Linking carbohydrate structure with function in the human gut microbiome using hybrid metagenome assemblies

Anuradha Ravi[1,2*], Perla Troncoso-Rey[1*], Jennifer Ahn-Jarvis[1*], Kendall R. Corbin[1,3], Suzanne Harris[1,4], Hannah Harris[1], Alp Aydin[1], Gemma L. Kay[1], Thanh Le Viet[1], Rachel Gilroy[1], Mark J. Pallen[1], Andrew J. Page[1], Justin O'Grady[1,5,*], Frederick J. Warren[1,*,+]

[1]Quadram Institute Bioscience, Norwich Research Park, Norwich, NR4 7UQ, UK.

[2]Current address: Gemini centre for Sepsis Research, Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, Trondheim, Norway

[3]Department of Horticulture, University of Kentucky, Lexington, Kentucky, USA.

[4]Current address: The Francis Crick Institute, 1 Midland Road, London, NW1 1AT, UK

[5]University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK.

*Contributed equally

+Corresponding author: fred.warren@quadram.ac.uk

Keywords: Nanopore, Sequencing, CAZymes, colonic fermentation, metagenomics, starch

Abstract [200 words]

Complex carbohydrates that escape digestion in the small intestine, are broken down in the large intestine by enzymes encoded by the gut microbiome. This is a symbiotic relationship between particular microbes and the host, resulting in metabolic products that influence host gut health and are exploited by other microbes. However, the role of carbohydrate structure in directing microbiota community composition and the succession of carbohydrate-degrading microbes, is not fully understood. In this study we evaluate species-level compositional variation within a single microbiome in response to six structurally distinct carbohydrates in a controlled model gut using hybrid metagenome assemblies. We identified 509 high-quality metagenome-assembled genomes (MAGs) belonging to ten bacterial classes and 28 bacterial families. We found dynamic variations in the microbiome amongst carbohydrate treatments, and over time. Using these data, the MAGs were characterised as primary (0h to 6h) and secondary degraders (12h to 24h). Recent advances in sequencing technology allowed us to identify significant unexplored diversity amongst starch degrading species in the human gut microbiota including CAZyme profiles for novel MAGs.

Microbial diversity within the microbiome and its interactions with host health and nutrition are now widely studied[1]. An important role of the human gut microbiome is the metabolic breakdown of complex carbohydrates derived from plants and animals (e.g. legumes, seeds, tissue and cartilage)[2]. Short chain fatty acids (SCFA) are the main products of carbohydrate fermentation by gut microbiota and provide a myriad of health benefits through their systemic effects on host metabolism.[3,4] However, we still do not have a complete picture of the range of microbial species involved in fermentation of complex carbohydrates to produce SCFA. Understanding the intricacies of complex carbohydrate metabolism by the gut microbiota is a significant challenge. The function of many 'hard to culture species' remains obscure and while advances in sequencing technology are beginning to reveal the true diversity of the human gut microbiota, there is still much to be learned.[5]

A key challenge is understanding the influence of structural complexity of carbohydrates on microbiota composition. Carbohydrates possess immense structural diversity, both at the chemical composition level (monomer and sugar linkage composition) and at the mesoscale. Individual species, or groups of species, within the gut microbiota are highly adapted to defined carbohydrate structures[6]. Starch is representative of the structural diversity found amongst carbohydrates and serves as a good model system as starches are readily fermented by several different species of colonic bacteria.[7] The gut microbiota is repeatedly presented with starches of diverse structures from the diet.[8] Consistent in starch is an α-1→4 linked glucose back bone, interspersed with α-1→6 linked branch points. Despite this apparent structural simplicity, starches botanical origin and subsequent processing (e.g. cooking) impacts its physicochemical properties, particularly crystallinity and recalcitrance to digestion.[7] It has been shown *in vitro*[7], in animal models[9]

and in human interventions[8], that altering starch structure can have a profound impact on gut microbiome composition.

The microbiome is known to harbour a huge repertoire of carbohydrate-active enzymes (CAZymes) that can degrade diverse carbohydrate structures.[10,11] However, it is a formidable challenge to study this functionality in complex microbial communities due to limitations in the depth of sequencing and coverage of all members in the community. While metagenomic sequencing has become a key tool, identifying genomes and functional pathways within the microbiome remains challenging in second generation sequencing due to limitations associated with short (~300bp) reads. Third generation sequencing such as nanopore sequencing (Oxford Nanopore Technologies (ONT)) promises to circumvent these difficulties by providing longer reads (> 3 kilobase pairs [kbp]). This technology has become popular in clinical metagenomics for rapid pathogen diagnosis[12] and in human genomics research.[13] Long-read sequences can help bridge inter-genomic repeats and produce better *de novo* assembled genomes.[14] While the MinION platform from ONT has been used for metagenomic studies,[15] it cannot provide sufficient sequencing depth and coverage to sequence the many hundreds of genomes present in the human gut microbiome. PromethION (ONT) is capable of producing far greater numbers of sequences compared to either MinION or GridION, averaging four-five times more data per flow cell and the capacity to run up to 48 flow cells in parallel; this makes it suitable for metagenomics and microbiome studies. For example, PromethION has been used for long-read sequencing of environmental samples such as wastewater sludge, demonstrating its potential to recover large numbers of metagenome-assembled genomes (MAGs) from diverse mixtures of microbial species.[16] However, long error-prone reads aren't ideal for species resolution

metagenomics, therefore, a hybrid approach using short and long read data has been found to be most effective for generating accurate MAGs.[17]

To achieve species-level resolution of the microbes present in the gut microbiome during complex carbohydrate utilisation, we conducted a genome-resolved metagenomics study in a controlled gut colon model. *In vitro* fermentation systems have been used extensively to model changes in the gut microbial community as a result of external inputs, e.g., changes in pH, protein and carbohydrate supply[7,18,19]. We measured the dynamic changes in bacterial populations during fermentation of six structurally contrasting substrates: two highly recalcitrant starches (native Hylon VII ("Hylon") and native potato starch ("potato")); two accessible starches (native normal maize starch ("n.maize") and gelatinized then retrograded maize starch ("r.maize"); an insoluble fibre (cellulose) resistant to fermentation ("Avicel"); and a highly fermentable soluble fibre ("inulin"). By generating hybrid assemblies using PromethION and NovaSeq data, we obtained 509 MAGs. The dereplicated set consisted of 151 genomes belonging to ten bacterial classes and 28 bacterial families. Using genome-level information and read proportions data, we identified several species that have novel putative starch-degrading properties.

Results

PromethION and NovaSeq sequencing of model gut samples enriched for carbohydrate degrading species. Fermentation of six contrasting carbohydrate substrates (inulin, Hylon, n.maize, potato, r.maize and Avicel; see methods section) was initiated by inoculation of the model colon with a carbohydrate and faecal material and the gut microbial community composition was monitored over time (0h, 6h, 12h and 24h) by sequencing as shown in

Figure 1. In total, 23 samples and a negative control were sequenced (see Supplementary Table 1 for the PromethION and NovaSeq summary sequencing statistics).

PromethION sequencing: The two sequencing runs generated 144 giga base pairs (Gbp) of raw sequences. In the first run, all 23 samples were analysed while in the second batch, 12 samples from hylon, inulin and r.maize were selected. The first run produced 7.87 million reads with an average read length of 3419 ± 57 bp and the second run generated 21.6 million reads with an average read length of 4707 ± 206 bp . Consolidating the runs, trimming and quality filtering resulted in the removal of 33.3 ± 14.7 % of reads (Supplementary Table 1). Median read lengths after trimming were 4972.5 ± 229 bp and the median quality score was 9.7 ± 0.9.

Illumina sequencing: All 23 samples provided high quality sequences (Q value > 30) generating a mean of 27 million reads per sample. Quality and read length (<60 bp) filtering removed 2.96 % of reads (Supplementary Table 1).

Dynamic shifts in taxonomic profiles among carbohydrate treatments. Hierarchical clustering for the taxonomic profiling using MetaPhlAn3 for each sample is shown in Supplementary Table 2. At baseline (0h), profiles of the top 30 selected species by clustering (using Bray-Curtis distances for samples and species, and a complete linkage) is similar for all treatments, as expected (**Error! Reference source not found.**). This uniform profile was distinct from the water control sample (a.k.a. 'the kitome'). The water blank also had less than 3% (NovoSeq) and less than 0.2% (PromethION) of the reads of the samples. Microbiome shifts were apparent from 6h in the n.maize treatment which showed a very high abundance of *E. coli,* indicating contamination. After 12h, the profiles changed further

129 with a higher abundance of *E. coli* and *B. animalis* in the n.maize treatment while the
130 r.maize and inulin treatment profiles were similar, as were the potato and Hylon treatment
131 profiles. By the last sampling point (24h), potato and hylon had similar profiles which are
132 also similar to r.maize. The most abundant species in all the substrates was consistently
133 *Prevotella copri* which decreased in abundance over time but remained one of the most
134 abundant species throughout. After 6h and 12h, *Ruminococcus bromii* (a keystone starch
135 degrader) and *Bifidobacterium adolescentis* increased in abundance in the r.maize, potato
136 and Hylon treatments. *Faecalibacterium praunitzii* decreased in abundance in inulin at 6h
137 and 12h and then increased in abundance for inulin and avicel at 24h.
138 Dynamic shifts in the microbiome were estimated using PCoA (Supplementary Figure
139 1), with 77% of total variance being explained by the first two components. As expected, the
140 0h profiles clustered closely together. The most distinct taxonomic change in microbial
141 community composition was apparent in the Avicel treatment after 24h. Inulin and r.maize
142 profiles clustered more closely together than potato and Hylon profiles. Inverse Simpson
143 index results followed a similar pattern for changes in diversity, which decreased after 0h
144 followed by a gradual increase (Supplementary Figure 2). However, in the Avicel treatment
145 there was a different pattern of taxonomic shifts with a large number of taxa increasing in
146 abundance after 12h.
147
148 **Hybrid metagenome assemblies vs short-read only assemblies.** Using Opera-MS, we
149 combined PromethION reads with Illumina assemblies to produce hybrid assemblies. The
150 assembly statistics for short-read-only and hybrid assemblies are shown in Supplementary
151 Table 3 and **Error! Reference source not found.**. The longest N50 and the largest contig per

