Open Access
Asian-Australas J Anim Sci
Vol. 31, No. 5:738-747 May 2018
https://doi.org/10.5713/ajas.17.0174
pISSN 1011-2367 eISSN 1976-5517

AJAS
Asian-Australasian Journal of Animal Sciences

Metagenomic analysis of bacterial community structure and diversity of lignocellulolytic bacteria in Vietnamese native goat rumen

Thi Huyen Do[1],*, Trong Khoa Dao[1,2], Khanh Hoang Viet Nguyen[2,3], Ngoc Giang Le[1,4], Thi Mai Phuong Nguyen[1], Tung Lam Le[1], Thu Nguyet Phung[1], Nico M. van Straalen[4], Dick Roelofs[4], and Nam Hai Truong[1,*]

* **Corresponding Authors:**
Thi Huyen Do
Tel: +84-24-37917021, **Fax:** +84-24- 38363144,
E-mail: dohuyen@ibt.ac.vn
Nam Hai Truong
Tel: +84-24-37917980, **Fax:** +84-24- 38363144,
E-mail: tnhai@ibt.ac.vn

[1] Institute of Biotechnology, Vietnam Academy of Science and Technology, Cau Giay, Ha Noi 100000, Vietnam
[2] Graduate University of Science and Technology, Vietnam Academy of Science and Technology, Cau Giay, Ha Noi 100000, Vietnam
[3] Institute of New Technology, Academy of Military Science and Technology, Cau Giay, Hanoi 100000, Vietnam
[4] Department of Ecological Science, Vrije Universiteit Amsterdam, De Boelelaan 1085, Amsterdam 1081 HV, The Netherlands

ORCID
Thi Huyen Do
https://orcid.org/0000-0003-0412-8872
Trong Khoa Dao
https://orcid.org/0000-0002-4501-3353
Khanh Hoang Viet Nguyen
https://orcid.org/0000-0001-5226-4831
Ngoc Giang Le
https://orcid.org/0000-0001-6660-557X
Thi Mai Phuong Nguyen
https://orcid.org/0000-0003-2686-5638
Tung Lam Le
https://orcid.org/0000-0002-3322-3134
Thu Nguyet Phung
https://orcid.org/0000-0002-7617-3237
Nico M. van Straalen
https://orcid.org/0000-0003-4219-7297
Dick Roelofs
https://orcid.org/0000-0003-3954-3590
Nam Hai Truong
https://orcid.org/0000-0001-9513-3479

Submitted Mar 3, 2017; Revised Jun 15, 2017; Accepted Sept 4, 2017

Objective: In a previous study, analysis of Illumina sequenced metagenomic DNA data of bacteria in Vietnamese goats' rumen showed a high diversity of putative lignocellulolytic genes. In this study, taxonomy speculation of microbial community and lignocellulolytic bacteria population in the rumen was conducted to elucidate a role of bacterial structure for effective degradation of plant materials.
Methods: The metagenomic data had been subjected into Basic Local Alignment Search Tool (BLASTX) algorithm and the National Center for Biotechnology Information non-redundant sequence database. Here the BLASTX hits were further processed by the Metagenome Analyzer program to statistically analyze the abundance of taxa.
Results: Microbial community in the rumen is defined by dominance of Bacteroidetes compared to Firmicutes. The ratio of Firmicutes versus Bacteroidetes was 0.36:1. An abundance of Synergistetes was uniquely identified in the goat microbiome may be formed by host genotype. With regard to bacterial lignocellulose degraders, the ratio of lignocellulolytic genes affiliated with Firmicutes compared to the genes linked to Bacteroidetes was 0.11:1, in which the genes encoding putative hemicellulases, carbohydrate esterases, polysaccharide lyases originated from Bacteroidetes were 14 to 20 times higher than from Firmicutes. Firmicutes seem to possess more cellulose hydrolysis capacity showing a Firmicutes/Bacteroidetes ratio of 0.35:1. Analysis of lignocellulolytic potential degraders shows that four species belonged to Bacteroidetes phylum, while two species belonged to Firmicutes phylum harbouring at least 12 different catalytic domains for all lignocellulose pretreatment, cellulose, as well as hemicellulose saccharification.
Conclusion: Based on these findings, we speculate that increasing the members of Bacteroidetes to keep a low ratio of Firmicutes versus Bacteroidetes in goat rumen has resulted most likely in an increased lignocellulose digestion.

Keywords: Bacterial Structure; Illumina *De novo* Sequencing; Lignocellulolytic Bacteria; Vietnamese Goat

INTRODUCTION

Cellulose, hemicellulose and lignin are the major components of lignocellulosic biomass, they interact together to form a rigid structure against enzymatic decomposition. However, this kind of biomass can be effectively digested by rumen microorganisms in natural ecosystems. The rumen harbours a complex anaerobic microbial ecosystem composed of bacteria, archaea, fungi, and protozoa, they are well-known to secret enzymes essential for conversion of lignocellulose into simple sugars. The released sugars will be taken up by the host and used as energy source, building block of volatile fatty acids, or used for protein synthesis thereby providing the host with various nutrients [1,2]. Thus these microbes influence the ruminant's health and productivity of meat and milk. In addition, industrial hydrolysis of

cellulosic biomass into mono-sugars is a rate-limiting step in producing of biofuel and many other value-added products [3]. Therefore, the rumen microbiome has attracted substantial interest in the pursuit of improved feed utilization efficiency of the livestock industry and cost effective production of biomass-derived products such as bio-fuel [4-8].

An earlier study showed that effective lignocellulose degrading consortia comprise mainly anaerobic cellulolytic bacterial taxa in stable coexistence with various non-cellulolytic members. The non-cellulolytic bacteria play an important role in metabolite dependence and environment modification, leading to enhancement in biomass hydrolysis [9], while, cellulolytic bacteria are recognized by the presence of endoglucanase and beta-glucosidase [10] or/and hemicellulases, carbohydrate esterases (CEs), polysaccharide lyases (PLs) to help lignocellulose digestion. Thus, for elucidation of an effective degradation environment such as rumen, an overview of bacterial community structure (including non-cellulolytic and cellulolytic bacteria) should be analyzed simultaneously with analysis of bacterial community harbouring lignocellulolytic genes.

