

ORIGINAL ARTICLE

New roles in hemicellulosic sugar fermentation for the uncultivated Bacteroidetes family BS11

Lindsey M Solden¹, David W Hoyt², William B Collins³, Johanna E Plank⁴, Rebecca A Daly¹, Erik Hildebrand⁵, Timothy J Beavers¹, Richard Wolfe¹, Carrie D Nicora², Sam O Purvine², Michelle Carstensen⁵, Mary S Lipton², Donald E Spalinger⁶, Jeffrey L Firkins⁴, Barbara A Wolfe⁷ and Kelly C Wrighton¹

¹Department of Microbiology, The Ohio State University, Columbus, OH, USA; ²Pacific Northwest National Laboratory, Richland, WA, USA; ³Alaska Department of Fish and Game, Division of Wildlife Conservation, Palmer, AK, USA; ⁴Department of Animal Sciences, The Ohio State University, Columbus, OH, USA; ⁵Minnesota Department of Natural Resources, Division of Fish and Wildlife, Wildlife Health Program, Forest Lake, MN, USA; ⁶Department of Biology, University of Alaska Anchorage, Anchorage, AK, USA and ⁷Department of Veterinary Preventative Medicine, College of Veterinary Medicine, The Ohio State University, Columbus, OH, USA

Ruminants have co-evolved with their gastrointestinal microbial communities that digest plant materials to provide energy for the host. Some arctic and boreal ruminants have already shown to be vulnerable to dietary shifts caused by changing climate, yet we know little about the metabolic capacity of the ruminant microbiome in these animals. Here, we use meta-omics approaches to sample rumen fluid microbial communities from Alaskan moose foraging along a seasonal lignocellulose gradient. Winter diets with increased hemicellulose and lignin strongly enriched for BS11, a Bacteroidetes family lacking cultivated or genomically sampled representatives. We show that BS11 are cosmopolitan host-associated bacteria prevalent in gastrointestinal tracts of ruminants and other mammals. Metagenomic reconstruction yielded the first four BS11 genomes; phylogenetically resolving two genera within this previously taxonomically undefined family. Genome-enabled metabolic analyses uncovered multiple pathways for fermenting hemicellulose monomeric sugars to short-chain fatty acids (SCFA), metabolites vital for ruminant energy. Active hemicellulosic sugar fermentation and SCFA production was validated by shotgun proteomics and rumen metabolites, illuminating the role BS11 have in carbon transformations within the rumen. Our results also highlight the currently unknown metabolic potential residing in the rumen that may be vital for sustaining host energy in response to a changing vegetative environment.

The ISME Journal (2017) 11, 691–703; doi:10.1038/ismej.2016.150; published online 13 December 2016

Introduction

Ruminants are herbivorous mammals that depend on their microbial symbionts for the degradation of plant biomass, providing the host with energy in the form of short-chain fatty acids (SCFA; Van Soest, 1994). A recent census of microbial membership from over thirty ruminant species from across the globe revealed seven core bacterial groups conserved across 94% of the samples (Henderson *et al.*, 2015). Notably three of these seven bacterial taxa (the BS11 gut group, BF311 and an unclassified Bacteroidales family) represent uncultivated families with the

phylum Bacteroidetes. Although the Bacteroidetes are inferred to have key roles in rumen carbon transformations (Dodd *et al.*, 2010; Rosewarne *et al.*, 2014), these prevalent three lineages remain elusive due to a lack of genomic sampling and physiological characterization.

Due to recent genomic efforts we have gained new metabolic insight into ruminant microbes. The Hungate 1000 project was critical in advancing our knowledge of ruminant physiology, sequencing the genomes of over hundreds of isolated rumen microorganisms (Creevey *et al.*, 2014). Recent metagenomics studies have focused on the vast carbon degradation potential that lies within the rumen microbiome from a variety of hosts (Brulc *et al.* 2009; Findley *et al.* 2011; Hess *et al.* 2011; Pope *et al.*, 2012; Naas *et al.*, 2014; Singh *et al.*, 2014). However, with a few exceptions (Hess *et al.* 2011), most rumen metagenomics projects have not reconstructed

Correspondence: KC Wrighton, Department of Microbiology, The Ohio State University, 484 W 12th Avenue, Columbus, OH 43201, USA.

E-mail: wrighton.1@osu.edu

Received 7 March 2016; revised 29 September 2016; accepted 5 October 2016; published online 13 December 2016

genomes from metagenomic information, and thus we have little insight into the uncultivated, but prevalent Bacteroidetes families distributed across ruminants.

Given the abundance of unclassified Bacteroidales in browsing ruminants (Henderson *et al.*, 2015), we selected moose as our model, as the diet of these animals naturally spans a gradient of carbon complexity (from highly digestible grasses and leaves in spring to predominantly woody biomass in winter). This diet variability offers a greater opportunity for uncovering understudied rumen microbial taxa, which may be enriched under specific diet regimes. We used Alaskan moose fitted with rumen cannulae that were naturally foraging along a seasonal ligno-cellulose gradient, gaining real-time access to the rumen microbiome. Using metagenomics, we uncovered the first four genomes from the BS11 gut group. Metaproteomic data, in tandem with measured rumen NMR metabolites, confirmed that members of the BS11 gut group have the potential and express genes for degrading hemicellulose monomeric sugars in the rumen. These results shed further light on the metabolic versatility of the rumen, and link uncultivated Bacteroidetes members to essential carbon degradation processes.

Materials and methods

Experimental design and sample collection

Moose of good health and body condition were released into a 12-acre enclosure for two weeks during the spring (May), summer (August) and winter (January) in 2014–2015 at the University of Alaska's Matanuska Research Center (Supplementary Figure S1). Two of these naturally foraging moose are fitted with rumen cannulae, offering a unique, first of its kind opportunity to continuously sample rumen fluid in real-time and in response to diet. On the basis of prior studies (Spalinger *et al.*, 2010) moose were allowed to acclimate to the new diet for 7 days before our first sample collection. Rumen fluid was collected at three time points during the second week of the diet, with corresponding fecal samples collected at the same time. Samples were collected via two copper tubes cemented into the plug of the rumen cannula. These tubes were each attached to a tygon hose placed inside the rumen either 10 cm long (to sample the dorsal sac) or 30 cm long (sampling the ventral sac). These samples were combined for microbial and chemical analyses. Samples were immediately frozen and stored at -20°C until shipment from Alaska, and -80°C until DNA extraction and corresponding analyses in Ohio. Tissue collection from wild and captive animals is described in detail in the Supplementary Methods.

DNA extraction and 16S rRNA gene sequencing

Genomic DNA was extracted from rumen fluid, feces and tissue samples (0.5 g each) using a MoBio PowerSoil DNA Isolation Kit following

manufacturer's protocol, with the additional preparation of initially pre-heating PowerBead tubes at 70°C for 10 min before placing in a Thermomixer at 2000 r.p.m. for 10 min. DNA was sequenced at Argonne National Laboratory at the Next Generation Sequencing facility with a single lane of Illumina MiSeq using 2×251 bp paired end reads following established HMP protocols (Caporaso *et al.*, 2011). Data processing was performed using QIIME 1.9.0 unless otherwise noted, with analyses details included in Supplementary Methods. Our QIIME pipeline and NMDS workflow in R are available on github (<https://github.com/lmsolden/Qiime-pipeline>, <https://github.com/lmsolden/NMDS-with-loadings>).

Forage and rumen chemical characterization

Following visual observation of plant species consumed by moose, representative samples were collected from leaves, stems or whole plants, immediately frozen and subsequently lyophilized. Sequential detergent fiber analyses (neutral detergent fiber (NDF), acid detergent fiber (ADF) and acid detergent lignin (ADL)) were performed on all plants following the procedure outlined by Spalinger *et al.* (2010), with only lignin corrected for ash (see Supplementary Methods for more details). For rumen content chemical analyses, lyophilized digesta samples were chopped and ground by mortar and pestle to obtain a uniform particle size. Subsamples previously dried at 105°C were analyzed for CP ($N \times 6.25$) by Kjeldahl determination using a Foss Tecator digestion system (Foss Tecator AB, Höganäs, Sweden). Subsamples previously dried at 55°C were analyzed sequentially for NDF, ADF and ADL using standard protocols (AOAC, 1990), described in detail in the Supplementary Methods. Lignin was determined as the fraction of ash-free ADF insoluble in 72% sulfuric acid. Cellulose and hemicellulose were estimated as the mass loss from ADF to ADL and NDF to ADF, respectively. The entire ADF filtrate from one winter rumen fluid was concentrated to 10 ml for ^1H NMR analyses of hemicellulosic monomers. SCFA in filtered ($0.2 \mu\text{m}$) rumen fluids were detected on a Shimadzu HPLC fitted with an Aminex HPX-87H organic acid column using manufacturer's protocol.

