Cell Host & Microbe
Review

Bacteriophages of the Human Gut: The "Known Unknown" of the Microbiome

Andrey N. Shkoporov[1,*] and Colin Hill[1]
[1]APC Microbiome Ireland & School of Microbiology, University College Cork, Co. Cork, Ireland
*Correspondence: andrey.shkoporov@ucc.ie
https://doi.org/10.1016/j.chom.2019.01.017

The human gut microbiome is a dense and taxonomically diverse consortium of microorganisms. While the bacterial components of the microbiome have received considerable attention, comparatively little is known about the composition and physiological significance of human gut-associated bacteriophage populations (phageome). By extrapolating our knowledge of phage-host interactions from other environments, one could expect that >10^{12} viruses reside in the human gut, and we can predict that they play important roles in regulating the complex microbial networks operating in this habitat. Before delving into their function, we need to first overcome the challenges associated with studying and characterizing the phageome. In this Review, we summarize the available methods and main findings regarding taxonomic composition, community structure, and population dynamics in the human gut phageome. We also discuss the main challenges in the field and identify promising avenues for future research.

Introduction

The human body has been referred to as a "superorganism" (Aziz et al., 2013) in which microbial cells are present in numbers (~10^{14}) comparable to human cells (Sender et al., 2016). An overwhelming majority (>99%) of these microbes are located in the distal segments of the gastrointestinal tract (GIT). They occupy different ecological niches in the gut lumen and on mucosal surfaces, forming complex biochemical interaction networks between themselves and with the host organism. The dynamic equilibrium of the gut microbiome is essential for normal host physiology. For instance, gut microbes participate in host metabolic processes (Everard and Cani, 2013), stimulate normal development of immunity and brain functions in early ontogenesis (Dinan and Cryan, 2017), provide a barrier against incoming pathogens, and balance local immune responses throughout life (Belkaid and Hand, 2014). This has led to an appreciation of the human gut microbiome as a "forgotten organ," an essential, albeit genetically and antigenically foreign, component of the human body (O'Hara and Shanahan, 2006). The gut microbiome contains all three domains of cellular life, *Bacteria*, *Archaea*, and *Eukarya*, as well as viruses, albeit at very different relative concentrations (Figure 1). *Bacteria* and *Archaea* account for more than 99% of the unique characterized gene repertoire and biomass (Qin et al., 2010; Yatsunenko et al., 2012; Sender et al., 2016; Wampach et al., 2017) and have received most of the attention in human microbiome studies. At the same time, recent works have also highlighted the role of fungi and protozoa, microbial eukaryotes that constitute a smaller but potentially important part of the gut microbiome (Hoffmann et al., 2013; Huseyin et al., 2017; Laforest-Lapointe and Arrieta, 2018).

It is often postulated that viruses of bacteria are the most numerous biological entities on the planet and in many environments outnumber the counts of their prokaryotic hosts by a factor of 10 (Wommack and Colwell, 2000). The original hypothesis of linear virus-to-microbe ratio (VMR), based on early data from marine and freshwater microbial communities (Weinbauer, 2004; Thingstad et al., 2008), has been revised recently, with power law and unimodal models seeming to more accurately reflect extensive variation in VMR (2.6–160 in the oceans) (Knowles et al., 2016; Wigington et al., 2016). It has long been known that abundant and diverse communities of non-pathogenic viruses, mainly tailed bacteriophages, colonize the mammalian gut (Dhillon et al., 1976). Up until the last decade, however, the phageome remained the "known unknown" of the gut microbiome. This was mainly due to a very limited toolkit, which included direct observation and counting of virus-like particles (VLPs) using transmission electron (TEM) and epi-fluorescence microscopy (EFM) techniques, as well as isolation of individual bacteriophages infecting specific host strains in culture. Microscopic methods helped to reveal a large diversity of viral morphotypes (up to several tens per individual) with total counts of bacterial viruses in human feces, caecal contents, and colonic mucosa reaching ~10^9–10^{10} VLPs g^{-1} (Hoyles et al., 2014; Lepage et al., 2008). These were largely members of the *Caudovirales* order, represented by the *Siphoviridae*, *Podoviridae*, and *Myoviridae* families. Culture-based methods were mainly used to isolate bacteriophages against a limited set of model and clinically important microorganisms such as *Escherichia*/*Shigella* (Dhillon et al., 1976; Martinez-Castillo et al., 2013), *Enterococcus faecalis* (Bonilla et al., 2010), *Clostridioides difficile* (Hargreaves and Clokie, 2014), and a few other bacteria. Because >95% of bacteria residing in the distal gut, including non-pathogenic strict anaerobes belonging to families *Bacteroidaceae*, *Prevotellaceae*, *Ruminococcaceae*, *Lachnospiraceae*, etc., are difficult to culture, the available collections of phage strains of human fecal origin clearly still do not reflect the true diversity of human gut bacteriophages.

The advent of high-throughput metagenomic sequencing technology has allowed us to appreciate the complexity and richness of human gut bacteriophage populations (Breitbart et al., 2003, 2008). The first metagenomic studies of fecal viromes revealed that most bacterial viruses in the gut (81%–93%) are novel and can be neither assigned a taxonomic

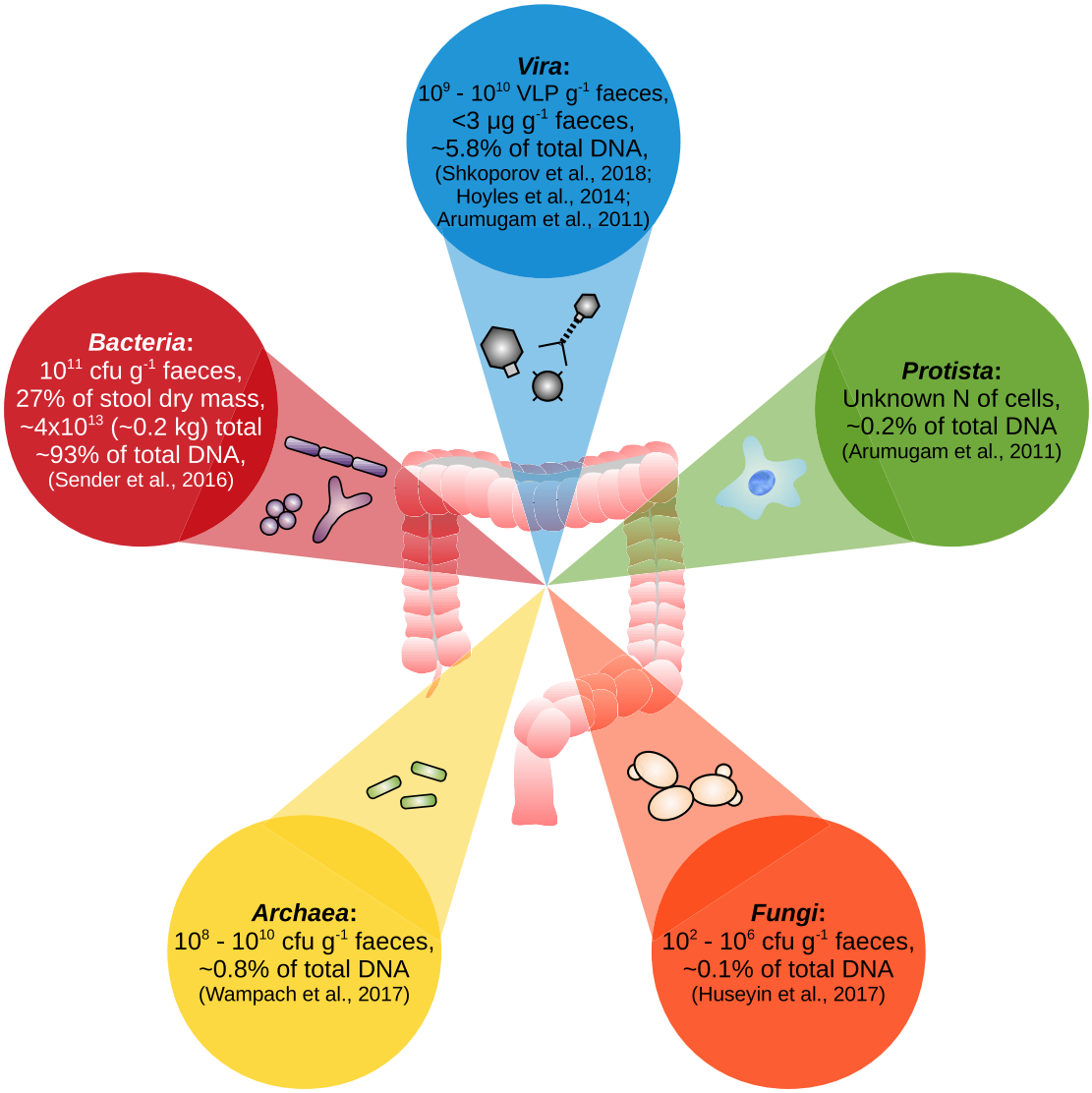

Figure 1. Main Taxonomic Groups of the Human Gut Microbiome and the Domain/Kingdom Level

position nor linked to a bacterial host (Manrique et al., 2016; Reyes et al., 2010). This is further complicated by the fact that human gut phageomes are highly individual specific, with only a small overlap between subjects (Manrique et al., 2016). The term "viral dark matter" has been coined to describe the existing gap in knowledge about the taxonomic composition and population structure of the gut phageome (Aggarwala et al., 2017). Description of the viral landscape in the gut would be incomplete without mentioning minority populations of circular, replication initiator protein (Rep) encoding, single-stranded DNA (CRESS-DNA) eukaryotic viruses (Lim et al., 2015; Reyes et al., 2015), and even pathogenic plant RNA viruses, which are likely of dietary origin but retain infectivity during transit through the gut (Zhang et al., 2006).

