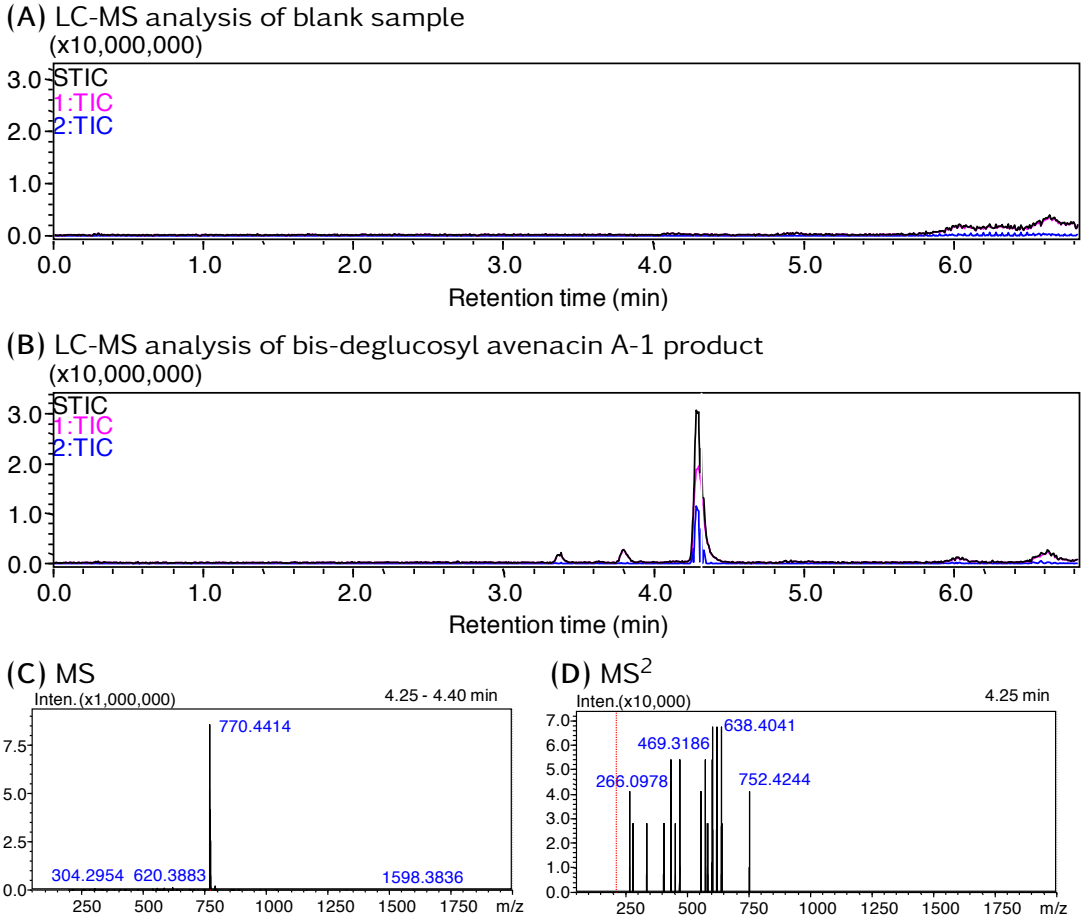


Figure 4.5: IT-ToF analysis of bis-deglucosyl avenacin A-1 purified by preparative TLC in positive mode. Total ion count (TIC) of: **(A)**, blank sample and **(B)**, bis-deglucosyl avenacin A-1 product. **(C)** The expected mass ion $[M+H]^+$ for bis-deglucosyl avenacin A-1 is 770.4474. The mass spectrum (MS) of the bis-deglucosyl avenacin A-1 peak between 4.25-4.39 minutes shows a major mass ion m/z 770.4414, which has an error of -6.0mDa, -7.79ppm. **(D)** The MS^2 spectrum of the fragmentation of this mass ion (precursor m/z = 770.9512) gave fragments consistent with the loss of one pentose (-132, m/z = 638.4041) although the error was greater than ±5 ppm.

4.3.2 Screening of UGT candidates for avenacin glycosyltransferase activity

AsUGT91, AsGT27f7, AsGT14h20, AsGT11i11, AsGT05827 and AsGT24a3 were prioritised in Chapter 3 as candidate GGTs for avenacin biosynthesis.

AsGT14h20, AsGT11i11 and AsGT05827 had already been cloned from *A. strigosa* during the search for the avenacin arabinosyltransferase (Louveau, 2013). The Gateway system (Hartley et al., 2000) was used to transfer these UGTs into the pH9GW plasmid, a vector for the heterologous expression of N-terminally 9xHis-tagged protein in *E. coli* under the control of the Lac promoter. AsUGT91, AsGT24a3 and AsGT27f7 were cloned from *A. strigosa* as in Materials and Methods section 4.5.4 using the same strategy.

Expression of AsUGT91, AsGT14h20, AsGT05827 and AsGT27f7 in *E. coli* (Materials and Methods section 2.18) yielded substantial amounts of soluble protein. AsGT11i11 did not express well as previously noted (Louveau, 2013). Upon IPTG induction, the culture appeared to stop actively dividing as the optical density (OD_{600nm}) did not increase past the induction time point, suggesting expression of this construct may be toxic to the cells. This particular enzyme may therefore need to be tested in a different heterologous system such as *N. benthamiana*, where it might be better tolerated. Recombinant expression of AsGT24a3 yielded a very small amount of soluble protein, requiring further optimisation of expression conditions. Therefore, neither AsGT11i11 or AsGT24a3 were included in the first round of *in vitro* screens.

The UGT candidates that expressed well in *E. coli* (AsUGT91, AsGT14h20, AsGT27f7 and AsGT05827) were partially purified using Ni-NTA agarose beads (Materials and Methods section 2.18) (Supplementary Figure S3). Proteins were quantified by Bradford assay and SDS-PAGE gel analysis. Protein preparations were then used in assays with bis-deglucosyl avenacin A-1 as an acceptor and UDP-glucose as the sugar donor (Material and Methods section 4.5.5). Control reactions (boiled protein preparations) were set up with enzyme preparations that had been denatured at 95°C for 10 minutes.

HPLC-UV-MS analysis of the assay (Figure 4.6) showed that one enzyme, AsUGT91, yielded a new more polar product peak at 4.8 minutes (Figure 4.6B).

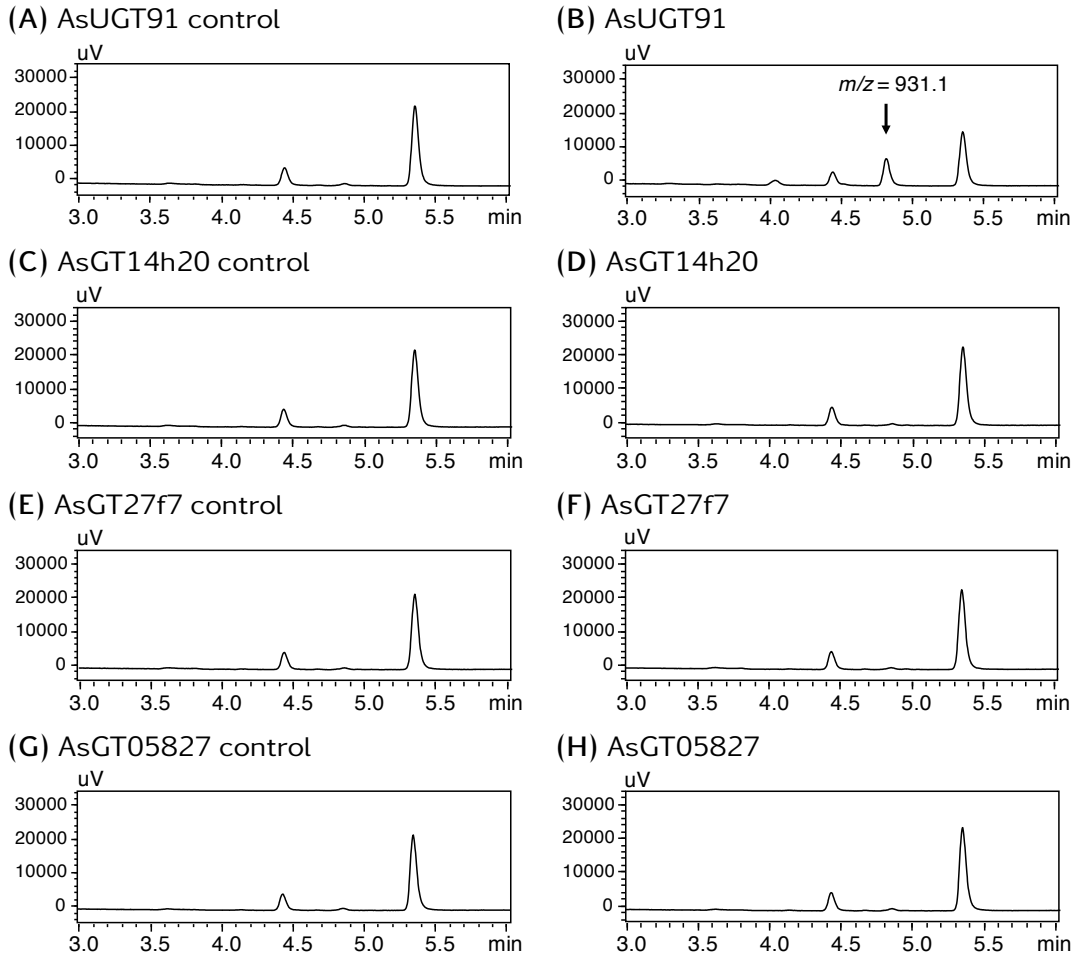


Figure 4.6: AsUGT91 is active against bis-deglucosyl avenacin A-1.
HPLC-UV-MS analyses of the enzymatic reactions with bis-deglucosyl avenacin A-1 and UDP-glucose with: (left), boiled protein preparations (control reactions) of **(A)** AsUGT91, **(C)** AsGT14h20, **(E)** AsGT27f7, **(G)** AsGT05827; and (right), protein preparations of **(B)** AsUGT91, **(D)** AsGT14h20, **(F)** AsGT27f7, **(H)** AsGT05827. The enzymatic reaction with AsUGT91 resulted in a new peak at 4.8 minutes (m/z = 931.1). Protein preparations for control reactions had been boiled at 95°C for 10 minutes. Data are representative of two separate experiments. Absorbance was measured at 357nm (Begley et al., 1986).

Mass-spectrometry analysis of the new peak at 4.8 minutes showed prominent signals of the same mass as the mass ion and chloride and formate adducts of bis-deglucosyl avenacin A-1 with the addition of a glucose molecule (Figure 4.7).

This suggests that AsUGT91 is an active glucosyltransferase which can glycosylate triterpene glycosides.

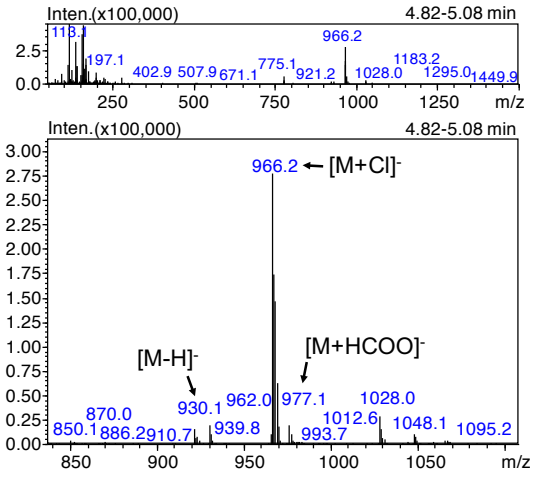


Figure 4.7: HPLC-single quadrupole mass spectrometry analysis of the AsUGT91 product peak. The mass spectrum (negative mode) of the AsUGT91 product peak between 4.8-5.1 minutes showed ions consistent with the mass ion ($[M-H]^-$, m/z = 930.1), chloride adduct ($[M+Cl]^-$, m/z = 966.2) and formate adduct ($[M+HCOO]^-$, m/z = 977.1) of bis-deglucosyl avenacin A-1 with the addition of a glucose molecule.

4.3.3 Transient co-expression of AsUGT91 with avenacin pathway enzymes in *N. benthamiana* leaves

Transient expression in *N. benthamiana* leaves has been used to characterise avenacin biosynthetic enzymes (Mugford et al., 2013; Geisler et al., 2013). The benefits of this heterologous system for plant triterpene biosynthetic gene expression include the presence of metabolic precursors and co-enzymes, correct mRNA and protein processing, protein and metabolite localisation and its speed, as it takes six days from agro-infiltration to the analysis of leaf extracts (Reed et al., 2017; Geisler et al., 2013; Thimmappa et al., 2014). In addition, this system can be used with vacuum-infiltrated whole plants to rapidly generate sufficient quantities of compounds for NMR structural analysis (Reed et al., 2017).

The *in vitro* assay indicated that AsUGT91 is able to glucosylate the triterpene glycoside, bis-deglucosyl avenacin A-1. To investigate this activity further and to corroborate the *in vitro* results, the function of AsUGT91 was also evaluated in *N. benthamiana* by transient plant expression.

Enhancing triterpene accumulation in the *N. benthamiana* system

Triterpene production in heterologous expression systems can be enhanced by increasing the availability of metabolic precursors (Reed et al., 2017; Schaller et al., 1995; Harker et al., 2003). The enzyme 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR) catalyses a key rate-limiting step in the sterol and triterpenoid biosythetic pathway (Figure 4.8). Co-expression with HMGR can boost triterpene accumulation in the *N. benthamiana* system and therefore an *Agrobacterium* strain containing an *A. strigosa* HMGR expression construct was included in the transient expression experiments.

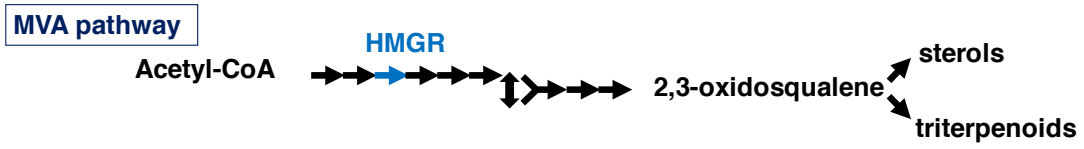


Figure 4.8: HMGR is a rate-limiting step in the melavonate pathway. 2,3-oxidosqualene is synthesised from acetyl CoA in the mevalonate pathway, and is the precursor to sterols and triterpenes. The rate-limiting 3-hydroxy, 3-methylglutaryl-coenzyme A reductase (HMGR) of the mevalonate pathway is highlighted in blue.

AsUGT91 was cloned into pEAQ-*HT*-DEST1, a Gateway-compatible binary vec-

tor for high-level transient expression of proteins in *N. benthamiana* (Sainsbury et al., 2012). pEAQ-*HT*-DEST1-AsUGT91 was transformed into *A. tumefaciens* LBA4404 and co-infiltrated with *A. tumefaciens* strains containing avenacin pathway genes. Infiltrated leaves were harvested after 5 days and the extracted metabolites were analysed by HPLC-CAD-MS.

The oxidosqualene cyclase of the avenacin biosynthetic pathway, AsbAS1 (SAD1), forms the β-amyrin backbone and this is modified by the cytochrome P450 AsCYP51H10 (SAD2) to form EpHβA[1](Figure 4.9A) (Haralampidis et al., 2001; Qi et al., 2006). Co-expression of AsbAS1 and AsCYP51H10 in *N. benthamiana* leaves results in the accumulation of EpHβA (Geisler et al., 2013), however this molecule is not in the the correct polarity range to be detected in this HPLC-MS analysis (Figure 4.9B).

The avenacin arabinosyltransferase, AsAAT1, adds an α-L-arabinose moiety to EpHβA to yield EpHβA-3-*O*-Ara[1] (Figure 4.9A) (Louveau et al., manuscript in preparation). Co-expression of AsbAS1, AsCYP51H10 and AsAAT1 results in the accumulation of EpHβA-3-*O*-Ara which has a mass ion $[M\text{-}H]^-$ of 589.4 and elutes at 19.9 minutes in the LC-MS analysis (Figure 4.9B).

AsUGT91 appears to be catalytically active in this system, as co-infiltration of AsUGT91 with AsbAS1, AsCYP51H10 and AsAAT1 results in the consumption of the EpHβA-3-*O*-Ara peak, and the appearance of four new more polar peaks: two consistent with the mass of EpHβA-3-*O*-Ara with the addition of a hexose (15.2 and 17.4 minutes), and two peaks (14 and 16.1 minutes) consistent with the mass of EpHβA-3-*O*-Ara with the addition of two hexoses (Figure 4.9B). These latter peaks may be the result of endogenous modifications of the AsUGT91 product by *N. benthamiana* glycosyltransferases, rather than the action of AsUGT91 alone.

[1]Abbreviations: EpHβA = 12,13-β-epoxy, 16β-hydroxy-β-amyrin; EpHβA-3-*O*-Ara = 12,13-β-epoxy, 16β-hydroxy-β-amyrin-3-*O*-α-L-arabinose; C-21-hydroxy-EpHβA = 12,13-β-epoxy, 16β,21β-dihydroxy-β-amyrin; C-21-hydroxy-EpHβA-3-*O*-Ara = 12,13-β-epoxy, 16β,21β-dihydroxy-β-amyrin-3-*O*-α-L-arabinose.

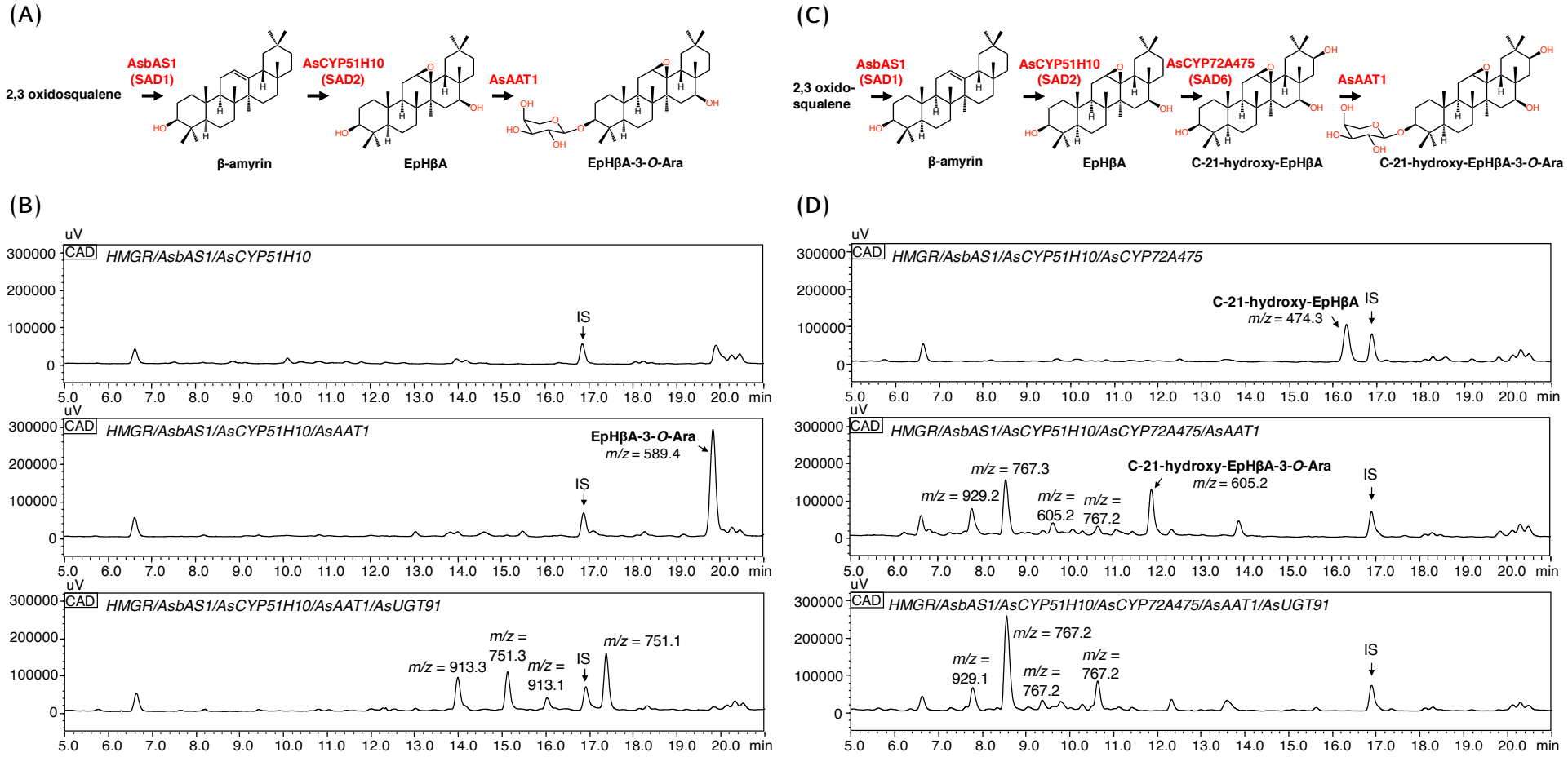


Figure 4.9: Expression of HMGR, AsbAS1, AsCYP51H10, AsCYP72A475, AsAAT1 and AsUGT91 in *N. benthamiana* leaves.
(A), Formation of EpHβA-3-*O*-Ara from 2,3-oxidosqualene by AsbAS1 (SAD1), AsCYP51H10 (SAD2) and AsAAT1 (Haralampidis et al., 2001; Qi et al., 2006; Louveau et al., manuscript in preparation). **(B)**, HPLC-CAD-MS analysis of extracts from infiltrated *N. benthamiana* leaves. Accumulation of EpHβA-3-*O*-Ara was detected in leaves expressing AsbAS1, AsCYP51H10 and AsAAT1. The addition of AsUGT91 resulted in lower levels of EpHβA-3-*O*-Ara and the accumulation of new peaks with the mass of EpHβA-3-*O*-Ara with the addition of one [*m/z* = 751, Rt = 15.2 and 17.4 min], or two [*m/z* = 913, Rt = 14.0 and 16.1 min] hexoses, respectively. **(C)**, Formation of C-21-hydroxy-EpHβA-3-*O*-Ara from 2,3-oxidosqualene by AsbAS1 (SAD1), AsCYP51H10 (SAD2), AsCYP72A475 (SAD6) and AsAAT1 (Haralampidis et al., 2001; Qi et al., 2006; Reed, 2016; Leveau et al., manuscript in preparation; Louveau et al., manuscript in preparation). **(D)**, HPLC-CAD-MS analysis of extracts from infiltrated *N. benthamiana* leaves. Co-expression of AsbAS1, AsCYP51H10 and AsCYP72A475 resulted in the accumulation of C-21-hydroxy-EpHβA. The addition of AsAAT1 leads to the accumulation of C-21-hydroxy-EpHβA-3-*O*-Ara along with C-21-hydroxy-EpHβA-3-*O*-Ara glycoside peaks due to endogenous *N. benthamiana* glycosyltransferases. Co-expression of AsbAS1, AsCYP51H10, AsCYP72A475, AsAAT1 and AsUGT91 resulted in the reduction of the C-21-hydroxy-EpHβA-3-*O*-Ara peak and an increase in the peaks with the mass of C-21-hydroxy-EpHβA-3-*O*-Ara with an additional hexose [*m/z* = 767.2, Rt = 8.6 and 10.6 min]. Data is representative of two replicates. IS = internal standard (digitoxin). EpHβA = 12,13-β-epoxy, 16β-hydroxy-β-amyrin; EpHβA-3-*O*-Ara = 12,13-β-epoxy, 16β-hydroxy-β-amyrin-3-*O*-α-L-arabinose; C-21-hydroxy-EpHβA = 12,13-β-epoxy, 16β,21β-dihydroxy-β-amyrin; C-21-hydroxy-EpHβA-3-*O*-Ara = 12,13-β-epoxy, 16β,21β-dihydroxy-β-amyrin-3-*O*-α-L-arabinose.

An additional avenacin pathway enzyme, the cytochrome P450 AsCYP72A475 (SAD6), oxidises EpHβA at the C-21 position to yield C-21-hydroxy-EpHβA[1] (Figure 4.9C) (Leveau et al., manuscript in preparation). The activity of AsUGT91 with combinations of avenacin pathway enzymes including AsCYP72A475 was also evaluated (Figure 4.9D).

Co-expression of AsbAS1, AsCYP51H10 and AsCYP72A475 results in the production of C-21-hydroxy-EpHβA[1], which elutes at 16.2 minutes ($[M\text{-}H]^-$ = 474.3) in the LC-CAD-MS analysis (Figure 4.9D). The arabinosyltransferase, AsAAT1, glycosylates C-21-hydroxy-EpHβA to result in C-21-hydroxy-EpHβA-3-*O*-Ara[1] (Figure 4.9C) (Louveau et al., manuscript in preparation).

Coexpression of AsbAS1, AsCYP51H10, AsCYP72A475 and AsAAT1 results in the accumulation of C-21-hydroxy-EpHβA-3-*O*-Ara (Louveau et al., manuscript in preparation), which appears as a new peak at 12 minutes in the HPLC-CAD chromatogram (m/z = 605.2) (Figure 4.9D). C-21-hydroxy-EpHβA-3-*O*-Ara is modified by endogenous glycosyltransferases, as can be seen by the additional appearance of several more polar peaks consistent with the addition of one (m/z = 767.3) or two (m/z = 929.2) hexoses ((Figure 4.9D)

Co-expression of AsUGT91 with AsbAS1, AsCYP51H10, AsCYP72A475 and AsAAT1 results in the increase of the peaks with the mass of the C-21-hydroxy-EpHβA-3-*O*-Ara with one hexose, with the concomitant reduction in the C-21-hydroxy-EpHβA-3-*O*-Ara peak eluting at 12 minutes ((Figure 4.9D), suggesting that AsUGT91 is converting the monoglycoside C-21-hydroxy-EpHβA-3-*O*-Ara to the diglycoside.

Thus AsUGT91 appears able to add hexose moieties to avenacin pathway intermediates with or without the C-21 modification in the *N. benthamiana* system.

The activities of the C21 P450 AsCYP72A475 and the glycosyltransferases AsAAT1 and AsUGT91 are dependent on the presence of the triterpene scaffold, EpHβA, as no new peaks are seen if AsbAS1 and AsCYP51H10 are not co-infiltrated with these three enzymes (Figure S4).

Additionally, the activity of AsUGT91 is dependent on the arabinosyltransferase AsAAT1, as coexpression of a combination of AsbAS1, AsCYP51H10, AsCYP72A475 and AsUGT91 without AsAAT1 results only in the un-glycosylated C-21-hydroxy-EpHβA (Figure S4), suggesting that arabinose moiety is required for binding specificity or that the hexose is added on to the arabinoside moiety, and not directly on to the C-21-hydroxy-EpHβA scaffold.

To investigate the structure of the AsUGT91 product produced *in planta*, the *N.*

benthamiana infiltration was scaled up using a vacuum infiltration system (Reed et al., 2017) to obtain 4.5 mg of the purified molecule for structural analysis. The combination of AsbAS1, AsCYP51H10, AsAAT1 and AsUGT91 was used, as this did not have a background of peaks due to endogenous *N. benthamiana* glycosyltranferases, such as is seen with the C-21 modified arabinoside (Figure 4.9D).

Freeze-dried leaves from 150 *N. benthamiana* plants provided by Thomas Louveau (JIC) were extracted (Materials and Methods section 4.5.7). HPLC-CAD-MS analysis of the extract revealed one peak with a mass consistent with the diglycoside EpHβA-3-*O*-Ara[1] with the addition of a hexose (Figure 4.10A, *m/z* = 751.3, 9.2 minutes), and one peak with the mass of the triglycoside EpHβA-3-*O*-Ara with the addition of two hexoses (Figure 4.10A, *m/z* = 913.3, 8 minutes).

The AsUGT91 diglycoside product peak at 9.2 minutes from the large-scale extraction was purified as in Materials and Methods section 4.5.7 and sent for analysis by NMR. Analysis by Michael Stephenson (JIC) showed that this product has a triterpene scaffold, with a disaccharide moiety linked at the C-3 position (Figure 4.11). The hexose could be identified as being linked to the pentose moiety in a β-1,2-configuration. Therefore this strongly suggests that AsUGT91 has triterpene 3-*O* arabinose 2-*O* glucosyltransferase activity.

Instead of the expected 12,13 epoxide, the product has a carbonyl at the C-12 position. This degradation of the avenacin epoxide has been previously observed; early structures of avenacin aglycones described 12-ketones (Crombie et al., 1984a; Begley et al., 1986; Crombie et al., 1986a) before it was realised that the carbonyl forms were a result of the acid hydrolysis used to remove the sugar chain (Crombie et al., 1984b) and mild acid conditions that leave the sugar chain intact are sufficient to cause the rearrangement (Geisler et al., 2013). This suggests that the additional peak of the same mass present with co-expression of AsbAS1, AsCYP51H10, AsAAT1 and AsUGT91 in the small-scale analysis of *N. benthamiana* leaves (Figure 4.9D) may be the epoxide isomer, and that this peak was not seen in the large-scale extract as it may have fully degraded during the extraction process.

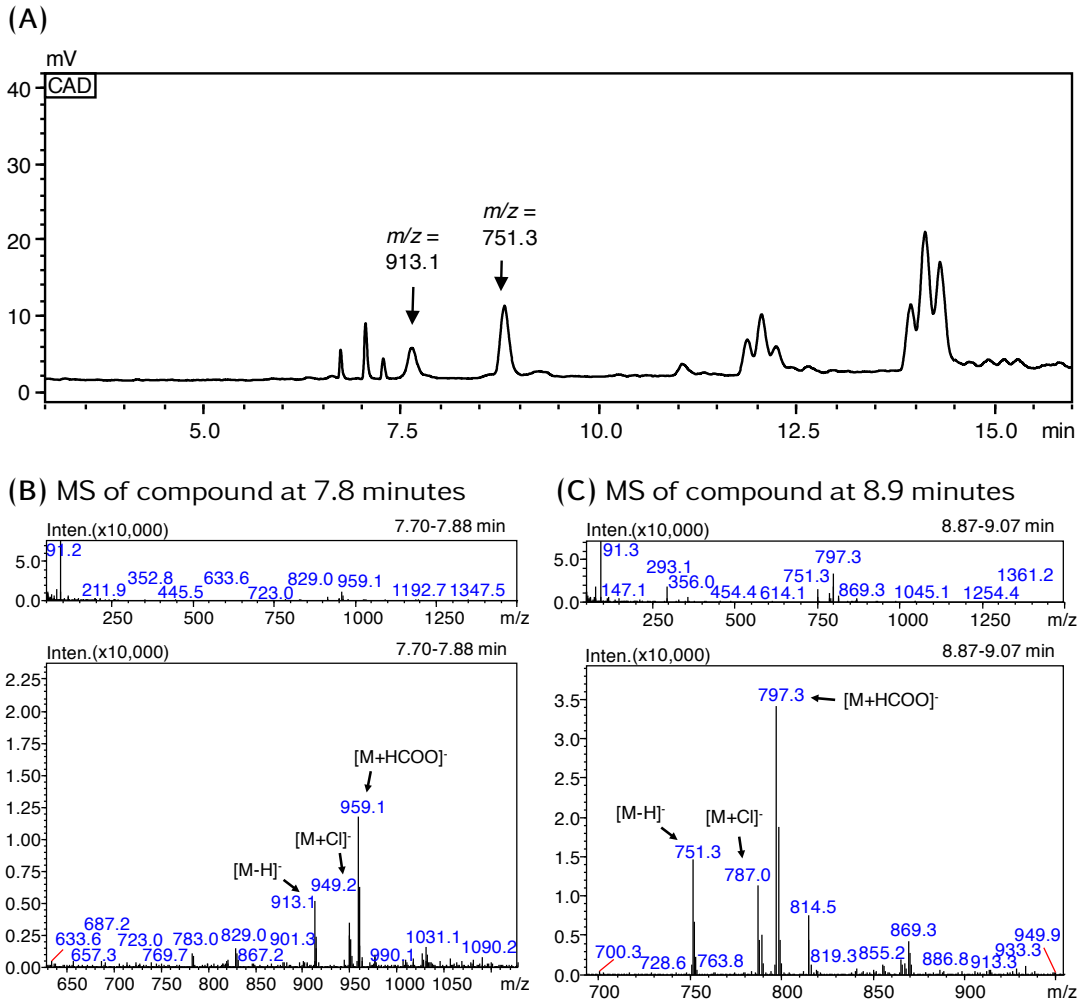


Figure 4.10: HPLC-CAD-MS analysis of the large-scale extraction of the AsUGT91 product.
(A) HPLC-CAD chromatogram of the extract from 150 *N. benthamiana* plants co-expressing AsbAS1, AsCYP51H10, AsAAT1 and AsUGT91. The accumulation of two product peaks can be detected with the mass of 12,13-β-epoxy, 16β-hydroxy-β-amyrin-3-*O*-α-L-arabinose (EpHβA-3-*O*-Ara) with the addition of one hexose [*m/z* = 751.3, retention time = 9.2 minutes] or the addition of two hexoses [*m/z* = 913.1, retention time = 7.8 minutes]. The mass spectra in negative mode of these two peaks **(B)** and **(C)** show signals consistent with the mass ion (-1, $[M-H]^-$), the chloride adduct (+35, $[M+Cl]^-$) and the formate adduct (+45, $[M+HCOO]^-$) of these two products.

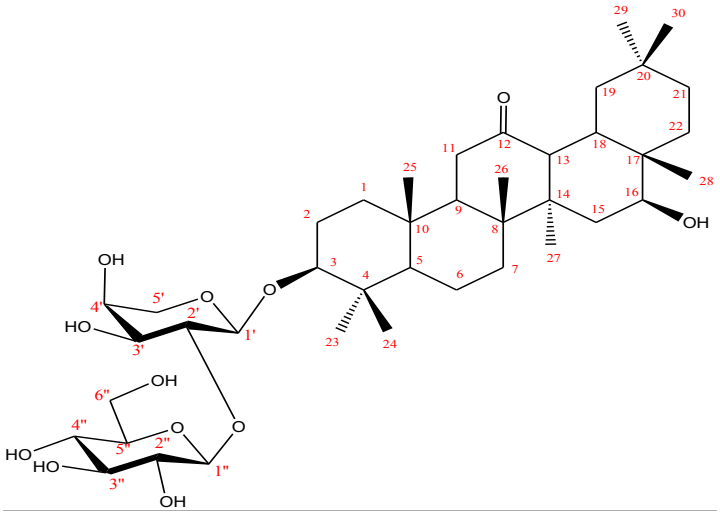


Figure 4.11: Structure of the AsUGT91 diglycoside product as analysed by NMR by Michael Stephenson, John Innes Centre.

4.3.4 Characterisation of *AsUGT91* mutants

The *in vitro* and *in planta* investigations reported in sections 4.3.2 and 4.3.3 indicate that AsUGT91 has triterpene-3-*O*-1,2-glycosyltransferase activity. To validate whether AsUGT91 is involved in the avenacin pathway, its role in *A. strigosa* was investigated by examining *A. strigosa* mutants.

Ten saponin-deficient (*sad*) mutants were found in the initial screen of sodium azide-induced mutants of *A. strigosa* that are deficient in avenacin production and have reduced root fluorescence under UV light (Papadopoulou et al., 1999). Further screening of the sodium azide-mutagenised population has revealed additional mutants (Qi et al., 2006). Some of these mutants are known to correspond to characterised *Sad* genes (Mylona et al., 2008; Mugford et al., 2009; Qin et al., 2010), but others correspond to unknown loci.

To determine whether any of these unknown mutants are avenacin-deficient due to mutations in the *AsUGT91* gene, the *AsUGT91* gene was amplified from 49 uncharacterised mutants (Materials and Methods section 4.5.8).

Table 4.1: Summary of SNPs in the AsUGT91 gene found for the uncharacterised mutants

Mutant	Mutation Event	Amino acid change
#85	G963A	W321-STOP
#543	G375A	W125-STOP
#1073	G776A	G259E
#1473	G775A	G259R

Sequencing of the *AsUGT91* gene revealed four mutants with single nucleotide polymorphisms (SNPs) (Table 4.1). Mutants #85 and #543 have SNPs that are predicted to result in premature termination of translation. Mutants #1073 and #1473 have undergone mutations in adjacent nucleotides, resulting in a change in the same amino acid, changing it from a glycine residue to a glutamic acid or an arginine, respectively. Sequence alignments to crystallised plant UGTs show that this residue is located in a loop in the protein C-terminal domain (the C1 loop) that contributes to the formation of the acceptor binding pocket and is associated with sugar-donor binding (Osmani et al., 2009). The glycine residue is part of a short sequence motif, a glycine followed by serine/threonine/arginine, that is conserved in many GT-B fold GTs, including bacterial enzymes (Osmani et al., 2009; Hu and Walker, 2002; Ha et al., 2000; Hoffmeister et al., 2001). The second residue in the motif is proposed to interact directly with the sugar-donor which may involve the mobility of the C1 loop (Osmani

et al., 2009). Therefore, mutations in the glycine residue might have an impact on substrate binding and catalytic activity, and could explain the loss of function of these mutants.

Previous LC-MS analysis of root extracts of mutants #85, #543, #1073 and #1473 showed an altered acyl donor profile, no detectable fully-glycosylated avenacins and two new peaks that are less polar than the avenacins (Rachel Melton, unpublished).

To verify the previous evaluation of these mutants, root extracts were compared to those of characterised *sad* mutants by HPLC-CAD-MS analysis (Materials and Methods section 4.5.9). The CAD chromatograms of all of the mutants tested is shown in Figure 4.12. Wild type (S75) root extracts accumulate the four avenacins, avenacin A-1, A-2, B-1 and B-2, the most abundant of which are avenacins A-1 and A-2 (Figure 4.12). *sad1* mutants are mutated in *AsbAS1* (Haralampidis et al., 2001) which catalyses the first committed step in the avenacin biosynthetic pathway, and therefore do not accumulate any avenacins (Figure 4.12). The genes responsible for the *sad3* and *sad4* mutant phenotypes are unknown; however *sad3* mutants accumulate mono-deglucosyl avenacins missing the 1,4-linked D-glucose, and *sad4* mutants accumulate a mixture of these mono-deglucosyl avenacins as well as fully glycosylated avenacins (Figure 4.12) (Papadopoulou et al., 1999; Mylona et al., 2008).

Analysis of the root extracts confirmed that mutants #85, #543, #1073 and #1473 share similar root metabolite profiles that are distinct from the wild type, *sad1*, *sad3* and *sad4* profiles. These mutants do not accumulate any avenacins, and instead accumulate two new products (m/z = 887.1 at 18.3 minutes and m/z = 916.1 at 20.5 minutes) that are less polar than either the avenacins or the mono-deglucosyl avenacins. The root extracts were also monitored by UV, showing that the peak at 20.5 minutes is UV-active like avenacins A-1 and B-1 (data not shown). To analyse the accurate mass of the two peaks, the root extract of mutant #85 was analysed by HPLC-MS-IT-TOF. This showed that these two peaks have masses consistent with avenacin A-2 and A-1 with the loss of one hexose and a further loss of 14 Da (Figure 4.13).

One modification that could account for the mass difference of 14 Da is the loss of the oxidation at the C-30 position. In the *N. benthamiana* system, the C-30 P450, AsCYP72A476, was unable to completely convert an EpHβA substrate (Reed, 2016). As other members of the same P450 class act on glycosylated substrates, it was speculated that AsCYP72A476 might be active after glycosylation of the avenacin scaffold (Reed, 2016). Root extracts of a mutant of the arabinosyltransferase *AsAAT1* also contain avenacin intermediates with a mass difference consistent with the loss of the C-30 oxidation (Louveau et al., manuscript in preparation) further supporting the hypothesis

that the C-30 P450 may not be active on avenacin intermediates that are insufficiently glycosylated.

Therefore mutants #85, #543, #1073 and #1473, which have SNPs in the *AsUGT91* gene, are all deficient in avenacin production and accumulate compounds that, based on mass and polarity, are likely to be avenacin intermediates missing a hexose.

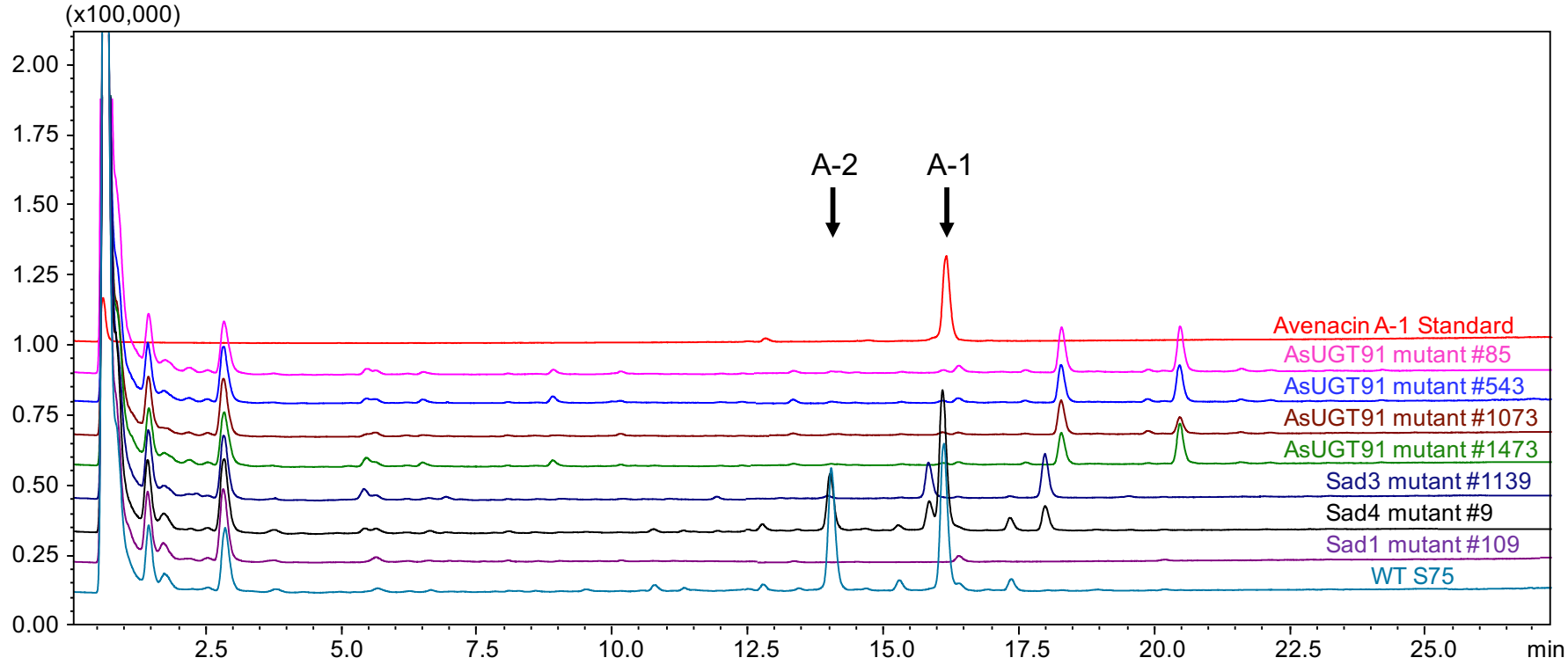


Figure 4.12: HPLC-CAD analysis of *sad* mutant root extracts.
HPLC-CAD analysis of methanolic root extracts of: WT (S75) roots contain the four avenacins: avenacin A-1, A-2, B-1 and B2 (turquoise trace) (Crombie and Crombie, 1986). *sad1* mutants (#109) are mutated in the first committed step of the avenacin biosynthetic pathway and do not accumulate any avenacins (purple trace). *sad4* mutants (#9) accumulate the four avenacins and mono-deglucosylated forms (black trace). *sad3* mutants (#1139) accumulate mono-deglucosyl avenacins (dark blue trace). Mutants (#85, #543, #1073 and #1473) do not accumulate avenacins but accumulate two major products (18.3 min and 20.5 min) that are less polar than the avenacins (green, dark red, blue and pink traces). Avenacin A-1 standard is shown in red. Avenacin A-1 (A-1) and Avenacin A-2 (A-2) are indicated with arrows. *sad* mutant numbers are as described in Papadopoulou et al. (1999).

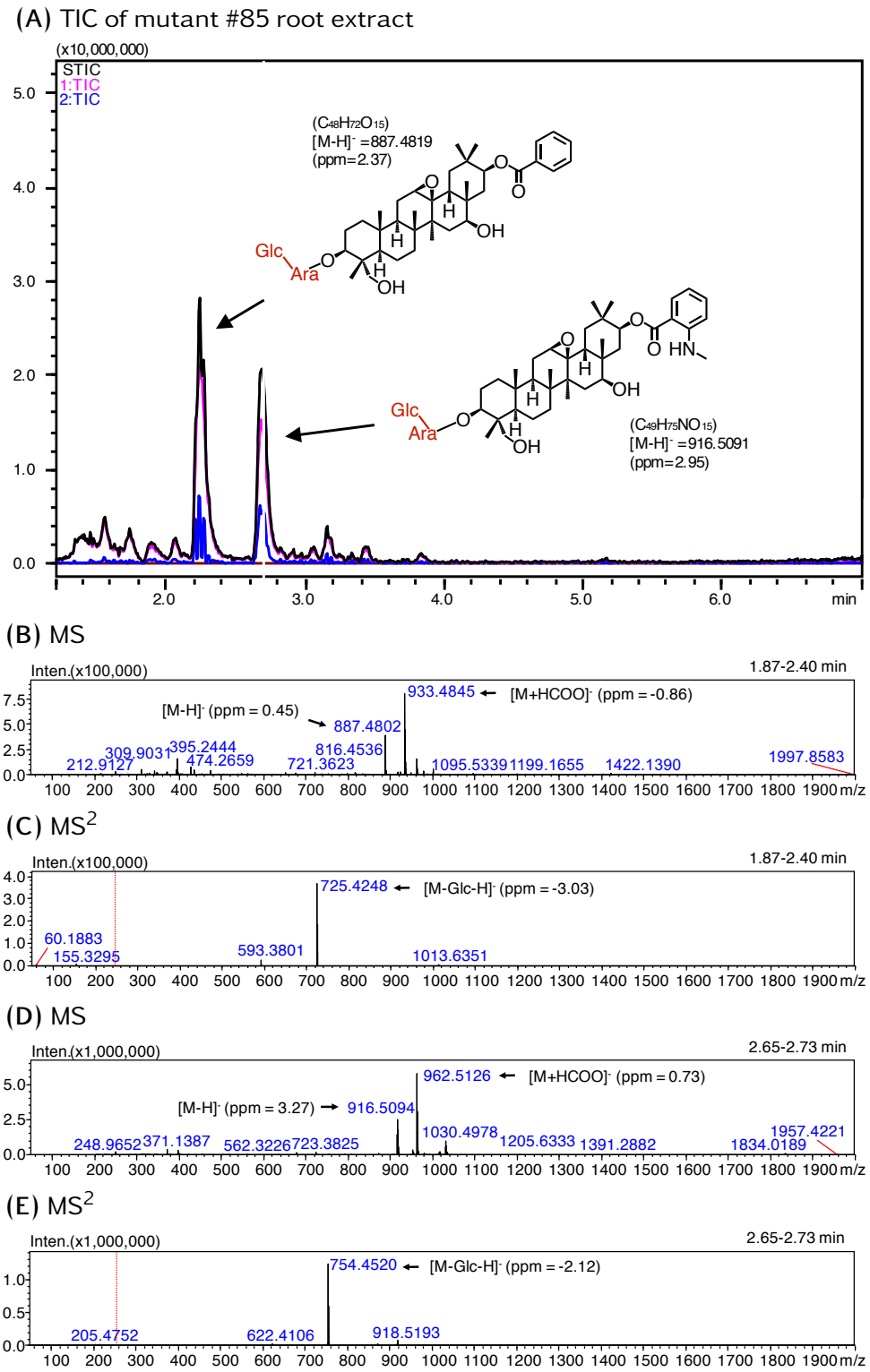


Figure 4.13: HPLC-IT-ToF analysis of mutant #85 root extract.
(A), Total ion chromatogram (TIC) of the #85 mutant root extract in negative mode shows two major peaks at 2.2 minutes and 2.7 minutes with the mass of avenacin A-2 and avenacin A-1 respectively with the loss of a hexose and the loss of the oxidation at the C-30 position. Predicted product structures are shown, Ara = L-arabinose, Glc = D-glucose.
The mass spectra (MS) of the peak **(B)** between 1.87-2.40 minutes and the peak **(D)** between 2.65-2.73 minutes shows signals consistent with the mass ion ($[M-H]^-$) and the formate adduct ($[M+HCOO]^-$) of the predicted product structures.
The MS^2 spectrum of the fragmentation of the mass ion of each predicted product structure **(C)** (precursor mass ion for fragmentation = *m/z* 887.4843) and **(E)** (precursor mass ion for fragmentation = *m/z* 916.5106) showed signals consistent with the loss of a glucose molecule ($[M-Glc-H]^-$). The difference in observed and expected mass values in parts per million is indicated (ppm).

To confirm whether the SNPs in the *AsUGT91* gene are directly responsible for the observed phenotype, F_2 progeny from mutant lines crossed to the *A. strigosa* wild type parent (S75) were analysed to test whether the respective SNP and the phenotype co-segregated. F_2 progeny of crosses to the wild type parent (S75) were available for two of the mutants, #543 and #1473 (Institute of Grasslands and Environmental Research, Aberystwyth, Wales, United Kingdom).

The roots of mutants #85, #543, #1073 and #1473 have reduced fluorescence as expected, since these mutants were originally identified in a screen for reduced root fluorescence. They also have shorter roots (Figure 4.14A). Analysis of F_2 seedlings from crosses of mutant #543 and #1473 with the wild type revealed that the reduced fluorescence phenotype consistently co-segregated with the short root phenotype (192 seedlings phenotyped for each F_2 population).

Table 4.2: F_2 progeny phenotype ratios

Mutant	wild type phenotype	Mutant phenotype	X^2 Wildtype:Mutant = 3:1 ($P > 0.05$)
#543	134	58	2.778
#1473	144	48	0

The segregation ratios for two F_2 populations phenotyped for root length were statistically consistent with the expected 3:1 ratio for a recessive mutation (Table 4.2). Subsequent sequencing of the phenotyped mutants showed that the mutation in the *AsUGT91* gene co-segregated with the short root phenotype in each population (n=192) (Figure 4.14). Together with the biochemical data, these results indicate that the mutations in the *AsUGT91* gene are likely to cause the observed phenotypes in these mutants, and that AsUGT91 is required for the biosynthesis of avenacin A-1.

The avenacin intermediates accumulated by the *AsUGT91* mutants may have toxic effects, which would account for the stunted roots of these mutants. Root growth defects were also present in *sad3* and *sad4* mutants, which were shown to be caused by the accumulation of toxic mono-deglucosylated avenacin intermediates (Papadopoulou et al., 1999; Mylona et al., 2008).

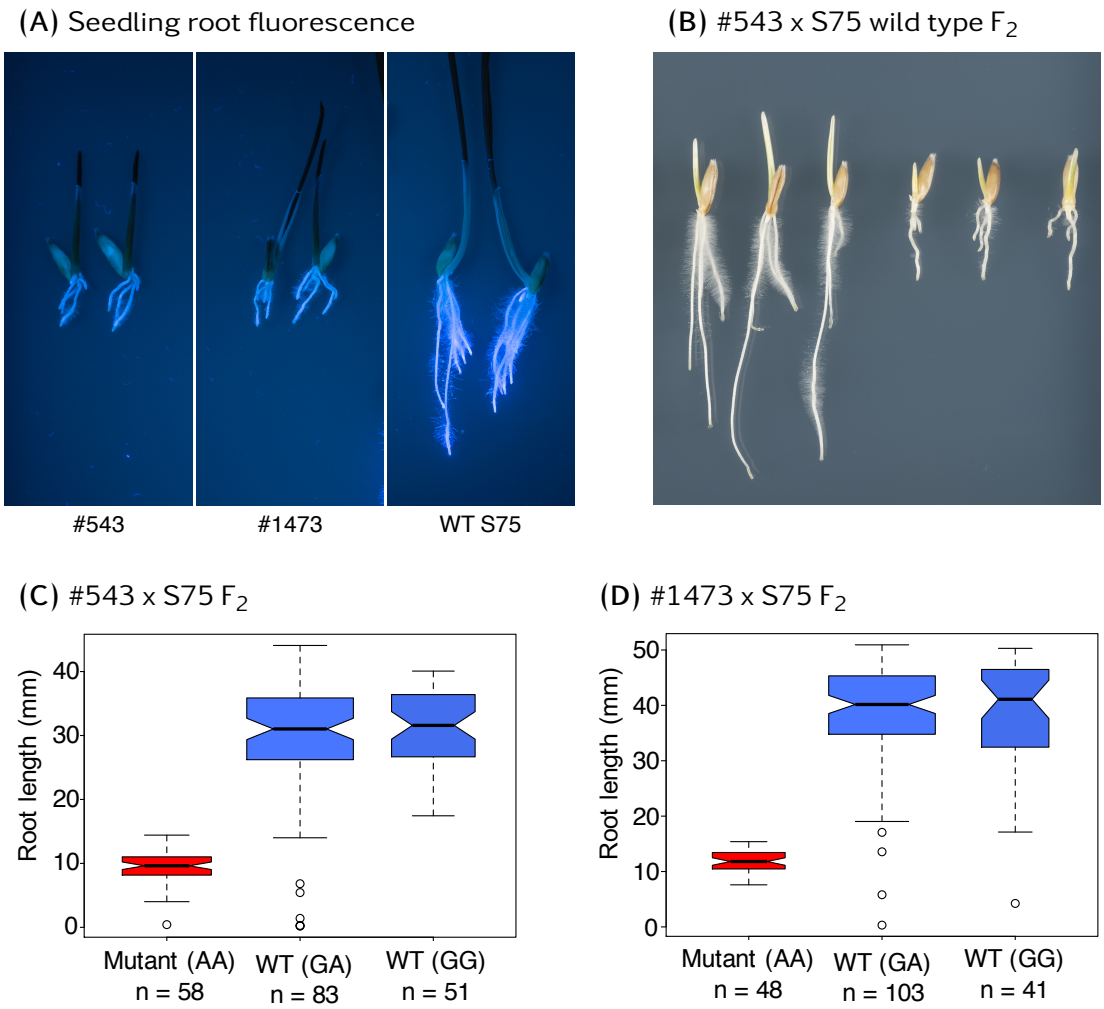


Figure 4.14: Root phenotype of mutants. **(A)**, Reduced fluorescence of roots of mutants #543 and #1473 compared to wild type seedlings (S75). Seedlings were grown by Rachel Melton. **(B)**, The roots of the *AsUGT91* mutant phenotype are significantly shorter than wild type roots. Four-day-old seedlings from an AsUGT91 #543 x S75 wild type F_2 population, showing wild type phenotype (left), and mutant phenotype of short roots (right). **(C)**, Boxplot of root lengths of the phenotypes of #543 x S75 F_2 population, **(D)**, boxplot of root lengths of the phenotypes of #1473 x S75 F_2 population. The genotypes of these seedlings were determined and are indicated in brackets, homozygous (GG) and heterozygous (GA) wild type seedlings have been separated into two groups. Photo credits: Andrew Davis, John Innes Centre.

The characterised *sad* mutants are compromised in their disease resistance to the wheat pathogen *Gaeumannomyces graminis* var. *tritici* compared to wild type seedlings (Papadopoulou et al., 1999). Therefore the disease resistance of the *AsUGT91* mutants was investigated.

The disease resistance of homozygous *AsUGT91* mutant lines to the take-all pathogen *G. graminis* var. *tritici* isolate T5 (Bryan et al., 1999) was compared to wild type *A. strigosa* S75 seedlings; *sad1* mutants, which do not synthesise avenacins (Haralampidis et al., 2001); *AsAAT1* mutants, which accumulate avenacin intermediates and a reduced amount of avenacins (Louveau et al., manuscript in preparation); *sad3* mutants, which only accumulate mono-deglucosylated avenacins and *sad4* mutants, which accumulate both fully glycosylated and mono-deglucosylated avenacins (Mylona et al., 2008).

Seedlings were incubated with plugs of actively growing *G. graminis* var. *tritici* and scored after 21 days based on the method in Bowyer et al. (1995) (Materials and Methods Section 4.5.11).

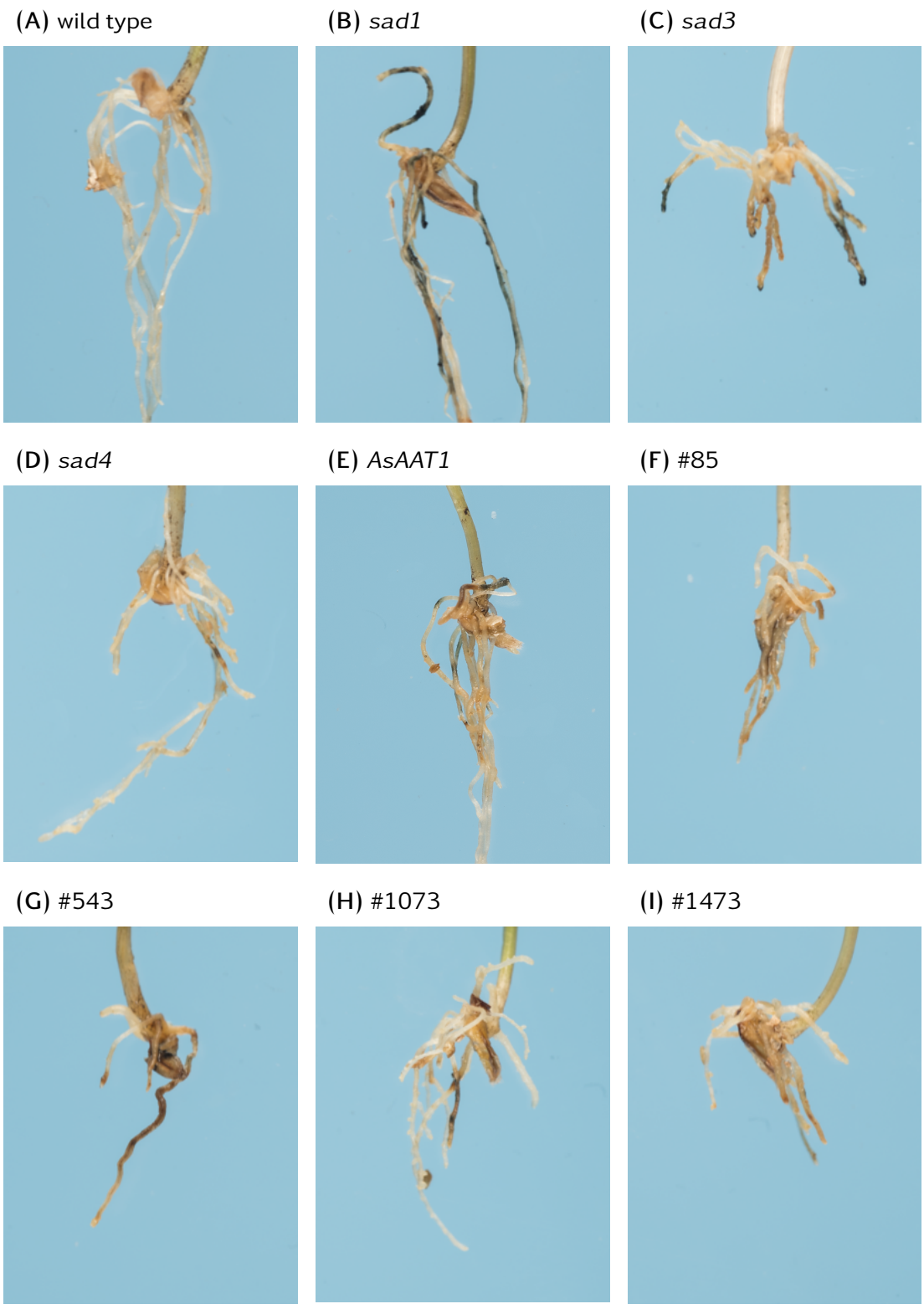


Figure 4.15: Representative take-all disease symptoms of WT and *sad* mutants.
Disease symptoms of seedlings incubated with plugs of actively growing *G. graminis* var. *tritici* for 21 days: **(A)**, wild type (S75); **(B)**, *sad1* mutant #109; **(C)**, *sad3* mutant #1139; **(D)**, *sad4* mutant #109; **(E)**, *AsAAT1* mutant #807; **(F)**, *AsUGT91* mutant #85; **(G)**, *AsUGT91* mutant #543; **(H)**, *AsUGT91* mutant #1073; **(I)**, *AsUGT91* mutant #1473. Photo credit: Andrew Davis, John Innes Centre

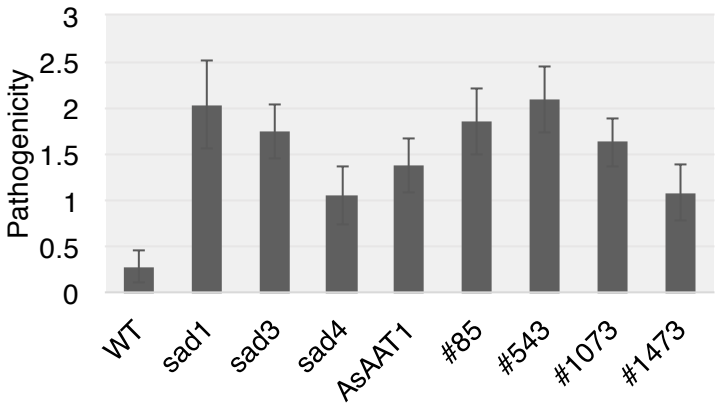


Figure 4.16: Pathogenicity of *G. graminis* var. *tritici* (isolate T5) to wild type and mutant oat seedlings. Graph shows mean pathogenicity scores (31-35 seeds per treatment; error bars represent 95% confidence limits).

All of the mutants including the *AsUGT91* mutant lines showed higher rates of disease when incubated with *G. graminis* var. *tritici* compared to wild type seedlings (Figures 4.15 and 4.16). *sad4* and *AsAAT1* mutants were less affected, presumably because these mutants still synthesise some avenacins.

4.4 Chapter 4 summary

A. strigosa UGT candidates were identified and prioritised in Chapter 3 by mining an oat root transcriptome database, phylogenetic analysis and expression profiling.

In this chapter, these candidates were heterologously expressed in *E. coli* and assayed for activity *in vitro* against bis-deglucosyl avenacin A-1. This revealed one candidate, AsUGT91, with positive activity. Transient expression of AsUGT91 in *N. benthamiana* and the structural determination of its product by NMR confirmed triterpene 3-*O* arabinose 1,2-glucosyltransferase activity, an activity which has not previously been described.

Mutants in the *AsUGT91* gene were identified in a sodium azide-induced *A. strigosa* mutant library. These were found to have root developmental defects, be deficient in avenacin production and show increased susceptibility to the take-all fungal pathogen, *G. graminis* var. *tritici*. Analysis of F_2 progeny of two *AsUGT91* SNP mutants backcrossed to the S75 wild type parent showed that *AsUGT91* is directly responsible for the observed phenotypes.

This evidence indicates that AsUGT91 is required for the biosynthesis of avenacins in *A. strigosa* and is necessary for disease resistance to the take-all pathogen, *G. graminis* var. *tritici*.