Studies on phylogenetic diversity and community structure based on 16S rRNA genes demonstrated that rumen microbiome can be shaped by both diet and host genotype. The effects of host genotype on the rumen microbiome were described in the rumen microbiome of African ruminant [11], reindeer [12], yak [13], sheep, deer, and cattle [14]. Other effects such as changing in diets an age of host also impacts the microbial population [15,16]. Assembly of bacterial consortia formed by host genotype and adapted with diet in Salbard raindeer rumen, pasture-fed sheep rumen, hay-fed cow rumen is different when compared to typical industrial biogas reactor fed with maize silage, cow manure, chicken manure or herbivore faeces. Güllert et al [17] suggested a ratio of Firmicutes/Bacteroidetes in bacterial communities that may be optimal for biomass hydrolysis. This ratio is typically high in biogas fermenters and associates with lack of cellulolytic glycoside hydolase (GH) enzymes. In contrast, this ratio is low in rumen communities and associates with high diversity of cellulolytic GH enzymes [17]. Such data indicate that hydrolysis of biomass will be more optimal if the ratio is kept low, thereby increasing the capacity of biomass degradation by a higher diversity of cellulolytic GH enzymes.

The rumen from Vietnamese native goats is a typical environment for effective lignocellulose decomposition. The goat Co and Bach Thao are domestic, native popular breeds within Vietnam and important for high quality meat and milk production. Bach Thao is hybrid generated by an ancient cross between Beetal and Jamnapari. The diet of both domestic goats grazing in the high rocky mountains consists of hay and a variety of grasses and leaves from mountain trees as well as crop residues at night. Analysis of ~9 Gb metagenome of bacteria in the goats' rumen (five Co and five Bach Thao animals) sequenced by Illumina technology resulted in 172,981 contigs and predicted 164,644 open reading frames (ORFs), of which 122,304 ORFs were annotated. A total of 39,579 ORFs could be linked to National Center for Biotechnology Information (NCBI) taxonomy, 99.8% belonged to bacteria. According to functional annotation results speculated by Swiss-Prot, Kyoto Encyclopedia of Genes and Genomes, Non-supervised Orthologous Groups, Cluster of Orthologous Groups and Carbohydrate-Active enZYmes databases, the absolute amount and diversity of hemicellulases (2,252 genes), lignocellulose pretreatment enzymes (821 ORFs) and carbohydrate binding modules (CBMs) (763 genes) were much higher in our data [18] when compared to a study in Korean goat rumen [4]. We hypothesize that: i) the Vietnamese diet is the main driver of this diversification apart from the host genotype; ii) ratio between Firmicutes and Bacteroidetes is an important indicator to assess the effectuality of lignocellulose degradation: a lower ratio indicates the presence of more diverse lignocellulolytic GH enzymes originating from Bacteroidetes.

MATERIALS AND METHODS

Animal care

The animal experimental protocol of this research was reviewed, discussed and approved by the institutional ethics committee of the Institute of Genome Research Institutional Review Board (IGR IRB), Vietnam (Approval number: No. 03/QD-NCHG).

Materials

In previous study, The goats Co (five animals) and Bach Thao (five animals) living on natural hay in high rocky mountains at private goat farms at Ninh Binh, Thanh Hoa in Vietnam were selected and used for harvesting bacteria in the goats' rumen [18].

The 39,579 ORFs from the metagenomic data that could be linked to NCBI taxonomy [18] were used for taxonomic assessment of the bacterial community structure in Vietnamese goats' rumen. The ORFs related to lignocellulose degradation including 821 ORFs for CEs and PLs, 816 ORFs for cellulases, and 2,252 ORFs for hemicellulases [18] were used for assessment of cellulolytic bacteria as described in the following section.

Taxonomic assignment

In previous study [18], a classified group of 39,579 ORFs was obtained from comparison of 164,644 ORFs from the metagenomic data of bacteria in goats' rumen against the NCBI non-redundant protein (NR) database using the BLASTX algorithm with an E-value of less than 10^{-5}. In this study, the ORFs were subjected to the Metagenome Analyzer program (MEGAN) (version 4.6) [19] for taxonomic assignment into

NCBI taxonomy using lowest common ancestor (LCA) algorithm. The LCA-based algorithm assigns genes to taxa in the way that the taxonomical level of the assigned taxon reflects the level of conservation of the gene. So sequences placed at the deeper nodes of the phylogeny represent more conserved genes relative to the genes on the leaves the tree that are more specific to that species. Thus all ORFs (represented in gene code) in the classified group were assigned in the detail in every taxon.

Analysis of overall bacterial community structure in Vietnamese goats' rumen

From 39,579 ORFs classified in MEGAN, we separated the taxon level: kingdom, phylum, class, order, family, genus, species in separated files. After that, we counted for abundance of each taxonomy level and then summarized for the most abundances in every levels. We also used the ratio of Firmicutes/Bacteroidetes to assess the bacterial structure in our goats' rumen in comparison with the ratio in other bacterial consortia in different animal rumens in previous studies.

Analysis of bacteria harbouring lignocellulolytic genes in Vietnamese goats' rumen

From 821 ORFs encoding CEs and PLs involved in lignocellulose pretreatment, 816 ORFs encoding cellulases, 2,252 ORFs encoding for hemicellulases had been annotated from 164,644 ORFs [18], we filtered all these ORFs accompanied with the corresponding gene codes and functional annotation to a new file. Based on the gene codes, we used formula Vlookup in exel file to assign the genes with the corresponding taxonomic levels from taxonomic profile of total 39,579 ORFs to generate a file with gene code, functional and taxonomical information combined. The every taxonomic level related to the lignocellulolytic genes were separated in to each file for counting abundance bacteria associated with GHs for cellulases, hemicellulases and CEs, PLs for lignocellulose degradation. We also used the ratio of Firmicutes/Bacteroidetes for assessment of bacterial role in lignocellulose degradation.

This project was deposited in the DNA Databank of Japan with the accession ID PSUB006562.

