Do rumen Bacteroidetes utilize an alternative mechanism for cellulose degradation?

A.E. Naas[1], A.K. Mackenzie[1], J. Mravec[2], J. Schückel[2], W.G.T. Willats[2], V.G.H. Eijsink[1], P.B. Pope[1†]

1. Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, 1432 NORWAY.
2. Department of Plant Biology and Biotechnology, University of Copenhagen, Copenhagen, DK-1871 DENMARK

Running title: *Rumen bacteria harbour alternate cellulolytic systems*

Word count abstract: 141

Word count text: 2569

Length estimation: 3.98 pages

†Corresponding Author: Phillip B. Pope
Department of Chemistry, Biotechnology and Food Science
Norwegian University of Life Sciences
Post Office Box 5003
1432, Ås
Norway
Phone: +47 6496 6232
Email: phil.pope@nmbu.no

ABSTRACT

Uncultured and therefore uncharacterized Bacteroidetes lineages are ubiquitous in many natural ecosystems which specialize in lignocellulose degradation. However, their metabolic contribution remains mysterious as well-studied cultured Bacteroidetes have only been shown to degrade soluble polysaccharides within the human distal gut and herbivore rumen. We have interrogated a reconstructed genome from an uncultured Bacteroidetes phylotype that dominates a switchgrass-associated community within the cow rumen. Importantly, this characterization effort has revealed the first preliminary evidence for Polysaccharide Utilization Locus (PUL)-catalysed conversion of cellulose. Based on these findings we propose a further expansion of the PUL paradigm and the saccharolytic capacity of rumen Bacteroidetes species to include cellulose, the most abundant terrestrial polysaccharide on earth. Moreover, the perspective of a cellulolytic PUL lays the foundation for PULs to be considered as an alternative mechanism for cellulose degradation, next to cellulosomes and free enzyme systems.

KEYWORDS

Polysaccharide Utilization Loci / Cellulases / Metagenomics / Microbiome

INTRODUCTION: Uncultured Bacteroidetes lineages dominate many lignocellulose degrading communities. A comprehensive understanding of how plant biomass deconstruction occurs in nature has far reaching implications, related to mammalian health and nutrition as well as development of sustainable bio-based economies. Our current understanding is severely impeded by the inability to cultivate and thus examine the majority of microbes that perform the key metabolic processes of interest. For example, the rumen of herbivores represents one of nature's most proficient plant biomass degrading ecosystems, however it is controlled largely by uncharacterized microbes that belong to a limited number of frequently observed bacterial phyla (1). Degradation of the most abundant plant polysaccharide (cellulose) within ruminal ecosystems has for the greater part been attributed to the metabolic capabilities of species affiliated to the bacterial phyla Firmicutes and Fibrobacteres. These species produce one or more well-known cellulases that are structurally assembled on the cell-surface as a cellulosome or secreted as free enzymes (2). The ruminal Bacteroidetes represent another numerically dominating phylum, which is not associated with cellulose degradation but whose saccharolytic reputation is based on limited case studies of non-cellulolytic *Prevotella* rumen isolates (3) and renowned culturable human gut representatives such as *Bacteroides thetaiotaomicron* and *B. ovatus* (4). The saccharolytic machineries of gastrointestinal Bacteroidetes species have thus far been attributed to Polysaccharide Utilization Loci (PULs), gene-clusters that encode cell-envelope associated enzyme systems that enable the bacterium to respond to, bind and degrade specific glycans, and import released oligosaccharides (5).

The numerical predominance of uncultured Bacteroidetes species in lignocellulose degrading ecosystems (1, 6) and the observed abundance and diversity of PUL-encoded carbohydrate-active enzymes within Bacteroidetes genomes suggest that there is much to learn about the

contribution of these enzymatic complexes to polysaccharide metabolism. Here we propose an alternative hypothesis regarding cellulose degradation, which was generated by the biochemical characterization of a simplistic cellulase-encoding PUL previously annotated in a high-coverage uncultured Bacteroidetes phylotype (hereafter referred to as *AC2a*) inherent to the cow rumen microbiome (6, 7). The gene organisation of the *AC2a* PUL indicates a direct targeting towards cellulose, which is unique for PULs that have been described and characterised to date (**Fig. 1, Fig. S1**).

EXPERIMENTAL RATIONALE: Biochemical characterisation of a putative cellulolytic PUL encoded within the uncultured *AC2a* Bacteroidetes phylotype. The *AC2a* draft genome sequence (~76% complete) was one of 16 genomes previously binned using tetranucleotide signatures from a switchgrass-degrading metagenome recovered from the cow rumen (7). The high assembly coverage of the genome (284x coverage; third highest) indicated that *AC2a* is likely a numerically abundant organism in the rumen microbiome. Our own *de novo* predictions using Support Vector Machine classifiers (8) identified *AC2a* as a potential cellulolytic Bacteroidetes species, which challenges the current idea that Bacteroidetes drive only non-cellulosic metabolism in the rumen. *AC2a*'s cellulolytic capabilities was predicted to be dependent on a relatively simple eight gene PUL encoding two putative cellulases (glycoside hydrolase (GH) family: GH5 and GH9) and a cellobiose phosphorylase (GH94) (**Fig 1a**) (6). Sequence analysis and comparisons of gene-organisation with the model starch utilization system (Sus) of *B. thetaiotaomicron* led to identification of a SusC-like (TonB-dependent) outer membrane transporter, SusD-like and SusE-positioned lipoproteins that putatively bind to the substrate (9, 10), an inner membrane sugar transporter and an inner-membrane sensor (4) (**Fig. 1**). We predicted that the GH5 and GH9 could degrade cellulose to cellobiose, which would be transported to the periplasm via the SusC-

