Epigenetic phase variation in the gut microbiome enhances bacterial adaptation

Mi Ni[1], Yu Fan[1], Yujie Liu[1*], Yangmei Li[1*], Wanjin Qiao[1*], Lauren E. Davey[2,3*], Xue-Song Zhang[4*], Magdalena Ksiezarek[1], Edward A. Mead[1], Alan Tourancheau[1], Wenyan Jiang[1], Martin J. Blaser[4], Raphael H. Valdivia[2], and Gang Fang[1#]

[1]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
[2]Department of Integrative Immunobiology, Duke Microbiome Center, Duke University School of Medicine, Durham, NC, USA.
[3]Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada.
[4]Center for Advanced Biotechnology and Medicine, Rutgers University, New Brunswick, NJ, USA.

** These authors contributed equally to this work.*

[#] Address correspondence to gang.fang@mssm.edu

Abstract

The human gut microbiome within the gastrointestinal tract continuously adapts to variations in diet, medications, and host physiology. A strategy for bacterial genetic adaptation is epigenetic phase variation (ePV) mediated by bacterial DNA methylation, which can regulate gene expression, enhance clonal heterogeneity, and enable a single bacterial strain to exhibit variable phenotypic states. Genome-wide and site-specific ePVs have been characterized in human pathogens' antigenic variation and virulence factor production. However, the role of ePV in facilitating adaptation within the human microbiome remains poorly understood. Here, we comprehensively cataloged genome-wide and site-specific ePV in human infant and adult gut microbiomes. First, using long-read metagenomic sequencing, we detected genome-wide ePV mediated by complex structural variations of DNA methyltransferases, highlighting those associated with antibiotics or fecal microbiota transplantation. Second, we analyzed a collection of public short-read metagenomic sequencing datasets, uncovering a great prevalence of genome-wide ePV in the human gut microbiome. Third, we quantitatively detected site-specific ePVs using single-molecule methylation analysis to identify dynamic variation associated with antibiotic treatment or probiotic engraftment. Finally, we performed an in-depth assessment of

an *Akkermansia muciniphila* isolate from an infant, highlighting that ePVs can regulate gene expression and enhance the bacterial adaptive capacity by employing a bet-hedging strategy to increase tolerance to differing antibiotics. Our findings indicate that epigenetic modifications are a common strategy used by gut bacteria to adapt to the fluctuating environment.

Introduction

The human gut microbiome is continuously subjected to a myriad of changes including those due to changes in diet, medications, and physiological states.[1-3] In response to these alterations, the gut microbiome shows a remarkable adaptive capacity.[4-6] Characterizing this adaptive capacity is crucial for understanding the dynamic relationship between the gut microbiome and host physiology, especially in the context of human health and disease. While many microbiome studies are focused on profiles of microbial species and the relevant functions across different human diseases,[4,5,7-11] great progress has also been made with regard to the molecular mechanisms that underpin the microbiome's adaptive capacity including point mutations,[12-14] copy number variation,[15,16] structural variation (SV),[17,18] gene amplification,[19,20] invertible promoters[21] and intragenic inversions.[22] These genomic variations provide a single bacterial strain with the plasticity of expressing alternative, heritable, and reversible phenotypic states, including phase variation.[23,24] This mechanism increases the chances of bacterial survival and adaptation by producing a genetically diverse population that can withstand multiple stresses.[21,22]

Another mechanism of bacterial adaptation involves DNA methylation, which can regulate gene expression, enhance clonal heterogeneity and mediate epigenetic phase variation (ePV, intra-strain epigenetic variation that leads to phenotypic differences, **Fig. 1A**).[25-28]

Bacterial DNA contains three primary types of DNA methylation: N6-methyladenine (6mA), N4-methylcytosine (4mC), and 5-methylcytosine (5mC).[29] These DNA modifications are catalyzed by DNA methyltransferases (MTases) that are either part of restriction-modification systems or orphan enzymes without cognate restriction genes.[25,29] ePVs driven by DNA methylation have been characterized in multiple human pathogens, where it facilitates processes such as antigenic variation and virulence factor production,[30-32] which can impact how the bacterium interacts with its environment and host.[28,33-35]

ePVs characterized across diverse human pathogens can be categorized into two types: genome-wide or site-specific.[36-39] A genome-wide ePV is usually mediated by genetic phase variation affecting an MTase or its corresponding specificity unit (which determines the recognition specificity, or sequence motif, of an MTase). In a diverse population, DNA MTase phase variation results in genome-wide activity of the MTase that alternates between ON/OFF states, leading to varying levels of methylation across the population, typically approaching either ~100% or ~0% methylation of its target motif in each bacterial cell (**Fig. 1B**).[33,38,39] The phase variation of the specificity unit associated with a DNA MTase can create two or more alternative genome-wide recognition motifs (**Fig. 1B**).[30,40] Both types of genome-wide ePVs can affect the expression of multiple downstream genes via epigenetic regulation, which provide greater flexibility than single-gene phase variation in creating transcriptional heterogeneity (e.g. *Streptococcus pneumoniae,*[41] *Haemophilus influenzae,*[42] *Helicobacter pylori,*[43] *etc.*) and selective advantages to bacterial pathogens (e.g. *Mycobacterium tuberculosis*[44]). In contrast to genome-wide ePVs, site-specific ePVs involve a small number (one or few) methylation sites and are usually driven by the competitive binding between a DNA MTase and other DNA binding proteins such as a transcription factor (TF) (**Fig. 1C**). Two well-characterized examples are the ePV of the *pap* operon[45] and the *agn43* gene[46] in *Escherichia coli*, where the 6mA MTase targeting 5'-GATC-3' sites compete with Lrp and OxyR, respectively. These site-specific

epigenetic states are inheritable and reversible and have the advantage of mediating phase variation without altered genomic sequences, thereby mitigating mutation costs, for example, associated with antibiotic resistance.[47] Both genome-wide and site-specific ePVs facilitate the manifestation of diverse phenotypic states within a single bacterial strain population, enhancing the ability to adapt to changing conditions.[48]

DNA methylation is highly abundant and diverse among the species constituting the human gut microbiome,[49] yet the role of ePV and its impact on bacterial adaptation in the human gut remains largely unexplored. In this study, we aimed to address three major questions: (i) How prevalent are genome-wide and site-specific ePVs in the human gut microbiome? (ii) What are the functions of genes regulated by ePVs? (iii) How does an ePV enhance gut bacteria's adaptation to environmental (e.g. antibiotics) and individual-specific (e.g. fecal microbiota transplantation, FMT) selective pressures?

To address these questions, we systematically investigated the occurrence and implications of ePV in the human infant and adult gut microbiome (**Fig. 1D-E**). Using long-read metagenomic and meta-epigenomic analysis, we detected genome-wide ePVs mediated by complex SVs in DNA MTases and their specificity units. We observed intra-strain ePVs associated with antibiotic use within individuals, as well as intra-strain ePVs between FMT donors and recipients. Furthermore, we analyzed a collection of publicly available short-read metagenomic sequencing datasets (2,345 unique samples) to estimate the prevalence of genome-wide ePVs mediated by structural variations in DNA MTases and their specificity units. In addition, we identified site-specific ePVs associated with probiotic engraftment and antibiotic treatment using single-molecule quantitative methylation analysis. Importantly, we observed that an *Akkermansia muciniphila* strain in an infant gut microbiome had an ePV detected following a

course of amoxicillin (AMX); we recapitulated this shift *in vitro* and functionally characterized how this ePV enhances the adaptive capacity of the *A. muciniphila* strain in response to different antibiotics classes (**Fig. 1E**).

Results

Genome-wide epigenetic phase variations (ePVs) in the gut microbiome

To detect genome-wide ePVs, we performed joint metagenomic and meta-epigenomic analysis using long-read DNA sequencing (**Fig. S1A**; **Methods**) on both publicly available long-read metagenomic sequencing data and new data we generated for this study (**Fig. S1B**, **Table S1**). We focused on detecting structural variations (SVs) involving DNA MTases and/or their corresponding specificity units, including invertible repeats, and complex SVs spanning one or multiple coding regions (**Methods**). These SVs have been linked to genome-wide ePV in bacterial pathogens[41-44] but not in the gut microbiome. While short-read DNA sequencing can detect invertible repeats, long-read DNA sequencing can resolve complex SVs by assigning individual long reads to a different possible genome rearrangement built from each reference genome (**Fig. S1B**, **Methods**).[22,50-52] We used long-read mapping tools (pbalign / NGMLR[52]) with additional quality controls to ensure read homogeneity at genomic regions flanking the SVs that reflect intra-strain variation (**Methods**). For example, from the microbiome sample from a healthy adult,[53] we detected a complex SV within an MTase in *Phocaeicola vulgatus* (formerly *Bacteroides vulgatus*), mediated by three paired/matching inverted repeats, resulting in three possible genotypes (g1-g3; **Fig. 2A**). Individual long reads can be unambiguously assigned to each of the three genotypes, facilitating genotype-specific quantitation (**Fig. 2A**). Across the 221 long-read metagenomic samples analyzed, we detected 294 SVs involving DNA MTase and/or specificity units (**Table S2**), which encode 2-8 alternative genotypes based on long-read

mapping analysis (**Fig. 2B**, **Fig. S1C**). One of these SVs was detected from *Gemmiger qucibialis* [within class Clostridia] in the context of an FMT trio (donor: D1; donor 5 years: D1.5y; and recipient post-FMT: R1-postFMT; **Fig. 2C**). Specifically, this SV has two alternative genotypes (g1 and g2) which demonstrate different patterns when colonizing donor and recipient. While g1 is the dominant genotype in the *G. qucibialis* strain in the FMT donor sample (very stable over a five-year time window, **Table S3**), g2 becomes the dominant genotype in the recipient one year after FMT (**Fig. 2C**, **Methods**), suggesting differing host selective pressures between the FMT donor and recipient. We built on our recently described method for precise tracking of individual strains, performed rigorous long-read mapping analysis and strain-level *de novo* metagenomic assembly (**Fig. S2**, **Supplementary Data**) to confirm the SV in *G. qucibialis* was indeed from the same donor strain that stably engrafted the recipient post-FMT.[54]

Because PacBio sequencing data support direct detection of 6mA and 4mC, *de novo* methylation motif and quantitative methylation analysis at single molecule resolution (**Methods**),[29,55,56] we were able to directly assess ePVs mediated by the SVs of DNA MTase genes and/or their specificity units. We focused on nine fecal samples (from three infants, ages 1-3) collected as part of the Early Childhood Antibiotics and the Microbiome (ECAM) study.[3] The three infants were selected based on their exposure to amoxicillin treatment, fecal samples availability both before and after administration of the antibiotics, which had substantial impact on microbiome composition in each case (**Fig. 2D-2F**). To illustrate how the SV of an MTase and specificity unit mediated a genome-wide ePV, we describe one example detected from a *Bacteroides xylanisolvens* strain in Infant 1. Two pairs of inverted repeats in the specificity unit create four possible genotypes via genome rearrangement, with four expected methylation recognition motifs (**Fig. 2G**). For these four genotypes, *de novo* methylation analysis resolved specific methylation motifs, and ePVs estimated from methylation analysis correlate with the genotype frequencies estimated from long-read mapping (**Fig. 2G**; **Methods**). Across the three

infants, we observed bacterial strains whose methylation motifs were stable over time (**Fig. 2H-2J**), as well as those with genome-wide ePV in at least one methylation motif following exposure to amoxicillin (**Fig. 2H-2J**), which suggests possible antibiotics-induced selective pressure of genome-wide ePVs.

Although long-read sequencing is powerful for metagenomic SV analysis, most existing microbiome studies were based on short-read sequencing platforms. While short-read sequencing has limited ability to resolve complex SVs, they are reliable for detecting relatively simple SVs (e.g. with only two alternative genotypes).[21,22] Building on PhaseFinder,[21] we adapted a pipeline for detecting putative ePVs from short-read metagenomic shotgun sequencing data (**Fig. 3A-3B; Methods**) and applied it to 2,345 unique metagenomic samples from nine studies including the Human Microbiome Project[57] and additional cohorts (**Table S4**). Across 1,308 unique species, we detected 2,556 putative genome-wide ePVs with at least one alternative genotype based on SVs in MTases or specificity units (**Fig. 3C**). Four orders (Clostridiales, Bacteroidales, Enterobacterales, and Bifidobacteriales) were enriched for ePVs (**Fig. 3C**; **Fig. S3A-S3B**), although we observed broad prevalence of potential ePVs (inverted repeats in MTase or specificity units, with neighboring invertase) across many other orders (**Fig. 3C**). Among the detected putative ePVs, some are associated with the use of antibiotics. For example, a *B. uniformis* strain from an infant cohort[58] has a putative ePV with two alternative MTase genotypes, which had clear changes upon amoxicillin treatment, across four infants (**Fig. 3A**, **3D**). The MTase gene and SV involved in this ePV was detected in 23 of 153 complete *B. uniforms* genomes available at NCBI, with each of these 23 strains exhibiting one of the two ePV genotypes (**Fig. 3E**).

Site-specific epigenetic phase variations (ePV) in the gut microbiome

Differing from the genome-wide ePVs resulting from changes in MTase activity or specificity, site-specific ePVs involve only one or several methylation sites, and are usually driven by the site-specific binding competition between a DNA MTase and another DNA binding protein such as a TF (**Fig. 4A**). For a specific sequence motif recognized by a DNA MTase in a bacterial strain, while most motif sites are stably methylated (~100%) across bacterial cells at a certain condition, a few motif sites are only methylated in a small percentage (for example, <50%) of bacterial cells (**Fig 4A**). We built on previous work for PacBio DNA sequencing-based single-molecule methylation analysis,[29,59-61] now with enhanced quality control to ensure site-specific ePV reflects variation within a single strain rather than a mixture of different strains (**Methods**). Methods evaluation supported that the pipeline can reliably quantify the extent of partial methylation of individual methylation motif sites (**Fig. S4A**).

We performed quantitative site-specific ePV detection on 12 ECAM infant samples and **8** FMT fecal samples, and discovered 885 site-specific ePVs with < 50% site-specific methylation across 32 strains (representing 27 unique species), including some motif sites that were mostly (<25% or even <10%) unmethylated (boxplot in **Fig. 4B**; **Fig. S4B**). The ePV sites were classified as either intergenic or intra-genic based on their positions relative to protein-coding genes. Of the 885 site-specific ePVs, 174 (19.7%) are in intergenic regions (**Fig. S4C**). This enrichment of ePV sites in intergenic regions (Chi-square p < 10e-5) is consistent with the proposed mechanism of site-specific competitive binding between MTases and transcription factors (or repressors),[45,46] although intragenic DNA methylation might also be associated with the regulation of gene expression in bacteria.[29,31] The genes flanking site-specific ePVs have functions across diverse categories related to metabolic pathways, cellular processes, and

environmental processing including several genes associated with host interactions, antibiotic resistance, and biofilm formation (**Fig. 4C**).

To determine whether certain site-specific ePVs are associated with the selective pressure of environmental factors, we examined the 12 ECAM cohort fecal samples from infants with amoxicillin exposures as described above. We detected ePVs with >40% differences in their methylation status between two consecutive sampling time points. In Infant 3 (**Fig. 4D**), we found three site-specific ePVs in *Bacteroides ovatus* (**Fig. S4D**) whose methylation levels significantly increased upon the infant's treatment with amoxicillin (**Fig. 4E-4F**). The first ePV is in the 5' upstream of the essential sigma factor σ^{70}, which regulates genes involved in basic cellular processes including protein synthesis, energy production, and cell wall biosynthesis.[62] The second ePV is located near the genes encoding SusC/SusD family outer membrane proteins, which have been identified in various *Bacteroides* species as phase-variable factors.[63,64] The third ePV is located in the *phoB* gene, which encodes a response regulator protein essential for the PhoBR two-component signal transduction system. This system regulates the expression of multiple genes under phosphate-limiting conditions[65] and plays a critical role in biofilm formation and environmental stress response.[66] The presence of site-specific ePVs in these genes likely contributes to transcriptional heterogeneity, providing additional adaptive capacity under amoxicillin exposure. Similarly, we examined the FMT fecal samples as described above. Between the FMT donor and post-FMT recipient samples **(Fig. 4G)**, we detected ePV sites with differentially methylation ≥ 50% (**Fig. 4H**). In donor 2, we found four site-specific ePVs in *Phocaeicola plebeius* whose methylation levels had major decreases 5 years after FMT (**Fig. 4H-4I**, **Fig. S4E**). One of the four sites is located in the *MBL* gene encoding a β-lactamase, which are emerging as the most notable resistance determinants in Enterobacteriaceae[67,68], while another site is located between *susD* and *TonB*, both were reported to be phase variable sites in several *Bacteroides* species.[54,63,64]

To determine whether site-specific ePVs may be affected by selective pressure in the gut environment differing from *in vitro* culture, we fed a commercially available probiotic capsule [Glucose Control, Pendulum] with five strains (*A. muciniphila*, *Bifidobacterium longum*, *Clostridium beijerinckii*, *Clostridium butyricum*, and *Anaerobutyricum hallii*) to Germ-free B6 mice and collected fecal samples at day 7 (**Fig. 4J**). PacBio long-read metagenomic sequencing data showed that *A. muciniphila* stably colonized all four tested mice (**Fig. S5A**). Site-specific methylation analysis found 38 ePVs across 15 unique RA6mATTY sites in the probiotic *A. muciniphila* strain, showing >50% differential methylation between the inoculum (capsule) and the fecal (*in vivo*) samples (**Fig. 4J-4L**). These findings were further supported by independent validation based on restriction digestion (**Fig. S5B**). In contrast, no ePV events were observed at A6mACN$_6$TAC sites (**Fig. S5C**). Five of the 16 genes flanking the 15 unique RA6mATTY sites exhibited host colonization defects: *hxlR*, *MASP*, *GLUL*, *fhaB* and Amuc_0952.[69] Additionally, another gene identified exclusively in this study, *pmp10*, is also associated with bacteria-host interaction[70,71] (**Fig 4L**). Some other genes involved in these site-specific ePVs are known to be associated with stress responses that help bacteria adapt to challenging conditions (*prmA*[72], *PAP*[73], **Fig. 4L**). The selection for *A. muciniphila* cells with methylated ePV states of these genes suggest they are associated with the engraftment of the probiotic strain in the gut environment.