152 treatment were generated using hybrid assemblies as expected (figure 3b & 3c). The overall
153 length of assembled sequences was similar for both approaches (Figure 3d).
154 The reads from each treatment and collective T0 were co-assembled into hybrid
155 assemblies and binned into MAGs. In total we binned and refined 509 MAGs that met the
156 MIMAG quality score criteria[20] of which 65% (n=333) were high-quality (Figure 4;
157 Supplementary table 4). From the co-assemblies, thirty-five MAGs had an N50 of > 500,000
158 Mbp and 158 MAGs were assembled into < 30 scaffolds. The MAGs were dereplicated into
159 primary and secondary clusters according to Average Nucleotide identity (ANI) (primary
160 clusters <97%; secondar clusters <99%). In total, we identified 151 MAG secondary clusters
161 (Supplementary table 5). Each genome cluster consisted of between one and seven
162 genomes based on their genome similarity.
163
164 **Taxonomic annotation of MAGs.** Proposed bacterial taxonomy using GTDb was represented
165 in existing bacterial families: All MAG clusters had > 99% identity to existing genera
166 (Supplementary Table 6). By directly comparing the MAG assembly statistics for the MAGs in
167 the present study to the representative assemblies in GTDb (Supplementary Table 8) we
168 found that while the average overall assembly length was almost similar (an average of
169 2,250,870 bp in the present study vs. 2,541,312bp in GTDb), there were far fewer contigs in
170 our assemblies (an average of 67 contigs in the present study vs. 160 in GTDb), and
171 therefore may be considered to be of higher quality. Therefore, using the approach
172 described in Pallen *et al*[21], Supplementary Table 7 provides proposed taxonomy for the 70
173 species which do not currently have Latin binomial names.
174

Carbohydrate structure drives progression of bacterial diversity. Relative abundance of each MAG within treatments was calculated and log fold change of abundance between treatments was used to estimate change in relative abundance (Supplementary table 9). In total, 36 of 151 clusters exhibited ≥2-log fold increase in relative abundance for all treatments. Specifically, ≥2-log fold change in abundance was seen in 6, 12, 11 and 18 MAGs for Avicel, Hylon, potato and r.maize treatments, respectively (Figure 5). The genomes were partitioned as early (0h up to 6h) and late degraders (12h to 24h) according to when they first showed an increase in relative abundance (Supplementary Table 10).

Relative abundance of all MAGs from each treatment was aggregated and plotted for each time period (Supplementary figure 3). Relative abundance was constant for Avicel throughout indicating low activity of the MAGs in utilising crystalline cellulose, likely reflecting the very limited fermentability of microcrystalline cellulose. As for other maize starches (hylon, r.maize and potato), the read proportions showed an overall reduction in abundance, with only starch degrading MAGs increasing in abundance.

CAZyme family interplay with the carbohydrate treatments. For identifying CAZymes in the MAGs, genome-predicted proteins identified by Prodigal were compared with the CAZy database using dbCAN2 (Supplementary table 11). CAZyme counts specifically for Glycoside hydrolases (GH) and Carbohydrate binding modules (CBM) for all clusters showed a high representation of the profiles with GH13, GH2 and GH3 accounting for 34.1% of all counts (Supplementary Figure 4). CAZyme profiles for MAGs with > 2-log fold change are highlighted in Supplementary table 12 and Figure 6. Although six genomes were identified as associated with the degradation of cellulose, none contained any characteristic cellulose active CAZy proteins indicating multiple cross feeders. *Collinsella aerofaciens*_J (cluster

29_1), *Candidatus Minthovivens enterohominis* (cluster 81_1) are novel genomes that showed a 2x log -fold increase when in the presence of inulin and also harboured multiple copies of inulinases (GH32). *Bacteroides uniformis*, a known inulin degrader also contained multiple copies of GH32. We identified a large representation of the amylolytic (starch degrading) gene family GH13 in Hylon (counts= 88), potato (counts=50) and r.maize (counts=77) treatments. As expected, GH13 was weakly represented in Avicel (counts=19) and inulin (counts=29) treatments (Figure 6). The presence of GH13 in MAGs was closely associated with CBM48, which is commonly appended to starch degrading GH13 enzymes.[22] In total, we identified several novel degraders and previously discovered degraders of the different carbohydrate treatments which are highlighted in Supplementary table 10.

Discussion

Using a hybrid assembly approach (i.e., combining NovaSeq short-read and PromethION long-read metagenomic data), we report species-level resolved taxonomic data identifying distinct changes in microbiome composition in response to different substrates. The large number of high-quality near-complete MAGs that we generated using this approach enabled us to functionally annotate the CAZymes in the MAGs and identify potential carbohydrate degrading species. Several of these species have not previously been identified as playing a role in starch fermentation (Figure 6 and Supplementary Table 9).

High quality DNA for long-read sequencing was extracted using a bead beating protocol

The N50 for the PromethION reads was 4,972 bp, which is comparable with another recent study using bead-beating-based DNA extraction and provided adequate read lengths to be

223 useful for assembly of MAGs.[17] A recent publication by Moss *et al.*[23] and associated protocol
224 paper[24] suggested that bead beating DNA extraction protocols were unsuitable for long-
225 read sequencing as they led to excessive shearing of DNA and therefore enzymatic cell lysis
226 followed by phenol-chloroform purification were preferred to recover high molecular
227 weight (HMW) DNA. This was not reflected in our experience. The N50's obtained by Moss
228 *et al*. for sequencing DNA extracted from stool samples by phenol-chloroform on the
229 PromethION platform ranged from 1,432 bp to 5,205 bp, which on average was shorter than
230 the N50 we obtained using comparable samples extracted by a bead beating protocol. This
231 is in agreement with Bertrand et al.[14] who directly compared commercial bead beating and
232 phenol-chloroform extraction protocols for extracting HMW DNA from stool samples for
233 MinION sequencing and found that while phenol-chloroform gave higher molecular weights
234 of DNA, the DNA was of low integrity compromising sequencing quality.
235
236 **Hybrid assemblies allow generation of near complete MAGs and identification of novel**
237 **microbial species.** We found larger N50s and longest contigs when using hybrid assemblies
238 compared with short-read assemblies; this is in agreement with previous benchmarking
239 data using a combined MinION and Illumina hybrid approach to sequence mock
240 communities, human gut samples,[14] and rumen gut microbiota samples.[17] This allowed us to
241 assemble 509 MAGs across all the major phylogenetic groups (Supplementary file 5), with
242 representatives from ten bacterial classes and 28 families, including both Gram-positive and
243 Gram-negative species. Bertand et al.[14] found that phenol-chloroform extractions led to
244 underrepresentation of 'hard to sequence' gram-positive species such as those of the genus
245 *Bifidobacterium*. In the present study near-complete MAG's were recovered from 5
246 different species of *Bifidobacterium*, in contrast to Moss *et al.*[23] who were unable to

247 recover *Bifidobacterium* MAG's from the PromethION data produced using their enzyme
248 and phenol-chloroform based extraction method (although they were able to recover
249 *Bifidobacterium* MAG's from short-read data which was obtained following a bead beating
250 based DNA extraction of the same samples). This indicates that bead beating is necessary to
251 obtain accurate representations of the microbial community in human stool samples. The
252 bead beating DNA extraction protocol used in this study was also recommended by the
253 Human Microbiome Project to avoid biases in microbiome samples.[25,26]
254 We have provided *Candidatus* names to 70 bacterial species which do not currently have
255 representative Latin binomial names in the GTDB database (Supplementary Table 7). Our
256 decision to provide names for these species reflects the higher quality of MAGs compared
257 to those currently represented in the databases (Supplementary Table 8).
258
259 **Structural diversity in substrates drives changes in microbial communities.** Over the 24h
260 fermentation, microbial communities rapidly diverged depending on substrate. The smallest
261 change in community composition occurred in the Avicel treatment, as would be expected,
262 given that Avicel was the most recalcitrant substrate evaluated, with very limited
263 fermentability.[27] Each substrate resulted in distinct changes in microbial community
264 composition, supporting previous findings that chemically-identical but structurally-diverse
265 starches can result in distinct changes in microbial community composition.[7,8]
266
267 **Changes in microbial composition are related to the ability to degrade structurally diverse**
268 **substrates.** To better understand potential mechanisms driving the changes in microbial
269 species composition in response to different substrates, we explored the CAZyme profiles of
270 our microbial community .[28] We found the greatest number and diversity of CAZyme genes

were in the genomes of *Bacteroidetes* (Figure 6 and Supplementary Figure 4), as has previously been computationally estimated for the human gut microbiome.[10,29] This is in contrast to rumen microbiomes where *Fibrobactares* are the primary fibre-degrading bacterial group.[17]

We identified genomes that increased in abundance during either early or late stages of fermentation suggesting that their involvement in substrate degradation was either as primary (early) or secondary (late) degraders (Figure 5). We also identified differences in abundance of particular CAZyme-encoding genes amongst species which may reflect their specialisation to specific substrates (Figure 6). *Bacteroides uniformis* has been characterised as an inulin-degrading species,[30] and in our analysis it was identified during inulin fermentation and had three copies of the GH32 (inulinase) gene and a gene encoding the inulin binding domain, CBM38. *Candidatus Minthovivens enterohominis* also increased in abundance early in inulin degradation, and its genome contained five copies of the GH32 gene. *Faecalibacterium prausnitzii* increased in abundance with inulin supplementation and has been shown to have the ability to degrade inulin when co-cultured with primary degrading species.[31,32] *F. prausnitzii* was also found to increase in abundance for cellulose, but not for the starch based substrates.

