Phase-variable capsular polysaccharides and lipoproteins modify bacteriophage susceptibility in *Bacteroides thetaiotaomicron*

Nathan T. Porter[1],[*], **Andrew J. Hryckowian**[2],[*],[^], **Bryan D. Merrill**[2], **Jaime J. Fuentes**[1], **Jackson O. Gardner**[2], **Robert W. P. Glowacki**[1], **Shaleni Singh**[1], **Ryan D. Crawford**[3], **Evan S. Snitkin**[1], **Justin L. Sonnenburg**[2], **Eric C. Martens**[1],[^]

[1]Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI 48109, USA

[2]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA 94305

[3]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA

Abstract

A variety of cell surface structures dictate interactions between bacteria and their environment, including their viruses (bacteriophages). Members of the human gut Bacteroidetes characteristically produce several phase-variable capsular polysaccharides (CPS), but their contributions to bacteriophage interactions are unknown. To begin to understand how CPS impact *Bacteroides*-phage interactions, we isolated 71 *B. thetaiotaomicron*-infecting bacteriophages from two locations in the United States. Using *B. thetaiotaomicron* strains that express defined subsets of CPS, we show that CPS dictates host tropism for these phages and that expression of non-permissive CPS variants are selected under phage predation, enabling survival. In the absence of CPS, *B. thetaiotaomicron* escapes bacteriophage predation by altering expression of 8 distinct phase-variable lipoproteins. When constitutively expressed, one of these lipoproteins promotes resistance to multiple bacteriophages. Our results reveal important roles for *Bacteroides* CPS and

other cell surface structures that allow these bacteria to persist under bacteriophage predation and hold important implications for using bacteriophages therapeutically to target gut symbionts.

Introduction

The human gut microbiome is dominated by a diverse population of bacteria, with hundreds of different species typically coexisting within an individual[1,2]. Frequent diet changes, host immune responses and bacteriophage infections are among the many intermittent perturbations to the gut ecosystem. Despite these perturbations, an individual's microbiome generally remains stable over long time periods[3], suggesting that bacteria have evolved strategies to persist despite frequent disturbances. One mechanism that may promote ecosystem resilience is the ability of some bacteria to produce multiple capsular polysaccharides (CPS), cell surface components that have been diversified in the genomes of gut-dwelling Bacteroidetes and other phyla[4,5]. Previous work showed that CPS from Bacteroidetes and other phyla play roles in evading or modulating host immunity[6–10], but the diversity of CPS synthesis loci in gut bacteria suggests that they could fill other roles[5,8,11,12].

The phylum Bacteroidetes—within which members of the genus *Bacteroides* are typically the most abundant Gram-negative gut symbionts in industrialized human populations[2,13]— provides excellent models to study persistence and competition mechanisms, including CPS. The type strains of the well-studied species *Bacteroides thetaiotaomicron* and *Bacteroides fragilis* each encode 8 different CPS[14,15] and there is broad genetic diversity of *cps* loci among different strains within these species[8]. In *Bacteroides*, CPS structures appear to surround the entire cell[16,17] and the biosynthetic loci that encode these surface coatings are often under the control of phase variable promoters[8,15,18]. In conjunction with other regulators, phase variable CPS expression generates phenotypic heterogeneity within an otherwise isogenic population that may facilitate survival during ecological disturbances[8,15,19,20].

Bacterial viruses or bacteriophages (phages), like the bacteria on which they prey, vary greatly across individual gut microbiomes and are responsive to host dietary changes and disease states[21–25]. Compared to bacteria, far less is understood about the phages of the gut microbiome, especially the mechanisms governing phage-bacteria interactions. Here, we tested the hypothesis that CPS mediate *Bacteroides*-phage interactions. Our results support this hypothesis, but also reveal that *B. thetaiotaomicron* possesses additional phase-variable phage-evasion strategies in addition to CPS. Our results provide a glimpse into the intricacy of bacterial-phage interactions in the human gut and provide a foundation for future work to leverage these interactions to manipulate the gut microbiome.

Results

Bacteriophages infect *B. thetaiotaomicron* in a CPS-dependent fashion

The genomes of human gut Bacteroidetes frequently encode multiple CPS[5] (Extended Data 1). To test the hypothesis that *Bacteroides* CPS mediate interactions with phages, we isolated

71 phages that infect *B. thetaiotaomicron* VPI-5482 from two locations in the United States (Ann Arbor, Michigan and San Jose, California). Phages were isolated using the wild-type strain that encodes 8 genetically distinct CPS (Extended Data 1)[14] and a panel of engineered strains that includes 8 strains that each express a single CPS (designated "cps1" through "cps8")[8] and an acapsular strain (*cps*)[26] (see Methods and Supplemental Table 1). Plaque morphologies varied among these 71 isolates, ranging in size from <1 mm to >3 mm and in opacity from very turbid to clear (Extended Data 2).

To determine if infection by these phages was impacted by CPS, we tested each phage against each of the 10 host strains in a plaque assay to measure host range. Hierarchical clustering of the infection profiles revealed a cladogram with 3 main branches that each encompasses phages from both collection sites (Fig. 1, Extended Data 3). Phages in Branches 1 and 2 generally infect the acapsular strain and are blocked by some, but not all, CPS. ARB154 exclusively infected cps8, an uncommon CPS among *B. thetaiotaomicron* strains that appears to be contained in a mobile element[8]. Phages in Branch 3 generally infect wild-type, cps1, cps2, and cps3 well and this branch contained most of a subset of phages that fails to infect the acapsular strain, suggesting that they require CPS for infection.

Elimination of specific CPS alters bacterial susceptibility to phages

For the phages described above that robustly infect the acapsular strain, resistance in only a subset of single cps strains indicates that capsule-independent receptor(s) mediate infection and only a subset of "non-permissive" CPS block it. For phages that inefficiently infect the acapsular strain, one or more CPS may serve as a direct phage receptor. To further define these roles, we investigated a subset of 6 phages (ARB72, ARB78, ARB82, ARB101, ARB105, and ARB25; marked in blue text in Fig. 1). These phages infect wild-type *B. thetaiotaomicron* and 5 of them infect the acapsular strain poorly or not at all. We first tested the hypothesis that some CPS are required for infection by deleting only the subsets of CPS encoding permissive capsules based on our prior experiments. For ARB72, simultaneous elimination of the two most permissive capsules CPS1 and CPS3 from wild-type *B. thetaiotaomicron* reduced infection below the limit of detection (Fig. 2a). Likewise, elimination of the most permissive CPS for ARB78, ARB82, ARB101 and ARB105 significantly reduced infection by these phages. In some cases, this occurred even when other permissive CPS were still present (Fig. 2b–e).

For ARB25, which promiscuously infects *B. thetaiotaomicron* cps variants (Fig. 1), some single and compounded *cps* deletions reduced infection rates (Fig. 2f). While individual deletion of four permissive CPS (CPS1,6,7,8) led to partially reduced infection, so did single eliminations of either of two CPS initially determined to be non-permissive (CPS3 and CPS4). Moreover, combinatorial deletion of non-permissive CPS4 and permissive CPS1 completely eliminated detectable infection, suggesting more complicated regulatory interactions, which are known to occur with *Bacteroides* CPS[19,20]. Interestingly, strains lacking CPS4 or CPS1/CPS4 compensated by significantly increasing relative expression of the non-permissive *cps2* locus, which could contribute to ARB25 resistance (Fig. 2g).

A strain expressing only two non-permissive CPS (CPS2 and CPS3) could not be detectably infected by ARB25 (Fig. 2f, "2,3 only"). However, a strain expressing CPS2,3,4 regained

some susceptibility (Fig. 2f, "2,3,4 only"), indicating that when CPS4 is present it is capable of mediating some infection by ARB25, which is different than the observation made in Figure 1 and discussed further below. In contrast to sole expression of CPS2 and CPS3, deletion of just the *cps2* and *cps3* loci led to dominant expression of *cps1* and *cps4*, which increased infection efficiency and production of clearer plaques (Fig. 2f–h). ARB25 formed smaller and more turbid plaques on the *cps4* strain, demonstrating that loss of CPS4 expression alone modifies ARB25 infection (Fig. 2h). Additional experiments with another *B. thetaiotaomicron* strain that encodes homologs of *cps2, cps5* and cps6 support that elimination of these permissive capsules reduces, in some cases, bacteria-phage interaction (Extended Data 4). Moreover, the exogenous presence of a stoichiometric excess of purified, non-permissive CPS2 did not inhibit ARB25 infection of the acapsular strain, suggesting that CPS does not block infection in trans (Extended Data 5).

Liquid cultures of each of the CPS-expressing strains inoculated with live or heat-killed ARB25 largely replicated our initial plaque assays (Fig. 3). Initial growth perturbations were characterized by extended lag or drop in density, but liquid cultures of ARB25-susceptible strains eventually either re-grew to high (wild-type, acapsular, cps1, cps5) or intermediate (cps4, cps7, cps8) densities, suggesting outgrowth of resistant bacteria that was not attributable to inactivity of the phage present (Extended Data 5). In contrast to plate-based assays, liquid cultures of cps4 were sensitive to growth inhibition by ARB25, while cultures of cps6 were not. A similar correlation between plaque assays and liquid infections was observed with SJC01, a Branch 2 phage with an infection profile similar to ARB25 (Extended Data 5). Importantly, re-isolation of ARB25-free isolates and re-infection with the same phage resulted in most (69%) of the re-isolates exhibiting susceptibility, revealing that resistance is not predominantly caused by permanent mutations and is reversible (Supplemental Table 2).

Phage-resistant, wild-type *B. thetaiotaomicron* populations exhibit altered *cps* locus expression

We hypothesized that wild-type cells that are pre-adapted by expressing non-permissive capsules would be positively selected in the presence of phage. We infected wild-type *B. thetaiotaomicron* with ARB25 and monitored bacterial growth. Cultures treated with a high multiplicity of infection (MOI $\approx$1) displayed similar growth kinetics as observed previously, with an apparently resistant population emerging after 3–4 hours (Fig. 4a). Notably, cultures originating from different single colonies displayed variable growth kinetics in the presence of ARB25. Next, we measured if infection with ARB25 resulted in altered CPS expression by the surviving bacteria. *B. thetaiotaomicron* exposed to heat-killed phage predominantly expressed CPS3 and CPS4. Treatment with live ARB25 resulted in a significant decrease of *cps*1 and *cps*4 expression and a significant increase of expression of the non-permissive *cps*3 locus (Fig. 4b; Dirichlet regression $p < 0.01$). Similar growth and expression phenotypes occurred in cultures treated with a low ($\approx 10^{-4}$) MOI (Extended Data 6). Notably, the most resistant bacterial clones in both experiments (e.g. faster outgrowth post-infection; Fig. 4, Extended Data 6) expressed lower levels of permissive *cps*1 and *cps*4 and higher levels of non-permissive *cps*3 in heat-killed phage treatment groups (Extended Data 6). Therefore,

pre-existing variation in CPS expression may contribute to the ability of some clones to mount resistance during phage challenge.

Multiple layers of phase-variable features equip *B. thetaiotaomicron* to survive phage predation

The results described above show that variations in CPS can pre-adapt a sub-population to survive phage challenge. However, *B. thetaiotaomicron* Δ*cps* also grows after infection by ARB25 in liquid culture (Fig. 3) and most recovered Δ*cps* clones regained susceptibility to ARB25 (Supplemental Table 2), suggesting that reversible, CPS-independent resistance mechanisms exist. We therefore performed RNA-seq to measure differences in gene expression between ARB25 post-infected wild-type and acapsular *B. thetaiotaomicron*.

As expected, the transcriptomes of wild-type bacteria surviving ARB25 infection (n=3) were largely characterized by alterations in CPS expression (Fig. 5a, Supplemental Table 3). Sixty-three of 83 genes with significant expression changes ≥3-fold, belonged to 4 *cps* loci, with *cps1/cps4* decreased and *cps2/cps3* increased. Two additional gene clusters encoding different outer-membrane "Sus-like systems", which are responsible for import and degradation of carbohydrates and other nutrients[27,28], were also decreased after infection. Notably, these loci encode TonB-dependent transporters (similar to *E. coli* TonA, the first described phage receptor[29]), suggesting that the proteins encoded by these genes might be part of the receptor for ARB25.

In acapsular *B. thetaiotaomicron*, 118 genes showed significant expression changes after ARB25 challenge and most (85%) were upregulated (Fig. 5b, Supplemental Table 4). One of the two Sus-like systems (*BT2170–73*) that was decreased in ARB25-exposed wild-type was similarly decreased in acapsular *B. thetaiotaomicron*. Among the most highly upregulated genes after infection (28 genes with ≥10-fold increase and an adjusted p-value ≤0.01), 6 genes in the well-characterized starch-utilization system (Sus)[27] were increased, suggesting that surviving bacteria consume glycogen released from lysed siblings.

An additional 17 up-regulated genes belong to 8 loci that encode predicted outer membrane S-layer lipoproteins and OmpA β-barrel proteins. One of these (BT1927) was previously found to be phase-variable and increase *B. thetaiotaomicron* resistance to complement-mediated killing when locked in the "on" state[30]. The remaining S-layer clusters share both syntenic organization and weak homology to BT1927–25. The promoter regions of all 8 loci are also flanked by a pair of imperfect 17 nucleotide palindromic repeats (Fig. 5c). Three of these repeats are identical to those that mediate recombination of the *BT1927* promoter[30] and the remaining 4 only vary by a trinucleotide in the middle of each imperfect palindrome (Fig. 5c). Finally, amplicon sequencing of the promoter regions using directionally oriented primers confirmed recombination in 5 of the 7 loci, while two did not generate PCR products (Extended Data 7).

Among the genes that were differentially regulated in ARB25-infected acapsular *B. thetaiotaomicron*, we identified a recombinational shufflon similar to systems described in *B. fragilis*[18]. Two genes in this shufflon encode TonB-dependent transporters (BT1040, BT1046) with truncated 5' ends relative to BT1042 (Fig. 5d,e, Extended Data 8). We

detected by PCR and by amplicon sequencing the presence of all 5 predicted alternative recombination states (Fig. 5d,e, Extended Data 8) and mutation of the associated tyrosine recombinase (BT1041) locked the corresponding mutant into the native genomic architecture. Interestingly, deletion of BT1033–52 led to variable plaquing efficiency by ARB25, as compared to the acapsular parent strain (Extended Data 8), suggesting that loss of these genes may exert indirect global effects that mediate phage susceptibility while revealing it is not the only receptor.

We gained additional insights into the complexity of the transcriptional response of *B. thetaiotaomicron* to phage infection by performing RNA-seq on *B. thetaiotaomicron* cps1 challenged with ARB25. Interestingly, the cps1 strain that is forced to express a permissive capsule can also survive ARB25 infection (Fig. 3) and mainly does so by increased expression of the S-layer proteins identified above (Extended Data 9, Supplemental Table 5). This suggests that, at least for CPS1, co-expression of capsule and S-layer proteins is not mutually exclusive.

We also performed RNA-seq on *B. thetaiotaomicron* wild-type and acapsular strains challenged with phage SJC01. Wild-type bacteria infected with SJC01 exhibited similar alterations in CPS expression as seen with ARB25 (Extended Data 9, Supplemental Table 6). In wild-type bacteria infected with SJC01, 61 of 67 differentially expressed genes belonged to CPS1 and CPS4 (both downregulated) or CPS3 (upregulated), which is consistent with CPS3 being non-permissive for SJC01. As observed with ARB25, nutrient-utilizing Sus-like systems were down-regulated in SJC01-infected cells, including previously described systems for ribose[31] and fungal cell wall α-mannan utilization[32]. In SJC01-infected acapsular cultures, expression of 4 of the 8 S-layer proteins was prominent, with the *BT1927–25* locus being the most highly expressed. Additionally, 2 genes (*BT4014–13*) encoding predicted phase-variable restriction endonucleases were up-regulated between 6–16 fold (Extended Data 9, Supplemental Table 7).

S-layer expression promotes resistance to multiple phages and is a prominent evasion strategy *in vivo*

The gene encoding the canonical outer membrane S-layer protein (BT1927), and its downstream genes were among the most highly activated in acapsular *B. thetaiotaomicron* after ARB25 or SJC01 challenge (Fig. 5b, Extended Data 9). We therefore focused on the effects of these proteins on phage infection. We mutated the recombination site upstream of the phase-variable promoter as described previously[30] to create S-layer "off" and S-layer "on" variants of acapsular *B. thetaiotaomicron*. Growth of acapsular S-layer "off" cells was more effectively inhibited by ARB25 relative to acapsular S-layer "on" cells (Fig. 6a). The strength of this effect was altered by the age of the colonies used for subsequent liquid culture experiments to test phage infectivity (Extended Data 10), suggesting hysteretic effects on the expression or function of this S-layer. In addition to ARB25, constitutive expression of the BT1927 S-layer promotes resistance to 3 additional phage isolates (Fig. 6b–d). In combination with previous findings that BT1927 promotes complement resistance[30], these observations suggest that BT1927 and perhaps the 7 other *B. thetaiotaomicron* S-layers (Fig. 5c) promote resistance to a variety of disturbances.

