THE BOVINE OTHER GENOME: A CATALOG OF GENES OF THE BOVINE RUMEN MICROBIOME

Junhua Li[1, 2, 3], Huanzi Zhong[1, 2], Yuliaxis Ramayo-Caldas[4,5*], Nicolas Terrapon[6, 7], Vincent Lombard[6, 7], Gabrielle Potocki-Veronese[8], Jordi Estellé[4], Milka Popova[9], Ziyi Yang[1, 2], Hui Zhang[1, 2], Fang Li[1, 2], Shanmei Tang[1, 2], Weineng Chen[1], Bing Chen[1, 2], Jiyang Li[1], Jing Guo[1, 2], Cécile Martin[9], Emmanuelle Maguin[10], Xun Xu[1, 2], Huanming Yang[1, 11], Jian Wang[1, 11], Lise Madsen[1, 12,13] Karsten Kristiansen[1, 13], Bernard Henrissat [6, 7,13], Stanislav D. Ehrlich[14,15], Diego P. Morgavi[9]

[1]BGI-Shenzhen, Shenzhen 518083, China. [2]China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China. [3]School of Bioscience & Bioengineering, South China University of Technology, Guangzhou, China. [4]INRA, Institut National de la Recherche Agronomique, Génétique Animale et Biologie Intégrative, AgroParisTech, Université Paris-Saclay, 78350, Jouy-en-Josas, France. [5]Animal Breeding and Genetics Program, Institute for Research and Technology in Food and Agriculture (IRTA), Torre Marimon, Caldes de Montbui, 08140, Spain. [6] CNRS UMR 7257, Aix-Marseille University, 13288 Marseille, France. [7]INRA, USC 1408 AFMB, 13288 Marseille, France. [8]LISBP, Universite de Toulouse, CNRS, INRA, INSA, 31077, Toulouse, France. [9]INRA, UMR Herbivores, Université Clermont Auvergne, VetAgro Sup, F-63122 Saint-Genès Champanelle, France. [10] INRA, Micalis Institute, AgroParisTech, Université Paris-Saclay, 78350, Jouy-en-Josas, France. [11]James D. Watson Institute of Genome Sciences, Hangzhou 310058, China. [12] Laboratory of Genomics and Molecular Biomedicine, Institute of Marine Research (IMR), Postboks 1870 Nordnes, 5817 Bergen, Norway. [13] Laboratory of Genomics and Molecular Biomedicine, Department of Biology, University of Copenhagen, 2100 Copenhagen Ø, Denmark. [14] MGP MetaGenoPolis, INRA, Université Paris-Saclay, 78350 Jouy en Josas, France, [15] Centre for Host Microbiome Interactions, Dental Institute, King's College London, UK. * presenting author.

Introduction

The rumen microbiota is a unique feature of ruminants and a thorough knowledge of the genetic potential of rumen symbiotic microbes provides opportunities for improving the sustainability of ruminant production systems. In this study, we established a catalog of rumen prokaryotes genes that enabled us to decipher functions of the microbiome, in particular the capacity to deconstruct structural carbohydrates from forages. In addition we performed a comparison with human, pig and mouse gut metagenome catalogs.

Material and Methods

Total rumen content samples from 10 animals used for deep sequencing metagenome were taken at the experimental slaughterhouse of the INRA Centre Auvergne-Rhône-Alpes. Total rumen content samples from 77 animals were also collected. These 77 animals, from two different genetic stocks, were fed diets characteristics of beef and milk production systems. Beef cattle, represented by Charolais breed were fed fattening diets high (n=16) and low (n=18) in starch and lipids; whereas Holstein dairy cows were fed a corn silage and concentrate diet (n=23) or grazed a natural prairie (n=20). DNA was extracted following the method described by Yu and Morrison[1]. Paired-end (PE) metagenomic libraries were constructed and sequenced following Illumina HiSeq2000's instruction. Quality control and bovine DNA removal for each sample were independently processed using MOCAT pipeline as previously described[2]. Three public rumen microbial datasets were also used in this study: (i) a cow rumen microbiome sequenced at DOE's Joint Genome Institute (JGI) in 2011 (NCBI accession number SRA023560), (ii) eight rumen metagenomics samples from beef steers[3] (European Bioinformatics Institute (EBI),

PRJEB10338), (iii) 409 reference rumen microbial genome sequences and one rumen metagenomic dataset from the Hungate1000 Project[4,5].