RESULTS

Bacterial community structure and the dominance of the Bacteroidetes

In this study, from ~9 Gb of rumen bacterial metagenomics data, 39,579 ORFs (accounted for 24.03%) were classified. The majority of the organisms belonged to bacteria (99.8%). The 39,501 bacterial ORFs were affiliated to 28 phyla, 41 classes, 95 orders, 181 families, 571 genera and 1,634 species (Table 1, Supplementary Table S1). The most dominant phylum was Bacteroidetes (63.5%) followed by Firmicutes (22.6%), Proteobacteria (7.5%), and Synergistetes (3.1%) (Table 1). Typical ratio of Firmicutes versus Bacteroidetes in the the bacterial community was 0.36:1 (Figure 1A). Further analysis showed that, within Bacteroidetes phylum, the most abundant order was Bacteroidales (60.8%). The most represented order within the Firmicutes phylum was Clostridiales (16.4%) followed by Selenomonadales (2.8%) (Figure 1). Synergistales—a unique member of Synergistetes phylum represented about 3.1% of the taxonomic classified ORFs and Aeromonadales—a member of Proteobacteria, were also abundant in this data. With regard to the taxonomy at genera level, *Prevotella* (35.3%) and *Bacteroides* (16%) were most abundant which belong to Bacteroidales. Analysis of the 13 most abundant species resulted in eight species belonging to *Prevotella*, one species belonging to *Bacteroides*, two species assigned to *Clostridiales* and only one species of *Synergistales* could be identified (Figure 1B).

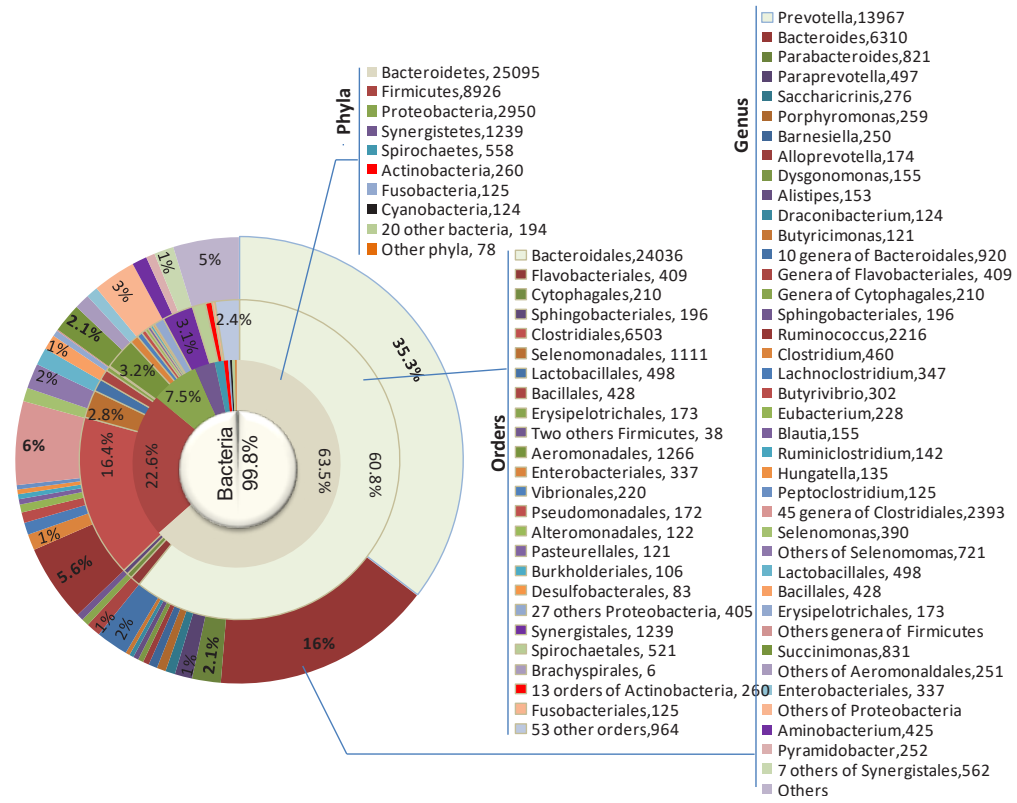
Bacterial community harbouring lignocellulolytic genes

In the previous publication, we indicated that lignocellulolytic genes were highly diverse in Vietnamese goats' rumen [18] if compared to another study on goat rumen microbial diversity [4]. We supposed that the diversity is an adaptation of the rumen bacteria to goat's diet. As such, we affiliated all the ORFs related to lignocellulose degradation in the goats' rumen. As a result, among 816 ORFs genes for cellulases, 2,252 ORFs for hemicellulases and 821 ORFs encoding enzymes involved in lignocelluloses pretreatment, that were annotated by different databases [18], 221 genes for cellulases, 544 genes for hemicellulases and 226 genes for lignocellulose pretreatment were affiliated by NCBI taxonomic classification. That means about 24% to 27% lignocellulolytic genes were successfully assigned to taxon and a large portion of the genes could not be affiliated to taxonomy. In general, these gene groups were mostly found

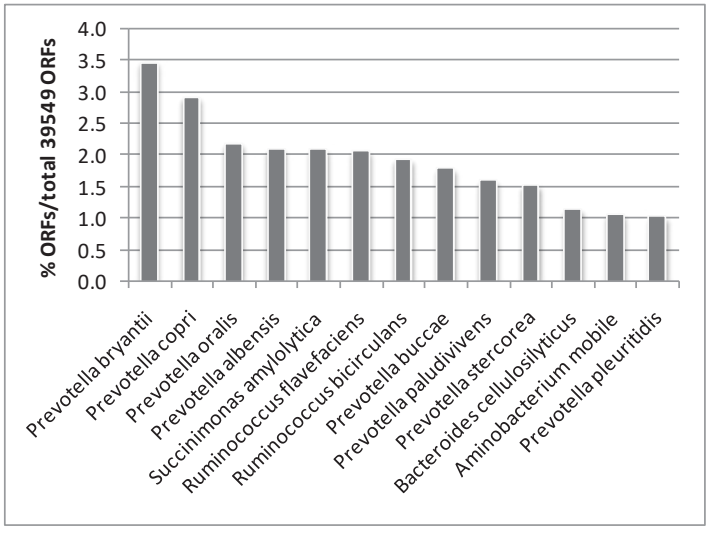
Table 1. Inventory of affiliation of 39,579 ORFs in NCBI taxomomic classification by MEGAN

	Genes	Percentage (%)	Phylum	Class	Order	Family	Genus	Species
Bacteria	39,501	99.80	28	41	95	181	571	1,634
Archaea	67	0.17	4	6	13	13	23	30
Eukaryota	9	0.02	3	2	4	4	4	5
Viruses	2	0.01	0	0	1	1	0	2
Sum	39,579	100.00	35	49	113	199	598	1,671

ORFs, open reading frames; NCBI, National Center for Biotechnology Information; MEGAN, Metagenome Analyzer program.



