

Microbial ecology along the buffalo digestive tract provides insights to their functions and interactions with the host

Qingyou Liu (✉ qyliu-gene@gxu.edu.cn)

Guangxi University <https://orcid.org/0000-0003-3265-540X>

Wei-Hua Chen

Huazhong University of Science and Technology

Guohua Hua

Huazhong Agricultural University

Tong Feng

State Key Laboratory for Conservation and Utilization of Subtropical Agro- bioresources, Guangxi University <https://orcid.org/0000-0002-6056-0590>

Teng Wang

Huazhong University of Science and Technology

Na L Gao

Huazhong University of Science and Technology

Ziying Liu

Guangxi University

Kuiqing Cui

Guangxi University

Yiqian Duan

Huazhong University of Science and Technology

Sicheng Wu

Huazhong University of Science and Technology Ezhou Industrial Technology Research Institute

Yuhong Luo

Guangxi University

Zhipeng Li

Guangxi University

Chengjian Yang

Buffalo Research Institute, Chinese Academy of Agricultural Sciences

Yixue Xu

Guangxi University

Bo Lin

Guangxi University

Liguo Yang

Huazhong Agricultural University

Alfredo Pauciullo

Department of Agricultural, Forest and Food Sciences (DISAFA) University of Torino Largo Paolo Braccini

Deshun Shi

Guangxi University

Article

Keywords: buffalo, microbiome, MAGs, rumen, gut, digestive tract

Posted Date: March 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-294016/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Communications on February 10th, 2022. See the published version at <https://doi.org/10.1038/s41467-022-28402-9>.

Abstract

Buffalo is an important livestock in Asia. Like other ruminants, its digestive tract (DT) is the key to the quality and wellbeing of buffalo and heavily interact with microbes. Here, we present a comprehensive survey on the microbial ecology along buffalo's DT, including eight sites in three sections (i.e., stomach, intestine and rectum). We collected 695 samples, performed metagenomic sequencing and obtained 4,960 high-quality metagenome-assembled genomes (MAGs), to which ~ 85% of the raw reads could be mapped. 90.7% of the MAGs are previously unidentified at species level. Overall, Firmicutes and Bacteroidetes are the most abundant phyla; their ratios showed an increasing trend along the DT, consistent with their functions in the DT sections. We identified known interactions between microbes and DT sites including the enrichment of fiber-digesting and methane-producing microbes in the stomach. Strikingly, archaea were highly abundant in both stomach and intestine and showed positive correlations with *Fibrobacter*, indicating their roles in methane-production and fiber-degradation at both sections. We annotated 5,862,748 non-redundant proteins from the MAGs, many of which also showed different abundances and were related to site-specific functions. By comparing with the rumen microbiota of cattle, we found higher abundances of microbes in fiber degradation and but lower in methane production. Our catalog of microbial genomes and encoded-proteins provides insight to their functions and interactions with distinct DT sites, and pave the way to microbial interventions for better buffalo quality.

Main

Buffalo (*Bubalus bubalis*, also known as water buffalo) is a globally important domestic animal of immense worth to humans, providing economic value from milk, meat, leather and draft power ¹. Worldwide population estimates of more than 200 million heads are relied on by more than two billion people—more than any other domesticated animals ². Buffaloes are even-toed, hooved mammals of the Bovidae family, Genus *Bubalus* and tribe Bovini. There are two sub-species of the domesticated water buffalo including swamp buffalo (*Bubalus bubalis carabanesis*, 2 N = 48) and river buffalo (*Bubalus bubalis bubalis*, 2 N = 50) ^{3,4}.

The distribution of swamp buffaloes overlaps closely with rice agriculture in East and Southeast Asian countries (e.g. China, Vietnam, Thailand, etc.), where they have served as the primary draft animals for rice cultivation over thousands of years ⁵. Buffaloes are herbivores, and most of their diet is made up of plant fiber such as forage grass. It can possess a characteristic four-compartment stomach (rumen, reticulum, omasum, and abomasum) that is specialized for transforming the forage with low nutritional value into high-quality animal protein and adapting to the life of forage grass ⁶. Digestibility of organic matter was greatest by buffalo when animals were fed the hay ⁷.

The importance of the microbiota at different sites of the digestive tract is now well recognized and has been linked to the functions and physiology of the corresponding DT sites ⁸. For example, rumen, the first and most important stomach site of ruminants, contains abundant bacteria capable of digesting

cellulose, the main component of plant cell walls, including *Fibrobacter*⁹⁻¹², *Ruminococcus*^{13,14} and *Butyrivibrio*^{15,16}; similarly, *Prevotella*, a group of bacteria capable of degrading non-cellulose plant fibers are also abundantly found in rumen¹⁷⁻²⁰. Archaea are also abundant in the rumen and other sites of the stomach; however, they are the main producers of methane, a major greenhouse gas²¹⁻²³ and thus targets for elimination²⁴. In addition to their roles in food digestion, rumen microbiota has also been linked to more pronounced phenotypes such as milk production and quality of cattle^{25,26}. So far, researches on the microbial ecology along the DT in ruminants mostly focus on rumen²⁷⁻³⁵ with a few studies on gut/feces³⁶, while other sites have been largely overlooked, despite accumulating evidence support their important roles. For example, in human, gut microbiota have been linked to many aspects of human life, including health^{37,38} and diseases³⁹⁻⁴³, development^{44,45}, responses to drugs and treatments⁴⁶⁻⁴⁸.

The application of metagenomic next-generation sequencing (mNGS) has greatly facilitated the reconstruction of large numbers of MAGs for model organisms including cow⁴⁹, sheep³⁶, pig^{50,51}, mouse^{52,53} and chicken^{54,55}, and revealed their associations in health and disease of the hosts^{56,57}. In this study, we set out to fill in the gaps in ruminant research by presenting a comprehensive survey on the microbial ecology along buffalo's DT. We collected in total 695 samples from eight DT sites in three sections, namely stomach, intestine and rectum. We performed metagenomic sequencing and obtained 4,960 high-quality metagenome-assembled genomes (MAGs), which covered in total 85% of the raw sequencing reads. We identified known interactions between microbes and rumen, especially the enrichment of fiber-digesting and methane-producing microbes, but also novel and important interactions at other stomach sites and DT sections. We annotated 5,862,748 non-redundant proteins from the MAGs, many of which also showed different abundances and were related to site-specific functions. Thus, our taxonomic and functional characterization of the microbial ecology along buffalo's DT provides insight to their functions and interactions with distinct DT sites, and overcomes the barriers to manipulating the gastrointestinal microbiota to improve animal productivity for buffalo as well as the related ruminant animals.

Results

Generation and quality assessment of MAGs along the digestive tract of buffalo

To provide a comprehensive collection of microbes associated with the digestive systems of buffalo, we collected in total 695 samples along buffalo's digestive tract (DT) from six provinces of China (Supplementary Fig. 1a), including 211 samples from four distinct sites of the stomach (rumen, reticulum, omasum, abomasum), 85 samples from three sites of the intestine (cecum, colon, and jejunum) and 399 samples from the rectum (feces) (Fig. 1a). The buffalos varied in breeds (three main breeds: river buffalo, swap buffalo, and hybrid buffalo), age (calf, breeding, and adult), and sex (ox and cow) (Supplementary Data 1).

We submitted the 695 samples to metagenomics next-generation sequencing (mNGS) using Illumina NovaSeq 6000 with read-length of 150. After removing vector and low-quality sequences, contaminations from host and food genomic sequences, we obtained in total 11Tb of clean data for further analyses (see Methods for details). On average, we obtained 41,842,231 pairs of clean reads and 6,244,074,222 bases for each sample.

To obtain high-quality metagenome-assembled genomes (MAGs), we adopted a customized bioinformatic analysis workflow (see Supplementary Fig. 1b for a graphical representation; Methods). Briefly, the clean reads were assembled by metaSPAdes⁵⁸ and MEGAHIT⁵⁹; the resulting 109,471,448 contigs were grouped into 58,094 bins using metaBAT2 with default parameters. All bins were aggregated and dereplicated using dRep⁶⁰ (v.2.3.2); followed by CheckM⁶¹ (v.1.0.18) for quality assessment. In the end, we obtained a non-redundant set of 4,960 bins (MAGs) with completeness $\geq 80\%$ and contamination $\leq 10\%$. Rarefaction analysis indicated that the curves could be plateaued using samples from stomach, rectum or all combined, although not for intestines that had relatively fewer samples (Fig. 1b).

Among the resulting MAGs, 2,581(52.0%) were high-quality draft genomes as defined by Bowers *et al*⁶⁴ with $\geq 90\%$ completeness and $\leq 5\%$ contamination, while 2,222(44.8%) met the score criterion defined by Parks *et al*⁶² (completeness – (5 × contamination) ≥ 50) (Fig. 1c, d). The sizes of the MAGs ranged from 402 kilobases (Kb) to 6.1 megabases (Mb), with a median of 2.1 Mb (Fig. 1f). The MAGs contained 1 to 679 contigs with a median of 102. The contig sizes of the MAGs ranged from 2.5Kb to 1.5Mb, with N50 values (50% of assembled bases in contigs larger than the N50 value) ranging from 5.4 kb to 1.4 Mb (Fig. 1e, g). To further access the quality of the resulting MAGs, we annotated tRNA and 16S rRNA genes encoded by these MAGs. All MAGs encoded tRNA genes with on average 18.6 tRNA genes per MAG; in total 2,640 of the MAGs encoded all 20 tRNA genes, indicating the high-quality of the MAGs. However, only 285 MAGs encoded at least one complete 16S rRNA gene, with additional 394 MAGs contained a partial 16S rRNA gene (Supplementary Data 3); these results are consistent with Stewart *et al*^{β1}.

To check if our MAGs could improve the coverage of microbial genomes associated with buffalo's DT, we mapped all the metagenomic clean reads to the resulting MAGs and compared the overall mapping rates to MAGs of selected organisms (human⁶⁵, chicken⁵⁴, pig⁶⁶ and cattle³¹) and reference microbial genomes from public databases (NCBI RefSeq genomes plus the Hungate collection genomes⁶³, referred as to BFAP (bacterial, fungal, archaeal, and protozoan)). As shown in Fig. 1g, the buffalo raw reads showed the highest mapping rate to our MAGs with an average of 85%, followed by the cattle's Rumen Uncultured Genomes (RUGs) from Stewart *et al*^{β1}, BFAP and MAGs of human, pig and chicken.

