nature communications

Article

https://doi.org/10.1038/s41467-026-70381-8

Dynamics of phage-host interactions in *Bacteroides fragilis* resolved by single-cell transcriptomics

Received: 19 December 2025

Accepted: 19 February 2026

Published online: 16 March 2026

Check for updates

Anika Gupta[1,2,10], Norma Morella[3,10], Dmitry Sutormin[1,10], Naisi Li[3], Karl Gaisser[1], Alexander Robertson[3,4], Yaroslav Ispolatov[5], Georg Seelig[2,6,7], Neelendu Dey[3,8,9] ✉ & Anna Kuchina[1,2,6] ✉

The interactions between lytic phages and their hosts are typically studied in bulk culture, which obscures cell-cell differences in infection susceptibility or expression of protective factors. Here, we use bacterial single-cell RNA sequencing to profile the transcriptomes of ~50,000 cells from cultures of a human pathobiont, *Bacteroides fragilis*, infected with a lytic bacteriophage. From a single sampling, we quantified the asynchronous progression of phage infection in individual bacterial cells and reconstructed the infection timeline, characterizing both host and phage transcriptomic changes as infection unfolded. Further, we discovered phenotypic subpopulations of bacteria that remained uninfected. Each cell's vulnerability to phage infection was influenced by expression of multiple genetic loci, most prominently phase-variable capsular polysaccharide (CPS) biosynthesis pathways and an operon predicted to encode fimbrial genes. These findings uncovered genome-wide phase variation and stochasticity that enable bacterial survival and re-growth without acquiring additional mutations. Overall, we establish bacterial single-cell RNA sequencing as a powerful platform for investigating the dynamics of host-phage interactions and revealing the roles of phase variation and stochasticity in bacterial defenses.

Lytic bacteriophages (hereafter "phages") infect specific bacterial hosts[1], resulting in rapid host death. Bacteria employ multiple resistance mechanisms to defend themselves from phages, including structural barriers such as polysaccharide capsules and an arsenal of immunity systems that detect and interfere with phage infection[2]. Despite successes in the identification of bacterial immunity systems across diverse bacterial taxa[3], most studies of bacterial responses to phage infection require phage isolation and culturing, remaining laborious and low-throughput. Additionally, relatively few anti-phage defense systems have been investigated in their natural host, with the majority of them studied via heterologous expression in a model organism[4–6]. High constitutive induction of these systems in a model organism does not reflect their natural expression patterns, which frequently vary across individual bacterial cells even in pure cultures

[1]Institute for Systems Biology, Seattle, WA, USA. [2]Molecular Engineering and Sciences Institute, University of Washington, Seattle, WA, USA. [3]Translational Science and Therapeutics Division, Fred Hutchinson Cancer Center, Seattle, WA, USA. [4]Molecular and Cellular Biology Graduate Program, University of Washington, Seattle, WA, USA. [5]Department of Physics, Center for Interdisciplinary Research in Astrophysics and Space Science, University of Santiago, Santiago, Chile. [6]Department of Electrical and Computer Engineering, University of Washington, Seattle, WA, USA. [7]Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, USA. [8]Microbiome Research Initiative, Fred Hutchinson Cancer Center, Seattle, WA, USA. [9]Department of Medicine, Division of Gastroenterology, School of Medicine, University of Washington, Seattle, WA, USA. [10]These authors contributed equally: Anika Gupta, Norma Morella, Dmitry Sutormin. ✉e-mail: ndey@fredhutch.org; anna.kuchina@isbscience.org

under identical conditions[7,8]. This behavior, termed phenotypic heterogeneity, may enhance the population-level fitness of the bacterium, for instance, when changes in the environment confer a strong benefit upon specific phenotypes.

The bacterial species inhabiting the human gut are known to exhibit extensive phenotypic heterogeneity, contributing to defenses against phages[9–11]. One example is the phase-variable expression of capsular polysaccharide (CPS) genes in *Bacteroides* spp., mediated by the stochastic inversion of promoters controlling the expression of biosynthetic operons to produce each of the multiple CPS types[12,13], some of which provide protection against phages[9]. Emerging evidence suggests similar expression heterogeneity of other anti-phage defense systems, such as CRISPR-Cas[14]. However, due to the challenges of transcriptome-wide single-cell gene expression profiling, how this cell-cell variability shapes the susceptibility to phage infection has remained largely unexplored.

To address this gap, we apply microbial split-pool ligation transcriptomics[8] (microSPLiT), a bacterial single-cell RNA sequencing technology, to characterize the transcriptional landscape of lytic phage infection in individual cells of *Bacteroides fragilis*. This commensal gut bacterium and pathobiont has been implicated in multiple disease states such as colorectal cancer[15,16], abscesses[17,18], undernutrition[19], and inflammatory bowel disease[20]. Our approach enabled the reconstruction of a phage infection cycle from a single sampling, leveraging the asynchronicity of infection across single cells[21]. We characterized the host and phage concurrent gene expression changes along the infection timeline, paving the way to using single-cell transcriptomics for studying host-phage dynamics in natural communities. Further, we quantified the chances of infection for cells within transcriptionally distinct *B. fragilis* subpopulations, identifying subpopulations of cells that are phenotypically protected from infection and contribute to the emergence of phage resistance. These findings reveal how extensive genome-wide phase variation and stochasticity enables the survival and regrowth of bacteria expressing these phenotypes without acquiring additional mutations. Overall, we establish single-cell transcriptomics as a valuable tool for uncovering previously hidden heterogeneity in infection dynamics and defense responses.

Results

Single-cell RNA sequencing identifies phage-infected and uninfected bacterial subpopulations

We isolated a lytic bacteriophage from King County wastewater that we named *Bacteroides* phage Bf12P1, with species- and strain-level host specificity to *Bacteroides fragilis* NCTC 9343. Whole genome sequencing (WGS) and morphological characterization (Fig. 1a and Supplementary Fig. 1a) were used to identify Bf12P1 as a novel siphovirus. Bf12P1 has a circular 47,370 bp genome with 72 coding sequences (Fig. 1b) and is most closely related to viral sequences reconstructed from human gut metagenomes[22] and, in particular, *Bacteroides* phage Barc2635 (GenBank: MN078104.1) with a nucleotide similarity of 89.0% (Supplementary Figs. 1b and 2). Bf12P1 readily infects and lyses *B. fragilis* NCTC 9343, as determined through liquid growth curves and visible lysis on plates (Fig. 1c and Supplementary Fig. 1c), with a latency time of 40 minutes and an average burst size of 59 ± 0.2 phage particles (standard deviation, Supplementary Fig. 1d).

To investigate the differences in infection susceptibility across individual cells in a population of *B. fragilis*, we infected the sample at a multiplicity of infection (MOI) 4×10^{-5} and collected phage-treated cells after 2.5 h of infection ("Phage-treated"), just before the projected onset of phage-induced lysis. In parallel, a set of *B. fragilis* cultures was prepared from the same stock to which phage had not been added ("Untreated"), and samples were collected at the same time point and optical density (Fig. 1c). Finally, we collected a similar set of samples using media with added deoxycholic acid (DCA), a secondary bile acid common in the gut environment that modulates bacteria–phage interactions[23–27]. We immediately fixed the samples with formaldehyde and proceeded with microSPLiT, a method we previously developed for single-cell RNA sequencing of bacteria[8].

We performed two independent microSPLiT experiments, named "M30" and "F3". While yielding similar results, F3 retrieved substantially more cells with high transcript content due to workflow optimizations (Supplementary Note 1 and Supplementary Figs. 3 and 4, Methods). After filtering out low quality cells, the no-DCA dataset contained 53,300 cells from two replicate F3 libraries (Fig. 1d), with medians of 79 and 66 *B. fragilis* mRNA transcripts per cell for the untreated and phage-treated samples, respectively (Fig. 1e). We also recovered 27 median phage transcripts per cell for the phage-treated sample (Fig. 1f). The summed gene expression profile correlated with the bulk RNA sequencing performed earlier (Supplementary Fig. 5). Unsupervised clustering of both samples revealed two main subpopulations of cells, largely discriminated by the presence of the phage transcripts (Fig. 1g, h and Supplementary Fig. 4a, c). Approximately two-thirds of the cells from the phage-treated sample contained phage transcripts and clustered separately from the untreated sample. The remaining phage-treated cells clustered together with untreated cells and lacked phage transcripts. We named this subpopulation of cells from the phage-treated sample "uninfected," as opposed to the "infected" subpopulation of the same sample (Fig. 1i). The infected subpopulation displayed a reduced median of 50 *B. fragilis* mRNA UMIs/cell, whereas we retrieved 100 median mRNA UMIs/cell for the uninfected subpopulation (Fig. 1j). The Bf12P1 genes were detected at a median of 51 UMIs/cell for the infected subpopulation (Fig. 1k). A similar distinction was observed in samples exposed to DCA (Supplementary Fig. 6a).

Clustering on the basis of both host and phage gene expression could be primarily driven by the presence of phage reads. To reveal host-specific transcriptional changes in response to bacteriophage infection, we removed phage transcripts from our dataset and re-clustered the data (Fig. 1l). We observed a new cluster, populated only by cells from the phage-treated sample, that we assumed was exhibiting the transcriptional response to phage infection; thus, we named it "responders" (Fig. 1l and Supplementary Fig. 6b). Indeed, when we extracted the barcodes of cells in the "responders" cluster and retrieved phage transcripts associated with these cellular identifiers, the vast majority of cells (98.7%) had an associated phage load and were previously assigned to the infected subpopulation (Fig. 1m). This analysis confirmed that the majority of infected *B. fragilis* cells displayed a distinct host gene expression profile associated with phage infection.

Examining the transcriptomic profile in the "responders" cluster, we found that it was predominantly defined by the expression of RNA polymerase subunits, replication proteins, and multiple ribonucleoside reductases (Supplementary Note 2 and Supplementary Fig. 6c). While overall this gene expression pattern reflected the expected host response to phage infection[28–31], it was an average representation of cells at different points in the infectious cycle. Prior to sampling, the phage-treated cell population had likely undergone multiple infectious cycles, the timing of which presumably varied between individual cells due to the intrinsic heterogeneity of host-phage interactions on the single-cell level. Thus, we next turned to resolving cell-to-cell variability in the progression of phage infection within the infected cells.

Single-cell host and phage gene expression reveals infection timeline and uncovers early and late phage gene modules

Subclustering the phage-treated sample by itself unveiled a gradient of phage transcript density across cells with a clear separation between the infected and uninfected subpopulations (Fig. 2a). To learn whether this pattern reflected differences in infection development between cells, we computed a pseudotime trajectory of the phage-treated

https://doi.org/10.1038/s41467-026-70381-8

sample. The diffusion pseudotime algorithm reconstructs the ordering of cells along a biological process by modeling their transcriptomic similarity as a graph and assigning each cell a value between 0 and 1 based on its diffusion distance from a root cell, indicating the position of a cell within the infection timeline[32]. We used the cell-to-cell differences in both the host and phage transcript content for pseudotime calculations. The resulting pseudotime trajectory aligned well with the distribution of phage transcripts across cells (Fig. 2b). While *B. fragilis* transcript content declined with pseudotime, reaching almost negligible values above a pseudotime value of 0.7, phage transcripts increased concurrently (Fig. 2c). We detected minimal to no phage transcripts between values 0.0 to 0.2, as cells within this range were

