Cell Reports, Volume *32*

Supplemental Information

Infection with Bacteroides Phage BV01 Alters the Host Transcriptome and Bile Acid Metabolism in a Common Human Gut Microbe

Danielle E. Campbell, Lindsey K. Ly, Jason M. Ridlon, Ansel Hsiao, Rachel J. Whitaker, and Patrick H. Degnan

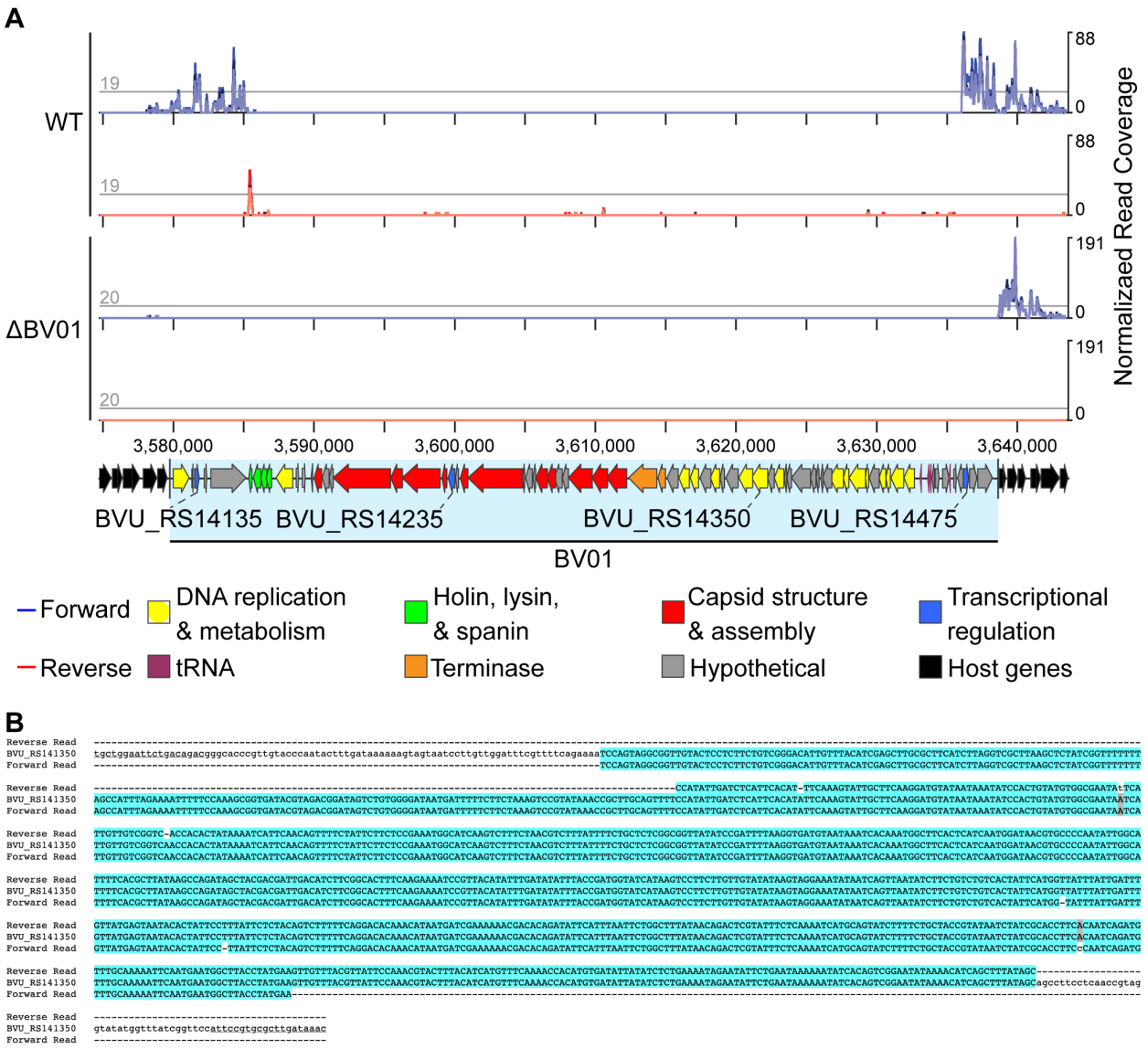

Figure S1. Transcriptional activity at BV01 attachment site and confirmation of free BV01 integrase amplicons.
Related to Figures 1, 2, and 3. (A) RNAseq reads from wild-type lysogen (WT) and cured lysogen (ΔBV01) strains were mapped to the region, and coverage was normalized to the total number of reads mapping to the genome. Locations of the phage marker gene (BVU_RS14350) and three phage transcriptional regulators are indicated: the putative phage repressor (BVU_RS14475), anti-repressor (BVU_RS14135), and an RNA polymerase sigma factor (BVU_RS14235). Three replicates shown. The average normalized read