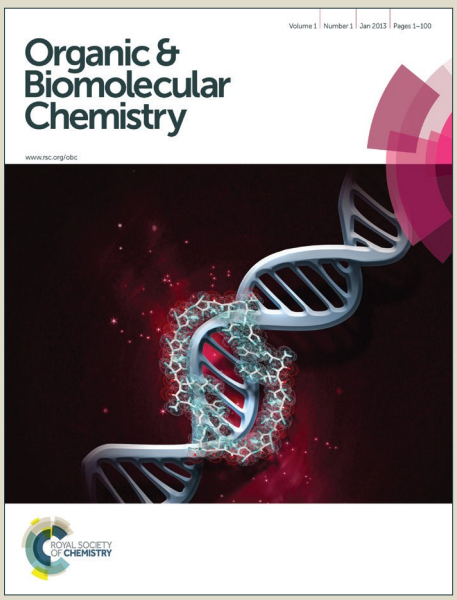
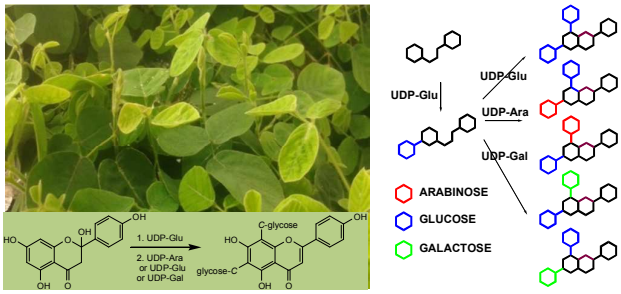
Organic & Biomolecular Chemistry

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/obc



Biosynthesis of allelopathic di-*C*-glycosylflavones in the plant *D. incanum*, occurs by UDP-glucosylation of a natural or analogue 2-hydroxyflavonoid substrate, followed by a second glycosylation with UDP-Glu, UDP-Ara, or UDP-Gal.

The biosynthesis of allelopathic di-*C*-glycosylflavones from the roots of *Desmodium incanum* (G. Mey.) DC.

Bing Hao,[a,c] John C. Caulfield,[b] Mary L. Hamilton,[b] John A. Pickett,[b] Charles A. O. Midega,[c] Zeyaur R. Khan,[c] Junru R. Wang,[a*] and Antony M. Hooper.[b*]

[a]College of Science, Northwest A&F University, Yangling, Shaanxi 712100, China

[b]Department of Biological Chemistry and Crop Protection, Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ, UK

[c]International Centre of Insect Physiology and Ecology, P.O. Box 30772, Nairobi, Kenya

Abstract

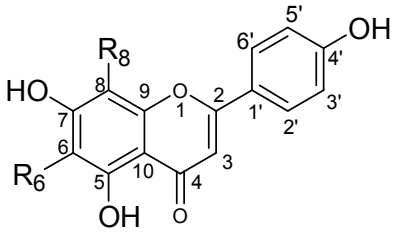
The allelopathic root exudate of the drought-tolerant subsistence cereal intercrop *D. incanum*, protecting against the parasitic weed *Striga hermonthica*, comprises a number of di-*C*-glycosylflavones specifically containing *C*-glucosyl, *C*-galactosyl and *C*-arabinosyl moieties. Here we demonstrate that the biosynthesis of all compounds containing a *C*-glucose involves *C*-glucosylation of 2-hydroxynaringenin with subsequent *C*-galactosylation, *C*-glucosylation or *C*-arabinosylation. In addition, the crude soluble enzyme extract converts two fluorinated 2-hydroxyflavanone analogues to corresponding mono- and di-*C*-glycosylflavones demonstrating that some differences in C-ring substitution can be tolerated by the plant enzymes. Elucidating the biosynthesis of these *C*-glycosylflavones (CGFs) has the potential to open up opportunities for transferring the enzymic and genetic basis for the *S. hermonthica* inhibiting allelopathic trait to food crop plants.

* Corresponding authors. Tel.: +44 1582 763133; fax: +44 1582 762595.
E-mail address: tony.hooper@rothamsted.ac.uk; john.caulfield@rothamsted.ac.uk; mary.hamilton@rothamsted.ac.uk; john.pickett@rothamsted.ac.uk; bing.hao@rothamsted.ac.uk; cmidega@icipe.org; zkhan@icipe.org.

Introduction

Parasitic weeds, or witchweeds, from the *Striga* genus devastate crops, specifically subsistence cereal crops, such as maize (*Zea mays* L.), sorghum (*Sorghum bicolor* (L.) Moench) and rain-fed rice, in sub-Saharan African where resource poor farmers are least well equipped, through lack of access to funds or technology, to deal with this biotic constraint to food production. Infestations of *Striga* can be so severe that communities are driven to abandon potentially productive agricultural land.[1] The research consortium of the International Centre of Insect Physiology and Ecology (*icipe*) and Rothamsted Research has developed an intervention called the push-pull technology that effectively inhibits *Striga* damage through the use of an intercrop from the *Desmodium* genus (animal forage legumes) that inhibits these parasites, a technology which has been adopted by more than 110,000 farmers in eastern Africa.[2-5] The mechanism of *Striga* suppression has been shown to arise from allelopathic root exudates from *Desmodium* that, along with the cereal crop, stimulate the germination of *Striga*,[6] but more importantly inhibit the subsequent development of the germinated parasitic seed,[7,8] so that in the field, almost no parasitism is observed.[9,10] In order to increase the resilience of the intervention to climate change and expand application of the push-pull technology into more arid areas or into areas already suffering as a consequence of climate change, we have identified *D. incanum* (G. Mey.) DC. as a suitable intercrop that provides the same protective chemistry as observed previously for non-drought tolerant species such as *D. uncinatum*.[11] The *Desmodium* intercrop provides a benefit for the farmer, not only in preventing *Striga* parasitism, but also in insect pest (cereal stemborer) control and providing fodder for zero-grazed livestock while improving soil quality through addition of organic carbon and fixed N.[12,13] However, there is a demand

for a human edible intercrop that can provide *Striga* control. Understanding the biosynthesis of the chemicals that comprise the allelopathic root exudates provides the foundation upon which subsequent identification of the enzymic and genetic basis for the trait can be elucidated, so that it may be introduced to crop plants themselves.[14] Our previous work has characterised the root exudate profile of a number of *Desmodium* species in hydroponic solution and shown that as the plants mature, their root exudates become similar in content with the major components being di-*C*-glycosylflavones possessing *C*-linked glucose, galactose and arabinose (Table 1).[11] We previously identified the precursor for *C*-glycosylflavone biosynthesis in *D. uncinatum* and *D. intortum* to be 2-hydroxynaringenin which is glucosylated with UDP-glucose[15] and subsequently arabinosylated with UDP-arabinose before dehydration of the 2-hydroxyflavanone product breaks the rotational flexibility and fixes the flavone A and B rings to yield the two 6,8-di-*C*-glycosylated isomers, isoschaftoside (**4**) and schaftoside (**5**).[16] There has been little previously reported regarding the biosynthesis of *C*-glycosylflavones and the multiple decoration of flavones required for di-*C*-glycosylflavones. This is despite the importance of this class of molecules in plant ecology as allelochemicals against *Striga*,[8] as the basis of protection against infection,[17] insect herbivory,[18,19] plant interactions with arbuscular mycorrhizal fungi[20] as well as many pharmacological properties.[21] Thus, the elucidation of the biosynthesis of the entire range of di-*C*-glycosylflavones found in *D. incanum* is now targeted.



	R6	R8
1	*C*-β-D-galactose	*C*-β-D-glucose
2	*C*-β-D-glucose	*C*-β-D-glucose
3	*C*-β-D-glucose	*C*-β-D-galactose
4	*C*-α-L-arabinose	*C*-β-D-glucose
5	*C*-β-D-glucose	*C*-α-L-arabinose
6	*C*-α-L-arabinose	*C*-β-D-galactose
7	*C*-β-D-galactose	*C*-α-L-arabinose

Table 1 Di-*C*-glycosylflavones characterised in the root exudates of *D. incanum*

Results and Discussion

First glycosylation of 2-hydroxyflavanone precursor

In previous work the first committed step to *C*-glycososylflavone biosynthesis in *D. uncinatum* was determined to be the *C*-glucosylation of the flavanone, 2-hydroxynaringenin.[15] We tested the first committed glycosylation step in *D. incanum* in the same way as it would be expected to be similar. The tetradeuterated substrate [2',3',5',6'-$^{2}H_4$]2-hydroxynaringenin was therefore synthesised as previously described.[15] The plant enzyme mixture used for the assays was the protein fraction soluble at 40% $(NH_4)_2SO_4$ that precipitated at 80% $(NH_4)_2SO_4$. Desalting provided the soluble protein fraction used in all assays. Both root and leaf proteins were isolated using the same protocol and were incubated separately with UDP-glucose and synthetic [2',3',5',6'-$^{2}H_4$]2-hydroxynaringenin and the assay quenched with dilute acid to promote chemical dehydration of the 2-

hydroxyflavanone products to the stable flavones. Control assays using boiled protein showed the presence of small quantities of di-*C*-glycosylflavones as they are water soluble and are co-isolated with the assay protein mixture. HPLC analysis of both the native root and leaf protein assays showed the presence of an enhanced peak for the *C*-glycosylflavone isovitexin (**9**) and an extra peak with the same retention time as the isomer vitexin (**8**), available as a standard. LCMS analysis using negative ion ESIMS revealed molecular ions of $[M+4-H]^-$ 435 for both vitexin and isovitexin which demonstrated clearly the incorporation of four deuterium atoms through *C*-glucosylation of the labelled 2-hydroxyflavanone substrate (Figure 1).