Genes, functions and selective advantage mediated by a genome-wide ePV in an *A. muciniphila* strain

Next, we focused on *A. muciniphila,* a promising next-generation probiotic associated with several positive health outcomes and drug responses.[74-78] This species was selected for its

prevalent genome-wide and site-specific ePVs (**Fig. 2I**, **4K-4L**) and its response to amoxicillin exposure observed in an infant fecal sample (**Fig. 5A**). This strain encodes a MTase specificity unit with three nested inverted repeats, creating eight different genotypes (g1-g8) and six unique methylation sequence recognition motifs, corresponding to six epigenetic phases m1–m6 (see **Fig. 5A; Methods**). Upon the use of amoxicillin, the subpopulation of g2 increased from 11.1% to 26.5% (**Fig. 5A; Methods**), consistent with a genome-wide loss of 6mA at m1:GG6mAN_9TGAA sites and genome-wide gain of 6mA at m2:G6mACN$_7$RTTC sites (**Fig. S6A**) due to the genome-wide change in MTase sequence recognition specificity.

To characterize the impact of this genome-wide ePV on bacterial fitness, we first isolated the *A. muciniphila* strain (**Methods**) and selected a clone (hereafter referred to as Akk_ECAM) that is dominated by the m1 genotype representing the same ePV state as in the infant's fecal sample before amoxicillin exposure (**Fig. S6B**). After ten passages of sub-cultures in liquid media with amoxicillin (**Fig. 5B**; **Fig. S6C**; **Methods**), we observed gradual population-level change from ePV phase m1 to m2 (**Fig. 5C**) along with genome-wide methylation changes from GG6mAN_9TGAA to G6mACN$_7$RTTC (**Fig. 5D**). These gradual epigenomic changes were not observed in controls without amoxicillin (**Fig. 5B-D**). The directionality of the changes (the m2 ePV state selected by amoxicillin) is consistent with the observations from the infant's fecal samples (**Fig. 5A**). However, in two additional clones, we did not observe the same changes in genome-wide ePV, suggesting that *A. muciniphila* may adapt to amoxicillin through multiple mechanisms.[79] Therefore, we performed a more comprehensive characterization of this ePV across a larger number (n=46) of m1 clones (**Fig. 5E**). Consistent with our initial observations, there was a variability among these 46 clones in terms of how the ePV state changed during the nine passages in the presence of amoxicillin: three clones changed to m2 and seven changed to m3 (**Fig. 5F**); the remaining clones remained as m1. In contrast, in the absence of amoxicillin, all clones remained at m1 after the passages. With the initial abundance of the non-m1 ePV

state < 0.1% at the beginning of the passage, that 10 out of 46 clones changed their ePV states in the presence of amoxicillin is statistically significant (Fisher's exact test p = 0.0011; **Fig. 5G**), suggesting that this genome-wide ePV may serve as one of several mechanisms that enhance the tolerance of the *A. muciniphila* strain to amoxicillin. Because m1 preserves methylation motif GG6mAN_9TGAA while m2 and m3 both lose the methylation motif (**Fig. 5A**), we hypothesized that the loss of GG6mAN_9TGAA methylation in m2 and m3 influences the expression of certain *A. muciniphila* genes that increases its tolerance to amoxicillin.

To test this hypothesis, we selected nine clones in total with preserved (m1) or lost (m2, m3) GG6mAN_9TGAA methylation and used RNA-seq to profile their gene expression in the last two passages in the presence of amoxicillin (**Fig. S7A**). The RNA-seq data analysis showed that the global gene expression of the m1 clones and m2/m3 clones have subtle but consistent differences (**Fig. 5H**). Differential gene expression analysis identified a single gene [Amuc_0914 homolog, herein named *mucC*, for *muciniphila* gene from cupin family] that is significantly (negative binomial test, p = 3.76E-07) and consistently up-regulated when GG6mAN_9TGAA methylation is lost in m2 and m3 clones (**Fig. 5I**, **Fig. S7B**). Consistently, the promoter of *mucC* contains a GG6mAN_9TGAA methylation site between its -35 region and translation initiation site (**Fig. 5J**) suggesting methylation-mediated regulation of gene expression. To test if some genetic variations might be selected by amoxicillin, we also performed whole genome shotgun sequencing on m2/m3 colonies: although we found several genetic variations across the clones, none were associated with the amoxicillin treatment (**Fig. S7C**). These findings suggest that the loss of GG6mAN_9TGAA methylation in the promoter of *mucC* gene is mediated by the genome-wide ePV changing from m1 to m2/m3, upregulating *mucC* expression, which enhances the tolerance of this *A. muciniphila* to amoxicillin.

The *mucC* gene is highly conserved across *A. muciniphila* strains (**Fig. S8A**, **Table S5**) encoding a small protein characterized by a cupin fold,[80] which consists of a conserved β-barrel structure (AlphaFold prediction,[81,82] **Fig. S8B**). MucC homologs are also found in other bacterial species (**Fig. S8C**). In *A. muciniphila*, *mucC* is immediately adjacent to Ribonuclease R (rnr), a member of the RNase II superfamily (**Fig. S8D**) that hydrolyze RNA in the 3' - 5' direction. RNase R is associated with selective mRNA degradation, particularly of mRNAs that lack a stop codon in bacteria[83] and are expressed under stress conditions.[84] Because the genetic tools of *A. muciniphila* are still limited[69], we expressed *mucC in trans* in *E. coli* cells (pZE21- *mucC*) and confirmed its protein expression by mass spectrometry (**Methods**; **Fig. S9A-S9D**). By comparing *mucC*$^{+}$ *E. coli* (pZE21- *mucC*) with negative controls (empty plasmid, pZE21), *mucC* expression conferred significantly greater tolerance to amoxicillin based on their MICs and survival rate (**Fig.5K**).

Finally, because ePV enhances non-genetic clonal heterogeneity and empowers the adaptive capacity of single cells in a population,[39,41-44] we hypothesized that while the ePV state with high *mucC* expression enhances tolerance to amoxicillin, other conditions may favor an ePV state with low *mucC* expression. High basal level of *mucC* expression doesn't create a disadvantageous fitness cost *in vitro*, as we found that the removal of amoxicillin didn't select for the m1 ePV state with lower *mucC* expression (**Fig. S10**). However, after testing several additional stress conditions, we discovered that ciprofloxacin selects the m1 ePV state with lower *mucC* expression (**Fig. 5L**, **Table S7**), demonstrating a bet hedging strategy of dynamic adaptation to different stresses. Further examination led to the observation that the selection of m1 epigenetic phase (low *mucC* expression) by ciprofloxacin varied between clones and replicates (**Fig. S11A**), which can be partially explained by additional SNPs (**Fig. S11B**) or structural variations (**Fig. S11C**). Ciprofloxacin is a fluoroquinolone antibiotic that targets bacterial DNA gyrase and topoisomerase IV, leading to double-stranded DNA breaks. This DNA

damage can trigger the bacterial SOS response, resulting in increased mutation rates and the potential emergence of resistance-conferring mutations. Nevertheless, our findings suggest that ePV provides an alternative pathway for the *A. muciniphila* strain to tolerate ciprofloxacin (**Fig. S11A**). These observations suggest that the *A. muciniphila* strain exhibits clonal heterogeneity involving both genetic and epigenetic variations, which collectively enhance its adaptation to antibiotics.

Overall, our findings indicate that a complex ePV detected in an *A. muciniphila* strain that originated from an infant gut microbiome enhances clonal heterogeneity via regulating *mucC* gene expression with several ePV states associated with high *mucC* expression selected by amoxicillin and low *mucC* expression selected by ciprofloxacin. The reversible selection by two different antibiotics highlights the advantage of ePV to empower the same *A. muciniphila* strain to epigenetically adapt to different stress conditions. Noteworthy, this evolutionary strategy, where a small subpopulation generates ePV heterogeneity sufficient to increase population-level adaptation across various stress conditions, underscores the adaptive advantage at the population level rather than individual cells.

Discussion

Our study highlights the prevalence of ePV across diverse bacterial species within the gut microbiota, underscoring the role of DNA methylation in enhancing clonal heterogeneity and enabling strains to adapt to environmental stressors such as antibiotic treatments or when colonizing new hosts after FMT.

We leveraged long-read metagenomic data to simultaneously detect complex SVs and DNA methylation, providing a detailed view of the epigenetic landscape of bacteria within gastrointestinal tract. Furthermore, our analysis of a collection of short-read metagenomic data has highlighted SVs near MTase and specificity units, suggesting the broad presence of ePV in various human cohorts. As long-read sequencing is more increasingly used in microbiome studies, we anticipate that our findings will encourage a new dimension to characterize how gut bacteria adapt to various stress conditions and host interactions through epigenetic mechanisms. The genes associated with ePVs in our analyses suggest that the biological functions mediated by ePV are broad, impacting microbe-host interactions, antibiotic resistance, and the metabolic capacities of gut bacteria. Importantly, the gene functions implicated by ePVs require further investigation to elucidate their mechanistic contributions. In pursuit of this, we have conducted a detailed characterization of an ePV identified in an *Akkermansia muciniphila* strain, revealing a specific gene that is regulated by DNA methylation and enhances tolerance to amoxicillin. As genetic tools for *Akkermansia muciniphila* continue to evolve, future research will likely uncover additional mechanistic insights into how this and other ePVs detected play a role in the adaptability and functionality of this vital commensal species.

Compared to genetic phase variation, ePV offers several advantages in enhancing clonal heterogeneity. The reversibility of ePV, without altering DNA sequence or incurring mutation costs serves as an additional way for individual bacterial strains to adapt to diverse stresses. Furthermore, genome-wide ePV can simultaneously regulate multiple genes, which provides coordinated activation or inhibition of multiple functionally related genes.[30,33] As shown in bacterial pathogens, regulation by global epigenetic modification highlights an intermediate state toward long-term adaptation to stress conditions and interactions with the human immune

system.[26,79,85] The diverse functions of the genes associated with both genome-wide and site-specific ePVs suggest that ePVs may regulate critical processes of gut bacteria to enhance their adaptation and resilience in the changing environment.

The ePV characterized in the *A. muciniphila* strain that regulated differential expression of the *mucC* gene upon exposure to two common antibiotics, amoxicillin and ciprofloxacin **(Fig. 5)**, is a great example of a new epigenetic bet-hedging mechanism of commensal bacteria. Our results indicate that in the human gut, ePVs may help bacterial populations regain heterogeneity after bottlenecks encountered during colonization of a new host or severe perturbations due to antibiotic exposures. These ePV-driven regulatory mechanisms open new opportunities for targeted epigenetic interventions to improve desired functions of beneficial bacteria. For example, by manipulating ePV, we may strategically boost the resilience and functional capabilities of beneficial bacteria, which might improve the success rates of probiotic engraftment and the efficacy of treatments for microbiota-associated conditions.

Despite that in our study, the epigenetically regulated *mucC* expression was an important factor enhancing tolerance of the *A. muciniphila* strain to amoxicillin, we observe certain variations across all clones, which suggest that the epigenetic bet hedging is not the only mechanism for the *A. muciniphila* strain to adapt to antibiotic or other stressors. As reported in other studies, genetic mutations can also confer increased fitness in bacterial species. For instance, a subpopulation shift in specific clones has been reported in *B. thetaiotaomicron*, which was associated with resistance to oxidative stress through alterations in the sequence of the IctA protein.[86] Our finding highlights that ePV is one additional way to study clonal heterogeneity and should be considered within a systemic view that includes other genetic, transcriptional, and translational forms of regulations.[15,21,22,57]

Data availability

All sequencing data generated in this study have been submitted to NCBI SRA with the accession number PRJNA1207694, and will be made publicly available upon publication.

Acknowledgments

We thank the Icahn School of Medicine at Mount Sinai colleagues Jeremiah J. Faith and Ilaria Mogno for helping with the characterization of the adult FMT fecal samples. This work was supported by grant no. R35 GM139655 (G.F.) from the National Institutes of Health and seed fund (G.F.) from the Icahn School of Medicine at Mount Sinai. L.E.D. lab is supported by grant no. RGPIN-2022-03915. This work was also supported in part through the computational resources and staff expertise provided by the Department of Scientific Computing at the Icahn School of Medicine at Mount Sinai.

Author contributions

M.N. and G.F. designed the methods. M.N. performed most of the computational analysis. Y.F., M.K. and A.T. assisted with computational analysis. M.N., Y.Liu, Y. Li, W.Q., L.E.D., X-S.Z, M.K. and E.A.M. performed experiments. M.N., Y.F., Y.Liu, Y. Li, W.Q., L.E.D., X-S.Z, M.K., E.A.M., W.J., M.J.B., R.H.V. and G.F. contributed to data interpretation. M.N. and G.F. wrote the manuscript with input and comments from all coauthors. G.F. conceived and supervised the project.

Declaration of interests

R.H.V. is a co-founder of Bloom Science (San Diego, CA)

References

1. Nandi, T., Lee, I.R., Ghosh, T., Ng, A.H.Q., Chng, K.R., Li, C., Tan, Y.H., Lye, D., Barkham, T., Chen, S., et al. (2018). Gut microbiome recovery after antibiotic usage is mediated by specific bacterial species. bioRxiv. 10.1101/350470.
2. Cox, Laura M., Yamanishi, S., Sohn, J., Alekseyenko, Alexander V., Leung, Jacqueline M., Cho, I., Kim, Sungheon G., Li, H., Gao, Z., Mahana, D., et al. (2014). Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences. Cell *158*, 705-721. 10.1016/j.cell.2014.05.052.
3. Bokulich, N.A., Chung, J., Battaglia, T., Henderson, N., Jay, M., Li, H., D. Lieber, A., Wu, F., Perez-Perez, G.I., Chen, Y., et al. (2016). Antibiotics, birth mode, and diet shape microbiome maturation during early life. Sci. Transl. Med. *8*, 343ra382-343ra382. 10.1126/scitranslmed.aad7121.
4. Wirbel, J., Pyl, P.T., Kartal, E., Zych, K., Kashani, A., Milanese, A., Fleck, J.S., Voigt, A.Y., Palleja, A., Ponnudurai, R., et al. (2019). Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. Nat. Med. *25*, 679-689. 10.1038/s41591-019-0406-6.
5. Franzosa, E.A., Sirota-Madi, A., Avila-Pacheco, J., Fornelos, N., Haiser, H.J., Reinker, S., Vatanen, T., Hall, A.B., Mallick, H., McIver, L.J., et al. (2019). Gut microbiome structure and metabolic activity in inflammatory bowel disease. Nat. Microbiol. *4*, 293-305. 10.1038/s41564-018-0306-4.
6. Nagy-Szakal, D., and Kellermayer, R. (2011). The remarkable capacity for gut microbial and host interactions. Gut Microbes *2*, 178-182. 10.4161/gmic.2.3.16107.
7. Demir, M., Lang, S., Hartmann, P., Duan, Y., Martin, A., Miyamoto, Y., Bondareva, M., Zhang, X., Wang, Y., Kasper, P., et al. (2022). The fecal mycobiome in non-alcoholic fatty liver disease. J. Hepatol. *76*, 788-799. 10.1016/j.jhep.2021.11.029.
8. Cammann, D., Lu, Y.M., Cummings, M.J., Zhang, M.L., Cue, J.M., Do, J., Ebersole, J., Chen, X.N., Oh, E.C., Cummings, J.L., and Chen, J.C. (2023). Genetic correlations between Alzheimer's disease and gut microbiome genera. Sci. Rep. *13*, 5258. 10.1038/s41598-023-31730-5.
9. Ferreiro, A.L., Choi, J., Ryou, J., Newcomer, E.P., Thompson, R., Bollinger, R.M., Hall-Moore, C., Ndao, I.M., Sax, L., Benzinger, T.L.S., et al. (2023). Gut microbiome composition may be an indicator of preclinical Alzheimer's disease. Sci. Transl. Med. *15*, eabo2984. 10.1126/scitranslmed.abo2984.
10. Smith, M., Dai, A.Q., Ghilardi, G., Amelsberg, K.V., Devlin, S.M., Pajarillo, R., Slingerland, J.B., Beghi, S., Herrera, P.S., Giardina, P., et al. (2022). Gut microbiome correlates of response and toxicity following anti-CD19 CAR T cell therapy. Nat. Med. *28*, 713-+. 10.1038/s41591-022-01702-9.
11. Valles-Colomer, M., Blanco-Miguez, A., Manghi, P., Asnicar, F., Dubois, L., Golzato, D., Pinto, F., Cumbo, F., Huang, K.D., Manara, S., et al. (2023). The person-to-person transmission landscape of the gut and oral microbiomes. Nature *614*, 125-135. 10.1038/s41586-022-05620-1.
12. Zhu, X., Qin, J.Y., Tan, C.Y., and Ning, K. (2021). The seasonal changes of the gut microbiome of the population living in traditional lifestyles are represented by characteristic species-level and functional-level SNP enrichment patterns. BMC Genomics *22*, 83. 10.1186/s12864-021-07372-0.

