Development of an oligosaccharide library to characterize the structural variation in glucuronoarabinoxylan in the cell walls of vegetative tissues in grasses.

Theodora Tryfona[a], Mathias Sorieul[a, 1], Carolina Feijao[a, 2], Katherine Stott[b], Denis V. Rubtsov[a,3], Nadine Anders [a], Paul Dupree[a, *]

*Correspondence: pd101@cam.ac.uk

[a] Department of Biochemistry, University of Cambridge, Hopkins Building, The Downing site, Tennis Court Road, Cambridge CB2 1QW, UK

Full list of author information is available at the end of the article.

Abstract

Background: Grass glucuronoarabinoxylan (GAX) substitutions can inhibit enzymatic degradation and are involved in the interaction of xylan with cell wall cellulose and lignin, factors which contribute to the recalcitrance of biomass to saccharification. Therefore, identification of xylan characteristics central to biomass biorefining improvement is essential. However, the task of assessing biomass quality is complicated and is often hindered by the lack of a reference for a given crop.

Results: In this study we created a reference library, expressed in glucose units, of *Miscanthus sinensis* GAX stem and leaf oligosaccharides, using DNA sequencer-Assisted Saccharide analysis in high throughput (DASH), supported by liquid chromatography (LC), nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS). Our analysis of a number of grass species highlighted variations in substitution type and frequency of stem and leaf GAX. In miscanthus, for example, the β-Xyl*p*-(1→2)-α-Ara*f*-(1→3) side chain is more abundant in leaf than stem.

Conclusions: The reference library allows fast identification and comparison of GAX structures from different plants and tissues. Ultimately, this reference library can be used in directing biomass selection and improving biorefining.

Keywords: Grass xylan; bioenergy; tissue-variation; species-variation; xylan branching; DASH.

Background

Plant xylan polysaccharides have attracted attention due to their numerous applications not only in the papermaking, baking and food industries but also in respect to bioenergy production. Branched xylan is the main hemicellulose in many crops. This xylan is made of a linear chain of β-(1→4)-linked xylopyranosyl (Xyl*p*) residues, which can be substituted by α-(1→2)-linked (4-*O*-methyl-)glucuronic acid ([Me]GlcA) and acetylation at the *O*-2 and/or *O*-3 positions [1]. In monocots, such as grasses, and also in gymnosperms, xylan is additionally modified by α-(1→3)-linked L-arabinofuranosyl residues (Ara*f*). The Ara*f* residues may be further substituted at *O*-2 with an Ara*f* or a Xyl*p* residue [2]. Cereal grain endosperm contains neutral arabinoxylan (AX), which is monosubstituted with α-(1→3)-linked Ara*f* residues or di-substituted with α-(1→2)-linked and α-(1→3)-linked Ara*f* residues [1, 3]. Ara*f* residues of both glucuronoarabinoxylan (GAX) and endosperm AX may be esterified with a feruloyl (Fer)- or coumaryl group at *O*-5 position [4, 5]. Feruloylation has been implicated in cross-linking of different xylan chains and in cross-linking to lignin [6, 7]. Feruloylated Ara*f* structures can be further substituted with β-(1→2)-linked Xyl*p* groups or additional sugars [8, 9]. Corn bran xylan was found to contain α-(1→2)-linked L-Gal*p* on the Xyl*p* residue of the Fer-Ara*f*-Xyl*p* oligomeric structure [10-12], and this was recently found to be present in other cereal grains as well [13, 14]. Among the ferulate-containing xylan side chain variants, 5-O-Fer-Ara*f* structure appears most abundant, followed by the Xyl*p*-[5-O-Fer]-Ara*f* structure [8]. Grass GAX and AX is acetylated but to a lesser extent than dicot glucuronoxylan. However, in addition to acetyl groups being added to the backbone Xyl*p* residues, the Ara*f* substituents can carry acetyl groups at *O*-2 [15].

Most of the energy in the lignocellulosic biomass is locked within the secondary cell walls in the form of cellulose and xylan, which form a dense matrix with lignin [16]. Lignocellulosic plant cell wall recalcitrance is a barrier to cost-effective cellulosic biofuel production. Cell wall

recalcitrance in regards to xylan is influenced by various factors: heavy substitution of xylan, which can impair the action of hydrolytic enzymes; specific branching points, which can serve as cross-linking sites with lignin; and the pattern of branching, which can affect the interaction of xylan with cellulose [17]. The importance of xylan in recalcitrance is illustrated by the finding that removal of xylan in switchgrass resulted in materials that achieved nearly 100% glucose yields in subsequent enzymatic hydrolysis [18].

In order to improve the biomass of bioenergy crops and optimise the biorefining processes, information on the structure of xylan, and its variation, is crucial. However, little is known about the variability of xylan structure of grasses in e.g. different tissues like stems and leaves. In addition, analysis methods of xylooligosaccharides have to be fast, accurate and robust. The majority of detailed structural information on xylan in grasses is based on analysis of cereal grain xylans such as corn cob, oat spelt, barley husks or wheat endosperm often using LC, NMR and MS [19-23]. More recently, structures of xylan of lignified tissues were analysed in grasses [24-26]. Oligosaccharides with different degree of polymerisation, glycosidic linkages and saccharide composition can be resolved with DNA sequencer-Assisted Saccharide analysis High throughput (DASH) which can analyse simultaneously 96 samples by capillary electrophoresis [27]. Our study provides the detailed structural characterisation of xylan oligosaccharides from *Miscanthus sinensis* stem cell walls hydrolysed with xylan specific glycosyl hydrolases (GH) from the families 10 and 11. This information was used to generate a DASH reference library of GAX oligosaccharides with their corresponding glucose units (GU) as mobility standards. This library allows the fast and quantitative comparison of GAX oligosaccharide structures using the high throughput method DASH, as shown by the comparison of oligosaccharide profiles of leaf and grass GAX from different grasses. Relative quantification of side chains can be achieved as exemplified by the more detailed analysis of miscanthus stem and leaf GAX.

Results

Development of a GAX oligosaccharide library from *Miscanthus sinensis* stem

In order to use DASH as a high throughput method to characterise GAX structures and achieve relative quantification of different xylan substitutions, a GAX oligosaccharide library had to be created to serve as a standard for structural analysis. To develop this library, *Miscanthus sinensis* stem GAX was hydrolysed with endo-β-xylanases GH10 and GH11, respectively. Most reported xylanases classify into these two families, which are described to have slightly different substrate specificities. Briefly, GH10 xylanases are more capable of hydrolysing adjacent to substitutions of the xylan backbone, while GH11 xylanases preferably act on relatively unsubstituted parts of xylan [28, 29].

To describe the various hydrolysis products we use the heteroxylan naming system suggested by Faure *et al.* (2009) with the exception of xylooligosaccharide standards which are described as X_1-X_6 corresponding to xylose, xylobiose, xylotriose, xylotetraose, xylopentaose and xylohexaose.

The DASH capillary electropherograms were first aligned using the internal mobility migration markers mixed with each sample to eliminate variation between capillaries [27]. The GAX oligosaccharide library was compiled by assigning the characterised oligosaccharide structures to DASH peaks based on their migration in DASH. The migration of the oligosaccharides was compared to the migration of dextran standards to provide migration information in a method we adapted here from liquid chromatography [30]. By providing the migration of the xylooligosaccharides in GU, the identification is more robust to any changes in capillary sequencer variation. Comparison of oligosaccharide GU migration therefore allows the fast and reliable annotation of GAX structures in unknown samples to all DASH users.

To characterise the detailed structure of GAX oligosaccharides, we separated the hydrolysis products by SEC and all fractions were analysed by DASH. Based on the result of the DASH analysis, SEC fractions were selected in which the respective oligosaccharides were highly abundant. Part of these fractions was subjected to secondary enzymatic hydrolysis followed by DASH analysis. The other part was used to separate possible structural isomers by Hydrophilic Interactions Liquid Chromatography (HILIC) followed by off line Matrix Assisted laser Desorption Ionisation (MALDI)-Mass Spectrometry. The oligosaccharides were then subjected to high energy MALDI Collision-Induced-Dissociation (CID) for detailed structural analysis. Ultimately, DASH peaks were matched to oligosaccharide structures characterised with HILIC-MALDI-MS/MS CID by combining data on peak abundance and enzyme sensitivity from different SEC fractions.

Based on the previously characterised composition of grass GAX [1], in the MALDI CID spectra we assign uronic acid substitutions as GlcA and furanosyl pentosyl substitutions as Ara*f* on a 1,4-linked Xyl*p* backbone. Given that no GlcA residues on xylan have been reported to carry a methyl modification on the *O*-6 we assign methyl group modifications at *O*-4. In addition, all oligosaccharides characterised here are generated by GH10 or GH11 endo-β-1,4 xylanases and therefore the non-reducing end backbone Xyl*p* residues can not be modified at the *O*-4. Glycosidic bond and cross-ring product ions are labelled according to the nomenclature of Domon and Costello (1988). The D, E, G and V ions are labelled according to previously established nomenclature [31-33].We used sequential digests of the oligosaccharides with enzymes to gain further insights into the structure and to confirm the results of the MALDI-CID: arabinofuranosidases GH51 and GH62 hydrolyse single Ara*f* substitutions but not Xyl*p* [34], which helps to distinguish the nature of pentosyl side chains; glucuronosidases GH67 and GH115 both remove [Me]GlcA substitutions. However, GH67 can only remove terminal GlcA from the non-reducing end, whereas GH115 preferably acts on

139 substitutions of internal regions although it will also cleave GlcA from non-reducing terminal
140 Xyl*p* residues albeit at a slower rate [35, 36].
141 An overview of the analysis is depicted in Figure 1.
142
143 **GAX oligosaccharide profiles of xylanase GH10 and GH11 hydrolysis of *Miscanthus***
144 ***sinensis***
145 In brief, a standard DASH profile was generated by APTS labelling of GH10 and GH11
146 hydrolysis products of *Miscanthus sinensis* stem Alcohol Insoluble Residue (AIR),
147 respectively, and analysis by DASH (Figure 2).
148 In order to partially separate the oligosaccharides for further structural analysis, the hydrolysis
149 products were subjected to Size Exclusion Chromatography (SEC). Eighty *Miscanthus sinensis*
150 GH10 and GH11 oligosaccharide SEC fractions (*Ms*10_1-80 and *Ms*11_1-80) were collected
151 and labelled with APTS and analysed by DASH, revealing separation of a number of
152 oligosaccharides by SEC (Figure 3 and Figure S1) in comparison to the standard DASH profile
153 prior to SEC (Figure 2). Figure 3 and Figure S1 also show the power of DASH to study a large
154 number of oligosaccharide mixtures, to aid interpretation of the SEC separation and selection
155 of fractions of interest.
156 Some oligosaccharides separated by DASH eluted at the same time as β-(1→4)-
157 xylooligosaccharide (X_1-X_6) standards; other oligosaccharides (namely XU^2XX, $XU^{(4Me)2}XX$
158 and XA^3XX) were previously structurally characterised [27-29]. The identity of their DASH
159 peaks was confirmed by MALDI-CID (data not shown). Remaining unknown peaks were
160 labelled N_{1-12} (Figure 2). The detailed results of the structural analysis of each oligosaccharide
161 are described below. Table 1 summarises the GAX oligosaccharide structures generated by
162 GH10 and GH11 hydrolysis and their corresponding GU. Table 2 summarise the
163 sensitivity/resistance to enzymatic hydrolysis of GAX oligosaccharides.