The identification of this enzyme increases the range of glycosyltransferases available to create novel triterpene glycosides by synthetic biology, and as one of the final unknown steps in the avenacin pathway, contributes towards engineering resistance in other crop species to the agriculturally important root disease, take-all.

NEW SECTION MATERIALS AND METHODS

4.5 Materials and Methods for Chapter 4

4.5.1 Extracting avenacin A-1 from oat roots

Seven litres of *Avena sativa* seeds were surface-sterilised for ten minutes in 14 litres of 0.5% sodium hypochlorite, rinsed thoroughly with tap water and drained. Seeds were distributed evenly onto sterilised aluminium gauze grids placed over plastic trays filled with tap water to one inch below the grids. Seeds were covered in two layers of Whatmann blotting paper dampened with sterile water, surrounded in foil and incubated at room temperature. The foil and blotting paper was removed after the seeds germinated. After seven days of growth, roots were harvested with a razor blade, freeze-dried and stored at -80°C. Root harvesting was repeated after a further seven days of growth. Freeze-dried roots were ground in liquid nitrogen and stored at -80°C.

Freeze-dried ground roots (2 x 17 g) were soaked overnight at 4°C in 500 ml 80% methanol, filtered through Miracloth (Merck) and Whatmann filter paper, and the methanol was evaporated with a rotary evaporator. The aqueous filtrate was precipitated overnight at 4 °C. The precipitate was collected by centrifugation at 3220 x g and freeze-dried. The precipitation process was repeated and the precipitates were combined.

The oat root filtrate precipitate was resuspended in methanol with sonication, dried onto diatomaceous earth (Celite, Sigma-Aldrich), and separated on a Biotage®SNAP C18 30g flash chromatography column with a flow rate of 25ml/min using a gradient of 5% methanol (Solvent A) and 95% methanol (Solvent B) as follows: 0-100% Solvent B over 750 ml and 100% Solvent B for 575 ml. Fluorescent fractions were combined and further purified as above with a gradient of 0-100% Solvent B over 1000 ml and 100% Solvent B for 525 ml. Fractions enriched for avenacin A-1 were determined by separation on a silica TLC plate and HPLC-UV-MS analysis as detailed below.

The oat root extracts were run on a Nexera/Prominence UHPLC system coupled to an IT-ToF mass spectrometer (Shimadzu) using a 50x2.1mm 2.6 µ Kinetex XB-C18 column (Phenomenex). Spectrometric detection was by UV absorbance collecting spectra from 190-300nm, at 6.25 Hz with a time-constant of 0.16 seconds. MS and MS^2 data were collected by electrospray in positive mode from *m/z* 170-1400. The column oven

was set at 40°C. The gradient was run at 0.6 ml/min with 0.1% TFA in water as Buffer A, 100% acetonitrile as Buffer B and was as follows: 20% Buffer B from 0- 3 minutes; 20-50% Buffer B from 3-15 minutes; 50-80% Buffer B from 15-23 minutes; a linear gradient between 23-27 minutes; 80 to 20% Buffer B from 27-28 minutes, and held at 20% Buffer B until 30 minutes.

4.5.2 Purification of avenacinase

Avenacinase was purified as described in Osbourn et al. (1991). Blocks of mycelium from actively growing colonies of *G. graminis* var. *avenae* strain A3 (Bryan et al., 1999) were placed on potato dextrose agar plates with 50 µg/ml streptomycin and 50 µg/ml ampicillin. After 5 days at 22 °C, colonies were scraped from the plates and homogenised with 1 ml Jermyn's medium (soluble starch: 1 g/L; K_2HPO_4: 6 g/L; NH_4Cl: 8 g/L; Yeast Extract: 1 g/L; $MgSO_4.7H_2O$: 1 g/L; $CaCl_2$: 0.02 g/L; $ZnSO_4.7H_2O$: 0.002 g/L; $MnSO_4.7H_2O$: 0.001 g/L) per colony. The homogenate was added to 2 litre flasks containing 500 ml Jermyn's medium with 50 µg/ml streptomycin and 50 µg/ml ampicillin (1 colony per 100 ml), and cultures were grown for 5 days at 22 °C with shaking at 200 rpm. Cultures were filtered through Miracloth (Merck) and two EDTA-free protease inhibitor tablets (Roche) per 500 ml filtrate was added. Filtrates were chilled to 4°C and ammonium sulphate added, with stirring to a final concentration of 580 l^{-1} of culture filtrate. The culture filtrate was centrifuged at 15 000 x g at 10 °C for 10 minutes, the supernatent was discarded and the pellet was resuspended in a minimum volume of ice-cold sterile water. The protein preparation was dialysed with four changes of buffer against 20 mM Tris-HCl pH 8 at 4 °C, centrifuged at 15000 x g at 10°C for ten minutes and frozen at -20°C.

4.5.3 Avenacin A-1 digestion and preparative thin-layer chromatography

The avenacin A-1 deglucosylation reaction contained the oat root extract enriched for avenacin A-1 (4.2 mg) (section 4.5.1), the avenacinase protein preparation (400 µl) (section 4.5.2) and 100 mM sodium acetate buffer pH 5 in a total volume of 13.4 ml. The reaction was incubated for 14 hours at 37°C, dried with a Genevac EZ-2 Elite centrifugal evaporator and stored at -20°C. The reaction was resuspended in 4.5 ml methanol and loaded onto the base of a 20x20 cm preparative silica thin layer chromatography (TLC) plate. The TLC plate was pre-run three times in 100% methanol 0.5 cm above the loading line, and then run in a mobile phase of dichloromethane:methanol:water (80:19:1; v:v:v). The position of each fluorescent

band was visualised under ultraviolet light and scraped off the plate with a scalpel blade, and filtered through filter paper with 15 ml methanol:ethyl acetate (25:75, v:v). The UV-active fractions were dried in a Genevac EZ-2 Elite centrifugal evaporator and stored at -20°C.

The bis-deglucosyl avenacin A-1 product was analysed on a Nexera/Prominence UHPLC system coupled to an IT-ToF mass spectrometer (Shimadzu) using a 50x2.1mm 2.6 µ Kinetex XB-C18 column (Phenomenex) and a column oven temperature of 30°C. MS and MS^2 data were collected by electrospray in positive mode from *m/z* 196-2000. Buffer A was 0.1% TFA in water and Buffer B was 100% acetonitrile. The gradient was: 0.5 ml/min; 25% Buffer B from 0-0.6 minutes; 25-80% Buffer B from 0.6-7 minutes; 80-100% Buffer B from 7-7.2 minutes; a linear gradient between 7.2-8 minutes; 100 to 25% Buffer B from 8-8.1 minutes, and held at 25% Buffer B until 10 minutes.

4.5.4 Gateway cloning of candidate enzymes

AsUGT91, AsGT24a3 and AsGT27f7 were amplified by two-step Gateway cloning (section 2.11.1) from genomic *Avena strigosa* S75 DNA (section 2.8) with gene-specific primers (Table 4.3). For *in vitro* assays, the amplified genes were cloned (sections 2.11.3 and 2.11.4) into the pH9-GW Gateway destination vector (section 2.6) and transformed into BL21 Rosetta *E. coli* cells (Invitrogen) (section 2.14). For transient expression in *N. benthamiana*, AsUGT91 was cloned (section 2.11.4) into the pEAQ-*HT*-DEST1 Gateway destination vector (section 2.6) and transformed into *A. tumefaciens* strain LBA4404 (section 2.15) .

Table 4.3: Primers used in Chapter 4

Name	Sequence
Fgw-UGT91	AAAAAGCAGGCTTAATGGCCGCCTCTGCTTCC
Rgw-UGT91	GAAAGCTGGGTATCAGTCCATGTAAGACGTGAGCTGCTG
Fgw-GT27f7	AAAAAGCAGGCTTAATGGCGACGCTGCCGGAGCTGCAC
Rgw-GT27f7	AGAAAGCTGGGTATTAAGACTGTACTGACAGTGC
Fgw-GT24a3	AAAAAGCAGGCTTAATGGGGAGCGAGCATCACGC
Rgw-GT24a3	AGAAAGCTGGGTATCACCTTGTGATGTAAGCGATGA
F-UGT91-0816	GCCCGCTACCTATTTGAATGGTGG
RoutUGT91	GTGTTGACCATGCACGAATCTCC
Rutr-UGT91-bis	GGAACCATATTGAAAAATCGCTTA
FUTR-UGT91-0516	TGTTTTTGTAAGCAGCGGGC
RUTR-UGT91-0516	AGGTAGTACACTCGCTCGCT
Rrt-UGT91	ACGACCAGCTGAAGCTTGCC

4.5.5 *In vitro* UGT enzymatic assays

Proteins were expressed as in Materials and Methods section 2.18. Reactions contained 50 mM Tris-HCl pH 7.5, 0.5 mM UDP-D-glucose (Sigma-Aldrich) and 150 µM bis- or mono-deglucosylated avenacin A-1 in a total volume of 50 µl. Reactions were started with the addition of 4 µg of the recombinant partially purified UGTs (section 2.18), incubated at 25°C overnight and were stopped by the addition of 50 µl methanol. Control reactions were set up as above, except UGT protein preparations had been boiled at 95°C for 10 minutes (inactive enzymes). The reaction mixtures were centrifuged and analysed by HPLC-MS (Section 2.18).

4.5.6 Agroinfiltrations

Agroinfiltrations were carried out using *A. tumefaciens* LBA4404 strains as in Material and Methods sections 2.19, 2.20 with detection by CAD-MS.

4.5.7 Large-scale extraction and purification of the AsUGT91 product

N. benthamiana plants (n=150) were vacuum infiltrated by Dr Thomas Louveau with a culture suspension mix of *A. tumefaciens* strain LBA4404 containing pEAQ-*HT*-DEST1-tHMGR; pEAQ-*HT*-DEST1-AsbAS1; pEAQ-*HT*-DEST1-AsCYP51H10; pEAQ-*HT*-DEST1-AsAAT1 and pEAQ-*HT*-DEST1-AsUGT91 as in Reed et al. (2017). After five days, the leaves were harvested and freeze-dried by Dr Thomas Louveau. The freeze-dried leaf material (40 g) was ground loosely in a pestle and mortar, and combined 1:1 v:v with quartz sand (0.3-0.9 mm). This mixture was layered in between 3 cm layers of quartz sand (0.3-0.9 mm) in a 120 ml extraction cell. Extraction was performed using a SpeedExtractor E-914 (Büchi) with 6 cycles at 90°C and 130 bar pressure. Cycle one (ethyl acetate) had zero hold time, cycle two (ethyl acetate) had five minutes hold time and cycles 3-6 (methanol) had five minutes hold time. The run finished with a two minute solvent flush and six minute N_2 flush.

The ethyl acetate and methanol fractions were analysed by HPLC-CAD-MS, using a 50x2.1mm 2.6 µ Kinetex XB-C18 column (Phenomenex) with a column oven temperature of 30°C. Detection was by charged aerosol detector (CAD, Corona Ultra RS from Dionex), as well as electrospray MS (Shimadzu LC-2020 dual source MS) collected in positive mode and negative mode from *m/z* 50 -1500. The gradient was run at 0.3 ml/min with 100% water as Buffer A, 100% acetonitrile as Buffer B and was as follows: 25% Buffer B from 0-1.5 minutes; 25-60% Buffer B from 1.5-21 minutes;

60-100% Buffer B from 21-21.5 minutes; a linear gradient between 21.5-23.5 minutes; 100 to 25% Buffer B from 23.5-24 minutes, and held at 15% Buffer B until 25 minutes.

The ethyl acetate and methanol fractions were combined and dried, resuspended in methanol, dried onto diatomaceous earth (Celite, Sigma-Aldrich), and separated through a column of silica gel 60 (Material Harvest) with DCM:MeOH (90:10, v:v) over 3 litres, DCM:MeOH (80:20, v:v) over 1 litre and DCM:MeOH (70:30, v:v) over 1 litre. Fractions containing the AsUGT91 product as assessed by thin layer chromatography were combined, dried onto diatomaceous earth (Celite, Sigma-Aldrich), and separated on a Biotage®SNAP KP-Sil 50 g column with a flow rate of 100 ml/min as follows; 100 % DCM (Solvent A) and 0% methanol (Solvent B) for 330 ml; 0-10% Solvent B over 500 ml; 10% Solvent B for 200 ml; 10-15% Solvent B over 260 ml; 15% Solvent B for 250 ml; 15-20% Solvent B over 230 ml and 20% Solvent B for 240 ml.

Fractions containing the AsUGT91 product as assessed by thin layer chromatography were combined, dried onto diatomaceous earth (Celite, Sigma-Aldrich), and separated on a Biotage®SNAP C18 30g column with a flow rate of 25 ml/min as follows: 45% water (Solvent A) and 55% methanol (Solvent B) for 165 ml; 55-80% Solvent B over 990 ml; 80-100% Solvent B over 33 ml; 100% Solvent B for 165 ml.

Fractions containing the AsUGT91 product as assessed by thin layer chromatography were combined, dried and sent for NMR analysis by Dr Michael Stephenson (JIC).

4.5.8 Screening uncharacterised mutants for mutations in the AsUGT91 gene

Purified gDNA (8-50 ng) (kindly provided by Rachel Melton) was amplified with primers F-UGT91-0816 and RoutUGT91 (Table 4.3) as Material and Methods Section 2.11, except with Q5®reaction buffer, Q5®High GC enhancer and Q5®High-Fidelity DNA Polymerase, an annealing temperature of 67°C and 40 amplification cycles. Purified PCR products were sequenced with F-UGT91-0816 and Rutr-UGT91-bis primers (Table 4.3) by GATC Biotech.

4.5.9 Root extract analysis

Mutant seedlings (20 per mutant) were grown on water agar plates as Materials and Methods Section 2.7 and whole roots were harvested from three-day-old seedlings (#85, #543, #1073 and #1473 mutant seedlings were 4-day-old due to short roots), flash-frozen in liquid N_2 and stored at -80°C. Frozen oat roots (26-27 mg) were

weighed into 2 ml screw-capped tubes. Two tungsten beads (3mm) were added with 500 µl 80% methanol and samples were incubated at 1400 rpm at 25°C for 1 hour, then the temperature was increased to 42°C for 30 minutes. Samples were centrifuged briefly at 16 000 x g and 450 µl was removed to a fresh 1.5 ml Eppendorf tube and partitioned twice with hexane. An aliquot (200 µl) of the methanolic fraction was dried down in a Genevac EZ-2 Elite centrifugal evaporator and resuspended in 100 µl methanol, filtered through Corning®Costar®Spin-X®centrifuge tube filters (Sigma-Aldrich). An aliquot (50 µl) of this filtrate was added to 50 µl of 50% methanol and analysed by HPLC-CAD-UV-MS using a 50x2.1mm 2.6 µ Kinetex XB-C18 column (Phenomenex) with a column oven temperature of 30°C. Detection was by charged aerosol detector (CAD, Corona Ultra RS from Dionex), UV/Vis absorbance collecting spectra from 200-500nm (Shimadzu SPD-M20A), and electrospray MS (Shimadzu LC-2020 dual source MS) collected in positive mode and negative mode from *m/z* 50-1500. The gradient was run at 0.3 ml/min with 100% water as Buffer A, 100% acetonitrile as Buffer B and was as follows: 20% Buffer B from 0-3 minutes; 20-60% Buffer B from 3-28 minutes; 60-100% Buffer B from 28-30 minutes; a linear gradient between 30-33 minutes; 100 to 20% Buffer B from 33-34 minutes, and held at 20% Buffer B until 35 minutes.

The *A. strigosa* mutant #85 root extract was run on a Nexera/Prominence UHPLC system coupled to an IT-ToF mass spectrometer (Shimadzu) using a 50x2.1mm 2.6 µ Kinetex XB-C18 column (Phenomenex). MS and MS^2 data were collected by electrospray in negative mode from *m/z* 200-2000. The column oven was set at 40°C. The gradient was run at 0.5 ml/min with 0.1% TFA in water as Buffer A and 100% acetonitrile as Buffer B and was as follows: 30-55% Buffer B from 0.01-6 minutes; 55-95% Buffer B from 6-8 minutes; a linear gradient between 8-9 minutes; 95 to 30% Buffer B from 9-9.1 minutes, and held at 30% Buffer B until 13.1 minutes.

4.5.10 F_2 population analysis

Seeds were grown as Materials and Methods section 2.7 on distilled water agar plates. Four-day-old seedlings were phenotyped for root length and imaged by Andrew Davis. Seedlings were grown for a further three days, and the phenotype verified before transfer to soil in 96-well trays. Seedlings were grown under glasshouse conditions for 1 week. Approximately 100 mg of leaf material was harvested and sent for genomic DNA extraction by Richard Goram. Genomic DNA (125-250 ng) was amplified as in section 4.5.8, except with FUTR-UGT91-0516 and RUTR-UGT91-0516

primers in a total volume of 25 µl (Table 4.3) and crude PCR products were sent for sequencing by PlateSeq Kit PCR (Eurofins) with Rrt-UGT91 primer (Table 4.3).

4.5.11 Take-all assays

Agar plugs containing actively-growing *G. graminis* var. *tritici* (isolate T5) inoculum were placed on 30 ml of loosely packed sterile moist vermiculite in 50-ml sterile plastic tubes and covered with a further 2 ml of wet vermiculite. Mock-inoculated tubes received plugs of sterile agar. Four-day old oat seedlings (section 2.7) were sown on top and covered with a further thin layer of vermiculite and the tubes sealed with Parafilm. Tubes were incubated at 22°C with a light-dark cycle of 16 hours of light and 8 hours of dark. Seedlings were carefully removed from the vermiculite 21 days after inoculation, and symptoms were scored. Pathogenicity was scored on an arbitrary score of 0 to 8; 0, no disease symptoms; 1, some browning of the roots, which may be nonspecific; 2, several lesions visible; 3, as 2, but with lesions confluent with seed; 4, as 3, with browning of the leaf sheath; 5, as 4, with more extensive browning of the leaf sheath; 6, extensive root necrosis and browning of the leaf sheath; 7, as 6 and leaves wilting and chlorotic; 8, as 6, and leaves brown and necrotic.

Chapter 5

Identification of AsTG: a triterpene transglucosidase

5.1 General introduction

In Chapter 4, AsUGT91 was identified as the triterpene 3-*O*-arabinoside 1,2-glucosyltransferase required for avenacin biosynthesis.

The recombinant UGT candidates that were screened in Chapter 4 were not able to glucosylate a mono-deglucosyl avenacin A-1 substrate that lacks the branching 1,4-D-glucose (data not shown), suggesting that none of these enzymes is the avenacin triterpene 3-*O*-arabinoside 1,4-glucosyltransferase that has yet to be characterised.

The triterpene 3-*O*-arabinoside 1,4-glucosyltransferase may be a UGT candidate that was not functionally assessed, or alternatively, the strategy in Chapter 3 of short-listing candidates based on sequence similarity to characterised triterpene glycosyltransferases may not have been effective in identifying this enzyme.

5.2 Aims

The aim of this work was to identify and characterise the enzyme responsible for the addition of the final D-glucose to the avenacin trisaccharide chain.

5.3 Results and Discussion

5.3.1 A glycosyl hydrolase family 1 member is immediately adjacent to the avenacin cluster

A method of identifying potential triterpene biosynthetic gene candidates is genomic linkage to the characterised genes of the avenacin A-1 biosynthetic pathway (Thimmappa et al., 2014), which are known to be clustered in the *A. strigosa* S75 genome (Qi et al., 2004).

The group of Dr Tim Langdon (IBERS, Aberystwyth University) is creating a genetic map of the *Avena* A-genome using a synteny-based approach (Mayer et al., 2009, 2011). Sequencing of parents and recombinant inbred lines (RILs) of crosses between two diploid species, *A. atlantica* and *A. strigosa*, has allowed the mapping of *A. atlantica* genomic sequences to large genomic segments. *AsUGT91* was identified as a gene adjacent to the closest markers to the avenacin pathway loci *Sad1* and *Sad2* in *A. atlantica*, mapping 0.68 centimorgans away (Dr Tim Langdon, personal communication). Thus it is likely that *AsUGT91* is in close proximity to the avenacin cluster in the genome.

The *A. strigosa* S75 genome has recently been sequenced (Bin Han, Chinese Academy of Sciences), confirming that the characterised avenacin biosynthetic genes map to the same scaffold, AS01_003827. AsUGT91 maps to a separate scaffold, AS01_006890, which could not be bridged to the AS01_003827 avenacin scaffold.

Table 5.1: Computationally predicted protein functions of genes on the AS01_006890 scaffold

Gene	predicted function
AS01_006890_0018824	UDP-glycosyltransferase (AsUGT91)
AS01_006890_0027185	Family 1 glycosyl hydrolase (AsGH1)
AS01_006890_0034277	RNA binding protein
AS01_006890_0095196	Myb domain DNA binding protein
AS01_006890_0101814	GH48 callose synthase

The AS01_006890 scaffold is 130 kilobases long and is predicted to encode five genes, the first of which is *AsUGT91* (Table 5.1). A predicted glycosyl hydrolase family 1 gene (AS01_006890_0027185), *AsGH1*, is located 7000 base pairs downstream from *AsUGT91*.

Intriguingly, this gene has a root tip-specific expression pattern as determined by RNA-seq and RT-PCR (Figures 5.1 and 5.2), which is similar to all of the characterised genes of the avenacin pathway (Haralampidis et al., 2001; Mugford et al., 2009; Owatworakit et al., 2013; Qi et al., 2006; Mugford et al., 2013). The last three predicted

genes on the scaffold, an RNA-binding protein, a Myb-domain DNA binding protein and a GH48 callose synthase, do not share this expression pattern (Figure 5.1).

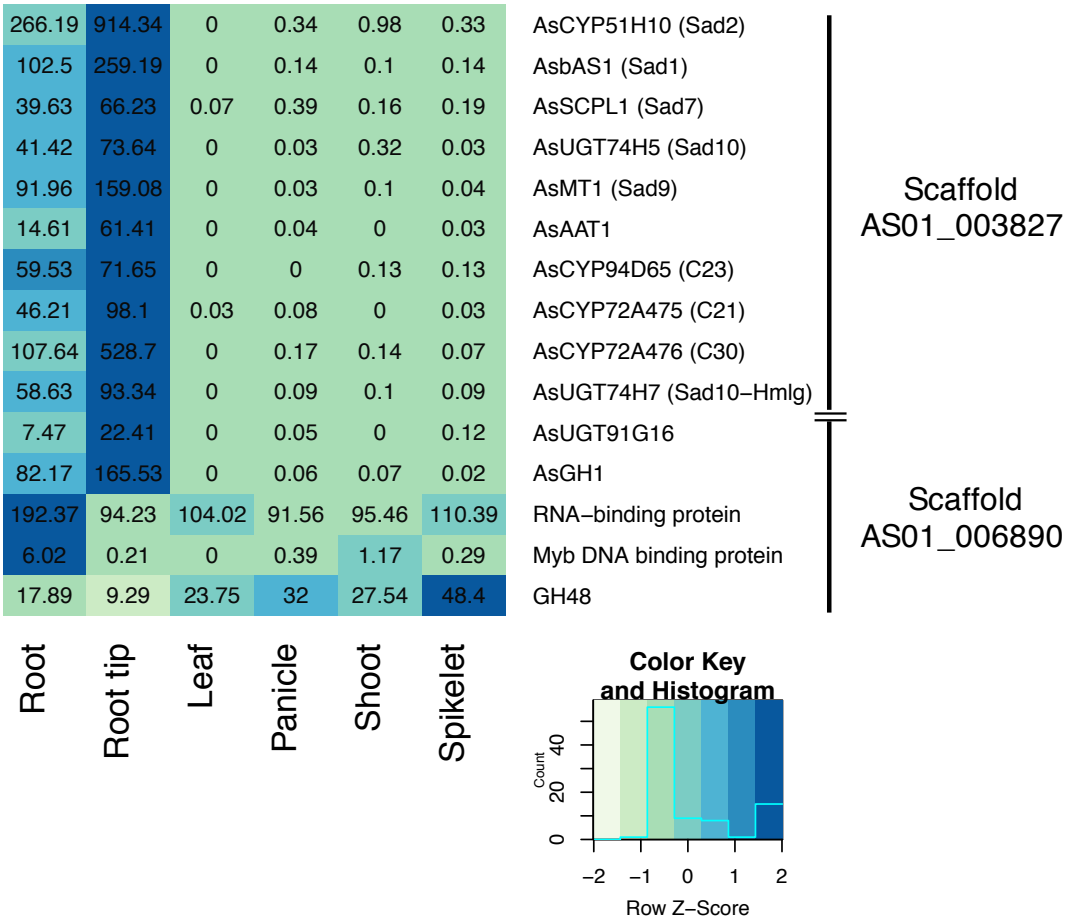


Figure 5.1: *AsGH1*, a gene predicted to encode a glycosyl hydrolase family 1 protein shares the same RNA-seq expression pattern as the avenacin cluster genes.
Heatmap of RNA-seq RPKM values (Bin Han, CAS) of genes on scaffolds AS01_003827 and AS01_006890. Genes are ordered according to their positions along their respective scaffolds. The characterised avenacin biosynthetic pathway genes cluster on the scaffold AS01_003827 and are expressed specifically in root tips. A predicted glycosyl hydrolase family 1 (GH1) enzyme, *AsGH1*, is located next to the avenacin sugar transferase *AsUGT91* on scaffold AS01_006890. Both of these genes share the root tip-specific expression pattern of the avenacin biosynthetic genes. RPKM values of each gene in each tissue are indicated.

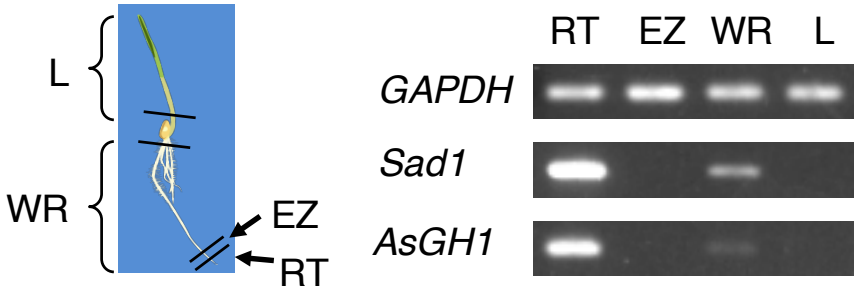


Figure 5.2: Analysis of the expression profile of the AsGH1 gene. RT-PCR expression profile of *AsGH1*. The profile of the previously characterised *AsbAS1* (*Sad1*) gene and the housekeeping oat glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) gene were included as positive controls. Oat tissues used are illustrated: root tips (RT), elongation zone (EZ), whole roots (WR) and young leaves (L). RNA was extracted from 3-day-old *A. strigosa* plants.

Considering that the genes on the AS01_006890 scaffold are genetically close to the avenacin cluster, this suggests that AsGH1 may play a role in the avenacin pathway.

GH1 enzymes hydrolyse glycosidic linkages with net retention of the anomeric configuration (retaining glycoside hydrolases) (Koshland, 1953). Retaining glycosyl hydrolases break glycosidic linkages in two steps, which allows these enzymes to catalyse both hydrolysis and transglycosylation. The first step is the reaction with the substrate to form a covalent glycosyl-enzyme intermediate with the release of a leaving group. The second step transfers the sugar to an acceptor molecule, which if water, leads to hydrolysis of the substrate, but in the case of other acceptor molecules leads to the formation of a new glycoside (Figure 5.3) (Sinnott, 1990; Bissaro et al., 2015).

Many retaining GHs have the ability to form glycosidic bonds, however they typically preferentially hydrolyse their substrates (Bissaro et al., 2015). Transglycosidase (TGs) that preferentially catalyse tranglycosylation are rare, but have arisen independently in multiple retaining GH families including GH Family 1 (Table 5.2).

GH1 TGs includes a galactosyltransferase, SENSITIVE TO FREEZING 2 (SFR2), which appears most closely related to GH1 enzymes from thermophilic bacteria and archaea (Thorlby et al., 2004; Moellering et al., 2010) and several specific transglucosidases that transfer D-glucose molecules with β-glycosidic linkages to small hydrophobic molecules such as plant specialised metabolites (Matsuba et al., 2010; Miyahara et al., 2012; Luang et al., 2013; Nishizaki et al., 2013; Miyahara et al., 2014). These latter enzymes fall into GH1 phylogenetic groups At/Os 6 and At/Os 7 as designated in Opassiri et al (2006) (Figure 5.5).

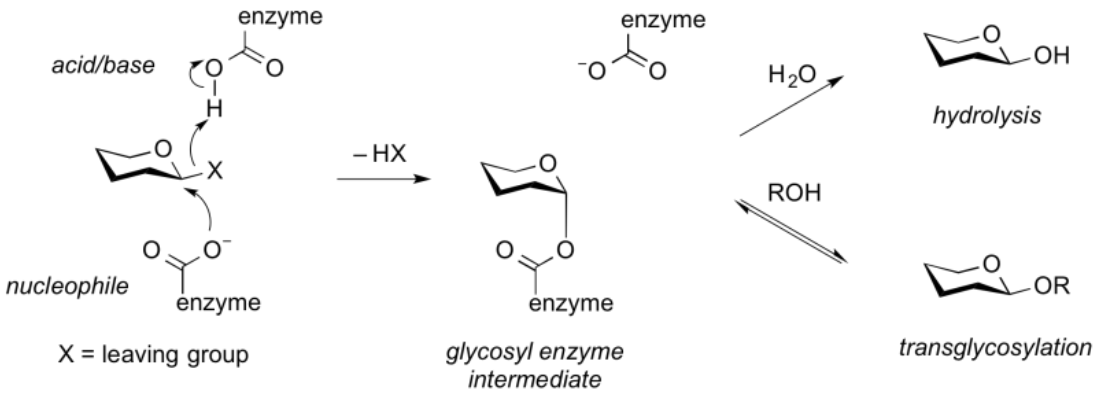


Figure 5.3: Generalised mechanism of a transglycosidase. Enzymatic cleavage of a substrate through a classical Koshland retaining mechanism results in formation of a glycosyl enzyme intermediate. This can partition to react with either water to cause hydrolysis (glycoside hydrolase activity) or to an alternative acceptor, often a sugar, to cause transglycosylation (transglycosylase activity). Figure and figure legend from cazypedia.org.

GH1 enzymes typically contain N-linked glycosylation sites and have predicted signal peptides to target them to the secretory pathway (Xu et al., 2004; Opassiri et al., 2006). Predicted cellular locations for GH1 enzymes in rice and *A. thaliana* include the cell exterior, cytoplasm, peroxisome, vacuole, ER lumen, ER membrane, plasma membrane and the mitochondrial matrix (Xu et al., 2004; Opassiri et al., 2006). Where this was investigated, all of the characterised TGs (excluding the chloroplastic SFR2) are predicted to be vacuolar (Matsuba et al., 2010; Luang et al., 2013; Nishizaki et al., 2013) (Table 5.2).

Analysis of the predicted protein sequence of AsGH1 shows that it contains the putative catalytic acid/base and nucleophilic glutamate residues of GH1 enzymes (Figure S8), and therefore has the potential to be catalytically active.

AsGH1 groups with the TG Os9Bglu31 in the phylogenetic group At/Os 6 (Figure5.5), suggesting that it may have transglycosylase activity and could therefore be directly involved in the glucosylation of avenacin A-1. Consistent with the vacuolar localisation of TGs, the AsGH1 protein contains an N-terminal 18-amino acid targeting sequence (MALLLCLFLFSLRLAALS) (SignalP 4.1 Server) which may target it to the secretory pathway (Figure 5.4).

```
AaAA7GT       ---MISYSLFFLLAFLFLYL-----VEFGISQSN--APKFSRDDFSSEFVFGAGTLAYQY
DgAA7GT       ---MCPS---FLVTLLLLQLSSLV-VVLVVWAEQLPEFNVRRDDFPSNFVFGAGTSALQV
CmAA7GT       M--LTQNQ---LKCHLHLLL--LV-VGVCTNNWDLTLADYSRLDFPSDFVFGAGTSAYQV
DcAA5GT       MNMSCKFEIVLLVSWWLLLV-----LVFGVESSM--FSEFDRLDFPKHFIFGASSCAYQV
Os9bglu31     ---MTPARVVFICCVVLL-------AAAAAAASSSTAAGITRADFPPEFIFGAGSSAYQV
AsGH1         M-------ALLLCLFLFS-------LRLAALSGDVVVAALTRRDFPDGFIFGAGTSSYQV
DgAA7BG-GT1   ---MGVMKIAYLVLDLFVVFNSIIFIPKPANPNQ-DSSAFDRNNFPVNFTFGVSSSAYQF
DgAA7BG-GT2   ---MGVMKLAYLIFDLFVMFNPIFFIPKPADHTELDSSALNRKSFPVNFTFGVASSAYQY
                        :                                * .*.  * **..: : *
```

Figure 5.4: AsGH1 and characterised transglucosidases have predicted N-terminal targeting sequences. N-terminal section of the full-sequence alignment between AsGH1 and the GH1 transglucosidases: BAM29304 AaAA7GT, from *Agapanthus africanus*; BAO96250 CmAA7GT from *Campanula medium*; E3W9M3 DgAA7GT, BAO04178 DgAA7BG-GT1 and BAO04181 DgAA7BG-GT2 from *Delphinium grandiflorum*; E3W9M2 DcAA5GT, from *Dianthus caryophyllus*; B7F7K7 Os9bglu31, from rice. Predicted N-terminal signal sequences are underlined in bold (Matsuba et al., 2010; Miyahara et al., 2012; Luang et al., 2013; Nishizaki et al., 2013; Miyahara et al., 2014).

Table 5.2: Characterised plant TGs

Enzyme name	Full name	Subcellular Localisation	Plant species	Reference
AtSFR2 (Q93Y07)	SENSITIVE TO FREEZING 2	chloroplast membrane	*Arabidopsis thaliana*	Moellering et al. (2010)
AaAA7GT (BAM29304)	Acyl-glucose-dependent anthocyanin 7-O-glucosyltransferase	not determined	*Agapanthus africanus*	Miyahara et al. (2012)
CmAA7GT (BAO96250)	Acyl-glucose dependent anthocyanin 7-O-glucosyltransferase	not determined	*Campanula medium*	Miyahara et al. (2014)
DgAA7GT (E3W9M3)	Acyl-glucose-dependent anthocyanin 7-O-glucosyltransferase	vacuolar	*Delphinium grandiflorum*	Matsuba et al. (2010)
DgAA7BG-GT1 (BAO04178)	Acyl-glucose-dependent anthocyanin acyl glucosyltransferase	vacuolar	*Delphinium grandiflorum*	Nishizaki et al. (2013)
DgAA7BG-GT2 (BAO04181)	Acyl-glucose-dependent anthocyanin acyl glucosyltransferase	vacuolar	*Delphinium grandiflorum*	Nishizaki et al. (2013)
DcAA5GT (E3W9M2)	Acyl-glucose-dependent anthocyanin 5-O-glucosyltransferase	vacuolar	*Dianthus caryophyllus*	Matsuba et al. (2010)
Os9bglu31 (B7F7K7)	Acyl-glucose-dependent flavonol/phytohormone /phenylpropanoid glucosyltransferase	vacuolar	*Oryza sativa Japonica Group*	Luang et al. (2013)

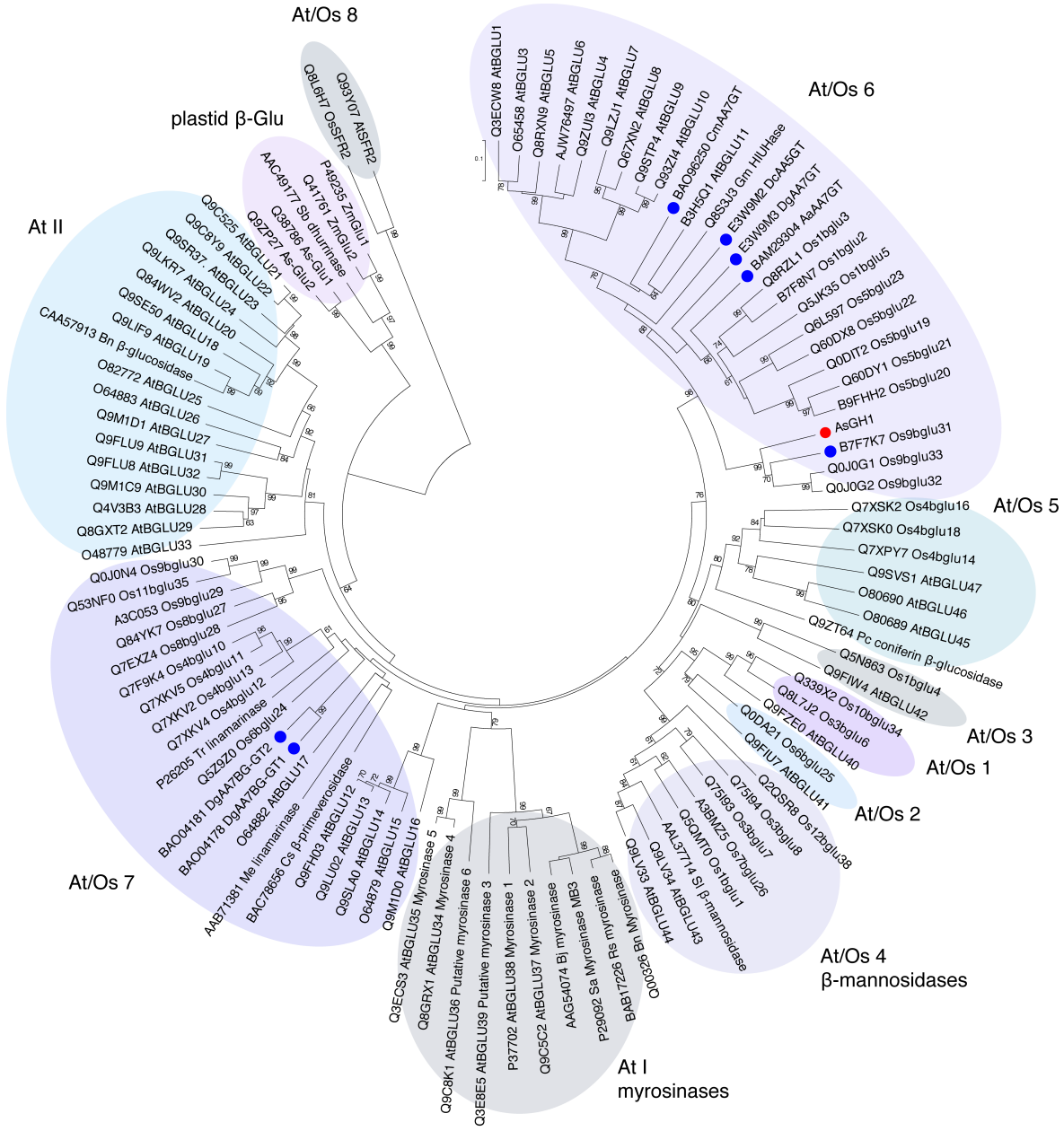


Figure 5.5: Phylogenetic tree of AsGH1 with other plant glycosyl hydrolase family 1 proteins. Protein sequences were aligned using MAFFT (https://mafft.cbrc.jp/alignment/software/) and the evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987) and analysed with 1000 bootstrap replicates which are shown at the nodes. The branches with <60% support were not marked. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016). GH1 family members of rice (Opassiri et al., 2006) and *Arabidopsis* (Xu et al., 2004) were included in the analysis, as well as sequences from other plant species as listed in Materials and Methods section 5.5.3. The phylogenetic clusters that include rice and Arabidopsis sequences (At/Os 1-8) and the Arabidopsis clusters (At I and II) are labelled as designated in (Opassiri et al., 2006), in addition to the monocot plastid β-glucosidases as in (Ketudat Cairns et al., 2012). The tree is drawn as an unrooted tree but is rooted by the outgroup, At/Os 8, which contains the chloroplastic *A. thaliana* AtSFR2 and the rice homologue, OsSFR2. Enzymes with reported transglycosylase activity are indicated (blue circles) and the *Avena strigosa* AsGH1 is highlighted (red circle).

5.3.2 Characterisation of recombinant AsGH1 *in vitro*

To assess the enzymatic activity of AsGH1, a similar strategy was used as for AsUGT91 in Chapter 4, by first investigating *in vitro* activity. *E. coli* was chosen as a heterologous host as successful expression of active His-tagged recombinant GH1 transglucosidases in *E. coli* strains has been previously described (Miyahara et al., 2012, 2014; Matsuba et al., 2010; Nishizaki et al., 2013; Luang et al., 2013).

The full-length AsGH1 coding sequence was amplified from *A. strigosa* root tip cDNA. The N-terminal 18-amino acid secretory targeting sequence (MALLLCLFLF-SLRLAALS) (SignalP 4.1 Server) was removed and replaced with a start codon, as the signal sequence of GH1 TGs has been reported to interfere with heterologous protein expression (Miyahara et al., 2014) and this approach has been used in previous cases of heterologous expression of GH1 TGs in *E. coli* (Miyahara et al., 2012, 2014; Matsuba et al., 2010; Nishizaki et al., 2013; Luang et al., 2013).

The resulting AsGH1 encoding sequence was cloned into the Gateway-compatible pH9-GW plasmid for IPTG-inducible expression of AsGH1 with an N-terminal 9xHis-tag, and this plasmid was introduced into BL21 Rosetta *E. coli* cells.

A preliminary assay was carried out to investigate suitable conditions to express the recombinant AsGH1 protein in *E. coli*. Six conditions were tested, varying the duration of protein expression after IPTG-induction and the temperature at which proteins were expressed (37°C for 1 hour, 2 hours or 5 hours; or 16°C for 2 hours, 5 hours or 21 hours).

Whole cell lysates of non-induced and IPTG-induced cultures were analysed by separation on an SDS-PAGE gel and staining with Coomassie blue. Recombinant AsGH1 was detectable as a ~58 kDa band in all induction conditions except 16°C for 2 hours (Figure 5.6A).

For lysates with detectable recombinant protein, the insoluble and soluble fractions were separated and the soluble fractions were purified further using Ni-NTA Agarose beads.

Analysis of the insoluble and partially purified soluble fractions showed a substantial amount of AsGH1 was soluble after induction for 5 hours at 16°C (Figure 5.6B). This induction condition yielded sufficient recombinant protein for *in vitro* enzymatic assays of crude extracts and therefore was used for subsequent expression of AsGH1 in *E. coli*. The relative amount of soluble recombinant protein was reduced with induction for 21 hours at 16°C, more of it being visible in the insoluble fraction (Figure 5.6B). Little of the recombinant protein was soluble with induction at 37°C (Figure

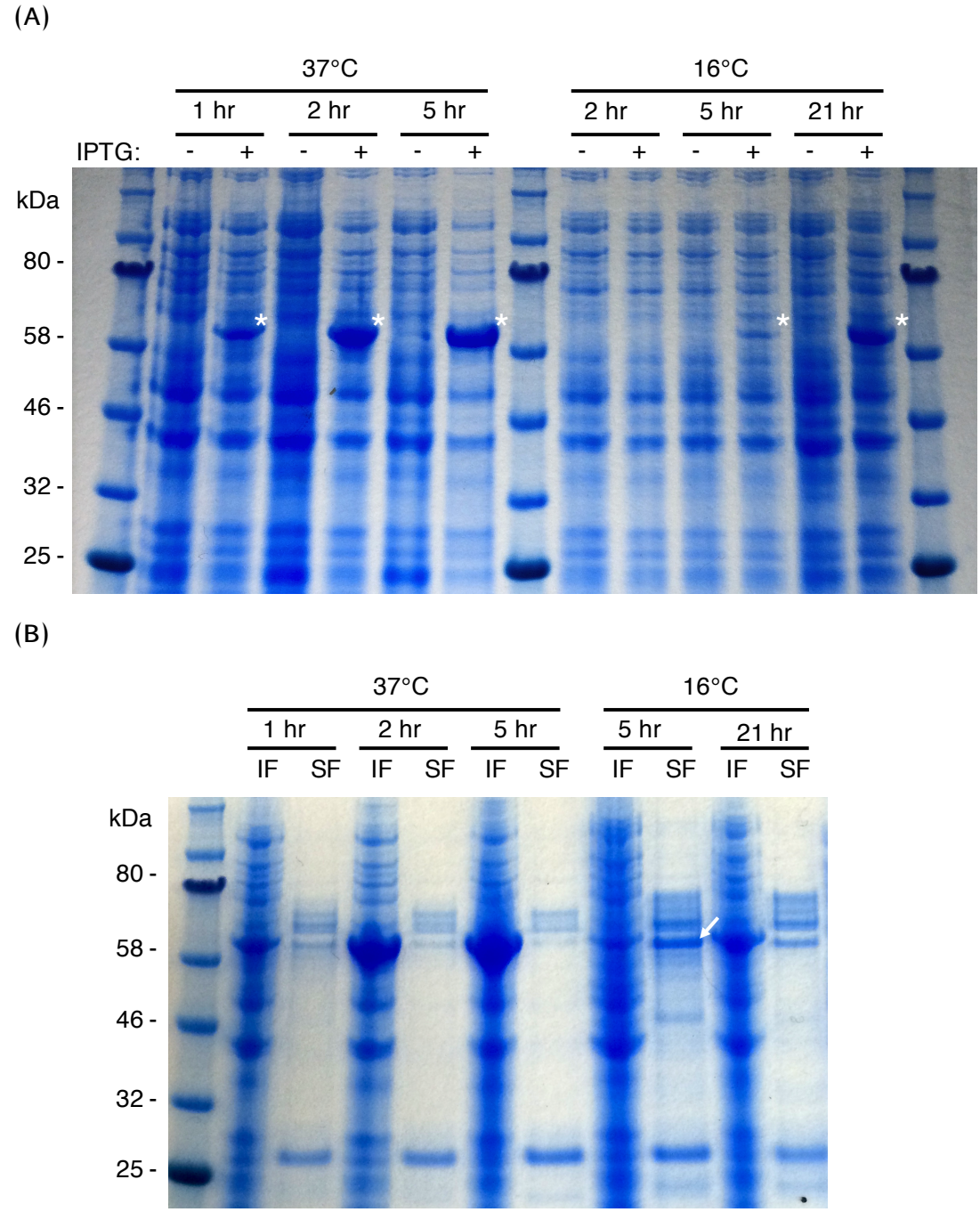


Figure 5.6: Evaluating conditions for the production of soluble recombinant AsGH1 in *E. coli*
E. coli Rosetta strains transformed with the pH9-GW-AsGH1 plasmid were grown to an OD_{600} of 0.5-0.6 at 37°C, then grown with (+) or without (-) the addition of 0.05 mM IPTG for 1, 2 or 5 hours at 37°C, or 2, 5 and 21 hours at 16°C. Cells induced at 16°C were acclimatised at 16°C for 30 minutes prior to IPTG addition.
(A), Cell lysates were visualised on a Coomassie-stained SDS-PAGE gel. Recombinant AsGH1 is expressed as a visible band (indicated with white asterisks) in all conditions except induction at 16°C for 2 hours.
(B), IPTG-induced cell cultures that were induced at 1, 2 or 5 hours at 37°C, or 5 and 21 hours at 16°C were lysed by sonication and the cell lysates were centrifuged at 12 000 x g for 20 minutes at 4°C. The soluble fractions were partially purified with Ni-NTA Agarose beads to enrich His-tagged proteins. The insoluble fractions (IF) and partially purified soluble fractions (SF) were visualised on a Coomassie-stained SDS-PAGE gel. The best conditions for obtaining soluble AsGH1 was induction at 16°C for 5 hours (white arrow). The estimated molecular weight of recombinant AsGH1 is 58 kDa.

5.6B), again, most of it being present in the insoluble fraction. This suggests that the protein might aggregate under these induction conditions.

The characterised GH1 TGs all use 1-*O*-acyl β-D-glucose esters as sugar donors (Miyahara et al., 2012, 2014; Matsuba et al., 2010; Nishizaki et al., 2013; Luang et al., 2013). Acyl-sugars are not commercially available, however transglycosylase activity was also detected *in vitro* for the rice TG, Os9BGlu31, using the readily available model β-glucosidase sugar donor substrate, 4-nitrophenyl β-D-glucoside (4NPGlc) (Figure 5.7) (Luang et al., 2013).

AsGH1 has close homology to Os9BGlu31, therefore there is a possibility that AsGH1 might be able to use this glycosidic substrate as a sugar-donor in an *in vitro* assay.

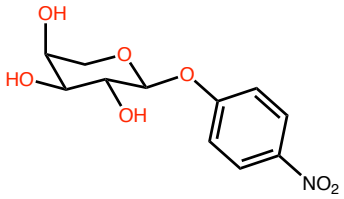


Figure 5.7: Structure of 4-nitrophenyl β-D-glucoside (4NPGlc)

The ability of AsGH1 to glycosylate the avenacin A-1 intermediates bis- and mono-deglucosyl avenacin A-1 (Chapter 4) when provided with 4NPGlc as a sugar donor was tested.

Previous *in vitro* assays for GH1 TGs have typically used citrate buffer at pH 5.5-6, at 30-35°C, with 100-200 µM of acceptor substrate and 1-2.25 mM of acyl-sugar donor (Miyahara et al., 2012, 2014; Matsuba et al., 2010). Five mM of 4NPGlc was used in the Os9BGlu31 assays (Luang et al., 2013). Therefore the preliminary assay to test AsGH1 was carried out at 30°C in 50 mM citrate buffer pH 5.75, using 150 µM of the acceptors bis- or mono-deglucosyl avenacin A-1 and 5 mM of the sugar donor 4NPGlc.

HPLC-UV-MS analysis showed that AsGH1 had negligible activity towards the acceptor bis-deglucosyl avenacin A-1 (Figure 5.8C). Nonetheless, AsGH1 yielded a new more polar peak at 4.4 minutes with mono-deglucosyl avenacin A-1 (Figure 5.8E). This peak has a mass ion ($[M-H]^-$) *m/z* of 1092.1, which is consistent with the mass of fully-glycosylated avenacin A-1.

This suggests that AsGH1 has transglycosidase activity towards the mono-deglucosyl avenacin A-1 intermediate *in vitro*. This enzyme was subsequently referred to as AsTG.

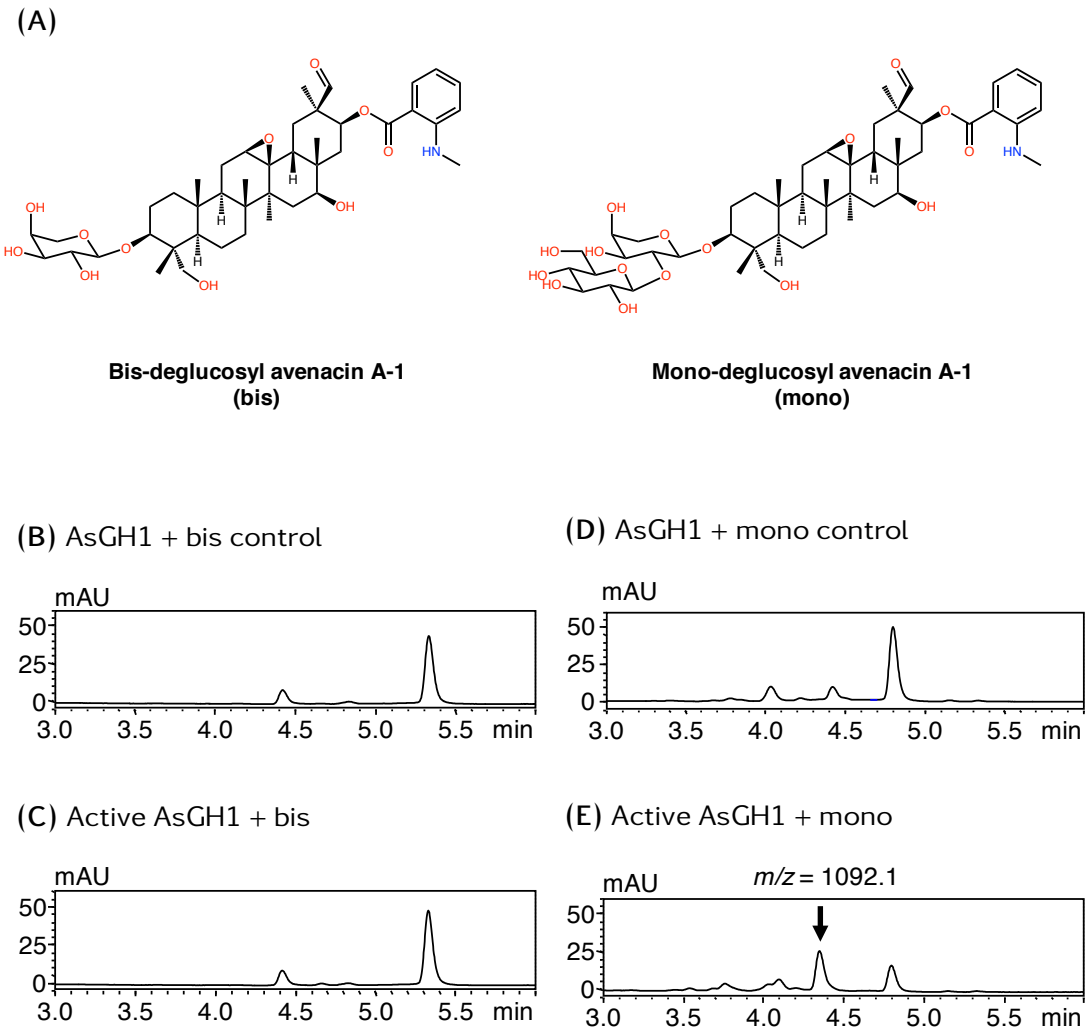


Figure 5.8: AsGH1 is active against mono-deglucosyl avenacin A-1.
(A), Structures of the sugar acceptors bis-deglucosyl avenacin A-1 (bis) and mono-deglucosyl avenacin A-1 (mono).
HPLC-UV-MS analyses of the enzymatic reactions with 4-nitrophenyl β-D-glucoside with: **(B)**, boiled protein preparation of AsGH1 (control reaction) and bis-deglucosyl avenacin A-1; **(C)**, AsGH1 and bis-deglucosyl avenacin A-1; **(D)**, boiled protein preparation of AsGH1 (control reaction) and mono-deglucosyl avenacin A-1; **(E)**, AsGH1 and mono-deglucosyl avenacin A-1.
The enzymatic reaction with active AsGH1 resulted in a new peak at 4.4 minutes ($[M\text{-}H]^-$ m/z = 1092.1). Protein preparations for control reactions had been boiled at 95°C for 10 minutes. Data are representative of two separate experiments. Absorbance was measured at 357nm (Begley et al., 1986).

5.3.3 Transient expression of AsTG with avenacin pathway enzymes in *N. benthamiana*

The activity of AsTG was evaluated by transient expression in *N. benthamiana*. Transient expression in *N. benthamiana* leaves of the avenacin oxidosqualene cyclase, AsbAS1 (SAD1); the two P450s, AsCYP51H10 (SAD2) and AsCYP72A475 (SAD6); and the two UGT glycosyltransferases, AsAAT1 and AsUGT91, leads to the accumulation of β-amyrin oxidised at three positions and glycosylated at the C-3 position with a dissaccharide sugar chain, (C-21-hydroxy-EpHβA-3-*O*-Ara-1,2-Glu[1]) (Figure 5.10 and Chapter 4).

Demonstration of activity in *N. benthamiana* following expression of the full length protein would indicate that AsTG is likely to be targeted to the correct subcellular location for activity. To investigate the role of the predicted N-terminal signal sequence, two AsTG constructs were cloned, both with and without the predicted N-terminal signal sequence (MALLLCLFLFSLRLAALS) (SignalP 4.1 Server) (Figure 5.9).

Figure 5.9: Schematic representation of the AsTG constructs used in the *N. benthamiana* expression assay. Construct diagrams are not drawn to scale. SP, AsTG predicted N-terminal signal peptide sequence. Grey boxes represent protein sequences introduced by Gateway cloning.

Co-expression of the full-length AsTG with AsbAS1, AsCYP51H10, AsCYP72A475, AsAAT1 and AsUGT91 resulted in the appearance of a new more polar peak in the HPLC-CAD-MS chromatogram with the expected mass ion of C-21-hydroxy-EpHβA-3-*O*-Ara-1,2-Glu with the addition of a hexose (Figure 5.10B, m/z = 929.2, 7.2 minutes).

[1] C-21-hydroxy-EpHβA-3-*O*-Ara-1,2-Glu = 12,13-β-epoxy, 16β,21β-dihydroxy-β-amyrin-3-*O*-α-L-arabinose-1,2-β-D-glucose. C-21-hydroxy-EpHβA-3-*O*-Ara = 12,13-β-epoxy, 16β,21β-dihydroxy-β-amyrin-3-*O*-α-L-arabinose.

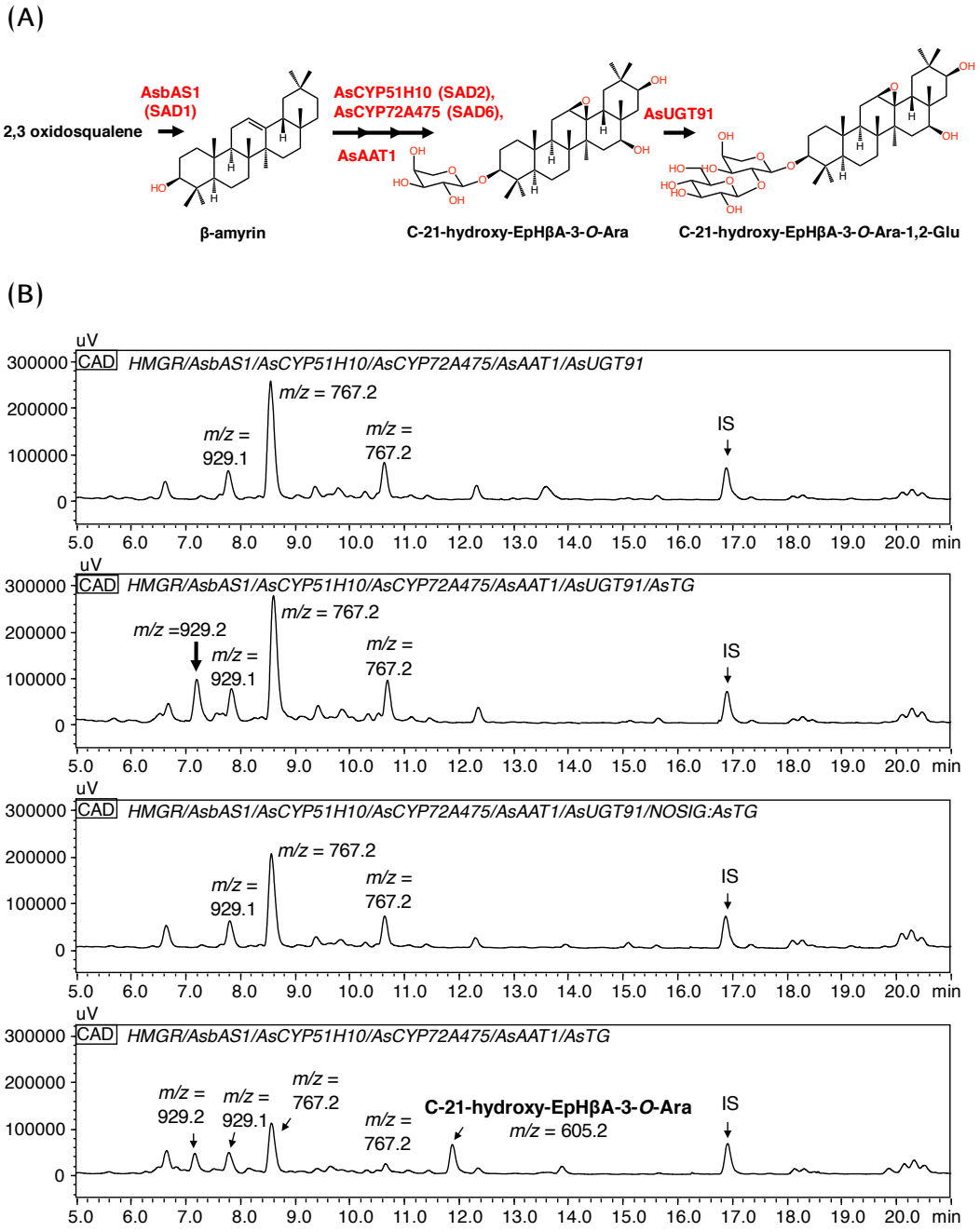


Figure 5.10: Expression of AsbAS1, AsCYP51H10, AsCYP72A475, AsAAT1, AsUGT91 and AsTG in *N. benthamiana* leaves.

(A) Formation of C-21-hydroxy-EpHβA-3-*O*-Ara-1,2-Glu from 2,3-oxidosqualene by AsbAS1 (SAD1), AsCYP51H10 (SAD2), AsCYP72A475 (SAD6), AsAAT1 and AsUGT91 (Chapter 4) (Haralampidis et al., 2001; Qi et al., 2006; Reed, 2016; Leveau et al., manuscript in preparation; Louveau et al., manuscript in preparation).

(B) HPLC-CAD analysis of extracts from infiltrated *N. benthamiana* leaves. Co-expression of AsbAS1, AsCYP51H10, AsCYP72A475, AsAAT1 and AsUGT91 results in the accumulation of peaks with the mass of C-21-hydroxy-EpHβA-3-*O*-Ara-1,2-Glu‡. The addition of AsTG results in a new peak with the mass of C-21-hydroxy-EpHβA-3-*O*-Ara-1,2-Glu with the addition of one hexose [*m/z* = 929.1, retention time = 7.2 min]. This peak was not present when the N-terminal signal peptide of AsTG was deleted. Expression of AsbAS1, AsCYP51H10, AsCYP72A475, AsAAT1 and AsTG without AsUGT91 lead to a reduction in the new peak at 7.2 minutes and an increase in the accumulation of the precursor C-21-hydroxy-EpHβA-3-*O*-Ara. Data are representative of two experiments (Supplementary Figure S4). IS = internal standard (digitoxin). C-21-hydroxy-EpHβA-3-*O*-Ara = 12,13-β-epoxy, 16β,21β-dihydroxy-β-amyrin-3-*O*-α-L-arabinose; C-21-hydroxy-EpHβA-3-*O*-Ara-1,2-Glu = 12,13-β-epoxy, 16β,21β-dihydroxy-β-amyrin-3-*O*-α-L-arabinose-1,2-β-D-glucose.

‡Top panel is in common with Figure 4.9D.

No new peaks were evident when the AsTG signal peptide was deleted (Figure 5.10B) suggesting that the signal peptide is necessary for AsTG activity. Without the signal peptide AsTG may be localised to a subcellular compartment where it is not active or in contact with the substrates, or it may not be properly folded *in planta*.

AsTG activity appears to be dependent on AsUGT91, although the analysis is complicated by the action of endogenous *N. benthamiana* glycosyltransferases on C-21-hydroxy-EpHβA-3-*O*-Ara[1] that have similar activity to AsUGT91 (see Chapter 4). When full-length AsTG is co-expressed with AsbAS1, AsCYP51H10, AsCYP72A475 and AsAAT1 without AsUGT91, there is accumulation of the C-21-hydroxy-EpHβA-3-*O*-Ara intermediate and the new product peak at 7.2 minutes is present but reduced. This suggests that endogenous *N. benthamiana* glycosyltransferases are able to glycosylate C-21-hydroxy-EpHβA-3-*O*-Ara to provide a suitable substrate for AsTG, but at a lower rate than AsUGT91, leading to the accumulation of the C-21-hydroxy-EpHβA-3-*O*-Ara precursor.

