

ORIGINAL ARTICLE

Open Access



# Metagenomic insights into the rumen microbial fibrolytic enzymes in Indian crossbred cattle fed finger millet straw

V. Lyju Jose<sup>1,2</sup>, Thulasi Appoorthy<sup>1\*</sup>, Ravi P. More<sup>3</sup> and A. Sha Arun<sup>4</sup>

## Abstract

The rumen is a unique natural habitat, exhibiting an unparalleled genetic resource of fibrolytic enzymes of microbial origin that degrade plant polysaccharides. The objectives of this study were to identify the principal plant cell wall-degrading enzymes and the taxonomic profile of rumen microbial communities that are associated with it. The cattle rumen microflora and the carbohydrate-active enzymes were functionally classified through a whole metagenomic sequencing approach. Analysis of the assembled sequences by the Carbohydrate-active enzyme analysis Toolkit identified the candidate genes encoding fibrolytic enzymes belonging to different classes of glycoside hydrolases (11,010 contigs), glycosyltransferases (6366 contigs), carbohydrate esterases (4945 contigs), carbohydrate-binding modules (1975 contigs), polysaccharide lyases (480 contigs), and auxiliary activities (115 contigs). Phylogenetic analysis of CAZyme encoding contigs revealed that a significant proportion of CAZymes were contributed by bacteria belonging to genera *Prevotella*, *Bacteroides*, *Fibrobacter*, *Clostridium*, and *Ruminococcus*. The results indicated that the cattle rumen microbiome and the CAZymes are highly complex, structurally similar but compositionally distinct from other ruminants. The unique characteristics of rumen microbiota and the enzymes produced by resident microbes provide opportunities to improve the feed conversion efficiency in ruminants and serve as a reservoir of industrially important enzymes for cellulosic biofuel production.

**Keywords:** Rumen, Fibrolytic enzymes, Plant polysaccharides, Metagenomics, Microbiome

## Introduction

Ruminants represent a substantial proportion of domesticated animal species worldwide and are the main source of milk, meat, and other dairy products. Ruminants have the ability to digest large amounts of plant polysaccharides by virtue of the composite microflora present in the rumen. The rumen has evolved into an efficient and effective fermentation vat for fiber degradation and the rumen is inhabited by a consortium of microorganisms consisting of bacteria, archaea, fungi, protozoa, and viruses (Miron et al. 2001; Pope et al. 2012), which interact and contribute significantly towards the health of ruminants. Among the various domains of microorganisms resident

in the rumen, bacteria are predominant representing about 95% of the total microbes (Mackie et al. 2000; Lin et al. 1997). The rumen fermentation process mediated by microbial communities affects the quality and composition of milk and meat and the productive performance of the host (Welkie et al. 2010; Stevenson and Weimer 2007; Sundset et al. 2009).

In tropical countries like India, ruminants are primarily fed on lignocellulose based agricultural crop residues. The extensive rumen microbiota are endowed with the potential to hydrolyze the plant polymers into simpler forms that provides nutrients to the host, predominantly in the form of volatile fatty acids and microbial proteins. The rumen essentially functioning as an anaerobic fermenter, has the ability to absorb the digested plant polysaccharides by the resident microflora (Jami and Mizrahi 2012). Studies on the symbiotic rapport between the rumen microbial communities and the mammalian host

\*Correspondence: thulasi.appoorthy@gmail.com

<sup>1</sup> Rumen Microbiology Laboratory, Animal Nutrition Division, National Institute of Animal Nutrition and Physiology, Bangalore, India  
Full list of author information is available at the end of the article

have posed a challenging area of research for the scientific community in the past due to the lack of adequate techniques to investigate and analyse such complex ecosystem.

Rumen microbes produce an array of fibrolytic enzymes called Carbohydrate-Active Enzymes (CAZymes), including exoglucanases, endoglucanases, glucosidases, and hemicellulases to deal with the complex plant polysaccharides. High Throughput Sequencing (HTS) technologies are extensively used to address the intricate process of lignocellulose degradation in ruminants. An improved understanding of the rumen microbial ecosystem could address the challenges in ruminant nutrition and environmental concerns in the livestock sector. Numerous metagenomic studies have reported on the diversity of fibrolytic enzymes from the rumen of yak (Dai et al. 2012), reindeer (Pope et al. 2012), Jersey cow (Wang et al. 2013), Angus cattle (Brulc et al. 2009), and buffalo (Singh et al. 2014). However, there are no comprehensive scientific reports available on metagenomic studies on the rumen CAZymes profile of Holstein–Friesian crossbred cattle, fed only finger millet straw. Due to the paucity of information on the CAZymes profile in HF cross fed only finger millet straw (a common crop residue fed to ruminants in Karnataka, India), this study was undertaken with the key objectives of deciphering CAZymes diversity in HF cross cattle and to enumerate the composition of metabolically active, CAZyme-contributing microbiota that is involved in the hydrolysis of plant polysaccharides. A comparative analysis of the data obtained in our study and other published herbivore metagenomes was also performed to identify whether any unique CAZyme families exist in the HF cross rumen ecosystem.

## Materials and methods

### Experimental design and rumen sampling

Three fistulated Holstein–Friesian crossbred steers with an average body weight of  $380 \pm 15$  kg were selected and maintained in individual stands for the feeding experiment at the Experimental Livestock Unit (ELU), National Institute of Animal Nutrition and Physiology, Bangalore, India. The animals were fed with finger millet straw, offered twice daily for a period of 21 days, at maintenance ration (ICAR 2013). The rumen contents were collected from all three animals prior to morning feeding on the last day of the experiment. Approximately 50 ml of rumen digesta samples were collected through the rumen fistula and immediately transported to the laboratory for further processing. Rumen digesta samples were mixed and strained through two layers of muslin cloth and immediately flash frozen in liquid nitrogen. Both liquid and solid portions of rumen digesta samples were then stored at  $-86^\circ\text{C}$  until further processing.

