HHS Public Access

Author manuscript

*Nat Microbiol*. Author manuscript; available in PMC 2020 September 09.

Published in final edited form as:

*Nat Microbiol*. 2020 May ; 5(5): 746–756. doi:10.1038/s41564-020-0683-3.

# Spatially distinct physiology of *Bacteroides fragilis* within the proximal colon of gnotobiotic mice

**Gregory P. Donaldson**[1,†,*], **Wen-Chi Chou**[2,†], **Abigail L. Manson**[2], **Peter Rogov**[2], **Thomas Abeel**[2,3], **James Bochicchio**[2], **Dawn Ciulla**[2], **Alexandre Melnikov**[2], **Peter B. Ernst**[4], **Hiutung Chu**[4], **Georgia Giannoukos**[2], **Ashlee M. Earl**[2,*], **Sarkis K. Mazmanian**[1,*]

[1]Divison of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA [2]Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, MA, USA [3]Delft Bioinformatics Lab, Delft University of Technology, Delft, The Netherlands [4]Department of Pathology, University of California, San Diego, CA, USA

## SUMMARY

A complex microbiota inhabits various microenvironments of the gut, with some symbiotic bacteria having evolved traits to invade the epithelial mucus layer and reside deep within intestinal tissue of animals. Whether these distinct bacterial communities across gut biogeographies exhibit divergent behaviors remains largely unknown. Global transcriptomic analysis to investigate microbial physiology in specific mucosal niches has been hampered technically by overabundance of host RNA. Herein, we employed hybrid selection RNA sequencing (hsRNA-Seq) to enable detailed spatial transcriptomic profiling of a prominent human commensal as it colonizes the colonic lumen, mucus or epithelial tissue of mice. Compared to conventional RNA-Seq, hsRNA-Seq increased reads mapping to the *Bacteroides fragilis* genome by 48- and 154-fold in mucus and tissue, respectively, allowing for high fidelity comparisons across biogeographic sites. Near the epithelium, *B. fragilis* up-regulated numerous genes involved in protein synthesis, indicating that bacteria inhabiting the mucosal niche are metabolically active. Further, a specific sulfatase (BF3086) and glycosyl hydrolase (BF3134) were highly induced in mucus and tissue compared to



*corresponding authors.
†equal contribution

AUTHOR CONTRIBUTIONS

G.P.D. and S.K.M. conceived the study. G.P.D., W.C.C., G.G., A.M.E., and S.K.M. designed the study. G.P.D. prepared samples for sequencing and performed mouse colonization and microbiological experiments. D.C., P.R., J.B., A.M., G.G. performed hybrid capture and sequencing experiments. W.C.C., A.L.M. and T.A. performed computational analysis. H.C. performed the colitis model and flow cytometry. P.B.E. scored sections for histology. G.P.D., W.C.C. and A.L.M. made the figures. A.M.E. and S.K.M. supervised the work. G.P.D., W.C.C., A.L.M., A.M.E., and S.K.M. wrote the paper. All authors provided input on the paper.

COMPETING INTERESTS

The authors declare no competing interests.

DATA AVAILABILITY

RNA-Seq and hsRNA-Seq data have been deposited in the NCBI Short Read Archive under a project accession number, PRJNA438372. The *B. fragilis* NCTC9343 genome used for mapping is available at Genbank GCA_000025985.1. All other source data are provided with the paper.

CODE AVAILABILITY

Analysis code is available at https://github.com/wenchichou/bugInHost.

bacteria in the lumen. In-frame deletion of these genes impaired *in vitro* growth on mucus as a carbon source, as well as mucosal colonization of mice. Mutants in either *B. fragilis* gene displayed a fitness defect in competing for colonization against bacterial challenge, revealing the importance of site-specific gene expression for robust host-microbial symbiosis. As a versatile tool, hsRNA-Seq can be deployed to explore the *in vivo* spatial physiology of numerous bacterial pathogens or commensals.

---

The mammalian gastrointestinal (GI) tract hosts an ecosystem of bacteria, protists, fungi, and viruses, which is assembled following birth to establish lifelong symbiosis. The microbial community is important for host digestion[1] and protection from enteric infection[2] (i.e., colonization resistance). Gut bacteria inhabit a variety of distinct microhabitats along the longitudinal and cross-sectional axes of the intestines[3]. Along the entire length of the gut, mucus provides a physical barrier that partitions the gut lumen from the intestinal surface[4]. Accordingly, the mucus and lumen compartments of the gut house distinct bacterial communities[5–8]. Some bacteria also directly associate with the gut epithelial surface, such as segmented filamentous bacteria[9,10], adherent *Lactobacillus*[11], and a diverse community of crypt-resident bacteria[12–14]. Studying the relevance of spatial microbiome structure is a challenge due to the dynamic nature of the gut, complexity of microbial communities, and difficulty in accessing gut tissues. Most current knowledge about the microbiome derives from DNA-based sequence profiling of low-volume, homogenized fecal samples, which is unlikely to capture information about bacteria closely associated with the host. While imaging can provide insights into microbial biogeography of the gut[15,16], the functional relevance of localization is often difficult to infer. Approaches that can determine the physiology of individual bacterial species, spatially within specific gut microenvironments, would offer unprecedented insights into host-microbial interactions.

Attempts to investigate bacterial transcriptomes in host tissues or mucus are complicated by the paucity of bacterial RNA present at these sites compared to host RNA. Existing dual RNA-Seq methods[17,18] address this issue by either sequencing combined host and bacterial RNA to extremely high depths, assisted by host ribosomal RNA depletion[19,20], and/or enriching for bacterial cells, prior to lysis, using methods such as fluorescence activated cell sorting (FACS)[21], density gradient centrifugation[5], or laser capture microdissection[22]. Deep sequencing for *in silico* separation is impractical for confident assessment of bacterial transcriptomes in host niches where the ratios of bacteria to host RNA are exceedingly small[21]. Bacterial cell enrichment prior to lysis can overcome the abundance limitation, but this involves greater processing time and conditions that induce bacterial stress responses and biases in RNA expression and degradation. As typical bacterial mRNAs have *in vivo* half-lives of just several minutes[23], RNA loss may be avoided by an alternative strategy: to separate bacterial transcripts from host transcripts at the molecular level, following rapid nucleic acid extraction. Towards this objective, we used hybrid selection to enrich bacterial reads from total RNA pools isolated from intact and unprocessed host tissues, then used RNA-Seq to measure the enriched transcriptome, a method we term ‘hybrid selection RNA sequencing’ (hsRNA-Seq). Following extraction of total RNA from various biogeographies in the gut of mice, reverse transcribed cDNA was enriched for bacterial sequences using biotinylated probes complementary to the entire genome of a prominent human gut

commensal bacteria (Fig. 1a). After probe hybridization to bacterial cDNA and capture using biotin, host cDNA was washed away, allowing elution of enriched bacterial cDNA for sequencing. Originally developed for re-sequencing targeted regions of the human genome[24], hybrid selection has been adapted to enrich eukaryotic pathogen[25] and RNA virus[26] genomes in *ex vivo* clinical DNA samples dominated by human genetic material. Herein, we validate and assess hsRNA-Seq to uncover *in vivo* interactions between a bacterial species of the gut microbiome and its mammalian host, with defined spatial resolution. We discover distinct transcriptomes of the same microorganism depending on its biogeography, and further identify and validate specific genes that aid in mucosal association and colonization resistance. This technique can potentially be applied to any sequenced bacterial species to explore its physiology during symbiotic colonization or pathogen infection in the gut or other body sites.

## RESULTS

### hsRNA-Seq enables spatial bacterial transcriptomics during commensal colonization

To explore the genetic underpinnings of host-microbial associations *in vivo*, we used a simplified model system of mice mono-colonized with *Bacteroides fragilis* (Fig. 1a), a symbiont found in close association with the epithelial surface, especially in the proximal (ascending) colon (Extended Data Figure 1a,b)[7,27–29]. The proximal colon was dissected and the lumen, mucus, and tissue were carefully separated (Fig. 1a, Extended Data Figure 1c), with total RNA isolated using a single standardized protocol. We performed RNA-Seq with or without hybrid selection to test for enrichment of bacterial transcripts relative to mouse RNA (Supplementary Table 1). Without hybrid selection, bacterial RNA represented 50% of the total RNA in the lumen, but only 0.6% and 0.1% of total RNA in mucus and tissue samples, respectively (Fig. 1b). After hybrid selection, we observed a reduction in unaligned reads and reads mapping to the mouse genome, with a corresponding increase in reads mapping to the *B. fragilis* genome (Supplementary Table 1); the percentage of total reads mapping to the *B. fragilis* genome increased 48 and 154-fold in the mucus and tissue, respectively (Fig. 1b). This enrichment resulted in a dramatic improvement in *B. fragilis* gene coverage in the mucus and tissue samples (Fig. 1c).

To assess the fidelity and performance of hybrid selection, we compared hybrid selected and non-hybrid selected transcriptomes from the same samples from mono-colonized animals. Within sample sites, normalized gene expression levels with and without hybrid selection were highly correlated when analyzed in bulk (Fig. 1d) or within individual mice (Extended Data Figure 2, Supplementary Table 2), indicating that the method did not globally skew the bacterial transcriptome. Our method also allows simultaneous study of the host transcriptome, as it remained largely unbiased after hybrid selection (Extended Data Figure 3), though we do not extend this analysis herein. In the lumen, which serves as a control sample that yields a quality bacterial transcriptome without hybrid selection, correlations between the same sample with and without hybrid selection were as good or better than correlations between biological replicates (Supplementary Table 2), indicating that the process did not alter the bacterial transcriptome. Including all sample sites, only 20 of the over 5,000 genes in the *B. fragilis* genome did not enrich as expected following hybrid

selection. These were mostly short (< 200 nucleotides) noncoding RNAs, including some tRNAs and 5S ribosomal RNAs (Supplementary Table 3). But the majority of tRNAs were not skewed, as 85% of them were enriched as expected (Extended Data Figure 4). These results demonstrate that hsRNA-Seq provides a significantly accurate transcriptome.

hsRNA-Seq enabled measurement of otherwise undetectable transcripts. Limited bacterial transcript coverage in mucus and tissue samples without hybrid selection led to a large number of genes with near-zero expression levels (555 in mucus and 1034 in tissue). Applying hybrid selection, the number of genes with near-zero expression levels decreased to 159 in mucus and 299 in tissue. Hybrid selection also substantially improved the correlation of gene expression levels between lumen and mucus, as well as lumen and tissue (Extended Data Figure 5). Importantly, hybrid selection dramatically increased the number of genes identified as differentially expressed between different microenvironments (Fig. 1e). We found that a conventional RNA-Seq method yielded several false positives which were ruled out when their low read counts were increased 20 to 100-fold by hybrid selection (Fig. 1e, Supplementary Tables 4 and 5). Thus, hsRNA-Seq enriched bacterial transcripts from host RNA without skewing the transcriptome, facilitating spatial comparisons of *B. fragilis* between different gut biogeographies.

## Differentially-expressed *B. fragilis* genes between gut microenvironments

hsRNA-Seq revealed a number of genes that were differentially expressed by *B. fragilis* based on its localization in the lumen, mucus, or epithelial tissue (Fig. 2, Supplementary Table 6). Compared to the lumen transcriptome, 26 *B. fragilis* genes were significantly up-regulated and 42 were down-regulated in the mucus, while in the tissue 52 genes were up-regulated and 47 were down-regulated (Fig. 2a, Supplementary Table 6). In total, we identified 130 genes spread across the genome that were differentially expressed based on localization, 37 of which changed in the same direction for both mucus and tissue (Fig. 2a). We confirmed the differential expression patterns of six genes of interest using quantitative real-time PCR as an independent method to validate the transcriptomics results (Fig. 2g–l). Pathway analyses to find groups of genes with related metabolic functions indicated that the glycosyl transferase group 1 Pfam[30] was upregulated in the lumen (Supplementary Table 7). As annotations of the *B. fragilis* NCTC9343 genome are limited, we also included additional functional annotations from multiple databases (listed with original and modern locus IDs in Supplementary Table 8). The STRING database[31] grouped differentially expressed genes based on multiple parameters including co-occurrence, co-expression, and related known functions (Fig. 2b, d).

In lumen samples which profiled *B. fragilis* gene expression in the fecal stream, we surprisingly observed patterns of stress response (Fig. 2b, c). A universal stress gene, *uspA* (BF2495), was more highly expressed in the lumen than mucus. An adenine-specific DNA methyltransferase (BF1252 or BF9343_RS05815) was the most highly expressed gene in the lumen compartment compared to tissue (Fig. 2c, g). DNA methylation in bacteria is widespread and important for genome protection[32]. Two bacterial histone-like proteins (BF3379 / BF9343_RS16290 and BF4220 / BF9343_RS20550) were also more highly expressed in the lumen (Fig. 2b, c, h), which are also involved in genome protection[33].

These patterns of gene expression indicate that environmental stress may cause a prioritization of genome protection for *B. fragilis* in the lumen, suggesting mucus or tissue microenvironments as more preferred habitats.

Our analysis also revealed the expression of a lumen-specific capsular polysaccharide locus. *B. fragilis* is capable of synthesizing at least eight distinct capsular polysaccharides that coat the surface of the bacterium and are important for gut colonization[29,34–36]. One of the most highly expressed genes in the lumen compared to mucus and tissue was the Polysaccharide G (PSG) flippase (BF0737 or BF9343_RS03450) (Fig. 2b, c, i), the enzyme responsible for transporting polysaccharides synthesized in the periplasm to the outer leaflet of the outer membrane. Four other genes in the PSG biosynthesis locus were also more highly expressed in the lumen (Fig. 2b, c). No genes in the seven loci for other capsular polysaccharide biosynthesis were differentially expressed, suggesting a unique, lumen-specific role for PSG in the physiology of *B. fragilis*. We speculate that the induction of a stress response, genome protection and PSG expression observed in the lumen may collectively prepare *B. fragilis* in the fecal stream for survival outside of the gut upon shedding.

