A noncanonical vacuolar sugar transferase required for biosynthesis of antimicrobial defense compounds in oat

Anastasia Ormea, Thomas Louveaub, Michael J. Stephensonb, Ingo Appelhagenb, Rachel Meltonb, Jitender Cheemac, Yan Li, Qiang Zhaob, Lei Zhanga, Danlin Fanb, Qinlin Tia, Robert J. Vickerstaffc,1, Tim Langдонb, Bin Hanb, and Anne Osbournb,2

aDepartment of Metabolic Biology, John Innes Centre, NR4 7UH Norwich, United Kingdom; bNational Centre for Gene Research, CAS-JIC Centre of Excellence for Plant and Microbial Science, Centre of Excellence for Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 200032 Shanghai, China; and cInstitute of Biological, Environmental, and Rural Sciences, Aberystwyth University, SY23 3FL Aberystwyth, United Kingdom

Edited by Wolf B. Frommer, Heinrich Heine University, Düsseldorf, Germany, and accepted by Editorial Board Member Joseph R. Ecker November 3, 2019 (received for review August 29, 2019)

Plants produce an array of natural products with important ecological functions. These compounds are often decorated with oligosaccharide groups that influence bioactivity, but the biosynthesis of such sugar chains is not well understood. Triterpene glycosides (saponins) are a large family of plant natural products that determine important agronomic traits, as exemplified byavenacins, antimicrobial defense compounds produced by oats. Avenacins have a branched trisaccharide moeity consisting of L-arabinose linked to β-D-glucose molecules that is critical for antifungal activity. Plant natural product glycosylation is usually performed by uridine diphosphate-dependent glycosyltransferases (UGTs). We previously characterized the arabinosyltransferase that initiates the avenacin sugar chain; however, the enzymes that add the 2 remaining β-glucose molecules have remained elusive. Here we characterize the enzymes that catalyze these last 2 glycosylation steps. AsUGT91G16 is a classical cytosolic UGT that adds a 1,2-linked β-glucose molecule to L-arabinose. Unexpectedly, the enzyme that adds the final 1,4-linked β-glucose (AsTG1) is not a UGT, but rather a sugar transferase belonging to Glycosyl Hydrodase family 1 (GH1). Unlike classical UGTs, AsTG1 is vacuolar. Analysis of oat mutants reveals that AsTG1 corresponds to SAD3, a previously uncharacterized locus shown by mutation to be required for avenacin biosynthesis. AsTG1 and AsUGT91G16 form part of the avenacin biosynthetic gene cluster. Our demonstration that a vacuolar transglucosidase family member plays a critical role in triterpene biosynthesis highlights the importance of considering other classes of carbohydrate-active enzymes in addition to UGTs as candidates when elucidating pathways for the biosynthesis of glycosylated natural products in plants.

Significance

Glycosylation is often critical for the biological activity of plant natural products. Here we identify 2 enzymes required for glycosylation of the antifungal triterpene glycoside avenacin A-1 and show that 1 of these (AsUGT91G16) is a classical cytosolic uridine diphosphate-dependent glycosyltransferase, while the other (AsTG1) is a noncanonical vacuolar sugar transferase belonging to glycosyl hydrodase family 1. The genes encoding these 2 enzymes form part of the avenacin biosynthetic gene cluster. Our demonstration elucidates the steps required for the biosynthesis of the branched oligosaccharide chain required for antifungal activity of avenacin A-1 and thus for disease resistance in oat. They further highlight the importance of noncanonical sugar transferases in natural product glycosylation.

Author contributions: A. Orme, B.H., and A. Osbourn designed research; A. Orme, T. Louveau, M.J.S., I.A., R.M., J.C., Y.L., Q.Z., L.Z., D.F., Q.T., R.I.V., and T. Langdon performed research; A. Orme, T. Louveau, M.J.S., I.A., R.M., J.C., Y.L., Q.Z., L.Z., D.F., Q.T., R.I.V., T. Langdon, B.H., and A. Osbourn analyzed data; and A. Orme, B.H., and A. Osbourn wrote the paper.

Competing interest statement: This work is subject to a patent filing.

This article is a PNAS Direct Submission. W.B.F. is a guest editor invited by the Editorial Board.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

Data deposition: Raw and assembled RNA-seq data can be accessed at http://db.ncgr.ac.cn/oat/RNAseq.php. Sequences have been deposited in the GenBank database (accession nos. MN396758–MN396761).

1Present address: Department of Genetics and Crop Improvement, East Malling Research, East Malling ME19 9BJ, United Kingdom.

2To whom correspondence may be addressed. Email: anne.osbourn@jic.ac.uk.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1914652116/DCSupplemental.

