The rumen contains a complex microbial ecosystem that degrades plant materials, such as cellulose and hemicellulose. We herein reconstructed 146 nonredundant, rumen-specific metagenome-assembled genomes (MAGs), with ≥50% completeness and <10% contamination, from cattle in Japan. The majority of MAGs were potentially novel strains, encoding various enzymes related to plant biomass degradation and volatile fatty acid production. The MAGs identified in the present study may be valuable resources to enhance the resolution of future taxonomical and functional studies based on metagenomes and metatranscriptomes.

Key words: Japanese cattle, microbiome, metagenomics, metagenome-assembled genomes, rumen
dereplicated and aggregated using DAS Tool version 1.1.4 (Sieber et al., 2018). The quality (completeness and contamination) of MAGs was evaluated in CheckM version 1.1.3 (Parks et al., 2015), retaining genomes with ≥50% completeness and <10% contamination. Retained MAGs were dereplicated in dRep version 3.2.2 (Olm et al., 2017) with the “dRep dereplicate” command at 99% and 95% average nucleotide identities (ANI) using a default clustering algorithm, which were hereafter called RUG1ANI99% and RUG1ANI95%, respectively, with MAGs in these groups being defined as nonredundant strains and species, respectively (Stewart et al., 2019).

Filtered reads were mapped against RUG1ANI99% MAGs using BamM (https://github.com/Ecogenomics/BamM) to prevent arbitrary mapping between similar genomes (Robbins et al., 2021). Genome relative abundance was calculated using the CoverM version 0.4.0 “genome” (https://github.com/wwood/CoverM) with the following parameters: -m relative_abundance --min-read-percent-identity 0.95 --min-read-aligned-percent 0.75. We taxonomically classified RUG1ANI99% MAGs in GTDB-tk version 2.1.1 (Chaumeil, P.A., et al., 2022). GTDB-Tk v2: memory friendly classification with the Genome Taxonomy Database. bioRxiv. https://doi.org/10.1101/2022.07.11.499641) with GTDB release 207. We then built a phylogenetic tree in PhyloPhlAn version 3.0.60 (Asnicar et al., 2020). The tree was visualized in iTOL version 6.1.2 (Letunic and Bork, 2021). To elucidate whether RUG1ANI99% MAGs represented potential novel species and strains, we compared them with the genomes of rumen prokaryotes (hereafter called RUG2), including Hungate1000 (Seshadri et al., 2018), and previously reported rumen MAGs (Stewart et al., 2019; Anderson and Fernando, 2021) using dRep (Olm et al., 2017). Genomes were defined as novel species (<95%) and strains (<99%) using the ANI outputs by dRep and GTDB-tk.

We predicted the proteins encoded by RUG1ANI99% MAGs with Prodigal version 2.6.3 (Hyatt et al., 2010), and then searched them against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database through GhostKOALA (Kanehisa et al., 2016). Carbohydrate-active enzyme (CAZyme) families, which are associated with cell wall degradation and essential for efficient lignocellulose processing in ruminants, were annotated using dbCAN2 (Zhang et al., 2018). This software integrates three tools (HMMER, DIAMOND, and Hotpep), and we only retained the CAZyme domains annotated by at least two of the three. The predicted protein sequences were then mapped to the Pfam database (Finn et al., 2016) using HMMER version 3.3.2 (Mistry et al., 2013) with an E value <10^{-3} to search cohesin (PF00963) and dockerin (PF00404) domains. We used PULpy (Stewart, R.D., et al. 2018) Open prediction of polysaccharide utilization loci (PUL) in 5,414 public Bacteroidetes genomes using PULpy. bioRxiv. https://doi.org/10.1101/421024) to predict polysaccharide utilization loci (PUL), linked gene clusters that encode the cell envelope-associated enzymes required for sensing, binding, and degrading polysaccharide substrates (Bjursell et al., 2006). All novel sequence raw data were deposited in DDBJ (accession number DRA014084). The 146 RUG1ANI99% MAGs are available at https://doi.org/10.6084/m9.figshare.20705449.

Filtered samples yielded 706 M reads, and we retained 291 bins. Of these, 202 MAGs had ≥50% completeness and <10% contamination, with 54 being high quality (≥90% completeness and <5% contamination). Following dereplication at 99% ANI, we generated 146 nonredundant RUG1ANI99% MAGs with 36 high-quality MAGs. The range of their sum genome length was 0.41–5.84 Mb, with N50 ranging between 1.66 and 93.2 kb (Table S2). A comparison with MAGs in the RUG2 (8,153 genomes) and GTDB (317,542 genomes) databases revealed 16 and 130 RUG1ANI99% MAGs that were potential novel species and strains, respectively (Table S2). Therefore, the RUG2 and GTDB databases did not completely cover the diversity of the rumen microbiota from the tested cattle.

