Metagenomic analysis of the cow, sheep, reindeer and red deer rumen

Authors: Laura Glendinning [1]*, Buğra Genç[2#], R John Wallace[2], Mick Watson[1]

[1]Genetics and Genomics, The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, Midlothian, United Kingdom

[2]Rowett Institute, University of Aberdeen, Ashgrove Rd W, Aberdeen, Aberdeenshire, United Kingdom

#Current address: Ondokuz Mayis University Veterinary Medicine Faculty Department of Laboratory Animals, Atakum, Samsun, Turkey

*Corresponding author: laura.glendinning@roslin.ed.ac.uk

Keywords: microbiota, deer, cow, sheep, reindeer, mag

Repositories: European Nucleotide Archive: PRJEB34458. Edinburgh DataShare: https://doi.org/10.7488/ds/2640.

Abstract:

The rumen microbiota comprises a community of microorganisms which specialise in the degradation of complex carbohydrates from plant-based feed. These microbes play a highly important role in ruminant nutrition and could also act as sources of industrially useful enzymes. In this study, we performed a metagenomic analysis of samples taken from the ruminal contents of cattle (*Bos Taurus*), sheep (*Ovis aries*), reindeer (*Rangifer tarandus*) and red deer (*Cervus elaphus*). We constructed 391 metagenome-assembled genomes originating from 16 microbial phyla. We compared our genomes to other publically available microbial genomes and found that they contained 279 novel species. We also found significant differences between the microbiota of different ruminant species in terms of the abundance of microbial taxonomies, carbohydrate-active enzyme genes and KEGG orthologs. However, we found that the vast majority of carbohydrate-active enzymes were present in all of our sample types, which may indicate that there is a core set of these enzymes which are present across ruminants and are independent of diet and environmental conditions. We present a dataset of rumen-derived genomes which in combination with other publicly-available rumen genomes can be used as a reference dataset in future metagenomic studies.

Data Summary:

The paired-read fastq files supporting the conclusions of this article are available in the European Nucleotide Archive repository (https://www.ebi.ac.uk/ena/browser/view/PRJEB34458). The RUG fasta files supporting the conclusions of this article are available in the Edinburgh DataShare repository (https://doi.org/10.7488/ds/2640).

Introduction:

The microbial communities which inhabit the rumen contain a mixture of bacteria, fungi, protozoa, viruses and archaea, and through fermentation are able to convert complex plant carbohydrates into short-chain volatile fatty acids. The microbial pathways present have a large impact on feed efficiency (1-3), alongside other important production traits such as milk and fat yield (4, 5). Understanding the processes by which food is digested in the rumen may allow us to improve feed efficiency in ruminants (1), either by the production of enzymes isolated from microbes (6) or by manipulating the microbiota through the use of pre- or probiotics (7). There are also other potential industrial uses for the proteins produced by ruminal microbes, for example in processing biofuels, bioremediation, processing pulp/paper and textile manufacturing (8-11). Ruminants are also the largest source of anthropogenic methane emissions and gaining a greater understanding of which microbes are important in methane production could lead to improved methane mitigation strategies (7, 12-16).

While inroads have been made towards culturing members of the ruminal microbiota (17, 18) there are still many members which have not been characterised. Metagenomics is a powerful tool which allows us to examine the entire genetic repertoire of the rumen microbiota without the need for culturing. Our group has previously published thousands of metagenome assembled genomes from cattle rumen samples (19, 20) and hundreds of genomes from chicken caecal samples (21), many of which were identified as novel species.

Several studies have examined the rumen microbiota using metagenomic techniques in cattle and sheep; however, less effort has been made to characterise the microbiota of other ruminant species which may be less commercially-important but which could harbour microbes which could be industrially useful. For example, wild ruminants are likely to consume a far more diverse diet than farm-raised individuals, and are therefore likely to contain microbes which are able to digest different substrates. In this paper we analyse rumen metagenomic data from four ruminant species:

63 cattle (*Bos Taurus*), sheep (*Ovis aries*), red deer (*Cervus elaphus*) and reindeer (*Rangifer tarandus*).
64 We compare the microbiota of these species taxonomically and functionally and construct 391
65 named rumen-uncultured genomes (RUGs), representing 372 putative novel strains and 279 putative
66 novel species.

67

68 **Methods:**

69 **Experimental design**

70 Reindeer (*Rangifer tarandus*: Grazing mixed vegetation, n=2) and red deer (*Cervus elaphus*: Grazing
71 mixed vegetation, n=4) were shot in the wild, and ruminal digesta samples were collected
72 immediately. Samples were taken from Holstein cattle (*Bos Taurus*: Fed total mixed ration (once a
73 day), n=4) and Finn-Dorset cross sheep (*Ovis aries*: Grazing mixed pasture, n=2) via a rumen cannula.
74 Samples were taken from sheep after morning grazing. Sheep sampling was performed as described
75 in McKain et al. (22). Cattle samples were taken 3 hours post feeding. Samples were collected from
76 the bovine rumen in the following locations: top near cannula, middle at the front of the rumen,
77 middle towards the back of the rumen and bottom (approximately 45cm down from the entrance to
78 the rumen). Digesta samples were mixed with buffer containing glycerol as a cryoprotectant (22).
79 The mixtures were kept on ice for 1-2 hours then frozen at -20°C. DNA extraction was performed
80 using repeated bead beating plus column filtration, as described in (23). Shotgun sequencing was
81 performed on an Ilumina Hiseq 2000, producing an average of 1626 million paired reads per sample,
82 of 100bp or 150bp in length.

83

84 **Bioinformatics**

85 Illumina adaptors were removed using trimmomatic (24) (v.0.36). IDBA-UD (25) (v.1.1.3) with the
86 options --num_threads 16 --pre_correction --min_contig 300 was used to perform single sample

assemblies. After indexing using BWA index (v.0.7.15), BWA-MEM was used to map reads to assemblies (26). BAM files were created by SAM tools (27) (v.1.3.1) and coverage was calculated using the command jgi_summarize_bam_contig_depths from the MetaBAT2 (v.2.11.1) software package (28). A coassembly was carried out on all samples using MEGAHIT (29) (v.1.1.1) with the options --continue --kmin-1pass -m 100e+10 --k-list 27,37,47,57,67,77,87 --min-contig-len 1000 -t 16. After filtering out reads which were <2kb, indexing and mapping were performed as for single assemblies.

Metagenomic binning was carried out using MetaBAT2 with the options --minContigLength 2000, --minContigDepth 2. From the single-assemblies, 1691 bins were created and from the co-assembly 2508 bins were created. Completeness and contamination of bins were calculated using CheckM (options: lineage_wf, -t 16, -x fa) (v.1.0.5), and the bins were dereplicated using dRep (30) (options: dereplicate_wf -p 16 -comp 80 -con 10 -str 100 –strW 0) (v.1.1.2). Thus, bins were discarded if their completeness was <80% or if they had contamination >10%. The dereplicated 'winning' bins are referred to below as RUGs. MAGpy was used to compare the RUGs to public datasets (31). Taxonomies were assigned to MAGs using GTDB-Tk (32). Trees produced by MAGpy were rerooted at the branch between archaea and bacteria using Figtree (33) (v.1.4.4) and visualised using GraPhlAn (34) (v.0.9.7). For submission to public repositories, our RUGs were named as the lowest taxonomic level at which NCBI and GTDB-Tk matched. The taxonomies assigned to RUGs were manually checked against the taxonomic tree and improved accordingly.

Carbohydrate-active enzymes (CAZymes) were identified using dbCAN2 (version 7, 24th August 2018) by comparing RUG proteins to the CAZy database (35). RUG proteins were compared to the KEGG database (downloaded on Sept 15th 2018) (36) using DIAMOND (37) (v0.9.21). KEGG hits for which the alignment length was ≥90% of the query length were retained. The likely KEGG ortholog group for each RUG protein was inferred from the DIAMOND search results and the KEGG database. CAZyme and KEGG ortholog abundances were calculated as the sum of the reads mapping to RUG

proteins within each group after using DIAMOND to align reads to the RUG proteins. PULpy was used to identify polysaccharide utilisation loci (38).

Statistical analyses were carried out within R (version 3.5.1). The ggplot2 (39) package was used to construct scatter plots and NMDS graphs. The vegan package (40) was used to create NMDS axes using the Bray–Curtis dissimilarity. The Adonis function from the vegan package was used to perform PERMANOVA analyses and DeSeq2 (41) was used to calculate differences in coverage for individual CAZymes, KEGG orthologs and RUGs. UpSet graphs were constructed using the UpSetR package (42). Taxonomies were assigned to paired sequence reads with Kraken (43) using a custom kraken database consisting of RefSeq complete genomes with our RUGs and the rumen superset (20) added. Prior to statistical analyses (excluding DeSeq2) and graph construction, data was subsampled. For RUGs, subsampling to the lowest sample coverage was performed. CAZymes and KEGG orthologs were subsampled to the lowest sample abundance.