Widespread bacteriophage predation and lysogenic conversion in bacterial populations plays a major role in regulating bacterial biomass, maintaining biodiversity, horizontal gene transfer and driving biogeochemical cycles in the Earth biosphere (Thingstad et al., 2008). With phage-bacterial ratios of ~1:1 in the human gut (Carding et al., 2017), we can expect that bacteriophage predation, lysogeny, and gene transfer will play major roles in controlling the density, diversity, and network interactions inside gut-associated symbiotic bacterial communities as well. Importantly, specific and lasting changes of phageome composition were detected in a number of diverse gut-related and systemic conditions such as inflammatory bowel disease (IBD), malnutrition, and AIDS (Norman et al., 2015; Monaco et al., 2016; Reyes et al., 2015). Additionally, evidence of the efficacy of sterile fecal filtrate transfer in the treatment of *C. difficile* infection (CDI) points toward the potential ability of gut phages to restrict pathobiont growth and promote normal richness of the gut microbiota (Ott et al., 2017). Interestingly, however, the majority of gut bacteriophages seem to engage in lysogenic interactions with their hosts, thereby persisting for prolonged

periods of time and with much slower evolution rates than the minority virulent bacteriophages (Minot et al., 2013). Lack of evidence of "kill-the-winner" dynamics and Red Queen co-evolution (Reyes et al., 2010) suggests that ecological strategies and modes of interaction between gut bacteriophages and their hosts are fundamentally different from those observed in other well-studied ecological systems, such as oceans, where lytic lifestyles prevail and play a central role in shaping and controlling bacterial populations (Silveira and Rohwer, 2016).

In this Review, we will focus on the taxonomic composition, dynamics, and spatial structure of bacteriophage populations, as well as certain aspects of their interaction with their hosts in the healthy human gut. We will also review the available methodology, specifically discussing the challenges associated with metagenomic analysis pipelines. We will not be discussing phageome interactions with the host immune system, its role in various disease states, as well as phageome-based therapeutic approaches to treat gut disorders, as these topics are covered by other reviews in this Phage Focus issue of *Cell Host & Microbe*.

Methodology and Main Challenges Associated with Studying the Human Phageome

An integrated gut virome analysis pipeline, which amalgamates different methods and approaches reported in the literature is presented in Figure 2. A crude VLP-containing fraction of feces (fecal fitrate; FF) is typically prepared by vigorous homogenization and subsequent centrifugation and microfiltration of the supernatant to remove bacterial cells and dietary debris. Absolute quantification of VLPs in FF and mucosal samples can be done by direct counting of particles stained with DNA/RNA intercalating dyes (SYBR green II, SYBR gold, DAPI) under EFM or using flow cytometry (Lepage et al., 2008; Brown et al., 2015). Flow cytometry has an added advantage in that specific fractions of particles selected on the basis of size, granularity, and fluorescence intensity can be collected and further analyzed (Džunková et al., 2015). To obtain more concentrated samples of VLPs, ultracentrifugation at ~120,000 g, precipitation with polyethylene glycol and NaCl or $ZnCl_2$, or tangential flow filtration can be employed (Castro-Mejía et al., 2015). A concentrated FF sample can then be enzymatically treated to remove free, capsid-unprotected DNA/RNA, yielding a viral fraction suitable for metagenomic sequencing (Shkoporov et al., 2018b). Alternatively, even purer VLP samples can be obtained by collecting fraction(s) of specific buoyant density after ultracentrifugation in CsCl step or continuous gradients (Kleiner et al., 2015). VLP fractions can be examined using TEM (Castro-Mejía et al., 2015), metagenomic DNA and cDNA sequencing, or metaproteomics.

Currently, deep sequencing using high-throughput short-read-based technologies (Roche 454, Illumina MiSeq/NextSeq/HiSeq/NovaSeq platforms, Ion Torrent platforms) remains the primary approach to characterizing unculturable viral communities in the gut. However, assembly, mapping, and classification of short reads arising from mostly novel and unknown viral genomes (viral dark matter) represent a considerable bioinformatic challenge (Roux et al., 2015a; Aggarwala et al., 2017). In recent years, two long-read sequencing technologies became available (Pacific Biosciences and Oxford Nanopore), which, despite considerably lower per base accuracy rates, could be an interesting complement to short-read sequencing. Specifically, they can be used to assist in scaffolding of large novel viral genomes, obtaining information on methylation patterns (potentially useful for host prediction; Beaulaurier et al., 2018), and for studying population structure at a single-virion level, since long reads can, in some cases, represent complete or near-complete viral genomes (Warwick-Dugdale et al., 2018). Additional future complementary approaches may include gut viral metatranscriptomics (RNA-seq) and viral metaproteomics.

Each of the available purification and analysis methods has its limitations and introduces a bias. For example, the use of glass beads for sample homogenization, high centrifugation speeds, and small pore filters leads to a dramatic reduction in large virions. Chloroform extraction of PEG-precipitated VLP samples removes enveloped viruses. Contrary to that, use of large pore filters and omitting the chloroform step improves recovery of some viruses but introduces considerable bacterial contamination (Conceição-Neto et al., 2015). Similarly, CsCl density gradient purification yields very pure samples, ideal for TEM and metaproteomic studies, but fails to recover enveloped viruses and those with atypical buoyant densities (Castro-Mejía et al., 2015; Kleiner et al., 2015). We will focus on some of the most significant biases and unsolved problems associated with metagenomic analysis of viral populations in the human gut.

Total Viral Counts in the Gut

A number of studies that employed direct counting of VLPs stained with DNA/RNA intercalating dyes, suggested viral counts from the human gut significantly lower than the expected level of 10^{12} VLPs g^{-1} (if the postulated approximate 10:1 phage:bacterium ratio was maintained). Hoyles et al. reported an average 3×10^9 VLPs obtained by filtration per gram of feces in healthy adult subjects (Hoyles et al., 2014), while a study of colonic mucosa biopsy samples revealed the presence of 1.2×10^8 VLPs per biopsy in healthy individuals and significantly higher viral loads of 2.9×10^9 VLPs per biopsy in IBD patients (Lepage et al., 2008). Recently, we employed viral metagenomics to quantify fecal bacteriophages by comparing total numbers of DNA reads in VLP-enriched fractions to numbers of reads aligned to a standardized number of exogenous phage deliberately spiked into the fecal samples (lactococcal phage Q33) (Shkoporov et al., 2018b). In a small sample size, we estimated viral loads to be 1.46×10^9–1.81×10^{10} VLPs g^{-1}. Taking 1×10^{10} VLPs g^{-1} as a rough estimate, one could conclude that the true VMR in the human gut is reversed relative to other environments and could be as low as 0.1.

In contrast, shotgun metagenomic studies of total community DNA in human feces imply a much higher proportion of bacteriophage sequences in the gut, from an average of 5.8% (Arumugam et al., 2011) to extremes of up to 22% (Dutilh et al., 2014). Assuming an average size of 50 kb per bacteriophage genome and an average size of bacterial genome of 4 Mb, this would imply a VMR of ~4.64. With shotgun metagenomic data, it is impossible to discern DNA packaged in phage particles from prophage sequences in bacterial genomes. At the same time, it seems likely that both EFM/flow cytometry counts and viral metagenomics methods tend to underestimate viral loads in the extremely dense microbial and viral communities of the gut due to a number of factors, including inefficient elution of VLPs from feces (Conceição-Neto et al., 2015). It is also possible