Our stomach samples showed the lowest mapping rates to the reference genomes and MAGs as compared with samples from intestines and rectum (Fig. 1g), with the cattle dataset being an exception because it consisted of MAGs from rumen metagenomes³¹. These results are consistent with the fact that stomach metagenomes were less covered in public databases and most metagenomic researches

focused on the gut. The overall mapping rates of the stomach samples to cattle and buffalo MAGs were close (Fig. 1g, orange boxes), suggesting their stomach metagenomes were similar; however, the mapping rates of buffalo intestine and rectum samples to cattle RUGs were significantly lower (less than 65%), suggesting gut metagenomes were under-represented in Stewart *et al*'s data³¹. Since the cattle and buffalo are closely related, we also mapped raw sequencing reads of cattle metagenomes obtained from³¹ to our MAGs. On average 71% of the cattle reads could be mapped to our MAGs, as compared to 82.5% to cattle RUGs (Supplementary Fig. 2).

Together, we obtained in total 4,960 high-quality MAGs that significantly improve the coverage of the microbes in buffalo's digestive tract, especially those in sections of intestine and rectum.

Taxonomic Annotation Of Buffalo Mags

We next assigned taxonomic classifications to the MAGs using GTDB-TK⁶⁷. According to GTDB-TK, genomes with < 99% average nucleotide identities (ANIs) belong to the same strains, while those with < 95% ANI belong to the same species. By these criteria, a total of 4,895 MAGs had < 99% ANI and 4,277 MAGs had < 95% ANI with GTDB-TK reference data, indicating potential new strains and species respectively. Of the 4,960 MAGs, almost all could be classified to known taxonomical classifications at the higher levels such as kingdom, phylum, class and order; however, at more refined levels especially the species level, only 460 (9.3%) of the MAGs could be classified as known species (Fig. 2a), indicating most the MAGs were novel (i.e., not present in the GTDB-TK database). Both bacteria and archaea showed similar classification results (Fig. 2b, c).

As shown in Fig. 2, the dominant phyla according to the numbers of MAGs could be assigned were Bacteroidota (n = 2,200) and Firmicutes_A (n = 1,736) (dominated by classes of Bacteroidia and Clostridia), followed by Verrucomicrobiota (n = 191), Firmicutes (n = 172), Spirochaetota (n = 164) and Proteobacteria (n = 147). The dominant orders include Bacteroidales (n = 2,194) and Oscillospirales (n = 1101), while the dominant families include Bacteroidaceae (n = 676), CAG-272 (n = 588), Rikenellaceae (n = 483), and UBA932 (n = 368). At the genus level, the dominant groups include *Alistipes*, *RC9*, and *Prevotella* (Supplementary Fig. 3). 74 archaeal MAGs were identified, belonging to three phyla, Halobacterota (n = 42), Thermoplasmata (n = 20) and Euryarchaeota (n = 12). The top three genera included *Methanocorpusculum* (n = 27), *Methanomicrobium* (n = 13), and *Methanobrevibacter_A* (n = 11), all of them are methanogens.

Taxonomic characteristics of MAGs in different sections along the digestive tract

To investigate the distributions of the 4,960 MAGs in different sections along the DT and their putative interactions with the host, we first determined the coverages and relative abundances of all the MAGs in each sample. We mapped the clean reads of a sample to all MAGs and calculated the coverage of a MAG as the total aligned bases divided by the total bases of the MAG⁶⁸, and relative abundance a MAG as the percentage of reads mapped to the MAG out of the total reads mapped to all MAGs (see Methods for details and Supplementary Table 3 for the results).

Using a coverage 1X as the cutoff of presence/absence, we found 3,032, 3,081 and 4,141 MAGs were present in at least one samples of the three sections, respectively (Fig. 3a). By this criterion, we found 1,692 (34% out 4,960) MAGs were present in all three sections, 1,910 MAGs (38.5%) in samples of two sections, while only small proportions of section-specific MAGs (499 in stomach and 859 in rectum; Fig. 3a). We did not find intestine-specific MAGs by this criteria, likely due to the fact that it positioned between stomach and rectum (Figs. 1a and 3a). At the phylum level, five phyla were stomach-specific, among which three (UBP3, UBP6, and Eremiobacterota) are candidatus, indicating unknown phyla. Among the other two, Euryarchaeota consists of species related to the methane metabolism, while Synergistota consists of oral bacteria in human and was first identified in the goat rumen, usually for the toxic compound degradation⁶⁹. The main rectum-specific phylum is Verrucomicrobiota, which was found in soil, water, and feces⁷⁰.

We further analyzed the relative abundances of the MAGs along the DT and found significantly different alpha diversities among all sections, measured by both the Shannon and Simpson indexes (Fig. 3b). Among which, rectum had the highest diversities, followed by stomach and intestine (Fig. 3b), in part consistent with the numbers of MAGs identified in the three sections (Fig. 3a). We then applied Principal coordinates analysis (PCoA) to show Bray-Curtis distances among the samples, and found that the overall microbial profiles of intestine and rectum were similar ($R = 0.048$; $P = 0.001$, pair-wise non-parametric MANOVA test) while both significantly different from that of the stomach ($R = 0.192, 0.270$, $P = 0.001, 0.001$, pair-wise non-parametric MANOVA test; Fig. 3c). Interestingly, samples from jejunum of the intestine formed their own cluster and were far away from other DT sites (Fig. 3d), indicating its distinctive microbial structure; this is also consistent with the alpha diversity of jejunum that was clearly different from other parts of the DT (Fig. 3b).

Distinctive patterns of MAGs along the digestive tract coincide with their functions

As shown in Fig. 3e, the overall microbial structures of stomach, intestine and rectum were different with each other (Fig. 3e), with Firmicutes and Bacteroidota being the two most abundant phyla. Firmicutes_all, including Firmicutes, Firmicutes_A, Firmicutes_B, and Firmicutes_C according to GTDB, on average accounted for 79%, 85%, and 90% of total microbial abundances in the three DT sections. Interestingly, Bacteroidota showed decreased abundances along the digestive tract (Fig. 4a), while Firmicutes_all showed the opposite (Fig. 4b). Consequently, Firmicutes to Bacteroidota ratio (F/B ratio, Fig. 4c) was lowest in stomach and highest in rectum. Previous results have linked increased F/B ratio with increased capacity for energy harvest from the diet^{71,72}, consistent with the physiological roles of the three sections. In addition, the F/B ratio in the rumen has been linked to milk fat yield in cows⁷³⁻⁷⁵.

The decreasing abundances of Bacteroidota along DT was in part due to *Prevotella*, the main genus of Bacteroidota that is mostly abundant in stomach, especially in rumen, and is significantly lower in other sections and DT sites (Fig. 4h), accounting for 33.1% total microbial abundances in rumen. *Prevotella* species are associated with non-cellulose plant fiber degradation, and are known the largest single bacterial group reported in the rumen of cattle and sheep under most dietary regimes⁷⁶. Most of the

species in this genus are unclassified, including the most abundant one (Supplementary Fig. 4). In addition to the role in plant degradation, *Prevotella* species also played important role to prevent ruminal acidosis⁷⁶⁻⁷⁸.

Similarly, bacteria capable of digesting cellulose, the main component of the cell wall of plants, were also significantly abundant in the stomach^{79,80}, and were less abundant in other DT sites (Fig. 4e, f, g). These included *Fibrobacter_all* (including *Fibrobacter* and *Fibrobacter_A* according to GTDB; Fig. 4e), *Ruminococcus_all* (including *Ruminococcus_E*, *Ruminococcus_A*, and *Ruminococcus* according to GTDB; Fig. 4f) and *Butyrivibrio_all* (including *Butyrivibrio_A* and *Butyrivibrio* according to GTDB; Fig. 4g). Among them, *Fibrobacter_all* was the most abundant taxon. Our results showed that the total abundances of *Fibrobacter_all* were significantly higher in all four stomach sites (i.e., rumen, reticulum, omasum and abomasum) than the other two sections; however, we found the highest abundances of *Fibrobacter_all* in omasum, indicated that the later may play important roles in cellulolytic digestion.

All the archaeal species we identified were methanogens; they were highly abundant in the stomach and intestine (Fig. 4d). These results contradicted our current understandings that only stomach especially rumen is the main organ of methane metabolism⁸¹, and highlight the importance of intestine in the methane metabolism. Strikingly, close examination revealed that the total abundances of these methanogens peaked in omasum (Fig. 4d), coincided with *Fibrobacter_all* (Fig. 4e); in fact, we found that the overall abundances of stomach and intestine had the highest correlations as compared with other categories (Fig. 4i). Together, these results highlighted the important roles of omasum in both methane metabolism and cellulose degradation, and possible functional link between the latter two⁸²

Functional characteristics of MAGs in different sections along the digestive tract

We next explored the proteomic contents of buffalo metagenome and their putative functions. We predicted in total 9,470,238 proteins from the 4,960 MAGs; after clustering by CD-HIT⁸³, we obtained a non-redundant proteome dataset of 5,862,748 proteins. We annotated these proteins by comparing the amino acid sequences with the eggNOG database⁸⁴ using eggNOG-mapper⁸⁵ and the CAZy⁸⁶ database using dbCAN2⁸⁷. As a result, 4,787,680 proteins (81.7% out of total) could be annotated according to one or both methods; among which, 114,989 and 4,672,691 were annotated by CAZy and eggNOG, respectively.

Only a small proportion of the proteins (114,989; 1.96% out of total) were annotated as enzymes according to CAZy (Fig. 5a). Among which, majority were involved in glycoside hydrolases (GH, 76,224, 66.29% out of 114,989), followed by glycosyl transferases (GT, 21,402, 18.61%) and carbohydrate-binding modules (CBM, 11,645, 10.13%); the remaining three categories contained much less proteins, including carbohydrate esterases (CE, 3,804, 3.31%), polysaccharide lyases (PL, 1,879, 1.63%) and predicted auxiliary activities (AA, 35; 0.03%). Overall, the annotated proteins showed decent sequence identities with those in the CAZy database with an average of 59.11% (Supplementary Fig. 6); however, only a small

proportion (3.8% out of the total annotated proteins) showed identities higher than 90%, indicating poor representation of our MAGs in public databases.