Based on our results, phase variants in CPS, S-layers, nutrient receptors, and restriction endonucleases are all selected for during phage predation. These mechanisms may help to explain previous observations that *Bacteroides* can co-exist with phage *in vitro*[33–35]. We hypothesized that if the phase-variable systems that promote resistance also spontaneously revert some cells to a susceptible state, the population could generate enough susceptible bacteria to maintain a phage population. To test this in an *in vivo* model, we colonized germfree Swiss Webster mice separately with either wild-type or acapsular *B. thetaiotaomicron* for 7 days, then introduced ARB25 by oral gavage. Both bacterial populations reached high colonization levels, were not noticeably perturbed upon addition of ARB25, and for 72 days, bacteria and phage co-existed at high levels (Fig. 6e).

Because the Δcps strain cannot evade ARB25 through alterations in CPS expression, we hypothesized that it would accrue mutations that promote full resistance after several weeks of constant ARB25 pressure. However, after isolating phage-free *B. thetaiotaomicron* from intestinal contents of mice, we noted that all of the Δcps clones regained susceptibility to ARB25 while 5/13 wild-type isolates remained resistant (See Methods and Supplemental Table 8). These results suggest that longer-term (perhaps permanent) resistance can occur after prolonged exposure to ARB25 *in vivo*, but this requires the presence of CPS.

Lastly, we performed RNA-seq on bacteria recovered from the cecal contents of mice after 72 days of *in vivo* co-culture with ARB25. Compared to the corresponding *in vitro* transcriptomes, many additional genes were induced *in vivo*, as expected based on previous studies of *B. thetaiotaomicron* metabolic adaptation to the gut[36–38]. Surprisingly, wild-type *B. thetaiotaomicron* that co-existed with ARB25 for 72d *in vivo* exhibited lower expression of non-permissive CPS3 (average 5-fold lower than uninfected wild-type grown *in vitro*). While expression of non-permissive CPS2 was increased, so was expression of several permissive CPS (CPS5, CPS6, CPS7 and CPS8, Extended Data 9, Supplemental Table 9), which may have been influenced by growth *in vivo*, which we previously showed to select for expression of CPS4, CPS5 and CPS6[8]. Wild-type *B. thetaiotaomicron* also increased expression of 6 of the S-layer loci, while repressing one.

As expected, acapsular *B. thetaiotaomicron* altered expression of some of its S-layers after 72 days of ARB25 exposure *in vivo* (Extended Data 9, Supplemental Table 10). Notably, one of the S-layers (BT1826) showed dominant expression (2,738-fold increased expression), suggesting that BT1826 confers optimal resistance to ARB25 in this strain background and *in vivo* growth condition. Surprisingly, acapsular *B. thetaiotaomicron* displayed high expression of another set of 3 genes (*BT0292–94;* increased 79–156-fold) and one of these genes BT0294 encodes a predicted lipoprotein. Adjacent to this locus is a predicted recombinase and we identified a near-consensus promoter, in the "off" orientation upstream of BT0292 and flanked by 18 bp repeats. We demonstrated via PCR/sequencing that this promoter is capable of undergoing phase-variation (Extended Data 9) bringing the total number of *B. thetaiotaomicron* phase-variable loci that show selection in response to phages to 19. Finally, the phase-variable restriction endonuclease system identified *in vitro* in SJC01 infected cells was also upregulated in wild-type and acapsular *B. thetaiotaomicron* after prolonged co-existence with ARB25 *in vivo* (Extended Data 9).

Discussion

Phages are the most abundant biological entities in the gut microbiome[39] and interest in their roles and identities has increased as metagenomic sequencing continue to unveil details of their dynamics during health and disease[21,22,25]. However, metagenomics-based approaches do not generate information on the definitive hosts or the mechanisms of individual bacteria-phage interactions, limiting dissection of their ecological roles in the gut. By isolating phages for a host of interest, experiments can be done to complement metagenomics-based studies and synergistically build a foundation for understanding phage-bacteria interactions in the gut and other complex ecosystems.

There are several ways CPS could promote or prevent phage infection[40]. First, CPS may sterically mask surface receptors to block phage binding, although additional specificity determinants must be involved because no individual phage that infects the acapsular strain is blocked by all CPS. Additional specificity determinants could be driven by CPS structure (physical depth on the cell surface, polysaccharide charge, permeability) or be actively circumvented by the presence of polysaccharide depolymerases on the phage particles, as has been described elsewhere (e.g., *E. coli* K1 and phiK1–5[41]). Alternatively, some CPS could serve as obligate receptors[42] or increase the affinity of a phage for the bacterial cell surface. This latter type of adherence to CPS might increase the likelihood of phage-receptor contact, similar to previous observations of phages in host mucus[39].

Using ARB25 and SJC01 as representatives from our larger collection, we demonstrate that infection with these phages does not fully eradicate their target *B. thetaiotaomicron* populations *in vitro* and *in vivo*. Phage-bacteria co-existence was previously observed in other *Bacteroides*-phage systems[35]. Given the prevalence and diversity of CPS in gut-resident Bacteroidetes (Extended Data 1), our results suggest that CPS influence phage-host co-existence across this prominent phylum. Analysis of transiently phage-resistant sub-populations of the acapsular and cps1 strain that emerge after ARB25 or SJC01 infection revealed additional phase-variable surface proteins, at least one of which (*BT1927–26*) confers increased resistance to phage infection when constitutively expressed (Fig. 6a–d, Extended Data 10). Previously, it was shown that 1:1000 *B. thetaiotaomicron* cells in a phage-free environment express *BT1927*[30]. In contrast to non-permissive CPS, which can comprise up to 40% of the expressed CPS (e.g. CPS3 in Fig. 4b) spontaneous *BT1927* "on" cells may only emerge after longer periods of phage exposure; such as those we modeled *in vivo* (Extended Data 9). Taken together, our data on CPS, surface proteins, and restriction enzymes as they relate to phage infection in *B. thetaiotaomicron* collectively reveal that that there are at least 19 independent loci that equip *B. thetaiotaomicron* to survive phage predation.

This work points to the existence of complex relationships between bacteria and phage in the gut microbiome. Considering that these interactions likely differ by bacterial species/strain and evolve differently within individuals or regionally distinct global populations, the landscape becomes even more complex. Given the diverse adaptive and counter-adaptive strategies present in *B. thetaiotaomicron* and its relatives, our findings hold important implications for using phages to therapeutically alter the composition or function of the gut

microbiota. While a cocktail of multiple phages could theoretically be harnessed together to elicit more robust alteration of target populations within a microbiome, the complexity of host tropisms and bacterial countermeasures that exist for *B. thetaiotaomicron* imply that a deliberate selection of complementary phage would be needed. Given these considerations, our findings contribute an important early step towards building a deep functional understanding of the bacterium-virus interactions that occur in the human gut microbiome.

Methods

Bacterial strains and culture conditions

The bacterial strains used in this study are listed in Supplemental Table 11. Frozen stocks of these strains were maintained in 25% glycerol at −80˚C and were routinely cultured in an anaerobic chamber or in anaerobic jars (using GasPak EZ anaerobe container system sachets w/indicator, BD) at 37˚C in *Bacteroides* Phage Recovery Medium (BPRM), as described previously[43]: per 1 liter of broth, 10 g meat peptone, 10 g casein peptone, 2 g yeast extract, 5 g NaCl, 0.5 g L-cysteine monohydrate, 1.8 g glucose, and 0.12 g $MgSO_4$ heptahydrate were added; after autoclaving and cooling to approximately 55 °C, 10 mL of 0.22 μm-filtered hemin solution (0.1% w/v in 0.02% NaOH), 1 mL of 0.22 μm-filtered 0.05 g/mL $CaCl_2$ solution, and 25 mL of 0.22μm-filtered 1 M Na_2CO_3 solution were added. For BPRM agar plates, 15 g/L agar was added prior to autoclaving and hemin and Na_2CO_3 were added as above prior to pouring the plates. For BPRM top agar used in soft agar overlays, 3.5 g/L agar was added prior to autoclaving. Hemin, $CaCl_2$, and Na_2CO_3 were added to the top agar as above immediately before conducting experiments. Bacterial strains were routinely recovered from the freezer stocks directly onto agar plates of Brain Heart Infusion supplemented with 10% horse blood (Quad Five, Rygate, Montana) (BHI-blood agar; or for the SJC phages used in Figure 1, on BPRM agar), grown anaerobically for up to 3 days and a single colony was picked for each bacterial strain, inoculated into 5 mL BPRM, and grown anaerobically overnight to provide the starting culture for experiments (note that for the BT1927 S-layer protein experiment shown in Figure 6, 3 days of growth on BPRM medium was determined to promote the greatest phage resistance).

For the experiment described in Figure 2g, liquid cultures of *B. thetaiotaomicron* were grown in BPRM using the pyrogallol method as described previously. Briefly, a sterile cotton ball was burned and then pushed midway into the tube, after which 200 μL of saturated $NaHCO_3$ and 200 μL of 35% pyrogallol solution were added to the cotton ball. A rubber stopper was used to seal the tubes, and tubes were incubated at 37 °C.

Bacteriophage isolation from primary wastewater effluent

The bacteriophages described in this study were isolated from primary wastewater effluent from two locations at the Ann Arbor, Michigan Wastewater Treatment Plant and from the San Jose-Santa Clara Regional Wastewater Treatment Facility. After collection, the primary effluent was centrifuged at 5,500 rcf for 10 minutes at room temperature to remove any remaining solids. The supernatant was then sequentially filtered through 0.45 μm and 0.22 μm polyvinylidine fluoride (PVDF) filters to yield "processed primary effluent." Initial screening for plaques was done using a soft agar overlay method[44] where processed primary

effluent was combined with 1 part overnight culture to 9 parts BPRM top agar and poured onto a BPRM agar plate (e.g. 0.5 mL overnight culture and 4.5 mL BPRM top agar was used for standard circular petri dishes [100 mm × 15 mm]). Soft agar overlays were incubated anaerobically at 37 ˚C overnight. Phages were successfully isolated using three permutations of this assay (see Supplemental Table 1): (1) Direct plating, where processed primary effluent was directly added to overnight culture prior to plating. (2) Enrichment, where 10 mL processed primary effluent was mixed with 10 mL 2XBPRM and 3 mL exponential phase *B. thetaiotaomicron* culture and grown overnight. The culture was centrifuged at 5500 rcf for 10 minutes and filtered through a 0.22 μm PVDF filter. (3) Size exclusion, where processed primary effluent was concentrated up to 500-fold via 30 or 100 kDa PVDF or polyethersulfone size exclusion columns. Up to 1 mL of processed primary effluent, enrichment, or concentrated processed primary effluent was added to the culture prior to adding BPRM top agar, as described above. To promote a diverse collection of phages, no more than 5 plaques from the same plate were plaque purified and a diversity of plaque morphologies were selected as applicable. When using individual enrichment cultures, only a single plaque was purified.

Single, isolated plaques were picked into 100 μL phage buffer (prepared as an autoclaved solution of 5 ml of 1 M Tris pH 7.5, 5 ml of 1 M MgSO4, 2 g NaCl in 500 ml with ddH$_2$O). Phages were successfully plaque purified using one of two methods: (1) a standard full plate method, where the diluted phage samples were combined with *B. thetaiotaomicron* overnight culture and top agar and plated via soft agar overlay as described above or (2) a higher throughput 96-well plate-based method, where serial dilutions were prepared in 96-well plates and 1 μL of each dilution was spotted onto a solidified top agar overlay. This procedure was repeated at least 3 times to purify each phage.

High titer phage stocks were generated by flooding a soft agar overlay plate that yielded a "lacey" pattern of bacterial growth (near confluent lysis). Following overnight incubation of each plate, 5 mL of sterile phage buffer was added to the plate to resuspend the phage. After at least 2 hours of incubation at room temperature, the lysate was spun at 5,500 rcf for 10 minutes to clear debris and then filter sterilized through a 0.22 μm PVDF filter.

Phylogenetic analysis of human gut Bacteroidetes and enumeration of cps biosynthetic gene clusters

Phylogenetic analysis was performed by creating a core gene alignment using a custom, publicly available software package, cognac, written for R (version 3.6.1) with C++ integration via Rcpp (version 1.0.3)[45]. Briefly, genbank files for the 53 isolates were parsed to extract the amino acid sequences and orthologous genes were identified with cd-hit (version 4.7) requiring at least 70% amino acid identity and ensuring that genes were of similar length[46]. The cd-hit output was parsed and core genes were identified as those present in a single copy in all genomes. Amino acid sequences were concatenated and aligned with MAFFT (v7.310)[46,47]. The concatenated gene alignment was then used as the input for fastTree (version 2.1.10) to generate an approximate maximum likelihood phylogeny[48]. The tree created from the core genome alignment was then midpoint rooted

and visualized using phytools (version 0.6.99) ape (version 5.3) R packages respectively[49,50].

To identify *cps* loci within each of the 53 genomes, previously annotated *cps* genes[8] from the type strains of *B. thetaiotaomicron* VPI-5482 (Extended Data 1), *B. fragilis* NCTC9343, and *B. vulgatus* ATCC 8482 *cps* loci were used to identify pfam models that correspond to the glycosyl transferases (GTs) they contain, which revealed pfam00534, pfam00535, pfam01755, pfam02397, pfam02485, pfam08759, pfam13439, pfam13477, pfam13579, pfam13692, and pfam14305. In addition, we searched for *upxY* and *upxZ* (pfam02357, pfam13614), and protein tyrosine kinase (PTK; pfam02706.) Genes corresponding to these pfam modules were extracted for the 53 genomes. In addition, because we found that a number of apparent *upxY/Z* homologs, which were in species more divergent than the *Bacteroides* noted above, we performed an additional search for homologous genes using the UpxY/Z amino acid sequences from the 3 species listed above. For this, we searched the Integrated Microbial Genomes (IMG) database IMG genome BLASTp tool and an E-value cutoff of 1e-5. Each *cps* locus was confirmed by visually comparing homologous genes within each gene locus neighborhood. Positive hits for the presence of a *cps* locus were required to contain at least one GT, along with at least one Upx or PTK homolog in the adjacent locus.

Quantitative host range assays

To accommodate the large number of phage isolates in our collection, we employed a spot titer assay for semi-quantitative comparisons of infectivity on each bacterial strain. High titer phage stocks were prepared on their "preferred host strain," which was the strain that yielded the highest titer of phages in a pre-screen of phage host range (see Figure 1, Supplemental Table 1). Lysates were then diluted to approximately 10^6 PFU/mL, were added to the wells of a 96-well plate, and further diluted to 10^5, 10^4, and 10^3 PFU/mL using a multichannel pipettor. One microliter of each of these dilutions was plated onto solidified top agar overlays containing single bacterial strains indicated in each figure. After spots dried, plates were incubated anaerobically for 15–24 hours prior to counting plaques. Phage titers were normalized to the preferred host strain. The heatmaps and dendrogram were generated using the "heatmap" function in the "stats" package of R (version 3.4.0) which employs unsupervised hierarchical clustering (complete linkage method) to group similar phage infection profiles, with branch length in the dendrogram at the left of Figure 1 indicating degree of similarity between infection profiles.

Images of phage plaques

To document the morphologies of plaques formed by the purified phages, two sets of plaque pictures were captured: the first set were taken with a Color QCount Model 530 (Advanced Instruments) with a 0.01 second exposure. Images were cropped to 7.5 mm^2 but were otherwise unaltered. The second set of images were taken on a ChemiDoc Touch instrument (BioRad) with a 0.5 second exposure. Images were cropped to 7.5 mm^2 and unnaturally high background pixels were removed (Image Lab, BioRad) to facilitate viewing of the plaques. Both sets of images are shown in Extended Data 2. Plaque images in Extended Data 4 were taken on a ChemiDoc Touch instrument (BioRad).

Incubation of ARB25 phage with extracted CPS

Approximately 50–100 PFU of ARB25 in 50 μL phage buffer were mixed with an equal volume of H_2O or capsule (2 mg/mL) extracted by the hot water-phenol method (as described in reference[8]) and incubated at 37 °C for 30 minutes. Samples were then plated on the acapsular strain, and plaques were counted after 15–24 hours anaerobic incubation at 37 °C. Counts from two replicates on the same day were then averaged, and the experiment was performed three times. While the size of individual CPS2 polymers is unknown, an estimate of 1,000 hexose sugars per molecule (180,000 Da) would be 9×10^{13} CPS glycans at 1mg/mL. If the CPS were only 10% pure, incubation with 10^3 ARB25 PFU/mL was estimated to provide at least 10^9-fold more CPS glycans than PFU.

