Microbiota-directed fibre activates both targeted and secondary metabolic shifts in the distal gut

Leszek Michalak[1], John Christian Gaby[1], Leidy Lagos[2], Sabina Leanti La Rosa[1], Torgeir R. Hvidsten[1], Catherine Tétard-Jones[3], William G.T. Willats[3], Nicolas Terrapon[4,5], Vincent Lombard[4,5], Bernard Henrissat[4,5,6], Johannes Dröge[7], Magnus Ø. Arntzen[1], Live Heldal Hagen[1], Margareth Øverland[2], **Phillip B. Pope[1,2]*, Bjørge Westereng[1]***

[1] Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, 1432 Ås, Norway

[2] Faculty of Biosciences, Norwegian University of Life Sciences, 1432 Ås, Norway

[3] School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, UK.

[4] Centre National de la Recherche Scientifique, Aix-Marseille Université, UMR7257, France

[5] Institut National de la Recherche Agronomique, USC1048 Architecture et Fonction des Macromolécules Biologiques, Marseille, France

[6] Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia

[7] Department for Mathematical Sciences, Chalmers University of Technology, Gothenburg, Sweden

* Contributed equally

Corresponding authors:

Bjorge Westereng bjorge.westereng@nmbu.no

Phillip B. Pope phil.pope@nmbu.no

John Christian Gaby john.christian.gaby@nmbu.no

ABSTRACT

Beneficial modulation of the gut microbiome has high-impact implications not only in humans, but also in livestock that sustain our current societal needs. In this context, we have tailored an acetylated galactoglucomannan (AcGGM) fibre to match unique enzymatic capabilities of *Roseburia* and *Faecalibacterium* species, both renowned butyrate-producing gut commensals. The accuracy of AcGGM was tested within the complex endogenous gut microbiome of pigs, wherein we resolved 355 metagenome-assembled genomes together with quantitative metaproteomes. In AcGGM-fed pigs, both target populations differentially expressed AcGGM-specific polysaccharide utilization loci, including novel, mannan-specific esterases that are critical to its deconstruction. However, AcGGM-inclusion also manifested a "butterfly effect", whereby numerous metabolic changes and interdependent cross-feeding pathways were detected in neighboring non-mannolytic populations that produce short-chain fatty acids. Our findings show that intricate structural features and acetylation patterns of dietary fibre can be customized to specific bacterial populations, with potential to create greater modulatory effects at large.

MAIN TEXT

Microbiota-directed foods (MDFs) have emerged as a strategy to modulate the gut microbiome, as diet has distinct and rapid effects on microbiome composition and function[1,2]. MDFs by definition are not broadly metabolized, but rather elicit a targeted metabolic response in specific indigenous microbiota that confers benefits to their host. This in itself presents a challenge; as many newly identified MDF target organisms, such as beneficial butyrate-producing (i.e. butyrogenic) *Roseburia* and *Faecalibacterium* spp.[3,4], have broad metabolic capabilities that are shared with the vast majority of fibre-fermenting microbiota in the gut ecosystem. Nevertheless, recent studies have revealed intimate connections between the enzymatic and mechanistic features of microorganisms and the glycan structures of the fibres they consume[5,6], which creates new conceptual MDF targets.

49 This is exemplified by discoveries of sophisticated polysaccharide-degrading apparatuses that enable
50 certain microbiota to consume fibre in a "selfish" manner, whereby complex glycan-structures (such
51 as β-mannans) are cleaved into large oligosaccharides at the cell surface, which are subsequently
52 transported into the cell and depolymerized into monomeric sugars[5,7,8]. Such a mechanism restricts
53 the release of sugars into the ecosystem for neighboring scavenging populations, thus giving a
54 selective metabolic advantage to the selfish-degrader in the presence of these highly complex glycans.

55 Beta-mannans are present in human and livestock diets, and depending on their plant origins, can
56 be heavily decorated with varying amounts of acetylation that protect the fibre from enzymatic
57 degradation[9]. We recently demonstrated that the human gut commensal *Roseburia intestinalis*
58 encodes a mannan-specific polysaccharide utilization locus (PUL), and "selfishly" converts highly
59 complex mannan substrates to butyrate[5]. Within this mannan PUL, a carbohydrate esterase (CE) family
60 2 (*Ri*CE2) removes 3-O-, and 6-O- acetylations on mannan, whereas a novel CE family 17 (*Ri*CE17)
61 removes the axially oriented 2-*O*-hydroxyl acetylations[9], which are distinctive features found in limited
62 mannan moieties and inaccessible to most of the characterized bacterial esterases present in the gut
63 microbiome. Closer genome examinations have revealed that putative CE2/CE17-containing mannan
64 PULs are in fact prominent within many butyrate-producers including *Roseburia* spp.,
65 *Faecalibacterium prausnitzii*, *Ruminococcus gnavus*, *Coprococcus eutactus* and *Butyrivibrio*
66 *fibrisolvens*[5,10]. It is well known that the metabolic attributes of these populations are highly desirable
67 in the gastrointestinal tract, and that their depletion is implicated in colorectal cancer, Crohn's disease,
68 inflammatory bowel syndrome, ulcerative colitis, forms of dermatitis and several other diseases[11,12].
69 These collective findings thus raised the question: could a custom MDF fibre that was tailored to match
70 these specialized enzymatic capabilities selectively engage butyrate-producers in a complex
71 microbiome ecosystem?

72 2-*O*-acetylated mannans are found in a limited number of characterized western dietary fibre
73 sources (i.e. tomatoes[13] and coffee[5]), however 2-*O*-acetylations are present in acetylated

74 galactoglucomannan (AcGGM), which is the main hemicellulose in the secondary cell wall of Norway
75 spruce (*Picea abies*)[14]. Here, we have utilized controlled steam explosion (SE), followed by
76 ultrafiltration (UF) fractionation to extract complex AcGGM from spruce wood. Processing conditions
77 were selected to tailor the fibre with a high degree of galactose branching's and 2-*O*-, 3-*O*- and 6-*O*-
78 acetylations[15], which is amenable to inclusion as an MDF in animal feed production. Previously, the
79 MDF concept has matched polysaccharides with *Bacteroides*-encoded PULs to demonstrate the
80 creation of exclusive metabolic niches[1,16]. In particular, Shepherd et al engrafted exogenous strains in
81 mice via their rare PUL-encoded enzymatic capabilites[16], whereas Patnode et al[1] illustrated that
82 bioactive carbohydrates were found to target particular *Bacteroides* species in a defined consortium
83 of 20 human gut microbial species in gnotobiotic mice. However, what is less understood from MDF
84 studies to date, is (1) can the MDF concept be applied to target indigenous populations within a
85 complex endogenous microbiome, and (2) what are the broader secondary community effects if a
86 targeted population is stimulated, i.e. are new niches created and/or existing ones closed?

87 Here, we tested whether our AcGGM fibre could specifically target beneficial Firmicutes
88 species *Roseburia* and *Faecalibacterium* within a "real-world" gut ecosystem that consists of 100-
89 1000's of different species. To do this, we fed weaned piglets diets containing varying AcGGM levels
90 over a 28-day period that extended from their first meal after sow separation until an adapted, fibre-
91 degrading microbiome was established. Temporal changes in the microbiome were monitored using
92 metagenomics, which phylogenetically and functionally resolved the genomes of indigenous
93 microbiota. Finally, quantitative metaproteomic analysis and carbohydrate microarrays were used to
94 visualize the metabolic and enzymatic responses of the different microbiota to the varying AcGGM
95 exposure. This approach demonstrates how the activity of specific beneficial microbiota can be
96 directly stimulated while simultaneously deciphering the secondary, trophic effects on other
97 populations and metabolic niches, with both aspects having broader implications for developing
98 strategies to effectively modulate the gut microbiome.

99

100 **RESULTS AND DISCUSSION**

101 **Steam explosion allows scalable production of highly complex dietary mannan fibres from wood.**
102 Spruce galactoglucomannan consists of a backbone of β-(1,4)-linked mannose and glucose residues,
103 decorated with α-(1,6) linked galactose branching, and a large degree of esterification of the mannose
104 residues by 2-*O*- and 3-*O*-, and 6-*O*- acetylations[14] (**Fig 1a**). A crucial part of this study was the
105 development of an efficient, large-scale extraction process entailing SE as well as ultra- and nano-
106 filtration, which ultimately provided high quantities at high purity whilst not damaging the complexity
107 of the AcGGM fibre (**Fig. 1b-c**). A total of 700kg of dry Norway spruce chips was processed using SE at
108 conditions corresponding to a combined severity factor (R'_0) of 1.70. We produced 50 kg of
109 oligo/polysaccharides for feed production (**Fig. 1c-e**), with a monosaccharide (Man: Glc: Gal) ratio of
110 4:1:0.6. The degree of polymerization (DP) of the AcGGM fibre ranged from β-mannooligosaccharides
111 with DP of 2-10 to manno-polysaccharides (DP of ≥11), with both exhibiting degree of acetylation (DA
112 = 0.35). Crucially, this DA value was higher than previous iterations of the fibre (DA = 0.28)[7], and its
113 acetylation patterns (previously determined[9]) matched the enzymatic capabilities of mannan PULs
114 encoded in human gut *Roseburia* and *Faecalibacterium* spp.[5,10]. We therefore predicted that our
115 AcGGM fibre would match representatives of the same populations that are indigenous to porcine gut
116 ecosystems[17,18] (**Fig. 1f**).

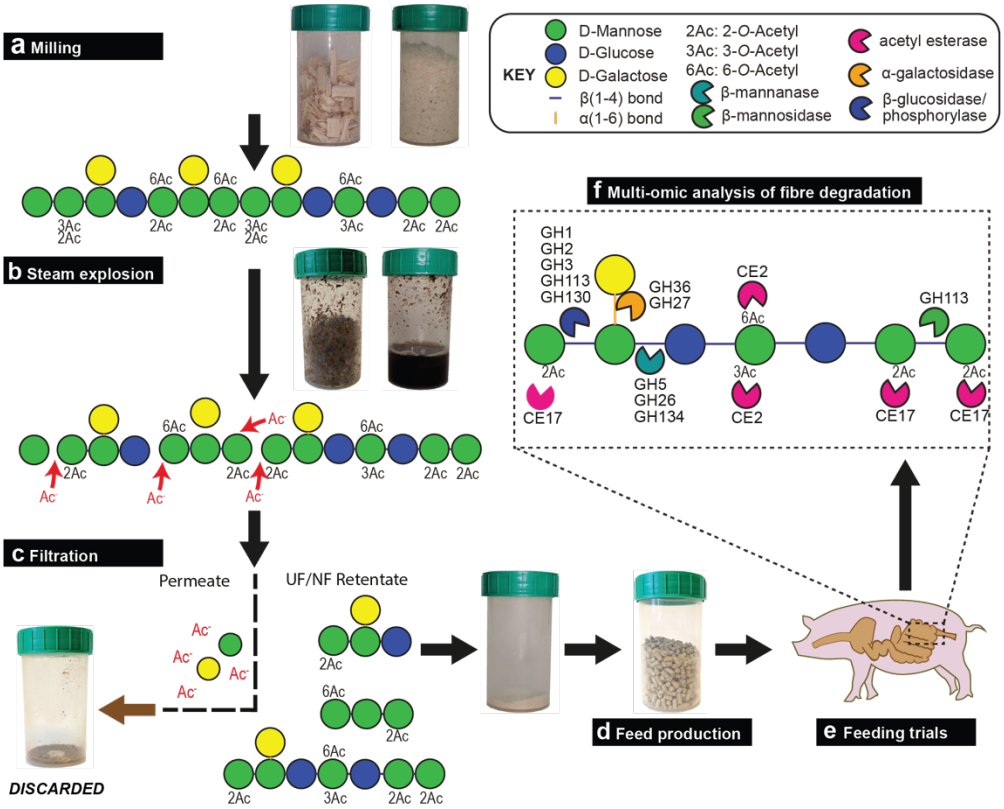


Fig. 1. Schematic representation and graphic illustration of the production pipeline for Norway spruce AcGGM. **a**, Wood chips were milled to increase the surface area exposed for hydrothermal extraction and washing the released mannooligosaccharides. Spruce mannan isolated by mild extraction is long chained galactoglucomannan (DP200-400), highly acetylated DA~0.65, predominantly 2-*O* and 3-*O* acetylations[19]. **b**, During hydrothermal pretreatment, release of acetic acid promoted hydrolysis of glycosidic bonds, thereby relinquishing monosaccharides, oligosaccharides and other breakdown products. **c**, Ultra filtration retained the longer, complex oligosaccharides and discarded the monosaccharides, acetic acid and other steam explosion byproducts. **d**, The purified mannan was incorporated into feed pellets at varying inclusion levels, produced by a conventional feed pelleting process. **e**, The growth performance experiment and a feeding trial were conducted in a randomized block design, with four inclusion levels of AcGGM. **f**, Multi-omic approaches were used to analyze the porcine gut microbiome in response to AcGGM and determine if indigenous mannan PULs matched the glycan structure of the AcGGM fibre. Glycosidic bonds between the β-(1,4)-mannose and glucose units in the backbone of AcGGM require hydrolysis by glycoside hydrolases (GH) from families GH5 and GH26. GH36 and GH27 α-galactosidases are required to remove the α-(1,6)-galactose decorations. Single mannose and glucose units are removed from the non-reducing end of the oligosaccharides by enzymes from families GH1, GH2, GH3, GH113 and GH130, while mannose in the reducing end can be removed by GH113 family mannosidases. 3-*O*-, and 6-*O*- acetylations on mannan are removed by esterases from family CE2. A unique feature of particular beta-mannans is the axially oriented 2-*O*- acetylation on mannose, which is the prevalent form of acetylation present on AcGGM used in this study. 2-*O*-acetylations are removed by esterases homologous to the *Ri*CE17 mannan-specific esterase from *Roseburia intestinalis*, which was recently characterized by our group[5].