### Total DNA extraction from rumen digesta and quantification

The frozen rumen samples were thawed at room temperature and the solid rumen digesta samples were resuspended in phosphate buffered saline (Amresco, Solon, USA), for 2 h with vortexing to liberate the microbes adhering to feed particles, and mixed with the rumen fluid sample. The rumen fluid samples were then centrifuged at 4000 rpm for 5 min and the supernatant obtained was used further for the DNA extraction. In brief, the rumen fluid was centrifuged at 14,000 rpm for 10 min and the pelleted cells were resuspended in a mix of 800  $\mu\text{l}$  of CTAB lysis buffer (2% CTAB, 1.4 M NaCl, 20 mM EDTA and 100 mM Tris–HCl), (Amresco, Solon, USA) and 0.2 g of glass beads (0.1 mm), (Biospec products Inc, Bartlesville, USA) and kept in a Mini bead beater (Biospec products Inc, Bartlesville, USA) for 3 min. 10  $\mu\text{l}$  of 20 mg/ml proteinase K (Amresco, Solon, USA) and 10 mg/ml lysozyme (Amresco, Solon, USA) were added to the above mixture and incubated at  $37^\circ\text{C}$  for 1 h. The tubes were then incubated at  $70^\circ\text{C}$  for 30 min with intermittent mixing. An equal volume of Phenol:Chloroform:Isoamyl alcohol (25:24:1) (Amresco, Solon, USA) was added to the above lysate and mixed by inverting until a thick milky white emulsion was formed. After centrifugation at 14,000 rpm for 10 min, the supernatant was transferred to a fresh tube and total DNA was precipitated using 0.3 volumes of chilled ethanol (Merck, Kirkland, Canada). The precipitated DNA was then washed twice with 70% ethanol and the pellet was finally dried using a vacuum concentrator (Concentrator 5301) (Eppendorf, Hamburg, Germany). The quality of extracted genomic DNA was assessed by running it in 0.8% agarose gel electrophoresis for a single intact band, and A260/280 ratio was determined by Nanodrop 8000 (Thermo Fisher Scientific, Waltham, USA). Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, USA) was used to measure the quantity of DNA.

### Metagenome library preparation and sequencing

The paired-end sequencing library was prepared using Illumina, Truseq Nano DNA LT Library Preparation Kit (Illumina, California, United States). Subsequently, 200 ng of genomic DNA was fragmented by Covaris (Covaris Inc, Massachusetts, USA) to generate a mean fragment distribution of 550 bp. The fragments were then subjected to end repair using end repair mix and indexing adapters were ligated to the ends of the DNA fragments. The ligated products were purified using SP beads supplied in the kit. The size-selected product was PCR amplified as described in the kit protocol. The amplified library was analyzed in Bioanalyzer 2100 (Agilent Technologies, California, USA) using a High

Sensitivity DNA chip (Agilent Technologies, California, USA) as per the manufacturer's instructions. The library was then loaded onto the Illumina MiSeq platform for cluster generation and subjected to paired-end sequencing.

#### Metagenome assembly and bioinformatic analysis

De novo assembly of high quality data was accomplished using the CLC Genomics workbench 6.0 (Qiagen, USA) at default parameters (minimum contig length: 200, automatic word size: yes, perform scaffolding: yes, mismatch cost: 2, insertion cost: 3, deletion cost: 3, length fraction: 0.5, similarity fraction: 0.8). Bioinformatic analysis of the metadata was performed with Metagenome Rapid Annotation using Subsystem Technology (MG-RAST) (Meyer et al. 2008) server. The quality of the uploaded sequences was checked using MG-RAST quality filters and the sequences, which failed QC, were removed from further analysis. The metadata were functionally categorized via an RPS-BLAST comparison with the Subsystem database, (Overbeek et al. 2014), and KEGG databases (Kanehisa and Goto 2000).

#### Carbohydrate-active enzyme annotation and taxonomic profiling

The fibrolytic gene encoding contigs from the metadata were identified and classified based on the carbohydrate-active enzymes database (Cantarel et al. 2009) (<http://www.cazy.org>) by the carbohydrate-active enzyme analysis toolkit (CAT) (Park et al. 2010) at an E value of  $1 \times 10^{-5}$ . Putative plant cell wall polysaccharide-degrading enzymes belonging to different CAZy families were identified and classified based on sequence-based annotation. The CAZyme encoding contigs were analyzed manually for different classes of CAZymes: glycoside hydrolases (GHs), glycosyltransferases (GTs), carbohydrate esterases (CEs), carbohydrate-binding modules (CBMs), polysaccharide lyases (PLs), and auxiliary activities (AAs). The phylogenetic analysis of putative contigs encoding different CAZyme classes (24891 contigs) was performed in parallel to identify their microbial origin. The CAZyme encoding contigs from HF cross metagenome were uploaded on the MG-RAST server v 3.2 (Meyer et al. 2008) for phylogenetic analysis by the M5NR database using the BLASTX algorithm (Wilke et al. 2012) with a minimum identity of 60% and an E-value cutoff of  $1 \times 10^{-5}$ . The CAZymes obtained in the present study were compared with other accessible metagenomic datasets, cow rumen (Hess et al. 2011), jersey cow (Wang et al. 2013), reindeer (Pope et al. 2012), macropod (Pope et al. 2010), and termite gut (Warnecke et al. 2007).

