



Figures and figure supplements

Resurrection of a global, metagenomically defined gokushovirus

Paul C Kirchberger and Howard Ochman

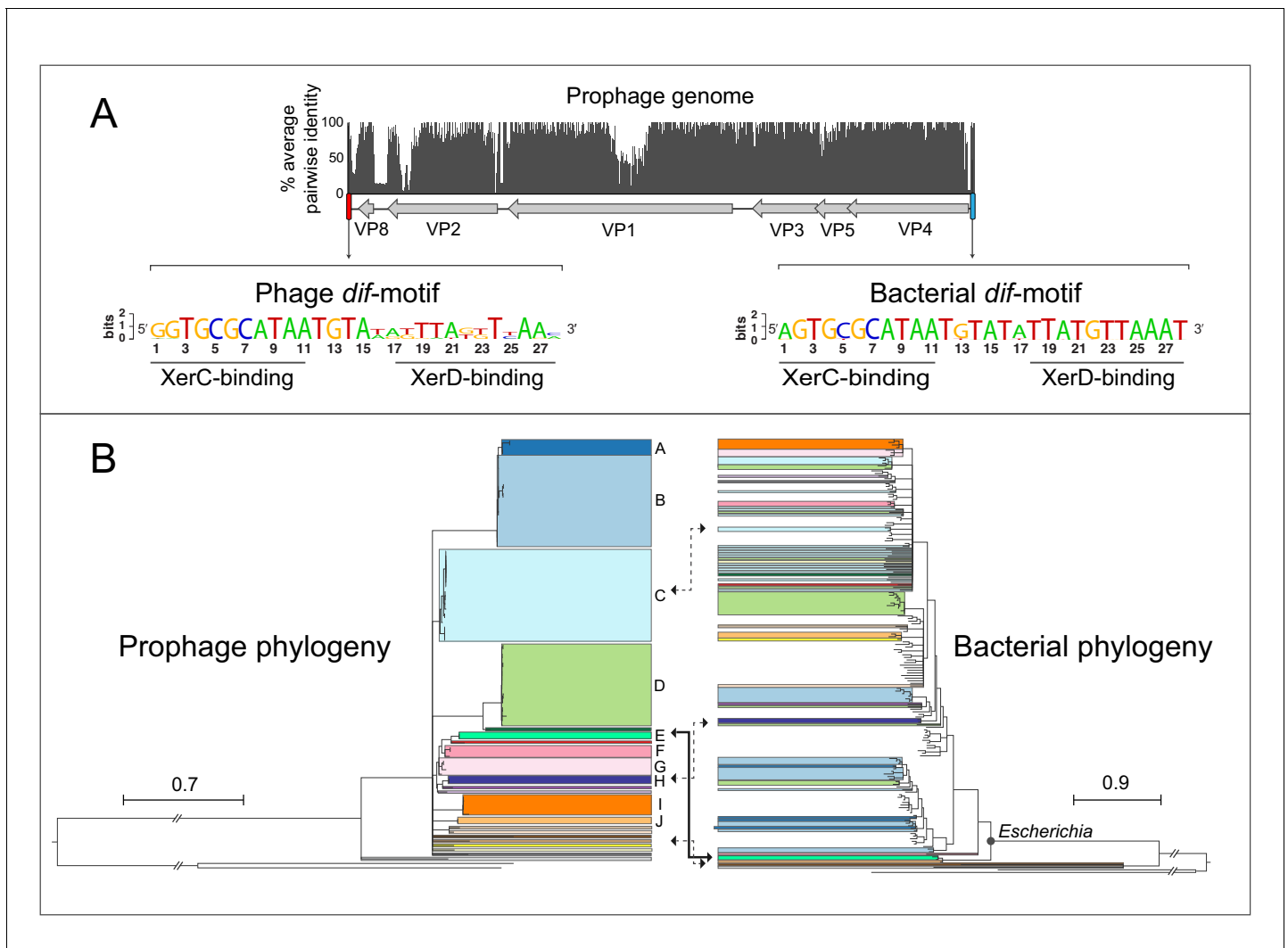


Figure 1. Gokushovirus prophages of Enterobacteria. (A) Genome organization and average pairwise nucleotide identities of the gokushovirus prophages detected in *Escherichia*. Integrated prophage genomes range from 4316 to 4692 bp in length, with genes indicated by grey arrows, and flanking phage and bacterial *dif*-motifs indicated by red and blue bars. Nucleotide sequence logos of phage and bacterial *dif*-motifs, with corresponding XerC- and XerD-binding sites are shown. (B) Phylogeny of gokushovirus prophages and their enterobacterial hosts. Single lineages or clades comprising strains sharing >95% average nucleotide identity are individually colored, with colors in the bacterial phylogeny corresponding to those of their associated prophages. Clades not colored in the bacterial phylogeny correspond to *Escherichia coli* collection reference (ECOR) strains. Clades with bootstrap support values below 70% are collapsed. Arrows denote prophage-host pairs in which the corresponding phage was tested against *E. coli* K-12, with the solid arrow indicating the prophage-host pair whose phage formed plaques in *E. coli* K-12 and subsequently used in experimental analyses. Tree scale bars correspond to nucleotide or amino acid substitutions/site for prophage and host trees, respectively, and ancestral branches with hatch-marks are truncated by the length of two scale bars. Accession numbers and details of prophages and their corresponding hosts are listed in **Supplementary files 1 and 3**.