Avicel is a highly crystalline cellulose that is resistant to fermentation; the human gut microbiota has a very limited capacity to degrade celluloses.[33] Interestingly, the largest increase in abundance we observed was for *Blautia hydrogenotrophica*; which has been reported in association with cellulose fermentation since it acts as an acetogen using hydrogen produced by primary degraders of cellulose.[34]

In all starch treatments, there were large increases in the proportion of identified genes that encoded GH13 (the major amylolytic gene family including α-amylase, α-

295 glucosidase and pullulanase) reflecting selection for starch-degrading species (Figure 6); this
296 was also the case for CBM48 which is also involved in starch degradation (Figure 6).[22] Our
297 analysis identified several well-known starch degrading species, most notably *R. bromii and*
298 *B. adoloscentis* (Figure 5). *R. bromii* is a well characterised specialist on highly recalcitrant
299 starch,[35] possessing specialised starch-degrading machinery termed the 'amylosome'; it was
300 only identified in the most recalcitrant starch treatments (Hylon and potato). Previous
301 genome sequencing of an *R. bromii* isolate reported 15 GH13 genes;[35] 14 GH13 genes were
302 identified in the *R. bromii* MAG assembled in this study. In the potato treatment another
303 closely related but less well characterised *Rumminococcus* species with ten GH13 genes and
304 one CBM48 gene was identified.
305 A previously uncultured *Blautia* species was identified possessing eight GH13 and
306 three CBM48 genes which increased in abundance in response to Hylon and potato. *Blautia*
307 species have previously been shown to increase in abundance in response to resistant
308 starch.[36,37] We also identified four further previously-uncharacterised species that increased
309 in abundance and had more than five GH13 genes: *Candidatus Cholicenecus caccae*,
310 *Candidatus Eisenbergiella faecalis*, *Candidatus Enteromorpha quadrami* and *Candidatus*
311 *Aphodonaster merdae*.
312 Maize starch treatments (r.maize and Hylon) showed increases in abundance of
313 *Bifidobacterium* species. Previous studies have characterized *Bifidobacterium* as a starch-
314 degrading genus.[38] The only *Bifidobacterium* species to increase in abundance in response
315 to Hylon was *B. adolescentis*, which is known to utilise to this hard-to-digest starch better
316 than other *Bifidobacterium* species,[39]; a broader range of *Bifidobacterium* species increased
317 in abundance in response to the more accessible r.maize.
318

Conclusion

We have demonstrated that deep long- and short-read metagenomic sequencing and hybrid assembly has great potential for studying the human gut microbiota. We identified species-level resolved changes in microbial community composition and diversity in response to carbohydrates with different structures over time, identifying succession of species within the fermenter. To provide functional information about these species we obtained over 500 MAGs from a single human stool sample. Annotating CAZyme genes in MAGs from species enriched for by fermentation of different carbohydrates allowed us to identify species specialised in degradation of defined carbohydrates, increasing our knowledge of the range of species potentially involved in starch metabolism in the human gut.

Material and Methods

A schematic overview of the workflow and experimental design is displayed in Figure 1.

Substrates. Native maize starch (catalogue no. S4126), native potato starch (catalogue no. 2004), Avicel PH-101 (catalogue no. 11365) and chicory inulin (catalogue no. I2255) were purchased from Sigma-Aldrich, (Gillingham, UK). Hylon VII® was kindly provided as a gift by Ingredion Incorporated (Manchester, UK).

Retrograded maize starch was prepared from 40g of native maize starch in 400 mL of deionized water. The slurry was stirred continuously at 95°C in a water bath for 20 minutes. The resulting gel was cooled to room temperature for 60 minutes, transferred to aluminium pots (150 mL, Ampulla, Hyde UK), and stored at 4°C for 48 hours. The retrograded gel was then frozen at -80°C for 12 hours and freeze-dried (LyoDry, MechaTech Systems Ltd, Bristol, UK) for 72 hours.

342 Each substrate (0.500 ± 0.005g, dry weight) was weighed in sterilized fermentation
343 bottles (100 mL) prior to start of the experiment.

344 **Inoculum collection and preparation.** A single human faecal sample was obtained from one
345 adult (≥ 18 years old), free-living, healthy donor who had not taken antibiotics in the 3
346 months prior to donation and was free from gastrointestinal disease. Ethical approval was
347 granted by Human Research Governance Committee at the Quadram Institute (IFR01/2015)
348 and London - Westminster Research Ethics Committee (15/LO/2169) and the trial was
349 registered on clinicaltrials.gov (NCT02653001). A signed informed consent was obtained
350 from the participant prior to donation. The stool sample was collected by the participant,
351 stored in a closed container under ambient conditions, transferred to the laboratory and
352 prepared for inoculation within 2 hours of excretion. The faecal sample was diluted 1:10
353 with pre-warmed, anaerobic, sterile phosphate buffer saline (0.1M, pH 7.4) in a double
354 meshed stomacher bag (500 mL, Seward, Worthing, UK) and homogenized using a
355 Stomacher 400 (Seward, Worthing, UK) at 200 rpm for two cycles, each of 60 seconds
356 length.

357 **Batch fermentation in the model colon.** Fermentation vessels were established with media
358 adapted from Williams *et al.*,[40] In brief, each vessel (100 mL) contained an aliquot (3.0 mL)
359 of filtered faecal slurry, 82 mL of sterilized growth medium, and one of the six substrates for
360 experimental evaluation: native Hylon VII or native potato starch (highly recalcitrant
361 starches); native maize starch or gelatinized, retrograded maize starch (accessible starches);
362 Avicel PH-101 (insoluble fibre; negative control); and chicory inulin (fermentable soluble
363 fibre; positive control). There was also a media only control with no inoculum (blank)
364 making a total of seven fermentation vessels.

For each fermentation vessel the growth medium contained 76 mL of basal solution, 5 mL vitamin phosphate and sodium carbonate solution, and 1 mL reducing agent. The composition of the various solutions used in the preparation of the growth medium is described in detail in **Supplementary Table 13**. A single stock (7 litres) of growth medium was prepared for use in all vessels. Vessel fermentations were pH controlled and maintained at pH 6.8 to 7.2 using 1N NaOH and 1N HCl regulated by a Fermac 260 (Electrolab Biotech, Tewkesbury, UK). A circulating water jacket maintained the vessel temperature at 37°C. Magnetic stirring was used to keep the mixture homogenous and the vessels were continuously sparged with nitrogen (99% purity) to maintain anaerobic conditions. Samples were collected from each vessel at 0 (5 min), 6, 12, and 24 hours after inoculation. The biomass from two 1.8 mL aliquots from each sample were concentrated by refrigerated centrifugation (4°C; 10,000 g for 10 min), the supernatant removed, and the pellets stored at -80°C prior to bacterial enumeration and DNA extraction; one pellet was used for enumeration and one for DNA extraction.

Bacterial cell enumeration. All materials used for bacterial cell enumeration were purchased from Sigma-Aldrich (Gillingham, UK), unless specified otherwise. To each frozen pellet, 400 µL of PBS and 1100 µL of 4% paraformaldehyde (PFA) were added and gently thawed at 20°C for 10 minutes with gentle mixing. Once thawed, each resuspension was thoroughly mixed and incubated overnight at 4°C for fixation to occur. The resuspensions were then centrifuged for 10 minutes at 8000 x g, the supernatant removed, and the residual pellet washed with 1 mL 0.1% Tween-20. This pellet then underwent two further washes in PBS to remove any residual PFA and was then resuspended in 600 µL PBS: ethanol (1:1).

The fixed resuspensions were centrifuged for 3 minutes at 16000 x g, the supernatant removed, and the pellet resuspended in 500 µL 1 mg/mL lysozyme (100 µL 1M Tris HCl at pH 8, 100 µL 0.5 M EDTA at pH 8, 800 µL water, and 1 mg lysozyme, catalogue no. L6876) and incubated at room temperature for 10 minutes. After thorough mixing and centrifugation for 3 minutes at 16000 x g, the supernatant was removed, and the pellet washed with PBS. The resulting pellet was then resuspended in 150 µL of hybridisation buffer (HB, per mL: 180 µL 5 M NaCl, 20 µL 1M Tris HCl at pH 8, 300 µL Formamide, 499 µL water, 1 µL 10% SDS), centrifuged, the supernatant removed and the remaining pellet resuspended again in 1500 µL of HB and stored at 4°C prior to enumeration. For bacterial enumeration, 1 µL of Invitrogen SYTO 9 (catalogue no. S34854, Thermo Fisher Scientific, Loughborough, UK) was added to 1 mL of each fixed and washed resuspension. Within 96-well plate resuspensions were diluted to 1:1000 and the bacterial populations within them enumerated using flow cytometry (Luminex Guava easyCyte 5) at wavelength of 488nm and Guava suite software, version 3.3.