Genes more highly expressed in mucus and tissue samples provide a glimpse into the biology of *B. fragilis* during close association with the intestinal surface. Though an obligate anaerobe, *B. fragilis* is well-equipped to contend with reactive oxygen species[37], which may allow it to associate with the oxygenic epithelial surface of the gut[8]. Indeed, both subunits of alkyl hydroperoxide reductase (a reactive oxygen species resistance enzyme), *ahpC* (BF1210 or BF9343_RS05610) and *ahpF* (BF1209 or BF9343_RS05605)[38], were induced in mucus and tissue samples (Fig. 2d, e, j). The previously defined transcriptomic response of *B. fragilis* to oxygen during growth in laboratory media includes suppression of dozens of genes involved in protein synthesis[39]. In contrast, we find evidence of increased protein synthesis by *B. fragilis* in the mucus and tissue during colonization of mice. 14 tRNAs were up-regulated in mucus and 26 in tissue compared to the lumen, but not a single tRNA was more highly expressed in the lumen (Fig. 2f, Supplementary Table 6). Ribonuclease P (RNase P, BF0076 or BF9343_RS00335), the ribozyme that cleaves the precursor RNA on tRNAs to form mature tRNAs, was also up-regulated in both the mucus and tissue (Fig. 2f). Additionally, a number of ribosome-related genes were up-regulated in the tissue, including the 30S and 50S subunits (Fig. 2d). Importantly, none of genes were outliers that appeared to be affected by the hybrid selection process (Supplementary Table 3). Though we did not measure protein synthesis directly, these data suggest that mucus-associated bacteria are metabolically active. We speculate that decreased protein synthesis in the lumen may reflect a relatively nutrient-poor environment, as a starvation response involving the shutdown of protein synthesis in the lumen by *B. thetaiotaomicron* was recently found to be important for colonization[40]. In contrast, *B. fragilis* tolerates the oxygenic stress of the epithelium through the deployment of alkyl-hydroperoxide reductase while expanding its capacity for protein synthesis in the mucus and tissue.

We observed spatially differentiated expression patterns in a small, defined set of individual genes, which we chose to further explore in the remainder of this study. The two most up-regulated genes in both mucus and tissue were BF3086 (or BF9343_RS14795) (Fig. 2e, k), one of 17 annotated sulfatases encoded in the *B. fragilis* genome (Fig. 2e), and BF3134 (or

BF9343_RS15035), annotated as a glycosyl hydrolase (Fig. 2e, l). Functional annotation and structural modeling indicated that BF3134 likely encodes a cyclo-malto-dextrinase, a member of the glycosyl hydrolase family 13 (Extended Data Figure 6a,b). BF3086 likely encodes an acetylglucosamine-6-sulfatase (Extended Data Figure 6c,d), an enzyme previously implicated in mucin desulfation, and important for mucosal glycan foraging by *B. thetaiotaomicron*[41]. Together, the structural modelling and functional predictions indicate that the BF3086 gene product is likely utilized for the catabolism of mucus glycans, which have diverse structures and are often sulfated[42]. BF3086 and BF3134 are highly conserved in *B. fragilis* genomes (Extended Data Figure 7a). Average pairwise nucleotide identities between 13 *B. fragilis* homologues were 99.5% (BF3086) and 99.7% (BF3134), while the average pairwise identity with other *Bacteroides* and *Parabacteroides* homologues was 69% and 66%, respectively (Extended Data Figure 7a). We also observed a common potential regulatory feature, a 36 bp motif (Extended Data Figure 7b) upstream of both BF3086 and BF3134 that was conserved in 13 other *B. fragilis* genomes, suggesting that both genes are part of the same regulon. Together, the unique spatial expression patterns of these genes, their conservation across multiple strains of *B. fragilis*, and their consistent expression profile across biological replicates (Extended Data Figure 8a) motivated us to investigate the function of BF3086 and BF3134 during colonization of mice.

## Discovery of candidate mucosal colonization factors in *B. fragilis*

To explore potential mechanisms in *B. fragilis* colonization of the gut, we generated in-frame deletion mutants in the sulfatase (ΔBF3086) and glycosyl hydrolase (ΔBF3134) genes. When initially assessing growth in minimal media with defined carbon sources, mutant strains exhibited similar growth in minimal glucose (Fig. 3a), several dietary polysaccharides (Extended Data Figure 8b–d), and pig stomach mucins (Extended Data Figure 8e) compared to their parent wild-type *B. fragilis*. We modeled mucosal growth using mucus harvested from the colons of germ-free mice, and discovered that wild-type *B. fragilis* grew robustly in culture, whereas both mutants exhibited growth defects, as measured by colony-forming units (CFUs) / ml (Fig. 3b). While wild-type and complemented *B. fragilis* strains reached stable stationary phase, both ΔBF3086 and ΔBF3134 entered a death phase after logarithmic growth on mucus (Fig. 3b). Lack of persistence in stationary phase may reflect inability of mutant strains to utilize less accessible or lower abundance mucosal glycans that remain after log-phase growth, because in minimal glucose all strains entered a rapid death phase (Fig. 3a). Ectopic expression of BF3086 or BF3134 under the control of their native promoters on a plasmid to complement the respective mutants fully recovered mucus growth phenotypes similar to wild-type levels (Fig. 3b), ruling out polar effects of the gene deletions.

To determine whether BF3086 or BF3134 play a role in *B. fragilis* colonization, groups of germ-free mice were associated with wild-type or mutant bacteria and all strains reached the same stable CFU level in feces after 4 weeks of mono-colonization (Fig. 3c). Spatially, although there were similar bacterial numbers in the colonic lumen (Fig. 3d), both ΔBF3086 and ΔBF3134 were reduced in colonization fitness of the colonic mucus (Fig. 3e). This defect was fully recovered by trans-complementation of BF3134, with a trending recovery for BF3086 (Fig. 3e). Impaired mucosal colonization was not due to effects on the

expression of genes previously shown to be involved in mucosal colonization: the commensal colonization factors (*ccf*) or capsular polysaccharides B and C (PSB and PSC) (Extended Data Figure 8f–h)[29]. Collectively, these data reveal genes employed by *B. fragilis* for growth on mucus *in vitro* and robust mucosal colonization *in vivo*.

### *B. fragilis* genes for mucosal association enhance competitive colonization

As a stringent test for *in vivo* function, we determined whether BF3086 or BF3134 contribute to *B. fragilis* fitness in competitive colonization studies. When mice were orally gavaged with an equal mixture of wild-type and mutant *B. fragilis*, ΔBF3134 steadily decreased as a proportion of the total population (Fig. 4a). In contrast, ΔBF3086 did not show a competitive disadvantage in this direct competition model (Fig. 4b), possibly because the de-sulfating activity of wild-type bacteria liberated enough mucosal glycans to also support the co-colonizing mutants.

Other *B. fragilis* mutants exhibiting defects in mucosal colonization are unable to exclude competitors of the same species[28,29]. In horizontal transmission assays between pairs of mice, animals colonized for 4 weeks with wild-type *B. fragilis* (WT initial) displayed robust colonization resistance, while mice initially colonized for 4 weeks with either ΔBF3086 or ΔBF3134 were substantially invaded by wild-type bacteria following co-housing (Fig. 4c and 4d). This outcome is consistent with a previously proposed model whereby saturation of a mucosal niche prevents invasion by a foreign strain[28,29]. Accordingly, the model predicts long-term colonization by a single strain of *B. fragilis*, which remarkably has been observed in longitudinal microbiome profiling studies in humans[43,44].

*B. fragilis* has been shown to protect from symptoms and gut pathology in multiple preclinical models of colitis, via the induction of anti-inflammatory interleukin (IL)-10 production by Foxp3+ regulatory T cells ($T_{REGS}$)[45,46]. This effect requires delivery to intestinal dendritic cells of *B. fragilis* capsular polysaccharide A (PSA)[46], which was not differentially expressed across microenvironments (Extended Data Figure 10a). *B. fragilis* strains were tested in the Dinitrobenzene sulfonic acid (DNBS) model of experimentally-induced colitis. Compared to mice mono-colonized with wild-type *B. fragilis* or BF3086 mutants, those colonized with ΔBF3134 exhibited worsened weight loss (Extended Data Figure 9a), gross inflammation of the tissue (Extended Data Figure 9b), shorter colon length (Extended Data Figure 9c), and higher pathology scoring (Extended Data Figure 9d). Following sham treatment, animal weight and colon length were similar in mice colonized with wild-type, ΔBF3086, or ΔBF3134 strains (Extended Data Figure 10b, c). Both mutants expressed a critical gene for PSA production at levels equal to or higher than wild-type bacteria (Extended Data Figure 9e), and promoted the development of similar amounts of pro-inflammatory IL-17 producing $T_{REGS}$ (Extended Data Figure 9f, Extended Data Figure 10d, e), but significantly less anti-inflammatory IL-10 producing Foxp3+ $T_{REGS}$ (Extended Data Figure 9g, Extended Data Figure 10d, e). We were unable to test whether trans-complementation of ΔBF3134 restored protection from colitis. These data suggest that expression of BF3134 may assist in PSA delivery to (or sensing by) the animal host[45–47] by positioning the bacteria in closer proximity to surveilling gut immune cells.

To expand these findings beyond gnotobiotic mice, which have poorly-developed mucus[48], we gavaged 8 week-old excluded flora (EF) mice, which have a complex but controlled microbiota, with $10^9$ CFU of *B. fragilis* strains. Two weeks later, we assessed colonization. In the feces (Fig. 4e) and lumen of the proximal colon (Fig. 4f), *B. fragilis* ΔBF3134 colonized at a higher level than ΔBF3086, while neither differed from wild-type bacteria. Importantly, both mutants colonized the proximal colon mucus at significantly lower levels than wild-type *B. fragilis* (Fig. 4g), indicating that mucosal colonization defects observed in gnotobiotic mice are maintained in the context of a complex microbial community.

## DISCUSSION

Despite numerous sequence-based studies exploring the structure of fecal microbial communities across geographies, diets, diseases, and lifestyles, investigations into fundamental bacterial functions have largely not accounted for the spatial organization of the microbiome within the gut[3]. This is in large part due to technical limitations. Here we show that hsRNA-Seq can be used to profile global RNA expression for *B. fragilis* in samples that contain overwhelming amounts of host RNA at various mucosal sites without skewing transcriptomes, providing a glimpse into bacterial physiologies with spatial resolution. Compared to previously published dual RNA-Seq studies of host-associated bacteria[49–52], the proportions of bacterial to host RNA was lower in this study, by an order of magnitude on average. To measure the transcriptomes reported here without using hybrid selection would require at least 400 billion additional base pairs of sequencing per sample. Thus, hsRNA-Seq provides an alternative to ultra-deep sequencing that is preferable when studying host-associated niches where the bacterial load is comparably low. A small number of RNAs did not enrich as expected, most of which were short (<200 nt) noncoding RNAs, therefore a degree of caution is warranted when interpreting the short transcript data. However, hsRNA-Seq requires only the minimal amount of RNA needed to create a cDNA sequencing library. While we used 2 μg of cDNA from a pool of libraries as the input for hybrid selection, lower amounts would likely work as well. Given that hsRNA-Seq effectively captured cDNAs across a wide dynamic range and that adjustments could be made to the protocol to accommodate lower input (e.g. increasing amount of bait, reducing hybridization volume, or increasing hybridization time), we expect that input amounts could be reduced substantially, giving hsRNA-Seq a further advantage over existing methods.

hsRNA-Seq could be extended to distinguish transcriptomes from individual species within a complex community, potentially enabling the study of low-abundance species behaviors that cannot be resolved from meta-transcriptomes. The level of enrichment of the targeted sequences from polymicrobial communities can potentially be increased with multiple rounds of the hybrid selection procedure. Homologous genes in different organisms present a challenge for hybrid selection, though synthetic bait designs that take into account sequence similarities among cohabiting microbes[53] could help overcome this potential issue. hsRNA-Seq could be applied to other tissues as well to explore microbiomes of the oral cavity, lungs, skin and other interfaces of host-microbial symbiosis.

The spatial gene expression patterns revealed herein provide several insights into the biology of the model gut symbiont, *B. fragilis*. Previous transcriptomics studies with members of the

prominent *Bacteroides* genus have revealed sets of genes which are upregulated during *in vivo* colonization compared to growth in laboratory culture[54,55], during growth in purified mucin glycans[56], or in animals exposed to different diets[54,57]. Gene expression profiles of *B. thetaiotaomicron* were investigated in the lumen and mucus of the colon using a density-gradient centrifugation method to separate bacteria from host cells, and confirmed the upregulation of polysaccharide utilization loci in the mucus layer that were previously shown to be induced by mucins *in vitro*[5]. We found that in the case of *B. fragilis*, bacterial populations proximal to tissue display increased metabolic activity, perhaps indicative of replication and/or production of microbial molecules that interact with the host. In some cases, genes previously shown to be induced by mucins *in vitro* and highly expressed *in vivo* compared to growth in laboratory culture (such as *don*[55] and *ccf*[28,58]) did not exhibit spatial expression patterns, either because the activating signal for these genes is present in the lumen or they are suppressed in culture. During colon mucus and tissue colonization, we discovered that *B. fragilis* up-regulates a set of candidate colonization factors, including genes encoding a sulfatase (BF3086) and a glycosyl hydrolase (BF3134). BF3086 was previously found to be induced during growth in mucosal O-glycans[28,58], implicating its importance to host mucus degradation. Expression of several other sulfatases were also induced in the previous *in vitro* model, while the glycosyl hydrolase BF3134 was not induced, highlighting the importance of *in vivo* models for study of gut bacterial physiology. We observed that *in vivo* mucosal association enabled by these gene products is beneficial to *B. fragilis* through increased colonization robustness, consistent with previous reports[28,29]. While association with the intestinal surface may be perilous for other bacteria, *B. fragilis* appears well adapted to thrive within the mucosal environment as it exhibits lower stress responses, increases protein synthesis while tolerating oxygen, and performs poorly in competitive colonization assays without genes enabling robust host association. We conclude that *B. fragilis* has evolved to dynamically modulate its behavior at distinct sites within the gut, deploying a specific genetic program during intimate association with its host. This concept and the tools to determine spatial transcriptomes in the gut can be applied to study the many bacterial species, both commensal and pathogenic, that associate with mucosal surfaces of animals[3].

## METHODS

### Bacterial strains and media

*B. fragilis* NCTC9343 was cultured in Brain Heart Infusion (BD) with 5 μg/ml hemin (Frontier Scientific) and 5 μg/ml vitamin K1 (Sigma) or in a defined minimal media[59,60] in an 80% nitrogen, 10% carbon dioxide, and 10% hydrogen atmosphere. For growth in mouse mucus, crude mucus was isolated as described below (**Separation of colon lumen, mucus, and tissue**) from the entire colon of germ-free mice into defined minimal media at a concentration of one whole colon of mucus per 5 ml of final media. Where appropriate, 200 μg/ml gentamicin, 10 μg/ml erythromycin, 2 μg/ml tetracycline, or 10 μg/ml chloramphenicol were used in selective media. For competitive colonization and horizontal transmission, marker plasmids pFD340-Chlor (providing resistance to erythromycin and chloramphenicol) or pFD340-Tet (providing resistance to erythromycin and tetracycline) were used to distinguish two strains, as described before[28]. Scar-less, in-frame deletions of

BF3086 (1305 bp deleted) and BF3134 (1686 bp deleted) were made using allelic exchange with the pKNOCK suicide vector[61] using a previously described method[28]. Briefly, flanking regions were cloned into the pKNOCK backbone, which was then conjugated into *B. fragilis* using erythromycin selection. Co-integrates were passaged without erythromycin until they lost resistance (following a second recombination event), and these colonies were screened for loss of the targeted gene using PCR. The mutants were complemented by expressing the full-length genes under control of their native promoters on the pFD340 shuttle vector (all primers are listed in Supplementary Table 9).