13. Ma, C.C., Wasti, S., Huang, S., Zhang, Z., Mishra, R., Jiang, S.M., You, Z.K., Wu, Y.X., Chang, H.B., Wang, Y.Y., et al. (2020). The gut microbiome stability is altered by probiotic ingestion and improved by the continuous supplementation of galactooligosaccharide. Gut Microbes *12*. 10.1080/19490976.2020.1785252.
14. Zahavi, L., Lavon, A., Reicher, L., Shoer, S., Godneva, A., Leviatan, S., Rein, M., Weissbrod, O., Weinberger, A., and Segal, E. (2023). Bacterial SNPs in the human gut microbiome associate with host BMI. Nat. Med. 10.1038/s41591-023-02599-8.
15. Greenblum, S., Carr, R., and Borenstein, E. (2015). Extensive strain-level copy-number variation across human gut microbiome species. Cell *160*, 583-594. 10.1016/j.cell.2014.12.038.
16. Maddamsetti, R., Yao, Y., Wang, T., Gao, J., Huang, V.T., Hamrick, G.S., Son, H.-I., and You, L. (2024). Duplicated antibiotic resistance genes reveal ongoing selection and horizontal gene transfer in bacteria. Nat. Commun. *15*, 1449. 10.1038/s41467-024-45638-9.
17. Zeevi, D., Korem, T., Godneva, A., Bar, N., Kurilshikov, A., Lotan-Pompan, M., Weinberger, A., Fu, J., Wijmenga, C., Zhernakova, A., and Segal, E. (2019). Structural variation in the gut microbiome associates with host health. Nature *568*, 43-48. 10.1038/s41586-019-1065-y.
18. Zhernakova, D.V., Wang, D.M., Liu, L., Andreu-Sánchez, S., Zhang, Y., Ruiz-Moreno, A.J., Peng, H.R., Plomp, N., Del Castillo-Izquierdo, A., Gacesa, R., et al. (2024). Host genetic regulation of human gut microbial structural variation. Nature *625*. 10.1038/s41586-023-06893-w.
19. Xu, J., Mahowald, M.A., Ley, R.E., Lozupone, C.A., Hamady, M., Martens, E.C., Henrissat, B., Coutinho, P.M., Minx, P., Latreille, P., et al. (2007). Evolution of symbiotic bacteria in the distal human intestine. PLoS Biol. *5*, 1574-1586, e156. 10.1371/journal.pbio.0050156.
20. Levan, S.R., Stamnes, K.A., Lin, D.L., Panzer, A.R., Fukui, E., McCauley, K., Fujimura, K.E., McKean, M., Ownby, D.R., Zoratti, E.M., et al. (2019). Elevated faecal 12,13-diHOME concentration in neonates at high risk for asthma is produced by gut bacteria and impedes immune tolerance. Nat. Microbiol. *4*, 1851-1861. 10.1038/s41564-019-0498-2.
21. Jiang, X., Hall, A.B., Arthur, T.D., Plichta, D.R., Covington, C.T., Poyet, M., Crothers, J., Moses, P.L., Tolonen, A.C., Vlamakis, H., et al. (2019). Invertible promoters mediate bacterial phase variation, antibiotic resistance, and host adaptation in the gut. Science *363*, 181-187. 10.1126/science.aau5238.
22. Chanin, R.B., West, P.T., Wirbel, J., Gill, M.O., Green, G.Z.M., Park, R.M., Enright, N., Miklos, A.M., Hickey, A.S., Brooks, E.F., et al. (2024). Intragenic DNA inversions expand bacterial coding capacity. Nature. 10.1038/s41586-024-07970-4.
23. Jonsson, A.B., Nyberg, G., and Normark, S. (1991). Phase variation of gonococcal pili by frameshift mutation in pilC, a novel gene for pilus assembly. EMBO J. *10*, 477-488-488. 10.1002/j.1460-2075.1991.tb07970.x.
24. Segal, E., Billyard, E., So, M., Storzbach, S., and Meyer, T.F. (1985). Role of chromosomal rearrangement in *N. gonorrhoeae* pilus phase variation. Cell *40*, 293-300. 10.1016/0092-8674(85)90143-6.
25. Sánchez-Romero, M.A., and Casadesús, J. (2020). The bacterial epigenome. Nat. Rev. Microbiol. *18*, 7-20. 10.1038/s41579-019-0286-2.
26. Gopalan-Nair, R., Coissac, A., Legrand, L., Lopez-Roques, C., Pécrix, Y., Vandecasteele, C., Bouchez, O., Barlet, X., Lanois, A., Givaudan, A., et al. (2024). Changes in DNA methylation contribute to rapid adaptation in bacterial plant pathogen evolution. PLoS Biol. *22*, e3002792. 10.1371/journal.pbio.3002792.

27. Forni, D., Pozzoli, U., Mozzi, A., Cagliani, R., and Sironi, M. (2024). Depletion of CpG dinucleotides in bacterial genomes may represent an adaptation to high temperatures. NAR genom. bioinform. *6*, lqae088. 10.1093/nargab/lqae088.
28. Atack, J.M., Tan, A., Bakaletz, L.O., Jennings, M.P., and Seib, K.L. (2018). Phasevarions of bacterial pathogens: Methylomics sheds new light on old enemies. Trends Microbiol. *26*, 715-726. 10.1016/j.tim.2018.01.008.
29. Beaulaurier, J., Schadt, E.E., and Fang, G. (2019). Deciphering bacterial epigenomes using modern sequencing technologies. Nat. Rev. Genet. *20*, 157-172. 10.1038/s41576-018-0081-3.
30. Manso, A.S., Chai, M.H., Atack, J.M., Furi, L., De Ste Croix, M., Haigh, R., Trappetti, C., Ogunniyi, A.D., Shewell, L.K., Boitano, M., et al. (2014). A random six-phase switch regulates pneumococcal virulence via global epigenetic changes. Nat. Commun. *5*, 5055. 10.1038/ncomms6055.
31. Oliveira, P.H., Ribis, J.W., Garrett, E.M., Trzilova, D., Kim, A., Sekulovic, O., Mead, E.A., Pak, T., Zhu, S., Deikus, G., et al. (2020). Epigenomic characterization of *Clostridioides difficile* finds a conserved DNA methyltransferase that mediates sporulation and pathogenesis. Nat. Microbiol. *5*, 166-180. 10.1038/s41564-019-0613-4.
32. Nahar, N., Tram, G., Jen, F.E.-C., Phillips, Z.N., Weinert, Lucy A., Bossé, Janine T., Jabbari, Jafar S., Gouil, Q., Du, Mei R.M., Ritchie, Matthew E., et al. (2023). *Actinobacillus pleuropneumoniae* encodes multiple phase-variable DNA methyltransferases that control distinct phasevarions. Nucleic Acids Res. 10.1093/nar/gkad091.
33. Srikhanta, Y.N., Fox, K.L., and Jennings, M.P. (2010). The phasevarion: phase variation of type III DNA methyltransferases controls coordinated switching in multiple genes. Nat. Rev. Microbiol. *8*, 196-206. 10.1038/nrmicro2283.
34. Ben-Assa, N., Coyne, M.J., Fomenkov, A., Livny, J., Robins, William P., Muniesa, M., Carey, V., Carasso, S., Gefen, T., Jofre, J., et al. (2020). Analysis of a phase-variable restriction modification system of the human gut symbiont *Bacteroides fragilis*. Nucleic Acids Res. 10.1093/nar/gkaa824.
35. Huang, X.T., Wang, J.J., Li, J., Liu, Y.N., Liu, X., Li, Z.Y., Kurniyati, K., Deng, Y.J., Wang, G.L., Ralph, J.D., et al. (2020). Prevalence of phase variable epigenetic invertons among host-associated bacteria. Nucleic Acids Res. *48*, 11468-11485. 10.1093/nar/gkaa907.
36. Brunet, Y.R., Bernard, C.S., Gavioli, M., Lloubès, R., and Cascales, E. (2011). An epigenetic switch involving overlapping fur and DNA methylation optimizes expression of a type VI secretion gene cluster. PLoS Genet. *7*, e1002205. 10.1371/journal.pgen.1002205.
37. Lim, H.N., and van Oudenaarden, A. (2007). A multistep epigenetic switch enables the stable inheritance of DNA methylation states. Nat. Genet. *39*, 269-275. 10.1038/ng1956.
38. Jen, F.E.C., Scott, A.L., Tan, A., Seib, K.L., and Jennings, M.P. (2020). Random switching of the ModA11 type III DNA methyltransferase of *Neisseria meningitidis* regulates entner-doudoroff aldolase expression by a methylation change in the *eda* promoter region. J. Mol. Biol. *432*, 5835-5842. 10.1016/j.jmb.2020.08.024.
39. Seib, K.L., Srikhanta, Y.N., Atack, J.M., and Jennings, M.P. (2020). Epigenetic regulation of virulence and immunoevasion by phase-variable restriction-modification systems in bacterial pathogens. In Annual Review of Microbiology, Vol 74, 2020, S. Gottesman, ed. pp. 655-671. 10.1146/annurev-micro-090817-062346.
40. Furuta, Y., Namba-Fukuyo, H., Shibata, T.F., Nishiyama, T., Shigenobu, S., Suzuki, Y., Sugano, S., Hasebe, M., and Kobayashi, I. (2014). Methylome Diversification through Changes in DNA Methyltransferase Sequence Specificity. PLoS Genet. *10*, e1004272. 10.1371/journal.pgen.1004272.

41. Li, J., Li, J.W., Feng, Z.X., Wang, J.J., An, H.R., Liu, Y.N., Wang, Y., Wang, K.L., Zhang, X.G., Miao, Z., et al. (2016). Epigenetic switch driven by DNA inversions dictates phase variation in *Streptococcus pneumoniae*. PLoS Path. *12*, e1005762. 10.1371/journal.ppat.1005762.
42. De Bolle, X., Bayliss, C.D., Field, D., Van De Ven, T., Saunders, N.J., Hood, D.W., and Moxon, E.R. (2000). The length of a tetranucleotide repeat tract in *Haemophilus influenzae* determines the phase variation rate of a gene with homology to type III DNA methyltransferases. Mol. Microbiol. *35*, 211-222. 10.1046/j.1365-2958.2000.01701.x.
43. de Vries, N., Duinsbergen, D., Kuipers Ernst, J., Pot Raymond, G.J., Wiesenekker, P., Penn Charles, W., van Vliet Arnoud, H.M., Vandenbroucke-Grauls Christina, M.J.E., and Kusters Johannes, G. (2002). Transcriptional phase variation of a type iii restriction-modification system in *Helicobacter pylori*. J. Bacteriol. *184*, 6615-6623. 10.1128/jb.184.23.6615-6624.2002.
44. Li, H.-c., Chen, T., Yu, L., Guo, H.-x., Chen, L., Chen, Y.-h., Chen, M., Zhao, J., Yan, H.-m., Zhou, L., and Wang, W. (2020). Genome-wide DNA methylation and transcriptome and proteome changes in *Mycobacterium tuberculosis* with para-aminosalicylic acid resistance. Chem. Biol. Drug Des. *95*, 104-112. 10.1111/cbdd.13625.
45. Braaten, B.A., Blyn, L.B., Skinner, B.S., and Low, D.A. (1991). Evidence for a methylation-blocking factor (mbf) locus involved in pap pilus expression and phase variation in *Escherichia coli*. J. Bacteriol. *173*, 1789-1800. 10.1128/jb.173.5.1789-1800.1991.
46. Kaminska, R., and van der Woude, M.W. (2010). Establishing and maintaining sequestration of dam target sites for phase variation of *agn43* in *Escherichia coli*. J. Bacteriol. *192*, 1937-1945. 10.1128/jb.01629-09.
47. Ghosh, D., Veeraraghavan, B., Elangovan, R., and Vivekanandan, P. (2020). Antibiotic Resistance and Epigenetics: More to It than Meets the Eye. Antimicrob. Agents Chemother. *64*, 10.1128/aac.02225-02219.
48. Carasso, S., Zaatry, R., Hajjo, H., Kadosh-Kariti, D., Ben-Assa, N., Naddaf, R., Mandelbaum, N., Pressman, S., Chowers, Y., Gefen, T., et al. (2024). Inflammation and bacteriophages affect DNA inversion states and functionality of the gut microbiota. Cell Host Microbe. 10.1016/j.chom.2024.02.003.
49. Won, C., and Yim, S.S. (2024). Emerging methylation-based approaches in microbiome engineering. Biotechnology for Biofuels and Bioproducts *17*, 96. 10.1186/s13068-024-02529-x.
50. English, A.C., Salerno, W.J., and Reid, J.G. (2014). PBHoney: identifying genomic variants via long-read discordance and interrupted mapping. BMC Bioinformatics *15*, 180. 10.1186/1471-2105-15-180.
51. Cretu Stancu, M., van Roosmalen, M.J., Renkens, I., Nieboer, M.M., Middelkamp, S., de Ligt, J., Pregno, G., Giachino, D., Mandrile, G., Espejo Valle-Inclan, J., et al. (2017). Mapping and phasing of structural variation in patient genomes using nanopore sequencing. Nat. Commun. *8*, 1326. 10.1038/s41467-017-01343-4.
52. Sedlazeck, F.J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A., and Schatz, M.C. (2018). Accurate detection of complex structural variations using single-molecule sequencing. Nat. Methods *15*, 461-468. 10.1038/s41592-018-0001-7.
53. Chen, L., Zhao, N., Cao, J., Liu, X., Xu, J., Ma, Y., Yu, Y., Zhang, X., Zhang, W., Guan, X., et al. (2022). Short- and long-read metagenomics expand individualized structural variations in gut microbiomes. Nat. Commun. *13*, 3175. 10.1038/s41467-022-30857-9.
54. Fan, Y., Ni, M., Aggarwala, V., Mead, E.A., Ksiezarek, M., Kamm, M.A., Borody, T.J., Paramsothy, S., Kaakoush, N.O., Grinspan, A., et al. (2024). Long-read metagenomics empowers precise tracking of bacterial strains and their genomic changes after fecal microbiota transplantation. bioRxiv. 10.1101/2024.09.30.615906.

55. Beaulaurier, J., Zhang, X.-S., Zhu, S., Sebra, R., Rosenbluh, C., Deikus, G., Shen, N., Munera, D., Waldor, M.K., Chess, A., et al. (2015). Single molecule-level detection and long read-based phasing of epigenetic variations in bacterial methylomes. Nat. Commun. *6*, 7438. 10.1038/ncomms8438.
56. Kong, Y., Cao, L., Deikus, G., Fan, Y., Mead, E.A., Lai, W., Zhang, Y., Yong, R., Sebra, R., Wang, H., et al. (2022). Critical assessment of DNA adenine methylation in eukaryotes using quantitative deconvolution. Science *375*, 515-522. 10.1126/science.abe7489.
57. Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J.H., Chinwalla, A.T., Creasy, H.H., Earl, A.M., FitzGerald, M.G., Fulton, R.S., et al. (2012). Structure, function and diversity of the healthy human microbiome. Nature *486*, 207-214. 10.1038/nature11234.
58. Yassour, M., Vatanen, T., Siljander, H., Hämäläinen, A.-M., Härkönen, T., Ryhänen, S.J., Franzosa, E.A., Vlamakis, H., Huttenhower, C., Gevers, D., et al. (2016). Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. Sci. Transl. Med. *8*, 343ra381-343ra381. 10.1126/scitranslmed.aad0917.
59. Zhu, S., Beaulaurier, J., Deikus, G., Wu, T.P., Strahl, M., Hao, Z., Luo, G., Gregory, J.A., Chess, A., He, C., et al. (2018). Mapping and characterizing N6-methyladenine in eukaryotic genomes using single-molecule real-time sequencing. Genome Res. 10.1101/gr.231068.117.
60. Beaulaurier, J., Zhu, S., Deikus, G., Mogno, I., Zhang, X.-S., Davis-Richardson, A., Canepa, R., Triplett, E.W., Faith, J.J., Sebra, R., et al. (2018). Metagenomic binning and association of plasmids with bacterial host genomes using DNA methylation. Nat. Biotechnol. *36*, 61-69. 10.1038/nbt.4037.
61. Kong, Y., Mead, E.A., and Fang, G. (2023). Navigating the pitfalls of mapping DNA and RNA modifications. Nat. Rev. Genet. *24*, 363-381. 10.1038/s41576-022-00559-5.
62. Park, J., and Wang, H.H. (2021). Systematic dissection of σ^{70} sequence diversity and function in bacteria. Cell Reports *36*, 109590. 10.1016/j.celrep.2021.109590.
63. Nakayama-Imaohji, H., Hirakawa, H., Ichimura, M., Wakimoto, S., Kuhara, S., Hayashi, T., and Kuwahara, T. (2009). Identification of the site-specific DNA invertase responsible for the phase variation of susC/susD family outer membrane proteins in *Bacteroides fragilis*. J. Bacteriol. *191*, 6003-6011. 10.1128/jb.00687-09.
64. Joglekar, P., Sonnenburg, E.D., Higginbottom, S.K., Earle, K.A., Morland, C., Shapiro-Ward, S., Bolam, D.N., and Sonnenburg, J.L. (2018). Genetic variation of the susC/susD homologs from a polysaccharide utilization locus underlies divergent fructan specificities and functional adaptation in *Bacteroidesthetaiotaomicron* strains. Msphere *3*, e00185-18. 10.1128/mSphereDirect.00185-18.
65. Sultan, S.Z., Silva, A.J., and Benitez, J.A. (2010). The PhoB regulatory system modulates biofilm formation and stress response in El Tor biotype Vibrio cholerae. FEMS Microbiol. Lett. *302*, 22-31. 10.1111/j.1574-6968.2009.01837.x.
66. Cai, X.H., Yang, S.Y., Peng, Y.H., Tan, K.N., Xu, P., Wu, Z.H., Kwan, K.Y., and Jian, J.C. (2024). Regulation of PhoB on biofilm formation and hemolysin gene hlyA and ciaR of Streptococcus agalactiae. Vet. Microbiol. *289*, 109961. 10.1016/j.vetmic.2023.109961.
67. Logan, L.K., and Bonomo, R.A. (2016). Metallo-β-Lactamase (MBL)-producing Enterobacteriaceae in United States children. OFID *3*, ofw090. 10.1093/ofid/ofw090.
68. Cherak, Z., Loucif, L., Moussi, A., Bendjama, E., Benbouza, A., and Rolain, J.M. (2022). Emergence of metallo-β-lactamases and OXA-48 carbapenemase producing gram-negative bacteria in hospital wastewater in Algeria: A potential dissemination pathway into the environment. Microb. Drug Resist. *28*, 23-30. 10.1089/mdr.2020.0617.