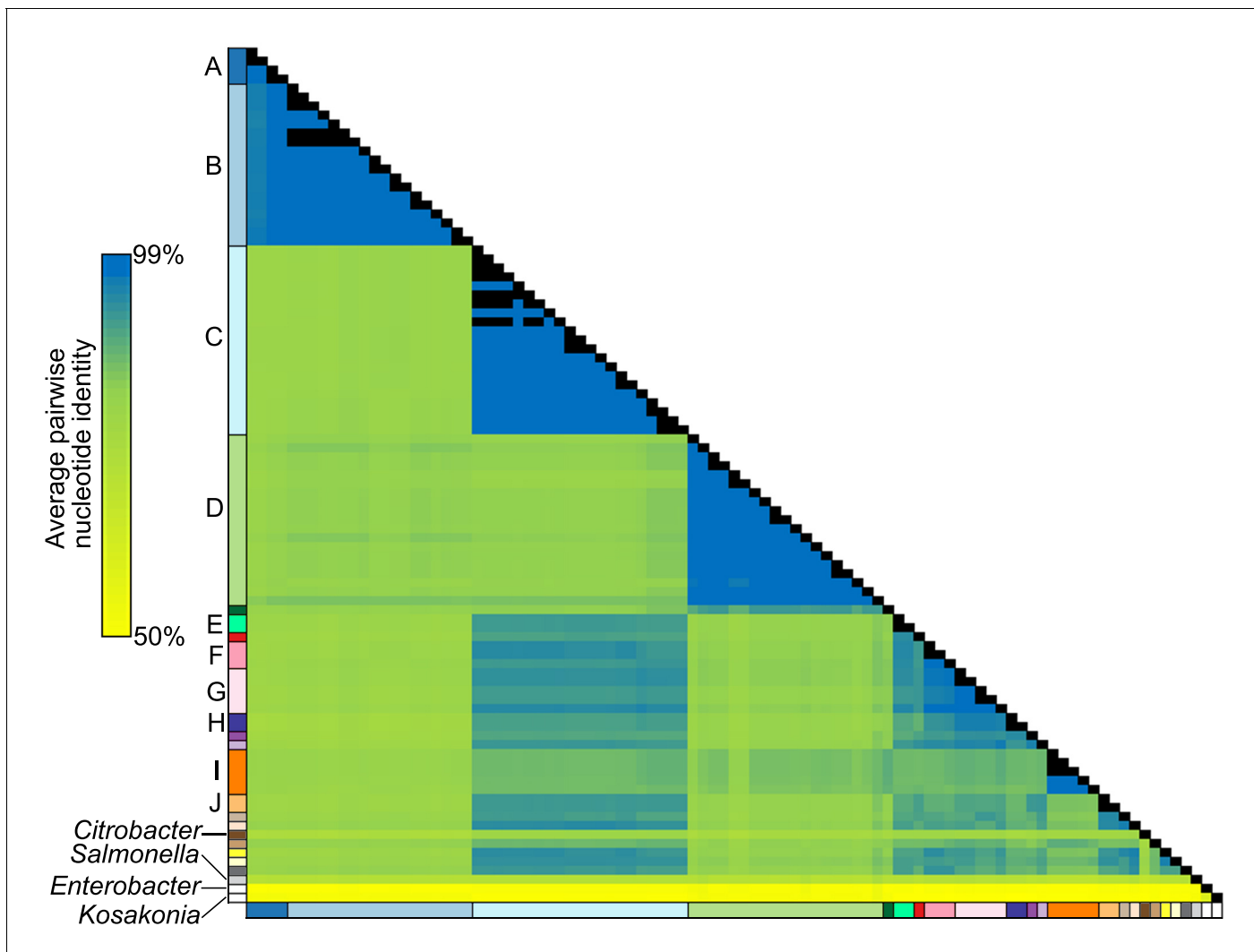


Figure 1—figure supplement 1. Sequence similarities among gokushoviral prophages. Average nucleotide identity values calculated from pairwise comparisons of whole prophage genomes indicated on a yellow-blue scale, with identical prophages indicated in black. Colors of individual lineages (or clades) bordering the similarity matrix correspond to those used in **Figure 1B**, with non-*Escherichia* prophages indicated separately.