## Mice and colonization experiments

All animal experiments were performed in accordance with ethical guidelines in the NIH Guide for the Care and Use of Animals and protocols were approved by the Caltech Institutional Animal Care and Use Committee (protocol #1550). Swiss Webster and C57BL/6 mice from Taconic Farms were C-section re-derived germ-free and bred in flexible film isolators. Sample size was based on previous studies that used similar methods[28,29,46]. Blinding was used for histology scoring only. Mice were randomly assigned to groups. For most experiments, germ-free mice were transferred at 6–8 weeks of age to sterile micro-isolator cages with autoclaved food (LabDiet 5010) and water. Mice were mono-colonized by a single oral gavage with $10^8$ CFU of *B. fragilis* in 100 μl of HBSS with 1.5% sodium bicarbonate (or a 1:1 mix of $10^8$ CFU each of two strains, for competitive colonization). Mice were maintained mono-colonized for 4 weeks prior to subsequent experimentation. Colonization was monitored in fresh fecal samples which were weighed, mashed, and vortexed in 1 ml BHI and diluted for plating CFU. Only female mice were used for most experiments. Half male and half female mice were used for the fecal, colon lumen, and mucus colonization assays in Fig. 3c–e. For mucosal CFU plating, mucus was isolated as described below (**Separation of colon lumen, mucus, and tissue**). For competitive colonization and horizontal transmission assays, water was supplemented with 100 μg/ml gentamicin (*Bacteroides* are naturally resistant to gentamicin) and 10 μg/ml erythromycin for plasmid selection. Horizontal transmission was assayed as previously[28], by co-housing mice for 4 hours in new sterile cages and then separating mice into single-housing.

## Separation of colon lumen, mucus, and tissue

First, the proximal colon was dissected. A 1 cm segment of the proximal (ascending) colon, starting at the cecal junction, was taken for sampling. The segment was opened longitudinally and ~100 mg of lumen content was collected for the “lumen” sample. The rest of the lumen content was removed. The tissue was washed by vigorously shaking with forceps in a dish of sterile HBSS for 20 seconds (changing forceps grip-point 3 times during the process). The washed tissue was carefully observed to ensure no lumen content remained. Tissue was dabbed in a dry sterile petri dish to remove excess buffer. Mucus was scraped from the surface of the tissue using a sterile plastic 1.8 cm cell scraper (BD Falcon) and collected as the “mucus” sample. The remaining tissue was collected as the “tissue” sample. The resulting three samples appeared as in Fig. 1a. Animals were sacrificed one at a time and samples were quickly processed through the bead-beating lysis step of the RNA purification protocol below before starting the next animal (< 10 minutes from sacrifice to lysis). Lysed samples were kept at 4°C until all samples were collected.

## RNA purification

All samples were subjected to the same RNA preparation protocol, which was the only method we found to reproducibly provide high-quality and high-yield RNA from lumen, mucus, and tissue. Samples were immediately lysed by bead-beating for 1 minute in 2 ml lysing matrix B tubes (MP Biomedicals) with 500 µl buffer (0.2 M NaCl and 20 mM EDTA), 210 µl 20% SDS (Ambion AM9820), and 500 µl phenol, chloroform, isoamyl alcohol mixture (Ambion AM9720). After centrifugation at 13,000 rpm for 3 minutes at 4°C, the aqueous phase was added to a new microcentrifuge tube with 500 µl of the phenol, chloroform, isoamyl alcohol mixture, and mixed by inversion. The centrifugation and isolation of aqueous phase was repeated, yielding approximately 300 µl of aqueous phase. Next, 30 µl of 3 M sodium acetate (Ambion AM9740) and 300 µl of −20 C 100% ethanol was mixed in by inversion. Samples were left on ice for 20 minutes, then centrifuged at 13,000 rpm for 20 minutes at 4°C. The supernatant was decanted and 500 µl of −20°C 70% ethanol was added and vortexed before centrifuging at 13,000 rpm for 5 minutes at 4°C. The supernatant was decanted and the tubes were inverted and air-dried for 5 minutes after wiping the lips of tubes dry on Kimwipes. 100 µl of water was added to the dried pellets, which were then frozen at −20°C. The next day, samples were thawed and 350 µl RLT buffer (Qiagen) with 1% 2-mercaptoethanol was added. Samples were vortexed for 20 minutes, at which point the pellets were completely dissolved. The samples were then loaded on Qiagen RNeasy mini columns and purified according to the manufacturer's instructions. Nucleic acids were eluted in 50 µl of water and quantified using a NanoDrop. Up to 10 µg of the nucleic acids were treated with 4 µl Turbo DNase (Ambion AM2238) in a 60 µl reaction at 37°C for 1 hour. 40 µl of water and 350 µl buffer RLT were added and the samples were loaded onto a second Qiagen RNeasy mini column. The second column purification was performed according to the manufacturer's instructions including the Qiagen on-column DNase digest. The final total RNA was eluted in 50 µl water.

## Preparation of whole-transcriptome fragment libraries (pond) for hybrid selection

Libraries for hybrid selection RNAseq were created as previously described[62]. The isolated RNA was first quantified and qualified by Qubit and Agilent Bioanalyzer. The RNA was then fragmented by FastAP (Thermo Scientific) and linked to barcoded adapters. The fragmented and barcoded RNA was pooled by sample type (lumen, mucus, tissue) to perform ribosomal RNA depletion using Ribo-Zero Magnetic Gold Epidemiology Kit (Epicentre/Illumina). The cDNA was generated from the RNA through template-switching RT-PCR. After Exonuclease I treatment and PCR enrichment, the cDNA was used for hybrid selection.

## Hybrid selection probes (bait) construction

Whole genome bait (WGB) was generated at the Broad Institute. For input, 3 µg of *B. fragilis* NCTC 9343 DNA was sheared for 4 minutes on a Covaris E210 instrument set to duty cycle 5, intensity 5 and 200 cycles per burst. The mode of the resulting fragment size distribution was 250 bp. End repair, addition of a 3'-A, adaptor ligation and reaction clean-up followed the Illumina's genomic DNA sample preparation kit protocol, except that the adapter consisted of oligonucleotides 5'-

TGTAACATCACAGCATCACCGCCATCAGTCxT-3' ('x' refers to an exonuclease I-resistant phosphorothioate linkage) and 5'-[PHOS]GACTGATGGCGCACTACGACACTACAATGT-3'. The ligation products were cleaned up (Qiagen), amplified by 8 to 12 cycles of PCR on an ABI GeneAmp 9700 thermocycler in Phusion High-Fidelity PCR master mix with HF buffer (NEB Ipswich, Massachusetts, United States) using PCR forward primer 5'-CGCTCAGCGGCCGCAGCATCACCGCCATCAGT-3' and reverse primer 5'-CGCTCAGCGGCCGCGTCGTAGTGCGCCATCAGT-3' (ABI Carlsbad, California, United States). Initial denaturation was 30 s at 98°C. Each cycle was 10 s at 98°C, 30 s at 50°C and 30 s at 68°C. PCR products were size-selected on a 4% NuSieve 3:1 agarose gel followed by QIAquick gel extraction. To add a T7 promoter, size-selected PCR products were re-amplified as above using the forward primer 5'-GGATTCTAATACGACTCACTATAGGGCGCTCAGCGGCCGCAGCATCACCGCCATCAGT-3'. Qiagen-purified PCR product was used as template for whole genome biotinylated RNA bait preparation with the MEGAshortscript T7 kit (Ambion)[24,25].

## Hybrid selection

Using the designed baits, hybridization was conducted at 65°C for 66 h with 2 μg of 'pond' libraries carrying standard or indexed Illumina paired-end adapter sequences and 500 ng of bait in a volume of 30 μl. After hybridization, captured DNA was pulled down using streptavidin Dynabeads (Invitrogen Carlsbad, California, United States). Beads were washed once at room temperature for 15 minutes with 0.5 ml 1 × SSC/0.1% SDS, followed by three 10-minute washes at 65°C with 0.5 ml pre-warmed 0.1 × SSC/0.1% SDS, re-suspending the beads once at each washing step. Hybrid-selected DNA was eluted with 50 μl 0.1 M NaOH. After 5 minutes at room temperature, the beads were pulled down, the supernatant transferred to a tube containing 70 μl of 1 M Tris-HCl, pH 7.5, and the neutralized DNA was desalted and concentrated on a QIAquick MinElute column and eluted in 20 μl.

## Sequencing

Pooled, indexed samples were sequenced on Illumina HiSeq2500 at the Broad Institute to produce 101-bp paired-end reads. Sequence data have been deposited in the NCBI Short Read Archive under a project accession number, PRJNA438372.

## RNAseq read processing and mapping

Identifiers for all RNAseq experiments are listed in Supplementary Table 1. The RNAseq reads were trimmed with a Phred quality score cut-off of 20 by the program fastq_quality_trimmer from the FASTX toolkit, version 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/). Reads shorter than 20 bp after adaptor- and poly(A)-trimming were discarded before mapping. Trimmed RNAseq reads were aligned to the *B. fragilis* NCTC 9343 genome (NC_003228.3) and the mouse genome (genome build GRCm38.p4) in parallel. RNAseq read mapping to the bacterial and mouse genome were performed by Bowtie2[63] and STAR[64], respectively. The mapping results were used to calculate read counts over each gene of the bacterial and mouse genomes by bedtools[65] and HTseq[66], respectively. We also calculated read coverage and transcripts per million (TPM) for each bacterial gene for examining the distribution of bacterial gene expression.

### Differential gene expression analysis

Read counts for each bacterial gene were used to analyze differential gene expression using the edgeR package[67]. Bacterial genes with more than 10 uniquely mapped reads in each of three replicates were considered to be detected and were retained for the differential gene expression analysis. Genes with an adjusted p-value < 0.05 in the edgeR analysis were considered differentially expressed.

### Functional annotation and structural modeling

We used the Broad Institute prokaryotic annotation pipeline to perform detailed functional analysis of genes of interest in *B. fragilis* NCTC9343. Protein-coding genes were predicted with Prodigal[68] and filtered to remove genes with >=70% overlap to tRNAs or rRNAs. The gene product names were assigned based on top blast hits against SwissProt protein database (>=70% identity and >=70% query coverage). Additional annotations were made using PFAM[30], KEGG[69], GO[70], and Enzyme Commission (EC) databases. Structural modeling was performed using Phyre[71].

### Functional enrichment and other statistical analysis

We identified predicted protein domains within the bacterial genes using Pfam categories[30] as part of the Broad Institute's prokaryotic annotation pipeline[72]. To assess functional enrichment of differentially expressed genes, we calculated statistical significance using the hypergeometric function with adjustment for multiple hypothesis testing. Adjusted p-values < 0.05 were considered enriched. We calculated Pearson correlation coefficients to determine if the expression of genes (in TPM) were comparable between two samples. We used the Wilcoxon signed-rank test to compare the expression of a gene family in two different samples. p-values < 0.05 were considered significant. STRING analysis of relatedness of differentially expressed genes was performed using the online database[31]. Default settings were used: minimum interaction score of 0.4 with interaction sources including text mining, experiments, databases, co-expression, neighborhood, gene fusion, and co-occurrence.

### Comparative genomics and motif scanning of 92 *Bacteroides* and *Parabacteroides* genomes

Using a comparative analysis of 92 diverse genome sequences related to *B. fragilis*[73], which included 23 *Bacteroides* and 5 *Parabacteroides* species, we identified 43 BF3134 orthologs in 43 strains and 117 BF3086 orthologs in 83 strains. We constructed multiple alignments of the nucleotide sequences of these groups of orthologs for BF3086 and BF3134, which we used to calculate pairwise sequence identities to measure conservation levels. We searched the upstream regions of BF3086 and BF3134 in *B. fragilis* NCTC 9343 for conserved motifs, or potential binding sites for transcription factors, using GLAM2[74] from the MEME suite[75]. We further examined the presence and conservation of this potential regulatory motif using GLAM2Scan[74] on our set of 92 diverse *Bacteroides* and *Parabacteroides* genome sequences (Extended Data Figure 7)[73]. Presence of the motif was defined by having 2 or less mismatches. Of the 16 *B. fragilis* genomes, we used multiple sequence alignments to confirm that the 3 divergent *B. fragilis* were missing the predicted motif (Extended Data

Figure 7). Excluding three same-patient samples, the *B. fragilis* genomes containing the motif had a pairwise average nucleotide identity (ANI) value of 98%, indicating that these strains were members of the same species, but not clonally related. In contrast, pairwise ANI values between *B. fragilis* strains with and without the motif averaged 86%, below the threshold commonly used to describe species[72,76].

## Quantitative real-time PCR

RNA purified as described above was used to generate cDNA with the Bio-Rad iScript cDNA synthesis kit according to the manufacturer instructions. Real-time PCR reactions with Applied Biosystems' Power SYBR Green Master Mix were run on an ABI PRISM 7900HT Fast Real-Time PCR System (Applied Biosystems). Relative quantification was calculated using the ΔΔCt method with one control (wild-type or lumen) sample as the calibrator (set to 1) and DNA gyrase (*gyrB*) as the housekeeping gene control for normalization. The mean Ct value of 3 technical replicates was used for each sample. Primers are listed in Supplementary Table 9.

## DNBS colitis

Female mice were randomly shuffled between cages into groups and mono-colonized at 3 weeks of age (just after weaning). Four weeks later, mice were anesthetized using isofluorane and 5% DNBS in 50% ethanol (or 50% ethanol only) was administered rectally through a 3.5F catheter (Instech Solomon) inserted 4 cm into the colon. Mice were subsequently kept upside-down for 1 minute to prevent leakage. Mice were weighed every 24 hours and sacrificed at 72 hours post-induction. The whole colon was dissected, fixed in 10% buffered formalin, paraffin embedded, sectioned, and then stained with hematoxylin and eosin (HE). Colitis histopathologic scores were evaluated in a blinded analysis using a method similar to previous work[46]. Briefly, tissue thickening was assessed 0 – 1. Invasion of polymorphonuclear cells was scored from 0 – 3 and that of mononuclear cells was scored from 0 – 3. Tissue damage was scored in the epithelium (0 – 3), the submucosa (0 – 3) and the muscularis (0 – 3). Total score 0 – 16 was assessed separately for the proximal, medial, and distal colon and then summed (maximum possible score 48).

