nature microbiology

Article

https://doi.org/10.1038/s41564-023-01407-w

A genetic system for *Akkermansia muciniphila* reveals a role for mucin foraging in gut colonization and host sterol biosynthesis gene expression

Received: 21 March 2022

Accepted: 10 May 2023

Published online: 19 June 2023

Check for updates

Lauren E. Davey[1,2,3] ✉, Per N. Malkus[1,2], Max Villa[1,2], Lee Dolat[1,2], Zachary C. Holmes[1,2], Jeff Letourneau[1,2], Eduard Ansaldo[4], Lawrence A. David[1,2], Gregory M. Barton[4] & Raphael H. Valdivia[1,2] ✉

Akkermansia muciniphila, a mucophilic member of the gut microbiota, protects its host against metabolic disorders. Because it is genetically intractable, the mechanisms underlying mucin metabolism, gut colonization and its impact on host physiology are not well understood. Here we developed and applied transposon mutagenesis to identify genes important for intestinal colonization and for the use of mucin. An analysis of transposon mutants indicated that de novo biosynthesis of amino acids was required for *A. muciniphila* growth on mucin medium and that many glycoside hydrolases are redundant. We observed that mucin degradation products accumulate in internal compartments within bacteria in a process that requires genes encoding pili and a periplasmic protein complex, which we term mucin utilization locus (MUL) genes. We determined that MUL genes were required for intestinal colonization in mice but only when competing with other microbes. In germ-free mice, MUL genes were required for *A. muciniphila* to repress genes important for cholesterol biosynthesis in the colon. Our genetic system for *A. muciniphila* provides an important tool with which to uncover molecular links between the metabolism of mucins, regulation of lipid homeostasis and potential probiotic activities.

Akkermansia muciniphila is a Gram-negative bacterium that colonizes the human gastrointestinal tract and has been associated with various beneficial health effects, including protection from metabolic disease[1], neurological disorders[2] and infection[3] and an enhanced response to cancer immunotherapy[4]. Some factors that contribute to *A. muciniphila*'s metabolic and immunomodulatory activity have been identified[1,5,6]; however, most of the genes that contribute to colonization of the gastrointestinal tract and regulate interactions with the host remain to be identified.

A. muciniphila uses mucins as its preferred nutrient source. Mucins are large, highly glycosylated proteins that make up the bulk of the intestinal mucus layer that separates the host epithelia from the microbiota[7]. The predominant mucins of the gastrointestinal tract—Muc2 in the colon and Muc5AC in the gastric mucosa—consist

[1]Department of Molecular Genetics and Microbiology, Duke University, Durham, NC, USA. [2]Duke Microbiome Center, Duke University, Durham, NC, USA. [3]Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia, Canada. [4]Division of Immunology and Pathogenesis, Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA, USA. ✉e-mail: laurendavey@uvic.ca; raphael.valdivia@duke.edu

of a protein backbone rich in serine, threonine and proline residues, decorated with *O*-linked glycans. 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[8]. *A. muciniphila* is particularly adept at using mucins, reaching a high optical density when cultured in liquid medium with gastric mucin as the sole carbon and nitrogen source (Extended Data Fig. 1a). *A. muciniphila* encodes multiple putative glycoside hydrolases and other mucolytic enzymes, some of which have been characterized in vitro[9–16]. *A. muciniphila*'s repertoire of mucin-degrading enzymes includes glycoside hydrolases belonging to the GH20 (refs. 14,17), GH35 (refs. 15,16) and GH16 (ref. 12) enzyme families, an aspartic protease[13], peptidases[9] and *O*-glycopeptidases[10,11]. While several *A. muciniphila* glycoside hydrolases have been characterized, the intractability of *Akkermansia* species to molecular genetic manipulation has limited our understanding of the role the metabolism of mucin by this microbe plays in colonization.

In this Article, we apply transposon (Tn) mutagenesis to identify genetic determinants of mucin utilization and colonization in *A. muciniphila*. We found that *A. muciniphila* accumulates mucin in intracellular structures and identified a mucin transport system required for mucin uptake. Using mutants unable to grow on mucin, we show that access to mucin is required for *A. muciniphila* to compete against other gut microbes and for stable engraftment in the gastrointestinal tract. Finally, we demonstrate that mucin metabolism represses the expression of lipid biosynthesis genes in the colon, revealing a link between this key characteristic of *A. muciniphila* and host physiology.

Results

Akkermansia accumulates mucins in intracellular compartments

To visualize mucin uptake, we used fluorescein-labelled mucin glycans generated from porcine gastric mucin. Commercial porcine gastric mucin, which is pepsin digested and partially purified, consists of 13 major *O*-linked glycans composed of the mucin sugars *N*-acetylglucosamine, *N*-acetylgalactosamine, galactose and fucose[18]. Additional mucin modifications include sialylation and sulfation. We observed an accumulation of these labelled mucins or their degradation products in intracellular compartments using super-resolution stimulated emission depletion (STED) microscopy (Fig. 1a and Supplementary Video 1). The formation of these intracellular structures, which we operationally term mucinosomes, appeared to be specific to *Akkermansia* species as we did not observe them in *Bacteroides thetaiotaomicron*, which also metabolizes mucin (Extended Data Fig. 1b).

To examine the kinetics of mucin uptake, we pulsed *A. muciniphila* with fluorescein-labelled mucin and analysed the fluorescence associated with bacteria over time by flow cytometry. The association with fluorescein-labelled mucin occurred rapidly and peaked by 6 h after labelling, followed by a subsequent decrease between 6 and 24 h (Fig. 1b). Next, we used live-cell imaging to monitor the uptake of fluorescein-labelled mucin in real time (Fig. 1c). Mucinosome formation occurred within minutes and was a highly dynamic process (Supplementary Video 2). These structures did not form after incubation with fluorescein-labelled dextran (Extended Data Fig. 1c) and were diminished in the presence of the ionophore carbonyl cyanide 3-chlorophenylhydrazone (CCCP), which dissipates the cell's proton motive force (Extended Data Fig. 1d,e and Supplementary Video 2). Carboxyfluorescein diacetate (CFDA), a cell-permeable dye that fluoresces green upon cleavage by esterases, confirmed that cells remained viable during CCCP treatment and continued to passively take up molecules (Extended Data Fig. 1f). Overall, these observations suggest that the acquisition of mucin glycans by *A. muciniphila* is a selective and energy-dependent process.

Development of a system for Tn mutagenesis

To define the genetic requirements for growth on mucin, we developed a system for Tn mutagenesis after identifying a suitable selectable marker (chloramphenicol acetyl transferase), defining conditions for DNA conjugation with an *Escherichia coli* donor and codon optimizing the Himar1c9 transposase[19] for expression in *Akkermansia* (Extended Data Fig. 2a,b). The absence of sequences from the conjugative plasmid in *A. muciniphila* chloramphenicol acetyl transferase-resistant colonies confirmed that transposition had occurred (Extended Data Fig. 2c) and a Southern blot analysis of total DNA extracted from these strains indicated that most mutants had a single Tn insertion (Extended Data Fig. 2d).

We assembled two mutant libraries, one consisting of a mixed pool of Tn mutants and one in which 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[19]. The pooled Tn mutant library consisted of 2,680 unique insertions in 721 genes (33% of the total coding sequence), 198 intergenic regions and five transfer RNAs. The arrayed Tn library consisted of 1,063 unique insertions in 406 genes, 58 intergenic regions and two transfer RNA genes. We then used a Cartesian pooling strategy[20] to map the locations of the majority of the Tn insertions to individual mutants in the arrayed collection.

Proteins of unknown function enable growth on mucin

To identify genes required for growth in 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[21]. 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 that in 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. 1d, Extended Data Fig. 3a and Supplementary Data 1). This result was not surprising given that these amino acids are not abundant in the protein core of gastrointestinal mucins[22]. Mutants with Tn insertions in genes required for assimilatory sulfate metabolism, which are required for Cys and Met biosynthesis[23], were also significantly depleted. The growth defect of these mutants could be rescued by the addition of protein hydrolysates (Extended Data Fig. 3b and Supplementary Data 1). These results indicate that *A. muciniphila* relies on de novo biosynthesis of amino acids during growth in mucin, presumably because some amino acids in the mucin protein backbone become rate limiting to support the bacterium's anabolic needs.

We anticipated that glycoside hydrolases would be required for *A. muciniphila* growth on mucin. *A. muciniphila* strain Muc^T is predicted to encode 60 glycoside hydrolases belonging to 24 different families. Our Tn mutant collection included insertions in genes encoding 38 glycoside hydrolases, representing 20 families[24]. Following growth in mucin medium, only eight glycoside hydrolases had a $\log_2$[fold change] decrease in abundance of >2, indicating that most glycoside hydrolases (30/38) were not essential under these conditions. Several mutations displayed a slight growth advantage, including mutations in genes encoding previously characterized enzymes such as a β-galactosidase[15], a β-hexosaminidase[25,26], a β-*N*-acetyl hexosaminidase[14] and a predicted mucin-binding chitinase[27]. Nonetheless, mutations in a subset of genes encoding glycoside hydrolases resulted in profound growth defects in mucin medium (Extended Data Fig. 3c,d and Supplementary Data 1), including those in genes predicted to encode a β-galactosidase (GH2), an α-*N*-acetylglucosaminidase (GH89), an α-amylase (GH13), a galactosidase (GH43) and a β-hexosaminidase (GH20). Mutations in genes encoding enzymes that catalyse the initial steps in mucin breakdown resulted in particularly strong growth defects. These included genes encoding a sialidase (GH33)[25], a fucosidase (GH95)[25] and an outer

https://doi.org/10.1038/s41564-023-01407-w

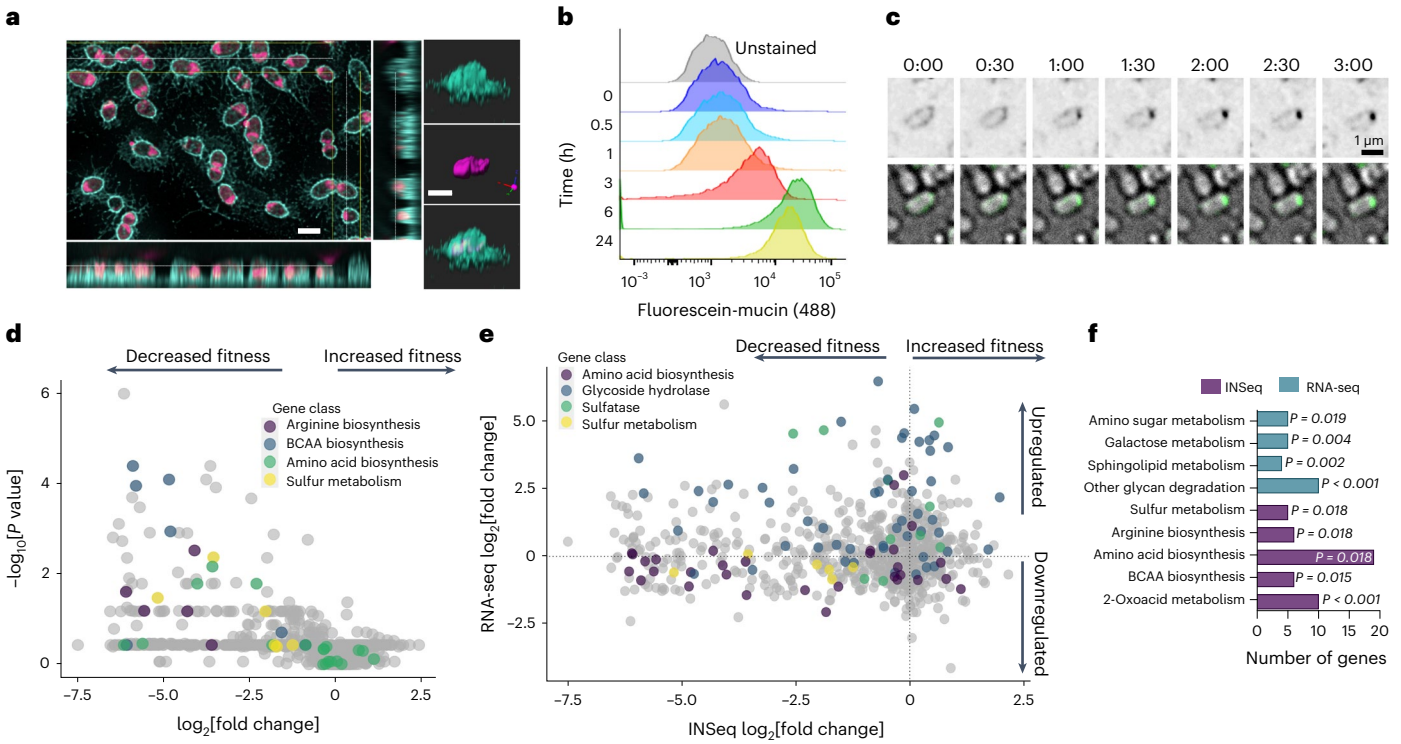


Fig. 1 | *A. muciniphila* accumulates mucin glycans and requires amino acid biosynthesis for replication in mucin as a sole carbon and nitrogen source. **a**–**c**, Mucin accumulates within *A. muciniphila* intracellular compartments. **a**, STED imaging of *A. muciniphila* showing mucin glycans and mucin degradation products within intracellular compartments (mucinosomes). Bacteria grown with fluorescein-labelled mucin (purple) were stained with anti-*Akkermansia* antisera (cyan). The bottom panel shows orthogonal views of *A. muciniphila* (white lines denote the orthogonal plane) and the right panel shows a 3D reconstruction of a single cell. The images are representative of three independent experiments. Scale bar, 1 µm. **b**, Flow cytometry of *A. muciniphila* incubated with fluorescein-labelled mucin over 24 h. **c**, Live-cell imaging over 3 min showing the accumulation of fluorescein-labelled mucin (green) at the *A. muciniphila* cell pole (Supplementary Video 2). **d**–**f**, Amino acid biosynthesis is required for optimal growth of *A. muciniphila* on mucin. **d**, 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. BCAA: branched chain amino acid. Unadjusted *P* values were determined by Mann–Whitney *U*-test. **e**, Comparison of the transcriptional response to growth on mucin (RNA-seq) and the fitness of the corresponding mutants (INSeq). Each dot represents one gene and only genes that were detected in both the INSeq and RNA-seq datasets are shown. The *y* axis shows the $\log_2$[fold change] in expression in mucin versus synthetic media, whereas the *x* axis shows the $\log_2$[fold change] in abundance between the input Tn mutant pools and output. The dotted lines represent no change. **f**, Overrepresentation analysis to identify differentially abundant KEGG pathways ($\log_2$[fold change] > 1.5; $P < 0.05$) detected by INSeq and RNA-seq following growth in mucin. The *P* values were calculated by hypergeometric distribution.

membrane-associated endo *O*-glycanase (GH16) that can break down mucin in vitro[12].

The overlap between genes required for mucin utilization by *A. muciniphila*, as determined by metabolic modelling[28] and insertion sequencing (INSeq) analysis, was low, suggesting that most of the enzymes predicted to have roles in mucin degradation are functionally redundant (Supplementary Data 1).

It is possible that the low representation of mutations in glycoside hydrolase genes among strains defective for the consumption of mucin is the result of *trans*-complementation by neighbouring cells. To test whether cross-feeding in batch cultures led us to miss important factors needed for mucin utilization, we performed droplet Tn-Seq[29] to analyse *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 were largely independent of whether the bacteria were grown in isolation (Extended Data Fig. 3e and Supplementary Data 1), suggesting that any cross-feeding is limited. Exceptions to this observation include genes involved in capsule production and genes encoding a predicted sialidase, fucosidase and sulfatase—all enzymes that can act on the terminal motifs of mucin glycans. These results are consistent with previous findings indicating that fucosidase activity is associated with the *Akkermansia* outer membranes[30].

