ARTICLES
https://doi.org/10.1038/s41564-022-01093-0
nature microbiology

Check for updates

Mechanistic insights into consumption of the food additive xanthan gum by the human gut microbiota

Matthew P. Ostrowski[1,13], Sabina Leanti La Rosa[2,3,13], Benoit J. Kunath[2], Andrew Robertson[4], Gabriel Pereira[1], Live H. Hagen[2], Neha J. Varghese[5], Ling Qiu[1], Tianming Yao[6], Gabrielle Flint[1], James Li[7], Sean P. McDonald[7], Duna Buttner[1], Nicholas A. Pudlo[1], Matthew K. Schnizlein[1], Vincent B. Young[1,8], Harry Brumer[7], Thomas M. Schmidt[1], Nicolas Terrapon[9,10], Vincent Lombard[9,10], Bernard Henrissat[11,12], Bruce Hamaker[6], Emiley A. Eloe-Fadrosh[5], Ashootosh Tripathi[4], Phillip B. Pope[2,3]✉ and Eric C. Martens[1]✉

Processed foods often include food additives such as xanthan gum, a complex polysaccharide with unique rheological properties, that has established widespread use as a stabilizer and thickening agent. Xanthan gum's chemical structure is distinct from those of host and dietary polysaccharides that are more commonly expected to transit the gastrointestinal tract, and little is known about its direct interaction with the gut microbiota, which plays a central role in digestion of other dietary fibre polysaccharides. Here we show that the ability to digest xanthan gum is common in human gut microbiomes from industrialized countries and appears contingent on a single uncultured bacterium in the family Ruminococcaceae. Our data reveal that this primary degrader cleaves the xanthan gum backbone before processing the released oligosaccharides using additional enzymes. Some individuals harbour *Bacteroides intestinalis* that is incapable of consuming polymeric xanthan gum but grows on oligosaccharide products generated by the Ruminococcaceae. Feeding xanthan gum to germfree mice colonized with a human microbiota containing the uncultured Ruminococcaceae supports the idea that the additive xanthan gum can drive expansion of the primary degrader Ruminococcaceae, along with exogenously introduced *B. intestinalis*. Our work demonstrates the existence of a potential xanthan gum food chain involving at least two members of different phyla of gut bacteria and provides an initial framework for understanding how widespread consumption of a recently introduced food additive influences human microbiomes.

Evidence is accumulating that food additives impact the symbiosis between humans and their associated gut microbiomes, in some cases promoting intestinal inflammation and metabolic syndrome[1] or promoting certain pathogens[2]. Often used as thickeners and emulsifiers, polysaccharides are a prominent subset of these food additives. Non-starch dietary polysaccharides typically transit the upper intestinal tract undigested and influence host health by altering the composition and function of the microbiome[3,4]. Given 'generally recognized as safe' (GRAS) approval in 1968, xanthan gum (XG) is an exopolysaccharide produced by *Xanthamonas campestris* that has been increasingly used in the food supply[5]. Unlike gum Arabic[6] and carrageenan[7], which are readily consumed by some gut bacteria and have either been in human food for over a thousand years or are similar to plant fibres, XG does not resemble any common dietary fibre[8] (Fig. 1a), suggesting that consuming it may create a novel nutrient niche. XG is typically added at concentrations of 0.05–0.5% to foods[9] and used as a replacement for gluten in a gluten-free diet, which is a vital component of managing coeliac disease[10]. XG can reach gram quantities per serving in gluten-free baked goods. Although small doses of XG have not been connected to immediate health impacts, its fate in the digestive tract is unknown[11]. The low-level but constant consumption of XG by much of the industrialized world and higher intake by specific subpopulations highlight the need to understand its effects on the human gut microbiota.

A human gut microbe degrades XG

To identify XG-degrading bacteria in the human gut microbiome, we surveyed 80 adults[12] using a bacterial culture strategy designed to enrich for Bacteroidetes (using the antibiotic gentamicin for selection), a phylum that generally harbours numerous polysaccharide-degrading enzymes[13]. On the basis of increased

[1]Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI, USA. [2]Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway. [3]Faculty of Biosciences, Norwegian University of Life Sciences, Ås, Norway. [4]Life Sciences Institute: Natural Products Discovery Core, University of Michigan, Ann Arbor, MI, USA. [5]DOE Joint Genome Institute, Berkeley, CA, USA. [6]Department of Food Science and Whistler Center for Carbohydrate Research, Purdue University, West Lafayette, IN, USA. [7]Michael Smith Laboratories, University of British Columbia, Vancouver, British Columbia, Canada. [8]Department of Internal Medicine, Infectious Diseases Division, University of Michigan Medical School, Ann Arbor, MI, USA. [9]Centre National de la Recherche Scientifique, Aix-Marseille Univ, Marseille, France. [10]Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement, Marseille, France. [11]Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia. [12]Technical University of Denmark, DTU Bioengineering, Lyngby, Denmark. [13]These authors contributed equally: Matthew P. Ostrowski, Sabina Leanti La Rosa. ✉e-mail: phil.pope@nmbu.no; emartens@umich.edu

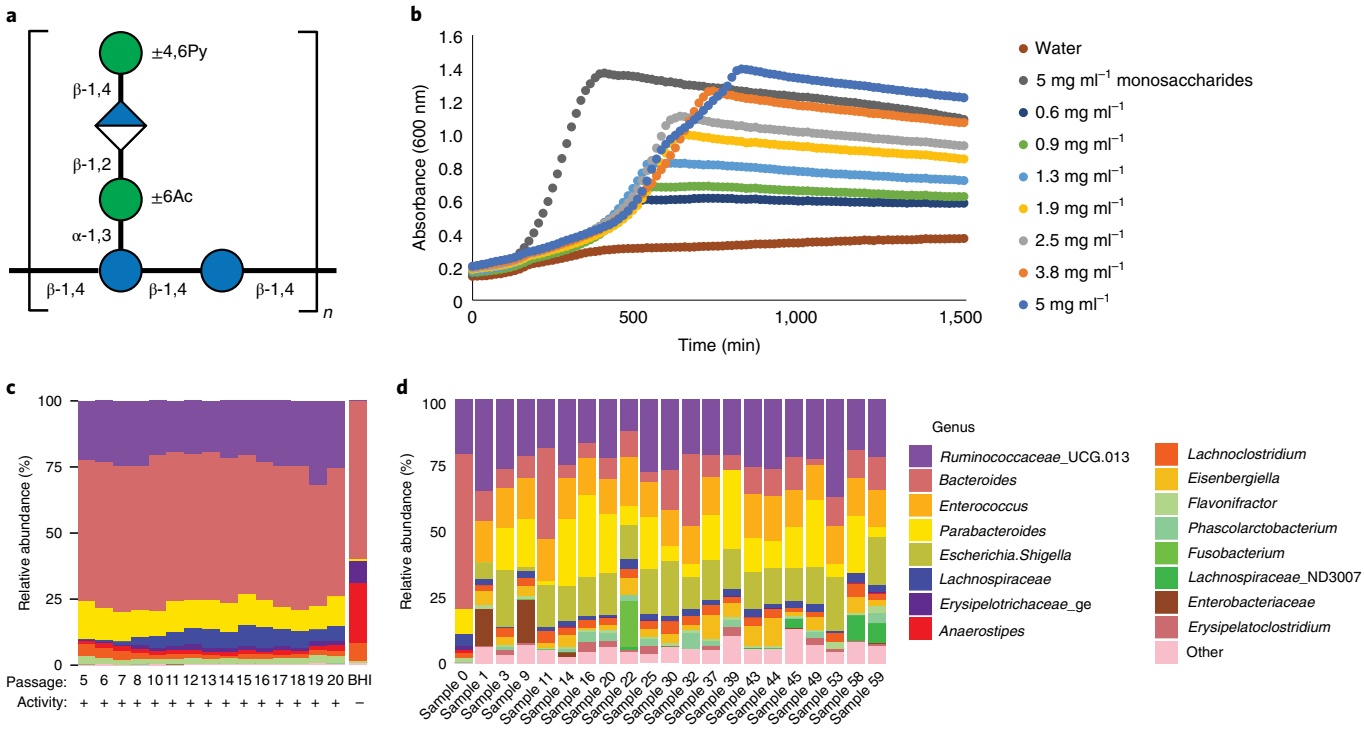


Fig. 1 | R.UCG13 was a common factor across xanthan gum degrading cultures. a, Xanthan gum has a β-1,4-linked backbone that is heavily decorated with trisaccharide branches on alternating glucose residues (blue circles) consisting of α-1,3-mannose (green circle), β-1,2-glucuronic acid (blue and white diamond) and terminal β-1,4-mannose (green circle). The terminal β-D-mannose and the inner α-D-mannose are variably pyruvylated at the 4,6-position or acetylated at the 6-position, respectively, with amounts determined by specific *X. campestris* strains and culture conditions. **b**,**c**, Growth curves (**b**) of the original xanthan-degrading culture (Sample 0) showed greater culture density as xanthan gum concentration was increased (points represent means of $n = 12$ biological replicates with s.e.m. ≤3% at every timepoint), and displayed relatively stable composition over sequential passaging (**c**). Passaging the culture on BHI-blood plates resulted in a loss of R.UCG13 as well as xanthan-degrading activity. **d**, An additional 20 samples were sequentially passaged in xanthan containing media (10×) and analysed for composition by 16S rRNA sequencing (16 of the most abundant genera are displayed for clarity). All cultures shared an abundant OTU, classified as R.UCG13.

bacterial culture turbidity and decreased viscosity of medium containing XG as the main carbon source, our initial survey yielded a single positive culture. Growth was dependent on the amount of XG provided, demonstrating nutrient specificity (Fig. 1b). Enrichment by sequential passaging yielded a mixed microbial culture with at least 12 distinct operational taxonomic units (OTUs; Fig. 1c), hereafter referred to as 'Sample 0'. Passaging on a commonly used anaerobic solid medium (brain–heart infusion (BHI) with 10% horse blood) resulted in the loss of two previously abundant OTUs (loss defined as <0.01% relative abundance), including a Ruminococcaceae from uncultured genus 13 (herein, R.UCG13)[14], and corresponded with loss of the XG-degrading phenotype upon re-inoculation into XG media (Fig. 1c).

Despite R.UCG13 and a *Bacteroides* OTU being present at >20% relative abundance, we repeatedly failed to isolate pure cultures that could degrade XG. Dilution of Sample 0 to extinction in medium supplemented with either XG or its component monosaccharides resulted in loss of growth on XG at more concentrated dilutions than on simple sugars (Extended Data Fig. 1). This suggests that the ability to degrade XG in the medium we employed requires multiple OTUs, which could be explained by multiple species being directly involved or the necessity of other species to promote sufficient growth conditions for the XG degrader[15].

We surveyed 60 new adults using the same media but without selective antibiotics and found that the ability to degrade XG is substantially more frequent, as approximately half of the cultures grew to appreciable levels on XG and decreased its characteristic viscosity. We passaged 20 positive samples 10 times each (one 1:100 dilution per day) and analysed the resulting community structure. While all cultures contained multiple OTUs (12–22 OTUs with relative abundance >0.5%) and commonalities at the genus level, the only OTU common across all cultures at >0.5% was R.UCG13 (Fig. 1d and Supplementary Table 1). These results suggest that a member of an uncultured *Ruminococcaceae* genus is necessary for XG degradation but may be unable to grow in isolation in our media.

Sequencing identifies two putative XG utilization loci

To identify XG-degrading genes within our bacterial consortium, we performed metagenomics and metatranscriptomics on Sample 0, using samples collected throughout growth (Extended Data Fig. 2a). We reconstructed 18 metagenome-assembled genomes (MAGs), 7 of which were of high quality (completion >90% and contamination <5%) (Supplementary Table 2). To connect 16S ribosomal RNA genes to MAGs, we performed long-read sequencing that yielded 2 MAGs with complete circular chromosomes including R.UCG13 (with 4 complete 16S rRNA operons, 3 of which were identical to the R.UCG13 OTU). This R.UCG13 genome is distantly related (47.26% average amino acid identity, AAI) to the recently cultured *Monoglobus pectinolyticus*[16]. Annotation of carbohydrate-active enzymes (CAZymes) in the R.UCG13 MAG revealed a locus encoding highly expressed enzymes that are candidates for XG degradation (Fig. 2 and Extended Data Fig. 2b). These included a polysaccharide lyase family 8 (PL8) with distant homology to known xanthan lyases from soil bacteria *Paenibacillus*

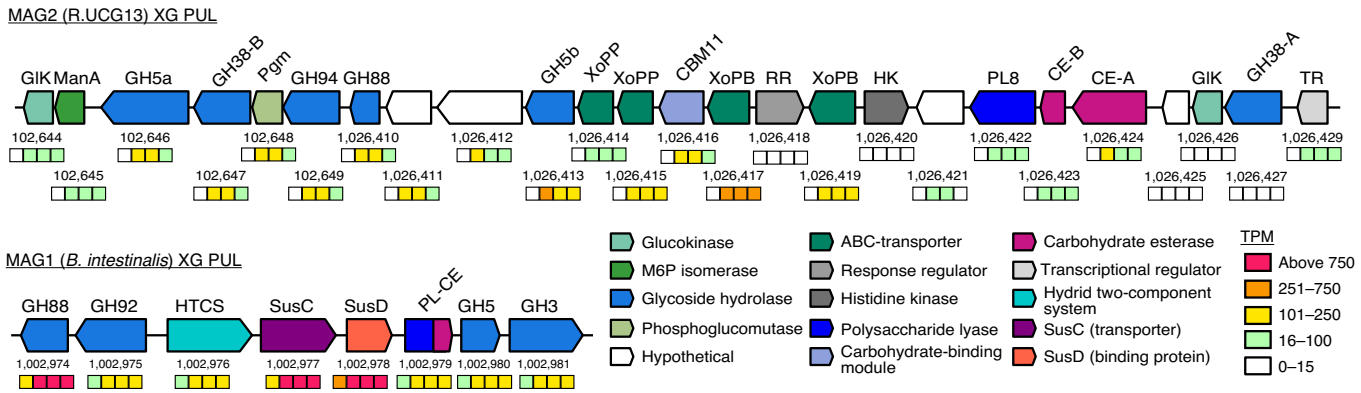


Fig. 2 | Metagenomics, metatranscriptomics and activity-guided proteomics identified two putative xanthan-gum-degrading loci. Putative xanthan utilization loci colour-coded and annotated by predicted protein family. Numbers below each gene show the locus tag in the corresponding genome. The four boxes below each gene are coloured to represent expression levels in transcripts per million (TPM) of each gene at timepoints taken throughout the culture's growth on xanthan gum. MAG taxonomy is indicated in parentheses.

alginolyticus XL-1[17] (36% identity/73% coverage) and *Bacillus* sp. GL1[18] (32% identity/81% coverage; Fig. 2). Xanthan lyases typically remove the terminal pyruvylated mannose before XG depolymerization[19,20]. The locus contained two GH5 enzymes with the potential to cleave the XG backbone[21], a GH88 to remove the unsaturated glucuronic acid residue produced by the PL8[22], and two GH38s to cleave the α-D-mannose[23]. Two carbohydrate esterases (CEs) could potentially remove the acetylation from the mannose[24].

Co-localization of genes that saccharify a polysaccharide into discrete polysaccharide utilization loci (PULs) is common in Bacteroidetes[25]. Although not present in most cultures, we obtained a MAG affiliated to *Bacteroides intestinalis*, which was the most abundant OTU in Sample 0 (Fig. 1). This MAG contained a putative PUL expressed during growth on XG (Fig. 2 and Extended Data Fig. 2b) with a putative polysaccharide lyase (PL) remotely related to alginate lyases[26,27], as well as several other GHs that could cleave the glycosidic bonds in XG. Despite extensive work to characterize the substrate specificity of PULs demonstrated by hundreds of characterized and predicted PULs in the PUL database (PUL-DB)[28], this database only harboured a single partially related homologue of the *B. intestinalis* PUL (*B. salyersiae* WAL 10018 PUL genes HMPREF1532_01924-HMPREF1532_01938).

Neutral monosaccharide analysis of Sample 0 over time showed a relatively stable 1:1 ratio of glucose:mannose in residual polysaccharide, implying that lyase-digested XG was not accumulating as growth progressed (Supplementary Table 3). This could be due to fast depolymerization and cellular importation of the XG polymer following lyase removal of the terminal mannose or XG depolymerization before subsequent hydrolysis[29,30].

R.UCG13 produces an endo-acting xanthanase

To investigate the cellular location of the enzymes responsible for xanthan degradation, Sample 0 was separated into cell-free supernatant, washed cells that were resuspended or lysed, or lysed cells with supernatant. Cell-free supernatant depolymerized XG into large oligosaccharides, while the intracellular fraction was required to further saccharify these products (Extended Data Fig. 3). Liquid chromatography–mass spectrometry (LC–MS) analysis of cell-free supernatant incubated with XG revealed pentameric and decameric oligosaccharides (Fig. 3a), supporting the model in which endo-cleaving enzymes depolymerize XG before further saccharification.

To identify XG-degrading enzymes, we collected protein fractions from cell-free supernatants and fractionated them (defined in Methods), sequentially collecting and pooling xanthan-degrading fractions. Proteomic analysis of three independent experiments, each using a different activity-guided fractionation strategy, yielded 33 proteins common across all experiments (Supplementary Table 4), notably including *Ru*GH5a from R.UCG13.

*Ru*GH5a consists of an N-terminal signal peptide sequence, its main catalytic domain, and 3 uncharacterized domains (UD) that were not assigned in our CAZyme annotation but were annotated as sugar-binding proteins or carbohydrate-binding modules (CBMs, which are often associated with CAZymes and can facilitate polysaccharide degradation) by the European Bioinformatics Institute hmmscan tool (Fig. 3b)[31]. The protein also contains *Listeria-Bacteroides* repeat domains (pfam[32] PF09479), a β-grasp domain originally characterized from the invasion protein InlB used by *Listeria monocytogenes* for host cell entry[33,34]. To test the activity of *Ru*GH5a on XG, we expressed recombinant forms of the protein (*Ru*GH5a full), the GH5 domain only (*Ru*GH5a GH5 only), and the GH5 domain with either one (*Ru*GH5a GH5 + UD-A), two (*Ru*GH5a GH5 + UD-A/B), or three of the UDs (*Ru*GH5a GH5 + UD-A/B/C, hereafter referred to as *Ru*GH5a for simplicity). Only constructs with the GH5 and all three UDs showed activity on XG, suggesting a critical role in enzyme stability, conformation, or catalysis for these domains (Extended Data Fig. 4b). The alternate GH5 (*Ru*GH5b) was also expressed in several forms but lacked detectable activity (Extended Data Fig. 4).

*Ru*GH5a reactions produced pentasaccharide repeating units of XG, with various acetylation and pyruvylation (including di-acetylation[35]). While isolation of homogenous pentameric oligosaccharides proved difficult, coincubation of XG with *Ru*GH5a and a *Bacillus* sp. PL8 facilitated isolation of tetrasaccharide (Supplementary Figs. 1–10), which was useful in determining the regiospecificity of *Ru*GH5a hydrolysis that occurs at the reducing end of the non-branching glucose (Fig. 3c and Supplementary Figs. 1–10). This is in contrast to characterized xanthanases (for example, *Paenibacillus nanensis* GH9[36] or *Bacillus* β-D-glucanase[18]), which hydrolyse XG at the reducing end of the branching glucose. *Ru*GH5a displayed little activity on other polysaccharides (Extended Data Fig. 4b–g) and hydrolysed native and lyase-treated XG with comparable specificity, in contrast to most previously known xanthanases, which show ≥600-fold preference for the lyase-treated substrate[36] (Fig. 3d). One exception is the xanthanase from *Microbacterium* sp. XT11, which also cleaves native and lyase-treated xanthan gum with similar kinetic specificity; however, this enzyme only produces intermediate XG fragments from the complete polysaccharide, whereas RuGH5a can cleave XG down to its repeating pentasaccharide unit.