Results:

Construction of RUGs from rumen sequencing data

We produced 979G of Illumina sequencing data from 12 samples then performed a metagenomic assembly of single samples and a co-assembly of all samples. This created a set of 391 dereplicated genomes (99% ANI (average nucleotide identity)) with estimated completeness ≥80% and estimated contamination ≤10% (**Additional File 1: Fig 1**). 284 of these genomes were produced from the single-sample assemblies and 107 were produced from the co-assemblies. 172 genomes were >90% complete with contamination <5%, and would therefore be defined as high-quality draft genomes by Bower et al. (44). The distribution of these RUGs between our samples can be found in **Additional file 2** (based on coverage). **Additional file 3** contains the predicted taxonomic assignment for each RUG while **Fig 1** shows a phylogenetic tree of the genomes. The tree is dominated by the

136 *Bacteroidota* (136 RUGs: All order *Bacteroidales*) and the *Firmicutes_A* (121 RUGs), followed by
137 lesser numbers of the *Firmicutes_C* (40 RUGs), *Synergistota* (20 RUGs: All family *Aminobacteriaceae*),
138 *Firmicutes* (19 RUGs), *Proteobacteria* (15 RUGs), *Cyanobacteriota* (9 RUGs: All family
139 *Gastranaerophilaceae*), *Actinobacteriota* (7 RUGs), *Euryarchaeota* (7 RUGs: All family
140 *Methanobacteriaceae*), *Spirochaetota* (5 RUGs), *Elusimicrobiota* (3 RUGs: All family
141 *Endomicrobiaceae*), *UBP6* (3 RUGs: All genus *UBA1177*), *Fibrobacterota* (2 RUGs: All genus
142 *Fibrobacter*), *Riflebacteria* (2 RUGs: All family *UBA8953*), *Chloroflexota* (1 RUGs: family
143 *Anaerolineaceae*) and *Desulfobacterota* (1 RUGs: genus *Desulfovibrio*). All members of the phylum
144 *Firmicutes_A* belonged to the *Clostridia* class: orders *4C28d-15* (n=9), *CAG-41* (n=3),
145 *Christensenellales* (n=4), *Lachnospirales* (n=56), *Oscillospirales* (n=45), *Peptostreptococcales* (n=2)
146 and *Saccharofermentanales* (n=2). *Firmicutes_C* contains the orders *Acidaminococcales* (n=8) and
147 *Selenomonadales* (n=32). The phylum *Firmicutes* contained the orders *Acholeplasmatales* (n=3),
148 *Erysipelotrichales* (n=1), *Izimaplasmatales* (n=1), *ML615J-28* (n=1), *Mycoplasmatales* (n=1). *RFN20*
149 (n=7) and *RF39* (n=5), The *Actinobacteria* contained the orders *Actinomycetales* (n=1) and
150 *Coriobacteriales* (n=6). The *Proteobacteria* phylum contains the orders *Enterobacterales* (n=4),
151 *Paracaedibacterales* (n=1), *RF32* (n=8) and *UBA3830* (n=2). The *Spirochaetota* contains the orders
152 *Sphaerochaetales* (n=1) and *Treponematales* (n=4).

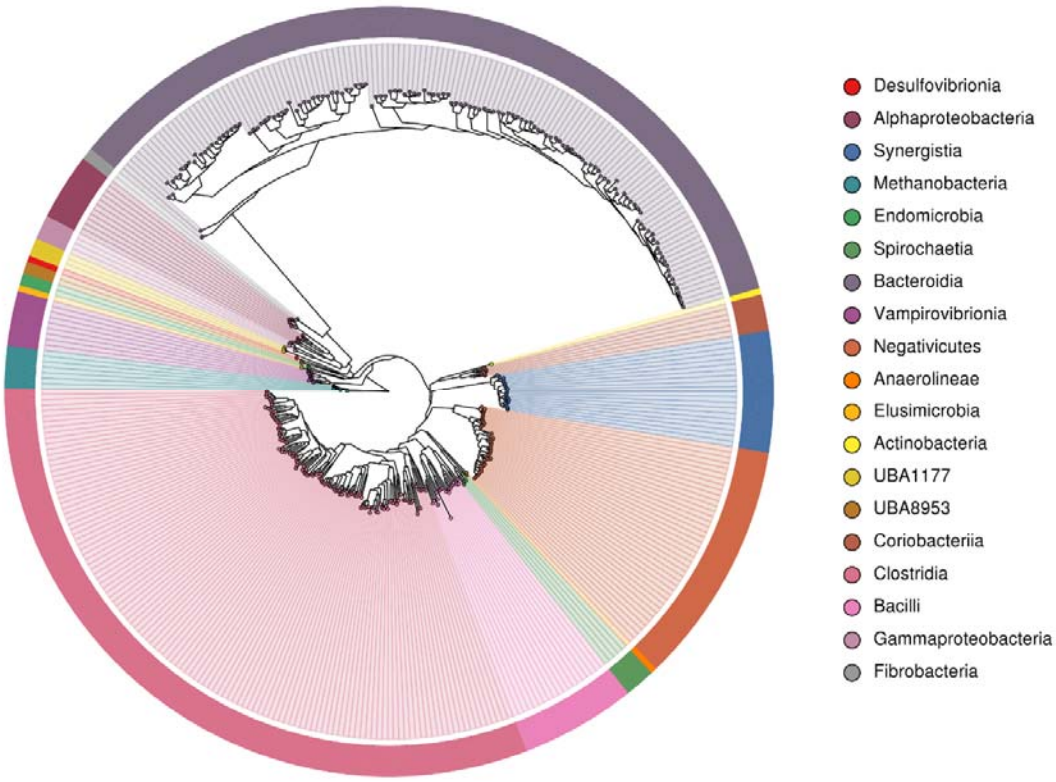


Fig 1: Phylogenetic tree of the 391 draft microbial genomes from rumen samples, labelled by taxonomic class. Taxonomies were defined by GTDB-Tk.

After sub-sampling, we found that samples from different ruminant species clustered significantly separately by abundance of RUGs (PERMANOVA: P = 3e-05). This may be due to the fact that the vast majority of RUGs were only found in a single host species (**Fig 2**), including 111 RUGs in red deer, 78 RUGs in reindeer, 40 RUGs in cow and 31 RUGs in sheep. Only 3 RUGs were found in ≥1X average coverage in all species: uncultured *Bacteroidaceae sp. RUG30019*, uncultured *Prevotella sp. RUG30028* and uncultured *Prevotella sp. RUG30114*.

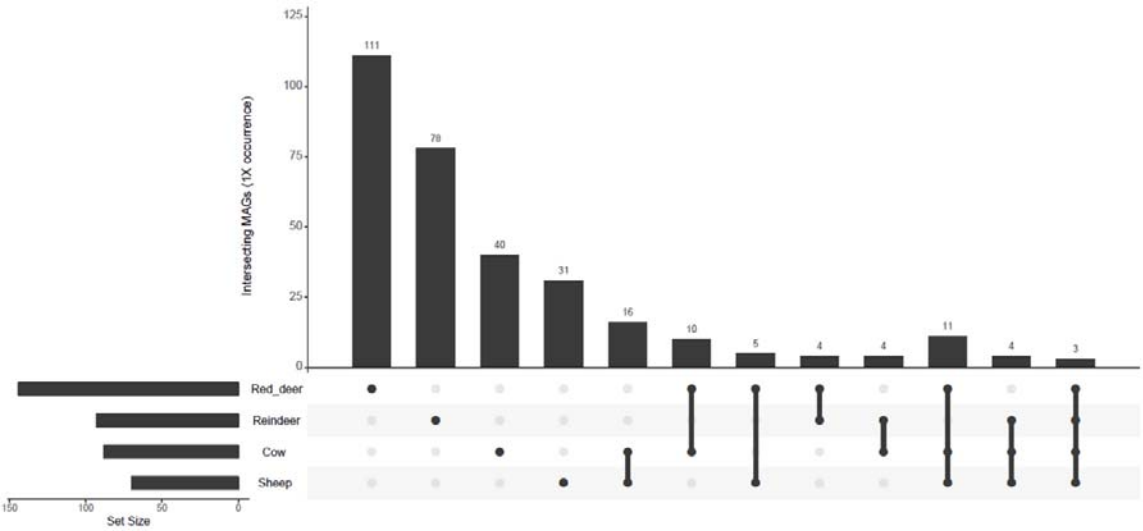


Fig 2: UpSetR graph showing the number of shared microbial genomes at average 1X coverage (after sub-sampling to equal depth) within four ruminant species.

We compared our RUGs to microbial genomes which had previously been sequenced from the rumen to determine if we had discovered any novel strains or species. We dereplicated our RUGs at 99% and 95% ANI to a "superset" of genomes containing rumen RUGs previously produced by our group (20), Hess et al. (11), Parks et al. (45), Solden et al. (46) and Svartström et al. (47) and the genomes from the Hungate collection (17). After dereplication at 99% and the removal of any RUGS with ≥99% ANI to an existing genome (as assigned by GTDB-Tk) or which clustered with members of the superset, 372 of our RUGs remained, representing putative novel strains. After dereplication at 95% and the removal of any RUGS with ≥95% ANI to an existing genome (assigned by GTDB-tk) or which clustered with members of the superset, 279 of our RUGs remained, representing putative novel species. The majority of these species originated from single-sample assemblies: 110 from red deer samples, 68 from reindeer samples, 23 from sheep samples and 1 from cattle samples, suggesting that many novel microbial species remain to be discovered from non-cattle ruminant hosts. These novel species are taxonomically diverse, with members belonging to the phyla *Bacteroidota* (n = 97), *Firmicutes_A* (n = 85), *Firmicutes_C* (n = 27), *Firmicutes* (n = 16), *Synergistota*

(n = 14), *Proteobacteria* (n = 11), *Cyanobacteriota* (n = 9), *Actinobacteriota* (n = 5), *Spirochaetota* (n = 4), *Euryarchaeota* (n = 3), *Elusimicrobiota* (n = 3), *Riflebacteria* (n = 2), *Chloroflexota* (n = 1), *Desulfobacterota* (n = 1) and *UBP6* (n = 1).

31 of our total RUGs were able to be taxonomically identified to species level and these contain bacteria which are commonly isolated from the rumen including novel strains of *Bacteroidales bacterium UBA1184* (45), *Bacteroidales bacterium UBA3292* (45), *Butyrivibrio fibrisolvens*, *Escherichia coli*, *Fibrobacter sp. UWB2* (48), *Lachnospiraceae bacterium AC3007* (17), *Lachnospiraceae bacterium UBA2932* (45), *Methanobrevibacter sp. UBA188* (45), *Methanobrevibacter sp. UBA212* (45), *Prevotella sp. UBA2859* (45), *Ruminococcaceae bacterium UBA3812* (45), *Ruminococcus sp. UBA2836* (45), *Sarcina sp. DSM 11001* (17), *Selenomonas sp. AE3005* (17), *Succiniclasticum ruminis* and *Succinivibrio dextrinosolvens*.