DNA extraction. Each pellet was resuspended in 500 µL (samples collected at 0 and 6 hr) or 650 µL (samples collected at 12 and 24 hr) with chilled (4°C) nuclease-free water (Sigma-Aldrich, Gillingham, UK). The resuspensions were frozen overnight at -80°C, thawed on ice and an aliquot (400 µL) used for bacterial genomic DNA extraction. FastDNA® Spin Kit for Soil (MP Biomedical, Solon, US) was used according to the manufacturer's instructions which included two bead-beating steps of 60s at a speed of 6.0m/s (FastPrep24, MP Biomedical, Solon, USA). DNA concentration was determined using the Quant-iT™ dsDNA Assay Kit, high sensitivity kit (Invitrogen, Loughborough, UK) and quantified using a FLUOstar Optima plate reader (BMG Labtech, Aylesbury, UK).

411 **Illumina NovaSeq Library preparation and sequencing.** Genomic DNA was normalised to 5
412 ng/µL with elution buffer (10mM Tris-HCl). A miniaturised reaction was set up using the
413 Nextera DNA Flex Library Prep Kit (Illumina, Cambridge, UK). 0.5 µL Tagmentation Buffer 1
414 (TB1) was mixed with 0.5 µL Bead-Linked Transposomes (BLT) and 4.0 µL PCR-grade water in
415 a master mix and 5 µL added to each well of a chilled 96-well plate. 2 µL of normalised DNA
416 (10 ng total) was pipette-mixed with each well of tagmentation master mix and the plate
417 heated to 55°C for 15 minutes in a PCR block. A PCR master mix was made up using 4 µL
418 kapa2G buffer, 0.4 µL dNTP's, 0.08 µL Polymerase and 4.52 µL PCR grade water, from the
419 Kap2G Robust PCR kit (Sigma-Aldrich, Gillingham, UK) and 9 µL added to each well in a 96-
420 well plate. 2 µL each of P7 and P5 of Nextera XT Index Kit v2 index primers (catalogue No.
421 FC-131-2001 to 2004; Illumina, Cambridge, UK) were also added to each well. Finally, the 7
422 µL of Tagmentation mix was added and mixed. The PCR was run at 72°C for 3 minutes, 95°C
423 for 1 minute, 14 cycles of 95°C for 10s, 55°C for 20s and 72°C for 3 minutes. Following the
424 PCR reaction, the libraries from each sample were quantified using the methods described
425 earlier and the high sensitivity Quant-iT dsDNA Assay Kit. Libraries were pooled following
426 quantification in equal quantities. The final pool was double-SPRI size selected between 0.5
427 and 0.7X bead volumes using KAPA Pure Beads (Roche, Wilmington, US). The final pool was
428 quantified on a Qubit 3.0 instrument and run on a D5000 ScreenTape (Agilent, Waldbronn,
429 DE) using the Agilent Tapestation 4200 to calculate the final library pool molarity. qPCR was
430 done on an Applied Biosystems StepOne Plus machine. Samples quantified were diluted 1 in
431 10,000. A PCR master mix was prepared using 10 µL KAPA SYBR FAST qPCR Master Mix (2X)
432 (Sigma-Aldrich, Gillingham, UK), 0.4 µL ROX High, 0.4 µL 10 µM forward primer, 0.4 µL 10
433 µM reverse primer, 4 µL template DNA, 4.8 µL PCR grade water. The PCR programme was:
434 95°C for 3 minutes, 40 cycles of 95°C for 10s, 60°C for 30s. Standards were made from a 10

nM stock of Phix, diluted in PCR-grade water. The standard range was 20 pmol, 2 pmol, 0.2 pmol, 0.02 pmol, 0.002 pmol, 0.0002 pmol. Samples were then sent to Novogene (Cambridge, UK) for sequencing using an Illumina NovaSeq instrument, with sample names and index combinations used. Demultiplexed FASTQ's were returned on a hard drive.

Nanopore library preparation and PromethION sequencing. Library preparation was performed using SQK-LSK109 (Oxford Nanopore Technologies, Oxford, UK) with barcoding kits EXP-NBD104 and EXP-NBD114. The native barcoding genomic DNA protocol by Oxford Nanopore Technologies (ONT) was followed with slight modifications. Starting material for the End-Prep/FFPE reaction was 1 µg per sample in 48 µL volume. 3.5 µL NEBNext FFPE DNA Repair Buffer (NEB, New England Biolabs, Ipswich, USA), 3.5 µL NEB Ultra II End-prep Buffer, 3 µL NEB Ultra II End-prep Enzyme Mix and 2 µL NEBNext FFPE DNA Repair Mix (NEB) were added to the DNA (final volume 60 µL), mixed slowly by pipetting and incubated at 20°C for 5 minutes and then 65°C for 5 minutes. After a 1X bead wash with AMPure XP beads (Agencourt, Beckman Coulter, High Wycombe, UK), the DNA was eluted in 26 µL of nuclease-free water. 22.5 µL of this was taken forward for native barcoding with the addition of 2.5 µL barcode and 25 µL Blunt/TA Ligase Master Mix (NEB) (final volume 50 µL). This was mixed by pipetting and incubated at room temperature for 10 minutes. After another 1X bead wash (as above), samples were quantified using Qubit dsDNA BR Assay Kit (Invitrogen, Loughborough, UK). In the first run, samples were equimolar pooled to a total of 900 ng in a volume of 65 µL. In the second run, samples were pooled to 1700 ng followed by a 0.4X bead wash to achieve the final volume of 65 µL. 5 µL Adapter Mix II (ONT), 20 µL NEBNext Quick Ligation Reaction Buffer (5X) and 10 µL Quick T4 DNA Ligase (NEB) were added (final volume 100 µL), mixed by flicking, and incubated at room temperature for 10 minutes. After bead washing with 50 µL of AMPure XP beads and two washes in 250 µL of

Long Fragment Buffer (ONT), the library was eluted in 25 µL of Elution Buffer and quantified with Qubit dsDNA BR and TapeStation 2200 using a Genomic DNA ScreenTape (Agilent Technologies, Edinburgh, UK). 470 ng of DNA was loaded for sequencing in the first run and 400 ng in the second run. The final loading mix was 75 µL SQB, 51 µL LB and 24 µL DNA library.

Sequencing was performed on a PromethION Beta using FLO-PRO002 PromethION Flow Cells (R9 version). The sequencing runtime was 57 hours for Run 1 and 64 hours for Run 2. Flow cells were refuelled with 0.5X SQB (75 µL SQB and 75 µL nuclease free water) 40 hours into both runs.

Bioinformatics analysis. The bioinformatics analysis was performed using default options unless specified otherwise.

Nanopore basecalling: Basecalling was performed using Guppy version 3.0.5+45c3543 (ONT) in high accuracy mode (model dna_r9.4.1_450bps_hac), and demultiplexed with qcat version 1.1.0 (Oxford Nanopore Technologies, https://github.com/nanoporetech/qcat).

Sequence quality: For Nanopore, sequence metrics were estimated by Nanostat version 1.1.2[41]. In total, 22 million sequences were generated with a median read length of 4500 bp and median quality of 10 (phred). Quality trimming and adapter removal was performed using Porechop version 0.2.3 (https://github.com/rrwick/Porechop). For Illumina, quality control was done for paired-end reads using *fastp*, version 0.20.0.[42] to remove adapter sequences and filter out low-quality (phred quality < 30) and short reads (length < 60 bp). After quality control, the average number of reads in the samples was over 26.1 million reads, with a minimum of 9.7 million reads; the average read length was 148 bp.

Taxonomic profiling: Trimmed and high-quality short reads are processed using MetaPhlAn3 version 3.0.2,[43] to estimate both microbial composition to species level and also the relative

abundance of species from each metagenomic sample. MetaPhlAn3 uses the latest marker information dataset, CHOCOPhlAn 2019, which contains ~1 million unique clade-specific marker genes identified from ~100,000 reference genomes; this includes bacterial, archaeal and eukaryotic genomes. Hclust2 was used to plot the hierarchical clustering of the different taxonomic profiles at each time point [https://github.com/SegataLab/hclust2]. The results of the microbial taxonomy were analysed in RStudio Version 1.1.453 (http://www.rstudio.com/).

Principle Coordinate analyses using the pcoa function in the ape package version 5.3 (https://www.rdocumentation.org/packages/ape/versions/5.3) and the vegan package was used to identify differences in microbiome profiles amongst treatments.

Hybrid assembly: Trimmed and high-quality Illumina reads were merged per treatment, and then used in a short-read-only assembly using Megahit version 1.1.3.[44,45] Then OPERA-MS[46] version 0.8.2, was used to combine the short-read only assembly with high-quality long reads, to create high-quality hybrid assemblies. By combining these two technologies, OPERA-MS overcomes the issue of low-contiguity of short-read-only assemblies and the low base-pair quality of long-read-only assemblies.