Hemicellulose metabolites identified by ^1H NMR

Filtered ($0.2 \mu\text{m}$) rumen fluid samples from each season were sent to EMSL at the Pacific Northwest National Laboratory for NMR metabolite analysis. A total of 270 μl of the filtered sample (rumen fluid or hemicellulose acid extraction fluid) was mixed with 30 μl of 5 mM 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) in D_2O (Weljie *et al.*, 2006). NMR data was acquired on a Varian Direct Drive 600 MHz spectrometer operating VNMRJ 4.0 software (Agilent/Varian Inc., Palo Alto, CA, USA). On each sample, a one

dimensional ^1H nuclear Overhauser effect spectroscopy with presaturation experiment was collected following standard Chemomx data collection guidelines (see Supplementary Methods for more details). These 1D ^1H spectra were collected with either 512 transients (filtered rumen fluid) or 2048 transients (hemicellulose acid extraction). Collected spectra were analyzed using Chemomx 8.1 software with quantification based on spectral intensities relative to the 0.5 mM DSS internal standard. Two-dimensional spectra were acquired on the rumen fluid samples and aided the 1D ^1H assignments.

Metagenomic assembly, annotation and binning

One winter rumen fluid sample was separated into a pellet of plant material (gentle centrifugation for 5 mins at 3000 g) and the supernatant was sequentially filtered through a 0.8 μm filter and then onto a 0.2 μm filter. DNA was extracted from four fractions: the pellet (1g), half of the biomass retained on each of the 0.8 and 0.2 μm filters, and the filtrate that passed through the 0.2 μm filter. DNA was sequenced with Illumina Hi-Seq 2500 (Columbus, OH, USA) at The Ohio State University. 16S rRNA gene sequences were reconstructed from the Illumina trimmed unassembled reads using EMIRGE (Miller *et al.*, 2011). Trimmed reads were assembled *de novo* to generate genome fragments using IDBA-UD (Peng *et al.*, 2012). Genes were called, annotated and analyzed as previously described by Wrighton *et al.* (2012) (see Supplementary Methods for details). A combination of phylogenetic signal, coverage and GC content was used to identify BS11 genomic bins (Sharon *et al.*, 2013). Additional assembly and binning methods and validation information are available in the Supplementary Methods. Genomic completion of the BS11 bins was assessed based on the presence of a core gene set that typically occurs only once per genome and is widely conserved among bacteria and archaea (Wu and Eisen, 2008). For sequence-based comparison, average amino acid identity (AAI) and average nucleotide identity (ANI) values were calculated using the ANI and AAI calculators from the Kostas lab calculator (<http://enve-omics.ce.gatech.edu/>).

Existing reference datasets for the 11 ribosomal proteins chosen as single-copy phylogenetic marker genes (RpL2, 3, 4, 6, 14, 15, 16 and 18, and RpS8, 17 and 19) were augmented with sequences mined from sequenced genomes from the Bacteroidales phyla from the NCBI and JGI IMG databases (August 2015). Each individual protein dataset was aligned using MUSCLE 3.8.31 and then manually curated to remove end gaps (Edgar, 2004). Alignments were concatenated to form an 11-gene, 63 taxa alignment and then run through ProtPipeliner, a python script developed in-house for generation of phylogenetic trees (<https://github.com/lmsolden/protpipeliner>).

The pipeline runs as follows: alignments are curated with minimal editing by GBLOCKS (Talavara and Castresana, 2007), and model selection conducted via ProtTest 3.4 (Darriba *et al.*, 2011). A maximum likelihood phylogeny for the concatenated alignment was conducted using RAxML version 8.3.1 under the LG model of evolution with 100 bootstrap replicates (Stamatakis, 2014) and visualized in iTOL (Letunic and Bork, 2007). Identified glycoside hydrolases of selected functional classes (for example, chitin, hemicellulose and debranching) were identified by a Pfam HMM search. Briefly, Pfam search was performed and parsed into an output table organized by function per genome. In addition, we manually identified genes for central carbon metabolism, motility and fermentation product generation in all genomes.

Metaproteomic extraction, spectral analysis and data acquisition

The other half of the same filter used for metagenomics was selected for metaproteomic analyses. Filters were sonicated in SDS-lysis buffer and water bath sonication. Proteins in the supernatant were precipitated with protein pellets washed twice with acetone, and the pellet lightly dried under nitrogen. Filter Aided Sample Preparation (FASP) kits were used for protein digestion according to the manufacturer's instructions. Resultant peptides were snap frozen in liquid N_2 , digested again overnight and concentrated to $\sim 30 \mu\text{l}$ using a SpeedVac (Labconco, Kansas City, MO, USA). Final peptide concentrations were determined using a bicinchoninic acid (BCA) assay. All mass spectrometric data were acquired using a Q-Exactive Pro (Thermo Scientific, Waltham, MA, USA) connected to an ACQUITY UPLC M-Class liquid chromatography system (Waters, Milford, MA, USA) via in-house column packed using Phenomenex Jupiter 3 μm C18 particles (Torrence, CA, USA) and in-house built electrospray apparatus. MS/MS spectra were compared with the predicted protein collections using the search tool MSGF+ (Kim and Pevzner, 2014). Contaminant proteins typically observed in proteomics experiments were also included in the protein collections searched. The searches were performed using ± 15 p.p.m. parent mass tolerance, parent signal isotope correction, partially tryptic enzymatic cleavage rules and variable oxidation of Methionine. In addition, a decoy sequence approach (Elias and Gygi, 2010) was employed to assess false discovery rates. Data were collated using an in-house program, imported into a SQL server database, filtered to $\sim 1\%$ FDR (peptide to spectrum level) and combined at the protein level to provide unique peptide count (per protein) and observation count (that is, spectral count) data. Protein identification was based on two or more peptides per protein, however we reported single peptides with low confidence in gray that were manually

confirmed by examining the spectra. Details are included in supplementary text.

Results and discussion

Microbial members respond to increases in woody biomass

We assessed dietary plant chemistry, rumen content chemistry and rumen microbial membership from moose foraging along a seasonal lignocellulose gradient. Across all time points, the diet was composed primarily of deciduous shrubs (for example, birch and willow) (Supplementary Dataset S1). Winter plants were statistically different in chemistry, with threefold higher lignin, 40% higher cellulose and hemicellulose and 63% lower protein concentrations compared with spring plants (Figure 1; Supplementary Figure S2). Reflecting winter increases in dietary woody biomass, rumen fluid transitioned from green colored fluids in spring to brown fluids in winter (Supplementary Figure S1). Winter rumen fluid samples also had significantly higher cellulose and lignin concentrations with lower crude protein concentrations than spring rumen fluid samples (Supplementary Table S1).

16S rRNA gene analyses showed that microbial communities in rumen fluids and feces changed in response to seasonal plant chemistry (Supplementary Figure S3). Winter rumen microbial communities were statistically different from spring and summer, and all three treatments could be distinguished from pelleted rations ($P=0.001$) (Figure 2a; Supplementary Figure S3). Notably, spring and summer microbial communities could not be statistically differentiated from one another. Increased dietary woody biomass (for example, cellulose and lignin) and decreased SCFA concentrations were correlated to changes in microbial structure, not alpha diversity metrics, in winter (envfit, $P<0.05$).