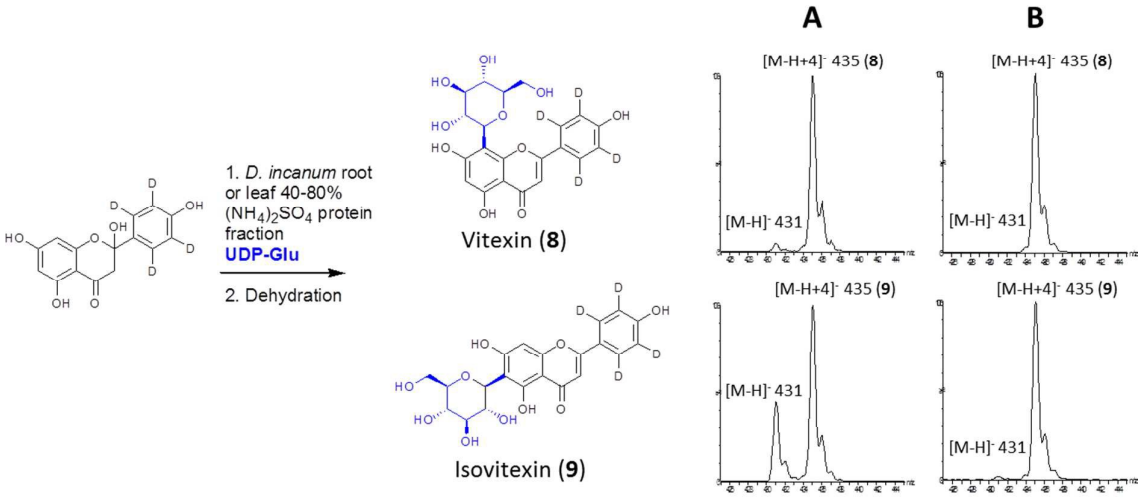


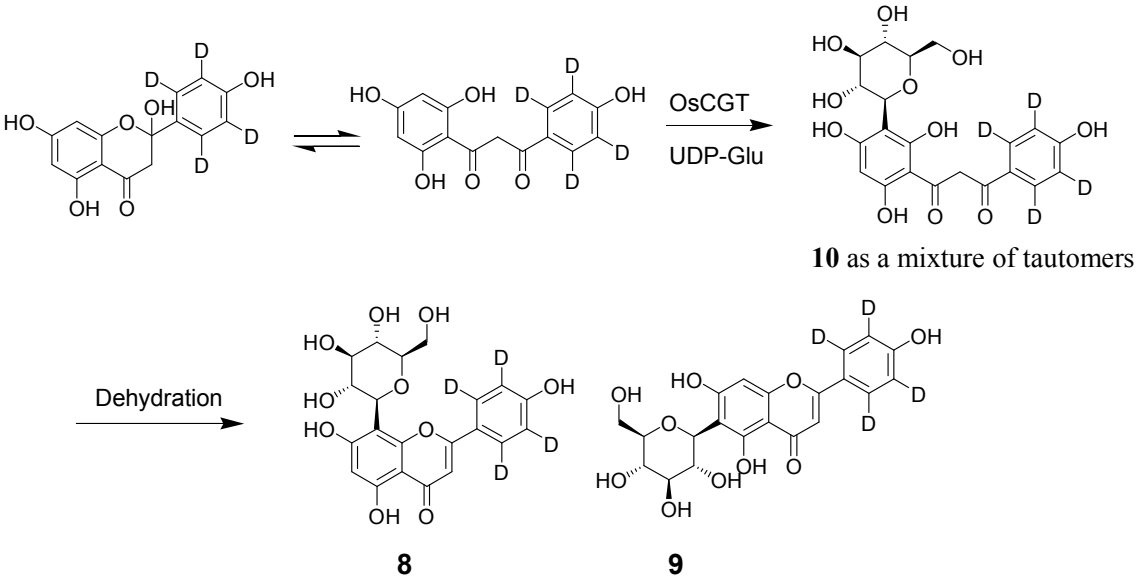
Fig. 1 Incorporation of [2',3',5',6'-$^2H_4$]2-hydroxynaringenin and UDP-glucose into vitexin (**8**) and isovitexin (**9**) by the soluble leaf (**A**) and root (**B**) protein fraction of *D. incanum*. Ions labelled with four deuterium atoms in the negative ESI mass spectra are at $[M-H+4]^-$ 435.

No molecular ion $[M+4-H]^-$ 435 was observed in the control experiments showing the transformation is enzyme mediated. Molecular ions visible in the spectrum at $[M-H]^-$ 431 represent small amounts of metabolite co-isolated from the plant tissue with the assay proteins. The flavone, apigenin, as expected, was not a substrate for glucosylation. When

the experiments were repeated with UDP-galactose, assays with both the native root and leaf protein produced the same results, generating $^{2}H_4$-labelled vitexin and isovitexin and no deuterium incorporation was observed when the protein mixture was denatured by boiling. The production of labelled vitexin and isovitexin suggested the UDP-hexoses, UDP-glucose and UDP-galactose, were interconverting through a UDP-glucose 4-epimerase. As neither labelled vitexin nor isovitexin were converted to a *C*-galactosylapigenin when reintroduced to the enzyme mixture and as the assay generated no additional peaks consistent with a *C*-galactosylapigenin, it was concluded that 4-epimerase activity was the likely origin of the metabolites rather than isomerisation of the sugar after *C*-galactosylation to *C*-glucose. This phenomenon was investigated further and it was demonstrated to be the case (see below). Increasing the cone-voltage during LCMS analysis showed the peaks identified as vitexin and isovitexin through standards were indeed *C*-glucosylated as fragments of $[M+4-H-90]^-$ and $[M+4-H-120]^-$, characteristic of *C*-linked glycosyl groups attached to flavones, were detected (*Supplementary figure 1*).[22] When 2-hydroxynaringenin was incubated with the soluble root or soluble leaf protein fraction and UDP-arabinose, no ion of mass $[M-H]^-$ 401 could be detected and so no mono-*C*-arabinosylflavone was biosynthesised in the assay. In summary, the only initial *C*-glycosylation step observable on 2-hydroxynaringenin in these assays was *C*-glucosylation with UDP-glucose.

Second glycosylation of 2-hydroxyflavanone precursor

Assays to examine the second glycosylation required the product of the first glucosylation, intermediate **10** (Scheme 1).



Scheme 1 *In vitro* production of the labelled mono-*C*-glycosyl-2-hydroxyflavanone intermediate (**10**) from 2-hydroxynaringenin and the flavone products from chemical dehydration.

To obtain this intermediate, synthetic [2',3',5',6'-$^{2}H_4$]2-hydroxynaringenin was incubated with recombinant His-tagged OsCGT and UDP-glucose overnight before quenching with ice cold methanol, which prevents dehydration to the flavone.[16,23] Insoluble material was removed by centrifugation, the solution concentrated under a stream of nitrogen and redissolved in water for subsequent assays. In the first instance, substrate **10** was incubated with UDP-arabinose and *D. incanum* soluble protein fraction from root or leaf tissue. In both root and leaf enzyme assays, LCMS analysis showed the incorporation of four deuterium labels demonstrated by $[M+4-H]^-$ 567 ions that were identified, by comparison with fully characterised natural product standards, as isoschaftoside (**4**) and schaftoside (**5**) (Figure 2).

