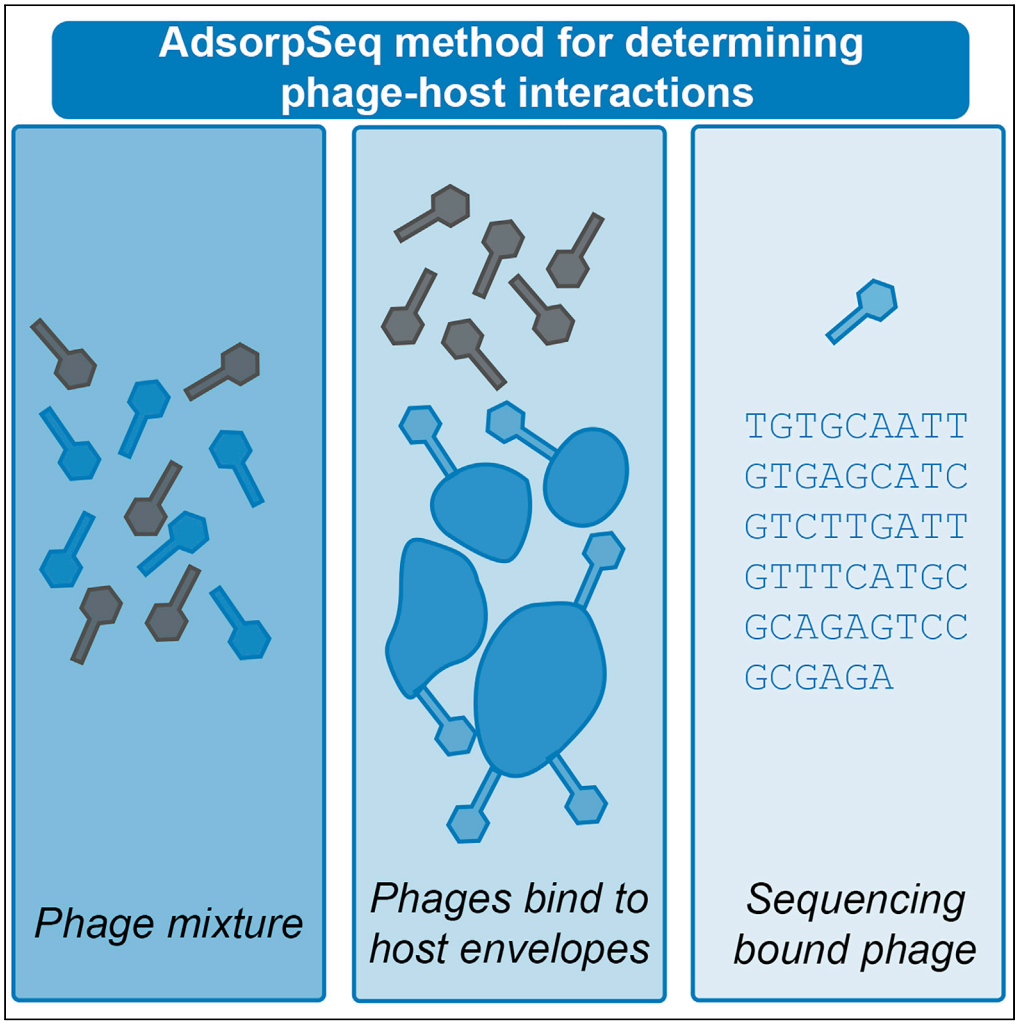
iScience

CellPress
OPEN ACCESS

Article

Adsorption Sequencing as a Rapid Method to Link Environmental Bacteriophages to Hosts



Patrick A. de Jonge, F.A. Bastiaan von Meijenfeldt, Ana Rita Costa, Franklin L. Nobrega, Stan J.J. Brouns, Bas E. Dutilh

bedutilh@gmail.com

HIGHLIGHTS

AdsorpSeq allows rapid determination of bacteria-bacteriophage interactions

Model phages can be differentially sequenced based on binding ability with AdsorpSeq

With AdsorpSeq 26 new phage-host interactions were determined in hospital wastewater

de Jonge et al., iScience *23*, 101439
September 25, 2020
https://doi.org/10.1016/j.isci.2020.101439

iScience

CellPress
OPEN ACCESS

Article

Adsorption Sequencing as a Rapid Method to Link Environmental Bacteriophages to Hosts

Patrick A. de Jonge,[1,2] F.A. Bastiaan von Meijenfeldt,[1] Ana Rita Costa,[2] Franklin L. Nobrega,[2] Stan J.J. Brouns,[2] and Bas E. Dutilh[1,3,*]

SUMMARY

An important viromics challenge is associating bacteriophages to hosts. To address this, we developed adsorption sequencing (AdsorpSeq), a readily implementable method to measure phages that are preferentially adsorbed to specific host cell envelopes. AdsorpSeq thus captures the key initial infection cycle step. Phages are added to cell envelopes, adsorbed phages are isolated through gel electrophoresis, after which adsorbed phage DNA is sequenced and compared with the full virome. Here, we show that AdsorpSeq allows for separation of phages based on receptor-adsorbing capabilities. Next, we applied AdsorpSeq to identify phages in a wastewater virome that adsorb to cell envelopes of nine bacteria, including important pathogens. We detected 26 adsorbed phages including common and rare members of the virome, a minority being related to previously characterized phages. We conclude that AdsorpSeq is an effective new tool for rapid characterization of environmental phage adsorption, with a proof-of-principle application to Gram-negative host cell envelopes.

INTRODUCTION

Bacteriophages (viruses that infect bacteria) are omnipresent and impact every ecosystem (Cobián Güemes et al., 2016). Their impact on microbial communities makes phages both useful and detrimental. On the one hand, they are potential bioengineered drug delivery systems (Karimi et al., 2016) and alternatives to antimicrobials (Nobrega et al., 2015). On the other hand, they spread bacterial pathogenicity (Chen and Novick, 2009) and disrupt food production chains like milk fermentations used in the dairy industry (Marcó et al., 2012). Phages also affect ecosystems at larger scales by controlling bacterial evolution and community structure, affecting, e.g., our microbiomes (De Sordi et al., 2019; Manrique et al., 2016), marine nutrient cycling through bacterial lysis (Corinaldesi et al., 2012; Danovaro et al., 2008), and global oxygen production by encoding photosynthesis genes that are expressed in cyanobacterial hosts (Sharon et al., 2009). Although their global importance makes understanding phage-host interactions crucial, most remain undetermined (Cobián Güemes et al., 2016).

A major reason for the mass of undetermined phage-host interactions is a shortage of readily applicable viromics techniques that can simultaneously (1) identify phages in an environmental sample and (2) link them to host(s). Unstudied phage genomes can be identified with metagenomics, but despite constant improvements (e.g., Ahlgren et al., 2017; Galiez et al., 2017; Liu et al., 2019; Mihara et al., 2016; Villarroel et al., 2016; Zhang et al., 2017) it remains challenging to predict to which hosts these phages adsorb, especially at low taxonomic levels (Edwards et al., 2016). Although CRISPR-Cas memory (spacers) and prophage regions can result in reliable host predictions (Edwards et al., 2016), many bacterial lineages do not have CRISPR-Cas systems (Burstein et al., 2016) and not all phages form prophages. Beyond computational approaches, phages can be linked to their host with isolation techniques like double-layer agar plates. Such techniques depend on lytic phages to form visible plaques (Abedon and Yin, 2009; Łoś et al., 2008; Serwer et al., 2007) and can be biased to phages with narrow host ranges (Guyader and Burch, 2008; Kauffman et al., 2018). These assays furthermore often employ a few highly related hosts, each needing a separate assay (Hyman and Abedon, 2010). Thus, available information on phage host range is limited. Other proposed methods include meta3C, which infers interactions based on physical proximity of phage and host DNA (Marbouty

[1]Theoretical Biology and Bioinformatics, Science4Life, Utrecht University, 3584 CH Utrecht, the Netherlands

[2]Department of Bionanoscience, Kavli Institute of Nanoscience, Delft University of Technology, 2629 HZ Delft, the Netherlands

[3]Lead Contact

*Correspondence: bedutilh@gmail.com

https://doi.org/10.1016/j.isci.2020.101439

et al., 2017), and those summarized in an earlier review (Edwards et al., 2016). However, such methods are generally cumbersome and thereby hard to implement.

An alternative approach to determining phage-host interactions is by focusing on the first step of the phage infection cycle, the adsorption of the phage to bacterial surface receptors. Although phage adsorption is not always followed by successful phage infection, it is a crucial step for successful infections and often specific (de Jonge et al., 2019a). Utilizing phage adsorption specificity could thus allow studies of phage-host interactions in environmental samples. This was recently shown through viral tagging (Deng et al., 2012; Džunková et al., 2019) where fluorescently labeled phages are added to bacteria. Bacteria bound by fluorescently labeled phages are isolated with fluorescence-activated cell sorting, and phage-bacterium pairs are sequenced. This approach allows abundant viruses to be linked to hosts, but it can remain challenging to identify phage-host links for rare members of the virome. Finally, viral tagging requires a specialized experimental setup.

Here, we rapidly identify phage-host pairings by linking cell envelope adsorption to phage sequencing and statistical analysis (adsorption sequencing or AdsorpSeq). AdsorpSeq allows identification of novel phages and their host interactions by exploiting differential migration of phages bound to host receptors and unbound phages in agarose gel electrophoresis. This enables selective sequencing of phages based on their interaction with cell envelopes of a specific host. Thereby multiple phages that interact with a given host can be rapidly and simultaneously identified. We show that model phages can be differentially identified based on the presence of their receptor molecule. Subsequently, we apply AdsorpSeq on a hospital wastewater virome and the cell envelopes of nine taxonomically distinct Gram-negative bacteria, uncovering 26 novel phage-host interactions with a range of abundances in the virome.

RESULTS AND DISCUSSION

Identification of Model Phages Based on Adsorption to Their Hosts

AdsorpSeq aims to selectively sequence phages based on their adsorption to bacterial cell envelopes. This is achieved by a five-step process (steps 1–5 in Figure 1A). First, a phage mixture is added to a cell envelope suspension that was isolated from a bacterium of interest (step 1 in Figure 1A). An incubation then allows phage adsorption to their receptors (2). Next, agarose gel electrophoresis separates bound and unbound phages (3). Unlike unbound phages, phages bound to cell envelope suspensions will migrate slower into agarose gels owing to the larger size and altered charge of the adsorption complex. The result is rapid separation of phages based on adsorption abilities. Finally, genomic material of bound phages is isolated from the gel (4) and sequenced (5). To validate AdsorpSeq, we tested the method with *Escherichia* phage λ and *Salmonella* phage P22 as two model phages with well-described adsorption properties. These two phages differ in host, morphology, and the receptor type. The receptor of *Siphoviridae* phage λ is the *E. coli* maltose pore protein LamB (Wang et al., 2000), whereas the *S. enterica* subspecies *enterica* (hereafter: *S. enterica*) lipopolysaccharide O-antigen chain serves as receptor for *Podoviridae* phage P22 (Andres et al., 2010). Because absorption to the bacterial cell envelope does not guarantee successful infection (de Jonge et al., 2019a) AdsorpSeq may detect phage-bacteria interactions beyond infecting host range. In addition, the isolation of bacterial membrane fractions may alter the interactions of the viruses with these membranes. For example, membrane fragments can form vesicles both in normal and inverted conformations (Coakley et al., 1977; Poole, 1993), therewith supplying the phage with the ability to bind to the cell envelope side that is normally pointed inward. However, as both phages are specific to their respective receptors (Andres et al., 2010; Randall-Hazelbauer and Shwartz, 1973), we could gauge preservation of adsorption specificity in AdsorpSeq.

As a first test, we added either phage λ or P22 to *E. coli* and *S. enterica* cell envelope suspensions. Upon adding non-host cell envelopes (e.g., adding *S. enterica* to λ), phage particles migrated into agarose gels, whereas adding host cell envelopes (e.g., adding *E. coli* to λ) resulted in phage particle retention around the sample slot at the origin of electrophoresis (Figure 1B). These gel regions consequently contained significantly more DNA when using host cell envelopes than when using non-host cell envelopes (two-tailed t test, $p < 0.05$). This showed that adsorption of phage particles to host cell envelopes prevented migration into agarose gels. To confirm receptor specificity, we repeated the experiment with a LamB-knockout strain (ΔLamB) (Baba et al., 2006), which resulted in phage λ losing adsorption ability (Figure 1C). These results agree with earlier studies (Andres et al., 2010) showing slower migration of phage particles into agarose gels after addition of purified phage receptor particles.

