

## EXPLORING THE METABOLICALLY ACTIVE RUMEN MICROBIOTA AND ITS FIBROLYTIC POTENTIAL IN CROSSBRED CATTLE FED ON FIBROUS DIET THROUGH METATRANSCRIPTOMICS

Lyju Jose V.<sup>1,2</sup>, Ravi P. More<sup>3</sup>, Pradeep Kumar Malik<sup>4\*</sup>, Atul Purushottam Kolte<sup>4</sup>, Shraddha Trivedi<sup>4</sup>, Sha Arun<sup>2</sup>, G. Thirumalaisamy, Raghavendra Bhatta<sup>4</sup>

### Address(es):

<sup>1</sup> Rumen Microbiology Laboratory, Animal nutrition Division, National Institute of Animal Nutrition and Physiology, Bangalore, India.

<sup>2</sup> PhD Scholar, Department of Biotechnology, Jain University, Bangalore, India.

<sup>3</sup> Bioinformatician, National Centre for Biological Sciences, Bangalore, India.

<sup>4</sup> Energy Metabolism Laboratory, Bioenergetics and Environmental Sciences Division, National Institute of Animal Nutrition and Physiology, Bangalore, India.

\*Corresponding author: [malikndri@gmail.com](mailto:malikndri@gmail.com)

doi: 10.15414/jmbfs.2020.10.2.182-189

### ARTICLE INFO

Received 30. 4. 2019

Revised 9. 6. 2020

Accepted 11. 6. 2020

Published 1. 10. 2020

Regular article



### ABSTRACT

The rumen wharves conglomerate of symbiotic fibrolytic microflora with the potential to produce repertoire of carbohydrate-active enzymes to elicit the degradation of recalcitrant plant lignocellulosic complex into simpler forms that can be utilized by the host. A metatranscriptomics approach was adopted to explore the metabolically active microbiota and the transcripts involved in the deconstruction of lignocellulose in Indian crossbred cattle rumen. The experimental animals were fed with mixed diet for a period of twenty-one days and the total RNA was isolated from the rumen digesta using a modified total RNA extraction protocol. After mRNA enrichment, the cDNA libraries were subjected to sequencing under Illumina HiSeq platform and generated about 2.89 GB of raw sequences. The contig assembly was accomplished using two different tools and the redundant sequences were removed before further processing. A total of 133,930 orfs were predicted from the assembled contigs using MetaGeneMark tool and the taxonomic affiliation of orfs were achieved using MEGAN 6.0 Community Edition. A total of 17 bacterial, 5 eukaryotic and 1 archaeal phylum were identified from rumen metatranscriptomic dataset. Phylum *Bacteroidetes* were acknowledged as the most abundant and metabolically active rumen bacterial populations in cattle rumen with 36,414 orfs corresponding to it, followed by *Firmicutes* with 6,349 orfs. Further, MEGAN annotation of metatranscriptomic data at species level revealed that *Prevotella brevis* and *Prevotella ruminicola* as the most metabolically active bacterial populations representing 4,022 and 2,725 orfs each. The functional annotation of metatranscriptomic data using COG database revealed that a large number of transcripts (~16%) were corresponding to translation, ribosomal structure and biogenesis and carbohydrate transport and metabolism (15%). dbCAN annotation of rumen metatranscriptomic data identified and classified 5,267 transcripts belonging to 168 CAZyme families (GH-52%, GT-26%, CBM-14%, CE-6%, PL-2% and AA-0%). The microbial community analysis of the CAZyme encoding transcripts using M5NR database revealed that a significant proportion of CAZymes were contributed by genera *Prevotella* (37%) and *Bacteroides* (21%). The cattle rumen microbiome metatranscriptome analysis presented in this study facilitated the detection of large number of transcripts encoding diverse CAZymes that actively take part in the hydrolysis of plant lignocellulose complex that may be useful for improving livestock and biofuel generation.

**Keywords:** Rumen, Metatranscriptomics, ORFs, MEGAN, CAZymes

### INTRODUCTION

Rumen-the foremost chamber in ruminant's stomach-encompasses a plethora of obligate anaerobic microorganisms including bacteria, archaea, fungi and protozoa that are essential for the bioconversion of plant lignocellulosic biomass into simpler forms that can be utilized by the host (Deng *et al.*, 2008). The highly complex and diverse microbiota present in the rumen has evolved into a symbiotic association with the host and can produce a battery of lignocellulolytic enzymes (CAZymes) that in turn hydrolyse the complex plant polysaccharides and release volatile fatty acids and microbial proteins which are the major sources of carbon, nitrogen and energy for the host (Purushe *et al.*, 2010; Youssef *et al.*, 2013). The rumen microbial community structure had been subjected to extensive studies by various researchers and the results designate that the rumen bacteria play a central role in cellulolysis due to their numerical predominance and metabolic diversity (Cheng *et al.*, 1991). The rumen microbial ecosystem has fascinated the scientific community due to their direct involvement in both economically and environmentally significant traits like feed conversion efficiency in livestock (Nkrumah *et al.*, 2008; Garg *et al.*, 2013), their impact on global warming due to methanogenesis (Zhou *et al.*, 2009, O'Brien *et al.*, 2014) and also the recent discovery about the potential of rumen microbes and its fibrolytic enzymes in the production of biofuels from cellulosic biomass (Azman *et al.*, 2015; Carrillo-Reyes *et al.*, 2016).

Advances in DNA sequencing technologies and bioinformatics are transforming our understanding of complex microbial ecosystems, particularly the gastrointestinal tract of mammals. The DNA based culture independent techniques form the majority of rumen microbial community studies, as traditional culture dependent methods can symbolise approximately 11% of total microbiota (Fernando *et al.*, 2010). Likewise, the sequencing of genomes from several cultured rumen bacterial and archaeal species is providing detailed information about their physiology (Nathani *et al.*, 2015; Kelly *et al.*, 2016). Lately, metagenomics, mainly aimed at understanding the rumen microbial community structure and its enzymatic machinery involved in the degradation of plant structural polysaccharides brings new insights by allowing access to the total community and overcoming the limitations imposed by cultivation and other PCR based assays (Li 2015; Pitta *et al.*, 2016; Jose *et al.*, 2017). Studies on rumen fibrolytic microbial genomes revealed they harbor a wide range of carbohydrate-active enzymes belonging to different classes that act synergistically to deconstruct the dietary fibre (Wang *et al.*, 2013; Ribeiro *et al.*, 2016).

While metagenomics brings an insight into the overall microbial and fibrolytic potential of rumen microbial ecosystem, metatranscriptomics approach is considered to be more reliable in scrutinizing the metabolically dynamic microbial communities and its functional transcripts encoding fibrolytic enzymes. Since metatranscriptomic approach focuses on the transcripts that are getting

expressed at the time of sampling, it can bring light into the active metabolic and functional profile of a microbial community. Rumen metatranscriptomic studies can provide a snapshot of the transcriptional profiles that correspond to the discrete populations, which are not essentially the dominant populations within the ecosystem. Qi *et al.*, in 2011 published the first ever rumen metatranscriptomic study from musk oxen rumen but limited the study to the eukaryotic microbial communities. Recently, Li & Guan (2017) and Comtet-Marre *et al.*, (2017) adopted RNA-sequencing technologies to identify the fibrolytic microbial communities in beef cattle and dairy cow, respectively. In present study, metatranscriptomic sequencing approach was employed to provide an insight into the metabolically active rumen microbiota and its lignocellulolytic enzyme profiles in HF cross cattle fed mixed diet.

**MATERIALS AND METHODS**

**Animal diet and rumen sample collection**

The study was conducted in three fistulated *Holstein Friesian* (HF) crossbred cattle (BW 380±15 kg) maintained at the Experimental Livestock Unit (ELU) in the National Institute of Animal Nutrition and Physiology. All experimental procedures adopted in the study were approved by the local institute ethical committee. The animals were fed with finger millet straw, green fodder and concentrate mixture in 60:30:10 for a period of twenty-one days. The ingredient composition of concentrate mixture is given in Table S1 & S2. The animals had free access to water twice a day. On the last day of feeding trial, the rumen digesta samples were collected through the rumen fistula, before morning feeding from all three experimental animals. The entire rumen digesta samples comprising both solid and liquid fractions were collected in cryovials and flash frozen in liquid nitrogen and immediately transported to the laboratory in a portable cool rack.