We then calculated the relative abundances of all proteins in each sample and compared their distributions (see Methods). As shown in Fig. 5b-g, all the six above-mentioned CAZy families showed significant differences in terms of relative abundances (the sum abundance of all proteins is 100%) among the three DT sections. Interestingly, all CAZy protein families except Auxiliary Activity (AA) showed the highest abundances in stomach, followed by rectum and then intestine, supporting the central role of stomach in food digestion and processing (Fig. 5b-g); rumen often had the highest abundances of these CAZy families among all DT sites (except for carbohydrate-binding module (CBM), omasum displayed the highest abundances), although other stomach sites also contained significantly higher CAZy proteins than other DT sections (Fig. 5b-g). Again, we identified jejunum as an outlier which contained the lowest percentages of all six CAZy families (Fig. 5b-g), likely due to its low microbial diversity (Fig. 3b).

Comparisons Of Rumen Microbiota Between Buffalo And Cattle

Rumen often is considered the most important stomach site in ruminants. As the first section of DT, rumen is the largest compared with other stomach sites, governs the first steps of feedstuffs degradation and is the main site for methane-production²⁶. The microbiota in rumen played important roles in its function and were the first to be studied by researchers^{26,88}. We thus compared the taxonomic and functional profiles of buffalo's rumen microbiota with that of cattle which was recently made available by *Stewart et al*³¹. To make a fair comparison, we obtained in total 4,941 Rumen Uncultured Genomes (RUGs, similar to MAGs; see ref³¹), 4,879,163 non-redundant proteins, and calculated their abundances in each sample using the same methods used in this study (Methods).

As shown in Fig. 6a, we found that buffalo and cattle differed significantly in the two most dominant phyla (Firmicutes_all and Bacteroidota) and consequently the F/B ratios (Fig. 6a). Bacteroidota species, especially those in its dominant genus *Prevotella* are capable of degrading non-cellulose plant fibers. Therefore, the higher abundances of Bacteroidota and *Prevotella* in buffalo rumen than cattle, as well as the similar levels of *Fibrobacter_all* (responsible for cellulolytic plant fiber digestion) between the two (Fig. 6a), suggested that buffalo was more adapted to coarse forage than cattle⁸¹. Conversely, we found significantly higher levels of Archaea, *Butyrivibrio_all* and *Ruminococcus_all* in cattle, all of which played important role in methanogenesis through biohydrogenation and glycolysis (Fig. 6b), suggesting that buffalo may produce significantly less methane than cattle.

We also compared the abundances of rumen protein families (i.e., CAZy) between buffalo and cattle. Surprisingly, all six families were significant abundant in buffalo (Fig. 6c). This result further suggested that buffalo had better capacity in carbohydrate metabolism. Besides, the comparison based on all CAZy proteins (Supplementary Fig. 7) displayed significant abundance of GH and CBM in cattle, indicating that rumen of cattle had more proportion of enzymes that relative to the formation of glycosidic bonds.

Discussion

As an important livestock, buffalo provides humans with milk, meat, leather and draft power. Like other ruminants, its digestive tract (DT) is the key to the quality and wellbeing of buffalo and heavily interact with microbes. However, the lack of microbial reference genomes at different sites of the DT greatly hindered our understanding of the functional interactions between DT sites and their microbial ecology, and our ability to modulate buffalo's physiology and economically important phenotypes through DT microbiota. More importantly, a comprehensive profiling on the methanogenic microbes that are presumably inhabit in rumen may provide us with insights on reducing emission of methane, an important source of greenhouse gases. Recent studies on ruminant microbiota have most focused on rumen, while the exploring for the digestive tracts was still missing. To fill in these gaps, we presented a comprehensive survey on the microbial ecology along buffalo's DT. We collected in total 695 samples from eight DT sites in three sections, namely stomach, intestine and rectum. To further increase the representativeness of our study, we took samples from six different locations (Guangxi, Henan, Anhui, Yunnan, Hainan, Hubei; Supplementary Tables 1, 4; Supplementary Fig. 1a), three breeds (river, swamp and hybrid; Supplementary Tables 1, 2), both sexes and two developmental stages (Supplementary Tables 1, 5).

We performed metagenomic next generation sequencing (mNGS) on these samples and obtained 4,960 high-quality metagenome-assembled genomes (MAGs). These MAGs greatly improved the coverage of the raw sequencing reads from 64% of the public databases (the combination of reference genomes datasets including bacterial, fungal, archaeal and protozoan genomes from NCBI RefSeq, BAFP) to 85%. Taxonomic annotation revealed that all MAGs could be classified into known phyla; however, more than 90% of the MAGs are novel at species level. Thus, our dataset represents a great expansion of buffalo microbiomes.

Sampling at different sites of the DT allowed us to better understand the functional associations between the microbial ecology and the DT sites. For example, we found that Firmicutes and Bacteroidota, the two most dominant phyla, showed distinct abundance patterns along the DT: Firmicutes were increased along the digestive tract (Fig. 4a), while Bacteroidota showed the opposite (Fig. 4b). Consequently, Firmicutes to Bacteroidota ratio (F/B ratio, Fig. 4c) was lowest in stomach and highest in rectum. F/B ratio has been shown to be related to energy harvesting^{71,72}; its trend thus coincides the physiological transition from food digestion (stomach) to energy harvesting (intestine) along the DT. Our data also allowed us to validate known interactions between microbes and DT sites, including the enrichment of fiber-digesting and methane-producing microbes in the stomach. Surprisingly, we found that *Fibrobacter*, a group of cellulolytic bacteria known to colonize mainly in rumen showed higher relative abundance in omasum; its distributions coincided with archaea, the main methane-producers whose abundances also peaked in omasum. Archaea were also highly abundant in both stomach and intestine and showed positive-correlations with *Fibrobacter*, indicating their roles in methane-production at both sections. Based on the abundance of the microbes at phylum and genus level, we identified several features that

abundant in stomach, intestine, and rectum. These results highlighted the importance of having samples from all DT sites, especially those other than rumen.

We also evaluated the functional capacities of the microbial ecology of the MAGs at the DT sites by annotating the protein-coding genes from the MAGs and comparing them against CAZy and eggNOG databases. We found all of the six CAZy families showed significant differences among DT sections, suggesting their different roles associated with distinct sections of the DT. Our data showed poor sequence identity with public data, indicating previously unidentified protein sequences and thus novel functions encoded in the DT-associated MAGs.

Rumen is the most important section in the DT; its microbiota has been extensively explored recently in cattle. We thus took the opportunity to compare the rumen microbes between the two closely related model ruminants. Our results showed the significant differences between buffalo and cattle. For example, we found higher abundances of microbes with fiber degradation capacity in buffalo's rumen than cattle. In addition, the relative abundance of methane-producing archaeal species in buffalo was significantly lower than cattle, indicating the less production of methane and the more fully use of energy from feedstuff.

Together, our catalogs of microbial genomes and their encoded-proteins represented the largest effort so far to characterize the microbial ecology along all major sections of the DT; our study provided insight to the microbial functions and interactions with distinct DT sites, and valuable resources for the community who are interested in microbial interventions for better buffalo quality.

Methods

Sample collection

A total of 695 samples were collected for metagenome sequencing. To ensure the diversity of the samples, they were collected from three breeds (153, 430, 112 samples from river, swamp and hybrid buffaloes), six regions (Guangxi, Henan, Anhui, Yunnan, Hainan, Hubei from China), two sexes (females and males respectively) and two developmental stages (Adult buffaloes and calves; see Supplementary Data 1 for details). These samples included 296 content and 399 fecal samples; the content samples were taken from rumen, reticulum, omasum, abomasum, jejunum, cecum and colon, while the fecal samples taken from rectum (Supplementary Table 3). In this study, the above eight DT sites were divided into three sections, namely stomach (rumen, reticulum, omasum, and abomasum), intestine (jejunum, cecum, and colon) and rectum; see also Fig. 1. All samples were immediately frozen after collection in liquid nitrogen and stored at -80°C until DNA extraction.

DNA Extraction, Library Construction And Metagenomics Sequencing

Three grams of each sample were taken for DNA extraction. DNA was extracted by a bead-beating method using a mini-bead beater (Biospec Products; Bartlesville, USA), followed by phenol-chloroform extraction. The solution was precipitated with ethanol, and the pellets were suspended in 50 μL of Tris-

EDTA buffer. DNA was quantified using a NanoPhotometer® (IMPLEN, CA, USA) following staining using a Qubit® 2.0 Fluorometer (Life Technologies, CA, USA). DNA samples were stored at – 80°C until further processing.

Library preparation was performed according to the TruSeq DNA Sample Preparation Guide (Illumina, 15026486 Rev. C) method and procedure using 500 ng DNA as template. Qualified libraries were selected and subjected to the Illumina NovaSeq 6000 for pair-ended sequencing with read-length of 150 base pairs (PE150).

Quality Control And Removal Host- And Food-associated Genomes

Raw sequencing reads were first trimmed by Trimmomatic⁸⁹ (v.0.35) to remove vectors and low-quality bases. Sequences longer than 110 bases and average base quality larger than 30 after trimming were retained for further analysis. To remove possible reads from the host and food, genomic sequences of buffalo⁹⁰, *Glycine max*⁹¹, *Zea Mays*⁹², and *Medicago truncatula*⁹³ were downloaded from NCBI and used as references by bowtie2⁹⁴ (v.2.3.3) with the options ‘-p 10 –very-sensitive’. Reads that aligned concordantly to references were removed as contamination. As a result, average 20.3% bases were removed. The remaining “clean reads” were used for further analysis.

