

1 **Analysis of the bovine rumen microbiome reveals a diversity of Sus-like polysaccharide**  
2 **utilization loci from the bacterial phylum *Bacteroidetes***

3

4 Carly P. Rosewarne,<sup>a,\*</sup> Phillip B. Pope,<sup>b</sup> Jane L. Cheung,<sup>c</sup> and Mark Morrison <sup>c,†</sup>

5

6 <sup>a</sup> CSIRO Animal, Food and Health Sciences, Riverside Life Sciences Centre, North Ryde,

7 New South Wales, 2113, Australia

8 <sup>b</sup> Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life

9 Sciences, Ås, Norway

10 <sup>c</sup> CSIRO Animal, Food and Health Sciences, Queensland Biosciences Precinct, St Lucia,

11 Queensland, 4068, Australia

12

13 \* Corresponding author

14 Carly Rosewarne

15 CSIRO Animal, Food and Health Sciences

16 PO Box 52, North Ryde NSW, 1670

17 Australia

18 Phone: +61294905170

19 Fax: +61294908499

20 Email: [carly.rosewarne@csiro.au](mailto:carly.rosewarne@csiro.au)

21

22 † Current address

23 The University of Queensland Diamantina Institute

24 37 Kent Street, Woolloongabba QLD, 4102, Australia

25

1 **Abstract**

2 Several unique Sus-like polysaccharide utilization loci (PULs) were identified from bacteria  
3 resident in bovine rumen microbiomes through functional screening of a fosmid library. The  
4 loci were phylogenetically assigned to the genus *Prevotella* within the phylum *Bacteroidetes*.  
5 These findings were augmented by a bioinformatic re-evaluation of ruminal *Prevotella*  
6 genomes, revealing additional loci not previously reported in the literature. Analysis of  
7 Bacteroidales-affiliated genomes reconstructed from a bovine rumen metagenome in a  
8 previous study further expanded the diversity of Sus-like PULs resident in this microbiome.  
9 Our findings suggest that Sus-like systems represent an important mechanism for degradation  
10 of a range of plant-derived glycans in ruminants.

11

## 1 **Introduction**

2 Mammalian gut microbial communities are dominated by members of the *Firmicutes*  
3 and *Bacteroidetes* [14]. The ability of bacteria from these phyla to metabolize otherwise  
4 indigestible complex polysaccharides is important for gut function and host health. The  
5 starch utilization system (Sus) of *Bacteroides thetaiotamicron* (Phylum *Bacteroidetes*)  
6 represents a well-studied gene cluster from a human gut symbiont [19]. In this system, SusD  
7 enables cell surface starch binding, and in association with the outer membrane proteins SusE  
8 and SusF, the affinity of starch binding is enhanced [28]. The oligosaccharides released by  
9 the extracellular  $\alpha$ -amylase (SusG) are imported into the periplasmic space by SusC, which is  
10 a member of the TonB-dependent receptor family. Further maltooligosaccharide metabolism  
11 then occurs in the periplasmic space, mediated by SusA and SusB [7].

12 The review by Martens *et al* [18] highlights that Sus-like PULs can be found in the  
13 genomes of many sequenced *Bacteroidetes*. Characteristic features of these cell-envelope  
14 systems include proteins with homology to SusC and SusD co-located with a diversity of  
15 carbohydrate active enzymes, providing evidence for their role in targeting a broad range of  
16 substrates. While previous studies have focused mainly on the importance of Sus-like PULs  
17 in glycan catabolism in the human gut, emerging data suggests that these loci are involved in  
18 co-ordinating plant biomass degradation in herbivore microbiomes. For example, *Prevotella*  
19 *bryantii* B<sub>14</sub> upregulates expression of a xylanolytic Sus-like PUL when grown on wheat  
20 arabinoxylan [8]. Functional screening of metagenomic libraries derived from the Svalbard  
21 reindeer rumen [23] and Tammar wallaby foregut [24] revealed a diversity of carbohydrate  
22 active enzymes affiliated with the *Bacteroidetes* arranged in Sus-like PULs with activity  
23 against polysaccharide substrates *in vitro*. The results presented in this manuscript build on  
24 these observations by describing the hydrolytic potential encoded on Sus-like PULs harbored  
25 by members of the *Bacteroidetes* in the rumen of cattle. We first recovered a number of loci

