

# New mechanistic insight into the digestion of complex dietary fibre by rumen microbiota using combinatorial high-resolution glycomic and transcriptomic analyses

**Ajay Badhan**

Agriculture and Agri-Food Canada

**Kristin Low**

Agriculture et Agroalimentaire Canada Centre de recherche et de développement de Lethbridge:  
Agriculture and Agri-Food Canada Lethbridge Research and Development Centre

**Darryl Jones**

Agriculture and Agri-Food Canada

**Xiaohui Xing**

Agriculture and Agri-Food Canada

**Mohammad Raza Marami Milani**

Agriculture and Agri-Food Canada

**Rodrigo OrtegoPolo**

Agriculture and Agri-Food Canada

**D. Wade Abbott**

Agriculture and Agri-Food Canada

**Sivasankari Venketachalam**

Oak Ridge National Laboratory

**Michael Hahn**

University of Georgia

**Tim McAllister** (✉ [tim.mcallister@canada.ca](mailto:tim.mcallister@canada.ca))

Agriculture and Agri-Food Canada <https://orcid.org/0000-0002-8266-6513>

---

## Research

**Keywords:** glycoside hydrolase, CAZymes, rumen microbiome, glycome profiling, linkage analysis, transcriptome, differential gene expression, nutrient utilization, feed recalcitrance, dietary polysaccharides

**Posted Date:** November 13th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-104578/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Computational and Structural Biotechnology Journal on December 1st, 2021. See the published version at <https://doi.org/10.1016/j.csbj.2021.12.009>.

## Abstract

## Background

The rumen microbial community is considered the most efficient anaerobic digestive ecosystem known, yet less than half of the energy in low quality forages is actually metabolized. There is a knowledge gap regarding the specific factors that impede the ruminal digestion of plant cell walls or if rumen microbiota have the functional potential and activities to overcome these constraints. To address these issues, innovative experimental methods may provide a high-resolution understanding of the cell wall chemistries and higher-order structures that are resistant to microbial digestion and how they interact with the functional activities of the rumen microbial community.

## Results

With this goal, we characterized the total tract indigestible residue (TTIR) from cattle fed a high-forage diet containing low-quality straw using two comparative glycomic approaches: ELISA-based glycome profiling and glycosidic linkage analysis. Using these techniques, we successfully detected numerous and diverse cell wall glycan epitopes in barley straw and TTIR and determined their relative abundance pre- and post-intestinal digestion. Of these, xyloglucans and heteroxylans were the most recalcitrant to digestion. Linkage analysis identified indigestible linkages consistent with the polysaccharide epitopes identified by ELISA-based glycome analysis. To determine if residual plant polysaccharides within TTIR could be metabolised, rumen microbiota from cannulated cattle fed barley straw were incubated with barley straw and TTIR in *in vitro* batch cultures. Transcript coding for carbohydrate-active enzymes (CAZymes) were identified and characterized for their contribution to cell wall digestion based on glycomic analyses, comparative gene expression profiles, and associated CAZyme families. High-resolution phylogenetic fingerprinting of these sequences revealed encoded enzymes with activities predicted to cleave the primary linkages within heteroxylan and arabinan.

## Conclusion

This experimental platform provides unprecedented precision in the understanding of forage structure and digestibility, which can inform next-generation solutions to improve the growth of ruminants fed low quality forages and enhance the use of crop residues as a feedstock.

## Background

Lignocellulosic biomass can be converted into a broad suite of fuels, chemicals, and biomaterials. The recalcitrance of cellulosic material and high cost of hydrolysis to simple sugars is a major barrier to the commercial viability of bioproducts developed from cellulosic biomass. With an ever-growing human population and more affluent societies, the global demand for food, meat and milk is projected to increase

exponentially [1, 2]. Ruminant livestock are in a unique position to satisfy the growing demand for high quality protein, as ruminants can produce milk and meat via the microbial fermentation of cellulose-rich forages, crop residues, and food by-products. In this light, the rumen microbiota represents an underexploited repository of carbohydrate active enzymes (CAZymes) and microorganisms for applications in animal nutrition, biofuel, and bio-based chemical industries.

Globally, more than 73.9 million metric tons of crop residues are produced annually [3, 4]. Although these residues could serve as feed for ruminants, usually less than half of the biomass they contain is digested. Supplementing ruminant diets with exogenous enzymes has the potential to improve the utilization of crop residues for meat and milk production [5, 6, 7, 8]. However, current commercial enzyme products have been largely developed for biorefinery and bioprocessing applications and were not intended to function in the gastrointestinal tract of ruminants. Physiological conditions within the rumen are not favorable for some enzymes within cellulase mixtures [7, 8], with many of these products possessing enzyme activities that are redundant to those already produced by the rumen microbial community. Quantifying the digestibility of neutral detergent fiber (NDF) and acid detergent fiber (ADF) in ruminants provides only cursory insights into the factors that limit plant cell wall digestion in the rumen. Immunofluorescence microscopy coupled with cell wall specific monoclonal antibodies (mAbs) has proven to be a more informative method to investigate the polysaccharide composition of plant cell walls [9, 10], yet an informed and more tailored approach is needed to investigate ruminal fiber digestion. Such an approach should be anchored in a higher understanding of cell wall recalcitrance to ruminal enzymatic digestion and identification of rate limiting enzyme activities. Once rate-limiting enzymatic activities within the rumen are identified, CAZymes with synergistic activities could be used to enhance the digestion of fiber by rumen microbiota.

Here, we have developed a platform for coupling glycomic analyses to comparative metatranscriptomes of rumen microbiota in batch cultures with barley straw and total tract indigestible residue (TTIR) as substrates. We aimed to characterize recalcitrant cell wall structures and identify candidate enzyme activities with the potential to enhance the rumen digestion of barley straw. Using this approach, we identified indigestible cell wall moieties within barley straw, and characterized the linkages that contributed to recalcitrance. A suite of target genes was selected for downstream production of recombinant enzymes that could potentially overcome the constraints to the digestion of barley straw plant cell walls. Although this study focused on the digestion of barley straw, the experimental pipeline holds high promise for the improvement of lignocellulosic feedstock from diverse plant sources in support of the sustainable conversion of lignocellulose biomass into industrial feedstock.

## Results

# Glycome profiling of barley straw and total tract indigestible residue

The sequential extraction of cell wall material leads to selective enrichment of various cell wall glycans into fractions based on the degree of integration into the plant cell wall [11]. Oxalate (AO) and carbonate

(SC) fractions contain loosely bound pectic and arabinogalactan polysaccharides, while extraction with 1 M and 4 M KOH solubilizes hemicelluloses (e.g., xyloglucans and xylans) and tightly bound pectic polysaccharides. Chlorite (CH) treatment releases lignin and lignin-associated polysaccharides, while a subsequent 4 M KOH extraction after delignification recovers remaining recalcitrant hemicellulose and pectin fractions embedded within crystalline cellulose. Chemical fractionation of barley straw and TTIR followed by enzyme-linked immunosorbent assay (ELISA) identified differences in the glycome profile of these substrates (Fig. 1). One major difference between barley straw and TTIR is the amount of carbohydrate recovered in the oxalate (AO) extracts of these two materials; about two and half times as much carbohydrate was recovered in the AO extract from TTIR as compared with barley straw. Yet, with the exception of the de-esterified homogalacturonan epitopes, which were enriched in the TTIR AO extract in comparison with the AO of barley straw (Fig. 1; box 1), there was very little antibody reactive material in the AO extract of TTIR. This suggests that most of the carbohydrate extracted by AO is too small to bind to the ELISA plate, or too small to include the full epitopes recognized by the antibodies used in this study. In general, the AO and SC fractions of TTIR had minimal or lacked most xyloglucan (XG), xylan, type I rhamnogalacturonan (RG-I), and arabinogalactan (AG) moieties, suggesting that these carbohydrates were digested during passage through the ruminant digestive tract. In addition, there was a notable reduction in the AG-3 and AG-4 epitopes and in the lignin-associated MeGlcA-substituted xylans (Fig. 1, Box 7) in TTIR compared with barley straw. In contrast, the somewhat loosely associated MeGlcA-substituted xylans present in the SC extract appear resistant to degradation during passage through the bovine digestive tract (Fig. 1, Box 2).

In the remainder of the glycome profiles, there appear to be selective enrichments in the TTIR for particular epitopes in comparison with barley straw. For example, there was a relative increase in the abundance of Gal-XG epitopes in the 1M and 4 M KOH extracts of TTIR compared with barley straw (Fig. 1, Boxes 3, 4), as well an increased enrichment of the Xylan-3 and 2-Ara-substituted xylans (Fig. 1, Boxes 5, 6). With diverse 6-linked  $\beta$ -galactan epitopes (common in RG-I and AGPs), there appears to be an enrichment in these epitopes in the 4M KOH extract (Fig. 1, Box 9), but a decreased presence of these epitopes in the 1 M KOH extract (Fig. 1, Box 8) when compared with barley straw. There also appears to be enrichment in some RG-I epitopes in the 4 M KOH extract from TTIR (Fig. 1, Box 10) compared with the corresponding extract from barley straw. Lastly, there is an overall enrichment in the TTIR biomass for the most tightly bound glycans released in the 4 M KOH post-chlorite extraction (PC) compared with barley straw. It is important to note that CH and PC represent the most recalcitrant cell wall fractions. The persistent presence of cross-linked (glucurono) arabinoxylan, xyloglucan, and cellulose-embedded pectins (galactans, RG-I, and AG-2) in the KOH, CH and PC extracts of TTIR indicates their recalcitrance to microbial digestion within the ruminant digestive tract.

## **Glycosidic linkage analysis of barley straw and total tract indigestible residue**

To further define the limits of plant cell wall digestibility by rumen microbiota, we undertook a glycosidic linkage analyses of barley straw and TTIR. The relative abundance of individual monosaccharide linkages

in these fraction was determined (Additional File 1) and a ratio of linkage abundance in the barley straw control over the TTIR representing the overall digestibility was calculated (Fig. 2A). A number of linkages within the pectin-rich fraction (EDTA + Na<sub>2</sub>CO<sub>3</sub>); 2,4-Xylp, t-GalAp, 4-GalAp, 3,6-Galp, 3-Araf, 5-Araf, 2,3,5-Araf, 4-Manp, and 3-Glcp linkages were more readily digested by rumen microbiota. Linkages from the hemicellulose fraction (4 M KOH), t-Araf, t-Xylp, 2-Xylp, 4-Xylp, 2,4-Xylp, 3,4-Xylp, t-GalAp, 4-GalAp, 4-Manp, and 3-Glcp, were freely digested. As expected, the final recalcitrant cellulosic fraction exhibited the lowest level of digestibility for most linkages in all fractions (Fig. 2A).

In order to evaluate the recalcitrance of barley straw to digestion by the rumen microbiota and identify rate-limiting enzymatic activities, the ten most abundant indigestible linkages were calculated as a ratio of abundance in TTIR over that of barley straw, and ranked highest to lowest (Fig. 2B). In the AIR fraction, t-Xylp was found to be highly abundant ( $1.6 \pm 0.6$ ), whereas 3,4-Xylp ( $5.4 \pm 0.4$ ), t-Manp ( $2.4 \pm 0.7$ ), and 2-Araf ( $2.6 \pm 0.9$ ) were the most indigestible linkages identified within the EDTA + Na<sub>2</sub>CO<sub>3</sub>, 4 M KOH, and cellulosic residue fractions, respectively.

Polysaccharide composition was estimated by assigning the linkage composition data to classes of polysaccharides according to Pettolino *et al.* [12] (Fig. 2C, Table 1, Additional File 1b). Differences in polysaccharide composition were observed between barley straw and TTIR samples. TTIR contained more arabinan (AB), HX and XG than barley straw in AIR and the final cellulosic residual fractions. The EDTA + Na<sub>2</sub>CO<sub>3</sub> fraction of TTIR showed higher HX and XG content. AG-2 content within the TTIR 4 M KOH fraction was observed to be higher as compared to barley straw.

Table 1  
Prediction of polysaccharide composition (%) based on glycosidic linkage data.

Polysaccharides <sup>1</sup>	AIR		EDTA + Na <sub>2</sub> CO <sub>3</sub>		4M KOH		Residue	
	Barley	TTIR	Barley	TTIR	Barley	TTIR	Barley	TTIR
Arabinan	1.9	2.1	6.3	5.3	2.1	2.5	0.2	0.4
Heteroxylan (HX)	30.8	33.8	17.1	32.4	79.6	76.2	4.7	7.9
Type I arabinogalactan (AG-I)	1.2	1.5	2.9	3.7	0.3	0.4	t.r.	0.1
Type II arabinogalactan (AG-II)	2.2	2.2	12.1	11.6	2.9	4.3	0.3	0.6
Xyloglucan (XG)	5.1	5.5	5.5	8.2	9.0	9.8	2.3	3.0
Mixed linkage glucan (MLG)	3.3	1.9	23.7	21.0	4.2	4.8	1.2	1.4
Cellulose	49.7	47.2	n.a.	n.a.	n.a.	n.a.	84.4	78.3
Homogalacturonan (HG)	0.2	0.3	7.0	2.4	0.4	0.3	t.r.	t.r.
Heteromannan (HM)	0.9	0.8	9.6	5.7	0.5	0.5	0.5	0.8
Unassigned (UA)	4.8	4.7	15.8	9.8	1.1	1.1	6.1	7.4
Note: "t.r." means trace amount (mol%<0.1%). "n.a." means linkage not assigned to polysaccharide. No 4-Glcp linkage was assigned to cellulose as EDTA + Na <sub>2</sub> CO <sub>3</sub> , and 4 M KOH solutions do not extract this fraction [60].								
<sup>1</sup> The estimation of polysaccharide was according to Pettolino <i>et al.</i> 2012 [12].								