Spruce AcGGM altered the gut microbiome of weaned piglets in a dose-response manner. We previously demonstrated that varieties of AcGGM can be metabolized by pure cultures of *Roseburia intestinalis* L1-82[5], *in vitro* enrichments with human gut butyrate-producers[10] and "mini-microbiota" within gnotobiotic mice[5]. Here, we wanted to test our AcGGM fibre's accuracy and ability to elicit a specific response in indigenous representatives of our target populations within a highly complex and competitive endogenous microbiome. In total, four separate cohorts of twelve weaned piglets were

given a pelleted basal feed semi-*ad libitum*, which contained either 0% (control), 1%, 2% or 4% AcGGM to additionally determine the level necessary to elicit an effect on both the host and its microbiome. We chose a dietary supplementation strategy, which is commonly applied to piglets during the post-weaning period to offset their particularly low feed intake and enhance gut health and growth performance[20]. Cautious measures were made to eliminate potential pen biases (see details in materials and methods), and fecal samples as well as animal performance metrics were taken before AcGGM administration (when piglets were assigned to pens), and subsequently at days 7, 14, 21 and 27 during the feeding trial. At day 28, the piglets were sacrificed and host gut tissue and digesta samples taken from all intestinal regions (duodenum, jejunum, ileum, cecum and colon) for down-stream analysis (Supplementary Table 1, Supplementary Dataset 1). Measurements of major short-chain fatty acids (SCFAs) in the cecum and colon showed a trend of incremental increases of absolute butyrate levels as AcGGM levels were increased (**Fig. 2**). Butyrate is commonly measured in the colon at levels that are approximately one-third of acetate, the most abundant SCFA[21]. Compared to control animals, the mean molar ratio of butyrate to acetate in samples from pigs fed 4% AcGGM increased in both the cecum (from 0.29:1 to 0.37:1 mM, p-value: 0.03) and colon (from 0.35:1 to 0.45:1 mM, p-value: 0.002) (Supplementary Dataset 1), suggesting that AcGGM inclusion promotes butyrogenic fermentation. While changes in SCFA and microbiome composition (**Fig. 2-3**) resulted from AcGGM inclusion, we observed no adverse effects on the host's physiology, with the average weight, feed conversion ratio, blood cell composition, T cell population, and colon morphology not differing between the control and AcGGM treatments (Supplementary Fig. 1, Supplementary Dataset 1).

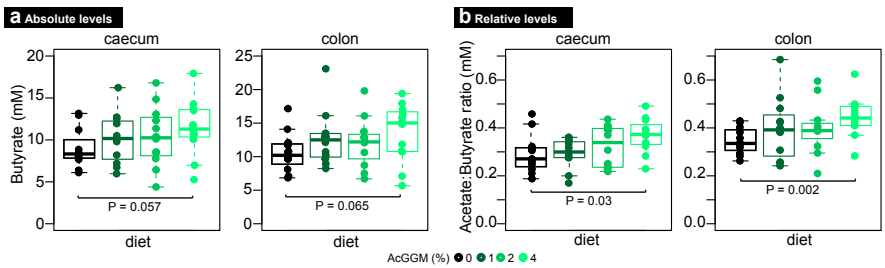


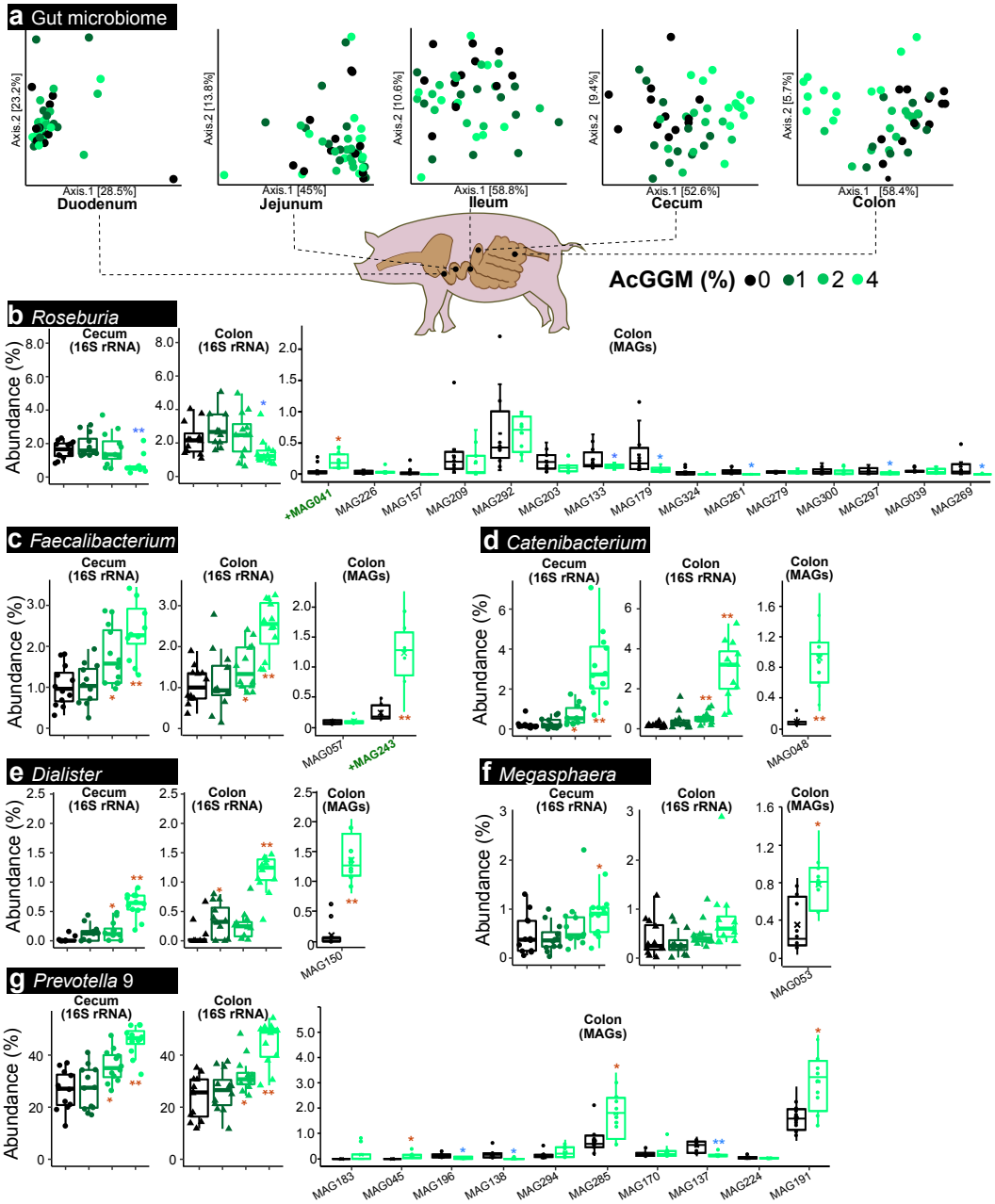
Fig. 2 Absolute levels of butyrate detected in the cecum and colon digesta of pigs fed the four different AcGGM diets with varying inclusion levels (0-4%). Raw SCFA measurements are presented in Supplementary Dataset 1. Bars represent median and boxes interquartile range determined from 12 animals analyzed per dietary group. Data were analyzed using a two-tailed *t* test.

168

169 Monitoring of temporal microbiome changes using 16S rRNA gene analysis over the month-
170 long trial indicated that inclusion of AcGGM into the piglets feed caused a pronounced shift in the
171 microbiome structural composition from the 21st day of the trial onwards (Supplementary Fig. 2b,
172 Supplementary Fig. 3). Spatial changes were also examined at the final sampling day and showed
173 typical patterns, whereby the structure of the gut microbiome varied by gut region (Supplementary
174 Fig. 2a, Supplementary Fig. 3, Supplementary Fig. 4). As expected, the AcGGM-effect was more
175 pronounced in the fibre-fermenting distal regions (cecum, colon) of the gut, where the relative
176 abundance of hundreds of phylotypes was observed to change (adjusted $p<0.05$) in response to
177 varying inclusion levels (**Fig. 3a**, Supplementary Dataset 2). To determine the effect AcGGM had on
178 microbiome function, we also closely examined the 355 metagenome assembled genomes (MAGs,
179 Supplementary Fig. 5a) that were reconstructed from metagenomic data generated from the colon
180 samples of each pig fed the control and 4% AcGGM diets (Supplementary Table 2). Taxonomic
181 affiliation of the MAGs was calculated via GTDB-Tk (Supplementary Dataset 3), while phylogeny was
182 inferred from a concatenated ribosomal protein tree (Newick format available in Supplementary
183 Dataset 4) that was constructed using MAGs from this study and 293 closely related reference
184 genomes.

185 Our target butyrogenic populations produced mixed results, whereby the 16S rRNA gene
186 relative abundance of *Faecalibacterium* affiliated phylotypes increased in response to increasing levels
187 of AcGGM (**Fig. 3c**, Supplementary Fig. 3c), whereas *Roseburia* affiliated phylotypes seemingly
188 decreased (**Fig. 3b**, Supplementary Fig. 3b). However, detailed analysis of *Roseburia*- affiliated MAGs
189 showed that specific phylotypes that encoded AcGGM-specific PULs were indeed stimulated by the
190 AcGGM fibre (**Fig. 3b**). Reputable fibre-fermenting populations affiliated to *Prevotella*[22] also showed
191 varying responses (**Fig. 3g**), with 16S rRNA gene relative abundance of individual phylotypes increasing
192 from 4% to 12% between the control and 4% AcGGM inclusion in both colon and cecum

193 (Supplementary Fig. 3a). Interestingly, relative abundance estimates of both 16S-rRNA gene OTUs and
194 MAGs indicated that phylotypes affiliated to non-fibre degrading taxa, such as *Catenibacterium*[23],
195 *Dialister*[24] and *Megasphaera*, demonstrated some of the highest dose-dependent increases in relative
196 abundance in response to AcGGM **(Fig. 3d-f)**, indicating that other underlying factors are likely
197 dictating microbiome structure, besides fibre degradation.



198

199 **Fig. 3. Effect of AcGGM-containing diets on the pig gut microbiome.** 16S rRNA gene amplicon analysis was used to monitor
200 the effect the AcGGM fibre had on the gut microbiome structure of weaned piglets. **a** Ordination plots of Bray-Curtis
201 distances between microbial communities from pigs feed either the control or AcGGM diets (at varying inclusion levels: 1%,
202 2% or 4%, 12 animals per dietary group). Samples were collected at day 27 of the feeding trials from various sections of the
203 small and large intestine. The AcGGM-effect was more pronounced in the fibre-fermenting distal regions (cecum, colon) of

204 the gut. **b** Relative 16S rRNA gene abundance of *Roseburia*-affiliated phylotypes (agglomerated at the genus level) in either
205 cecum or colon samples collected from pigs fed AcGGM diets with varying inclusion levels. MAG relative abundances were
206 determined from colon samples only, collected from pigs fed either the control or 4% AcGGM diets. Despite being suspected
207 as an active mannan degrader, 16S rRNA gene abundance of *Roseburia*-affiliated phylotypes decreased with increasing %
208 AcGGM. However, MAG relative abundance analysis of *Roseburia* populations showed that specific phylotypes that encoded
209 AcGGM-specific PULs (indicated by green text) were indeed stimulated by the AcGGM fibre. **c** The relative 16S rRNA gene
210 abundance of the genus *Faecalibacterium* was enhanced by the inclusion of AcGGM, as were affiliated MAGs in the colon.
211 The same trends were observed for phylotypes affiliated to *Catenibacterium* (**d**), *Dialister* (**e**), *Megasphaera* (**f**) and *Prevotella*
212 group 9 (**g**). Due to space constraints only a partial representation of *Prevotella*-affiliated MAG abundances is illustrated in
213 panel **g.** Bars represent median and boxes interquartile range determined from 12 animals analyzed per dietary group. Both
214 16S rRNA and MAG relative abundance data were analyzed using a two-tailed t test, with significant increases (red) and
215 decreases (blue) indicated between pigs fed either the control or AcGGM diets (* denotes $p<0.05$, ** denotes $p<0.001$).