R.UCG13 encodes enzymes required for XG saccharification

To validate other enzyme activities in the R.UCG13 locus, we set up reactions with *Ru*GH5a produced XG oligosaccharides (XGOs; Extended Data Fig. 5a,b, R0) and combinations of enzymes. In contrast to characterized xanthan lyases[17,20,37], R.UCG13 PL8 showed no activity on polymeric XG but was active on XGOs (Extended Data Fig. 5c,d, R3). Both R.UCG13 CEs removed acetyl groups from XGOs (Extended Data Fig. 5a,b, R1, R2). The tetrasaccharide produced by PL8 was processed by GH88 (Extended Data Fig. 5e,f, R6) and both GH38s were able to hydrolyse the resulting trisaccharide (Extended Data Fig. 5g, R7, R8). GH94 catalysed phosphorolysis of cellobiose, completing saccharification of XG (Extended Data Fig. 5l). Additional support for the involvement of this locus in XG degradation is provided by RNA-seq analysis, which showed the induction of genes in this cluster when the community was grown on XG compared with another polysaccharide (polygalacturonic acid, PGA) that supports R.UCG13 abundance (Extended Data Fig. 6a).

Human gut *Bacteroides* cross-feed on XG oligosaccharides

Although R.UCG13 was recalcitrant to culturing, we isolated several bacteria from Sample 0, including the abundant *B. intestinalis* (Fig. 1c). While unable to grow on XG, it showed robust growth on XGOs (Fig. 4a). Isolates of *P. distasonis* and *B. clarus* showed little or no growth (Extended Data Fig. 6b,c). All genes in the *B. intestinalis* PUL were activated >100-fold during growth on XGOs compared with a glucose-grown reference (Fig. 4b). RNA-seq of the *B. intestinalis* strain grown on XGOs revealed that the identified PUL was the most highly upregulated in the genome (Extended Data Fig. 6g). Interestingly, *Ru*GH5a XGOs treated with PL8 continued to support *B. intestinalis* growth, but an isomeric tetramer generated from the *P. nanensis* GH9 and PL8 failed to support any growth (Extended Data Fig. 6d). Growth was rescued by glucose but not *Ru*GH5a XGOs (Extended Data Fig. 6d), suggesting that the transporters or downstream enzymes are incapable of processing this substrate.

We also tested the *B. salyersiae* strain with a PUL homologous to the one identified in the *B. intestinalis* isolate. Although it did not grow on XGO pentasaccharides, it did grow on lyase-produced tetrasaccharides and showed substantial upregulation of its PUL relative to glucose-grown cultures (Extended Data Fig. 6e,f).

To further test the role of the *B. intestinalis* PUL in XGO degradation, we tested recombinant forms of its enzymes and confirmed several activities. The CE domain C terminal to the PL-CE protein removed acetyl groups from XGOs (Extended Data Fig. 5a,b, B2). While we were unable to detect lyase activity for the PL-CE enzyme on XG or XGOs, the GH88 was able to remove the corresponding 4,5 unsaturated glucuronic acid residue from tetrasaccharide generated by a lyase from another organism (Extended Data Fig. 5f, B4). The GH92 hydrolysed trisaccharide produced by the GH88 (Extended Data Fig. 5f,g, B5). Finally, the GH3 was active on cellobiose (Extended Data Fig. 5l). To test whether the *B. intestinalis* contained a functional lyase, we grew cultures on XGOs, fractionated the culture and incubated fractions with fresh XGOs. Pelleted and lysed cells showed an increase in lyase-produced tetrasaccharide (Extended Data Fig. 7), verifying that *B. intestinalis* has a

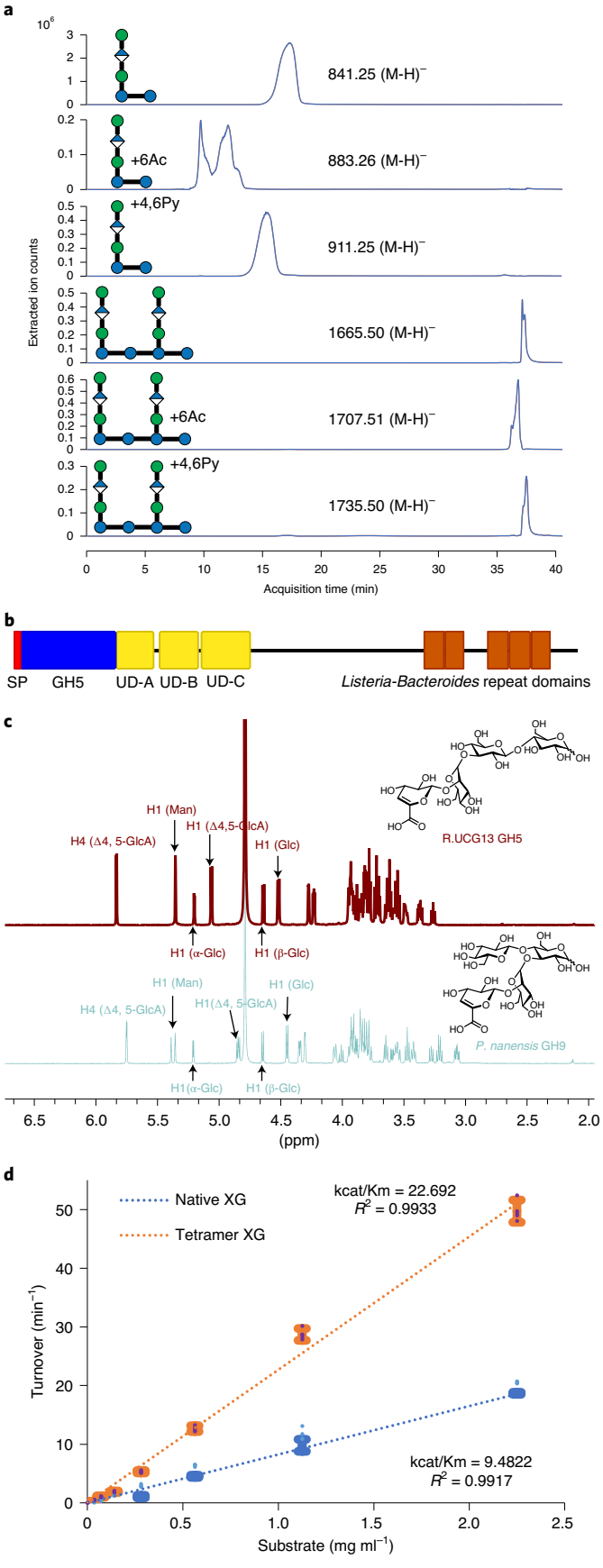


Fig. 3 | R.UCG13 encodes a GH5 that depolymerizes native xanthan gum. **a**, Representative extracted ion chromatograms showing various acetylated and pyruvylated penta- and decasaccharides produced by incubating sample 0 culture supernatant with XG. Repeated incubations showed similar results. **b**, Annotated domains of the xanthan-degrading *Ru*GH5a, showing its signal peptide (SP), three uncharacterized domains (UDs) and multiple *Listeria-Bacteroides* repeat domains. UD-A and -C yielded Pfam predictions of CBM family 11, while all three UDs were predicted as galactose-binding lectins by Gene3D. **c**, Proton NMR contrasting tetrasaccharide products obtained from incubating lyase-treated xanthan gum with either *Ru*GH5a or *P. nanensis* GH9. **d**, Kinetics of *Ru*GH5a on native and lyase-treated xanthan gum (mean ± s.d.; $n = 4$ distinct samples).

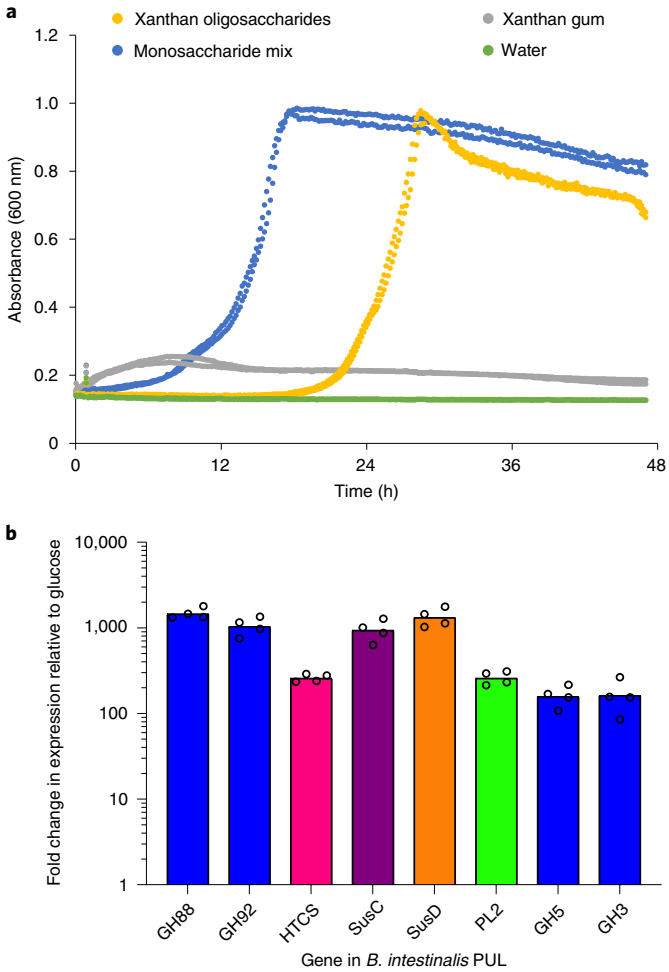


Fig. 4 | *B. intestinalis* cross-feeds on xanthan oligosaccharides. a, Growth curves of *B. intestinalis* isolated from the Sample 0 culture on various substrates at 5 mg ml^{-1} ($n = 2$ biological replicates). **b**, Fold change in expression of *B. intestinalis* genes when grown on xanthan oligosaccharides relative to glucose (mean of $n = 4$ biological replicates, with individual replicates shown as open circles).

cell-associated enzyme(s) that carries out the lyase reaction on XGOs. Despite extensive testing of the candidate PL-CE enzyme, we were unable to detect lyase activity with this construct, suggesting that there may be one or more genes outside of the PUL that provide(s) this function.

Metagenomics suggests additional cross-feeding modalities

To determine if the Sample 0 consortium was representative of all XG-degrading cultures, we performed metagenomic sequencing on 20 additional XG-degrading communities and retrieved 16 high-quality and 3 low-quality R.UCG13 MAGs, as well as an unbinned contig affiliated with R.UCG13 (Supplementary Table 2). The R.UCG13 XG locus is extremely well conserved across these cultures with one variation in gene content—the insertion of a GH125 coding gene—and >95% amino acid identity (Extended Data Fig. 8a). The GH125 gene was observed in most loci (14/17), suggesting that this gene provides a complementary but non-essential function[38]. Only a subset of samples (4/17) contained the *B. intestinalis* XGOs PUL, which showed essentially complete conservation (Extended Data Fig. 8b). Additional microbes may cross-feed on products released by R.UCG13; for example, we found that samples 1, 22 and 59 contained microbes with a GH88, GH92 and GH3, suggesting that they may metabolize XG-derived tetramers (Extended Data Fig. 8c).

Xanthan utilization is widespread in modern microbiomes

Next, we asked whether XG inclusion in the modern diet could have increased the prevalence of the R.UCG13 and *B. intestinalis* xanthan loci. All industrialized populations sampled had the R.UCG13 locus in their faecal metagenomes (Fig. 5). The *B. intestinalis* locus was less prevalent, with two industrialized datasets (Japan and Denmark/Spain) lacking any incidence. The non-industrialized populations sampled (Yanomami, Hadza and Burkina Faso) had no detected presence of either locus.

The size of the non-industrialized datasets is relatively small, which could mean that R.UCG13 and *B. intestinalis* might be present in these microbiomes but were undetected because of methodological limitations. Excluding the possibility of a false negative, our results suggest that inclusion of XG in the modern diet may have driven the expansion of R.UCG13 (and to a lesser extent *B. intestinalis*) in gut communities of numerous human populations. This is in concordance with the observation that a set of volunteers fed XG for an extended period produced stool with increased probability and degree of xanthan degradation[39]. Alternatively, the modern microbiome is drastically different from that of hunter-gatherers and these differences may simply correlate with the abundance of R.UCG13, rather than any causal effect of dietary XG. Microbiomes of hunter-gatherer populations may also degrade XG utilizing distinct microbes and pathways.

To probe for the presence of the identified XG utilization genes in other environments and possibly determine their origin, we conducted an expanded LAST search[40] in 72,491 sequenced bacterial isolates, 102,860 genome bins extracted from 13,415 metagenomes, and 21,762 metagenomes that are part of the Integrated Microbial Genomes and Microbiomes[41] (IMG/M) database using fairly stringent thresholds of 70% alignment over the query and 90% nucleotide identity. This search yielded 35 hits of the R.UCG13 locus in human microbiome datasets, including senior adults, children and an infant (12 months old, Ga0169237_00111) (Supplementary Table 5). We also found 12 hits for the *B. intestinalis* XGO locus, all in human microbiome samples, except for a single environmental sample from a fracking water sample from Oklahoma, USA (81% coverage, 99% identity) (Extended Data Fig. 8f and Supplementary Table 5). XG and other polysaccharides are used in oil industry processes, and genes for guar gum catabolism have previously been found in oil well microbial communities[42]. Since most samples searched were non-gut-derived and most of our positive hits were human gut samples, this suggests that XG-degrading R.UCG13 and XGO-degrading *B. intestinalis* loci may be largely confined to the human gut microbiome.

The mouse microbiome harbours a xanthan utilization locus

To investigate the prevalence of XG-degrading populations beyond the human gut microbiome, we used samples from a previous mouse experiment in which animals were fed XG[43]. Culturing mouse faeces from this experiment in XG-containing liquid media resulted in loss of viscosity and increased turbidity over multiple passages, confirming these communities' ability to grow and degrade XG. We used metagenomics to characterize two samples (M1741 and M737), revealing a microbe related to R.UCG13 (AAI values between the human R.UCG13 and the mouse R.UCG13 were 75.7% and 75.2% for M1741 and M737, respectively) and a locus with strikingly similar genetic architecture to our previously characterized human XG locus (Extended Data Fig. 8d and Supplementary Table 6). Although several genes were conserved across both the human and mouse loci, the respective *Ru*GH5a proteins diverged

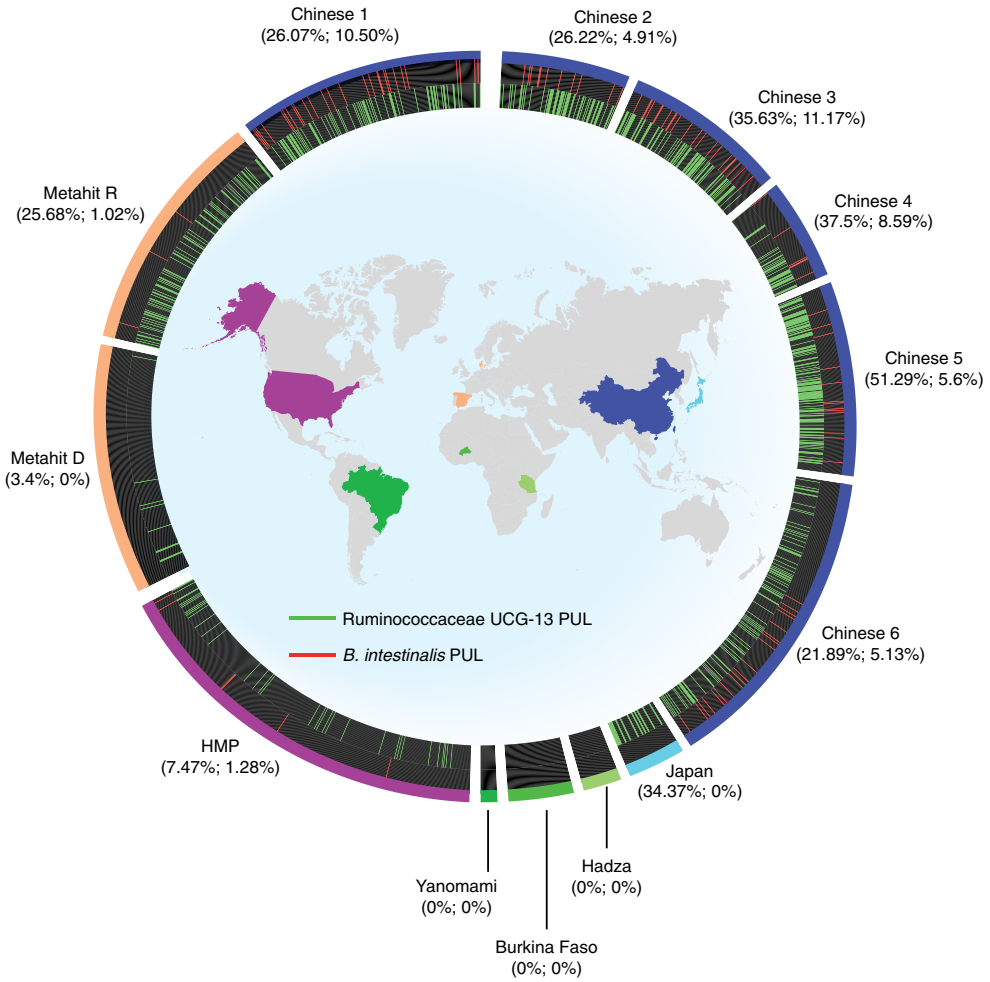


Fig. 5 | Xanthan-degrading loci are present in modern human microbiomes but not in hunter-gatherers. Multiple microbiome metagenome datasets were searched for the presence or absence of the R.UCG13 and *B. intestinalis* xanthan loci. Map colours correspond to where populations were sampled for each dataset displayed on the outside of the figure. Circle segments are sized proportionately to total number of individuals sampled for each dataset. Lines represent presence of either the R.UCG13 xanthan locus (green) or the *B. intestinalis* xanthan locus (red). Percentages display the total abundance of R.UCG13 or *B. intestinalis* locus in each dataset.

substantially (Extended Data Fig. 8e). Recombinant versions of the mouse *Ru*GH5a hydrolysed XG (Extended Data Fig. 8h). These data suggest that the R.UCG13 locus is more broadly present in mammalian gastrointestinal microbiomes.

XG drives colonization of a humanized mouse microbiome

To test whether XG drives expansion of R.UCG13 and *B. intestinalis* in microbiomes, we colonized germfree mice with a complex human microbiome that contained R.UCG13 but lacked the XGO-consuming *B. intestinalis*. Two cages of mice were maintained on an XG diet during initial gavages with the same inoculum, then Group 2 was switched to a nearly identical diet without XG for ~3 weeks before reintroducing XG (Fig. 6a). Mice were colonized with ~100 OTUs and total number of OTUs did not vary significantly between groups throughout the experiment (Supplementary Table 7). Although essentially undetectable by 16S rRNA community analysis performed at the end of the experiment, anaerobic culture of faecal samples in XG media revealed that a subject in Group 1 had an XG-degrading community as early as day 13 (Fig. 6c). This early colonization appears to have resulted in differential R.UCG13 colonization at day 25 despite identical treatments, possibly due to cage effects and the slow rate of R.UCG13 expansion even with constant XG support (Fig. 6c). Indeed, several other genus features diverged at day 25 and were consistently different between groups for the remainder of the experiment (for example, *Lachnoclostridium*, *Hungatella*, *Blautia* and several low-abundance features such as *Tyzzerella*, *Eisenbergiella*, *Butyricimonas* and *Ruminiclostridium 5*); notably, *Terrisporobacter* reached 4–6% relative abundance in Group 1, but was essentially absent in Group 2 mice (Fig. 6b and Supplementary Table 7). Although the abundance of R.UCG13 diverged before the diet switch, half of Group 2 subjects had XG-degrading communities at day 25, suggesting successful R.UCG13 colonization, albeit at very low levels (Fig. 6c). The data show that R.UCG13 increased to substantial levels in a complex microbiome and that this was dependent on dietary XG.