We noted that four operational taxonomic units (OTUs) demonstrated the largest change in relative abundance on the winter diet compared with spring. Of these four OTUs, which all represented at least 0.8% average relative abundance across the winter samples, the two most dynamic OTUs were members of the ‘uncultivated BS11 gut group’, whereas the other two OTUs were members of the *Prevotella* genus. The most enriched OTU, a BS11, increased 800 fold from spring (0.005% abundance) to winter (4.4%) (Figure 2b). This OTU was a low abundant member in spring rumen fluids (1026 OTU rank) and became the third most abundant OTU detected in winter rumen fluid (Supplementary Dataset 2). Here we targeted the BS11 for genomic reconstructions, given that this prevalent ruminant bacterial family lacks any physiological or phylogenetic information. In addition, these uncultivated organisms represent conditionally rare members of the rumen community

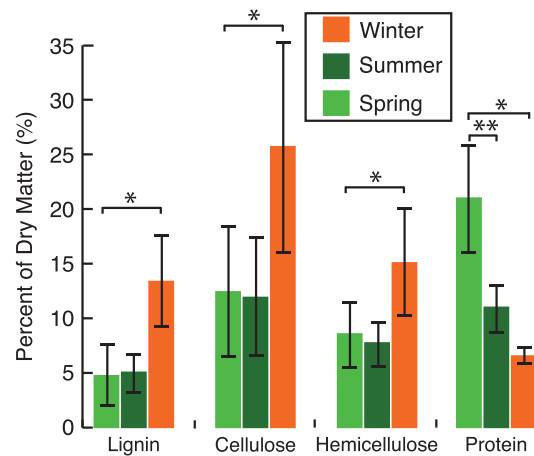


Figure 1 Winter plants have lower protein and higher lignin, cellulose and hemicellulose concentrations compared with spring and summer. Data includes the average of insoluble fiber determined by sequential fiber analyses on primary plant species consumed across all seasons, with s.d. shown. Values are reported as a percent of dry matter that was remaining at each sequential step. Significant differences ($P<0.05$, students *t*-test) between spring and summer compared with winter is denoted by *, whereas significant differences between spring and summer are denoted by **.

that responded to the increase in dietary woody biomass suggesting a unique role in rumen carbon cycling (Figure 2b).

Members of the BS11 are poorly described, yet prevalent in mammalian gastrointestinal tracts

On the basis of our findings in Alaska, we sampled the entire gastrointestinal intestinal tracts (GIT) from deceased moose from different geographic regions to better understand the biogeography of BS11. A majority of these GIT were from wild animals opportunistically collected in Minnesota as part of a study identifying factors involved in the 58% decline in moose over the past decade (Wunschmann *et al.*, 2015). We demonstrated that BS11 is present in rumen fluids from Ohio and Minnesota moose, in addition to Alaska (Supplementary Dataset S3).

Similar to the results of the seasonal diet experiment, the BS11 were the most dominant members in five wild Minnesota moose and moose from the Columbus Zoo whose cause of death was inferred to be illness, and all appeared nutritionally deprived (Supplementary Information). This increase at the family level was driven largely by changes in one OTU that was not highly similar (<90% identical) to the OTUs enriched on the Alaskan winter diet (Figure 2). The increased relative abundance of the BS11 family in rumen fluids in the sick moose (Supplementary Figure S4) could not be attributed to season due to lack of sampling but was shown to be independent of host geography, captivity status, diet, age and sex but was higher with more complex dietary carbon (Alaska winter) or diminished health

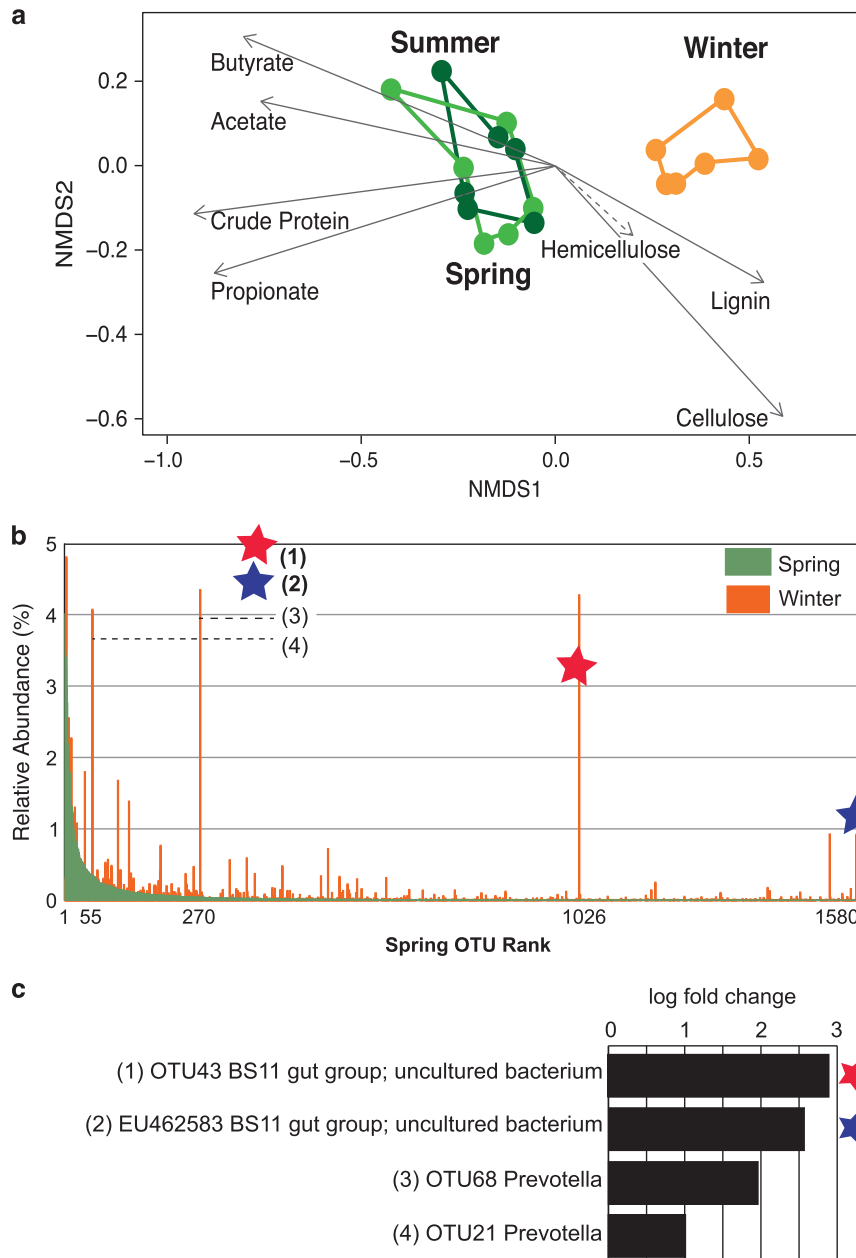


Figure 2 Rumen fluid microbial communities are statistically different in winter. **(a)** Non-metric multidimensional scaling (NMDS) of Bray–Curtis similarity metric shows a statistically significant (mrpp $P=0.001$) separation of rumen microbial communities from winter to spring in both moose. Diet is indicated by color, with spring (light green), summer (dark green) and winter (orange) denoted. Vectors are fitted with envfit and the length represents strength of the correlation (solid lines significant at $P<0.005$, dotted lines $P>0.05$). See Supplementary Table S1 for raw data. **(b)** Rank abundance of the all OTUs in spring and the corresponding abundance of the same OTUs on winter diet. The top two most enriched OTUs in winter are starred, with the third and fourth most enriched OTUs in winter highlighted with a dotted line. **(c)** Log-fold change from spring of most enriched OTUs in winter (x-axis). Corresponding OTU taxonomy is represented on the y-axis. Red and blue stars correspond to increasing OTUs in the winter diet in rank abundance curve **(b)**.

status (wild MN and Columbus Zoo) (Supplementary Dataset S3).

