Carbohydrate Active Enzyme Profile of Bar-Headed Goose (Anser indicus) Gut Metagenome

Wang Wen¹, Zheng Si Si¹,², Sharshov Kirill³, Sun Hao¹, Yang Fang¹, Wang Xue Lian¹ and Li Lai Xing¹*$

¹Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xi'ning, China; ²University of Chinese Academy of Sciences, Beijing, China; ³Research Institute of Experimental and Clinical Medicine, Novosibirsk, Russia

*Corresponding author: lxli@nwipb.cas.cn

ABSTRACT

Carbohydrate Active enzymes (CAZymes), encoded by microbes that inhabit the gastrointestinal tract, play a crucial role in breakdown of complex dietary carbohydrates into components that can be absorbed by host intestinal epithelium. Bar-headed goose (Anser indicus), an herbivorous bird, has become one of the most popular wild birds for artificial rearing industries in several provinces of China. To understand how CAZymes in the gut of artificially reared Bar-headed geese are influenced by artificial diets, we describe here analysis of CAZymes from gut metagenomic data from both wild and artificially reared Bar-headed geese. A total of 172 and 215 CAZymes were found in wild and artificially reared Bar-headed geese, respectively. Glycoside hydrolases and glycosyl transferases were found to be the two most abundant categories of CAZymes in both groups. Comparative study showed a total of 22 significantly different CAZymes between wild and artificially reared group. Further, these significantly different CAZymes were observed to be specifically abundant in the Firmicutes phyla in wild group, whereas Bacteroidetes phyla in artificially reared group. These results provide a global view of CAZyme profiles of Bar-headed goose, and make an original contribution to the artificial diets management for rearing this bird.

Key words: Bar-headed goose, Gut metagenome, Carbohydrate active enzyme, High-throughput sequencing, Artificial rearing

INTRODUCTION

Advances in next-generation sequencing technologies coupled with new bioinformatic developments permit the study of the various microbiome (the complex collection of microorganisms, their genes, and their metabolites) of the human and animal bodies at an unprecedented scale (Human Microbiome Project, 2012, Qin et al., 2010, Weinstock, 2012). These microorganisms are no longer considered as disease-producing pathogens, rather they are now considered as a co-evolutionary partner contributing to both host health and disease (Backhed et al., 2005, Ley et al., 2008, Shapira, 2016). The gastrointestinal tract contains the largest collection of microorganisms, which are collectively named “gut microbiome”. A growing number of studies have shown that gut microbiome play an important role in several fundamental and crucial processes such as host development (Malmuthuge et al., 2015), immune homeostasis (Ahern et al., 2014), nutrient assimilation (Kau et al., 2011), vitamins synthesis and sterols metabolism in the host (O'mahony et al., 2015), and diseases (e.g., obesity, diabetes, and cancer) in humans and other animals (Kinross et al., 2011, Lee and Hase, 2014).

The gut microbiota of herbivores is also the main driver of plant cell wall degradation as genomes of these animals do not encode most of the enzymes needed to degrade the structural polysaccharides present in plant material. In general and as expected, the gut microbiome of herbivorous animals encoded for high numbers of carbohydrate active enzymes (CAZymes). Thereby herbivores can gain 70% of their energy from microbial polysaccharide breakdown (Flint et al., 2012). CAZymes designate the ensemble of the enzymes that catalyze the assembly, breakdown or modification of oligosaccharides, polysaccharides and glycoconjugates. They are currently comprising 135 families of glycoside hydrolases (GHs), 24 polysaccharide lyases (PLs), 16 carbohydrate esterases (CEs), and 98 glycosyl transferases (GTs) (Lombard et al., 2014). In addition there are currently 64 families of...
carbohydrate binding modules (CBMs) that are indirectly associated with carbohydrate metabolism (Cockburn and Svensson, 2013), and in order to simplify our results and discussions, they will not be described in this manuscript.

The identification of these CAZymes is constrained due to the fact that the majority of gut microbes are uncultivable. Techniques for mining metagenomes have developed rapidly in recent years, and the huge output data are valuable resources for identifying putative CAZymes from both culturable and unculturable gut microbes (El Kaoutari et al., 2013, Vidal-Melgosa et al., 2015). Among avians, gut metagenome research mainly focused on commercially farmed species such as chicken (Sergeant et al., 2014), turkey (Lu and Domingo, 2008) and ostrich (Matsui et al., 2010). Only a very limited number of wild birds' metagenome have been reported in the literature (Roggenbuck et al., 2014, Waite and Taylor, 2015). However, very limited information is available for estimating gut microbes' functional activities because metagenome analyses tend to reflect bacterial composition, but not bacterial activity.

