Infection with novel *Bacteroides phage BV01* alters host transcriptome and bile acid metabolism in a common human gut microbe

Danielle E. Campbell[a#], Lindsey K. Ly[b,c], Jason M. Ridlon[b,c,d], Ansel Hsiao[e], Rachel J. Whitaker[a,d], Patrick H. Degnan[a,e]

[a]Department of Microbiology, University of Illinois, Urbana, Illinois, USA

[b]Division of Nutritional Sciences, University of Illinois, Urbana, Illinois, USA

[c]Department of Animal Sciences, University of Illinois, Urbana, Illinois, USA

[d]Carl R. Woese Institute for Genomic Biology, University of Illinois, Urbana, Illinois, USA

[e]Department of Microbiology and Plant Pathology, University of California, Riverside, California, USA

Running Head: Prophage BV01 alters transcriptome and bile metabolism

[#]Address correspondence to Danielle E. Campbell, campbe28@illinois.edu

Abstract Word Count: 233

Text Word Count: 6552

ABSTRACT

The bacterial genus *Bacteroides* is among the most abundant and common taxa in the human gut, yet little is known about the phages infecting the group. *Bacteroides phage BV01* (BV01) was identified as a prophage integrated on the chromosome of its host, *Bacteroides vulgatus* ATCC 8482. Phage BV01 is actively produced, and infects susceptible *B. vulgatus* hosts in the mouse gut. Infection with BV01 causes a generalized repression of the *B. vulgatus* transcriptome, downregulating 103 transcripts and upregulating only 12. Integration of BV01 disrupts the promoter sequence of a downstream gene encoding a putative tryptophan-rich sensory protein (*tspO*). Deletion of *tspO* and subsequent RNAseq analysis revealed that more than half of the differentially-regulated transcripts are shared with the BV01 lysogen, suggesting the transcriptomic response to BV01 is linked to *tspO*. Among these differentially-regulated transcripts are two encoding bile salt hydrolases. Bile acid deconjugation assays show that BV01 represses its host's ability to hydrolyze bile acids in a *tspO*-dependent manner. Analysis of 256 published healthy human gut metagenomes suggests that phage integration adjacent to *B. vulgatus*-like *tspO* genes is rare within an individual, but common among humans. Finally, this work proposes a novel phage family that includes BV01, the *Salyersviridae*, whose host range spans the *Bacteroides* and is detectable in human-associated samples. Together, these findings highlight the importance of phage-host interactions to our understanding of how gut microbes sense and interact with their environment.

IMPORTANCE

The links between human disease and the gut microbiome are numerous. Most mechanisms by which most gut microbes and their activities change and impact human health remain elusive. Phages, viruses that infect bacteria, are hypothesized to play a central role in modulating both community dynamics and functional activities. Here we have characterized an active prophage, BV01, which infects a pervasive and abundant human gut-associated species. BV01 infection alters its host's transcriptional profile including its metabolism of bile acids, molecules implicated in mediating health and disease states in the gut. This highlights that prophages and other components of the variable genome should not be overlooked in bacterial genomes because they may dramatically alter host phenotypes. Furthermore, BV01 represents a new family of phages infecting human gut symbionts, providing a foundation for future investigations of phage-host interactions in these clinically-relevant but underexplored hosts.

INTRODUCTION

The human gut is colonized by a dense and diverse microbial community comprised of bacterial, archaeal, and fungal cells, as well as the viruses that infect them. This gut microbiome is vital to human health and development, and is linked to an increasingly long list of disease states. Recent work has specifically implicated the gut phageome in disease, including inflammatory bowel disease, malnutrition, AIDS, colorectal cancers, and hypertension (1–5). Broadly, gut phages act as important modulators of bacterial community structure (4, 6, 7) and metabolism (8). Despite their apparent importance, little is known about how most gut-associated phages interact with their bacterial hosts.

The *Bacteroides* is one of the most common and abundant bacterial genera in the distal human gut. The genus is known to degrade a diversity of complex carbohydrates (9, 10), and interact with host immune cells (11, 12). Within a single human host, many *Bacteroides* species and strains coexist, competing for nutrients under changing environmental conditions caused by host diet (13), host metabolites (14), host immune system activities (15, 16), and phage predation (17). Moreover, horizontal gene transfer plays an important role in shaping the evolution and function of *Bacteroides* genomes (18, 19). How the diversity of *Bacteroides* strains in the human gut persists over time in such a dense, dynamic, and competitive environment is likely multi-fold, and perhaps afforded by their highly plastic genomes.

Phage diversity and phage-host interactions within most commensal gut-associated bacteria, including *Bacteroides* species, is underexplored. Currently, the most abundant gut-associated phages are the crAssphages (20, 21), a group of related lytic phages that infect *Bacteroides intestinalis* and potentially other species. CrAssphages demonstrate how traditional phage techniques (*e.g.,* agar overlay plaque assays) are not reliable for *Bacteroides* hosts (22), likely due to heterogeneity in capsular polysaccharide composition within isogenic cultures (23, 24). In fact, deletion of all capsular polysaccharide synthesis loci allows for the isolation of many phages on the host *Bacteroides thetaiotaomicron* VPI-5482 (17). Most phages isolated against *Bacteroides* hosts thus far exhibit an obligately lytic lifestyle (17, 22, 25, 26), despite the potentially large role of lysogeny in phage-host interactions in the gut, where at least 17% of the gut phageome is predicted to be temperate (27, 28).

Prophage–host interactions have the potential to cause complex alterations to the host phenotype by virtue of the temperate lifestyle having two distinct phases: lysogeny and lysis. Unlike most strictly lytic phages, temperate phages may horizontally transfer beneficial genes between hosts, such as antibiotic resistance genes (29) and auxiliary metabolic genes (30, 31). Some phage regulatory machinery expressed from prophages can modulate the transcription of host genes, resulting in altered phenotypes (32, 33). Integration of prophages into the host genome may also disrupt or enhance the activity of surrounding chromosomal genes (34–36). While much is known about how these prophage–host interactions contribute to virulence in pathogens (37–41), exceedingly little is about how temperate phages modulate the activities of commensals.

Here we have identified an active prophage, *Bacteroides phage BV01,* in a genetically tractable host strain, *B. vulgatus* ATCC 8482, and characterized its effects on the host's transcriptome and phenotype. Further we determine that BV01 represents a larger group of *Bacteroides*-associated prophages comprising the proposed phage family *Salyersviridae*, which are common members of the human gut phageome. This work provides the first insights to how the *Bacteroides* react to temperate phage infection, and establishes a model system for exploring complex phage–host interactions in an important human gut symbiont.

RESULTS

***Bacteroides phage BV01* is a prophage in *B. vulgatus* ATCC 8482.** *Bacteroides phage BV01* was first partially predicted with ProPhinder and deposited in the ACLAME database (42, 43). Through comparative genomics and annotation of this region, we

115 extended the predicted BV01 prophage to 58.9 Kb (NC_009614.1: 3,579,765..3,638,687),
116 comprised of 72 predicted ORFs (Fig. 1), a prediction that agrees with an observation of
117 prophage induction from *B. vulgatus* ATCC 8482 in a gnotobiotic mouse model (44).
118 BV01 encodes genes suggesting a temperate lifestyle, with a putative CI-like repressor
119 and a Cro-like anti-repressor, as well as a holing-lysin-spanin operon (Fig. 1).

120 BV01 is detectable outside of host cells in the supernatants of *in vitro* cultures as
121 a DNase-protected, dsDNA genome by both sequencing (Fig. 1) and PCR (Fig. 2A). An
122 isogenic cured lysogen (ΔBV01) strain constructed by replacement with the
123 corresponding chromosomal region of an uninfected *B. vulgatus* strain (*attB*) does not
124 release free BV01 (Fig. 1, 2A). Assembly of sequencing reads from free BV01 phage
125 DNA results in a circular contig that spans the phage attachment site (*attP*). The BV01
126 *attP* is identical to the left and right attachment sites (*attL* and *attR*), a pair of 25-bp direct
127 repeat sequences (5'-GTCTAGTTTAGTTTTTGTGTTGTAA-3'), suggesting BV01
128 enters a circular intermediate before replication.

129 To confirm that the release of DNase-protected BV01 genomes from host cells is
130 a phage-encoded and directed process, we sought to identify the phage integrase. BV01
131 encodes three genes with integrase domains (PF00589); we hypothesized that the gene
132 BVU_RS14130, adjacent to the phage attachment site, was the most likely candidate for
133 catalyzing integration and excision of BV01. A BVU_RS14130 deletion mutant (Δ*int*)
134 was constructed (Fig. 1) and its activity assayed by PCR of paired cell pellets and DNase-
135 treated culture supernatants (Fig. 2A). Phage DNA was not detected in the supernatants
136 of the Δ*int* strain. Furthermore, the Δ*int* mutant does not yield an amplicon for the
137 circularization of the BV01. Expression of the *int* gene *in trans* from a pNBU2 plasmid

complements both the circularization and release phenotypes (Fig. 2A). These results demonstrate that the integrase encoded by BVU_RS14130 is necessary for phage excision, circularization, and release from the host. Furthermore, they suggest that BV01 is an intact prophage capable of directing its own mobilization.

Despite numerous attempts, we have not identified *in vitro* inducing conditions for BV01. This has included treatments with UV light and sub-inhibitory concentrations of norfloxacin and mitomycin C (data not shown). BV01 supernatants never produced a plaque on any of 10 *B. vulgatus* isolates tested, the *B. vulgatus* cured lysogen, *B. thetaiotaomicron* VPI-5482, or *Bacteroides dorei* DSM 17855, on any of four media: TYG, BHI-HB, BHI-HM, and TYG_S. To test for new lysogenic infections, BV01 was tagged with a copy of *tetQ*, conferring tetracycline resistance, using allelic exchange. Transduction of BV01 or BV01-*tetQ* via culture supernatants or in co-culture with erythromycin resistant hosts tagged with pNBU2-*bla*-*ermG*$_b$ has never generated transductants. We conclude that BV01 is a latent prophage in culture, which is further supported by transcriptional data showing that BV01 exists in a largely repressed state (Fig. S1).

The only infectious conditions that have been identified for BV01 are in a gnotobiotic mouse model (Fig. 2B). Within a single day of gavage, BV01-*tetQ* transductants were identified on doubly-selective media from mouse pellets from 4 of the 7 mice. Over the course of the 11-day experiment transductants were eventually observed in all animals, in all cages. The average frequency of transduction ranged from 1.9×10^{-6} to 3.6×10^{-9} per animal. These results support the hypothesis that an unknown mammalian host factor is required for novel BV01 infection.

Lysogeny with BV01 alters the host transcriptome. It was hypothesized that lysogeny with a prophage such as BV01 could alter the activities of the *B. vulgatus* host. RNA sequencing (RNAseq) of the *B. vulgatus* wild-type lysogen and the isogenic cured lysogeny was performed to identify transcripts differentially regulated in response to BV01 lysogeny (Fig. 3).

Analysis of RNAseq data revealed 115 host transcripts differentially regulated in response to lysogeny with BV01 (Fig. 3A), 103 of which (89%) are up-regulated in the cured lysogen (Table S1). These transcriptional changes occur across the host genome (Fig. 1B). Functional analysis of these transcripts revealed that most function in metabolism and cellular processes and signaling (Fig. 3C). Pathway analysis using the Kyoto Encyclopedia of Genes and Genomes Pathway Database (45), however, failed to yield pathway-level differences, which is likely a reflection of the level of annotation of the *B. vulgatus* genome. Taken together, these results indicate BV01 represses a diverse array of its host's metabolic activities, suggesting it is acting through one or more transcriptional regulators.

One possible explanation for the widespread transcriptomic response to BV01 lysogeny is that a phage product directly alters the transcriptional activity of host genes. BV01 encodes two candidate genes that might act in this way: a predicted transcriptional regulator (BVU_RS14475) and a predicted sigma factor-like protein (BVU_RS14235) (Fig. S1). The transcriptional regulator encoded by BVU_RS14475 is the most highly transcribed gene in the BV01 prophage, encoding the putative CI-like repressor protein, which might directly interact with host promoters. Transcription of BVU_RS14235 is very low, so it is less likely to play a major role in the observed transcriptional response

(Fig. S1). It is also possible that the observed transcriptional response to BV01 is the result of a host response to infection. A universal stress protein (*uspA*) homolog (BVU_RS16570) is upregulated in the BV01 lysogen, though it is not clear if that is a primary or secondary effect of infection.

BV01 alters bile acid metabolism by disrupting the *tspO* promoter. Notably, integration of BV01 at the *attB* is correlated with a 23-fold down-regulation of the adjacent downstream transcript (BVU_RS14490), through an apparent disruption of the gene's promoter (Fig. 4A). A low level of expression at this gene is observed in the wild-type lysogen, perhaps a result of readthrough from phage transcripts. This gene encodes a predicted tryptophan rich sensory protein (TspO) homolog (Fig. 4B), an intramembrane protein whose endogenous ligand is unknown, but which is broadly implicated in metabolic regulation and stress response in other bacteria (46–48). Although TspO is conserved in many bacteria, archaea, and eukaryotes, it is considered an accessory protein. Indeed, not all gut-associated members of the family Bacteroidales nor the genus *Bacteroides* encode *tspO* (Fig. 4C). Within the *Bacteroides*, *tspO* is restricted to the clade including *B. vulgatus* and *B. dorei*. Among *B. vulgatus* strains TspO is highly conserved (Fig. 4D), suggesting it plays a specialized role in regulating the cellular activities unique to *B. vulgatus*.