Comparing microbial taxonomies, CAZymes and KEGG orthologs between ruminant species

We assigned taxonomies to paired sequence reads using our custom kraken database containing RefSeq complete genomes, our RUGs, and the superset of rumen isolated microbial genomes. After subsampling we compared the abundance of members of the microbiota in different ruminant species at multiple taxonomic levels. Averaging reads across rumens species, the vast majority of reads mapped to bacteria (Sheep: 97%, Cow: 97%, Reindeer: 92%, Red deer: 98%) with smaller amounts of archaea (Sheep: 2.3%, Cow: 2.1%, Reindeer: 6.3%, Red deer: 1.9%) and Eukaryota (Sheep: 0.23%, Cow: 1.3%, Reindeer: 1.8%, Red deer: 0.56%). Eukaryota reads originated primarily from fungi and protists. In all ruminants, *Bacteroidetes* was the most abundant phylum (Sheep: 64%, Cow: 65% Reindeer: 54% Red deer: 52%), with *Firmicutes* being the second most abundant (Sheep: 29%, Cow: 26% Reindeer: 26% Red deer: 38%). Using PERMANOVA, significant differences in the abundance of taxonomies between ruminant species were found at both high (Kingdom: P =

205 0.01058, Phylum: P = 0.00017) and low (Family: P = 1e-05, Genus: P = 3e-05) taxonomic levels
206 (**Additional File 1: Fig 2**).

207 We also compared the abundance of genes encoding for specific CAZymes between species. These
208 enzymes are responsible for the synthesis, binding and metabolism of carbohydrates. The
209 carbohydrate esterases (CEs), glycoside hydrolases (GHs), glycosyltransferases (GTs) and
210 polysaccharide lyases (PLs) act to degrade cellulose, hemicellulose and other carbohydrates which
211 could otherwise not be digested by the host. Non-catalytic carbohydrate-binding modules (CBMs)
212 bind to specific carbohydrates, increasing the efficiency of enzymatic degradation (49). The auxiliary
213 activities (AAs) redox enzymes are reclassified CBMs which are lytic polysaccharide monooxygenases
214 (50). In our samples we found the following numbers of these CAZyme families: 6 AAs redox
215 enzymes, 39 CBMs, 14 CEs, 191 GHs, 61 GTs and 27 PLs. The ten most abundant GHs in the different
216 ruminant species were: for cows GH2, GH3, GH31, GH97, GH28, GH51, GH43_10, GH105, GH10 and
217 GH95; for sheep GH2, GH3, GH28, GH31, GH97, GH32, GH51, GH77, GH78 and GH95; for red deer
218 GH2, GH3, GH31, GH97, GH77, GH32, GH51, GH109, GH28 and GH78; and for reindeer GH2, GH3,
219 GH92, GH109, GH97, GH13, GH31, GH78, GH28 and GH77. Different ruminant species were found to
220 have significantly differently abundant CAZyme genes (PERMANOVA: P = 1e-05, **Additional File 1: Fig**
221 **3**). However, it should be noted that the vast majority of CAZyme families were found in all sample
222 types (**Fig 3**), indicating that there exists a set of CAZymes which are present across ruminant species
223 consuming different diets and living in vastly different conditions.

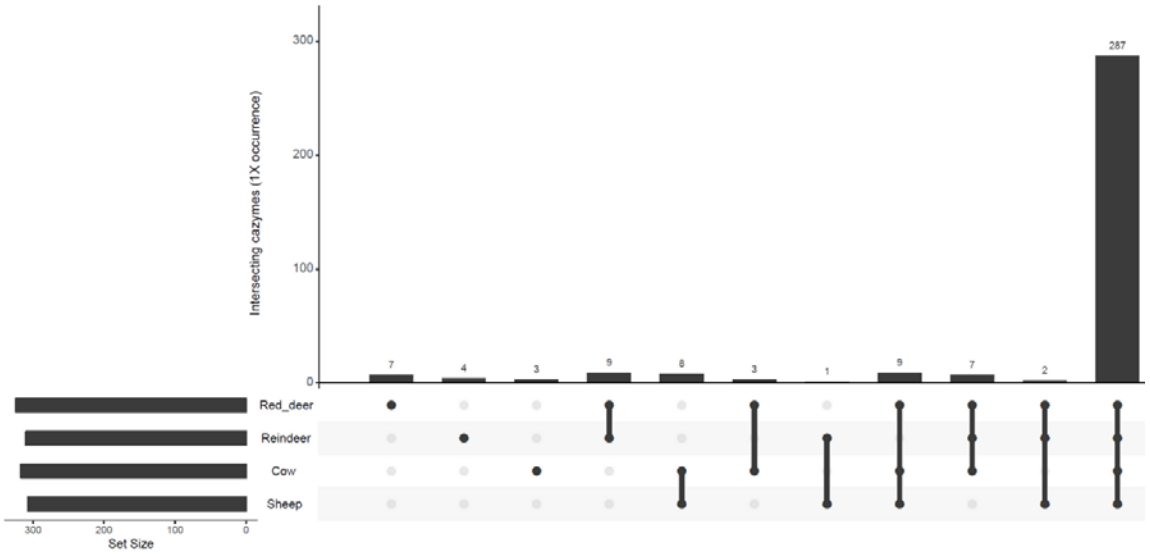


Fig 3: UpSetR graph showing the number of shared CAZyme families at average 1X coverage within four ruminant species.

DeSeq2 was used to identify specific CAZymes which were significantly more abundant in one ruminant species vs another (**Additional file 4**). Those CAZymes which were consistently more abundant in specific species when compared to other species are listed in **Tables 1-4**.

231 **Table 1: CAZymes (Carbohydrate-active enzymes) which were consistently more abundant in**

232 **cattle.**

CAZyme	**Deer - adjusted p-value**	**Sheep - adjusted p-value**	**Reindeer - adjusted p-value**
Carbohydrate-binding modules			
CBM11	0.000113	1.81E-05	8.41E-05
CBM22	0.000996	0.000958	0.004915
CBM25	0.000571	0.010859	0.005573
CBM6	4.32E-11	0.02129	1.91E-25
CBM74	4.51E-06	0.009696	0.00389
Carbohydrate esterases			
CE15	1.03E-07	0.000403	0.000357
CE6	6.57E-13	0.000634	1.07E-27
Glycoside hydrolases			
GH105	8.55E-10	0.0195	1.22E-14
GH11	0.000134	0.000687	2.24E-05
GH115	0.000737	0.018632	1.91E-17
GH13_28	7.54E-06	0.012113	0.042438
GH130	1.02E-08	2.69E-06	0.02559
GH146	0.011892	0.032424	3.57E-06
GH31	4.20E-08	0.000243	7.31E-10
GH35	3.11E-06	0.003975	5.19E-10
GH36	0.001473	0.002527	1.22E-13
GH43_1	5.48E-12	6.03E-05	2.36E-22

GH43_24	3.66E-06	0.000849	0.022236
GH43_29	5.69E-10	0.000779	3.84E-65
GH43_35	1.76E-06	0.018491	4.78E-29
GH43_5	4.79E-16	0.011903	2.98E-40
GH43_7	2.36E-09	0.012453	1.60E-36
GH45	1.13E-05	0.000541	0.001412
GH5_10	9.78E-05	0.000118	0.000164
GH5_38	1.36E-11	0.022559	1.78E-37
GH5_39	6.13E-05	0.000849	0.02559
GH5_52	0.00487	0.015377	0.003089
GH51	6.44E-08	0.011587	2.36E-12
GH67	1.50E-11	0.001929	1.45E-31
GH8	4.01E-10	0.000212	2.96E-31
GH9	1.12E-06	1.08E-07	0.002527
GH94	1.45E-05	5.96E-13	0.000853

233

234 **Table 2: CAZymes (Carbohydrate-active enzymes) which were consistently more abundant in**

235 **sheep.**

CAZyme	**Cow - adjusted p-value**	**Deer - adjusted p-value**	**Reindeer - adjusted p-value**
Carbohydrate-binding modules			
CBM27	0.019613	1.62E-07	1.13E-05
Carbohydrate esterases			
CE8	0.021156	1.65E-08	4.29E-14
Glycoside hydrolases			
GH144	6.75E-05	0.003343	0.034267
GH30_8	0.026986	0.0002685	2.49E-32
GH53	0.000983	6.98E-14	9.02E-14
Glycosyltransferases			
GT69	7.60E-05	5.21E-05	4.40E-06
GT7	3.50E-08	0.0002904	0.000314
Polysaccharide lyases			
PL10	0.022559	2.52E-07	3.35E-22

236

237 **Table 3: CAZymes (Carbohydrate-active enzymes) which were consistently more abundant in red**

238 **deer.**

CAZyme	Cow - adjusted p-value	Reindeer - adjusted p-value	Sheep - adjusted p-value
Carbohydrate-binding modules			
CBM34	2.18E-22	3.69E-13	0.049585
CBM42	0.005095	0.003958	0.000579
CBM54	2.87E-14	4.60E-07	0.000154
CBM58	5.65E-05	0.00122	0.000348
Carbohydrate esterases			
CE13	0.002535	5.30E-06	0.006324
Glycoside hydrolases			
GH13_20	3.56E-09	0.014494	0.021084
GH13_29	2.22E-40	4.90E-08	0.006428
GH13_4	9.39E-08	0.000614	0.000579
GH147	0.00095	0.000495	0.000842
GH148	5.24E-10	1.92E-08	0.0105
GH24	5.09E-23	0.00023	1.24E-05
GH43	0.010658	0.000118	0.024375
GH43_11	4.01E-10	0.004229	0.024919
GH43_21	3.74E-06	0.004636	0.000375
GH5_44	8.22E-23	0.000382	0.001021
Glycosyltransferases			
GT23	2.31E-11	0.00713	0.036234
Polysaccharide lyases			