## RNA-seq output and *de novo* assembly

RNA-seq and differential gene expression analyses were conducted to study the composition of rumen microbial communities *in vitro*, and to compare microbial CAZyme expression profiles of rumen microbes cultured on TTIR or barley straw. Rumen microbial samples were batch-cultured in triplicate *in vitro* for each substrate (TTIR and barley straw), as reported previously [13], and transcriptomics analyses were performed. A total of 364,583,860 reads were used for quantitative RNA-Seq analysis. The number of reads per sample ranged from 56.95 million to 66.20 million. The transcriptome was assembled *de novo* using the Trinity assembler [14] and led to 1,998,343 distinct transcripts with a median transcript length of 336 (bp) and N50 of 533 (bp) (Additional File 2).

### Microbial taxonomic classification based on putative mRNA.

RNA-seq raw reads were submitted as input to the Kaiju [15] webserver for taxonomical assignment at the whole transcriptome level. The results of the taxonomic classification with Kaiju were further processed and visualized with Krona [16]. Bacteria represented the majority (88%) of rumen transcripts (Table 2). Phylum *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, *Fibrobacteres*, *Proteobacteria*, and *Spirochaetae* were at a relative abundance above 1%. At the family level, *Lachnospiraceae* (9%) *Ruminococcaceae* (7%), *Spirochaetaceae* (5%), *Prevotellaceae* (4%), *Clostridiaceae* (4%), *Fibrobacteraceae* (2%) were most abundant. The transcripts associated with Eukaryotes accounted for 8% of the total transcripts with *Ascomycota*, *Basidiomycota*, *Neocallimastigomycota*, and *Intramacronucleata* being major phyla (Table 2).

Table 2  
Taxonomic analyses of microbiota transcriptome

Domain	Phylum	Family	
Bacteria (88%)	<i>Firmicutes</i> (41%)	<i>Lachnospiraceae</i> (9%)	
		<i>Ruminococcaceae</i> (7%)	
		<i>Clostridiaceae</i> (4%)	
		<i>Unclassified Clostridiales</i> (4%)	
		<i>Paenibacillaceae</i> (1%)	
	<i>Bacteroidetes</i> (17%)	<i>Prevotellaceae</i> (4%)	
		<i>Bacteroidaceae</i> (2%)	
		<i>Unclassified Bacteroidales</i> (2%)	
		<i>Rikenellaceae</i> (1%)	
		<i>Porphyromonadaceae</i> (1%)	
	<i>Actinobacteria</i> (9%)	<i>Streptomycetaceae</i> (3%)	
		<i>Fibrobacteres</i> (2%)	<i>Fibrobacteraceae</i> (2%)
		<i>Proteobacteria</i> (12%)	<i>Pseudomonadaceae</i> (0.6%)
<i>Spirochaetae</i> (5%)	<i>Spirochaetaceae</i> (5%)		
	<i>Lentisphaerae</i> (0.9%)	<i>Unclassified Lentisphaerae</i> (0.9%)	
	Eukaryota (8%)	<i>Opisthokonta</i> (64%)	<i>Ascomycota</i> (32%)
<i>Neocallimastigomycota</i> (6%)			
<i>Basidiomycota</i> (18%)			
<i>Alveolata</i> (9%)			<i>Intramacronucleata</i> (2%)

Note: Taxonomical affiliations of whole transcriptomic data was obtained from taxonomic analyses of mRNA by Kaiju [15]. Relative abundance (%) at the domain level were calculated as a percentage of total mRNA transcripts, whereas relative abundance at the phyla and family level are calculated as a percentage of mRNA transcripts within the affiliated domain. Only those phyla and families with relative abundance higher than 0.6% are reported.

## CAZyme expression under *in vitro* batch culture

To investigate the functional activities of these rumen samples, the RNA-seq raw reads for both barley straw and TTIR cultures were assembled into a total of 1,993,243 transcripts with an average length of 492 bp (Additional File 2) and parsed through BLASTx and the CAZy database (v2017-07-20) [17]. Several transcripts were annotated as CAZymes in the following families: 126 glycoside hydrolase (GH), 27 polysaccharide lyase (PL), and 16 carbohydrate esterase (CE) (Additional File 3). The overall transcriptome contained all the major GH families putatively involved in cellulose and hemicellulose digestion.

Substrates for individual GH, PL, and CE families were assigned according to Couger, *et al.* [18]. A total of 13 GH families possibly involved in cellulose degradation were identified, collectively accounting for ~ 18% of total CAZymes (Fig. 3A, Additional File 4). Major cellulose-targeting GH families included GH5, GH9, GH45, GH48, and GH124 endoglucanases; GH6 and GH48 cellobiohydrolases; and GH1 and GH3  $\beta$ -glucosidases. Similarly, transcripts targeting hemicellulose degradation included GH10, GH11, GH43, GH5, GH26, and GH8; debranching enzymes belonging to GH2, GH3, GH43, GH51, GH95 and GH43 and acetyl xylan esterase of CE1 family. Out of 39 hemicellulose targeting GH families, only 12 had a relative abundance greater than 1% (accounting for ~ 34% of CAZyme expression, Fig. 3B). Twelve GH, 6 PL, and 3 CE putatively involved in pectin degradation were identified, including polygalacturonases (GH28), rhamnosidases (GH78), galactanases (GH53), pectin lyases (PL1) and pectin acetyl esterases (CE12) (Fig. 3C). In addition, transcripts encoding enzymes that degrade starch (pertaining to GH13, GH31, and GH77 families) were also observed (Fig. 3D, Additional File 4).

Transcripts predicted to be involved in cellulose, hemicellulose, pectin and starch digestion (GH + PL + CE) were taxonomically classified to gain insight into the predominant species involved in cell wall degradation. The majority of the cellulose-targeting transcripts were associated with the Firmicutes (42%; *Ruminococcus*, 9%, *Lachnospiraceae*, 5%), Bacteroidetes (23%; *Prevotella*, 7%, *Bacteroides*, 3%), and Fibrobacteres (4%; *Fibrobacter* 3%) (Fig. 4A, Additional File 5). Fungi contributed ~ 4% of transcripts that were predicted to target cellulose digestion (Fig. 4A). Similarly, Firmicutes contributed ~ 38% (*Clostridium* 6%, *Ruminococcus* 5%) and *Bacteroidetes* 26% (*Prevotella* 7%) to the pool of transcripts targeting hemicellulose (Fig. 4, Additional File 5). Likewise, transcripts for families putatively involved in dismantling pectin were extensively assigned to bacteria and primarily to Firmicutes (39%; *Ruminococcus*, 5%) and Bacteroidetes (24%; *Prevotella*, 6%). In contrast, starch-degrading transcripts mostly originated from Bacteroidetes (34%; with *Prevotella* contributing 11% of total starch digesting transcripts). In total, the majority of Firmicutes and Bacteroidetes-assigned CAZyme transcripts targeted hemicellulose and cellulose digestion (Fig. 4B), with Bacteroidetes associated with the transcripts targeting starch digestion. Out of the total *Fibrobacter* CAZyme transcripts, 59%, 30%, and 11% putatively targeted hemicellulose, cellulose, and pectin, respectively (Fig. 4B).

## Gene expression at CAZyme family level

Metatranscriptomic data was used to evaluate the contribution of specific CAZyme families to cell wall degradation. Thus, for those CAZyme family transcripts known to mediate the deconstruction of major cell wall constituents, total expression levels of transcripts for the individual CAZyme family was calculated as a percentage of total expression levels of all CAZyme transcripts (Additional File 6). Endoglucanase and  $\beta$ -glucosidase (GH5, GH8, and GH9) transcripts were higher ( $p > 0.001$ ) in TTIR cultures (Additional File 6A), while barley straw cultures had a higher abundance of cellobiohydrolase (GH6, GH48) transcripts (Additional File 6B). Interestingly, both GH6 and GH48 families contributed towards the cellobiohydrolase pool for barley straw degradation, whereas GH48 was the only major family of cellobiohydrolases associated with TTIR. Similar transcript abundance was seen for xylan degradation families (GH10, GH11, GH43, and GH39) for TTIR and barley straw (Additional File 6D). Putative pectin-degrading GH families showed higher expression on barley straw with GH78 as the major CAZyme family, whereas GH28 and GH53 were the major pectin-associated GH families expressed with TTIR (Additional File 6F).

## Differential gene expression at individual transcript level

Identifying transcripts that catalyze the cleavage of carbohydrate linkages that limit ruminal digestion could provide insight into improving the digestibility of barley straw. Therefore, we focused on differentially expressed transcripts between barley straw and TTIR that are classified as CAZymes. Differential expression analyses identified a total of 88 glycoside hydrolases (GH) that were upregulated when rumen microbiota were cultured with TTIR, while in the presence of barley straw, 130 GH transcripts were ( $p < 0.01$ ) upregulated (Table 3, Additional File 7). In TTIR samples, 7 PL and 9 CE transcripts were upregulated ( $p < 0.01$ ), whereas 9 PL transcripts and 21 CE transcripts were upregulated ( $p < 0.01$ ) with barley straw. At the family level (the total collection of differentially expressed transcripts within a given CAZy family), higher expression was observed in the presence of TTIR of CAZy families known to be putatively involved in degradation of cellulose (GH5), xylan (CE15, GH67), pectin (GH105, PL11) and mannan (GH26), but were not found to be statistically significant due to the high variation in expression levels across replicate batch cultures (Fig. 5, Additional File 7).

Table 3  
Total number of upregulated CAZyme transcripts on barley straw and TTIR substrates.

Upregulated Transcript Enzyme Class	Number of transcripts	
	BS	TTIR
Glycoside hydrolase (GH)		
( $p < 0.01$ )	130	88
( $p < 0.1$ )	26	3
Polysaccharide lyase (PL)		
( $p < 0.01$ )	5	7
( $p < 0.1$ )	0	1
Carbohydrate esterase (CE)		
( $p < 0.01$ )	21	9
( $p < 0.1$ )	6	3

To increase the resolution of functional annotation, the nucleotide sequences for differentially expressed (DE) transcripts were re-submitted to the recent version of the dbCAN2 server [19] and manually curated for CAZyme annotation. To distinguish dominant individual DE transcripts, the transcriptional expression pattern for DE transcripts between TTIR and barley straw cultures was studied for CAZy families of interest known to target XG, HX, and pectin (Fig. 6). Transcript TR363754|c1\_g3 (GH11), TR744494|c3\_g1 (GH16), and TR463944|c0\_g2 (CE15) contributed significantly towards TTIR overexpressed DE transcripts as their relative abundance was lower in the presence of barley straw. Whereas other transcripts, such as TR400632|c0\_g1 (GH11), TR849235|c0\_g1 (GH11), and TR300921|c2\_g2 (GH43), were observed to be upregulated in barley straw cultures as compared to TTIR cultures.

## Prediction of enzyme activities for rumen DE transcript CAZymes

The differentially expressed transcripts represent uncharacterized rumen microbial CAZyme sequences. As HX and AB were identified as key recalcitrant components by glycome and linkage analyses, statistically significant DE transcripts from relevant polyspecific families GH11, GH43, and CE15 were selected for high-resolution functional prediction with SACCHARIS [20]. Phylogenetic trees generated for biochemically characterized members of GH 11 (Fig. 7A) indicated that TTIR overexpressed transcripts (TR363754|c1\_g3,

TR668965|c0\_g1, TR196282|c0\_g1, TR190685|c0\_g1, and TR83509|c0\_g1) partitioned into distantly-related clades, as compared to those overexpressed in barley straw (TR476599|c0\_g1, TR849235|c0\_g1, TR327227|c0\_g1, TR291133|c0\_g1, TR400632|c0\_g1, and TR1014827|c1\_g1). All GH11 transcripts are predicted to function as *endo*- $\beta$ -1,4-xylanases, the dominant enzymatic activity for the family. Similarly, in GH43 (Fig. 7B), TTIR-overexpressed transcript TR712918|c0\_g1 was classified as an *endo*- $\alpha$ -1,5-L-arabinanase, while barley straw-overexpressed transcripts TR400300|c0\_g2 and TR451123|c0\_g1 were predicted to be putative arabinofuranosidases. Interestingly, TR941931|c2\_g1 and TR300921|c2\_g2 were both overexpressed in barley straw and were predicted to be *endo*- $\alpha$ -1,5-L-arabinanases. All three TTIR-overexpressed CE15 transcripts were classified as acetyl xylan esterases (Fig. 7C).

## Discussion

Developing efficient technologies to enhance the conversion of low-quality forages into available energy within the rumen is pivotal for sustainable dairy and beef production [21]. Potentially, such technologies can reduce feed costs and improve the competitiveness of cattle management, while reducing the environmental footprint of meat and milk production [22]. Supplementing ruminant diets with exogenous enzymes has the potential to enhance feed efficiency [7, 8]. However, current commercial enzymes have been tailored for bio-refinery and bioprocessing applications and as such are not optimized for the physiological conditions of the rumen [23, 24, 25]. Likewise, commercial enzymes often contain redundant enzyme activities already encoded within the rumen microbiota. Understanding the rate-limiting steps in ruminal fiber digestion from the perspective of plant cell wall chemistry and endogenous catalytic potential is critical for the successful design of technologies based on enzyme addition, programming of the microbiome, and breeding plant cell walls that are more amenable to digestion by microorganisms within the ruminant digestive tract. Here we present an informed experimental platform (Fig. 8) that couples high-resolution cell wall structural analysis with targeted meta-transcriptomic analysis of rumen microbiota so as to expand the present understanding of the factors that limit plant cell wall degradation.