Given TspO's important role in regulating cellular activities in other bacterial systems, it was hypothesized that it may be responsible for some of the differential regulation observed in response to prophage BV01. A *tspO* deletion mutant was constructed in the cured lysogen background (ΔBV01Δ*tspO*) and its transcriptome sequenced alongside that of the wild-type and cured lysogen strains (Fig. 3). The

207 predicted *tspO* regulon extends far beyond the differential expression observed in the
208 cured lysogen (Fig. 3A, 3B), suggesting the small amount of *tspO* transcription in the
209 BV01 lysogen exerts effects on the rest of the genome. Transcripts differentially
210 regulated between the BV01 wild-type lysogen and cured lysogen which returned to wild
211 type-like levels upon further deletion of *tspO* were classified as *tspO*-dependent
212 transcripts (Fig. 3C). Of the 115 transcripts differentially regulated in response to BV01,
213 69 (60%) are *tspO*-dependent. Consistent with TspO's role in regulating stress, many
214 *tspO*-dependent transcripts fall into the COG category for post-translation modification,
215 protein turnover, and chaperones, including several thioredoxins, peroxidases, and
216 protein chaperones. *tspO*-dependent transcripts also account for the majority of metabolic
217 genes differentially regulated in response to BV01.

218 Two *tspO*-dependent transcripts that are down-regulated in response to lysogeny
219 with BV01 encode putative bile salt hydrolases (BVU_RS13575, 5.38-fold change, $q <$
220 10^{-100}; BVU_RS20010, 6.58-fold change, $q < 10^{-200}$). It was hypothesized these
221 transcriptional differences might reflect enzyme activities. To this end, *B. vulgatus* strains
222 were grown in the presence of bile acids and the deconjugation of those bile acids
223 measured by LC/MS (Fig. 5).

224 LC/MS results show that the wild-type *B. vulgatus* lysogen does not significantly
225 deconjugate glycocholic acid (GCA) to cholic acid (CA), and may exhibit modest
226 deconjugation of taurocholic acid (TCA) (Fig. 5). This agrees with a previous study that
227 showed *B. vulgatus* ATCC 8482 can deconjugate TCA but cannot deconjugate GCA over
228 a 48 hr incubation (49). Importantly, CA is clearly detectable only in the cured lysogen
229 background, consistent with our prediction based on RNAseq data. This bile acid

deconjugation phenotype is ablated with the deletion of *tspO*, further supporting the hypothesis that *tspO* activates transcription of bile salt hydrolases, resulting in their increased enzymatic activity.

To see if *tspO*-disrupted lysogens occur in natural human gut microbiomes, read mapping from 256 healthy human gut metagenomes was performed. Starting with reads which mapped to *tspO* in the reverse orientation, read mates were checked for mapping to Phage BV01 and its relatives (Fig. 6). All samples had reads mapping to *tspO*, with an average of 0.0004% of metagenome reads mapping, indicating the corresponding population of *B. vulgatus* and *B. dorei* encoding *tspO* is relatively abundant (Fig. S2A). Incidence of *tspO* associated with BV01 or a related phage is also common; 13.3% of samples (*n*=34) contained read pairs mapping to *tspO* and an adjacent prophage. Within an individual microbiome, incidence of *tspO*-disrupted lysogens appears rare, usually comprising 3% or less of the combined *B. vulgatus* and *B. dorei* population, although for some individuals this incidence rate can be more than 10% (Fig. S2B). In addition to these patterns, carriage of *tspO*-disrupted lysogens appears stable over time, as indicated by individuals sampled at multiple time points (Table S2). Together, these data suggest that BV01 and other phages' effects on downstream phenotypes via *tspO* are likely quite common among humans, though these lysogens comprise the minority of the overall microbiome.

BV01 represents the proposed viral family *Salyersviridae*. While searching for potential new hosts for BV01, its predicted 25-bp *attB* was queried against 154 gut-associated Bacteroidales genomes. It was found that all *Paraprevotella* and *Bacteroides* genomes had at least the first 21 bases of the attachment site conserved (Fig. 6A). In 20

genomes, two copies of the *att* site were found, and all are associated with putative prophages. In all instances, the putative *attL* and *attR* sites are direct repeats, as is true for BV01. Importantly, only *B. vulgatus* lysogens encode *tspO* (Fig. 4); other prophages and *attB* sites occur in an alternative genomic context. Alignment of these putative prophage-associated *att* sites finds that the first 22 basepairs are always conserved, and an additional 3 basepairs are variable among lysogen genomes (Fig. 6B).

Integration at the same *att* site suggests these prophages are genetically related. To assess relatedness, VICTOR (50) was used to build a genome tree from all of the identified prophages (Fig. 6C) and OPTSIL clustering (51) was used to predict taxonomic groups, which we named *Salyersviridae*, *Salyersvirinae*, and *Salyersvirus* (Table 1). Taxonomic clustering also defined phage species, four of which have more than one member. Interestingly, the phage species BX01 has representatives in two distantly related hosts, *Bacteroides xylanisolvens* XB1A (phage BX01) and *B. thetaiotaomicron* 1_1_6 (phage BT01), suggesting that at least some Salyersviruses have broad host ranges. Phage BC01 was placed outside the subfamily *Salyersvirinae*, which is consistent with its considerable sequence divergence (Fig. S3). Further examination of the BC01 *attL* and *attR* sites found that the 22-bp consensus sequence is only a portion of the full 69-bp repeat flanking the BC01 prophage, which further supports placement of BC01 outside the *Salyersvirinae*. Also included in the analysis were three outgroup sequences from the known *Bacteroides*-infecting lytic phages: B124-14, B40-8, and crAssphage. Surprisingly, OPTSIL taxonomic clustering placed phages B124-14 and B40-8 in the family *Salyersviridae*. This similarity is not detectable at the level of nucleotide sequence

(Fig. S3), but is likely driven by similarities in several proteins, including homologs to the predicted lysin, excisionase, and ssDNA binding protein from BV01.

To check that the proposed *Salyersviridae* clade was comprised of active phages, we searched for evidence of activity in culture and in wastewater samples. Sequencing and PCR were used to test the activity of six additional Salyersviruses (Fig. S4). Phages BV02, BV05, and BV06 were found to be released from host cells in culture, while BV03 is likely inactivated by genomic rearrangement. We did not observe activity for BV04 or BO03 from *B. ovatus* ATCC 8483, both of which appear complete based on synteny with intact Salyersviruses. This may be the result of inactivating mutations or not being induced in the growth conditions tested (Fig. S4D).

Furthermore, we used read mapping to search for evidence of all 20 *Salyersviridae* phages in a wastewater metavirome (Fig. S5A). Reads mapped to all *Salyersviridae* genomes, however it may be the result of sequence conservation within the family, as reads often accumulate at the most conserved regions of each phage genome. Further, very little read mapping occurred in the distal portion of the inactivated prophage BV04, which resembles host chromosome more than phage sequence, indicating that virome processing removed most cellular DNA prior to sequencing (Fig. S5A). *De novo* assembly of the individual wastewater viromes finds contigs which align with high identity but imperfectly to each *Salyersviridae* phage, supporting the findings seen by read mapping, and suggesting the real diversity of the family *Salyersviridae* is far greater than what has been observed integrated in cultured host genomes so far. This analysis found that phages infecting *B. vulgatus* are more abundant than other *Salyersviridae* phages based on maximum normalized read coverages. A similar

comparison concludes that most individual temperate *Salyersviridae* phages are approximately equal in abundance to the lytic *Salyersviridae* phage B124-14 and B40-8, and approximately ten-fold less abundant than crAssphage, in these wastewater viromes (Fig. S5A). Searches for *Salyersviridae* phages in individual healthy human fecal metagenomes refine this conclusion, showing that in most human samples, *Salyersviridae* phages are at least as abundant as crAssphage, though crAssphage can reach very high abundances in a subset of individuals (Fig. S5B). Although confirmation of most *Salyersviridae* activities will require better sampling and *in vitro* testing, these results indicate the phage family is active in human-associated communities.

DISCUSSION

Here, we characterize a complex phage-host interaction between *Bacteroides phage BV01* and its host *B. vulgatus*. We first demonstrate that BV01 is an intact prophage capable of directing its own excision, and it is transducible *in vivo* in a gnotobiotic mouse model. Using a combination of genetics, RNAseq, and analytical chemistry, we show that BV01 decreases its host's ability to deconjugate bile acids by disrupting the transcription of the gene adjacent to the *attB* encoding a TspO homolog. Furthermore, we show that *tspO* disruption by phage integration is common among, but rare within, healthy human gut microbiomes, and can be mediated by BV01 or its relatives. Together, these findings elucidate a complex mechanism by which a phage alters its host's activities.

The repression of bile acid deconjugation as a consequence of BV01 integration is particularly relevant in the context of the mammalian gut. Mammals secrete conjugated

primary bile acids into the small intestine, where they reach concentrations as high as 1 mM (52); though the majority of bile acids secreted into the small intestine are readsorbed, they can still accumulate to concentrations of 0.2-1 mM in the colon (53). While bile acids are broadly capable of damaging lipid membranes, generally *Bacteroides* species are considered bile-resistant (54), the mechanism of which is unknown. Bile acid deconjugation is a common activity encoded by gut-associated microbes, though its direct benefit to those microbes is unclear. Microbial modification of the bile acid pool can be linked to beneficial changes in the human host metabolism (49, 55) and varied epithelial susceptibility to viral pathogens (56). The link between BV01 and bile acid metabolism suggests a heretofore undescribed mechanism by which gut phages might influence mammalian host phenotypes.

Here, bile acid deconjugation in *B. vulgatus* is dependent on a putative TspO. Bacterial TspOs are important for regulating metabolic switches and stress regulation in at least three diverse systems (46–48), though the mechanism of action for the protein is unknown. The crystal structure of TspO shows a periplasm-facing binding pocket distinct from the intramembrane cholesterol recognition consensus sequence, which may bind or degrade porphyrins (57–59). Both the porphyrin degrading and cholesterol transporting functions of TspO, however, have been disputed (60, 61). Despite this, it is notable that cholesterol is structurally similar to bile acids, being their biosynthetic precursor. In at least one other gut-associated microbe, *tspO* is up-regulated by bile acids, suggesting TspO may be involved in bile acid metabolism in gut microbes more broadly (62). The regulatory link described here between bile acid hydrolysis and TspO suggests a

343 hypothesis where the *B. vulgatus* TspO might be a sensor and regulator of bile acid
344 interactions.

345 Induction of BV01 from its integrated state and infection of new hosts remains
346 enigmatic. Prophage induction is canonically linked to stress-dependent pathways, as is
347 the case for lambdoid phages that respond to DNA damage via RecA-dependent cleavage
348 of the CI repressor protein (63). It is possible that prophages in *Bacteroides* hosts respond
349 to alternative stimuli, as is the case for CTnDOT, a well-studied *Bacteroides* conjugative
350 transposon, whose excision is inducible only by teteracycline (64). Neither DNA damage
351 nor antibiotics induce prophage BV01 *in vitro*, so all experiments here relied on an
352 apparently low rate of spontaneous prophage induction. Similarly, no infecting conditions
353 or susceptible hosts have been identified for BV01 *in vitro.* We demonstrate that BV01 is
354 transducible in a gnotobiotic mouse model, suggesting that an unknown mammalian host
355 factor is required for novel BV01 infection. Enigmatic infection dynamics may be the
356 result of the phase variable polysaccharide capsule, as recent work suggests heterogeneity
357 in capsular composition hinders phage infection on population-scales (17). Indeed, it has
358 long been observed that finding phages in the *Bacteroides* using traditional techniques is
359 difficult or impossible for most host strains (65, 66), making the host-first approach to
360 phage discovery used here especially appealing.

361 Finally, phage BV01 is the first representative of a broad family of phages that
362 spans an entire host genus, and includes lytic and temperate members. *Salyersviridae* is
363 common and diverse among natural human samples, but rare within individuals,
364 suggesting lysogenization may confer frequency-dependent advantages to the bacterial
365 host. The genetic context of non-*B. vulgatus Salyersviridae* lysogens remains unexplored,

366 providing ample opportunity for further discovery of novel phage-host interactions. The
367 absence of *tspO* in these other host systems may provide the ideal background for
368 studying more direct impacts of these phages on their hosts. Certainly, other interactions
369 between BV01 and its host remain to be studied, though they were overshadowed here by
370 the enormous effects of *tspO*. Future studies should also examine the role of
371 *Salyersviridae* phages on bacterial host fitness and evolution (67, 68), as these phages
372 directly impact their bacterial hosts and those interactions likely have important ripple
373 effects throughout the microbiome and on the mammalian host that remain to be
374 elucidated.
375
376 **MATERIALS AND METHODS**
377 **Strains and culture conditions.** All strains and plasmids used in the study are
378 listed in Table S3. *Escherichia coli* S17-1 *λ pir* was used for all routine recombinant
379 DNA cloning, and grown aerobically in Lysogeny Broth (LB) at 37°C. *B. vulgatus* strains
380 were cultured anaerobically in a vinyl anaerobic chamber using 70% N_2, 20% CO_2, and
381 10% H_2 gas mixture (Coy Laboratory Products, Grass Lake, MI). All *B. vulgatus* cultures
382 were grown on Difco Brain Heart Infusion (BHI) agar supplemented with 10%
383 defibrinated horse blood (BHI-HB; Quad Five, Ryegate, MT), or in tryptone-yeast
384 extract-glucose (TYG) broth (69) at 37°C. When necessary, ampicillin (100 µg/mL),
385 gentamicin (200 µg/mL), erythromycin (25 µg/mL), 5'-flourodeoxyuridine (FUdR; 20
386 µg/mL), or tetracycline (2 µg/mL) were supplemented in the media. Infection assays were
387 performed on BHI supplemented with 50 µg/mL hemin and 0.5 µg/mL menadione (BHI-
388 HM) and TYG_S (70).