69. Beck, L.C., Masi, A.C., Young, G.R., Vatanen, T., Lamb, C.A., Smith, R., Coxhead, J., Butler, A., Marsland, B.J., Embleton, N.D., et al. (2022). Strain-specific impacts of probiotics are a significant driver of gut microbiome development in very preterm infants. Nat. Microbiol. *7*, 1525-1535. 10.1038/s41564-022-01213-w.
70. Vandahl, B.B., Pedersen, A.S., Gevaert, K., Holm, A., Vandekerckhove, J., Christiansen, G., and Birkelund, S. (2002). The expression, processing and localization of polymorphic membrane proteins in Chlamydia pneumoniae strain CWL029. BMC Microbiol. *2*, 36. 10.1186/1471-2180-2-36.
71. Pedersen, A.S., Christiansen, G., and Birkelund, S. (2001). Differential expression of Pmp10 in cell culture infected with Chlamydia pneumoniae CWL029. FEMS Microbiol. Lett. *203*, 153-159. 10.1111/j.1574-6968.2001.tb10834.x.
72. Patrauchan Marianna, A., Miyazawa, D., LeBlanc Justin, C., Aiga, C., Florizone, C., Dosanjh, M., Davies, J., Eltis Lindsay, D., and Mohn William, W. (2012). Proteomic Analysis of Survival of Rhodococcus jostii RHA1 during Carbon Starvation. Appl. Environ. Microbiol. *78*, 6714-6725. 10.1128/AEM.01293-12.
73. Mishra, S., and Verma, S.K. (2021). Differential expression of proteins in Pseudomonas mendocina SMSKVR-3 under arsenate stress. J. Basic Microbiol. *61*, 351-361. 10.1002/jobm.202000671.
74. Ansaldo, E., Slayden, L.C., Ching, K.L., Koch, M.A., Wolf, N.K., Plichta, D.R., Brown, E.M., Graham, D.B., Xavier, R.J., Moon, J.J., and Barton, G.M. (2019). *Akkermansia muciniphila* induces intestinal adaptive immune responses during homeostasis. Science *364*, 1179-1184. doi:10.1126/science.aaw7479.
75. Bae, M., Cassilly, C.D., Liu, X., Park, S.-M., Tusi, B.K., Chen, X., Kwon, J., Filipčík, P., Bolze, A.S., Liu, Z., et al. (2022). Akkermansia muciniphila phospholipid induces homeostatic immune responses. Nature *608*, 168-173. 10.1038/s41586-022-04985-7.
76. Derosa, L., Routy, B., Thomas, A.M., Iebba, V., Zalcman, G., Friard, S., Mazieres, J., Audigier-Valette, C., Moro-Sibilot, D., Goldwasser, F., et al. (2022). Intestinal *Akkermansia muciniphila* predicts clinical response to PD-1 blockade in patients with advanced non-small-cell lung cancer. Nat. Med. *28*, 315-324. 10.1038/s41591-021-01655-5.
77. Yoon, H.S., Cho, C.H., Yun, M.S., Jang, S.J., You, H.J., Kim, J.-h., Han, D., Cha, K.H., Moon, S.H., Lee, K., et al. (2021). *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. 10.1038/s41564-021-00880-5.
78. Zheng, Y., Wang, T., Tu, X., Huang, Y., Zhang, H., Tan, D., Jiang, W., Cai, S., Zhao, P., Song, R., et al. (2019). Gut microbiome affects the response to anti-PD-1 immunotherapy in patients with hepatocellular carcinoma. JITC *7*, 193. 10.1186/s40425-019-0650-9.
79. Motta, S.S., Cluzel, P., and Aldana, M. (2015). Adaptive resistance in bacteria requires epigenetic inheritance, genetic noise, and cost of efflux pumps. PLoS One *10*, e0118464. 10.1371/journal.pone.0118464.
80. Dunwell, J.M., Purvis, A., and Khuri, S. (2004). Cupins: the most functionally diverse protein superfamily? Phytochemistry *65*, 7-17. 10.1016/j.phytochem.2003.08.016.
81. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature *630*, 493-500. 10.1038/s41586-024-07487-w.
82. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583-589. 10.1038/s41586-021-03819-2.

83. Cheng, Z.-F., and Deutscher, M.P. (2005). An important role for RNase R in mRNA decay. Mol. Cell *17*, 313-318. 10.1016/j.molcel.2004.11.048.
84. Chen, C., and Deutscher, M.P. (2005). Elevation of rnase r in response to multiple stress conditions. J. Biol. Chem. *280*, 34393-34396. 10.1074/jbc.C500333200.
85. Bryant, J.M., Brown, K.P., Burbaud, S., Everall, I., Belardinelli, J.M., Rodriguez-Rincon, D., Grogono, D.M., Peterson, C.M., Verma, D., Evans, I.E., et al. (2021). Stepwise pathogenic evolution of *Mycobacterium abscessus*. Science *372*, eabb8699. doi:10.1126/science.abb8699.
86. Tawk, C., Lim, B., Bencivenga-Barry, N.A., Lees, H.J., Ramos, R.J.F., Cross, J., and Goodman, A.L. (2023). Infection leaves a genetic and functional mark on the gut population of a commensal bacterium. Cell Host Microbe *31*, 811-826.e816. 10.1016/j.chom.2023.04.005.

Figure captions

Figure 1. An integrated framework for epigenetic phase variation (ePV) detection and function validation from the human gut microbiome

(**A**) A specific bacterial strain may have sub-populations that exhibit heterogeneous methylation patterns, leading to distinct gene expression and phenotypes. This may enhance the ability of a single strain to adapt to different stress conditions. (**B**) Schematic diagram of genome-wide ePV events, mediated by the structural variation located within MTase genes (ON/OFF states) or their specificity units (alters recognition motifs). (**C**) Schematic diagram of site-specific ePV events, resulting from the competition between MTase and transcription factors (TFs). (**D**) Prevalence and phylogenetic distribution of ePV were explored across various healthy and clinical-related cohorts. Phylogenetic tree is color-coded at phylum level. (**E**) Workflow for functional analysis of genes involved in ePV events. Bacterial strains carrying the ePVs were isolated from fecal samples. An *in vitro* passage assay was performed under conditions mimicking the *in vivo* environment, followed by whole-genome shotgun and targeted DNA sequencing. MTase genotypes were quantified and associated with treatment conditions. RNA sequencing was conducted to profile transcriptomic differences between genotypes. Differentially expressed genes (DEGs) were overexpressed further in native strains or model species, and phenotypic tests were carried out to confirm whether the altered phenotype was due to the expression of the gene of interest (GOI).

Figure 2. Long-read metagenomic sequencing deciphers complex ePV events

(**A**) An ePV observed in a *Phocaeicola vulgatus* strain from a healthy adult gut microbiome, mediated by three paired/matching inverted repeats (inverted repeats), resulting in three distinct MTase genotypes (g1-g3). The alignment plots show raw reads aligned with their corresponding reference genomes. The proportion of each genotype is displayed on the right side of the read alignment chart. (**B**) Additional ePV events were identified in various bacterial species. Gene diagrams illustrate the rearrangement of MTase gene fragments, while pie charts show the proportion of each genotype variant. (**C**) ePV events identified in an FMT cohort. The top panel presents metadata explaining sample and subject information (see **Table S3**). The bottom panel shows an ePV in a *Gemmiger qucibialis* strain, detected in both donor and post-FMT samples. (**D-F**) Metagenomic analysis of fecal samples from three infants in the ECAM cohort. The top panel presents the timeline of sample collection and amoxicillin (AMX) exposure. The bottom panel displays species-level relative abundance profiles, including only species with a sequencing depth > 10X. (**G**) An ePV event was observed in a *Bacteroides xylanisolvens* strain, mediated by two pairs of IR features, resulting in four specificity unit genotypes (g1-g4). The genotype was profiled in pre- and post-AMX treatment samples, alongside their corresponding methylome (m1-m4). AMX treatment suppressed the population of g2 (highlighted in purple) while promoting the growth of g1 (highlighted in grey). Simultaneously, the methylome shifted from m2 to m1, mirroring the genotype changes. (**H-J**) Additional methylation changes and ePV events in different bacterial species in Infant 1 (**H**), Infant 2 (**I**) and Infant 3 (**J**). The methylation profiles were generated based on the single-site methylation analysis (see **Methods**) and manually curated. (**A-C**, **G**) M: MTase represented by green arrows; S1, specificity unit 1 represented by red arrows; S2, specificity unit 2 represented by yellow arrows. The pink and black rectangles above the gene schematic diagram are inverted repeats.

Figure 3. Genome-wide ePV events are widespread across different cohorts and associated with clinical treatments

(**A**) An ePV observed in a *Bacteroides uniforms* strain from an adult gut microbiome. Due to read length limitation, short-read data can only detect simple ePV (or “inverton”, as defined in a previous study[21]), which are mediated by a pair of IR features and result in two distinct MTase genotypes (g1, g2). The proportion of each genotype is shown to the right of the gene diagram. The read alignment chart illustrates how paired-end (PE) reads were used to detect the inverton-like ePV. (**B**) Additional inverton-like ePVs detected across various bacterial species. Gene diagrams show the rearrangement of MTase gene fragments, while pie charts depict the proportion of each genotype variant. (**C**) Phylogenetic tree of inverton-like ePVs. Potential ePV (based solely on the reference genome) and confirmed ePV (supported by both the reference

genome and metagenomic read) across three types of Restriction-Modification Systems (I, II, III) (**Fig. S3**) and are projected onto the phylogenetic tree. ePV enriched orders are highlighted according to the colors scheme in the legend. (**D**) Response of *B. uniforms* ePV to AMX treatment. The *B. uniformis* ePV mentioned in **Fig. 3A** was detected in four infants from the DIABIMMUNE cohort, both pre- and post-AMX treatment. The MTase genotype shifted from g1 to g2 in response to treatment. (**E**) Prevalence analysis of the *B. uniformis* ePV. Among 153 complete *B. uniforms* genomes from NCBI, 23 were found to carry this ePV. The bar plot on the right shows the frequency of the dominant ePV genotype across the 23 genomes. (**A**, **B**, **E**) M: MTase represented by green arrows; S1, specificity unit 1 represented by red arrows; S2, specificity unit 2 represented by yellow arrows. The pink rectangles above the gene schematic are inverted repeats.

Figure 4. Site-specific ePV events are widespread across different cohorts and associated with bacterial adaptation to environmental stress

(**A**) For a specific bacterial strain, only a subset of cells exhibit methylation at a specific promoter, while the remainder remain unmethylated (left panel). The schematic in the middle panel depicts a transcription factor (TF) competing with an MTase at this specific site. The ePV site (highlighted within the dashed rectangle) was identified from multiple molecules using single-molecule methylation analysis. The right panel illustrates the deconvolution results obtained from this single-site methylation analysis. (**B**) Summary of all ePV sites detected in this study. Values were normalized based on the strain number within the order (diamond points highlighted in red). The methylation percentage within the order were also determined (box plot). Orders with fewer than five strains were merged into the 'Others' category. (**C**) Function categories of the genes affected by ePV sites. Heatmap showing the number of genes regulated by ePV, with host strain taxonomy (strain) as column label and gene function category (pathway) as row label. The color scale represents the number of genes, indicating the occurrence of each functional category per strain. (**D**) The top panel presents the time-line of AMX administration and sample collection for Infant 3. The bottom panel shows the involved methylation-related genes in a *B. ovatus* strain from Infant 3. (**E**) Distribution of partially methylated G6mATC sites across the *B. ovatus* genome. The genome consists of three contigs, randomly ordered along the x axis. The top three changed partially methylated sites were selected and labeled with numbers. (**F**) Changes in partially methylated G6mATC sites in response to AMX treatment. The motif sites were examined on both strands, and only sites that showed partial methylation on both strands were considered final ePV sites. The same criteria

were applied to all other partially methylated sites in Fig. **4I**,**L**. The gene diagram shows the location of each site within the affected gene. **(E,F)** Pre- (black) and post-AMX (red) methylations were highlighted. (**G**) Sample collection information for Donor 2 and Recipient 2 from FMT cohort. (**H**) Distribution of partially methylated G6mATC sites across the *Phocaeicola plebeius* genome. The top four partially methylated sites were selected and labeled with numbers. (**I**) Changes in partially methylated G6mATC sites in response to FMT. The gene diagram shows the location of each site within the affected gene. **(H**,**I)** Pre- (black) and post-FMT (red) methylations were highlighted. (**J**) Design and sample collection information for Akk-fed mouse assay. (**K**) Distribution of partial methylated RA6mATTY sites across the *A. muciniphila* genome. Three of the partially methylated sites, specifically lowly methylated in capsule, were identified and labeled with numbers. (**L**) Changes in partially methylated RA6mATTY sites from capsule to *in vivo* colonization (mouse stool). The gene diagram shows the location of each site within the affected gene. **(K**,**L)** Capsule (black) and *in vivo* (red) methylations were highlighted.

Figure 5. Characterization of a complex genome-wide ePV and its role in *Akkermansia* tolerance to AMX

(**A**) An ePV event observed in an *A. muciniphila* strain from Infant 2 in the ECAM cohort, mediated by three pairs of IR features, resulting in eight MTase genotypes (g1-g8) and six unique methylation sequence recognition motifs (epigenetic phases m1-m6). The genotype was profiled in pre- and post-AMX treatment samples, AMX treatment suppressed the population of g1 (highlighted in grey) while promoting the growth of g2 (highlighted in purple) or other non g1 populations. (**B**) Experimental design of the *in vitro* passage assay. *A. muciniphila* was isolated from the fecal sample of Infant 2 (hereafter referred to as Akk_ECAM). After recovery on the BHI agar plates, a culture derived from a single colony was subjected to 2 mg / L AMX treatment for 10 passages. Cultures without AMX served as negative controls. (**C**) Epigenetic phase profiling using targeted sequencing. In response to AMX treatment, the epigenetic phase changed from m1 dominance to m2 dominance, whereas m1 remained dominant through passage 10 in the absence of AMX. Only epigenetic phases with an abundance > 10% are shown in legend. (**D**) Methylation motif changed in the Akk_ECAM strain upon AMX treatment. Genomic DNA was extracted, and whole genome shotgun sequencing was performed using Pacbio Sequel II platform. Single-molecule methylation analysis was then conducted to investigate the methylation level of the two motifs of interest. The results align with the *in vivo* findings, where AMX treatment decreased in the m1 methylation subpopulation (highly methylated at GG6mAN_9TGAA) and increased the non-m1 methylation subpopulation (highly

methylated at G6mACN$_7$RTTC). (**E**) Experimental design of the 2nd batch of the *in vitro* passage assay. After recovery on BHI agar plates, 46 colonies were subjected to 2 mg / L AMX treatment for nine passages. Culture without AMX served as negative controls. (**F**) Epigenetic phase profiling with targeted sequencing. The epigenetic phase changed from m1 dominance to non-m1 (m2/m3) dominance in response to AMX treatment, while m1 remained dominant through passage 9 in the absence of AMX treatment. Only epigenetic phases with an abundance > 10% are shown in legend. (**G**) Summary of the 2nd batch *in vitro* passage assay. Epigenetic phases were determined by targeted sequencing and categorized into two groups, m1 and non-m1 (m2/m3). Fisher's exact test was used to assess the statistical significance of the difference in proportions between pre- and post-AMX treatment. The results indicate a significant association ($p = 0.0011$), suggesting that non-m1 phases were selected by AMX treatment. (**H**) Principal Component Analysis (PCA) analysis of RNA-seq-based expression data from in vitro passage samples. Under AMX treatment, m2/m3 samples were clearly separated from m1 samples. Overall sample variation was evaluated using PCA, with the explained variation labeled as percentages on the axes. (**I**) DEGs were identified by comparing m1 and m2/m3 samples under AMX treatment. Thresholds for DEGs were set at $\log_2$ fold changes > 1 or < -1, with an adjusted p value < 0.05. *mucC* was the only upregulated gene (highlighted in red). *tRNA-Phe* was the only downregulated gene (highlighted in green). (**J**) The promoter elements and methylation motif (GG6mAN$_9$TGAA) were represented in the diagram. The fraction of m1 methylation motif in the promoter of *mucC* decreased with AMX treatment. (**K**) AMX tolerance of *E. coli* and overexpression of the *mucC* gene. The *mucC* gene was codon-optimized, synthesized and fused with a 6 × HIS tag at the N-terminus. It was inserted into a medium-copy plasmid with a constitutive promoter, and AMX tolerance was assessed using the ETEST and CFU counting methods respectively. Each strain was tested in triplicate. (**L**) MTase specificity reversibility assay. m2/m3 dominant strains from the 2nd batch of the *in vitro* assay were treated under various conditions. The point plot shows the epigenetic phase profiling results obtained through targeted sequencing, while the zoomed-in bar plot indicates that ciprofloxacin treatment can reverse non-m1 (characterized by high *mucC* expression) back to the M1 genotype (with low *mucC* expression). . Each condition was tested in duplicate.