As one of the dominant waterfowl species in wetland areas in Qinghai-Tibetan Plateau, artificial rearing of Bar-headed geese (Anser indicus) is increasing in several provinces of China since year 2003 for the purpose of conservation and economic development. In our previous study (Wang et al., 2016b), we showed that the core gut microbiota of wild Bar-headed geese were dominated by Firmicutes, Proteobacteria, Actinobacteria, and Bacteroidetes. Furthermore, by comparative analysis of the gut microbiome of Bar-headed geese in different rearing conditions, we found that the artificially reared Bar-headed geese had significantly more Bacteroidetes compared to the wild ones (Wang et al., 2016a). As an herbivorous bird, the nourishment is composed of highly fibrous plant material, mainly grass, leaves, twigs and seeds (Middleton and van der Valk, 1987). However, the CAZymes profile of this species remains unknown.

Therefore, the first aim of the present study was to provide an overview of the CAZymes present in the gut microbiota of Bar-headed geese. To the best of the authors' knowledge, this is also the first data set to report CAZymes profile for the Bar-headed geese gut metagenome. The second aim was to evaluate the variation of CAZymes between wild Bar-headed geese group and artificially reared group.

MATERIALS AND METHODS

Ethics statement

This study was carried out in strict accordance with the Animal Management Rule of the National Health and Family Planning Commission, People’s Republic of China (Documentation 55, 2001). The research protocol was reviewed and approved by the Animal Care and Use Committee of the Chinese Academy of Sciences. The manuscript does not contain experiments using animals and human studies.

Faecal samples collection

Two groups of faecal samples were obtained in Qinghai province, China, on 2nd July, 2014. The wild Bar-headed geese faecal samples (n = 2) were collected from Ha Da-tan wetland (37°07′41.3″N, 99°43′39.9″E, elevation 3,100 m). The artificially reared (abbreviation: AR) Bar-headed geese faecal samples (n = 2) were derived from Bu Ha River Estuary (36°58′25.5″N, 99°50′19.2″E, elevation 3,197 m) in Qinghai Lake. The AR populations lived freely in both wild and captivity environments, fed on both natural and artificial diets (commercial blended feed for chicken). These populations were not treated with antibiotics. About 1 g of faecal samples were collected from faecal balls, avoiding collection of faecal material that was touching the ground. All samples were placed in sterile containers and transported to the laboratory in car-carried refrigerator. In laboratory, faecal samples were kept frozen at -80°C until processing.

DNA extraction and shotgun metagenomic sequencing

Genomic DNA was isolated from approximately 1 g of faecal sample using the E.Z.N.A. ® stool DNA Kit (Omega Bio-tek, Norcross, GA, USA) following the manufacturer’s instruction. DNA concentration and quality were determined using QuantiFluor™ - ST (Promega, Madison City, WI, USA) and gel electrophoresis respectively. With the extracted DNA, library construction was performed on an Illumina Hiseq2500 platform according to the standard protocols. Metagenomes data are now available at NCBI under the Sequence Read Archive (SRA) database with accession No.SRP072790 and No.SRP072793.

Bioinformatic analysis of sequencing data

Raw sequences obtained from 4 metagenomic samples were subjected to a quality check using the FastQC software (version v0.11.3) (Andrews, 2012). All samples showed satisfactory values for each parameter tested. Next, the sequences were run through Trimomatic (version 0.33) (Bolger et al., 2014) to remove low quality base pairs using these parameters (SLIDINGWINDOW: 4:15 MINLEN: 36). Further, the host specific and other eukaryotic sequences were removed by parsing the NCBI non-redundant protein database (NCBI-nr) taxonomy assignment using the lowest common ancestor (LCA) algorithm in MEGAN (Huson et al., 2007). The protein-encoding open reading frame (ORF) from these resulting cleaned sequences were predicted using Prodigal (version 2.6.2) (Hyatt et al., 2012). CAZymes were identified from these protein coding sequences using dbCAN (Yin et al., 2012), a web resource that implements hidden Markov models (HMMs) for automated signature domain annotations representative of each individual category and family. For the dbCAN assignments, a minimum e-value cut-off of 1 × 10⁻3 was used. Two-sided Welch's t-test (Parks and Beiko, 2010) in STAMP software package was applied to test the CAZymes categories differences between the AR and Wild group. Categories difference with a p value of <0.05 were considered to be significant. All figures were generated with customized R scripts.

RESULTS

Diversity profile of CAZymes found in Bar-headed geese gut metagenome

In the wild group, a total of 172 CAZymes were found (Table S1). They were 14 CEs (4,188 reads), 93
GHs (30,896 reads), 50 GTs (11,356 reads) and 15 PLs (246 reads). A total of 215 CAZymes were found in the AR group (Table S2), including 16 CEs (30,950 reads), 116 GHs (331,970 reads), 64 GTs (77,767 reads) and 19 PLs (9,687 reads). The average proportion of each class of the CAZymes revealed higher proportion of GHs and GTs in both groups. The average detection frequency of these totally five CAZymes among the total clean reads generated from gut metagenome was 0.94% and 1.91% in the wild and AR group, respectively (Table S3).