We subsequently gavaged an XGO-consuming *B. intestinalis* strain into all mice to test whether it could colonize with support from R.UCG13 and dietary XG. Interestingly, the donor microbiome we selected contained a native *Bacteroides* affiliated to *B. intestinalis*, which presumably lacked the XGO locus because the donor sample was PCR negative for genes in that PUL, and this strain colonized well before the gavage (~3–5% relative abundance). After the gavage, overall *B. intestinalis* levels were similar between groups, but qPCR specific primers for the *B. intestinalis* XGO PUL revealed that essentially all the endogenous *B. intestinalis* was displaced by

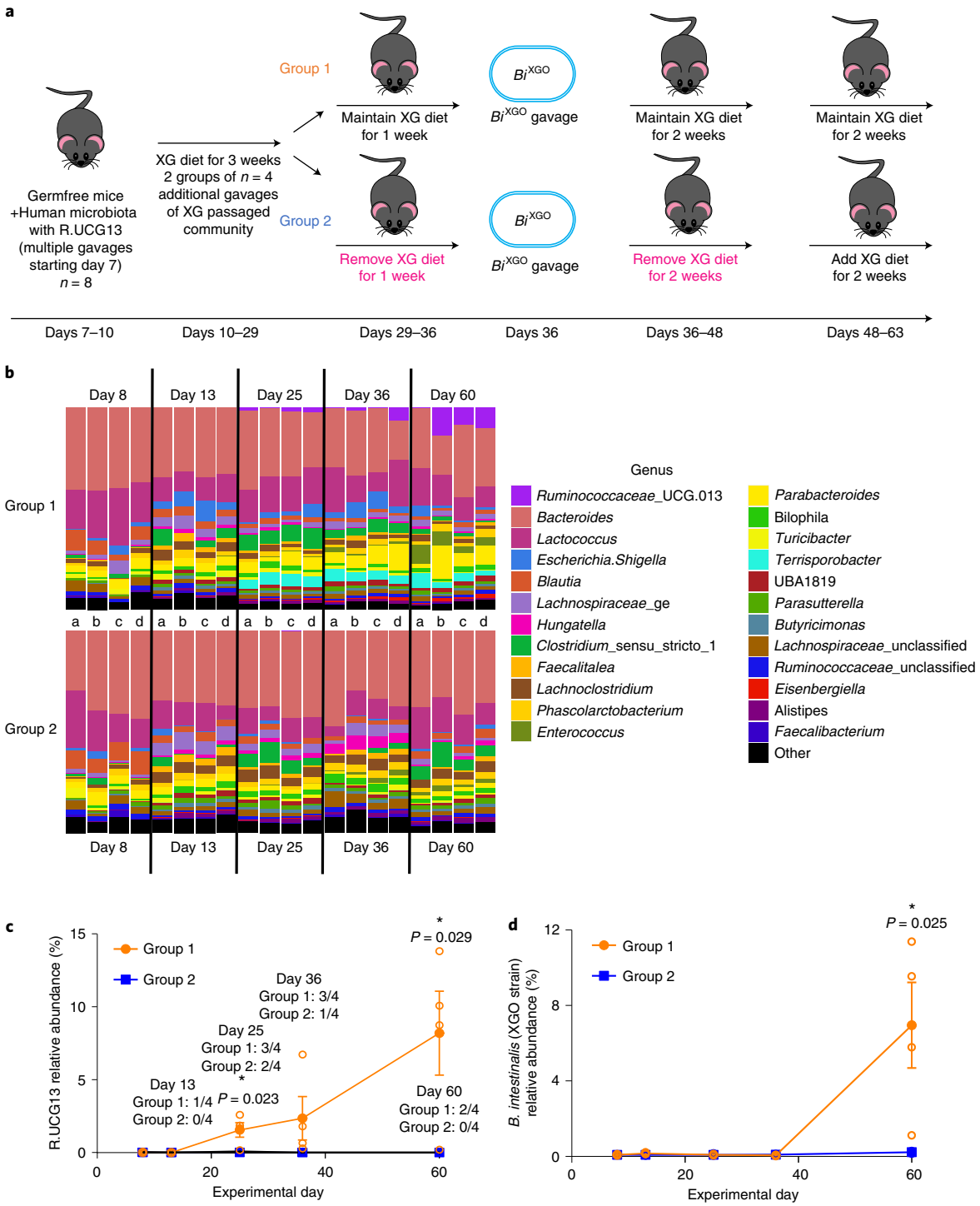


Fig. 6 | Dietary XG supports the expansion of R.UCG13 and *B. intestinalis* in vivo. a, Schematic of the experiment with timing of gavages and diet changes (see Methods for additional description). **b**, Genus-level features of each subject's faecal microbiome community (any genus present at >2% relative abundance is shown). Letters identify individual mouse subjects sampled at multiple timepoints throughout the experiment. **c**, Relative abundance of R.UCG13 over time. Text above datapoints shows how many of each group's faecal samples were positive for XG growth and degradation after inoculation into anaerobic XG media. **d**, Relative abundance of the XGO-consuming *B. intestinalis* over time as measured by qPCR of the XGO locus. Mean ± s.e.m.; $n = 4$ biological replicates (open circles are individual data points). Unpaired two-sided *t*-tests, $^{*}P < 0.05$.

the XGO-consuming *B. intestinalis* strain in the XG-group (Fig. 6d). This evidence supports the idea that XG can promote the colonization and expansion of R.UCG13 and *B. intestinalis*, with *B. intestinalis* colonization likely dependent on the presence of R.UCG13 to release XGOs for its consumption.

Discussion

Our results demonstrate a multi-phylum food chain for XG that can support R.UCG13 and *B. intestinalis* expansion (Extended Data Fig. 9). Population-wide consumption of XG may support higher levels of these microbes in industrialized human microbiomes. The

absence of these XG degraders in pre-industrialized microbiomes and their variable presence across post-industrialized populations suggest that XG-driven modulation of human microbiomes may be an ongoing process, although human studies will be required to definitively test this hypothesis. The wide range in levels of XG consumption, variable presence of XG-degrading microbes across human populations, and our finding that R.UCG13 can colonize infants highlight the impacts that XG may be having on the assembly, stability and evolution of industrialized human microbiomes.

The discovery of XG loci in an environmental sample and in mouse microbiome raises ecological questions about the transfer and evolution of XG utilization. Mice are affiliated with humans as pests, representing one route of XG exposure, and mammalian microbiomes could be exposed through consumption of *X. campestris*-infected plants or domesticated animal foodstuffs (for example, in calf milk replacers[44]). Another hypothesis is that the R.UCG13 XG locus evolved to target exopolysaccharides produced by endogenous gut microbes; this raises the possibility that dietary XG did not create a novel nutrient niche and only amplified one that already existed. Thus, R.UCG13 may be a low-abundance but native microbe in numerous microbiomes and XG can support its expansion but may not be necessary for its colonization.

While many questions remain about the ecological, functional and health-relevant impacts of XG on the human microbiome, our study demonstrates that the abrupt introduction of a new dietary component may drive changes in microbiome ecology with potentially broad impacts. Our work also demonstrates a workflow using enrichment cultures in combination with multiple analytical methods to understand the biology of microbes that consume novel substrates, yet resist isolation attempts.

Methods

Human subjects. For screening of faecal samples for xanthan gum degradation, study participants were recruited through the Authentic Research Sections of the introductory biology laboratory course at the University of Michigan (BIO173) as previously described[12]. Briefly, individuals who had taken antibiotics within the last 6 months or with self-reported history of inflammatory bowel syndrome, inflammatory bowel disease or colorectal cancer were excluded from the study. All participants gave written informed consent before participating in the study. Participants under the age of 18 were granted permission by a parent or legal guardian. Participants ranged in age from 17 to 29 years old, with a median age of 19 years old. This study was approved by the Institutional Review Board of the University of Michigan Medical School (HUM00094242 and HUM00118951) and was conducted in compliance with the Helsinki Declaration.

For the humanized mouse experiment, new faecal samples were screened for degradation of XG and absence of an XGO-consuming *B. intestinalis* strain by PCR. Human subjects volunteered to donate their faecal matter and verbally consented to study participation. Participants had no history of gastrointestinal diseases and no history of antibiotic treatment in the previous 6 months. Participants ranged in age from 19 to 38 years old, with a median age of 30 years old. This study was approved by the Institutional Review Board of the University of Michigan Medical School (HUM00141992) and was conducted in compliance with the Helsinki Declaration.

Isolation, culture and phylogenetic analysis of xanthan-degrading cultures. Faecal samples were collected into pre-reduced phosphate buffered saline, then transferred to a 37 °C anaerobic chamber (10% H_2, 5% CO_2 and 85% N_2; Coy Manufacturing, Grass Lake, MI) within 24 h. Faecal suspensions were used to inoculate cultures and passaged using partially Defined Medium (DM), which was generally prepared as a 2x stock, then mixed 1:1 with 10 mg ml^{-1} carbon source (for example, xanthan gum). In the first survey of 80 healthy adults, media contained 200 µg ml^{-1} gentamicin to enrich for members of the Bacteroidetes phylum. After isolation of the original XG-degrading culture and for all other surveys and cultures, gentamicin was excluded. Samples were passaged by serial dilution (1:100 dilution into fresh media) and monitored for bacterial growth by increased turbidity or loss of the gel-like viscosity that is characteristic of XG in solution.

Each litre of prepared DM medium (pH 7.2) contained 13.6 g KH_2PO_4 (Fisher, P284), 0.875 g NaCl (Sigma, S7653), 1.125 g $(NH_4)_2SO_4$ (Thermo Fisher, A702), 2 mg each of adenine, guanine, thymine, cytosine and uracil (Sigma, A2786, G11950, T0895, C3506, U1128, prepared together as 100x solution), 2 mg of each of the 20 essential amino acids (prepared together as 100x solution), 1 mg vitamin K3 (menadione, Sigma M5625), 0.4 mg $FeSO_4$ (Sigma, 215422), 9.5 mg $MgCl_2$ (Sigma, M8266), 8 mg $CaCl_2$ (Sigma, C1016), 5 µg Vitamin B12 (Sigma, V2876), 1 g L-cysteine, 1.2 mg haematin with 31 mg histidine (prepared together as 1,000x solution), 1 ml Balch's vitamins, 1 ml trace mineral solution and 2.5 g beef extract (Sigma, B4888).

Each litre of Balch's vitamins was prepared with 5 mg *p*-Aminobenzoic acid, 2 mg folic acid (Sigma, F7876), 2 mg biotin (Sigma, B4501), 5 mg nicotinic acid (Sigma, N4126), 5 mg calcium pantothenate (Sigma, P2250), 5 mg riboflavin (Sigma, R7649), 5 mg thiamine HCl (Sigma, T4625), 10 mg pyridoxine HCl, 0.1 mg cyanocobalamin and 5 mg thioctic acid. Prepared Balch's vitamins was adjusted to pH 7.0, filter-sterilized with 0.22 µm PES filters and stored in the dark at 4 °C.

Each litre of trace mineral solution was prepared with 0.5 g EDTA (Sigma, ED4SS), 3 g $MgSO_4{\cdot}7H_2O$, 0.5 g $MnSO_4{\cdot}H_2O$, 1 g NaCl (Sigma, S7653), 0.1 g $FeSO_4{\cdot}7H_2O$ (Sigma, 215422), 0.1 g $CaCl_2$, 0.1 g $ZnSO_4{\cdot}7H_2O$, 0.01 g $CuSO_4{\cdot}5H_2O$, 0.01 g H_3BO_3 (Sigma, B6768), 0.01 g $Na_2MoO_4{\cdot}2H_2O$, 0.02 g $NiCl_2{\cdot}6H_2O$. Prepared trace mineral solution was adjusted to pH 7.0, filter-sterilized with 0.22 µm PES filters and stored at room temperature.

Samples that showed growth on xanthan gum, as evidenced by loss of viscosity and increased culture density, were subcultured 10 times by diluting an active culture 1:100 into fresh DM-XG medium. For the original culture (that yielded Sample 0), multiple samples were stored for genomic DNA (gDNA) extraction and analysis over the course of passaging; for the second sample set, samples were stored after 10 passages. Samples were collected by centrifugation, decanted and stored at −20 °C until further processing.

Frozen cell pellets were resuspended in 500 µl buffer A (200 mM NaCl, 200 mM Tris-HCl, 20 mM EDTA) and combined with 210 µl SDS (20% w/v, filter-sterilized), 500 µl phenol:chloroform (alkaline pH) and ~250 µl acid-washed glass beads (212–300 µm; Sigma). Samples were bead-beaten on high for 2–3 min with a Mini-BeadBeater-16 (Biospec Products), then centrifuged at 18,000 *g* for 5 min. The aqueous phase was recovered and mixed by inversion with 500 µl phenol:chloroform, centrifuged at 18,000 *g* for 3 min, and the aqueous phase was recovered again. The sample was mixed with 500 µl chloroform, centrifuged at 18,000 *g* for 3 min, and then the aqueous phase was recovered and mixed with 0.1 volume 3 M sodium acetate (pH 5.2) and 1 volume isopropanol. The sample was stored at −80 °C for ≥30 min, then centrifuged at ≥20,000 *g* for 20 min at 4 °C. The pellet was washed with 1 ml room-temperature 70% ethanol, centrifuged at 18,000 *g* for 3 min, decanted and allowed to air dry before resuspension in 100 µl sterile water. Resulting samples were additionally purified using the DNeasy blood and tissue kit (QIAGEN).

Illumina sequencing, including PCR and library preparation, were performed by the University of Michigan Microbiome Core as described by Kozich et al.[45]. Barcoded dual-index primers specific to the 16S rRNA V4 region were used to amplify the DNA. PCR reactions consisted of 5 µl 4 µM equimolar primer set, 0.15 µl AccuPrime *Taq* DNA high-fidelity polymerase, 2 µl 10x AccuPrime PCR buffer II (Thermo Fisher, 12346094), 11.85 µl PCR-grade water and 1 µl DNA template. The PCR conditions used consisted of 2 min at 95 °C, followed by 30 cycles of 95 °C for 20 s, 55 °C for 15 s and 72 °C for 5 min, followed by 72 °C for 10 min. Each reaction was normalized using the SequalPrep normalization plate kit (Thermo Fisher, A1051001), then pooled and quantified using the Kapa Biosystems Library qPCR MasterMix (ROX Low) quantification kit for Illumina platforms (KK4873). After confirming the size of the amplicon library using an Agilent Bioanalyzer and a high-sensitive DNA analysis kit (5067-4626), the amplicon library was sequenced on an Ilumina MiSeq platform using the 500 cycle MiSeq V2 reagent kit (MS-102-2003) according to the manufacturer's instructions with modifications of the primer set, with custom read 1/read 2 and index primers added to the reagent cartridge. The 'Preparing libraries for sequencing on the MiSeq' (part 15039740, Rev. D) protocol was used to prepare libraries with a final load concentration of 5.5 pM, spiked with 15% PhiX to create diversity within the run.

Sequencing FASTQ files were analysed with mothur (v.1.40.5)[46] using the Silva reference database[14]. OTUs with the same genus were combined and displayed using R[47] with the packages reshape2 v1.4.4[48], RColorBrewer v1.1-2[49] and ggplot2 v3.3.0[50].

Dilution to extinction experiment. An overnight culture was serially diluted in 2x DM. Serial dilutions were split into two 50 ml tubes and mixed 1:1 with either 10 mg ml^{-1} xanthan gum or 10 mg ml^{-1} monosaccharide mixture (4 mg ml^{-1} glucose, 4 mg ml^{-1} mannose, 2 mg ml^{-1} sodium glucuronate), both of which also had 1 mg ml^{-1} L-cysteine. Each dilution and carbon source was aliquoted to fill a full 96-well culture plate (Costar 3370) with 200 µl per well. Plates were sealed with Breathe-Easy gas-permeable sealing membrane for microtitre plates (Diversified Biotech, BEM-1). Microbial growth was measured for at least 60 h by monitoring OD_{600} using a Synergy HT plate reader (Biotek Instruments) and BIOSTACK2WR plate handler (Biotek Instruments)[51].

Maximum optical density (OD) for each substrate was measured for each culture. Full growth on substrates was conservatively defined as a maximum OD_{600} of >0.7. For each unique 96-well plate of substrate and dilution factor, the fraction of wells exhibiting full growth was calculated. Fractional growth was plotted against the dilution factor for each substrate. Data were fit to the Hill equation by minimizing squared differences between the model and experimental values using Solver (GRG nonlinear) in Excel. For each experiment, a 50% growth dilution factor (GDF 50) was calculated for each substrate at which half of the wells would be predicted to exhibit full growth.

Metagenomic analysis. Seven samples (15 ml) were collected at four timepoints (Extended Data Fig. 2; referred to as T1, T2, T3 and T4) during growth of two biological replicates of the Sample 0 culture. Cells were collected by centrifugation at 14,000 × *g* for 5 min and stored at −20 °C until further use. A phenol:chloroform:isoamyl alcohol and chloroform extraction method was used to obtain high-molecular-weight DNA as previously described[52]. The gDNA was quantified using a Qubit fluorometer and the Quant-iT double-stranded DNA (dsDNA) broad range (BR) assay kit (Invitrogen), and the quality was assessed with a NanoDrop One instrument (Thermo Fisher). Samples were subjected to metagenomic shotgun sequencing using the Illumina HiSeq 3000 platform at the Norwegian Sequencing Center (NSC, Oslo, Norway). Samples were prepared with the TrueSeq DNA PCR-free preparation and sequenced with paired ends (2 × 150 bp) on one lane. Quality trimming of the raw reads was performed using Cutadapt[53] v1.3 to remove all bases on the 3′-end with a Phred score lower than 20 and exclude all reads shorter than 100 nucleotides, followed by quality filtering using the FASTX toolkit v.0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/). Retained reads had a minimum Phred score of 30 over 90% of the read length. Reads were co-assembled using metaSPAdes[54] v3.10.1 with default parameters and *k*-mer sizes of 21, 33, 55, 77 and 99. The resulting contigs were binned with MetaBAT[55] v0.26.3 in 'very sensitive mode'. The quality (completeness, contamination and strain heterogeneity) of the metagenome-assembled genomes (MAGs) was assessed by CheckM[56] v1.0.7 with default parameters. Contigs were submitted to the Integrated Microbial Genomes and Microbiomes system for open reading frames (ORFs) prediction and annotation[57]. Additionally, the resulting ORFs were annotated for CAZymes using the CAZy annotation pipeline[58]. This MAG collection was used as a reference database for mapping of the metatranscriptome data, as described below. Taxonomic classifications of MAGs were determined using both MiGA[59] and GTDB-Tk[60].

Human faecal samples (20) from a second enrichment experiment (unbiased towards the cultivation of *Bacteroides*) as well as two enrichments with mouse faecal samples were processed for gDNA extraction and library preparation exactly as described above. Metagenomic shotgun sequencing was conducted on two lanes of both Illumina HiSeq 4000 and Illumina HiSeq X Ten platforms (Illumina) at the NSC (Oslo, Norway), and reads were quality trimmed, assembled and binned as described above. Open reading frames were annotated using PROKKA[61] v1.14.0 and resulting ORFs were further annotated for CAZymes using the CAZy annotation pipeline and expert human curation[58]. Completeness, contamination and taxonomic classifications for each MAG were determined as described above. AAI comparison between the human R.UCG13 and the R.UCG13 found in the two mouse samples was determined using CompareM (https://github.com/dparks1134/CompareM).