Materials and methods

Strains, media, and growth conditions

Escherichia coli str. K-12 substr. MG1655 was used for heterologous expression of the gene of interest. *E. coli* strains were cultured in LB broth (LB broth Miller, Becton) or LB agar (LB agar Miller, Becton) at 37°C. When grown in liquid medium, the cultures were aerated with shaking at 220 rpm, unless otherwise specified.

Akkermansia muciniphila strains employed in this work include Akk_ECAM and Akk_probiotics. The Akk_ECAM strain was isolated in this study (details provided below), while the Akk_probiotics strain was derived from a commercial probiotic product (Glucose Control, Pendulum). All *Akkermansia* strains were routinely grown statically under an anaerobic condition (5% hydrogen, 5% carbon dioxide, and 90% nitrogen) using a type C anaerobic chamber with Indicator (#7250000, Coy laboratory) in BHI media (Brain Heart Infusion Agar, Gibco) supplemented with 0.25% v/v porcine gastric mucin (M1778, Sigma-Aldrich). The minimum inhibitory concentration (MIC) for amoxicillin was determined using an ETEST assay. A gradient of diluted concentrations (0.5–16 mg/L) was tested. For the selection of amoxicillin (AMX)-tolerant Akkermansia, amoxicillin sodium salt was used at a concentration of 2 mg/L.

Stool specimen collection

ECAM samples were obtained from a previous study with an Institutional Review Board–approved study which was conducted on infants of healthy, pregnant mothers in New York City from 2011 to 2014.[1] FMT samples were obtained from a previous study with Institutional Review Board approval.[2]

Isolation of Akkermansia strains from infant stool

The *A. muciniphila* strain Akk_ECAM was isolated using a modified version of a previously described protocol[3] based on growth in mucin medium (3 mM KH_2PO_4, 3 mM Na_2HPO_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). To enrich for *A. muciniphila*, a loopful of stool sample (~20 mg) was added into 2 mL of mucin medium and incubated at 37°C for 48 h under anaerobic conditions. The culture was passaged in mucin media three times by

inoculating 100 µL into 2 mL of mucin medium and incubating until turbid. After three sequential passages, a suspension sample was streaked on BHI agar supplemented with 0.25% mucin and incubated for one week. Single colonies were re-streaked on BHI to purify, and the resulting purified colonies were then grown in 2 mL of mucin medium. To identify the isolated bacteria, DNA was isolated from 1 mL of culture (DNeasy Blood and Tissue Kit, Qiagen), and a portion of the 16S rRNA gene was amplified by PCR and sequenced (see primer sequences listed in **Table S6**).

DNA extraction

Extraction of DNA from fecal samples

Fecal DNA was extracted with a QIAamp PowerFecal Pro DNA kit (#51804, Qiagen) according to the manufacturer's instructions. Briefly, 50-100 mg of fecal sample was first homogenized by beads-beating using a Retsch mixer mill (MM400, Verder Scientific Inc., Newtown, PA) for 10 min at 25 Hz speed (5 min each side) and then subjected to column-based extraction steps as described in the kit protocol.

Extraction of DNA from bacterial cells

Genomic DNA was extracted from bacterial cells using either a Wizard Genomic DNA Purification Kit (A1120, Promega) or a Qiagen DNeasy Blood and Tissue kit (#69504, Qiagen) following the manufacturer’s instruction. The DNA was eluted in 100 µL elution buffer and stored at -20 °C.

RNA extraction

The bacterial cell pellets were collected from 0.5 mL of cell cultures grown in BHI media supplemented with 0.25% mucin after 48h of incubation and saved in -80 °C. Total RNA was extracted with an RNeasy Mini kit (#74104, Qiagen). Briefly, cells were first lysed in RLT buffer (from the kit) with the addition of 10 µL/mL β-mercaptoethanol. The lysate was further processed by adding the content of a PowerBead tube (glass, 0.1 mm; Qiagen) and homogenizing the mixture for 5 min at a speed of 30 Hz using a Retsch mixer mill. Subsequent steps followed the

manufacturer's instructions. Contaminating genomic DNA was depleted using on-column DNase treatments with DNase (#79254, Qiagen). RNA quality was assessed with a Qubit 4 Fluorometer (Invitrogen) using the Qubit 1× dsDNA HS Assay Kit to detect DNA contamination and the Qubit RNA HS Assay Kit to measure total RNA. Ribosomal RNA was depleted using NEBNext rRNA Depletion Kit (Bacteria) (#E7850, NEB). The quality of total RNA was checked using an Agilent 2100 Bioanalyzer (#5067-4626, Agilent Technologies USA).

Real-time quantitative RT-PCR

One microgram of total RNA was converted to cDNA using SuperScript™ VILO™ MasterMix kit (#11755050, Invitrogen). Gene-specific primer pairs are provided in **Table S6**. SYBR-based RT-qPCR was performed in 384-well plates (#4306737; Applied Biosystems) and analyzed by a qPCR system (QuantStudio 6 Flex, Thermo Fisher Scientific). The thermal cycling protocol was as follows: 10 min at 50°C, 1 min at 90°C, followed by 40 cycles of 30 s at 60°C and 1 min at 90°C, with a final melt curve stage. SYBR signal was normalized to the ROX reference dye. Cycle of quantification (Cq) values were automatically determined using QuantStudio Design & Analysis Software (v1.3). Based on the melting curves, the primers were specifically bound to the target genome. All the samples were run in replicates technical replicates.

Library preparation and sequencing

Oxford Nanopore whole genome sequencing

gDNA samples were sheared to an average length of 10 kb using g-Tubes (#520079, Covaris) and purified with 0.4 × AMPure XP beads (A63882, Beckman Coulter Inc.) to remove short fragments. Native libraries were prepared using Oxford Nanopore Technologies (ONT) ligation sequencing kit and native barcoding kit (SQK-LSK109, and SQK-NBD112-24, respectively). PCR barcoding was performed using the Native Barcoding Expansion 1-12 kit (EXP-NBD104) and/or 13-24 kit (EXP-NBD114) for sample multiplexing. Whole genome amplification (WGA) were conducted with the REPLI-g Mini kit (#150023, Qiagen) using 10 ng of input DNA and 16 h of incubation. WGA samples were treated with T7 endonuclease I (M0302S, NEB) to maximize sequencing yield, as per ONT documentation. Native or amplified gDNA samples were sequenced on R9.4.1 and R10.4.1 MinION flow cells using MinION MK1B or MK1C instruments.

Basecalling was performed with Guppy v.5.0.7 (ONT) in GPU mode using the dna_r9.4.1_450bps_sup.cfg model for R9.4 data or Dorado v0.4.1 (ONT) with dna_r10.4.1_e8.2_400bps_sup.cfg model for R10.4 data.

ONT targeted sequencing

The RMS region was amplified using specific forward and reverse primers (**Table S6**) to generate a 6.6-kb amplicon, which included sequences enable sample barcoding. PCR was performed in a 50 μL reactions containing 250 ng of gDNA and 0.7 μM of primers. Specific barcodes were assigned to the PCR products using the PCR Barcoding Expansion 1–96 kit (EXP-PBC096, ONT). A second-round PCR was performed with 0.5 nM purified PCR product, 1 × LongAmp PCR buffer, 2 mM of $MgCl_2$, 0.3 mM of dNTPs and 5 units of LongAmp Taq DNA polymerase (NEB) in a 50 μL reaction. The thermal cycling protocol was 95°C for 3 min, followed by 15 cycles of 95°C for 15 s, 62°C for 15 s, 65°C for 7 min, with a final extension at 65°C for 7 min. Amplicons were sequenced on R9.4.1 / R10.4.1 Flongle flow cells.

PacBio WGS sequencing

The SMRTbell library was prepared according to the manufacturer's instructions. DNA was sheared to ~5,000 bp using a Covaris LE220-plus Focused-ultrasonicator (#500569, Covaris). The samples underwent damage and end repair, followed by ligation of single-stranded barcoded adaptors. Sequencing primer was annealed to the SMRTbells and incubated with templates for 1 hour at 20 ºC using a primer-to-template ratio of 20:1. The Sequel 2.1 polymerase was bound to the annealed complex at a polymerase to SMRTbell ratio of 30:1 and incubated for 4 h at 30 °C. The binding complex was cleaned, diluted to the final loading concentration (125-175 pM), and held at 4 °C until diffusion-based loading on the Sequel II instrument (PacBio). The sequencing runs used 8M sequencing chips, with 30 h movies recorded following the manufacturer's guidelines. Most metagenome samples except for some FMT samples, were sequenced on the latest Revio instrument (PacBio). Circular consensus sequencing (CCS) reads were generated using CCS v.3.4.1 (https://github.com/PacificBiosciences/ccs) with default settings.

Illumina RNA sequencing

RNA-seq library construction for Illumina sequencing was performed using the NEBNext Ultra II RNA Library Prep Kit for Illumina (#E7770, NEB). Library quality and size were assessed using

an Agilent 4200 TapeStation system (Agilent Technologies, Inc.) and quantified with a Qubit 4 Fluorimeter using Qubit DNA HS Assay Kit (Invitrogen). All samples were equitably pooled, size-selected, and sequenced on an Illumina HiSeq 2500 platform, generating paired-end reads of 150 base pairs (bp) each.

Phylogenetic analysis of HMP reference genomes

The bacterial genomics sequences and proteome files were downloaded from the NIH Human Microbiome Project (HMP) legacy data browser website (https://www.hmpdacc.org/hmp/HMRGD/). A phylogenetic tree was constructed using PhyloPhlAn v3.0.67[4] with proteome files as input and high--diversity configuration. The tree was visualized and edited using the interactive tree of life (iTOL) software (https://itol.embl.de).

***De novo* assembly of metagenomes**

ONT read metagenomic assembly and taxonomic annotation

Basecalled reads were processed using metaFlye tool v2.9[5] with the parameters ---nano-raw --meta --keep-haplotypes --iterations 2. Meta-assemblies were divided into species-level bins using the MaxBin 2.0 software.[6] Strainberry v1.1[7] was used to phase haplotypes and separate strain level contigs. Contigs were subsequently binned using the "binning" module of metaWRAP v1.3.2.[8] The quality of the reconstructed MAGs was evaluated using the 'bin_refinement' module of metaWRAP v1.3.2, and taxonomic assignment was performed using GTDB-tk v3.8.[9]

SMRT read metagenomic assembly and taxonomic annotation

HiFi sequencing reads were assembled using hifiasm_meta 0.13-r308 (https://github.com/xfengnefx/hifiasm-meta) with default parameters, focusing on primary contigs. To maximize MAGs recovery, the HiFi-MAG-Pipeline (https://github.com/PacificBiosciences/pb-metagenomics-tools/tree/master/HiFi-MAG-Pipeline) was deployed, increasing both the total number of MAGs and single-contig circular MAGs.

Hybrid assembly

Nanopore-based long reads and Illumina paired-end short reads were used for a hybrid assembly with MUFFIN,[10] producing high-quality and complete bins.

Gene annotation and RMS prediction

Gene annotation for bacterial genomes and MAGs was performed using Prokka v1.14.6.[11] RMS identification followed previously established methods.[12] Curated reference protein sequences for types I, II, IIC, and III RMS, and type IV REases were downloaded from the REBASE[13] “gold standards” dataset (accessed November 2016). A dedicated BLAST database was created for the specific gene (HsdS) of Type I RMS using 39,729 HsdS proteins from REBASE. The TRD domains in hsdS genes were identified using HMMER v3.3 (http://hmmer.org) with a profile built from 31 TRD domains.

Detection of SNPs and small Indels in antibiotic-treated *A. muciniphila* strains

Genome variations were identified from the aligned Nanopore reads to the reference genome using Clair3[14] with default ONT settings, except for the options --no_phasing_for_fa, --include_all_ctgs, and --haploid_precise. Variants called by Clair3 were filtered with bcftools[15] based on the criteria "QUAL > 15 and DP > 10". Subsequently, SNPs located within structural variation regions were excluded. The potential effects of the candidate SNPs were then annotated using snpEFF.[16]

Differential-expression analysis of RNA-seq data

Read quality was assessed using FastQC v.0.11.9 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Adapter sequences and low-quality reads were removed using Trimmomatic v.0.36[17] with the following parameters: PE, -phred33, ILLUMINACLIP:<adapters.fa>:2:30:10:8: keepBothReads, SLIDINGWINDOW:4:15, LEADING:20, TRAILING:20, and MINLEN:50. rRNA sequences were then filtered out using SortMeRNA v.2.1b [18] with the SILVA 16S/23S[19] and the Rfam 5S[20] rRNA database. Non-rRNA reads were aligned to the reference genome using BWA-MEM v.0.7.17-r1198.[21] Read assignment was performed with FeatureCounts v.2.0.1,[22] excluding multi-mapping and multi-overlapping reads. Normalization and differential expression testing were conducted using

DESeq2 v.1.18.1,[23] with differentially expressed genes defined as those with an FDR-corrected $p < 0.05$ and $|\log2[FC]| > 1$.

Gene set enrichment analysis

Gene function was annotated by eggmapper v2.1.6 against the defaults eggNOG database.[24] GO and KEGG enrichment analyses were performed using the R package clusterProfiler v.3.18.1. [25] KEGG pathways and modules were filtered for UniRef90 v.2019_06 entries, downloaded from UniProt website.[26] Enrichments at different taxonomic levels were calculated and visualized as heatmap using ComplexHeatmap package v.2.6.2.[27]

Genome-wide epigenetic phase variation (ePV) detection in long-read data

Structural variation (SV) detection

Basecalled ONT reads were aligned to the de novo assembled MAGs with NGMLR v 0.2.6.[28] HiFi PacBio reads were aligned to the de novo assemblies with pbalign (SMRTlink v10, https://downloads.pacbcloud.com/public/software/installers/smrtlink_10.1.0.119588.zip). The resulting alignment files were processed with sniffles v2.0.7[29] for structural variant calling. Detected SVs were mapped to their corresponding genome locations, and SVs overlapping predicted RMS regions / orphan MTases (within a 2,000 bp window) were defined as potential ePVs. The ePV list was refined through manually curation.

Quantification of ePV with raw reads

To overcome the limitations of SV detection tools in quantifying read counts assigned to genomic variants in complex rearrangement regions, we developed an in-house pipeline: (1) predicted all possible genomic variants base on the reference genome and the read alignment; (2) created an *in silico* reference for each variants and merged them; (3) realigned reads to the new reference genomes; (4) applied strict filtered thresholds, including “MAPQ 60 & mg > 99” and “-max softsclip 300”, (5) counted the reads uniquely mapped to each genotype.

Site-specific ePV detection for PacBio data

Preprocess of PacBio raw data

PacBio sequencing data contained both DNA sequence and kinetic data, which were used to detect single-molecule methylation signal. Raw SMRT Sequel II data was processed into CCS reads and demultiplexed using SMRTlink v10.

Genome-wide aggregated methylation analysis

Subreads and CCS reads were aligned to MAGs using pbalign (SMRTlink v10) with default parameters, and alignments were separated by MAG. The IPD ratio was calculated with ipdSummary (SMRTlink v10) for each MAGs. De novo methylation motifs were identified using motifMaker (SMRTlink v10) with default parameters. All of the above analyses aggregate the CCS data based on the MAG genomes, as opposed to single-molecule analysis, which is based on each individual CCS reads.

Single molecule level methylation analysis

Subreads were split based on CCS read ID and mapped to their corresponding CCS reads using pbalign with default parameters. Alignments were filtered with strict thresholds, including "MAPQ 60 & mg > 99" and "-max softsclip 300". Single-molecule IPD ratio was calculated with ipdSummary. For Revio data, raw data lacked subread information and require an in-house scripts for single-molecule analysis (released the detailed scripts on Zenodo).

Identification of single site ePV

IPD ratio files from single-molecule methylation signals were processed in R v4.0.3. Low methylation sites (< 50%) were identified on both strands, and only sites with low methylation on both strands were considered partially methylated sites. The methylation variation sites were visualized across MAGs, with loci and potentially affected genes highlighted in a gene diagram chart.

Genome-wide ePV detection in Illumina short-read data

For short-read metagenomic data, ePVs exclusively referred to invertons, as defined in a previous study.[30] Invertons across the metagenomic samples were identified using the PhaseFinder[30] pipeline, with HMP reference genomes (https://hmpdacc.org/hmp/HMRGD/) as reference. Invertons overlapping predicted RMS regions / orphan MTase (within a 2,000 bp window) were classified as potential ePVs.

Protein extraction, purification and identification

Protein preparation

E. coli MG1655 cells were harvested by centrifugation at 5,000 × g for 10 minutes at 4°C. The supernatant was collected in 50 mL aliquots. The cell pellet was resuspended in lysis buffer containing 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 2 mM tris (2-carboxyethyl)phosphine (TCEP, #68957, Sigma-Aldrich) and a protein inhibitor cocktail (P8340, Sigma-Aldrich). The suspension was lysed by sonication on ice (15-second pulses, 10 times, with 2-min intervals). The lysate was clarified by centrifugation at 12,000 × g for 30 min at 4°C to remove cell debris. The supernatant from cell culture was concentrated to 1 mL using an Amicon Ultra Centrifugal Filter unit (3 kDa, Millipore) and subjected to sonication as described above. The resulting crude protein solution was purified using a Capturem His-Tagged Purification kit (#635719, TaKaRa Bio Inc.) following the manufacturer's instructions.