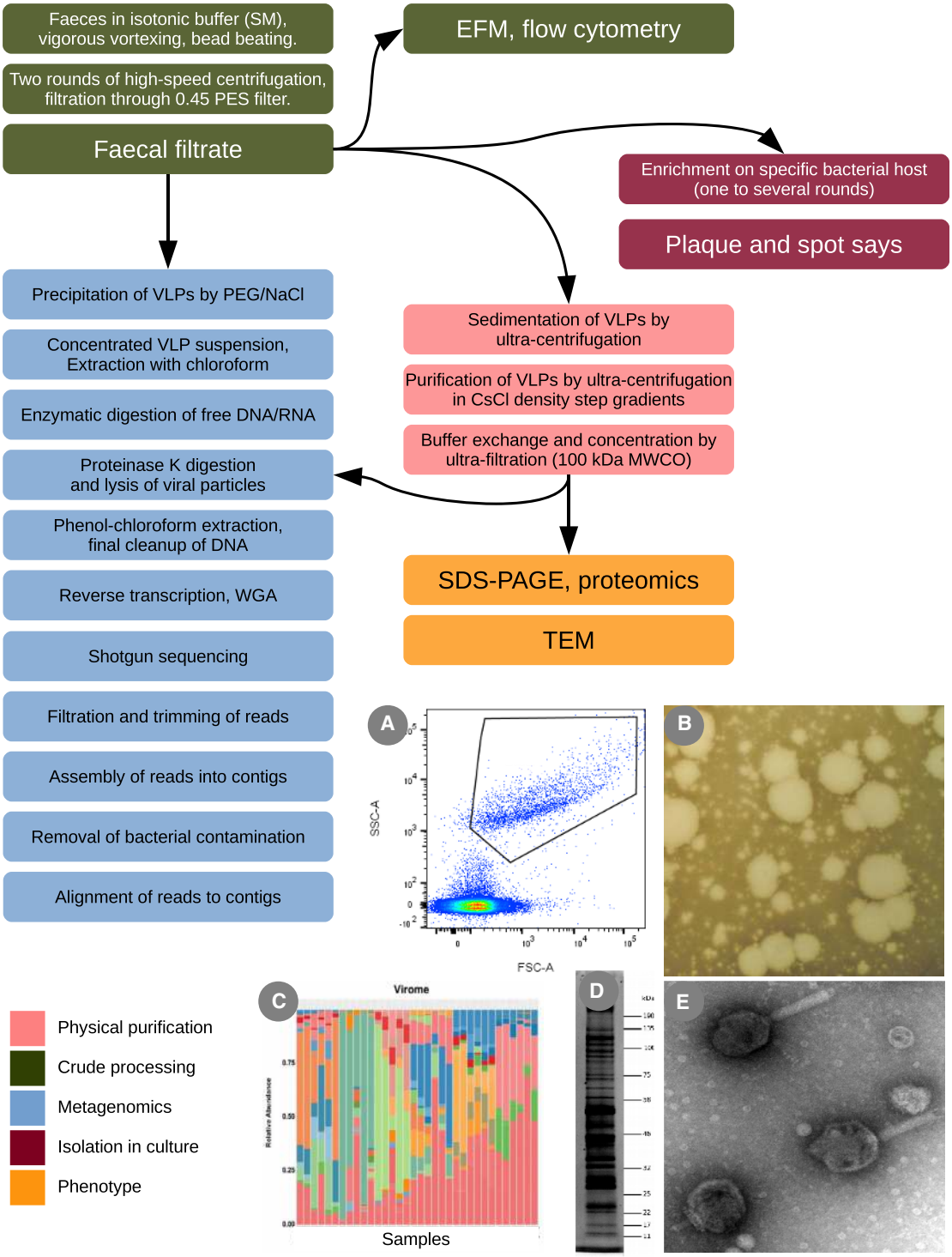

Figure 2. A Generic Gut Phageomics Workflow Combining Metagenomic, Culture-Based, Microscopic, and Proteomic Approaches
(A) Flow cytometric enumeration of phage particles, obtained from human fecal filtrate.
(B) Agar overlay showing plaques of different morphologies on *E. coli* from enriched cattle manure sample.
(C) Relative abundance of metagenomic contigs obtained by shotgun DNA sequencing of human fecal VLP fractions, (reproduced from Shkoporov et al., 2018b).
(D) SDS-PAGE gel electrophoresis displaying predominant polypeptide fractions in CsCl-gradient-purified human fecal VLP fraction (reproduced from Guerin et al., 2018).
(E) TEM of phage particles in CsCl-gradient-purified human fecal VLP fraction (A.N.S. and C.H., unpublished data).

that binding of phages to bacterial cells and cell debris and aggregation of VLPs could lead to underestimated EFM counts.

DNA Amplification Bias

If we were to assume that the gut phageome is composed entirely of coliphage T4 (genome of ~170 kb; an example of a relatively large *Myoviridae* phage) with a titer of 1×10^{10} cfu g^{-1}, the total mass of virion heads (MW = 1.94×10^8 Da) would be 3.23 μg per g of feces (Yap and Rossmann, 2014). Consequently, the total mass of its 1.04×10^8 Da genomic double-stranded DNA (dsDNA) would be 1.7 μg per g of feces. In a similar way, should the entire fecal phage population consist of coliphage X174 (genome of ~5 kb; an example of small *Microviridae* phage), the total genomic single-stranded DNA (ssDNA) content would be equal to 30 ng per g of faeces. Since the human gut phage community represents a complex mixture of species with different genome sizes, with the types of nucleic acids varying as well, the actual DNA yields can be as high as 250–500 ng viral DNA per gram of feces (Shkoporov et al., 2018b), albeit with some samples yielding as little as 4–5 ng. Thus, whole-genome amplification (WGA) techniques (typically multiple displacement amplification [MDA] with phage φ29 DNA polymerase) will be required to obtain sufficient DNA for downstream processing (Reyes et al., 2010; Minot et al., 2013). If a reverse transcription step is included, MDA gives the added advantage of converting single-stranded cDNA into a double-stranded form, compatible with common sequencing library preparation techniques (Shkoporov et al., 2018b).

The use of WGA results in a significant distortion in viral taxonomic composition, especially if the starting DNA concentration was extremely low. For instance, use of the popular φ29 DNA polymerase-based kits results in a significant expansion of small circular ssDNA genome viruses (phage families *Microviridae*, *Inoviridae*, and eukaryotic CRESS-DNA viruses such as *Circoviridae* and *Anelloviridae*) and plasmids (Norman et al., 2015; Roux et al., 2016). In addition, amplification leads to a general reduction of diversity and obscures detection of some rare viral groups (Kim and Bae, 2011). Therefore, the use of a new generation of library prep kits compatible with Illumina sequencing platforms, suitable for extremely low dsDNA and ssDNA inputs, and minimizing amplification steps should be adopted as standard practice in gut virome/phageome research (Roux et al., 2016). Single-virion genomics (SGV) will probably be more widely used in future studies (Martinez-Hernandez et al., 2017; Warwick-Dugdale et al., 2018).

Viral Dark Matter and Insufficiency of Viral Sequence Databases

Perhaps the most critical shortcoming of the metagenomic approach to studying the human gut phageome is the severe discrepancy between the demonstrable diversity of gut viruses and the number of genomes of known gut-associated bacteriophages and eukaryotic viruses in public databases. Viral metagenomics of the human gut yields between 75% and 99% of reads that do not produce significant alignments to any known viral genome (Aggarwala et al., 2017). This is in stark contrast with the current status of human gut bacteriome research, where the Human Microbiome Project (Qin et al., 2010) and other efforts have allowed for the isolation and complete genome sequencing of >1,000 predominant gut bacterial species (Rajilić-Stojanović and de Vos, 2014), accounting for >90% of total gut microbial diversity at the species level (Browne et al., 2016). Despite this obvious shortfall, many studies of human gut virome in health and disease published so far have relied on alignment of individual reads or assembled contigs to these nascent viral sequence databases and hence were able to interpret only a small minority of the sequencing data (Minot et al., 2013; Norman et al., 2015; Lim et al., 2015). In a study with just 13 human donors, we were able to assemble 8,920 putative non-redundant complete and partial viral genomes, of which only 161 (1.8%) could be assigned to known viral taxa (with >50% identity over 90% of contig length). Of these, 157 were bacteriophages, two were human papillomaviruses, and one was a plant RNA virus (Shkoporov et al., 2018b).

Given that the majority of reads cannot be aligned to a closed-reference database, the alternative is an open-reference approach where viral reads are assembled into contigs, which are then classified and annotated through reference-based and *de novo* annotation steps. Reads are then aligned back to assembled contigs, and alignment counts matrices are built. This gives an opportunity to quantify viral species and perform α- and β-diversity analyses independent of the taxonomic position of any of these contigs (Manrique et al., 2016; Reyes et al., 2015; Shkoporov et al., 2018b). Quality assembly of short reads is a significant hurdle in viral metagenomics. The right choice of assembler software and algorithm parameters is critical and can lead to dramatic differences in the results (Roux et al., 2017; Sutton et al., 2019). Bacteriophage populations of the human gut represent a particularly challenging target for *de novo* assembly due to (1) high levels of diversity (Manrique et al., 2016); (2) modular structure of bacteriophage genomes and high levels of genetic mosaicism (Lima-Mendez et al., 2011); (3) population microdiversity and high heterogeneity at strain level (Minot et al., 2013; Martinez-Hernandez et al., 2017); (4) high incidence of repeat and hypervariable regions (Minot et al., 2012); and (5) wide variation of relative abundance and hence sequence coverage (Sutton et al., 2019). This leads to a high degree of assembly fragmentation and hampers annotation and interpretation of alignment results. The most radical solution to the "metagenomic assemblies" conundrum could be physical separation of individual viral particles and sequencing of individual genomes—i.e., single virion genomics or use of long reads spanning nearly complete viral genomes (Martinez-Hernandez et al., 2017; Warwick-Dugdale et al., 2018).

Combating Bacterial DNA Contamination in Viral Metagenomics

Methods based on density gradient ultracentrifugation are able to deliver phage nucleic acid samples that are virtually free from contaminating bacterial DNA (Reyes et al., 2010). These methods are impractical for routine use because of high manual workloads and hence operator-to-operator variability, low throughput, and a tendency to introduce bias by excluding viruses with atypical buoyant densities (Castro-Mejía et al., 2015; Conceição-Neto et al., 2015). Therefore, most metagenomic studies of the human gut virome use more practical methods based on filtration, with subsequent precipitation or ultrafiltration of VLPs (Lim et al., 2015; Minot et al., 2013; Norman et al., 2015). These protocols can introduce considerable amounts of residual bacterial DNA into the sample (Shkoporov et al., 2018b). Combined with the incompleteness of

bacteriophage genomic databases and the need to *de novo* identify viral genomes against a bacterial background, this can lead to frequent misinterpretations in gut virome studies. For instance, a study of the murine gut virome exposed to antibiotic treatment claimed an expansion of the resistome and other functions potentially beneficial to bacterial hosts in the phage metagenome (Modi et al., 2013). These claims, however, were refuted by a subsequent independent re-examination, which demonstrated that the majority of these genes were likely to be associated with bacterial DNA contamination (Enault et al., 2017). A number of metrics are suggested to measure the amount of contaminating bacterial DNA in virome samples, including percentage of reads aligned to bacterial 16S rRNA and *cpn60* gene databases (Roux et al., 2013; Shkoporov et al., 2018b).

