The metabolic influence of the core ciliate *Entodinium caudatum* within the rumen microbiome

T.O. Andersen[1], I. Altshuler[1], A.V.P. de Leon[2], J. Walter[2] E. McGovern[3], K. Keogh[3], C. Martin[4], L. Bernard[4], H. Fougère[4], D.P. Morgavi[4], T. Park[5,6], J.L. Firkins[6], Z. Yu[6], T.R. Hvidsten[2], S.M. Waters[3], M. Popova[4], M.Ø. Arntzen[2], L.H. Hagen[2], P.B. Pope[1,2*]

[1] Department of Animal and Aquacultural Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, Ås, Norway

[2] Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway

[3] Teagasc, Animal and Bioscience Research Department, Animal and Grassland Research and Innovation Centre, Teagasc, Grange, Dunsany, County Meath, Ireland

[4] Université Clermont Auvergne, INRAE, VetAgro Sup, UMR Herbivores, F-63122 Saint-Genes-Champanelle, France

[5] Department of Animal Science and Technology, Chung-Ang University, Anseong-si, Gyeonggi-do 17546, Republic of Korea

[6] Department of Animal Sciences, The Ohio State University, Columbus, OH, United States

* Correspondence: phil.pope@nmbu.no

ABSTRACT

Protozoa comprise a major fraction of the microbial biomass in the rumen microbiome, of which the genus *Entodinium* has been consistently observed to be dominant across a diverse genetic and geographical range of ruminant hosts. Despite the apparent core role that species such as *Entodinium caudatum* exert, their major biological and metabolic contributions to rumen function remain largely undescribed. Here, we have leveraged (meta)genome-centric metaproteomes from rumen fluid samples originating from both cattle and goats fed diets with varying inclusion levels of lipids and starch, to detail the specific metabolic niches that *E. caudatum* occupies in the context of its microbial co-habitants. Initial proteome estimations via total protein counts and label-free quantification highlight that *E. caudatum* comprises an extensive fraction of the total rumen metaproteome. Our analysis also suggested increased microbial predation and volatile fatty acid (VFA) metabolism by *E. caudatum* to occur in high methane-emitting animals, although with no apparent direct metabolic link to methanogenesis. Despite *E. caudatum* having a well-established reputation for digesting starch, it was unexpectedly less detectable in low methane emitting-animals fed high starch diets, which were instead dominated by propionate/succinate-producing bacterial populations suspected of being resistant to predation irrespective of host. Finally, we reaffirmed our abovementioned observations in geographically independent datasets, thus illuminating the substantial metabolic influence that

under-explored eukaryotic populations have in the rumen, with greater implications for both digestion and methane metabolism.

Keywords: Protozoa, *Entodinium caudatum*, metaproteomics, metagenomics, multi-omics, methane, rumen

BACKGROUND

Ruminants operate in symbiosis with their intrinsic rumen microbiome, which is responsible for the degradation of forage into nutrients, in the form of volatile fatty acids (VFAs), supplying ~70% of net energy for the host[1]. The rumen microbiome itself is a complex assemblage of bacterial, fungal, archaeal, viral, and protozoal microorganisms whose intricate composition and function is connected to host productivity traits, such as feed efficiency, milk yield, animal health and greenhouse gas (GHG) emissions[2-5]. Large collaborative research efforts have been made to identify and characterize the core rumen microbiome including creating a publicly available catalogue for cultivated and sequenced genomes[6-8]. In the rumen, bacteria is estimated to constitute 50-90 %, protozoa 10-50 %, fungi 5-10% and archaea less than 4% of the total microbial biomass[9,10]. Due to the difficulties of axenically culturing rumen eukaryotic populations and their complex genomic features that are obstinate to current metagenomic technologies, the reconstruction of the rumen microbiome has been heavily biased towards bacterial and archaeal members, whereas the fungal and protozoal contributions of the rumen currently remain poorly characterized. While anaerobic fungi have a reputable role as fibre degraders in the rumen, only 18 anaerobic gut fungi from herbivores are currently described, with only 11 genomes available[11-13]. Similarly, to date few rumen protozoal genomes are sequenced and publicly available, chief among them, the rumen ciliate protozoa *Entodinium caudatum*[14].

Entodinium represents one of the most dominant genera of rumen protozoa, previously being detected in more than 99% of 592 rumen samples at a mean protozoal relative abundance of ~38% (2015 rumen census: 32 animal species, 35 countries)[15]. While *E. caudatum* has been previously observed to stimulate methane production[16], it is believed that this protozoa lacks hydrogenosomes, and instead encodes mitosomes and iron hydrogenases that may indicate hydrogen production, beneficial for methanogenic endosymbionts[17]. Here, we present a genome-centric metaproteomics analysis of the rumen microbiome from two different host species Holstein dairy cows (*Bos taurus*) and alpine goats (*Capra hircus*) that were fed diets of first-cut grassland hay with a 45:55 forage:concentrate ratio, with concentrates supplemented with either no additional lipid (CTL), or corn oil and cracked-wheat starch grains (COS) [5,18,19]. Moreover, metadata revealed that animals fed COS displayed reduced methane emissions, irrespective of host. To describe how these diets affect digestion and production of methane and VFA's, we sought to investigate changes in function and

70 composition in the complex rumen microbiome. By using shotgun metagenomic sequencing we
71 recovered in total 244 prokaryote metagenome-assembled genomes (MAGs) that together with
72 selected isolate-derived eukaryote genomes[14,20-25] formed the database for our integrated functional
73 analysis of *E. caudatum*. Despite *E. caudatum* having the genetic ability to degrade plant
74 polysaccharides such as starch and produce hydrogen, our analysis showed contrasting data that
75 suggests *E. caudatum* is less metabolically active in the rumen microbiome of animals fed a starch-
76 rich diet. In such a scenario other starch-degrading and/or propionate and succinate producing
77 bacterial genera, as *Prevotella* and *Fibrobacter* and members of the families *Succinivibrionaceae* and
78 *Aminobacteriaceae* appeared to be more prevalent. In concert, our analysis showed that reduced
79 methane production in both cattle and goats eating feeds supplemented with COS is likely caused via
80 a redirection of hydrogen to succinate and propionate production instead of methanogenesis. Finally,
81 by analysing a secondary, geographically independent dataset, we reaffirmed our primary
82 observations of *E. caudatum* dominance and starch-related metabolism, thus supporting our
83 hypothesis that this protozoal species plays a core role in rumen microbiome function.
84

85 ***METHODS***

86 ***Animal trial and sample handling***

87 The experimental procedures were approved by the Auvergne-Rhône-Alpes Ethics Committee for
88 Experiments on Animals (France; DGRI agreement APAFIS#3277–2015121411432527 v5) and
89 complied with the European Union Directive 2010/63/EU guidelines. Experiments were performed at
90 the animal experimental facilities of HerbiPôle site de Theix at the Institut National de la Recherche
91 pour l'Agriculture, l'Alimentation l'Environnement (INRAE, Saint-Genès-Champanelle, France) from
92 February to July 2016. Experimental design, animals and diets were previously described by Fougère
93 *et al.*[19] and Martin *et al.*[5]. Briefly, 4 Holstein cows and 4 Alpine goats, all lactating, were enrolled in
94 respectively two 4 x 4 Latin square design trials to study the effects of 4 diets over four 28-d
95 experimental periods. The original study included a control diet and 3 experimental diets with various
96 lipid sources[19]. However, in this work, we only focused on the CTL (grass hay and concentrates
97 containing no addition lipids) and COS (CTL diet supplemented with corn oil and wheat starch) diets
98 which were associated with the most extreme methane (CH_4) emission profiles in both ruminant
99 species. In the present study, we focused on the two following diets: a diet composed of grass hay ad
100 libitum with concentrates containing no additional lipid (CTL) or corn oil (5.0% total dry matter intake
101 (DMI)) and cracked-wheat starch -5.0 % of total DMI (COS) (**Table 1**). Corn oil (Olvea, Saint Léonard,
102 France) was added to the concentrate, at 5% of total DMI and contained (g/kg of total FA): 16:0 (114),
103 18:0 (16.4), cis-9 18:1 (297), cis-11 18:1 (6.30), 18:2n-6 (535), 18:3n-3 (7.57), 20:0 (3.48), 22:0 (1.0),

24:0 (1.5), and total FA (1000 g/kg). Detailed diet composition is available in Martin *et al.*[5]. Each experimental period lasted for 28 days. Diets were offered as 2 equal meals at 0830 and 1600h. Animals had access to a constant supply of freshwater ad libitum. Concentrate and hay refusals were weighed daily. The amounts of feed offered the following day was adjusted regarding to refusals to maintain the targeted the dietary 45 % dry matter (DM) forage and 55 % DM concentrate ratio.

Table 1. Summary of the animal host, dietary conditions and accompanying metadata that are linked to the rumen samples used to explore *E. caudatum* function. Effects of control diet (CTL) and diet supplemented with corn oil and wheat starch (COS) on average VFA concentration in the percentage of total VFA concentration (mmol/l) and average methane yield (CH_4 g/kg DMI) in addition to average total protozoa cell counts and average cell counts for small *Entodiniomorph* protozoa for cows (n=4) and goats (n=4) (**Supplementary Table S1**)[5]. Total protozoal cell counts, and cell counts of small *Entodiniomorph* protozoa were based on ciliate protozoa morphology in microscopy[5]. Measurements on VFA concentrations and methane yield were determined by gas chromatography and respiration chambers, respectively.