First published December 5, 2019.
Triterpene glycosides are one of the largest and most structurally diverse groups of plant natural products. These molecules are also referred to as saponins (a name derived from sape, the Latin word for soap) because of their surfactant properties. The names of some triterpene glycoside-producing plant species—for example, soapwort (Saponaria officinalis), soapberry (Sapindus species), and soap bark (Quillaja saponaria)—reflect their original use as sources of soap. Triterpene glycosides display a tremendous range of structural diversity and biological activities. They protect plants against attack by pests and pathogens and can determine other agriculturally important traits, such as flavor; for example, they are associated with bitterness in soybean. They are also of considerable interest for industrial and pharmaceutical applications (8, 9). Examples include the sweetener glycyrrhizin, from licorice, the vaccine adjuvant QS-21 from Yaphe and ginsenosides, heath-beneficial triterpene glycosides from ginseng (10).

Triterpene glycosides commonly have 1 or more simple or branched sugar chains, often consisting of 3 or more sugars, attached to the triterpene scaffold. These sugars are critical for many of the bioactive properties of these compounds (2). Despite the importance of glycosylation for biological activity, the enzymes required for the biosynthesis of triterpene sugar chains are not well understood. To date, 26 triterpene UGTs have been previously defined in a forward mutagenesis screen as being required for avenacin pathway genes (Fig. 1C) along with characterized UGTs from other plants, including those previously reported to glycosylate triterpenoids (SI Appendix, Table S5). The phylogenetic tree approximately recapitulated the monophyletic groups A to P previously defined by Ross et al. (37) and Caputi et al. (38). Four oat UGTs from groups A and D were prioritized for further investigation, since both of these groups contain enzymes known to extend triterpene sugar chains (SI Appendix, Table S6). Gene expression analysis revealed that the genes for 2 of these (AsUGT91G16 and AsUGT99C8) had expression patterns closely resembling those of the characterized avenacin pathway genes (Fig. 1D).

AsUGT91G16 and AsUGT99C8 were expressed in Erichesia coli as N-terminal 9X histidine-tagged proteins. Protein preparations enriched for the recombinant enzymes were prepared using immobilized metal affinity chromatography and incubated with bis-deglucosyl avenacin A-1 (2) (SI Appendix, Table S1) (generated by the enzymatic hydrolysis of avenacin A-1) in the presence of UDP-glucose. AsUGT91G16 generated a more polar product with a mass consistent with addition of a hexose when analyzed by HPLC-MS (Fig. 2A), while AsUGT99C8 did not yield any detectable product (SI Appendix, Fig. S3).

We next expressed AsUGT91G16 in Nicotiana benthamiana leaves by Agrobacterium-mediated transient expression, in combination with AsbAS1/SAD1, AsCYP51H10/SAD2, and AsAAT1/UGT99D1, which together compose the arabinosylated triterpene scaffold 3β-(α-L-arabinopyranosyloxy)-12,13β-epoxy,16β-hydroxy-β-amyrin (3) (11) (Fig. 2B). Coexpression of AsUGT91G16 together with these enzymes resulted in the disappearance of (3) and the appearance of 1 major and 3 minor polar peaks. The major peak (4a) (retention time 16.8 min) and 1 of the minor peaks (4b) (retention time 14.4 min) both had masses consistent with the addition of a single hexose to (3), and the other 2 minor peaks (retention times 13.3 and 15.3 min) had masses consistent with the addition of 2 hexoses. We speculated that (4b) is a breakdown product of (4a) formed by opening of the C12-C13 epoxide, which is known to be unstable (39) (Fig. 2B). The latter 2 minor peaks may be the result of modifications of the AsUGT91G16 product by endogenous N. benthamiana glycosyltransferases.

We then scaled up our N. benthamiana infiltration experiments using vacuum infiltration following a method that we developed previously (40), agroinfiltrated 150 plants, and extracted the leaves to purify the AsUGT91G16 product for structural determination by NMR. Under these conditions, the major extracted product was (4b) (SI Appendix, Fig. S4). NMR
analysis of (4b) confirmed that AsUGT91G16 is indeed a triterpene 3-O-arabinose-β-[1,2]-glycosyltransferase (Fig. 2B and SI Appendix, Table S7). The analysis also identified the scaffold of the purified compound 4b as 12-keto,16β-hydroxy-β-amyrin, consistent with breakdown of the epoxide group of (4a) during large-scale extraction.