PL4	4.20E-69	1.94E-19	2.12E-44

239

240 **Table 4: CAZymes (Carbohydrate-active enzymes) which were consistently more abundant in**

241 **reindeer.**

CAZyme	**Cow - adjusted p-value**	**Deer - adjusted p-value**	**Sheep - adjusted p-value**
Carbohydrate-binding modules			
CBM32	4.15E-29	8.21E-08	4.63E-11
CBM41	0.00392	1.73E-05	1.71E-11
CBM62	5.74E-06	1.52E-05	1.95E-09
CBM66	2.12E-29	1.66E-13	3.43E-11
CBM67	6.33E-11	5.31E-05	0.000125
CBM68	5.01E-05	8.99E-06	0.000182
CBM9	8.88E-13	1.44E-11	2.22E-12
Carbohydrate esterases			
CE3	1.68E-10	7.47E-05	0.00062
CE9	5.83E-19	7.37E-05	0.007305
Glycoside hydrolases			
GH109	1.80E-18	9.23E-08	1.92E-13
GH117	1.25E-05	0.004221	0.003013
GH123	0.003063	0.000606	0.001734
GH125	4.14E-19	1.33E-16	1.59E-15
GH128	2.16E-14	9.97E-07	4.18E-05
GH13	0.004306	1.47E-07	0.007748

GH13_16	7.00E-05	0.00063	3.51E-05
GH13_2	0.005224	0.0006	0.026906
GH13_36	0.002131	0.003061	0.027897
GH13_40	0.000141	1.57E-05	4.25E-07
GH133	0.030687	1.47E-07	0.000337
GH16	7.62E-10	6.74E-06	0.000368
GH18	1.28E-07	2.39E-07	0.000111
GH20	6.60E-12	0.000614	1.03E-05
GH38	9.43E-46	1.51E-18	3.04E-43
GH39	2.69E-07	0.000556	1.43E-10
GH43_26	1.57E-05	0.000556	0.000556
GH43_28	1.35E-16	3.00E-08	4.40E-06
GH43_3	5.69E-08	7.45E-07	8.84E-19
GH43_31	5.43E-07	0.000583	2.26E-07
GH43_33	1.58E-06	2.24E-05	1.01E-13
GH5	1.57E-22	1.80E-08	1.55E-15
GH5_35	8.84E-13	0.001397	4.49E-06
GH57	0.003543	0.005297	9.37E-05
GH59	0.000148	0.003958	0.008335
GH64	1.91E-09	1.87E-07	2.57E-09
GH76	2.09E-20	1.55E-27	3.29E-16
GH85	0.001061	0.002853	0.008951
GH87	4.47E-10	2.62E-07	1.42E-07
GH88	0.001577	0.00356	0.006143
GH92	8.02E-20	3.58E-16	8.82E-23

GH93	1.65E-08	4.49E-06	3.88E-11
GH99	2.06E-07	2.73E-09	0.000163
Glycosyltransferases			
GT1	1.86E-05	1.86E-10	1.93E-05
GT10	2.68E-08	0.012407	0.005445
GT2_Glyco_tranf_2_4	5.94E-21	0.000115	2.57E-07
GT3	0.008785	0.001394	0.001478
GT39	0.002494	0.021868	0.003876
GT4	7.61E-09	0.000752	0.018262
GT47	0.000453	0.011114	0.002988
GT66	2.68E-06	5.25E-05	0.008382
GT70	9.38E-07	0.000433	0.001094
GT77	0.002527	0.014307	0.009975
GT8	5.78E-21	0.000151	0.02682
GT81	1.55E-06	8.26E-06	0.025595
GT83	4.25E-39	4.13E-05	1.77E-06
Polysaccharide lyases			
PL17_2	2.55E-05	0.000285	0.00041
PL22	1.27E-06	0.005756	0.000133
PL8	0.003063	0.002491	0.006191

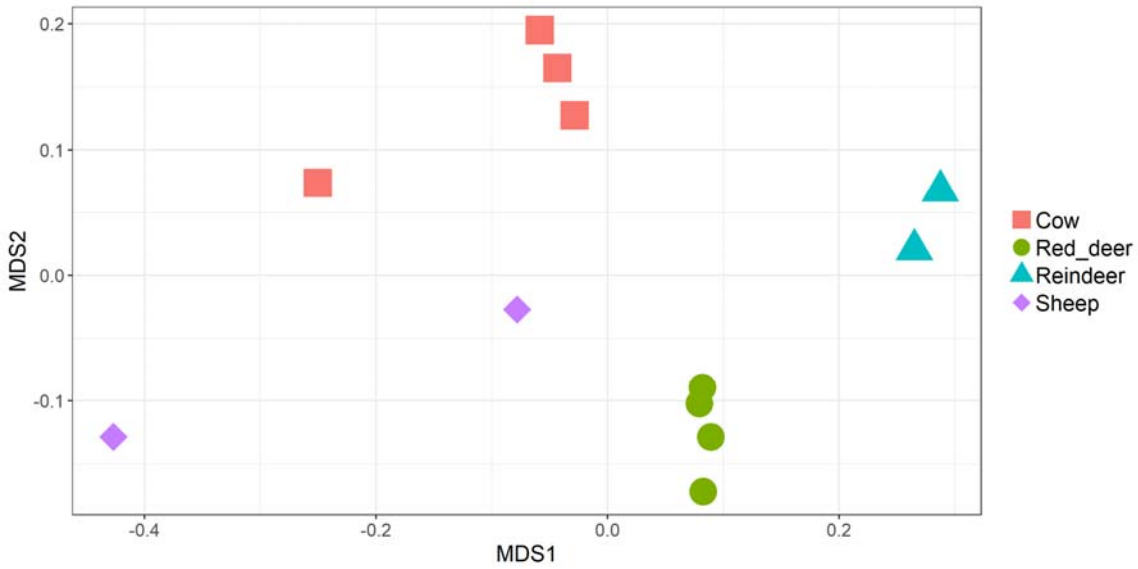
242

243 CAZymes are often found organised into Polysaccharide Utilization Loci (PUL) which comprise a set
244 of genes that enable the binding and degradation of specific carbohydrates or multiple
245 carbohydrates. We used the software PULpy to predict PULs which were present in our
246 *Bacteroidales* RUGs. Of the 136 RUGs which belong to the taxonomy *Bacteroidales*, 112 contain

247 putative PULs. Within these RUGs we identified 970 PULs, with numbers of PULs per RUG ranging
248 from 1 to 35. The largest quantity of PULs originating from one RUG was 35 from uncultured
249 *Bacteroidales sp. RUG30227*; these encoded a wide range of CAZymes. This RUG was more abundant
250 in reindeer samples than samples from other ruminants. Of the 970 PULs, 332 of these were a single
251 susC/D pair. A summary of identified PULs can be found in **Additional file 5** (**Additional File 1: Fig 4**).

252 We also examined the abundance of genes which belonged to specific KEGG orthologs. KEGG
253 orthologs represent a wide range of molecular functions and are defined by a network-based
254 classification. We found that, as for CAZymes, ruminant species clustered significantly by the
255 abundance of genes with specific KEGG orthologs (PERMANOVA: P = 1e-05, **Fig 4**) and that the vast
256 majority of orthologs were found in all ruminant species (**Fig 5**). However, the large amount of
257 orthologs (n=729) which were only found in the two domesticated species (cattle and sheep) is also
258 worthy of note. It should also be noted that the two sheep samples did not cluster visually to the
259 same extent as the samples originating from the other ruminant speices (**Fig 4**). DeSeq2 was used to
260 identify many KEGG orthologs which were significantly more abundant in one ruminant species vs
261 another (**Additional file 6**). Those orthologs which were consistently more abundant in specific
262 ruminant species (Adjusted p-value <0.05) are listed in **Additional file 7**.

263

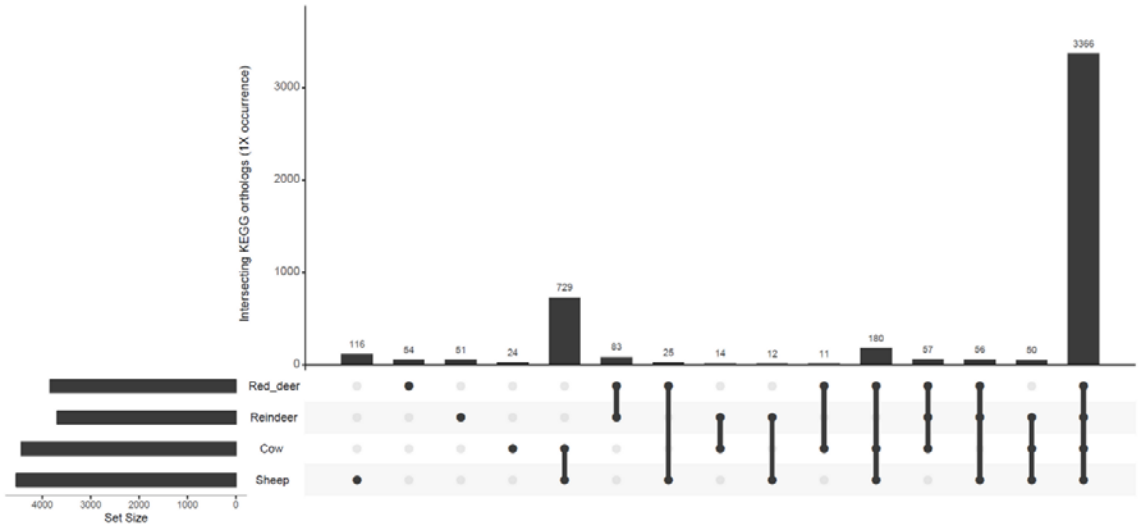


264

265 **Fig 4: NMDS of ruminal samples clustered by abundance of KEGG orthologs, using Bray–Curtis**

266 **dissimilarity values (PERMANOVA; P = 1e-05).**

267



268

269 **Fig 5: UpSetR graph showing the number of shared KEGG orthologs families at average 1X**

270 **coverage within four ruminant species.**

Discussion:

The rumen microbiota plays a crucial role in the ability of ruminants to efficiently digest feed while the rumen microbiota and their products also have a potential use in diverse industrial applications. The ruminal microbiota of red deer and reindeer have previously been studied using 16S rRNA gene sequencing (51-54). However, metagenomic studies of these species are limited, with only one study in reindeer (55) and no studies in red deer.