AsTG was also able to glucosylate 12,13-β-epoxy, 16β-hydroxy-β-amyrin-3-*O*-α-L-arabinose-1,2-β-D-glucose (EpHβA-3-*O*-Ara-1,2-Glu) which does not have the oxidation at the C-21 position (Figure S5).

Therefore AsTG shows transglycosidase activity towards avenacin pathway intermediates when transiently co-expressed in *N. benthamiana*. It also is able to utilise unknown endogenous *N. benthamiana* sugar donors that may differ from the *A. strigosa* donor(s).

5.3.4 Investigation of the subcellular localisation of AsTG

Analysis of fluorescent protein fusions to full-length or to N-terminal signal sequences of previously characterised GH1 transglucosidases have indicated that these enzymes are vacuolar (Luang et al., 2013; Matsuba et al., 2010; Nishizaki et al., 2013). The transglucosidase activity of AsTG *in planta* appears to rely on its subcellular location as deletion of the predicted N-terminal signal sequence abolishes its activity *in planta* (Section 5.3.3). AsUGT91 does not have any predicted signal sequences and is predicted to be cytosolic.

To investigate the subcellular localisation of AsUGT91 and AsTG, a series of fluorescent protein constructs using green fluorescent protein (GFP) and red fluorescent protein (mRFP1) under the control of the CaMV 35S promoter were made (Figure 5.11) (Curtis and Grossniklaus, 2003; Karimi et al., 2002). AsTG constructs with mRFP1 were made with and without the predicted N-terminal signal sequence (Figure 5.11) to investigate the location of the AsTG protein in each case.

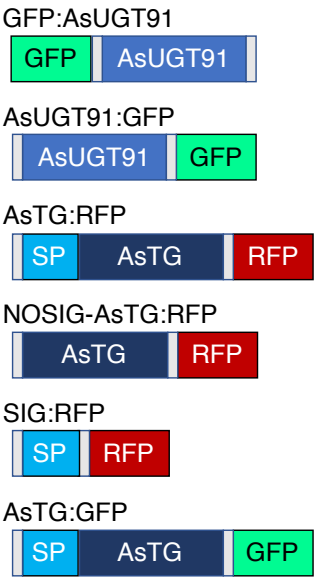


Figure 5.11: Schematic representation of fluorescent protein fusion constructs. Construct diagrams are not drawn to scale. SP, AsTG N-terminal signal peptide sequence; GFP, green fluorescent protein; RFP, monomeric red fluorescent protein. Grey boxes represent linkers introduced by Gateway cloning.

The fluorescent protein fusion constructs were introduced into *A. tumefaciens* and the fluorescent fusions to AsUGT91 or AsTG were evaluated for enzymatic activity in the *N. benthamiana* expression system (Figure 5.12).

When transiently co-expressed with the avenacin biosynthetic pathway en-

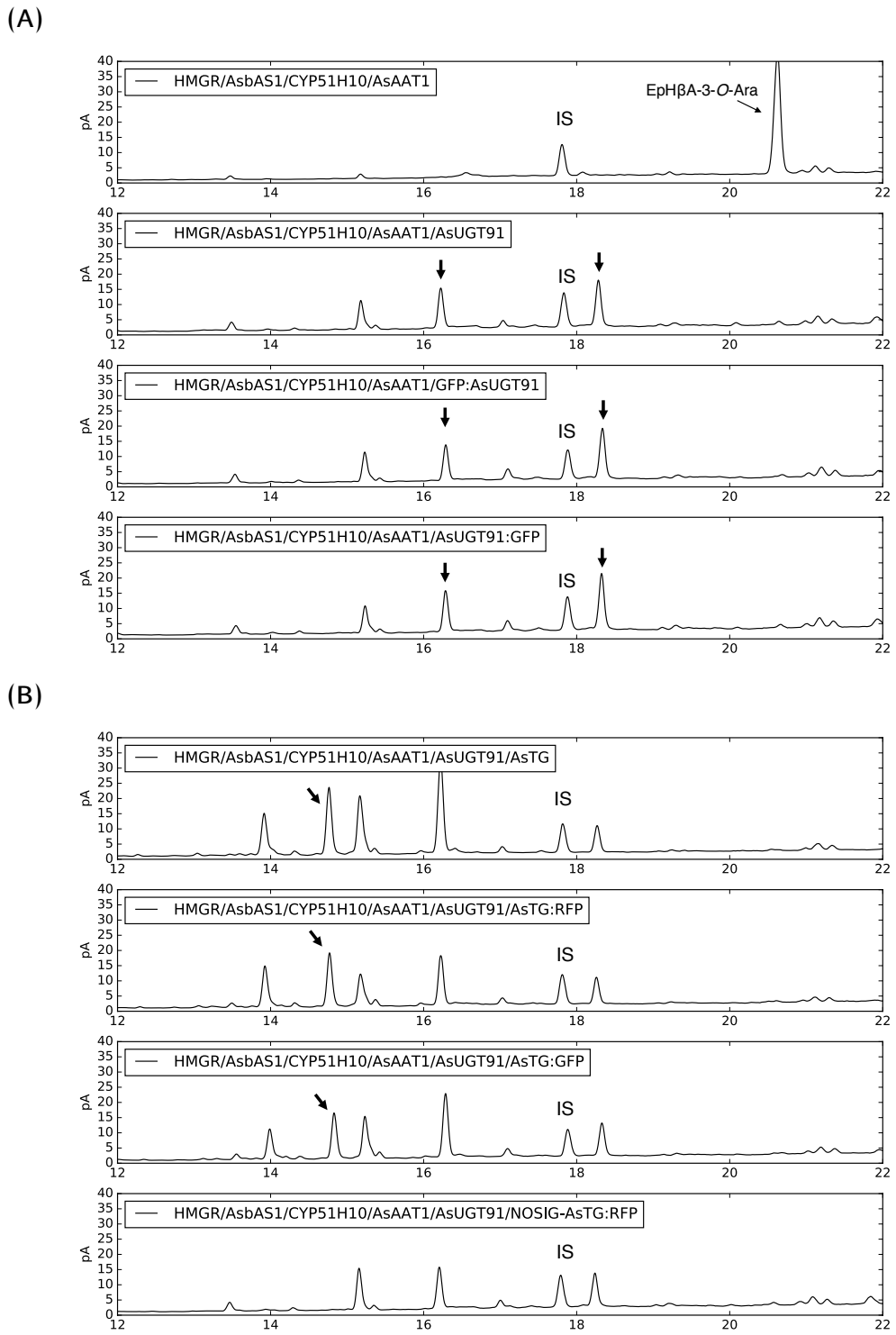


Figure 5.12: Expression of fluorescent protein fusions in *N. benthamiana*.
HPLC-CAD analysis of extracts from infiltrated *N. benthamiana* leaves. **(A)** The GFP-tagged AsUGT91 fusion constructs are active in *N. benthamiana* and accumulate the same compounds as AsUGT91 with no fluorescent tag. Accumulation of 12,13-β-epoxy, 16β-hydroxy-β-amyrin-3-*O*-α-L-arabinose (EpHβA-3-*O*-Ara) was detected in leaves expressing AsbAS1, AsCYP51H10 and AsAAT1. Co-expression of AsbAS1, AsCYP51H10 and AsAAT1 with: AsUGT91, GFP:AsUGT91 or AsUGT91:GFP show the accumulation of AsUGT91 products (black arrows) .
(B) Co-expression of AsbAS1, AsCYP51H10, AsAAT1 and AsUGT91 with: AsTG, AsTG:RFP, AsTG:GFP and NOSIG-AsTG:RFP. Fusions of GFP or mRFP1 to the C-terminus of AsTG (AsTG:RFP and AsTG:GFP) accumulate the same product peak (black arrow) as AsTG without a fusion tag. The RFP fusion protein to AsTG without the N-terminal signal sequence (NOSIG-AsTG:RFP) was not active.
Additional more polar peaks are present which are likely to be due to the action of endogenous *N. benthamiana* glycosyltransferases. Data are representative of two experiments. IS = internal standard (digitoxin).

zymes AsbAS1 (SAD1), AsCYP51H10 (SAD2) and AsAAT1 in *N. benthamiana*, both N-terminal and C-terminal GFP fusions to AsUGT91 showed similar activity to AsUGT91 without a fusion tag (Figure 5.12) (Section 5.3.3). This suggests that these proteins are folded correctly and are in the correct subcellular location for enzymatic activity.

The C-terminal GFP and RFP fluorescent fusion constructs of the full-length AsTG (containing the N-terminal signal peptide) were also active in the *N. benthamiana* expression system, showing similar activity to AsTG without a fusion tag when co-expressed with AsbAS1, AsCYP51H10, AsAAT1 and AsUGT91 (Figure 5.12) (Section 5.3.3). The fluorescent fusion constructs of AsTG with the N-terminal signal peptide deleted were not active (Figure 5.12), suggesting that these constructs are mis-targeted within the cell or are inactive for other reasons such as mis-folding. Therefore the fluorescent fusion tags do not affect the activity or trafficking to the correct compartment of AsUGT91 or the full-length AsTG.

To investigate the subcellular localisation of AsUGT91, *N. benthamiana* leaves were infiltrated with *A. tumefaciens* strains containing expression constructs for free mRFP1 (Moglia et al., 2014) and the two AsUGT91 GFP fusion proteins. Confocal microscopy of leaves two days post-infiltration showed that both the N-terminal and C-terminal GFP fusions to AsUGT91 were co-localised with free mRFP1 in the cytoplasm and nucleus (Figure S6). The predicted combined size of the N- or C-terminal GFP tags to AsUGT91 (76 kDa and 78 kDa respectively) are within the molecular size range that is known to passively diffuse through the nuclear pores (Wang and Brattain, 2007). Leaves imaged after three days showed the same co-localisation pattern with mRFP1 (data not shown).

To investigate the localisation of the AsTG constructs in comparison to AsUGT91, *N. benthamiana* leaves were infiltrated with the N-terminal fusion of GFP to AsUGT91 (GFP:AsUGT91) and the RFP fusion proteins to AsTG with and without the N-terminal signal sequence (AsTG:RFP and NOSIG-AsTG:RFP).

Three days post-infiltration, the RFP fusion protein to AsTG (AsTG:RFP) could be seen in the vacuole, as well as a signal in the apoplastic space in between the cells (white arrow, Figure 5.13B). In contrast, the RFP fusion to AsTG with the N-terminal signal sequence deleted (NOSIG-AsTG:RFP) co-localised with GFP:AsUGT91 in the cytoplasm and nucleus (Figure 5.13C). This suggests that the N-terminal signal sequence is necessary for the localisation of AsTG, and that the loss of enzymatic activity of the AsTG constructs without the N-terminal signal sequences *in planta* is likely to be due to mis-localisation of AsTG in the cytoplasm.

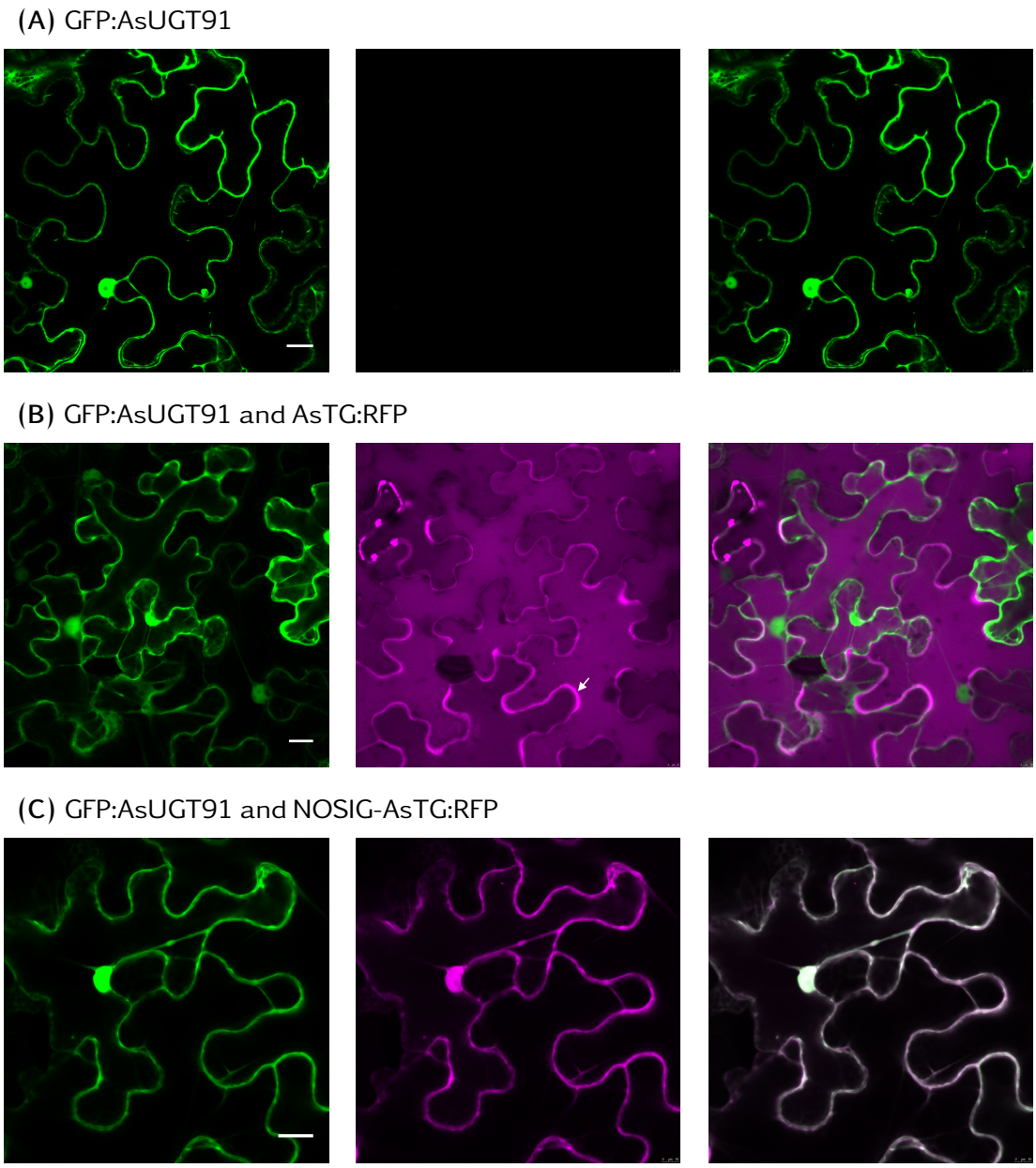


Figure 5.13: AsTG localises to the vacuole and the apoplast in *N. benthamiana* leaves.
Co-expression of an N-terminal GFP fusion to AsUGT91 (GFP:AsUGT91): **(A)**, alone **(B)**, with AsTG:RFP, and **(C)**, with NOSIG-AsTG:RFP. An RFP fusion to AsTG localises to the vacuole and the apoplast (white arrow) and does not co-localise with a GFP fusion protein to AsUGT91 in the cytoplasm. An RFP fusion to AsTG without the N-terminal signal peptide (NOSIG-AsTG:RFP) co-localises with the GFP fusion to AsUGT91 in the cytoplasm and nucleus. No signal is seen in the RFP channel when no RFP construct is co-infiltrated. GFP fusions are shown in green (left); RFP fusions are shown in magenta (middle) and merged images are shown in white (right). Images are taken three days post-infiltration. Bar = 20 µm. Image credits: Ingo Appelhagen, John Innes Centre.

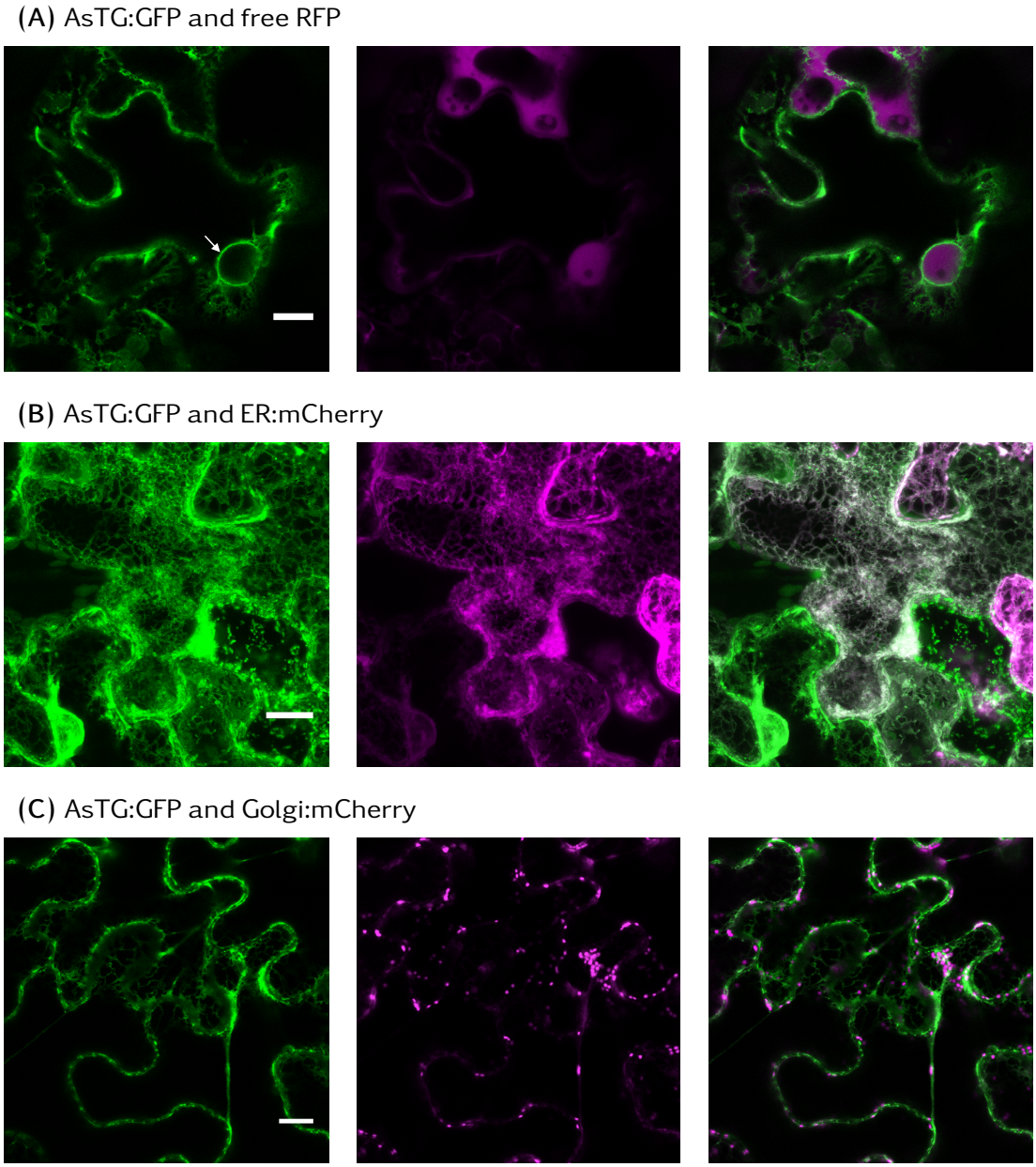


Figure 5.14: AsTG traffics through the ER.
Co-expression of an C-terminal GFP fusion to AsTG (AsTG:GFP) with: **(A)**, free RFP; **(B)**, ER:mCherry, and **(C)**, Golgi:mCherry. One day post-infiltration, a GFP fusion to AsTG co-localises with a marker for the ER, and does not co-localise with free RFP in the cytoplasm or nucleus. No co-localisation was seen with a marker for the Golgi. GFP fusions are shown in green (left); RFP/mCherry fusions are shown in magenta (middle) and merged images are shown in white (right). Images are taken one day post-infiltration. Scale bars = 10 µm. Image credits: Ingo Appelhagen, John Innes Centre.

Analysis of infiltrated leaves after one and two days post-infiltration suggested that AsTG:RFP might be also be localised in the ER (data not shown). This was investigated further by co-infiltrating the GFP protein fusion to AsTG (AsTG:GFP) with an mCherry marker targeted to the ER (ER:mCherry) or the Golgi (Golgi:mCherry) (Nelson et al., 2007).

The ER is continuous with the nucleus and surrounds it on all sides, and is seen as a lace-like cortical network adjacent to the plasma membrane and as large dynamic subcortical strands (Staehelin, 1997; Collings, 2013; Nelson et al., 2007). One day post-infiltration, AsTG:GFP can be seen surrounding the nucleus (arrow in Figure 5.14A) and co-localised with the ER:mCherry marker in the ER (Figure 5.14B). AsTG:GFP did not co-localise with free RFP in the cytoplasm and nucleus (Figure 5.14A) or Golgi:mCherry in the Golgi, which can be seen as multiple individual small (<1µm) stacks (Figure 5.14C) (Nelson et al., 2007).

Two days post-infiltration, AsTG:GFP could still be seen in the ER (data not shown). After two days, the Golgi:mCherry marker was mis-localised in other subcellular compartments which is indicative of over-expression (Sparkes et al., 2006; Collings, 2013) and it was not clear if the AsTG:RFP localises to this compartment (data not shown).

AsTG appears to be targeted to the endomembrane system, as it is visible in the ER, the vacuole and in the apoplast. The classical route to the apoplast or the vacuole in plants involves trafficking through the ER and Golgi, however multiple alternative routes are possible for both destinations, including routes that bypass the Golgi (Rose and Lee, 2010; Xiang et al., 2013). As the fluorescently tagged AsTG is enzymatically active when expressed in *N. benthamiana* leaves, this suggests that one of these compartments is physiologically relevant. The early steps of the avenacin biosynthetic pathway occur on the cytosolic face of the ER and in the cytoplasm (Wegel et al., 2009; Reed, 2016; Leveau et al., manuscript in preparation; Louveau et al., manuscript in preparation). AsTG is predicted to be a soluble protein and is likely to be present in the ER lumen (Rapoport, 2007) and therefore it is unlikely that AsTG is functionally active in this location. The avenacin serine carboxypeptidase-like acyltransferase AsSCPL1 (SAD7) is located in the vacuole (Mugford et al., 2013) where avenacin A-1 accumulates (Mylona et al., 2008) suggesting that the vacuole is a viable and likely location for AsTG. Partial secretion to the apoplast has been reported for vacuolar proteins due to the saturation of the vacuolar targeting machinery (daSilva et al., 2005; Frigerio et al., 1998; Pereira et al., 2013).

The N-terminal signal peptide is not sufficient for AsTG localisation in this system.

Co-infiltration of the RFP construct preceded by the AsTG N-terminal signal peptide (SIG:RFP) with a free GFP construct (Materials and Methods 5.5.4) showed that both two and three days post-infiltration the SIG:RFP construct was still largely cytoplasmic, but was also targeted to the apoplast (Figure S7). The general default location for proteins targeted to the plant endomembrane system is secretion to the apoplast (Rojo and Denecke, 2008; Xiang et al., 2013; Craddock et al., 2008). This suggests that the SIG:mRFP1 is partially targeted into the endomembrane system, yet a positive signal required for vacuolar localisation (Pereira et al., 2013) is missing or obscured.

5.3.5 *AsTG* is synonymous with *Sad3*

Mutants defective in avenacin glucosylation, *sad3* and *sad4*, were identified in the initial screen for sodium azide-induced mutants of *A. strigosa* that are deficient in avenacin production and have reduced root fluorescence (Papadopoulou et al., 1999).

sad3 and *sad4* mutants were found to accumulate avenacin A-1 without the β-1,4-D-glucose (Mylona et al., 2008). These mutants had morphological defects such as stunted root growth, fewer root hairs, membrane trafficking defects, and exhibited stress responses such as callose deposition in their roots. These defects were rescued by double mutations with *sad1*, suggesting the phenotypes were a direct result of the toxic effects of incompletely glucosylated avenacins.

Sad4 is unlinked to the avenacin gene cluster and mutants at this locus have a less severe phenotype. These mutants accumulate a mixture of mono-deglucosyl avenacin A-1 and avenacin A-1 and show fewer root growth defects than *sad3* mutants. *sad4* mutants are additionally defective in the glucosylation of avenacosides, steroidal glycosides in oat leaves.

The *Sad3* locus is likely to have a direct role in avenacin A-1 biosynthesis as it is closely linked to the avenacin cluster, and the *sad3* mutant is fully compromised in β-1,4-D-glucose linkage. Therefore it is possible that the *Sad3* locus corresponds to the *AsTG* gene.

Four mutants, #105, #368, #891, and #1139, have been shown to be independent mutant alleles of *sad3* (Mylona et al., 2008). Examination of an extended collection of uncharacterised reduced fluorescence mutants identified four further candidate *sad3* mutants, #986, #1136, #1429 and #1804 (Rachel Melton and Xue Qiao, unpublished).

To investigate a link between the *Sad3* locus and *AsTG*, the *AsTG* gene was amplified from genomic DNA from these eight mutants. DNA sequencing analysis showed that all of these mutants had single-point mutations in the *AsTG* gene (Table 5.3). Four of these had mutations at intron-exon boundaries that might cause splicing errors; one had a predicted premature termination of translation mutation; and three had predicted amino acid substitutions (Table 5.3 and Figure S8). The amino acid substitutions are all responsible for an increase in side chain length and steric bulk and may therefore affect protein folding and stability or catalytic activity.

The F_2 progeny from a cross of #1139 to the *A. strigosa* wild type parent (S75) were analysed to establish whether the *sad3* phenotype and the *AsTG* gene SNP co-segregated.

Three-day-old seedlings (n=190) were phenotyped for wild type or reduced flu-

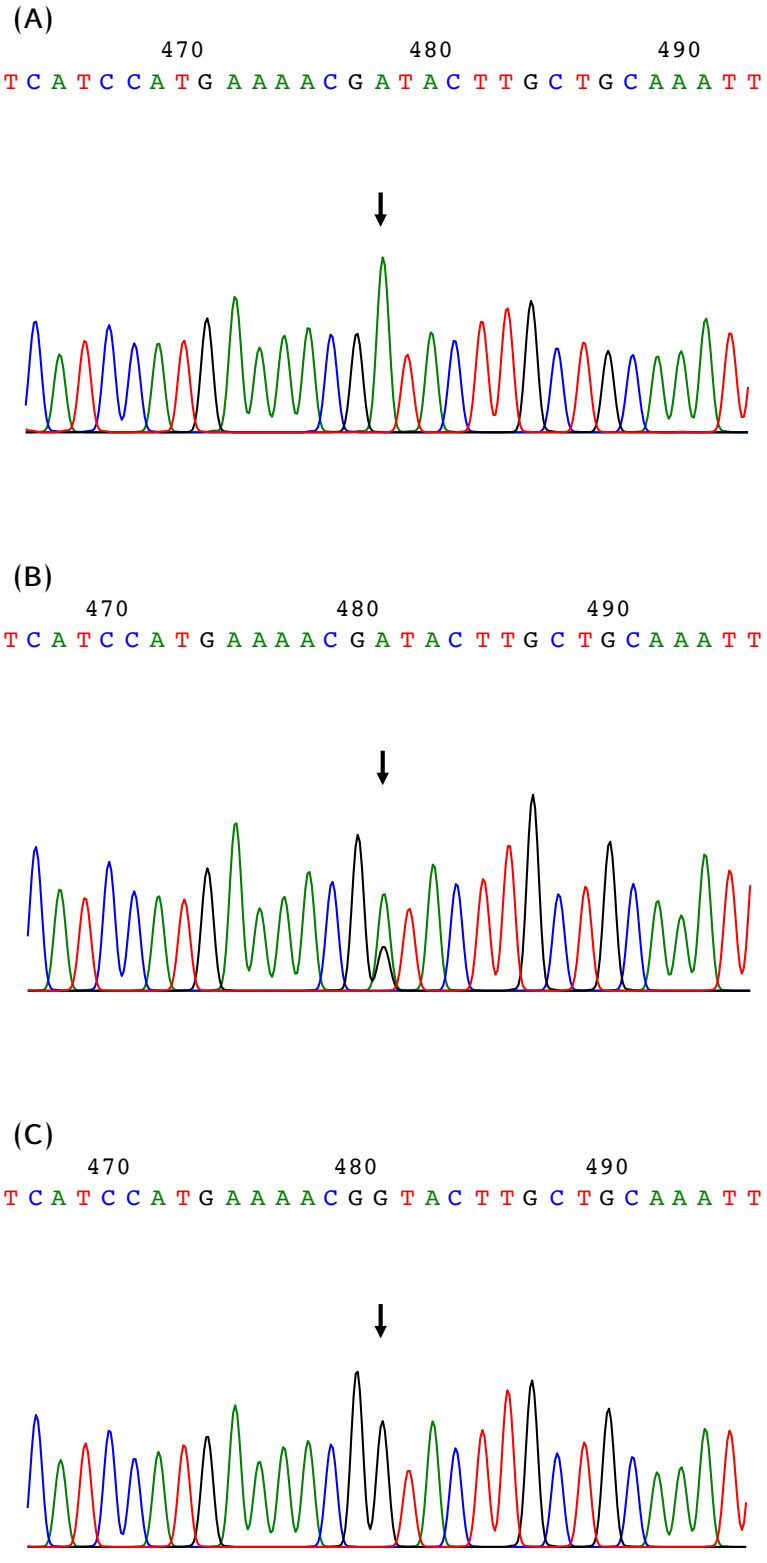


Figure 5.15: Analysis of the *AsTG* gene in #1139 x wild type (WT) F_2 progeny.
Representative electropherograms of the *AsTG* gene in #1139 x wild type (WT) F_2 progeny with **(A)** mutant phenotype; or **(B)** and **(C)**, wild type phenotype. The #1139 mutant has a single nucleotide polymorphism at position +1800 (G1800A) (black arrow).

Table 5.3: Summary of *sad3* SNPs.

Mutant	Mutation Event	Predicted amino acid change
#1139	G1800A	intron/exon
#105	G1705A*	intron/exon
#368	G216A	Val-29 Met
#891	C481T	Ala-88 Val
#986	G1705A*	intron/exon
#1136	G2049A	intron/exon
#1429	G945A	Val-166 Met
#1804	G1729A	Trp-360 STOP

*Identical mutation, but likely to be independent mutation events as the mutants were isolated from different M2 families.

orescence as in Papadopoulou et al (1999). The segregation ratio was statistically consistent with the expected 3:1 ratio for a single recessive Mendelian mutation (χ^2 = 0.007, p > 0.05). These phenotyped mutants were then sequenced to analyse the *AsTG* genotype (Figure 5.15).

The #1139 mutant SNP and the reduced root fluorescence phenotype absolutely co-segregated, indicating that the mutation in AsTG is responsible for the *sad3* mutant phenotype. This includes the inability of *sad3* mutants to produce fully-glucosylated avenacins, with an associated increase in susceptibility to the fungal pathogen *G. graminis* var. *tritici*, and developmental defects such as stunted roots and callose deposition in the roots due to the toxic effects of these avenacin intermediates (Papadopoulou et al., 1999; Mylona et al., 2008).

5.4 Summary

This chapter reports the identification and characterisation of AsTG, the enzyme that adds the final sugar onto the avenacin trisaccharide chain. The *AsTG* gene was identified by genetic linkage to the avenacin cluster genes as it is adjacent to *AsUGT91*, the gene that encodes the avenacin triterpene-3-*O*-arabinose 1,2-glucosyltransferase. RNA-seq analysis shows that *AsTG* shares the same expression profile as the characterised avenacin genes, and it is phylogenetically related to recently characterised GH1 enzymes that catalyse the transfer of glucose moieties to plant specialised metabolites.

Functional characterisation of recombinant AsTG showed that it has transglucosidase activity towards avenacin intermediates both *in vitro* and in an *in planta* expression system, and is the first instance of a GH1 transglucosidase reported to be involved in triterpene glycosylation. *AsTG* was found to be synonymous with *Sad3*, a locus that had been previously defined genetically as being required for avenacin biosynthesis but not yet cloned, confirming its role in the avenacin biosynthetic pathway. This indicates a wider role for glycosyl hydrolase family enzymes in the glycosylation of plant specialised metabolites, and may facilitate the identification of unknown GTs in other plant natural product pathways.

An investigation into the subcellular localisation of AsTG by transient expression of fluorescent fusion proteins in *N. benthamiana* leaves was consistent with a vacuolar localisation of AsTG. This also increases the range of possible glycosylation events that are possible in heterologous systems, as molecules that are transferred to the vacuole and are inaccessible to cytosolic UGTs can be further decorated by vacuolar TGs.

The identification of AsTG completes the knowledge of the avenacin biosynthetic pathway. This will enable the future engineering of this pathway into other agriculturally important crops to confer take-all resistance, and opens up opportunities to access a whole new swathe of carbohydrate-active enzymes for metabolic engineering in heterologous systems.

5.5 Materials and Methods for Chapter 5

5.5.1 Heatmap

Heatmap was visualised in R (R Core Team, 2017) using the package gplots (https://CRAN.R-project.org/package=gplots).

5.5.2 RT-PCR profiling

RT-PCR profiling was carried out as in Materials and Methods Section 2.10 with cDNA kindly provided by Dr Thomas Louveau, with an annealing temperature of 55 °C and 30 PCR cycles. Primers for glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) are listed in Table 3.6. Primers for *AsbAS1* (*Sad1*) were Sad1-1-5 (ATGTGGAGGCTAACAATAGG) and Sad1-2-3 (TATCTCATGACGATGTTCCG) and primers for *AsTG* were F-AsTG-8 (CTCGGGAGTCTACTCGACCA) and R-AsTG-8(GGGTGTTTCCATTTGCGAGC). PCR products were analysed on an 1.2% agarose gel.

5.5.3 Phylogenetic analysis

Protein sequences were aligned using MAFFT (https://mafft.cbrc.jp/alignment/software/) and the evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and are in the units of the number of amino acid substitutions per site. The analysis involved 104 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 214 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016).

GH1 family members of rice (Opassiri et al., 2006) and *Arabidopsis* (Xu et al., 2004) were included in the analysis, as well as the following protein sequences from other plant species: BAM29304 AaAA7GT, *Agapanthus africanus* acyl-glucose-dependent anthocyanin 7-*O*-glucosyltransferase; BAO96250 CmAA7GT, *Campanula medium* acyl-glucose dependent anthocyanin 7-*O*-glucosyltransferase; E3W9M3 DgAA7GT, *Delphinium grandiflorum* acyl-glucose-dependent anthocyanin 7-*O*-glucosytransferase; BAO04178 DgAA7BG-GT1, *Delphinium grandiflorum* acyl-glucose-dependent anthocyanin glucosyltransferase; BAO04181 DgAA7BG-GT2, *Delphinium grandiflorum* acyl-glucose-dependent anthocyanin glucosyltransferase; E3W9M2 DcAA5GT, *Dianthus caryophyllus* acyl-glucose-dependent anthocyanin 5-

O-glucosytransferase; Q8S3J3 Gm HIUHase, *Glycine max* hydroxyisourate hydrolase; Q9ZT64 Pc coniferin β-glucosidase, *Pinus contorta* coniferin β-glucosidase; AAL37714 Sl β-mannosidase, *Solanum lycopersicum* β-mannosidase; AAC49177 Sb dhurrinase, *Sorghum bicolor* dhurrinase; Q38786 As-Glu1, *Avena sativa* 26-desgluco-avenacosidase 1; Q9ZP27 As-Glu2, *Avena sativa* 26-desgluco-avenacosidase 2; P49235 ZmGlu1, *Zea mays* β-glucosidase 1; Q41761 ZmGlu2, *Zea mays* β-glucosidase 2; AAB71381 Me linamarinase, *Manihot esculenta* cassava linamarase; P26205 Tr linamarinase, *Trifolium repens* white clover linamarase; CAA57913 Bn β-glucosidase, *Brassica napus* zeotin-*O*-glucosidase; AAG54074 Bj myrosinase, *Brassica juncea* myrosinase; Q00326 Bn Myrosinase, *Brassica napus* myrosinase; P29092 Sa Myrosinase MB3, *Sinapis alba* myrosinase; BAC78656 Cs β-primeverosidase, *Camellia sinensis* β-primeverosidase.

5.5.4 Gateway cloning

AsTG was amplified from cDNA kindly provided by Dr James Reed, by two-step Gateway cloning (Materials and Methods section 2.11.1) with Fgw-AsTG and Rgw-AsTG primers (Table 5.4), and cloned into the pDONR207 entry vector (Section 2.11.3) to generate the plasmid pDONR207-AsTG.

To generate a plasmid with AsTG missing the predicted N-terminal signal sequence (AsTG-NOSIG), 60 ng of the pDONR207-AsTG plasmid was used as a template with Fgw-nosigAsTG and Rgw-AsTG primers (Table 5.4) by two-step Gateway cloning (section 2.11.1). Before the PCR purification step, 1 µl of DpnI (New England Biolabs) was added to the amplified PCR product and the mixture was incubated at 37°C for 1 hour to digest the pDONR207-AsTG DNA template. The purified PCR product was then cloned into the pDONR207 entry vector (section 2.11.3) to generate pDONR207-NOSIG-AsTG.

For expression in *E. coli*, NOSIG-AsTG was transferred from pDONR207-NOSIG-AsTG into the pH9-GW Gateway destination vector (Section 2.11.4, Table 2.2) to generate pH9-GW-NOSIG-AsTG. pH9-GW-NOSIG-AsTG was transformed into BL21 Rosetta (Invitrogen) (section 2.14).

For transient expression in *N. benthamiana*, pDONR207-AsTG and pDONR207-NOSIG-AsTG were cloned (section 2.11.4) into the pEAQ-*HT*-DEST1 Gateway destination vector (Table 2.2) and transformed into *A. tumefaciens* strain LBA4404 (section 2.15).

For p35S-driven C-terminal RFP fusion constructs, AsTG and AsTG-NOSIG were amplified from pEAQ-*HT*-DEST1-AsTG and pEAQ-*HT*-DEST1-NOSIG-AsTG as DNA

templates respectively, in one-step Gateway PCR reactions (section 2.11.2) with assistance from Freddie Morrison. These sequences were amplified using the forward primers, Fgw-AsTGsig-FULL and Fgw-AsTG-NOSIG-FULL, respectively, with the reverse primer, Rgw-AsTG-NOSTOP-FULL, to remove the stop codons of each sequence (Table 5.4). The predicted N-terminal signal sequence of AsTG was amplified from pEAQ-*HT*-DEST1-AsTG in a one-step Gateway PCR reaction using primers Fgw-AsTGsig-FULL and Rgw-TGSIGNALONLY-FULL (Table 5.4) as section 2.11.2, except that the annealing temperature was 60 °C and the extension time was 15 seconds. The above PCR products were cloned into the pDONR207 entry plasmid (Section 2.11.3)and subsequently transferred into the pB7RWG2 destination vector (Table 2.2, Section 2.11.4).

For p35-driven N- and C-terminal fusion contructs of AsUGT91, AsUGT91 was amplified from the pH9-GW-AsUGT91 plasmid with AsUGT91-NTGW and Rgw-UGT91, or Fgw-GTUGT91 and Rgw-UGT91-NOSTOP primers (Tables 4.3 and 5.4) and cloned into pMDC45 and pMDC83, respectively (Table 2.2, sections 2.11.1, 2.11.3 and 2.11.4.)

To generate the 35S:GFP control, GFP was amplified using the pMDC45 plasmid as a template, with SpeI-GFP-For and GFP+STOP+SacI-Rev primers (Table 5.4) to introduce a SpeI cleavage site at the N-terminus and a STOP codon and a SacI restriction site at the C-terminus (PCR conditions were as section 2.11.2, except with Q5®High-Fidelity DNA polymerase and Q5®High-Fidelity buffer and an elongation time of 45 seconds). The PCR insert was digested with 2 µl of SpeI (Thermo Scientific) and 1 µl of SacI (Thermo Scientific) in 1x Sure/Cut^{TM} Buffer A (Roche) in a total volume of 50 µl at 37 °C for 4 hours, then purified using a PCR Cleanup kit (Qiagen). The Gateway cassette of the pMDC83 vector was removed by cleavage at SpeI and SacI sites. The pMDC83 vector (750 ng) was digested with 1 µl of SpeI (Thermo Scientific) in 1x Sure/Cut^{TM} Buffer A (Roche) in a total volume of 25 µl at 37 °C for 2 hours. Cleavage at the SpeI site was verified by DNA electrophoresis (section 2.4). SacI (Thermo Scientific) (1 µl) was added and the reaction volume made up to 50 µl in 1x Sure/Cut^{TM} Buffer A (Roche) and incubated at 37 °C for a further 2 hours. The digested pMDC83-derived backbone was gel extracted and purified using a QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's instructions. The purified insert (30 ng) and the purified backbone (100 ng) were ligated with T4 DNA ligase (Promega) in 1x T4 DNA ligase buffer at room temperature overnight. The reaction (5 µl) was used to transform *E. coli* strain DH5α (Section 2.14) for plasmid extraction (2.13).

Fluorescent fusion protein constructs were verified using RFP- or GFP-specific se-

quencing primers (Table 5.4) and transformed into *A. tumefaciens* strain GV3101 (section 2.16).

Table 5.4: Gateway primers

Name	Sequence	
Fgw-AsTG	AAAAAGCAGGCTTAATGGCACTGCTGCTCTGC	Two-step Gateway forward primer for AsTG
Rgw-AsTG	AGAAAGCTGGGTATCACGCAGAGTCGTAATATTGT	Two-step Gateway reverse primer for AsTG/AsTG-NOSIG
Fgw-nosigAsTG	AAAAAGCAGGCTTAATGGGAGACGTTGTGGTGGCG	Two-step Gateway forward primer for AsTG-NOSIG
Fgw-AsTGsig-FULL	GGGGACAAGTTTGTACAAAAAAGCAGGCTTA-ATGGCACTGCTGCTCTGC	One-step Gateway forward primer for AsTG/signal sequence of AsTG
Fgw-AsTG-NOSIG-FULL	GGGGACAAGTTTGTACAAAAAAGCAGGCTTA-ATGGGAGACGTTGTGGTGGCG	One-step Gateway forward primer for AsTG-NOSIG
Rgw-AsTG-NOSTOP-FULL	GGGGACCACTTTGTACAAGAAAGCTGGGTAC-GCAGAGTCGTAATATTGTTTC	One-step Gateway reverse primer for AsTG/AsTG-NOSIG with no stop codon
Rgw-TGSIGNALONLY-FULL	GGGGACCACTTTGTACAAGAAAGCTGGGTAC GAGAGGGCGGCAAGCC	One-step Gateway reverse primer for signal sequence of AsTG
AsUGT91-NTGW	AAAAAGCAGGCTTATGGCCGCCTCTGCTTCC	Two-step Gateway forward primer for AsUGT91 compatible with pMDC45
Rgw-UGT91-NOSTOP	AGAAAGCTGGGTAGTCCATGTAAGACGTGAGCTGCTG	Two-step Gateway reverse primer for AsUGT91 with no stop codon
SpeI-GFP-For	TCTAGAACTAGTCCGGGTACCGGTAGAAAAAATGAGTAAAGG	
GFP+STOP+SacI-Rev	AAATTCGAGCTCTTATTTGTATAGTTCATCCATGCC	
midRFP-Rev	GAGCCGTACTGGAACTGAGG	
midGFP-Rev	GTAGTTCCCGTCGTCCTTGA	
midGFP-For	TCAAGGAGGACGGAAACATC	
midRFP-For	CATCCCCGACTACTTGAAGC	

5.5.5 *In vitro* AsTG enzymatic assays

Recombinant AsTG was expressed as in Materials and Methods Section 2.18. Reactions contained 50 mM citrate buffer pH 5.75, 5 mM *p*-nitrophenyl glucose and 150 µM bis- or mono-deglucosylated avenacin A-1 in a total volume of 50 µl, with two biological repeats. Reactions were started with the addition of approximately 1 µg of the recombinant partially purified AsTG (Materials and Methods Section 2.18), incubated at 30°C overnight and were stopped by the addition of 50 µl methanol. Control reactions were set up as above, except that the AsTG protein preparations had been boiled at 95°C for 10 minutes. The reaction mixtures were centrifuged and analysed by HPLC-MS (Materials and Methods Section 2.18).

5.5.6 Expression in *N. benthamiana* leaves

Analysis of AsTG activity *in planta*

Agroinfiltrations were carried out using *A. tumefaciens* LBA4404 strains as in Sections 2.19 and 2.20 with detection by HPLC-CAD-MS.

Subcellular localisation assay

A. tumefaciens strain GV3101 was used for all fluorescent protein fusion assays. pEAQ-*HT*-DEST1-AsUGT91 and pEAQ-*HT*-DEST1-AsTG were transformed into *A. tumefaciens* strain GV3101 by Dr Aymeric Leveau (JIC). To assess enzymatic activity of the fluorescent protein fusion constructs, agroinfiltrations were carried out as in Material and Methods Sections 2.19 and 2.20 with detection by HPLC using a 50x2.1mm 2.6 µ Kinetex XB-C18 column (Phenomenex) with a column oven temperature of 25°C. Detection was by Charged Aerosol detector (CAD, Corona Ultra RS from Dionex). The gradient was run at 0.3 ml/min with 100% water as Buffer A,100% acetonitrile as Buffer B and was as follows: 10% Buffer B from 0-1.5 minutes; 10-50% Buffer B from 1.5-21 minutes; 50-95% Buffer B from 21-21.5 minutes; a linear gradient between 21.5-23.5 minutes; 95 to 10% Buffer B from 23.5-24 minutes, and held at 10% until 25 minutes.

To assess the expression of fusion constructs in *N. benthamiana* leaves, agroinfiltrations were carried out as in Material and Methods Section 2.19 and imaged by Dr Ingo Appelhagen (JIC) with a Leica TCS SP8X confocal microscope.

5.5.7 Screening *sad3* mutants for mutations in the AsTG gene

Purified gDNA (50 ng) (provided by Aymeric Leveau, Xue Qiao and Rachel Melton) was amplified in duplicate for each mutant, with primers F-AsTG-3 and R-AsTG-tot-1 (Table 5.5) as per the conditions for one-step Gateway cloning (2.11) except with an extension time of three minutes and with 40 amplification cycles. PCR products were excised from a 1.2% agarose gel using a QIAquick gel extraction kit (Qiagen) as per manufacturer's instructions. Purified PCR products were sequenced with primers R-AsTG-3, F-AsTG-4, F-AsTG-5, R-AsTG-6, F-AsTG-7, R-AsTG-7, F-AsTG-8, R-AsTG-9, F-AsTG-10 and R-AsTG-10 (Table 5.5).

Table 5.5: Primers for *sad3* mutant analysis

Name	Sequence
R-AsTG-tot-1	GCGCGGTCTCAAACTTGTTT
F-AsTG-3	TGTCTTCCAGGCTAGTGGGA
R-AsTG-3	TGCTGCAACATCTCCGGTAG
F-AsTG-4	TGTTTCTCTTCAGCCTCCGG
F-AsTG-5	GATTATAAGCAAGCCGCCGC
R-AsTG-6	GCTTGAGATTGAAGGCGTGC
F-AsTG-7	GGACTACCCTCCGGTGATGA
R-AsTG-7	CAGCCCGTCCTGAATGAAGT
F-AsTG-8	CTCGGGAGTCTACTCGACCA
R-AsTG-9	TCTTGCCGACAAAGAGCCAT
F-AsTG-10	ACTCCGCCAGATGGTACTCT
R-AsTG-10	GTTGTTGGACCACCTAGCGA

5.5.8 F_2 population analysis

F_2 seeds from a #1139 x wild type S75 cross were grown on sterile filter paper as Section 2.7. Three-day-old seedlings were phenotyped for reduced fluorescence and transfered to soil in 96-well trays and grown under glasshouse conditions. Approximately 100 mg of leaf material was harvested from two-week-old seedlings and sent for genomic DNA extraction by Richard Goram. 250-500 ng of DNA was amplified with F-AsTG-7 and R-AsTG-7 primers (Table 5.5) as per the conditions for one-step Gateway cloning (2.11) except with iProof HF-buffer, an annealing temperature of 66°C, an elongation step of 40 seconds and with 40 PCR cycles. Crude PCR products were sent for sequencing by PlateSeq Kit PCR (Eurofins) with the R-AsTG-7 primer (Table 5.5).

Chapter 6

General discussion

Plant specialised metabolites, such as triterpene glycosides, determine important crop traits such as pest and pathogen resistance, or flavour. They are also of great potential value in medicine, industry and agriculture (Osbourn et al., 2011; Moses et al., 2014). Their use is restricted, however, by the difficulty of purifying these compounds from natural sources. The identification of the genes and enzymes for these pathways will allow the metabolic engineering of the metabolite biosynthetic pathways in heterologous hosts, and the engineering of novel traits into crops.

A challenge in delineating metabolic pathway genes in plants is that the enzymes involved, such as P450s and UGTs, are members of vast superfamilies (Nelson and Werck-Reichhart, 2011; Caputi et al., 2012; Hamberger and Bak, 2013). Candidate pathway genes can be prioritised based on the relatedness of their predicted products to characterised enzymes with similar activities; co-expression patterns; and/or physical linkage of genes to metabolic gene clusters in plant genomes (Thimmappa et al., 2014).

The work presented in this thesis reports the discovery of two enzymes required for the addition of the β-1,2-D- and β-1,4-D-glucose molecules of the avenacin trisaccharide chain. The UDP-dependent glycosyltransferase, AsUGT91, was identified by homology to characterised UGT enzymes that extend plant metabolite sugar chains, and by the similarity of the *AsUGT91* gene expression pattern to those of characterised avenacin pathway genes. The enzyme that adds the β-1,4-linked D-glucose, AsTG, was identified by virtue of the fact that the *AsTG* gene is tightly physically linked to *AsUGT91*, and shares the same expression profile as other genes in the avenacin cluster. Unexpectedly, this glycosyltransferase is a member of glycosyl hydrolase family 1 (GH1) and is the first transglucosidase (TG) enzyme identified to be involved in

triterpene biosynthesis.

GH1 enzymes show broad-spectrum β-D-glycosidase activity and have a wide range of biological and enzymatic functions across the domains of life (Ketudat Cairns and Esen, 2010; Ketudat Cairns et al., 2012). Archaea, bacteria and fungi typically only have one or a few GH1 genes. However, like other enzyme superfamilies involved in the biosynthesis of specialised metabolites, GH1 enzymes have greatly expanded in plants (Ketudat Cairns and Esen, 2010; Ketudat Cairns et al., 2012; Lombard et al., 2014). There are 30-50 GH1 enzymes in rice, *A. thaliana*, *Brachypodium distachyon* and sorghum (Xu et al., 2004; Opassiri et al., 2006; Tyler et al., 2010).

Plant GH1 enzymes include β-D-glucosidases that can hydrolyse phytohormone glucosides (Luang et al., 2013; Brzobohaty et al., 1993; Lee et al., 2006) and monolignol glucosides (Dharmawardhana et al., 1999). Others hydrolyse plant defence compounds such as the steroidal saponin avenacosides in oat leaves (Gus-Mayer et al., 1994), and the cyanogenic glycosides dhurrin in sorghum and linamarin in white clover (Hosel et al., 1987; Barrett et al., 1995), thereby activating these compounds by transforming them into antimicrobial agents. Other activities include β-D-mannosidases (Xu et al., 2004), dissaccharidases (Mizutani et al., 2002; Ahn et al., 2004), thioglucosidases, such as the myrosinases that cleave *S*-glucosides (Bednarek et al., 2009), and even non-glycosidic bonds in the case of the soybean purine hydrolase hydroxyisourate hydrolase (HIUH) (Raychaudhuri and Tipton, 2002).

The first two GH1 TGs to be identified were anthocyanin glycosyltransferases that are involved in the flower pigmentation of carnations and delphiniums (Matsuba et al., 2010). After a fruitless search to find possible UDP-sugar dependent UGT candidates, the authors reasoned that the carnation glycosyltransferase might not use UDP-sugars as donors. An acyl-glucose donor (1-*O*-acyl β-D-glucose ester) was identified by screening protein-free carnation petal extracts for activity with crude carnation petal protein extracts. The GH1 family enzyme, DcAA5GT, was identified by subsequent fractionation and screening of the carnation petal protein extracts. The second TG, the delphinium DgAA7GT, was identified by homology to the carnation *DcAA5GT* cDNA (Matsuba et al., 2010). Recombinant expression of DcAA5GT and DgAA7GT both *in vitro* and *in planta* showed that they catalyse the 5- and 7-*O* glucosylation of anthocyanidin 3-*O*-glucosides, respectively, with little hydrolysis activity.

Two DgAA7GT homologues, AaAA7GT and CmAA7GT, were found in the monocot *Agapanthus africanus* and a core eudicot, *Campanula medium*, and also showed acyl-glucose dependent anthocyanin 7-*O*-glucosyltransferase activity in *in vitro* assays (Miyahara et al., 2012, 2014). A further two TGs were identified from delphinium,

both of which are able to glucosylate the 7-*O* anthocyanin side chain to form viodelphin (Nishizaki et al., 2013). The *A. thaliana* enzymes, AtBGLU6 and AtBGLU10, have also been implicated as anthocyanin acyl-glucose dependent transglycosidases by co-expression analyses and genetic experiments (Miyahara et al., 2011; Ishihara et al., 2016; Miyahara et al., 2013).

A rice GH1 TG, Os9Bglu31, was identified when a glycosyl hydrolase enzymatic assay unexpectedly showed that recombinant Os9Bglu31 did not release glucose as a product but instead transferred it to acetate, which was being used to buffer the assay (Luang et al., 2013). This TG has a broader specificity than the anthocyanin TGs, being able to transfer glucose between phenolic acids, phytohormones and flavonoids. While it is proposed to use fatty acid glucose esters as glucosyl donor substrates in rice and is induced in rice seedlings in response to drought, stress and phytohormone treatment, the biological role of Os9Bglu31 has yet to be determined (Komvongsa et al., 2015b).

Thus very little is known about the activities of TGs outside of anthocyanin biosynthesis. The discovery of AsTG from oat indicates that these enzymes likely play a much wider role in plant specialised metabolism. The characterisation of GH1 enzymes (particularly in groups At/Os 6 and At/Os 7) is likely to identify novel TGs of plant metabolite pathways.

AsTG is synonymous with *Sad3*, an *A. strigosa* locus that is necessary for the β-1,4-glucosylation step in avenacin trisaccharide formation (Papadopoulou et al., 1999; Mylona et al., 2008). The *Sad4* locus is also involved in this glucosylation step, with the roots of *sad4* mutants accumulating a mixture of avenacins and intermediates lacking this glucose molecule (Papadopoulou et al., 1999; Mylona et al., 2008). *Sad4* may be involved in the formation of the native sugar donor for AsTG, accounting for the partial inhibition of this glucosylation step in *sad4* mutants. These mutants are also affected in the glucosylation of the steroidal glycoside, avenacoside A, produced in oat leaves (Papadopoulou et al., 1999). This suggests that the *Sad4* locus may play a similar role in providing acyl sugar donors for a GH1 TG enzyme in the avenacoside biosynthetic pathway, additionally implicating TGs in steroidal glycoside biosynthesis.

A BLAST (tBLASTn) search of predicted *A. strigosa* coding DNA sequences (sequences provided by the group of Bin Han, CAS) using the AsTG protein sequence revealed 52 oat GH1 proteins, including a homologue of the *A. thaliana* AtSFR2 enzyme. A phylogenetic analysis of the enzyme sequences shows that AsTG groups with seven other oat enzymes, whose genes have a range of expression profiles as anal-

ysed by RNA-seq (Figure S9). In the future it will be interesting to characterise the physiological roles of these enzymes in oats, and to determine whether they are active glycosyl hydrolases and/or transglucosidases. The factors determining the switch from glycosyl hydrolase activity to transglycosylase activity are currently unknown, but once elucidated could enable the production of mutant enzymes with novel activities (Ketudat Cairns et al., 2012; Bissaro et al., 2015; Komvongsa et al., 2015a). Reports of GH enzymes with transglycosidase activity and little hydrolysis are rare, and the characterisation of novel TGs in comparison with related enzymes with GH activity may contribute to knowledge in this area.

All of the characterised TGs of GH1 transfer glucose molecules onto their acceptors. GH1 enzymes show a range of activities, including β-D-fucosidase, β-D-galactosidase, β-D-mannosidase, and β-D-disaccharidase activities (Ketudat Cairns et al., 2012), and therefore it is possible that TGs of this family may be able to transfer other saccharide molecules onto acceptor substrates. Additionally, other GH families that hydrolyse glycosidic linkages with net retention of the anomeric configuration (retaining glycoside hydrolases) and show specificity towards plant specialised metabolites (such as GH3) may also be found to be involved in the biosynthesis of these compounds.

All of the characterised GH1 TGs have been proposed to be vacuolar (Matsuba et al., 2010; Luang et al., 2013; Nishizaki et al., 2013). Parallels have been made with acyltransferases, another class of enzyme with two evolutionarily distinct types, in which one type is localised in the cytoplasm whilst the other is active in the vacuole (Sasaki and Nakayama, 2015). The cytoplasmic UGT glycosyltransferases and BAHD family acyltransferases use different classes of donor molecules (UDP-sugar and acyl-CoA, respectively), whilst vacuolar GH1 family transglycosidases and SCPL family acyltransferases both use acyl-glucose molecules as sugar-donors. The same acyl-glucose donor *p*-hydroxybenzoyl-glucose has been shown to be used in four successive enzymatic reactions with two vacuolar acyltransferases and two vacuolar TGs in the synthesis of cyanodelphin in delphinium (Nishizaki et al., 2013). The use of a different class of glycosyltransferase may contribute to the diversity of plant specialised metabolites, as vacuolar enzymes allow the elaboration of compounds that have been transported to this subcellular compartment.

Localisation assays of AsTG by monitoring fluorescent (mRFP1) protein fusions in *N. benthamiana* were consistent with a vacuolar location for this enzyme although fluorescent fusions to AsTG were also present in the apoplast. Targetting through the endomembrane system is prone to generating artefacts (Collings, 2013; Moore and

Murphy, 2009) and the partial secretion to the apoplast may have been due to the saturation of the vacuolar targeting machinery (daSilva et al., 2005; Frigerio et al., 1998; Pereira et al., 2013).

Optimisation of the expression level of AsTG:RFP in this system could be carried out to attempt to reduce the possible saturation of the system; or alternative localisation assays (including subcellular fractionation, immunofluorescence staining or fluorescent fusion protein assays in other systems such as onion epidermal cells) could be carried out to complement the *N. benthamiana* assays. Preliminary enzymatic assays *in vitro* in two buffer systems, (citrate buffer pH 3-6) and (acetic acid/MES/Tris buffer pH 4.5-7.5) suggest that the optimal pH of AsTG is at pH 5.5-pH6, which is consistent with a vacuolar localisation.

The flexibility of AsUGT91 and AsTG in the *N. benthamiana* system has yet to be investigated. Both enzymes are active towards triterpene scaffolds that differ in their oxidation and acylation states, and they may therefore be able to accept a range of triterpene (and possibly also other) scaffolds as substrates. Both activities, the transfer of a glucose to the 2-*O* and 4-*O* positions of an triterpene-3-*O*-arabinoside have not been previously described, and these enzymes provide new activities that could be used in combinatorial biosynthesis to create novel triterpene glycosides. In addition, the specificity of these enzymes towards different sugar donors is as yet unexplored. The identification of the sugar donor of AsTG would shed light on this aspect of triterpene glycoside biosynthesis in oats and may provide methods to increase the yield of the AsTG product by increasing the availability of these donors in heterologous expression systems.

The discovery of AsUGT91 and AsTG completes the biosynthetic steps to form avenacin A-1. Knowledge of this pathway will enable the genetic engineering of the avenacin cluster into agriculturally important crops such as barley or wheat to confer resistance to the fungal 'take-all' pathogen.

The eleven avenacin biosynthetic genes are all physically linked in the *A. strigosa* genome, and include a diverse range of enzymes from different enzyme superfamilies. The discovery of specialised metabolism gene clusters has proved highly useful in the identification of genes for entire pathways, and algorithms for mining plant genomes for biosynthetic gene clusters have recently been developed for facilitating pathway discovery (Medema and Osbourn, 2016; Kautsar et al., 2017). The new understanding of the types of enzymes that contribute to the formation of these compounds, such as TGs as presented here, will rapidly accelerate our ability to discover and harness the chemical engineering capabilities of plants.

Bibliography

Achnine, L., Huhman, D. V., Farag, M. A., Sumner, L. W., Blount, J. W., and Dixon, R. A. (2005). Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume *Medicago truncatula*. *Plant J.*, 41:875–87.

Ahn, Y. O., Mizutani, M., Saino, H., and Sakata, K. (2004). Furcatin hydrolase from *Viburnum furcatum* Blume is a novel disaccharide-specific acuminosidase in glycosyl hydrolase family 1. *J. Biol. Chem.*, 279(22):23405–23414.

Armah, C. N., Mackie, A. R., Roy, C., Price, K., Osbourn, A. E., Bowyer, P., and Ladha, S. (1999). The membrane-permeabilizing effect of avenacin A-1 involves the reorganization of bilayer cholesterol. *Biophys. J.*, 76:281–90.

Asada, K., Salim, V., Masada-Atsumi, S., Edmunds, E., Nagatoshi, M., Terasaka, K., Mizukami, H., and De Luca, V. (2013). A 7-deoxyloganetic acid glucosyltransferase contributes a key step in secologanin biosynthesis in Madagascar periwinkle. *Plant Cell*, 25:4123–34.

Augustin, J. M., Drok, S., Shinoda, T., Sanmiya, K., Nielsen, J. K., Khakimov, B., Olsen, C. E., Hansen, E. H., Kuzina, V., Ekstrom, C. T., Hauser, T., and Bak, S. (2012). UDP-glycosyltransferases from the UGT73C subfamily in *Barbarea vulgaris* catalyze sapogenin 3-*O*-glucosylation in saponin-mediated insect resistance. *Plant Physiol.*, 160:1881–95.

Augustin, J. M., Kuzina, V., Andersen, S. B., and Bak, S. (2011). Molecular activities, biosynthesis and evolution of triterpenoid saponins. *Phytochemistry*, 72:435–57.

Babst, B. A., Chen, H.-Y., Wang, H.-Q., Payyavula, R. S., Thomas, T. P., Harding, S. A., and Tsai, C.-J. (2014). Stress-responsive hydroxycinnamate glycosyltrans-

ferase modulates phenylpropanoid metabolism in *Populus*. *J. Exp. Bot.*, 65:4191–200.

Barrett, T., Suresh, C. G., Tolley, S. P., Dodson, E. J., and Hughes, M. A. (1995). The crystal structure of a cyanogenic β-glucosidase from white clover, a family 1 glycosyl hydrolase. *Structure*, 3:951–60.

Bednarek, P., Pislewska-Bednarek, M., Svatos, A., Schneider, B., Doubsky, J., Mansurova, M., Humphry, M., Consonni, C., Panstruga, R., Sanchez-Vallet, A., Molina, A., and Schulze-Lefert, P. (2009). A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum antifungal defense. *Science*, 323:101–6.

