Research Article

Journal of Molecular Microbiology and Biotechnology

J Mol Microbiol Biotechnol 2015;25:292–299
DOI: 10.1159/000437265

Published online: August 22, 2015

# Functional Characterization Reveals Novel Putative Coding Sequences in *Prevotella ruminicola* Genome Extracted from Rumen Metagenomic Studies

Neelam M. Nathani[a, b] Ramesh K. Kothari[a] Amrutlal K. Patel[b]
Chaitanya G. Joshi[b]

[a]Department of Biosciences, Saurashtra University, Rajkot, and [b]Department of Animal Biotechnology, College of Veterinary Science and Animal Husbandry, Anand Agricultural University, Anand, India



**Key Words**
Rumen microbiome · Bacteroidetes · *Prevotella ruminicola* · Carbohydrate active enzymes · Glycoside hydrolases

**Abstract**
***Aim:*** To reassemble *Prevotella ruminicola* genome from rumen metagenomic data of cattle and buffalo and compare with the published reference genome. ***Method:*** Rumen microbial communities from Mehsani buffaloes (n = 8) and Kankrej cattle (n = 8), each adapted to different proportions of a dry or green roughage diet, were subjected to metagenomic sequencing by Ion Torrent PGM, and subsequent reads were analyzed by MG-RAST. Using reference-guided assembly of the sequences against the published *P. ruminicola* strain 23, draft genomes of 2.56 and 2.46 Mb were reconstructed from Mehsani buffalo and Kankrej cows, respectively. The genomes were annotated using the RAST Server and carbohydrate active enzyme (CAZyme) analysis. ***Results:*** Taxonomic analysis by MG-RAST revealed *P. ruminicola* to be the most abundant species present among the rumen microflora. Functional annotation of reconstructed genomes using the RAST Server depicted the maximum assignment of coding sequences involved in the subsystems amino acid and derivatives and carbohydrate metabolism. CAZyme profiling revealed the glycoside hydrolases (GH) family to be the most abundant. GH family subclassification revealed that the extracted genomes had more sequence hits for GH2, GH3, GH92 and GH97 as compared to the reference. ***Conclusion:*** The results reflect the metabolic significance of rumen-adapted *P. ruminicola* in utilizing a coarse diet for animals based on acquisition of novel genetic elements.



## Introduction

Increased availability of genomic data has led to a concomitant increase in the area of comparative studies for exploring the microbial diversity of complex communities and diverse environmental niche like animal and human hosts. Next-generation sequencing technology is playing a major role in modernizing the study of natural microbial communities [Konstantinidis et al., 2009; Quince et al., 2009]. Due to user-friendly Web-based bioinformatic tool analysis and visualization of large-scale data, outputs by next-generation sequencing have been


1464–1801/15/0254–0292$39.50/0

E-Mail karger@karger.com
www.karger.com/mmb

Prof. Dr. Ramesh K. Kothari
Department of Biosciences
Saurashtra University
Rajkot, Gujarat 360005 (India)
E-Mail rkkothari1971@gmail.com, kothari1971@gmail.com

simplified. As the numbers of available whole genome sequences continue to increase, comparative genomics approaches facilitate the identification of candidate homologous genes, their functions and evolutionary studies. Genome comparison is also applied to find genes unique in a given species incurring phenotypic variation to the organism [Lu et al., 2006]. Recently, the drastic cost reduction of whole genome shotgun sequencing and assembly techniques has resulted in rapid enhancement of genome sequencing research [Margulies et al., 2005]. Such development in sequencing of bacterial genomes provides further access to numerous genomes of closely related species, which are further analyzed for phylogenetic, functional and genetic inferences.

When analyzing the metagenomic sequences, assembling an individual genome is a challenging issue for environmental studies. The presence of closely related species in samples complicates the assembly of a single genotype. One of the major objectives of metagenomic studies is to recover a complete or draft genome sequence of a species. Next-generation sequencing technologies provide high-throughput data; however, using these technologies to robustly recover individual genomes from complex communities requires several improvements. Nowadays, several studies have been able to recover novel bacterial and viral genomes from metagenomic studies on bioreactors and complex environments; however, such work on the rumen environment is scarcely reported due to the extremely diverse population of the rumen microbes.

Ruminants harbor a diverse and dense microbial population in the rumen, which play a vital role in the bioconversion of feeds for the host digestive system. This complex microbial community is comprised of bacteria, protozoa, fungi [Nathani et al., 2013; Singh et al., 2012], methanogenic archaea [Morvan et al., 1996] and bacteriophages [Klieve and Bauchop, 1988]. Among these, the bacteria are most abundant and perform a considerable part of the degradation of plant fiber [Koike and Kobayashi, 2009]. Studies including comparative sequence analysis of rumen bacterial 16S rRNA gene clone libraries have shown the dominance of two phyla in the rumen: low GC Gram-positive bacteria and the *Bacteroides* group. Within the *Bacteroides* group, *Prevotella*-related sequences were found to be predominantly associated with the rumen [Koike et al., 2003; Whitford et al., 1998]. Earlier studies based on 16S rRNA gene analysis of rumen bacterial diversity revealed that *Prevotella ruminicola* sequences are the single most abundant operational taxonomic unit [Edwards et al., 2004]. The *Prevotella* genus involves a group of bacteria that are less characterized, highlighting their role in the gastrointestinal tract, providing a need for further genetic characterization of these species [Purushe et al., 2010]. It includes four characterized rumen *Prevotella* species: *P. ruminicola*, *P. bryantii*, *P. albensis* and *P. brevis* [Avgustin et al., 1997]. Cultivated rumen *Prevotella* strains have been reported to exhibit a higher degree of genetic divergence [Mannarelli et al., 1991; Ramsak et al., 2000].