## Results

### Metagenome sequence data statistics and phylogenetic abundance

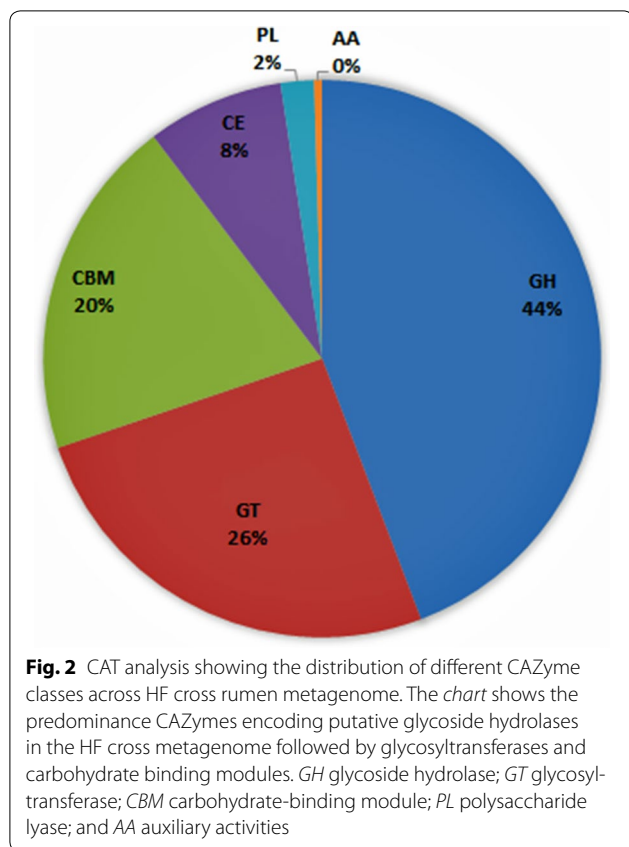
The ultimate objective of our study was to elucidate the fibrolytic potential of the rumen microbial community in Indian crossbred (HF) cattle fed finger millet straw. The whole metagenome sequencing of the total DNA from cattle rumen digesta generated about 1.8 gigabases of raw sequences. De novo assembly of the raw sequencing reads after quality check (CLC Genomics Workbench 6.0) (Qiagen, USA) resulted in 171,594 contigs with an average length of 838 bp. The statistical elements of the assemblies were calculated by in-house perl scripts and the metagenomic data analysis statistics are given in Table 1. Contig-7574 was the largest contig with a length of 25,731 bps. In order to validate the contig assembly, 16 contigs ( $\geq 600$  bp) from the glycoside hydrolase family were randomly selected and primers were designed to amplify the target gene fragment (Additional file 1: Table S1). Fifteen of the sixteen contigs were successfully amplified using at least one set of the primers and the sequences showed >99% identity to the assembled contigs. The metagenomic dataset was uploaded on the MGRAST server for further bioinformatic analysis. Phylogenetic analysis of metagenomic data at the domain level revealed that 97.5% of sequences binned to bacteria, 1.3% to archaea, and 0.9% to eukaryota (Fig. 1). At the genus level, the most predominant genera were *Prevotella*, *Bacteroides*, *Clostridium*, *Ruminococcus*, and *Parabacteroides*, representing more than 47% of the total sequences (Additional file 1: Table S2). The functional annotation using the SEED subsystem (Overbeek et al. 2014) has identified 73,886 predicted functions, out of

**Table 1 Rumen metagenome data assembly analysis statistics by using in house perl scripts**

Parameters	Number of sequences
Total number of bases uploaded	147,749,531
Total number of sequences uploaded	171,594
Mean sequence length bp uploaded	838 $\pm$ 481
Mean GC count uploaded	46 $\pm$ 10%
Artificial duplicate reads	14
Number of sequences failed QC	9310
Total number of bases post QC	125,833,189
Total number of sequences post QC	162,284
Mean sequence length post QC	775 $\pm$ 234
Mean GC count post QC	45 $\pm$ 10%
Predicted protein features	201,967
Predicted rRNA features	244
Identified protein features	97,723
Identified functional categories	58,691







of different CAZyme classes that are mainly involved in the catalytic hydrolysis of plant cell wall polysaccharides inside cattle rumen.

The GH catalytic modules corresponding to 11,010 sequences were the most predominant and represented 96 different families altogether. The contigs encoding enzymes belonging to functional class GTs were the second most prominent (25.58% of CAZymes) group in HF cattle rumen and contained 6366 sequences from 40 different families. The enzymes from GT2, GT4, GT51, and GT28 families were present in larger proportion, representing 72.93% of the total GTs. CBM modules included 4945 sequences from 35 families. The CE1 family, encoding feruloyl esterases, which are essential for the solubilization of plant polymer lignin, was the principal CE family with 390 contigs in cattle rumen. There were 1975 contigs associated with thirteen CE families in the CAZy database, representing 7.93% of CAZymes. CAT analysis also detected the presence of fifteen families of PLs (1.93%) and six families of AAs (0.46%) in HF cross rumen metagenome (Additional file 1: Table S4). Out of 24,891 CAZyme encoding contigs, there were 2517 contigs (10.11%) that had two or more distinct CAZy domains. CBM domains were predominant among the multi-domain enzyme families accounting for

about 1275 contigs (50.66%). Class glycoside hydrolases had 807 (32.06%) multi-domain contigs, largely associated with different carbohydrate-binding modules. CEs, GTs, and PLs represented 8.03, 5.44 and 3.77% of multi-domain enzymes, respectively, most of which were linked to CBMs.

A majority of the cattle rumen microbiome cellulases identified were classified as families GH5, GH6, GH9, GH44, GH45, and GH48. Gene fragments coding for putative cellulases belonging to families GH5 and GH9 accounted for 4% of the total GHs and were found to be the major families of cellulases present in cattle rumen. Likewise, 11 contigs encoding endoglucanases (GH45) and cellobiohydrolases (GH48) were also documented. Comparatively lesser number of cellulases degrading the main chains of plant cell wall components (10.67% of total GHs), were prevalent in HF rumen, while the enzymes involved in hydrolysis of side chains of plant polymers and oligosaccharides (GH1, GH2, GH3, GH29, GH35, GH38, GH39, GH42, GH43, and GH52), were in abundance and represented a majority of (60.11%) GHs. Putative exo- $\beta$ -1, 4-glucanases (GH1 and GH3) and endo- $\beta$ -1, 4-glucanases (GH5, and GH12) involved in the hydrolysis of  $\beta$ -1, 4-linked glucose residues from cellulose were represented by  $\sim$ 20 and  $\sim$ 4% of the total GHs, respectively. A large number of endo-hemicellulase degrading enzymes from GH8, GH10, GH11, GH12, GH26, GH28, and GH53 families were also identified and classified. CAT analysis also identified the presence of debranching enzymes belonging to family GH51 ( $\alpha$ -L-arabino furanosidases), GH67 ( $\alpha$ -glucuronidases), and GH 78 ( $\alpha$ -L-rhamnosidases) that play a key role in depolymerization of hemicellulose. The enzymes effecting the hydrolysis of xylan main chains such as endo-1, 4- $\beta$ -xylanase and endo-1, 3- $\beta$ -xylanase were represented by GH10 and GH11 families. These accounted for about 2.09 and 0.3% of the total GHs, respectively (Table 2). Among the 96 GH families reported in this study, families GH3, GH2, and GH43 encoding oligosaccharide-degrading enzymes were most abundant and represented by 993, 927, and 739 contigs, respectively (Additional file 1: Table S4).