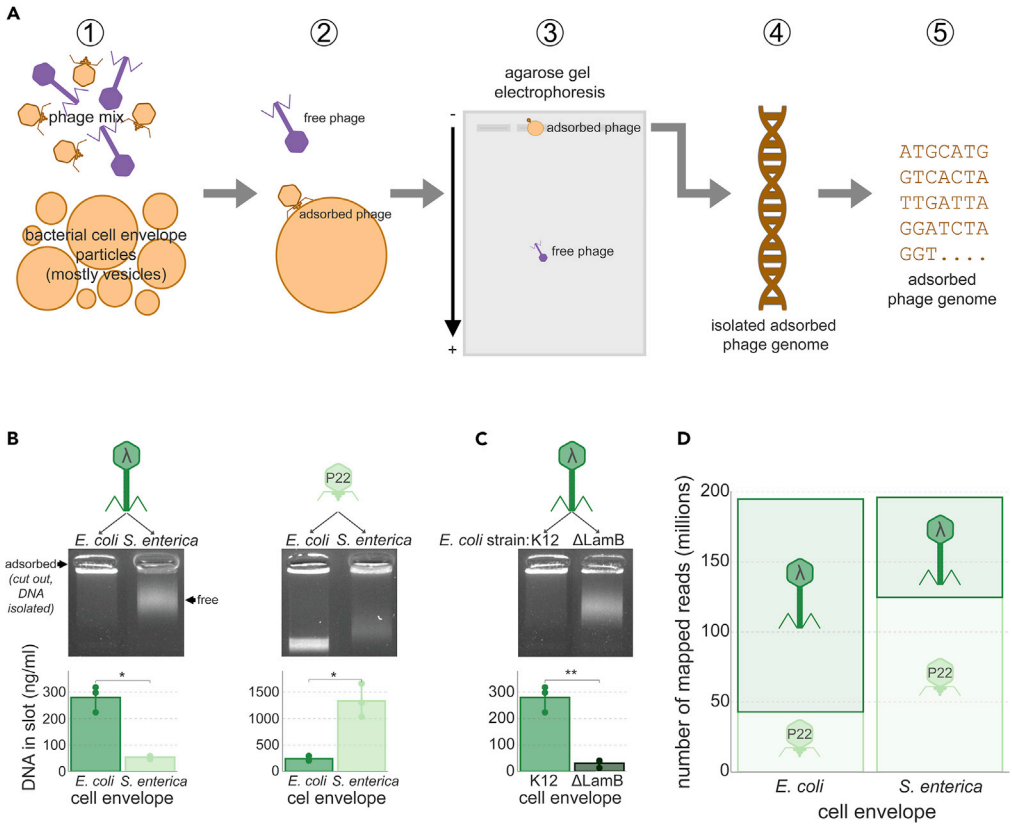


Figure 1. AdsorpSeq Allows the Selective Sequencing of Model Phages Based on Adsorption

(A) Schematic of AdsorpSeq. It shows the main steps of (1) mixing phages with bacterial cell envelopes, (2) allowing phages to adsorb to cell envelopes, (3) separating phages using agarose gel electrophoresis based on adsorbing capability, (4) isolating the genomes of adsorbed phages, and (5) sequencing genomes of adsorbed phages isolated from gels.

(B) Adsorption of phages λ and P22 to host cell envelopes hinders their migration into agarose gels. Agarose gels of phages λ and P22 after being added to cell envelope suspensions of *E. coli* K12 and *S. enterica* S1400, and bar graphs showing DNA quantities that were isolated from the gel slots at the top of the gels. Arrows indicate the location of free phages (migrated into the gel) and adsorbed phages (in the gel slot) in the first gel. This is identical in the other gels.

(C) AdsorpSeq maintains receptor molecule specificity of phage λ. Agarose gel of phage λ after being added to *E. coli* strain K12, to which it can adsorb, and *E. coli* ΔLamB, to which it cannot adsorb. Bar graph depicts DNA isolated from gel slots at the top of the gel. Note: although the smear seems visually stronger in the K12 lane, significantly more DNA was retained in the well containing the K12 envelope fraction than in the ΔLamB envelope fraction (see bar graphs).

(D) Applying AdsorpSeq to a mixture of phages leads to differentiation based on adsorbing capacity. Stacked bar graph showing the number of reads mapped to phages λ and P22 after AdsorpSeq was applied using an equal mixture of the two phages and cell envelopes from either *E. coli* K12 or *S. enterica* S1400. Significance levels according to a paired t test, error bars depict standard deviations, points are biological replicates. *$p < 0.05$, **$p < 0.01$.

After establishing that presence of host cell envelopes alters phage migration in agarose gels, we used this property to preferentially sequence phage genomes from a mixture based on the presence of host cell envelopes. We performed AdsorpSeq with a mixture of equal parts phage λ and P22, which we added to cell envelopes of either *E. coli* or *S. enterica*. Upon sequencing of genomic material isolated from agarose slices, two to three times more reads mapped to the phage λ genome than the P22 genome when *E. coli* cell envelopes were added and vice versa (Figure 1D). The resulting difference in phage genome abundance was highly significant (Fisher's exact test, $p = 2.5 \times 10^{-16}$). As AdsorpSeq was thus capable of discerning phage-host associations in a simple phage mixture, we next applied it to a complex environmental phage mixture.

AdsorpSeq Results in Selection of Unique Phage Subsets

We next applied AdsorpSeq to identify phages targeting specific bacterial cell envelopes in a complex virome derived from a hospital wastewater influent pipe (Figure S1A). In such environments, phages can be

found at concentrations of 10^8–10^{10} particles per milliliter (Otawa et al., 2007; Rosario et al., 2009; Tamaki et al., 2012; Wu and Liu, 2009). Phage adsorption targets consisted of cell envelope suspensions from nine taxonomically diverse Gram-negative bacteria, including three *Enterobacterales* (*Escherichia coli, Citrobacter freundii*, and *Klebsiella pneumoniae*), two *Pseudomonadales* (*Pseudomonas aeruginosa* and *Acinetobacter baumannii*), two *Bacteroidales* (*Bacteroides fragilis* and *Bacteroides dorei*), one *Burkholderiales* (*Ralstonia pickettii*), and one *Fusobacterales* (*Fusobacterium necrophorum*). All these bacteria are either part of the healthy human gut microbiome (e.g., *B. dorei*) (Huttenhower et al., 2012) or pathogens linked with hospital infections (e.g., *K. pneumoniae*) (Rice, 2008). We expect that AdsorpSeq may be most useful for Gram-negative bacteria, because these bacteria possess an outer membrane that is both easily isolated and contains the main receptors targeted during phage infections (Nobrega et al., 2018; Silva et al., 2016). Phage receptors in Gram-positive bacteria are often associated with the thick outer peptidoglycan layer (Dowah and Clokie, 2018) that can be difficult to break and isolate. Thus, in the following validation and application of AdsorpSeq, we focused our efforts on Gram-negative bacteria. Next to the virome treated with the cell envelopes of these nine bacteria, we sequenced the full untreated virome as control. As we increased DNA quantities of all samples by multiple displacement amplification (MDA), which alters apparent viral community compositions (Pinard et al., 2006), the full virome was sequenced both before and after MDA. This allowed us to gauge and correct for MDA effects during data analysis (below).

Sequencing of the nine cell envelope-treated samples and the two viromes (pre- and post-MDA) resulted in 138 Mbp of read-level data. A cross-assembly resulted in 23,373 contigs longer than 2,500 bp, representing 71.4% of the total dataset, as determined by mapping the reads back to the contigs (for annotated contig metadata and contig abundances, see Table S1). Taxonomic classification showed that the cross-assembly contained 1,111 viral contigs and a further 8,921 contigs that were taxonomically unclassified, whereas the remaining 13,341 contigs were mostly derived from bacteria, with a minority of archaeal and eukaryote contigs. The group of unclassified contigs likely reflects the large numbers of unstudied human gut phages (Shkoporov and Hill, 2019). We therefore combined the viral and unclassified sets to create a dataset of 10,032 confirmed and suspected viral contigs, which represented 51.5% of total reads. The percentage of reads represented by selected contigs fluctuated across the cell envelope-treated samples, ranging from 41.0% (*E. coli* treated) to 81.9% (*B. fragilis* treated) of reads (median: 70.2%, Figure S1B). These differences between the samples were a first indication that phage adsorption depended on the cell envelope suspension used.

Although there were 842 contigs with identical ends in the selected dataset representing putatively complete and circular genomes, many contigs also represented likely genomic fragments. We therefore assigned contigs with similar tetranucleotide usage patterns and read depth patterns across the nine cell envelope-treated samples to 1,058 viral populations. Binning of contigs is based on nucleotide usage and abundance signals. Thus, the bins do not necessarily represent single phage genomes, but rather groups of contigs with similar characteristics, which we refer to as viral populations. Some viral populations may contain fragments of different phages specifically binding to a host and similar phages that non-specifically bind to the same host. Jaccard distances based on viral population relative abundances showed distinct dissimilarity between the cell envelope-treated samples (Figure S1C). This indicated that each sample contained a unique set of viral populations and thus selected different phages. The viral populations with the highest abundance across the samples were also highly abundant in both virome controls (Figures S2A and S2B) suggesting that some phage particles are retained in the wells through non-specific interactions, as also observed for λ and P22 (Figure 1D). Perhaps some phages are physically prevented from migrating into agarose gels by the vesicles that cell envelope suspensions likely form (Poole, 1993). Although we did not achieve perfect separation between bound and unbound phage particles, we concluded from the evident dissimilarity in composition between the samples (Figure S1C) that selection of phages by AdsorpSeq is dependent on the type of bacterial cell envelope used.

Selection of Phages with Putative Adsorption Activity

To identify phages that specifically adsorb to one or more of the nine bacterial species, we selected viral populations that were overrepresented in one or more of the samples. To focus on the strongest adsorbing phages first, we defined overrepresented viral populations based on outlier analysis (see Methods for details), which meant that we selected viral populations with a relative abundance of at least 1.58 times higher in one sample than in the other eight samples. Relative abundance values from samples of the same bacterial taxonomic order were discounted when determining overrepresented viral populations to allow for

phages with a broad adsorption or infection host range. In total, 123 viral populations represented phages that specifically adsorb to the cell envelope fractions of one of the nine bacterial species (Figure S3A).

We next refined the selection of putatively adsorbing viral populations by applying two filters. The first filter removed viral populations that were positively selected for by MDA. MDA can result in efficient rolling circle amplification but has also been shown to lead to a bias for small single-stranded DNA (ssDNA) phage genomes (Kim and Bae, 2011; Probst et al., 2015; Yilmaz et al., 2010). This held true in our dataset, as comparing viral populations in the virome before and after MDA showed 10- to 100-fold higher amplification of ssDNA *Microviridae* than other phage families (Figure S3B). Note that 23,564 reads from the unamplified virome mapped to *Microviridae* contigs, showing that these small circular DNA viruses are abundant in the hospital virome and their detection does not fully depend on MDA. To reduce the impact of MDA bias, we thus filtered out 79 viral populations with strong MDA selection, leaving 44 viral populations that passed the MDA selection filter (Figure S3A).

In addition to MDA selection, we tested if certain phages were universally selected for by the AdsorpSeq technique (Figure S3C). This identified 18 putative adsorbing viral populations for which the relative abundance in all nine samples was higher than in post-MDA virome (Figure S3A). This methodological bias was highest among *Inoviridae* and *Microviridae* (Figure S3C). Although *in silico* evidence suggests that some phages may have very broad host ranges (Paez-Espino et al., 2016; Roux et al., 2016), most phages likely have a narrow host range spanning a few closely related strains within the same species or genus (de Jonge et al., 2019a; Džunková et al., 2019). We thus interpreted our findings as a methodological selection bias that may reflect the inability of a phage particle to migrate into agarose gels owing to large size, low charge, or non-specific interactions with bacterial cell envelopes. This may explain the stronger methodological selection pressure on *Inoviridae*, which have lipid membrane-adsorbing coat proteins (Stopar et al., 2003). The 18 viral populations that were under strong methodological bias were filtered out of the final selection of adsorbing vial populations.

After applying MDA and methodological selection filters, 26 viral populations with predicted adsorbing activity remained (Figure S3A). All 26 selected viral populations were highly specific to cell envelopes of a single bacterium and thus represented phages with a single predicted host. This was despite our allowance of broad host range at the order level in selecting viral populations but agrees with a recent report that found that broad host-range phages are rare in gut viromes based on single-cell viral tagging experiments (Džunková et al., 2019).

Notably, 22 of the 26 selected viral populations putatively adsorb to *Proteobacteria* cell envelopes, consistent with recent findings that *Proteobacteria* are the dominant bacterial phylum in global wastewater treatment plant microbiomes and are abundant in wastewater influent (Petrovich et al., 2019; Wu et al., 2019). It is thus likely that *Proteobacteria* phages are common in wastewater microbiomes, which supports our findings.

Selected Viral Populations Are Rare and Similar to *Proteobacteria* Phages

Next, we examined the final selection of 26 viral populations with adsorption predictions. Their relative abundance in the virome ranged from 0.0001% to 1%, covering the spectrum from relatively rare to relatively abundant in the virome before and after MDA (Figure 2A). Characterization of the ORFs by direct homology searches showed that the majority of ORFs (64% of total) in selected viral populations had no significant similarity to protein sequences in the National Institute for Biotechnology Information (NCBI) non-redundant (nr) database (Agarwala et al., 2017) (Figure 2B). From these findings we concluded that AdsorpSeq can be used to identify adsorption hosts of both common and rare uncharacterized environmental phages.

To assess AdsorpSeq host predictions, we performed a whole genome clustering of all viral populations (including those without adsorption predictions) with all characterized phages whose genome sequences were available in the NCBI bacterial and viral RefSeq V85 database (Pruitt et al., 2007) (Figure 2C). First, we observed that the contigs of the viral populations identified with AdsorpSeq clustered together, confirming our contig binning approach. Second, most of the 22 selected viral populations that were predicted to adsorb to *Proteobacteria* species were similar to characterized *Proteobacteria* phages. This supports the notion that the viral populations that were detected with AdsorpSeq on the cell envelopes of specific bacterial pathogens reflect naturally occurring phages capable of infecting these bacteria.