Generation And Quality Assessment Of Metagenomics-assembled Genomes (mags)

MEGAHIT (v.1.2.8) and metaSPAdes (v.3.13.0) were used for single-sample assemblies. Before each run of metaSPAdes, the k-mer parameter was tested with a range of 21 to 141; then the k-mer with longest N50 and total reads length was chosen (if the result of N50 and total length were inconsistent, the k-mer with longest total length was chosen). For MEGAHIT, we used default parameters to assemble the reads. Co-assemblies were performed for each the three sections by combining all samples as input only using MEGAHIT because its less consumption of time and memory compared with metaSPAdes. However, As shown in Supplementary Fig. 8, metaSPAdes could generate contigs with longer N50 than MEGAHIT through comparison with N50, among other measurements for assembly qualities. Thus, we combined the assembly results from both tools to increase the coverage and quality of the resulting contigs.

MetaBAT2⁹⁵ (2.12.1) was used to group contigs into bins. First, BWA-MEM⁹⁶ (v.0.7.15) was used to map reads to the contigs (MEGAHIT) and scaffolds (metaSPAdes) to get the depths of contigs (MEGAHIT) or scaffolds (metaSPAdes) in each sample. The results were saved in SAM files. Second, Samtools⁹⁷ (v.1.8) was used to convert SAM files to BAM format. Last, MetaBAT2 was used to calculate coverage from the resulting BAM files and output the results of bins. As a result, Single-sample binning produced a total of 58,041 bins, and while additional 53 bins were obtained from co-assembly binning. We referred these bins to as metagenome-assembled genomes (MAGs).

All bins were dereplicated using dRep⁶⁰ (v.2.3.2) with the option ‘dereplicate_wf -p 16 -comp 80 -con 10 -str 100 -strW 0’. During this process, CheckM⁶¹ (v.1.0.18) was first used to assess the quality of the resulting MAGs. After removing MAGs with completeness < 80% or contamination > 10%, the remaining

high-quality MAGs were processed by two clustering steps to remove replicates with default parameters. The first is a rapid primary algorithm (Mash, ANI = 0.9), and the second is a more sensitive algorithm (ANI = 0.99). After that, we removed MAGs which size larger than 10Mb. At the end, 4960 non-redundant MAGs were obtained.

To calculate the coverage of each MAG in each sample, clean reads of each sample were mapped to the 4,960 MAGs using BWA-MEM with default parameters. After converting the resulted SAM files to BAM format using Samtools⁹⁷ (v.1.8), BEDTools⁶⁸ (v.2.27.1) was used to calculate the coverage of MAGs, which defined as the total bases mapped to a MAG in a sample divided by its length.

Comparisons with reference microbial genomes and MAGs associated with model organisms

To check if our MAGs could improve the coverage of microbial genomes associated with buffalo's DT, MAGs of model organisms including human⁶⁵, chicken⁵⁴, pig⁶⁶ and cattle³¹ were downloaded from their respective sources. In addition, a BFAP dataset was created to include reference microbial genomes (bacterial, fungal, archaeal, and protozoan) from the NCBI RefSeq genome database, and the Hungate collection genomes⁶³.

BWA-MEM was used to map the "clean reads" to the above datasets and our MAGs as references. A mapping rate was calculated for each sample as the percentage of clean reads mapped to each of the reference datasets.

Taxonomic Assignments Of Buffalo Mags

Taxonomic assignments of the 4,960 MAGs were performed using the GTDB-TK tool⁶⁷ with its "classify_wf" workflow. The results were visualized in GraPhlAn⁹⁸ as a phylogenetic tree.

Annotation And Functional Characterization Of Mag-encoded Proteins

MAG-encoded proteins were predicted using Prodigal⁹⁹ (v2.6.3). CD-HIT⁸³ was used to cluster predicted proteins with the option '-c 0.95 -n 10 -d 0 -M 16000 -T 8'. The resulting non-redundant proteins were searched against the CAZy database using dbCAN2⁸⁷ and the EggNOG database using eggNOG-mapper (v.4.5)⁸⁴.

Estimation of relative abundances of MAGs and proteins in each sample

To enable the identification of differentially abundant MAGs and proteins in different sites/sections along DT, the relative abundance of MAGs and proteins were calculated based on their respective coverage in each sample. The coverage of each protein was calculated as same as the coverage of MAGs already described in the above, and relative abundance of a MAG (gene) in each sample was calculated as the percentage of the coverage of MAG (protein) out of the sum of coverages of all MAGs (proteins).

Statistics

All processed data, if not otherwise stated, were loaded into R (v.3.6.3, <https://www.r-project.org/>), analyzed and/or visualized.

Data Availability

The raw sequencing data were submitted to the NCBI SRA database under the accession ID PRJNA656389; the sequences and annotations of the 4,960 MAGs are available at ENA under the accession ID ERZ1741894. These data will be public after the manuscript is published.

Declarations

Acknowledgments

We acknowledge Prof. Jiaxiang Huang, Dr. Guangsheng Qin and Hui Li for providing partial buffalo samples. We also acknowledge Prof. Jianghua Shang, Asisitant Prof. Chunyan Yang, Dr. Jieping Huang, Dr. Hui Li and Miss Siwen Wu for manuscript polish and editorial assistance.

Funding

This research is supported by the National Natural Science Fund (U20A2051, 31760648 and 31860638), and Guangxi Natural Science Foundation (AB18221120), and Guangxi Distinguished scholars Program (201835).

Author information

Feng Tong, Teng Wang, and Na L. Gao are co-first authors.

Affiliations

State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, Guangxi University, Nanning 530005, China

Feng Tong, Ziying Liu, Kuiqing Cui, Yuhong Luo, Zhipeng Li, Yixue Xu, Bo Lin, Deshun Shi & Qingyou Liu

Key Laboratory of Molecular Biophysics of the Ministry of Education, Hubei Key Laboratory of Bioinformatics and Molecular-imaging, Center for Artificial Biology, Department of Bioinformatics and Systems Biology, College of Life Science and Technology, Huazhong University of Science and Technology, 430074 Wuhan, Hubei, China

Teng Wang, Na L. Gao, Yiqian Duan, Sicheng Wu & Wei-Hua Chen.

Key Lab of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education, College of Animal Science and Technology, Huazhong Agricultural University, Wuhan, Hubei Province, China; International Joint Research Centre for Animal Genetics, Breeding and Reproduction, Wuhan 430070, China

Liguo Yang & Guohua Hua

Buffalo Research Institute, Chinese Academy of Agricultural Sciences, 24-1Yongwu Road, Nanning City530001, P. R. China

Chengjian Yang

College of Life Science, Henan Normal University, 453007 Xinxiang, Henan, China

Wei-Hua Chen

Department of Agricultural, Forest and Food Sciences (DISAFA) University of Torino Largo Paolo Braccini, 210095 Grugliasco (TO) – Italy

Alfredo Pauciullo

Contributions

QL, WHC, and GH designed and directed the research; FT, ZL, KC, YL, ZL, YX, BL, DS, and QL participated in sampling and background investigation of buffalo; TW, NLG, and FT analyzed data and wrote the paper with results from all authors. All authors read and approved the final manuscript.

Corresponding authors

Correspondence to [Qingyou Liu](#), [Wei-Hua Chen](#) or [Guohua Hua](#).

Ethics declarations

Ethics approval

The investigation was approved by the Experimental Animal Ethics Committee, College of Animal Science and Technology, Guangxi University, under reference number Gxu-2021-010.

Competing interests

The authors declare that they have no competing interests.

References

- 1 Cockrill, W. R., Fao, R. & AGA. The husbandry and health of the domestic buffalo. *Tropical Animal Health & Production* **7** (1975).
- 2 Scherf, B. & Scherf, B. World watch list for domestic animal diversity. *World Watch List for Domestic Animal Diversity* (2000).

- 3 Tanaka, K., Solis, C. D., Masangkay, J. S., Maeda, K. I. & Namikawa, T. Phylogenetic relationship among all living species of the genus *Bubalus* based on DNA sequences of the cytochrome b gene. *Biochemical Genetics* **34**, 443-452 (1996).
- 4 Kumar, S. *et al.* Mitochondrial DNA analyses of Indian water buffalo support a distinct genetic origin of river and swamp buffalo. *Animal Genetics* (2007).
- 5 Cockrill, W. R., Fao, R., Campaign, S. & Aga. The buffaloes of China. *XF2006131266* (1978).
- 6 Lei, Y., Zhang, K., Guo, M., Li, G. & Wang, X. Exploring the Spatial-Temporal Microbiota of Compound Stomachs in a Pre-weaned Goat Model. *Frontiers in Microbiology* **9** (2018).
- 7 Batista, H. A. M., Autrey, K. M. & Von Tiesenhausen, I. M. E. V. Comparative In Vitro Digestibility of Forages by Buffalo, Zebu, and Holstein Cattle. *Journal of Dairy Science* **65**, 746-748 (1982).
- 8 Sommer, F. & Bäckhed, F. The gut microbiota – masters of host development and physiology. *Nature Reviews Microbiology*.
- 9 Burnet, M. C. *et al.* Evaluating Models of Cellulose Degradation by *Fibrobacter succinogenes* S85. *PLoS One* **10**, e0143809, doi:10.1371/journal.pone.0143809 (2015).
- 10 Raut, M. P., Couto, N., Karunakaran, E., Biggs, C. A. & Wright, P. C. Deciphering the unique cellulose degradation mechanism of the ruminal bacterium *Fibrobacter succinogenes* S85. *Sci Rep* **9**, 16542, doi:10.1038/s41598-019-52675-8 (2019).
- 11 Neumann, A. P. & Suen, G. The Phylogenomic Diversity of Herbivore-Associated *Fibrobacter* spp. Is Correlated to Lignocellulose-Degrading Potential. *mSphere* **3**, doi:10.1128/mSphere.00593-18 (2018).
- 12 Palevich, N. *et al.* Comparative Genomics of Rumen *Butyrivibrio* spp. Uncovers a Continuum of Polysaccharide-Degrading Capabilities. *Applied and Environmental Microbiology* **86**, e01993-01919, doi:10.1128/aem.01993-19 (2019).
- 13 Anju, K. *et al.* Impact of levels of total digestible nutrients on microbiome, enzyme profile and degradation of feeds in buffalo rumen. *Plos One* **12**, e0172051 (2017).
- 14 Hongbin *et al.* Microbial and metabolic alterations in gut microbiota of sows during pregnancy and lactation. *Faseb Journal Official Publication of the Federation of American Societies for Experimental Biology* (2019).
- 15 Derakhshani, H. *et al.* Linking Periparturient Dynamics of Ruminal Microbiota to Dietary Changes and Production Parameters. *Frontiers in Microbiology* **7**, doi:10.3389/fmicb.2016.02143 (2017).
- 16 Vargas, J. E. *et al.* Effect of Sunflower and Marine Oils on Ruminal Microbiota, In vitro Fermentation and Digesta Fatty Acid Profile. *Frontiers in Microbiology* **8**, doi:10.3389/fmicb.2017.01124

(2017).