Furthermore, *P. ruminicola* synergistically participate in plant cell wall degradation along with cellulolytic bacteria [Osborne and Dehority, 1989]. Previously, culture-based studies have shown the abundance of *Prevotella* (60%) in total cultivable bacteria from the cow rumen [Van Gylswyk, 1990]. Restriction enzyme profiling has reported the relative abundance of rumen *Prevotella/Bacteroides* ribotypes in the total 16S rRNA gene as ranging from 12 to 62% [Wood et al., 1998]. Recent qPCR-based quantification studies have also shown that *Prevotella* was present in 12–60% of the total bacteria in the rumen [Stevenson and Weimer, 2007]. Based on the phenogenotypic diversity of cultured *Prevotella* spp., it is likely that differences at the functional level among the uncultured *Prevotella* occur [Bekele et al., 2010]. India is known to possess several indigenous breeds of cattle and buffalo. We considered Mehsani buffalo and Kankrej cattle breeds for *P. ruminicola* species study. Both of the breeds are native to North Gujarat localities, whereby the breeding tracts of Mehsani buffalo include the Banaskantha and Sabarkantha districts, while the Kankrej cattle are a native of the Kankrej town of Banaskantha district. The climate of the region is usually arid to semiarid with an average annual rainfall of about 700–800 mm and a temperature range of 11–27°C during winters and 28–41°C in summers. The Mehsani breed is considered to be a cross between Surti and Murrah buffalo and is famous as a persistent milker and regular breeder. Kankrej cattle are considered one of the heaviest breeds in India and are well known for milk and draught capabilities, which make it a very suitable livestock choice in India.

The major aim of the present study was to extract a draft genome of the *P. ruminocola* species from the metagenomic studies of the Kankrej breed of cattle (n = 8) and Mehsani breed of buffalo (n = 8) fed on varied proportions of roughage and concentrate diet. The genomes thus extracted were mined for the presence of various carbohydrate active enzymes (CAZymes). The resultant hits obtained were compared with each other and also with the already published complete genome of *P. ruminicola* strain 23 for genome-specific traits present in the extracted genomes due to host specificity, adaptation and other variations beneficial to the hosts metabolic functioning.

**Table 1.** Mapping statistics of the two partial genomes extracted from metagenomic data of 8 Mehsani buffalo and 8 Kankrej cattle

| Extracted draft genome | Input reads | Input bases | Contigs, n | Genome size, Mb | Largest contig, bp | $N_{50}$ contig size |
|---|---|---|---|---|---|---|
| PRM | 20406103 | 3.26 Gb | 3,871 | 2.56 | 18,022 | 1,964 |
| PRK | 21944989 | 4.60 Gb | 3,683 | 2.46 | 14,745 | 2,007 |

**Table 2.** Genome features of extracted draft genomes PRM and PRK as predicted by RAST

| Extracted draft genome | Genome size, Mb | GC content, % | Protein-encoding genes, n | RNA, n | Subsystems Involved, n |
|---|---|---|---|---|---|
| PRM | 2.56 | 49.1 | 3,546 | 87 | 227 |
| PRK | 2.46 | 49.6 | 3,122 | 88 | 216 |

## Results

*Metagenomic Shotgun Sequencing*

Metagenomic DNA with a concentration of >100 ng/μl was further processed for library preparation and sequencing. Out of a total of 12 shotgun sequencing runs (4 samples per run), 6 of each breed generated 3.26 and 4.6 Gb of data from Kankrej and Mehsani rumen metagenomics. These were analyzed using MG-RAST for bacterial taxonomic assignments to samples of each breed. We could observe that at the phyla level, Bacteroidetes and Firmicutes were abundantly represented in all of the samples (online suppl. fig. 2; for all online suppl. material, see www.karger.com/doi/10.1159/000437265). In liquid fractions, Bacteroidetes dominated by accounting for as much as 35–53% of the total bacteria, while in the majority of the solid fractions, Firmicutes dominated with about 29–45%, followed by a slightly lesser percentage of Bacteroidetes (28–44%). When further classification was studied, we observed a major abundance for all 48 samples of each breed as follows: class, Bacteroidia (27–35%); order, Bacteroidales (27–34%); family: Prevotellaceae (12–20%), and genus, *Prevotella* (12–75%). The metagenomic results showed the maximum hits for *P. ruminicola* species (>70% of total *Prevotella*) for all the samples in cattle as well as buffalo (online suppl. fig. 3).

*Draft Genome Reconstruction*

Reference-guided assembly using GS Reference Mapper V2.6 of the metagenomic reads from shotgun sequencing of the Mehsani (PRM) and Kankrej (PRK) rumen metagenomes against the reference genome of *P. ruminicola* strain 23 (PR23) resulted in draft genome reassembly of *P. ruminicola* with a size of 2.56 and 2.46 Mb, respectively (detailed mapping statistics in table 1).

