Mucin foraging enables Akkermansia muciniphila to compete against other microbes in the gut and to modulate host sterol biosynthesis

Lauren Davey
Duke University Medical Center
Per Malkus
Duke University Medical Center
Max Villa
Duke University Medical Center
Lee Dolat
Duke University Medical Center
Zachary Holmes
Duke University
Jeffrey Letourneau
Duke University https://orcid.org/0000-0002-5878-0914
Eduard Ansaldo
UC Berkeley https://orcid.org/0000-0002-7128-3373
Lawrence David
Duke University https://orcid.org/0000-0002-3570-4767
Greg Barton
University of California Berkeley
Raphael Valdivia (raphael.valdivia@duke.edu)
Duke University Medical Center https://orcid.org/0000-0003-0961-073X

Article

Keywords:

Posted Date: March 31st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1475049/v1

Mucin foraging enables *Akkermansia muciniphila* to compete against other microbes in the gut and to modulate host sterol biosynthesis

Lauren Davey[1,2], Per N Malkus[1,2], Max Villa[1,2], Lee Dolat[1,2], Zachary C. Holmes[1,2], Jeff Letourneau[1,2], Eduard Ansaldo[3], Lawrence A. David[1,2], Gregory M. Barton[3], and Raphael H. Valdivia[1,2]

[1]Department of Molecular Genetics and Microbiology, Duke University, Durham, North Carolina, USA

[2]Duke Microbiome Center, Duke University, Durham, North Carolina, USA

[3]Division of Immunology & Pathogenesis, Department of Molecular and Cell Biology, University of California, Berkeley, California, USA

Akkermansia muciniphila, a prominent member of the gastrointestinal tract (GI) microbiota, uses mucins as a sole source of carbon and nitrogen. *A. muciniphila* is considered a next-generation probiotic because its abundance in humans positively correlates with protection from metabolic syndrome and obesity. However, *A. muciniphila* is intractable to genetic analysis and thus the molecular mechanisms underlying the metabolism of mucin, its colonization of the GI tract, and its impact on host physiology are poorly understood. Here, we developed and applied transposon mutagenesis to identify *A. muciniphila* factors important for the use of mucin and determined that mucin degradation products accumulate in internal compartments through a process that requires pili and a periplasmic protein complex. We further determined that the degradation of mucin and related proteoglycans is important for colonization of the GI by *A. muciniphila* but only in the context of competing microbes. In germ free mice, mucin use by *A. muciniphila* repressed the expression of host genes required for mevalonate and cholesterol biosynthesis in the colon, providing a molecular link between *A. muciniphila* metabolism of mucins, the regulation of lipid homeostasis and potential probiotic activities.

Akkermansia accumulates mucins in intracellular compartments

Intestinal mucins are large, highly glycosylated proteins that provide a physical barrier between the host epithelia and the microbiota[1]. The complex linkages of mucin glycans and terminal sulfation make them a challenging substrate for microbes to degrade. Nonetheless, a range of intestinal microbes display enhanced growth in media supplemented with mucin[2]. *A. muciniphila* is particularly adept at using mucins, reaching a high optical density when cultured in liquid medium with gastric mucin as a sole carbon and nitrogen source compared to other intestinal microbes (**Fig. S1a**). The *A. muciniphila* genome encodes multiple putative glycoside hydrolases and enzymes that may degrade mucins, some of which have been characterized in vitro[3–10]. However, because *Akkermansia* spp. are intractable to molecular genetic manipulation, little is known as to how mucins are scavenged, transported, cleaved, or what role the metabolism of mucin plays in the colonization or physiology of their hosts.

A. muciniphila, like other members of the *Planctomycetes-Verrucomicrobia-Chlamydiae* (PVC) super-phylum, has an enlarged periplasmic space termed a paryphoplasm[11,12]. In ocean dwelling

Planctomyces, large chondroitin sulfate glycans are transported across the outer membrane[13], while *Gemmata obscuriglobus* can take up large proteins into the paryphoplasm[14]. Similarly, fresh-water *Verrucomicrobes* take up laminarin and xylan polysaccharides[15] We predicted that *A. muciniphila* may use a strategy similar to other members of the PVC super-phylum to acquire glycoproteins. After incubating live *A. muciniphila* with fluorescein (FL)-labeled pepsin-digested mucin glycans we observed an accumulation of the labeled mucins or its degradation products in intracellular compartments as assessed by super resolution STED microscopy (**Fig. 1a, Movie S1**). The formation of these mucin glycan-containing intracellular structures, which we term mucinosomes, appear to be specific to *Akkermansia* spp., as we did not observe them in *Bacteroides thetaiotaomicron*, which can also metabolize mucin (**Fig. S1b**).

To examine the kinetics of mucin uptake, we pulsed *A. muciniphila* with FL-labeled mucin and analyzed fluorescence associated with bacteria over time by flow cytometry. The association with FL-mucin occurred rapidly and peaked by 3h post labeling, followed by a subsequent decrease between 6 and 24h (**Fig. 1b**). In parallel, we used SDS-urea agarose polyacrylamide gel electrophoresis (SDS-agPAGE) to visualize high molecular weight (MW) FL-glycans associated with bacteria. Consistent with the flow cytometric findings, we observed an accumulation of FL-labeled mucins peaking at 3h, followed by a decrease in the size and abundance of high MW proteoglycans indicative of degradation (**Fig. 1c**). In contrast, the apparent MW of FL-mucin precipitated from the corresponding culture supernatants remained constant (**Fig. 1c**).

Next, we used live-cell imaging to monitor the uptake of FL-mucin in real-time. The distribution of FL-mucin within the cells was consistent with what we observed in fixed cells, with accumulation occurring at the cell pole and at the midzone (**Fig. 1d**). Mucinosome formation occurred within minutes and was a highly dynamic process with an initial accumulation of FL-mucin-rich foci forming throughout the bacterial cell surface (**Movie S2**). These structures did not form after incubation with FL-labeled dextran (**Fig. S1c**) and were diminished in the presence of the ionophore carbonyl cyanide 3-chlorophenylhydrazone (CCCP) which dissipates the cell's proton motive force (**Fig. S1d-e, Movie S2**), suggesting that the acquisition of mucin glycans by *A. muciniphila* is a selective and energy-dependent process.

Development of a system for transposon (Tn) mutagenesis in *A. muciniphila*

To define the genetic requirements for growth on mucin, we developed a system for Tn mutagenesis after identifying suitable selectable markers, defining conditions for DNA conjugation with an *E. coli* donor, and codon-optimizing the Himar1c9 transposase[16] for expression in *Akkermansia* (**Fig. S2a-b**). Transposition events, as assessed by the number of chloramphenicol (Cm) resistant mutants, occurred at a low frequency, and required prolonged incubation. The absence of plasmid sequences in *A. muciniphila* Cm^R colonies confirmed that transposition had occurred (**Fig. S2c**) and a Southern blot analysis of total DNA extracted from these strains indicated that most mutants had a single Tn insertion (**Fig. S2d**).

We assembled two mutant libraries, one consisting of a mixed pool of Tn mutants and one where individual Tn mutants were arrayed in 96-well plates. The Tn insertion sites in both pools were identified by DNA sequencing on an Illumina platform as previously described[17]. The pooled Tn mutant library consisted of 2680 unique insertions in 721 genes (33% of the total CDS), 198 intergenic regions, and 5 tRNAs. The arrayed Tn library consisted of 1063 unique insertions in 406 genes, 58 intergenic regions, and two tRNA genes. We then used a Cartesian pooling

strategy[18] to map the locations of the majority of the Tn insertions to individual mutants in the arrayed collection.

Broad amino acid prototrophies and multiple proteins of unknown function are required for *A. muciniphila* to grow in mucin

To identify genes required for the metabolism of mucins, Tn mutants from the mixed pool were grown for eight generations in a defined medium with gastric mucin as the sole source of carbon and nitrogen[19]. The relative abundance of each Tn insertion was determined using a modified version of INSeq[16] to account for the relatively low density of Tn insertions in the mutant libraries. The impact of mutations in each gene on bacterial growth was expressed as the Log_2 fold change between the normalized abundance in the input inoculum pool and the output pool, where negative values are indicative of mutants with growth defects. We observed a significant decrease in the abundance of mutants with Tn insertions in genes required for amino acid biosynthesis, particularly branched chain amino acids and Arg (**Fig.1e, Fig. S3a** and **Table S1**). This result was not surprising given that these amino acids are not abundant in the protein core of GI mucins, which are primarily composed of Ser, Thr, and Pro[20]. Mutants with Tn insertions in genes required for assimilatory sulfate metabolism were also significantly depleted, as they are required for the de novo generation of Cys and Met[21]. The addition of protein hydrolysates to the mucin medium rescued most mutants in the pathways for branched chain amino acids, sulfate reduction, and phenylalanine biosynthesis (**Fig. S3b** and **Table S1**). These results indicate that *A. muciniphila* relies on de novo biosynthesis of amino acid during growth in mucin, presumably because some amino acids in the protein backbone become rate limiting to support the bacterium's anabolic needs.

We anticipated that glycoside hydrolases (GH) would be required for *A. muciniphila* to grow on mucin. *A. muciniphila* strain Muc^T is predicted to encode 60 GHs, belonging to 24 different families. Our Tn mutant collection included insertions in genes encoding 38 GHs, representing 20 families[22]. Following growth in mucin medium, only 8 GHs had a Log_2 FC decrease in abundance > 2, indicating that most of the enzymes (30/38) were not essential for growth in mucin medium. Several mutants displayed a slight growth advantage, including previously characterized enzymes such as a β-galactosidase (Amuc_1686, GH35)[5], a β-hexosaminidase (Amuc_2136, GH20)[23,24], a β-N-acetyl hexosaminidase (Amuc_0868, GH20)[4], and a predicted mucin binding chitinase (Amuc_2164, GH18)[25]. In contrast, mutations in a sialidase (Amuc_1835, GH33)[23] and a fucosidase (Amuc_1120, GH95)[23], both of which are likely involved in cleaving the terminal sugars on mucin glycans, and a different β-galactosidase (Amuc_0290, GH2), an α-N-acetylglucosaminidase (Amuc_1220, GH89), and a β-hexosaminidase (Amuc_2148, GH20), displayed profound growth defects in mucin medium (**Fig. S3c** and **Table S1**) and to a lesser extent in synthetic medium (**Fig. S3d**).

The overlap between genes required for mucin utilization by *A. muciniphila* as determined by metabolic modeling[26] and INSeq analysis is low, suggesting that most of the enzymes predicted to have roles in mucin degradation are functionally redundant in liquid culture. For instance, mutants with Tn insertions in genes predicted to be involved in the transport and metabolism of mucin monosaccharides such as fucose (Amuc_0146, Amuc_0392, Amuc_1832-33), sialic acid (Amuc_0825, Amuc_1547), and galactose (Amuc_0829, Amuc_1187, Amuc_1086, Amuc_0969) were only modestly depleted during growth in mucin medium. An exception was mutants defective for N-acetyl glucosamine (GlcNAc) metabolism (Amuc_1220Amuc_0030)

which was expected as *A. muciniphila* lacks GlmS, an enzyme required for glucosamine-6-phosphate formation and peptidoglycan synthesis[27]. We also determined that several genes involved in the fermentation of pyruvate to propionic acid via the succinate pathway[26,28], including the propionyl-CoA:succinate CoA transferase Amuc_0206, and both subunits of succinyl-CoA ligase (Amuc_1712 and Amuc_1713), were preferentially required for growth in mucin medium (**Table S1**).

It is possible that the low representation of GH mutants among strains defective for the consumption of mucin is the result of trans complementation of GH activities by neighboring cells. To test if cross-feeding interactions in batch cultures led us to miss important GHs (and other factors) needed for mucin utilization, we performed Droplet Tn-seq[29] to analyze *A. muciniphila* Tn mutants grown in microdroplets. A comparison of the abundance of mutants grown in batch culture versus microdroplets indicated that the growth defects in mucin medium were largely independent of whether the bacteria were grown in isolation (**Fig. S3e** and **Table S1**), suggesting that any cross-feeding is limited. Exceptions to this observation include genes involved in capsule production and a predicted sialidase, fucosidase, and sulfatase - all enzymes that can act on the terminal motifs of mucin glycans. These results are consistent with prior findings indicating fucosidase activity associated with the *Akkermansia* outer membranes [30] and our own observations that while some trimming may occur extracellularly, mucin degradation is primarily associated with intact bacteria (**Fig 1c**).

We performed an RNAseq analysis of *A. muciniphila* cultured in the mucin medium used for INSeq selections and in the mucin-free synthetic medium used to propagate the Tn mutant libraries. In response to mucin, 103 genes were significantly upregulated (fold change > 4, adjusted P value < 0.05), and a gene set enrichment analysis showed that pathways involved in glycan degradation and galactose metabolism dominated this response (**Fig. 1f**) as previously observed[26]. Genes that were highly expressed in response to mucin included the O-glycopeptidase Amuc_0627[31], putative sulfatases (Amuc_1182, Amuc_0491), and 24 GH enzymes. The correlation between the genes that were required for growth in mucin medium and those that were strongly upregulated by mucin was low (**Fig. 1f-g**). These observations indicate that, unlike the polysaccharide utilization (Pul) genes of *Bacteroidetes*[32], the *Akkermansia* genes that are essential for growth in mucin are not preferentially induced by mucins.

Approximately 35% of the *A. muciniphila* genome encodes proteins with no predicted function[25]. We mapped Tn insertions to 226 genes annotated as either 'conserved hypothetical' or 'hypothetical' proteins, which accounts for 31.3% of all the mutated coding sequences in our pooled library. Mutations in 54/179 genes encoding hypothetical proteins that were either *Akkermansia* specific or had homologs restricted to the PVC superphylum led to decreases in abundance of > 2-fold during growth in mucin medium (**Fig. S4a**). For some of these genes, a BLAST search revealed homologs in related PVC members with more informative annotations (**Table S1**). Manual curation of these proteins of unknown function for associated Pfam[2] motifs and Conserved Domains[34] indicated that 19% contained either a tetratricopeptide repeat domain, which mediate protein-protein interactions[35], or were associated with pili assembly (**Fig. S4, Table S1**). Proteins potentially involved in pili or type II secretion included Amuc_1100 and Amuc_1102, which have been characterized for their anti-obesity activity[36] and whose structures have been solved[37,38]. Overall, these findings suggest that mucin utilization by *A. muciniphila* requires pili-like proteins and several poorly characterized hypothetical proteins whose expression is not regulated by mucin.

The ability of *Akkermansia* to metabolize mucin in the GI tract enables successful competition against other microbes

To identify *A. muciniphila* genes required to colonize the GI tract, we performed an INSeq analysis of the Tn arrayed mutant pool in four mouse models of colonization: germ-free (GF), mice stably colonized with the eight-member Altered Schadler flora (ASF), conventionally raised (CONV) mice, and Muc2$^{-/-}$ mutant mice lacking the most prominent intestinal mucin[39]. Mice raised with a conventional microbiota were pre-treated with tetracycline to displace their endogenous *Akkermansia*, which prevents the engraftment of the Muc^{T} *A. muciniphila* strain[40]. Mice were then orally gavaged with a pool of Tn mutants and cecal material was collected at 14 days. The relative fitness of the arrayed Tn mutants in the caecum or after growth in mucin medium was assessed by INSeq (**Fig. 2a**).

Our analysis indicated that approximately half of the genes whose disruption led to a > 2-fold decrease in abundance in the GI tract of mice, were also required for growth in mucin (**Fig. 2a, d**). As the complexity of the microbiota increased from GF to CONV mice, additional genes became conditionally essential (**Fig. 2a-e**). As we had observed in mucin medium, Tn insertions in genes required for amino acid homeostasis were significantly overrepresented among mutants that failed to colonize the GI tract (**Fig. 2b, c**), especially in branched chain amino acids. In addition, mutants defective for Ala, Asp, Glu, and Arg biosynthesis were depleted in ASF and CONV mice (**Fig. 2c-e**). Mutants in the glyoxylate pathway, particularly components of the Gly cleavage system (Amuc_0446, Amuc_0447, Amuc_0448), also had significant growth defects in both conventional and Muc2$^{-/-}$ mice. The Gly cleavage system generates 5,10-methylene-tetrahydrofolate, a one carbon donor used in Ser biosynthesis, which may be particularly important for colonization of Muc2$^{-/-}$ mice because they lack this Ser-rich mucin. Overall, these findings indicate that the ability to synthesize key amino acids, especially in the context of an complex microbiota, is required for *A. muciniphila* to successfully colonize the GI.

Metabolic and cellular pathways overrepresented in *A. muciniphila* mutants with fitness defects in mice indicated that Tn insertions in central processes such as transcription, protein metabolism, and RNA metabolism led to similar growth defects in the GI tract as in mucin medium (**Fig. 2c**). In contrast, mutations in translation and nucleotide biosynthesis displayed more substantial growth defects in CONV and Muc2$^{-/-}$ mice, indicative of the additional metabolic pressures as the competing microbiotas become more complex. Several genes were required for survival in the GI, but not for growth in mucin medium. These included putative exopolysaccharide and capsule biosynthesis genes, such as Amuc_2077, Amuc_1412, and Amuc_1413, and a locus harboring glycosyl transferases belonging to the GT2, GT11, and GT8 families (Amuc_0939 to Amuc_0945) (**Fig. 2f**), which are orthologous to the *Bacillus* biofilm genes *epsH* and *epsJ* [41]. These findings are consistent with *A. muciniphila* cell surface modifications playing a prominent role in intestinal colonization.

We identified eight genes that were selected against in mucin medium and GF mice but rescued in mice with a microbiota (**Fig. 2d**). Three genes mapped to components of the assimilatory sulfate reduction (ASR) pathway (**Fig. 2g**), which is required for the reduction of sulfate to hydrogen sulfide and the biosynthesis of cysteine, methionine, and sulfur containing metabolites. The growth defect of these ASR^{-} mutants was partially complemented in GF mice, indicating that *A. muciniphila* can harvest hydrogen sulfide or cysteine from the host or its diet. In both ASF and CONV colonized mice, ASR mutants had a slight growth advantage (**Fig. 2g**), but this

was context dependent as ASR mutants displayed significant growth defects in Muc2$^{-/-}$ mice (**Fig 2g**).