In this study we constructed 391 rumen microbial genomes from metagenomic data from cattle, sheep, red deer and reindeer. We assigned taxonomies to our RUGs using GTDB-Tk rather than NCBI based taxonomies as this improves the classification of uncultured bacteria due to the use of a genome-based taxonomy (32). We have also previously found less need to manually correct taxonomic assignments when using GTDB-Tk (21). Our microbes predominantly belonged to the *Bacteroidota* and *Firmicutes_A*, with lesser numbers of 14 other phyla. We dereplicated our genomes alongside a superset of rumen bacterial genomes (20) and used the results output by GTDB-Tk to identify RUGs which represent novel microbial strains and species. Amongst our genomes we identified 372 novel strains and 279 novel species. These microbes were taxonomically diverse, belonging to 15 phyla. Only 31 RUGs were assigned an identity at species level.

The vast majority of our total RUGs were only present on average at ≥1 coverage in one ruminant species. However, we found that at higher taxonomic levels taxonomies were shared between sample types. When comparing the abundance of taxonomies between samples we found that ruminant species clustered separately by both higher (kingdom and phylum) and lower (family and genus) taxonomic levels. We are aware that the sample sizes for our study are small and therefore any conclusions about differences between the microbiota of ruminant species should be drawn cautiously. However, our data are supportive of the hypothesis that there are host species-specific rumen microbiota at the strain and species level but that these differences do not necessarily translate into large differences in the types of CAZymes expressed.

296 While we found that there were significant differences between the abundances of CAZymes
297 between ruminant species, most CAZymes were present in all ruminant species. These results also
298 reflected those which we found when analysing the abundance of KEGG orthologs. We also
299 identified 970 PULs in our *Bacteriodales* RUGs, with numbers of PULs per genome ranging from 1 to
300 35. The RUG containing 35 PULs was found most abundantly in reindeer samples, emphasising the
301 potential for the discovery of novel carbohydrate-active enzymes in lesser studied ruminant species,
302 as also highlighted by a previous study which identified multiple PULs in metagenomic samples from
303 reindeer (55). Unfortunately due to the nature of our samples, with red deer and reindeer samples
304 originating from animals eating a non-regimented diet, we are not able to provide metadata as to
305 the exact nutritional composition of our animals' diets, therefore a more in depth analysis of dietary
306 carbohydrates vs CAZyme/PUL abundance is not possible.

307 While several thousand RUGs have previously been published that originate from the rumen
308 microbiota, the vast majority of these originate from cattle. By investing more effort in exploring the
309 metagenome of less well studied ruminants we will be able to identify even more microbes and
310 microbial products that are of industrial-interest. In conclusion, we present a dataset of RUGs from
311 four ruminant species which can be used as a reference dataset in future metagenomic studies and
312 to aid in the design of culture based studies.

313

314 **Author statements:**

315 **Authors and contributors:**

316 LG contributed to methodology, formal analysis, data curation, writing (original draft preparation)
317 and visualisation. BG and RJW contributed to conceptualisation, methodology, investigation,
318 resources and writing (review and editing). RJW also contributed to supervision and project
319 administration. MW contributed to conceptualisation, methodology, formal analysis, writing (review

and editing), visualisation, supervision and project administration. All authors read and approved the final manuscript.

Conflicts of interest:

The authors declare that there are no conflicts of interest

Funding information:

The Roslin Institute forms part of the Royal (Dick) School of Veterinary Studies, University of Edinburgh. This project was supported by the Biotechnology and Biological Sciences Research Council, including institute strategic programme and national capability awards to The Roslin Institute (BBSRC: BB/P013759/1, BB/P013732/1, BB/J004235/1, BB/J004243/1), and the Technology Strategy Board (TS/J000108/1, TS/J000116/1). The Rowett Institute is funded by the Rural and Environment Science and Analytical Services Division (RESAS) of the Scottish Government. The funding bodies had no role in the study design, the collection, analysis, and interpretation of data or the writing of the manuscript. Sequencing was carried out by Edinburgh Genomics.

Ethical approval:

Cattle projects were carried out under Home Office PPL 30/2579. Sheep experimentation was carried out under the conditions set out by UK Home Office licence no. 604028, procedure reference number 8.

Consent for publication:

342 Not applicable.

343

344 **Acknowledgments:**

345 BG thanks the Higher Education Council of the Turkish Republic for the award of a travelling
346 fellowship. We thank Bob Mayes and Dave Hamilton of the James Hutton Institute for their
347 permission and help in sampling the sheep digesta, Kevin Shingfield for the provision of reindeer
348 digesta samples, and Euan Munro for red deer digesta samples. We also thank Nest McKain of RI for
349 technical assistance.

350

351 **References**:

352 1. Lima J, Auffret MD, Stewart RD, Dewhurst RJ, Duthie CA, Snelling TJ, et al. Identification of
353 rumen microbial genes involved in pathways linked to appetite, growth, and feed conversion
354 efficiency in cattle. Front Genet. 2019;**10**: 701 doi: 10.3389/fgene.2019.00701
355 2. Ben Shabat SK, Sasson G, Doron-Faigenboim A, Durman T, Yaacoby S, Miller MEB, et al.
356 Specific microbiome-dependent mechanisms underlie the energy harvest efficiency of ruminants.
357 Isme J. 2016;**10**: 2958-72 doi: 10.1038/ismej.2016.62
358 3. Patil RD, Ellison MJ, Wolff SM, Shearer C, Wright AM, Cockrum RR, et al. Poor feed efficiency
359 in sheep is associated with several structural abnormalities in the community metabolic network of
360 their ruminal microbes. J Anim Sci. 2018;**96**: 2113-24 doi: 10.1093/jas/sky096
361 4. Jami E, White BA, Mizrahi I. Potential role of the bovine rumen microbiome in modulating
362 milk composition and feed efficiency. PLoS One. 2014;**9**: e85423 doi: 10.1371/journal.pone.0085423
363 5. Scharen M, Frahm J, Kersten S, Meyer U, Hummel J, Breves G, et al. Interrelations between
364 the rumen microbiota and production, behavioral, rumen fermentation, metabolic, and
365 immunological attributes of dairy cows. J Dairy Sci. 2018;**101**: 4615-37 doi: 10.3168/jds.2017-13736

366 6. Ribeiro GO, Gruninger RJ, Badhan A, McAllister TA. Mining the rumen for fibrolytic feed
367 enzymes. Anim Front. 2016;**6**: 20-6 doi: 10.2527/af.2016-0019
368 7. Huws SA, Creevey CJ, Oyama LB, Mizrahi I, Denman SE, Popova M, et al. Addressing global
369 ruminant agricultural challenges through understanding the rumen microbiome: Past, present, and
370 future. Front Microbiol. 2018;**9**: 33 doi: 10.3389/fmicb.2018.02161
371 8. Solbak AI, Richardson TH, McCann RT, Kline KA, Bartnek F, Tomlinson G, et al. Discovery of
372 pectin-degrading enzymes and directed evolution of a novel pectate lyase for processing cotton
373 fabric. J Biol Chem. 2005;**280**: 9431-8 doi: 10.1074/jbc.M411838200
374 9. Singh B, Gautam SK, Verma V, Kumar M, Singh B. Metagenomics in animal gastrointestinal
375 ecosystem: Potential biotechnological prospects. Anaerobe. 2008;**14**: 138-44 doi:
376 10.1016/j.anaerobe.2008.03.002
377 10. Ufarte L, Laville E, Duquesne S, Morgavi D, Robe P, Klopp C, et al. Discovery of carbamate
378 degrading enzymes by functional metagenomics. PLoS One. 2017;**12**: e0189201 doi:
379 10.1371/journal.pone.0189201
380 11. Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, et al. Metagenomic discovery
381 of biomass-degrading genes and genomes from cow rumen. Science. 2011;**331**: 463-7 doi:
382 10.1126/science.1200387
383 12. Wallace RJ, Rooke JA, McKain N, Duthie CA, Hyslop JJ, Ross DW, et al. The rumen microbial
384 metagenome associated with high methane production in cattle. BMC Genomics. 2015;**16**: 839 doi:
385 10.1186/s12864-015-2032-0
386 13. Difford GF, Plichta DR, Lovendahl P, Lassen J, Noel SJ, Hojberg O, et al. Host genetics and the
387 rumen microbiome jointly associate with methane emissions in dairy cows. PLoS Genet. 2018;**14**:
388 e1007580 doi: 10.1371/journal.pgen.1007580
389 14. Shi WB, Moon CD, Leahy SC, Kang DW, Froula J, Kittelmann S, et al. Methane yield
390 phenotypes linked to differential gene expression in the sheep rumen microbiome. Genome Res.
391 2014;**24**: 1517-25 doi: 10.1101/gr.168245.113