The entire draft genome had a total GC content of about 49%. RAST annotation showed that the genomes comprised more than 3,000 protein-encoding sequences (table 2). Out of these, 40% [1,415 coding sequences (CDS) in PRM and 1,224 CDS in PRK] were classified into various defined subsystems, with 24 and 28 hypothetical proteins in PRM and PRK, respectively. Among the classified CDS, both PRM and PRK showed major CDS hits for enzymes involved in the subsystems amino acids and derivatives and the carbohydrate metabolism (fig. 1). Further, to get a better understanding of the host-specific species PRM and PRK, we studied the CDS for metabolic pathway-related features that might be specific to the extracted genomes because of host adaptation (table 1). We could observe CDS for enzymes utilizing various monosaccharide and oligosaccharide molecules like arabinogalactans.

*CAZyme Prediction*

As the functional annotation depicted the carbohydrate metabolism to be most abundantly represented, we further proceeded for mining the carbohydrate-associated enzymes present in *P. ruminicola* of the rumen environment using a Pfam-HMM-based search to identify various CAZymes. All the assembled contigs translated into 6 frames and were uploaded for CAZy annotation, which resulted in a total of 1,551 hits with an average of 500 hits per genome, i.e. PRM, PRK and PR23. After nor-

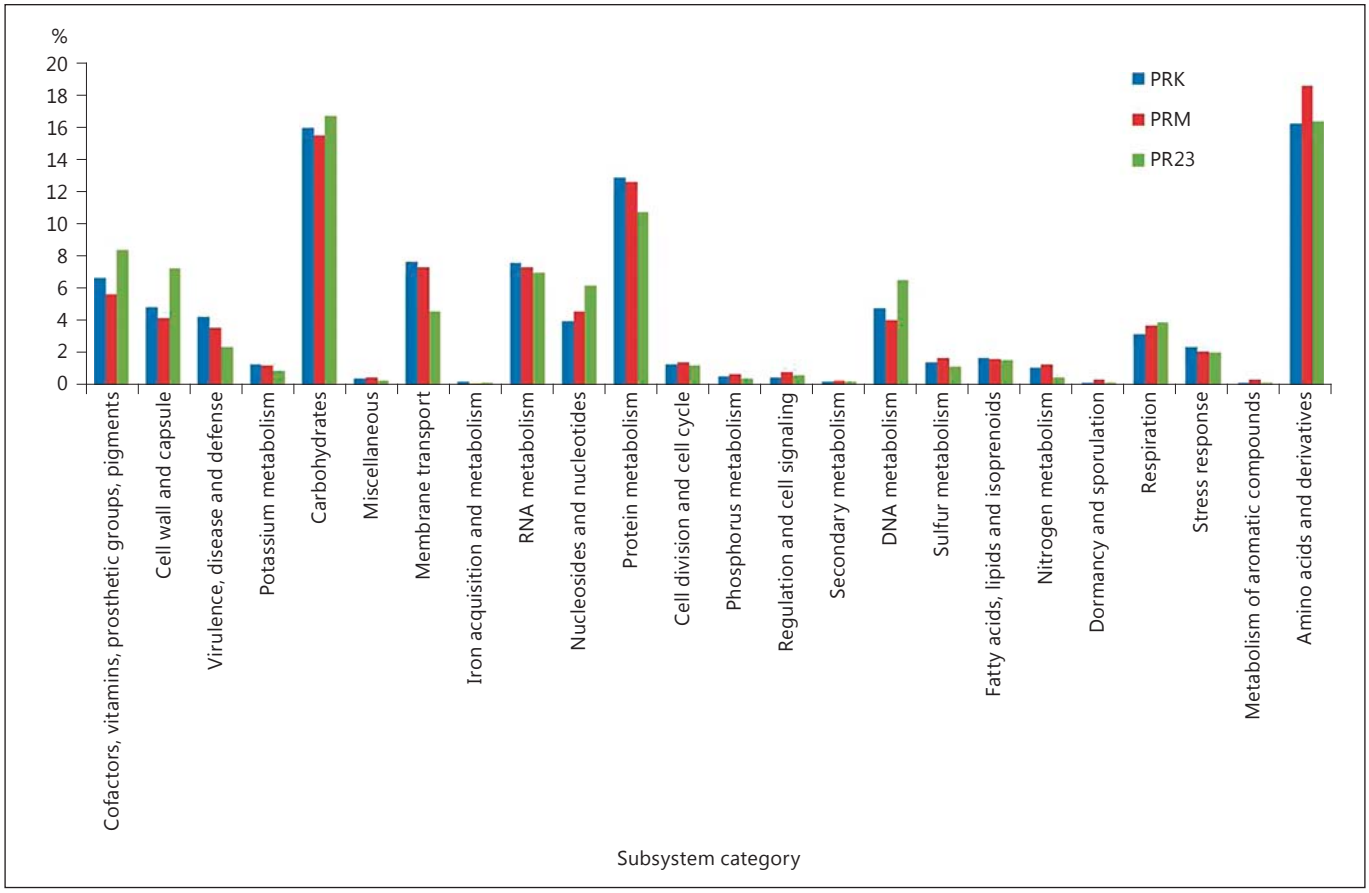


**Fig. 1.** Subsystem category distribution as per RAST annotation results.

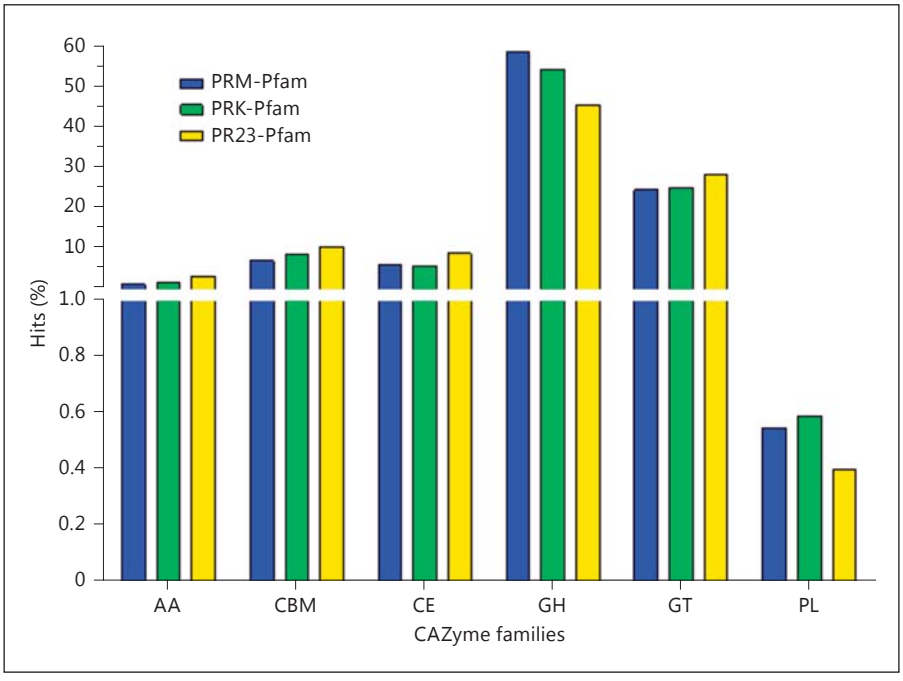


**Fig. 2.** Pfam-based CAZyme family classification. GT = Glycosyltransferases; CE = carboxylesterase; PL = polysaccharide lyase, AA = auxiliary activities; CBM = carbohydrate-binding modules.

