Metabolic influence of core ciliates within the rumen microbiome

Thea O. Andersen[1], Ianina Altshuler[1], Arturo V.P. de Leon[2], Juline Walter[2], Emily McGovern[3], Kate Keogh[3], Cécile Martin[4], Laurence Bernard[4], Diego P. Morgavi[4], Tansol Park[5,6], Zongjun Li[7], Yu Jiang[7], Jeffrey L. Firkins[6], Zhongtang Yu[6], Torgeir R. Hvidsten[2], Sinead M. Waters[3], Milka Popova[4], Magnus Ø. Arntzen[2], Live H. Hagen[2], Phillip 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

[7] Center for Ruminant Genetics and Evolution, College of Animal Science and Technology, Northwest A&F University, Yangling, 712100, China

* Correspondence: phil.pope@nmbu.no

ABSTRACT

Protozoa comprise a major fraction of the microbial biomass in the rumen microbiome, of which the entodiniomorphs (order: *Entodiniomorphida*) and holotrichs (order: *Vestibuliferida*) are consistently observed to be dominant across a diverse genetic and geographical range of ruminant hosts. Despite the apparent core role that protozoal species exert, their major biological and metabolic contributions to rumen function remain largely undescribed *in vivo*. 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 protozoa occupy in the context of their microbial co-habitants. Initial proteome estimations via total protein counts and label-free quantification highlight that entodiniomorph species *Entodinium* and *Epidinium* as well as the holotrichs *Dasytricha* and *Isotricha* comprise an extensive fraction of the total rumen metaproteome. Proteomic detection of protozoal metabolism such as hydrogenases (*Dasytricha*, *Isotricha, Epidinium, Enoploplastron*), carbohydrate-active enzymes (*Epidinium, Diplodinium, Enoploplastron*, *Polyplastron*), microbial predation (*Entodinium*) and volatile fatty acid production (*Entodinium* and *Epidinium*) was observed at increased levels in high methane-emitting animals. Despite certain protozoal species having well-established reputations for digesting starch, they were

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, entodiniomorphs, holotrichs, metaproteomics, 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 are 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 assembly and binning 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]. More recently, single-cell amplified genomes (SAGs) have been recovered from rumen microbiome samples representative of 5 holotrich and 14 entodiniomorph species spanning 13 genera including *Isotricha*, *Dasytricha*, *Diplodinum*, *Enoploplastron*, *Metadinium*, *Eremoplastron*, *Ostracodinium*, *Polyplastron*, *Ophryoscolex*, *Epidinum* and *Entodinium*[15].

Entodinium and *Epidinium* are entodiniomorphs of the order *Entodiniomorphida* and represent two 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% and 16% respectively (2015 rumen

71 census: 32 animal species, 35 countries)[16]. Entodiniomorphs have been previously observed to
72 stimulate methane production[17], but while some possess hydrogenosomes such as *Epidinium*
73 *caudatum*[18], it is believed the numerically dominant *Entodinium caudatum* does not, and instead
74 encodes mitosomes and iron hydrogenases that may indicate hydrogen production, beneficial for
75 methanogenic endosymbionts[19]. Holotrich species of the order *Vestibuliferida* such as *Dasytricha* and
76 *Isotricha* are generally found at lower abundances compared to entodiniomorphs but are renowned
77 for their fermentation of soluble carbohydrates and their production of VFAs and hydrogen[20,21]. Here,
78 we present a genome-centric metaproteomics analysis of the rumen microbiome from two different
79 host species; Holstein dairy cattle (*Bos taurus*) and alpine goats (*Capra hircus*), that were fed diets of
80 first-cut grassland hay with a 45:55 forage:concentrate ratio, with concentrates supplemented with
81 either no additional lipid (CTL), or corn oil and wheat starch (COS)[5,22,23]. Previous microscopy analysis
82 revealed high counts of entodiniomorphs and to a lesser extent holotrichs, while metadata revealed
83 that animals fed COS displayed reduced methane emissions, irrespective of host. To describe how
84 these diets affect digestion and production of methane and VFAs, we sought to investigate changes in
85 function and composition in the complex rumen microbiome. By using shotgun metagenomic
86 sequencing we recovered in total 244 prokaryote metagenome-assembled genomes (MAGs) that
87 together with selected isolate- and SAG-derived eukaryote genomes[14,15,24-29] formed the database for
88 our genome-centric metaproteomic analysis of rumen protozoal species. Key protozoal metabolisms
89 were detected related to plant fiber and starch degradation, bacterial predation, and hydrogen and
90 VFA production. Despite numerous protozoal species having the genetic ability to degrade starch, our
91 analysis showed contrasting data that suggests rumen protozoa were less metabolically active in the
92 rumen microbiome of animals fed a starch-rich diet. In such a scenario other starch-degrading and/or
93 propionate and succinate producing bacterial genera, as *Prevotella* and *Fibrobacter* and members of
94 the families *Succinivibrionaceae* and *Aminobacteriaceae* appeared to be more prevalent. In concert,
95 our analysis showed that reduced methane production in both cattle and goats eating feeds
96 supplemented with COS is likely caused via a redirection of hydrogen to succinate and propionate
97 production instead of methanogenesis. Finally, by analysing a secondary, geographically independent
98 dataset, we reaffirmed our primary observations of protozoal dominance and patterns concerning
99 hydrogen, plant fiber and starch-related metabolism, thus supporting our hypothesis that this
100 protozoal species plays a core role in rumen microbiome function.
101
102 ***METHODS***
103 ***Animal trial and sample handling***

The experimental procedures were approved by the Auvergne-Rhône-Alpes Ethics Committee for Experiments on Animals (France; DGRI agreement APAFIS#3277–2015121411432527 v5) and complied with the European Union Directive 2010/63/EU guidelines. Experiments were performed at the animal experimental facilities of HerbiPôle site de Theix at the Institut National de la Recherche pour l'Agriculture, l'Alimentation l'Environnement (INRAE, Saint-Genès-Champanelle, France) from February to July 2016. Experimental design, animals and diets were as described elsewhere[5,23]. Briefly, four Holstein cows and four Alpine goats, all lactating, were enrolled in respectively two 4 x 4 Latin square design trials to study the effects of 4 diets over four 28-d experimental periods (**Supplementary Table S1**). The original study included four different experimental grassland hay basal diets with concentrates supplemented with various lipid sources; control diet with no added lipids (CTL), diet supplemented with corn oil and wheat starch (COS), diet supplemented with marine algae powder (MAP) and diet supplemented with hydrogenated palm oil (HPO) (**Supplementary Table S2**)[23]. In the present study, we focused on the CTL and COS diets, which were associated with the most extreme methane (CH_4) emission profiles in both ruminant species. The CTL diet composed of grass hay *ad libitum* with concentrates containing no additional lipid, whereas COS contained corn oil (5.0% total dry matter intake (DMI)) and wheat starch -5.0 % of total DMI (COS) (**Table 1**). Corn oil (Olvea, Saint Léonard, France) was added to the concentrate, at 5% of total DMI and contained (g/kg of total FA): 16:0 (114), 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 two 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. We acknowledge that by only including one sampling point for rumen fluid we are reducing the power that the original experimental Latin square design gives regarding minimizing individual error between the animals.

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 cattle, 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-

138 beating procedure[30]. In total, 32 rumen fluid samples (four cows and four goats fed four diets included
139 in the original study[23]) were sent to the Norwegian University of Life Sciences (NMBU) for
140 metagenomic and metaproteomic analysis. Respiration chambers were used to measure methane
141 emissions over a 5-day period, while VFA and NH_3 concentrations were determined by gas
142 chromatography using a flame ionization detector[31]. Protozoa were counted by microscopy and
143 categorized as either small entodiniomorphs (<100 µm), large entodiniomorphs (>100 µm) or as
144 holotrichs *Dasytricha* and *Isotricha*[9]. Further specifics about diets and measurements can be found in
145 Martin *et al.*[5] and VFA and methane measurements are summarized in **Table 1**.
146
147 ***Metagenomic sequencing and analysis***
148 Metagenomic shotgun sequencing was performed at the Norwegian Sequencing Centre on two lanes
149 of the HiSeq 3/4000 (Illumina) generating 150 bp paired-end reads in both lanes. Sequencing libraries
150 were prepared using the TruSeq DNA PCR-Free High Throughput Library Prep Kit (Illumina) prior to
151 sequencing. All 32 samples (four cows and four goats fed four diets included in the original study[23])
152 were run on both lanes to prevent potential lane-to-lane sequencing bias. FASTQ files were quality
153 filtered and Illumina adapters removed using Trimmomatic[32] (v. 0.36) with parameters -phred33 for
154 base quality encoding, leading and trailing base threshold set to 20. Sequences with an average quality
155 score below 15 in a 4-base sliding window were trimmed and the minimum length of reads was set to
156 36 bp. MEGAHIT[33] (v.1.2.9) was used to co-assemble reads originating from samples collected from
157 cattle and goats separately, with options –kmin-1pass, --k-list 27,37,47,57,67,77,87, --min-contig-len
158 1000 in accordance with[7]. Bowtie2[34] (v. 2.3.4.1) was used to map reads back to the assemblies and
159 SAMtools[35] (v. 1.3.1) was used to convert SAM files to BAM format and index sorted BAM files.
160
161 The two co-assemblies (one from the samples originating from cattle and the other originating from
162 the samples of goats) were binned using Maxbin2[36], MetaBAT2[37] and CONCOCT[38]. MetaBAT2 (v.
163 2.12.1) was run using parameters –minContig 2000 and –numThreads 4, Maxbin2 (v. 2.2.7) ran with
164 default parameters and -thread 4, min_contig_length 2000, and CONCOCT (v. 1.1.0) ran with default
165 parameters and –length_threshold 2000. Further, bins were filtered, dereplicated and aggregated
166 using DASTool[39](v. 1.1.2) with the parameters –write_bins 1, --threads 2 and BLAST[40] as search engine.
167 This resulted in a total of 244 dereplicated MAGs across the two host species (104 originating from
168 cattle and 140 from goats). CheckM[43](v. 1.1.3) lineage workflow was used to calculate completeness
169 and contamination of each MAG, with parameters –threads 8, --extension fa, and CoverM (v. 0.5.0)
170 (https://github.com/wwood/CoverM) was used to calculate relative abundance of each MAG, while
171 GTDB-tk[44,45] (v. 1.3.0) was used for taxonomic annotation. Approximately 90% (219 of 244) of the

172 recovered MAGs were considered high or medium quality MAGs according to MIMAGs threshold for
173 completeness and contamination for genome reporting standards[46] (**Supplementary Table S3**). Gene
174 calling and functional annotation of the final MAGs were performed using the DRAM[47] pipeline with
175 the databases dbCAN[48], Pfam[49], Uniref90[50], Merops[51], VOGdb and KOfam[52]. The translated amino acid
176 sequences from the publicly available drafted *En. caudatum* macronucleus genome were annotated
177 with the KEGG metabolic pathway database using BlastKOALA[53] by Park *et al.*[14]. Proteins with resulting
178 KEGG Orthology (KO) numbers were functionally assigned to metabolic pathways using KEGG Mapper
179 Reconstruct Tool[54].
180
181 The resulting protein sequences for each MAG, as well as those from the host genomes of goat (*Capra*
182 *hircus*, NCBI ID: 10731) and cattle (*Bos taurus*, NCBI ID: 82) were compiled into two databases, from
183 now on referred to as sample specific RUmen DataBase for Goat (RUDB-G) and sample specific RUmen
184 DataBase for Cattle (RUDB-C). In addition, both databases were supplemented with the genome of
185 the protozoa *Entodinium caudatum*[14], as well as 18 SAGs from Li et al.[15] including *Dasytricha*
186 *ruminantium* SAG3, *Diplodinium dentatum* SAGT1, *Diplodinium flabellum* SAG1, *Enoploplastron*
187 *triloricatum* SAGT1, *Entodinium bursa* SAG3, *Entodinium longinucleatum* SAG4, *Epidinium cattanei*
188 SAG3, *Epidinium caudatum* SAG1, *Eremoplastron rostratum* SAG2, *Isotricha intestinalis* SAGT2,
189 *Isotricha prostoma* SAG3, *Isotricha* sp YL-2021a SAG1, *Isotricha* sp YL-2021b SAG3, *Metadinium*
190 *minomm* SAG1, *Ophryoscolex caudatus* SAGT3, *Ostracodinium dentatum* SAG1, *Ostracodinium gracile*
191 SAG1 and *Polyplastron multivesiculatum* SAGT3. The genome of *Fibrobacter succinogenes* S85 (NCBI
192 ID: 932) was also added to both rumen databases since it is well recognized as a primary cellulolytic
193 bacterium in the rumen microbiome and has previously been observed as an active microorganism in
194 similar studies yet was not thoroughly binned as a unique MAG in this study. Nine available fungal
195 genomes downloaded from Joint Genome Institute (JGI) Mycocosm[55] were added to RUDB-C
196 (*Anaeromyces* sp. S4[25], *Caecomyces churrovis*[28], *Neocallimastix californiae*[25], *Neocallimastix lanati*[29],
197 *Piromyces finnis*[25], *Piromyces* sp. E2[25], *Piromyces* UH3-1[24,25,27], *Piromyces* eukMAG[56], *Orpinomyces*
198 sp.[26]). In total the complete databases consisted of 1 291 083 and 1 190 938 protein entries for RUDB-
199 G and RUDB-C, respectively.
200
201 ***Metaproteomic data generation***
202 To 300 µL of rumen fluid sample (in total 32 rumen fluid samples originating from four cows and four
203 goats fed four diets included in the original study[23]) 150 µL lysis buffer (30 mM DTT, 150 mM Tris-HCl
204 (pH=8), 0.3% Triton X-100, 12% SDS) was added together with 4 mm glass beads (≤ 160 µm), followed
205 by brief vortexing and resting on ice for 30 mins. Lysis was performed using FastPrep-24 Classic Grinder