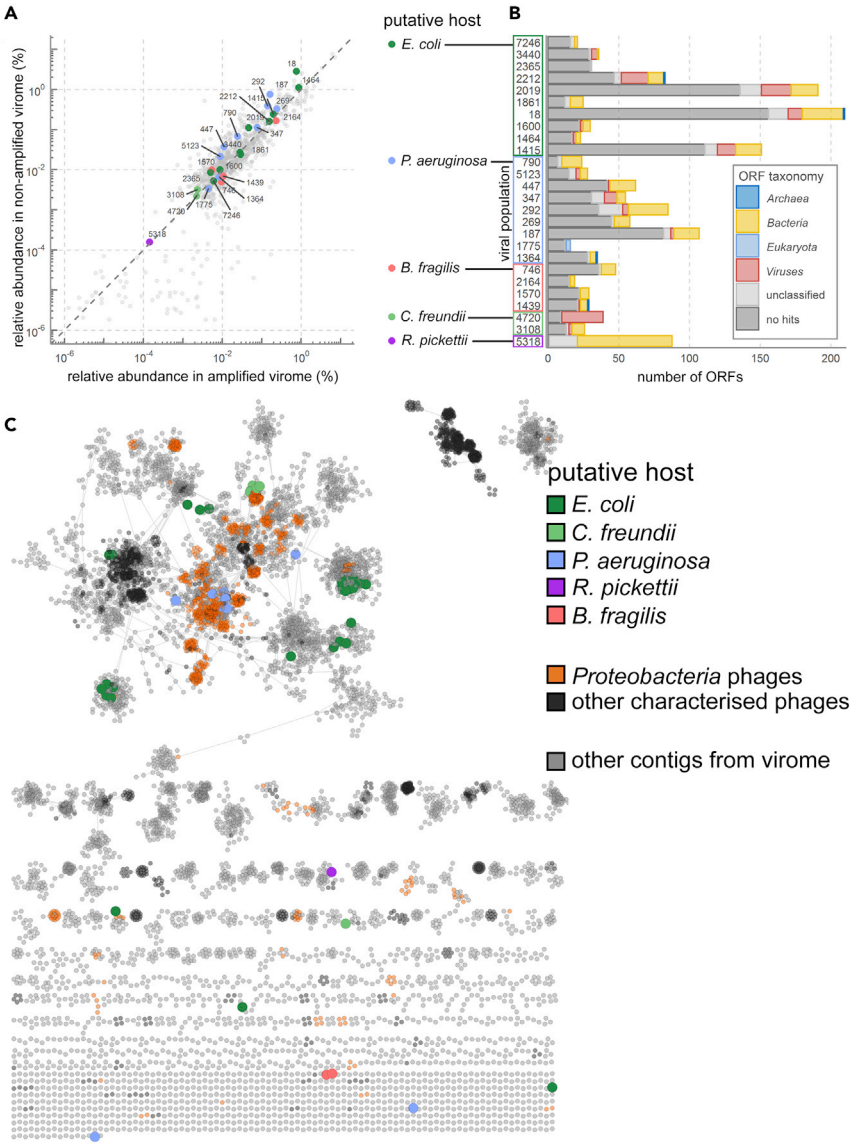


Figure 2. Most Selected Viral Populations Represent Rare and Uncharacterized Viral Sequences

(A) Relative abundance of selected viral populations with adsorption predictions in the virome before and after MDA shows that AdsorpSeq is not biased for abundant or rare phage sequences. Numbers next to data points show the viral population number.

(B) ORF-level taxonomical predictions using CAT show most ORFs from selected viral populations have no similarities in the NCBI nr protein sequence database (dark gray). Some contigs had database hits but could not be classified because the hits involved proteins from different superkingdoms. These are labeled as unclassified (light gray).

(C) The hospital wastewater virome contained a large diversity of uncharacterized phage sequences, as shown by a gene-sharing network of 10,032 viral contigs and all phage genomes in the NCBI viral RefSeq database (Pruitt et al., 2007), made using vContact2 (Bin Jang et al., 2019). Large colored contigs represent those in the final selection of 26 putative adsorbing viral populations. *Proteobacteria*-infecting characterized phages are orange.

Selected Viral Populations with Similarity to Characterized Phages

To further assess the phage-host associations detected by AdsorpSeq, we searched for matches between contigs in selected viral populations and bacterial CRISPR-Cas spacers. We identified 1.4 million spacers predicted from bacteria in the Pathosystems Resource Integration Center (PATRIC) database (Wattam et al., 2014) and queried them against contigs in the selected viral populations. No full-length identical protospacers were detected on any of the viral contigs. Two contigs from viral populations adsorbing to

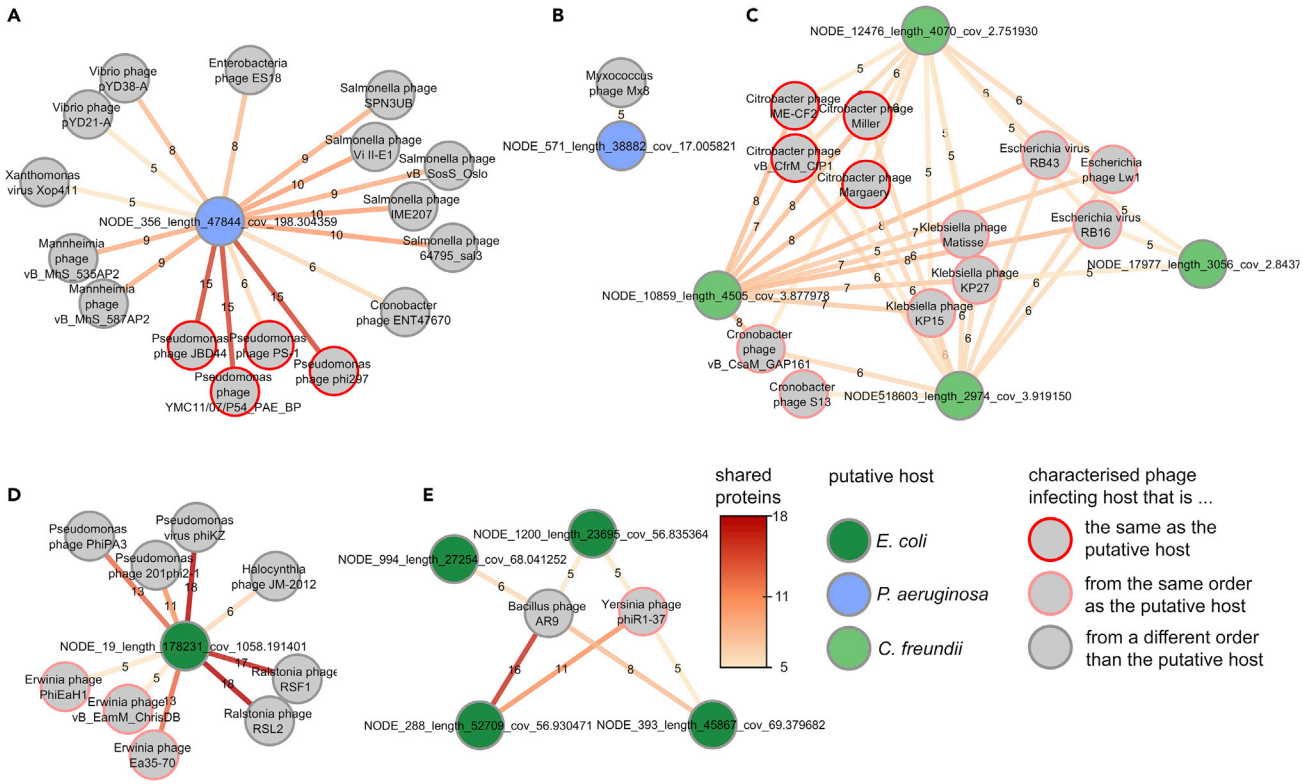


Figure 3. Selected Viral Populations Related to Characterized Phages

Protein-sharing networks of viral populations show their relationships to characterized phages. ORFs from selected viral populations were used in BLASTp searches against proteins of phages in the viral RefSeq database (Pruitt et al., 2007). Bubbles are phages. Edge color and labels show similar protein counts (E-value $\leq 10^{-5}$).

(A) *P. aeruginosa*-adsorbing viral population 292.

(B) *P. aeruginosa*-adsorbing viral population 447.

(C) *C. freundii*-adsorbing viral population 4720. One additional contig did not share protein similarity to characterized phages.

(D) *E. coli*-adsorbing viral population 18.

(E) *E. coli*-adsorbing viral population 2019.

Pseudomonas aeruginosa contained spacer hits with a single mismatch each. These hits originated from CRISPR-Cas arrays encoded on the genomes of *Aeromonas caviae* and *Tolumonas auensis*, which are members of the same taxonomical class as *P. aeruginosa* (*Gammaproteobacteria*). Several examples have been observed of phage genera that adsorb to differently related bacteria, such as Tequatroviruses (Nolan et al., 2006), Gap227likeviruses (Wang et al., 2016), and Plpelikeviruses (Comeau et al., 2012; Walker et al., 2019). Alternatively, cell envelope adsorption does not necessitate successful infection, as many factors play a role in the completion of the infection cycle (de Jonge et al., 2019a), including differences in transcriptional programs between hosts (Howard-Varona et al., 2018) or the presence of molecular defense systems (Hampton et al., 2020; Labrie et al., 2010). No other selected viral populations contained contigs with spacer hits with fewer than five mismatches. As five mismatches corresponds to a ~50% false discovery rate at the species level (Edwards et al., 2016), hits with more mismatches were not considered for analysis.

Next, we placed the viral populations identified with AdsorpSeq in the context of characterized phages by using protein sharing networks. We narrowed our search to contigs that contained over five ORFs with known homologs, yielding five selected viral populations with extensive similarity to characterized phages (Figure 3 and Table S3).

First, *P. aeruginosa*-adsorbing viral population 292 contained a single circular contig on which 15/85 ORFs (17.6%) were similar to proteins from the known *Pseudomonas* phages JBD44, phi297, and YMC11/07/P54_PAE_BP (Figure 3A). Notably, phi297 infects the *P. aeruginosa* strain used in this study (Bourkal'tseva et al., 2011). The fifteen shared proteins included the often conserved terminase (Low et al., 2019; Serwer

et al., 2004) and several adjacent genes (Figure 4A and Table S4), stressing the relatedness between these phages. Following earlier practice (Lavigne et al., 2008), this protein similarity would classify these phages into the same taxonomical family. Although the conservative CAT ORF-level predictions provided no taxonomic classifications for most ORFs in viral population 292, direct homology searches found significant similarity (BLASTp, e-value $\leq 10^{-5}$) to proteins from *Pseudomonas* bacteria for 45 of the 85 ORFs (53%, Figure 4A and Table S4) from viral population 292. This viral population thus represents a novel *Pseudomonas* phage identified in the hospital sewage inlet by AdsorpSeq with *P. aeruginosa* cell envelopes as bait.

A second viral population that also adsorbed to *P. aeruginosa* (447) contained a contig of 38,882 bp that had 5/62 ORFs with similarity to *Myxococcus* phage Mx8 (Figure 3B), which translates to less than 5% of 85 Mx8 ORFs. The similarity between viral population 447 and phage Mx8 is thus limited to a small number of genes, which, although not adjacent, all belong to the structural section of the Mx8 genome (Table S3). We suggest that these limited shared proteins reflect shared gene cassettes (Hatfull and Hendrix, 2011; Lavigne et al., 2008).

In addition to the *Pseudomonas*-adsorbing viral populations, viral population 4720 represents a *Citrobacter*-adsorbing phage. This viral population contains five short (2,974–4,505 bp) contigs. The four longest of these had significant protein sequence similarity (BLASTp, e-value $\leq 10^{-5}$) to at least five ORFs from several *Citrobacter* phages (Figure 3C), whereas all five contigs showed full-length sequence similarity to *Citrobacter* phage Margaery (tBLASTx, e-value ≤ 0.001, Figure 4B). Combined, they shared similarity to 25/280 (9%) *Citrobacter* phage Margaery proteins and several other T4-like *Citrobacter* phages (Figure 4B). As these known *Citrobacter* phages have genomes over 150,000 bp, we suggest that viral population 4720 may represent fragments of a larger *Citrobacter* phage genome.

Finally, two selected *E. coli*-adsorbing viral populations (18 and 2019) shared five to eighteen ORFs with several known jumbo phages (Figures 3D and 3E). These included *Yersinia* phage phiR1-37 and *Erwinia* phage Ea35-70, both of which infect bacteria from the same taxonomic order as *E. coli* (*Gammaproteobacteria*). Notably, *E. coli* phages like T4 (Tétart et al., 1996) and T3 (Garcia et al., 2003) need only limited genomic alterations to extend their host range to *Yersinia* species. To gain added insight into the relation of these phages to other jumbo phages, we built a phylogeny of their terminase genes (Figure 4C). Earlier analysis showed that jumbo phages with phylogenetically closely related terminases often infect related hosts (Yuan and Gao, 2017). For our analysis, we gathered 148 sequences from the NCBI nr database (Agarwala et al., 2017) and 74 jumbo phage proteins from a recent study (Al-Shayeb et al., 2020) that had significant homology to viral population 18 and 2019 terminases (BLASTp, e-value $\leq 10^{-5}$). The resulting tree placed viral population 2019 in a branch that mostly held recently described jumbo phages (Al-Shayeb et al., 2020), with hosts predicted to belong to the *Firmicutes*. Viral population 18 terminase belonged to a more diverse clade with multiple *Ralstonia* phages, a single *Alteromonadaceae* bacterial sequence (a *γ-proteobacteria* species), and multiple jumbo phages previously predicted to infect proteobacterial hosts (Al-Shayeb et al., 2020). These phages might interact with a surface element that is common to all these hosts. This hypothesis is reinforced by the fact that *Ralstonia* were long assigned to the genus *Pseudomonas* within the γ-proteobacteria, despite differences in membrane composition (Yabuuchi et al., 1995) and lipopolysaccharide structure (Zdorovenko et al., 2008). Alternatively, in similar fashion to *Enterobacteria* phage phi92 (Schwarzer et al., 2012), they may harbor multiple receptor-adsorbing proteins in their large genomes. Besides the terminase, the circular contig in viral population 18 notably also contained an FtsZ homolog. In *Pseudomonas* jumbo phages, FtsZ-proteins are part of a nucleus-like defense mechanism (Chaikeeratisak et al., 2017; Mendoza et al., 2019), together with a nucleus-forming protein. Interestingly, protein homology searches did not identify similar nucleus-forming proteins in viral population 18, although this population consists of a single circular contig. It may thus contain a yet unknown system like the nucleus-like defense mechanism found in *Pseudomonas* jumbo phages or use the FtsZ homolog in a system that is different altogether.