Animal	Diet	VFA sum mmol/l	Acetate (% sum)	Propionate (% sum)	Butyrate (% sum)	CH_4 g/kg dry matter intake	Standard deviation	Total protozoa (10^3 cells/ml)	Small entodiniomorphs (<100µm) (10^3 cells/ml)
COW	CTL	61.83	72.02	14.21	10.80	19.51	2.18	92	51
	COS	68.06	69.12	21.32	6.39	14.45	2.89	100	97
GOAT	CTL	33.70	65.73	15.62	13.66	19.52	3.96	1382	1346
	COS	25.27	64.17	19.07	9.90	13.47	4.14	898	898

Rumen fluid was collected through stomach-tubing before the morning feeding on day 27 of each experimental period. The stomach tube consisted of a flexible 23 mm external diameter PVC hose fitted to a 10 cm-strainer at the head of the probe for cows, and a flexible 15 mm PVC hose with a 12 cm-strainer for goats. The first 200 ml of rumen fluid was discarded from to minimize contamination from saliva. Samples were filtered through a polyester monofilament fabric (280 µm pore size), dispatched in 2-ml screw-cap tubes, centrifuged at 15000 x g for 10 mins and the pellet snap-frozen in liquid nitrogen. Samples were stored at -80°C until DNA extraction using the Yu and Morrison bead-beating procedure [26]. In total, 32 rumen fluid samples (4 cattle and 4 goats fed four diets included in the original study[19]) were sent to the Norwegian University of Life Sciences (NMBU) for metagenomic and metaproteomic analysis. Respiration chambers were used to measure methane emissions over a 5-day period, while VFA and NH_3 concentrations were determined by gas chromatography using a flame ionization detector[27]. Protozoa were counted by microscopy and categorized as either small entodiniomorphs (<100 µm), large entodiniomorphs (>100 µm) or as holotrichs *Dasytricha* and *Isotricha*[9]. Further specifics about diets and measurements can be found in Martin *et al.*[5] and VFA and methane measurements are summarized in **Table 1**.

137 ***Metagenomic sequencing and analysis***

138 Metagenomic shotgun sequencing was performed at the Norwegian Sequencing Centre on two lanes
139 of the Illumina HiSeq 3/4000 generating 150 bp paired-end reads in both lanes. Sequencing libraries
140 were prepared using the TruSeq DNA PCR-Free High Throughput Library Prep Kit (Illumina, Inc) prior
141 to sequencing. All 32 samples (4 cattle and 4 goats fed four diets included in the original study[19]) were
142 run on both lanes to prevent potential lane-to-lane sequencing bias. FASTQ files were quality filtered
143 and Illumina adapters removed using Trimmomatic [28] (v. 0.36) with parameters -phred33 for base
144 quality encoding, leading and trailing base threshold set to 20. Sequences with an average quality
145 score below 15 in a 4-base sliding window were trimmed and the minimum length of reads was set to
146 36 bp. MEGAHIT [29] (v.1.2.9) was used to co-assemble reads originating from samples collected from
147 cow and goats separately, with options –kmin-1pass, --k-list 27,37,47,57,67,77,87, --min-contig-len
148 1000 in accordance with [7]. Bowtie2[30] (v. 2.3.4.1) was used to map reads back to the assemblies and
149 SAMtools[31] (v. 1.3.1) was used to convert SAM files to BAM format and index sorted BAM files.

150

151 The two co-assemblies (one from the samples originating from cattle and the other originating from
152 the samples of goats) were binned using Maxbin2[32], MetaBAT2[33] and CONCOCT[34]. MetaBAT2 (v.
153 2.12.1) was run using parameters –minContig 2000 and –numThreads 4, Maxbin2 (v. 2.2.7) ran with
154 default parameters and -thread 4, min_contig_length 2000, and CONCOCT (v. 1.1.0) ran with default
155 parameters and –length_threshold 2000. Further, bins were filtered, dereplicated and aggregated
156 using DASTool[35](v. 1.1.2) with the parameters –write_bins 1, --threads 2 and BLAST[36] as search engine.
157 This resulted in a total of 244 dereplicated MAGs across the two host species (104 originating from
158 cow and 140 from goat). CheckM[39](v. 1.1.3) lineage workflow was used to determine completeness
159 and contamination of each MAG, with parameters –threads 8, --extension fa, and CoverM (v. 0.5.0)
160 (https://github.com/wwood/CoverM) was used to calculate relative abundance of each MAG, while
161 GTDB-tk[40,41] (v. 1.3.0) was used for taxonomic annotation. 90% (219 of 244) of the recovered MAGs
162 were considered high or medium quality MAGs according to MIMAGs threshold for completeness and
163 contamination for genome reporting standards[42]. Gene calling and functional annotation of the final
164 MAGs were performed using the DRAM[43] pipeline with the databases dbCAN[44], Pfam[45], Uniref90[46],
165 Merops[47], VOGdb and KOfam[48]. The translated amino acid sequences from the publicly available
166 drafted *E. caudatum* macronucleus genome were annotated with the KEGG metabolic pathway
167 database using BlastKOALA[49] by Park *et al.* [14]. Proteins with resulting KEGG Orthology (KO) numbers
168 were functionally assigned to metabolic pathways using KEGG Mapper Reconstruct Tool [50].

169

170 The resulting protein sequences for each MAG, as well as those from the host genomes of goat (*Capra*
171 *hircus*, NCBI ID: 10731) and cattle (*Bos taurus*, NCBI ID: 82) were compiled into two databases, from
172 now on referred to as sample specific RUmen DataBase for Goat (RUDB-G) and sample specific RUmen
173 DataBase for cattle (RUDB-C). In addition, the genomes of the protozoa *Entodinium caudatum* [14] and
174 *Fibrobacter succinogenes* S85 (NCBI ID: 932) was added to both rumen databases. *F. succinogenes* is
175 well recognized as a primary cellulolytic bacterium in the rumen microbiome and has previously been
176 observed as an active microorganism in similar studies yet was not thoroughly binned as a unique
177 MAG in this study. Nine available fungal genomes downloaded from Joint Genome Institute (JGI)
178 Mycocosm[51] were added to RUDB-C (*Anaeromyces* sp. S4[21], *Caecomyces churrovis*[24], *Neocallimastix*
179 *californiae*[21], *Neocallimastix lanati*[25], *Piromyces finnis*[21], *Piromyces* sp. E2[21], *Piromyces* UH3-1[20,21,23],
180 *Piromyces* eukMAG[52], *Orpinomyces* sp.[22]). Because of database size restrictions in downstream
181 analysis of metaproteomics data[53], only four of the nine fungal genomes were added to RUDB-G
182 (*Anaeromyces* sp. S4, *Caecomyces churrovis, Neocallimastix lanati, Piromyces* UH3-1). In total the
183 complete databases consisted of 452 073 and 431 023 protein entries for RUDB-G and RUDB-C,
184 respectively.
185
186 ***Metaproteomic data generation***
187 To 300 µL of rumen fluid sample (in total 32 rumen fluid samples originating 4 cattle and 4 goats fed
188 four diets included in the original study[19]) 150 µL lysis buffer (30 mM DTT, 150 mM Tris-HCl (pH=8),
189 0.3% Triton X-100, 12% SDS) was added together with 4 mm glass beads (≤ 160 µm), followed by brief
190 vortexing and resting on ice for 30 mins. Lysis was performed using FastPrep-24™ Classic Grinder (MP
191 Biomedical, Ohio, USA) for 3 × 60 seconds at 4.0 meter/ second[54]. Samples were centrifuged at 16 000
192 × *g* for 15 minutes at 4°C and lysate was carefully removed. Protein concentration was measured using
193 the Bio-Rad DC™ Protein Assay (Bio-Rad, California USA) with bovine serum albumin as standard.
194 Absorbance of sample lysates was measured at A750 on BioTek™ Synergy™ H4 Hybrid Microplate
195 Reader (Thermo Fisher Scientific Inc., Massaschusetts, USA). 40-50 µg of protein was prepared in SDS-
196 buffer, heated in water bath for 5 minutes at 99°C and analysed by SDS-PAGE using Any-kD Mini-
197 PROTEAN TGX Stain-Free™ gels (Bio-Rad, California, USA) in a 2-minute run for sample clean-up
198 purposes, before staining with Coomassie Blue R-250. The visible bands were carefully excised from
199 the gel and divided as 1×1 mm pieces before reduction, alkylation, and digestion with trypsin. Peptides
200 were concentrated and eluted using C18 ZipTips (Merck Millipore, Darmstadt, Germany) according to
201 manufacturer's instructions, dried, and analysed by nano-LC-MS/MS using a Q-Exactive hybrid
202 quadrapole Orbitrap MS (Thermo Fisher Scientific Inc., Massaschusetts, USA) as previously described[55]
203

204 ***Metaproteomic data analysis***

205 Acquired MS raw data were analysed using MaxQuant[56] (v. 1.6.17.0) and searched against the RUDB's.
206 The MaxLFQ algorithm was used to quantify proteins[57]. Detected protein groups were explored in
207 Perseus[58] (v. 1.6.8.0). Both RUDB's were supplemented with contaminant protein entries, such as
208 human keratin, trypsin, and bovine serum albumin, in addition to reversed sequences of all protein
209 entries for estimation of false discovery rates (FDR). Oxidation of methionine, protein N-terminal
210 acetylation, deamination of asparagine and glutamine, and conversion of glutamine to pyroglutamic
211 acids were used as variable modifications, while carbomidomethylation of cysteine residues was used
212 as fixed modification. Trypsin was chosen as digestive enzyme and maximum missed cleavages
213 allowed was one. Protein groups identified as potential contaminants were removed. Proteins were
214 filtered to 1% FDR and considered valid if they had at least one unique peptide per protein and at least
215 one valid value in total. One sample originating from goat fed HPO diet (14201 P2 (Goat 2 fed HPO P2,
216 sample no. 22)) was deemed as a technical outlier, mapping significantly fewer protein groups in
217 metaproteomic analysis compared to the rest of the data and was therefore removed from the
218 downstream analysis. After filtration, we resolved 1081 unique protein groups across the 16 samples
219 from cattle and 1632 unique protein groups across 15 samples from goats. Box plots for **Figure 1** were
220 made with ggplot2[59] in R (v. 4.2.0)[60]. To determine which expressed metabolic pathways *E. caudatum*
221 were significantly enriched for in each diet/animal, we used the hyperR package[61] in R which employs
222 the hypergeometric test. The 'geneset' for hyperR was generated by using the KEGGREST R package
223 to retrieve entries from the KEGG database and determine which pathways the *E. caudatum* KOs
224 belong to. The geneset was then manually curated to only include metabolic pathways of interest (i.e.,
225 we remove pathways such as "Huntington disease"). For the 'background' setting in hyperR, to be
226 conservative, we used the total number of unique KOs (7592) in the *E. caudatum* genome that could
227 possibly be expressed.

228

229 ***Animal trial, sample handling and metagenomic data generation for independent validating dataset***

230 Samples were also analysed from previously performed feeding experiments with Holstein Friesian
231 bulls[62]. In brief, these bulls were subjected to either *ad libitum* or restricted feeding regime in a
232 compensatory growth model detailed in Keogh *et al.*, 2015[62]. Both feeding groups received the same
233 ratio of concentrate and grass silage, respectively 70% and 30%, of which the concentrate was mainly
234 composed of starch-rich rolled barley (72.5%) and soya (22.5%). Rumen samples were collected at
235 slaughter and stored at -80C prior to metagenomics and metaproteomic analysis in this study.