206 (MP Biomedical, Ohio, USA) for 3 × 60 seconds at 4.0 meter/ second[57]. Samples were centrifuged at
207 16 000 × *g* for 15 minutes at 4°C and lysate was carefully removed. Protein concentration was
208 measured using the Bio-Rad DC™ Protein Assay (Bio-Rad, California USA) with bovine serum albumin
209 as standard. Absorbance of sample lysates was measured at A750 on BioTek Synergy H4 Hybrid
210 Microplate Reader (Thermo Fisher Scientific Inc., Massaschusetts, USA). 40-50 µg of protein was
211 prepared in SDS-buffer, heated in water bath for 5 minutes at 99°C and analysed by SDS-PAGE using
212 Any-kD Mini-PROTEAN TGX Stain-Free gels (Bio-Rad, California, USA) in a 2-minute run for sample
213 clean-up purposes, before staining with Coomassie Blue R-250. The visible bands were carefully
214 excised from the gel and divided as 1×1 mm pieces before reduction, alkylation, and digestion with
215 trypsin. Peptides were concentrated and eluted using C18 ZipTips (Merck Millipore, Darmstadt,
216 Germany) according to manufacturer's instructions, dried, and analysed by nano-LC-MS/MS using a
217 Q-Exactive hybrid quadrapole Orbitrap MS (Thermo Fisher Scientific Inc., Massaschusetts, USA) as
218 previously described[58]
219
220 ***Metaproteomic data analysis***
221 Mass spectrometry (MS) raw data were analysed with FragPipe version 19 and searched against the
222 sample-specific protein sequence database (1 291 083 and 1 190 938 protein entries for RUDB-G and
223 RUDB-C, respectively) with MSFragger[59] (**Supplementary Table S4**). The database was supplemented
224 with contaminant protein entries, such as human keratin, trypsin, and bovine serum albumin, in
225 addition to reversed sequences of all protein entries for estimation of false discovery rates (FDR)[60].
226 Oxidation of methionine and protein N-terminal acetylation were used as variable modifications, while
227 carbomidomethylation of cysteine residues were used as fixed modification. Trypsin was chosen as
228 digestive enzyme, maximum missed cleavages allowed was one and matching tolerance levels for both
229 MS and MS/MS were 20 ppm. The results were filtered to 1% FDR and quantification was done using
230 IonQuant[61] including normalization between samples and the feature 'match between runs' to
231 maximize protein identifications. Perseus[62] version 1.6.2.3 was used for further analysis. A protein
232 group was considered valid if it was quantified in at least 50% of the replicates in at least one
233 condition. Protein groups identified as potential contaminants were removed. Calculations of
234 individual genome/SAG/MAG contributions were calculated using the workflow outlined by Kleiner et
235 al.[63], which sums protein abundances (label free quantification (LFQ) values) for all proteins assigned
236 to each genome/SAG/MAG, and differential abundance between diets were detected by an unpaired
237 two-sided Student's t-test (*p*-value <0.05).
238

239 After filtration, we resolved 3657 unique protein groups across the 16 samples from cattle and 3818
240 unique protein groups across 15 samples from goats (**Supplementary Table S5** and **Supplementary**
241 **Table S6**). Dot plots for **Figure 1** were made with ggplot2[64] in R (v. 4.2.0)[65]. To determine which
242 expressed metabolic pathways *En. caudatum* were significantly enriched for in each diet/animal, we
243 used the hypeR() function from the hyperR package[66] in R which employs the hypergeometric test for
244 significance. The 'geneset' for hyperR was generated by using the keggGet() functionof KEGGREST R
245 package to retrieve entries from the KEGG database and determine which pathways the *En. caudatum*
246 KOs belong to. The geneset was then manually curated to only include metabolic pathways of interest
247 (i.e., we remove pathways such as "Huntington disease"). For the 'background' setting in hyperR(), to
248 be conservative, we used the total number of unique KOs (7592) in the *En. caudatum* genome that
249 could possibly be expressed. Significantly enriched KOs with *p*-value < 0.05 were manually filtered.
250 Both *p*-values and FDR-adjusted *p*-values are available in **Supplementary Table S7**, with FDR- adjusted
251 *p*-values < 0.05 highlighted in bold.
252
253 ***Animal trial, sample handling and metagenomic data generation for independent validating dataset***
254 Samples were also analysed from previously performed feeding experiments with Holstein Friesian
255 bulls[67]. In brief, these bulls were subjected to either *ad libitum* or restricted feeding regime in a
256 compensatory growth model detailed in Keogh *et al.*, 2015[67]. Both feeding groups received the same
257 ratio of concentrate and grass silage, respectively 70% and 30%, of which the concentrate was mainly
258 composed of starch-rich rolled barley (72.5%) and soya (22.5%). Rumen samples were collected at
259 slaughter and stored at -80C prior to metagenomics and metaproteomic analysis in this study.
260
261 Sample preparation, cell lysis and extraction of DNA was carried out as previously described by
262 McCabe *et al.*[68] Quality check of fastq files and removal of low-quality reads was performed using fastp
263 (V.0.19.5). Sequence reads were mapped against the bovine genome (ARS-UCD1.3) using minimap2
264 (V.2.16), and host sequences were removed. Reads were co-assembled using Megahit (V1.2.6) with
265 "−meta-large" pre-set option as the metagenome was complex. Metagenomic binning was applied to
266 the co-assembly using MetaBAT2 using standard parameters (V.2.12.1). MAGs were then dereplicated
267 using dRep (V.1.4.3), and the resulting MAGs were taxonomically assigned using Bin Annotation Tool
268 (BAT), available on (https://github.com/dutilh/CAT). This tool internally uses prodigal (V.2.6.3) for
269 gene prediction and DIAMOND (V.0.9.14) for the alignment against the non-redundant (nr) protein
270 database (as of Feb 2020).
271

272 Sample preparation for metaproteomics was done by lysing cells with bead beating with two glass
273 bead sizes (≤106 µm and 0.5 mm), in 100 mM Tris, pH8, 5% SDS and 10 mM DTT. A FastPrep 24
274 instrument was operated for 3 × 45 seconds at a speed of 6.5 m/s. The samples were centrifuged for
275 15 minutes at 20.000 × *g* and the protein extracts were cleaned by Wessel-Flügge precipitation[69];
276 pellets were dissolved in 5% SDS, 100 mM Tris-Cl, pH8, 10 mM DTT and kept at -20 ^{o}C until further
277 processing. Protein digestion was performed using suspension trapping (STrap)[70], dried in a SpeedVac
278 (Eppendorf Concentrator Plus) and re-dissolved in 0.05 % trifluoroacetic acid, 2% acetonitrile for
279 peptide concentration estimation using a Nanodrop One instrument, and subsequent MS/MS-
280 analysis. The samples were analyzed using an Ultimate3000 RSLCnano UHPLC coupled to a QExactive
281 hybrid quadrupole-orbitrap mass spectrometer (Thermo Fisher, Bremen, Germany) as described
282 previously[58]. MS raw data were analysed with FragPipe and searched against the Holstein Friesian
283 bulls sample-specific protein sequence database (2 533 362 protein entries) using the same workflow
284 described above.
285
286 ***RESULTS AND DISCUSSION***
287 ***Protozoal populations have large proteomes in the rumen microbiome***
288 Because of their large size protozoal species can comprise a significant fraction of the microbial
289 biomass in the rumen[9]. While the total count and diversity of protozoal species are lesser than their
290 bacterial counterparts in the rumen, their genome size and total number of genes encoded within are
291 considerably larger. In addition, due to alternative splicing and post-translational modifications, the
292 protein representation, or proteome, of protozoal populations will be larger than the number of genes
293 in their protozoal genome[71]. Thus, the amount of protein in a protozoa species can be expected to far
294 exceed the amount of protein in a bacterial species that can be identified and quantified in proteomic
295 studies. In this context, our metaproteomic data showed an extensive fraction of detectable proteins
296 affiliated to protozoal species, and other closely related species, in both dairy cattle and goats in
297 proportion to the combined bacterial species that were represented in our genome databases (**Figure**
298 **1a**). Protein detection intensity of proteins assigned to protozoa were also proportionally greater than
299 that of the bacterial fraction of the rumen microbiome in animals fed the control (CTL) diet (**Figure 1b**)
300 further supporting the significant levels of protozoal detection in our samples. Proteomes from
301 entodiniomorph species were detected at high levels in both animal hosts, with *Ep. cattanei* and *En.*
302 *caudatum* dominating the rumen metaproteome of cattle and goats respectively (**Figure 2**). In cattle
303 we also detected entodinomorphs *D. dentatum*, *En. bursa* and *Ep. caudatum* (**Figure 2a**,
304 **Supplementary Table S6**) at high levels whereas in goats *Entodinium* species *En. caudatum*, *En.*
305 *longinucleatum* and *En. bursa* where among the highest detected entodinomorphs (**Figure 2c**). More

306 than three times as many proteins from *En. caudatum* were detected in goats than cattle, however
307 this was somewhat expected given the 7x higher counts of entodiniomorph concentration (cells/mL)
308 previously observed in the goat samples compared to cattle (**Table 1**)[5]. In agreement with microscopic
309 cell counts of holotrichs, we detected proteins from both *Dasytricha* spp. and *Isotricha* spp (**Table 1**)
310 in both cattle and goats (**Figure 2a and 2c**).
311
312 ***Entodiniomorphs and holotrichs have influential roles in rumen carbohydrate metabolism***
313 Rumen protozoa have long been characterized as degraders of plant fibers as well as non-structural
314 storage polysaccharides such as starch[72]. Closer examination of the proteomes from both highly
315 detected *Epidinium species Ep. cattanei* and *Ep. caudatum* in cattle highlighted a plethora of detected
316 carbohydrate active enzymes (CAZymes) targeting lignocellulosic polysaccharides such as cellulose
317 (GH5, GH9), xylan (GH10, GH11, GH30, CE1), mannan (GH26) and fiber-derived oligosaccharides (GH2,
318 GH3, GH94) (**Figure 2a**, **Supplementary Table S6**). In goats similar CAZy families were detected but
319 from different entodinomorphs species, such as *E. triloricatum* and *P. multivesiculatum* (**Figure 2c**,
320 **Supplementary Table S6**). Our *in vivo* CAZyme observations were in strong agreement to long standing
321 *in vitro* data, which have shown entodinomorphs such as *Epidinium*, *Polyplastron* and *Enoploplastron*
322 have greater endoglucanase and xylanase activity, while *Entodinium* spp. have only weak activity[72].
323 Similarly, it has previously been shown *in vitro* that holotrichs such as *Dasytricha* spp. have glucosidase
324 and cellobiosidase activity but negligible fibrolytic activity[73], which our *in vivo* data also supported
325 (**Figure 2a and 2c**). The composition of protozoal species metabolising starch (both degrading and
326 synthesizing) also varied between the two animal hosts with goats dominated by *Entodinium* spp. and
327 *D. ruminantium* while protozoal starch metabolism in cattle was *Epidinium* spp., *D. ruminantium and*
328 *Isotricha* spp (**Figure 2**, **Supplementary Table S6**). While bacterial CAZymes targeting lignocellulose
329 were also detected (e.g. GH48 cellulases from *Ruminococcaceae* populations in cattle), CAZymes
330 targeting lignocellulose and starch were detected at higher frequency and protein quantification levels
331 in protozoal species suggesting they have a central role in ruminal fibre digestion.
332
333 ***Metabolism of En. caudatum shows predatory activity and metabolism of VFAs***
334 *En. caudatum* is renowned for its predatory activity and is acknowledged as the most abundant
335 protozoa in the rumen, whereby it has been estimated that 0.1% of rumen prokaryotes are digested
336 by the rumen protozoal population every minute[74]. The dominance of *En. caudatum* in our data
337 presented an opportunity to better understand the metabolic influence of this universal species. While
338 previous efforts have investigated the genome and transcriptome of *En. caudatum* grown in
339 monoculture[14,75], our metaproteomic analysis sought to reveal *in vivo* metabolism and functions of *En.*