164

165 **Table 1.** Structures of GAX oligosaccharides and their migration positions by DASH expressed

166 in glucose units (GU).

Nomenclature	Structure	Unknown structure	MALDI-CID or NMR Figure	Glucose Units
$U^{(4Me)2}X$	4-*O*-Me-α-Glc*p*A-(1→2)-β-Xyl*p*-(1→4)-Xyl*p*	N_1	S3	0.72
X	Xyl*p*			0.80
$U^{2}XX$	α-Glc*p*A-(1→2)-β-Xyl*p*-(1→4)-β-Xyl*p*-(1→4)-Xyl*p*	N_2		1.40
$U^{(4Me)2}XX$	4-*O*-Me-α-Glc*p*A-(1→2)-β-Xyl*p*-(1→4)-β-Xyl*p*-(1→4)-Xyl*p*	N_3	S4	1.48
X_2	β-Xyl*p*-(1→4)-Xyl*p*			1.57
$XU^{2}XX$	β-Xyl*p*-(1→4)-[α-Glc*p*A-(1→2)]-β-Xyl*p*-(1→4)-β-Xyl*p*-(1→4)-Xyl*p*			2.03
$XU^{(4Me)2}XX$	β-Xyl*p*-(1→4)-[4-*O*-Me-α-Glc*p*A-(1→2)]-β-Xyl*p*-(1→4)-β-Xyl*p*-(1→4)-Xyl*p*			2.07
X_3	β-Xyl*p*-(1→4)-β-Xyl*p*-(1→4)-Xyl*p*			2.24
$A^{3}X$	α-Ara*f*-(1→3)-β-Xyl*p*-(1→4)-Xyl*p*	N_4	S5	2.31
$A^{3}U^{(4Me)2}XX$	α-Ara*f*-(1→3)-β-Xyl*p*-(1→4)-[4-*O*-Me-α-Glc*p*A-(1→2)]-β-Xyl*p*-(1→4)-β-Xyl*p*-(1→4)-Xyl*p*	N_{10}	S8	2.74
$XA^{3}X$	β-Xyl*p*-(1→4)-[α-Ara*f*-(1→3)]-β-Xyl*p*-(1→4)-Xyl*p*	N_5	S6	2.87
$B^{2,3}U^{(4Me)2}XX$ or $D^{2,3}U^{(4Me)2}XX$	Ara*f*-(1→2)-α-Ara*f*-(1→3)-β-Xyl*p*-(1→4)-[α-Glc*p*A-(1→2)]-β-Xyl*p*-(1→4)- β-Xyl*p*-(1→4)-Xyl*p* β-Xyl*p*-(1→2)-α-Ara*f*-(1→3)-β-Xyl*p*-(1→4)-[α-Glc*p*A-(1→2)]-β-Xyl*p*-(1→4)- β-Xyl*p*-(1→4)-Xyl*p*	N_{11}	S9	3.09
$D^{2,3}X$	β-Xyl*p*-(1→2)-α-Ara*f*-(1→3)-β-Xyl*p*-(1→4)-Xyl*p*	M_1	S12	3.12
X_4	β-Xyl*p*-(1→4)-β-Xyl*p*-(1→4)-β-Xyl*p*-(1→4)-Xyl*p*			3.17
$A^{3}A^{3}X$	α-Ara*f*-(1→3)-β-Xyl*p*-(1→4)-[α-Ara*f*-(1→3)]-β-Xyl*p*-(1→4)-Xyl*p*	N_6 (Z_1)	5A	3.70
$A^{3}XXX$	α-Ara*f*-(1→3)-β-Xyl*p*-(1→4)-β-Xyl*p*-(1→4)-β-Xyl*p*-(1→4)-Xyl*p*	N_7 (Z_2)	5B	3.85
$XA^{3}XX$	β-Xyl*p*-(1→4)-[α-Ara*f*-(1→3)]-β-Xyl*p*-(1→4)-β-Xyl*p*-(1→4)-Xyl*p*			3.92
$D^{2,3}XX$	β-Xyl*p*-(1→2)-α-Ara*f*-(1→3)-β-Xyl*p*-(1→4)-β-Xyl*p*-(1→4)-Xyl*p*	N_8 (Z_3)	5C 5D	3.97

$XA^3XU^{(4Me)2}XX$	β-Xyl*p*-(1→4)-[α-Ara*f*-(1→3)]-β-Xyl*p*-(1→4)-β-Xyl*p*-(1→4)-[4-*O*-Me-α-Glc*p*A-(1→2)]-β-Xyl*p*-(1→4)-β-Xyl*p*-(1→4)-Xyl*p*	N_{12}	S10	4.01
X_5	β-Xyl*p*-(1→4)-β-Xyl*p*-(1→4)-β-Xyl*p*-(1→4)-β-Xyl*p*-(1→4)-Xyl*p*			4.32
XA^3A^3X	β-Xyl*p*-(1→4)-[α-Ara*f*-(1→3)]-β-Xyl*p*-(1→4)-[α-Ara*f*-(1→3)]-β-Xyl*p*-(1→4)-Xyl*p*	N_9	S7	4.47
$XD^{2,3}XX$	β-Xyl*p*-(1→4)-[β-Xyl*p*-(1→2)-α-Ara*f*-(1→3)]-β-Xyl*p*-(1→4)-β-Xyl*p*-(1→4)-Xyl*p*	M_2	S11	4.83
X_6	β-Xyl*p*-(1→4)-β-Xyl*p*-(1→4)-β-Xyl*p*-(1→4)-β-Xyl*p*-(1→4)-β-Xyl*p*-(1→4)-Xyl*p*			5.54

Table 2. Enzymatic analysis of GAX oligosaccharides.

Oligosaccharide	**Unknown structure**	**GH10 product**	**GH11 product**	**Enzyme sensitivity**				
				***Cg*GH3 xylosidase**	***Pc*GH51 α-arabinosidase**	***Pa*GH62 α-arabinosidase**	***Cj*GH67 α-glucuronidase**	***Bo*GH115 α-glucuronidase**
$U^{(4Me)2}X$	N_1	●●	●	-	-	-	-	+
U^2XX	N_2	●		-	-	-	+	+
$U^{(4Me)2}XX$	N_3	●●	●	-	-	-	+	+
A^3X	N_4	●●		-	+	+	**NT**	-
XU^2XX	-		●	-	-	-	+	+
$XU^{(4Me)2}XX$	-		●	-	-	-	+	+
XA^3X	N_5	●●	●●	**NT**	+	+	**NT**	-
A^3A^3X	N_6 (Z_1)	●		**NT**	+	-	**NT**	-
A^3XXX	N_7 (Z_2)	●		**NT**	+	+	**NT**	**NT**
XA^3XX	-		●●	-	**NT**	+	**NT**	-
$D^{2,3}XX$	N_8 (Z_3)	●		+	-	-	**NT**	-
XA^3A^3X	N_9	●		-	+	-	**NT**	**NT**
$A^3U^{(4Me)2}XX$	N_{10}		●	**NT**	+	-	**NT**	-
$B^{2,3}U^{(4Me)2}XX$ or $D^{2,3}U^{(4Me)2}XX$	N_{11}		●	-	-	**NT**	**NT**	-
$XA^3XU^{(4Me)2}XX$	N_{12}		●	**NT**	+	+	**NT**	+
$D^{2,3}X$	M_1	●		+	-	-	**NT**	-
$XD^{2,3}XX$	M_2		●	+	-	-	**NT**	-

The dots indicate, which xylanase produces the oligosaccharide; one dot (●) indicates minor products while two dots (●●) indicate major products . Oligosaccharides sensitive to GH enzymes are marked with a plus (+), those resistant with a minus (-); NT: not tested.

Oligosaccharides $XU^{2}XX$, $X^{(4Me)2}XX$ and $XA^{3}XX$ have been previously structurally characterised [27-29].

Structural characterisation of the oligosaccharides N_1-N_9 from the xylanase GH10 digest of *Miscanthus sinensis*

After GH10 hydrolysis three oligosaccharides co-eluted with X_1, X_2 and X_3 of the xylooligosaccharide standards in DASH. We identified nine additional peaks from oligosaccharides of unknown structure which were named N_1-N_9 (Figure 2). The [Me]GlcA modified hydrolysis products mainly accumulated in earlier SEC fractions independent of their molecular size (Figure 3, fractions *Ms*10_24-46) because of the negative charge of the Bio-Gel P2, which results in the exclusion of uronic acids [37].

The oligosaccharide N_1 in the DASH trace was analysed from the unfractionated GH10 miscanthus stem hydrolysis products (oligosaccharide mix prior to SEC fractionation). The $[M+Na]^+$ ion at *m/z* 616.0 corresponds to a $Pent_2$ structure modified with one MeGlcA. The MALDI-CID reveals the structure of N_1 as $U^{(4Me)2}X$ (Figure S3):

The Y_2 ion (*m/z* 426.0) shows there are two Xyl*p* residues in the backbone. The presence of the significant Y_1 ion (*m/z* 294.1) indicates that the reducing end Xyl*p* is unsubstituted. The $^{0,2}X_1$ ion (*m/z* 523.9) indicates that the non-reducing-end Xyl*p* is modified at the *O*-2 with a MeGlcA, while the presence of the V_3 product ion (*m/z* 539.9) indicates that the GlcA is modified with a methyl group at the *O*-4. Consistent with this, N_1 oligosaccharide was sensitive to GH115 glucuronidase digestion, although resistant to GH67 hydrolysis due to it not being an ideal substrate for the GH67 enzyme [35].

The oligosaccharide N_2 was in low abundance in the DASH trace and MALDI-CID showed it is the unmethylated counterpart of the N_3 oligosaccharide (data not shown).