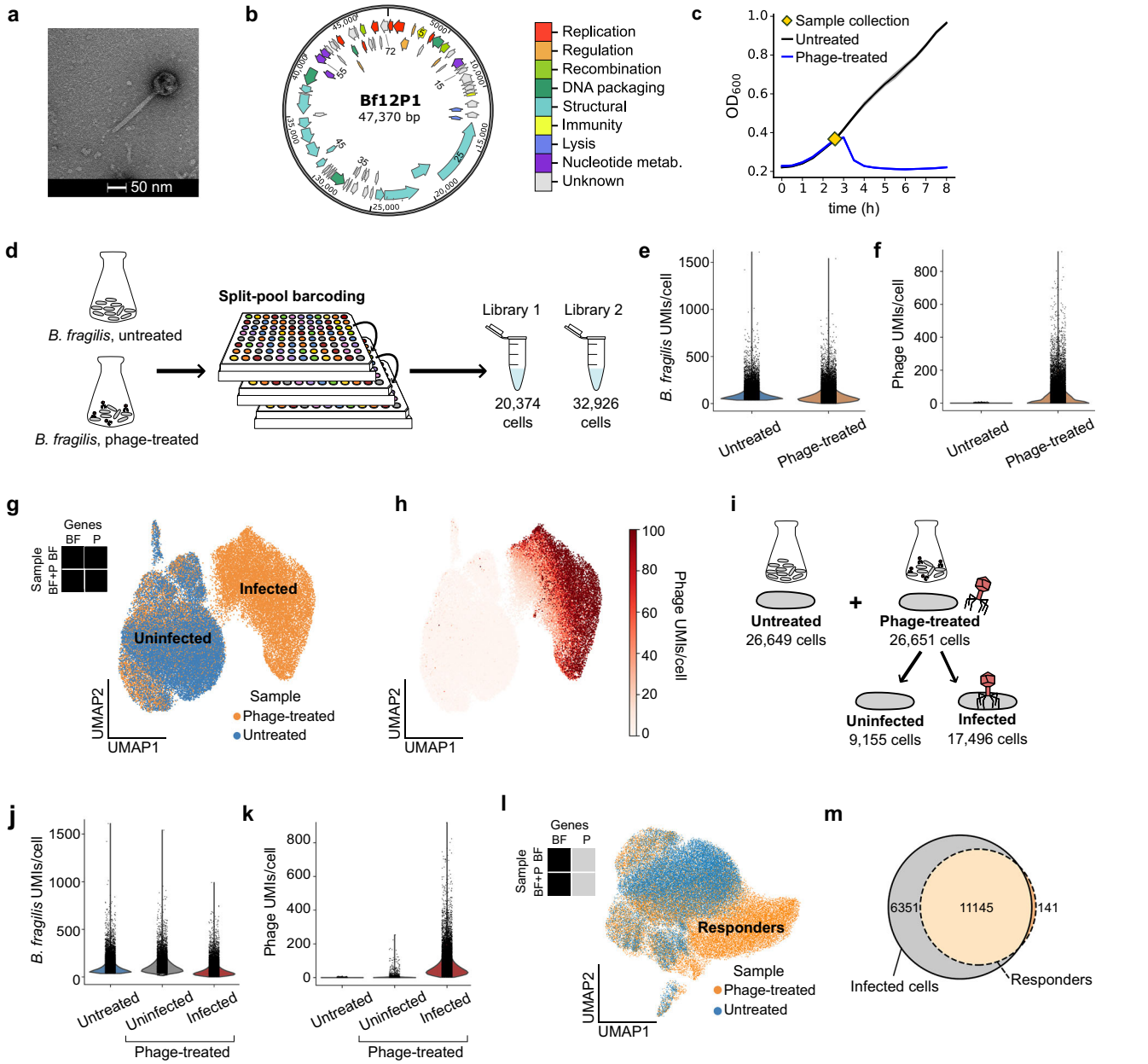

Fig. 1 | Single-cell transcriptomics of *B. fragilis* treated with Bf12P1 phage detects infected and uninfected cells within a single sample. a Electron microscopy image of siphovirus Bf12P1, representative of at least five independent experiments. **b** Bf12P1 genome with feature colors corresponding to putative gene functions and numerical labels indicating locations of each gene. **c** Growth curves of untreated (black) and phage-treated (blue) *B. fragilis* from an experiment prior to microSPLiT sampling (MOI of 4 x 10^{-5}). Symbol marks the time of sample collection chosen for the future M30 and F3 experiments. **d** Summary of F3 microSPLiT experiment on both untreated and phage-treated *B. fragilis* samples producing two replicate sequencing libraries. **e** *B. fragilis*, and **f** Bf12P1 unique molecular identifiers (UMIs)/cell, corresponding to individual detected transcripts, for untreated and phage-treated samples (F3 experiment, combined replicates). **g** UMAP embedding of both treated and untreated samples, using both *B. fragilis* and Bf12P1 genes, colored by sample. The two distinct parts of the UMAP plot representing the infected or uninfected cells are labeled accordingly. **h** Phage UMIs/cell overlaid on the UMAP from panel (**g**). **i** Sample definitions. The phage-treated sample after sequencing is divided into uninfected and infected subpopulations based on the location on the UMAP plot with indicated cell numbers. **j**, **k** *B. fragilis* and Bf12P1 UMIs/cell, respectively, for all subpopulations of cells, including uninfected and infected cells. **l** UMAP embedding of both phage-treated and untreated samples, constructed using only *B. fragilis* transcripts, colored by sample. A distinct cluster of cells from the phage-treated sample is labeled "responders." **m** Venn diagram showing the overlap between cells in the "responders" cluster with the cells in the infected subpopulation of cells defined in panel (**g**).

https://doi.org/10.1038/s41467-026-70381-8

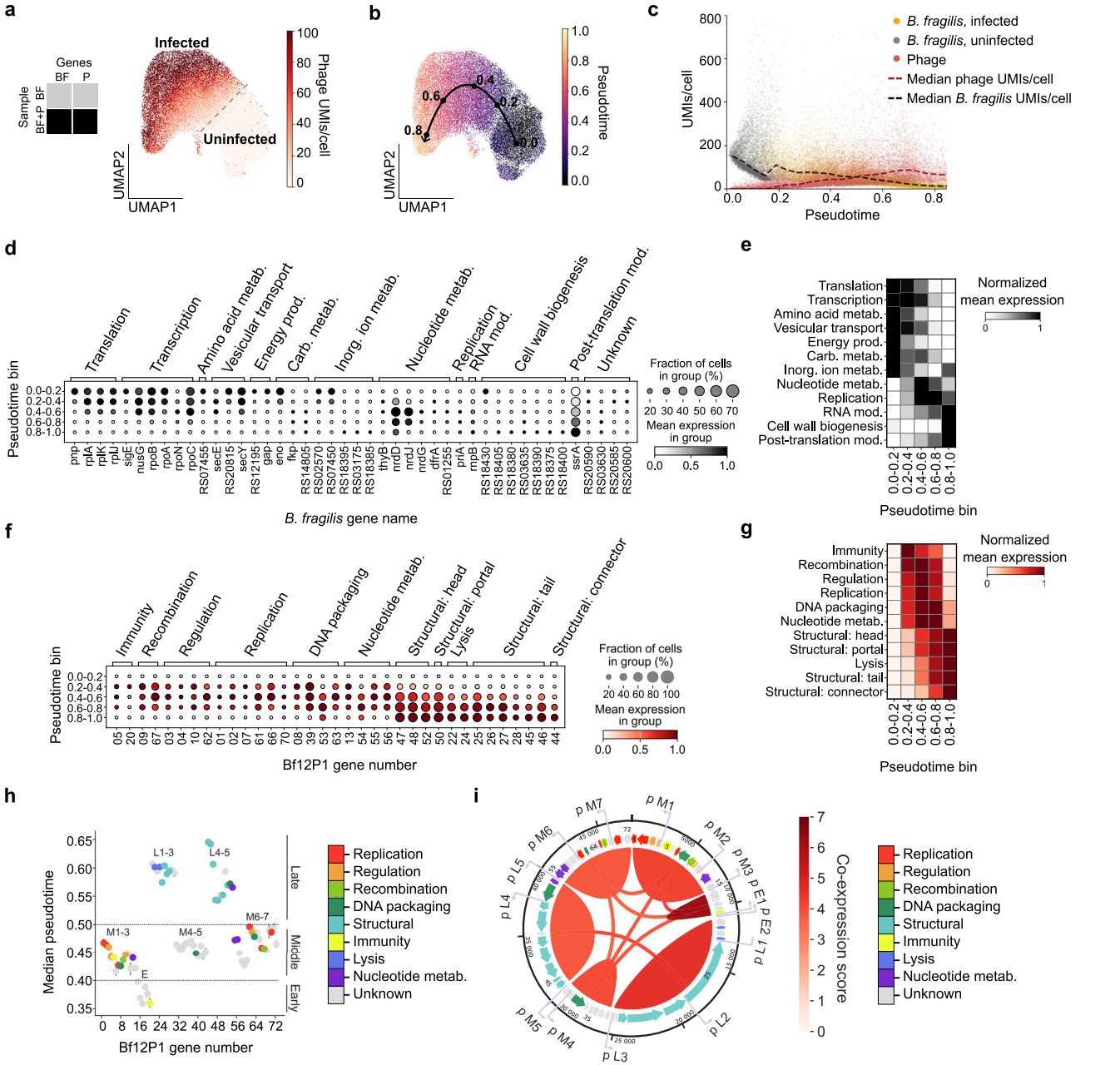

Fig. 2 | Host and phage gene expression profiles along the calculated infection trajectory. a UMAP embedding of the phage-treated sample using both *B. fragilis* and Bf12P1 genes, divided between uninfected and infected cell subpopulations. Phage UMIs/cell are shown in red. **b** UMAP plot from panel (**a**) colored by calculated diffusion pseudotime values, overlaid with a trajectory connecting cell regions of approximately 0.0, 0.2, 0.4, 0.6, and 0.8 pseudotime values. **c** *B. fragilis* UMIs/cell (gray and yellow for uninfected and infected cells, respectively) and phage UMIs/cell (red) in the phage-treated sample plotted against the pseudotime values with each dot representing a single cell. Median UMIs/cell across pseudotime are shown as a black line for *B. fragilis* and a red line for phage. **d** Expression of top 10 differentially expressed genes in the phage-treated sample across 5 equally sized pseudotime bins, categorized by function. **e** Summed raw *B. fragilis* gene expression for each functional category in each pseudotime bin, normalized by bin. **f** Expression of all phage genes in the phage-treated sample across five pseudotime bins, grouped according to function. **g** Summed raw Bf12P1 gene expression for each functional category in each pseudotime bin, normalized by bin. **h** Median pseudotime of expression for each phage gene is shown as a dot colored according to functional category and arranged by genomic location. Error bars are vertical 95% confidence intervals for each median pseudotime value. Late, middle, and early stages of pseudotime are indicated according to numerical y-axis values, and gene clusters are referenced. **i** Diagram of Bf12P1 genome showing the predicted putative phage promoters named according to expression stage as shown in (**h**). Co-expression between gene modules is indicated by the arcs colored by the calculated co-expression score.

primarily uninfected. These transcriptomic patterns show that the calculated pseudotime trajectory reflects the progression of phage infection within *B. fragilis* cells, with an initial rapid shutdown of most host gene expression followed by induction of cellular machinery for phage replication, and degradation of the host, accompanied by an increase in phage abundance.

Next, we investigated how *B. fragilis* transcriptomes changed along the infection trajectory. Differential expression analysis of the infected sample across pseudotime intervals revealed a pattern of *B. fragilis* gene expression in which certain genes were primarily expressed at early or late stages of infection (Fig. 2d, e and Supplementary Fig. 7a, c). In early pseudotime (0.2 to 0.4), we observed high

expression of genes encoding RNA polymerase subunits (*rpoA*, *rpoB*, *rpoC*), a SigE sigma factor (*sigE*), the transcription termination factor NusG (*nusG*), and ribosomal proteins (*rplA* and *rplK*). Expression of these transcription- and translation-associated genes reflects the phage-induced upregulation of host cellular machinery to facilitate expression of its genome[33].

In the mid-pseudotime range (0.4 to 0.6), we saw an upregulation of expression of *thyB* and *dfrA* genes responsible for conversion of deoxyribonucleotide dUMP to dTMP, as well as ribonucleoside reductase (RNR) genes (*nrdJ*, *nrdD*, and *nrdG*) encoding two different classes of RNRs and representing two alternative pathways of ribonucleoside conversion to deoxyribonucleosides for DNA synthesis[34] (Fig. 2d, e). Thus, in mid-infection, Bf12P1 is likely mediating the conversion of NTPs into dNTPs to promote its replication. A similar gene expression pattern with strong upregulation of RNRs has been observed in a different siphoviral infection of a marine bacterium[31]. Since Bf12P1 does not encode *nrd* genes like many other phages[35], it may be repurposing *B. fragilis* genes for replication of its DNA.

In late pseudotime (above 0.6), we observed only a small number of host genes that were selectively enriched, including the *priA* gene encoding a primosomal protein involved in replication of some phages[36]. We also observed strong upregulation of transfer-messenger RNA (tmRNA) or SsrA, a component of the damaged protein degradation system[37,38](Fig. 2d, e). This could indicate that the phage is using SsrA for maturation of some of its proteins, or the overall damage accumulation in bacterial cells approaching lysis. Alternatively, this housekeeping RNA can be more stable than mRNA in late stages of phage infection amid widespread degradation of the host transcriptome, an observation supported by literature on other phage-bacteria systems[39,40].