Tn insertions in hypothetical genes accounted for 23.6% (48/203) of the mutants with strong growth defects (Log_2 fold-change > 5) in GF mice and 27.3% (67/245) in CONV mice. Many of the hypothetical proteins required for growth on mucin medium were also required to colonize the GI tract, with the greatest growth defects occurring in genes encoding proteins with predicted N-methyl domains, a hallmark of pili proteins[42], or TPR domains (**Fig. S4b-c**).

Two genetic loci define a mucin transport system in *A. muciniphila*

Amuc_0544 is a predicted secreted protein of unknown function found only in *Akkermansia* and other *Verrucomicrobia*. The abundance of TPR domains in Amuc_0544 is reminiscent of the polysaccharide transport protein SusD in the *Bacteroides* Starch Utilization System (SUS)[43]. *Amuc_0543* and *Amuc_0544* are part of a larger gene cluster that includes two putative ExbD biopolymer transport proteins (Amuc_546, Amuc_547), and an ExbB-like proton channel (Amuc_0548). ExbBD are components of TonB dependent transporters (TBDT), also found in SUS, which transduce the proton motive force to transport substrates across the outer membrane[44]. Reads mapped by RNAseq indicate that the region is likely expressed as at least two mRNA transcripts (**Fig. S5a**). We retrieved *A. muciniphila* strains with Tn insertions in *Amuc_0543* and *Amuc_0544* and confirmed that they were impaired for growth in mucin medium (**Fig. 3a, d, S5c**), leading us to name the genomic region spanning Amuc_0543 to Amuc_0550 (**Fig. 3a, 2a**) as Mucin Utilization Locus I (*MUL1*) (**Fig. 3a**).

A second major group of genes required for mucin utilization were annotated as encoding pili, including type IV-like pili proteins encoded by the *Amuc_1098* to *Amuc_1102* locus[24,36]. These proteins are among the most abundant produced by *A. muciniphila*[24,30] which includes Amuc_1100, a TLR2 agonist[24,36] and modulator of host metabolism[36]. The structures of Amuc_1100[38,45] and Amuc_1102[37] indicate that these proteins are related to archaeal type IV pili. Amuc_1101 is annotated as the cell division protein FtsA, but it belongs to the PilM Pfam, suggesting a more likely role in pili biogenesis. We renamed this locus, spanning *Amuc_1098* to *Amuc_1102*, as *MUL2* (**Fig. 3a, 2a, S5b**) given that Tn insertions in *MUL2* genes exhibited strong growth defects in mucin (**Fig. 3c, Table S1**). Additional *A. muciniphila* potential *mul* loci are scattered throughout the genome (**Fig. 2a**), including genes encoding additional potential pili proteins with N-methylphenylalanine domains typical of type IV pilins[46] (Amuc_0394, Amuc_1524, Amuc_0745), further suggesting a broader role for pili in mucin acquisition (**Fig. S4b-c**).

The largest protein in the *MUL1* locus, Mul1A (Amuc_0544), is predicted to have a type I signal peptide and to localize to the periplasm. Given its relatively large size and multiple TPR domains, we hypothesized that Mul1A serves a structural role in the assembly of protein complexes important for the transport of mucin. We generated antisera specific for Mul1A and performed native immunoprecipitations from lysates of wild type *A. muciniphila* or *mul1A*::Tn mutants, coupled to quantitative mass spectrometry (**Fig. S5d**). Mul1A showed cleavage of the N-terminal signal sequence confirming that Mul1A is secreted across the cytoplasmic membrane. The top interacting protein was Amuc_0543, designated Mul1B, which efficiently co-precipitated with Mul1A as a complex (**Fig. 3h, Table S1**), indicating that both proteins function in the same step in the transport of mucin.

Peptides belonging to the pili subunit protein Amuc_1102 (*mul2A*), were also significantly enriched with Mul1A complexes suggesting that pili could play a key role in binding and acquisition of mucin, which is consistent with the disproportionately large number of Tn insertions in putative pili genes that lead to growth defects in mucins (**Fig. S4b-c**). Similar observations have been made in the related bacteria *Planctomycetes limnophila* and *Gemmata obscuriglobus*, which bind dextran glycans with pili-like fibers[12]. These findings also suggest that components of the *MUL1* and *MUL2* loci may cooperate to transport mucin fragments. We next tested if these *mul* mutants would be impaired for the import of mucins across the outer membrane. Indeed, Tn mutants in either *mul1A or mul1B* were unable to acquire or internalize FL-mucin (**Fig. 3f-g**). The loss of FL-mucin labeling in the *mul* mutants is not an indirect consequence of their inability to metabolize mucins as a Tn mutant in *amuc_0029*, encoding a UDP-glucose 4-epimerase in the Leloir pathway for galactose degradation[47], also exhibited an extreme growth defect in mucin medium (**Fig. 3e, Table S1**), but still displayed high intracellular staining with FL-mucin (**Fig. 3f-g**).

Mul1A interacted with several GH enzymes, two sulfatases (Amuc_0491 and Amuc_0451), and two disulfide reductases (Amuc_1841 and Amuc_0260), which are predicted to cleave sulfated glycans and reduce disulfide linkages, respectively. We also detected peptides from a putative sodium-solute symporter transporter (SSS), Amuc_0970, which in other microbes can import sialic acid[48] and galactose[49] across the inner membrane. The SSS transporter is a potential route for mucin components to cross the inner membrane, a function performed by major facilitator superfamily (MFS) transporters in other PUL systems[50]. A second putative transporter with a conserved outer membrane protein beta-barrel domain, Amuc_1687, was also enriched, suggesting a potential route for the transport of mucin glycan fragments (**Fig. 3h, Table S1**), although the precise mechanism remains to be determined. Comparison of the mass spectrometry data to the transcriptional analysis in response to mucin indicates that core components of the *MUL* systems, including *mul1A*, *mul1B,* and *mul2A* are constitutively expressed, while MUL-accessory proteins such as GHs are inducible (**Fig. 1f, 3h**). Finally, we detected peptides belonging to the Von Willebrand type D domain and the C-terminal cysteine knot domain of Muc5AC specifically co-precipitating with anti-Mul1A antisera (**Fig. S5e**), indicating that Mul1A or associated proteins can engage the non-glycosylated portion of the gastric mucin backbone.

Based on these observations, we propose that the MUL system functions analogously to the PUL systems of *Bacteroides* species (**Fig. 3i**) but with associated pili, perhaps in conjunction with additional *mul* loci not yet identified, to capture extracellular mucins that ultimately accumulate in mucinosomes.

A. muciniphila mutants unable to use mucin induce the expression of genes important in cholesterol biosynthesis in the colon

We next tested the requirement for the MUL system to colonize the GI after mono association challenges of GF or CONV mice (**Fig. 4a-4b**). Consistent with our INSeq results, mutants in *mul1A* and *mul2A* failed to colonize CONV mice (**Fig. 4a**). Unexpectedly, the *mul1A* mutant colonized GF mice as efficiently as wild type *A. muciniphila* (**Fig. 4b**). This suggests that mucin utilization is required for *A. muciniphila* to colonize in the context of other microbes, but in the absence of microbial competition including from wild type *Akkermansia*, it is dispensable to establish residence in the mouse GI.

High level colonization by *mul1A* mutants in GF mice indicates that *A. muciniphila* can access other sources of nutrients in the GI and is consistent with observations that Muc2$^{-/-}$ mice are highly colonized by an endogenous mouse *Akkermansia* strain[51]. We found that both wild-type and *mul1A* or *mul1B* mutants colonized the GI of antibiotic treated Muc2$^{-/-}$ mice at high levels (**Fig. S6a**). Thus, the requirement for mucin to colonize the GI is context dependent and *Akkermansia* can use other nutrients from the diet, host, or the microbiota. However, *mul* mutants were still outcompeted by wild-type *A. muciniphila* in Muc2$^{-/-}$ mice (**Fig. S6b**), suggesting that the MUL system is required to scavenge other related proteoglycans. Muc2$^{-/-}$ mice experience severe inflammation[51], which can increase the production of other metabolites not normally found in a healthy lower GI tract including fecal amino acids[52–54]. Paradoxically, Tn mutants in the ASR pathway, whose growth is normal in ASF and conventional mice, fail to survive in Muc2$^{-/-}$ mice (**Fig. 2g**), further emphasizing the unusual nutritional environment in the GI tract of these mucin-deficient mice and the influence other microbes can have on *A. muciniphila* colonization.

Although the nutrient source for *mul1A* mutants in GF mice is unknown, we reasoned that these mutants would produce different fermentation products compared to wild type *A. muciniphila*. We found that GF animals colonized with the *mul1A* mutant or wild type *A. muciniphila* differed in their cecal short chain fatty acids (SCFA) levels. GF mice colonized with wild-type *A. muciniphila* had higher levels of total acetate and propionate compared to mice colonized with the *mulA* mutant (**Fig. 4c, S7c**) but a lower ratio of acetate to propionate (**Fig. 4c**). Thus, mucin utilization could impact host physiology either as a direct result of mucin consumption or through metabolites produced by *A. muciniphila* during mucin metabolism.

A. muciniphila modulates transcriptional responses in the host, and differential expression of host genes in response to various components of *A. muciniphila*, including live and pasteurized cultures[55], purified surface proteins[36], and conditioned supernatants[56] have been reported. Wild type and *mul1A* mutants colonized the GI of female GF mice at similar levels (**Fig. S7a**), thus we chose to test the host's transcriptional responses to mucin foraging by *A. muciniphila* in these mice. An RNA-seq analysis of colonic tissue from GF mice colonized for two weeks revealed differentially expressed genes in the category of immunoglobulin genes, indicating a generalized defense response to bacteria (**Fig. 4d-e**). However, the most statistically significant difference in the colonic transcriptional response between GF mice colonized with wild type *A. muciniphila* versus *mul1A* mutants occurred in female mice and was in pathways associated with lipid metabolism (**Fig. 4e-f, Table S1**). Expression of the cholesterol biosynthetic genes Sqle, Hmgcs, Hmgcr, as well as most genes along the cholesterol/mevalonate synthesis pathway were repressed in mice colonized with wild type *A. muciniphila*, as were genes that modulate cholesterol uptake, including Ldlr and Pcsk9 (**Fig. 4d-g, S7b**). Publicly available single cell RNAseq datasets suggest that cholesterol/mevalonate genes are expressed in colonic intestinal epithelial cells and goblet cells[57] (**Fig. S7d**), which are the most likely to respond directly to *A. muciniphila* and its metabolites. These observations are consistent with reports that *A. muciniphila* metabolites, including SCFA, repress lipid metabolism in intestinal organoids[56], and that treatment with *A. muciniphila* lowers serum cholesterol in mice[36,55,58] and humans[59]. While control GF mice showed intermediate levels of expression (**Fig. S7b**) of sterol biosynthesis genes, the same pathways were significantly elevated in mice colonized with the *mul1A* mutant (**Fig. 4d-g, S7b**). This observation suggests that the colonic transcriptional response to *A. muciniphila* is directly influenced by its ability to metabolize mucins and related proteoglycans.

Overall, our findings indicate that *A. muciniphila* can access multiple nutrient sources in the GI, but the ability to use mucin via the MUL complex gives the bacterium a competitive advantage against other members of the intestinal microbiota to enable stable colonization. Furthermore, *A. muciniphila* mucin fermentation repressed cholesterol biosynthesis in the colons of GF mice. Therefore, in addition to the immunomodulatory activities that have been assigned to components of *A. muciniphila* and their links to health benefits, the active catabolism of mucin by *Akkermansia* may provide additional health benefits by regulating the expression of genes involved in lipid biosynthesis.

Acknowledgements

We are thankful to Ozge Kuddar and Eliud Rivas for support in the assembly of arrayed transposon mutant libraries, Josh Granek for the base trimming code, and members of the Valdivia laboratory for critical reading of the manuscript. This work was supported by NIH awards AI142376 and DK110496 to RHV, AHA award 18POST34070017 to LD and a fellowship from the Natural Sciences and Engineering Research Council of Canada PDF4878642016 to LD.

Methods

Media and strains

Bacteria were grown in an anaerobic chamber (Coy Laboratory) with the following gaseous characteristics: 5% hydrogen, 5% carbon dioxide, 90% nitrogen. *A. muciniphila* was grown in mucin medium based on previous work[19] (3 mM KH_2PO_4, 3 mM Na_2PO_4, 5.6 mM NH_4Cl, 1 mM $MgCl_2$, 1 mM $Na_2S{\cdot}9H_2O$, 47 mM $NaHCO_3$, 1 mM $CaCl_2$ and 40 mM HCl, trace elements and vitamins [60], and 0.25% porcine gastric mucin (Type III, Sigma-Aldrich)). Additional media used in this study included synthetic medium, where porcine gastric mucin was replaced with 0.2% GlcNAc, 0.2% glucose, 16g/L of soy peptone (Amresco) or Phytone (BD) and 4g of threonine/L[36], and Brain Heart Infusion (BHI, BD) supplemented with 2% tryptone and 1x hemin and vitamin K (Hardy Diagnostics). *A. muciniphila* BAA-835[19], *Bacteroides vulgatus*, and *B. thetaiotaomicron* were obtained from ATCC and additional strains, including *A. glycaniphila, Bifidobacterium longum, Bifidobacterium bifidum, Ruminococcus gnavus, Ruminococcus torques*, and *Peptostreptococcus russellii* were obtained from DSMZ. *E. coli* S17 harboring the pSAM_Bt plasmid was provided by Andrew Goodman at Yale University. When required, antibiotics were used at the following concentrations for *Akkermansia*: chloramphenicol 7 μg/ml, gentamicin 10 μg/ml, kanamycin 12 μg/ml. *E. coli* was cultured in LB medium, and antibiotics were added as required at the following concentrations: ampicillin 100 μg/ml, chloramphenicol 35 μg/ml, kanamycin 30 μg/ml.

Labeling of mucin glycans

Porcine gastric mucin was labelled with 6-aminofluorescein by adapting a previously described protocol to label polysaccharides [61]. Prior to labelling, commercial pepsin digested gastric mucin (Type III mucin, Sigma) was filtered through a 0.45 μM filter and transferred to a 5 ml tube. In a fume hood, 1 ml of filtered mucin was combined with 30 mg of CNBr diluted in 350 μl of water. The reaction was carried out for 25 min while monitoring the pH using paper strips. Aliquots of 0.25 M NaOH were added as needed to maintain the pH > 10 for the duration of the reaction. Following activation, excess CNBr was removed using a Bio-Rad 10DG desalting column equilibrated with 0.2M sodium borate buffer, pH 8. Large mucin glycoproteins were eluted in the void volume into a tube containing 2 mg 6-aminofluorescein and reacted overnight in the dark. The labelled mucin glycans were then dialyzed against 16 L of 20 mM sodium phosphate buffer, pH 8, using an 8 kDa MWCO membrane and quantified using the phenol sulfuric acid method[62].

Microscopy

To test *Akkermansia*'s interaction with labeled glycans by microscopy, cultures were grown overnight to saturation and sub-cultured 1:1 in fresh mucin medium for 5 hours. The actively growing cultures (600 μl) were combined with 1.3 ml fresh synthetic media and either 20 μg carbohydrate/ml fluorescein-mucin or 20 μg/ml fixable Fluoro-Emerald (fluorescein) dextran (10,000 MW, Thermo D1820). For experiments involving carbonyl cyanide m-chlorophenylhydrazone (CCCP), the cultures were pretreated with 50 μM CCCP for 30 min prior to the addition on FL-mucin. Cultures were then incubated at 37°C for 3h and subsequently harvested by centrifugation at 14 000 xg for 5 min, washed once with PBS, and fixed with 4% formaldehyde solution for 30 min on ice. The cells were washed with PBS to remove formaldehyde, resuspended in GTE buffer (50 mM glucose, 25 mM Tris, 10 mM EDTA), and applied to poly-lysine coated coverslips. Unbound cells were removed by washing with PBS.

The cells were then treated with lysozyme (1 μg/ml) for 2 min, washed with PBS, and blocked with 2% BSA in PBS for 10 min at room temperature. Finally, the cells were stained with anti-*Akkermansia* antisera overnight at 4°C. After staining with fluorescently labeled secondary antibodies (goat anti-rabbit IgG Alexa Fluor 647, Invitrogen A21244). The cells were imaged on an inverted confocal laser scanning microscope (LSM 880; Zeiss) equipped with a motorized stage, diode (405, 561 nm), argon ion (488 nm), and helium-neon (633 nm) lasers, and an Airyscan detector module with Airyscan Fast capability. Images were acquired using a Plan Apo 63x/1.4 NA objective (Zeiss) and the Zen software (Zeiss) with Airyscan deconvolution.

To test fluorescent mucin uptake by *B. thetaiotaomicron*, starter cultures were grown to an OD_{600} of 0.4-0.5 in chopped meat media supplemented with hemin and vitamin K, or for *A. muciniphila*, cultures were grown in synthetic medium. The cells were washed once and suspended in PBS. The cultures (600 μl) were combined with 1.3 ml of a modified version of synthetic media prepared with 0.25% mucin instead of glucose/GlcNAc and supplemented with 100 μl FL-mucin (final concentration 20 μg carbohydrate/ml). Labeled mucin accounts for approximately 0.8% of the total carbohydrate content of the media. The cultures were incubated for 3h at 37°C. Membranes were stained with 10 μg/ml fixable FM4-64fx (Invitrogen, F34653) and the cells were fixed with 4% formaldehyde and prepared for microscopy as described above.

For super-resolution microscopy, *A. muciniphila* was grown with FL-labeled mucin and prepared for imaging as described above, except that lysozyme was omitted and the cells were stained with anti-*Akkermansia* antisera followed by STED compatible fluorescently labeled secondary antibodies (goat anti-rabbit IgG Alexa Fluor 594, Invitrogen A32740). Super resolution microscopy was carried out using a Leica TCS SP8 STED (STimulated Emission Depletion) microscope equipped with a Leica DMi8 inverted motorized stage, a pulsed white light laser, high-sensitivity GaAsP HyD detectors, and STED depletion lasers (592nm, 660nm, 775nm). For each image, 40 stacks in the Z plane were acquired with a 100x/1.4 objective HCX PL APO OIL DIC WD 90 μm (Leica) and LAS X software (Leica), followed by deconvolution with Huygens Professional (SVI) and 3D rendering with Imaris (v9.5, Oxford Instruments).