Genome binning, quality, dereplication and comparative genomics of hybrid assemblies: The hybrid co-assemblies from Opera-MS[46] were used for binning. Here, Illumina reads for each time period were mapped to the co-assembled contigs to obtain a coverage map. Bowtie2 version 2.3.4.1 was used for mapping, and samtools to convert SAM to BAM format. MaxBin2 version 2.2.6[47] and MetaBat2 version 2.12.1[48] which uses sequence composition and coverage information, was used to bin probable genomes using default parameters. The binned genomes and co-assembled contigs were integrated into Anvi'o version 6.1 for manual refinement and visual inspection of problematic genomes. In

507 particular, we used the scripts: 'anvi-interactive' to visualise the genome bins; 'anvi-run-
508 hmms' to estimate genome completeness and contamination; 'anvi-profile' to estimate
509 coverage and detection statistics for each sample; and 'anvi-refine' to manually refine the
510 genomes. All scripts were run using default parameters. Additionally, DAS tool version 1.1.2
511 [49] was used to aggregate high-quality genomes from each treatment by using single copy
512 gene-based scores and genome quality metrics to produce a list of good quality genomes for
513 every treatment. Additionally, checkM version 1.0.18[50] was used on all final genomes to
514 confirm completion and contamination scores. In general, genomes with a 'quality satisfying
515 completeness - 5*contamination > 50 score' and/or with a '>60% completion and <10%
516 contamination' score according to CheckM, were selected for downstream analyses.
517 *Dereplication into representative clusters:* In order to produce a dereplicated set of
518 genomes across all treatments, dRep version 2.5.0[51] was used. Pairwise genome
519 comparisons or Average Nucleotide Identity (ANI) was used for clustering. dRep clusters
520 genomes with ANIs of 97% were regarded as primary clusters, and genomes with ANI of 99
521 % regarded as secondary clusters. A representative genome is provided for each of the
522 secondary clusters.
523 *Relative abundance of genomes:* Since co-assemblies were used for binning, relative
524 abundance was calculated as the proportion of reads recruited to that bin across all time
525 periods for each treatment. This provides an estimate of which time period recruited the
526 most reads. To provide this estimate in relative terms, the value is normalised to the total
527 number of reads that was recruited for that genome. As for Avicel that misses the time 0h, a
528 mean relative abundance from each MAG in the cluster at time 0h was used. The relative
529 abundance scores was provided by 'anvi-summarize' (from the Anvi'o package) as relative
530 abundance. Further, fold changes were calculated between the relative abundance at time

0h to the corresponding relative abundance at 6h, 12h and 24h using gtools R package version 3.5.0. Fold changes provide an estimate of change in MAG abundance which might be a result from utilisation of a particular carbohydrate. Fold changes were converted to log ratios. MAGs with a fold change of 2x (log_2 foldchange=1) were regarded as an active carbohydrate utiliser.

Metagenomic assignment and phylogenetic analyses: Genome bins that passed quality assessment were analysed for their closest taxonomic assignment. To assign taxonomic labels, the genome set was assigned into the microbial tree of life using GTDB version 0.3.5 and database R95 to identify the closest ancestor and obtain a putative taxonomy assignment for each genome bin. For genomes where the closest ancestor could not be determined, the Relative Evolutionary Distance (RED) to the closest ancestor and novel taxa names were provided. Using these genome bins, a phylogenetic tree was constructed using Phylophlan version 0.99 and visually inspected using iTOL version 4.3.1 and ggtree from package https://github.com/YuLab-SMU/ggtree.git. The R packages ggplot2 version 3.3.2, dplyr version 1.0.2, aplot, ggtree version 2.2.4 and inkscape version 1.0.1 were used for illustrations

Carbohydrate metabolism analyses: All representative genome clusters were annotated for CAZymes using dbCAN.[52] The genome's nucleotide sequences were processed with Prodigal to predict protein sequences, and then three tools were used for automatic CAZyme annotation: a) HMMER[53] to search against the dbCAN HMM (Hidden Markov Model) database; b) DIAMOND[54] to search against the CAZy pre-annotated CAZyme sequence database; and c) Hotpep[55] to search against the conserved CAZyme PPR (peptide pattern recognition) short peptide library. To improve annotation accuracy, a filtering step

554 was used to retain only hits to CAZy families found by at least two tools. The R packages
555 ggplot2, dplyr, ComplexHeatmap version 2.4.3 and inkscape were used for illustrations.

556 **Acknowledgements**

557 We thank Dave J. Baker for assisting with sequencing and the anonymous donor who
558 provided faecal material for this study. We thank Dr. Judith Pell for assistance with editing
559 the manuscript.

560 **Author contributions**

561 All authors read and contributed to the manuscript. AR, PR and JAJ are joint first authors.
562 FJW conceived and designed the study. AR led on the preparation of the manuscript. AA and
563 GLK prepared the sequencing libraries and did the sequencing. AR and PR did the sequence
564 and bioinformatics analysis. TLV did the post-sequencing analysis. JAJ, KC and SH did the
565 model colon experiments and DNA extractions. HH enumerated the bacterial cells. RG and
566 MJP assisted with bioinformatic analysis and taxonomic descriptions. JOG provided long-
567 read sequencing and molecular biology expertise; AJP provided bioinformatics expertise;
568 and FJW provided expertise in carbohydrate structure and model colon protocols. FJW, JOG,
569 AJP secured funding, provided management oversight and scientific direction.

570 **Ethical approval**

571 Ethical approval was granted by the Human Research Governance Committee at the
572 Quadram Institute (IFR01/2015) and the London - Westminster Research Ethics Committee
573 (15/LO/2169). The trial is registered on clinicaltrials.gov (NCT02653001). A signed informed
574 consent was obtained from the participant prior to donation.

575 **Funding statements**

576 The authors gratefully acknowledge the support of the Biotechnology and Biological
577 Sciences Research Council (BBSRC). This research was funded by: the BBSRC Institute

Strategic Programme (ISP) Food Innovation and Health BB/R012512/1 and its constituent projects (BBS/E/F/000PR10343, BS/E/F/000PR10346); the BBSRC ISP Microbes in the Food Chain BB/R012504/1 and its constituent projects (BBS/E/F/000PR10348, BBS/E/F/000PR10349, BBS/E/F/000PR10352); and the BBSRC Core Capability Grant (BB/CCG1860/1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability

Raw read data from the PromethION and NovoSeq sequencing runs can be accessed through the NCBI SRA project number PRJNA722408. GenBank accession numbers for individual MAG's within this ProjectID can be found in Supplementary Table 5.

Figure legends

Figure 1. Workflow for bioinformatics analysis of combined Illumina NovoSeq and Oxford Nanopore PromethION metagenomics data collected in a model colon study of the fermentation of different carbohydrate substrates with contrasting structures (Avicel, Inulin, Normal maize (N.maize), Retrograded maize (R.maize), Potato and Hylon) by the gut microbiota present in a human stool sample**.**

Figure 2. **Hierarchical clustering** of the top 30 selected gut microbial species present after fermentation of Avicel, Inulin, N.maize, R.maize, Potato and Hylon at 0h, 6h, 12h and 24h in the model colon. The hierarchical clustering also includes a water sample ("the kitome").

Figure 3: **Comparison of Illumina short read assemblies and hybrid assemblies:** a) shows the number of contigs per treatment, b) shows the N50, c) statistics on the largest contig, d) size of the total assembly for each carbohydrate treatment.

Figure 4: MAG quality. Dots represent each MAG. Completeness and contamination scores were estimated using CheckM. Colours are based on the MAG standards (high quality as >90% completeness & <5% contamination; good quality as <90%- 60% completeness and >5% - 10% contamination. The horizontal and vertical bar charts provide the number of genomes with high completeness and low contamination scores.

Figure 5: Phylogenomic tree and fold changes. The phylogenetic tree was **constructed from concatenated protein sequences using PhyloPhlAn and illustrated using ggtree.** Clades belonging to similar bacterial family and bacterial genus were collapsed. The colour strips represent the phylum-level distribution of the phylogenetic tree. Dot plot shows the decrease (negative $\log_2$ fold change; blue shades) and increase (positive $\log_2$ fold change; red shades) of read proportions from 0h to 6h, 0h to 12h and 0h to 24h for all treatments.

Figure 6: CAZyme profiles of selected-MAGs. The colour strip represents the phylum-based taxonomy annotation. The heat map represents the number of proteins identified for each CAZy protein family.

Supplementary Files

Supplementary Table 1: Read stats and quality metrics for PromethION and Illumina sequence data

Supplementary Table 2: Taxonomy profiles of relative abundances for all treatments using MetaPhlAn3.

Supplementary Table 3: Assembly stats for short read assemblies using Megahit and hybrid assemblies using OPERA-MS

Supplementary Table 4: MAG genomic stats, assembly features, closest taxonomy annotation and relative evolutionary distance for novel genus and species.

Supplementary Table 5: Dereplicated MAGs with representative cluster names and their taxonomy annotations

Supplementary Table 6: Stats showing the diversity of GTDb taxonomy within MAGs.

Supplementary Table 7: Novel latin binomials for MAGs and taxa names submitted to Genbank

Supplementary Table 8: Comparison of genome stats between MAGs from this study and GTDb corresponding representative MAG cluster

Supplementary Table 9: Relative abundance, fold change and log ratio foldchange for all MAGs

Supplementary Table 10: Genomes depicted as early and late degraders according to the time the genomes showed a 2x fold change.

Supplementary Table 11: MAGs and their CAZyme profiles.

635 **Supplementary Table 12:** CaZymes counts for selected MAG clusters

636 **Supplementary Table 13:** media preparation materials, sources and quantity

637

638 **Supplementary Figure 1:** Principle Component Analysis (PCoA) showing the dynamics of the

639 microbiome during the different time points and between the Carbohydrate treatment. PC1

640 and PC2 represent the percentage of variance explained by Principle Component (PC) 1 and

641 2.

642 **Supplementary Figure 2:** Changes in inverse Simpson index between time periods of the

643 substrates.

644 **Supplementary figure 3: Box plots showing the dynamic shifts in read proportions for all**

645 **binned MAGs after 0h, 6h, 12h and 24h fermentation in the model colon.** The box

646 represents the interquartile range (IQR) (25^{th} and 75th percentile); the median is shown

647 within the box. The whiskers indicate minimum and maximum Inter Quartile Range (IQR);

648 dots represent outliers.

649 **Supplementary Figure 4:** Distribution of CAZy families per substrate and in all the genome

References

1 Kau, A. L., Ahern, P. P., Griffin, N. W., Goodman, A. L. & Gordon, J. I. Human nutrition, the gut microbiome and the immune system. *Nature* **474**, 327-336 (2011).

2 Koropatkin, N. M., Cameron, E. A. & Martens, E. C. How glycan metabolism shapes the human gut microbiota. *Nature Reviews Microbiology* **10**, 323-335 (2012).