236

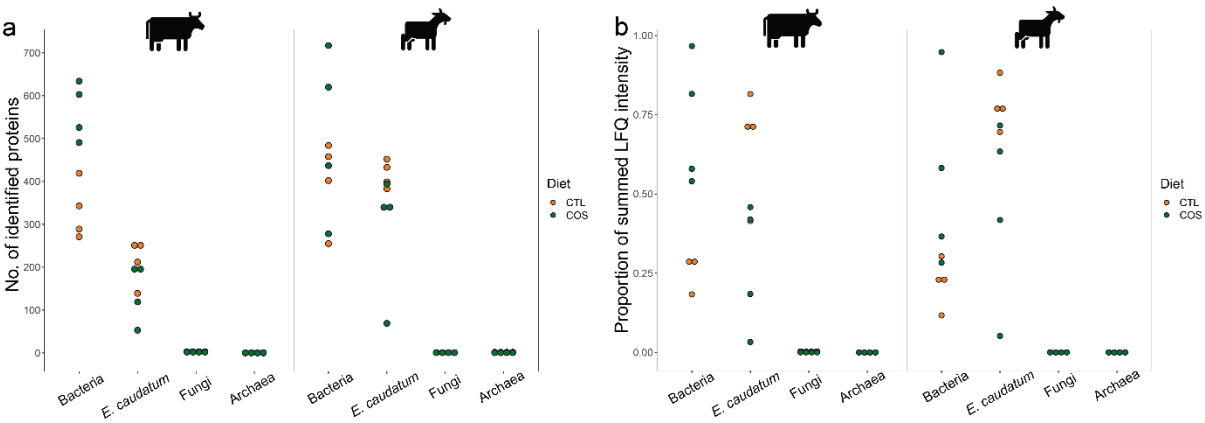
237 Sample preparation, cell lysis and extraction of DNA was carried out as previously described by
238 McCabe *et al.*[63] Quality check of fastq files and removal of low-quality reads was performed using fastp
239 (V.0.19.5). Sequence reads were mapped against the bovine genome (ARS-UCD1.3) using minimap2
240 (V.2.16), and host sequences were removed. Reads were co-assembled using Megahit (V1.2.6) with
241 "–meta-large" pre-set option as the metagenome was complex. Metagenomic binning was applied to
242 the co-assembly using MetaBAT2 using standard parameters (V.2.12.1). MAGs were then dereplicated
243 using dRep (V.1.4.3), and the resulting MAGs were taxonomically annotation using Bin Annotation
244 Tool (BAT), available on (https://github.com/dutilh/CAT). This tool internally uses prodigal (V.2.6.3)
245 for gene prediction and DIAMOND (V.0.9.14) for the alignment against the non-redundant (nr) protein
246 database (As of Feb 2020).
247
248 Sample preparation for metaproteomics was done by lysing cells with bead beating with two glass
249 bead sizes (≤106 µm and 0.5 mm), in 100 mM Tris, pH8, 5% SDS and 10 mM DTT. A FastPrep 24
250 instrument was operated for 3 × 45 seconds at a speed of 6.5 m/s. The samples were centrifuged for
251 15 minutes at 20.000 × *g* and the protein extracts were cleaned by Wessel-Flügge precipitation[64];
252 pellets were dissolved in 5% SDS, 100 mM Tris-Cl, pH8, 10 mM DTT and kept at -20 ^{o}C until further
253 processing. Protein digestion was performed using suspension trapping (STrap)[65], dried in a SpeedVac
254 (Eppendorf Concentrator Plus) and re-dissolved in 0.05 % trifluoroacetic acid, 2% acetonitrile for
255 peptide concentration estimation using a Nanodrop One instrument, and subsequent MS/MS-
256 analysis. The samples were analyzed using an Ultimate3000 RSLCnano UHPLC coupled to a QExactive
257 hybrid quadrupole-orbitrap mass spectrometer (Thermo Fisher, Bremen, Germany) as described
258 previously[55].
259 Mass spectrometry raw data were analysed with a sequence of software in the Galaxy software suite
260 (usegalaxy.eu). Initially, they were searched against the sample-specific protein sequence database
261 (1.773.447 protein sequences) with SearchGUI[66] utilizing the X!Tandem search engine[67] version
262 Vengeance. The database was supplemented with contaminant protein entries, such as human
263 keratin, trypsin, and bovine serum albumin, in addition to reversed sequences of all protein entries
264 for estimation of false discovery rates (FDR). Oxidation of methionine and protein N-terminal
265 acetylation were used as variable modifications, while carbomidomethylation of cysteine residues
266 were used as fixed modification. Trypsin was chosen as digestive enzyme, maximum missed cleavages
267 allowed was one and matching tolerance levels for MS and MS/MS were 10 ppm and 20 ppm,
268 respectively. PeptideShaker[68] was used to filter the results to 1% FDR and quantification was done
269 using FlashLFQ[69] including normalization between samples and the feature 'match between runs' to
270 maximize protein identifications. Perseus[58] version 1.6.2.3 was used for further analysis. A protein

271 group was considered valid if it had at least one unique peptide identified and being quantified in at
272 least 50% of the replicates in at least one condition (7 restricted and 8 Ad lib). Protein groups identified
273 as potential contaminants were removed. Calculations of MAG abundances were done by summing
274 LFQ values for all proteins belonging to each MAG and differential abundance between diets were
275 detected by a two-sided Student's t-test ($p<0.05$).

276 ***RESULTS AND DISCUSSION***

277 ***Protozoal populations have large proteomes in the rumen microbiome***

278 Because of their large size protozoal species can comprise a significant fraction of the microbial
279 biomass in the rumen[9]. While the total number and diversity of protozoal species are lesser than their
280 bacterial counterparts in the rumen, their genome size and total gene count are considerably larger
281 and due to alternative splicing and post-translational modifications, the protein representation of
282 protozoal populations will be larger than the number of genes in the genome[70]. Thus, the amount of
283 protein in a protozoa species can be expected to far exceed the amount of protein in a bacterial
284 species that can be identified and quantified in proteomic studies. In this context, our metaproteomic
285 data showed an extensive fraction of detectable proteins affiliated to *E. caudatum*, and other closely
286 related species, in both cows (26.3%) and goats (31.5%) in proportion to the combined bacterial
287 species that were represented in our genome databases (**Figure 1a**). In addition, the label free
288 quantification (LFQ) of *E. caudatum* proteins, which is indicative of protein detection intensity, was
289 proportionally higher than the bacterial fraction of the rumen microbiome further supporting the
290 dominance of protozoal activity in our samples (**Figure 1b**). Twice as many *E. caudatum* proteins were
291 detected in goat than cows (mean: 514 vs 284), however this was somewhat expected given the 7x
292 higher counts of entodiniomorph concentration (cells/mL) previously observed in the goat samples
293 compared to cows (**Table 1**)[5].



294

295 **Figure 1. Quantities of proteins identified as *E. caudatum* in the rumen microbiome vary depending on host**
296 **animal and dietary conditions.** Boxplots for total proteins identified (**a**), and average recovered metaproteomic
297 expression (**b:** presented as proportion of summed LFQ intensities) belonging to protozoal (*E. caudatum*),
298 bacterial, archaeal, or fungal species across the control diet (CTL) and diets supplemented with corn oil and

299 wheat starch (COS) for dairy cows (n=4) and goats (n=4). Detected proteins LFQ intensities for protozoal and
300 bacterial populations can be found in **Supplementary Table S2**.

301

302 ***Metabolism of E. caudatum shows predatory activity and metabolism of VFAs***

303 While previous efforts have investigated the genome and transcriptome of *E. caudatum* grown in
304 monoculture[14,71], our metaproteomic analysis sought to reveal *in vivo* metabolism and functions of *E.*
305 *caudatum* within the rumen microbiome. In accordance with Wang *et al.*[71], our metaproteomic
306 analysis revealed expressed proteins significantly enriched in metabolic pathways such as carbon
307 metabolism, glycolysis/gluconeogenesis, starch and sucrose (and glycogen) metabolism, pyruvate
308 metabolism, oxidative phosphorylation and production of alcohol (**Supplementary Table S3**). Wang *et*
309 *al.* found that as for most rumen microbes, *E. caudatum* uses carbohydrates such as starch as its
310 primary substrate, as well as cellulose and hemicellulose to a certain degree[71], and their transcript
311 analyses showed that *E. caudatum* had high levels of expression of amylases and low-level expression
312 of hemicellulases, cellulases and pectinases. Similarly, our metaproteomic analysis reveals expression
313 of amylases by *E. caudatum* that are predicted to enable *E. caudatum* to engulf and degrade starch
314 granules to simpler sugars and to produce glycogen, its most important storage carbohydrate[72].
315 However, no detection of *E. caudatum* carbohydrate active enzymes (CAZymes) related to
316 hemicellulose or pectin were observed in any of our metaproteomes, suggesting that it is not engaging
317 in the deconstruction of these carbohydrates at the time our samples were collected for analysis
318 (before feeding). It should be noted that ruminal fermentation activity as well as production of VFA's
319 and methane will be at its highest after feeding, as a result of an increased availability of fermentable
320 substrate[73]. While sampling time can influence the recovered microbial composition and hence
321 function, any differences in metabolic parameters or species abundance in this study is relative across
322 both diets given the consistent sampling times.