Live cell imaging of mucin acquisition

Cells were imaged under 1.5% agarose pads. The pads were prepared by melting 0.3 g agarose (Apex General Purpose LE Agarose) in 20 ml synthetic media and spotting ~200 μl of molten agarose onto glass bottom dishes. The pads were solidified and equilibrated for 2 h under anaerobic conditions prior to use. To prepare for live cell imaging, *A. muciniphila* cultures were grown overnight in mucin medium (OD_{600} ~0.05) and subsequently diluted into a 1:1 mix of fresh synthetic and mucin medium. The cultures were incubated for an additional 3 h at 37°C to ensure that the cells were actively growing. For CCCP treated cells, 50 μM CCCP was added during the last 15 min of incubation. Cells from 1.5 ml of culture were then harvested by centrifugation at 10 000 x*g* for 3 min and resuspended in 50 μl synthetic media supplemented with 15 μl of FL-labeled mucin (1.4 μg). Aliquots of the cells (10 μl) were immediately plated on 35 mm glass-bottom dishes (Cellvis), overlaid with an agarose pad, and sealed with parafilm under anaerobic conditions. The cells were imaged on a Nikon Ti2 inverted microscope equipped with a motorized stage, LED light source (Sola Light Engine), an ORCA-Flash4.0 V3 sCMOS digital camera (Hamamatsu), and an environmental chamber (Okolab) set to 37°C. Images were acquired every 30 sec for 30 min using a Plan Apo 100x/1.4 NA oil objective (Nikon) with 1.5x

zoom in the NIS-Elements software. The images (resolution = 43.33 nm/pixel) were deconvolved in the NIS-Elements using the automatic 2D deconvolution function and bleach corrected with histogram matching in ImageJ (NIH), converted to 8-bit TIFFs and reconstructed in Adobe Photoshop where only linear adjustments were made to fluorescence intensity. Annotations were added using ImageJ[63].

Kinetics of mucin uptake and processing

A. muciniphila was grown in mucin medium to an OD_{600} of 0.5 and mixed 1:1 with fresh synthetic media and 96 μg/ml fluorescein labelled mucin. Cultures were incubated anaerobically at 37°C. At each timepoint, the OD_{600} was measured and 1.5 ml of culture was collected and the bacteria were harvested by centrifugation at 14 000 xg, 3 min. The resulting cell pellets were either fixed with 4% formaldehyde and used in flow cytometry as described below, or they were boiled in 50 μl SDS-PAGE sample buffer for analysis using SDS-urea agarose polyacrylamide composite gel electrophoresis (SDS-agPAGE)[64], which is a modified type of protein gel designed for running high molecular weight mucin proteoglycans. The spent medium supernatants were transferred to a new tube and precipitated on ice with 5% TCA for 30 min, pelleted by centrifugation, and washed with cold acetone. Loading volumes for the cell lysates were standardized based on the culture optical density, and samples were resolved on a standard SDS-PAGE gel and stained with Coomassie Blue to test for equal protein loading. The same samples were then run on an SDS-agPAGE gel at 80v for 120 min followed by 150v for 30 min. Gels were then imaged unstained using the 600 nm channel in a LI-COR Odyssey Fc imaging system (LI-COR Biosciences) with a 2 min exposure time to detect fluorescein-labeled mucin glycans.

Flow cytometry

Flow cytometry was used to measure the fluorescent intensity of bacteria grown in the presence of fluorescein-mucin. Cultures were grown in mucin medium overnight, subcultured 1:1 into fresh mucin media and grown for an additional 5h. The actively growing cultures (600 μl) were combined with 1.3 ml synthetic media and 20 μg carbohydrate/ml fluorescein-mucin. For mutants with mucin growth defects, synthetic media was used throughout. Cultures were grown in triplicate and fixed as described above. The cells were then incubated with rabbit anti-*Akkermansia* antisera in PBS/2% BSA (1:100 dilution) overnight at 4°C, washed with PBS, followed by incubation with anti-rabbit IgG Alexa 647 antibodies (1:1000, Invitrogen A-21244). Cells were analyzed on a BD Canto II flow cytometer and 50 000 events were collected for each sample. Additional analysis was run using FlowJo v10 software. The samples were gated on the population of single cells that stained positive with anti-*Akkermansia*/Alexa-647, and the mean intensity of fluorescein-mucin staining was recorded.

Growth curve assays

Growth kinetics of wild-type *A. muciniphila* and selected Tn mutants was measured in both mucin and synthetic media. Starter cultures were prepared by growing bacteria in synthetic medium, supplemented with chloramphenicol (7 μg/ml) for mutants and diluted 1:1 into fresh medium and grown for an additional 5 h. The resulting cultures were then diluted 1:25 into fresh media (OD_{600} 0.01 – 0.05) and 150 ml aliquots were dispensed into 96-well microplates. Each well was covered with 100 μl of paraffin oil and incubated at 37°C in a BMG SpectroStar Nano plate reader under anaerobic conditions. The optical density (OD_{600}) was measured at 1 h intervals for 72 h. Results were obtained from three biological replicates per strain. Assays with

intestinal microbes other than *Akkermansia* were carried out as describe above, except that the starter cultures were prepared in Brain Heart Infusion medium (BHI, BD) with 1x hemin vitamin K and 2% tryptone. The BHI media for *A. muciniphila* was also supplemented with 0.25% mucin. Growth curves were run using either mucin with 1x hemin vitamin k or in BHI with hemin vitamin K and tryptone. Results were obtained from two biological replicates per strain.

Generation of anti-*Akkermansia* anti-sera

Polyclonal anti-sera was raised against whole *A. muciniphila* cells. To prepare the antigen, overnight cultures of *A. muciniphila* were washed three times with PBS and incubated at 4°C aerobically for 24h. The cells were then fixed with 4% formaldehyde, washed, and suspended in PBS. Antibodies were raised in New Zealand white rabbits. The antigen was mixed 1:1 with Freund's complete adjuvant and 450 µl was used for subcutaneous inoculation, followed by three additional injections using Freund's incomplete adjuvant.

Generation of anti-Amuc_0544 anti-sera

Polyclonal anti-sera was raised against a 47 kDa fragment of Amuc_0544 protein that spans the TPR domains (from 1216 bp to 2457 bp). The fragment was amplified with primers #127 and #129 (**Table S1**) using Phusion polymerase and 5X HF Buffer (Thermo Scientific) and cleaned with a QIAquick PCR purification kit (Qiagen). The resulting PCR product and the expression vector pET28a (EMD Biosciences) were digested with EcoRI-HF and HindIII-HF (NEB #R3101, #R3104), gel extracted using a QIAquick Gel Extraction kit, and ligated overnight with T4 DNA ligase (NEB, #M0202). The ligated DNA was cloned into competent *E. coli* XL-1 Blue (Agilent), and after the construct was confirmed by sequencing, the final plasmid was transformed into *E. coli* BL21(DE3) (NEB, #C2527I) for protein expression.
The Amuc_0544 fragment was produced as a soluble protein with an N-terminal hexahistidine tag. *E. coli* BL21(DE3) was grown in LB medium with 30 µg/ml kanamycin to mid-exponential phase and induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 3 h. The cells were harvested by centrifugation (5000 xg, 10 min) and the pellet was suspended in buffer (50 mM sodium phosphate, 300 mM NaCl, 20 mM imidazole; pH 7.4) and lysed by sonication with 30 cycles, 30 s intervals alternating with 30s on ice (Branson Sonifier 450). Cell debris was removed by centrifugation (15 000 xg, 45 min, 4°C), and the supernatant was applied to a Ni-NTA agarose bead column (Prometheus Protein Biology Products). The column was washed with 25 ml of buffer containing 40 mM imidazole, and protein was eluted in 1 ml fractions with 300 mM imidazole. Antibodies were raised in rabbits as described above using 2.5 mg of purified protein per injection.

RNA-seq analysis of *A. muciniphila* grown in mucin and synthetic media

RNA-seq was used to assess bacterial transcriptional responses to growth in mucin and synthetic media. For each medium to be tested, 25 ml cultures were prepared in triplicate and grown to mid-log (OD_{600} ranging from 0.35 to 0.5). Cells were harvested by centrifugation at 10 000 x*g*, 5 min, 4°C. RNA was isolated using a Qiagen RNeasy kit following the enzymatic digestion protocol (proteinase K + lysozyme). The resulting RNA was concentrated by precipitation with 2 volumes of isopropyl alcohol and 0.3M sodium acetate, pH 5. rRNA was depleted using a Ribominus Bacteria kit (Thermo). RNA quality was assessed by running on an Agilent Bioanalyzer. Finally, 100 ng per sample was used as input RNA to prepare libraries with a NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB #E7760S) and NEBNext

Multiplex Oligos for Illumina. Libraries were sequenced on a Hiseq 4000 in a 50 bp SR run. Following sequencing, adaptor sequences were trimmed and filtered using the program fastq-mcf[65] and read quality was assessed with FastQC[66]. The reads were then mapped to the *A. muciniphila* genome (accession: CP001071.1) using STAR[67] with the parameter --alignIntronMax set to 1. Differential expression analysis was carried out with DESeq2[68] and pathway analysis was run using ClusterProfiler[69].

pSAM_Akk plasmid construction

Transposon mutagenesis was carried out with a modified version of pSAM_Bt[16], adapted for use in *Akkermansia*. First, the *ermG* gene was replaced with a *cat* gene to confer chloramphenicol resistance after observing that *A. muciniphila* could acquire spontaneous resistance to erythromycin. The *cat* gene and its promoter was amplified from pRL1342[70] with the primers #80 and #81 (**Table S1**) and introduced into pSAM_Bt using the MfeI and XbaI cut sites. Cloning was carried out in the *pir*+ strains *E. coli* Pir2 (Invitrogen) and *E. coli* S17. Next, the *himar1C9* transposase was codon optimized for expression in *Akkermansia*. To generate a codon table for *Akkermansia*, we used a selection of representative genes (Amuc_R0036, *rpoD, secA*, *eftu*, DNA pol I, DNA pol III alpha, *ftsZ, ftsY, yidC*, GAPDH, GroEL, enolase) and the tools provided in Sequence Manipulation Suite[71]. Comparison between *A. muciniphila* codon usage and *himar1C9* revealed rare codons for arginine and serine. These were substituted for codons frequently used in *Akkermansia*, and the resulting gene was synthesized as a gBlock (IDT). The codon optimized *himar1C9* was then ligated into pSAM_Bt using the NdeI and NotI cut sites to produce pSAM_Akk. Replacement of the original *himar1C9* gene was confirmed by sequencing.

Transposon mutagenesis

Basic protocol. Initial attempts at mutagenesis using pSAM_Akk were unsuccessful, and we observed that the plasmid can transiently remain in *A. muciniphila*, even in the absence of the *pir* genes required for the R6K origin of replication. To facilitate transposition, cure the donor plasmid, and remove potential residual donor *E. coli* contamination, several outgrowths in the absence of selection were required. To carry out conjugations, *Akkermansia* and *E. coli* were combined at a 10:1 ratio. *A. muciniphila* starter cultures were grown to saturation in 5 ml synthetic medium and subsequently diluted 1:10 into 20 ml synthetic medium. The cultures were grown to an OD_{600} = 1 and harvested by centrifugation at 15 000 xg, 5 min. The *E. coli* donor was prepared by growing in LB medium (Fisher Bio-reagents, #BP1426-2) with ampicillin (100 μg/ml). Overnight cultures were subcultures 1:50 and grown to an OD_{600} = 0.4 to 0.6. Cells were harvested by centrifugation at 1 500 xg and washed with LB media to remove residual antibiotics. The pellets were combined to create a slurry of cells that was distributed in 100 μl puddles on pre-reduced synthetic media plates. Conjugations were incubated aerobically at 37°C for 14 to 16h. Following conjugation, the puddles were scraped into a 1:1 mixture of PBS and 50% glycerol and resuspended. The cell suspension was diluted 1:10 into fresh synthetic media supplemented with gentamicin (10 μg/ml) and kanamycin (12 μg/ml) to counter select against the *E. coli* donor. Cultures were grown anaerobically for an additional 48 h to allow *Akkermansia* to recover, and subsequently sub-cultured 1:10 into fresh media two more times, at 24 h intervals. After the final subculture, 150 μl of the suspension was plated on synthetic media containing gentamicin, kanamycin, and chloramphenicol (10 μg/ml, 12 μg/ml, 7 μg/ml). Colonies appeared after 4 to 6 days incubation at 37°C. Colonies were routinely tested for contamination from residual *E. coli* donor by growing aerobically at 37°C.

To confirm that transposition had occurred, initial quality control tests were performed by PCR to screen for the presence of plasmid backbone by amplification of the *bla* gene (primers #11/12), *Akkermansia* specific 16S (primers #75/76), and the *cat* gene located on the transposon (primers #80/81). Once transposition was confirmed, arbitrary PCR was used for small-scale mapping of Tn insert locations[72], using the following sequential PCR settings: Reaction 1, run with primers #22/88: 94°C for 5 min; 6 cycles of 94°C × 30 s, 30°C × 30 s, 72°C for 1 min; 30 cycles of 94°C × 30 s, 45°C × 30 s, 72°C for 1 min; and 72°C for an additional 5 min. Reaction 2 was run using 2 ul of Reaction 1 as template and primers #23/89: 95 °C for 1 min followed by 30 cycles of 95°C × 30 s, 45°C × 30 s, 72°C for 2 min; 72°C for 5 min. The resulting PCR products were sequenced with primer #89. Finally, to confirm that the Tn inserted randomly into the genome, a set of representative mutants were analyzed by Southern blot. DIG labelled probes were generated to detect the *cat* gene and the blot was run using a DIG High-Prime Labeling and Detection II kit (Sigma).

Construction of Tn libraries

Tn libraries were prepared in two formats for INSeq analysis. **1. Pooled Tn library:** The first library consisted of approximately 25 000 mutant colonies which were prepared by scraping agar plates and pooling the resulting cell suspensions. This library was used for in vitro INSeq experiments. **2. Arrayed Tn library:** The second approach used a smaller, but well-defined library constructed with equal inputs of know mutants from a library arrayed in 96-well plates. This library was used for all in vivo INSeq experiments. By generating pools of defined mutants to be used for INSeq analysis, we were able to filter the data against known inputs and remove spurious reads generated by sequencing errors. A subset of the arrayed library was used for a Cartesian mapping strategy, described below, to facilitate retrieval of mutants of interest.

i. Pooled Tn mutant Library. Because of the low transposition efficiency, we ran 15 conjugation reactions and then subdivided each reaction into 16 independent outgrowths using 96 deep-well plates. The outgrowths were carried out as described above by culturing conjugation mixes for an initial 48h followed and two additional subculture steps at 24h intervals in the synthetic medium with gentamicin and kanamycin to inhibit *E. coli* growth. To select for transconjugants, the reactions were plated on synthetic media with gentamicin, kanamycin, and chloramphenicol and incubated anaerobically at 37°C for 6 days. Colonies were scraped into PBS with 20% glycerol and the number of colonies and volumes used were recorded for each reaction. This information was used to determine volumes required to mix equal numbers of colonies from each reaction. In total, the large pool was estimated to represent approximately 25 137 colonies (2680 insertions in 198 intergenic regions and 725 genes). The pooled cell suspensions were aliquoted and stored at -80°C and used as inoculum for in vitro INSeq experiments.

ii. Arrayed Tn mutant library. To generate an arrayed library of *Akkermansia* Tn mutants, mutagenesis was carried out as described above, except that individual colonies were picked and arrayed into a total of 54 x 96-well plates. Colonies were grown in synthetic medium with antibiotics until all wells were turbid. Glycerol was added to a final concentration of 20% for storage at -80C. To identify the Tn insertion locations for each mutant, we used the INseq library and mapping protocol. For each plate, 60 ul per well was pooled and DNA was extracted with Qiagen's DNeasy Bloods and Tissue kit. The resulting DNA was used to prepare sequencing

libraries, enabling the identification of each Tn insertion on a plate. Cultures from the arrayed plates were then combined to create a single pooled library comprised of all of the arrayed mutants in approximately equal amounts. The pooled cell suspensions were aliquoted and stored at -80°C and used as inoculum for in vitro and in vivo INSeq experiments.
Because the collection of 54 individual 96-well plates was derived from multiple conjugations and the extensive outgrowths required to obtain *A. muciniphila* Tn-mutants, plates were first sequenced separately. Three plates displayed extreme levels of clonality and were not used further. We chose to use the Arrayed Tn mutant library for the majority of experiments, because this ensured that any Tn-mutants identified from INseq experiments could potentially be retrieved for follow-up experiments.

Cartesian Mapping strategy to map Tn insertion sites in the *A. muciniphila* arrayed collection

Bacteria from individual 96-well plates were pooled, assigned a barcode, and the Tn insertion sites sequenced as previously described[17]. This approach provided very high sequencing depth for each 96-well plate and provided confidence in distinguishing true sequence reads derived from Tn-insertions on the 96-well plate from PCR amplification and sequencing artifacts. As the signal-to-noise varied between individual samples and sequencing runs, a Matlab script (INseq_read_filter_v3.m) was used for viewing data and setting thresholds. This process was used to build a table of Tn-insertion sites present on each plate, and collectively in the library. A high degree of clonal redundancy was observed in the Arrayed Collection (mean of 3.7 replicates of each clone), likely as a result of the extensive outgrowths that were required to effectively remove the E. coli donor strain prior to selection of transconjugants.