Comparison of CAZymes between wild and artificially reared Bar-headed geese

The proportion of each CAZymes was tested statistically with STAMP using Welch’s t-test. As shown in Figure 1, GH73, 90, 119, 126 and GT45, 96 were found to be significantly higher in wild group (P<0.05) compared to AR group. Another 16 CAZymes had higher proportions in AR group (P<0.05), including GH10, 30, 35, 51, 53, 55, 77, 82, 84, 98, 121, 127, 130, PL11, 12 and CE12.

Phylogenetic classification of significantly different CAZymes

These 22 significantly different CAZymes were further analyzed for their microbial origin. In the wild group, all the 6 CAZymes were highly enriched in phylum Firmicutes (Table 1). In the AR group, 15 CAZymes were found to be enriched in phylum Bacteroidetes, while GH77 were found to be only highly enriched in phylum Firmicutes (Table 2). These results indicate that different types of gut microbes contribute differently to the occurrence of CAZymes.

![Fig. 1](image-url) Significant CAZymes differences as a result of Welch’s t-test between the AR and Wild group conducted with the STAMP program. Difference with a p value of <0.05 were considered to be significant.

<table>
<thead>
<tr>
<th>CAZymes</th>
<th>Known activities</th>
<th>Corresponding phylum (average proportion)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GH73</td>
<td>mannosyl-glycoprotein endo-β-N-acetylglucosaminidase</td>
<td>Firmicutes (80.45%)</td>
</tr>
<tr>
<td>GH90</td>
<td>endorhamnosidase</td>
<td>Firmicutes (83.335%)</td>
</tr>
<tr>
<td>GH119</td>
<td>α-amylase</td>
<td>Firmicutes (63.72%)</td>
</tr>
<tr>
<td>GH126</td>
<td>α-amylase</td>
<td>Firmicutes (99.00%)</td>
</tr>
<tr>
<td>GT45</td>
<td>α-N-acetylglucosaminyltransferase</td>
<td>Firmicutes (92.96%)</td>
</tr>
<tr>
<td>GT96</td>
<td>peptidyl serine α-galactosyltransferase</td>
<td>Firmicutes (100.00%)</td>
</tr>
</tbody>
</table>

Table 1: The taxonomic assignment at the phylum level of each significantly higher CAZymes in wild group.
Table 2: The taxonomic assignment at the phylum level of each significantly higher CAZymes in AR group

<table>
<thead>
<tr>
<th>CAZymes</th>
<th>Known activities</th>
<th>Corresponding phylum (average proportion)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GH10</td>
<td>endo-1,4-β-xylanase</td>
<td>Bacteroidetes (77.84%)</td>
</tr>
<tr>
<td>GH30</td>
<td>endo-β-1,4-xylanase</td>
<td>Bacteroidetes (76.81%)</td>
</tr>
<tr>
<td>GH35</td>
<td>β-galactosidase</td>
<td>Bacteroidetes (83.99%)</td>
</tr>
<tr>
<td>GH51</td>
<td>endo-β-1,4-xylanase</td>
<td>Bacteroidetes (71.14%)</td>
</tr>
<tr>
<td>GH53</td>
<td>endo-β-1,4-galactanase</td>
<td>Bacteroidetes (85.14%)</td>
</tr>
<tr>
<td>GH55</td>
<td>endo-β-1,3-glucanase</td>
<td>Bacteroidetes (62.47%)</td>
</tr>
<tr>
<td>GH77</td>
<td>amylo maltase or 4-α-glucanotransferase</td>
<td>Firmicutes (48.47%)</td>
</tr>
<tr>
<td>GH82</td>
<td>l-carrageenase</td>
<td>Bacteroidetes (91.56%)</td>
</tr>
<tr>
<td>GH84</td>
<td>N-acetyl β-glucoaminidase</td>
<td>Bacteroidetes (67.04%)</td>
</tr>
<tr>
<td>GH98</td>
<td>endo-β-1,4-xylanase</td>
<td>Bacteroidetes (76.71%)</td>
</tr>
<tr>
<td>GH121</td>
<td>β-L-arabinobiosidase</td>
<td>Bacteroidetes (66.25%)</td>
</tr>
<tr>
<td>GH127</td>
<td>β-L-arabinofuranosidase</td>
<td>Bacteroidetes (72.74%)</td>
</tr>
<tr>
<td>GH130</td>
<td>β-1,4-mannnosylglucose phosphorylase</td>
<td>Bacteroidetes (78.13%)</td>
</tr>
<tr>
<td>PL11</td>
<td>rhamnogalacturonan lyase</td>
<td>Bacteroidetes (79.18%)</td>
</tr>
<tr>
<td>PL12</td>
<td>heparin-sulfate lyase</td>
<td>Bacteroidetes (68.98%)</td>
</tr>
<tr>
<td>CE12</td>
<td>pectin acetyl esterase</td>
<td>Bacteroidetes (87.96%)</td>
</tr>
</tbody>
</table>