SDS-PAGE

Protein samples were mixed with 4× Sodium dodecyl sulfate (SDS) loading buffer (Bio-Rad Laboratories, Inc) and heated at 98°C for 5 min. SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was performed using Mini-PROTEIN TGX Gels (4–20%, Bio-Rad Laboratories, Inc). Gels were stained with Coomassie Brilliant Blue R-250, and protein bands were visualized using a gel documentation system (Bio-Rad Laboratories, Inc).

Intact Protein Mass Spectrometry

For intact protein mass analysis, purified protein samples were buffer exchanged into 1 M ammonium acetate solution using a 3 kDa Amicon Ultra Centrifugal Filter unit (Millipore). Qualitative analysis of recombinant protein was performed with a 1260 HPLC system coupled to a 6545 XT Q-TOF MS instrument operated in positive ESI mode (Agilent Technologies, Inc.).

The acquired spectra were extracted and deconvoluted with Bioconfirm A 2.0 software (Agilent Technologies) with the default parameters, except that the mass range was set to be 10 kDa to 100 kDa.

In vitro passage assay of *A. muciniphila*

AMX selection assay

An *A. muciniphila* isolate with the g1 RMS genotype was selected as the starting material and cultured on BHI agar plates supplemented with 0.25% mucin for three days. 46 single colonies were picked and subcultured in 3 mL BHI broth supplemented with 0.25% mucin for two additional passages. The above subcultures were diluted to an initial OD_{600} of 0.01 and a 50 μL aliquot was inoculated in 150 μL BHI liquid medium supplemented with 0.25% mucin, and with or without 2 mg / L AMX. The subculture process was repeated for nine passages. Cells from each passage were collected and stored. gDNA extraction and targeted sequencing were performed as described above.

RMS Reversible assay

To assess the conditions under which the RMS genotype could revert from non-g1 to g1, *A. muciniphila* colonies with the non-g1 RMS genotype from the AMX selection assay were stressed with various microbiological mediums, antibiotics, and atmospheric conditions (see detailed design and conditions in **Table S7**). RMS genotypes were determined across the assay using targeted sequencing. Additional replicates for ciprofloxacin treatment were performed following the same procedure but with an increased number of replicates. The minimum inhibitory concentration (MIC) for each antibiotic was determined using a preliminary CFU count assay. Antibiotic concentrations were selected based on their respective MIC values.

Expression of the *mucC* gene in *E. coli*

To overexpress the *mucC* gene, we synthesized the gene with codons optimized for expression in *E. coli*. The gene was inserted into a medium-copy-number plasmid containing a constitutive expression promoter. The recombination plasmid was then transformed into *E. coli*. Overnight cell cultures were diluted 1:100 into fresh LB medium containing different concentrations of AMX. The growth curve of each group of cells (OD_{600}) was monitored using a microplate spectrophotometer (SpectraMax M5, USA).

AMX sensitivity test

AMX test strips (AES, Liofilchem) were placed on the surface of a BHI agar plate supplemented with 0.25% mucin and inoculated with a 500 µL bacterial suspension. The plates were incubated at 37°C for 48 h under anaerobic conditions. The MIC was determined by the intersection of the lower part of the ellipse-shaped growth inhibition area with the test strip.

Akkermansia-fed Mouse model

A probiotic supplement including *Clostridium beijerinckii* WB-STR-0005, *Clostribium butyricum* WB-STR-0006, *A. muciniphila* WB-STR-0001 (introduced earlier as Akk_probiotics), *Anaerobutyricum hallii* WB-STR-0008, and *Bifidobacterium infantis* 100 (Pendulum, USA) was administered to healthy male germ-free (GF) C57BL/6J mice (JAX stock #000664) (n=4). 12-week-old GF mice were fed 10 mg probiotic supplement via Mini Capsule Kit (Braintree Scientific XKITM) for the first two days. Fecal samples were collected at day 7.

Quantification and statistical analysis

Statistical analyses were performed using R (v.4.0.3). Statistical significance was determined using unpaired two-tailed Mann–Whitney–Wilcoxon test, unpaired two-tailed Welch's t-test, and unpaired Student's t-test, as indicated in figure legends. Data are shown as mean ± s.d. $P < 0.05$ was considered statistically significant. Information regarding the number of samples and statistical tests are provided in the figure legends.

References for Methods

1. Bokulich, N.A., Chung, J., Battaglia, T., Henderson, N., Jay, M., Li, H., D. Lieber, A., Wu, F., Perez-Perez, G.I., Chen, Y., et al. (2016). Antibiotics, birth mode, and diet shape microbiome maturation during early life. Sci. Transl. Med. *8*, 343ra382-343ra382. 10.1126/scitranslmed.aad7121.
2. Aggarwala, V., Mogno, I., Li, Z., Yang, C., Britton, G.J., Chen-Liaw, A., Mitcham, J., Bongers, G., Gevers, D., Clemente, J.C., et al. (2021). Precise quantification of bacterial strains after fecal microbiota transplantation delineates long-term engraftment and explains outcomes. Nat. Microbiol. *6*, 1309-1318. 10.1038/s41564-021-00966-0.

3. Derrien, M., Vaughan, E.E., Plugge, C.M., and de Vos, W.M. (2004). *Akkermansia muciniphila* gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. Int. J. Syst. Evol. Microbiol. *54*, 1469-1476. 10.1099/ijs.0.02873-0.
4. Asnicar, F., Thomas, A.M., Beghini, F., Mengoni, C., Manara, S., Manghi, P., Zhu, Q., Bolzan, M., Cumbo, F., May, U., et al. (2020). Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. Nat. Commun. *11*, 2500. 10.1038/s41467-020-16366-7.
5. Kolmogorov, M., Bickhart, D.M., Behsaz, B., Gurevich, A., Rayko, M., Shin, S.B., Kuhn, K., Yuan, J., Polevikov, E., Smith, T.P.L., and Pevzner, P.A. (2020). metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat. Methods *17*, 1103-1110. 10.1038/s41592-020-00971-x.
6. Wu, Y.-W., Simmons, B.A., and Singer, S.W. (2015). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics *32*, 605-607. 10.1093/bioinformatics/btv638.
7. Vicedomini, R., Quince, C., Darling, A.E., and Chikhi, R. (2021). Strainberry: automated strain separation in low-complexity metagenomes using long reads. Nat. Commun. *12*, 4485. 10.1038/s41467-021-24515-9.
8. Uritskiy, G.V., DiRuggiero, J., and Taylor, J. (2018). MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome *6*, 158. 10.1186/s40168-018-0541-1.
9. Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2022). GTDB-Tk v2: memory friendly classification with the genome taxonomy database. Bioinformatics *38*, 5315-5316. 10.1093/bioinformatics/btac672.
10. Van Damme, R., Hölzer, M., Viehweger, A., Müller, B., Bongcam-Rudloff, E., and Brandt, C. (2021). Metagenomics workflow for hybrid assembly, differential coverage binning, metatranscriptomics and pathway analysis (MUFFIN). PLoS Comp. Biol. *17*, e1008716. 10.1371/journal.pcbi.1008716.
11. Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics *30*, 2068-2069. 10.1093/bioinformatics/btu153.
12. Oliveira, P.H., Ribis, J.W., Garrett, E.M., Trzilova, D., Kim, A., Sekulovic, O., Mead, E.A., Pak, T., Zhu, S., Deikus, G., et al. (2020). Epigenomic characterization of *Clostridioides difficile* finds a conserved DNA methyltransferase that mediates sporulation and pathogenesis. Nat. Microbiol. *5*, 166-180. 10.1038/s41564-019-0613-4.
13. Roberts, R.J., Vincze, T., Posfai, J., and Macelis, D. (2014). REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res. *43*, D298-D299. 10.1093/nar/gku1046.
14. Zheng, Z., Li, S., Su, J., Leung, A.W.-S., Lam, T.-W., and Luo, R. (2022). Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. Nature Computational Science *2*, 797-803. 10.1038/s43588-022-00387-x.
15. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. 10.1093/gigascience/giab008.
16. Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly *6*, 80-92. 10.4161/fly.19695.
17. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114-2120.
18. Kopylova, E., Noé, L., and Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics *28*, 3211-3217.

19. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. *41*, D590-D596.
20. Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S.R. (2003). Rfam: an RNA family database. Nucleic Acids Res. *31*, 439-441.
21. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754-1760.
22. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923-930.
23. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 1-21.
24. Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). Eggnog-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol. Biol. Evol. *38*, 5825-5829. 10.1093/molbev/msab293.
25. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284-287. 10.1089/omi.2011.0118.
26. Consortium, T.U. (2022). UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. *51*, D523-D531. 10.1093/nar/gkac1052.
27. Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32*, 2847-2849. 10.1093/bioinformatics/btw313.
28. Sedlazeck, F.J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A., and Schatz, M.C. (2018). Accurate detection of complex structural variations using single-molecule sequencing. Nat. Methods *15*, 461-468. 10.1038/s41592-018-0001-7.
29. Smolka, M., Paulin, L.F., Grochowski, C.M., Horner, D.W., Mahmoud, M., Behera, S., Kalef-Ezra, E., Gandhi, M., Hong, K., Pehlivan, D., et al. (2024). Detection of mosaic and population-level structural variants with Sniffles2. Nat. Biotechnol. 10.1038/s41587-023-02024-y.
30. Jiang, X., Hall, A.B., Arthur, T.D., Plichta, D.R., Covington, C.T., Poyet, M., Crothers, J., Moses, P.L., Tolonen, A.C., Vlamakis, H., et al. (2019). Invertible promoters mediate bacterial phase variation, antibiotic resistance, and host adaptation in the gut. Science *363*, 181-187. 10.1126/science.aau5238.

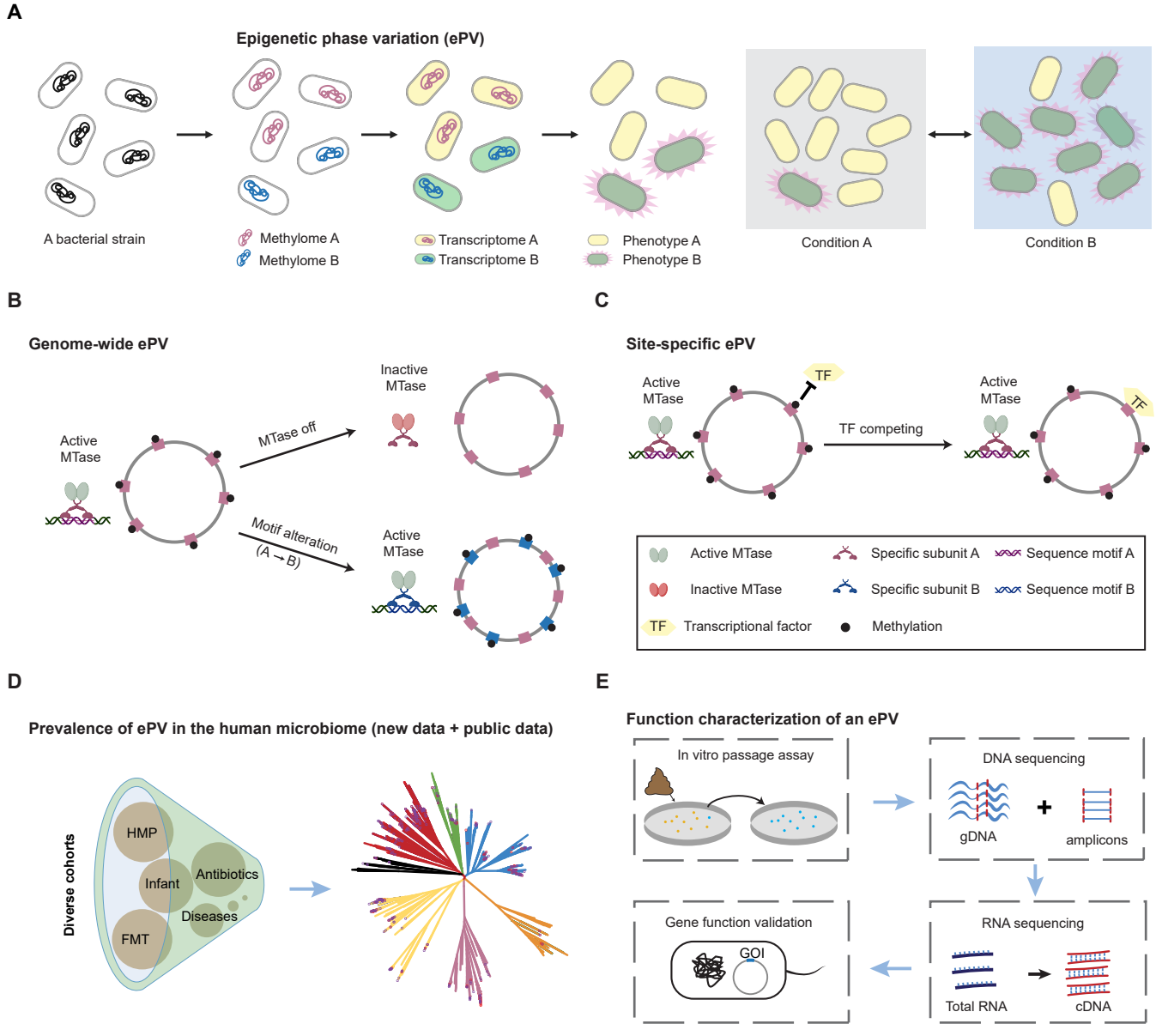
A
Epigenetic phase variation (ePV)
A bacterial strain
Methylome A
Methylome B
Transcriptome A
Transcriptome B
Phenotype A
Phenotype B
Condition A
Condition B
B
Genome-wide ePV
Active MTase
MTase off
Inactive MTase
Motif alteration (A → B)
Active MTase
C
Site-specific ePV
Active MTase
TF
TF competing
Active MTase
TF
Active MTase
Inactive MTase
TF Transcriptional factor
Specific subunit A
Specific subunit B
Methylation
Sequence motif A
Sequence motif B
D
Prevalence of ePV in the human microbiome (new data + public data)
Diverse cohorts
HMP
Infant
Antibiotics
Diseases
FMT
E
Function characterization of an ePV
In vitro passage assay
DNA sequencing
gDNA
amplicons
RNA sequencing
Total RNA
cDNA
Gene function validation
GOI

Figure 1. An integrated framework for epigenetic phase variation (ePV) detection and function validation from the human gut microbiome

(**A**) A specific bacterial strain may have sub-populations that exhibit heterogeneous methylation patterns, leading to distinct gene expression and phenotypes. This may enhance the ability of a single strain to adapt to different stress conditions. (**B**) Schematic diagram of genome-wide ePV events, mediated by the structural variation located within MTase genes (ON/OFF states) or their specificity units (alters recognition motifs). (**C**) Schematic diagram of site-specific ePV events, resulting from the competition between MTase and transcription factors (TFs). (**D**) Prevalence and phylogenetic distribution of ePV were explored across various healthy and clinical-related cohorts. Phylogenetic tree is color-coded at phylum level. (**E**) Workflow for functional analysis of genes involved in ePV events. Bacterial strains carrying the ePVs were isolated from fecal samples. An *in vitro* passage assay was performed under conditions mimicking the *in vivo* environment, followed by whole-genome shotgun and targeted DNA sequencing. MTase genotypes were quantified and associated with treatment conditions. RNA sequencing was conducted to profile transcriptomic differences between genotypes. Differentially expressed genes (DEGs) were overexpressed further in native strains or model species, and phenotypic tests were carried out to confirm whether the altered phenotype was due to the expression of the gene of interest (GOI).