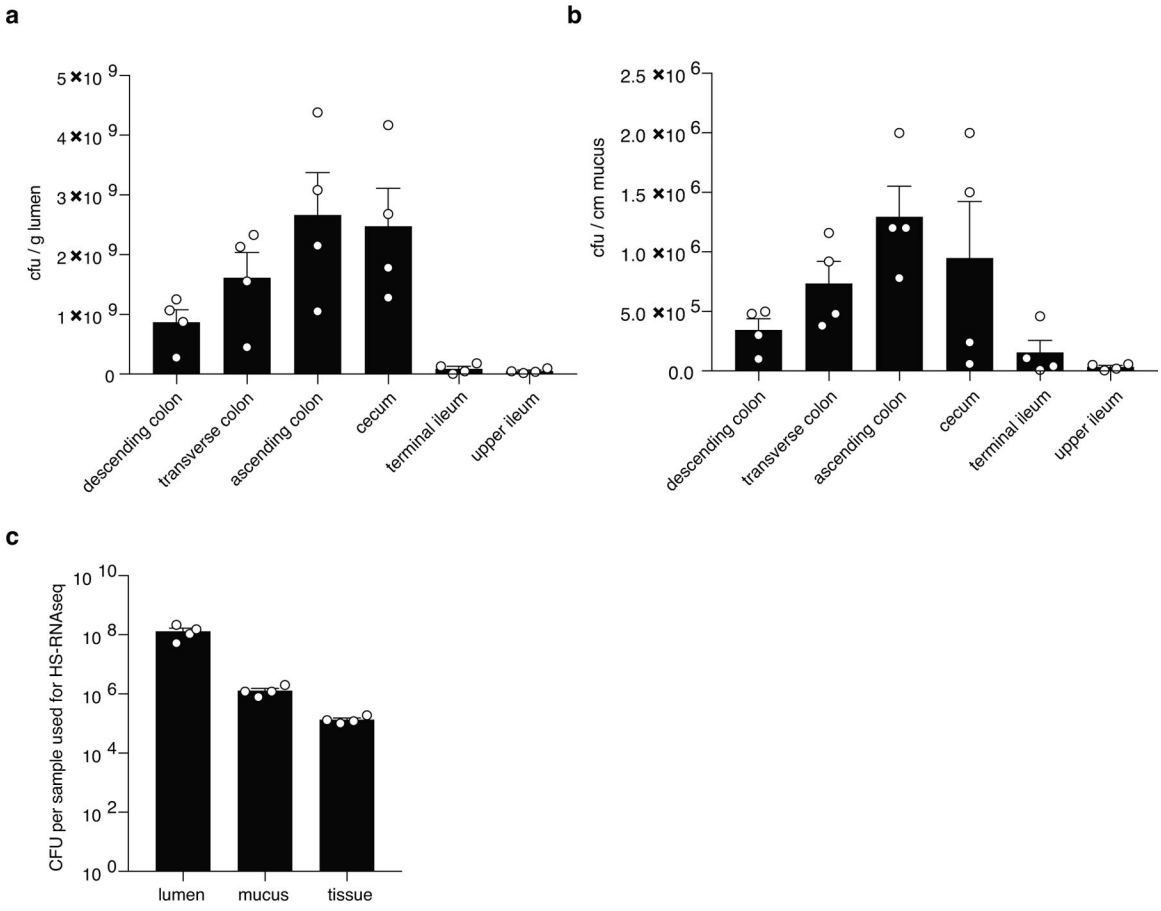
## Isolation of mesenteric lymph node lymphocytes and flow cytometry

Mesenteric lymph nodes (MLN) were isolated and processed by dissociating tissues through a 70 μm cell strainer (BD Falcon) to generate single cell suspensions. Cells were washed in ice cold PBS. For flow cytometry analysis, cells were labelled with the LIVE/DEAD fixable violet dead cell stain kit (Life Technologies), with empirically titrated concentration of PE-Cy7-conjugated anti-mouse CD4 (RM4–5, eBioscience). For intracellular staining, cells were fixed and permeabilized with the Foxp3/Transcription factor buffer kit (eBioscience), followed by staining with the following antibodies: FITC-conjugated anti-mouse IFNγ (XMG1.2, eBioscience), PE-conjugated anti-mouse IL-10 (JES5–16E3, eBioscience), PerCP-Cy5.5 anti-mouse IL-17A (eBio17B7, eBioscience), and APC-conjugated anti-mouse Foxp3 (FJK16s, eBioscience). Cell acquisition was performed on a Miltenyi MACSQuant (Miltenyi), and data was analyzed using FlowJo software suite (TreeStar).

### Statistical Analysis

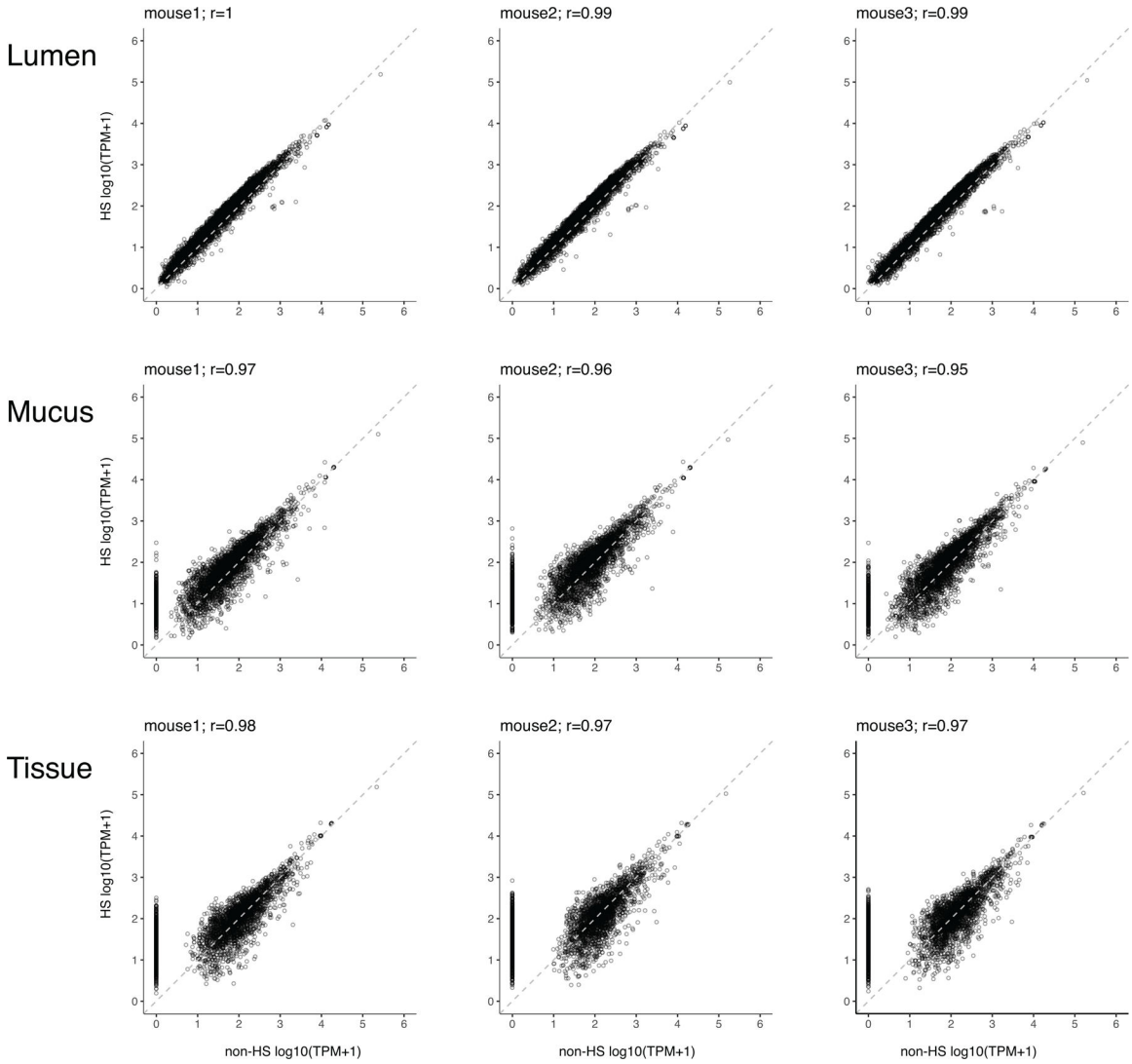
Unless otherwise stated, mean and standard error are plotted in all figures. In general, standard one or two-way ANOVA tests were used for comparing means with post-hoc correction for multiple comparisons. For dynamic colonization experiments (Fig. 4a–d), data were log-transformed (Y = Log(Y)) prior to statistical testing. In all figures, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Details for all statistical comparisons, including effect sizes, confidence intervals, and exact p-values are included in Supplementary Table 10.

## Extended Data



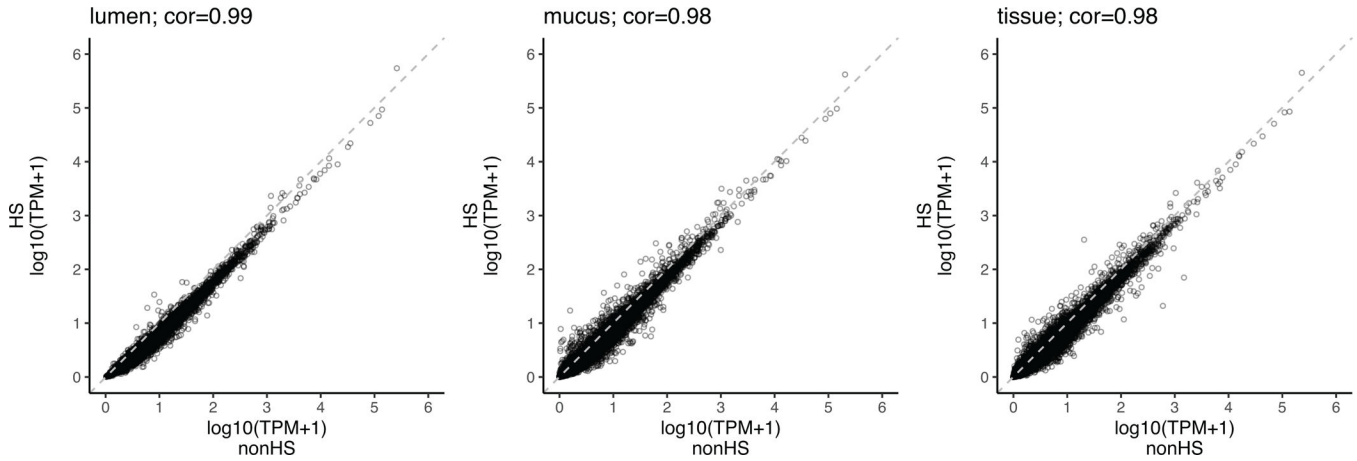

**Extended Data Figure 1 |. Intestinal biogeography of *Bacteroides fragilis* during mono-colonization.**

**a**, CFU per gram of lumen content and **b**, CFU per cm of mucus from indicated regions of intestine after 4 weeks of mono-colonization with wild-type *B. fragilis* (mean and standard error, n = 4 animals). **c**, CFU per sample in lumen, mucus, and tissue samples of the proximal intestine of mice mono-colonized for 4 weeks with wild-type *B. fragilis* (mean and standard error, n = 4 animals). These samples were collected using the same dissection method used to prepare samples for RNA-Seq (Fig. 1a).



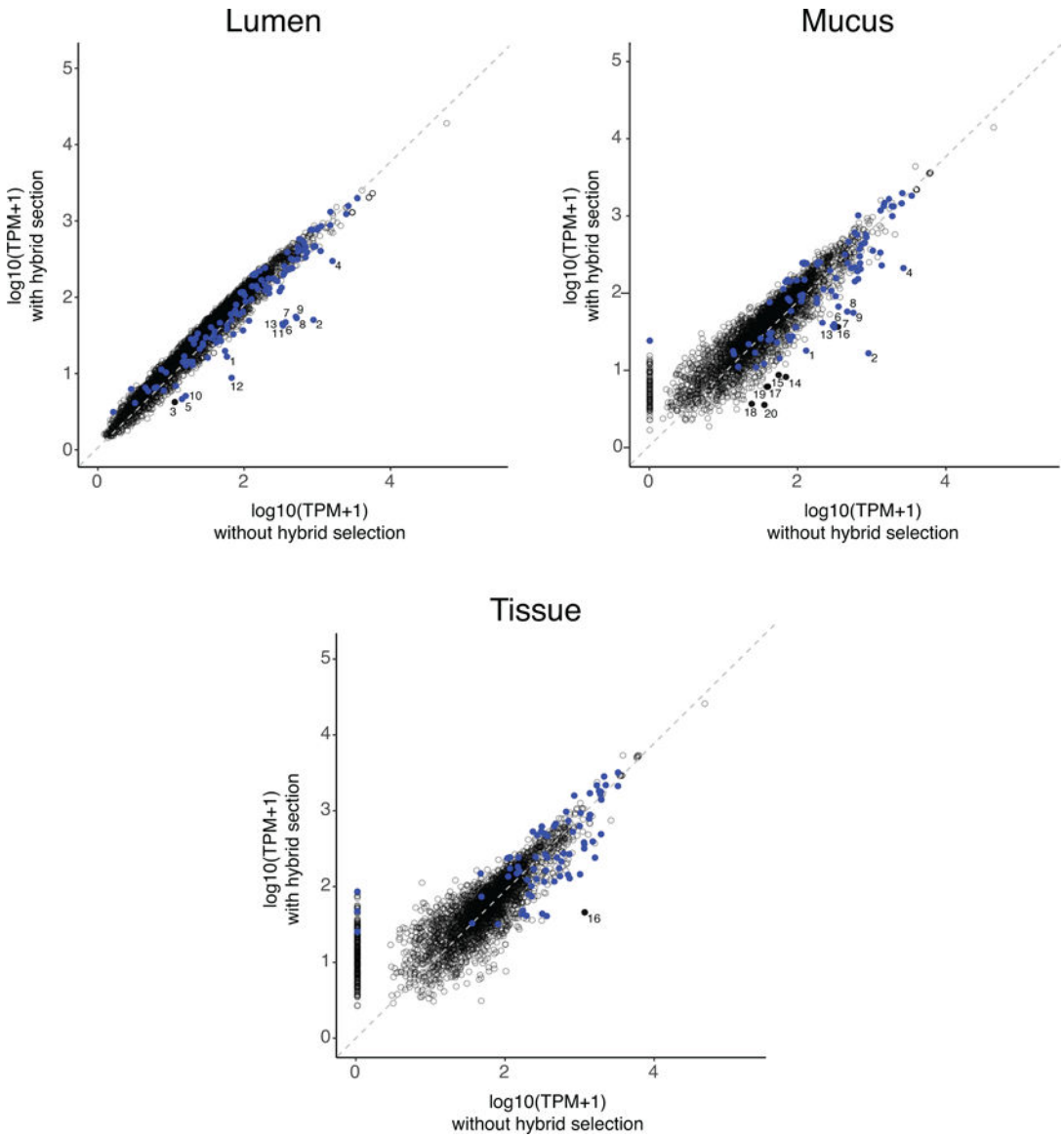

**Extended Data Figure 2 |. Individual mouse correlation plots to assess hybrid selection performance.**

Correlation plots for HS vs non-HS in individual mice (3 individual-mouse samples from lumen, 3 from mucus, and 3 from tissue, Pearson's r). Each dot represents a single gene.