#### Microbial community analysis of CAZyme encoding contigs

To determine the phylogenetic origin of core microbial populations that significantly contribute CAZymes, all 24,891 CAZyme encoding regions, representing different classes of GHs, GTs, CBMs, CEs, PLs, and AAs, were analyzed separately. Phylogenetic analysis of CAZyme contigs on the MG-RAST server using the M5NR database revealed that the bacterial species belonging to genera *Prevotella*, *Bacteroides*, *Clostridium*, *Fibrobacter*, and *Ruminococcus* contributed a

**Table 2 Overview of the comparative analysis of putative carbohydrate-active enzymes belong to members of various GH families targeting plant structural polysaccharides identified in HF cross rumen with other herbivore metagenomes**

GH family <sup>a</sup>	Major activity	HF cross	Jersey cow <sup>b</sup>	Cow <sup>c</sup>	Reindeer <sup>d</sup>	Termite <sup>e</sup>	Macropod <sup>f</sup>
Cellulases							
GH5	Cellulases	6.95	10.53	7.88	5.56	14.62	3.72
GH6	Endoglucanase	0.10	0.44	0.00	0.00	0.00	0.00
GH7	Endoglucanase	0.00	0.00	0.01	0.00	0.00	0.00
GH9	Endoglucanase	3.16	3.51	4.32	2.11	2.35	0.00
GH44	Endoglucanase	0.09	0.00	0.54	0.10	1.57	0.00
GH45	Endoglucanase	0.19	0.00	0.62	0.00	1.04	0.00
GH48	Cellobiohydrolases	0.19	0.00	0.02	0.10	0.00	0.00
Subtotal		10.67	14.47	13.39	7.87	19.58	3.72
Endo-hemicellulases							
GH8	Endoxylanase	1.14	0.00	1.79	0.68	1.31	0.37
GH10	Endo-1,4- $\beta$ -xylanases	2.03	15.35	5.57	3.68	12.01	4.09
GH11	Xylanases	0.29	0.00	0.90	0.16	3.66	0.00
GH12	Xyloglucanases	0.53	0.44	0.00	0.00	0.00	0.00
GH26	$\beta$ -Mannase and xylanases	1.16	0.44	2.00	2.97	3.92	1.86
GH28	Galacturonases	2.57	0.00	2.56	2.33	1.57	0.74
GH53	Endo-1,4- $\beta$ -galactanases	10.19	7.89	2.62	2.42	3.13	3.35
Subtotal		17.92	24.12	15.45	12.23	25.59	10.41
Xylanoglucanases							
GH16	Xyloglucanases	2.53	0.00	2.62	2.25	0.26	1.49
GH74	Xyloglucanases	0.45	0.00	2.09	0.85	1.83	0.37
Subtotal		2.98	0.00	4.72	3.10	2.09	1.86
Debranching enzymes							
GH51	$\alpha$ -L-Arabinofuranosidases	0.48	0.44	6.79	9.46	4.70	4.46
GH54	$\alpha$ -L-Arabinofuranosidases	3.47	0.00	0.41	0.45	0.00	0.00
GH62	$\alpha$ -L-Arabinofuranosidases	0.00	0.00	0.01	0.00	0.00	0.00
GH67	$\alpha$ -Glucuronidases	0.98	0.00	0.65	1.43	2.61	1.86
GH78	$\alpha$ -L-Rhamnosidases	3.38	5.70	6.85	6.07	0.00	9.29
Subtotal		8.31	6.14	14.70	17.40	7.31	15.61
Oligosaccharide degrading enzymes							
GH1	$\beta$ -Glucosidases	2.17	4.39	1.37	2.36	5.74	22.68
GH2	$\beta$ -Galactosidases	15.99	7.02	7.80	13.88	6.01	8.92
GH3	$\beta$ -Glucosidases	17.12	21.05	15.45	16.36	18.02	26.77
GH29	$\alpha$ -L-Fucosidases	2.45	1.32	5.10	5.19	0.00	0.74
GH35	$\beta$ -Galactosidases	1.12	0.88	0.86	0.76	0.78	1.12
GH38	$\alpha$ -Mannosidases	0.79	0.44	1.48	2.25	2.87	1.12
GH39	$\beta$ -Xylosidases	7.28	7.89	1.71	1.47	0.78	0.37
GH42	$\beta$ -Galactosidases	0.45	0.00	2.03	1.84	6.27	2.97
GH43	Arabino/xylosidases	12.74	12.28	15.93	15.25	4.18	3.72
GH52	$\beta$ -Xylosidases	0.00	0.00	0.01	0.04	0.78	0.00

**Table 2 continued**

GH family <sup>a</sup>	Major activity	HF cross	Jersey cow <sup>b</sup>	Cow <sup>c</sup>	Reindeer <sup>d</sup>	Termite <sup>e</sup>	Macropod <sup>f</sup>
Subtotal		60.11	55.26	51.75	59.40	45.43	68.40
Total		100.00	100.00	100.00	100.00	100.00	100.00
Metagenome size		1.8 Gb	0.28 Gb	268 Gb	0.30 Gb	0.062 Gb	0.054 Gb

Table lists the percentages of different glycoside hydrolase families targeting plant cell wall polysaccharides and their major activities. GHs are divided into different groups based on their activity and hydrolysis of plant polysaccharides. Enzymes belonging to oligosaccharide degrading cluster were observed as the predominant class representing a major proportion of GHs

<sup>a</sup> Based on CAZy data base (<http://www.cazy.org>)

<sup>b</sup> Wang et al. (2013) fed on Timothy grass hay

<sup>c</sup> Hess et al. (2011) fed on switch grass

<sup>d</sup> Pope et al. (2012) natural grazing on winter pastures, Norway

<sup>e</sup> Warnecke et al. (2007) study on wood feeding termite

<sup>f</sup> Pope et al. (2010) fed on Timothy canary grass and commercial pellet mix