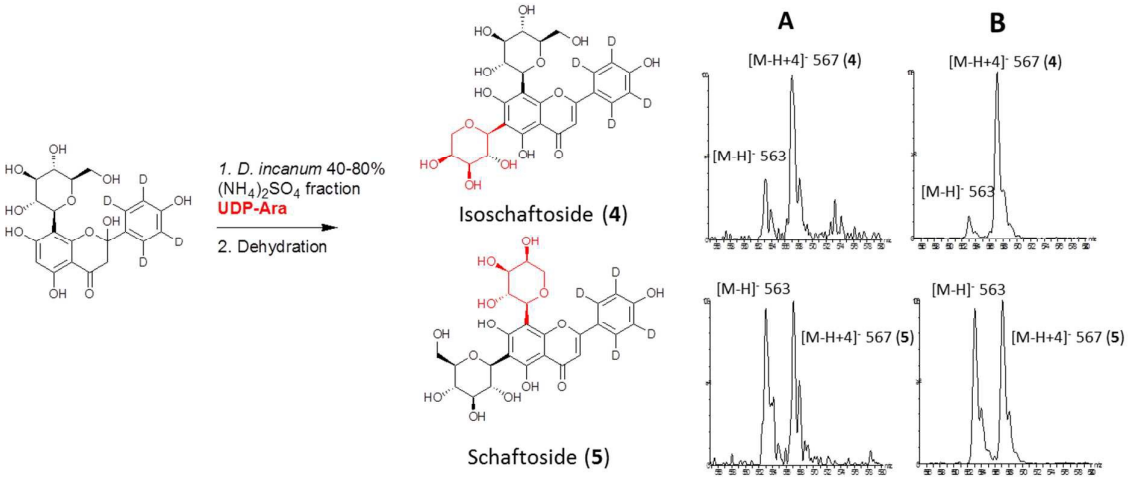


Fig. 2 Incorporation of [2',3',5',6'-$^{2}H_4$]6-*C*-glucosyl-2-hydroxynaringenin and UDP-arabinose into isoschaftoside (**4**) and schaftoside (**5**) by the soluble leaf (**A**) and root (**B**) protein fraction of *D. incanum*. Ions labelled with four deuterium atoms in the negative ESI mass spectra are at $[M-H+4]^-$ 567.

Molecules with molecular ion $[M-H]^-$ 563 are explained by the co-isolation of unlabelled **4** and **5** from plant tissue with the assay proteins. The *C*-linkage was verified by increasing the ESIMS cone voltage to cause fragmentation and generate characteristic ions at $[M+4-H-60]^-$, $[M+4-H-90]^-$,$[M+4-H-120]^-$ and $[M+4-H-210]^-$ proving the *C*-linkage of both arabinose (loss of 60 and 90 amus) and glucose moieties (loss of 90 and 120 amus, *Supplementary figure 2*). When using boiled protein, no labels were incorporated showing the transformation is enzyme mediated. Neither vitexin (**8**) nor isovitexin (**9**) were substrates, so the transformation requires the mono-glucosylated 2-hydroxyflavonone intermediate **10**.

Incubation of the *D. incanum* root or leaf soluble protein fraction with glucosylated substrate **10** and UDP-glucose generated three labelled products characterised through

natural product standards. These were $^{2}H_4$-labelled 6-*C*-galactosyl-8-*C*-glucosylapigenin **1**, 6,8-di-*C*-glucosylapigenin (vicenin-2) **2** and 6-*C*-glucosyl-8-*C*-galactosylapigenin **3** (Figure3).

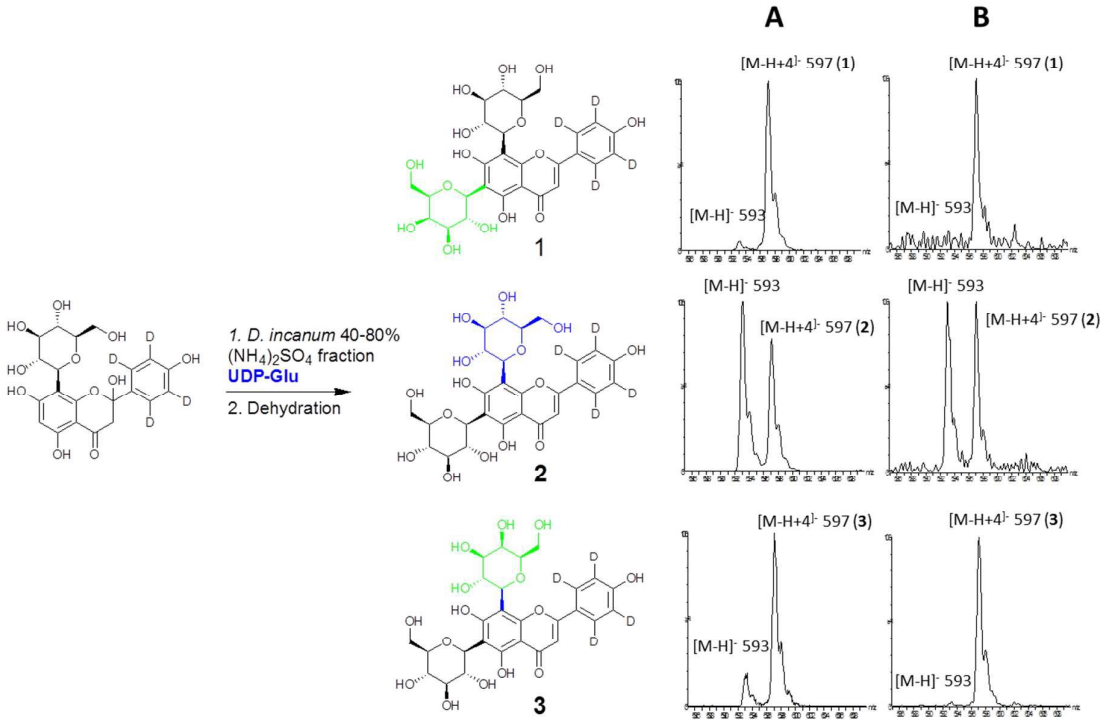


Fig. 3 Incorporation of [2',3',5',6'-$^{2}H_4$]6-*C*-glucosyl-2-hydroxynaringenin and UDP-glucose into 6-*C*-galactosyl-8-*C*-glucosylapigenin **1**, 6,8-di-*C*-glucosylapigenin (vicenin-2) **2** and 6-*C*-glucosyl-8-*C*-galactosylapigenin **3** by the soluble leaf (**A**) and root (**B**) protein fraction of *D. incanum*. Ions labelled with four deuterium atoms in the negative ESI mass spectra are at $[M-H+4]^-$ 597.

The result was the same using root tissue or leaf tissue as the protein source. Again, unlabelled material was observed with molecular ion $[M-H]^-$ 593 due to co-isolation of metabolites from plant tissue with the assay proteins. When the incubation was repeated with UDP-galactose, the same three compounds, **1-3** were similarly labelled, implying a UDP-glucose 4-epimerase might be interconverting the UDP-hexoses during the incubation (Figure 4).

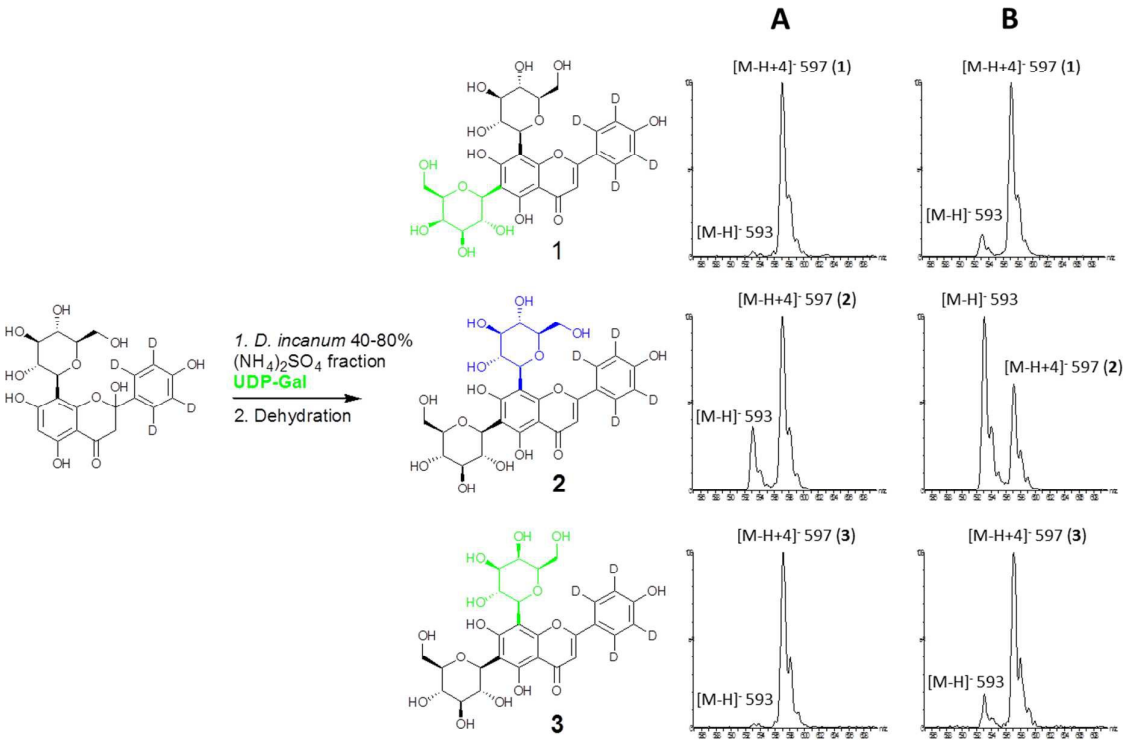


Fig. 4 Incorporation of [2',3',5',6'-$^{2}H_4$]6-*C*-glucosyl-2-hydroxynaringenin and UDP-galactose into 6-*C*-galactosyl-8-*C*-glucosylapigenin **1**, 6,8-di-*C*-glucosylapigenin (vicenin-2) **2** and 6-*C*-glucosyl-8-*C*-galactosylapigenin **3** by the soluble leaf (**A**) and root (**B**) protein fraction of *D. incanum*. Ions labelled with four deuterium atoms in the negative ESI mass spectra are at $[M-H+4]^-$ 597.