Bacterial growth curves with phages

For growth curve experiments, 3 or more individual colonies of each indicated strain were picked from agar plates and grown overnight in BPRM. Then, for experiments in Figures 3, 6, Extended Data 5 and Extended Data 10 each clone was diluted 1:100 in fresh BPRM and 100 μL was added to a microtiter plate. 10 μL of approximately $5*10^6$ PFU/mL live or heat-killed phage were added to each well, plates were covered with an optically clear gas-permeable membrane (Diversified Biotech, Boston, MA) and optical density at 600 nm (OD_{600}) values were measured using an automated plate reading device (BioTek Instruments). Phages were heat killed by heating to 95 °C for 30 minutes, and heat-killed phage had no detectable PFU/mL with a limit of detection of 100 PFU/mL.

In Extended Data 5, wild-type *B. thetaiotaomicron* was infected with live or heat-killed ARB25, and bacterial growth was monitored via optical density at 600 nm (OD_{600}) on an automated plate reader for 12 hours. At 0, 6.02, 8.36, and 11.7 hours post inoculation, replicate cultures were vortexed in 1:5 volume chloroform, centrifuged at 5,500 rcf at 4 °C for 10 minutes, and the aqueous phase was titered on the acapsular strain. No phages were detected in heat-killed controls.

Generation of phage-free bacterial isolates and determination of their phage susceptibility

To isolate phage-free bacterial clones from ARB25-infected cultures (Supplemental Tables 2 and 8), each culture was plated on a BHI-blood agar plate using the single colony streaking method. Eighteen individual colonies were picked from each plate, and each of these clones was re-isolated on a separate BHI-blood agar plate. One colony was picked from each of these secondary plates and was inoculated into 150 μL BPRM broth and incubated anaerobically at 37 °C for 2 days. Only one of the clones (a *cps*4 isolate) failed to grow in liquid media. To determine whether cultures still contained viable phage, 50 μL of each culture was vortexed with 20 μL chloroform, then centrifuged at 5,500 rcf for 10 minutes. 10 μL of the lysate was spotted on BPRM top agar containing naïve acapsular bacteria and was incubated anaerobically overnight at 37 °C. Loss of detectable phage in the twice passaged clones was confirmed for most of the clones (79/89, 89%) by the absence of plaques on the naïve acapsular strain.

To determine whether the resulting phage-free isolates were resistant to ARB25 infection, each culture was diluted 1:100 in fresh BPRM, 100 μL was added to a microtiter plate, and

10 μL of either live or heat-killed ARB25 (approximately $5*10^6$ PFU/mL) was added. Plates were incubated anaerobically at 37 °C for 48 hours, and OD_{600} was measured as described above. Cultures were determined to be susceptible to ARB25 by demonstration of delayed growth or drop in OD_{600} compared to heat-killed controls.

Measurement of cps gene expression

For Figures 2g, 4b, and Extended Data 6, overnight cultures were diluted into fresh BPRM to an OD_{600} of 0.01. For Figure 4b, 200 μL of approximately 2×10^8 PFU/mL live phage or heat killed phage were added to 5 mL of the diluted cultures. For Extended Data 6, 200 μL of approximately 2×10^5 PFU/mL live phage or heat killed phage were added to 5 mL of the diluted cultures. Bacterial growth was monitored by measuring OD_{600} every 15–30 minutes using a GENESYS 20 spectrophotometer (Thermo Scientific). Cultures were briefly mixed by hand before each measurement. For determination of relative *cps* gene expression, cultures were grown to OD_{600} 0.6–0.8, centrifuged at 7,700 rcf for 2.5 minutes, the supernatant was decanted, and the pellet was immediately resuspended in 1 mL RNA-Protect (Qiagen). RNA-stabilized cell pellets were stored at −80 °C.

Total RNA was isolated using the RNeasy Mini Kit (Qiagen) then treated with the TURBO DNA-free Kit (Ambion) followed by an additional isolation using the RNeasy Mini Kit. cDNA was synthesized using SuperScript III Reverse Transcriptase (Invitrogen) according to the manufacturer's instructions using random oligonucleotide primers (Invitrogen). qPCR analyses for *cps* locus expression were performed on a Mastercycler ep realplex instrument (Eppendorf). Expression of each of the 8 *cps* synthesis loci was quantified using primers to a single gene in each locus (primers are listed in Supplemental Table 12) and normalized to a standard curve of DNA from wild-type *B. thetaiotaomicron*. The primers used were selected to target a gene specific to each *cps* locus and were previously validated against the other strains that lack the target *cps* locus for specificity[8]. Relative abundance of each *cps*-sepcific transcript was then calculated for each locus. A custom-made SYBR-based master mix was used for qPCR: 20 μL reactions were made with ThermoPol buffer (New England Biolabs), and contained 2.5 mM $MgSO_4$, 0.125 mM dNTPs, 0.25 μM each primer, 0.1 μL of a 100 X stock of SYBR Green I (Lonza), and 500 U Hot Start *Taq* DNA Polymerase (New England Biolabs). 10 ng of cDNA was used for each sample, and samples were run in duplicate. A touchdown protocol with the following cycling conditions was used for all assays: 95 °C for 3 minutes, followed by 40 cycles of 3 seconds at 95 °C, 20 seconds of annealing at a variable temperature, and 20 seconds at 68 °C. The annealing temperature for the first cycle was 58 °C, then dropped one degree each cycle for the subsequent 5 cycles. The annealing temperature for the last 34 cycles was 52 °C. These cycling conditions were followed by a melting curve analysis to determine amplicon purity.

Transcriptomic analysis of *B. thetaiotaomicron* after phage infection

Whole genome transcriptional profiling of wild-type and acapsular *B.thetaiotaomicron* infected with live or heat-killed ARB25 or SJC01, or from *in vivo* samples, was conducted using total bacterial RNA that was extracted the same as described above (Qiagen RNAEasy, Turbo DNA-free kit) and then treated with Ribo-Zero rRNA Removal Kit (Illumina Inc.) and concentrated using RNA Clean and Concentrator −5 kit (Zymo Research Corp, Irvine, CA).

Sequencing libraries were prepared using TruSeq barcoding adaptors (Illumina Inc.), and 24 samples were multiplexed and sequenced with 50 base pair single end reads in one lane of an Illumina HiSeq instrument at the University of Michigan Sequencing Core. Demultiplexed samples were analyzed using SeqMan NGen and Arraystar software (DNASTAR, Inc.) using EdgeR normalization and >98% sequence identity for read-mapping. Changes in gene expression in response to live ARB25 infection were determined by comparison to the heat-killed reference: retained were genes with ≥3-fold expression changes up or down and EdgeR adjusted P value ≤0.01. All RNA-seq data have been deposited in the publicly available NIH gene expression omnibus (GEO) database as project number GSE147071.

PCR and sequencing of phase variable *B. thetaiotaomicron* chromosomal loci

We found that each of the 8 chromosomal loci shown in Figure 5c had nearly identical 301 bp promoter sequences, including both of the imperfect palindromes that we predict to mediate recombination and the intervening sequence at each locus. While the 8 S-layer genes and the 7/8 of the upstream regions encoding putative tyrosine recombinases (all but the BT1927 region) shared significant nucleotide identity and gene orientation, we were able to design primers that were specific to regions upstream and downstream of each invertible promoter and used these to generate an amplicon for each locus that spanned the predicted recombination sites. After gel extracting a PCR product of the expected size for each locus, which should contain promoter orientations in both the "on" and "off" orientations, we performed a second PCR using a universal primer that lies within the 301 bp sequence of each phase-variable promoter and extended to unique primers that anneal within each S layer protein encoding gene. Bands of the expected size were excised from agarose gels, purified and sequenced using the primer that anneals within each S layer encoding gene to determine if the predicted recombined "on" promoter orientation is detected. (Note that the assembled *B. thetaiotaomicron* genome architecture places all of these promoters in the proposed "off" orientation. We were able to detect 6/8 of these loci in the "on" orientation in ARB25-treated cells by this method, Extended Data 7). Similar approaches were used to determine the re-orientation of DNA fragments in the *B. thetaiotaomicron* PUL shufflon shown in Figure 5d and restriction enzyme and additional lipoprotein system shown in Extended Data 9. For shufflon gene orientation, we used PCR primer amplicons positioned according to the schematic in Figure 5d followed by sequencing with the primer on the "downstream" end of each amplicon according to its position relative to the shuffled promoter sequence. For a list of primers used see Supplemental Table 12.

Construction of acapsular *B. thetaiotaomicron* S-layer 'ON' and S-layer 'OFF' mutants

Acapsular *B. thetaiotaomicron* S-layer 'ON' and 'OFF' mutants (Δ*cps* BT1927-ON and Δ*cps* BT1927-OFF, respectively) were created using the Δ*tdk* allelic exchange method[51]. To generate homologous regions for allelic exchange, the primers BT_1927_XbaI-DR and BT_1927_SalI-UF were used to amplify the BT1927-ON and BT1927-OFF promoters from the previously-constructed BT1927-ON and BT1927-OFF strains[30] via colony PCR using Q5 High Fidelity DNA polymerase (New England Biolabs). Candidate Δ*cps* BT1927-ON and Δ*cps* BT1927-OFF mutants were screened and confirmed by PCR using the primer pair BT1927_Diagnostic_R and BT1927_Diagnostic_F and confirmed by Sanger sequencing

using these diagnostic primers. All plasmids and primers are listed in Supplemental Tables S11 and S12, respectively.

Construction of *B. thetaiotaomicron* mutants lacking one or more cps loci

All publically available bacterial genomes in NCBI GenBank were queried via MultiGeneBlast[52] to identify fully sequenced bacteria with *B. thetaiotaomicron* VPI-5482-like cps loci. *B. thetaiotaomicron* 7330 was identified as having sytenic copies of VPI-5482-like *cps2*, *cps5*, and *cps6* loci with corresponding genes in the same order and 98% amino acid identity. Mutants of strain 7330 lacking one or more *cps* loci constructed for this study (Supplemental Table 11) were created using the *tdk* allelic exchange method[51]. The *B. thetaiotaomicron* 7330 *tdk-* strain was generated by UV mutagenesis by exposing a liquid culture of 7330 to 320 nm ultraviolet light from a VWR-20E transilluminator (VWR) for 60 seconds and plating onto BHIS-Blood agar supplemented with 200 micrograms/mL of 5-fluoro-2'-deoxyuridine (FUdR). All plasmids and primers used to construct these strains are listed in Supplemental Tables 11 and 12, respectively.

Germfree mouse experiments

All experiments were approved by the University of Michigan Institutional Animal Care and Use Committee and animals were monitored by a veterinarian. A group of 6, 6–8 week old germfree Swiss Webster mice were randomly assigned to two different groups containing 2 females and 1 male each, then gavaged with either wild-type or acapsular *B. thetaiotaomicron* for 7 days of mono-colonization. After 7 days, mice were gavaged with 1M sodium bicarbonate followed immediately by 2×10^8 PFU of ARB25 as previously described[24]. Feces were monitored for both colony forming units (CFU) or plaque forming units (PFU) every 7 days by plating fecal homogenates in SM buffer or fecal homogenate supernatant and serial dilutions in SM buffer on BPRM top agar plates. Sample size was selected based on using the fewest animals required to observe significant effects on colonization and gene expression in previous, similar experiments . The experimenters were not blinded to the different treatment (*B. thetaiotaomicron* strain) that each group was receiving.

Data representation and statistical analysis

The heatmaps for Figures 1 and Extended Data 3 and the dendrogram for Figure 1 were generated using the "heatmap" function in the "stats" package of R (version 3.4.0) which employs unsupervised hierarchical clustering (complete linkage method) to group similar phage infection. Other graphs were created in Prism software (GraphPad Software, Inc., La Jolla, CA). Statistical significance in this work is denoted as follows unless otherwise indicated: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$. All precise P values are either provided on each figure in or in Source Data. Statistical analyses other than Dirichlet regression were performed in Graphpad Prism. Dirichlet regression was performed in R using the package "DirichletReg" (version 0.6–3), employing the alternative parameterization as used previously[8,53]. Briefly, the parameters in this distribution are the proportions of relative *cps* gene expression and the total *cps* expression, with *cps7* expression used as a reference since we previously determined this *cps* to be poorly activated and not subject to phase-variable expression[8]. The variable of interest used in Figure 2g is

bacterial strain, whereas the variable of interest used in Figure 4b is phage viability (live versus heat-killed phage). Precision was allowed to vary by group given this model was superior to a model with constant precision, as determined by a likelihood ratio test at significance level $p < 0.05$. P-values for each condition were determined by z-test.

Extended Data

Extended Data Fig. 1. Diversification and structure of cps gene clusters in human gut Bacteroidetes.

Diversification and structure of *cps* gene clusters in human gut Bacteroidetes. (a) The genomes of 53 different human gut Bacteroidetes (predominantly named type strains) were searched for gene clusters that contain two or more different protein families indicative of

cps loci (see Methods). The number of cps loci detected in each genome is shown in the context of phylogenetic tree derived from the core genome of the 53 species used for this analysis; species for which *cps* loci were not detected using our search criteria are marked with a red "X". Due to gaps in several genomes, which often occur at *cps* loci, the numbers shown are likely to be an underestimate. (b) Schematics of the 8 annotated cps loci in *B.thetaiotaomicron* VPI-5482, which are singly present in the cps1-cps8 strains used in this study, or completely eliminated in the acapsular strain. Genes are color coded according to the key at the bottom and additional Pfam family designations are provided under most genes. The four main protein families used for informatics analysis are marked with asterisks and highlighted in bold in the key.
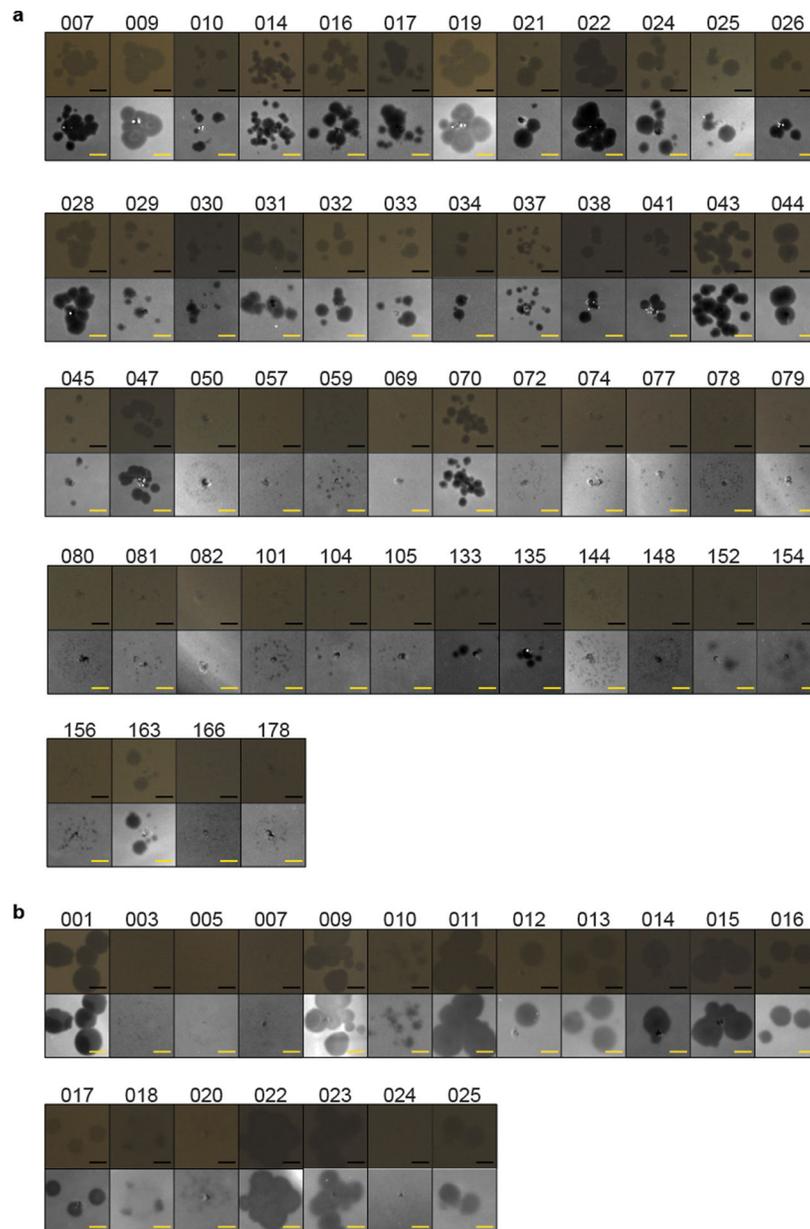
Extended Data Fig. 2. Representative pictures of phage plaques for all phages in this study: (a) phages from Ann Arbor (ARB); (b) phages from San Jose (SJC).
Representative pictures of phage plaques for all phages in this study: (a) phages from Ann Arbor (ARB); (b) phages from San Jose (SJC). The top row of images for each phage are unaltered; background and saturated pixels were removed from images in the bottom row to facilitate viewing of the plaques. Experiment was performed only once for photographing plaques, no major variations were observed in additional experiments with the same phage on the same host strain. Scale bar = 2 mm
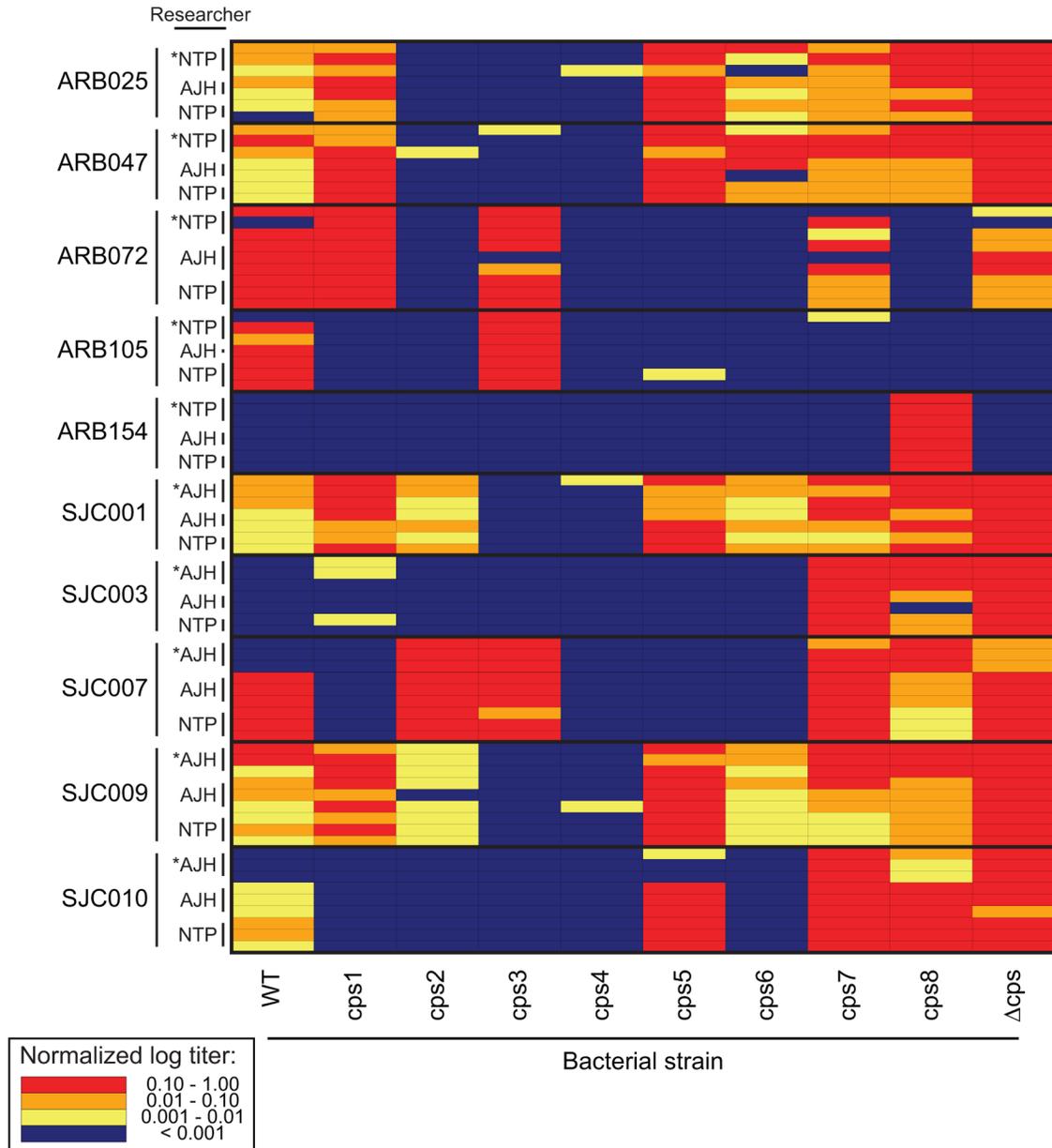
Extended Data Fig. 3. Replication of a subset of host range assays of *B. thetaiotaomicron*-targeting phages on strains expressing different CPS types.