198 The oligosaccharide N_3 in the DASH trace was analysed from fraction *Ms*10_40. The $[M+Na]^+$
199 ion at *m/z* 748.1 corresponds to a $Pent_3$ structure modified with one MeGlcA. The MALDI-
200 CID reveals the structure of N_3 as $U^{(4Me)2}XX$ (Figure S4): The series of Y and cross-ring $^{1,5}X$
201 ions gives crucial sequence information showing the positioning of the [Me]GlcA on the xylan
202 backbone. The cross-ring $^{0,2}X_2$ ion (*m/z* 655.9) indicates that the non-reducing-end Xyl*p* is
203 modified at the *O*-2 with a methylated uronic acid. The non-reducing end E_2 (*m/z* 461.0)
204 product ion confirms this linkage assignment and the presence of V_4 (*m/z* 671.9) product ion
205 indicates that the GlcA is modified with a methyl group at the *O*-4. The N_3 oligosaccharide
206 was susceptible to digestion with GH67 and GH115 glucuronidase (Table 2 and Figure S2A),
207 with the GH67-sensitivity confirming that the GlcA substitution is at the non-reducing end.
208 The oligosaccharide N_4 in the DASH trace was analysed from fraction *Ms*10_75. The $[M+Na]^+$
209 ion at *m/z* 558.1 corresponds to a $Pent_3$ structure, but it does not co-migrate with Xyl_3 in DASH,
210 suggesting it is a substituted Xyl_2. The MALDI-CID reveals the structure of N_4 as A^3X (Figure
211 S5): A major Y_1 ion (*m/z* 294.1) is indicative of an unsubstituted reducing-end Xyl*p*. The
212 presence of the $^{0,2}X_1$ ion (*m/z* 336.1) and the E_2 product ion (*m/z* 271.1) indicates that the second
213 Xyl*p* from the reducing-end is unsubstituted at the *O*-2. The presence of the G_3 ion (*m/z* 510.7)
214 indicates the pentose substitution is Ara*f*. Consistent with this, N_4 oligosaccharide was sensitive
215 to GH62 and GH51 arabinosidase digestions (Table 2).

216 The oligosaccharide N_5 in the DASH trace was analysed from fraction *Ms*10_65. The $[M+Na]^+$
217 ion at *m/z* 690.1 corresponds to a $Pent_4$ structure, but it does not co-migrate with Xyl_4 in DASH
218 suggesting it is a substituted Xyl_3. The MALDI-CID reveals the structure of N_5 as XA^3X
219 (Figure S6):
220 The $^{1,5}X$ ion shows that the 2-AA derivatised reducing-end Xyl*p* is not modified. The cross-
221 ring $^{0,2}X_1$ ion (*m/z* 336.1) indicates that the *O*-2 in the middle Xyl*p* is also not substituted. This
222 is supported by the presence of the E_2 (*m/z* 403.1) product ion, while the presence of the W_2

sugar lactone ion (*m/z* 424.1) [29] indicates that the middle Xyl*p* is substituted at *O*-3 with a pentose residue. The presence of the $G_{3\alpha}$ ion (*m/z* 642.0) indicates the substitution is Ara*f*. Consistent with this, the N_5 oligosaccharide was sensitive to GH62 and GH51 arabinosidase digestions (Table 2).

The oligosaccharides N_6-N_8 in the DASH trace were analysed from fraction *Ms*10_55 (Figure 4A, the structure N_9 was more abundant in fraction *Ms*10_45 and corresponds to a $Pent_6$ oligosaccharide, see below). N_6-N_8 oligosaccharides had the same *m/z* 822.0 $[M + Na]^+$ corresponding to $Pent_5$ structural isomers (Figure 4B). The Extracted Ion Chromatogram (EIC) from off-line HILIC-MALDI-CID Mass Spectrometry of 2-AA labelled SEC fractions showed that *Ms*10_55, contained three structural isomers of $Pent_5$ observed as $[M + Na]^+$ at *m/z* 822.0 (Z_1-Z_3, Figure 4C).

Arabinofuranosidase digestion of the oligosaccharides followed by DASH and HILIC-MALDI showed that the oligosaccharides Z_1, Z_2 and Z_3 correspond to N_6, N_7 and N_8 (Table 2).

The MS/MS spectra for N_6, N_7 and N_8 oligosaccharides are shown in Figure 5.

The presence of the G ions (*m/z* 773.9) in the N_6, N_7 and N_8 spectra indicate Ara*f* substitution. The presence of the $^{1,5}X_1$ cross-ring ion showed that the 2-AA labelled reducing-end Xyl*p* is not substituted in any of the three structural isomers. The significant Y_2, $^{1,5}X_2$ and $^{0,2}X_2$ ions (*m/z* 426.1, 454.0, and 468.1 respectively) in the spectra of N_7 and N_8, indicate that the second Xyl*p* residue from the reducing end is also not substituted. In contrast, the absence of these ions from the N_6 spectrum and the concomitant presence of the W_2 (*m/z* 424.1) sugar lactone ion indicates that in this isomer the second Xyl*p* residue from the reducing-end is substituted at *O*-3. The presence of the G_3 product ion at *m/z* 656.0 in the spectrum of N_6 indicates that the third Xyl*p* residue from the reducing end is also substituted at the *O*-3 with Ara*f*. This is confirmed by the presence of the E_2 (*m/z* 271.1) product ion which indicates that the third Xyl*p* from the reducing end is not substituted at *O*-2. Hence the N_6 structure was identified as A^3A^3X.

Consistent with this, the N_6 oligosaccharide was sensitive to GH51 arabinosidase digestion. Possibly due to steric hindrance of Ara*f* substitutions on two consecutive Xyl*p* residues, N_6 was resistant to GH62 hydrolysis (Table 2).

The presence of the series of B, Y and ^{1,5}X ions differing by 132 Da in the MALDI-CID spectra of the N_7 and N_8 structural isomers, indicate the linear sequence of pentoses. Based on the MALDI-CID spectra alone the precise structure of N_7 and N_8 oligosaccharides could not be identified. However, N_7 was found to be sensitive to both GH62 and GH51 arabinofuranosidases, while N_8 was resistant to both enzymes. Taking the MALDI-CID data and the information from enzyme sensitivity together, N_7 was identified as A^3XXX. N_8 was further analysed by Nuclear Magnetic Resonance (NMR) to gain further insight into the detailed structure of this oligosaccharide (Figure 5D and Table S1).

Chemical-shift assignments were obtained using 2D ^{1}H-^{1}H TOCSY (TOtal Correlated SpectroscopY) and ROESY (Rotating-frame Overhauser Effect SpectroscopY) alongside 2D ^{13}C HSQC (Heteronuclear Single Quantum Correlation) and HSQC-TOCSY experiments. The non-reducing-end Xyl*p* residue was readily identified and the chemical shifts of the H-1 and C-1 were consistent with a β configuration. The Xyl*p* (1→2) linkage to α-Ara*f* was apparent from the intense NOE from Xyl*p* H-1 to Ara*f* H-2 taken together with the downfield shift of Ara*f* C-2 characteristic of a glycosidic bond. The Ara*f*-(1→3)-Xyl*p* and Xyl*p*-(1→4)-Xyl*p* links were also confirmed by a combination of NOEs and the downfield ^{13}C shifts of the linked carbon. H3 and H4 assignments of β-Xyl*p* were apparent from relative TOCSY and NOESY intensities of the cross-peak connecting to H1 (both were stronger for H3). The reducing-end Xyl*p* could not be identified due to peak broadening and overlap (the peaks for the reducing-end-adjacent Xyl*p* are also significantly broadened). Our chemical shift assignments are in accordance with previous ^{1}H and ^{13}C NMR analysis of a feruloylated $D^{2,3}X$ oligosaccharide from shoots of wiregrass (*Cynodon dactylon*) [38], except the nuclei involved in the feruoyl

linkage. Therefore, structure N_8 was identified as $D^{2,3}XX$. In addition, this structure was found to be sensitive to a GH3 β-xylosidase (*Cg*GH3) that cleaves terminal Xyl*p* residues from β-Xyl*p*-(1→2)-Ara*f*-(1→3)-Xyl*p* structures, confirming the assignment.

The oligosaccharide N_9 in the DASH trace was analysed from fraction *Ms*10_45. The $[M+Na]^+$ ion at *m/z* 954.3 corresponds to a $Pent_6$ structure. The MALDI-CID reveals the structure of N_9 as XA^3A^3X (Figure S7):

The $^{1,5}X_1$ (*m/z* 322.1) and the Y_1 (*m/z* 294.2) ions show that the reducing-end Xyl*p* residue is unsubstituted. The reducing-end G_2 (*m/z* 392.2) and G_3 (*m/z* 656.1) product ions show that the second and third Xyl*p* residues from the reducing end are substituted at *O*-3. The presence of the $G3_\alpha/G4_\beta$ ion (*m/z* 906.4, loss of 48 Da) indicate the presence of at least one terminal Ara*f*. GH51 arabinosidase hydrolysis resulted in an oligosaccharide co-migrating with xylotetraose (X_4; data not shown), confirming that both pentose substitutions are Ara*f* residues. Similar to N_6 oligosaccharide, N_9 was resistant to GH62 arabinosidase hydrolysis (Table 2 and Figure S2C).

Structural characterisation of the oligosaccharides N_{10}, N_{11} and N_{12} from the xylanase GH11 digest of *Miscanthus sinensis*

After GH11 hydrolysis four oligosaccharides co-eluted with X_1, X_2, X_3 and X_4 of the xylooligosaccharide standards, three oligosaccharides were structurally characterised with high-energy MALDI-CID and were identical to the previously characterised xylooligosaccharides: XU^2XX and $XU^{(4Me)2}XX$ and XA^3XX. Furthermore three oligosaccharides, N_1, N_3 and N_5, were identified based on their GU unit assigned from the GH10 library. However, we also identified three additional peaks from oligosaccharides of unknown structure which were named N_{10}-N_{12}. For the detailed structural characterisation of these oligosaccharides, the GH11 oligosaccharide mixture was loaded on a SEC column

(Figure S1 fractions *Ms*11_01 to *Ms*11_80) and unknown structures were consequently analysed by high-energy MALDI-CID.

The oligosaccharide N_{10} in the DASH trace was analysed from fraction *Ms*11_05. The $[M+Na]^+$ ion at *m/z* 1012.8 corresponds to a $Pent_5$ structure modified with one MeGlcA. The MALDI-CID reveals the structure of N_{10} as $A^3U^{(4Me)2}XX$ (Figure S8): The presence of the Y_1 (*m/z* 294.3) and Y_2 (*m/z* 426.3) ions indicate that the reducing end and the adjacent Xyl*p* are unsubstituted. The E_3 ion (*m/z* 403.4) and H_3 (*m/z* 435.3) sugar lactone indicate that the third Xyl*p* from the reducing end is modified at *O*-2 with a methylated glucuronic acid, while the $V_{3\alpha}$ product ion (*m/z* 936.3) indicates that the GlcA is modified with a methyl group at *O*-4. The series of Y_3 (*m/z* 748.3) and Y_4 (*m/z* 880.2) indicate that two pentosyl residues are present at the non-reducing end while the absence of the non-reducing end $^{3,5}A_2$ ion (*m/z* 199.0) indicates that the non-reducing end Xyl*p* is substituted by a pentose. Finally the presence of the G_4 reducing end ion (*m/z* 846.3) is indicative of a substitution at *O*-3 of the reducing end Xyl*p* residue. The N_{10} oligosaccharide was found sensitive to GH51 hydrolysis indicating that the pentosyl modification on the fourth Xyl*p* residue from the reducing end is an Ara*f* residue.