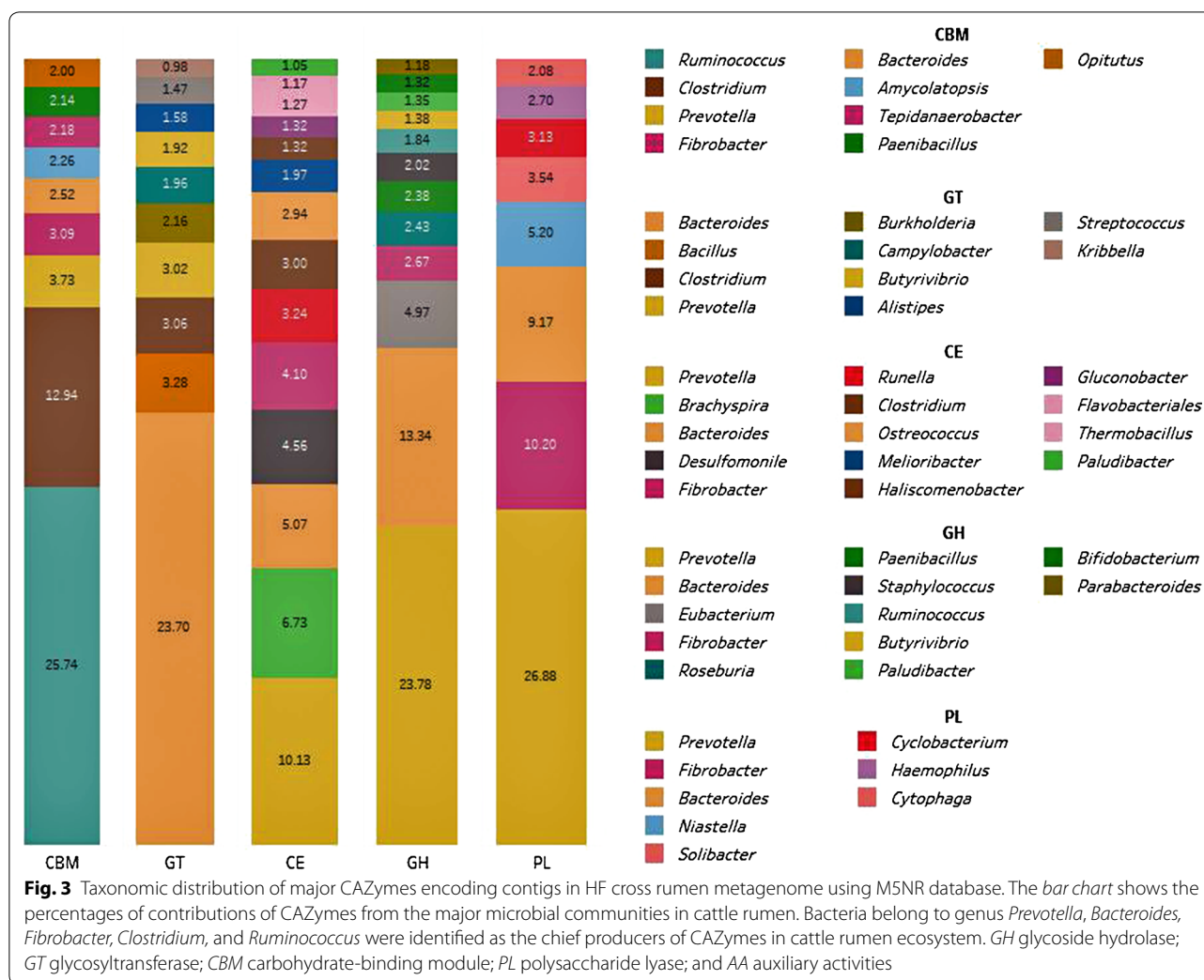
greater part of CAZyme encoding gene fragments in the HF cross rumen metagenome. Genera *Bacteroides* and *Prevotella* were the most dominant, producing a substantial amount of four major classes of CAZymes (GHs, GTs, CEs, and CBMs) accounting for about 44.63 and 36.1%, respectively. A significant proportion of GTs (23.7%), GHs (13.34%), and CEs (5.07%) were assigned to genus *Bacteroides* alone, whereas genus *Prevotella* was found to be the dominant contributor of GHs (19.22%), CEs (10.13%), and CBMs (3.73%). Genera *Fibrobacter* and *Clostridium* were found to encode three classes of CAZymes each: CBM, CE, GH and CBM, CE, GT, respectively. A major portion of CBMs were corresponding to members of genera *Ruminococcus* (25.74%) and *Clostridium* (12.94%) (Fig. 3).

## Discussion

### Rumen microbiota and CAZymes

Agricultural crop residues represent a significant proportion of renewable carbohydrate resources of energy for ruminants. The host as such cannot produce any enzymes that degrade the plant material, so they take advantage of the symbiotic association with rumen microflora to release the energy in form of carbohydrates and sugars from the recalcitrant plant polysaccharides. Rumen microbial populations are proven to play a vital role in plant fiber degradation in different ruminant species. A considerable variation in rumen microbiome has been observed across individual animals (Jami and Mizrahi 2012) and rumen microbial community structure also tends to vary significantly depending on the animal's age and diet (Kittlmann and Janssen 2011; Li et al. 2012). The rumen microbial profile is also moderately influenced by breed, gender, and other ecological factors like grazing location (Jami and Mizrahi 2012). Nevertheless, most of the published studies on rumen microbial communities have

identified the predominance of bacteria belonging to phyla *Bacteroidetes* and *Firmicutes*, irrespective of the change in diet, gender, breed, and other ecological factors (Kittlmann and Janssen 2011; Li et al. 2012; Petri et al. 2013; Stevenson and Weimer 2007; Thoetkiattikul et al. 2013). The results obtained in our studies confirm the earlier reports where the bacteria belonging to phyla *Bacteroidetes* and *Firmicutes* are the major contributors of different classes of CAZymes in the cattle rumen ecosystem. This result may indicate that the bacterial populations inhabiting the cattle rumen are largely similar to those of other ruminants in terms of *Bacteroidetes* and *Firmicutes* populations. The enzymes produced by these microbial communities are reported to have the potential to digest plant polymers like xylan, pectin, and starch (Stevenson and Weimer 2007). Genus *Prevotella* which has been established to account for a significant part of genetic and metabolic diversity of microbial communities in ruminants (Purushe et al. 2010). This genera was found to contribute a substantial proportion of CAZymes (>36%) in our study, when the animals were fed with only finger millet straw, that is rich in fiber. *Prevotella* has already been reported to play a vital part when there was a shift from a high-calorie diet to high-fiber diet (Jami et al. 2013). Microbes belonging to both genera *Bacteroides* and *Prevotella* were found to possess a large repertoire of CAZymes; *Bacteroides* were the chief manufacturers of CAZyme class GTs, whereas *Prevotella* were the principal contributors of GHs in cattle rumen. The natural capability of rumen microbiota to produce an array of potential enzymes that hydrolyze the rigid lignocellulose biomass has been successfully employed to treat different systems like agricultural residues and straw waste (Hu et al. 2007; Barnes and Keller 2004). The application of different artificial rumen systems for organic waste conversion (Yue et al. 2013) was also being extensively studied. Plant biomass containing