**Table S1** Composition of concentrate mixture

Ingredient	Quantity (%)
Maize	37
GNC*	32
WB**	28
Mineral mix <sup>#</sup>	2
salt	1

<sup>#</sup>DCP: 52.5%; Calcite: 20%; ZnSO4: 6.5%; CuSO4: 1%; MgO: 2.5%; NaCl: 17.5%, \*GNC: Ground nut cake; \*\*WB: Wheat bran

**Table S2** Chemical composition of finger millet straw, para grass and concentrate mixture

Chemical composition (%) ++	Concentrate Mixture	Finger millet	Paragrass
		Straw	
DM	91.71	92.84	29.5
CP	20.49	4.87	8.9
EE	1.14	1.88	2.0
NDF	27.96	74.32	72.5
ADF	17.20	44.75	38.9
Total ash	3.52	6.88	8.1

++ On dry matter basis except for dry matter

**RNA extraction and purification**

The total RNA was isolated from the rumen digesta samples on the same day. All centrifugation steps were performed at 20 °C until and otherwise specified. The frozen rumen digesta samples were crushed in liquid nitrogen using a motor and pestle and all three samples were pooled together. Into the pooled rumen digesta sample, 1 ml of TRIzol reagent (M/s Thermo Fisher Scientific, USA) was added. The sample was transferred to 2 ml screw cap tube containing 0.1 g of acid washed 0.1mm glass beads and 0.1mm zirconia beads (M/s Biospec products, USA) and subjected to bead beating (2 x 1 minutes at room temperature, 2 min in ice flakes between bead beating) using a Mini-Beadbeater (M/s Biospec products, USA). The homogenate was then centrifuged at 5,000 rpm for 10 min. The supernatant was then transferred to a fresh 2 ml tube and centrifuged at 14,000 rpm for 10 minutes. The pellet obtained was then resuspended in 1 ml of TRIzol reagent (M/s Thermo Fisher Scientific, USA), mixed by pipetting and incubated at 60 °C for 3 min. Then, 50 µL of lysozyme (500 µg/ml) (M/s Amresco, USA) was added, mixed by pipetting and incubated for 10 min at 37 °C. The lysate obtained was further processed with RiboPure RNA Purification Kit (M/s Thermo Fisher Scientific, USA) as per manufacturer’s instructions. On Colum DNase treatment was performed using the RNase free DNase supplied along with the kit by adding 80 µl of PureLink DNase mixture to the spin cartridge and incubated at room temperature for 15 minutes. The quality and quantity of isolated total RNA was checked using Experion Automated Electrophoresis

System (M/s Bio-Rad, USA) using Experion RNA StdSens Analysis Kit (M/s Bio-Rad, USA). The total RNA was further quantified using Qbit fluorometer before processing (M/s Invitrogen, USA).

**mRNA enrichment and Illumina library preparation**

The total RNA samples were initially treated with MICROExpress Bacterial mRNA Enrichment Kit (Thermo Fisher Scientific, Waltham, USA) to deplete the ribosomal RNA and enrich the mRNA population. The metatranscriptome libraries from the enriched mRNA were prepared using Illumina Truseq mRNA Stranded Library Preparation Kit (Illumina, San Diego, USA) as per manufacturer's instructions

**Quality and quantity check of library, cluster generation and sequencing**

Library quantification was performed using Quant-iT dsDNA High Sensitivity Assay Kit, (Thermo Fisher Scientific, Waltham, USA). The amplified libraries were further analyzed in Bioanalyzer 2100 using High Sensitivity (HS) DNA chip (Agilent Technologies, Santa Clara, USA) as per manufacturer's instructions. After obtaining the Qubit concentration for the libraries and the mean peak size from Bioanalyzer profile libraries were loaded into Illumina HiSeq (2 x 250) platform for cluster generation and paired-end sequencing. The library molecules bind to complementary adapter oligos on paired-end flow cell. The adapters are designed to allow selective cleavage of the forward strands after re-synthesis of the reverse strand during sequencing. The copied reverse strand was used to sequence from the opposite end of the fragment.

**Metatranscriptomic contig assembly, rRNA and ORF prediction**

The raw reads obtained were checked for its quality and the presence of adapter sequences using in-house Perl scripts. *De novo* assembly of high-quality PE reads were accomplished using 2 different algorithms- Metavelvet (Namiki *et al.*, 2012) and CLC Genomics Workbench 7.0- at default parameters. Further, both the assemblies were combined, and the redundant sequences were removed to get the final set of transcripts. The statistical elements of the assemblies were calculated by in-house perl scripts. The assembled contigs were analyzed against in-house RNA database using BlastN algorithm (Zhao & Chu, 2014) at similarity threshold 100% and E-value cut-off 1e-06, to identify the ribosomal RNA reads. The contigs generated from metatranscriptomic data were also analysed against the SSU and LSU rRNA sequences in RDP database. The reads that were matching with the rRNA database were considered as rRNA reads and other reads are considered as non rRNA or mRNA sequences. The open reading frames (orfs) from the contigs were predicted using MetaGeneMark tool (Zhu *et al.*, 2010). The predicted orfs along with the rRNA reads were uploaded on MG-RAST web server (Meyer *et al.*, 2008) for further analysis.

**Taxonomic affiliation and functional annotation of metatranscriptomic data**

The function based taxonomic affiliation for the transcripts was achieved using the MEGEN 6.0 Community Edition (CE) (Huson *et al.*, 2016). The functional annotation and homology search for the assembled transcript sequences were performed against COG database proteins downloaded from the database (<https://www.ncbi.nlm.nih.gov/COG/>) (Galperin *et al.*, 2015) using NCBI-blast 2.2.29 for homology search between the sequences.

**COG annotation, Gene ontology and WEGO (Web Gene Ontology Annotation Plot) analysis**

The functional annotation and homology search for the assembled transcript sequences were performed against COG (cluster of orthologous genes) database proteins downloaded from the database (<https://www.ncbi.nlm.nih.gov/COG/>) using NCBI-blast2.2.29 for homology search between the sequences. COG analysis (Galperin *et al.*, 2015) was performed to classify the transcripts according to their homologous relationships from the assembled transcripts from the cattle rumen metatranscriptomic data. The functional gene ontology (GO) annotation of rumen metatranscriptomic dataset set was performed using BLAST2GO tool (Götz *et al.*, 2008) and the GO annotation results were further analyzed using WEGO program (Ye *et al.*, 2006), to classify the protein encoding transcripts under three major functional classes viz. cellular location, molecular function and biological process. Since the WEGO web tool can process only one lakh GO annotation sequences, the output from BLAST2GO annotation were divided into two files and uploaded on WEGO server. The WEGO results from two separate files were then merged and represented graphically.

**CAZyme annotation and taxonomic assignment of CAZyme encoding transcripts**

To identify and classify different CAZyme families that are involved in lignocellulose degradation in cattle rumen, the predicted orfs from the metatranscriptome data were further analyzed using a Pfam based Hidden

Markov Model (HMMs) in dbCAN database (Yin et al., 2012). The dbCAN was downloaded from dbCAN was downloaded (<http://csbl.bmb.uga.edu/dbCAN/download/CAZyDB.03172015.fa>). The predicted orfs were subjected to blastX search against the dbCAN sequences to identify and classify the metabolically active transcripts encoding different CAZymes. The transcripts which showed resemblance with the dbCAN database were then analyzed manually in detail to categorize them under different CAZyme families. The transcripts encoding principal plant cell-wall degrading enzymes were extracted from the metatranscriptomic dataset using Perl scripts. Taxonomic affiliation of rumen metatranscriptomic sequences encoding cellulases, endo-hemicellulases, xyloglucanase, debranching and oligosaccharide degradation enzymes that are predominantly involved in the hydrolysis of plant polymers were determined using M5NR database on MG-RAST server v3.2 (Meyer et al., 2008). The CAZymes encoding transcripts identified through metatranscriptomic approach in present study was also compared with other publically accessible herbivore metagenomes and metatranscriptomes.