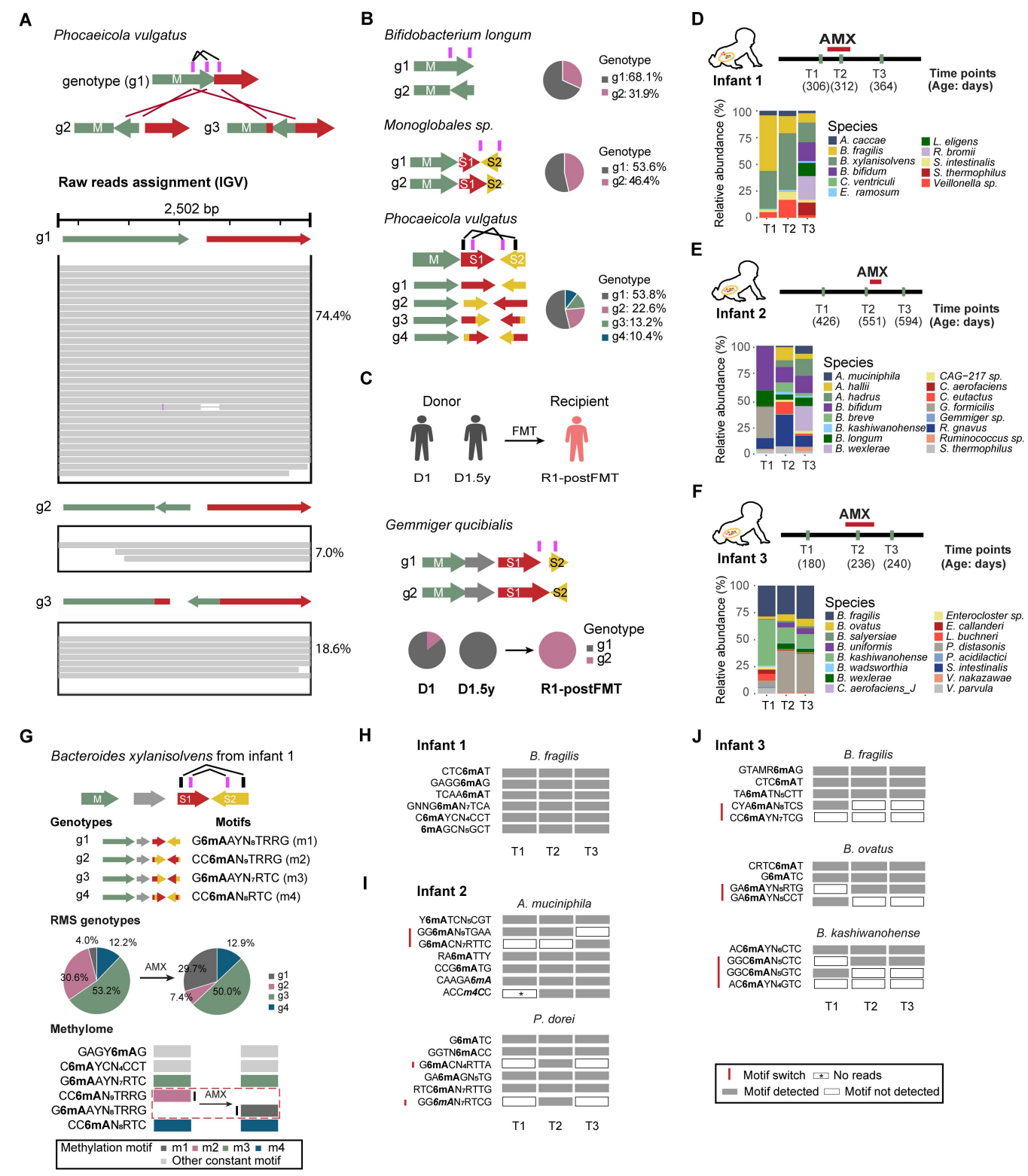
A
Phocaeicola vulgatus
genotype (g1)
g2
g3
Raw reads assignment (IGV)
2,502 bp
g1
74.4%
g2
7.0%
g3
18.6%
B
Bifidobacterium longum
Genotype
g1:68.1%
g2: 31.9%
Monoglobales sp.
Genotype
g1: 53.6%
g2: 46.4%
Phocaeicola vulgatus
Genotype
g1: 53.8%
g2: 22.6%
g3:13.2%
g4:10.4%
C
Donor
Recipient
FMT
D1
D1.5y
R1-postFMT
Gemmiger qucibialis
Genotype
g1
g2
D1
D1.5y
R1-postFMT
D
Infant 1
AMX
T1 (306)
T2 (312)
T3 (364)
Time points (Age: days)
Relative abundance (%)
Species
A. caccae
B. fragilis
B. xylanisolvens
B. bifidum
C. ventriculi
E. ramosum
L. eligens
R. bromii
S. intestinalis
S. thermophilus
Veillonella sp.
E
Infant 2
AMX
T1 (426)
T2 (551)
T3 (594)
Time points (Age: days)
Relative abundance (%)
Species
A. muciniphila
A. hallii
A. hadrus
B. bifidum
B. breve
B. kashiwanohense
B. longum
B. wexlerae
CAG−217 sp.
C. aerofaciens
C. eutactus
G. formicilis
Gemmiger sp.
R. gnavus
Ruminococcus sp.
S. thermophilus
F
Infant 3
AMX
T1 (180)
T2 (236)
T3 (240)
Time points (Age: days)
Relative abundance (%)
Species
B. fragilis
B. ovatus
B. salyersiae
B. uniformis
B. kashiwanohense
B. wadsworthia
B. wexlerae
C. aerofaciens_J
Enterocloster sp.
E. callanderi
L. buchneri
P. distasonis
P. acidilactici
S. intestinalis
V. nakazawae
V. parvula
G
Bacteroides xylanisolvens from infant 1
Genotypes
Motifs
g1
G6mAAYN8TRRG (m1)
g2
CC6mAN9TRRG (m2)
g3
G6mAAYN7RTC (m3)
g4
CC6mAN8RTC (m4)
RMS genotypes
4.0%
12.2%
30.6%
53.2%
AMX
12.9%
29.7%
7.4%
50.0%
g1
g2
g3
g4
Methylome
GAGY6mAG
C6mAYCN4CCT
G6mAAYN7RTC
CC6mAN9TRRG
AMX
G6mAAYN8TRRG
CC6mAN8RTC
Methylation motif
m1
m2
m3
m4
Other constant motif
H
Infant 1
B. fragilis
CTC6mAT
GAGG6mAG
TCAA6mAT
GNNG6mAN7TCA
C6mAYCN4CCT
6mAGCN5GCT
T1
T2
T3
I
Infant 2
A. muciniphila
Y6mATCN5CGT
GG6mAN9TGAA
G6mACN7RTTC
RA6mATTY
CCG6mATG
CAAGA6mA
ACCm4CC
P. dorei
G6mATC
GGTN6mACC
G6mACN4RTTA
GA6mAGN5TG
RTC6mAN7RTTG
GG6mAN7RTCG
T1
T2
T3
J
Infant 3
B. fragilis
GTAMR6mAG
CTC6mAT
TA6mATN5CTT
CYA6mAN8TCS
CC6mAYN7TCG
B. ovatus
CRTC6mAT
G6mATC
GA6mAYN8RTG
GA6mAYN5CCT
B. kashiwanohense
AC6mAYN6CTC
GGC6mAN5CTC
GGC6mAN5GTC
AC6mAYN4GTC
T1
T2
T3
Motif switch
No reads
Motif detected
Motif not detected

Figure 2. Long-read metagenomic sequencing deciphers complex ePV events

(**A**) An ePV observed in a *Phocaeicola vulgatus* strain from a healthy adult gut microbiome, mediated by three paired/matching inverted repeats (inverted repeats), resulting in three distinct MTase genotypes (g1-g3). The alignment plots show raw reads aligned with their corresponding reference genomes. The proportion of each genotype is displayed on the right side of the read alignment chart. (**B**) Additional ePV events were identified in various bacterial species. Gene diagrams illustrate the rearrangement of MTase gene fragments, while pie charts show the proportion of each genotype variant. (**C**) ePV events identified in an FMT cohort. The top panel presents metadata explaining sample and subject information (see **Table S3**). The bottom panel shows an ePV in a *Gemmiger qucibialis* strain, detected in both donor and post-FMT samples. (**D-F**) Metagenomic analysis of fecal samples from three infants in the ECAM cohort. The top panel presents the timeline of sample collection and amoxicillin (AMX) exposure. The bottom panel displays species-level relative abundance profiles, including only species with a sequencing depth > 10X. (**G**) An ePV event was observed in a *Bacteroides xylanisolvens* strain, mediated by two pairs of IR features, resulting in four specificity unit genotypes (g1-g4). The genotype was profiled in pre- and post-AMX treatment samples, alongside their corresponding methylome (m1-m4). AMX treatment suppressed the population of g2 (highlighted in purple) while promoting the growth of g1 (highlighted in grey). Simultaneously, the methylome shifted from m2 to m1, mirroring the genotype changes. (**H-J**) Additional methylation changes and ePV events in different bacterial species in Infant 1 (**H**), Infant 2 (**I**) and Infant 3 (**J**). The methylation profiles were generated based on the single-site methylation analysis (see **Methods**) and manually curated. (**A-C**, **G**) M: MTase represented by green arrows; S1, specificity unit 1 represented by red arrows; S2, specificity unit 2 represented by yellow arrows. The pink and black rectangles above the gene schematic diagram are inverted repeats.

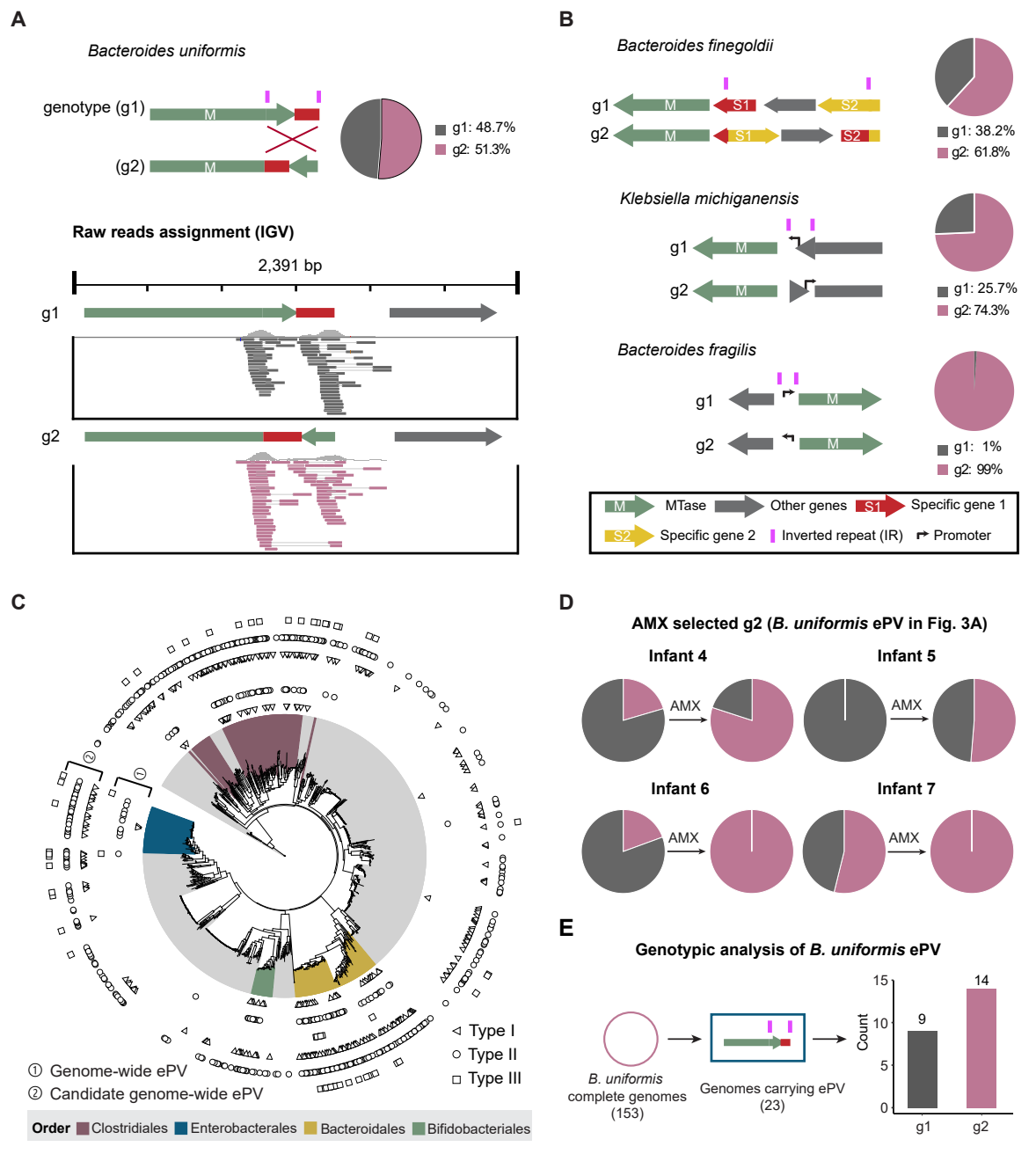
A
Bacteroides uniformis
genotype (g1)
(g2)
M
g1: 48.7%
g2: 51.3%
Raw reads assignment (IGV)
2,391 bp
g1
g2
B
Bacteroides finegoldii
g1
g2
M
S1
S2
g1: 38.2%
g2: 61.8%
Klebsiella michiganensis
g1: 25.7%
g2: 74.3%
Bacteroides fragilis
g1: 1%
g2: 99%
MTase
Other genes
Specific gene 1
Specific gene 2
Inverted repeat (IR)
Promoter
C
Type I
Type II
Type III
① Genome-wide ePV
② Candidate genome-wide ePV
Order
Clostridiales
Enterobacterales
Bacteroidales
Bifidobacteriales
D
AMX selected g2 (B. uniformis ePV in Fig. 3A)
Infant 4
Infant 5
Infant 6
Infant 7
AMX
E
Genotypic analysis of B. uniformis ePV
B. uniformis complete genomes (153)
Genomes carrying ePV (23)
Count
15
10
5
0
9
14
g1
g2

Figure 3. Genome-wide ePV events are widespread across different cohorts and associated with clinical treatments

(**A**) An ePV observed in a *Bacteroides uniforms* strain from an adult gut microbiome. Due to read length limitation, short-read data can only detect simple ePV (or "inverton", as defined in a previous study[21]), which are mediated by a pair of IR features and result in two distinct MTase genotypes (g1, g2). The proportion of each genotype is shown to the right of the gene diagram. The read alignment chart illustrates how paired-end (PE) reads were used to detect the inverton-like ePV. (**B**) Additional inverton-like ePVs detected across various bacterial species. Gene diagrams show the rearrangement of MTase gene fragments, while pie charts depict the proportion of each genotype variant. (**C**) Phylogenetic tree of inverton-like ePVs. Potential ePV (based solely on the reference genome) and confirmed ePV (supported by both the reference genome and metagenomic read) across three types of Restriction-Modification Systems (I, II, III) (**Fig. S3**) and are projected onto the phylogenetic tree. ePV enriched orders are highlighted according to the colors scheme in the legend. (**D**) Response of *B. uniforms* ePV to AMX treatment. The *B. uniformis* ePV mentioned in **Fig. 3A** was detected in four infants from the DIABIMMUNE cohort, both pre- and post-AMX treatment. The MTase genotype shifted from g1 to g2 in response to treatment. (**E**) Prevalence analysis of the *B. uniformis* ePV. Among 153 complete *B. uniforms* genomes from NCBI, 23 were found to carry this ePV. The bar plot on the right shows the frequency of the dominant ePV genotype across the 23 genomes. (**A**, **B**, **E**) M: MTase represented by green arrows; S1, specificity unit 1 represented by red arrows; S2, specificity unit 2 represented by yellow arrows. The pink rectangles above the gene schematic are inverted repeats.