cellulose and hemicellulose on earth is considered to be one of the largest sources of fermentable sugars and energy that could be utilized for various industrial applications, like ethanol production (Jorgensen et al. 2007). Despite the large-scale availability of these plant polysaccharides, the major challenge for accessing these fermentable sugars is the presence of a highly resilient aromatic compound called lignin (Alvira et al. 2010). However, the ruminants with their highly complex efficient microbial communities can produce a broad range of fibrolytic enzymes that can facilitate the hydrolysis of lignocellulose biomass (Hess et al. 2011). However, it should be noted that the NGS-based Illumina-MiSeq sequencing platform employed in this study cannot be used for the quantitative interpretation of the rumen microbial community structure and size due to the bias inherent with the PCR amplification and sequencing of 16s rRNA gene, as it may result in misinterpretation of active community members (Schloss et al. 2011).

### Cattle rumen has highly diverse and complex CAZymes

The CAZy database (<http://www.cazy.org>) exclusively deals with the diverse group of enzymes that actively contribute to synthesis and degradation of complex carbohydrates and glycoconjugate. The CAZy database provides manually curated information for all CAZyme families: glycoside hydrolases, glycosyltransferases, polysaccharide lyases, carbohydrate esterases, carbohydrate-binding modules, and auxiliary activities; it now allows one to examine all known families and enzymes involved in cellulolysis, hemicellulolysis, and pectinolysis (Cantarel et al. 2009). Enzymes coding for GH families are highly abundant in most of the genomes and they account for about 50% of the enzymes classified in the CAZy database. Among all major classes of CAZymes that have been reported till date, glycoside hydrolases were observed to be the most predominant and diverse group of catalytic enzymes involved in the hydrolysis of plant polymers in Indian crossbred cattle rumen metagenome.



GHS representing glycosidases and transglycosidases are responsible for the hydrolysis of the glycosidic bonds linking two or more carbohydrates or a carbohydrate and a non-carbohydrate moiety accounting for about 6.5% of the total sequences. The total number of contigs obtained in each CAZyme family gives an indication of the abundance of that particular group of enzymes in the HF cross cattle rumen. Out of 135 CAZy families that are reported in the CAZy database till date, 96 families were represented by cattle rumen metagenome. Patel et al. (2014) reported 72 families of GHS from the buffalo rumen ecosystem. The substantially large number of enzymes from the GH family (96 families) in our study could be either due to the difference in NGS sequencing platform or the modifications in the analysis pipe line. Considerably higher number of GH families obtained in our study also indicates that HF cross rumen might have a more intricate process of lignocellulose breakdown than other reported rumen metagenomes.

In parallel to previous reports (Wang et al. 2013; Stevenson and Weimer 2007; Pope et al. 2012), genes encoding cellulases belonging to families GH5 and GH9 were present in higher proportion than other cellulase coding families. Dai et al. (2012) reported the exoglucanase activity of enzymes corresponding to the GH5 family which is known for endoglucanase, mannanase and endo-xylanase activities. The large number of putative cellulase genes identified in the GH5 family (403 contigs) in our study suggests that cattle rumen microbes may have novel strategies for degradation of plant polysaccharides present in the feed. Previous studies on rumen metagenomes have not reported any contigs assigned to the GH6 family, other than one contig from the fosmid library of jersey cow rumen metagenome (Wang et al. 2013). Conversely, Indian crossbred cattle rumen microbiome was found to exhibit a marginally higher amount of endoglucanases corresponding to the GH6 family (6 contigs). Consistent with earlier herbivore-related metagenome and metatranscriptome studies (Brulc et al. 2009; Hess et al. 2011; Pope et al. 2012; Warnecke et al. 2007; Qi et al. 2011) no contigs associated with the GH7 cellulase family were found. Consequently, the scanty representation (11 contigs) of enzymes belonging to the GH48 family displays the possibilities of the non-cellulosomal mode of plant cell wall hydrolysis in cattle rumen. Additionally, the contigs which were affiliated to diverse CAZyme modules (~13%), but could not be classified under any of the known or reported Pfam domains, indicated that the cattle rumen microbiome produces a large number of unique CAZymes which are meagerly characterized. Complete depolymerization of plant polysaccharides requires collective activities of different groups of enzymes other than GHS and these include GTs, CBMs,

CEs, PLs, and AAs. GTs, reported to be involved in the catalysis of glycosidic bonds to form disaccharides, oligosaccharides, and polysaccharides from phospho-activated sugar donors (Coutinho et al. 2003; Lairson et al. 2008), were the second most abundant CAZy family HF rumen (~4% of total contigs). Out of the forty GT families identified in cattle rumen metagenome, enzymes belonging to families GT2 and GT4 (cellulose synthase, chitin synthase,  $\alpha$ -glucosyltransferase, etc.) represented a significant proportion (>62%) of the total GTs. CBMs which have no reported enzymatic activity on their own, but can potentiate the activities of all other CAZymes (GHS, CEs, and auxiliary enzymes) or act as an appendix module of CAZymes (Tomme et al. 1995; Boraston et al. 2004), accounting for 2.9% of contigs.

In conclusion, high throughput sequencing-based whole metagenomic approach was used to investigate the natural biomass-converting microbial communities in the cattle rumen ecosystem. Comprehensive analysis of metagenome data strongly indicates that the complex digestive system of Indian crossbred cattle possesses a very high degree of a deeply branched and extremely diverse group of enzymes of microbial origin which have the ability to degrade the recalcitrant lignocellulose plant biomass. The CAZymes corresponding to class glycoside hydrolases were found to be the most abundant and diverse group of CAZymes in cattle rumen. Bacteria belonging to genera *Prevotella*, *Bacteroides*, *Clostridium*, *Fibrobacter*, and *Ruminococcus* were identified as the key contributors of CAZymes inhabiting the cattle rumen. This study provides a substantially expanded catalogue of enzymes that participate in the deconstruction of plant cellulosic biomass in Indian cattle rumen, which provides an opportunity to improve ruminant nutrition and also to develop proficient fermentation systems for bioconversion of plant biomass into biofuels.