Straw residues that escape ruminal total tract digestion (TTIR) can provide valuable insight into nutrient utilization by rumen and lower-tract microbes and into plant cell wall components that confer recalcitrance to the hydrolysis of straw. Presented here is the first report using ELISA-based glycome profiling to study the progression of fiber utilization and recalcitrance during total tract digestion of fiber in ruminants. This methodology has been extensively used to characterize plant cell walls including those of genetically modified plants, and microbially and chemically treated plant biomass [9, 10, 26]. Glycome profiling provided us with an unprecedented ability to decipher the architecture of native barley straw and its residual fraction that is recalcitrant to digestion in cattle (TTIR) as it can detect diverse structural features within plant cell wall polysaccharide classes (Fig. 1). For example, a comprehensive suite of more than 25 mAbs that recognize different substructures, including each of the major side-chains of xyloglucans was applied, while another set of approximately 30 mAbs that recognize diverse xylan epitopes including Ara, GlcA and MeGlcA substitutions as well as various lengths of backbone residues [27] were used to assay xylans (Fig. 1). This not only allowed us to identify recalcitrant polymers with high precision, but also to differentiate polysaccharide subclasses that were removed during passage through the ruminant digestive

tract or that exhibit high recalcitrance and were enriched in the TTIR. For instance, 4-O-Me-GlcA xylan epitopes that are lignin bound, as well as those xylans loosely associated with the cell walls in barley straw were removed during passage through the digestive tract (Fig. 1). Alternatively, 2-Ara-substituted xylans, as well as xylan epitopes recognized by the Xylan-2 and Xylan-3 groups of antibodies, were enriched in TTIR, suggesting that these epitopes directly contribute to recalcitrance. Relative proportions of xyloglucan and xylan epitopes in TTIR remained largely unchanged from the proportions of these epitopes in native barley straw (Fig. 1). However, it was clear that glycome profiling alone could not fully describe the glycan composition of TTIR. For example, there was a significant fraction of glycans that were loosely associated with the TTIR (AO extract), yet, with the exception of de-esterified HG epitopes (DP > 4), there was no antibody binding to this extracted fraction from TTIR. This lack of antibody binding may be attributed to the fact that short-chain oligosaccharides released during digestion fail to bind to ELISA plates [10]. Alternatively, while the antibody collection used here recognizes a large and diverse set of plant cell wall glycan epitopes, there may be other epitopes native to barley straw or generated during passage through the ruminant digestive tract that are not recognized by the current antibody library. Therefore, to confirm glycome profiling results we quantified individual glycosidic linkages using partially methylated acetyl alditol derivatization, which measures all glycan linkages present. The linkage data showed that the TTIR was enriched in HX, XG, and AB when compared with barley straw at the whole cell wall level (Fig. 2, Table 1), suggesting that these components were more recalcitrant than other components within this biomass. Fractionation of the walls showed that the EDTA + carbonate extract was strongly enriched in HX and XG (Fig. 2, Table 1), which was in contrast to the results from glycome profiling, where the equivalent extract (AO) only showed enhanced HG content (Fig. 1). This may reflect the composition of the large fraction of cell wall glycans that are loosely associated with TTIR, but which were not assayable in the glycome profiling ELISAs for the reasons mentioned above. Consistent abundance of HX and XG within the whole cell wall and the EDTA/carbonate and KOH extracts of TTIR suggest that these epitopes contribute to the recalcitrance of barley straw within the ruminant digestive tract. Apart from cell wall composition, linkage analysis also provided useful insight into the predominant bonds within identified recalcitrant polysaccharides. The results presented in this study could prove to be a valuable resource for crop breeders to develop feed stocks that exhibit increased digestibility in ruminants.

Identification of predominant recalcitrant bonds coupled to transcriptomic and differential expression analyses, provided two complementary approaches to identify candidate linkages to target for improved plant cell wall digestion. To identify potential enzymes tailored for the digestion of recalcitrant linkages, TTIR and barley straw were used as substrates to culture rumen microbial communities in batch cultures. Incubation with TTIR provided a unique experimental approach that enabled the rumen microbial digestion of recalcitrant plant cell walls to be studied. This is the first such attempt where recalcitrant residues like TTIR have been used to study rumen community composition and CAZyme expression. Nutritional value and effective fermentation of TTIR was confirmed by ELISA results and through comparison of total gas production using barley straw and TTIR as substrates in *in vitro* cultures (Additional File 8). Firmicutes (*Lachnospiraceae*, *Ruminococcaceae*, *Clostridiaceae*), Bacteroidetes (*Prevotellaceae*, *Bacteroidaceae* and unclassified *Bacteroidiales*), Actinobacteria, Fibrobacteres, Proteobacteria and Spirochaetae were found to be dominant bacterial phyla at the transcriptome level

(Table 2). These results are in agreement with previously reported metatranscriptomic analysis, such as Dai et al. [28], that reported *Firmicutes*, *Bacteroidetes*, *Spirochaetes*, *Proteobacteria*, *Actinobacteria* and *Fibrobacteres* are the most abundant bacterial phyla. Likewise, Comtet-Marre et al. [29] found that bacteria represented 77.5% of rumen ssu rRNA reads with *Firmicutes*, *Bacteroidetes*, *Fibrobacteres*, *Proteobacteria*, *Spirochaetae* and *Lentisphaerae* as predominant phyla. Although the present study found a similar core microbiome, there was variation in the abundance of taxonomic families as compared to previous literature [28, 29]. This variation may be attributed to the shorter incubation time (48 h), and absence of host effects in closed *in vitro* batch culture system as compared to *in vivo* animal-based studies. Shorter incubation times favor rapidly growing bacteria over the slower growing fungal and protozoal members of ruminal microbial communities. In contrast to the host environment where fermentation by-products like VFA and ammonia are absorbed across the intestinal epithelium, closed batch culture systems accumulate by-products, which can alter the fermentation process towards certain types of microbial communities. Furthermore, host genetics have also been reported to influence microbial populations [30], as a region on chromosome 6 in cattle has recently been shown to be associated with *Actinobacteria*, *Euryarchaeota* and *Fibrobacteres* densities [31]. While whole animal experiments can not specifically study the digestion of TTIR, *in vitro* batch cultures provided us with the level of control over experimental variables required to investigate those factors that limit plant cell wall digestion.

Gene expression analyses of rumen microbial communities cultured on barley straw and TTIR demonstrated substrate-specific modes of digestion, as evidenced by the expression levels of CAZy family transcripts. At the family level, higher expression of putative *endo*-glucanases GH5, GH9, and GH8 on TTIR indicated an abundance of metabolizable cellulose within TTIR, as confirmed in the linkage analyses. Furthermore, GH48 cellobiohydrolases were the dominant *exo*-glucanases expressed during the digestion of TTIR by rumen microbes. In contrast, GH6 and GH48 were the predominant families associated with the microbial digestion of barley straw (Additional File 6B). Both GH6 and GH48 are known to efficiently hydrolyze the crystalline regions of cellulose [28, 32]; while cellobiohydrolases of family GH48 act on the reducing ends of cellulose chains, and GH6 attack non-reducing ends, generating cellobiose [33]. The high expression levels of these *exo*-glucanases suggest there is a greater abundance of amorphous cellulose at the reducing ends of micro-fibrils in TTIR as compared to barley straw, as result of its exposure to CAZymes during passage through the ruminant digestive tract. The level of expression of GH43 hemicellulases in both TTIR and barley straw was higher than other hemicellulose-targeting activities (Additional File 6E). Among the 501 bacteria in the Hungate catalogue, the GH43 family was reported as the most abundant hemicellulase [34]. Metagenomic [35–37] and metatranscriptomic [28, 29, 38–40] studies have suggested that members of the GH43 family are the principal hemicellulases within the rumen. GH43 encode a range of debranching enzymes, including arabinofuranosidases, xylanases, galactanases, arabinanases and  $\beta$ -xylosidases, which aid in the degradation of arabinoxylan [41]. Indeed, glycomics revealed that arabinoxylan was enriched in TTIR, highlighting its resistance to ruminal digestion. Additionally, abundance of branched linkages like 3,4-Xylp, t-GlcAp, 2,3,4-Xylp, t-Araf in TTIR also confirm high resistance of heteroxylan to degradation by rumen microbiota (Fig. 2). Metatranscriptomic analyses reflected the efforts of the rumen microbiota to degrade this polysaccharide, with the GH10 family being the dominant family of xylanases (Additional File 6D). This family has been previously

reported to digest heteroxylan and release xylotriose with a methylglucuronic acid (MeGA) moiety at the non reducing end which is hydrolyzed by members of the GH5 family of xylanases [42]. The GH5 xylanases have been shown to be active against glucuronoxylan and MeGA substituents on the xylan backbone [42]. Pectin polysaccharides differed between barley straw and TTIR (Figs. 1 and 2). Accordingly, expression levels of pectin-targeting CAZy families differed between rumen microbiota incubated with TTIR vs barley straw (Additional File 6F). Extracted rumen microbial communities cultured on barley straw expressed more GH78s, a family that contains  $\alpha$ -l-rhamnosidases that act on rhamnogalacturonan and arabinogalactan-protein linkages [43]. In contrast, with TTIR there was a higher expression of GH28, a family that encodes polygalacturonases and rhamnogalacturonases that hydrolyse the  $\alpha$ -1,4 galacturonate linkage in HG and RG-I respectively; and GH53s, which encode for  $\beta$ -1,4-galactanases [44].

Metatranscriptomics of barley straw and TTIR cultures provided a broad picture of the enzymatic potential of the rumen in terms of CAZy families. However, to gain more insight into the enzymatic processes of ruminal digestion, differential expression analyses, a more nuanced approach with in-depth analyses was used. Based on differential expression, several GH families putatively involved in cellulose, hemicellulose, and pectin digestion were identified (Fig. 5). Informed by glycomics, the relative contribution of dominant DE transcripts was determined for candidate CAZyme families with a focus on those activities important for digestion of recalcitrant plant cell walls (Fig. 6). Since candidate transcripts were expressed by the microbial community in the presence of TTIR, it is highly likely that these transcripts coded for enzymes targeted at inaccessible or non-hydrolyzable linkages within recalcitrant polymers of TTIR. As identified transcripts represent uncharacterized microbial genes, phylogenetic analysis using SACCHARIS was employed in an attempt to postulate enzyme function and specificity for DE transcripts from the selected families GH5, GH11, GH43, and CE15 important for XG, HX, and AB saccharification. No significant difference was seen between barley straw and TTIR in the relative contribution of GH5 and GH9 family DE transcripts (Fig. 6), suggesting that rumen microbes may lack specialized xyloglucanases for the utilization of XG, despite its abundance within TTIR. Thus, xyloglucanases may represent a rate-limiting enzyme activity in the rumen, and exogenous supplementation of xyloglucanases may promote straw digestion within this environment. Although no xyloglucanases were identified, substrate-specific gene expression targeted towards HX and AB digestion was evident from phylogenetic trees of GH11, GH43 and CE15 DE transcripts, whereby the expression of specialized xylanases and arabinanases were induced within the rumen microbial community for the digestion of TTIR. Interestingly, GH11 transcripts that were overexpressed in TTIR cultures clustered distantly from the barley straw overexpressed GH11 transcripts, and some of the early diverging sequences, are GH11-like, and may represent members of a new family (Fig. 7A). The divergence of TTIR-overexpressed transcripts also suggests that changes in expression levels is a result of changes of the predominant bacterial species, differing between TTIR and barley straw microbes for utilizing recalcitrant HX. Furthermore, the TTIR-overexpressed GH11 transcript TR83509|c0\_g1 was observed to contain a unique arrangement with four tandem GH11-like domains (Fig. 7A), possibly the first report of a novel repetitive xylanase architecture. The phylogenetic analyses of the overexpressed GH43 transcripts revealed that both TTIR (TR712918|c0\_g1) and barley straw (TR941931|c2\_g1 and TR300921|c2\_g2) clustered together with *endo*- $\alpha$ -1,5-l-arabinanases (Fig. 7B), suggesting differential regulation of arabinanase activity in rumen microbiota cultured on barley straw vs

TTIR. The recalcitrance of arabinan to microbial digestion was also reflected by linkage analysis and glycome profiling (Figs. 1 and 2; Table 1). Transcripts TR400300|c0\_g2 and TR451123|c0\_g1 were overexpressed in incubations with barley straw and clustered with arabinofuranosidases, reflective of the cross-linked nature of HX in barley straw (Fig. 7B). Likewise, the three overexpressed CE15 transcripts with TTIR were embedded among acetyl xylan esterase activity, further reflecting the recalcitrant nature of HX. The results shown in this study also identified 5-linked Ara*f* within arabinan, 3-linked Ara*f* within arabinoxylan and Xyl-3AC and Xyl-2AC within acetyl-xylan as linkages that resisted microbial digestion (Fig. 7). These results suggest that rumen microorganisms are quite efficient in utilizing loosely bound pectin and hemicellulose, but are limited in their ability to digest cross-linked core pectin and hemicellulose.

Previous metatranscriptomic studies have been mostly limited to GH family levels with limited or absent information regarding substrate specificity of identified transcripts [28, 29, 38–40]. In this study, transcript-level SACCHARIS analyses in combination with glycome profiling and linkage analysis suggested that substrate-specific CAZyme transcript expression is inducible. The TTIR-overexpressed transcripts and putative activities may identify those CAZymes with the greatest potential to improve the rate and extent of recalcitrant plant cell wall residues by the rumen microbial consortia. Identified target CAZymes of rumen origin are likely to stand a better chance of success when applied as additive enzymes as they are optimized for the complex physiological conditions within the rumen. Future recombinant production and characterization of these candidate CAZymes and their screening is needed to ascertain their synergy to the natural complement of CAZymes produced by rumen microbiota. In addition, the barley straw overexpressed transcripts and their closely clustered CAZymes signifies important enzyme activities for effective barley straw digestion and hold high intellectual value for their wide application to industrial sectors such as bioenergy, and green chemistry. Here we have shown that integrating high-resolution analytical methods, such as cell wall structural characterization, *in vitro* batch culture, RNA-sequencing and DE analysis (Fig. 8), can be successfully used to identify recalcitrant polymers and resistant linkages within feeds, as well as inform the biological processes involved in ruminal digestion of polysaccharides.