323

324 While monoculture cultures of *E. caudatum* have not been established to verify the VFA's it can
325 produce, Wang *et al.* found transcripts of enzymes involved in fermentative formation of acetate and
326 butyrate[71]. Similarly, we detected proteins inferred in metabolism of acetate, butyrate, and alcohol in
327 *E. caudatum*. Irrespective of host, animals fed the control (CTL) diet had a higher proportion of *E.*
328 *caudatum* proteins and concurrently had increased relative levels of acetate and butyrate compared
329 to animals fed the corn oil and wheat starch diet (COS), which had fewer *E. caudatum* proteins and
330 lower acetate/butyrate levels (**Figure 1** and **Table 1**). As *E. caudatum* was seemingly most abundant
331 in goats fed the CTL diet, we used these metaproteomes to reconstruct metabolic features (**Figure 2**).
332 Of the 514 *E. caudatum* proteins identified in goats, 454 had unique KO numbers assigned, from which
333 KEGG Mapper reconstructions[50] enabled functional assignment of 268 proteins to metabolic

334 pathways. Our metabolic reconstructions showed expressed proteins involved in endocytosis,
335 phagosome and lysosome processes for predatory activity, engulfment, and digestion of bacteria
336 (**Supplementary Table S2**). Interestingly, for the rumen samples used in this study Martin *et al.*
337 previously observed higher NH_3 concentrations in goats compared to cows[5] and hypothesised that it
338 might have resulted from increased bacterial protein breakdown and feed protein degradability due
339 to higher density of entodiniomorphs known for their predatory activity[5]. In support of these
340 observations, we performed metaproteomic pathway enrichment analysis of *E. caudatum* (**Figure 2,**
341 **Supplementary Table S3**), which revealed significantly enriched nitrogen metabolism, in addition to
342 purine and pyridine metabolism in goats but not in cows. Other biological processes such as signalling
343 and metabolism of amino acids, and amino and nucleotide sugars represented significantly enriched
344 pathways in *E. caudatum* (**Supplementary Table S3**). In the previous transcriptome study by Wang *et*
345 *al.*, transcripts for a ferredoxin hydrolase and an iron hydrogenase were recovered and are suspected
346 to be involved in production of hydrogen. Here in our metaproteomic analysis, we identified only one
347 of its eight iron hydrogenases (in goats, it was absent in cows, **Supplementary Table S2**), which
348 showed no contrasting changes in LFQ intensity in either the CTL or COS datasets, contributing to the
349 uncertainty that *E. caudatum* is a major producer of hydrogen, despite previous reports associating
350 its abundance with higher methane levels [74-76].
351

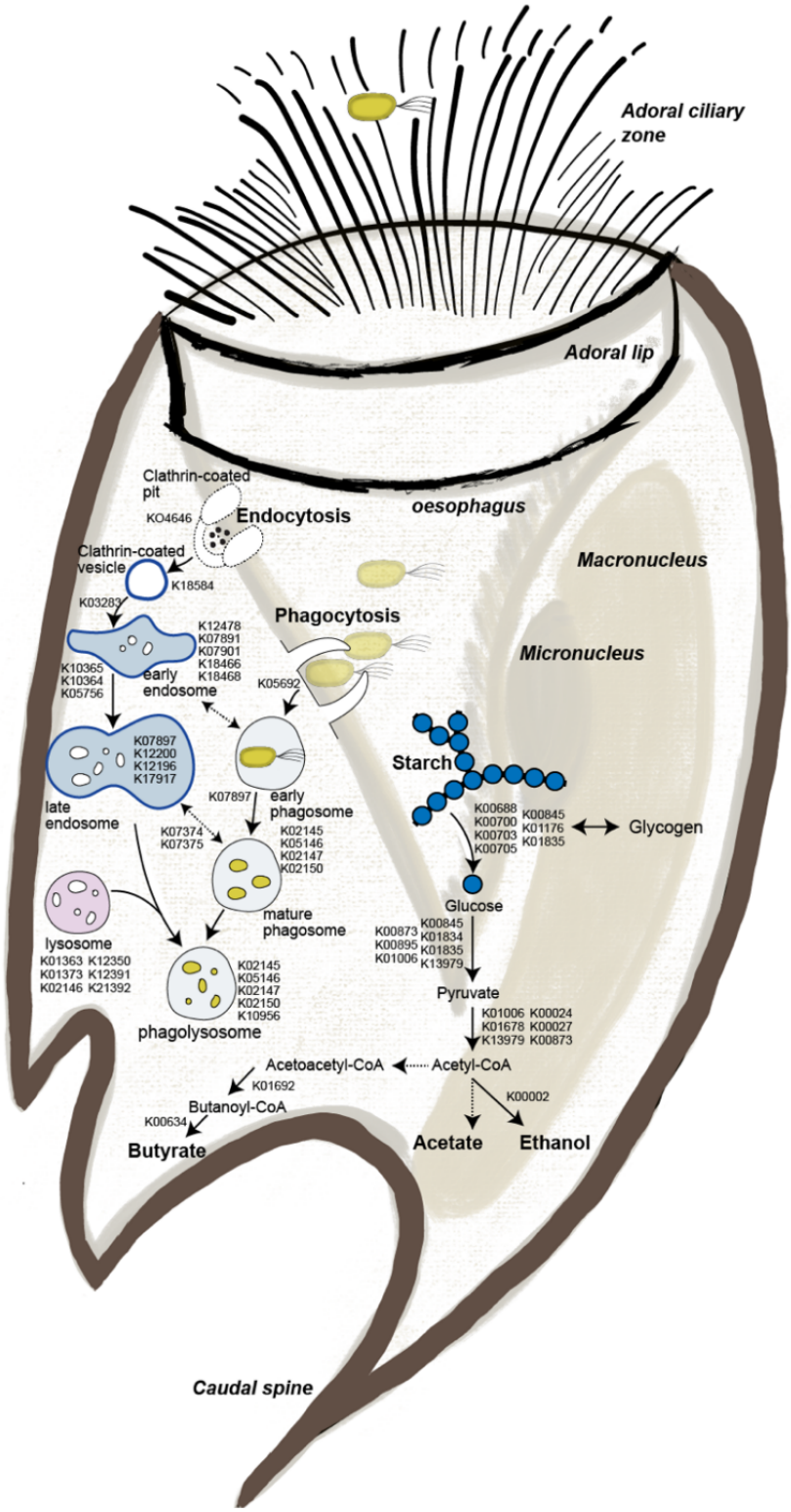


Figure 2. Reconstructed phagolysosome formation and starch metabolism of *Entodinium caudatum* within the rumen microbiome of goats fed the control diet (CTL) based on metaproteomic analysis. KO identifiers for identified proteins were analysed via KEGG mapper to reconstruct expressed key features in the metabolism of *E. caudatum*. Dashed arrows represent proteins or pathways that were not detected in our metaproteomes but are key steps in their respective pathways. Detailed information connecting KO identifiers to their respective gene ID, LFQ and animal/diet can be found in **Supplementary Table S2**.

360 ***E. caudatum is less active in diets supplemented with starch regardless of its starch degrading***
361 ***reputation.***
362 The changes in VFA and methane levels in animals fed the high starch COS diet, previously measured
363 by Martin *et al.*[5], suggested significant alterations in composition and thus functions of the rumen
364 microbiome irrespective of host species. In particular, a decrease in proportions of acetate and
365 butyrate, decrease in the acetate:propionate ratio and an increase in proportional propionate levels
366 were observed in animals fed the COS diet, compared to the CTL diet (**Table 1**). Diets that are high in
367 starch content or with low forage:concentrate ratios have previously been shown to result in higher
368 production of propionate and succinate, as they are easily fermented in the rumen and accordingly
369 have high passage rates[77,78]. We therefore leveraged our genome-centric metaproteomic data from
370 both cows (**Figure 3a-c)** and goats (**Figure 3d-f**) fed either the COS or CTL diet to gain an overview of
371 protein expression from individual populations. We specifically focused on pathways involved in the
372 degradation of starch (CTL: corn starch, COS: corn + wheat starch) to pyruvate through glycolysis and
373 finally formation of acetate, butyrate, and propionate (via succinate). Irrespective of host, and despite
374 its starch-degrading reputation[71,72], *E. caudatum* had a lower abundance and less proteins involved in
375 starch degradation in animals fed the COS diet compared to those fed the CTL diet (**Figure 3a** and **3d**).
376 Further, we observed opposing patterns for *E. caudatum* proteins involved in glycolysis, and
377 production of pyruvate, acetate, and butyrate, which were detected in higher levels in both cows and
378 goats fed the CTL diet compared to the starch and corn oil (COS) supplemented diet.

379 While several putative *E. caudatum* amylases were detected across all animals and diets, their
380 quantification levels (i.e., LFQ intensities) did not increase as expected when higher levels of starch
381 were available (**Figure 3a** and **d**). We therefore hypothesized that the observed shift in VFA profiles in
382 response to increased starch was additionally influenced by the bacterial fraction of the rumen
383 microbiome. In contrast to lower *E. caudatum* levels in the animals fed the COS diet, we observed an
384 increase in suspected starch-degrading bacterial species, and succinate- and propionate-producing
385 bacterial species irrespective of host (**Figure 3c** and **3f**). For example, starch fermentation pathways
386 from population genomes affiliated with the *Succinivibrionaceae* family, *Prevotella* species,
387 *Fibrobacter* species and, additionally for goats, members of the *Selenomonadaceae* and
388 *Aminobacteriaceae* families, were detected at higher proteomic levels in the animals fed the COS diet
389 compared to those fed the CTL diet (**Figure 3c** and **3f**).