We previously carried out a forward genetic screen for avenacin-deficient mutants of A. strigosa following sodium azide
mutagenesis and identified a collection of approximately 100 mutants, roughly one-half of which have been assigned to characterized loci (22, 26, 28, 31–33, 41). We sequenced AsUGT91G16 in 49 uncharacterized avenacin-deficient mutants and identified 4

mutants with single nucleotide polymorphisms (SNPs) in this gene (SI Appendix, Table S8). Root extracts of wild-type A. strigosa contain the major avenacin A-1 (1), along with 3 other structural variants: A-2 (5), B-1 (6), and B-2 (7) (Fig. 2C and SI Appendix, Table S8).
Table S1) (39). Synthesis of all 4 avenacins is abolished in sad1 mutants, which are blocked in the first step in the pathway (AsbAS1/SAD1) (30). We previously showed that mutants affected at 2 other uncharacterized loci, Sad3 and Sad4, lack avenacins and accumulate monodeglucosyl avenacins missing the 1,4-linked α-glucose, of which monodeglucosyl avenacin A-1 (8) and monodeglucosyl avenacin A-2 (9) are the most abundant (22, 26). HPLC analysis of root extracts of the 4 ugt91g16 mutants revealed that these mutants share a common metabolite profile that is distinct from that of the wild-type line and also from sad1, sad3, and sad4 mutants. The ugt91g16 mutants accumulate 2 new, less polar compounds, (10) and (11) (Fig. 2C). These 2 peaks have masses consistent with avenacins A-1 and A-2 minus a hexose and also lacking oxidation at the C-30 position (Fig. 2C and SI Appendix, Fig. S5). A lack of C-30 oxidation was also previously observed in AsAAT1 mutants, suggesting that the introduction of this modification is dependent on previous glycosylation (11). Collectively, our data indicate that AsUGT91G16 adds the 1,2-linked α-glucose molecule to the sugar moiety of avenacin A-1.

All of the ugt91g16 mutants had reduced root fluorescence and stunted roots (SI Appendix, Fig. S6). Analysis of seedlings of F2 progeny derived from crosses between the wild-type and each of 2 ugt91g16 mutant lines (#543 and #1473) confirmed that the reduced avenacin phenotype cosegregated with both the SNPs in these lines and with the short root phenotype, and that these phenotypes are likely due to recessive mutations at AsUGT91G16 (SI Appendix, Table S9). The stunted root phenotype is most likely due to accumulation of toxic pathway intermediates. Root growth defects were previously observed for sad3 and sad4 mutants and attributed to the accumulation of incompletely glycosylated avenacin pathway intermediates (26). Asugt91g16 mutants, like other characterized avenacin-deficient mutants, were compromised in resistance to the root-infesting pathogen Gaeumannomyces graminis var. tritici, which causes take-all disease of cereals (SI Appendix, Fig. S7) (22).

Neither AsUGT91G16 nor AsUGT99C8 is able to complete the sugar chain of avenacin A-1 by adding the 1,4-linked α-glucose (SI Appendix, Fig. S8). We have previously shown that 5 of the 7 characterized avenacin biosynthetic genes are located on an ~400-kb bacterial artificial chromosome contig in A. strigosa (28) (SI Appendix, Fig. S1). Genetic analysis indicates that the other 2 characterized avenacin pathway genes (Sad6 and AsAAT1) and other loci that we have identified by mutagenesis but not yet cloned (including Sad3) are also likely to be clustered with these 5 genes (11, 22, 26, 33). A genome sequence is not yet available for A. strigosa; however, another avenacin-producing diploid oat species, Avena atlantica, has been subjected to whole-genome shotgun sequencing, and the assembled contigs have been mapped by survey resequencing of recombinant inbred progeny derived from a cross between this accession and A. strigosa accession Cc7651 (IBERS) as described in Louveau et al. (11). Examination of these assembled contigs revealed that AsUGT91G16 is located on a short (22.9 kb) scaffold next to a gene predicted to encode a member of the GH1 family (Fig. 3A and SI Appendix, Fig. S9 and Table S10). Plant GH1 hydrolases have previously been reported to function in plant development and stress responses, for example, by modulating iron deficiency and rhizobacteria-induced systematic resistance responses and by activating glucosinolate defense compounds and plant growth hormones by hydrolysis (42–44).

GH1 enzymes cleave glycosidic linkages in a 2-step mechanism, retaining the anomeric configuration of the product with respect to the sugar donor (45, 46) (www.cazy.org/glycosyltransferases.html). The first step cleaves the sugar donor to form a covalent glycosyl-enzyme intermediate. The second step transfers the sugar to an acceptor molecule, which is typically water, resulting in hydrolysis of the sugar donor. However, if the acceptor molecule is an alternative sugar acceptor, then transglycosylation can occur (47, 48).