## Conclusion

A comparative study involving changes in cell wall glycan composition, glycosidic linkages, and the relative extractability of polysaccharides within TTIR and barley straw provided an unprecedented level of understanding of nutrient utilization by rumen microbes and helped to delineate the role of matrix polysaccharides in cell wall recalcitrance towards microbial enzymatic digestion. Adopting TTIR as a substrate for batch culture of rumen samples and the application of next generation sequencing to study gene expression by the rumen microbial community, successfully identified a number of CAZyme genes displaying potential for downstream enzyme product development and degradation of recalcitrant residues. Furthermore, incubation of rumen microbial communities on TTIR has proven a powerful approach for enzyme discovery. In this regard, characterization of the small oligosaccharide fraction associated with TTIR may allow targeted searches from other organisms for hydrolytic enzymes capable of digesting these oligosaccharides, whose addition may enhance the rumen digestion of recalcitrant

forages. Taken together, the results presented in this study suggest that the cell wall layered structure hinders enzymatic accessibility to recalcitrant cell wall components like HX and XG. Considering the rumen is one of the most efficient biomass digestive systems known, the results of this study hold high value for bioproduct and bioprocess advances with lignocellulosic feedstocks. The integration of high-resolution glycomics with targeted metatranscriptomics analyses is an innovative experimental platform that is expandable to other feed-host systems and industrial fermentation processes as it holds vast promise for addressing rate-limiting reactions in digestive ecosystems.

## Methods

### Experimental setup and collection of samples.

All animal procedures and protocols used in this study were reviewed and approved by the Lethbridge Research and Development Center Animal Care Committee (ACC number 1501) in accordance with the guidelines of Canadian Council on Animal Care (CCAC, 2009). Overview of the experimental setup is graphically represented in Figure 8.

*In vitro* batch culture was used to culture mix rumen microbes on barley straw and total tract indigestible residues (TTIR) as substrates, as described previously [12]. TTIR was obtained from faecal samples collected from heifers (six) fed barley straw-based diets containing 70:30 forage-to-concentrate on a dry matter (DM) basis [45]. Collected faecal samples were pooled and extensively washed (6-7 times) with 50 mM citrate buffer to remove soluble particles. Given the high digestibility of the concentrate, it was assumed that the majority of TTIR originated from barley straw. The fecal residue obtained was dried for 72 h in an oven at 40 °C and defined as TTIR for use in batch culture experiments. Barley straw was dried at 40°C and ground to pass through a 1 mm screen.

*In vitro* batch cultures were set up in 3 replicate serum vials. Barley straw and TTIR were weighed (0.7g/bag) into acetone-washed and pre-weighted filter bags (F57 ANKOM bag, Ankom Technology Corp., Macedon, NY) that were then heat sealed. Individual bags were then placed in 125 mL serum vials. Rumen fluid from 4-5 different sites within rumen was collected from four ruminally cannulated Angus x Hereford heifers fed 50% grass hay, 30% barley straw, 15% corn dried distillers grains plus solubles and 5% mineral/vitamin supplement (DM basis) [13]. Collected rumen fluid was strained through cheesecloth and equal volumes from each cow were combined. The inoculum was prepared by mixing rumen fluid 1:4 with mineral buffer [13]. Inoculum (65 mL) was transferred to each vial under a stream of O<sub>2</sub>-free CO<sub>2</sub>. Vials were sealed with rubber stoppers and placed in a rotary shaker (120 rpm) in an incubator at 39 °C. Gas pressure in each vial was measured at 3, 6, 12, 24 and 48 h of incubation by inserting a 23 gauge (0.6 mm) needle attached to a pressure transducer (model PX4200-015GI, Omega Engineering, Inc., Laval, QC, Canada) and used to estimate gas production according to Romero-Pérez and Beauchemin (2018) [46]:  
Gas volume = 4.7047 × (gas pressure) + 0.0512 × (gas pressure<sup>2</sup>).

### RNA extraction and sequencing

Total RNA from the solid bound ruminal microbes were extracted as described previously [47]. Briefly, the solid contents within nylon bags from *in-vitro* batch cultures were recovered upon completion of 48 h of incubation and manually ground to a fine powder using a mortar and pestle for 5 min in liquid nitrogen. Ground samples (~ 200 mg) were placed in 2 mL microfuge tubes and each was mixed with 1.5 mL of TRIzol reagent. The samples were thawed, incubated at room temperature for 5 min, and subsequently the RNA was extracted using the acid guanidinium-phenol-chloroform (AGPC) method [48]. Total RNA was further purified with MEGAclear kit according to manufacturer instructions (Applied Biosystems/Ambion). The RNA concentration and integrity were estimated using an Agilent 2100 bioanalyzer (Agilent Technologies, Mississauga, Ontario, Canada) and RNA 6000 Nano kit (Agilent Technologies) according to the manufacturer recommendations. RNA sequencing was conducted on rRNA depleted library using the HiSeq 4000 PE100 platform at McGill University-Genome Quebec Innovation center.

### **Data analysis and bioinformatics**

*RNA-seq raw reads were submitted as input to the Kaiju [15] webserver for taxonomical assignment at the whole transcriptome level. Kaiju was run in greedy mode with a minimum match length of 11 and a minimum match score of 55, with 5 mismatches allowed. The results of the taxonomic classification with Kaiju were visualized with Krona [16].*

The raw sequenced reads were further processed using MUGQIC RNA-Seq De Novo assembly and differential analysis pipeline [49]. Raw sequenced reads were trimmed and clipped for sequencing adapters, low quality and short sequences were filtered using Trimmomatic [50], normalized and assembled using Trinity normalization utility inspired by the Diginorm algorithm [51]. The sequenced reads were assembled by the *de novo* transcriptomic assembly program Trinity. Transcripts from Trinity were aligned against the uniprot\_sprot.trinotate\_v2.0 protein database using the BLASTx program. The Trinotate suite was used for homology searches to known sequence data (BLAST+/SwissProt/Uniref90), protein domain identification (HMMER/PFAM), protein signal peptides and transmembrane domain prediction (signal/tmHMM), and for comparison to currently curated annotation databases (EMBL Uniprot eggnog/GO Pathways databases). The derived functional annotation data were integrated into SQLite database to prepare the annotation report. To identify putative carbohydrate active enzymes, all sequences were profiled against the CAZy database [16]. Gene abundance was estimated using RSEM and differentially expressed transcripts between substrates were identified using DEseq [52] and edgeR [53] package from Bioconductor. Sequence assemblies were deposited in GenBank under BioProject PRJNA673210.

For DE transcript functional analyses, DE transcript nucleotide sequences were re-submitted to the dbCAN2 meta server [19], manually curated, and categorized by CAZy family. The amino acid sequences from the dbCAN-annotated CAZymes were submitted as input to SACCHARIS for phylogenetic analyses [20]. Sequences and accession numbers of characterized GH11, GH43, and CE15 enzymes were extracted from the CAZy database [17], and ProtTest [54] was used for best-fit model selection using the sequence alignment. FastTree [55] was used to generate the phylogenetic trees which were then annotated using iTOL [56].

## Glycome profiling of barley straw and total tract indigestible residue

Alcohol insoluble residues from barley straw and TTIR samples were sequentially extracted using 50 mM ammonium oxalate, 50 mM sodium carbonate [with 0.5% (w/v) NaBH<sub>4</sub>], 1 M KOH [with 1% (w/v) NaBH<sub>4</sub>], 4 M KOH [with 1% (w/v) NaBH<sub>4</sub>], sodium chlorite solution (acidified by glacial acetic acid), and 4M KOH [with 1% (w/v) NaBH<sub>4</sub>] resulting in five wall extracts designated AO, SC, 1M KOH, 4M KOH, CH, and 4M KOHPC, respectively [57]. These plant cell wall extracts were probed with a collection of 154 cell wall glycan-directed mAbs [58] using an enzyme-linked ELISA, and binding intensities from this experiment presented as a heat map as described previously [57].

## Linkage analysis of barley straw and total tract indigestible residue

Alcohol insoluble residues (AIRs) were prepared according to Wood *et al.* [59]. The dried AIRs were de-starched by incubation with porcine pancreatic  $\alpha$ -amylase in 10 mM Tris-maleate buffer (pH 6.9) following Pettolino *et al.* [11]. The de-starched AIRs were further fractionated based on the literature [60-63] with modifications. Briefly, the cell walls (~55 mg) were treated with 10 mL of 0.25 M sodium borodeuteride (NaBD<sub>4</sub>, 99 atom % D, Alfa Aesar) deionized water solution at 4°C for 24 h, followed by quenching the excess reductant by dropwise addition of 10% (v/v) acetic acid and adjusting pH to neutral. The suspensions were then centrifuged (3000 × g, 0.5 h), and resulting supernatants were pooled with three deionized water washes of the pellets (with centrifugation between each wash), and a pooled solution (Solution 1) was stored at 4 °C. The pellets were then subjected to sequential extraction of polysaccharides using 15 mL of each of the following solutions: 50 mM EDTA (pH 6.5), 50 mM Na<sub>2</sub>CO<sub>3</sub> (containing 25 mM NaBD<sub>4</sub>), and 4 M KOH (containing 25 mM NaBD<sub>4</sub>) over 24 h at room temperature with gentle magnetic stirring. After extraction, soluble fractions were separated from residues by centrifugation (3000 × g, 0.5 h), followed by neutralizing the supernatants, washing the residues three times with deionized water (with centrifugation between washes), and pooling all washes into corresponding fractions. The EDTA and Na<sub>2</sub>CO<sub>3</sub> extracts and the aforementioned Solution 1 were pooled (designated F<sub>EDTA+Na<sub>2</sub>CO<sub>3</sub></sub>). The 4 M KOH fraction and the final residue left after strong alkaline extraction were designated F<sub>4MKOH</sub> and F<sub>Residue</sub>, respectively. All fractions were dialyzed against deionized water and lyophilized. For linkage analysis of the whole cell walls and the isolated fractions, uronic acids in the samples (~2 mg) were first converted to their 6,6-dideuterio neutral sugars using carbodiimide activation at pH 4.75 followed by NaBD<sub>4</sub> reduction at pH 7.0. Samples were then dialyzed against deionized water, and freeze-dried. The carboxyl-reduced samples were converted to their partially methylated alditol acetate (PMAA) derivatives by permethylation with iodomethane and sodium hydroxide in dimethyl sulfoxide [64,65] with 2 M trifluoroacetic acid at 121°C for 1.5 h, reduction with NaBD<sub>4</sub>, and peracetylation with acetic anhydride [12]. Deutero-methylation using iodomethane-d<sub>3</sub> (≥99.5 atom % D, Sigma) was applied to the whole cell wall and F<sub>4MKOH</sub> fractions in order to identify and quantify the linkages from endogenously *O*-methylated sugars (e.g., 4-*O*-methyl glucuronic acids) [66-68]. The PMAAs were then subjected to GC-MS and GC-FID analyses on an Agilent 7890A-5977B GC-MS/FID system (Agilent Technologies, Santa Clara, CA). The PMAAs were separated using a medium polarity Supelco SP-2380 capillary column (30 m × 0.25 mm × 0.20 μm, Sigma-Aldrich)

with a constant column outlet helium flow rate of 0.8 mL/min. Sample solutions were injected at an inlet temperature of 250°C with a split ratio of 10:1. The oven temperature was programmed to start at 120 °C (hold 4 min) followed by increasing at 8 °C/min to 175 °C, 0.5 °C/min to 183 °C (hold 8 min), 0.5 °C/min to 195 °C, 4 °C/min to 210 °C, and 20 °C/min to 255 °C (hold 8 min). The transfer line temperature was kept at 280 °C. The mass spectrometer was operated in electron ionization (EI), full-scan mode (ionization energy, 70 eV; source temperature, 230 °C; quad temperature, 150 °C). The FID detector was operated at 300 °C (H<sub>2</sub> flow, 30 mL/min; air flow, 400 mL/min; N<sub>2</sub> makeup flow, 25 mL/min). The PMAAs were then identified by comparing their MS fragmentation patterns with those of reference derivatives and the literature [69], and quantified based on the previously reported FID response factors calculated using the effective carbon number concept [70]. Glycosidic linkages were assigned to relevant cell wall polysaccharides, and the relative compositions of each type of polysaccharides were then estimated by summing up corresponding linkage compositions [12,71]. Six separate experiments were conducted for the whole cell walls and the F<sub>4MKOH</sub> fractions, of which half were used for deuterio-methylation analyses and the other half for methylation analyses. The F<sub>Residue</sub> fractions were subjected to six separate experiments (methylation analysis only). Three separate methylation analyses were conducted for the F<sub>EDTA+Na<sub>2</sub>CO<sub>3</sub></sub> fractions.

## Abbreviations

AIR: alcohol insoluble residue, ADF: acid detergent fiber, AG: arabinogalactan, AO: ammonium oxalate, CAZyme: carbohydrate active enzymes, CE: carbohydrate esterases, CH: chlorite, ELISA: *enzyme-linked immunosorbent assay*, GH: glycosyl hydrolase, HG: homogalacturonan, HX: hetero xylan, mAbs: monoclonal antibodies, NDF: neutral detergent fiber, PC: post-chlorite, PL: pectin lyases, RG: rhamnogalacturonan, SC: sodium carbonate, TTIR: total tract indigestible residue, XG: xyloglucan.