A



B

Figure 1. Analysis of the Vietnamese native goat rumen bacterial community structure at kingdom, phylum, order and genus levels (A) and the thirteen most abundant species (B). The numbers indicate the number of genes.

to be associated with the Bacteroidetes phylum (854 ORFs of the total 991 ORFs; accounting for 86.2%). The second most abundant phylum was Firmicutes (94 ORFs; accounting for 9.5% of ligocellulolytic genes) (Figure 2A). The ratio of Firmicutes/Bacteroidetes is therefore quite low (0.11:1) and lower than the ratio in bacterial structure (0.36:1). Looking at order level, the most lignocellulolytic associated ORFs were affiliated to Bacteroidales, and Clostridiales (Figure 2, Supplementary

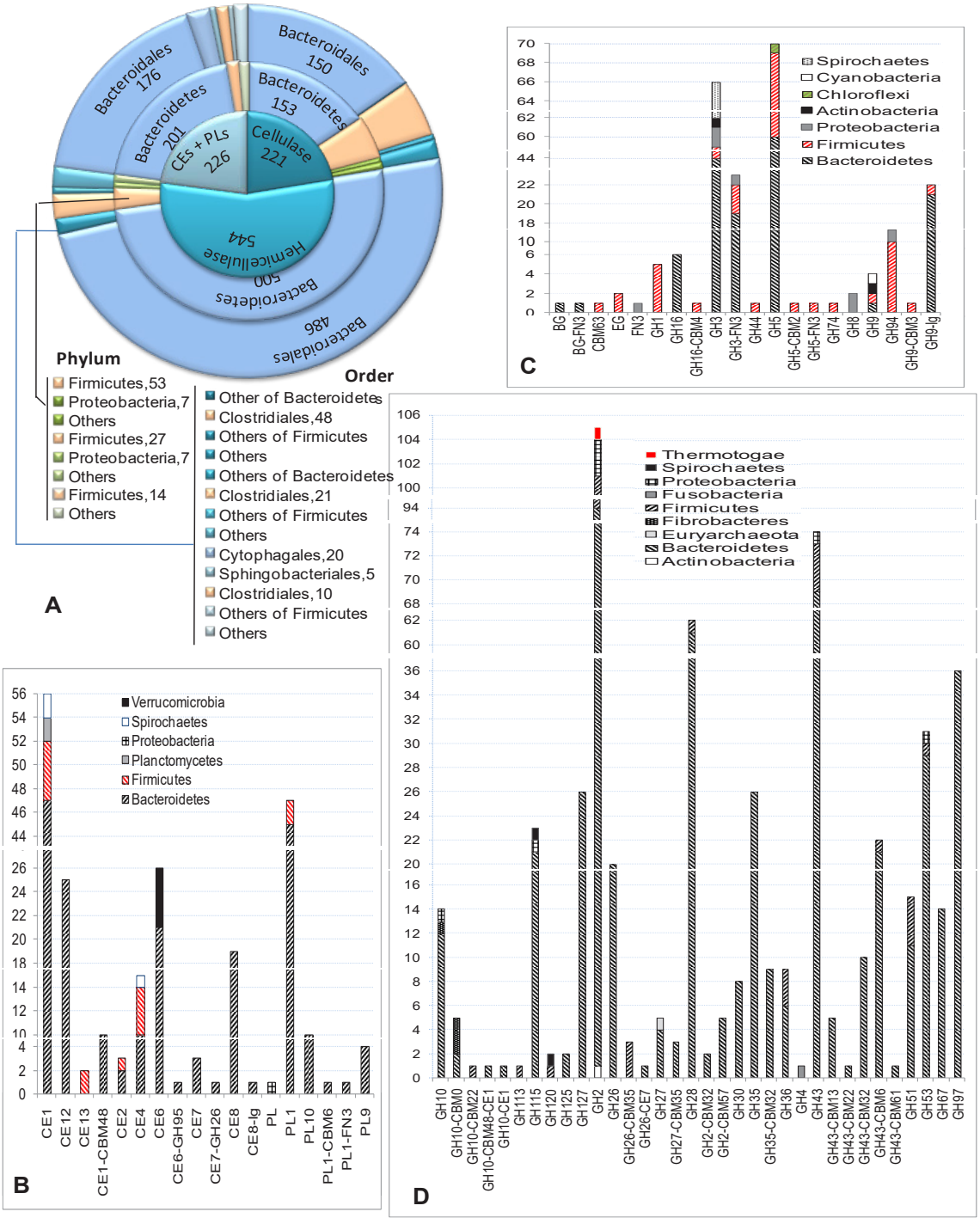


Figure 2. Analysis of community structure of rumen goat bacteria harbouring genes for lignocellulose degradation at phylum, and order level (A), bacterial phyla bearing genes encoding enzymes (CEs, PLs) for lignocellulose pretreatment (B), cellulases (C) and hemicellulases (D). The numbers indicate the number of genes. GH, glycoside hydrolase family; CE, carbohydrate esterase; PL, polysaccharide lyases; CBM, carbohydrate binding model; FN3, fibronectin 3; Ig, immunoglobulin-like domain; BG, beta-glucosidase; EG, endoglucanase.

Table S2). Analysis in genus level, 398 ORFs were affiliated to *Provotella*, 300 ORFs were belonged to *Bacteroides*, 39 ORFs were predicted to be derived from *Paraprevotella*, 39 ORFs were originated from *Ruminococcus*. Specially *Aerophaga* was predicted to harbour ORFs for only hemicellulase (7 ORFs), while *Algoriphagus* was predicted to carry ORFs for only CEs and PLs.