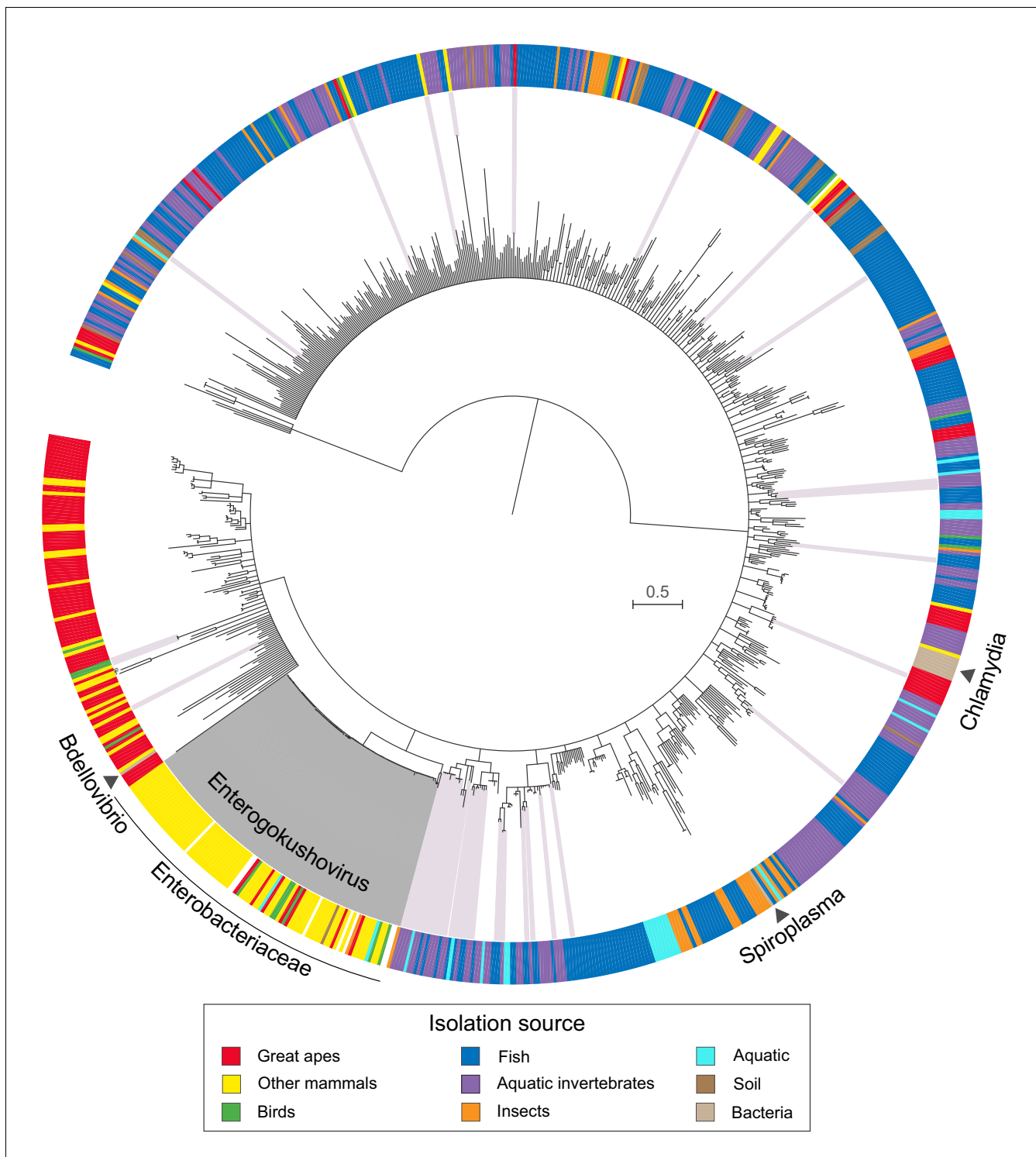


Figure 2. Phylogeny and sources of *Gokushovirinae*. Maximum likelihood tree built from concatenated alignments of VP1 and VP4 protein sequences of 855 gokushovirus genomes. Tree is midpoint rooted, and branch support estimated with 100 bootstrap replicates. Branches with bootstrap support values below 70% are collapsed. Clades highlighted in grey indicate *Enterogokushovirus* prophages with recognizable *dif*-motifs; those highlighted in pink possess *dif*-motifs identified through an iterative HMM search. Outer ring indicates isolation source, with black triangles denoting the phylogenetic positions of officially described gokushoviruses. Scale bar corresponds to amino acid substitutions per site. Sample accession numbers are listed in **Supplementary file 4**.