Replication of a subset of host range assays of *B. thetaiotaomicron*-targeting phages on strains expressing different CPS types. Ten bacteriophages isolated and purified on the wild-type, acapsular, or the 8 single CPS-expressing strains were re-tested in a spot titer assay to determine phage host range. 10-fold serial dilutions of each phage ranging from approximately 10^6 to 10^3 plaque-forming units (PFU) / ml were spotted onto top agar plates containing the 10 bacterial strains. Plates were then grown overnight, and phage titers were calculated. Titers are normalized to the titer on the preferred host strain for each replicate. Each row in the heatmap corresponds to a replicate for an individual phage, whereas each column corresponds to one of the 10 host strains. One to three replicates of the assay were conducted for each phage by the two lead authors (AJH and NTP). Assays were carried out at the same time, and each author used the same set of cultures and phage stocks. For comparison, individual replicates from Figure 1 are included (marked with *). Experiment was conducted once during a visit of AJH to the University of Michigan research site to compare reproducibility between experimenters with the replication described above.
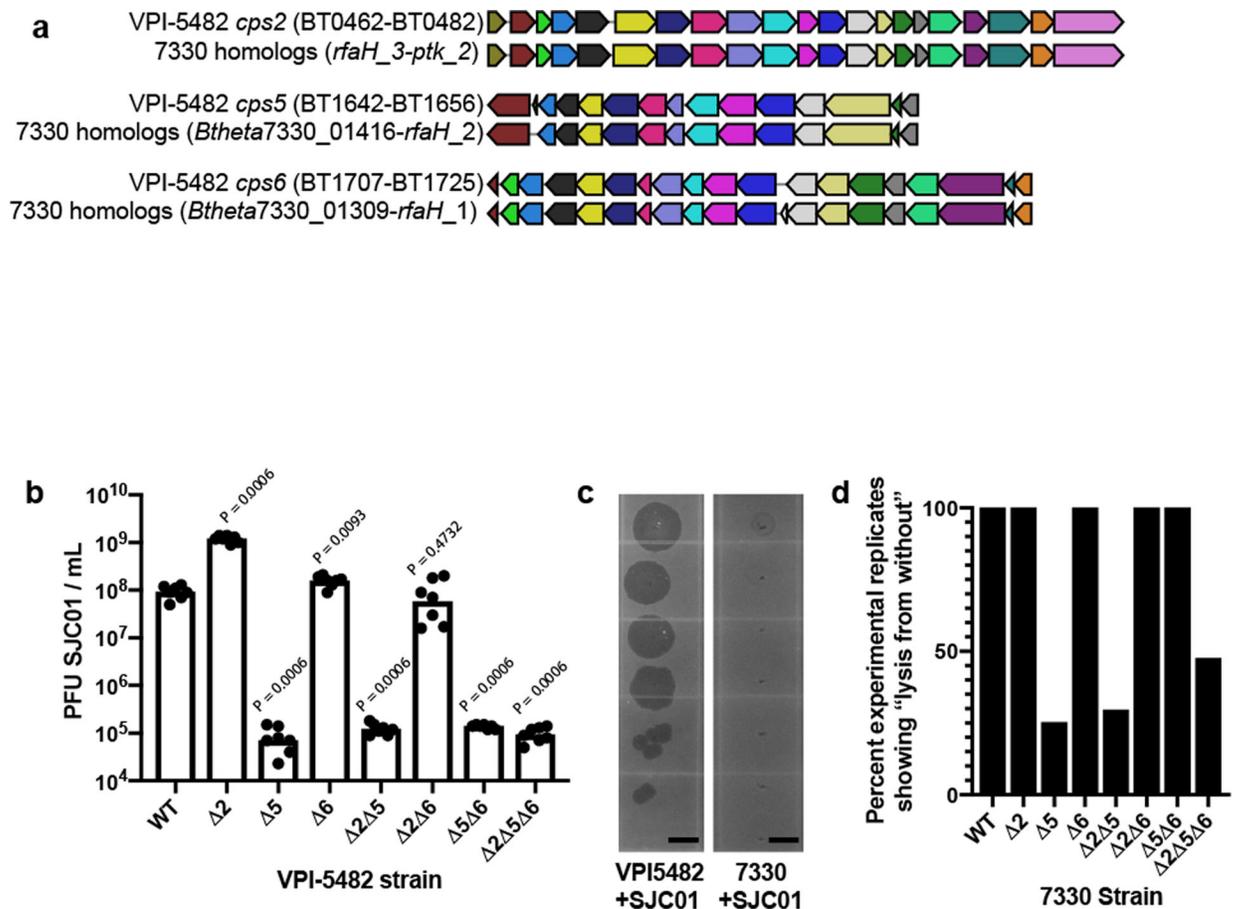

Extended Data Fig. 4. Effects of eliminating permissive CPS from another *B. thetaiotaomicron* strain.

Effects of eliminating permissive CPS from another *B. thetaiotaomicron* strain. (a) We identified *B. thetaiotaomicron* 7330[56] as the only sequenced and genetically tractable strain that contains VPI-5482-like *cps* loci (*cps2, cps5, and cps6*). Gene colors illustrate syntenic

genes with >98% amino acid identity but do not indicate function. Please see Extended Data 1 for functional annotation of *B. thetaiotaomicron* VPI-5482 cps loci. We also observed that the Branch 2 phage SJC01did not yield productive infection in *B. thetaiotaomicron* 7330, but could partially clears lawns of *B. thetaiotaomicron* 7330 at high titers. This ability to clear established lawns is a previously described phenomenon known as "lysis from without[57]".
(b) Deletion of permissive capsules (*cps2, cps5, and cps6*) either alone or in combination affects VPI-5482 infection by SJC01 (n=7 biological replicates per strain; bars represent the geometric mean). (c) Deletion of *B. thetaiotaomicron* VPI-5482-like cps loci from *B. thetaiotaomicron* 7330 affects the "lysis from without" phenotype. While SJC01 plaques on WT *B. thetaiotaomicron* VPI-5482, it does not form plaques on wild-type B. thetaiotaomicron 7330. However, SJC01 does exhibit a "lysis from without" clearing phenotypes at high densities of phage (top two spots, made with 1 microliter of 1e8 and 1e7 PFU per mL, according to titers observed on wild-type VPI-5482). Experiment was independently performed 7 times with VPI-5482 and 8 times with 7330 with similar results.
(d) *B. thetaiotaomicron* 7330 strains lacking cps5 (with the exception of 7330 Δcps5 Δcps6) show the lysis from without phenomenon less frequently than strains that have intact *cps5*. Experiments were performed with the following number of replicates per strain with similar results: wild-type 7330 (n=8), Δ2 (n=8), Δ5 (n=8), Δ6 (n=8), Δ2,5 (n=17), Δ2,6 (n=5), Δ5,6 (n=3), Δ2,5,6 (n=19). For panel b, significant differences in phage titers on each mutant strain were compared to wild type via two-tailed Mann-Whitney test with actual P values shown. For panel c, scale bars = 0.5 cm.
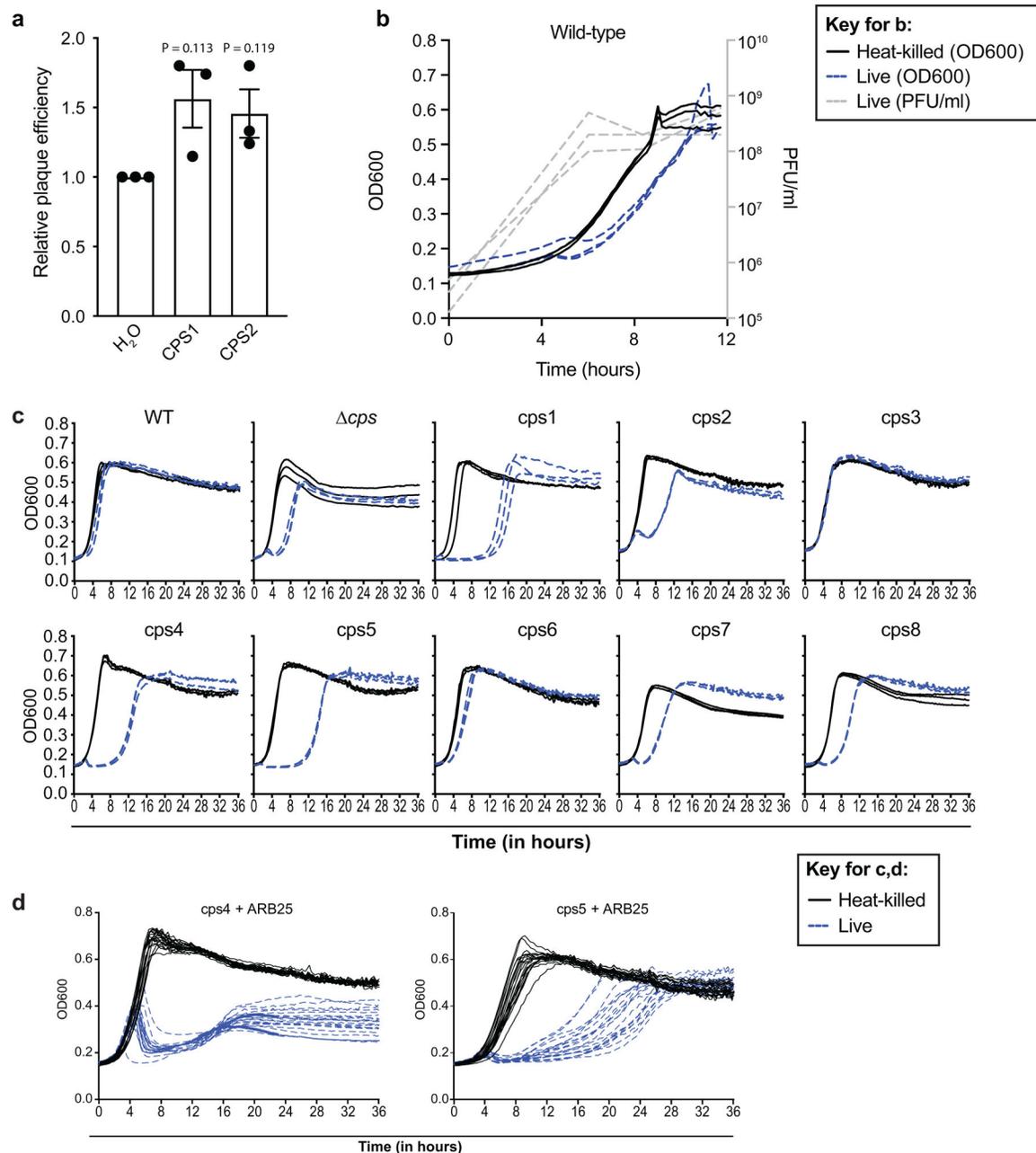
Extended Data Fig. 5. Free CPS does not inhibit ARB25 infection when provided *in trans* and effect of CPS to phage infection on bacterial growth.