Phylogenetic analysis indicates that AsGH1 belongs to the At/Os6 subfamily of GH1 enzymes, as defined by Opasiri et al. (49) (Fig. 3B and SI Appendix, Table S11). Intriguingly, several members of this subfamily have previously been shown to glucosylate anthocyanins using acyl sugars as sugar donors (SI Appendix, Table S12) (50–52). Therefore, we hypothesized that the AsGH1 enzyme could be the missing avenacin sugar transferase. Consistent with this idea, the AsGH1 gene was seen to be preferentially expressed in the root tips, like the other characterized avenacin biosynthetic genes (Fig. 3C). Unlike classical UDP-dependent sugar transferases, which are normally localized in the cytosol, the oat GH1 enzyme has an N-terminal 18-aa leader sequence, suggesting that it may be targeted for secretion (SI Appendix, Fig. S10).

His-tagged AsGH1 was expressed in E. coli and assayed for activity toward avenacin A-1 with the 1,4-linked α-glucose removed (8) in the presence of the model β-glucosidase sugar substrate 4-nitropheny1-β-glucose (4NPGlc) (53). HPLC-UV-MS analysis revealed conversion of the monodeglucosylated substrate (8) to a product with a retention time and mass consistent with that of avenacin A-1 (1) (Fig. 3D); therefore, we named this enzyme A. sativa transglucosylase 1 (AsTG1).

Next we evaluated the activity of AsTG1 by transient expression in N. benthamiana, as we had done previously for AsUGT91G16. Coexpression of AsbAS1/SAD1, AsCYP51H10/SAD2, and AsAAT1/UGT99D1 and the β-[1-2]-glucosyltransferase AsUGT91G16 yielded (4a) and its breakdown product (4b) (Fig. 3E), as seen previously (Fig. 2B). Inclusion of AsTG1 led to the appearance of a new, more polar peak (12b) on HPLC-charged aerosol detector (CAD)-MS chromatography with a mass consistent with addition of a hexose to (4a/4b), a retention time of 12.7 min, and m/z = 913.3 (Fig. 3E). In addition, the peak at 14.4 min corresponding to the retention time of (4b) was slightly larger and the mass data revealed another molecular ion (m/z = 913.3), indicating coelution of a new compound (12a) with (4b). AsTG1 activity is dependent on AsUGT91G16, since when AsTG1 was coexpressed with AsbAS1/SAD1, AsCYP51H10/SAD2, and AsAAT1/UGT99D1 in the absence of AsUGT91G16, only accumulation of 3β-(α-L-arabinopyranosyloxy)-12,13β-epoxy-16β-hydroxy-β-amin (3) was observed (SI Appendix, Fig. S11). AsTG1 activity was also dependent on the presence of the signal sequence (Fig. 3E).

Vacuum infiltration was used to scale up agroinfiltration of N. benthamiana and the peak at 12.7 min (12b) purified following extraction from 150 plants. NMR analysis confirmed the scaffold of (12b) as 12-keto,16β-hydroxy-β-amin, and verified that the trisaccharide moiety was consistent with that of avenacin A-1 (SI Appendix, Table S13). By analogy with the AsUGT91G16 products, (12b) is likely the C12 ketone degradation product of (12a) following loss of the C12-C13 epoxide.

Localization of fluorescence-tagged proteins in N. benthamiana leaves by confocal microscopy suggests that AsUGT91G16 is cytosolic, as expected for a GT1 family UGT, and that AsTG1 is localized to the vacuole (Fig. 3F and SI Appendix, Figs. S12 to S14). The localization of AsTG1 is dependent on its signal sequence, as deletion of this sequence resulted in colocalization with AsUGT91G16 in the cytosol and nucleus (Fig. 3F).

Avenacin-deficient oat mutants affected at the uncharacterized loci, Sad3 and Sad4, accumulate monodeglycosyl avenacins missing the 1,4-linked α-glucose (26). Sad3 is genetically linked to the avenacin biosynthetic gene cluster (22), and mutations at this locus result in complete loss of avenacin production (22, 26). In contrast, Sad4 is unlinked, sad4 mutants are only partially compromised in avenacin glycosylation, and these mutants are also affected in ability to glycosylate other oat metabolites (22, 26). Like the ugt91g16 mutants, sad3 and sad4
mutants have stunted root growth (26). DNA sequence analysis revealed that all 4 sad3 mutants had an SNP in the AsTG1 gene (SI Appendix, Table S14). When F2 progeny derived from a cross between the A. strigosa wild type and a sad3 mutant (#1139) were examined, the reduced root fluorescence phenotype behaved as a single recessive Mendelian mutation (SI Appendix, Table S15).