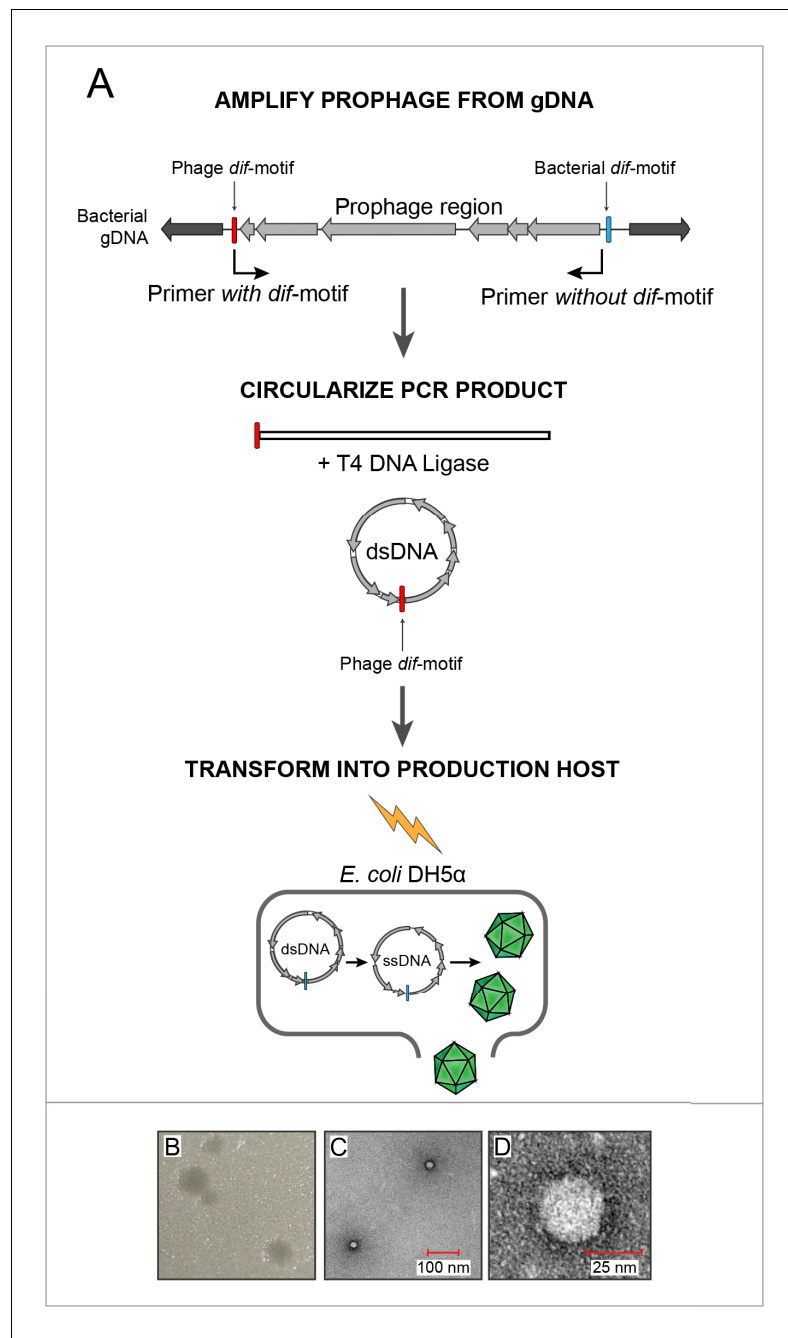


Figure 3. In vitro assembly and revival of enterobacterial gokushoviruses. **(A)** Scheme used to produce viable phage from prophage inserts. The prophage region, with genes colored in grey, is amplified from the bacterial genome (black) using primers that incorporate the phage *dif*-motif but exclude the bacterial *dif*-motif (indicated by bent arrows). Circularization of the amplification product results in a molecule corresponding to the replicative dsDNA form of the phage. Transformation of this circular molecule into electrocompetent *E. coli* DH5α cells leads to expression of phage proteins, replication and packaging of ssDNA into infective virions. **(B)** Plaques formed by constructed bacteriophage EC6098 after infecting *E. coli* BW25113. **(C)** TEM images of bacteriophage EC6098 viewed at 175,000x magnification. **(D)** TEM images of bacteriophage EC6098 viewed at 300,000x magnification.