As expected, arabinosylated natural products were not labelled in these assays, only hexosylated di-*C*-glycosylflavones. In control experiments, boiled proteins did not cause incorporation of a second glycosyl group and neither vitexin nor isovitexin were substrates for the second glycosylation which required the 2-hydroxyflavanone intermediate **10** as the substrate.

Interconversion of UDP-Glu and UDP-Gal by *D. incanum* protein mixture

HPLC analysis of the root and leaf protein mixtures used for the assays showed they did not contain any UDP-glucose or UDP-galactose. In addition, when UDP-hexoses were added to the crude protein mixture before desalting, they were removed by the desalting process (PD-10 column) so UDP-hexoses could not have come from the plant, only from addition to the assays. The inter-conversion of UDP-hexoses was proven by incubating the assay protein fraction with UDP-glucose and detecting both UDP-glucose and UDP-galactose. As expected when the same assay was performed with UDP-galactose, both UDP-galactose and UDP-glucose were detected (*Supplementary figure 3*). However, UDP-glucose could be a contaminant as it was used in the preparation of substrate **10** and could be a source of UDP-glucose in the experiments where UDP-galactose was the added sugar donor. To confirm that UDP-hexoses incorporated were coming from those added in the second incubation and not contaminants from the incubation of 2-hydroxynaringenin with OsCGT that generated **10**, the assays testing the second glycosylation were repeated using UDP-α-D-[UL-$^{13}C_6$]glucose and UDP-α-D-[UL-$^{13}C_6$]galactose. The presence of molecular ion $[M+10-H]^-$ 603 due to four deuterium and six carbon-13 isotopes demonstrated incorporation of the UDP-hexose into all three di-*C*-hexosylflavones compounds **1-3** as expected. This was the case when using either UDP-[UL-$^{13}C_6$]hexose and when using either root tissue protein or leaf tissue protein demonstrating the interconversion of the hexoses in the assays (Figure 5).

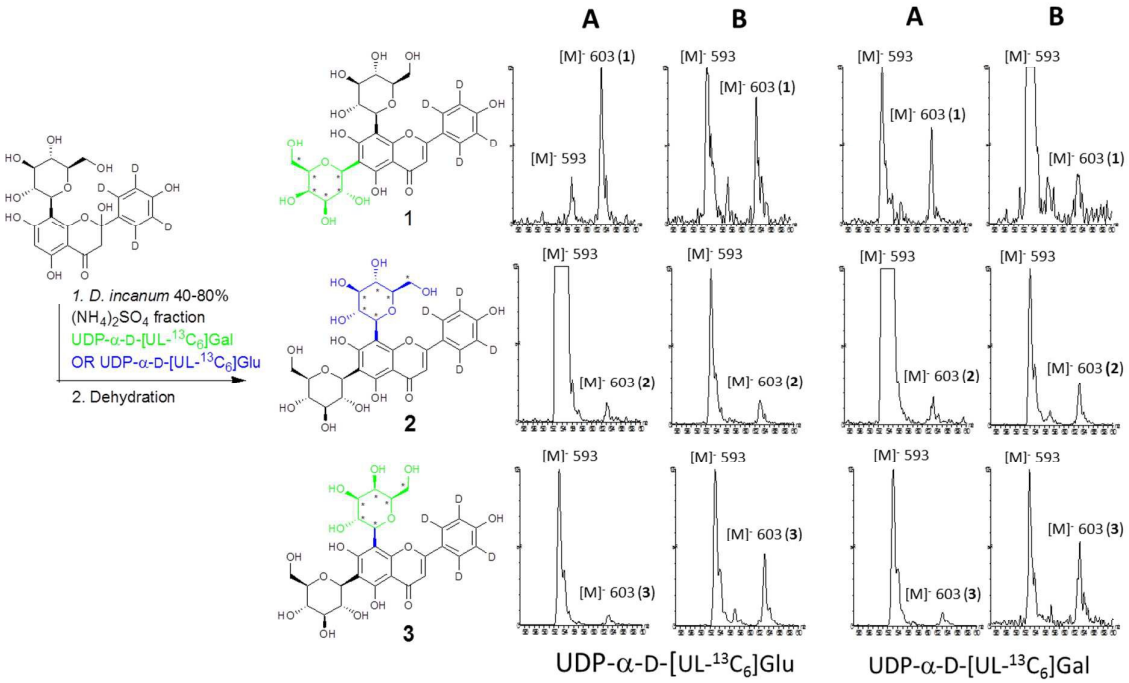


Fig. 5 Incorporation of [2',3',5',6'-$^{2}H_4$]6-*C*-glucosyl-2-hydroxynaringenin and UDP-α-D-[UL-$^{13}C_6$]glucose (**left**) or UDP-α-D-[UL-$^{13}C_6$]galactose (**right**) into 6-*C*-galactosyl-8-*C*-glucosylapigenin **1**, 6,8-di-*C*-glucosylapigenin (vicenin-2) **2** and 6-*C*-glucosyl-8-*C*-galactosylapigenin **3** by the soluble leaf (**A**) and root (**B**) protein fraction of *D. incanum.* Ions labelled with four deuterium atoms and six ^{13}C atoms in the negative ESI mass spectra are at $[M-H+10]^-$ 603 and unlabelled metabolites at $[M-H]^-$ 593.

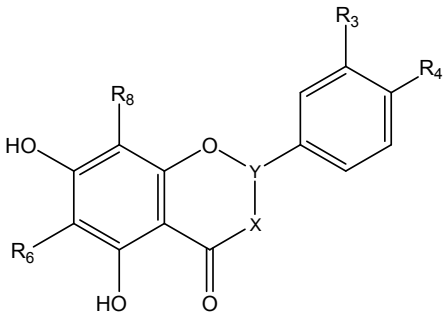
Where visible, molecular ions at $[M+4-H\}^-$ 597 are a result of incorporation derived from trace amounts of unlabelled UDP-glucose contaminant and tetradeuterated **10** into **1**, **2** and **3**.

To demonstrate further the interconversion of the UDP-hexoses occurs before incorporation, $^{2}H_4$-labelled **1**, **2** and **3** were isolated from previous assays by HPLC and re-introduced into enzyme assays without addition of any UDP-hexose cofactors. In all cases (root and leaf tissue), no label was found in any other metabolite than that introduced so

demonstrating these di-*C*-glycosylflavones cannot be interconverted in the assay and so interconversion of the hexoses is most likely due to a UDP-glucose 4-epimerase before incorporation into the metabolites. Interconversion of UDP-glucose and UDP-galactose by the UDP-glucose-4-epimerase requires NADH which must stay bound to the enzyme during the protein isolation process.

Novel mono- and di-*C*-glycosides biosynthesised by *D. incanum* soluble proteins

The results of the enzyme assays shown in Figure 5 reveal the presence of natural di-*C*-glycosyflavone metabolites, co-isolated with the assay proteins, with molecular ions $[M-H]^-$ 593 that dwarf the desired labelled molecular ions. This is because of the very small quantity of multiply labelled UDP-sugars used. To remove this interference and more clearly demonstrate incorporation of the UDP-hexose as the second glycosyl donor, two fluorinated analogues of substrate **10** were prepared containing slight difference on the C ring. 2-Hydroxyflavanones **11** and **12** were prepared by modification of the synthetic route described previously (Table 2).[16]



R_3	R_4	R_6	R_8	X-Y	
H	F	H	H	CH_2-C(OH)	**11**
H	F	*C*-β-D-glucoside	H	CH_2-C(OH)	**11a**
H	F	H	*C*-β-D-glucoside	CH=C	**11b**
H	F	*C*-β-D-glucoside	H	CH=C	**11c**
F	OH	H	H	CH_2-C(OH)	**12**

F	OH	*C*-β-D-glucoside	H	CH_2-C(OH)	**12a**
F	OH	H	*C*-β-D-glucoside	CH=C	**12b**
F	OH	*C*-β-D-glucoside	H	CH=C	**12c**

Table 2 Analogue 2-hydroxyflavanones, 6-*C*-glucosyl-2-hydroxyflavanones and *C*-glycosylflavones.

Incubation of **11** and **12** with root or leaf protein preparations from *D. incanum* gave two new metabolites in each case with molecular ions $[M-H]^-$ 433 consistent with the expected analogues **11b** and **11c** and $[M-H]^-$ 449 for **12b** and **12c** (Figure 6).