positioned transporter, where the well-known GH94 activity would generate monomeric-sugars for transport into the cytoplasm (**Fig. 1b**). To test this prediction, we have biochemically characterized the GH5 and GH9 enzymes and determined the functionality of the putative SusD-like and SusE-positioned glycan-binding lipoproteins.

Initial screens with chromogenic substrates showed that the GH5 and GH9 glycoside hydrolases in the *AC2a* PUL are active on β-(1,4) linked glucose units in amorphous cellulose and β-glucan (barley) (**Fig. S2**). The enzymes showed weak side activities on xyloglucan and xylans (**Fig. S2**) and were not active on β-(1,3) glucan (Pachyman). These observations are consistent with the subfamily four classification of the GH5 enzyme (11), which typically encompasses extracellular bacterial enzymes that exhibit one or more activities categorized as endoglucanase, xyloglucan-specific endoglucanase, xylanase and licheninase. Typical for endoglucanases, both GH5 and GH9 demonstrated higher activity on soluble cellulose and highly accessible natural β-glucan substrates than recalcitrant crystalline cellulose (**Table S1**). The *AC2a* PUL is distinct from barley β-glucan PULs characterised from *B. cellulosilyticus* (12) and *B.ovatus* (4) in that it does not contain a GH3 (β-(1-3)-glucosidase) or a GH16 (β-(1,3)-glucanase) (**Fig. S1c**). HPAEC-PAD analysis demonstrated that GH5 hydrolysis of filter paper produced dimer and trimer cellodextrins, whereas GH9 hydrolysis produced dimers and monomers (**Fig. 2a**). Interestingly, upon combining the two enzymes the filter paper was converted to dimers and monomers only (**Fig. 2a**), indicating synergism to produce cellobiose, which ultimately would be degraded by the periplasmic GH94 cellobiose phosphorylase. Further analysis of cellodextrin (DP2-6) hydrolysis revealed that the GH5 cannot degrade cellotriose or cellobiose, and produces only cellobiose from cellotetraose (**Fig. S3**). GH5 hydrolysis of cellopentaose and cellohexaose produced dimers and trimers. In contrast, the GH9 degraded cellotriose and produced cellobiose and a small

amount of glucose from DP4-6. This indicates that the two cellulases have different, complementary roles. The degradative effect of the *AC2a* PUL enzymes could easily be observed by monitoring the partial solubilisation of filter paper discs (**Fig 2a**).

Pull-down binding assays showed that SusD-like and SusE-positioned proteins from the *AC2a PUL* bind to crystalline cellulose (Fig. 2b), while also binding weakly to β-glucan (**Fig. S4**). To further visualize the ability of the SusD-like and SusE-positioned proteins to interact with plant cell walls, both proteins were used for indirect immunofluorescence-labelling of *Arabidopsis thaliana* cross-sections. The two proteins demonstrated clear binding to various sections of the plant cell walls, including xylem, phloem and cortical parenchyma, with the SusE-positioned protein giving weaker signals than the SusD-like protein (**Fig. 2c)**. CBM3a from *Clostridium thermocellum* (13) was included as a positive control, and showed binding to cellulose-rich secondary cell walls (SCW) in the xylem and adhered faces of adjacent pith parenchyma (PP) cells. The SusD-like protein bound to similar regions, whereas binding of the SusE-positioned protein appeared limited to the intercellular junctions of adjacent PP cells (**Fig. 2c)**.

THE HYPOTHESIS: PULs represent an alternative mechanism for cellulose degradation, next to cellulosomes and free enzyme systems. The insatiable interest in the human gastrointestinal microbiome has provided detailed accounts on the diversity and mechanisms of PULs that are central to plant polysaccharide degradation. However, this understanding has been limited by a reliance on well-known cultivated Bacteroidetes species, which represent a significant minority in many saccharolytic ecosystems. By specifically targeting uncultured microbiota resident in the cow rumen with approaches that go beyond predictive annotation, we reveal a possible alternative mechanism for microbial cellulose