The results discussed above represent 6 of 26 AdsorpSeq-selected viral populations with at least five protein similarities to characterized phages. The ORFs encoded on the remaining 20 selected viral populations represented novel or highly divergent proteins, unrelated to previously characterized phages. Together, our results underscore the ability of AdsorpSeq to uncover environmental phages and their potential host associations, without bias for known or abundant phages.

Conclusion

Despite recent advances in viromics, linking environmental phage sequences to hosts remains problematic (Edwards et al., 2016). Here, we presented AdsorpSeq, a rapid method for detecting phage adsorption to

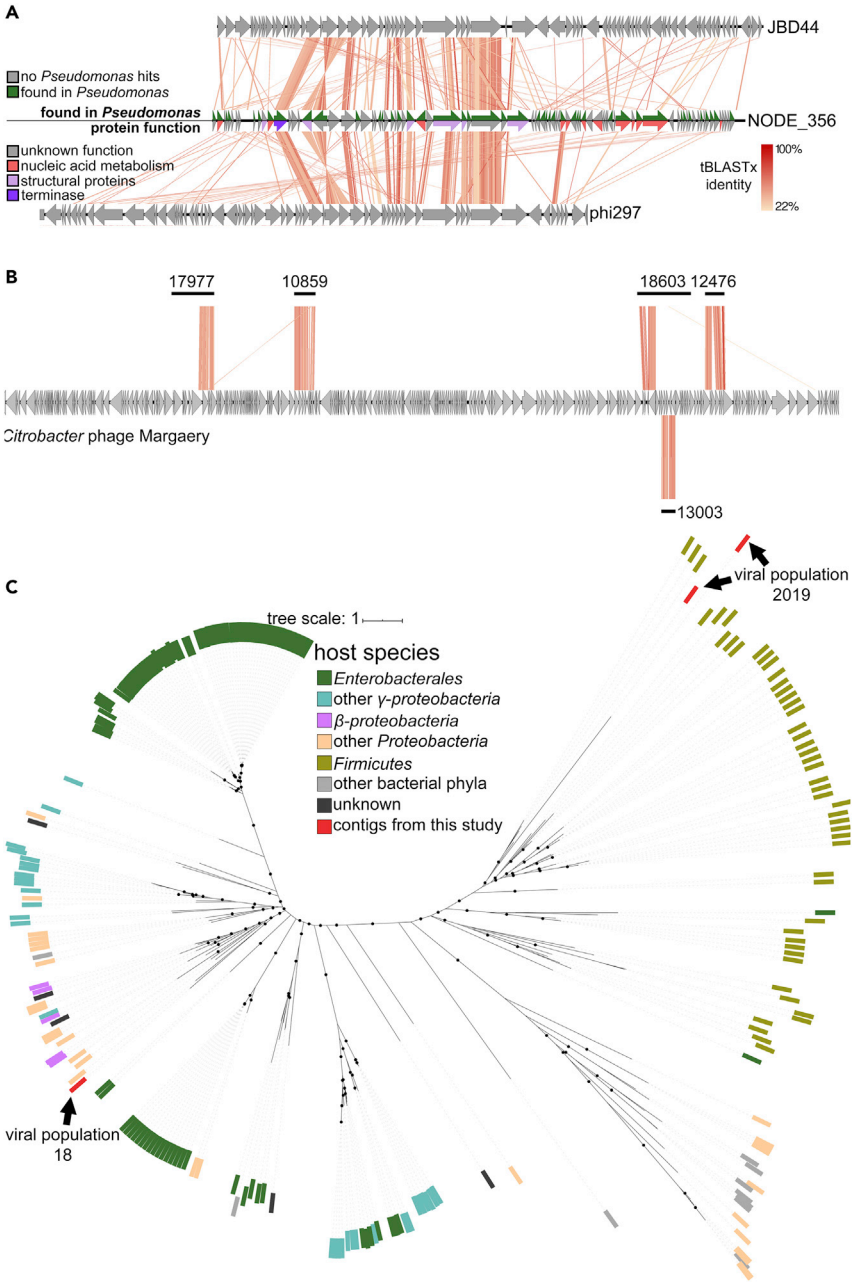


Figure 4. Several Viral Populations and their Relations to Characterized Families

(A) Similarity of viral population 447, containing contig 356, to *Pseudomonas* phages JBD44 and phi297. Depicted is a whole genome comparison made using Easyfig (Sullivan et al., 2011). In the line representing contig 356, the top half shows ORFs with BLASTp hit against *Pseudomonas* bacteria proteins in the NCBI nr database, whereas the bottom half shows protein function.

(B) Similarity between five contigs from *C. freundii*-adsorbing viral population 4720 and T4-like *Citrobacter* phage Margaery, as shown by genome comparisons made using Easyfig (Sullivan et al., 2011). Numbers indicate contig numbers, contig 13,003 was placed below phage *Margaery* as it overlaps with contig 18,603. Colors indicate tBLASTx hits and use the same legend as (A).

(C) The relation of *E. coli*-adsorbing viral populations 18 and 2019 to jumbo phages displayed in an unrooted approximate maximum likelihood tree of jumbo phage terminases. Dots on branches represent ultrafast bootstrap support of ≥85 (Hoang et al., 2018).

host cellular envelopes. AdsorpSeq is easily implementable, as besides sequencing it uses only commonly available laboratory methods, such as cell disruption and agarose gel electrophoresis.

In the current study, we validated AdsorpSeq on model bacteriophages. This showed that AdsorpSeq can separate phages based on the presence of a specific host receptor in bacterial cell envelopes. Future improvements to AdsorpSeq could exploit this feature, for instance, through heterologous expression of certain bacterial features of interest and uncovering phages that adsorb to them. Therewith, AdsorpSeq could aid in uncovering interactions between phages and microbial surface molecules like antimicrobial efflux pumps. AdsorpSeq uncouples analysis of phage adsorption and phage infection, potentially allowing for analysis of phage-bacterium interactions that do not result in infection, for instance, owing to intracellular defense systems (Hampton et al., 2020).

The experiments laid out in this study identified 26 phage-host interactions in a hospital wastewater virome. Most of these were predictions of phages adsorbing to *Proteobacteria*, which is consistent with recent findings that these bacteria are ubiquitous and abundant in global waste water communities (Petrovich et al., 2019; Wu et al., 2019). Several of the putative adsorbing viral populations represented rare members of the sampled hospital wastewater virome. Although rare here, these viruses may still be important in another time and place as they become transiently dominant through ecological dynamics (Arkhipova et al., 2018). Moreover, expanding our knowledge to include rare members of the virosphere is important to address fundamental questions about the evolution of viral genes and genomes (de Jonge et al., 2019b; Mavrich and Hatfull, 2017), to identify candidate viruses with potentially promising genomic properties for phage therapy (Nobrega et al., 2015) and for targeted monitoring (Metsky et al., 2019). This underscores the necessity for novel methods such as AdsorpSeq with which phage-host interactions can be rapidly assessed, even for uncommon phages within a complex environmental mixture.

Limitations of the Study

Here we presented AdsorpSeq as a method to link environmental bacteriophages to their bacterial hosts. An important caveat for future AdsorpSeq applications is that it determined whether phages bind to the cell surface of a host cell, which is not necessarily followed by a successful infection. Investigations on the ability of phages with AdsorpSeq binding predictions to complete infections, such as phage isolations, would thus be a useful addition of the method. Of course, targeted phage isolation is a non-trivial task, and an inability to observe infection in methods such as plaque assays does not necessarily indicate an inability to infect, as phage infections can be (pseudo-)lysogenic or chronic and as not all lytic phages form plaques (Serwer et al., 2007). Additionally, we focused our efforts on Gram-negative bacterial envelopes, whereas the applicability of AdsorpSeq to Gram-positive bacteria remains to be tested.

Resource Availability

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Bas E. Dutilh (bedutilh@gmail.com)

Materials Availability

Materials and protocols used in this study are available from the authors upon request. This study did not generate new unique reagents.

Data and Code Availability

The accession number for the sequencing data project reported in this paper is ENA: PRJEB37817. The accession number of sequencing reads reported in this paper are ENA: ERS4427880–ERS4427890, whereas the accession number of cross-assembled contigs reported in this paper is ENA: ERZ1305919.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101439.

ACKNOWLEDGMENTS

We thank Utrecht Sequencing Facility for providing sequencing service and data. Utrecht Sequencing Facility is subsidized by the University Medical Centre Utrecht, Hubrecht Institute, Utrecht University and The Netherlands X-omics Initiative (Netherlands Organisation for Scientific Research (NWO) project 184.034.019). We further thank Reinier de Graaf hospital in Delft (the Netherlands) for providing wastewater samples and B.E. Estrada Bonilla (TU Delft) for assisting in sample filtration. P.A.d.J., F.A.B.v.M., and B.E.D. were supported by NWO Vidi grant 864.14.004. F.A.B.v.M. and B.E.D. were supported by and ERC Consolidator grant 865694: DiversiPHI. F.L.N. was supported by NWO Veni grant 016.Veni.181.092, and S.J.J.B. was supported by Vici grant VI.C.182.027 and European Research Council (ERC) Stg grant 639707.

AUTHOR CONTRIBUTIONS

P.A.d.J., F.A.B.v.M., and B.E.D. designed and performed the experiments. P.A.d.J. and F.A.B.v.M. analyzed the data. A.R.C., F.L.N., and S.J.J.B. assisted with experimental design and data interpretation. P.A.d.J. and B.E.D. prepared the figures and wrote the paper with input from all other authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 11, 2020
Revised: July 23, 2020
Accepted: August 3, 2020
Published: September 25, 2020

REFERENCES

Abedon, S.T., and Yin, J. (2009). Bacteriophage plaques: theory and analysis. In Bacteriophages: Methods and Protocols, Volume 1: Isolation, Characterization, and Interactions, M.R.J. Clokie and A. Kropinski, eds. (Humana Press), pp. 161–174.

Agarwala, R., Barrett, T., Beck, J., Benson, D.A., Bollin, C., Bolton, E., Bourexis, D., Brister, J.R., Bryant, S.H., Canese, K., et al. (2017). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. *45*, D12–D17.

Ahlgren, N.A., Ren, J., Lu, Y.Y., Fuhrman, J.A., and Sun, F. (2017). Alignment-free d2* oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. Nucleic Acids Res. *45*, 39–53.

Al-Shayeb, B., Sachdeva, R., Chen, L.X., Ward, F., Munk, P., Devoto, A., Castelle, C.J., Olm, M.R., Bouma-Gregson, K., Amano, Y., et al. (2020). Clades of huge phages from across Earth's ecosystems. Nature *578*, 425–431.

Andres, D., Hanke, C., Baxa, U., Seul, A., Barbirz, S., and Seckler, R. (2010). Tailspike interactions with lipopolysaccharide effect DNA ejection from phage P22 particles in vitro. J. Biol. Chem. *285*, 36768–36775.

Arkhipova, K., Skvortsov, T., Quinn, J.P., McGrath, J.W., Allen, C.C.R., Dutilh, B.E., McElarney, Y., and Kulakov, L.A. (2018). Temporal dynamics of uncultured viruses: a new dimension in viral diversity. ISME J. *12*, 199–211.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. *2*, 2006.0008.

Bin Jang, H., Bolduc, B., Zablocki, O., Kuhn, J.H., Roux, S., Adriaenssens, E.M., Brister, J.R., Kropinski, A.M., Krupovic, M., Lavigne, R., et al. (2019). Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. Nat. Biotechnol. *37*, 632–639.

Bourkal'tseva, M.V., Krylov, S.V., Kropinski, A.M., Pleteneva, E.A., Shaburova, O.V., and Krylov, V.N. (2011). Bacteriophage phi297, a new species of Pseudomonas aeruginosa temperate phages with a mosaic genome: potential use in phage therapy. Russ. J. Genet. *47*, 794–798.

Burstein, D., Sun, C.L., Brown, C.T., Sharon, I., Anantharaman, K., Probst, A.J., Thomas, B.C., and Banfield, J.F. (2016). Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. Nat. Commun. *7*, 1–8.

Chaikeeratisak, V., Nguyen, K., Khanna, K., Brilot, A.F., Erb, M.L., Coker, J.K.C., Vavilina, A., Newton, G.L., Buschauer, R., Pogliano, K., et al. (2017). Assembly of a nucleus-like structure during viral replication in bacteria. Science *355*, 194–197.

Chen, J., and Novick, R.P. (2009). Phage-mediated intergeneric transfer of toxin genes. Science *323*, 139–141.

Coakley, W.T., Bater, A.J., and Lloyd, D. (1977). Disruption of micro-organisms. Adv. Microb. Physiol. *16*, 279–341.

Cobián Güemes, A.G., Youle, M., Cantú, V.A., Felts, B., Nulton, J., and Rohwer, F. (2016). Viruses as winners in the game of life. Annu. Rev. Virol. *3*, 197–214.