Next, we performed an RNA sequencing (RNA-seq) analysis of *A. muciniphila* cultured in the same mucin medium as was 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 gene set enrichment analysis showed that pathways involved in glycan degradation and galactose metabolism dominated this response (Fig. 1e,f), as previously observed[28]. Genes that were highly expressed in response to mucin included those encoding an *O*-glycopeptidase[9], putative sulfatases and 24 glycoside hydrolase enzymes. The correlation between the genes required for growth in mucin and those that were strongly upregulated by mucin was low (Fig. 1e,f). These observations indicate that, unlike the polysaccharide utilization genes of *Bacteroidetes*[31], 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[27]. We mapped Tn insertions to 226 genes annotated as encoding either conserved hypothetical or hypothetical proteins, accounting for 31.3% of all of the mutated coding sequences in our pooled library. Mutations in 54/179 genes encoding hypothetical proteins that were either *Akkermansia* specific or had homologues restricted to the PVC superphylum (a superphylum of

https://doi.org/10.1038/s41564-023-01407-w

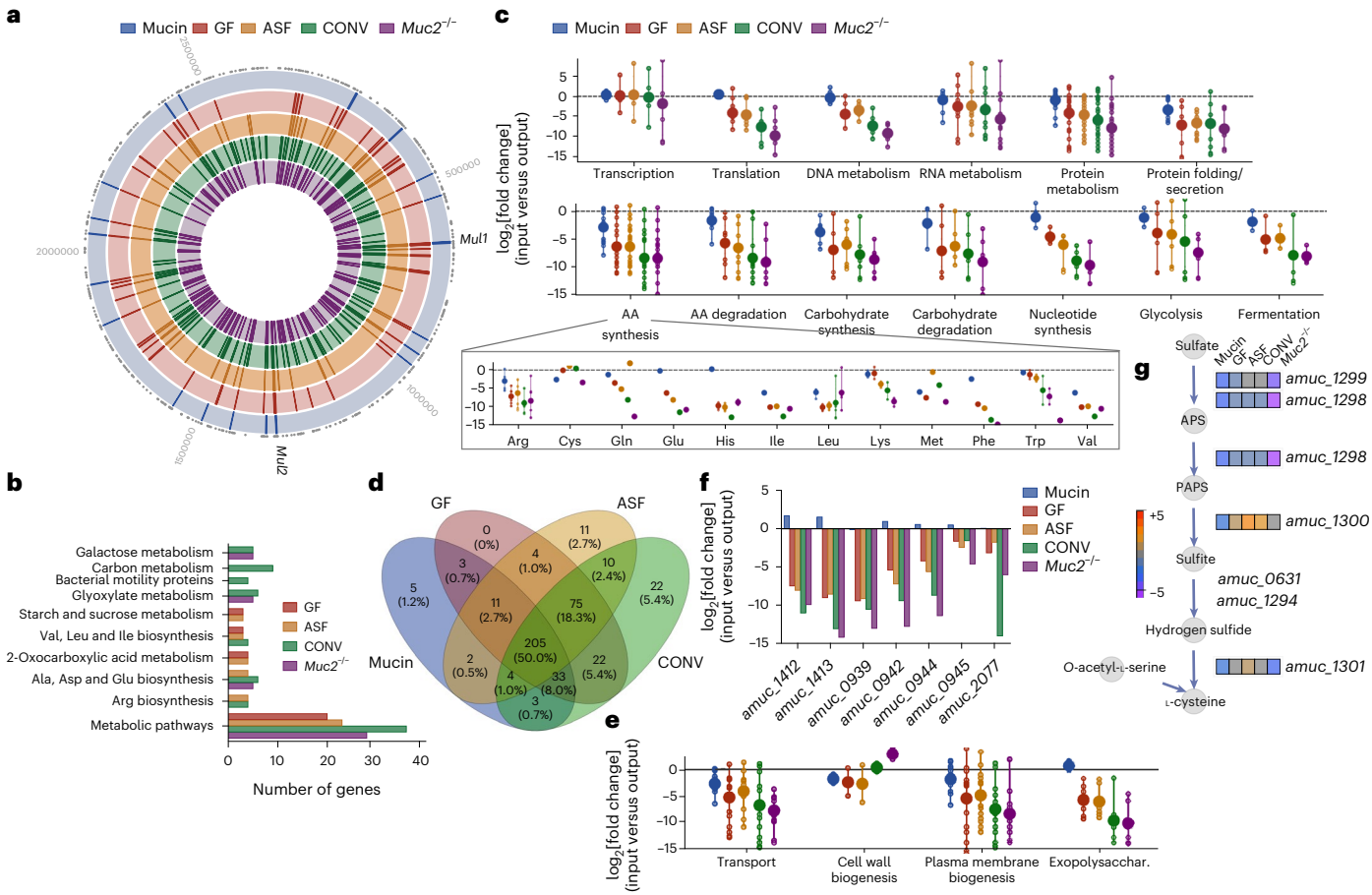


Fig. 2 | The metabolic requirements for *A. muciniphila* to colonize the gastrointestinal tract increase as the host microbiota becomes more complex. a, Chromosomal locations of genes with Tn insertions that produced a significant change in abundance after growth in mucin medium (Mucin) or in the caecum of the following mouse models: germ free (GF), ASF, CONV and *Muc2*$^{-/-}$. The lines represent the locations of Tn inserts that resulted in growth defects ($\log_2$[fold change] > 5; $P < 0.05$; unadjusted Mann–Whitney *U*-test). The dots along the outer ring indicate the locations of all Tn insertions and the numbers represent chromosomal location (bp). **b**, KEGG pathways overrepresented ($q < 0.05$) among mutants that displayed significant fitness defects. **c**, Global analysis of the abundance of mutants with fitness defects in mucin medium, as determined by INSeq and analysed with Omics Dashboard[76]. 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. The connecting lines show the range of values for each pathway. The boxed part shows the mean $\log_2$[fold change] for individual amino acid biosynthesis pathways. **d**, Venn diagram showing overlapping genes required for growth under various conditions. For each condition, genes with a $\log_2$[fold change] in abundance of >1 were compared. The number of shared genes and the percentage of the total mutant pool are shown for each condition. **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, cell wall biogenesis and the 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 ASR pathway. The heatmaps represent the $\log_2$[fold change] of input versus output for mutants in the pathway. APS, adenosine 5′-phosphosulfate; PAPS, 3′-phosphoadenosine-5′-phosphosulfate.

bacteria named after its three important members: Planctomycetota, Verrucomicrobiota and Chlamydiota) led to decreases in abundance of greater than twofold during growth in mucin medium (Extended Data Fig. 4a and Supplementary Data 1). Manual curation of these proteins of unknown function for associated Pfam[32] motifs and conserved domains[33] indicated that 19% either contained a tetratricopeptide repeat domain (which are known to mediate protein–protein interactions[34]) or were associated with pili assembly (Extended Data Fig. 4b–d and Supplementary Data 1). Proteins potentially involved in pili or type II secretion included Amuc_1100 and Amuc_1102, which have been characterized for their antiobesity activity and whose structures have been solved[35,36]. Overall, these findings suggest that mucin utilization by *A. muciniphila* requires genes encoding pili-like proteins and several poorly characterized hypothetical proteins whose expression is not regulated by mucin.

Mucin metabolism enables *Akkermansia* to compete in the gut

To identify *A. muciniphila* genes required to colonize the gastrointestinal tract, we performed an INSeq analysis of the Tn arrayed mutant pool in four mouse models: germ-free mice; 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[37]. Mice raised with a conventional microbiota had to be pretreated with tetracycline to displace their endogenous *Akkermansia*, which prevents the engraftment of the Muc^{T} *A. muciniphila* strain[38]. Mice were then orally gavaged with a pool of Tn mutants. The relative fitness of the Tn mutants in the caecum or after growth in mucin medium was assessed by INSeq (Fig. 2a).

While growth in mucin medium and mice is not directly comparable due to contributions from the host environment, diet and the microbiota, our analysis indicated that approximately half of the genes whose

disruption led to a greater than twofold decrease in abundance in the gastrointestinal tract of mice were also required for growth in mucin (Fig. 2a). As the complexity of the microbiota increased from germ-free to CONV mice, additional genes became conditionally essential (Fig. 2a–e). As we observed in mucin medium, Tn insertions in genes required for amino acid homeostasis were significantly overrepresented among mutants that failed to colonize the gastrointestinal tract (Fig. 2b,c). In addition, mutations in genes encoding proteins defective for Ala, Asp, Glu and Arg biosynthesis were depleted in ASF and CONV mice (Fig. 2c–e). Mutations in genes encoding proteins involved in the glyoxylate pathway—particularly components of the Gly cleavage system—also resulted in significant growth defects. The Gly cleavage system generates 5,10-methylene-tetrahydrofolate, a one carbon donor used in Ser biosynthesis, which may be particularly important for the colonization of *Muc2*$^{-/-}$ mice because they lack this Ser-rich mucin. Overall, these findings indicate that the synthesis of key amino acids, especially in the context of a complex microbiota, is required for *A. muciniphila* to successfully colonize the gastrointestinal tract.

Metabolic and cellular pathways overrepresented in *A. muciniphila* mutants with fitness defects in mice indicated that Tn insertions in genes encoding proteins involved in central processes such as transcription, protein metabolism and RNA metabolism led to similar growth defects in the gastrointestinal tract as in mucin medium (Fig. 2c). In contrast, several genes were required for survival in the gastrointestinal tract, but not for growth in mucin medium. These included genes encoding proteins involved in putative exopolysaccharide and capsule biosynthesis and a locus harbouring glycosyl transferases (Fig. 2f), which are orthologous to the *Bacillus* biofilm genes *epsH* and *epsJ*[39]. 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 germ-free mice but rescued in mice with a microbiota (Fig. 2d). Three of these genes encoded 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. Mutations in a fourth gene in the ASR pathway, *Amuc_1295*, led to strong growth defects under all conditions. The growth defects of these ASR^{-} mutants were partially complemented in germ-free mice, indicating that *A. muciniphila* can harvest hydrogen sulfide or cysteine from the host or its diet. In both ASF and CONV colonized mice, mutants with Tn insertions in ASR genes had a slight growth advantage (Fig. 2g), but this was context dependent as the 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 germ-free 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 gastrointestinal tract, with the greatest growth defects occurring in genes encoding proteins with predicted *N*-methyl domains (a hallmark of pili proteins) or tetratricopeptide repeat domains (Extended Data Fig. 4c,d).

Two genetic loci define a mucin transport system

We identified a cluster of Tn mutations in and around *Amuc_0544* causing strong mucin growth defects. This gene encodes a predicted secreted protein of unknown function found only in *Akkermansia* and other *Verrucomicrobia* (Fig. 3a). Amuc_0544 has multiple tetratricopeptide repeat domains, reminiscent of the polysaccharide transport protein SusD in the *Bacteroides* starch utilization system[40]. *Amuc_0543* and *Amuc_0544* are part of a larger gene cluster that includes two genes annotated as encoding putative ExbD biopolymer transporters, as well as an ExbB-like proton channel. ExbBD proteins are components of TonB-dependent transporters, which are also found in starch utilization systems; they transduce the proton motive force to transport substrates across the outer membrane[41]. Reads mapped by RNA-seq indicate that the region is probably expressed as at least two messenger RNA (mRNA) transcripts (Extended Data Fig. 5a). 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. 3b–e and Extended Data Fig. 5b), leading us to name the genomic region spanning *Amuc_0543* to *Amuc_0550* 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[1,26]. These proteins are among the most abundant produced by *A. muciniphila*[26] and include Amuc_1100, a Toll-like receptor 2 agonist[1,26] and modulator of host metabolism[1]. The structure of Amuc_1102 (ref. 35) indicates that it is related to archaeal type IV pili. Amuc_1101 is annotated as the cell division protein FtsA, but it belongs to the PilM family as categorized by Pfam, suggesting a more likely role in pili biogenesis. We renamed this locus, spanning *amuc_1098* to *amuc_1102*, *MUL2* (Fig. 3a and Extended Data Fig. 5a), given that Tn insertions in *MUL2* genes exhibited strong growth defects in mucin medium (Fig. 3c, Extended Data Fig. 5b and Supplementary Data 1). Additional potential *mul* loci are scattered throughout the genome (Fig. 2a), including in genes encoding additional potential pili proteins with *N*-methylphenylalanine domains typical of type IV pilins, further suggesting a broader role for pili in mucin acquisition (Extended Data Fig. 4b–d).

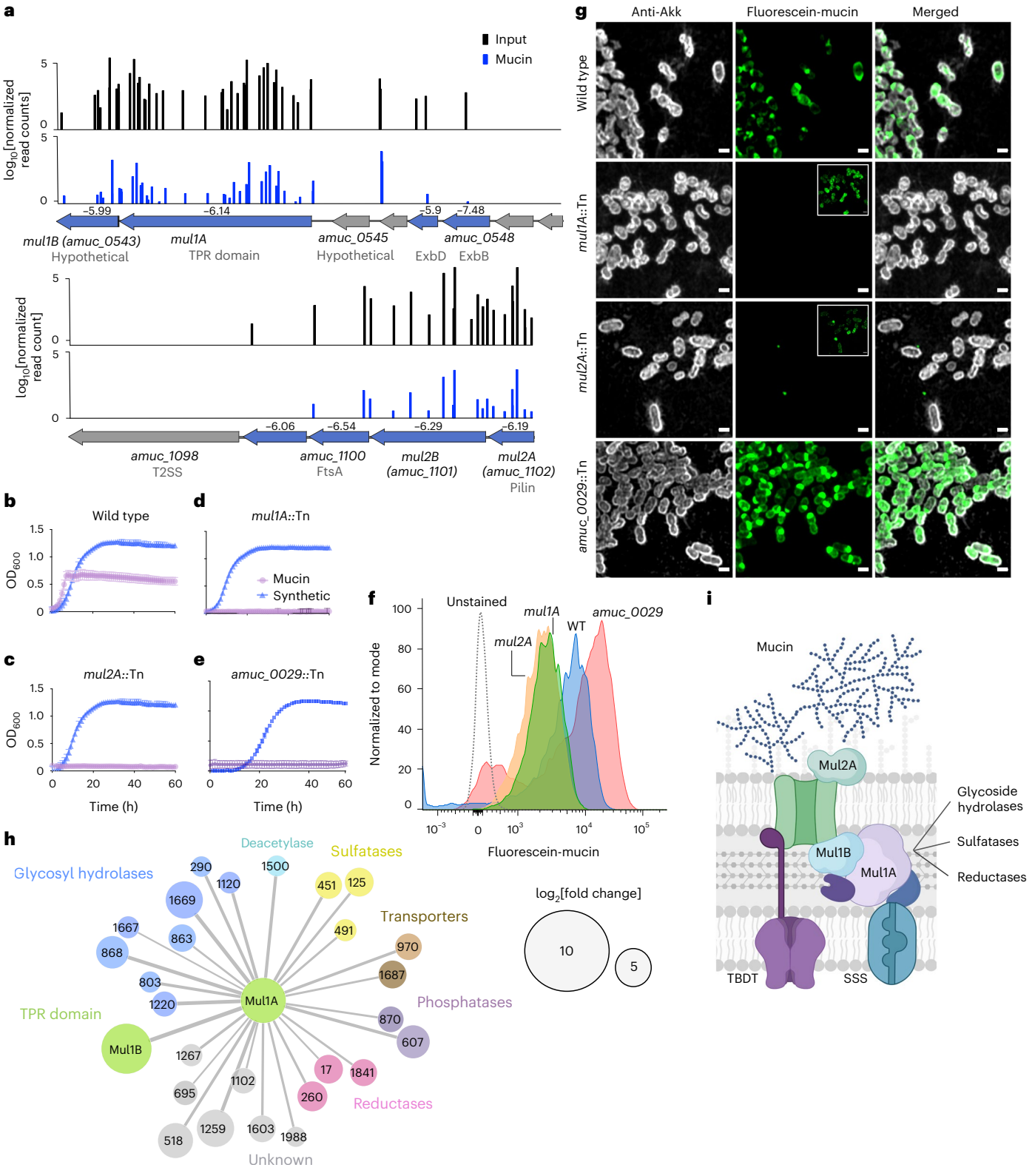
The largest gene in the *MUL1* locus, *mul1A* (*Amuc_0544*), is predicted to encode a protein with a type I signal peptide. Given its relatively large size and multiple predicted tetratricopeptide repeat domains, we hypothesized that Mul1A serves a structural role in the assembly of protein complexes important for the transport of mucin. We generated antisera specific for Mul1A and performed native

Fig. 3 | MULs encode a mucin transport complex in *A. muciniphila*. **a**, Map of the MULs *mul1* (top; *amuc_0543* to *amuc_0550*) and *mul2* (bottom; *amuc_1098* to *amuc_1102*). The 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. The bars represent the mean normalized reads for each Tn insert in the input pool and the output after growth in mucin. **b**–**g**, Characterization of *A. muciniphila* mutants with transposon insertions in *mul* genes. Mutants defective for mucin utilization display different patterns of association with fluorescein-labelled mucin. **b**–**e**, Growth curves of the wild type (**b**) and *mul2A*::Tn (**c**), *mul1A*::Tn (**d**) and *amuc_0029*::Tn (**e**) grown in synthetic or mucin medium, where mucin was the sole carbon and nitrogen source. **f**, Flow cytometric analysis of fluorescein-labelled mucin acquisition by *mul* transposon mutants. The mean fluorescein intensity was quantified for bacteria detected with anti-*Akkermansia* antibodies. WT, wild type. **g**, Mutants lacking *mul1A* and *mul2A*, but not a galactose epimerase (*amuc_0029*::Tn), failed to accumulate mucin or mucin degradation products in intracellular compartments. Cultures were grown in synthetic medium supplemented with fluorescein-labelled mucin before staining with anti-*Akkermansia* antibodies (anti-Akk). The insets show the corresponding image with enhanced brightness to visualize fluorescein-labelled mucin in *mul1A* and *mul2A mutants*. The images are representative of three independent experiments. Scale bars, 1 µm. **h**, Multiple proteins associate with the tetratricopeptide repeat domain protein Mul1A, as determined by LC/MS/MS analysis of co-immunoprecipitation. The numbers refer to the gene ID (*amuc_*). The node size reflects the $\log_2$[fold change] in normalized spectral counts over immunoprecipitations performed with *mul1A* mutants. The line thickness is scaled to the $-\log_{10}$[*P* value]. The nodes are colour coded based on Pfam association. Statistical analysis of the peptides used two-tailed heteroscedastic *t*-test on $\log_2$-transformed data. **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, glycoside hydrolases and potential inner and outer membrane transporters (SSS and TonB-dependent transporter (TBDT)). Panel **i** created with BioRender.com.