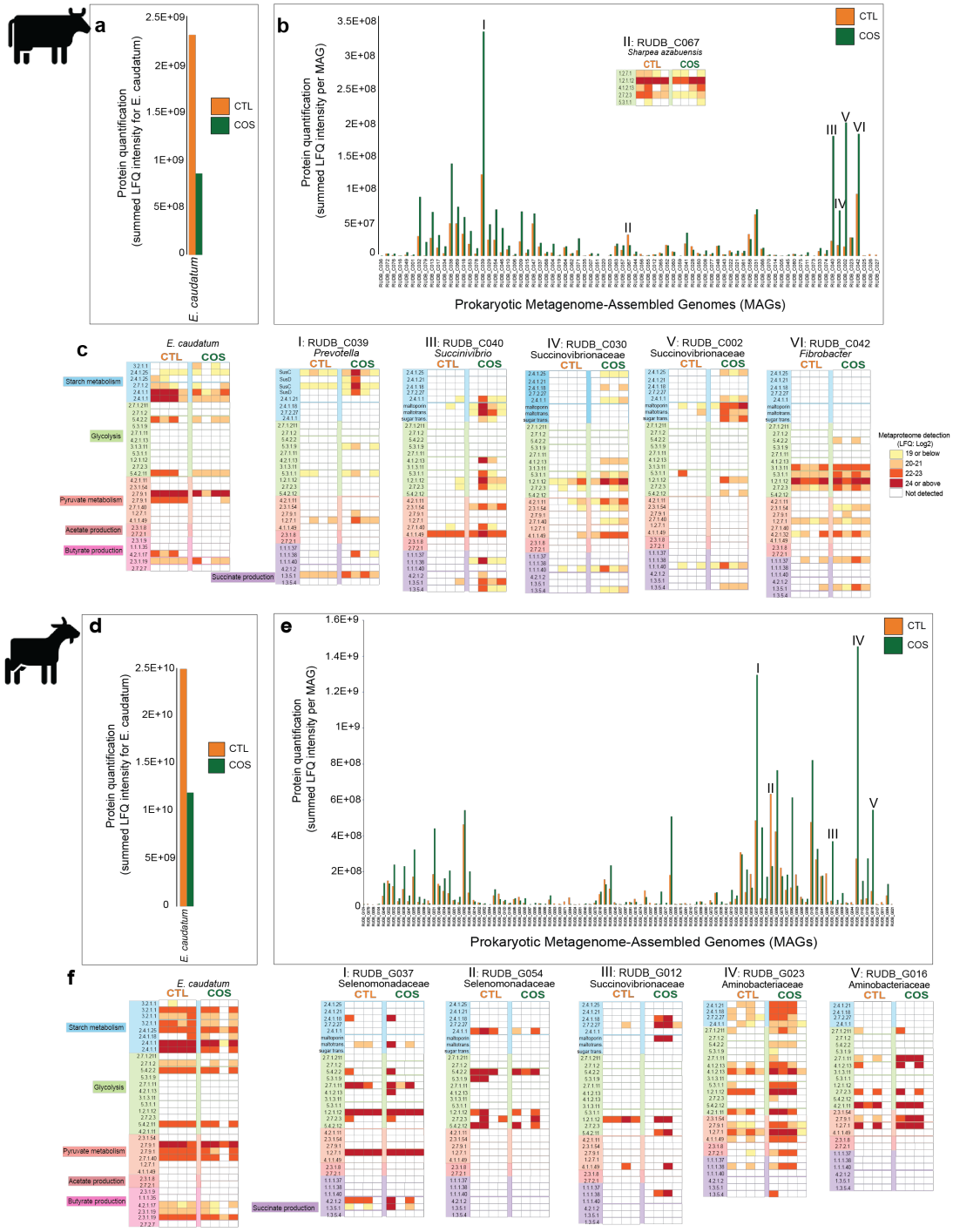


Figure 3. Detected proteins mapped to the genome of *E. caudatum* and bacterial metagenome-assembled genomes (MAGs) in the rumen microbiome of dairy cattle (n=4) and goats (n=4) fed either a control diet (CTL) or one supplemented with corn oil and wheat starch (COS). The figure displays metabolically active populations (as genomes or MAGs), with selected expressed proteins (presented as Enzyme Commission (EC) number) active in starch degradation, glycolysis and production of pyruvate, butyrate, acetate, and succinate in cows (**a-c**) and goats (**d-f**) fed CTL or COS diets. Panels **a** and **d** depict *E.caudatum* proteomes that were detected in cattle and goats respectively are presented separately to bacteria (panels **b** and **e**) as the scale of their protein quantification values were ~10x larger. Protein quantification values (y-axis) were calculated by considering both the number of proteins detected per MAG/genome and their LFQ intensity: we averaged LFQ intensities for each detected protein across biological replicates for each dietary condition (CTL: green or COS: orange), which were subsequently summed for all detected proteins per MAG/genome. Panels **c** and **f** show selected MAGs (I-VI) with metabolically active proteins, presented as EC numbers, recovered from cows (RUDB-C) and goats (RUDB-G) fed

403 either CTL or COS diets. MAGs are presented with their MAG ID and taxonomic annotation from GTDB-tk.
404 Genome annotations and LFQ intensities used to create panels **a-f** can be found in **Supplementary Table S2**.

405

406 ***E. caudatum is less active in animals that produce lower methane yield***

407 For the animals sampled in this study, Martin *et al.* demonstrated a ~25-30% reduction in methane
408 emissions in both cows and goats fed the COS diet compared to the control (**Table 1**)[5]. While our
409 proteomic evidence clearly showed a lower *E. caudatum* activity in COS-fed low-methane producing
410 animals, a specific mechanism that explains this phenomenon is still elusive. Previous comparisons
411 between defaunated and faunated animals have shown decrease in methane production in protozoa-
412 free ruminants, suggesting symbiotic interactions between methanogenic archaea and protozoal
413 species[74]. Methanogen's epi- and endo-symbiotic relationships with protozoa have also been
414 suggested to contribute to 9-37% of rumen methanogenesis[74,75,79,80]. Moreover, studying microcosms
415 with the presence and absence of protozoal species Solomon *et al.* reported higher levels of acetate
416 and butyrate in microcosms with protozoa present in addition to increased methane emissions[74],
417 which supports the main findings of animals fed the CTL diet in this study (**Figure 4**). It is tempting to
418 speculate that such protozoal-methanogen relationships in this study are centred on hydrogen
419 transfer. However, we observed minimal evidence in our proteomic data that *E. caudatum* makes
420 major contributions to ruminal hydrogen production that is linked to methane levels, with only one of
421 its eight iron hydrogenases detected in goats (absent in cattle), which showed no changes in LFQ
422 intensity in either the high (CTL) or low (COS) methane yielding animals.

423 Increases in dietary starch for ruminants is known to stimulate the propionate and succinate pathways
424 of starch-degrading bacteria, which due to their net incorporation of metabolic hydrogen [H]
425 represent a [H] sink in rumen fermentation besides hydrogenotrophic methanogenesis[79,81]. In addition
426 to starch, in a study conducted by Zhang *et al.*[82], goats fed corn oil as a supplement decreased ruminal
427 H_2 concentrations and total methane emissions. Nevertheless, there was seemingly no effect on
428 rumen protozoal populations, which suggests that corn oil does not act as an anti-protozoal agent,
429 with the dose of corn oil used in this study[82]. Furthermore, supplementation of dietary lipids can
430 decrease plant fibre degradation and hence levels of acetate and butyrate at the expense of
431 propionate production, as lipid-derived long-chain fatty acids can be toxic to keystone fibre degrading
432 gram-positive bacterial species[83,84]. These findings were in agreement with the decreased CH_4
433 production in cows and goats in this study fed the COS diet, which was observed to additionally impact
434 other ruminal fermentation parameters, such as increased propionate and decreased butyrate and
435 acetate levels (**Figure 4**)[5].

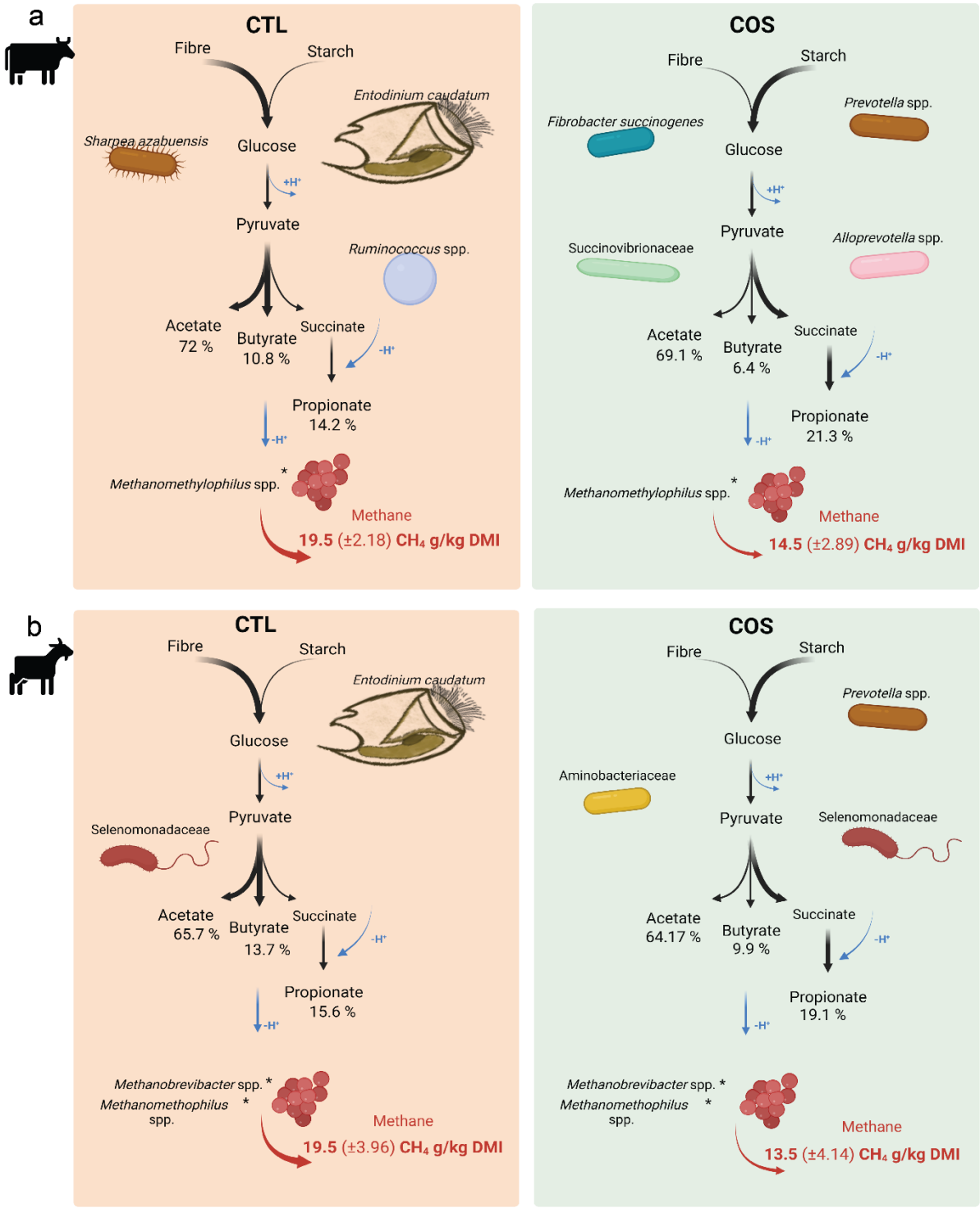


Figure 4. Schematic overview of the overall predicted metabolism in the rumen microbiome of cows (Panel a) and goats (Panel b) fed the control diet (CTL) compared to those fed the diet supplemented with corn oil and wheat starch (COS). The dominant microorganisms, as determined via MAG-centric metaproteomics, were predicted to convert starch into volatile fatty acids (VFAs) and methane in the rumen microbiome. However, the taxonomic and VFA profile of the rumen microbiomes differed when their host animals were subjected to different levels of starch in their diet. Thicker arrows represent higher abundance of metabolites. VFA concentrations were determined using gas chromatography. Percentages of VFA concentrations can be found in Supplementary Table S1. Percentages of VFA concentrations and methane yield are presented at average values, and numbers in parentheses are standard deviation for measured values for methane yield from each diet. * Represented by few detected proteins in the metaproteomic analysis.