389 **Genetic manipulation.** All primers used to construct genetic mutants are listed in
390 Table S3. Markerless deletion mutants in *B. vulgatus* were achieved by allelic exchange
391 using a system analogous to that developed in *B. thetaiotaomicron,* (71) and confirmed
392 by PCR and whole genome sequencing. The *tdk* gene (BVU_RS09305), encoding
393 thymidine kinase, was deleted from *B. vulgatus* ATCC 8482 by allelic exchange,
394 conferring resistance to the toxic nucleotide analog FUdR. Cloning was performed as
395 described by Degnan *et al.* (72). Briefly, the 3.5 Kb regions flanking either side of *tdk*
396 were amplified with Kapa HiFi Taq MasterMix (Kapa Biosystems, Wilmington, MA)
397 and joined by splicing overlap exchange (SOE) PCR. The SOE product was purified,
398 restriction digested and ligated into the suicide vector pKNOCK-*bla-ermG*b in *E. coli*,
399 and conjugated into *B. vulgatus*. Single recombinant merodiploids were selected for on
400 BHI-HB supplemented with gentamicin and erythromycin, and double recombinant
401 deletion mutants subsequently selected for on BHI-HB with FUdR. The counterselectable
402 suicide vector pExchange-*tdk*BV was constructed by amplifying *tdk* from *B. vulgatus* and
403 cloning the *tdk* amplicon into pKNOCK-bla-ermG$_b$ by the same methods used to clone
404 the SOE product above.
405 Subsequent deletions were accomplished similarly as described for *tdk*, except
406 using pExchange-*tdk*BV and flanking regions of ~1 Kb to create the deletion alleles
407 (*tspO, int*). For deletion of the entire BV01 provirus, an empty attachment site (*attB*) and
408 the flanking 800 bp were cloned from *B. vulgatus* VPI-4506, which has 99.9% nucleotide
409 identity to the analogous regions flanking BV01 in *B. vulgatus* ATCC 8482.
410 Complementation of the BV01 integrase (BVU_RS14130) was accomplished by
411 cloning the gene and its native promoter into the integrative plasmid pNBU2-*bla-ermG*b,

412 which has a single integration site in the *B. vulgatus* genome (*attN*;
413 NC_009614.1:3152550..3152572). This construct was conjugated into *B. vulgatus* and
414 transconjugants selected for on BHI-HB supplemented with gentamicin and erythromycin
415 as described elsewhere (72).

416 The BV01-*tetQ* strain was constructed by inserting *tetQ* from pNBU2-*bla*-*tetQ*
417 immediately downstream of the stop codon of BVU_RS14265, upstream of a predicted
418 transcriptional terminator. As was done for deletion constructs, the desired region was
419 constructed on the pExchange-*tdk*BV plasmid and moved into the wild-type *B. vulgatus*
420 strain by allelic exchange. First, a ~2 Kb region surrounding the BVU_RS14265 stop
421 codon was amplified in two pieces with SOE primers designed to insert adjacent SpeI and
422 BamHI cut sites downstream of the stop codon and ligated into pExchange-*tdk*BV. This
423 construct was confirmed by Sanger sequencing before *tetQ* and its promoter were
424 amplified from CTnDOT, and ligated into the SpeI and BamHI cut sites. Tetracycline
425 was used to select for mutants, and release of BV01-*tetQ* phages confirmed by PCR.

426 Select mutant strains were confirmed by whole genome sequencing and analyzed
427 with Breseq (73) aligned to the wild-type *B. vulgatus* ATCC 8482 genome
428 (NC_009614.1), and summarized in Table S4.

429 **Genome sequencing.** Cells were pelleted from 5 mL overnight culture in TYG by
430 centrifugation at 4,000 × *g* for 5 min at 4°C, resuspended in 0.5 mL TE buffer (10 mM
431 Tris, 1 mM EDTA), and lysed by adding sodium dodecyl sulfate (SDS) and proteinase K
432 (GoldBio, Olivette, MO) to final concentrations of 0.07% and 300 µg/mL, respectively,
433 and incubating for 2 hr at 55°C. Cellular material was removed by washing twice in an
434 equal volume of buffered phenol, phenol-chloroform-isoamyl alcohol (VWR, Radnor,

435 PA), and DNA precipitated with 100% ethanol in the presence of 0.3M sodium acetate at
436 -20°C overnight. DNA pellets were washed with 70% ethanol, dried, and resuspended in
437 TE buffer.

438 Phage DNA was prepared from overnight TYG culture supernatants collected
439 after centrifugation and concentrated by centrifugation with 30,000 MWCO Corning
440 Spin-X UF 20 Concentrators (Corning, NY) or by tangential flow filtration with a
441 Vivaflow 50R 30,000 MWCO Hydrosart filter (Sartorius, Gottingen, Germany).
442 Supernatants were treated with 200 µg/mL DNase I and 1 µg/mL RNase A for 1 hr at
443 room temperature to remove unprotected DNA and RNA. Virions were disrupted with
444 1% SDS and 1 mg/mL proteinase K for 2 hr at 55°C. DNA was further isolated using the
445 same phenol-chloroform-isoamyl alcohol extraction and ethanol precipitation procedures
446 as for cellular DNA.

447 DNA libraries were constructed with the Nextera XT Library Preparation Kit and
448 Index Kit (Illumina, San Diego, CA). DNA libraries were pooled and sequenced on both
449 the Illumina HiSeq5000 and HiSeq2500 and fastq files were generated from
450 demultiplexed reads with bcl2fastq Conversion Software (Illumina, San Diego, CA).
451 Reads were trimmed and assembled using the A5ud pipeline (74). Sequencing methods
452 and assembly data are summarized in Table S5.

453 **Genome annotation.** Annotation of cellular genomes was accomplished with a
454 custom Perl script that calls protein coding genes with Prodigal (75) and RNA coding
455 genes with tRNAscan-SE (77), Rnammer (78) and Infernal (78). Functional predictions
456 are assigned by searching against Kyoto Encyclopedia of Genes and Genomes (45),

Cluster of Orthologous Genes (80), Pfam (80), and TIGRFAM (81) databases, and subCELlular LOcalization predictor (82) is used to predict cellular localization.

For phage genomes, genes were called by Prodigal and the gene calling tool within Artemis (83). Functional predictions were made as above except with relaxed search parameters (cut_tc in hmmscan), plus using Basic Local Search Alignment Tool (84) with the Genbank virus database (85), Phyre2 (86) to identify conserved protein folds, and iVireons (87) to predict structural proteins, and manually comparing and combining results.

Integrase activity assays. Integrase activity was assayed through PCR of DNase-treated supernatant DNA. Briefly, free phage DNA was prepared as for DNA sequencing, and amplified with Kapa HiFi Taq MasterMix with primers specific to BV01 (BVU_RS14350) or spanning the circularized *attP* (Table S3). Free phage DNAs were checked for the presence of contaminating cellular DNA by amplifying the 16S rRNA gene with universal primers. Amplicons were cleaned with a Qiagen PCR Cleanup kit (Hilden, Germany) and run on an agarose gel in 0.5X Tris-borate-EDTA buffer at 70V alongside 1 Kb ladder (New England BioLabs, Ipswich, MA) or GeneRule Express DNA Ladder (Thermo Scientific, Waltham, MA) and stained with GelRed (VWR, Radnor, PA). Amplicons generated with *attP*-flanking primers were sequence confirmed by Sanger sequencing performed by ACGT, Inc (Wheeling, IL).

Gnotobiotic mice. All experiments using mice were performed using protocols approved by the University of California Riverside Institutional Animal Care and Use Committee. Germfree C57BL/6J mice were maintained in flexible plastic gnotobiotic isolators with a 12-hr light/dark cycle. Animals caged individually (n=1, female) or in

pairs of litter mates (*n*=6, males) were provided with standard, autoclaved mouse chow (5K67 LabDiet, Purina) *ad libitum*. With no *a priori* reason to expect age to influence transduction rates, animals ranged from 7 weeks to nearly 12 months old. Individually antibiotic resistance marked bacterial strains were grown individually for ~20h in TYG medium with appropriate antibiotics and frozen at -80ºC in anaerobic cryovials. Cell viability was tested by plating and viable CFU counts were used to combine equal parts of the wild-type *B. vulgatus* lysogen tagged with pNBU2-*bla-ermGb* and *B. vulgatus* BV01-*tetQ* tagged with pNBU2-*bla-cfx*. Approximately 4 x 10^6 CFUs of the combined strains were administered to each animal by oral gavage. Fecal samples were collected on days 1, 3, 7 and 11 from each animal. Fecal pellets were processed by adding 500µl of TYG+20% Glycerol to each tube and vigorously shaking in a beadbeater without beads for 1m 30s. Fecal slurries were spun down at 2,000 x g for 1 s, followed by serial dilution on selective media (BHI+Tet+Gn, BHI+Erm+Gn, BHI+Erm+Tet+Gn) to determine CFUs. Animals were sacrificed on d11 following final fecal collection.

Transcriptomic response to lysogeny. *B. vulgatus* was grown overnight in 5 mL TYG medium. Each culture was pelleted (4,000 × *g* for 5 min at 4°C), supernatant decanted, and washed in an equal volume of TYG three times. Cells were normalized to an OD_{600} of ~0.3 and used to inoculate cultures in 10 mL TYG at a final dilution of 1:1000 in biological triplicate. Cell growth was monitored and cells were harvested at an OD_{600} of ~0.4. Total RNA was prepared with a Qiagen RNeasy kit (Hilden, Germany) and treated on-column with RNase-free DNase (Qiagen, Hilden, Germany). RNA was quantitated with a Qubit 2.0 fluorometer (Thermo Fisher, Waltham, MA) and stored at -80°C.

RNA was submitted to the W. M. Keck Center for Comparative and Functional Genomics at the University of Illinois at Urbana-Champaign for quality analysis, rRNA depletion with the RiboZero Bacteria kit (Illumina, San Diego, CA), library construction with the TruSeq Stranded mRNAseq Sample Prep kit (Illumina, San Diego, CA), and sequencing on an Illumina NovaSeq 6000 with the NovaSeq S4 reagent kit. Fastq files of demultiplexed reads were prepared with the bcl2fastq v2.20 Conversion Software (Illumina, San Diego, CA).

RNAseq reads were quality filtered and trimmed with trim_galore v0.4.4 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Rockhopper (88) was used to identify differentially expressed transcripts between isogenic mutants (≥2-fold change, $q \leq 0.01$).

Bile salt deconjugation assay & LC/MS. *B. vulgatus* strains were inoculated in TYG liquid supplemented with 50 μM glycocholic acid (GCA) or 50 μM taurocholic acid (TCA) and allowed to grow for 16 or 28 hr, respectively. Grown cultures were brought to a pH 2.0-3.0 with 10 N hydrochloric acid, centrifuged for 5 min at 4,000 x *g*, and the pellets discarded. Bile acids were isolated by solid phase extraction over Sep-Pak tC18 500 mg cartridges (Waters Corp., Milford, MA). Cartridges were preconditioned by serial washes with 6 mL hexane, 3 mL acetone, 6 mL methanol, and 6 mL water (pH = 3.0). Acidified supernatants were loaded before washing with 3 mL 40% methanol. The column was allowed to dry, then bile acids eluted in 3 mL methanol. Samples were evaporated under nitrogen, resuspended in 20 μL methanol, and centrifuged before analysis by liquid chromatography-mass spectroscopy (LC/MS).

LC/MS for all samples was performed on a Waters Aquity UPLC coupled with a Waters Synapt G2-Si ESI MS. Chromatography was performed using a Waters Cortecs UPLC C18 column (1.6 µm particle size) (2.5 mm x 50 mm) with a column temperature of 40° C. Samples were injected at 1 µl. Solvent A consisted of 95% water, 5% acetonitrile, and 0.1% formic acid. Solvent B consisted of 95% acetonitrile, 5% water, and 0.1% formic acid. The initial mobile phase was 90% Solvent A, 10% Solvent B and increased linearly until the gradient reached 50% Solvent A and 50% Solvent B at 7.5 min. Solvent B was increased linearly again until it was briefly 100% at 8.0 min until returning to the initial mobile phase (90% Solvent A, 10% Solvent B) over the next 2 min. The total run was 10 min with a flow rate of 10 µL/min. MS was performed in negative ion mode. Nebulizer gas pressure was maintained at 400° C and gas flow was 800 L/hour. The capillary voltage was set at 2,000 V in negative mode. MassLynx was used to analyze chromatographs and mass spectrometry data.

Taxonomic nomenclature. The family, subfamily, and generic names were chosen to honor the microbiologist Abigail A. Salyers, who made significant contributions to the understanding of function and genetics of human gut anaerobes and the importance of their mobile genetic elements.

Wastewater collection, processing, and viromics. From the Urbana & Champaign Sanitary District Northeast Plant (Urbana, IL), 1 L of unprocessed wastewater was collected at each of three time points: May 25, 2016, June 23, 2016, and October 3, 2016.

Wastewater samples were transported on ice, immediately centrifuged at 2,500 × *g* for 10 min at 4°C and filtered through a 0.4 µm polyethersulfone filter to remove large

548 particulate and cellular matter. The sample was split into three aliquots and processed
549 three ways. One aliquot was not processed further (F). Another aliquot was filtered a
550 second time through a 0.22 μm polyethersulfone filter (DF). The last aliquot was washed
551 three times with an equal volume of chloroform (FC). All aliquots were concentrated
552 100-fold and virome DNA was isolated from each as described for genome sequencing of
553 phages.