3 Chambers, E. S. *et al.* Dietary supplementation with inulin-propionate ester or inulin improves insulin sensitivity in adults with overweight and obesity with distinct effects on the gut microbiota, plasma metabolome and systemic inflammatory responses: a randomised cross-over trial. *Gut* **68**, 1430-1438 (2019).

4 Blaak, E. *et al.* Short chain fatty acids in human gut and metabolic health. *Beneficial Microbes* **11**, 411-455 (2020).

5 Lloyd-Price, J. *et al.* Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* **550**, 61-66 (2017).

6 Martens, E. C., Kelly, A. G., Tauzin, A. S. & Brumer, H. The devil lies in the details: how variations in polysaccharide fine-structure impact the physiology and evolution of gut microbes. *Journal of molecular biology* **426**, 3851-3865 (2014).

7 Warren, F. J. *et al.* Food starch structure impacts gut microbiome composition. *Msphere* **3** (2018).

8 Deehan, E. C. *et al.* Precision microbiome modulation with discrete dietary fiber structures directs short-chain fatty acid production. *Cell Host & Microbe* (2020).

9 Carmody, R. N. *et al.* Cooking shapes the structure and function of the gut microbiome. *Nature microbiology* **4**, 2052-2063 (2019).

10 Lapébie, P., Lombard, V., Drula, E., Terrapon, N. & Henrissat, B. Bacteroidetes use thousands of enzyme combinations to break down glycans. *Nat. Commun.* **10**, 1-7 (2019).

11 Kujawska, M. *et al.* Succession of Bifidobacterium longum strains in response to the changing early-life nutritional environment reveals specific adaptations to distinct dietary substrates. (2020).

12 Charalampous, T. *et al.* Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. *Nat. Biotechnol.* **37**, 783-792 (2019).

13 De Coster, W. *et al.* Structural variants identified by Oxford Nanopore PromethION sequencing of the human genome. *Genome Res.* **29**, 1178-1187 (2019).

14 Bertrand, D. *et al.* Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. *Nat. Biotechnol.* **37**, 937-944 (2019).

15 Arumugam, K. *et al.* Annotated bacterial chromosomes from frame-shift-corrected long-read metagenomic data. *Microbiome* **7**, 61 (2019).

16 Singleton, C. M. *et al.* Connecting structure to function with the recovery of over 1000 high-quality activated sludge metagenome-assembled genomes encoding full-length rRNA genes using long-read sequencing. *bioRxiv* (2020).

17 Stewart, R. D. *et al.* Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nature biotechnology* **37**, 953 (2019).

18 Walker, A. W., Duncan, S. H., Leitch, E. C. M., Child, M. W. & Flint, H. J. pH and peptide supply can radically alter bacterial populations and short-chain fatty acid

ratios within microbial communities from the human colon. *Appl. Environ. Microbiol.* **71**, 3692-3700 (2005).

19 Leitch, E. C. M., Walker, A. W., Duncan, S. H., Holtrop, G. & Flint, H. J. Selective colonization of insoluble substrates by human faecal bacteria. *Environ. Microbiol.* **9**, 667-679 (2007).

20 Bowers, R. M. *et al.* Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725-731 (2017).

21 Pallen, M. J., Telatin, A. & Oren, A. The Next Million Names for Archaea and Bacteria. *Trends Microbiol.* (2020).

22 Machovič, M. & Janeček, Š. Domain evolution in the GH13 pullulanase subfamily with focus on the carbohydrate-binding module family 48. *Biologia* **63**, 1057-1068 (2008).

23 Moss, E. L., Maghini, D. G. & Bhatt, A. S. Complete, closed bacterial genomes from microbiomes using nanopore sequencing. *Nature Biotechnology*, 1-7 (2020).

24 Maghini, D. G., Moss, E. L., Vance, S. E. & Bhatt, A. S. Improved high-molecular-weight DNA extraction, nanopore sequencing and metagenomic assembly from the human gut microbiome. *Nature Protocols* **16**, 458-471 (2021).

25 Aagaard, K. *et al.* The Human Microbiome Project strategy for comprehensive sampling of the human microbiome and why it matters. *The FASEB Journal* **27**, 1012-1022 (2013).

26 Methé, B. A. *et al.* A framework for human microbiome research. *Nature* **486**, 215 (2012).

27 Campbell, J. M., Fahey Jr, G. C. & Wolf, B. W. Selected indigestible oligosaccharides affect large bowel mass, cecal and fecal short-chain fatty acids, pH and microflora in rats. *The Journal of nutrition* **127**, 130-136 (1997).

28 Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research* **42**, D490-D495 (2014).

29 El Kaoutari, A., Armougom, F., Gordon, J. I., Raoult, D. & Henrissat, B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nature Reviews Microbiology* **11**, 497-504 (2013).

30 Benítez-Páez, A., Gómez del Pulgar, E. M. & Sanz, Y. The glycolytic versatility of Bacteroides uniformis CECT 7771 and its genome response to oligo and polysaccharides. *Frontiers in cellular and infection microbiology* **7**, 383 (2017).

31 Moens, F., Weckx, S. & De Vuyst, L. Bifidobacterial inulin-type fructan degradation capacity determines cross-feeding interactions between bifidobacteria and Faecalibacterium prausnitzii. *International journal of food microbiology* **231**, 76-85 (2016).

32 Ramirez-Farias, C. *et al.* Effect of inulin on the human gut microbiota: stimulation of Bifidobacterium adolescentis and Faecalibacterium prausnitzii. *British Journal of Nutrition* **101**, 541-550 (2008).

33 Flint, H. J., Scott, K. P., Duncan, S. H., Louis, P. & Forano, E. Microbial degradation of complex carbohydrates in the gut. *Gut microbes* **3**, 289-306 (2012).

34 Bui, T. P. N. *et al.* Mutual Metabolic Interactions in Co-cultures of the Intestinal Anaerostipes rhamnosivorans With an Acetogen, Methanogen, or Pectin-Degrader Affecting Butyrate Production. *Frontiers in microbiology* **10**, 2449 (2019).

35 Ze, X. *et al.* Unique organization of extracellular amylases into amylosomes in the resistant starch-utilizing human colonic Firmicutes bacterium Ruminococcus bromii. *MBio* **6** (2015).

36 Upadhyaya, B. *et al.* Impact of dietary resistant starch type 4 on human gut microbiota and immunometabolic functions. *Scientific reports* **6**, 28797 (2016).

37 Xie, Z. *et al.* In vitro fecal fermentation of propionylated high-amylose maize starch and its impact on gut microbiota. *Carbohydrate polymers* **223**, 115069 (2019).

38 Ryan, S. M., Fitzgerald, G. F. & van Sinderen, D. Screening for and identification of starch-, amylopectin-, and pullulan-degrading activities in bifidobacterial strains. *Applied and Environmental Microbiology* **72**, 5289-5296 (2006).

39 Crittenden, R. *et al.* Adhesion of bifidobacteria to granular starch and its implications in probiotic technologies. *Applied and Environmental Microbiology* **67**, 3469-3475 (2001).

40 Williams, B. A., Bosch, M. W., Boer, H., Verstegen, M. W. & Tamminga, S. An in vitro batch culture method to assess potential fermentability of feed ingredients for monogastric diets. *Anim. Feed Sci. Technol.* **123**, 445-462 (2005).

41 De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M. & Van Broeckhoven, C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **34**, 2666-2669, doi:10.1093/bioinformatics/bty149 (2018).

42 Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890 (2018).

43 Truong, D. T. *et al.* MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nature methods* **12**, 902-903 (2015).

44 Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674-1676 (2015).

45 Li, D. *et al.* MEGAHIT v1. 0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* **102**, 3-11 (2016).

46 Bertrand, D. *et al.* Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. *Nat. Biotechnol.* **37**, 937-944, doi:10.1038/s41587-019-0191-2 (2019).

47 Wu, Y.-W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605-607 (2016).

48 Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165, doi:10.7717/peerj.1165 (2015).

49 Sieber, C. M. K. *et al.* Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat. Microbiol.* **3**, 836-843, doi:10.1038/s41564-018-0171-1 (2018).

50 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043-1055, doi:10.1101/gr.186072.114 (2015).

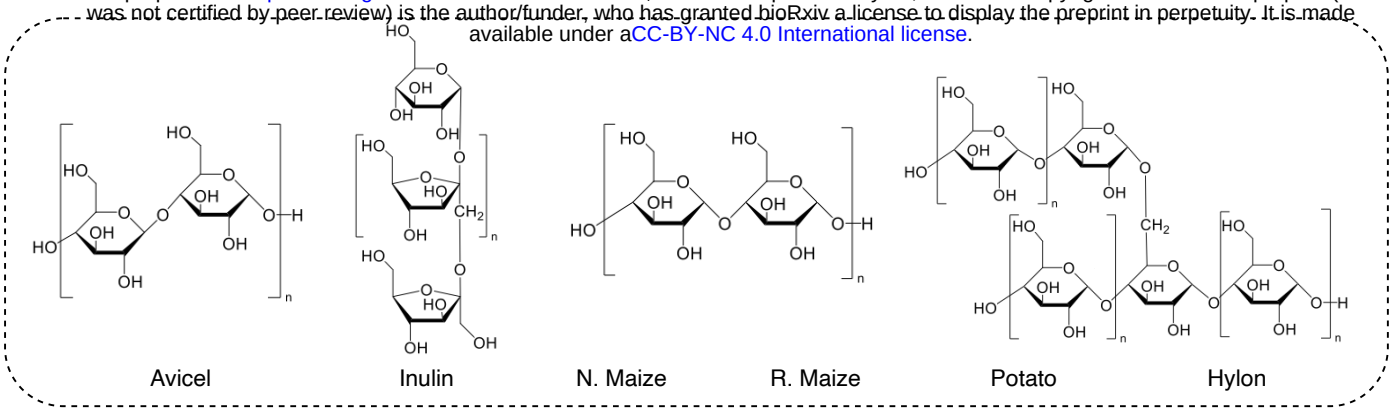
51 Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME journal* **11**, 2864-2868 (2017).