The oligosaccharide N_{11} in the DASH trace was analysed from fraction *Ms*11_70. The $[M+Na]^+$ ion at *m/z* 1144.8 corresponds to a $Pent_6$ structure modified with one MeGlcA. The MALDI-CID combined with enzymatic hydrolysis were not conlusive but indicate two putative structures of N_{11} as $B^{2,3}U^{(4Me)2}XX$ or $D^{2,3}U^{(4Me)2}XX$ (Figure S9):

The presence of the Y_1 (*m/z* 294.3) and Y_2 (*m/z* 426.3) ions indicated that the reducing end and the adjacent Xyl*p* are unsubstituted. The E_4 ion (*m/z* 535.4) and H_4 (*m/z* 567.3) sugar lactone indicate that the third Xyl*p* from the reducing end is modified at *O*-2 with a MeGlcA. This assignment is also confirmed by the presence of the G_3 reducing end ion (*m/z* 714.3). The $V_{3\alpha}$ product ion (*m/z* 1068.3) indicated that the GlcA was modified with a methyl group at the *O*-4

position. The presence of a series of Y ions, Y_4 (*m/z* 880.2) and Y_5 (*m/z* 1012.2) indicate that the reducing end sugar sequence of this oligosaccharide is a linear pentose chain. However, the presence of G_4 ion (*m/z* 846.3) indicated that the fourth Xyl*p* residue from the reducing end is modified at *O*-3. Furthermore the reducing end $^{0,2}X_4$ ion (*m/z* 1054.3) is indicative of a modification at *O*-2 of the fifth pentosyl group from the reducing end and therefore pointing to a structure similar to the $D^{2,3}XX$ (Figure 5C). N_{11} oligosaccharide was resistant to *Cg*GH3 xylosidase which would be in accordance with the $B^{2,3}U^{(4Me)2}XX$ assignment or could be explained by steric hindrance by MeGlcA modification on the adjacent Xyl*p* residue in the case of $D^{2,3}U^{(4Me)2}XX$ assignment (Table 2).

The oligosaccharide N_{12} in the DASH trace was analysed from fraction Ms11_05. The $[M+Na]^+$ ion at *m/z* 1276.9 corresponds to a $Pent_7$ structure modified with one MeGlcA. The MALDI-CID combined with enzymatic hydrolysis reveals the structure of N_{12} as $XA^3XU^{(4Me)2}XX$ (Figure S10):

The series of Y ions (Y_1, Y_2, Y_3, Y_4 and Y_5; *m/z* 294.4, *m/z* 426.4, *m/z* 748.3, *m/z* 880.3 and *m/z* 1144.2, respectively) indicate that the oligosaccharide is modified on the third Xyl*p* residue from the reducing end with [Me]GlcA and also on the fifth Xyl*p* residue from the reducing end with a pentosyl group. The presence of H_3 (*m/z* 699.4) sugar lactone confirms that the third Xyl*p* residue is modified by a [Me]GlcA and indicates that this modification is at *O*-2. Additionally, the presence of G_5 (*m/z* 978.3) reducing end ions, indicates that the fifth Xyl*p* residue is modified at *O*-3. The $V_{3\alpha}$ product ion (*m/z* 1200.4) indicates that the GlcA was modified with a methyl group at *O*-4. This oligosaccharide was found sensitive to GH51 and GH62 hydrolysis, indicating that the pentosyl modification on the fifth Xyl*p* residue from the reducing end is an Ara*f* residue, and to GH115 hydrolysis confirming that the uronic acid modification on the third Xyl*p* from the reducing end is a [Me]GlcA (Table 2 and Figure S2A).

Comparative structural analysis of GAX oligosaccharides derived from different grasses and different tissues

The developed reference library was used to characterise variation of GAX in agriculturally important grasses in different aerial tissues. We analysed the xylooligosaccharide products of GH10 and GH11 hydrolysis from stems and leaves of brachypodium, maize, rice, sugar cane, wheat, leaves of miscanthus and stems of barley as well as sugar cane bagasse (Figure 6 and 7).

The DASH profiles of the xylooligosaccharides of the GH10 and the GH11 digests of all grasses analysed look very similar and nearly all peaks can be annotated based on the GU of their oligosaccharides. The main peaks are X_1, X_2, A^3X and XA^3X, respectively. Overall, arabinosylation appears more abundant than glucuronosylation and the latter seems more variable in abundance across grasses. In addition, the ratio of methylated and unmethylated GlcA seems to widely differ in grasses with rice GAX being rather unmethylated, whereas major proportion of GlcA in e.g. miscanthus and sugar cane GAX is methylated. There is also a slight variance in abundance of the more complex oligosaccharides. Moreover, we can detect additional peaks: The most characteristic peaks, M_1 and M_2, are present in the DASH profiles of leaf GAX from all grasses analysed, but very rare in GAX of stem with the exception of rice stem GAX, in which M_1 and M_2 are substantial.

Structural characterisation of the oligosaccharides M_1 and M_2 from leaf GAX

In order to identify the unknown oligosaccharides M_1 and M_2, maize leaf GH10 and rice leaf GH11 oligosaccharide products, respectively, were separated by SEC. The suitable SEC fractions (*Zm*10_65 for structure M_1 and *Os*11_60 for structure M_2) were further analysed by

MALDI-ToF/ToF Mass Spectrometry, structural isomers were separated by HILIC and structurally characterised by high energy MALDI-CID (Figure S11).

The oligosaccharide M_1 in the DASH trace was analysed from fraction *Zm*10_65 (Figure S11A). The unknown oligosaccharide had an *m/z* 690.00 $[M + Na]^+$, corresponding to a $Pent_4$ structure (Figure S11B). It did not co-migrate with Xyl_4 in DASH, suggesting it is a substituted xylooligosaccharide. Off-line HILIC-MALDI-MS separated the two structural isomers with *m/z* 690.00 (M_1 and XA^3X, Figure S11C) and the respective MALDI-CID spectra revealed the structure of M_1 as $D^{2,3}X$ (Figure S12): The presence of a series of Y ions (Y_1, Y_2 and Y_3; *m/z* 294.3, 426.3 and 558.3 respectively) and the presence of $^{1,4}X_2$ and $^{1,5}X_3$ ions (*m/z* 454.3 and 586.3, respectively) indicate that the M_1 is a linear DP4 pentose chain. The presence, however, of the G_3 product ion (*m/z* 642.4, loss of 48Da) indicates the existence of an Ara*f* residue in the structure. In addition, the presence of the G_4 product ion (*m/z* 656.5) gives evidence of a terminal Xyl*p* residue on the structure. Finally, the presence of the $^{0,2}X_2$ cross-ring ion (*m/z* 600.3) suggests that a pentose is 2-linked on the penultimate Xyl*p* residue. The assignment of M_1 as $D^{2,3}X$ was confirmed by the finding that the structure is resistant to the hydrolysis with GH62 arabinosidase, GH67 and GH115 glucuronidases, but sensitive to *Cg*GH3 β-xylosidase hydrolysis (Table 2).

The oligosaccharide M_2 in the DASH trace was analysed from fraction *Os*11_60 (Figure S11D). The unknown oligosaccharide had an *m/z* 954.5 $[M + Na]^+$, corresponding to a $Pent_6$ structure (Figure S11E). It did not co-migrate with Xyl_6 in DASH, suggesting it is a substituted xylooligosaccharide. Off-line HILIC-MALDI-CID revealed that this was the only structural isomer present in the sample (Figure S11F). The MALDI-CID revealed the structure of M_2 as $XD^{2,3}XX$ (Figure S13): The significant Y_1 ion (*m/z* 294.4) and the $^{1,5}X_1$ ion (*m/z* 322.4) indicate that the 2-AA labelled Xyl*p* residue is not substituted. The Y_2 and Y_3 ions (*m/z* 426.4 and 690.4, respectively) show that the second Xyl*p* is unsubstituted but that the third Xyl*p* residue from

the reducing end is modified with two pentose residues. The reducing-end G_3 (*m/z* 524.5) product ion and the concomitant presence of the W_3 (*m/z* 556.4) sugar lactone indicate that the substitution on the third Xyl*p* from the reducing end is at *O*-3. This assignment is further verified by the presence of the D_2 (*m/z* 271.4) product ion. The existence of the G_4 (*m/z* 906.5) product ion indicates the presence of an Ara*f* substitution. Finally the presence of the $^{3,5}A_2$ non-reducing end cross-ring fragment ion indicates the presence of an unsubstituted Xyl*p* at the non-reducing end. The assignment of M_2 as $XD^{2,3}XX$ was confirmed by the finding that the structure is resistant to the hydrolysis with GH62 arabinosidase, GH67 and GH115 glucuronidases, but sensitive to *Cg*GH3 β-xylosidase hydrolysis (Table 2).

Relative quantitative differences in oligosaccharide structure and abundance in *Miscanthus sinensis* stem and leaf GAX

By DASH the relative quantity of reducing-end labelled oligosaccharides within a sample can be determined by the relative fluorescence intensity of the corresponding electropherogram peaks [27]. To analyse the difference in GAX structure in stems and leaves in greater detail, we quantified the frequency of substitution of all characterised library oligosaccharides in the GH10 and GH11 digests of three biological replicates of miscanthus using the DASHboard software (Figure 8). The substitution frequency was calculated after normalisation of values by comparing the abundance of side chains to the Xyl*p* residues in the backbone.

First of all, the overall substitution frequency of xylan with [Me]GlcA, Ara*f* and $D^{2,3}$ calculated from the two different digests with GH10 and GH11, respectively, is consistent with only a minor difference for the $D^{2,3}$ structure in leaves, showing a slightly higher frequency in the GH10 digest only (1.3% versus 1%, p-value 0.01).

In both, stems and leaves of miscanthus, arabinosylation of xylan is more frequent than glucuronosylation, ranging between 10-12% of Ara*f* substitutions versus 3-6% of [Me]GlcA

substitutions. However, there is a significant difference between tissues regarding the amount of $D^{2,3}$ substitution both in the GH10 (p-value 0.002) and the GH11 (p-value 0.02) digests. In leaves the $D^{2,3}$ substitution frequency of the GH10 and GH11 digest is 1.3% and 1%, respectively, whereas in stem it is down to 0.2% and 0.4%, respectively. Concomitantly, with the presence of the additional $D^{2,3}$ structure ($D^{2,3}X$ and $XD^{2,3}XX$) in leaf xylan, our data show that the frequency of the $D^{2,3}$ structure is increased in leaves.