Comeau, A.M., Tremblay, D., Moineau, S., Rattei, T., Kushkina, A.I., Tovkach, F.I., Krisch, H.M., and Ackermann, H.W. (2012). Phage morphology recapitulates phylogeny: the comparative genomics of a new group of myoviruses. PLoS One *7*, 1–11.

Corinaldesi, C., Dell'Anno, A., and Danovaro, R. (2012). Viral infections stimulate the metabolism and shape prokaryotic assemblages in submarine mud volcanoes. ISME J. *6*, 1250–1259.

Danovaro, R., Dell'Anno, A., Corinaldesi, C., Magagnini, M., Noble, R., Tamburini, C., and Weinbauer, M. (2008). Major viral impact on the functioning of benthic deep-sea ecosystems. Nature *454*, 1084–1087.

Deng, L., Gregory, A., Yilmaz, S., Poulos, B.T., Hugenholtz, P., and Sullivan, M.B. (2012). Contrasting life strategies of viruses that infect photo- and heterotrophic bacteria, as revealed by viral tagging. MBio *3*, e00373-12.

Dowah, A.S.A.A., and Clokie, M.R.J.J. (2018). Review of the nature, diversity and structure of bacteriophage receptor binding proteins that target Gram-positive bacteria. Biophys. Rev. *1*, 1–8.

Džunková, M., Low, S.J., Daly, J.N., Deng, L., Rinke, C., and Hugenholtz, P. (2019). Defining the human gut host–phage network through single-cell viral tagging. Nat. Microbiol. *4*, 2192–2203.

Edwards, R.A., McNair, K., Faust, K., Raes, J., and Dutilh, B.E. (2016). Computational approaches to predict bacteriophage-host relationships. FEMS Microbiol. Rev. *40*, 258–272.

Galiez, C., Siebert, M., Enault, F., Vincent, J., and Söding, J. (2017). WIsH: who is the host?

Predicting prokaryotic hosts from metagenomic phage contigs. Bioinformatics *33*, 3113–3114.

Garcia, E., Elliott, J.M., Ramanculov, E., Chain, P.S.G.G., Chu, M.C., and Molineux, I.J. (2003). The genome sequence of *Yersinia pestis* bacteriophage ϕA1122 reveals an intimate history with the coliphage T3 and T7 genomes. J. Bacteriol. *185*, 5248–5262.

Guyader, S., and Burch, C.L. (2008). Optimal foraging predicts the ecology but not the evolution of host specialization in bacteriophages. PLoS One *3*, e1946.

Hampton, H.G., Watson, B.N.J., and Fineran, P.C. (2020). The arms race between bacteria and their phage foes. Nature *577*, 327–336.

Hatfull, G.F., and Hendrix, R.W. (2011). Bacteriophages and their genomes. Curr. Opin. Virol. *1*, 298–303.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. Molecular biology and evolution. Mol. Biol. Evol. *35*, 518–522.

Howard-Varona, C., Hargreaves, K.R., Solonenko, N.E., Markillie, L.M., White, R.A., Brewer, H.M., Ansong, C., Orr, G., Adkins, J.N., and Sullivan, M.B. (2018). Multiple mechanisms drive phage infection efficiency in nearly identical hosts. ISME J. *12*, 1605–1618.

Huttenhower, C., Fah Sathirapongsasuti, J., Segata, N., Gevers, D., Earl, A.M., Fitzgerald, M.G., Young, S.K., Zeng, Q., Alm, E.J., Alvarado, L., et al. (2012). Structure, function and diversity of the healthy human microbiome. Nature *486*, 207–214.

Hyman, P., and Abedon, S.T. (2010). Bacteriophage host range and bacterial resistance. In Advances in Applied Microbiology (Elsevier Inc.), pp. 217–248.

de Jonge, P.A., Nobrega, F.L., Brouns, S.J.J., and Dutilh, B.E. (2019a). Molecular and evolutionary determinants of bacteriophage host range. Trends Microbiol. *27*, 51–63.

de Jonge, P.A., von Meijenfeldt, F.A.B., van Rooijen, L.E., Brouns, S.J.J., and Dutilh, B.E. (2019b). Evolution of BACON domain tandem repeats in crAssphage and novel gut bacteriophage lineages. Viruses *11*, 1085.

Karimi, M., Mirshekari, H., Moosavi Basri, S.M., Bahrami, S., Moghoofei, M., and Hamblin, M.R. (2016). Bacteriophages and phage-inspired nanocarriers for targeted delivery of therapeutic cargos. Adv. Drug Deliv. Rev. *106*, 45–62.

Kauffman, K.M., Hussain, F.A., Yang, J., Arevalo, P., Brown, J.M., Chang, W.K., VanInsberghe, D., Elsherbini, J., Sharma, R.S., Cutler, M.B., et al. (2018). A major lineage of non-tailed dsDNA viruses as unrecognized killers of marine bacteria. Nature *554*, 118–122.

Kim, K.H., and Bae, J.W. (2011). Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. Appl. Environ. Microbiol. *77*, 7663–7668.

Labrie, S.J., Samson, J.E., and Moineau, S. (2010). Bacteriophage resistance mechanisms. Nat. Rev. Microbiol. *8*, 317–327.

Lavigne, R., Seto, D., Mahadevan, P., Ackermann, H.W., and Kropinski, A.M. (2008). Unifying classical and molecular taxonomic classification: analysis of the Podoviridae using BLASTP-based tools. Res. Microbiol. *159*, 406–414.

Liu, D., Ma, Y., Jiang, X., and He, T. (2019). Predicting virus-host association by Kernelized logistic matrix factorization and similarity network fusion. BMC Bioinformatics *20*, 1–10.

Low, S.J., Džunková, M., Chaumeil, P.-A., Parks, D.H., and Hugenholtz, P. (2019). Evaluation of a concatenated protein phylogeny for classification of tailed double-stranded DNA viruses belonging to the order Caudovirales. Nat. Microbiol. *4*, 1306–1315.

Łoś, J.M., Golec, P., Węgrzyn, G., Węgrzyn, A., and Łoś, M. (2008). Simple method for plating *Escherichia coli* bacteriophages forming very small plaques or no plaques under standard conditions. Appl. Environ. Microbiol. *74*, 5113–5120.

Manrique, P., Bolduc, B., Walk, S.T., van der Oost, J., de Vos, W.M., and Young, M.J. (2016). Healthy human gut phageome. Proc. Natl. Acad. Sci. U S A *113*, 10400–10405.

Marbouty, M., Baudry, L., Cournac, A., and Koszul, R. (2017). Scaffolding bacterial genomes and probing host-virus interactions in gut microbiome by proximity ligation (chromosome capture) assay. Sci. Adv. *3*, e1602105.

Marcó, M.B., Moineau, S., and Quiberoni, A. (2012). Bacteriophages and dairy fermentations. Bacteriophage *2*, 149–158.

Mavrich, T.N., and Hatfull, G.F. (2017). Bacteriophage evolution differs by host, lifestyle and genome. Nat. Microbiol. *2*, 1–9.

Mendoza, S.D., Nieweglowska, E.S., Govindarajan, S., Leon, L.M., Berry, J.D., Tiwari, A., Chaikeeratisak, V., Pogliano, J., Agard, D.A., and Bondy-Denomy, J. (2019). A bacteriophage nucleus-like compartment shields DNA from CRISPR nucleases. Nature *577*, 244–248.

Metsky, H.C., Siddle, K.J., Gladden-Young, A., Qu, J., Yang, D.K., Brehio, P., Goldfarb, A., Piantadosi, A., Wohl, S., Carter, A., et al. (2019). Capturing sequence diversity in metagenomes with comprehensive and scalable probe design. Nat. Biotechnol. *37*, 160–168.

Mihara, T., Nishimura, Y., Shimizu, Y., Nishiyama, H., Yoshikawa, G., Uehara, H., Hingamp, P., Goto, S., and Ogata, H. (2016). Linking virus genomes with host taxonomy. Viruses *8*, 10–15.

Nobrega, F.L., Costa, A.R., Kluskens, L.D., and Azeredo, J. (2015). Revisiting phage therapy: new applications for old resources. Trends Microbiol. *23*, 185–191.

Nobrega, F.L., Vlot, M., de Jonge, P.A., Dreesens, L.L., Beaumont, H.J.E., Lavigne, R., Dutilh, B.E., and Brouns, S.J.J. (2018). Targeting mechanisms of tailed bacteriophages. Nat. Rev. Microbiol. *16*, 760–773.

Nolan, J.M., Petrov, V., Bertrand, C., Krisch, H.M., and Karam, J.D. (2006). Genetic diversity among five T4-like bacteriophages. Virol. J. *3*, 1–15.

Otawa, K., Lee, S.H., Yamazoe, A., Onuki, M., Satoh, H., and Mino, T. (2007). Abundance, diversity, and dynamics of viruses on microorganisms in activated sludge processes. Microb. Ecol. *53*, 143–152.

Paez-Espino, D., Eloe-Fadrosh, E.A., Pavlopoulos, G.A., Thomas, A.D., Huntemann, M., Mikhailova, N., Rubin, E., Ivanova, N.N., and Kyrpides, N.C. (2016). Uncovering Earth's virome. Nature *536*, 425–430.

Petrovich, M.L., Ben Maamar, S., Hartmann, E.M., Murphy, B.T., Poretsky, R.S., and Wells, G.F. (2019). Viral composition and context in metagenomes from biofilm and suspended growth municipal wastewater treatment plants. Microb. Biotechnol. *12*, 1324–1336.

Pinard, R., de Winter, A., Sarkis, G.J., Gerstein, M.B., Tartaro, K.R., Plant, R.N., Egholm, M., Rothberg, J.M., and Leamon, J.H. (2006). Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. BMC Genomics *7*, 216.

Poole, R.K. (1993). The isolation of membranes from bacteria. Methods Mol. Biol. *19*, 109–122.

Probst, A.J., Weinmaier, T., DeSantis, T.Z., Santo Domingo, J.W., and Ashbolt, N. (2015). New perspectives on microbial community distortion after whole-genome amplification. PLoS One *10*, 1–16.

Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. *35*, 61–65.

Randall-Hazelbauer, L., and Shwartz, M. (1973). Isolation of the bacteriophage lambda receptor from E. coli. J. Bacteriol. *116*, 1436–1446.

Rice, L.B. (2008). Federal funding for the study of antimicrobial resistance in nosocomial pathogens: No ESKAPE. J. Infect. Dis. *197*, 1079–1081.

Rosario, K., Nilsson, C., Lim, Y.W., Ruan, Y., and Breitbart, M. (2009). Metagenomic analysis of viruses in reclaimed water. Environ. Microbiol. *11*, 2806–2820.

Roux, S., Brum, J.R., Dutilh, B.E., Sunagawa, S., Duhaime, M.B., Loy, A., Poulos, B.T., Solonenko, N., Lara, E., Poulain, J., et al. (2016). Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. Nature *537*, 689–693.

Schwarzer, D., Buettner, F.F.R., Browning, C., Nazarov, S., Rabsch, W., Bethe, A., Oberbeck, A., Bowman, V.D., Stummeyer, K., Muhlenhoff, M., et al. (2012). A multivalent adsorption apparatus explains the broad host range of phage phi92: a comprehensive genomic and structural analysis. J. Virol. *86*, 10384–10398.

Serwer, P., Hayes, S.J., Zaman, S., Lieman, K., Rolando, M., and Hardies, S.C. (2004). Improved isolation of undersampled bacteriophages:

finding of distant terminase genes. Virology *329*, 412–424.

Serwer, P., Hayes, S.J., Thomas, J.A., and Hardies, S.C. (2007). Propagating the missing bacteriophages: a large bacteriophage in a new class. Virol. J. *4*, 21.

Sharon, I., Alperovitch, A., Rohwer, F., Haynes, M., Glaser, F., Atamna-Ismaeel, N., Pinter, R.Y., Partensky, F., Koonin, E.V., Wolf, Y.I., et al. (2009). Photosystem I gene cassettes are present in marine virus genomes. Nature *461*, 258–262.

Shkoporov, A.N., and Hill, C. (2019). Bacteriophages of the human gut: the "known unknown" of the microbiome. Cell Host Microbe *25*, 195–209.

Silva, J.B., Storms, Z., Sauvageau, D., Bertozzi Silva, J., Storms, Z., and Sauvageau, D. (2016). Host receptors for bacteriophage adsorption. FEMS Microbiol. Lett. *363*, 1–11.

De Sordi, L., Lourenço, M., and Debarbieux, L. (2019). The battle within: interactions of bacteriophages and bacteria in the gastrointestinal tract. Cell Host Microbe *25*, 210–218.

Stopar, D., Spruijt, R.B., Wolfs, C.J.A.M., and Hemminga, M.A. (2003). Protein-lipid interactions of bacteriophage M13 major coat protein. Biochim. Biophys. Acta *1611*, 5–15.

Sullivan, M.J., Petty, N.K., and Beatson, S.A. (2011). Easyfig: a genome comparison visualizer. Bioinformatics *27*, 1009–1010.

Tamaki, H., Zhang, R., Angly, F.E., Nakamura, S., Hong, P.-Y., Yasunaga, T., Kamagata, Y., and Liu, W.-T. (2012). Metagenomic analysis of DNA viruses in a wastewater treatment plant in tropical climate. Environ. Microbiol. *14*, 441–452.

Tétart, F., Repoila, F., Monod, C., and Krisch, H.M. (1996). Bacteriophage T4 host range is expanded by duplications of a small domain of the tail fiber adhesin. J. Mol. Biol. *258*, 726–731.