Extracted DNA from a second enrichment experiment on XG using Sample 0 was prepared for long-reads sequencing using Oxford Nanopore Technologies (ONT) ligation sequencing kit (SQK-LSK109) according to the manufacturer's protocol. The DNA library was sequenced with the ONT MinION sequencer using an R9.4 flow cell. The sequencer was controlled by the MinKNOW software v3.6.5 running for 6 h on a laptop (Lenovo ThinkPad P73 Xeon with data stored to 2TB SSD), followed by base calling using Guppy v3.2.10 in 'fast' mode. This generated 3.59 GB of data in total. The Nanopore reads were further processed using Filtlong v0.2.0 (https://github.com/rrwick/Filtlong), discarding the poorest 5% of the read bases and reads shorter than 1,000 bp.

The quality processed Nanopore long reads were assembled using CANU[62] v1.9 with the parameters 'corOutCoverage = 10000 corMinCoverage = 0 corMhapSensitivity = high genomeSize = 5 m redMemory = 32 oeaMemory = 32 batMemory = 200'. An initial polishing of the generated contigs was carried out using error-corrected reads from the assembly with minimap2[63] v2.17 '-x map-ont' and Racon[64] v1.4.14, with the argument '–include-unpolished'. The racon-polished contigs were further polished using Medaka v1.1.3 (https://github.com/nanoporetech/medaka), with the commands 'medaka_consensus–model r941_min_fast_g303_model.hdf5'. Finally, Minimap2 '-ax sr' was used to map quality processed Illumina reads to the medaka-polished contigs, followed by a final round of error correction using Racon with the argument '–include-unpolished'. Circular contigs were identified by linking the contig identifiers in the polished assembly back to 'suggestCircular = yes' in the initial contig header provided by CANU. These contigs were quality-checked using CheckM[56] v1.1.3 and BUSCO[65] v4.1.4. Circular contigs likely to represent chromosomes (>1 Mbp) were further gene-called and functionally annotated using PROKKA[61] v1.13 and taxonomically classified using GTDB-tk[60] v1.4.0 with the 'classify_wf' command. Barrnap v0.9 (https://github.com/tseemann/barrnap) was used to predict ribosomal RNA genes. Average nucleotide identity (ANI) was measured between the short-reads and long-reads MAGs using FastANI[66] v1.1 with default parameters. Short-reads MAGs were used as query, while long-reads MAGs were set as reference genomes. Short-reads MAG1 showed an ANI of 99.98% with the long-reads ONT_Circ01, while short-reads MAG2 showed an ANI of 99.99% with the long-reads ONT_Circ02 (Supplementary Table 2). Phylogenetic analysis revealed that ONT_Circ02 encoded four complete 16S rRNA operons, three of which were identical to the aforementioned R.UCG13 OTU.

Temporal metatranscriptomic analysis of the Sample 0 XG-degrading community. Cell pellets from 6 ml samples collected at T1–T4 during growth of two biological replicates of the Sample 0 culture were supplemented with RNAprotect bacteria reagent (QIAGEN) following the manufacturer's instructions and kept at −80 °C until RNA extraction. Messenger RNA extraction and purification were conducted as described in Kunath et al.[67]. Samples were processed with the TruSeq stranded RNA sample preparation, which included the production of a complementary DNA (cDNA) library, and sequenced on one lane of the Illumina HiSeq 3000 system (NSC, Oslo, Norway) to generate 2 × 150 paired-end reads. Before assembly, RNA reads were quality-filtered with Trimmomatic[68] v0.36, whereby the minimum read length was required to be 100 bases with an average Phred threshold of 20 over a 10 nt window, and rRNA and transfer RNA were removed using SortMeRNA[69] v.2.1b. Reads were pseudo-aligned against the metagenomic dataset using kallisto pseudo '–pseudobam'[70]. Of the 58,089 ORFs (that encode proteins with >60 aa) identified from the metagenome of the Sample 0 community, 7,549 (13%) were not found to be expressed, whereas 50,540 (87%) were expressed, resulting in a reliable quantification of the expression due to unique hits (reads mapping unambiguously against one unique ORF).

Neutral monosaccharide analysis. The hot-phenol extraction method originally described by Massie & Zimm[71] and modified by Nie[72] was used for collecting and purifying the polysaccharides remaining at different timepoints. Samples were heated to 65 °C for 5 min, combined with an equal volume of phenol, incubated at 65 °C for 10 min, then cooled to 4 °C and centrifuged at 4 °C for 15 min at 12,000 *g*. The upper aqueous layer was collected and re-extracted using the same procedure, dialysed extensively against deionized water (2,000 Da cutoff) and freeze-dried. Neutral monosaccharide composition was obtained using a previously described method[73]. Briefly, sugar alditol acetates were quantified by gas chromatography using the capillary column SP-2330 (Supelco) with the following conditions: injector volume, 2 µl; injector temperature, 240 °C; detector temperature, 300 °C; carrier gas (helium); velocity, 1.9 m s^{-1}; split ratio, 1:2; temperature programme was 160 °C for 6 min, then 4 °C min^{-1} to 220 °C for 4 min, then 3 °C min^{-1} to 240 °C for 5 min, and then 11 °C min^{-1} to 255 °C for 5 min.

Thin layer chromatography (TLC) for localization of enzyme activity. Overnight cultures were collected at 13,000 *g* for 10 min. Supernatant fractions were prepared by vacuum filtration through 0.22 µm PES filters. Cell pellet fractions were prepared by decanting supernatant, washing with phosphate buffered saline (PBS), spinning at 13,000 *g* for 3 min, decanting and resuspending in PBS. Intracellular fractions were prepared by taking cell pellet fractions and bead beating for 90 s with acid-washed glass beads (G1277, Sigma) in a Biospec mini beadbeater. Lysed culture fractions were prepared by directly bead beating unprocessed culture.

Each culture fraction was mixed 1:1 with 5 mg ml^{-1} xanthan gum and incubated at 37 °C for 24 h. Negative controls were prepared by heating culture fractions to 95 °C for 15 min, then centrifuging at 13,000 *g* for 10 min before the addition of xanthan gum. All reactions were halted by heating to ≥85 °C for 15 min, then spinning at 20,000 *g* for 15 min at 4 °C. Supernatants were stored at −20 °C until analysis by TLC.

Samples (3 µl) were spotted twice onto a 10 × 20 cm TLC plate (Millipore TLC silica gel 60, 20 × 20 cm aluminum sheets), with intermediate drying using a Conair 1875 hairdryer. Standards included malto-oligosaccharides of varying lengths (even: 2, 4, 6; odd: 1, 3, 5, 7), glucuronic acid and mannose. Standards were prepared at 10 mM and 3 µl of each was spotted onto the TLC plate. Plates were run in ~100 ml of 2:1:1 butanol:acetic acid:water, dried, then run an additional time. After drying, plates were incubated in developing solution (100 ml ethyl acetate, 2 g diphenylamine, 2 ml aniline, 10 ml ~80% phosphoric acid, 1 ml ~38% hydrochloric acid) for ~30 s, then dried and developed by holding over a flame until colours were observed.

Proteomic analysis. Approximately 1 l of xanthan gum culture was grown until it had completely liquified (~2–3 d). Supernatant was collected by centrifuging at 18,000 *g* and vacuum filtering through a 0.2 µm PES filter. Ammonium sulfate (4 M) was added to 200–400 ml of filtrate to a final concentration of 2.4 M and incubated for 30–60 min at r.t. or, for one sample, overnight at 4 °C. Precipitated proteins were collected by centrifugation at 18,000 *g* for 30–60 min, then resuspended in 50 mM sodium phosphate (pH 7.5). Three different fractionation protocols were followed, but after every fractionation step, active fractions were identified by mixing ~500 µl with 10 mg ml^{-1} xanthan and incubating at 37 °C overnight; active fractions were identified by loss of viscosity or production of xanthan oligosaccharides as visualized by TLC (method previously described).

1. Resuspended protein was filtered and applied to a HiTrapQ column, running a gradient from 0 to 100% buffer B (buffer A: 50 mM sodium phosphate, pH 7.5; buffer B: 50 mM sodium phosphate, 1 M NaCl, pH 7.5). Active fractions were pooled and concentrated with a 10 kDa MWCO centricon and injected onto an S-200 16/60 column equilibrated in 50 mM sodium phosphate, 200 mM NaCl, pH 7.5. The earliest fractions to elute with substantial A280 absorbance were also the most active fractions; these were pooled and submitted for proteomics.
2. Resuspended protein was filtered and applied to an S-500 column equilibrated in 50 mM sodium phosphate, 200 mM NaCl, pH 7.5. Active fractions eluted in the middle of the separation were pooled and submitted for proteomics.

3. Resuspended protein was filtered and applied to an S-500 column equilibrated in 50 mM sodium phosphate, 200 mM NaCl, pH 7.5. Pooled fractions were applied to a 20 ml strong anion exchange column running a gradient from 0 to 100% buffer B (buffer A: 50 mM sodium phosphate, pH 7.5; buffer B: 50 mM sodium phosphate, 1 M NaCl, pH 7.5). Active fractions were pooled and applied to a 1 ml weak anion exchange column running a gradient from 0 to 100% buffer B (buffer A: 50 mM sodium phosphate, 10% glycerol, pH 7.5; buffer B: 50 mM sodium phosphate, 1 M NaCl, 10% glycerol, pH 7.5). Active fractions were pooled and submitted for proteomics.

Cysteines were reduced by adding 50 ml 10 mM dithiothreitol and incubating at 45 °C for 30 min. Samples were cooled to room temperature and alkylation of cysteines was achieved by incubating with 65 mM 2-Chloroacetamide under darkness for 30 min at room temperature. An overnight digestion with 1 µg sequencing-grade modified trypsin was carried out at 37 °C with constant shaking in a Thermomixer. Digestion was stopped by acidification and peptides were desalted using SepPak C18 cartridges following the manufacturer's protocol (Waters). Samples were completely dried using vacufuge. Resulting peptides were dissolved in 8 ml 0.1% formic acid/2% acetonitrile solution, and 2 µl of the peptide solution was resolved on a nanocapillary reverse-phase column (Acclaim PepMap C18, 2 µ, 50 cm, Thermo Fisher) using a 0.1% formic acid/2% acetonitrile (buffer A) and 0.1% formic acid/95% acetonitrile (buffer B) gradient at 300 nl min^{-1} over a period of 180 min (2–25% buffer B in 110 min, 25–40% in 20 min, 40–90% in 5 min, followed by holding at 90% buffer B for 10 min and requilibration with buffer A for 30 min). Eluent was directly introduced into a Q exactive HF mass spectrometer (Thermo Fisher) using an EasySpray source. MS1 scans were acquired at 60 K resolution (automatic gain control target, 3×10^6; max injection time, 50 ms). Data-dependent collision-induced dissociation MS/MS spectra were acquired using top speed method (3 s) following each MS1 scan (normalized collision energy, ~28%; 15 K resolution; automatic gain control target, 1×10^5; max injection time, 45 ms).

Proteins were identified by searching the MS/MS data against a database of all proteins identified in the Sample 0 culture metagenomes using Proteome Discoverer v2.1 (Thermo Fisher). Search parameters included MS1 mass tolerance of 10 ppm and fragment tolerance of 0.2 Da; two missed cleavages were allowed; carbamidimethylation of cysteine was considered a fixed modification, and oxidation of methionine and deamidation of asparagine and glutamine were considered as potential modifications. False discovery rate (FDR) was determined using Percolator and proteins/peptides with an FDR of ≤1% were retained for further analysis.

Plasmid design and protein purification. Plasmid constructs to produce recombinant proteins were made with a combination of synthesized DNA fragments (GenScript Biotech) and PCR amplicons using extracted culture gDNA as a template. In general, sequences were designed to remove N-terminal signalling peptides and add a histidine tag for immobilized metal affinity chromatography (IMAC) (in many cases using the Lucigen MA101-Expresso-T7-Cloning-&-Expression-System). Plasmid assembly and protein sequences are described in Supplementary Table 8.

Constructs were transformed into HI-Control BL21(DE3) cells and single colonies were inoculated in 5 ml overnight Luria-Bertani broth cultures at 37 °C. Cultures (5 ml) were used to inoculate 1 l of terrific broth (TB) with selective antibiotic, grown to OD_{600} ~0.8–1.1 at 37 °C, and induced with 250 µM isopropyl β-D-1-thiogalactopyranoside. *B. intestinalis* enzymes were expressed at r.t., while R.UCG13 enzymes were expressed at 18 °C overnight. Cells were collected by centrifugation at ≥12,000 *g* for 20 min at 4 °C and pellets were stored at −80 °C until further processing. Proteins were purified using standard IMAC purification procedures, employing sonication to lyse cells. R.UCG13 proteins were purified using 50 mM sodium phosphate and 300 mM sodium chloride at pH 7.5; *B. intestinalis* proteins were purified using 50 mM Tris and 300 mM sodium chloride at pH 8.0. All proteins were eluted from cobalt resin using buffer with the addition of 100 mM imidazole, then buffer exchanged to remove imidazole using Zeba 2 ml 7 kDa MWCO desalting columns. Protein concentrations were determined by measuring A280 absorbance and converting to molarity using calculated extinction coefficients.

Characterization and isolation of xanthan gum degradation products. In general, pentameric xanthan oligosaccharides were produced by incubating ≥0.1 mg ml^{-1} *Ru*GH5a with 5 mg ml^{-1} xanthan gum in PBS in approximately 1 l total volume. For xanthan tetrasaccharides, ~0.5 U ml^{-1} xanthan lyase (E-XANLB, Megazyme) was included. After incubating for 2–3 d at 37 °C to allow complete liquefication, reactions were heat-inactivated, centrifuged at ≥10,000 *g* for 30 min, and the supernatant was vacuum-filtered through 0.22 µm PES sterile filters. Supernatants were loaded onto a column containing ~10 g graphitized carbon (Supelclean ENVI-Carb, 57210-U, Supelco), washed extensively with water to remove salt and unbound material, then eluted in a stepwise fashion with increasing concentrations of acetonitrile. Fractions were dried, weighed and analysed by LC–MS and fractions that contained the largest yield of desired products were combined.

Highly pure products were obtained by reconstituting samples in 50% water:acetonitrile and applying to a Luna 5 µm HILIC 200 Å LC column (250 ×10 mm) (00G-4450-N0, Phenomenex). A gradient was run from 90 to 20% acetonitrile, with peaks determined through a combination of evaporative light scattering, UV and post-run analytical LC–MS (Agilent qToF 6545) of resulting fractions.

Nuclear magnetic resonance (NMR) spectra were collected using an Agilent 600 NMR spectrometer (^{1}H: 600 MHz, ^{13}C: 150 MHz) equipped with a 5 mm DB AUTOX PFG broadband probe and a Varian NMR System console. All data analysis was performed using MestReNova NMR software. All chemical shifts were referenced to residual solvent peaks [^{1}H (D_2O): 4.79 ppm].

Enzyme reaction analysis. Most enzymes were tested for activity by mixing different combinations of purified enzymes together with XGOs generated from *Ru*GH5a and measuring the products generated after a 16 h incubation at 37 °C (Extended Data Fig. 5a–h). These enzyme reactions were carried out in 15–25 mM sodium phosphate buffer, 100–150 mM sodium chloride, and sometimes included up to 0.01 mg ml^{-1} bovine serum albumin (B9000S, NEB) to limit enzyme adsorption to pipettes and tubes. All R.UCG13 or *B. intestinalis* enzymes were tested at concentrations from 1 to 10 µM. Reactions used 2.5 mg ml^{-1} pentasaccharide (produced using *Ru*GH5a) and were carried out at pH 6.0. Reactions were incubated for 16 h at 37 °C, halted by heating at ≥95 °C for 5–10 min, and centrifugation at ≥20,000 *g* for 10 min. Supernatants were mixed with 4 parts acetonitrile to yield an 80% acetonitrile solution, centrifuged for 10 min at ≥20,000 *g* and transferred into sample vials. Each sample (15 µl) was injected onto a Luna Omega 3 µm HILIC 200 Å LC column (100 ×4.6 mm) (00D-4449-E0, Phenomenex). An Agilent 1290 Infinity II high performance liquid chromatography (HPLC) system was used to separate the sample using solvent A (100% water, 0.1% formic acid) and solvent B (95% acetonitrile, 5% water, with 0.1% formic acid added) at a flow rate of 0.4 ml min^{-1}. Before injection and following each sample injection, the column was equilibrated with 80% solvent B. After injection, samples were eluted with a 30 min isocratic step at 80% solvent B, a 10 min gradient decreasing solvent B from 80% to 10%, and a final column wash for 2 min at 10% solvent B. Spectra were collected in negative mode using an Agilent 6545 LC/qToF. Data were analysed using MassHunter Qualitative Analysis v10.0.

To test enzyme activity against the smaller substrate cellobiose, individual enzymes were tested using 1 mM cellobiose at pH 7.5 (Extended Data Fig. 5l), also using 15–25 mM sodium phosphate buffer, 100–150 mM sodium chloride, and sometimes included up to 0.01 mg ml^{-1} bovine serum albumin. All R.UCG13 or *B. intestinalis* enzymes were tested at concentrations from 1 to 10 µM. After incubating for 16 h at 37 °C, reactions were analysed using TLC as previously described.

Kinetics of RuGH5a (bicinchoninic acid assay). Lyase-treated xanthan gum was generated by mixing 5 mg ml^{-1} xanthan gum with 0.5 U ml^{-1} of *Bacillus* sp. xanthan lyase (E-XANLB, Megazyme) in 30 mM potassium phosphate buffer (pH 6.5). After incubating overnight at 37 °C, an additional 0.5 U ml^{-1} of xanthan lyase was added. Both lyase-treated and native xanthan gum were dialysed extensively against deionized water, heated in an 80 °C water bath to inactivate the lyase and centrifuged at 10,000 *g* for 20 min to remove particulates. Supernatants were collected and stored at 4 °C until use.

Kinetic measurements were conducted using a slightly modified version of the low-volume bicinchoninic acid assay for glycoside hydrolases used by Arnal et al.[74]. Briefly, anion exchange chromatography (AEX) and size exclusion chromatography (SEC) purified *Ru*GH5a was diluted to a 10x stock of 5 µM enzyme, 50 mM sodium phosphate, 300 mM sodium chloride and 0.1 mg ml^{-1} bovine serum albumin, at pH7.5. Reactions were 20 µl of enzyme stock mixed with 180 µl of various concentrations of xanthan gum. Negative controls were conducted with heat-inactivated enzyme stock. Timepoints were taken by quenching reactions with dilute ice-cold bicinchoninic acid working reagent. Reactions and controls were run with 4 independent replicates and compared to a glucose standard curve. Enzyme-released reducing sugar was calculated by subtracting controls from reaction measurements.

Growth curves of isolates on XG oligos. Pure isolates from the xanthan culture or type strains were obtained by streaking an active culture onto a variety of agar plates including Luria-Bertani broth and brain–heart infusion, with the optional addition of 10% defibrinated horse blood (Colorado Serum) and gentamicin (200 µg ml^{-1}). After passaging isolates twice on agar plates, individual colonies were picked and grown overnight in tryptone-yeast extract-glucose (TYG) broth medium, then stocked by mixing with 0.5 volumes each of TYG and molecular-biology-grade glycerol and storing at −80 °C.

DM without beef extract (DM^{-BE}), with the addition of a defined carbon source, was used to test isolates for growth on xanthan oligosaccharides. Some isolates (for example *Parabacteroides distasonis*) required the inclusion of 5 mg ml^{-1} beef extract (Sigma, B4888) to achieve robust growth on simple monosaccharides; in these cases, beef extract was included across all carbon conditions. Unless otherwise specified, carbon sources were provided at a final concentration of 5 mg ml^{-1}. To test tetrasaccharide utilization, XGOs were incubated with lyase as previously described before their use in growth media. Isolates were grown

overnight in TYG media, subcultured 1:50 into DM^{-BE}-glucose and grown overnight, then subcultured 1:50 into DM^{-BE} with various carbon sources. Final cultures were monitored for growth by measuring increase in absorbance (600 nm) using 96-well plates as previously described.