degradation, which implies that rumen Bacteroidetes utilize PUL-based machinery, rather than (or in addition to) well-known mechanisms such as cellulosomes and free-enzyme systems. Broader genomic comparisons of the AC2a PUL with publicly-available metagenomes and Bacteroidetes genomes identified sequence homology and synteny with a partial metagenomic fragment derived from the Tammar wallaby foregut, a marsupial herbivore whose diet is rich in lignocellulose (14) (**Fig. S1a**). Partial synteny was observed with the "core" components of a well-characterised xyloglucan PUL that encodes both GH5 and GH9 representatives but which targets only xyloglucans, whilst lacking activity against any other hemicellulose or cellulose substrate (**Fig. S1a**)(15). Closer inspection of the proteins occurring in both these PULs revealed low sequence similarity (Fig. S1a) as well as different Pfam-predicted domain organisations for the GH5 (AC2a lacking the BACON domain at the N-terminus) and the SusD-like lipoprotein (AC2a: PF12771, BACOVA_02651: PF14322/PF07980). Whilst degrading activity for soluble cellulose-analogues has been described for several endoglucanases encoded within large hemicellulosic PULs, these enzymes are devoid of activity on recalcitrant cellulose and the PULs in question bear no resemblance to the AC2a PUL (**Fig. S1b**) (6, 14). Interestingly the SusD-like lipoprotein from the AC2a PUL exhibited very low sequence identity to two cellulose-binding SusD-like representatives we previously characterised from a hemicellulose-degrading PUL reconstructed from an uncultured phylotype (both exhibited less than 23% alignment coverage and 31% ID) (10). This would suggest that functional differences cannot necessarily be detected by the binding assays done in the present and past studies.

Collectively, these findings expand current perceptions regarding the overall saccharolytic capacity of rumen Bacteroidetes-affiliates, which so far have only been coupled to the degradation of non-cellulosic polysaccharides. Furthermore, it adds the Earth's most abundant

organic polymer to an already impressive catalogue of PUL target substrates including starch, alginate, various hemicelluloses and host-mucin glycans (3, 4, 15-17). To conclusively determine that *AC2a* is indeed capable of sustaining cell metabolism and growth on recalcitrant cellulosic substrates, knock-out mutagenesis studies on pure culture representatives are required. Whilst isolating deeply branched novel affiliates of the Bacteroidetes has proved extremely difficult in herbivore microbiomes, the ability to mine the *AC2a* genome for growth requirements provides a unique opportunity to reconstruct a custom enrichment media and isolation strategy. Similar metagenome-directed isolation approaches have ultimately proved successful for gut microbiomes in the past (18), and form the basis of our ongoing efforts.

MATERIALS AND METHODS

Gene annotation of the *AC2a* genome. The *AC2a* genome was previously reconstructed from metagenome sequencing data generated from the microbiota in the cow rumen (pH7.0) (7). Assembled and unprocessed DNA reads previously assigned to *AC2a* based on tetranucleotide frequencies were retrieved from http://portal.nersc.gov/project/jgimg/CowRumenRawData/submission/cow_rumen_genome_bins.tar.gz and annotated via the RAST server (19). The cellular localization of proteins was predicted using PSORTb 3.0 (20) and LipoP 1.0 (21).

Heterologous expression and purification of enzymes. Genes encoding signal peptide-free versions of *AC2a* GH5, GH9, SusD-like and SusE-positioned proteins were synthesized and cloned into the pNIC-CH expression vector by ligation independent cloning (LIC) using the primers listed in **Table S2** (22). Transformants were verified by sequencing. *Escherichia coli* BL21 harbouring the plasmids were pre-cultured for eight hours in Luria-Bertani Broth and

inoculated to 1 % in an overnight culture at 18 °C. Expression was induced by adding IPTG to a final concentration of 0.75 mM at OD_{600} 0.5-1.0, followed by incubation for 24 hours at 18 °C. Cells were harvested by centrifugation (5 000 rpm, 10 min) and resuspended in lysis buffer (100 mM Tris-HCl, pH 8.0, 500 mM NaCl, 5 mM imidazole, 0.1 mg/ml lysozyme) before 30 minutes incubation on ice. Cells were disrupted by pulsed sonication, and debris removed by centrifugation (8 000 g, 10 min) with the supernatant filtered using 0.45 and 0.22 μm syringe filters. Proteins were loaded onto 5 mL HisTrap HP Ni Sepharose columns (GE Healthcare) and eluted with a linear gradient of 100 mM Tris-HCl, pH 8.0, 500 mM NaCl, 500 mM imidazole. The eluted fractions were concentrated and the buffer changed to 100 mM Tris-HCl pH 8.0 using Sartorius Vivaspin concentrators with a 10 kDa cutoff. Further purification steps were performed using ion exchange chromatography (GH5, GH9 and SusD-like proteins) with a 5ml HiTrap DEAE FF column (GE Healthcare) and gel filtration (HiLoad Superdex 75, GE Healthcare) in 50 mM Tris-HCl with 200 mM NaCl (SusE). Proteins were concentrated and the buffer exchanged to 10mM Tris-HCl pH7.5. Protein purity was analyzed by sodium dodecyl sulfate polyacrylamide gel-electrophoresis, and protein concentration was estimated by measuring A_{280} and using the proteins' molar extinction coefficients.

Chromogenic substrates. AZCL-substrates (**Table S3**) partly dissolved in isopropanol (10mg/ml) were added to 135μL buffer (50 mM potassium phosphate, pH 7.5, or positive control's preferred pH). Plates were sealed with adhesive PCR plate seals (Thermo Scientific, AB-0558) and incubated with overhead rotation (~20rpm, room temperature) for one hour. Plates were spun down (4000 rpm, 10 minutes) and the absorption of the filtrate was measured against negative controls at 590 nm. Values reported are relative absorbance values calculated against absorbance values of the positive controls listed in **Table S3**.