Discussion

In this study we used a combination of SEC, DASH, HILIC and MALDI-CID to elucidate the detailed structure of grass GAX oligosaccharides released by GH10 and GH11 xylanases. Based on these data, we developed a DASH reference library of structurally identified oligosaccharides and expressed their migration in GU. It has been reported previously that DASH can separate oligosaccharides with minor structural differences like methylation of GlcA [27]. Here, we show that DASH can successfully resolve structural isomers as demonstrated by the separation of the three DP5 oligosaccharides (*m/z* 822.0). Few approaches allow for wider screening of cell wall polysaccharide structures. Microarray based glycan profiling techniques [39, 40], for example, integrate the sequential extraction of glycans with the generation of microarrays, which are probed with monoclonal antibodies (mAbs) or carbohydrate binding molecules (CBMs) with specificities for cell wall components. Although Microarray based glycan profiling is a powerful technique offering high throughput analysis of a wider variety of cell wall polymers, it is limited by the availability of mAbs and CBMs and their epitope specificity. A non-destructive high throughput method for the compositional analysis of plant cell wall polymers is Fourier Transformed Mid-Infrared (FT-IR) spectroscopy [41]. This approach can provide structural information about substitution nature and frequency of cereal arabinoxylan [42]. The DASH reference library generated here allows the fast and

445 robust comparison of a large number of GAX samples (96 samples in 50 min), providing
446 detailed structural and quantitative information on biomass structural variation. Information
447 gathered from DASH analysis will ultimately greatly facilitate the biorefining selection process
448 of appropriate biomass.

449 We analysed the GAX structure of a number of different grass species (miscanthus, barley,
450 brachypodium, maize, rice, sugar cane and wheat). The DASH profiles of all grasses analysed
451 and the quantification data of xylan substitution of *Miscanthus sinensis* showed that Ara*f* side
452 chains are more frequent than [Me]GlcA substitutions, which is consistent with sugar
453 composition analysis of wheat straw [26, 43] and MS and NMR analysis of miscanthus, rice
454 and brachypodium of the entire actively growing aerial portions [26]. Unlike cereal grain AX
455 where 2-linked Ara*f* residues are abundant [1, 44-46], we did not detect any such substitutions,
456 which is consistent with earlier reports on the structure of grass GAX in lignified tissues [47-
457 49]. The overall xylan structure of the different grass species and of the different tissues is
458 remarkably conserved, which is in line with earlier studies on grass xylan structure [50].
459 However, differences between species and tissues are detectable. Interestingly, the xylan
460 structure of rice stems appears more similar to the xylan structure in leaves than to other xylan
461 stem structures, forming an exception in the grasses analysed, perhaps because of the immature
462 developmental stage of the culms collected. Plants exploit a number of variations in the xylan
463 structure that were not studied in this work, e.g. the pattern of substitutions along the xylan
464 backbone, feruloylation and coumaroylation of Ara*f* and acetylation of backbone Xyl*p*. These
465 could be studied by exploitation of additional carbohydrate active enzymes and generation of
466 additional standards in the DASH mobility library.

467 The most characteristic tissue-specific differences were identified in form of the two
468 oligosaccharides $D^{2,3}X$ and $XD^{2,3}XX$, which are (apart from in rice) scarce in stem xylan but
469 abundant in leaf xylan. Quantification of the overall frequency of the $D^{2,3}$ side chain in

miscanthus suggests that it is significantly more abundant in leaves than in stem. The $D^{2,3}$ structure has been linked to feruloylation of xylan, and hence crosslinking and reduced digestibility [8]. Interestingly, the lignin amount and composition also differs in cell walls of miscanthus stems versus leaves [51]. Therefore xylan structural changes in different tissues might reflect specific levels and positioning of cross-linking, which can be a way of adjusting polymer structures to different functional requirements of the cell wall depending on its particular composition. A disaccharide side chain on xylan has been reported in sorgum [52] and switchgrass [53] although not detected by Bowman *et al.* (2014) in a similar xylan analysis of switchgrass. The presence of this disaccharide side chain can not be excluded although with our approach all disaccharide side chain modifications positively identified were β-Xyl*p*-(1→2)-α-Ara*f*-(1→3) structures. The only exception was the GH11 product, putatively assigned as $D^{2,3}U^{(4Me)2}XX$, which was found resistant to *Cg*GH3 β-xylosidase. Resistance to *Cg*GH3 β-xylosidase could either be due to steric hindrance of the enzyme or could indicate the presence of an Ara*f*-(1→2)-α-Ara*f*-(1→3) side chain on this GH11 hydrolysis product ($B^{2,3}U^{(4Me)2}XX$). However, the Ara*f*-(1→2)-Ara*f*-(1→3)- adjacent to a Xyl*p* modified by a GlcA residue would be in a different substitution context to the Mazumder *et al.* (2010) and Verbruggen *et al.* (1998) reported oligosaccharides. If the latter is true then we could hypothesise that GH10 and GH11 xylanases hydrolyse slightly different parts of GAX. It is also, however, possible that the Ara*f*-(1→2)-α-Ara*f*-(1→3) side chain is not present in the tissues and plant species analysed here, or that it is present in amounts below the detection level.

It has been reported that the degree of methylation of GlcA on xylan varies between grass species and that in miscanthus methylation of xylan is more predominant than in, for example, rice or Brachypodium [48, 49]. This finding is consistent with our data on miscanthus xylan.

Some qualitative differences of oligosaccharides reflect enzyme specificity and distinct tolerance of the xylanases GH10 and GH11 to substitutions [28], for example A^3X from GH10 and XA^3XX from GH11. However, some oligosaccharide such as $A^3U^{(4Me)2}XX$, $D^{2,3}U^{(4Me)2}XX$ (or $B^{2,3}U^{(4Me)2}XX$) and $XA^3XU^{(4Me)2}XX$, although only minor products, were only detectable in GH11 digestion and might indicate that GH10 and GH11 act on different domains of grass xylan as they resemble different substitution patterns. The fact that the level of 3-linked Ara*f* and [Me]GlcA substitutions are remarkably similar independent of the xylanase used, does not necessarily support this hypothesis. Surprisingly, some of these minor digestion products of GH11 harbour substitutions at the non-reducing end of the oligosaccharide, which are not predicted as products and might be the result of either using the enzymes in excess or that the GH11 enzyme preparation used in this study was contaminated with small amounts of either GH10 xylanase or β-xylosidase. Alternatively, these oligosaccharides could derive from the non-reducing end of the xylan polymer and would be produced by a single cleavage at the reducing end of the oligosaccharide.

Conclusions

As characterised here in grass stems and leaves from several species, mainly three GAX sugar side chains, 3-linked Ara*f*, 2-linked GlcA/MeGlcA and 3-linked Xyl*p*-(1→2)-Ara*f*, are utilised to decorate the xylan molecule. Our GAX oligosaccharide reference library of DASH mobilities, developed in this study, provides a means to study structural aspects of xylan and might help to shed light on how structural changes of xylan correlate with the interaction of this polysaccharide with other cell wall components, how this influences its biological function, its mechanical properties and recalcitrance of the cell wall. Furthermore it provides a high-throughput quantitative method for the selection of suitable lignocellulosic biomass and tailoring of biorefining processes.

519

520 **Methods**

521 **Plant material**

522 The plant materials used in this study were fresh material from *Miscanthus sinensis*
523 (miscanthus; Wageningen University, Netherlands, Luisa Trindade & Oene Dolstra),
524 *Hordeum vulgare* (barley; University of Dundee, Claire Halpin), *Brachypodium distachyon*
525 (Brachypodium; (grown at University of Cambridge greenhouse), *Zea mays* (maize; University
526 of Cambridge, Paul Dupree), *Oryza sativa* (rice; grown at University of Cambridge
527 greenhouse), *Saccharum officinarum* (sugar cane; EMBRAPA-Brazil, Christiane Farinas),
528 sugar cane bagasse (University of Sao Paulo-Brazil, Marcos Buckeridge) and *Triticum*
529 *aestivum* (wheat; University of Nottingham, Greg Tucker).

530

531 **Extraction of Alcohol Insoluble Residue (AIR)**

532 Plant stems and leaves were harvested, submerged in 96% (v/v) ethanol, and boiled at 70˚C for
533 30 min to inactivate enzymes. Following homogenisation using a ball mixer mill (Glen
534 Creston), the pellet was collected by centrifugation (4000 x g for 15 min) and was washed with
535 100% (v/v) ethanol, twice with chloroform:methanol (2:1), followed by successive washes
536 with 65% (v/v), 80% (v/v) and 100% (v/v) ethanol. The remaining pellet of AIR was air dried.
537 Aqueous suspensions (0.5mg/ml; at 21°C) of AIR were prepared using a glass homogeniser
538 and kept for further analysis.

539

540 **Hemicellulose extraction**

541 Hemicelluloses were extracted by treating AIR preparations (2 g for Size exclusion
542 chromatography (SEC) fractionation; 50 µg for small scale digestions) with a small volume of
543 4M NaOH (5 ml for SEC fractions; 20 µl for small scale digestions) for 1h at room temperature

before the pH was adjusted to about pH 6.0 with 1N HCl; 50 mM ammonium acetate buffer pH 6.0 was added (200 ml for SEC fractionation; 0.5 ml for small scale digestions). Note: Alkali treatments results in the removal of acetylation and feruloylation.

Enzymatic hydrolysis and enzymes

Glycoside hydrolases of different Carbohydrate Active enZYme families were used in this study: GH10 endo-β-1,4-xylanase *Cj*GH10A from *Cellvibrio japonicus* [54]; GH11 endo-β-1,4-xylanase *Np*GH11A from *Neocallimastix patriciarum* [55]; GH67 α-glucuronidase *Cj*GH67 from *Cellvibrio japonicus* [35, 56]; and GH115 α-glucuronidase *Bo*GH115 from *Bacteroides ovatus* [36, 57]; GH62 α-arabinofuranosidase *Pa*GH62 from *Penicillium aurantiogriseum* [58], a gift from Novozymes; GH51 α-arabinofuranosidase *Pc*GH51 from *Pseudomonas cellulose* [34]; GH3 xylosidase *Cg*GH3 from *Chaetomium globosum* (NS39127) and GH3 β-1,4 xylosidase *Tr*GH3 from *Trichoderma reesei* [59] were both gifts from Novozymes. All enzymes were added at a final concentration of 2 μM and incubated at 21 °C under constant shaking for 24 h. Enzymatic hydrolysis progression was monitored by Polysaccharide Analysis Using Carbohydrate Gel Electrophoresis [60] and if necessary, the enzyme amount was adjusted to ensure complete hydrolysis.