Fig. 3. Characterization of AsTG1. (A) The AsUGT91G16 and AsGH1 genes are adjacent in the oat genome. (B) Phylogenetic tree of AsGH1 and other plant GH1 sequences. Rice (49) and Arabidopsis thaliana (63) GH1 enzymes were included in the analysis, along with characterized sequences from other plant species (SI Appendix, Table S11). The phylogenetic clusters that include the rice and Arabidopsis sequences (At/Os 1 to 8) and the Arabidopsis clusters (At I and II) are labeled as designated in Opapissiri et al. (49), and the monocot plastidial β-glucosidases are labeled as in Ketudat Cairns et al. (64). Enzymes with reported transglycosidase activity (SI Appendix, Table S12) are indicated in bold type, and the A. strigosa GH1 is shown in red. The tree was constructed using the neighbor-joining method with 1,000 bootstrap replicates (% support for branch points is shown). The scale bar shows 0.1 substitutions per site at the amino acid level. (C) RT-PCR expression profile of AsGH1 in different oat tissues. AsbAS1/Sad1 and the AsGAPDH housekeeping gene were included as controls. RNA was extracted from 3-d-old A. strigosa seedlings. (D) Analysis of AsGH1 activity following heterologous expression in E. coli. HPLC-UV-M5 traces are shown following incubation of protein preparations with 4-nitrophenyl β-D-glucoside and avenacin A-1 lacking the 1,4-linked D-glucose (8). A boiled protein preparation served as a control. The data are representative of 2 separate experiments. Absorbance was measured at 357 nm. (E) Transient expression in N. benthamiana. HPLC-CAD-M5 traces for extracts from agroinfiltrated leaves are shown. Data are representative of 2 experiments. IS, internal standard (digitoxin). (F) Localization of AsUGT91G16 and AsTG1 in N. benthamiana leaves. Coexpression of an N-terminal AsUGT91G16 GFP fusion (GFP:AsUGT91G16) with a C-terminal mRFP fusion to AsTG1 (TG1:RFP) (Upper) and a C-terminal mRFP fusion to AsTG1 without the N-terminal signal peptide (NOSIG:TG1:RFP) (Lower). AsTG1:RFP localizes to the vacuole and the apoplast and does not colocalize with GFP:UGT91G16 in the cytoplasm. NOSIG-AsTG:RFP1 colocalizes with the GFP fusion to AsUGT91 in the cytoplasm and nucleus. GFP fusions are shown in green (Left), RFP fusions are shown in magenta (Middle), and merged images are shown in white (Right). Images were obtained at 3 d postinfiltration. (Scale bar: 35 μm.)
and cosegregated absolutely with the short root phenotype and the SNP, indicating that AsTG1 is synonymous with Sad3.

**Conclusion**

Here we have characterized the 2 enzymes required for the final glucosylation steps in biosynthesis of the antifungal defense compound, avenacin A-1. We show that 1 of these enzymes, AsUGT91G16, is a classical UDP sugar transferase belonging to family A of the GT1 family. In contrast, the enzyme that adds the final sugar (AsTG1) is an unusual vacuolar sugar transferase belonging to the glycosyl hydrolase family 1 (GH1) family of carbohydrate-active enzymes. Most of the characterized avenacin pathway enzymes are either known or likely to be endoplasmic reticulum-associated or cytosolic (11, 28, 33); however, the pathway and product avenacin A-1, like many other plant specialized metabolites, is localized in the vacuole (28). We have previously shown that AsSCPL1 (SAD7), the serine carboxypeptidase-like acyltransferase responsible for avenacin acylation, is located in the vacuole (28). Our results indicate that the partially glycosylated precursor \(3\beta-\{(\beta-D\text{-}glucopyranosyl-(1\rightarrow2)\alpha-L\text{-}arabinopyranosyl)\text{-}oxy\}\cdot12,13\beta\text{-}epoxy,16\beta\text{-}hydroxy-\beta\text{-}amyrin (4a), which is the product of AsUGT91G16, is transported from the cytosol to the vacuole, where it is then further glucosylated by AsTG1 and is also acylated (Fig. 4). This transport process is not specific to oat, since it also occurs in *N. benthamiana*. Several GH1 enzymes have previously been reported to glucosylate anthocyanins in plants using acyl sugars as sugar donors (50–52). Our study provides evidence of a role for GH1 enzymes in triterpene glycosylation and hints at a broader significance for this family of carbohydrate-active enzymes in glycosylation of plant natural products. Of note, it will be important to consider the alternative (non-UDP) sugar donor preferences and vacuolar localization of GH1 family sugar transferases when designing strategies for metabolic engineering of plant pathways in yeast and other microbial heterologous expression hosts. We further show that the AsUGT91G16 and AsTG1 genes are located next to each other
on a short genome contig, and that AsTG1 corresponds to Sad3, a locus that we previously defined genetically as being part of the avenacin biosynthetic gene cluster in a forward screen for avenacin-deficient mutants (22, 26). Therefore, these 2 genes form part of the extended gene cluster, although the physical distance between AsUGT91G16/AsTG1 and the other characterized pathway genes (SI Appendix, Fig. S1) is not known. The generation of a high-quality genome sequence for A. strigosa will shed light on the organization of the complete avenacin biosynthetic gene cluster.