After aligning filtered reads against RUG1ANI99% MAGs, we identified 32 MAGs that were present (relative abundance >0) in >90% of the tested cattle (n=20), suggesting that they are core rumen bacteria in Japan (Table S3). The average mapping rate was 43.9% per sample, indicating that they are core rumen bacteria in Japan (Table S3). This software integrates three tools (HMMER, DIAMOND, and Hotpep), and we only retained the CAZyme domains annotated by at least two of the three. The predicted protein sequences were then mapped to the Pfam database (Finn et al., 2016) using HMMER version 3.3.2 (Mistry et al., 2013) with an E value <10^{-3} to search cohesin (PF00963) and dockerin (PF00404) domains. We used PULpy (Stewart, R.D., et al. 2018) Open prediction of polysaccharide utilization loci (PUL) in 5,414 public Bacteroidetes genomes using PULpy. bioRxiv. https://doi.org/10.1101/421024) to predict polysaccharide utilization loci (PUL), linked gene clusters that encode the cell envelope-associated enzymes required for sensing, binding, and degrading polysaccharide substrates (Bjursell et al., 2006). All novel sequence raw data were deposited in DDBJ (accession number DRA014084). The 146 RUG1ANI99% MAGs are available at https://doi.org/10.6084/m9.figshare.20705449.

The 32 and 50 RUG1ANI99% MAGs were classified as members of Bacteroidota and Firmicutes_A, respectively (Fig. 1 and Table S2); this taxonomic assignment is consistent with that reported in a previous study in which many MAGs reconstructed from the cattle rumen were classified into these phyla (Gharechahi et al., 2021). The remaining RUG1ANI99% MAGs were assigned to the following 12 phyla: Firmicutes_C (23 MAGs), Actinobacteriota (10 MAGs), Firmicutes (9 MAGs), Verrucomicrobiota (7 MAGs), Spirochaetota (4 MAGs), Patescibacteria (3 MAGs), Proteobacteria (2 MAGS), Methanobacteriota (2 MAGs), Planctomycetota (1 MAG), Synergistota (1 MAG), Desulfbacterota_1 (1 MAG), and Elusimicrobiota (1 MAG). In Bacteroidota MAGs, 10 RUG1ANI99% MAGs were classified as Prevotella, which is the most abundant bacterial genus in the rumen (Henderson et al., 2015), and its abundance increases with high-concentrate diets (Luo et al., 2017). Most of the animals tested were offered high-concentrate diets (Table S1), resulting in the reconstruction of Prevotella MAGs. In the present study, 14 Lachnospiraceae MAGs were reconstructed; however, we previously recovered only 3 Lachnospiraceae MAGs (Sato et al., 2021). Lachnospiraceae is associated with feed efficiency in ruminants (Li and Guan, 2017) and is crucial for shaping the characteristics of JB rumen microorganisms (Ogata et al., 2019; Sato et al., 2021). Therefore, the Lachnospiraceae MAGs obtained herein may be useful for metagenomic and metatranscriptomic research in ruminants, particularly JB cattle.

Prodigal (Hyatt et al., 2010) predicted 330,575 CDS from RUG1ANI99% MAGs, with a range of 537–8,276 (Table S2). We found 156,599 CDS (47.4%) in the KEGG Orthology (KO) database. Consistent with our previous findings (Sato et al., 2021), a KO analysis indicated that most RUG1ANI99% MAGs were roughly clustered according to their phylum (Fig. S1). We focused on genes related to VFA production
because VFAs are a major energy source for ruminants (Table 1 and Fig. S2). We identified 113 RUG1\_ANI99\% MAGs (77.4%) with at least one gene related to acetate production (K00925: acetate kinase, K00625: phosphate acetyltransferase). These genes were absent from Firmicutes\_C RUG1\_ANI99\% MAGs. Rumen propionate is produced via the succinate and acrylate pathways (Jeyanathan et al., 2014). As markers of the succinate pathway, we selected genes encoding methylmalonyl-CoA mutase (K01847, K01848, and K01849), methylmalonyl-CoA epimerase (K05606), and methylmalonyl-CoA decarboxylase (K11264); acrylate pathway markers were genes encoding lactoyl-CoA dehydratase (K20626 and K20627). Succinate pathway marker genes were present in 52 RUG1\_ANI99\% MAGs; most were MAGs belonging to Bacteroidota and Firmicutes\_C, including Prevotella and Succinoclasticum MAGs (10 and 15 MAGs, respectively). Prevotella (Wirth et al., 2018) and Succinoclasticum (Van Gylswyk, 1995) both produce propionate from succinate in the rumen. In contrast, none of the RUG1\_ANI99\% MAGs possessed the acrylate pathway marker gene.