De novo identification of novel viral genomes in the metagenomic datasets against a background of bacterial and eukaryotic DNA contamination presents an extremely challenging task. Viral lineages are polyphyletic by origin and fast evolution of many viruses leads to further sequence diversification, sometimes to an extreme degree where no DNA sequence similarity can be captured even between members of the same viral family (Guerin et al., 2018). Unlike in cellular life forms, lack of any conserved phylogenetic marker genes prevents easy identification and taxonomic assignment of novel uncultured viruses. We utilize a rigorous multi-stage approach to filtering out bacterial DNA contamination from VLP-enriched metagenomic sequencing samples. Putative viral contigs are identified via several selection steps, including positive results with the VirSorter classifier (Roux et al., 2015b), alignments to viral genomes in NCBI RefSeq database and our in-house database of crAss-like phages (Guerin et al., 2018), as well as the presence of multiple genes with above-threshold similarities to conserved bacteriophage proteins in pVOGs database (Grazziotin et al., 2017) and/or circular contig topology (Figure 2).

A number software tools, databases, and websites have been specifically designed for processing high-throughput virome sequencing data. A concise selection of such software, along with some general-purpose tools useful for the steps from read filtration, trimming, and assembly to gene prediction, host prediction, taxonomic classification, and multivariate analysis of community composition, are listed in Table 1.

Diversity and Individuality of Human Gut Phageomes

Microscopic studies have shown that gut bacteriophages are almost exclusively represented by tailed viruses with icosahedral capsids, belonging to the order *Caudovirales*. In the majority of cases, they can be robustly classified into families based on tail morphology: *Siphoviridae* with long flexible non-contractile tails, *Myoviridae* with long stiff contractible tails, and *Podoviridae* with very short tails (Figure 3). Unique assemblages of up to several tens of distinct morphotypes of these three families can be recognized in feces collected from different human donors (Hoyles et al., 2014; Castro-Mejía et al., 2015). Some studies report detecting bacteriophage families other than those of the order *Caudovirales*, e.g., *Cystoviridae*, *Inoviridae*, and *Microviridae* (Guerin et al., 2018; Lepage et al., 2008). It should be noted, however, that recognizing small non-tailed icosahedral or filamentous capsids in fecal samples against a background of dietary debris can be a challenging task (Figure 3).

These findings were further corroborated by metagenomic studies of the gut phageome involving sequencing of both viral genomic DNA and RNA. While >80% of viral sequences did not match against closed-reference databases, most of the classifiable metagenomic reads aligned to *Siphoviridae* phage genomes (Breitbart et al., 2003; Reyes et al., 2010). In more recent studies focusing on DNA viruses, only 7%–13% of recovered viral contigs could be assigned to known viral families—mostly of the order *Caudovirales* (dsDNA genomes) and family *Microviridae* (ssDNA genomes)—based on the presence of family-specific hallmark genes (Minot et al., 2013; Manrique et al., 2016; McCann et al., 2018). The abundance of *Microviridae* was, however, likely to be overestimated through the use of MDA amplification as described earlier. In a recent study, we found that the majority of viral contigs identifiable to family level belonged to the *Caudovirales* (Shkoporov et al., 2018b). However, large numbers of *Microviridae* contigs were also detectable. The presence of lysogeny genes in the majority of the complete contigs of *Caudovirales* suggests temperate lifestyles. We were unable to detect RNA bacteriophages (e.g., family *Leviviridae*) in gut phage communities, likely because of low viral loads (*Leviviridae* virions are very small and can be resistant to precipitation) and nucleic acid extraction procedures (Shkoporov et al., 2018b).

In addition to bacteriophages, cryptic human and microeukaryotic CRESS-DNA viruses are consistently detected (*Anelloviridae*, *Circoviridae*, and *Genomoviridae*), as well as human *Herpesviridae* and *Papillomaviridae* (Lim et al., 2015; McCann et al., 2018), but these fall outside of the scope of this Review. Interestingly, the only group of RNA viruses consistently found in the healthy gut are plant viruses belonging to family *Virgaviridae* (Shkoporov et al., 2018b). These viruses of dietary origin are able to maintain infectivity upon passage through the human GIT (Zhang et al., 2006). The presence of giant amoebal viruses (Aherfi et al., 2016) in the gut has never been reported but cannot be completely ruled out because methods used by most studies are incapable of recovering them. The same would be true of some "jumbo" bacteriophages (with genome size of >500 kbp; Devoto et al., 2018) and viruses with unusual morphologies and physical properties, such as *Autolykiviridae*, a family of highly prevalent ocean phages, which avoided detection until very recently (Kauffman et al., 2018). Archaeal viruses include some morphotypes shared with bacteriophages, as well as the unusual ones specific to the archaeal domain (Prangishvili et al., 2017; Krupovic et al., 2018). None have been detected in human gut samples to date.

As discussed above, the scarcity of bacteriophage genomes in the reference databases makes it impossible to directly stratify metagenomic sequences of the human gut phageome by their position in viral taxonomy systems, functional properties or their specificity toward bacterial hosts. Using network-based *de novo* clustering approaches (e.g., vConTACT framework) could be an alternative. Such methods, based on analysis of the content of conserved protein-coding genes, attempt to represent evolutionary and functional relationships between both cultured and uncultured viral genomes in a reticulate fashion and categorize them into clusters, agnostic from established taxonomy (Lima-Mendez et al., 2008; Bolduc et al., 2017). However, further efforts will be required to reconcile these viral clusters with established

Table 1. Selected Software Tools and Databases Useful in Phage and Viral Metagenomic Analysis

Tool Name	Description	Reference
Software		
Trimmomatic	Trimming and filtering of short DNA sequencing reads (command line)	https://github.com/timflutre/trimmomatic
Bowtie 2	Fast alignment of DNA reads to reference contigs (command line)	https://github.com/BenLangmead/bowtie2
SPAdes, metaSPAdes	Assembly algorithms for high-throughput sequencing data (short and long reads likewise, command line)	Nurk et al., 2017
Sunbeam	Modular pipeline for processing of microbial metagenomic data (command line)	Clarke et al., 2018
VirSorter	Detection of viral sequences in metagenomic data (web based and command line)	Roux et al., 2015b
vConTACT	*De novo* classification of dsDNA viral genomes (web based and command line)	Bolduc et al., 2017
PHASTER	Detection of prophages in draft or complete bacterial genomes and metagenomes (web based)	Arndt et al., 2016
Prokka	Rapid annotation of prokaryotic genomes (command line)	Seemann, 2014
RASTtk	Modular annotation pipeline suitable for viral genomes and metagenomic contigs (web based)	Brettin et al., 2015
Metavir 2	Annotation and comparison of assembled viromes (web based)	Roux et al., 2014
PhageTerm	Determination of phage genome termini and packaging mechanisms (command line or Galaxy based)	Garneau et al., 2017
iVIREONS	Neural network-based identification of bacteriophage structural protein genes from genome sequences (web based)	Seguritan et al., 2012
mauve	Multiple genome alignment (GUI and command line)	Darling et al., 2004
PyANI	Pairwise average nucleotide identity between genomes (command line)	https://github.com/widdowquinn/pyani
Easyfig	Linear BLAST comparison of multiple genomic loci (GUI)	Sullivan et al., 2011
GView	Visualization of circular and linear genome maps (GUI and command line)	https://www.gview.ca/
vegan	Multivariate analysis of microbial and viral communities (R package)	https://cran.r-project.org/web/packages/vegan/index.html
ape	Phylogenetic and evolutionary analysis of DNA and protein sequence data (R package)	Paradis and Schliep, 2018
Databases		
NCBI RefSeq Viral and Viral Genome Browser	Comprehensive database of cultured and/or characterized viral genomes with taxonomic annotation.	https://www.ncbi.nlm.nih.gov/genome/viruses/
IMG/VR	Integrated database of viral metagenomic sequences from various habitats complete with predicted host information and tentative taxonomic classification.	Paez-Espino et al., 2016, 2017
pVOGs	A COG framework-based database of orthologous protein families encoded by prokaryotic virus genomes.	Grazziotin et al., 2017

taxonomic systems or to predict biological properties and host ranges of newly discovered unknown bacteriophages.