52 Zhang, H. *et al.* dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **46**, W95-W101 (2018).

53 Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic acids research* **39**, W29-W37 (2011).

54 Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nature methods* **12**, 59-60 (2015).

55 Busk, P. K., Pilgaard, B., Lezyk, M. J., Meyer, A. S. & Lange, L. Homology to peptide pattern for annotation of carbohydrate-active enzymes and prediction of function. *BMC bioinformatics* **18**, 214 (2017).



Substrate

Human faecal sample

Inoculum Collection and Preparation

Faecal Slurry

growth medium

Fermentation Experiment for each substrate

Colon Model

0h 6h 12h 24h

Sample Collecction

DNA Extraction

Illumina NovaSeq Sequencing

Nanopore PromethION Sequencing

QC

fastp

Nanopore basecalling

Guppy qcat

Taxonomic Classification

Metaphlan3

Assembly

Megahit

QC

Nanostat Porechop

Microbial Composition

Hybrid Assembly

Opera-MS

Binning

Metabat MaxBin2

Dereplication

DAS Tools dRep

Refining

Anvi'o

QC for MAGs

CheckM Anvi'o

PPR Library

CAZy database

HMM database

Genome Annotation

GTDb

CAZyme Annotation

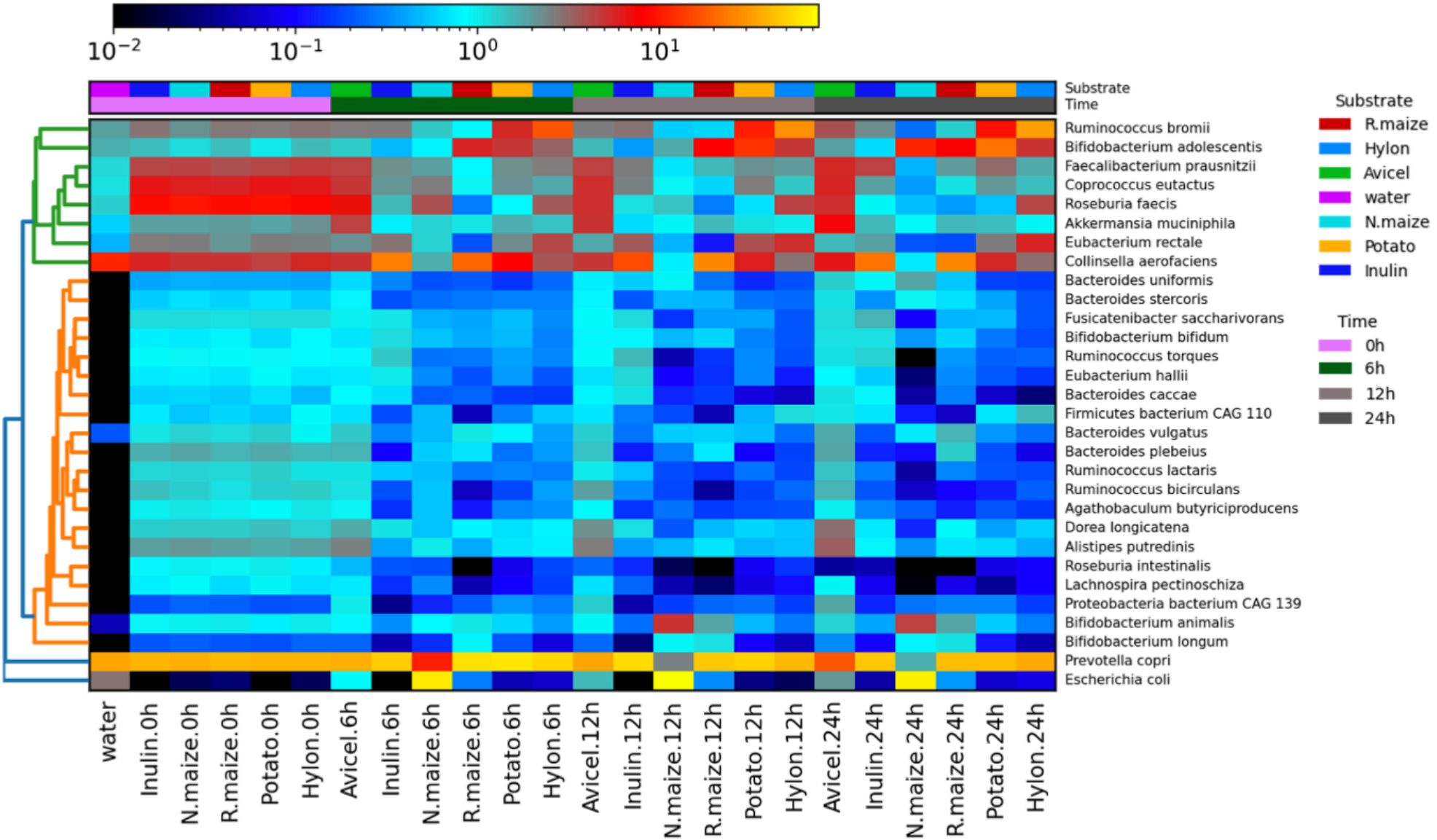
dbCAN HMMER Diamond Hotpep

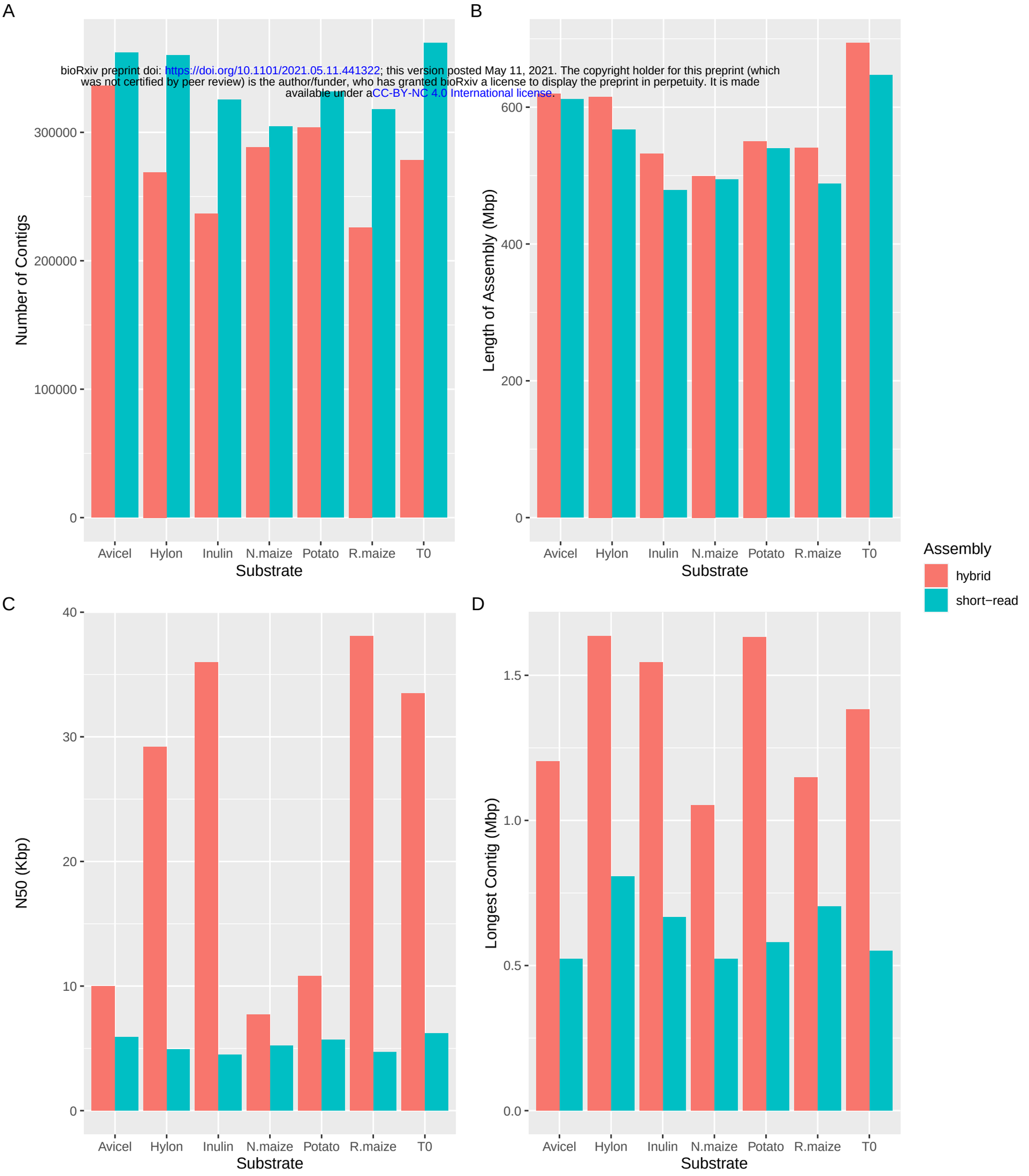
CAZyme Profiles

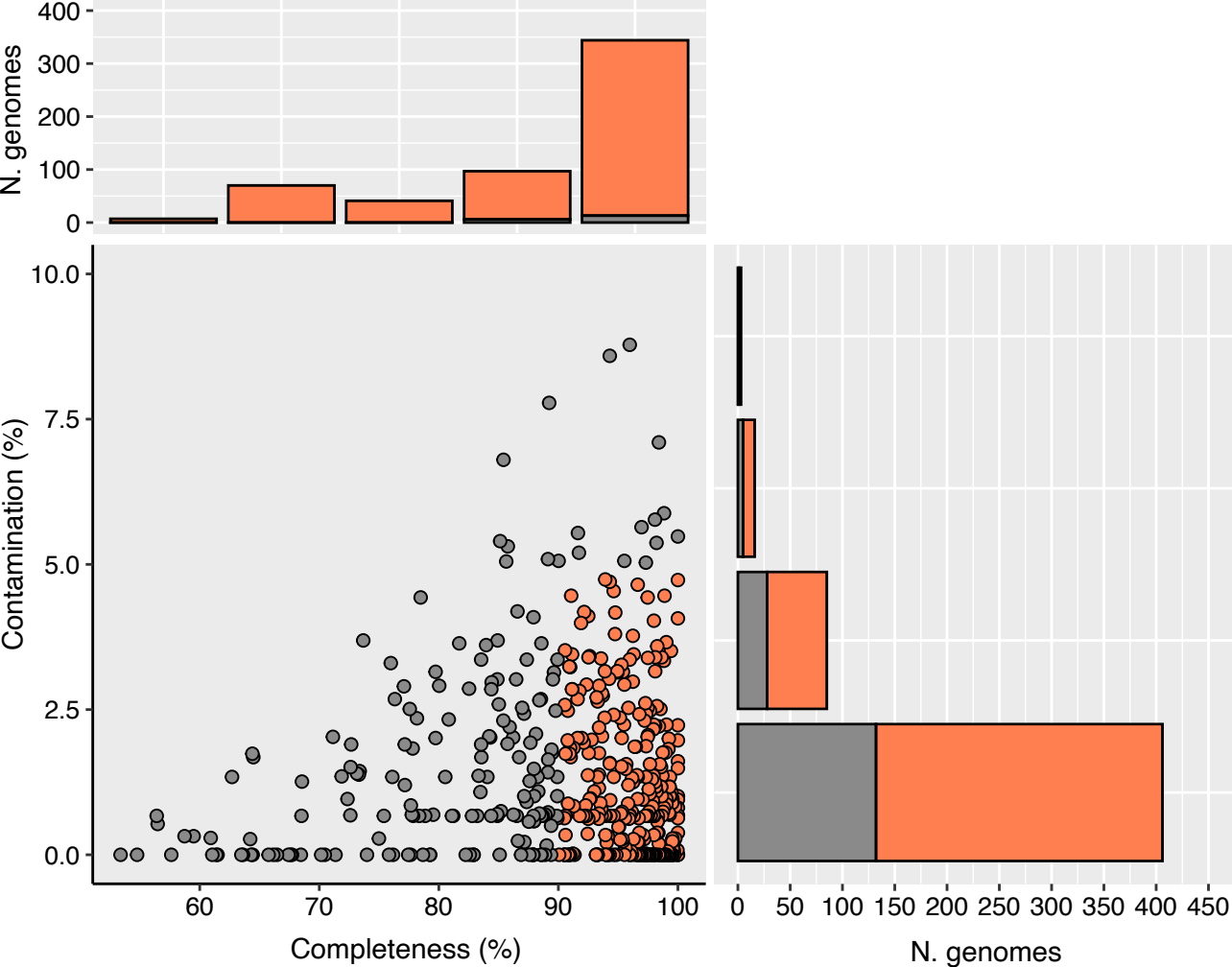
Phylogenetic Analysis

PhyloPhlAn

MAGs (including novel species)


10^{-2}
10^{-1}
10^{0}
10^{1}
Substrate
Time
Ruminococcus bromii
Bifidobacterium adolescentis
Faecalibacterium prausnitzii
Coprococcus eutactus
Roseburia faecis
Akkermansia muciniphila
Eubacterium rectale
Collinsella aerofaciens
Bacteroides uniformis
Bacteroides stercoris
Fusicatenibacter saccharivorans
Bifidobacterium bifidum
Ruminococcus torques
Eubacterium hallii
Bacteroides caccae
Firmicutes bacterium CAG 110
Bacteroides vulgatus
Bacteroides plebeius
Ruminococcus lactaris
Ruminococcus bicirculans
Agathobaculum butyriciproducens
Dorea longicatena
Alistipes putredinis
Roseburia intestinalis
Lachnospira pectinoschiza
Proteobacteria bacterium CAG 139
Bifidobacterium animalis
Bifidobacterium longum
Prevotella copri
Escherichia coli
water
Inulin.0h
N.maize.0h
R.maize.0h
Potato.0h
Hylon.0h
Avicel.6h
Inulin.6h
N.maize.6h
R.maize.6h
Potato.6h
Hylon.6h
Avicel.12h
Inulin.12h
N.maize.12h
R.maize.12h
Potato.12h
Hylon.12h
Avicel.24h
Inulin.24h
N.maize.24h
R.maize.24h
Potato.24h
Hylon.24h
Substrate
R.maize
Hylon
Avicel
water
N.maize
Potato
Inulin
Time
0h
6h
12h
24h




N. genomes
400
300
200
100
0
Contamination (%)
10.0
7.5
5.0
2.5
0.0
60
70
80
90
100
Completeness (%)
0
50
100
150
200
250
300
350
400
450
N. genomes

A.

Candidatus Minthonaster sp.

Dysosmobacter sp.(4)

Candidatus Enteromorpha sp.(7).

Candidatus Colihabitans sp.

Agathobaculum butyriciproducens (6)