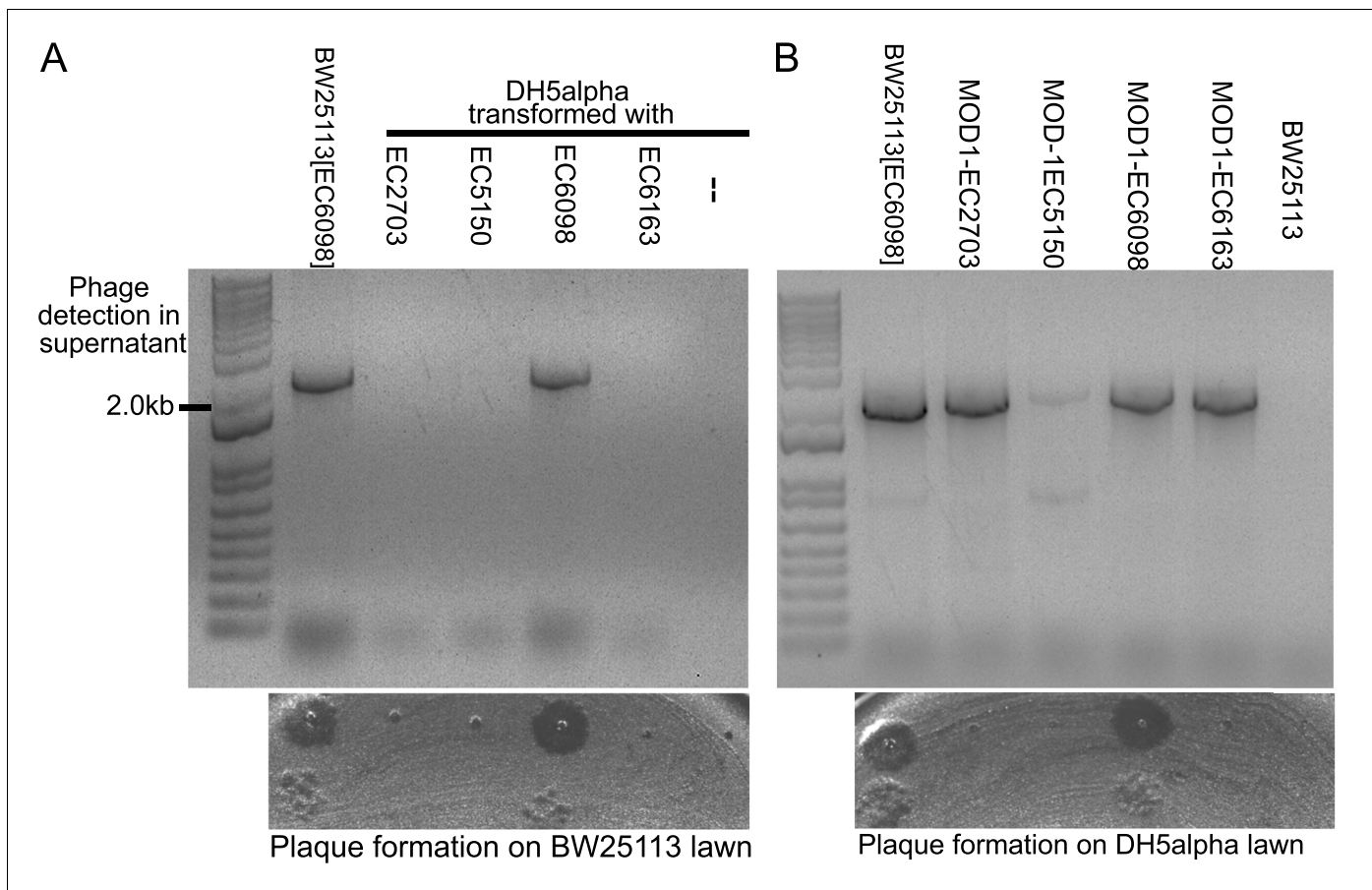


Figure 3—figure supplement 1. Synthesis of gokushoviruses in *Escherichia coli* DH5 α and natural host strains. **(A)** Transformation and growth of four synthesized gokushovirus genomes (EC2703, EC5150, EC6098, EC6163) in *E. coli* DH5 α . Gel image shows assay for circularized phage genomes in liquid cultures grown overnight at 37°C. Agar-plate pictured below gel shows screen for plaque formation in the corresponding host-phage combinations when plated on a lawn of *E. coli* BW25113. Plaques are evident in the positive control (+) of BW25113 producing EC6098 and in DH5 α transformed with EC6098. – denotes negative control of DH5 α without transformed genome. **(B)** Growth of four gokushoviruses (EC2703, EC5150, EC6098, EC6163) in their natural hosts. Gel image shows assay for circularized phage genomes in liquid cultures grown overnight at 37°C. Agar-plate pictured below gel shows screen for plaque formation in the corresponding host-phage combinations when plated on a lawn of *E. coli* DH5 α . Plaques are evident in the positive control (+) of *E. coli* BW25113 producing EC6098 and in MOD1-EC6098. BW25113 in absence of gokushoviral genomes serves as negative control.