Clonal redundancy creates a challenge for mapping Tn-insertions to specific plate-well locations using methods that are based on orthogonal pooling and high throughput sequencing because the presence of multiple representatives of a clone in a pool either eliminates information on location (binary code[16]) or makes that information ambiguous (Cartesian pooling and Sudoku[18]). Thus, we faced a trade-off between capturing the full diversity of Tn insertions in the collection, and fewer plates to derive more useful location mapping information. Simulations were used to model this trade-off, and to optimize the choice of 96-well plates taken for location mapping, with the aim of capturing the highest number of unique clones and the lowest level of redundancy.

i. Simulations: An optimization routine was implemented in Matlab (sudoku_plate_compare_v1.m) to find the subset of 96-well plates that would yield the highest diversity and lowest redundancy when combined. Intra-plate redundancy was first simulated for each plate to generate an in silico set of 96-well plates. The number of clonal replicates on a plate was taken as 96 minus the number of unique Tn-insertions identified on that plate by deep sequencing. Replicates were then assigned to specific Tn-insertions identified on a plate by uniform random sampling with replacement. A full set of in silico 96-well plates generated by this process was then used for inter-plate comparisons designed to capture the series of plate-sets that minimizes clonal redundancy. The first in this series is simply the plate with the most unique clones (lowest intra-plate redundancy). In subsequent steps, all remaining plates are tested one at a time in combination with the set of previously chosen plates (1..i-1), to find the plate (i) that creates a new set (1..i) that has the lowest overall redundancy.

ii. Evaluating the trade-off between coverage-redundancy when choosing a subset of plates suitable for Cartesian mapping*:* Simulation data for each optimized set of plates was then used to evaluate the trade-off between coverage and redundancy. The number of unique Tn-insertions and the mean redundancy increase for optimized plates sets were determined for a size of 1 through 40 (Fig. **S2e**). The plate set that yields the largest diversity-for-redundancy is found where the distance between these two curves is greatest (set size 16: 873 unique Tn-insertions, mean redundancy = 1.76).

The degree of redundancy per clone changes with increasing plate-set size in the simulated data (**Fig S2f**). This plot was used to further inform our choice and establish a cut off for the plate-set series used for Cartesian mapping. Cartesian mapping is precise when a clone is present in the library in one copy, allowing unambiguous identification of the Plate-Row-Column location for each Tn-insertion directly from the sequencing data, without any need for retrieval and secondary evaluation. This is particularly useful for high-throughput screening (forward genetics), where the genetic nature of hits in the screen can be immediately identified from the Cartesian mapping table. Clones present in two and three copies in the library still provide useful location information for Cartesian mapping, but this information is not precise and manual retrieval and secondary confirmation by PCR are necessary to match Tn-insertion information with a Plate-Row-Col identifier. The search space for redundant clones scales as the <number of copies> to the power of the <number of pooling dimensions>, so a clone present three times in the collection would be mapped to 27 potential Plate-Row-Col locations (3^3). In most cases, only one of the three copies needs to be retrieved for further work, effectively eliminating one pooling dimension and reducing the search space to nine Plate-Row-Col locations. Thus, hypothesis-directed retrieval of clones from the library (reverse genetics) also benefits from reducing the degree of clonal redundancy. Striking a balance between genome coverage and useful location mapping, we chose the set of 19 optimized plates (Akk SudSet-19) for location mapping. This set of plates captures 950 of the 1277 Tn-insertions in the entire collection, 86% of disrupted genes (insertion occurred with the 90th percentile of the CDS pos=0.9), and reduces the mean redundancy from 3.7 to 1.9 copies per clone

IN-seq analysis of conditionally essential genes in mucin medium

The mixed pooled mutant library was inoculated at 1:20 dilution in Synthetic Medium containing 10 ug/ml gentamycin, 12 ug/ml kanamycin, 5 ug/ml chloramphenicol and grown to saturation. Cells were washed twice with media and used to inoculate 5 ml test media at a 1:100 dilution. Growth in mucin was tested in mucin medium with a final concentration of 0.5% porcine gastric mucin. Growth in mucin with added nitrogen was tested in mucin medium supplemented with 0.5% mucin, 3 g/L threonine and 16 g/L soy-peptone. As growth rates varied significantly in the different test media, cultures were monitored regularly for OD_{600} by withdrawing a small volume of culture. To obtain samples that had gone through a similar number of generations after inoculation with the Input pool, cultures were collected as they reached saturation, rather than after a fixed length of time.

INSeq analysis of conditionally essential genes in mice

INSeq analysis was carried out using four C57BL6/J mouse models of colonization: germ-free (GF, n = 12), Altered Schaedler Flora colonized (ASF, n = 7), conventional (CONV, n = 12), and $Muc2^{-/-}$ mice[39] (n = 4). Colonization of CONV and Muc2-/- mice required pre-treatment with tetracycline to eliminate endogenous *Akkermansia* and allow the mutant library to engraft (3 g/L

tetracycline suspended in distilled water with 1% sucrose for two weeks). DNA was then extracted from fecal pellets using a Qiagen Fast Stool Mini kit and tested for residual *Akkermansia* by PCR with *Akkermansi*a specific 16S primers (#75/76). Once the *Akkermansia* was cleared, the antibiotics were withdrawn 48 h prior to the delivery of Tn mutant libraries by oral gavage. GF and ASF mice did not receive antibiotics.

Tn mutant libraries (Arrayed Collection) were prepared for gavage by diluting frozen library stocks 1:10 into synthetic media and incubating at 37°C for 36 h. Cells were harvested by centrifugation at 10 000 xg, 5 min, 4C, washed once with reduced PBS before being resuspended in $1/10^{th}$ of the initial volume. Mice were inoculated by intragastric gavage with 150 µl of the Tn mutant pool (~10^9 bacteria), and the mutant library was allowed to colonize for 14 days. Mice were euthanized and cecal contents were collected and stored at -80°C until they were processed for DNA extraction. While most samples could be used directly for DNA extraction, the ASF mice had lower yields of *Akkermansia* DNA, and an enrichment step was added to increase recovery of *A. muciniphila*. Cecal contents from ASF mice (~200 mg) were combined with 10 ml of synthetic medium containing chloramphenicol, kanamycin, and gentamicin and grown for 8 generations (19 h). DNA was extracted from 9 ml culture per mouse and subsequent steps of INseq library preparation were carried out using the same approach for all samples.

Droplet INSeq

Starter cultures were prepared by diluting frozen Tn library stocks (Arrayed library) 1:10 into synthetic media and grown for 36 h. To determine the dilution factor required for droplet single cell loading, serial dilutions of cultures were made on mucin-supplemented agar plates for a CFU assay. Given the OD_{600} of the starter culture and the plate CFU counts, the CFU/mL concentration could be estimated from an O_{600} measurement prior to droplet loading[73]. Droplets were generated using syringe pumps and a T- junction microfluidic chip with six droplet generators (Dolomite Microfluidics) with media-containing cell suspension as the aqueous phase and 2% Pico-Surf surfactant (Sphere Fluidics) in NOVEC 7500 (3M) as the oil phase. Mucin media was passed through a 40 µM filter to prevent clogging of the microfluidic chip. All droplet generation and culturing were performed in an anaerobic chamber outfitted with a portable microscope with LCD screen for viewing droplet generation on the microfluidic chip (Celestron). Following droplet encapsulation, cultures were grown at 37°C in anaerobic conditions for 72h. qPCR with *Akkermansia*-specific primers[74] (#75/76, **Table S1**) was used to assess cell abundance between droplet cultures. To retrieve cells, the droplet emulsions were broken with an equal volume of PFO, vortexed and spun down briefly at 300 g. The aqueous phase was placed in a new tube, cells were pelleted at 10000 g for 1 minute, the supernatant was removed, and the cell pellet was stored at -20 C for DNA extraction. Nine replicate cultures for each condition were pooled to generate sufficient DNA for INSeq library preparation.

INSeq library preparation

Sequencing libraries were prepared as described in Goodman et al.[17] with minor modifications. Libraries prepared from in vitro samples used 2 µg of input DNA, while libraries prepared from mice used 0.5 µg of *A. muciniphila* input DNA as determined by qPCR. Linear PCR was performed with Pfx polymerase (Invitrogen) and the primer Tru-BioSamA (**Table S1**). Two 50 µl reactions were run per sample using the following cycling conditions: 94°C for 2 min followed by 50 cycles of 94°C for 15 sec and 68°C 1 min. The resulting PCR products were

cleaned using a Qiagen QiaQuick PCR cleanup kit and eluted in 50 µl elution buffer. The resulting biotinylated PCR products were then bound to Dynabeads M-280 Streptavidin magnetic resin (Thermo). Washing, second strand synthesis, MmeI digestion, and adaptor ligation were performed as described in Goodman et al.[17], except that six-base pair adaptor sequences[75] were used. Finally, the samples were amplified with Pfx polymerase using the buffer at 2x concentration, additional $Mg2^{+}$ (4 mM final), and the primer pair TruP7-PCR_5 and TruP5-PCR_3. PCR was performed with the following conditions: 94°C for 2 min followed by 18 cycles of: 94°C 15 sec, 60°C 1 min, 68°C 2 min, followed by 68°C 4 min. The resulting PCR products were size selected on 2% E-Gel SizeSelect II agarose gels (Thermo). The DNA was then quantified using a Qubit dsDNA HS assay kit (Thermo) and sequenced on an Illumina Hiseq 4000 50 bp SR flow cell, samples from ASF and $Muc2^{-/-}$ mice were sequenced on a NovaSeq 6000 using an S1 50 bp lane. Primer sequences were modified to produce libraries that were compatible with sequencing on Illumina Hiseq 4000/Novaseq 6000 sequencers. All sequencing was performed at the Duke Sequencing and Genomic Technologies Shared Resource.

INSeq mapping and quantification of transposon insertion junctions and normalization
Pre-processing: Raw sequencing data was processed using a python script (repair_barcodes.py) to format raw sequence data files into sample-specific fasta files. These served as the input for Perl scripts (INseq_pipeline_v3.pl, minor modifications from Ref[17]) that identify sequence reads containing the Mariner-Tn, and extract the bounding genomic sequence, which are grouped and counted. Bowtie v1.2.2 was used for mapping 16 bp reads to the *Akkermansia* MucT genome. Reads that failed to map to a unique genomic locus were discarded. Output read count tables used for subsequent analysis steps contain the Tn insertion coordinate and the number of instances a sequence read was mapped to this site.

The NCBI *Akkermansia* Muc^{T} genome sequence contains an erroneous G nucleotide at position 1704819 that results in a frame shift in the highly conserved gene *recG* (Amuc_1422). Corrected sequence and annotation files that restored the Amuc_1422 reading frame were used for mapping and annotating INseq sequence reads.

Quality control: A MATLAB script (INseq_QCsummary_v3.m) was used to extract and plot quality control metrics for all samples from a sequencing run. We encountered problems with broadly distributed noise as well as small numbers of spurious sequences that dominated read depth. Quality control metrics were used to establish criteria for sample inclusion in the data set. [Depth] QC metric 1: Samples were excluded when more than 50% of the total sequencing data represented sequence reads not found in the deeply sequenced Input Sample.

[Noise] QC metric 2: The INseq library preparation should generate genomic DNA fragments from both sides (L and R) of the Mariner-Tn insertion site. High quality data showed very little variation in the relative abundance of reads mapping to the L and R sides of individual Tn insertion sites. However, some samples had a high degree of L-to-R read variability. This was quantified by calculating the variance of log10(ratLR), as follows: 1) a psuedocount of 1 was added when either L or R read count was zero; 2) for each insertion site, 'ratLR' was calculated using the larger of L or R, so that all ratios of side-bias are greater than one; 3) calculate the sample-wide variance in log10(ratLR). Samples were excluded when this metric exceeded 0.2.

Filtering: For the Arrayed Library, Tn-insertion sites identified by sequencing individual 96-well plates provided a high-confidence set of sites for filtering data sets acquired using this

collection. An analogous set of high-confidence Tn-sites was identified in the Pooled Library, by performing multiple deep sequencing runs of the Input library and combining the data for additional depth. Dynamic filtering scripts (INseq_read_filter_v3.m, INseq_read_filter_v4.m) allow users to plot the data and view quality control metrics, apply thresholds, re-plot after thresholding, and ultimately output a high-confidence set of Tn-insertion sites that could be used for filtering sequencing data from lower quality samples.

Cartesian mapping of Tn-insertions to Plate-Row-Column locations provided information about which clones had more than one Tn-insertion. Mutants with a single Tn insertion accounted for 77.5% of the arrayed library, while 17.3% were predicted to have two insertions, and the remaining 5.3% had ≥ 3 insertions. Tn-insertion sites present in multi-Tn clones in the arrayed pool were excluded from analysis, as the phenotype attributed to disruption of a gene by one Tn, could be driven by Tn disruption of another gene in the same clone.

Normalization: We applied the normalization method used by the ARTIST package to account for changes in library complexity that occur in different test conditions[76]. We made minor changes to the ARTIST method to account for our low-coverage library (ARTIST assumes high coverage and uses all TA sites in the genome for complexity calculations), and to improve computational efficiency (Artist_setup.m, Artist_run_v2.m). ARTIST is designed for single Sample-Input pairs, so an additional multi-sample normalization step (Artist_multiSample_normalization_exact.m) was required when using TRANSIT for analysis of multiple Samples (e.g. biological replicates) to account for variation in sequencing depth between Samples. Normalization features in TRANSIT were disabled.

Analysis of transposon mutant populations

The data analysis package TRANSIT[77] was used to identify genes with altered relative abundance in input and output populations. Normalized counts were analyzed using the TRANSIT Mann-Whitney U-test in command line mode with the flag -n nonorm and insertions were mapped to a protein table generated from the *A. muciniphila* ATCC BAA-835 complete genome (GenBank: CP001071.1, corrected as noted above). The output from TRANSIT was used for all downstream analyses.

Pathway analysis was carried out with the ClusterProfiler[69] program enrichKEGG to analyze genes with Tn inserts that generated an absolute Log_2 fold change > 5 in input vs output samples and an uncorrected P value < 0.05. Enriched pathways with a q-value < 0.05 were considered significantly enriched. BioCyc[78] was used to analyze global pathways in the INSeq data. For this analysis, data was analyzed using the Log_2 fold change values for all genes and mapped to the *A. muciniphila* genome. Analysis of the assimilatory reduction pathway was carried out using a combination of BioCyc and KEGG[79]. The KEGG pathway for assimilatory sulfate reduction was not included in the BioCyc database, thus the genes from the KEGG pathway amu_M00176 were retrieved in BioCyc and the Log_2 fold change for each gene was displayed using the Pathway Collage option. The Circos plot was created with Circa (omgenomics.com/circa) using INSeq data generated from mouse experiments and from INSeq data generated from growth in mucin medium using the same Tn arrayed library. Genes with a Log_2 fold change > 5 in input vs output samples and an uncorrected P value < 0.05 were mapped onto the Circos plot. To assess genes that were required for growth in vivo and in vitro, a Venn diagram was generated using the

R package ggvenn[80]. For this analysis, a less stringent cut-off of a Log_2 fold change > 1 was used to capture all genes with modest growth defects.

To identify putative glycoside hydrolases (GH) in the INSeq datasets, the list of predicted GH enzymes for *A. muciniphila* BAA-835/Muc^T was retrieved from the CAZY database[22]. Genes in additional pathways, including amino acid biosynthesis genes, were obtained from KEGG. INSeq data was plotted using the R package ggPlot2[81] and specific genes of interest were highlighted using the package gghighlight[82].

Colonization of mice with wild-type *A. muciniphila* and Tn mutants

All mouse experiments were approved by Duke's Institutional Animal Care and Use Committee. In vivo colonization experiments were carried out in GF C57BL/6J mice obtained from Duke's Microbiome Core Facility, C57BL/6J mice obtained from Jackson Laboratories, and Muc2$^{-/-}$ mice[39] (kind gift from Leonard Augenlicht, Albert Einstein College of Medicine) were bred in Duke's Division of Laboratory Animal Resources Breeding Core Facility. Mice were between 8 and 12 weeks old, and were a mix of males and females, except for the CONV mice which were all females. CONV and Muc2$^{-/-}$ mice were pre-treated with 3 g/L tetracycline suspended in distilled water with 1% sucrose for two weeks. Following antibiotic treatment, clearance of residual endogenous mouse *Akkermansia* was confirmed by PCR using *Akkermansia*-specific 16S rRNA primers[74]. Antibiotics were replaced with regular water for two days prior to gavage.

To prepare the inoculum for colonization, *A. muciniphila* cultures were standardized by optical density and suspended in PBS containing 20% glycerol. The inoculum was stored at -80°C and thawed immediately before oral gavage. All mice were inoculated by intragastric gavage with 150 – 200 µl containing ~10^9 CFU. For the competition assay, wild-type *A. muciniphila* was mixed with the *mul1A*::Tn mutant at a 1:1 ratio and mice were gavaged with 150 µl of a mixture containing 0.5 x 10^8 CFU of each strain. GF mice received a single gavage, while all other groups were gavaged once after an overnight fast and a second time 24 h later. To measure *A. muciniphila* abundance, fecal pellets were collected at regular intervals and DNA was extracted using a Qiagen Fast Stool DNA kit.

qPCR to assess *A. muciniphila* colonization levels in mice

A standard curve was generated using *A. muciniphila* genomic DNA isolated from a pure culture with a DNeasy Blood and Tissue kit (Qiagen), quantified with Qubit, and used to make a five-point standard curves with 10-fold dilutions of a 10 ng/µl starting concentration. *A. muciniphila* specific 16S rRNA primers were used to measure total *Akkermansia* loads in all experiments, and in the competition experiment levels of the mul1A::Tn mutant were determined by amplification of the *cat* gene. PCR was run with PowerUp SYBR Master Mix (Thermo) reagent on a QuantStudio 3 real time PCR system (Applied Biosystems) using fast cycling mode. The abundance of *A. muciniphila* was calculated as copy per gram fecal material or cecal content.