554 DNA libraries of virome DNA were prepared using the same methods as
555 described for genome sequencing and were sequenced on an Illumina HiSeq 2500
556 sequencer with a HiSeq v4 SBS sequencing kit (Illumina, San Diego, CA) producing
557 2×160-bp paired-end reads. Fastq files of demultiplexed reads were generated with the
558 bcl2fastq v2.17.1.14 Conversion Sotfware (Illumina, San Diego, CA). Reads were
559 trimmed and quality filtered using Trimmomatic 0.38 (89) and assembled with
560 metaSPAdes v3.13.0 using default parameters (90). Sequencing and assembly data for
561 wastewater viromes is summarized in Table S5. Read mapping to phage genomes was
562 performed with bwa (91).

563 Human Microbiome Project Healthy Human Subjects Study samples were
564 downloaded with portal_client (Table S2). Read mapping was performed with bwa (91).

565 **Data availability.** Trimmed RNAseq reads from this study are deposited in the
566 NCBI SRA under PRJNA622597; sample accession numbers are SAMN14522273 for
567 the wild-type lysogen (WT), SAMN14522274 for the cured lysogen (ΔBV01), and
568 SAMN14522275 for the cured lysogen *tspO* deletion (ΔBV01Δ*tspO*). Trimmed
569 wastewater virome reads are deposited in the NCBI SRA under PRJNA622299. Ten new
570 assembled *B. vulgatus* genomes are deposited in NCBI GenBank under PRJNA622758.

571

572 **ACKNOWLEDGMENTS**

573 We thank Nadja Shoemaker and Abigail Salyers for access to an impressive
574 collection of *Bacteroides* isolates; Ken Ringwald for critical review of the manuscript;
575 Jim Imlay for insightful discussion on metabolism and stress; Alvaro Hernandez and
576 Chris Wright for DNA and RNA sequencing; Bruce Rabe for aid in wastewater
577 collection; Jonathan Mitchell for maintenance and animal care at the UCR vivarium.

578 This research was supported by initial complement funding to PHD from UIUC
579 and UCR and DC was supported by the Department of Microbiology from UIUC. LL
580 was supported by the National Science Foundation Graduate Research Fellowship.
581 Gnotobiotic mouse work and AH were supported by National Institute of General
582 Medical Sciences grant R35GM124724. RJW is supported by the Allen Foundation with
583 an Allen Distinguished Investigator Award.

584

585 **REFERENCES**

586 1. Norman JM, Handley SA, Baldridge MT, Droit L, Liu CY, Keller BC, Kambal A,
587 Monaco CL, Zhao G, Fleshner P, Stappenbeck TS, McGovern DPB, Keshavarzian A,
588 Mutlu EA, Sauk J, Gevers D, Xavier RJ, Wang D, Parkes M, Virgin HW. 2015.
589 Disease-Specific Alterations in the Enteric Virome in Inflammatory Bowel Disease.
590 Cell 160:447–460.

591 2. Reyes A, Blanton LV, Cao S, Zhao G, Manary M, Trehan I, Smith MI, Wang D,
592 Virgin HW, Rohwer F, Gordon JI. 2015. Gut DNA viromes of Malawian twins
593 discordant for severe acute malnutrition. Proc Nat Acad Sci USA 112:11941–11946.

3. Monaco CL, Gootenberg DB, Zhao G, Handley SA, Ghebremichael MS, Lim ES, Lankowski A, Baldridge MT, Wilen CB, Flagg M, Norman JM, Keller BC, Luévano JM, Wang D, Boum Y, Martin JN, Hunt PW, Bangsberg DR, Siedner MJ, Kwon DS, Virgin HW. 2016. Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome. Cell Host & Microbe 19:311–322.

4. Hannigan GD, Duhaime MB, Ruffin MT, Koumpouras CC, Schloss PD. 2018. Diagnostic Potential and Interactive Dynamics of the Colorectal Cancer Virome. mBio 9:e02248-18.

5. Han M, Yang P, Zhong C, Ning K. 2018. The Human Gut Virome in Hypertension. Front Microbiol 9.

6. Moreno-Gallego JL, Chou S-P, Di Rienzi SC, Goodrich JK, Spector TD, Bell JT, Youngblut ND, Hewson I, Reyes A, Ley RE. 2019. Virome Diversity Correlates with Intestinal Microbiome Diversity in Adult Monozygotic Twins. Cell Host & Microbe 25:261-272.e5.

7. Khan Mirzaei M, Khan MdAA, Ghosh P, Taranu ZE, Taguer M, Ru J, Chowdhury R, Kabir MdM, Deng L, Mondal D, Maurice CF. 2020. Bacteriophages Isolated from Stunted Children Can Regulate Gut Bacterial Communities in an Age-Specific Manner. Cell Host Microbe 27:199-212.e5.

8. Hsu BB, Gibson TE, Yeliseyev V, Liu Q, Lyon L, Bry L, Silver PA, Gerber GK. 2019. Dynamic Modulation of the Gut Microbiota and Metabolome by Bacteriophages in a Mouse Model. Cell Host & Microbe 25:803-814.e5.

9. Salyers AA, Vercellotti JR, West SE, Wilkins TD. 1977. Fermentation of mucin and plant polysaccharides by strains of Bacteroides from the human colon. Appl Environ Microbiol 33:319–322.

10. Kaoutari AE, Armougom F, Gordon JI, Raoult D, Henrissat B. 2013. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. Nature Rev Microbiol 11:497–504.

11. Hickey CA, Kuhn KA, Donermeyer DL, Porter NT, Jin C, Cameron EA, Jung H, Kaiko GE, Wegorzewska M, Malvin NP, Glowacki RWP, Hansson GC, Allen PM, Martens EC, Stappenbeck TS. 2015. Colitogenic Bacteroides thetaiotaomicron Antigens Access Host Immune Cells in a Sulfatase-Dependent Manner via Outer Membrane Vesicles. Cell Host & Microbe 17:672–680.

12. Shen Y, Torchia MLG, Lawson GW, Karp CL, Ashwell JD, Mazmanian SK. 2012. Outer Membrane Vesicles of a Human Commensal Mediate Immune Regulation and Disease Protection. Cell Host & Microbe 12:509–520.

13. Tuncil YE, Xiao Y, Porter NT, Reuhs BL, Martens EC, Hamaker BR. 2017. Reciprocal Prioritization to Dietary Glycans by Gut Bacteria in a Competitive Environment Promotes Stable Coexistence. mBio 8:e01068-17.

633 14. Ridlon JM, Harris SC, Bhowmik S, Kang D-J, Hylemon PB. 2016. Consequences of
634 bile salt biotransformations by intestinal bacteria. Gut Microbes 7:22–39.

635 15. Cullen TW, Schofield WB, Barry NA, Putnam EE, Rundell EA, Trent MS, Degnan
636 PH, Booth CJ, Yu H, Goodman AL. 2015. Antimicrobial peptide resistance mediates
637 resilience of prominent gut commensals during inflammation. Science 347:170–175.

638 16. Planer JD, Peng Y, Kau AL, Blanton LV, Ndao IM, Tarr PI, Warner BB, Gordon JI.
639 2016. Development of the gut microbiota and mucosal IgA responses in twins and
640 gnotobiotic mice. Nature 534:263–266.

641 17. Porter NT, Hryckowian AJ, Merrill BD, Gardner JO, Singh S, Sonnenburg JL,
642 Martens EC. 2019. Multiple phase-variable mechanisms, including capsular
643 polysaccharides, modify bacteriophage susceptibility in Bacteroides
644 thetaiotaomicron. bioRxiv 521070.

645 18. Coyne MJ, Zitomersky NL, McGuire AM, Earl AM, Comstock LE. 2014. Evidence
646 of Extensive DNA Transfer between Bacteroidales Species within the Human Gut.
647 mBio 5:e01305-14.

648 19. Lange A, Beier S, Steimle A, Autenrieth IB, Huson DH, Frick J-S. 2016. Extensive
649 Mobilome-Driven Genome Diversification in Mouse Gut-Associated Bacteroides
650 vulgatus mpk. Genome Biol Evol 8:1197–1207.

651 20. Yutin N, Makarova KS, Gussow AB, Krupovic M, Segall A, Edwards RA, Koonin
652 EV. 2018. Discovery of an expansive bacteriophage family that includes the most
653 abundant viruses from the human gut. Nat Microbiol 3:38–46.

654 21. Dutilh BE, Cassman N, McNair K, Sanchez SE, Silva GGZ, Boling L, Barr JJ, Speth
655 DR, Seguritan V, Aziz RK, Felts B, Dinsdale EA, Mokili JL, Edwards RA. 2014. A
656 highly abundant bacteriophage discovered in the unknown sequences of human faecal
657 metagenomes. Nat Commun 5:1–11.

658 22. Shkoporov AN, Khokhlova EV, Fitzgerald CB, Stockdale SR, Draper LA, Ross RP,
659 Hill C. 2018. ΦCrAss001 represents the most abundant bacteriophage family in the
660 human gut and infects Bacteroides intestinalis. Nat Commun 9:1–8.

661 23. Krinos CM, Coyne MJ, Weinacht KG, Tzianabos AO, Kasper DL, Comstock LE.
662 2001. Extensive surface diversity of a commensal microorganism by multiple DNA
663 inversions. Nature 414:555–558.

664 24. Patrick S, Blakely GW, Houston S, Moore J, Abratt VR, Bertalan M, Cerdeño-
665 Tárraga AM, Quail MA, Corton N, Corton C, Bignell A, Barron A, Clark L, Bentley
666 SD, Parkhill J. 2010. Twenty-eight divergent polysaccharide loci specifying within-
667 and amongst-strain capsule diversity in three strains of Bacteroides fragilis.
668 Microbiology, 156:3255–3269.

669 25. Tartera C, Jofre J. 1987. Bacteriophages active against Bacteroides fragilis in
670 sewage-polluted waters. Appl Environ Microbiol 53:1632–1637.

671 26. Jofre J, Blanch AR, Lucena F, Muniesa M. 2014. Bacteriophages infecting
672 Bacteroides as a marker for microbial source tracking. Water Research 55:1–11.

673 27. Ogilvie LA, Bowler LD, Caplin J, Dedi C, Diston D, Cheek E, Taylor H, Ebdon JE,
674 Jones BV. 2013. Genome signature-based dissection of human gut metagenomes to
675 extract subliminal viral sequences. Nat Commun 4:1–16.

676 28. Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, Lewis JD, Bushman FD.
677 2011. The human gut virome: Inter-individual variation and dynamic response to
678 diet. Genome Res 21:1616–1625.

679 29. Abeles SR, Ly M, Santiago-Rodriguez TM, Pride DT. 2015. Effects of Long Term
680 Antibiotic Therapy on Human Oral and Fecal Viromes. PLoS One 10:e0134941.

681 30. Sullivan MB, Coleman ML, Weigele P, Rohwer F, Chisholm SW. 2005. Three
682 Prochlorococcus Cyanophage Genomes: Signature Features and Ecological
683 Interpretations. PLoS Biology 3:e144.

684 31. Anantharaman K, Duhaime MB, Breier JA, Wendt KA, Toner BM, Dick GJ. 2014.
685 Sulfur Oxidation Genes in Diverse Deep-Sea Viruses. Science 344:757–760.

686 32. Berger P, Kouzel IU, Berger M, Haarmann N, Dobrindt U, Koudelka GB, Mellmann
687 A. 2019. Carriage of Shiga toxin phage profoundly affects Escherichia coli gene
688 expression and carbon source utilization. BMC Genomics 20:504.

689 33. Hernandez-Doria JD, Sperandio V. 2018. Bacteriophage Transcription Factor Cro
690 Regulates Virulence Gene Expression in Enterohemorrhagic Escherichia coli. Cell
691 Host & Microbe 23:607-617.e6.

692 34. Rabinovich L, Sigal N, Borovok I, Nir-Paz R, Herskovits AA. 2012. Prophage
693 Excision Activates Listeria Competence Genes that Promote Phagosomal Escape and
694 Virulence. Cell 150:792–802.

695 35. Carey JN, Mettert EL, Fishman-Engel DR, Roggiani M, Kiley PJ, Goulian M. 2019.
696 Phage integration alters the respiratory strategy of its host. eLife 8:e49081.

697 36. Chen Y-Y, Wang J-T, Lin T-L, Gong Y-N, Li T-H, Huang Y-Y, Hsieh Y-C. 2019.
698 Prophage excision in streptococcus pneumoniaeserotype 19A ST320 promote
699 colonization: Insight into its evolution from the ancestral clone Taiwan 19F-14
700 (ST236). Frontiers in Microbiology 10:205.

701 37. Nakayama K, Kanaya S, Ohnishi M, Terawaki Y, Hayashi T. 1999. The complete
702 nucleotide sequence of φCTX, a cytotoxin-converting phage of Pseudomonas
703 aeruginosa: implications for phage evolution and horizontal gene transfer via
704 bacteriophages. Mol Microbiol 31:399–419.

705 38. Winter SE, Thiennimitr P, Winter MG, Butler BP, Huseby DL, Crawford RW,
706 Russell JM, Bevins CL, Adams LG, Tsolis RM, Roth JR, Bäumler AJ. 2010. Gut
707 inflammation provides a respiratory electron acceptor for Salmonella. Nature
708 467:426–429.