340 *caudatum* within the rumen microbiome. In accordance with Wang *et al.*[75], our metaproteomic
341 analysis revealed expressed proteins attributed to metabolic pathways such as carbon metabolism,
342 glycolysis/gluconeogenesis, starch and sucrose (and glycogen) metabolism, pyruvate metabolism,
343 oxidative phosphorylation and production of alcohol (**Supplementary Table S6**). Wang *et al.* found
344 that as for most rumen microbes, *En. caudatum* uses carbohydrates such as starch as its primary
345 substrate, as well as cellulose and hemicellulose to a certain degree[75], and their transcript analyses
346 showed that *En. caudatum* had high levels of expression of amylases and low-level expression of
347 hemicellulases, cellulases and pectinases. Similarly, our metaproteomic analysis reveals expression of
348 amylases by *En. caudatum* that are predicted to enable *En. caudatum* to engulf and degrade starch
349 granules to simpler sugars and to produce glycogen, its most important storage carbohydrate[76].
350 However, no detection of *En. caudatum* CAZymes related to hemicellulose or pectin were observed in
351 any of our metaproteomes, suggesting that it is not engaging in the deconstruction of these
352 carbohydrates at the time our samples were collected for analysis (before feeding). It should be noted
353 that ruminal fermentation activity as well as production of VFAs and methane will be at its highest
354 after feeding, as a result of an increased availability of fermentable substrate[77]. While sampling time
355 can influence the recovered microbial composition and hence function, any differences in metabolic
356 parameters or species abundance in this study is comparable across both diets given the consistent
357 sampling times.
358
359 While monoculture cultures of *En. caudatum* have not been established to verify the VFAs it can
360 produce, Wang *et al.* found transcripts of enzymes involved in fermentative formation of acetate and
361 butyrate[75]. Similarly, we detected proteins inferred in metabolism of acetate, butyrate, and alcohol in
362 *En. caudatum*. Goats fed the CTL diet had a significantly higher proportion of proteins from *En.*
363 *caudatum* and concurrently had increased relative levels of acetate and butyrate compared to animals
364 fed the corn oil and wheat starch diet (COS), which had fewer *En. caudatum* proteins and lower
365 acetate/butyrate levels (**Figure 2** and **Table 1**). As *En. caudatum* populations were seemingly most
366 abundant in goats fed the CTL diet, we used these metaproteomes to reconstruct metabolic features
367 (**Figure 3**). Of the 538 *En. caudatum* proteins identified in CTL-fed goats, 244 had unique KO numbers
368 assigned, from which KEGG Mapper reconstructions[54] enabled functional assignment of 217 proteins
369 to metabolic pathways. Our metabolic reconstructions showed expressed proteins involved in
370 endocytosis, phagosome and lysosome processes for predatory activity, engulfment, and digestion of
371 bacteria (**Figure 3**, **Supplementary Table S6**). For the rumen samples used in this study Martin *et al.*
372 previously observed higher NH_3 concentrations in goats compared to cattle and hypothesised that it
373 might have resulted from increased bacterial protein breakdown and feed protein degradability due

374 to higher density of entodiniomorphs known for their predatory activity[5]. In support of these
375 observations, we performed metaproteomic pathway enrichment analysis of *En. caudatum* (**Figure 3,**
376 **Supplementary Table S7**) proteins detected in goats, which revealed significantly enriched
377 endocytosis metabolism. Other biological processes such as glycolysis and starch and sucrose
378 metabolism were significantly enriched pathways in *En. caudatum* (**Supplementary Table S7**).
379 Although suspected of having metabolic interactions with methanogenic archaea, several protozoal
380 populations such as *Entodinium* are hypothesized as having associations with certain members of the
381 Gram-negative *Gammaproteobacteria*, which multiple studies have speculated are resistant to
382 protozoal engulfment[78-80]. In contrast, Gutierrez and Davis previously demonstrated that *Entodinium*-
383 species engulf Gram positive starch-degraders[80]. In the context of our data, we speculate that CTL fed
384 animals provided *En. caudatum*-like populations optimal conditions for predation, whereas increased
385 starch levels in the COS diets facilitated increases in Gram-negative species that are possibly resistant
386 to protozoal engulfment and/or reduced pH levels, leading to sub-optimal conditions for *Entodinium*
387 and a decrease in its levels (**Figure 2**).
388
389 ***Protozoa were less active in diets supplemented with starch regardless of their starch metabolising***
390 ***reputations.***
391 The changes in VFA and methane levels in animals fed the high starch COS diet, previously measured
392 by Martin *et al.*[5], suggested significant alterations in composition and thus functions of the rumen
393 microbiome irrespective of host species. In particular, a decrease in proportions of acetate and
394 butyrate, decrease in the acetate:propionate ratio and an increase in relative propionate levels were
395 observed in animals fed the COS diet, compared to the CTL diet (**Table 1**). Diets that are high in starch
396 content or with low forage:concentrate ratios have previously been shown to result in higher
397 production of propionate and succinate, as they are easily fermented in the rumen and accordingly
398 have high passage rates[81,82]. We therefore leveraged our genome-centric metaproteomic data from
399 both cattle (**Figure 2a-b)** and goats (**Figure 2c-d**) fed either the COS or CTL diet to gain an overview of
400 protein expression from individual populations. We specifically focused on pathways involved in the
401 degradation of starch (CTL: corn starch, COS: corn + wheat starch) to pyruvate through glycolysis and
402 finally formation of acetate, butyrate, and propionate (via succinate). Irrespective of host, and despite
403 their starch-metabolising reputation[75,76], *Entodinium*, *Epidinium*, *Isotricha* and *Dasytricha* spp had
404 lower abundance and less proteins involved in starch degradation in animals fed the COS diet
405 compared to those fed the CTL diet (**Figure 2a** and **2c**). Further, we observed opposing patterns for
406 proteins identified as *Entodinium* and *Epidinium*-spp. involved in glycolysis, and production of

407 pyruvate, acetate, and butyrate, which were detected in higher levels in both cattle and goats fed the
408 CTL diet compared to the COS diet.
409
410 While several putative protozoal amylases were detected across all animals and diets, their
411 quantification levels (i.e., protein abundances) did not increase as expected when higher levels of
412 starch were available (**Figure 2a** and **2c**). We therefore hypothesized that the observed shift in VFA
413 profiles in response to increased starch was additionally influenced by the bacterial fraction of the
414 rumen microbiome. In contrast to lower levels of protozoa in the animals fed the COS diet, we
415 observed an increase in bacterial proteome detection irrespective of host (**Figure 4**), including
416 suspected starch-degrading and succinate- and propionate-producing bacterial species (**Figure 2b** and
417 **2d**). For example, starch fermentation pathways from population genomes affiliated with the
418 *Succinivibrionaceae* family, *Prevotella* species and, additionally for goats, members of the
419 *Aminobacteriaceae* families, were detected at higher proteomic levels in the animals fed the COS diet
420 compared to those fed the CTL diet (**Figure 2b** and **2d**).
421
422 ***Protozoa are less active in animals that produce lower methane yield***
423 For the animals sampled in this study, Martin *et al.* demonstrated a ~25-30% reduction in methane
424 emissions in both cattle and goats fed the COS diet compared to the control (**Table 1**)[5]. Previous
425 comparisons between defaunated and faunated animals have shown decrease in methane production
426 in protozoa-free ruminants, suggesting symbiotic interactions between methanogenic archaea and
427 protozoal species[78]. Methanogen's epi- and endo-symbiotic relationships with protozoa have also
428 been suggested to contribute to 9-37% of rumen methanogenesis[78,83-85]. Moreover, studying
429 microcosms with the presence and absence of protozoal species Solomon *et al.* reported higher levels
430 of acetate and butyrate in microcosms with protozoa present in addition to increased methane
431 emissions[78], which supports the main findings of animals fed the CTL diet in this study (**Table 1**).
432 Several of the protozoal species detected in our metaproteomic data are known to have
433 hydrogenosomes and closer examination showed that [FeFe]-hydrogenases from both holotrich and
434 entodinomorph genera, such as *Dasytricha, Isotricha, Diplodinium, Epidinium, Entodinium* and
435 *Enoploplastron*, were detected at higher quantification levels in animals fed the CTL compared to
436 those fed COS (**Figure 2a and 2c**, **Supplementary Table S6**). We also observed partial evidence in our
437 proteomic data that the hydrogenosome-lacking *En. caudatum* makes contributions to ruminal
438 hydrogen production, supporting previous reports associating its abundance with higher methane
439 levels[78,85,86]. Only one of its eight [FeFe]-hydrogenases were detected in goats (absent in cattle),
440 though it showed no changes in protein abundance in either the high (CTL) or low (COS) methane

441 yielding animals. Given our observations, we speculate that predicted protozoal-methanogen
442 relationships that lead to increased methane levels in this study are centred on hydrogen transfer.
443
444 The addition of lipids in diets for ruminants have been largely studied as a methane mitigation
445 strategy[87,88]. However, it is widely believed that lipids are not fermented in the rumen and rather
446 modified/hydrogenated and hence do not contribute to hydrogen production or methane production
447 in the rumen. While biohydrogenation of unsaturated fatty acids such as corn oil can serve as a
448 hydrogen sink, it has been observed that very small amounts of metabolic hydrogen (1-2%) are used
449 for biohydrogenation[87,88]. Several studies have also shown that the CH_4 mitigation effects of lipids are
450 dependent on both dose and fatty acid composition[89,90]. For example, medium chain fatty acids
451 (MCFAs), such as lauric acid, have been proven to be a more effective inhibitor of protozoa than long
452 chain polyunsaturated fatty acids (PUFAs), such as corn oil, which mainly consists of linoleic acid[88-91].
453 In a study conducted by Zhang *et al.*[90], goats fed corn oil as a supplement decreased ruminal hydrogen
454 concentrations and total methane emissions. Nevertheless, there was seemingly no effect on rumen
455 protozoal populations, which suggests that the corn oil dosage used in the Zhang study did not act as
456 an anti-protozoal agent[90]. In addition, high levels of PUFAs in the diet will also likely impact the
457 metabolism of various bacterial and archaeal species[88-90], particularly keystone fibre degrading Gram-
458 positive bacterial species that produce various levels of VFAs and hydrogen[92,93]. Collectively, these
459 previous findings were in agreement with the decreased CH_4 production in cattle and goats in this
460 study fed the COS diet, which was observed to additionally impact other bacterial metabolism and
461 ruminal fermentation parameters, such as increased propionate and decreased butyrate and acetate
462 levels (**Table 1**)[5]. Thus, while we cannot unequivocally rule out that diets supplemented with lipids at
463 higher levels could be directly affecting the abundance of protozoa, we believe it is also possible that
464 corn oil is influencing other fibre-degrading and hydrogen-producing species in the rumen, hence
465 effecting ruminal fermentation and methane emissions[88,94]. Nonetheless, validation of these results
466 under different conditions and extended sampling is necessary to verify our findings.
467
468 Increases in dietary starch for ruminants is known to stimulate the propionate and succinate pathways
469 of starch-degrading bacteria, which due to their net incorporation of metabolic hydrogen [H]
470 represent a [H] sink in rumen fermentation besides hydrogenotrophic methanogenesis[83,95]. Diets rich
471 in starch are more fermentable in the rumen, which can decrease the ruminal pH to levels that can
472 inhibit methanogenic archaea and fibre-degrading bacterial species[96,97]. Yet, lowered pH levels in the
473 rumen can also lead to clinical (or sub-clinical in most production scenarios) ruminal acidosis[98,99].
474 Hence, high concentrate diets, which increase production of propionate at the expense of methane,