216

217 **Targeted mannan PULs and butyrogenic activities were detected in *Faecalibacterium*- and**
218 ***Roseburia*-affiliated populations within the colon of AcGGM-fed pigs.**

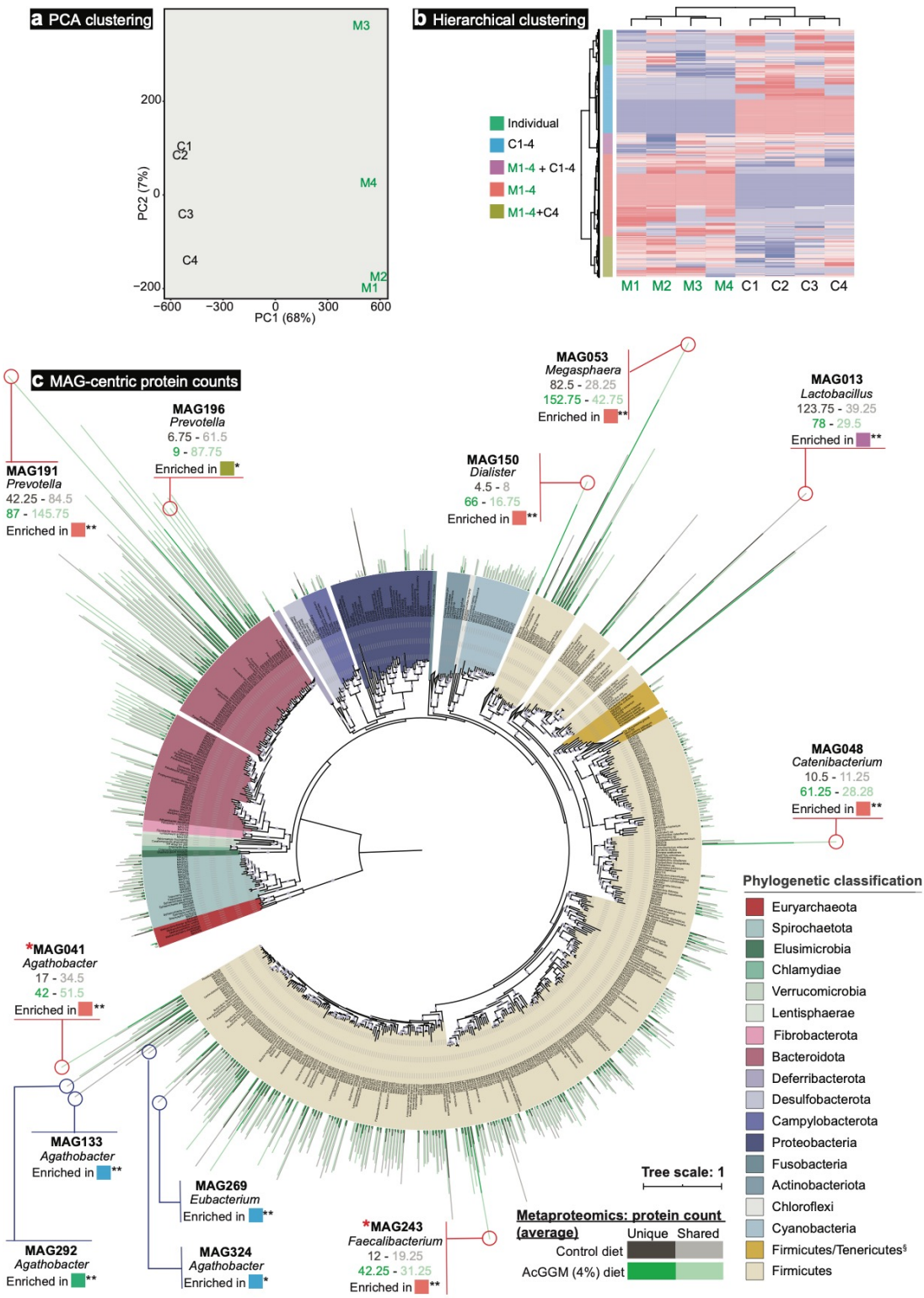
219 Because our primary goal was to elucidate whether our target, butyrogenic populations were
220 activated in response to AcGGM, we annotated MAGs affiliated to butyrate-producers and conducted
221 metaproteomic analysis with label-free quantification (LFQ) on randomly selected colon samples from
222 four control and four 4% AcGGM fed pigs (**Fig. 4a-b**), and mapped 8515 detected protein groups back
223 to our MAGs to identify functionally active populations **(Fig. 4c**) (Supplementary Datasets 5-6).
224 Community-wide analysis of the MAG genetic content (Supplementary Fig. 5b) from each sample and
225 clustering analysis of their detected proteins (**Fig. 4a-b**) further supported our 16S rRNA gene analysis,
226 reiterating that the microbiomes from piglets fed the control and 4% AcGGM diets were distinct.

227 Our MAG-centric multi-omic approach gave clear indications as to what effect the AcGGM fibre
228 had on putative butyrogenic *Roseburia* and *Faecalibacterium* populations in the distal gut of pigs.
229 Fifteen MAGs clustered with representative *Roseburia* spp. genomes (**Fig. 3b, Fig. 4**), which reflected
230 the multiple *Roseburia*-affiliated phylotypes that were predicted with our 16S rRNA gene analysis
231 (Supplementary Fig. 3b, Supplementary Dataset 2). In general, the relative abundance of *Roseburia*-
232 affiliated MAGs (**Fig. 3b**) and detected proteins (**Fig. 4c**) were observed at either static or lower levels
233 in AcGGM-fed pigs, reiterating our initial 16S rRNA gene observations that AcGGM negatively affected
234 *Roseburia* populations (**Fig. 3b**). However, one specific *Roseburia*-affiliated population (MAG041) was

235 detected at significantly higher abundance levels (p=0.0016) (**Fig. 3b**), and its detected proteins were

236 enriched (p<0.001) in the 4% AcGGM pig samples compared to the control (**Fig. 4b-c**, Supplementary

237 Dataset 7).



238

239 **Fig. 4. Genome-centric metaproteomic analysis of colon samples collected from pigs fed either the control of 4% AcGGM**
240 **diet. a** Principle Component Analysis of metaproteomes generated from randomly selected colon samples collected from

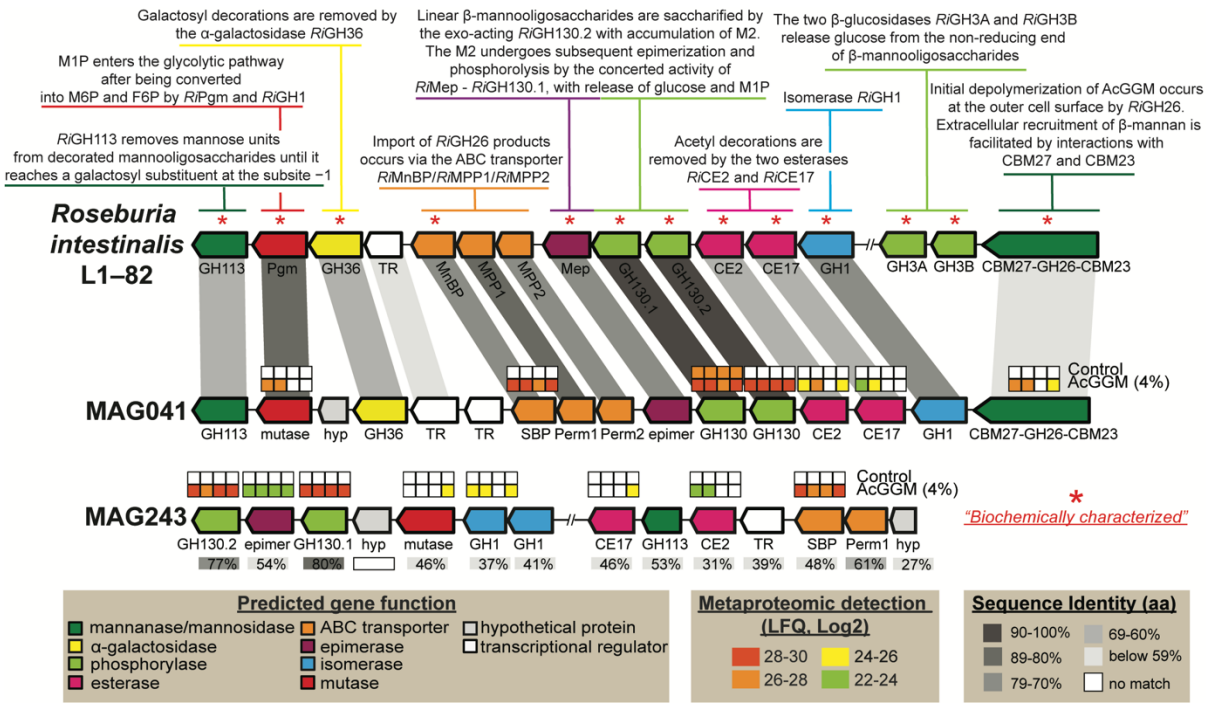
241 pigs fed either the control (C1-4) or 4% AcGGM (M1-4) diet. Clear separation was observed between the two diets,
242 highlighting that the detected proteins in the colon microbiomes from piglets fed the control and 4% AcGGM diets were
243 distinct. **b** Hierarchal clustering and heatmap of detected protein group abundance profiles. Rows are scaled so that red
244 represents the highest abundance for that protein group and blue the lowest. Five different clusters were observed, with
245 protein groups differentially detected in AcGGM fed pigs (M1-4: red), control pigs (C1-4: blue), all pigs (M1-4 + C1-4: purple),
246 AcGGM fed pigs plus one control (M1-4 + C4: brown) and only in individual pigs (Individual: green). **c** Phylogeny and
247 metaproteomic detection of 355 MAGs sampled from the colon of weaned piglets. This maximum likelihood tree is based on
248 an alignment of 22 concatenated ribosomal proteins from the 355 MAGs reconstructed in this study from 24 colon
249 metagenomes (12 control pigs, 12 fed 4% AcGGM diet), in addition to 239 reference genomes closely related to the MAGs.
250 Branches are shaded with colour to highlight phylum-level affiliations (see legend). Coloured bars on the outside of the tree
251 depict the average number of unique and shared protein groups detected for each MAG in four randomly selected samples
252 that were analyzed from pigs fed either the control- (grey) or 4% AcGGM-diet (green). Total number of proteins for each
253 MAG in each pig is detailed in Supplementary Dataset 5. Purple circles on the inside of the tree represent nodes with
254 bootstrap support ≥ 70%, relative to size. MAG041 and MAG243 were found to encode CE2/CE17-containing mannan PULs
255 are indicated by *. All MAGs depicted in **Fig. 6 (blue circles) and Fig. 7 (red circles)** are listed, with average detected protein
256 counts in both diets, which metaproteomic expression cluster they are enriched in (part **b**, adjusted p-values are indicated:
257 * denotes p<0.05, ** denotes p<0.001), and their MAG taxonomic affiliation determined via GTDB-Tk. Recently reclassified
258 *Roseburia* species are denoted as *Agathobacter*. The full tree in Newick format is provided in Supplementary Dataset 4.