Enzyme deactivation and removal of undigested material

Enzymes were then deactivated by boiling for 30 min undigested cell wall material was removed by centrifugation. In case of SEC, centrifugation was followed by filtration (Whatman 45 μm). Samples were then dried in a centrifugal evaporator.

Size exclusion chromatography

Dried hydrolysed cell wall material (corresponding to 2 g of AIR) were resuspended in 2ml distilled water. Sample (2 ml) was applied onto the column and eluted with distilled water. SEC was performed on a gravity driven BioGel P-2 column (190 x 2.5 cm, BioRad) as previously described [61]. 2 ml fractions were collected, concentrated to 100 µl in a centrifugal evaporator and 10 µl were analysed by DASH as described below. A total of 80 SEC fractions were collected for each grass species and xylanase hydrolysis analysed. The SEC fraction naming system includes information on the species (*Ms: Miscanthus sinensis*, *Zm: Zea mays* and *Os: Oryza sativa*), the xylanase (10: GH10 and 11: GH11) and the fraction number (_01 to _80). The fractions with xylooligosaccharides of interest were subjected to secondary enzymatic hydrolysis or dried, desalted, reductively aminated with 2-anthranilic acid (2-AA), purified by HILIC and structurally characterised by high energy MALDI-CID.

APTS labelling and analysis by DASH

The derivatisation of oligosaccharides with 8-aminopyrene-1,4,6-trisulfonate (APTS) was performed according to previously developed protocols [27]. A set of 7 fluorophore (DY-481XL-NHS ester) labelled amino acids and peptides was used as electrophoretic mobility standards (Asp-Asp-Asp-Asp; Asp-Asp-Asp; Glu-Glu; Cysteic acid; L-2-Aminoadipic acid; Glycine; Gly-Gly-Gly) to align the electropherograms. These electrophoretic mobility standards were mixed with each sample prior to DASH separation serving as internal mobility markers. DASH data generated by the DNA sequencer were processed with the DASHboard software [27]. Control experiments without the substrates were performed under the same conditions in order to identify any non-specific compounds in the enzymes or labelling reagents. An APTS derivatised dextran ladder (0.1 M TFA hydrolysis at 100 °C for 2 h; 50 µg μl^{-1} dextran in 200 µl TFA solution) was simultaneously separated by DASH and was used to provide the GU migration positions.

Desalting and clean up for HILIC separation

Xylooligosaccharides from SEC fractions were desalted using HyperSep Hypercarb cartridges (Thermo-Hypersil-Keystone, Runcorn, Cheshire, UK) as previously described [29]. Oligosaccharides were lyophilised and then derivatised with 2-aminobenzoic acid (2-AA) as described below.

Reductive amination and purification for HILIC separation

SEC purified xylooligosaccharides were reductively aminated with 2-AA (Sigma) and then purified from the reductive amination reagents using a Glyko Clean S cartridge (Prozyme, San Leandro, CA) as previously described [62].

HILIC-MALDI-MS and MALDI-MS/MS CID analysis

Capillary HILIC was carried out using an LC-Packings Ultimate system (Dionex, CA, USA) equipped with an amide-80 column (300 μm x 25 cm; 3 μm particle size; Dionex) as previously described [62]. Briefly, the LC system was used to generate the gradient that flowed at 3 μL min^{-1}. Solvent A was 50 mm ammonium formate adjusted to pH 4.4 with formic acid. Solvent B was 5% solvent A in acetonitrile. The labelled oligosaccharides dissolved in 95% acetonitrile were loaded onto the column (20 μl) and eluted with increasing aqueous concentrations. The following gradient conditions were applied: 0 min, 5% solvent A, 95% solvent B; 6 min, 25% solvent A, 75% solvent B; 86 min, 45% solvent A, 55% solvent B. The system operated at ambient temperature. The column eluent passed through a capillary UV detector (set at 254 nm) to the MALDI sample spotter. For HILIC-MALDI-ToF/ToF Mass Spectrometry a Probot sample fraction system (Dionex) was employed for automated spotting of the HPLC eluent onto a MALDI target at 20 s intervals. After air drying, the sample spots were overlaid with

0.5 µl 2,5-DHB matrix (1 mg ml^{-1} in 50% aqueous methanol) and analysed by MALDI-ToF/ToF-MS on an AB-Sciex 4700. The MS spectra were obtained in automatic mode with an average 1500 laser shots/spectrum (mass range 400-2500 Da). The oligosaccharide molecular ions $[M+Na]^+$ were identified in the MALDI data and their HILIC elution positions were determined by carrying out an extracted ion chromatogram (EIC). High energy MALDI-CID spectra were acquired with an average 10,000 laser shots/spectrum, using a high collision energy (1 kV). The oligosaccharide ions were allowed to collide in the CID cell with argon at a pressure of 2×10^{-6} Torr.

NMR analysis

Saponified miscanthus AIR was hydrolysed with GH10 xylanase, followed by GH115 xylan glucuronosidase, GH51 arabinofuranosidase and *Tr*GH3 β-1,4-xylanase using enzyme hydrolysis conditions described above. The resulting N_8 oligosaccharide was then isolated by SEC as described above. Consequently, SEC fractions containing the N_8 oligosaccharide were pooled, solubilised in 0.6 ml D_2O and analysed by NMR.

NMR spectra were recorded at 298 K with a Bruker AVANCE III spectrometer operating at 600 MHz equipped with a TCI CryoProbe. Two-dimensional ^1H-^1H TOCSY, ROESY, ^{13}C HSQC and HSQC-TOCSY experiments were performed, using established methods [63]; the mixing times were 70 ms and 200 ms for the TOCSY and ROESY experiments, respectively. Chemical shifts were measured relative to internal acetone ($\delta H = 2.225$, $\delta C = 31.07$ ppm). Data were processed using the Azara suite of programs (v. 2.8, copyright 1993-2014, Wayne Boucher and Department of Biochemistry, University of Cambridge, unpublished) and chemical-shift assignment was performed using Analysis v2.2 [64].

642 **Oligosaccharide naming system**

643 The various hydrolysis products are named according to the Faure *et al.* (2009) naming system.
644 This naming system, utilises a single letter code where the uppercase letters are identifying the
645 substituents of the main xylan chain. The letter "A" is attributed to single Ara*f* substitution, the
646 letter "U" is used for single GlcA substitution and the letter "X" for unsubstituted Xyl*p*
647 residues. The superscript numbers indicate substitution linkage position on Xyl*p*. Information
648 of side chain modifications is included in the superscript part of the name, for example "Me"
649 for methylation. Finally, further substitutions of the side chains receive a new letter assignment,
650 for example, the Ara*f*-(1→2)-α-Ara*f*-(1→3) side chain has been designated the "$B^{2,3}$" character
651 and the β-Xyl*p*-(1→2)-α-Ara*f*-(1→3) substitution has been designated the "$D^{2,3}$" character
652 [65].

653

654 **DASHboard software and substitution frequency quantitation**

655 Data generated by the DNA sequencer were processed in DASHboard software [27] which was
656 developed to complement the profiling technique and perform tasks such as visualisation of
657 data, alignment of electropherograms, peak area quantification and export to Excel for further
658 analysis.

659 Relative quantitation of substitution frequency was calculated after normalisation of values by
660 comparing the abundance of side chains to the Xyl*p* residues in the backbone. Value
661 normalisation allowed the accurate peak area calculation by DASH software because
662 electropherograms from higher dilution were utilised for the peak area calculation of highly
663 abundant mono- and oligosaccharides (such as xylose and xylobiose) and the
664 electropherograms of same samples but of lower dilution were utilised for peak area
665 calculation of less abundant oligosaccharides. The peak area ratio between each of the xylanase

oligosaccharide products and a reference oligosaccharide (XA^3X and XA^3XX for GH10 and GH11 digestions, respectively) resulted in the normalised values for each oligosaccharide. The total backbone Xyl*p* present in the digested xylan was calculated as the sum of the relative quantity of each of these digestion products multiplied by the number of Xyl*p* residues present in each structure (total backbone Xyl*p*). For each specific side chain, side chain substitution was calculated as the sum of the relative quantity of each of these side chains (Ara*f*, [Me]GlcA and $D^{2,3}$) multiplied by the number of side chains present in each structure (side chain substitution). Hence, side chain substitution frequency was calculated as the ratio between (side chain substitution) and (total backbone Xyl*p*).

Abbreviations

DASH: DNA sequencer-Assisted Saccharide analysis in High throughput; Ara*f*: Arabinofuranose/arabinofuranosyl; Xyl*p*: Xylopyranose/xylopyranosyl; [Me]GlcA: [Methyl]GlucuronicAcid/glucuronyl; (G)AX: (Glucurono)arabinoxylan; AIR: Alcohol Insoluble Residue; GH: Glycosyl hydrolase; GU: Glucose units

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

All authors read and approved the final manuscript.

Availability of data and materials

688 The dataset supporting the conclusions of this article are included within the article and its
689 additional files.

690 **Competing interests**

691 The authors declare that they have no competing interests.

692 **Funding**

693 The work was funded by the BBSRC (BSBEC BIOMASS BB/G016216/1 and BB/K005537/1)
694 and the European Community's Seventh Framework Programme SUNLIBB (FP7/2007–2013)
695 under the grant agreement no. 251132 to P.D. The Cambridge NMR facility infrastructure is
696 funded by the BBSRC and the Wellcome Trust. We are grateful for plant material from
697 Aberystwyth, Rothamsted and Wageningen University, NL.

698 **Authors' contributions**

699 TT designed the experimental part and carried out the analysis of xylooligosaccharides. MS
700 and CF supported the analysis. KS performed the NMR analysis. DVR updated the
701 DASHboard software. TT, NA and PD analysed the data and wrote the manuscript.

702 **Acknowledgements**

703 We thank Dr Tina Theys and Dr Rita Marques for technical support.

704 **Authors' information**

705 [a] Department of Biochemistry, University of Cambridge, Hopkins Building, The Downing site,
706 Tennis Court Road, Cambridge CB2 1QW, UK; [b] Department of Biochemistry, University of
707 Cambridge, Sanger Building, 80 Tennis Court Road, Cambridge CB2 1GA, UK; [1] Present
708 address: Scion, 49 Sala Street, Private Bag 3020, Rotorua 3046, New Zealand; [2] Present
709 address: Frontiers, WeWork, 1 Fore St, London EC2Y 5EJ, UK; [3] Present address: ideaSpace
710 South, Cambridge Biomedical Campus, Bay 13 Hills Road, Cambridge, CB2 0SP, UK.