475 does not necessarily opt for a viable methane mitigation strategy in the long term. Our results suggests
476 that decreased methanogenesis in COS-fed animals is likely due to a decrease in available hydrogen
477 and/or decrease in pH levels, which we predict is caused by the lower levels of hydrogen-producing
478 protozoa and bacteria as well as the metabolism of dominant bacterial populations that likely do not
479 produce exogenous hydrogen due to their own [H]-utilizing succinate and propionate metabolism.
480 This prediction was partially supported by the detection of several putative hydrogenases believed to
481 use molecular hydrogen in putative fumarate-reducing *Aminobacteriaceae* populations that were
482 dominant in the COS-fed goat rumen. In particular, two [FeFe]-hydrogenases were detected in COS-
483 fed animals from RUDB_G023 (ORF_4456741_13) and RUDBG016 (ORF_ 1143158_19) that were
484 annotated as NADP-reducing hydrogenase subunits (EC:1.12.1.3) (**Figure 2d**, **Supplementary Table**
485 **S6**).
486
487 ***Protozoal dominance is validated in geographically independent datasets***
488 To further test our hypothesis that protozoal species play a central role in the rumen ecosystem, we
489 explored additional metagenome-centric metaproteomic datasets originating from an independent
490 feeding experiment performed in Ireland on 60 Holstein Friesian bulls[67]. In brief, these bulls were
491 subjected to the same ratio of concentrate and grass silage at either an *ad libitum* or restricted feeding
492 regime in a compensatory growth model detailed in Keogh et al, 2015[67]. We applied the same strategy
493 as for the described Holstein dairy cattle and alpine goats to resolve the metaproteomic dataset for a
494 subset of 15 animals (7 restricted and 8 *ad libitum*) against 781 reconstructed sample-specific MAGs
495 (RUDB-HF), which were supplemented with the genome of *En. caudatum,* 18 protozoal SAGs[15] as well
496 as genomes of available anaerobic fungi. This collection of microbial prokaryote and eukaryote
497 genomes was then used as a sequence database for the generated protein spectra. Consistent with
498 our previous observation in dairy cattle, a substantial proportion of the detected proteins were
499 affiliated to *Epidinium* (*Ep. cattanei* and *Ep. caudatum*), *Entodinium* (*En. caudatum* and *En. bursa*) and
500 *Isotricha* spp. providing further support that protozoa are an important and metabolically active
501 contributor to the rumen microbiome (**Figure 5, Supplementary Table S8**). The protein quantification
502 (measured as sum of LFQ intensities affiliated to each MAG/genome, averaged for each diet) was also
503 higher in the rumen sample from bulls on the restricted diet, which likely had less starch available
504 compared to the *ad libitum* group and a higher retention time (**Figure 5a, Supplementary Table S8**).
505 A previously published 16S rRNA amplicon investigation of the phylogenetic differences between the
506 rumen microbiomes of these two diet groups highlighted an increase in *Succinivibrionaceae* in the
507 starch-rich *ad libitum* diet[68]. Our metaproteomic analysis confirmed a significantly higher (unpaired t-
508 test, p-value < 0.05) proteomic detection of several *Succinivibrionaceae*-MAGs under the *ad libitum*

509 group (**Supplementary Table S8**), accompanied with a reduced acetate:propionate ratio in the rumen,
510 which is often associated with increased feed efficiency and reduced production of methane[68]. These
511 observations largely mirror the dominance of *Succinivibrionaceae*-MAGs in the dairy cattle and goats
512 fed the COS diet, further strengthening our hypothesis that *Entodinium* and *Epidinium* spp do not
513 metabolically respond to increases in available starch in the host animals' diets and have other roles
514 than being a primary starch degrader.
515
516 To assess whether similar patterns of diet-induced microbiome shifts were consistent across the two
517 geographically distinct datasets used in this study, we compared ratios of protein abundance level for
518 commonly observed taxonomic families. Spearman's rank correlation coefficients were calculated for
519 the comparisons, where two tests were performed: correlation between common taxonomic families
520 (rarely observed families with < 2 MAGs with detected proteins across both conditions were excluded)
521 in the rumen of beef (RUDB_HF) and dairy (RUDB_C) cattle as well as beef cattle (RUDB_HF) and alpine
522 goats (RUDB_G), respectively. Changes in protein abundance levels from a total of 16 taxonomic
523 families were found in common between the rumen microbiomes of beef and dairy cattle (significant
524 correlation: p-value < 0.05 and correlation coefficient rho of 0.68), further strengthening our
525 hypothesis that similar structural rumen microbiome shifts occur across independent datasets of
526 cattle subjected to diets of high and low starch (**Figure S1, Supplementary Table S9**). However,
527 unsurprisingly no significant correlations were observed between the rumen microbiomes of goats
528 (RUDB_G) and beef cattle (RUDB_HF) (p-value > 0.05 and correlation coefficient rho of 0.19,
529 **Supplementary Table S9, Supplementary Figure S1**), despite similar trends in protein abundance
530 levels across taxa being observed (**Figure 2, Figure 5,** and **Supplementary Figure S1)**.
531
532 In conclusion, by using a (meta)genome-centric metaproteomics approach we primarily investigated
533 the role of the rumen protozoa in the rumen microbiome of beef and dairy cattle as well as dairy goats
534 that were subjected to varying dietary conditions. We showed that the proteomes of core
535 entodinomorph and holotrich genera such as *Entodinium*, *Epidinium*, *Dasytricha* and *Isotricha*
536 constitute a substantial fraction of the recovered rumen microbial proteome, which supports previous
537 16S/18S rRNA gene-based rumen census data that have highlighted their global dominance across a
538 plethora of ruminant species. In both animal hosts, protozoal CAZymes targeting lignocellulose were
539 detected though more frequently and at higher quantification levels in animals fed the CTL diet
540 compared to the COS supplemented diet, presumably due to lower cell density in the later (**Table 1**).
541 Proteins identified as *En. caudatum*, *Epidinium, Isotricha* and *Dasytricha* spp. were surprisingly
542 detected at lower levels in animals that were fed increased levels of wheat starch, despite their

543 reputable starch-metabolising capabilities (**Figures 2-5**). We hypothesize that protozoa were being out
544 competed by Gram-negative bacterial species (e.g. *Succinivibrionaceae* in cattle and/or
545 *Aminobacteriaceae* in goats) in the COS-fed animals, which were possibly resistant to protozoal
546 engulfment[78-80] and/or lower pH levels, creating sub-optimal conditions for starch-degrading
547 protozoal populations. We also observed lower detection of protozoal [FeFe]-hydrogenases in COS-
548 fed animals that measured a ~25-30% lower CH_4 yield at the time of sampling in this study (prior to
549 feeding), further suggesting the specific conditions that enable succinate- and propionate-producing
550 populations to flourish subsequently exert an impact on hydrogen and methane metabolisms in the
551 rumen microbiome. For animals that were additionally fed corn oil, it is acknowledged that diets
552 supplemented with lipids can have effects on the decreased abundance of protozoa, however our
553 study did not provide evidence of a direct effect. While much work is still needed to confirm our
554 abovementioned hypotheses, our integrated metaproteomics approaches have demonstrated the
555 future importance of including eukaryote populations for accurate and meaningful analyses of the
556 rumen microbiome and its impact on GHG mitigation strategies and host productivity traits.
557

558 **Acknowledgements**

559 PBP and TOA are grateful for support from The Research Council of Norway (FRIPRO program, PBP:
560 250479), as well as the European Research Commission Starting Grant Fellowship (awarded to PBP:
561 336355 - MicroDE), and the Novo Nordisk Foundation (awarded to PBP: 0054575 - SuPAcow). LHH was
562 supported by The Research Council of Norway (FRIPRO program, LHH: 302639 – SeaCow), while MØA
563 was supported by the Novo Nordisk Foundation, project No. NNF20OC006131. The experimental trial
564 was financed by APIS-GENE (Paris, France) as part of the NutriLip project. Gas emission measurements
565 were funded by UMR 1213 Herbivores (INRAE, Saint-Genès-Champanelle, France). The sequencing
566 service was provided by the Norwegian Sequencing Centre (www.sequencing.uio.no), a national
567 technology platform hosted by the University of Oslo and supported by the "Functional Genomics"
568 and "Infrastructure" programs of the Research Council of Norway and the Southeastern Regional
569 Health Authorities. The authors acknowledge the Orion High Performance Computing Center at the
570 Norwegian University of Life Sciences and Sigma2 - the National Infrastructure for High Performance
571 Computing and Data Storage in Norway for providing computational resources that have contributed
572 to meta-omics computations reported in this paper. Mass spectrometry-based proteomic analyses
573 were performed by The MS and Proteomics Core Facility, Norwegian University of Life Sciences
574 (NMBU). This facility is a member of the National Network of Advanced Proteomics Infrastructure
575 (NAPI), which is funded by the Research Council of Norway INFRASTRUKTUR-program (project
576 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[100] partner repository with the dataset identifiers PXD040467, PXD040454 and PXD040349. NCBI accession numbers for the protozal SAGs used in this study are as follows: *Dasytricha ruminantium* SAG3 (JAJJKK000000000), *Diplodinium dentatum* SAGT1 (JAJJLX000000000), *Diplodinium flabellum* SAG1 (JAJJLB000000000), *Enoploplastron triloricatum* SAGT1 (JAJNAB000000000), *Entodinium bursa* SAG3 (JAJJKR000000000), *Epidinium cattanei* SAG3 (JAJJKG000000000), *Epidinium caudatum* SAG1 (JAJMZR000000000), *Eremoplastron rostratum* SAG2 (JAJMZO000000000), *Isotricha intestinalis* SAGT2 (JAJJLT000000000), *Isotricha prostoma* SAG3 (JAJMZS000000000), *Isotricha* sp. YL-2021a SAG1 (JAJJLH000000000), *Isotricha* sp. YL-2021b SAG3 (JAJJKW000000000), *Metadinium minomm* SAG1 (JAJMZT000000000), *Ophryoscolex caudatus* SAGT3 (JAJJLZ000000000), *Ostracodinium dentatum* SAG1 (JAJJKF000000000), *Ostracodinium gracile* SAG1 (JAJJKZ000000000), *Polyplastron multivesiculatum* SAGT3 (JAJJLS000000000).

Data access for reviewers:

SRA:
https://dataview.ncbi.nlm.nih.gov/object/PRJNA844951?reviewer=7q1stta7ga42mq5tq27ifj7n85

Figshare - https://figshare.com/s/cec40f220249f3ef848c
- https://figshare.com/s/ad4eb7e1b7db17f0535b

PRIDE - Reviewer must download PRIDE Inspector Tool
https://www.ebi.ac.uk/pride/markdownpage/prideinspector

- MS Rawfiles for samples from CATTLE:
  Username: reviewer_pxd040467@ebi.ac.uk
  Password: KPtwrE0J
- MS Rawfiles for samples from GOAT:
  Username: reviewer_pxd040454@ebi.ac.uk
  Password: FjGXlkSf
- MS Rawfiles for samples from Holstein Friesian BULLS:
  Username: reviewer_pxd040349@ebi.ac.uk
  Password: I1rcyaFi