## Declarations

### Ethics approval and consent to participate

All animal procedures and protocols used in this study were reviewed and approved by the Lethbridge Research and Development Center Animal Care Committee (ACC number 1501) in accordance with the guidelines of Canadian Council on Animal Care (CCAC, 2009).

### Consent for publication

No consent was required for this publication

### Availability of data and material

Sequenced transcripts have been deposited into the NCBI SRA database BioProject accession number: PRJNA673210.

### Competing interests

Authors declare no competing interest.

## Funding

This work was funded by Alberta Agriculture and Forestry, Canada: project ID 2018F182R, and Agriculture and Agri-Food Canada: project ID: J-001589. Glycome profiling of biomass samples was supported by the Center for Bioenergy Innovation (Oak Ridge National Laboratory), US Department of Energy (DOE) Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

## Authors' contributions

XX, AKB, DRJ, SV, and MGH contributed for experiment and data analysis for cell wall characterization. AKB, ROP performed *in vitro* batch culture, RNA extraction, transcriptome sequence analysis. KEL and MR contributed towards data processing, statistical support and graphical representation of data. AKB, DWA, and TAM were funded and conceived the study. AKB, DWA and TAM wrote the manuscript and all authors reviewed and approved the manuscript.

## Acknowledgements

The authors thank the Lethbridge Research and Development Centre (AAFC) staff for their technical and animal care assistance.

## Authors' information (optional)

## References

1. Alexandratos N, Bruinsma J. World agriculture towards 2030/2050: the 2012 revision. In: ESA Working paper No. 12-03. Rome. FAO. 2012. <http://www.fao.org/3/a-ap106e.pdf>. Accessed 26 Oct 2020.
2. Elam TE. Projections of global meat production through 2050. 2006. <https://docplayer.net/21349988-Projections-of-global-meat-production-through-2050.html> Accessed 26 Oct 2020.
3. Sokhansanj S, Mani S, Stumborg M, Samson, R, Fenton J. Production and distribution of cereal straw on the Canadian Prairies. *Can. Biosys. Engineering*. 2006; 48: 3.39-3.46.
4. Kim S, Dale BE. Global potential bioethanol production from wasted crops and crop residues. *Biomass Bioenerg*. 2004; 26:361-375.
5. Hatfield RD, Ralph J, Grabber JH. Cell Wall Structural Foundations: Molecular basis for Improving forage digestibilities. *Crop Sci*. 1999; 39:27-37.
6. Badhan A, Wang Y, Gruninger RJ, Patton D, Powlowski J, Tsang A et al. Formulation of enzyme blends to maximize the hydrolysis of alkaline peroxide pretreated alfalfa hay and barley straw by rumen enzymes and commercial cellulases. *BMC Biotechnol*. 2014; 14:31.
7. Ribeiro GO, Gruninger RJ, Badhan A, McAllister TA. Mining the rumen for fibrolytic feed enzymes. *Anim. Front* . 2016; 6:20–26.

8. Meale SJ, Beauchemin KA, Hristov AN, Chaves AV, McAllister TA. Board-invited review: Opportunities and challenges in using exogenous enzymes to improve ruminant production. *J Anim. Sci.* 2014; 92:427–442.
9. Sun X, Joblin K, Andrew I, Hoskin S, Harris P. Degradation of forage chicory by ruminal fibrolytic bacteria. *J. Appl. Microbiol.* 2008; 105:1286-1297
10. Pattathil S, Avci U, Zhang T, Cardenas CL, Hahn MG. Immunological approaches to biomass characterization and utilization. *Front. Bioeng. Biotechnol.* 2015; 3: 173
11. Pattathil S, Hahn MG, Dale BE, Chundawat SPS. Insights into plant cell wall structure, architecture, and integrity using glycome profiling of native and AFEX™-pre-treated biomass. *Exp. Bot.* 2015; 66: 4279–94.
12. Pettolino FA, Walsh C, Fincher GB, Bacic A. Determining the polysaccharide composition of plant cell walls. *Nat. Protoc.* 2012; 7:1590-1607.
13. Ribeiro GO, Badhan A, Huang JL, Beauchemin KA, Yang WZ, Wang YX, et al. New recombinant fibrolytic enzymes for improved *in-vitro* ruminal fiber degradability of barley straw. *J. Anim. Sci.* 2018; 96:3928–3942.
14. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I. et al. Full-Length Transcriptome Assembly from RNA-Seq Data Without a Reference Genome. *Biotechnol.* 2011; 29: 644–652.
15. Menzel P, Ng K, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Commun.* 2016; 7: 11257.
16. Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics.* 2011;12:
17. Lombard V, Ramulu HG, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 2014; 42: D490–D495
18. Couger MB, Youssef NH, Struchtemeyer CG, Ligenstoffer AS, Elshahed ME. Transcriptomic analysis of lignocellulosic biomass degradation by the anaerobic fungal isolate *Orpinomyces* strain C1A. *Biotechnol. Biofuels.* 2015; 8: 208.
19. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 2018; 46: W95–W101
20. Jones DR, Thomas D, Alger N, Ghavidel A, Inglis GD, Abbott DW. SACCHARIS: an automated pipeline to streamline discovery of carbohydrate active enzyme activities within polyspecific families and de novo sequence datasets. *Biotechnol. Biofuels.* 2018; 11: 27.
21. Beigh YA, Ganai AM, Ahmad HA. Prospects of complete feed system in ruminant feeding: A review. *World.* 2017;10: 424–437.
22. Terry S, Badhan A, Wang Y, Chaves VA, McAllister TA. Fibre digestion by rumen microbiota – a review of recent metagenomic and metatranscriptomic studies. *Can. J. Ani. Sci.* 2019;94: 678-692.
23. Adesogan AT, Ma ZX, Romero JJ, Arriola KG. RUMINANT NUTRITION SYMPOSIUM: Improving cell wall digestion and animal performance with fibrolytic enzymes. *Ani. Sci.* 2014;92: 1317–1330

24. Vicini JL, Bateman HG, Bhat MK, Clark JH, Erdman RA, Phipps RH, et al. Effect of feeding supplemental fibrolytic enzymes or soluble sugars with malic acid on milk production. *J. Dairy. Sci.* 2003; 576–585.
25. Colombatto D, Mould FL, Bhat MK, Phipps RH, Owen E. *In-vitro* evaluation of fibrolytic enzymes as additives for maize (*Zea mays* L.) silage: II. Effects of ensiling temperature, enzyme source and addition level. *Anim. Feed Sci. Technol.* 2004; 86:111–128.
26. Pattathil S, Ingwers MW, Victoriano OL, Kandemkavil S, McGuire MA, Teskey RO, Aubrey DP. Cell Wall Ultrastructure of Stem Wood, Roots, and Needles of a Conifer Varies in Response to Moisture Availability. *Front. Plant Sci.* 2016 ;7:882.
27. Ruprecht C, Bartetzko MP, Senf D, Dallabernardina P, Boos I, Andersen MCF, Kotake T, Knox JP, Hahn MG, Clausen MH, Pfrenge F. A synthetic glycan microarray enables epitope mapping of plant cell wall glycan-directed antibodies. *Plant Physiol.* 2017; 175: 1094-1104.
28. Dai X, Tian Y, Li J, Su X, Wang X, Zhao S, et al. Metatranscriptomic Analyses of Plant Cell Wall Polysaccharide Degradation by Microorganisms in the Cow Rumen. *Appl. Environ. Microbiol.* 2015;81: 1375-1386.
29. Comtet-Marre S, Parisot N, Lepercq P, Chaucheyras-Durand F, Mosoni P, Peyretailade E, Bayat, et al. Metatranscriptomics Reveals the Active Bacterial and Eukaryotic Fibrolytic Communities in the Rumen of Dairy Cow Fed a Mixed Diet. *Front. Micro.* 2017 ;8: 67.
30. Tapio I, Shingfield KJ, Mckain N, Bonin A, Fischer D, Bayat AR, et al. Oral Samples as non-invasive proxies for assessing the composition of the rumen microbial community. *PLoS ONE.* 2016;11: e0151220.
31. Golder HM, Thomson JM, Denman SE, Mcsweeney CS, Lean IJ. Genetic markers are associated with the ruminal microbiome and metabolome in grain and sugar challenged dairy heifers. *Genet.* 2018; 9:62.
32. Brunecky R, Alahuhta M, Xu Q, Donohoe BS, Crowley MF, Kataeva IA, et al. Revealing nature's cellulase diversity: the digestion mechanism of *Caldicellulosiruptor bescii* CelA. 2013;342: 1513–1516
33. Barr BK, Hsieh YL, Ganem B, Wilson DB. Identification of two functionally different classes of exocellulases. 1996; 35:586–592
34. Seshadri R, Leahy SC., Attwood GT, Teh KH, Lambie SC, Cookson AL, et al. Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nat. Biotechnol.* 2018;36: 359.
35. Brulc JM, Antonopoulos DA, Berg-Miller ME, Wilson MK, Yannarell AC, Dinsdale EA, et al. Gene-centric metagenomics of the fibre-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proc. Natl. Acad. Sci.* 2009; 106:1948-1953.
36. Dai X, Zhu Y, Luo Y, Song L, Liu D, Liu L, et al. Metagenomic Insights into the Fibrolytic Microbiome in Yak Rumen. *PLoS ONE.* 2012;7: e40430
37. Pope PB, Denman SE, Jones M, Tringe SG., Barry K, Malfatti SA, et al. Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside hydrolase profiles different from other herbivores. *Proc. Natl. Acad. Sci.* 2010; 107:14793-14798.

38. Qi M, Wang P, O'Toole N, Barboza PS, Ungerfeld E, Leigh MB, et al. Snapshot of the Eukaryotic Gene Expression in Muskoxen Rumen—A Metatranscriptomic Approach. *PLoS ONE*. 2011; e20521
39. Shinkai T, Mitsumori M, Sofyan A, Kanamori H, Sasaki H, Katayose Y, et al. Comprehensive detection of bacterial carbohydrate-active enzyme coding genes expressed in cow rumen. *Anim. Sci. J.* 2016; 87:1363–1370.
40. Li F and Guan LL. Metatranscriptomic profiling reveals linkages between the active rumen microbiome and feed efficiency in beef cattle. *Appl. Environ. Microbiol.* 2017; e00061-17.
41. Mewis K, Lenfant N, Lombard V, Henrissat B. Dividing the large glycoside hydrolase family 43 into subfamilies: a motivation for detailed enzyme characterization. *Appl. Environ. Microbiol.* 2016 ; 82 :1686 –1692.
42. Pollet A, Delcour JA, Courtin CM. Structural determinants of the substrate specificities of xylanase from different glycoside hydrolase families. *Crit. Rev. Biotechnol.* 2010; 30:176-191.
43. O'Neill EC, Stevenson CE, Paterson MJ, et al. Crystal structure of a novel two domain GH78 family  $\alpha$ -rhamnosidase from *Klebsiella oxytoca* with rhamnose bound. *Proteins*. 2015; 83:1742–1749
44. Sprockett DD, Piontkivska H, Blackwood CB. Evolutionary analysis of glycosyl hydrolase family 28 (GH28) suggests lineage-specific expansions in necrotrophic fungal pathogens. *Gene*. 2011;479: 29-36.
45. Ribeiro GO, Oss DB, He Z, Gruninger RJ, Elekwachi C, Forster RJ, et al. Repeated inoculation of cattle rumen with bison rumen contents alters the rumen microbiome and improves nitrogen digestibility in cattle. *Rep.* 2017; 7:1276.
46. Romero-Perez, A., Beauchemin, K.. Estimating gas volume from headspace pressure in a batch culture system. *Can. J. Anim. Sci.* 2018 ; 98:593-596.
47. Wang P, Qi M, Barboza P, Leigh MB, Ungerfeld E, Selinger LB, et al. Isolation of high-quality total RNA from rumen anaerobic bacteria and fungi, and subsequent detection of glycoside hydrolases *J. Microbiol.* 2011; 57:590-598.
48. Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Biochem.* 1987; 162:156-159.
49. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J et al. De Novo Transcript Sequence Reconstruction from RNA-Seq Using the Trinity Platform for Reference Generation and Analysis. *Protoc.* 2013; 8:1494–1512.
50. Bolger AM, Lohse M, Usadel B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics*.2014; 30:2114–20.
51. Titus B, Howe CA, Zhang Q, Pyrkosz AB, Brom TH. A Reference-Free Algorithm for Computational Normalization of Shotgun Sequencing Data. *org.* 2012 <http://arxiv.org/abs/1203.4802>.
52. Ander S, Huber W. Differential Expression Analysis for Sequence Count Data. *Genome Biology*. 2010; 11: R106.
53. Robinson MD, McCarthy DJ, Smyth G.K. EdgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. 2010;26: 139–40.