#### Metabolic pathway predictions from rumen metatranscriptomic data

The protein coding regions that were predicted from rumen metatranscriptomic dataset were analyzed using KEGG Mapper tool. The occurrence of enzymes that are part of different key metabolic pathways in carbohydrate metabolism (glycolysis, starch and sucrose metabolism) was predicted using KEGG-Mapper tool (Kanehisa et al., 2011).

## RESULTS AND DISCUSSION

### Metatranscriptome sequencing from cattle rumen digesta samples

Rumen-the foremost chamber in ruminant's stomach, encompasses a plethora of obligate anaerobic microorganisms including bacteria, archaea, fungi, and protozoa that are essential for the bioconversion of plant lignocellulosic biomass into simpler forms that can be utilized by the host (Deng et al., 2008). The highly complex and diverse microbiota present in the rumen has evolved into a symbiotic association with the host and can produce a battery of lignocellulolytic enzymes (CAZymes) that in turn hydrolyze the complex plant polysaccharides and release volatile fatty acids and microbial proteins which are the major sources of carbon, nitrogen, and energy for the host (Youssef et al., 2013). The microorganisms residing in the rumen execute numerous critical biochemical processes that are advantageous to the host. These include the fibrolytic responses and the mechanisms that change over the recalcitrant plant polymers into short-chain unsaturated fats and vitamins and furthermore the detoxification of secondary compounds in the plant feed to a specific degree. Plant lignocellulosic biomass encompasses cellulose and hemicellulose making it a rich source of renewable energy (Sheehan & Himmel, 1999). The primary objective of our study was to provide an insight into the transcripts involved in the deconstruction of plant cell wall polysaccharides and to identify the metabolically active bacterial communities residing in the rumen. The plateau shown in the rarefaction curve (Supplementary file 4) after cattle rumen metatranscriptome sequencing in present study indicates that the sample had been sequenced to adequate depth to categorize maximum number of metabolically active microbiota from the sample.

### Rumen metatranscriptome data statistics

The Next Generation Sequencing of the rumen metatranscriptome sample using Illumina HiSeq platform generated ~2.98 GB of raw sequences comprising about 17,101,702 paired-end reads accounting a total of 47,043,232 bases. The *de novo* assembly of the raw sequences was performed using two different algorithms: (i) CLC genomics workbench 7.0 and (ii) MetaVelvet program (Namiki et al., 2012) using the default parameters, except the minimum contig length. The minimum contig length was set at 100 bp in both the programs. The contigs generated from both the assembly tools were merged together and the redundant data was removed before further processing. The metatranscriptomic data statistics were generated using in house Perl scripts. Out of 143,666 contigs that generated after *de novo* assembly, 6,388 contigs (4.4%) failed the QC pipeline. Among these 1,431 sequences were identified as artificial duplicate reads and therefore removed from further analysis. Among the sequences that passed QC 1,419 sequences were identified as 16S rRNA gene sequences. A total of 133,930 orfs were predicted from the metatranscriptomic dataset using MetaGeneMark tool (Zhu et al., 2010) (Table 1).

**Table 1** Summary of NGS Illumina HiSeq metatranscriptomic sequencing data statistics

Sequencing parameters	Data statistics
Data in GB	2.98
No of bases	47,043,232
Total no of reads	17,101,702
Total no of processed reads	16,645,780
Mean GC content	52±10%
Artificial duplicate reads	1,431,431
Total no of contigs	143,666
Minimum Contig length	200
Post QC sequence count	137,278
Maximum Contig length	8,539
Average Contig length	327± 177 bp
Total Contig length	47,043,232 (47.0 MB)
Number of ORFs predicted	133,930
Contigs > 100 b	143,666
Contigs > 500 b	12,669
Contigs > 1 Kb	1,467
rRNA contigs	1,419

### Taxonomic mining using MEGAN

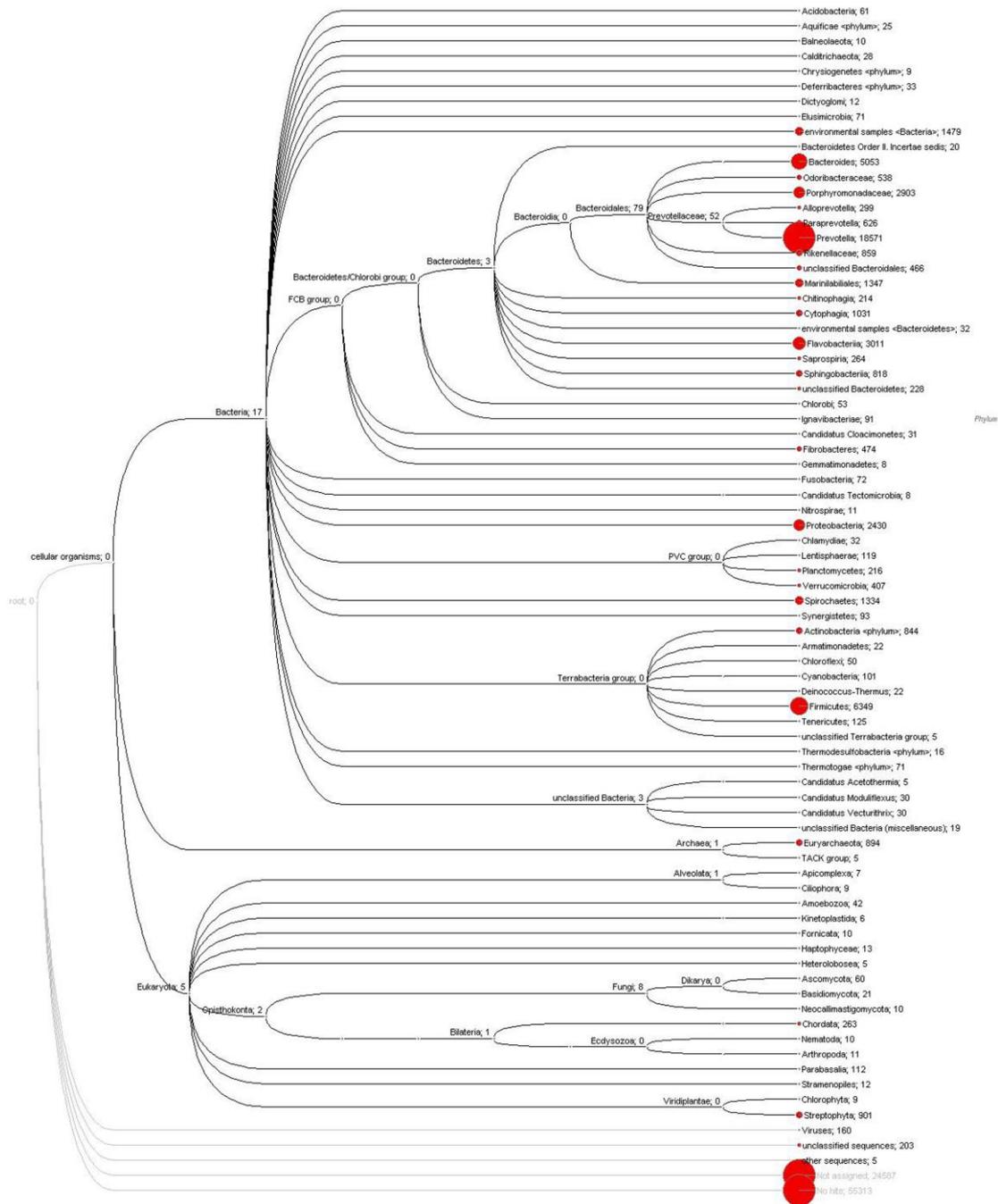
Generally, the 16S and 18S rRNA reads obtained during metatranscriptome sequencing is widely used for the taxonomic mining of microbial communities from diverse ecosystems. A similar strategy was also adopted by the earlier reports from rumen metatranscriptome studies (Dai et al., 2015; Comtet-Marre et al., 2017). The main reason for this can possibly be the presence of higher proportion of the rRNA reads in total metatranscriptomic data sets, despite performing either an mRNA enrichment or rRNA gene removal step. This is probably the first cattle rumen metatranscriptome study which reports ~99% of reads were corresponding to coding sequences (mRNAs) rather than rRNAs. This was accomplished using a modified and highly efficient total RNA isolation protocol adopted in present study from rumen digesta sample which ensured the isolation of intact rRNAs-the degraded rRNAs cannot be removed efficiently since the probe will not bind efficiently to degraded rRNAs- that were removed during mRNA enrichment, resulting in large number of mRNA transcript sequences. In a recent metatranscriptomic study on cow rumen by Dai et al. (2015), a total of 75% of mRNA reads were obtained, after performing an additional rRNA removal step. Due to the presence of substantially lower number of reads from rRNA genes in present study, they were excluded from further analysis and the coding sequences (mRNA reads) were used for the prediction of metabolically active cattle rumen microbiome using METaGenome ANalyzer (MEGAN) Community Edition.