- 17 Chen, T. *et al.* Fiber-utilizing capacity varies in Prevotella- versus Bacteroides-dominated gut microbiota. *Scientific Reports* **7**, 2594, doi:10.1038/s41598-017-02995-4 (2017).
- 18 Stanislawski, M. A., Dabelea, D., Lange, L. A., Wagner, B. D. & Lozupone, C. A. Gut microbiota phenotypes of obesity. *npj Biofilms and Microbiomes* **5**, 18, doi:10.1038/s41522-019-0091-8 (2019).
- 19 De Vadder, F. *et al.* Microbiota-Produced Succinate Improves Glucose Homeostasis via Intestinal Gluconeogenesis. *Cell Metabolism* **24**, 151-157, doi:<https://doi.org/10.1016/j.cmet.2016.06.013> (2016).
- 20 Rubino, F. *et al.* Divergent functional isoforms drive niche specialisation for nutrient acquisition and use in rumen microbiome. *The ISME Journal* **11**, 932-944, doi:10.1038/ismej.2016.172 (2017).
- 21 Zhang, Q. *et al.* Bayesian modeling reveals host genetics associated with rumen microbiota jointly influence methane emission in dairy cows. *The ISME Journal* **14** (2020).
- 22 Kamke, J. *et al.* Rumen metagenome and metatranscriptome analyses of low methane yield sheep reveals a Sharpea -enriched microbiome characterised by lactic acid formation and utilisation. *Microbiome* **4**, 1-16 (2016).
- 23 Borrel, G., Brugère, J.-F., Gribaldo, S., Schmitz, R. A. & Moissl-Eichinger, C. The host-associated archaeome. *Nature Reviews Microbiology* **18**, 622-636, doi:10.1038/s41579-020-0407-y (2020).
- 24 Misiukiewicz, A. *et al.* Review: Methanogens and methane production in the digestive systems of nonruminant farm animals. *Animal* **15**, 100060, doi:10.1016/j.animal.2020.100060 (2020).
- 25 Zhang, J., Xu, C., Huo, D., Hu, Q. & Peng, Q. Comparative study of the gut microbiome potentially related to milk protein in Murrah buffaloes (*Bubalus bubalis*) and Chinese Holstein cattle. *Sci Rep* **7**, 42189, doi:10.1038/srep42189 (2017).
- 26 Matthews, C. *et al.* The rumen microbiome: a crucial consideration when optimising milk and meat production and nitrogen utilisation efficiency. *Gut Microbes* **10**, 115-132, doi:10.1080/19490976.2018.1505176 (2019).
- 27 Morais, S. & Mizrahi, I. The Road Not Taken: The Rumen Microbiome, Functional Groups, and Community States. *Trends in Microbiology* **27**, 538-549, doi:<https://doi.org/10.1016/j.tim.2018.12.011> (2019).
- 28 Stewart, R. D. *et al.* Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nature Communications* **9**, 870, doi:10.1038/s41467-018-03317-6 (2018).
- 29 Mingyuan *et al.* Assessment of rumen microbiota from a large cattle cohort reveals the pan and core bacteriome contributing to varied phenotypes. *Applied and environmental microbiology* (2018).

- 30 Franco-Lopez, J., Duplessis, M., Bui, A., Reymond, C. & Ronholm, J. Correlations between the Composition of the Bovine Microbiota and Vitamin B 12 Abundance. *mSystems* **5** (2020).
- 31 Stewart, R. D. *et al.* Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nat Biotechnol* **37**, 953-961, doi:10.1038/s41587-019-0202-3 (2019).
- 32 Zebeli, Q., Ghareeb, K., Humer, E., Metzler-Zebeli, B. U. & Besenfelder, U. Nutrition, rumen health and inflammation in the transition period and their role on overall health and fertility in dairy cows. *Res Vet Sci* **103**, 126-136, doi:10.1016/j.rvsc.2015.09.020 (2015).
- 33 Stuart *et al.* Metagenomic analysis of the rumen microbial community following inhibition of methane formation by a halogenated methane analog. *Frontiers in Microbiology* (2015).
- 34 Cremonesi, P. *et al.* Evaluation of the effects of different diets on microbiome diversity and fatty acid composition of rumen liquor in dairy goat. *Animal*, 1-11 (2018).
- 35 Kyung-Tai *et al.* Metagenomic mining and functional characterization of a novel KG51 bifunctional cellulase/hemicellulase from black goat rumen. *Journal of Agricultural & Food Chemistry* (2018).
- 36 Peng, X. *et al.* Genomic and functional analyses of fungal and bacterial consortia that enable lignocellulose breakdown in goat gut microbiomes. *Nat Microbiol*, doi:10.1038/s41564-020-00861-0 (2021).
- 37 Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174-180, doi:10.1038/nature09944 (2011).
- 38 Wu, S. *et al.* GMrepo: a database of curated and consistently annotated human gut metagenomes. *Nucleic Acids Res* **48**, D545-D553, doi:10.1093/nar/gkz764 (2020).
- 39 Pedersen, H. K. *et al.* Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature* **535**, 376-381, doi:10.1038/nature18646 (2016).
- 40 Qin, J. *et al.* A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55-60, doi:10.1038/nature11450 (2012).
- 41 Li, J. *et al.* Gut microbiota dysbiosis contributes to the development of hypertension. *Microbiome* **5**, 14, doi:10.1186/s40168-016-0222-x (2017).
- 42 Wirbel, J. *et al.* Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat Med* **25**, 679-689, doi:10.1038/s41591-019-0406-6 (2019).

- 43 Dai, D., Wang, T., Wu, S., Gao, N. L. & Chen, W. H. Metabolic Dependencies Underlie Interaction Patterns of Gut Microbiota During Enteropathogenesis. *Frontiers in microbiology* **10**, 1205, doi:10.3389/fmicb.2019.01205 (2019).
- 44 Stewart, C. J. *et al.* Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* **562**, 583-588, doi:10.1038/s41586-018-0617-x (2018).
- 45 Pronovost, G. N. & Hsiao, E. Y. Perinatal Interactions between the Microbiome, Immunity, and Neurodevelopment. *Immunity* **50**, 18-36, doi:10.1016/j.immuni.2018.11.016 (2019).
- 46 Forslund, K. *et al.* Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* **528**, 262-266, doi:10.1038/nature15766 (2015).
- 47 Routy, B. *et al.* Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science* **359**, 91-97, doi:10.1126/science.aan3706 (2018).
- 48 Jiang, P., Lai, S., Wu, S., Zhao, X.-M. & Chen, W.-H. Host DNA contents in fecal metagenomics as a biomarker for intestinal diseases and effective treatment. *BMC Genomics* **21**, 348, doi:10.1186/s12864-020-6749-z (2020).
- 49 Stewart, R. D. *et al.* Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nature Communications* **9** (2018).
- 50 Xiao, L. *et al.* A reference gene catalogue of the pig gut microbiome. *Nature Microbiology* **1**, 16161 (2016).
- 51 Chen, C. *et al.* Expanded catalog of microbial genes and metagenome-assembled genomes from the pig gut microbiome. *Nature communications* **12**, 1106, doi:10.1038/s41467-021-21295-0 (2021).
- 52 Xiao, L. *et al.* A catalog of the mouse gut metagenome. *Nat Biotechnol* **33**, 1103-1108, doi:10.1038/nbt.3353 (2015).
- 53 Lavrinienko, A., Tukalenko, E., Mousseau, T. A., Thompson, L. R. & Watts, P. C. Two hundred and fifty-four metagenome-assembled bacterial genomes from the bank vole gut microbiota. *entific Data* **7** (2020).
- 54 Huang, P. *et al.* The chicken gut metagenome and the modulatory effects of plant-derived benzylisoquinoline alkaloids. *Microbiome* **6**, 211, doi:10.1186/s40168-018-0590-5 (2018).
- 55 Maki, J. J., Bobeck, E. A., Sylte, M. J. & Looft, T. Eggshell and environmental bacteria contribute to the intestinal microbiota of growing chickens. *J Anim Sci Biotechnol* **11**, 60, doi:10.1186/s40104-020-00459-w (2020).

- 56 Jian, X., Zhu, Y., Ouyang, J., Lei, Q. & Zhou, W. Alterations of Gut Microbiome Accelerate Multiple Myeloma Progression By Increasing the Relative Abundances of Nitrogen Recycling Bacteria. *Blood* **134**, 688-688 (2019).
- 57 Ridaura, V. K. *et al.* Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* **341**, 1241214, doi:10.1126/science.1241214 (2013).
- 58 Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* **27**, 824-834, doi:10.1101/gr.213959.116 (2017).
- 59 Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674-1676, doi:10.1093/bioinformatics/btv033 (2015).
- 60 Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J* **11**, 2864-2868, doi:10.1038/ismej.2017.126 (2017).
- 61 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**, 1043-1055, doi:10.1101/gr.186072.114 (2015).
- 62 Parks, D. H. *et al.* Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life (vol 2, pg 1533, 2017). *Nature Microbiology* **3**, 253-253, doi:10.1038/s41564-017-0083-5 (2018).
- 63 Seshadri, R. *et al.* Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nat Biotechnol* **36**, 359-367, doi:10.1038/nbt.4110 (2018).
- 64 Bowers, R. M. *et al.* Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea (vol 35, pg 725, 2017). *Nature Biotechnology* **36**, 660-660, doi:DOI 10.1038/nbt0718-660a (2018).
- 65 Almeida, A. *et al.* A unified sequence catalogue of over 280,000 genomes obtained from the human gut microbiome. (2019).
- 66 Xiao, L. *et al.* A reference gene catalogue of the pig gut microbiome. *Nat Microbiol* **1**, 16161, doi:10.1038/nmicrobiol.2016.161 (2016).
- 67 Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*, doi:10.1093/bioinformatics/btz848 (2019).
- 68 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).