617 **REFERENCES**

618 1 Bergman, E. N. Energy contributions of volatile fatty acids from the gastrointestinal tract in
619 various species. *Physiol Rev* **70**, 567-590 (1990).
620 2 Wallace, R. J. *et al.* The rumen microbial metagenome associated with high methane
621 production in cattle. *BMC Genomics* **16**, 839 (2015).
622 3 Jami, E., White, B. A. & Mizrahi, I. Potential role of the bovine rumen microbiome in
623 modulating milk composition and feed efficiency. *PLoS One* **9**, e85423 (2014).
624 4 McCann, J. C. *et al.* Induction of Subacute Ruminal Acidosis Affects the Ruminal Microbiome
625 and Epithelium. *Front Microbiol* **7**, 701 (2016).
626 5 Martin, C. *et al.* Diets supplemented with corn oil and wheat starch, marine algae, or
627 hydrogenated palm oil modulate methane emissions similarly in dairy goats and cows, but
628 not feeding behavior. *Anim Feed Sci Technol* **272**, 114783 (2021).
629 6 Seshadri, R. *et al.* Cultivation and sequencing of rumen microbiome members from the
630 Hungate1000 Collection. *Nat Biotechnol* **36**, 359-367 (2018).
631 7 Stewart, R. D. *et al.* Compendium of 4,941 rumen metagenome-assembled genomes for
632 rumen microbiome biology and enzyme discovery. *Nat Biotechnol* **37**, 953-961 (2019).
633 8 Xie, F. *et al.* An integrated gene catalog and over 10,000 metagenome-assembled genomes
634 from the gastrointestinal microbiome of ruminants. *Microbiome* **9**, 137 (2021).
635 9 Williams, A. G. & Coleman, G. S. in *The rumen microbial ecosystem* 73-139 (Springer,
636 1997).
637 10 Lin, C., Raskin, L. & Stahl, D. A. Microbial community structure in gastrointestinal tracts of
638 domestic animals: comparative analyses using rRNA-targeted oligonucleotide probes. *FEMS*
639 *Microbiol Ecol* **22**, 281-294 (1997).
640 11 Hagen, L. H. *et al.* Proteome specialization of anaerobic fungi during ruminal degradation of
641 recalcitrant plant fiber. *ISME J* **15**, 421-434 (2021).
642 12 Saye, L. M. G. *et al.* The Anaerobic Fungi: Challenges and Opportunities for Industrial
643 Lignocellulosic Biofuel Production. *Microorganisms* **9**, 694 (2021).
644 13 Stabel, M. *et al.* Aestipascuomyces dupliciliberans gen. nov, sp. nov., the First Cultured
645 Representative of the Uncultured SK4 Clade from Aoudad Sheep and Alpaca.
646 *Microorganisms* **8**, 1734 (2020).
647 14 Park, T., Wijeratne, S., Meulia, T., Firkins, J. L. & Yu, Z. The macronuclear genome of
648 anaerobic ciliate *Entodinium caudatum* reveals its biological features adapted to the distinct
649 rumen environment. *Genomics* **113**, 1416-1427 (2021).
650 15 Li, Z. *et al.* Genomic insights into the phylogeny and biomass-degrading enzymes of rumen
651 ciliates. *ISME J* **16**, 2775–2787 (2022).
652 16 Henderson, G. *et al.* Rumen microbial community composition varies with diet and host, but
653 a core microbiome is found across a wide geographical range. *Sci Rep* **5**, 14567 (2015).
654 17 Ranilla, M. J., Jouany, J. P. & Morgavi, D. P. Methane production and substrate degradation
655 by rumen microbial communities containing single protozoal species in vitro. *Lett Appl*
656 *Microbiol* **45**, 675-680 (2007).
657 18 Yarlett, N., Coleman, G. S., Williams, A. G. & Lloyd, D. Hydrogenosomes in known species of
658 rumen entodiniomorphid protozoa. *FEMS Microbiol Lett* **21**, 15-19 (1984).
659 19 Park, T., Meulia, T., Firkins, J. L. & Yu, Z. Inhibition of the Rumen Ciliate Entodinium
660 caudatum by Antibiotics. *Front Microbiol* **8** (2017).
661 20 Belanche, A., de la Fuente, G. & Newbold, C. J. Study of methanogen communities associated
662 with different rumen protozoal populations. *FEMS Microbiol Ecol* **90**, 663-677 (2014).
663 21 Hillman, D., Lloyd, K., Yarlett, N. & Williams, A. G. Hydrogen Production by Rumen Holotrich
664 Protozoa: Effects of Oxygen and Implications for Metabolic Control by In Situ Conditions. *J*
665 *Protozool* **36**, 205-213 (1989).

666 22 Fougere, H. & Bernard, L. Effect of diets supplemented with starch and corn oil, marine
667 algae, or hydrogenated palm oil on mammary lipogenic gene expression in cows and goats:
668 A comparative study. *J Dairy Sci* **102**, 768-779 (2019).
669 23 Fougere, H., Delavaud, C. & Bernard, L. Diets supplemented with starch and corn oil, marine
670 algae, or hydrogenated palm oil differentially modulate milk fat secretion and composition
671 in cows and goats: A comparative study. *J Dairy Sci* **101**, 8429-8445 (2018).
672 24 Solomon, K. V. *et al.* Early-branching gut fungi possess a large, comprehensive array of
673 biomass-degrading enzymes. *Science* **351**, 1192-1195 (2016).
674 25 Haitjema, C. H. *et al.* A parts list for fungal cellulosomes revealed by comparative genomics.
675 *Nat Microbiol* **2**, 17087 (2017).
676 26 Youssef, N. H. *et al.* The genome of the anaerobic fungus Orpinomyces sp. strain C1A reveals
677 the unique evolutionary history of a remarkable plant biomass degrader. *Appl Environ*
678 *Microbiol* **79**, 4620-4634 (2013).
679 27 Hooker, C. A. *et al.* Hydrolysis of untreated lignocellulosic feedstock is independent of S-
680 lignin composition in newly classified anaerobic fungal isolate, Piromyces sp. UH3-1.
681 *Biotechnol Biofuels* **11**, 293 (2018).
682 28 Brown, J. L. *et al.* Cocultivation of the anaerobic fungus Caecomyces churrovis with
683 Methanobacterium bryantii enhances transcription of carbohydrate binding modules,
684 dockerins, and pyruvate formate lyases on specific substrates. *Biotechnol Biofuels* **14**, 234
685 (2021).
686 29 Wilken, S. E. *et al.* Experimentally Validated Reconstruction and Analysis of a Genome-Scale
687 Metabolic Model of an Anaerobic Neocallimastigomycota Fungus. *mSystems* **6**, e00002-
688 00021 (2021).
689 30 Yu, Z. & Morrison, M. Improved extraction of PCR-quality community DNA from digesta and
690 fecal samples. *Biotechniques* **36**, 808-812 (2004).
691 31 Morgavi, D. P., Jouany, J.-P. & Martin, C. Changes in methane emission and rumen
692 fermentation parameters induced by refaunation in sheep. *Aust J Exp Agric* **48**, 69-72 (2008).
693 32 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence
694 data. *Bioinformatics* **30**, 2114-2120 (2014).
695 33 Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. MEGAHIT: an ultra-fast single-node
696 solution for large and complex metagenomics assembly via succinct de Bruijn graph.
697 *Bioinformatics* **31**, 1674-1676 (2015).
698 34 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**,
699 357-359 (2012).
700 35 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-
701 2079 (2009).
702 36 Wu, Y. W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to
703 recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605-607 (2016).
704 37 Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome
705 reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
706 38 Alneberg, J. *et al.* Binning metagenomic contigs by coverage and composition. *Nat Methods*
707 **11**, 1144-1146 (2014).
708 39 Sieber, C. M. K. *et al.* Recovery of genomes from metagenomes via a dereplication,
709 aggregation and scoring strategy. *Nat Microbiol* **3**, 836-843 (2018).
710 40 Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Res* **36**, W5-9 (2008).
711 41 Hyatt, D., LoCascio, P. F., Hauser, L. J. & Uberbacher, E. C. Gene and translation initiation site
712 prediction in metagenomic sequences. *Bioinformatics* **28**, 2223-2230 (2012).
713 42 Matsen, F. A., Kodner, R. B. & Armbrust, E. V. pplacer: linear time maximum-likelihood and
714 Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC*
715 *Bioinformatics* **11**, 538 (2010).

716 43 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM:
717 assessing the quality of microbial genomes recovered from isolates, single cells, and
718 metagenomes. *Genome Res* **25**, 1043-1055 (2015).
719 44 Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny
720 substantially revises the tree of life. *Nat Biotechnol* **36**, 996-1004 (2018).
721 45 Parks, D. H. *et al.* A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat*
722 *Biotechnol* **38**, 1079-1086 (2020).
723 46 Bowers, R. M. *et al.* Minimum information about a single amplified genome (MISAG) and a
724 metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* **35**, 725-
725 731 (2017).
726 47 Shaffer, M. *et al.* DRAM for distilling microbial metabolism to automate the curation of
727 microbiome function. *Nucleic Acids Res* **48**, 8883-8900 (2020).
728 48 Zhang, H. *et al.* dbCAN2: a meta server for automated carbohydrate-active enzyme
729 annotation. *Nucleic Acids Res* **46**, W95-W101 (2018).
730 49 Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res* **49**, D412-D419
731 (2021).
732 50 Suzek, B. E. *et al.* UniRef clusters: a comprehensive and scalable alternative for improving
733 sequence similarity searches. *Bioinformatics* **31**, 926-932 (2015).
734 51 Rawlings, N. D. *et al.* The MEROPS database of proteolytic enzymes, their substrates and
735 inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids*
736 *Res* **46**, D624-D632 (2018).
737 52 Aramaki, T. *et al.* KofamKOALA: KEGG Ortholog assignment based on profile HMM and
738 adaptive score threshold. *Bioinformatics* **36**, 2251-2252 (2020).
739 53 Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for
740 Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol* **428**, 726-
741 731 (2016).
742 54 Kanehisa, M., Sato, Y. & Kawashima, M. KEGG mapping tools for uncovering hidden features
743 in biological data. *Protein Sci* **31**, 47-53 (2022).
744 55 Nordberg, H. *et al.* The genome portal of the Department of Energy Joint Genome Institute:
745 2014 updates. *Nucleic Acids Res* **42**, D26-31 (2014).
746 56 Peng, X. *et al.* Genomic and functional analyses of fungal and bacterial consortia that enable
747 lignocellulose breakdown in goat gut microbiomes. *Nat Microbiol* **6**, 499-511 (2021).
748 57 Mićić, M., Whyte, J. D. & Karsten, V. in *Sample Preparation Techniques for Soil, Plant, and*
749 *Animal Samples* (ed Miodrag Micic) 99-116 (Springer New York, 2016).
750 58 Michalak, L. *et al.* Microbiota-directed fibre activates both targeted and secondary metabolic
751 shifts in the distal gut. *Nat Commun* **11**, 5773 (2020).
752 59 Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D. & Nesvizhskii, A. I.
753 MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry–based
754 proteomics. *Nat. Methods* **14**, 513-520 (2017).
755 60 da Veiga Leprevost, F. *et al.* Philosopher: a versatile toolkit for shotgun proteomics data
756 analysis. *Nat Methods* **17**, 869-870 (2020).
757 61 Yu, F., Haynes, S. E. & Nesvizhskii, A. I. IonQuant Enables Accurate and Sensitive Label-Free
758 Quantification With FDR-Controlled Match-Between-Runs. *Mol Cell Proteomics* **20**, 100077
759 (2021).
760 62 Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of
761 (prote)omics data. *Nat Methods* **13**, 731-740 (2016).
762 63 Kleiner, M. *et al.* Assessing species biomass contributions in microbial communities via
763 metaproteomics. *Nat Commun* **8**, 1558 (2017).
764 64 Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (2016).
765 65 R Core Team. R: A language and environment for statistical computing. (2013).