54. Darriba D, Guillermo L, Doallo TR, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. *2011*; 27, 1164–1165.
55. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE*; 2010: 5(3), e9490.
56. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.*; 2019: 47, W256–W259.
57. Pattathil S, Avci U, Miller JS, Hahn MG. Immunological Approaches to Plant Cell Wall and Biomass Characterization: Glycome Profiling. In: Himmel M, editors. *Biomass Conversion. Methods in Molecular Biology (Methods and Protocols)*. Humana Press, Totowa, NJ. 2012. p 908.
58. Pattathil S, Avci U, Baldwin D, Swennes AG, McGill JA, Popper Z, et al. A comprehensive toolkit of plant cell wall glycan-directed monoclonal antibodies. *Plant Physiol.* 2010;153: 514-525.
59. Wood JA, Tan HT, Collins HM, Yap K, Khor SF, Lim WL, et al. Genetic and environmental factors contribute to variation in cell wall composition in mature desi chickpea (*Cicer arietinum*) cotyledons. *Plant, Cell Environ.* 2018; 41: 2195-2208.
60. Hsieh YSY, Chien C, Liao SKS, Liao SF, Hung WT, Yang WB, Lin CC, Cheng TJ, Chang CC, Fang JM, Wong CH. Structure and bioactivity of the polysaccharides in medicinal plant *Dendrobium huoshanense*. *Bioorgan. Med. Chem.* 2008; 16: 6054-6068.
61. Li J, Wang DM, Xing X, Cheng TJR, Liang PH, Bulone V, Park JH, Hsieh YSY. Structural analysis and biological activity of cell wall polysaccharides extracted from *Panax ginseng* *Int. J. Biol. Macromol.* 2019;135: 29-37.
62. Bacic A, Moody SF, Clarke AE. Structural analysis of secreted root slime from maize (*Zea mays*). *Plant Physiol.* 1986; 80: 771-777.
63. Kim JB, Carpita NC. Esterification of maize cell wall pectins related to cell expansion. *Plant Physiol.* 1992; 98: 646-653.
64. Ciucanu I, Kerek F. A simple and rapid method for the permethylation of carbohydrates. *Carbohydr. Res.* 1984; 131: 209-217.
65. Albersheim P, Nevins DJ, English PD, Karr A. A method for the analysis of sugars in plant cell wall polysaccharides by gas liquid chromatography. *Res.* 1967; 5: 340-345.
66. Darvill AG, McNeil M, Albersheim P. Structure of plant cell walls VIII. A new pectic polysaccharide. *Plant Physiol.* 1978; 62: 418-422.
67. Dell A, Khoo KH, Panic M, McDowell RA, Etienne A, Reason AJ, Morris HR. FAB-MS and ESI-MS of glycoproteins. In: Fukuda M, Kobata A, editor. *Glycobiology, A Practical Approach*. Oxford University Press; 1989 p. 187-222.
68. Chiovitti A, Bacic A, Craik DJ, Kraft GT, Liao ML. A nearly idealized 6'-O-methylated  $\kappa$ -carrageenan from the Australian red alga *Claviclonium ovatum* (Acrotylaceae, Gigartinales). *Carbohydr. Res.* 2004; 339: 1459-1466.
69. Carpita NC, Shea EM. Linkage structure of carbohydrates by gas chromatography-mass spectrometry (GC-MS) of partially methylated alditol acetates. In: Biermann CJ, McGinnis GD, editors. *Analysis of*

Carbohydrates by GLC and MS. CRC Press. 1988. p. 157-216.

70. Sweet DP, Shapiro RH, Albersheim P. Quantitative analysis by various G.L.C. response-factor theories for partially methylated and partially ethylated alditol acetates. *Carbohydr. Res.* 1975; 40: 217-225.
71. Sowinska EE, Gilbert S, Lamb E, Carpita NC. Linkage structure of cell-wall polysaccharides from three duckweed species. *Carbohydr. Polym.* 2019; 223: 115-119.
72. Jones P, Binns D, Chang HU, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. 2014; 30: 1236–1240.

## Additional Files

**Additional file 1:** (A) The relative abundance of the glycosidic linkages as determined by gas chromatography – mass spectrometry/flame ionization detection (GC-MS/FID) of partially methylated acetylated alditol (PMAA) derivatives. (B) Assignment of linkages to the polysaccharides according to Pettolino *et al.* (2012) [12].

**Additional file 2:** Trinity de novo assembly metrics.

**Additional file 3:** Identified CAZyme transcripts based on gene annotation results. Transcripts from Trinity were aligned against the uniprot\_sprot.trinotate\_v2.0 protein database using the BLASTx program. The Trinotate suite was used for homology searches to known sequence data (BLAST+/SwissProt/Uniref90), protein domain identification (HMMER/PFAM), protein signal peptides and transmembrane domain prediction (signal/tmHMM), and for comparison to currently curated annotation databases (EMBL Uniprot eggNOG/GO Pathways databases). To identify putative carbohydrate active enzymes, all sequences were profiled against the CAZy database [16].

**Additional file 4:** The relative abundance of CAZy family and relative contribution of each family towards total CAZyme expression. The relative abundance is defined as the proportion of transcripts for a given family in the total of all CAZyme (GH, CE, and PL only) transcripts. Percentage of expression for transcripts for GH families putatively out of the total CAZyme (GH, CE, and PL only) expression was quantified. FPKM values were normalized by library size (as calculated by DEseq).

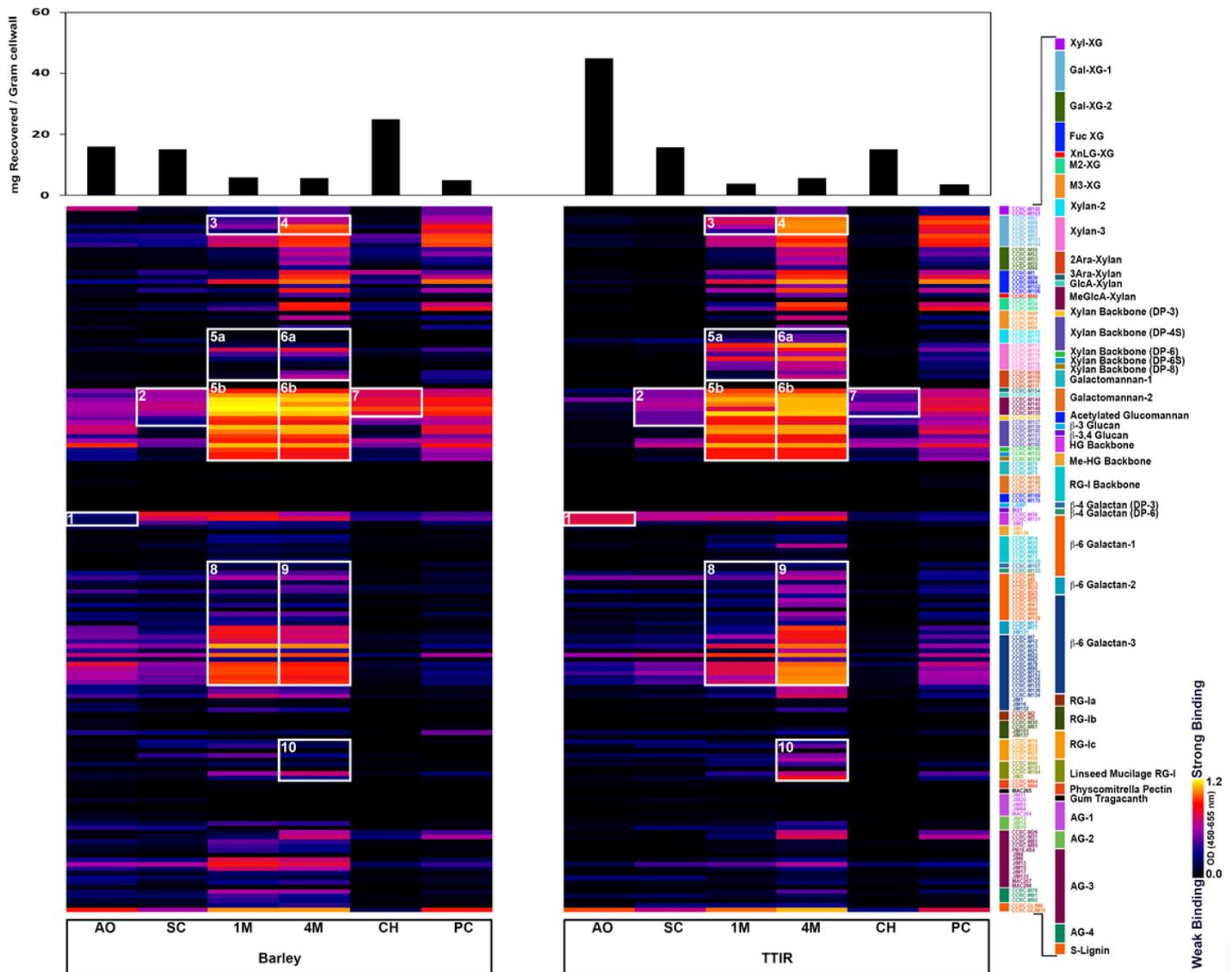
**Additional file 5:** Taxonomic affiliation of transcripts according to Kaiju analysis.

**Additional file 6:** Relative contribution of GH family transcripts for *in vitro* BS and TTIR-supplemented bacterial cultures.

**Additional file 7:** Predicting CAZyme activities from individual *differentially expressed (DE) transcripts by dbCAN2*.

**Additional file 8.** Cumulative gas production by rumen microbiota when incubated with BS and TTIR under *in vitro* batch culture.

# Figures



**Figure 1**

Glycome profile of barley straw and total tract indigestible residue (TTIR). Barley straw and total tract indigestible residue (TTIR) samples were sequentially extracted with ammonium oxalate (AO), sodium carbonate (SC), 1 M (1M) and 4 M (4M) KOH, chlorite (CH), and post-chlorite 4 M KOH (PC). The total carbohydrate recovered per gram of cell wall is indicated by bar graph (top). Cell wall extracts were analyzed through an ELISA-based screen against a collection of plant cell wall glycan-directed monoclonal antibodies (mAbs), and results obtained were shown as a heat map, where each row reflects the binding of a single mAb against the different extracts. The colour of each element in the heat map represents the strength of the ELISA signal, as indicated (black = no binding, yellow = strong binding). The panel on the right hand side depicts the clade of mAbs that are specific to various plant cell wall glycans. Numbered

white boxes represent major differences in mAb binding intensities between barley straw and TTIR samples.

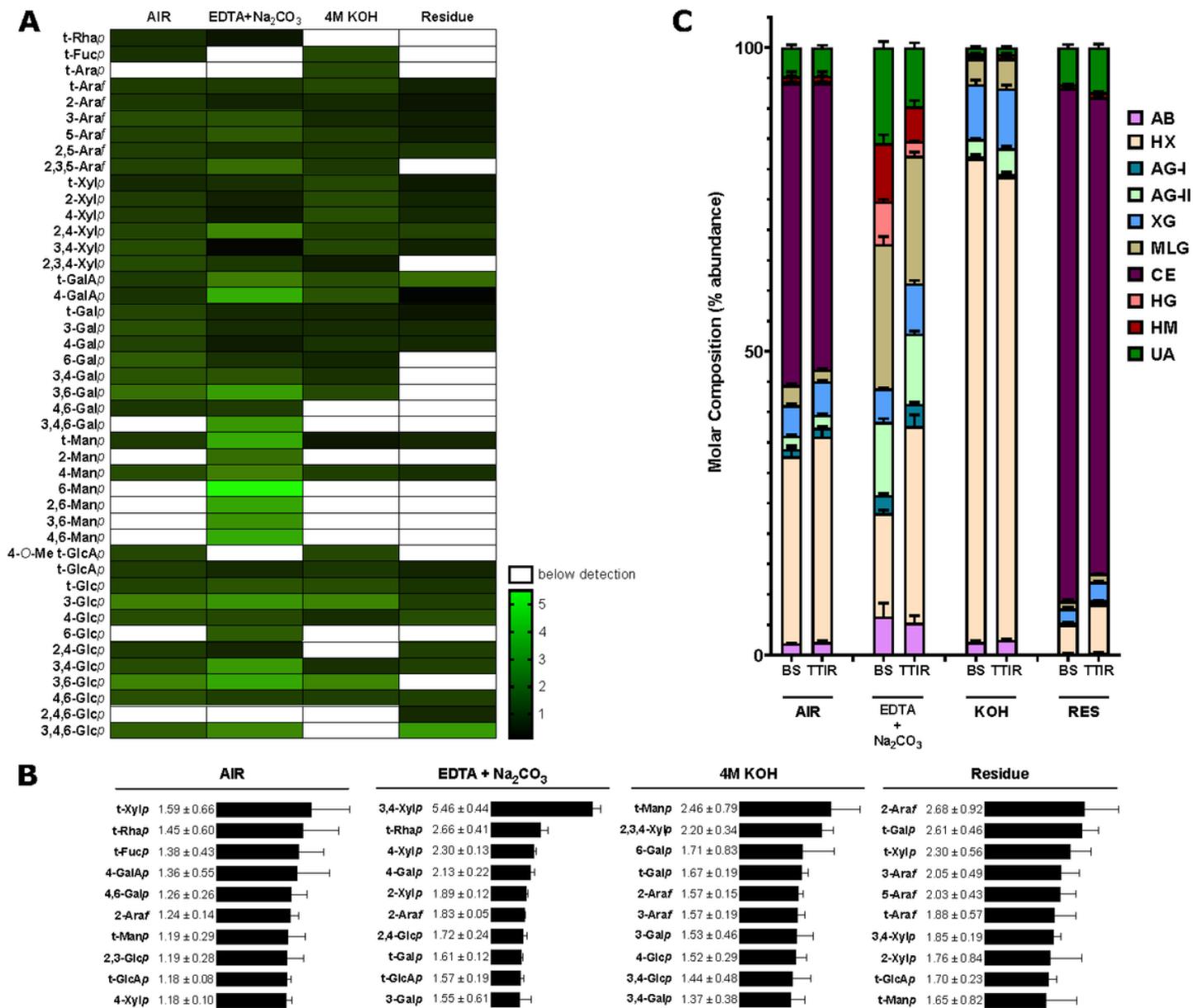
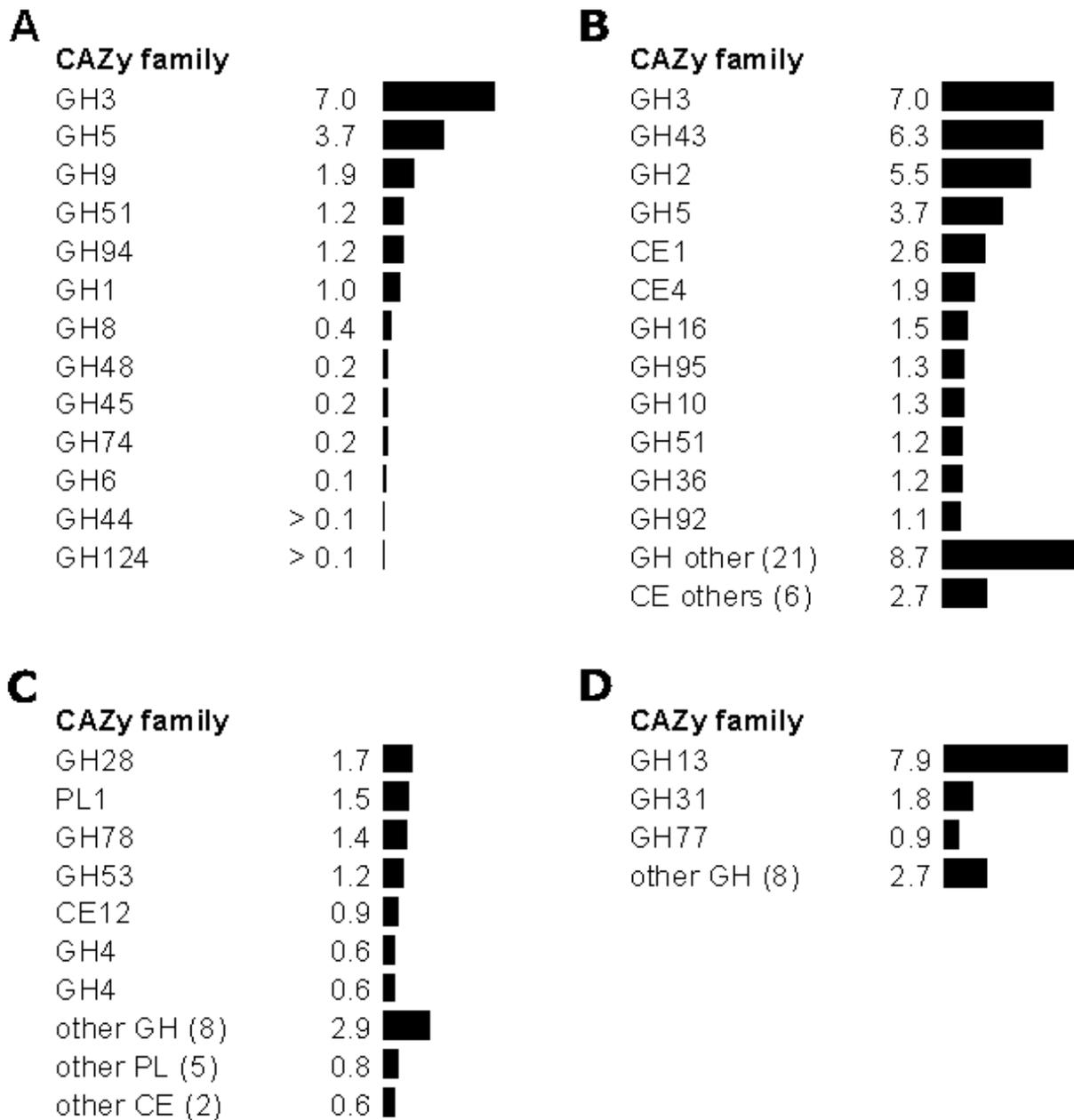