Alternatively, the BS11 family was not enriched in four healthy wild Minnesota moose where vehicle collisions were the cause of death ($2.1\% \pm 1$), nor was it prevalent in rumen fluids from a dairy cow (Supplementary Figure S4). The abundance of BS11 in our sampled cow rumen (3%) was similar to the abundance of BS11 in a microbial community

analysis of 394 cattle (1.4%; Henderson *et al.*, 2015). In addition to the rumen fluids, we also examined the GIT tissue for BS11 and found that high abundances were confined to reticulum tissue from moose that died of illness, but were not significantly enriched in other tissues from the rumen, omasum, abomasum, small intestine or colon and feces from healthy or cachectic moose (Supplementary Figure S5; Dataset S3). The low abundance of BS11 in

moose rumen tissue is consistent with a survey from hunter kills (Ishaq and Wright, 2014). The increase in BS11 under specific conditions is further support that these bacteria may represent dormant metabolic potential that respond to nutritional, environmental or health related stressors in ruminants.

Given that a microbial rumen census reported BS11 as the fifth most abundant taxon sampled in rumen fluid, and prevalent in 94% of 32 different hosts, this family appears highly adept at colonizing rumen fluids (Henderson *et al.*, 2015). We mined the dominant BS11 16S rRNA gene in our Alaska samples (Figure 2b) from other ruminant datasets. This OTU was detected in rumen fluid surveys from reindeer, elk, white tailed deer and muskoxen, as well as yak, cow, camel, goats and sheep (Koike *et al.*, 2003; Kong *et al.*, 2010; Pope *et al.*, 2012; Gruninger *et al.*, 2014; McCann *et al.*, 2014; Guo *et al.*, 2015; Gharechahi *et al.*, 2015). Analogous to our findings, these ruminants consumed a high fiber diet, with the BS11 often associated with the plant adherent fraction of the rumen (Koike *et al.*, 2003; Gruninger *et al.*, 2014). BS11 also colonized monogastric hosts and was detected in feces from a wide variety of zoo animals consuming a plant-based diet (for example, elephants, orangutans and gorillas) (Ley *et al.*, 2008; Yamono *et al.*, 2008), as well as in human feces (Leach *et al.*, 2012) and periodontal disease pockets (Silva accession number, HE681238). On the basis of our analysis, BS11 is likely host associated (Quast *et al.*, 2013), as we failed to identify 16S rRNA gene sequences originating from a non-animal source; that is, no sequences were found from other high woody biomass anoxic habitats, such as wetland soils

(see Supplementary Information for additional details).

BS11 remain enigmatic due to poor taxonomic assignment and a lack of metabolic information. On the basis of the ARB Silva (v123) database, the BS11 family contains 3,344 unique 16S rRNA gene sequences (resolved at >97% identity) but lacks any finer level of taxonomic resolution (for example, genera) (Quast *et al.*, 2013). Other 16S rRNA gene databases fail to recognize the BS11, for example commonly assigning sequences as ‘unclassified members of the phylum Bacteroidetes’ (RDP) (Wang *et al.*, 2007). This inconsistent taxonomic assignment across databases may contribute to the misidentification of BS11 in 16S rRNA surveys in hosts. Furthermore, the metabolic roles for the BS11 are unknown. For example, the most enriched BS11 OTU in our dataset shared only 85% 16S rRNA gene identity with the nearest isolated bacterium (a member of Porphyromonadaceae). Despite the lack of genomic sampling in the BS11, the broad host distribution in animals consuming a vegetarian based diet suggests a role in the degradation of plant-based compounds.

Multiple near-complete genomes resolve the family BS11 and identify two new candidate genera

To define the phylogeny and metabolic role of BS11, we sequenced four different size fractions from a winter rumen fluid sample (a loosely-centrifuged pellet of plant material, biomass contained on a 0.8 μ m filter, biomass on contained on a 0.2 μ m filter and a post-0.2 μ m filtrate). Near full-length 16S rRNA genes were reconstructed from metagenomic reads (Miller *et al.*, 2011) and phylogenetic analyses showed

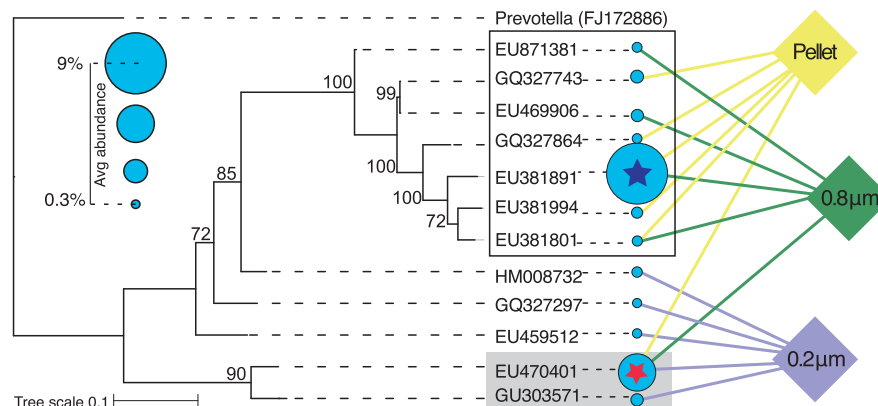


Figure 3 Different BS11 populations in rumen fluid are associated with different size fractions. Maximum likelihood tree of the reconstructed near full-length BS11 16S rRNA gene OTUs (circles, 97% similarity). Branches are named by corresponding Silva accession number, with *Prevotella* (EU259391) as the outgroup. Network analyses (right) highlight the distribution of the reconstructed 16S rRNA sequences in the different size fractions (diamonds). Circle size represents the 16S rRNA gene average relative abundance across the detected fractions (detailed in Supplementary Table S2). The gray box indicates sequences corresponding to *Candidatus Alcium*, whereas the white box indicates sequences corresponding to *Candidatus Hemicelluliticus*. The blue and red stars indicate the EMIRGE 16S sequence that match identically over the V4 region to the enriched BS11 OTUs in Figure 2b (blue and red stars, respectively). In addition, the red starred sequence was also recovered from the *Candidatus Alcium* genome scaffolds (Supplementary Text; Supplementary Figure S11).

that two BS11 genera were dominant in these samples (Figure 3). In our metagenomic dataset, the most dominant BS11 OTU (EU381891) was detected in the plant pellet and the 0.8 μm filter fractions and not on smaller sized soluble contents, whereas the second most abundant BS11 OTU (EU470401), constituting a different genus, was recovered from all size fractions greater than 0.2 μm in the rumen fluid (Figure 3; Supplementary Table S2). Both dominant OTUs were identical to the two OTUs in our amplicon sequencing (highlighted in Figure 2b with red and blue stars) that responded to the winter diet treatment (Supplementary Figure S6). Our metagenomic findings show that the two dominant BS11 genera in moose rumen fluid are associated with different size fractions, suggesting a possible niche partitioning.

We reconstructed four BS11 genomes with an estimated completion of 87–>99% and an average genome size of 2.9 Mbp (Supplementary Table S3; Supplementary Figure S7). We also recovered a BS11 scaffold containing 28 ribosomal proteins (0.155 Mbp), and a separate bin that contained BS11 scaffolds that could not be confidently assigned to one of the four BS11 genomes (1.2 Mbp). To evaluate the presence of BS11 in other rumen systems, we searched assembled ruminant metagenomes publically available (img.jgi.doe.gov) for BS11 ribosomal genes, identifying putative BS11 genomic fragments in sheep, cow and reindeer rumen fluids (Supplementary Discussion; Supplementary Figure S8) (Hess *et al.*, 2011; Pope *et al.*, 2012; GOLD project ID Ga0008901). However, these genomic fragments and the functional properties contained on the scaffolds were unbinned and not taxonomically assigned to the BS11 because of a lack of available reference genomes.

Phylogenetic analyses using concatenated alignments of 11 ribosomal proteins from the four near-complete genomes and the one scaffold supported the BS11 as a monophyletic family within the order Bacteroidales (Figure 4a; Supplementary Figure S9). This topology was also confirmed in single gene analyses. The average amino acid identity across the genomes (Supplementary Figure S10) and 16S rRNA gene linkages (Supplementary Text; Supplementary Figure S11) revealed these four genomes could be assigned to the two BS11 genera identified in our reconstructed 16S rRNA phylogeny (Figure 3; Supplementary Text). The *Candidatus* Alcium genus includes the OTU that most increased in abundance in response to a winter diet (red star corresponding to red star in Figure 2), whereas the *Candidatus* Hemicellulyticus genus is the second most enriched OTU relative to the spring diet (blue star, Figures 2c and 3). Following the recent naming protocol for near-complete (>95%) genomes from metagenomics (Konstantinidis and Rossello-Mora, 2015), we propose the name *Candidatus* Alcium based on the mammalian source (*Alces alces*) for this widely distributed OTU. For the second genus, composed of two near-complete genomes (>96%) and one partial genomic fragment, we propose the name

Candidatus Hemicellulyticus based on the inferred metabolism described below.