259

260 Closer examination of MAG041 revealed a putative CE2/CE17-containing mannan-degrading PUL
261 that was absent in the other *Roseburia*-affiliated MAGs and was differentially expressed in the AcGGM
262 diet **(Fig. 5**). Importantly, the MAG041 mannan PUL exhibited gene synteny to the *R. intestinalis* strain
263 L1-82 PUL whose biochemical properties we recently characterized in detail[5] **(Fig. 5**). The predicted
264 multi-modular mannanase (CBM27-GH26-CBM23) in the MAG041 mannan PUL is homologous to the
265 GH26 in *R. intestinalis* L1-82 (48% identity over 87% of the sequence), and can be presumed to fulfill
266 the same function – "selfishly" breaking down AcGGM fibres at the cell surface prior to intracellular
267 transport. Besides the detection of GH26 and esterases in AcGGM-fed pigs, other mannan-specific
268 enzymes also responded to the dietary shift within the MAG041 mannan PUL, including a
269 phosphoglucomutase, a multiple-sugar binding protein, a GH130.1 4-*O*-β-D-mannosyl-D-glucose
270 phosphorylase and a GH130.2 β-1,4-manno-oligosaccharide phosphorylase (**Fig. 5**).

271 In contrast to *Roseburia*-affiliated MAGs, only one MAG clustered with *F. prausnitzii* (MAG243,
272 **Fig. 4**, Supplementary Dataset 3), implying that the multiple phylotypes that were predicted with our
273 16S rRNA gene data (Supplementary Fig. 3c, Supplementary Dataset 2) encode high genome similarity
274 and coverage and thus likely co-assembled into a representative population-level MAG. The relative
275 abundance of MAG243 in the distal gut of pigs fed 4% AcGGM increased approximately 5-fold to
276 ~1.24% (p<0.001) (**Fig. 3c**), while metaproteomic hierarchical clustering analysis showed that MAG243

277 proteins were enriched in the same samples (**Fig. 4b-c**, Supplementary Dataset 7). Similar to MAG041,
278 MAG243 was found to encode a CE2/CE17-containing mannan PUL, which was broadly detectable in
279 the presence of AcGGM but absent in the control samples (**Fig. 5**, Supplementary Dataset 6). While
280 the MAG243 mannan PUL contained two GH130 manno-oligophosphorylases, a mannose 6-
281 phosphate isomerase, phosphoglucomutase and two carbohydrate esterase (CE17 and CE2), it lacked
282 a GH26 mannanase representative, which suggests that *F. prausnitzii* is likely preferentially targeting
283 the shorter acetylated manno-oligosaccharides that form part of the AcGGM structure (**Fig. 1**). In
284 addition to the mannan-PULs of MAG041 and MAG243 being activated in AcGGM-fed pigs, their
285 butyrogenic pathways were also detected at high levels, based on label-free quantification (LFQ)
286 scores of detected proteins **(Fig. 6-7**, Supplementary Dataset 6), suggesting that both populations can
287 convert mannan to butyrate (Supplementary Dataset 6).



288

289 **Fig. 5. Metaproteomic detection of CE2/CE17-containing mannan-PULs encoded in *Roseburia*- (MAG041) and**
290 ***Faecalibacterium*- (MAG243) affiliated MAGs in pigs fed with either the control or 4% AcGGM diet.** Predicted gene
291 organization and annotated gene function is colour-coded and largely derived from previous biochemical and structural
292 characterization of the mannan degradation cluster (characterized genes indicated with *) in *R. intestinalis* L1-82[5]. Gene
293 synteny and identity % between mannan PULs found in in *R. intestinalis* L1-82, MAG041 and MAG243 are indicated in grey
294 boxes. Heat maps above detected enzymes show the LFQ detection levels for the four replicates sampled in control and 4%
295 AcGGM-fed pigs. LFQ values of proteins from both clusters are in Supplementary Dataset 6. The predicted multi-modular
296 mannanase (CBM27-GH26-CBM23) from MAG041 was the only extracellular protein in the locus, and the only extracellular
297 mannanase expressed in response to AcGGM inclusion.

298

299 **Removal of 2-*O*-, 3-*O*- and 6-*O*-mannose acetylations are the key enzymatic activities required for**
300 **accessing AcGGM.**

301 A crucial step in the utilization of mannans as an energy source is the deacetylation of 2-*O*-, 3-*O*-
302 and 6-*O*-mannose residues, which allows the subsequent breakdown of the sugar-containing
303 backbone of the fibre. In *R. intestinalis* L1-82, AcGGM deacetylation occurs via the synergistic actions
304 of two carbohydrate esterases (*Ri*CE2 and *Ri*CE17) that exert complementary specificities[9]. MAG041
305 and MAG243 both encoded CE2 homologues within their mannan PULs, sharing 63% and 31% identity
306 (respectively) to *Ri*CE2, which has demonstrated activity on 3-*O*-, (4-*O*-) and 6-*O*-acetylations, and is
307 mannan specific[9]. For CE17, MAG041 and MAG243 homologs shared 65% and 46% identity
308 (respectively) with *Ri*CE17, including the active site residues and the aromatic stacking tryptophan
309 (Trp326), which in *Ri*CE17 are associated with 2-*O*- acetylation specificity[9]. Broader screens of our
310 MAG data revealed other CE2/CE17-containing PULs within Firmicute-affiliated MAGs from the pig
311 colon microbiome (Supplementary Fig. 6), however aside from MAG041 and MAG243, they originated
312 from populations whose MAG relative abundance was very low (<0.05%) and metabolic activity was
313 undetectable via metaproteomics in any of the control or AcGGM diets (**Fig. 4**). Finally, the differential
314 proteomic detection of MAG041 and MAG243 CEs in pigs fed AcGGM diets (**Fig. 5**), strengthened our
315 hypothesis that both these populations can accommodate the unique features of the AcGGM fibre
316 and are actively engaging in its utilization *in vivo*.

317

318 **Beyond the target populations consuming AcGGM, we also observed a pronounced "butterfly**
319 **effect"**

320 MDFs studies to date have eloquently highlighted that metabolic symmetry between individual
321 fibres and microbiota can be used to stimulate specific populations[1,16]. However, deeper microbiome

322 effects that result from a target species being enriched via MDFs are poorly understood. Although the
323 specificity of the AcGGM fibre matched selected mechanistic features of our target populations, our
324 data showed that AcGGM dietary intervention reverberated further down the microbial trophic
325 networks that support conversion of dietary fibre into keystone SCFAs that are of nutritional value to
326 the host animal. In particular, we observed varying effects upon different non-AcGGM-degrading
327 butyrogenic populations belonging to clostridial cluster XIVa (Lachnospiraceae: *Roseburia* spp.,
328 *Eubacterium rectale*, *Butyrivibrio* and *Pseudobutyrivibrio*) and cluster IV (Ruminococcaceae:
329 *Faecalibacterium prausnitzii*-related bacteria)[25] (**Fig. 6**, Supplementary Dataset 6). Specific populations
330 that were closely related to AcGGM-degrading MAG041 were found to be either unaffected (MAG292)
331 or metabolically suppressed (MAG133) in the presence of 4% AcGGM. Indeed, detected proteins from
332 MAG133 were enriched in pigs fed the control diet (adj. $p<0.001$, Supplementary Dataset 7), and
333 enzymes associated with butyrate production were largely undetected in 4% AcGGM-fed pigs (**Fig. 6**,
334 Supplementary Dataset 6). In most cases, non-AcGGM degrading, butyrogenic populations were
335 seemingly utilizing sugars found in starch, arabinoxylan and/or arabinogalactan fibres that were
336 detected in the basal feed components using Micro Array Polymer Profiling (MAPP; Supplementary
337 Dataset 8). Broader analysis of our omic data identified multiple abundant populations with similar
338 metabolic capabilities, suggesting that suppressed butyrogenic populations were possibly being
339 outcompeted by Prevotella-affiliated populations such as MAG285 (see below). While SCFA
340 measurements indicated an increase in relative butyrate levels in AcGGM-fed pigs (**Fig. 2b**), the
341 indirect effects of our MDF on this dynamic functional group are likely dampening an even greater
342 influence of AcGGM-stimulated *Roseburia* and *Faecalibacterium* populations upon the broader
343 microbiome and host animal (Fig. 6).

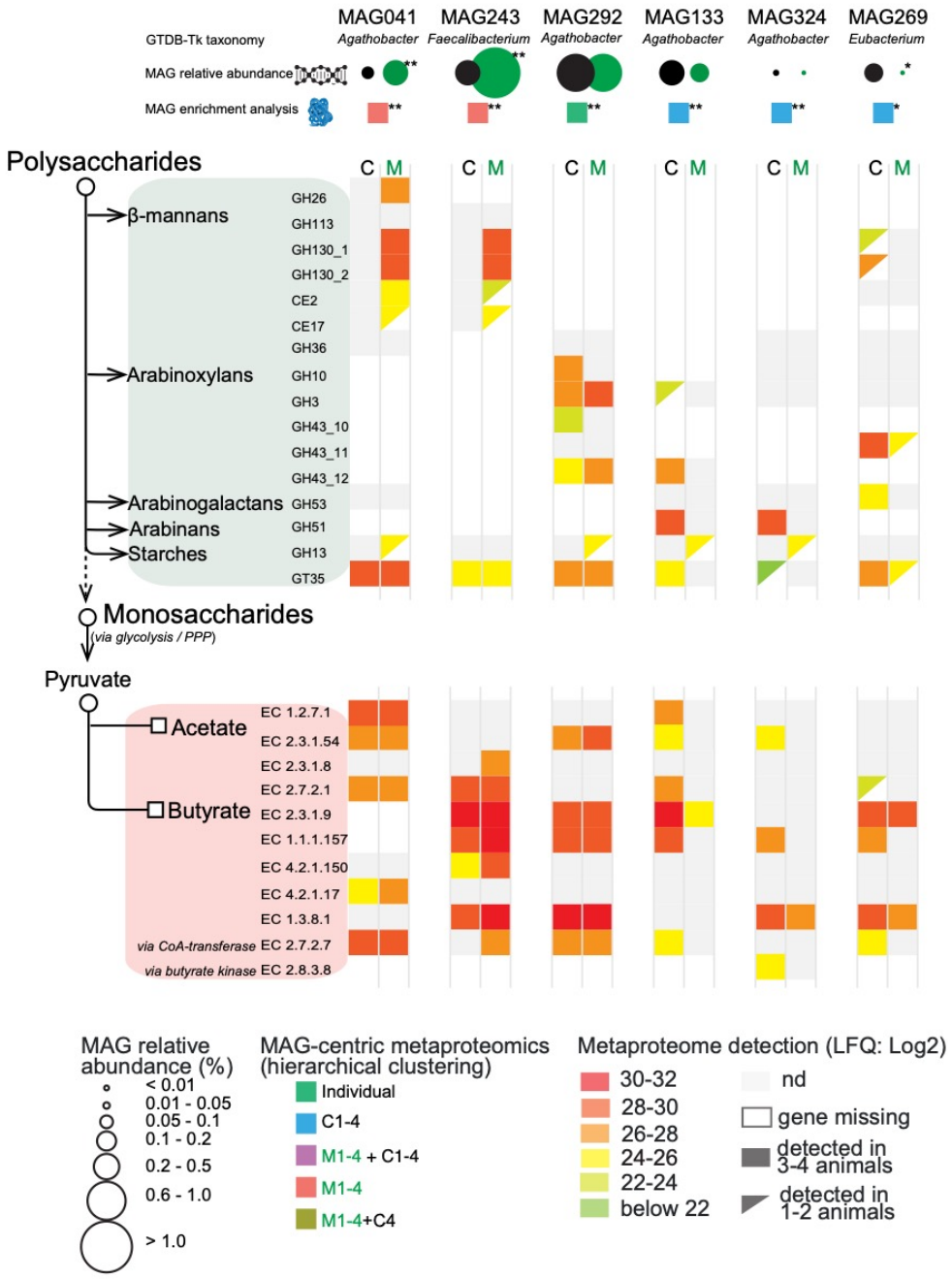


Fig. 6. The differing metabolisms of butyrate-producers in the colon microbiome of pigs fed AcGGM. A plethora of MAGs affiliated with putative butyrate-producers were recovered from the pig colon metagenome, which varied in their fibre degrading capacity and detected activity in either the control (C) or 4% AcGGM (M) diets. Here, we present a representative selection of populations that: encoded specific AcGGM degrading capabilities (MAG041 and MAG243), were abundant across both diets with no apparent effect (MAG292), were significantly enriched in pigs feed the control diet (MAG133, MAG324 and MAG269), and encoded an alternative pathway for butyrate production (via Butyrate kinase) (MAG324). GTBD-Tk inferred taxonomy (recently reclassified *Roseburia* species are denoted as *Agathobacter*), MAG relative abundance (control: black circles, AcGGM: green circles) and MAG-centric metaproteomic enrichment analysis is indicated (see **Fig. 4c**, significant differences denoted by adjusted p-values: * p<0.05, ** p<0.001).