Caudovirales

Members of this viral group have linear dsDNA genomes ranging in length from ~16 kb (streptococcal podovirus C1) to hundreds of kilobases in large *Myoviridae*. Temperate phages of this order engage in lysogenic interaction with their hosts by integrating their genome into the host chromosome (e.g., coliphages λ and μ) or persisting through generations as an autonomously replicating episome (e.g., coliphage P1). The frequent occurrence of prophages in gut commensal bacteria makes it possible to identify hosts for some of the temperate phage genomes detected in metagenomic surveys (Minot et al., 2013; Norman et al., 2015). The host range of the order *Caudovirales* is very broad and includes all major bacterial phyla found in the gut: *Firmicutes*, *Bacteroidetes*, *Proteobacteria*, and *Actinobacteria*. The typical gut *Caudovirales* member *Siphoviridae* (identified *de novo* or using reference databases) have linear genomes of moderate size (~35–50 kb), often containing a lysogeny module with a gene for a serine or tyrosine integrase (Minot et al., 2013; Shkoporov

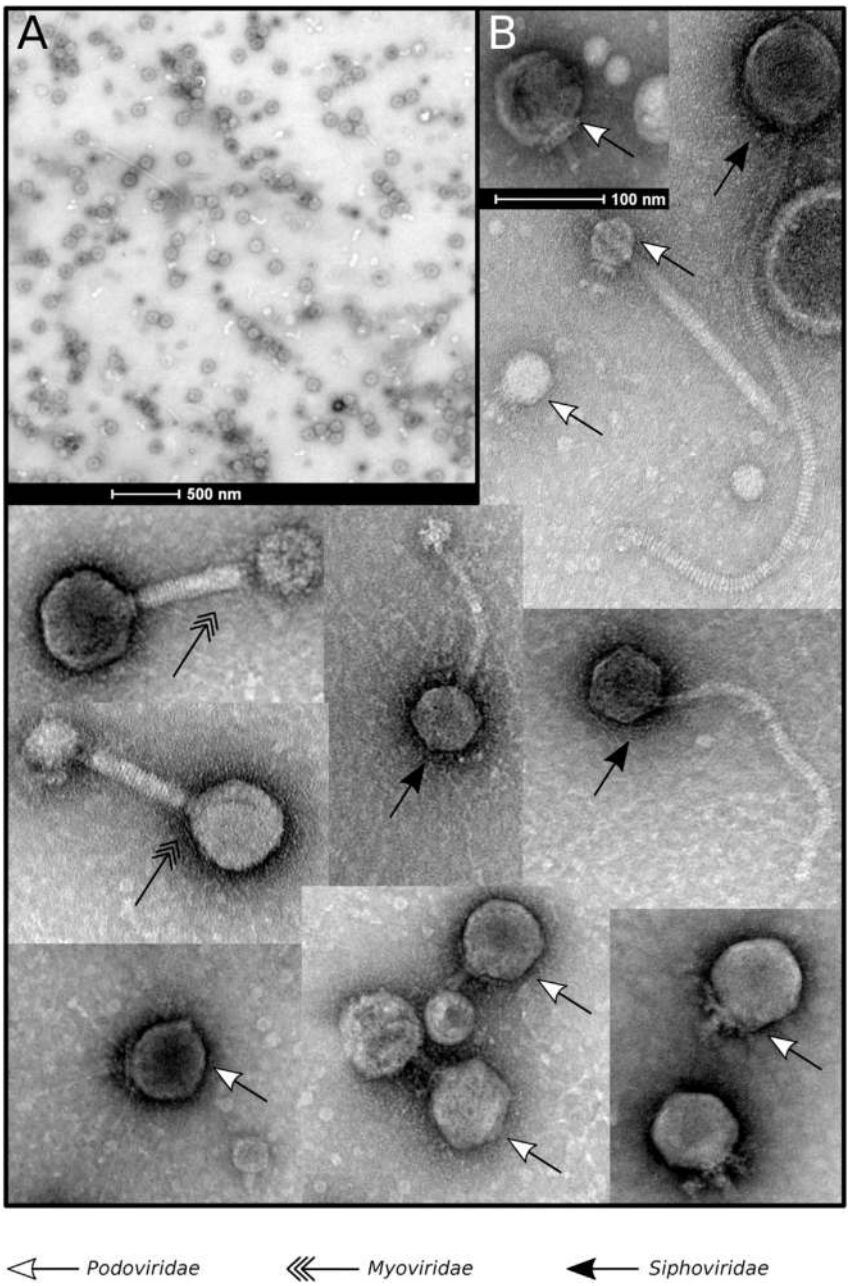

Figure 3. Main Bacteriophage Morphological Types Detectable Using TEM in a Human Fecal Sample

(A) Overview of bacteriophage content in CsCl-gradient-purified human feecal VLP fraction at low magnification (20,500×).

(B) Examples of morphotypes detected in the same sample at high magnification (160,000×).

et al., 2018b). Some representatives of this group of viruses infecting common members of the human gut bacteriome, such as *Bacteroides* and *Clostridium*, have been isolated in culture (Puig and Gironés, 1999; Ogilvie et al., 2012; Hargreaves and Clokie, 2014). However, the presence of large *Siphoviridae*-type virions in TEM images (sometimes with tails >1 μm long) suggests the existence of potentially virulent viruses with larger genomes, since virion volume can serve as an (imperfect) predictor of genome size (Cui et al., 2014). Furthermore, virulent-tailed phages tend to have larger genomes in order to supply the necessary complement of nucleic acid metabolism genes for efficient lytic cycles. Data on *Myoviridae* and *Podoviridae* phages remain scarce, but results suggest a wide variation of virion size, genome length, and potentially host ranges (Hoyles et al., 2014). Recently, an uncultured megaphage with a genome length of >540 kbp, predicted to infect *Prevotella* and showing properties of the *Myoviridae* family, was detected in the microbiota of humans from Bangladesh and Tanzania (Devoto et al., 2018). *Faecalibacterium* and *Bifidobacterium* are among the most predominant bacteria in healthy human microbiota in adulthood and infancy, respectively (Yatsunenko et al., 2012), but efforts to isolate bacteriophages infecting these predominant bacterial genera have proven unsuccessful, despite the presence of numerous prophages in their genomes. Recently, induction of prophages and secretion of *Siphoviridae* and *Myoviridae*-type viral particles from *Faecalibacterium* and *Bifidobacterium* have been reported (Cornuault et al., 2018; Mavrich et al., 2018).

crAss-like Phages

In 2014, a previously unrecognized 97 kb dsDNA phage genome was described in human gut metagenomic datasets and termed crAssphage (crAss, cross-assembly) (Dutilh et al., 2014). Its high relative abundance (up to 90% of total viral load in the gut of individuals) and wide representation in human population (>50% of Western population colonized) attracted a lot of attention, because of both its apparent significance for the human health and its potential use as fecal pollution marker (Liang et al., 2016; Cinek et al., 2017; García-Aljaro et al., 2017). Despite no significant homology to any known bacteriophages, it was predicted to infect *Bacteroides* based on co-abundance and CRISPR spacer hits (Dutilh et al., 2014; Cinek et al., 2017). Analysis of the genome assigned functions to ~50% of genes, predicted a *Podoviridae*-like morphology, and identified a whole family of similar crAss-like bacteriophages present in diverse environments such as human and insect guts, oceans, terrestrial and groundwater samples (Yutin et al., 2018). We recently described an expansive collection of uncultured, human gut-associated crAss-like bacteriophage genomes, which could be classified into ten candidate genera and four subfamily-level taxa. Taken together, representatives of this novel proposed viral family are present in 77% of individuals in diverse human populations

with relative abundances of up to 95% of the total viral load in the gut (Guerin et al., 2018). In a separate study, we reported the propagation of the first member (φcrAss001) of the family in a pure culture of *Bacteroides intestinalis* (Shkoporov et al., 2018a). Preliminary results suggest that members of this family maintain stable colonization of the human gut and can engage in unusual carrier state-type of interaction with their bacterial hosts both *in vitro* and *in* vivo (Guerin et al., 2018; Shkoporov et al., 2018a). Reyes et al. reported stable engraftment and long-term persistence of two human fecal viruses (φHSC04 and φHSC05), which we later identified as crAss-like phages belonging to candidate genera I and VII, in mice colonized by a defined microbial community consisting of 15 strains of anaerobic bacteria, including 8 different species of *Bacteroides* (Reyes et al., 2013). Further, a recent report highlighted engraftment and persistence for up to 1 year of allogeneic crAssphage in humans during fecal microbial transplantation (FMT) (Draper et al., 2018). These interesting phenomena provide additional clues regarding the host range of this unusual viral family and also support *in vitro* observations of their unusual ability to persist long term in the presence of a sensitive host.

Microviridae

These phages are an ever-present component of the human gut microbiome, even though the relative abundance of this family in the phageome remains controversial (Norman et al., 2015; Manrique et al., 2016). These viruses possess small circular ssDNA genomes (4–7 kb) packaged into icosahedral capsids. A large diversity of these viruses was detected in both marine and animal-associated microbiomes (Roux et al., 2012). However, only a small sample of *Microviridae* phages, mainly enterobacteria phages belonging to the *Bullavirinae* subfamily, and *Chlamydia* and *Bdellovibrio* phages in the *Gokushovirinae* subfamily, has been isolated in culture. The latter subfamily, along with subfamily *Alpavirinae*, is especially predominant in the human gut (Stockdale et al., 2018; Shkoporov et al., 2018b). As the only *Microviridae* group capable of a temperate lifestyle, prophages of family *Alpavirinae* are frequently detected in the genomes of *Bacteroides* and *Prevotella*, integrated through an unconfirmed mechanism that possibly involves cellular chromosome dimer resolution machinery (Krupovic and Forterre, 2011). The host range of gut *Gokushovirinae* cannot be directly established. However, acquisition of possible *Gokushovirinae* peptidase genes by some strains of predominant gut *Firmicutes*, in particular *Faecalibacterium prausnitzii*, suggests that these anaerobes may serve as the host to some strains of *Gokushovirinae* (Roux et al., 2012).