Materials and Methods

Oat Material. The wild-type A. strigosa accession S75 and avenacin-deficient mutants have been described previously (22).

RNA Extraction and Analysis. cDNA was generated from 3-d-old seedlings of A. strigosa accession S75 for whole roots, root tips (terminal 0.2 cm of the root), root elongation zones (from 0.2 cm to the first root hair), and young leaves. Expression of candidate genes was analyzed by mRNA RT-PCR as described previously (11) and normalized relative to the transcript levels of the housekeeping gene encoding GAPDH (AsGAPDH). The gene-specific primers used for PCR amplification are listed in SI Appendix, Table S16. RNA-seq analysis was performed using material collected from 6 different A. strigosa tissues: root tips, whole roots, shoots, leaves, panicles, and spikelets. Root tips were harvested from 3-d-old seedlings as described above. Tissue from roots and shoots were collected from 3-wk-old glasshouse-grown plants. Mature leaves were harvested before earing, and spikelets and panicles (large spikes) were collected when the spacing between the flag leaf and top 2 seed leaves was 4–12 cm and 4 cm, respectively. For each tissue, ≥100 μg of total RNA was isolated using TRIzol reagent (Invitrogen). Poly(A)+ mRNA was purified using the DynaBeads mRNA Purification Kit (Life Technologies). Paired-end cDNA libraries were constructed using the RNA-Seq NGS Library Preparation Kit from Whole-Transcriptome Discovery (Gnomegen) with 10 cycles of PCR amplification. The resulting paired-end cDNA libraries were sequenced using the Illumina HiSeq 2500 system with a read length of 100 to 150 bp. More information on the RNA-seq analysis is provided in SI Appendix. Raw and assembled data are available at http://db.ncgr.ac.cn/oat/RNAseq.php. Self-organizing maps (SOMs) were implemented and visualized in R (54) using the Kohonen package (34, 35) (SI Appendix and Datasets S1 and S2).

Phylogenetic Analysis. For UGT sequences, amino acid sequences were deduced from the predicted full-length coding sequences of the A. strigosa UGTs identified in the SOM analysis (SI Appendix, Table S54). Representative amino acid sequences of characterized family 1 UGTs from other plant species (SI Appendix, Table S55) were obtained from the National Center for Biotechnology Information (NCBI) database and incorporated into the phylogenetic analysis. For each sequence, representative amino acid sequences of characterized plant GH1 family members (SI Appendix, Table S11) were obtained from the NCBI database. Protein sequences were aligned using MAFFT (http://mafft.cbrc.jp/alignment/software/). Unrooted trees were constructed in MEGA7 by the neighbor-joining method with 1,000 bootstrap replicates (55, 56).

Cloning and In Vitro Analysis of Candidate Glucosyltransferases. Expression constructs were created using Gateway technology (Invitrogen). Coding sequences (CDSs), including stop codons of A. strigosa UGT91G16 and AsTG1 (AsGH1), were amplified from oat genomic DNA and root tip cDNA, respectively, with a 2-step PCR method. The first amplification step used gene-specific primers (SI Appendix, Table S16) to attach partial Aatt adapters to the full-length CDS. The second amplification step attached the full Aatt site to each end of the CDS. Amplified fragments were purified from 1% agarose gels, digested with DpnII, and inserted into pDONR207 vector using BP Clonase II (Invitrogen). Then the plasmid CDS fragments were transferred into the pDONR207 vector using BP Clonase II Enzyme Mix (Invitrogen) according to the manufacturer’s instructions. To generate a plasmid with AsTG1 missing the predicted N-terminal signal sequence (AsTG1-NOSIG), 60 ng of the pDONR207-AsTG1 plasmid was used as a template with Fgw-nosigaSTG1 and Rgw-AsTG1 primers (SI Appendix, Table S16) by 2-step Gateway cloning. Before the PCR purification step, 1 μL of PfuTurbo (Agilent) was added to the PCR amplification mixture, and the mixture was then incubated at 37 °C for 1 h to digest the pDONR207-AsTG1 DNA template. Sequence-verified coding sequences were then transferred by LR clonase II into pH9GW (Invitrogen), a Gateway-compatible version of pET-28 encoding 9 N-terminal histidines (57). AsUGT99C8 (originally referred to as AsGT14n20) had previously been cloned into pH9GW as described above as part of an earlier investigation of avenacin glycosyltransferases (58). Constructs were transformed into E. coli Rosetta strain DE3 (Novagen), and recombinant partially purified enzymes were assayed for their ability to glycosylate 3β-(bisdeglucosyl-(1→2)-(1→4)-)avenacin A-1 (9) and 3β-(deglucosyl-(1→4)-)avenacin A-1 (9) (SI Appendix).