coverage for each genome is displayed as the grey line. Maximum read coverage for the region is indicated on the y-axis. Forward transcription (red) and reverse transcription (blue) were plotted separately. (B) Nucleotide alignment of BVU_RS14350 with trimmed Sanger sequencing data generated from DNase-treated cell-free BV01 supernatant. Despite appearing larger in size by gel electrophoresis (Fig. 2B), no large insertions were detected. Locations of primers used to amplify BVU_RS14350 and for Sanger sequencing denoted by underlined sequence.

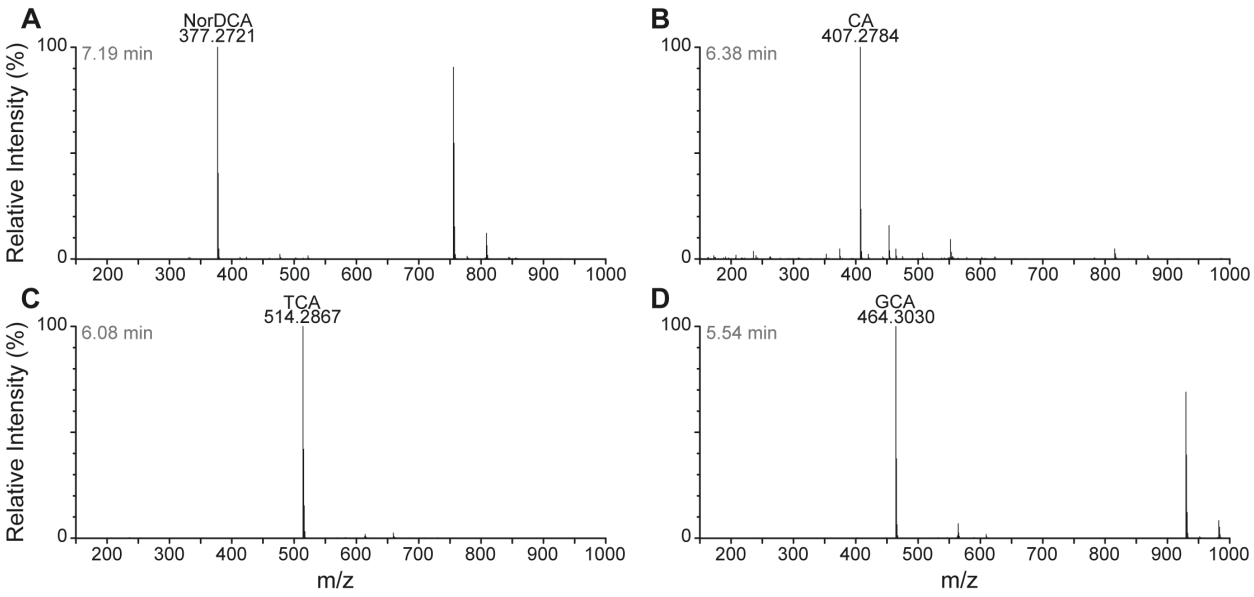

Figure S2. Representative extracted ion count (EIC) chromatograms for the bile acid species of interest. Related to Figure 5. EIC are shown for (A) the internal control nordeoxycholic acid (NorDCA), (B) the product of bile acid deconjugation, cholic acid (CA), taurocholic acid (TCA), and glycocholic acid (GCA). Times of flight indicated in grey.