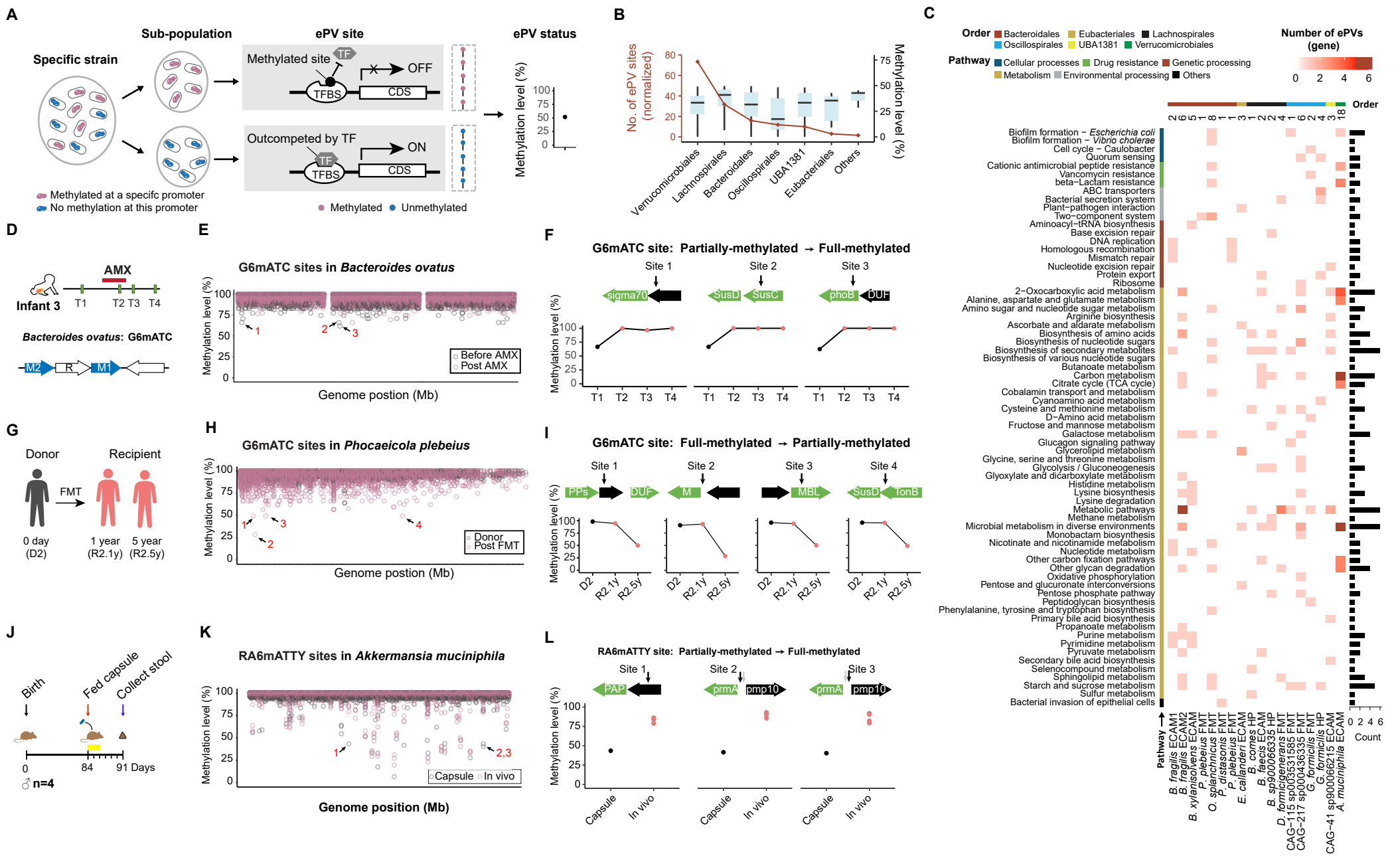
A
Sub-population
ePV site
ePV status
Specific strain
Methylated site
TF
TFBS
CDS
OFF
Outcompeted by TF
ON
Methylation level (%)
Methylated at a specifc promoter
No methylation at this promoter
Methylated
Unmethylated
B
No. of ePV sites (normalized)
Methylation level (%)
Verrucomicrobiales
Lachnospirales
Bacteroidales
Oscillospirales
UBA1381
Eubacteriales
Others
C
Order
Bacteroidales
Eubacteriales
Lachnospirales
Oscillospirales
UBA1381
Verrucomicrobiales
Pathway
Cellular processes
Drug resistance
Genetic processing
Metabolism
Environmental processing
Others
Number of ePVs (gene)
Biofilm formation – Escherichia coli
Biofilm formation – Vibrio cholerae
Cell cycle – Caulobacter
Quorum sensing
Cationic antimicrobial peptide resistance
Vancomycin resistance
beta–Lactam resistance
ABC transporters
Bacterial secretion system
Plant–pathogen interaction
Two–component system
Aminoacyl–tRNA biosynthesis
Base excision repair
DNA replication
Homologous recombination
Mismatch repair
Nucleotide excision repair
Protein export
Ribosome
2–Oxocarboxylic acid metabolism
Alanine, aspartate and glutamate metabolism
Amino sugar and nucleotide sugar metabolism
Arginine biosynthesis
Ascorbate and aldarate metabolism
Biosynthesis of amino acids
Biosynthesis of nucleotide sugars
Biosynthesis of secondary metabolites
Biosynthesis of various nucleotide sugars
Butanoate metabolism
Carbon metabolism
Citrate cycle (TCA cycle)
Cobalamin transport and metabolism
Cyanoamino acid metabolism
Cysteine and methionine metabolism
D–Amino acid metabolism
Fructose and mannose metabolism
Galactose metabolism
Glucagon signaling pathway
Glycerolipid metabolism
Glycine, serine and threonine metabolism
Glycolysis / Gluconeogenesis
Glyoxylate and dicarboxylate metabolism
Histidine metabolism
Lysine biosynthesis
Lysine degradation
Metabolic pathways
Methane metabolism
Microbial metabolism in diverse environments
Monobactam biosynthesis
Nicotinate and nicotinamide metabolism
Nucleotide metabolism
Other carbon fixation pathways
Other glycan degradation
Oxidative phosphorylation
Pentose and glucuronate interconversions
Pentose phosphate pathway
Peptidoglycan biosynthesis
Phenylalanine, tyrosine and tryptophan biosynthesis
Primary bile acid biosynthesis
Propanoate metabolism
Purine metabolism
Pyrimidine metabolism
Pyruvate metabolism
Secondary bile acid biosynthesis
Selenocompound metabolism
Sphingolipid metabolism
Starch and sucrose metabolism
Sulfur metabolism
Bacterial invasion of epithelial cells
Pathway
B. fragilis ECAM1
B. fragilis ECAM2
B. xylanisolvens ECAM
P. plebeius FMT
O. splanchnicus FMT
P. distasonis FMT
P. plebeius FMT
E. callanderi ECAM
B. comes HP
B. faecis ECAM
B. sp900066335 HP
D. formicigenerans FMT
CAG–115 sp900353158 FMT
CAG–217 sp900043635 FMT
G. formicilis FMT
G. formicilis HP
CAG–41 sp900066215 ECAM
A. muciniphila ECAM
Count
D
AMX
Infant 3
T1
T2 T3 T4
Bacteroides ovatus: G6mATC
M2
R
M1
E
G6mATC sites in Bacteroides ovatus
Methylation level (%)
Before AMX
Post AMX
Genome postion (Mb)
F
G6mATC site: Partially-methylated → Full-methylated
Site 1
Site 2
Site 3
sigma70
SusD
SusC
phoB
DUF
G
Donor
Recipient
FMT
0 day (D2)
1 year (R2.1y)
5 year (R2.5y)
H
G6mATC sites in Phocaeicola plebeius
Donor
Post FMT
I
G6mATC site: Full-methylated → Partially-methylated
Site 4
PPs
M
MBL
TonB
D2
R2.1y
R2.5y
J
Birth
Fed capsule
Collect stool
0
84
91 Days
n=4
K
RA6mATTY sites in Akkermansia muciniphila
Capsule
In vivo
Genome position (Mb)
L
RA6mATTY site: Partially-methylated → Full-methylated
PAP
prmA
pmp10

Figure 4. Site-specific ePV events are widespread across different cohorts and associated with bacterial adaptation to environmental stress

(**A**) For a specific bacterial strain, only a subset of cells exhibit methylation at a specific promoter, while the remainder remain unmethylated (left panel). The schematic in the middle panel depicts a transcription factor (TF) competing with an MTase at this specific site. The ePV site (highlighted within the dashed rectangle) was identified from multiple molecules using single-molecule methylation analysis. The right panel illustrates the deconvolution results obtained from this single-site methylation analysis. (**B**) Summary of all ePV sites detected in this study. Values were normalized based on the strain number within the order (diamond points highlighted in red). The methylation percentage within the order were also determined (box plot). Orders with fewer than five strains were merged into the 'Others' category. (**C**) Function categories of the genes affected by ePV sites. Heatmap showing the number of genes regulated by ePV, with host strain taxonomy (strain) as column label and gene function category (pathway) as row label. The color scale represents the number of genes, indicating the occurrence of each functional category per strain. (**D**) The top panel presents the time-line of AMX administration and sample collection for Infant 3. The bottom panel shows the involved methylation-related genes in a *B. ovatus* strain from Infant 3. (**E**) Distribution of partially methylated G6mATC sites across the *B. ovatus* genome. The genome consists of three contigs, randomly ordered along the x axis. The top three changed partially methylated sites were selected and labeled with numbers. (**F**) Changes in partially methylated G6mATC sites in response to AMX treatment. The motif sites were examined on both strands, and only sites that showed partial methylation on both strands were considered final ePV sites. The same criteria were applied to all other partially methylated sites in Fig. **4I**,**L**. The gene diagram shows the location of each site within the affected gene. **(E,F)** Pre- (black) and post-AMX (red) methylations were highlighted. (**G**) Sample collection information for Donor 2 and Recipient 2 from FMT cohort. (**H**) Distribution of partially methylated G6mATC sites across the *Phocaeicola plebeius* genome. The top four partially methylated sites were selected and labeled with numbers. (**I**) Changes in partially methylated G6mATC sites in response to FMT. The gene diagram shows the location of each site within the affected gene. **(H**,**I)** Pre- (black) and post-FMT (red) methylations were highlighted. (**J**) Design and sample collection information for Akk-fed mouse assay. (**K**) Distribution of partial methylated RA6mATTY sites across the *A. muciniphila* genome. Three of the partially methylated sites, specifically lowly methylated in capsule, were identified and labeled with numbers. (**L**) Changes in partially methylated RA6mATTY sites from capsule to *in vivo* colonization (mouse stool). The gene diagram shows

the location of each site within the affected gene. **(K**,**L)** Capsule (black) and *in vivo* (red) methylations were highlighted.

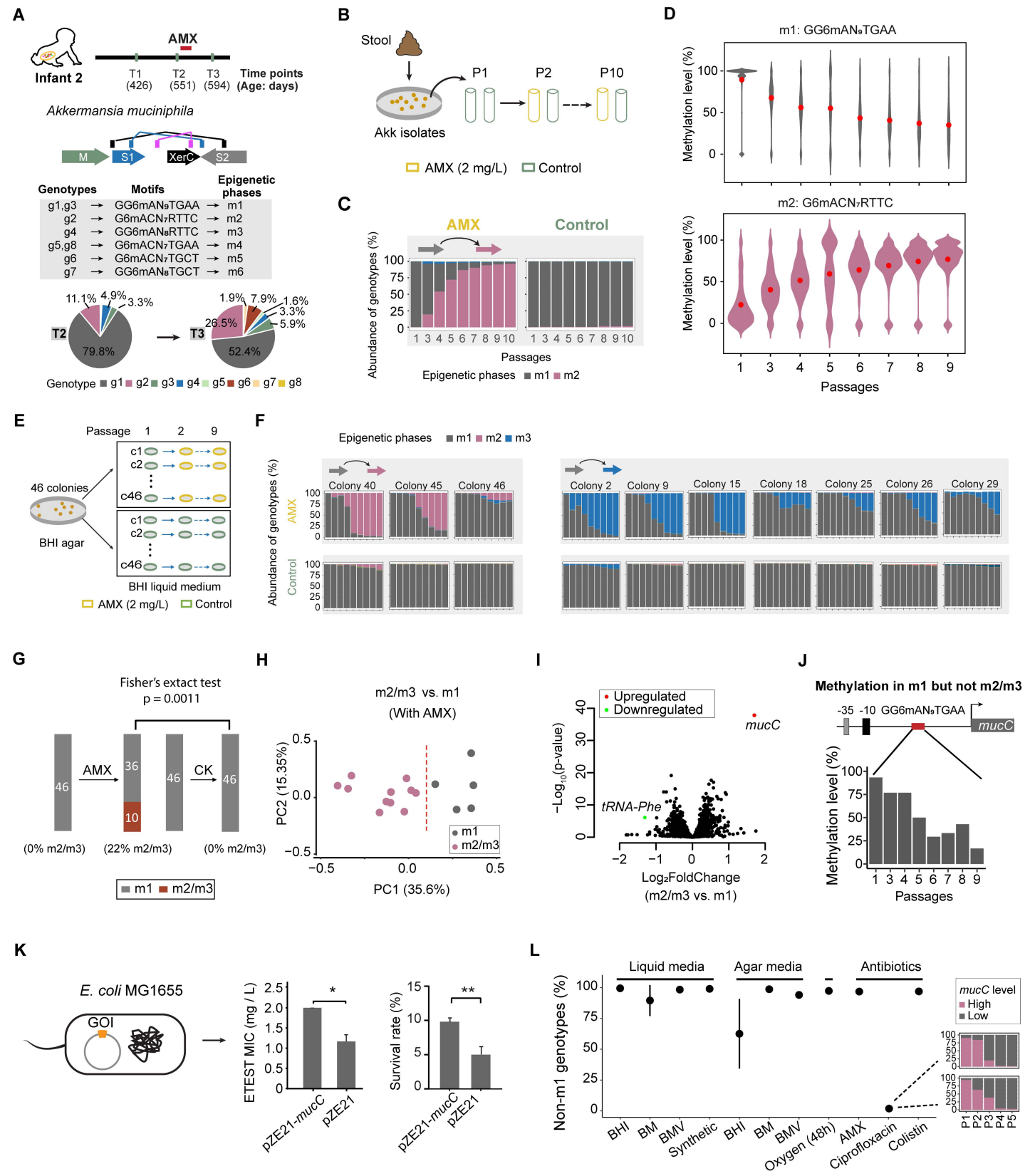
A
Infant 2
AMX
T1 (426)
T2 (551)
T3 (594)
Time points (Age: days)
Akkermansia muciniphila
M
S1
XerC
S2
Genotypes
Motifs
Epigenetic phases
g1,g3 → GG6mAN9TGAA → m1
g2 → G6mACN7RTTC → m2
g4 → GG6mAN8RTTC → m3
g5,g8 → G6mACN7TGAA → m4
g6 → G6mACN7TGCT → m5
g7 → GG6mAN8TGCT → m6
11.1%
4.9%
3.3%
T2
79.8%
1.9%
7.9%
1.6%
3.3%
5.9%
26.5%
T3
52.4%
Genotype g1 g2 g3 g4 g5 g6 g7 g8
B
Stool
P1
P2
P10
Akk isolates
AMX (2 mg/L)
Control
C
AMX
Control
Abundance of genotypes (%)
Passages
Epigenetic phases m1 m2
D
m1: GG6mAN9TGAA
m2: G6mACN7RTTC
Methylation level (%)
Passages
E
Passage 1 2 9
46 colonies
BHI agar
c1
c2
c46
BHI liquid medium
AMX (2 mg/L)
Control
F
Epigenetic phases m1 m2 m3
Abundance of genotypes (%)
AMX
Control
Colony 40
Colony 45
Colony 46
Colony 2
Colony 9
Colony 15
Colony 18
Colony 25
Colony 26
Colony 29
G
Fisher's extact test
p = 0.0011
AMX
CK
46
36
10
(0% m2/m3)
(22% m2/m3)
(0% m2/m3)
m1 m2/m3
H
m2/m3 vs. m1
(With AMX)
PC2 (15.35%)
PC1 (35.6%)
m1
m2/m3
I
Upregulated
Downregulated
mucC
tRNA-Phe
−Log10(p-value)
Log2FoldChange
(m2/m3 vs. m1)
J
Methylation in m1 but not m2/m3
-35 -10 GG6mAN9TGAA
mucC
Methylation level (%)
Passages
K
E. coli MG1655
GOI
ETEST MIC (mg / L)
Survival rate (%)
pZE21-mucC
pZE21
L
Liquid media
Agar media
Antibiotics
Non-m1 genotypes (%)
mucC level
High
Low
BHI
BM
BMV
Synthetic
BHI
BM
BMV
Oxygen (48h)
AMX
Ciprofloxacin
Colistin
P1
P2
P3
P4
P5

Figure 5. Characterization of a complex genome-wide ePV and its role in *Akkermansia* tolerance to AMX

(**A**) An ePV event observed in an *A. muciniphila* strain from Infant 2 in the ECAM cohort, mediated by three pairs of IR features, resulting in eight MTase genotypes (g1-g8) and six unique methylation sequence recognition motifs. The genotype was profiled in pre- and post-AMX treatment samples, AMX treatment suppressed the population of g1 (highlighted in grey) while promoting the growth of g2 (highlighted in purple) or other non g1 populations. (**B**) Experimental design of the *in vitro* passage assay. *A. muciniphila* was isolated from the fecal sample of Infant 2 (hereafter referred to as Akk_ECAM). After recovery on the BHI agar plates, a culture derived from a single colony was subjected to 2 mg / L AMX treatment for 10 passages. Cultures without AMX served as negative controls. (**C**) MTase genotype profiling using targeted sequencing. In response to AMX treatment, the MTase genotype changed from g1 dominance to g2 dominance, whereas g1 remained dominant through passage 10 in the absence of AMX. (**D**) Methylation motif changed in the Akk_ECAM strain upon AMX treatment. Genomic DNA was extracted, and whole genome shotgun sequencing was performed using Pacbio Sequel II platform. Single-molecule methylation analysis was then conducted to investigate the methylation level of the two motifs of interest. The results align with the *in vivo* findings, where AMX treatment decreased in the g1 methylation subpopulation (highly methylated at GG6mAN_9TGAA) and increased the non-g1 methylation subpopulation (highly methylated at G6mACN_7RTTC). (**E**) Experimental design of the 2nd batch of the *in vitro* passage assay. After recovery on BHI agar plates, 46 colonies were subjected to 2 mg / L AMX treatment for nine passages. Culture without AMX served as negative controls. (**F**) MTase genotype profiling with targeted sequencing. The MTase genotype changed from g1 dominance to non-g1 (g2-g4) dominance in response to AMX treatment, while g1 remained dominant through passage 9 in the absence of AMX treatment. (**G**) Summary of the 2nd batch *in vitro* passage assay. Genotypes were determined by targeted sequencing and categorized into two groups, g1 and non-g1 (g2-g4). Fisher's exact test was used to assess the statistical significance of the difference in proportions between pre- and post-AMX treatment. The results indicate a significant association ($p = 0.0037$), suggesting that g2-g4 genotypes were selected by AMX treatment. (**H**) Principal Component Analysis (PCA) analysis of RNA-seq-based expression data from in vitro passage samples. Under AMX treatment, g2/g4 samples were clearly separated from g1/g3 samples. Overall sample variation was evaluated using PCA, with the explained variation labeled as percentages on the axes. (**I**) DEGs were identified by comparing g1/g3 and g2/g4 samples under AMX treatment. Thresholds for DEGs were set at $\log_2$ fold changes > 1 or

< -1, with an adjusted p value < 0.05. *mucC* was the only upregulated gene (highlighted in red). *tRNA-Phe* was the only downregulated gene (highlighted in green). (**J**) The promoter elements and m1 methylation motif (GG6mAN_9TGAA) were represented in the diagram. The fraction of m1 methylation motif in the promoter of *mucC* decreased with AMX treatment. (**K**) AMX tolerance of *E. coli* and overexpression of the *mucC* gene. The *mucC* gene was codon-optimized, synthesized and fused with a 6 × HIS tag at the N-terminus. It was inserted into a medium-copy plasmid with a constitutive promoter, and AMX tolerance was assessed using the ETEST and CFU counting methods respectively. Each strain was tested in triplicate. (**L**) MTase specificity reversibility assay. g2/g4 dominant strains from the 2nd batch of the *in vitro* assay were treated under various conditions. The point plot shows the targeted sequencing results, while the zoomed-in bar plot indicates that ciprofloxacin treatment can reverse non-g1 (g2/g4) genotypes back to g1 genotype. Each condition was tested in duplicate.