https://doi.org/10.1038/s41564-023-01407-w

immunoprecipitations from lysates of wild-type *A. muciniphila* or *mul1A*::Tn mutants, coupled to quantitative mass spectrometry (Extended Data Fig. 5c). Mul1A showed cleavage of the amino-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 and Supplementary Data 1), 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 may play a key role in the acquisition of mucin (Fig. 3h). Similar observations have been made in the related bacteria *Planctomycetes limnophila* and *Gemmata obscuriglobus*, which bind dextran glycans with pili-like fibres[42]. These findings also suggest that components of the *MUL1* and *MUL2* loci may cooperate to transport mucin fragments. Next, we tested whether these mutants would



https://doi.org/10.1038/s41564-023-01407-w

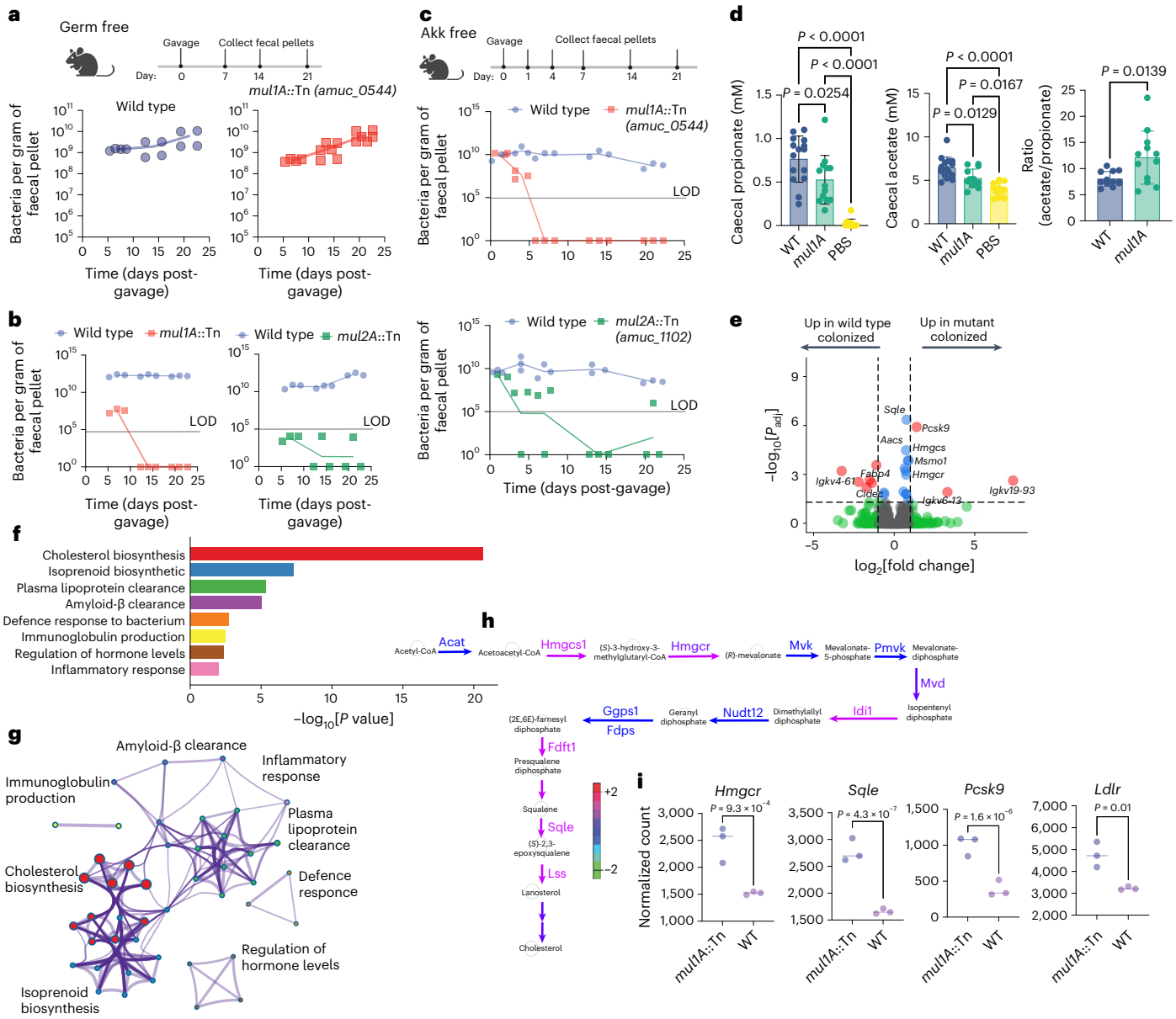


Fig. 4 | Mucin utilization enables *A. muciniphila* to compete against members of the microbiota and leads to repression of genes in cholesterol biosynthesis. a–**c**, Mucin utilization gives *A. muciniphila* a competitive advantage. **a**,**b**, In germ-free mice, growth on mucin is not required for mono-colonization (**a**), but *mul1A*::Tn and *mul2A*::Tn are outcompeted by wild-type *A. muciniphila* (**b**). **c**, Similarly, *mul1A*::Tn and *mul2A*::Tn were outcompeted in Akk-free mice. Each point represents the average count of *A. muciniphila* per gram of faeces per cage ($n = 5$ (germ free) or 6 (Akk free)). **d**, SCFA content of caecal contents from germ-free mice colonized with wild-type *A. muciniphila* ($n = 15$), *mul1A*::Tn ($n = 13$) or vehicle controls (PBS; $n = 12$). Acetate and propionate were analysed by one-way analysis of variance (F statistic). The ratio of propionate to acetate was analysed by two-tailed Student's t-test. The data are presented as means ± s.d. **e**–**h**, Comparison of the transcriptional profiles of colonic tissue from germ-free mice colonized with *mul1A*::Tn or wild-type *A. muciniphila*. **e**, RNA-seq of colon tissue. Each point represents a gene. The colours indicate a $\log_2$[fold change] > 1 (green; outside the two vertical dashed lines), a Benjamini–Hochberg-adjusted P value (P_{adj}) < 0.05 (blue; above the horizontal dashed line) or both (red). Significance was determined using the Wald test. **f**,**g**, MetaScape enrichment plot (**f**) and network visualization (**g**) showing pathways enriched in mice colonized with *mul1A*::Tn. The node colours correspond to cluster annotation and the line thickness denotes the relatedness of the pathways. Benjamini–Hochberg-adjusted P values were determined in MetaScape with a hypergeometric test. **h**, Differential expression of genes along the mevalonate and cholesterol biosynthesis super pathways, as visualized with BioCyc. The enzyme names are colour coded to indicate the relative expression levels of mice colonized with *mul1A*::Tn versus wild-type *A. muciniphila* (see coloured scale bar). **i**, Expression of selected genes in response to *mul1A*::Tn (Wald test with Benjamini–Hochberg-adjusted P values). LOD, limit of detection. Mouse icons in **a** and **c** created with BioRender.com.

be impaired for the import of mucins across the outer membrane. Indeed, Tn mutants with mutations in either *mul1A or mul1B* were unable to internalize fluorescein-labelled mucin (Fig. 3f,g). The loss of fluorescein-labelled mucin labelling in the mutants is not an indirect consequence of their inability to metabolize mucins as a Tn mutant with mutated *amuc_0029*, encoding a uridine diphosphate-glucose 4-epimerase in the Leloir pathway for galactose degradation[43], could not grow in mucin medium (Fig. 3e and Supplementary Data 1), yet displayed high intracellular staining with fluorescein-labelled mucin (Fig. 3f,g).

Mul1A interacted with several glycoside hydrolase enzymes, two sulfatases and two disulfide reductases. We also detected peptides from a putative sodium solute symporter (SSS) transporter, Amuc_0970, which in other microbes can import galactose[44] 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 transporters in other polysaccharide utilization systems. 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 and Supplementary Data 1), although the precise mechanism remains to be determined. Comparison of the mass spectrometry data with the transcriptional analysis in response to mucin indicated that core components of the MUL system are constitutively expressed, while MUL accessory proteins such as glycoside hydrolases are inducible (Figs. 1e and 3h). Finally, we detected peptides belonging to the von Willebrand type D domain and the carboxy-terminal cysteine knot domain of Muc5AC specifically co-precipitating with anti-Mul1A antisera (Extended Data Fig. 5d), indicating that Mul1A or associated proteins can engage the non-glycosylated portion of the gastric mucin backbone. On the basis of these observations, we propose that the MUL system functions as a complex with associated pili (Fig. 3i), perhaps in conjunction with additional *mul* loci not yet identified, to capture extracellular mucins or mucin fragments and import them into the cell, where they ultimately accumulate in mucinosomes.

Mucin utilization modulates host transcriptional responses

We tested the requirement for the MUL system to colonize the gastrointestinal tract in germ-free mice. Unexpectedly, when administered alone, *mul1A*::Tn colonized germ-free mice as efficiently as wild-type *A. muciniphila* (Fig. 4a), although wild-type *A. muciniphila* readily outcompeted *mul1A*::Tn and *mul2A*::Tn when the strains were co-administered (Fig. 4b). Next, we tested the ability of *mul* mutants to colonize the gastrointestinal tract in the presence of an intact microbiota by using a mouse colony that was naturally devoid of *Akkermansia* (Akk free). These mice have an intestinal microbiota like that of conventional mice (Extended Data Fig. 6a–d) and do not require pretreatment with tetracycline to enable the engraftment of *Akkermansia*. Both mutants exhibited more pronounced defects in their ability to colonize mice with complex microbiotas (Fig. 4c). Similar results were obtained in antibiotic-pretreated CONV mice (Extended Data Fig. 6e). We also performed competitions between wild-type *A. muciniphila* and *amuc_0029*::Tn, which binds to and acquires mucin, but cannot grow in it (Fig. 3e–g). Like the *mul* mutants, *amuc_0029*::Tn was rapidly outcompeted by wild-type *A. muciniphila* (Extended Data Fig. 6e), suggesting that the colonization defects of mutants lacking functional *mul* genes are not simply because they cannot bind to mucins in the gastrointestinal tract. Taken together, these findings indicate that the MUL system and mucin metabolism are required for *A. muciniphila* to establish residence in the mouse gastrointestinal tract, but only in the context of other microbes.

Colonization by *mul1A*::Tn in germ-free mice indicates that *A. muciniphila* can access other nutrient sources and is consistent with observations that *Muc2*$^{-/-}$ mice are colonized by an endogenous *Akkermansia* strain. We found that both the wild type and mutants lacking *mul1A* or *mul1B* colonized the gastrointestinal tract of tetracycline-treated *Muc2*$^{-/-}$ mice (Extended Data Fig. 6f). Thus, the requirement for mucin to colonize the gastrointestinal tract is context dependent and *Akkermansia* can use other nutrients from the diet, host or microbiota. However, the mutants were still outcompeted by wild-type *A. muciniphila* in *Muc2*$^{-/-}$ mice (Extended Data Fig. 6g), suggesting that the MUL system is required to scavenge other related glycoproteins.

We reasoned that *mul1A*::Tn would produce different fermentation products compared with wild-type *A. muciniphila* since they produce varying ratios of acetate and propionate in response to different carbon sources[28]. We found that germ-free animals colonized with *mul1A*::Tn or wild-type *A. muciniphila* differed in their caecal short-chain fatty acid (SCFA) levels. Germ-free mice colonized with wild-type *A. muciniphila* had higher levels of total acetate and propionate compared with mice colonized with *mul1A*::Tn (Fig. 4d), but a lower ratio of acetate to propionate (Fig. 4d). It is unknown whether these differences reflect SCFA production by *A. muciniphila* or adsorption by the host. 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.

Differential expression of host genes in response to various components of *A. muciniphila*, including live and pasteurized cultures[45], purified surface proteins[1] and conditioned supernatants[46], has been reported. Because the wild type and *mul1A*::Tn colonized the gastrointestinal tract of female germ-free mice at similar levels (Extended Data Fig. 7a), 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 germ-free mice revealed differentially expressed genes in the category of immunoglobulin genes, indicating a generalized defence response to bacteria (Fig. 4e–g). However, the most statistically significant difference in the colonic transcriptional response between germ-free mice colonized with wild-type *A. muciniphila* and *mul1A*::Tn occurred in female mice and was in pathways associated with lipid metabolism (Fig. 4e–g and Supplementary Data 1). The expression of the cholesterol biosynthetic genes *Sqle*, *Hmgcs* and *Hmgcr*, as well as most genes along the cholesterol/mevalonate synthesis pathway, was repressed in mice colonized with wild-type *A. muciniphila*, as was the expression of genes encoding proteins that modulate cholesterol uptake, including *Ldlr* and *Pcsk9* (Fig. 4d–i and Extended Data Fig. 7b,c). Single-cell RNA-seq datasets suggest that cholesterol/mevalonate biosynthetic genes are expressed in colonic intestinal epithelial cells and goblet cells[47] (Extended Data Fig. 7d), where cholesterol biosynthesis contributes to stem cell proliferation[48] and is required to maintain the integrity of the intestinal epithelium[49]. Colonic epithelial cells are the most likely to respond directly to *A. muciniphila* and its metabolites. These observations are consistent with reports that *A. muciniphila* metabolites, including SCFAs, repress lipid metabolism in intestinal organoids[46] and that treatment with *A. muciniphila* lowers serum cholesterol in mice[1,45] and humans[50]. While germ-free mice showed intermediate levels of expression of sterol biosynthesis genes (Extended Data Fig. 7b), these pathways were significantly elevated in germ-free mice mono-colonized with *mul1A*::Tn (Fig. 4d–i and Extended Data Fig. 7b). This observation suggests that the colon's transcriptional response to *A. muciniphila* is directly influenced by the bacterium's ability to take up or metabolize mucins and related glycoproteins.

Discussion

A. muciniphila is a prevalent member of the intestinal microbiota and a potential next-generation probiotic. However, we have an incomplete understanding of *A. muciniphila* biology and how it interacts with its host due to a lack of genetic tools and because *Akkermansia* proteins share limited homology with other prominent gut microbes. Here we established methods for Tn mutagenesis and applied INSeq[51] to identify *A. muciniphila* genes required for mucin utilization and colonization. This led to the discovery of genes encoding hypothetical proteins, (that is, MULs) that were essential for growth on mucin. Unlike the substrate-inducible polysaccharide utilization systems that are common among *Bacteroides* and other intestinal microbes[52], many of the *mul* genes were constitutively expressed, possibly reflecting *A. muciniphila*'s specialized niche using mucin as its preferred nutrient source. The MUL complex was required for active transport of mucin across the outer membrane and appears to be a specialized system tailored to mucin acquisition, and possibly other glycans, in *A. muciniphila*.

Our findings indicate that *A. muciniphila* can access multiple nutrient sources in the gastrointestinal tract, yet the ability to use mucin via the MUL system provides a competitive advantage against other members of the intestinal microbiota. Furthermore, we were able to determine that *A. muciniphila* mucin consumption repressed the expression of cholesterol biosynthesis genes in the colons of germ-free mice. Therefore, in addition to the beneficial immunomodulatory activities that have been assigned to cellular components of *A. muciniphila*[1,5], the active catabolism of mucin by *Akkermansia* may provide additional health benefits by regulating the expression of genes involved in lipid biosynthesis.

Methods

Animal studies

Animal care and study protocols were approved by Duke University's Institutional Animal Care and Use Committee. All experiments with mice were performed under protocol A118-20-05.