BS11 ferments hemicellulosic monomers to produce SCFA

We infer a fermentation-based lifestyle for the four sampled BS11, as these genomes lack *c*-type cytochromes, a linked electron transport chain, a complete tricarboxylic acid cycle and pyruvate dehydrogenase. The metabolic capabilities for one *Candidatus* Hemicellulyticus genome (F08-3) may be more extensive, as this genome had a partial NADH dehydrogenase (subunits D, E, F, C, B, A), succinate dehydrogenase and a putative b1 oxidase. We note however, that the NADH dehydrogenase may be an annotation artifact as the homologs identified had a low identity to known proteins and the remaining required subunits (M, N, L, J, K, H, I and G) were not detected in the genome. Members of *Candidatus* Hemicellulyticus and Alcium had bacterioferritin (cytochrome b1 oxidase), but given the lack of other electron transport chain complexes, we suggest this enzyme may have a role in iron assimilation (Carrondo, 2003). All four genomes are inferred to have a gram-negative cell envelope and lacked flagella, but encode an 11-gene gliding motility complex that in other organisms enables attachment and movement across surfaces with low water tension (for example, biofilm) (Mignot *et al.*, 2007; McBride and Zhu, 2013). This attachment-based motility may be a physiological adaptation to adhere to plant material in rumen fluid.

To put the global carbon-degrading metabolism of BS11 in a phylogenetic context, we profiled the glycoside hydrolases (GH) in genomes from across the Bacteroidales order. Although we report the average GH genes per genome of each family (Figure 4), a detailed analyses for each genome is included (Supplementary Table S5). Each BS11 genome encodes up to 25 GH, which when compared with other families is less than the average (mean 52.875 ± 38.6) (Supplementary Dataset S4). The BS11 genomes contained up to 11 chitin associated genes, which is the most sampled across the order. Although chitin is not commonly associated with a plant-based diet in ruminants, several of these genes (shown in red) confer the degradation of glycoproteins (for example, human mucin) in other gut bacteria (Macfarlane and Gibson, 1991; Roberts *et al.*, 2000; Etzold and Juge, 2014). Thus, it is possible these genes may have an alternative role in BS11 as well (Fredericksen *et al.*, 2013). Like most other members of the Bacteroidales, all members of the BS11 sampled to date encode the capacity for releasing hemicellulosic monomeric sugars.

In contrast to other families in this order, the BS11 lack a single genome representative with the capacity for cellulose degradation (Supplementary Dataset S4). However, similar to the gene prevalence in the Bacteroidetes, both *Candidatus* Hemicellulyticus

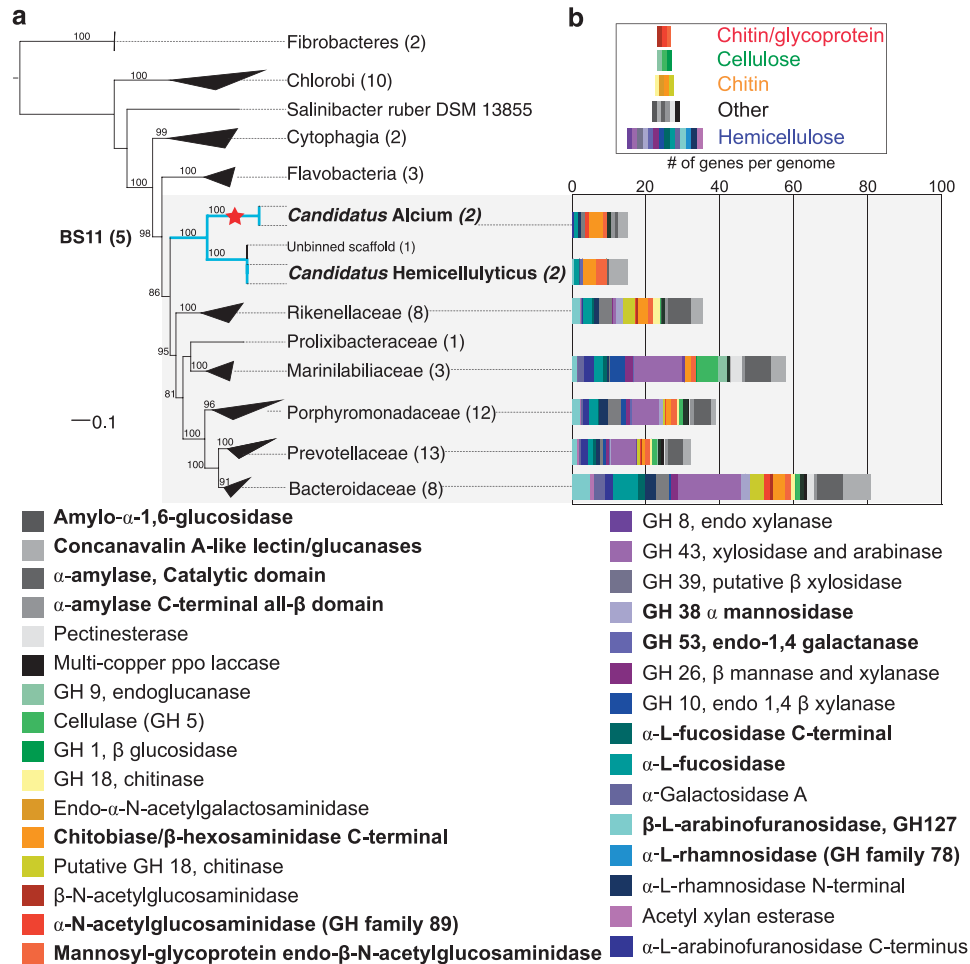


Figure 4 BS11 is a novel family in the Bacteroidales. (a) Concatenated ribosomal protein tree of the four BS11 genomes and one BS11 scaffold in comparison to the other 45 known genomic representatives from the order Bacteroidales with the six other families besides BS11 noted (gray box). An additional 18 reference genome sequences from other Bacteroidetes, Chlorobi and Fibrobacteres (outgroup) are also included. The number of genomes in each Bacteroidales family is denoted in parentheses next to the name. The near-complete BS11 sequences are shown in light blue with *Candidatus Alcium* highlighted with a red star to represent the recovery of a full-length 16S rRNA sequence in these genomic bins. Numbers on the node represent bootstrap support, using 100 bootstraps. The full maximum likelihood tree is provided in Supplementary Figure S8. (b) Glycoside hydrolase (GH) genes in genomes belonging to the order Bacteroidales are reported as the average number of GH genes/per genome for each family. The x-axis denotes the number of GH genes per genome, with specific Pfams clustered into functional category (denoted by color gradient). Gene names in bold were detected in the BS11 genomes. The cluster labeled as ‘other’ includes α -amylase, pectinesterase, concanavalin and polyphenol oxidoreductase laccase. The complete dataset from all genomes and accompanying PFAM numbers used to construct Figure 4b is included (Supplementary Dataset S4).

genomes contained Sus-like polysaccharide utilization loci (PULs) (Supplementary Text). The gene organization was similar to prior reports from rumen metagenomes or Bacteroidetes isolate genomes (Naas et al., 2014; Rosewarne et al., 2014), with SusC and SusD colocalized. Each *Candidatus Hemicellulyticus* genome contained two scaffolds with this structure and the Sus genes were surrounded by either peptidases or glycoside hydrolases (Supplementary Table S4), suggesting these loci may be involved in coordinating plant biomass degradation.