Individual Specificity and Age-Dependent Variation of the Gut Phageome

A number of studies reported very high levels of individual specificity of the gut phageome, with inter-individual differences being the primary source of variance at the population level (Minot et al., 2011; Norman et al., 2015; Shkoporov et al., 2018b). Despite that, the identification of a common set of bacteriophages found in 20%–50% of individuals formed the basis for the healthy core gut phageome concept (Manrique et al., 2016). A significant fraction of bacteriophages was found to be shared between twins and their mothers, as well as between IBD-affected and healthy members of the same household (Norman et al., 2015; Reyes et al., 2010, 2015). A common presumption is that newborns are born sterile; therefore, bacterial viruses would not be expected to be present in their gut. Rapid colonization of the newborn gut in the first days of life by a dynamic assembly of bacteriophages was reported (Breitbart et al., 2008). The neonatal gut phageome is complex and relatively unstable, preying on a low abundance of microbial hosts (Lim et al., 2015; Manrique et al., 2017). However, progressive maturation of the infant gut microbiome leads to a reduction of viral abundance and diversity, accompanied by an increase in abundance and diversity of the bacterial component. Interestingly, abundance and diversity of *Caudovirales* and *Microviridae* show opposite trends in the early postnatal ontogenesis with a gradual decrease of the former and increase of the latter up to 2.5 years of life. Birth mode has a profound effect on phageome composition, still detectable at 1 year of age (McCann et al., 2018). Little is known about phageome progression later in life. However, unusually high abundances of *Gokushovirinae* were detected in one cohort of elderly subjects, possibly reflecting a shift toward *Firmicutes* in their bacteriomes (Stockdale et al., 2018). Interesting but understudied aspects of individual specificity in the human gut phageome are geographic and ethnic differences. We have observed stark contrasts in crAss-like phage composition in Western and African populations, with the former being predominantly colonized by candidate genus I (likely host *Bacteroides*), while the latter is colonized by candidate genera VIII and IX (likely host *Prevotella*) (Guerin et al., 2018).

Spatial Structure and Dynamics of Bacteriophage Populations in the Gut

The ocean microbiome provides a classical ecological model to study general principles of population dynamics and ecological-evolutionary relationships between bacteriophages and their bacterial prey (Thingstad et al., 2008). A number of models have been developed explaining certain aspects of phage population dynamics and phage-host co-evolution. Among others, there is the "arms race" model (continuous Red Queen-like directional selection of mutations leading to broadly resistant hosts and highly infective parasites), the fluctuating selection model (density dependent fluctuating selection based on a trade-off between benefits of resistance and its metabolic costs), and "kill-the-winner" (extension of fluctuating selection model, taking into account abiotic factors), each with their own assumptions and limitations (Avrani et al., 2012).

The mammalian gut environment presents a much more complicated system than the ocean, with a number of biotic and abiotic factors at play, such as the complex anatomy of the gut at both macroscopic (longitudinal segmentation, valves, haustra, peristalsis and mass movement, secretion of bile, and pancreatic juice) and microscopic (villi, microvilli, intestinal and colonic glands, M cells, glycocalyx, and secreted mucus) levels, the action of local immune system (secretion of sIgA into the lumen), a constant influx of new phages, and their hosts from the environment, as well as chemical composition, amounts, and consistency of dietary residue. This results in phage population dynamics that are fundamentally different to ocean ecosystems. Lack of observable biomass control from phages leads to bacterial densities (10^{11} cfu g^{-1} faeces) reaching the carrying capacity of the habitat, while phage titers and VMR remain comparatively low (<0.1). An unresolved question is why in

such a dense community do we not see frequent phage bursts leading to much higher VMR levels?

Knowles et al. analyzed data on various microbial ecosystems with a focus on coral reef microbiota and demonstrated that VMR dependency on total bacterial load is unimodal, peaking at $\sim 10^6$ cfu mL^{-1} (Knowles et al., 2016). This implies that phages choose the lysogenic cycle or similar modes of temperate infection not only in case of low concentration of their hosts as was traditionally postulated, but also at very high host titers. The evolutionary logic of the latter decision for the phage is to take advantage of rapid bacterial proliferation and replicate together with their successful hosts—to "piggyback the winner." Recently, an extension of this theory has been proposed, describing bacteriophage population dynamics on generic mucosal surfaces (Silveira and Rohwer, 2016). According to the proposed model, bacterial colonization levels are highest in the top mucin layer, where "piggyback-the-winner" dynamics occurs. Deeper into the mucin layer, bacterial colonization becomes scarcer, favoring "kill-the-winner" dynamics with high VMR. The high loads of phage particles in the dense mucin adjacent to the epithelial lining provides an additional barrier against bacterial invaders. At the same time, lysogenic infection favored at the apical surface of the mucin layer can potentially facilitate lateral gene transfer and lysogenic conversion of commensal bacteria. Beneficial traits provided through lysogeny can include genes improving adherence and expanding the metabolic capacity of a strain, resistance to super-infection by related bacteriophages, and lysis of taxonomically related competitor populations by controlled lysogen induction (Manrique et al., 2017). Therefore, despite its metabolic costs, lysogeny can potentially improve overall fitness, colonization ability, and competitive exclusion of incoming pathogens and pathobionts. This theoretical model is well supported by the experimental data. For instance, Barr et al. observed that host-associated bacteriophages can specifically bind to and accumulate in mucin secretions in organisms ranging from cnidarians to humans (Barr et al., 2013). This binding is mediated by immunoglobulin (Ig)-like capsid proteins binding to glycan residues in mucin glycoproteins. The resulting model of bacteriophage-mediated immunity was termed BAM (bacteriophage adherence to mucus). In concurrence with these findings, TEM observations of colonic mucosa samples have reported higher levels of bacteriophage colonization (Lepage et al., 2008) than that observed for feces and cecal contents (Hoyles et al., 2014). Interestingly, many Ig-like protein genes detected in the human gut virome were subject to rapid diversification over time (Minot et al., 2012). In addition, it was found that the majority of luminal bacteriophages are indeed capable of lysogenic infection (Reyes et al., 2010) or lysogeny-like interaction (pseudolysogeny, carrier state interaction) with their hosts, which, despite efficient infection and replication of the phage, does not impede bacterial host proliferation (Siringan et al., 2014; Shkoporov et al., 2018a).

Contradictory to the fluctuating selection model and in agreement with "piggyback-the-winner," gut phageome composition is stable over time, with 80%–95% of phage contigs retained in a single individual over 1- to 2.5-year observation periods (Minot et al., 2013; Reyes et al., 2010). As predicted by the latter model, mutation accumulation rates were low for temperate *Caudovirales* phages but significantly higher for obligately virulent *Microviridae* phages (>1 nt substitution per 100 nt per day).

"Piggyback-the-winner" offers an elegant explanation of low VMR despite high bacterial counts in the gut. This model, however, fails to explain all observed phenomena. For example, strictly virulent *Microviridae* phages are able to persist in the gut for extended periods of time (Minot et al., 2013). At the same time, virulent *Myoviridae* phage (φHSC03) and crAss-like phages (φHSC04 and φHSC05) were able to engraft and stably persist in high amounts in mice colonized with a 15-strain artificial bacterial community without significant changes to the latter (Reyes et al., 2013). In the same experiment, a temperate *Siphoviridae* phage φHSC01 and a *Microviridae* phage φHSC02 were capable of only briefly colonizing mice, causing a transient decline of their hosts (*Bacteroides caccae* and *Bacteroides ovatus*) followed by rapid recovery of microbiota and elimination of phages.

A number of studies have focused on phage-host dynamics in the gut utilizing germ-free mice monocolonized with either non-pathogenic or pathogenic strains of *E. coli* challenged with well-characterized strictly virulent bacteriophages (Weiss et al., 2009; Maura et al., 2012). Despite dramatic expansion of bacteriophage populations, little or no decrease was seen in *E. coli* colonization levels. Interestingly, while phage T4 colonization was transient, bacteriophage T7 was maintained at $\sim 10^{11}$ cfu g^{-1} of feces with its host being stable at $\sim 10^{10}$ cfu g^{-1} of feces (Weiss et al., 2009). Similar dynamics were reported with a three-phage cocktail, which showed no detectable changes in colonization levels (Maura et al., 2012). Interestingly, eradication of phage T4 from the gut, despite the presence of large numbers of sensitive host, was not due to genetic resistance in *E. coli*. Only 20% of *E. coli* clones became resistant on prolonged *in vivo* exposure to phage T7, suggesting, on the one hand, the metabolic cost of such resistance and, on the other hand, additional factors at play in the living gut preventing the virulent phage from completely wiping out its host. We observed rapid emergence of resistance to a crAss-like phage in an *in vitro* co-cultivation system with its host, but this never resulted in complete takeover by resistant clones (Shkoporov et al., 2018a). Interestingly, the high mutation rate suggests that a genetic switch mechanism rather than random point mutations may be responsible for resistance. Furthermore, some of the mutants were readily able to revert to sensitive phenotype, again suggesting a metabolic cost associated with resistance.