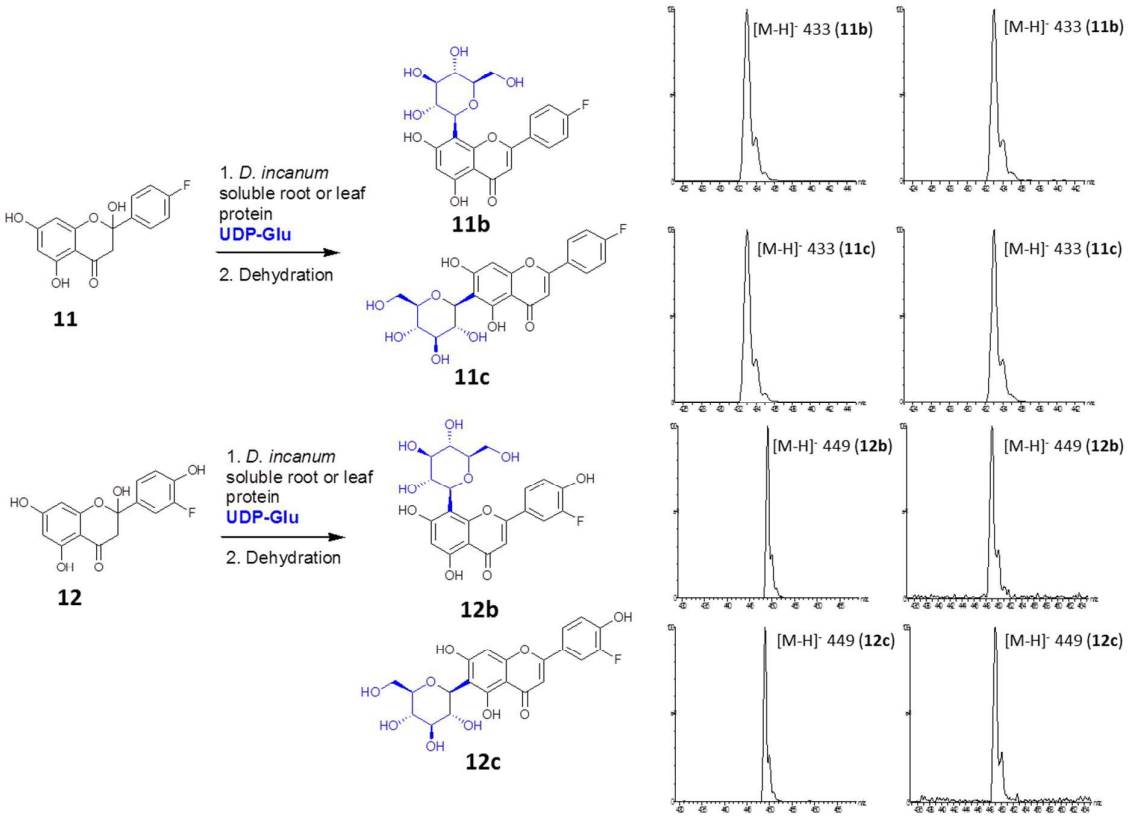
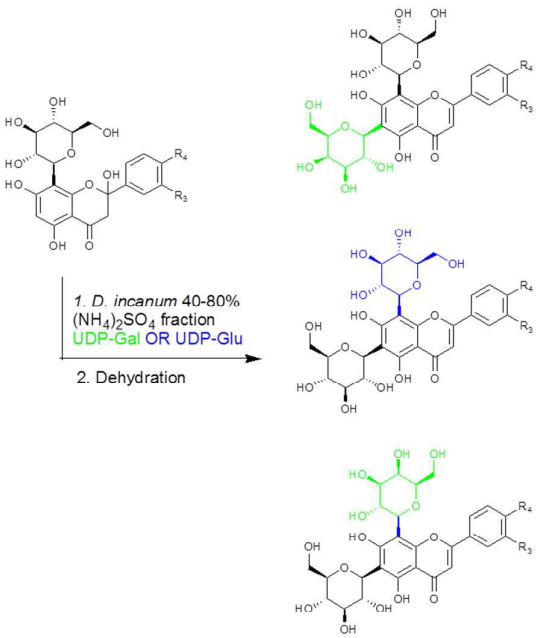


Fig. 6 Incorporation of **11** or **12** and UDP-glucose into C-glycosylflavones by the soluble leaf (**A**) and root (**B**) protein fraction of *D. incanum.* Molecular ions of *C*-glucosylated, fluorinated substrates in the negative ESI mass spectra are at $[M-H]^-$ 433 (**11b**, **11c**) and 449 (**12b**, **12c**).

The metabolites could also be fragmented in high cone voltage ESIMS experiments to demonstrate a *C*-linked glucose (*Supplementary figure 4*). Incubation of **11** and **12** with recombinant OsCGT and UDP-glucose generated enough material to characterise **11b**, **11c**, **12b** and **12c** by NMR spectroscopy after HPLC purification and confirm the metabolite structures and show that both *D. incanum C*-glycosyltransferase proteins and OsCGT can *C*-glucosylate **11** and **12**. The monoglucosyl-2-hydroxyflavanone substrates **11a** and **12a** were therefore accessible through incubation of **11** or **12** with OsCGT and UDP-glucose and so were prepared. Substrate **11a** was incubated with UDP-glucose or UDP-galactose and the root protein fraction of *D. incanum*. In both cases, LCMS analysis of the assay products showed the presence of three new compounds (**13-15**) with molecular ion $[M-H]^-$ 595 (*Supplementary figure 5A*). The products were shown to contain two *C*-linked hexoses by fragmentation through high cone voltage ESIMS (*Supplementary figure 5*) and so deduced to be the analogues of the di-*C*-glycosylflavones **1-3** (Figure 7).



R_3	R_4	substrate	product
H	F	11a	13,14,15
F	OH	12a	16,17,18

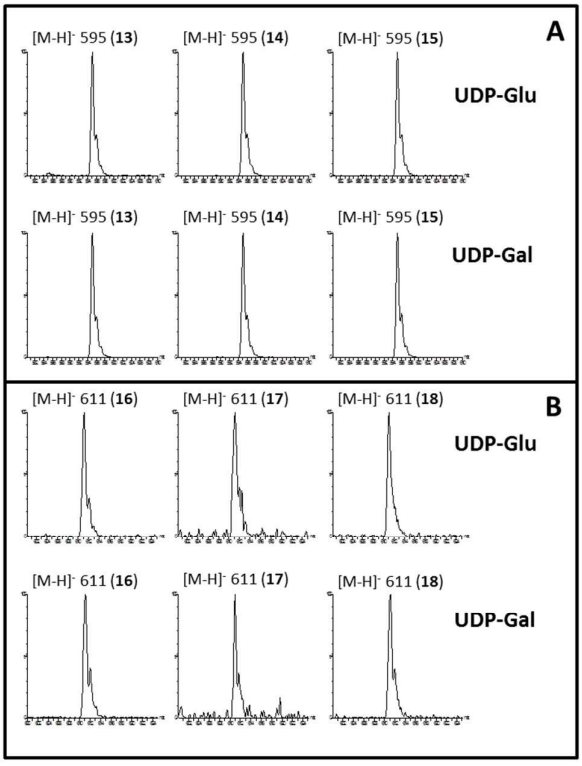


Fig. 7 Incorporation of fluorinated *C*-glucosyl-2-hydroxyflavanone analogue **11a** (**A**) or **12a** (**B**) and UDP-glucose or UDP-galactose into novel di-*C*-glycosylflavonoid products by the soluble root protein fraction of *D. incanum.* Molecular ions of di-*C*-glycosylated, fluorinated substrates in the negative ESI mass spectra are at $[M-H]^-$ 595 (**13**, **14**, **15**) and $[M-H]^-$ 611 (**16**, **17**, **18**).

Likewise, incubation of **12a** with either UDP-galactose or UDP-glucose and the leaf proteins of *D. incanum* generated three new metabolites (**16-18**) with molecular ions of $[M-H]^-$ 611. These were also shown to contain two *C*-linked hexoses (*Supplementary figure 5B*) and determined to be analogues of **1-3** (Figure 8). As novel metabolites, the analysis is not complicated by ions derived from the natural compounds and LCMS analysis gives a clear molecular ion envelope. As before, substrates **11a** and **12a** were not separated from trace amounts of UDP-glucose used in their preparation and so the incubations were repeated using UDP-α-D-[UL-$^{13}C_6$]glucose or UDP-α-D-[UL-$^{13}C_6$]galactose to demonstrate the source of the second glycosyl unit. When **11a** was incubated with the *D. incanum* leaf protein fraction and UDP-α-D-[UL-$^{13}C_6$]glucose or UDP-α-D-[UL-$^{13}C_6$]galactose, the same three new metabolites were again found by LCMS analysis with the expected $[M-H+6]^-$ 601 molecular ion through incorporation of a UDP-[UL-$^{13}C_6$]hexose. Ions detected at $[M-H]^-$ 595 are derived from contamination of **11a** with the UDP-glucose used in its preparation (*Supplementary figure 6A*). In the case of substrate **12a**, assays with root protein also produced the same three metabolites with molecular ion $[M-H+6]^-$ 617 as expected with ions at $[M-H]^-$ 611 derived from contaminant UDP-glucose (*Supplementary figure 6B*).