- 69 Allison, M. J., Mayberry, W. R., Mcsweeney, C. S. & Stahl, D. A. *Synergistes jonesii*, gen. nov., sp.nov.: A Rumen Bacterium That Degrades Toxic Pyridinediols. *System.appl.microbiol* **15**, 522-529 (1992).
- 70 Chiang, E. *et al.* Verrucomicrobia are prevalent in north-temperate freshwater lakes and display class-level preferences between lake habitats (vol 13, e0195112, 2018). *Plos One* **13**, doi:ARTN e0206396 10.1371/journal.pone.0206396 (2018).
- 71 Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027-1031, doi:10.1038/nature05414 (2006).
- 72 Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444**, 1022-1023, doi:10.1038/4441022a (2006).
- 73 Jami, E., White, B. A. & Mizrahi, I. Potential role of the bovine rumen microbiome in modulating milk composition and feed efficiency. *Plos One* **9**, e85423, doi:10.1371/journal.pone.0085423 (2014).
- 74 Jewell, K. A., McCormick, C. A., Odt, C. L., Weimer, P. J. & Suen, G. Ruminant Bacterial Community Composition in Dairy Cows Is Dynamic over the Course of Two Lactations and Correlates with Feed Efficiency. *Appl Environ Microb* **81**, 4697-4710, doi:10.1128/Aem.00720-15 (2015).
- 75 Lima, F. S., Oikonomou, G., Lima, S. F., Bicalho, M. L. S. & Bicalho, R. C. Prepartum and Postpartum Rumen Fluid Microbiomes: Characterization and Correlation with Production Traits in Dairy Cows. *Appl Environ Microb* **81**, 1327-1337 (2014).
- 76 Flint, H. J. & Stewart, C. S. in *Encyclopedia of Food Microbiology* (ed Richard K. Robinson) 198-203 (Elsevier, 1999).
- 77 Bandarupalli, V. V. K. & St-Pierre, B. Identification of a Candidate Starch Utilizing Strain of *Prevotella albensis* from Bovine Rumen. *Microorganisms* **8**, doi:10.3390/microorganisms8122005 (2020).
- 78 Avgustin, G., Wallace Rj Fau - Flint, H. J. & Flint, H. J. 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*.
- 79 Ransom-Jones, E., Jones, D. L., McCarthy, A. J. & McDonald, J. E. The Fibrobacteres: an important phylum of cellulose-degrading bacteria. *Microb Ecol* **63**, 267-281, doi:10.1007/s00248-011-9998-1 (2012).
- 80 Brown, D. W. & Moore, W. E. C. Distribution of *Butyrivibrio Fibrisolvens* in Nature. *Journal of Dairy ence* **43**, 1570-1574 (1960).

- 81 Wang, Q. *et al.* A comparative study on rumen ecology of water buffalo and cattle calves under similar feeding regime. *Vet Med Sci* **6**, 746-754, doi:10.1002/vms3.302 (2020).
- 82 Prins, R. A., Hungate, R. E. & Prast, E. R. Function of the omasum in several ruminant species. *Comparative Biochemistry and Physiology Part A: Physiology* **43**, 155-163, doi:[https://doi.org/10.1016/0300-9629\(72\)90477-X](https://doi.org/10.1016/0300-9629(72)90477-X) (1972).
- 83 Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152, doi:10.1093/bioinformatics/bts565 (2012).
- 84 Huerta-Cepas, J. *et al.* eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research* **44**, D286-D293, doi:10.1093/nar/gkv1248 (2016).
- 85 Huerta-Cepas, J. *et al.* Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol Biol Evol* **34**, 2115-2122, doi:10.1093/molbev/msx148 (2017).
- 86 Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* **37**, D233-238, doi:10.1093/nar/gkn663 (2009).
- 87 Zhang, H. *et al.* dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **46**, W95-W101, doi:10.1093/nar/gky418 (2018).
- 88 McCann, J. C., Wickersham, T. A. & Loor, J. J. High-throughput Methods Redefine the Rumen Microbiome and Its Relationship with Nutrition and Metabolism. *Bioinform Biol Insights* **8**, 109-125, doi:10.4137/BBI.S15389 (2014).
- 89 Bolger, A. M., Marc, L. & Bjoern, U. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
- 90 Luo, X. *et al.* Understanding divergent domestication traits from the whole-genome sequencing of swamp- and river-buffalo populations. *National Science Review* **7**, 686-701, doi:10.1093/nsr/nwaa024 (2020).
- 91 Shen, Y. *et al.* Update soybean Zhonghuang 13 genome to a golden reference. *Sci China Life Sci* **62**, 1257-1260, doi:10.1007/s11427-019-9822-2 (2019).
- 92 Jiao, Y. *et al.* Improved maize reference genome with single-molecule technologies. *Nature* **546**, 524-527, doi:10.1038/nature22971 (2017).
- 93 Tang, H. *et al.* An improved genome release (version Mt4.0) for the model legume *Medicago truncatula*. *BMC Genomics* **15**, 312, doi:10.1186/1471-2164-15-312 (2014).

- 94 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
- 95 Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359, doi:10.7717/peerj.7359 (2019).
- 96 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv* **1303** (2013).
- 97 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
- 98 Asnicar, F., Weingart, G., Tickle, T. L., Huttenhower, C. & Segata, N. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ* **3**, e1029, doi:10.7717/peerj.1029 (2015).
- 99 Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119, doi:10.1186/1471-2105-11-119 (2010).

Figures

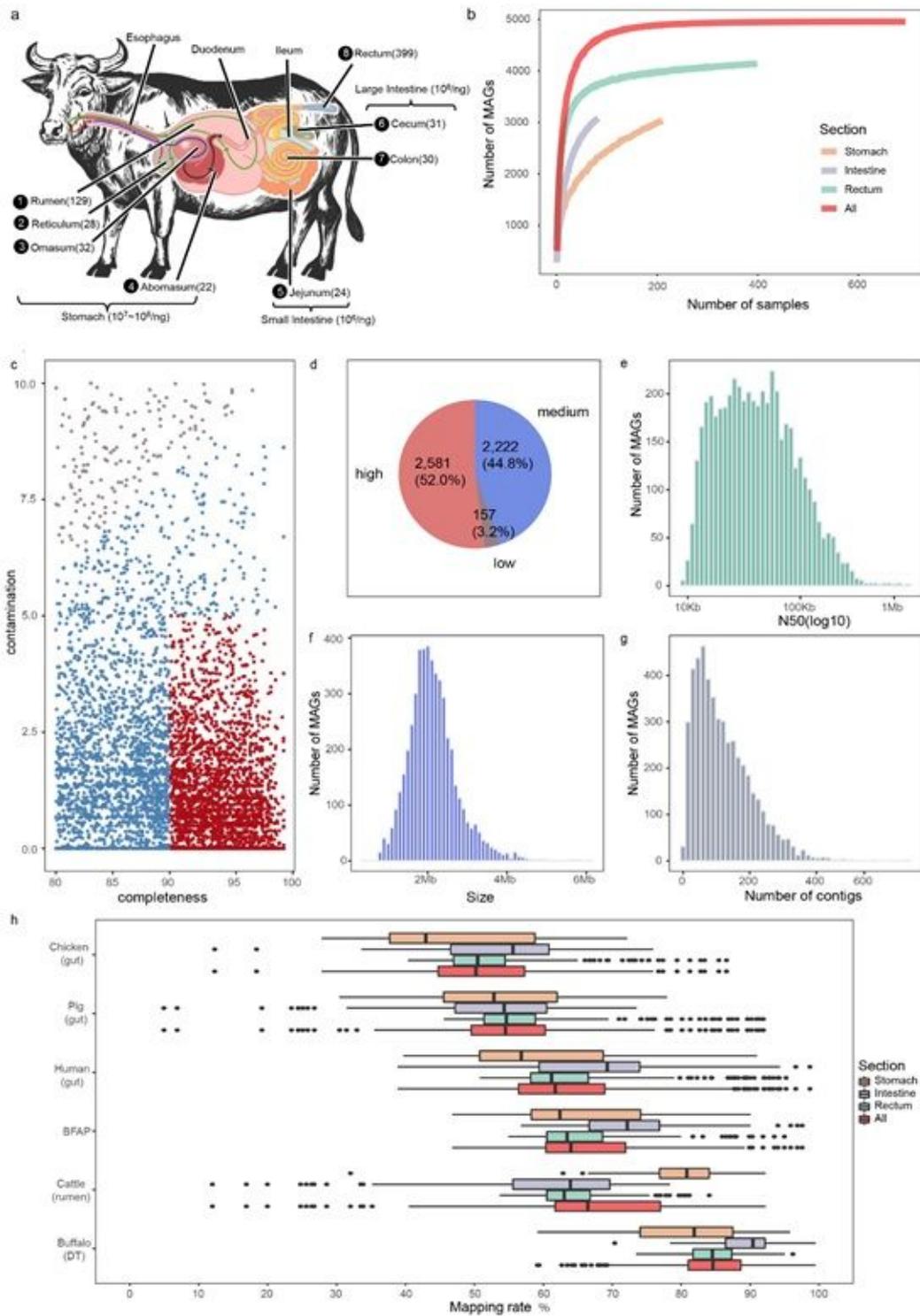


Figure 1

Generation and quality assessment of 4,960 metagenome-assembled genomes (MAGs) along the digestive tract of buffalo. a) A graphical representation of samples collected from distinct sites along the digestive tract of buffalo. Distinct sites were labeled; the arrows inside the digestive tract indicate the flow of the food. The numbers in parentheses after the labels indicate the number of samples obtained for this study. The numbers before the labels indicate the order of food flow; please note only the sites with

samples collected were ordered. The number beside curly braces indicate the estimated numbers of microbes per nanogram DNA in Stomach, Small intestine and Large intestine of DT. b) Rarefaction curves of assembled genomes (MAGs) as a function of input samples. Here the Y-axis represents the median number of non-redundant high-quality MAGs (those shown in panel c) obtained from a specific number of samples (X-axis) using 100-times random sampling with replacement. c) Quality assessment (i.e., completeness and contamination statistics) for the 4,960 non-redundant high-quality MAGs. Each point represents a MAG. Red points indicate the highest quality genomes with $\geq 90\%$ completeness and $\leq 5\%$ contamination. All other MAGs are $>80\%$ complete and $\leq 10\%$ contaminated. Those in blue have a quality score ≥ 50 as defined by Parks et al 62, whereas those in grey have a quality score <50 . d) Pie chart shows the numbers and relative proportions of the red, blue and grey MAGs in c). Histograms in e), f) and g) show the distributions of N50, genome size and the number of contig per genome respectively for the 4,960 MAGs. h) Increased coverage of metagenomics reads by our MAGs as compared with reference microbial genomes and MAGs of other model organisms. Shown here are percentages of reads from stomach (orange), intestine (purple), rectum (green) and all samples that could be mapped to the collected datasets. In addition to our MAGs (buffalo), chicken, pig and human MAGs were obtained from gut metagenomes [9, 11, 35]; BFAP was the combination of reference genomes datasets including bacterial, fungal, archaeal and protozoan genomes from NCBI RefSeq and the genomes from the Hungate collection 63; cattle MAGs were obtained from rumen metagenomes [37].