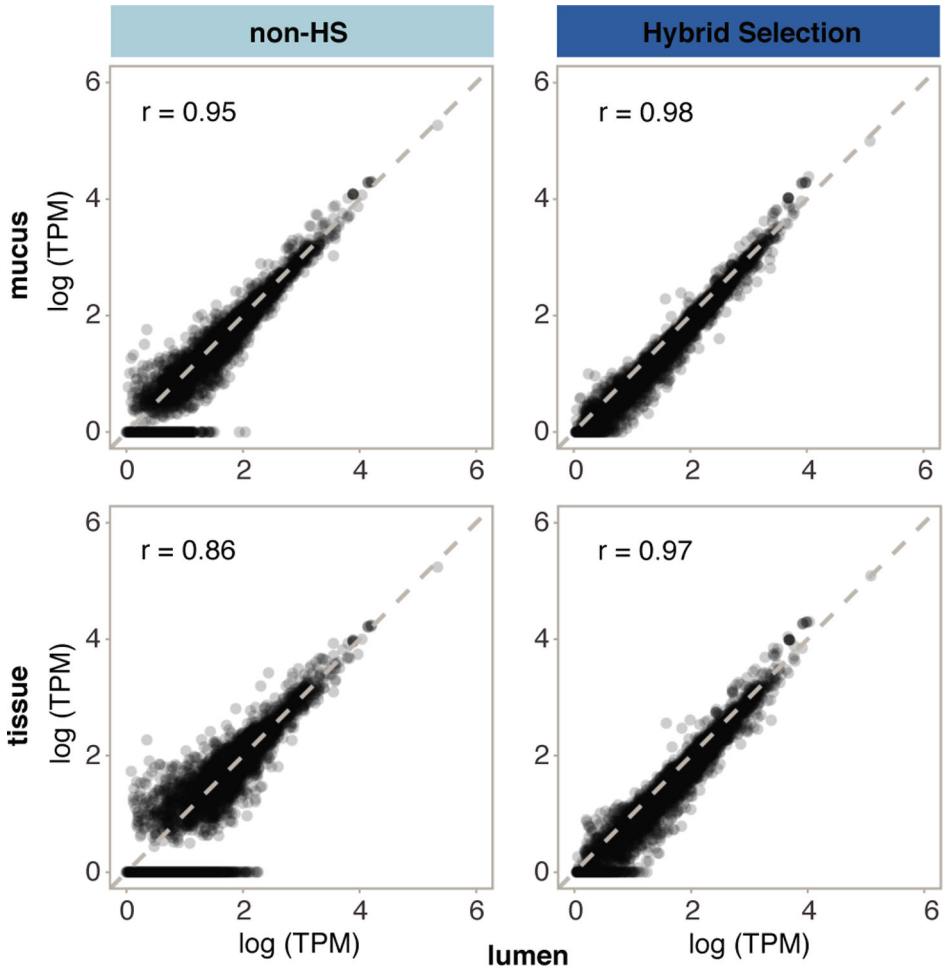

**Extended Data Figure 3 |. Host gene expression comparisons between samples with and without hybrid selection.**

Total RNA-Seq reads were mapped to mm10 mouse genome using STAR, and the mapped reads were converted into read counts for each gene by HTSeq. After excluding genes with <10 reads mapping across any sample, the read counts for each sample were normalized by TPM (Transcripts Per Million). Each dot represents a single gene. The average TPM for each gene is shown from non-hybrid selected libraries (x-axis) and hybrid selected libraries (y-axis) (n = 3 animals, Pearson's r).



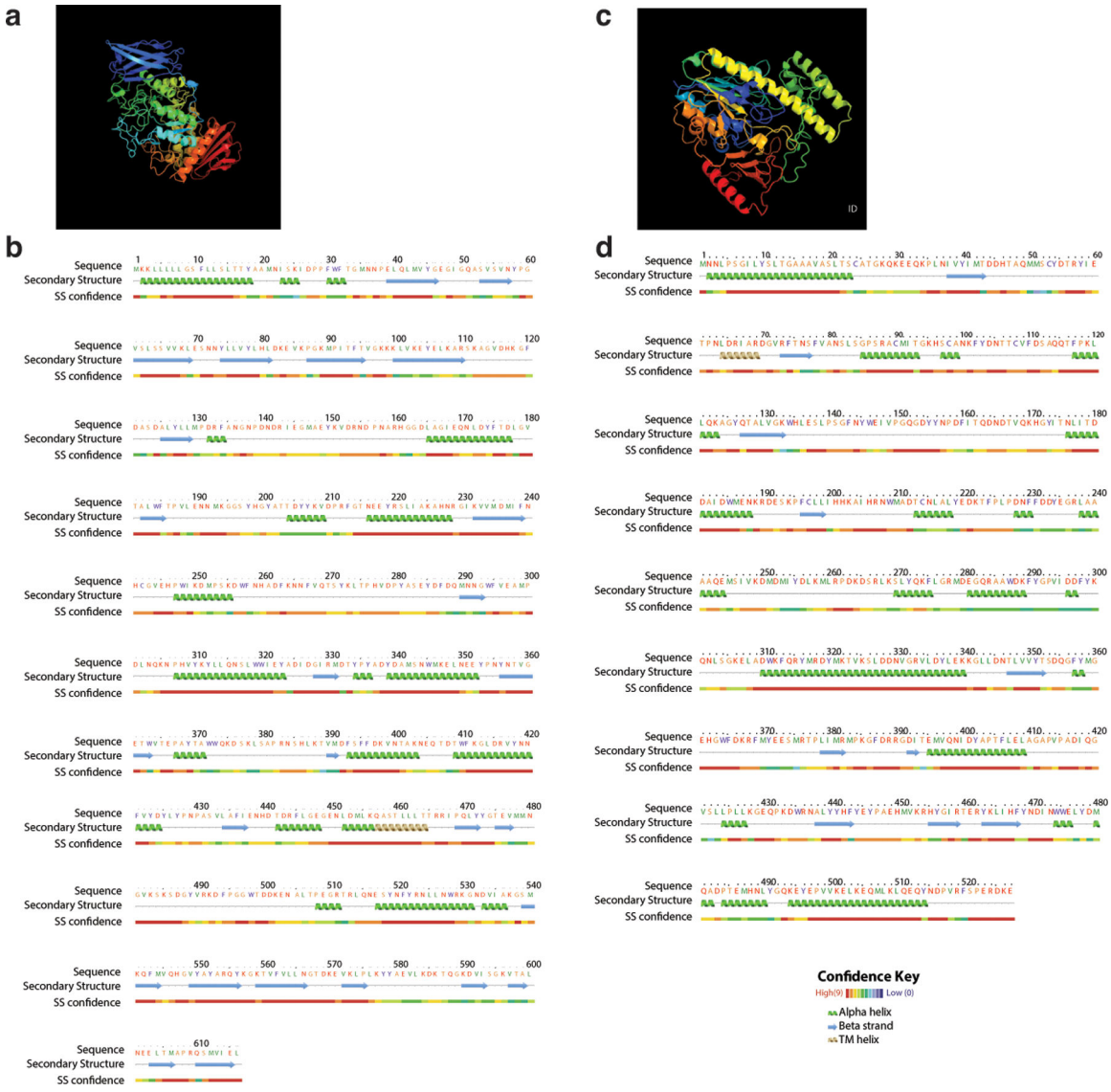

**Extended Data Figure 4 |. Normalized gene expression levels with and without hybrid selection are highly correlated with few outliers.**

Each gene is represented by a single dot. The correlation coefficients for lumen, mucus, and tissue are 0.99, 0.96, and 0.98, respectively. Outliers where the difference between the HS and non-HS values is larger than three standard deviations are numbered and listed in Supplementary Table 3. These represent primarily short genes (median length 110 nucleotides), particularly tRNA and 5s rRNA genes. Short genes (<200 nt) are colored blue, showing that most protein-coding genes are enriched properly.




**Extended Data Figure 5 |. Correlation in gene expression between different sample sites was improved with hybrid selection.**

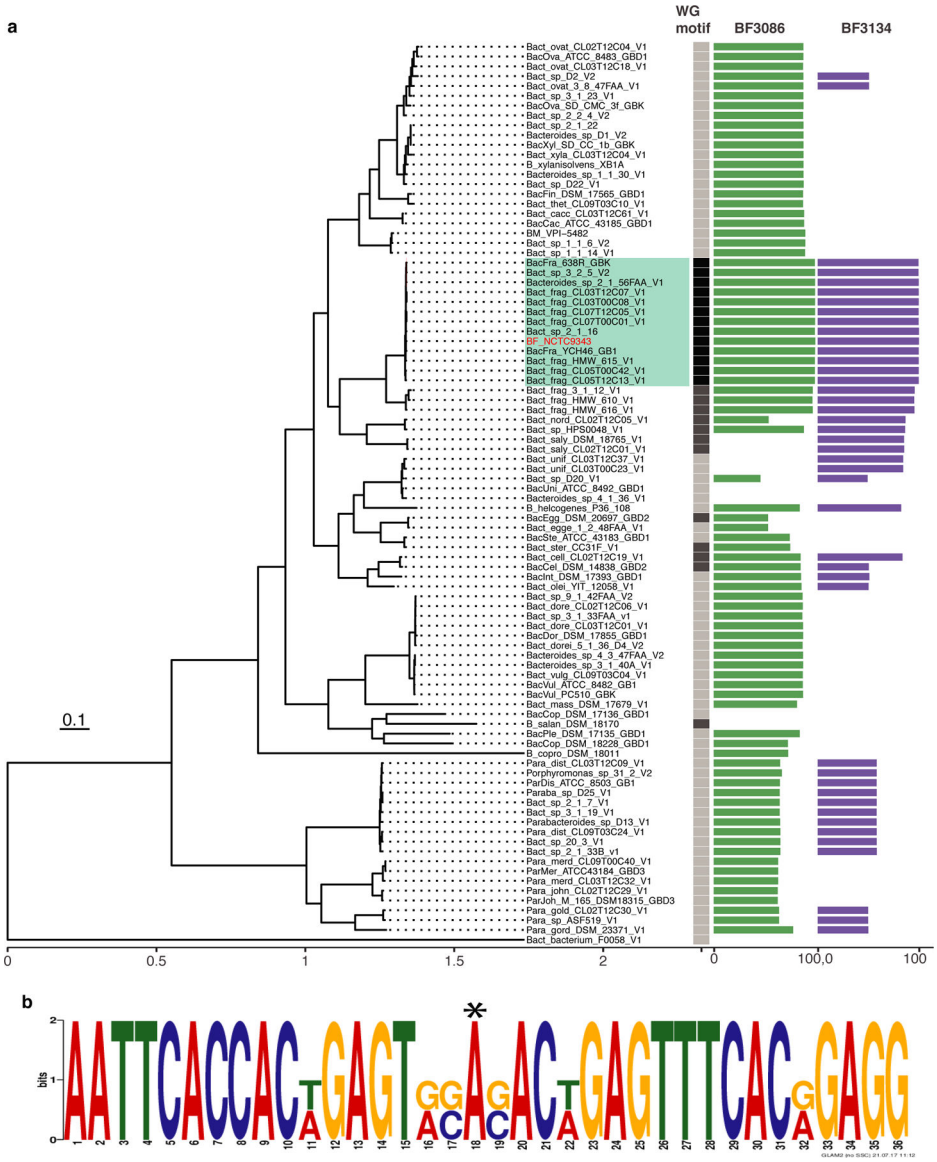
Each dot represents a single gene with all genes plotted (n = 3 animals, Pearson's r).




**Extended Data Figure 6 |. Structural modeling for genes of interest using Phyre.**

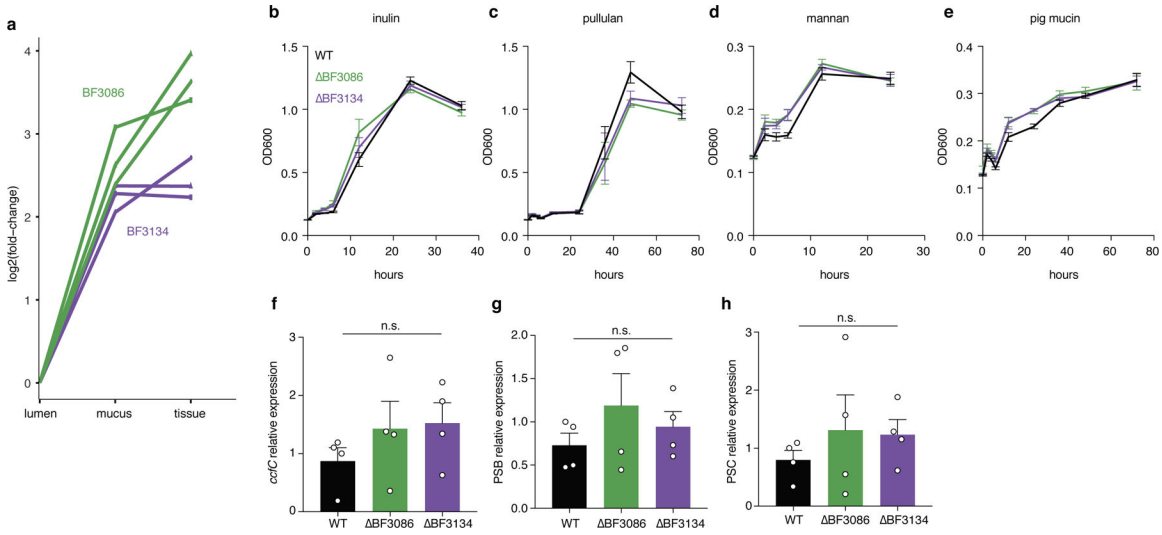
**a**, The predicted structure for BF3134, modeled using Phyre[71], indicated that BF3134 is a likely cyclo-malto-dextrinase, closely related to neopullulanase and maltogenic amylase and a member of glycosyl hydrolase family 13 (96% of the sequence was modeled with 100% confidence to the cyclo-malto-dextrinase template c3edeB, with 42% identity). **b**, Secondary structure prediction for BF3134 using Phyre. Pfam domain analysis for BF3134 also indicated the presence of an N-terminal cyclo-malto-dextrinase domain (PF09087), a central alpha-amylase domain (glycosyl hydrolase family 13; PF00128), and a C-terminal cyclo-malto-dextrinase domain (PF10438). **c**, The predicted structure for BF3086 indicated a role as an acetylglucosamine-6-sulfatase (93% of the sequence was modeled with 100% confidence by the single highest scoring template, c5g2va, an n-acetylglucosamine-6-sulfatase, with 51% identity). **d**, Secondary structure prediction for BF3086. Pfam domain analysis indicated the presence of a sulfatase domain, in addition to a domain of unknown function (DUF4976) downstream of the sulfatase domain. The region aligned by Phyre with

the c5g2va template included both the regions encompassed by the Pfam sulfatase domain, as well as the Pfam domain of unknown function (DUF4976).