Table S1: List of all CAZymes that were found in wild Bar-headed geese gut metagenomic data sets

<table>
<thead>
<tr>
<th>CAZymes</th>
<th>Wild_1</th>
<th>Wild_2</th>
<th>CAZymes</th>
<th>Wild_1</th>
<th>Wild_2</th>
<th>CAZymes</th>
<th>Wild_1</th>
<th>Wild_2</th>
<th>CAZymes</th>
<th>Wild_1</th>
<th>Wild_2</th>
<th>CAZymes</th>
<th>Wild_1</th>
<th>Wild_2</th>
</tr>
</thead>
</table>
It is widely acknowledged that the vertebrate gut microbiome play critical roles in host health and disease (Zhang et al., 2015), which is now attracting increasing attention in the wild birds research (Waite and Taylor, 2014). Metagenome sequencing approaches, which rapidly produce millions of whole genome shotgun sequencing reads that enable the investigation on a culture independent basis, are now popular for exploring microbial community (Jovel et al., 2016). In the present study, for the first time, we outline CAZymes profile of Bar-headed goose metage-nome and we find variations of these CAZymes between wild Bar-headed geese group and artificially reared group.

Our previous studies have demonstrated that the variations in Bar-headed goose gut microbiota diversity...
and structure is mostly due to the rearing conditions (Wang et al., 2016a). Furthermore, in the present study, the different gut microbes contribute to the generation of different CAZymes. In the wild group, the significantly increased CAZymes totally belonged to Firmicutes phyla, while the chief contributors of the significantly increased CAZymes in the AR group are Bacteroidetes. In Firmicutes, degradative capacity is largely restricted to the cell surface and involves elaborate cellulose complexes in specialized cellulosytic species. By contrast, in the Bacteroidetes, utilization of soluble polysaccharides entails outer membrane binding proteins, and degradation is largely periplasmic or intracellular. In general, Bacteroidetes are well reported for their starch, pectin and xylan digestion (Thomas et al., 2011), while Firmicutes for their cellulose and hemicellulose digestion (Flint et al., 2012). Further, previous study has shown presence of higher saccharolytic potential in Bacteroidetes as compared to Firmicutes (El Kaoutari et al., 2013). The AR populations had unrestricted access to fly away to seek natural food and were also fed on artificial diets (blends of 60% corn flour, 20% soybean flour and some vegetables). The members of phylum Bacteroidetes may therefore confer more efficient extraction of energy from both natural and artificial food resources for the artificially reared Bar-headed geese. The microbial communities of the wild Bar-headed geese (Wang et al., 2016a) were mainly dominated by Firmicutes (60.67%), with very low relative abundances of Bacteroidetes (3.33%). Therefore, the digestion of dietary polysaccharides contained in natural diets was mainly conducted by the CAZymes produced by the Firmicutes phyla.

GHs are modular enzymes consisting of different combinations of catalytic modules and helper modules. About 1% of the genome of any organism encodes for GHs (Vocadlo and Davies, 2008). Of the total detected GHs, the proportions of 4 GHs were significantly increased in wild group, with the capacity of digestion cellulose (Flint et al., 2008). 13 GHs were found to be significantly increased in AR group. 4 of them (GH10, 30, 51, 98) are xylanases involved in xylan breakdown. GH35, 53, 55, 77, 84 and PL11 enzymes encompass beta-galactosidases, beta-glucosidases. These enzymes are involved in the breakdown of a large variety of oligosaccharides. Taken together, these increased CAZymes in each group indicate enrichment of gut microbes harboring CAZymes and thereby playing beneficial role in utilization of complex plant polysaccharides.

This data set is not without limitations. First, the number of samples used for sequencing was low. Second, the content of artificial diets as compared to natural diets was not precisely defined. The relationship among different dietary elements, gut microbes and CAZymes should be documented in the future. At this time, however, this is the most comprehensive data set of its type, and provides more knowledge of the Bar-headed geese microbiome and how it may be impacted by dietary intervention.

Conclusions

In conclusion, the present study reveals the global picture of CAZyme profiles of both wild and artificially reared Bar-headed geese. This study can form a basis for further investigations into CAZyme profiles and the gut microbes harboring them, on much larger flocks of Bar-headed geese.

Acknowledgement

This work was supported by the State Forestry Administration Program (no. Y311351B01), the National Key Basic Research Program of China (973 Program, no. 2010CB530301) and China Biodiversity Observation Networks (Sino BON).

REFERENCES