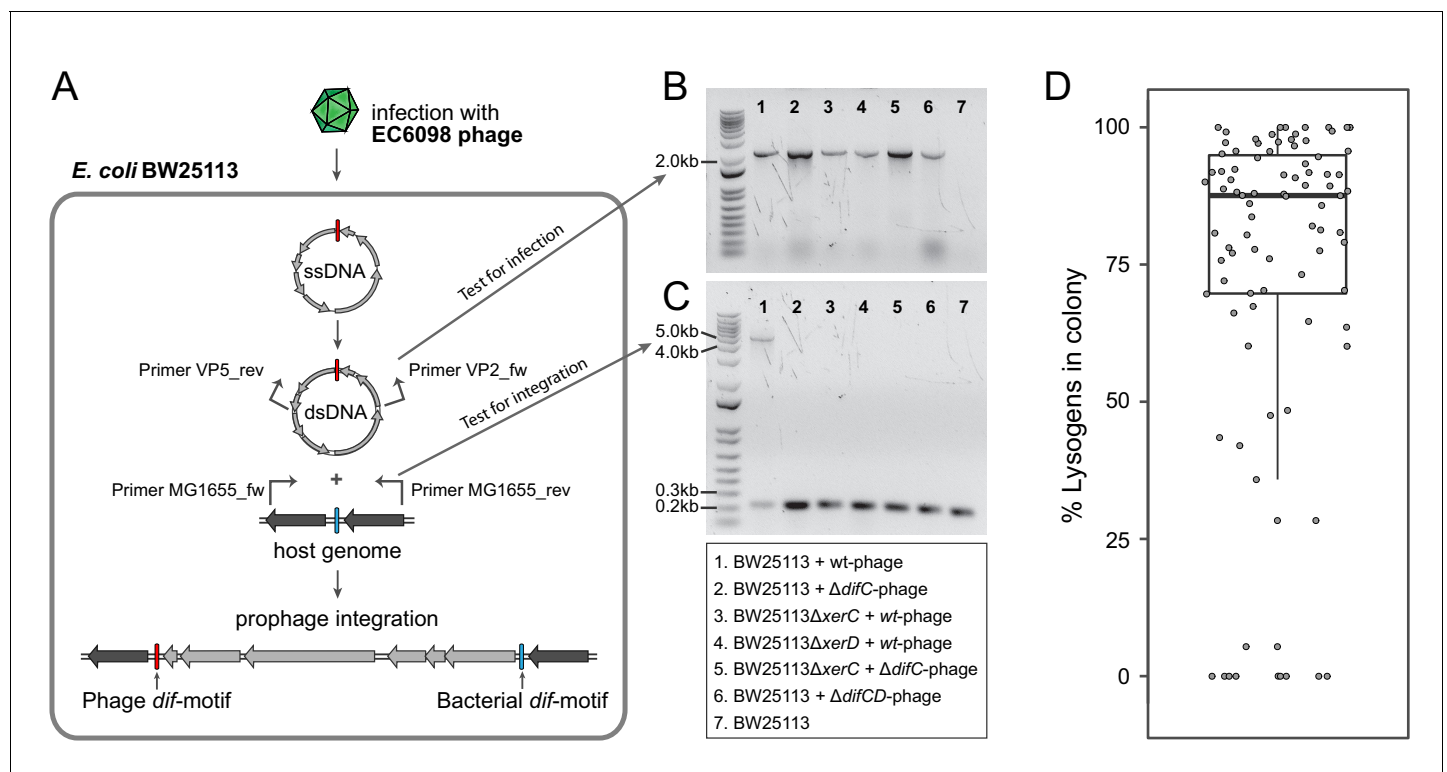


Figure 4. Integration of gokushoviruses into *E. coli* host genome. **(A)** Schematic representation of phage integration process, and detection of circularized EC6098 phage genome and integrated prophages in *E. coli* BW25113. Upon infection, phage EC6098 releases ssDNA, with *dif*-motif denoted in red, leading to formation of dsDNA replicative genomes that integrate downstream of the bacterial *dif*-motif (blue). Primers VP2_rev and VP5_fw (indicated by bent arrows on EC6098 genome) anneal to genes flanking the phage *dif*-motif and amplify a ~2.1 kb product from closed circular phage genomes (corresponding to bands in panel B). Primers MG1655dif_fw and MG1655dif_rev (indicated by bent arrows on host genome) anneal to sites flanking the bacterial *dif*-motif and amplify either a 210 bp region of bacterial DNA when there is no phage integration or a ~5 kb region denoting the presence of an integrated prophage (corresponding to bands in panel C). **(B)** Detection of fragments indicating the presence of circularized phage. Numbered lanes correspond to samples listed in the box below the gel photographs. **(C)** Detection of fragments indicating the presence or absence of integrated phage from lysogenic colonies after infection of BW25113 strains with wild type or mutant phage. Numbered lanes are the same as in panel B, and correspond to samples listed in the box below the gel photograph. **(D)** Proportion of cells with integrated prophages in clonal lysogenic colonies. Box-and-whiskers plot shows median, 25th and 75th percentiles, and 1.5 inter-quartile range as well as individual datapoints for 87 independently sampled clonal colonies.