Villarroel, J., Kleinheinz, K.A., Jurtz, V.I., Zschach, H., Lund, O., Nielsen, M., and Larsen, M.V. (2016). HostPhinder: a phage host prediction tool. Viruses *8*, 1–22.

Walker, P.J., Siddell, S.G., Lefkowitz, E.J., Mushegian, A.R., Dempsey, D.M., Dutilh, B.E., Harrach, B., Harrison, R.L., Hendrickson, R.C., Junglen, S., et al. (2019). Changes to virus taxonomy and the international code of virus classification and nomenclature ratified by the international committee on taxonomy of viruses (2019). Arch. Virol. *164*, 2417–2429.

Wang, J., Hofnung, M., and Charbit, A. (2000). The C-terminal portion of the tail fiber protein of bacteriophage lambda is responsible for binding to LamB, its receptor at the surface of Escherichia coli K-12. J. Bacteriol. *182*, 508–512.

Wang, J.-B., Lin, N.-T., Tseng, Y.-H., and Weng, S.-F. (2016). Genomic characterization of the novel Aeromonas hydrophila phage Ahp1 suggests the derivation of a new subgroup from phiKMV-like family. PLoS One *11*, e0162060.

Wattam, A.R., Abraham, D., Dalay, O., Disz, T.L., Driscoll, T., Gabbard, J.L., Gillespie, J.J., Gough, R., Hix, D., Kenyon, R., et al. (2014). PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res. *42*, 581–591.

Wu, Q., and Liu, W.-T. (2009). Determination of virus abundance, diversity and distribution in a municipal wastewater treatment plant. Water Res. *43*, 1101–1109.

Wu, L., Ning, D., Zhang, B., Li, Y., Zhang, P., Shan, X., Zhang, Q., Brown, M., Li, Z., Van Nostrand, J.D., et al. (2019). Global diversity and biogeography of bacterial communities in wastewater treatment plants. Nat. Microbiol. *4*, 1183–1195.

Yabuuchi, E., Yano, I., Hotta, H., Nishiuchi, Y., Kosako, Y., Yano, I., Hotta, H., and Nishiuchi, Y. (1995). Transfer of two burkholderia and an alcaligenes species to Ralstonia. Gen. Nov. Microbiol. Immunol. *39*, 897–904.

Yilmaz, S., Allgaier, M., and Hugenholtz, P. (2010). Multiple displacement amplification compromises quantitative analysis of metagenomes. Nat. Methods *7*, 943–944.

Yuan, Y., and Gao, M. (2017). Jumbo bacteriophages: an overview. Front. Microbiol. *8*, 1–9.

Zdorovenko, E.L., Vinogradov, E., Wydra, K., Lindner, B., and Knirel, Y.A. (2008). Structure of the oligosaccharide chain of the SR-type lipopolysaccharide of *Ralstonia solanacearum* toudk-2. Biomacromolecules *9*, 2215–2220.

Zhang, M., Yang, L., Ren, J., Ahlgren, N.A., Fuhrman, J.A., and Sun, F. (2017). Prediction of virus-host infectious association by supervised learning methods. BMC Bioinformatics *18*, 60.

iScience, Volume *23*

Supplemental Information

Adsorption Sequencing as a Rapid Method to Link Environmental Bacteriophages to Hosts

Patrick A. de Jonge, F.A. Bastiaan von Meijenfeldt, Ana Rita Costa, Franklin L. Nobrega, Stan J.J. Brouns, and Bas E. Dutilh

Transparent methods

General data reporting

All metagenomics tools used default parameters, except where explicitly stated otherwise. All graphs were plotted using the ggplot (Gómez-Rubio, 2017) v3.2.1 package in R. All chemicals were obtained from Sigma-Aldrich, unless explicitly stated otherwise.

Data availability

All genomic data has been uploaded to the European Nucleotide Archive (ENA) under project PRJEB37817. Reads are available under ENA accession numbers ERS4427880-ERS4427890, while the cross-assembled contigs are available under ENA accession number ERZ1305919.

Bacterial cultivation and phage stock preparation

We tested the validity of AdsorpSeq with two model phages and their hosts. These were *Escherichia* phage λ (DSMZ #4499) infecting *Escherichia coli* K12 BW25113 (DSMZ #27469) and *Salmonella* phage P22 (DSM #18523), infecting *Salmonella enterica* subsp. *enterica* serovar Enteritidis S1400/94 (hereafter called *Salmonella enterica*) (Allen-Vercoe et al., 1997). For additional tests of phage λ specificity, we used *E. coli* JW3996 (ΔLamB) from the Keio strain collection (Baba et al., 2006), which lacks the λ protein receptor (Wang et al., 2000). For subsequent AdsorpSeq application on a hospital wastewater virome, we used nine bacterial strains. These were *Acinetobacter baumannii* (DSMZ #300007), *Klebsiella pneumoniae* (ATCC #11296), *Ralstonia pickettii* (DSMZ #6297), *Pseudomonas aeruginosa* PA01 (DSMZ #22644), *Fusobacterium necrophorum* D12, *Citrobacter freundii* 4_7_47CFAA, *Escherichia coli* 4_1_47FAA , *Bacteroides fragilis* 3_1_12, and *Bacteroides dorei* 5_1_36/D4. The latter five strains were provided by the reference catalogue of the Human Microbiome Project (HMP) from the University of Guelph in Guelph, Canada. All bacteria, except for the obligate anaerobic *Bacteroides* and *Fusobacterium* strains, were aerobically cultivated in lysogeny broth (LB) at 37^{o}C while under agitation. The obligate anaerobic strains were cultivated at 37^{o}C in anaerobic Columbia broth. This medium was made anaerobic by boiling and cooling under a stream of nitrogen gas, followed by dispensation in bottles which were closed with rubber stoppers and aluminium crimp caps. Bottles were subjected to three rounds of vacuum and nitrogen gas and autoclaved at 121^{o}C for 20 minutes.

Phage λ and P22 stocks were produced with the soft-agar overlay method as described (Kutter, 2009). To determine phage titres, an aliquot of 0.1 ml exponentially growing bacterial culture was added to 5 ml 0.7% (w/v) LB agarose. This mixture was layered on top of 1.5% (w/v) LB agar and allowed to dry. Phage stock was diluted in a series of 10-fold dilutions using SM buffer (100 mM NaCl, 8 mM $MgSO_4$ x 7 H_2O, 50 mM Tris-HCl pH 7.5) and 10 µl of each dilution was placed on the plates. After 16 h incubation at 37^{o}C, plaques were counted.

Bacterial cell envelope isolations

To isolate bacterial cell envelopes, bacteria were first grown overnight (aerobic) or for 3 days (anaerobic) as described under "bacterial cultivation and phage stock preparation". The resulting cultures were centrifuged at 10,000 x *g*, 4^{o}C for 15 minutes. Supernatant was discarded and cell pellets were washed with the original volume of lysis buffer (50 mM Tris HCl pH 7.5, 2 mM $MgCl_2$). This was centrifuged again, the supernatant was discarded, and the pellet was re-suspended in 4 volumes of lysis buffer per gram wet cell weight with the addition of 1 tablet of cOmplete EDTA-free protease inhibitor. Cells were lysed by thrice passing the suspension through a model CF1 Cell Disruptor (Constant Systems) at 1.5 kBar. After removal of cell debris by centrifugation at 12,000 x *g*, 4^{o}C for 15 minutes, cell envelopes were collected by centrifugation at 225,000 x *g* ,4^{o}C for 1 hour. The supernatant was discarded, cell envelope pellets were re-suspended in lysis buffer and centrifuged again. Soluble proteins were removed by resuspension of the cell envelope pellet in 200 mM NaCl + 20 mM Tris HCl, pH 7.5 and a third centrifugation. Supernatant was removed and cell envelope pellets were dissolved in 20 mM Tris HCl pH 7.5 + 200 mM NaCl at a concentration of 10 mg/ml and stored at -20^{o}C until further use.

Agarose assays with model phages

AdsorpSeq was first tested using phage λ and P22 and bacterial cell envelope suspensions from their hosts. Equal volumes of 10^{11} pfu/ml phage titre and 10 mg/ml cell envelope suspension were mixed. To allow phages to adsorb to bacterial cell envelope suspensions, the mixture was incubated at room temperature for 20 minutes. Bound and unbound phages were separated by applying the mixtures on a 1% (w/v) agarose gel and applying a current of 20 V/cm for 20 minutes using a Mupid One gel electrophoresis system (Eurogentec). The slots at the top of the gel, which contain bound phages, were cut out of the gel using a fresh scalpel knife and DNA was isolated using a Zymoclean gel DNA recovery kit (Zymo Research). The recovered DNA was quantified with a Qubit dsDNA HS assay kit and Qubit fluorometer (Thermo Fisher Scientific).

Hospital virome preparation

Material for the virome consisted of two litres of wastewater influent kindly provided by the Reinier de Graaf hospital wastewater treatment facility in Delft, the Netherlands. To remove debris and cellular material, the material was filtered in stages using coffee filters, 0.45 µm filters, and 0.2 µm filters. This may have selected against giant bacteriophages. Subsequently the virome was concentrated to 70 ml (i.e. approximately 30 times) using a Vivaflow tangential flow filter with a 100 kDa cut-off (Sartorius). The sample was stored at 4°C until further use.

To isolate phage DNA from the virome, viral capsids were first broken by incubating a 1 ml virome aliquot with 10 µg/ml proteinase K and 0.2 % SDS at 56°C for 1 hour. Afterward, DNA was purified by phase separations using consecutively phenol, phenol/chloroform, and chloroform. The aqueous phase was collected and 0.1 volume 3 M sodium acetate (pH 5.2) and 2.5 volumes ice-cold absolute ethanol were added. The mixture was incubated overnight at -20°C and DNA was pelleted by centrifugation at 21,000 x *g* for 15 minutes at 4°C. The pellet was washed with one volume of 70% ethanol, and re-pelleted by repeating the centrifugation. The DNA pellet was dissolved in TE buffer (1 mM EDTA, 10 mM Tris HCl pH 8) and the DNA concentration was ascertained using a Qubit dsDNA HS assay kit and Qubit fluorometer (Thermo Fisher Scientific).

Sample preparation and sequencing

Two sequencing experiments were performed. The first was a confirmation that AdsorpSeq can be used with a mixture of phage λ and P22, while the second applied AdsorpSeq to the hospital wastewater virome and nine bacterial strains (see “bacterial cultivation and phage stock preparation”). For the first sequencing experiment, phage λ and P22 were mixed at a titre of approximately $1 \cdot 10^8$ pfu/ml each. Incubation with cell envelopes of either *E. coli* BW25113 or *S. enterica* and DNA isolation were performed as described under *“*agarose assays with model phages”. For the second sequencing experiment, the same methodology was used on samples prepared from the hospital wastewater virome and cell envelope preparations. To increase DNA quantities, the samples were subjected to multiple displacement amplification (MDA) using an Illustra genomiphi v3 φ29 polymerase kit (GE Lifesciences) according to the instructions. Primers were removed from the amplified samples using AMpure XP beads (Beckman Coulter) and the final amplicons were eluted in 50 µl of TE buffer. DNA concentrations were determined using a Qubit dsDNA HS assay kit and Qubit fluorometer (Thermo Fisher Scientific). All samples with cell envelope suspensions added to them were prepared in biological quadruplicates. The whole procedure was repeated for the

samples including both cell envelopes and viromes, for a total of two technical duplicates that each consisted of biological quadruplicates. Virome-only controls only had biological duplicates. For both experiments, these virome-only controls consisted of DNA isolated from the phages both before and after MDA.

Library preparation and sequencing of all samples was performed at the Utrecht sequencing facility (USEQ) in Utrecht, the Netherlands. Libraries were prepared using the TruSeq DNA nano kit (Illumina), and samples were sequenced on a NextSeq500 run with 2x150 bp paired reads (Illumina).

Sequencing data analysis

To increase the read quality from our sequencing experiments, Illumina reads were quality trimmed, poly-G tails were removed, and remaining adapters were removed using fastp v0.20.0 (Chen et al., 2018) (options -g, -x). For the application of AdsorpSeq on the hospital waste water virome, trimmed reads of the unamplified virome and all bacterial samples were cross-assembled into 1,013,501 contigs using metaSPAdes v3.11.0 (Nurk et al., 2017). Reads from all samples were mapped to the genomes of the model phages and their hosts (model phage experiment) or the cross-assembled contigs (hospital waste water virome experiment) using the Burrows-Wheeler Aligner v0.7.12-r1039 (Li and Durbin, 2009). The number of reads that mapped to each genome or contig were determined using the idxstats tool in samtools v1.454 (Li et al., 2009).

For the experiment using hospital waste water, mean read depths of each contig in each sample were determined using the JGI summarize bam contig depths tool in metaBAT v2.12.1 (Kang et al., 2015). Additionally, all reads from the amplified virome were mapped against the contigs to determine which contigs were selected for by the amplification procedure. Contigs belonging to the host genomes were identified by a megaBLAST search against the host genomes using BLAST v2.6.0+ (Camacho et al., 2009). The 12,496 contigs with over 50% coverage and over 50% identity were removed as host contigs. Subsequently, contigs were taxonomically annotated using the contig annotation tool (CAT) v5.0.3 (Von Meijenfeldt et al., 2019), which uses homology searches of ORFs against the National Centre for Biotechnology Information (NCBI) non-redundant protein database (nr) (Agarwala et al., 2017) to predict contig taxonomy. CAT used prodigal v2.6.3 (Hyatt et al., 2010) to predict ORFs. Contigs with a superkingdom classification of “Viruses” and a score of 1 (which indicates that all predicted ORFs in a contig were classified in that superkingdom), those that had other superkingdom classifications with scores below 1, and those that could not be classified at all were selected for further analysis. This resulted in a dataset of 13,032 (putative) viral contigs. Putatively completed circular contigs were identified using a custom script that checked

whether the start and end of a contig contained identical sequences, following earlier studies (Jahn et al., 2019; Roux et al., 2017).