392 15. Auffret MD, Stewart R, Dewhurst RJ, Duthie CA, Rooke JA, Wallace RJ, et al. Identification,
393 comparison, and validation of robust rumen microbial biomarkers for methane emissions using
394 diverse *Bos taurus* breeds and basal diets. Front Microbiol. 2018;**8**: 2642 doi:
395 10.3389/fmicb.2017.02642
396 16. Roehe R, Dewhurst RJ, Duthie CA, Rooke JA, McKain N, Ross DW, et al. Bovine host genetic
397 variation influences rumen microbial methane production with best selection criterion for low
398 methane emitting and efficiently feed converting hosts based on metagenomic gene abundance.
399 PLoS Genet. 2016;**12**: 20 doi: 10.1371/journal.pgen.1005846
400 17. Seshadri R, Leahy SC, Attwood GT, Teh KH, Lambie SC, Cookson AL, et al. Cultivation and
401 sequencing of rumen microbiome members from the Hungate1000 Collection. Nat Biotechnol.
402 2018;**36**: 359-67 doi: 10.1038/nbt.4110
403 18. Creevey CJ, Kelly WJ, Henderson G, Leahy SC. Determining the culturability of the rumen
404 bacterial microbiome. Microb Biotechnol. 2014;**7**: 467-79 doi: 10.1111/1751-7915.12141
405 19. Stewart RD, Auffret MD, Warr A, Wiser AH, Press MO, Langford KW, et al. Assembly of 913
406 microbial genomes from metagenomic sequencing of the cow rumen. Nat Commun. 2018;**9**: 870
407 doi: 10.1038/s41467-018-03317-6
408 20. Stewart RD, Auffret MD, Warr A, Walker AW, Roehe R, Watson M. Compendium of 4,941
409 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. Nat
410 Biotechnol. 2019;**37**: 953-61 doi: 10.1038/s41587-019-0202-3
411 21. Glendinning L, Stewart RD, Pallen MJ, Watson KA, Watson M. Assembly of hundreds of novel
412 bacterial genomes from the chicken caecum. Genome Biol. 2020;**21**: 16 doi: 10.1186/s13059-020-
413 1947-1
414 22. McKain N, Genc B, Snelling TJ, Wallace RJ. Differential recovery of bacterial and archaeal 16S
415 rRNA genes from ruminal digesta in response to glycerol as cryoprotectant. J Microbiol Methods.
416 2013;**95**: 381-3 doi: 10.1016/j.mimet.2013.10.009

417 23. Yu ZT, Morrison M. Improved extraction of PCR-quality community DNA from digesta and
418 fecal samples. Biotechniques. 2004;**36**: 808-12 doi: 10.2144/04365st04
419 24. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data.
420 Bioinformatics. 2014;**30**: 2114-20 doi: 10.1093/bioinformatics/btu170
421 25. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: a de novo assembler for single-cell and
422 metagenomic sequencing data with highly uneven depth. Bioinformatics. 2012;**28**: 1420-8 doi:
423 10.1093/bioinformatics/bts174
424 26. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013
425 arXiv:13033997
426 27. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
427 Alignment/Map format and SAMtools. Bioinformatics. 2009;**25**: 2078-9 doi:
428 10.1093/bioinformatics/btp352
429 28. Kang DWD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately
430 reconstructing single genomes from complex microbial communities. PeerJ. 2015;**3**: 15 doi:
431 10.7717/peerj.1165
432 29. Li DH, Liu CM, Luo RB, Sadakane K, Lam TW. MEGAHIT: An ultra-fast single-node solution for
433 large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;**31**:
434 1674-6 doi: 10.1093/bioinformatics/btv033
435 30. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic
436 comparisons that enables improved genome recovery from metagenomes through de-replication.
437 Isme J. 2017;**11**: 2864-8 doi: 10.1038/ismej.2017.126
438 31. Stewart RD, Watson M, Auffret MD, Roehe R, Snelling TJ. MAGpy: A reproducible pipeline for
439 the downstream analysis of metagenome-assembled genomes (MAGs). Bioinformatics. 2018;**35**:
440 2150–2 doi: 10.1093/bioinformatics/bty905

441 32. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, et al. A
442 standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life.
443 Nat Biotechnol. 2018;**36**: 996-1004 doi: 10.1038/nbt.4229
444 33. Rambaut A. FigTree v1. 4. http://treebioedacuk/software/figtree20122012.
445 34. Asnicar F, Weingart G, Tickle TL, Huttenhower C, Segata N. Compact graphical
446 representation of phylogenetic data and metadata with GraPhlAn. PeerJ. 2015;**3**: e1029 doi:
447 10.7717/peerj.1029
448 35. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-
449 Active EnZymes database (CAZy): An expert resource for glycogenomics. Nucleic Acids Res. 2009;**37**:
450 D233-D8 doi: 10.1093/nar/gkn663
451 36. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res.
452 2000;**28**: 27-30 doi: 10.1093/nar/28.1.27
453 37. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat
454 Methods. 2015;**12**: 59-60 doi: 10.1038/nmeth.3176
455 38. Stewart RD, Auffret MD, Roehe R, Watson M. Open prediction of polysaccharide utilisation
456 loci (PUL) in 5414 public *Bacteroidetes* genomes using PULpy. preprint in bioRxiv. 2018: 421024 doi:
457 10.1101/421024
458 39. Wickham H. ggplot2: elegant graphics for data analysis. http://ggplot2org20162016.
459 40. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, Minchin DMPR, et al. vegan:
460 Community Ecology Package. https://CRANR-projectorg/package=vegan20132018.
461 41. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-
462 seq data with DESeq2. Genome Biol. 2014;**15**: 550 doi: 10.1186/s13059-014-0550-8
463 42. Conway JR, Lex A, Gehlenborg N. UpSetR: An R package for the visualization of intersecting
464 sets and their properties. Bioinformatics. 2017;**33**: 2938-40 doi: 10.1093/bioinformatics/btx364
465 43. Wood DE, Salzberg SL. Kraken: Ultrafast metagenomic sequence classification using exact
466 alignments. Genome Biol. 2014;**15**: R46 doi: 10.1186/gb-2014-15-3-r46

467 44. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al.
468 Minimum information about a single amplified genome (MISAG) and a metagenome-assembled
469 genome (MIMAG) of bacteria and archaea. Nat Biotechnol. 2017;**35**: 725-31 doi: 10.1038/nbt.3893
470 45. Parks DH, Rinke C, Chuvochina M, Chaumeil PA, Woodcroft BJ, Evans PN, et al. Recovery of
471 nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol.
472 2017;**2**: 1533-42 doi: 10.1038/s41564-017-0012-7
473 46. Solden LM, Naas AE, Roux S, Daly RA, Collins WB, Nicora CD, et al. Interspecies cross-feeding
474 orchestrates carbon degradation in the rumen ecosystem. Nat Microbiol. 2018;**3**: 1274-84 doi:
475 10.1038/s41564-018-0225-4
476 47. Svartstrom O, Alneberg J, Terrapon N, Lombard V, de Bruijn I, Malmsten J, et al. Ninety-nine
477 de novo assembled genomes from the moose (*Alces alces*) rumen microbiome provide new insights
478 into microbial plant biomass degradation. Isme J. 2017;**11**: 2538-51 doi: 10.1038/ismej.2017.108
479 48. Neumann AP, Suen G. The phylogenomic diversity of herbivore-associated *Fibrobacter* spp.
480 is correlated to lignocellulose-degrading potential. mSphere. 2018;**3**: e00593-18 doi:
481 10.1128/mSphere.00593-18
482 49. Shoseyov O, Shani Z, Levy I. Carbohydrate binding modules: Biochemical properties and
483 novel applications. Microbiol Mol Biol Rev. 2006;**70**: 283-95 doi: 10.1128/mmbr.00028-05
484 50. Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. Expansion of the enzymatic
485 repertoire of the CAZy database to integrate auxiliary redox enzymes. Biotechnol Biofuels. 2013;**6**:
486 41 doi: 10.1186/1754-6834-6-41
487 51. Ostbye K, Wilson R, Rudi K. Rumen microbiota for wild boreal cervids living in the same
488 habitat. FEMS Microbiol Lett. 2016;**363**: fnw233 doi: 10.1093/femsle/fnw233
489 52. Salgado-Flores A, Hagen LH, Ishaq SL, Zamanzadeh M, Wright ADG, Pope PB, et al. Rumen
490 and cecum microbiomes in reindeer (*Rangifer tarandus tarandus*) Are changed in response to a
491 lichen diet and may affect enteric methane emissions. PLoS One. 2016;**11**: e0155213 doi:
492 10.1371/journal.pone.0155213

493 53. Sundset M, Edwards J, Cheng Y, Senosiain R, Fraile M, Northwood K, et al. Molecular
494 diversity of the rumen microbiome of Norwegian reindeer on natural summer pasture. Microb Ecol.
495 2009;**57**: 335-48 doi: 10.1007/s00248-008-9414-7
496 54. Qian WX, Ao WP, Jia CH, Li ZP. Bacterial colonisation of reeds and cottonseed hulls in the
497 rumen of Tarim red deer (*Cervus elaphus yarkandensis*). Antonie Van Leeuwenhoek. 2019;**112**: 1283-
498 96 doi: 10.1007/s10482-019-01260-0
499 55. Pope PB, Mackenzie AK, Gregor I, Smith W, Sundset MA, McHardy AC, et al. Metagenomics
500 of the Svalbard reindeer rumen microbiome reveals abundance of polysaccharide utilization loci.
501 PLoS One. 2012;**7**: e38571 doi: 10.1371/journal.pone.0038571

502

Supporting data:

Additional file 1.docx: S1 Figures - Supplementary figures 1 to 4.

Additional file 2.xls: Dataset 1 - Average coverage of RUGs in all samples. Coverage was calculated by mapping the RUG scaffolds to adaptor-trimmed Illumina reads.