711

References

1. Ebringerová A, Heinze T: **Xylan and xylan derivatives – biopolymers with valuable properties, 1. Naturally occurring xylans structures, isolation procedures and properties**. *Macromolecular Rapid Communications* 2000, **21**:542-556.
2. Saulnier L, Vigouroux J, Thibault J-F: **Isolation and partial characterization of feruloylated oligosaccharides from maize bran**. *Carbohydr Res* 1995, **272**(2):241-253.
3. Tian L, Gruppen H, Schols HA: **Characterization of (Glucurono)arabinoxylans from Oats Using Enzymatic Fingerprinting**. *J Agric Food Chem* 2015, **63**(50):10822-10830.
4. Hatfield RD, Rancour DM, Marita JM: **Grass Cell Walls: A Story of Cross-Linking**. *Frontiers in Plant Science* 2017, **7**(2056).
5. Mueller-Harvey I, Hartley RD, Harris PJ, Curzon EH: **Linkage of p-coumaroyl and feruloyl groups to cell-wall polysaccharides of barley straw**. *Carbohydr Res* 1986, **148**(1):71-85.
6. de O. Buanafina MM: **Feruloylation in Grasses: Current and Future Perspectives**. *Molecular Plant*, **2**(5):861-872.
7. Ishii T: **Structure and functions of feruloylated polysaccharides**. *Plant Science* 1997, **127**(2):111-127.
8. Wende G, Fry SC: **2-O-β-d-xylopyranosyl-(5-O-feruloyl)-l-arabinose, a widespread component of grass cell walls**. *Phytochemistry* 1997, **44**(6):1019-1030.
9. Chiniquy D, Sharma V, Schultink A, Baidoo EE, Rautengarten C, Cheng K, Carroll A, Ulvskov P, Harholt J, Keasling JD *et al*: **XAX1 from glycosyltransferase family 61 mediates xylosyltransfer to rice xylan**. *Proc Natl Acad Sci USA* 2012, **109**(42):17117-17122.
10. Appeldoorn MM, de Waard P, Kabel MA, Gruppen H, Schols HA: **Enzyme resistant feruloylated xylooligomer analogues from thermochemically treated corn fiber contain large side chains, ethyl glycosides and novel sites of acetylation**. 2013, **381**:33-42.
11. Allerdings E, Ralph J, Steinhart H, Bunzel M: **Isolation and structural identification of complex feruloylated heteroxylan side-chains from maize bran**. *Phytochemistry* 2006, **67**(12):1276-1286.

12. Appeldoorn MM, Kabel MA, Van Eylen D, Gruppen H, Schols HA: **Characterization of Oligomeric Xylan Structures from Corn Fiber Resistant to Pretreatment and Simultaneous Saccharification and Fermentation**. *J Agric Food Chem* 2010, **58**(21):11294-11301.
13. Coelho E, Rocha MAM, Moreira ASP, Domingues MRM, Coimbra MA: **Revisiting the structural features of arabinoxylans from brewers' spent grain**. *Carbohydr Polymers* 2016, **139**:167-176.
14. Schendel RR, Meyer MR, Bunzel M: **Quantitative Profiling of Feruloylated Arabinoxylan Side-Chains from Graminaceous Cell Walls**. *Frontiers in Plant Science* 2016, **6**(1249).
15. Ishii T: **Acetylation at O-2 of arabinofuranose residues in feruloylated arabinoxylan from bamboo shoot cell-walls**. *Phytochemistry* 1991, **30**(7):2317-2320.
16. McCann MC, Carpita NC: **Designing the deconstruction of plant cell walls**. *Current Opinion in Plant Biology* 2008, **11**(3):314-320.
17. Simmons TJ, Mortimer JC, Bernardinelli OD, Pöppler A-C, Brown SP, deAzevedo ER, Dupree R, Dupree P: **Folding of xylan onto cellulose fibrils in plant cell walls revealed by solid-state NMR**. 2016, **7**:13902.
18. DeMartini JD, Pattathil S, Miller JS, Li HJ, Hahn MG, Wyman CE: **Investigating plant cell wall components that affect biomass recalcitrance in poplar and switchgrass**. *Energy & Environmental Science* 2013, **6**(3):898-909.
19. Hoije A, Sandstrom C, Roubroeks JP, Andersson R, Gohil S, Gatenholm P: **Evidence of the presence of 2-O-β-d-xylopyranosyl-α-l-arabinofuranose side chains in barley husk arabinoxylan**. *Carbohydr Res* 2006, **341**(18):2959-2966.
20. Marcotuli I, Hsieh YSY, Lahnstein J, Yap K, Burton RA, Blanco A, Fincher GB, Gadaleta A: **Structural Variation and Content of Arabinoxylans in Endosperm and Bran of Durum Wheat (Triticum turgidum L.)**. *J Agric Food Chem* 2016, **64**(14):2883-2892.
21. Pastell H, Virkki L, Harju E, Tuomainen Pi, Tenkanen M: **Presence of α-linked 2-O-β-d-xylopyranosyl-α-l-arabinofuranosyl side chains in cereal arabinoxylans**. *Carbohydr Res* 2009, **344**(18):2480-2488.
22. Toole G, Wilson R, Parker M, Wellner N, Wheeler T, Shewry P, Mills E: **The effect of environment on endosperm cell-wall development in Triticum aestivum during grain filling: an infrared spectroscopic imaging study**. *Planta* 2007, **225**(6):1393-1403.

23. Ordaz-Ortiz JJ, Devaux M-F, Saulnier L: **Classification of Wheat Varieties Based on Structural Features of Arabinoxylans As Revealed by Endoxylanase Treatment of Flour and Grain**. *J Agric Food Chem* 2005, **53**(21):8349-8356.
24. Bowman MJ, Dien BS, Vermillion KE, Mertens JA: **Structural characterization of (1□2)-β-xylose-(1□3)-α-arabinose-containing oligosaccharide products of extracted switchgrass (Panicum virgatum, L.) xylan after exhaustive enzymatic treatment with α-arabinofuranosidase and β-endo-xylanase**. *Carbohydr Res* 2014, **398**:63-71.
25. Kulkarni A, Pena MJ, Avci U, Mazumder K, Urbanowicz B, Pattathil S, Yin Y, O'Neill MA, Roberts AW, Hahn MG *et al*: **The ability of land plants to synthesize glucuronoxylans predates the evolution of tracheophytes**. *Glycobiology* 2011, **22**(3):1-15.
26. Peña MJ, Kulkarni AR, Backe J, Boyd M, O'Neill MA, York WS: **Structural diversity of xylans in the cell walls of monocots**. *Planta* 2016, **244**(3):589-606.
27. Li X, Jackson P, Rubtsov D, Faria-Blanc N, Mortimer J, Turner S, Krogh K, Johansen K, Dupree P: **Development and application of a high throughput carbohydrate profiling technique for analyzing plant cell wall polysaccharides and carbohydrate active enzymes**. *Biotechnology for Biofuels* 2013, **6**(1):94.
28. Biely P, Vrsanská M, Tenkanen M, Kluepfel D: **Endo-[beta]-1,4-xylanase families: differences in catalytic properties**. *J Biotechnol* 1997, **57**(1-3):151-166.
29. Maslen SL, Goubet F, Adam A, Dupree P, Stephens E: **Structure elucidation of arabinoxylan isomers by normal phase HPLC-MALDI-TOF/TOF-MS/MS**. *Carbohydr Res* 2007, **342**(5):724-735.
30. Guile GR, Rudd PM, Wing DR, Prime SB, Dwek RA: **A Rapid High-Resolution High-Performance Liquid Chromatographic Method for Separating Glycan Mixtures and Analyzing Oligosaccharide Profiles**. *Anal Biochem* 1996, **240**(2):210-226.
31. Chai W, Piskarev V, Lawson AM: **Negative-Ion Electrospray Mass Spectrometry of Neutral Underivatized Oligosaccharides**. *Analytical Chemistry* 2001, **73**(3):651-657.
32. Bromley JR, Busse-Wicher M, Tryfona T, Mortimer JC, Zhang Z, Brown DM, Dupree P: **GUX1 and GUX2 glucuronyltransferases decorate distinct domains of glucuronoxylan with different substitution patterns**. *Plant J* 2013, **74**(3):423-434.

33. Spina E, Sturiale L, Romeo D, Impallomeni G, Garozzo D, Waidelich D, Glueckmann M: **New fragmentation mechanisms in matrix-assisted laser desorption/ionization time-of-flight/time-of-flight tandem mass spectrometry of carbohydrates**. *Rapid Commun Mass Spectrom* 2004, **18**(4):392-398.
34. Beylot MH, Emami K, McKie VA, Gilbert HJ, Pell G: **Pseudomonas cellulosa expresses a single membrane-bound glycoside hydrolase family 51 arabinofuranosidase**. *Biochem J* 2001, **358**(3):599-605.
35. Nurizzo D, Nagy T, Gilbert HJ, Davies GJ: **The Structural Basis for Catalysis and Specificity of the Pseudomonas cellulosa α-Glucuronidase, GlcA67A**. *Structure* 2002, **10**(4):547-556.
36. Rogowski A, Basle A, Farinas CS, Solovyova A, Mortimer JC, Dupree P, Gilbert HJ, Bolam DN: **Evidence That GH115 α-Glucuronidase Activity, Which Is Required to Degrade Plant Biomass, Is Dependent on Conformational Flexibility**. *J Biol Chem* 2014, **289**(1):53-64.
37. Thibault J-F: **Separation of a-d-galacturonic acid oligomers by chromatography on polyacrylamide gel**. *Journal of Chromatography A* 1980, **194**(3):315-322.
38. Himmelsbach D, Hartley RD, Borneman W, Poppe L, Van Halbeek H: **Structure of a feruloylated arabinoxylan tetrasaccharide that contains the β-D-Xylp-(1 → 2)-α-L-Araf element by1H and13C NMR spectrocopy.** *Magnetic Resonance in Chemistry* 1994, **32**(3):158-165.
39. Moller I, Marcus SE, Haeger A, Verhertbruggen Y, Verhoef R, Schols H, Ulvskov P, Mikkelsen JrD, Knox JP, Willats W: **High-throughput screening of monoclonal antibodies against plant cell wall glycans by hierarchical clustering of their carbohydrate microarray binding profiles**. *Glycoconjugate J* 2008, **25**(1):37-48.
40. da Costa RM, Pattathil S, Avci U, Lee SJ, Hazen SP, Winters A, Hahn MG, Bosch M: **A cell wall reference profile for Miscanthus bioenergy crops highlights compositional and structural variations associated with development and organ origin**. *New Phytol* 2017, **213**(4):1710-1725.
41. Badhan A, Wang Y, McAllister TA: **Analysis of Complex Carbohydrate Composition in Plant Cell Wall Using Fourier Transformed Mid-Infrared Spectroscopy (FT-IR)**. In: *Protein-Carbohydrate Interactions: Methods and Protocols*. Edited by Abbott DW, Lammerts van Bueren A. New York, NY: Springer New York; 2017: 209-214.