We binned the selected contigs according to their tetranucleotide usage patterns and their read depth patterns across the nine bacterial samples using metaBAT v2.12.1 (Kang et al., 2015). Because viral genomes are smaller than the bacterial genomes for which metaBat was originally built, we lowered the minimal contig length allowance to 2,500 bp and the minimum bin size to 10,000 bp. In addition, we decreased the minimum mean coverage necessary in each sample for binning to 0.001 to allow for contigs that might not be present in all samples (i.e. options -m 2500, -s 10000, and -x 0.001). Contigs that metaBAT could not bin, whether due to low coverage or short length, were discarded. We binned genomic fragments into viral populations (Gregory et al., 2019) based on similarity in tetranucleotide usage and abundance patterns by using metaBat resulting in 1158 viral populations containing a total of 6,572 contigs.

Selection of overrepresented viral populations

To select putative cell envelope-adsorbing viral populations, the abundance per viral population in each sample was calculated with the following formula:

$$\text{abundance}_{\text{viral population}} = \frac{\sum\left(\text{read depth}_{\text{viral population}} \cdot \text{contig length}_{\text{viral population}}\right)}{\sum \text{contig length}_{\text{viral population}}} \quad \text{(Equation 1)}$$

This abundance was then divided by the total abundance of all viral populations in the sample and expressed in a percentage to obtain the relative abundance for each viral population in each sample. For each viral population, the highest relative abundance across all cell envelope-treated samples was divided by the next highest relative abundance from a sample that used cell envelope suspensions from different taxonomical order. Positive outliers among the fractions between the top two samples of all viral populations were determined with the following formula:

$$\text{outlier} \geq 75\%_{\text{all viral populations}} + \text{interquartile range}_{\text{all viral populations}} \cdot 1.5 \quad \text{(Equation 2)}$$

The boundary for inclusion as putative adsorbing viral population established by this was 1.58 (Supplementary Figure 4A). A total of 123 putative adsorbing viral populations that constituted as outliers were selected.

MDA and methodological selection filters

To determine the extent that viral populations were selected for by MDA, we divided the relative abundance of each viral population in the post-MDA virome by that in the pre-MDA virome. To ascertain the relationship between phage taxonomy and MDA selection factor, we plotted the

values of all viral populations with a CAT classification at the family level that belonged to a phage lineage. From the resulting MDA selection factors, positive outliers were determined using equation 2. The 79 viral populations with MDA selection factors above 3.8, as defined by equation 2, were discounted (Supplementary Figure 4B).

In addition to the MDA selection factor, we also determined which viral populations were universally selected for by AdsorpSeq. For this, we calculated a methodological selection factor by dividing the relative abundance of each viral population per sample by the relative abundance in the amplified virome. As a result, each viral population had one methodological selection factor for each cell envelope-treated sample for a total of nine per viral population. After determining positive outliers with equation 2, the 18 viral populations which were above the outlier threshold of 3.43 in all nine samples were removed from the dataset (Supplementary Figure 4C). After accounting for these two biases, our dataset contained 26 putatively adsorbing viral populations which combined contained 83 contigs.

Analysis of putatively adsorbing viral populations

To uncover the extent of viral 'dark matter' among the putatively adsorbing viral populations, we plotted the ORF-specific taxonomical classifications obtained with CAT (Von Meijenfeldt et al., 2019) of all ORFs found in putative adsorbing viral populations. While CAT is built to classify contigs, it first predicts ORF-specific taxonomy and subsequently uses these results to predict contig taxonomy. Relatedness of the putative adsorbing viral populations and characterised phages was established by performing a BLASTp all vs all on a dataset composed of all ORFs from putative adsorbing viral populations and all ORFs from characterised phages in the NCBI bacterial and viral RefSeq V85 database (Pruitt et al., 2007). BLASTp searches were performed using diamond v0.9.29.130 (Buchfink et al., 2014), with a bit score significance cut-off of 50. Links between viral populations and characterised phages were displayed using Cytoscape v3.7.2 (Shannon et al., 2003). For further analysis of the links between putative adsorbing viral populations and known phages, we performed a taxonomical assignment on all selected contigs using vContact2 v.0.9.8 (Bin Jang et al., 2019), while the resulting network was visualised using Cytoscape v3.7.2 (Shannon et al., 2003).

Contigs from putative adsorbing viral populations were annotated using PROKKA v1.11 [80], with the metagenomic option enabled and with both the viral and bacterial settings (i.e. options--metagenome and both --kingdom Bacteria and --kingdom Viruses). In addition, distant homologs to ORFs were identified by performing searches against prokaryotic viral orthologous groups (pVOGs) (Grazziotin et al., 2017) using hmmsearch v3.1b2 (Eddy, 2011). To compare contig sequences to characterised phage genomes they were aligned using

Easyfig v2.2.3 (Sullivan et al., 2011), which used BLAST v2.9.0+ (Camacho et al., 2009) to perform tBLASTx searches.

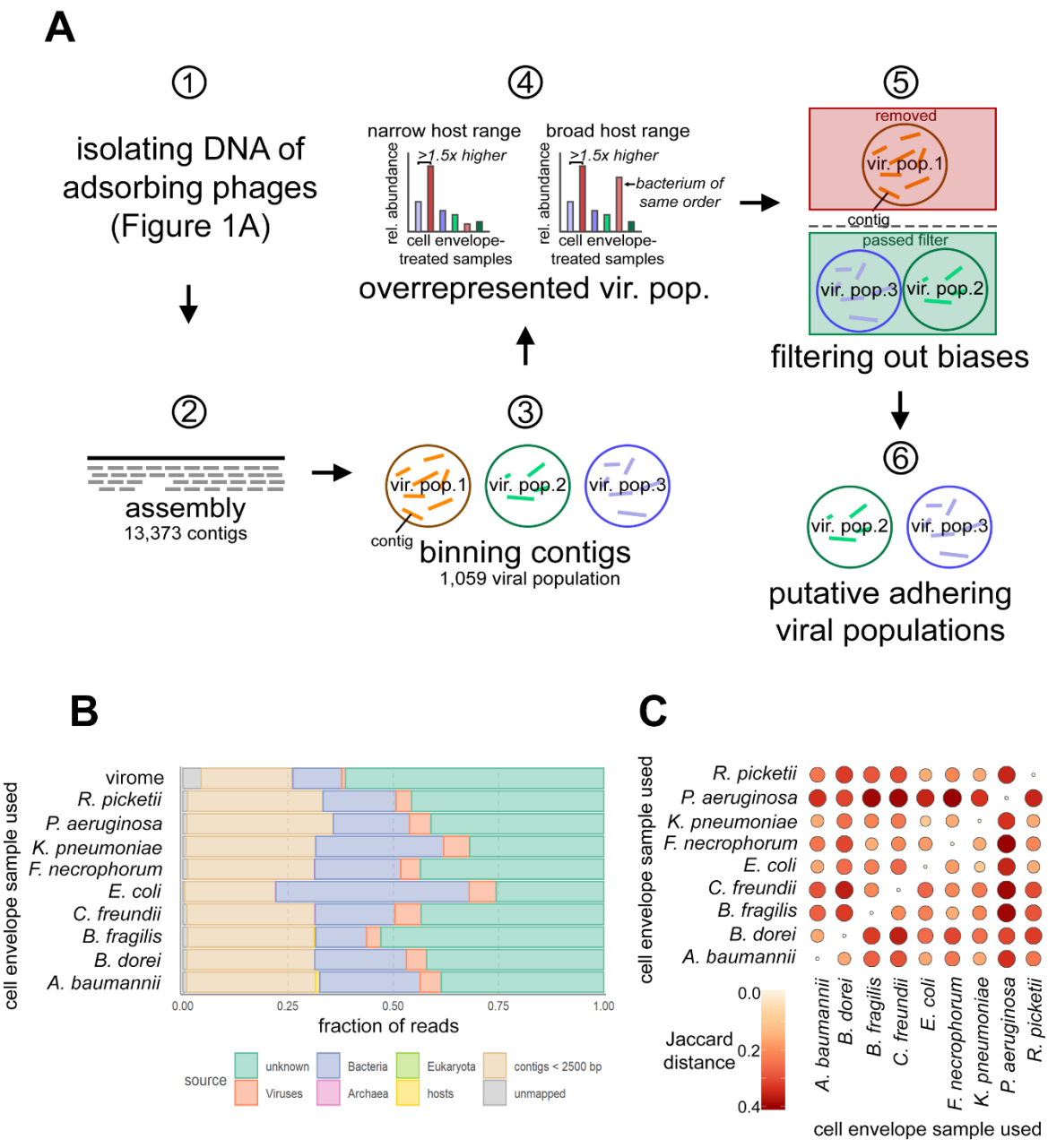
CRISPR-Cas spacer analysis

To search for CRISPR-Cas spacer hits against the putative adsorbing viral populations, spacers were identified in all bacteria Pathosystems Resource Integration Centre (PATRIC) database (Wattam et al., 2014) (accessed February 2019) using detect 2.2.1. This resulted in a dataset of 1,473,418 spacers. These spacers were used as query against contigs from putative adsorbing viral populations in a BLASTn with the option for short sequences enabled (option -task blastn-short) in BLAST v2.9.0+ (Camacho et al., 2009). Only spacer hits with fewer than five mismatches were taken into consideration. In practice, there were no spacer hits with between one and five mismatches.

Jumbo phage terminase phylogeny

Some of the putative adsorbing viral populations contained (fragments of) jumbo phage genomes. It was previously shown that in a terminase phylogenetic tree these phages largely clustered according to host (Yuan and Gao, 2017). Three terminases that we identified in the putative adsorbing viral populations were used as query for a BLASTp search against the NCBI nr-database using the NCBI webserver (Johnson et al., 2008) on January 14 2020. In addition, the three terminases were used as query for a BLASTp search against jumbo phage proteins from Al-Shayeb *et al* (Al-Shayeb et al., 2020). All 222 total hits above a bit score threshold of 50 were combined with the three queries. Duplicated sequences were removed and the remainder was aligned using Clustal Omega v1.2.1 (Sievers and Higgins, 2018). Aligned positions that consisted of more than 90% gaps were trimmed with trimAl v1.2 (Capella-Gutiérrez et al., 2009) (option -gt 0.1). A maximum likelihood tree was constructed using IQ-Tree v1.6.12 (Nguyen et al., 2015) using model finder (Kalyaanamoorthy et al., 2017) and performing 1000 iterations of both SH-like approximate likelihood ratio test and the ultrafast bootstrap approximation (UFBoot) (Hoang et al., 2018). In addition, ten iterations of the tree were separately constructed, as has been recommended (Zhou et al., 2018) (IQ-Tree options -bb 1000, -alrt 1000, and --runs 10). The iteration with the best log-likelihood score, for which model finder selected LG+F+R6, was used for analysis. The tree was visualised using interactive Tree of Life v5.5 (Letunic and Bork, 2019). In addition, we selected the putative FtsZ homolog from viral population 18 (ORF 17) and used it to query the NCBI nr-database using the NCBI webserver (Johnson et al., 2008) on January 14 2020. This did not result in significant hits.

Supplementary figures

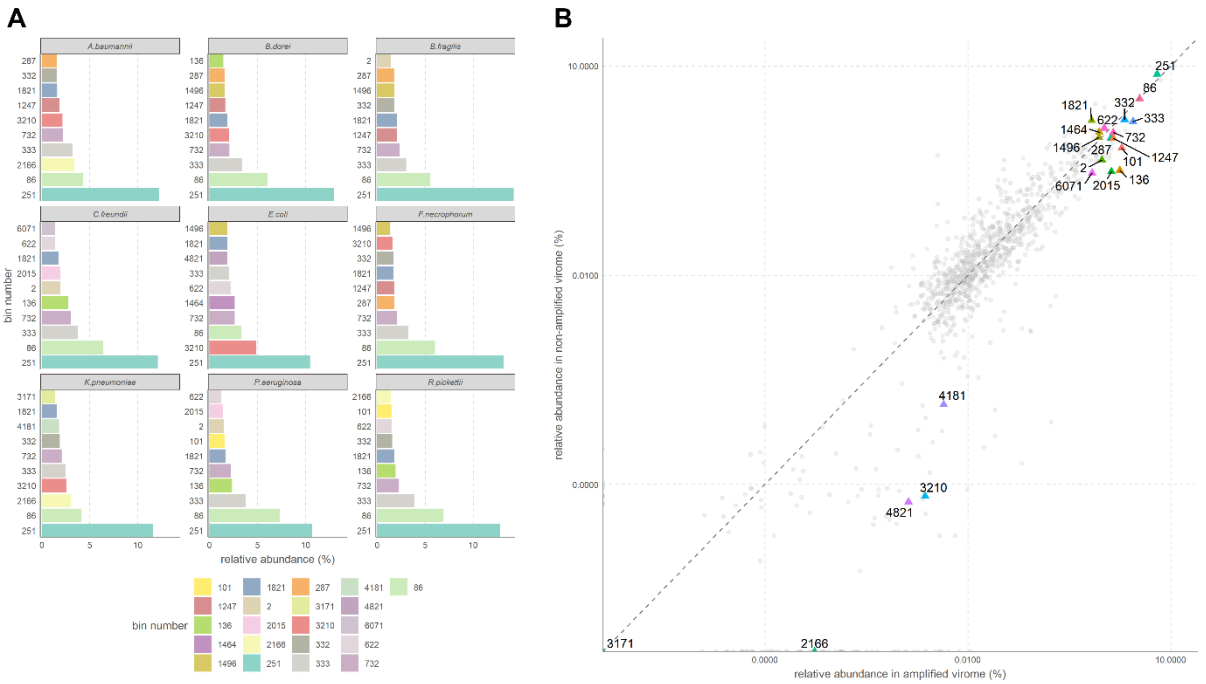


Supplementary Figure 1: Application of AdsorpSeq to a hospital wastewater virome, related to Figure 1 and 2.