## Additional file

**Additional file 1.** Additional tables and figures.

## Abbreviations

CAZymes: carbohydrate active enzymes; HTS: high throughput sequencing; HF: Holstein Friesian; ELU: experimental livestock unit; CTAB: cetyl trimethylammonium bromide; MG-RAST: Metagenome Rapid Annotation using Subsystem Technology; QC: quality control; RPS BLAST: reversed position specific basic local alignment search tool; KEGG: Kyoto Encyclopedia of Genes and Genomes; CAT: carbohydrate-active enzyme analysis toolkit; GHS: glycoside hydrolases; GTs: glycosyl transferases; CEs: carbohydrate esterases; CBMs: carbohydrate-binding modules; PLs: polysaccharide lyases; AAs: auxiliary activities.

## Authors' contributions

LJV and TA designed and performed the experiment and wrote the manuscript. RPM and ASA analyzed the data. All authors read and approved the final manuscript.

**Author details**

<sup>1</sup> Rumen Microbiology Laboratory, Animal Nutrition Division, National Institute of Animal Nutrition and Physiology, Bangalore, India. <sup>2</sup> Department of Biotechnology, Jain University, Bangalore, India. <sup>3</sup> Bioinformatics Laboratory, Division of Molecular Entomology, National Bureau of Agricultural Insect Resources, Bangalore, India. <sup>4</sup> Wildlife SOS, Bannerghatta Bear Rescue Centre, Bangalore, India.

**Acknowledgements**

The authors would like to acknowledge the Director, NIANP, Bangalore, and Jain University, Bangalore, for supporting the research program.

**Competing interests**

The authors declare that they have no competing interests.

**Availability of data and materials**

The raw data obtained had been deposited in the NCBI Sequence Read Archive (SRA) with the Bio Project ID- PRJNA291680, Accession: SAMN03963232, ID: 3963232.

**Ethical approval**

All applicable institutional guidelines for the care and use of animals were followed and ethical committee at National Institute of Animal Nutrition and Physiology, Bangalore, India, has approved the feeding experiment and sample collection.

**Funding**

This research program was carried out with financial support from, the Indian Council of Agricultural Research (ICAR), New Delhi, India, by the National Centre on Veterinary Type Culture Collection-Rumen Microbes (NCVTCC-RM) component.

Received: 14 December 2016 Accepted: 19 December 2016

Published online: 03 January 2017

**References**

- Alvira P, Tomás-Pejó E, Ballesteros M, Negro MJ (2010) Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: a review. *Bioresour Technol* 101:4851–4861
- Barnes SP, Keller J (2004) Anaerobic rumen SBR for degradation of cellulosic material. *Water Sci Technol* 50:305–311
- Boraston AB, Bolam DN, Gilbert HJ, Davies GJ (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J* 382:769–778. doi:10.1042/BJ20040892
- Brucic JM, Antonopoulos DA, Miller MEB, Wilson MK, Yannarell AC, Dinsdale EA, Edwards RE, Frank ED, Emerson JB, Wacklin P, Coutinho PM (2009) Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proc Natl Acad Sci USA* 106:1948–1953. doi:10.1073/pnas.0806191105
- Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycomics. *Nucleic Acids Res* 37:D233–D238. doi:10.1093/nar/gkn663
- Coutinho PM, Deleury E, Davies GJ, Henrissat B (2003) An evolving hierarchical family classification for glycosyltransferases. *J Mol Biol* 328:307–317
- Dai X, Zhu Y, Luo Y, Song L, Liu D, Liu L, Chen F, Wang M, Li J, Zeng X, Dong Z (2012) Metagenomic insights into the fibrolytic microbiome in yak rumen. *PLoS ONE* 7(7):e40430. doi:10.1371/journal.pone.0040430
- Hess M, Szczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, Luo S, Clark DS, Chen F, Zhang T, Mackie RI (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* 331:463–467. doi:10.1126/science.1200387
- Hu ZH, Liu SY, Yue ZB, Yan LF, Yang MT, Yu HQ (2007) Microscale analysis of in vitro anaerobic degradation of lignocellulosic wastes by rumen microorganisms. *Environ Sci Technol* 42:276–281. doi:10.1021/es071915h
- ICAR (2013) Nutrient requirements of cattle and buffaloes, 3rd edn. Indian Council of Agricultural Research New Delhi, New Delhi
- Jami E, Mizrahi I (2012) Composition and similarity of bovine rumen microbiota across individual animals. *PLoS ONE* 7(3):e33306. doi:10.1371/journal.pone.0033306
- Jami E, Israel A, Kotsler A, Mizrahi I (2013) Exploring the bovine rumen bacterial community from birth to adulthood. *ISME J*:1069–1079. doi:10.1038/ismej.2013.2
- Jorgensen H, Vibe-Pedersen J, Larsen J, Felby C (2007) Liquefaction of lignocellulose at high-solids concentrations. *Biotechnol Bioeng* 96:862–870. doi:10.1002/bit.21115
- Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28:27–30. doi:10.1093/nar/28.1.27
- Kittelmann S, Janssen PH (2011) Characterization of rumen ciliate community composition in domestic sheep, deer, and cattle, feeding on varying diets, by means of PCR-DGGE and clone libraries. *FEMS Microbiol Ecol* 75:468–481
- Lairson LL, Henrissat B, Davies GJ, Withers SG (2008) Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem* 77:521. doi:10.1146/annurev.biochem.76.061005.092322
- Li RW, Connor EE, Li C, Baldwin VI, Ransom L, Sparks ME (2012) Characterization of the rumen microbiota of pre-ruminant calves using metagenomic tools. *Environ Microbiol* 14:129–139. doi:10.1111/j.1462-2920.2011.02543.x
- Lin C, Raskin L, Stahl DA (1997) Microbial community structure in gastrointestinal tracts of domestic animals: comparative analyses using rRNA-targeted oligonucleotide probes. *FEMS Microbiol Ecol* 22:281–294
- Mackie RI, Aminov RI, White BA, McSweeney CS (2000) Molecular ecology and diversity in gut microbial ecosystems. *Ruminant physiology: digestion, metabolism, growth and reproduction*. CAB International, Oxford, pp 61–77
- Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J (2008) The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinform*. doi:10.1186/1471-2105-9-386
- Miron J, Ben-Ghedalia D, Morrison M (2001) Invited review: adhesion mechanisms of rumen cellulolytic bacteria. *J Dairy Sci* 84:1294–1309. doi:10.3168/jds.S0022-0302(01)70159-2
- Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V (2014) The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res* 42:D206–D214. doi:10.1093/nar/gkt1226
- Park BH, Karpinets TV, Syed MH, Leuze MR, Uberbacher EC (2010) CAZymes analysis toolkit (CAT): web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZY database. *Glycobiology* 20:1574–1584. doi:10.1093/glycob/cwq106
- Patel DD, Patel AK, Parmar NR, Shah TM, Patel JB, Pandya PR, Joshi CG (2014) Microbial and carbohydrate active enzyme profile of buffalo rumen metagenome and their alteration in response to variation in the diet. *Gene* 545:88–94
- Petri RM, Schwaige T, Penner GB, Beauchemin KA, Forster RJ, McKinnon JJ, McAllister TA (2013) Characterization of the core rumen microbiome in cattle during transition from forage to concentrate as well as during and after an acidotic challenge. *PLoS ONE* 8(12):e83424. doi:10.1371/journal.pone.0083424
- Pope PB, Denman SE, Jones M, Tring SG, Barry K, Malfatti SA, McHardy AC, Cheng JF, Hugenholtz P, McSweeney CS, Morrison M (2010) Adaptation to herbivory by the Tamar wallaby includes bacterial and glycoside hydrolase profiles different from other herbivores. *Proc Natl Acad Sci* 107:14793–14798. doi:10.1073/pnas.1005297107
- Pope PB, Mackenzie AK, Gregor I, Smith W, Sundset MA, McHardy AC, Morrison M, Eijsink VG (2012) Metagenomics of the Svalbard reindeer rumen microbiome reveals abundance of polysaccharide utilization loci. *PLoS ONE* 7(6):e38571. doi:10.1371/journal.pone.0038571
- Purushe J, Fouts DE, Morrison M, White BA, Mackie RI, Coutinho PM, Henrissat B, Nelson KE, North American Consortium for Rumen Bacteria (2010) Comparative genome analysis of *Prevotella ruminicola* and *Prevotella bryantii*: insights into their environmental niche. *Microb Ecol* 60:721–729. doi:10.1007/s00248-010-9692-8
- Qi M, Wang P, O'Toole N, Barboza P, Ungerfeld E, Leigh MB, Selinger LB, Butler G, Tsang A, McAllister TA, Forster RJ (2011) Snapshot of the eukaryotic gene expression in muskoxen rumen—a metatranscriptomic approach. *PLoS ONE* 6(5):e20521. doi:10.1371/journal.pone.0020521