Transient Expression in N. benthamiana. N. benthamiana plants were grown in greenhouses maintained at 23 to 25 °C with 16 h of supplementary light per day, as described previously (59). Candidate glucosyltransferase CDSs were cloned into the pEAQ-HT-DEST1 expression vector (59) using LR clonase II (Invitrogen) according to the manufacturer’s instructions. Agroinfiltration was carried out as described previously (40, 59). In brief, expression constructs were transformed into Agrobacterium tumefaciens strain LBA4404 and the strains infiltrated into N. benthamiana leaves. To coexpress combinations of genes, A. tumefaciens strains containing different expression constructs were diluted to 0.9 OD600 and mixed in equal volumes to result in a final concentration per strain of 0.15 OD600. A strain containing a construct for expression of the feedback-insensitive form of HMG-CoA reductase, thHMRG, was included in all combinations to increase transcription of downstream genes. After an overnight incubation, a green fluorescent protein (GFP) expression construct was used as a “filler” to maintain the same concentration of bacteria in all infiltration combinations. Details of metabolite analysis, purification, and structural determination by NMR are provided in SI Appendix.

Characterization of Oat Mutants and Segregating F2 Populations. Oat genomic DNA was extracted based on the method described by Pallota et al. (60). The AsUGT91G16 gene was sequenced in 47 uncharacterized sodium azide-generated avenacin-deficient mutants (22, 31). The AsTG1 gene was sequenced in 4 independent sad3 mutants (26). The primers used for amplification and sequencing are listed in SI Appendix, Table S16. The purified PCR products were sequenced by GATC Biotech. To analyze segregating progeny, 3-d-old seedlings of F2 progeny from crosses between mutant lines and a wild-type variety were typed for rol gene expression by PCR and then transferred to soil in 96-well trays under glasshouse conditions. Genomic DNA was extracted from 2-wk-old seedlings, and genotyping was performed using the amplification and sequencing primers listed in SI Appendix, Table S16.

Metabolite Analysis of Oat Root Extracts. Roots were harvested from 3-d-old seedlings of wild-type and mutant A. strigosa lines and then flash-frozen in liquid N2. Aliquots (~25 mg) were extracted in 0.5 mL of 80% methanol with two 3-mm tungsten beads at 1,400 rpm at 25 °C for 1 h, after which the temperature was increased to 42 °C for a further 30 min. Samples were centrifuged briefly, and 0.45 mL of supernatant was partitioned twice with 0.3 mL of hexane. Aliquots (0.2 mL) of the methanolic fractions were dried in a Genevac EZ-2 Elite centrifugal evaporator, resuspended in 100 μL of methanol, and filtered through a 0.2-μm filter (Costar, Sigma-Aldrich), after which 50 μL of the filtrate was combined with 50 μL of 50% methanol. Samples were analyzed by HPLC-CAD-UV-MS using a 50 × 2.1-mm, 2.6-μm Kinetex XB-C18 column (Phenomenex) with a column oven temperature of 30 °C. Detection was by CAD (Corona Ultra RS; Dionex); UV/Vis absorbance (Shimadzu SPD-M20A), collecting spectra from 200 to 500 nm; and electrospray MS (Shimadzu LC-2020 dual-source MS), collecting in positive and negative modes from m/z 50 to 1,500. The gradient was run at 0.3 mL/min with 100% water as Buffer A and 100% acetonitrile as Buffer B as follows: 20% Buffer B from 0 to 3 min, 20 to 60% Buffer B from 3 to 28 min, 60 to 100% Buffer B from 28 to 30 min, a linear gradient between 30 and 3 min, 100 to 20% Buffer B from 33 to 34 min, and held at 20% Buffer B until 35 min.

Pathogenicity Tests. Pathogenicity tests to evaluate root infection of the A. strigosa wild-type and avenacin-deficient mutant lines with the fungal pathogen G. graminis var. tritici (isolate T5) were performed as described previously (22). Seedlings were scored for root lesions at 21 d after inoculation using a 7-point scale (11).