Observed changes in the Bf12P1 gene expression over pseudotime intervals corresponded well with the *B. fragilis* transcriptome changes (Fig. 2f, g and Supplementary Fig. 7b, d). Genes enriched in the pseudotime range between 0.2 and 0.6 were related to DNA synthesis, DNA binding, and recombination, marking the period of phage replication in the host. As pseudotime values increased above 0.6, we found predominant expression of lysis, structural, and capsid maturation genes (Fig. 2f, g).

To better understand the relationship between the expression and genomic localization of phage genes, we next calculated the median pseudotime associated with the expression of each phage gene. We then mapped the genomic co-localization of genes (Fig. 2h). This analysis revealed that phage genes are organized in three distinct modules: early, middle, and late (Fig. 2h and Supplementary Data 7). Genes within each module are usually colocalized in the phage genome and were highly co-expressed (Fig. 2h, i). By predicting the putative phage promoters, we further resolved one gene cluster in the early gene module regulated by two putative promoters, three gene clusters in the middle module regulated by seven putative promoters, and two gene clusters in the late module regulated by five putative promoters, where gene clusters likely represent operons (Fig. 2i and Supplementary Figs. 8 and 9).

Given the lack of known functions for the early module phage genes (16–20), we performed structural modeling with AlphaFold 3[41] and searched for similar protein structures, revealing that the gene 20 product is similar to the AcrIF9 anti-CRISPR protein, an inhibitor of the I-F CRISPR system[42] (Supplementary Data 8). We found that gene 20 was conserved and had homologs in 70 closely related gut phages[22] (see Supplementary Fig. 2). We speculate that the earliest expressed Bf12P1 genes are responsible for the initial host takeover by inhibiting its defense systems. Several conserved genes of the middle module are involved in nucleotide metabolism: gene 13 encodes putative thymidylate synthase and gene 54 encodes putative RNase E. These genes are likely involved in active conversion of host transcripts into ribonucleotides and provide substrates for host-encoded RNRs, mirroring the corresponding upregulation of these genes in the host in the middle pseudotime range. Two genes (44-45) within the late module gene cluster 2 are characterized by the latest expression in the infection trajectory (median pseudotime values >0.64) (see Supplementary Fig. 9). These genes encode a tail protein (gene 45) and a tail terminator protein (also known as connector, gene 44), presumably involved in the latest stage of phage particle assembly[43,44].

Combined, these observations demonstrate that microSPLiT enables reconstruction of the bacteriophage infection timeline from individual cells, profiling the concurrent transcriptomic changes in the host and the phage. Moreover, the pseudotime approach reveals early and late phage gene modules, with the former likely responsible for the rapid disabling of the host defenses to promote infection.

Phase-variable genomic loci coordinate cellular susceptibility to bacteriophage infection

To study the effects of differences in gene expression on phage susceptibility, we examined transcriptional heterogeneity in our dataset. Within the untreated and uninfected cells, we discovered several transcriptional clusters (Fig. 3a). A distinct cluster was marked by the expression of an operon transcribed from a phase-variable promoter activated by the tyrosine site-specific recombinase, Tsr15. This operon encodes components of type 3 fimbria that, when the promoter region downstream of Tsr15 is in the ON state, promote aggregation and biofilm formation[45]. Another separate cluster was defined by the expression of several co-directional genes organized in an operon (BF9343_RS20580-20610, Supplementary Data 3). A similarity search based on predicted protein structures using the AlphaFold database indicates that the operon presumably encodes a variant of fimbriae. This operon is adjacent to genes encoding a putative AraC-family transcriptional regulator ("TR", BF9343_RS20615) and recombinase (BF9343_RS20620, identified as Tsr16[46]). The TR gene is flanked by perfect 23 bp-long inverted repeats; by performing in-depth analysis of F3 sequencing data, based on reads alignment, we discovered the inversion of the TR gene in cells expressing the Tsr16-adjacent operon (Fig. 3b and Supplementary Fig. 10a). We also found that the BF9343_RS20580-20610 operon was transcriptionally active when the TR gene was co-oriented with the rest of the operon, and silent when the gene was inverted (Fig. 3b and Supplementary Fig. 10b). We conclude that the Tsr16 gene cluster represents another phase-variable region in the *B. fragilis* genome controlled by the Tsr16 recombinase.

Finally, two distinct clusters were marked by the expression of genes belonging to distinct capsular polysaccharide (CPS) operons controlled by phase-variable promoters: namely, PSF and PSG[12]. Overall, *B. fragilis* has eight operons encoding genes for biosynthesis of distinct capsule types, named PSA through PSH. Seven of them, excluding PSC, are expressed from a promoter which can be stochastically inverted by the Mpi recombinase and is only active in the "ON" orientation[47]. When plotting the combined expression of genes within each CPS operon on the cluster map for the entire dataset, we observed distinct patterns of expression (Supplementary Fig. 11). We saw no co-expression between any two CPS types in the same cell (Supplementary Fig. 12), consistent with prior literature showing that despite the frequent presence of multiple CPS operon promoters in the "ON" state per cell simultaneously, only a single capsule type is produced in each cell[48,49]. Meanwhile, genes within each operon were highly co-expressed with clear operon boundaries indicated by a drop in co-expression score (Fig. 3c and Supplementary Note 3). The identification of multiple transcriptionally distinct subpopulations of *B. fragilis* within the untreated sample highlights the ubiquitous phase variation creating extensive phenotypic heterogeneity in this bacterium.

Since PSF and PSG expression defined two clear clusters for uninfected and untreated cells, we hypothesized that these phase-variable regions may protect against phage infection. CPS biosynthesis

https://doi.org/10.1038/s41467-026-70381-8

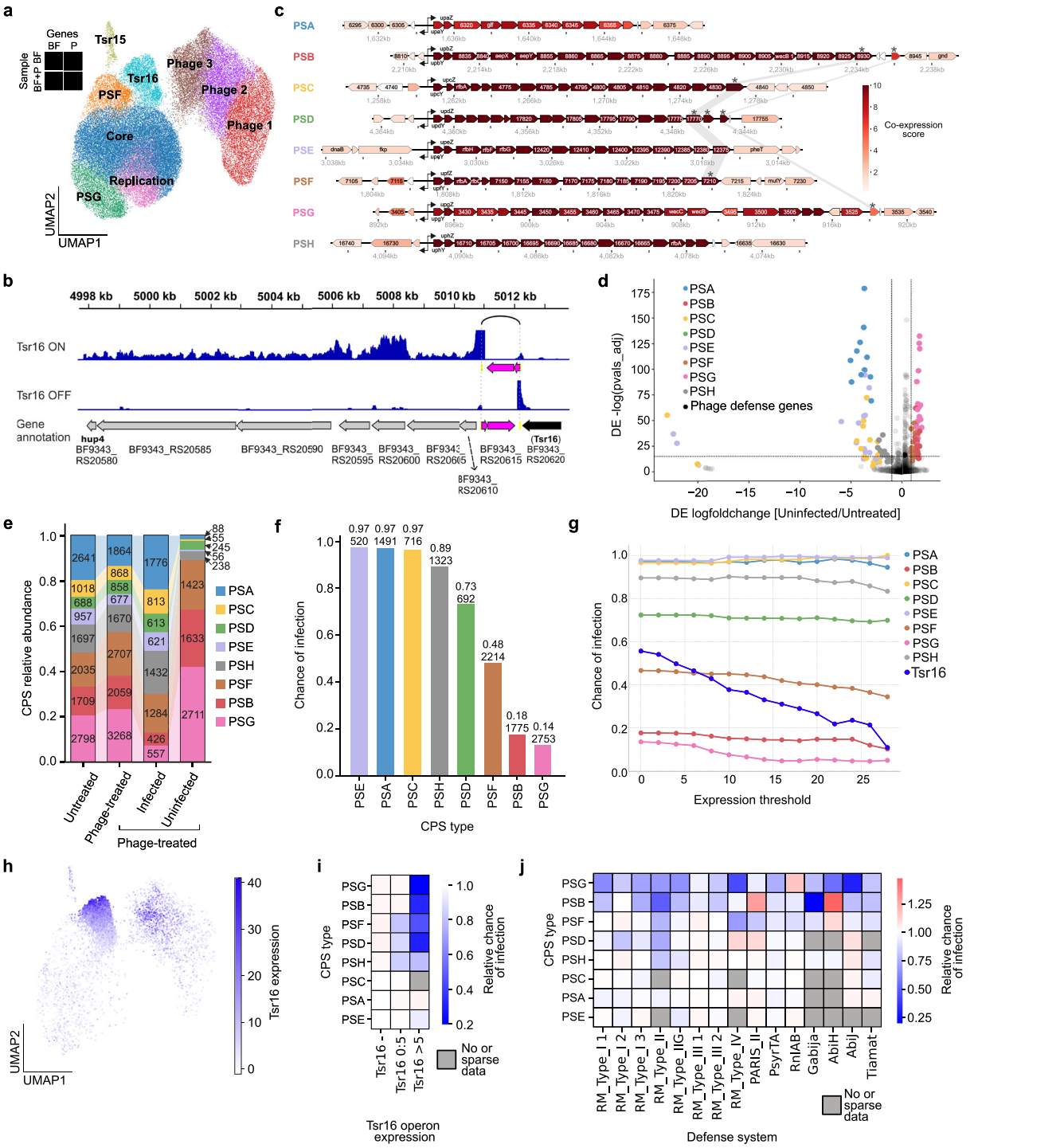

Fig. 3 | Cells expressing distinct capsular polysaccharide (CPS) types and high levels of the Tsr16-adjacent region are protected from infection. a UMAP of clusters for phage-treated and untreated samples, using host and phage genes, labeled by expression patterns. **b** Density of microSPLiT reads for the Tsr16-associated gene cluster. Read density is shown for the reference genome (top) and a sequence with an inverted regulatory region (bottom). Repeat regions presumably involved in site-specific recombination are within black dotted lines and connected with a black arch. **c** Operon maps for eight CPS types, with co-expression values between each gene and all other genes within the operon shown by color (see Supplementary Data 2). Three flanking genes per operon were included. Uniquely identified genes at each operon end with high co-expression are indicated with an asterisk. Gene paralogs with >80% sequence identity are highlighted. **d** Differentially expressed genes between uninfected and untreated cells were identified using a two-sided *t*-test with FDR correction and filtered by fold-change and test score. Defense genes were annotated using DefenseFinder[50]; genes are colored by CPS type. **e** Proportions of cells expressing each CPS for untreated, phage-treated, infected, and uninfected cells. **f** Infection chance of cells expressing each CPS, calculated as the proportion of infected cells in the group. **g** Calculated infection chance for cells expressing the Tsr16 gene cluster or CPS above the indicated expression thresholds. **h** Summed expression for the Tsr16-adjacent operon shown on a UMAP of all samples with *B. fragilis* and phage transcripts. **i**, **j** Relative infection chance for cells co-expressing each CPS type and **i** Tsr16 cluster with indicated expression thresholds or **j** anti-phage defense mechanisms simultaneously. Infection chances were calculated for cells with >0 expression of the indicated defense systems. Subpopulations with <10 cells were excluded (gray). For significance values, see Supplementary Fig. 16 and Supplementary Note 5.

genes were the most differentially expressed gene group between infected and uninfected cells, with PSF, PSB, and PSG capsular types strongly enriched in the uninfected cells (Supplementary Fig. 13). Meanwhile, none of the known anti-phage defense systems which we annotated in *B. fragilis* using DefenseFinder[50], nor Tsr15 and Tsr16 regions, were differentially expressed (Supplementary Figs. 13 and 14).

In the comparison between uninfected and infected samples, the differentially expressed genes may reflect both the changes induced by phage and the selection of protective phenotypes from the pre-existing heterogeneous pool of states. To remove the phage-induced changes from our analysis, we compared gene expression between uninfected and untreated cells (Fig. 3d). Again, the CPS biosynthesis genes were the strongest differentially expressed group (Fig. 3d), suggesting that a portion of cells were not uninfected by chance, but instead protected against phage infection by upregulation of these genes.