Among 146 RUG1\_ANI99\% MAGs, 136 and 112 possessed the genes encoding CAZymes GH13 (α-amylase) and GH77 (4-α-glucanotransferase), respectively, both of which are involved in starch degradation (Fig. S3 and Table S4). Furthermore, oligosaccharide-degrading GH2 (β-galactosidase) and GH3 (β-glucosidase) were present in 92 and 98 RUG1\_ANI99\% MAGs, respectively. GH2, GH3, and GH13

---

**Table 1.** The number of RUG1\_ANI99\% MAGs associated with volatile fatty acid production

<table>
<thead>
<tr>
<th>Phylum</th>
<th>Acetate</th>
<th>Propionate</th>
<th>Succinate pathway</th>
<th>Acrylate pathway</th>
<th>Butyrate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actinobacteriota (10)</td>
<td>9</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Bacteroidota (32)</td>
<td>29</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>Desulfbacterota I (1)</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Elusimicrobiota (1)</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Firmicutes (9)</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Firmicutes A (50)</td>
<td>48</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>Firmicutes C (23)</td>
<td>0</td>
<td>23</td>
<td>0</td>
<td>0</td>
<td>22</td>
</tr>
<tr>
<td>Patescibacteria (3)</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Planctomycetota (1)</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Proteobacteria (2)</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Spirochaetota (4)</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Synergistota (1)</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Verrucomicrobiota (7)</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Methanobacteriota (2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

1 The number after the phylum name shows MAG numbers of the phylum in RUG1\_ANI99\% MAGs.
2 Numbers of MAGs that have genes assigned as K00925 and/or K00625.
3 Numbers of MAGs that possessed at least one succinate pathway marker gene (K01847, K01848, K01849, K11264, and K05606) or acrylate pathway marker gene (K20626 and K20627).
4 Numbers of MAGs with at least one gene related to butyrate production (K00634, K00929, K01034, K00626, and K00074).
were among the most dominant CAZymes in the cattle rumen (Wang et al., 2019). Additionally, Prevotella RUG1\textsubscript{ANI99\%} MAGs had multiple genes related to CAZymes (Fig. S3), which is in accordance with the ability of the genus to degrade fiber sources, such as hemicellulose and pectin (Stewart et al., 1997). All Prevotella RUG1\textsubscript{ANI99\%} MAGs possessed genes encoding GH2 (β-galactosidase), GH3 (β-glucosidase), GH5 (endo-β-1,4-glucanase, endo-β-1,4-xylanase, and β-mannosidase), GH13 (α-amylase), GH26 (endo-β-1,4-mannanases), GH32 (invertase), GH36 (α-galactosidase), GH43 (β-xilosidase), GH73 (lysozyme), GH94 (cellobiose phosphorylase), and GH97 (α-glucosidase).

Cellulosomes (Bayer et al., 2004; Artzi et al., 2017) and amylosomes (Ze et al., 2015) consist of scaffolding subunits that contain cohesion modules and dockerin modules, which append to an enzyme for the degradation of lignocellulose and starch, respectively. Cohesin modules specifically interact with dockerin modules, and the cohesin–dockerin interaction is responsible for the integration of enzymes into the MAGs. The proteins containing cohesin and dockerin domains are summarized in Table S5. In total, 200 dockerin and 18 cohesin domains were present in 25 RUG1\textsubscript{ANI99\%} MAGs, including 2 Ruminococcus and 11 Ruminococcus\textsubscript{E} MAGs. Sixteen dockerin-containing proteins carried CAZyme domains, mostly including GH families containing α-amylase (GH13_14, GH13_19, and GH13_28). Among the 25 RUG1\textsubscript{ANI99\%} MAGs, 13 had both cohesin- and dockerin-containing proteins, while 9 had only dockerin-containing proteins. Dassa et al. (2014) reported that Ruminococcus albus 8, which degrades cellulose substrates, harbored no cohesion-containing protein. Therefore, we cannot rule out the possibility that RUG1\textsubscript{ANI99\%} MAGs, which had no cohesion- and some dockerin-containing proteins, used an alternative mechanism for the immobilization of dockerin-containing enzymes onto carbohydrates, similar to Ruminococcus albus 8 (Dassa et al., 2014).

In Bacteroidetes, CAZymes are often organized in PULs. We herein identified 215 PULs in Bacteroidota RUG1\textsubscript{ANI99\%} MAGs, and 93 lacked any known CAZyme domains (Table S6). The most common PUL was a single susC/susD-like pair, which was consistent with previous findings (Stewart et al., 2019). Among 32 Bacteroidota MAGs, 29 had at least one PUL. The number of PULs per Bacteroidota MAG ranged between 1 and 18, with the highest number being observed in Prevotella MAG and Cryptobacteroides MAG (JB\textsubscript{AI} 91\_bin_120 and F1\textsubscript{AI} 424\_bin_68, respectively). Ninety-eight and 80 PULs were identified in Cryptobacteroides MAG and Prevotella MAGs, respectively, and contained various CAZymes. Overall, the present results suggest that Cryptobacteroides and Prevotella are important for rumen function (specifically carbon degradation) in cattle in Japan.

In summary, we reconstructed 146 rumen-specific MAGs from cattle in Japan, with many being potentially novel. These MAGs are valuable resources for enhancing the resolution of future metagenome- and metatranscriptomic-based taxonomical and functional studies.

Acknowledgements

The present study was partly supported by a Grant-in-Aid from the Japan Society for the Promotion of Science (JSPS) Research Fellow Number 20J15021 and JSPS KAKENHI Grant Number JP21H05057. Super-computing resources were provided by the Human Genome Center, Institute of Medical Science, University of Tokyo. We thank members of the Iwate Agricultural Research Center Animal Industry Research Institute for helping us collect rumen samples.

References


Metagenomics of Japanese Cattle Rumen