Media and strains

Bacteria were grown in an anaerobic chamber (Coy Laboratory) with the following gaseous characteristics: 5% hydrogen, 5% carbon dioxide and 90% nitrogen. *A. muciniphila* was grown in mucin medium based on previous work[21] (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 $CaCl_2$, 40 mM HCl, trace elements and vitamins 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, 16 g l^{-1} soy peptone (Amresco) or phytone (BD Biosciences) and 4 g l^{-1} threonine) and brain heart infusion medium (BHI; BD Biosciences) supplemented with 2% tryptone and 1× haemin and vitamin K (Hardy Diagnostics). *A. muciniphila* BAA-835 (ref. 21), *Bacteroides vulgatus* and *B. thetaiotaomicron* were obtained from the American Type Culture Collection, and additional strains, including *Akkermansia glycaniphila*, *Bifidobacterium longum*, *Bifidobacterium bifidum*, *Ruminococcus gnavus*, *Ruminococcus torques* and *Peptostreptococcus russellii*, were obtained from the German Collection of Microorganisms and Cell Cultures. *E. coli* S17 harbouring the pSAM_Bt plasmid was provided by A. Goodman at Yale University. When required, antibiotics were used at the following concentrations for *Akkermansia*: 7 mg ml^{-1} chloramphenicol, 10 mg ml^{-1} gentamicin and 12 mg ml^{-1} kanamycin. *E. coli* was cultured in LB medium and antibiotics were added as required at the following concentrations: 100 mg ml^{-1} ampicillin, 35 mg ml^{-1} chloramphenicol and 30 µg ml^{-1} kanamycin.

Labelling of mucin glycans

Porcine gastric mucin was labelled with 6-aminofluorescein by adapting a previously described protocol to label polysaccharides[53]. Before labelling, commercial pepsin-digested gastric mucin (Type III; Sigma–Aldrich) was filtered through a 0.45 µM filter and transferred to a 5 ml tube. In a fume hood, 1 ml filtered mucin was combined with 30 mg CNBr diluted in 350 µl 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 a pH of >10 for the duration of the reaction. Following activation, excess CNBr was removed using a Bio-Rad 10DG desalting column equilibrated with 0.2 M 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 dialysed against 16 l 20 mM sodium phosphate buffer (pH 8) using an 8 kDa molecular weight cut-off membrane and quantified using the phenol sulfuric acid method.

Microscopy

To test *Akkermansia*'s interaction with labelled glycans by microscopy, cultures were grown overnight to saturation and subcultured 1:1 in fresh mucin medium for 5 h. The actively growing cultures (600 ml) were combined with 1.3 ml fresh synthetic media and either 20 mg carbohydrate per ml fluorescein-labelled mucin or 20 mg ml^{-1} fixable fluoro-emerald (fluorescein) dextran (10,000 MW; D1820; Thermo Fisher Scientific). For experiments involving CCCP, the cultures were pretreated with 50 µM CCCP for 30 min before the addition on fluorescein-labelled mucin. Cultures were then incubated at 37 °C for 3 h and subsequently harvested by centrifugation at 14,000*g* for 5 min, washed once with phosphate-buffered saline (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 and 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^{-1}) for 2 min, washed with PBS and blocked with 2% bovine serum albumin 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 labelled secondary antibodies (goat anti-rabbit IgG Alexa Fluor 647; A-21244; Invitrogen). The cells were imaged on an inverted confocal laser scanning microscope (LSM 880; Zeiss) equipped with a motorized stage, a diode (405 or 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-Apochromat 63×/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 optical density measured at a wavelength of 600 nm (OD_{600}) of 0.4–0.5 in chopped meat media supplemented with haemin and vitamin K. 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 fluorescein-labelled mucin (final concentration: 20 µg ml^{-1} carbohydrate). Labelled mucin accounted for approximately 0.8% of the total carbohydrate content of the media. The cultures were incubated for 3 h at 37 °C. Membranes were stained with 10 µg ml^{-1} fixable FM 4-64FX (F34653; Invitrogen) and the cells were fixed with 4% formaldehyde and prepared for microscopy as described above.

For super-resolution microscopy, *A. muciniphila* was grown with fluorescein-labelled mucin and prepared for imaging as described above, except that lysozyme was omitted and the cells were stained with anti-*Akkermansia* antisera (1:100 dilution) followed by STED-compatible fluorescently labelled secondary antibodies (goat anti-rabbit IgG Alexa Fluor 594; A32740; Invitrogen). Super-resolution microscopy was carried out using a Leica TCS SP8 STED microscope equipped with a Leica DMi8 inverted motorized stage, a pulsed white light laser, high-sensitivity GaAsP HyD detectors and STED depletion lasers (592, 660 and 775 nm). For each image, 40 stacks in the *Z* plane were acquired with a 100×/1.4 HCX PL APO OIL DIC WD 90 µm objective (Leica) and LAS X software (Leica), followed by deconvolution with Huygens Professional software (Scientific Volume Imaging) and three-dimensional (3D) rendering with Imaris (version 9.5; 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 ml molten agarose onto glass-bottom dishes. The pads were solidified and equilibrated for 2 h under anaerobic conditions before 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 culture were then harvested by centrifugation at 10,000*g* for

3 min and resuspended in 50 µl synthetic media supplemented with 15 ml fluorescein-labelled 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, a light-emitting diode 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 s for 30 min using a Plan Apochromat 100×/1.4 NA oil objective (Nikon) with 1.5× zoom in the NIS-Elements software. The images (resolution: 43.33 nm per pixel) were deconvolved in NIS-Elements using the automatic 2D deconvolution function and bleach corrected with histogram matching in ImageJ (National Institutes of Health), then converted to eight-bit TIFFs and reconstructed in Adobe Photoshop where only linear adjustments were made to the fluorescence intensity. Annotations were added using ImageJ.

Kinetics of mucin uptake

A. muciniphila was grown in mucin medium to an OD_{600} of 0.5 and mixed 1:1 with fresh synthetic media and 96 mg ml^{-1} fluorescein-labelled mucin. Cultures were incubated anaerobically at 37 °C. At each time point, the OD_{600} was measured, 1.5 ml culture was collected and the bacteria were harvested by centrifugation at 14,000*g* for 3 min. The resulting cell pellets were fixed with 4% formaldehyde and used in flow cytometry as described below.

To test CCCP inhibition of mucin transport, cultures were treated with 50 µM CCCP for 15 min before the addition of mucin. As a control for active transport, the cell-permeant esterase CFDA (C1354; Thermo Fisher Scientific) was used in place of mucin. CFDA stains and killed controls were prepared as described previously[54]. CFDA fluoresces at 488 nm in viable cells.

Flow cytometry

Flow cytometry was used to measure the fluorescence intensity of bacteria grown in the presence of fluorescein-labelled mucin. Cultures were grown in mucin medium overnight, subcultured 1:1 into fresh mucin media and grown for an additional 5 h. The actively growing cultures (600 ml) were combined with 1.3 ml synthetic media and 20 mg carbohydrate per ml fluorescein-labelled 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% bovine serum albumin (1:100 dilution) overnight at 4 °C and washed with PBS, followed by incubation with anti-rabbit IgG Alexa 647 antibodies (1:1,000; A-21244; Invitrogen). Cells were analysed on a BD FACSCanto II flow cytometer and 50,000 events were collected for each sample. Flow cytometry data were acquired using FACSDiva software (version 9; BD Biosciences) and additional analysis was run using FlowJo software (versions 9 and 10). The samples were gated on the population of single cells that stained positive with anti-*Akkermansia*/Alexa Fluor 647, and the mean intensity of fluorescein-labelled mucin staining was recorded.

Growth curve assays

The growth kinetics of wild-type *A. muciniphila* and selected Tn mutants were measured in both mucin and synthetic media. Starter cultures were prepared by growing bacteria in synthetic medium, supplemented with chloramphenicol (7 µg ml^{-1}) for mutants, 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 ml 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 described above, except that the starter cultures were prepared in BHI (BD Biosciences) with 1× haemin and vitamin K and 2% tryptone. The BHI media for *A. muciniphila* was also supplemented with 0.25% mucin. Growth curves were assessed in mucin medium[21] supplemented with 1× haemin and vitamin K, where mucin was the sole source of carbon and nitrogen, or in BHI with haemin, vitamin K and tryptone. The results reflect two biological replicates per strain.

Generation of anti-*Akkermansia* antisera

Polyclonal anti-serum 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 24 h. 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 a 450 ml volume was used for subcutaneous inoculation, followed by three additional injections using Freund's incomplete adjuvant.

Generation of anti-Amuc_0544 antisera

Polyclonal antisera were raised against a 47 kDa fragment of Amuc_0544 protein spanning the TPR domains (from 1,216–2,457 base pairs (bp)). The fragment was amplified with primers 127 and 129 (Supplementary Data 1) using Phusion Polymerase and 5× HF Buffer (Thermo Fisher Scientific) and cleaned using 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 (R3101 and R3104, respectively; NEB), gel extracted using a QIAquick Gel Extraction kit and ligated overnight with T4 DNA Ligase (M0202; NEB). The ligated DNA was cloned into competent *E. coli* XL1-blue (Agilent) and, after the construct was confirmed by sequencing, the final plasmid was transformed into *E. coli* BL21(DE3) (C2527I; NEB) for protein expression.

The Amuc_0544 fragment was produced as a soluble protein with an amino-terminal hexahistidine tag. *E. coli* BL21(DE3) was grown in LB medium with 30 µg ml^{-1} kanamycin to the mid-exponential phase and induced with 1 mM isopropyl-β-D-thiogalactopyranoside for 3 h. The cells were harvested by centrifugation (at 5,000*g* for 10 min) and the pellet was suspended in buffer (50 mM sodium phosphate, 300 mM NaCl and 20 mM imidazole (pH 7.4)) and lysed by sonication with 30 cycles of 30-s intervals alternating with 30 s on ice (Branson Sonifier 450). Cell debris were removed by centrifugation (15,000*g* for 45 min at 4 °C) and the supernatant was applied to an Ni-NTA agarose bead column (Prometheus Protein Biology Products). The column was washed with 25 ml buffer containing 40 mM imidazole and the protein was eluted in 1 ml fractions with 300 mM imidazole. Antibodies were raised in rabbits as described above using 2.5 mg 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 0.35–0.50). Cells were harvested by centrifugation at 10,000*g* and 4 °C for 5 min. RNA was isolated using a Qiagen RNeasy kit following the enzymatic digestion protocol (proteinase K plus lysozyme). The resulting RNA was concentrated by precipitation with two volumes of isopropyl alcohol and 0.3 M sodium acetate (pH 5). Ribosomal RNA (rRNA) was depleted using a RiboMinus Bacteria kit (Thermo Fisher Scientific). The 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 (E7760S; NEB) and NEBNext Multiplex Oligos for Illumina. Libraries were sequenced on a HiSeq 4000 in a 50 bp single-read run. Following sequencing, adaptor sequences were trimmed and filtered using the program fastq-mcf[55]. The reads were

then mapped to the *A. muciniphila* genome (accession CP001071.1) using STAR[56] with the parameter --alignIntronMax set to 1. Differential expression analysis was carried out with DESeq2 (ref. 57) and pathway analysis was run using clusterProfiler[58].

pSAM_Akk plasmid construction

Transposon mutagenesis was carried out with a modified version of pSAM_Bt[19] 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 were amplified from pRL1342 (ref. 59) with the primers 80 and 81 (Supplementary Data 1) 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 and their encoded proteins (Amuc_R0036, RpoD, SecA, EF-Tu, DNA pol I, DNA pol III alpha, FtsZ, FtsY, YidC, GAPDH, GroEL and enolase) and the tools provided in the Sequence Manipulation Suite[60]. 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} of 1 and harvested by centrifugation at 15,000*g* for 5 min. The *E. coli* donor was prepared by growing in LB medium (BP1426-2; Fisher Bioreagents) with ampicillin (100 μg ml^{-1}). Overnight cultures were subcultured 1:50 and grown to an OD_{600} of 0.4–0.6. Cells were harvested by centrifugation at 1,500*g* and washed with LB media to remove residual antibiotics. The pellets were combined to create a slurry of cells that were distributed in 100 μl puddles on pre-reduced synthetic media plates. Conjugations were incubated aerobically at 37 °C for 14–16 h. 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^{-1}) and kanamycin (12 μg ml^{-1}) to counter select against the *E. coli* donor. Cultures were grown anaerobically for an additional 48 h to allow *Akkermansia* to recover, then subsequently subcultured 1:10 into fresh media twice more at 24 h intervals. After the final subculture, 150 μl of the suspension was plated on synthetic media containing gentamicin, kanamycin and chloramphenicol (10, 12 and 7 μg ml^{-1}, respectively). Colonies appeared after 4–6 d of 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 and 12), *Akkermansia*-specific 16S (primers 75 and 76) and the *cat* gene located on the Tn (primers 80 and 81). Once transposition was confirmed, arbitrary PCR was used for small-scale mapping of Tn insert locations[61] using the following sequential PCR settings. Reaction 1 was run with primers 22 and 88 (94 °C for 5 min, then six cycles of 94 °C for 30 s, 30 °C for 30 s and 72 °C for 1 min; then 30 cycles of 94 °C for 30 s, 45 °C for 30 s and 72 °C for 1 min; and finally 72 °C for an additional 5 min). Reaction 2 was run using 2 μl reaction 1 as a template and primers 23 and 89 (95 °C for 1 min; then 30 cycles of 95 °C for 30 s, 45 °C for 30 s and 72 °C for 2 min; then 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 analysed by Southern blot. DIG-labelled probes were generated to detect the *cat* gene and the blot was run using a DIG-High Prime DNA Labeling and Detection Starter Kit II (Sigma–Aldrich). Southern blots, as well as DNA gels, were imaged with a LI-COR Odyssey Imager and images were labelled using Adobe Photoshop version 24.1.1.

Construction of Tn libraries

Tn libraries were prepared in two formats for INSeq analysis: (1) as a pooled Tn library; and (2) as an arrayed 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 the in vitro INSeq experiments. The second approach used a smaller but well-defined library constructed with equal inputs of known mutants from a library arrayed in 96-well plates. This library was used for all of the 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, as described below, to facilitate the retrieval of mutants of interest.

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-well deep-well plates. The outgrowths were carried out as described above by culturing conjugation mixes for an initial 48 h followed by two additional subculture steps at 24 h 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 d. 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 (2,680 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 the vitro INSeq experiments.

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 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 −80 °C. 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 Blood 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 comprising 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 the 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, the 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 the 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 and assigned a barcode, then the Tn insertion sites were sequenced as previously described[51] This approach provided a 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 artefacts. As the signal-to-noise ratio varied between individual samples and sequencing runs, a MATLAB script (INseq_read_filter_v3.m) was used to view the data and set 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 (a mean of 3.7 replicates of each clone), probably as a result of the extensive outgrowths that were required to effectively remove the *E. coli* donor strain before the 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 the location (binary code[19]) or makes that information ambiguous (Cartesian pooling and Sudoku[20]). Thus, we faced a trade-off between capturing the full diversity of Tn insertions in the collection and using 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.

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 was simply the plate with the most unique clones (lowest intra-plate redundancy). In subsequent steps, all remaining plates were tested one at a time in combination with the set of previously chosen plates (1&hillip;$i-1$) to find the plate (i) that creates a new set (1…i) that has the lowest overall redundancy.

Evaluating the trade-off between coverage and redundancy when choosing a subset of plates suitable for Cartesian mapping. Simulation data for each optimized set of plates were 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 (Extended Data Fig. 2e). The plate set that yielded the largest diversity for redundancy was found where the distance between these two curves was greatest (set size 16: 873 unique Tn insertions; mean redundancy = 1.76).