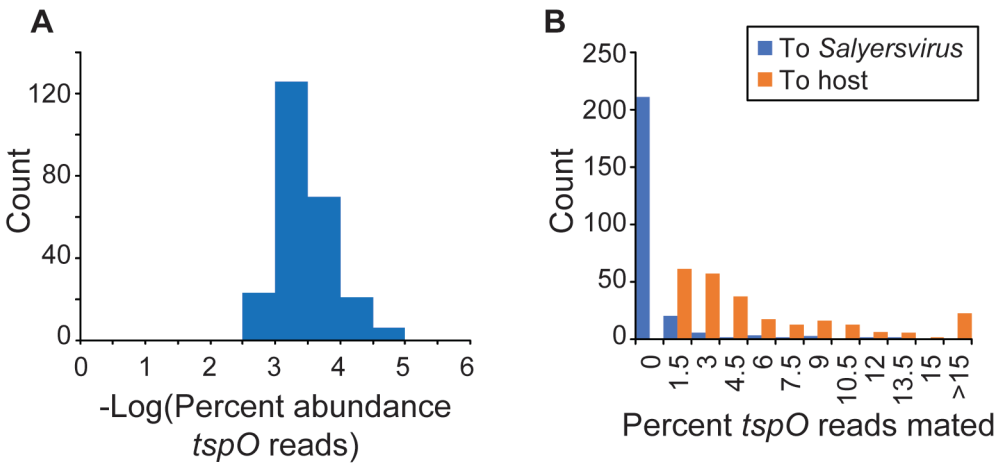

Figure S3. Prevalence of prophage insertion adjacent to *tspO* in human gut samples. Related to Figures 4 and 6. Reads from 246 healthy human gut metagenomes were obtained from the Human Microbiome Project Healthy Human Subjects Study. (A) Reads were first mapped to representative sequences of *tspO* 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 either the empty *B. vulgatus attB* (To host) or to BV01 and its *Salyersvirus* relatives (To *Salyersvirus*) (Fig. 6). Percent *tspO* reads mated was calculated as the read pairs bridging *tspO* and the sequence of interest, sequence divided by the total number of reads mapping antisense to *tspO*. Histogram shows counts of samples.

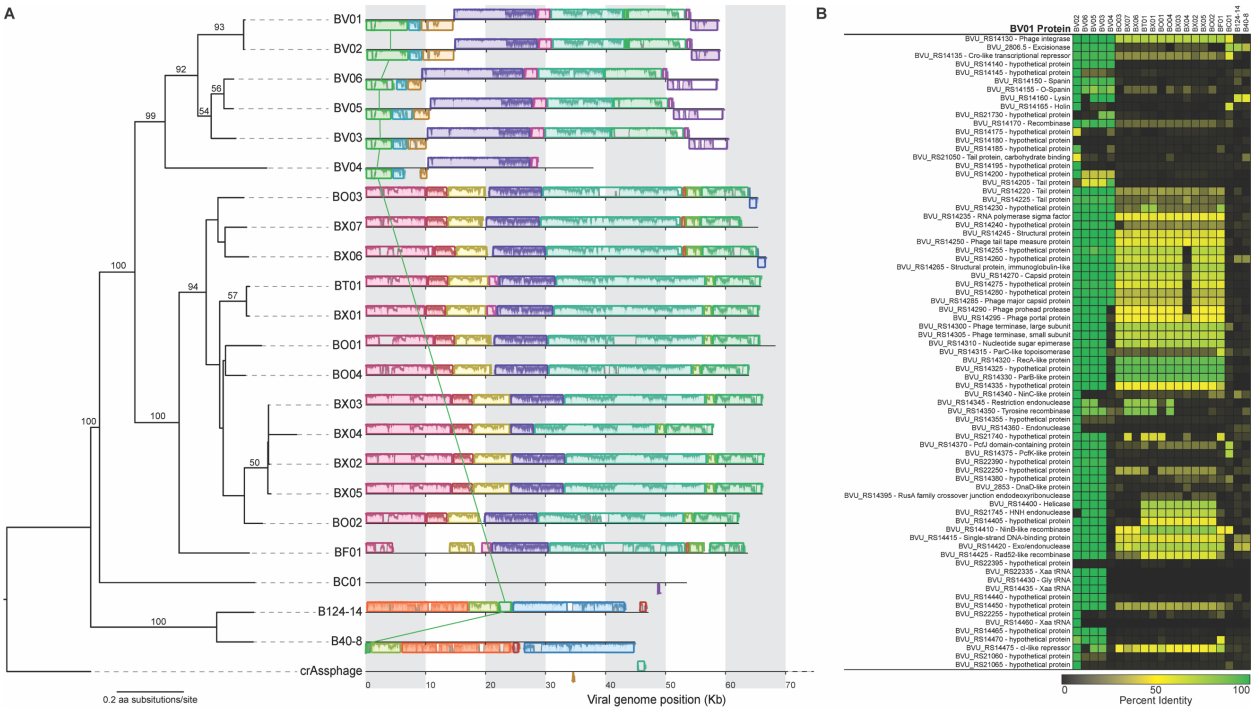

Figure S4. Whole genome tree, nucleotide alignment and proteomic similarity of *Salyersviridae* phages. Related to Figure 6. (A) Phylogenomic Genome-BLAST Distance Phylogeny implemented with the VCTOR online tool (Meier-Kolthoff and Göker, 2017) using amino acid data from all phage ORFs. For consistency, all phage genomes were annotated with MetaGeneAnnotator (Noguchi et al., 2008) implemented via VirSorter (Roux et al., 2015). 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. (B) Gene similarity matrix for BV01-encoded genes among the S*alyersviridae*. Genes for non-BV01 predicted phages were predicted with Prodigal (Hyatt et al., 2010). Similarity was computed as amino acid percent identity for protein coding genes and nucleotide percent identity for tRNA genes.