Enzymatic assays. The optimum pH was determined to be approximately pH 6.6 for both enzymes, and 20 mM BisTris pH 6.6 was used for all enzyme assays. Enzyme activities were determined for carboxymethyl-cellulose (CMC) (Sigma-Aldrich), filter paper (Whatman no. 1), Avicel (Sigma-Aldrich) and barley β-glucan (Megazyme). CMC (1% w/v) and β-glucan (0.5% w/v) were incubated at 40°C, with 900 rpm horizontal shaking, with 25 nM and 10 nM GH5, respectively, for 10 minutes in a total volume of 500 µl. The reactions were stopped by adding an equal amount of DNS reagent, and the amount of reducing sugars relative to a glucose standard curve was determined using the DNS assay (23). A Unit of enzyme activity was defined as the amount of enzyme releasing one µmol of reducing sugars per minute. For GH9, the enzyme concentration was increased to 100 nM, and the incubation time was 15 minutes. For filter paper and Avicel, the conditions were 1% substrate (w/v), 100 nM GH5 or GH9, with an incubation time of 30 minutes. The reactions were stopped by boiling (5 min), before soluble cellodextrins were quantified by HPAEC-PAD as described below. A Unit of enzyme activity was defined as the amount of enzyme releasing one µmol of soluble products per minute. The time-course analysis of degradation of filter paper was performed using 5% (w/v) substrate and 3µM enzyme (GH5+GH9: 1.5 + 1.5µM enzyme). Soluble cellodextrin products were quantified against a standard curve of cellodextrins (DP1-3) by high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) using a Dionex ICS-3000 with a CarboPac PA1 column at 0.25 ml/min 0.1M NaOH. Oligosaccharides were eluted by a multi-step linear gradient going from 0.1M NaOH to 0.1M NaOH/0.1M NaOAc in 10 min, to 0.1M NaOH/0.18M NaOAc in 8 minutes, to 0.1M NaOH/0.3M NaOAc in 1 minute, and to 0.1M NaOH/1.0 M NaOAc in 1 minute, before column reconditioning by 0.1 M NaOH for 14 minutes. Visual assessment of the degradation of filter paper discs was performed using the same conditions as above in glass tubes in a

total volume of 1 mL, with 0.8 U/mL β-glucosidase (Megazyme) added to avoid potential cellobiose inhibition.

Binding assays. Filter paper (Whatman no. 1) milled to 0.5 mm size, Avicel (Sigma-Aldrich) and the insoluble fraction (room temperature) of Barley β-glucan (Megazyme), were washed twice in MES buffer (20 mM, pH 6.0), suspended to 6% w/v in a total volume of 200 µL along with 0.1 mg/ml protein and incubated at 40 °C with horizontal shaking (900 rpm). The substrate and bound protein were pelleted by centrifugation and the supernatant containing unbound protein (referred to as flowthrough) was carefully removed. The pellet was washed with 200 µl buffer for 15 minutes and the supernatant was again removed by centrifugation. To elute the proteins, the pellets were resuspended in 200 µl 8 M urea and boiled for 10 minutes (filter paper and Avicel), or incubated with 200 µl 2% SDS and incubated with shaking for 10 minutes (β-glucan). The flowthrough-, wash-, and elution fractions were analyzed by SDS-PAGE.

Binding to plant material was tested by probing transverse sections through *Arabidopsis thaliana* stems. Hand-cut sections through the stems of 4–5-week-old plants were labelled using a $His_{(6)}$-tag based three stage procedure essentially as previously described (24), in which binding was detected using a fluorescein isothiocyanate conjugated tertiary antibody. Cellulose-binding CBM3a from *Clostridium thermocellum* (13) was included as a positive control.

ACKNOWLEDGEMENTS

We are grateful for support from The Research Council of Norway's FRIPRO program (214042) and the European Commission Marie Curie International Incoming Fellowship

(awarded to PBP; PIIF-GA-2010-274303). AM was supported by a grant from the Norwegian Research Council (190965/S60).