447

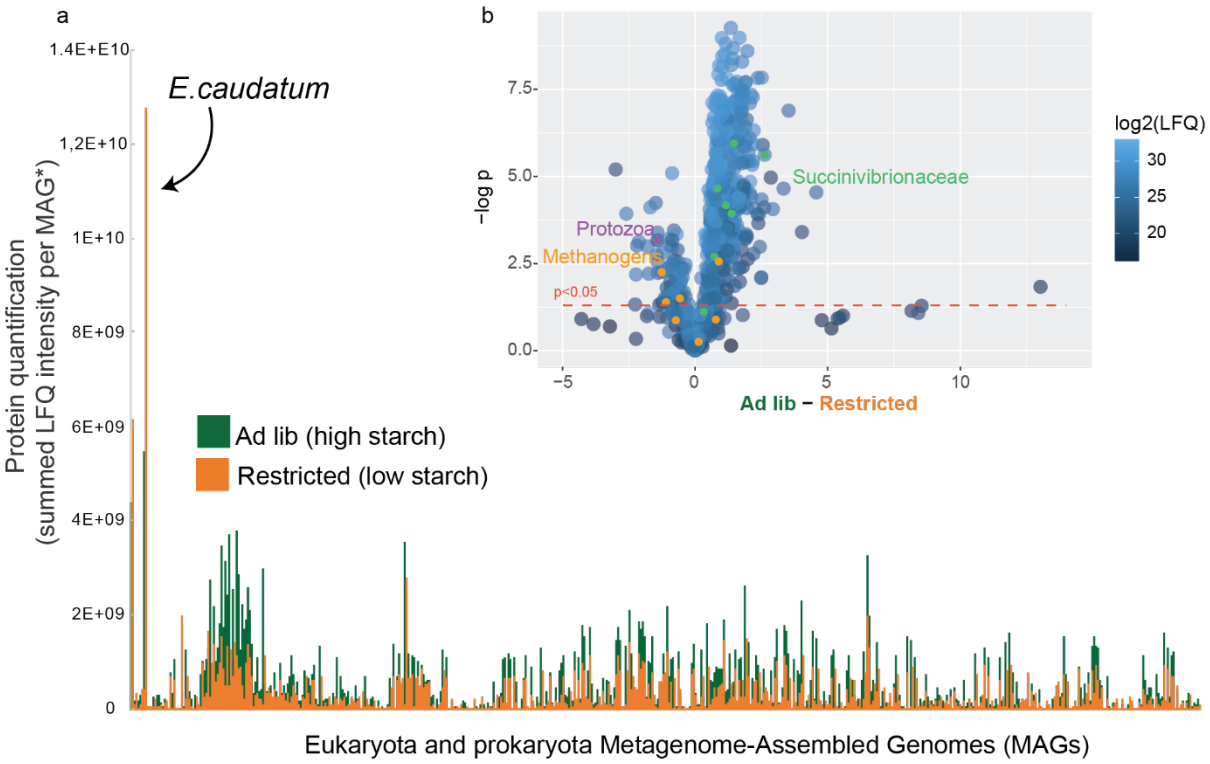
448 Diets rich in starch are more fermentable in the rumen, which can decrease the ruminal pH to levels
449 that can inhibit methanogenic archaea and fibre-degrading bacterial species[85,86]. Yet, lowered pH
450 levels in the rumen can also lead to clinical (or sub-clinical in most production scenarios) ruminal
451 acidosis[87,88]. Hence, high concentrate diets, which increase production of propionate at the expense
452 of methane, does not necessarily opt for a viable methane mitigation strategy in the long term. Our
453 results suggests that decreased methanogenesis in COS-fed animals is likely due to a decrease in
454 available hydrogen and/or decrease in pH levels, which we predict is caused by the metabolism of
455 dominant wheat starch-degrading populations that likely do not produce exogenous hydrogen due to
456 their own [H]-utilizing succinate and propionate metabolism (**Figure 4**).

457

458 ***E. caudatum dominance is validated in geographically independent datasets***

459 To further test our hypothesis that *E. caudatum* plays a central role in the rumen ecosystem, we
460 explored additional metagenome-centric metaproteomic datasets originating from an independent
461 feeding experiment performed in Ireland on 60 Holstein Friesian bulls[62]. In brief, these bulls were
462 subjected to the same ratio of concentrate and grass silage at either an *ad libitum* or restricted feeding
463 regime in a compensatory growth model detailed in Keogh et al, 2015[62]. We applied the same strategy
464 as for the described Holstein dairy cows and alpine goats to resolve the metaproteomic dataset for a
465 subset of 15 animals (7 restricted and 8 Ad libitum) against 781 reconstructed sample-specific MAGs
466 (RUDB-HF), which were supplemented with the genome of *E. caudatum,* as well as genomes of
467 available anaerobic fungi. This collection of microbial prokaryote and eukaryote genomes was then
468 used as a sequence database for the generated protein spectra. Consistent with our previous
469 observation, a substantial proportion of the detected proteins were affiliated to *E. caudatum*
470 providing further support that *Entodinium* is an important and metabolically active contributor to the
471 rumen microbiome. Intriguingly, the protein quantification (measured as sum of LFQ intensities
472 affiliated to each MAG/genome, averaged for each diet) was twice as high in the rumen sample from
473 bulls on the restricted diet, which likely had less starch available compared to the *ad libitium* group
474 and a higher retention time (**Figure 5a**). A previously published 16S rRNA amplicon investigation of
475 the phylogenetic differences between the rumen microbiomes of these two diet groups highlighted
476 an increase in *Succinivibrionaceae* in the starch-rich *ad libitium* diet[63]. Our metaproteomic analysis
477 confirmed a significantly higher ($p < 0.05$) proteomic detection of several *Succinivibrionaceae*-MAGs
478 under the *ad libitium* group (**Figure 5b**), accompanied with a reduced acetate:propionate ratio in the
479 rumen, which is often associated with increased feed efficiency and reduced production of methane[63].
480 These observations largely mirror the dominance of *Succinivibrionaceae*-MAGs in the dairy cattle and

481 goats fed the COS diet, further strengthening our hypothesis that *E. caudatum* does not metabolically
482 respond to increases in available starch in the host animals' diets and has other roles than being a
483 primary starch degrader.
484



485

486 **Figure 5. The proteomes of *Entodinium caudatum* and other rumen microbiome populations from Holstein-**
487 **Friesian beef cattle are affected by high starch diets.** A total of 60 beef cattle were subjected to two dietary
488 contrasting condition: 30 animals with *ad libitum* feeding and 30 subjected to 125 days of feed restriction.
489 Dietary components in both treatments consisted of 70% concentrate, and 30% grass silage, with the
490 concentrate containing rolled barley 72.5%, soya 22.5%, molasses 3% and calf mineral 2%. Rolled barley is high
491 in energy and starch content (~50%). **a.** Metaproteomes for a subset of 15 animals (7 restricted and 8 *ad libitium*)
492 were analysed against a database of 781 MAGs and isolate genomes, including eukaryotic representatives such
493 as *E. caudatum*. Protein quantification values (y-axis) were calculated by considering both the number of
494 proteins detected per MAG (and *E. caudatum*) and their LFQ intensity; we averaged LFQ intensities for each
495 detected protein across biological replicates for each dietary condition (Ad lib: green or restricted: orange),
496 which were subsequently summed for all detected proteins per MAG. Similar to our observations in Holstein
497 dairy heifers and Alpine goats (**Figure 3)**, the proteome of *E. caudatum* was a major fraction of the total rumen
498 metaproteome, however it was substantially reduced in dietary conditions where starch content was higher. **b.**
499 Volcano plot indicating different rumen microbiome proteins that displayed both large magnitude of fold-
500 changes in LFQ intensities (x axis) and high statistical significance (-log10 of p values using a t-test, y axis, dashed
501 horizontal line denote a $p<0.05$ cut-off). Supporting our results in dairy cattle and goats, we observed in high-
502 starch conditions an increase in protein detection for populations affiliated to the *Succinivibrionaceae* (green)
503 compared to the *E. caudatum* (purple) and methanogens (orange), which were detected at higher LFQ intensities
504 in restricted dietary conditions. LFQ intensities used to create panel **a** can be found in **Supplementary Table S4**.
505

506 ***E. caudatum seemingly has preferential bacterial species it will predate.***

507 *E. caudatum* is renowned for its predatory activity and is acknowledged as the most abundant
508 protozoa in the rumen, whereby it has been estimated that 0.1% of rumen prokaryotes are digested

509 by the rumen protozoal population every minute [89]. Although suspected of having metabolic
510 interactions with methanogenic archaea, several protozoal populations such as *Entodinium* are
511 hypothesized as having associations with certain members of the Gram negative
512 *Gammaproteobacteria*, which multiple studies have speculated are resistant to protozoal
513 engulfment[74,90,91]. In contrast, Gutierrez and Davis previously demonstrated that *Entodinium*-species
514 engulf Gram positive starch-degraders[91]. In the context of our data, we speculate that CTL fed animals
515 provided *E. caudatum* optimal conditions for predation, whereas increased starch levels in the COS
516 diets facilitated Gram-negative starch-degraders resistant to protozoal engulfment and/or reduced
517 pH levels. Such a scenario would enable populations of *Succinivibrionaceae* in cattle and/or
518 *Aminobacteriaceae* in goats to exploit the "predation free" COS diet and could plausibly explain the
519 observations of higher propionate levels, less methane, and lower activity of *E. caudatum*.