66 Federico, A. & Monti, S. hypeR: an R package for geneset enrichment workflows. *Bioinformatics* **36**, 1307-1308 (2020).
67 Keogh, K., Waters, S. M., Kelly, A. K. & Kenny, D. A. Feed restriction and subsequent realimentation in Holstein Friesian bulls: I. Effect on animal performance; muscle, fat, and linear body measurements; and slaughter characteristics. *J Anim Sci* **93**, 3578-3589 (2015).
68 McCabe, M. S. *et al.* Illumina MiSeq Phylogenetic Amplicon Sequencing Shows a Large Reduction of an Uncharacterised Succinivibrionaceae and an Increase of the *Methanobrevibacter gottschalkii* Clade in Feed Restricted Cattle. *PLOS ONE* **10**, e0133234 (2015).
69 Wessel, D. & Flugge, U. I. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem* **138**, 141-143 (1984).
70 Zougman, A., Selby, P. J. & Banks, R. E. Suspension trapping (STrap) sample preparation method for bottom-up proteomics analysis. *Proteomics* **14**, 1006-1000 (2014).
71 Harper, J. W. & Bennett, E. J. Proteome complexity and the forces that drive proteome imbalance. *Nature* **537**, 328-338 (2016).
72 Williams, A. G. & Coleman, G. S. *The Rumen Protozoa*. (1992).
73 Takenaka, A., Tajima, K., Mitsumori, M. & Kajikawa, H. Fiber Digestion by Rumen Ciliate Protozoa. *Microbes Environ* **19**, 203-210 (2004).
74 Coleman, G. S. & 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*. *J Agric Sci* **92**, 729-742 (1979).
75 Wang, L. *et al.* The transcriptome of the rumen ciliate *Entodinium caudatum* reveals some of its metabolic features. *BMC Genomics* **20**, 1008 (2019).
76 Belzecki, G., McEwan, N. R., Kowalik, B., Michalowski, T. & Miltko, R. Effect of Entodinium caudatum on starch intake and glycogen formation by Eudiplodinium maggii in the rumen and reticulum. *Eur J Protistol* **57**, 38-49 (2017).
77 Allen, M. S. Relationship between fermentation acid production in the rumen and the requirement for physically effective fiber. *J Dairy Sci* **80**, 1447-1462 (1997).
78 Solomon, R. *et al.* Protozoa populations are ecosystem engineers that shape prokaryotic community structure and function of the rumen microbial ecosystem. *ISME J* **16**, 1187-1197 (2022).
79 Park, T. & Yu, Z. Do Ruminal Ciliates Select Their Preys and Prokaryotic Symbionts? *Front Microbiol* **9**, 1710 (2018).
80 Gutierrez, J. & Davis, R. E. Bacterial Ingestion by the Rumen Ciliates Entodinium and Diplodinium. *J Protozool* **6**, 222-226 (1959).
81 Bauman, D. E., Davis, C. L. & Bucholtz, H. F. Propionate production in the rumen of cows fed either a control or high-grain, low-fiber diet. *J Dairy Sci* **54**, 1282-1287 (1971).
82 Jiao, H. P. *et al.* Effect of concentrate feed level on methane emissions from grazing dairy cows. *J Dairy Sci* **97**, 7043-7053 (2014).
83 Martin, C., Morgavi, D. P. & Doreau, M. Methane mitigation in ruminants: from microbe to the farm scale. *Animal* **4**, 351-365 (2010).
84 Finlay, B. J. *et al.* Some rumen ciliates have endosymbiotic methanogens. *FEMS Microbiol Lett* **117**, 157-161 (1994).
85 Newbold, C. J., Lassalas, B. & Jouany, J. P. The importance of methanogens associated with ciliate protozoa in ruminal methane production in vitro. *Lett Appl Microbiol* **21**, 230-234 (1995).
86 Newbold, C. J., de la Fuente, G., Belanche, A., Ramos-Morales, E. & McEwan, N. R. The Role of Ciliate Protozoa in the Rumen. *Front Microbiol* **6**, 1313 (2015).
87 Hristov, A. N. *et al.* Special topics--Mitigation of methane and nitrous oxide emissions from animal operations: I. A review of enteric methane mitigation options. *J Anim Sci* **91**, 5045-5069 (2013).

88 Jenkins, T. C., Wallace, R. J., Moate, P. J. & Mosley, E. E. Board-invited review: Recent advances in biohydrogenation of unsaturated fatty acids within the rumen microbial ecosystem. *J Anim Sci* **86**, 397-412 (2008).
89 Martin, C., Morgavi, D. P. & Doreau, M. Methane mitigation in ruminants: from microbe to the farm scale. *Animal* **4**, 351-365 (2010).
90 Zhang, X. M. *et al.* Corn oil supplementation enhances hydrogen use for biohydrogenation, inhibits methanogenesis, and alters fermentation pathways and the microbial community in the rumen of goats. *J Anim Sci* **97**, 4999-5008 (2019).
91 Guyader, J. *et al.* Influence of rumen protozoa on methane emission in ruminants: a meta-analysis approach. *Animal* **8**, 1816-1825 (2014).
92 Niu, P. *et al.* A Basic Model to Predict Enteric Methane Emission from Dairy Cows and Its Application to Update Operational Models for the National Inventory in Norway. *Animals (Basel)* **11**, 1891 (2021).
93 McAllister, T. A., Cheng, K.-J., Okine, E. K. & Mathison, G. W. Dietary, environmental and microbiological aspects of methane production in ruminants. *Can J Anim Sci* **76**, 231-243 (1996).
94 Gruninger, R. J. *et al.* Application of 3-nitrooxypropanol and canola oil to mitigate enteric methane emissions of beef cattle results in distinctly different effects on the rumen microbial community. *Anim Microbiome* **4**, 35 (2022).
95 Popova, M. *et al.* Effect of fibre- and starch-rich finishing diets on methanogenic Archaea diversity and activity in the rumen of feedlot bulls. *Anim Feed Sci Technol* **166-167**, 113-121 (2011).
96 Van Kessel, J. A. S. & Russell, J. B. The effect of pH on ruminal methanogenesis. *FEMS Microbiol Ecol* **20**, 205-210 (1996).
97 Russell, J. B., Muck, R. E. & Weimer, P. J. Quantitative analysis of cellulose degradation and growth of cellulolytic bacteria in the rumen. *FEMS Microbiol Ecol* **67**, 183-197 (2009).
98 Russell, J. B. The importance of pH in the regulation of ruminal acetate to propionate ratio and methane production in vitro. *J Dairy Sci* **81**, 3222-3230 (1998).
99 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).
100 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).

850 **Table 1. Summary of the animal host, dietary conditions and accompanying metadata that are linked to**
851 **the rumen samples used to explore protozoal functions.** Effects of control diet (CTL) and diet
852 supplemented with corn oil and wheat starch (COS) on average VFA concentration in the percentage of
853 total VFA concentration (mmol/l) and average methane yield (CH_4 g/kg DMI) in addition to average total
854 protozoa cell counts and average cell counts for small entodiniomorph protozoa for cattle (n=4) and goats
855 (n=4) (**Supplementary Table S2**)[5]. Total protozoal cell counts, and cell counts of entodiniomorphs and
856 holotrichs were based on ciliate protozoa morphology in microscopy[5]. Measurements on VFA
857 concentrations and methane yield were determined by gas chromatography and respiration chambers,
858 respectively.

Animal	Cattle		Goat	
Diet	CTL	COS	CTL	COS
VFA sum mmol/l	61.8	68.3	33.7	25.3
Acetate (% sum)	72.0	69.1	65.7	64.2
Propionate (% sum)	14.2	21.3	15.6	19.1
Butyrate (% sum)	10.8	6.4	13.7	9.9
CH_4 g/kg dry matter intake	19.5	14.5	19.5	13.5
Standard deviation	2.2	2.9	4.0	4.1
Total protozoa (10^3 cells/ml)	92	100	1382	898
Small entodiniomorphs (<100µm) (10^3 cells/ml)	51	97	1346	898
Large entodiniomorphs (>100µm) (10^3 cells/ml)	25.1	1.7	4.9	0.2
Holotrichs				
Isotricha spp	4.1	1.0	4.6	0.2
Dasytricha spp	11.8	0.1	26.7	0.2

859

860 **Figure legends**

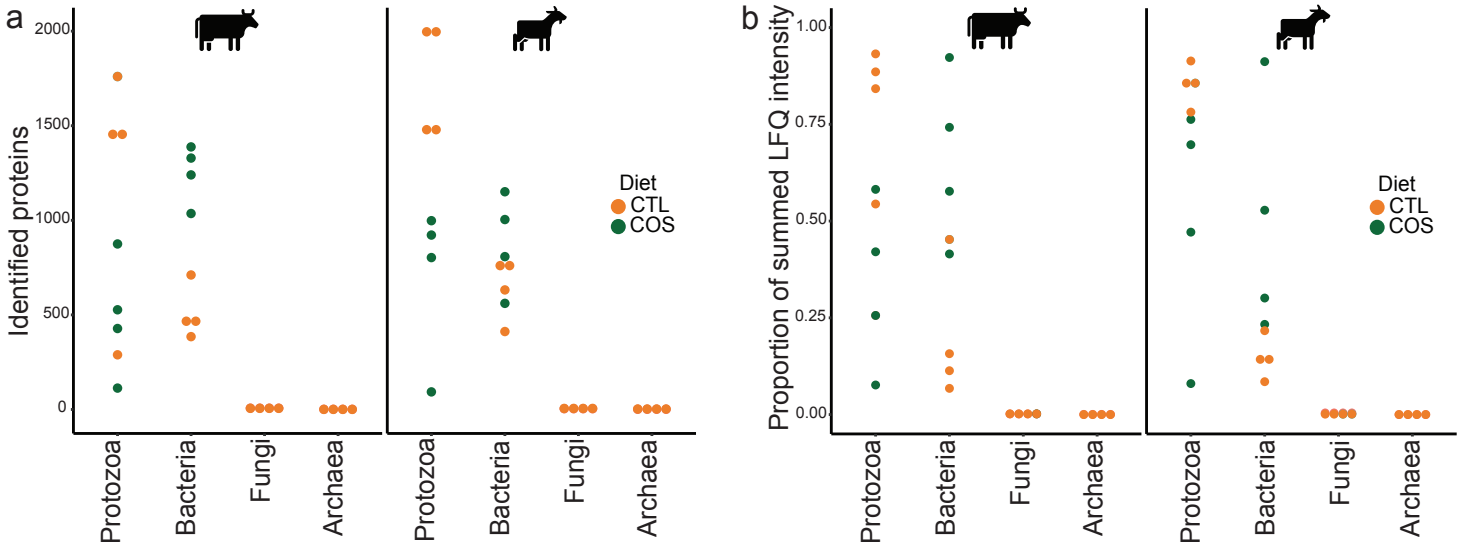
861 **Figure 1. Quantities of identified protozoal proteins in the rumen microbiome vary depending on**
862 **host animal and dietary conditions.** Dot plots for total proteins identified (**a**), and average recovered
863 metaproteomic expression (**b:** presented as proportion of summed LFQ intensities) belonging to
864 protozoal, bacterial, archaeal, or fungal species across the control diet (CTL) and diets supplemented
865 with corn oil and wheat starch (COS) for dairy cattle (n=4) and goats (n=4). Detected protein
866 abundances for protozoal and bacterial populations can be found in **Supplementary Table S6**.

867

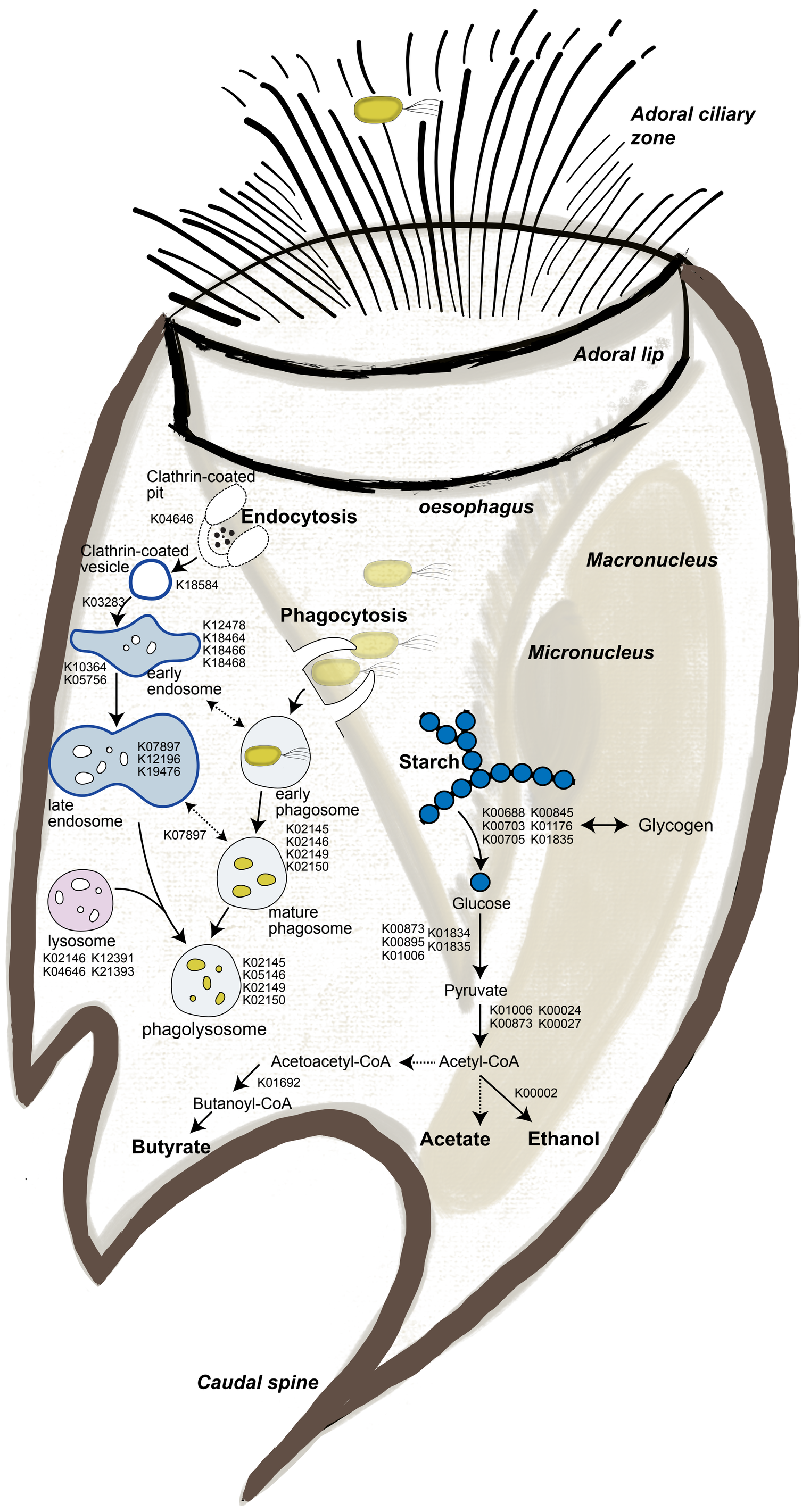
868 **Figure 2. Detected proteins mapped to protozoal genomes/SAGs and bacterial metagenome-**
869 **assembled genomes (MAGs) in the rumen microbiome of dairy cattle (n=4) and goats (n=4) fed**
870 **either a control diet (CTL) or one supplemented with corn oil and wheat starch (COS)**. The figure
871 displays metabolically active populations (as genomes, SAGs or MAGs), with selected expressed
872 proteins (presented as Enzyme Commission (EC) number with short enzyme descriptions) active in