REFERENCES

1. **Edwards JE, Mcewan NR, Travis AJ, Wallace RJ.** 2004. 16S rDNA library-based analysis of ruminal bacterial diversity. Antonie Van Leeuwenhoek. **86:**263-281.
2. **Flint HJ, Bayer EA, Rincon MT, Lamed R, White BA.** 2008. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. Nat. Rev. Microbiol. **6:**121-131.
3. **Dodd D, Moon YH, Swaminathan K, Mackie RI, Cann IK.** 2010. Transcriptomic analyses of xylan degradation by Prevotella bryantii and insights into energy acquisition by xylanolytic bacteroidetes. J. Biol. Chem. **285:**30261-30273.
4. **Martens EC, Lowe EC, Chiang H, Pudlo NA, Wu M, McNulty NP, Abbott DW, Henrissat B, Gilbert HJ, Bolam DN, Gordon JI.** 2011. Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. PLoS Biol **9:**e1001221.
5. **Martens EC, Koropatkin NM, Smith TJ, Gordon JI.** 2009. Complex glycan catabolism by the human gut microbiota: The bacteroidetes Sus-like paradigm. J Biol. Chem. **284:**24673-24677.
6. **Pope PB, Mackenzie AK, Gregor I, Smith W, Sundset MA, McHardy AC, Morrison M, Eijsink VGH.** 2012. Metagenomics of the Svalbard reindeer rumen microbiome reveals abundance of Polysaccharide Utilization Loci. PLoS ONE **7:**e38571.
7. **Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, Luo S, Clark DS, Chen F, Zhang T, Mackie RI, Pennacchio LA, Tringe SG, Visel A, Woyke T, Wang Z, Rubin EM.** 2011. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science **331:**463-467.
8. **Weimann A, Trukhina Y, Pope PB, Konietzny SG, McHardy AC.** 2013. De novo prediction of the genomic components and capabilities for microbial plant biomass degradation from (meta-)genomes. Biotechnol. Biofuels **6:**24.
9. **Cameron EA, Maynard MA, Smith CJ, Smith TJ, Koropatkin NM, Martens EC.** 2012. Multidomain Carbohydrate-binding Proteins Involved in Bacteroides thetaiotaomicron Starch Metabolism. J. Biol. Chem. **287:**34614-36425.

10. **Mackenzie AK, Pope PB, Pedersen HL, Gupta R, Morrison M, Willats WG, Eijsink VGH.** 2012. Two SusD-like proteins encoded within a polysaccharide utilization locus of an uncultured ruminant Bacteroidetes phylotype bind strongly to cellulose. Appl. Environ. Microbiol. **78:**5935-5937.
11. **Aspeborg H, Coutinho PM, Wang Y, Brumer H, Henrissat B.** 2012. Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). BMC Evol. Biol. **12:**186.
12. **McNulty NP, Wu M, Erickson AR, Pan C, Erickson BK, Martens EC, Pudlo NA, Muegge BD, Henrissat B, Hettich RL, Gordon JI.** 2013. Effects of diet on resource utilization by a model human gut microbiota containing Bacteroides cellulosilyticus WH2, a symbiont with an extensive glycobiome. PLoS Biol. **11:**e1001637.
13. **Blake AW, McCartney L, Flint JE, Bolam DN, Boraston AB, Gilbert HJ, Knox JP.** 2006. Understanding the biological rationale for the diversity of cellulose-directed carbohydrate-binding modules in prokaryotic enzymes. J. Biol. Chem. **281:**29321-29329.
14. **Pope PB, Denman SE, Jones M, Tringe SG, Barry K, Malfatti SA, McHardy AC, Cheng J-F, Hugenholtz P, McSweeney CS, Morrison M.** 2010. Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside hydrolase profiles different to other herbivores. Proc. Natl Acad. Sci. USA **107:**14793-14798.
15. **Larsbrink J, Rogers TE, Hemsworth GR, McKee LS, Tauzin AS, Spadiut O, Klinter S, Pudlo NA, Urs K, Koropatkin NM, Creagh AL, Haynes CA, Kelly AG, Cederholm SN, Davies GJ, Martens EC, Brumer H.** 2014. A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. Nature **506:**498-502.
16. **Hehemann JH, Kelly AG, Pudlo NA, Martens EC, Boraston AB.** 2012. Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. Proc. Natl Acad. Sci. USA **109:**19786-19791.
17. **Sonnenburg ED, Zheng H, Joglekar P, Higginbottom SK, Firbank SJ, Bolam DN, Sonnenburg JL.** 2010. Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. Cell **141:**1241-1252.
18. **Pope PB, Smith W, Denman SE, Tringe SG, Barry KH, P., McSweeney CS, McHardy AC, Morrison M.** 2011. Isolation of Succinivibrionaceae implicated in low methane emissions from Tammar wallabies. Science **333:**646-648.
19. **Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL,**

Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. **9:**75.

20. **Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FS.** 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics **26:**1608-1615.
21. **Juncker AS, Willenbrock H, Von Heijne G, Brunak S, Nielsen H, Krogh A.** 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Sci. **12:**1652-1662.
22. **Aslanidis C, de Jong PJ.** 1990. Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res. **18:**6069-6074.
23. **Miller GL.** 1959. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. Anal. Chem. **31:**426-428.
24. **McCartney L, Gilbert HJ, Bolam DN, Boraston AB, Knox JP.** 2004. Glycoside hydrolase carbohydrate-binding modules as molecular probes for the analysis of plant cell wall polymers. Anal. Biochem. **326:**49-54.

FIGURE LEGENDS

Figure 1. A putatively cellulolytic PUL recovered from the *AC2a* genome inherent to the cow rumen. a. Gene organization of a cellulase (GH5 and GH9) containing PUL identified in *AC2a*, which was selected for in-depth biochemical characterization. Gene identification numbers can be found in **Table S2**. **b.** A hypothetical model, based on predicted protein locations and analogies to the model starch utilization system (Sus) of *B. thetaiotaomicron* (5), depicts that glucans are bound and hydrolyzed via outer membrane lipoproteins and enzymes, whereas the generated cellobiose is transported to the periplasm, converted to glucose and imported to the cytoplasm for cellular metabolism (see text for more details). Proteins marked * were subjected to biochemical characterization. Proteins marked with a "tail" are predicted to be membrane-associated. The annotations of SusC (TBDR: *TonB*-dependent receptor) and SusD are based on significant hits using Pfam and these two proteins are therefore referred to as "SusC-like" and "SusD-like" in the main text. SusE is not recognized by Pfam and its annotation is thus based on position only; this protein is referred to as "SusE-positioned" (17).