Free CPS does not inhibit ARB25 infection when provided *in trans* and effect of CPS to phage infection on bacterial growth. (a) ARB25 was incubated with purified CPS1 or CPS2 (1 mg/ml, an estimated 10^9 molar excess of CPS molecules to phage, see Methods) before plating on the acapsular strain, and plaques were counted after overnight incubation. Titers are normalized to mock (H_2O) treatment. No significant differences in titers were found compared to mock treatment, as determined by Welch's t test, 2-tailed (n=3 biological replicates, bars represent mean ± SEM). (b) Post ARB25-infected, surviving cultures still contain infectious phages. Wild-type *B. thetaiotaomicron* was infected with live or heat-

killed ARB25, and bacterial growth was monitored via optical density at 600 nm (OD_{600}). At 0, 6.02, 8.36, and 11.7 hours post inoculation, replicate cultures were removed and phage levels were titered (n=3 and individual replicate curves are shown). No phages were detected in heat-killed controls. Note that the PFU/mL do not increase substantially after the initial "burst" corresponding to decreased bacterial culture density prior to re-growth. (c) Ten strains: the wild-type (WT), the acapsular strain (cps), or the eight single CPS-expressing strains were infected with either live or heat-killed SJC01. (d) 20 different colonies of cps4 or cps5 strains were infected with ARB25. Growth was monitored via optical density at 600 nm (OD_{600}) on an automated plate reading instrument as described in Methods and individual growth curves for live and heat-killed phage exposure are shown separately.
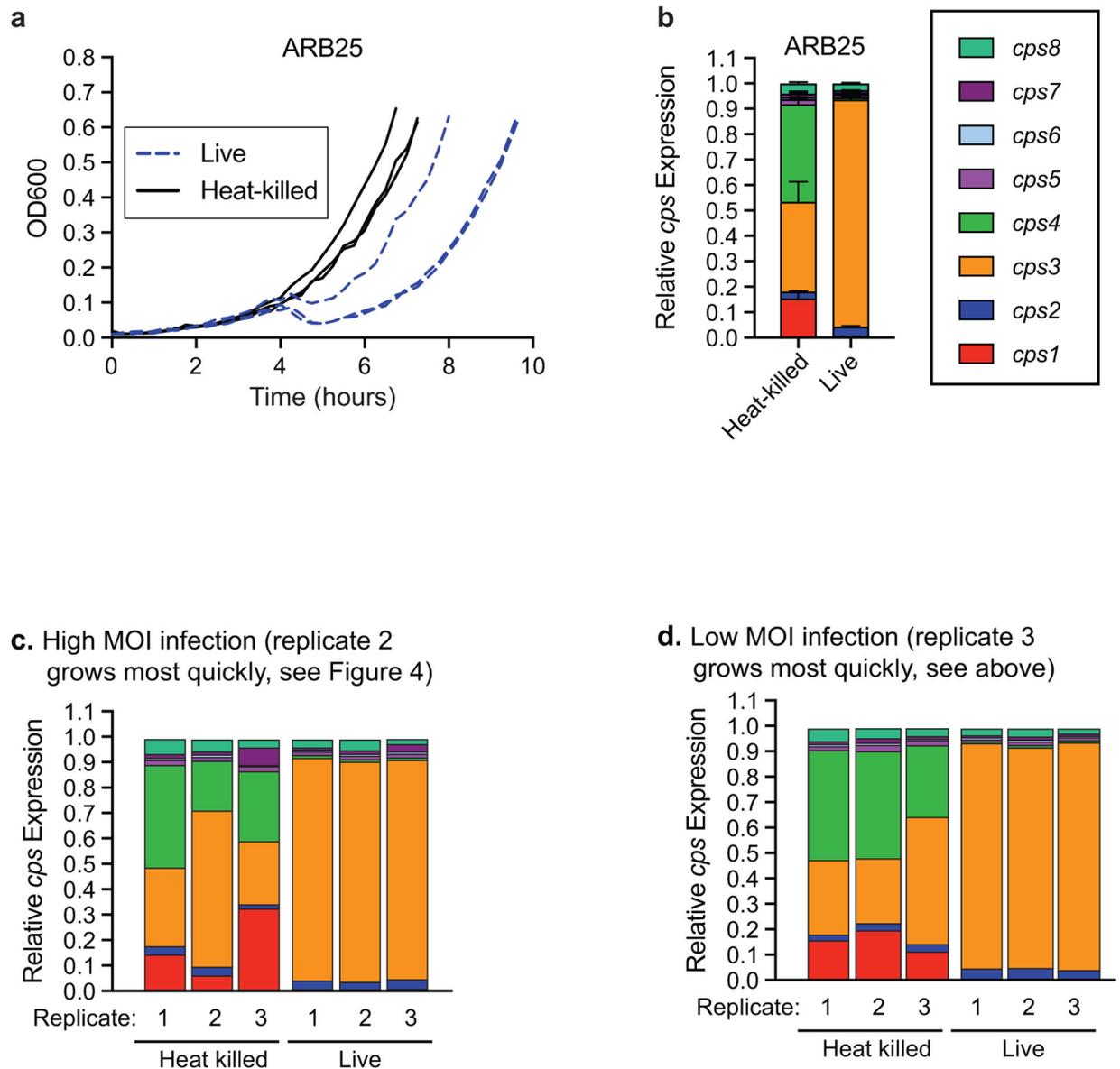
Extended Data Fig. 6. Infection of wild-type *B. thetaiotaomicron* at a low multiplicity of infection and subsequent effects on *cps* gene expression.

Infection of wild-type *B. thetaiotaomicron* at a low multiplicity of infection and subsequent effects on *cps* gene expression. (a) The wild-type (WT) strain was infected at a low multiplicity of infection (MOI = 1×10^{-4}) of live or heat-killed ARB25, and bacterial growth was monitored via OD_{600} (n=3 biological replicates and separate curves are shown). (b) RNA was harvested from cultures after reaching an OD_{600} of 0.6-0.7, cDNA was generated, and relative expression of the 8 *cps* loci was determined by qPCR (histogram bars are mean +± SEM of 3 biological replicates). Individual replicates of high MOI (c) and low MOI (d). In the high MOI experiment, replicate 2 showed higher starting expression of the non-permissive CPS3 compared to others. In the low MOI experiment, replicate 3 showed higher starting expression of the non-permissive CPS3. In both experiments, post phage-exposed replicates displayed nearly identical CPS expression profiles characterized by high expression of CPS3. The experiments in a-c were repeated one time with three parallel biological replicates started from single *B. thetaiotaomicron* wild-type colonies picked from the same plate.
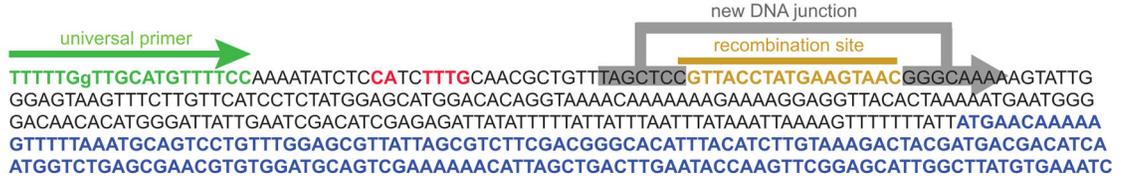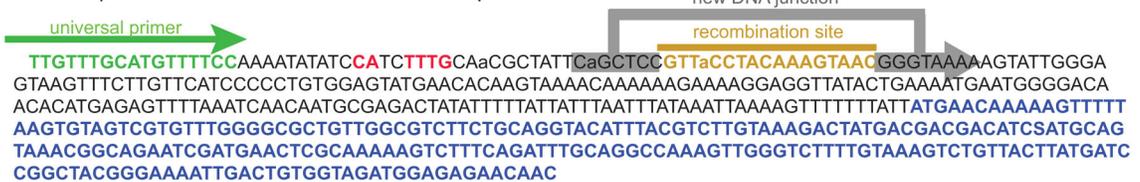
BT1927 promoter "on" recombination PCR sequence:

new DNA junction (published)

recombination site

universal primer

TTTTGTTTGCATGTTTTCCAAAATATACC**CA**TAt**TTG**CaGCGTCTTTTCTTCCc**GTTaCCTACAAAGTAAC**GGAA
AGACAGACCACCAAGATGAGATGGATTATGAAATGACAGCAAACAAATTTTTTATTTATTAATTGTATAACATTTAATTT
ACTTTTATGAGAAAATGGACGTATTTGGTCGCAACCTTGCTA**ATGGCGGGTACGACAGCTACATTCACGGGATGAA**

BT1826 promoter "on" recombination PCR sequence:

new DNA junction

recombination site

universal primer

TTTTTGTTTGCATGTTTTCCAAAATATGTC**CA**TA**TTTG**CAGTGCTATTCAGTCCC**GTTACCTACCAAGTAAC**GGGACAGGCATTATTG
AAACTTAGTGTTTAATAATATATTATTAATGTAATTTTTAAGTAGTT**ATGAAAAGAAAATTTGTAAAAGTGATGTTCTTCGGGGCGTTAG**
CACTTTCTACCGTCACCTACGTAGGCTGTAAGGACTACGATGATGACATCAAAAACGTGCAGGAACAGATTGATGCCATCAACAAG
AAAGGTGCTGATGTTACAACTGAAGCAATGACTGCTGCTATTAATAGCGCAGTAGCCGGATTGCAGACACAATTGGATGCTATCGC

BT1502 promoter "on" recombination PCR sequence:

new DNA junction

recombination site

universal primer

TTTTTGgTTGCATGTTTTCCAAAATATCTC**CA**TC**TTTG**CAACGCTGTTTAGCTCC**GTTACCTATGAAGTAAC**GGGCAAAAAGTATTG
GGAGTAAGTTTCTTGTTCATCCTCTATGGAGCATGGACACAGGTAAAACAAAAAAGAAAAGGAGGTTACACTAAAAATGAATGGG
GACAACACATGGGATTATTGAATCGACATCGAGAGATTATATTTTTATTATTTAATTTATAAATTAAAAGTTTTTTTATT**ATGAACAAAAA**
GTTTTTAAATGCAGTCCTGTTTGGAGCGTTATTAGCGTCTTCGACGGGCACATTTACATCTTGTAAAGACTACGATGACGACATCA
ATGGTCTGAGCGAACGTGTGGATGCAGTCGAAAAAAACATTAGCTGACTTGAATACCAAGTTCGGAGCATTGGCTTATGTGAAATC

BT1507 promoter "on" recombination PCR sequence:

new DNA junction

recombination site

universal primer

TTGTTTGCATGTTTTCCAAAATATATC**CA**TC**TTTG**CAaCGCTATTCaGCTCC**GTTaCCTACAAAGTAAC**GGGTAAAAAGTATTGGGA
GTAAGTTTCTTGTTCATCCCCCTGTGGAGTATGAACACAAGTAAAACAAAAAAGAAAAGGAGGTTACATTACTGAAAATGAATGGGGACA
ACACATGAGAGTTTTAAATCAACAATGCGAGACTATATTTTTTATTATTTAATTTATAAATTAAAAGTTTTTTTATT**ATGAACAAAAAGTTTTT**
AAGTGTAGTCGTGTTTGGGGCGCTGTTGGCGTCTTCTGCAGGTACATTTACGTCTTGTAAAGACTATGACGACGACATCSATGCAG
TAAACGGCAGAATCGATGAACTCGCAAAAAGTCTTTCAGATTTGCAGGCCAAAGTTGGGTCTTTTGTAAAGTCTGTTACTTATGATC
CGGCTACGGGAAAATTGACTGTGGTAGATGGAGAGAACAAC

BT4481 promoter "on" recombination PCR sequence:

new DNA junction

recombination site

universal primer

CATGTTTTCCAAAATATATC**CA**TC**TTTG**CAACGCTaTTTAGCTCC**GTTACCTAGaAAGTAAC**GGGTAAAGGCTGTT
TGGCCCGTATGCTTCATAGTGTATTATATAAGGGAAATAAAGGTTGAATGAATTAAGAAAAGGAGGTCGCACGAGAAATGAACGG
AACAACACATGGGATTTAAAATCAAAAGTGAGAAATTATATTTTTATTATTTTTTTAATTTATAAATTAAAAGTTTTTTTATT**ATGAACAA**
AAAGTTTTTAAATGCTGTCCTGTTTGGAGCGCTGATAGCGTCCTCGGCCAGGTACATTTACTTCTTGTAAAGACTACGATGMCG
ACATCAAGAGTCTGCAGGAACAAATTGACAAAAGCGGAAGCACCGTTGGCGATCTTCAGACTCAGCTGACAACGCTGAAGA
GTGCTGCAGAAGCTGCAtCAGGCTGCTGCTGATGCTGCAAAGAgCAGCTGCTGCAGAAGCTAAGACTGCCGCAGAAGCTGC

BT2486 promoter "on" recombination PCR sequence:

new DNA junction

recombination site

TCCTTGCGCGTTTTCTcCCC**GTT-CCTACAA-GTAAC**GGGCGAATATTATTTG
AACTTAGTGTTTAATAATATATTATTAATGTAATTTTTAAGTAGTT**ATGAAGAAAAAATTTATTAGAGTGATGTTTTTCGGAGCATTG**
GCGCTAACCACATTTTCATATGTAGGTTGTAAGGACTACGATGACGATATCGATCGGTTAGAGCAGAAGATAACCGAAAATGC
CAACGCTATTGCGGAAATCAACAAGCTGATCGAAGGGGGATCTGTGATCACAAAAGTTGAGAACATTACCGGTGGTGTGGA
AGTAACATTGAGCAACGGTAAGAGCTTTACCATTAAAAATGGTGATGCCGGAAAAGATGGTACTCAGTGGACGATTGGTGAA
GACGGATACTGGTATCAAGATGGTAAAAA

Extended Data Fig. 7. Determination of phase-variable promoter switching for six loci encoding putative S-layer proteins.

Determination of phase-variable promoter switching for six loci encoding putative S-layer proteins. The hypothesis that the promoters associated with seven additional *B. thetaiotaomicron* S-layer like lipoproteins was validated using a PCR amplicon sequencing strategy. Because of high nucleotide identity in both the regions flanking the 7 additional loci, a nested PCR approach was required to specifically amplify and sequence each site. In the first step, a primer lying in each S-layer gene (Supplemental Table 5 "S-layer gene" primers) was oriented towards the promoter and used in a PCR extension to a primer in the upstream recombinase gene (Supplemental Table 5 "recombinase gene 3" primer). The products of this PCR were purified without gel extraction and used in a second reaction with a nested primer that lies internal to the previous recombinase gene primer (Supplemental

Table 5 "recombinase 2" primer). The expected PCR products from this reaction, which are ~1 kb and span promoter sequences in both the ON and OFF orientations, were excised and used for an orientation-specific PCR using the original S-layer gene primer for each site and a universal primer (green schematic) that was designed for each promoter and is oriented to extend upstream of the S-layer gene (e.g., OFF orientation). Resulting products from this third reaction, which should correspond to the ON orientation if a promoter inversion has occurred in some cells, were obtained for 5/7 of the additional identified loci and the BT1927 S-layer locus as a control. In all cases in which an amplicon and sequence were obtained, the expected recombination occurred between the inverted repeat site proximal to the S-layer gene start (new DNA junction), which would orient the promoter to enable expression of the downstream S-layer gene. The sequences shown are the consensus between forward and reverse reads for each amplicon. The putative core promoter -7 sequence is shown in bold/red text, the coding region of each S-layer gene is shown in bold/blue text and the S-layer gene proximal recombination site is noted and highlighted in bold/gold text. Note that the 5'-end of the sequenced amplicon was not resolved for the BT2486 locus.
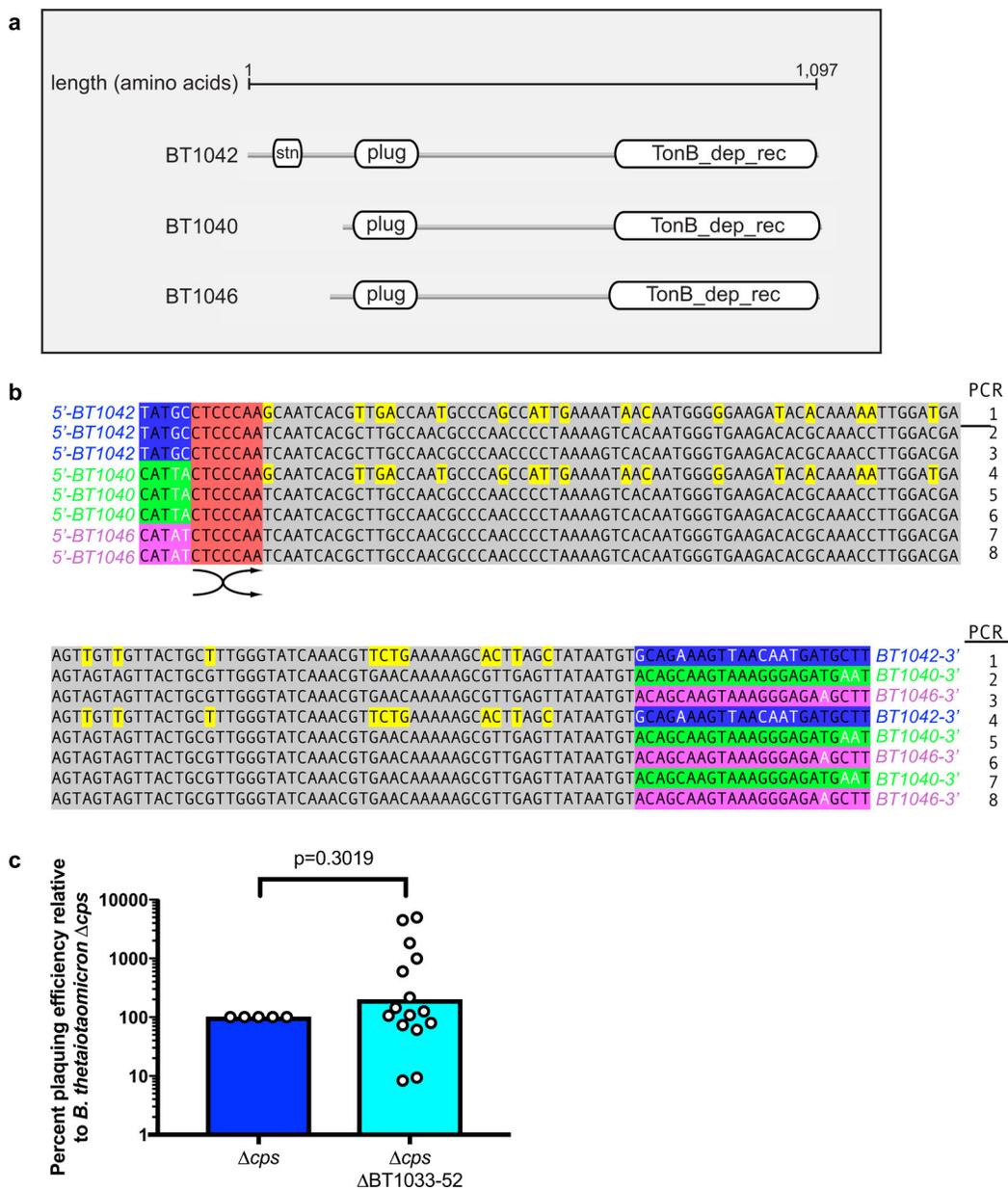
Extended Data Fig. 8. Recombination between the genes BT1040, BT1042, and BT1046 and effect of BT1033-52 locus.