Begley, M. J., Crombie, L., Crombie, W. M. L., and Whiting, D. A. (1986). The isolation of avenacins A-1, A-2, B-1, and B-2, chemical defences against cereal 'take-all' disease. Structure of their 'aglycones', the avenestergenins, and their anhydro dimers. *J. Chem. Soc. Perkin Trans. 1*, pages 1905–1915.

Bissaro, B., Monsan, P., Faure, R., and O'Donohue, M. J. (2015). Glycosynthesis in a waterworld: new insight into the molecular basis of transglycosylation in retaining glycoside hydrolases. *Biochem. J.*, 467(1):17–35.

Bowles, D., Isayenkova, J., Lim, E.-K., and Poppenberger, B. (2005). Glycosyltransferases: managers of small molecules. *Curr. Opin. Plant Biol.*, 8(3):254 – 263.

Bowles, D. and Lim, E.-K. (2010). Glycosyltransferases of small molecules: Their roles in plant biology. *eLS*. doi: https://doi.org/10.1002/9780470015902.a0021537.

Bowles, D., Lim, E.-K., Poppenberger, B., and Vaistij, F. E. (2006). Glycosyltransferases of lipophilic small molecules. *Annu. Rev. Plant Biol.*, 57:567–597.

Bowyer, P., Clarke, B., Lunness, P., Daniels, M., and Osbourn, A. (1995). Host range of a plant pathogenic fungus determined by a saponin detoxifying enzyme. *Science*, 267(5196):371–374.

Brazier-Hicks, M., Evans, K. M., Gershater, M. C., Puschmann, H., Steel, P. G., and Edwards, R. (2009). The *C*-glycosylation of flavonoids in cereals. *J. Biol. Chem.*, 284:17926–34.

Brazier-Hicks, M., Gershater, M., Dixon, D., and Edwards, R. (2017). Substrate specificity and safener inducibility of the plant UDP-glucose-

dependent family 1 glycosyltransferase super-family. *Plant Biotechnol. J.* doi: https://doi.org/10.1111/pbi.12775.

Brazier-Hicks, M., Offen, W. A., Gershater, M. C., Revett, T. J., Lim, E.-K., Bowles, D. J., Davies, G. J., and Edwards, R. (2007). Characterization and engineering of the bifunctional *N*- and *O*-glucosyltransferase involved in xenobiotic metabolism in plants. *Proc. Natl. Acad. Sci. U.S.A.*, 104:20238–43.

Brugliera, F., Holton, T. A., Stevenson, T. W., Farcy, E., Lu, C. Y., and Cornish, E. C. (1994). Isolation and characterization of a cDNA clone corresponding to the Rt locus of *Petunia hybrida*. *Plant J.*, 5:81–92.

Bryan, G. T., Labourdette, E., Melton, R. E., Nicholson, P., Daniels, M. J., and Osbourn, A. E. (1999). DNA polymorphism and host range in the take-all fungus, *Gaeumannomyces graminis*. *Mycol. Res.*, 103(3):319 – 327.

Brzobohaty, B., Moore, I., Kristoffersen, P., Bako, L., Campos, N., Schell, J., and Palme, K. (1993). Release of active cytokinin by a β-glucosidase localized to the maize root meristem. *Science*, 262:1051–4.

Burge, C. and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.*, 268:78–94.

Campbell, J. A., Davies, G. J., Bulone, V., and Henrissat, B. (1997). A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochem. J.*, 326 (Pt 3):929–39.

Caputi, L., Lim, E.-K., and Bowles, D. J. (2008). Discovery of new biocatalysts for the glycosylation of terpenoid scaffolds. *Chemistry*, 14:6656–62.

Caputi, L., Malnoy, M., Goremykin, V., Nikiforova, S., and Martens, S. (2012). A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land. *Plant J.*, 69:1030–42.

Carter, J. P., Spink, J., Cannon, P. F., Daniels, M. J., and Osbourn, A. E. (1999). Isolation, characterization, and avenacin sensitivity of a diverse collection of cereal-root-colonizing fungi. *Appl. Environ. Microbiol.*, 65(8):3364–3372.

Chen, H.-Y. and Li, X. (2017). Identification of a residue responsible for UDP-sugar donor selectivity of a dihydroxybenzoic acid glycosyltransferase from *Arabidopsis* natural accessions. *Plant J.*, 89:195–203.

Chu, H. Y. (2013). *Adaptive metabolic gene clusters as toolkits for chemical innovation: Investigation of the origin of the avenacin gene cluster for synthesis of defense compounds in oats*. PhD thesis, University of East Anglia.

Chu, H. Y., Wegel, E., and Osbourn, A. (2011). From hormones to secondary metabolism: the emergence of metabolic gene clusters in plants. *Plant J.*, 66:66–79.

Collings, D. A. (2013). Subcellular localization of transiently expressed fluorescent fusion proteins. In: Rose R. (eds) Legume genomics. Methods in molecular biology (methods and protocols), vol 1069. Humana Press, Totowa, NJ. pages 227–258.

Coutinho, P. M., Deleury, E., Davies, G. J., and Henrissat, B. (2003). An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.*, 328:307–17.

Craddock, C. P., Hunter, P. R., Szakacs, E., Hinz, G., Robinson, D. G., and Frigerio, L. (2008). Lack of a vacuolar sorting receptor leads to non-specific missorting of soluble vacuolar proteins in *Arabidopsis* seeds. *Traffic*, 9:408–16.

Crombie, L., Crombie, W. M. L., and Whiting, D. A. (1984a). Isolation of avenacins A-1, A-2, B-1, and B-2 from oat roots: structures of their 'aglycones', the avenestergenins. *J. Chem. Soc. Chem. Commun.*, 4:244–246.

Crombie, L., Crombie, W. M. L., and Whiting, D. A. (1984b). Structures of the four avenacins, oat root resistance factors to 'take-all' disease. *J. Chem. Soc., Chem. Commun.*, 4:246–248.

Crombie, L., Crombie, W. M. L., and Whiting, D. A. (1986a). Structures of the oat root resistance factors to 'take-all' disease, avenancins A-1, A-2, B-1 and B-2 and their companion substances. *J. Chem. Soc. Perkin Trans. 1*, pages 1917–1922.

Crombie, L. M. W. and Crombie, L. (1986). Distribution of avenacins A-1, A-2, B-1 and B-2 in oat roots: Their fungicidal activity towards 'take-all' fungus. *Phytochemistry*, 25(9):2069 – 2073.

Crombie, W. M. L., Crombie, L., Green, J. B., and Lucas, J. A. (1986b). Pathogenicity of 'take-all' fungus to oats: Its relationship to the concentration and detoxification of the four avenacins. *Phytochemistry*, 25(9):2075 – 2083.

Curtis, M. D. and Grossniklaus, U. (2003). A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol.*, 133(2):462–469.

Dai, L., Liu, C., Zhu, Y., Zhang, J., Men, Y., Zeng, Y., and Sun, Y. (2015). Functional characterization of cucurbitadienol synthase and triterpene glycosyltransferase involved in biosynthesis of mogrosides from *Siraitia grosvenorii*. *Plant Cell Physiol.*, 56:1172–82.

Dai, X., Zhuang, J., Wu, Y., Wang, P., Zhao, G., Liu, Y., Jiang, X., Gao, L., and Xia, T. (2017). Identification of a flavonoid glucosyltransferase involved in 7-OH site glycosylation in tea plants (*Camellia sinensis*). *Sci. Rep.*, 7:5926.

daSilva, L. L. P., Taylor, J. P., Hadlington, J. L., Hanton, S. L., Snowden, C. J., Fox, S. J., Foresti, O., Brandizzi, F., and Denecke, J. (2005). Receptor salvage from the prevacuolar compartment is essential for efficient vacuolar protein targeting. *Plant Cell*, 17:132–48.

Dharmawardhana, D. P., Ellis, B. E., and Carlson, J. E. (1999). cDNA cloning and heterologous expression of coniferin β-glucosidase. *Plant Mol. Biol.*, 40:365–72.

Di, S., Yan, F., Rodas, F. R., Rodriguez, T. O., Murai, Y., Iwashina, T., Sugawara, S., Mori, T., Nakabayashi, R., Yonekura-Sakakibara, K., Saito, K., and Takahashi, R. (2015). Linkage mapping, molecular cloning and functional analysis of soybean gene *Fg3* encoding flavonol 3-*O*-glucoside/galactoside (1 –> 2) glucosyltransferase. *BMC Plant Biol.*, 15:126.

Eudes, A., Bozzo, G. G., Waller, J. C., Naponelli, V., Lim, E.-K., Bowles, D. J., Gregory, J. F. r., and Hanson, A. D. (2008). Metabolism of the folate precursor *p*-aminobenzoate in plants: glucose ester formation and vacuolar storage. *J. Biol. Chem.*, 283:15451–9.

Falcone Ferreyra, M. L., Rodriguez, E., Casas, M. I., Labadie, G., Grotewold, E., and Casati, P. (2013). Identification of a bifunctional maize *C*- and *O*-glucosyltransferase. *J. Biol. Chem.*, 288:31678–88.

Field, B., Fiston-Lavier, A.-S., Kemen, A., Geisler, K., Quesneville, H., and Osbourn, A. E. (2011). Formation of plant metabolic gene clusters within dynamic chromosomal regions. *Proc. Natl. Acad. Sci. U.S.A.*, 108:16116–21.

Field, B. and Osbourn, A. (2012). Order in the playground: Formation of plant gene clusters in dynamic chromosomal regions. *Mob. Genet. Elements.*, 2:46–50.

Field, B. and Osbourn, A. E. (2008). Metabolic diversification–independent assembly of operon-like gene clusters in different plants. *Science*, 320:543–7.

Ford, C. M., Boss, P. K., and Hoj, P. B. (1998). Cloning and characterization of *Vitis vinifera* UDP-glucose:flavonoid 3-*O*-glucosyltransferase, a homologue of the enzyme encoded by the maize *Bronze-1* locus that may primarily serve to glucosylate anthocyanidins in vivo. *J. Biol. Chem.*, 273:9224–33.

Fraissinet-Tachet, L., Baltz, R., Chong, J., Kauffmann, S., Fritig, B., and Saindrenan, P. (1998). Two tobacco genes induced by infection, elicitor and salicylic acid encode glucosyltransferases acting on phenylpropanoids and benzoic acid derivatives, including salicylic acid. *FEBS Lett.*, 437:319–23.

Franks, T. K., Yadollahi, A., Wirthensohn, M. G., Guerin, J. R., Kaiser, B. N., Sedgley, M., and Ford, C. M. (2008). A seed coat cyanohydrin glucosyltransferase is associated with bitterness in almond (*Prunus dulcis*) kernels. *Funct. Plant Biol.*, 35(3):236–246.

Frigerio, L., de Virgilio, M., Prada, A., Faoro, F., and Vitale, A. (1998). Sorting of phaseolin to the vacuole is saturable and requires a short C-terminal peptide. *Plant Cell*, 10(6):1031–1042.

Frydman, A., Liberman, R., Huhman, D. V., Carmeli-Weissberg, M., Sapir-Mir, M., Ophir, R., W Sumner, L., and Eyal, Y. (2013). The molecular and enzymatic basis of bitter/non-bitter flavor of citrus fruit: evolution of branch-forming rhamnosyltransferases under domestication. *Plant J.*, 73:166–78.

Fukuchi-Mizutani, M., Okuhara, H., Fukui, Y., Nakao, M., Katsumoto, Y., Yonekura-Sakakibara, K., Kusumi, T., Hase, T., and Tanaka, Y. (2003). Biochemical and molecular characterization of a novel UDP-glucose:anthocyanin 3'-*O*-glucosyltransferase, a key enzyme for blue anthocyanin biosynthesis, from gentian. *Plant Physiol.*, 132:1652–63.

Gandia-Herrero, F., Lorenz, A., Larson, T., Graham, I. A., Bowles, D. J., Rylott, E. L., and Bruce, N. C. (2008). Detoxification of the explosive 2,4,6-trinitrotoluene in *Arabidopsis*: discovery of bifunctional O- and C-glucosyltransferases. *Plant J.*, 56:963–74.

Geisler, K., Hughes, R. K., Sainsbury, F., Lomonossoff, G. P., Rejzek, M., Fairhurst, S., Olsen, C.-E., Motawia, M. S., Melton, R. E., Hemmings, A. M., Bak, S., and Osbourn, A. (2013). Biochemical analysis of a multifunctional cytochrome P450 (CYP51) enzyme required for synthesis of antimicrobial triterpenes in plants. *Proc. Natl. Acad. Sci. U.S.A.*, 110(35):E3360–7.

Goodwin, R. H. and Pollock, B. M. (1954). Studies on roots. I. Properties and distribution of fluorescent constituents in *Avena* roots. *Am. J. Bot.*, 41(6):516–520.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., and Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.*, 38:W695–W699.

Griesser, M., Vitzthum, F., Fink, B., Bellido, M. L., Raasch, C., Munoz-Blanco, J., and Schwab, W. (2008). Multi-substrate flavonol *O*-glucosyltransferases from strawberry (*Fragaria x ananassa*) achene and receptacle. *J. Exp. Bot.*, 59:2611–25.

Güçlü-Üstündağ, Ö. and Mazza, G. (2007). Saponins: properties, applications and processing. *Crit. Rev. Food Sci. Nutr.*, 47:231–58.

Gus-Mayer, S., Brunner, H., Schneider-Poetsch, H. A., and Rüdiger, W. (1994). Avenacosidase from oat: purification, sequence analysis and biochemical characterization of a new member of the BGA family of β-glucosidases. *Plant Mol. Biol.*, 26:909–21.

Ha, S., Walker, D., Shi, Y., and Walker, S. (2000). The 1.9 Å crystal structure of *Escherichia coli* MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis. *Protein Sci.*, 9:1045–52.

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

Hansen, K. S., Kristensen, C., Tattersall, D. B., Jones, P. R., Olsen, C. E., Bak, S., and Moller, B. L. (2003). The *in vitro* substrate regiospecificity of recombinant UGT85B1, the cyanohydrin glucosyltransferase from *Sorghum bicolor*. *Phytochemistry*, 64:143–51.

Haralampidis, K., Bryan, G., Qi, X., Papadopoulou, K., Bakht, S., Melton, R., and Osbourn, A. (2001). A new class of oxidosqualene cyclases directs synthesis of antimicrobial phytoprotectants in monocots. *Proc. Natl. Acad. Sci. U.S.A.*, 98(23):13431–13436.

Harker, M., Holmberg, N., Clayton, J. C., Gibbard, C. L., Wallace, A. D., Rawlins, S., Hellyer, S. A., Lanot, A., and Safford, R. (2003). Enhancement of seed phytosterol levels by expression of an N-terminal truncated *Hevea brasiliensis* (rubber tree) 3-hydroxy-3-methylglutaryl-CoA reductase. *Plant Biotechnol. J.*, 1:113–21.

Hartley, J. L., Temple, G. F., and Brasch, M. A. (2000). DNA cloning using *in vitro* site-specific recombination. *Genome Res.*, 10(11):1788–1795.

He, F., Chen, W.-K., Yu, K.-J., Ji, X.-N., Duan, C.-Q., Reeves, M. J., and Wang, J. (2015). Molecular and biochemical characterization of the UDP-glucose: Anthocyanin 5-*O*-glucosyltransferase from *Vitis amurensis*. *Phytochemistry*, 117:363–72.

He, X.-Z., Wang, X., and Dixon, R. A. (2006). Mutational analysis of the *Medicago* glycosyltransferase UGT71G1 reveals residues that control regioselectivity for (iso)flavonoid glycosylation. *J. Biol. Chem.*, 281:34441–7.

Hefner, T., Arend, J., Warzecha, H., Siems, K., and Stockigt, J. (2002). Arbutin synthase, a novel member of the NRD1βglycosyltransferase family, is a unique multifunctional enzyme converting various natural products and xenobiotics. *Bioorg. Med. Chem.*, 10:1731–41.

Heng, L., Vincken, J.-P., van Koningsveld, G., Legger, A., Gruppen, H., van Boekel, T., Roozen, J., and Voragen, F. (2006). Bitterness of saponins and their content in dry peas. *J. Sci. Food Agric.*, 86(8):1225–1231.

Hirade, Y., Kotoku, N., Terasaka, K., Saijo-Hamano, Y., Fukumoto, A., and Mizukami, H. (2015). Identification and functional analysis of 2-hydroxyflavanone *C*-glucosyltransferase in soybean (*Glycine max*). *FEBS Lett.*, 589:1778–86.

Hirotani, M., Kuroda, R., Suzuki, H., and Yoshikawa, T. (2000). Cloning and expression of UDP-glucose: flavonoid 7-*O*-glucosyltransferase from hairy root cultures of *Scutellaria baicalensis*. *Planta*, 210:1006–13.

Hoffmeister, D., Ichinose, K., and Bechthold, A. (2001). Two sequence elements of glycosyltransferases involved in urdamycin biosynthesis are responsible for substrate specificity and enzymatic activity. *Chem. Biol.*, 8:557–67.

Home-Grown Cereals Association (2006). Take-all in winter wheat – management guidelines.

Hosel, W., Tober, I., Eklund, S. H., and Conn, E. E. (1987). Characterization of β-glucosidases with high specificity for the cyanogenic glucoside dhurrin in *Sorghum bicolor* (l.) Moench seedlings. *Arch. Biochem. Biophys.*, 252:152–62.

Hou, B., Lim, E.-K., Higgins, G. S., and Bowles, D. J. (2004). *N*-glucosylation of cytokinins by glycosyltransferases of *Arabidopsis thaliana*. *J. Biol. Chem.*, 279:47822–32.

Hsu, Y.-H., Tagami, T., Matsunaga, K., Okuyama, M., Suzuki, T., Noda, N., Suzuki, M., and Shimura, H. (2017). Functional characterization of UDP-rhamnose-dependent rhamnosyltransferase involved in anthocyanin modification, a key enzyme determining blue coloration in *Lobelia erinus*. *Plant J.*, 89:325–337.

Hu, Y. and Walker, S. (2002). Remarkable structural similarities between diverse glycosyltransferases. *Chem. Biol.*, 9:1287–96.

Hughes, J. and Hughes, M. A. (1994). Multiple secondary plant product UDP-glucose glucosyltransferase genes expressed in cassava (*Manihot esculenta* crantz) cotyledons. *DNA Seq.*, 5:41–9.

Ikeda, T., Udayama, M., Okawa, M., Arao, T., Kinjo, J., and Nohara, T. (1998). Partial hydrolysis of soyasaponin I and the hepatoprotective effects of the hydrolytic products. Study of the structure-hepatoprotective relationship of soyasapogenol B analogs. *Chem. Pharm. Bull. (Tokyo)*, 46(2):359–361.

Ikegami, A., Akagi, T., Potter, D., Yamada, M., Sato, A., Yonemori, K., Kitajima, A., and Inoue, K. (2009). Molecular identification of 1-Cys peroxiredoxin and anthocyanidin/flavonol 3-*O*-galactosyltransferase from proanthocyanidin-rich young fruits of persimmon (*Diospyros kaki* Thunb.). *Planta*, 230:841–55.

Isayenkova, J., Wray, V., Nimtz, M., Strack, D., and Vogt, T. (2006). Cloning and functional characterisation of two regioselective flavonoid glucosyltransferases from *Beta vulgaris*. *Phytochemistry*, 67:1598–612.

Ishihara, H., Tohge, T., Viehover, P., Fernie, A. R., Weisshaar, B., and Stracke, R. (2016). Natural variation in flavonol accumulation in *Arabidopsis* is determined by the flavonol glucosyltransferase BGLU6. *J. Exp. Bot.*, 67(5):1505–1517.

Itkin, M., Davidovich-Rikanati, R., Cohen, S., Portnoy, V., Doron-Faigenboim, A., Oren, E., Freilich, S., Tzuri, G., Baranes, N., Shen, S., Petreikov, M., Sertchook, R.,

Ben-Dor, S., Gottlieb, H., Hernandez, A., Nelson, D. R., Paris, H. S., Tadmor, Y., Burger, Y., Lewinsohn, E., Katzir, N., and Schaffer, A. (2016). The biosynthetic pathway of the nonsugar, high-intensity sweetener mogroside V from *Siraitia grosvenorii*. *Proc. Natl. Acad. Sci. U.S.A.*, 113:E7619–E7628.

Itkin, M., Heinig, U., Tzfadia, O., Bhide, A. J., Shinde, B., Cardenas, P. D., Bocobza, S. E., Unger, T., Malitsky, S., Finkers, R., Tikunov, Y., Bovy, A., Chikate, Y., Singh, P., Rogachev, I., Beekwilder, J., Giri, A. P., and Aharoni, A. (2013). Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. *Science*, 341:175–9.

Ito, T., Fujimoto, S., Suito, F., Shimosaka, M., and Taguchi, G. (2017). *C*-glycosyltransferases catalyzing the formation of di-*C*-glucosyl flavonoids in citrus plants. *Plant J.*, 91:187–198.

Jones, D. M., Pullen, N., Wells, R., Trick, M., Irwin, J., and Morris, R. (2017). Regulatory divergence of flowering time genes in the allopolyploid *Brassica napus*. *bioRxiv*. doi: https://doi.org/10.1101/178137.

Jones, P., Messner, B., Nakajima, J.-I., Schaffner, A. R., and Saito, K. (2003). UGT73C6 and UGT78D1, glycosyltransferases involved in flavonol glycoside biosynthesis in *Arabidopsis thaliana*. *J. Biol. Chem.*, 278:43910–8.

Jones, P. and Vogt, T. (2001). Glycosyltransferases in secondary plant metabolism: tranquilizers and stimulant controllers. *Planta*, 213:164–74.

Jugde, H., Nguy, D., Moller, I., Cooney, J. M., and Atkinson, R. G. (2008). Isolation and characterization of a novel glycosyltransferase that converts phloretin to phlorizin, a potent antioxidant in apple. *FEBS J.*, 275:3804–14.

Jung, S.-C., Kim, W., Park, S. C., Jeong, J., Park, M. K., Lim, S., Lee, Y., Im, W.-T., Lee, J. H., Choi, G., and Kim, S. C. (2014). Two ginseng UDP-glycosyltransferases synthesize ginsenoside Rg3 and Rd. *Plant Cell Physiol.*, 55:2177–88.

Kaminaga, Y., Sahin, F. P., and Mizukami, H. (2004). Molecular cloning and characterization of a glucosyltransferase catalyzing glucosylation of curcumin in cultured *Catharanthus roseus* cells. *FEBS Lett.*, 567:197–202.

Kannangara, R., Motawia, M. S., Hansen, N. K. K., Paquette, S. M., Olsen, C. E., Moller, B. L., and Jorgensen, K. (2011). Characterization and expression profile

of two UDP-glucosyltransferases, UGT85K4 and UGT85K5, catalyzing the last step in cyanogenic glucoside biosynthesis in cassava. *Plant J.*, 68:287–301.

Karimi, M., Inzé, D., and Depicker, A. (2002). Gateway vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.*, 7(5):193–195.

Kautsar, S. A., Suarez Duran, H. G., Blin, K., Osbourn, A., and Medema, M. H. (2017). plantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters. *Nucleic Acids Res.* doi: https://doi.org/10.1093/nar/gkx305.

Kemen, A. C., Honkanen, S., Melton, R. E., Findlay, K. C., Mugford, S. T., Hayashi, K., Haralampidis, K., Rosser, S. J., and Osbourn, A. (2014). Investigation of triterpene synthesis and regulation in oats reveals a role for β-amyrin in determining root epidermal cell patterning. *Proc. Natl. Acad. Sci. U.S.A.*, 111:8679–84.

Ketudat Cairns, J. R. and Esen, A. (2010). β-Glucosidases. *Cell. Mol. Life Sci.*, 67(20):3389–3405.

Ketudat Cairns, J. R., Pengthaisong, S., Luang, S., Sansenya, S., Tankrathok, A., and Svasti, J. (2012). Protein-carbohydrate interactions leading to hydrolysis and transglycosylation in plant glycoside hydrolase family 1 enzymes. *J. Appl. Glycosci.*, 59(2):51–62.

Kim, J. H., Kim, B. G., Park, Y., Ko, J. H., Lim, C. E., Lim, J., Lim, Y., and Ahn, J.-H. (2006). Characterization of flavonoid 7-*O*-glucosyltransferase from *Arabidopsis thaliana*. *Biosci. Biotechnol. Biochem.*, 70:1471–7.

King, A. J., Brown, G. D., Gilday, A. D., Larson, T. R., and Graham, I. A. (2014). Production of bioactive diterpenoids in the Euphorbiaceae depends on evolutionarily conserved gene clusters. *Plant Cell*, 26:3286–98.

Kita, M., Hirata, Y., Moriguchi, T., Endo-Inagaki, T., Matsumoto, R., Hasegawa, S., Suhayda, C. G., and Omura, M. (2000). Molecular cloning and characterization of a novel gene encoding limonoid UDP-glucosyltransferase in *Citrus*. *FEBS Lett.*, 469:173–8.

Ko, J. H., Kim, B. G., Kim, J. H., Kim, H., Lim, C. E., Lim, J., Lee, C., Lim, Y., and Ahn, J.-H. (2008). Four glucosyltransferases from rice: cDNA cloning, expression, and characterization. *J. Plant Physiol.*, 165:435–44.

Kohara, A., Nakajima, C., Hashimoto, K., Ikenaga, T., Tanaka, H., Shoyama, Y., Yoshida, S., and Muranaka, T. (2005). A novel glucosyltransferase involved in steroid saponin biosynthesis in *Solanum aculeatissimum*. *Plant Mol. Biol.*, 57:225–39.

Kohara, A., Nakajima, C., Yoshida, S., and Muranaka, T. (2007). Characterization and engineering of glycosyltransferases responsible for steroid saponin biosynthesis in *Solanaceous* plants. *Phytochemistry*, 68:478–86.

Kohonen, T. (2013). Essentials of the self-organizing map. *Neural Net.*, 37:52–65.

Komvongsa, J., Luang, S., Marques, J. V., Phasai, K., Davin, L. B., Lewis, N. G., and Ketudat Cairns, J. R. (2015a). Active site cleft mutants of Os9BGlu31 transglucosidase modify acceptor substrate specificity and allow production of multiple kaempferol glycosides. *Biochim. Biophys. Acta*, 1850:1405–14.

Komvongsa, J., Mahong, B., Phasai, K., Hua, Y., Jeon, J.-S., and Ketudat Cairns, J. R. (2015b). Identification of fatty acid glucose esters as Os9BGlu31 transglucosidase substrates in rice flag leaves. *J. Agric. Food. Chem.*, 63:9764–9.

Koshland, D. E. (1953). Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev.*, 28(4):416–436.

Kramer, C. M., Prata, R. T. N., Willits, M. G., De Luca, V., Steffens, J. C., and Graser, G. (2003). Cloning and regiospecificity studies of two flavonoid glucosyltransferases from *Allium cepa*. *Phytochemistry*, 64:1069–76.

Kubo, A., Arai, Y., Nagashima, S., and Yoshikawa, T. (2004). Alteration of sugar donor specificities of plant glycosyltransferases by a single point mutation. *Arch. Biochem. Biophys.*, 429(2):198 – 203.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.*, 33:1870–4.

Kwak, Y.-S. and Weller, D. M. (2013). Take-all of wheat and natural disease suppression: A review. *Plant Pathol. J.*, 29:125–35.

Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008). Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.*, 77:521–55.

Lanot, A., Hodge, D., Jackson, R. G., George, G. L., Elias, L., Lim, E.-K., Vaistij, F. E., and Bowles, D. J. (2006). The glucosyltransferase UGT72E2 is responsible for monolignol 4-*O*-glucoside production in *Arabidopsis thaliana*. *Plant J.*, 48:286–95.

Lee, K. H., Piao, H. L., Kim, H.-Y., Choi, S. M., Jiang, F., Hartung, W., Hwang, I., Kwak, J. M., Lee, I.-J., and Hwang, I. (2006). Activation of glucosidase via stress-induced polymerization rapidly increases active pools of abscisic acid. *Cell*, 126:1109–20.

Letunic, I. and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.*, 44:W242–5.

Li, Y., Baldauf, S., Lim, E. K., and Bowles, D. J. (2001). Phylogenetic analysis of the UDP-glycosyltransferase multigene family of *Arabidopsis thaliana*. *J. Biol. Chem.*, 276:4338–43.

Lim, C. E., Choi, J. N., Kim, I. A., Lee, S. A., Hwang, Y.-S., Lee, C. H., and Lim, J. (2008). Improved resistance to oxidative stress by a loss-of-function mutation in the Arabidopsis UGT71C1 gene. *Mol. Cells*, 25:368–75.

Lim, E.-K., Ashford, D. A., Hou, B., Jackson, R. G., and Bowles, D. J. (2004). *Arabidopsis* glycosyltransferases as biocatalysts in fermentation for regioselective synthesis of diverse quercetin glucosides. *Biotechnol. Bioeng.*, 87:623–31.

Lim, E.-K., Baldauf, S., Li, Y., Elias, L., Worrall, D., Spencer, S. P., Jackson, R. G., Taguchi, G., Ross, J., and Bowles, D. J. (2003). Evolution of substrate recognition across a multigene family of glycosyltransferases in *Arabidopsis*. *Glycobiology.*, 13:139–45.

Lim, E.-K., Doucet, C. J., Li, Y., Elias, L., Worrall, D., Spencer, S. P., Ross, J., and Bowles, D. J. (2002). The activity of Arabidopsis glycosyltransferases toward salicylic acid, 4-hydroxybenzoic acid, and other benzoates. *J. Biol. Chem.*, 277:586–92.

Lim, E. K., Li, Y., Parr, A., Jackson, R., Ashford, D. A., and Bowles, D. J. (2001). Identification of glucosyltransferase genes involved in sinapate metabolism and lignin synthesis in *Arabidopsis*. *J. Biol. Chem.*, 276:4344–9.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.*, 42(Database issue):D490–5.

Louveau, T. (2013). *Investigation of glycosyltransferases from oat*. PhD thesis, University of East Anglia.

Louveau, T., Leitao, C., Green, S., Hamiaux, C., van der Rest, B., Dechy-Cabaret, O., Atkinson, R. G., and Chervin, C. (2011). Predicting the substrate specificity of a glycosyltransferase implicated in the production of phenolic volatiles in tomato fruit. *FEBS J.*, 278:390–400.

Luang, S., Cho, J.-I., Mahong, B., Opassiri, R., Akiyama, T., Phasai, K., Komvongsa, J., Sasaki, N., Hua, Y.-l., Matsuba, Y., Ozeki, Y., Jeon, J.-S., and Cairns, J. R. K. (2013). Rice Os9BGlu31 is a transglucosidase with the capacity to equilibrate phenylpropanoid, flavonoid, and phytohormone glycoconjugates. *J. Biol. Chem.*, 288(14):10111–10123.

Lunkenbein, S., Bellido, M., Aharoni, A., Salentijn, E. M. J., Kaldenhoff, R., Coiner, H. A., Munoz-Blanco, J., and Schwab, W. (2006). Cinnamate metabolism in ripening fruit. Characterization of a UDP-glucose:cinnamate glucosyltransferase from strawberry. *Plant Physiol.*, 140:1047–58.

Ma, L.-Q., Liu, B.-Y., Gao, D.-Y., Pang, X.-B., Lu, S.-Y., Yu, H.-S., Wang, H., Yan, F., Li, Z.-Q., Li, Y.-F., and Ye, H.-C. (2007). Molecular cloning and overexpression of a novel UDP-glucosyltransferase elevating salidroside levels in *Rhodiola sachalinensis*. *Plant Cell Rep.*, 26:989–99.

Mackenzie, P. I., Owens, I. S., Burchell, B., Bock, K. W., Bairoch, A., Belanger, A., Fournel-Gigleux, S., Green, M., Hum, D. W., Iyanagi, T., Lancet, D., Louisot, P., Magdalou, J., Chowdhury, J. R., Ritter, J. K., Schachter, H., Tephly, T. R., Tipton, K. F., and Nebert, D. W. (1997). The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. *Pharmacogenetics.*, 7:255–69.

Mageroy, M. H., Jancsik, S., Man Saint Yuen, M., Fischer, M., Withers, S. G., Paetz, C., Schneider, B., Mackay, J., and Bohlmann, J. (2017). A conifer UDP-sugar dependent glycosyltransferase contributes to acetophenone metabolism and defense against insects. *Plant Physiol.*, 175:641–651.

Martin, R. C., Mok, M. C., Habben, J. E., and Mok, D. W. (2001). A maize cytokinin gene encoding an *O*-glucosyltransferase specific to cis-zeatin. *Proc. Natl. Acad. Sci. U.S.A.*, 98:5922–6.

Martin, R. C., Mok, M. C., and Mok, D. W. (1999). A gene encoding the cytokinin enzyme zeatin *O*-xylosyltransferase of *Phaseolus vulgaris*. *Plant Physiol.*, 120:553–8.

Masada, S., Terasaka, K., Oguchi, Y., Okazaki, S., Mizushima, T., and Mizukami, H. (2009). Functional and structural characterization of a flavonoid glucoside 1,6-glucosyltransferase from *Catharanthus roseus*. *Plant Cell Physiol.*, 50:1401–15.

Mato, M., Ozeki, Y., Itoh, Y., Higeta, D., Yoshitama, K., Teramoto, S., Aida, R., Ishikura, N., and Shibata, M. (1998). Isolation and characterization of a cDNA clone of UDP-galactose: flavonoid 3-*O*-galactosyltransferase (UF3GaT) expressed in *Vigna mungo* seedlings. *Plant Cell Physiol.*, 39:1145–55.

Matsuba, Y., Sasaki, N., Tera, M., Okamura, M., Abe, Y., Okamoto, E., Nakamura, H., Funabashi, H., Takatsu, M., Saito, M., Matsuoka, H., Nagasawa, K., and Ozeki, Y. (2010). A novel glucosylation reaction on anthocyanins catalyzed by acyl-glucose-dependent glucosyltransferase in the petals of carnation and delphinium. *Plant Cell*, 22(10):3374–3389.

Matsuda, F., Hirai, M. Y., Sasaki, E., Akiyama, K., Yonekura-Sakakibara, K., Provart, N. J., Sakurai, T., Shimada, Y., and Saito, K. (2010). AtMetExpress development: a phytochemical atlas of *Arabidopsis* development. *Plant Physiol.*, 152:566–78.

Mayer, K. F. X., Martis, M., Hedley, P. E., Simkova, H., Liu, H., Morris, J. A., Steuernagel, B., Taudien, S., Roessner, S., Gundlach, H., Kubalakova, M., Suchankova, P., Murat, F., Felder, M., Nussbaumer, T., Graner, A., Salse, J., Endo, T., Sakai, H., Tanaka, T., Itoh, T., Sato, K., Platzer, M., Matsumoto, T., Scholz, U., Dolezel, J., Waugh, R., and Stein, N. (2011). Unlocking the barley genome by chromosomal and comparative genomics. *Plant Cell*, 23:1249–63.

Mayer, K. F. X., Taudien, S., Martis, M., Simkova, H., Suchankova, P., Gundlach, H., Wicker, T., Petzold, A., Felder, M., Steuernagel, B., Scholz, U., Graner, A., Platzer, M., Dolezel, J., and Stein, N. (2009). Gene content and virtual gene order of barley chromosome 1H. *Plant Physiol.*, 151:496–505.

McCue, K. F., Allen, P. V., Shepherd, L. V. T., Blake, A., Maccree, M. M., Rockhold, D. R., Novy, R. G., Stewart, D., Davies, H. V., and Belknap, W. R. (2007). Potato glycosterol rhamnosyltransferase, the terminal step in triose side-chain biosynthesis. *Phytochemistry*, 68:327–34.

Medema, M. H. and Osbourn, A. (2016). Computational genomic identification and functional reconstitution of plant natural product biosynthetic pathways. *Nat. Prod. Rep.*, 33:951–62.

Meesapyodsuk, D., Balsevich, J., Reed, D. W., and Covello, P. S. (2007). Saponin biosynthesis in *Saponaria vaccaria*. cDNAs encoding β-amyrin synthase and a triterpene carboxylic acid glucosyltransferase. *Plant Physiol.*, 143:959–69.

Michlmayr, H., Malachova, A., Varga, E., Kleinova, J., Lemmens, M., Newmister, S., Rayment, I., Berthiller, F., and Adam, G. (2015). Biochemical characterization of a recombinant UDP-glucosyltransferase from rice and enzymatic production of deoxynivalenol-3-*O*-β-D-glucoside. *Toxins*, 7:2685–700.

Milkowski, C., Baumert, A., and Strack, D. (2000). Identification of four *Arabidopsis* genes encoding hydroxycinnamate glucosyltransferases. *FEBS Lett.*, 486:183–4.

Mittasch, J., Mikolajewski, S., Breuer, F., Strack, D., and Milkowski, C. (2010). Genomic microstructure and differential expression of the genes encoding UDP-glucose:sinapate glucosyltransferase (UGT84A9) in oilseed rape (*Brassica napus*). *Theor. Appl. Genet.*, 120:1485–500.

Miyahara, T., Matsuba, Y., Ozeki, Y., and Sasaki, N. (2011). Identification of genes in *Arabidopsis thaliana* with homology to a novel acyl-glucose dependent glucosyltransferase of carnations. *Plant Biotechnology*, 28(3):311–315.

Miyahara, T., Sakiyama, R., Ozeki, Y., and Sasaki, N. (2013). Acyl-glucose-dependent glucosyltransferase catalyzes the final step of anthocyanin formation in *Arabidopsis*. *J. Plant Physiol.*, 170(6):619–624.

Miyahara, T., Takahashi, M., Ozeki, Y., and Sasaki, N. (2012). Isolation of an acyl-glucose-dependent anthocyanin 7-*O*-glucosyltransferase from the monocot *Agapanthus africanus*. *J. Plant Physiol.*, 169(13):1321–1326.

Miyahara, T., Tani, T., Takahashi, M., Nishizaki, Y., Ozeki, Y., and Sasaki, N. (2014). Isolation of anthocyanin 7-*O*-glucosyltransferase from canterbury bells (*Campanula medium*). *Plant Biotechnology*, 31(5):555–559.

Mizutani, M., Nakanishi, H., Ema, J.-i., Ma, S.-J., Noguchi, E., Inohara-Ochiai, M., Fukuchi-Mizutani, M., Nakao, M., and Sakata, K. (2002). Cloning of β-primeverosidase from tea leaves, a key enzyme in tea aroma formation. *Plant Physiol.*, 130(4):2164–2176.

Modolo, L. V., Blount, J. W., Achnine, L., Naoumkina, M. A., Wang, X., and Dixon, R. A. (2007). A functional genomics approach to (iso)flavonoid glycosylation in the model legume *Medicago truncatula*. *Plant Mol. Biol.*, 64:499–518.

Moellering, E. R., Muthan, B., and Benning, C. (2010). Freezing tolerance in plants requires lipid remodeling at the outer chloroplast membrane. *Science*, 330(6001):226–228.

Moglia, A., Lanteri, S., Comino, C., Hill, L., Knevitt, D., Cagliero, C., Rubiolo, P., Bornemann, S., and Martin, C. (2014). Dual catalytic activity of hydroxycinnamoyl-coenzyme A quinate transferase from tomato allows it to moonlight in the synthesis of both mono- and dicaffeoylquinic acids. *Plant Physiol.*, 166(4):1777–1787.

Montefiori, M., Espley, R. V., Stevenson, D., Cooney, J., Datson, P. M., Saiz, A., Atkinson, R. G., Hellens, R. P., and Allan, A. C. (2011). Identification and characterisation of F3GT1 and F3GGT1, two glycosyltransferases responsible for anthocyanin biosynthesis in red-fleshed kiwifruit (*Actinidia chinensis*). *Plant J.*, 65:106–118.

Moore, I. and Murphy, A. (2009). Validating the location of fluorescent protein fusions in the endomembrane system. *Plant Cell*, 21:1632–6.

Moraga, A. R., Mozos, A. T., Ahrazem, O., and Gomez-Gomez, L. (2009). Cloning and characterization of a glucosyltransferase from *Crocus sativus* stigmas involved in flavonoid glucosylation. *BMC Plant Biol.*, 9:109.

Moraga, A. R., Nohales, P. F., Perez, J. A. F., and Gomez-Gomez, L. (2004). Glucosylation of the saffron apocarotenoid crocetin by a glucosyltransferase isolated from *Crocus sativus* stigmas. *Planta*, 219:955–66.

Morita, Y., Hoshino, A., Kikuchi, Y., Okuhara, H., Ono, E., Tanaka, Y., Fukui, Y., Saito, N., Nitasaka, E., Noguchi, H., and Iida, S. (2005). Japanese morning glory dusky mutants displaying reddish-brown or purplish-gray flowers are deficient in a novel glycosylation enzyme for anthocyanin biosynthesis, UDP-glucose:anthocyanidin 3-*O*-glucoside-2"-*O*-glucosyltransferase, due to 4-bp insertions in the gene. *Plant J.*, 42:353–63.

Moses, T., Papadopoulou, K. K., and Osbourn, A. (2014). Metabolic and functional diversity of saponins, biosynthetic intermediates and semi-synthetic derivatives. *Crit. Rev. Biochem. Mol. Biol.*, 49:439–62.

Mugford, S. T., Louveau, T., Melton, R., Qi, X., Bakht, S., Hill, L., Tsurushima, T., Honkanen, S., Rosser, S. J., Lomonossoff, G. P., and Osbourn, A. (2013). Modularity of plant metabolic gene clusters: a trio of linked genes that are collectively required for acylation of triterpenes in oat. *Plant Cell*, 25(3):1078–1092.

Mugford, S. T., Qi, X., Bakht, S., Hill, L., Wegel, E., Hughes, R. K., Papadopoulou, K., Melton, R., Philo, M., Sainsbury, F., Lomonossoff, G. P., Roy, A. D., Goss, R. J., and Osbourn, A. (2009). A serine carboxypeptidase-like acyltransferase is required for synthesis of antimicrobial compounds and disease resistance in oats. *Plant Cell*, 21(8):2473–2484.

Mylona, P., Owatworakit, A., Papadopoulou, K., Jenner, H., Qin, B., Findlay, K., Hill, L., Qi, X., Bakht, S., Melton, R., and Osbourn, A. (2008). *Sad3* and *Sad4* are required for saponin biosynthesis and root development in oat. *Plant Cell*, 20(1):201–212.

Nagatomo, Y., Usui, S., Ito, T., Kato, A., Shimosaka, M., and Taguchi, G. (2014). Purification, molecular cloning and functional characterization of flavonoid *C*-glucosyltransferases from *Fagopyrum esculentum* M. (buckwheat) cotyledon. *Plant J.*, 80:437–48.

Nagatoshi, M., Terasaka, K., Nagatsu, A., and Mizukami, H. (2011). Iridoid-specific glucosyltransferase from *Gardenia jasminoides*. *J. Biol. Chem.*, 286:32866–74.

Nagatoshi, M., Terasaka, K., Owaki, M., Sota, M., Inukai, T., Nagatsu, A., and Mizukami, H. (2012). UGT75L6 and UGT94E5 mediate sequential glucosylation of crocetin to crocin in *Gardenia jasminoides*. *FEBS Lett.*, 586:1055–61.

Nakatsuka, T. and Nishihara, M. (2010). UDP-glucose:3-deoxyanthocyanidin 5-*O*-glucosyltransferase from *Sinningia cardinalis*. *Planta*, 232:383–92.

Naoumkina, M. A., Modolo, L. V., Huhman, D. V., Urbanczyk-Wochniak, E., Tang, Y., Sumner, L. W., and Dixon, R. A. (2010). Genomic and coexpression analyses predict multiple genes involved in triterpene saponin biosynthesis in *Medicago truncatula*. *Plant Cell*, 22:850–66.

Nelson, B. K., Cai, X., and Nebenführ, A. (2007). A multicolored set of in vivo organelle markers for co-localization studies in *Arabidopsis* and other plants. *Plant J.*, 51(6):1126–1136.

Nelson, D. and Werck-Reichhart, D. (2011). A *P450*-centric view of plant evolution. *Plant J.*, 66:194–211.

Nishizaki, Y., Sasaki, N., Yasunaga, M., Miyahara, T., Okamoto, E., Okamoto, M., Hirose, Y., and Ozeki, Y. (2014). Identification of the glucosyltransferase gene that supplies the *p*-hydroxybenzoyl-glucose for 7-polyacylation of anthocyanin in delphinium. *J. Exp. Bot.*, 65:2495–506.

Nishizaki, Y., Yasunaga, M., Okamoto, E., Okamoto, M., Hirose, Y., Yamaguchi, M., Ozeki, Y., and Sasaki, N. (2013). *p*-hydroxybenzoyl-glucose is a zwitter donor for the biosynthesis of 7-polyacylated anthocyanin in *Delphinium*. *Plant Cell*, 25(10):4150–4165.

Noda, N., Yoshioka, S., Kishimoto, S., Nakayama, M., Douzono, M., Tanaka, Y., and Aida, R. (2017). Generation of blue chrysanthemums by anthocyanin B-ring hydroxylation and glucosylation and its coloration mechanism. *Sci. Adv.*, 3:e1602785.

Noguchi, A., Fukui, Y., Iuchi-Okada, A., Kakutani, S., Satake, H., Iwashita, T., Nakao, M., Umezawa, T., and Ono, E. (2008). Sequential glucosylation of a furofuran lignan, (+)-sesaminol, by *Sesamum indicum* UGT71A9 and UGT94D1 glucosyltransferases. *Plant J.*, 54:415–27.

Noguchi, A., Horikawa, M., Fukui, Y., Fukuchi-Mizutani, M., Iuchi-Okada, A., Ishiguro, M., Kiso, Y., Nakayama, T., and Ono, E. (2009). Local differentiation of sugar donor specificity of flavonoid glycosyltransferase in *Lamiales*. *Plant Cell*, 21:1556–72.

Noguchi, A., Saito, A., Homma, Y., Nakao, M., Sasaki, N., Nishino, T., Takahashi, S., and Nakayama, T. (2007). A UDP-glucose:isoflavone 7-*O*-glucosyltransferase from the roots of soybean (*Glycine max*) seedlings. purification, gene cloning, phylogenetics, and an implication for an alternative strategy of enzyme catalysis. *J. Biol. Chem.*, 282:23581–90.

Nomura, T., Ishihara, A., Imaishi, H., Endo, T. R., Ohkawa, H., and Iwamura, H. (2002). Molecular characterization and chromosomal localization of cytochrome P450 genes involved in the biosynthesis of cyclic hydroxamic acids in hexaploid wheat. *Mol. Genet. Genomics*, 267:210–7.

Nomura, T., Ishihara, A., Imaishi, H., Ohkawa, H., Endo, T. R., and Iwamura, H. (2003). Rearrangement of the genes for the biosynthesis of benzoxazinones in the evolution of *Triticeae* species. *Planta*, 217:776–82.

Nutzmann, H.-W. and Osbourn, A. (2014). Gene clustering in plant specialized metabolism. *Curr. Opin. Biotechnol.*, 26:91–9.

Offen, W., Martinez-Fleites, C., Yang, M., Kiat-Lim, E., Davis, B. G., Tarling, C. A., Ford, C. M., Bowles, D. J., and Davies, G. J. (2006). Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. *EMBO J.*, 25(6):1396–1405.

Ogata, J., Kanno, Y., Itoh, Y., Tsugawa, H., and Suzuki, M. (2005). Plant biochemistry: anthocyanin biosynthesis in roses. *Nature*, 435:757–8.

Ono, E., Fukuchi-Mizutani, M., Nakamura, N., Fukui, Y., Yonekura-Sakakibara, K., Yamaguchi, M., Nakayama, T., Tanaka, T., Kusumi, T., and Tanaka, Y. (2006). Yellow flowers generated by expression of the aurone biosynthetic pathway. *Proc. Natl. Acad. Sci. U.S.A.*, 103:11075–80.

Ono, E., Ruike, M., Iwashita, T., Nomoto, K., and Fukui, Y. (2010). Co-pigmentation and flavonoid glycosyltransferases in blue *Veronica persica* flowers. *Phytochemistry*, 71:726–35.

Ono, N. N., Qin, X., Wilson, A. E., Li, G., and Tian, L. (2016). Two UGT84 family glycosyltransferases catalyze a critical reaction of hydrolyzable tannin biosynthesis in pomegranate (*Punica granatum*). *PLoS One*, 11:e0156319.

Opassiri, R., Pomthong, B., Onkoksoong, T., Akiyama, T., Esen, A., and Ketudat Cairns, J. R. (2006). Analysis of rice glycosyl hydrolase family 1 and expression of Os4bglu12 beta-glucosidase. *BMC Plant Biol.*, 6:33.

Osbourn, A., Clarke, B., Dow, J., and Daniels, M. (1991). Partial characterization of avenacinase from *Gaeumannomyces graminis* var. *avenae*. *Physiol. Mol. Plant Pathol.*, 38(4):301 – 312.

Osbourn, A., Clarke, B., Lunness, P., Scott, P., and Daniels, M. (1994). An oat species lacking avenacin is susceptible to infection by *Gaeumannomyces graminis* var. *tritici*. *Physiol. Mol. Plant Pathol.*, 45(6):457 – 467.

Osbourn, A., Goss, R. J. M., and Field, R. A. (2011). The saponins: polar isoprenoids with important and diverse biological activities. *Nat. Prod. Rep.*, 28:1261–8.

Osmani, S. A., Bak, S., Imberty, A., Olsen, C. E., and Moller, B. L. (2008). Catalytic key amino acids and UDP-sugar donor specificity of a plant glucuronosyltransferase,

UGT94B1: molecular modeling substantiated by site-specific mutagenesis and biochemical analyses. *Plant Physiol.*, 148:1295–308.

Osmani, S. A., Bak, S., and Moller, B. L. (2009). Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling. *Phytochemistry*, 70(3):325–347.

Ostrowski, M., Hetmann, A., and Jakubowska, A. (2015). Indole-3-acetic acid UDP-glucosyltransferase from immature seeds of pea is involved in modification of glycoproteins. *Phytochemistry*, 117:25–33.

Owatworakit, A., Townsend, B., Louveau, T., Jenner, H., Rejzek, M., Hughes, R. K., Saalbach, G., Qi, X., Bakht, S., Roy, A. D., Mugford, S. T., Goss, R. J. M., Field, R. A., and Osbourn, A. (2013). Glycosyltransferases from oat (*Avena*) implicated in the acylation of avenacins. *J. Biol. Chem.*, 288(6):3696–3704.

Owens, D. K. and McIntosh, C. A. (2009). Identification, recombinant expression, and biochemical characterization of a flavonol 3-*O*-glucosyltransferase clone from *Citrus paradisi*. *Phytochemistry*, 70:1382–91.

Pang, Y., Peel, G. J., Sharma, S. B., Tang, Y., and Dixon, R. A. (2008). A transcript profiling approach reveals an epicatechin-specific glucosyltransferase expressed in the seed coat of *Medicago truncatula*. *Proc. Natl. Acad. Sci. U.S.A.*, 105:14210–5.

Papadopoulou, K., Melton, R. E., Leggett, M., Daniels, M. J., and Osbourn, A. E. (1999). Compromised disease resistance in saponin-deficient plants. *Proc. Natl. Acad. Sci. U.S.A.*, 96(22):12923–12928.

Paquette, S., Moller, B. L., and Bak, S. (2003). On the origin of family 1 plant glycosyltransferases. *Phytochemistry*, 62:399–413.

Payne, R. M. E., Xu, D., Foureau, E., Teto Carqueijeiro, M. I. S., Oudin, A., Bernonville, T. D. d., Novak, V., Burow, M., Olsen, C.-E., Jones, D. M., Tatsis, E. C., Pendle, A., Ann Halkier, B., Geu-Flores, F., Courdavault, V., Nour-Eldin, H. H., and O'Connor, S. E. (2017). An NPF transporter exports a central monoterpene indole alkaloid intermediate from the vacuole. *Nat. Plants*, 3:16208.

Pereira, C., Pereira, S., Satiat-Jeunemaitre, B., and Pissarra, J. (2013). Cardosin A contains two vacuolar sorting signals using different vacuolar routes in tobacco epidermal cells. *Plant J.*, 76:87–100.

Priest, D. M., Ambrose, S. J., Vaistij, F. E., Elias, L., Higgins, G. S., Ross, A. R. S., Abrams, S. R., and Bowles, D. J. (2006). Use of the glucosyltransferase UGT71B6 to disturb abscisic acid homeostasis in *Arabidopsis thaliana*. *Plant J.*, 46:492–502.

Qi, X., Bakht, S., Leggett, M., Maxwell, C., Melton, R., and Osbourn, A. (2004). A gene cluster for secondary metabolism in oat: Implications for the evolution of metabolic diversity in plants. *Proc. Natl. Acad. Sci. U.S.A.*, 101(21):8233–8238.

Qi, X., Bakht, S., Qin, B., Leggett, M., Hemmings, A., Mellon, F., Eagles, J., Werck-Reichhart, D., Schaller, H., Lesot, A., Melton, R., and Osbourn, A. (2006). A different function for a member of an ancient and highly conserved cytochrome P450 family: From essential sterols to plant defense. *Proc. Natl. Acad. Sci. U.S.A.*, 103(49):18848–18853.

Qin, B., Eagles, J., Mellon, F. A., Mylona, P., Peña-Rodriguez, L., and Osbourn, A. E. (2010). High throughput screening of mutants of oat that are defective in triterpene synthesis. *Phytochemistry*, 71(11):1245 – 1252.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rapoport, T. A. (2007). Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature*, 450:663–9.

Rask, L., Andreasson, E., Ekbom, B., Eriksson, S., Pontoppidan, B., and Meijer, J. (2000). Myrosinase: gene family evolution and herbivore defense in *Brassicaceae*. *Plant Mol. Biol.*, 42:93–113.

Raychaudhuri, A. and Tipton, P. A. (2002). Cloning and expression of the gene for soybean hydroxyisourate hydrolase. Localization and implications for function and mechanism. *Plant Physiol.*, 130:2061–8.

Reed, J., Stephenson, M. J., Miettinen, K., Brouwer, B., Leveau, A., Brett, P., Goss, R. J. M., Goossens, A., O'Connell, M. A., and Osbourn, A. (2017). A translational synthetic biology platform for rapid access to gram-scale quantities of novel drug-like molecules. *Metab. Eng.*, 42:185–193.

Reed, J. A. (2016). *Transient expression for engineering triterpenoid diversity in plants*. PhD thesis, University of East Anglia.

Richman, A., Swanson, A., Humphrey, T., Chapman, R., McGarvey, B., Pocs, R., and Brandle, J. (2005). Functional genomics uncovers three glucosyltransferases involved in the synthesis of the major sweet glucosides of *Stevia rebaudiana*. *Plant J.*, 41:56–67.

Rojas Rodas, F., Di, S., Murai, Y., Iwashina, T., Sugawara, S., Mori, T., Nakabayashi, R., Yonekura-Sakakibara, K., Saito, K., and Takahashi, R. (2016). Cloning and characterization of soybean gene Fg1 encoding flavonol 3-*O*-glucoside/galactoside (1–>6) glucosyltransferase. *Plant Mol. Biol.*, 92:445–456.

Rojas Rodas, F., Rodriguez, T. O., Murai, Y., Iwashina, T., Sugawara, S., Suzuki, M., Nakabayashi, R., Yonekura-Sakakibara, K., Saito, K., Kitajima, J., Toda, K., and Takahashi, R. (2014). Linkage mapping, molecular cloning and functional analysis of soybean gene Fg2 encoding flavonol 3-*O*-glucoside (1 –> 6) rhamnosyltransferase. *Plant Mol. Biol.*, 84:287–300.

Rojo, E. and Denecke, J. (2008). What is moving in the secretory pathway of plants? *Plant Physiol.*, 147:1493–503.

Rose, J. K. and Lee, S.-J. (2010). Straying off the highway: Trafficking of secreted plant proteins and complexity in the plant cell wall proteome. *Plant Physiol.*, 153(2):433–436.

Ross, J., Li, Y., Lim, E., and Bowles, D. J. (2001). Higher plant glycosyltransferases. *Genome Biol.*, 2:REVIEWS3004.

Rozen, S. and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.*, 132:365–386.

Sainsbury, F., Saxena, P., Geisler, K., Osbourn, A., and Lomonossoff, G. P. (2012). Using a virus-derived system to manipulate plant natural product biosynthetic pathways. *Methods Enzymol.*, 517:185 – 202. Natural Product Biosynthesis by Microorganisms and Plants, Part C.

Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, 4:406–25.

Sasaki, K., Takase, H., Kobayashi, H., Matsuo, H., and Takata, R. (2015a). Molecular cloning and characterization of UDP-glucose: furaneol glucosyltransferase gene from grapevine cultivar Muscat Bailey A (*Vitis labrusca x V. vinifera*). *J. Exp. Bot.*, 66:6167–74.

Sasaki, N. and Nakayama, T. (2015). Achievements and perspectives in biochemistry concerning anthocyanin modification for blue flower coloration. *Plant Cell Physiol.*, 56(1):28–40.

Sasaki, N., Nishizaki, Y., Yamada, E., Tatsuzawa, F., Nakatsuka, T., Takahashi, H., and Nishihara, M. (2015b). Identification of the glucosyltransferase that mediates direct flavone *C*-glucosylation in *Gentiana triflora*. *FEBS Lett.*, 589:182–7.

Sasaki, N., Wada, K., Koda, T., Kasahara, K., Adachi, T., and Ozeki, Y. (2005). Isolation and characterization of cDNAs encoding an enzyme with glucosyltransferase activity for cyclo-DOPA from four o'clocks and feather cockscombs. *Plant Cell Physiol.*, 46:666–70.

Sawada, S., Suzuki, H., Ichimaida, F., Yamaguchi, M.-A., Iwashita, T., Fukui, Y., Hemmi, H., Nishino, T., and Nakayama, T. (2005). UDP-glucuronic acid:anthocyanin glucuronosyltransferase from red daisy (*Bellis perennis*) flowers. Enzymology and phylogenetics of a novel glucuronosyltransferase involved in flower pigment biosynthesis. *J. Biol. Chem.*, 280:899–906.

Sayama, T., Ono, E., Takagi, K., Takada, Y., Horikawa, M., Nakamoto, Y., Hirose, A., Sasama, H., Ohashi, M., Hasegawa, H., Terakawa, T., Kikuchi, A., Kato, S., Tatsuzaki, N., Tsukamoto, C., and Ishimoto, M. (2012). The Sg-1 glycosyltransferase locus regulates structural diversity of triterpenoid saponins of soybean. *Plant Cell*, 24:2123–38.

Schaller, H., Grausem, B., Benveniste, P., Chye, M. L., Tan, C. T., Song, Y. H., and Chua, N. H. (1995). Expression of the *Hevea brasiliensis* (H.B.K.) Mull. Arg. 3-hydroxy-3-methylglutaryl-coenzyme A reductase 1 in tobacco results in sterol overproduction. *Plant Physiol.*, 109:761–770.

Schweiger, W., Boddu, J., Shin, S., Poppenberger, B., Berthiller, F., Lemmens, M., Muehlbauer, G. J., and Adam, G. (2010). Validation of a candidate deoxynivalenol-inactivating UDP-glucosyltransferase from barley by heterologous expression in yeast. *Mol. Plant. Microbe Interact.*, 23:977–86.

Shao, H., He, X., Achnine, L., Blount, J. W., Dixon, R. A., and Wang, X. (2005). Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from *Medicago truncatula*. *Plant Cell*, 17(11):3141–3154.

Shibuya, M., Nishimura, K., Yasuyama, N., and Ebizuka, Y. (2010). Identification and characterization of glycosyltransferases involved in the biosynthesis of soyasaponin i in *Glycine max*. *FEBS Lett.*, 584:2258–64.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, 7(1).

Sinnott, M. L. (1990). Catalytic mechanism of enzymic glycosyl transfer. *Chem. Rev.*, 90(7):1171–1202.

Song, C., Hong, X., Zhao, S., Liu, J., Schulenburg, K., Huang, F.-C., Franz-Oberdorf, K., and Schwab, W. (2016). Glucosylation of 4-hydroxy-2,5-dimethyl-3(2H)-furanone, the key strawberry flavor compound in strawberry fruit. *Plant Physiol.*, 171:139–51.

Sparg, S. G., Light, M. E., and van Staden, J. (2004). Biological activities and distribution of plant saponins. *J. Ethnopharmacol.*, 94:219–43.

Sparkes, I. A., Runions, J., Kearns, A., and Hawes, C. (2006). Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nat. Protoc.*, 1:2019–25.

Staehelin, L. A. (1997). The plant ER: a dynamic organelle composed of a large number of discrete functional domains. *Plant J.*, 11:1151–65.

Sun, H.-X., Xie, Y., and Ye, Y.-P. (2009). Advances in saponin-based adjuvants. *Vaccine*, 27:1787–96.

Sun, W., Liang, L., Meng, X., Li, Y., Gao, F., Liu, X., Wang, S., Gao, X., and Wang, L. (2016). Biochemical and molecular characterization of a flavonoid 3-*O*-glycosyltransferase responsible for anthocyanins and flavonols biosynthesis in *Freesia hybrida*. *Front. Plant Sci.*, 7:410.

Suzuki, H., Hayase, H., Nakayama, A., Yamaguchi, I., Asami, T., and Nakajima, M. (2007). Identification and characterization of an *Ipomoea nil* glucosyltransferase which metabolizes some phytohormones. *Biochem. Biophys. Res. Commun.*, 361:980–6.

Szerszen, J. B., Szczyglowski, K., and Bandurski, R. S. (1994). *iaglu*, a gene from *Zea mays* involved in conjugation of growth hormone indole-3-acetic acid. *Science*, 265:1699–701.

Taguchi, G., Yazawa, T., Hayashida, N., and Okazaki, M. (2001). Molecular cloning and heterologous expression of novel glucosyltransferases from tobacco cultured cells that have broad substrate specificity and are induced by salicylic acid and auxin. *Eur. J. Biochem.*, 268:4086–94.

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

Thorlby, G., Fourrier, N., and Warren, G. (2004). The SENSITIVE TO FREEZING2 gene, required for freezing tolerance in *Arabidopsis thaliana*, encodes a β-glucosidase. *Plant Cell*, 16(8):2192–2203.

Tian, L., Blount, J. W., and Dixon, R. A. (2006). Phenylpropanoid glycosyltransferases from osage orange (*Maclura pomifera*) fruit. *FEBS Lett.*, 580:6915–20.

Tiwari, P., Sangwan, R. S., and Sangwan, N. S. (2016). Plant secondary metabolism linked glycosyltransferases: An update on expanding knowledge and scopes. *Biotechnol. Adv.*, 34:714–739.

Tognetti, V. B., Van Aken, O., Morreel, K., Vandenbroucke, K., van de Cotte, B., De Clercq, I., Chiwocha, S., Fenske, R., Prinsen, E., Boerjan, W., Genty, B., Stubbs, K. A., Inze, D., and Van Breusegem, F. (2010). Perturbation of indole-3-butyric acid homeostasis by the UDP-glucosyltransferase UGT74E2 modulates *Arabidopsis* architecture and water stress tolerance. *Plant Cell*, 22:2660–79.

Trapero, A., Ahrazem, O., Rubio-Moraga, A., Jimeno, M. L., Gomez, M. D., and Gomez-Gomez, L. (2012). Characterization of a glucosyltransferase enzyme involved in the formation of kaempferol and quercetin sophorosides in *Crocus sativus*. *Plant Physiol.*, 159:1335–54.