The metabolically active rumen microbial communities were predicted from the orfs generated from rumen metatranscriptomic data, using MEGAN 6.0 Community Edition (Huson et al., 2016) (<https://ab.inf.uni-tuebingen.de/software/megan6>). A total of 133,930 protein coding sequences from rumen metadata were used for MEGAN microbial community analysis. At domain level 51,235 orfs were corresponding to bacteria, 900 orfs with archaea, 1,527 orfs with eukaryote, and 160 orfs were matching with viruses. A total of 203 orfs were falling under unclassified sequences whereas, the remaining 79,905 orfs could not be either assigned to any of the above class or had no similarity hits during MEGAN analysis (Supplementary file 5). The microbial community analysis of the annotated transcripts at phylum level could identify 17 different phyla under bacteria, 1 phylum under archaea and 5 phyla under eukaryota domains. MEGAN analysis revealed the abundance of bacteria belonging to phylum *Bacteroidetes* as the metabolically most active group of microorganisms in cattle rumen with 36,414 orfs corresponding to it. Similarly, phylum *Firmicutes* were the second most predominant phyla with 6,349 orfs followed by phylum *Proteobacteria* and *Spirochaetes* with 2,430 and 1,334 orfs, respectively, assigned to it.

In depth analysis of rumen metatranscriptomic data at genus level displayed the predominance of genera *Prevotella* (18,571 orfs), *Bacteroides* (5,053 orfs), *Porphyromonas* (680 orfs), *Flavobacterium* (658 orfs), *Parabacteroides* (640 orfs), and *Fibrobacter* (464 orfs) under phylum *Bacteroidetes*. Interestingly, 823 orfs were affiliated with genus *Treponema* under phylum *Spirochaetes*. Under the phylum *Firmicutes*, genera *Butyrivibrio*, *Ruminococcus* and *Clostridium* were observed as the predominant group of bacteria with 749, 703 and 407 orfs, respectively. Strikingly, a large number of transcripts were assigned with unclassified *Lachnospiraceae* family (821 orfs) under the order *Clostridiales* and phylum *Firmicutes*. Under phylum *Euryarchaeota*, genera *Methanoplanus* and *Methanolacinia* were the largest group with signifying 114 and 109 orfs, respectively (Figure 1).

Further MEGAN analysis of rumen metatranscriptome data revealed the metabolically active bacterial population in Indian crossbred cattle rumen at species level. Maximum numbers of orfs at species level were assigned to *Prevotella brevis* and *Prevotella ruminicola* with 4,022 and 2,725 orfs

respectively, consigned to it. Similarly, maximum number of orfs (275 orfs) was matching with *Bacteroides fragilis* under genus *Bacteroides*.



**Figure 1** Taxonomic affiliation of rumen metatranscriptomic dataset using MEGAN. The cladogram showing the microbial diversity of Indian crossbred cattle rumen metatranscriptome sample at genus level along with the number of transcripts

The protein coding regions from the assembled contigs were predicted using the GeneMark program and the taxonomic affiliation of predicted orfs from rumen metatranscriptomic datasets using RefSeq protein database at domain level indicated that bacteria are highly abundant in cattle rumen and the majority of transcripts identified in the present study were assigned to bacteria. This result was in consistent with other published metagenomic studies from the authors (Jose et al., 2017a; 2017b) and other published metatranscriptomic study from the rumen. The taxonomic binning of the predicted orfs using MEGAN Community Edition program unravelled the microbial communities that contribute these transcripts. The results obtained in present cattle rumen metatranscriptomic study are in parallel to the earlier metagenomic study conducted by the authors (Jose et al., 2017a) under the same feeding regimen were the maximum number of transcripts were corresponding to phylum *Bacteroidetes*. The results obtained from the metatranscriptomic analysis in our study authenticate the abundance of this phylum in cattle rumen. The maximum number of transcripts matching with phylum *Bacteroidetes* designates the role of different microbes belonging to this phylum in different metabolic activities in

the rumen. The MEGAN analysis of microbial diversity was performed up to the species level. Under phylum *Bacteroidetes*, genus *Prevotella* was the most abundant and at species level *Prevotella brevis* was the most metabolically involved organism followed by *Prevotella ruminicola*.

**Functional annotation of rumen metatranscriptome data**

To reconnoitre the distribution of transcripts that involved in different metabolisms and functional class, we assigned the contigs obtained from the cattle rumen metatranscriptomic data to COG database based on BLAST search. Out of the 137,278 contigs that were analyzed, 49,117 contigs (about 36%) were annotated by COG database and the remaining contigs did not match with any sequences in COG data base. Among the COG annotated transcripts (49,117), 88% of sequences were corresponding to 13 major functional classes. The functional annotation and classification of rumen metatranscriptomic dataset based on COG database showed that maximum numbers of transcripts were corresponding to translation, ribosomal structure and biogenesis (16% of

transcripts). Carbohydrate transport and metabolism was the second most prominent metabolism representing about 15% of sequences. Transcripts belonging to cell wall/membrane/envelope biogenesis and amino acid transport and metabolism represented 8% and 7% of total annotated transcripts, respectively. The pie chart depicting the COG functional class annotation and the percentage level involvement of all major transcripts in diverse metabolisms are shown in figure 2a.

To explore the potential metabolisms involved in the rumen microbiome, the assembled contigs generated from our study was annotated with COG database. The reads that were obtained from rRNA genes were excluded from COG annotation. The search was done by using BLASTX program against COG database. Less than 50% of sequences obtained from metatranscriptomic dataset could be annotated by COG database and the rest of the sequences could not assign any of the functional classes in COG database. This could be primarily because of the lack of an exclusive rumen microbial metatranscriptome database for the data annotation and analysis. We also assume that the sequences that could not be assigned to any functions might code or represent novel dataset that are not reported so far. Transcripts encoding translation, carbohydrate, and transport metabolism, and replication, recombination and repair were seen be the prominent functional classes that identified in present study. COG annotation of metatranscriptomic data is widely being used for human gut microbiome studies and quite surprisingly the above-mentioned metabolisms were also seen to be dominant in a human gut metatranscriptomic study conducted by Gosalbes et al. (2011).

The Gene Ontology (GO) annotation of the predicted orfs followed by WEGO analysis of GO terms revealed the diverse functional classes and categorized the transcripts into three major groups viz. cellular component, molecular function and biological process. Under the cellular component category, maximum numbers of transcripts were assigned to cell, cell part, macromolecular complex, and organelle related functions. Transcripts encoding the catalytic activity and binding functions were the most predominant under molecular functions set. Under the biological process category, the transcripts encoding metabolic process were prominent followed by cellular process and establishment of localization. WEGO analysis displayed that the maximum number of transcripts in cattle rumen was represented by catalytic activity and accounted about 16.2% of total predicted orfs from metatranscriptomic data (Figure 2b).