In an analysis of the distributions of cells with distinct capsular phenotypes across our dataset, we did not observe substantial differences in CPS operon expression composition between untreated and phage-treated *B. fragilis* samples, suggesting that short-term exposure to phage did not induce CPS switching. The uninfected and infected subpopulations, however, had markedly different CPS distributions (Fig. 3e and Supplementary Fig. 15), suggesting phage-mediated selection for specific capsule types. By calculating infection chances for cells expressing each capsule, we found that infection chances span a wide range between 0.14-0.18 for the most protective capsules, PSG and PSB, to 0.97 for the most vulnerable capsules, PSC, PSA, and PSE (Fig. 3f). Interestingly, two capsule types, PSF and PSD, had intermediate infection chances of 0.48 and 0.73, respectively, suggesting that the infection outcome is more complex than a simple selection of fully "vulnerable" capsules over fully "protective" ones by phage (Fig. 3f).

We next asked whether the degree of phage susceptibility or resistance for each capsule type varied proportionally with the strength of CPS or Tsr16-adjacent operon expression. The capsule type produced, rather than the CPS operon expression level, determined phage infection chance (Fig. 3g). However, unlike the capsular phenotypes, Tsr16-mediated protection strongly increased in a dose-dependent manner, with infection chance dropping essentially to zero with high expression levels (Fig. 3g, h). To investigate the combined effects of Tsr16 gene cluster expression in the context of each capsule, we calculated the chance of infection for cells of each capsule type expressing varying levels of the Tsr16 gene cluster (Fig. 3i). This analysis revealed a strong drop of infection chance in cells with high expression of Tsr16-adjacent region together with protective capsules PSG (4.1-fold, to 0.034) and PSB (2.7-fold, to 0.067). Fimbrial components, the predicted products of genes in this operon, could potentially act as phage decoys or promote bacterial aggregation that shields some cells from phage, especially in concert with a protective capsule[51].

We also examined whether known anti-phage defense mechanisms contributed to protection from phage in the context of each capsule by calculating infection chances for cells expressing these systems concurrently with each capsular operon. *B. fragilis* NCTC 9343 encodes 15 putative anti-phage defense systems, including multiple restriction-modification systems (RM) of types I, II, III, and IV, DNA degradation machinery (Gabija), abortive defense mechanisms (AbiJ, AbiH), toxin-antitoxin (TA) systems (PARIS, PsyrTA, RnlAB), and some systems with yet unknown mechanisms (Tiamat) (Supplementary Data 7). Quantifying infection chances for cells expressing these systems revealed that expression of several anti-phage systems enhanced phage protection in cells with protective capsule types (Fig. 3j). In particular, the expression of the AbiJ system simultaneously with the PSG capsule, and the expression of Gabija together with the PSB capsule, was associated with essentially no chance of infection (<0.05). AbiJ, classified into "Abi" or abortive infection proteins, consists of N-terminal and C-terminal domains, the latter of which is a HEPN domain containing two catalytic residues required for nucleolytic activity[52]. Gabija is a complex of GajA and GajB proteins involved in DNA cleavage and nucleotide hydrolysis, respectively, activated by foreign DNA binding[53,54]. Cells expressing RM type-IV[5,55] together with PSG also showed very low infection chances (0.059) (Fig. 3j). Altogether, these data suggest that variable expression of select genomic loci, such as the Tsr16 gene cluster and certain anti-phage systems, may act in concert with capsular variation to protect the cells from phage, although the details of these interactions remain to be characterized.

Finally, to determine if the discovered protective factors had a similar effect under different environmental conditions, we calculated the infection chances for cells with different capsule types from the F3 dataset samples obtained in the presence of DCA. Concordant to the obtained results without DCA, we found that cells expressing PSB and PSG had low infection chances (Supplementary Fig. 17a), and the Tsr16-adjacent operon expression protected from phage infection in a dose-dependent manner (Supplementary Fig. 17b). Interestingly, infection chances were overall lower in presence of DCA which indicates that this bile acid may act as a protective agent, for example, affecting phage binding to the cell surface[26].

Taken together, these observations show that the expression of distinct phase-variable loci such as CPS operons and the Tsr16 gene cluster strongly influences *B. fragilis* vulnerability to Bf12P1 infection across conditions.

Selection for protective CPS types underlies Bf12P1 phage resistance

To experimentally confirm that certain capsule types indeed conferred protection against the Bf12P1 phage, we used *B. fragilis* NCTC 9343 'phase-locked' strains unable to switch between CPS types[55]. Specifically, we used four previously characterized strains: "PSA/PSE", "PSB/PSG", "PSD/PSH", and "PSC", referring to the CPS operon promoters locked in the "ON" orientation in each strain[55]. We confirmed the promoter orientations in these strains and demonstrated that on a transcriptional level, each strain expressed a single dominating type of capsule: PSA in the "PSA/PSE", PSG in the "PSB/PSG", PSH in the "PSD/PSH", and PSC in the "PSC" strain (Supplementary Fig. 18).

In a liquid culture phage infection assay, we observed that "PSA/PSE", "PSD/PSH", and "PSC" strains, all expressing phage-sensitive capsules according to the microSPLiT data, exhibited a sharp drop in optical density and failed to recover when incubated with the Bf12P1 phage (Fig. 4a). The wild-type culture, challenged with the same phage titer, was lysed after 5–6 h of incubation but then resumed growth, presumably due to propagation of phage-resistant bacteria (Fig. 4a). The growth of the "PSB/PSG" strain, however, was unaffected by phage. A phage plaque assay similarly showed that all tested CPS 'phase-locked' strains were highly sensitive to phage except for the "PSB/PSG" strain, for which no plaques were observed even at the highest MOI (Fig. 4b and Supplementary Fig. 19a). These results demonstrate that expression of a capsular type predicted to be the most protective by microSPLiT data, PSG, is rendering the bacteria resistant to phage infection.

For additional confirmation, we collected phage-resistant bacteria in various conditions, including the presence of bile acids (deoxycholic or taurocholic acids) (Supplementary Note 4, Supplementary Data 5). Specifically, we collected eighteen aliquots of wild-type culture (referred to later as bulk culture samples or BC) and isolated colonies (referred to later as the resistant isolates) from the outgrowth following phage exposure (Fig. 4c). Using WGS, we examined the CPS invertible promoter orientations in resistant isolates and bulk culture samples, finding that the majority of samples harbored the PSG operon

https://doi.org/10.1038/s41467-026-70381-8

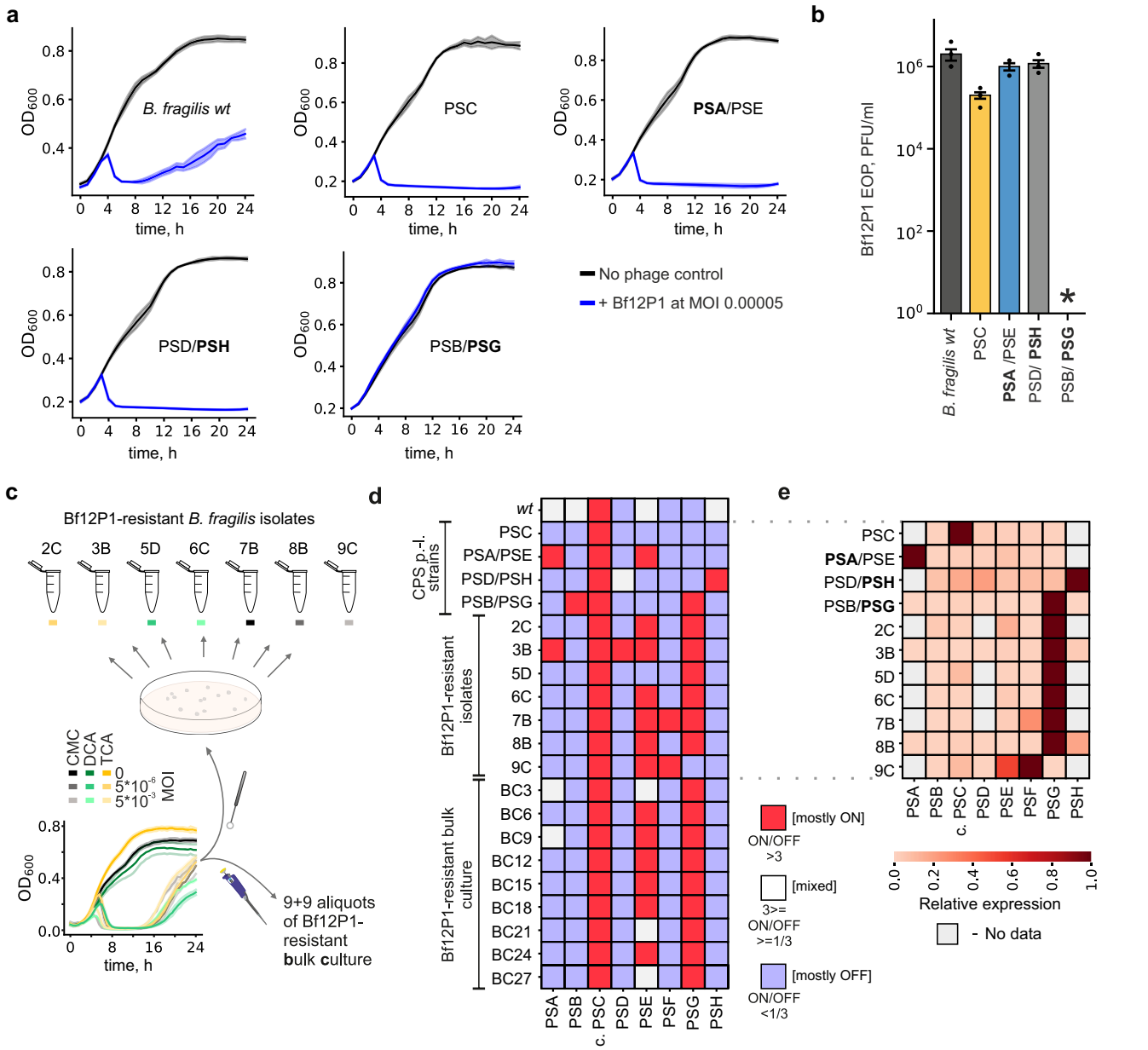

Fig. 4 | Specific capsular polysaccharide types provide resistance to Bf12P1 phage. a Growth curves for the liquid culture phage assay performed with CPS 'phase-locked' *B. fragilis* strains and the wild-type (*wt*) *B. fragilis* strain. Bf12P1 was added at time point 0 at MOI of 4 x 10^{-5} (blue curves), and untreated samples are shown as controls (black curves). Shaded curves represent a 0.95 confidence interval of the mean of four biological replicates. **b** Quantification of a phage plaque assay performed with CPS 'phase-locked' *B. fragilis* strains and the *wt B. fragilis* strain. Bars, error bars, and black dots represent, respectively, the mean of three biological replicates, the standard error of the mean, and individual data points. Black asterisk indicates no plaques observed for the "PSB/PSG" 'phase-locked' strain. **c** Schematic of an experiment for selection of Bf12P1-resistant isolates and bulk culture aliquots. **d** CPS operon promoter states in the *wt B. fragilis*, CPS 'phase-locked' strains, Bf12P1-resistant isolates, and Bf12P1-resistant bulk culture aliquots after incubation with phage at 0.005 MOI measured from WGS data. "c.PSC" denotes the non-invertible, "constantly ON" promoter for the PSC operon. **e** Relative expression of CPS operons in CPS 'phase-locked' strains and Bf12P1-resistant isolates, normalized to the operon with maximal expression for every strain or isolate.

promoter in an "ON" orientation. The only exception was resistant isolate 9C, which nonetheless had the PSF operon promoter in the "ON" state (Fig. 4d and Supplementary Fig. 19b), encoding a capsule type predicted to provide 'intermediate' protection (see Fig. 3f). We confirmed that despite multiple CPS operon promoters in the "ON" state in both resistant isolates and bulk culture samples, only PSF and PSG operons were expressed (Fig. 4e). Since the initial culture had PSG and PSF promoters mostly in the "OFF" position, these findings indicate strong selection for the protective capsular types in the presence of Bf12P1 phage.

To check whether any of the isolates acquired mutations which could be underlying resistance against Bf12P1, we performed variant calling on WGS data, which revealed one non-ancestral mutation present in five resistant isolates and the majority of resistant BC samples (Supplementary Data 5). This mutation, however, was not present in the "PSB/PSG" strain resistant to phage, suggesting it was not essential for resistance against Bf12P1. Of note, the "PSB/PSG" strain did not have any additional mutations compared to the ancestral strain (Supplementary Data 5). We also found no reads mapped to the Bf12P1 genome in the WGS data for resistant isolates or

BC samples, dismissing lysogeny with Bf12P1 as a potential resistance mechanism[56].