Turner, E. M. (1953). The nature of the resistance of oats to the take-all fungus. *J. Exp. Bot.*, 4(11):264–271.

Turner, E. M. C. (1961). An enzymic basis for pathogenic specificity in *Ophiobolus graminis*. *J. Exp. Bot.*, 12(34):169–175.

Tyler, L., Bragg, J. N., Wu, J., Yang, X., Tuskan, G. A., and Vogel, J. P. (2010). Annotation and comparative analysis of the glycoside hydrolase genes in *Brachypodium distachyon*. *BMC Genomics*, 11:600.

Veach, Y. K., Martin, R. C., Mok, D. W. S., Malbeck, J., Vankova, R., and Mok, M. C. (2003). *O*-glucosylation of *cis*-zeatin in maize. Characterization of genes, enzymes, and endogenous cytokinins. *Plant Physiol.*, 131:1374–80.

Vincken, J.-P., Heng, L., de Groot, A., and Gruppen, H. (2007). Saponins, classification and occurrence in the plant kingdom. *Phytochemistry*, 68:275–97.

Vogt, T., Grimm, R., and Strack, D. (1999). Cloning and expression of a cDNA encoding betanidin 5-*O*-glucosyltransferase, a betanidin- and flavonoid-specific enzyme with high homology to inducible glucosyltransferases from the *Solanaceae*. *Plant J.*, 19:509–19.

Vogt, T. and Jones, P. (2000). Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. *Trends Plant Sci.*, 5(9):380–386.

von Rad, U., Huttl, R., Lottspeich, F., Gierl, A., and Frey, M. (2001). Two glucosyltransferases are involved in detoxification of benzoxazinoids in maize. *Plant J.*, 28:633–42.

Vranova, E., Coman, D., and Gruissem, W. (2013). Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annu. Rev. Plant Biol.*, 64:665–700.

Wang, B., Jin, S.-H., Hu, H.-Q., Sun, Y.-G., Wang, Y.-W., Han, P., and Hou, B.-K. (2012). UGT87A2, an *Arabidopsis* glycosyltransferase, regulates flowering time via FLOWERING LOCUS C. *New Phytol.*, 194:666–75.

Wang, R. and Brattain, M. G. (2007). The maximal size of protein to diffuse through the nuclear pore is larger than 60kDa. *FEBS Lett.*, 581:3164–70.

Wang, X. (2009). Structure, mechanism and engineering of plant natural product glycosyltransferases. *FEBS Lett.*, 583(20):3303–3309.

Wang, X., Li, C., Zhou, C., Li, J., and Zhang, Y. (2017). Molecular characterization of the *C*-glucosylation for puerarin biosynthesis in *Pueraria lobata*. *Plant J.*, 90:535–546.

Wegel, E., Koumproglou, R., Shaw, P., and Osbourn, A. (2009). Cell type-specific chromatin decondensation of a metabolic gene cluster in oats. *Plant Cell*, 21(12):3926–3936.

Wehrens, R. and Buydens, L. M. C. (2007). Self- and Super-organizing Maps in R: The kohonen package. *J. Stat. Softw.*, 021(i05).

Wei, W., Wang, P., Wei, Y., Liu, Q., Yang, C., Zhao, G., Yue, J., Yan, X., and Zhou, Z. (2015). Characterization of *Panax ginseng* UDP-glycosyltransferases catalyzing protopanaxatriol and biosyntheses of bioactive ginsenosides F1 and Rh1 in metabolically engineered yeasts. *Mol. Plant*, 8:1412–24.

Witte, S., Moco, S., Vervoort, J., Matern, U., and Martens, S. (2009). Recombinant expression and functional characterisation of regiospecific flavonoid glucosyltransferases from *Hieracium pilosella* L. *Planta*, 229:1135–46.

Wu, B., Gao, L., Gao, J., Xu, Y., Liu, H., Cao, X., Zhang, B., and Chen, K. (2017). Genome-wide identification, expression patterns, and functional analysis of UDP glycosyltransferase family in peach (*Prunus persica* L. Batsch). *Front. Plant Sci.*, 8:389.

Xiang, L., Etxeberria, E., and Van den Ende, W. (2013). Vacuolar protein sorting mechanisms in plants. *FEBS J.*, 280:979–93.

Xie, K., Chen, R., Li, J., Wang, R., Chen, D., Dou, X., and Dai, J. (2014). Exploring the catalytic promiscuity of a new glycosyltransferase from *Carthamus tinctorius*. *Org. Lett.*, 16(18):4874–4877.

Xu, G., Cai, W., Gao, W., and Liu, C. (2016a). A novel glucuronosyltransferase has an unprecedented ability to catalyse continuous two-step glucuronosylation of glycyrrhetinic acid to yield glycyrrhizin. *New Phytol.*, 212:123–35.

Xu, Z., Escamilla-Trevino, L., Zeng, L., Lalgondar, M., Bevan, D., Winkel, B., Mohamed, A., Cheng, C.-L., Shih, M.-C., Poulton, J., and Esen, A. (2004). Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 1. *Plant Mol. Biol.*, 55(3):343–367.

Xu, Z.-J., Nakajima, M., Suzuki, Y., and Yamaguchi, I. (2002). Cloning and characterization of the abscisic acid-specific glucosyltransferase gene from adzuki bean seedlings. *Plant Physiol.*, 129:1285–95.

Xu, Z.-S., Ma, J., Wang, F., Ma, H.-Y., Wang, Q.-X., and Xiong, A.-S. (2016b). Identification and characterization of DcUCGalT1, a galactosyltransferase responsible for anthocyanin galactosylation in purple carrot (*Daucus carota* L.) taproots. *Sci. Rep.*, 6:27356.

Yahyaa, M., Davidovich-Rikanati, R., Eyal, Y., Sheachter, A., Marzouk, S., Lewinsohn, E., and Ibdah, M. (2016). Identification and characterization of UDP-glucose:phloretin 4'-*O*-glycosyltransferase from Malus x domestica Borkh. *Phytochemistry*, 130:47–55.

Yamazaki, M., Gong, Z., Fukuchi-Mizutani, M., Fukui, Y., Tanaka, Y., Kusumi, T., and Saito, K. (1999). Molecular cloning and biochemical characterization of a novel anthocyanin 5-*O*-glucosyltransferase by mRNA differential display for plant forms regarding anthocyanin. *J. Biol. Chem.*, 274:7405–11.

Yonekura-Sakakibara, K., Fukushima, A., Nakabayashi, R., Hanada, K., Matsuda, F., Sugawara, S., Inoue, E., Kuromori, T., Ito, T., Shinozaki, K., Wangwattana, B., Yamazaki, M., and Saito, K. (2012). Two glycosyltransferases involved in anthocyanin modification delineated by transcriptome independent component analysis in *Arabidopsis thaliana*. *Plant J.*, 69:154–67.

Yonekura-Sakakibara, K., Nakabayashi, R., Sugawara, S., Tohge, T., Ito, T., Koyanagi, M., Kitajima, M., Takayama, H., and Saito, K. (2014). A flavonoid 3-*O*-glucoside:2"-*O*-glucosyltransferase responsible for terminal modification of pollen-specific flavonols in *Arabidopsis thaliana*. *Plant J.*, 79:769–82.

Yonekura-Sakakibara, K., Tohge, T., Niida, R., and Saito, K. (2007). Identification of a flavonol 7-*O*-rhamnosyltransferase gene determining flavonoid pattern in *Arabidopsis* by transcriptome coexpression analysis and reverse genetics. *J. Biol. Chem.*, 282:14932–41.

Yu, H.-S., Ma, L.-Q., Zhang, J.-X., Shi, G.-L., Hu, Y.-H., and Wang, Y.-N. (2011). Characterization of glycosyltransferases responsible for salidroside biosynthesis in *Rhodiola sachalinensis*. *Phytochemistry*, 72:862–70.

Zhao, C. L., Chen, Z. J., Bai, X. S., Ding, C., Long, T. J., Wei, F. G., and Miao, K. R. (2014). Structure-activity relationships of anthocyanidin glycosylation. *Mol. Diversity*, 18:687–700.

Zhong, Y., Xue, X., Liu, Z., Ma, Y., Zeng, K., Han, L., Qi, J., Ro, D.-K., Bak, S., Huang, S., Zhou, Y., and Shang, Y. (2017). Developmentally regulated glucosylation of bitter triterpenoid in cucumber by the UDP-glucosyltransferase UGT73AM3. *Mol. Plant*, 10:1000–1003.

Zhou, J., Li, C.-L., Gao, F., Luo, X.-P., Li, Q.-Q., Zhao, H.-X., Yao, H.-P., Chen, H., Wang, A.-H., and Wu, Q. (2016). Characterization of three glucosyltransferase genes in Tartary buckwheat and their expression after cold stress. *J. Agric. Food. Chem.*, 64:6930–8.

Zuckerkandl, E. and Pauling, L. (1965). *Evolutionary Divergence and Convergence in Proteins*, volume 97, pages 97 – 166. Academic Press.

Appendix

Table S1: *A. strigosa* contigs mined from the 454 root tip transcriptome database.

contig00101	length=1687	numreads=270	contig12913	length=871	numreads=8
contig00128	length=1205	numreads=115	contig13120	length=601	numreads=6
contig00180	length=953	numreads=47	contig13140	length=58	numreads=12
contig00243	length=1136	numreads=16	contig13141	length=312	numreads=31
contig00260	length=1678	numreads=102	contig13142	length=58	numreads=31
contig00733	length=1562	numreads=76	contig13899	length=670	numreads=22
contig00931	length=923	numreads=32	contig14612	length=787	numreads=16
contig01092	length=1594	numreads=116	contig15351	length=1186	numreads=41
contig01194	length=1280	numreads=22	contig16327	length=764	numreads=94
contig01332	length=1581	numreads=95	contig16496	length=1318	numreads=33
contig01341	length=1363	numreads=19	contig16525	length=288	numreads=117
contig01577	length=890	numreads=45	contig17424	length=705	numreads=14
contig01599	length=795	numreads=8	contig17576	length=1785	numreads=136
contig01670	length=1611	numreads=61	contig17673	length=290	numreads=18
contig01781	length=1709	numreads=126	contig17930	length=891	numreads=10
contig01799	length=1660	numreads=139	contig18035	length=966	numreads=12
contig01989	length=1651	numreads=575	contig18257	length=1277	numreads=59
contig02132	length=712	numreads=16	contig18279	length=427	numreads=98
contig02314	length=1302	numreads=12	contig18280	length=637	numreads=91
contig02362	length=1725	numreads=79	contig18535	length=1616	numreads=55
contig02436	length=1552	numreads=85	contig19358	length=251	numreads=91
contig02480	length=91	numreads=88	contig20313	length=147	numreads=204
contig02699	length=588	numreads=9	contig20631	length=442	numreads=12
contig02980	length=721	numreads=21	contig21401	length=438	numreads=22
contig03158	length=1785	numreads=67	contig21862	length=1184	numreads=11
contig03745	length=586	numreads=26	contig22121	length=1562	numreads=61
contig03883	length=780	numreads=10	contig22388	length=1435	numreads=93
contig04347	length=790	numreads=15	contig22538	length=1554	numreads=118
contig04598	length=737	numreads=7	contig23002	length=1061	numreads=23
contig05602	length=1073	numreads=14	contig23339	length=302	numreads=5
contig05740	length=1712	numreads=38	contig23340	length=346	numreads=18
contig05827	length=1577	numreads=65	contig23453	length=466	numreads=7
contig06218	length=690	numreads=9	contig23586	length=208	numreads=205
contig06492	length=1619	numreads=44	contig23781	length=1814	numreads=44
contig06751	length=1693	numreads=77	contig23818	length=1629	numreads=147
contig07600	length=1667	numreads=71	contig24248	length=117	numreads=23
contig07784	length=1426	numreads=24	contig24249	length=140	numreads=22
contig08700	length=506	numreads=19	contig24696	length=830	numreads=49
contig08947	length=1708	numreads=76	contig24951	length=123	numreads=17
contig09778	length=1614	numreads=620	contig25505	length=431	numreads=99
contig10086	length=799	numreads=20	contig26167	length=1494	numreads=23
contig10188	length=523	numreads=7	contig26778	length=546	numreads=10
contig10189	length=552	numreads=7	contig26925	length=358	numreads=11
contig10190	length=990	numreads=39	contig26962	length=1548	numreads=116
contig10326	length=1791	numreads=91	contig27009	length=1115	numreads=92
contig10378	length=1459	numreads=88	contig27468	length=358	numreads=108
contig10759	length=50	numreads=26	contig27585	length=322	numreads=49
contig10760	length=51	numreads=20	contig27586	length=830	numreads=205
contig10761	length=79	numreads=26	contig28415	length=22	numreads=115
contig10762	length=125	numreads=20	contig28561	length=185	numreads=92
contig10763	length=171	numreads=20	contig28584	length=27	numreads=108
contig10772	length=1466	numreads=87	contig28651	length=428	numreads=223
contig11637	length=1360	numreads=82	contig28904	length=19	numreads=115
contig11873	length=1194	numreads=223	contig28981	length=92	numreads=49
contig12842	length=745	numreads=166	contig29220	length=146	numreads=49

Table S1: *A. strigosa* contigs mined from the 454 root tip transcriptome database

Table S2: Summary of *A. strigosa* UGT sequences mined from the 454 root tip transcriptome database

Enzyme	contig	UGT name	GeneID
AsUGT91	N/A	UGT91G16	AS01_006890_0018824
AsGT01194	contig01194		AS01_012363_0009496
AsGT01332	contig01332	UGT91F11	AS01_003718_0016409
AsGT18280	contig18280		AS01_002796_0056524
AsGT19358	contig19358		AS01_013835_0105701
AsGT18279	contig18279	UGT91G15	AS01_002740_0084538
AsGT00243	contig00243		n.d.
AsGT22027	contig22027		n.d.
AsGT00733	contig00733	UGT90G1	n.d.
AsGT01461	contig01461		n.d.
AsGT02436	contig02436	UGT701A5	AS01_000232_0183840
AsGT16f23	contig01781	UGT98B4	AS01_015816_0099017
AsGT12842	contig12842		AS01_011433_0183300
AsGT14h21	contig13141	UGT99C4	AS01_000040_0580230
AsAAT1	contig15351	UGT99D1	AS01_003827_0329694
AsGT000892-1	chimera†		AS01_000892_0475555
AsGT000892-2	chimera†		AS01_000892_0649070
AsGT27f7	contig16327	UGT99B9	AS01_018433_0019113
AsGT23781	contig23781	UGT99A6	AS01_002760_0003285
AsGT24i2	contig26925		AS01_009157_0340895
AsGT10433	contig10433		AS01_001764_0065978
AsGT11099	contig11099		AS01_003427_0107181
AsGT23141	contig23141		AS01_004255_0482187
AsGT14h20	N/A‡		AS01_000040_0343206
AsGT11i11	contig02362	UGT705A4	AS01_008603_0020929
AsGT3i21	contig02980		AS01_003018_0175304
AsGT17424	contig17424		AS01_003018_0175304
AsGT03295	contig03295		AS01_001569_0429119
AsGT16525	contig16525	UGT707A21	n.d.
AsGT22538	contig22538		n.d.
AsGT27009	contig27009		n.d.
AsGT28561	contig28561		n.d.
AsGT00931	contig00931		n.d.
AsGT06751	contig06751	UGT72J6	n.d.
AsGT08700	contig08700		n.d.
AsGT10772	contig10772		n.d.
AsGT16496	contig16496		n.d.
AsGT17930	contig17930		n.d.

AsGT01315	contig01315		n.d.
AsGT28947A	contig28947A	UGT706F6	n.d.
AsGT28947B	contig28947B	UGT706F7	n.d.
AsGT20n10	contig00101	UGT88C4	n.d.
AsGT01092	contig01092	UGT88C11	n.d.
AsGT05602	contig05602		n.d.
AsGT06492	contig06492	UGT88C12	n.d.
AsGT10189	contig10189		n.d.
AsGT25n16	contig10190		n.d.
AsGT17673	contig17673		n.d.
AsGt21862	contig21862		n.d.
AsGT23340	contig23340		n.d.
AsGT23586	contig23586		n.d.
AsGT23818	contig23818	UGT706E6	n.d.
AsGT27586	contig27586		n.d.
AsGT00260	contig00260	UGT85F24	n.d.
AsGT02132	contig02132		n.d.
AsGT05740	contig05740		n.d.
AsGT8i4	contig08947	UGT85F25	n.d.
AsGT16h6	contig09778	UGT85B2	n.d.
AsGT18035	contig18035		n.d.
AsGT26167	contig26167		n.d.
AsGT10811	contig10811		n.d.
AsGT01670	contig01670	UGT710C7	n.d.
AsGT04347	contig04347		n.d.
AsGT04598	contig04598		n.d.
AsGT4h2	contig17583		n.d.
AsGT03883	contig03883		n.d.
AsGT11637	contig11637		n.d.
AsGT18535	contig18535	UGT83E3	n.d.
AsGT23002	contig23002		n.d.
AsGT24951	contig24951		n.d.
AsGT17328	contig17328		n.d.
AsGT26962	contig26962	UGT87B3	n.d.
AsGT10326	contig10326	UGT86B4	n.d.
AsGT15a11	contig01799	UGT74H7	AS01_003827_0587827
AsGT01989	contig01989	UGT74H11	AS01_007207_0097636
AsGT02699	contig02699		AS01_004391_0093438
SAD10	contig07600	UGT74H5	AS01_003827_0182130
AsGT07784	contig07784		AS01_011633_0017154
AsGT01599	contig01599		AS01_001003_0114322
AsGT03158	contig03158	UGT75K6	AS01_001483_0570300

AsGT10759	contig10759		AS01_003944_0076156
AsGT28b19	contig10760		AS01_015483_0023201
AsGT17576	contig17576	UGT75J3	AS01_001959_0203229
AsGT11140	contig11140		AS01_014177_0041609
AsGT15275	contig15275		AS01_003944_0076156
AsGT1a15	contig00180		AS01_000525_0200999
AsGT01341	contig01341		AS01_000815_0292318
AsGT05827	contig05827	UGT93B16	AS01_000207_0282310
AsGT18257	contig18257		AS01_000815_0141040
AsGT27a12	contig22121	UGT93B8	AS01_009625_0058710
AsGT22388	contig22388		AS01_000815_0187839
AsGT24248	contig24248		AS01_002123_0323801
AsGT24a3	contig24249		AS01_000815_0222774
AsGT06218	contig06218		n.d.
AsGT12o13	contig10086		n.d.

† chimera with contig16327 (AsGT27f7) and in RACE PCR of AsGT27f7 (Louveau, 2013).

‡ (Louveau, 2013)

n.d. = not determined

Table S2: Summary of *A. strigosa* UGT sequences mined from the 454 root tip transcriptome database.

Table S3: Characterised UGTs used in phylogenetic analysis

Enzyme	Accession Number	UGT family	Group	*Plant species*	Reported activity	Reference
AtUGT79B1	Q9LVW3	UGT79	A	*Arabidopsis thaliana*	Anthocyanidin 3-*O*-glucoside 1,2-xylosyltransferase	Yonekura-Sakakibara et al. (2012)
AtUGT79B6	Q9FN26	UGT79	A	*Arabidopsis thaliana*	Flavonol 3-*O*-galactoside 1,2-glucosyltransferase	Yonekura-Sakakibara et al. (2014)
BpUGT94B1	Q5NTH0	UGT94	A	*Bellis perennis*	Anthocyanidin 3-*O*-glucoside 1,2-glucuronosyltransferase	Sawada et al. (2005)
CaUGT3	BAH80312	UGT94	A	*Catharanthus roseus*	Flavonoid 3-*O*-glucoside 1,6-glucosyltransferase (processive)	Masada et al. (2009)
Cm1-2RhaT1	AAL06646	UGT94	A	*Citrus maxima*	Flavonoid 7-*O*-glucoside 1,2-rhamnosyltransferase	Frydman et al. (2013)
Cs1-6RhaT	ABA18631	UGT79	A	*Citrus sinensis*	Flavonoid 7-*O*/3-*O* glucoside 1,6-rhamnosyltransferase	Frydman et al. (2013)
GjUGT94E5	F8WKW8	UGT94	A	*Gardenia jasminoides*	Apocarotenoid glucoside 1,6-glucosyltransferase	Nagatoshi et al. (2012)
GmUGT79A7	BAV56172	UGT79	A	*Glycine max*	flavonol-3-*O*-glucoside/galactoside 1,6-glucosyltransferase	Rojas Rodas et al. (2016)
GmUGT79A6	BAN91401	UGT79	A	*Glycine max*	Flavonol 3-*O*-glucoside/galactoside 1,6-rhamnosyltransferase	Rojas Rodas et al. (2014)
GmUGT79B30	BAR88077	UGT79	A	*Glycine max*	Flavonol 3-*O*-glucoside/galactoside 1,2-glucosyltransferase	Di et al. (2015)
GmUGT91H4	BAI99585	UGT91	A	*Glycine max*	Triterpene 3-*O*-galactoside 1,2-rhamnosyltransferase	Shibuya et al. (2010)
LeABRT2	BAU68118	UGT79	A	*Lobelia erinus*	Delphinidin 3-*O*-glucoside 1,6-rhamnosyltransferase	Hsu et al. (2017)
LeABRT4	BAU68119	UGT79	A	*Lobelia erinus*	Delphinidin 3-*O*-glucoside 1,6-rhamnosyltransferase	Hsu et al. (2017)
PgUGT94Q2	AGR44632	UGT94	A	*Panax ginseng*	Triterpene 3-*O*-glucoside 1,2-glucosyltransferase	Jung et al. (2014)
PhA3G1-6RhaT	Q43716	UGT79	A	*Petunia hybrida*	Anthocyanidin 3-*O*-glucoside 1,6-rhamnosyltransferase	Brugliera et al. (1994)
In3GGT	Q53UH4	UGT91	A	*Ipomoea nil*	Anthocyanidin 3-*O*-glucoside 1,2-glucosyltransferase	Morita et al. (2005)
SiUGT94D1	BAF99027	UGT94	A	*Sesamum indicum*	Sesaminol 2'-*O*-glucoside 1,6-glucosyltransferase	Noguchi et al. (2008)
SgUGT94-289-3		UGT94	A	*Siraitia grosvenorii*	Triterpene 24-*O*-glucoside/3-*O* glucoside 1,2-/1-6-glucosyltransferase	Itkin et al. (2016)
SlGAME18	XP_004243636		A	*Solanum lycopersicum*	Steroidal alkaloid 3-*O*-glucoside 1,2-glucosyltransferase	Itkin et al. (2013)
VpUGT94F1	BAI44133	UGT94	A	*Veronica persica*	Flavonoid 3-*O*-glucoside 1,2-glucosyltransferase	Ono et al. (2010)
UGT89A2-Col-0	Q9LZD8	UGT89	B	*Arabidopsis thaliana*	Dihydroxybenzoic acid xylosyltransferase	Chen and Li (2017)
UGT89A2-Col24		UGT89	B	*Arabidopsis thaliana*	Dihydroxybenzoic acid xylosyltransferase/glucosyltransferase	Chen and Li (2017)
AtUGT89B1	NP_177529	UGT89	B	*Arabidopsis thaliana*	Flavonoid/phenolic acid glucosyltransferase	Lim et al. (2004)
AtUGT89C1	AAF80123	UGT89	B	*Arabidopsis thaliana*	Flavonol 7-*O*-rhamnosyltransferase	Yonekura-Sakakibara et al. (2007)
AtUGT90A1	Q9ZVX4	UGT90	C	*Arabidopsis thaliana*		Ross et al. (2001)
PoUGT90A7	ACB56926	UGT90	C	*Pilosella officinarum*	flavonol glucosyltransferase	Witte et al. (2009)
AlcUGT73G1	AAP88406	UGT73	D	*Allium cepa*	Flavonoid glucosyltransferase	Kramer et al. (2003)
AtUGT73B3	AAM47999	UGT73	D	*Arabidopsis thaliana*	Flavonoid-7-*O*-glucosyltransferase	Kim et al. (2006)
AtUGT73C1	AEC09294	UGT73	D	*Arabidopsis thaliana*	Cytokinin glucosyltransferase 1	Gandia-Herrero et al. (2008)
BavUGT73C10	AFN26666	UGT73	D	*Barbarea vulgaris*	Triterpene-3-*O*-glucosyltransferase	Augustin et al. (2012)
BvUGT73A4	AAS94329	UGT73	D	*Beta vulgaris*	Flavonoid glucosyltransferase	Isayenkova et al. (2006)
CsUGT73A20	ALO19886	UGT73	D	*Camellia sinensis*		
CaUGT2	BAD29722	UGT73	D	*Catharanthus roseus*	Diarylheptanoid-3-*O*-glucosyltransferase	Kaminaga et al. (2004)
CbBet5OGT	CAB56231	UGT73	D	*Cleretum bellidiforme*	betanidin-5-*O*-glucosyltransferase	Vogt et al. (1999)
CsUGT73AM3	KGN59015	UGT73	D	*Cucumis sativus*	Triterpene-3-*O*-glucosyltranferase	Zhong et al. (2017)
DicGT4	BAD52006	UGT73	D	*Dianthus caryophyllus*		
FaGT7	Q2V6J9	UGT73	D	*Fragaria x ananassa*	Flavonol 3-*O*-glucosyltransferase 7	Griesser et al. (2008)
GtA3OGT	Q8H0F2	UGT73	D	*Gentiana triflora*	Anthocyanin 3'-*O*-beta-glucosyltransferase	Fukuchi-Mizutani et al. (2003)
GmUGT73F2	BAM29362	UGT73	D	*Glycine max*	Triterpene 22-*O*-arabinoside 1,3-glucosyltransferase	Sayama et al. (2012)
GmUGT73F4	BAM29363	UGT73	D	*Glycine max*	Triterpene 22-*O*-arabinoside 1,3-xylosyltransferase	Sayama et al. (2012)
GmUGT73P2	BAI99584	UGT73	D	*Glycine max*	Triterpene 3-*O*-glucoronide 1,2-galactosyltransferase	Shibuya et al. (2010)
GuUGAT	ANJ03631	UGT73	D	*Glycyrrhiza uralensis*	Triterpene 3-*O* glucoronosyltransferase/Triterpene 3-*O*-glucuronide 1,2-glucuronosyltransferase	Xu et al. (2016a)
LbUGT73A10	BAG80536	UGT73	D	*Lycium barbarum*	(+)-catechin 4'-*O*-textbeta-textscd-glucosyltransferase	Noguchi et al. (2008)
MtUGT73K1	AAW56091	UGT73	D	*Medicago trunctula*	Triterpene glucosyltransferase	Achnine et al. (2005)

MtUGT73F3	ACT34898	UGT73	D	*Medicago trunctula*	Triterpene 28-*O*-glucosyltransferase	Naoumkina et al. (2010)
PgUGT73AP1	ASU43996	UGT73	D	*Picea glauca*	Acetophenone 4-*O* glucosyltranferase	Mageroy et al. (2017)
PgUGT73AP3	ASU43997	UGT73	D	*Picea glauca*	Acetophenone 3-*O* glucosyltranferase	Mageroy et al. (2017)
RhsUGT73B6	AAS55083	UGT73	D	*Rhodiola sachalinensis*	Tyrosol glucosyltransferase	Ma et al. (2007)
ScbUBGT	BAA83484	UGT73	D	*Scutellaria baicalensis*	Baicalein 7-*O*-glucosyltransferase	Hirotani et al. (2000)
SlUGT73L4	ADQ37966	UGT73	D	*Solanum lycopersicum*	Steroidal alkaloid 3-*O*-glucoside 1,3-xylosyltransferase	Itkin et al. (2013)
StSGT3	ABB84472	UGT73	D	*Solanum tuberosum*	Steroidal alkaloid 3-*O*-glucoside/galactoside 1,2-rhamnosyltransferase	McCue et al. (2007)
VaAOG	Q8W3P8	UGT73	D	*Vigna angularis*	Abscisate beta-glucosyltransferase	Xu et al. (2002)
AmUGT88D3	ABR57234	UGT88	E	*Antirrhinum majus*	chalcone 4'-*O*-glucosyltransferase	Ono et al. (2006)
AtUGT71B6	NP_188815	UGT71	E	*Arabidopsis thaliana*	Abscisate beta-glucosyltransferase	Priest et al. (2006)
AtUGT71C1	NP_180536	UGT71	E	*Arabidopsis thaliana*	UDP-glucosyl transferase 71C1	Lim et al. (2008)
AtUGT72B1	Q9M156	UGT72	E	*Arabidopsis thaliana*	UDP-glycosyltransferase 72B1	Brazier-Hicks et al. (2007)
AtUGT88A1	AEE75831	UGT88	E	*Arabidopsis thaliana*	UDP-glucosyl transferase 88A1	Lim et al. (2004)
AtUGT72E2	AED98252	UGT72	E	*Arabidopsis thaliana*	Hydroxycinnamate 4-beta-glucosyltransferase	Lanot et al. (2006)
BvUGT71F1	AAS94330	UGT71	E	*Beta vulgaris*	Flavonoid glucosyltransferase	Isayenkova et al. (2006)
FcCGT	BBA18062	UGT708	E	*Citrus paradisi/Fortunella crassifolia*	Flavonoid 3'-C/5'-C-glucosyltransferase	Ito et al. (2017)
FeUGT708C1	BAP90360		E	*Fagopyrum esculentum*	Flavonoid C-glucosyltransferase	Nagatomo et al. (2014)
FaUGT71K3	XP_004294260	UGT71	E	*Fragaria x ananassa*	4-hydroxy-2,5-dimethyl-3(2H)-furanone glycosyltransferase	Song et al. (2016)
GtUF6CGT1	BAQ19550		E	*Gentiana triflora*	Flavonoid 6-C-glucosyltransferase	Sasaki et al. (2015b)
GmUGT708D1	I1L3T1		E	*Glycine max*	Hydroxyflavanone-2-C-glucosyltransferase	Hirade et al. (2015)
GmIF7GT	NP_001235161	UGT88	E	*Glycine max*	Isoflavone 7-*O*-glucosyltransferase	Noguchi et al. (2007)
HpUGT72B11	ACB56923	UGT72	E	*Hieracium pilosella*	Flavonoid glucosyltransferase	Witte et al. (2009)
InGTase1	BAF75917	UGT71	E	*Ipomoea nil*	Phytohormone glucosyltransferase	Suzuki et al. (2007)
Lp4CGT	BAE48240	UGT88	E	*Linaria vulgaris*	Chalcone 4'-glucosyltransferase	Ono et al. (2006)
MpUGT88A4	ABL85471	UGT88	E	*Maclura pomifera*	Benzopyrone glucosyltransferase	Tian et al. (2006)
MdPGT1	B3TKC8	UGT88	E	*Malus domestica*	Phloretin 2'-*O*-glucosyltransferase	Jugde et al. (2008)
MtUGT72L1	ACC38470	UGT72	E	*Medicago truncatula*	Proanthocyanidin precursor-specific UDP-glycosyltransferase	Pang et al. (2008)
MtUGT71G1	AAW56092	UGT71	E	*Medicago trunctula*	Triterpene glucosyltransferase	Achnine et al. (2005)
NtGT1b	BAB60721	UGT71	E	*Nicotiana tabacum*	Flavonoid glucosyltransferase	Taguchi et al. (2001)
OsCGT	CAQ77160		E	*Oryza sativa*	C-glucosyltransferase	Brazier-Hicks et al. (2009)
OsUGT707A3	BAC83989	UGT71	E	*Oryza sativa*	Flavonoid 3-*O*-glycosyltransferase	Ko et al. (2008)
PgUGT71A27	AIZ00429	UGT71	E	*Panax ginseng*	Triterpene 20-*O*-glucosyltransferase	Jung et al. (2014)
PfUGT88A7	BAG31949	UGT88	E	*Perilla frutescens*	Flavonoid glucosyltransferase	Noguchi et al. (2009)
PlUGT43	A0A172J2G3	UGT708	E	*Pueraria lobata*	Daidzein C-glucosyltransferase	Wang et al. (2017)
RfAS	Q9AR73	UGT72	E	*Rauvolfia serpentina*	Hydroquinone glucosyltransferase	Hefner et al. (2002)
RsUGT72B14	ACD87062	UGT72	E	*Rhodiola sachalinensis*	Tyrosol glucosyltransferase	Yu et al. (2011)
RhGT1	BAD99560	UGT88	E	*Rosa hybrida*	Anthocyanidin 5/3-*O*-glucosyltransferase	Ogata et al. (2005)
ScUGT5	BAJ11653	UGT88	E	*Sinningia cardinalis*	UDP-glucose:3-deoxyanthocyanidin 5-*O*-glucosyltransferase	Nakatsuka and Nishihara (2010)
SlUGT5	NP_001307116	UGT72	E	*Solanum lycopersicum*	Hydroquinone glucosyltransferase	Louveau et al. (2011)
SclUGT88D5	BAG31946	UGT88	E	*Scutellaria laeteviolacea var. yakusimensis*	Flavonoid glucuronyltransferase	Noguchi et al. (2009)
VpUGT88D8	BAH47552	UGT88	E	*Veronica persica*	flavonoid 7-*O*-glycosyltransferase	Ono et al. (2010)
ZmUGT708A6	A0A096SRM5	UGT708		*Zea mays*	Flavonoid 2-C/7-*O* glucosyltransferase	Falcone Ferreyra et al. (2013)
CrsUGT707B1	CCG85331	UGT707	E	*Crocus sativus*	Flavonol 3-*O*-glucoside 1,2-glucosyltransferase	Trapero et al. (2012)
AcF3GT1	ADC34700	UGT78	F	*Actinidia chinensis*	Anthocyanidin 3-*O*-galactosyltransferase	Montefiori et al. (2011)
AtUGT78D1	Q9S9P6	UGT78	F	*Arabidopsis thaliana*	Flavonol 3-*O*-glucosyltransferase	Jones et al. (2003)
ArcGaT	BAD06514	UGT78	F	*Aralia cordata*	Flavonoid 3-*O*-galactosyltransferase	Kubo et al. (2004)

CpF3OGT	ACS15351	UGT78	F	*Citrus x paradisi*	flavonol 3-*O*-glucosyltransferase	Owens and McIntosh (2009)
Ctan35GT	BAF49289	UGT78	F	*Clitoria ternatea*		Noda et al. (2017)
DcUCGalT1	AKI23632	UGT78	F	*Daucus carota L.*	Cyanidin 3-*O*-galactosyltransferase	Xu et al. (2016b)
DicGT1	BAD52003	UGT78	F	*Dianthus caryophyllus*		
DkF3galtase	BAI40148	UGT78	F	*Diospyros kaki*	Flavonoid 3-*O*-galactosyltransferase	Ikegami et al. (2009)
FtUFGT3	AOS85164	UGT78	F	*Fagpyrum tataricum*	Anthocyanidin 3-*O*-glucosyltransferase	Zhou et al. (2016)
FtUFGT1	AOS85162	UGT78	F	*Fagpyrum tataricum*	Anthocyanidin 3-*O*-glucosyltransferase	Zhou et al. (2016)
Fh3GT1	ADK75021	UGT78	F	*Freesia hybrid cultivar*	Anthocyanidin 3-*O*-glucosyltransferase	Sun et al. (2016)
SocGT4A	BAD89042	UGT73	F	*Solanum culeatissimum*	Steroidal-3-*O*-glucosyltransferase	Kohara et al. (2005)
VmUF3GaT	BAA36972	UGT78	F	*Vigna mungo*	Flavonoid 3-*O*-galactosyltransferase	Mato et al. (1998)
VvGT1	AAB81683	UGT78	F	*Vitis vinifera*	Anthocyanidin 3-*O*-glucosyltransferase	Ford et al. (1998)
AtUGT85A1	AAF18537	UGT85	G	*Arabidopsis thaliana*	Cytokinin-*O*-glucosyltransferase 2	Hou et al. (2004)
CrUGT85A23	F8WLS6	UGT85	G	*Catharanthus roseus*	Monoterpenoid glucosyltransferase	Asada et al. (2013)
GjUGT85A24	BAK55737	UGT85	G	*Gardenia jasminoides*	Monoterpene (iridoid) glucosyltransferase	Nagatoshi et al. (2011)
MeUGT85K4	AEO45781	UGT85	G	*Manihot esculenta*	Cyagenic glucoside glucosyltransferase	Kannangara et al. (2011)
MtUGT85H2	ABI94024	UGT85	G	*Medicago truncatula*	Isoflavonoid glucosyltransferase	Modolo et al. (2007)
PdUGT85A19	ABV68925	UGT85	G	*Prunus dulcis*	Cyanohydrin glucoside 1,6-glucosyltransferase	Franks et al. (2008)
SbUGT85B1	AAF17077	UGT85	G	*Sorghum bicolor*	Cyanohydrin glycosyltransferase UGT85B1	Hansen et al. (2003)
SrUGT85C2	AAR06916	UGT85	G	*Stevia rebaudiana*	Diterpenoid 13-*O*-glucosyltransferase	Richman et al. (2005)
VUGT85K14		UGT85	G	*Vitis labrusca x Vitis vinifera*	Furanone glucosyltransferase	Sasaki et al. (2015a)
AtUGT76C1	Q9FI99	UGT76	H	*Arabidopsis thaliana*	Cytokinin-7-*N*-glucosyltransferase 1	Hou et al. (2004)
AtUGT76C2	AED90933	UGT76	H	*Arabidopsis thaliana*	Cytokinin-*N*-glucosyltransferase 2	Hou et al. (2004)
AtUGT76D1	AEC07843	UGT76	H	*Arabidopsis thaliana*	Flavonoid-7-*O*-glucosyltransferase	Lim et al. (2004)
AtUGT76E1	AED97208	UGT76	H	*Arabidopsis thaliana*	UDP-glycosyltransferase	Lim et al. (2004)
CrUGT76A2	U3UA11	UGT76	H	*Catharanthus roseus*	UDP-glucose monoterpene glucosyltransferase	Asada et al. (2013)
DicGT5	BAD52007	UGT76	H	*Dianthus caryophyllus*		
OsUGT709A4	BAC80066	UGT76	H	*Oryza sativa*	Isoflavonoid-7-*O*-glucosyltransferase	Ko et al. (2008)
		UGT76	H	*Prunus persica L. Batsch*	Volatile ester glucosyltransferase	Wu et al. (2017)
SrUGT76G1	AAR06912	UGT76	H	*Stevia rebaudiana*	Diterpenoid 13-*O*-glucoside 1,3-glucosyltransferase	Richman et al. (2005)
ZmBx8	AAL57037	UGT76	H	*Zea mays*	UDP-glucosyltransferase BX8	von Rad et al. (2001)
AtUGT83A1	Q9SGA8	UGT83	I	*Arabidopsis thaliana*		Ross et al. (2001)
AtUGT87A1	O64732	UGT87	J	*Arabidopsis thaliana*		Ross et al. (2001)
AtUGT87A2	NP_001077979	UGT97	J	*Arabidopsis thaliana*		Wang et al. (2012)
AtUGT86A1	Q9SJL0	UGT86	K	*Arabidopsis thaliana*		Ross et al. (2001)
AtUGT74E2	NP_172059	UGT74	L	*Arabidopsis thaliana*	Auxin (IBA) glycosyltransferase	Tognetti et al. (2010)
AtUGT75B1	AEE27854	UGT75	L	*Arabidopsis thaliana*	Indole-3-acetate beta-glucosyltransferase 1	Eudes et al. (2008)
AtUGT84A1	Q5XF20	UGT84	L	*Arabidopsis thaliana*	Hydroxycinnamate glucosyltransferase 2	Milkowski et al. (2000)
AsSAD10	ACD03250	UGT74	L	*Avena strigosa*	*N*-methyl anthranilate glucosyltransferase	Owatworakit et al. (2013)
AsUGT74H6	ACD03261	UGT74	L	*Avena strigosa*	Aromatic acid glucosyltransferase	Owatworakit et al. (2013)
BnUGT84A9a	CAS03354	UGT84	L	*Brassica napus*	Sinapate/hydroxycinnamate glucosyltransferase	Mittasch et al. (2010)
CuLGT	BAA93039		L	*Citrus unshui*	Triterpene (limonoid)-17-*O*-glucosyltransferase	Kita et al. (2000)
CrsGT45	ACM66950	UGT75	L	*Crocus sativus*	Flavonoid 7-*O*-glucosyltransferase	Moraga et al. (2009)
CrsGLT2	Q6X1C0	UGT74	L	*Crocus sativus*	Apocarotenoid (crocetin) glucosyltransferase 2	Moraga et al. (2004)
DgUGT1	BAO66179	UGT84	L	*Delphinium grandiflorum*	UDP-glucose dependent *p*-hydroxybenzoic acid glucosyltransferase	Nishizaki et al. (2014)
EpGT-1	BAD90934	UGT75	L	*Eucalyptus perriniana*		
FaGT2	Q66PF4	UGT84	L	*Fragaria ananassa*	Cinnamate β-D-glucosyltransferase	Lunkenbein et al. (2006)
GjUGT75L6	F8WKW0	UGT75	L	*Gardenia jasminoides*	Apocarotenoid glucosyltransferase	Nagatoshi et al. (2012)

HvUGT13248	ADC92550	UGT84	L	*Hordeum vulgare*	Epoxy-sesquiterpenoid-3-*O*-glucosyltransferase	Schweiger et al. (2010)
MpUGT75L4	ABL85474	UGT75	L	*Maclura pomifera*	Flavonoid glycosyltransferase	Tian et al. (2006)
VhUGT74M1	ABK76266	UGT74	L	*Vaccaria hispanica*	Triterpene carboxylic acid28 -*O*- glucosyltransferase	Meesapyodsuk et al. (2007)
MdUGT75L17	AAX16493	UGT75	L	*Malus x domestica Brokh.*	UDP-glucose:phloretin 4'-*O*-glycosyltrasferase	Yahyaa et al. (2016)
NtToGT1	Q9AT54	UGT73	L	*Nicotiana tabacum*	Scopoletin glucosyltransferase	Fraissinet-Tachet et al. (1998)
OsUGT79	XP_015635481	UGT74	L	*Oryza sativa Japonica*	Epoxy-sesquiterpene 3-*O*-glucosyltransferase	Michlmayr et al. (2015)
			L	*Pisum sativum*	Auxin (IAA) glycosyltransferase	Ostrowski et al. (2015)
PgUGT74A1	AGR44631	UGT74	L	*Panax ginseng*	Triterpene-3-*O*-glucosyltransferase	Jung et al. (2014)
Pf3R4	BAA36421	UGT75	L	*Perilla frutescens*	Anthocyanin 5-*O*-glucosyltransferase	Yamazaki et al. (1999)
PfxPaUGT84A17	AII32448	UGT84	L	*Populus fremontii x Populus angustifolia*	Hydroxycinnamate glycosyltransferase	Babst et al. (2014)
PugUGT84A23	ANN02875	UGT84	L	*Punica granatum*	Gallic acid glucosyltransferase	Ono et al. (2016)
PcF7GT	AAY27090	UGT75	L	*Pyrus communis*		
RhsUGT74R1	ABP49574	UGT74	L	*Rhodiola sachalinensis*		
SgUGT74AC1	AEM42999	UGT74	L	*Siraitia grosvenorii*		
SlGtsatom	CAI62049	UGT74	L	*Solanum lycopersicum*		
SrUGT74G1	Q6VAA6	UGT74	L	*Stevia rebaudiana*	Diterpenoid-19-*O*-glucosyltransferase	Richman et al. (2005)
Via5GT	AHL68667	UGT75	L	*Vitis amurensis Rupr. cv. 'Zuoshanyi*	Anthocyanin 5-*O*-glucosyltransferase	He et al. (2015)
ZmIAGT	AAA59054	UGT74	L	*Zea mays*	Auxin glucosyltransferase	Szerszen et al. (1994)
AtUGT92A1	Q9LXV0	UGT92	M	*Arabidopsis thaliana*		Ross et al. (2001)
CcDOPA5GT	BAD91804	UGT92	M	*Celosia cristata*	Cyclo-DOPA 5-*O*-glucosyltransferase	Sasaki et al. (2005)
MjcDOPA5GT	BAD91803	UGT92	M	*Mirabilis jalapa*	Cyclo-DOPA 5-*O*-glucosyltransferase	Sasaki et al. (2005)
AtUGT82A1	Q9LHJ2	UGT82	N	*Arabidopsis thaliana*		Ross et al. (2001)
	ONM15823.1	UGT82	N	*Zea mays*		
PlZOG1	Q9ZSK5	UGT93	O	*Phaseolus lunatus*	Trans-zeatin *O*-beta-D-glucosyltransferase	Martin et al. (1999)
PvZOX1	P56725	UGT93	O	*Phaseolus vulgaris*	Zeatin *O*-textbeta -textscd-xylosyltransferase	Martin et al. (1999)
ZmcisZog1	AAK53551	UGT93	O	*Zea mays*	cis-zeatin *O*-glucosyltransferase	Martin et al. (2001)
ZmcisZog2	AAL92460	UGT93	O	*Zea mays*	cis-zeatin *O*-glucosyltransferase 2	Veach et al. (2003)
SlGAME17	XP_004243637	UGT93	O	*Solanum lycopersicum*	Steroidal alkaloid 3-*O*-galactoside 1,4-glucosyltransferase	Itkin et al. (2013)
PoUGT95A1	ACB56927	UGT95		*Pilosella officinarum*		
SgUGT720-269-1		UGT720	P	*Siraitia grosvenorii*	Triterpene 24-*O*/3-*O* glucosyltransferase	Itkin et al. (2016)

Table S3: Characterised UGTs used in phylogenetic analysis

Table S4: Summary of *A. strigosa* genes co-expressed with the avenacin biosynthetic cluster

Clustered with avenacin genes	Gene name	Annotation
100%	AS01_012650_0384107	2-oxo_acid_dehydrogenase,_lipoyl-binding_site
28%	AS01_004745_0032580	2Fe-2S_ferredoxin,_iron-sulphur_binding_site
39%	AS01_012875_0060530	2Fe-2S_ferredoxin,_iron-sulphur_binding_site
100%	AS01_015805_0033457	2Fe-2S_ferredoxin,_iron-sulphur_binding_site
100%	AS01_003804_0162096	3-oxoacyl-(acyl-carrier-protein)_reductase
65%	AS01_006876_0227474	3-Oxoacyl-[acyl-carrier-protein_(ACP)]_synthase_III_C-terminal
17%	AS01_005097_0073724	3-oxoacyl-[acyl-carrier-protein]_synthase_2
100%	AS01_008687_0025539	30s_ribosomal_protein_S13,_C-terminal
17%	AS01_001440_0369202	40S_ribosomal_protein_S1/3,_eukaryotes
20%	AS01_002388_0760385	40S_ribosomal_protein_S1/3,_eukaryotes
35%	AS01_006009_0008645	5'-3'_exonuclease,_C-terminal_domain
100%	AS01_003660_0027543	50S_ribosomal_protein_L30e-like
5%	AS01_006211_0161273	50S_ribosomal_protein_L30e-like
18%	AS01_010567_0422925	50S_ribosomal_protein_L30e-like
100%	AS01_011855_0007290	50S_ribosomal_protein_L30e-like
86%	AS01_012796_0066794	6-phosphogluconate_dehydrogenase,_C-terminal
5%	AS01_001149_0185212	6-phosphogluconate_dehydrogenase,_C-terminal-like
100%	AS01_004734_0258615	6-phosphogluconate_dehydrogenase,_C-terminal-like
11%	AS01_006190_0078154	6-phosphogluconate_dehydrogenase,_C-terminal-like
6%	AS01_008962_0173231	60S_ribosomal_protein_L4,_C-terminal_domain
100%	AS01_000035_0176986	AAA+_ATPase_domain
100%	AS01_000125_0035510	AAA+_ATPase_domain
100%	AS01_000944_0072793	AAA+_ATPase_domain
20%	AS01_001240_0007383	AAA+_ATPase_domain
100%	AS01_002154_0076095	AAA+_ATPase_domain
100%	AS01_003033_0023116	AAA+_ATPase_domain
100%	AS01_003056_0065509	AAA+_ATPase_domain
1%	AS01_006037_0318233	AAA+_ATPase_domain
97%	AS01_007568_0393072	AAA+_ATPase_domain
100%	AS01_007887_0064712	AAA+_ATPase_domain
100%	AS01_009919_0048873	AAA+_ATPase_domain
12%	AS01_010369_0586680	AAA+_ATPase_domain
20%	AS01_012820_0083732	AAA+_ATPase_domain
100%	AS01_006614_0101831	Acid_phosphatase,_type_5
95%	AS01_007359_0089494	Acyl-CoA_N-acyltransferase
3%	AS01_011461_0217222	Adenylate_kinase_subfamily
100%	AS01_006447_0270330	Aldo/keto_reductase_subgroup
100%	AS01_006447_0613930	Aldo/keto_reductase_subgroup
100%	AS01_010045_0236967	Aldo/keto_reductase_subgroup
20%	AS01_018997_0014671	Aldo/keto_reductase_subgroup
100%	AS01_000507_0622140	Aldolase-type_TIM_barrel
100%	AS01_000776_0267412	Aldolase-type_TIM_barrel
20%	AS01_003211_0263616	Aldolase-type_TIM_barrel
20%	AS01_003993_0256058	Aldolase-type_TIM_barrel
3%	AS01_007333_0106136	Aldolase-type_TIM_barrel
20%	AS01_008796_0005394	Aldolase-type_TIM_barrel
100%	AS01_009111_0117041	Aldolase-type_TIM_barrel
100%	AS01_004753_0260854	Aldose_1-/Glucose-6-phosphate_1-epimerase
100%	AS01_000041_0115475	Alkaline_phosphatase-like,_alpha/beta/alpha
100%	AS01_011528_0022677	Allergen_V5/Tpx-1-related,_conserved_site
100%	AS01_013634_0090131	Allergen_V5/Tpx-1-related,_conserved_site
100%	AS01_004802_0212317	Alpha_crystallin/Hsp20_domain
100%	AS01_000032_0289442	Alpha/Beta_hydrolase_fold
100%	AS01_001096_0030884	Alpha/Beta_hydrolase_fold

5%	AS01_002172_0096998	Alpha/Beta_hydrolase_fold
100%	AS01_002725_0041619	Alpha/Beta_hydrolase_fold
100%	AS01_003827_0163235	Alpha/Beta_hydrolase_fold
100%	AS01_003832_0115743	Alpha/Beta_hydrolase_fold
68%	AS01_004947_0056780	Alpha/Beta_hydrolase_fold
51%	AS01_005116_0037590	Alpha/Beta_hydrolase_fold
20%	AS01_005914_0177849	Alpha/Beta_hydrolase_fold
100%	AS01_005997_0231296	Alpha/Beta_hydrolase_fold
64%	AS01_007090_0386988	Alpha/Beta_hydrolase_fold
100%	AS01_008190_0164813	Alpha/Beta_hydrolase_fold
96%	AS01_008303_0036296	Alpha/Beta_hydrolase_fold
100%	AS01_010577_0074969	Alpha/Beta_hydrolase_fold
100%	AS01_010684_0114869	Alpha/Beta_hydrolase_fold
100%	AS01_011357_0030722	Alpha/Beta_hydrolase_fold
20%	AS01_011744_0114519	Alpha/Beta_hydrolase_fold
99%	AS01_011749_0018765	Alpha/Beta_hydrolase_fold
100%	AS01_011749_0038591	Alpha/Beta_hydrolase_fold
90%	AS01_012070_0003689	Alpha/Beta_hydrolase_fold
100%	AS01_015502_0037686	Alpha/Beta_hydrolase_fold
100%	AS01_006585_0127444	Alternative_oxidase
100%	AS01_013747_0096713	Amidohydrolase_1
63%	AS01_007110_0026533	Amidophosphoribosyl_transferase
100%	AS01_013716_0017526	Amine_oxidase
95%	AS01_006242_0127424	Amino_acid_transporter,_transmembrane
23%	AS01_012779_0135997	Amino_acid_transporter,_transmembrane
30%	AS01_003494_0235687	Aminoacyl-tRNA_synthetase,_class_II
100%	AS01_004403_0297824	Aminotransferase,_class_I/classII
100%	AS01_007573_0328460	Aminotransferase,_class_I/classII
100%	AS01_010095_0371969	Aminotransferase,_class_I/classII
100%	AS01_020319_0004566	Aminotransferase,_class_I/classII
41%	AS01_013674_0107888	Aminotransferase,_class_IV
100%	AS01_010873_0184751	AMP-binding_enzyme_C-terminal_domain
100%	AS01_016552_0074338	AMP-binding_enzyme_C-terminal_domain
100%	AS01_005234_0398423	Anaphase-promoting_complex,_subunit_10/DOC_domain
15%	AS01_006011_0366501	Ankyrin_repeat
33%	AS01_008964_0016764	Ankyrin_repeat
30%	AS01_011289_0157053	Annexin
17%	AS01_009996_0040896	Anticodon-binding
84%	AS01_000367_0050374	AP2/ERF_domain
38%	AS01_001457_0265120	AP2/ERF_domain
100%	AS01_002444_0599709	AP2/ERF_domain
100%	AS01_003363_0303028	AP2/ERF_domain
100%	AS01_005531_0558003	AP2/ERF_domain
100%	AS01_005574_0035438	AP2/ERF_domain
96%	AS01_009261_0010169	AP2/ERF_domain
96%	AS01_010396_0421106	AP2/ERF_domain
100%	AS01_010801_0037805	AP2/ERF_domain
9%	AS01_014420_0216424	AP2/ERF_domain
100%	AS01_004712_0196161	APOBEC/CMP_deaminase,_zinc-binding
100%	AS01_002554_0186416	Apple-like
100%	AS01_006156_0230463	Aquaporin-like
100%	AS01_004055_0133683	Arabinogalactan_peptide,_AGP
98%	AS01_000923_0123126	Archaeal_RpoH_/eukaryotic_RPB5_RNA_polymerase_subunit
2%	AS01_012700_0306674	Armadillo-like_helical
100%	AS01_000190_0038695	Armadillo-type_fold
100%	AS01_010145_0027800	Armadillo-type_fold
100%	AS01_010792_0213288	Armadillo-type_fold
98%	AS01_013986_0154719	Aromatic-L-amino-acid_decarboxylase
100%	AS01_003938_0244634	Aromatic-ring_hydroxylase-like
1%	AS01_006159_0233551	Aromatic-ring_hydroxylase-like

100%	AS01_009022_0331534	Aromatic-ring_hydroxylase-like
100%	AS01_009022_0394291	Aromatic-ring_hydroxylase-like
100%	AS01_009625_0493932	Aromatic-ring_hydroxylase-like
100%	AS01_006816_0093962	Aspartic_peptidase
100%	AS01_008231_0280456	Aspartic_peptidase
100%	AS01_002678_0139164	Aspartic_peptidase_domain
20%	AS01_012841_0004484	Aspartic_peptidase_domain
100%	AS01_012841_0040135	Aspartic_peptidase_domain
100%	AS01_014294_0075114	Aspartic_peptidase_domain
100%	AS01_014294_0077943	Aspartic_peptidase_domain
20%	AS01_015558_0056647	Aspartic_peptidase_domain
96%	AS01_012439_0015632	ATP_synthase_subunit_alpha-like_domain
100%	AS01_003595_0396783	ATP11
100%	AS01_002780_0159282	ATPase,_F0_complex,_B_chain/subunit_B/MI25
22%	AS01_008314_0067066	ATPase,_F1_complex_alpha/beta_subunit,_N-terminal_domain
100%	AS01_001651_0757622	ATPase,_F1/V1/A1_complex,_alpha/beta_subunit,_nucleotide-binding_domain
68%	AS01_004978_0017297	ATPase,_V1/A1_complex,_subunit_E
1%	AS01_004022_0301133	AUX/IAA_protein
100%	AS01_004022_0448106	AUX/IAA_protein
100%	AS01_011708_0157643	AUX/IAA_protein
16%	AS01_004255_0259283	Auxin_efflux_carrier
100%	AS01_008385_0061731	Auxin_efflux_carrier
100%	AS01_016090_0062519	Auxin-induced_protein,_ARG7
1%	AS01_002300_0833320	B3_DNA_binding_domain
100%	AS01_007066_0174088	B3_DNA_binding_domain
61%	AS01_012678_0094694	B3_DNA_binding_domain
57%	AS01_020031_0015278	B3_DNA_binding_domain
100%	AS01_000647_0292208	Bacterial_Fmu_(Sun)/eukaryotic_nucleolar_NOL1/Nop2p
64%	AS01_008510_0500030	Bacterial_transferase_hexapeptide_repeat
92%	AS01_010387_0080736	Bacterial_transferase_hexapeptide_repeat
50%	AS01_002657_0284720	BAG_domain
100%	AS01_000616_0259449	Band_7_protein
100%	AS01_002668_0088967	Barwin-like_endoglucanase
100%	AS01_003355_0226793	Barwin-like_endoglucanase
100%	AS01_004301_0383567	Barwin-like_endoglucanase
100%	AS01_004301_0480482	Barwin-like_endoglucanase
100%	AS01_004370_0425388	Barwin-like_endoglucanase
100%	AS01_004463_0367549	Barwin-like_endoglucanase
100%	AS01_005205_0071559	Barwin-like_endoglucanase
20%	AS01_009022_0311118	Barwin-like_endoglucanase
100%	AS01_019922_0004306	Barwin-like_endoglucanase
100%	AS01_002551_0093977	BCNT-C_domain
100%	AS01_001726_0200492	Berberine/berberine-like
100%	AS01_001853_0064263	Beta-glucanase
1%	AS01_005225_0039216	Beta-glucanase
100%	AS01_005806_0349684	Beta-glucanase
100%	AS01_009779_0071484	Beta-glucanase
100%	AS01_012446_0016688	Beta-hexosaminidase
100%	AS01_002168_0177302	Bifunctional_inhibitor/plant_lipid_transfer_protein/seed_storage_helical_domain
100%	AS01_002168_0180486	Bifunctional_inhibitor/plant_lipid_transfer_protein/seed_storage_helical_domain
100%	AS01_002979_0006917	Bifunctional_inhibitor/plant_lipid_transfer_protein/seed_storage_helical_domain
100%	AS01_003726_0085665	Bifunctional_inhibitor/plant_lipid_transfer_protein/seed_storage_helical_domain
100%	AS01_003726_0233183	Bifunctional_inhibitor/plant_lipid_transfer_protein/seed_storage_helical_domain
100%	AS01_005384_0062831	Bifunctional_inhibitor/plant_lipid_transfer_protein/seed_storage_helical_domain
100%	AS01_005428_0072851	Bifunctional_inhibitor/plant_lipid_transfer_protein/seed_storage_helical_domain
100%	AS01_010694_0140762	Bifunctional_inhibitor/plant_lipid_transfer_protein/seed_storage_helical_domain
100%	AS01_012005_0115021	BING4,_C-terminal_domain
100%	AS01_001295_0044052	Biopterin_transport-related_protein_BT1
20%	AS01_004491_0256843	Biopterin_transport-related_protein_BT1

62%	AS01_006043_0169886	Brix_domain
2%	AS01_002787_0239101	BTB/POZ_fold
100%	AS01_011555_0005620	BTB/POZ_fold
100%	AS01_002580_0683241	BURP_domain
100%	AS01_011341_0291411	BURP_domain
100%	AS01_003531_0070455	Bystin
93%	AS01_007410_0071151	C2_domain
70%	AS01_013823_0028601	C2_domain
100%	AS01_003827_0230264	Caffeate_O-methyltransferase_(COMT)_family
100%	AS01_004403_0268098	Caffeate_O-methyltransferase_(COMT)_family
100%	AS01_002888_0792374	Calcineurin-like_phosphoesterase_domain,_apaH_type
100%	AS01_009473_0327103	Calcineurin-like_phosphoesterase_domain,_apaH_type
24%	AS01_006293_0788762	Calcineurin-like_phosphoesterase_domain,_lpxH_type
80%	AS01_001084_0143020	Calcium/proton_exchanger
14%	AS01_013089_0421516	Caleosin
61%	AS01_010299_0121696	Calmodulin-binding_domain,_plant
20%	AS01_002399_0556291	Calponin_homology_domain
64%	AS01_008249_0461786	Carbon-nitrogen_hydrolase
100%	AS01_006961_0011778	Casein_kinase_II,_regulatory_subunit
99%	AS01_000403_0439915	CBF1-interacting_co-repressor_CIR,_N-terminal_domain
100%	AS01_001149_0180558	CBF1-interacting_co-repressor_CIR,_N-terminal_domain
100%	AS01_007055_0225447	CBS_domain
100%	AS01_001189_0009733	Chalcone/stilbene_synthase,_C-terminal
100%	AS01_001189_0026525	Chalcone/stilbene_synthase,_C-terminal
100%	AS01_005097_0185387	Chalcone/stilbene_synthase,_C-terminal
98%	AS01_001618_1123740	Chaperone_DnaK
100%	AS01_014761_0007636	Chaperone_DnaK
10%	AS01_001488_0444216	Chaperonin_Cpn10
13%	AS01_003305_0014848	Chaperonin_Cpn60
8%	AS01_009286_0043762	Chaperonin_Cpn60
100%	AS01_008197_0020875	CheY-like_superfamily
100%	AS01_000317_0636117	Chloramphenicol_acetyltransferase-like_domain
97%	AS01_002142_0018192	Chloramphenicol_acetyltransferase-like_domain
100%	AS01_004187_0024720	Chloramphenicol_acetyltransferase-like_domain
100%	AS01_005853_0649187	Chloramphenicol_acetyltransferase-like_domain
100%	AS01_007800_0004197	Chloramphenicol_acetyltransferase-like_domain
100%	AS01_010669_0054541	Chloramphenicol_acetyltransferase-like_domain
100%	AS01_012659_0004513	Chloramphenicol_acetyltransferase-like_domain
35%	AS01_014148_0094215	Chromatin_assembly_factor_1_subunit_A
20%	AS01_000082_0035546	CO_dehydrogenase_flavoprotein-like,_FAD-binding,_subdomain_2
100%	AS01_001739_0115042	CO_dehydrogenase_flavoprotein-like,_FAD-binding,_subdomain_2
100%	AS01_001853_0001565	Concanavalin_A-like_lectin/glucanase,_subgroup
100%	AS01_001853_0121656	Concanavalin_A-like_lectin/glucanase,_subgroup
100%	AS01_001864_0168294	Concanavalin_A-like_lectin/glucanase,_subgroup
100%	AS01_002182_0382999	Concanavalin_A-like_lectin/glucanase,_subgroup
100%	AS01_003250_0130327	Concanavalin_A-like_lectin/glucanase,_subgroup
100%	AS01_003504_0294638	Concanavalin_A-like_lectin/glucanase,_subgroup
4%	AS01_006362_0377680	Concanavalin_A-like_lectin/glucanase,_subgroup
71%	AS01_007479_0321283	Concanavalin_A-like_lectin/glucanase,_subgroup
20%	AS01_007502_0073832	Concanavalin_A-like_lectin/glucanase,_subgroup
100%	AS01_008662_0004501	Concanavalin_A-like_lectin/glucanase,_subgroup
3%	AS01_009734_0030843	Concanavalin_A-like_lectin/glucanase,_subgroup
100%	AS01_013978_0064871	Concanavalin_A-like_lectin/glucanase,_subgroup
100%	AS01_000098_1043704	Concanavalin_A-like_lectin/glucanases_superfamily
20%	AS01_007167_0318719	Copper_amine_oxidase
23%	AS01_016324_0035181	Copper_amine_oxidase,_N2/N3-terminal
23%	AS01_004460_0246630	CRAL-TRIO_domain
100%	AS01_008292_0041576	CRAL-TRIO_domain