RNA-seq of mouse tissues

To investigate host responses to *A. muciniphila* colonization and mucin grazing, RNA-seq was performed on gastrointestinal tissues from GF mice colonized for 14 days with either wild-type *A. muciniphila* or the *mul1A*::Tn mutant. The intestinal contents were collected to quantify *A. muciniphila* levels along the GI tract and tissue was collected to prepare RNA. The entire intestinal tract was removed, and the contents of the colon, cecum, ileum, jejunum and

duodenum were collected and stored at -80°C. Colonic tissue was rinsed in cold PBS and immediately transferred to RNALater (Invitrogen). To prepare RNA, the tissue samples (~25 mg per mouse) were placed in 0.6 ml TRIzol in soft tissue homogenizing tubes (Precelly), and the samples were lysed using a bead beater set to 5000 rmp, 15 s, x 3 cycles. RNA was then isolated using Zymo's Direct-zol RNA miniprep kit (Zymo, R2050) according to the manufacturer's instructions. RNA was quantified with Qubit RNA Br kit and the RIN score was tested on an Agilent Bioanlyzer. Poly-A purification of mRNA was carried out using a NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB) and used to prepare libraries with the NebNext Ultra II directional RNAseq kit. Libraries were sequenced on a NovaSeq 6000 S-Prime 50bp PE flow cell. Sequencing data was trimmed and filtered as described for the bacterial samples, except that the reads were mapped to *Mus musculus* genome (assembly GRCm38) using STAR with the flag --sjdbOverhang 49. Differential expression was analyzed with DESeq2[68] and pathway analysis was run using MetaScape[83]. Differentially expressed genes were plotted with the R package Enhanced Volcano[84].

Co-immunoprecipitation and mass spectrometry

Native immunoprecipitation coupled with LC-MS/MS was used to identify proteins that interact with Mul1A (Amuc_0544). Bacterial cultures were prepared with either wild type or the *mul1A*::Tn mutant as a negative control. Samples were prepared with the wild type strain grown in both mucin and synthetic media, and the mutant was grown in synthetic media only, and each sample was prepared in triplicate. Cultures were grown in 30 ml media to an OD_{600} of 0.6 to 0.8 and the cells were harvested by centrifugation at 4000 xg, 20 min. The supernatant was discarded, and the cell pellets were stored at -80°C overnight, and subsequently thawed in 10 ml of a lysis buffer (150 mM NaCl, 50 mM Tris, pH 8, 1% dodecyl-beta-maltoside, 1 mM PMSF). The cells were lysed by sonication at 30 s intervals on ice and remaining cell debris was pelleted by centrifugation at 15 000 xg, 25 min, 4°C. The resulting supernatants were used for immunoprecipitation of Mul1A. Affinity purified anti-Mul1A was bound to protein A magnetic beads following the manufacturers protocol (Bio-Rad), and 50 µl of the resin was then added to cell lysates and incubated for 48h at 4°C with rotation. The beads were collected and washed four times with 50 mM NaCl, 25 mM Tris pH 8 buffer. Proteins were eluted with 25 µl 62.5mM Tris 2% SDS at 60°C for 10 min and submitted to the Duke Proteomics Core Facility for quantitative LC/MS/MS analysis.

Samples were reduced with 10 mM dithiothreitol for 30 min at 80°C and alkylated with 20 mM iodoacetamide for 30 min at room temperature. Next, they were supplemented with a final concentration of 1.2% phosphoric acid and 273 µL of S-Trap (Protifi) binding buffer (90% MeOH/100mM TEAB). Proteins were trapped on the S-Trap, digested using 20 ng/µl sequencing grade trypsin (Promega) for 1 hr at 47C, and eluted using 50 mM TEAB, followed by 0.2% FA, and lastly using 50% ACN/0.2% FA. All samples were then lyophilized to dryness and resuspended in 12 µL 1%TFA/2% acetonitrile containing 12.5 fmol/µL yeast alcohol dehydrogenase (ADH_YEAST). From each sample, 3 µL was removed to create a QC Pool sample which was run periodically throughout the acquisition period.

Quantitative LC/MS/MS was performed on 3 µL of each sample, using a nano Acquity UPLC system (Waters Corp) coupled to a Thermo Orbitrap Fusion Lumos high resolution accurate mass tandem mass spectrometer (Thermo) via a nanoelectrospray ionization source. Briefly, the

sample was first trapped on a Symmetry C18 20 mm × 180 µm trapping column (5 µl/min at 99.9/0.1 v/v water/acetonitrile), after which the analytical separation was performed using a 1.8 µm Acquity HSS T3 C18 75 µm × 250 mm column (Waters Corp.) with a 90-min linear gradient of 5 to 30% acetonitrile with 0.1% formic acid at a flow rate of 400 nanoliters/minute (nL/min) with a column temperature of 55C. Data collection on the Fusion Lumos mass spectrometer was performed in a data-dependent acquisition (DDA) mode of acquisition with a r=120,000 (@ m/z 200) full MS scan from m/z 375 – 1500 with a target AGC value of 2e5 ions. MS/MS scans were acquired at Rapid scan rate (Ion Trap) with an AGC target of 5e3 ions and a max injection time of 200 ms. The total cycle time for MS and MS/MS scans was 2 sec. A 20s dynamic exclusion was employed to increase depth of coverage. The total analysis cycle time for each sample injection was approximately 2 hours. Following 13 total UPLC-MS/MS analyses (excluding conditioning runs but including 4 replicate QC injections; **Table S1**), data was imported into Proteome Discoverer 2.2 (Thermo Scientific Inc.), and analyses were aligned based on the accurate mass and retention time of detected ions ("features") using Minora Feature Detector algorithm in Proteome Discoverer. Relative peptide abundance was calculated based on area-under-the-curve (AUC) of the selected ion chromatograms of the aligned features across all runs. The MS/MS data was searched against the TrEMBL *A. muciniphila* database (downloaded in Apr 2017) and an equal number of reversed-sequence "decoys" for false discovery rate. The database was customized to include the *Sus scrofa* MUC5AC protein sequence (XP_020938242.1), the main component of pig gastric mucin. Mascot Distiller and Mascot Server (v 2.5, Matrix Sciences) were used to produce fragment ion spectra and to perform the database searches. Database search parameters included fixed modification on Cys (carbamidomethyl) and variable modifications on Meth (oxidation) and Asn and Gln (deamidation). Peptide Validator and Protein FDR Validator nodes in Proteome Discoverer were used to annotate the data at a maximum 1% protein false discovery rate.

The analysis identified 448 Mul1A interacting proteins. To further reduce this list to the most significant candidate proteins, we implemented several filtering steps. First, this list of proteins was filtered to only include proteins that were significantly enriched in the wild-type strain vs the *mul1A* mutant negative control ($P < 0.01$) and to have > 1 peptide hit. The resulting 81 proteins were then screened for the presence of type I and type II signal peptides using the programs SignalP v3 and v4 and LipoP[85]. The resulting list of 32 proteins that were predicted to be secreted, and thus have a greater likelihood of co-localizing with Mul1A, were selected as probable Mul1A interactors. To visualize the interacting proteins, genes were plotted in Cytoscape[86]. The size of each node was set to represent the Log_2 fold change normalized peptide abundance in wild type vs the *mul1A* mutant negative control and edge thickness represents the -Log_{10}(P value), the color of the nodes was used to represent their respective Pfam[33] associations.

Analysis of short chain fatty acids in the mouse caecum

Cecal contents were used to prepare 20% (w/v) slurries in 0.5 ml PBS. The slurries were acidified to a pH of <3 with 25 µl of 6N HCL and thoroughly mixed. Debris was removed by centrifugation at 14 000 x *g* for 5 min at 4°C, and the resulting supernatant was filtered through a 0.22 µm spin column. The filtrate was then transferred to an autosampler vial and analyzed on an Agilent 7890 gas chromatograph (GC) equipped with a flame-ionization detector (FID) and an Agilent HP-FFAP free fatty-acid column. The concentrations of acetate and propionate in the samples were determined using an 8-point standard curve (0.1 mM to 16 mM).

Statistical analysis
Unless otherwise noted, statistical analysis and plots were generated with GraphPad Prism version 9.0.0.

Code availability
Code for analyzing INseq data is available at https://github.com/pmalkus/Akk_INseq_paper

Figure Legends

Fig. 1

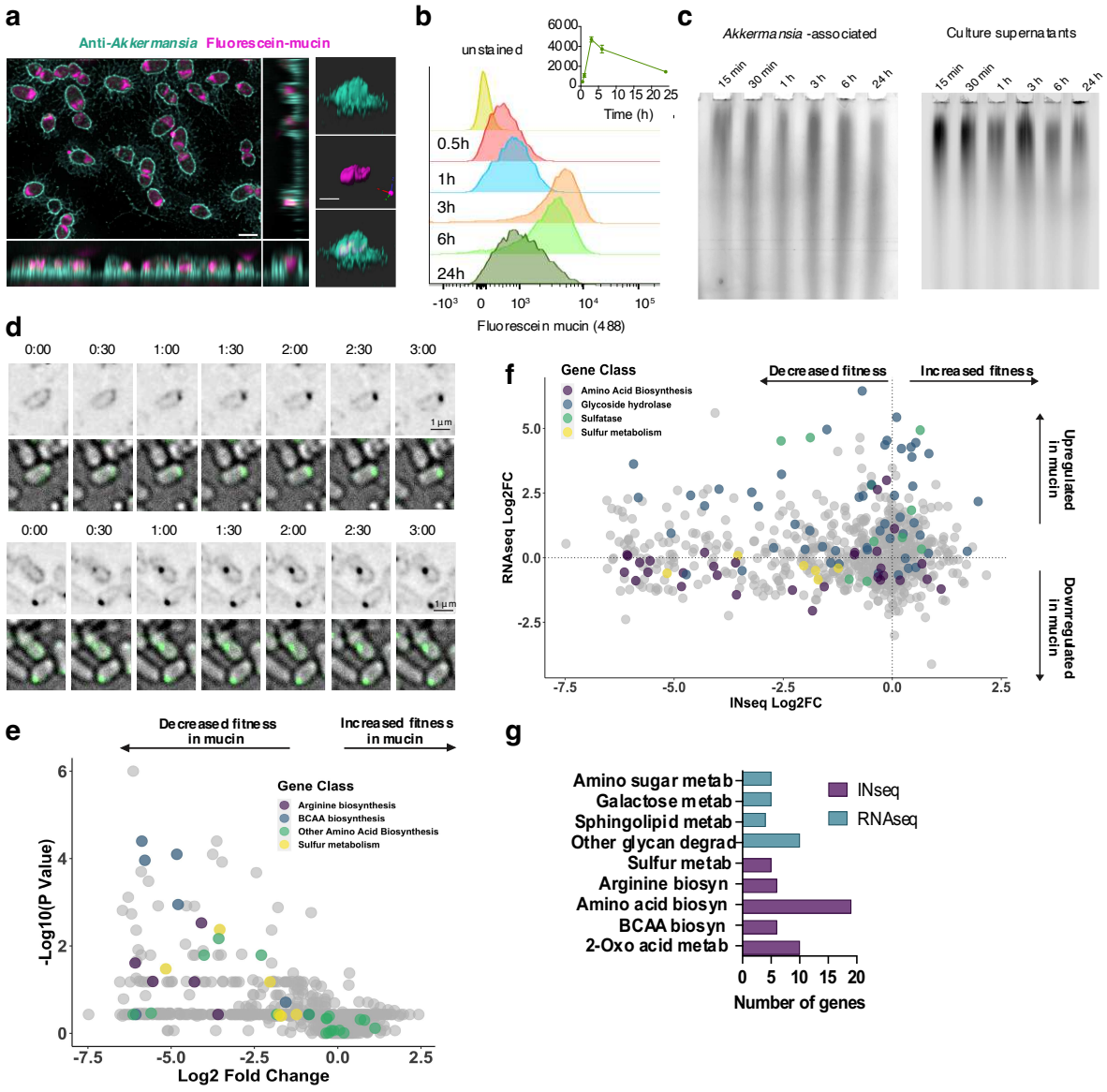

Figure 1. ***A. muciniphila*** **accumulates mucin glycans and requires amino acid biosynthesis for replication in mucin as a sole carbon and nitrogen source.** (**a-d**) Mucin glycans accumulate within *A. muciniphila* intracellular compartments. (**a**) STED reconstructions of *A. muciniphila* showing mucin glycans or mucin degradation products within intracellular compartments ("mucinosomes"). Bacterial cultures grown in the presence of fluorescein (FL) labeled mucin (purple) were immunostained with a polyclonal anti-*Akkermansia* antisera (anti-Akk, cyan). The lower panel shows orthogonal views of *A. muciniphila* and the right panel shows a 3D reconstruction of a single cell. (**b**) Flow cytometric analysis of *A. muciniphila* cultures incubated with FL-mucin over a 24 h time period. The inset shows the mean fluorescence intensity per cell at each time point. (**c**) *A. muciniphila* associated high molecular

weight mucin glycans display evidence of degradation over time. FL-labeled high molecular weight mucin glycans associated with bacterial pellets (left panel) and the corresponding culture supernatants (right panel) were resolved on an SDS-agPAGE composite gel. (**d**) Live cell imaging over a 3 min period shows accumulation of FL-mucin at the *A. muciniphila* cell pole (top) and at the midzone (bottom) (**Movie S2**). (**e-g**) Amino acid biosynthesis is required for optimal growth of *A. muciniphila* on mucin as a sole source of carbon and nitrogen. A complex Tn mutant pool was grown in synthetic medium (Input) or mucin medium for eight generations (Output). Each dot represents the pooled Tn insertions for each gene as determined by INSeq €. A comparison of the transcriptional response to growth on mucin (RNAseq) to the fitness of the corresponding mutants (IN-Seq) shows low correlation between genes expressed in response to mucin and those required for growth in mucin (**f**). Each dot represents one gene, and only genes that were detected in both the INSeq and RNA-seq datasets are shown. RNA-seq represents the Log_2 fold change in expression in mucin vs synthetic media, while INSeq represents the Log_2 fold change in the abundance between input Tn mutant pools grown in synthetic media versus the output. KEGG overrepresentation analysis using ClusterProfiler[69] to identify differentially abundant pathways (Log_2 fold change > 1.5, P < 0.05) detected by INSeq and RNA-seq following growth in mucin (**g**). (Scale bars indicate 1μm unless otherwise noted).

Fig. 2

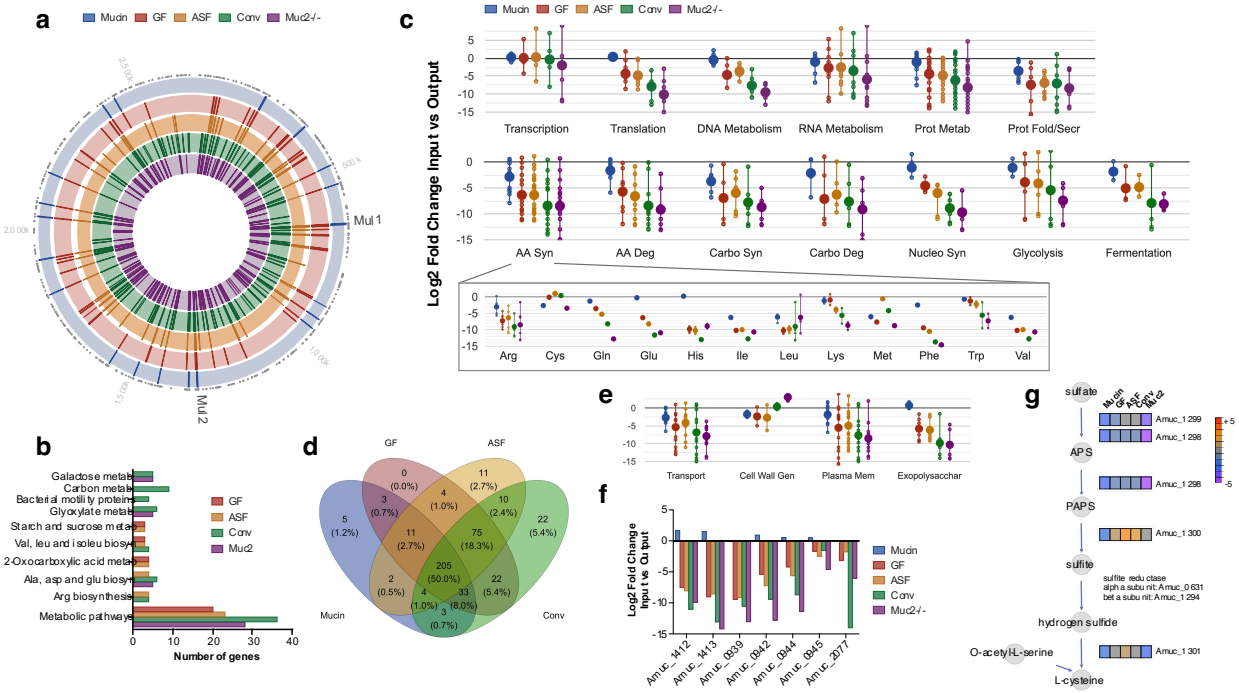

Figure 2. The metabolic requirements for *A. muciniphila* to colonize the GI tract increase as the host microbiota becomes more complex. (**a**) Chromosomal location of genes with Tn insertions from an arrayed mutant pool that produced a significant change in abundance of the corresponding mutants after growth in mucin medium (Mucin), or in the cecum of the following mouse models: germ-free (GF), stably colonized with an Altered Schaedler Flora (ASF), raised

under conventional conditions (CONV), and Muc2-deficient ($Muc2^{-/-}$) (absolute Log_2 fold change > 5, $P < 0.05$). Dots along the outer ring indicate locations of Tn insertions. (**b**) KEGG pathways overrepresented ($q < 0.05$) among mutants that displayed significant fitness defects. (**c**) Global analysis of the abundance of shared cellular systems among mutants with fitness defects in mucin medium as determined by INSeq and analyzed with Omics Dashboard[87]. The larger nodes represent the mean Log_2 fold change for components of a metabolic pathway, and the smaller nodes represent individual genes within each pathway. (**d**) Venn diagram showing overlapping genes required for growth under various conditions. For each condition, genes with a Log_2 fold change in abundance >1 were compared. (**e**) Omics Dashboard analysis of genes encoding putative cell surface components. Genes with annotations corresponding to predicted functions in capsule and exopolysaccharide biosynthesis were manually curated (Exopolysacchar). Additional functions related to plasma membrane biogenesis (Plasma Mem), cell wall biogenesis (Cell Wall Gen), and transport of amino acids, carbohydrate, and ions across the cell wall (Transport) were predicted in BioCyc. (**f**) Detailed view of exopolysaccharide biosynthesis or capsule genes and the relative fitness of respective Tn mutants in each environment. (**g**) INSeq analysis of genes in the assimilatory sulfate reduction (ASR) pathway. Heatmaps represent the Log_2 fold change of input versus output for mutants in the pathway.