B. fragilis NCTC 9343 also encodes two putative CRISPR-Cas systems of class I subtype IB and class 2 subtype IIC. Notably, WGS analysis of phage-resistant isolates and bulk culture samples revealed that a spacer perfectly matching the Bf12P1 genome was newly acquired by the CRISPR cassette associated with the IIC system in the 6C isolate (Supplementary Fig. 20). Interestingly, we found another spacer in this cassette aligning to the Bf12P1 genome with two mismatches. This partially matching spacer is present in the ancestral strain and all its derivatives, supporting the existence of a history of interactions between Bf12P1, or closely related phages, and *B. fragilis* NCTC 9343. Acquisition of a new spacer demonstrated that the CRISPR-Cas system in this strain is functional and likely provides additional adaptive immunity to bacteriophage in cells that were protected from the initial phage attack by phase variation.

Combined, these data indicate that selection for protective capsular phenotypes, rather than accumulation of specific mutations, underlies resistance to Bf12P1.

Discussion

Bacterial single-cell transcriptomics is a powerful tool for unraveling key aspects of lytic bacteriophage infection from a single sampling prior to lysis. For instance, we reconstructed the stages of bacteriophage infection using the asynchronicity of infection across single cells[21]. We classified phage genes into functional modules which contained uncharacterized genes without homologs alongside genes with inferred functions. Consequently, we assigned putative function to novel phage genes, including genes that are expressed at the earliest point in infection and are presumably involved in disabling host defenses. By analyzing both the phage and host gene expression patterns concurrently as infection unfolded, we demonstrated how this approach can be used to decipher complex regulatory programs involved in host-bacteriophage interactions. Our work also opens the door toward direct high-throughput profiling of phage infections in a complex environment, potentially bypassing the need to isolate and culture the phage and its host. This capacity may help overcome one of the primary challenges of studying bacteria–phage interactions.

We also demonstrated the utility of microSPLiT for studying phenotypic heterogeneity in anaerobic bacterial species found in human gut microbiota, identifying multiple phase-variable regions active in different cell subpopulations. These encompass capsular polysaccharide and fimbrial loci and an uncharacterized operon (Tsr16 gene cluster), which was heterogeneously expressed during exponential growth in culture. This operon affected phage susceptibility in a dose-dependent manner and led to almost complete phage resistance at high levels of expression in cells concurrently expressing PSG and PSB capsules. Moreover, we used microSPLiT data to predict the regulation of expression of this operon, likely expressing fimbrial components, by an adjacent phase-variable transcription factor.

Protection from phage infection for only a few cells in a population could arise from immunity mechanisms heterogeneously induced in cells by exposure to phage. Alternatively, or in combination, it could result from preexisting heterogeneity of cellular attributes irrespective of phage treatment, leading to different degrees of vulnerability to phage. Here, we found that the preexisting phase-variable expression of different CPS types and either select phase-variable loci (Tsr16-adjacent operon) or anti-phage defense systems, such as AbiJ and Gabija, provided near-complete protection against Bf12P1 phage. While not all defense systems are expected to act synergistically with a protective CPS, our proof-of-concept experiments suggest that by acting together, some of these variable traits may allow the population to survive phage attack by selection and propagation of small subpopulations carrying protective combinations. In line with that, the estimated size of a resistant population (337 cells or 1.3% of the "Phage-treated" sample) matched well with a 0.5–1.5% prediction made by modeling of *B. fragilis* culture growth in the presence of Bf12P1 phage (Supplementary Note 6 and Supplementary Fig. 21). Interestingly, when we explored whether this phenotypic landscape was sensitive to environmental conditions, such as growth in different media, we found that while the vulnerability of different CPS types to phage diminished after bile acid exposure, the shape of the overall protective landscape remained relatively stable (see Supplementary Fig. 17).

The novel siphovirus Bf12P1 that we isolated in this study is closely related to other recently described lytic phages targeting *B. fragilis*: VA7[57], vB_BfrS_23[58], and Barc2635[11] (see Supplementary Fig. 1). Of these, phage VA7 has been positively evaluated for therapeutic potential against enterotoxigenic *B. fragilis* infection in colonic epithelial cell culture models[57]. Interestingly, mice monocolonized with *B. fragilis* and Barc2635 exhibited reduced PSA capsule expression[11], suggesting that this related phage targets similar sensitive capsular phenotypes. As PSA is the capsule type most associated with abscess formation by B. fragilis in vivo[59,60], it is noteworthy that we identified PSA as the most vulnerable capsule type to Bf12P1 infection. Mapping the phage-resistant host phenotypes for diverse phages can enable the rational design of phage cocktails to prevent bacterial resistance and potentiate phage therapies.

In conclusion, our study offers a novel conceptual lens to interrogate phage-host interactions and bacterial anti-phage defense mechanisms, highlighting phenotypic heterogeneity of these systems as a primary driver of phage evasion for organisms such as gut *Bacteroidales*. The widespread phase variation throughout the *Bacteroidales* genomes ensures the existence of multiple subpopulations of cells within the same species primed to withstand attacks by multiple phages, which could be targeting cells with distinct gene expression profiles. As phage infections typically progress very quickly, this variation of a bet-hedging strategy would provide immediate protection to a subpopulation of cells expressing the most resilient phenotypes, before other, potentially slower, anti-phage defense mechanisms such as CRISPR-Cas systems have an opportunity to engage. This mechanism would therefore ensure enhanced fitness for *Bacteroidales* in the unpredictable and competitive environment of the human intestinal tract.

Methods

Phage isolation and genome sequencing

Bacteriophage Bf12P1 was isolated from twice 0.22-μm filtered (PES membrane) primary effluent wastewater (King County Wastewater Treatment Division–West Point Treatment Plant, Seattle, WA). Phage was isolated using *Bacteroides fragilis* NCTC 9343 ATCC 25285 (hereafter referred to as *B. fragilis*). *B fragilis* was grown anaerobically at 37 °C in "Mega media" (see below) for 24 h. About 1 mL of fresh culture was added to 5 mL of filtered wastewater and 5 mL of fresh media. After incubating overnight at 37 °C, the culture was filtered using a 0.22-μm syringe filter to remove bacterial cells. The filtrate was plated on a lawn of *B. fragilis* soft agar and incubated overnight at 37 °C. The following day, potential phage isolates, which appear as zones of bacterial clearing in the soft agar (plaques), were picked and reinoculated into overnight cultures of *B. fragilis*. The following day, the cultures were re-filtered, and the filtrate was diluted and plated again onto soft agar lawns. From these plated dilutions, individual plaques, or phage isolates, were again re-amplified overnight in liquid culture. Phage isolates are stored at 4 °C in glass containers in the dark.

An aliquot of each potential phage isolate was used for DNA extraction and sequencing. Phage DNA was extracted using the Norgen Phage DNA Isolation kit. Phage DNA was sequenced by the Microbial Genome Sequencing Center (MiGS; Pittsburgh, PA) using Illumina NextSeq 2000. DNA sequences were quality trimmed to Q30 using the PHRED algorithm using BBDuk (BBMap v38.92[61]). Reads were de novo assembled into contigs using MEGAHIT[62] v 1.2.9, and genome coding

sequences (CDS) were predicted and annotated using Prokka v1.14.5[63] and Pharokka v1.7.3[64]. Functional annotation of CDS prediction was performed manually using NCBI BLASTp v2.9.0 on the NCBI protein sequence database[65], using phold v0.2.0[66], and PhageScope[67]. Additional annotation of unknown genes was performed by gene products modeling with AlphaFold 3[41], with a subsequent structural similarity search versus the PDB database[68] using Dali[69]. Closest relatives of Bf12P1 were determined using nucleotide BLAST search[65] on the nucleotide collection (nr/nt) standard database and iLund4u[22]. Intergenomic similarity of Bf12P1 to its closest relative was calculated using VIRIDIC Web[70]. Putative Bf12P1 promoters were predicted using four criteria: (a) a putative promoter should precede a block of co-expressed genes, as identified from the microSPLiT data; (b) a putative promoter should be located in an intergenic region; (c) a putative promoter should contain a putative transcription start site (TSS) mark visible as a sharp increase in a microSPLiT read density; (d) a putative promoter should have a well-defined −7 and −33 elements, containing the TTTG motifs, known for promoters in *Bacteroides* species[71].

Rich anaerobic bacterial growth media (mega media) preparation

A 100x dry microsalts mixture was prepared by mixing $NaHCO_3$ (4 g), NaCl (8 g), $MgSO_4{\cdot}7H_2O$ (2 g), $CaCl_2$ (0.08 g), and $FeSO_4$ (0.04 g). $FeSO_4$ pellets were crushed prior to weighing for uniform distribution. A 20x TYG super mix was prepared by combining BBL tryptone peptone (100 g), Bacto yeast extract (50 g), glucose (20 g), L-cysteine (5 g), KH_2PO_4 (41 g), K_2HPO_4 (122 g), and 1.012 g of the 100x dry microsalts. To prepare 500 mL of "Mega media", 17 g of TYG super mix was dissolved in deionized water with meat extract (2.5 g), cellobiose (0.5 g), maltose (0.5 g), fructose (0.5 g), sodium acetate (0.5 g), sodium sulfate (1 g), and malic acid (0.5 g). Neat tween 80 (100 μL) was then added, and the volume was adjusted to 500 mL. Media was sterilized by filtration or autoclaving, followed by the addition of vitamin K (500 μL of a 1 mg/mL stock), ATCC vitamin mix (500 μL), ATCC mineral mix (500 μL), and hematin. Prepared media was equilibrated in an anaerobic chamber for a minimum of 2 days prior to use.

Phage transmission electron microscopy

Transmission electron microscopy was performed through the Fred Hutch Cellular Imaging Shared Resource. Bf12P1 samples were fixed in ½ strength Karnovsky's fixative overnight at 4 °C. 200 mesh Formvar/Carbon coated glow discharged grids were floated over 25–50 μL bacterial suspension for 20 min. Grids were briefly rinsed with ½ strength Karnovsky's, 0.1 M cacodylate buffer and four changes of distilled water. Grids were negatively stained with 1% uranyl acetate, wicked with damp filter paper, and allowed to air dry before imaging. Samples were examined on a JEOL JEM 1400 transmission microscope (JEOL, Tokyo) at 120 kV. Images were acquired with a Gatan Rio 4k × 4k digital camera system.

Phage culture conditions and quantification

Phage Bf12P1 stocks were prepared by inoculation of a liquid culture of *B. fragilis* with 20 μL of high-titer phage stock, overnight incubation to allow infection and lysis, and double 0.22 μm filtration of lysate. Phage stocks were immediately quantified using the soft agar overlay technique[72] and stored in glass vials in the dark at 4 °C. Dilutions of phage stocks were performed in sterilized spent *B. fragilis* growth media. All phage quantification for experiments was performed using the soft agar overlay method. Briefly, 7 mL of "Mega media" soft agar was combined with 500 μL of an overnight *B. fragilis* culture, poured onto a hard agar plate, and allowed to solidify. Phage dilutions were spotted onto the plate (10 μL). Plates were incubated for 24–48 h, and phage titer was quantified and reported as plaque-forming units (PFU) per mL. Titers were routinely reconfirmed to ensure consistency between experiments.

Bacteria–phage growth curve (MOI infection dynamics)

For growth curve experiments used to characterize infection dynamics at varying MOIs, conditions were as follows: a 24-h culture of *B. fragilis* was used (OD_{600} of 0.7–0.8). Phage stocks with known concentrations were used. Filtered chopped meat carbohydrate (CMC) media (Anaerobe systems) was used for the following experiments: selection of phage-resistant isolates; cultures used for RNA sequencing; cultures used for single-cell RNA sequencing. CMC media was further prepared by serially filtering through 0.45 μm and then 0.22 μm syringe filters to remove all pieces of meat. For 96-well plate growth curves, 200 μL cultures were prepared. About 20 μL of *B. fragilis* was added to 170 μL of media and 10 μL of a known concentration of phage. A "Breathe-Easy" film was added to the top to prevent evaporation and allow gas exchange. A BioTek Epoch 2 microplate spectrophotometer was used to incubate at 37 °C and collect OD_{600} measurements for 24 h.