When we analysed specific classes of lignocellulolytic genes, we observed divergent patterns of taxonomic distribution, depending on enzyme function. For pretreatment enzymes (CEs

and PLs), the most abundant phylum is again Bacteroidetes (201 ORFs), followed by Firmicutes (14 ORFs) (Figure 2A). In other words, Bacteroidetes hold 88.9% genes for lignocellulose pretreatment, Firmicutes kept 6.2%. Thus the ratio between Firmicutes versus Bacteroidetes is very low (0.07:1). However, genes encoding for esterase enzymes showed a different taxonomic pattern. The 56 ORFs encoding CE1 were affiliated to many bacteria orders such as Bacteroidetes, Firmicutes, Planctomycetes, Spirochaetes, while CE1 domains collocated with a carbohydrate binding domain (CBM) CBM48 (10 ORFs) were only assigned to Bacteroidetes. This pattern was similar for CE6 and PL1, CE6 (21 ORFs), which were predicted to be produced in Bacteroidetes and Verrucomicrobia, while PL1 (45 ORFs) could be linked to Bacteroidetes and Firmicutes. However, the bi-fuctional enzymes CE6-GH95 and PL1 carrying CBM6 domain or a fibronectin like 3 domain (FN3), which help for increase the affinity between enzyme and corresponding substrate, were again assigned to Bacteroidetes only. Moreover, several enzymes such as CE12 (25 ORFs), CE7, CE7-CH26, CE8, CE8-Ig (immunoglobulin like domain), PL9, PL10 were predicted to be produced only in Bacteroidetes, while CE13 was only generated in Firmicutes (Figure 2B). At species level, *Prevotella buccae* was the most abundant species among lignocellulose pretreatment enzymes.

The most cellulase ORFs were affiliated to Bacteroidetes (153 ORFs; 69.2%), followed by Firmicutes (53 ORFs, ~24.0%), and then Proteobacteria (7 ORFs) (Figure 2A). Thus ratio of Firmicutes/Bacteroidetes is again low (0.35:1) but higher than the ratio in CEs and PLs producers (0.07:1). We found that Bacteroidetes and Firmicutes were predicted to produce many enzymes such as GH3, GH3-FN3, GH5, GH9, GH9-Ig. More specifically, GH3 is predicted to have beta-glucosidase activity, while GH5 and GH9 exhibit endoglucanase activity. Within cellulases, we also identified genes uniquely assigned to Firmicutes. For instance, GHs (GH6, GH5, GH9) accompanied with CBMs, such as CBM2, CBM3, CBM4, CBM63 were only identified among Firmicutes, while no GHs collocated with CBMs in this groups could be assigned to Bacteroidetes. Furthermore, GH1, GH44, GH5-FN3, GH74 were also uniquely associated with Firmicutes (Figure 2C). Further analysis of species harbouring cellulolytic genes showed that, *Bacteroides uniformis*, *Ruminococcus bicirculans*, *Eubacterium siraeum* were the most dominant to bear putative cellulases (Figure 3, Supplementary Table S2).

The hemicellulase group is the most diverse compared to cellulases and lignocellulose pretreatment genes, with 20 different GHs activity domains (Figure 2D). Within this group, the most abundant phylum was again Bacteroidetes with 500 ORFs accounting for 91.9%, followed by Firmicutes (27 ORFs), and then by Proteobacteia (7 ORFs) (Figure 2A). Therefore,

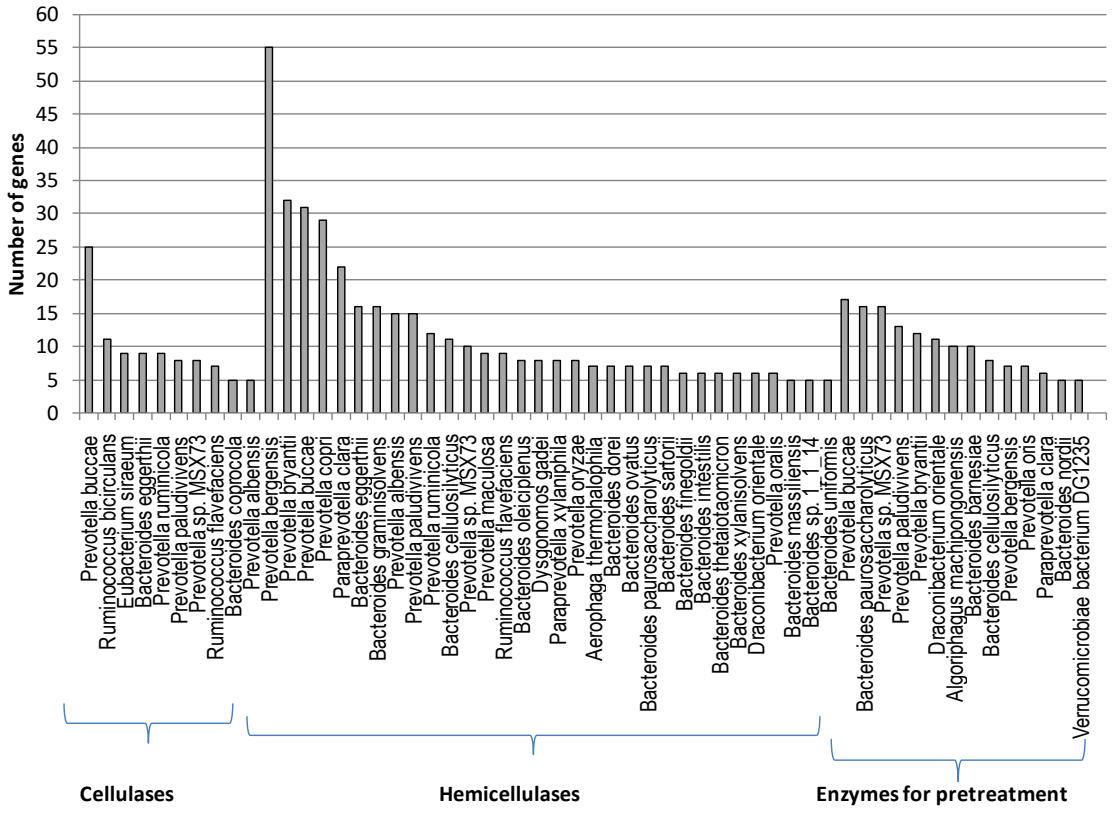


Figure 3. Species abundance with regard to genes involved in cellulases, hemicellulases, and enzymes for lignocelluloses pretreatment.

the ratio of Firmicutes versus Bacteroidetes was again low being 0.05:1. Here, GH2 was estimated to be produced from the most diverse phyla such as Actinobacteria, Bacteroidetes, Firmicutes, Proteobacteia, Thermotogae, followed by GH10, GH15, GH53. Firmicutes and Bacteroidetes were also associated with genes encoding six GH enzymes (GH2, GH28, GH36, GH6, GH51, GH53). However, many other GHs including bi-functional enzymes containing also a CE domain or a CBM domain, and were predicted to be produced solely associated with Bacteroidetes, while GH113, GH26-CBM35 were only linked only to Firmicutes (Figure 2D). At the species level, *Prevotella bergensis*, *P. bryantii*, *P. copri*, *Paraprevotella clara*, *B. graminisolvens*, *P. albensis*, *P. paludivivens*, *B. cellulosilyticus*. *P. buccae*, *B. eggerthii*, *P. ruminicola* were the most abundances for producing hemicellulases.