Fig. 3

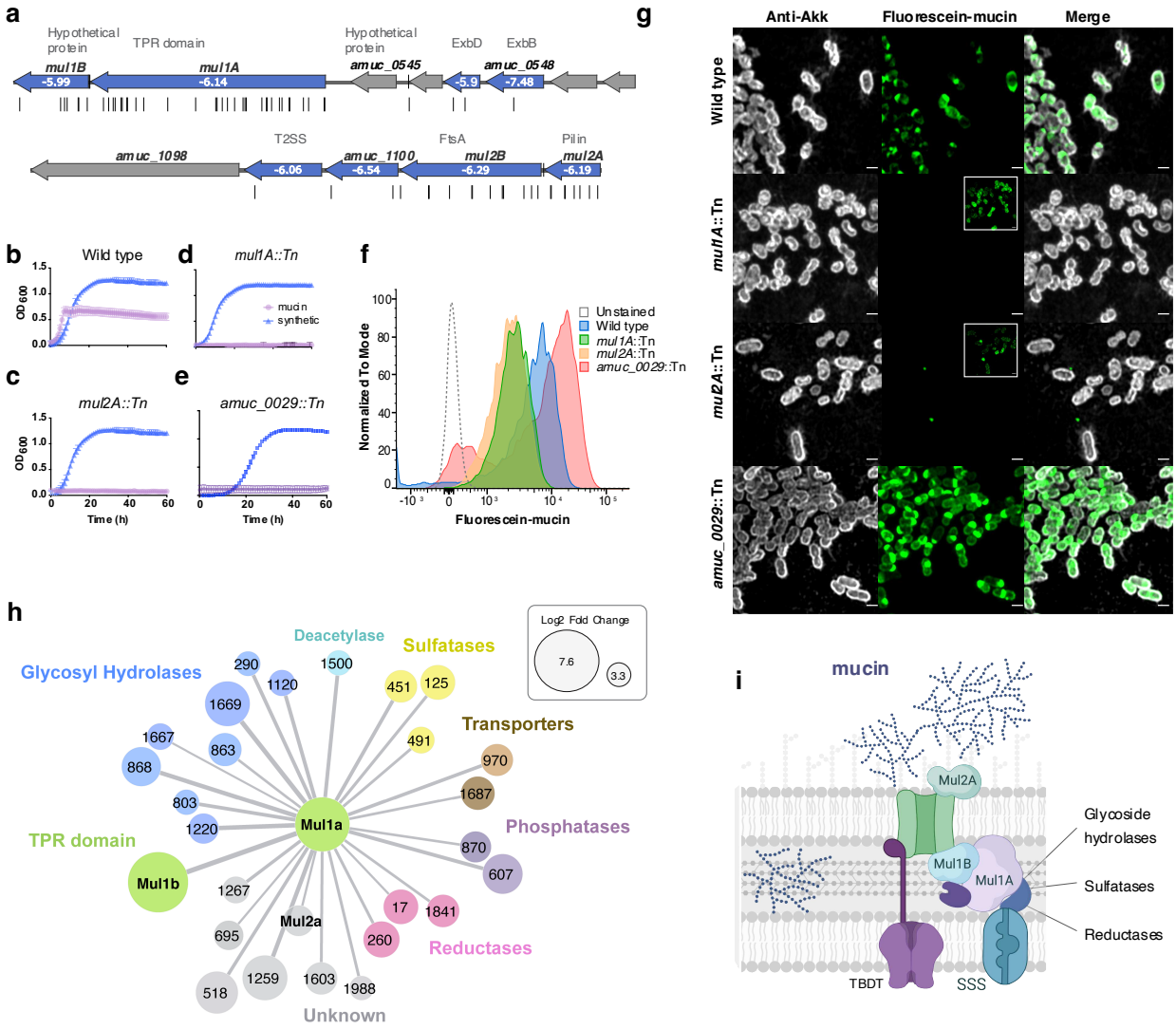

Figure 3. MUL loci encode for a mucin transport complex in *A. muciniphila*. (**a**) Map of the Mucin Utilization Loci, *MUL1* (Amuc_0543 – Amuc_0550) and *MUL2* (Amuc_1098 – Amuc_1102). Blue arrows represent genes with Tn inserts and their Log_2 fold change in abundance after culturing in mucin. Genes without Tn inserts are represented as grey arrows. Vertical lines denote the locations of Tn inserts in the complex pool. (**b-g**) Characterization of *A. muciniphila mul* mutants defective for mucin utilization display different patterns of association with FL-mucin. Growth curves of representative mutants grown in synthetic or mucin medium (**b-e**), where mucin is the sole carbon and nitrogen source, and (**f**) flow cytometric analysis of fluorescein-mucin acquisition by *mul* mutants. Mean fluorescein intensity was quantified for bacteria detected with anti-*Akkermansia* antibodies. (**g**) Mutants lacking *mul1A* and *mul2A*, but not a galactose epimerase (Amuc_0029), failed to accumulate mucin or mucin degradation products in intracellular compartments. Cultures were grown in synthetic medium supplemented

with FL-mucin prior to staining with anti-*Akkermansia* antibodies (anti-Akk). The insets show the corresponding image with enhanced brightness to visualize FL-mucin in *mul1A* and *mul2A mutants*. The scale bar is 1 **µ**m. (**h**) Multiple proteins associate with the TPR domain protein Mul1A. Mass spectrometric (LC-MS/MS) analysis of proteins that co-immunoprecipitated with Mul1A. Node size reflects the Log_2 fold change in normalized spectral counts over immunoprecipitations performed with *mul1A* mutants and edge thickness is scaled to the -Log_{10}(P value). Nodes are color coded based on Pfam associations. (**i**) Proposed model of the MUL transporter that imports mucin glycans or mucin degradation products. Mul1A and Mul1B form a complex with accessory proteins that include sulfatases, GHs, and potential inner and outer membrane transporters (sodium solute symporter (SSS) and TonB dependent transporter (TBDT)).

Fig. 4

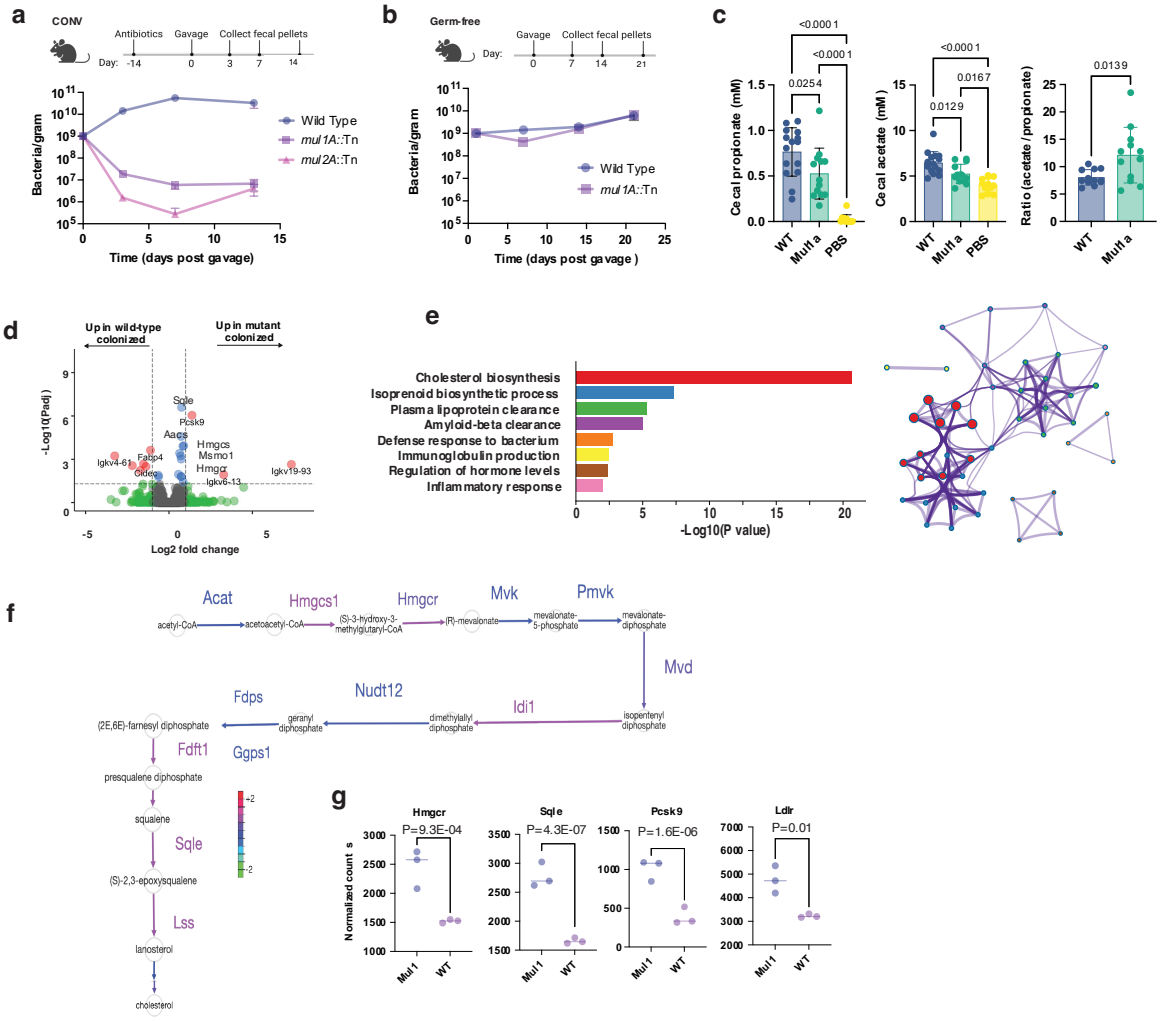

Figure. 4. Mucin consumption enables *A. muciniphila* to compete against members of the GI microbiota and leads to the repression of genes in cholesterol biosynthesis. (**a-b**) Differential colonization of mice with a conventional microbiota (CONV) (**a**) or GF (**b**) with wild type *A. muciniphila* or *mul* mutants, CONV mice were pre-treated with antibiotics to allow

engraftment of the added *A. muciniphila*. Both CONV and GF mice were inoculated by oral gavage and bacterial loads in fecal pellets assessed by qPCR. Each point represents the average *A. muciniphila* per gram of feces (n = 9-10 for CONV and n = 4-6 for GF). (**c**) SCFA content of cecal contents from GF mice (n=12-15) colonized with wild-type *A. muciniphila* (WT), *mul1A* Tn mutants, or vehicle controls (PBS). (**d-g**) Comparison of the transcriptional profiles of colonic tissue from GF mice colonized with *mul1A* mutants or wild-type *A. muciniphila*. Each point represents a gene (grey), and colors indicate a Log_2 fold change > 1 (green), an adjusted P value < 0.05 (blue), or both (red) (**d**). MetaScape enrichment plot and network visualization showing pathways enriched in mice colonized with the *mul1A* mutants. Node colors correspond to cluster annotation and edge thickness denotes relatedness of the pathways (**e**). Differential expression of genes along the mevalonate and cholesterol biosynthesis super pathways visualized with BioCyc. Enzyme names are color coded to indicate relative expression levels between mice colonized with *mul1A* mutants versus wild-type *A. muciniphila* (**f**). (**g**) Normalized expression of selected mouse genes in response to *mul1A* mutants. Benjamini-Hochberg adjusted P values are based on DEseq2 analysis.

Supplementary Figures

Fig. S1

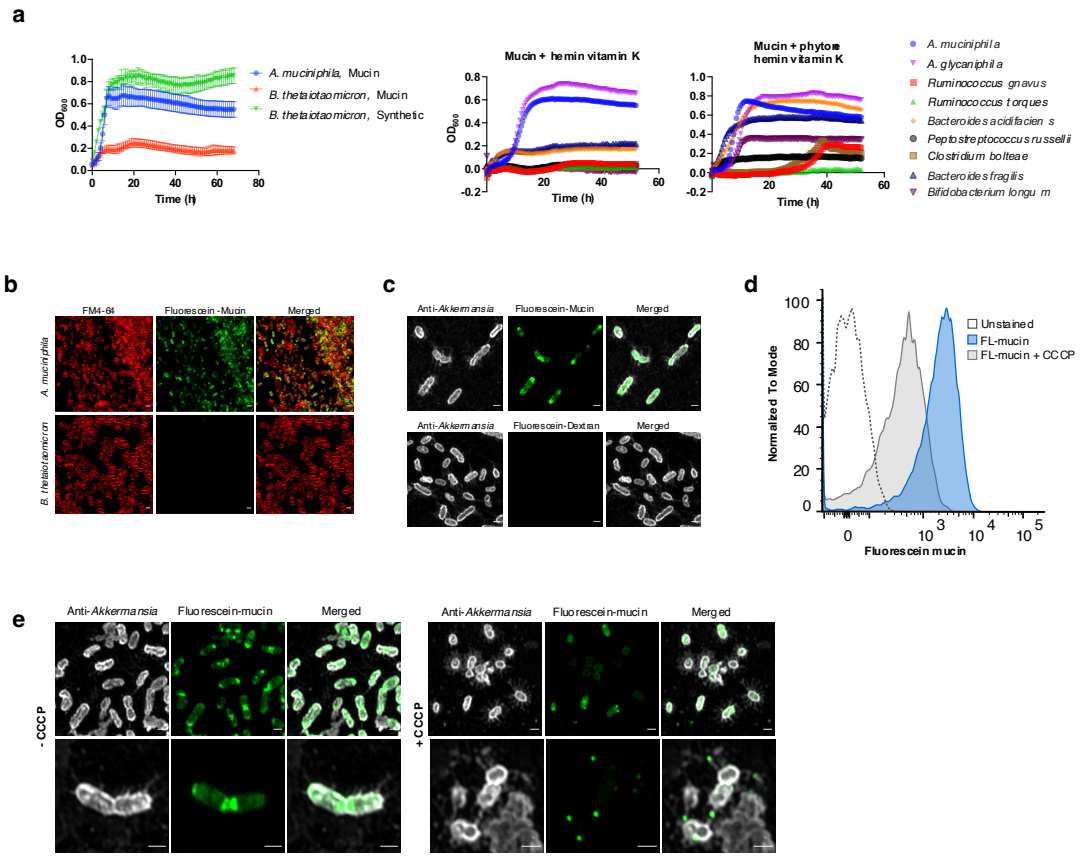

Figure S1. ***Akkermansia* sp. are mucin specialists and the acquisition of mucin by *A. muciniphila* is selective and energy dependent**. (**a**) Growth curves, as assessed by optical density (OD_{600}) of a range of Gram-positive and Gram-negative mucin-degrading intestinal microbes, including *A. muciniphila* and *A. glycaniphila*, in the indicated medium. (**b**) *A. muciniphila* and *Bacteroides thetaiotaomicron* grown with fluorescein-mucin. The cells were grown with fluorescein mucin in a modified version of synthetic media with 0.25% mucin as the sole carbon source. Membranes were labeled with FM4-64. (**c-d**) Mucin uptake is a specific and active process. *A. muciniphila* grown in the presence of either fluorescein-mucin or fluorescein-dextran for 3 h and stained with anti-*Akkermansia* anti-sera (anti-Akk) (**c**). Flow cytometric analysis of cells grown in the presence of fluorescein-mucin for 3h, with or without pre-treatment with CCCP. Cells for flow cytometry were gated for the anti-*Akkermansia* positive population and the numbers under each curve represent the mean fluorescent intensity of fluorescein-mucin (**d**). *A. muciniphila* grown with fluorescein-mucin for 3h without CCCP or with CCCP treatment (**e**). Scale bar = 1 μm. Error bars represent the standard error of the mean.