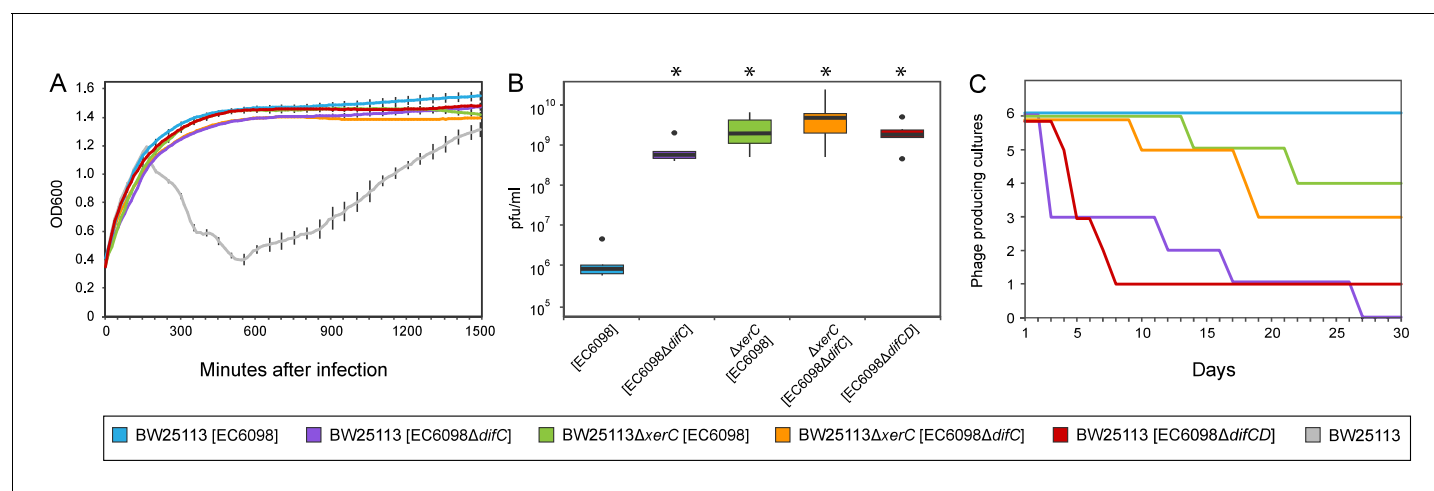


Figure 5. Quantifying the effects of gokushovirus carriage. **(A)** Superinfection immunity offered by integrated or non-integrated phages. Unimpeded growth in the presence of infecting EC6098 phage is shown for host cultures carrying lysogenic or lysogeny-deficient EC6098. Lysis of host by EC6098 in the absence of superinfection immunity is shown in grey for comparison. Averages and standard deviations are based on three replicates. **(B)** Phage production by lysogenic and non-lysogenic hosts. Plaque-forming units in culture supernatant were determined from six independent cultures of BW25113 or BW25113ΔxerC, each grown overnight at 37°C from an initial concentration of OD₆₀₀ = 0.7. Box-and-whiskers plots show median, 25th and 75th percentiles (upper and lower hinges) and 1.5 inter-quartile range (whiskers). Outliers are shown as individual dots. P-values of <0.01 (Wilcoxon Rank-Sum test) in comparison to BW25113[EC6098] are indicated with asterisks. **(C)** Continuous phage production over the course of a month. Formation of plaques from supernatants of six replicate cultures of hosts carrying integrated or non-integrated phages is checked daily before 1:1000 dilution and growth for 18–24 hr at 37°C. In all cases, BW25113ΔfhuA served as the host strain for agar overlay assays.