42. Robert P, Marquis M, Barron C, Guillon F, Saulnier L: **FT-IR Investigation of Cell Wall Polysaccharides from Cereal Grains. Arabinoxylan Infrared Assignment**. *J Agric Food Chem* 2005, **53**(18):7014-7018.
43. Sun X-F, Sun, Fowler P, Baird MS: **Extraction and Characterization of Original Lignin and Hemicelluloses from Wheat Straw**. *J Agric Food Chem* 2005, **53**(4):860-870.
44. Bengtsson S, Aman P, Andersson RE: **Structural studies on water-soluble arabinoxylans in rye grain using enzymatic hydrolysis**. *Carbohydr Polymers* 1992, **17**(4):277-284.
45. Gruppen H, Kormelink FJM, Voragen AGJ: **Water-unextractable Cell Wall Material from Wheat Flour. 3. A Structural Model for Arabinoxylans**. *J Cereal Sci* 1993, **18**(2):111-128.
46. Verbruggen MA, Beldman G, Voragen AGJ: **Enzymic degradation of sorghum glucuronoarabinoxylans leading to tentative structures**. *Carbohydr Res* 1998, **306**(1-2):275-282.
47. Bowman MJ, Dien BS, Vermillion KE, Mertens JA: **Structural characterization of (1-->2)-beta-xylose-(1-->3)-alpha-arabinose-containing oligosaccharide products of extracted switchgrass (Panicum virgatum, L.) xylan after exhaustive enzymatic treatment with alpha-arabinofuranosidase and beta-endo-xylanase**. *Carbohydr Res* 2014, **398**:63-71.
48. Pena MJ, Kulkarni AR, Backe J, Boyd M, O'Neill MA, York WS: **Structural diversity of xylans in the cell walls of monocots**. *Planta* 2016:1-18.
49. Kulkarni A, Pattathil S, Hahn MG, York WS, O'Neill MA: **Comparison of Arabinoxylan Structure in Bioenergy and Model Grasses** *Industrial Biotechnology* 2012, **8**(4):222-229.
50. Xiao C, Zhang T, Zheng Y, Cosgrove DJ, Anderson CT: **Xyloglucan deficiency disrupts microtubule stability and cellulose biosynthesis in Arabidopsis, altering cell growth and morphogenesis**. *Plant Physiol* 2015, **170**(1):234-249.
51. da Costa RMF, Lee SJ, Allison GG, Hazen SP, Winters A, Bosch M: **Genotype, development and tissue-derived variation of cell-wall properties in the lignocellulosic energy crop Miscanthus**. *Annals of Botany* 2014, **114**(6):1265–1277.
52. Verbruggen MA, Spronk BA, Schols HA, Beldman G, Voragen AGJ, Thomas JR, Kamerling JP, Vliegenthart JFG: **Structures of enzymically derived oligosaccharides from sorghum glucuronoarabinoxylan**. *Carbohydr Res* 1998, **306**(1):265-274.

877 53. Mazumder K, York WS: **Structural analysis of arabinoxylans isolated from ball-**
878 **milled switchgrass biomass**. *Carbohydr Res* 2010, **345**(15):2183-2193.
879 54. Charnock SJ, Spurway TD, Xie H, Beylot M-Hln, Virden R, Warren RAJ, Hazlewood
880 GP, Gilbert HJ: **The Topology of the Substrate Binding Clefts of Glycosyl**
881 **Hydrolase Family 10 Xylanases Are Not Conserved**. *J Biol Chem* 1998,
882 **273**(48):32187-32199.
883 55. Gilbert HJ, Hazlewood GP, Laurie JI, Orpin CG, Xue GP: **Homologous catalytic**
884 **domains in a rumen fungal xylanase: evidence for gene duplication and**
885 **prokaryotic origin**. *Molecular Microbiology* 1992, **6**(15):2065-2072.
886 56. Nagy T, Nurizzo D, Davies GJ, Biely P, Lakey JH, Bolam DN, Gilbert HJ: **The α-**
887 **Glucuronidase, GlcA67A, of Cellvibrio japonicus Utilizes the Carboxylate and**
888 **Methyl Groups of Aldobiouronic Acid as Important Substrate Recognition**
889 **Determinants**. *J Biol Chem* 2003, **278**(22):20286-20292.
890 57. Ryabova O, Vršanská M, Kaneko S, van Zyl WH, Biely P: **A novel family of**
891 **hemicellulolytic α-glucuronidase**. *FEBS Letters* 2009, **583**(9):1457-1462.
892 58. Peng W, Pedersen NR, Pettersson D, Ekloff JM, Nymand-Grarup S, Palmen LG,
893 Monrad RN, Spodsberg N, Stringer MA, Blom C *et al*: **Compositions comprising**
894 **polypeptides having xylanase activity and polypeptides having**
895 **arabinofuranosidase activity**. (2017), US 2017/0335302 A1
896 59. Margolles-Clark E, Tenkanen M, Nakari-Setälä T, Penttilä M: **Cloning of genes**
897 **encoding alpha-L-arabinofuranosidase and beta-xylosidase from Trichoderma**
898 **reesei by expression in Saccharomyces cerevisiae**. *Appl Environ Microbiol* 1996,
899 **62**(10):3840-3846.
900 60. Goubet F, Jackson P, Deery MJ, Dupree P: **Polysaccharide Analysis Using**
901 **Carbohydrate Gel Electrophoresis: A Method to Study Plant Cell Wall**
902 **Polysaccharides and Polysaccharide Hydrolases**. *Anal Biochem* 2002, **300**(1):53-68.
903 61. Tryfona T, Theys TE, Wagner T, Stott K, Keegstra K, Dupree P: **Characterisation of**
904 **FUT4 and FUT6 β-(1-->2)-Fucosyltransferases Reveals that Absence of Root**
905 **Arabinogalactan Fucosylation Increases Arabidopsis Root Growth Salt**
906 **Sensitivity**. *PLoS ONE* 2014, **9**(3):e93291.
907 62. Tryfona T, Stephens E: **Analysis of carbohydrates on proteins by offline Normal-**
908 **Phase Liquid Chromatography MALDI-ToF/ToF-MS/MS**. *Meth Mol Biol* 2010,
909 **658**(2):137-151.

63. Cavanagh J, Fairbrother WJ, Palmer AG, N.J S: **Protein NMR Spectroscopy: Principles and Practice.** . San Diego, CA, USA: Academic Press; 1996.

64. Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED: **The CCPN data model for NMR spectroscopy: Development of a software pipeline**. *Proteins: Structure, Function, and Bioinformatics* 2005, **59**(4):687-696.

65. Faure R, Courtin CM, Delcour JA, Dumon C, Faulds CB, Fincher GB, Fort S, Fry SC, Halila S, Kabel MA *et al*: **A Brief and Informationally Rich Naming System for Oligosaccharide Motifs of Heteroxylans Found in Plant Cell Walls**. *Australian Journal of Chemistry* 2009, **62**(6):533-537.

Figure captions

Figure 1. Overview of the technical procedure used to characterise GAX structures. Creation of the GAX oligosaccharide library from *Miscanthus sinensis* including extensive structural analysis (solid lines), generation of standard DASH profiles prior to SEC (dashed lines). White boxes indicate optional steps.

Figure 2. DASH profile of GAX oligosaccharides after hydrolysis with xylanases prior to SEC separation. GH10 (top panel), GH11 (middle panel), β-(1→4)-xylooligosaccharide standards X_1-X_6 (bottom panel). In addition, xylooligosaccharides XU^2XX, $XU^{(4Me)2}XX$ and XA^3XX as well as unknown peaks N1-N12 are labelled. Oligosaccharide migration is expressed in glucose units (GU). Asterisks (*) mark off-scale peaks.

Figure 3. DASH profile of SEC fractions from GH10 hydrolysis of miscanthus stem. 80 SEC fractions were separated and analysed by DASH (blue and grey traces), xylooligosaccharide standards X_1-X_6 (red trace). Area marked with yellow dashed box corresponds to background noise. Unknown peaks N3-N9 are labelled.

Figure 4. Characterisation of oligosaccharides in SEC fraction *Ms*10_55. (A) DASH capillary electropherograms showing four unknown oligosaccharides N_6, N_7, N_8 and N_9. (B) MALDI-ToF-MS of the 2-AA labelled oligosaccharides, showing a major sodiated and doubly sodiated ion corresponding to $Pent_5$ (*m/z* 822.0, *m/z* 844.0 respectively) and a minor ion corresponding to $Pent_6$ (*m/z* 954.0 and 976.0, respectively). (C) HILIC separation of structural isomers of *m/z* 822.0 $[M + Na]^+$ followed by off line-MALDI-ToF-MS, results in the EIC showing three structural isomers Z_1, Z_2 and Z_3, respectively.

Figure 5. High energy MALDI-CID spectra of oligosaccharides in SEC fraction *Ms*10_55. (A) Z_1 structure (N_6 in DASH) A^3A^3X. Inset: proposed chemical structure for W product ion [29]; (B) Z_2 structure (N_7 in DASH) A^3XXX; (C) Z_3 structure (N_8 in DASH) $D^{2,3}XX$. (D) NMR analysis of the Z_3 structure. H-1 strip plots from 2D ^1H-^1H TOCSY (blue) and ROESY (red) spectra showing the NOE connectivity arising from the β-Xyl*p*-(1→2)-α-Ara*f*-(1→3)-β-Xyl*p*-(1→4)-β-Xyl*p* glycosidic linkages.

Figure 6. Comparison of DASH profiles of GAX oligosaccharides from stem of various grasses. (A) GH10 hydrolysis products (B) GH11 hydrolysis products. Respective Miscanthus

957 oligosaccharide library (bottom panel). Note the presence of two unknown peaks M_1 and M_2
958 in the rice DASH profile.
959
960 **Figure 7.** Comparison of DASH profiles of GAX oligosaccharides from leaves of various
961 grasses. (A) GH10 hydrolysis products (B) GH11 hydrolysis products. Respective Miscanthus
962 oligosaccharide library (bottom panel). Note the abundance of the two unknown peaks M_1 and
963 M_2 in profile of leave material.
964
965
966 **Figure 8.** Quantification of substitution frequency of miscanthus stem and leaf GAX. All
967 characterised GH10 and GH11 hydrolysis products were analysed by DASH using the
968 DASHboard software for quantification. The substitution frequency was calculated after
969 normalisation of values by comparing the abundance side chain to the Xyl*p* residues in the
970 backbone. Error bars represent mean ± SD (n=3). A significant difference was observed in the
971 $D^{2,3}$ substitution frequency of the GH10 digest ($p = 0.002$) and of the GH11 digest ($p = 0.2$)
972 between stems and leaves and between the GH10 and G11 digest in leaves only ($p = 0.01$)
973 using the paired t-test for two-tailed distribution, marked with asterisk.
974