The degree of redundancy per clone changed with increasing plate set size in the simulated data (Extended Data Fig. 2f). 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–column 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–column 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–column 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 1,277 Tn insertions in the entire collection and 86% of disrupted genes (insertion occurred with the 90th percentile of the coding sequence pos = 0.9) and reduces the mean redundancy from 3.7 to 1.9 copies per clone

INSeq analysis of conditionally essential genes in mucin medium

The mixed pooled mutant library was inoculated at 1:20 dilution in synthetic medium containing 10 μg ml^{-1} gentamycin, 12 μg ml^{-1} kanamycin and 5 μg ml^{-1} 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^{-1} threonine and 16 g l^{-1} 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 ($n = 4$), ASF ($n = 7$), CONV ($n = 11$) and *Muc2*$^{-/-}$ mice[37] ($n = 4$). Colonization of CONV and *Muc2*$^{-/-}$ mice required pretreatment with tetracycline to eliminate endogenous *Akkermansia* and to allow the mutant library to engraft (3 g l^{-1} tetracycline suspended in distilled water with 1% sucrose for 2 weeks). DNA was then extracted from faecal pellets using a Qiagen DNA Fast Stool Mini kit and tested for residual *Akkermansia* by PCR with *Akkermansia*-specific 16S primers (75 and 76). Once the *Akkermansia* was cleared, the antibiotics were withdrawn 48 h before the delivery of Tn mutant libraries by oral gavage. Germ-free 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*g* and 4 °C for 5 min and washed once with reduced PBS before being resuspended in one-tenth 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 d. Mice were euthanized and caecal 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 the

recovery of *A. muciniphila*. Caecal contents from ASF mice (~200 mg) were combined with 10 ml synthetic medium containing chloramphenicol, kanamycin and gentamicin and grown for eight 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 colony-forming unit (c.f.u.) assay. Given the OD_{600} of the starter culture and the plate c.f.u. counts, the c.f.u. per ml concentration could be estimated from an OD_{600} measurement before droplet loading[62]. 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 (3 M) 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 an LCD screen for the viewing of droplet generation on the microfluidic chip (Celestron). Following droplet encapsulation, cultures were grown at 37 °C under anaerobic conditions for 72 h. Quantitative PCR with *Akkermansia*-specific primers (75 and 76; Supplementary Data 1) was used to assess the cell abundance between droplet cultures. To retrieve cells, the droplet emulsions were broken with an equal volume of perfluorooctanol, vortexed and spun down briefly at 300*g*. The aqueous phase was placed in a new tube, cells were pelleted at 10,000*g* for 1 min, 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 by Goodman et al.[19], with minor modifications. Libraries prepared from in vitro samples used 2 µg input DNA, whereas libraries prepared from mice used 0.5 µg *A. muciniphila* input DNA as determined by quantitative PCR. Linear PCR was performed with Pfx polymerase (Invitrogen) and the primer Tru-BioSamA (Supplementary Data 1). 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 s and 68 °C for 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 Fisher Scientific). Washing, second strand synthesis, MmeI digestion and adaptor ligation were performed as described by Goodman et al.[19], except that 6-bp adaptor sequences[63] were used. Finally, the samples were amplified with Pfx polymerase using the buffer at 2× concentration, additional Mg^{2+} (4 mM final) and the primer pair TruP7-PCR_5 and TruP5-PCR_3. PCR was performed under the following conditions: 94 °C for 2 min, followed by 18 cycles of 94 °C for 15 s, 60 °C for 1 min and 68 °C for 2 min, followed by 68 °C for 4 min. The resulting PCR products were size selected on 2% E-Gel SizeSelect II Agarose Gels (Thermo Fisher Scientific). The DNA was then quantified using a Qubit dsDNA HS assay kit (Thermo Fisher Scientific) and sequenced on an Illumina HiSeq 4000 50-bp single-read flow cell, and samples from ASF and *Muc2*$^{-/-}$ mice were sequenced on a NovaSeq 6000 system 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 of the sequencing was performed at the Duke University Sequencing and Genomic Technologies Shared Resource.

INSeq mapping and quantification of Tn insertion junctions and normalization

Pre-processing. Raw sequencing data were 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, with minor modifications from Goodman et al.[19]) that identified sequence reads containing the mariner Tn and extracted the bounding genomic sequence, which were grouped and counted. Bowtie version 1.2.2 was used to map 16-bp reads to the *Akkermansia* Muc^T genome. Reads that failed to map to a unique genomic locus were discarded. Output read count tables used for subsequent analysis steps contained the Tn insertion coordinate and the number of instances a sequence read was mapped to this site.

The National Center for Biotechnology Information *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 to map and annotate 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 the read depth. Quality control metrics were used to establish criteria for sample inclusion in the dataset.

For quality metric 1 (depth), samples were excluded when >50% of the total sequencing data represented sequence reads not found in the deeply sequenced input sample.

For quality control metric 2 (noise), 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 $\log_{10}$[ratLR] as follows: (1) a psuedocount of 1 was added when either the L or R read count was 0; (2) for each insertion site, ratLR was calculated using the larger of L or R, so that all ratios of side bias were greater than 1; and (3) the sample-wide variance in $\log_{10}$[ratLR] was calculated. 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 datasets 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 and INseq_read_filter_v4.m) allow users to plot the data, 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 to filter 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 under different test conditions[64]. 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 the computational efficiency (Artist_setup.m or 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 the analysis of multiple samples (for example, biological replicates) to account for variation in the sequencing depth between samples. Normalization features in TRANSIT were disabled.

Analysis of Tn mutant populations

The data analysis package TRANSIT[65] was used to identify genes with altered relative abundance in input and output populations. Normalized counts were analysed 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* American Type Culture Collection BAA-835 complete genome (CP001071.1; GenBank, corrected as noted above). The output from TRANSIT was used for all downstream analyses and *P* values were adjusted by the Benjamini–Hochberg method.

Pathway analysis was carried out with the clusterProfiler[58] program enrichKEGG to analyse genes with Tn inserts that generated an absolute $\log_2$[fold change] > 5 in input versus output samples and an uncorrected *P* value < 0.05. Enriched pathways with a *q* value < 0.05 were considered significantly enriched. BioCyc[66] was used to analyse global pathways in the INSeq data. For this analysis, data were analysed 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 the Kyoto Encyclopedia of Genes and Genomes (KEGG). The KEGG pathway for ASR 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 INSeq data generated from growth in mucin medium using the same Tn arrayed library. Genes with a $\log_2$[fold change] > 5 in input versus 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. 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 in the INSeq datasets, the list of predicted glycoside hydrolase enzymes for *A. muciniphila* BAA-835/Muc^T was retrieved from the CAZy database[24]. Genes in additional pathways, including amino acid biosynthesis genes, were obtained from the KEGG. INSeq data were plotted using the R package ggPlot2 and specific genes of interest were highlighted using the R package gghighlight.

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

All of the mouse experiments were approved by Duke University's Institutional Animal Care and Use Committee.

In vivo colonization experiments were carried out in germ-free C57BL/6J mice obtained from Duke University's Microbiome Core Facility, C57BL/6J mice obtained from The Jackson Laboratory, *Muc2*$^{-/-}$ mice[37] (a kind gift from L. Augenlicht at the Albert Einstein College of Medicine) and *Akkermansia*-free mice bred in Duke University's Division of Laboratory Animal Resources Breeding Core Facility. To generate an *Akkermansia*-free mouse colony, 5-week-old conventional C57BL/6J mice obtained from The Jackson Laboratory were treated with tetracycline for 2 weeks to eliminate endogenous *A. muciniphila*. We then reconstituted the microbiota with a faecal slurry prepared using pellets from a previously described line of *Akkermansia*-free mice[67]. The mice were fasted overnight and gavaged with 200 µl faecal slurry (0.03 g ml^{-1}). These mice were then used to start a breeding colony that we have maintained in sterile cages since January 2020. The colony is routinely tested for contamination using *Akkermansia*-specific 16S rRNA primers[68] and all mice are tested for contamination immediately before their use in experiments. Akk-free mice were maintained on autoclaved water and autoclaved LabDiet 5K67, and CONV and *Muc2*$^{-/-}$ mice received reverse osmosis water and LabDiet 5353. Mice were between 8 and 12 weeks of age and were a mix of males and females, except for the CONV mice, which were all female. CONV and *Muc2*$^{-/-}$ mice were pretreated with 3 g l^{-1} tetracycline suspended in distilled water with 1% sucrose for 2 weeks. Following antibiotic treatment, clearance of residual endogenous mouse *Akkermansia* was confirmed by PCR using *Akkermansia*-specific 16S rRNA primers. Antibiotics were replaced with regular water for 2 d before gavage. No antibiotic treatment was used for the germ-free or *Akkermansia*-free mice.

Mice were colonized with *mul1A*::Tn (*amuc_0544*::Tn; insert at 644982 bp), *mul1B*::Tn (*amuc_0543*::Tn; insert at 642725 bp), *mul2A*::Tn (*amuc_1102*::Tn; insert at 1317866 bp) or a mutant with a Tn insertion immediately upstream of the *amuc0029* start codon (insert at 38253 bp). To prepare the inoculum for colonization, *A. muciniphila* cultures were grown for 36 h, concentrated by centrifugation and suspended in PBS containing 20% glycerol. Cultures were standardized by optical density and plate counting. 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 c.f.u. For the competition assay, wild-type *A. muciniphila* was mixed with the mutants at a 1:1 ratio and mice were gavaged with 150 µl of a mixture containing approximately 0.5×10^8 c.f.u. of each strain. To partially reconstitute the microbiota in antibiotic-treated mice, the gavage material for CONV mice included a faecal slurry prepared from faecal pellets from *Akkermansia*-free mice. The slurry was prepared under anaerobic conditions by homogenizing faecal pellets in sterile PBS (0.1 g ml^{-1}). Large particles were removed by centrifugation (300*g* for 1 min) and the resulting supernatant was used to prepare the gavage material (100 µl per mouse). All mice received a single gavage after an overnight fast. To measure the *A. muciniphila* abundance, faecal pellets were collected at regular intervals and DNA was extracted using a Qiagen Fast DNA Stool kit or Qiagen PowerSoil Plus.

Quantitative PCR 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 produce five-point standard curves with tenfold dilutions of a 10 ng ml^{-1} starting concentration. *A. muciniphila*-specific 16S rRNA primers were used to measure total *Akkermansia* loads, while strain-specific primers designed to flank the Tn insertion site, to measure total loads for each mutant, were used for the competition experiments. To validate the primers for the competition experiments and ensure that they were specific to their intended target, each primer pair was tested using serial dilutions of the target strain diluted in a 10 µg ml^{-1} solution of the competing strain. The primers are listed in Supplementary Data 1. PCR was run with the PowerUp SYBR Master Mix (Thermo Fisher Scientific) 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 faecal material or caecal 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 germ-free mice colonized for 14 d with either wild-type *A. muciniphila* or the *mul1A*::Tn mutant. The intestinal contents were collected to quantify *A. muciniphila* levels along the gastrointestinal tract, and tissue was collected to prepare RNA. The entire intestinal tract was

removed and the contents of the colon, caecum, 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 (Precellys) and the samples were lysed using a bead beater set to 5,000 RMP for 15 s and three cycles. RNA was then isolated using Zymo's Direct-zol RNA Miniprep kit (R2050; Zymo) according to the manufacturer's instructions. RNA was quantified with a 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 RNA-seq kit. Libraries were sequenced on a NovaSeq 6000 S Prime 50-bp paired-end flow cell. Sequencing data were trimmed and filtered as described for the bacterial samples, except that the reads were mapped to the *Mus musculus* genome (assembly GRCm38) using STAR with the flag --sjdbOverhang 49. Differential expression was analysed with DESeq2 (ref. 57) and pathway analysis was run using MetaScape[69]. *P* values were adjusted using the Benjamini–Hochberg method. To adjust for multiple testing, differentially expressed genes were plotted with the R package EnhancedVolcano.

Co-immunoprecipitation and mass spectrometry

Native immunoprecipitation coupled with liquid chromatography/tandem mass spectrometry (LC/MS/MS) was used to identify proteins that interact with Mul1A (Amuc_0544). Bacterial cultures were prepared with either the 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. Each sample was prepared in triplicate. Cultures were grown in 30 ml media to an OD_{600} of 0.6–0.8 and the cells were harvested by centrifugation at 4,000*g* for 20 min. The supernatant was discarded, and the cell pellets were stored at −80 °C overnight, then subsequently thawed in 10 ml of a lysis buffer (150 mM NaCl, 50 mM Tris (pH 8), 1% dodecyl-β-maltoside and 1 mM phenylmethylsulfonyl fluoride). The cells were lysed by sonication at 30-s intervals on ice and remaining cell debris were pelleted by centrifugation at 15,000*g* for 25 min at 4 °C. The resulting supernatants were used for immunoprecipitation of Mul1A. Affinity-purified anti-Mul1A was bound to protein A magnetic beads following the manufacturer's protocol (Bio-Rad), then 50 µl of the resin was added to cell lysates and incubated for 48 h at 4 °C with rotation. The beads were collected and washed four times with 50 mM NaCl and 25 mM Tris (pH 8) buffer. Proteins were eluted with 25 µl 62.5 mM Tris 2% SDS at 60 °C for 10 min and submitted to the Duke University 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 S-Trap (ProtiFi) binding buffer (90% MeOH and 100 mM TEAB). Proteins were trapped on the S-Trap, digested using 20 ng μl^{-1} sequencing-grade trypsin (Promega) for 1 h at 47 °C and eluted using 50 mM triethyl ammonium bicarbonate (TEAB), followed by 0.2% formic acid, and lastly 50% acetonitrile and 0.2% formic acid. All of the samples were then lyophilized to dryness and resuspended in 12 µl 1% formic acid and 2% acetonitrile containing 12.5 fmol μl^{-1} yeast alcohol dehydrogenase. From each sample, 3 µl was removed to create a quality control pool sample that was run periodically throughout the acquisition period.

Quantitative LC/MS/MS was performed on 3 µl of each sample using a nanoACQUITY ultra-performance liquid chromatography system (Waters Corporation) coupled to a Thermo Orbitrap Fusion Lumos high-resolution accurate mass tandem mass spectrometer (Thermo Fisher Scientific) via a nanoelectrospray ionization source. Briefly, the sample was first trapped on a Symmetry C18 20 mm × 180 µm trapping column (5 µl min^{-1} at 99.9/0.1 vol/vol water/acetonitrile), after which analytical separation was performed using a 1.8 µm ACQUITY HSS T3 C18 75 µm × 250 mm column (Waters Corporation) with a 90-min linear gradient of 5–30% acetonitrile with 0.1% formic acid at a flow rate of 400 nl min^{-1} and a column temperature of 55 C. Data collection on the Fusion Lumos mass spectrometer was performed in data-dependent acquisition mode with an $r = 120{,}000$ (at m/z 200) full mass spectrometry scan from m/z 375–1,500 and a target AGC value of 2×10^5 ions. MS/MS scans were acquired at the rapid scan rate (ion trap) with an automatic gain control (AGC) target of 5×10^3 ions and a max injection time of 200 ms. The total cycle time for mass spectrometry and MS/MS scans was 2 s. A 20-s dynamic exclusion was employed to increase the depth of coverage. The total analysis cycle time for each sample injection was ~2 h. Following 13 total ultra-performance LC/MS/MS analyses (excluding conditioning runs but including four replicate quality control injections; Supplementary Data 1), the data were imported into Proteome Discoverer 2.2 (Thermo Fisher Scientific) and analyses were aligned based on the accurate mass and retention time of detected ions (features) using the Minora Feature Detector algorithm in Proteome Discoverer. The relative peptide abundance was calculated based on the areas under the curve of the selected ion chromatograms of the aligned features across all runs. The MS/MS data were searched against the TrEMBL *A. muciniphila* database (downloaded in April 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 (version 2.5; Matrix Science) were used to produce fragment ion spectra and to perform the database searches. The database search parameters included fixed modification on Cys (carbamidomethyl) and variable modifications on Met (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 protein false discovery rate of 1%.

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 versus the *mul1A* mutant negative control ($P < 0.01$) and to have more than one peptide hit. The resulting 81 proteins were then screened for the presence of type I and II signal peptides using the programs SignalP (versions 3 and 4) and LipoP[70]. 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[71]. The size of each node was set to represent the $\log_2$[fold change] normalized peptide abundance in the wild type versus the *mul1A* mutant negative control and the line thickness represents the $-\log_{10}$[*P* value]. The colour of the nodes was used to represent their respective Pfam[32] associations.

16S rRNA gene sequencing

Genomic DNA was extracted from faecal pellets collected from mice housed in separate cages using a Qiagen DNeasy PowerSoil DNA extraction kit (Qiagen). Amplicon sequencing was performed with primers targeting the V4 region of the 16S rRNA gene, as previously described[72], at Duke University's Sequencing and Genomic Technologies core facility using an Illumina MiSeq with paired-end 250-bp reads.

Taxonomy assignment and bioinformatics analysis

Demultiplexed sequences were obtained from the sequencing core. We used the DADA2 (version 1.25.2)[73] R package to filter and trim the reads, keeping bases between positions 30 and 240. DADA2 was then used to denoise, dereplicate and merge the forward and reverse reads, and to remove bimeras. Taxonomy was assigned using the SILVA database (version 138.1). Additional analyses were performed using the R packages phyloseq[74], ape, vegan and microshades. Comparisons of microbial

community composition were run using weighted UniFrac distances at the genus level. Linear discriminant analysis effect size (LEfSe)[75] was run using default settings and accessed through the Huttenhower Galaxy website at https://huttenhower.sph.harvard.edu/galaxy.

Analysis of SCFAs in the mouse caecum

Caecal contents were used to prepare 20% (wt/vol) slurries in 0.5 ml PBS. The slurries were acidified to a pH of <3 with 25 µl of 6 N HCL and thoroughly mixed. Debris were removed by centrifugation at 14,000*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 analysed on an Agilent 7890 gas chromatograph equipped with a flame-ionization detector and an Agilent HP-FFAP free fatty acid column. The concentrations of acetate and propionate in the samples were determined using an eight-point standard curve (0.1–16 mM). The ratio of propionate to acetate was analysed by two-tailed Student's *t*-test after removing two outliers identified by the robust regression and outlier removal (ROUT) method in GraphPad Prism with a *Q* coefficient value of 1%. Differences in acetate and propionate levels were analysed by one-way analysis of variance (*F* statistic) and Tukey's post-hoc test.