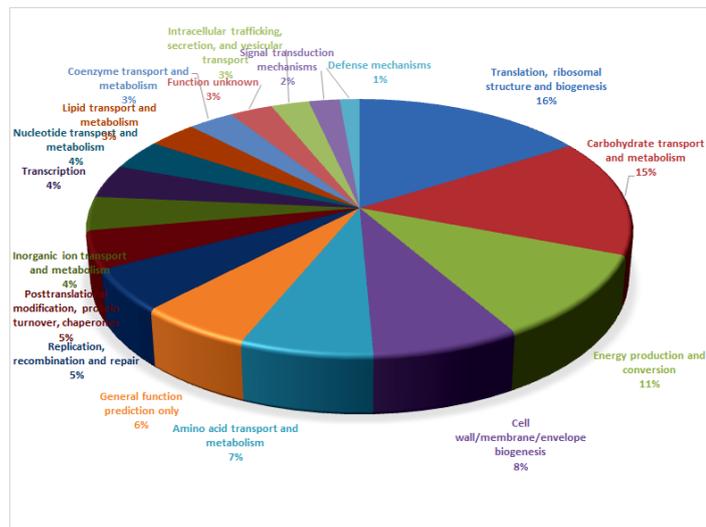


Figure 2a The functional annotation of rumen metatranscriptomic dataset

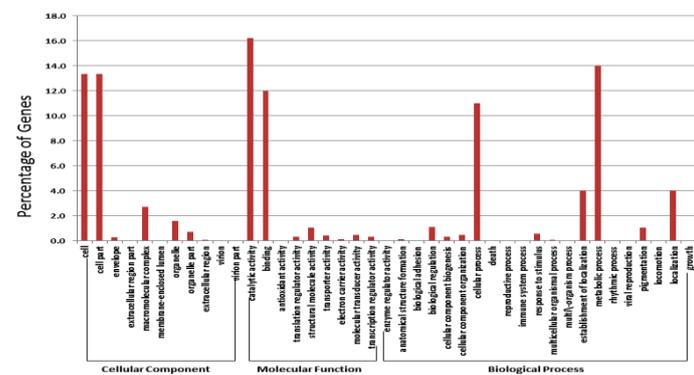


Figure 2b WEGO analysis of cattle rumen metatranscriptomic dataset

Carbohydrate-Active enZyme (CAZyme) annotation of transcripts

To identify and annotate the transcripts encoding different classes of CAZymes from the Indian crossbred cattle rumen metatranscriptome dataset by the Pfam based Hidden Markov Model (HMMs), the predicted open reading frames by the GeneMark program were analyzed using the dbCAN database (Yin et al., 2012) (<http://csbl.bmb.uga.edu/dbCAN/>) using the default parameters. The results obtained were manually analyzed further to find the abundance of each CAZyme class in the rumen metatranscriptomic dataset. The dbCAN annotation of metatranscriptomic dataset could identify a total of 5,267 transcripts belonging to 168 CAZyme families from different classes of CAZymes viz. glycoside hydrolases, glycosyltransferases, carbohydrate-binding modules, carbohydrate esterases, auxiliary activities, and polysaccharide lyases. Among these diverse CAZyme families that were identified and classified from the transcripts, GHs were the most abundant CAZyme class with 2,764 transcripts corresponding to 84 GH families and they accounted about 52% of total CAZymes from cattle rumen metatranscriptome. The transcripts belong to CAZyme class GTs were the second most abundant (1,346 transcripts) and 35 GT families were identified, and it represented about 26% of total CAZyme encoding transcripts. The transcripts encoding other major CAZymes classes like CBMs (23 families), CEs (12 families), PLs (8 families), and AAs (5 families) were represented by 738, 301, 108 and 10 transcripts, respectively (Supplementary file 7). Among the principal plant cell wall degrading GH families, 5 families (GH5, GH9, GH44, GH45, and GH48) are known for its exclusive cellulose degrading ability and these families signified a total of 1,466 transcripts in cattle rumen. Amongst these, the transcripts representing GH5 (108 transcripts) and GH9 (42 transcripts) families encoding endo-β-1, 4-glucanases that hydrolyze the β-1, 4-linked glucose residues from cellulose, were relatively high in number as compared to other cellulase encoding GH families. Transcripts encoding eleven GH families under the principal plant cell wall degrading category (GH8, GH10, GH11, GH12, GH26, GH28, GH51, GH53, GH54, GH67, and GH78) are efficient endo-hemicellulose degrading and debranching enzymes and a total of 410 transcripts were corresponded by these CAZyme families. dbCAN annotation revealed the presence of number of transcripts encoding α-l-arabinofuranosidases (GH51) (74 transcripts), α-glucuronidases (GH67) (9 transcripts) and, α-l-rhamnosidases (GH78) (30 transcripts), that efficiently involved in depolymerization of hemicellulose. A large number of transcripts (137 transcripts) encoding GH10 family and 8 transcripts encoding GH11 families representing endo-1,4-β-xylanase and endo-1, 3-β-xylanase that degrade the xylan main chains were also defined from the cattle rumen metatranscripts. Maximum number of principal plant cell wall degrading CAZymes encoding transcripts (about 60.5%) was identified under oligosaccharide degrading enzymes category and these were represented by 9 GH families (GH1, GH2, GH3, GH29, GH35, GH38, GH39, GH43 and GH94). Among these 3 GH families (GH2, GH3, and GH43) represented >50% of total CAZyme encoding transcripts that were identified in present study (Table 2). Although cellulose is chemically homogeneous in nature, it is highly resilient and cannot be hydrolyzed by a lone enzyme. It requires the synchronized or concerted action of an assemblage of enzymes or multi-enzyme complexes to break it down into simpler forms that can be utilized. Enzymatic degradation of cellulose is a slow and incomplete process in the rumen. The rumen microflora is competent of hydrolyzing only 60-65% of cellulose present in the feed (Flint et al., 2012). However, termites are affirmed to have almost 90% efficiency in breaking down this complex polymer (Breznak & Brune, 1994). In general, the process of cellulolysis is accomplished by the contemporaneous actions of three genre of enzymes: (i) endoglucanases (EC:3.2.1.4) that cleave the polymeric chain at random, (ii) exoglucanases that liberate D-glucose (EC:3.2.1.74) and D-cellobiose (EC:3.2.1.91) from β-glucan and cellodextrins, and (iii) β-glucosidases (EC:3.2.1.21) that release D-glucose from cellodextrins (Henrissat et al., 1998; Corpet et al., 1999).

The cellulases that are mainly involved in the deconstruction of plant polymers are belonging to different families and they have different substrate specificities and mechanism of action of action. The efficient degradation or hydrolysis of plant cell wall polymers requires the synergistic action of many CAZyme families altogether (Lynd et al., 2002). Not surprisingly, our study on rumen microbial metatranscriptomics could identify transcripts representing a wide range of CAZyme families and it showed some substantial difference with the previous published rumen metatranscriptomic studies (Table 2). To the best of our knowledge, any previous rumen metagenomic and metatranscriptomic studies have ever reported any contigs or orfs belonging GH12 family. This is the first rumen metatranscriptomic study where we report the metabolically active transcripts belonging to GH12 family encoding xylanoglucanases in Indian crossbred cattle rumen ecosystem. The major differences among the transcripts that were reported in all three studies could be mainly due to (i) the difference in the geographical location (ii) diet (iii) mode of sampling and processing of samples (iv) analysis pipeline followed. In the very first rumen metatranscriptomic study conducted by Qi et al. (2011) on Muskoxen rumen and the second report by Dai et al. (2015) on cow rumen did not report any transcripts belong to this family. Similarly, many transcripts belonging to GH28, GH51, and GH78 families, which are mainly concerned with the degradation of

endo-hemicellulose, are also recognized in our study. Likewise, a higher percentage of oligosaccharide degrading enzyme encoding families, GH2, GH3, GH29, and GH43 were also identified. The presence of high degree of endo-hemicellulose and oligosaccharide degrading enzyme encoding transcripts in our metatranscriptomic data set could be purely due to the dietary manipulations employed in our study and also due to the differences in sequencing platforms and advanced analysis pipelines used in this study. Broadly, our study based on rumen metatranscriptome sequencing could identify and annotate comparatively larger proportion of endo-hemicellulose and oligosaccharide degrading enzymes families, indicating their abundance in cattle rumen ecosystem. Most of these enzyme families are the primary degraders that attack the side chains of plant cell-wall polysaccharides.

In one of the most recent publications about cow rumen metatranscriptome, Comtet-Marre et al. (2017), reported 96 GH, 14 CE, and 13 PL families. Present study on Indian crossbred cattle rumen could identify 84 GH, 12 CE and 8 PL CAZyme families. In consistent with the earlier study, our samples also had the prevalence of CAZymes encoding transcripts targeting hemicellulose and oligosaccharides. The abundance of the later was also been reported in several rumen metagenomic studies and by previous published studies by the authors (Brulc et al., 2009; Hess et al., 2011; Jose et al., 2017a).