Besides butyrate-producers, MAG-centric metaproteomic enrichment analysis showed that fibre-degrading *Prevotella*-affiliated populations experienced contrasting effects as a result of AcGGM

357 inclusion. For example, MAG191 was found to account for the highest levels of detectable proteins in
358 our datasets (**Fig. 4c**, Supplementary Dataset 6), which were enriched in clusters differently detected
359 in the presence of 4% AcGGM ($p<0.001$, Supplementary Dataset 6). Pathway annotation of AcGGM-
360 enriched *Prevotella* populations (such as MAG191, MAG196, MAG285, see **Fig. 4c**, **Fig. 7**,
361 Supplementary Dataset 6-7) indicated active metabolism of dietary fibres such arabinoxylans, starch,
362 glucans (e.g. cellobiose), α-galactans and mannose sugars (detected via MAPP analysis,
363 Supplementary Dataset 8) as well as acetate, succinate and/or propionate production, which were all
364 detected with higher LFQ scores in AcGGM-fed pigs (**Fig. 7**, Supplementary Dataset 6). However, many
365 CAZymes and fermentation enzymes from the aforementioned MAGs were also detected across both
366 diets (albeit at varying LFQ values), while other *Prevotella* populations were specifically enriched in
367 pigs fed the control diet (i.e. MAG034: adj. $p<0.001$, Supplementary Dataset 7), suggesting that
368 *Prevotella*-driven baseline consumption of basal feed fibre was occurring irrespective of AcGGM
369 inclusion (**Fig. 7**, Supplementary Dataset 6).

370 Several mannan-targeting PULs were identified in *Prevotella*-affiliated MAGs that were configured
371 in an archetypical "Bacteroidetes-format", which combines outer-membrane transport and
372 carbohydrate-binding SusC/D-like proteins as well as CAZymes[26] (Supplementary Fig. 5a). In particular,
373 a PUL recovered from MAG196 encoded predicted SusC/D-like proteins, mannanases (GH26, GH5_7),
374 mannosyl-phosphorylases (GH130) and an esterase, although neither the mannanases nor the
375 esterase were detected in the metaproteomes recovered from the AcGGM-fed pigs (**Fig. 7**,
376 Supplementary Fig. 6, Supplementary Dataset 6). In addition, we speculate that MAG196 and MAG191
377 are perhaps capable of metabolizing elements of the AcGGM fibre such as the α-galactose side-chain
378 or deacetylated manno-oligosaccharides, which was inferred via detected GH36 and GH130
379 representatives (**Fig. 7**, Supplementary Fig. 6, Supplementary Dataset 6).

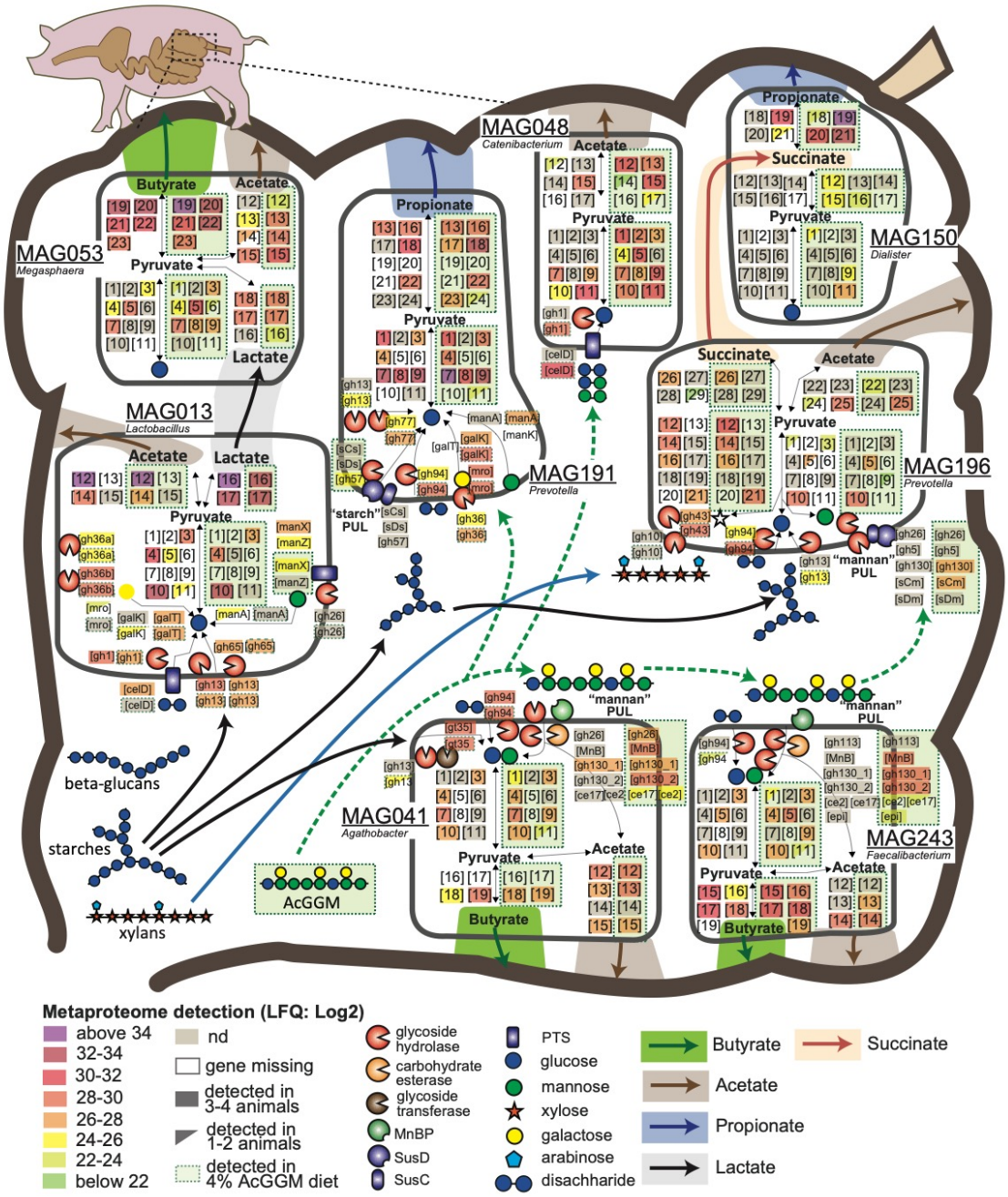


Fig. 7. Selected metabolic features of the porcine colon microbiome in response to AcGGM dietary intervention, as inferred from genome and proteome comparisons. The different metabolic pathways (fibre deconstruction, glycolysis, pentose-phosphate pathway and SCFA production) are displayed for each population MAG. Graphical representation of pathways, enzymes, CAZymes, and cellular features are based on functional annotations that are depicted as numbered or abbreviated gene boxes, which are additionally listed in Supplementary Dataset 6. Metaproteomic analysis (detected genes and enzyme complexes) is highlighted according to the different LFQ values of the detected proteins sampled at a single time-point (28 days) either from animals fed the control or 4% AcGGM diet (green perforated boxes). The main dietary fibres (starches, xylans, glucans and mannans) and SCFAs (butyrate, acetate, propionate, lactate and succinate) are represented by large colored arrows. GTBD-Tk inferred taxonomy is included (recently reclassified *Roseburia* species are denoted as *Agathobacter*). Gene names and abbreviations are also provided in Supplementary Dataset 6.

392 Dietary inclusion of AcGGM also unexpectedly led to pronounced metabolic shifts in numerous
393 phenotypically and phylogenetically diverse microbial populations that are not normally associated
394 with fibre hydrolysis. Mirroring our 16S rRNA gene analysis, MAGs affiliated to the genera *Dialister*
395 (MAG150), *Catenibacterium* (MAG048) and *Megasphaera* (MAG053) were the three highest enriched
396 populations in expression clusters differentially detected in the AcGGM diet (adj. $p<0.001$), although
397 none were found to encode CE2/CE17-containing mannan PULs (**Fig. 3-4**, **Fig. 7**, Supplementary
398 Dataset 6). In particular, the MAG048 proteome included a putative sugar phosphotransferase system
399 (PTS) and GH1 phospho-β-glucosidases (EC 3.2.1.86) that are predicted to catalyze the
400 phosphorylation of disaccharides (such as cellobiose and mannobiose[27,28]) and hydrolyze the PTS-
401 transported sugars into D-glucose and D-glucose 6-phosphate. Concomitantly, glycolysis and
402 acetogenesis pathways from MAG048 were also highly detected in AcGGM-fed pigs (**Fig. 7**,
403 Supplementary Dataset 6). Together, this data suggests that certain *Catenibacterium* populations are
404 advantageously consuming AcGGM-derived disaccharides that were either available as a fraction of
405 the original AcGGM fibre preparation, have been generated via the actions of other fibre-degrading
406 populations or have become available via new ecological niches that have been created via the
407 AcGGM-derived structural shifts in the microbiome.

408 Both MAG053 (*Megasphaera*) and MAG150 (*Dialister*) were predicted via our multi-omics
409 approach to metabolize SCFAs such as lactate and succinate that were generated endogenously by
410 *Lactobacillus*- (i.e. MAG013) and *Prevotella*- (i.e. MAG285) affiliated populations, and produce
411 butyrate and propionate, respectively (**Fig. 7**, Supplementary Dataset 6). Neither of these cross-
412 species relationships had any evidence of being directly influenced by AcGGM and are presumably
413 being driven by underlying availabilities of fermentation intermediates (e.g. lactate and succinate)
414 resulting from as-yet uncharacterized niche fluctuations. Collectively, these data highlight that while
415 a highly tailored MDF can directly activate enzymatically equipped populations in an endogenous
416 microbiome, they can also inadvertently alter the metabolisms of other non-mannolytic populations
417 creating effects that should be quantified to ascertain the true modulation potential of an MDF. The

418 community dynamics that drive fibre-fermentation is not well understood and is likely a contributory
419 factor that restricts broader prebiotic fibre development.

420

421 **CONCLUSION**

422 Here, we have advanced the MDF concept by showing that the defined relationship between
423 fibre and microbe can be used to directly target specific microbiota in highly complex and competitive
424 gut ecosystems. This was achieved by matching the structural features of our characterized AcGGM
425 fibre to a specific genetic locus in our target microbiota, for which there exists comprehensive
426 biochemical information pertaining to the encoded enzymatic functions.

427 Using integrated multi-omics analysis, we illustrate the impact of the AcGGM structural configuration
428 on microbial uptake and metabolism within the distal regions of the porcine digestive tract, with the
429 key driver of AcGGM selectivity being the presence of acetylations of mannan, as well as carbohydrate
430 composition and size[29]. Preserving the complexity of AcGGM resulted in a highly specific shift in the
431 composition of the colon microbiome from weaned piglets, with increases in the relative
432 butyrate:acetate ratios and no adverse effect on host growth performance or health status. We
433 showed that the AcGGM fibre activated a metabolic response in specific *Roseburia* and
434 *Faecalibacterium* populations *in vivo*, as it did in the previous *in vitro* experiments[5,10], with both
435 populations expressing proteins from unique CE2/CE17-containing mannan PULs that are homologous
436 to a biochemically characterized representative in *R. intestinalis* L1-82. In conclusion, our data provide
437 a foundation for modulatory strategies to design and match custom dietary fibres to unique enzymatic
438 features of their target organisms. However, they also underscore the fact that the greater network
439 of interconnected metabolic exchanges and trophic structures inherent to the gut microbiome are
440 highly susceptible to minor dietary interventions. Ultimately, further research is required to determine
441 whether AcGGM MDFs can confer improvements to animal performance metrics (such as growth and
442 disease resistance) in the context of full-scale livestock production.

443

444 **DATA AVAILABILITY**

445 All sequencing reads have been deposited at the NCBI sequence read archive under BioProject
446 PRJNA574295, with specific numbers listed in Supplementary Table 2. All annotated MAGs are publicly
447 available via doi: 10.6084/m9.figshare.9816581. The proteomics data has been deposited to the
448 ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner
449 repository[30] with the dataset identifier PXD015757.

450

451 **ACKNOWLEDGEMENTS**

452 We are grateful for support from The Research Council of Norway (Bionær program 244259,
453 208674/F50, 270038, FRIPRO program, PBP: 250479), as well as the European Research Commission
454 Starting Grant Fellowship (awarded to PBP; 336355 - MicroDE). The sequencing service was provided
455 by the Norwegian Sequencing Centre (www.sequencing.uio.no), a national technology platform
456 hosted by the University of Oslo and supported by the "Functional Genomics" and "Infrastructure"
457 programs of the Research Council of Norway and the Southeastern Regional Health Authorities. We
458 thank Sivert Dæhli for performing the detailed analysis of degree of polymerization of the AcGGM.