Statistical analysis

Unless otherwise noted, GraphPad Prism version 9.0.0 was used for statistical analysis and plot generation. The data distribution was assumed to be normal but this was not formally tested. Data collection and analysis were not performed blind to the conditions of the experiments. No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those reported in previous publications[38,51].

Reporting summary

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

Data availability

Analysed data, including INSeq, RNA-seq and mass spectrometry outputs, as well as primer and adaptor sequences are available in Supplementary Data 1. Sequencing data used for this study can be found in the National Center for Biotechnology Information BioProject database with the accession code PRJNA955715. Unprocessed data from mass spectrometry and additional supporting data are available upon reasonable request from the corresponding authors. Source data are provided with this paper.

Code availability

Code for analysing the INSeq data is available at https://github.com/pmalkus/Akk_INseq_paper.

References

1. 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).
2. Blacher, E. et al. Potential roles of gut microbiome and metabolites in modulating ALS in mice. *Nature* **572**, 474–480 (2019).
3. Xie, J. et al. *Akkermansia muciniphila* protects mice against an emerging tick-borne viral pathogen. *Nat. Microbiol.* **8**, 91–106 (2023).
4. Derosa, L. et al. Intestinal *Akkermansia muciniphila* predicts clinical response to PD-1 blockade in patients with advanced non-small-cell lung cancer. *Nat. Med.* **28**, 315–324 (2022).
5. Yoon, H. S. et al. *Akkermansia muciniphila* secretes a glucagon-like peptide-1-inducing protein that improves glucose homeostasis and ameliorates metabolic disease in mice. *Nat. Microbiol.* **6**, 563–573 (2021).
6. Zhang, Q. et al. Genetic mapping of microbial and host traits reveals production of immunomodulatory lipids by *Akkermansia muciniphila* in the murine gut. *Nat. Microbiol.* **8**, 424–440 (2023).
7. 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. USA* **108**, 4659–4665 (2011).
8. Wlodarska, M. et al. Indoleacrylic acid produced by commensal *Peptostreptococcus* species suppresses inflammation. *Cell Host Microbe* **22**, 25–37 (2017).
9. Shon, D. J. et al. An enzymatic toolkit for selective proteolysis, detection, and visualization of mucin-domain glycoproteins. *Proc. Natl Acad. Sci. USA* **117**, 21299–21307 (2020).
10. Trastoy, B., Naegeli, A., Anso, I., Sjögren, J. & Guerin, M. E. Structural basis of mammalian mucin processing by the human gut *O*-glycopeptidase OgpA from *Akkermansia muciniphila*. *Nat. Commun.* **11**, 4844 (2020).
11. Medley, B. J. et al. A previously uncharacterized *O*-glycopeptidase from *Akkermansia muciniphila* requires the Tn-antigen for cleavage of the peptide bond. *J. Biol. Chem.* **298**, 102439 (2022).
12. Crouch, L. I. et al. Prominent members of the human gut microbiota express endo-acting *O*-glycanases to initiate mucin breakdown. *Nat. Commun.* **11**, 4017 (2020).
13. Meng, X. et al. A purified aspartic protease from *Akkermansia muciniphila* plays an important role in degrading Muc2. *Int. J. Mol. Sci.* **21**, 72 (2020).
14. 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).
15. Kosciow, K. & Deppenmeier, U. Characterization of a phospholipid-regulated β-galactosidase from *Akkermansia muciniphila* involved in mucin degradation. *MicrobiologyOpen* **8**, e00796 (2019).
16. Guo, B.-S. 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).
17. 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).
18. Pruss, K. M. et al. Mucin-derived *O*-glycans supplemented to diet mitigate diverse microbiota perturbations. *ISME J.* **15**, 577–591 (2021).
19. 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).
20. 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).
21. 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).
22. Hansson, G. C. Mucins and the microbiome. *Annu. Rev. Biochem.* **89**, 769–793 (2020).
23. Lensmire, J. M. & Hammer, N. D. Nutrient sulfur acquisition strategies employed by bacterial pathogens. *Curr. Opin. Microbiol.* **47**, 52–58 (2019).
24. 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).
25. Kostopoulos, I. et al. *Akkermansia muciniphila* uses human milk oligosaccharides to thrive in the early life conditions in vitro. *Sci. Rep.* **10**, 14330 (2020).

26. Ottman, N. et al. Pili-like proteins of *Akkermansia muciniphila* modulate host immune responses and gut barrier function. *PLoS ONE* **12**, e0173004 (2017).
27. 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**, e16876 (2011).
28. 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**, e01014–e01017 (2017).
29. Thibault, D. et al. Droplet Tn-Seq combines microfluidics with Tn-Seq for identifying complex single-cell phenotypes. *Nat. Commun.* **10**, 5729 (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. Schwalm, N. D. & Groisman, E. A. Navigating the gut buffet: control of polysaccharide utilization in *Bacteroides* spp. *Trends Microbiol.* **25**, 1005–1015 (2017).
32. Mistry, J. et al. Pfam: the protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419 (2021).
33. Marchler-Bauer, A. et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* **45**, D200–D203 (2017).
34. Cortajarena, A. L. & Regan, L. Ligand binding by TPR domains. *Protein Sci.* **15**, 1193–1198 (2006).
35. 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).
36. Mou, L. et al. Crystal structure of monomeric Amuc-1100 from *Akkermansia muciniphila*. *Acta Crystallogr. F Struct. Biol. Commun.* **76**, 168–174 (2020).
37. Velcich, A. et al. Colorectal cancer in mice genetically deficient in the mucin Muc2. *Science* **295**, 1726–1729 (2002).
38. Becken, B. et al. Genotypic and phenotypic diversity among human isolates of *Akkermansia muciniphila*. *mBio* **12**, e00478-21 (2021).
39. 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).
40. Glenwright, A. J. et al. Structural basis for nutrient acquisition by dominant members of the human gut microbiota. *Nature* **541**, 407–411 (2017).
41. 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).
42. Boedeker, C. et al. Determining the bacterial cell biology of *Planctomycetes*. *Nat. Commun.* **8**, 14853 (2017).
43. 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).
44. Faham, S. et al. The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na^+/sugar symport. *Science* **321**, 810–814 (2008).
45. 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).
46. 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**, e01438–14 (2014).
47. Schaum, N. et al. Single-cell transcriptomics of 20 mouse organs creates a *Tabula Muris*. *Nature* **562**, 367–372 (2018).
48. Wang, B. et al. Phospholipid remodeling and cholesterol availability regulate intestinal stemness and tumorigenesis. *Cell Stem Cell* **22**, 206–220 (2018).
49. McFarlane, M. R. et al. Scap is required for sterol synthesis and crypt growth in intestinal mucosa. *J. Lipid Res.* **56**, 1560–1571 (2015).
50. 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).
51. 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).
52. Grondin, J. M., Tamura, K., Déjean, G., Abbott, D. W. & Brumer, H. Polysaccharide utilization loci: fueling microbial communities. *J. Bacteriol.* **199**, e00860-16 (2017).
53. 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).
54. Pan, Y. & Kaatz, L. Use of image-based flow cytometry in bacterial viability analysis using fluorescent probes. *Curr. Protoc. Microbiol.* **27**, 2C.5.1–2C.5.11 (2012).
55. Aronesty, E. Comparison of sequencing utility programs. *Open Bioinform. J.* **7**, 1–8 (2013).
56. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
57. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
58. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS J. Integr. Biol.* **16**, 284–287 (2012).
59. 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).
60. Stothard, P. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *BioTechniques* **28**, 1102–1104 (2000).
61. O'Toole, G. A. et al. Genetic approaches to study of biofilms. *Methods Enzymol.* **310**, 91–109 (1999).
62. Najah, M., Griffiths, A. D. & Ryckelynck, M. Teaching single-cell digital analysis using droplet-based microfluidics. *Anal. Chem.* **84**, 1202–1209 (2012).
63. Shames, S. R. et al. Multiple *Legionella pneumophila* effector virulence phenotypes revealed through high-throughput analysis of targeted mutant libraries. *Proc. Natl Acad. Sci. USA* **114**, E10446–E10454 (2017).
64. Pritchard, J. R. et al. ARTIST: high-resolution genome-wide assessment of fitness using transposon-insertion sequencing. *PLoS Genet.* **10**, e1004782 (2014).
65. 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).
66. Karp, P. D. et al. The BioCyc collection of microbial genomes and metabolic pathways. *Brief. Bioinformatics* **20**, 1085–1093 (2018).
67. Ansaldo, E. et al. *Akkermansia muciniphila* induces intestinal adaptive immune responses during homeostasis. *Science* **364**, 1179–1184 (2019).
68. 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).
69. Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).

70. 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).
71. Shannon, P. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
72. Thompson, L. R. et al. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**, 457–463 (2017).
73. Callahan, B. J. et al. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
74. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217 (2013).
75. Segata, N. et al. Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60 (2011).
76. Paley, S. et al. The omics dashboard for interactive exploration of gene-expression data. *Nucleic Acids Res.* **45**, 12113–12124 (2017).

Acknowledgements

We are thankful to O. Kuddar and E. Rivas for support with the assembly of arrayed Tn mutant libraries; J. Granek for the base trimming code; A. Sharma for preparing sequencing libraries; and members of the R.H.V. laboratory for critical reading of the manuscript. We thank L. Augenlicht at the Albert Einstein College of Medicine for providing the Muc2$^{-/-}$ mice. This work was supported by National Institutes of Health awards AI142376 and DK110496 (to R.H.V.), American Heart Association award 18POST34070017 (to L.E.D.) and a fellowship from the Natural Sciences and Engineering Research Council of Canada (PDF4878642016 to L.E.D.).

Author contributions

R.H.V., L.E.D. and P.N.M. designed the research. L.E.D., P.N.M., M.V., L.D. and E.A. performed the experiments and analysed the data. L.E.D. prepared the figures. P.M.N. wrote the INSeq analysis code. Z.C.H. and J.L. contributed to running the SCFA analysis. M.V. performed the microfluidic droplet experiments. L.D. performed the live imaging experiments. L.E.D. and R.H.V. wrote the manuscript. R.H.V. supervised the project. All authors reviewed and edited the manuscript.

Competing interests

R.H.V. is a founder of Bloom Science.

Additional information

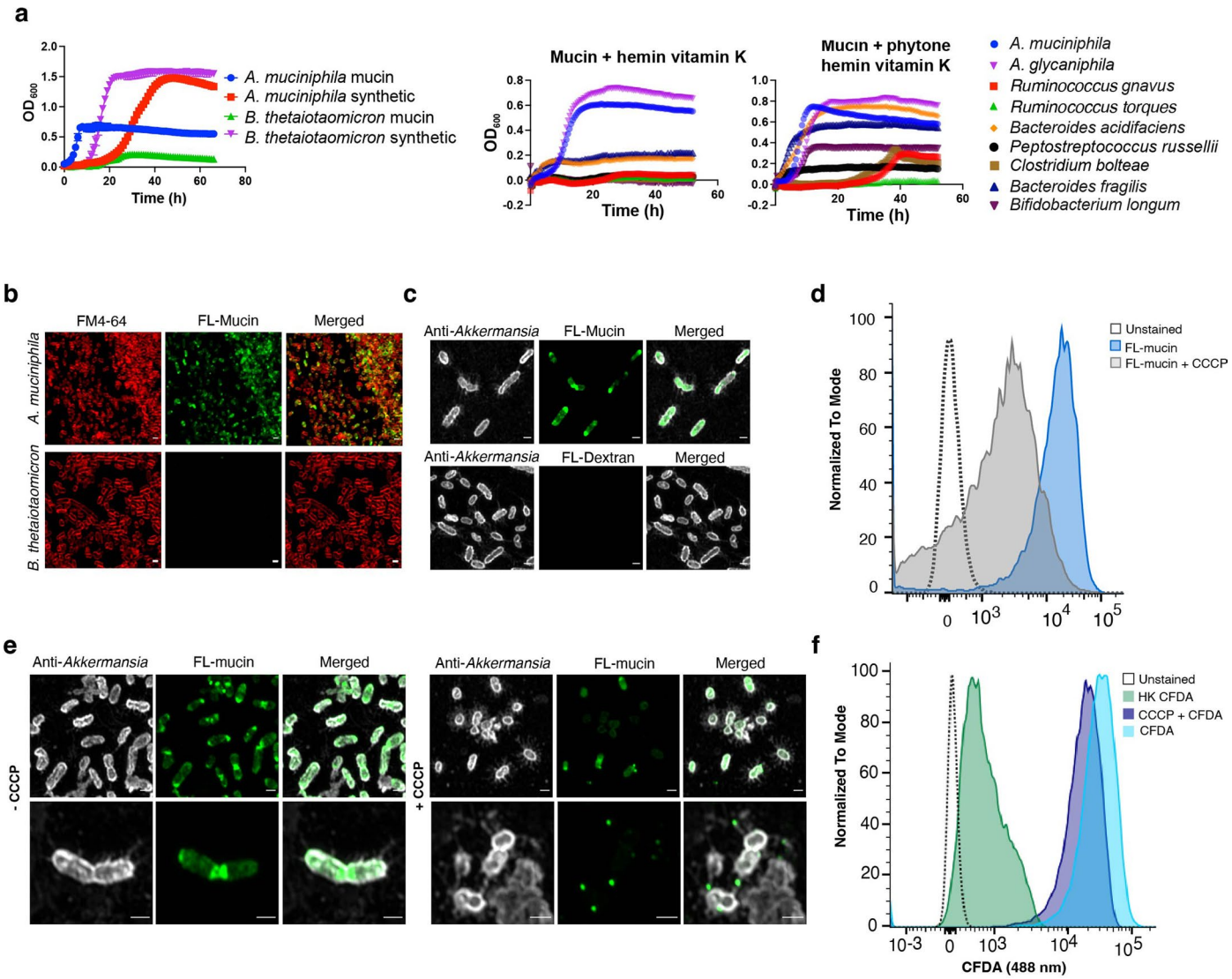
Extended data is available for this paper at https://doi.org/10.1038/s41564-023-01407-w.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41564-023-01407-w.

Correspondence and requests for materials should be addressed to Lauren E. Davey or Raphael H. Valdivia.

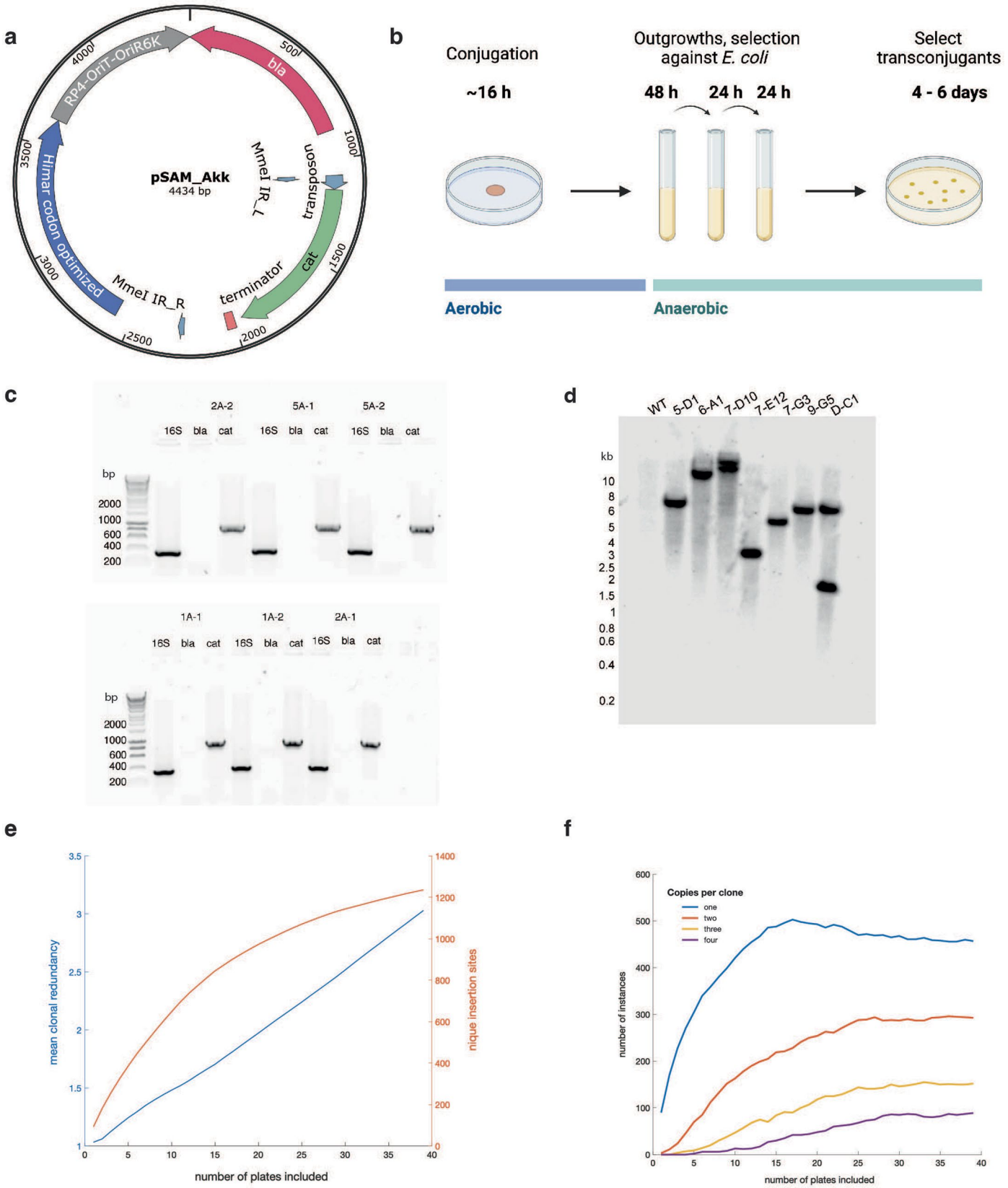
Peer review information *Nature Microbiology* thanks Matthew Waldor and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

https://doi.org/10.1038/s41564-023-01407-w



Extended Data Fig. 1 | *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 labelled with FM4-64. Experiments were repeated twice. (**c-d**) Mucin uptake is a specific and active process. *A. muciniphila* grown in the presence of either fluorescein-mucin or fluorescein-dextran (green) for 3 h and stained with anti-*Akkermansia* anti-sera (anti-Akk). All microscopy was performed at least three times (**c**). Flow cytometric analysis of cells grown in the presence of fluorescein-mucin for 3 h, 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 3 h without CCCP or with CCCP treatment (**e**). Flow cytometry analyses of *A. muciniphila* grown with the cell permanent esterase carboxyfluorescein diacetate (CFDA) in the presence and absence of CCCP, and after heat inactivation (**f**). Scale bar, 1 µm. Error bars represent the standard error of the mean.