1 through functional screening of an enrichment culture derived metagenomic fosmid library  
2 for clones expressing hydrolytic activity against carboxymethylcellulose. Additional loci  
3 were recovered bioinformatically from sequenced rumen isolates belonging to the genus  
4 *Prevotella*, and from Bacteroidales-affiliated genomes reconstructed from a bovine rumen  
5 metagenomic dataset.

6

## 7 **Materials and Methods**

8 Anaerobic culture of bovine rumen samples

9

10 Rumen grab samples were collected from six fistulated *Bos indicus* steers consuming Rhodes  
11 grass (*Chloris gayana*) in Rockhampton (QLD, Australia) using protocols approved by the  
12 Rendel Laboratory Animal Experimentation and Ethics Committee. In an anaerobic  
13 chamber, rumen samples were pooled before the digesta particles were separated from the  
14 rumen fluid by squeezing through two layers of cheesecloth. Particles were resuspended in  
15 an equal volume of sterile phosphate buffered saline and homogenised with an IKA  
16 UltraTurrax® TP18/10. Homogenate (0.01 vol. of 1:10 dilution) was added to 10 mL liquid  
17 anaerobic medium 1 [3] containing 10% (w/v) powdered Rhodes grass in triplicate, and  
18 incubated at 39 °C with overpressure of 100kPa H<sub>2</sub> until an OD<sub>600</sub> of 0.5 was reached  
19 (approximately 16 hours). The resulting cultures are therefore reflective of rumen microbes  
20 that can divide rapidly under these conditions. Culture material (comprised of grass substrate  
21 and microbial cells) was pelleted at 10,000 x g for 10 minutes at 4 °C. Genomic DNA was  
22 extracted from the enrichment cultures and aliquots of the homogenated inoculum using the  
23 NML method [27].

24

25 Microbiota community profiling and fosmid library construction and screening

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25

Amplicons of the 16S ribosomal RNA gene were prepared for 454 GS FLX Titanium sequencing as previously described (amplified from 20 ng of input DNA using primers 8F15B and 515R14AM; [1]). The data was analysed using QIIME 1.1.0 [5], based on sequence clustering at 97% similarity, taxonomic assignment using RDP and rarefaction to provide 4,000 sequences per sample. The remaining metagenomic DNA from the triplicate enrichment cultures was pooled and used to prepare a large insert fosmid library (pCC1FOS; Epicentre Corp.) using methods as previously described [27]. Clones capable of carboxymethylcellulose (CMC) hydrolysis were identified using a Congo red assay [23]. A total of  $\sim 2 \times 10^4$  clones were screened. Specific selected clones were subsequently screened using the same assay for their ability to degrade xylan (from birchwood), xyloglucan (from tamarind), lichenan (from Icelandic moss) and glucomannan (from konjac; low viscosity). Substrates were purchased from Sigma-Aldrich or Megazyme International, and were prepared for use according to manufacturer's instructions; before inclusion into LB agar plates at a concentration of 0.05% (w/v).

Functional annotation and analysis of sequences from Sus-like PULs

Fosmids from individual clones of interest were induced to high copy number according to the manufacturer's protocols. Fosmid DNA was extracted using the QIAprep spin miniprep kit (Qiagen), incorporating modifications for purification of large plasmids. The positional insertion of the cloned DNA within the vector was determined by Sanger sequencing with primers EPIFOS-F and EPIFOS-R (Epicentre Corp.). Equimolar amounts of each clone (quantified using the Invitrogen Quant-IT dsDNA BR kit and Qubit fluorometer) were pooled and a 3 kilobase paired-end library was prepared and subjected to pyrosequencing using 454