Construction of the rumen microbial gene catalog

High quality reads from 10 deep sequenced samples were processed with MOCAT[2] including de novo individual assembly. The assembled contigs with length equal to or greater than 100 bp were followed by gene prediction using MetaGeneMark[6] and redundant genes were removed (CD-HIT12, ≥ 95% identity and ≥ 90% overlap). Taxonomic assignments of genes from rumen, mouse, pig and human guts were performed using CARMA3[7] on the basis of BLASTP[8] (V2.2.24) against the NCBI-NR database (v20130906 for rumen, mouse, pig guts; v20160219 for human gut). Microbiome from these four species were compared at different taxonomic levels. Functional annotation based on Kyoto Encyclopedia of Genes and Genomes (KEGG) database was performed using an in-house pipeline. Annotation of the carbohydrate-active enzymes (CAZymes) of each catalog was performed by comparing the predicted protein sequences to those in the CAZy database and to Hidden Markov models (HMMs) built from each CAZy family[9]. The gene profiles of 77 rumen samples were generated by aligning high-quality clean reads to the gene catalogue. The relative abundance of each KEGG orthologous group (KO) and CAZy enzyme was calculated from the abundance of its genes. Finally, annotation of antibiotic resistance genes (ARGs) was done as previously reported in the pig metagenome catalogue[10] by using the ARDB data base[11].

Results and Discussions

Using deep metagenome sequencing we identified 13,825,880 non-redundant prokaryote genes from the bovine rumen and constructed 324 metagenomic species (MGS). Compared to the first rumen gene catalog published by Hess et al[11], the number of non-redundant genes discovered in this study is more than 5 fold larger. The mapping reads from 77 additional rumen samples obtained in this study increased from ~10% using the previous catalog to ~40% (11-51%) (Figure 1). This confirms that the representativeness of the rumen catalog has been improved, even if the mapping efficiency was relatively low, as compared to 80% for the human gut microbiome[13].

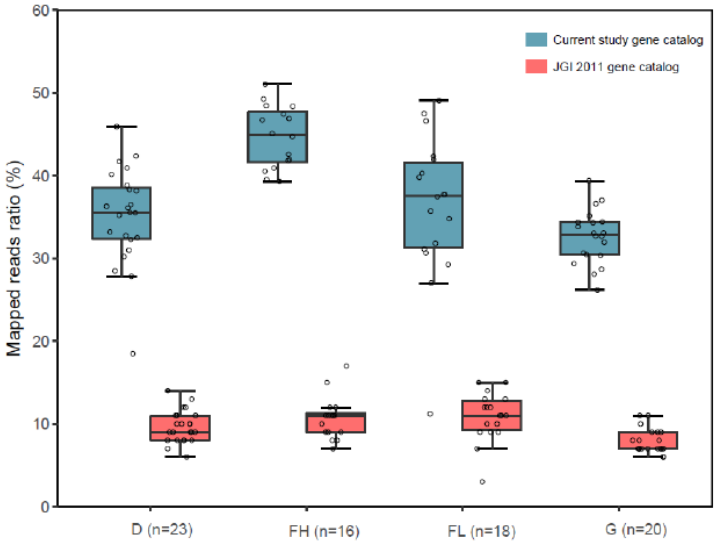

Figure 1. Percentage of total reads in the current study (n=77 samples) in four different diet groups that could be mapped to current study gene catalog and JGI 2011 gene catalog.

Genes were taxonomically classified and compared to genes from human, mouse and pig gut catalogs [10,13-14]. Up to 42.7% of rumen genes could be annotated to known phyla. This value is similar to pig gut (41.3%) but lower than the human (55.9%) and mouse gut metagenomes (59.6%). Firmicutes and Bacteroidetes were predominant in all catalogs representing 84-94% of assigned genes. For the rumen, however, the proportion of Firmicutes and that of Bacteroidetes was lower and higher, respectively, than for the other three catalogs. *Prevotella* was the most abundant rumen genus with 39% of genus-annotated genes assigned. Other abundant genera were *Treponema*, *Butirivibrio*, *Methanobrevibacter* and *Ruminococcus* that were absent or at lower proportions in other catalogs, particularly in human. Based on abundance profiles and clustering methods 324 metagenomic species (MGS), with an average size of 1.8 Mbp were identified. For genomes from the Hungate 1000 project[4], which are representative of the diversity of cultured rumen bacteria and archaea, mapping rate was 5.4%. In contrast, only 0.1% of reads mapped to the metagenomic species described by Hess et al.[12]. Only 23 MGS were similar to genomes from the Hungate 1000 project, highlighting the novelty represented by our MGS as more than 90% have no close cultured representative.