Since strictly virulent bacteriophages are also available in the gut lumen, without causing any significant disturbance to the bacteriome, models other than "piggyback-the-winner" are required to explain the ecological and evolutionary forces driving maintenance of an equilibrium in the tripartite host-bacteriome-phageome system. CRISPR-Cas systems provide a powerful tool for bacterial cells to rapidly acquire resistance at population or community level upon initial contact with a new bacteriophage. Of special interest are CRISPR arrays encoded by temperate bacteriophages, adding an additional layer of complexity to the system in the form of phage versus phage antibiosis (Minot et al., 2013). The enormous gene pool of the gut microbiome and high frequency of lateral gene transfer promotes the generation of diversity in both bacteriophages and their hosts, with host switches occurring at rates much higher than those seen in reductionist *in vitro* systems (De Sordi et al.,

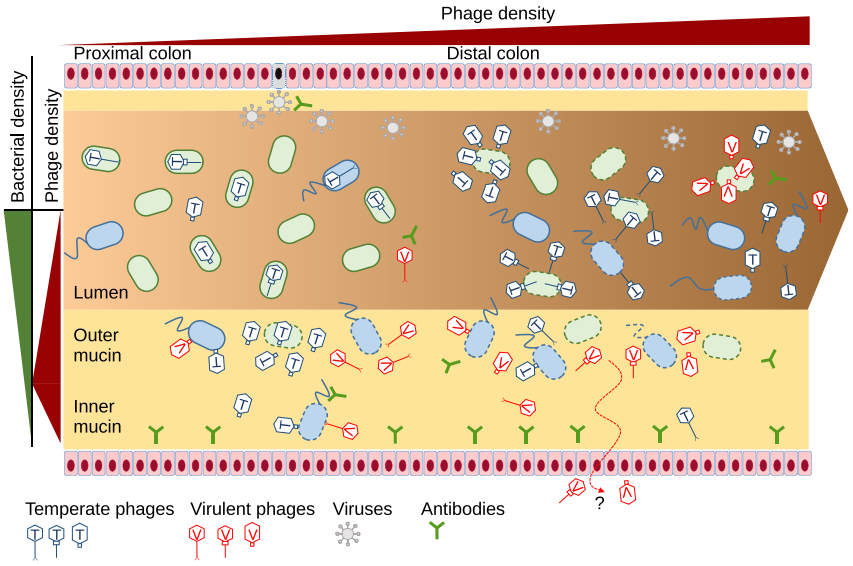

Figure 4. Bacteriophage Production in the Human Gut

The luminal contents contain dense bacterial populations, propelled in the distal direction by peristalsis and mass movement. Lysogeny if favored in the gut lumen over lytic cycle ("piggyback-the-winner" model) resulting in low virus to microbe ratio (VMR). Toward the terminal colon prophage induction is more likely due to nutrient starvation and, possibly, oxidative stress. In the thick mucin layer, bacterial density is kept to relatively low levels, causing a density-dependent switch to lytic cycle in temperate bacteriophages ("kill-the-winner" model). Large amounts of phage particles become attached to mucin where they provide BAM immunity and could potentially translocate into lamina propria and sub-mucosal layer triggering an anti-phage immune response.

2017). Pseudolysogeny, which seems to be a common mechanism of persistence for gut bacteriophages, was shown to promote accumulation of mutations in infected hosts and a build-up of resistance (Latino et al., 2016). However, despite rapid diversification, there is no evidence of existence of Red Queen dynamics in the gut, which would otherwise lead to continuous directional selection of both multi-resistant hosts and generalist bacteriophages. Instead, it seems that the metabolic cost of resistance leads to slower growth of bacteria, while the generally lower efficiency of generalist phages toward a particular host prevents them from taking over the specialists (Howard-Varona et al., 2018). Furthermore, the "enhanced infection" model predicts that metabolic cost-less resistance mutations (e.g., alterations of cell envelope) can render a bacterial clone sensitive to a different phage, leading to passive host switching (Avrani et al., 2012).

The available experimental data, however, suggest that physiological and epigenetic resistance (growth phase, expression of surface receptors), as well as physical "abiotic" factors of the gut biome (refuge model), can play a decisive role in the protection of bacterial cells from extermination by virulent bacteriophages (Lourenço et al., 2018; Reyes et al., 2013; Weiss et al., 2009). There is a significant deficit of knowledge regarding how the gut anatomy at macroscopic and microscopic levels restricts phage-host interactions. Further studies should focus on biogeographic aspects of the gut phageome along both longitudinal and radial axes. It is of special interest to investigate mechanisms of physiological and epigenetic regulation of phage infection with prominent members of gut phageome, such as crAss-like phages and *Microviridae* phages infecting *Bacteroides*, *Prevotella*, and *Faecalibacterium*.

Figure 4 illustrates available experimental data in the context of the "piggyback-the-winner" model of bacteriophage colonization dynamics in the human colon. The gut lumen is a spatially structured habitat with lytic-lysogenic switches occurring in the radial direction. The central cylinder of luminal contents contains little mucin and dense bacterial populations, propelled in the distal direction by peristalsis and mass movement. A moderate extent of phage production in the lumen happens as a result of restricted lytic cycles in case of virulent and pseudo-temperate phages, as well as by spontaneous induction of prophages. The latter event is more likely to happen as the colonic contents progress toward the distal colon due to nutrient starvation and, possibly, oxidative stress. A thick layer of mucus, protecting the epithelial lining of the colon restricts bacterial density to relatively low levels, causing a density-dependent switch from lysogenic to lytic cycle in temperate (and potentially in pseudo-temperate) bacteriophages. Large amounts of phage particles become trapped in mucus, where they provide BAM immunity and could potentially translocate into the lamina propria and sub-mucosal layer triggering an anti-phage immune response.

Conclusions and Future Perspectives

During the last decade viral metagenomics has helped to shed some light onto the "known unknown" component of the gut microbiome and to enable insights into its taxonomic composition, dynamics, and importance to human gut homeostasis. Deep sequencing of bacteriophages in the GIT has uncovered previously overlooked viral populations of high complexity with potential roles in regulation of overall microbiome composition and in the onset, progression, and treatment of gut and systemic disorders. Indeed, ample evidence exists for specific and lasting changes in phageome composition in such diverse conditions as IBD, malnutrition, and AIDS (Monaco et al., 2016; Norman et al., 2015; Reyes et al., 2015). Co-transfer and stable engraftment of bacteriophages during FMT further highlights their potential stabilizing role in the microbiome (Draper et al., 2018). Successful "correction" of antibiotic-damaged microbiomes in CDI by sterile FF transfer provides the first tantalizing evidence that phageome manipulation may be an effective therapeutic strategy (Ott et al., 2017). In fact, the gut phageome is already being used as a source of individual phages with potential therapeutic applications (Hargreaves and Clokie, 2014).

A major obstacle toward application of advanced phage-based diagnostics and therapeutics is our incomplete understanding of the structure, dynamics, and function of the normal gut phageome. Cross-sectional studies on a much larger scale are required to address our lack of knowledge on the impact of

age, sex, genetic background, and geographic variations of the healthy phageome at the population level. At an organismal and organ level, meticulous biogeographic analysis will be required to understand how phage infection shapes microbial populations in various sections of the GIT along both longitudinal and radial axes. In the light of recent discovery of bacteriophage transfer and even phage-mediated microbiota correction in FMT, it seems important to study the ability of bacteriophages to spread horizontally or vertically (from mother to infant) in human populations, as well as consequences of such spread for microbiota composition.

While cross-sectional virome studies are important to understand population-level variance, longitudinal studies will be needed to identify persistent and accessory virome fractions per individual and relationships between temporal shifts in microbiome and virome in different life stages. Traditionally for metagenomic studies, the composition of the virome is reflected as the sum of relative abundances of the individual viral sequences. Recent microbiome studies put forward the importance of absolute quantification of species and assessing total microbial loads for understanding of host-microbial interactions in health and disease (Vandeputte et al., 2017). Truly quantitative viromics with total viral load measures could be just as revelatory. In addition, new models will be required to explain the unexpectedly low VMR in the gut, as well as the unusually high temporal stability and low mutational rates of bacteriophage populations. The ability of bacterial populations in the gut to stably co-exist with high titers of virulent phages also needs further investigation. Isolation of previously uncultured but highly abundant and important phages, such as crAss-like bacteriophages and members of family *Microviridae* will help us understand their biological properties and roles in the maintenance of a dynamic equilibrium in the gut.

Finally, novel sequencing technologies and improved bioinformatic pipelines in combination with phage culture techniques and phage proteomics methods will help to shed light on the viral dark matter, link previously uncultured bacteriophages to their hosts, and help us to explore the functional potential of phage genomes.

ACKNOWLEDGMENTS

The authors thank Dr. Dimitri Scholz, Tiina O'Neill, (UCD Conway Institute of Biomolecular and Biomedical Research, Dublin, Ireland), and Dr. Andrei Bolocan (APC Microbiome Ireland, Cork, Ireland) for their help with transmission electron microscopy imaging and Dr. Stephen Stockdale and Dr. Aonghus Lavelle (APC Microbiome Ireland, Cork, Ireland) for their critical remarks during preparation of this manuscript. This work was supported by Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2273, a Science Foundation Ireland's Spokes Programme, which is co-funded under the European Regional Development Fund under Grant Number SFI/14/SP APC/B3032, and a research grant from Janssen Biotech, Inc.

REFERENCES

Aggarwala, V., Liang, G., and Bushman, F.D. (2017). Viral communities of the human gut: metagenomic analysis of composition and dynamics. Mob. DNA *8*, 12.

Aherfi, S., Colson, P., La Scola, B., and Raoult, D. (2016). Giant viruses of amoebas: an update. Front. Microbiol. *7*, 349.

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., and Wishart, D.S. (2016). PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. *44* (W1), W16–W21.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., Fernandes, G.R., Tap, J., Bruls, T., Batto, J.-M., et al.; MetaHIT Consortium (2011). Enterotypes of the human gut microbiome. Nature *473*, 174–180.

Avrani, S., Schwartz, D.A., and Lindell, D. (2012). Virus-host swinging party in the oceans: Incorporating biological complexity into paradigms of antagonistic coexistence. Mob. Genet. Elements *2*, 88–95.

Aziz, Q., Doré, J., Emmanuel, A., Guarner, F., and Quigley, E.M.M. (2013). Gut microbiota and gastrointestinal health: current concepts and future directions. Neurogastroenterol. Motil. *25*, 4–15.