1 GS FLX chemistry (Macrogen Inc., Korea). The scaffolds from the resulting data were  
2 assembled using Newbler v2.6 and annotated using IMG/M-ER [16]. ORFs in reconstructed  
3 genomes from a published metagenomic survey of the bovine rumen [11] (genomes AC2a,  
4 AGa, AJ, AH and AQ; available at  
5 <http://portal.nersc.gov/project/jgim/CowRumenRawData/submission/>) were identified using  
6 MetaGeneMark [32]. ORFs in the genome of *Prevotella bryantii* B14 and *Prevotella*  
7 *ruminicola* 23 were identified in a previous study [26]. Carbohydrate active enzymes  
8 (CAZymes) and carbohydrate binding modules (CBMs) were identified in protein coding  
9 sequences from these various sources using dbCAN [31], a web resource that implements  
10 hidden Markov Models (HMMs) for automated signature domain-based annotations  
11 representative of each individual family. Sus-like PULs were defined as operons encoding  
12 proteins with Pfam domains [25] belonging to the TonB-dependent receptor (PF00593) and  
13 SusD-like families (PF12771, PF14721, or PF14322), collocated with proteins encoding  
14 known CAZyme domains. For both the dbCAN and Pfam HMM assignments, a minimum e-  
15 value cutoff of  $1 \times 10^{-5}$  was used. All putative Sus-like PULs that were recovered from the *in*  
16 *silico* analysis have been described in this manuscript.

17

18 Nucleotide accession numbers

19

20 Annotated sequences from five fosmid scaffolds have been deposited in GenBank under  
21 accession numbers JX424616-28.

22

## 23 **Results and Discussion**

24

1 Bacterial composition of fiber-adherent bovine rumen microbiome and cultures enriched on  
2 Rhodes grass  
3  
4 Amplicon pyrosequencing of 16S rRNA genes revealed differences in community structure  
5 resulting from *in vitro* cultivation of the fiber-adherent bovine rumen microbiome. The  
6 material used to inoculate the enrichments was comprised predominantly of sequences from  
7 the phyla *Bacteroidetes* (47.2%) and *Firmicutes* (42.3%). The *Bacteroidetes* fraction was  
8 dominated by sequences from the order Bacteroidales (71% of *Bacteroidetes*; 33.5% of total),  
9 while the *Firmicutes* fraction was dominated by sequences from the order Clostridiales (91%  
10 of *Firmicutes*; 38.6% of total). Much of this diversity could not be classified at the genus  
11 level. These results are comparable to other studies that have profiled the fiber-adherent  
12 fraction of the rumen microbiome from cattle consuming high-forage diets [3, 11, 22]. The  
13 cultures were dominated by representatives of recognised genera, including *Prevotella*  
14 (31.4%), *Selenomonas* (19.1%), *Psuedobutyrvibrio* (11.1%), *Streptococcus* (8.4%), and  
15 *Fibrobacter* (6.6%). These organisms have been enriched in this experiment because they  
16 are more amenable to *in vitro* growth under the specific culture conditions used. Sequences  
17 affiliated to the genus *Prevotella* were distributed into 16 operational taxonomic units  
18 (OTUs). Ten of these OTUs contained over 95% of the total *Prevotella* sequences in the  
19 enrichment dataset. Representative sequences from these 10 OTUs were compared to data  
20 from cultured isolates in Ribosomal Database Project Release 10.32 ([6]; Table S1). This  
21 indicated that the two predominant *Prevotella* OTUs (approximately 55% of total *Prevotella*  
22 sequences) were closely related to *P. ruminicola* 23 and *P. bryantii* B<sub>14</sub>, two rumen species  
23 for which genome sequences are available [26]. The remaining OTUs were more similar to  
24 unsequenced isolates, suggesting that metagenomic DNA extracted from the cultures contains  
25 material from both sequenced and “novel” *Prevotella* strains.