459

460 **AUTHOR CONTRIBUTIONS**

461 L.M., J.C.G., P.B.P. and B.W. conceived the study, performed the primary analysis of the data and wrote
462 the paper (with input from all authors). L.L. and M.Ø. designed, performed and analyzed the animal
463 experiments. S.L.L.R., L.H.H., M.Ø.A., T.R.H. and J.D. generated the data and/or contributed to the data
464 analyses. W.G.T.W. and C.T-J. generated and analyzed MAPP data. N.T., V.L. and B.H. annotated and
465 curated the MAGs and identified carbohydrate-active enzymes.

466

467 **COMPETING INTERESTS**

468 The authors declare there are no competing financial interests in relation to the work described.

469

470 **MATERIALS AND METHODS**

471 **Animals, diets and experimental design.**

472 Animal care protocols and experimental procedures were approved by the Norwegian Animal
473 Research Authority, approval no. 17/9496, FOTS ID 11314 and treated according to institutional
474 guidelines. A total of 48 cross bred piglets (Landrace x Yorkshire), 24 male and 24 female, with an
475 average initial body weight (BW) of 9,8 ± 0,5 kg and weaned at 28 days of age, were sorted by litter,
476 sex and weight and randomly divided into twelve pens of four animals, with three pens fed each of
477 the four diets. The piglets in each pen were trained to feed at individual stations containing their
478 respective feed. The animals were housed in an environmentally controlled facility with plastic flooring
479 and a mechanical ventilation system. The temperature of the room was maintained at 22°C.

480 Piglets were fed cereal-based diets containing increasing levels of AcGGM in the diets (1, 2
481 and 4%). Inclusion levels were selected based on in-house experience from previous feeding trials and
482 communication with the feed industry, where an inclusion level of 1-4% of MDF in feed would be
483 realistic. The highest inclusion level at 4% was selected in order to test a dose response. Diets were
484 pelleted with a 3 mm diameter feed formulated to meet the requirements for indispensable amino
485 acids and all other nutrients (NRC, 2012). The composition of diets is listed in Supplementary Table 3.
486 Pigs were fed semi-ad libitum twice a day at a feeding level equal to about 5% of body weight. To
487 evaluate growth performance, the BW of each pig was recorded at the beginning and once a week.
488 Feed consumption was recorded for each individual pig during the experiment to calculate individual
489 weight gain and feed intake. After each meal, feed leftovers were registered, dried and subtracted
490 from the total feed intake.

491

492 **Production of AcGGM.**

493 AcGGM oligosaccharides for the feeding trial were produced from Norway spruce chips milled
494 with a hammer mill to <2 mm size. Wood chips were then steam-exploded on a pilot scale steam
495 explosion rig (100L reactor vessel) at the Norwegian University of Life Sciences (NMBU). The steam
496 explosion was conducted in batches of approximately 6kg dry matter, 14.5 bar pressure (equivalent
497 to 200° C), with 10 minutes residence time. The pH in the collected biomass slurry after the steam
498 explosion was ~3.7, which corresponds to a combined severity factor R'_0=1.70 for the process. The
499 severity was calculated by $R'_0 = (10^{-pH}) \times (t \times e^{(Texp-100)/14.75})$[31]. Steam exploded wood was collected in
500 50 L plastic buckets that were topped up with hot (~70° C) water. The slurry was transferred to a
501 60Lcider press (Speidel, Germany) and the liquid fraction was pressed out. Milled wood was collected,
502 soaked in hot water again, and pressed for the second time. The liquid fraction was collected and
503 recirculated through a bag filter 50µm pore WE50P2VWR (Allied filter systems, England) partly filled
504 with the wood particles as a filter aid. Once free of floating wood particles, the liquid fraction of
505 hemicellulose was filtered through a 5-kDa spiral wound Polysulphone/polyethersulphone
506 ultrafiltration membrane, GR99PE polyester (Alfa Laval, Denmark) that was deliberately fouled to
507 prevent larger oligosaccharides from running through the permeate, using a GEA pilot-scale filtration
508 system Model L (GEA, Denmark). The fraction retained by the membrane was concentrated by
509 nanofiltration using a TriSep XN 45, which had a higher efficiency for permeating water. The filtrate
510 was further concentrated by vacuum evaporation (set to 65 °C) and the concentrate was freeze-dried
511 and homogenized with a grain mill. The final product consisted of 0.9% rhamnose, 2.7% arabinose,
512 13.7% xylose, 58.9% mannose, 14.9% glucose and 9.4% galactose (determined by gas chromatography
513 as alditol acetates after sulfuric acid hydrolysis as described previously[32]). AcGGM contained 0.73 %
514 ash and 2.4% protein (quantified from total nitrogen by the Kjeldahl method). The Man:Glc:Gal ratio
515 in the mannan was 4:1:0.6, and the DA=0.35[9] (determined by acetate release from NaOH treated
516 AcGGM by the same method as described in the SCFA section below). Moreover, previous two-
517 dimensional (2D) ^{1}H-^{13}C heteronuclear single quantum correlation (HSQC) NMR analysis of the AcGGM
518 used in this study, showed the distribution of acetylations on mannose units, indicating that spruce

519 mannan hydrolysate without enzyme addition contained prevalently 2-O-, some 3-O- acetylations,
520 and a lower degree of 4-O- and 6-O-acetylations[9]. The dry matter content was determined by drying
521 0.2g of sample at 105°C for 20 hours. The remaining sample was burned at 600°C for 24 hours in an
522 oven (Carbolite, Sheffield, England) to determine ash content. All measurements were performed in
523 triplicates.

524

525 **Fecal scoring.**

526 During the experiment, fecal consistency was assessed using a scoring system developed by Pedersen
527 and Toft[33] to improve and help standardize current protocols for clinical characterization of fecal
528 consistency. The scoring was based on the following 4 consistency categories: score 1 = firm and
529 shaped, score 2 = soft and shaped, score 3 = loose and score 4 = watery. Samples with score 3 or 4 are
530 considered diarrheic. Daily fecal scores for each pen were recorded throughout the trial.

531

532 **pH measurements.**

533 The pH of digesta samples from duodenum, jejunum, ileum, cecum and colon were measured
534 immediately after slaughter. Samples were placed in universal containers and pH measurements made
535 using an Inolab pH7110 pH meter (WTW, Germany).

536

537 **Blood sampling and flow cytometry.**

538 Blood samples were collected from the same six piglets per diet at 0, 7 and 27 feeding days. The blood
539 samples were taken 1-2 hours post-prandial by venipuncture in the jugular vein while pigs were kept
540 on their backs. Non-heparinised and K_3EDTA vacuum tubes (Beckman Dickson Vacutainer System)
541 were used to recollect serum and whole blood. Serum was isolated immediately by centrifugation at

542 1,500 x *g* at 4°C for 15min. Serum samples were split in PCR-tubes (200 µL) and stored at -80°C until
543 analysis. Hematological and clinical analyses were performed with an Advia® 2120 Hematology System
544 using Advia 2120 MultiSpecies System Software and clinical chemistry analyses were performed with
545 Advia 1800 Chemistry System (both from Siemens AG Healthcare Sector).

546 For flow cytometry analysis, whole blood was diluted 1:1 in RPMI 1640 and kept on ice until single
547 cells isolation. For the isolation of peripheral blood mononuclear cells (PBMCs) blood was purified by
548 centrifugation in a Ficoll gradient (Kreuzer et al. 2012). Then, isolated PBMCs were incubated with
549 Fixable Yellow Dead Cell Stain Kit (Life Technologies, Thermo Fisher Scientific Inc.) followed by primary
550 monoclonal antibodies (mAbs), brief incubation with 30% normal pig serum to block Fc-receptors, and
551 finally fluorescence-labeled secondary antibodies (Abcam plc, UK). To detect the intracellular CD3
552 epitope, surface-labeled cells were permeabilized with Intracellular Fixation and Permeabilization
553 Buffer Set (eBioscience, Affymetrix Inc.) according to the manufacturer's instructions. Labeled cells
554 were analyzed on a Gallios Flow Cytometer (Beckman Coulter, Inc.) and data were processed using
555 Kaluza 1.5 software (both Beckman Coulter, Inc.). Cell gates were designed to select for single and
556 viable mononuclear cells. Defined markers were used to identify the different immune
557 subpopulations. For monocytes, antibodies against CD45, CD3, CD14, CD163 and MHCII were used. To
558 analyze regulatory T cells (T reg) the following antibodies were used: CD45, CD3, TCR γ/δ, CD4, CD8,
559 FOXp3 and CD25, while CD45, CD8, NKp46, CD4, CD8, Ki67 and CD27 were used to identify T and NK
560 cells. The gating strategy used for flow cytometric data is presented in Supplementary Figure 7.

561

562 **Analysis of Serum Cytokines: MULTIPLEX**

563 Expression of GMCSF, IFNG, IL-1A, IL1B, IL-1RA, IL-2, IL-4, IL-6, IL-8, IL-10, IL-12, IL-18 and TNFα were
564 measured in serum samples using MILLIPLEX MAP Porcine Cytokine and Chemokine Magnetic Bead
565 Panel - Immunology Multiplex Assay (Merck Millipore) following the manufacturer instructions. The
566 measurement was performed using a Bio-Plex MAGPIX Multiplex Reader (BIO-RAD).

567

568 **Small Intestine Morphology.**

569 The samples of the small intestine were collected on day 0 and 28 for determination of intestinal
570 morphology and integrity. Intestinal morphological measurements included the following indices:
571 villus height (VH), crypt depth (CD) and VH:CD. Mean values of VH, CD and their ratio were calculated.
572 Histology evaluation was performed by the Veterinary Histophalogy Center, VeHiCe, Chile.

573

574 **SCFA analysis.**

575 Samples of digesta from duodenum, jejunum, ileum, cecum and colon of individual pigs were collected
576 for SCFA analysis. 250 mg or 250 µL of a sample, depending on the source site were mixed 1:1 with
577 4mM H_2SO_4, homogenized by shaking at room temperature for one hour, and centrifuged at 12000xg
578 for 10 minutes. The supernatant was collected with a syringe and filtered through a 0.22 µm pore
579 syringe filter. Samples were stored at -20 °C and centrifuged at 12000xg before transferring aliquots
580 into HPLC vials for analysis. SCFA content was analyzed by HPLC using a REZEX ROA-Organic Acid H+
581 (Phenomenex, Torrance, California, USA) 300x7.8 mm ion exclusion column, isocratic elution with 0.6
582 mL/min 4 mM H_2SO_4 at 65 °C. Eluting analytes were detected by UV at 210 nm. The same HPLC method
583 was used to determine the degree of acetylation, by measuring the acetate released from samples of
584 AcGGM dissolved in 100mM KOH and relating the values to the mannose content of the sample. All
585 data were analyzed using a two-tailed *t* test.

586

587 **Micro Array Polymer Profiling (MAPP)**

588 All cell wall preparations were analysed similar to Moller et al[34]. Cell-wall glycans were sequentially
589 extracted from 10mg of each sample by incubation with 2ug/ml cellulase, pH8.9 in 20mM tris buffer
590 for 16hrs, 45°C to release residual glycans bound to cellulose. Each extract was mixed 50/50 with

591 glycerol buffer (55.2% glycerol, 44% water, 0.8% Triton X-100) and spotted with four 5-fold dilutions
592 and two technical replicates onto nitrocellulose membrane with a pore size of 0.45µm (Whatman,
593 Maidstone, UK) using an Arrayjet Sprint (Arrayjey, Roslin, UK). The arrays were probed using
594 antibodies from the JIM, LM (Plant Probes), BioSupplies, series and a Carbohydrate Binding Module
595 (CBM, NZYTech) detected using anti-rat (JIM, LM), anti-mouse (BioSupplies) or anti-His (CBM)
596 secondary antibodies conjugated to alkaline phosphatase (Sigma Aldrich) as described in Pedersen et
597 al[35]. The microarrays were developed using an NBT/BCIP substrate containing Nitro Blue Tetrazolium
598 (NBT) and 5-bromo-4-chloro-3-indolylphosphate p-toludine salt (BCIP) in Tris buffer (0.5mM $MgCl_2$,
599 100 mM diethanol-amine, pH 9.5). The arrays where scanned using a desktop scanner (CanoScan 8800
600 F, Canon) at 2400 dpi and the intensity of NBT/BCIP stained spots quantified using Array-Pro Analyzer
601 6.3 (Media Cybernetics). This produces data for the normalised intensity of spots relative to the
602 background intensity of the nitrocellulose membrane (to control for background staining). We used
603 this data to create a heatmap, for average spot intensity across biological replicates and sample
604 dilutions.
605
606 **Microbial Sampling**

607 Fecal samples were collected from 6 piglets per experimental group (n=12) at days 0, 7, 14,
608 21, and 27 post-weaning. At the end of the trial, all piglets (n=48) were sacrificed, and distinct samples
609 were collected from the lumen of the duodenum, jejunum, ileum, cecum, and colon. Samples were
610 obtained within the first 15 minutes after the piglets were sacrificed and the samples were flash-
611 frozen in liquid nitrogen and stored at -80°C until DNA extraction.