Recombination between the genes BT1040, BT1042, and BT1046 and effect of BT1033-52 locus. (a) Pfam domain schematics of the amino acid sequences of these three genes highlighting that BT1040 and BT1046, as originally assembled in the *B. thetaiotaomicron* genome sequence, lack additional N-terminal sequences that are present on BT1042. (b) Sequencing of the 8 PCR amplicons schematized in Figure 5d. Amplicons 1, 5 and 8 represent the original genome architecture, while the others represent inferred recombination events that are validated here by sequencing. The 5' and 3' ends of the BT1042, BT1040 and BT1046 genes are color-coded to assist in following their connectivity changes after recombination. A series of single-nucleotide polymorphisms (SNPs) present in BT1042, downstream of the proposed recombination site, are highlighted in yellow. The transfer of

these SNPs to a fragment containing the 5' end of BT1040 (Amplicon 4) was used to narrow the recombination region to the 7 nucleotide sequence highlighted in red. Additional SNPs that are specific to the regions upstream of this recombination site are shown in white text for each sequence. Susceptibility of acapsular *B. thetaiotaomicron* to ARB25 without the BT1022-52 locus is not affected. (c) Ten-fold serial dilutions of ARB25 were spotted onto lawns of *B. thetaiotaomicron* Δ*cps* (n=5) and *B. thetaiotaomicron* Δ*cps* ΔBT1033-52 (n=5). Each of the 5 biological replicates contained 3 technical replicates from independent clones and all 15 replicates are shown individually. Plaquing efficiency was determined by normalizing plaque counts on *B. thetaiotaomicron* Δ*cps* ΔBT1033-52 relative to plaque counts on *B. thetaiotaomicron* Δ*cps* for each replicate. Statistical significance was determined using the 2-tailed Mann-Whitney test and bars represent the geometric mean.
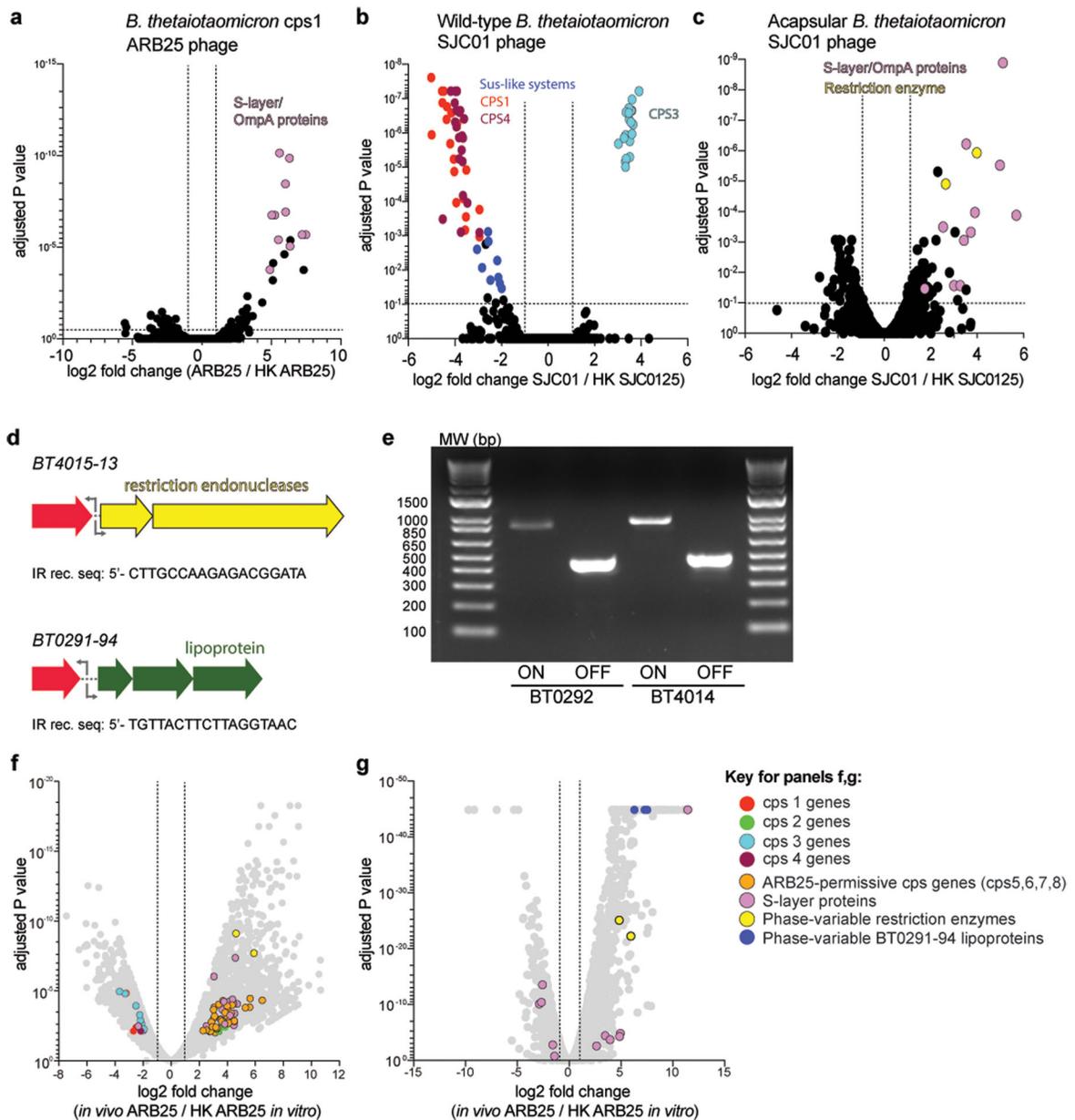
Extended Data Fig. 9. Whole genome transcriptional analyses of several additional *B. thetaiotaomicron* strain and phage combinations.
Whole genome transcriptional analyses of several additional *B. thetaiotaomicron* strain and phage combinations. (a) Infection of the cps1 strain with ARB25, revealing a post-infection response that is largely characterized by increased expression of S-layer/OmpA proteins. (b) Infection of wild-type *B. thetaiotaomicron* with SJC01, revealing that, as with ARB25/wild-type, the bacteria survive phage infection by mostly altering CPS expression. Expression of the non-permissive CPS3 is prominently increased. (c) Infection of acapsular *B. thetaiotaomicron* with SJC01, revealing that in the absence of CPS survival is mostly promoted by increased S-layer/OmpA expression and expression of a phase-variable restriction enzyme system. Transcript abundance values in panels a-c were compared between live and HK treatments to generate fold change (x axis), which is plotted against the

adjusted P value (EdgeR) for each gene that was generated using an exact test adapted for overdispersed data[55] (n=3 biological replicates for each panel, peformed with similar results). Please see Supplemental Tables 5-7 for actual adjusted P values for panels a-c. (d) Gene schematic of the phase-variable restriction enzyme system (top) and a lipoprotein contain locus (bottom) that is different from the 8 S-layer loci also revealed in this study. The inverted repeat sequence that was determined to mediate recombination in each locus is shown. (e) PCR analysis of the restriction enzyme system and additional lipoprotein promoter orientations with primers designed to detect phase variation from off to on states. Amplicons were sequenced to confirm the re-orientation to the on orientation (not shown). Experiment was performed once. (f) Global transcriptional responses of wild-type *B. thetaiotaomicron* in the ceca of mice after 72 d of co-existence with ARB25. Note that shifts in CPS expression are mostly characterized by increases in permissive CPS, which may be dictated by growth *in vivo* selecting for these capsules or against the non-permissive CPS3. Correspondingly, wild-type shows increased expression of some but not all S-layer/OmpA systems and the phase-variable restriction enzyme. (g) Global transcriptional responses of acapsular *B. thetaiotaomicron* in the ceca of mice after 72 d of co-existence with ARB25. In the absence of CPS, surviving bacteria show increased expression of only a subset of the identified S-layer/OmpA proteins, with BT1826 expressed most dominantly, along with the BT0291-94 locus and expression of the restriction enzyme system. Experiments in panels f-g are the results of 3 separate biological replicates, statistical tests are identical to those described in panels a-c. The dashed lines in panels a,b,c,f,g represent the adjusted P value cutoff (0.01) that was used to establish significance, which was generated using an exact test adapted for overdispersed data[55]

a

b

- acapsular BT1927 locked on, heat-killed phage
- acapsular BT1927 locked off, heat killed phage
- acapsular BT1927 locked on, live phage
- acapsular BT1927 locked off, live phage
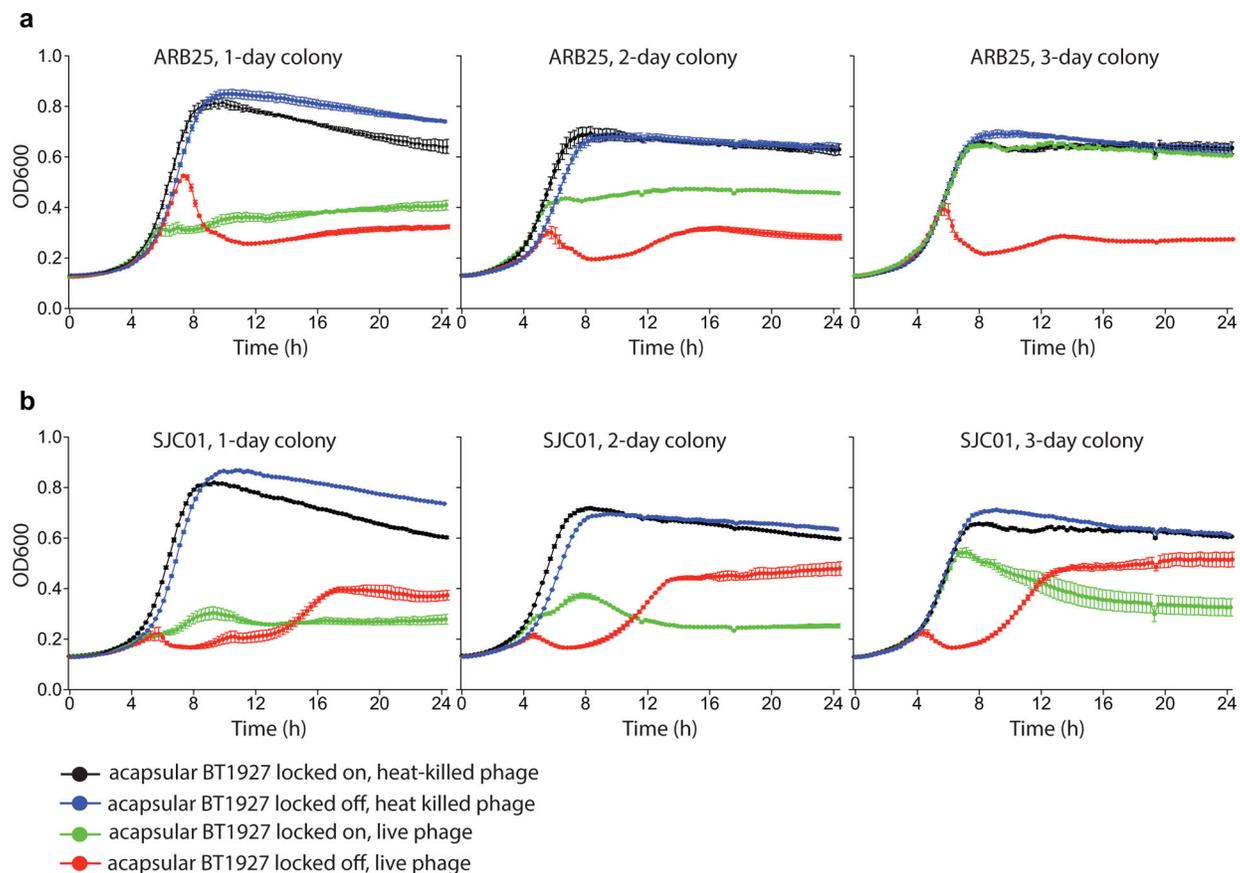
Extended Data Fig. 10. ARB25 (a) or SJC01(b) infection of the acapsular BT1927 locked on and off strains after 1, 2 or 3 days of growth on BPRM.

ARB25 (a) or SJC01(b) infection of the acapsular BT1927 locked on and off strains after 1, 2 or 3 days of growth on BPRM (n=6 biological replicates per treatment, performed with similar results). Three separate colonies were picked each day as a biological replicate, grown overnight and used to setup infection cultures that were monitored for 24 hours in an automated plate reader. Colonies picked after only 1 day show the least resistance to either phage when BT1927 is locked on. After 2 days, resistance is increased and this continues to increase after 3 days, becoming almost complete (compared to HK controls for ARB25). Growth curves represent the mean of 6 biological replicates ± standard error.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Eckburg PB et al. Diversity of the human intestinal microbial flora. Science (New York, N.Y.) 308, 1635–1638, doi:10.1126/science.1110591 (2005).

2. Qin J et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464, 59–65 (2010). [PubMed: 20203603]

3. Faith JJ et al. The long-term stability of the human gut microbiota. Science (New York, N.Y.) 341, 1237439, doi:10.1126/science.1237439 (2013).

4. Donia MS et al. A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. Cell 158, 1402–1414, doi:10.1016/j.cell.2014.08.032 (2014). [PubMed: 25215495]

5. Coyne MJ & Comstock LE Niche-specific features of the intestinal bacteroidales. Journal of bacteriology 190, 736–742, doi:10.1128/jb.01559-07 (2008). [PubMed: 17993536]

6. Neff CP et al. Diverse Intestinal Bacteria Contain Putative Zwitterionic Capsular Polysaccharides with Anti-inflammatory Properties. Cell host & microbe 20, 535–547, doi:10.1016/j.chom.2016.09.002 (2016). [PubMed: 27693306]

7. Peterson DA, McNulty NP, Guruge JL & Gordon JI IgA response to symbiotic bacteria as a mediator of gut homeostasis. Cell host & microbe 2, 328–339, doi:10.1016/j.chom.2007.09.013 (2007). [PubMed: 18005754]

8. Porter NT, Canales P, Peterson DA & Martens EC A Subset of Polysaccharide Capsules in the Human Symbiont Bacteroides thetaiotaomicron Promote Increased Competitive Fitness in the Mouse Gut. Cell host & microbe 22, 494–506.e498, doi:10.1016/j.chom.2017.08.020 (2017). [PubMed: 28966055]

9. Round JL et al. The Toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. Science (New York, N.Y.) 332, 974–977, doi:10.1126/science.1206095 (2011).

10. Shen Y et al. Outer membrane vesicles of a human commensal mediate immune regulation and disease protection. Cell host & microbe 12, 509–520, doi:10.1016/j.chom.2012.08.004 (2012). [PubMed: 22999859]

11. Patrick S et al. Twenty-eight divergent polysaccharide loci specifying within- and amongst-strain capsule diversity in three strains of Bacteroides fragilis. Microbiology (Reading, England) 156, 3255–3269, doi:10.1099/mic.0.042978-0 (2010).

12. Porter NT & Martens EC The Critical Roles of Polysaccharides in Gut Microbial Ecology and Physiology. Annual review of microbiology 71, 349–369, doi:10.1146/annurev-micro-102215-095316 (2017).

13. Pasolli E et al. Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. Cell 176, 649–662 e620, doi:10.1016/j.cell.2019.01.001 (2019). [PubMed: 30661755]

14. Xu J et al. A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. Science (New York, N.Y.) 299, 2074–2076, doi:10.1126/science.1080029 (2003).

15. Krinos CM et al. Extensive surface diversity of a commensal microorganism by multiple DNA inversions. Nature 414, 555–558, doi:10.1038/35107092 (2001). [PubMed: 11734857]

16. Martens EC, Roth R, Heuser JE & Gordon JI Coordinate regulation of glycan degradation and polysaccharide capsule biosynthesis by a prominent human gut symbiont. The Journal of biological chemistry 284, 18445–18457, doi:10.1074/jbc.M109.008094 (2009). [PubMed: 19403529]

17. Hsieh S et al. Polysaccharide Capsules Equip the Human Symbiont Bacteroides thetaiotaomicron to Modulate Immune Responses to a Dominant Antigen in the Intestine. The Journal of Immunology, ji1901206, doi:10.4049/jimmunol.1901206 (2020).