(A) Schematic of the approach taken to apply AdsorpSeq to an environmental sample. Step (1) consists of the process depicted in Figure 1A. Subsequently, data analysis consisted of (2) assembly of read-level data into contigs, (3) binning contigs into viral populations based on their relative abundance across the samples, (4) selection of viral populations that were overrepresented in one sample, (5) remove viral populations under multiple displacement amplification (MDA) or methodological bias, and (6) the final selection of viral populations with putative adhering capabilities.

(B) For contigs over 2500 bp, a majority of reads across the all but one of the samples map to unknown or viral contigs. Stacked bar charts of the fragment of reads from each sample that are unmapped, map to short contigs below 2,500 bp, map to one of the nine bacterial hosts, map to contigs that are classifiable at the superkingdom level, and map to unknown/unclassifiable contigs.

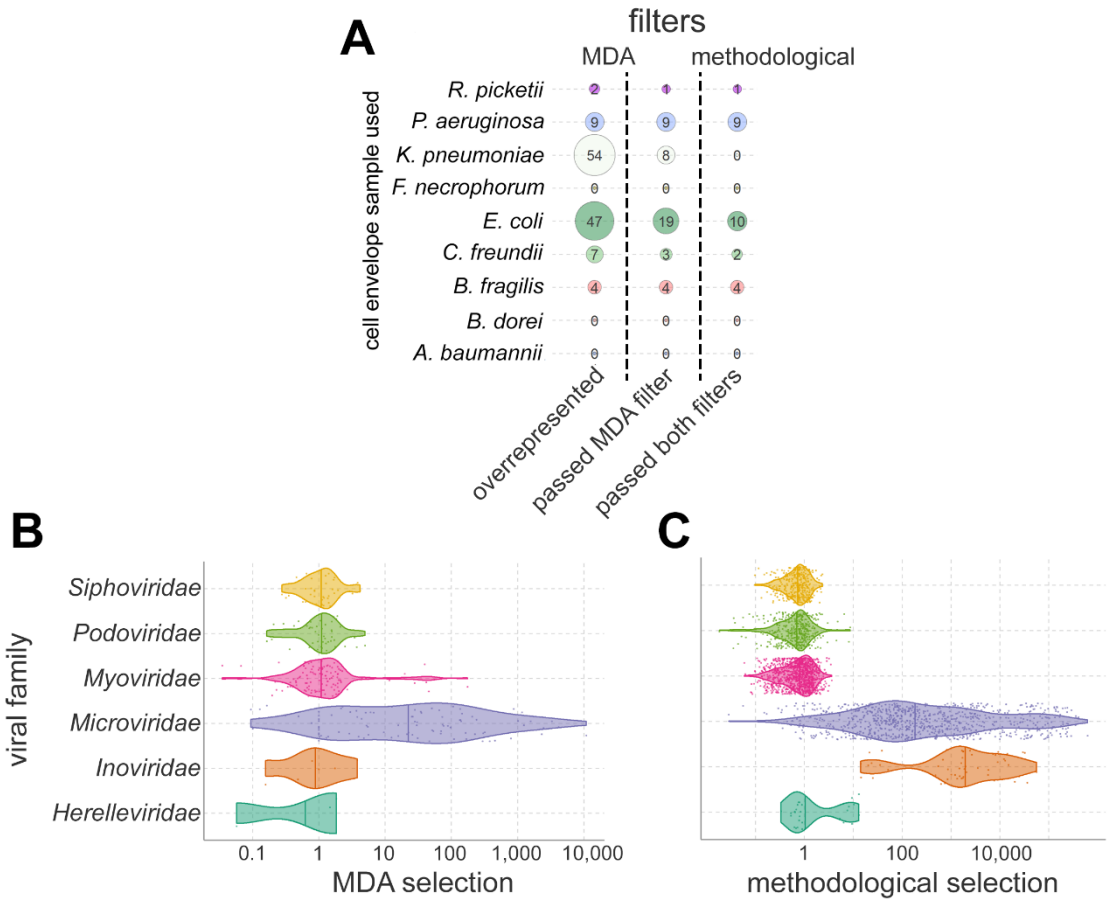
(C) Selection of phages by AdsorpSeq is dependent on the species of bacterial cell envelope that is used. A correlogram of the Jaccard distance between the membrane-treated samples, based on 1,058 viral populations. Higher Jaccard distance means greater dissimilarity. Size and colour indicated Jaccard dissimilarity.



Supplementary Figure 2: The most abundant genome viral populations across the cell envelope-treated samples are similar, related to Figure 2.

(A) Bar charts showing the ten most abundant genome viral populations in each of the nine cell envelope-treated samples and their relative abundances.

(B) The relative abundances of the genome viral populations depicted in (A) in the virome before (y-axis) and after (x-axis) multiple displacement amplification. Numerical labels refer to the genome viral populations.

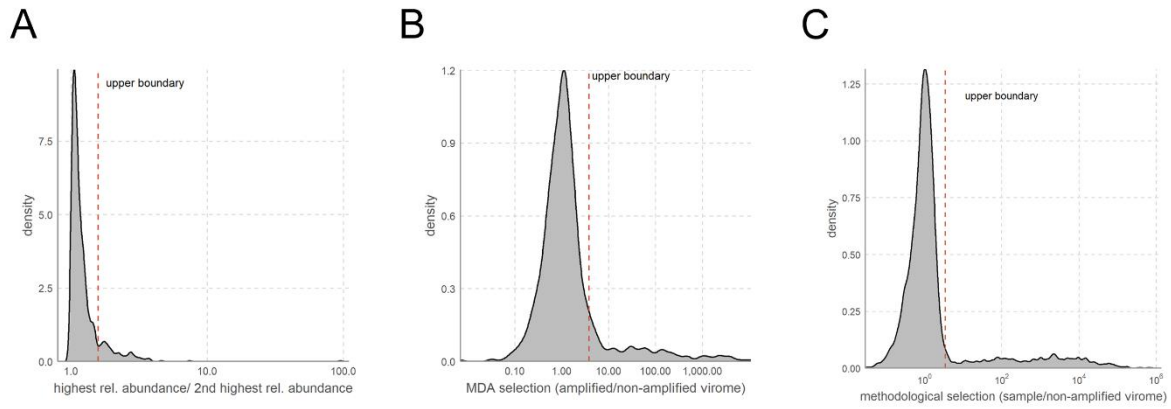


Supplementary Figure 3: Selection of putative adhering viral populations from a hospital wastewater virome, related to Figure 2 and 3.

(A) After filtering for MDA and methodological selection, 26 viral populations with putative adsorbing capacity were selected from the virome. Bubble area shows number of viral populations after sequentially selecting based on being overrepresented in one sample, filtering for MDA selection, and filtering for methodological selection. The third column shows the selected viral populations with putative adhesion activity.

(B) *Microviridae* are highly selected for by MDA, as shown by viral populations with positive CAT predictions within a bacteriophage family. MDA selection is defined as the ratio in relative abundance between the virome before and after MDA. Points in the charts represent individual viral populations.

(C) *Inoviridae* and *Microviridae* undergo high methodological selection, which is defined by the ratio between the relative abundance of a viral population in each sample and in the amplified virome. As there were nine cell envelope-treated samples, each viral population is represented in the plot by nine data points.



Supplementary Figure 4: Distributions used in selection of viral populations that represent adhering viral populations, related to Figure 2.

(A) Distribution of the ratios between the cell envelope-treated sample with the highest and second highest relative abundance among the viral populations. All viral populations with a ratio above the red dashed line labelled “upper boundary” were selected as being overrepresented in a sample.

(B) Distribution of the ratios in relative abundance between the virome before and after multiple displacement amplification (MDA). All viral populations with a ratio above the red dashed line labelled “upper boundary” were under strong MDA selection.

(C) Distribution of the ratios between the relative abundances in the samples and in the non-amplified virome for each viral population. Each viral population is represented with nine datapoint in this plot; one for each cell envelope-treated sample. All viral populations with a ratio above the red dashed line labelled “upper boundary” were under strong methodological selection.

Supplementary references

Agarwala, R., Barrett, T., Beck, J., Benson, D.A., Bollin, C., Bolton, E., Bourexis, D., Brister, J.R., Bryant, S.H., Canese, K., et al., (2017). Database Resources of the National Center for Biotechnology Information. Nucleic Acids Res. 45, D12–D17.

Al-Shayeb, B., Sachdeva, R., Chen, L.X., Ward, F., Munk, P., Devoto, A., Castelle, C.J., Olm, M.R., Bouma-Gregson, K., Amano, Y., et al., (2020). Clades of huge phages from across Earth’s ecosystems. Nature 578, 425–431.

Allen-Vercoe, E., Dibb-Fuller, M., Thorns, C.J., Woodward, M.J., (1997). SEF17 fimbriae are essential for the convoluted colonial morphology of Salmonella enteritidis. FEMS Microbiol. Lett. 153, 33–42.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., Mori, H., (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. 2, 2006.0008.

Bin Jang, H., Bolduc, B., Zablocki, O., Kuhn, J.H., Roux, S., Adriaenssens, E.M., Brister, J.R., Kropinski, A.M., Krupovic, M., Lavigne, R., et al., (2019). Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. Nat. Biotechnol.

37, 632–639.
Buchfink, B., Xie, C., Huson, D.H., (2014). Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60.
Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., (2009). BLAST+: Architecture and applications. BMC Bioinformatics 10, 1–9.
Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973.
Chen, S., Zhou, Y., Chen, Y., Gu, J., (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890.
Eddy, S.R., (2011). Accelerated profile HMM searches. PLoS Comput. Biol. 7.
Gómez-Rubio, V., (2017). ggplot2 - Elegant Graphics for Data Analysis (2nd Edition) . J. Stat. Softw. 77, 3–5.
Grazziotin, A.L., Koonin, E. V., Kristensen, D.M., (2017). Prokaryotic Virus Orthologous Groups (pVOGs): A resource for comparative genomics and protein family annotation. Nucleic Acids Res. 45, D491–D498.
Gregory, A.C., Zayed, A.A., Conceição-Neto, N., Temperton, B., Bolduc, B., Alberti, A., Ardyna, M., Arkhipova, K., Carmichael, M., Cruaud, C., et al., (2019). Marine DNA Viral Macro- and Microdiversity from Pole to Pole. Cell 177, 1109-1123.e14.
Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecular biology and evolution. Mol. Biol. Evol. 35, 518–522.
Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11.
Jahn, M.T., Arkhipova, K., Markert, S.M., Stigloher, C., Lachnit, T., Pita, L., Kupczok, A., Ribes, M., Stengel, S.T., Rosenstiel, P., et al., (2019). A Phage Protein Aids Bacterial Symbionts in Eukaryote Immune Evasion. Cell Host Microbe 26, 542-550.e5.
Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., Madden, T.L., (2008). NCBI BLAST: a better web interface. Nucleic Acids Res. 36, 5–9.
Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589.
Kang, D.D., Froula, J., Egan, R., Wang, Z., (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 2015, 1–15.
Kutter, E., (2009). Phage host range and efficiency of plating. Methods Mol. Biol., Methods in Molecular Biology 501, 141–9.
Letunic, I., Bork, P., (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new

developments. Nucleic Acids Res. 47, W256–W259.
Li, H., Durbin, R., (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.
Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274.
Nurk, S., Meleshko, D., Korobeynikov, A., Pevzner, P.A., (2017). MetaSPAdes: A new versatile metagenomic assembler. Genome Res. 27, 824–834.
Pruitt, K.D., Tatusova, T., Maglott, D.R., (2007). NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 35, 61–65.
Roux, S., Emerson, J.B., Eloe-Fadrosh, E.A., Sullivan, M.B., (2017). Benchmarking viromics: An in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. PeerJ 2017, 1–26.
Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–504.
Sievers, F., Higgins, D.G., (2018). Clustal Omega for making accurate alignments of many protein sequences. Protein Sci. 27, 135–145.
Sullivan, M.J., Petty, N.K., Beatson, S.A., (2011). Easyfig: A genome comparison visualizer. Bioinformatics 27, 1009–1010.
Von Meijenfeldt, F.A.B., Arkhipova, K., Cambuy, D.D., Coutinho, F.H., Dutilh, B.E., (2019). Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. Genome Biol. 20, 1–14.
Wang, J., Hofnung, M., Charbit, A., (2000). The C-terminal portion of the tail fiber protein of bacteriophage lambda is responsible for binding to LamB, its receptor at the surface of Escherichia coli K-12. J. Bacteriol. 182, 508–512.
Wattam, A.R., Abraham, D., Dalay, O., Disz, T.L., Driscoll, T., Gabbard, J.L., Gillespie, J.J., Gough, R., Hix, D., Kenyon, R., et al., (2014). PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res. 42, 581–591.
Yuan, Y., Gao, M., (2017). Jumbo bacteriophages: An overview. Front. Microbiol. 8, 1–9.
Zhou, X., Shen, X.X., Hittinger, C.T., Rokas, A., (2018). Evaluating fast maximum likelihood-based phylogenetic programs using empirical phylogenomic data sets. Mol. Biol. Evol. 35, 486–503.