Given the increase in dietary hemicellulose when BS11 were enriched and the presence of GH for releasing hemicellulose, we used proton NMR to identify the sugar compositions associated with the different size fractions on the winter diet. Unlike cellulose (glucose subunits), hemicelluloses are

heterogeneous plant polymers containing multiple monomeric sugar units (for example, xylose, mannose, galactose, rhamnose and arabinose) arranged in different proportions (Scheller and Ulvskov, 2010). The polymeric properties of hemicelluloses vary depending on plant species and phenological progression (Van Soest, 1994), with much less known about the structural characteristics of arctic or boreal forages. Our NMR results show that soluble rumen fluids and plant materials contain a different combination of hemicellulosic sugars. Specifically the soluble fraction contained fucose (37%), galactose (24%), arabinose (19%), xylose (14%) and rhamnose (7%), whereas solidified plant material contained xylose (79%), arabinose (12%), fucose (7%) and galactose sugars (2%) (Figures 5a and b).

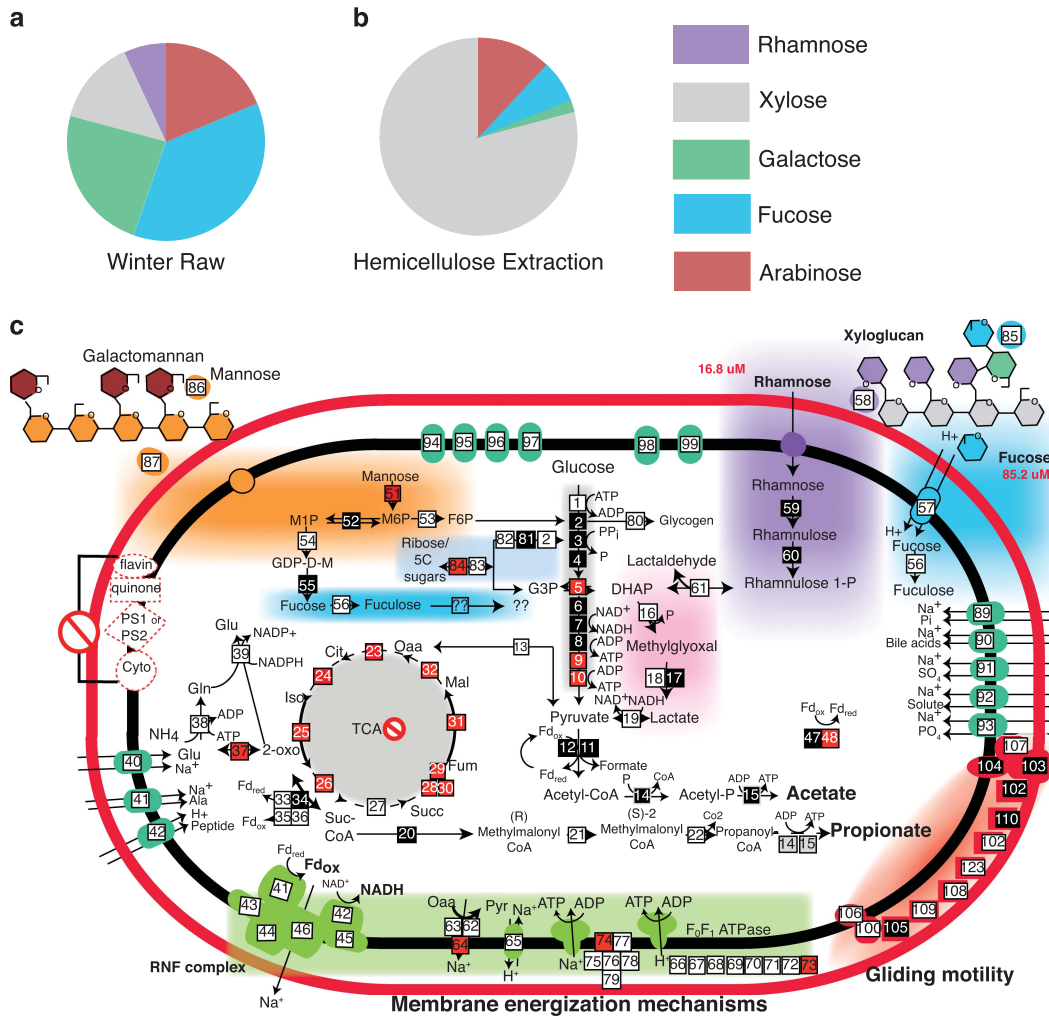


Figure 5 Genome-enabled metabolic potential of uncultivated BS11 *Candidatus Alcium*. **(a)** ^1H NMR analyses of hemicellulose sugars detected in raw filtered rumen fluid from winter sample. **(b)** ^1H NMR analyses of rumen fluid acid extract to specifically release hemicellulose sugars from solid plant material. **(c)** Metabolic predictions for BS11 *Candidatus Alcium* with putative pathways displayed in different colored boxes. Hemicellulose structures are displayed on the outside of the cell membrane (red oval), with concentrations of sugars detected with ^1H NMR listed in red. The metabolic pathway of each hemicellulose sugar is represented in orange (mannose), purple (rhamnose) or blue (fucose). Genes detected are represented in colored boxes. White boxes indicate presence of the enzyme in the genome, red boxes indicate absence, black boxes indicate genomic and proteomic support, and grey boxes indicate the presence in two pathways. For full enzyme annotation see Supplementary Table 6 Abbreviations: M1P: mannose-1-phosphate, M6P: mannose-6-phosphate, GDP-D-M: GDP-D-mannose, F6P: fructose-6-phosphate, G3P: glyceraldehyde-3-phosphate, DHAP: dihydroxyacetone phosphate, Oaa: oxaloacetate, Cit: citrate, Iso: isocitrate, 2-oxo: 2-oxoglutarate, Suc-CoA: succinyl-CoA, Succ: succinate, Fum: fumarate, Mal: malate, Gln: glutamate, Glu: glutamine, Pyr: Pyruvate. A detailed comparison of the metabolic capacities in *Candidatus Alcium* and *Hemicellulolyticus* is provided in Supplementary Table S5.

Our data suggest that the two BS11 genera are specialized for the fermentation of different hemicellulosic monomers. Although all four BS11 genomes are inferred to ferment mannose, the *Candidatus Alcium* genomes exclusively contain genes for releasing and utilizing more accessible hemicellulosic side-chain residues (for example, rhamnose and fucose) (Scheller and Ulvskov, 2010). Fucose fermentation may be particularly important to energy generation, as the most complete *Candidatus Alcium* genome has four copies of α -L-fucosidase, two fucose permeases and a fucose isomerase. Consistent with the distribution of *Candidatus Alcium* genomes in all size fractions (that

is, not necessarily associated with plant biomass), the genomes encode the capacity to ferment hemicellulose monomers most abundant in the bulk rumen fluid (fucose and rhamnose) (Figure 5a).

Unlike the broadly distributed *Candidatus Alcium*, the *Candidatus Hemicellulolyticus* genomes were recovered primarily from fractions containing solid plant biomass (Figure 3). Metabolite analysis of this plant biomass fraction revealed enrichment of xylose relative to fucose and rhamnose, likely reflecting the xylose backbone in hardwood hemicellulose (Scheller and Ulvskov, 2010). Consistent with the metabolite and abundance data from the plant fractions, the *Candidatus Hemicellulolyticus*

genomes exclusively encoded genes for xylose fermentation, but lacked genes for fucose or rhamnose fermentation, sugars in higher abundance in the soluble fractions (Supplementary Table S5). Specifically these BS11 may utilize the xylose isomerase pathway to convert xylose to xylulose 5-phosphate, which is ultimately degraded by a complete pentose phosphate pathway also present in the most complete *Candidatus Hemicellulolyticus* genome (Supplementary Table S5). Taking the genomic insights together, the preference for different hemicellulosic monomers by *Candidatus Alcium* (mannose, fucose and rhamnose) and *Candidatus Hemicellulolyticus* (mannose, galactose and xylose) likely explains the distribution of these two genera in the rumen fractions (Figure 3).