1

2 Recovery of Sus-like PULs from fosmid library sequencing and the genomes of ruminal  
3 *Prevotella* isolates

4

5 Metagenomic DNA extracted from the enrichment cultures was pooled and used to construct  
6 a large insert fosmid library. A total of 142 clones expressing hydrolytic activity towards  
7 carboxymethylcellulose were recovered from a screen of  $2 \times 10^4$  clones. It is probable that  
8 additional loci capable of CMC hydrolysis are present in the enrichment cultures but were not  
9 recovered due to biases associated with functional screening of fosmid clones in *E. coli*.  
10 Pooled DNA from each of the fosmid clones was sequenced and assembled into scaffolds.  
11 Mapping of insertion sites (based on Sanger sequencing with vector specific primers)  
12 indicated that several scaffolds were assembled from multiple fosmid clones (data not  
13 shown). The closest match to each scaffold larger than 10 kB (n=51) was determined using  
14 Standard Nucleotide BLAST (BlastN) of the NCBI nr database [12]. Approximately 41% of  
15 scaffolds were most closely affiliated to the genus *Prevotella* (n=21), followed by  
16 *Fibrobacter* (29%; n=15) and *Streptococcus* (6%; n=3). The *Fibrobacter* derived scaffolds  
17 have a high degree of similarity to regions within the genome of *F. succinogenes* S85 [29]  
18 and were not analysed further (data not shown).

19

20 A total of ten scaffolds encoding Sus-like PULs were recovered from the fosmid library.  
21 Each PUL encodes a putative SusC-like protein exhibiting homology to the TonB-dependent  
22 receptor family (PF07715) and a protein with homology to a SusD-like family (PF12771,  
23 PF14721, or PF14322). PhylopythiaS [21] was used to establish that each of these scaffolds  
24 was derived from a *Prevotella* genome. Insertion site mapping identified five scaffolds  
25 (designated Sc00026, 28, 33, 44 and 66) where the complete Sus-like PUL region could be



1 accurately assigned to at least one fosmid clone (Fig. S1). For the remaining five scaffolds,  
2 both sites of insertion into the vector could not be mapped to an individual clone and as a  
3 result were excluded from further analysis. The genomic architecture of these loci is  
4 described in more detail in Fig. S2.

5

6 Concomitant with the diversity of predicted carbohydrate active enzymes (CAZymes [31];  
7 Fig. 1), functional activity screening (Table 1) indicated that the five Sus-like PULs selected  
8 for further analysis have different substrate ranges. The substrates for this screen were  
9 chosen based on the CAZyme profiles of the fosmid clones. The Sc00026 PUL targets both  
10 carboxymethylcellulose and glucomannans. This scaffold encodes a 16S rRNA gene with  
11 99% sequence identity to *P. bryantii* B<sub>14</sub>. Comparison of the PUL-encoding contig from  
12 Sc00026 to the genome of *P. bryantii* B<sub>14</sub> revealed a high level of similarity to a region of  
13 B<sub>14</sub> previously shown to encode  $\beta$ -1,4-endoglucanase (GH5) and mannanase (GH26)  
14 activities ([10]; Genbank accession U96771 and Table S2). Association of the B<sub>14</sub> glycoside  
15 hydrolases with a Sus-like PUL has not previously been reported, although the  $\beta$ -1,4-  
16 endoglucanase activity has been extensively studied. The B<sub>14</sub> locus is unusual in that two  $\beta$ -  
17 1,4-endoglucanases with different molecular weights are transcribed from two adjacent open  
18 reading frames with a -1 frameshift [10, 20]. The Sc00026 PUL appears to have a similar  
19 structure, with the ORFs shown on the right of this locus in Figure 1 potentially also  
20 encoding two endoglucanases that share a common N-terminal domain.