709 39. Bensing BA, Siboo IR, Sullam PM. 2001. Proteins PblA and PblB of Streptococcus
710 mitis, Which Promote Binding to Human Platelets, Are Encoded within a Lysogenic
711 Bacteriophage. Infection and Immunity 69:6186–6192.

712 40. Jermyn WS, Boyd EF. 2002. Characterization of a novel *Vibrio* pathogenicity island
713 (VPI-2) encoding neuraminidase (*nanH*) among toxigenic *Vibrio cholerae* isolates.
714 Microbiology, 148:3681–3693.

715 41. Sekulovic O, Fortier L-C. 2015. Global Transcriptional Response of *Clostridium*
716 *difficile* Carrying the ϕCD38-2 Prophage. Appl Environ Microbiol 81:1364–1374.

717 42. Leplae R, Lima-Mendez G, Toussaint A. 2010. ACLAME: A CLAssification of
718 Mobile genetic Elements, update 2010. Nucleic Acids Res 38:D57–D61.

719 43. Lima-Mendez G, Van Helden J, Toussaint A, Leplae R. 2008. Prophinder: a
720 computational tool for prophage prediction in prokaryotic genomes. Bioinformatics
721 24:863–865.

722 44. Reyes A, Wu M, McNulty NP, Rohwer FL, Gordon JI. 2013. Gnotobiotic mouse
723 model of phage–bacterial host dynamics in the human gut. Proc Nat Acad Sci USA
724 110:20236–20241.

725 45. Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes.
726 Nucleic Acids Res 28:27–30.

727 46. Yeliseev AA, Kaplan S. 1999. A Novel Mechanism for the Regulation of
728 Photosynthesis Gene Expression by the TspO Outer Membrane Protein of
729 *Rhodobacter sphaeroides* 2.4.1. J Biol Chem 274:21234–21243.

730 47. Davey ME, Bruijn FJ de. 2000. A Homologue of the Tryptophan-Rich Sensory
731 Protein TspO and FixL Regulate a Novel Nutrient Deprivation-Induced
732 *Sinorhizobium meliloti* Locus. Appl Environ Microbiol 66:5353–5359.

733 48. Balsemão-Pires E, Jaillais Y, Olson BJ, Andrade LR, Umen JG, Chory J, Sachetto-
734 Martins G. 2011. The *Arabidopsis* translocator protein (AtTSPO) is regulated at
735 multiple levels in response to salt stress and perturbations in tetrapyrrole metabolism.
736 BMC Plant Biol 11:108.

737 49. Yao L, Seaton SC, Ndousse-Fetter S, Adhikari AA, DiBenedetto N, Mina AI, Banks
738 AS, Bry L, Devlin AS. 2018. A selective gut bacterial bile salt hydrolase alters host
739 metabolism. eLife.

740 50. Meier-Kolthoff JP, Göker M. 2017. VICTOR: genome-based phylogeny and
741 classification of prokaryotic viruses. Bioinformatics 33:3396–3404.

742 51. Göker M, García-Blázquez G, Voglmayr H, Tellería MT, Martín MP. 2009.
743 Molecular Taxonomy of Phytopathogenic Fungi: A Case Study in Peronospora. PLoS
744 One 4: e6319.

745 52. Northfield TC, McColl I. 1973. Postprandial concentrations of free and conjugated
746 bile acids down the length of the normal human small intestine. Gut 14:513–518.

747 53. Hamilton JP, Xie G, Raufman J-P, Hogan S, Griffin TL, Packard CA, Chatfield DA,
748 Hagey LR, Steinbach JH, Hofmann AF. 2007. Human cecal bile acids: concentration
749 and spectrum. Am J Phys-Gastr L 293:G256–G263.

750 54. Citron DM, Baron EJ, Finegold SM, Goldstein EJ. 1990. Short prereduced
751 anaerobically sterilized (PRAS) biochemical scheme for identification of clinical
752 isolates of bile-resistant Bacteroides species. J Clin Microbiol 28:2220–2223.

753 55. Joyce SA, MacSharry J, Casey PG, Kinsella M, Murphy EF, Shanahan F, Hill C,
754 Gahan CGM. 2014. Regulation of host weight gain and lipid metabolism by bacterial
755 bile acid modification in the gut. Proc Nat Acad Sci USA 111:7421–7426.

756 56. Grau KR, Zhu S, Peterson ST, Helm EW, Philip D, Phillips M, Hernandez A, Turula
757 H, Frasse P, Graziano VR, Wilen CB, Wobus CE, Baldridge MT, Karst SM. 2020.
758 The intestinal regionalization of acute norovirus infection is regulated by the
759 microbiota via bile acid-mediated priming of type III interferon. Nature Microbiol
760 5:84–92.

761 57. Li F, Liu J, Zheng Y, Garavito RM, Ferguson-Miller S. 2015. Crystal structures of
762 translocator protein (TSPO) and mutant mimic of a human polymorphism. Science
763 347:555–558.

764 58. Guo Y, Kalathur RC, Liu Q, Kloss B, Bruni R, Ginter C, Kloppmann E, Rost B,
765 Hendrickson WA. 2015. Structure and activity of tryptophan-rich TSPO proteins.
766 Science 347:551–555.

767 59. Zeno S, Veenman L, Katz Y, Bode J, Gavish M, Zaaroor M. 2012. The 18 kDa
768 mitochondrial translocator protein (TSPO) prevents accumulation of protoporphyrin
769 IX. Involvement of reactive oxygen species (ROS). Curr Mol Med 12:494–501.

770 60. Li F, Liu J, Garavito RM, Ferguson-Miller S. 2015. Evolving understanding of
771 translocator protein 18kDa (TSPO). Pharmacol Res 99:404–409.

772 61. Rone MB, Midzak AS, Issop L, Rammouz G, Jagannathan S, Fan J, Ye X, Blonder J,
773 Veenstra T, Papadopoulos V. 2012. Identification of a Dynamic Mitochondrial
774 Protein Complex Driving Cholesterol Import, Trafficking, and Metabolism to Steroid
775 Hormones. Mol Endocrinol 26:1868–1882.

776 62. Devendran S, Shrestha R, Alves JMP, Wolf PG, Ly L, Hernandez AG, Méndez-
777 García C, Inboden A, Wiley J, Paul O, Allen A, Springer E, Wright CL, Fields CJ,
778 Daniel SL, Ridlon JM. 2019. Clostridium scindens ATCC 35704: Integration of
779 Nutritional Requirements, the Complete Genome Sequence, and Global
780 Transcriptional Responses to Bile Acids. Appl Environ Microbiol 85.

781 63. Roberts JW, Roberts CW, Craig NL. 1978. *Escherichia coli recA* gene product
782 inactivates phage lambda repressor. Proc Nat Acad Sci USA 75:4714–4718.

783 64. Moon K, Shoemaker NB, Gardner JF, Salyers AA. 2005. Regulation of Excision
784 Genes of the *Bacteroides* Conjugative Transposon CTnDOT. Journal of Bacteriology
785 187:5732–5741.

786 65. Puig M, Gironés R. 1999. Genomic structure of phage B40-8 of *Bacteroides fragilis*.
787 Microbiol, 145:1661–1670.

788 66. Salyers AA, Bonheyo G, Shoemaker NB. 2000. Starting a New Genetic System:
789 Lessons from *Bacteroides*. Methods 20:35–46.

790 67. Oh J-H, Lin XB, Zhang S, Tollenaar SL, Özçam M, Dunphy C, Walter J, Pijkeren J-P
791 van. 2019. Prophages in *Lactobacillus reuteri* Are Associated with Fitness Trade-
792 Offs but Can Increase Competitiveness in the Gut Ecosystem. Appl Environ
793 Microbiol 86.

794 68. Zhao S, Lieberman TD, Poyet M, Kauffman KM, Gibbons SM, Groussin M, Xavier
795 RJ, Alm EJ. 2019. Adaptive Evolution within Gut Microbiomes of Healthy People.
796 Cell Host Microbe 25:656-667.e8.

797 69. Holdeman LV, Cato EP, Moore WEC. 1977. Anaerobic Laboratory Manual, 4th
798 Edition. Anaerobe Laboratory, Virginia Polytechnic Institute and State University,
799 Blacksburg.

800 70. Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, Knight R,
801 Gordon JI. 2009. Identifying Genetic Determinants Needed to Establish a Human Gut
802 Symbiont in Its Habitat. Cell Host & Microbe 6:279–289.

803 71. Koropatkin NM, Martens EC, Gordon JI, Smith TJ. 2008. Starch Catabolism by a
804 Prominent Human Gut Symbiont Is Directed by the Recognition of Amylose Helices.
805 Structure 16:1105–1115.

806 72. Degnan PH, Taga ME, Goodman AL. 2014. Vitamin B12 as a Modulator of Gut
807 Microbial Ecology. Cell Metab 20:769–778.

808 73. Deatherage DE, Barrick JE. 2014. Identification of mutations in laboratory-evolved
809 microbes from next-generation sequencing data using breseq. Methods Mol Biol
810 1151:165–188.

811 74. Riley AB, Kim D, Hansen AK. 2017. Genome Sequence of "*Candidatus Carsonella*
812 *ruddii*" Strain BC, a Nutritional Endosymbiont of Bactericera cockerelli. Genome
813 Announc 5:e00236-17.

814 75. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010.
815 Prodigal: prokaryotic gene recognition and translation initiation site identification.
816 BMC Bioinformatics 11:119.

817 76. Lowe TM, Eddy SR. 1997. tRNAscan-SE: A Program for Improved Detection of
818 Transfer RNA Genes in Genomic Sequence. Nucleic Acids Res 25:955–964.

819 77. Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW. 2007.
820 RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids
821 Res 35:3100–3108.

822 78. Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches.
823 Bioinformatics 29:2933–2935.

824 79. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov
825 DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov
826 AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. 2003. The COG database: an updated
827 version includes eukaryotes. BMC Bioinformatics 4:41.

828 80. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A,
829 Marshall M, Moxon S, Sonnhammer ELL, Studholme DJ, Yeats C, Eddy SR. 2004.
830 The Pfam protein families database. Nucleic Acids Res 32:D138–D141.

831 81. Haft DH, Selengut JD, White O. 2003. The TIGRFAMs database of protein families.
832 Nucleic Acids Res 31:371–373.

833 82. Yu C-S, Cheng C-W, Su W-C, Chang K-C, Huang S-W, Hwang J-K, Lu C-H. 2014.
834 CELLO2GO: A Web Server for Protein subCELlular LOcalization Prediction with
835 Functional Gene Ontology Annotation. PLoS ONE 9:e99368.

836 83. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. 2012. Artemis: an
837 integrated platform for visualization and analysis of high-throughput sequence-based
838 experimental data. Bioinformatics 28:464–469.

839 84. Mount DW. 2007. Using the Basic Local Alignment Search Tool (BLAST). Cold
840 Spring Harb Protoc 2007:pdb.top17.

841 85. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2009. GenBank.
842 Nucleic Acids Res 37:D26–D31.

843 86. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web
844 portal for protein modeling, prediction and analysis. Nat Protoc 10:845–858.

845 87. Seguritan V, Jr NA, Arnoult M, Raymond A, Lorimer D, Jr ABB, Salamon P, Segall
846 AM. 2012. Artificial Neural Networks Trained to Detect Viral and Phage Structural
847 Proteins. PLoS Comput Biol 8:e1002657.

848 88. McClure R, Balasubramanian D, Sun Y, Bobrovskyy M, Sumby P, Genco CA,
849 Vanderpool CK, Tjaden B. 2013. Computational analysis of bacterial RNA-Seq data.
850 Nucleic Acids Res 41:e140–e140.

851 89. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina
852 sequence data. Bioinformatics 30:2114–2120.

853 90. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new
854 versatile metagenomic assembler. Genome Res 27:824–834.

855 91. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler
856 transform. Bioinformatics 25:1754–1760.

857 92. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and
858 high throughput. Nucleic Acids Res 32:1792–1797.

859 93. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – Approximately Maximum-
860 Likelihood Trees for Large Alignments. PLOS ONE 5:e9490.

861 94. Noguchi H, Taniguchi T, Itoh T. 2008. MetaGeneAnnotator: Detecting Species-
862 Specific Patterns of Ribosomal Binding Site for Precise Gene Prediction in
863 Anonymous Prokaryotic and Phage Genomes. DNA Res 15:387–396.

864 95. Roux S, Enault F, Hurwitz BL, Sullivan MB. 2015. VirSorter: mining viral signal
865 from microbial genomic data. PeerJ 3:e985.

866 96. Simon R, Priefer U, Pühler A. 1983. A Broad Host Range Mobilization System for In
867 Vivo Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria. Nat
868 Biotechnol 1:784–791.

869 97. Eggerth AH, Gagnon BH. 1933. The *Bacteroides* of Human Feces. Journal of
870 Bacteriology 25:389–413.

871 98. Johnson JL. 1978. Taxonomy of the *Bacteroides*. International Journal of Systematic
872 and Evolutionary Microbiology, 28:245–256.

873 99. Whittle G, Hund BD, Shoemaker NB, Salyers AA. 2001. Characterization of the 13-
874 Kilobase *ermF* Region of the *Bacteroides* Conjugative Transposon CTnDOT. Appl
875 Environ Microbiol 67:3488–3495.

876 100. Shoemaker NB, Vlamakis H, Hayes K, Salyers AA. 2001. Evidence for Extensive
877 Resistance Gene Transfer amongBacteroides spp. and among Bacteroides and Other
878 Genera in the Human Colon. Appl Environ Microbiol 67:561–568.