The most potential species for lignocellulose degradation

Potential degraders of the cellulosic biomass usually have both beta-glucosidase and endoglucanase [10], while the potential opportunists possess only beta-glucosidase to hydrolyze cellobiose or short polysaccharides to glucose for their own consumption. In accordance with this, the most potential opportunist bacteria in Vietnamese native goats' rumen were proposed to be *Bacteroides fluxus*, *Lachnospiraceae bacterium* JC7, *Bifidobacterium breve*, *Bacteroides caccae*, *Prevotella nceiensis*, *Hungatella hathewayi*, *Pseudomos syringae* and *Treponema* sp. JC4 (Supplementary Table S3) that produce beta-glucosidase only.

Many species were classified as potential lignocellulose degraders in this study, of which 56 species harboured at least 2 different catalytic domains. For cellulose degradation, *R. bicirculans* and *R. flavefaciens* were associated with six different catalytic enzymes. *R. bicirculans* harbored GH3-FN3, GH44, GH5, GH5-CBM2, GH5-FN3, GH74, while *R. flavefaciens* was predicted to contain CBM63, GH16-CBM4, GH5, GH9, GH94, GH9-CBM3 (Supplementary Table S3).

For effective hemicellulose decomposition, *B. graminisolvens* possesses nine different catalytic domains, *P. copri* and *P. paludivivens* had eight different catalytic domains (Supplementary Table S3). In the bacterial group for lignocellulose pretreatment, *Prevotella* sp. MSX73 was predicted to the most potential degrader (Supplementary Table S3) to harbour six different catalytic domains. It is interesting that so many species harboured genes for so many functions of lignocelluose degradation. For instance, *Prevottella albensis* was predicted to bear both cellulases and hemicellulases when this bacteria was predicted to possess both cellulase domains GH3-FN3, GH9-Ig and hemicellulase domains GH10, GH2, GH30, GH43-CBM13, GH43-CBM6, GH97. Seven species contained both GHs for hemicellulases and CEs and PLs for lignocelluose pretreatment. More specifically, *B. cellulosilyticus* and *P. bryantii* had 10 genes for different catalytic domains (Supplementary Table S3).

Remarkably, seven species (*Draconibacterium orientale*, *P. buccae*, *P. copri*, *P. oryzae*, *P. paludivivens*, *R. bicirculans*, *R. flavefaciens*) contained all domains conducting all steps for lignocellulose degradation: pretreatment, cellulose and hemicellulose saccharification. Most domains (14 domains) were identified in one species *P. paludivivens*, while four other species harboured 13 catalytic domains (Supplementary Table S3).

DISCUSSION

The Vietnamese rumen microbiome is mostly adapted to diet

To get an insight of the bacterial community and their role in lignocellulose digestion in Vietnamese native goats' rumen, the enriched bacterial metagenomic DNA was sequenced by Illumina platform. The result from NCBI taxonomic classification showed significant amount of unclassified genes that accounted for 50.25%. This suggests that the major part of bacteria in the goats' rumen was novel or the threshold set for classification was too conservative.

Among 39,579 ORFs affiliated to NCBI taxonomic classification, we found Bacteroidetes and Firmicutes were the most abundant phyla, accounted for 63.5% and 22.6% respectively (Figure 1). In general, the phylogenetic structure of our sample rumen appeared to be similar to the structures described in already published studies of animal rumens. Also, the ratio between Firmicutes and Bacteroidetes was comparable. In our study, ratio Firmicutes/Bacteroidetes was 0.36:1, that is the same ratio as in the rumen of Svalbard reindeer [7], sheep [20], Jinnan cattle [13] and relevant to the ratio 0.3:1 found in goat rumen at 110 days old after adaptation to plant materials [15] (Table 2). However this ratio is strongly dependent on diet throughout the life cycle of goats, and the low ratio between Firmicutes and Bacteroidetes only establishes later in life when the diet consists of mainly plant materials. This was elegantly shown in a previous study [15], they analyzed changes of bacterial community in goats' rumen after weaning at 60-day old and getting adaptation with plant materials and showed that the ratio Firmicutes/ Bacteroidetes changed from 2.1:1 at 80 days old to 1.7:1 at 100 days old and to 0.3:1 at 110 days old [15]. This demonstrates that for adaptation to plant diet the bacterial structure in rumen typifies a greater abundance of Bacteroidetes if compared to Firmicutes. So, the ratio of Firmicutes/Bacteroidetes changes from higher to lower level. Thus in this study, the low ratio of Firmicutes/Bacteroidetes (0.36:1) indicates that bacterial structure in Vietnamese goats' rumen adapted well with the diverse of lignocellulose materials in high mountain pastures and dry crop residues. That also explains the diverse lignocellulolytic genes observed in our previous publication [18] compared with Korean goats' rumen [4]. And this diversity is apparently associated with the increase in Bac-

Table 2. Ratio of the phyla Firmicutes versus Bacteroidetes in rumen of herbivorous animals and biogas fermenters

Microbiome/community	Firmicutes versus Bacteroidetes ratio in		References
	Bacterial community structure	Bacterial community for lignocellose degradation	
Vietnamese native goats' rumen	0.36:1	Mean: 0.11:1 Cellulases: 0.35:1 Hemicellulases: 0.05:1 CEs, PLs: 0.07:1	This study
Biogas fermenter	5.2:1	Cellulases: 2.8:1 CEs: 2.4:1	[17]
Hay-fed cow rumen	0.8-1.7:1	Cellulases: 1.4:1 CEs: 1.3:1	[17,20]
Pasture grass-, green bajra- fed goat rumen	1:1		[28]
Svalbard reindeer rumen	0.4-0.5:1		[7,17]
Pasture-fed sheep rumen	0.3-0.5:1		[17,20]
Giraffe rumen	1.6:1		[17,21]
Jak rumen	1.7:1		[13]
Jinnan cattle rumen	0.6:1		
Goat rumen (80 days old)	2.1:1		[15]
Goat rumen (100 days old)	1.7:1		
Goat rumen (110 days old)	0.3:1		

teroidetes.