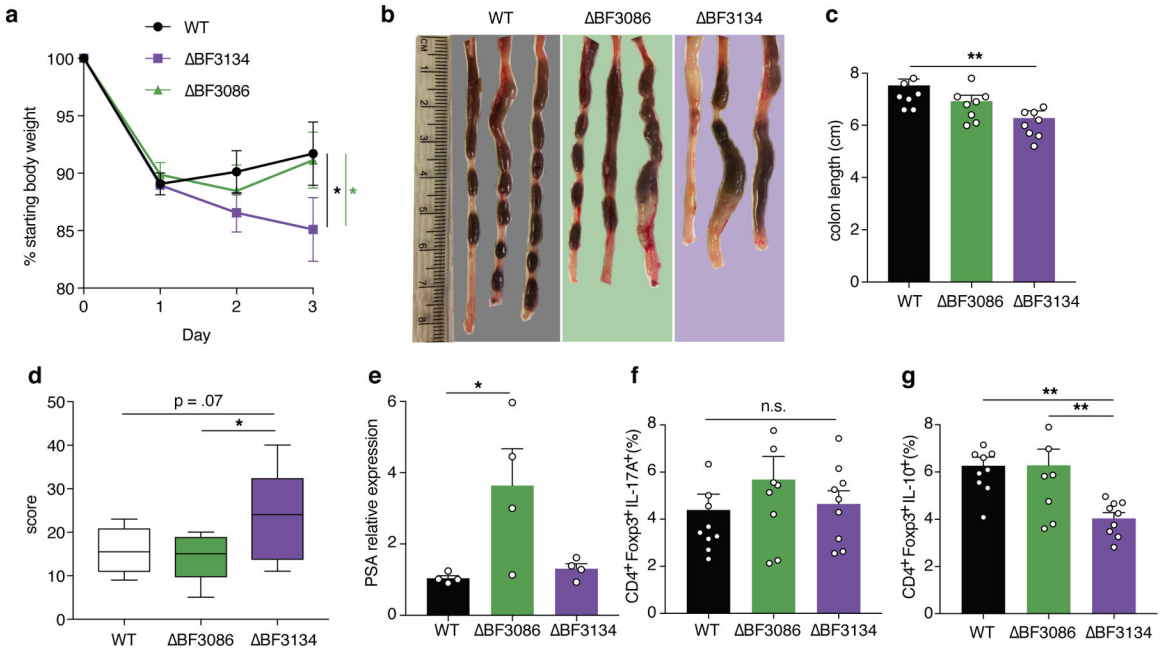

**Extended Data Figure 7 |. BF3086 and BF3134 are conserved and share a potential regulatory motif.**

**a**, Phylogeny of 92 *Bacteroides* and *Parabacteroides* strains[73] showing the presence of BF3086 and BF3134 orthologues, with horizontal bar graphs indicating the percent protein sequence identity to the studied type strain (NCTC9343, highlighted with red font). The teal box indicates strains that can be confidently assigned to the *B. fragilis* species (average pairwise ANI[77] between them is 98%, whereas it falls below 95% for the next-closest strains also labeled as *B. fragilis*). The black squares indicate the presence of the conserved upstream motif (0–2 mismatches), using the GLAM2Scan algorithm[74]. **b**, Sequence of the conserved motif upstream of both genes. The asterisk (*) at position 18 indicates a position that differs between the upstream regions of the glycosyl hydrolase (BF3086) and the sulfatase (BF3134). The glycosyl hydrolase upstream region has an “A” at this position, whereas the sulfatase upstream region has a deletion at this position.



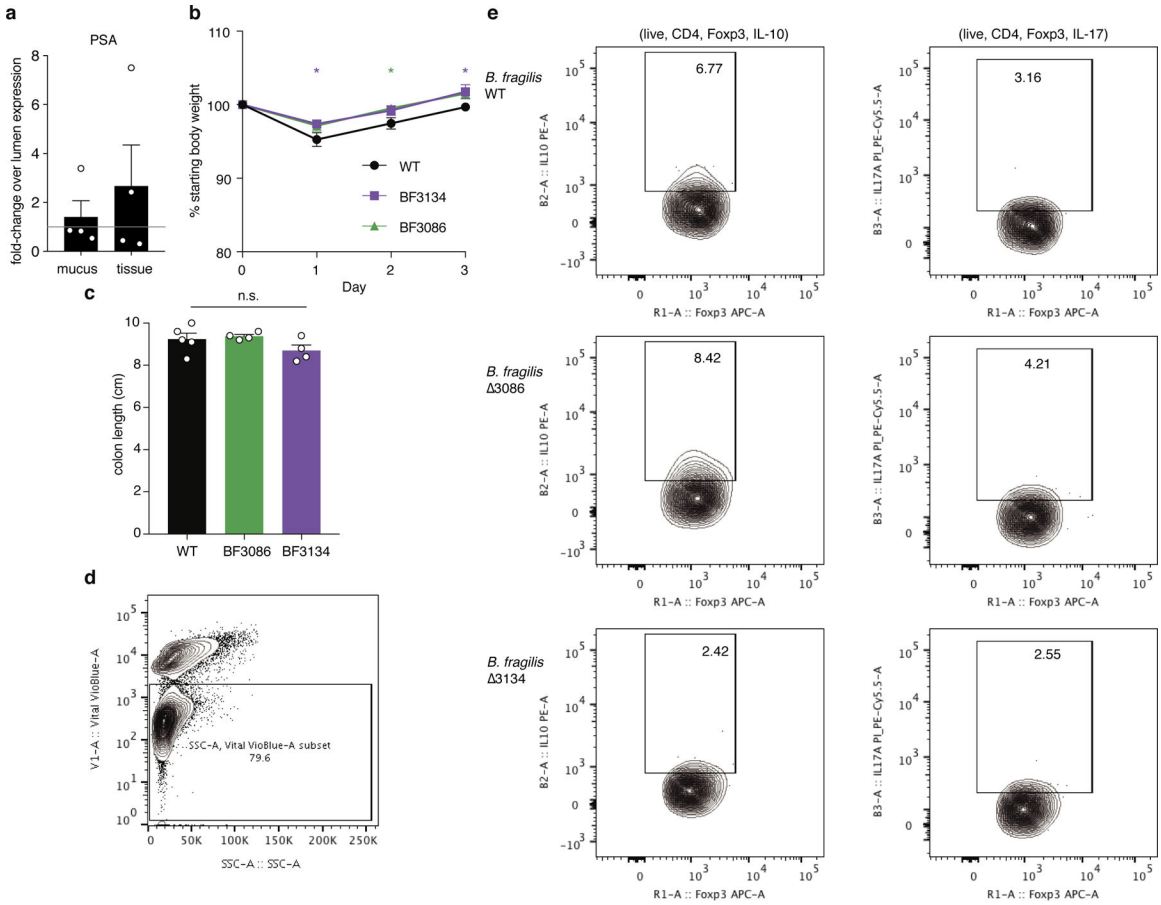

**Extended Data Figure 8 |. Additional *in vitro* and *in vivo* phenotypes of ΔBF3086 and ΔBF3134.**

**a,** BF3086 and BF3134 biological replicates. Fold-change for individual mice indicate consistently induced expression of BF3086 and BF3134 in the mucus and tissue relative to the lumen. **b–e**, Growth of individual *B. fragilis* strains in a defined minimal medium with **b**, inulin, **c**, pullulan, **d**, mannan, or **e**, pig mucin (mean and standard error, n = 8 independent cultures). **f–h**, Quantitative real-time PCR (ΔΔCt method normalized to *gyrB*) on fecal samples of mice mono-colonized with indicated strains of *B. fragilis*, assessing the expression of **f**, *ccfC* (BF3581), **g**, PSB flippase (BF1900), and **h**, PSC flippase (BF1014) (mean and standard error, Tukey ANOVA, n = 4 animals).




**Extended Data Figure 9 |. BF3134 is required for *B. fragilis* protection from experimental colitis.**

**a**, Mice were mono-colonized with *B. fragilis* strains at weaning (3 weeks of age) before inducing DNBS colitis at 7 weeks of age. Body weights of mice were measured every 24 hours and are represented as a percentage of their starting weight on day 0 (Tukey 2-way ANOVA, n = 10, 9, 9, representative of two independent experiments). **b**, 72 hours after colitis induction, mice were sacrificed and the length of the colon from rectum to the cecal junction was dissected (representative images of 3 colons per group, images normalized to size using rulers and then cropped around the colon) and **c,** colon length measured (Tukey ANOVA, n = 10, 9, 9). **d**, Histopathologic scores of whole colons (max 48, mean and interquartile range, Tukey ANOVA, n = 10, 9, 9). **e**, Quantitative real-time PCR (ΔΔCt method normalized to *gyrB*) on fecal samples of mice mono-colonized with indicated strains of *B. fragilis*, assessing the expression of the PSA flippase (BF1369) (Tukey ANOVA, n = 4 animals). **f**, Lymphocytes isolated from mesenteric lymph nodes of mono-colonized, DNBS-induced mice were analyzed using flow cytometry. IL-17A-producing T cells quantified as a percent of total CD4+Foxp3+ regulatory T cells (Tukey ANOVA, n = 10, 9, 9 animals). **g**, IL-10-producing T cells quantified as a percent of total CD4+Foxp3+ regulatory T cells (Tukey ANOVA, n = 10, 9, 9 animals, representative of two independent experiments) (all panels unless noted: mean and standard error, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).




**Extended Data Figure 10 |. Control experiments and flow cytometry methods for DNBS colitis.**

**a**, Quantitative real-time PCR (ΔΔCt method normalized to *gyrB*) for PSA flippase (BF1369) in lumen, mucus and tissue samples (mean and standard error, n = 4 animals). Fold-change between sample sites was quantified within each mouse individually. **b**, Mice mono-colonized with indicated strains of *B. fragilis* for one month were treated with 50% ethanol, the vehicle control for DNBS colitis induction. Mice were weighed every 24 hours, graphed as a percentage of their weight at day 0 (Tukey 2-way ANOVA, n = 5, 4, 4). **c**, 72 hours after treatment the mice were sacrificed and the length of the colon was measured from rectum to the cecal junction (Tukey 2-way ANOVA, n = 5, 4, 4) **d**, Example live cell gating for flow cytometry in Extended Data Figure 9f and 9g (representative from two independent experiments with similar results). **e**, Example flow plots (1 from each group) for assessing the proportion of IL-10 and IL-17 positive regulatory T cells, as quantified in Extended Data Figure 9f and 9g (representative from two independent experiments with similar results, mean and standard error in graphs, * $p < 0.05$).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

We would like to thank Drs. Elaine Hsiao, Eric Martens, Dirk Gevers, Christopher Desjardins, Brian Haas and Jonathan Livny for helpful discussions, and members of the Mazmanian laboratory for comments. G.P.D. was supported by an NIH training grant (5T32 GM07616), NSF Graduate Research Fellowship (DGE-1144469) and the Center for Environmental Microbial Interactions at Caltech. The project was funded by NIH grant U19AI110818 to the Broad Institute; NIH grant DK110534 to H.C.; NIH grants GM099535 and DK078938, and the Heritage Medical Research Institute to S.K.M.