Figure 2. Biochemical characterization of cellulases and binding proteins encoded within the *AC2a* PUL. a. Enzymatic activities of the GH5 and GH9 proteins determined by HPAEC-PAD analysis of products generated from filter paper (5% w/v, 3μM total enzyme concentration, pH 6.6). Error bars represent standard deviations between three replicates. The image to the right visualizes partial solubilization of filter paper discs after 6 days incubation (discs were diluted 3:1 prior to image capture). **b.** SDS-PAGE gel analysis of fractions from pull-down assays using cellulosic substrates. (FT) marks unbound protein from supernatant fractions collected after 1 hr incubation and centrifugation. (W) marks the wash fraction, containing protein washed off the substrate, while (E) marks eluted protein fractions where

protein was released from the polysaccharides by boiling in urea. (C) marks control lanes, where only the protein was loaded on the gel. (M) marks a molecular weight standard. **c.** Immunofluorescence labelling of *A. thaliana* cross sections using crystalline cellulose binding CBM3a (positive control) as well as the SusD-like and SusE-positioned proteins from the *AC2a* cellulose-active PUL**.** Fluorescence from anti-*his* in red indicates bound protein, while auto fluorescence, mainly in the interfascicular tissue, is colored blue. The SusD-like protein bound to cortical parenchyma (Cp), phloem (p), xylem (x) and pith parenchyma (Pp) adjacent cell walls. The SusE-positioned protein showed weaker binding than SusD-like protein, and particularly targeted the intercellular junctions in the xylem tissue. The positive control images were taken at lower gain due to high signals, while negative control images were taken at higher gain.

SUPPLEMENTAL MATERIAL

Fig. S1. Identification of partially homologous/syntenic PULs encoding GH5 and/or GH9 representatives. Genomic comparisons of the AC2a PUL were made with publicly-available metagenomes, Bacteroidetes genomes and previously characterised PULs that encode either a GH5 or GH9 representative. **a.** Sequence homology and synteny was observed with a partial metagenomic fragment (bracket indicates fosmid terminus) derived from the Tammar wallaby foregut and the "core" region (indicated by grey box) of a xyloglucan PUL from a human gut bacterium. Sequence identity is indicated as a percentage in parentheses. **b.** PULs that encode GH5 and/or GH9 representatives; these PULs all act on hemicellulose and there is no synteny. **c.** Example of previously characterised barley beta-glucan PULs (gene expression) which have been shown to encode a GH3 (β-(1-3)-glucosidase) or a GH16 (β-(1,3)-glucanase) and no GH5 and/or GH9 representatives. Text boxes to the left indicate the source of the PUL and the gene locus/accession number. Text boxes to the right provide a characterisation summary,

including a listing of substrates for which activity has been demonstrated or inferred: black print, biochemical characterisation; red print, inferred from gene expression only; green print, fosmid screening. Green ORFs indicate SusE/F-positioned and other hypothetical proteins. Black ORFs indicate putative transcriptional regulators. "Trans" indicates an inner membrane sugar transporter. Dashed lines indicate genes the products of which have been characterized. References from main text are indicated in parentheses.

[a] only CMC hydrolysis tested.

[b] only SusC and SusD gene expression analysed.

[c] A.K. Mackenzie, A.E. Naas, J. Mravec, S. Kracun, J. Schückel, J. Fangel, J.W. Agger, W.G.T. Willats, V.G.H. Eijsink and P.B. Pope, submitted for publication, 2014.

Fig. S2. Substrates specificity screening of *AC2a* outer membrane enzymes. Substrate specificities of the GH5 and GH9 enzymes were determined by AZCL substrate screening. Values are reported as relative absorbance calculated against the activity of the respective positive control enzymes at 1 U/ml (for the controls the absorbance value was set to 1.0). An overview of which enzymes were used as positive controls for the various substrates in the substrate specificity screens is provided in **Table S3**.

Fig. S3. Analysis of products generated by *AC2a* GH5 and GH9 from oligomeric substrates. The figure shows HPAEC-PAD chromatograms of product mixtures obtained from DP3-DP6 cellodextrins digested with GH5 or GH9 at pH 6.6. Enzyme assays were performed for 30 minutes at 40°C and enzymes were then inactivated by boiling for five minutes. Negative controls without added enzymes and containing 0.1 mg/ml cellodextrins are included.

Fig. S4. Binding of *AC2a* SusD-like and SusE-positioned proteins to barley β-glucan. SDS-PAGE gel analysis of fractions from pull-down assays. (FT) marks unbound protein from supernatant fractions collected after 1 hr incubation and centrifugation. (W) marks the wash fraction, containing protein in washed off the substrate, while (E) marks eluted protein fractions where protein was released from the polysaccharides by incubation with 2% SDS. (C) marks control lanes, where only the protein was loaded on the gel. (M) marks a molecular weight standard (BenchMark, Life Technologies).