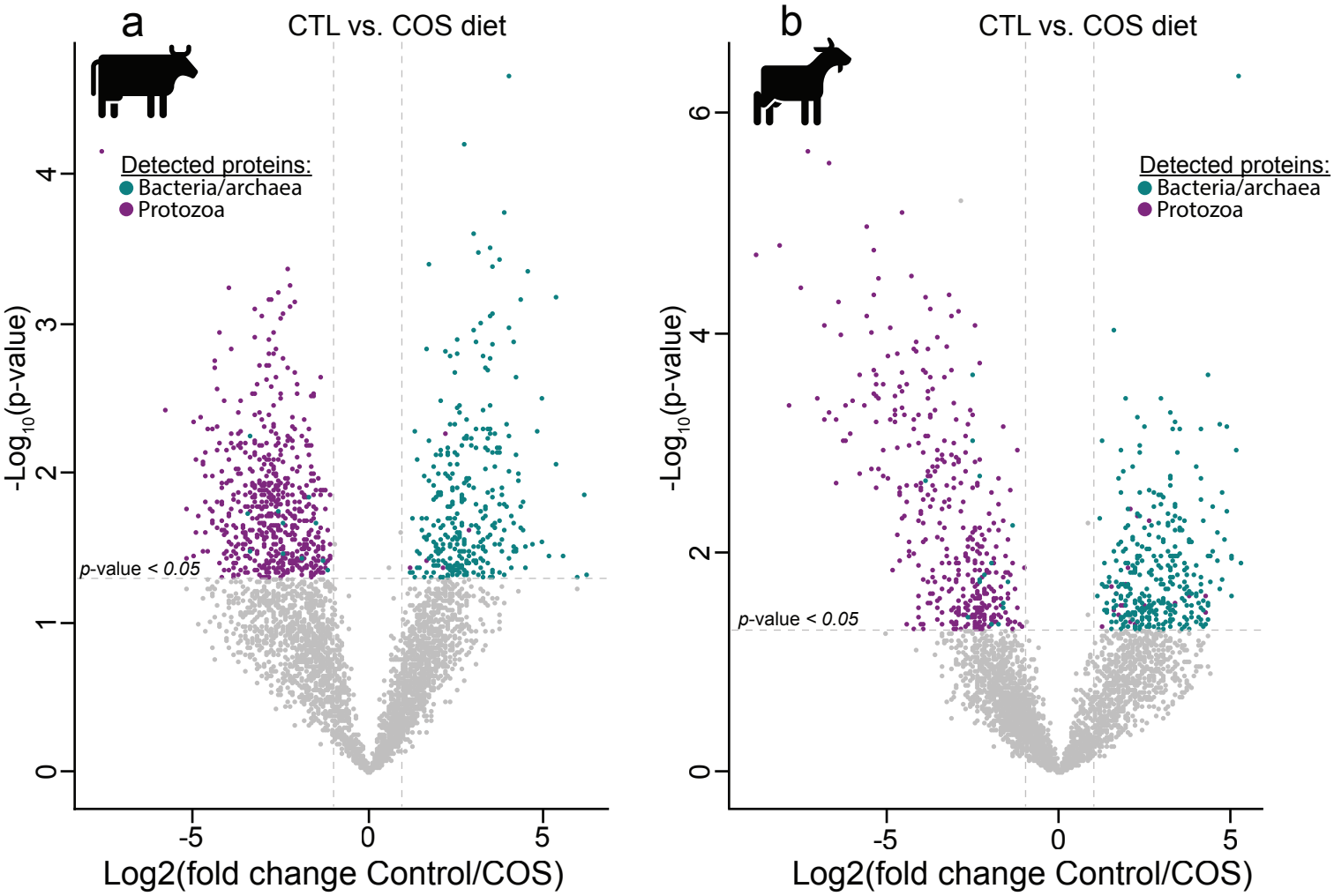
873 starch degradation, glycolysis and production of pyruvate, butyrate, acetate, and succinate in cattle
874 (**a-b**) and goats (**c-d**) fed CTL or COS diets. Panels **a** and **c** depict protozoal proteomes that were
875 detected in cattle and goats respectively are presented separately to bacteria (panels **b** and **d**) as the
876 scale of their protein quantification values were ~10x larger. Protein quantification values (y-axis)
877 were calculated by considering both the number of proteins detected per MAG/SAG/genome and
878 their LFQ intensity: we averaged LFQ intensities for each detected protein across biological replicates
879 for each dietary condition (CTL: green or COS: orange), which were subsequently summed for all
880 detected proteins per MAG/SAG/genome. Heatmaps show selected MAG/SAG/genome with
881 metabolically active proteins, presented as EC numbers, recovered from cattle (RUDB-C) and goats
882 (RUDB-G) fed either CTL or COS diets. Where average protein abundance level for a given
883 MAG/SAG/genome was significantly different (determined by paired Wilcoxon test, *p*-value < 0.05)
884 between the COS and CTL diet, the given MAG/SAG/genome has been marked with a red star. MAGs
885 are presented with their MAG ID and taxonomic annotation from GTDB-tk. Genome annotations and
886 LFQ intensities used to create heatmaps can be found in **Supplementary Table S6**.
887
888 **Figure 3. Reconstructed phagolysosome formation and starch metabolism of *Entodinium caudatum***
889 **within the rumen microbiome of goats fed the control diet (CTL) based on metaproteomic analysis.**
890 KO identifiers for identified proteins were analysed via KEGG mapper to reconstruct expressed key
891 features in the metabolism of *En. caudatum*. Dashed arrows represent proteins or pathways that
892 were not detected in our metaproteomes but are key steps in their respective pathways. Detailed
893 information connecting KO identifiers to their respective gene ID, LFQ and animal/diet can be found
894 in **Supplementary Table S6**.
895
896 **Figure 4. Volcano plots indicating different rumen microbiome proteins from dairy cattle (a) and**
897 **goats (b) fed either the control or COS diets and which displayed both large magnitude of fold-**
898 **changes in LFQ intensities (x axis) and high statistical significance (-log10 of nominal *p*-values using**
899 **an unpaired t-test, y axis).** Dashed horizontal line denotes a *p*-value <0.05 cut-off. For both animal
900 hosts we observed an increase in protein detection for bacterial populations (greenblue) in animals
901 fed high-starch diets compared to protozoal populations (purple), which were detected at higher LFQ
902 intensities in animals fed the control diet.
903
904 **Figure 5. The proteomes of rumen microbiome populations from Holstein-Friesian beef cattle are**
905 **affected by high starch diets.** A total of 60 beef cattle were subjected to two dietary contrasting
906 condition: 30 animals with *ad libitum* feeding and 30 subjected to 125 days of feed restriction. Dietary

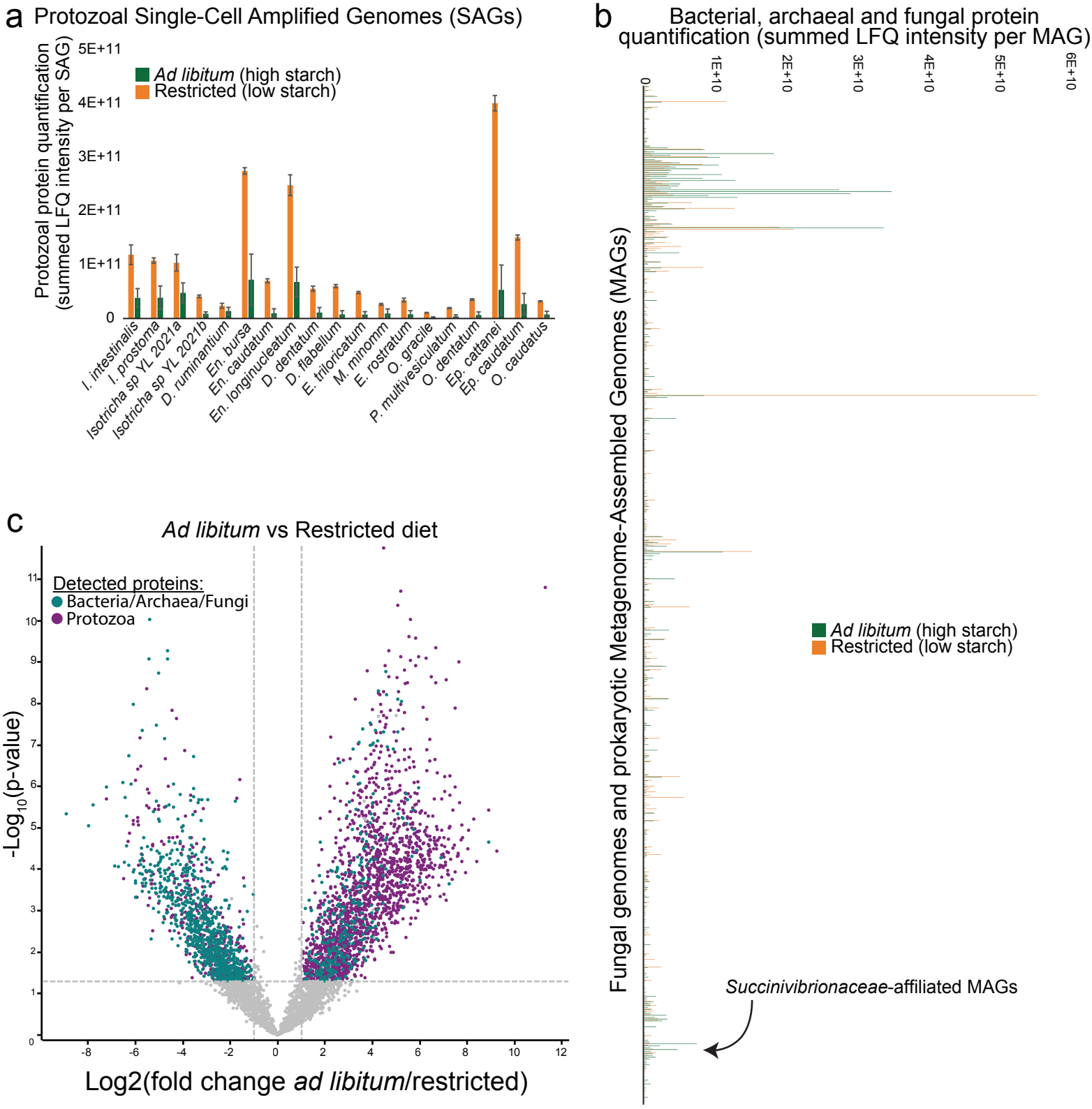
907 components in both treatments consisted of 70% concentrate, and 30% grass silage, with the
908 concentrate containing rolled barley 72.5%, soya 22.5%, molasses 3% and calf mineral 2%. Rolled
909 barley is high in energy and starch content (~50%). Metaproteomes for a subset of 15 animals (7
910 restricted and 8 *ad libitum*) were analysed against a database that combined 19 protozoal
911 SAGs/genome (**a**) and 781 MAGs and isolate genomes (**b**), including fungal representatives. Protozoal
912 proteomes are presented separately as the scale of their protein quantification values were ~10x
913 larger. Protein quantification values were calculated by considering both the number of proteins
914 detected per MAG/SAG/genome and their abundance; we averaged LFQ intensities for each detected
915 protein across biological replicates for each dietary condition (*ad libitum*: green or restricted: orange),
916 which were subsequently summed for all detected proteins per MAG/SAG/genome. Similar to our
917 observations in Holstein dairy heifers and Alpine goats (**Figure 2)**, the proteomes of protozoal species
918 were a major fraction of the total rumen metaproteome, however they were substantially reduced in
919 dietary conditions where starch content was higher. Further supporting our results in dairy cattle, we
920 observed in high-starch conditions an increase in protein detection for bacterial populations affiliated
921 to the *Succinivibrionaceae*. **c.** Volcano plot indicating different rumen microbiome proteins that
922 displayed both large magnitude of fold-changes in LFQ intensities (x axis) and high statistical
923 significance (-log10 of nominal *p*-values using an unpaired t-test, y axis). Dashed horizontal line
924 denotes a *p*-value<0.05 cut-off. Volcano plots showed an increase in protein detection for bacterial
925 populations (greenblue) in animals fed *ad libitum* compared to protozoal populations (purple), which
926 were detected at higher LFQ intensities in animals under restricted conditions. LFQ intensities used to
927 create panel **a** and **b** can be found in **Supplementary Table S8**.
928
929 **Supplementary Figure S1. Protein abundance level ratios calculated for commonly observed families**
930 **between animals fed high/low starch diets.** Log 2-Fold change ratios between average LFQ intensities
931 for dairy cattle (**a**), goats (**b**) or beef cattle from an independent dataset (**c**) fed high or low starch diets.
932 Positive ratios show higher average protein abundance level for a given taxonomic family in animals
933 that were fed high starch diets (COS or *ad libitum*). Negative ratios show higher average protein
934 abundance levels for a given taxonomic family in animals that were fed the low starch diets (CTL or
935 Restricted). Rarely observed families with < 2 MAGs with detected proteins across both conditions
936 were excluded from the analysis. All ratios, and subsequent correlation tests are available in
937 **Supplementary Table S9**.
938