Fig. S2

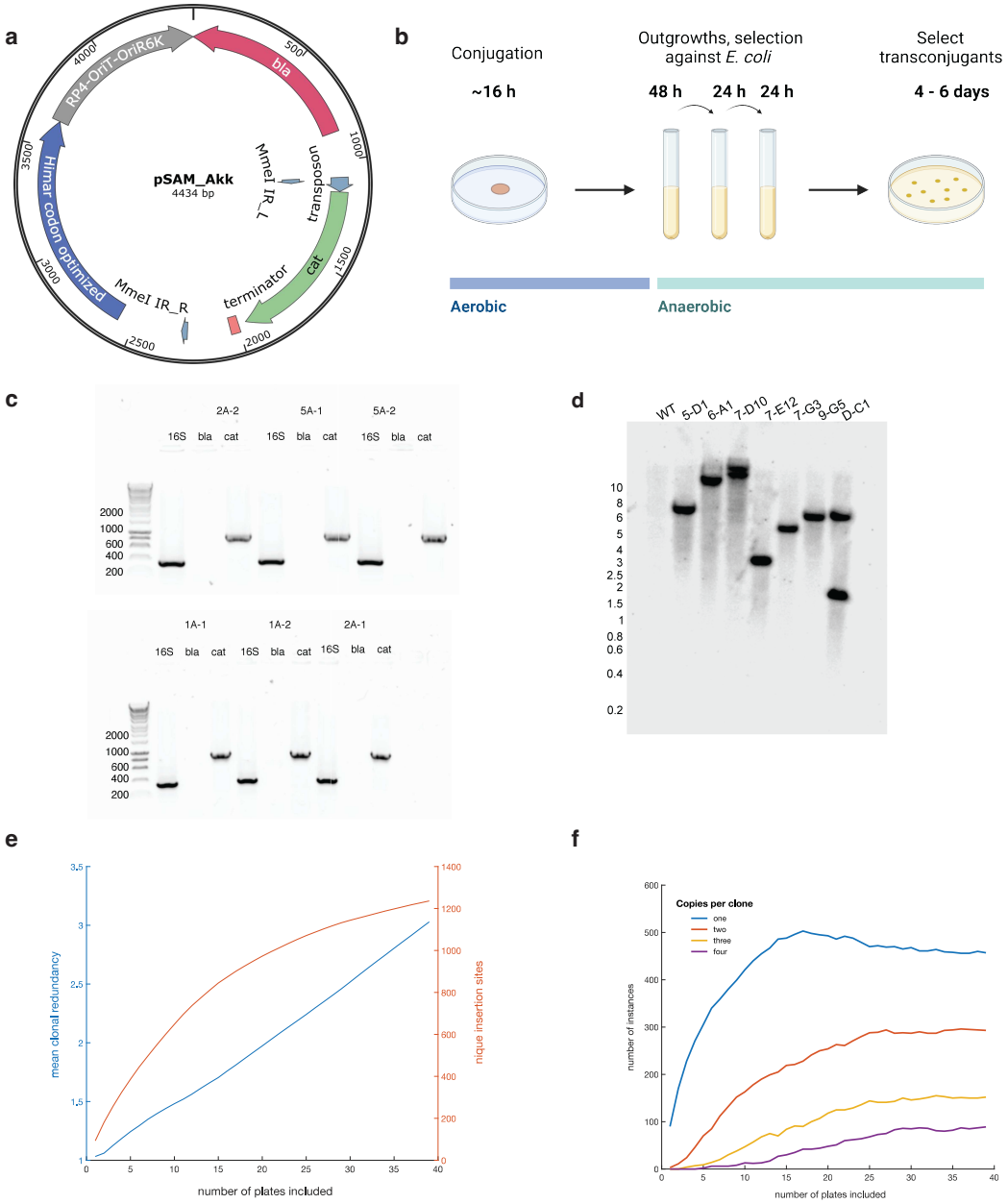

Figure S2. Transposon mutagenesis in *A. muciniphila*. (**a**) Map of the *A. muciniphila* optimized INSeq plasmid. (**b**) Overview of the *A. muciniphila* conjugation protocol. (**c**) PCR analysis confirming transposition. DNA from representative Tn mutants was amplified with primers for *A. muciniphila* specific 16S rRNA, the *bla* gene located on the delivery plasmid backbone, and the *cat* gene located with the transposon. (**d**) Southern blot analysis of Tn mutant DNA digested with *Hin*dIII and probed with DIG-labeled probes that recognize the *cat* gene in the Tn insert. (**e-f**) A Cartesian mapping strategy to identify Tn insertions. (e) Trade-off between genome coverage and clonal redundancy, using simulated subsets of the arrayed collection optimized for low redundancy. A series of 96-well plates drawn from the arrayed collection that minimizes clonal redundancy was identified by simulation. The tradeoff between increasing

genome coverage (orange) and increasing clonal redundancy (blue) as the number of plates included from the optimized series grows (X-axis). (**f**) Estimating location mapping accuracy for different sizes of the optimized library. The distribution of clonal replicates for increasing sizes of the optimized library was drawn from the simulation. The estimated number of clones present in one, two, three, and four replicates are shown as a function of increasing collection size. For orthogonal pooling and Cartesian location mapping the search space for an individual clone scales as the number of replicates to the power of the number of pooling dimensions. A clone present in only one well has a unique plate-well address (1^3), while a clone with present three times in the collection would be mapped to 27 potential Plate-Row-Col locations (3^3).

Fig. S3

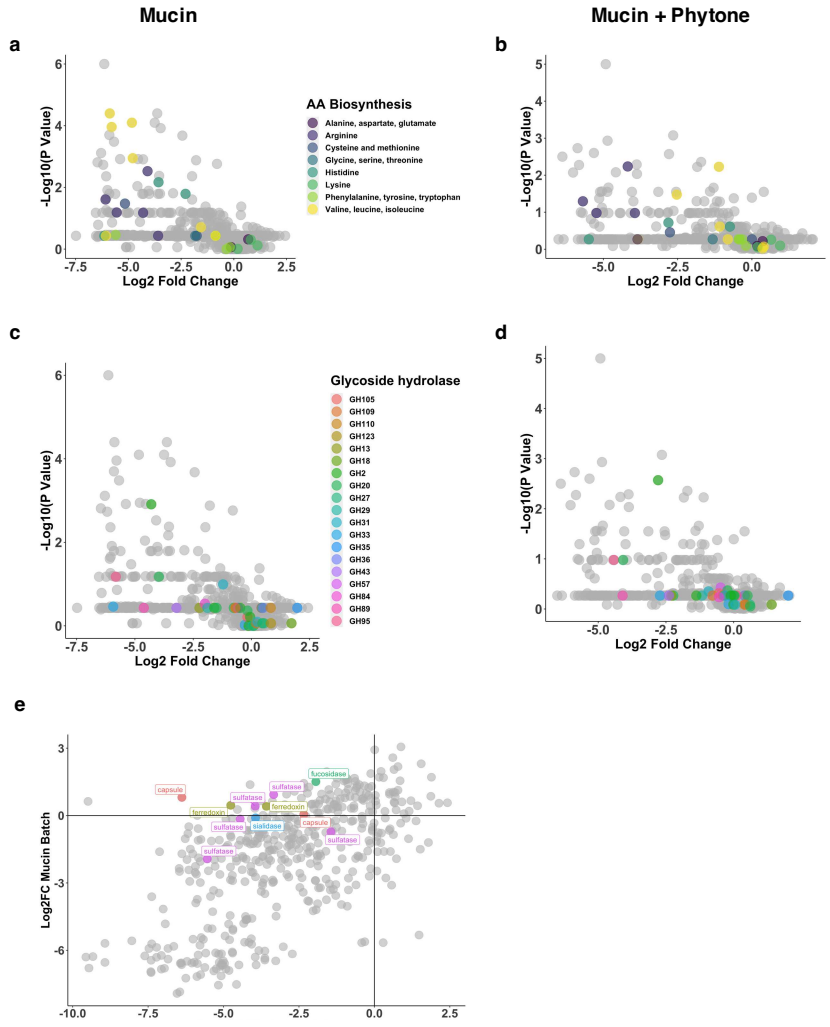

Figure S3. INSeq analysis of relative nutritional requirements for *A. muciniphila* to grow in mucin medium and the role of putative glycan hydrolases. Plot of INSeq data from Tn mutant pools grown for eight generations in mucin medium where each dot represents all inserts in a specific gene. Genes that belong to KEGG amino acid biosynthesis pathways are highlighted for cultures grown in (**a**) mucin medium and (**b**) mucin medium supplemented with Phytone. Predicted glycosyl hydrolases for *A. muciniphila* BAA-835 were identified using the CAZy

database and highlighted on the INSeq plot for cultures grown in (**c**) mucin and in (**d**) mucin medium with Phytone. ∈ Droplet-seq analysis of *A. muciniphila* grown in mucin medium microdroplets. Tn mutants (Arrayed Pool) were injected into a microfluidic device at a low density to generate on average less than one bacterium per droplet. The graph displays the INSeq analysis and Log_2 fold change for cultures grown in mucin in batch culture (8 generations) versus single cell growth in droplets (72h). Selected genes that were depleted in one condition relative to the other are highlighted on the plot.

Fig. S4

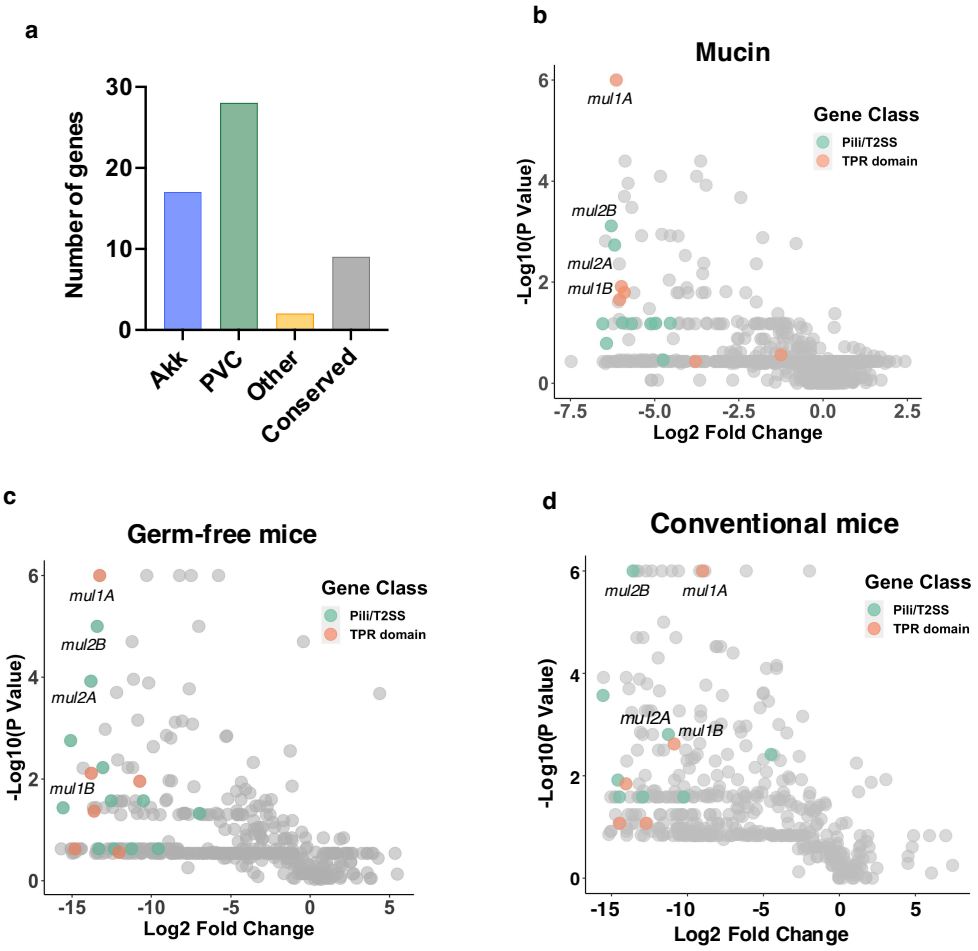

Figure S4. A significant proportion of *A. muciniphila* genes required for growth in mucin medium are specific to *Akkermansia/Verrucomicrobia*. (**a**) Number of genes required for optimal *A. muciniphila* growth in mucin medium that lack functional annotations. Genes corresponding to Tn mutants with a $Log_2 > 2$ fold decrease in abundance in mucin medium were used as the query for a BLAST search to identify potential homologs. The plot represents the number of genes encoding hypothetical proteins that were unique to *Akkermansia* spp. (Akk), homologs in other members of the PVC super phylum (PVC), homologs in other bacteria (other), and genes annotated as conserved hypothetical proteins (conserved). (**b**) Distribution of genes with Pfam designations belonging to pili or type II secretion families (Pili/T2SS), or TPR

families in the INSeq analysis of genes required for growth in mucin medium in vitro, (**c**) in the cecum of germ-free mice, and (**d**) in the cecum of conventional mice.

Fig. S5

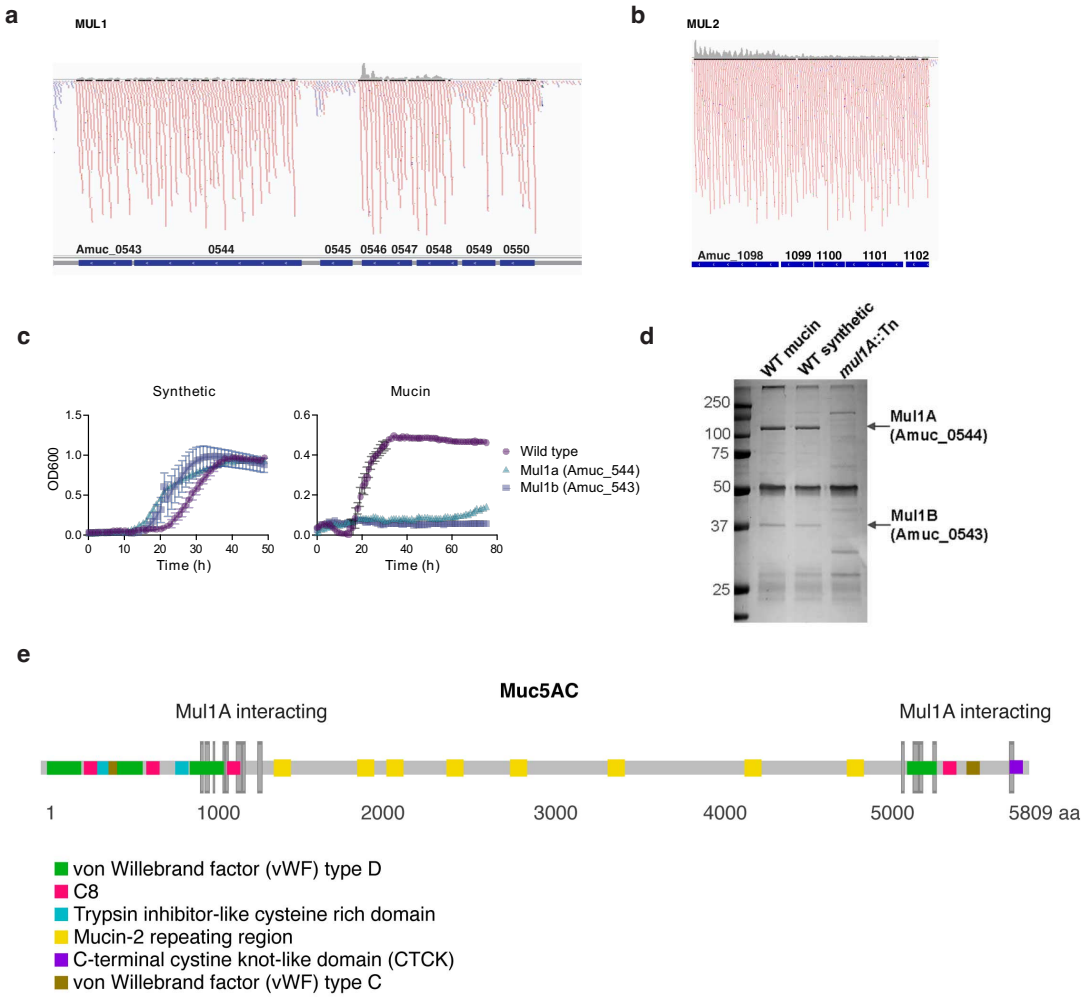

Figure S5. Evidence for the presence of a stable Mul1A-Mul1B protein complex. Transcriptional analysis of *Mul1* operons. View of RNA-seq reads generated from wild-type *A. muciniphila* grown in mucin medium mapped to genes in the (**a**) MUL1 and (**b**) MUL2 loci. (**c**) Growth curves for wild-type *A. muciniphila* and mutants in *mul1A* and *mul1B* grown in synthetic medium or with mucin as the sole carbon and nitrogen source. (**d**) Coomassie blue stained SDS-PAGE gel showing eluted proteins following immunoprecipitation with anti-Mul1 antibodies. Immunoprecipitations were performed with cell lysates from wild-type *A. muciniphila* and in *mul1A* mutants. (**e**) Depiction of Conserved Domains (colors) in Muc5AC and locations of peptides identified as co-precipitating with Mul1A (vertical bars).

Fig. S6

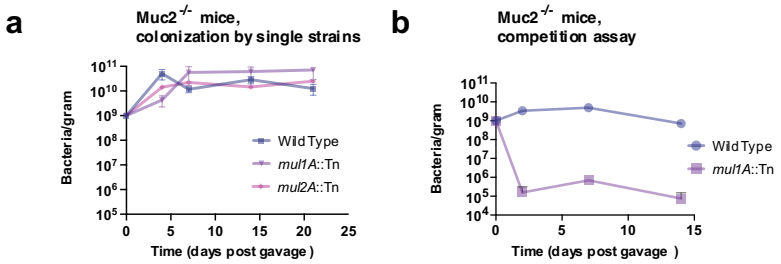

Fig. S7

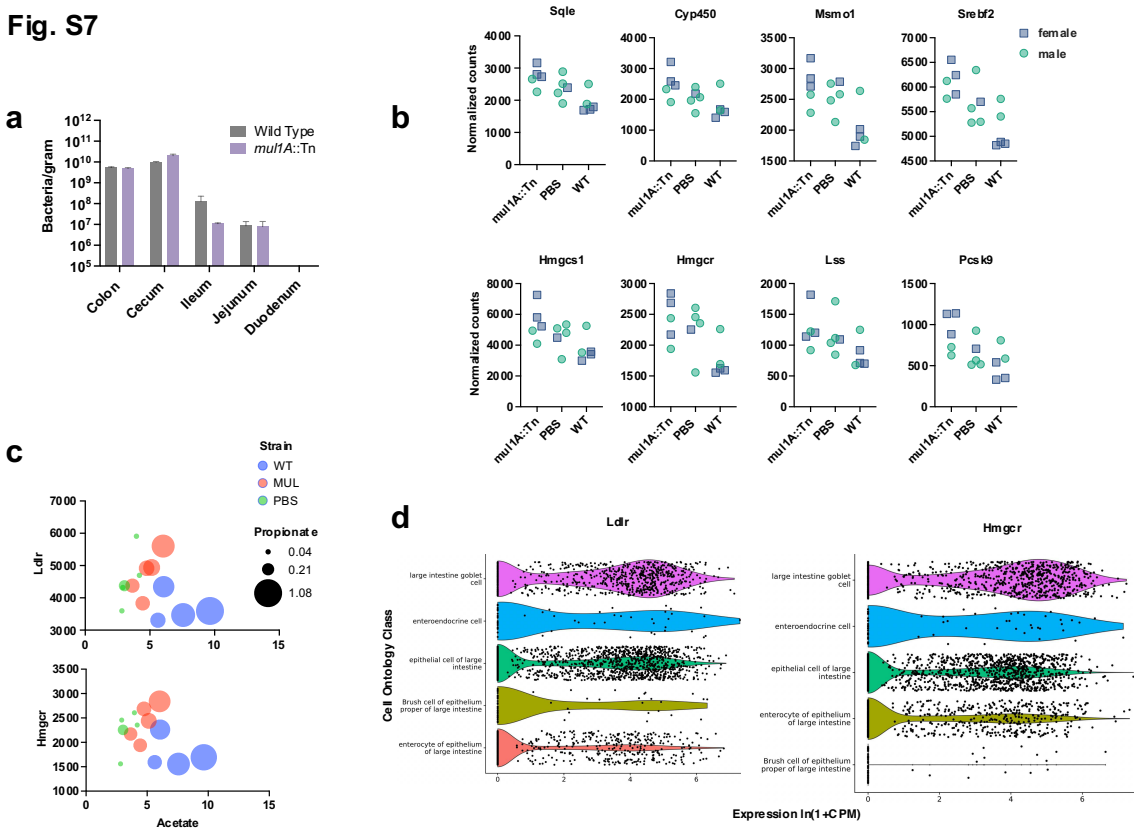

Figure S6. Muc2$^{-/-}$ mice can be colonized by *mul1* mutants. (a) Colonization of mucin deficient Muc2$^{-/-}$ mice with *A. muciniphila*. Each point represents the average *A. muciniphila* per gram of feces (n = 4-6). (**b**) Competition between wild-type *A. muciniphila* and the *mul1A*::Tn mutant in Muc2$^{-/-}$ mice. Mice were gavaged with a 1:1 mix of wild-type and mutant and abundance was monitored over time using strain specific primers. Each point represents the average amount of *A. muciniphila* (n = 4), error bars represent the standard error.