Table S1. Activity of the AC2a PUL endoglucanases on various glycans. Previously calculated specificities for example endoglucanases characterized from rumen bacteria (Cel9B from *Fibrobacter succinogenes* and CelA from *Butyrivibrio fibrisolvens*) are listed.

[a] one unit of enzyme activity is the amount that produces 1 μmol of product per min.

[b] assays performed in triplicate.

[c] substrate specificity measurements: M. Qi, H-S. Jun, and C.W. Forsberg, Appl. Environ. Microbiol, **73**:6098-6105, 2007. Cel9B is a major cellulase secreted by the rumen bacterium *F. succinogenes* S85 accounting for approximately 32% of the total endoglucanase activity present in the nonsedimentable fraction (M. McGavin, and C.W. Forsberg, J. Bacteriol. **170**:2914–2922, 1988).

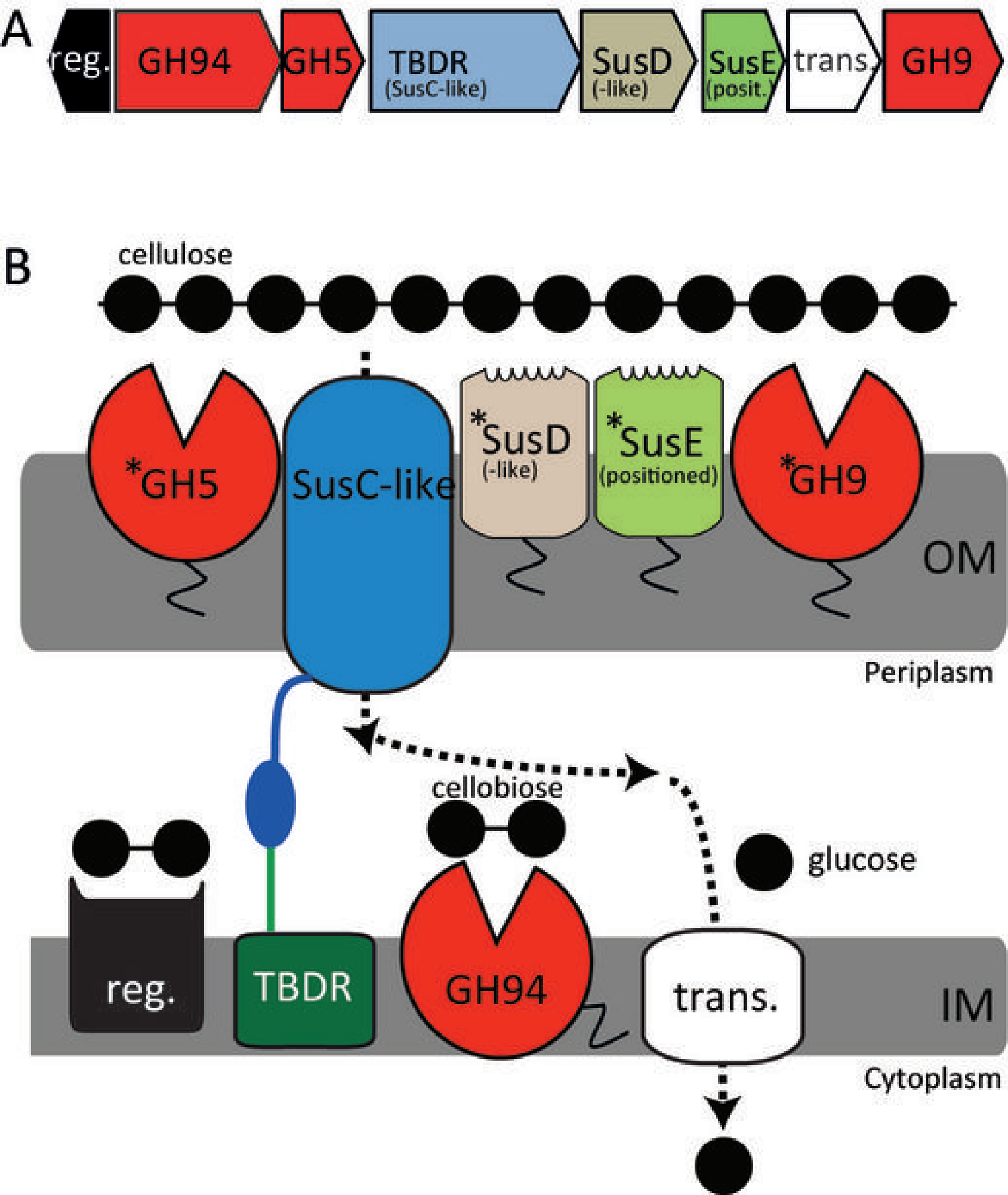
[d] substrate specificity measurements: G.P. Hazlewood, K. Davidson, J.I. Laurie, M.P. Romaniec, and H.J. Gilbert, J Gen Microbiol, **136**:2089-2097, 1990.