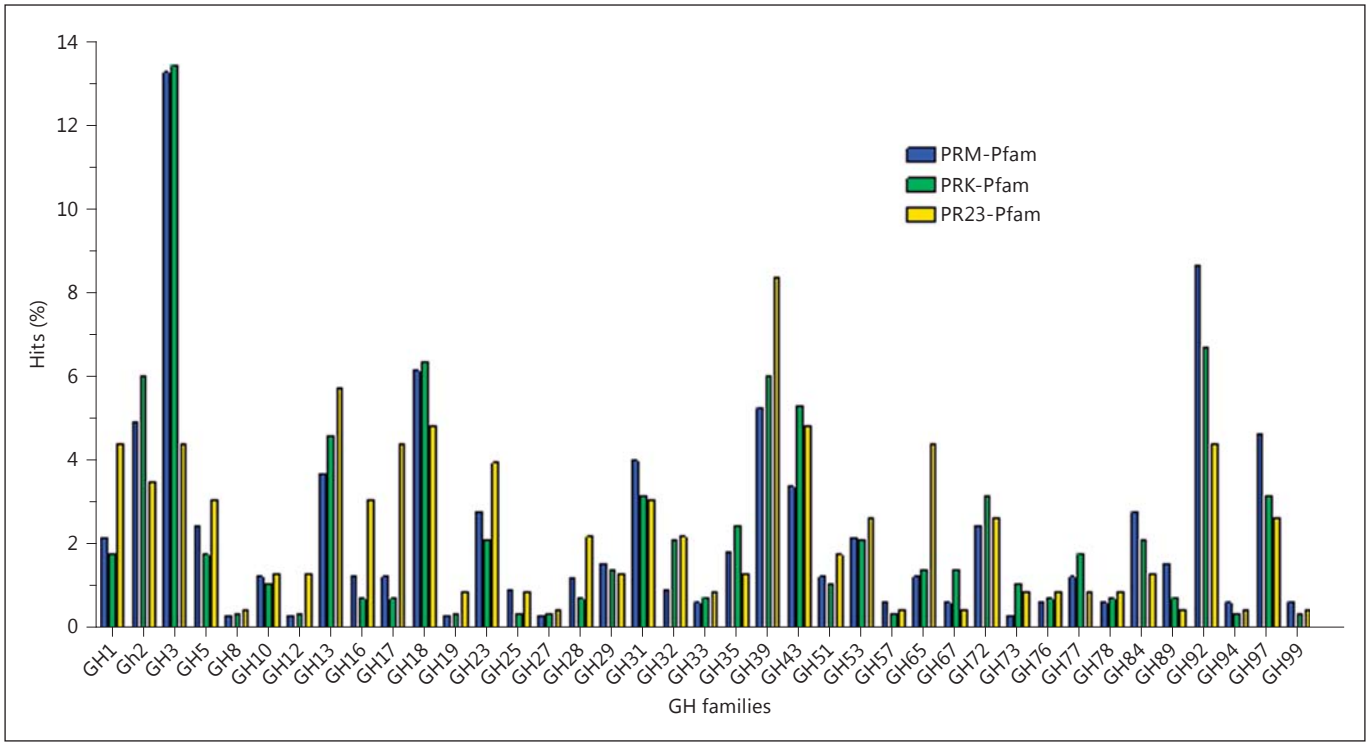


**Fig. 3.** GH family distribution of hits.

malizing the results, we could observe that the glycoside hydrolases (GH) group dominated followed by other enzyme families such as glycosyltransferases, carboxylesterase, polysaccharide lyase, auxiliary activities and carbohydrate-binding modules (fig. 2).

Hits in the GH corresponded to 39 different families with GH3 being the most abundant family followed by GH92 and GH18; a more than 3% presence was shown by the GH2, GH13, GH31, GH39, GH43 and GH97 families (fig. 3). We also observed that the GH2, GH3, GH18, GH31, GH77, GH84, GH89, GH92 and GH97 families were abundant in the extracted genomes as compared to the published reference genome (fig. 3). Overall, the results showed that the draft genomes comprised a higher proportion (~10%) of GH family encoding sequences as compared to the published reference genome.

## Discussion

*Metagenomic Shotgun Sequencing*

The dominance of the Bacteroidetes and the Firmicutes phyla observed in the metagenomic results was in accordance with a number of earlier studies [Jami et al., 2014; Sundset et al., 2007] related to gut and rumen of cattle [Kalyuzhnaya et al., 2008; Kong et al., 2010] and buffalo [Singh et al., 2012]. Of these, relatively few analyzed bacterial diversity using high-throughput next-generation sequencing. Although Bacteroidetes was the most abundant in a majority samples, some of the samples in the solid phase of rumen revealed a higher Firmicutes proportion, similar to the results of a bacterial diversity study in solid and liquid phase of Canadian cervids [Gruninger et al., 2014]. Bacteroidetes accounted for a majority of the genera *Prevotella*, *Bacteroides* and *Parabacteroides*. Earlier studies have reported that a high percentage of the bacteria belonging to these phylum have not been further classified to species level, suggesting a great demand of studies related to characterization of such bacterial species within the complex rumen environment [Fernando et al., 2010]. *Fibrobacter succinogens* has been reported to be in high abundance in the rumen ecosystem due to its important role as major cellulolytic species [Nathani et al., 2013]. They are known to improve the dietary fiber metabolizing efficiency of the ruminants [Shi and Weimer, 1996] and thus they have frequently been observed to be in high proportions the rumen of buffalo [Wanapat and Cherdthong, 2009]. Therefore, the metagenomic shotgun sequencing

gave a clear idea of the abundance of various bacterial species within the rumen and any variation in the bacterial composition based on diet changes and in the solid and liquid fractions of the rumen fluid in cattle and buffalo. Detailed characterization of the Mehsani and Kankrej rumen has been reported at the taxonomic and functional level in earlier studies [Parmar et al., 2014; Patel et al., 2014]. However, there have been very few reports on rumen metagenomic studies with a focus on extracting individual genomes of microbes and their study at the species level. One such study includes a cow rumen metagenomic-based discovery of biomass-degrading enzymes, wherein the group assembled 15 uncultured microbial genomes of cellulolytic bacteria from cow rumen metagenomic sequences [Hess et al., 2011].

*Draft Genome Reconstruction*

A major challenge in metagenomic studies is the reconstruction of constituent taxa genomes from the obtained data in the form of reads [Carr et al., 2013]. A large number of methods have been employed to obtain deeper insight into the diverse taxa from shotgun metagenomic data. These methods include homology-based approaches like alignment to reference genomes, taxonomic classification, assembly and binning. For ecosystems that have a well-established number of reference genomes in databases like the human microbiome [Fodor et al., 2012], reference-guided assembly with reference genomes provides a way to determine various species within the microbiome [Martin et al., 2012; Qin et al., 2010]. Such data can further be used to study the unique traits in species related to host-specific adaptations and variations between different samples or different conditions of same sample. Such an approach can be applied to those community members whose whole genome type strains have been formerly sequenced. Taking into account the immense microbial diversity and the challenges related to culturing efforts, this approach can be applied on a large scale to a limited number of microbes. Thus, mining of individual genomes of bacterial species from community members of a complex environment is gaining a lot of attention from researchers worldwide.