21

22 The Sus-like PUL encoded by Sc00066 only expresses activity against  
23 carboxymethylcellulose (Table 1). This locus encodes two proteins with GH5 domains and  
24 one with a GH94 domain, indicative of cellobiosidase and cellobiose phosphorylase activities  
25 respectively. Accordingly, GH5a from Sc00066 is homologous (BlastX [12]; 75% identity

1 over 255 residues) to M40-2 (ACA61171), a protein cloned from buffalo rumen with  
2 degradative activity against p-nitrophenyl-D-cellobioside [9].

3

4 The PULs encoded on scaffolds Sc00028, 33 and 44 are capable of targeting  
5 carboxymethylcellulose and xyloglucans; while the Sc00028 and Sc00044 PULs are also able  
6 to degrade lichenan. According to the classification scheme proposed by Aspeborg *et al* [2],  
7 the GH5 enzymes encoded by these PULs belong to subfamily 4. Members of this subfamily  
8 with activities against xyloglucan and lichenan have been reported previously [2]. There are  
9 two Sus-like PULs with GH5 family proteins in the genome of *P. ruminicola* 23 (Fig. 1).  
10 Comparison of this genome to the Sus-like PUL-encoding contigs of Sc00028, 33, 44 and 66  
11 using BlastN [12] revealed 50-90% coverage with 73-85% sequence identity (Table S2).  
12 The PUL encoded by genes PRU\_2514-19 is most similar to that found on Sc00033, while  
13 the locus at PRU\_2222-32 includes three SusC and SusD-like proteins, a structure not found  
14 in any of the fosmid clones. A PUL with CMCase activity that encodes multiple Sus-like  
15 proteins was cloned from Svalbard reindeer rumen in a previous study [23]. Functional  
16 characterisation of the two SusD-like proteins in this PUL indicated that different  
17 mechanisms are used to interact with various forms of cellulose, and that they are able to bind  
18 to distinct features in the plant cell wall microstructure [15]. An increasing body of evidence  
19 suggests that Sus-like PULs may be capable of cellulose degradation in herbivores, although  
20 additional targeted experiments are required in order to confirm this hypothesis.

21

22 Sus-like PULs recovered from a published bovine rumen metagenome

23

24 As part of their analysis into the structural and functional diversity of the fiber-adherent  
25 bovine rumen microbiome, Hess *et al* [11] describe reconstruction of 15 bacterial genomes

1 from a substantial metagenomic dataset. Five of these reconstructed genomes (AC2a, AH,  
2 AJ, AGa and AQ) are from deeply branching members of the order Bacteroidales. Searches  
3 of these genomes were undertaken to recover additional Sus-like PULs. Five of these loci  
4 have been included in Fig. 1. Three Sus-like PULs from the genome of AC2a are shown - the  
5 cellulase encoding PUL (AC2a\_PUL1) has been described previously [11, 23]. The  
6 remaining two PULs from AC2a are predicted to target pectin (AC2a\_PUL2) and  $\beta$ -glucan  
7 (AC2a\_PUL3). The draft genomes of AJ and AH contain PULs with two presumptive GH10  
8 endoxylanases, suggesting a role in the degradation of hemicellulose similar to that found in  
9 *P. bryantii* B<sub>14</sub> and other related bacteria [8].

10

11 A further 36 Sus-like PULs were uncovered in *P. bryantii* B<sub>14</sub>, *P. ruminicola* 23 and three of  
12 the five reconstructed rumen Bacteroidales genomes (AC2a, AH and AJ). Several additional  
13 SusC and SusD homologs were annotated in the genome sequences, but were not associated  
14 with recognised carbohydrate active enzymes in the dbCAN database (data not shown). The  
15 putative PULs (Table S3) are predicted to target a diverse array of glycans. This expands the  
16 potential substrate range of these systems in the bovine rumen to include hemicelluloses,  
17 starch, fructans, glucans, pectins, chitin, alginate and various oligosaccharides. Several loci  
18 described in Table S3 (e.g. PBR\_0326 to PBR\_0345, PBR\_0377 to PBR\_0398, PRU\_2666 to  
19 PRU\_2716 and AJ\_902 to AJ\_918) are predicted to encode a large number of ORFs and  
20 include multiple Sus-like systems. Clustering of polysaccharide degradation genes is a  
21 strategy used by many bacteria to co-ordinate enzyme production and metabolic activities  
22 [13, 17]. Additional studies are required to confirm substrate specificity and to determine  
23 how the Sus-like PULs described in this study are regulated and expressed in response to the  
24 presence of different complex carbohydrates.