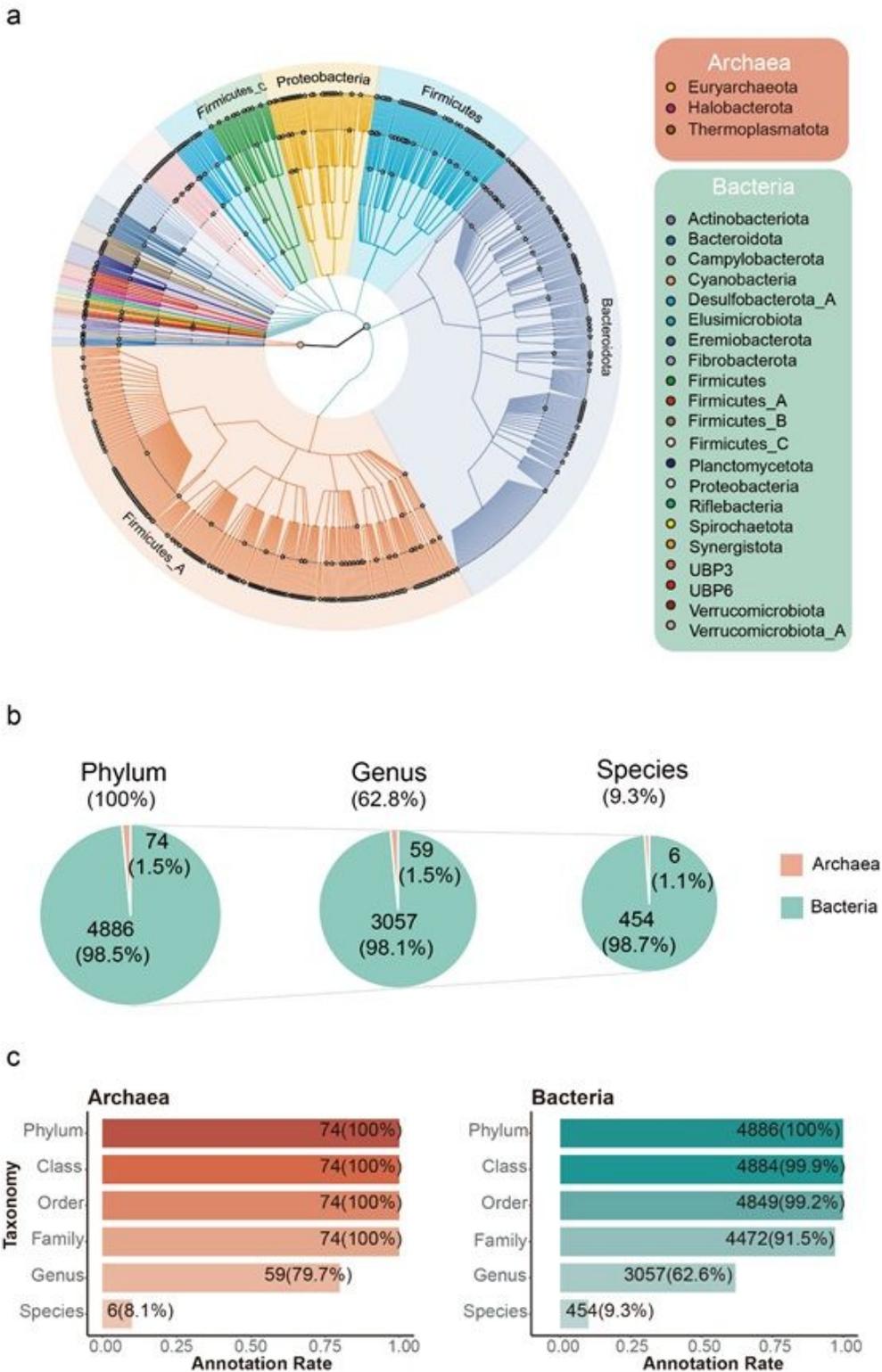


Figure 2

Classification of the MAGs to known taxonomy at different taxonomic levels. a) Phylogenetic relationships and taxonomic classifications of the 4,960 MAGs from the buffalo digestive tract. Shown here is the circular cladogram representation of the phylogenetic relationships of the 4,960 MAGs. The stars at the internal and leaf branches indicates novel branches that were not presented in GTDB-TK 67. Phyla were highlighted with different background colors, with the outermost labels indicate selected

phyla with top numbers of MAGs. The right panel lists all the phyla belonging to Archaea and Bacteria; the filled colors of the dots before the phylum labels correspond to the branch colors in the phylogenetic tree. b) The classification rates of 4,960 MAGs at different taxonomic levels. The numbers above the pie charts indicate the percentages of MAGs (out of 4,960) could be annotated at the respective levels; the numbers inside the pie charts indicate the percentages of archaea (orange) and bacteria (green) of each pie. c) The classification rates of archaea (left) and bacteria (right) at different taxonomic levels. The numbers indicate the amounts of MAGs classified.

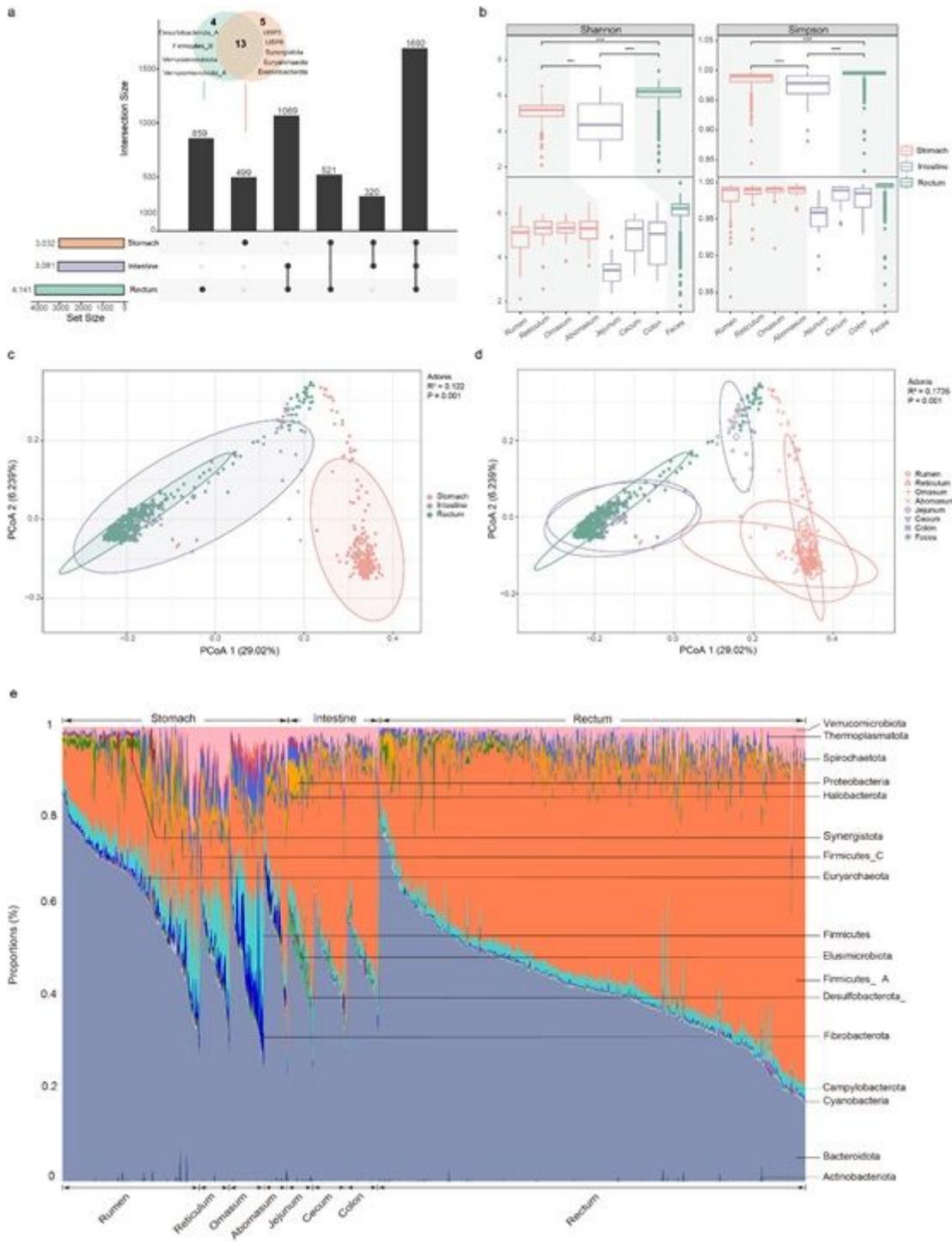


Figure 3

Taxonomic characterization of the buffalo MAGs along the digestive tract (DT). a) The distributions of MAGs in different parts along the DT. Here a MAG is considered to present in a section if its coverage is higher than $1\times$ in more than one samples of the section. The coverage is defined as the total bases mapped to a MAG in a sample divided by its length. b) Boxplots showing the alpha diversity measurements including Shannon (left panel) and Simpson (right panel) indexes of the DT sections (upper panel) and sites (lower panel). Pair-wise Wilcoxon Rank Sum Test was used to compare between the sections; * $P \leq 0.05$, *** $p \leq 0.001$, **** $p \leq 0.0001$. c) Principal coordinates analysis (PCoA) analysis of Bray-Curtis distances among the samples for the three sections. The p-value and R2 were reported by the 'adonis' function from the 'vegan' package of R using a non-parametric MANOVA test (see Materials and Methods for details). d) Principal coordinates analysis (PCoA) analysis of Bray-Curtis distances among the samples for the eight DT sites. e) The stream graph displaying the distribution of phyla along the DT of buffalo. X-axis indicating samples clustered by the sampled sites along the DT. Y-axis indicating the relative abundances of the phyla in each sample.