879 101. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh
880 PJ, Fierer N, Knight R. 2011. Global patterns of 16S rRNA diversity at a depth of
881 millions of sequences per sample. PNAS 108:4516–4522.

882

883 **FIGURE LEGENDS**

884 **Figure 1. DNA sequencing reads mapped to the BV01 prophage region.** Reads from
885 shotgun sequencing of *B. vulgatus* genomic DNA (WT) or isolated phage DNA (Free
886 BV01) were normalized to the total number of reads after trimming, and represented as a
887 coverage curve. A cured lysogen (ΔBV01) and integrase deletion mutant (Δ*int*) of *B.*
888 *vulgatus* were confirmed by shotgun sequencing of genomic DNA. The discrete coverage
889 peak at position ~3,588,000 nt from ΔBV01 is attributed to a homologous sequence
890 elsewhere on the *B. vulgatus* chromosome. Putative functions of BV01 genes are
891 indicated by the colors in the legend.

892

Figure 2. Prophage BV01 is an intact prophage. (A) Excision and circularization activities of the BV01 integrase are confirmed by PCR. The presence of phage DNA was detected by amplification of a phage marker gene (BVU_RS14350). Amplification across the phage attachment site (*attP*) indicates circularization of the BV01 genome; *attP* amplicons from the integrase complement strain (Δ*int* + pNBU2-*int*) are ~1.2 Kb shorter than wild-type amplicons due to deletion of the integrase gene. Supernatant fractions were treated with DNase, eliminating all contaminating host genomic DNA, as demonstrated by the amplification of a host marker gene (16S rRNA). Note that despite apparent size shift of BVU_RS14350 amplicons from the pellets and supernatants, Sanger sequencing validated that the products are in fact identical. PCR amplicons were visualized by agarose gel electrophoresis alongside NEB 1 Kb DNA ladder (BVU_RS14350, *attP*) or GeneRuler Express DNA ladder (16S rRNA); ladder band sizes shown in Kb. (B) BV01 can transduce uninfected hosts in a gnotobiotic mouse. Germ-free mice (n=7) were gavaged with an equal mixture of a BV01-*tetQ* lysogen and an erythromycin-tagged cured lysogen (Day 0). Recipient, donor, and transductant cells were identified by plating on Brain Heart Infusion (BHI) media with antibiotic selection: erythromycin (Erm) or tetracycline (Tet).

Figure 3. Differential regulation of the host transcriptome in response to BV01 lysogeny. (A) Count of differentially regulated transcripts as compared to the wild-type *B. vulgatus* lysogen (fold change ≥ 2, q-value ≤ 0.01). (B) Chromosomal localization of the differentially expressed genes. Each dot represents a differentially expressed transcript on a $\log_2$ scale; genes below the 2-fold change cutoff (yellow) and within the

916 BV01 prophage (blue) not shown. Positive fold change values correspond to increased
917 transcription in the second background listed. (C) General functional assignment of genes
918 differentially expressed between wild-type and cured lysogen strains was accomplished
919 using the Clusters of Orthologous Groups (COGs). Transcripts which are not
920 differentially expressed in other strain comparisons are shown in black. *tspO*-dependent
921 transcripts are marked on right (•). Positive fold change values correspond to increased
922 transcription in the second background listed. Letters correspond to COG categories: cell
923 wall/membrane/envelope biogenesis (M), post-translation modification, protein turnover,
924 and chaperones (O), signal transduction mechanisms (T), transcription (K), energy
925 production and conversion (C), amino acid transport and metabolism (E), carbohydrate
926 transport and metabolism (G), coenzyme transport and metabolism (H), lipid transport
927 and metabolism (I), inorganic ion transport and metabolism (P), secondary metabolite
928 biosynthesis, transport, and catabolism (Q), general function prediction only (R), function
929 unknown (S).
930
931 **Figure 4. BV01 integration disrupts transcription of *tspO*.** (A) Transcriptional activity
932 of the *tspO* gene region as it exists in the cured lysogen background. RNAseq reads from
933 wild-type (WT) and cured lysogen (ΔBV01) *B. vulgatus* were mapped to the region, and
934 coverage was normalized to the total number of reads mapping to the genome. The
935 average normalized read coverage for each genome is displayed as the y-axis maximum
936 (grey line). Maximum read coverage for the region is displayed as the grey dashed line.
937 (B) Amino acid alignment of *B. vulgatus* TspO with known TspO sequences was
938 generated with MUSCLE (92). Identical and similar residues are colored blue and gray,

939 respectively. Shown are TspO protein sequences from *B. vulgatus* (WP_005843416.1),
940 *Bacillus cereus* (GCF80909.1), *Rhodobacter sphaeroides* (AAF24291.1), *Sinorhizobium*
941 *meliloti* (AAF01195.1), human (NP_001243460.1), and rat (NP_036647.1). Secondary
942 structures, cholesterol recognition/interaction amino acid consensus (CRAC) sequence,
943 and critical residues (•) from *R. sphaeroides* TspO crystal structure are shown (57). (C)
944 The search for TspO homologs in the family *Bacteroidales* was accomplished with a
945 BLAST-based approach, using the *Bacillus cereus* copy of TspO (GCF80909.1) as a
946 query against a database of 154 gut-associated *Bacteroidales* genomes, 122 of which are
947 from the genus *Bacteroides*. Genome counts are indicated within categories. (D) Gene
948 tree estimated from TspO sequences across the *Bacteroidales*. All *B. vulgatus* and *B.*
949 *dorei* genomes included in the search encode *tspO*. Clade for *B. vulgatus* TspO sequences
950 is displayed as a polytomy; all *B. vulgatus* TspO sequences are at least 98% identical to
951 each other. Numbers above branches represent bootstrap values; only bootstraps over 50
952 shown. The gene tree was estimated using FastTree (93).
953
954 **Figure 5. BV01 alters host interactions with bile acids in a *tspO*-dependent manner.**
955 (A) Representative LC/MS traces showing *B. vulgatus* deconjugates taurocholic acid
956 (TCA) to cholic acid (CA) in the cured lysogen background (ΔBV01), but little or no
957 activity CA is detectable in the wild-type (WT) or cured lysogen *tspO* deletion
958 (ΔBV01Δ*tspO*) backgrounds. *B. vulgatus* cultures were incubated with 50 μM TCA for
959 16 hr prior to bile acid extraction. (B) Representative LC/MS traces showing *B. vulgatus*
960 deconjugates glycocholic acid (GCA) to CA in the ΔBV01 background, but not in the
961 WT or ΔBV01Δ*tspO* backgrounds. *B. vulgatus* cultures were incubated with 50 μM GCA

962 for 48 hr prior to bile acid extraction. Nordeoxycholic acid (norDCA) was added to a
963 final concentration of 15 μM as an internal standard after incubation. Peaks labeled for
964 their metabolites based on m/z; TCA = 514.29, GCA = 464.30, CA = 407.28, norDCA =
965 377.27.
966
967 **Figure 6. *Salyersviridae* occur throughout the *Bacteroides* genus.** (A) *Bacteroides*
968 phylogeny and occurrence of *Salyersviridae att* site. All duplications of the *att* site are
969 associated with a putative integrated prophage. Host phylogeny estimated by maximum
970 likelihood from concatenated alignment of 13 core genes. (B) Consensus *att* site for
971 *Salyersvirinae*. The *attP* is duplicated upon integration of a *Salyersvirinae* prophage,
972 resulting in direct repeats. Image made with the WebLogo online tool. (C) Phylogenomic
973 Genome-BLAST Distance Phylogeny implemented with the VCTOR online tool (50)
974 using amino acid data from all phage ORFs. For consistency, all phage genomes were
975 annotated with MetaGeneAnnotator (94) implemented via VirSorter (95). Support values
976 above branches are GBDP pseudo-bootstrap values from 100 replications. Family (F),
977 subfamily (Subf), genus (G), and species (Sp) assigned by OPTSIL clustering (51) (Table
978 1). Each leaf of the tree represents a unique phage species, except where indicated by
979 colored boxes. Active prophages confirmed by sequencing and/or PCR indicated with
980 "•"; prophages confirmed to have been inactivated by genome rearrangement indicated
981 with "x"; prophages which were tested for activity with inconclusive results indicated
982 with "*" (Fig. S4).
983
984 **TABLES**

985 **Table 1. *Bacteroides* phage taxonomy as determined by whole genome clustering.**

Phage	Host	Family	Subfamily	Genus	Species
BV01	*B. vulgatus* ATCC 8482	*Salyersviridae*	*Salyersvirinae*	*Salyersvirus*	*BV01*
BV02	*B. vulgatus* VPI-4245	*Salyersviridae*	*Salyersvirinae*	*Salyersvirus*	*BV01*
BV03	*B. vulgatus* VPI-2365	*Salyersviridae*	*Salyersvirinae*	*Salyersvirus*	*BV03*
BV04	*B. vulgatus* VPI-6186	*Salyersviridae*	*Salyersvirinae*	*Salyersvirus*	*BV04*
BV05	*B. vulgatus* VPI-5710	*Salyersviridae*	*Salyersvirinae*	*Salyersvirus*	*BV05*
BV06	*B. vulgatus* DH4096S	*Salyersviridae*	*Salyersvirinae*	*Salyersvirus*	*BV06*
BF01	*B. finegoldii* CL09T03C10	*Salyersviridae*	*Salyersvirinae*	*Salyersvirus*	*BF01*
BT01	*Bacteroides* sp. 1_1_6	*Salyersviridae*	*Salyersvirinae*	*Salyersvirus*	*BX01*
BO01	*B. ovatus* ATCC 8483	*Salyersviridae*	*Salyersvirinae*	*Salyersvirus*	*BO01*
BO02	*B. ovatus* CL02T12C04	*Salyersviridae*	*Salyersvirinae*	*Salyersvirus*	*BO02*
BO03	*Bacteroides*	*Salyersviridae*	*Salyersvirinae*	*Salyersvirus*	*BO03*

	sp. 2_2_4				
BO04	*Bacteroides* sp. D2	*Salyersviridae*	*Salyersvirinae*	*Salyersvirus*	*BO04*
BX01	*B. xylanisolvens* XB1A	*Salyersviridae*	*Salyersvirinae*	*Salyersvirus*	*BX01*
BX02	*Bacteroides* sp. D1	*Salyersviridae*	*Salyersvirinae*	*Salyersvirus*	*BX02*
BX03	*Bacteroides* sp. 2_1_22	*Salyersviridae*	*Salyersvirinae*	*Salyersvirus*	*BX02*
BX04	*B. ovatus* SD CC 2a	*Salyersviridae*	*Salyersvirinae*	*Salyersvirus*	*BX02*
BX05	*B. xylanisolvens* SD CC 1b	*Salyersviridae*	*Salyersvirinae*	*Salyersvirus*	*BX02*
BX06	*Bacteroides* sp. D22	*Salyersviridae*	*Salyersvirinae*	*Salyersvirus*	*BX06*
BX07	*B. ovatus* SD CMC 3f	*Salyersviridae*	*Salyersvirinae*	*Salyersvirus*	*BX07*
BC01	*B. clarus* YIT 12056	*Salyersviridae*	?	?	*BC01*
B124-14	*B. fragilis*	*Salyersviridae*	?	?	*B124-14*

B40-8	*B. fragilis*	*Salyersviridae*	?	?	*B40-8*
crAssphage	*B. intestinalis*	?	?	?	?

SUPPLEMENTARY FIGURE LEGENDS

Figure S1. Transcriptional activity of the BV01 prophage and surrounding chromosome. RNAseq reads from wild-type lysogen (WT) and cured lysogen (ΔBV01) strains were mapped to the region, and coverage was normalized to the total number of reads mapping to the genome. One representative of three replicates shown for each. The average normalized read coverage for each genome is displayed as the y-axis maximum (grey line). Maximum read coverage for the region is indicated on the y-axis. Forward reads (red) and reverse reads (blue) were plotted separately. Locations of two putative BV01-encoded transcriptional regulators are indicated (BVU_RS14235, BVU_RS14475).

Figure S2. Prevalence of prophage insertion adjacent to *tspO* in human gut samples. Reads from 256 healthy human gut metagenomes were obtained from the Human Microbiome Project Healthy Human Subjects Study. (A) Reads were first mapped to representative sequences of *tspO* from *B. vulgatus* and *B. dorei*. Percent abundance *tspO* reads was calculated on a per sample basis as the number of reads mapping to *tspO* divided by the total number of reads. Histogram shows counts of samples. (B) Reads mapping to *tspO* were filtered to only include reads antisense to *tspO*, predicted to point toward the *attB* based on the known genomic architecture. Mates to those reads were subsequently mapped to BV01 and its *Salyersvirus* relatives (Fig. 6). Only samples with

read pairs bridging *tspO* and a phage sequence are shown (*n*=34). Percent *tspO* reads mated with *Salyersvirus* reads was calculated as the read pairs bridging *tspO* and a phage sequence divided by the total number of reads mapping antisense to *tspO*. Histogram shows counts of samples.

Figure S3. Paired whole genome tree and nucleotide alignment of *Salyersviridae* phages. Phylogenomic Genome-BLAST Distance Phylogeny implemented with the VCTOR online tool (50) using amino acid data from all phage ORFs. For consistency, all phage genomes were annotated with MetaGeneAnnotator (94) implemented via VirSorter (95). Support values above branches are GBDP pseudo-bootstrap values from 100 replications. Genome alignment of all phages made with MAUVE. One locally collinear block (LCB) connects phages B124-14 and B40-8 to the *Salyersviridae* at the nucleotide level (green). Other LCB connecting lines removed for clarity.