612

613 **DNA Extraction**

614 DNA was extracted with a MagAttract PowerMicrobiome DNA/RNA Kit (MO BIO Laboratories
615 Inc., Carlsbad, CA, USA) according to the manufacturer instructions, except for the bead beating step

616 where we used a FastPrep-96 Homogenizer (MP Biomedicals LLC., Santa Ana, CA, USA) at maximum
617 intensity for a total of 2 minutes in 4 pulses of 30 s with a 5 minute cooling period between each pulse.
618 A KingFisher Flex DNA extraction robot was used for the automated steps of the protocol. The
619 extracted nucleic acids were quantified with a Qubit Fluorimeter and the Qubit dsDNA BR Assay Kit
620 (Thermo Fisher Scientific, Waltham, MA, USA) and stored at -80°C.

621

622 **16S Amplicon Sequencing and Analysis**

623 16S amplicon sequence data was obtained for all fecal and intestinal samples. The V3-V4
624 region of the 16S rRNA gene was PCR amplified using the primers Pro341F (5'-CCT ACG GGN BGC ASC
625 AG-3') and Pro805R (5'-GAC TAC NVG GGT ATC TAA TCC-3'), to which the MiSeq adaptors were
626 additionally incorporated on the 5' ends[36]. The 25 µL PCR reactions consisted of 1X iProof High-Fidelity
627 Master Mix (Biorad, Hercules, CA, USA), 0.25 µM primers, and 5 ng template DNA. PCR thermal cycling
628 began with a hot start step at 98 °C for 180 s and was followed by 25 cycles of 98 °C denaturation for
629 30 s, 55 °C annealing for 30 s, and 72 °C extension for 30 s, followed by a final, 300 s extension step at
630 72 °C. Amplicons were individually purified with AMPure XP beads (Beckman Coulter, Indianapolis, IN,
631 USA) and indexed with the Nextera XT Index Kit v2 (Illumina, San Diego, CA, USA) according to the
632 Illumina protocol for 16S metagenomic sequencing library preparation. Next, equal volumes from each
633 indexing reaction were pooled together, and the pool was purified with AMPure XP beads. The purified
634 amplicon pool was then quantified with a Qubit Fluorimeter, diluted, mixed with 15% PhiX Control v3
635 (Illumina), and denatured according to the aforementioned Illumina protocol. The denatured library
636 was sequenced on the Illumina MiSeq platform using the MiSeq Reagent Kit v3 (600 cycle). Data were
637 output from the sequencer as demultiplexed FASTQ format files.

638 Processing of the data was done with a combination of standalone programs, QIIME[37]
639 MOTHUR[38] and the R package Phyloseq[39]. To process the data, the paired end reads for each sample
640 were merged with PEAR[40], specifying a minimum assembly length 400, maximum assembly length 575,

641 minimum overlap 50, and no statistical test. Then, PRINSEQ[41] version 0.20.4 was used to filter low
642 quality reads by requiring a minimum quality score of 10 for all bases and a minimum mean quality of
643 30. Primer sequences were trimmed in MOTHUR version 1.36.1, and chimeric sequences were
644 identified and filtered out using QIIME version 1.9.1. Next, open reference $OTU_{0.97}$ clustering[42] was
645 performed with VSEARCH[43] version 2.3.2 and the Silva database[44] release 128 as the taxonomy
646 reference. Then, the QIIME core diversity analyses script was run. Differentially abundant phylotypes
647 were identified in both cecum and colon for the control vs. 4% AcGGM samples using both the
648 MetagenomeSeq fitZIG and DESeq2 negative binomial algorithms via the QIIME wrapper. The OTU
649 table, phylogenetic tree, representative sequences, and taxonomy from QIIME were incorporated
650 along with the sample metadata into a Phyloseq version 1.22.3 object in R for data exploration and
651 visualization.

652

653 **Whole Metagenome Sequencing and Analysis**

654 Whole metagenome sequencing was performed at the Norwegian Sequencing Centre on 2
655 lanes of the Illumina HiSeq 4000 to generate 2 X 150 paired-end reads. TruSeq PCR-free libraries were
656 prepared for 12 control and 12 AcGGM (4%) samples from the colon. All 24 samples were run in both
657 lanes to eliminate the potential for lane-specific sequencing bias. FASTQ format files were received
658 from the sequencing center, and prior to assembly, these were quality filtered with Trimmomatic[45]
659 version 0.36 whereby TruSeq adaptor sequences were eliminated, sequences were required to have
660 an average quality score above 20, leading and trailing bases with quality below 20 were removed,
661 sequences with average quality score below 15 in a 4-base sliding window were trimmed, and the
662 minimum read length was required to be 36 bases. Individual sample assembly was accomplished with
663 metaSPAdes[46] version 3.11.1. MegaHIT[47] version 1.1.3 was used for co-assembly of all 24 samples
664 together as well as co-assembly of the 12 control samples together and the 12 4% AcGGM samples
665 together. MetaBAT[48] version 0.26.3 was used to bin the assemblies, and dRep[49] was used to

666 dereplicate the multiple assembly and binning combinations to produce an optimal set of MAGs.
667 MASH[50] version 2.0 used to compare the similarity of the 24 metagenomes by calculating pairwise
668 Jaccard distances which were imported into R for NMDS ordination and visualization. Completeness
669 and contamination was determined for each MAG using CheckM[51] version 1.0.7. Taxonomic
670 classifications of MAGs was performed using GTDB-Tk[52], and relative abundance estimations were
671 generated using CoverM (https://github.com/wwood/CoverM). Feature and functional annotation
672 were completed with the Prokka pipeline[53] version 1.12, and the predicted protein sequences from all
673 355 MAGs were concatenated to create the metaproteomics reference database. Resulting annotated
674 open reading frames (ORFs) were retrieved, further annotated for CAZymes using the CAZy annotation
675 pipeline with libraries from the July 2018 database release[54,55], and subsequently used as a reference
676 database for the metaproteomics (with the exception of glycosyltransferases).

677

678 **Metaproteomics**

679 Proteins were extracted from each sample by the following method. An aliquot (1 g) of colon digesta
680 from pigs fed either a control diet or a diet supplemented with 4% β-mannan was dissolved 1:1 (w/v)
681 in 50 mM TrisHCl, pH 8.4.

682 Lysis was performed using a bead-beating approach whereby glass beads (size ≤ 106 µm) were added
683 to the colon digesta slurry and cells were disrupted in 3 x 60 second cycles using a FastPrep24 (MP
684 Biomedicals, Santa Ana, CA, USA). Debris were removed by centrifugation at 16.600 x *g* for 20 minutes
685 and proteins were precipitated overnight in 16% ice-cold TCA. The next day, proteins were dissolved
686 in 100 µL 50 mM TrisHCl, pH 8.4 and concentration was determined using the Bradford protein assay
687 (Bradford Laboratories, USA) using bovine serum albumin as a standard. Fifty milligrams of protein
688 was prepared in SDS sample buffer, separated by SDS-PAGE using an Any-kD Mini-PROTEAN gel (Bio-
689 Rad Laboratories, Hercules, CA, USA) and stained using Coomassie Brilliant Blue R250. The gel was cut
690 into 6 slices and reduced, alkylated and digested as described previously[56]. Prior to mass spectrometry,

691 peptides were desalted using C_{18} ZipTips (Merck Millipore, Darmstadt, Germany) according to the
692 manufacturer's instructions.

693 The peptides were analyzed by nanoLC-MS/MS as described previously, using a Q-Exactive hybrid
694 quadrupole orbitrap mass spectrometer (Thermo Scientific, Bremen, Germany)[57], and the acquired
695 raw data was analyzed using MaxQuant[58] version 1.4.1.2. Proteins were quantified using the MaxLFQ
696 algorithm[59]. The data was searched against a sample-specific database (602.947 protein sequences),
697 generated from the 355 metagenome assembled genomes (MAGs), and against the genome of pig,
698 (*Sus scrofa domesticus)* (40.708 sequences). In addition, common contaminants such as human
699 keratins, trypsin and bovine serum albumin were concatenated to the database as well as reversed
700 sequences of all protein entries for estimation of false discovery rates. Protein N-terminal acetylation,
701 oxidation of methionine, conversion of glutamine to pyro glutamic acid, and deamination of
702 asparagine and glutamine were used as variable modifications, while carbamidomethylation of
703 cysteine residues was used as a fixed modification. Trypsin was used as digestion enzyme and two
704 missed cleavages were allowed. All identifications were filtered in order to achieve a protein false
705 discovery rate (FDR) of 1% using the target-decoy strategy. A total of 8515 protein groups (20 350
706 shared proteins) were considered valid, which required the protein group to be both identified and
707 quantified in at least two replicates, and in addition, we required at least one unique peptide per
708 protein and at least two peptides in total for every protein group. In cases where a protein group
709 consisted of two or more homologues protein identifications, both unique and shared proteins are
710 indicated (i.e. **Fig. 4c** and **Supplementary Dataset 6**). The output from MaxQuant was further explored
711 in Perseus version 1.6.0.7 where filtering, data transformation, and imputation were performed.

712 For metaproteomic enrichment analysis, missing values in the proteomics data were set to zero, and
713 proteins with zero variance across the eight samples, or with a non-zero value in only one replicate
714 animal in either the control or AcGGM treatment were removed. This resulted in 4562 unique protein
715 groups assigned to 12 535 MAG ORFs (i.e. shared detected proteins). A Principle Component Analysis

716 (PCA) was performed on the 4562x8 matrix using the R function *prcomp*. Hierarchal clustering
717 dendrograms were constructed using the R functions *hclust* with Pearson correlation and *method =*
718 *"ward.D"*, and a heatmap was drawn using the function *heatmap.2* with *scale = "row"*. The number of
719 protein clusters were selected by visual inspection and the clusters were selected using the *cutree*
720 function. Enzyme commission (EC) annotation and MAG enrichment analysis was performed using the
721 hypergeometric distribution function *phyper* in R.

722

723 **Genome tree**

724 Phylogenetic analysis was performed using a block of 22 universal ribosomal proteins (30S ribosomal
725 protein L1, L2, L4-L6, L10, L11, L14, L15,L18 and 50S ribosomal protein S3, S5, S7-S13, S15, S17,
726 S19)[60,61]. In addition to the MAGs, we recruited 239 reference genomes for phylogenetic resolution.
727 These genomes were selected based on preliminary examination of the assembled metagenome using
728 metaQUAST[62]. The reference genomes were annotated using the Prokka pipeline in the same manner
729 as for the MAGs. All identified ribosomal protein sequences were aligned separately with MUSCLE
730 v3.8.31[63], and manually checked for duplications and misaligned sequences. Divergent regions and
731 poorly aligned positions were further eliminated using GBlocks[64], and the refined alignment were
732 concatenated using catfasta2phyml.pl (https://github.com/nylander/catfasta2phyml) with the
733 parameter '-c' to replace missing ribosomal proteins with gaps (-). The maximum likelihood-based
734 phylogeny of the concatenated ribosomal proteins was inferred using RAxML version 8.2.12[65]
735 (raxmlHPC-SSE3 under PROTGAMMA distributed model with WAG substitution matrix) and support
736 values determined using 100 bootstrap replicates. The tree was rooted to the Euryarchaeota phylum
737 and visualized using iTOL[66]. Clades of reference genomes with only distant phylogenetic relation to
738 the MAGs were collapsed to refine the final tree in **Fig. 3**. The complete tree is available in Newick
739 format as Supplementary Dataset 1.

740

741 **REFERENCES**

742 1 Patnode, M. L. *et al.* Interspecies Competition Impacts Targeted Manipulation of Human Gut
743 Bacteria by Fiber-Derived Glycans. *Cell.* **179**, 59-73 (2019).