91%	AS01_008920_0079813	CRAL-TRIO_domain
65%	AS01_003531_0828943	Cullin_repeat-like-containing_domain
100%	AS01_003137_0384196	Cupredoxin
21%	AS01_005103_0138409	Cupredoxin
100%	AS01_011449_0262871	Cupredoxin
20%	AS01_013466_0153454	Cupredoxin
100%	AS01_002688_0280947	Cyclin_P/U
100%	AS01_003262_0509019	Cyclophilin-like_domain
100%	AS01_009182_0052374	Cyclophilin-like_domain
27%	AS01_008624_0142756	Cysteine_alpha-hairpin_motif_superfamily
6%	AS01_011377_0116240	Cysteine_alpha-hairpin_motif_superfamily
100%	AS01_000207_0151615	Cysteine_peptidase,_asparagine_active_site
100%	AS01_005367_0213228	Cysteine_peptidase,_asparagine_active_site
100%	AS01_000207_0128881	Cysteine_peptidase,_cysteine_active_site
93%	AS01_009779_0062527	Cysteine_synthase_A
100%	AS01_001099_0015309	Cytochrome_b5,_heme-binding_site
3%	AS01_001243_0307648	Cytochrome_b5,_heme-binding_site
1%	AS01_008475_0012989	Cytochrome_b561,_eukaryote
20%	AS01_000040_0285788	Cytochrome_P450
100%	AS01_000207_0258224	Cytochrome_P450
100%	AS01_000207_0273082	Cytochrome_P450
100%	AS01_001100_0713812	Cytochrome_P450
100%	AS01_001219_0224457	Cytochrome_P450
18%	AS01_001693_0328482	Cytochrome_P450
67%	AS01_002181_0025085	Cytochrome_P450
100%	AS01_002420_0411327	Cytochrome_P450
11%	AS01_002793_0697039	Cytochrome_P450
100%	AS01_003274_0135372	Cytochrome_P450
100%	AS01_003749_0204332	Cytochrome_P450
100%	AS01_003827_0017073	Cytochrome_P450
100%	AS01_003827_0458945	Cytochrome_P450
100%	AS01_003827_0541552	Cytochrome_P450
100%	AS01_003827_0568353	Cytochrome_P450
100%	AS01_003906_0073703	Cytochrome_P450
100%	AS01_004403_0510975	Cytochrome_P450
100%	AS01_004926_0026134	Cytochrome_P450
100%	AS01_005146_0088305	Cytochrome_P450
20%	AS01_006213_0009007	Cytochrome_P450
100%	AS01_008603_0009691	Cytochrome_P450
20%	AS01_009487_0225226	Cytochrome_P450
100%	AS01_011324_0069982	Cytochrome_P450
100%	AS01_012751_0110146	Cytochrome_P450
100%	AS01_012751_0140393	Cytochrome_P450
100%	AS01_012751_0168282	Cytochrome_P450
100%	AS01_014689_0131873	Cytochrome_P450
100%	AS01_017455_0016538	Cytochrome_P450
71%	AS01_018556_0038357	Cytochrome_P450
98%	AS01_003159_0323495	DBC1/CARP1_catalytically_inactive_NUDIX_hydrolase_domain
6%	AS01_005705_0259143	DBC1/CARP1_catalytically_inactive_NUDIX_hydrolase_domain
20%	AS01_002411_0043270	DDT_domain
91%	AS01_003087_0058547	DDT_domain
100%	AS01_000362_1362948	DEAD/DEAH_box_helicase_domain
68%	AS01_002330_0246214	DEAD/DEAH_box_helicase_domain
10%	AS01_004045_0075244	DEAD/DEAH_box_helicase_domain
73%	AS01_005298_0190929	DEAD/DEAH_box_helicase_domain
100%	AS01_005997_0154560	DEAD/DEAH_box_helicase_domain
10%	AS01_007331_0193737	DEAD/DEAH_box_helicase_domain
100%	AS01_010684_0238832	Decaprenyl_diphosphate_synthase-like
100%	AS01_004187_0043293	Dehydrogenase,_E1_component
1%	AS01_013835_0433202	DEK,_C-terminal

100%	AS01_014461_0137826	DEP_domain
2%	AS01_013568_0052777	Dihydrodipicolinate_reductase,_C-terminal
100%	AS01_007821_0161127	Diphosphomevalonate_decarboxylase
9%	AS01_018451_0022598	Diphthine_synthase
100%	AS01_000810_0869139	Disulphide_isomerase
100%	AS01_004480_0159147	Disulphide_isomerase
18%	AS01_007235_0054765	DNA-binding_WRKY
7%	AS01_011924_0108747	DNA-directed_RNA_polymerase_III_subunit_RPC4
100%	AS01_000490_0175072	DNA-directed_RNA_polymerase_M,_15kDa_subunit,_conserved_site
100%	AS01_001863_0650690	DNA-directed_RNA_polymerase,_RBP11-like_dimerisation_domain
100%	AS01_018221_0004024	DNA-directed_RNA_polymerase,_RBP11-like_dimerisation_domain
100%	AS01_008424_0033869	Domain_of_unknown_function_DUF125,_transmembrane
100%	AS01_019301_0008839	Domain_of_unknown_function_DUF1336
21%	AS01_002017_1178736	Domain_of_unknown_function_DUF296
88%	AS01_011618_0333255	Domain_of_unknown_function_DUF4094
100%	AS01_006676_0637298	Domain_of_unknown_function_DUF4110
3%	AS01_012837_0076729	Domain_of_unknown_function_DUF4187
100%	AS01_000533_0041695	Domain_of_unknown_function_DUF4220
100%	AS01_002529_0078425	Domain_of_unknown_function_DUF4220
100%	AS01_006774_0286434	Domain_of_unknown_function_DUF4408
100%	AS01_005720_0317124	Domain_of_unknown_function_DUF828
100%	AS01_011989_0093299	DOMON_domain
3%	AS01_003208_0155726	Drug/metabolite_transporter
100%	AS01_008877_0167519	Drug/metabolite_transporter
100%	AS01_010542_0286845	Drug/metabolite_transporter
1%	AS01_002181_0118992	Dual_specificity_phosphatase,_catalytic_domain
100%	AS01_007623_0242836	DVL
100%	AS01_014926_0018823	Dynamin_GTPase_effector
100%	AS01_001693_0210784	Dyskerin-like
100%	AS01_003814_0176942	Dyskerin-like
17%	AS01_014940_0049012	EF_hand_associated,_type-1
100%	AS01_002301_0102771	EF-Hand_1,_calcium-binding_site
5%	AS01_005134_0094044	EF-Hand_1,_calcium-binding_site
4%	AS01_006194_0083028	EF-Hand_1,_calcium-binding_site
5%	AS01_443019_0003520	EF-Hand_1,_calcium-binding_site
98%	AS01_010543_0249141	Elongation_factor_Tu-type_domain
40%	AS01_001953_0032324	Elongation_factor,_GTP-binding_domain
5%	AS01_004452_0028567	Elongation_factor,_GTP-binding_domain
99%	AS01_001661_0531571	Embryo-specific_3
99%	AS01_004307_0025167	Emopamil-binding_protein
100%	AS01_012448_0021728	Enamine/imine_deaminase_YjgF-like
38%	AS01_013667_0157660	Endonuclease/exonuclease/phosphatase
30%	AS01_001634_0217334	Endosulphine
98%	AS01_004528_0196408	Endosulphine
100%	AS01_004540_0070596	ERAP1-like_C-terminal_domain
100%	AS01_005776_0235745	ERGosterol_biosynthesis_methyltransferase_(ERG6)_family
2%	AS01_009535_0109677	Eukaryotic_porin/Tom40
24%	AS01_016848_0036587	Eukaryotic_porin/Tom40
94%	AS01_021783_0038169	Eukaryotic_porin/Tom40
99%	AS01_007933_0166634	Eukaryotic_rRNA_processing
5%	AS01_006083_0107058	Eukaryotic_translation_initiation_factor_3_subunit_A
25%	AS01_008289_0057675	Eukaryotic_translation_initiation_factor_3_subunit_A
20%	AS01_008103_0127443	Eukaryotic_translation_initiation_factor_3_subunit_J
100%	AS01_012236_0131022	Exonuclease
100%	AS01_001954_0102318	Exostosin-like
100%	AS01_004288_0131207	Exostosin-like
100%	AS01_004288_0134775	Exostosin-like
100%	AS01_007961_0219700	Exostosin-like

100%	AS01_008690_0377742	Exostosin-like
100%	AS01_008777_0190352	Exostosin-like
100%	AS01_009996_0281745	Exostosin-like
100%	AS01_010271_0149316	Exostosin-like
100%	AS01_012504_0070924	Exostosin-like
15%	AS01_013617_0068074	Exostosin-like
100%	AS01_015347_0101649	Exostosin-like
44%	AS01_004897_0013256	F-box_domain
1%	AS01_019290_0041719	F-box_domain
100%	AS01_004143_0195342	FAD-dependent_pyridine_nucleotide-disulphide_oxidoreductase
29%	AS01_002168_0527521	FAD-linked_oxidoreductase-like
100%	AS01_007473_0078919	FAS1_domain
97%	AS01_006791_0110393	Fatty_acid_desaturase,_type_1
100%	AS01_012232_0059895	Fatty_acid_desaturase,_type_1
100%	AS01_009832_0045460	Fatty_acid_desaturase,_type_2
100%	AS01_001861_0086592	Fatty_acyl-CoA_reductase
73%	AS01_004660_0160413	Ferredoxin_reductase-type_FAD-binding_domain
3%	AS01_005349_0043853	Ferredoxin_reductase-type_FAD-binding_domain
2%	AS01_006982_0007598	Ferredoxin_reductase-type_FAD-binding_domain
100%	AS01_015341_0144766	Fibrillarin
100%	AS01_012548_0126536	Fibronectin_type_III-like_domain
1%	AS01_008081_0045336	Flavin_monooxygenase-like
21%	AS01_001180_0095021	G-protein_beta_WD-40_repeat
100%	AS01_001451_0200044	G-protein_beta_WD-40_repeat
100%	AS01_003370_0062201	G-protein_beta_WD-40_repeat
99%	AS01_005313_0495838	G-protein_beta_WD-40_repeat
100%	AS01_006400_0105185	G-protein_beta_WD-40_repeat
2%	AS01_007542_0220406	G-protein_beta_WD-40_repeat
96%	AS01_008158_0053006	G-protein_beta_WD-40_repeat
20%	AS01_011663_0185273	G-protein_beta_WD-40_repeat
98%	AS01_006447_0648447	GAF_domain
82%	AS01_002074_0146806	Galactose-binding_domain-like
100%	AS01_002136_0037965	Galactose-binding_domain-like
100%	AS01_002136_0071459	Galactose-binding_domain-like
100%	AS01_004346_0056065	Galactose-binding_domain-like
100%	AS01_007675_0135149	Galactose-binding_domain-like
100%	AS01_008475_0124453	Galactose-binding_domain-like
100%	AS01_010766_0013394	Galactose-binding_domain-like
93%	AS01_007263_0478096	Galactosyl_transferase
63%	AS01_008595_0056264	Galactosyl_transferase
100%	AS01_009113_0005975	Galactosyl_transferase
100%	AS01_013604_0142055	Galactosyl_transferase
100%	AS01_013604_0145899	Galactosyl_transferase
100%	AS01_018174_0038340	GDP-mannose_4,6-dehydratase
20%	AS01_003681_0277226	General_substrate_transporter
20%	AS01_004115_0168885	General_substrate_transporter
100%	AS01_004869_0065887	General_substrate_transporter
100%	AS01_007531_0132795	General_substrate_transporter
10%	AS01_017716_0125323	General_substrate_transporter
100%	AS01_006158_0224429	GH3_auxin-responsive_promoter
100%	AS01_000449_0014361	Glucose-6-phosphate_dehydrogenase
20%	AS01_000153_0468965	Glucose/ribitol_dehydrogenase
99%	AS01_000317_0619732	Glucose/ribitol_dehydrogenase
100%	AS01_010382_0180664	Glucose/ribitol_dehydrogenase
100%	AS01_011849_0161352	Glucose/ribitol_dehydrogenase
100%	AS01_000996_0014962	Glucosidase_II_beta_subunit,_N-terminal
87%	AS01_004816_0096945	Glutaredoxin
100%	AS01_000246_0663325	Glutathione_S-transferase,_C-terminal
93%	AS01_000852_0298651	Glutathione_S-transferase,_C-terminal
100%	AS01_000852_0326867	Glutathione_S-transferase,_C-terminal

100%	AS01_004111_0018952	Glutathione_S-transferase,_C-terminal
100%	AS01_004232_0103576	Glutathione_S-transferase,_C-terminal
100%	AS01_004232_0113686	Glutathione_S-transferase,_C-terminal
100%	AS01_004234_0009923	Glutathione_S-transferase,_C-terminal
100%	AS01_007221_0030169	Glutathione_S-transferase,_C-terminal
19%	AS01_008951_0118887	Glutathione_S-transferase,_C-terminal
100%	AS01_009750_0019856	Glutathione_S-transferase,_C-terminal
100%	AS01_009750_0028109	Glutathione_S-transferase,_C-terminal
100%	AS01_011025_0006494	Glutathione_S-transferase,_C-terminal
20%	AS01_016610_0037082	Glutathione_S-transferase,_C-terminal
100%	AS01_001509_0589721	Glutathione_S-transferase,_C-terminal-like
100%	AS01_003144_0075826	Glutathione_S-transferase,_C-terminal-like
100%	AS01_003816_0014282	Glutathione_S-transferase,_C-terminal-like
100%	AS01_005061_0145903	Glutathione_S-transferase,_C-terminal-like
20%	AS01_008424_0351035	Glutathione-dependent_formaldehyde-activating_enzyme/centromere_protein_V
97%	AS01_008424_0371130	Glutathione-dependent_formaldehyde-activating_enzyme/centromere_protein_V
2%	AS01_012005_0277191	Glyceraldehyde_3-phosphate_dehydrogenase,_NAD(P)_binding_domain
3%	AS01_016526_0010952	Glyceraldehyde_3-phosphate_dehydrogenase,_NAD(P)_binding_domain
100%	AS01_003965_0174447	Glycoside_hydrolase_family_3_C-terminal_domain
100%	AS01_001496_0332320	Glycoside_hydrolase,_catalytic_domain
100%	AS01_004977_0031218	Glycoside_hydrolase,_catalytic_domain
100%	AS01_006890_0027185	Glycoside_hydrolase,_catalytic_domain
100%	AS01_007960_0078364	Glycoside_hydrolase,_catalytic_domain
100%	AS01_013300_0024526	Glycoside_hydrolase,_catalytic_domain
100%	AS01_008721_0244631	Glycoside_hydrolase,_family_19
100%	AS01_017973_0045706	Glycoside_hydrolase,_family_85
100%	AS01_002301_0135172	Glycoside_hydrolase,_family_9
100%	AS01_004204_0366133	Glycosyl_transferase,_family_1
100%	AS01_001148_0268953	Glycosyl_transferase,_family_14
100%	AS01_008807_0113901	Glycosyl_transferase,_family_14
3%	AS01_003683_0215223	Glycosyl_transferase,_family_29
100%	AS01_000791_0186845	Glycosyl_transferase,_family_8
100%	AS01_002456_0097315	Glycosyltransferase_AER61,_uncharacterised
100%	AS01_003401_0160351	Glycosyltransferase_AER61,_uncharacterised
100%	AS01_004886_0273678	Glycosyltransferase_AER61,_uncharacterised
100%	AS01_004886_0315146	Glycosyltransferase_AER61,_uncharacterised
100%	AS01_004886_0404415	Glycosyltransferase_AER61,_uncharacterised
100%	AS01_017111_0020903	Glycosyltransferase_AER61,_uncharacterised
100%	AS01_006939_0072579	GNS1/SUR4_membrane_protein
100%	AS01_001531_0173402	GRAM_domain
97%	AS01_001483_0720148	GTP_binding_domain
7%	AS01_001962_0055933	GTP_binding_domain
100%	AS01_002555_0065221	GTP_cyclohydrolase_I_domain
92%	AS01_000390_0043113	H/ACA_ribonucleoprotein_complex,_subunit_Gar1/Naf1
100%	AS01_001829_0643521	HAD-like_domain
100%	AS01_001465_0439824	Haem_peroxidase
100%	AS01_002992_0064727	Haem_peroxidase
96%	AS01_002992_0069767	Haem_peroxidase
20%	AS01_002992_0126490	Haem_peroxidase
100%	AS01_003719_0381526	Haem_peroxidase
100%	AS01_004198_0110302	Haem_peroxidase
100%	AS01_004267_0150687	Haem_peroxidase
100%	AS01_005142_0063396	Haem_peroxidase
20%	AS01_005609_0047680	Haem_peroxidase
100%	AS01_005773_0088323	Haem_peroxidase
3%	AS01_005943_0133361	Haem_peroxidase
100%	AS01_007334_0017056	Haem_peroxidase
100%	AS01_007334_0037390	Haem_peroxidase

100%	AS01_008021_0075360	Haem_peroxidase
20%	AS01_008334_0093582	Haem_peroxidase
20%	AS01_008875_0315884	Haem_peroxidase
63%	AS01_009197_0173397	Haem_peroxidase
100%	AS01_011367_0234451	Haem_peroxidase
100%	AS01_014319_0116183	Haem_peroxidase
6%	AS01_015227_0105780	Haem_peroxidase
65%	AS01_006085_0219572	Harbinger_transposase-derived_protein,_plant
4%	AS01_001040_0547581	Heat_shock_chaperonin-binding
26%	AS01_014850_0032402	Heat_shock_protein_70_family
100%	AS01_004946_0074264	Heat_shock_protein_Hsp90_family
100%	AS01_014454_0107645	Heavy_metal-associated_domain,_HMA
100%	AS01_014361_0144048	Helix-turn-helix_motif
8%	AS01_009999_0138652	High_mobility_group_box_domain
100%	AS01_002689_0217237	Histidine_kinase-like_ATPase,_C-terminal_domain
68%	AS01_009664_0065806	Histidine_phosphatase_superfamily
8%	AS01_000921_0484840	Histone_H2A
13%	AS01_002393_0035422	Histone_H2A
100%	AS01_007934_0087237	Histone_H2A
2%	AS01_011538_0029980	Histone_H2A
100%	AS01_011538_0032585	Histone_H2A
5%	AS01_014043_0089437	Histone_H2A
29%	AS01_004826_0324488	Histone_H2B
100%	AS01_006468_0093753	Histone_H3/CENP-A
100%	AS01_010737_0014594	Histone_H4
18%	AS01_006050_0138630	HnRNP-L/PTB/hephaestus_splicing_factor
20%	AS01_001766_0204701	Homeobox_domain
74%	AS01_000191_0172634	Homeodomain-like
99%	AS01_001246_0392399	Homeodomain-like
50%	AS01_004126_0288466	Homeodomain-like
100%	AS01_012985_0035331	Homeodomain-like
100%	AS01_012533_0043369	Homocysteine_S-methyltransferase
100%	AS01_003993_0233129	HR-like_lesion-inducer
100%	AS01_004040_0593368	Hyaluronan/mRNA-binding_protein
100%	AS01_009016_0264253	Hydroxymethylglutaryl-CoA_reductase,_N-terminal
100%	AS01_009779_0374650	Hydroxymethylglutaryl-CoA_synthase,_eukaryotic
99%	AS01_004403_0209072	Immunoglobulin_E-set
3%	AS01_009159_0129253	Initiation_factor_2B_alpha/beta/delta
2%	AS01_012748_0250004	Inosine/uridine-preferring_nucleoside_hydrolase_domain
18%	AS01_010591_0047172	Isocitrate_dehydrogenase_NAD-dependent
89%	AS01_002468_0177196	Isopenicillin_N_synthase
10%	AS01_006278_0083961	Isopenicillin_N_synthase
100%	AS01_009472_0036123	Isopenicillin_N_synthase
60%	AS01_000317_0606615	Isopenicillin_N_synthase-like
78%	AS01_005401_0149179	Isopenicillin_N_synthase-like
100%	AS01_007221_0028088	Isopenicillin_N_synthase-like
100%	AS01_010092_0153600	Isopenicillin_N_synthase-like
100%	AS01_013667_0258499	Isopenicillin_N_synthase-like
20%	AS01_441659_0000201	Isopenicillin_N_synthase-like
100%	AS01_004436_0156754	Isopentenyl-diphosphate_delta-isomerase,_type_1
100%	AS01_007068_0018876	JmjC_domain
92%	AS01_008158_0184051	KOW
98%	AS01_022893_0005593	KOW
57%	AS01_001289_0011748	KRR1_interacting_protein_1
20%	AS01_017973_0025070	Lateral_organ_boundaries,_LOB
100%	AS01_009976_0463564	Leucine_rich_repeat_4
100%	AS01_010849_0017611	Leucine_rich_repeat_4
100%	AS01_013329_0302123	Leucine_rich_repeat_4
60%	AS01_001671_0010374	Leucine-rich_repeat
86%	AS01_004027_0169510	Leucine-rich_repeat

3%	AS01_006012_0181851	Leucine-rich_repeat
100%	AS01_006364_0022893	Leucine-rich_repeat
30%	AS01_006614_0337775	Leucine-rich_repeat
22%	AS01_009058_0451457	Leucine-rich_repeat
100%	AS01_012554_0177711	Leucine-rich_repeat
1%	AS01_004585_1050737	Like-Sm_(LSM)_domain
27%	AS01_007359_0085125	Like-Sm_(LSM)_domain
20%	AS01_000303_0304288	Lipase,_GDSL
100%	AS01_001488_0374005	Lipase,_GDSL
100%	AS01_002101_0756857	Lipase,_GDSL
100%	AS01_002168_0505582	Lipase,_GDSL
100%	AS01_005364_0006604	Lipase,_GDSL
100%	AS01_005399_0014172	Lipase,_GDSL
100%	AS01_005956_0368959	Lipase,_GDSL
100%	AS01_010665_0009225	Lipase,_GDSL
25%	AS01_020432_0047816	Lipase,_GDSL
100%	AS01_002988_0532963	Lipopolysaccharide-modifying_protein
84%	AS01_001929_0053765	LisH_dimerisation_motif
47%	AS01_004126_0578254	Longin-like_domain
1%	AS01_002301_0298318	Lupus_La_protein
72%	AS01_003981_0033480	Magnesium_transporter_NIPA
100%	AS01_008158_0185780	Magnesium_transporter_NIPA
100%	AS01_000648_0394667	Major_facilitator_superfamily
5%	AS01_005097_0208155	Major_facilitator_superfamily
100%	AS01_003741_0178550	Major_facilitator_superfamily_domain,_general_substrate_transporter
88%	AS01_004403_0011598	Major_facilitator_superfamily_domain,_general_substrate_transporter
1%	AS01_007232_0113041	Major_facilitator_superfamily_domain,_general_substrate_transporter
100%	AS01_009154_0097911	Major_facilitator_superfamily_domain,_general_substrate_transporter
21%	AS01_009398_0150337	Major_facilitator_superfamily_domain,_general_substrate_transporter
100%	AS01_010757_0070432	Major_facilitator_superfamily_domain,_general_substrate_transporter
100%	AS01_002712_0210677	Mannose-6-phosphate_isomerase
100%	AS01_002712_0228917	Mannose-6-phosphate_isomerase
100%	AS01_004210_0022121	Mannose-binding_lectin
100%	AS01_004210_0071474	Mannose-binding_lectin
100%	AS01_004522_0917654	Mannose-binding_lectin
100%	AS01_007204_0072139	Mannose-binding_lectin
100%	AS01_008367_0126202	Mannose-binding_lectin
100%	AS01_008367_0225057	Mannose-binding_lectin
100%	AS01_008367_0280245	Mannose-binding_lectin
100%	AS01_010097_0137152	Mannose-binding_lectin
100%	AS01_011412_0025680	Mannose-binding_lectin
100%	AS01_011981_0083458	Mannose-binding_lectin
100%	AS01_018058_0027362	Mannose-binding_lectin
100%	AS01_010050_0237502	Mediator_complex,_subunit_Med19,_metazoa
1%	AS01_006541_0401630	Metalloenzyme,_LuxS/M16_peptidase-like
2%	AS01_002930_0337681	Metallopeptidase,_catalytic_domain
98%	AS01_014143_0054295	Metallopeptidase,_catalytic_domain
98%	AS01_003400_0075256	Methyltransferase_type_11
100%	AS01_007628_0061383	Methyltransferase_type_12
55%	AS01_001674_1372961	Mitochondrial_carrier_domain
88%	AS01_005545_0049816	Mitochondrial_carrier_domain
99%	AS01_007800_0378398	Mitogen-activated_protein_(MAP)_kinase,_conserved_site
100%	AS01_004706_0070726	Mpv17/PMP22
100%	AS01_000754_0116907	mRNA_splicing_factor,_Cwf18
100%	AS01_002007_0259435	mRNA_splicing_factor,_Cwf21
100%	AS01_001605_0165902	Multi_antimicrobial_extrusion_protein
100%	AS01_004403_0090433	Multi_antimicrobial_extrusion_protein
100%	AS01_005097_0197866	Multi_antimicrobial_extrusion_protein
7%	AS01_010405_0034216	Multi_antimicrobial_extrusion_protein
100%	AS01_012751_0131129	Multi_antimicrobial_extrusion_protein

4%	AS01_008086_0004031	Myb-like_domain
100%	AS01_000037_0019810	Myc-type,_basic_helix-loop-helix_(bHLH)_domain
100%	AS01_000882_0382230	Myc-type,_basic_helix-loop-helix_(bHLH)_domain
100%	AS01_005176_0059912	Myc-type,_basic_helix-loop-helix_(bHLH)_domain
100%	AS01_006181_0012120	Myc-type,_basic_helix-loop-helix_(bHLH)_domain
100%	AS01_006196_0238284	Myc-type,_basic_helix-loop-helix_(bHLH)_domain
100%	AS01_006196_0565524	Myc-type,_basic_helix-loop-helix_(bHLH)_domain
100%	AS01_006203_0093271	Myc-type,_basic_helix-loop-helix_(bHLH)_domain
100%	AS01_006516_0110236	Myc-type,_basic_helix-loop-helix_(bHLH)_domain
20%	AS01_006813_0102035	Myc-type,_basic_helix-loop-helix_(bHLH)_domain
12%	AS01_007821_0248906	Myc-type,_basic_helix-loop-helix_(bHLH)_domain
100%	AS01_011352_0041474	Myc-type,_basic_helix-loop-helix_(bHLH)_domain
100%	AS01_013114_0143392	Myc-type,_basic_helix-loop-helix_(bHLH)_domain
100%	AS01_012605_0053403	Mycolic_acid_cyclopropane_synthase
100%	AS01_003492_0325324	NAC_domain
100%	AS01_000696_0155012	NAD(P)-binding_domain
36%	AS01_000772_0070982	NAD(P)-binding_domain
100%	AS01_001472_0640352	NAD(P)-binding_domain
100%	AS01_002270_0024303	NAD(P)-binding_domain
100%	AS01_004690_0356404	NAD(P)-binding_domain
81%	AS01_007344_0020694	NAD(P)-binding_domain
100%	AS01_009046_0051532	NAD(P)-binding_domain
100%	AS01_009438_0043314	NAD(P)-binding_domain
100%	AS01_011836_0053149	NAD(P)-binding_domain
100%	AS01_011836_0122321	NAD(P)-binding_domain
100%	AS01_007069_0037999	NADP-dependent_oxidoreductase_domain
100%	AS01_439711_0000193	NADP-dependent_oxidoreductase_domain
20%	AS01_000803_0334541	NB-ARC
100%	AS01_002297_0075165	NB-ARC
20%	AS01_002477_0064603	NB-ARC
20%	AS01_004994_0170890	NB-ARC
20%	AS01_006178_0024248	NB-ARC
100%	AS01_009712_0272392	NB-ARC
100%	AS01_009808_0037741	NB-ARC
100%	AS01_010602_0092157	NB-ARC
20%	AS01_016212_0194698	NB-ARC
80%	AS01_003267_0275399	Nicotinamide_N-methyltransferase-like
57%	AS01_018545_0025505	NMD3
100%	AS01_000833_0262202	NOP5,_N-terminal
100%	AS01_004747_0534007	NOP5,_N-terminal
100%	AS01_005353_0118830	NOP5,_N-terminal
70%	AS01_000733_0424885	NUC153
99%	AS01_003670_0038151	Nucleic_acid-binding,_OB-fold
9%	AS01_013831_0043039	Nucleic_acid-binding,_OB-fold
29%	AS01_010225_0009338	Nucleolar_complex_protein_2
4%	AS01_001380_0087811	Nucleolar_protein_12
100%	AS01_005151_0164603	Nucleolar_protein_14
93%	AS01_002948_0133086	Nucleolar,_Nop52
97%	AS01_001100_0856467	Nucleophile_aminohydrolases,_N-terminal
18%	AS01_012236_0074012	Nucleophile_aminohydrolases,_N-terminal
100%	AS01_003430_0536468	Nucleoplasmin_core_domain
11%	AS01_000049_0013388	Nucleotide-binding,_alpha-beta_plait
100%	AS01_000733_0378110	Nucleotide-binding,_alpha-beta_plait
20%	AS01_003448_0059775	Nucleotide-binding,_alpha-beta_plait
81%	AS01_003514_0635294	Nucleotide-binding,_alpha-beta_plait
100%	AS01_003552_0062058	Nucleotide-binding,_alpha-beta_plait
100%	AS01_005656_0112369	Nucleotide-binding,_alpha-beta_plait
100%	AS01_000058_0193081	Nucleotide-diphospho-sugar_transferase
100%	AS01_000058_0195819	Nucleotide-diphospho-sugar_transferase
100%	AS01_006449_0086955	Nucleotide-diphospho-sugar_transferase

100%	AS01_010666_0035487	Nucleotide-diphospho-sugar_transferase
23%	AS01_016090_0086659	Nucleotide-diphospho-sugar_transferase
94%	AS01_001431_0028359	Nucleotide-diphospho-sugar_transferases
100%	AS01_002043_0666416	Nucleotide-diphospho-sugar_transferases
100%	AS01_009449_0015969	Nucleotide-diphospho-sugar_transferases
100%	AS01_020482_0001913	Nucleotide-diphospho-sugar_transferases
2%	AS01_005934_0178912	Oligosaccharyl_transferase_complex,_subunit_OST3/OST6
2%	AS01_012413_0013846	Origin_recognition_complex,_subunit_6
4%	AS01_012652_0164617	Ovarian_tumour,_otubain
100%	AS01_000335_0098999	P-loop_containing_nucleoside_triphosphate_hydrolase
100%	AS01_003166_0415328	P-loop_containing_nucleoside_triphosphate_hydrolase
100%	AS01_005244_0205607	P-loop_containing_nucleoside_triphosphate_hydrolase
93%	AS01_006811_0253271	P-loop_containing_nucleoside_triphosphate_hydrolase
20%	AS01_007161_0089843	P-loop_containing_nucleoside_triphosphate_hydrolase
100%	AS01_007648_0046953	P-loop_containing_nucleoside_triphosphate_hydrolase
1%	AS01_008968_0421328	P-loop_containing_nucleoside_triphosphate_hydrolase
100%	AS01_010125_0204755	P-loop_containing_nucleoside_triphosphate_hydrolase
88%	AS01_011177_0123989	P-loop_containing_nucleoside_triphosphate_hydrolase
100%	AS01_011177_0155579	P-loop_containing_nucleoside_triphosphate_hydrolase
100%	AS01_013667_0262007	P-loop_containing_nucleoside_triphosphate_hydrolase
8%	AS01_013986_0049428	P-loop_containing_nucleoside_triphosphate_hydrolase
94%	AS01_012748_0273344	PapD-like
100%	AS01_000336_0041187	PC-Esterase
100%	AS01_001725_0971424	PC-Esterase
100%	AS01_006512_0211342	PC-Esterase
20%	AS01_007526_0097216	PC-Esterase
3%	AS01_012341_0035232	PC-Esterase
99%	AS01_014447_0000235	PC-Esterase
99%	AS01_002992_0409168	PDZ_domain
99%	AS01_003576_0641686	Pentatricopeptide_repeat
100%	AS01_003704_0013319	Pentatricopeptide_repeat
3%	AS01_008141_0011564	Pentatricopeptide_repeat
100%	AS01_015165_0081942	Pentatricopeptide_repeat
100%	AS01_010593_0082031	Peptidase_A22B,_signal_peptide_peptidase
22%	AS01_001291_0023190	Peptidase_C13,_legumain
23%	AS01_009230_0100785	Peptidase_C48,_SUMO/Sentrin/Ubl1
6%	AS01_001208_0230685	Peptidase_M24,_methionine_aminopeptidase
78%	AS01_000949_0132525	Peptidase_S54,_rhomboid_domain
2%	AS01_010125_0224604	Peptidase_S54,_rhomboid_domain
100%	AS01_001588_0431554	Peptidase_S8,_subtilisin,_Ser-active_site
100%	AS01_007362_0194990	Peptidase_S8,_subtilisin,_Ser-active_site
100%	AS01_007362_0244924	Peptidase_S8,_subtilisin,_Ser-active_site
100%	AS01_017111_0016296	Peptidase_S9A,_N-terminal_domain
100%	AS01_002036_0968308	Peptidyl-prolyl_cis-trans_isomerase,_FKBP-type,_domain
100%	AS01_007283_0288885	Permease,_cytosine/purines,_uracil,_thiamine,_allantoin
4%	AS01_004972_0434444	Phosducin,_thioredoxin-like_domain
100%	AS01_013122_0056934	Phosphate-induced_protein_1
20%	AS01_001982_0167923	Phospholipase_C/P1_nuclease_domain
100%	AS01_006675_0074170	Plant_disease_resistance_response_protein
100%	AS01_017985_0000375	Plant_disease_resistance_response_protein
100%	AS01_017985_0006786	Plant_disease_resistance_response_protein
100%	AS01_018076_0022700	Plant_disease_resistance_response_protein
73%	AS01_009240_0072052	Plant_organelle_RNA_recognition_domain
78%	AS01_011337_0038674	Plant_organelle_RNA_recognition_domain
36%	AS01_005074_0206970	PLC-like_phosphodiesterase,_TIM_beta/alpha-barrel_domain
100%	AS01_000712_0074399	Pollen_Ole_e_1_allergen/extensin
100%	AS01_000712_0163812	Pollen_Ole_e_1_allergen/extensin
100%	AS01_001608_0217978	Pollen_Ole_e_1_allergen/extensin
100%	AS01_013835_0442961	Pollen_Ole_e_1_allergen/extensin
3%	AS01_009594_0081800	PPPDE_putative_peptidase_domain

100%	AS01_003827_0085775	Prenyltransferase/squalene_oxidase
4%	AS01_008595_0090478	Programmed_cell_death_protein_2,_C-terminal
100%	AS01_009666_0126202	Proteasome_assembly_chaperone_2
100%	AS01_008296_0049391	Proteasome_maturation_factor_UMP1
100%	AS01_001450_0239020	Protein_arginine_N-methyltransferase
100%	AS01_000726_0160166	Protein_kinase_domain
37%	AS01_001449_0062878	Protein_kinase_domain
20%	AS01_001724_0171163	Protein_kinase_domain
100%	AS01_002046_0019779	Protein_kinase_domain
100%	AS01_003488_0073064	Protein_kinase_domain
76%	AS01_005045_0249477	Protein_kinase_domain
100%	AS01_008158_0041861	Protein_kinase_domain
21%	AS01_009659_0161521	Protein_kinase_domain
4%	AS01_011198_0070410	Protein_kinase_domain
18%	AS01_002318_0093294	Protein_kinase-like_domain
1%	AS01_017479_0024144	Protein_kinase-like_domain
100%	AS01_009301_0070815	Protein_N-terminal_glutamine_amidohydrolase,_alpha_beta_roll
84%	AS01_016996_0014533	Protein_notum_homologue
68%	AS01_010205_0035464	Protein_of_unknown_function_DUF106,_transmembrane
100%	AS01_001957_0026708	Protein_of_unknown_function_DUF1218
100%	AS01_001231_0198278	Protein_of_unknown_function_DUF241,_plant
100%	AS01_001231_0201927	Protein_of_unknown_function_DUF241,_plant
100%	AS01_001231_0299663	Protein_of_unknown_function_DUF241,_plant
100%	AS01_006424_0326128	Protein_of_unknown_function_DUF241,_plant
100%	AS01_006424_0337742	Protein_of_unknown_function_DUF241,_plant
100%	AS01_006424_0339619	Protein_of_unknown_function_DUF241,_plant
100%	AS01_013787_0024598	Protein_of_unknown_function_DUF241,_plant
100%	AS01_014026_0011580	Protein_of_unknown_function_DUF241,_plant
20%	AS01_014026_0079048	Protein_of_unknown_function_DUF241,_plant
100%	AS01_015456_0038292	Protein_of_unknown_function_DUF241,_plant
100%	AS01_017011_0046369	Protein_of_unknown_function_DUF241,_plant
100%	AS01_017011_0056397	Protein_of_unknown_function_DUF241,_plant
100%	AS01_020007_0016616	Protein_of_unknown_function_DUF241,_plant
100%	AS01_000431_0358822	Protein_of_unknown_function_DUF2921
20%	AS01_012409_0051098	Protein_of_unknown_function_DUF2921
100%	AS01_013559_0034071	Protein_of_unknown_function_DUF2921
100%	AS01_001048_0286066	Protein_of_unknown_function_DUF3339
100%	AS01_009058_0369955	Protein_of_unknown_function_DUF4228,_plant
100%	AS01_012177_0150096	Protein_of_unknown_function_DUF4598
100%	AS01_019523_0022441	Protein_of_unknown_function_DUF4602
100%	AS01_000563_0771215	Protein_of_unknown_function_DUF538
100%	AS01_010759_0295032	Protein_of_unknown_function_DUF597
100%	AS01_000818_0378760	Protein_of_unknown_function_DUF620
100%	AS01_015781_0005107	Protein_of_unknown_function_DUF707
100%	AS01_003508_0178621	Protein_of_unknown_function_DUF716_(TMEM45)
27%	AS01_000906_1356458	Protein_of_unknown_function_DUF761,_plant
100%	AS01_003454_0444520	Protein_of_unknown_function_DUF761,_plant
85%	AS01_011412_0049760	Protein_of_unknown_function_DUF819
97%	AS01_001312_0034198	Protein_transport_protein_SecG/Sec61-beta/Sbh1
100%	AS01_002712_0393011	Proteinase_inhibitor_I12,_Bowman-Birk
100%	AS01_003131_0184404	Proteinase_inhibitor_I12,_Bowman-Birk
100%	AS01_008437_0026193	Proteinase_inhibitor_I12,_Bowman-Birk
3%	AS01_002036_0697460	Pseudouridine_synthase,_RsuA/RluB/C/D/E/F
21%	AS01_005306_0156600	PUB_domain
100%	AS01_009843_0066074	Putative_methionine_gamma-lyase
100%	AS01_006257_0198590	Putative_polysaccharide_biosynthesis_protein
100%	AS01_443003_0000273	Putative_S-adenosyl-L-methionine-dependent_methyltransferase_MidA
100%	AS01_000996_0097696	Pyruvate_kinase
100%	AS01_004240_0039452	Rab-GTPase-TBC_domain
100%	AS01_000433_0224146	Reverse_transcriptase_zinc-binding_domain

100%	AS01_010249_0102300	Rhodanese-like_domain
43%	AS01_011139_0020001	Ribonuclease_H-like_domain
100%	AS01_000541_0306779	Ribonuclease_T2,_active_site
100%	AS01_002299_0175920	Ribosomal_protein_L1,_2-layer_alpha/beta-sandwich
100%	AS01_007948_0368383	Ribosomal_protein_L1,_2-layer_alpha/beta-sandwich
100%	AS01_008994_0076275	Ribosomal_protein_L10/L12
63%	AS01_000499_0478160	Ribosomal_protein_L10e
59%	AS01_009594_0261512	Ribosomal_protein_L14_conserved_site
14%	AS01_020753_0014542	Ribosomal_protein_L15e
100%	AS01_008635_0171473	Ribosomal_protein_L18/L5
100%	AS01_001416_0020527	Ribosomal_protein_L18e,_conserved_site
1%	AS01_005425_0448345	Ribosomal_protein_L19/L19e
43%	AS01_005703_0036613	Ribosomal_protein_L21e
1%	AS01_002857_0465165	Ribosomal_protein_L22/L17
34%	AS01_003433_0368198	Ribosomal_protein_L25
87%	AS01_010470_0038069	Ribosomal_protein_L29
3%	AS01_013131_0119040	Ribosomal_protein_L3
4%	AS01_027550_0002230	Ribosomal_protein_L3
100%	AS01_008783_0173936	Ribosomal_protein_L30,_N-terminal
13%	AS01_008602_0126403	Ribosomal_protein_L34Ae
13%	AS01_021816_0010771	Ribosomal_protein_L36e
2%	AS01_009594_0324230	Ribosomal_protein_L37ae/L37e
6%	AS01_003094_0424552	Ribosomal_protein_L5
100%	AS01_008085_0099188	Ribosomal_protein_L9,_C-terminal
56%	AS01_001435_0064900	Ribosomal_protein_S13/S15,_N-terminal
4%	AS01_006194_0086850	Ribosomal_protein_S13/S15,_N-terminal
100%	AS01_002821_0044013	Ribosomal_protein_S2
5%	AS01_007508_0342128	Ribosomal_protein_S26e
18%	AS01_010244_0142823	Ribosomal_protein_S5_domain_2-type_fold
100%	AS01_001270_0225433	Ribosomal_protein_S5/S7,_eukaryotic/archaeal
47%	AS01_014576_0037858	Ribosomal_protein_S8e
100%	AS01_005489_0066019	Ribosomal_RNA-processing_protein_14,_N-terminal
100%	AS01_012341_0143309	Ribosomal_RNA-processing_protein_7
10%	AS01_007997_0120921	Ribosomal_S11,_conserved_site
100%	AS01_000825_0129716	Ribosome_recycling_factor_domain
100%	AS01_015527_0052069	Ribosome-inactivating_protein
5%	AS01_001002_0127386	RNA-binding_protein_Lupus_La
100%	AS01_000690_0048783	RNA-binding_S4_domain
18%	AS01_005674_0214058	RNA-binding_S4_domain
100%	AS01_003347_0399659	Root_cap
100%	AS01_005928_0097668	Root_cap
100%	AS01_006541_0032896	Root_cap
100%	AS01_008717_0030093	Root_cap
100%	AS01_008717_0035391	Root_cap
46%	AS01_010213_0142088	RWD_domain
57%	AS01_002084_0411265	S-adenosyl-L-methionine-dependent_methyltransferase-like
100%	AS01_004948_0015698	S-adenosyl-L-methionine-dependent_methyltransferase-like
100%	AS01_013128_0066475	S-adenosyl-L-methionine-dependent_methyltransferase-like
47%	AS01_011771_0071450	Sas10_C-terminal_domain
87%	AS01_000997_0198251	SEC-C_motif
100%	AS01_003721_0006110	SGNH_hydrolase-type_esterase_domain
52%	AS01_011663_0095892	Signal_transduction_histidine_kinase,_phosphotransfer_(Hpt)_domain
100%	AS01_005097_0171762	Six-bladed_beta-propeller,_TolB-like
100%	AS01_013267_0148644	SMAD/FHA_domain
20%	AS01_003346_0634538	Small-subunit_processome,_Utp12
17%	AS01_004349_0506917	Sodium/calcium_exchanger_membrane_region
32%	AS01_001394_0276109	Solute_carrier_family_35_member_SLC35F1/F2/F6
17%	AS01_000549_0109556	Steadiness_box
100%	AS01_006328_0036868	Stigma-specific_protein_Stig1
20%	AS01_006537_0047479	Target_SNARE_coiled-coil_domain

100%	AS01_011433_0174469	Terpene_synthase,_N-terminal_domain
39%	AS01_014701_0060587	Tetraspanin
100%	AS01_001116_0355431	Tetratricopeptide_repeat
20%	AS01_005602_0054176	Tetratricopeptide_repeat-containing_domain
100%	AS01_001793_0645337	Tetratricopeptide-like_helical_domain
100%	AS01_000855_0406535	Thaumatin
86%	AS01_010023_0074080	Thiamin_diphosphate-binding_fold
100%	AS01_002101_0199082	Thionin
26%	AS01_000340_0877618	THUMP
97%	AS01_008212_0028794	Tim10/DDP_family_zinc_finger
99%	AS01_000058_0255493	Transcription_factor_E2F/dimerisation_partner_(TDP)
100%	AS01_010996_0083087	Transcription_factor_MYC/MYB_N-terminal
100%	AS01_005791_0017162	Transcription_factor_TCP_subgroup
69%	AS01_005791_0082952	Transcription_factor_TCP_subgroup
3%	AS01_002058_0274905	Transcription_factor,_GATA,_plant
100%	AS01_005394_0450633	Transcription_factor,_K-box
28%	AS01_010170_0233342	Transposase,_Tc1-like
97%	AS01_000272_0020994	Triose-phosphate_transporter_domain
100%	AS01_002224_0820646	Triose-phosphate_transporter_domain
100%	AS01_003363_0188331	Triose-phosphate_transporter_domain
47%	AS01_016752_0059051	Triose-phosphate_transporter_domain
15%	AS01_003162_0139372	Tryptophan_synthase_beta_chain/beta_chain-like
81%	AS01_005173_0208498	UAA_transporter
23%	AS01_015676_0004033	Ubiquitin-conjugating_enzyme/RWD-like
100%	AS01_000040_0343206	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_000040_0580230	UDP-glucuronosyl/UDP-glucosyltransferase
88%	AS01_000207_0282310	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_000232_0183840	UDP-glucuronosyl/UDP-glucosyltransferase
39%	AS01_000525_0200999	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_000815_0222774	UDP-glucuronosyl/UDP-glucosyltransferase
20%	AS01_001240_0004470	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_001384_0023180	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_001959_0203229	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_002110_0512100	UDP-glucuronosyl/UDP-glucosyltransferase
20%	AS01_002123_0323801	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_002760_0003285	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_002774_0633711	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_002796_0048708	UDP-glucuronosyl/UDP-glucosyltransferase
20%	AS01_002796_0056524	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_003018_0175304	UDP-glucuronosyl/UDP-glucosyltransferase
94%	AS01_003251_0510695	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_003718_0016409	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_003827_0182130	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_003827_0329694	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_003827_0587827	UDP-glucuronosyl/UDP-glucosyltransferase
48%	AS01_003866_0515441	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_003944_0076156	UDP-glucuronosyl/UDP-glucosyltransferase
38%	AS01_004898_0191830	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_005173_0292209	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_006545_0031520	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_006890_0018824	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_007006_0066560	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_007207_0097636	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_008603_0020929	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_009143_0044829	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_009157_0340895	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_009625_0051477	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_009625_0058710	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_009625_0484622	UDP-glucuronosyl/UDP-glucosyltransferase
89%	AS01_009997_0231261	UDP-glucuronosyl/UDP-glucosyltransferase

100%	AS01_010493_0015704	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_011433_0183300	UDP-glucuronosyl/UDP-glucosyltransferase
4%	AS01_011449_0125658	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_012751_0126886	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_014560_0017746	UDP-glucuronosyl/UDP-glucosyltransferase
57%	AS01_015483_0023201	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_015816_0099017	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_017784_0074517	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_018433_0019113	UDP-glucuronosyl/UDP-glucosyltransferase
100%	AS01_002580_0523439	Uncharacterised_protein_family_UPF0136,_Transmembrane
100%	AS01_011059_0112822	Uncharacterised_protein_family_UPF0197
18%	AS01_012199_0063099	UPF0014_family
100%	AS01_010006_0058583	Uricase
100%	AS01_008372_0455841	UVR_domain
100%	AS01_007972_0019878	Vitamin_B6_photo-protection_and_homoeostasis
26%	AS01_004055_0772783	Voltage-dependent_anion_channel
100%	AS01_002610_0137126	Wax_synthase
100%	AS01_004244_0095864	WD40_repeat
100%	AS01_000197_0034569	WEB_family
22%	AS01_004001_0636034	WRC
35%	AS01_003214_0682270	Xyloglucan_fucosyltransferase
100%	AS01_003214_0701912	Xyloglucan_fucosyltransferase
100%	AS01_003214_0734254	Xyloglucan_fucosyltransferase
100%	AS01_003214_0743100	Xyloglucan_fucosyltransferase
100%	AS01_015781_0073597	Xyloglucan_fucosyltransferase
18%	AS01_009144_0172215	zf-FLZ_domain
87%	AS01_004078_0207470	Zinc_finger_C2H2-type/integrase_DNA-binding_domain
44%	AS01_001022_0114682	Zinc_finger,_C2H2
100%	AS01_009500_0114448	Zinc_finger,_C2H2
21%	AS01_004680_0066394	Zinc_finger,_RING-type
80%	AS01_006668_0132587	Zinc_finger,_RING-type

Table S4: Summary of *A. strigosa* genes co-expressed with the avenacin biosynthetic cluster

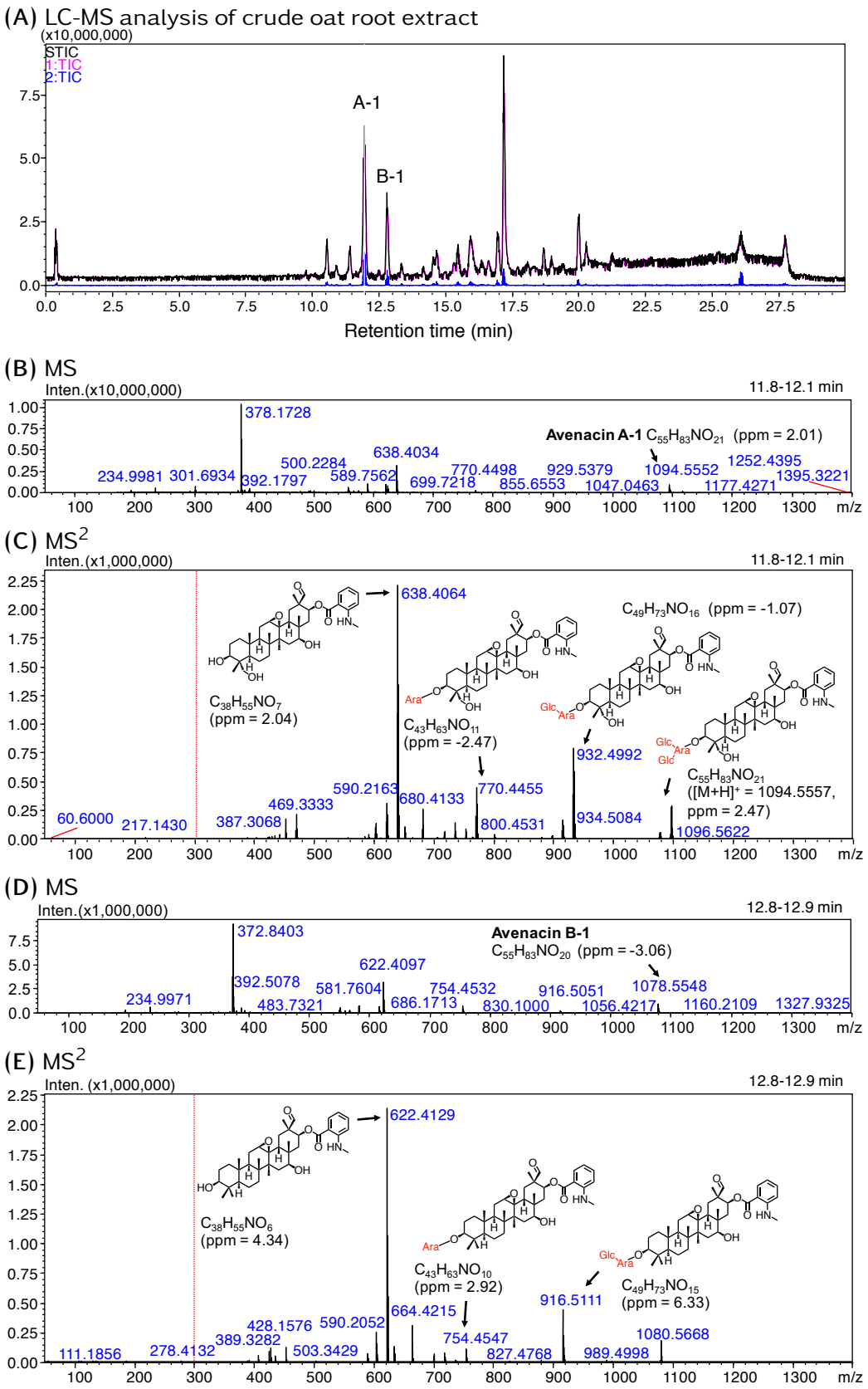


Figure S1: HPLC-IT-ToF analysis of crude oat root extract precipitate. (**A**), HPLC-IT-ToF analysis of the crude oat root methanolic extract in positive mode shows that it contained peaks corresponding to avenacin A-1 (A-1) and avenacin B-1 (B-1). The mass spectra (MS) of the peak (**B**) between 11.8-12.1 minutes and the peak (**D**) between 12.8-12.9 minutes shows signals consistent with the mass ion ($[M\text{-}H]^+$) of avenacin A-1 and avenacin B-1, respectively. The MS^2 spectrum of the fragmentation of the mass ion of (**C**) avenacin A-1 (precursor mass ion for fragmentation = *m/z* 1094.5531) and (**D**) avenacin B-1 (precursor mass ion for fragmentation = *m/z* 1078.5513 showed signals consistent with the successive loss of glucose (Glc, -162) and arabinose (Ara, -132) molecules. The difference in observed and expected mass values in parts per million is indicated (ppm).

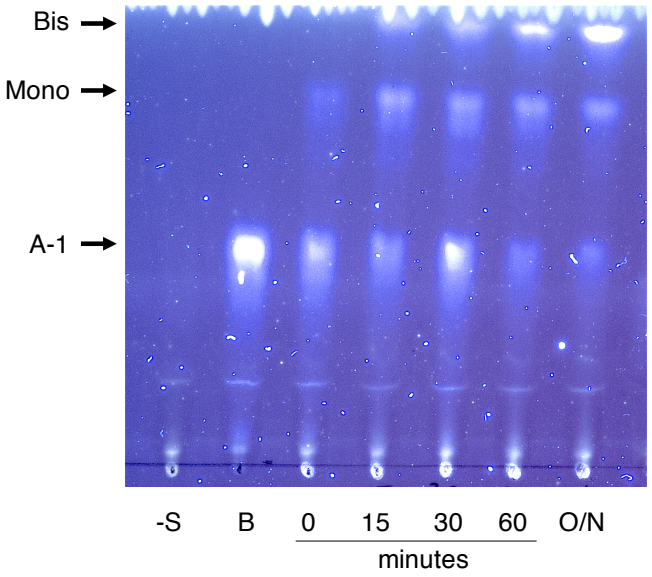


Figure S2: Timecourse experiment of extracted avenacinase on avenacin A-1 standard. -S = no substrate, B = boiled enzyme, 0 = 0 min, 15 = 15 min, 30 = 30 min, 60 = 60 min, O/N = overnight, A-1 = avenacin A-1; Bis = bis-deglucosylated avenacin A-1, Mono = mono-deglucosylated avenacin A-1. Experiment was carried out in 100 mM sodium acetate buffer pH 5 at 37°C, separated by TLC ($CHCl_3$:MeOH:H_2O, 13:6:1) and visualised under UV light.

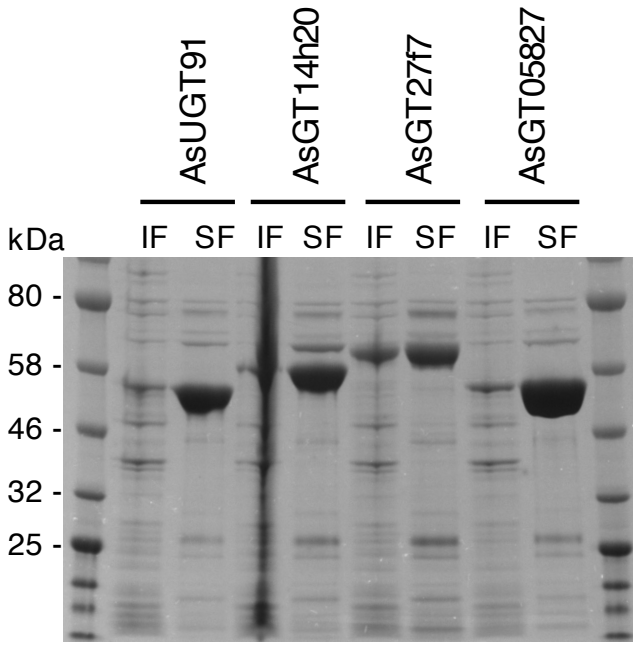


Figure S3: Analysis of insoluble and soluble fractions of *E. coli* expressing recombinant UGTs. UGTs were expressed in *E. coli* Rosetta cells as in Materials and Methods section 2.18. The insoluble fractions (IF) and soluble fractions (partially purified with Ni-NTA Agarose beads to enrich His-tagged proteins) (SF) of the cell lysates were visualised on a Coomassie-stained SDS-PAGE gel. The estimated molecular weights of the recombinant proteins are: AsUGT91 = 57.6 kDa; AsGT14h20 = 54.4 kDa; AsGT27f7 = 56.2 kDa; AsGT05827 = 51 kDa.

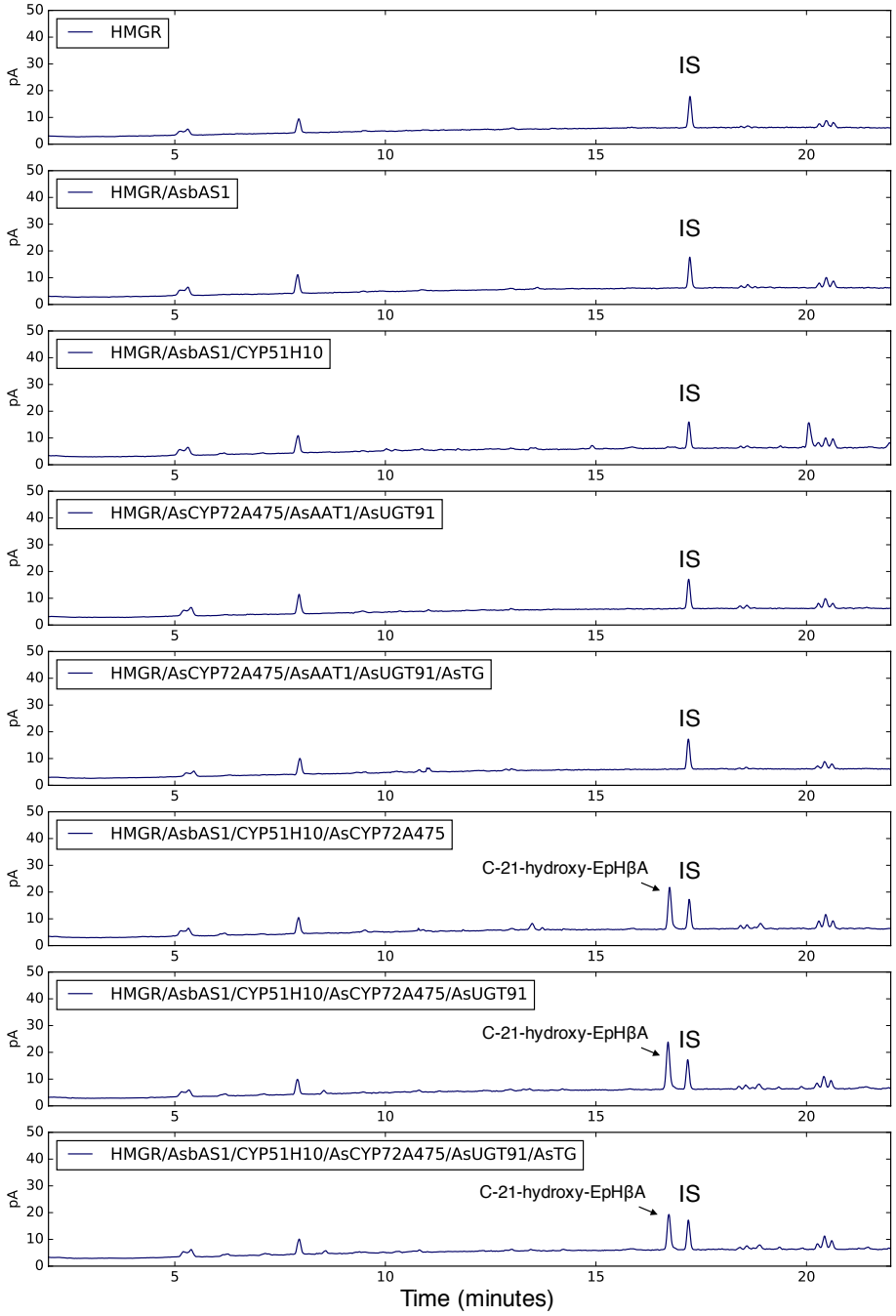


Figure S4: Analysis of *N. benthamiana* leaf extracts expressing HMGR, AsbAS1, AsCYP51H10, AsCYP72A475, AsAAT1, AsUGT91 and AsTG by HPLC-CAD.
AsbAS1 and AsCYP51H10 form the triterpene compound, 2,13β-epoxy-16β-hydroxy-β-amyrin (EpHβA) (Haralampidis et al., 2001; Qi et al., 2006) which is insufficiently polar to be detected in the HPLC analysis. The P450 AsCYP72A475 oxidises EpHβA at the C-21 position to form 12,13β-epoxy, 16β,21β-dihydroxy-β-amyrin (C-21-hydroxy-EpHβA) (Reed, 2016; Leveau et al., manuscript in preparation). Coexpression of HMGR, AsbAS1, AsCYP51H10 and AsCYP72A475 results in the accumulation of C-21-hydroxy-EpHβA. The addition of AsUGT91 or both AsUGT91 and AsTG does not result in the accumulation of any new peaks, suggesting AsUGT91 and AsTG are not active on C-21-hydroxy-EpHβA. Coexpression of HMGR, the P450 AsCYP72A475, the arabinosyltransferase AsAAT1 (Louveau et al., manuscript in preparation) and AsUGT91 or both AsUGT91 and AsTG without AsbAS1 and AsCYP51H10 do not result in the accumulation of any new peaks, suggesting that these enzymes do not modify endogenous *N. benthamiana* compounds. IS = internal standard (digitoxin).