Testing for lyase activity in *B. intestinalis* cell fractions. *B. intestinalis* was grown overnight in TYG, then diluted 1:50 into fresh DM^{-BE}-XGO media. Triplicate cultures were grown for 6 h until OD_{600} of 1.1 ± 0.1 (s.e.m.), then collected by centrifugation and processed analogously to previous culture fractionation methods. Individual fractions (including heat-inactivated controls) were mixed 1:1 with 5 mg ml^{-1} XGOs and incubated at 37 °C. Timepoints were taken at 30 and 60 min as well as overnight by heating reactions at 85 °C for 15 min, centrifuging at 20,000 *g* for 10 min and storing supernatant at −20 °C. Analysis of reaction products was identical to enzyme reaction analysis.

qPCR on *B. intestinalis* and *B. salyersiae*. For qPCR, *B. intestinalis* or *B. salyersiae* were grown as before, but cells were collected by centrifugation at ≥16,000 *g* at mid-exponential phase, mixed with RNA Protect (QIAGEN) and stored at −80 °C until further processing. At collection, *B. intestinalis* average OD_{600} values were ~0.8 and ~0.6 for glucose- and oligosaccharide-grown cultures, respectively. For *B. salyersiae*, OD_{600} values were ~0.7 and ~0.5 for glucose- and lyase-treated oligosaccharide-grown cultures, respectively. RNeasy mini kit buffers (QIAGEN) were used to extract total RNA, which were then purified with RNA-binding spin columns (Epoch), treated with DNase I (NEB) and additionally purified using the RNeasy mini kit. SuperScript III reverse transcriptase and random primers (Invitrogen) were used to perform reverse transcription. Target transcript abundance in the resulting cDNA was quantified using a homemade qPCR mix as described previously[75] and gene-specific primers (Supplementary Table 9). Each 20 ul reaction contained 1x Thermopol reaction buffer (NEB), 125 uM dNTPs, 2.5 mM MgSO4, 1X SYBR Green I (Lonza), 500 nM gene-specific or 65 nM 16S rRNA primer and 0.5 units Hot Start *Taq* polymerase (NEB), and 10 ng of template cDNA. All samples were run in technical duplicates; in rare instances where a cycle threshold (Ct) value was not measured for one of the technical duplicates, analysis was completed using a single replicate. Results were processed using the delta-delta-Ct (ddCt) method in which raw values were normalized to 16S rRNA values, then xanthan oligosaccharide values were compared to those from glucose to calculate fold change in expression.

RNA-seq on *B. intestinalis* and Sample 0 community. For RNA-seq, total RNA was used from the *B. intestinalis* growths used for qPCR. For the community grown on XG or PGA, 5 ml cultures of DM-XG or DM-PGA were inoculated with a 1:100 dilution of a fully liquified DM-XG culture. PGA cultures were collected at mid-log phase at OD_{600} ~0.85, whereas XG cultures were collected at late-log phase at OD_{600} ~1.2 to allow liquefaction of XG, which was necessary to extract RNA from these cultures. As before, cultures were collected by centrifugation, mixed with RNA Protect (QIAGEN) and stored at −80 °C until further processing. RNA was purified as before except that multiple replicates of DM-XG RNA were pooled together and concentrated with Zymo RNA Clean and Concentrator−25 to reach acceptable concentrations for RNA depletion input. rRNA was depleted twice from the purified total RNA using the MICROBExpress kit, each followed by a concentration step using the Zymo RNA Clean and Concentrator-25. About 90% rRNA depletion was achieved for all samples. *B. intestinalis* RNA was sequenced using NovaSeq and community RNA was sequenced using MiSeq. The resulting sequence data were analysed for differentially expressed genes following a previously published protocol[76]. Briefly, reads were filtered for quality using Trimmomatic v0.39[68]. Reads were aligned to each genome using BowTie2 v2.3.5.1[77]. For the *Bacteroides intestinalis* transcriptome, reads were aligned to its genome, while for the community data, reads were aligned to either the *B. intestinalis* genome or the closed Ruminococcaceae UCG13 metagenome-assembled genome (MAG). Reads mapping to gene features were counted using htseq-count (release_0.11.1)[78]. Differential expression analysis was performed using the edgeR v3.34.0 package in R v.4.0.2 (with the aid of Rstudio v1.3.1093). The trimmed mean of M-values (TMM) method was used for library normalization[79]. Coverage data were visualized using Integrated Genome Viewer (IGV)[80].

Extended metagenome analysis/comparison methodology. Individual MAGs in each sample were searched by BlastP for the presence of proteins similar to those encoded by the XG-degrading PUL of R.UCG13 and *B. intestinalis*. This was done using the amino acid sequences of the proteins in the R.UCG13 and *B. intestinalis* PULs as the search homologues; both BlastP probes were searched against all the individual MAGs in the different samples, with the default threshold *e*-value of 1×10^{-5}.

Looking for R.UCG13 and *B. intestinalis* XG Loci in metagenomes. Available cohorts of human gut metagenomic sequence data (National Center for Biotechnology Information projects: PRJNA422434[81], PRJEB10878[82], PRJEB12123[83], PRJEB12124[84], PRJEB15371[85], PRJEB6997[86], PRJDB3601[87], PRJNA48479[88], PRJEB4336[89], PRJEB2054[90], PRJNA392180[91] and PRJNA527208[92]) were searched for the presence of xanthan locus nucleotide sequences from R.UCG13 (92.7 kb) and *B. intestinalis* (17.9 kb) using the following workflow: each xanthan locus nucleotide sequence was used separately as a template and then magic-blast v1.5.0[93] was used to recruit raw Illumina reads from the available metagenomic datasets, with an identity cutoff of 97%. Next, the alignment files were used to generate a coverage map using bedtools v2.29.0[94] to calculate the percentage coverage of each sample against each individual reference. We considered a metagenomic data sample to be positive for a particular xanthan locus if it had at least 70% of the corresponding xanthan locus nucleotide sequence covered.

The R.UCG13 locus and *B. intestinalis* XG locus were used as the query in a large-scale search against the assembled scaffolds of isolates, metagenome-assembled genomes (bins) and metagenomes included into the Integrated Microbial Genomes and Microbiomes (IMG/M) comparative analysis system[41]. Within the LAST software package (version 1066), the 'lastal' tool was used with default thresholds to search the 2 loci against 72,491 public high-quality isolate genomes, 102,860 bins from 13,415 public metagenomes and 21,762 public metagenomes in IMG/M. Metagenome bins were generated using the previously described binning analysis method[95].

Humanized mouse experiment. All animal experiments followed protocols approved by the University of Michigan Committee for the Use and Care of Animals. None of the mice used were involved in any previous experiments or treatments before their use in this experiment. Subjects were randomized to experimental treatment by cages to minimize the number of cages needed to conduct the experiment. Subjects were housed in Iso-Positive Tecniplast cages and light cycles were 12 h on, 12 h off, with lights on between 6:00 and 18:00. Temperature was maintained at 70–73 °F and humidity was maintained at 40–60%. A donor microbiome was chosen on the basis of its ability to degrade XG in anaerobic cultures and its lack of the *B. intestinalis* XGO locus as determined by PCR of extracted gDNA. Germfree Swiss Webster female mice aged 7–8 weeks were put on a high-XG diet (5% XG) at experimental day 0 and maintained on this diet during initial microbiome gavages. The 5% XG diet was a modified 'fiber-free' diet that has been previously used, with 50 g kg^{-1} XG added at the expense of dextrose (XG diet Teklad/Envigo TD.18017)[96]. Faecal material from the donor was resuspended in pre-reduced PBS at ~1:10 dilution and 200 µl was gavaged into each subject on days 7, 8 and 9. Faecal material was also passaged 3x in DM-XG to enrich for R.UCG13, then 200 µl of these cultures were gavaged into each subject on days 21 and 22. On day 29, one cage of mice (Group 2) was switched to a diet without XG (TD.130343)[97]. On day 36, all mice were gavaged with 200 µl of the XGO-consuming *B. intestinalis* isolate that had been grown overnight in TYG media. Finally, the mice that had been switched to the diet without XG (Group 2) were put back on the high-XG diet. Fresh faecal samples were taken at several timepoints, stored at 4 °C and used within 48 h to inoculate DM-XG media in an anaerobic chamber as discussed previously. Cultures were monitored and determined positive for XG degradation if they showed increased turbidity and loss of XG's characteristic viscosity. Faecal samples for gDNA extraction were collected throughout the experiment and stored at −20 °C until processing. DNA extractions and 16S rRNA community sequencing were carried out as previously described.

To determine the amount of XGO-consuming *B. intestinalis* in each of the samples, gene-specific primers for the XGO locus and qPCR were used to quantify the amount of *B. intestinalis* DNA using a standard curve of isolated *B. intestinalis* DNA. qPCR was carried out as previously described, except that only 0.5 ng of input DNA was used due to low extraction yields from mouse faecal pellets (possibly due to high residual XG concentrations). Input gDNA was quantified using a Qubit fluorometer and the Quant-iT dsDNA BR assay kit (Invitrogen) as before and universal primers confirmed that total input DNA was similar across samples. Quantities of *B. intestinalis* DNA were converted to relative abundance by dividing by the total input DNA and averaged across all genes tested (*B. intestinalis* XGO genes: GH88, GH92, GH3 and PL-CE) to give the final relative abundance of the XGO-consuming *B. intestinalis*. Unpaired *t*-tests were run in Graphpad Prism v9.3.0 to determine significance of differences, with significant *P* values (<0.05) noted.

Statistics and reproducibility. No power calculations were made before experiments to arrive at sample size for assays, including in vivo experiments. Sample sizes for growth curves, enzyme reactions, expression analysis and in vivo experiments were typically performed at least in triplicate, which in past studies by our groups and others have been sufficient to detect significant differences between experimental conditions[98–100]. Specific experiments were performed only in duplicate either because of limited material availability and/or results of experiments performed in this manner had already yielded significant and informative results. Data distributions were assumed to be normal, but this was not formally tested. No data were excluded from the analysis. The experiments were not randomized. The Investigators were not blinded to allocation during experiments and outcome assessment.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All sequencing reads have been deposited at the European Nucleotide Archive under BioProject PRJEB44146. All annotated MAGs are publicly available via Figshare (DOIs: 10.6084/m9.figshare.14494602, 10.6084/m9.figshare.14494536, 10.6084/m9.figshare.14494677, 10.6084/m9.figshare.14494683 and 10.6084/m9.figshare.14494689). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[101] partner repository with the dataset identifier PXD031522. There are no restrictions on data or biological resource availability. Data and biological resources can be requested from the corresponding authors. Source data are provided with this paper.

Code availability

No new codes were developed or compiled in this study.

Received: 3 November 2021; Accepted: 24 February 2022;
Published online: 1 April 2022

References

1. Chassaing, B. et al. Dietary emulsifiers impact the mouse gut microbiota promoting colitis and metabolic syndrome. *Nature* **519**, 92–96 (2015).
2. Collins, J. et al. Dietary trehalose enhances virulence of epidemic *Clostridium difficile*. *Nature* **553**, 291–294 (2018).
3. Laudisi, F. et al. The food additive maltodextrin promotes endoplasmic reticulum stress–driven mucus depletion and exacerbates intestinal inflammation. *Cell. Mol. Gastroenterol. Hepatol.* **7**, 457–473 (2019).
4. Etienne-Mesmin, L. et al. Experimental models to study intestinal microbes–mucus interactions in health and disease. *FEMS Microbiol. Rev.* **43**, 457–489 (2019).
5. García-Ochoa, F., Santos, V. E., Casas, J. A. & Gómez, E. Xanthan gum: production, recovery, and properties. *Biotechnol. Adv.* **18**, 549–579 (2000).
6. Cartmell, A. et al. A surface endogalactanase in *Bacteroides thetaiotaomicron* confers keystone status for arabinogalactan degradation. *Nat. Microbiol.* **3**, 1314–1326 (2018).
7. Pudlo, N. A. et al. Diverse events have transferred genes for edible seaweed digestion from marine to human gut bacteria. *Cell Host & Microbe* **30**, 314–328 (2022).
8. Casas, J. A., Santos, V. E. & García-Ochoa, F. Xanthan gum production under several operational conditions: molecular structure and rheological properties. *Enzym. Microb. Technol.* **26**, 282–291 (2000).
9. Sworn, G. in *Handbook of Hydrocolloids* (Third edition), edited by Phillips, G. O. and Williams, P. A. 833–853 (Elsevier, 2021).
10. King, J. A. et al. Incidence of celiac disease is increasing over time. *Am. J. Gastroenterol.* https://doi.org/10.14309/ajg.0000000000000523 (2020).
11. Mortensen, A. et al. Re-evaluation of xanthan gum (E 415) as a food additive. *EFSA J.* **15**, e04909 (2017).
12. Baxter, N. T. et al. Dynamics of human gut microbiota and short-chain fatty acids in response to dietary interventions with three fermentable fibers. *mBio* **10**, e02566–18 (2019).
13. Hehemann, J.-H., Kelly, A. G., Pudlo, N. A., Martens, E. C. & Boraston, A. B. Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. *Proc. Natl Acad. Sci. USA* **109**, 19786–19791 (2012).
14. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, 590–596 (2013).
15. Goodman, A. L. et al. Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proc. Natl Acad. Sci. USA* **108**, 6252–6257 (2011).
16. Kim, C. C. et al. Genomic insights from *Monoglobus pectinilyticus*: a pectin-degrading specialist bacterium in the human colon. *ISME J.* **13**, 1437–1456 (2019).
17. Ruijssenaars, H. J., de Bont, J. A. M. & Hartmans, S. A pyruvated mannose-specific xanthan lyase involved in xanthan degradation by *Paenibacillus alginolyticus* XL-1. *Appl. Environ. Microbiol.* **65**, 2446–2452 (1999).
18. Nankai, H., Hashimoto, W., Miki, H., Kawai, S. & Murata, K. Microbial system for polysaccharide depolymerization: enzymatic route for xanthan depolymerization by *Bacillus* sp. strain GL1. *Appl. Environ. Microbiol.* **65**, 2520–2526 (1999).
19. Hashimoto, W., Nankai, H., Mikami, B. & Murata, K. Crystal structure of *Bacillus* sp. GL1 xanthan lyase, which acts on the side chains of xanthan. *J. Biol. Chem.* **278**, 7663–7673 (2003).
20. Jensen, P. F. et al. Structure and dynamics of a promiscuous xanthan lyase from *Paenibacillus nanensis* and the design of variants with increased stability and activity. *Cell Chem. Biol.* **26**, 191–202.e6 (2019).
21. Aspeborg, H., Coutinho, P. M., Wang, Y., Brumer, H. & Henrissat, B. Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol. Biol.* **12**, 1–16 (2012).
22. Jongkees, S. A. K. & Withers, S. G. Unusual enzymatic glycoside cleavage mechanisms. *Acc. Chem. Res.* **47**, 226–235 (2014).
23. Rovira, C., Males, A., Davies, G. J. & Williams, S. J. Mannosidase mechanism: at the intersection of conformation and catalysis. *Curr. Opin. Struct. Biol.* **62**, 79–92 (2020).
24. Kool, M. M. et al. Characterization of an acetyl esterase from *Myceliophthora thermophila* C1 able to deacetylate xanthan. *Carbohydr. Polym.* **111**, 222–229 (2014).
25. Grondin, J. M., Tamura, K., Déjean, G., Abbott, D. W. & Brumer, H. Polysaccharide utilization loci: fueling microbial communities. *J. Bacteriol.* **199**, e00860–16 (2017).
26. Pilgaard, B., Vuillemin, M., Holck, J., Wilkens, C. & Meyer, A. S. Specificities and synergistic actions of novel PL8 and PL7 alginate lyases from the marine fungus *Paradendryphiella salina*. *J. Fungi* **7**, 80 (2021).
27. Zhu, B. & Yin, H. Alginate lyase: review of major sources and classification, properties, structure-function analysis and applications. *Bioengineered* **6**, 125–131 (2015).
28. Terrapon, N. et al. PULDB: the expanded database of polysaccharide utilization loci. *Nucleic Acids Res.* **46**, D677–D683 (2018).
29. Sun, Z., Liu, H., Wang, X., Yang, F. & Li, X. Proteomic analysis of the xanthan-degrading pathway of *Microbacterium* sp. XT11. *ACS Omega* **4**, 19096–19105 (2019).
30. Yang, F. et al. Novel endotype xanthanase from xanthan-degrading *Microbacterium* sp. strain XT11. *Appl. Environ. Microbiol.* **85**, e01800–18 (2019).
31. Guillén, D., Sánchez, S. & Rodríguez-Sanoja, R. Carbohydrate-binding domains: multiplicity of biological roles. *Appl. Microbiol. Biotechnol.* **85**, 1241–1249 (2010).
32. Mistry, J. et al. Pfam: the protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419 (2021).
33. Ebbes, M. et al. Fold and function of the InlB B-repeat. *J. Biol. Chem.* **286**, 15496–15506 (2011).
34. Bleymüller, W. M. et al. MET-activating residues in the B-repeat of the *Listeria monocytogenes* invasion protein InlB. *J. Biol. Chem.* **291**, 25567–25577 (2016).
35. Kool, M. M., Gruppen, H., Sworn, G. & Schols, H. A. Comparison of xanthans by the relative abundance of its six constituent repeating units. *Carbohydr. Polym.* **98**, 914–921 (2013).
36. Moroz, O. V. et al. Structural dynamics and catalytic properties of a multi-modular xanthanase. *ACS Catal.* **8**, 6021–6034 (2018).
37. Yang, F. et al. Production and purification of a novel xanthan lyase from a xanthan-degrading *Microbacterium* sp. strain XT11. *ScientificWorldJournal* **2014**, A368434 (2014).
38. Gregg, K. J. et al. Analysis of a new family of widely distributed metal-independent α-mannosidases provides unique insight into the processing of N-linked glycans. *J. Biol. Chem.* **286**, 15586–15596 (2011).
39. Daly, J., Tomlin, J. & Read, N. W. The effect of feeding xanthan gum on colonic function in man: correlation with in vitro determinants of bacterial breakdown. *Br. J. Nutr.* **69**, 897–902 (1993).
40. Kiełbasa, S. M., Wan, R., Sato, K., Horton, P. & Frith, M. C. Adaptive seeds tame genomic sequence comparison. *Genome Res.* **21**, 487–493 (2011).
41. Chen, I. M. A. et al. The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. *Nucleic Acids Res.* **49**, D751–D763 (2021).
42. Liang, R. et al. Metabolic capability of a predominant *Halanaerobium* sp. in hydraulically fractured gas wells and its implication in pipeline corrosion. *Front. Microbiol.* **7**, 988 (2016).
43. Schnizlein, M. K., Vendrov, K. C., Edwards, S. J., Martens, E. C. & Young, V. B. Dietary xanthan gum alters antibiotic efficacy against the murine gut microbiota and attenuates *Clostridioides difficile* colonization. *Msphere* **5**, e00708–19 (2020).
44. Katzbauer, B. Properties and applications of xanthan gum. *Polym. Degrad. Stab.* **59**, 81–84 (1998).
45. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* **79**, 5112–5120 (2013).
46. Schloss, P. D. et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).
47. Core R Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2020).
48. Wickham, H. Reshaping Data with the reshape Package. *J. Stat. Softw.* **21**, 1–20 (2007).
49. Neuwirth, E. RColorBrewer: ColorBrewer Palettes (2014); https://cran.r-project.org/package=RColorBrewer
50. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Second edition). (Springer, 2016).