## REFERENCES

1. Topping DL & Clifton PM Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides. Physiol. Rev 81, 1031–1064 (2001). [PubMed: 11427691]
2. Buffie CG & Pamer EG Microbiota-mediated colonization resistance against intestinal pathogens. Nat. Rev. Immunol 13, 790–801 (2013). [PubMed: 24096337]
3. Donaldson GP, Lee SM & Mazmanian SK Gut biogeography of the bacterial microbiota. Nat. Rev. Microbiol 14, 20–32 (2015). [PubMed: 26499895]
4. Johansson MEV & Hansson GC Immunological aspects of intestinal mucus and mucins. Nat. Rev. Immunol (2016). 10.1038/nri.2016.88
5. Li H et al. The outer mucus layer hosts a distinct intestinal microbial niche. Nat Commun 6, 8292 (2015). [PubMed: 26392213]
6. Wang Y et al. Regional mucosa-associated microbiota determine physiological expression of TLR2 and TLR4 in murine colon. PLoS ONE 5, e13607 (2010). [PubMed: 21042588]
7. Yasuda K et al. Biogeography of the intestinal mucosal and lumenal microbiome in the rhesus macaque. Cell Host Microbe 17, 385–391 (2015). [PubMed: 25732063]
8. Albenberg L et al. Correlation Between Intraluminal Oxygen Gradient and Radial Partitioning of Intestinal Microbiota in Humans and Mice. Gastroenterology 147, 1055–63.e8 (2014). [PubMed: 25046162]
9. Davis CP & Savage DC Habitat, succession, attachment, and morphology of segmented, filamentous microbes indigenous to the murine gastrointestinal tract. Infect. Immun 10, 948–956 (1974). [PubMed: 4426712]
10. Ivanov II et al. Induction of intestinal Th17 cells by segmented filamentous bacteria. Cell 139, 485–498 (2009). [PubMed: 19836068]
11. Vélez MP, De Keersmaecker SCJ & Vanderleyden J Adherence factors of Lactobacillus in the human gastrointestinal tract. FEMS Microbiol. Lett 276, 140–148 (2007). [PubMed: 17888009]
12. Savage DC Microbial ecology of the gastrointestinal tract. Annu Rev Microbiol 31, 107–133 (1977). [PubMed: 334036]
13. Rowan F et al. Bacterial colonization of colonic crypt mucous gel and disease activity in ulcerative colitis. Ann. Surg 252, 869–875 (2010). [PubMed: 21037444]
14. Pédron T et al. A crypt-specific core microbiota resides in the mouse colon. MBio 3, (2012).
15. Earle KA et al. Quantitative Imaging of Gut Microbiota Spatial Organization. Cell Host Microbe 18, 478–488 (2015). [PubMed: 26439864]
16. Mark Welch JL, Hasegawa Y, McNulty NP, Gordon JI & Borisy GG Spatial organization of a model 15-member human gut microbiota established in gnotobiotic mice. Proc Natl Acad Sci USA 114, E9105–E9114 (2017). [PubMed: 29073107]
17. Marsh JW, Humphrys MS & Myers GSA A Laboratory Methodology for Dual RNA-Sequencing of Bacteria and their Host Cells In Vitro. Front Microbiol 8, 1830 (2017). [PubMed: 28983295]
18. Westermann AJ, Barquist L & Vogel J Resolving host-pathogen interactions by dual RNA-seq. PLoS Pathog 13, e1006033 (2017). [PubMed: 28207848]
19. Humphrys MS et al. Simultaneous transcriptional profiling of bacteria and their host cells. PLoS ONE 8, e80597 (2013). [PubMed: 24324615]
20. Mavromatis CH et al. The co-transcriptome of uropathogenic Escherichia coli-infected mouse macrophages reveals new insights into host-pathogen interactions. Cell. Microbiol 17, 730–746 (2015). [PubMed: 25410299]

21. Westermann AJ & Vogel J in Bacterial Regulatory RNA: Methods and Protocols (eds. Arluison V & Valverde C) 59–75 (Springer New York, 2018).
22. Vannucci FA, Foster DN & Gebhart CJ Laser microdissection coupled with RNA-seq analysis of porcine enterocytes infected with an obligate intracellular pathogen (Lawsonia intracellularis). BMC Genomics 14, 421 (2013). [PubMed: 23800029]
23. Bernstein JA, Khodursky AB, Lin P-H, Lin-Chao S & Cohen SN Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. Proc Natl Acad Sci USA 99, 9697–9702 (2002). [PubMed: 12119387]
24. Gnirke A et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol 27, 182–189 (2009). [PubMed: 19182786]
25. Melnikov A et al. Hybrid selection for sequencing pathogen genomes from clinical samples. Genome Biol 12, R73 (2011). [PubMed: 21835008]
26. Matranga CB et al. Enhanced methods for unbiased deep sequencing of Lassa and Ebola RNA viruses from clinical and biological samples. Genome Biol 15, 519 (2014). [PubMed: 25403361]
27. Huang JY, Lee SM & Mazmanian SK The human commensal Bacteroides fragilis binds intestinal mucin. Anaerobe 17, 137–141 (2011). [PubMed: 21664470]
28. Lee SM et al. Bacterial colonization factors control specificity and stability of the gut microbiota. Nature 501, 426–429 (2013). [PubMed: 23955152]
29. Donaldson GP et al. Gut microbiota utilize immunoglobulin A for mucosal colonization. Science 360, 795–800 (2018). [PubMed: 29724905]
30. Finn RD et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44, D279–85 (2016). [PubMed: 26673716]
31. Szklarczyk D et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res 45, D362–D368 (2017). [PubMed: 27924014]
32. Blow MJ et al. The Epigenomic Landscape of Prokaryotes. PLoS Genet 12, e1005854 (2016). [PubMed: 26870957]
33. Burroughs AM, Kaur G, Zhang D & Aravind L Novel clades of the HU/IHF superfamily point to unexpected roles in the eukaryotic centrosome, chromosome partitioning, and biologic conflicts. Cell Cycle 16, 1093–1103 (2017). [PubMed: 28441108]
34. Liu CH, Lee SM, Vanlare JM, Kasper DL & Mazmanian SK Regulation of surface architecture by symbiotic bacteria mediates host colonization. Proc Natl Acad Sci USA 105, 3951–3956 (2008). [PubMed: 18319345]
35. Coyne MJ, Chatzidaki-Livanis M, Paoletti LC & Comstock LE Role of glycan synthesis in colonization of the mammalian gut by the bacterial symbiont Bacteroides fragilis. Proc Natl Acad Sci USA 105, 13099–13104 (2008). [PubMed: 18723678]
36. Round JL et al. The Toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. Science 332, 974–977 (2011). [PubMed: 21512004]
37. Baughn AD & Malamy MH The strict anaerobe Bacteroides fragilis grows in and benefits from nanomolar concentrations of oxygen. Nature 427, 441–444 (2004). [PubMed: 14749831]
38. Rocha ER & Smith CJ Role of the alkyl hydroperoxide reductase (ahpCF) gene in oxidative stress defense of the obligate Anaerobe bacteroides fragilis. Journal of Bacteriology 181, 5701–5710 (1999). [PubMed: 10482511]
39. Sund CJ et al. The Bacteroides fragilis transcriptome response to oxygen and H2O2: the role of OxyR and its effect on survival and virulence. Molecular Microbiology 67, 129–142 (2008). [PubMed: 18047569]
40. Schofield WB, Zimmermann-Kogadeeva M, Zimmermann M, Barry NA & Goodman AL The Stringent Response Determines the Ability of a Commensal Bacterium to Survive Starvation and to Persist in the Gut. Cell Host Microbe 1–20 (2018). 10.1016/j.chom.2018.06.002
41. Benjdia A, Martens EC, Gordon JI & Berteau O Sulfatases and a radical S-adenosyl-L-methionine (AdoMet) enzyme are key for mucosal foraging and fitness of the prominent human gut symbiont, Bacteroides thetaiotaomicron. J Biol Chem 286, 25973–25982 (2011). [PubMed: 21507958]

42. Thomsson KA et al. Detailed O-glycomics of the Muc2 mucin from colon of wild-type, core 1- and core 3-transferase-deficient mice highlights differences compared with human MUC2. Glycobiology 22, 1128–1139 (2012). [PubMed: 22581805]
43. Scholz M et al. Strain-level microbial epidemiology and population genomics from shotgun metagenomics. Nat. Methods 13, 435–438 (2016). [PubMed: 26999001]
44. Yassour M et al. Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. Science Translational Medicine 8, 343ra81–343ra81 (2016).
45. Mazmanian SK, Round JL & Kasper DL A microbial symbiosis factor prevents intestinal inflammatory disease. Nature 453, 620–625 (2008). [PubMed: 18509436]
46. Chu H et al. Gene-microbiota interactions contribute to the pathogenesis of inflammatory bowel disease. Science 352, 1116–1120 (2016). [PubMed: 27230380]
47. Round JL & Mazmanian SK Inducible Foxp3+ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. Proc Natl Acad Sci USA 107, 12204–12209 (2010). [PubMed: 20566854]
48. Jakobsson HE et al. The composition of the gut microbiota shapes the colon mucus barrier. EMBO Rep 16, 164–177 (2015). [PubMed: 25525071]
49. Thänert R, Goldmann O, Beineke A & Medina E Host-inherent variability influences the transcriptional response of Staphylococcus aureus during in vivo infection. Nat Commun 8, 14268 (2017). [PubMed: 28155859]
50. Nuss AM et al. Tissue dual RNA-seq allows fast discovery of infection-specific functions and riboregulators shaping host-pathogen transcriptomes. Proc Natl Acad Sci USA 114, E791–E800 (2017). [PubMed: 28096329]
51. Stapels DAC et al. Salmonella persisters undermine host immune defenses during antibiotic treatment. Science 362, 1156–1160 (2018). [PubMed: 30523110]
52. Aprianto R, Slager J, Holsappel S & Veening J-W Time-resolved dual RNA-seq reveals extensive rewiring of lung epithelial and pneumococcal transcriptomes during early infection. Genome Biol 17, 198 (2016). [PubMed: 27678244]
53. Metsky HC et al. Capturing sequence diversity in metagenomes with comprehensive and scalable probe design. Nat Biotechnol 37, 160–168 (2019). [PubMed: 30718881]
54. Sonnenburg JL et al. Glycan foraging in vivo by an intestine-adapted bacterial symbiont. Science 307, 1955–1959 (2005). [PubMed: 15790854]
55. Cao Y, Rocha ER & Smith CJ Efficient utilization of complex N-linked glycans is a selective advantage for Bacteroides fragilis in extraintestinal infections. Proc Natl Acad Sci USA 111, 12901–12906 (2014). [PubMed: 25139987]
56. Martens EC, Chiang HC & Gordon JI Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. Cell Host Microbe 4, 447–457 (2008). [PubMed: 18996345]
57. Kashyap PC et al. Genetically dictated change in host mucuscarbohydrate landscape exerts a diet-dependenteffect on the gut microbiota. Proc Natl Acad Sci USA (2013). 10.1073/pnas.1306070110
58. Pudlo NA et al. Symbiotic Human Gut Bacteria with Variable Metabolic Priorities for Host Mucosal Glycans. MBio 6, (2015).
59. Varel VH & Bryant MP Nutritional features of Bacteroides fragilis subsp. fragilis. Appl Microbiol 28, 251–257 (1974). [PubMed: 4853401]
60. Kotarski SF & Salyers AA Isolation and characterization of outer membranes of Bacteroides thetaiotaomicron grown on different carbohydrates. Journal of Bacteriology 158, 102–109 (1984). [PubMed: 6715279]
61. Alexeyev MF The pKNOCK series of broad-host-range mobilizable suicide vectors for gene knockout and targeted DNA insertion into the chromosome of gram-negative bacteria. BioTechniques 26, 824–6– 828 (1999).
62. Shishkin AA et al. Simultaneous generation of many RNA-seq libraries in a single reaction. Nat. Methods 12, 323–325 (2015). [PubMed: 25730492]
63. Langmead B & Salzberg SL Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012). [PubMed: 22388286]

64. Dobin A et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013). [PubMed: 23104886]

65. Quinlan AR & Hall IM BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010). [PubMed: 20110278]

66. Anders S, Pyl PT & Huber W HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169 (2015). [PubMed: 25260700]

67. Robinson MD, McCarthy DJ & Smyth GK edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140 (2010). [PubMed: 19910308]

68. Hyatt D et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11, 119 (2010). [PubMed: 20211023]

69. Kanehisa M, Furumichi M, Tanabe M, Sato Y & Morishima K KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 45, D353–D361 (2017). [PubMed: 27899662]

70. The Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. Nucleic Acids Res 45, D331–D338 (2017). [PubMed: 27899567]

71. Kelley LA, Mezulis S, Yates CM, Wass MN & Sternberg MJE The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10, 845–858 (2015). [PubMed: 25950237]

72. Lebreton F et al. Emergence of epidemic multidrug-resistant Enterococcus faecium from animal and commensal strains. MBio 4, (2013).

73. Sefik E et al. Individual intestinal symbionts induce a distinct population of ROR$\gamma^+$ regulatory T cells. Science 349, 993–997 (2015). [PubMed: 26272906]

74. Frith MC, Saunders NFW, Kobe B & Bailey TL Discovering sequence motifs with arbitrary insertions and deletions. PLoS Comput. Biol 4, e1000071 (2008). [PubMed: 18437229]

75. Bailey TL et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37, W202–8 (2009). [PubMed: 19458158]

76. Lebreton F et al. Tracing the Enterococci from Paleozoic Origins to the Hospital. Cell 169, 849–861.e13 (2017). [PubMed: 28502769]

77. Chan JZ-M, Halachev MR, Loman NJ, Constantinidou C & Pallen MJ Defining bacterial species in the genomic era: insights from the genus Acinetobacter. BMC Microbiol 12, 302 (2012). [PubMed: 23259572]

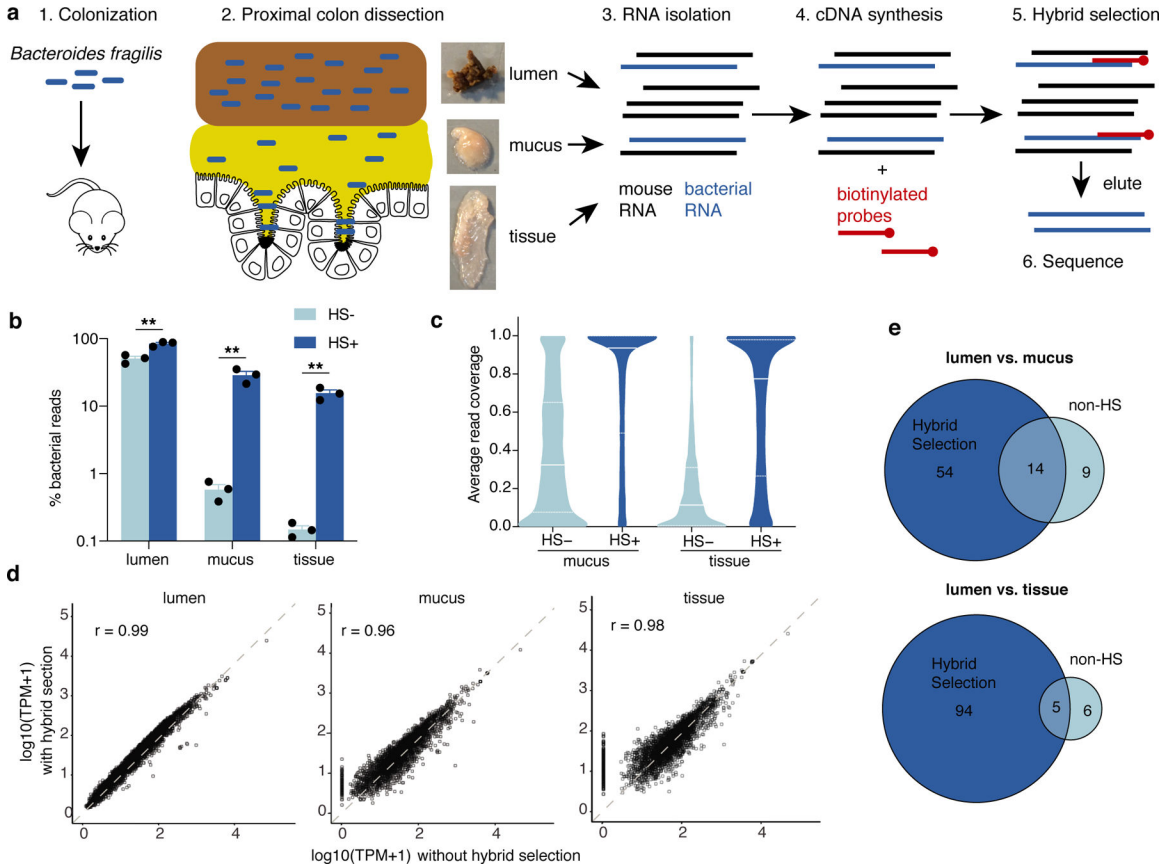



**Fig. 1 |. Hybrid selection (hsRNA-Seq) enables spatial bacterial transcriptomics during commensal colonization.**

**a**, Proximal colon lumen, mucus, and tissue samples were collected from mice mono-colonized with *B. fragilis* (3 mice, originating from different cages). RNA was purified from each sample type using the same protocol. After cDNA synthesis, biotinylated whole genome baits (probes) were used to select *B. fragilis* cDNA to the exclusion of mouse cDNA. The eluted bacterial cDNA libraries were then sequenced. **b**, Percentage of RNA-Seq reads mapping to the *B. fragilis* genome increases with hybrid selection (HS) (mean and standard error, one-sided t tests, n = 3, ** $p < 0.01$). **c**, Distribution of genes by average read coverage (in 3 animals) shifts dramatically with HS for mucus and tissue transcriptomes (median and interquartile range indicated, 4306 genes plotted). **d**, Normalized gene expression levels with and without HS are highly correlated (Pearson's r) within microenvironments. Each gene is represented by a single dot. **e**, HS increases the number of genes identified in differential expression analyses between various sample sites.