In BS11, we predict hemicellulose sugars are shunted to a complete glycolysis pathway in *Candidatus Hemicellulolyticus*, whereas both *Candidatus Alcium* genomes have a partial EMP-pathway lacking enolase, pyruvate kinase and triose-phosphate isomerase homologs (Figure 5c; Supplementary Table S5). This missing functionality in *Candidatus Alcium* is likely rescued via the full methylglyoxal detoxification pathway, in a proposed pathway similar to bovine rumen *Butyrivibrios* (Kelly et al., 2010; Hackman and Firkins, 2015). This alternative pathway, present in only the *Candidatus Alcium* genomes is consistent with the selective ability of this genus to use rhamnose, as intermediates from the degradation of this sugar yield dihydroxyacetone phosphate, the intermediate of the methylglyoxal detoxification pathway. The absence of triose-phosphate isomerase and the presence of D-lactate dehydrogenase suggest that rhamnose may be degraded fully to pyruvate via this metabolism. Together, the metabolite and genomic data support the degradation of hemicellulosic monomers as a shared metabolic feature of BS11 (Figure 5c), which explains the increase in BS11 during winter when woody plant tissue (and hemicellulose) comprises a much greater proportion of the moose diet. In addition, organismal distributions on different size fractions and sugar metabolite profiles suggest the two genera may be functionally specialized for specific hemicellulose monomers.

We infer the Rnf complex identified in all four BS11 genomes supports reverse electron flow (Biegel and Muller, 2010), using the sodium gradient to oxidize NADH and generate reduced ferredoxin. These reducing equivalents can then be disposed of by the generation of hydrogen (Schut and Adams, 2009) via FeFe hydrogenases. In addition to hydrogen generation, BS11 genomes also encode the production of SCFA to generate ATP via substrate-level phosphorylation. *Candidatus Hemicellulolyticus* has the potential to produce butyrate, propionate and acetate, whereas *Candidatus Alcium* can only produce propionate and acetate.

The production of SCFA by microbial symbionts is vital to ruminant nutrition, and supplies as much as 80% of the host energy and gluconeogenic substrates (Van Soest, 1994). Although the BS11 may contribute to SCFA, the lower absolute concentration produced on the winter diet is likely a result of the poorer quality diet. This capacity for metabolic exchange with the host hints at why BS11 may be maintained across a broad host exchange.

Metaproteomics supported functional genomic interpretations; the key genes in metabolic pathways identified by metagenomics were expressed in the rumen (Supplementary Table S5 and Dataset S5; Figure 5c). In *Candidatus Alcium*, proteins for rhamnose degradation, methylglyoxyl shunt and acetate production were detected (Figure 5c). Alternatively, in *Candidatus Hemicellulolyticus*, we detected proteins for butyrate and hydrogen production, but not acetate (Supplementary Table S5). Surprisingly, many genes annotated as proteins of unknown function were highly expressed comprising three of the five most highly detected proteins in *Candidatus Alcium*. This large number of highly expressed but unannotated genes indicates the large portion of metabolic information still to be discovered in relatively well-annotated phyla like the Bacteroidetes.

Conclusion

In summary, we uncovered important functional qualities for uncultivated, yet ubiquitous host-associated Bacteroidetes. Using >95% complete genomes we resolved at least two new genera within the BS11 family, here named *Candidatus Alcium* and *Candidatus Hemicellulolyticus*. Together, our metagenomic, metaproteomic and metabolomic data indicate BS11 are specialized to ferment many different hemicellulosic monomers (xylose, fucose, mannose and rhamnose), producing acetate and butyrate for the host. In the context of adaptation to changing environmental conditions, we have identified BS11 that become dominant in the rumen when the host consumes a high woody biomass diet. Furthermore the release and use of hemicellulosic monomeric sugars may help explain broad distribution in other animals, as this represents one of the most abundant food sources on this planet constituting the fiber in foliage, fruits, vegetables and grains (Baker et al., 1979).

Keystone ruminants such as the North American moose (*Alces alces*) and caribou (*Rangifer tarandus*) are experiencing nutritional limitation and severe population declines in response to climate related changes in the plant chemical landscape (Lenart et al., 2002; McArt et al., 2009). Climate change in arctic and boreal ecosystems is increasing the abundance of woody deciduous shrubs (Tape et al., 2006) and potentially making the carbon more recalcitrant via the production of secondary metabolites like

condensed tannins, which may impact rumen microbial metabolism (Kaarlejärvi *et al.*, 2012; Lavola *et al.*, 2013). Currently, the connection between dietary carbon composition, rumen microbial metabolism and host nutrition is largely unexplored in wild ruminants from climate-sensitive regions (Ishaq and Wright, 2014). The fact that necessary host metabolites are produced under new dietary regimes suggest conditionally rare taxa like BS11, which are generally concealed within the metabolic reservoir of the rumen, may be one of the many important, yet understudied players enabling wild herbivores to nutritionally adapt to a rapidly changing world.

Accession codes

Amplicon sequencing and metagenomic sequencing data have been deposited in the sequence read archive under BioProject PRJNA301235. The draft BS11 genomes have been deposited in NCBI GenBank and have been assigned accession numbers SAMN04328017, SAMN04328115, SAMN04328100 and SAMN04327460.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

Portions of this work were performed under a Science Theme Proposal (proposal ID: 48859, K.C. Wrighton) using EMSL, a national scientific user facility sponsored by DOE's Office of Biological and Environmental Research and located at PNNL. PNNL is operated by Battelle for the DOE under Contract DE-AC05-76RL01830. We thank the Minnesota Department of Natural Resources and Columbus Zoo for sample collection and ongoing collaboration. In addition, we would like to thank J John for making the preparation of this manuscript possible.

References

AOAC (Association of Official Analytical Chemists). (1990). *Fiber (Acid Detergent) and Lignin in Animal Feed (973.18). Official Methods of Analyses*. AOAC International: Arlington, VA, USA.

Baker DN, Norris KH, Li BW. (1979). *Food Fiber Analysis: Advances in Methodology*. In: Inglett GE, Falkehaug SI (eds), *Dietary fibers: chemistry and nutrition*. Academic Press: New York, 67–78.

Biegel E, Muller V. (2010). Bacterial Na⁺-translocating ferredoxin: NAD⁺ oxidoreductase. *Proc Natl Acad Sci USA* **107**: 18138–18142.

Brulc JM, Antonopoulos DA, Berg Miller ME, Wilson MK, Yannarell AC, Dinsdale EA *et al.* (2009). Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific

glycoside hydrolases. *Proc Natl Acad Sci USA* **106**: 1948–1953.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ *et al.* (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci USA* **108**: 4516–4522.

Carrondo MA. (2003). Ferritins, iron uptake and storage from the bacterioferritin viewpoint. *EMBO J* **22**: 1959–1968.

Creevey CJ, Kelly WJ, Henderson G, Leahy SC. (2014). Determining the culturability of the rumen bacterial microbiome. *Microbial Biotech* **7**: 467–479.

Darriba D, Taboada GL, Doallo R, Posada D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**: 1164–1165.

Dodd D, Mackie RI, Cann IKO. (2010). Xylan degradation, a metabolic property shared by rumen and human colonic Bacteroidetes. *Mol Microbiol* **79**: 292–304.

Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.

Elias JE, Gygi SP. (2010). Target-decoy search strategy for mass spectrometry-based proteomics. *Methods Mol Biol* **604**: 55–71.

Etzold S, Juge N. (2014). Structural insights into bacterial recognition of intestinal mucins. *Curr Opin in Struct Biol* **28**: 23–31.

Findley SD, Mormile MR, Sommer-Hurley A, Zhang XC, Tipton P, Arnett K *et al.* (2011). Activity-based metagenomic screening and biochemical characterization of bovine ruminal protozoan glycoside hydrolases. *Appl Environ Microbiol* **77**: 8106–8113.

Frederiksen RF, Paspaliari DK, Larsen T, Storgaard BG, Larsen MH, Ingmer H *et al.* (2013). Bacterial chitinases and chitin-binding proteins as virulence factors. *Microbiology* **159**: 833–847.

Gharechahi J, Zahiri HS, Noghabi KA, Salekdeh GH. (2015). In-depth diversity analysis of the bacterial community resident in the camel rumen. *Syst Appl Microbiol* **38**: 67–76.