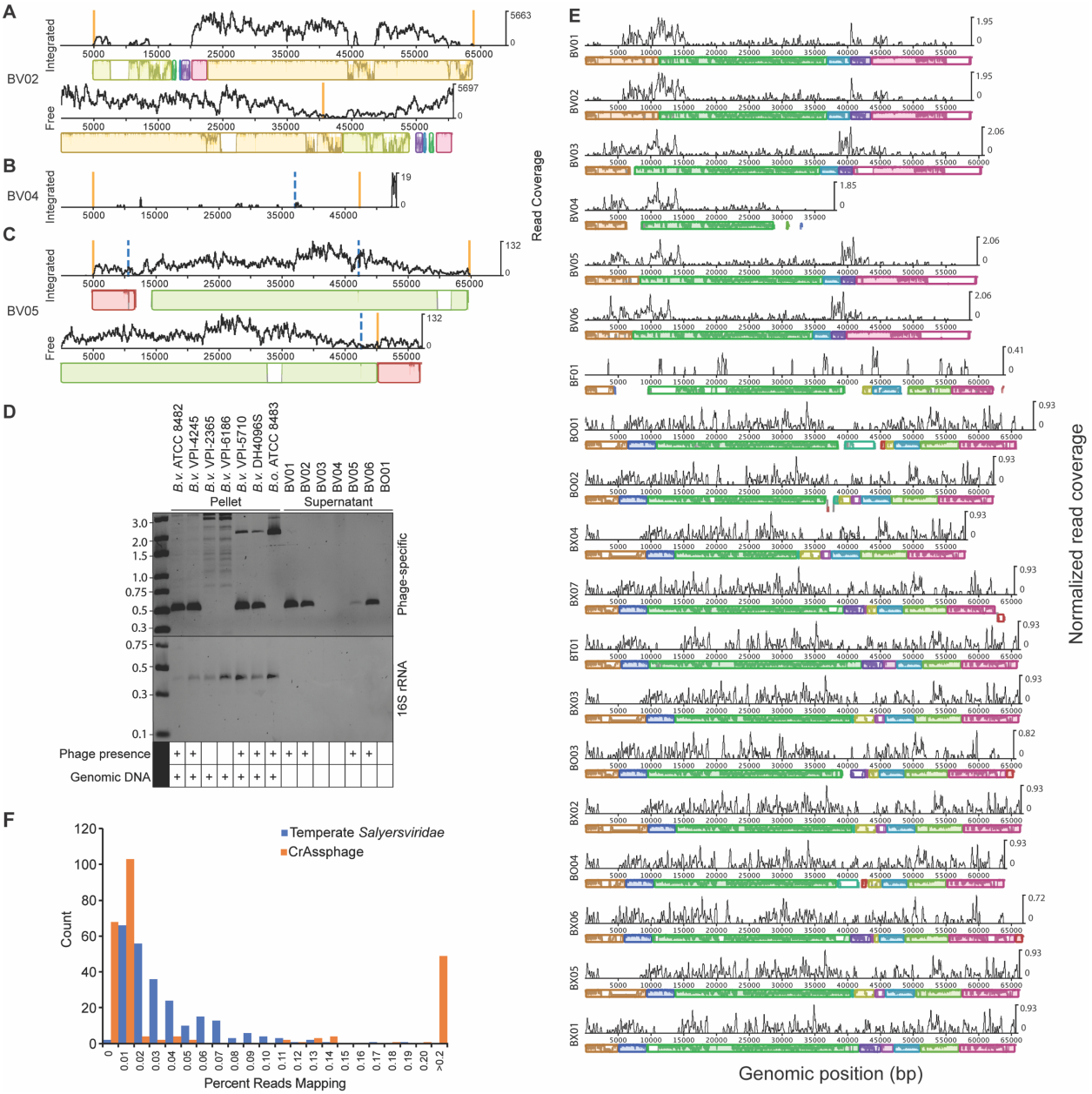

Figure S5. Confirmation of activity of *Salyersviridae* phages *in vitro* and in human-associated samples. Related to Figure 6. DNase-treated culture supernatants for predicted Salyersviruses BV02 (A), BV04 (B), and BV05 (C) were sequenced. Assembly resulted in contigs corresponding to the free form 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. (E) 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. (F) Reads from 246 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.