The lignolytic activities of most enzymes belong to diverse GH families is highly influenced by the catalytic modules CBM, as they can maintain the catalytic domain near the substrate. In the present study, a total of 23 CBMs were detected from cattle rumen metatranscriptome and among the CBMs identified CBM37 family represents the maximum number of transcripts (251). *Ruminococcus albus* is the only bacterial strain that is reported to have the ability to produce CBM37 enzymes and therefore facilitate the attachment of rumen bacteria to the substrate. Therefore, the large number transcripts from CBM37 family suggest the contribution of this bacterium in rumen fibrolytic activities.

Noticeably, the taxonomic assignment of the major CAZymes identified in the present study revealed that most of them were contributed by a limited number of genera and typically by the numerically prominent genera. The taxonomic affiliation of sequences encoding different cellulases, endo-hemicellulases, debranching and oligosaccharide degrading enzymes revealed that the microbes belonging to genera *Prevotella*, *Bacteroides*, *Ruminococcus*, *Spirosoma*, *Plaudibacter*, *Parabacteroides*, and *Fibrobacter* contribute a major proportion of these enzymes. These results are not surprising as the bacteria belonging to genera *Prevotella* and *Bacteroides* are proven to be numerically predominant and they encode diverse CAZyme families and therefore take part actively in the mechanism of plant cell-wall hydrolysis in the rumen. However, Dai et al. (2015) in their cow rumen metatranscriptomic study reported they found ~62% of cellulases were contributed by genera *Ruminococcus*, *Eubacterium*, and *Fibrobacteres*. Our results are contradictory to the earlier observation and were we report the contributions from microbes belong to these genera were in much lower percentages than genera *Prevotella* and *Bacteroides* in the Indian crossbred cattle rumen.

The chief plant cell wall-degrading transcripts that were identified in present cattle rumen metatranscriptomic study were compared with other published rumen metatranscriptomic and metagenomic datasets. Transcripts belonging to twenty-four GH families were identified from the HF cross rumen metatranscriptome that were inferred to be involved principally in cell wall polysaccharide degradation. The transcripts representative of GH43 family that encodes xylosidases was one of the largest enzyme families that were identified in present study (Table 2).

**Table 2** Comparative analysis of major plant cell wall degrading enzymes belonging to different GH families identified from HF cross rumen with other large herbivore metatranscriptomes. Putative GHs are divided into different groups based on their major enzyme activities

CAZyme family*	Major activity	HF cross rumen <sup>#</sup>	Holstein cow <sup>a</sup>	Muskoxen rumen <sup>b</sup>
<b>Cellulases</b>				
GH5	Cellulases	7.54	9.40	12.30
GH6	Endoglucanase	0.00	0.35	9.02
GH7	Endoglucanase	0.00	0.00	0.00
GH9	Endoglucanase	2.93	13.71	8.52
GH44	Endoglucanase	0.14	0.35	0.00
GH45	Endoglucanase	0.00	3.94	9.51
GH48	Cellobiohydrolases	0.28	7.65	8.85
<b>Subtotal</b>		10.89	35.40	48.20
<b>Endo-hemicellulases and</b>				

<b>debranching</b>				
GH8	endoxyylanase	1.33	1.11	0.98
GH10	endo-1,4-β-xylanases	9.56	8.48	11.64
GH11	xylanases	0.56	4.78	7.38
GH12	xyloglucanases	0.63	0.00	0.00
GH26	β-mannase and xylanases	0.63	3.12	2.13
GH28	galacturonases	6.56	1.26	0.16
GH53	endo-1,4-β-galactanases	1.05	1.84	0.66
GH51	α-L-arabinofuranosidases	5.16	1.50	0.16
GH54	α-L-arabinofuranosidases	0.42	0.08	0.00
GH62	α-L-arabinofuranosidases	0.00	0.00	0.00
GH67	α-glucuronidases	0.63	0.43	0.16
GH78	α-L-rhamnosidases	2.09	1.80	0.82
<b>Subtotal</b>		28.61	24.40	24.09
<b>Oligosaccharide degrading enzymes</b>				
GH1	β-glucosidases	1.19	5.07	3.93
GH2	β-galactosidases	13.54	6.01	1.97
GH3	β-glucosidases	14.24	12.20	8.20
GH29	α-L-fucosidases	2.72	1.64	0.00
GH35	β-galactosidases	2.58	0.67	0.00
GH38	α-mannosidases	0.49	0.47	0.16
GH39	β-xylosidases	1.33	0.37	0.00
GH42	β-galactosidases	0.00	0.79	0.00
GH43	Arabino/xylosidases	22.26	9.53	13.28
GH94	Cellobiose phosphorylase	2.16	3.45	0.17
<b>Subtotal</b>		60.50	40.20	27.71
<b>Total</b>		100	100	100

Table lists the percentages of different carbohydrate-active enzyme families targeting plant structural polysaccharides along with its principal activity. In addition to enzymes belonging to glycoside hydrolases table shows the numbers of two more proteins presumably involved in cellulosome mediated plant biomass conversion

\*CAZy database (<http://www.cazy.org>)

<sup>#</sup> Samples used in present study, <sup>a</sup>Dai et al., (2015), <sup>b</sup>Qi et al., (2011)

**Phylogenetic analysis of CAZyme encoding transcripts**

The primary plant cell wall degrading enzymes encoding orfs were extracted from the metatranscriptomic datasets using perl scripts and the microbial community analysis was performed using the M5NR database. A significant proportion of all these CAZyme classes were contributed by bacteria belonging to genera *Prevotella* and *Bacteroides*. About 37% of principal plant cell wall degrading CAZymes was contributed by genus *Prevotella* whereas 21% of CAZymes were contributed by genus *Bacteroides*. Present study also acknowledged the microbes belonging to the genus *Spirosoma* contributed about 3% of CAZymes, and genera *Paludibacter*, *Parabacteroides*, *Fibrobacter*, *Clostridium*, and *Opitutus* produced about 2% CAZymes each, in Indian crossbred cattle rumen. The figure showing the percentage level contributions principal plant cell-wall degrading CAZymes by different bacterial genera is given figure 3.

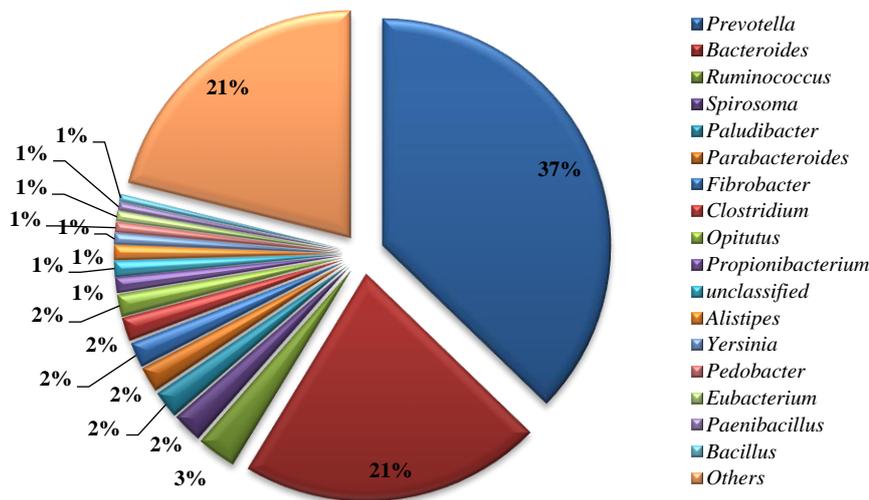


Figure 3 Taxonomic affiliation of major plant cell wall degrading enzyme encoding orfs using M5NR database

Metabolic pathways prediction using KEGG Mapper

The KEGG annotation and the metabolic pathway analysis using KEGG Mapper from the predicted orfs from the rumen metatranscriptomic dataset revealed the presence of different metabolic enzymes that were involved in the carbohydrate metabolism. The metabolic pathways and the enzymes involved in glycolysis,

starch and sucrose metabolisms are shown in figure 4a and 4b. The presence of transcripts involved in TCA cycle, pentose phosphate pathway, and oxidative phosphorylation were also predicted using the KEGG Mapper tool (Supplementary file 8a, 8b and 8c, respectively). The metabolic pathway analysis revealed that the Indian crossbred cattle rumen possess wide range of enzymes that actively take part in diverse metabolic pathways.