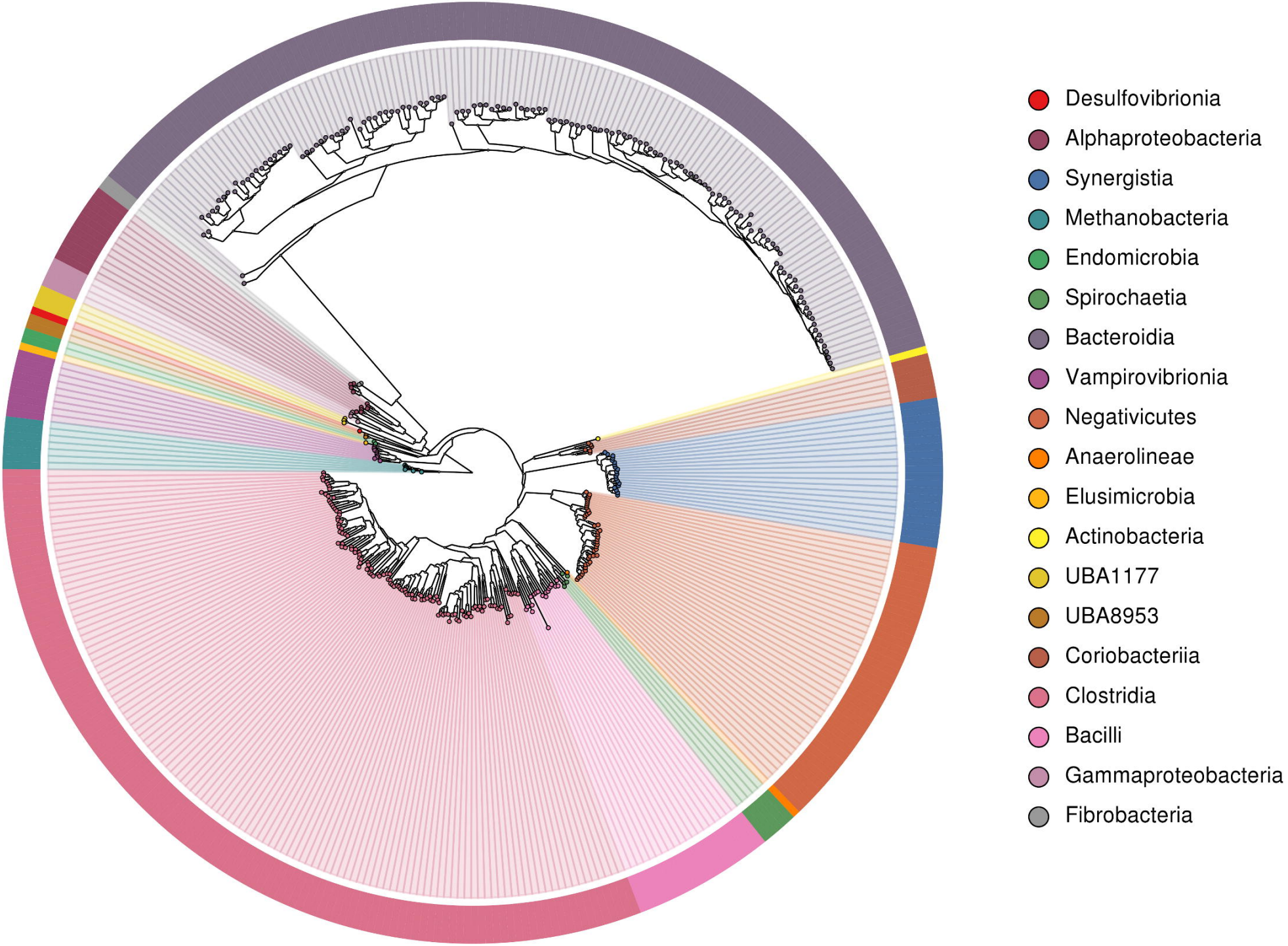
Additional file 3.xls: Dataset 2 - Description of each RUG, including taxonomy, novelty, genome completeness, genome contamination, genome size, number of ambiguous bases, number of scaffolds, number of contigs, N50 (scaffolds), N50 (contigs), mean scaffold length, mean contig length, longest contig and GC content.

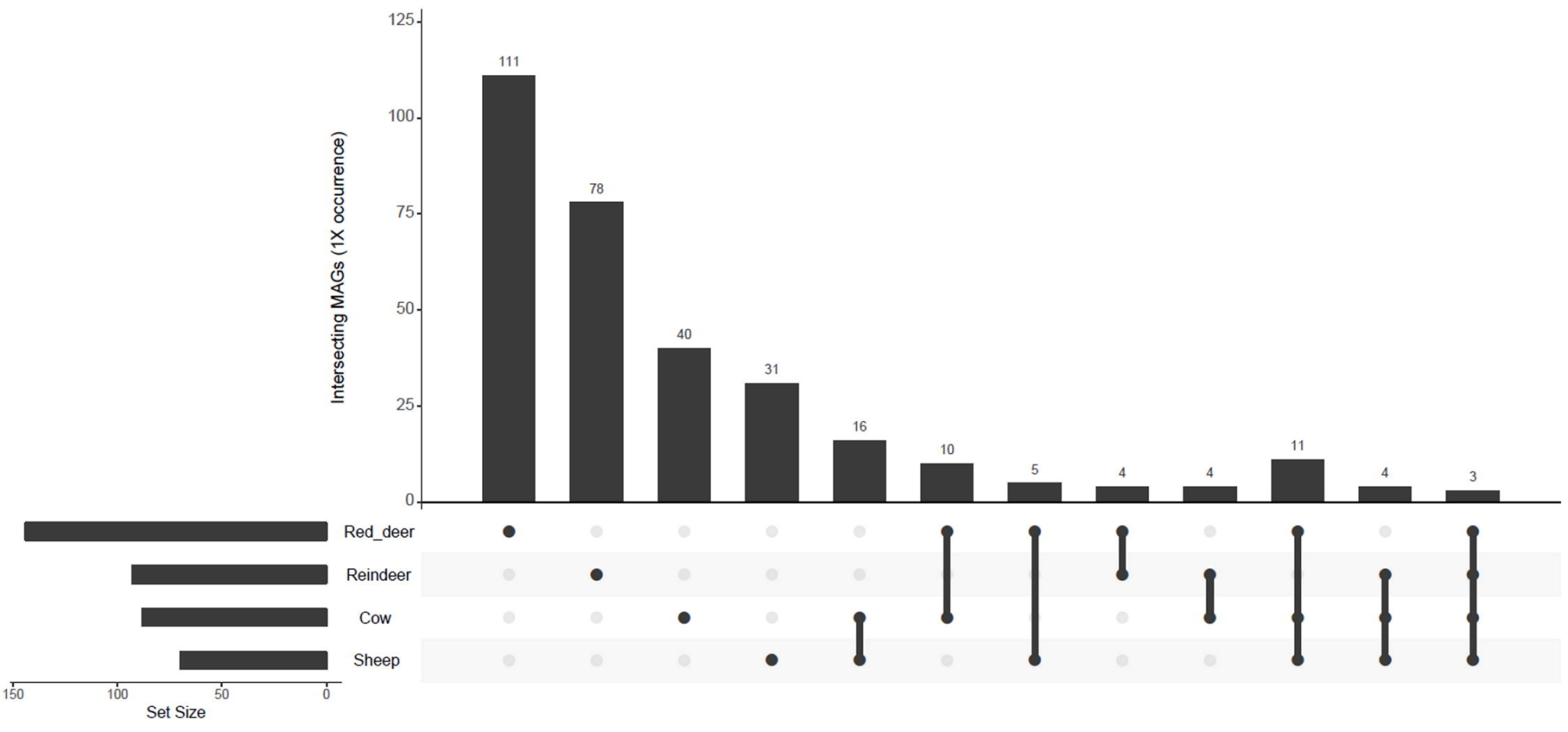
Additional file 4.xls: Dataset 3 - CAZymes which were more abundant in sheep, cattle, reindeer or red deer (DeSeq2: Adjusted p-value <0.05).

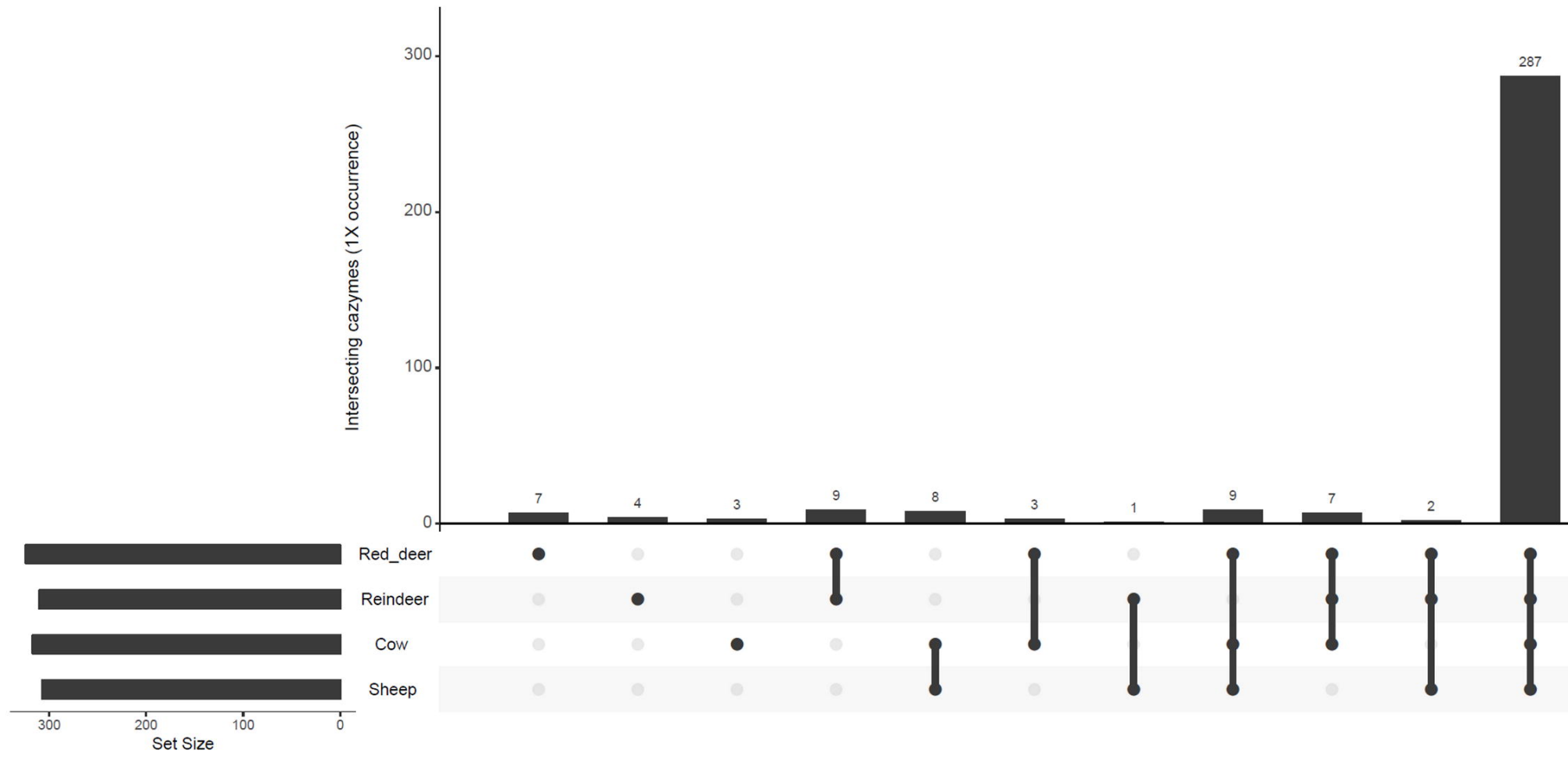
Additional file 5: Dataset 4 - Summary of polysaccharide utilisation loci found in *Bacteroidales* RUGs.