Bulk RNA-seq of *Bf12P1*-infected and uninfected cultures

200 uL *B. fragilis* overnight culture was mixed with 1.75 mL CMC Media (Anaerobe Systems) and 50 uL of Bf12P1 at MOI of 0.00004. For non-phage conditions, 50 μL of spent *B. fragilis* CMC media was used instead of the phage. Cultivation was performed anaerobically in a deep-well 2 mL plate at 37 °C with shaking. At the time of collection (t = 2.15 h), RNA-Protect (Qiagen) was added in equal volume to each culture in the anaerobic chamber. After addition and RNA-Protect, cultures were removed from the anaerobic chamber and the remaining steps were carried out in the biosafety cabinet. Cells were pelleted at 5000×g for 10 min, RNA was extracted using the RNeasy kit (Qiagen) and treated with DNase I sequencing library was prepared using TruSeq RNA Library Prep Kit (Illumina) and sequenced using Illumina NextSeq 500 at Microbial Sequencing and Analysis Center. RNA-Seq data were aligned to the *B. fragilis* reference genome using STAR (v2.7.10a) with --quantMode GeneCounts option, and calculated gene counts were used for further analysis. RNA-Seq experiments were performed in three biological replicates.

Selection of phage-resistant isolates

B. fragilis seed cultures were grown and prepared for the experiment as described. To prepare the experimental culture conditions: 20 μL of overnight *B. fragilis* culture was added to 170 μL twice-filtered CMC media and 10 μL of phage to achieve final MOIs of 5×10^{-6} ("low-MOI") and 5×10^{-3} ("high-MOI"). Additionally, two different bile acid growth conditions were generated by adding taurocholic acid (TCA), deoxycholic acid (DCA), or a solvent control (methanol) to cultures at a final concentration of 100 μM. The "Breathe-Easy" film was added to the top of the flat-bottom 96-well plate to prevent evaporation and allow gas exchange. A BioTek Epoch 2 Microplate Spectrophotometer was used to incubate at 37 °C and collect OD_{600} measurements for 24 h. After 24 h, the emergence of phage resistance was visually confirmed by the regrowth of cultures infected with Bf12P1 after the lysis stage. Aliquots of bulk culture were collected via centrifugation and stored at −20 °C for shotgun sequencing. To isolate individual phage-resistant isolates, an aliquot of each 24-h culture was streaked out on solid agar Mega media plates and incubated at 37 °C for 48 h. After 48 h, four individual colonies from each condition of interest were picked using an inoculation loop and inoculated into "Mega media" and incubated at 37 °C for 48 h. After 48 h, bacterial cultures were tested for resistance against Bf12P1 using a cross-streaking assay. Briefly, 20 μL of high-titer phage stock Bf12P1 was dispensed at the top of an agar plate in triplicate. The plate was tilted at a 45-degree angle, allowing the aliquot to run down the plate. Phage streaks were then allowed to dry. Plates were turned perpendicular, and 12 μL of bacterial culture was applied in the same manner; effectively, cross-streaking with phage. Plates were then incubated at 37 °C for 48 h. Isolates were determined to be resistant or sensitive by a visual examination of whether they grew across the phage streak. Resistant and sensitive isolates were further confirmed

to be resistant through liquid growth curve assays with Bf12P1 as previously described.

Preparation of samples for single-cell RNA sequencing

A turbid 24-h culture of *B. fragilis* was diluted in CMC media (filtered through 0.45 μm filter units) with Bf12P1 at an approximate MOI of 4×10^{-5} (total PFU/total CFU as estimated by starting bacterial culture OD_{600}) and incubated at 37 °C. For bile acid samples, a final concentration of 100 μM DCA or TCA was added to the media. The collection timepoint was based on previous growth curves and an expected time to lysis at this MOI. Sample OD_{600} measurements were monitored every 30 min to ensure lysis had not yet occurred at the time of collection. At 2.5–3 h and OD_{600} of 0.23, cells were collected on ice, centrifuged, and resuspended in 1.5 mL of 4% formaldehyde. For the M30 experiment, we added *B. subtilis* cells to a dedicated set of first barcode wells for validating the single-cell resolution. *B. subtilis* samples and reads were excluded from all further analysis.

Permeabilization, polyA tailing, and reverse transcription

Fixed cells were centrifuged at 10,000×*g* for 10 min at 4 °C, resuspended in cold 100 mM Tris-HCl+RI (100 mM Tris-HCl, pH = 8, 20 U/mL SUPERase-In RNase Inhibitor (Thermo Fisher Scientific)), and spun down again in the same conditions. Cells were resuspended in 250 μL permeabilization solution (0.04% Tween-20 in PBS), incubated for 3 min on ice, and after 1 mL addition of PBS + RI (1x PBS, 6 U/mL SUPERase-In RNase Inhibitor, 24 U/mL Enzymatics RNase Inhibitor (Qiagen)), spun down at 10,000×*g* for 10 min at 4 °C. Cells were aspirated and resuspended in 200 μL digestion mix (100 mM Tris-HCl, pH = 7, 50 mM EDTA, 7.5 μg of lysozyme (Thermo Fisher Scientific), 250 U/mL SUPERase-In RNase Inhibitor), incubated for 15 min at 37 °C, immediately added to 1 mL PBST (0.01% Tween-20 in PBS + RI), and centrifuged at 10,000×*g* for 10 min at 4 °C. Samples were then resuspended in 50 μL PBS + RI per sample post-permeabilization. Cells were counted by hemocytometer, and 10 million cells were taken per plate for polyadenylation. About 50 μL polyadenylate polymerase I solution (1x *E. coli* Poly(A) polymerase buffer, 800 U/mL SUPERase-In RNase Inhibitor, 50 mM ATP, 25U of *E. coli* Poly(A) polymerase) was added, and each sample was incubated at 37 °C for 30 min, spun down with 1 mL PBST, and resuspended in PBS + RI. Cells were counted again for verification of the number of cells per plate.

Reverse transcription (RT) mix was created on ice with final concentrations of 1X reverse transcription buffer, 0.25 U/μL Enzymatics RNase Inhibitor, 0.25 U/μL SUPERase-In RNase Inhibitor, 500 μM dNTPs, 10 U/μL Maxima H Minus Reverse Transcriptase, and 7.5% PEG8000 with a final volume of 1320 μL for a 96-well plate. Reverse transcription mix was added (12 μL) to wells, already containing 4 μL mixture of 12.5 μM random hexamer and 12.5 μM 15dT primer in each well; cells were vortexed and 4 μL suspension was added to wells for a final volume of 20 μL. Each sample was loaded into different wells of the reverse transcription plate to demultiplex them during analysis. Plates were incubated at 23 °C for 10 min and at 50 °C for 50 min at 0.42×*g* (500 rpm using the Eppendorf Thermomixer C). Then, 40 μL PBST was added to each well on ice, samples were pooled in a single tube with 24.96 μL 10% Tween-20, and the total reaction was centrifuged for 10 min at 10,000×*g* at 4 °C. The sample was aspirated and resuspended in 2.2 mL PBS + RI and disaggregated with both 10 μm and 1 μm pluriStrainer filters. The composition of the 96-well plate and barcoded oligonucleotides used for RT were as previously described[73].

Ligation barcoding

The ligation master mix was made on ice, containing final concentrations of 1X T4 Ligase Buffer 10X, 0.58 U/μL Enzymatics RNase Inhibitor, 0.05 U/μL SUPERase-In RNase Inhibitor, 7.5% PEG8000, and 8 U/μL T4 DNA Ligase, with a final volume of 2040 μL. About 2 mL cell suspension was added to the ligation mix and 40 μL aliquots were transferred into each well of a Round 2 ligation barcode plate before incubating for 30 min at 37 °C and adding 10 μL of round 2 blocking solution (26.4 μM BC_0340 (Supplementary Data 1), 2.5X Ligase Buffer). The plate was incubated again for 30 min at 37 °C, and cells were pooled, disaggregated, and barcoded again (Round 3 blocking solution: 11.5 μM BC_0066 (Supplementary Data 1), 125 mM EDTA). After the final filtering step, cells were washed (4000 μL 1X PBS, 40 μL 10% Tween-20, 10 μM SUPERase-In RNase Inhibitor) and aliquoted into libraries with PBS + RI to a final volume of 50 μL. Again, the compositions of the 96-well plates and oligonucleotides used for the barcoding step were as previously described[73].

Sequencing library preparation

Aliquoted libraries were mixed with 50 μL lysis buffer (20 mM Tris-HCl, pH = 8, 400 mM NaCl, 100 mM EDTA, pH = 8, 4.4% SDS) and 10 μL of 20 mg/mL Proteinase K for protein digestion prior to incubation at 55 °C for 2 h with 0.42×*g* (500 rpm using the Eppendorf Thermomixer C) shaking. The resulting biotinylated cDNA was captured on MyOne C1 Dynabeads. For template switching, a common 5′ sequence was added to full-length cDNA by resuspending the beads in 200 μL Template Switch Mix (1X Maxima RT Buffer, 7.5% PEG8000, 1 mM of each dNTP, 500 U/mL SUPERase-In RNase Inhibitor, 2.5 μM TSO primer (Supplementary Data 1), 10 U/μL Maxima H minus RT) and incubating at room temperature for 20 min at 0.60×*g* (600 rpm using the Eppendorf Thermomixer C) before further incubation at 42 °C for 90 min with agitation. Then, beads were washed in 250 μL Tris-T + RI (10 mM Tris-HCl, 0.1% Tween-20, 50 U/mL SUPERase-In RNase Inhibitor), decanted, and resuspended again with 100 μL per sample of terminal deoxynucleotidyl transferase tailing (TdT) master mix (1X TdT buffer, 0.25 mM $CoCl_2$, 1.2 mM dCTP, 0.06 mM ddCTP, 2 U/μL TdT enzyme (NEB), 0.1 U/μL RNase H (NEB)). Following incubation for 1 h at 37 °C and a wash with Tris-T + RI, cDNA was amplified via qPCR using primers BC_0108/M_0002 and BC_0062 (Supplementary Data 1), fragmented, and ligated with adapters prior to Illumina Sequencing[73].

Data preprocessing

Sequencing was performed on NovaSeq X Plus Series (PE150, Novogene), and a genome was generated by concatenating reference genomes for *Bacteroides fragilis* NCTC 9343 (GCF_000025985.1) and Bf12P1. With the generated genome, barcode lists for each of the three indexing rounds, and output sequencing files, STARsolo[74] (STAR-2.7.10b) was run for genome alignment, allowing read multimapping with the uniform count model. The raw count matrix was loaded into AnnData[75] (version 0.10.9); after assigning reads to barcode combinations and sample wells, log-log barcode rank plots were used to determine an initial filtering threshold of 400 UMIs/cell. Remaining ribosomal (r)RNA and transfer (t)RNA reads were filtered out, and the dataset was filtered again by 40 UMIs/cell.

Single-cell analysis

Two separate libraries were concatenated after filtering, log-normalized by 600 counts per cell, and scaled to unit variance and zero mean. Subsequently, ComBat[76] was run for batch effect correction. Normalized expression data were dimensionally reduced using principal component analysis (PCA). Shared neighbor graphs and uniform manifold approximation representations (UMAP[77]) were calculated with the first 12 principal components. All subsequent calculations were run in Python using Scanpy[78] (version 1.10.3) documentation for single-cell analysis.

Differential gene expression analysis

Scanpy gene ranking functions (sc.tl.rank_genes_groups and sc.get.rank_genes_groups_df) were used to analyze and retrieve statistical data between two groups of interest within the annotated data object. The output parameters included the names of all genes, *z*-score, log-

fold change, *p* values, and adjusted *p* values. To create a volcano plot from this data, either the |score| or −log(adjusted *p* value) was plotted on the y-axis against the log-fold change on the x-axis. Lists of defense genes from DefenseFinder[50] (Supplementary Data 6), host response genes upregulated after phage treatment, and CPS operon genes were used to assign colors to each point. Host response genes were taken as the top 15 distinct genes identified in the phage-treated sample, only clustering analysis relative to the untreated sample. Initial CPS operon annotations were mapped by and received from Dr. Laurie Comstock (University of Chicago).

Calculation of co-expression score and infection chance

To assess the co-expression between two genes A and B, first, probabilities of expression of individual genes were calculated as fractions of cells having above-zero normalized non-scaled expression values - p(A) and p(B), respectively. Then, the chance of simultaneous expression of the two genes, p(A&B), was calculated as a fraction of cells having above-zero normalized non-scaled expression values for both inspected genes. Co-expression score was calculated as a ratio of p(A&B) to the multiplication of p(A) and p(B). Values close to 1 indicate independent expression of genes; values above 1 indicate co-expression, and values below 1 indicate mutually exclusive expression of genes. To define boundaries of a CPS operon, mean co-expression values were calculated between genes adjacent to the CPS operon and core genes of the CPS operon. Co-expression between gene sets (e.g., CPS operons) was calculated similarly to genes, with the chance of expression calculated as a fraction of cells having above-zero normalized non-scaled expression values for any of the genes within a set. Infection chance was computed as N(infected)/(N(infected) + N(uninfected)).