Figure S4. Confirmation of activity of three additional Salyersviruses. DNase-treated culture supernatants for predicted Salyersviruses BV02 (A), BV04 (B), and BV05 (C) were sequenced. Assembly resulted in contigs corresponding to the free form of phages BV02 and BV05; BV04 did not yield any contigs corresponding to the putative prophage region, suggesting it is inactivated. Assembled free phage contigs were aligned to their integrated prophage region with Mauve (A-C). Sequence reads were mapped back to their free and integrated forms and represented as coverage curves (A-C). Vertical orange lines indicate the location of *att* sequences; vertical dashed blue lines indicate the location of contig breaks. (D) PCR amplification with phage-specific primers tests for phage

presence in pellet and supernatant fractions for 7 predicted Salyersviruses. Supernatant fractions were treated with DNase, eliminating all contaminating host genomic DNA, as demonstrated by the amplification of a host marker gene (16S rRNA). BV04 is not detectable in supernatant, supporting the conclusion that it is an inactivated prophage. PCR amplicons were visualized by agarose gel electrophoresis alongside GeneRuler Express DNA ladder (16S rRNA); ladder band sizes shown in Kb.

Figure S5. *Salyersviridae* sequence is detectable in human-associated samples. (A) Wastewater viromes were collected, and processed in three ways prior to sequencing (see Methods). Resulting reads were trimmed, pooled, and mapped to all *Salyersviridae* genomes and crAssphage. Only *Salyersvirinae* genomes shown in alignment to better demonstrate conservation, constructed with Mauve. Read coverages normalized to total number of reads in the metavirome. The maximum normalized read coverage for BC01 is 4.1, B40-8 and B124-14 is 1.13, and crAssphage is 7.33. (B) Reads from 256 healthy human gut metagenomes were obtained from the Human Microbiome Project Healthy Human Subjects Study. Reads were mapped to all 20 temperate *Salyersviridae* phages and crAssphage. Percent reads mapping was calculated on a per sample basis as the number of reads mapping to any virus divided by the total number of reads. Histogram shows counts of samples.

SUPPLEMENTARY TABLE TITLES

Table S1. Rockhopper results for RNAseq from B. vulgatus WT (602), ΔBV01 (853) and ΔBV01ΔtspO (1662).

1053

1054 **Table S2.** Detection of *Salyersviridae* in Human Microbiome Project Healthy Human
1055 Subjects Study samples.

1056

1057 **Table S3.** Bacterial strains, plasmids, and primers used in this study.

1058

1059 **Table S4.** Breseq read mapping results of *B. vulgatus* WT and mutant strains used in this
1060 study.

1061

1062 **Table S5.** DNA sequencing data generated for B. vulgatus isolates, cell-free
1063 Salyersviridae phages, and wastewater viromes used in this study.

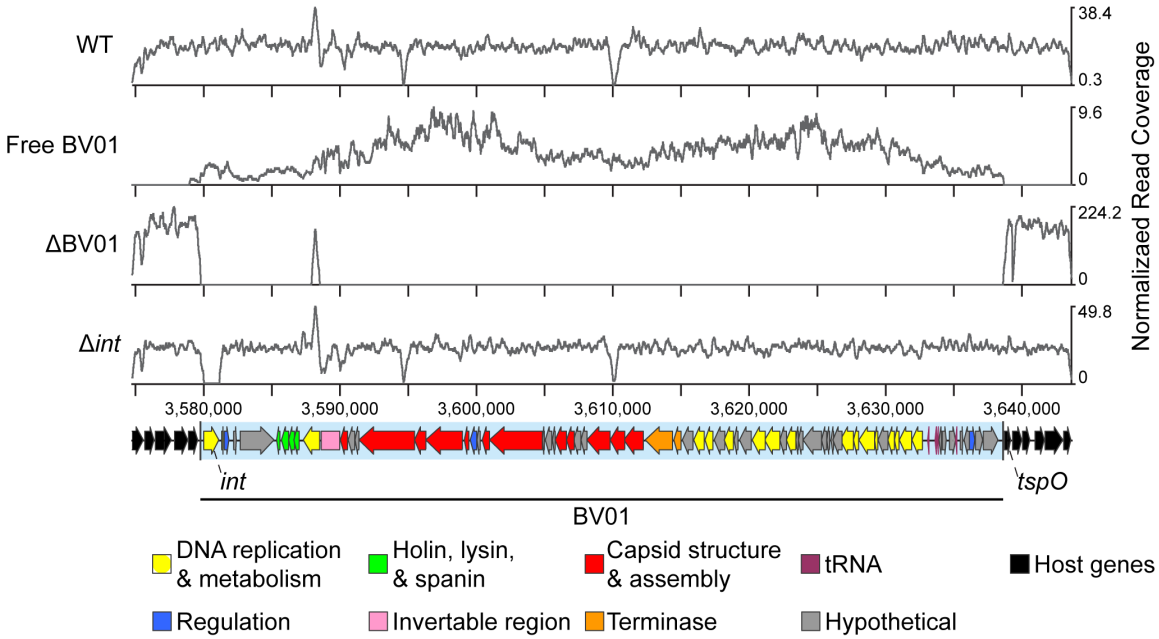

Figure 1. DNA sequencing reads mapped to the BV01 prophage region. Reads from shotgun sequencing of *B. vulgatus* genomic DNA (WT) or isolated phage DNA (Free BV01) were normalized to the total number of reads after trimming, and represented as a coverage curve. A cured lysogen (ΔBV01) and integrase deletion mutant (Δ*int*) of *B. vulgatus* were confirmed by shotgun sequencing of genomic DNA. The discrete coverage peak at position ~3,588,000 nt from ΔBV01 is attributed to a homologous sequence elsewhere on the *B. vulgatus* chromosome. Putative functions of BV01 genes are indicated by the colors in the legend.

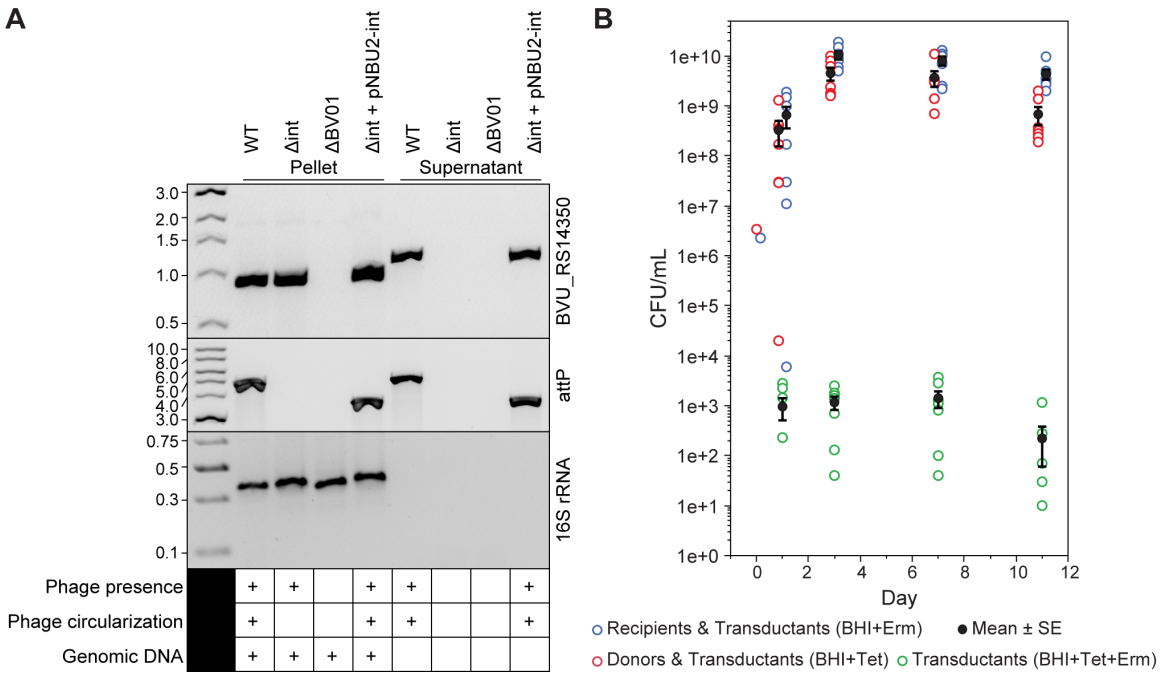

Figure 2. Prophage BV01 is an intact prophage. (A) Excision and circularization activities of the BV01 integrase are confirmed by PCR. The presence of phage DNA was detected by amplification of a phage marker gene (BVU_RS14350). Amplification across the phage attachment site (*attP*) indicates circularization of the BV01 genome; *attP* amplicons from the integrase complement strain (Δ*int* + pNBU2-*int*) are ~1.2 Kb shorter than wild-type amplicons due to deletion of the integrase gene. Supernatant fractions were treated with DNase, eliminating all contaminating host genomic DNA, as demonstrated by the amplification of a host marker gene (16S rRNA). Note that despite apparent size shift of BVU_RS14350 amplicons from the pellets and supernatants, Sanger sequencing validated that the products are in fact identical. PCR amplicons were visualized by agarose gel electrophoresis alongside NEB 1 Kb DNA ladder (BVU_RS14350, *attP*) or GeneRuler Express DNA ladder (16S rRNA); ladder band sizes shown in Kb. (B) BV01 can transduce uninfected hosts in a gnotobiotic mouse. Germ-free mice (*n*=7) were gavaged with an equal mixture of a BV01-*tetQ* lysogen and an erythromycin-tagged cured

lysogen (Day 0). Recipient, donor, and transductant cells were identified by plating on Brain Heart Infusion (BHI) media with antibiotic selection: erythromycin (Erm) or tetracycline (Tet).

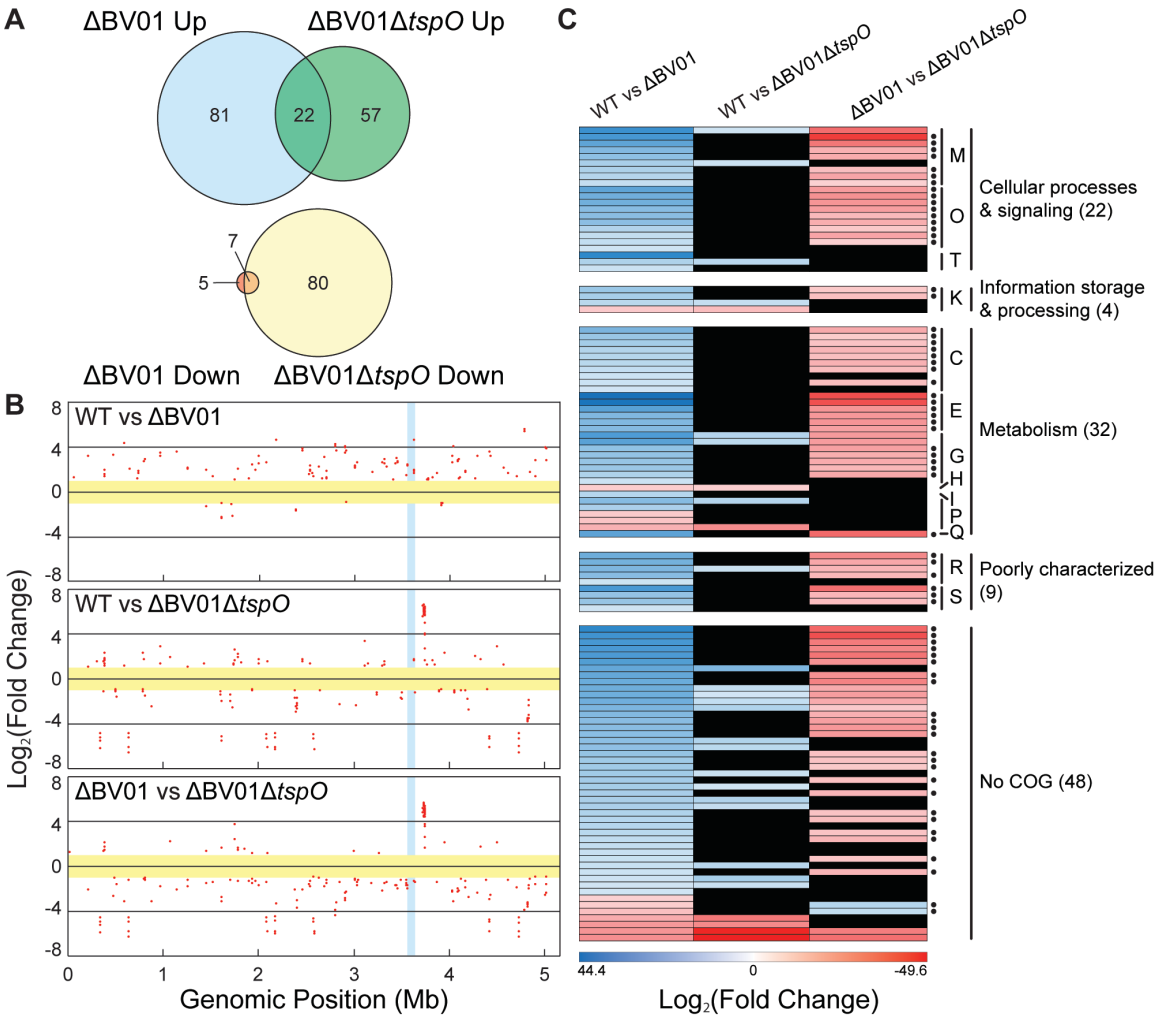

Figure 3. Differential regulation of the host transcriptome in response to BV01 lysogeny. (A) Count of differentially regulated transcripts as compared to the wild-type *B. vulgatus* lysogen (fold change ≥ 2, *q*-value ≤ 0.01). (B) Chromosomal localization of the differentially expressed genes. Each dot represents a differentially expressed transcript on a $\log_2$ scale; genes below the 2-fold change cutoff (yellow) and within the BV01 prophage (blue) not shown. Positive fold change values correspond to increased transcription in the second background listed. (C) General functional assignment of genes differentially expressed between wild-type and cured lysogen strains was accomplished using the Clusters of Orthologous Groups (COGs). Transcripts which are not differentially expressed in other strain comparisons are shown in black. *tspO*-dependent

transcripts are marked on right (•). Positive fold change values correspond to increased transcription in the second background listed. Letters correspond to COG categories: cell wall/membrane/envelope biogenesis (M), post-translation modification, protein turnover, and chaperones (O), signal transduction mechanisms (T), transcription (K), energy production and conversion (C), amino acid transport and metabolism (E), carbohydrate transport and metabolism (G), coenzyme transport and metabolism (H), lipid transport and metabolism (I), inorganic ion transport and metabolism (P), secondary metabolite biosynthesis, transport, and catabolism (Q), general function prediction only (R), function unknown (S).