520 In conclusion, by using a (meta)genome-centric metaproteomics approach we primarily investigated
521 the role of the rumen protozoa *E. caudatum* in the rumen microbiome of beef and dairy cattle as well
522 as dairy goats that were subjected to varying dietary conditions. We showed that the proteome of *E.*
523 *caudatum* constitutes a substantial fraction of the recovered rumen microbial proteome, which
524 supports previous 16S/18S rRNA gene-based rumen census data that have highlighted its global
525 dominance across a plethora of ruminant species. However, *E. caudatum* proteins were surprisingly
526 detected at lower levels in animals that were fed increased levels of wheat starch, despite its reputable
527 starch-degrading capabilities (**Figures 3-5**). We hypothesize that this scenario is likely caused by the
528 out competition of *E. caudatum* by Gram-negative starch-degrading bacterial species that are possibly
529 resistant to protozoal engulfment and/or lower pH levels, creating sub-optimal conditions for *E.*
530 *caudatum*. We also observed limited evidence of *E. caudatum* metabolism being directly linked to
531 higher CH_4 yield at the time of sampling in this study (prior to feeding). However, the abundance of *E.*
532 *caudatum* in high methane-emitting animals may be indirectly fuelled in instances where preferential
533 pH conditions also support methanogens and fibrolytic bacteria that are also known to produce
534 hydrogen. Similarly, our data further support the theories that certain Gram-negative bacterial species
535 are resistant to predation by *E. caudatum*, which could enable specific niches for succinate- and
536 propionate-producing populations to flourish, subsequently exerting a larger impact on hydrogen and
537 methane metabolisms in the rumen microbiome. While much work is still needed to confirm our
538 abovementioned hypotheses, our integrated metaproteomics approaches have demonstrated the
539 future importance of including eukaryote populations for accurate and meaningful analyses of the
540 rumen microbiome and its impact on GHG mitigation strategies and host productivity traits.
541
542

Acknowledgements

PBP and TOA are grateful for support from The Research Council of Norway (FRIPRO program, PBP: 250479), as well as the European Research Commission Starting Grant Fellowship (awarded to PBP: 336355 - MicroDE), and the Novo Nordisk Foundation (awarded to PBP: 0054575 - SuPAcow). LHH was supported by The Research Council of Norway (FRIPRO program, LHH: 302639 – SeaCow), while MØA was supported by the Novo Nordisk Foundation, project No. NNF20OC006131. The experimental trial was financed by APIS-GENE (Paris, France) as part of the NutriLip project. Gas emission measurements were funded by UMR 1213 Herbivores (INRAE, Saint-Genès-Champanelle, France). The sequencing service was provided by the Norwegian Sequencing Centre (www.sequencing.uio.no), a national technology platform hosted by the University of Oslo and supported by the "Functional Genomics" and "Infrastructure" programs of the Research Council of Norway and the Southeastern Regional Health Authorities. The authors acknowledge the Orion High Performance Computing Center at the Norwegian University of Life Sciences and Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Norway for providing computational resources that have contributed to meta-omics computations reported in this paper. Mass spectrometry-based proteomic analyses were performed by The MS and Proteomics Core Facility, Norwegian University of Life Sciences (NMBU). This facility is a member of the National Network of Advanced Proteomics Infrastructure (NAPI), which is funded by the Research Council of Norway INFRASTRUKTUR-program (project number: 295910).

The authors declare no conflicts of interest.

Data Availability

Raw shotgun metagenomic data has been deposited in the National Center for Biotechnology Sequence Read Archive (NCBI-SRA) under accessions numbers SRR19524239 to SRR19524270 with links to BioProject accession number PRJNA844951. All annotated prokaryote MAGs are available publicly at DOI: 10.6084/m9.figshare.20066972.v1. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[92] partner repository with the dataset identifiers PXD034544, PXD034779 and PXD034642.

REFERENCES

1 Bergman, E. N. Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiol Rev* **70**, 567-590 (1990).

2 Wallace, R. J. *et al.* The rumen microbial metagenome associated with high methane production in cattle. *BMC Genomics* **16**, 839 (2015).

578 3 Jami, E., White, B. A. & Mizrahi, I. Potential role of the bovine rumen microbiome in
579 modulating milk composition and feed efficiency. *PLoS One* **9**, e85423 (2014).
580 4 McCann, J. C. *et al.* Induction of Subacute Ruminal Acidosis Affects the Ruminal Microbiome
581 and Epithelium. *Front Microbiol* **7**, 701 (2016).
582 5 Martin, C. *et al.* Diets supplemented with corn oil and wheat starch, marine algae, or
583 hydrogenated palm oil modulate methane emissions similarly in dairy goats and cows, but
584 not feeding behavior. *Animal Feed Science and Technology* **272**, 114783 (2021).
585 6 Seshadri, R. *et al.* Cultivation and sequencing of rumen microbiome members from the
586 Hungate1000 Collection. *Nat Biotechnol* **36**, 359-367 (2018).
587 7 Stewart, R. D. *et al.* Compendium of 4,941 rumen metagenome-assembled genomes for
588 rumen microbiome biology and enzyme discovery. *Nat Biotechnol* **37**, 953-961 (2019).
589 8 Xie, F. *et al.* An integrated gene catalog and over 10,000 metagenome-assembled genomes
590 from the gastrointestinal microbiome of ruminants. *Microbiome* **9**, 137 (2021).
591 9 Williams, A. G. & Coleman, G. S. in *The rumen microbial ecosystem* 73-139 (Springer,
592 1997).
593 10 Lin, C. Z., Raskin, L. & Stahl, D. A. Microbial community structure in gastrointestinal tracts of
594 domestic animals: Comparative analyses using rRNA-targeted oligonucleotide probes. *Fems*
595 *Microbiology Ecology* **22**, 281-294 (1997).
596 11 Hagen, L. H. *et al.* Proteome specialization of anaerobic fungi during ruminal degradation of
597 recalcitrant plant fiber. *ISME J* **15**, 421-434 (2021).
598 12 Saye, L. M. G. *et al.* The Anaerobic Fungi: Challenges and Opportunities for Industrial
599 Lignocellulosic Biofuel Production. *Microorganisms* **9**, 694 (2021).
600 13 Stabel, M. *et al.* Aestipascuomyces dupliciliberans gen. nov, sp. nov., the First Cultured
601 Representative of the Uncultured SK4 Clade from Aoudad Sheep and Alpaca.
602 *Microorganisms* **8**, 1734 (2020).
603 14 Park, T., Wijeratne, S., Meulia, T., Firkins, J. L. & Yu, Z. The macronuclear genome of
604 anaerobic ciliate Entodinium caudatum reveals its biological features adapted to the distinct
605 rumen environment. *Genomics* **113**, 1416-1427 (2021).
606 15 Henderson, G. *et al.* Rumen microbial community composition varies with diet and host, but
607 a core microbiome is found across a wide geographical range. *Sci Rep* **5**, 14567 (2015).
608 16 Ranilla, M. J., Jouany, J. P. & Morgavi, D. P. Methane production and substrate degradation
609 by rumen microbial communities containing single protozoal species in vitro. *Lett Appl*
610 *Microbiol* **45**, 675-680 (2007).
611 17 Park, T., Meulia, T., Firkins, J. L. & Yu, Z. Inhibition of the rumen ciliate Entodinium caudatum
612 by antibiotics. *Frontiers in microbiology* **8**, 1189 (2017).
613 18 Fougere, H. & Bernard, L. Effect of diets supplemented with starch and corn oil, marine
614 algae, or hydrogenated palm oil on mammary lipogenic gene expression in cows and goats:
615 A comparative study. *J Dairy Sci* **102**, 768-779 (2019).
616 19 Fougere, H., Delavaud, C. & Bernard, L. Diets supplemented with starch and corn oil, marine
617 algae, or hydrogenated palm oil differentially modulate milk fat secretion and composition
618 in cows and goats: A comparative study. *J Dairy Sci* **101**, 8429-8445 (2018).
619 20 Solomon, K. V. *et al.* Early-branching gut fungi possess a large, comprehensive array of
620 biomass-degrading enzymes. *Science* **351**, 1192-1195 (2016).
621 21 Haitjema, C. H. *et al.* A parts list for fungal cellulosomes revealed by comparative genomics.
622 *Nat Microbiol* **2**, 17087 (2017).
623 22 Youssef, N. H. *et al.* The genome of the anaerobic fungus Orpinomyces sp. strain C1A reveals
624 the unique evolutionary history of a remarkable plant biomass degrader. *Appl Environ*
625 *Microbiol* **79**, 4620-4634 (2013).
626 23 Hooker, C. A. *et al.* Hydrolysis of untreated lignocellulosic feedstock is independent of S-
627 lignin composition in newly classified anaerobic fungal isolate, Piromyces sp. UH3-1.
628 *Biotechnol Biofuels* **11**, 293 (2018).

629 24 Brown, J. L. *et al.* Cocultivation of the anaerobic fungus Caecomyces churrovis with
630 Methanobacterium bryantii enhances transcription of carbohydrate binding modules,
631 dockerins, and pyruvate formate lyases on specific substrates. *Biotechnol Biofuels* **14**, 234
632 (2021).
633 25 Wilken, S. E. *et al.* Experimentally Validated Reconstruction and Analysis of a Genome-Scale
634 Metabolic Model of an Anaerobic Neocallimastigomycota Fungus. *mSystems* **6** (2021).
635 26 Yu, Z. & Morrison, M. Improved extraction of PCR-quality community DNA from digesta and
636 fecal samples. *Biotechniques* **36**, 808-812 (2004).
637 27 Morgavi, D. P., Jouany, J. P. & Martin, C. Changes in methane emission and rumen
638 fermentation parameters induced by refaunation in sheep. *Australian Journal of*
639 *Experimental Agriculture* **48**, 69-72 (2008).
640 28 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence
641 data. *Bioinformatics* **30**, 2114-2120 (2014).
642 29 Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. MEGAHIT: an ultra-fast single-node
643 solution for large and complex metagenomics assembly via succinct de Bruijn graph.
644 *Bioinformatics* **31**, 1674-1676 (2015).
645 30 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**,
646 357-359 (2012).
647 31 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-
648 2079 (2009).
649 32 Wu, Y. W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to
650 recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605-607 (2016).
651 33 Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome
652 reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
653 34 Alneberg, J. *et al.* Binning metagenomic contigs by coverage and composition. *Nat Methods*
654 **11**, 1144-1146 (2014).
655 35 Sieber, C. M. K. *et al.* Recovery of genomes from metagenomes via a dereplication,
656 aggregation and scoring strategy. *Nat Microbiol* **3**, 836-843 (2018).
657 36 Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Res* **36**, W5-9 (2008).
658 37 Hyatt, D., LoCascio, P. F., Hauser, L. J. & Uberbacher, E. C. Gene and translation initiation site
659 prediction in metagenomic sequences. *Bioinformatics* **28**, 2223-2230 (2012).
660 38 Matsen, F. A., Kodner, R. B. & Armbrust, E. V. pplacer: linear time maximum-likelihood and
661 Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC*
662 *Bioinformatics* **11**, 538 (2010).
663 39 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM:
664 assessing the quality of microbial genomes recovered from isolates, single cells, and
665 metagenomes. *Genome Res* **25**, 1043-1055 (2015).
666 40 Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny
667 substantially revises the tree of life. *Nat Biotechnol* **36**, 996-1004 (2018).
668 41 Parks, D. H. *et al.* A complete domain-to-species taxonomy for Bacteria and Archaea. *Nature*
669 *biotechnology* **38**, 1079-1086 (2020).
670 42 Bowers, R. M. *et al.* Minimum information about a single amplified genome (MISAG) and a
671 metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* **35**, 725-
672 731 (2017).
673 43 Shaffer, M. *et al.* DRAM for distilling microbial metabolism to automate the curation of
674 microbiome function. *Nucleic Acids Res* **48**, 8883-8900 (2020).
675 44 Zhang, H. *et al.* dbCAN2: a meta server for automated carbohydrate-active enzyme
676 annotation. *Nucleic Acids Res* **46**, W95-W101 (2018).
677 45 Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res* **49**, D412-D419
678 (2021).