Diffusion pseudotime analysis

The data was first subset into the phage-treated sample alone. A root cell was defined using a data.uns["iroot"] = np.flatnonzero(adata.obs["louvan"]==0)[1], which selected a random indexed cell from the untreated cluster within all phage-treated cells. Scanpy diffusion maps were created prior to running the existing diffusion pseudotime tool with 0 branchings and ten diffusion components. For downstream analysis using pseudotime values, five bins of equal size were created to group cells into pseudotime ranges (0.0–0.2, 0.2–0.4, 0.4–0.6, 0.6–0.8, and 0.8–1.0). Differential expression analysis was run to identify the top ten distinct *B. fragilis* genes within each pseudotime bin (Supplementary Data 4), and the Clusters of Orthologous Groups (COGs) database was used to define functional categories for these genes. Duplicate genes were removed. Raw mean expression for each of these genes, grouped by functional category, was calculated using the adata.raw.X matrix.

Verification of capsular operons promoter state via PCR

The procedure was adapted from the Troy et al. study[79] characterizing "ON" and "OFF" promoter states for the PSA locus, in addition to the recommended PCR protocol for use of KAPA Biosystems HiFi HotStart Ready Mix containing a type B DNA polymerase, optimized buffer, $MgCl_2$, and dNTPs. In the anaerobic chamber, *B. fragilis* glycerol stocks were streaked onto blood agar plates and incubated at 37 °C overnight. A single colony was transferred to 1 mL of BHI broth and grown overnight at 37 °C in the chamber; 100 μL of culture was then transferred to 10 mL of BHI broth into a sealed Balch-type tube and grown to an OD_{600} of 0.5–0.6 at ~1.4×*g* (225 rpm using the New Brunswick Innova 44 Shaker Incubator), 37 °C. After growth, DNA was extracted from each sample using the DNA extraction protocol provided by Qiagen and concentration was measured via Qubit v4.

For PCR, all primers and PCR conditions were used based on Troy et al. study[79] (Supplementary Data 1). The PCR mix contained 12.5 μL 2X Kapa HiFi HotStart Ready Mix (KAPA Biosystems), 0.75 μL of 10 μM forward primer, 0.75 μL of 10 μM reverse primer, 1.0 μL template DNA (~10 ng), and remaining molecular biology grade water to a final volume of 25 μL. For all CPS operon promoters, annealing temperature was 55–59 °C and set at 60 °C; for PSD, annealing was at 67 °C and set to 70 °C. The Monarch PCR cleanup kit was then used for purification, and 1.0 ng of each DNA product was visualized by Agilent TapeStation.

About 250 ng of purified PCR products were digested with 0.5 μL of the respective restriction enzyme (PSA: SspI-HF; PSB: RsaI; PSD: DraI; PSE: SspI-HF; PSF: EarI; PSG: PacI; PSH: SfcI)[79], 2.5 μL digest buffer, and molecular biology grade water to a final volume of 25 μL for 2 h at 37 °C. Post-digestion, products were mixed with gel loading dye and run on a 1.5% agarose gel at 100 V for 30 min with a 1 Kb standard DNA ladder. Bands were visualized and quantified on the gel to verify CPS operon promoter orientation for each of the 'phase-locked' *B. fragilis* mutants.

Verification of capsular operons expression via RT-qPCR

Overnight *B. fragilis* cultures with 4 mL basal metabolic broth were started for each of the samples, directly from original glycerol stocks, using wild-type *B. fragilis* as a control. Cultures were incubated anaerobically at approximately 1.4×*g* (225 rpm using the New Brunswick Innova 44 Shaker Incubator) and 37 °C overnight in Balch-type tubes to an approximate OD_{600} 0.6–1.0. Approximately 1.5–2.0 mL of culture was spun down at 10,000×*g* for 30 s, and supernatant was discarded. A general TRIzol protocol from Invitrogen was followed to extract RNA from each culture pellet. Briefly, 1 mL of TRIzol was added to ~10^6 cells for each sample, homogenized, and incubated for 5 min on ice. Next, 0.2 mL of chloroform was added, mixed via shaking, and incubated for an additional 3–4 min on ice. Samples were centrifuged for 15 min at 12,000×*g*, and the aqueous phase containing RNA was transferred to a new tube. To precipitate the RNA, 0.5 mL of isopropanol was added to the aqueous phase and incubated for 10 min on ice. Samples were centrifuged for 10 min at 12,000×*g*, then kept at −20 °C for 30 min and spun down again at 12,000×*g* for 20 min. Supernatant was discarded carefully, and the pellet was resuspended in 1 mL of 75% EtOH by vortexing. Then, samples were centrifuged for 5 min at 7500×*g*, and the pellet was air dried for 5–10 min. Obtained RNA was dissolved in 25 μL RNase-free water and incubated at 55 °C for 15 min before Qubit concentration measurements. All samples were treated with 2 μL of DNase I from (NEB) and re-purified using the NEB Monarch RNA cleanup kit with an elution volume of 30 μL.

For first-strand cDNA synthesis, 20 μL template RNA (2–4 μg) was mixed with 3 μL of 100 μM random hexamer primer (IDT), 3 μL of 10 mM dNTP mix, 12 μL 5X RT buffer, 1.5 μL SUPERase-In RNase inhibitor, 1 μL Maxima H Minus RT, and nuclease-free water to a final volume of 60 μL. Reaction mixture was incubated for 10 min at 25 °C, 30 min at 50 °C, and terminated for 5 min at 85 °C. For qPCR, two replicates were run for each cDNA sample along with two replicates of untreated RNA and corresponding non-template controls. RT-qPCR primers used for capsular operons expression detection are listed in Supplementary Data 1. The qPCR program was set to 98 °C for 3 min, followed by 33 cycles at 98 °C for 12 s, 65 °C for 15 s, 72 °C for 10 s, and a final extension at 72 °C for 5 min.

gDNA extraction, library preparation, and sequencing

Overnight cultures were pelleted via centrifugation, supernatant removed, and DNA extracted using the NEB Monarch Spin gDNA Extraction Kit per the manufacturer's instructions. Concentration and purity were measured using NanoDrop and Qubit, and DNA integrity was confirmed on an Agilent TapeStation. For library preparation, 100 ng gDNA was fragmented using the WGS Fragmentation Mix (Qiagen), and sequencing adapters (Supplementary Data 1) were ligated using the T4 ligase (NEB). Indexed PCR amplification was performed using KAPA polymerase, 10 μM P5 and uniquely barcoded P7

extension primers (Supplementary Data 1). Thermocycling conditions included initial denaturation at 95 °C for 3 min, followed by 8–12 cycles of 98 °C for 20 s, 55 °C for 20 s, and 72 °C for 3 min. Indexed libraries were purified using a double-sided size selection with SPRI beads at a 0.6–0.8x ratio. Final libraries were quantified by Qubit and TapeStation and pooled at equimolar ratios before submitting for Illumina sequencing. Sequencing was performed in a PE 150 + 150 mode using Illumina NovaSeq X Plus Series at Novogene or by the Microbial Genome Sequencing Center using their Illumina Whole Genome Sequencing service.

Variant calling

Variant calling was performed using snippy v 5.2[80] with default conditions using the *Bacteroides fragilis* NCTC 9343 GCF_000025985.1 ASM2598v1 as a reference genome. Medium-quality variants having SNP score >100 were extracted from the raw.vcf data.

Identification of CRISPR spacer acquisition from WGS data for *B. fragilis* isolates

WGS reads were de novo assembled using SPAdes v3.15.5[81] with default conditions. CRISPR cassettes were predicted in the obtained contigs using Minced v0.4.2[82,83].

Reporting summary

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

Data availability

MicroSPLiT sequencing data generated in this study have been deposited in the Gene Expression Omnibus series under accession code GSE309504. The bulk RNA sequencing data are available under GSE318809. WGS data for *Bacteroides fragilis* strains and Bf12P1 phage are available in the Sequence Read Archive under BioProject PRJNA1335357. Sequencing datasets generated in this study are summarized in Supplementary Data 9. The Bf12P1 genome is deposited in GenBank under accession number PX436277. Source data are provided with this paper.

Code availability

Scripts used for data analysis and visualization are available on GitHub: https://github.com/anika-gupta-8/B_fragilis_phage_microSPLiT[84].

References

1. Shkoporov, A. N. & Hill, C. Bacteriophages of the human gut: the "known unknown" of the microbiome. *Cell Host Microbe* **25**, 195–209 (2019).
2. Murtazalieva, K., Mu, A., Petrovskaya, A. & Finn, R. D. The growing repertoire of phage anti-defence systems. *Trends Microbiol.* **32**, 1212–1228 (2024).
3. Georjon, H. & Bernheim, A. The highly diverse antiphage defence systems of bacteria. *Nat. Rev. Microbiol.* **21**, 686–700 (2023).
4. Wu, Y. et al. Bacterial defense systems exhibit synergistic anti-phage activity. *Cell Host Microbe* **32**, 557–572.e6 (2024).
5. Vassallo, C. N., Doering, C. R., Littlehale, M. L., Teodoro, G. I. C. & Laub, M. T. A functional selection reveals previously undetected anti-phage defence systems in the *E. coli* pangenome. *Nat. Microbiol.* **7**, 1568–1579 (2022).
6. Gao, L. et al. Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science* **369**, 1077–1084 (2020).
7. Spratt, M. R. & Lane, K. Navigating environmental transitions: the role of phenotypic variation in bacterial responses. *mBio* **13**, e02212–e02222 (2022).
8. Kuchina, A. et al. Microbial single-cell RNA sequencing by split-pool barcoding. *Science* **371**, eaba5257 (2021).
9. Porter, N. T. et al. Phase-variable capsular polysaccharides and lipoproteins modify bacteriophage susceptibility in Bacteroides thetaiotaomicron. *Nat. Microbiol.* **5**, 1170–1181 (2020).
10. Shkoporov, A. N. et al. Long-term persistence of crAss-like phage crAss001 is associated with phase variation in Bacteroides intestinalis. *BMC Biol.* **19**, 163 (2021).
11. Carasso, S. et al. Inflammation and bacteriophages affect DNA inversion states and functionality of the gut microbiota. *Cell Host Microbe* **32**, 322–334.e9 (2024).
12. Krinos, C. M. et al. Extensive surface diversity of a commensal microorganism by multiple DNA inversions. *Nature* **414**, 555–558 (2001).
13. Nakayama-Imaohji, H. et al. DNA inversion regulates outer membrane vesicle production in *Bacteroides fragilis*. *PLoS ONE* **11**, e0148887 (2016).
14. Stoltzfus, M. J., Workman, R. E., Keith, N. C. & Modell, J. W. A dynamic subpopulation of CRISPR-Cas overexpressers allows *Streptococcus pyogenes* to rapidly respond to phage. *Nat. Microbiol.* **9**, 2410–2421 (2024).
15. Boleij, A. et al. The *Bacteroides fragilis* toxin gene is prevalent in the colon mucosa of colorectal cancer patients. *Clin. Infect. Dis.* **60**, 208–215 (2015).
16. Kordahi, M. C. et al. Genomic and functional characterization of a mucosal symbiont involved in early-stage colorectal cancer. *Cell Host Microbe* **29**, 1589–1598.e6 (2021).
17. Gibson, F. C. 3rd, Onderdonk, A. B. & Kasper, D. L. Cellular mechanism of intraabdominal abscess formation by *Bacteroides fragilis*. *J. Immunol.* **160**, 5000–5006 (1998).
18. Tzianabos, A. O., Kasper, D. L. & Onderdonk, A. B. Structure and function of *Bacteroides fragilis* capsular polysaccharides: relationship to induction and prevention of abscesses. *Clin. Infect. Dis.* **20**, S132–S140 (1995).
19. Wagner, V. E. et al. Effects of a gut pathobiont in a gnotobiotic mouse model of childhood undernutrition. *Sci. Transl. Med.* **8**, 366ra164 (2016).
20. Wick, E. C. et al. Stat3 activation in murine colitis induced by enterotoxigenic *Bacteroides fragilis*. *Inflamm. Bowel Dis.* **20**, 821–834 (2014).
21. Trapnell, C. et al. Pseudo-temporal ordering of individual cells reveals dynamics and regulators of cell fate decisions. *Nat. Biotechnol.* **32**, 381 (2014).
22. Egorov, A. A., Hauryliuk, V. & Atkinson, G. C. Systematic annotation of hyper-variability hotspots in phage genomes and plasmids. Preprint at *bioRxiv* https://doi.org/10.1101/2024.10.15.618418 (2024).
23. Ridlon, J. M., Harris, S. C., Bhowmik, S., Kang, D.-J. & Hylemon, P. B. Consequences of bile salt biotransformations by intestinal bacteria. *Gut Microbes* **7**, 22–39 (2016).
24. Gabig, M. et al. The cell surface protein Ag43 facilitates phage infection of *Escherichia coli* in the presence of bile salts and carbohydrates. *Microbiology* **148**, 1533–1542 (2002).
25. Scanlan, P. D., Bischofberger, A. M. & Hall, A. R. Modification of *Escherichia coli*–bacteriophage interactions by surfactants and antibiotics in vitro. *FEMS Microbiol. Ecol.* **93**, fiw211 (2017).
26. Scanlan, J. G., Hall, A. R. & Scanlan, P. D. Impact of bile salts on coevolutionary dynamics between the gut bacterium *Escherichia coli* and its lytic phage PP01. *Infect., Genet. Evol.* **73**, 425–432 (2019).
27. Netter, Z., Dunham, D. T. & Seed, K. D. Adaptation to bile and anaerobicity limits *Vibrio cholerae* phage adsorption. *mBio* https://doi.org/10.1128/mbio.01985-23 (2023).
28. Mojardín, L. & Salas, M. Global transcriptional analysis of virus-host interactions between phage ϕ29 and *Bacillus subtilis*. *J. Virol.* **90**, 9293–9304 (2016).