Gruninger RJ, Sensen CW, Mcallister TA, Forster RJ. (2014). Diversity of rumen bacteria in canadian cervids. *PLoS One* **9**: e89682.

Guo W, Li Y, Wang L, Wang J, Xu Q, Yan T *et al.* (2015). Evaluation of composition and individual variability of rumen microbiota in yaks by 16S rRNA high-throughput sequencing technology. *Anaerobe* **34**: 74–79.

Hackmann TJ, Firkins JL. (2015). Electron transport phosphorylation in rumen butyrovibrios: unprecedented ATP yield for glucose fermentation to butyrate. *Front Microbiol* **6**: 622.

Henderson G, Cox F, Ganesh S, Jonker A, Young W, Global Rumen Census *et al.* (2015). Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Sci Rep* **5**: 14567.

Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G *et al.* (2011). Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* **331**: 463–467.

Ishaq SL, Wright AD. (2014). High-throughput DNA sequencing of the ruminal bacteria from moose (*Alces alces*) in Vermont, Alaska, and Norway. *Microb Ecol* **68**: 185–195.

- Kaarlejärvi E, Baxter R, Hofgaard A, Hytteborn H, Khitun O, Molau U *et al.* (2012). Effects of warming on shrub abundance and chemistry drive ecosystem-level changes in a forest–Tundra ecotone. *Ecosystems* **15**: 1219–1233.
- Kelly WJ, Leahy SC, Altermann E, Yeoman CJ, Dunne JC, Kong Z *et al.* (2010). The glyco biome of the rumen bacterium *Butyrivibrio proteoclasticus* B316^T highlights adaptation to a polysaccharide rich environment. *Plos One* **5**: e11942.
- Kim S, Pevzner PA (2014). MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat Commun* **5**: 5277.
- Koike S, Yoshitani S, Kobayashi Y, Tanaka K. (2003). Phylogenetic analysis of fiber-associated rumen bacterial community and PCR detection of uncultured bacteria. *FEMS Microbiol Lett* **229**: 23–30.
- Kong Y, Teather R, Forster R. (2010). Composition, spatial distribution, and diversity of the bacterial communities in the rumen of cows fed different forages. *FEMS Microbiol Ecol* **74**: 612–622.
- Konstantinidis KT, Rossello-Mora R. (2015). Classifying the uncultivated microbial majority: a place for metagenomic data in the Candidatus proposal. *Syst Appl Microbiol* **38**: 223–230.
- Lavola A, Nybakken L, Rousi M, Pusenius J, Petrelius M, Kellomaki S *et al.* (2013). Combination treatment of elevated UVB radiation, CO and temperature has little effect on silver birch (*Betula pendula*) growth and phytochemistry. *Physiol Plant* **149**: 499–514.
- Leach AL, Chong JP, Redeker KR. (2012). SSuMMo: rapid analysis, comparison and visualization of microbial communities. *Bioinformatics* **28**: 679–686.
- Lenart EA, Bowyer RT, Hoef JV, Ruess RW. (2002). Climate change and caribou: effects of summer weather on forage. *Canadian Journal of Zoology* **80**: 664–678.
- Letunic I, Bork P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**: 127–128.
- Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS *et al.* (2008). Evolution of mammals and their gut microbes. *Science* **320**: 1647–1651.
- Macfarlane GT, Gibson GR. (1991). Formation of glycoprotein degrading enzymes by *Bacteroides fragilis*. *Fems Microbiol Letters* **61**: 289–294.
- McArt SH, Spalinger DE, Collins WB, Schoen ER, Stevenson T, Bucho M. (2009). Summer dietary nitrogen availability as a potential bottom-up constraint on moose in south-central Alaska. *Ecology* **90**: 1400–1411.
- McBride MJ, Zhu Y. (2013). Gliding motility and Por secretion system genes are widespread among members of the phylum bacteroidetes. *J Bacteriol* **195**: 270–278.
- McCann JC, Wiley LM, Forbes TD, Rouquette FM, Tedeschi LO. (2014). Relationship between the rumen microbiome and residual feed intake-efficiency of Brahman bulls stocked on bermudagrass pastures. *PLoS One* **9**: e91864.
- Mignot T, Shaevitz JW, Hartzell PL, Zusman DR. (2007). Evidence that focal adhesion complexes power bacterial gliding motility. *Science* **315**: 853–856.
- Miller CS, Baker BJ, Thomas BC, Singer SW, Banfield JF. (2011). EMERGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biol* **12**: R44.
- Naas AE, Mackenzie AK, Mravec J, Schuckel J, Willats WGT, Eijsink VGH *et al.* (2014). Do rumen *Bacteroidetes* utilize an alternative mechanism for cellulose degradation? *mBio* **5**: e01401–e01414.
- Peng Y, Leung HC, Yiu SM, Chin FY. (2012). IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**: 1420–1428.
- Pope PB, Mackenzie AK, Gregor I, Smith W, Sundset MA, Mchardy AC *et al.* (2012). Metagenomics of the Svalbard reindeer rumen microbiome reveals abundance of polysaccharide utilization loci. *PLoS One* **7**: e38571.
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P *et al.* (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590–D596.
- Roberts G, Tarelli E, Homer KA, Philpott-Howard J, Beighton D. (2000). Production of an endo-beta-N-acetylglucosaminidase activity mediates growth of *Enterococcus faecalis* on a high-mannose-type glycoprotein. *J Bacteriol* **182**: 882–890.
- Rosewarne CP, Pope PB, Cheung JL, Morrison M. (2014). Analysis of the bovine rumen microbiome reveals a diversity of Sus-like polysaccharide utilization loci from the bacterial phylum *Bacteroidetes*. *J Ind Microbiol Biotechnol* **41**: 601–606.
- Scheller HV, Ulvskov P. (2010). Hemicelluloses. *Annu Rev Plant Biol* **61**: 263–289.
- Schut GJ, Adams MW. (2009). The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production. *J Bacteriol* **191**: 4451–4457.
- Sharon I, Morowitz MJ, Thomas BC, Costello EK, Relman DA, Banfield JF. (2013). Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. *Genome Res* **23**: 111–120.
- Singh KM, Reddy B, Patel AK, Panschasara H, Parmar N, Patel AB *et al.* (2014). Metagenomic analysis of buffalo rumen microbiome: effect of roughage diet on dormancy and sporulation genes. *Meta Gene* **2**: 252–268.
- Spalinger DE, Collins WB, Hanley TA, Cassara NE, Carnahan AM. (2010). The impact of tannins on protein, dry matter, and energy digestion in moose (*Alces alces*). *Can J Zool* **88**: 977–987.
- Stamatakis A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.
- Talavera G, Castresana J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* **56**: 564–577.
- Tape K, Sturm M, Racine C. (2006). The evidence for shrub expansion in Northern Alaska and the Pan-Arctic. *Global Change Biol* **12**: 686–702.
- Van Soest P. (1994). *Nutritional Ecology of the Ruminant*. Cornell University: London, UK.
- Wang Q, Garrity GM, Tiedje JMCole JR. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261–5267.

- Weljie AM, Newton J, Mercier P, Carlson E, Slupsky CM. (2006). Targeted profiling: quantitative analysis of ¹H NMR metabolomics data. *Anal Chem* **78**: 4430–4442.
- Wrighton KC, Thomas BC, Sharon I, Miller CS, Castelle CJ, Verberkmoes NC *et al.* (2012). Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science* **337**: 1661–1665.
- Wu M, Eisen JA. (2008). A simple, fast, and accurate method of phylogenomic inference. *Genome Biol* **9**: R151.
- Wunschmann A, Armien AG, Butler E, Schrage M, Stromberg B, Bender JB *et al.* (2015). Necropsy findings in 62 opportunistically collected free-ranging moose (*Alces alces*) from Minnesota, USA (2003–13). *J Wildl Dis* **51**: 157–165.
- Yamano H, Koike S, Kobayashi Y, Hata H. (2008). Phylogenetic analysis of hindgut microbiota in

Hokkaido native horses compared to light horses. *Anim Sci J* **79**: 234–242.



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>

Supplementary Information accompanies this paper on *The ISME Journal* website (<http://www.nature.com/ismej>)