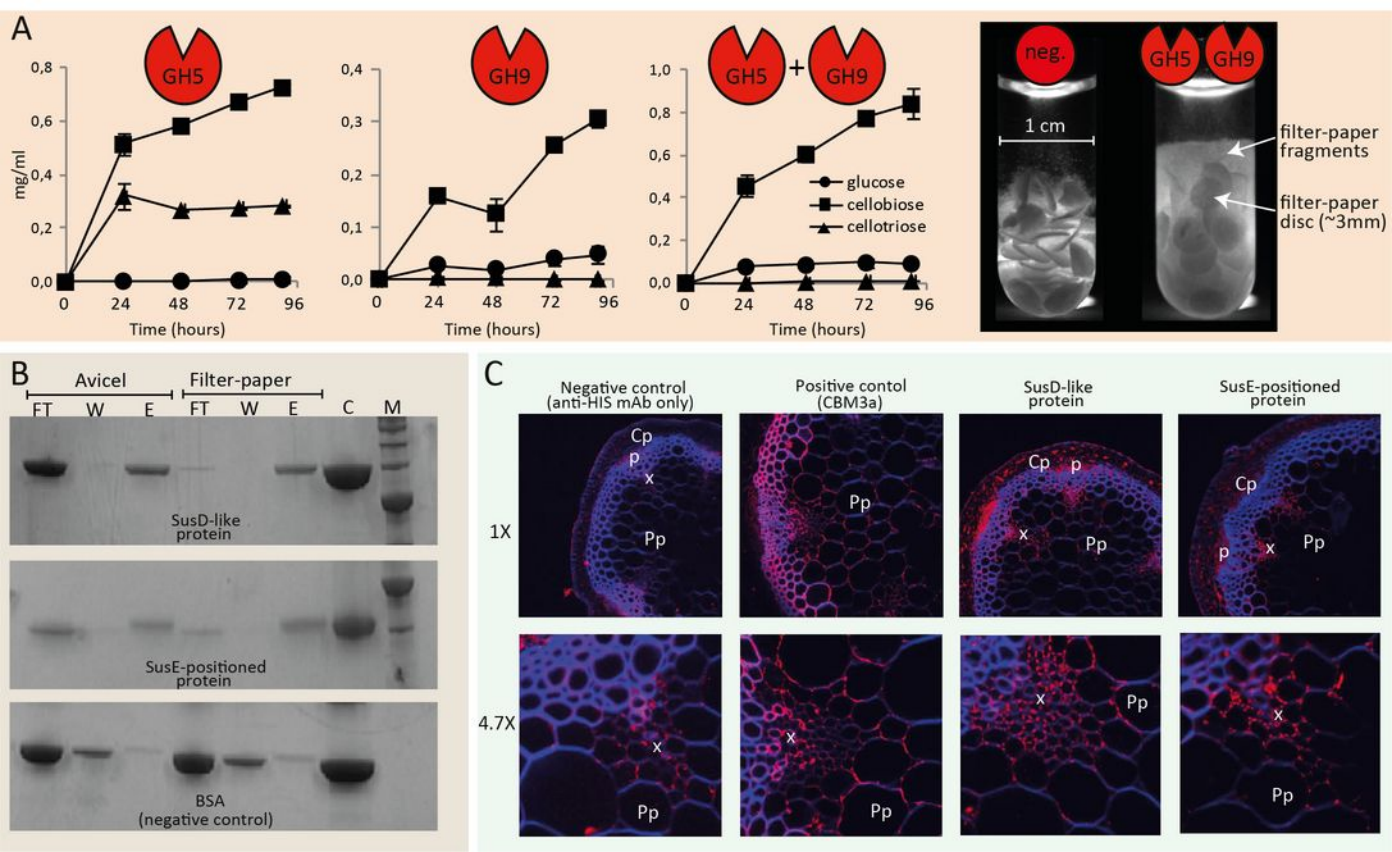
Table S2. Primers used to clone AC2a proteins.

[a] Corresponds to Feature ID from the RAST genome submission 171549.4: undefined AC2a (Taxonomy ID: 171549)

Table S3. Commercial enzymes used as positive controls for AZCL substrate specificity screening.

The preferred pH of the enzyme was used for the respective positive controls during the assay.


A
reg.
GH94
GH5
TBDR
(SusC-like)
SusD
(-like)
SusE
(posit.)
trans.
GH9
B
cellulose
*GH5
SusC-like
*SusD
(-like)
*SusE
(positioned)
*GH9
OM
Periplasm
cellobiose
glucose
reg.
TBDR
GH94
trans.
IM
Cytoplasm


A
GH5
GH9
GH5 + GH9
mg/ml
Time (hours)
glucose
cellobiose
cellotriose
neg.
1 cm
filter-paper fragments
filter-paper disc (~3mm)
B
Avicel
Filter-paper
FT W E FT W E C M
SusD-like protein
SusE-positioned protein
BSA (negative control)
C
Negative control (anti-HIS mAb only)
Positive contol (CBM3a)
SusD-like protein
SusE-positioned protein
1X
4.7X
Cp
p
x
Pp