Figure S7. The impact of mucin utilization by *A. muciniphila* in colonization along the GI tract, SCFA production and transcriptional responses. (a) Abundance of *A. muciniphila* wild type and *mul1A* mutants along the GI tract of female GF mice (n =3). Intestinal contents were scraped from sections along the GI tract and *A. muciniphila* levels were quantified by qPCR. The

analysis was carried out with the same female mice that were used for RNA-seq. (**b**) Expression of cholesterol biosynthesis genes in male and female mice, and control mice gavaged with sterile PBS. (**c**) Normalized expression of genes that are pivotal to cholesterol biosynthesis (Hmgcr) and uptake (Ldlr) in relation to cecal acetate and propionate levels. (**d**) Representative single cell RNAseq expression data from the *Tabula Muris*[57]. Violin plots show expression of Ldlr and Hmgcr in mouse colonic epithelial and goblet cells.

Movie S1. View through an orthogonal section of a 3D-reconstructed STED image of a mucinosome inside *A. muciniphila*. Fluorescein mucin is shown in magenta and the *A. muciniphila* cell surface (anti-Akk) is cyan.

Movie S2. Live imaging of *A. muciniphila* grown with fluorescein labeled mucin under an agarose pad, with and without the addition of 50 μM CCCP. Images were captured every 30 s for 20 min. Arrows indicate cells with active mucinosome formation (untreated), and locations where mucin accumulates at the cell surface but fails to form mucinosomes in the presence of CCCP.

Table S1. Data tables with INSeq, RNAseq, mass spectrometry outputs, and primer and adaptor sequences.

References

1. Johansson, M. E. V, Larsson, J. M. H. & Hansson, G. C. The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proc. Natl. Acad. Sci. U. S. A.* **108 Suppl**, 4659–65 (2011).
2. Wlodarska, M. *et al*. Indoleacrylic acid produced by commensal *Peptostreptococcus* species suppresses inflammation. *Cell Host Microbe* **22**, 25-37.e6 (2017).
3. Zhang, X. *et al*. Substrate specificity of the galactokinase from the human gut symbiont *Akkermansia muciniphila* ATCC BAA-835. *Enzyme Microb. Technol.* **139**, 109568 (2020).
4. Xu, W., Yang, W., Wang, Y., Wang, M. & Zhang, M. Structural and biochemical analyses of β-N-acetylhexosaminidase Am0868 from *Akkermansia muciniphila* involved in mucin degradation. *Biochem. Biophys. Res. Commun.* **529**, 876–881 (2020).
5. Kosciow, K. & Deppenmeier, U. Characterization of a phospholipid-regulated β-galactosidase from *Akkermansia muciniphila* involved in mucin degradation. *Microbiologyopen* **8**, 1–11 (2019).
6. Malaker, S. A. *et al*. The mucin-selective protease StcE enables molecular and functional analysis of human cancer-associated mucins. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 7278–7287 (2019).
7. Chen, X. *et al*. Crystallographic evidence for substrate-assisted catalysis of β-N-acetylhexosaminidas from *Akkermansia muciniphila*. *Biochem. Biophys. Res. Commun.* **511**, 833–839 (2019).
8. Guo, B. *et al*. Cloning, purification and biochemical characterisation of a GH35 beta-

1,3/beta-1,6-galactosidase from the mucin-degrading gut bacterium *Akkermansia muciniphila*. *Glycoconj. J.* **35**, 255–263 (2018).
9. Meng, X. *et al*. A purified aspartic protease from *Akkermansia muciniphila* plays an important role in degrading muc2. *Int. J. Mol. Sci.* **21**, (2020).
10. Meng, X., Zhang, J., Wu, H., Yu, D. & Fang, X. *Akkermansia muciniphila* aspartic protease amuc_1434* Inhibits human colorectal cancer LS174T cell viability via TRAIL-mediated apoptosis pathway. *Int. J. Mol. Sci.* **21**, (2020).
11. Lee, K.C. *et al*. Phylum *Verrucomicrobia* representatives share a compartmentalized cell plan with members of bacterial phylum *Planctomycetes*. *BMC Microbiol.* **9**, 5 (2009).
12. Boedeker, C. *et al*. Determining the bacterial cell biology of *Planctomycetes*. *Nat. Commun.* **8**, 14853 (2017).
13. Reintjes, G., Arnosti, C., Fuchs, B. M. & Amann, R. An alternative polysaccharide uptake mechanism of marine bacteria. *ISME J.* **11**, 1640–1650 (2017).
14. Lonhienne, T. G. A. *et al*. Endocytosis-like protein uptake in the bacterium *Gemmata obscuriglobus*. *Proc. Natl. Acad. Sci.* **107**, 12883–12888 (2010).
15. Martinez-Garcia, M. *et al*. Capturing single cell genomes of active polysaccharide degraders: An unexpected contribution of *Verrucomicrobia*. *PLoS One* **7**, 1–11 (2012).
16. Goodman, A. L. *et al*. Identifying Genetic Determinants Needed to Establish a Human Gut Symbiont in Its Habitat. *Cell Host Microbe* **6**, 279–289 (2009).
17. Goodman, A. L., Wu, M. & Gordon, J. I. Identifying microbial fitness determinants by insertion sequencing using genome-wide transposon mutant libraries. *Nat. Protoc.* **6**, 1969–1980 (2011).
18. Anzai, I. A., Shaket, L., Adesina, O., Baym, M. & Barstow, B. Rapid curation of gene disruption collections using Knockout Sudoku. *Nat. Protoc.* **12**, 2110–2137 (2017).
19. Derrien, M., Vaughan, E. E., Plugge, C. M. & de Vos, W. M. *Akkermansia muciniphila* gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int. J. Syst. Evol. Microbiol.* **54**, 1469–1476 (2004).
20. Hansson, G. C. Mucins and the Microbiome. *Annu. Rev. Biochem.* **89**, 769–793 (2020).
21. Lensmire, J. M. & Hammer, N. D. Nutrient sulfur acquisition strategies employed by bacterial pathogens. *Curr. Opin. Microbiol.* **47**, 52–58 (2019).
22. 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).
23. Kostopoulos, I. *et al*. *Akkermansia muciniphila* uses human milk oligosaccharides to thrive in the early life conditions in vitro. *Sci. Rep.* **10**, 1–17 (2020).
24. Ottman, N. *et al*. Pili-like proteins of *Akkermansia muciniphila* modulate host immune responses and gut barrier function. *PLoS One* **12**, e0173004 (2017).
25. van Passel, M. W. J. *et al*. The genome of *Akkermansia muciniphila*, a dedicated intestinal mucin degrader, and its use in exploring intestinal metagenomes. *PLoS One* **6**, (2011).
26. Ottman, N. *et al*. Genome-Scale Model and Omics Analysis of Metabolic Capacities of *Akkermansia muciniphila* Reveal a Preferential Mucin-Degrading Lifestyle. *Appl. Environ. Microbiol.* **83**, 1–15 (2017).
27. van der Ark, K. C. H. *et al*. Model-driven design of a minimal medium for *Akkermansia muciniphila* confirms mucus adaptation. *Microb. Biotechnol.* **11**, 476–485 (2018).
28. Reichardt, N. *et al*. Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. *ISME J.* **8**, 1323–1335 (2014).

29. Thibault, D. *et al*. Droplet Tn-Seq combines microfluidics with Tn-Seq for identifying complex single-cell phenotypes. *Nat. Commun.* **10**, (2019).
30. Ottman, N. *et al*. Characterization of Outer Membrane Proteome of *Akkermansia muciniphila* Reveals Sets of Novel Proteins Exposed to the Human Intestine. *Front. Microbiol.* **7**, 1157 (2016).
31. Shon, D. J. *et al*. An enzymatic toolkit for selective proteolysis, detection, and visualization of mucin-domain glycoproteins. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 21299–21307 (2020).
32. Schwalm, N. D. & Groisman, E. A. Navigating the Gut Buffet: Control of Polysaccharide Utilization in *Bacteroides* spp. *Trends Microbiol.* **25**, 1005–1015 (2017).
33. Mistry, J. *et al*. Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419 (2021).
34. Marchler-Bauer, A. *et al*. CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* **45**, D200–D203 (2017).
35. Cortajarena, A. L. & Regan, L. Ligand binding by TPR domains. *Protein Sci.* **15**, 1193–1198 (2006).
36. Plovier, H. *et al*. A purified membrane protein from *Akkermansia muciniphila* or the pasteurized bacterium improves metabolism in obese and diabetic mice. *Nat. Med.* **23**, 107–113 (2017).
37. Xiang, R., Wang, J., Xu, W., Zhang, M. & Wang, M. Amuc_1102 from *Akkermansia muciniphila* adopts an immunoglobulin-like fold related to archaeal type IV pilus. *Biochem. Biophys. Res. Commun.* **547**, 59–64 (2021).
38. Mou, L. *et al*. Crystal structure of monomeric Amuc-1100 from *Akkermansia muciniphila*. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **76**, 168–174 (2020).
39. Velcich, A. *et al*. Colorectal cancer in mice genetically deficient in the mucin Muc2. *Science*. **295**, 1726–1729 (2002).
40. Becken, B. *et al*. Genotypic and Phenotypic Diversity among Human Isolates of *Akkermansia muciniphila*. *MBio* **12**, 1–21 (2021).
41. Roux, D. *et al*. Identification of Poly-N-acetylglucosamine as a major polysaccharide component of the *Bacillus subtilis* biofilm matrix. *J. Biol. Chem.* **290**, 19261–19272 (2015).
42. Hospenthal, M. K., Costa, T. R. D. & Waksman, G. A comprehensive guide to pilus biogenesis in Gram-negative bacteria. *Nat. Rev. Microbiol.* **15**, 365–379 (2017).
43. Glenwright, A. J. *et al*. Structural basis for nutrient acquisition by dominant members of the human gut microbiota. *Nature* **541**, 407–411 (2017).
44. Bolam, D. N. & van den Berg, B. TonB-dependent transport by the gut microbiota: novel aspects of an old problem. *Curr. Opin. Struct. Biol.* **51**, 35–43 (2018).
45. Wang, J. *et al*. The variable oligomeric state of Amuc_1100 from *Akkermansia muciniphila*. *J. Struct. Biol.* **212**, 107593 (2020).
46. Strom, M. S., Nunn, D. N. & Lory, S. A single bifunctional enzyme, PilD, catalyzes cleavage and N-methylation of proteins belonging to the type IV pilin family. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 2404–2408 (1993).
47. Holden, H. M., Rayment, I. & Thoden, J. B. Structure and Function of Enzymes of the Leloir Pathway for Galactose Metabolism. *J. Biol. Chem.* **278**, 43885–43888 (2003).
48. Wahlgren, W. Y. *et al*. Substrate-bound outward-open structure of a Na+-coupled sialic acid symporter reveals a new Na+ site. *Nat. Commun.* **9**, 1–14 (2018).

49. Faham, S. *et al*. The Crystal Structure of a Sodium Galactose Transporter Reveals Mechanistic Insights into Na+/Sugar Symport. *Science*. **321**, 810–814 (2008).
50. Terrapon, N. *et al*. PULDB: The expanded database of Polysaccharide Utilization Loci. *Nucleic Acids Res*. **46**, D677–D683 (2018).
51. Wu, M. *et al*. The Dynamic Changes of Gut Microbiota in Muc2 Deficient Mice. *Int. J. Mol. Sci*. **19**, 2809 (2018).
52. Bosch, S. *et al*. Fecal Amino Acid Profiles Exceed Accuracy of Serum Amino Acids in Diagnosing Pediatric Inflammatory Bowel Disease. *J. Pediatr. Gastroenterol. Nutr*. **71**, 371–375 (2020).
53. Bjerrum, J. T. *et al*. Metabonomics of human fecal extracts characterize ulcerative colitis, Crohn's disease and healthy individuals. *Metabolomics* **11**, 122–133 (2015).
54. Sugihara, K., Morhardt, T. L. & Kamada, N. The role of dietary nutrients in inflammatory bowel disease. *Front. Immunol*. **10**, 1–16 (2019).
55. Depommier, C. *et al*. Pasteurized *Akkermansia muciniphila* increases whole-body energy expenditure and fecal energy excretion in diet-induced obese mice. *Gut Microbes* **11**, 1231–1245 (2020).
56. Lukovac, S. *et al*. Differential modulation by *Akkermansia muciniphila* and faecalibacterium prausnitzii of host peripheral lipid metabolism and histone acetylation in mouse gut organoids. *MBio* **5**, 1–10 (2014).
57. Schaum, N. *et al*. Single-cell transcriptomics of 20 mouse organs creates a *Tabula Muris*. *Nature* **562**, 367–372 (2018).
58. Shen, J. *et al*. Low-Density Lipoprotein Receptor Signaling Mediates the Triglyceride-Lowering Action of *Akkermansia muciniphila* in Genetic-Induced Hyperlipidemia. *Arterioscler. Thromb. Vasc. Biol*. **36**, 1448–1456 (2016).
59. Depommier, C. *et al*. Supplementation with *Akkermansia muciniphila* in overweight and obese human volunteers: a proof-of-concept exploratory study. *Nat. Med*. **25**, 1096–1103 (2019).
60. Stams, A. J. M., Van Dijk, J. B., Dijkema, C. & Plugge, C. M. Growth of syntrophic propionate-oxidizing bacteria with fumarate in the absence of methanogenic bacteria. *Appl. Environ. Microbiol*. **59**, 1114–1119 (1993).
61. Arnosti, C. Fluorescent derivatization of polysaccharides and carbohydrate-containing biopolymers for measurement of enzyme activities in complex media. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci*. **793**, 181–191 (2003).
62. DuBois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. A. & Smith, F. Colorimetric Method for Determination of Sugars and Related Substances. *Anal. Chem*. **28**, 350–356 (1956).
63. Daetwyler, S., Modes, C. D. & Fiolka, R. Fiji plugin for annotating movies with custom arrows. *Biol. Open* **9**, 3–6 (2020).
64. Issa, S. M. A., Schulz, B. L., Packer, N. H. & Karlsson, N. G. Analysis of mucosal mucins separated by SDS-urea agarose polyacrylamide composite gel electrophoresis. *Electrophoresis* **32**, 3554–3563 (2011).
65. Aronesty, E. Comparison of Sequencing Utility Programs. *Open Bioinforma. J*. **7**, 1–8 (2013).
66. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. (2010).
67. Dobin, A. *et al*. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21

(2013).
68. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
69. Yu, G., Wang, L.G., Han, Y. & He, Q.Y. ClusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omi. A J. Integr. Biol.* **16**, 284–287 (2012).
70. Wolk, C. P. *et al*. Paired cloning vectors for complementation of mutations in the cyanobacterium *Anabaena* sp. strain PCC 7120. *Arch. Microbiol.* **188**, 551–563 (2007).
71. Stothard, P. The Sequence Manipulation Suite: JavaScript Programs for Analyzing and Formatting Protein and DNA Sequences. *Biotechniques* **28**, 1102–1104 (2000).
72. O'Toole, G. A. *et al*. Genetic approaches to study of biofilms. *Methods Enzym.* **310**, 91–109 (1999).
73. Najah, M., Griffiths, A. D. & Ryckelynck, M. Teaching single-cell digital analysis using droplet-based microfluidics. *Anal. Chem.* **84**, 1202–1209 (2012).
74. Collado, M. C., Derrien, M., Isolauri, E., De Vos, W. M. & Salminen, S. Intestinal integrity and *Akkermansia muciniphila*, a mucin-degrading member of the intestinal microbiota present in infants, adults, and the elderly. *Appl. Environ. Microbiol.* **73**, 7767–7770 (2007).
75. Shames, S. R. *et al*. Multiple *Legionella pneumophila* effector virulence phenotypes revealed through high-throughput analysis of targeted mutant libraries. *Proc. Natl. Acad. Sci.* **114**, E10446–E10454 (2017).
76. Pritchard, J. R. *et al*. ARTIST: High-Resolution Genome-Wide Assessment of Fitness Using Transposon-Insertion Sequencing. *PLoS Genet.* **10**, e1004782 (2014).
77. DeJesus, M. A., Ambadipudi, C., Baker, R., Sassetti, C. & Ioerger, T. R. TRANSIT - A Software Tool for Himar1 TnSeq Analysis. *PLOS Comput. Biol.* **11**, e1004401 (2015).
78. Karp, P. D. *et al*. The BioCyc collection of microbial genomes and metabolic pathways. *Brief. Bioinform.* **20**, 1085–1093 (2018).
79. Tanabe, M. & Kanehisa, M. Using the KEGG database resource. *Current Protocols in Bioinformatics* (2012).
80. Linlin, Y. ggVenn: Venn Diagram by ggplot2 [R package]. (2020).
81. Wickham, H. *ggplot2: elegant graphics for data analysis*. (Springer-Verlag, 2016).
82. Yutani, H. gghighlight. [R package] (2021).
83. Zhou, Y. *et al*. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).
84. Blighe, K. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling [R package]. (2019). doi:10.18129/B9.bioc.EnhancedVolcano
85. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786 (2011).
86. Shannon, P. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **13**, 2498–2504 (2003).
87. Paley, S. *et al*. The omics dashboard for interactive exploration of gene-expression data. *Nucleic Acids Res.* **45**, 12113–12124 (2017).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- MovieS1STED.mp4
- MovieS2CCCP.mp4
- MovieS2A.mp4
- MovieS2BCCCP.mp4
- TableS1ALLwDroplets.xlsx