51. Martens, E. C. et al. Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol.* **9**, e1001221 (2011).
52. Pope, P. B. et al. Isolation of Succinivibrionaceae implicated in low methane emissions from Tammar wallabies. *Science* **333**, 646–648 (2011).
53. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).
54. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. MetaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
55. 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 (2015).
56. 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 (2015).
57. Chen, I. M. A. et al. IMG/M: integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res.* **45**, D507–D516 (2017).
58. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, 490–495 (2014).
59. Rodriguez-R, L. M. et al. The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic Acids Res.* **46**, W282–W288 (2018).
60. Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* **36**, 1925–1927 (2020).
61. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
62. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
63. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
64. Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).
65. Seppey, M., Manni, M. & Zdobnov, E. M. *BUSCO: Assessing Genome Assembly and Annotation Completeness BT - Gene Prediction: Methods and Protocols* (Humana Press, 2019).
66. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).
67. Kunath, B. J. et al. From proteins to polysaccharides: lifestyle and genetic evolution of *Coprothermobacter proteolyticus*. *ISME J.* **13**, 603–617 (2019).
68. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
69. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).
70. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
71. Massie, H. R. & Zimm, B. H. The use of hot phenol in preparing DNA. *Proc. Natl Acad. Sci. USA* **54**, 1641–1643 (1965).
72. Nie, X. *Relationships Between Dietary Fiber Structural Features and Growth and Utilization Patterns of Human Gut Bacteria*. Doctoral dissertation, Purdue University (2016).
73. Tuncil, Y. E., Thakkar, R. D., Marcia, A. D. R., Hamaker, B. R. & Lindemann, S. R. Divergent short-chain fatty acid production and succession of colonic microbiota arise in fermentation of variously-sized wheat bran fractions. *Sci. Rep.* **8**, 16655 (2018).
74. Arnal, G., Attia, M. A., Asohan, J. & Brumer, H. in *Protein-Carbohydrate Interactions. Methods and Protocols* (eds Abbott, D. W. & Lammerts van Bueren, A.) 209–214 (Springer, 2017).
75. Speer, M. A. *Development of a Genetically Modified Silage Inoculant for the Biological Pretreatment of Lignocellulosic Biomass* (Pennsylvania State Univ., 2013).
76. Anders, S. et al. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat. Protoc.* **8**, 1765–1786 (2013).
77. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
78. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
79. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).
80. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
81. Wang, J. et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55–60 (2012).
82. Yu, J. et al. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut* **66**, 70–78 (2017).
83. Liu, R. et al. Gut microbiome and serum metabolome alterations in obesity and after weight-loss intervention. *Nat. Med.* **23**, 859–868 (2017).
84. Gu, Y. et al. Analyses of gut microbiota and plasma bile acids enable stratification of patients for antidiabetic treatment. *Nat. Commun.* **8**, 1–12 (2017).
85. He, Q. et al. Two distinct metacommunities characterize the gut microbiota in Crohn's disease patients. *Gigascience* **6**, 1–11 (2017).
86. Zhang, X. et al. The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nat. Med.* **21**, 895–905 (2015).
87. Nishijima, S. et al. The gut microbiome of healthy Japanese and its microbial and functional uniqueness. *DNA Res.* **23**, 125–133 (2016).
88. Lloyd-Price, J. et al. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* **550**, 61–66 (2017).
89. Le Chatelier, E. et al. Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541–546 (2013).
90. Qin, J. et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
91. Smits, S. A. et al. Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. *Science* **357**, 802–805 (2017).
92. Conteville, L. C., Oliveira-Ferreira, J. & Vicente, A. C. P. Gut microbiome biomarkers and functional diversity within an Amazonian semi-nomadic hunter-gatherer group. *Front. Microbiol.* **10**, 1743 (2019).
93. Boratyn, G. M., Thierry-Mieg, J., Thierry-Mieg, D., Busby, B. & Madden, T. L. Magic-BLAST, an accurate RNA-seq aligner for long and short reads. *BMC Bioinformatics* **20**, 405 (2019).
94. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
95. Clum, A. et al. DOE JGI Metagenome Workflow. *Msystems* **6**, e00804–20 (2021).
96. Schnizlein, M. K., Vendrov, K. C., Edwards, S. J., Martens, E. C. & Young, V. B. Dietary xanthan gum alters antibiotic efficacy against the murine gut microbiota and attenuates *Clostridioides difficile* colonization. *mSphere* **5**, e00708–19 (2020).
97. Desai, M. S. et al. A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. *Cell* **167**, 1339–1353.e21 (2016).
98. Larsbrink, J. et al. A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature* **506**, 498–502 (2014).
99. Foley, M. H., Martens, E. C. & Koropatkin, N. M. SusE facilitates starch uptake independent of starch binding in *B. thetaiotaomicron*. *Mol. Microbiol.* **108**, 551–566 (2018).
100. Luis, A. S. et al. A single sulfatase is required to access colonic mucin by a gut bacterium. *Nature* **598**, 332–337 (2021).
101. Perez-Riverol, Y. et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **50**, D543–D552 (2022).

Acknowledgements

We thank S. Theide for growth curve analysis suggestions; T. Johannessen and A. Lysberg for help with Nanopore metagenomics; the University of Michigan Proteomics Resource Facility, Germfree Animal Core, Microbiome Core, and Natural Products Discovery Core for their support in completion of this project; the University of Michigan Center for Gastrointestinal Research (UMCGR, NIDDK 5P30DK034933) for financial support with proteomics. This work was supported by the US National Institutes of Health (DK118024, DK125445 to E.C.M. and UL1TR002240 to M.P.O. and AI124255 to V.B.Y.) and the Research Council of Norway (FRIPRO programme, P.B.P. and S.L.L.R.: 250479; L.H.H.: 302639). The work conducted by the US Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, was supported under Contract No. DE-AC02-05CH11231.

Author contributions

M.P.O., S.L.L.R., P.B.P. and E.C.M. designed experiments and wrote the manuscript. T.M.S. provided access to cohort faecal samples for initial enrichment cultures by D.B. and E.C.M. M.P.O. and E.C.M. ran all additional culture experiments. S.L.L.R., B.J.K., L.H.H. and P.B.P. carried out metagenomic and metatranscriptomic analysis on cultured microbes. T.Y. and B. Hamaker carried out neutral monosaccharide analysis. M.P.O., S.L.L.R. and G.F. carried out recombinant enzyme studies. M.P.O., A.R. and A.T. isolated pure tetrasaccharide and A.R. characterized it with NMR. J.L., S.P.M. and H.B. provided isomeric tetrasaccharide from *P. nanensis* GH9. M.P.O., L.Q., G.P. and N.A.P. carried out qPCR and RNA-seq. M.K.S. and V.B.Y. provided faecal samples from mice fed xanthan gum. N.T., V.L. and B. Henrissat carried out CAZyme annotation. N.J.V., G.P., S.L.L.R., M.P.O., P.B.P. and E.A.E.-F. carried out bioinformatic searches for loci of interest in metagenomes and genomes from global gut and environmental samples. M.P.O. and E.C.M. carried out mouse experiments. All authors read and approved the manuscript.

Competing interests

E.C.M., P.B.P., M.P.O., S.L.L.R. and B.J.K. are listed as inventors on a patent application (number PCT/US2021/050494) from the Norwegian University of Life Sciences and The Regents of the University of Michigan. Recombinant enzyme constructs for processing xanthan gum and oligosaccharides are covered in this patent application and discussed in this paper. E.C.M. serves as consultant and Scientific Advisory Board member of Novome Biotechnologies. All other authors declare no competing interests.

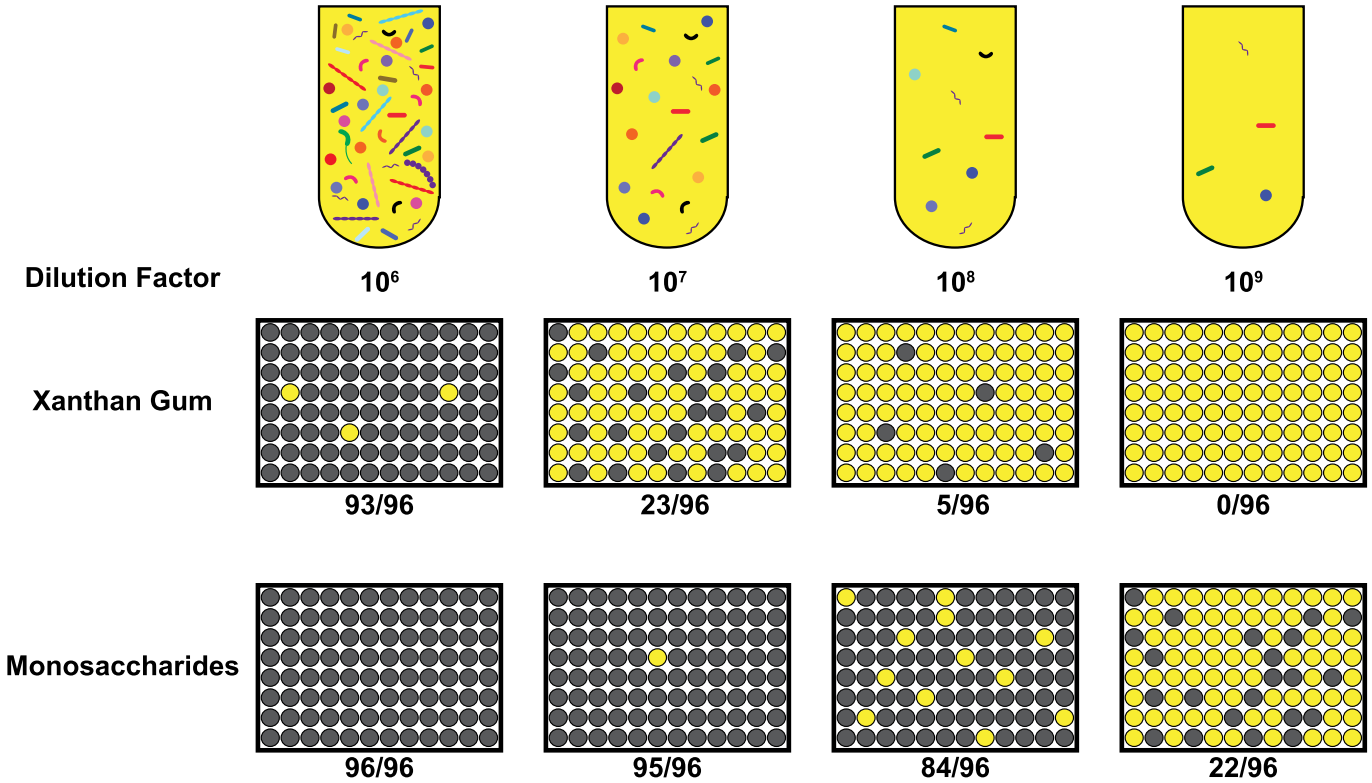
Additional information

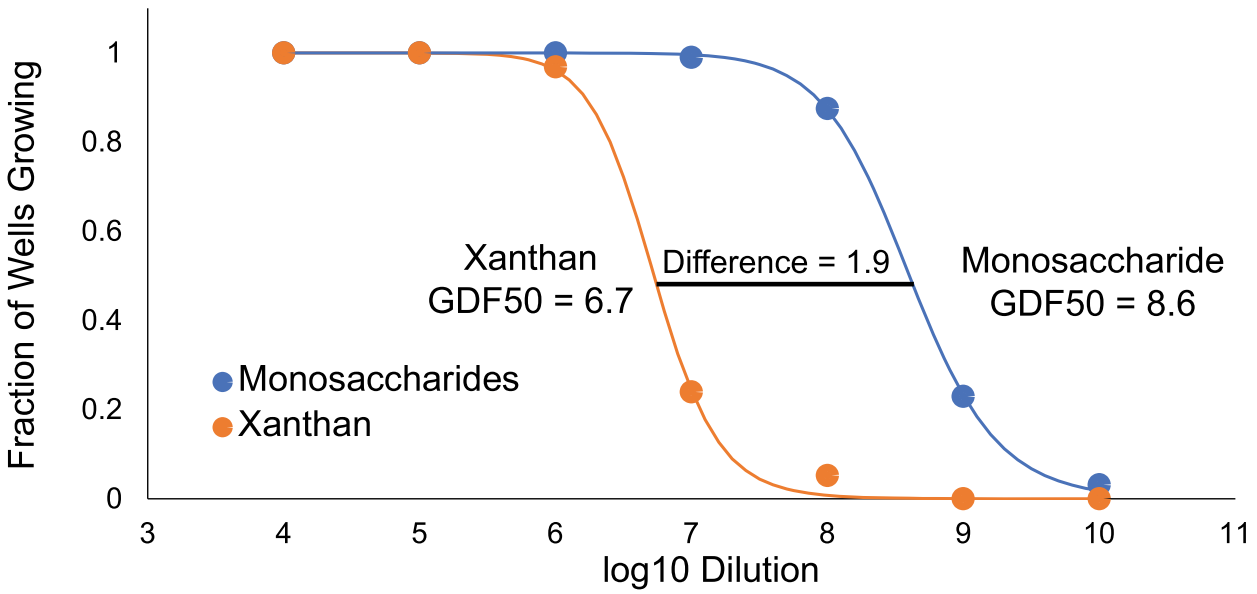
Extended data is available for this paper at https://doi.org/10.1038/s41564-022-01093-0.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41564-022-01093-0.

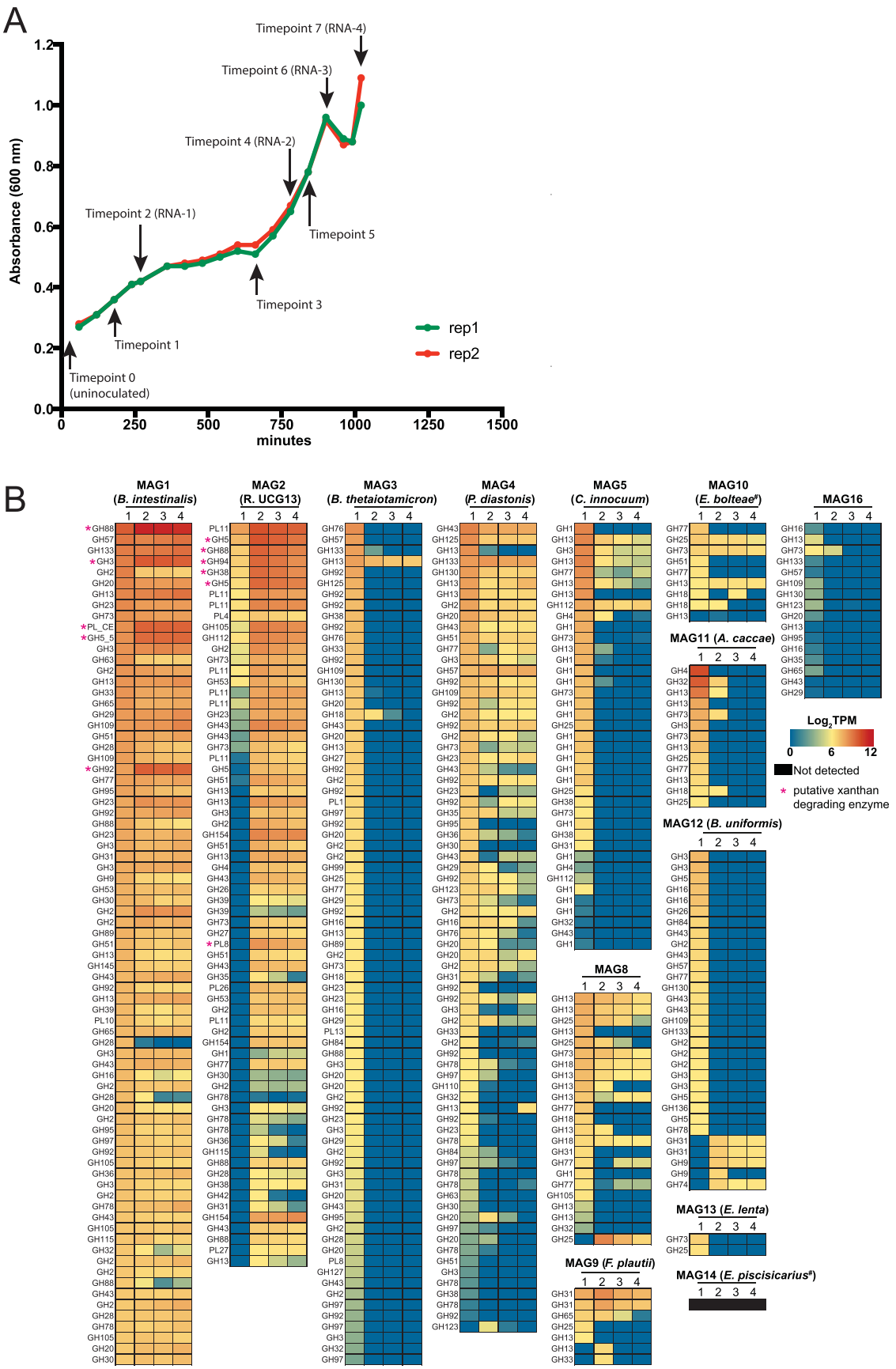
Correspondence and requests for materials should be addressed to Phillip B. Pope or Eric C. Martens.

Peer review information *Nature Microbiology* thanks Harry Gilbert, Seth Rakoff-Nahoum and the other anonymous, reviewer(s) for their contribution to the peer review of this work.



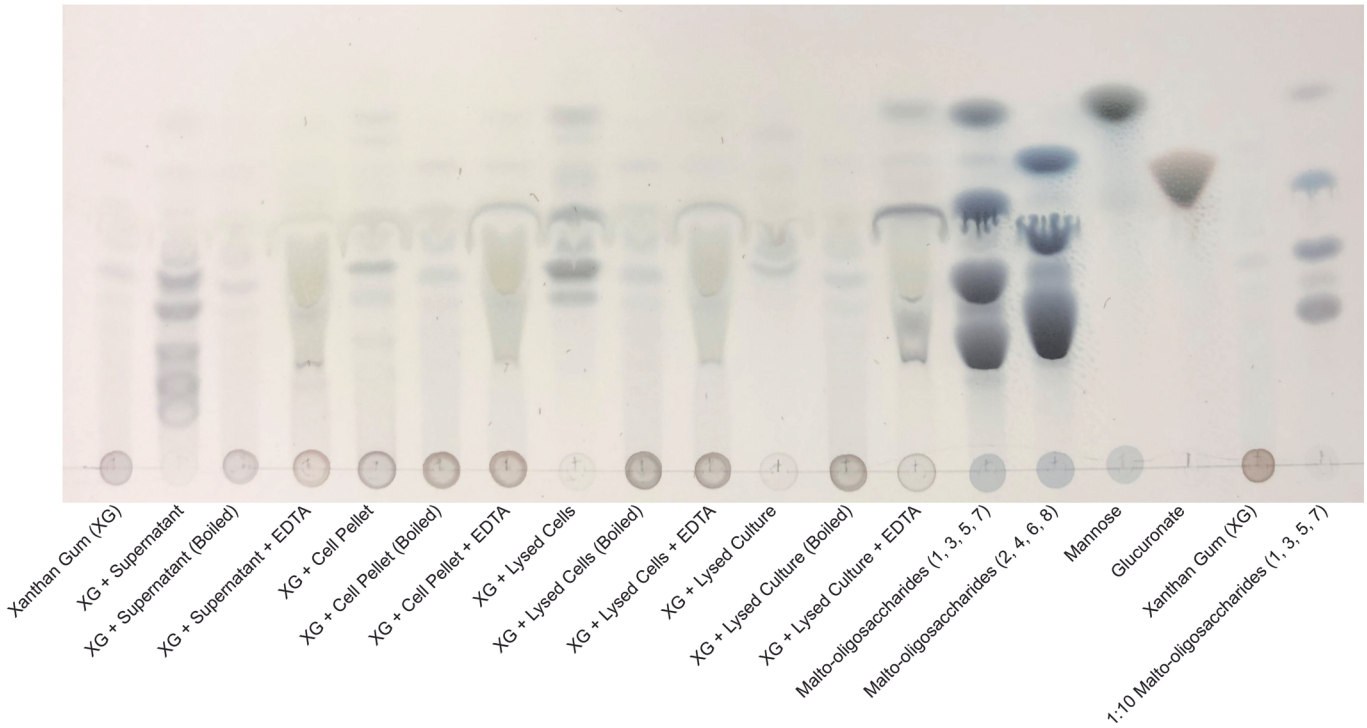


Extended Data Fig. 1 | Dilution to extinction suggests that xanthan gum degradation is a multi-species phenotype. An active xanthan culture was diluted in 2x defined media without a major carbon source, then divided and diluted 1:1 with either 2x xanthan gum or 2x monosaccharide mix (2:2:1 mannose;glucose;glucuronic acid), then aliquoted into 200 uL cultures in 96-well plates. Each datapoint represents the fraction of cultures (out of 96) growing above OD600 0.7 at each dilution, grown in either xanthan gum or monosaccharide mix media. Data were fit to the Hill equation to calculate a 50% growth dilution factor (GDF 50) at which half of the cultures would grow above OD 0.7. Across 5 independent experiments, there was an average GDF50 difference of 1.8 (standard deviation = 0.4, SEM = 0.2). This demonstrates that at comparable dilutions, microbes were present that could grow on monosaccharides but were unable to grow using XG, suggesting that several microbes are required in this media to allow growth on XG. In the schematic representation of 96-well plates, yellow wells represent non-growing media while gray wells represent a growing culture.