In contrast, the ratio of Firmicutes/Bacteroidetes in our sample is quite different from the observation in rumen of Indian goats (1:1) adapted to Pasture grass, green bajra or rumen of hay-fed cow (0.8-1.7:1) [20], giraffe (1.6:1) [21], as well as rumen from jak (1.7:1) [13] (Table 2). This confirms that bacterial structure in rumen depends strongly on diet.

While, Firmicutes, Bacteroidetes are the two most abundant groups, Proteobacteria and Synergistetes were ranked third and fourth in abundance in our goat rumen. Synergistetes was observed to be abundant in goat rumen in a previous study but not abundant in other rumens [15]. Our study confirms this abundance, and supports the hypothesis that Synergistetes may be unique in goat rumen, potentially reflecting the influence of the host genotype on the bacterial structure.

Regarding the taxonomy at genera level, the two most abundant genera in our sample were *Prevotella* (35.3%) then *Bacteroides* (16%) belonging both to *Bacteroidales*. *Prevotella* was also the most abundant genus in cattle rumens [22,23] and other goat rumen studies [15,22]. However in that particular case, the *Prevotella* was reduced during adaptation to other plant matters [15] and went down to lower than 40%.

High abundance of Bacteroidetes as indicator of effective lignocelluloses digestion

In previous study, the low ratio Firmicutes/Bacteroidetes was defined to be a parameter for efficient design of a technical system to degrade lignocelluloses on an industrial scale [17], as well as for assessing the optimality of natural cellulolytic systems by herbivorous animals. However, very limited studies showed evidence for this. Here, we elucidated the association between the ratio of Firmicutes versus Bacteroidetes in a functional context to provide scientific evidence for the use of Firmicutes/Bacteroidetes as an indicator of lignocellulose breakdown capacity.

This study indeed confirms that the ratio of Firmicutes/Bacteroidetes was very low (0.11:1) in the bacterial community responsible for lignocellulose breakdown. In other words, the number of lignocellulases associated with Bacteroidetes is ~10 times when compared to Firmicutes or Bacteroidetes plays an important role in lignocellulose degradation, especially in the Vietnamese rumen. In contrast to this, biogas fermenters [17] show a ratio of 2.6:1, while hay-fed cow rumens [20] show a ratio of ~1.4:1. Overall, the ratios Firmicutes/Bacteroidetes in these studies were higher (Table 2).

Functional domains involved in lignocellulose pretreatment and hemicellulose hydrolysis depend on Bacteroidetes

The lowest ratios of Firmicutes/Bacteroidetes were observed in communities predicted to carry ORFs for hemicellulases and CEs, PLs (0.05:1 and 0.07:1 respectively) (Table 2). Again, the ratios differ between earlier studies done on other biomass degrading environments (biogas fermenter [17] and hay-fed cow rumens [20]) (Table 2). This implicates that the number of CEs and PLs for lignocellulotic pretreament, hemicellulases, are mainly found among Bacteroidetes (Figure 2B). Additionally, Bacteroidetes also harboured very high diverse GHs for hemicellulases (14 families only produced in Bacteroidetes) accompanied with CBMs (Figure 2D). This is the first time CBMs allocated with catalytic domains were so abundantly associated with Bacteroidetes.

Although Bacteroidetes is the most dominant group associated with cellulases, Firmicutes harboured some very specific cellulases GHs domains, accompanied with many CBMs including CBM2, CBM3, CBM4, and CBM63. That is contrary to the CBMs situated in hemicellulases catalytic domains. CBM63 domain found in this study also contained expansin domain that disrupts hydrogen bonding between cellulose microfibrils and enhance the accessibility of cellulases to cellulose substrates [24]. The CBM63 is usually seen in various bacteria [24] but this is the first time observed in *R. flavefaciens*. CBM63 was highly expressed in *Escherichia coli* and showed activity to increase activity of cellulase (CAS No. 9012548, Sigma-Aldrich, Toluca de Lerdo, Mexico) [18].

Typical microbial species representing lignocellulose breakdown

We also investigated whether typical species were associated with lignocellulase activity in the goat rumen, but the species distribution turned out to deviate from a previous study investigating goat rumen microbial diversity [4].

In the present study we identified *B. uniformis*, *R. bicirculans*, *E. siraeum* to be the most dominant species bearing cellulases. In contrast, *Prevotella ruminicola*, *Butyrivibrio proteoclasticus* were the two most abundant species associated with cellulases activity in the Korean goat rumen study [4]. However, the dominant species producing cellulases were not abundant species in overall bacterial structure in Vietnamese goats' rumen.

In the group producing hemicellulases, we found a correlation in abundant species when compared to the overall microbiome in the Vietnamese goats' rumen. As such, the most six abundant species producing hemicellulases also were in the list of 13 most abundant bacteria in the goat rumen community. *Clostridium* is known to inhabit the gastro-intestinal tract of ruminants and produce a range of lignocellulolytic enzymes in a multi-enzyme complex designated as cellulosome. However in this study, we did not observe the presence of *Clostridium* sp, and confirms the absence of this species in another study of goat rumen [15].

In this study, we found many potential lignocellulolytic degraders harbouring many different catalytic domains (up to 14 different catalytic domains) that function for all lignocellulose pretreatment and hemicellulose, cellulose saccharification. These degraders were affiliated in both Bacteroidetes and Firmicutes phyla. Remarkably, *R. flavefaciens* belonged to *Firmicutes* is the only gut bacterium so far shown to produce a cellulosome-type enzyme complex [25] was also found in the present study.

Recent research has uncovered that bacteria belonging to Bacteroidetes contain very versatile polysaccharide utilization loci (PULs), that play an important role in cellulose degradation in rumen, gut and faecal samples of herbivores [26]. A metagenomic data investigating the bacteria producing lignocellulolytic enzymes in cow rumen showed several PULs that contain up to 10 domains at one locus. All of these PULs were affiliated with Bacteroidetes [27]. Thus we hypothesize that the species harbouring many catalytic domains in this study may produce many enzymes in the same way as described for PULs previously. However, in this study we also emphasize the significance of lignocellulolytic degrading catalytic domains in Firmicutes as well. This novel insight will be further elucidated in future studies.