a
Identified proteins
2000
1500
1000
500
0
Protozoa
Bacteria
Fungi
Archaea
Protozoa
Bacteria
Fungi
Archaea
Diet
CTL
COS
b
Proportion of summed LFQ intensity
1.00
0.75
0.50
0.25
0.00
Protozoa
Bacteria
Fungi
Archaea
Protozoa
Bacteria
Fungi
Archaea
Diet
CTL
COS

a
CTL
COS
[FeFe] hydrogenase
Ep. cattanei gene 2999
D. dentatum gene 4291
I. intestinalis gene 23360
Protein quantification (summed LFQ intensity - Protozoa)
holotrichs
entodiniomorphs
I
II
III
I. intestinalis
I. prostoma
Isotricha sp YL 2021a
Isotricha sp YL 2021b
D. ruminantium
En. bursa
En. caudatum
En. longinucleatum
D. dentatum
D. flabellum
E. triloricatum
M. minomm
E. rostratum
O. gracile
P. multivesiculatum
O. dentatum
Ep. cattanei
Ep. caudatum
O. caudatus
Protozoal Single-Cell Amplified Genomes (SAGs)
I: Epidinium cattanei
II: Diplodinium dentatum
III: Dasytrichia ruminantium
fiber digestion
Starch metabolism
Glycolysis
Pyruvate metabolism
Acetate production
Butyrate production
cellulase (GH5)
cellulase (GH9)
xylanase (GH10)
xylanase (GH11)
xylanase (GH30)
mannanase (GH26)
cellobiose phosphorylase (GH94)
beta-glucosidase (GH3)
mannosidase (GH2)
acetyl xylan esterase (CE1)
alpha-amylase (GH13)
4-alpha-glucanotransferase (GH77)
4-alpha-glucanotransferase
starch synthase (GT5)
glycogen phosphorylase (GT35)
glycogen phosphorylase (GT35)
sucrose phosphotransferase
glucokinase
phosphoglucomutase
glucose-6-phosphate isomerase
6-phosphofructokinase
fructose-bisphosphate aldolase
fructose-bisphosphatase
triose-phosphate isomerase
glyceraldehyde-3-phosphate dehyd.
phosphoglycerate kinase
phosphoglycerate mutase
enolase
pyruvate formate-lyase
pyruvate, phosphate dikinase
pyruvate, phosphate dikinase
pyruvate synthase
pyruvate kinase
phosphoenolpyruvate carboxykinase
phosphate acetyltransferase
acetate kinase
acetyl-CoA C-acetyltransferase
3-hydroxyacyl-CoA dehydrogenase
enoyl-CoA hydratase
phosphate butyryltransferase
phosphate butyryltransferase
butyrate kinase
b
Protein quantification (summed LFQ intensity - MAG)
Prokaryotic Metagenome-Assembled Genomes (MAGs)
Metaproteome detection (LFQ: Log2)
19 or below
20-23
24-27
28 or above
Not detected
I: RUDB_C039 Prevotella
II: RUDB_C040 Succinivibrio
III: RUDB_C030 Succinovibrionaceae
IV: RUDB_C002 Succinovibrionaceae
Succinate production
4-alpha-glucanotransferase (GH77)
glucan 1,4-alpha-glucosidase (GH97)
1,4-alpha-glucan branch. enz. (GH13)
glycogen phosphorylase (GT35)
glycogen phosphorylase (GT35)
maltotransporter
maltotransporter
maltotransporter
sucrose phosphotransferase
glucokinase
phosphoglucomutase
glucose-6-phosphate isomerase
6-phosphofructokinase
fructose-bisphosphate aldolase
fructose-bisphosphatase
triose-phosphate isomerase
glyceraldehyde-3-phosphate dehyd.
phosphoglycerate kinase
phosphoglycerate mutase
enolase
pyruvate formate-lyase
pyruvate, phosphate dikinase
pyruvate synthase
pyruvate kinase
phosphoenolpyruvate carboxykinase
phosphoenolpyruvate carboxykinase
phosphate acetyltransferase
acetate kinase
malate dehydrogenase
malate dehydrogenase
malate dehydrogenase
fumarate hydratase
succinate dehydrogenase
c
CTL
COS
[FeFe] hydrogenase
En. caudatum gene 20303
D. ruminantium gene 12261
E. triloricatum gene 6464
I. prostoma gene 877
D. flabellum gene 4923
Protein quantification (summed LFQ intensity - Protozoa)
holotrichs
entodiniomorphs
I
II
III
I. intestinalis
I. prostoma
Isotricha sp YL 2021a
Isotricha sp YL 2021b
D. ruminantium
En. bursa
En. caudatum
En. longinucleatum
D. dentatum
D. flabellum
E. triloricatum
M. minomm
E. rostratum
O. gracile
P. multivesiculatum
O. dentatum
Ep. cattanei
Ep. caudatum
O. caudatus
Protozoal Single-Cell Amplified Genomes (SAGs)
I: Entodinium caudatum
II: Enoploplastron triloricatum
III: Dasytrichia ruminantium
fiber digestion
Starch metabolism
Glycolysis
Pyruvate metabolism
Acetate production
Butyrate production
cellulase (GH5)
cellulase (GH9)
xylanase (GH10)
xylanase (GH11)
xylanase (GH30)
mannanase (GH26)
cellobiose phosphorylase (GH94)
beta-glucosidase (GH3)
xylan esterase
alpha-amylase (GH13)
alpha-amylase (GH13)
alpha-amylase (GH13)
1,4-alpha-glucan branching enzyme (GH13)
4-alpha-glucanotransferase (GH77)
starch synthase (GT5)
glycogen phosphorylase (GT35)
glycogen phosphorylase (GT35)
sucrose phosphotransferase
glucokinase
phosphoglucomutase
glucose-6-phosphate isomerase
6-phosphofructokinase
fructose-bisphosphate aldolase
fructose-bisphosphatase
triose-phosphate isomerase
glyceraldehyde-3-phosphate dehyd.
phosphoglycerate kinase
phosphoglycerate mutase
enolase
pyruvate formate-lyase
pyruvate, phosphate dikinase
pyruvate, phosphate dikinase
pyruvate synthase
pyruvate kinase
phosphoenolpyruvate carboxykinase
phosphate acetyltransferase
acetate kinase
acetyl-CoA C-acetyltransferase
3-hydroxyacyl-CoA dehydrogenase
enoyl-CoA hydratase
phosphate butyryltransferase
phosphate butyryltransferase
butyrate kinase
d
CTL
COS
[FeFe] hydrogenase
RUDB_G023 gene13 (Aminobacteriaceae)
RUDB_G023 gene31 (Aminobacteriaceae)
RUDB_G016 gene19 (Aminobacteriaceae)
Protein quantification (summed LFQ intensity - MAG)
Prokaryotic Metagenome-Assembled Genomes (MAGs)
Metaproteome detection (LFQ: Log2)
19 or below
20-23
24-27
28 or above
Not detected
I : RUDB_G032 Prevotella
II RUDB_G037 Selenomonadaceae
III: RUDB_G023 Aminobacteriaceae
IV: RUDB_G016 Aminobacteriaceae
Succinate / propionate production
4-alpha-glucanotransferase (GH77)
starch glucosyltransferase (GT4)
1,4-alpha-glucan branching enzyme (GH13)
glucose-1-phosphate adenylyltransferase
glycogen phosphorylase (GT35)
maltoporin
maltotransporter
sugar transporter
phosphotransferase system
glucokinase
phosphoglucomutase
glucose-6-phosphate isomerase
6-phosphofructokinase
fructose-bisphosphate aldolase
fructose-bisphosphate aldolase
fructose-bisphosphatase
triose-phosphate isomerase
glyceraldehyde-3-phosphate dehyd.
phosphoglycerate kinase
phosphoglycerate mutase
enolase
pyruvate kinase
pyruvate, phosphate dikinase
pyruvate synthase
phosphoenolpyruvate carboxykinase
phosphate acetyltransferase
acetate kinase
malate dehydrogenase
malate dehydrogenase
malate dehydrogenase
fumarate hydratase
fumarate reductase
acetate kinase
propanoyl-CoA:succinate CoA transferase


Adoral ciliary zone
Adoral lip
oesophagus
Macronucleus
Micronucleus
Clathrin-coated pit
K04646
Endocytosis
Clathrin-coated vesicle
K18584
K03283
K12478
K18464
K18466
K18468
K10364
K05756
early endosome
Phagocytosis
K07897
K12196
K19476
late endosome
early phagosome
K07897
K02145
K02146
K02149
K02150
mature phagosome
lysosome
K02146 K12391
K04646 K21393
K02145
K05146
K02149
K02150
phagolysosome
Starch
K00688 K00845
K00703 K01176
K00705 K01835
Glycogen
Glucose
K00873 K01834
K00895 K01835
K01006
Pyruvate
K01006 K00024
K00873 K00027
Acetoacetyl-CoA
Acetyl-CoA
K01692
K00002
Butanoyl-CoA
Butyrate
Acetate
Ethanol
Caudal spine


a
CTL vs. COS diet
Detected proteins:
Bacteria/archaea
Protozoa
p-value < 0.05
-Log10(p-value)
Log2(fold change Control/COS)
b
CTL vs. COS diet
Detected proteins:
Bacteria/archaea
Protozoa
p-value < 0.05
-Log10(p-value)
Log2(fold change Control/COS)


a
Protozoal Single-Cell Amplified Genomes (SAGs)
Protozoal protein quantification (summed LFQ intensity per SAG)
5E+11
4E+11
3E+11
2E+11
1E+11
0
Ad libitum (high starch)
Restricted (low starch)
I. intestinalis
I. prostoma
Isotricha sp YL 2021a
Isotricha sp YL 2021b
D. ruminantium
En. bursa
En. caudatum
En. longinucleatum
D. dentatum
D. flabellum
E. trioricatum
M. minomm
E. rostratum
O. gracile
P. multivesiculatum
O. dentatum
Ep. cattanei
Ep. caudatum
O. caudatus
b
Bacterial, archaeal and fungal protein quantification (summed LFQ intensity per MAG)
0
1E+10
2E+10
3E+10
4E+10
5E+10
6E+10
Fungal genomes and prokaryotic Metagenome-Assembled Genomes (MAGs)
Ad libitum (high starch)
Restricted (low starch)
Succinivibrionaceae-affiliated MAGs
c
Ad libitum vs Restricted diet
Detected proteins:
Bacteria/Archaea/Fungi
Protozoa
-Log10(p-value)
Log2(fold change ad libitum/restricted)