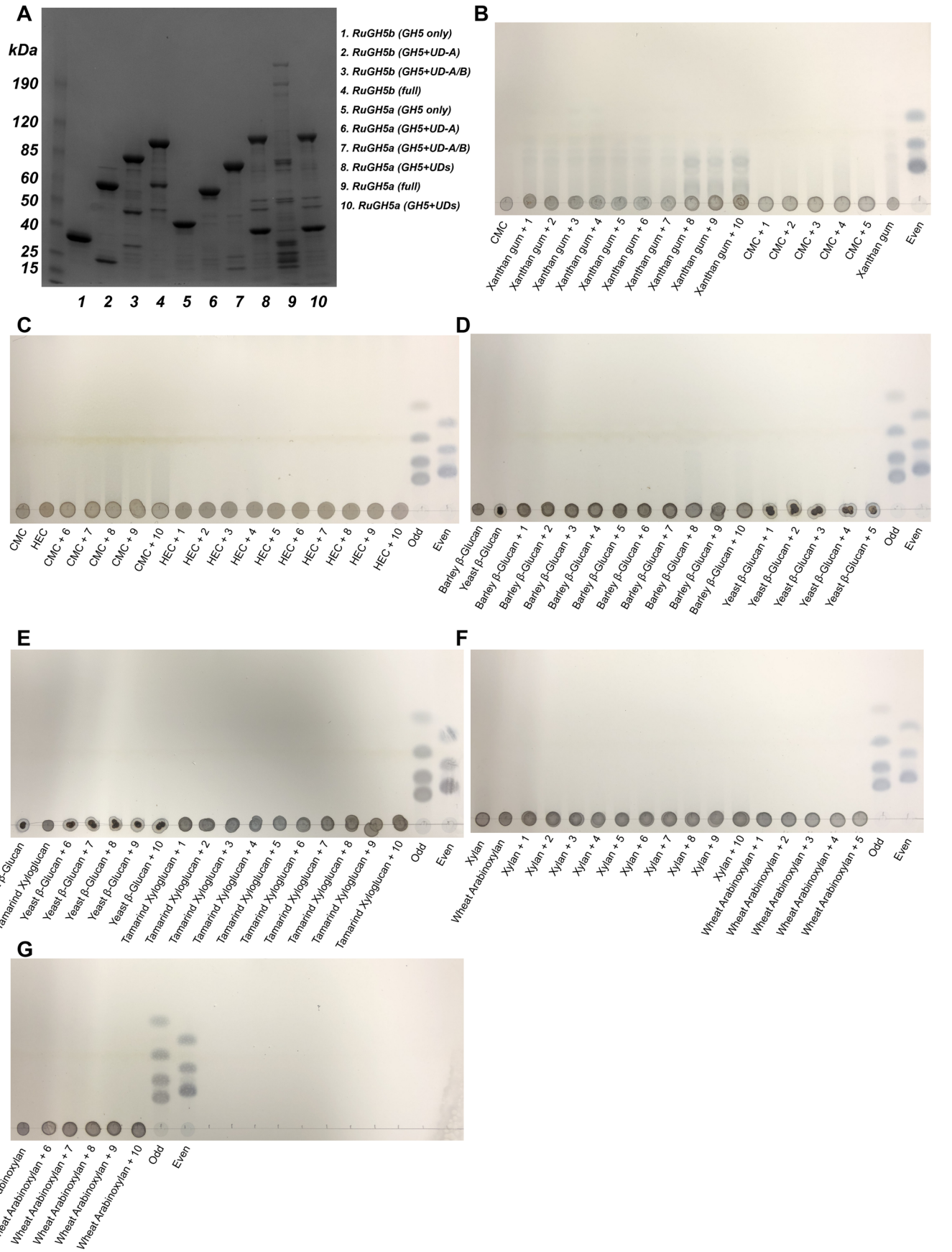


Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Sample 0 growth and metatranscriptomic analysis. a, Two distinct replicates of the Sample 0 culture were grown and sampled at multiple timepoints for **b**, metatranscriptomic analysis of annotated CAZymes in each of the MAGs (completeness value > 75%) reconstructed from metagenomic data from the enrichment culture. Values are transcripts per million (TPM). MAG taxonomy (Supplementary Table 2) is indicated in parentheticals. An '#' indicates a low AAI%.

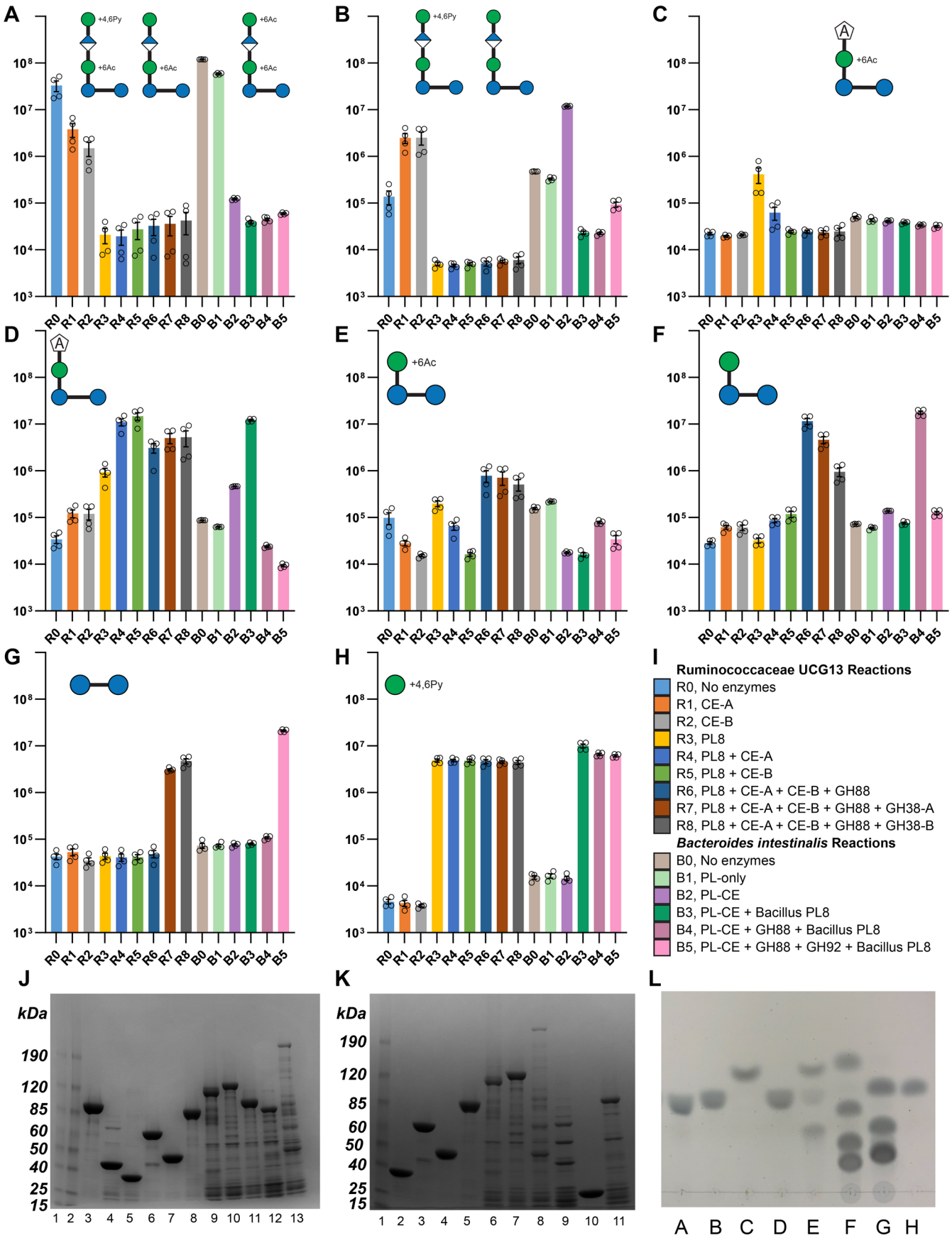


Extended Data Fig. 3 | Culture supernatant contains enzymes capable of depolymerizing xanthan gum, while intracellular contents are required for complete saccharification. Representative thin layer chromatography of xanthan gum incubated with different fractions of an active xanthan gum culture (supernatant, washed cell pellet, lysed cell pellet, or lysed culture). Negative controls were prepared by heating fractions at 95 C for 15 minutes prior to initiating with xanthan gum. EDTA was added to a final concentration of ~50 mM to determine the necessity of divalent cations for enzyme activity. Strong color development in circles at baseline is undigested polysaccharide while bands that migrated with solvent are digested oligosaccharides and monosaccharides. Repeated experiments showed similar results.



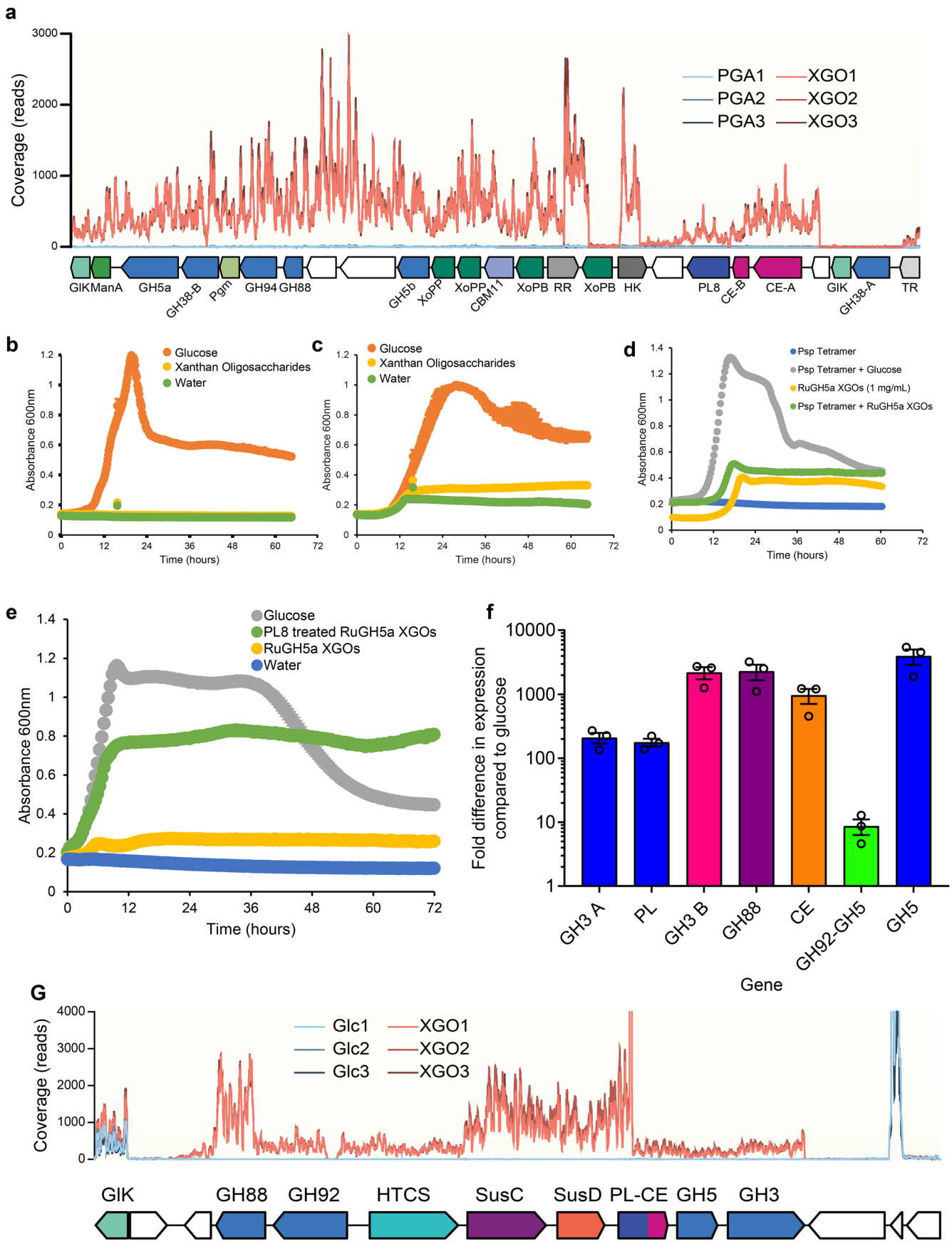
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Activity of R.UCG13 GH5 enzymes on various polysaccharides. a, SDS-PAGE gel of purified GH5 constructs and their resultant activity as assessed by TLC on **b**, xanthan gum, **b-c**, carboxymethyl cellulose (CMC), **c**, hydroxyethyl cellulose (HEC), **d**, barley β-glucan, **d-e**, yeast β-glucan, **e**, tamarind xyloglucan, **f**, xylan, and **f-g**, wheat arabinoxylan. Enzymes are 1, *Ru*GH5b (GH5 only); 2, *Ru*GH5b (GH5 with UD-A); 3, *Ru*GH5b (GH5 with UD-A/B); 4, *Ru*GH5b (full protein); 5, *Ru*GH5a (GH5 only); 6, *Ru*GH5a (GH5 with UD-A); 7, *Ru*GH5a (GH5 with UD-A/B); 8, *Ru*GH5a (GH5 with UD-A/B/C); 9, *Ru*GH5a (full protein); 10, replicate of 8. Strong color development in circles at baseline is undigested polysaccharide while bands or streaking that migrated with solvent are digested oligosaccharides and monosaccharides. Although minor streaking appears in some substrates due to residual oligosaccharides, comparing untreated substrate with enzyme incubated substrate allows determination of enzyme activity. *Ru*GH5a constructs with all 3 UDs (8-10) showed clear activity on XG. Repeated experiments showed similar results.



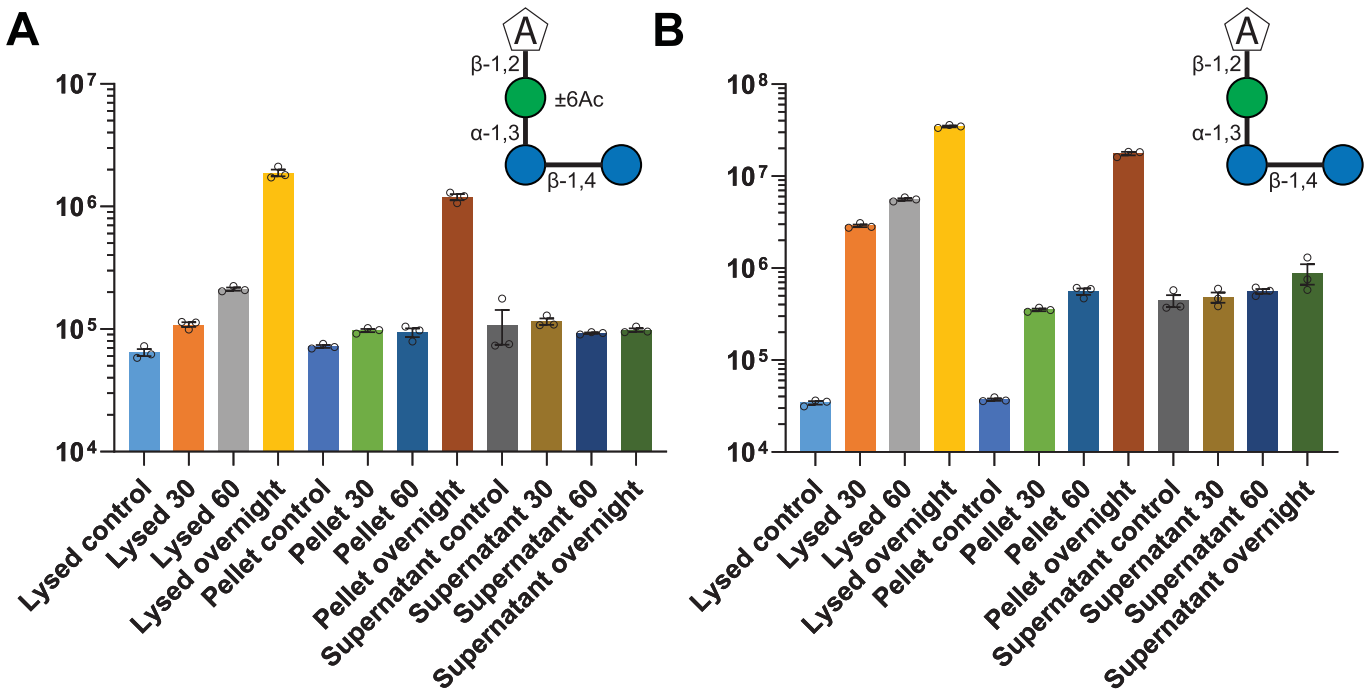
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Activity of R.UCG13 and B. intestinalis enzymes. Enzyme activities were tested by incubating different combinations of enzymes with 2.5 mg/mL XGOs (generated by *Ru*GH5a) at 37 °C for 16 hours. LC-MS analysis was used to track relative increases and decreases of intermediate oligosaccharides with the addition of enzymes, verifying their abilities to successively cleave XG pentasaccharides to their substituent monosaccharides. Integrated extracted ion counts (mean of n=4 distinct enzyme reactions and error bars displaying SEM; individual reactions shown as open circles) that correlate with compound abundance are shown for **a**, acetylated pentasaccharide (M-H ions: 883.26, 953.26, 925.27), **b**, deacetylated pentasaccharide (M-H ions: 841.25, 911.25), **c**, acetylated tetrasaccharide (2M-H ion: 1407.39), **d**, tetrasaccharide (M-H ion: 661.18), **e**, acetylated trisaccharide (M+Cl ion: 581.15), **f**, trisaccharide (M+Cl ion: 539.14), **g**, cellobiose (M+Cl ion: 377.09), and **h**, pyruvylated mannose (M-H ion: 249.06). Reactions were carried out using xanthan oligosaccharides produced by the *Ru*GH5a to test activities of the R.UCG13 (R0-R8) and *B. intestinalis* (B0-B5) enzymes. R.UCG13 enzymes were tested in reactions that included (R0) no enzyme, (R1) R.UCG13 CE-A, (R2) R.UCG13 CE-B, (R3) R.UCG13 PL8, (R4) R.UCG13 PL8 and CE-A, (R5) R.UCG13 PL8 and CE-B, (R6) R.UCG13 PL8, both CEs, and GH88, (R7) R.UCG13 PL8, both CEs, GH88, and GH38-A, (R8) R.UCG13 PL8, both CEs, GH88, and GH38-B. *B. intestinalis* enzymes were tested in reactions that included (B0) no enzyme, (B1) Bi PL-only, (B2) Bi PL-CE, (B3) Bi PL-CE and Bacillus PL8, (B4) Bi PL-CE and GH88 and Bacillus PL8, (B5) Bi PL-CE, GH88, and GH92 and Bacillus PL8. **i**, Legend of enzymes included in each reaction. Recombinant enzymes were purified and analyzed for expression and purity by SDS-PAGE. Proteins generally expressed well with a single dense band indicating overexpression of the enzyme at its predicted molecular weight as compared to a size ladder. Exceptions included the R.UCG13 GH88 and CE-A, both of which had bands at the predicted enzyme size but also showed bands of comparable density at other sizes resulting from either proteolysis or co-purification of undesired *E. coli* proteins. **j**, SDS-PAGE gel of purified enzymes with 4.5 μg loaded, including (1-2) ladder, (3) *B. intestinalis* GH3, (4) *B. intestinalis* GH5, (5) *B. intestinalis* PL-only, (6) *B. intestinalis* PL-CE, (7) *B. intestinalis* GH88, (8) *B. intestinalis* GH92, (9) R. UCG 13 GH38-A, (10) R.UCG13 GH38-B, (11) R.UCG13 GH94, (12) R.UCG13 PL8, (13) R.UCG13 CE-A. Repeated purifications showed similar results. **k**, SDS-PAGE gel of purified enzymes with 4.5 μg loaded, including (1) ladder, (2) *B. intestinalis* PL-only, (3) *B. intestinalis* PL-CE, (4) *B. intestinalis* GH88, (5) *B. intestinalis* GH92, (6) R.UCG13 GH38-A, (7) R.UCG13 GH38-B, (8) R.UCG13 CE-A, (9) R.UCG13 GH88, (10) R.UCG13 CE-B, (11) R.UCG13 PL8. Repeated purifications showed similar results. **l**, TLC analysis showed that R.UCG13 GH94 and *B. intestinalis* GH3 are active on cellobiose. From left to right lane show (A) *Ru*GH5b (full protein), (B) *Ru*GH5a (full protein), (C) *B. intestinalis* GH3, (D) *B. intestinalis* GH5, (E) R.UCG13 GH94, (F) odd standards, (G) even standards, (H) cellobiose. Odd and even standards are maltooligosaccharides with 1, 3, 5, and 7 hexoses or 2, 4, and 6 hexoses, respectively. While the *B. intestinalis* GH3 only produced one product, the R.UCG13 GH94 produced two, one matching the approximate Rf of glucose while the other had a much lower Rf which presumably is phosphorylated glucose (matching the known phosphorylase activity of the GH94 family). Repeated experiments showed similar results.

