Systematic and Applied Microbiology 40 (2017) 160–170

Contents lists available at ScienceDirect

Systematic and Applied Microbiology

journal homepage: www.elsevier.de/syapm

In depth analysis of rumen microbial and carbohydrate-active enzymes profile in Indian crossbred cattle☆

V. Lyju Jose[a,b], Ravi P. More[c,1], Thulasi Appoothy[a,*], A. Sha Arun[d]

[a] Rumen Microbiology Laboratory, Animal Nutrition Division, National Institute of Animal Nutrition and Physiology, Bangalore, India
[b] Department of Biotechnology, Jain University, Bangalore, India
[c] Bioinformatics Laboratory, Division of Molecular Entomology, National Bureau of Agricultural Insect Resources, Bangalore, India
[d] Wildlife SOS, Bannerghatta Bear Rescue Centre, Bangalore, India

ARTICLE INFO

Article history:
Received 23 September 2016
Received in revised form 18 February 2017
Accepted 20 February 2017

Keywords:
Rumen
HF cross
Pfam
Metagenomics
CAZymes
MiSeq

ABSTRACT

Rumen houses a plethora of symbiotic microorganisms empowering the host to hydrolyze plant lignocellulose. In this study, NGS based metagenomic approach coupled with bioinformatic analysis was employed to gain an insight into the deconstruction of lignocellulose by carbohydrate-active enzymes (CAZymes) in Indian crossbred Holstein-Friesian cattle. Cattle rumen metagenomic DNA was sequenced using Illumina-MiSeq and 1.9 gigabases of data generated with an average read length of 871 bp. Analysis of the assembled sequences by Pfam-based Carbohydrate-active enzyme Analysis Toolkit identified 17,164 putative protein-encoding CAZymes belonging to different families of glycoside hydrolases (7574), glycosyltransferases (5185), carbohydrate-binding modules (2418), carbohydrate esterases (1516), auxiliary activities (434) and polysaccharide lyases (37). Phylogenetic analysis of putative CAZymes revealed that a significant proportion of CAZymes were contributed by bacteria belonging to the phylum Bacteroidetes (40%), Firmicutes (30%) and Proteobacteria (10%). The comparative analysis of HF cross rumen metagenome with other herbivore metagenomes indicated that Indian crossbred cattle rumen is endowed with a battery of CAZymes that may play a central role in lignocellulose deconstruction. The extensive catalog of enzymes reported in our study that hydrolyzes plant lignocellulose biomass, can be further explored for the better feed utilization in ruminants and also for different industrial applications.

Introduction

The rumen is a natural habitat and a specialized bioreactor possessing a unique genetic resource of potent carbohydrate-active enzymes, which have evolved into an efficient biocatalyst system for the breakdown of plant cell-wall polysaccharides produced by the resident rumen microorganisms [13,23]. However, the majority of these microbes are uncultivable. Rumen microbiota primarily mediates the bioconversion of crop residues into simpler compounds viz., short-chain fatty acids that can be utilized by the host [14]. Rumen microbial fermentation is of enormous significance wherein the solar energy trapped in the plant material gets converted into other sources of food (milk, meat etc.). Additionally, the rumen microbes live in a symbiotic relationship with the host contribute significantly toward its health and improved production performance [29]. Approximately 50% of the ingested lignocellulosic biomass passes through the rumen undigested due to its inherent recalcitrant nature and slow retention time in the rumen [45]. The scarcity of competent lignocellulolytic enzymes is the major bottleneck for the large scale bioconversion of plant biomass into biofuels [32]. Considerable efforts have been made in the past to understand the rumen microbial community structure and to decode lignocellulose degrading mechanism in the rumen. A battery of carbohydrate-active enzymes (CAZymes) functioning in a concerted fashion is required for the complete hydrolysis of plant cell-wall polysaccharides in the rumen [4]. Rumen microbiota and its multifaceted foregut fermentation process that degrade lignocellulose have attracted tremendous attention and efforts are being made to improve both the feed utilization efficiency in livestock and the production of economical cellulosic biofuel [8,4].

Metagenomic sequencing technologies are extensively used to address the complex process of lignocellulose degradation in

☆ The raw data obtained had been deposited in the NCBI Sequence Read Archive (SRA) with the Bio Project ID-PRJNA291680.

* Corresponding author. National Institute of Animal Nutrition and Physiology, Bangalore, India. Tel.: +080 25711304; fax: +080 25711420.
E-mail address: thulasi.appoothy@gmail.com (T. Appoothy).
[1] Present Address: National Centre for Biological Sciences (NCBS), Bangalore, India.

http://dx.doi.org/10.1016/j.syapm.2017.02.003
0723-2020/

ruminants. Although different metagenomic studies have been conducted to elucidate the carbohydrate-active enzyme profiles from the rumen of cow [13], yak [7], reindeer [37], Jersey cow [43], Angus cross [4], and buffalo [39], there are no comprehensive scientific reports are available on metagenomic evaluation of CAZyme diversity in HF crossbred cattle from India. This study was undertaken with the objectives of deciphering CAZymes diversity and to determine the microbes contributing to this diversity of enzymes in HF cross cattle fed finger millet straw along with para grass and concentrate mixture. A comparative analysis of our data with published herbivore metagenomes was undertaken to identify whether any exclusive CAZyme families prevalent in the HF cross rumen ecosystem.