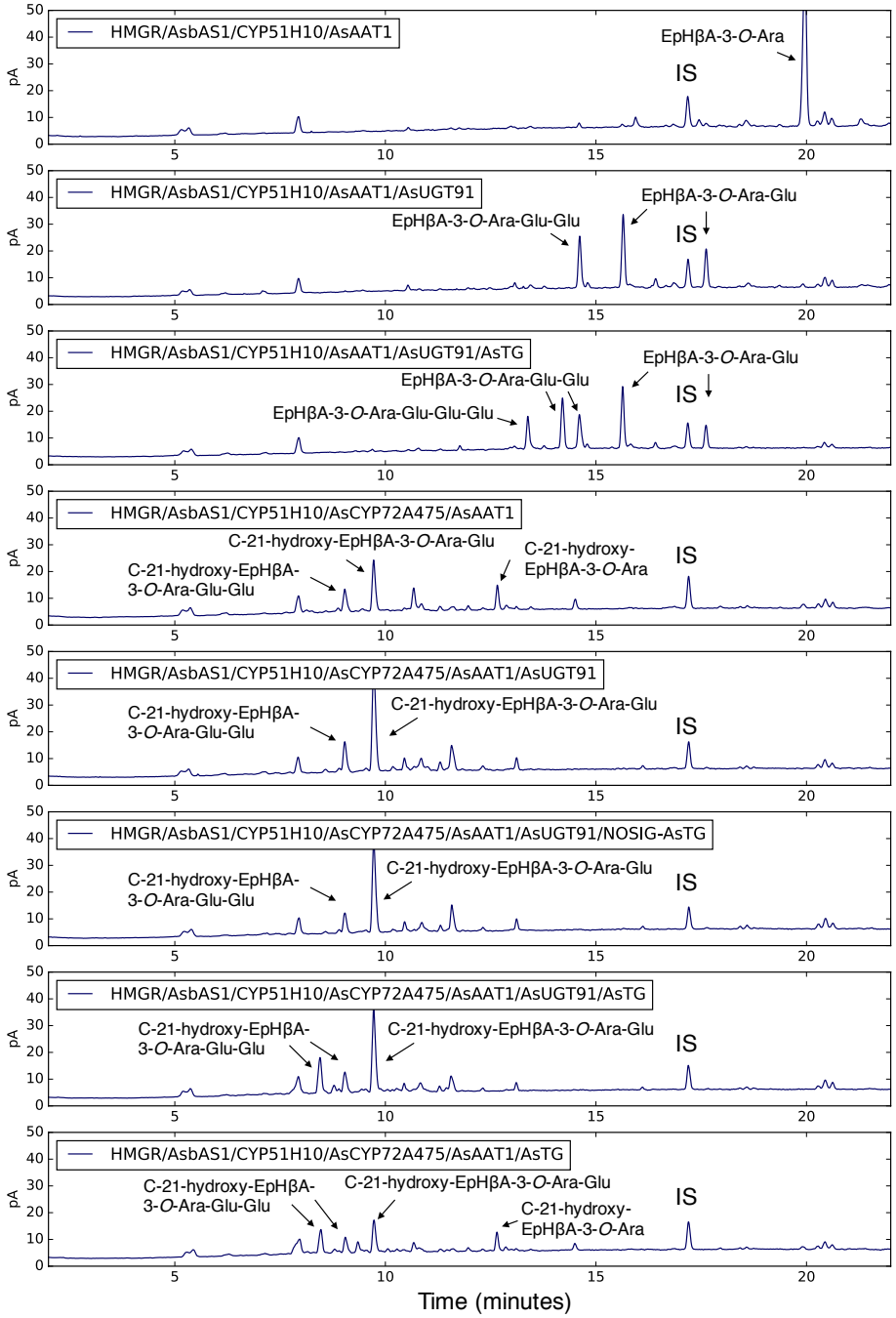


Figure S4: Analysis of *N. benthamiana* leaf extracts expressing HMGR, AsbAS1, AsCYP51H10, AsCYP72A475, AsAAT1, AsUGT91 and AsTG by HPLC-CAD (continued). AsbAS1 (SAD1), AsCYP51H10 (SAD2) and AsAAT1 form the triterpene glycoside, 12,13β-epoxy, 16β-hydroxy-β-amyrin-3-*O*-α-L-arabinose (EpHβA-3-*O*-Ara) (Haralampidis et al., 2001; Qi et al., 2006; Louveau et al., manucript in preparation). EpHβA-3-*O*-Ara accumulation is detected in leaf extracts expressing HMGR, AsbAS1, AsCYP51H10 and AsAAT1. The addition of AsUGT91 and AsTG resulted in the accumulation of new peaks with the mass of EpHβA-3-*O*-Ara with the addition of one or two glucose molecules, EpHβA-3-*O*-Ara-Glu and EpHβA-3-*O*-Ara-Glu-Glu, respectively. Additional peaks due to endogenous *N. benthamiana* glycosyltransferases (EpHβA-3-*O*-Ara-Glu-Glu and EpHβA-3-*O*-Ara-Glu-Glu-Glu) were also present.
AsbAS1 (SAD1), AsCYP51H10 (SAD2), AsCYP72A475 (SAD6) and AsAAT1 form the triterpene glycoside 12,13β-epoxy, 16β,21β-dihydroxy-β-amyrin-3-*O*-α-L-arabinose (C-21-hydroxy-EpHβA-3-*O*-Ara) (Haralampidis et al., 2001; Qi et al., 2006; Reed, 2016; Leveau et al., manuscript in preparation; Louveau et al., manuscript in preparation). Coexpression of HMGR, AsbAS1, AsCYP51H10, AsCYP72A475 and AsAAT1 resulted in the accumulation of C-21-hydroxy-EpHβA-3-*O*-Ara and additional more polar peaks due to the modification of this product by endogenous *N. benthamiana* glycosyltransferases. The addition of AsUGT91 resulted in the reduction of the C-21-hydroxy-EpHβA-3-*O*-Ara peak and an increase in the C-21-hydroxy-EpHβA-3-*O*-Ara-Glu peak. Coexpression of AsTG with HMGR, AsbAS1, AsCYP51H10, AsCYP72A475, AsAAT1 and AsUGT91 resulted in a new more polar peak (C-21-hydroxy-EpHβA-3-*O*-Ara-Glu-Glu) which was not present if the N-terminal signal peptide of AsTG was deleted (NOSIG:AsTG). Expression of AsbAS1, AsCYP51H10, AsCYP72A475, AsAAT1 and AsTG without AsUGT91 lead to a reduction in the new peak and an increase in the accumulation of the precursor C-21-hydroxy-EpHβA-3-*O*-Ara.
IS = internal standard (digitoxin).

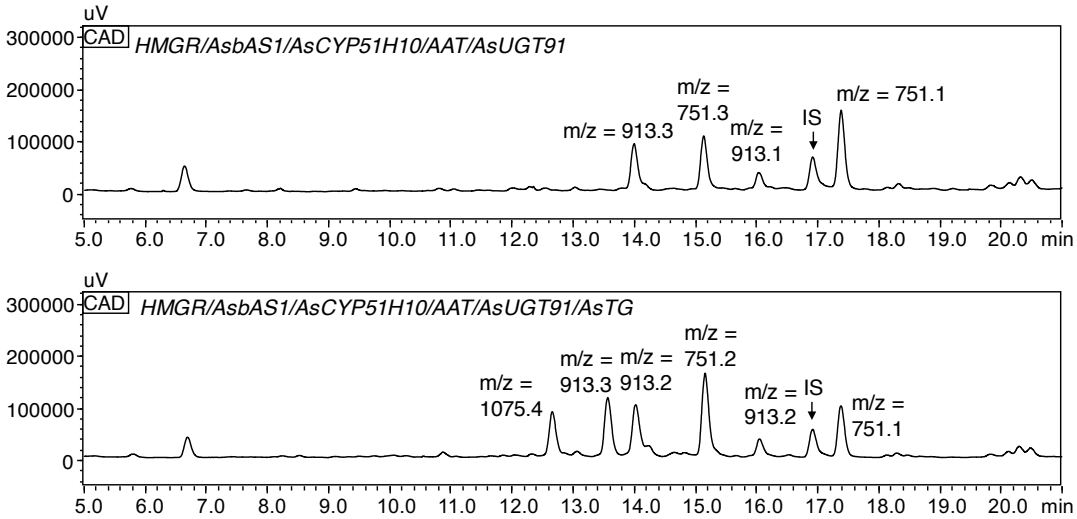
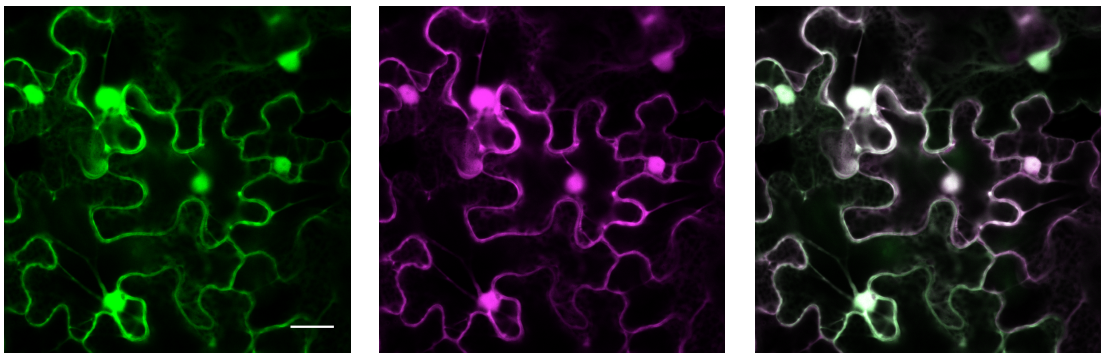


Figure S5: Expression of AsbAS1, AsCYP51H10, AAT, AsUGT91 and AsTG in *N. benthamiana* leaves. HPLC-CAD analysis of extracts from infiltrated *N. benthamiana* leaves. Accumulation of EpHβA-3-*O*-Ara-1,2-Glu is detected in *N. benthamiana* leaves co-expressing AsbAS1 (SAD1), AsCYP51H10 (SAD2), AsAAT1 and AsUGT91.

The addition of AsTG results in new peaks with the mass of EpHβA-3-*O*-Ara-1,2-Glu with the addition of one [*m/z* = 913.3, Retention time = 13.6 min] or two [*m/z* = 1075.4, Retention time = 12.7 min] hexoses. This latter peak is likely to be due to an additional glycosylation by an endogenous *N. benthamiana* enzyme. Data is representative of two replicates. IS = internal standard (digitoxin). EpHβA-3-*O*-Ara-1,2-Glu = EpHβA-3-*O*-Ara = 12,13β-epoxy,16β-hydroxy-β-amyrin-3-*O*-α-L-arabinose-1,2-β-D-glucose.

(A) GFP:AsUGT91 and free mRFP1



(B) AsUGT91:GFP and free mRFP1

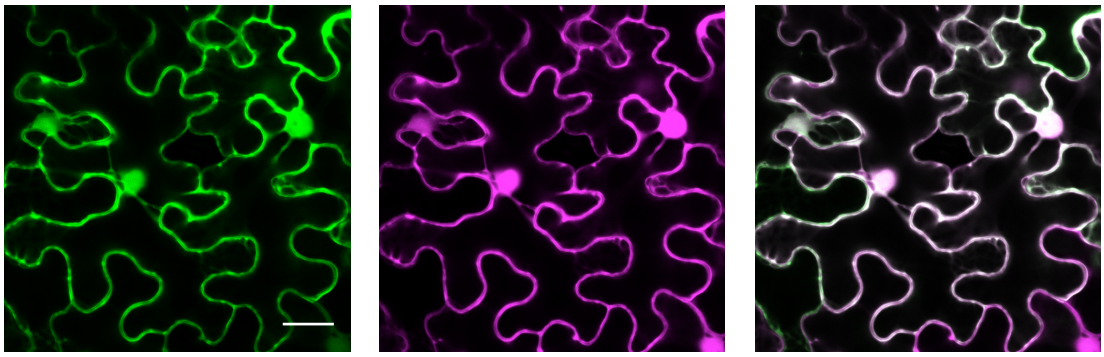
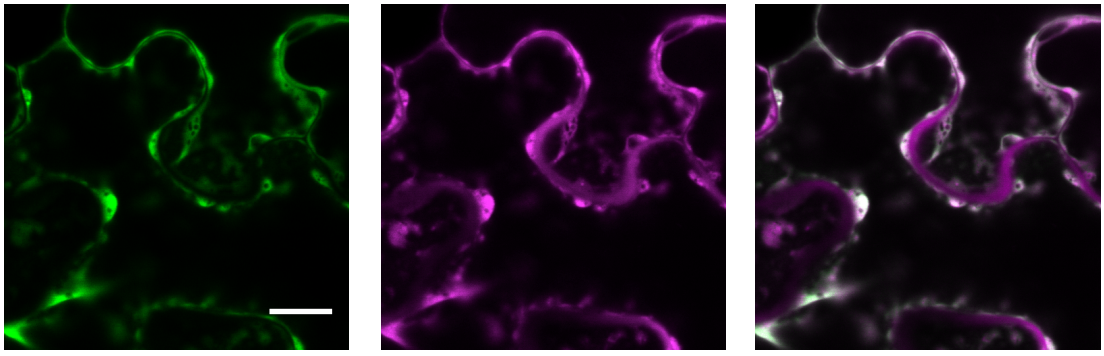


Figure S6: AsUGT91 localises to the cytosol in *N. benthamiana* leaves.
Co-localisation of free mRFP1 (Moglia et al., 2014) with: **(A)**, GFP:AsUGT91 and **(B)**, AsUGT91:GFP. The GFP fusions to AsUGT91 co-localise with free mRFP1 in the cytosol and nucleus. GFP fusions are shown in green (left); RFP is shown in magenta (middle) and merged images are shown in white (right). Images are taken two days post-infiltration. Bar = 20 µm. Image credits: Ingo Appelhagen, John Innes Centre.

(A) GFP and SIG:RFP; 2 days post-infiltration



(B) GFP and SIG:RFP; 3 days post-infiltration

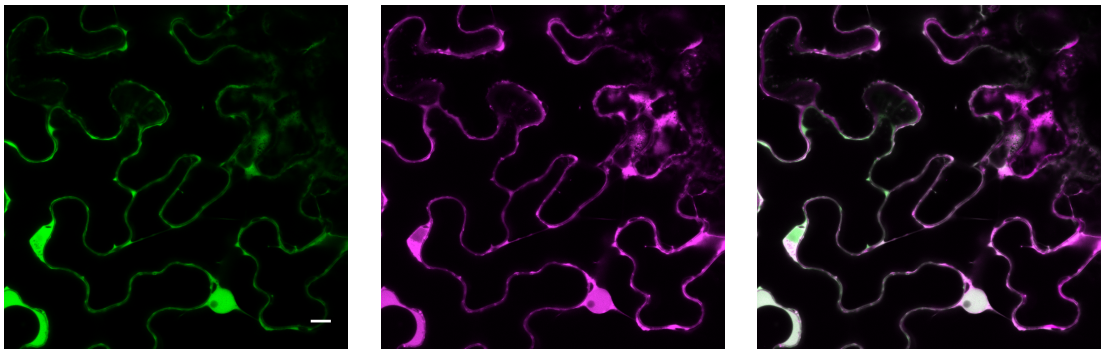


Figure S7: The N-terminal signal sequence is not sufficient for AsTG localisation. The AsTG N-terminal signal peptide fused to RFP is targeted both to the apoplast and to the cytoplasm where it co-localises with free GFP. Co-expression of a GFP construct with RFP preceded by the AsTG N-terminal signal peptide (SIG:RFP) **(A)**, two days post-infiltration and **(B)**, three days post-infiltration. GFP is shown in green (left); RFP fusions are shown in magenta (middle) and merged images are shown in white (right). Bar = 10 µm. Image credits: Ingo Appelhagen, John Innes Centre.

Figure S8: Protein sequence alignment of oat AsTG with GH1 enzymes. AsTG is aligned with the GH1 transglucosidases: BAM29304 AaAA7GT, from *Agapanthus africanus*; BAO96250 CmAA7GT from *Campanula medium*; E3W9M3 DgAA7GT, BAO04178 DgAA7BG-GT1 and BAO04181 DgAA7BG-GT2 from *Delphinium grandiflorum*; E3W9M2 DcAA5GT, from *Dianthus caryophyllus*; B7F7K7 Os9bglu31, from rice; and also the GH1 enzymes AJW76497 AtBGLU6 and Q93ZI4 AtBGLU10 from *Arabidopsis thaliana*. The catalytic acid/base and nucleophile are highlighted in bold text. The predicted N-terminal signal sequence of AsTG is underlined in blue bold text. The amino acid substitutions in *sad3* mutants #368, #891 and #1429 are underlined in red. Sequences were aligned with Clustal Omega (Goujon et al., 2010; Sievers et al., 2011).

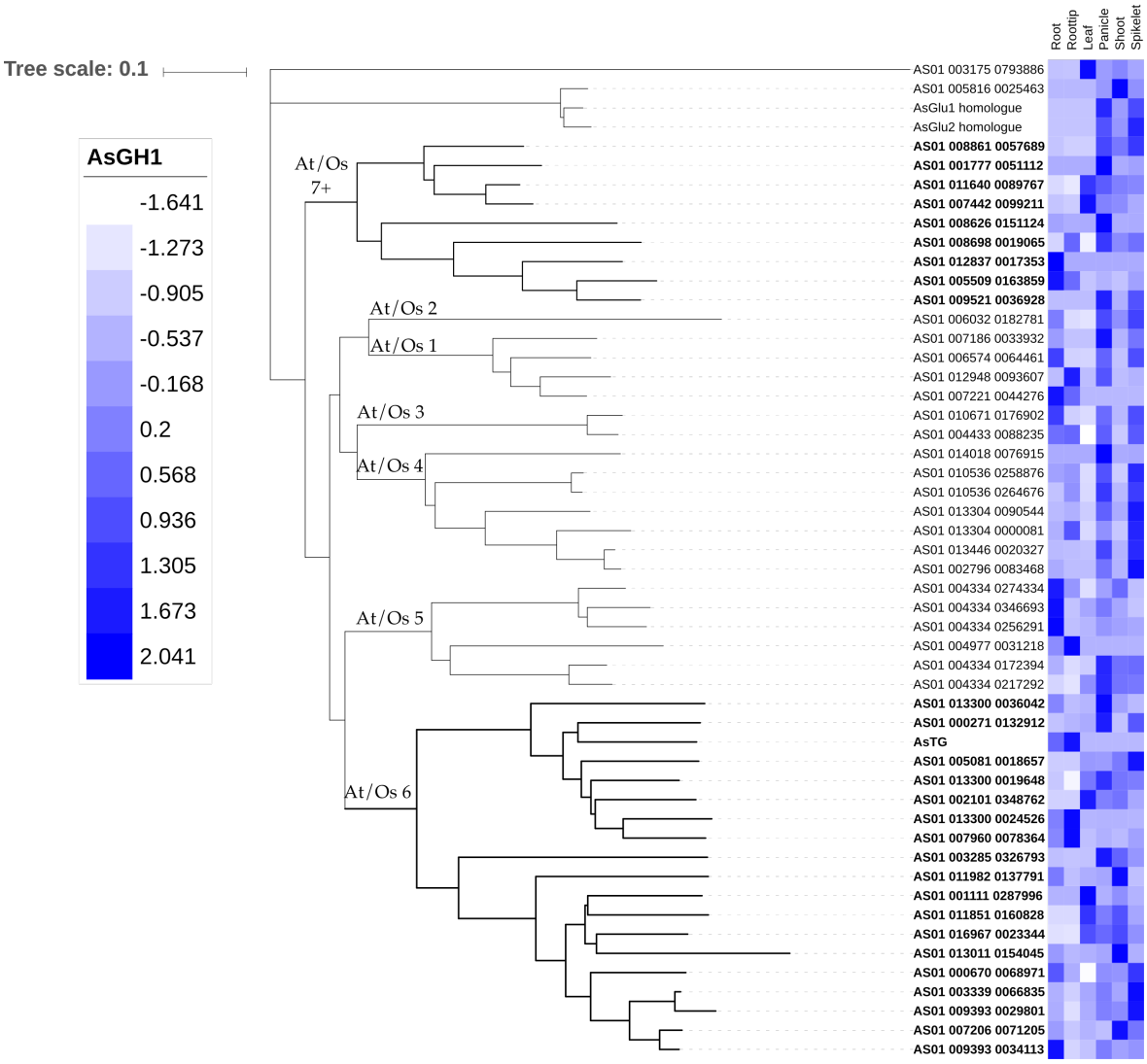


Figure S9: Expression profiles of *A. strigosa* glycosyl hydrolase family 1 (GH1) enzymes. 52 GH1 enzymes were identified by a BLAST (tblastn) search against predicted coding DNA sequences of the *A. strigosa* genome (CDS sequences provided by the group of Bin Han, CAS). The expression profiles of these enzymes as determined by RNA-seq are shown. Phylogenetic groups were determined by comparison with GH1 enzymes in other plant species (listed in Chapter 5, section 5.5.3; data not shown) and are labelled as designated in Luang et al. (2013). Phylogenetic analysis was carried out as in Chapter 5 section 5.5.3 and visualised using the Interactive tree of life (iTOL) software (Letunic and Bork, 2016) (https://itol.embl.de/).

A. strigosa UGT protein sequences

>AsUGT91
MAASASRPLHLVICPWLAFGHLLPCLELAHRLASRGHRVSFVSTPRNISRLPPLPPAVAPLVNFVALPLAQVPGL
PDGAEATSDVGDDKAELLRLASDGLAEPFSEFLRAAKKPDWLIVDIVNNWAAAAAAEHKVPCVALLQCAARMFTP
LAGPMSHNDDWMKTYTVESPASGVSIAERCAITLKACKLAALRSCLEWEPDAVPLVKTHPQSGTPVVTLGLLPPP
PPSADTRGKDDDDDATVRWLDAQPAKSVVYIALGSEVPLREEQVHEMALGLDLSGTRFLWALRRPRDAAPDADAD
DVLPPGFEERTRGRGLVVFGWVPQVSILAHGAVAAFLTHCGWSSTIEGLRFGRPLVMLPIATGDQWPNARLMEER
GVGLRVPRDGNDGSFHREGLAATVRAVTADQGGTFAANAGKLQLVVADRECHERCIDGFIQQLTSYMD
>AsGT01194
MHVVMLPWLAFGHILPFTEFAKRVARQGHRVTFLSTPNNTRRLIDIPAALAGLIRVVDVPLPRVHRLPEHAEATI
DLPSDALRPYLRRAYDAAFQDELSRLLQEEEESWPRPDWVIVDYASYWAPAAAARHGVPCAFLSLFGAAALSFFG
GPDALLGLGQHAKTEPAHFTVVPAYVPFPTTVAYRGYEARELFEPGMVPDDSGVSEAYRYAKTIQGCELVGIRSS
AQFEPEWLRLLGELYEKPVVPVGFFPPPPQQDVADHEATIRWLDEQAPGSVVYAAFGSEVKLTSAQLRRIALGLE
ASGLPFIWAFRPPTDARGGDGLPEGFEERLAGRGVVCRGWVPQVRFLAHASVGGFLTHAGWNSVAEGLAQGVRLV
LLPLVFDQGLNARNLVDKKIGVEVARDEQDGSFGPDDIAAALRRVMVEDEGEDFGAKVKELAKVFGDDEVNDQSV
RDFLRHLSEYIRKNEQ
>AsGT01332
MVFPWLAFGHMIPFLELSNRLARRGHFVTFVSTPRNVSRLRVVPPELSSNVRVVSLVLPEVDGLPEGAESTADVP
PEKVGLLKKAFDGLAAPFAGLVADACAAASSGDAAAGFTRKPSWIILDFAHNWIWPIAEEHEIPCAMFLIFPAAL
LAFVGPRHENQAHPRSTVEDYMVPPPWIDFPSTLAYRRHEAKWVAAMFRPNASGVSDVDRFWQMQRPCCRLILLR
SCPEIEPRLFPLLTGLFAKPAVPTGLLMPAVDDDTVQAPNADAAADQSLLAWLDEQPRKSVMYVALGSEAPVTAE
LLRELALGLELSGARFLWALRRPSGHAGELLPHGFESRLSGRGVVCMGWVPQVRVLAHGAVGAFLTHCGWGSTVE
SFRFGHALVMLPFVADQGLVAQAMAARGVGVEVARDYDDGSFHRDGVAAAVRRVMVEEEGKVLARNAKALQELLG
DRARQDRYVDQLVQDLHYYK
>AsGT18280
MDATPSACSSPPLHVVICPWLAFGHLLPCLDIAERLASRGHRVSFVSTPRNIARLPPVRPAVAPLVDFVALPLPH
VEGIPDGAESTNDVPYEKFELHRKAFDGLGASFSEFLGAACAEGGKKPDWILVDVYHHWAAAAANEHKVPCAMLL
LGAATVIADWVGRSGQEPSEAAQPPRFETERMKLMSTQRGSGMTVAERFSLTLMRCHFVAIRSCLEWEPQSVPRL
TTLWGKPVIPLGLLPPSPAGGRGGSKDGEDATVRWLDEQPAKSVVYVALGSEVPLAAEQVHELALGLELAGTRFL
WALRKPSGVADADVLPPGFEERKRGRGLVVTGWVPQISVLAHGAVAAFLTHCGWNSTIEGLLFGHPLIMLPIFGD
QGPNARLMEGRKVGVQVARNENDGSFNREDVAAAVQAVAVEEDTRKIFVANAKKMQEIVADSKCHERCIDGFIQQ
LRSYQE
>AsGT19358
MDAAGSSSSPPAPLRIVIVPWLAFGHLLPYLELAERLASRGHRVSYVSTPRNLARLPLPLRSHVDLVALPLPRVD
GLPDGAESTNDVPDDKRELHWKAFDGLAAPFAEFLAAACADEATRPHWVIADSFHHWAAAAASEHDVPLAMLLPT
AAMVATVPRPPPSPESAAYSFELQAAAAARAVPRYEREGMKMVFANDGASSGMSTEQRCLLTAQGCAVAAIRSCV
EWEPESFPLVAPLLGKPALPLGLLPPPPDGARRAAGTNREHATVRWLDAHPPGSVLYVALGSEVPLSAEQVHELA
LGLDLAGTRFLWALRKPSGAHADDDGDMLPPGFRDRTDGHGLVTTGWVPQMSILAHAAVGGFLTHCGRNSLIESL
LFGHPLVMLPIFGDQGPNARQMEAKKVGLQVARDEEDGSFDRHGVASAVRAVMLEGEARRGFVAGAARMQEVVAD
ADRHERYIDELVEHLRSYSDGASATTTDGNFTTAAIPTASS
>AsGT18279
MDAMFSPSSPPQEPLRIVIVPWLAFGHLLPYLELAERLASRGHRVSYVSTPRNLARLPPLRPAAAPRVDLVALPL
PRVEGLPEGAESTNDVPDDQRELHWTAFDGLAAPFAEFMAVLCADVDTKPHWIFADTFHHWVAASAIEHKVPCAL
LLPTAAVIAMSAQLPEHAPADHAETDQLLRLMTAPRYEQEALAPLFTNHGVSGMSTVQRCTLTKERCTIGVIRSC
VEWEPESFHVVPTRLGMPVVPLGLLPPPPDSGRRAASTNREHAAVRWLETQPPGSVLYVALGSEVPLRVEQVHEL
ALGLELAGTRFLWALRKPSGAAVLDDGADMLPPGFQERTRGQGMVTTGWLPQMSILAHAAVGGFLTHCGRNSLIE
GLLFGHPLVMLPIFGDQGPNARQMERKKVGLQVARDDNDGSFDRHGVATAVQAVMLDGEARRGFVAGAAKMQEVV
ADTERQERYIDELVQHLRSYTVHGDFATGPPTSSSS
>AsGT00243
MATTSGPGRSRPHVLVVPYPAQGHALPLLDLAALLASRGIRLTVVTTPANAPLLSPLLAAHPASVQPLVLPFPAH
PALPPGLENVKNCPPSYFAIFIHALAALHQPILAWAKSQSAAAQPVVAVLSDFFCGWTQPLAAELGVPRLVFSPS
GVLGTAVPHSLFRRLVKRPSEADDVVTFPAIPGERSYQWREISGTYINYIEGRMEEQVGAAVRQNFLWNLESWGF

VSNTLRALEGKYLNAPLQDLGFKRVWAVGPLAPETHTSGKLGGDLGEWLDGFSEGTVVYVCFGSQALLTPAVAAA
LGEALERSAVPFVWAVGAGDGVVLPEGFEARASTAKRGLVVRGWAPQVAILRHAAVGWFMTHCGWNSMLEAVAAG
VPMLAWPMAADQFVNARLLVDEARVAVRACAGGLGVAPDGGELAAVLADAVGEGGTGVRARAKELAAEAAVATKV
GGRSREDLEVLVEEIRKL
>AsGT22027
MAAAAAADCPAAPMPHILVVPYPAQGHTIPLLDLAYLLAARGLRLTVVATPATAPLLAPLLAAHPDGAVRALTLP
FPSHPAFPAGVESAKGCPPVLFGALIVAFSGLRSPLLSWARALSGSPDRVVAVLSDFFCGWTQPLAADLGVPRIA
FSSSAVYATAVLHSLLRRLPRREDESDDESPITFPGLPGIPAYPWRQLSLLYRTYKEEDEVAEGVRSNFLWNLDC
SSFVSNTFRRLEAPYLETPLADLGFRRVRAVGPLAPRDAVVSRGGETAVAAADLCAWLDGFEDGSVLYISFGSMA
VLQPAHAAALAAALERTGAAFVWAVGEAVTFPEGFEERTAAGGRGTVIRGWAPQVAALRHGAVGWFVTHCGWNSV
LESAAAGVTILAWPMTADQFVNARLLVDELRAAVPVCWGGLATAPSADELVSVLEATVAGKKEGREEAWGGVAAR
AKELAEEAAAAVREGGSSWREVEELARELRELGSEPKIAHGEASLRASI
>AsGT00733
MGSMAEAGVSCDASRELPHVVIFPFMARGHTIPLTQLANLLLRRRLATVTFFTTPGNAAFVRALLPAGADVVELP
FPAAQGAENVEGIASVSSFAGFAEATLALQPRFEEALASMRPAASLLVADAFLYWTGASAAALGVPRVSFLGTSA
FAHVMREMFVRDKPGSDATGGSTYTVPEFPHVQFPLADIPPLPISVLALDAKMGMAVAGSRGVIMNTFHEQESRY
IEHWNRHVGPRAWPIGPLCLARESSTDVDRVHGTQAPSLVVWLDEKAAAGQSVIFVALGTLLAVSDAQLGEVARG
LEEAQVNFLWALRPDDNVDLEFEERVQGRGMVTRGWVNQQAILQHDCVAGFLSHCGWNSVLESVCAGVPLAVWPM
VFDQPLNAKLVVDELGVGVRVKTSDDTTGGGGGLVKSGEISRAVREIMLGETRASALKQNAAVLAEQAALAMSAG
GSSWNMVEEMISELCVAADTLG
>AsGT01461
MATTTTADGAAAANHTCGRDHVVLFPFMAKGHTLPLLHFATALSTVHHKGPCLRVTVVTTPGNLAFVGSHVPSSV
RLAVLPFPSLPPLPAGVESTDALPSPSLHPTFMRAAALLRQPFADLMASLSSPPLVLVSDFFLGFTGRVAADAGV
RRVVFHGMSCFSIALGKSAKMGPLPSCAGRGESFHVPRMPAHLKFTAAEVPPDALPKTTDHEHEEDDTSSTRSVV
DKEEYSDASSWGVLVNSFAKLDQEFVATLESFYQPDARAWLVGPLFIAAGDIVEEQDPEGCLPWIDEQSEPVVYV
SFGTQAHVSDEQLGELARGLAQSGHPFLWAVRSETWSPPVDVGPRGRLVRGWVPQRSVLANRAVGGFVSHCGWNS
VMESLAAGKPVLAWPMMAEQPVNARHVADIVGAGIRMDVGDSSVVVGRAEVEKKIRRLMDTGGEEGSRIRARAAW
AQQAAMSAVSDGGTSRTALLKLLEVLQGSYDRDVVV
>AsGT02436
METSAKPHFVLVPWIGSISHIVPMTDIGCLLASHGASVAIITTPANAPLVQSRVDRVTPRGAVIGVTTIPFPTAE
AGLPDGCERLDLVRSPAMVPSFFKANKKFGEAVAQYCRREDAPRRPSCIIAGMCNTWTLPVARDLGVPCYIFHGF
GAFALLCIDHLYRQGRHEAIASAEELVDISVLQQFECKILGRQLPPHFLPSTSMGGGLMQEVRDFDVAVDGVVVN
SFDELEHGSAALLAAASGRKVLAVGPVSLCCAPSLDPQGDDARRCMAWLDGKEANSVVYVSFGSAGSLPPAQLIQ
LGMALVSCRWPVMWVIRGADSLPDDVNAWLGENTDPDGVADGKCLVVRGWAPQVAILAHPAVGGFMTHCGWGSTL
ESVAAGVPMLTWPLFAEQFVNEKLIVDVLAIGVSVGVTKPTENVLTASKLGSGEAKAEVGAERVEKALERLMDGA
GEGEDIRRRAAELKRKANAALEKGGSSYNNLENLIDSCA
>AsGT16f23
MTFARGNVHSGSASAHFVLVPMMAQGHTIPMTDMARLLAEHGAQVSFVTTPVNASRLAGFIADVEAAGLAIRFVE
LHFPTTEFGLPDGCENLDLIQAKGLFLNFMEACAALREPLMAHLREQHQLSPPSCIISDMMHWWTGDIARELGIP
RLTFIGFCGFSSLVRYIISQNNLLENMTDENELITIPGFPTHLELTKAKCPGSLCVPGMEKIREKMIEEELRSDG
EVINSFQELETVYIESFEQVAKKKAWTVGPMCLCHRDSNTMAARGSKASMDEAQCLQWLDSMKPGSVIFVSFGSL
AATTPQQLVELGLGLEASKKPFIWVIKAGPKFPEVEEWLADGFEERVKDRGMIIRGWAPQMMILWHQAIGGFMTH
CGWNSTVEGICAGVPMITWPHFAEHFLNEKLVVDVLKTGLEVGVKGVTQWGNTEQEVMVTRDAVETAVYTLMGEG
KAAEELRMRAKHYAIKARRAFDEEGSSYNNVRLLIQEMGNNTNACG
>AsGT12842
MAPPLPPPHFVLVPLLLQGHTIPMVDLGFLFAERGARVSIVTTTVTNAAARLQEGIVADRARRRANKLLLVEIVE
LPFPGPGLPPAARPCRSGSEKVDDFLPLLQALHKLAGPLEAYLRGLARRPSCIINDSLCSWTVSVARRVGVPRLS
FQAASCFFALCALNVHSHSGGGGGGGDHGYVVPGMPVRVEMSEDTWSSSFLTTPVWEAFLEEAREGVRAADGVVA
NTFLELEGEFVEGYEAALGKPVWALGPSCLGSQTQTQGDDDQLIIMAWLDAMDDSAVTYVSFGSLVRMLPSQLYE
VGHGLEQSGKPFLWVVKDDDKAETEEWLQALEARTAGRGLVARGWAPQLAILSHRAVGGFLTHCGWNSLVESVTL
GVPLLTWPHFADHFLNEQLAVQVLGVAVPVCRRRRHDVVRAEDIARAVSELMGGGEVAQDRRAKCKQYAEKAHRA
IQEGGSSYENMTQLLRTFTPSSANSQQH
>AsGT14h21

MAPTETAAPPPPPPHFVLVPLIGQGHTIPMGDLACLLAERGARVSLVTTPVNAARLQGVADRARRARLPLEIVEL
PLPPADDGLPPGSANVDNFLRLFLDLYRLAGPLEAYVRALPWRPSCIISDSCNPWMAGVARSVGVPRLFFNGPSC
FYSLCSHNVARHGLLHDGEGEGERDAYVVTGVPVRVEMTKDTWSAALLTCMPKWEAFLQDVREGMRTADGAVVNT
FLDLEEQFVACYRTALGKPVWALGPFFLGNRDEEAVAARGGKDKPSAVAQSAVTAWLETMDQSTVTYVCFGSFAR
MLPKQLYEVGHGLEDSGKPFLLALKESETALPEAQEWLQALEARTAGKGLVVRGWAPQLAILSHRAVGGFVTHCG
WNSLLESVAHGVPVVTWPHSGDQFLNERLAIEVLGVGAPVRGAVVPVTPFDESKAVAPVLRGHIAEAVSELMGGG
AVARERRRKCKEYGERAHAAIAKGGSSHENLTQLLQSFMRSGSKEQ
>AsAAT1
MGKPAAGEELAAAGYEGKQQQRAHFVFIPLMAQGHVIPSLDTALLLASQGAVCSMVATPWTAARIRPCIEQCGLE
VRLLEFPLEYVSDGADNLDNIPPERVVGYFQAVALLQAPIQERLGKLEPRVSCIVSDFSHPWTAPVAAALGVPRI
SFFPMCAFCALSEHNVHEYNKCLCSPGSEEVVAVPLLDATLLEMRRVEAPCFFRHASMGTLGEDIQRAHSQGAGM
IFNSFLELEPDYVRGLSSAWGGKKVWTVGPVSLHHQLAAATTCRGEEASMDDDCLQWLDGKEPGSVVYVSFGTIA
PKMEPEMLLELARALETSGHPFIWALSKADHPFAETSQELQELEARVAASGGGRIVRGWVPQLLILSHAAVGCLF
THSGWNSVMEAITAGKPAVTWPRLIGSDHFVNEKFTVEVLRIGVSVRPEDPHIQPVEVRREAIQAALTAVLEGGD
EGQERRNRVRDLSLKAKAAMQPGGSSHANLSDLVQRFTI
>AsGT000892-1
MGTLSELHFVLVPLAAQGHIIPMVDVARLIAARGPRVTVVTTPVNAARNRATVDGARRAGLPLELVELPFSGPEH
GLPEGMEAVDQLTAGAHELGMFLKFFQAIWNLAGPLEEYIRALPRPPICLVADLCNPWVAAVCERLGIPRLVMHC
PSAYFQLTVNCLITHGVYDRVEEMEPFEVPGFPVRAVGMKATMRGFFQYPGAEKEYRDTLDAEATADGLLFNTFR
GIEGTFLDAYAAALGKPTWAVGPTCASSTMADDADSKAGRGNTADVDAGHIVSWLDARPPASVLYVSFGSIAQLT
AKQLANLARGLEASGRPFVWAIKEAKGDAAVRALLDEEGFEARVKDRALLVRGWAPQVTILSHPAVGGFLTHCGW
NGTLEALSIGVPALTWPNIADQFCGEQLLVDVLGVGVRSGVKIPAMYLPAEAEGVQVESGDIEKAVAELMGDAPE
AAARRSRAKEIAAEAKAAMEEGGSSHSDLTDLINYVSELSRKKEHKNSTALPSAVAELGGKDTGKIEDDSVL
>AsGT000892-2
MAASTELHFLLVPLVAQGHIIPMMDLARILAARGPRVTVVTTPVNAARNRATVDAANRAGHAVELVEIPFPGPRL
GLPEGMEAIDQMVDKAMYLTFFHAIWKMAEPLEEYVRALPRKPDCLVADSCNPWTAGVCSAHGIPRLVMHCPSAY
FLLAVHNLSAHGVYDSVGDDEMAPFEVPDFPVRAVGNKATFRGFFQYPGVEKEQQDVIEAEASADGLLFNTFRGV
EGVFVDAYAAALGRRTWAVGPTCASPSMDADAKAGRGNRADVDVNHIVTWLDARPRESVLYISFGSIAQLPGKQL
AELARGLEKSGRSFVWAIKEAKASADVRALLDEEGLEDRVKDRGLLVRGWAPQVTILSHPAVGGFLTHCGWNATL
EAISHGVPALTWPSFADQFSSERLLVDVLGVGVRSGVKVPVMNVPEEAEGVQVESGDVEKAIAKLMDEGPEGAAR
RSKAKELAAEARAAMEEGGSSYSDLTDMIRYILDLSQKRSHQREASTTDLSMSFAAAELRNNKGNKIESNAALSV
QS
>AsGT27f7
MATLPELHFVMVPHPAQGHIIPMVDVARLIGARGPRVTVVTTPVNAARNRATVDGARRAGLPLELLELPFSGPQH
GLPEGIEAVDQLTGHELGMFLSFFHATWNLAGPLEEYIRALPRLPVCLVADMCNPWTAAVCERLGIPRLVMHCPS
AYFQLAVNCLTAHGVYDRVDEMEPFEVPGFPVRAVGTKATFRGFFQYPGVEKEYRDTIEAEATADGLLFNTFRGI
EGTFLDAYASALGKKTWAVGPTCASSTMADDADSKAGRGKPADVDAGHIVSWLDARPPASVLYVSFGSISQLTTK
QFAKLARGLEASGRPFVWAIKEAKADAAVRALLDEEGFEARVKDRALLVRGWAPQVTILSHPAVGGFLTHCGWNG
TLETLSLGVPTLTWPTIADQFCSEQLLVDVLGVGVRSGAKLPAWYLPTEAEGVQVESGDVEKAVAELMGDTPEAA
ARRSRAKELAAKARTAMEEGGSSYSDLTDMIRYVSELSRKRSLEIDAMPFAAAELGSNKGEKIEADAALSVQS
>AsGT23781
MANQTQAQAAGDEAKADEAPHFVFVPLMAQGHIIPAVDTALLLATHGALCTVVATPSTAARVRPTIDSARRSGLQ
VSLVDFPLDYAAVGLPDGVPGGADNMDNVPPEYMLAYYSAIALLCGPIETYLRAHAPRPPTCVVSDFCHPWTTKL
AASLGVPRLSFFSMCAFCVLCQHNVERFNAYDGVLDPNQPVVVPGLEKRFEVTMAQAPGFFRGWPGWEKFADDVE
RARVDADGVVMNTFEEMEPEYVAGYAAARGMKVWTVGPVSLYHQHARTLAARGATASIDTDDCVRWLDGKDPGSV
VYVSFGSIVHADPKQIMELGLGLEASGYPFVWVVKGAERHDEAALAFLRGLEERVAGRGLLVWGWAPQALILSHR
AAGAFVTHCGWNSTLEAVAAGLPVVTWPHFTDQFLNEKLAVEVLGIGVSVGVKEPVMYQVDKKEIVVGRSAVEAA
VRSAMDGGEDGQERRRRARELAVKARAAVSEGGSSHTNIRDLVKCFGVGASTQDAAE
>AsGT24i2
MAVKDEQQSPLHILLFPFLAPGHLIPIADMAALFASRGVRCTILTTPVNAAIIRSAVDRANDAFRGSDCPAIDIS
VVPFPDVGLPPGVENGNALTSPADRLKFFQAVAELREPFDRFLADNHPDAVVSDSFFHWSTDAAAEHGVPRLGFL
GSSMFAGSCNESTLHNNPLETAADDPDALVSLPGLPHRVELRRSQTMDPKKRPDHWALLESVNAADQKSFGEVFN
SFHELEPDYVEHYQTTLGRRTWLVGPVALASKDMAGRGSTSARSPDADSCLRWLDTKQPGSVVYVSFGTLIRFSP

AELHELARGLDLSGKNFVWVLGRAGPDSSEWMPQGFADLITPRGDRGFIIRGWAPQMLILNHRALGGFVTHCGWN
STLESVSAGVPMVTWPRFADQFQNEKLIVEVLKVGVSIGAKDYGSGIENHDVIRGEVIAESIGKLMGSSEESDAI
QRKAKDLGAEARSAVENGGSSYNDVGRLMDELMARRSSVKVGEDIIPTNDGL
>AsGT10433
MASTVPPHFVLVPLSSQGHIIPTSDLACLLAERGARVSLVTTPGDAARLQGVADRARRAKLPLEIVELPFQPADD
GVAPGSEGVDTVLRWFQSLYRLAGPLEEYVRALPRRPSCIISDLLNPWTVGVARNVGVPRLVYATPSCFYSLCDL
NVATHKDRPEGEFVVPGMPMRVELTKGTWPLAFFSPPSWEAFLEERNESLRTADGVVVNSFLDIEGQFAECLGAA
LGMPVWVLGPFFLNNRDEGALAERTDQYKPSASLDENAVTAWLDDMARRGTKVAYVNFGTVVGMCAGQLYEVGHG
LEDSGTSFVWVVREEETATPEAREWLEALEARTAGRGLIVRGWAPQVAILSHPAVGGFVTHCGWNSLLESITHGV
AVVTWPKFADQILNERLAVDVLGIGAPVSATPPVVGQLSVLRADVARAVSDLLGDGEAAQERRRKAKEYGQRAHA
AMAEGGSSYQNLTKLVQSFAQSGGNGTISTTPS
>AsGT11099
MATKDDEQEQPPLHILFFPFPAPGHLIPMADMAALFATRGARCTILTTPVNAAIIRPAVDRANAHSAAIEISVVP
FPCVGLPPGFENATGLTDQADRDKFARAIQLLRREPFAQFLADNKRNHSVDAVVSDSQFQWSVDAAAEHGVPRIAF
LGTSMFARACTDVMLRENPMTEGPSPDQDDPDATMVLLPGLPHRVELPRSQMPDPRTQAKVIAFFKVVNEQDQRS
YGELFNSFHELEPEYVEHYRATLGRRVWLVGPVARPHNMDARSRNNNNSTLSPDADGCLRWLDGKPDGSVAYVSF
GTLSNFSPDELRELARGLDLSGRCFLWVIGTGAITNAAPSSEWMPEGFPELLARGERGLVVRGWAPQVLVLNHPS
IGGFVTHCGWNSVLEAVSAGVPMVTWPRCGDQFFNEKLVVEVLKVGASVGAKDYASAIESHQVIGGEVVAESIGR
VMDDDAGYTVRSNAKELGVKARRAVEKGGSSYGDVGRLMEELMARRS
>AsGT23141
MAASELHFLLVPLVAQGHIIPMVDLARLIAARGPRVTVLTTPVNAARNRVAVEGAARAGLAVELVELPFPGPQLG
LPEGLENADQMVDRTMYYKFFEAIWKMAEPLEAYVRALPRRPDCLVADSCNPWTAGVCDALGVPRLVMHCPSAYF
LLAVHNLSAHGVYDRVGDDEDEEFEVPGFPVRAVGNKATFRGFFQWPGVEKEHQDVLHAEATADGLIMNTFRGIE
GAFVDAYAAALGTRTWAVGPTCAASGGAVGDDADARAGRGNRADVDAGRVVSWLDARPPASVLYISFGSIAQLPA
KQVEELARGLEASGRPFVWAIKEAKADAAVRALLDDEGFEERIKDRGLLVRGWAPQVTILSHPAVGGFLTHCGWN
ATLETISHGVPALTWPNFADQFCSERLLVDVLGVGVRSGVKVPAMNAAGVDVGVQVTSEDVERVVAELMDGGAEG
AARRSRAKELAAEARAAMEEGGSSYTDLEDMILYVSQLSRTRRHERDAGSTPVPSTTAELETKNGAEKIEADAAL
SVKS
>AsGT14h20
MAHTETTEKSMILMSSAAPPPPPHFVLIPLIGQGHTIPMVDLAYLLVERGARVSLVTTPVNAARLQGVADRARRA
MQPLEIVELPFPPADDGLSPGSANVDNFLRLFLDLYRLAGPLEAYLRALPRRPSCIISDSCNPWTAGVARSVGVP
RLFFHVASCFYSLCKLKAATHGLLLHDGNKDDAYVEVPGMPVRVEVTKDSWSSSYTTPEWEAFVEDARDGMRTAD
GAVLNTFLGLEGQFVKCFEAALGKPVWALGPFFLNNRDEDAVATRGDKDKPSAVDQDAVTAWLDAMDESAVTYVS
FGSLVRMPPEQLYEVGHGLVDSGKPFVWVVKESETASPEAREWLQALEARTAGRGLVVRGWVSQLAILSHRAIGG
FVTHCGWNSLLESVAHGVPVVTWPHFGDQFLNEQLVVEVLGVGVPVRGAAGPVVPVVREHIERAVSELMGGGAVA
QERRRKCKEFGERAHTAVAKGGSSHENLTQLVHSFVRSGSTEQQTTQDRNC
>AsGT11i11
MASNDNVPTAVTSSINKKLRVLLIPILATSHIGPFTELAISLAATNDAVEATVAVTPANVSIVQSMLEHRGGHSV
KVATYPFPAVDGLPEGVENFGSAATPEQSMCIMVATKSEALTRPVHETLIRSQSPDAVVTDMTFLWNSGIAAELG
VPCVVFSVMGAFSMLAMHHLEDAGVDRDDQDDDDDDDAVVEVPGFPGPPIRIPRTELPGFLRRPDYSITNLFISL
KAANCFGLAMNTSSELEKQYCELYTTPPEEGGGGLRRAYFLGPLALALPPPISSSSSSSSDCCSIMAWLDSKPSR
SVVYVSFGSMAHVKDVQLDELALGLETSGISFLWVVRGREEWSPPKGWEARVQDRGFIIRAWAPQISILGHHAAG
AFVTQCGWNSVLETVAAAVPMLTWPLAFEQFITERLVTDVLGIGVRLWPDGAGLRSESYQEHEVIPRQDVARALV
EFMRPAAGGPSSIRDMARTKLMDLSAKLHAAVAQGGSSHRDLHRLVDDLLMEAAAKRPRT
>AsGT3i21
MASTTTATRSSSSSSRSKKLRVLLIPFFATSHIEPFTDLAIRLAAAGSPSVAVEATVAVTPANVSIVQSLLERHY
GRQHDAAAESTIPVKIATYPFPAVDGLPRGVENLGRAAAADSWRIDVAAFSDTLMRPVQEALVREQAPDALVTDV
HFVWNVRVAAELGVPCVTFKVVGAFSSLAMRHLALVADVASSDPDVAVVPRFPGLPVRIPRTELPEFLRKKQEVD
YSTTNTFYAAQAACFGLAVNTSSDLEQEYCELHMREGYVKRAYFIGPVSLRPSPSLDAVGDSQHCVDWLDSKPAR
SVVYLCFGSFAPVSEAQLQELALGLEASGESFLWVVRSQEWTPPEGWEERVGDRGMVVTAWAPQTAILGHPAVGA
FVTHCGWNSVLETVAAGVPVLTWPMVFEQFITERLVTDVLGIGQRLWPHGAGIRSTRHIENEIVPAEAVARALMA
FMCPGGPGDSARNRVMRLAAKAHAAMAEGGSSHRDLRRLVDDLLEARGAAVAGGPKSVQG
>AsGT17424

MASTTTATRSSSSSSRSKKLRVLLIPFFATSHIEPFTDLAIRLAAAGSPSVAVEATVAVTPANVSIVQSLLERHY
GRQHDAAAESTIPVKIATYPFPAVDGLPRGVENLGRAAAADSWRIDVAAFSDTLMRPVQEALVREQAPDALVTDV
HFVWNVRVAAELGVPCVTFKVVGAFSSLAMRHLALVADVASSDPDVAVVPRFPGPPVRIPRTELPEFLRKKQEVD
YSTTNTFYAAQAACFGLAVNTSSDLEQEYCELHMREGYVKRAYFIGPVSLRPSPSLDAVGDSQHCVDWLDSKPAR
SVVYLCFGSFAPVSEAQLQELALGLEASGESFLWVVRSQEWTPPEGWEERVGDRGMVVTAWAPQTAILGHPAVGA
FVTHCGWNSVLETVAAGVPVLTWPMVFEQFITERLVTDVLGIGQRLWPHGAGIRSTRHIENEIVPAEAVARALMA
FMCPGGPGDSARNRVMRLAAKAHAAMAEGGSSHRDLRRLVDDLLEARGAAVAGGPKSVQG
>AsGT03295
NHDSLLAAGPPLIFVVDPLDHSKTRKTRNPEMASAEHGKKLRIMLIPFFATSHIGPHTDLAVRLAAARPDAVEMT
MAVTPANVSVVQAALQRHGASSASSAVKIATYPFPDVEGLPPGVENLTAAAGDAWRVDAAAMDEALTRPVQEALV
RELSPDAVFTDFHFFWNSIVAAELGVPCVTFSVIGPFSSLALRHLDGGGSDIQEMVVPGFPGPEIRIPRAELPEF
LRCQQKHDRFNPGPGVQGISRCFGIAFNTFFDMEQQYCELYARSADVKRAYFVGPVSLPLPPAGANAGESPCVSW
LGSMPSCSVVYVCFGTYASISDDQLRELALGLEASGKPFLWVLRADGWVPPAGWEERVEKRGMLVRGWAPQSEIL
SLSSIGAFLTHCGSSSLLEAAAAGVPMLTWPLVFDQFIEERLVTDVLKIGERVWSGARSTRYEEQVTVPAEAVAL
AVTRFLEPGGTGEAARVRAQELAVKAHAAVAEGGSSHSDICRLINDLIEAKEASGGGDTIVAGDVETSVVAAAL
>AsGT16525
MAPRPTVVLIPSWGSGHFMSALEAGKRLLAASGGAFSLTVLVMHAPTQAKASEVEGHVRREAASGLDIRFLQLPA
VEHPIDCVDPVEFASRYTQLHGPHVKAAIASLAPSFRAAAVVVDLFLTALFDVAHELAVPAYVYFASPAAFLALM
LRLPALREDLTGPGFEEMEGTVDVPGLPPVPPSYMPDCLVRRKIQSYDWFEYHGRRFMEARGLIVNTTIELEASV
LAAIADGRCVPGCPVPALHAIGPVIWFDPKDDDQRRHECMRWLDDQPPASVVFLCFGSMGSLDAAQVREVAAGLE
RSGHRFLWVLRGPPVAGTRFPTDANLDELLPEGFLDATAGRGLVWPAWAPQREILSHAAVGGFVTHCGWNSVLES
LWFGVPLVPWPLYGEQHLNAFELVAGVGVAVALEMDRKKGFFVEAAELERVVRSLMNGGSEEGRKARTKASETSA
EFRRAIGEGGSSCAALQRLVGEILDLPVGR
>AsGT22538
MSMLEAGKRLLARSGGALSLTVLVMQAPTESYRSDVAAHVRREEASGLDIRFRHLPAVEPPTEYVGGEDFVSRFV
ELHAAHVKAAISGLTCPVAALVIDFFGTTLLDASRELAVPCYVYFTASAAMYALFLRLPRLHEEVTVEFEEMEGN
VDVPGLPPVPPSALPTPMMDRTKPRYAWFVYHGRRFMEARGVIVNTAGELERSVLAAIADGRCTQGVPAPTVYPI
GPVVALNPPADPPHECVRWLDAQPPASVVFLCFGSGGFSTATQAHEIARGLERSGRRFLWVLRGQPAAGAGQPSD
ANLAELLPEGFLEKTKEKGMVWPTKAPQKEILAHAAVGGFVTHGGWNSVLESLWFGVPMVPWPLYAEQHLNAFTM
VAYLGVAVAMEVDRKRNNFVDASELERAVKTLMDGDSADGRKVREKCAEMKAVCRKAVEEGGSSYAALCRLSEEM
RKGAVRASK
>AsGT27009
MAAPTVVLLPVWAAGHLMSMLDAGKRLLARSGGALSLTVLVMQAPTEKDRSEVAAHIRREEASGLDIHFHHLPAV
EPPRDYVGIEEFVSRFIELHADHVKTAISALTCPVAAVVLDFFGTTLFDVCRELSVPAYVYFTADAAFYALLLRL
PALHEEVTVEFGEMEGMVDVPGLPPVPPSCLPVPVMDKKNPNYTWFVYHGRRFMEGDGVIINTVAELEQSVLAAI
ADGRCTRGIPAPTVYPVGPVLSLSPSPQGAEQPHECVRWLDAQPPASVVLLCFGSVGFFTMTQAHEVARGLERSG
HRFLWVLRGPPAAGTLQTTDANLGELLPEGFLERTKEKGLVWPTKAPQKEILAHAAVGGFVTHGGWNSTLESLWF
GVPMVPWPLYAEQHLNAFTLVEYMGAAVVMEVDRKRNNFVDASELERAVKALMDGDSDDGRKVREKATEMKAACR
KAVEEGGSSYSALGRLSEDIRKGAVRTSQ
>AsGT28561
MVTPTVVLLPIWGAGHLMSMLDAGKRLLARSGGALSLTVLVMQPPTENDRSEVAGHIRREEASGLDIRFHHLPAV
EPPTDWVGVEEFVSRFEELHADHVKAAISALTCPVAAVVLDLFGTTLLDVCRKLDVPAYVYFTANAASFALMLHL
SALHEEVTAEFEDMEGMVYLPGLPPVPPSSFPLPVMDKKNPNYTWFVYHGRRFMEADGIIINTASVLEQSVLAAI
ADGWCTHGVPAPTVYPVGPVLSLSMSSPAEQPHECVRWLDAQPPASVVFLCFGSGGFSTATQAHETAHGLERSGH
RFLWVLHGPPVAGTQQPSDANIRELLPEGFLDRTKGKGLVWPTKAPQKEILAHAAVGGFVTHGGWNSILESLWTL
WYGVPMVPWPLYAEQYFNAFTVVTYMGLAVAMEVDRKRNNFVHASELERAVKALMDSDSDDGRKAREKATEMKAA
CRKAVEEGGSSYFALGRLSEDTRKGAVRSRQ
>AsGT00931
MAPSYPSAHPTDATMSGDGDDARAPHVVLLSSPGMGHVVPVAELARRLHAEHGFTATVFTYASSDSAAQRAVLAS
LPPAVGSASLPEVPLDDLVAAGAAIETLLSVEAQRAVPALTALLADIGKASNLVAFVADLFGADTLPAARAAGVP
GYLFFPSNLLMLSLMLHLPRLDAEIEGEFRDLREPVRLPGCVAVPGADILGPLQDRTSDAYRWMLHHGERYRDAD
GILVNTFDAIEPGAAAVLRRQEPWRPPVYPVGPVIRQPDDGEDATGCIEWLDAQPDRSVLFVSFGSGGALPAAQM
DELARGLELSGQRFLWVVRSPTDSGADPGANYYDGSKSKDYPLRFLPFGFLERTKEVGLVVPSWAPQVRVLGHRA

TRAMLTHCGWNSVLESVMHGVPIIAWPLYAEQRENAVMLHEETKVALRPKVRGAQGLILGEDIAEVVNDMMNGKE
GEAARNKVVKLQEAARSGLTPSGISHETLTELVSKWKNT
>AsGT06751
FCQSKLILPDHDRSGAGACPHVVLLTSPGAGHVLPVAEFATRLAADHDFTATIVTYTNLSSHAHNSPLASLPPRV
SVAALPEVPLDDLPADAHILTRIFAVVNRTLPHLRDLLRSLLGSPSGSGVAAFLTDMLCPAALAVARELGVPRYV
FFTSSLTSLLSLLYTPELARTTTGDCRYLPEPVVLPGCVPLHGADLVEPVQNRSDPVYPLMVDLGHNYLLADGFI
VNTFDAMEHDTLVAFKELSDRGVYPPVYAVGPFIRSVSAGAGNKHSCLRWLDDQPDGSVLYVSFGSGGTLSTEQT
AELAAGLEASEQKFLWVVHFPSDKDRSAGYFGTSADQGNDDPLSYLPEGLVERTSGTGLAVPLWAPQVEILNHPA
VGGFLSHCGWNSTLESAAAGVPMVAWPLFAEQRVNAVLVSSERLGLALWERPSSVDNDGVVVPRGEVTELARELM
VGKKGALARNKAGQVRDGAQMALAPGGPQRRALAAVISKWNARP
>AsGT08700
MESLTSIGAERAASTEQPRPHVVLLASPGAGHLIPLAELARRLVEHHGFEATLVTFTDLISSPEALSGIPASVST
AALPSVPLHDLPAATPIETVLFELVRRSLPHLRTLLRSAAHLAALVPDFFCSAALPLAAELGVPAYVFIPTNLTS
LTVMRRTVDLHDGVPPGEHRDLPDPLELPGGVSLRRADLPHAFRSSHEPVYAYLLEEGRRYRHADGFLMNTFYEM
EPAAVEEFSQATTEGAWPPVFPVGPFVRSGSDETGESACIDWLDRQPTGSVVYVSFGSGGSLSVKQTAELAAGLE
ASSHRFLWVVRMPSMDGEVEEHRGDKNNNPLAWLPEGFLERTRDKGLAVAAWAPQVRVLSHPATAAFVSHCGWNS
TLESVSCGVPMIAWPLYAEQRLNAVLLEGSVGVALRVGTRARDADGDGMVVTRGEIADVVRELMEGAEKGRAVRR
QAGDMQQAAARAWAPEGSSRRALEKVAVTWKAWRGWRGEVTDVPASGER
>AsGT10772
MEEANGVCNGTVESRRPPHVAMLVTPGMGHLIPLAELAKRLAARHGVTATLITFASTASATQRAFLATLPPSISA
VSLPPVDLSDLAPDAAIETLMSEECVRLVPALTGILSGLRETTRLVAFVADLFGADSFDAAVAAGVPRRYLFFPT
NLHGLTLILHLPDLDVSLPGEFRDLAEPVRLPGCVPIPGTDILSPLQDKSNPSYRWMVHHGKRYRDADAILVNSF
DAVEPDAIAALRVPEQGRPPVYNIGPLTQTNAASPAPRAACLEWLDRQPPKSVIFVSFGSGGSLPREQMHELALG
LEQSGQRFLWVVRSPSDEGAVNANYYDAESKKDPLAYLPDGFVERTKDVGLLVPSWAPQIKVLAHEATGGFLVHC
GWNSVLESLVHGVPMVAWPLFAEQRQNAVMLTEGVGAAIRVPETKKKEEIAAVVREVMVGQGKGAEVRAKVAELQ
KAAEEGLGEGGAATSALDGVVHKWIGGEEK
>AsGT16496
MAATEGTAAPEPRRSTHVVLLVCPGTGHILPAAELARRIVAMDGFTVTLVSHTNFSSVEQYSSTFASLPPSVSTA
LLPEVSLADLPADARLETLVMTVVQRALPHLRDLLRSLLDSPTGVAALLPDQLCPWALEVAAELGIPGYLFCSTN
LMALSTMLHAPELDRTTTGEFRDLPELIRLPGCVPLHGAELLDTVQDRTNPAYALIVELGKLYLLAEGVLVNTLD
AMEHETEVAMKGLSDRGVYPPVYVVGPVTRRADGSCNHSSLRWLDQQQDGSVLYVCFGSGGTLSMEQTAELAAGL
EASQQKFLWVVQFVNNKDKYGSYFGGGGGHGDNSDSPVNYLPQGFVERTKGRGLVVPLWAPQVEILSHPAVGGFV
SHCGWNSTLEAVAAGVPTIAWPLFAEQRMNAKMLSSERAGMALQLKAVRAGLVRRVSPRPALTKAREEDGVVTRD
ELAAVARELMAGEKGASARKKVRELREELGKSLAPDGPSHKALEAVAGKWQAGGNGTTAA
>AsGT17930
MESLTSTGSAPAPSAARPHVVLLASPDAGHLIPLGELARRLVEHHGFAATLITFNSLSDPQALPSSLPSSVATAA
LPGVQIDDLPANAVRGTVLVEIIRRSLPSLRTLLLSISSTTPIAALVPDFLCFTALPLAAELGVPAYIFFTTNMN
FLYLTYRLVVLHDDASPGEYRDLPEPLEMPGGLSLRRAELPEGHRSSKAPVYARLLEGGQRYRLADGFLANTFNE
MEPASVEAFKQLADRGAFPPVFPVGPLVRSNSEEEASAGASPLLEWLHHQPARSVVYVSFGSGGALSVEQTAVLA
AGLEASGHRFLWVVRMPSLDGRSHAFGAGGDGNDPLAWLPEGFMQRTEGRGLAVAGWAPQVRVLSHPATAAFVSH
CGWNSTLESAVAGVPMVAWPLHTEQTMNALLLEESLGVALRPRAREDDGVVGREEVAAAVKELLEGENGRAVRRR
AEELQQAAARALRPEGSSCRALEDVAARWKAALIRERRSC
>AsGT01315
MESLTSIGAERAASTEQPRPHVVLLASPGAGHLIPLAELARRLVEHHGFEATLVTFTDLISSPEALSGIPASVST
AALPSVPLHDLPAATPIETVLFELVRRSLPHLRTLLRSAAHLAALVPDFFCSAALPLAAELGVPAYVFIPTNLTS
LTVMRRTVDLHDGVPPGEHRDLPDPLELPGGVSLRRADLPHAFRSSHEPVYAYLLEEGRRYRHADGFLMNTFYEM
EPAAVEEFSQATTEGAWPPVFPVGPFVRSGSDETGESACIDWLDRQPTGSVVYVSFGSGGSLSVKQTAELAAGLE
ASSHRFLWVVRMPSMDGEVEEHRGDKNNNPLAWLPEGFLERTRDKGLAVAAWAPQVRVLSHPATAAFVSHCGWNS
TLESVSCGVPMIAWPLYAEQRLNAVLLEGSVGVALRVGTRARDADGDGMVVTRGEIADVVRELMEGAEKGRAVRR
QAGDMQQAAARAWAPEGSSRRALEKVAVTWKAWRGWRGEVTDVPASGER
>AsGT28947A
MATEGTPAAPERSTHVVLLACPGTGHVLPVAELARRVAGLDGFTATLVTFTNFSSGDQYSSTFASLPPSVSTAVL
PEVSIDDLPADARAETRIIAVVRRALPHLRDLLRSLLDSPAGVAAFVPDLFCPWALEVSAELGVPGYVFCPTNLM