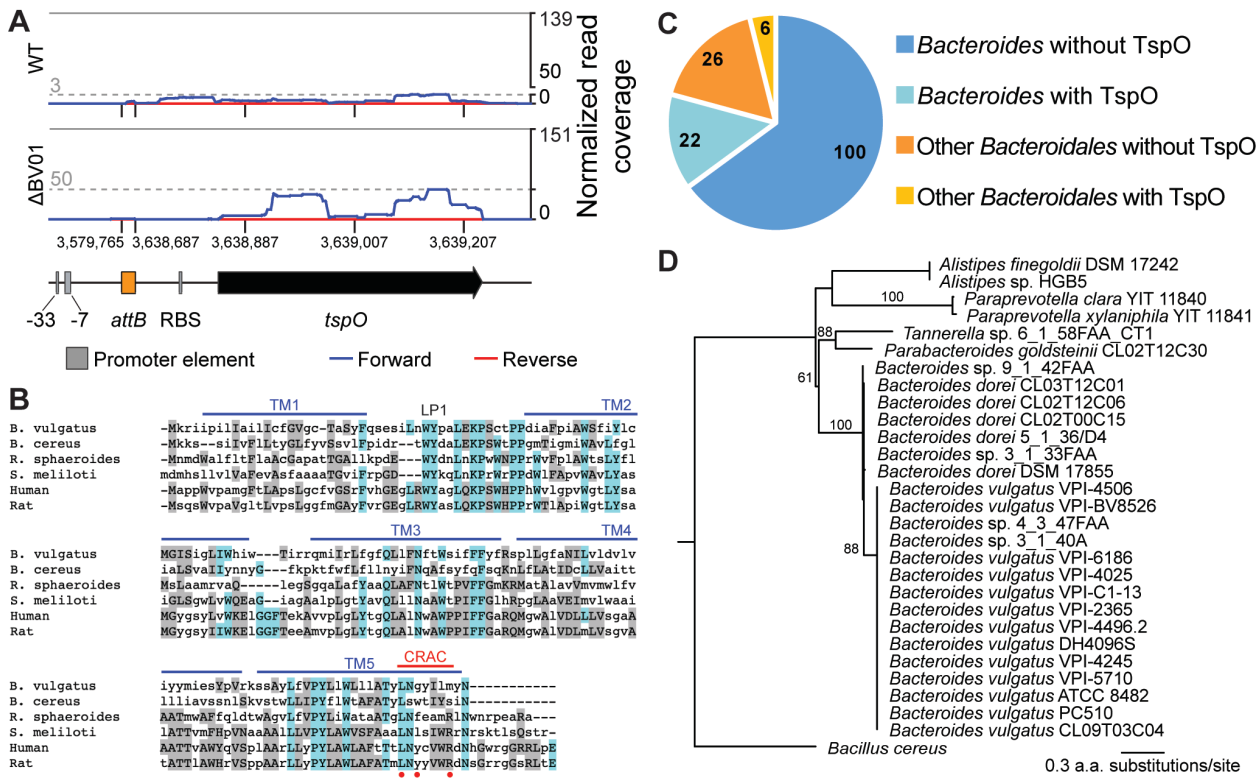

Figure 4. BV01 integration disrupts transcription of *tspO*. (A) Transcriptional activity of the *tspO* gene region as it exists in the cured lysogen background. RNAseq reads from wild-type (WT) and cured lysogen (ΔBV01) *B. vulgatus* were mapped to the region, and coverage was normalized to the total number of reads mapping to the genome. The average normalized read coverage for each genome is displayed as the y-axis maximum (grey line). Maximum read coverage for the region is displayed as the grey dashed line. (B) Amino acid alignment of *B. vulgatus* TspO with known TspO sequences was generated with MUSCLE (49). Identical and similar residues are colored blue and gray, respectively. Shown are TspO protein sequences from *B. vulgatus* (WP_005843416.1), *Bacillus cereus* (GCF80909.1), *Rhodobacter sphaeroides* (AAF24291.1), *Sinorhizobium meliloti* (AAF01195.1), human (NP_001243460.1), and rat (NP_036647.1). Secondary structures, cholesterol recognition/interaction amino acid consensus (CRAC) sequence, and critical residues (•) from *R. sphaeroides* TspO crystal structure are shown (50). (C) The search

for TspO homologs in the family *Bacteroidales* was accomplished with a BLAST-based approach, using the *Bacillus cereus* copy of TspO (GCF80909.1) as a query against a database of 154 gut-associated *Bacteroidales* genomes, 122 of which are from the genus *Bacteroides*. Genome counts are indicated within categories. (D) Gene tree estimated from TspO sequences across the *Bacteroidales*. All *B. vulgatus* and *B. dorei* genomes included in the search encode *tspO*. Clade for *B. vulgatus* TspO sequences is displayed as a polytomy; all *B. vulgatus* TspO sequences are at least 98% identical to each other. Numbers above branches represent bootstrap values; only bootstraps over 50 shown. The gene tree was estimated using FastTree (51).

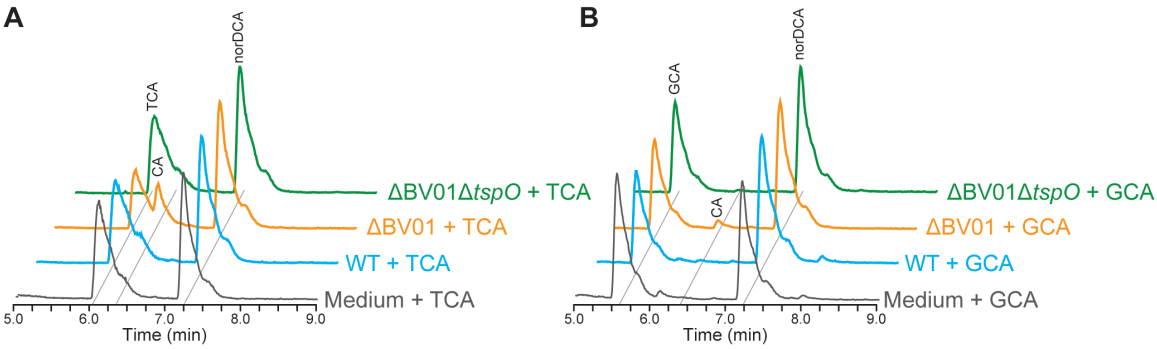

Figure 5. BV01 alters host interactions with bile acids in a *tspO*-dependent manner. (A) Representative LC/MS traces showing *B. vulgatus* deconjugates taurocholic acid (TCA) to cholic acid (CA) in the cured lysogen background (ΔBV01), but little or no activity CA is detectable in the wild-type (WT) or cured lysogen *tspO* deletion (ΔBV01Δ*tspO*) backgrounds. *B. vulgatus* cultures were incubated with 50 μM TCA for 16 hr prior to bile acid extraction. (B) Representative LC/MS traces showing *B. vulgatus* deconjugates glycocholic acid (GCA) to CA in the ΔBV01 background, but not in the WT or ΔBV01Δ*tspO* backgrounds. *B. vulgatus* cultures were incubated with 50 μM GCA for 48 hr prior to bile acid extraction. Nordeoxycholic acid (norDCA) was added to a final concentration of 15 μM as an internal standard after incubation. Peaks labeled for their metabolites based on m/z; TCA = 514.29, GCA = 464.30, CA = 407.28, norDCA = 377.27.

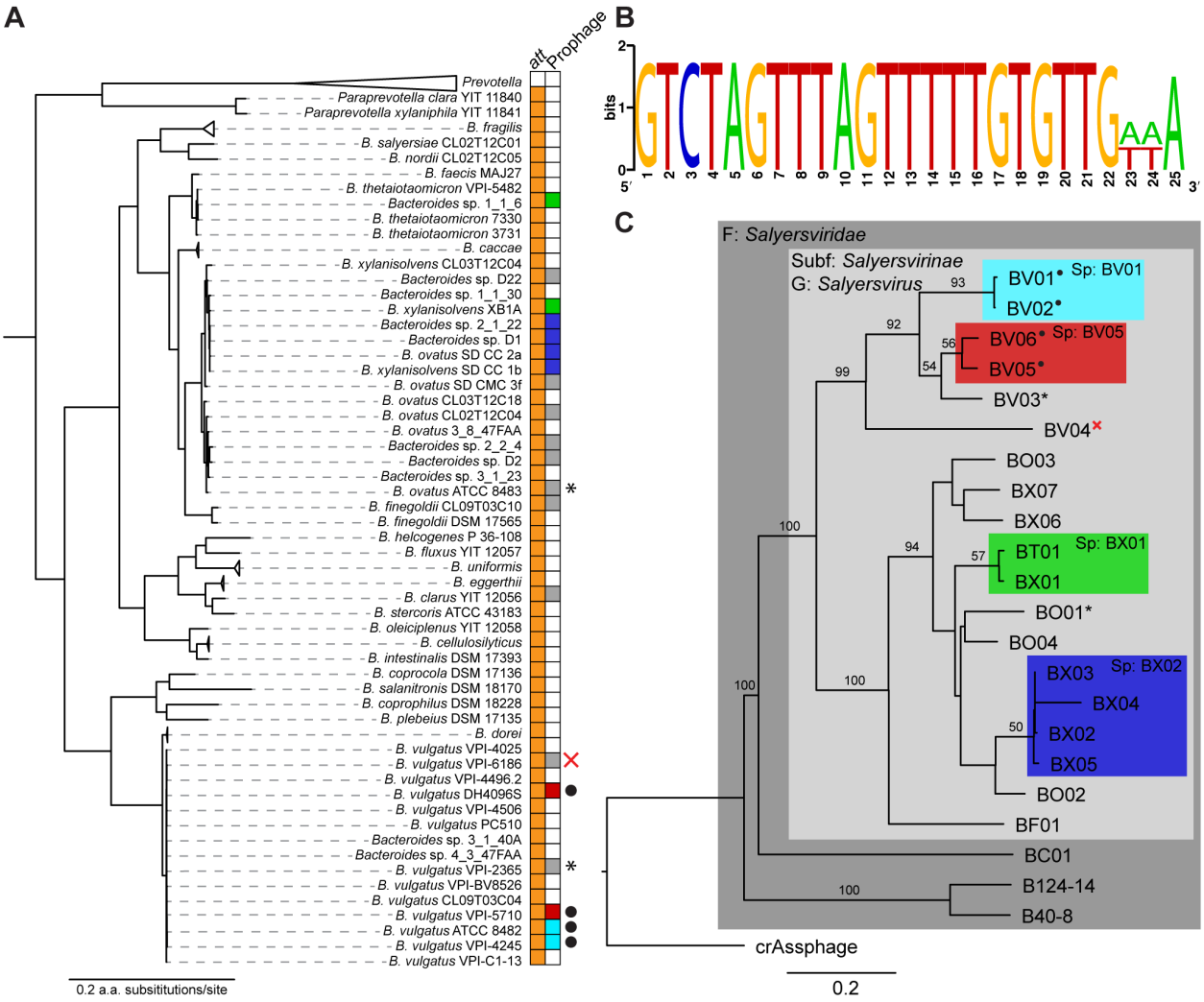

Figure 6. *Salyersviridae* occur throughout the *Bacteroides* genus. (A) *Bacteroides* phylogeny and occurrence of *Salyersviridae att* site. All duplications of the *att* site are associated with a putative integrated prophage. Host phylogeny estimated by maximum likelihood from concatenated alignment of 13 core genes. (B) Consensus *att* site for *Salyersvirinae*. The *attP* is duplicated upon integration of a *Salyersvirinae* prophage, resulting in direct repeats. Image made with the WebLogo online tool. (C) Phylogenomic Genome-BLAST Distance Phylogeny implemented with the VCTOR online tool (53) using amino acid data from all phage ORFs. For

consistency, all phage genomes were annotated with MetaGeneAnnotator (54) implemented via VirSorter (55). Support values above branches are GBDP pseudo-bootstrap values from 100 replications. Family (F), subfamily (Subf), genus (G), and species (Sp) assigned by OPTSIL clustering (56) (Table 1). Each leaf of the tree represents a unique phage species, except where indicated by colored boxes. Active prophages confirmed by sequencing and/or PCR indicated with "•"; prophages confirmed to have been inactivated by genome rearrangement indicated with "x"; prophages which were tested for activity with inconclusive results indicated with "*" (Fig. S4).