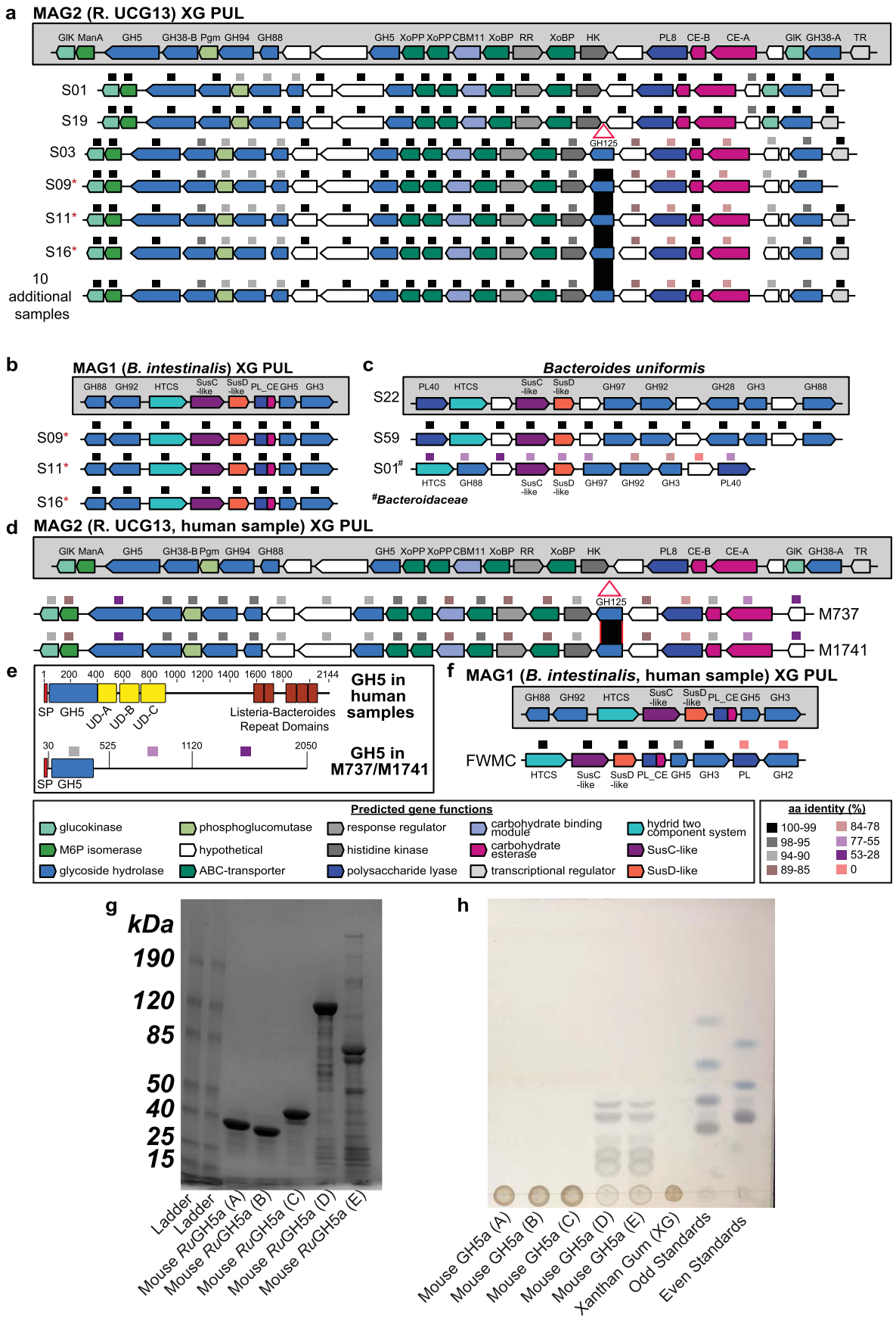


Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Expression and growth analysis of isolates and cultures grown on various substrates. a, Traces of RNA-seq expression data from n=3 biological replicates of the Sample 0 culture grown on either XG or polygalacturonic acid (PGA), illustrating overexpression of the XG PUL. **b**, *Bacteroides clarus* and **c**, *Parabacteroides distasonis* isolated from the Sample 0 culture did not grow on XGOs. **d**, *Bacteroides intestinalis* did not grow on tetramer generated with *P. nanensis* GH9 and PL8 (Psp Tetramer) even in the presence of 1 mg/mL *Ru*GH5a generated XGOs to activate the PUL. Growth on glucose confirmed that the Psp Tetramer was not inherently toxic to cells. **e**, *Bacteroides salyersiae* grew on lyase treated XGOs but not on full XGOs and **f**, overexpressed a PUL homologous to the *B. intestinalis* PUL when grown on lyase treated XGOs. All substrates were used at 5 mg/mL unless otherwise noted. Growths are n ≥ 2 (for **b** and **c**, n=2 biological replicates on xanthan oligosaccharides and n=4 biological replicates on water and glucose; for **d** n=3 biological replicates, for **e**, n=6 biological replicates, for **f** n=3 biological replicates), error bars show SEM (in most cases, smaller than the marker). **g**, Traces of RNA-seq expression data from n=3 biological replicates of *B. intestinalis* grown on either glucose (Glc) or XG oligosaccharides (XGOs), illustrating overexpression of the XGO PUL.

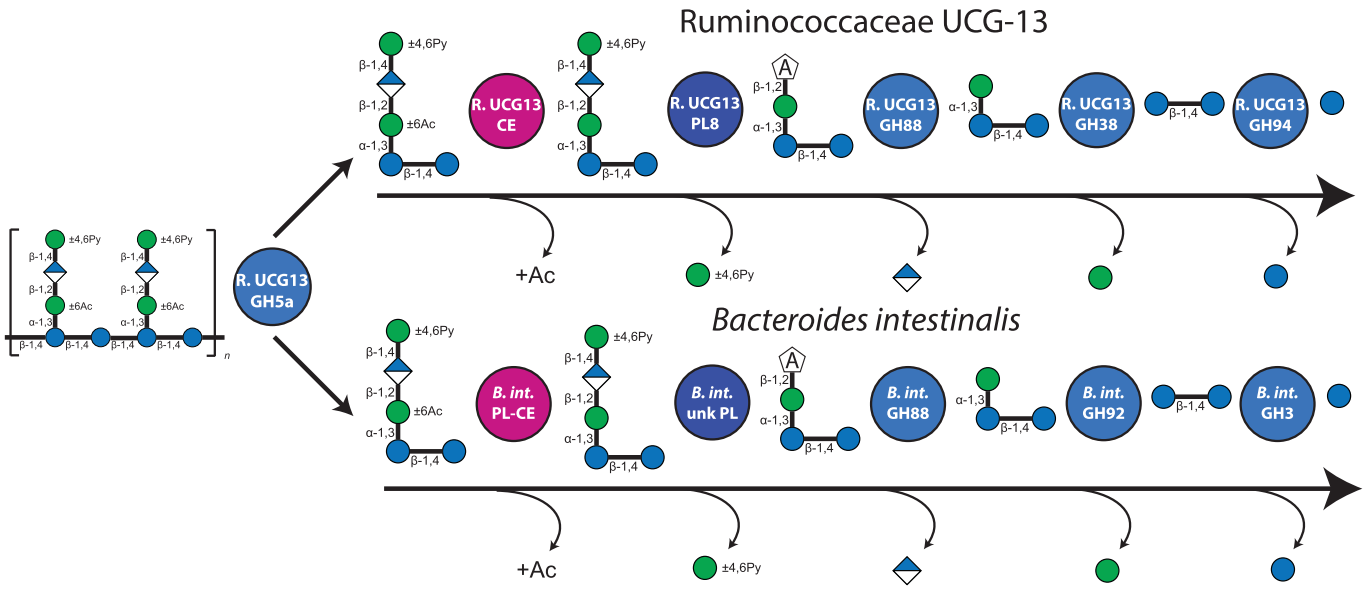


Extended Data Fig. 7 | *B. intestinalis* has a lyase that removes the terminal mannose from XGOs. *B. intestinalis* was grown on XGOs then fractionated into supernatant, washed cell pellets (Pellet), and washed then lysed cells (Lysed), then incubated with fresh XGOs. Timepoints were taken at 30 minutes, 60 minutes, and overnight and analyzed for the formation of **a**, lyase-produced acetyl tetrasaccharide and **b**, tetrasaccharide. Integrated extracted ion counts (mean of n=3 distinct enzyme reactions and error bars displaying SEM; individual reactions shown as open circles) that correlate with compound abundance are shown. Relative to boiled controls, lysed cells and pelleted cells showed clear increases in lyase-produced tetrasaccharide.



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Metagenomic and bioinformatic analysis of additional xanthan gum degrading loci. a, Metagenomic sequencing of additional 16 cultures (S, human fecal sample) that actively grew on and degraded xanthan gum revealed two architectures of the R.UCG13. The more prevalent locus contained a GH125 insertion. The 10 additional samples with this locus architecture include: S22, S25, S39, S43, S44, S45, S49, S53, S58, and S59. **b**, The *B. intestinalis* xanthan locus was present in 3 additional cultures. **c**, Additional members of the Bacteroideceae family harbor a PUL with a GH88, GH92 and GH3 that could potentially enable utilization of XG-oligosaccharides. **d**, The GH125-containing version of the R.UCG13 xanthan locus was detected in two mouse fecal samples (M, mouse fecal sample). **e**, Comparison of the human and mouse *Ru*GH5a aa sequence, showing the annotated signal peptide (SP), GH5 domain, three uncharacterized domains (UDs) with some homology to sugar-binding proteins, and multiple Listeria-Bacteroides repeat domains. **f**, Genetic organization and aa identity (%) between the *B. intestinalis* xanthan locus in the human sample and a PUL detected in a fracking water microbial community (FWMC) using LAST-searches. For a-f, the gray box delineates gene architecture and sequences derived from the human-derived Sample 0 culture. **g**, SDS-PAGE gel of purified enzymes with 4.5 μg loaded, including ladder and the different mouse *Ru*GH5a constructs. A, B, and C are all versions of the GH5 domain alone, D is a construct designed to terminate at a site homologous to the last UD in the human *Ru*GH5a, and E is a full-length construct of the mouse *Ru*GH5a. Repeated purifications showed similar results. **h**, TLC of each mouse *Ru*GH5a construct incubated with XG and also odd (1, 3, 5, and 7 residues) and even (2, 4, and 6 residues) malto-oligosaccharide standards. The GH5-only constructs did not degrade XG but constructs D and E (with regions homologous to the human *Ru*GH5a UDs) were able to hydrolyze XG. Repeated experiments showed similar results.



Extended Data Fig. 9 | Enzymatic model of xanthan degradation. The *Ru*GH5a enzyme initially depolymerizes XG into XGOs that are then saccharified by additional enzymes in R.UCG13 or *B. intestinalis*. Although the identified *B. intestinalis* PL-CE displayed acetylase functionality, we were unable to detect lyase activity with this construct. Lyase activity on XGOs was demonstrated with *B. intestinalis* pelleted or lysed cells suggesting that one or more unknown lyases can remove the terminal mannose from the pentasaccharide. Lyase activity generates a 4,5 unsaturated glucuronic acid (represented by a pentagon with an A), which is converted back to a standard glucuronic acid upon hydration and release by a GH88. The schematic uses standard nomenclature for glycans (blue circles, glucose; green circles, mannose; white and blue diamonds, glucuronic acid; Ac, acetylation; Py, pyruvylation) with enzymes represented as circles with interior labels.

nature portfolio

Corresponding author(s): Eric C Martens

Last updated by author(s): Feb 1, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	MinKNOW software v3.6.5; Bio-stack software (v2.02.3); Gen5 (v1.08.4)
Data analysis	Mothur (v.1.40.5); R (versions 3.6.3 and v.4.0.2) with packages reshape2 v1.4.4, RColorBrewer v1.1-2, ggplot2 v3.3.0, and edgeR v3.34.0; Microsoft® Excel® for Microsoft 365; Guppy v3.2.10; Filtlong v0.2.0; Cutadapt v1.3; FASTX-Toolkit v.0.0.14; metaSPAdes v3.10.1; MetaBAT v0.26.3; CheckM v1.1.3; MiGA (MiGA Online (microbial-genomes.org)); GTDB-Tk v1.4.0; PROKKA v1.14.0; CompareM v0.1.2; CANU v2.1.1; minimap2 v2.17; Racon v1.4.13; Medaka v1.1.3; BUSCO v4.1.4; FastANI v1.1; Trimmomatic v0.36; SortMeRNA v.2.1b; kallisto v0.43.1; Proteome Discoverer (v2.1, Thermo Scientific); MestReNova; MassHunter Qualitative Analysis v10.0; BowTie2 v2.3.5.1; htseq-count (release_0.11.1); Integrative Genomics Viewer (https://software.broadinstitute.org/software/igv/); magic-blast v1.5.0; bedtools v2.29.0; LAST software package (version 1066); GraphPad Prism v9.3.0; Barrnap v0.9; BLAST+ v2.2.31

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All sequencing reads have been deposited at the European Nucleotide Archive under BioProject PRJEB44146. All annotated MAGs are publicly available via Figshare

(DOIs: 10.6084/m9.figshare.14494602, 10.6084/m9.figshare.14494536, 10.6084/m9.figshare.14494677, 10.6084/m9.figshare.14494683 and 10.6084/m9.figshare.14494689). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD031522.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No power calculations were made prior to experiments to arrive at sample size for assays, including in vivo experiments. Sample sizes for growth curves, enzyme reactions, expression analysis, and in vivo experiments were typically performed at least in triplicate, which in past studies by our groups and others have been sufficient to detect significant differences between experimental conditions. Specific experiments were performed only in duplicate, either because of limited material availability and/or results of experiments performed in this manner had already yielded significant and informative results.
Data exclusions	No data points were excluded from analysis.
Replication	All attempts to replicate experiments were successful and this is stated as appropriate in the corresponding figure legends or methods section (including information on replicate numbers).
Randomization	For studies using germfree mice, available mice were randomly assigned to Group 1 or 2. To limit costs and handling, Group 1 and 2 were assigned to pre-existing cages of mice (i.e. litter mates). This may have resulted in cage-effects which are discussed within the paper.
Blinding	No investigators were blinded to experimental data.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Female 7-8 week old germfree Swiss Webster mice were utilized. Subjects were housed in Iso-Positive Tecniplast cages and light cycles were 12 on, 12 off, with lights on between 6 AM and 6 PM. Temperature was maintained at 70-73° F and humidity was maintained at 40-60%. This is stated in the corresponding section of Methods.
Wild animals	No wild animals were used in this study.
Field-collected samples	No field-collected samples were used in this study.
Ethics oversight	All experiments involving animals were approved by the University Committee on Use and Care of Animals at the University of Michigan (NIH Office of Laboratory Animal Welfare number A3114-01) and overseen by a veterinarian.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	For screening of fecal samples for xanthan degradation, participants ranged in age from 17 to 29 years old, with a median age of 19 years old (for HUM00094242 and HUM00118951). For fecal samples used in humanized mouse experiments (HUM00141992), participants ranged in age from 19 to 38 years old, with a median age of 30 years old. Other than exclusion criteria used during the recruitment phase, no population characteristics were recorded.
Recruitment	For HUM00094242 and HUM00118951, study participants were recruited through the Authentic Research Sections of the introductory biology laboratory course at the University of Michigan (BIO173). Because study participants were recruited as part of an introductory biology lab course there may be self-selection bias although this is not expected to influence the results. Individuals with self-reported history of inflammatory bowel syndrome, inflammatory bowel disease, or colorectal cancer were excluded from the study, as were individuals who had taken antibiotics within the last 6 months. Pre- or probiotic usage was not an exclusion criterion for the study, nor was the amount of fiber already being consumed. All participants gave written, informed consent prior to participating in the study. Participants under the age of 18 were granted permission by a parent or legal guardian. For HUM00141992, only legal adults over the age of 18 were allowed to participate. Individuals with self-reported history of inflammatory bowel syndrome, inflammatory bowel disease, or colorectal cancer were excluded from the study, as were individuals who had a gastrointestinal infection or had taken antibiotics within the last 6 months. Immunocompromised or immunosuppressed individuals were also excluded from the study. Human subjects volunteered to donate their fecal matter and verbally consented to study participation. Because subjects volunteered for this study, there may be self-selection bias although this is not expected to influence the results.
Ethics oversight	This study was approved by the Institutional Review Board (IRBMED) of the University of Michigan Medical School (HUM00094242, HUM00118951, and HUM00141992) and was conducted in compliance with the Helsinki Declaration.

Note that full information on the approval of the study protocol must also be provided in the manuscript.