744 2 Blanton, L. V., Barratt, M. J., Charbonneau, M. R., Ahmed, T. & Gordon, J. I. Childhood
745 undernutrition, the gut microbiota, and microbiota-directed therapeutics. *Science* **352**, 1533
746 (2016).

747 3 Gibson, G. R. *et al.* Expert consensus document: The International Scientific Association for
748 Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of
749 prebiotics. *Nat. Rev. Gastroenterol. Hepatol.* **14**, 491-502 (2017).

750 4 Sanders, M. E., Merenstein, D. J., Reid, G., Gibson, G. R. & Rastall, R. A. Probiotics and
751 prebiotics in intestinal health and disease: from biology to the clinic. *Nat. Rev. Gastroenterol.*
752 *Hepatol.* **doi: 10.1038/s41575-019-0173-3** (2019).

753 5 La Rosa, S. L. *et al.* The human gut Firmicute *Roseburia intestinalis* is a primary degrader of
754 dietary β-mannans. *Nat. Commun.* **10**, 905 (2019).

755 6 Larsbrink, J. *et al.* A discrete genetic locus confers select *Bacteroidetes* with a niche role in
756 xyloglucan metabolism in the human gut. *Submitted* (2013).

757 7 Cuskin, F. *et al.* Human gut Bacteroidetes can utilize yeast mannan through a selfish
758 mechanism. *Nature* **517**, 165-169 (2015).

759 8 Solden, L. M. *et al.* Interspecies cross-feeding orchestrates carbon degradation in the rumen
760 ecosystem. *Nat. Microbiol.* **3**, 1274-1284 (2018).

761 9 Michalak, L. *et al.* A pair of esterases from a commensal gut bacterium remove acetylations
762 from all positions on complex β-mannans. *Proc. Natl. Acad. Sci. U S A* **117**, 7122-7130 (2020).

763 10 La Rosa, S. L. *et al.* Wood-Derived Dietary Fibers Promote Beneficial Human Gut Microbiota.
764 *mSphere* **4**, e00554-00518 (2019).

765 11 Ferreira-Halder, C. V., Faria, A. V. S. & Andrade, S. S. Action and function of *Faecalibacterium*
766 *prausnitzii* in health and disease. *Best Pract. Res. Clin. Gastroenterol.* **31**, 643-648 (2017).

767 12 Lopez-Siles, M., Duncan, S. H., Garcia-Gil, L. J. & Martinez-Medina, M. *Faecalibacterium*
768 *prausnitzii*: from microbiology to diagnostics and prognostics. *ISME J.* **11**, 841-852 (2017).
769 13 Lahaye, M., Quemener, B., Causse, M. & Seymour, G. B. Hemicellulose fine structure is
770 affected differently during ripening of tomato lines with contrasted texture. *Int. J. Biol.*
771 *Macromol.* **51**, 462–470 (2012).
772 14 Xu, C. *et al.* Acetylation and characterization of spruce (Picea abies) galactoglucomannans.
773 *Carbohydrate research* **345**, 810-816 (2010).
774 15 Michalak, L., Knutsen, S. H., Aarum, I. & Westereng, B. Effects of pH on steam explosion
775 extraction of acetylated galactoglucomannan from Norway spruce. *Biotechnol. Biofuels* **11**,
776 311 (2018).
777 16 Shepherd, E. S., DeLoache, W. C., Pruss, K. M., Whitaker, W. R. & Sonnenburg, J. L. An exclusive
778 metabolic niche enables strain engraftment in the gut microbiota. *Nature* **557**, 434–438
779 (2018).
780 17 Chen, L. *et al.* The Maturing Development of Gut Microbiota in Commercial Piglets during the
781 Weaning Transition. *Front. Microbiol.* **8**, 1688 (2017).
782 18 Holman, D. B., Brunelle, B. W., Trachsel, J. & Allen, H. K. Meta-analysis To Define a Core
783 Microbiota in the Swine Gut. *MSystems* **2**, e00004.
784 19 Willför, S. *et al.* Characterisation of water-soluble galactoglucomannans from Norway spruce
785 wood and thermomechanical pulp. *Carbohydr. Polym.* **52**, 175-187 (2003).
786 20 Rattigan, R. *et al.* Laminarin-rich Extract Improves Growth Performance, Small Intestinal
787 Morphology, Gene Expression of Nutrient Transporters and the Large Intestinal Microbial
788 Composition of Piglets During the Critical Post-Weaning Period. *Br. J. Nutr.* **123**, 255-263
789 (2020).
790 21 Parada Venegas, D. *et al.* Short Chain Fatty Acids (SCFAs)-Mediated Gut Epithelial and Immune
791 Regulation and Its Relevance for Inflammatory Bowel Diseases. *Front Immunol.* **10**, 277 (2019).

792 22 Precup, G. & Vodnar, D. C. Gut *Prevotella* as a possible biomarker of diet and its eubiotic versus
793 dysbiotic roles: a comprehensive literature review. *Br. J. Nutr.* **122**, 131-140 (2019).
794 23 Kageyama, A. & Benno, Y. *Catenibacterium mitsuokai* gen. nov., sp. nov., a gram-positive
795 anaerobic bacterium isolated from human faeces. *Int. J Syst. Bacteriol.* **50**, 1595-1599 (2000).
796 24 Moore, L. V. & Moore, W. E. *Oribaculum catoniae* gen. nov., sp. nov.; *Catonella morbi* gen.
797 nov., sp. nov.; *Hallella seregens* gen. nov., sp. nov.; *Johnsonella ignava* gen. nov., sp. nov.; and
798 *Dialister pneumosintes* gen. nov., comb. nov., nom. rev., Anaerobic gram-negative bacilli from
799 the human gingival crevice. *Int J Syst Bacteriol.* **44**, 187-192 (1994).
800 25 Louis, P., Young, P., Holtrop, G. & Flint, H. J. Diversity of human colonic butyrate-producing
801 bacteria revealed by analysis of the butyryl-CoA:acetate CoA-transferase gene. *Environ.*
802 *Microbiol.* **12**, 304–314 (2010).
803 26 Martens, E. C., Koropatkin, N. M., Smith, T. J. & Gordon, J. I. Complex glycan catabolism by the
804 human gut microbiota: the Bacteroidetes Sus-like paradigm. *J. Biol. Chem.* **284**, 24673-24677
805 (2009).
806 27 Panwar, D. & Kapoor, M. Transcriptional analysis of galactomannooligosaccharides utilization
807 by *Lactobacillus plantarum* WCFS1. *Food Microbiol.* **86**, 103336 (2020).
808 28 Sadaie, Y., Nakadate, H., Fukui, R., Yee, L. M. & Asai, K. Glucomannan utilization operon of
809 *Bacillus subtilis*. *FEMS Microbiol. Lett.* **279**, 103-109 (2008).
810 29 Leth, M. L. *et al.* Differential bacterial capture and transport preferences facilitate co-growth
811 on dietary xylan in the human gut. *Nat. Micro.* **3**, 570-580 (2018).
812 30 Vizcaino, J. A. *et al.* The PRoteomics IDEntifications (PRIDE) database and associated tools:
813 status in 2013. *Nucleic Acids Res.* **41**, D1063-1069 (2013).
814 31 Chum, H. L., Johnson, D. K., Black, S. K. & Overend, R. P. Pretreatment Catalyst Effects and the
815 Combined Severity Parameter. *Appl. Biochem. Biotech.* **24**, 1-14 (1990).
816 32 Vestby, L. K., Møretrø, T., Ballance, S., Langsrud, S. & Nesse, L. L. Survival potential of wild type
817 cellulose deficient Salmonella from the feed industry. *BMC Veterinary Research* **5**, 43 (2009).

818 33 Pedersen, K. S. & Toft, N. Intra- and inter-observer agreement when using a descriptive
819 classification scale for clinical assessment of faecal consistency in growing pigs. *Prev Vet Med*
820 **98**, 288-291 (2011).
821 34 Moller, I. *et al.* High-throughput mapping of cell-wall polymers within and between plants
822 using novel microarrays. *Plant J.* **50**, 1118-1128 (2007).
823 35 Pedersen, H. L. *et al.* Versatile high resolution oligosaccharide microarrays for plant
824 glycobiology and cell wall research. *J. Biol. Chem.* **287**, 39429-39438 (2012).
825 36 Takahashi, S., Tomita, J., Nishioka, K., Hisada, T. & Nishijima, M. Development of a prokaryotic
826 universal primer for simultaneous analysis of Bacteria and Archaea using next-generation
827 sequencing. *PloS one* **9**, e105592 (2014).
828 37 Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data.
829 *Nature methods* **7**, 335-336 (2010).
830 38 Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-
831 supported software for describing and comparing microbial communities. *Appl Environ*
832 *Microbiol* **75**, 7537-7541 (2009).
833 39 McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and
834 graphics of microbiome census data. *PloS one* **8**, e61217 (2013).
835 40 Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End
836 reAd mergeR. *Bioinformatics* **30**, 614-620 (2014).
837 41 Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets.
838 *Bioinformatics* **27**, 863-864 (2011).
839 42 Rideout, J. R. *et al.* Subsampled open-reference clustering creates consistent, comprehensive
840 OTU definitions and scales to billions of sequences. *PeerJ* **2**, e545 (2014).
841 43 Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahe, F. VSEARCH: a versatile open source tool
842 for metagenomics. *PeerJ* **4**, e2584 (2016).

843 44 Yilmaz, P. *et al.* The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks.
844 *Nucleic acids research* **42**, D643-648 (2014).
845 45 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence
846 data. *Bioinformatics* **30**, 2114-2120 (2014).
847 46 Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile
848 metagenomic assembler. *Genome research* **27**, 824-834 (2017).
849 47 Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. MEGAHIT: an ultra-fast single-node solution
850 for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**,
851 1674-1676 (2015).
852 48 Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately
853 reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165 (2015).
854 49 Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic
855 comparisons that enables improved genome recovery from metagenomes through de-
856 replication. *Isme j* **11**, 2864-2868 (2017).
857 50 Ondov, B. D. *et al.* Mash: fast genome and metagenome distance estimation using MinHash.
858 *Genome Biology* **17**, 132 (2016).
859 51 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing
860 the quality of microbial genomes recovered from isolates, single cells, and metagenomes.
861 *Genome research* **25**, 1043-1055 (2015).
862 52 Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify
863 genomes with the Genome Taxonomy Database. *Bioinformatics.* **doi:**
864 **10.1093/bioinformatics/btz848** (2019).
865 53 Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-2069
866 (2014).
867 54 Terrapon, N., Lombard, V., Gilbert, H. J. & Henrissat, B. Automatic Prediction of Polysaccharide
868 Utilization Loci in Bacteroidetes Species. *Bioinformatics* **31**, 647-655 (2014).

869 55 Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The
870 carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490-495
871 (2014).
872 56 Arntzen, M. O., Karlskas, I. L., Skaugen, M., Eijsink, V. G. & G., M. Proteomic investigation of
873 the response of *Enterococcus faecalis* V583 when cultivated in urine. *PLoS One* **10**, e0126694
874 (2015).
875 57 Arntzen, M. O., Karlskas, I. L., Skaugen, M., Eijsink, V. G. & Mathiesen, G. Proteomic
876 Investigation of the Response of Enterococcus faecalis V583 when Cultivated in Urine. *PloS*
877 *one* **10**, e0126694 (2015).
878 58 Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-
879 range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* **26**,
880 1367-1372 (2008).
881 59 Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and
882 maximal peptide ratio extraction, termed MaxLFQ. *Molecular & cellular proteomics : MCP* **13**,
883 2513-2526 (2014).
884 60 Hug, L. A. *et al.* A new view of the tree of life. *Nat. Microbiol.* **1**, 16048 (2016).
885 61 Wu, D., Jospin, G. & Eisen, J. A. Systematic identification of gene families for use as “markers”
886 for phylogenetic and phylogeny-driven ecological studies of bacteria and archaea and their
887 major subgroups. *PloS One* **8**, e77033 (2013).
888 62 Mikheenko, A., Saveliev, V. & Gurevich, A. MetaQUAST: evaluation of metagenome
889 assemblies,. *Bioinformatics* **32**, 1088-1090 (2016).
890 63 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput.
891 *Nucleic acids research* **32**, 1792-1797 (2004).
892 64 Talavera, G. & Castresana, J. Improvement of Phylogenies after Removing Divergent and
893 Ambiguously Aligned Blocks from Protein Sequence Alignments. *Syst. Biol.* **56**, 564–577
894 (2007).

895 65 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
896 phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).
897 66 Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and
898 annotation of phylogenetic and other trees. *Nucleic acids research* **44**, W242-W245 (2016).

899