Figure 2

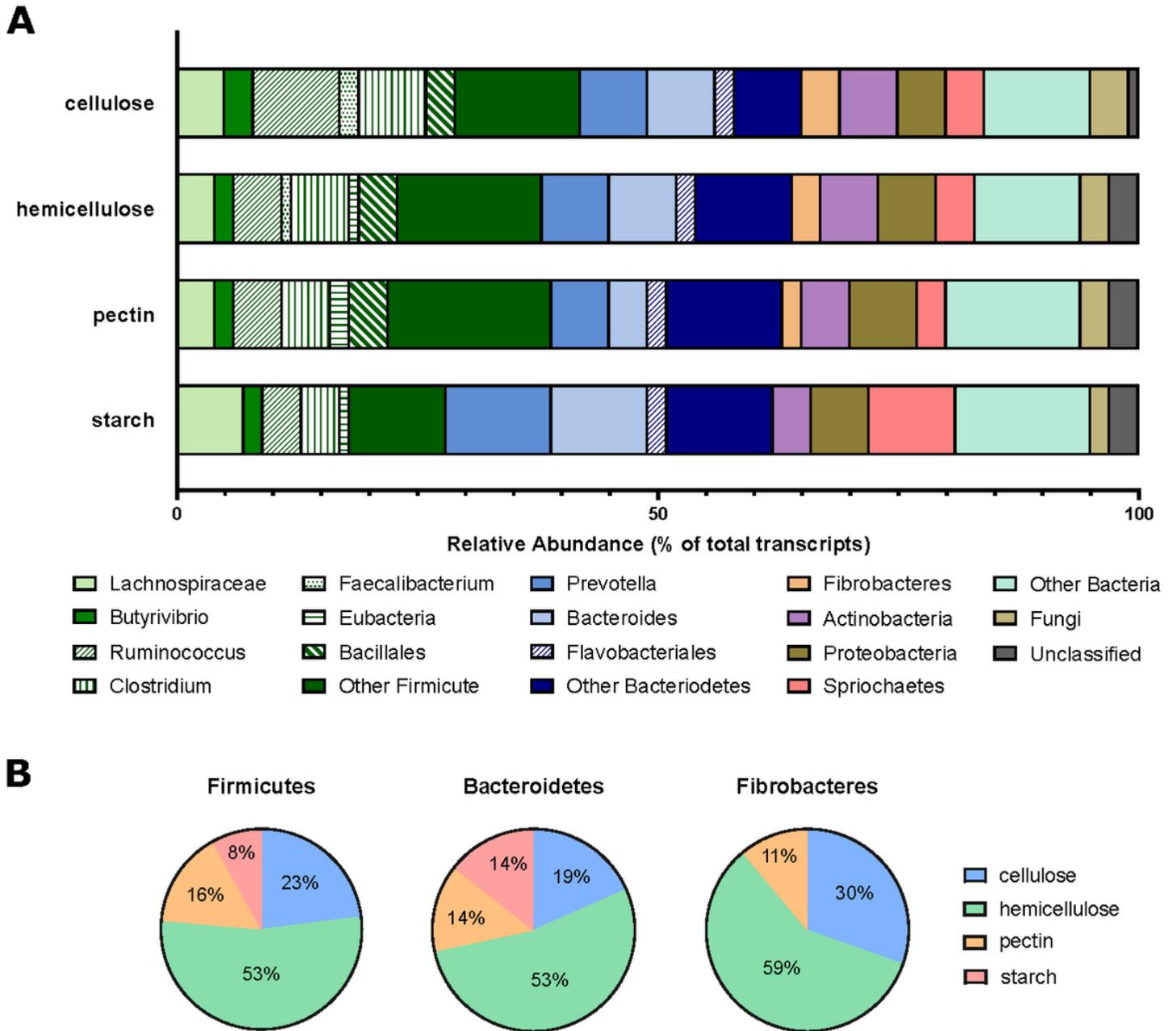
Digestibility of barley straw plant cell wall linkages. (A) Glycans were sampled from barley straw (BS) and total tract indigestible residue (TTIR) plant whole cell wall (de-starched alcohol insoluble residue, AIR) and isolated non-starch polysaccharides fractions (EDTA+Na<sub>2</sub>CO<sub>3</sub>, 4 M KOH, and cellulosic-residue). The relative abundance (molar %) of glycosidic linkages was determined by gas chromatography – mass spectrometry/flame ionization detection (GC-MS/FID) of partially methylated acetylated alditol (PMAA) derivatives. Heat maps were generated based on the calculated ratio of linkage abundance in the barley straw control over the digested TTIR (black = indigestible, bright green = most digestible). (B) Bar graphs were generated for the top 10 indigestible linkages within each plant cell wall fraction, representing the ratio of the highest average abundance calculated in TTIR over that of barley straw control. Linkages with

standard deviations of greater than 50% were excluded as they were considered unreliable. (C) Linkage composition data for barley straw and TTIR were assigned to classes of polysaccharides according to Pettolino, et al. [12].



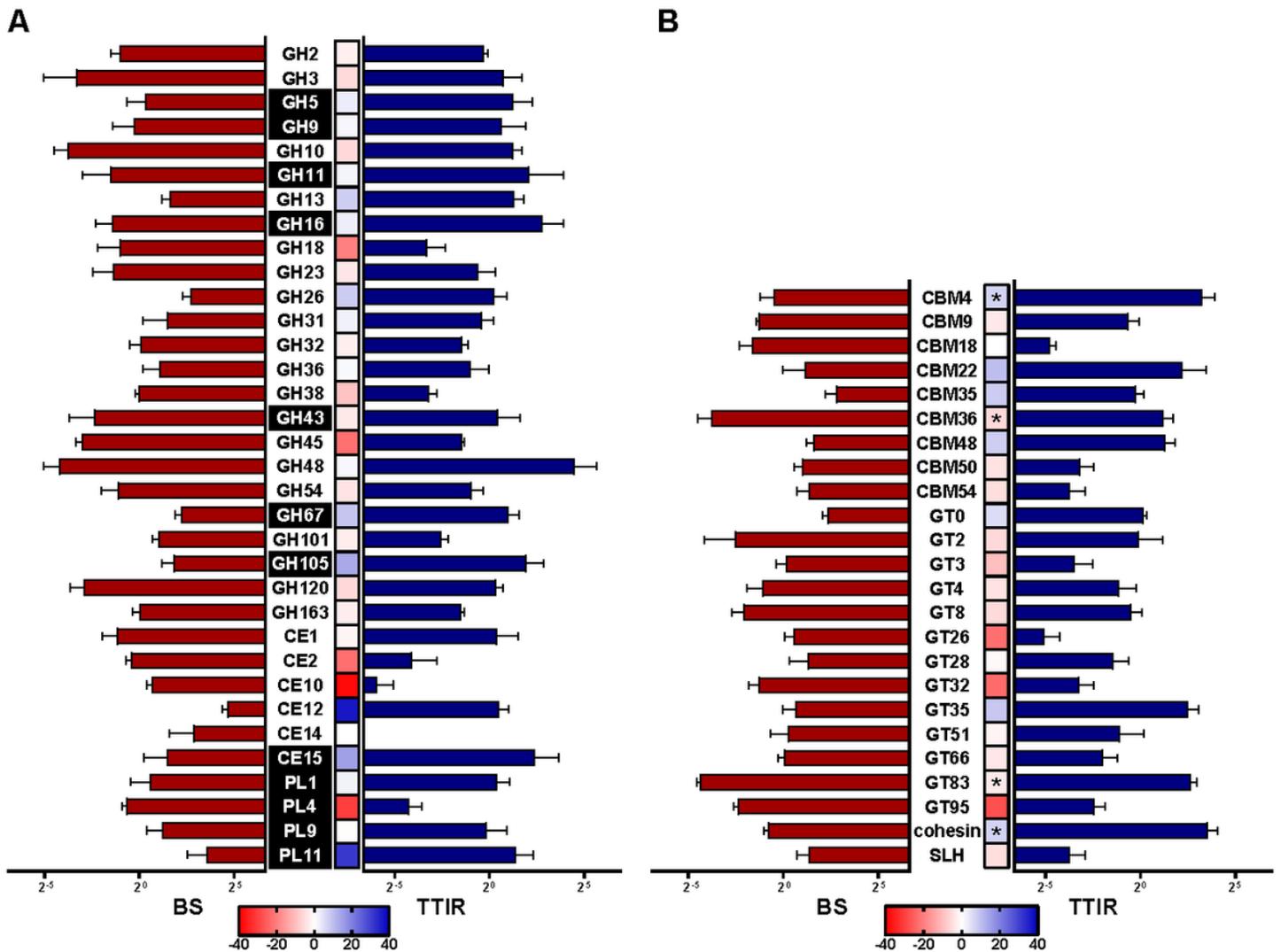
**Figure 3**

Proportional expression of CAZy families involved in cell wall degradation. The relative abundance of transcripts corresponding to CAZy families with known activities involved in the digestion of (A) cellulose, (B) hemicellulose, (C) pectin, and (D) starch. The relative abundance is defined as the proportion of transcripts for a given family in the total of all CAZyme (GH, CE, and PL only) transcripts targeting each plant polysaccharide.



**Figure 4**

Taxonomic classification of transcripts involved in the digestion of key plant cell wall components. Taxonomic affiliation and relative abundance of transcripts putatively known to be involved in the digestion of cellulose, hemicellulose, pectin, and starch as assigned by Kaiju. (A) The relative abundance is defined as the proportion of transcripts from given subphyla in the total of CAZyme (GH, CE, and PL only) transcripts targeting individual plant polysaccharides as indicated. (B) Percent of transcript originating from the phyla: Firmicutes, Bacteroidetes, or Fibrobacter putatively targeted to different cell wall component were shown.