Based on the results of MG-RAST, we successfully used a homology-based approach to reassemble the draft genome of *P. ruminicola* from Mehsani and Kankrej metagenomic data. The resultant contigs were in larger numbers as is usually the case in the assembly of short reads. The GC content was similar to the reference genome as well as that reported for *Prevotella* species [Avgustin et al., 1997]. Functional profiling of the genomes thus extracted was performed using RAST to gain a better understanding of their metabolic role in ruminants. Higher numbers of hits were observed for the carbohydrate and protein metabolism categories. The results reflect the well-established hypothesis that about 50–80% of the total absorbable protein for ruminants is due to microbial protein synthesis [Storm and Orskov, 1983]. Thus *P. ruminicola* plays an important role in the same. Cellulose is the major component of ruminant diet. Ruminant animals digest cellulose by completely relying on various cellulolytic and noncellulolytic microorganisms harbored in the rumen. Relatively few ruminal bacterial species (e.g. *Ruminococcus albus, Ruminococcus flavefaciens* and *Fibrobacter succinogens*) are found to exhibit direct cellulolytic activity; these bacteria provide the noncellulolytic organisms like *P. ruminicola* with oligosaccharides like the cellodextrins and cellobiose, resulting in cross-feeding interactions further enhancing fiber digestion [Johnson et al., 1982; Scheifinger and Wolin, 1973]. This confirms the high number of hits related to carbohydrate metabolism in PRM and PRK defining the high substrate utilization potentials. Presence of CDS utilizing various carbohydrate moieties further confirmed their role in the metabolism of dietary uptake by the host animal.

*CAZyme Prediction*

The highest percentage of contigs were found to be attributed to the GH family enzymes, confirming the role of *Prevotella* as utilizing several byproducts generated after the initial breakdown of complex polysaccharides. Earlier reports have shown that bacterial enzyme activities are species dependent, and the glycosidases associated with plant cell wall breakdown are mostly found high in the cellulolytic and hemicellulolytic species [Williams et al., 1984]. The presence of various different families of GH suggests their versatility in utilizing a wide variety of carbohydrate moieties as substrates. Specifically, the higher amount GH92 and GH2 family-related hits in PRM as well as PRK suggest the specific role of *P. ruminicola* species in galactoside, mannose, etc., utilization due to higher abundance of β-galactosidases, β-glucuronidases, β-mannosidases and exo-β-glucosaminidases. The higher amount of hits also suggests the presence of novel sequences coding highly expressed enzymes utilizing various substrates. Such enzymes can be further studied using cloning and expression-like experimental approaches.

The results in our study reveal the efficacy of metagenomics for individual species analysis in complex environments like rumen where the microflora plays a vital role in the host metabolism and digestion. Variation in

the abundance of a few GH family enzyme CDS also reflects the metabolic significance of rumen-adapted *P. ruminicola* in utilization potential of coarse diet by these animals based on the acquisition of novel genetic elements. These findings will further help develop strategies to facilitate in vitro growth of uncultivable rumen microbes by exploring enzymes specific to *P. ruminicola*.

## Experimental Procedures

*Metagenomic Shotgun Sequencing*

Rumen samples were collected from Mehsani buffalo (n = 8) and Kankrej cattle (n = 89) reared at Sardar Krushinagar Agricultural University and Anand Agricultural University for metagenomic analysis of ruminal microbes. The animals were kept on specific dietary regimes, and samples were collected in the last week of treatment using a flexible stomach tube from the rumen following standard procedures [Parmar et al., 2014]. The detail of sampling and entire design of the experiment is depicted in online supplementary figure 1. Total metagenomic DNA was extracted using the commercially available QIAmp DNA Stool Mini Kit (Qiagen, USA), and the concentrations were measured using NanoDrop Spectrophotometer ND 1000 (Thermo Scientific, USA). The DNA samples were subjected to sequencing based on Ion semiconductor technology using Ion Torrent PGM as per the manufacturer's instructions. The steps in brief included enzymatic fragmentation of DNA to obtain fragments in the range of 280–300 bp in size. The desired size fragments were ligated with the library adaptors and subjected to emulsion PCR, followed by recovery and loading onto an Ion PGM 316 Chip. Metagenomic sequence data reads of Kankrej and Mehsani were retrieved by downloading from the Ion Torrent server and after quality filtering by PRINSEQ; the sequences were uploaded as 48 individual files per breed on the MG-RAST Server [Meyer et al., 2008]. The analyzed data was studied for taxonomic diversity using the M5NR database with a maximum e-value cutoff of 1e-5 and minimum percent identity of 80. The published genomes of the organisms showing higher abundance in the rumen metagenome based on maximum best hits classification were downloaded from the NCBI ftp site.

*Draft Genome Reconstruction Using GS Reference Mapper*

Based on the analysis using MG-RAST, we reconstructed the draft genome of *P. ruminicola* using GS Reference Mapper V 2.3 by performing reference-guided assembly of the reads from the rumen metagenomic sequence run data of Kankrej cattle and Mehsani buffalo against the published genome *P. ruminicola* 23 used as the reference. All of the parameters were kept at default except the percent identity cutoff that was set to 97%. The extracted genomes PRM and PRK were also uploaded on the RAST server to study the functional annotation of the CDS into various subsystems as per the SEED classification [Aziz et al., 2008].

*CAZyme Prediction*

The contigs of each extracted genomes of PRK and PRM as well as the reference genome sequence were translated into 6 corresponding frame peptide sequences using EMBOSS Transeq [McWilliam et al., 2013; Rice et al., 2000]. All of the translated sequences were then submitted to the CAZyme Annotation Toolkit [Park et al., 2010] (http://mothra.ornl.gov/cgi-bin/cat/cat.cgi) for Pfam-based annotation with an e-value of 0.00001. Pfam-based annotation was preferred due to its higher sensitivity as compared to sequence-based annotation. The resulting hits obtained for each of the 3 genomes were compared for CAZyme family gene distribution and further classification of GH family was studied in detail.