To obtain insights into the efficient deconstruction of structural plant polysaccharides by symbiotic gastrointestinal microbes we analyzed carbohydrate active enzymes (CAZymes) in the rumen ecosystem. Glycoside hydrolases (GHs) and polysaccharide lyases (PLs) are the most relevant classes of CAZymes as they orchestrate the breakdown of plant material and of diverse polysaccharides which are encountered in the rumen ecosystem. The rumen catalog reported here encodes ~290,000 GHs (97.4%) and PLs (2.6%) modules, which belong to 114 distinct GHs families and 18 PLs famílies. The CAZyme profile in the rumen catalog was compared to the mouse, pig and human reference gut catalogs. The specific adaptation of the rumen catalog to herbivory was confirmed by comparing its GH+PL family counts against the human catalog after normalization (**Figure 2**). The most enriched GH families in the rumen are involved in the degradation of plant polysaccharides while the more depleted families of GHs are those degrading animal (host) glycans (**Figure 2**). These observations are not only in accord with cattle normal diet but they are also in agreement with the absence of a glycoprotein-rich mucus lining the rumen as opposed to the lower gastrointestinal tract.

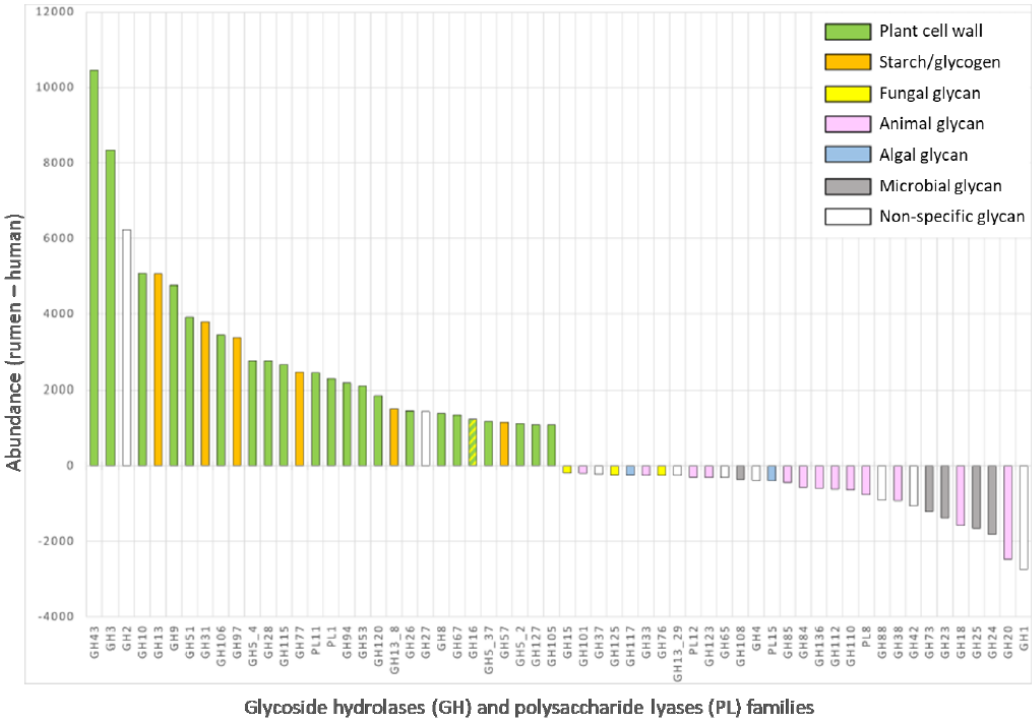

Figure 2. Enrichment or depletion of glycoside hydrolases and polysaccharide lyases in the bovine rumen as compared to human gut. Human counts were normalized to rumen catalog size before comparison.