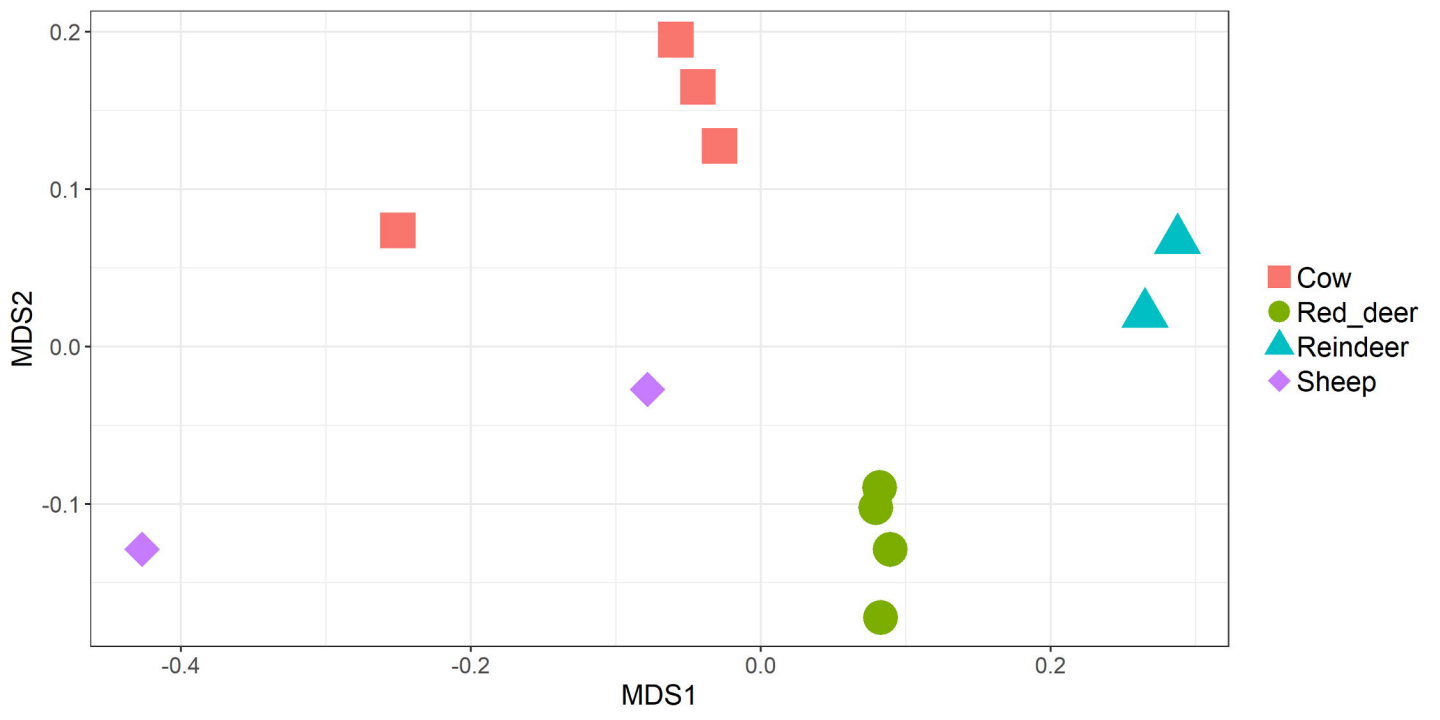
Additional file 6: Dataset 5 - KEGG orthologs which were more abundant in sheep, cattle, reindeer or red deer (DeSeq2: Adjusted p-value <0.05).

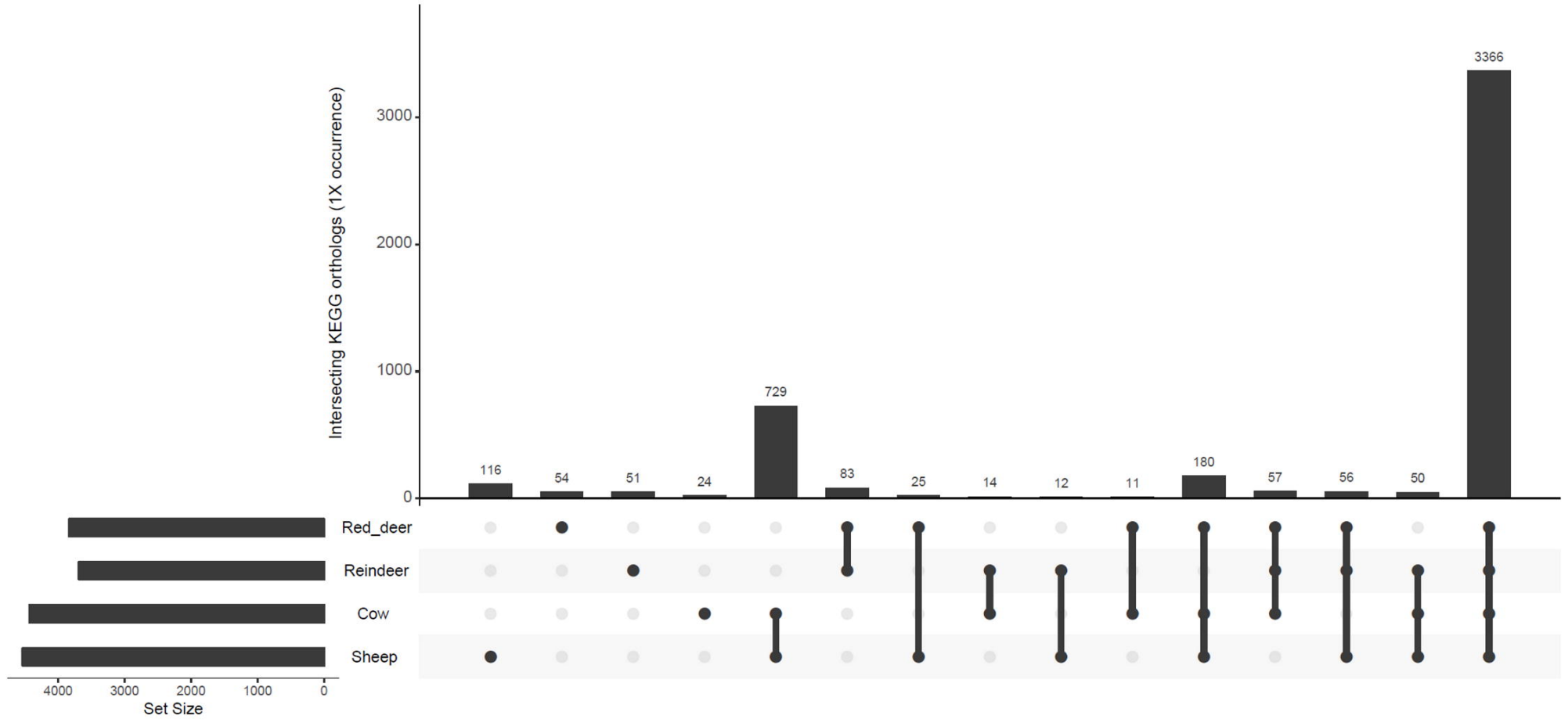
Additional file 7: Dataset 6 - KEGG orthologs which were consistently more abundant in sheep, cattle, reindeer or red deer (DeSeq2: Adjusted p-value <0.05).


Desulfovibrionia
Alphaproteobacteria
Synergistia
Methanobacteria
Endomicrobia
Spirochaetia
Bacteroidia
Vampirovibrionia
Negativicutes
Anaerolineae
Elusimicrobia
Actinobacteria
UBA1177
UBA8953
Coriobacteriia
Clostridia
Bacilli
Gammaproteobacteria
Fibrobacteria


Intersecting MAGs (1X occurrence)
125
100
75
50
25
0
111
78
40
31
16
10
5
4
4
11
4
3
Red_deer
Reindeer
Cow
Sheep
150
100
50
0
Set Size


Intersecting cazymes (1X occurrence)
300
200
100
0
287
7
4
3
9
8
3
1
9
7
2
Red_deer
Reindeer
Cow
Sheep
300
200
100
0
Set Size


0.2
0.1
0.0
-0.1
MDS2
-0.4
-0.2
0.0
0.2
MDS1
Cow
Red_deer
Reindeer
Sheep


Intersecting KEGG orthologs (1X occurrence)
3000
2000
1000
0
3366
729
116
54
51
24
83
25
14
12
11
180
57
56
50
Red_deer
Reindeer
Cow
Sheep
4000
3000
2000
1000
0
Set Size