Conclusions

The first glycosylation results are consistent with those previously described for *D. intortum* and *D. uncinatum*. In fact, through our work characterising the potential for other species to

generate these metabolites, we have identified the same biochemistry in *D. intortum*, *D. uncinatum*, *D. incanum*, *D. ramossisimum* (Fabacaea), *Lens culinaris* (lentil, Fabaceae), *Cajanus cajan* (pigeon pea, Fabaceae), *Fagopyrum esculentum* (buckwheat, Polygonaceae),[24] *Cerastium arvense* (field mouse ear, Caryophyllaceae), *Silene alba* (white campion, Caryophyllaceae) and it has been reported by others in *Oryza sativa* (rice, Poaceae)[23] and *Spirodella* (duckweed, Araceae).[25]

The second glycosylation requires the *C*-glucosyl-2-hydroxyflavanone substrate **10** and cofactors UDP-glucose, UDP-galactose or UDP-arabinose to complete the transformation. It is not clear from these assays whether a single enzyme is capable of introducing all of the UDP-sugars onto intermediate **10**, or whether a different enzyme is required for each specific glycosylation. The substrate specificity of the second glycosylation is being probed by fractionation of the enzyme mixture to determine whether different activities can be separated or are co-fractionated. In the assays described here, dehydration of the glycosylated 2-hydroxyflavanone intermediates was carried out chemically by the addition of dilute HCl but there is the possibility of involvement of an enzymic dehydratase, which could alter the ratio of regioisomers of the final metabolites favouring specific glycoside positions at C6 or C8. We have also demonstrated that there is some relaxed substrate specificity for the *C*-glucosytransferases that catalyse both the first and second sugar transfers. In the first step, fluorinated analogues of vitexin and isovitexin were generated that could be identified and shown to be identical to fully characterised metabolites from OsCGT incubation with the substrate. The second *C*-hexosylation using UDP-glucose or UDP-galactose also demonstrated some relaxed substrate specificity and three di-*C*-

glycosylflavones were characterised as the analogous fluorinated products by mass spectrometry only.

In our experiments, the only *C*-glycosylation of 2-hydroxynaringenin observed was with UDP-glucose to yield a *C*-glucosyl intermediate. However, two of the compounds observed in the root tissues and exudates of *D. incanum* are **6** and **7**, which contain *C*-arabinosyl and *C*-galactosyl groups but no *C*-glucose. It is still unknown how these metabolites are biosynthesised and the chemical synthesis of intermediates to probe possible pathways is in progress.

Experimental

General

Electrospray ionisation mass spectra (ESIMS) were recorded in positive ionisation mode on a VG Autospec spectrometer and in negative ionisation mode using the Micromass Quattro Ultima. LCMS analysis was performed on the Micromass Quattro Ultima bench top triple quadrupole mass spectrometer attached to Waters Acquity UPLC system (Ultra Performance Liquid Chromatography). NMR spectra were recorded using a Bruker Avance 500 MHz spectrometer. UDP-glucose was purchased from Sigma-Aldrich Company (UK), UDP-arabinose was purchased from Carbosource Services supported in part by NSF-RCN grant 0090281 and UDP-α-D-[UL-$^{13}C_6$]glucose/UDP-α-D-[UL-$^{13}C_6$]galactose were purchased from Omicron Biochemicals, Inc.

Plant material.

Seeds of *D. incanum* were obtained from the seed bank at the International Livestock Research Institute (Addis Ababa, Ethiopia) and grown in seed multiplication plots at the field station of the International Centre for Insect Physiology and Ecology, Thomas Odhiambo Campus (*icipe*-TOC) on the shores of Lake Victoria in western Kenya. *Desmodium* seeds were germinated in vermiculite and grown until approximately 4 cm high so that a significant root system had established. Subsequent removal of the plants was facilitated by this medium. The plants were washed and placed in a hydroponic solution (40% Long Ashton) in plastic tubs containing about 2 litres for approximately 50 plants. The plants were grown until a suitable size for harvesting root or leaf tissue for protein extraction.

Protein extraction

Leaves or roots of *D.incanum* were snap frozen in liquid nitrogen and blended in ice cold 5 v/w 100 mM HEPES pH 7.2, 2 mM DTT, 2.5% PVP, 0.5 mM EDTA, 1 mM benzamidine, 0.1 mM PMSF. The mixture was centrifuged (9,000 x *g*, 30 min, 0˚C) to remove plant debris. Ammonium sulfate was added to the crude extract to achieve 40% concentration. The solution was stirred for 45 min and then centrifuged (9,000 x *g*, 30 min, 0˚C). The supernatant was decanted and its ammonium sulphate concentration increased to 80% and the mixture incubated at 4˚C for 1 hour with occasional mixing. The protein was then pelleted by centrifugation (9,000 x g, 30 min, 0˚C) and the pellet kept at - 80˚C until use. The protein pellets were desalted in 20 mM HEPES-NaOH pH 8 containing 2 mM DTT using Sephadex G-25 gel filtration chromatography (PD-10 columns, GE Healthcare) and concentrated to achieve a concentration of approximately 100-150 μL per gram wet weight of desmodium tissue. Protein content was estimated spectrophotometrically (NanoDrop).

***D. incanum* enzyme assays**

Approximately 1 g wet plant tissue weight equivalent of protein extract was added to 100 μL buffer (20 mM HEPES-NaOH pH 8 containing 2 mM DTT) along with 40 μL of a 20 mmol solution of substrate and 1 mg of UDP-sugar. The mixture was shaken briefly and then incubated at 30 °C overnight. Enzyme reactions were terminated by the addition of 250 μL 1 M HCl. The mixture was centrifuged to remove insoluble debris and the supernatant decanted and evaporated under N_2. The residue was redissolved in 500 μL 3:1 H_2O/MeOH for HPLC analysis.

Due to the cost and availability of some of the sugar donors, when using UDP-arabinose, UDP-α-D-[UL-$^{13}C_6$]glucose and UDP-α-D-[UL-$^{13}C_6$]galactose, the assays were conducted on one tenth scale. 0.1 mg of the sugar donor in 4 μL buffer (20 mM HEPES-NaOH pH 8 containing 2 mM DTT) was incubated with 4 μL of a 20 mM solution of substrate in ethanol, 10 μL of additional buffer and 0.1 g wet weight equivalent of 40-80% $(NH_4)_2SO_4$ *D. incanum* protein fraction. The assay was incubated at 30 °C overnight, quenched with 1 M HCl, centrifuged to remove insoluble material, the supernatant decanted and evaporated under a stream of nitrogen and the residue dissolved in 100 μL MeOH for HPLC and LCMS analysis. UDP-arabinose was purchased from Carbosource Services supported in part by NSF-RCN grant 0090281. D-[UL-$^{13}C_6$]galactose and D-[UL-$^{13}C_6$]glucose were purchased from Omicron Biochemicals Inc., Indiana.

Generation of *C*-glucoside-2-hydroxyflavanone intermediates using recombinant OsCGT

2-Hydroxyflavanone (40 μL of a 20 mmol solution) in ethanol was added to 8 mM UDP-glucose and 50 μg recombinant OsCGT[16] dissolved in 200 μl 100 mM HEPES pH 7.5 buffer

and incubated at 30°C overnight. The reaction was terminated by addition of ice cold MeOH to precipitate protein which were removed by centrifugation at 4 °C. 1.5 assay equivalents of *C*-glucosyl-2-hydroxyflavanones were then used directly in each single subsequent *Desmodium* enzyme assays. In the case of substrates **11** and **12**, some assays were quenched with dilute HCl and the products isolated by semi-preparative HPLC, using the same method used for analysis, to yield 6- and 8-*C*-glucosylated products that could be characterized and used in *Desmodium* assay experiments for identification purposes.