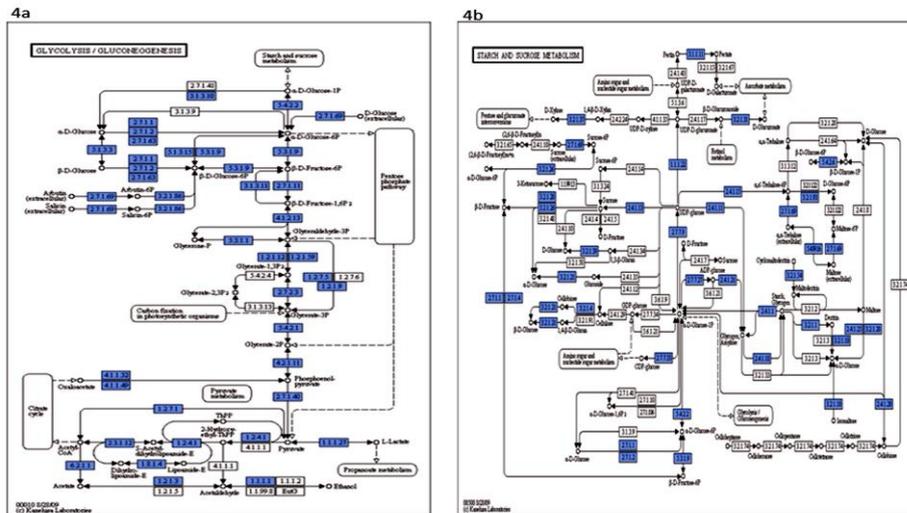


Figure 4a & 4b The KEGG mapper analysis indicating the occurrence of different key enzymes involved in glycolysis and starch and sucrose metabolisms

CONCLUSION

In nature, plant lignocellulose biomass remains mostly as an unexploited source of renewable energy presumably due to its recalcitrant nature. Currently, only limited understanding is available about the metabolically active rumen microbiome and transcripts involved in deconstruction of plant polysaccharides from Indian crossbred cattle. The comprehensive rumen metatranscriptomic approach used in present study brings an insight into the naturally active and functionally dynamic rumen microflora and its fermentation strategies in cattle rumen. Present study reports that, the bacteria belonging to phylum *Bacteroidetes* and *Firmicutes* represent a significant proportion of functionally active rumen microbiome and *Prevotella brevis* and *Prevotella ruminicola* as the most metabolically active rumen microbes residing in Indian crossbred cattle rumen. Unravelling the complex enzymatic phenomena behind the plant lignocellulose bioconversion is a formidable challenge and hence the functional transcripts corresponding to a large number of CAZyme families (168) reported in this study can be further reconnoitred for its potential applications in various industries and also for the development of a sustainable bio-based economy.

**Acknowledgments:**The authors are thankful to late Dr. A Thulasi for providing the technical inputs to conduct the research work. The authors also would like to acknowledge Jain University, Bangalore for supporting the research program.

REFERENCES

Azman, S., Khadem, A. F., Van Lier, J. B., Zeeman, G., & Plugge, C. M. (2015). Presence and role of anaerobic hydrolytic microbes in conversion of lignocellulosic biomass for biogas production. *Critical Reviews in Environmental Science and Technology*, 45(23), 2523-2564. doi.org/10.1080/10643389.2015.1053727

Brule, J. M., Antonopoulos, D. A., Miller, M. E. B., Wilson, M. K., Yannarell, A. C., Dinsdale, E. A., ... & Coutinho, P. M. (2009). Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proceedings of the National Academy of Sciences*, 106(6), 1948-1953. doi: 10.1073/pnas.0806191105

Carrillo-Reyes, J., Barragán-Trinidad, M., & Buitrón, G. (2016). Biological pretreatments of microalgal biomass for gaseous biofuel production and the potential use of rumen microorganisms: A review. *Algal research*, 18, 341-351. doi.org/10.1016/j.algal.2016.07.004

Cheng, K. J., Forsberg, C. W., Minato, H., & Costerton, J. W. (1991). Microbial ecology and physiology of feed degradation within the rumen. In *Physiological aspects of digestion and metabolism in ruminants* (pp. 595-624). Academic Press. doi.org/10.1007/978-94-009-1453-7\_12

Comtet-Marre, S., Parisot, N., Laperçq, P., Chaucheyras-Durand, F., Mosoni, P., Peyretailade, E., ... & Forano, E. (2017). Metatranscriptomics reveals the active bacterial and eukaryotic fibrolytic communities in the rumen of dairy cow fed a mixed diet. *Frontiers in microbiology*, 8, 67. DOI: 10.3389/fmicb.2017.00067