18. Kuwahara T et al. Genomic analysis of Bacteroides fragilis reveals extensive DNA inversions regulating cell surface adaptation. Proceedings of the National Academy of Sciences of the United States of America 101, 14919–14924, doi:10.1073/pnas.0404172101 (2004). [PubMed: 15466707]

19. Chatzidaki-Livanis M, Coyne MJ & Comstock LE A family of transcriptional antitermination factors necessary for synthesis of the capsular polysaccharides of Bacteroides fragilis. Journal of bacteriology 191, 7288–7295, doi:10.1128/jb.00500-09 (2009). [PubMed: 19801412]

20. Chatzidaki-Livanis M, Weinacht KG & Comstock LE Trans locus inhibitors limit concomitant polysaccharide synthesis in the human gut symbiont Bacteroides fragilis. Proceedings of the National Academy of Sciences of the United States of America 107, 11976–11980, doi:10.1073/pnas.1005039107 (2010). [PubMed: 20547868]

21. Duerkop BA et al. Murine colitis reveals a disease-associated bacteriophage community. Nature microbiology, doi:10.1038/s41564-018-0210-y (2018).

22. Manrique P et al. Healthy human gut phageome. Proceedings of the National Academy of Sciences of the United States of America 113, 10400–10405, doi:10.1073/pnas.1601060113 (2016). [PubMed: 27573828]

23. Minot S et al. The human gut virome: inter-individual variation and dynamic response to diet. Genome research 21, 1616–1625, doi:10.1101/gr.122705.111 (2011). [PubMed: 21880779]

24. Reyes A et al. Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature 466, 334–338, doi:10.1038/nature09199 (2010). [PubMed: 20631792]

25. Norman JM et al. Disease-specific alterations in the enteric virome in inflammatory bowel disease. Cell 160, 447–460, doi:10.1016/j.cell.2015.01.002 (2015). [PubMed: 25619688]

26. Rogers TE et al. Dynamic responses of Bacteroides thetaiotaomicron during growth on glycan mixtures. Molecular microbiology 88, 876–890, doi:10.1111/mmi.12228 (2013). [PubMed: 23646867]

27. Cockburn DW & Koropatkin NM Polysaccharide Degradation by the Intestinal Microbiota and Its Influence on Human Health and Disease. Journal of molecular biology 428, 3230–3252, doi:10.1016/j.jmb.2016.06.021 (2016). [PubMed: 27393306]

28. Martens EC, Koropatkin NM, Smith TJ & Gordon JI Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. The Journal of biological chemistry 284, 24673–24677, doi:10.1074/jbc.R109.022848 (2009). [PubMed: 19553672]

29. Braun V FhuA (TonA), the career of a protein. Journal of bacteriology 191, 3431–3436, doi:10.1128/jb.00106-09 (2009). [PubMed: 19329642]

30. Taketani M, Donia MS, Jacobson AN, Lambris JD & Fischbach MA A Phase-Variable Surface Layer from the Gut Symbiont Bacteroides thetaiotaomicron. mBio 6, e01339–01315, doi:10.1128/mBio.01339-15 (2015). [PubMed: 26419879]

31. Glowacki RWP et al. A Ribose-Scavenging System Confers Colonization Fitness on the Human Gut Symbiont Bacteroides thetaiotaomicron in a Diet-Specific Manner. Cell host & microbe 27, 79–92.e79, doi:10.1016/j.chom.2019.11.009 (2020). [PubMed: 31901520]

32. Cuskin F et al. Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. Nature 517, 165–169, doi:10.1038/nature13995 (2015). [PubMed: 25567280]

33. Keller R & Traub N The characterization of Bacteroides fragilis bacteriophage recovered from animal sera: observations on the nature of bacteroides phage carrier cultures. The Journal of general virology 24, 179–189, doi:10.1099/0022-1317-24-1-179 (1974). [PubMed: 4843748]

34. Lwoff A Lysogeny. Bacteriol Rev 17, 269–337 (1953). [PubMed: 13105613]

35. Shkoporov AN et al. PhiCrAss001 represents the most abundant bacteriophage family in the human gut and infects Bacteroides intestinalis. Nature communications 9, 4781, doi:10.1038/s41467-018-07225-7 (2018).

36. Bjursell MK, Martens EC & Gordon JI Functional Genomic and Metabolic Studies of the Adaptations of a Prominent Adult Human Gut Symbiont, Bacteroides thetaiotaomicron, to the Suckling Period. Journal of Biological Chemistry 281, 36269–36279, doi:10.1074/jbc.M606509200 (2006). [PubMed: 16968696]

37. Martens EC, Chiang HC & Gordon JI Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. Cell host & microbe 4, 447–457, doi:10.1016/j.chom.2008.09.007 (2008). [PubMed: 18996345]

38. Sonnenburg JL et al. Glycan Foraging in Vivo by an Intestine-Adapted Bacterial Symbiont. Science (New York, N.Y.) 307, 1955–1959, doi:10.1126/science.1109051 (2005).

39. Barr JJ et al. Bacteriophage adhering to mucus provide a non-host-derived immunity. Proceedings of the National Academy of Sciences of the United States of America 110, 10771–10776, doi:10.1073/pnas.1305923110 (2013). [PubMed: 23690590]

40. De Sordi L, Lourenço M & Debarbieux L The Battle Within: Interactions of Bacteriophages and Bacteria in the Gastrointestinal Tract. Cell host & microbe 25, 210–218, doi:10.1016/j.chom.2019.01.018 (2019). [PubMed: 30763535]

41. Tomlinson S & Taylor PW Neuraminidase associated with coliphage E that specifically depolymerizes the Escherichia coli K1 capsular polysaccharide. Journal of virology 55, 374–378 (1985). [PubMed: 3894684]

42. Gupta DS et al. Coliphage K5, specific for E. coli exhibiting the capsular K5 antigen. FEMS Microbiology Letters 14, 75–78, doi:10.1111/j.1574-6968.1982.tb08638.x (1982).

43. Tartera C, Araujo R, Michel T & Jofre J Culture and decontamination methods affecting enumeration of phages infecting Bacteroides fragilis in sewage. Applied and environmental microbiology 58, 2670–2673 (1992). [PubMed: 1514815]

44. Araujo R et al. Optimisation and standardisation of a method for detecting and enumerating bacteriophages infecting Bacteroides fragilis. Journal of virological methods 93, 127–136 (2001). [PubMed: 11311351]

45. Eddelbuettel D & François R Rcpp: Seamless R and C++ Integration. Journal of Statistical Software 40 (2011).

46. Fu L, Niu B, Zhu Z, Wu S & Li W CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28, 3150–3152, doi:10.1093/bioinformatics/bts565 (2012). [PubMed: 23060610]

47. Katoh K, Misawa K, Kuma K & Miyata T MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30, 3059–3066, doi:10.1093/nar/gkf436 (2002). [PubMed: 12136088]

48. Price MN, Dehal PS & Arkin AP FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 26, 1641–1650, doi:10.1093/molbev/msp077 (2009). [PubMed: 19377059]

49. Paradis E & Schliep K ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35, 526–528, doi:10.1093/bioinformatics/bty633 (2019). [PubMed: 30016406]

50. Revell LJ phytools: an R package for phylogenetic comparative biology (and other things). . Methods in Ecology and Evolution 3, 217–223 (2012).

51. Koropatkin N, Martens EC, Gordon JI & Smith TJ Structure of a SusD homologue, BT1043, involved in mucin O-glycan utilization in a prominent human gut symbiont. Biochemistry 48, 1532–1542, doi:10.1021/bi801942a (2009). [PubMed: 19191477]

52. Medema MH, Takano E & Breitling R Detecting sequence homology at the gene cluster level with MultiGeneBlast. Mol Biol Evol 30, 1218–1223, doi:10.1093/molbev/mst025 (2013). [PubMed: 23412913]

53. MJ M DirichletReg: Dirichlet Regression in R. (2015).

54. Robinson MD, McCarthy DJ & Smyth GK edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140, doi:10.1093/bioinformatics/btp616 (2010). [PubMed: 19910308]
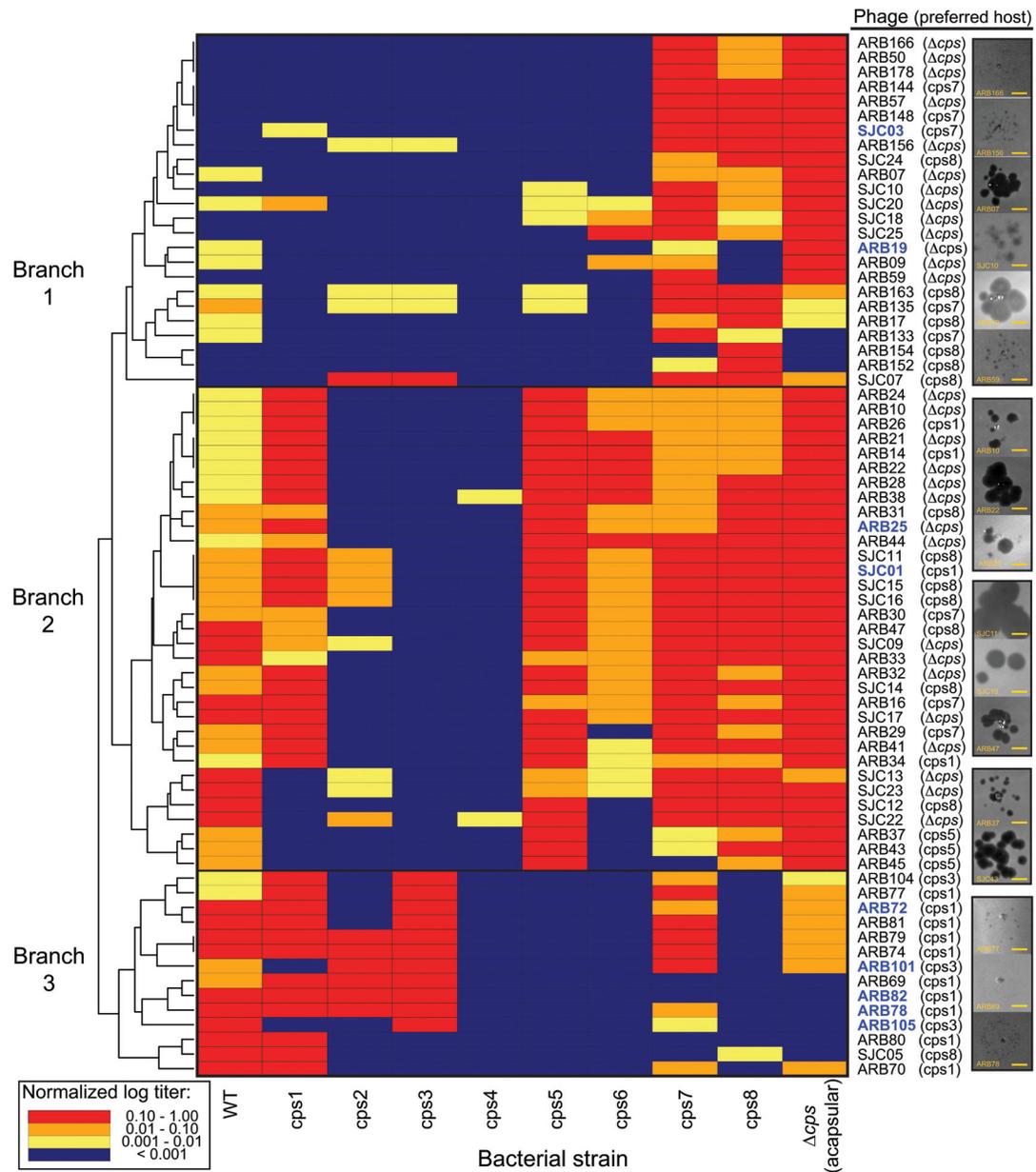
Figure 1.

Host range of *B. thetaiotaomicron* phages on strains expressing different CPS. Seventy-one bacteriophages were isolated and purified on the wild-type, *cps* (acapsular), or the 8 single CPS-expressing *B. thetaiotaomicron* strains. High titer phage stocks were prepared on their "preferred host strain", which was the strain that yielded the highest titer of phages in a pre-screen and is listed next to each phage. Phages were then tested in a quantitative host range assay. Phage titers were calculated for each bacterial host and normalized to the preferred host strain for each replicate, and 3 replicates averaged for each assay. The results clustered based on plaquing efficiencies (see Methods). The dendrogram on the left of the heatmap displays the clustering of phages based on similar plaquing efficiencies, with longer branch lengths representing greater distance between individual phages. Images at the far right of

the figure illustrate the range of plaque morphologies of select phages from the collection (see Extended Data 2 for images of plaques for all phages). Several phages that are the subjects of additional follow up studies are highlighted in blue text. Scale bar = 2mm. Each experiment was replicated with similar results a minimum of 3 separate times, including at both research sites at the University of Michigan and Stanford University.
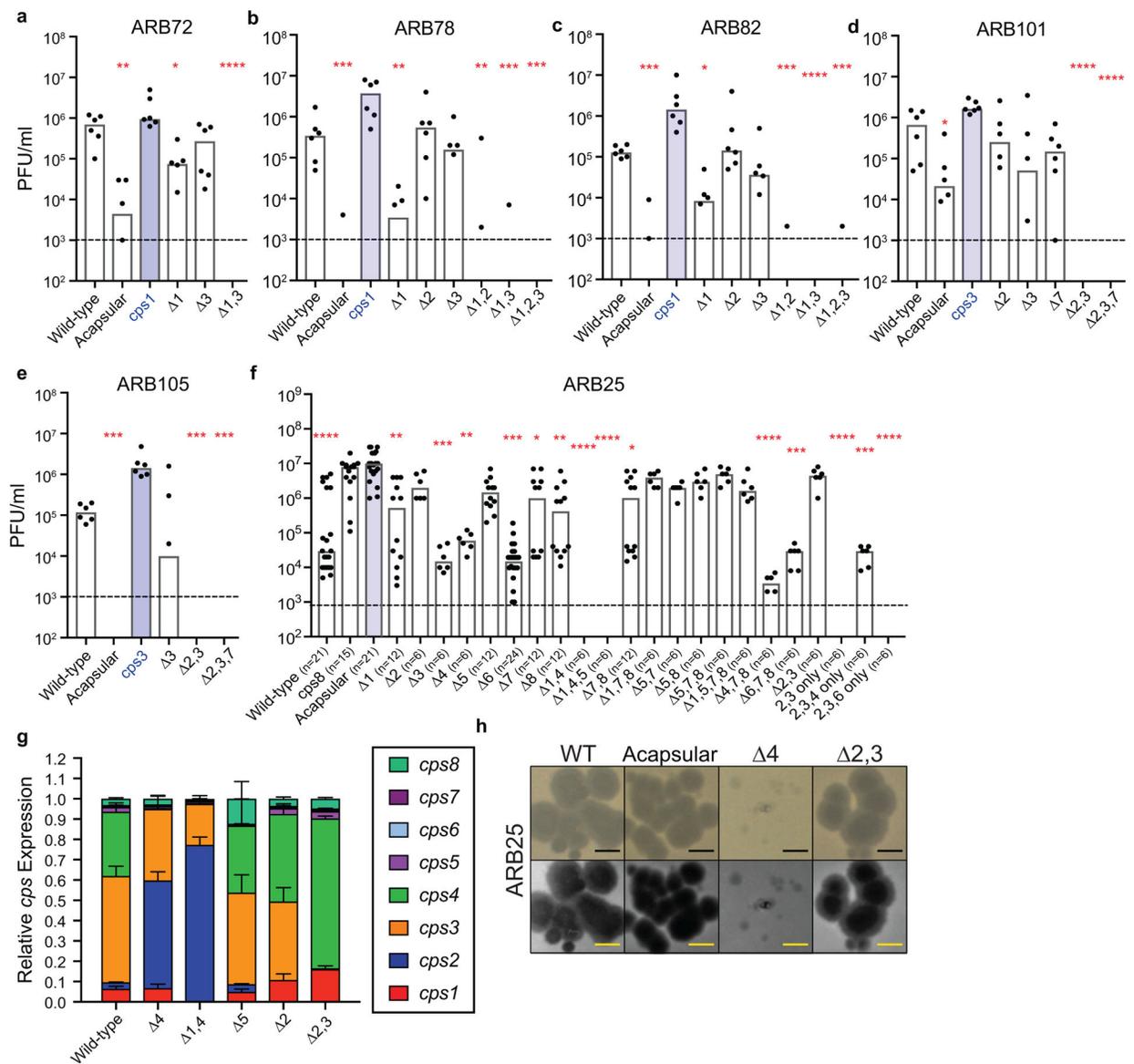
Figure 2.
Infection of various CPS mutant strains by Branch 3 phages ARB72 (a), ARB78 (b), ARB82 (c), ARB101 (d) and ARB105 (e) is inhibited by eliminating most or all of the permissive CPS from wild-type *B. thetaiotaomicron*. Each phage was tested on the wild-type strain, the acapsular strain, their respective preferred host strain (blue bars), and a set of bacterial strains harboring selected *cps* locus deletions that correspond to their predetermined host range (n = 6 replicates/phage). (f) Elimination of permissive CPS from Branch 2 phage ARB25 reduces infection, but complete reduction of infection only occurs in the context of deleting more than one permissive CPS. The number of replicates (n=6–21) conducted on each strain is annotated in parentheses next to the strain name. For panels a-f, each experiment was replicated with similar results a minimum of 6 separate times. The dashed black lines in panels a-f represent the limit of detection (1,000 pfu/ml) based on the dilution scheme used. (g) Relative *cps* locus expression of the 8 *cps* loci in the indicated