25

## 1 **Conclusion**

2

3 This paper describes more than 50 distinct Sus-like PULs from the bovine rumen  
4 microbiome, the most comprehensive collection described in herbivores to date. Each locus  
5 has been phylogenetically assigned to the predominant gut bacterial phylum *Bacteroidetes*.  
6 In *Prevotella ruminicola* 23, genes that were putatively associated with Sus-like PULs  
7 represent approximately 10% of the total protein coding sequences in the complete genome.  
8 Association of a high proportion of the total genomic CAZyme repertoire with Sus-like genes  
9 has been shown previously for human gut *Bacteroides* isolates [17]. For many members of  
10 the *Bacteroidetes*, Sus-like PULs represent an important mechanism for metabolism of a  
11 diverse range of plant and host-derived polysaccharides. The diversity found within loci as  
12 described in this manuscript and others [23, 24, 30], in concert with previously observed  
13 differences in the substrate binding properties of SusD-like proteins [15] suggests that there is  
14 much to learn about the contribution of these cell-envelope associated enzymatic complexes  
15 to plant biomass degradation in herbivores. This has implications for improving ruminant  
16 feed efficiency and discovery of novel systems for deconstruction of plant-based biofuel  
17 substrates.

18

## 19 **Acknowledgements**

20

21 This research was supported by funds provided to Meat and Livestock Australia project  
22 number B.CCH.1005, as part of the Reducing Emissions from Livestock Research Program,  
23 and by a Marie Curie International Incoming Fellowship from the European Commission  
24 (awarded to PBP; PIIF-GA-2010-274303). We are grateful to Honglei Chen for preparation

1 of amplicons for 454 pyrosequencing, to Ivan Gregor and Alice McHardy for assistance with  
2 PhyloPythiaS, and to Nigel Tomkins for assistance with sample collection.

3

#### 4 **References**

5

- 6 1. Aguirre de Carcer D, Denman SE, McSweeney CS, Morrison M (2011) Strategy for  
7 modular tagged high-throughput amplicon sequencing. *Appl Environ Microbiol* 77: 6310-  
8 6312
- 9 2. Aspeborg H, Coutinho PM, Wang Y, Brumer H 3<sup>rd</sup>, Henrissat B (2012) Evolution,  
10 substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5).  
11 *BMC Evol Biol* 12:186
- 12 3. Balch WE, Fox GE, Magrum LJ, Woese CR, Wolfe RS (1979) Methanogens: reevaluation  
13 of a unique biological group. *Microbiol Rev* 43: 260-296
- 14 4. Brulc JM, Antonopoulos DA, Berg Miller ME, Wilson MK, Yannarell AC, Dinsdale EA  
15 *et al* (2009) Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome  
16 reveals forage specific glycoside hydrolases. *Proc Nat Acad Sci USA* 106: 1948-1953
- 17 5. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al*  
18 (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Meth* 7:  
19 335-336
- 20 6. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ *et al* (2009) The Ribosomal  
21 Database Project: improved alignments and new tools for rRNA analysis. *Nucl Acids Res* 37:  
22 D141-D145