- Schloss PD, Gevers D, Westcott SL (2011) Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS ONE* 6(12):e27310. doi:10.1371/journal.pone.0027310
- Singh KM, Reddy B, Patel AK, Panchasara H, Parmar N, Patel AB, Shah TM, Bhatt VD, Joshi CG (2014) Metagenomic analysis of buffalo rumen microbiome: effect of roughage diet on Dormancy and Sporulation genes. *Meta Gene* 2:252–268
- Stevenson DM, Weimer PJ (2007) Dominance of *Prevotella* and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. *Appl Microbiol Biotechnol* 75:165–174. doi:10.1007/s00253-006-0802-y
- Sundset MA, Edwards JE, Cheng YF, Senosiain RS, Fraile MN, Northwood KS, Praesteng KE, Glad T, Mathiesen SD, Wright AD (2009) Molecular diversity of the rumen microbiome of Norwegian reindeer on natural summer pasture. *Microb Ecol* 57:335–348. doi:10.1007/s00248-008-9414-7
- Thoetkiattikul H, Mhuantong W, Laothanachareon T, Tangphatsornruang S, Pat-tarajinda V, Eurwilachitr L, Champreda V (2013) Comparative analysis of microbial profiles in cow rumen fed with different dietary fiber by tagged 16S rRNA gene pyrosequencing. *Curr Microbiol* 67:130–137. doi:10.1007/s00284-013-0336-3
- Tomme P, Warren RAJ, Gilkes NR (1995) Cellulose hydrolysis by bacteria and fungi. *Adv Microb Physiol* 37:1–81
- Wang L, Hatem A, Catalyurek UV, Morrison M, Yu Z (2013) Metagenomic insights into the carbohydrate-active enzymes carried by the microorganisms adhering to solid digesta in the rumen of cows. *PLoS ONE* 8(11):e78507. doi:10.1371/journal.pone.0078507
- Warnecke F, Luginbühl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, Cayouette M, McHardy AC, Djordjevic G, Aboushadi N, Sorek R, Tringe SG, Podar M, Martin HG, Kunin V, Dalevi D, Madejska J, Kirton E, Platt D, Szeto E, Salamov A, Barry K, Mikhailova N, Kyrpides NC, Matson EG, Ottesen EA, Zhang X, Hernandez M, Murillo C, Acosta LG, Rigoutsos I, Tamayo G, Green BD, Chang C, Rubin EM, Mathur EJ, Robertson DE, Hugenholtz P, Leadbetter JR (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* 450:560–565. doi:10.1038/nature06269
- Welkie DG, Stevenson DM, Weimer PJ (2010) ARISA analysis of ruminal bacterial community dynamics in lactating dairy cows during the feeding cycle. *Anaerobe* 16:94–100
- Wilke A, Harrison T, Wilkening J, Field D, Glass EM, Kyrpides N, Mavrommatis K, Meyer F (2012) The M5nr: a novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools. *BMC Bioinform* 13(1):141. doi:10.1186/1471-2105-13-141
- Yue ZB, Li WW, Yu HQ (2013) Application of rumen microorganisms for anaerobic bioconversion of lignocellulosic biomass. *Bioresour Technol* 128:738–744

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)

---