4'-fluoro-8-*C*-glucosylchrysin (11b). UV (λmax MeOH/H_2O) 269, 316 nm. ^{1}H NMR (500 MHz, d_4-MeOD, 325 K) δ 8.17 (2H, br t, *J* = 6.6 Hz, H2', H6'), 7.32 (2H, t, *J* = 8.7 Hz, H3', H5'), 6.72 (1H, s, H3), 6.32 (1H, s, H6), 6-*C*-β-Glu 5.06 (1H, d, *J* = 10.0 Hz, H1''), 4.10 (1H, t, *J* = 9.3 Hz, H2"), 3.96 (1H, dd, *J* = 2.0, 12.1, H6a"), 3.85 (1H, dd, *J* = 5.1, 12.0 Hz, H6b''), 3.67 (1H, br t, 9.4 Hz, H4''), 3.57 (1H, t, J = 8.8 Hz, H3''), 3.52 (1H, m, H5''). ^{13}C NMR (125 MHz, d_4-MeOD, 325 K) 182.6 (C4), 165.0 (d, *J* = 250 Hz, C4'), 164.0 (C5), 163.5 (C7), 163.5 (C2), 156.7 (C9), 129.2 (d, *J* = 8.8 Hz, C2', C6'), 128.0 (C1'), 115.7 (d, *J* = 22.5 Hz, C3', C5'), 104.4 (C10), 104.4 (C3), 104.2 (C8), 98.9 (C6), 81.4 (C5''), 78.9 (C3''), 74.5 (C1''), 71.6 (C2''), 70.2 (C4''), 61.4 (C6''). ^{19}F NMR (470 MHz, d_4-MeOD, 325 K) -109.9. ESIMS m/z (CV = XX, rel. int.): 433 $[M-H]^-$ (15), 343 $[M-H-90]^-$ (7), 313 $[M-H-120]^-$ (100). HRMS (negESI): calcd for $C_{21}H_{18}FO_{10}$ [M-H] 433.09349; found: 433.09299.

4'-fluoro-6-*C*-glucosylchrysin (11c). UV (λmax MeOH/H_2O) 271, 317 nm. ^{1}H NMR (500 MHz, d_4-MeOD, 325 K) δ 8.01 (2H, dd, *J* = 5.2, 8.9 Hz, H2', H6'), 7.27 (2H, t, *J* = 8.7 Hz, H3', H5'), 6.69 (1H, s, H3), 6.52 (1H, s, H6), 6-*C*-β-Glu 4.92 (1H, d, *J* = 9.9 Hz, H1''), 4.11 (1H, t, *J* = 8.8 Hz, H2"), 3.86 (1H, dd, *J* = 2.2, 12.0, H6a"), 3.75 (1H, dd, *J* = 5.1, 12.2 Hz, H6b''), 3.52-3.46 (2H, m, H3'', H4''), 3.44 (1H, m, H5''). ^{13}C NMR (125 MHz, d_4-MeOD, 325 K) 182.8 (C4), 165.3

(d, *J* = 247.5 Hz, C4'), 163.5 (C7), 163.5 (C2), 160.9 (C5), 157.8 (C9), 128.5 (d, *J* = 10.0 Hz, C2', C6'), 127.8 (C1'), 115.8 (d, *J* = 22.5 Hz, C3', C5'), 108.3 (C6), 104.8 (C10), 104.6 (C3), 94.1 (C8), 81.1 (C5''), 78.8 (C3''), 74.0 (C1''), 71.4 (C2''), 70.4 (C4''), 61.4 (C6''). ^{19}F NMR (470 MHz, d_4-MeOD, 325 K) -110.5 (dt, *J* = 8.3 5.7 Hz). ESIMS m/z (CV = 150, rel. int.): 433 $[M-H]^-$ (15), 343 $[M-H-90]^-$ (20), 313 $[M-H-120]^-$ (100). HRMS (negESI): calcd for $C_{21}H_{18}FO_{10}$ [M-H] 433.09349; found: 433.09282.

3'-fluorovitexin (12b). UV (λmax MeOH/H_2O) 269, 327 nm. ^{1}H NMR (500 MHz, d_4-MeOD, 325 K) δ 7.77 (1H, d, *J* = 13.0 Hz, H2'), 7.70 (1H, d, *J* = 8.7 Hz, H6'), 7.01 (1H, t, *J* = 8.7 Hz, H5'), 6.56 (1H, s, H3), 6.26 (1H, s, H6), 6-*C*-β-Glu 5.05 (1H, d, *J* = 9.9 Hz, H1''), 4.10 (1H, t, *J* = 9.4 Hz, H2"), 3.96 (1H, br d, *J* = 10.8, H6a"), 3.82 (1H, dd, *J* = 5.8, 12.1 Hz, H6b''), 3.65 (1H, t, *J* = 9.3 Hz, H4''), 3.56 (1H, t, *J* = 8.8 Hz, H3''), 3.54 (1H, m, H5''). ^{13}C NMR (125 MHz, d_4-MeOD, 325 K) 181.7 (C4), 163.8 (C2), 163.6 (C7), 161.2 (C5), 156.4 (C9), 151.4 (d, *J* = 151.4 Hz, C3'), 151.3 (C4''), 123.5 (C6'), 120.4 (C1'), 118.7 (C5'), 114.0 (C2'), 103.9 (C8), 103.5 (C10), 102.4 (C3), 99.3 (C6), 81.4 (C5''), 79.0 (C3''), 74.5 (C1''), 71.7 (C2''), 71.0 (C4''), 61.9 (C6''). ESIMS m/z (CV = 150, rel. int.): 449 $[M-H]^-$ (12), 359 $[M-H-90]^-$ (10), 329 $[M-H-120]^-$ (100). HRMS (negESI): calcd for $C_{21}H_{18}FO_{11}$ [M-H] 449.08840; found: 449.08748.

3'-fluoroisovitexin (12c). UV (λmax MeOH/H_2O) 271, 332 nm. ^{1}H NMR (500 MHz, d_4-MeOD, 310 K) δ 7.69 (1H, dd, *J* = 1.7, 12.2 Hz, H2'), 7.65 (1H, br d, *J* = 8.7 Hz, H6'), 7.04 (1H, t, *J* = 8.6 Hz, H5'), 6.61 (1H, s, H3), 6.52 (1H, s, H8), 6-*C*-β-Glu 4.93 (1H, d, *J* = 9.9 Hz, H1''), 4.18 (1H, t, *J* = 9.1 Hz, H2"), 3.90 (1H, dd, *J* = 2.0, 12.0, H6a"), 3.78 (1H, dd, *J* = 5.1, 12.1 Hz, H6b''), 3.54-3.49 (2H, m, H3'', H4''), 3.45 (1H, m, H5''). ^{13}C NMR (125 MHz, d_4-MeOD, 310 K) 182.0 (C4), 164.1 (C7), 163.2 (C2), 160.4 (C5), 156.8 (C9), 152.3 (d, *J* = 230.0 Hz, C3'), 151.3 (C4''), 123.1 (C6'), 121.4 (C1'), 118.2 (C5'), 113.6 (C2'), 107.8 (C6), 103.6 (C10), 102.9 (C3), 94.0 (C6), 81.2 (C5''), 78.8 (C3''), 74.0 (C1''), 71.3 (C2''), 70.4 (C4''), 61.5 (C6''). ESIMS m/z (CV = 150, rel.

int.): 449 $[M-H]^-$ (17), 359 $[M-H-90]^-$ (57), 329 $[M-H-120]^-$ (100). HRMS (negESI): calcd for $C_{21}H_{18}FO_{11}$ [M-H] 449.08840; found: 449.08760.

HPLC analysis of enzyme assays with *D. incanum* soluble protein

The residue from a single enzyme assay was dissolved in 500 μL 3:1 H_2O/MeOH for HPLC analysis on a Shimadzu VP series HPLC system using an ACEQ C-18 column (250 mm x 4.6 mm, 5 μm). The mobile phase used a gradient program at 1 mL/min, initially 95:5 (A:B), to 85:15 at 3 min, 75:25 at 13 min, 70:30 at 25 min, 45:55 at 35 min, 45:55 at 45 min, 5:95 at 46 min, 5:95 at 58 min, 95:5 at 60 min.[26]

LCMS analysis of products from *D. incanum* soluble protein assays.

The enzyme assay products (section *4.5*) were subject to LCMS either as they were or diluted tenfold in MeOH. The mass spectrometer was operated in negative ion mode, with a capillary voltage of 2.7KV, cone voltage 50-180 eV, mass range 50 – 1000 m/z. Source temperature 130 °C, desolvation temperature 350°C, desolvation gas flow 1000L/hr (nitrogen) and cone gas flow 60L/hr (nitrogen). Where required, selected ions were admitted to the collision cell for MSMS analysis with argon admitted at a pressure of 2.1 e^{-3} mbar, causing CID. Samples were injected *via* the Acquity sample manager, injecting 1 μl onto an Acquity UPLC BEC HSS C18 1.8 μm 2.1 x 150 mm column. Run time was 63 minutes at a flow rate of 0.4 mL/min. Solvents used are defined A (water, 0.05% formic acid) and B (methanol). The mobile phase used a gradient program, initially 95:5 (A:B), to 85:15 at 3 min, 75:25 at 13 min, 70:30 at 25 min, 45:55 at 35 min, 45:55 at 40 min, 5:95 at 46 min, 5:95 at 58 min, 95:5 at 60 min, 95:5 at 63 min.[26] Standards for comparison were obtained previously from *Desmodium* plant tissue and fully characterised by NMR spectroscopy.[11]

Acknowledgements

Rothamsted Research receives Grant-aided support from the Biotechnology and Biological Sciences Research Council (BBSRC). This project was funded through the Bill and Melinda Gates Foundation funded project 'Integrated *Striga* Management for Africa', and the European Union funded project 'Adapting Push-Pull Technology for Climate Change'.