- 1 7. D'Elia JN, Salyers AA (1996) Contribution of a neopullulanase, a pullulanase, and an  
2 alpha-glucosidase to growth of *Bacteroides thetaiotaomicron* on starch. J Bacteriol 178:  
3 7173-7179
- 4 8. Dodd D, Moon Y-H, Swaminathan K, Mackie RI, Cann IKO (2010) Transcriptomic  
5 analyses of xylan degradation by *Prevotella bryantii* and insights into energy acquisition by  
6 xylanolytic Bacteroidetes. J Biol Chem 285: 30261-30273
- 7 9. Duan CJ, Xian L, Zhao GC, Feng Y, Pang H, Bai XL *et al* (2009) Isolation and partial  
8 characterisation of novel genes encoding acidic cellulases from metagenomes of buffalo  
9 rumens. J Appl Microbiol 107: 245-56
- 10 10. Gardner RG, Wells JE, Fields MW, Wilson DB, Russell JB (1997) A *Prevotella*  
11 *ruminicola* B<sub>14</sub> operon encoding extracellular polysaccharide hydrolases. Curr Microbiol 35:  
12 274-277
- 13 11. Hess M, Sczyrba A, Egan R, Kim T-W, Chokhawala H, Schroth G *et al* (2011)  
14 Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science  
15 331: 463-467
- 16 12. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis and Madden TL  
17 (2008) NCBI BLAST: a better web interface. Nucl Acids Res 36: W5-9
- 18 13. Kelly W, Leahy SC, Altermann A, Yeoman CJ, Dunne JC, Kong Z *et al* (2010) The  
19 glycobiome of the rumen bacterium *Butyrivibrio proteoclasticus* B316<sup>T</sup> highlights adaptation  
20 to a polysaccharide-rich environment. PLoS One 5: e11942
- 21 14. Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS *et al* (2008)  
22 Evolution of mammals and their gut microbes. Science 320: 1647-1651
- 23 15. Mackenzie AK, Pope PB, Pedersen HL, Gupta R, Morrison M, Willats WGT, Eijsink  
24 VGH (2012) Two SusD-like proteins encoded within a polysaccharide utilization locus of an

1 uncultured ruminant *Bacteroidetes* bind strongly to cellulose. Appl Environ Microbiol 78:  
2 5935-5937

3 16. Markowitz VM, Chen IMA, Chu K, Szeto E, Palaniappan K, Grechkin Y *et al* (2012)  
4 IMG/M: the integrated metagenome data management and comparative analysis system.  
5 Nucl Acids Res 40: D123-9

6 17. Martens EC, Lowe EC, Chiang H, Pudlo NA, Wu M, McNulty NP *et al* (2011)  
7 Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts.  
8 PLoS Biol 9: e1001221

9 18. Martens EC, Koropatkin NM, Smith TJ, Gordon JI (2009) Complex glycan  
10 catabolism by the human gut microbiota: the *Bacteroidetes* Sus-like paradigm. J Biol Chem  
11 284: 24673-24677

12 19. Martens EC, Chiang HC, Gordon JI (2008) Mucosal glycan foraging enhances fitness  
13 and transmission of a saccharolytic human gut bacterial symbiont. Cell Host & Microbe 4:  
14 447-457

15 20. Matsushita O, Russell JB, Wilson DB (1991) A *Bacteroides ruminicola* 1,4- $\beta$ -D-  
16 endoglucanase is encoded in two reading frames. J Bacteriol 173: 6969-6926

17 21. Patil KR, Haider P, Pope PB, Turnbaugh PJ, Morrison M, Scheffer T, McHardy AC  
18 (2011) Taxonomic metagenome sequence assignment with structured output models. Nat  
19 Meth 8: 191-192

20 22. Pitta D, Pinchak W, Dowd S, Osterstock J, Gontcharova V, Youn E *et al* (2010)  
21 Rumen bacterial diversity dynamics associated with changing from bermudagrass hay to  
22 grazed winter wheat diets. Microb Ecol 59: 511-522

23 23. Pope PB, Mackenzie AK, Gregor I, Smith W, Sundset MA, McHardy AC *et al* (2012)  
24 Metagenomics of the Svalbard reindeer rumen microbiome reveals abundance of  
25 polysaccharide utilization loci. PLoS One 7: e38571