- Corpet, F., Gouzy, J., & Kahn, D. (1999). Recent improvements of the ProDom database of protein domain families. *Nucleic Acids Research*, 27(1), 263-267. doi: [10.1093/nar/27.1.263](https://doi.org/10.1093/nar/27.1.263)
- Dai, X., Tian, Y., Li, J., Su, X., Wang, X., Zhao, S., ... & Wang, J. (2015). Metatranscriptomic analyses of plant cell wall polysaccharide degradation by microorganisms in the cow rumen. *Appl. Environ. Microbiol.*, 81(4), 1375-1386. doi: [10.1128/AEM.03682-14](https://doi.org/10.1128/AEM.03682-14)
- Deng, W., Xi, D., Mao, H., & Wanapat, M. (2008). The use of molecular techniques based on ribosomal RNA and DNA for rumen microbial ecosystem studies: a review. *Molecular biology reports*, 35(2), 265-274. doi: [10.1007/s11033-007-9079-1](https://doi.org/10.1007/s11033-007-9079-1)
- Fernando, S. C., Purvis, H. T., Najar, F. Z., Sukharnikov, L. O., Krehbiel, C. R., Nagaraja, T. G., ... & DeSilva, U. (2010). Rumen microbial population dynamics during adaptation to a high-grain diet. *Appl. Environ. Microbiol.*, 76(22), 7482-7490. doi: [10.1128/AEM.00388-10](https://doi.org/10.1128/AEM.00388-10)
- Flint, H. J., Scott, K. P., Duncan, S. H., Louis, P., & Forano, E. (2012). Microbial degradation of complex carbohydrates in the gut. *Gut microbes*, 3(4), 289-306. doi: [10.4161/gmic.19897](https://doi.org/10.4161/gmic.19897)
- Galperin, M. Y., Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic acids research*, 43(D1), D261-D269. doi: [10.1093/nar/gku1223](https://doi.org/10.1093/nar/gku1223)
- Garg, M. R., Sherasia, P. L., Bhandari, B. M., Phondba, B. T., Shelke, S. K., & Makkar, H. P. S. (2013). Effects of feeding nutritionally balanced rations on animal productivity, feed conversion efficiency, feed nitrogen use efficiency, rumen microbial protein supply, parasitic load, immunity and enteric methane emissions of milking animals under field conditions. *Animal Feed Science and Technology*, 179(1-4), 24-35. doi: [doi.org/10.1016/j.anifeedsci.2012.11.005](https://doi.org/10.1016/j.anifeedsci.2012.11.005)
- Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., ... & Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic acids research*, 36(10), 3420-3435. doi: [10.1093/nar/gkn176](https://doi.org/10.1093/nar/gkn176)
- Henrissat, B., Teeri, T. T., & Warren, R. A. J. (1998). A scheme for designating enzymes that hydrolyse the polysaccharides in the cell walls of plants. *Febs Letters*, 425(2), 352-354. doi: [10.1016/s0014-5793\(98\)00265-8](https://doi.org/10.1016/s0014-5793(98)00265-8)
- Hess, M., Sczyrba, A., Egan, R., Kim, T. W., Chokhawala, H., Schroth, G., ... & Mackie, R. I. (2011). Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science*, 331(6016), 463-467. doi: [10.1126/science.1200387](https://doi.org/10.1126/science.1200387)
- Huson, D. H., Beier, S., Flade, I., Górski, A., El-Hadidi, M., Mitra, S., ... & Tappu, R. (2016). MEGAN community edition-interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS computational biology*, 12(6). doi: [doi.org/10.1371/journal.pcbi.1004957](https://doi.org/10.1371/journal.pcbi.1004957)
- Breznak, J. A., & Brune, A. (1994). Role of microorganisms in the digestion of lignocellulose by termites. *Annual review of entomology*, 39(1), 453-487. doi: [doi.org/10.1146/annurev.en.39.010194.002321](https://doi.org/10.1146/annurev.en.39.010194.002321)
- Jose, V. L., Appoohy, T., More, R. P., & Arun, A. S. (2017b). Metagenomic insights into the rumen microbial fibrolytic enzymes in Indian crossbred cattle fed finger millet straw. *AMB Express*, 7(1), 13. doi: [10.1186/s13568-016-0310-0](https://doi.org/10.1186/s13568-016-0310-0)
- Jose, V. L., More, R. P., Appoohy, T., & Arun, A. S. (2017a). In depth analysis of rumen microbial and carbohydrate-active enzymes profile in Indian crossbred cattle. *Systematic and applied microbiology*, 40(3), 160-170. doi: [10.1016/j.syapm.2017.02.003](https://doi.org/10.1016/j.syapm.2017.02.003)
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., & Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic acids research*, 40(D1), D109-D114. doi: [10.1093/nar/gkr988](https://doi.org/10.1093/nar/gkr988)
- Kelly, W. J., Li, D., Lambie, S. C., Cox, F., Attwood, G. T., Altermann, E., & Leahy, S. C. (2016). Draft genome sequence of the rumen methanogen *Methanobrevibacter olleyae* YLM1. *Genome Announc.*, 4(2), e00232-16. doi: [10.1128/genomeA.00232-16](https://doi.org/10.1128/genomeA.00232-16)
- Li, F. (2017). Metatranscriptomic profiling reveals linkages between the active rumen microbiome and feed efficiency in beef cattle. *Appl. Environ. Microbiol.*, 83(9), e00061-17. doi: [doi.org/10.1128/AEM.00061-17](https://doi.org/10.1128/AEM.00061-17)
- Li, R. W. (2015). Rumen metagenomics. In *Rumen Microbiology: From Evolution to Revolution* (pp. 223-245). Springer, New Delhi. doi: [doi.org/10.1007/978-81-322-2401-3](https://doi.org/10.1007/978-81-322-2401-3)
- Lynd, L. R., Weimer, P. J., Van Zyl, W. H., & Pretorius, I. S. (2002). Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol. Mol. Biol. Rev.*, 66(3), 506-577. doi: [10.1128/mmbr.66.3.506-577.2002](https://doi.org/10.1128/mmbr.66.3.506-577.2002)
- Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., ... & Wilkening, J. (2008). The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC bioinformatics*, 9(1), 386. doi: [10.1186/1471-2105-9-386](https://doi.org/10.1186/1471-2105-9-386)
- Namiki, T., Hachiya, T., Tanaka, H., & Sakakibara, Y. (2012). MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic acids research*, 40(20), e155-e155. doi: [doi.org/10.1093/nar/gks678](https://doi.org/10.1093/nar/gks678)
- Nathani, N. M., Kothari, R. K., Patel, A. K., & Joshi, C. G. (2015). Functional characterization reveals novel putative coding sequences in *Prevotella ruminicola* genome extracted from rumen metagenomic studies. *Journal of molecular microbiology and biotechnology*, 25(4), 292-299. doi: [10.1159/000437265](https://doi.org/10.1159/000437265)
- Guan, L. L., Nkrumah, J. D., Basarab, J. A., & Moore, S. S. (2008). Linkage of microbial ecology to phenotype: correlation of rumen microbial ecology to cattle's feed efficiency. *FEMS microbiology letters*, 288(1), 85-91. doi: [10.1111/j.1574-6968.2008.01343.x](https://doi.org/10.1111/j.1574-6968.2008.01343.x)
- O'Brien, M., Navarro-Villa, A., Purcell, P. J., Boland, T. M., & O'Kiely, P. (2014). Reducing in vitro rumen methanogenesis for two contrasting diets using a series of inclusion rates of different additives. *Animal production science*, 54(2), 141-157. doi: [10.1071/AN12204](https://doi.org/10.1071/AN12204)
- Pitta, D. W., Indugu, N., Kumar, S., Vecchiarelli, B., Sinha, R., Baker, L. D., ... & Ferguson, J. D. (2016). Metagenomic assessment of the functional potential of the rumen microbiome in Holstein dairy cows. *Anaerobe*, 38, 50-60. doi: [doi.org/10.1016/j.anaerobe.2015.12.003](https://doi.org/10.1016/j.anaerobe.2015.12.003)
- Purushe, J., Fouts, D. E., Morrison, M., White, B. A., Mackie, R. I., Coutinho, P. M., ... & North American Consortium for Rumen Bacteria. (2010). Comparative genome analysis of *Prevotella ruminicola* and *Prevotella bryantii*: insights into their environmental niche. *Microbial ecology*, 60(4), 721-729. doi: [10.1007/s00248-010-9692-8](https://doi.org/10.1007/s00248-010-9692-8)
- Qi, M., Wang, P., O'Toole, N., Barboza, P. S., Ungerfeld, E., Leigh, M. B., ... & Forster, R. J. (2011). Snapshot of the eukaryotic gene expression in muskoxen rumen—a metatranscriptomic approach. *PloS one*, 6(5). doi: [10.1371/journal.pone.0020521](https://doi.org/10.1371/journal.pone.0020521)
- Ribeiro, G. O., Gruninger, R. J., Badhan, A., & McAllister, T. A. (2016). Mining the rumen for fibrolytic feed enzymes. *Animal Frontiers*, 6(2), 20-26. doi: [doi.org/10.2527/af.2016-0019](https://doi.org/10.2527/af.2016-0019)
- Sheehan, J., & Himmel, M. (1999). Enzymes, energy, and the environment: a strategic perspective on the US Department of Energy's research and development activities for bioethanol. *Biotechnology Progress*, 15(5), 817-827. doi: [doi:10.1021/bp990110d](https://doi.org/10.1021/bp990110d)
- Wang, L., Hatem, A., Catalyurek, U. V., 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). doi: [doi.org/10.1371/journal.pone.0078507](https://doi.org/10.1371/journal.pone.0078507)
- Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., ... & Wang, J. (2006). WEGO: a web tool for plotting GO annotations. *Nucleic acids research*, 34(suppl\_2), W293-W297. doi: [10.1093/nar/gkl031](https://doi.org/10.1093/nar/gkl031)
- Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., & Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic acids research*, 40(W1), W445-W451. doi: [10.1093/nar/gks479](https://doi.org/10.1093/nar/gks479)
- Youssef, N. H., Couger, M. B., Struchtemeyer, C. G., Ligenstoffer, A. S., Prade, R. A., Najar, F. Z., ... & Elshahed, M. S. (2013). The genome of the anaerobic fungus *Orpinomyces* sp. strain C1A reveals the unique evolutionary history of a remarkable plant biomass degrader. *Appl. Environ. Microbiol.*, 79(15), 4620-4634. doi: [10.1128/AEM.00821-13](https://doi.org/10.1128/AEM.00821-13)
- Zhao, K., & Chu, X. (2014). G-BLASTN: accelerating nucleotide alignment by graphics processors. *Bioinformatics*, 30(10), 1384-1391. doi: [doi.org/10.1093/bioinformatics/btu047](https://doi.org/10.1093/bioinformatics/btu047)
- Zhou, M. I., & Hernandez-Sanabria, E. (2009). Assessment of the microbial ecology of ruminal methanogens in cattle with different feed efficiencies. *Appl. Environ. Microbiol.*, 75(20), 6524-6533. doi: [10.1128/AEM.02815-08](https://doi.org/10.1128/AEM.02815-08)
- Zhu, W., Lomsadze, A., & Borodovsky, M. (2010). Ab initio gene identification in metagenomic sequences. *Nucleic acids research*, 38(12), e132-e132. doi: [doi.org/10.1093/nar/gkq275](https://doi.org/10.1093/nar/gkq275)