Candidatus Ruminococcus anthropi
Ruminococcus bromii

Gemmiger sp.(3)

Faecalibacterium prausnitzii (2)

Bacteroidaceae (6)

Butyricimonas faecihominis (2)

Alistipes sp. (4)

Alistipes sp.

Burkholderiaceae

Escherichia coli

Adlercreutzia celatus (7)

Gordonibacter urolithinfaciens

Bifidobacterium sp.(5)

Streptococcus thermophilus (3)

Fusicatenibacter saccharivorans (4)

Blautia sp.
Candidatus Blautia hennigii.
Blautia hydrogenotrophica.

Dorea sp. (2)

Roseburia sp. (3).

Lachnospira sp.

Coprococcus sp. (4).

B.

C.

avicell | hylon | inulin | pstarch | rmaize

cluster_56_1 cluster_57_1 cluster_58_1 cluster_55_1 cluster_59_1 cluster_60_1 cluster_49_1 cluster_49_2 cluster_50_0 cluster_48_1 cluster_47_1 cluster_51_1 cluster_118_1 cluster_54_4 cluster_54_3 cluster_54_1 cluster_54_2 cluster_52_3 cluster_52_2 cluster_52_1 cluster_62_1 cluster_61_1 cluster_119_2 cluster_119_3 cluster_119_1 cluster_119_4 cluster_119_5 cluster_119_6 cluster_68_2 cluster_68_1 cluster_37_2 cluster_37_1 cluster_117_1 cluster_40_1 cluster_53_1 cluster_11_0 cluster_84_1 cluster_41_0 cluster_39_1 cluster_115_1 cluster_112_1 cluster_67_1 cluster_66_1 cluster_65_1 cluster_64_1 cluster_63_1 cluster_46_0 cluster_116_1 cluster_69_1 cluster_73_0 cluster_109_0 cluster_44_1 cluster_45_0 cluster_42_1 cluster_70_0 cluster_15_1 cluster_17_1 cluster_18_1 cluster_16_0 cluster_13_1 cluster_14_1 cluster_10_1 cluster_19_1 cluster_23_1 cluster_23_2 cluster_22_1 cluster_20_2 cluster_20_1 cluster_12_1 cluster_7_1 cluster_8_1 cluster_6_1 cluster_9_0 cluster_21_1 cluster_4_1 cluster_5_0 cluster_3_1 cluster_2_1 cluster_26_1 cluster_43_1 cluster_1_1 cluster_114_7 cluster_114_4 cluster_114_6 cluster_114_2 cluster_114_5 cluster_114_1 cluster_114_3 cluster_113_2 cluster_113_1 cluster_35_1 cluster_29_1 cluster_34_1 cluster_33_1 cluster_32_1 cluster_31_1 cluster_30_1 cluster_88_1 cluster_89_1 cluster_24_1 cluster_38_1 cluster_120_0 cluster_36_2 cluster_36_1 cluster_36_3 cluster_28_0 cluster_27_1 cluster_25_1 cluster_25_2 cluster_97_4 cluster_97_1 cluster_97_3 cluster_97_2 cluster_102_1 cluster_103_1 cluster_96_1 cluster_111_0 cluster_99_0 cluster_98_1 cluster_79_1 cluster_80_1 cluster_77_1 cluster_86_1 cluster_85_1 cluster_75_1 cluster_81_0 cluster_106_0 cluster_107_0 cluster_105_1 cluster_101_1 cluster_100_1 cluster_104_1 cluster_108_0 cluster_78_1 cluster_90_1 cluster_72_1 cluster_91_1 cluster_92_1 cluster_110_0 cluster_87_0 cluster_74_1 cluster_82_1 cluster_93_3 cluster_93_4 cluster_93_2 cluster_93_1 cluster_83_0 cluster_95_1 cluster_94_1 cluster_76_1 cluster_71_0

T6 T12 T24

Actinobacteria
Bacilli
Bacteroidia
Clostridia
Coriobacteriia
Desulfovibrionia
Gammaproteobacteria
Negativicutes
Verrucomicrobiae

Log_2 foldchange ratio
−8 −6 −5 −4 −3 −2 −1 1 2 3 4 5 6