- 1 24. Pope PB, Denman SE, Jones M, Tringe SG, Barry K, Malfatti SA *et al* (2010)  
2 Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside hydrolase  
3 profiles different from other herbivores. *Proc Nat Acad Sci USA* 107: 14793-14798
- 4 25. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C *et al* (2012) The  
5 Pfam protein families database. *Nucl Acids Res* 40: D290-301
- 6 26. Purushe J, Fouts D, Morrison M, White B, Mackie RI, Coutinho P *et al* (2010)  
7 Comparative genome analysis of *Prevotella ruminicola* and *Prevotella bryantii*: insights into  
8 their environmental niche. *Microb Ecol* 60: 721-729
- 9 27. Rosewarne CP, Pope PB, Denman SE, McSweeney CS, O’Cuiv P, Morrison M.  
10 (2011) High-yield and phylogenetically robust methods of DNA recovery for analysis of  
11 microbial biofilms adherent to plant biomass in the herbivore gut. *Microb Ecol* 61: 448-454
- 12 28. Shipman JA, Berleman JE, Salyers AA (2000) Characterization of four outer  
13 membrane proteins involved in binding starch to the cell surface of *Bacteroides*  
14 *thetaiotaomicron*. *J Bacteriol* 182: 5365-5372
- 15 29. Suen G, Weimer PJ, Stevenson DM, Aylward FO, Boyum J, Deneke J *et al* (2011)  
16 The complete genome sequence of *Fibrobacter succinogenes* S85 reveals a cellulolytic and  
17 metabolic specialist. *PLoS One* 6: e18814
- 18 30. Wang L, Hatem A, Catalyurek UV, Morrison M, Zu Y (2013) Metagenomic insights  
19 into the carbohydrate active enzymes carried by the microorganisms adhering to solid  
20 digesta in the rumen of cows. *PLoS One* 8: e78507
- 21 31. Yin Y, Mao X, Yang JC, Chen X, Mao F, Xu Y (2012) dbCAN: a web resource for  
22 automated carbohydrate-active enzyme annotation. *Nucl Acids Res* 40: W445-51
- 23 32. Zhu W, Lomsadze A, Borodovsky M (2010) *Ab initio* gene identification in  
24 metagenomic sequences. *Nucl Acids Res* 38: e132

25



1

2

	Sc00026 (H36 A20)	Sc00028 (H39 N13)	Sc00033 (H6 M11)	Sc00044 (H30 F13)	Sc00066 (H13 J8)	Control (no fosmid)
Carboxymethylcellulose	+	+	+	+	+	-
Xylan	-	-	-	-	-	-
Xyloglucan	-	+	+	+	-	-
Lichenan	-	+	-	+	-	-
Glucomannan	+	-	-	-	-	-

3

4

5 **Table 1.** Activity screening of fosmid clones on selected polysaccharide substrates. Clones  
6 encompassing the entire Sus-like PUL (as shown in Figure S1) were chosen for this assay as  
7 indicated in brackets below the scaffold ID. Substrates were incorporated into LB agar at a  
8 concentration of 0.05% (w/v). Zones of clearing were visualised using a Congo red assay  
9 and recorded as a positive or negative result. The *E. coli* EPI300<sup>TM</sup>-T1<sup>R</sup> strain from the  
10 fosmid library kit was included as a negative control

11

1 **Figure 1.** Gene organization of selected PULs from bacteria resident in the bovine rumen.  
2 Loci were identified from functional screening of a fosmid library derived from anaerobic  
3 cultures enriched on Rhodes grass (this study); in the genome of *Prevotella ruminicola* 23  
4 [26]; or from re-analysis of draft genomes reconstructed from a metagenomic survey [11].  
5 Green genes represent putative outer membrane proteins of unknown function; black genes  
6 encode putative response regulators; white genes encode putative transmembrane proteins;  
7 TonB indicates members of the TonB dependent receptor family of proteins that are predicted  
8 to transport solutes and macromolecules. GH: glycoside hydrolase (red); CBM: carbohydrate  
9 binding module; CE: carbohydrate esterase (orange); PL: polysaccharide lyase (yellow).