strains. Experiment was performed 3 separate times with similar results and measurements averaged in the plot displayed. (h) Representative pictures of phage plaques on the indicated host strains. The top row of images for each phage is unaltered; background and unnaturally saturated pixels were removed from images in in the bottom row to facilitate plaque visualization. Scale bar = 2mm. The pictures shown are representative results and were repeated 3 times with similar results. For panels a-f, significant differences in phage titers on the preferred host strain were calculated via Kruskal Wallis test followed by Dunn's multiple comparisons test. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$, **** $p < 0.0001$. For panel g, significant changes in *cps2* expression were observed in Δ4 and Δ1,4 strains ($p < 0.05$ determined by z-test for each change using Dirichlet regression). Please see Source Data for all actual P values. In panels a-f, bars are drawn at the median and individual points shown. In panel g, bars represent mean and error bars SEM, n=3.
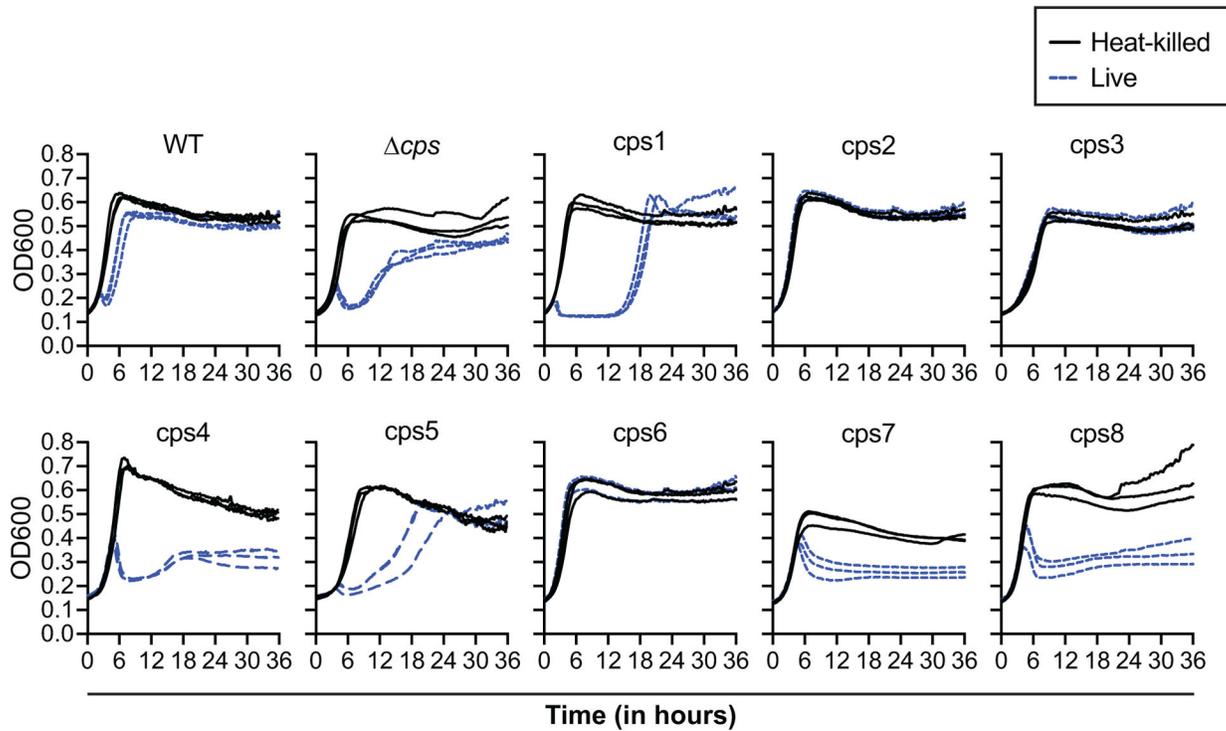
Figure 3.
Effects of ARB25 phage infection on growth of bacteria expressing different CPS. Ten strains: the wild-type (WT), the acapsular strain (*cps*), or the eight single CPS-expressing strains were infected with either live or heat-killed ARB25. Growth was monitored via optical density at 600 nm (OD_{600}) on an automated plate reading instrument as described in Methods and individual growth curves for live and heat-killed phage exposure are shown separately. Growth curves represent the mean of 3 biological replicates performed with similar results on separate days. Results from a minimum of 3, 200μl technical replicate cultures were averaged to generate each biological replicate.
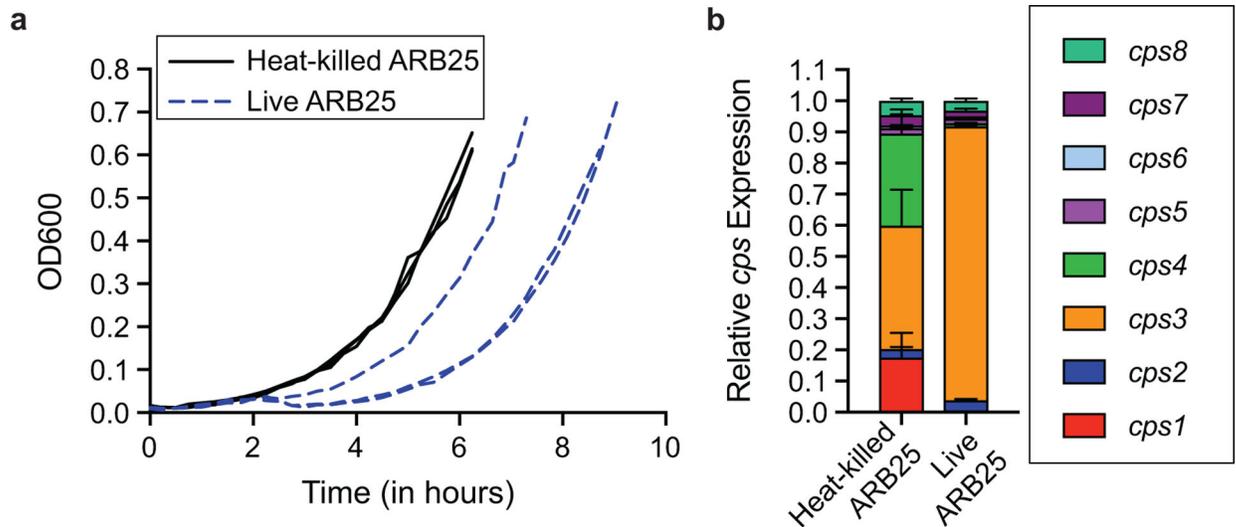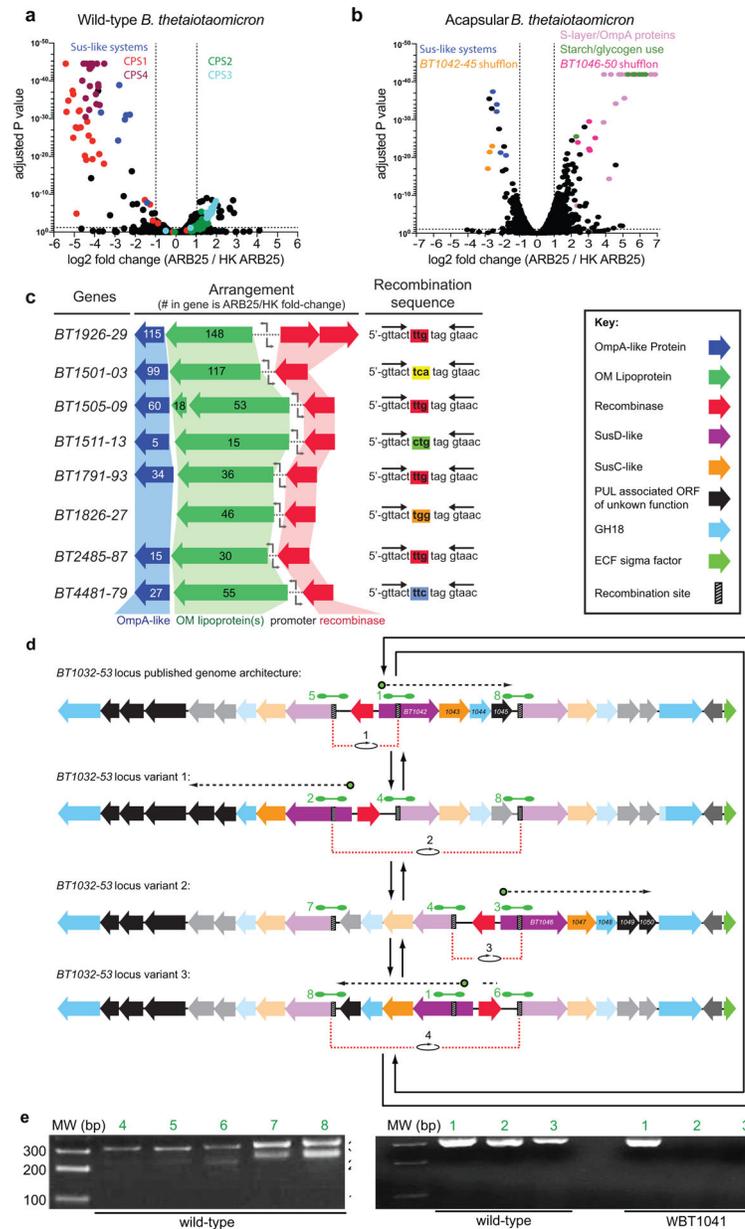
Figure 4.
ARB25 infection of wild-type *B. thetaiotaomicron* causes altered *cps* gene expression. Wild-type *B. thetaiotaomicron* was infected with live or heat-killed ARB25 at an MOI of ~1. (a) Growth was monitored by measuring OD_{600} every 15–30 minutes and individual growth curves for live and heat-killed phage exposure are shown separately (n=3). Experiment was conducted 3 times as show, plus a separate high MOI, with similar results. (b) *cps* gene transcript analysis was carried out by qPCR. The end of the growth curve in panel a represents the point at which cultures were harvested for qPCR analysis (i.e., the first observed time point where culture surpassed OD_{600} of 0.6). Significant changes in *cps1*, *cps3*, and *cps4* expression were observed between groups treated with live or heat-killed ARB25 (p < 0.01 determined by z-test for each change using Dirichlet regression; bars represent mean and error bars SEM, n=3). Please see Source Data for all actual P values. Individual replicates for high and low MOI experiments are displayed in Extended Data 6.

Figure 5.
Infection of acapsular *B. thetaiotaomicron* selects for increased expression of multiple phase-variable loci, whereas wild-type mostly alters CPS expression. (a) Wild-type *B. thetaiotaomicron* was infected with ARB25 or was alternatively exposed to heat-killed (HK) ARB25 and cultures were grown to OD_{600}=0.6–0.7. Cells were harvested and RNA-seq analysis was carried out as described in Methods (n=3 independent experiments for each treatment group). Transcript abundance was compared between live and HK treatments to generate fold change (x axis), which is plotted against the adjusted P value (EdgeR) for each gene that was generated using an exact test adapted for overdispersed data[54]. (b) Acapsular *B. thetaiotaomicron* was treated with ARB25 or HK ARB25 and fold change in transcript abundance was calculated, as described in panel a (n=3 independent experiments for each

treatment group). The dashed lines in panels a,b represent the adjusted fold-change cutoff (0.01) that was used to establish significance as described in panel A. Please see Supplemental Tables 3 and 4 for actual adjusted P values. (c) Among the genes with increased expression in post-infected acapsular *B. thetaiotaomicron*, 25 genes were part of 8 different gene clusters that encode predicted tyrosine recombinases along with outer membrane lipoproteins and OmpA-like proteins. These gene clusters are shown. The number inside the schematic for each gene represents the fold change in expression in ARB25-treated cells relative to those treated with HK ARB25. Flanking the promoters of each of these loci are pairs of imperfect, 17 nucleotide palindromic repeats. PCR analysis and amplicon sequencing of each orientation of these 8 promoters revealed expected confirmation of changes in orientation to the "ON" position in ARB25-exposed acapsular *B. thetaiotaomicron*, although we were unable to quantify the on/off ratios due to high levels of sequence similarity between the 8 loci. (d) Another chromosomal locus with signatures of phage-selected recombination was identified by RNA-seq. Specifically, 3 of 4 genes in an operon (*BT1042-BT1045*) were significantly (adjusted P value   0.01) down-regulated after exposure to phage and 5 genes in an adjacent operon (*BT1046-BT1051*) were up-regulated. The inset key defines color codes for genes in panels c-d. (e) PCR using oligonucleotides flanking direct repeats within the *BT1032-BT1053* locus (green dumbbells, panel d) were used to demonstrate locus architecture in wild type *B. thetaiotaomicron* and in a mutant lacking the tyrosine recombinase within this locus (*B. thetaiotaomicron*   BT1041). All RNAseq data is provided in Supplemental Table 3a-g. Experiment in panel e was completed once.
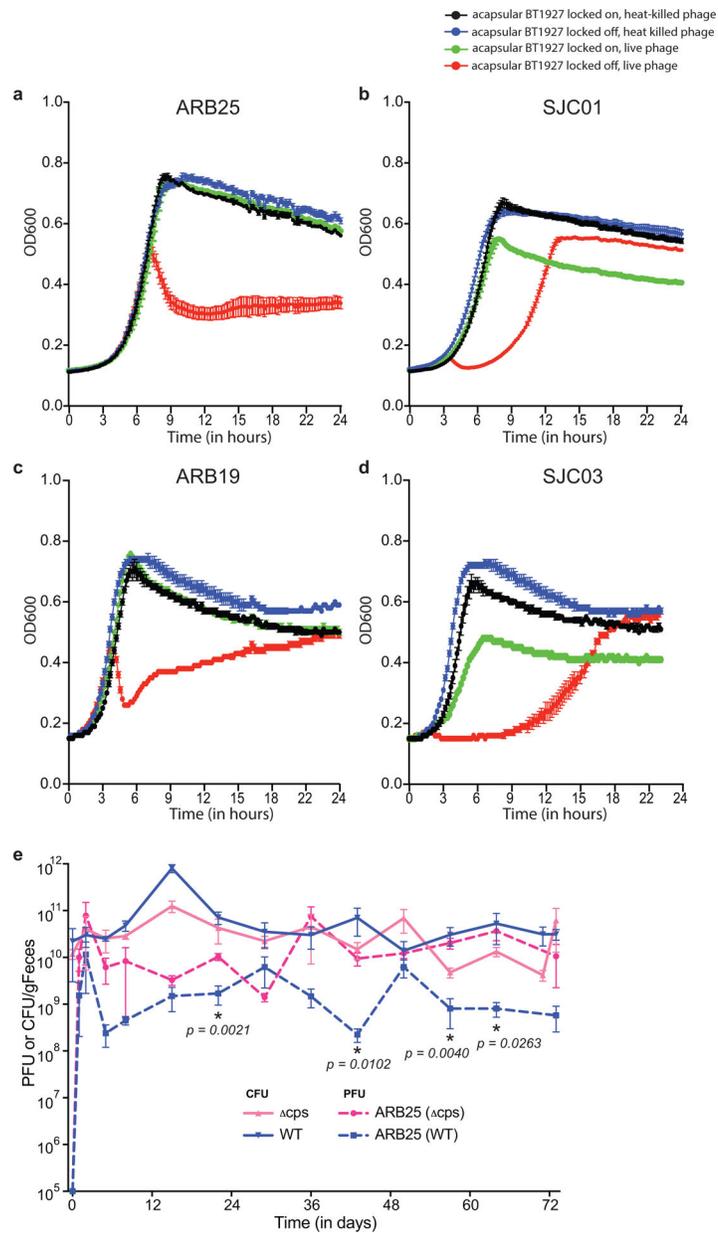
Figure 6.
Expression of the BT1927 S-layer increases *B. thetaiotaomicron* resistance to four different phages. Acapsular *B. thetaiotaomicron* S-layer 'ON' and 'OFF' mutants (*cps* BT1927-ON and *cps* BT1927-OFF, respectively) were infected with (a) ARB25, (b) SJC01 (c) ARB19, or (d) SJC03 in liquid culture. Growth was monitored via optical density at 600 nm (OD$_{600}$) as described in Methods and individual growth curves for live and heat-killed phage exposure are shown separately (n=3 for panel c; n=6 for panels a, b, d; curves represent the mean of the indicated number of replicates ± standard error of the mean). (e) CFU (solid line) or PFU (dashed line) per g Feces for each monocolonized Germ-free swiss webster mouse with either wild-type (blue) or acapsular (cps; pink) *B. thetaiotaomicron* and challenged with ARB25 phage (n=3 mice per group sex matched 2 female, 1 male per

group; actual P values are shown based on two-tailed student's t test; curves represent the mean of 3 replicates ± standard error).