679 46 Suzek, B. E. *et al.* UniRef clusters: a comprehensive and scalable alternative for improving
680 sequence similarity searches. *Bioinformatics* **31**, 926-932 (2015).
681 47 Rawlings, N. D. *et al.* The MEROPS database of proteolytic enzymes, their substrates and
682 inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids*
683 *Res* **46**, D624-D632 (2018).
684 48 Aramaki, T. *et al.* KofamKOALA: KEGG Ortholog assignment based on profile HMM and
685 adaptive score threshold. *Bioinformatics* **36**, 2251-2252 (2020).
686 49 Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for
687 Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol* **428**, 726-
688 731 (2016).
689 50 Kanehisa, M., Sato, Y. & Kawashima, M. KEGG mapping tools for uncovering hidden features
690 in biological data. *Protein Sci* **31**, 47-53 (2022).
691 51 Nordberg, H. *et al.* The genome portal of the Department of Energy Joint Genome Institute:
692 2014 updates. *Nucleic Acids Res* **42**, D26-31 (2014).
693 52 Peng, X. F. *et al.* Genomic and functional analyses of fungal and bacterial consortia that
694 enable lignocellulose breakdown in goat gut microbiomes. *Nature Microbiology* **6**, 499-+
695 (2021).
696 53 Andersen, T. O., Kunath, B. J., Hagen, L. H., Arntzen, M. O. & Pope, P. B. Rumen
697 metaproteomics: Closer to linking rumen microbial function to animal productivity traits.
698 *Methods* **186**, 42-51 (2021).
699 54 Mićić, M., Whyte, J. D. & Karsten, V. in *Sample Preparation Techniques for Soil, Plant, and*
700 *Animal Samples* 99-116 (Springer, 2016).
701 55 Michalak, L. *et al.* Microbiota-directed fibre activates both targeted and secondary metabolic
702 shifts in the distal gut. *Nat Commun* **11**, 5773 (2020).
703 56 Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-
704 range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-
705 1372 (2008).
706 57 Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and
707 maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **13**, 2513-2526
708 (2014).
709 58 Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of
710 (prote)omics data. *Nat Methods* **13**, 731-740 (2016).
711 59 ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag New York, 2016).
712 60 Team, R. C. R: A language and environment for statistical computing. (2013).
713 61 Federico, A. & Monti, S. hypeR: an R package for geneset enrichment workflows.
714 *Bioinformatics* **36**, 1307-1308 (2020).
715 62 Keogh, K., Waters, S. M., Kelly, A. K. & Kenny, D. A. Feed restriction and subsequent
716 realimentation in Holstein Friesian bulls: I. Effect on animal performance; muscle, fat, and
717 linear body measurements; and slaughter characteristics. *J Anim Sci* **93**, 3578-3589 (2015).
718 63 McCabe, M. S. *et al.* Illumina MiSeq Phylogenetic Amplicon Sequencing Shows a Large
719 Reduction of an Uncharacterised Succinivibrionaceae and an Increase of the
720 Methanobrevibacter gottschalkii Clade in Feed Restricted Cattle. *PLoS One* **10**, e0133234
721 (2015).
722 64 Wessel, D. & Flugge, U. I. A method for the quantitative recovery of protein in dilute solution
723 in the presence of detergents and lipids. *Anal Biochem* **138**, 141-143 (1984).
724 65 Zougman, A., Selby, P. J. & Banks, R. E. Suspension trapping (STrap) sample preparation
725 method for bottom-up proteomics analysis. *Proteomics* **14**, 1006-1000 (2014).
726 66 Barsnes, H. & Vaudel, M. SearchGUI: A Highly Adaptable Common Interface for Proteomics
727 Search and de Novo Engines. *J Proteome Res* **17**, 2552-2555 (2018).

728 67 Fenyo, D. & Beavis, R. C. A method for assessing the statistical significance of mass
729 spectrometry-based protein identifications using general scoring schemes. *Anal Chem* **75**,
730 768-774 (2003).
731 68 Vaudel, M. *et al.* PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nat*
732 *Biotechnol* **33**, 22-24 (2015).
733 69 Millikin, R. J., Solntsev, S. K., Shortreed, M. R. & Smith, L. M. Ultrafast Peptide Label-Free
734 Quantification with FlashLFQ. *J Proteome Res* **17**, 386-391 (2018).
735 70 Harper, J. W. & Bennett, E. J. Proteome complexity and the forces that drive proteome
736 imbalance. *Nature* **537**, 328-338 (2016).
737 71 Wang, L. L. *et al.* The transcriptome of the rumen ciliate Entodinium caudatum reveals some
738 of its metabolic features. *Bmc Genomics* **20**, 1-18 (2019).
739 72 Belzecki, G., McEwan, N. R., Kowalik, B., Michalowski, T. & Miltko, R. Effect of Entodinium
740 caudatum on starch intake and glycogen formation by Eudiplodinium maggii in the rumen
741 and reticulum. *Eur J Protistol* **57**, 38-49 (2017).
742 73 Allen, M. S. Relationship between fermentation acid production in the rumen and the
743 requirement for physically effective fiber. *Journal of Dairy Science* **80**, 1447-1462 (1997).
744 74 Solomon, R. *et al.* Protozoa populations are ecosystem engineers that shape prokaryotic
745 community structure and function of the rumen microbial ecosystem. *ISME J* **16**, 1187-1197
746 (2022).
747 75 Newbold, C. J., Lassalas, B. & Jouany, J. P. The importance of methanogens associated with
748 ciliate protozoa in ruminal methane production in vitro. *Lett Appl Microbiol* **21**, 230-234
749 (1995).
750 76 Newbold, C. J., de la Fuente, G., Belanche, A., Ramos-Morales, E. & McEwan, N. R. The Role
751 of Ciliate Protozoa in the Rumen. *Front Microbiol* **6**, 1313 (2015).
752 77 Bauman, D. E., Davis, C. L. & Bucholtz, H. F. Propionate Production in Rumen of Cows Fed
753 Either a Control or High-Grain, Low-Fiber Diet. *Journal of Dairy Science* **54**, 1282-& (1971).
754 78 Jiao, H. P. *et al.* Effect of concentrate feed level on methane emissions from grazing dairy
755 cows. *J Dairy Sci* **97**, 7043-7053 (2014).
756 79 Martin, C., Morgavi, D. P. & Doreau, M. Methane mitigation in ruminants: from microbe to
757 the farm scale. *Animal* **4**, 351-365 (2010).
758 80 Finlay, B. J. *et al.* Some Rumen Ciliates Have Endosymbiotic Methanogens. *Fems*
759 *Microbiology Letters* **117**, 157-162 (1994).
760 81 Popova, M. *et al.* Effect of fibre- and starch-rich finishing diets on methanogenic Archaea
761 diversity and activity in the rumen of feedlot bulls. *Animal Feed Science and Technology* **166-**
762 **67**, 113-121 (2011).
763 82 Zhang, X. M. *et al.* Corn oil supplementation enhances hydrogen use for biohydrogenation,
764 inhibits methanogenesis, and alters fermentation pathways and the microbial community in
765 the rumen of goats. *Journal of Animal Science* **97**, 4999-5008 (2019).
766 83 Niu, P. *et al.* A Basic Model to Predict Enteric Methane Emission from Dairy Cows and Its
767 Application to Update Operational Models for the National Inventory in Norway. *Animals*
768 *(Basel)* **11**, 1891 (2021).
769 84 McAllister, T. A., Okine, E. K., Mathison, G. W. & Cheng, K. J. Dietary, environmental and
770 microbiological aspects of methane production in ruminants. *Canadian Journal of Animal*
771 *Science* **76**, 231-243 (1996).
772 85 Van Kessel, J. A. S. & Russell, J. B. The effect of pH on ruminal methanogenesis. *FEMS*
773 *microbiology ecology* **20**, 205-210 (1996).
774 86 Russell, J. B., Muck, R. E. & Weimer, P. J. Quantitative analysis of cellulose degradation and
775 growth of cellulolytic bacteria in the rumen. *FEMS Microbiol Ecol* **67**, 183-197 (2009).
776 87 Russell, J. B. The importance of pH in the regulation of ruminal acetate to propionate ratio
777 and methane production in vitro. *J Dairy Sci* **81**, 3222-3230 (1998).

88 Kleen, J. L., Hooijer, G. A., Rehage, J. & Noordhuizen, J. P. Subacute ruminal acidosis (SARA): a review. *J Vet Med A Physiol Pathol Clin Med* **50**, 406-414 (2003).

89 Coleman, G. & Sandford, D. C. The engulfment and digestion of mixed rumen bacteria and individual bacterial species by single and mixed species of rumen ciliate protozoa grown in vivo. *The Journal of Agricultural Science* **92**, 729-742 (1979).

90 Park, T. & Yu, Z. Do Ruminal Ciliates Select Their Preys and Prokaryotic Symbionts? *Front Microbiol* **9**, 1710 (2018).

91 Gutierrez, J. & Davis, R. E. Bacterial Ingestion by the Rumen Ciliates Entodinium and Diplodinium. *Journal of Protozoology* **6**, 222-226 (1959).

92 Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **47**, D442-D450 (2019).