## Acknowledgements

The authors are thankful to Department of Science and Technology (DST), Government of India for providing grant to N.M.N. in the form of the INSPIRE fellowship. The work was supported by the Niche Area of Excellence Program on Metagenomic Analysis of Ruminal Microbes funded by Indian Council of Agriculture Research (ICAR), New Delhi, India.


## Disclosure Statement

The authors declare no conflict of interests.

## References


Avgustin G, Wallace RJ, Flint H: Phenotypic diversity among ruminal isolates of *Prevotella ruminicola*: proposal of *Prevotella brevis* sp. nov., *Prevotella bryantii* sp. nov., and *Prevotella albensis* sp. nov. and redefinition of *Prevotella ruminicola*. Int J Syst Bacteriol 1997;47:484–488.

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O: The RAST Server: rapid annotations using subsystems technology. BMC Genomics 2008;9:75.

Bekele AZ, Koike S, Kobayashi Y: Genetic diversity and diet specificity of ruminal Prevotella revealed by 16S rRNA gene-based analysis. FEMS Microbiol Lett 2010;305:49–57.

Carr R, Shen-Orr SS, Borenstein E: Reconstructing the genomic content of microbiome taxa through shotgun metagenomic deconvolution. PLoS Comput Biol 2013;9:e1003292.

Edwards J, McEwan N, Travis A, John Wallace R: 16S rDNA library-based analysis of ruminal bacterial diversity. Antonie van Leeuwenhoek 2004;86:263–281.

Fernando SC, Purvis HT, 2nd, Najar FZ, Sukharnikov LO, Krehbiel CR, Nagaraja TG, Roe BA, Desilva U: Rumen microbial population dynamics during adaptation to a high-grain diet. Appl Environ Microbiol 2010;76:7482–7490.

Fodor AA, DeSantis TZ, Wylie KM, Badger JH, Ye Y, Hepburn T, Hu P, Sodergren E, Liolios K, Huot-Creasy H, Birren BW, Earl AM: The 'most wanted' taxa from the human microbiome for whole genome sequencing. PloS One 2012;7:e41294.

Gruninger RJ, Sensen CW, McAllister TA, Forster RJ: Diversity of rumen bacteria in Canadian cervids. PloS One 2014;9:e89682.

Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, Luo S, Clark DS, Chen F, Zhang T, Mackie RI, Pennacchio LA, Tringe SG, Visel A, Woyke T, Wang Z, Rubin EM: Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science 2011;331: 463–467.

Jami E, White BA, Mizrahi I: Potential role of the bovine rumen microbiome in modulating milk composition and feed efficiency. PloS One 2014;9:e85423.

Johnson EA, Reeses ET, Demain AL: Inhibition of *Clostridium thermocellum* cellulase by end products of cellulolysis. J Appl Biochem 1982; 4:8.

Kalyuzhnaya MG, Lapidus A, Ivanova N, Copeland AC, McHardy AC, Szeto E, Salamov A, Grigoriev IV, Suciu D, Levine SR, Markowitz VM, Rigoutsos I, Tringe SG, Bruce DC, Richardson PM, Lidstrom ME, Chistoserdova L: High-resolution metagenomics targets specific functional types in complex microbial communities. Nat Biotechnol 2008;26:1029–1034.

Klieve AV, Bauchop T: Morphological diversity of ruminal bacteriophages from sheep and cattle. Appl Environ Microbiol 1988;54:1637–1641.

Koike S, Kobayashi Y: Fibrolytic rumen bacteria: their ecology and function. Asian Aust J Anim Sci 2009;22:131–138.

Koike S, Yoshitani S, Kobayashi Y, Tanaka K: Phylogenetic analysis of fiber-associated rumen bacterial community and PCR detection of uncultured bacteria. FEMS Microbiol Lett 2003; 229:23–30.

Kong Y, Teather R, Forster R: Composition, spatial distribution, and diversity of the bacterial communities in the rumen of cows fed different forages. FEMS Microbiol Ecol 2010;74:612–622.

Konstantinidis KT, Braff J, Karl DM, DeLong EF: Comparative metagenomic analysis of a microbial community residing at a depth of 4,000 meters at station ALOHA in the North Pacific subtropical gyre. Appl Environ Microbiol 2009;75: 5345–5355.

Lu G, Jiang L, Helikar RM, Rowley TW, Zhang L, Chen X, Moriyama EN: GenomeBlast: A web tool for small genome comparison. BMC Bioinformatics 2006;7(suppl 4):S18.

Mannarelli BM, Ericsson LD, Stack RJ: Taxonomic relationships among strains of the anaerobic bacterium *Bacteroides ruminicola* determined by DNA and extracellular polysaccharide analysis. Appl Environ Microbiol 1991;57:2975–2980.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM: Genome sequencing in microfabricated high-density picolitre reactors. Nature 2005;437:376–380.

Martin J, Sykes S, Young S, Kota K, Sanka R, Sheth N, Orvis J, Sodergren E, Wang Z, Weinstock GM, Mitreva M: Optimizing read mapping to reference genomes to determine composition and species prevalence in microbial communities. PloS One 2012;7:e36427.

McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R: Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Res 2013;41:W597–W600.

Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA: The metagenomics RAST Server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 2008;9:386.

Morvan B, Bonnemoy F, Fonty G, Gouet P: Quantitative determination of H2-utilizing acetogenic and sulfate-reducing bacteria and methanogenic archaea from digestive tract of different mammals. Curr Microbiol 1996;32: 129–133.

Nathani NM, Patel AK, Dhamannapatil PS, Kothari RK, Singh KM, Joshi CG: Comparative evaluation of rumen metagenome community using qPCR and MG-RAST. AMB Express 2013;3:55.

Osborne JM, Dehority BA: Synergism in degradation and utilization of intact forage cellulose, hemicellulose, and pectin by three pure cultures of ruminal bacteria. Appl Environ Microbiol 1989;55:2247–2250.

Park BH, Karpinets TV, Syed MH, Leuze MR, Uberbacher EC: CAZymes Analysis Toolkit (CAT): web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. Glycobiology 2010;20:1574–1584.

Parmar NR, Solanki JV, Patel AB, Shah TM, Patel AK, Parnerkar S, Kumar JI, Joshi CG: Metagenome of Mehsani buffalo rumen microbiota: an assessment of variation in feed-dependent phylogenetic and functional classification. J Mol Microbiol Biotechnol 2014;24:249–261.

Patel V, Patel AK, Parmar NR, Patel AB, Reddy B, Joshi CG: Characterization of the rumen microbiome of Indian Kankrej cattle *(Bos indicus)* adapted to different forage diet. Appl Microbiol Biotechnol 2014;98:9749–9761.

Purushe J, Fouts DE, Morrison M, White BA, Mackie RI, Coutinho PM, Henrissat B, Nelson KE: Comparative genome analysis of *Prevotella ruminicola* and *Prevotella bryantii*: insights into their environmental niche. Microb Ecol 2010;60:721–729.

Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Jian M, Zhou Y, Li Y, Zhang X, Qin N, Yang H, Wang J, Brunak S, Dore J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J, Bork P, Ehrlich SD: A human gut microbial gene catalogue established by metagenomic sequencing. Nature 2010;464:59–65.

Quince C, Lanzen A, Curtis TP, Davenport RJ, Hall N, Head IM, Read LF, Sloan WT: Accurate determination of microbial diversity from 454 pyrosequencing data. Nat Methods 2009;6:639–641.

Ramsak A, Peterka M, Tajima K, Martin JC, Wood J, Johnston EA, Aminov RI, Flint JH, Avgustin G: Unraveling the genetic diversity of ruminal bacteria belonging to the CFB phylum. FEMS Microbiol Ecol 2000;33:69–79.

Rice P, Longden I, Bleasby A: EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet 2000;16:276–277.

Scheifinger CC, Wolin MJ: Propionate formation from cellulose and soluble sugars by combined cultures of *Bacteroides succinogenes* and *Selenomonas ruminantium*. Appl Microbiol 1973; 26:789–795.

Shi Y, Weimer PJ: Utilization of individual cellodextrins by three predominant ruminal cellulolytic bacteria. Appl Environ Microbiol 1996;62: 1084–1088.

Singh KM, Ahir VB, Tripathi AK, Ramani UV, Sajnani M, Koringa PG, Jakhesara S, Pandya PR, Rank DN, Murty DS, Kothari RK, Joshi CG: Metagenomic analysis of Surti buffalo *(Bubalus bubalis)* rumen: a preliminary study. Mol Biol Rep 2012;39:4841–4848.

Stevenson DM, Weimer PJ: 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 2007;75:165–174.

Storm E, Orskov ER: The nutritive value of rumen micro-organisms in ruminants. 1. Large-scale isolation and chemical composition of rumen micro-organisms. Br J Nutr 1983;50:463–470.

Sundset MA, Praesteng KE, Cann IK, Mathiesen SD, Mackie RI: Novel rumen bacterial diversity in two geographically separated sub-species of reindeer. Microb Ecol 2007;54:424–438.

Van Gylswyk NO: Enumeration and presumptive identification of some functional groups of bacteria in the rumen of dairy cows fed grass-silage based diets. FEMS Microbiol Ecol 1990;73: 243–254.

Wanapat M, Cherdthong A: Use of real-time PCR technique in studying rumen cellulolytic bacteria population as affected by level of roughage in swamp buffalo. Curr Microbiol 2009;58: 294–299.

Whitford MF, Forster RJ, Beard CE, Gong J, Teather RM: Phylogenetic analysis of rumen bacteria by comparative sequence analysis of cloned 16S rRNA genes. Anaerobe 1998;4:153–163.

Williams AG, Withers SE, Coleman GS: Glycoside hydrolases of rumen bacteria and protozoa. Curr Microbiol 1984;10:7.

Wood J, Scott KP, Avgustin G, Newbold CJ, Flint HJ: Estimation of the relative abundance of different Bacteroides and Prevotella ribotypes in gut samples by restriction enzyme profiling of PCR-amplified 16S rRNA gene sequences. Appl Environ Microbiol 1998;64: 3683–3689.