ALSTILHTPELDRTTTCEFRDLPDLIRLPGCVPLHGSELLDTVQVRADPAYALIVDLGKLYLLARGFLVNTFDAM
EHETVAAFEELSDEGVYPPAYAVGPFIRPRSEKSSEHSSLRWLDEQPAGSVLYVCFGSGGTLSREQTDELAAGLE
ASGQRFLWVVQFPSDKDKSGAFFGDGGRGDGDSPVKYLPEGFVERTEGTGLAVPFWAPQVEVLSHPAVGGFVSHC
GWNSTLEAVAAGVPTVAWPLYAEQRMNAKMLSERAGMALRLEAREDGVVTRDEVAAVARELMAGEKGASARKKAR
QLREEMAKGLAPDGRSRMAFEAVAGRWKAAGNGTAD
>AsGT28947B
MAIEGTAAAERSTHVVLMACPGTGHVLPVAELARRVVEVDGFTATLVRFTNFSSGDQYSATFASLPPSVSTALLP
EVSIEDLPADARAETRIIAVVKRALPHLRDLLQSLVDAPAGVAAFVPDLFCPWALEVAAELGIPGYIFCPSNLMA
LSTVLHTPELDRTTTCEFRDLPGLIRLPGCVPLHGADLLDTVQDRANPAYALIVDLAKLYLLAQGFLVNTFDAME
HETIVAFKDLSDKGVYPPAYAVGPFIRASSDKSSKHSSLRWLDDQPDGSVLYVCFGSGGTLSTEQTSELAAGLEA
SGQRFLWVVQFPSDNDKCASFFGGGRGDGDSPINYLPEGFTERTEGTGLAVPLWAPQVEILSHRAVGGFVSHCGW
NSTLEAVAAGVPTIAWPLYAEQRMNAKMLSERAGMSLWPKAREDGIVTRDEVETAARELMTGKKGVAARKKAREL
REEMEKGLAPEGASRKAFEDVAGRWKAAGNGTAA
>AsGT20n10
MGVATISAVPRKLAVLYPSPGMGHIVSMIELGKIFVARGLAVTIVVIDLPNNTGSSATGPFLAGVSAANPSISFH
RLPQVKLPHVESRHIETLNFEVARAANPHLRDFLAGISPDIFIADFFCHVARDVASELGIPFYFFFTSGAEVLAV
LLHLPVLHSQSTASFQDMGEELVHVPGIPSFPASHSMLPVMDRDDAAYMAFVNVCSDLCRSQGIIVNTFSSFEPR
AIEAIAAGLCTPAGLPIPALHCIGPLIKSEEVGVKRGDECMAWLDTQPKDSVVFLCFGSLGRFSGKQIREVALGL
EASGQRFLWVVKSPPNDDPAKKFENPSEKPDLDALLPEGFLDRTKDKGLVVKSWAPQRDVLMHAAVGGFVTHCGW
NSVLESVMAGVPMLAWPLYAEQRMNKVFLEEELGLAVAVEGYDKEVVEAREVAAKVKWMMDSDGGRVIRERTQAA
MRQAKKGMGEGGESEVTLAGLVDAWTTHA
>AsGT01092
MDGTPTTTTMGPRKPVVLYPSPGMGHLVSMIELGKLFTARGLVITIVVIDPPNNTSATGPFLAGVSAANPSISFH
RLPQVKLGESERPEMLTFQLARLSNPHLHDFLAGASPAILVVDFFCSAAMDVAAELGIPAYFFCTSGAQVLAYFM
HLTVLHGKSTVSLREMGEELVHAPGIPSFPASHSIQPLMNRDGPAYKELLSVSAKLFGSQGIIVNTFRSLEPRAV
DTIVAGLCTPPGLPTPPVYCIGPLIKSEEVGVKRGDECLAWLDTQPKASVVFLCFGSLGRFSAKQTREVAAGLEA
SGQRFLWVVRSPPGEDDPAKTKFEKPAEPDLDALLPEGFLDRTKGWGLVVKSWAPQRDVLAHDAVGCFVTHCGWN
SVLESVMAGVPMLAWPLYAEQKMNKVFLEKEMRLAVAMEGYDKELVEAGEVAKKVRWVMDSEGGRALRERTLTAM
RQANDALLEGGESEATLAGLVDAWVHA
>AsGT05602
MTEKTVVLYPTLGVGHLNPMVQLAKAFLRRGVAVTIAVVDPPGKDPVLEAAVARLASACPSMTVRLLPIPPARAD
KPYSHPAMRMLDELSLASPVLREFLGSLPAVDALVVDMFCIDALDVAAELAVPAYIFYPSAAGDLAIYLQVPDCC
RAAPSSFKDMGKTALRFSGVPPVRALDMPDTMLDRESELCRRRVRQLARMPEARGILVNSFEELEPRALKALRDG
LCVPAGRSTPEIYCVGPLVDGGVSVDGESGERHACLEWLDSQPIKSVVFLCFGSRGVFSAAQLRETARGLEESGQ
RFLWTVRSPREEERSKFAEPDLEAFLPDGFLQRTVGRGLVVKNWAPQAEVVQHEAVGAFVTHCGWNSALEAIMAG
VPMICWPLYAEQRLNKVHLVEEMRIGVVVEGYEEELVKAEEVEAKVRLVMQSEEGKKLRDRATMAKEMAADAVKE
GGSSDVAFYSFLKHLEMRKLEQGLIRRSENP
>AsGT06492
MDGHATTRAARKQVVLYPSPGMGHLVSMIELGKLFAARGLAVTILIVKLPFEDTGARGPFLAGVTAANPSITFHS
FPQLDLPPVQSDHPEAITFEVARVSNPHLRDFLASASPAVLVVDFFCSVALDVAAELGIPGYCFFTSGAEALCFF
LYLPVLHAQTTASFRDMGEELVHVPGISSFPASHAIKPIMHRDDAAYRGFLSVSPYLCRSQGIIVNTSRSLEPRA
VDAIGAGLCTPPGLPTPPVYCIGPLIKSEEVGVTKHGEGCLAWLDSQPKDSVVFLCFGSLGVFGAKQIREIAVGL
EASGKRFLWVVRSPPSDDPARKFEKTPEPDLDALLPEGFLDRTKGRGLVVKSWAPQRDVLAHDAVGGFVTHCGWN
SILESVMAGVPMLAWPLYAEQRINRVFLEKELGLAVAVEGYDRDVVEADEVASKVSWMMDSDGGRVLRERTLEAM
RQAKEALREGGESDVTLARLVEGWTLA
>AsGT10189
MPAPQKTFVLYPSLGVGHLNPMVELAKHLLRHGHNAVIAVVDPPDTDAVSAGAVARLAAANPSIAFRLLPVPTSP
DPAAHPVKRSLDTLRLANPVLRDFLRSLPAHAIFLDMFCVDALDVAAELGVPAYFFFASAASALAVLLNLPYTYP
ELPSSFKDMGEALVRTPGMPPIRAVDMPLMMQDRESVTTKVRLHQFRRIPEGRGVLVNTFDWLEPTALRALADGV
CVRGRPTPRVYCIGPLVDGGGNEKKRHGHECLAWLDAQPRHSVVFLCFGSNGAFSAAQLKEIARGLESSGHRFLW
AVRSPPEEQGEFPEPDLERLLPAGFLERTRDRGTVVGSWVPQAEVVRHEAVGAFVTHCGWNSTLEAIMAGLPMVC
WPLYAEQAMNKVFMVEEMKIAVALEGYEAGIVKAEEVEAKVRLVMETEEGRKLRNTLVVAKKMALDAIGQGGSSQ
AAFADFLADMENSSS

>AsGT25n16
MAGMAPLAKTFVLYPSLGVGHLNPMVELAKFLVRQGHNVIVAVADPPDSDAVSADAVARLSAANPCIDFRRLPAP
PNPDPAAHPLQRILDTLRLCNPVLRDFLRSLPGPGAHALLLDMFCVHALDVAAELALPAYFFFVSPAGALAVLLN
LPHSYPEMPSFKDMGHQALVRFPGMPPFRAVDMPQGMHDKDSDLTKGLLYQFSRIPEGRGVLVNTLDWLEPTALR
ALGDGVCVPGRPTPPVFCIGLLVDGGYGEKSRPDGGANKCLAWLDKQPHRSVVFLCFGSQGAFSAAQLKEIALGL
ESSGHRFLWAVRSPPEQQGEPDLEGLLPAGFLERTRDRGMVLADWVPQAQVLRHEAVGAFVTHGGWNSAMEAIMS
GLPMICWPLYAEQALNKVFMVDEMKIAVEVAGYEEGMVKAEEVEAKVRLLMETEEGRKLREMLVVARKMALDANA
KGGSSQVAFAKFLCDLENSTST
>AsGT17673
MATARRVVLIPGQSISHLNPMMEFAAVCLRRGLAVTVVVPNPTLTAPALSSATGRYASQFPSLAVHSLPLPPSHQ
QSLGAVHPFIRMQAAFRSQAPGLRNFLRSLPAVHALVADMFAVYALDVAAEVGVPGHLFYCTGANNLAVFLQLPS
FCSGSDRDLKDLGDAPVSFPGVRTMPASHLVDGVLDSGTDLYAPVLEVFGRMAAARGILVNTFEALEGSVMAALR
SGSCLPDGANLSVYCVGPLIAEGEKEEERHPCLPWLDAQPERSVVFICFGSRCTVSLEQIREMAEGLDKSGHRFL
WVLRAPPGIAAAEPDAALSLLPEGFLTRTADRGLVVTASWVPQMDVLRHSSTGAFITHCGWNSTLEAVVTGMPMV
CWPLEAEQWMNKVYIVEEMKVGIEVRGYKPGALVTADNVDATVRQIMDLESEARRTVMKQAMAVKESAAAAWKED
GSSCAAFTEFVKQMG
>AsGt21862
MEKTVVLYPGFIVSHFVPMMRLAGVLIQHGYAVSVALIDPAVNLDDAFAAAVARAAASMPSVTFHRIPLVDGPLT
PGAHFILNYLDIISRHNDRFHDFLLRSSSSSRGGVSAVVVDSMSVEAFRVTRRLGIPGYVMFTSNAAALAAFVQL
PSVHAEARTSLKFKELRDAPLDFFGIAPMPASHLFGDIFDHPESDICKATVASLIGISEADGILVNTFESLDAPA
VATLGDPRCLPGRVMPPVYCVGPFVGGVGVGVQTSKQQHECLSWLDGQPDHSVVFLCFGSAGHNHPEEQVKEIAV
GLENSGHRFLWVVREPDANAVLPTGFLERTSGRGLVVKNWVPQADVLRHTATAVFVTHCGWNSVLEGVTAGVPML
CWPLYAEQKLNKMRMVGEMGIAAEMVVSQQGLVEAAEVERKVRLAMESEEGRELRTRVAAHREAAAVAWDDGGSS
RAAFVRFLSDIQSRRPQPARSGEGK
>AsGT23340
MSTARRVVVVPGQSISHLIPMLEFAAICLDRGLAVTVVVPDPSLTAPAFRSTVCRYASRLPSLSVHSLPPPPAQQ
QSLDAASAPAHPFIRMQAASRSQAPGLRDFLRSLPAVHALVADMLDVVCAVDVAAEVGIPGHLFFSTGATTLSVF
LGLPLFCSRSDRGLKDLGDAPVVFPGAPKMPASHLPDAVLDRGTDLYATSLDVFGRMAAPSGILVNTFEALESSA
LEALRDGRCLPDRATPPVYCIGPLFAEAEKGEERHPCLPWLDAQPERSVVFLCFGSRYTVSQEQIGEMAKGLEVS
GHRFLWVLRAPPAFAAAAGEPDAALSLLPDGFLVRTADRGLVVTASSSWVPQVEVLRHASTGAFVTHCGWNSTLE
GVAAGVPMVCWPLVAEQWMNKVYIVEEMKVGVEVRGNKRGGLVTADDVDAAVRQVMDMEPERRREMEERLMEMKE
SSAAAWKEGGSSRTAFAEFVKQME
>AsGT23586
MKQTVVLYPGAGGSHVAAMTELANVFLKHGYDVTMVLVEPPFKSSDSGATAIERIAASNPSISFHVLPPLPPPDF
AAAGNKNPFVLMFQLLLEYNELLEAFLRSIPRKRLHSVVLDMFCIHALDVCVKLGVPVYTFFASGASCLSVLTQF
PALIAGRQTGLKEIGDTPLDFLGVPPMPASHIIKELLEHPEDEMCKILTNMWKRNTETMGVLVNTFESLESRAVQ
SLRDPLCVPGRILPPIYCVGPLVGEGAKDGDGAERNECLAWLDSQPDRSVVFLCFGSKGTVSAEQLKEIAVGLER
SGQRFLWSVRTPAGSQDAKKYLEVRAEPDLDELMPEGFLERTKDKGLVIKSWAPQVDVLRHRATGAFVTHCGWNS
VLEAVSAGVPMLCWPLEAEQKMNKVCMTEDMGVAVELDGYMAGFVKADEVEAKVRLVIEGEDGRQLRARVAARKE
EAEAALEEGGASRAAFVQFLLDVENIGEQVRE
>AsGT23818
MKKTVVLYPGLGVGHLVPMVELAKVFLKHGVAVIVALIDPGAKSTDFSAAVARAKTSNPSVDFHVLPPPPPPPAP
VDSNSQQVAASTHHVTKIFKLLTAMNAPLRDFLRSLPSVDALILDMFCADAQDVAAELELPVYYFYASGAADLAI
FLNLPSRVAGNTSKMKELGDSVITFPGVPPFKASDLPNEVIGDGEALPAIVGMFDRMSRADGILVNSFDSLETRA
VQALRDDGLCVPGRATPPVYCVGPLVSGGGGGEKEHECLRWLDAQPDRSVVFLSFGSMGTFSGKQLLEIAVGLEK
SGERFLWVVRTPRSPDFSYGGPLPEPDLDALMPEGFLERTKDRGLVVKSWAPQVEVLRHRATGAFVTHCGWNSTL
EGITAGLPLLCWPLYAEQRVNKVHIVEEMKLGVEMRGYNQEVVKAEEVEEKVRWVMASEGGEALRERVAAAKEGA
AEALKEGGSSHLAFVQFLKDLDTATVQD
>AsGT27586
MKQTVVLYPGAGVSHVGPMTELANVFLKHGYDVTMVLVEPPFKSSDSGATAIERIAASNPSISFHVLPPLPPPDF
AAAGNKNPFVLMFQLLLEYNELLEAFLRSIPRKRLHSVVLDMFCIHALDVCVKLGVPVYTFFASGASCLSVLTQF
PALIAGRQTGLKEIGDTPLDFLGVPPMPASHLIKEMLEHPEEELCKEITNIWKRNTETMGVLVNTFESLESRAVQ
SLRDPLCVPGRTLPPIYCVGPLVGEGAKDGGGTERNKCLAWLDSQPDRSVVFLCFGSKGTLSAEQLKEIAVGLER

SGQRFLWSVRTPAGSDDAMKYLEVRPEPDLDALMPEGFLERTKDRGLVVKSWAPQVDVLRHRATGAFVTHCGWNS
VLEAVSAGVPMLCWPLEAEQKMNKVCMTEDMGVAVELDGYRTGFVKADEVEAKVRLVLESEEGRQLRAR
>AsGT00260
MGSMEQKKPHAVCVPFPAQGHITPMLKVAKLLHARGFHVTFVLTDYNYSRLLRSRGAAAFHGCPGFDFASIPDGL
PPSDAEATQDIPALCRSTMTTCLPHVSALLATLNGPASAVPPVTCLLCDACMSFAYDAAKEIGLPCAGLWTASGC
GFMAYNYYKNLVEQGLVPLKDEAQRTDGYLDTVVHGVPGVCDGFQLRDFPDFIRTTDSDDIMLNFLIRETARAAS
LPDAVIINSFDDLEQRELDAMRAILPPVCALGPLLLHVRRLVPEGSPLDVAVQSNLWKEQDGLLEWLDSHPPRSV
VYVNYGSITVMTNEQMLEFAWGLANSGYPFLWNVRPDLVKGDAAVLPPEFSAAIEGRGLLTTWCPQEKVIVQDAV
GVFLTHSGWNSTLESLCAGVPMLSWPFFAEQQTNCRYKRTEWGVGMEIGGEVRRAEVAAKIQEAMEGEKGKEMRR
RAAEWKEKAARATLPGGAAEANLDKLIDVLHGKTGQAVNASTGRN
>AsGT02132
GASNEAPGSPIICEHSPAQGHVTPMLHLAKALHARGFHVTFVNSEYNHRRLLRSRGPGSLDGADGFRFEAVPDGL
PPSDDDDVTQDIAALCLSTTKHSAAPFRELLARLNSSSTPGTPPVSCVIADGVMSFAQRVAEDMGILALVFWTTS
ACGFMGYLQFAELVRRGYVPLKDESDLTNGYLDTAIDWIPGMEGVRLKDIPSFIRTTDPDDVMLNYDGGEAQNAR
GARGLILNTYDALEQDVVDALRREFPRVYTVGPLPAFTPRGAELDAIGGNLWKEDASCLRWLDAQRQPGSVVYVN
FGSITVVTPAQLTEFAWGLASCGRPFLWVVRPDLVSGEKAVLPEEFVRGTKDRGVLVSWCPQELVLSHPSVGLFL
THCGWNSTLESVCAGVPMVCWPFFAEQPTNCRYACAKWGIGMEIGSGVAREEVARLVREAMDGERGKAMRVSALA
WKGRARAATVDGGTSSENMDRLIEFLSAGCV
>AsGT05740
MGSMGAEEKPHAVCLPYPAQGHITPMLNVAKLLHARGFHVTFVNSEYNHARLLRTRGAAALSGVPGFCFATIPDG
MPVIDDDDVTQDIPALCKSTTETCLEPFRRLLAEINGSASADGHPPVTCVISDVIMGFSMDAAKELGLPYVQLWT
ASAVSYLAYHHYRLLIDRGIFPLKDAKQLTDGYLDTPVEDMPGLRSMRVRDFPSFIRSTDPDEFMVGYAIKETGR
AASASAVIINSFADLEGQEVEAMETLLGLPKVYMLGPLPLVAAPHSTAIISGLSLWKEEAECLQWLHGKEAGSVV
YVNFGSIVVMTNEQMVEFAWGLANSGRQFIWIIRRDLVRGDAAVLPPEFLAATAERGFMASWCPQQQVLNHPAVG
AFLTHSGWNSALESICGGVPVISWPFFADQQTNCRYQCNEWGVGMEIDSNVQRDEVSDLITELMEGEKGKAMSKR
AEEWREKAVKAAQPDGSSHRNFDGLVRDVLLAKH
>AsGT8i4
MGSLPPGKRPHAVMIPYPAQGHITPMLQLAKLLHARGFHVTFVNNEFNHRRLLRSQGADTLGGVPAFRFAAIEDG
LPPSDPDATQDIPALCHSTMTTSLPRFKDLVAKLNGEAEASGGALPPVTCVVADSTMTFALRAARELGLRCATLW
TASACGFMCYFHYKDLRDRGIIPLKDEAQLSNGYLDDTTVDWIPVVPKDMRLRDLPSFVRTTDPDDIMFNFFIHE
TAGMSQASGVVINTFDELDAPLLDAMSRLLPRVYTAGPLHLTVRNNVPDESPAAALDSSLWKAQDAPLRWLDGRP
PRSVVYVNFGSITVMSNEQLLEFAWGLANTGYAFLWNVRPDLVKGDEAALPPEFSAETEGRSMLSTWCPQAKVLQ
HEAVGVFLTHSGWNSTLESISGGVPMVCWPFFAEQQTNCRYACTEWGVGMEIGDDVRRAQVEGMIREVMEGEKGR
EMRRRVTELRDCAVASAGRDGRSMRNVDRLINEVLLA
>AsGT16h6
MASRQYHVVMVPYPAQSHVAPLMQLARLLHARGAHVTFVHTQFNYRRLVDAKGEAAVRPSSSTGFCVEVIDDGLS
LSVQQHDVAAVVDALRRNCQGPFRALLRKLSSAMPPVTTVVADTVMTFAATEAREAGIPDVGFFTASACGLMGYF
QFGELIKRGLVPLQDASCLATPLHWVPGMNHMRLKDMPSFCHTTDPDDTMVAATLEQMNTALGAKAIVLNTFYEL
EKDVVDGLAAFFPPLYTVGPLAEVDSGGSDSLLGAIDISIWQEDAQCLAWLDDKKASSVVYVNFGSIHVMTAAQL
REFALGLASCGFPFLWIKRPDVVVDGEEDAVLPEEFLAAVARGAGLVVPWCAQPAVLKHPAVGLFVTHCGWNSLL
EAAAAGMPLLCWPLFAEQTTNCRQVCECWGNGAEIPKEVEHGAVSALVREMMEGELGREKRAKAAEWKAAAQTAI
VEGGSSCRSVDRLVEDILLIPSQRK
>AsGT18035
MARPHAVVLPHPGSGNINPALQLAKLLHRHGFYITFVNTEHNHGRVQATQGAAAVRGREGFRFEAIPDGLVEADR
DRDAEDYDLRLSAATSNRCAQPLRELLLRLNGTPGVPPVTCLVPTALMSFALDVARELDVPSMVLWGCSAAALMA
DMRLSELHQRGYLPLKDESWLTNGHLAKTVIGWIPGMPPMSLGEMSSFVRTTNPEDFGLRFSIREAEGCTKAGAL
IINTFDNLEADVLEALRAEYPRIYTIGPLGNHQHLDGNAATDDDSSSSGLSLWKQDTGCLEWLDTQKPSSVVYAN
FGSITVLTVDQLAEFAWGLAASGHPFLWSVRDNLVPGAKAGLGSLPPEFVAATAGRCFLTTWCPQEQVLRHTAVG
CFLTHNGWNSTCESVAAGVPMVCWPGFADQYTNCKYVCQVWGVGLRLDDEVTREQVAGHVRQAMDSEEMRRSARG
WKAKAKAAAAPGGSSYENLQSMVTALASFSSE
>AsGT26167
MAPRPHAVVLPYPGSGNINPALQIAKLLHRRGVYVTFVNTEHNHRKAGADAVRGHDGFRFESIPDGLDEAARGVQ
DYGRGLAISTSTRCAAPLRDLIARLNATPGVPAVTCVVPTALMSFALGVARELGIPSMAFWGGSAASLMGHMRLL

ELQERGYLPLEDESCLTNGHLETTVIDWIPGMPPISLGDISSFLRTTDPDDFGLRFNESEANNCTKAGALILNTF
DDLEADVLAALRAEYPCIYTIGPLGSLLKEVIAAGDVEESSTSGLSLWKQDTECITWLATQEASSVVYVNFGSHT
VLTPEQLAEFAWGLADSGHILLWSIRENLVRGGGLAALPPEFVAATEGRCCLTTWCPQEQVLQHPAVGCFLTHNG
WNSTCESLAAGVPMVCWPGFADQFTNCKYACEVWGVGLRLDDEVRREQVAGHVRKAMESEEIRRSAAGWKAKAAE
AAAPGGSSYQNLQSLVNALNSETDHL
>AsGT10811
MAMPHAVVVPYPGSGNINPALQLAKLLHRHGVYISFVNTEHNHRRVQATVGAAAVRGREGFRFEAIPDGLVEADR
DAGDYDLGLSAATSHRCAQPLKELVRRLNGTAGVPPVTCVMPTALMSFALDVARELGLPSMVLWGGSAASLMGHM
RLRELQERGYLPLKDKSFLTNGHLDKTVIDWIPGMPPISLNDISSFVRTTDPDDFGLRFNILEANGCTKAGALIV
NTFQDLEADVLEALGAEYPRIYTIGPLGSLLSHHITDDDSSSTGGLSLWKQDTDCLAWLDKQEPSSVVYANFGSL
TVLTADQLVEFAWGLADSGHRFLWSMRDNLVPSDGAGLSSLPPEFIAATAGRCCLTTWCPQEEVLRHPAVGCFVT
HNGWNSTCESVASGVPMVCWPGFADQYTNCKYVCEVWGVGLRLDDVVRREQVAGHVKQAIEAEEMRRSAMGWKEK
AEDAAAPGGSSYENLQSMVTALGSVSA
>AsGT01670
MASVNGDGRRARRVLVFPLPFQGHINPMMQLADVLHSRGLGVTVLHTRFNALDPALHPEYAFVAVPDGIPADVAA
SGSIISIILAMNAAMEASGAVHDVLASVLADGPAAACLFIDANLLAVQKVAAALGLPTMVLRTGSAACFSCFLAY
PMLHQKGYLPPKESQLYTPVDELPPLRVRDLFFSSSNNHEMVRQVLARATETVRNSHGLVINTFEALETAELDMI
RRELEVAVVLAAGPLHKLSTRGNGSSLLQQDRTCIEWLDTQAAGSVLYVSFGSLASMDPGELSEVAWGLANSGQP
FLWVVRPDLVRGSDGSGLSEGFHRAVEGRAKVIPWAPQQEVLAHSAVGGFWTHNGWNSTLESISEGHPMICRPQF
ADQMMNTRYVEAAWGVGFELEGKLERKKIEEAIRNLMKGSQGELARERARELKKKVISCLESDGSSSLAIDKLIE
HMSSL
>AsGT04347
MGHNAALAVAGRRRHVLLFPLGYQGHINPMFRLAGILHARGFEVTVFHTHFNAPDPSRHPQYRFVTVPDGISGPA
PVAIEDVVSHIFALNAACEAPFRDLLAAILDEYSRDAVACLVADGHLLSMVQVATRLSVPTLVLRTGSAACFSCF
LAYPLLIDRGYLPSELETEVSELLPYRVRDLMRVGSHDLTRELLARVVAAVDASAGVILNTFDALERSELEKLRR
DLTLQVFDIGPLHMFSPVTESSLLRQDRSCLEWLDARPTASVLYVSFGSLACMSSRDLVETAWGIASSGVPFIWV
VRPGLVPASVEEKETTATRLPDGFEAATRGRGTVVAWAPQEEVLRHPAVAGFWTHCGWNSTTESVCEGVPMLCRP
HFGDQMGNARYVEHVWKVGFEVGGDLERSAVEAAIRRLVTGSDGAEMRARASELKKAAKECTGKAGSSDLAIECY
RCT
>AsGT04598
MPNPGSATSTQNATTAIDSMGQEEGNTATGGLRRRRVLVFPLPYQGHINPMFQLAGLLHARGFAITVFHAHFNAP
DPSRHPAYDFVPLPDGMSAACTDDTAEPAVIATIKNIFAVNTSCEASFRESLAALLEAPGARDDVACLIADAHLL
TLVDIARQQGVPTLALRTGSAACFRCFLVNPMLCDKGYLPAQESQLDTPVRELPPYRVMDLMAISRNREEHDLTC
KLLSRAVEAARTSTGFILNTFDALETDELATVRQDLGLPVFDVGPLHKISPAASSSILPQDRACLEWLDAQAPAS
VLYVSFGSLANMSGADLAEMAWGIANAGQPFLWVLRSDLVRGAARVTLPDGFDAVTRDRGMVVDWAPQEEVLAHG
AVGGFWTHGGWNSTLEGVCGGVPMLCRPHFGDQMGNARYVELVWRTGLALEGELERGKVEEAIRRLMRSEEGDRM
RERAQELKRRAAEAITDDASGSSRLSIDKLVNHILSLSLSK
>AsGT4h2
MMYLCPISSRGPHDCDSSLDLQQLAIMSLLCDRSNGTGRRVVIIPLPYLGHMTPMFRLAAALHARGHAITVLHTE
LHAPDPASYPSDYRFVGVGVPAAELPAASEDIAAFLVALNDSCAAAFKDRLAAMLAAEGSVCCVVTDVVWFSAQA
AARELGVPALALMTSSAASFRTFMAYPALLANGHLPYDESRRDHLVEELPPFRVRDLQRIDTSSLDTFAGLLERF
VDGARRSSGLILNTFHSIEDQEVRNIRDGLAVPVFPVGPLNKISSSPPPLPQDQDQDQDCLILDWLDTKPTGSVL
FVSLGSVATVDAQELAELARGLADTGHPFLWVVRPGMIRGGPPDLDLELPADRGMVVPWAPQEEVLRHAAVGAFL
THSGWNSTVEALSEGVPMACLPCFGDQLGTARYACHVWRVGVEVQGIKRDTVRSAIHRLMGPGAIEEGKEIRERA
HDLKSKVGQSIAQGGSSHMALLGLLEKIACF
>AsGT03883
MAKGHVMVLPMPCQGHVIPLMELSHRLVDHGFEVTFVNTEVDHALVLAALGDGALHGIHLASIPDGLADDEDRKD
LNKLVDAYSRHMPGYLEALVGEMEAAGRPKVRWLVGDVNMGWSFPVAKRLGIRVASFWPASMACLAIMLKIPKLI
QDGVLDDKGWPVREETLQLAPGMPPLHTSLLSWNNAGAPEGQPIIYDLVCRNNKFSDLSEIVVTNSFQEAEAGAF
KLFPDILPIGPLFADGAFQKPVGSFLPEDERCLRWLDARAAGSVVYVAFGSMAIFDPRQFQELAEGLELTGRPFL
WVVRPDFTPGLDMTWLQAFQRRVAGTGMIVSWCSQQQVLAHRAVACFVSHCGWNSTLEAVRNGVPVLCWPYFCDQ
FLDRSYVTDVWRTGLAVAPGEDGVVAKEEVRGKVDKVIGDAGFRKRAGWLKDAAAECVGEDGSSYKNFTRFVELL
SE

>AsGT11637
MSPPVLVMPLPEQGHVIPLMELSHRLVEHGLEVTFVCTEHIHKLLLGGGATDEAALGGVRLVSVPDDLADSDDRR
DLGKVLDGFTRSIPAHMEELIRVTKAEWLVADSTIVGWCSEVAKKLGVRVACFFPASAAYLGTLLRIPQLIEEGI
FDANGFPKRQEALERAPNMPPLYPAQMPWSVAGGPEGQQAFFRLVSRNAEATRNADVVVCNTFLDAERTALGLYP
EILPIGPLLSRKKKSPVAGQFLPEDTGCLQWLDAHSDKSVVYVAFGTSSAFDPGEIRALAEGLERTGRPFLWVVR
PDFTADAAAGSGGGVWTKEWFDGFKERVAGKAMVVSWCAQQQVLAHPAVACFVTHCGWNSTMEAVMNGVPILCWP
YFLDHFTHQSYICDIWRTGLAVSPGQDGVGVTKEEVSSKVEQMVTDIGIAQRARSLSDAAHRGLAPGGSSFQNFN
TFLKLLGLKERPAAF
>AsGT18535
MAASDAAPAPPHPHPHALILPYPAQGHVIPFMELAQRFLDRGFHVTFVNTEFTHGRVMTADAGATSWTGSSSAAA
GGGARLRLVAVADGIEDAGDHENLVLLNAAMPEAIPPQLEPLLDGTGEGLGKVTCVVVDVAMSFALDVVKRRGIT
SAALWPSSAAVLSAMVNAEKLIRDGVIDDDGAPLNLVNNSFHLTEFKAPMDATFLAWNYMGNRDAERLVFHYLAT
IAQTAATKTDYLLCNTFADIEPIIFTDSTPTIIPIGPLRTWQRQTRHGPIGHFWHTEDVMCMSFLDAQPIGSVVY
VAFGSISIMTAEQVQELALGLQASGRPFLWVVRPEQASKLPVGFMDDVDESRKGKVVGWAPQEQVLGHSSVGCFV
THCGWNSTLEGIRNGLPMLCWPYFLDQVTNQTYICDIWKVGLRVAPAEGGVLVTKSRIMELLDGLMEDEGVKERV
LKLKETAESNMSEEGASLKNLNILMESFESKTF
>AsGT23002
MAAAAAAEPRVLGHVIPLMELSCKLAEHGIQVDFVNTEFNHDLILGAMADERAIPQGIRMISIPDGLGPEDDHAD
IGKFVRDLPAAMSGRLQEMIRSNKIKWVIVDVSMSWALEVATKAGARAASFSTYSAAVFALRVNLPKLIEDGVLD
ESEHGAYSGNVNRQARIQMMPPIDAAEIPWVSLASTSAPERRRNNIQNVLKTNLSMPLAEAIICNTSMEMEPDAL
ALLPNALPLGPLVAPASGPAGHFLSEDLACLTWIDAQGPGSVVYVAFGSSGFLDATQFQELADGLALSGRPFLWV
VRPKFTTGVGVGQDWFDAFKRRVEGRGLVVGWAPQQRVLSHPSVACFVSHCGWNSTVEGLLHGVPFLCWPYFADQ
FCNQSYVCNVWGTGVKLCADERGVVSKEEVKKKVEQVLGDEEVKARAAVWKDKARASIAEGGSSHQSLLKLVNLL
RQQ
>AsGT24951
MATSAPRVMVLPLPAQGHVTPLMELSHRLVDHGFEVTFVCTEPVLDALRQTTVDNDGIRLVSIPDGLADGDERRD
LGKVLDALSRCVPGYVEELIRKTKVKWLVADTNMGSLCFEVAKKLGVRVASFFPASAACLGTLFRVPQLIEDGFF
DEMGFPKRRGAFQLAPKMPPIYTSHMVWSVEGGPAVQHAAFQLVCQNNEASSHAEVVVCNSFLEAEATAFELFPN
ILPIGPLFADRRKPVGQFLPEDTSCLGWLDAQPDNSVVYVAFGTSTVFDPRQFKELAEGLELTGRPFLWVARPDF
TSGAGVSNKAWFDEFETRVAGKGMVVSWCSQQQVLAHRAVACFVSHCGWNSTMEGVRNGVPFLCWSTLKVDQYSN
RSYVCEIWRTGLAVSPGGEDGVVTKEEVRTRLEQVIGDHGIAERAQMLADAARMSIGEGGSSYENFNRFIGLLME
>AsGT17328
MVLPFPAQGHVMPLMKLSHRLVDHGLEVDFVNTDFNHARVLKVMAAETGAAAGGIHMVSFPDGMGPDGDRTDIAM
LGNGLPAAMLGPLEETIRSRKIKWVIADVSMSWAMELAAMAGVRVALFSTFSAAVVALRLHVPKLIDDGIIDECG
TVTRNEMIHLSPKMPPVEAAEIPWASLLGGSPDRMRVFIKNVLKTNPAISLAAAVICNTFEEIESEALDLVPNAL
PVGPLEAPLASSSSSAAGQFWPEDATCLAWLDAQTRGSVIYVAFGSFTVLDAAHFEELADGLELSGRPFLWAVRP
NFTDGVVGEGWLDAFKRRVEGKGLVVGWAPQQRVLAHPSVACFVSHCGWNSTMEALRHGVPFLCWPYFADQFCNQ
SYVCNVWGTGVKLLADKRGVVTKEEIKNKVEQLLSDEKIKARVAMWKHAACASISDGGSSHVNLLKLVNLLTEQ
>AsGT26962
MASLAAAEGARHVVALPYPGRGHINPMLAVCRLLVAADGALAVTVVVTEEWHTLLASTALPDRVGLATIPNVIPS
ERGRGVDHAGFIDAVYHKMGEPVERLLDGLERRPEAVVADTYLTWAVAAGARRGIPVCSLWTQPATFFLALFHFD
LWPPVDDRASEEELSIRSLEETVPVQCLSSVRLSDLKTFRAWKRPMEISAETFANIGRAQCVLFTSFHELEPCAM
DAVAESLPCPVYPVGPSIPQLALDGDDKILDEGHRGWLDAQPENSVLYVSFGSFVAMSPSQFEEISMGLRDSGVR
FFWVARDKAADVQQTCGDKGLAVPWCEQQKVLCHPSVGGFLSHCGWNSVLEAVCAGVPLLAFPVGWDQLVNARMV
ADEWKVGINLREQRGEGGTVGRAAISAAARKLMDFDSEVGQEMRRRAAELSHLSRGAVKEGGSSRRSLSGFLQDL
AEGKLDVTETPP
>AsGT10326
MDPLLPPMAMAREANGRRKKPHAVVIPYPMQGHVIPAAHLSLRLAARGFAVTFVNTESVHQRTADALGVDPRSYD
VFAGARCADVRYELVSDGFPLGFDRSLNHDQFMEGLLHVIPAHMEELLRRVVVDPASTCLVVDTFFVWPATLARK
LGVPYVSFWTEPALIFALYYNMDLLTKNGHFRCQEPRRDTITYIPGVPAIEPHELMSYLQETDTTSVVHRIIFKA
FEEAHGADYILCNTVEELEPSTIAALRLEKPFYAVGPIFPAGFAHSVVATSMWAESDCSGWLDAQPAGSVLYISF
GSYAHITKQELREIAGGVLASGARFLWVMRPDIVSSEDPDPLPEGFAAASAGRGLVVPWCCQVEVLSHAAVGGFL

THCGWNSVLESVWAAVPMLCFPLLTDQLTNRRIVVREWRVGVPVGDRGAVFAHEVRARIQGVMSGEEGEELRQAV
RKVRATLEAAAAPGGSSQRGFNEFVDELTRRCGG
>AsGT15a11 (AsUGT74H7)
MGAEPGHVYCDIHVLVLPYPLQGHINPMLQFAKRLARTQTHIGGGGGVRCTLAVTPYLLGQCQDPCPGAVHLAEI
SDGFDRAGFLEVGDVAAYLAQLESAGSRTLDELLRSEAEKGRKVCAVVYDSFLQPWAPPVARRHGAACVSFFTQA
PAVNLAYAHHARGGGTGGRLEGLPAGFEHEDLPTFLTMPDDCPPYLEMLLRQHVGLDAVDHVLVNSFHELQPLES
DYMASKWGAKTVGPTVPSAYLDKRIPDDVSYGFHLYTPTTATTTAWLDAQPPRSVAYVSFGSMAAPGPEQMAEMA
EGLHSSGKAFLWVVRASEASKIPDGFSEKVGTRGLVVPWVAQLEVLAHSAVGCFVTHCGWNSTMEALGAGLPMVA
VPQWSDQPTNAKYVEDVWCVGVRARRDPEAGVVRREEVERCVKEVMGADKQYARNASDWKEKAVRSMCQGGSSEK
NITEFLHALRSRKSQKSTQPKQI
>AsGT01989
MAAADSDEHGGEVHVLLLPYPSQGHINPILQFGKRLASHAGVRCTLAVTHFLLGQGRDPSPGAVHLAEISDGYDR
GGFAEAAGDVAAYLARLESVGSRTVDELLLSEAAQGRPVRVVVYDAFLQPWAPEVARRRGAACASFFTQAPAVDV
AYEHAWAGRMVVPVVGEVPSGLPGLPAGLEPADLPSFLTAPDDCPAYLDLLVKQFVGLDAADHVLVNSFHELQPL
ESEHMASTWGAKTVGPTVPSAYLDNRLPDDVSYGFHLHTPTTATTKSWLDAQPARSVAYVSFGSIAELGPAEMAE
VAEGLYSSGKAFLWVVRALETSKIPEGFADRAGGRGLIVTWTAQLEVLAHGAVGCFVTHCGWNSTTEALGAGVPM
VAMPHWSDQPTNAKYIEDVWRVGVRARRDDGGVVRREEVERCVREVMEGERSEEYVRNAVEWKEKAINAMGEGGS
SDTNITEFLRQLIRSKKPEQARKAESVV
>AsGT02699
MEAKPPLSVQADGSGGGGHVLLLAFPGAQGHLNPMLQFGRRLAYHGLRPTLVTTRHLLATLPPPAAPFRVAAISD
GFDAGGMAACPDFEEYVRRLAAAGSDTLEALIRSEAAAGRPVRALVYDPHLPWAGRVARAASVPTAALFSQPCAV
DVIYGEVYAGRVGLPVVDGSALRGLLSVELGPQDVPSFVAAPDSYRMFLDAVVGQFDGLEDADDVFVNSFHELER
KEADYLASTWRFKTIGPTLPSFYLDDDRLPSNKTYGFNLFSSTAPCMAWLDSQPPCSVVYASYGTVADLEPTQLE
EIGNGLCNSGKQFLWVVRSVDEHKLSEELRGKCKERGLIVSWCPQLEVLSHKATGCFLTHCGWNSTIEAITTGIP
LLAMPQWTDQPTTAKYVESAWVIGVRVHRDKEGVVRKEEVERCIREVLDGVRKEEYKKSAYIWMKKAKEAMQEGG
SSDKNIAEFAAKYASS
>SAD10
MGAEWEHVSDIHVLLLPYPVQGHINPMLQFGKRLAHIGGVGVRCTLAITPYLLRQCQDPCPGAVHLVEISDGFDS
AGFEEVGDVAAYLAGMESAGSRTLDELLRSEAEKGRPIHAVVYDAFLQPWVPRVARLHGAACVSFFTQAAAVNVA
YSRRVGKIEEGLPAGFEAEDLPTFLTLPLPYQDMLLSQFVGLDAVDHVLVNSFHELQPQESAYMESTWGAKTVGP
TVPSAYLDKRITDDVSYGFHLYTPMTATTKAWLDAQPPRSVTYVSFGSMATPGPTEMAEMAEGLHSSGKAFLWVV
RASEASKIPDGFQERVGGRGLVVTWVAQLEVLAHGAIGCFVTHCGWNSTMEALGAGVPMVAVPQWSDQPTNAKFV
EDVWCVGVRARRDPEGVVRREELERCIREVTGDDKYACNALDWKEKSKRAMSQGGSSDMNITEFLQALRRSRKSY
EAKPIEPLLVGLDA
>AsGT07784
ETPTKPKHTWRARRRPTPARARAAAARTCCSFRSRERRATRTRCSSSAGASRATACAPPRGHPLRALDHSPARRA
LPRGRHLRRLRRRRPGVLRGHGGVHLEPGGRGVQDAAGAPGVGGARGRPVRVLVYDPHMPWPPRVAREAGVLAAA
FFSQPCAVNVVYGEVWAGRMALPATDGRALVERGALGVELGPEDMPPFVAVPELKPEFLKTSIWQFEGLEDAADV
LVNSFSDIEPKEAKYMEDTWRAKMIGPSVPSFYLDDHRLPYNKSYGFNLFSGEDPCMDWMEKQSINSVVFVSYGT
FSEYEASQLEELGSGLCSSGKPFLWVVRSSEACKLSDELKAKCKKNGLIVSWCPQLEVLAHKSIGCFVTHCGWNS
TLEAIVNGVPLIGIPHWADQPTIAKYVQSLWGMGVRAKRGENGWVTREEVDRCIKEVMDGKMKDVYKRNAGKWMQ
KAKKAMQEGGSSDKNILEFVAKYS
>AsGT01599
MADRGGDGGRTCSFLVVAYGIQGHLNPARSLARRLAAIAGVTATLSVQIFSHRRMFPASSSSGDDAEIVSDGVIS
YLPFSDGQDDGSWPTGSDQERARRRAATVDSLSSVVRRLAAAGQPVSCVVCTLNVPAVVDLARAHGLPLAVYWIQ
PATALLAYYHHFHGHGEAIASHAADQAYEIALPGLRRPLRVRDLPSFIVDEGTGGGDELSRFIHQEFRRLFEQMD
EDKAVIVLVNTFEALEATALEGIRPYFDGGVFAVGASTIPLPGAGEEDGRIDLFDQDVDNEYMAWLDAQPARSVV
YVSFGSLLTYSAREAEEILIGLRRIGRPYLWVVRLEGRSPEVDRLLLSEATSASAEVGMVVTWCDQVRVLSHPSV
ACFVTHCGWNSTLEAVECGVPVVAAPSWSDQPMIAHLLEEEWGVGVRAERGADDGVLAGAELARCVELVVGDRGT
TATNVSAWKKRAREAVAAGGPSERSLGCFVKTVQELDGFVISTNLS
>AsGT03158
MPAMVAQEVERAAAAAAPHFLVVTYPAQGHINPARHLALRLLRATPGARVTLSTAVSACRKMFPGLEEEHVDGAG
VRYAPYSDGYDGGFDKEAHHHTDYMSQLKVVGARTLDGVLARLRDAGRPVTRVVYTVLLSWVAGVARAHGVPAAL

YWIQPATVLATYFHFFRGTDGFDEAVVAAAQDPWAEVRLPGFPAPLRMRDLPSFLTITSDDHPYAFILTAFRELL
DALDDDRQEDGRGRSATVLANTFDAMEPDAVATLRQHGVHVVPIGPVLSFLDAAPASSTNNNNDLFSQDGKGYLE
WLDAQEAASVVYISFGSMSSMSKRQITEVSRGMAESGRPFLWVLRKDNRGEVDGDDLCAGAGGGGMVVEWCDQGK
VLSHPAVGCFVTHCGWNSTLESVSCGVPVVGVPQWTDQGTNAWLVERQLGTGVRAAVCEKDGVLEAEELRRCIGF
ATSEMVRAKAALWREKARAAAAGGGSSERNLRAFVDLAGN
>AsGT10759
MAETELLPHPHGGGAGVRRHFLLVAYGIQSHLNPCRVLARRLVQLHDADGSDPVLATLSFPLFTHRRMFPSSGKG
ELEEEVADGVISYAPYSDGVDDGTIARDADGRARRRRASFESLSAVVARLAARGRPVTCIVCSLVLPCALDVARE
NAIPMAVFWIQPATVLAAYYHYFHGYGELIASHAADPSYEVALPGLCQPLWIRDFPSFLVDTTGGEVGKLVNDAF
RELFEFMDEQGHGAKFLVNTFDELEPAALAAMRRYLDVFAIGPLVGSSAEARIHLFDHAGADKKRYMEWLGAQPD
MSVVYVSFGSIWTYSKQQMEEIAKGLRQCGRPYLLVVRKDGRQEDVSSCLDDAVLKGQGMVVEWCDQPEVLAHPS
VGCFVTHCGWNSTLEAMTLGVPVVAVPSMFDQPTNAYLIEQCVAGVRGERNGNGVFAGAELARCVELVMGSGARA
VEIREGVEALKGMAREAAVSGGRADRNLRSFVTDTTIH
>AsGT28b19
MAEMERLPHPHGRSSNGERRHFLLVAYGSQSHLNPCRVLARSLVQLHDADGSGPVLATLSVPLFTHRRMFPSSCS
GVPEDEDTTDGVISYAPYSDGVDDGTNAKDAEGRARRRRASFESLSAIVARLAARGRPVTCVVCSLVLPYALDVA
REHAIPMAVFWIQPATVLAAYYHYFHGYGELITSHAADPAYEVTLPGLCQPLRTRDFPSFLVDTTGGEVAKSVND
LFRELFEFMDAQGQRAKFLVNTFEKLEPAALASMRQHLDVFAVGPVMGSSAVARIHLFHHAGADKKRYMEWLGAQ
PEISVVYISFGSVWTYSKQQMEEIMHGLRQCGRPYLLVVRKDDRQEDVSSCLDDVVREGRGMVVEWCDQPAVLSH
PSVGCFLTHCGWNSTLEAMAMGVPVVAAPSMLDQPTNAFLIEGEWKAGVRGECNGEGVFTGAELARCVEMVMSSG
ARALEIRQRAEALKGMAREAAASGGPAERSLRNFVTAASGADKTSTKDIPPTMSESI
>AsGT17576
MAPQHFLVVAFPGQGHINPARALAERLARTTPGARVTLSAAVSAHRLMFPSLASPDEEIHDGAISYIPYSDGYDH
GFRLFAGDGDDARRYNEAFARVGPETFSAVLDRLAARGRPVTCVVYALLMWWAAEVARERGLPRALYWNQPATML
AVYYHYFHGYERTVAEHAGDPEFTVAMPGLPPMAIRDLPSFFTDLTDGRIVEAFGNIRSTFQQLDLDVDQGSTGR
RKPMVLVNTVEALEAGAIASVPELDVFPIGPAVLSLFEQDGDNAIVGDLFEHDGKGYMEWLDTKPARSVVYVSFG
SMTASSRRQKEEMKRGLAASGRSYLWVVRKDNRDDDEDGDDDERSMVVEWCDQVRVLSHPAVGCFVTHCGWNSTL
ESMACGAPVVAVPQWSDQDTNARLVVEWGIGVRATIDADRFLDAEELTRCVEMVMGDTEEGAAIRRRSTACKAKV
QEAITDGGSSENNLRTFQGHFANDV
>AsGT11140
MSLRPHFLVLTFPLQGHIAPALRLARRLLAAAPDALVTFSTTEAGHRRMFSATPDAPDDDGRLELLPFSDGTETG
YVRSSEAASFNDYMASFHAAGARSVGELVDALAARGRPVTRVVYTMLLPWAADVARERGVPSALYWIQPAAVLAI
YYHYFHGHAAVVAEHRYDPSFLVRLPGLPPQAIRDLPSFITDSTDPSDFYHCVYVTTRDLFDALDRETPKATVII
NTCQELEEATLAALGAYNSLPIGPVLPAGDEACLFKQDDAGYIEWLDAKPANSVVYVSFGSLARMAKEQLDELLL
GLEESGRPYLCVVRKDNKAELAEAEAEMDARIKNGMVVEWCDQVRVLEHAAVGCFVTHCGWNSVLETVASGVPMV
CVPQMSDQRMNAWLVEYEWRVGTRAEVGSDGVLRAAELRRRVEEVMRGDARGTAAGWKRAVAEALGKGGSSDHNL
MAFVEGIGSDVQ
>AsGT15275
MAETELLPHPHGGGAGVRRHFLLVAYGIQSHLNPCRVLARRLVQLHDADGSDPVLATLSFPLFTHRRMFPSSGKG
ELEEEVADGVISYAPYSDGVDDGTIARDADGRARRRRASFESLSAVVARLAARGRPVTCIVCSLVLPCALDVARE
NAIPMAVFWIQPATVLAAYYHYFHGYGELIASHAADPSYEVALPGLCQPLWIRDFPSFLVDTTGGEVGKLVNDAF
RELFEFMDEQGHGAKFLVNTFDELEPAALAAMRRYLDVFAIGPLVGSSAEARIHLFDHAGADKKRYMEWLGAQPD
MSVVYVSFGSIWTYSKQQMEEIAKGLRQCGRPYLLVVRKDGRQEDVSSCLDDAVLKGQGMVVEWCDQPEVLAHPS
VGCFVTHCGWNSTLEAMTLGVPVVAVPSMFDQPTNAYLIEQCVAGVRGERNGNGVFAGAELARCVELVMGSGARA
VEIREGVEALKGMAREAAVSGGRADRNLRSFVTDTTIH
>AsGT1a15
MGEEAVVSTVAAAAMSSAPHLLLICFPGQGHVNPMLRLAKRFAAKGLLVTFSSTSDVXAKITASTGVEAGGDGVP
LGLGRIRFEFLDDHSEGLTDLDPLMRHLQTVGPPAFVELIRRQEEAGRPVSCVVGNPFLPWAIDVAHDAGIPSAV
LWVQSCAVFSLYYHHVHGLVEFPPEDDLEALVKLPGLPAMSVADVPSFLLPSNPYKLLANEILKQFRTIHKASWV
FVNSFSELERDVVDALPGVSPAPPPLIPVGPLVELAEDASVRGDMLKAADDCVGWLDTQAPRSVVYASLGSVVVL
SAEQLAELAYGLASSGRPFLWVVRPDSSAMLPEGYLESIAGRGMVVPWSPQDLVLAHPSTACFLTHCGWNSTLET
LAAGVPVVAFPQWGDQCTDAKYLVEEFKMGVRIGAPLRRDAVRDAVEDAVAGPDAAAMAENARAWSAAATTAVSS
GGSSDRHVQAFVDEVVARAGGARADKDHLAVEQKASIGMTSVAA

>AsGT01341
MAPAAATEVAPAAPVAIVAVPFPAQGHLNQMLHLSLLLASRGLDVHYAAPAPHVRQARARVQGWDARTVGSLHFA
ALDIPPYATPPPDPESPVAFPGHLQPLFDAFCDHAAAPLARLLEDLAASHSHRRVVVLHDNIMSFPAAEASRLPS
CEAYALQCVAQSFCAGFKDPAHPLVRALGLPVPPPESFLTAEFMELVQRQDGLGVPGAGLLLNSCRALEGDFIDA
LHEILSEDGKKLFTIGPLNPLLDLDVSTTPAPSPQPRHECLDWLDKQPVSSVLYVSFGTMTSLPGKQIEELAAAL
QGSGQRFIWVLRDADRADIFAEAGGKSRHGELLSEFTKQTEGKGLVITGWAPQLEILAHGATAAFVSHCGWNSLL
EGLGHGKPILAWPMHCDQPWNAGYVCGHLKAGIVVRPWEKNRETVPAKDIQEVIRRAMDSDEGAVLRKAAKALAE
DVRAAVAEGGSSWADMEAFIRHITK
>AsGT05827
MGIESMDSSVALVAVPFPAQGHLNQLMHLSLLVASRGLSVHYAAPAAHVRQAKSRVHGWDAKALASIHFHDLDVP
TFKSPDPDPAAASPFPSHLLPMWETYSAAARVPLASLLERLSATHRRVVVVYDHMNSFAAAEAARVDGEAYGLVC
VAISNHLAWMPDGHQLLRDRGLRSVPMDACMSKEFVEYMARVTTEAEGAGFLMNTCRALEGEFLDAVAEIPDIKR
QKRFAVGPLNPLLPLATEPDVVTTATARHDCMRWLDAQPPASVLYVSFGTTSSFLPEQIAELAAAIKGSRQRFIW
VLRDADRADIFAGNSGGDSRPRYEKLLSEEQAQGTGLVITGWAPQLEILAHGATAAFMSHCGWNSTMESLSHGKP
ILAWPMHSDQPWDAELLCNYLRVAMLVRPWEKHGEVVAAGAIQEVIEKAMLSEKGTALRQRAKLLGEAVRAAVAD
NGSSTKDLHDFVAHITRI
>AsGT18257
MAVDTLESVAFVAVPFPAQGHLNQLMHLSLLVASRGLSVHYAAPVPHVRQARSRVHGWDPNALASINFHDLDVST
YDSPPPDPTAPSPFPNHLMPMFETFTAAARAPLGRLVETLAATHRRVVVVHDKLNSFAAAEAARVSNGEAFALQC
VAISYNIGWLDREHRLLTDHGLQFLPIDACMSKEFVEYVFRTETETLQEGGGGVASSGLVMNTCRALEGEFIDALA
EHPEFKDQKLFAVGPLNPLLDVSARTPGKARHECMDWLDEQPPASVLYVSFGTTSSFRGEQIAEMAAALKGSKQR
FIWVLRDADRADIFAEAGEARHGEMLSQFTTETEGTGLVITGWAPQLEILAHGATAAFMSHCGWNSTMESLSHGK
PILAWPMHSDQPWDAEFLCKYLKVGLLVRPWEKHSEVVPAADIQEVIEEAMISENGIALRNRAKVLGEAVRAAVA
DGGSSSKGLDDFVAYITR
>AsGT27a12
MKDATATERVAVVAVPFPAQGHLNQMLHLSLQLASRGGLELEVHYAAPEQHVRQARARVHGWGDEALRSIHFYDL
GISSYISPPPDPTADSPFPSHLMPLYEAYTASARAPLAALLDTLSGSRRRVVVVHDRINAFAIQEAARLPNGEAF
GLHCLALSMLVGQADPDHRVLRDNGLVFSGVEQYATEEFLEYRRRARPSKQISPGAGVLTNTCRAIEGEFIDVVA
AHLAGDGKKKLFAIGPLNPLVAPAAAPKRSPHAHECLDWLDQQPPASVLYVSFGTSSTLPAEQIQELAAALRDSK
QRFIWVLRDADRGDIFMAEDGGESRRHEKLLPEFTQGTGMVITGWAPQLEILAHRATAAFMSHCGWNSTMESLSH
GKPILAWPMHCDQPWDAQLVCNYLKAGILVRPWEKHGEVIAANAIQRVIEDAMLSDKGMEVRRRAMALGDSVRAS
MADGGPSSKSQDDFIAYITSR
>AsGT22388
MAVDSLESVAFVAVPFPAQGHLNQLMHLSLLVASRGLSVHYAAPAAHVRQAQSRVHGWDPEALGAIHFHDLDVST
YESPAPDPTAASPFPSHLMPMWETFTAAARAPLADLLESLSATHRRVVVVYDNLNSFAAVEAARLRNGEAFGLQC
VAISYSLGWLDSEHKLLRDHDLQFLPIDACTTKEFVDYVSRAARDLLDRGGVPSSGLVMNTCRALEGEFIDAIAE
TPAFKDRKLFAVGPLNPLLDASARTPGRTRHECMDWLDKQPDASVLYVSFGTTTSFRLEQIAEMAAAIKGSRQRF
IWVLRDADRADIFAESGGETPYEKLLSQFTKDTEGTGLVITGWAPQLEILAHGATAAFMSHCGWNSTMESLSHGK
PILAWPMHSDQPWDADLLCKYLKVGFLVRPWEKHSEVVPAASIQVVIEEAMVSDNGMAVRQRAKMLGEGVRASVA
EGGSSSKGLDDFVAYITR
>AsGT24248
MAMESVAFVAVPFPAQGHLNQLMDMSLLVASRGLSVHYAAPAAHVRQARSRVHGWDPKALASVHFHDLDIATYES
PAPDPAAVSPFPSHLIPMFEAFTAGARAPLAVLLDSLSATHRRVVVVYDNINYFAAVEAARLGNGEAYGLQCVAM
SYNVAWLGAGPGHQLLRDHGLQCLPFDAPMSREFMEYMFRATAEARDAGGVARAGLVVNTCRVLEGEFIDAVAEH
AEYKGQKLFAVGPLNPLLDGARAVPGKTRHECMDWLDAQPPASVLYVSFGTTSSFREEQIAEIAAALQGSKQRFI
WVLRDADRADIFAESGESRHDKLLPEFTKETEGSGLVITEWAPQLEILAHGATAAFMSHCGWNSTLESLSHGKPM
LAWPMHSDQPWDAELLCKYLKVALLVRPWEKHAEVVPADAIKEVIEEAMITENGMALRERAKVLGDAIRASMAEG
GSSRKGLDDFVSYITR
>AsGT24a3
MGSEHHAMGSVAVVAVPFPAQGHLNQLLHLSLQLASRGADVHYAAPAPHIRQARARVHGWDQEALRSIQFHDLAI
SSYDSPPPDPTADSPFPSHIMPLTVAYITGARAPLAALLHELSGSRRRVVVVHDRINAFAHEEAARLPNGEAFGL
HCLAVSLLVGHIDAELLREKGLVYTGVERHATKEFLEWARRARPSRQISPGAGVLANTCRALEGEFVDVVAGHLA
PDGKKIFAIGPLNPVLPASASNQGKQQRHECLGWLDEQPPASVLYVSFGTTSSLRDEQIQELAAALRGSKQRFIW

VLRDADRGDIFAEAGESRHQKLLSEFTNNTEQTGLVITGWAPQLEILAHGATAAFMSHCGWNSTVESLSHGKPIL
AWPMHCDQPWDAELVCNYLKAGILVRPWEKHSEVIAAHAIQEVIEEAMLSDRGIAVRQRAKALGDAVRDSVADGG
SSRKDLDDFIAYITR
>AsGT06218
MAPVHVLVFPWPLQGHINSMLHFAAGLLGVGIHVSFVHTEHNLRRVDLAKAAASPRLRVMSLPDGLPDDHPRSVD
DLKDLAKSLITTGTVPYRALLASMLSAGSRAVDAADATDSGFPPLSCVVADGLLPFAIDIAEDLGVPALAFRTAS
ACSVLAYLSVPKLMEVGEVPIPVGADLDELVLGVPGMEDFLRRRDLPGTCRLRAETNAVDPLLHILVNYSVHSSK
ARALIFNSPASLEGLALAHIAPHMRDVFAIGPLHAISAAPVLPTSLWREDDGCMAWLDGHADRSVVYVSLGSLAV
ISLEQFTEFLSGLVGAGYAFVWVLRPDMVGESQRGVLQEAIDAAGNDKMRVVDWAPQRDVLRHRAVGCFLTHAGW
NSTLEGVVEGVPLVCWPFFADQQINSRFVGAVWGNGLDMKDVCERAVVEGMVREAMVSGELRMSAQALAQQVRRD
IAEGGSSATEFERLLCFIKELTAKGCSPK
>AsGT12o13
MNLSTKTTTPAQACTRSPTASAGMEPAAAAAAAHVLVFPWPLQGHINCMHHLATALLDAGLHVTFLHTHHNLRRL
ATKPAPAPSQPRLRLLSIPDGLPEDHPRSVAHLNDLMDSMRTTGSAAYRALLLASSSNKDGHPPVTCVIADGVMA
FAVDVAEEVGVPAIAFRTASACSFLTYLSVRRLVELGEFPFPSDQPVSGVPGMEGFLRRRDLPRAPRPAGSATDD
CGVDPMLLNMGECTVHSGEARALILNTSASMEGPALAQIAPHMRDVFSVGPLHVAAGTGTKSTAPTASLWREDDG
CMAWLDGQQDRSVVYVSLGSLTVISEEQLAEFLSGLAATGYAFLWVLRPDMVAGGTTSLAAVKTLVGEKARVVHW
APQRDVLRHPAVGCFLTHAGWNSTLEAAYEGVPMVCWTFFGDQLINSRFVDTVWKTGVDIKDVCDRAVVEKAVRE
AMESAQIRAAAQAMARQLRLDVADGGSSSSEIKRLVAFIRDLL