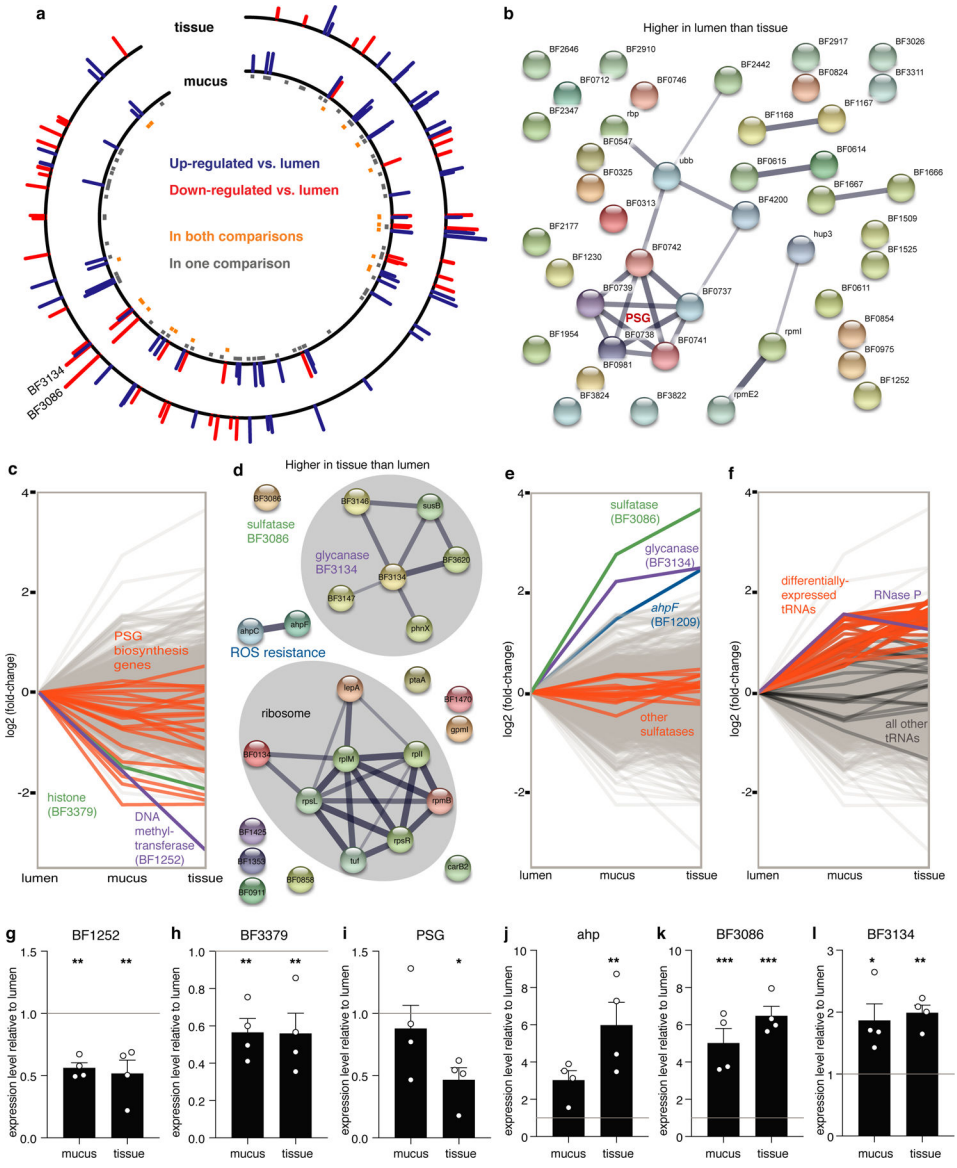



**Fig. 2 |. *B. fragilis* gene expression across gut microenvironments.**

**a**, Genes differentially expressed between the lumen and mucus (inner circle) and lumen and tissue (outer circle) during mono-colonization. Blue and red bars indicate magnitude of fold-change in gene expression. Squares in the innermost ring indicate genes differentially expressed in both (orange) or only one of the two comparisons (grey). **b** and **d**, STRING[31] network analyses of genes more highly expressed in the lumen compared to tissue (**b**) and those more highly expressed in the tissue compared to lumen (**d**). The thickness of connecting lines indicates confidence in relationships between genes and colors are arbitrary. **c**, **e**, and **f**, Fold change in expression of individual genes in mucus and tissue, with respect to the lumen. Indicated genes are highlighted within each figure: **c**, PSG biosynthesis genes and genes with lumen-specific expression patterns, **e**, genes with the tissue-specific expression patterns and all annotated sulfatases, **f**, tRNAs and the tRNA processing ribozyme, ribonuclease P. **g-l**, Quantitative real-time PCR (ΔΔCt method normalized to

*gyrB*) confirms differential expression of genes identified by hsRNA-Seq (grey horizontal line at y = 1 represents equal expression to lumen). Genes more highly expressed in the lumen: **g**, BF1252 DNA methyltransferase, **h**, BF3379 histone, **i**, polysaccharide G flippase. Genes more highly expressed in mucus and tissue: **j**, alkyl-hydroperoxide reductase, **k**, BF3086 sulfatase, and **l**, BF3134 glycosyl hydrolase (mean and standard error, Dunnett ANOVA, n = 4 animals). Fold-change between sample sites was quantified within each mouse individually (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

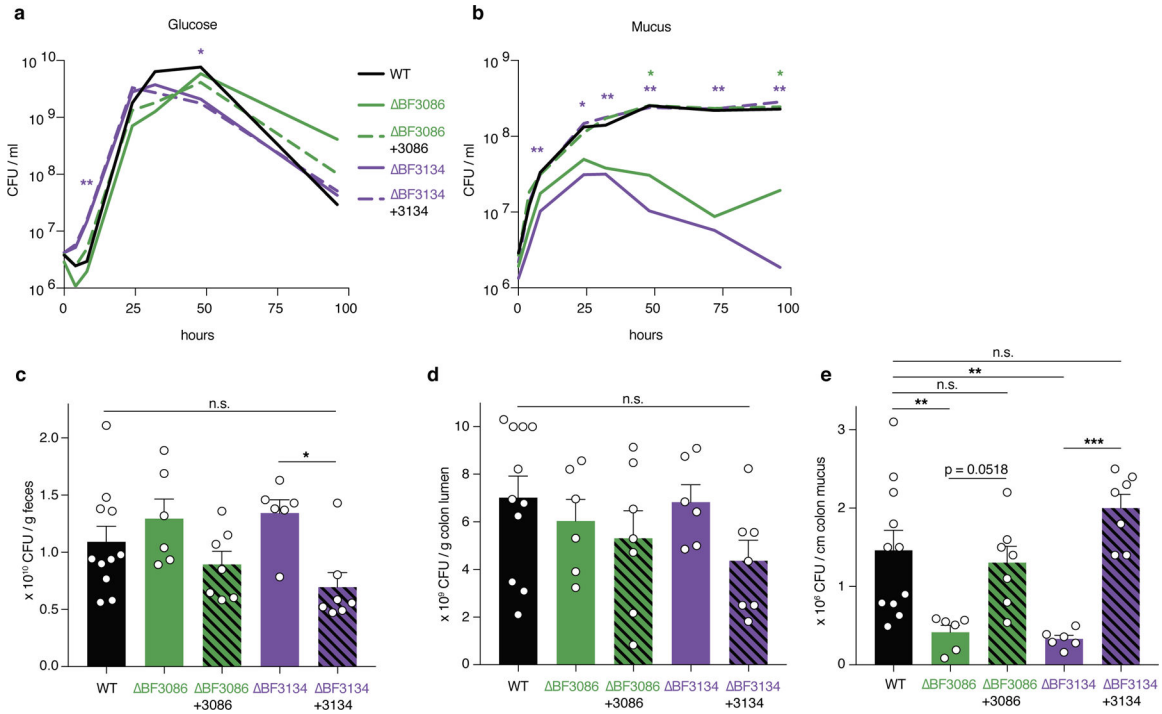



**Fig. 3 |. Discovery of candidate mucosal colonization factors in *B. fragilis*.**

**a**, Growth in minimal glucose media of *B. fragilis*, in-frame deletion mutants of BF3086 and BF3134, and these mutants complemented by expression on a plasmid (geometric mean, n = 4 independent cultures, representative of two independent experiments). **b**, Growth of *B. fragilis* strains in defined minimal media with mouse mucus from germ-free mice as the sole carbon source (geometric mean, Dunnett 2-way ANOVA on log-transformed data, n = 4 independent preparations of mucus from animals, representative of two independent experiments). **c**, Quantification of bacteria in feces, 4 weeks after mono-colonization with indicated strains of *B. fragilis*. Mice were sacrificed to quantify bacteria in the **d**, colonic lumen and **e**, colonic mucus (mean and standard error, Sidak ANOVA, n = 11, 6, 6, 7, 7 animals, pooled from two experiments) (all panels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

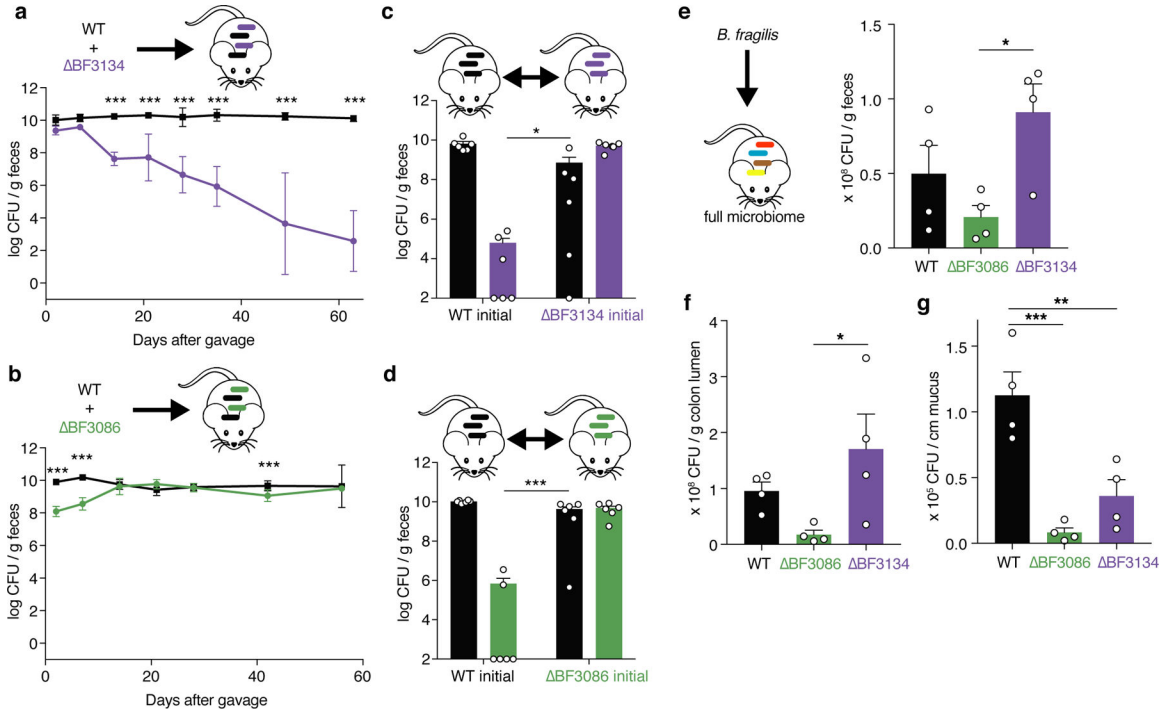



**Fig. 4 |. BF3086 and BF3134 promote robustness of *B. fragilis* colonization.**

**a**, Germ-free mice were orally gavaged with a 1:1 mixture of $10^8$ colony-forming units (CFU) each of wild-type and ΔBF3134 *B. fragilis*. Colonization of each strain was monitored using antibiotic resistance during microbiological plating (geometric mean and 95% C.I., Sidak 2-way ANOVA of log-transformed data, n = 4 animals, representative of two independent experiments) **b**, Germ-free mice were orally gavaged with a 1:1 mixture of $10^8$ colony-forming units (CFU) each of wild-type and ΔBF3086 *B. fragilis*. Colonization of each strain was monitored using antibiotic resistance during microbiological plating (geometric mean and 95% C.I., Sidak 2-way ANOVA of log-transformed data, n = 4 animals, representative of two independent experiments). **c**, Horizontal transmission between pairs of mice that had been mono-colonized for 4 weeks with either wild-type *B. fragilis* (WT initial) or ΔBF3134 (BF3134 initial). Two weeks after co-housing and separating the mice, levels of initial and invading strains are shown (mean and standard error, Tukey 2-way ANOVA of log-transformed data, n = 6 animals pooled from two independent experiments). **d**, Horizontal transmission with wild-type and ΔBF3086-colonized mice (mean and standard error, Tukey 2-way ANOVA of log-transformed data, n = 6 animals pooled from two independent experiments). **e**, Colonization by indicated *B. fragilis* strains in mice with a complex microbiome, 2 weeks after a single gavage of $10^9$ CFU. Mice were sacrificed to assess **f**, colon lumen and **g**, colon mucus colonization (mean and standard error, Tukey ANOVA, n = 4 animals, representative of two independent experiments). (all panels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).