Materials and methods

Experimental design, animals, and sample collection

Three fistulated crossbred steers with an average body weight of 380 ± 15 kg were maintained in individual stands for feeding experiment at the Experimental Livestock Unit (ELU), National Institute of Animal Nutrition and Physiology, Bangalore, India. The ethical committee at the National Institute of Animal Nutrition and Physiology, approved the animal feeding, sample collection and other experimental procedures used in this study. The animals were fed with a maintenance ration (ICAR) [16], composed of roughages (para grass and finger millet straw) and concentrate mixture (maize 37%, groundnut cake 32%, wheat bran 28%, mineral mixture 2% and salt 1%) for a period of 21 days. Rumen contents were collected from all three animals, 2 h after feeding at the end of the experiment. Approximately 50 ml of rumen digesta samples were collected through the rumen fistula and immediately transported to the laboratory for further processing. Rumen digesta samples were mixed and strained through two layers of muslin cloth and immediately flash frozen in liquid nitrogen. Both liquid and solid portions of rumen digesta samples were then stored at −86 °C until further processing.

Total DNA extraction and quantification

The frozen rumen samples were thawed at room temperature and the solid rumen digesta samples were resuspended in phosphate buffered saline (Amresco, USA), for 2 h with vortexing to liberate the microbes adhering to feed particles, and mixed with the rumen fluid sample. The rumen fluid samples were centrifuged at 4000 rpm for 5 min and the supernatant was used for total DNA extraction. In brief, the rumen fluid was centrifuged at 13,000 rpm for 10 min and the pelleted cells were resuspended in a mix of 800 μl of CTAB lysis buffer (2% CTAB, 1.4M NaCl, 20 mM EDTA and 100 mM Tris–HCl), (Amresco, USA) and 0.2 g of glass beads (0.1 mm), (Biospec products Inc, USA) was added and kept in a Mini bead beater (Biospec products Inc, USA) for 3 min. Subsequently, 10 μl of 20 mg/ml proteinase K (Amresco, USA) and 10 mg/ml lysozyme (Amresco, USA) were added to the above mixture and incubated at 37 °C for 1 h. The tubes were then incubated at 70 °C for 30 min with intermittent mixing. An equal volume of phenol:chloroform:isoamyl alcohol mixture (25:24:1) (Amresco, USA) was added to the above lysate and mixed by inverting until a thick milky white emulsion was formed. After centrifugation at 13,000 rpm for 10 min, the supernatant was transferred to a fresh tube and total DNA was precipitated using 0.3 volumes of chilled ethanol (Merck, Canada). The precipitated DNA was then washed twice with 70% ethanol and the pellet was dried in a vacuum concentrator (Concentrator 5301, Eppendorf, Germany). The quality of extracted genomic DNA was assessed by electrophoresis in 0.8% agarose gel for a single intact band, and the A260/280 ratio was determined using Nanodrop 8000 (Thermo Scientific, USA). Qubit 2.0 Fluorometer (Invitrogen, USA) was used to measure the quantity of DNA.

Preparation of 2 × 300 MiSeq libraries and whole metagenome sequencing

The paired-end sequencing library was prepared using Illumina, truseq Nano DNA LT Library Preparation Kit (Illumina, California, United States). About 200 ng of genomic DNA was fragmented by Covaris (CovarisInc, USA) to generate a mean fragment distribution of 550 bp. The fragments were subsequently subjected to end repair (converts the overhangs to blunt ends) using End Repair Mix. Indexing adapters were then ligated to the ends of DNA fragments and purified. The size-selected products were PCR amplified as per the kit protocol. The amplified library was analyzed in Bioanalyzer 2100 (Agilent Technologies, USA) using a High Sensitivity DNA chip (Agilent Technologies, USA) as per the manufacturer's instructions. The library was then loaded onto the Illumina-MiSeq platform for cluster generation and then subjected to paired-end sequencing.

Metagenome assembly and analysis

De novo assembly of high quality data was accomplished using the CLC Genomics workbench 6.0 at default parameters (Minimum contig length: 200, Automatic word size: yes, Perform scaffolding: yes, Mismatch cost: 2, Insertion cost: 3, Deletion cost: 3, Length fraction: 0.5, Similarity fraction: 0.8). The whole metagenome dataset (147,749,531 base pairs) was uploaded on Metagenome Rapid Annotation using Subsystem technology (MG-RAST) server [30], for further analysis. MG-RAST quality filters were used to evaluate the quality of the uploaded sequences. Sequences which did not fulfill the quality criteria, were removed from further analysis. The data were normalized before subsequent analysis to enable the analysis of the microbial diversity and the CAZyme repertoire within the rumen of HF crossbred steers.

Carbohydrate-active enzyme Analysis Toolkit (CAT) analysis

The putative protein-coding regions of rumen whole metagenome data were predicted using FragGeneScan [31], a gene-calling program that combines sequencing error models and is built on a Hidden Markov Model (HMM) to improve the prediction of the protein-coding region in short reads. Predicted putative protein coding regions were scrutinized for the presence of carbohydrate-active enzymes in the CAZy database [5] (http://www.cazy.org) and Pfam-based annotation of sequences were performed using the Carbohydrate-active enzyme Analysis Toolkit (CAT) [33], with an e-value of 1e-5. The amino acid sequences were thoroughly analyzed for different classes of CAZymes: GHs, GTs, PLs, CEs, CBMs and AAs. The results obtained were further analyzed manually to determine the proportions of the different CAZymes present in the rumen metagenome data.

Microbial community profile analysis of CAZyme contributors

The CAZymes obtained in the present study were compared with other accessible metagenomic data sets, Angus Simmental cross [4] (MG-RAST ID: 4441679.3) (0.027GB), Jersey cow [43] (Genbank accession numbers KC246771-KC247082) (0.28GB), and termite gut [44] (MG-RAST ID: 4442701.3) (0.062GB). The metagenomic data were downloaded from respective public domains and all datasets were normalized before further processing. CAT analysis for all three metagenomes were performed as mentioned earlier to identify the orfs encoding different classes of CAZymes. Putative