Bing Hao was supported through a China Scholarship Council Award.

Notes and References

1. Z. R. Khan, in *Encyclopedia of Pest Management*, ed. D. Pimentel, Markel Dekker, Inc., USA, 2002, pp. 155–158.
2. M. Fischler, *Impact assessment of push–pull technology developed and promoted by* icipe *and partners in eastern Africa,* Intercooperation, Switzwerland*.* 2010, http://www.push-pull.net/Impact_Assessment.pdf
3. C. A. O. Midega, T. J. A. Bruce, J. A. Pickett, J. O. Pittchar, A. Murage and Z. R. Khan, *Field Crops Res*., 2015, **180**, 118-125.
4. Z. R. Khan, C. A. O. Midega, J. O. Pittchar, A. W. Murage, M. A. Birkett, T. J. A. Bruce, J. A. Pickett, *Phil. Trans. Royal Soc.*-B, 2014, **1639**, 369.
5. Z. R. Khan, D. M. Amudavi, C. A. O. Midega, J. M. Wanyama and J. A. Pickett, *Crop protection*, 2008, **27**, 976-987.
6. Z. R. Khan, A. Hassanali, W. Overholt, T. M. Khamis, A. M. Hooper, J. A. Pickett, L. J. Wadhams and C. M. Woodcock, *J. Chem. Ecol*., 2002, **28**, 1871-1885.
7. M. K. Tsanuo, A. Hassanali, A. M. Hooper, Z. R. Khan, F. Kaberia, J. A. Pickett and L. J. Wadhams, *Phytochemistry*, 2003, **64**, 265-273.

8. A. M. Hooper, M. K. Tsanuo, K. Chamberlain, K. Tittcomb, J. Scholes, A. Hassanali, Z.R. Khan and J.A. Pickett, *Phytochemistry*, 2010, **71**, 904-908.
9. Z. R. Khan, C. A. O. Midega, T. J. A. Bruce, A. M Hooper and J. A. Pickett, *J. Exp. Bot*., 2010, **61**, 4185-4196.
10. J. A. Pickett, M. L. Hamilton, A. M. Hooper, Z. R. Khan and C. A. O. Midega, *Ann. Rev. Phytopath*., 2010, **48**, 161-177.
11. A. M. Hooper, J. C. Caulfield, B. Hao, J. A. Pickett, C. A. O. Midega and Z. R. Khan, *Phytochemistry*, 2015, **117**, 380-389.
12. Z. R. Khan, C. A. O. Midega, E. M. Njuguna, D. M. Amudavi, J. M. Wanyama and J. A. Pickett, *Crop Protection*, 2008, **27**, 1084-1097.
13. C. A. O. Midega, J. Pittchar, D. Salifu, J. A. Pickett and Z. R. Khan, *Crop Protection*, 2013, **44**, 44-49.
14. A. M. Hooper, A. Hassanali, K. Chamberlain, Z. R. Khan and J.A. Pickett, *Pest Manag. Sci*., 2009, **65**, 546-552.
15. M. L. Hamilton, J. C. Caulfield, J. A. Pickett and A. M. Hooper, *Tetrahedron Lett*., 2009, **50**, 5656-5659.
16. M. L. Hamilton, S. P. Kuate, M. Brazier-Hicks, J. C. Caulfield, R. Rose, R. Edwards, B. Torto, J. A. Pickett and A. M. Hooper, *Phytochemistry*, 2012, **84**, 169-176.
17. P. F. Byrne, M. D. McMullen, M. E. Snook, T. A. Musket, J. M. Theuri, N. W. Widstrom, B. R. Wiseman and E. H. Coe, *Proc. Natl. Acad. Sci*., 1996, **93**, 8820-8825.
18. K. Chul-Sa, M. A. Kabir, M. Sachi, T. Shin-ichi and K. Hen-Sik, *Z. naturforsch*, 2009, **63c**, 755-760.
19. D. J. McNally, K. V. Wurms, C. Labbe and R. R. Belanger, *Physiol. Mol. Plant Pathol*., 2003, **63**, 293-303.

20. K. Akiyama, H. Matsuoka and H. Hayashi, *Mol. Plant Microbe Int*., 2002, **15**, 334-340.

21. O. Talhi and A. M. S. Silva, *Curr. Org. Chem*., 2012, **16**, 859-896.

22. F. Cuyckens and M. Claeys. *J. Mass Spectrom*., 2004, **39**, 1-15.

23. M. Brazier-Hicks, K. M. Evans, M. C. Gershater, H. Puschmann, P. G. Steel and R. Edwards*, J. Biol. Chem*., 2009, **284**, 17926-17934.

24. F. Kerscher and G. Franz, *J. Plant Physiol*., 1988, **132**, 110-115.

25. J. W. Wallace and H. Griesbach, *Biochim Biophys Acta*, 1973, **304**, 837-841.

26. F. Ferreres, B. M. Silva, P. B. Andrade, R. M. Seabra, M. A. Ferreira, *Phytochem. Anal*., 2003, **14**, 352-359.

Figures and Legends

Fig. 1 Incorporation of [2',3',5',6'-$^{2}H_4$]2-hydroxynaringenin and UDP-glucose into vitexin (**8**) and isovitexin (**9**) by the soluble leaf (**A**) and root (**B**) protein fraction of *D. incanum*. Ions labelled with four deuterium atoms in the negative ESI mass spectra are at [M-H+4$^-$] 435.

Fig. 2 Incorporation of [2',3',5',6'-$^{2}H_4$]6-*C*-glucosyl-2-hydroxynaringenin and UDP-arabinose into isoschaftoside (**4**) and schaftoside (**5**) by the soluble leaf (**A**) and root (**B**) protein fraction of *D. incanum*. Ions labelled with four deuterium atoms in the negative ESI mass spectra are at [M-H+4$^-$] 567.

Fig. 3 Incorporation of [2',3',5',6'-$^{2}H_4$]6-*C*-glucosyl-2-hydroxynaringenin and UDP-glucose into 6-*C*-galactosyl-8-*C*-glucosylapigenin **1**, 6,8-di-*C*-glucosylapigenin (vicenin-2) **2** and 6-*C*-glucosyl-8-*C*-galactosylapigenin **3** by the soluble leaf (**A**) and root (**B**) protein fraction of *D. incanum*. Ions labelled with four deuterium atoms in the negative ESI mass spectra are at [M-H+4$^-$] 597.

Fig. 4 Incorporation of [2',3',5',6'-$^{2}H_4$]6-*C*-glucosyl-2-hydroxynaringenin and UDP-galactose into 6-*C*-galactosyl-8-*C*-glucosylapigenin **1**, 6,8-di-*C*-glucosylapigenin (vicenin-2) **2** and 6-*C*-glucosyl-8-*C*-galactosylapigenin **3** by the soluble leaf (**A**) and root (**B**) protein fraction of *D. incanum*. Ions labelled with four deuterium atoms in the negative ESI mass spectra are at [M-H+4$^-$] 597.

Fig. 5 Incorporation of [2',3',5',6'-$^{2}H_4$]6-*C*-glucosyl-2-hydroxynaringenin and UDP-α-D-[UL-$^{13}C_6$]glucose (**left**) or UDP-α-D-[UL-$^{13}C_6$]galactose (**right**) into 6-*C*-galactosyl-8-*C*-glucosylapigenin **1**, 6,8-di-*C*-glucosylapigenin (vicenin-2) **2** and 6-*C*-glucosyl-8-*C*-galactosylapigenin **3** by the soluble leaf (**A**) and root (**B**) protein fraction of *D. incanum.* Ions labelled with four deuterium atoms and six ^{13}C atoms in the negative ESI mass spectra are at [M-H+10]$^-$ 603 and unlabelled metabolites at [M-H]$^-$ 593.

Fig. 6 Incorporation of **11** or **12** and UDP-glucose into C-glycosylflavones by the soluble leaf (**A**) and root (**B**) protein fraction of *D. incanum.* Molecular ions of *C*-glucosylated, fluorinated substrates in the negative ESI mass spectra are at $[M-H]^-$ 433 (**11b**, **11c**) and 449 (**12b**, **12c**).

Fig. 7 Incorporation of fluorinated *C*-glucosyl-2-hydroxyflavanone analogue **11a** (**A**) or **12a** (**B**) and UDP-glucose or UDP-galactose into novel di-*C*-glycosylflavonoid products by the soluble root protein fraction of *D. incanum.* Molecular ions of di-*C*-glycosylated, fluorinated substrates in the negative ESI mass spectra are at $[M-H]^-$ 595 (**13**, **14**, **15**) and $[M-H]^-$ 611 (**16**, **17**, **18**).