**Figure 5**

Predicting CAZyme activities from individual transcripts. Differential expression of CAZyme genes in the presence of barley straw and TTIR, plotted on a Log<sub>2</sub> scale. Transcripts with a fold change of 2 or more, with an FDR-adjusted  $p < 0.05$ , are considered differentially expressed. Transcript nucleotide sequences were run through dbCAN2 [18], manually curated, and categorized by CAZy family: (A) GH, CE, and PL; (B) CBM, GT, cohesin, and SLH. Cultures were incubated in triplicate with barley straw (red) or TTIR (blue) were used for RNA extraction. A heat map was generated for the fold change of transcript levels for each CAZy family in TTIR over barley straw (blue), or vice versa (red). Statistical significance of the differences between TTIR and barley straw CAZy family transcripts were calculated by two-way ANOVA (\*  $p < 0.01$ ).

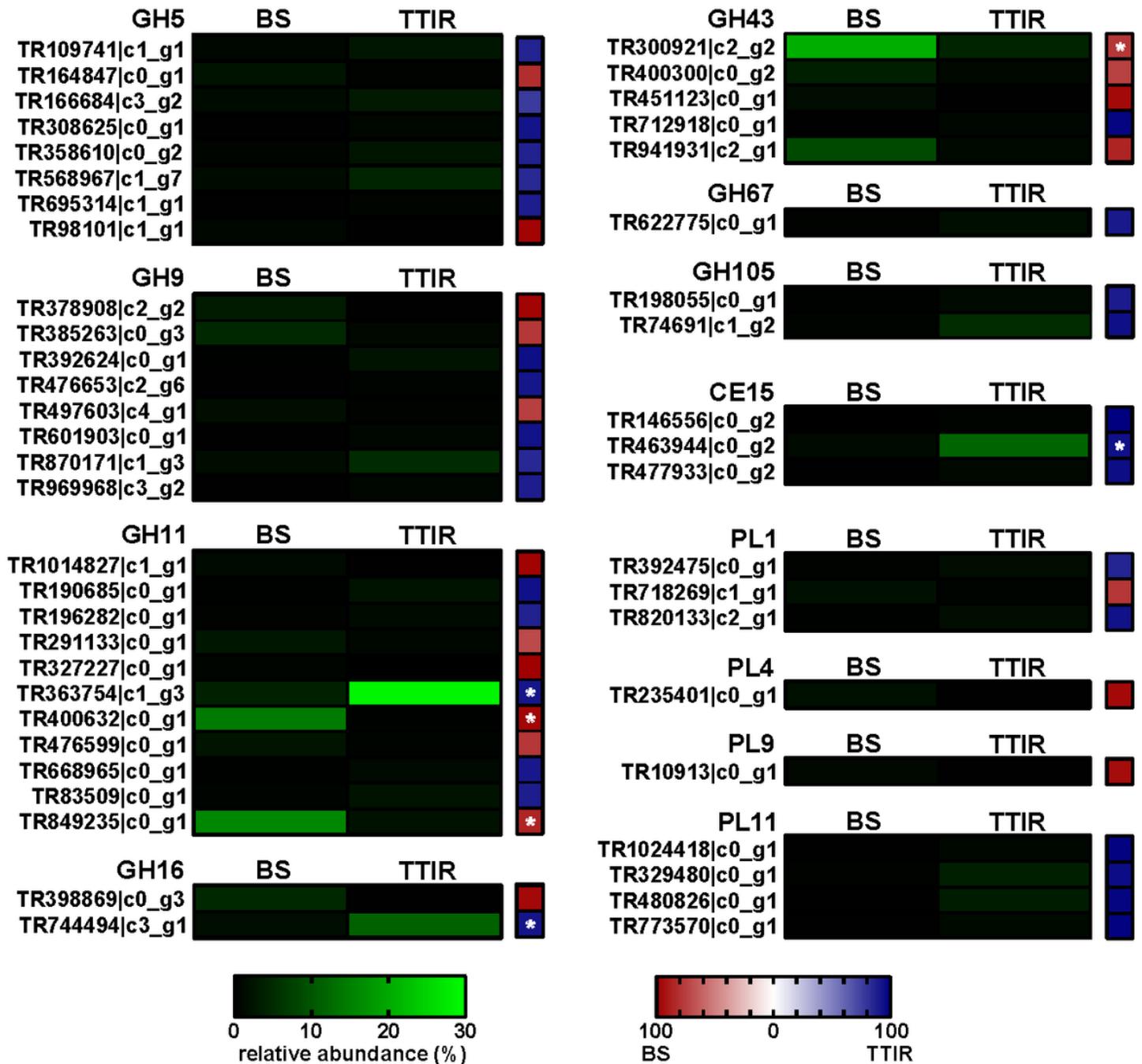
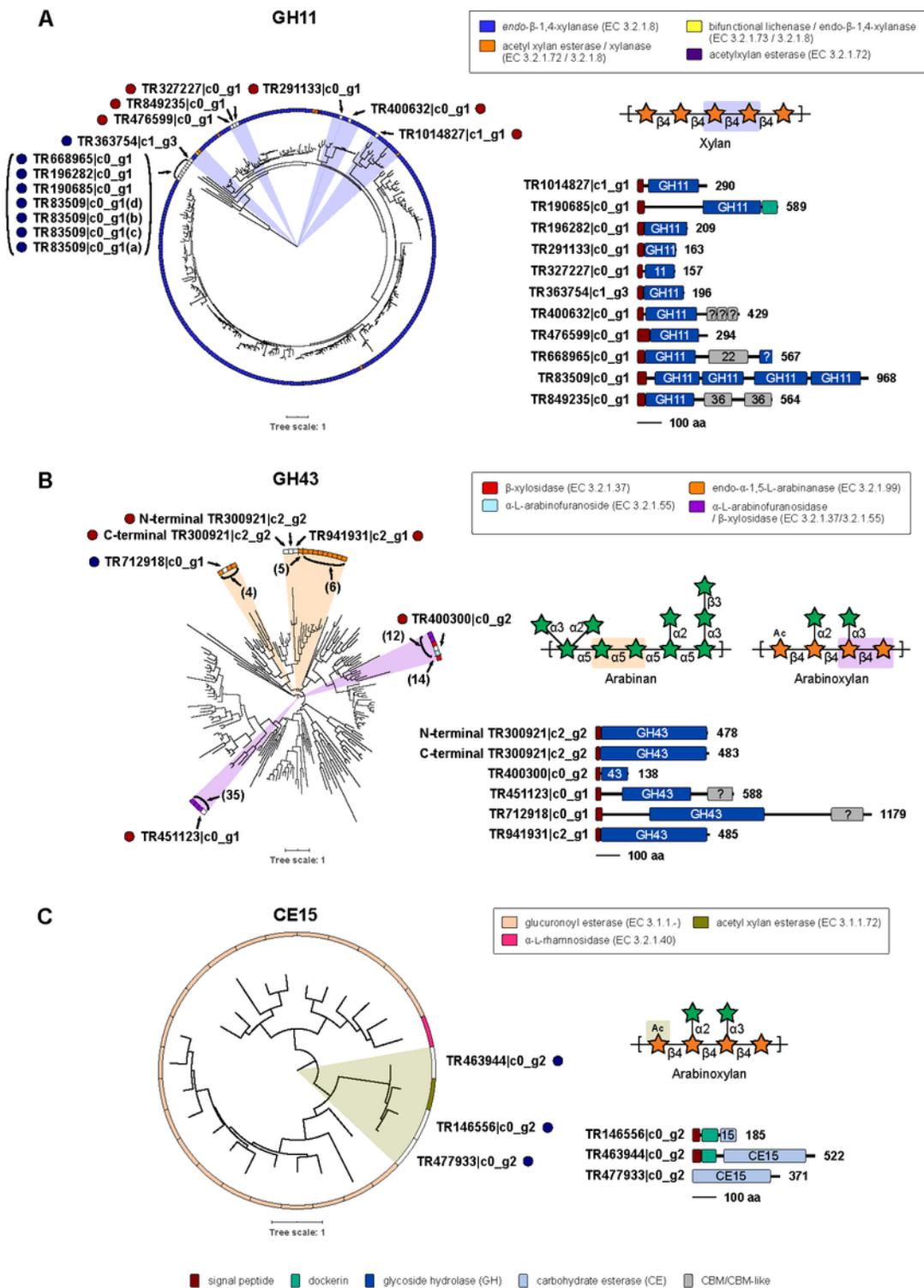


Figure 6

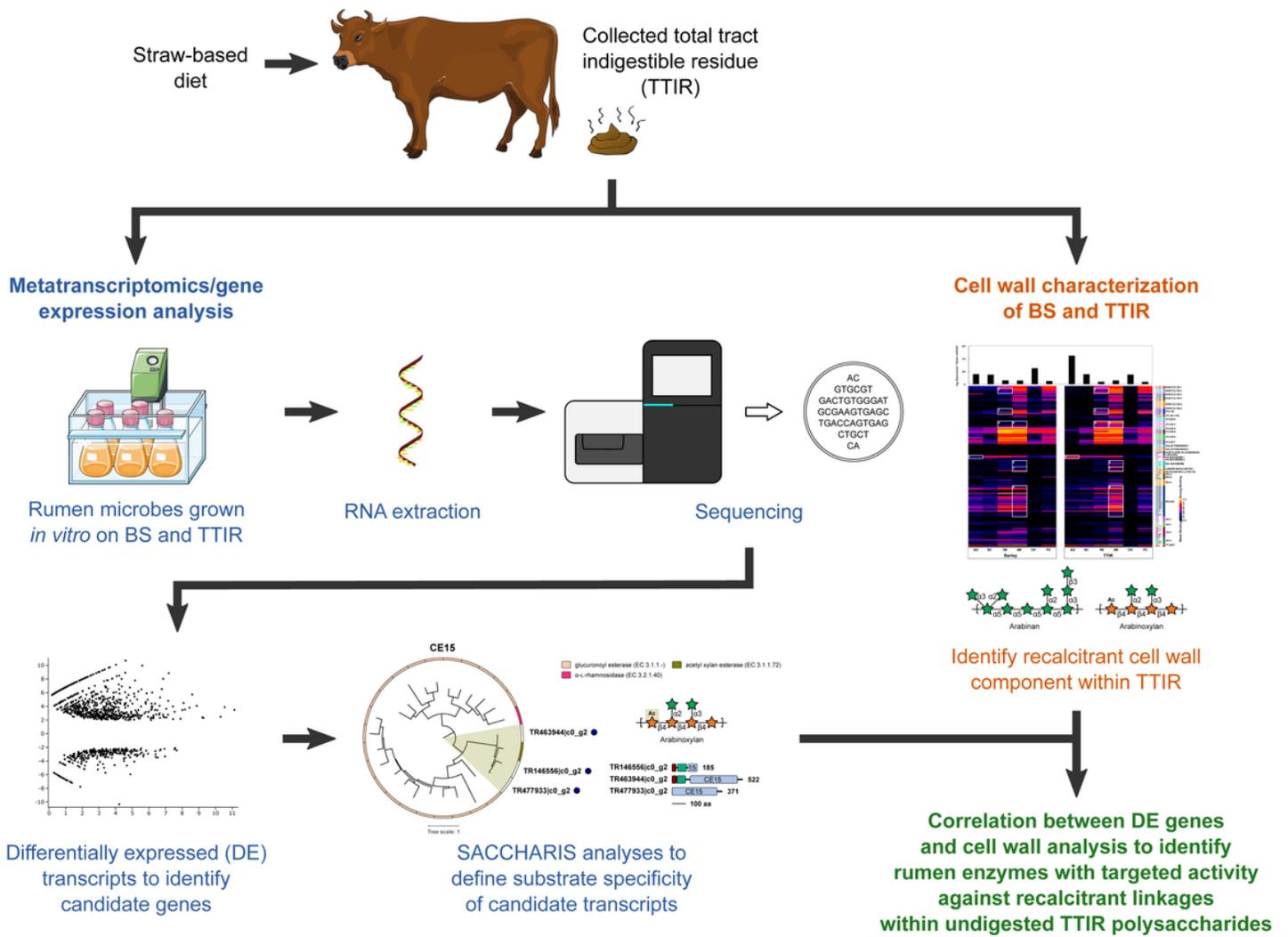
Relative individual DE transcript expression levels for GH families of interest. The contribution levels of DE individual transcripts for GH families of interest was calculated as the abundance (%) of the individual transcript relative to the expression levels of transcripts observed for those GH families listed, for each substrate (barley straw or TTIR). A heat map was generated for the relative total signal for each transcript and represented as % either barley straw (red) or TTIR (blue), and statistical significance was calculated by two-way ANOVA (\* p<0.0001).



**Figure 7**

Prediction of enzyme activities by CAZyme fingerprinting. Translated protein sequences annotated by dbCAN2 [19] were used as query sequence inputs for SACCHARIS [19] and embedded into phylogenetic trees of characterized GH11 (A), GH43 (B), and CE15 (C) enzymes. EC number and CAZy database annotated functions are colour-coded as indicated. Domain boundaries are based on predictions by dbCAN [19] and InterProScan [72], with all schematics to scale and colour-coded. The polypeptide length

for the open reading frame of each transcript is shown. Schematics of typical plant cell wall substrates, as identified by EC number, are shown with the targeted bond highlighted.



**Figure 8**

Here we have shown that integrating high-resolution analytical methods, such as cell wall structural characterization, *in vitro* batch culture, RNA-sequencing and DE analysis (Figure 8), can be successfully used to identify recalcitrant polymers and resistant linkages within feeds, as well as inform the biological processes involved in ruminal digestion of polysaccharides.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.pdf](#)
- [Additionalfile1B.pdf](#)
- [Additionalfile2.pdf](#)

- [Additionalfile3.xlsx](#)
- [Additionalfile4.xlsx](#)
- [Additionalfile5.xlsx](#)
- [Additionalfile6.docx](#)
- [Additionalfile7.xlsx](#)
- [Additionalfile8.pdf](#)