https://doi.org/10.1038/s41564-023-01407-w

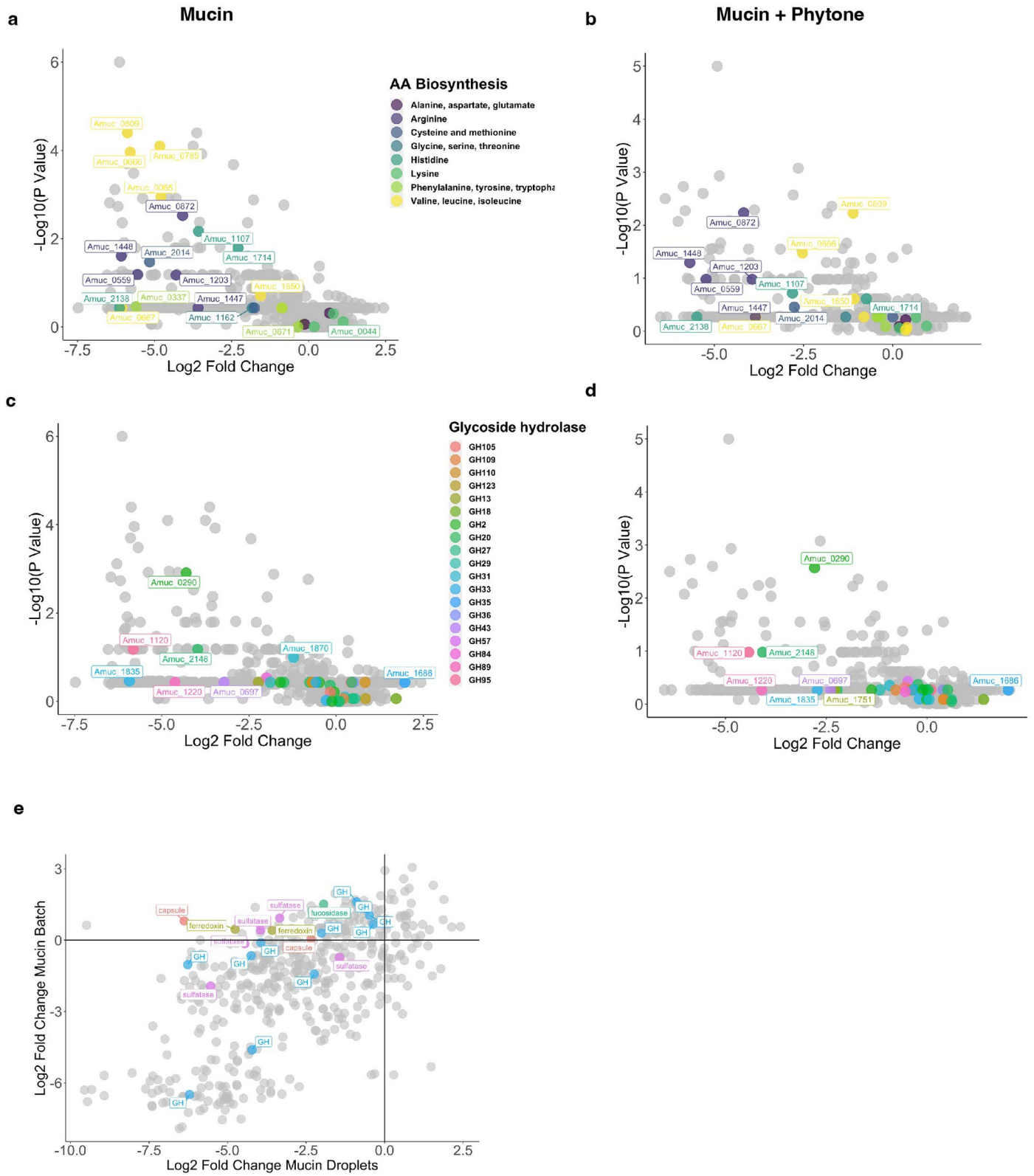


Extended Data Fig. 2 | See next page for caption.

https://doi.org/10.1038/s41564-023-01407-w

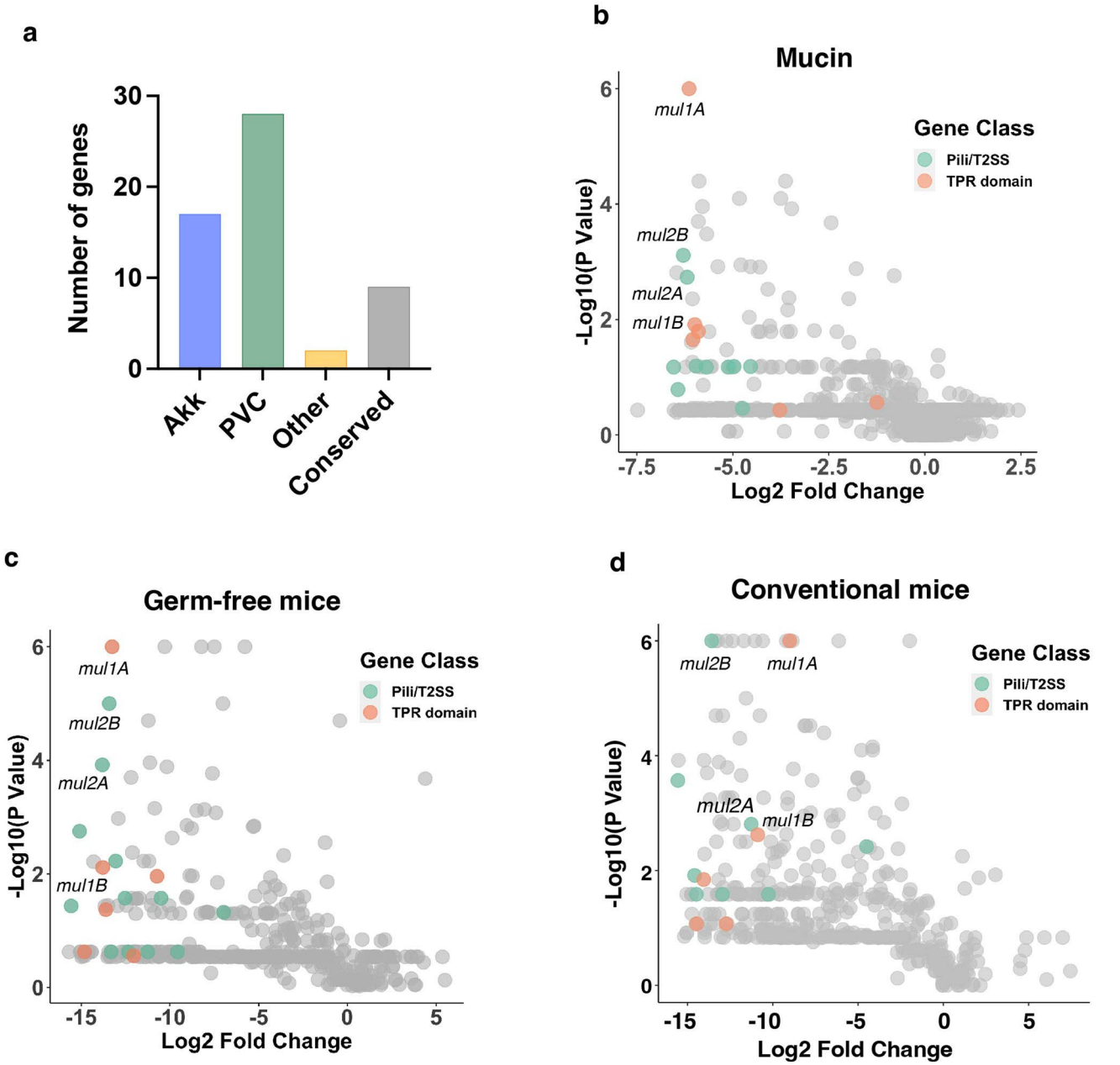
Extended Data Fig. 2 | Transposon mutagenesis in *A. muciniphila*. (**a**) Map of the *A. muciniphila* optimized INSeq plasmid. (**b**) Overview of the *A. muciniphila* conjugation protocol. (**c**) PCR analysis confirming transposition. DNA from representative Tn mutants was amplified with primers for *A. muciniphila* specific 16 S rRNA, the *bla* gene located on the delivery plasmid backbone, and the *cat* gene located with the transposon. This analysis was performed for every transposition experiment. (**d**) Southern blot analysis of Tn mutant DNA digested with *Hin*dIII and probed with DIG-labelled probes that recognize the *cat* gene in the Tn insert. Data is representative of two experiments with similar results. (**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).

https://doi.org/10.1038/s41564-023-01407-w



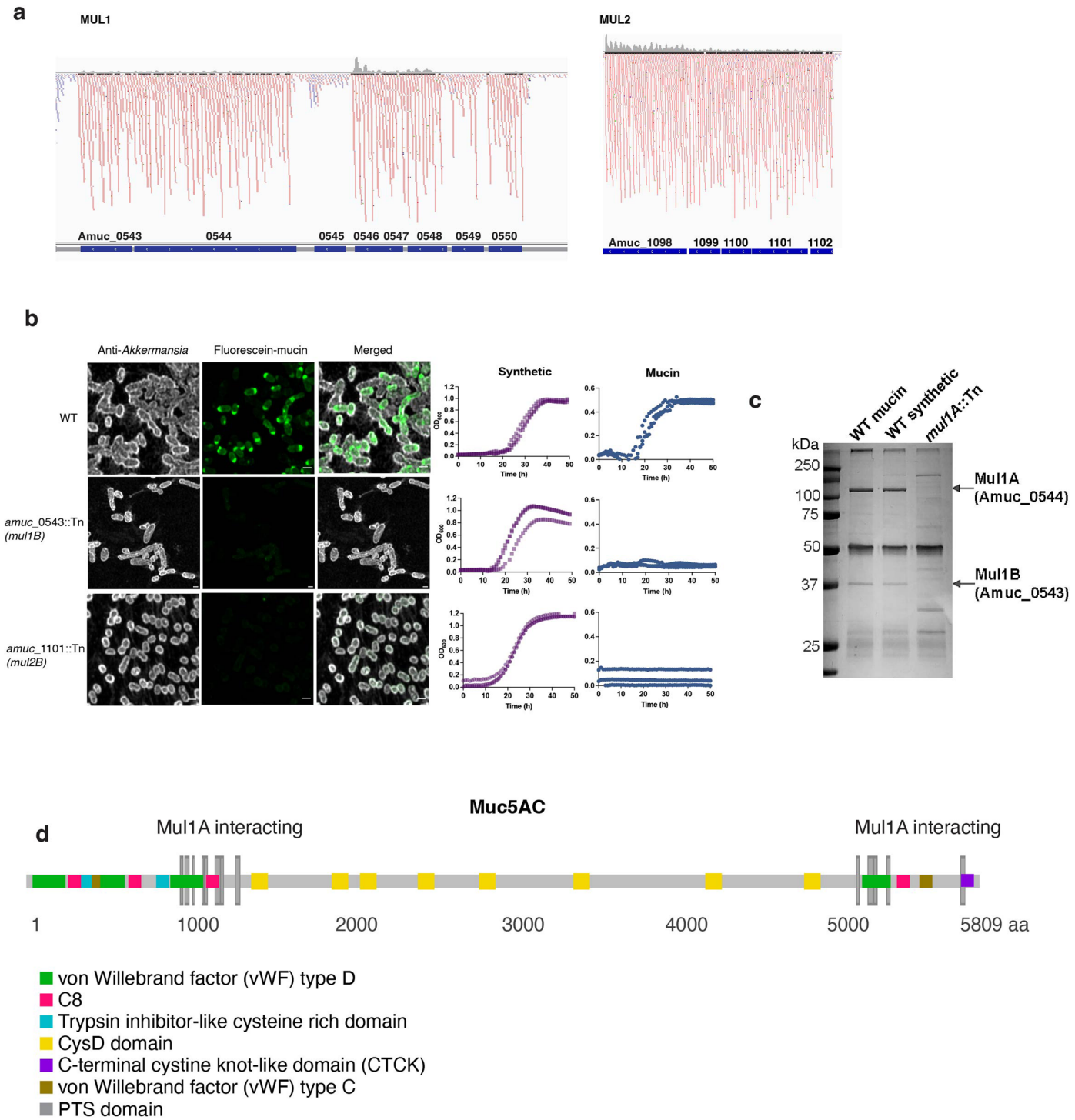
Extended Data Fig. 3 | INSeq analysis of relative nutritional requirements for *A. muciniphila* to grow in mucin medium and the role of putative glycan hydrolases. Plot of INSeq data from Tn mutant pools grown for eight generations in mucin medium where each dot represents all inserts in a specific gene. 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. Statistical analysis on INSeq data was performed with a Mann-Whitney Utest. **(e)** 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 (72 h). Selected genes that were depleted in one condition relative to the other are highlighted on the plot. GH, glycosyl hydrolase.