Finally, the presence of ARGs in the rumen gene catalogue was evaluated and 42 two ARGs encoding resistance to 27 antibiotics were detected (**Figure 3**). The most abundant resistances were to tetracycline and bacitracin with Charolais animals harboring globally a higher proportion of these genes (**Figure 3**). To be noted, antibiotics as growth promoters were never used on these animals. The diversity of ARG is low compared to pig feces where resistance to up to 52 antibiotics was reported[9], even in farms with no use of growth promoting antibiotics.

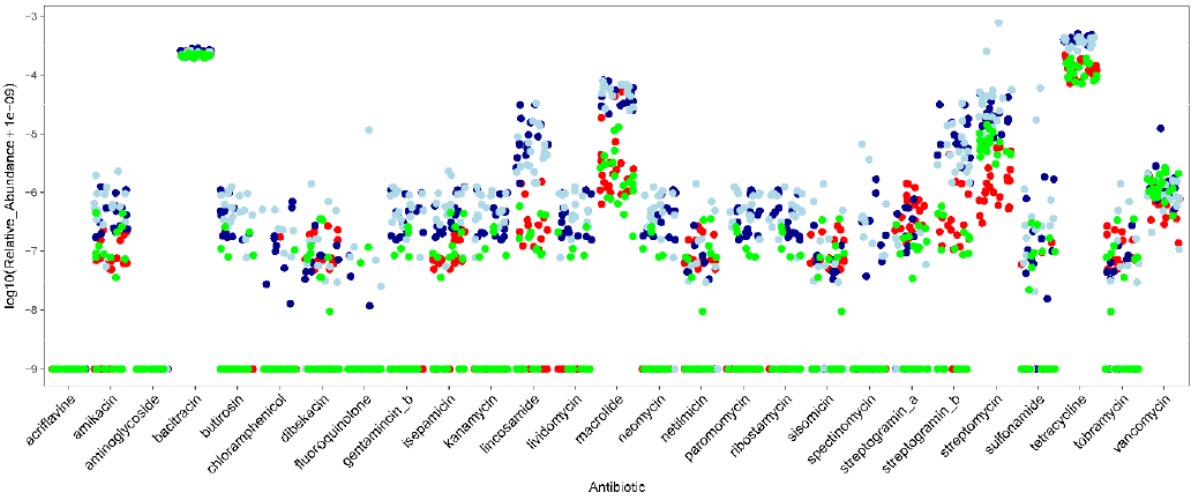

Figure 3. Relative abundances (log10 scale) of antibiotic types found in each of the 77 individual. In every column on the x axis each dot represents an animal, with colours according to diets: grazing (G; green), dairy (D; red), fattening high-start (FH; dark blue) and fattening low-starch (FL; light blue).

In summary, compared to human, pig and mouse gut metagenome catalogs, the rumen as expected is richer in functions and microbial species associated with the degradation of plant cell wall material and production of methane. Genes coding for enzymes that deconstruct plant polysaccharides showed a particularly high richness that is otherwise impossible to infer from available genomes or shallow metagenomics sequencing. These data bring new insights on functions, enzymes and microbes of the rumen, critical to understand phenotypes and biological processes.

References

[1] Yu & Morrison, 2004. Biotechniques 36, 808-812. [2] Kultima et al, 2012. PLoS One 7, e47656. [3] Wallace et al, 2015. BMC Genomics 16, 839. [4] Kelly, 2016. genome.jgi.doe.gov. [5] Seshadri et al, Nature Biotech 2018. [6]Zhu et al, 2010. Nucl Acids Res 38, e132. [7]Gerlach & Stoye, 2011. Nucl Acids Res 39, e91-e91. [8]Altschul et al, 1990. J Mol Biol 215, 403-410. [9]Lombard et al, 2014. Nucl Acids Res 42, D490-D495. [10]Xiao et al, 2016. Nat Microbiol 1, 16161. [11]Liu, 2009. Nucl Acids Res 37, D443-447. [12]Hess et al, 2011. Science 331, 463-467. [13]Li et al, 2014. Nat Biotech 32, 834-841. [14]Xiao et al, 2015. Nat Biotech 33, 1103-1108.