CONFLICT OF INTEREST

We certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

ACKNOWLEDGMENTS

This study was carried out with the financial support of the Project "Metagenome of some potential mini-ecologies for mining novel genes encoding effective lignocellulotic enzymes" code DTDLCN.15/14 managed by the Ministry of Science and Technology, Vietnam, in collaboration with Department of Ecological Science, VU University Amsterdam, Netherland, supported by the BE-BASIC consortium project numbers F07.003.05 and F07.003.07. We thank the National Key Laboratory of Gene Technology, Institute of Biotechnology, VAST, Vietnam for use of their facilities.

REFERENCES

1. Li RW, Wu S, Li W, Huang Y, Gasbarre LC. Metagenome plasticity of the bovine abomasal microbiota in immune animals in response to *Ostertagia ostertagi* infection. PloS One 2011;6:e24417.
2. Hess M, Sczyrba A, Egan R, et al. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science 2011;331:463-7.
3. Mosier N, Wyman C, Dale B, et al. Features of promising technologies for pretreatment of lignocellulosic biomass. Bioresour Technol 2005;96:673-86.
4. Lim S, Seo J, Choi H, et al. Metagenome analysis of protein domain collocation within cellulase genes of goat rumen microbes. Asian-Australas J Anim Sci 2013;26:1144-51.
5. Dai X, Zhu Y, Luo Y, et al. Metagenomic insights into the fibrolytic microbiome in yak rumen. PloS One 2012;7:e40430.
6. Del Pozo MV, Fernández-Arrojo L, Gil-Martínez J, et al. Microbial β-glucosidases from cow rumen metagenome enhance the saccharification of lignocellulose in combination with commercial cellulase cocktail. Biotechnol Biofuels 2012;5:73.
7. Pope PB, Mackenzie AK, Gregor I, et al. Metagenomics of the

Svalbard reindeer rumen microbiome reveals abundance of polysaccharide utilization loci. PloS One 2012;7:e38571.
8. Brulc JM, Antonopoulos DA, Miller MEB, et al. Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. Proc Natl Acad Sci USA 2009;106:1948-53.
9. Wongwilaiwalin S, Laothanachareon T, Mhuantong W, et al. Comparative metagenomic analysis of microcosm structures and lignocellulolytic enzyme systems of symbiotic biomass-degrading consortia. Appl Microbiol Biotechnol 2013;97:8941-54.
10. Berlemont R, Martiny AC. Phylogenetic distribution of potential cellulases in bacteria. Appl Environ Microbiol 2013;79:1545-54.
11. Nelson KE, Zinder SH, Hance I, et al. Phylogenetic analysis of the microbial populations in the wild herbivore gastrointestinal tract: insights into an unexplored niche. Environ Microbiol 2003;5:1212-20.
12. Sundset MA, Praesteng KE, Cann IKO, Mathiesen SD, Mackie RI. Novel rumen bacterial diversity in two geographically separated sub-species of reindeer. Microb Ecol 2007;54:424-38.
13. An D, Dong X, Dong Z. Prokaryote diversity in the rumen of yak (*Bos grunniens*) and Jinnan cattle (*Bos taurus*) estimated by 16S rDNA homology analyses. Anaerobe 2005;11:207-15.
14. Kittelmann S, Janssen PH. Characterization of rumen ciliate community composition in domestic sheep, deer, and cattle, feeding on varying diets, by means of PCR-DGGE and clone libraries. FEMS Microbiol Ecol 2011;75:468-81.
15. Han X, Yang Y, Yan H, et al. Rumen bacterial diversity of 80 to 110-day-old goats using 16S rRNA sequencing. PLoS ONE 2015;10:e0117811.
16. Tajima K, Aminov RI, Nagamine T, et al. Diet-dependent shifts in the bacterial population of the rumen revealed with real-time PCR. Appl Environ Microbiol 2001;67:2766-74.
17. Güllert S, Fischer MA, Turaev D, et al. Deep metagenome and metatranscriptome analyses of microbial communities affiliated with an industrial biogas fermenter, a cow rumen, and elephant feces reveal major differences in carbohydrate hydrolysis strategies. Biotechnol Biofuels 2016;9:121.
18. Do TH, Le NG, Dao TK, et al. Metagenomic insights into lignocellulose-degrading genes through Illumina-based *de novo* sequencing of the microbiome in Vietnamese native goats rumen. J Gen Appl Microbiol Year Month Date [Epub]. https://doi.org/10.2323/jgam.2017.08.004
19. Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. Genome Res 2007;17:377-86.
20. Henderson G, Cox F, Kittelmann S, et al. Effect of DNA extraction methods and sampling techniques on the apparent structure of cow and sheep rumen microbial communities. PLoS ONE 2013;8:e74787.
21. Roggenbuck M, Sauer C, Poulsen M, Bertelsen MF, Sørensen SJ. The giraffe (*Giraffa camelopardalis*) rumen microbiome. FEMS Microbiol Ecol 2014;90:237-46.
22. 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.
23. 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;75:165-74.
24. Martinez-Anaya C. Understanding the structure and function of bacterial expansins: a prerequisite towards practical applications for the bioenergy and agricultural industries. Microb Biotechnol 2016;9:727-36.
25. Flint HJ, Scott KP, Duncan SH, Louis P, Forano E. Microbial degradation of complex carbohydrates in the gut. Gut Microbes 2012;3:289-306.
26. Naas AE, Mackenzie AK, Mravec J, et al. Do rumen *Bacteroidetes* utilize an alternative mechanism for cellulose degradation? mBio 2014;5:e01401-14.
27. Wang L, Hatem A, Catalyurek UV, Morrison M, Yu Z. Metagenomic insights into the carbohydrate-active enzymes carried by the microorganisms adhering to solid digesta in the rumen of cows. PLOS ONE 2013;8:e78507.
28. Patel J, Jhala M, Soni P, et al. Molecular characterization and diversity of rumen bacterial flora in Indian goat by 16S rDNA sequencing. Online Vet J 2011;6:article 77.