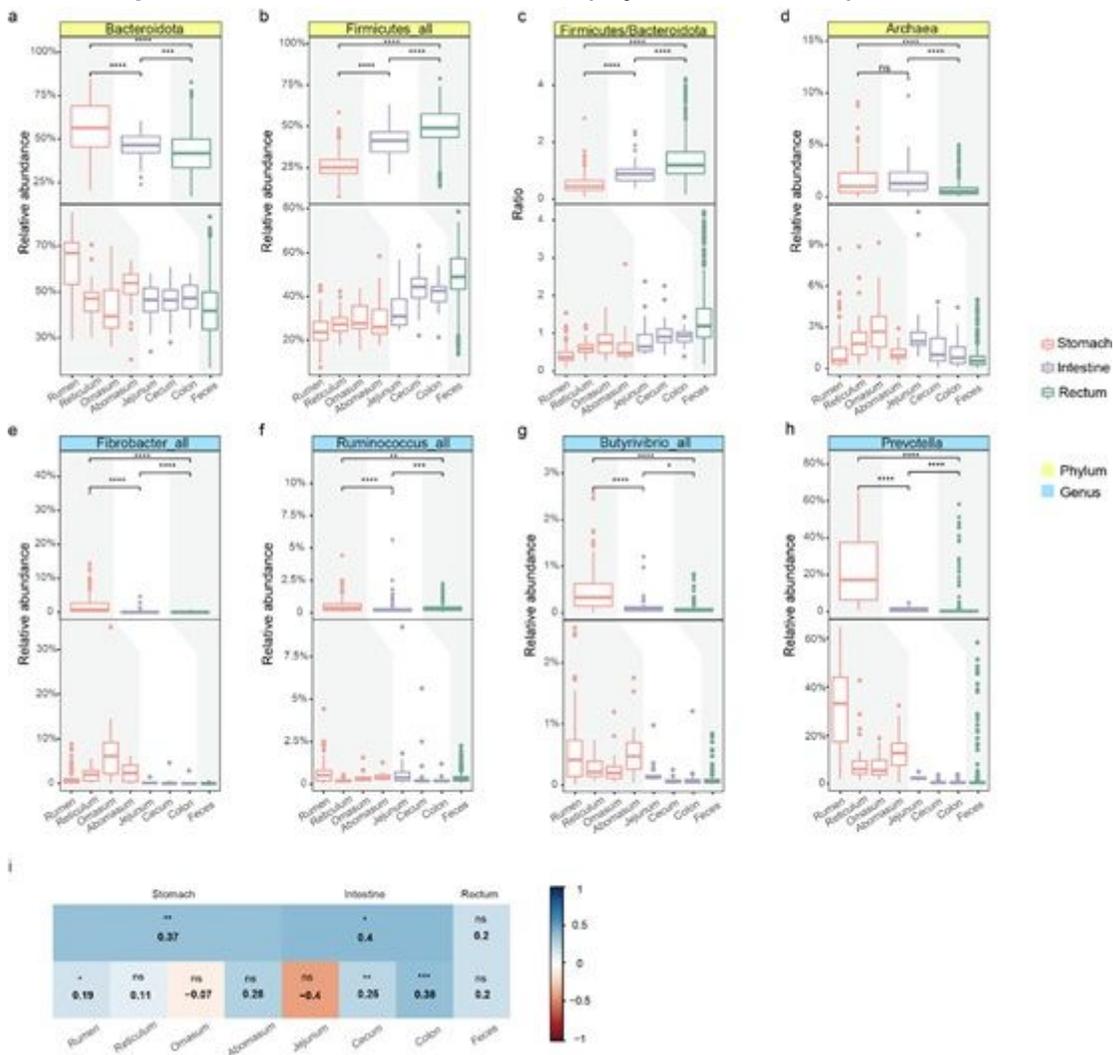


Figure 4

Distinct abundance patterns of the MAGs correlate with the functions of digestive tract. Panels a-h, except c) use boxplots to summarize the abundance distributions of taxa of interests in the three sections

(upper part) and eight sites (lower part); Y-axis for in these panels indicate the relative abundances, while that of c indicates the abundance ratios between Firmicutes and Bacteroidota. Background colors of the panel labels indicate classification levels of microbes while the different color of boxes indicate the different sections of DT. Wilcoxon Rank Sum Test was used to perform pairwise comparisons between sections (the upper part). i) Heatmap showing the Spearman correlation between archaea and fibrobacter_all along the DT; the number indicating the correlation coefficient between archaea and fibrobacter_all. The correlation of all phyla and genus from a-h can be seen in Supplementary Fig.5. ns $P \geq 0.05$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

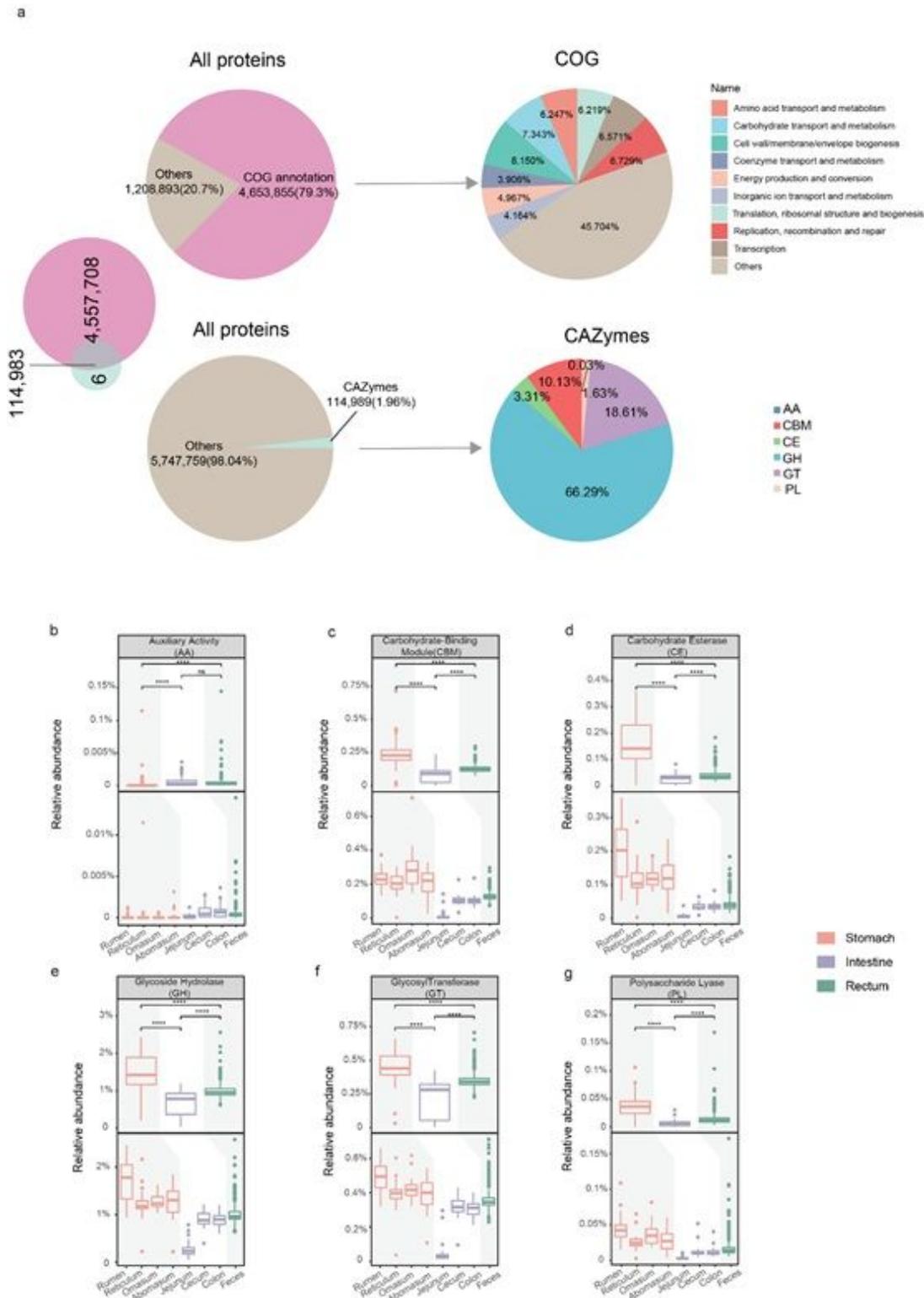


Figure 5

Functional annotation of the 5,789,172 non-redundant proteins encoded by buffalo MAGs and their abundance distributions along DT. a) functional annotations of buffalo microbial proteins. Annotation results using EggNOG-mapper (upper) and dbCAN2 (lower); pies show the proportions of proteins annotated by these two methods (left) and the overall categories (right). Boxplots summarize the abundance distributions of proteins (in functional groups) annotated by dbCAN2 including b) auxiliary

- [SupplementaryData1.xlsx](#)
- [SupplementaryData2.pdf](#)
- [SupplementaryData3.xlsx](#)