29. Li, X. et al. Temporal Transcriptional Responses of a *Vibrio alginolyticus* strain to *Podoviridae* phage HH109 revealed by RNA-seq. *mSystems* **7**, e00106–e00122 (2022).
30. Li, T. et al. Isolation and characterization of the novel phage JD032 and global transcriptomic response during JD032 infection of *Clostridioides difficile* ribotype 078. *mSystems* https://doi.org/10.1128/msystems.00017-20 (2020).
31. Howard-Varona, C. et al. Phage-specific metabolic reprogramming of virocells. *ISME J.* **14**, 881–895 (2020).
32. Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845–848 (2016).
33. Gerovac, M. et al. Phage proteins target and co-opt host ribosomes immediately upon infection. *Nat. Microbiol.* **9**, 787–800 (2024).
34. Torrents, E. Ribonucleotide reductases: essential enzymes for bacterial life. *Front. Cell. Infect. Microbiol.* **4**, 52 (2014).
35. Dwivedi, B., Xue, B., Lundin, D., Edwards, R. A. & Breitbart, M. A bioinformatic analysis of ribonucleotide reductase genes in phage genomes and metagenomes. *BMC Evol. Biol.* **13**, 33 (2013).
36. Weigel, C. & Seitz, H. Bacteriophage replication modules. *FEMS Microbiol. Rev.* **30**, 321–381 (2006).
37. Patrick, S. et al. A unique homologue of the eukaryotic protein-modifier ubiquitin present in the bacterium *Bacteroides fragilis*, a predominant resident of the human gastrointestinal tract. *Microbiology* **157**, 3071 (2011).
38. Fritze, J., Zhang, M., Luo, Q. & Lu, X. An overview of the bacterial SsrA system modulating intracellular protein levels and activities. *Appl. Microbiol. Biotechnol.* **104**, 5229–5241 (2020).
39. Vargas-Blanco, D. A. & Shell, S. S. Regulation of mRNA stability during bacterial stress responses. *Front. Microbiol.* **11**, 2111 (2020).
40. Ueno, H. & Yonesaki, T. Phage-induced change in the stability of mRNAs. *Virology* **329**, 134–141 (2004).
41. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
42. Hirschi, M. et al. AcrIF9 tethers non-sequence specific dsDNA to the CRISPR RNA-guided surveillance complex. *Nat. Commun.* **11**, 2730 (2020).
43. Aksyuk, A. A. & Rossmann, M. G. Bacteriophage assembly. *Viruses* **3**, 172 (2011).
44. Christie, G. E. & Calendar, R. Bacteriophage P2. *Bacteriophage* **6**, e1145782 (2016).
45. Weinacht, K. G. et al. Tyrosine site-specific recombinases mediate DNA inversions affecting the expression of outer surface proteins of *Bacteroides fragilis*. *Mol. Microbiol.* **53**, 1319–1330 (2004).
46. Coyne, M. J., Weinacht, K. G., Krinos, C. M. & Comstock, L. E. Mpi recombinase globally modulates the surface architecture of a human commensal bacterium. *Proc. Natl. Acad. Sci. USA* **100**, 10446–10451 (2003).
47. Liu, C. H., Lee, S. M., VanLare, J. M., Kasper, D. L. & Mazmanian, S. K. Regulation of surface architecture by symbiotic bacteria mediates host colonization. *Proc. Natl. Acad. Sci. USA* **105**, 3951–3956 (2008).
48. Saba, J. et al. Bacteroides expand the functional versatility of a conserved transcription factor and transcribed DNA to program capsule diversity. *Nat. Commun.* **15**, 10862 (2024).
49. Chatzidaki-Livanis, M., Weinacht, K. G. & Comstock, L. E. Trans locus inhibitors limit concomitant polysaccharide synthesis in the human gut symbiont *Bacteroides fragilis*. *Proc. Natl. Acad. Sci. USA* **107**, 11976–11980 (2010).
50. Tesson, F. et al. Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nat. Commun.* **13**, 2561 (2022).
51. Lacqua, A., Wanner, O., Colangelo, T., Martinotti, M. G. & Landini, P. Emergence of biofilm-forming subpopulations upon exposure of *Escherichia coli* to environmental bacteriophages. *Appl. Environ. Microbiol.* **72**, 956–959 (2006).
52. Anantharaman, V., Makarova, K. S., Burroughs, A. M., Koonin, E. V. & Aravind, L. Comprehensive analysis of the HEPN superfamily: identification of novel roles in intra-genomic conflicts, defense, pathogenesis and RNA processing. *Biol. Direct* **8**, 15 (2013).
53. Cheng, R. et al. Prokaryotic Gabija complex senses and executes nucleotide depletion and DNA cleavage for antiviral defense. *Cell Host Microbe* **31**, 1331–1344.e5 (2023).
54. Antine, S. P. et al. Structural basis of Gabija anti-phage defence and viral immune evasion. *Nature* **625**, 360–365 (2024).
55. Lan, F. et al. Single-cell analysis of multiple invertible promoters reveals differential inversion rates as a strong determinant of bacterial population heterogeneity. *Sci. Adv.* **9**, eadg5476 (2023).
56. Bucher, M. J. & Czyż, D. M. Phage against the machine: the SIE-ence of superinfection exclusion. *Viruses* **16**, 1348 (2024).
57. Bakuradze, N. et al. In vitro evaluation of the therapeutic potential of phage VA7 against enterotoxigenic *Bacteroides fragilis* infection. *Viruses* **13**, 2044 (2021).
58. Tariq, M. A. et al. Genome characterization of a novel wastewater *Bacteroides fragilis* bacteriophage (vB_BfrS_23) and its host GB124. *Front. Microbiol.* **11**, 583378 (2020).
59. Surana, N. K. & Kasper, D. L. The yin yang of bacterial polysaccharides: lessons learned from *B. fragilis* PSA. *Immunol. Rev.* **245**, 13–26 (2012).
60. Tzianabos, A. O., Onderdonk, A. B., Rosner, B., Cisneros, R. L. & Kasper, D. L. Structural features of polysaccharides that induce intra-abdominal abscesses. *Science* **262**, 416–419 (1993).
61. BBMap. SourceForge. https://sourceforge.net/projects/bbmap/ (2025).
62. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
63. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
64. Bouras, G. et al. Pharokka: a fast scalable bacteriophage annotation tool. *Bioinformatics* **39**, btac776 (2023).
65. Resource Coordinators, N. C. B. I. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **46**, D8–D13 (2018).
66. Bouras, G. et al. Protein structure-informed bacteriophage genome annotation with Phold. *Nucleic Acids Res.* **54**, gkaf1448 (2026).
67. Wang, R. H. et al. PhageScope: a well-annotated bacteriophage database with automatic analyses and visualizations. *Nucleic Acids Res.* **52**, D756–D761 (2024).
68. Berman, H. M. et al. The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
69. Holm, L. Dali server: structural unification of protein families. *Nucleic Acids Res.* **50**, W210–W215 (2022).
70. Moraru, C., Varsani, A. & Kropinski, A. M. VIRIDIC-a novel tool to calculate the intergenomic similarities of prokaryote-infecting viruses. *Viruses* **12**, 1268 (2020).
71. Bayley, D. P., Rocha, E. R. & Smith, C. J. Analysis of cepA and other *Bacteroides fragilis* genes reveals a unique promoter structure. *FEMS Microbiol. Lett.* **193**, 149–154 (2000).
72. Gratia, A. Des relations numériques entre bactéries lysogènes et particules de bactériophages. *Ann. Inst. Pasteur* **57**, 652–676 (1936).
73. Gaisser, K. D. et al. High-throughput single-cell transcriptomics of bacteria using combinatorial barcoding. *Nat. Protoc.* **19**, 3048–3084 (2024).
74. Kaminow, B., Yunusov, D. & Dobin, A. STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data. Preprint at *bioRxiv* https://doi.org/10.1101/2021.05.05.442755 (2021).
75. Virshup, I., Rybakov, S., Theis, F. J., Angerer, P. & Wolf, F. A. anndata: access and store annotated data matrices. *J. Open Source Softw.* **9**, 4371 (2024).

76. Behdenna, A. et al. pyComBat, a Python tool for batch effects correction in high-throughput molecular data using empirical Bayes methods. *BMC Bioinformatics* **24**, 459 (2023).
77. Healy, J. & McInnes, L. Uniform manifold approximation and projection. *Nat. Rev. Methods Prim.* **4**, 1–15 (2024).
78. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
79. Troy, E. B., Carey, V. J., Kasper, D. L. & Comstock, L. E. Orientations of the *Bacteroides fragilis* capsular polysaccharide biosynthesis locus promoters during symbiosis and infection. *J. Bacteriol.* **192**, 5832–5836 (2010).
80. Seemann, T. tseemann/snippy. (2024).
81. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
82. Skennerton, C. T. ctSkennerton/minced. (2024).
83. Bland, C. et al. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* **8**, 209 (2007).
84. Gupta, A. & Sutormin, D. anika-gupta-8/B_fragilis_phage_microSPLiT. *Zenodo* https://doi.org/10.5281/zenodo.18343042 (2026).

Acknowledgements

This research was supported by the following grants: Department of Energy Office of Science, Biological and Environmental Research (BER) Program, Grant DE-SC0023091 (A.K. and G.S.) National Institute of General Medical Sciences of the National Institutes of Health Grant R35GM150994 (A.K.) National Cancer Institute of the National Institutes of Health Grant U54 CA274374 (N.D.) Washington Research Foundation Postdoctoral Fellowship (N.M.) This research was supported by the Electron Microscopy Shared Resource RRID:SCR_022611 of the Cancer Consortium of Fred Hutch, University of Washington, and Seattle Children's Hospital (Cancer Center Support Grant P30 CA015704). We thank Steve MacFarlane for assistance with electron microscopy. We thank Jason Saba and Robert Landick (University of Wisconsin-Madison) for providing CPS 'phase-locked' strains.

Author contributions

A.G., N.M., D.S., G.S., N.D., and A.K. designed experiments; A.G., N.M., D.S., N.L., K.G., and A.R. performed experiments; A.G., N.M., D.S., and A.K. analyzed data; D.S. and Y.I. performed mathematical modeling; G.S., N.D., and A.K. obtained funding and provided supervisory support; A.G., N.M., D.S., G.S., N.D., and A.K. wrote the manuscript.

Competing interests

A.K. and G.S. are inventors on a patent application for microSPLiT filed by the University of Washington. G.S. is a cofounder and shareholder of Parse Biosciences, a single-cell RNA sequencing company. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-70381-8.

Correspondence and requests for materials should be addressed to Neelendu Dey or Anna Kuchina.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.