https://doi.org/10.1038/s41564-023-01407-w



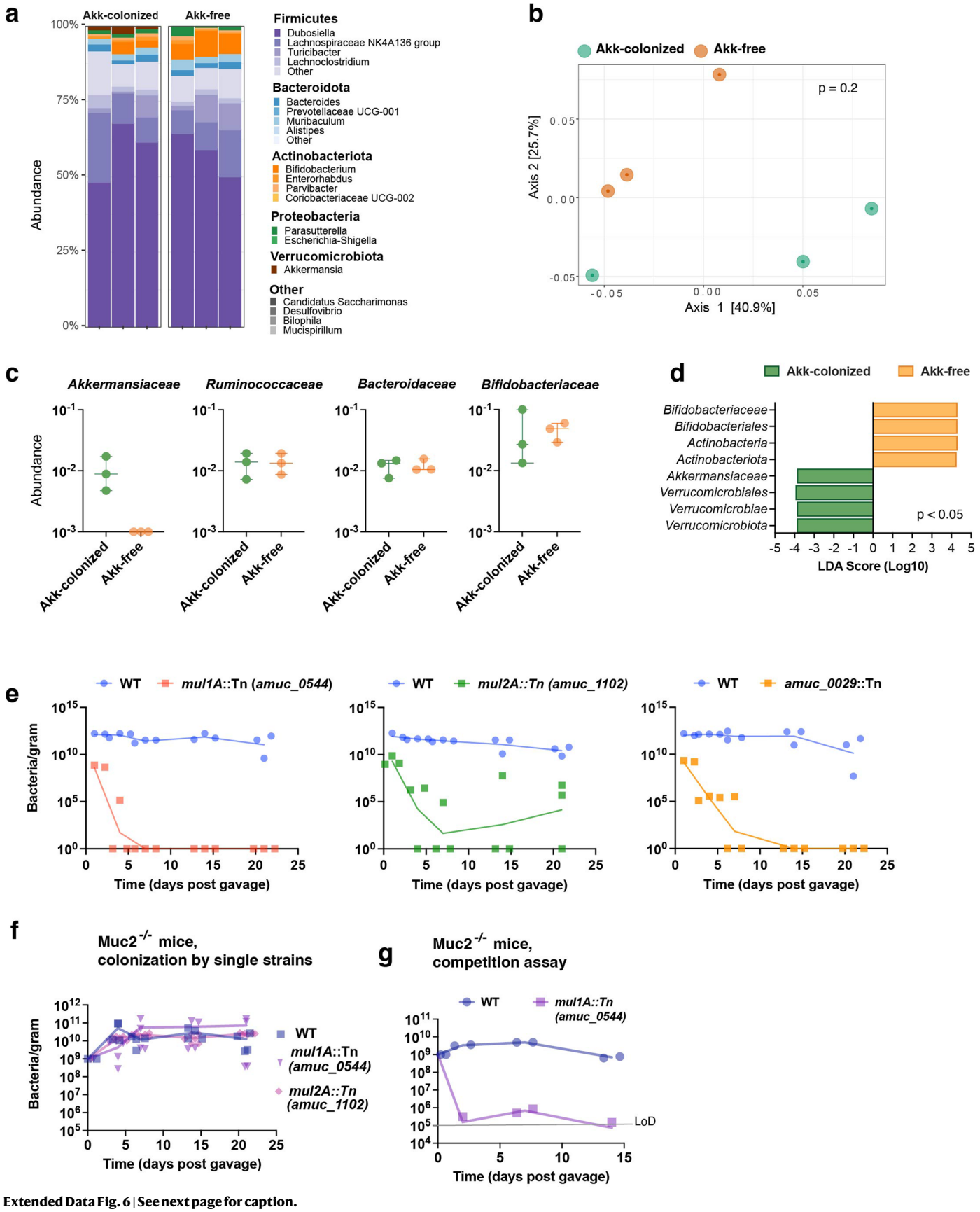
Extended Data Fig. 4 | A significant proportion of *A. muciniphila* genes required for growth in mucin medium are specific to *Akkermansia*/*Verrucomicrobia*. (**a**) Number of genes required for optimal *A. muciniphila* growth in mucin medium that lack functional annotations. Genes corresponding to 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.

https://doi.org/10.1038/s41564-023-01407-w



Extended Data Fig. 5 | Evidence for the presence of a stable Mul1A-Mul1B protein complex. Transcriptional analysis of *Mul1* operons. (**a**) View of RNA-seq reads generated from wild type *A. muciniphila* grown in mucin medium mapped to genes in the *mul1* and *mul2* loci. (**b**) Growth curves for wild type *A. muciniphila* and mutants in *mul1B* and *mul2B* grown in triplicate in synthetic medium or with mucin as the sole carbon and nitrogen source and corresponding microscopy with FL-mucin (green). Cells are stained with anti-*Akkermansia* antisera (white). The scale bar is 1 µm. (**c**) 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. (**d**) Depiction of Conserved Domains (colours) in Muc5AC and locations of peptides identified as co-precipitating with Mul1A (vertical bars). The experiment was performed in triplicate.

https://doi.org/10.1038/s41564-023-01407-w

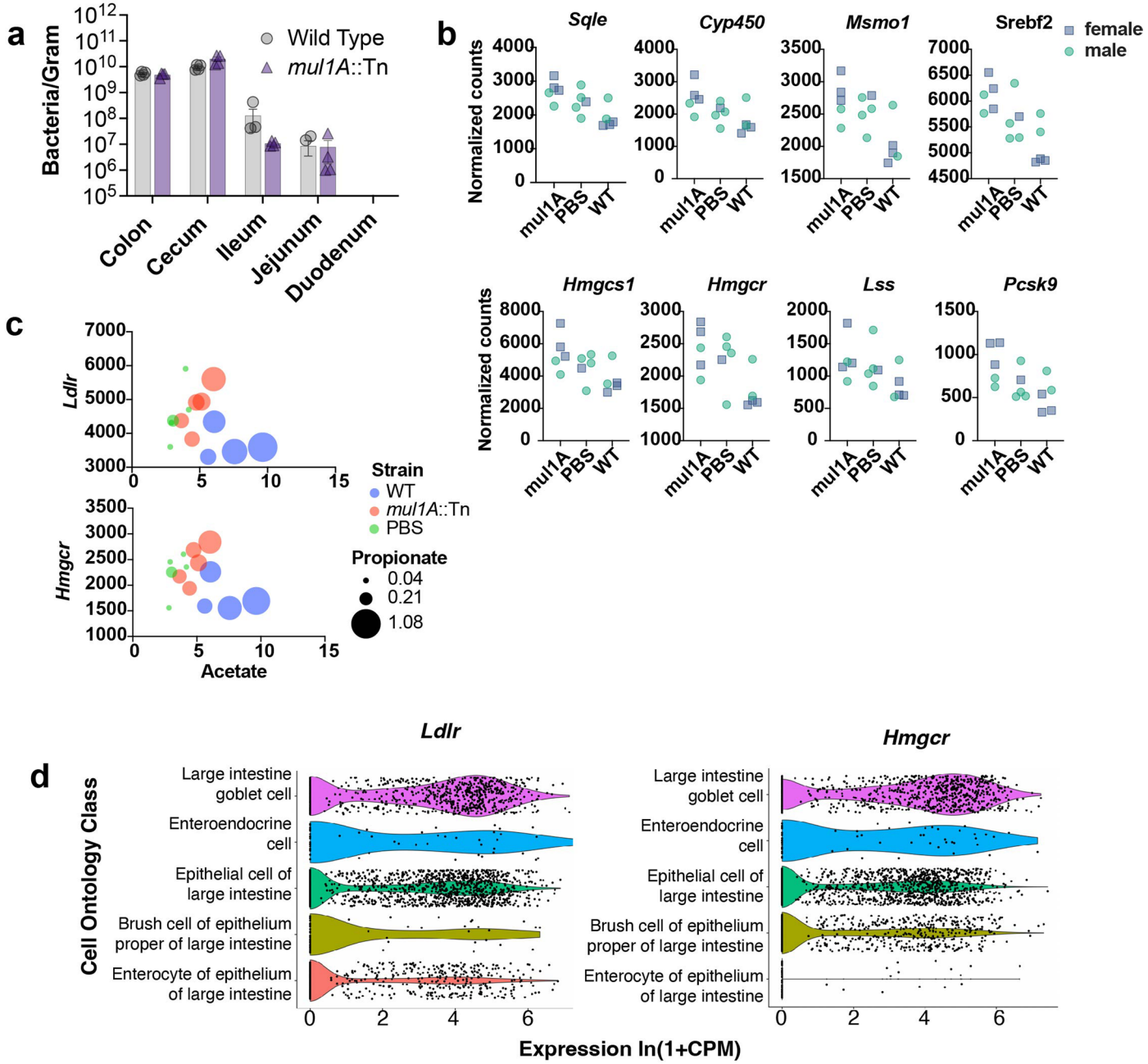


Extended Data Fig. 6 | See next page for caption.

https://doi.org/10.1038/s41564-023-01407-w

Extended Data Fig. 6 | Mucin utilization is required for *A. muciniphila* to compete in CONV mice and in *Muc2*$^{-/-}$ mice. A breeding colony of *Akkermansia*-free mice (*Akk*-free) was generated to facilitate mouse colonization without antibiotic pre-treatment. (**a-c**) Comparison of the microbiota of *Akkermansia* colonized (*Akk*-colonized) and *Akk*-free mice by 16 S rRNA gene sequencing. (**a**) Relative abundances of fecal bacteria at the genus level in *Akk*-colonized and *Akk*-free mice. Each sample was obtained from separately housed mice (n = 3 per group). (**b**) Principal Coordinates Analysis (PCoA) performed on weighted UniFrac distances. Statistical significance was determined by Permutational Multivariate Analysis of Variance (PERMANOVA). (**c**) Relative abundances of potential mucin-degrading taxa at the family level. The centre line is the mean, and the whiskers show the minimum and maximum. (**d**) Linear discriminant analysis Effect Size (LEfSe) analysis of *Akk*-colonized and *Akk*-free mice. The Kruskal-Wallis test was used to detect features with a significant differential abundance ($p < 0.05$). (**e**) CONV mice were pre-treated with antibiotics and gavaged with a 1:1 mix of WT and mutant *A. muciniphila* prepared with a fecal slurry from *Akk*-free mice to partially reconstitute the microbiota (n = 6 per group). Bacterial loads in fecal pellets were quantified by qPCR, each point represents one cage. **(f)** Colonization of mucin deficient *Muc2*$^{-/-}$ mice with *A. muciniphila*. Each point represents the average *A. muciniphila* per gram of feces (WT, n = 4; *mul1A*::Tn, n = 4; *mul2A*::Tn, n = 6). (**g**) 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.

https://doi.org/10.1038/s41564-023-01407-w



Extended Data Fig. 7 | 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. Data are presented as mean values +/- SEM. The analysis was carried out with the same female mice that were used for RNAseq. (**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*[47]. Violin plots show the expression of *Ldlr* and *Hmgcr* in mouse colonic epithelial and goblet cells.

nature portfolio

Corresponding author(s): Raphael H. Valdivia

Last updated by author(s): May 3, 2023

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	Microscopy images were collected with Zeiss Zen software (v2.3), NIS-ElementsAR (v5.20.02, Nikon), and Leica Application Suite X. Flow cytometry data was acquired using FACSDiva software (v9, BD)
Data analysis	1. Microscopy - STED images were deconvolved with Huygens Professional (no version given, SVI) and analyzed with Imaris (v9.5, Oxford Instruments) - All other images were analyzed with ImageJ2/FIJI (v2.3.0, NIH) - Flow cytometry data was analyzed using FlowJo (v9 and 10) 2. INSeq, RNA-seq, and 16s rRNA sequencing - INSeq data was mapped in MATLAB (vR2019b) using a published data analysis package (Goodman et al., PMID: 22094732) - Quality control and normalization was performed using custom code based on ARTIST (Pritchard et al., PMID: 25375795). The code is available on Github (https://github.com/pmalkus/Akk_INseq_paper) - INSeq data was analyzed with TRANSIT (v3.2.1, https://github.com/mad-lab/transit) - The programs fastq-mcf (part of the EA-Utils package, v1.05) and FASTQC (v0.11.9) were used to trim raw reads and to assess read quality - RNA-seq reads were mapped using STAR (v2.7.5c) - Differential gene expression was analyzed with DESeq2 (v1.30.1) - Data was further analyzed using Metascape (v3.5), ClusterProfiler (v4.2.2), and BioCyc (v19) - 16s rRNA sequencing data was processed with dada2 (v1.25.2) to filter, trim, dereplicate, and merge reads. - ASVs were analyzed using Phyloseq (v1.42.0) and the vegan (v2.6-4) function adonis2 was use to run PERMANOVA analysis - Stacked barplots were generated using microshades (v1.10) - LEfSe analysis was run through the Huttenhower Lab Galaxy module (http://huttenhower.sph.harvard.edu/galaxy, Hutlab Galaxy v1), and input files were formatted using phyloseqCompanion (v1.1) -DNA/SDS-PAGE gels and blots were labeled using Adobe Photoshop v.24.1.1

3. Proteomics
- SignalP (v3 and v4) was used to predict secreted proteins
- Protein data was visualized with Cytoscape (v3.8.2)

4. Graphing and statistical analysis was performed using
- GraphPad Prism (v9) was used for statistical analysis and graphing
- R Studio (v2021.09.0) was used for volcano plots and Venn diagrams, using the packages ggplot2 (v3.4.0) and gghighlight (v0.4.0)

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

Data

Policy information about availability of data

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

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

Data generated in this study are available from the corresponding author (raphael.valdivia@duke.edu)
Taxonomy was assigned in dada2 using the Silva database (v138.1)

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	Sample size was determined based on similar studies in the field (PMID: 34006653; PMID: 19748469)
Data exclusions	We used pre-established quality control metrics as criteria to exclude some INSeq data sets (described in detail in the methods section). These metrics were designed to detect samples with insufficient sequencing depth and a high left to right mapping ratio of sequences flanking the transposon, indicative of low sequencing quality. Two samples were omitted from SCFA analysis as they were identified as statistically significant outliers using Graphpad Prism's ROUT outlier test.
Replication	All experiments were performed multiple times to ensure reproducibility. INSeq experiments in mice included 4 - 12 mice per group, and in vitro samples were prepared in duplicate. In addition to the data included in the paper, we performed additional experiments both in vitro and in vivo at additional time points to ensure reproducibility. For the proteomics data, we performed quantitative mass spectrometry with biological triplicates, as well as confirmatory studies performed independently and analyzed separately. Flow cytometry experiments were run multiple times on different days. Similarly, samples for microscopy were prepared independently, on different days and with different batches of reagents, all with consistent results. None of the experiments in the manuscript include results that were not reproducible.
Randomization	Mice used in colonization experiments were randomly assigned to colonization by wild-type or mutant bacteria. For all RNAseq and INSeq animal experiments, mice were randomly assigned to different groups, ensuring that each group included mice housed in at least three separate cages to control for cage effects. All groups were designed to all have mice with similar ages and numbers of males and females.
Blinding	No blinding was performed in this study. Blinding was not feasible for our animal experiments because they required knowledge of colonization levels and detection of contaminants (eg. to ensure Akk-free mice remained Akkermansia free) for downstream experiments.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Antibodies

Antibodies used

1. Goat anti-rabbit IgG Alexa-Fluor-647 (A-21244, Invitrogen), (1:1000)
2. Antibodies against Amuc_0544 were generated in rabbits at Duke University's Division of Laboratory Animal Resources (1:100)
3. Antibodies against Akkermansia were generated in rabbits at Duke University's Division of Laboratory Animal Resources (1:100)
4. Goat anti-rabbit IgG Alexa Fluor-594, Invitrogen (1:1000, A32740)

Validation

Antibodies generated in-house were validated by testing with appropriate positive and negative controls:
1. Anti-Amuc_0544 anti-sera was tested against wild-type Akkermansia and mutants lacking Amuc_0544.
2. Anti-Akkermansia anti-sera was tested for specificity by comparison to non-Akkermansia bacteria and using colon tissue samples from mice with and without Akkermansia colonization.

Animals and other organisms

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

Laboratory animals

All animal experiments were carried out in C57BL/6J mice:
1. Experiments in conventional mice were carried out using 8-12 week old females obtained from The Jackson Laboratory.
2. Experiments in Muc2-/- mice used a mix of 8-12 week old male and female mice bred at Duke's Division of Laboratory Animal Resources Breeding Core.
3. Experiments in ASF mice used a mix of 8-12 week old male and female mice. Breeding and colonization experiments were run by the Barton lab at the University of California Berkeley.
4. Experiments in germ-free mice used a mix of 8-12 week old male and female mice. Breeding and colonization experiments were run through Duke's Division of Laboratory Animal Resources Gnotobiotic Core.

Wild animals

This study did not involve wild animals.

Field-collected samples

No field samples were collected.

Ethics oversight

All experiments were performed in accordance with the Institutional Animal Care and Use Committee guidelines at Duke University and the University of California Berkeley.

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

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Wild type and/or mutant Akkermansia were harvested by centrifugation, washed with PBS, and fixed with 4% paraformaldehyde. The cells were then washed twice with PBS and blocked with PBS with 2% BSA for 10 min. Rabbit anti-Akkermansia anti-sera was added at a 1:100 dilution and the cells were incubated overnight at 4C. The primary antibody was removed by washing with PBS and a 1:1000 dilution of goat anti-Rabbit IgG Alexa Flour-647 was added for 1 h at room temperature. After a final wash, the cells were suspended in PBS with 2% BSA and immediately run on the flow cytometer.

Instrument

BD FACSCanto II flow cytometer (338962)

Software	Flow cytometry data was acquired using FACSDiva software (v9, BD) and analyzed with FlowJo (v10)
Cell population abundance	For each sample, 50 000 events were acquired. For analysis of labeled mucin uptake by wild-type Akkermansia and transposon mutants, a range of 6200 to 7600 Akkermansia-647 positive cells were analyzed for each strain. For the time course of mucin uptake, each sample had between 6200 to 9400 Akkermansia-647 positive cells.
Gating strategy	The flow experiments involve sorting bacterial cells, which due to their small size, can appear similar to other debris. Therefore, the samples were first gated on the forward and side scatter, and subsequently gated on a population that was stained with anti-Akkermansia anti-sera (detected with goat anti-rabbit Alexa Fluor-647). The population that stained positive for Akkermansia was then used to determine the mean fluorescent intensity at 488nm, indicative of fluorescein -mucin. Unstained and singly stained controls were included in all experiments.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.