Subcellular Localization. To make N- and C-terminal AsUGT91G16 fusions with GFP, AsUGT91G16 was amplified from the pH9-GW-AsUGT91G16 plasmid with primers AsUGT91-NTGW and Rgw-UGT91 or with primers Fgw-GTUGT91 and Rgw-UGT91-NOSTOP (SI Appendix, Table S16) and then cloned into pMDCA5 and pMDC83, respectively (61). To make C-terminal
protein fusions with GFP or RFP, AsTG1 and AsTG1-NOSIG (i.e., AsTG1 lacking the N-terminal signal sequence) were amplified from pDEST1-AsTG or pGEG. AsTG1 and AsTG1-NOSIG were transformed into E. coli using Fgw-AsTG-NOSIG-FULL and Fgw-AsTG-NOSTOP-FULL or primers Fgw-AsTG-NOSIG-FULL and Rgw-AsTG-NOSTOP-FULL (61), respectively, and then cloned into pB7R-WG2 (62) for GFP constructs and pMDC83 (61) for GFP constructs. Fluorescent protein fusion constructs were verified using GFP- or RFP-specific sequencing primers (SI Appendix, Table S1).

A. tunefaciens strain GV3101 was used for all fluorescent protein fusion assays to assess the enzymatic activity of the fluorescent protein fusion constructs, agroinfiltrations were carried out following the method described for transient expression in N. benthamiana. Metabolites were detected by HPLC using a 50 × 2.1-mm, 2.6-µm Kinetex XB-C18 column (Phenomenex) with a column oven temperature of 25 °C. Detection was by CAD (Corona Ultra RS, Dionex). The gradient was run at 0.3 mL/min with 100% water as Buffer A and 100% acetonitrile as Buffer B with 0 to 1.5 min, 10 to 50% Buffer B from 1.5 to 21 min, 50 to 95% Buffer B from 21 to 21.5 min, a linear gradient between 21.5 and 23.5 min, 95 to 10% Buffer B from 23.5 to 24 min, and held at 10% until 25 min.

The subcellular localization of fusion constructs was assessed in N. benthamiana leaves. Agroinfiltration was carried out as described above but with the following changes: rhizobacteria were diluted to 0.6 OD600 before mixing in equal volumes to give a final concentration of 0.2 OD600, except for combinations of GFP-UGT191G16 with AsTG1-RFP and NOSIG:AsTG1-RFP, in which strains were further diluted to achieve a final concentration of 0.04 OD600 per strain. Leaves were imaged with a Leica TCS SPBX confocal microscope after the number of days indicated.

ACKNOWLEDGMENTS. This work was supported by the UK Biological Sciences Research Council (BBSRC) Doctoral Training Programme (A. Orme); the joint Engineering and Physical Sciences Research Council/Biotechnological and and BBSRC-funded OpenPlant Synthetic Biology Research Centre Grant BB/L011003/1 (to M.I.S. and A.O.; the Centre of Excellence for Plant and Microbial Sciences, established between the John Innes Centre and the Chinese Academy of Sciences and funded by the BBSRC and the Chinese Academy of Sciences; National Natural Science Foundation of China Grant 31771661; “Molecular, Diverse, Self–future Crops“ (to B.H.); BBSRC Strategic Research Programme Grant BB/P012523/1, “Molecules from Nature,” and the John Innes Foundation (A. Osbourn and R.M.).
52. T. Miyahara et al., Isolation of anthocyanin 7-O-glucosyltransferase from Canterbury bells (Campanula medium). Plant Biotechnol. 31, 555–559 (2014).

53. S. Luang et al., Rice Os9BGlU31 is a transglucosidase with the capacity to equilibrate phenylpropanoid, flavonoid, and phytohormone glycoconjugates. J. Biol. Chem. 288, 10111–10123 (2013).

54. R Core Team, R: A language and environment for statistical computing, version 3.5.2. https://www.R-project.org/. 2017. Accessed 20 December 2018.

55. N. Saitou, M. Nei, The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425 (1987).

56. S. Kumar, G. Stecher, K. Tamura, MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874 (2016).

57. P. E. O'Maille et al., Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases. Nat. Chem. Biol. 4, 617–623 (2008).

58. T. Louveau, “Investigation of glycosyltransferases from oat”, PhD thesis, University of East Anglia, UK (2014).

59. F. Sainsbury, P. Saxena, K. Geisler, A. Osbourn, G. P. Lomonossoff, Using a virus-derived system to manipulate plant natural product biosynthetic pathways. Methods Enzymol. 517, 185–202 (2012).

60. M. A. Pallotta et al., “Marker-assisted wheat breeding in the southern region of Australia” in Proceedings of the Tenth International Wheat Genetics Symposium, N. E. Pogna, M. Romano, E. A. Pogna, Z. Galterio, Eds. (Istituto Sperimentale per la Cerealicultura, 2003) pp. 789–791.

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

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

63. Z. Xu et al., Functional genomic analysis of Arabidopsis thaliana glycoside hydrolase family 1. Plant Mol. Biol. 55, 343–367 (2004).

64. J. R. Ketudat Cairns et al., Protein-carbohydrate interactions leading to hydrolysis and trannglycosylation in plant glycoside hydrolase family 1 enzymes. J. Appl. Glycosci. 59, 51–62 (2012).