Barr, J.J., Auro, R., Furlan, M., Whiteson, K.L., Erb, M.L., Pogliano, J., Stotland, A., Wolkowicz, R., Cutting, A.S., Doran, K.S., et al. (2013). Bacteriophage adhering to mucus provide a non-host-derived immunity. Proc. Natl. Acad. Sci. USA *110*, 10771–10776.

Beaulaurier, J., Zhu, S., Deikus, G., Mogno, I., Zhang, X.-S., Davis-Richardson, A., Canepa, R., Triplett, E.W., Faith, J.J., Sebra, R., et al. (2018). Metagenomic binning and association of plasmids with bacterial host genomes using DNA methylation. Nat. Biotechnol. *36*, 61–69.

Belkaid, Y., and Hand, T.W. (2014). Role of the microbiota in immunity and inflammation. Cell *157*, 121–141.

Bolduc, B., Jang, H.B., Doulcier, G., You, Z.-Q., Roux, S., and Sullivan, M.B. (2017). vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect *Archaea* and *Bacteria*. PeerJ *5*, e3243.

Bonilla, N., Santiago, T., Marcos, P., Urdaneta, M., Domingo, J.S., and Toranzos, G.A. (2010). Enterophages, a group of phages infecting Enterococcus faecalis, and their potential as alternate indicators of human faecal contamination. Water Sci. Technol. *61*, 293–300.

Breitbart, M., Hewson, I., Felts, B., Mahaffy, J.M., Nulton, J., Salamon, P., and Rohwer, F. (2003). Metagenomic analyses of an uncultured viral community from human feces. J. Bacteriol. *185*, 6220–6223.

Breitbart, M., Haynes, M., Kelley, S., Angly, F., Edwards, R.A., Felts, B., Mahaffy, J.M., Mueller, J., Nulton, J., Rayhawk, S., et al. (2008). Viral diversity and dynamics in an infant gut. Res. Microbiol. *159*, 367–373.

Brettin, T., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Olsen, G.J., Olson, R., Overbeek, R., Parrello, B., Pusch, G.D., et al. (2015). RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci. Rep. *5*, 8365.

Brown, M.R., Camézuli, S., Davenport, R.J., Petelenz-Kurdziel, E., Øvreås, L., and Curtis, T.P. (2015). Flow cytometric quantification of viruses in activated sludge. Water Res. *68*, 414–422.

Browne, H.P., Forster, S.C., Anonye, B.O., Kumar, N., Neville, B.A., Stares, M.D., Goulding, D., and Lawley, T.D. (2016). Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. Nature *533*, 543–546.

Carding, S.R., Davis, N., and Hoyles, L. (2017). Review article: the human intestinal virome in health and disease. Aliment. Pharmacol. Ther. *46*, 800–815.

Castro-Mejía, J.L., Muhammed, M.K., Kot, W., Neve, H., Franz, C.M.A.P., Hansen, L.H., Vogensen, F.K., and Nielsen, D.S. (2015). Optimizing protocols for extraction of bacteriophages prior to metagenomic analyses of phage communities in the human gut. Microbiome *3*, 64.

Cinek, O., Kramna, L., Lin, J., Oikarinen, S., Kolarova, K., Ilonen, J., Simell, O., Veijola, R., Autio, R., and Hyöty, H. (2017). Imbalance of bacteriome profiles within the Finnish Diabetes Prediction and Prevention study: Parallel use of 16S profiling and virome sequencing in stool samples from children with islet autoimmunity and matched controls. Pediatr. Diabetes *18*, 588–598.

Clarke, E.L., Taylor, L.J., Zhao, C., Connell, J., Bushman, F.D., and Bittinger, K. (2018). Sunbeam: an extensible pipeline for analyzing metagenomic sequencing experiments. bioRxiv. https://doi.org/10.1101/326363.

Conceição-Neto, N., Zeller, M., Lefrère, H., De Bruyn, P., Beller, L., Deboutte, W., Yinda, C.K., Lavigne, R., Maes, P., Van Ranst, M., et al. (2015). Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. Sci. Rep. *5*, 16532.

Cornuault, J.K., Petit, M.-A., Mariadassou, M., Benevides, L., Moncaut, E., Langella, P., Sokol, H., and De Paepe, M. (2018). Phages infecting Faecalibacterium prausnitzii belong to novel viral genera that help to decipher intestinal viromes. Microbiome *6*, 65.

Cui, J., Schlub, T.E., and Holmes, E.C. (2014). An allometric relationship between the genome length and virion volume of viruses. J. Virol. *88*, 6403–6410.

Darling, A.C.E., Mau, B., Blattner, F.R., and Perna, N.T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. *14*, 1394–1403.

De Sordi, L., Khanna, V., and Debarbieux, L. (2017). The gut microbiota facilitates drifts in the genetic diversity and infectivity of bacterial viruses. Cell Host Microbe *22*, 801–808.e3.

Devoto, A.E., Santini, J.M., Olm, M.R., Anantharaman, K., Munk, P., Tung, J., Archie, E.A., Turnbaugh, P.J., Seed, K.D., Blekhman, R., et al. (2018). Megaphage infect Prevotella and variants are widespread in gut microbiomes. bioRxiv. https://doi.org/10.1101/356790.

Dhillon, T.S., Dhillon, E.K., Chau, H.C., Li, W.K., and Tsang, A.H. (1976). Studies on bacteriophage distribution: virulent and temperate bacteriophage content of mammalian feces. Appl. Environ. Microbiol. *32*, 68–74.

Dinan, T.G., and Cryan, J.F. (2017). Gut instincts: microbiota as a key regulator of brain development, ageing and neurodegeneration. J. Physiol. *595*, 489–503.

Draper, L.A., Ryan, F.J., Smith, M.K., Jalanka, J., Mattila, E., Arkkila, P.A., Ross, R.P., Satokari, R., and Hill, C. (2018). Long-term colonisation with donor bacteriophages following successful faecal microbial transplantation. Microbiome *6*, 220.

Dutilh, B.E., Cassman, N., McNair, K., Sanchez, S.E., Silva, G.G.Z., Boling, L., Barr, J.J., Speth, D.R., Seguritan, V., Aziz, R.K., et al. (2014). A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat. Commun. *5*, 4498.

Džunková, M., D'Auria, G., and Moya, A. (2015). Direct sequencing of human gut virome fractions obtained by flow cytometry. Front. Microbiol. *6*, 955.

Enault, F., Briet, A., Bouteille, L., Roux, S., Sullivan, M.B., and Petit, M.-A. (2017). Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. ISME J. *11*, 237–247.

Everard, A., and Cani, P.D. (2013). Diabetes, obesity and gut microbiota. Best Pract. Res. Clin. Gastroenterol. *27*, 73–83.

García-Aljaro, C., Ballesté, E., Muniesa, M., and Jofre, J. (2017). Determination of crAssphage in water samples and applicability for tracking human faecal pollution. Microb. Biotechnol. *10*, 1775–1780.

Garneau, J.R., Depardieu, F., Fortier, L.-C., Bikard, D., and Monot, M. (2017). PhageTerm: a tool for fast and accurate determination of phage termini and packaging mechanism using next-generation sequencing data. Sci. Rep. *7*, 8292.

Grazziotin, A.L., Koonin, E.V., and Kristensen, D.M. (2017). Prokaryotic Virus Orthologous Groups (pVOGs): a resource for comparative genomics and protein family annotation. Nucleic Acids Res. *45* (D1), D491–D498.

Guerin, E., Shkoporov, A., Stockdale, S.R., Clooney, A.G., Ryan, F.J., Sutton, T.D.S., Draper, L.A., Gonzalez-Tortuero, E., Ross, R.P., and Hill, C. (2018). Biology and taxonomy of crAss-like bacteriophages, the most abundant virus in the human gut. Cell Host Microbe *24*, 653–664.e6.

Hargreaves, K.R., and Clokie, M.R.J. (2014). Clostridium difficile phages: still difficult? Front. Microbiol. *5*, 184.

Hoffmann, C., Dollive, S., Grunberg, S., Chen, J., Li, H., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2013). Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents. PLoS One *8*, e66019.

Howard-Varona, C., Hargreaves, K.R., Solonenko, N.E., Markillie, L.M., White, R.A., 3rd, 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.

Hoyles, L., McCartney, A.L., Neve, H., Gibson, G.R., Sanderson, J.D., Heller, K.J., and van Sinderen, D. (2014). Characterization of virus-like particles associated with the human faecal and caecal microbiota. Res. Microbiol. *165*, 803–812.

Huseyin, C.E., O'Toole, P.W., Cotter, P.D., and Scanlan, P.D. (2017). Forgotten fungi-the gut mycobiome in human health and disease. FEMS Microbiol. Rev. *41*, 479–511.

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.

Kleiner, M., Hooper, L.V., and Duerkop, B.A. (2015). Evaluation of methods to purify virus-like particles for metagenomic sequencing of intestinal viromes. BMC Genomics *16*, 7.

Knowles, B., Silveira, C.B., Bailey, B.A., Barott, K., Cantu, V.A., Cobián-Güemes, A.G., Coutinho, F.H., Dinsdale, E.A., Felts, B., Furby, K.A., et al. (2016). Lytic to temperate switching of viral communities. Nature *531*, 466–470.

Krupovic, M., and Forterre, P. (2011). Microviridae goes temperate: microvirus-related proviruses reside in the genomes of Bacteroidetes. PLoS One *6*, e19893.

Krupovic, M., Cvirkaite-Krupovic, V., Iranzo, J., Prangishvili, D., and Koonin, E.V. (2018). Viruses of archaea: structural, functional, environmental and evolutionary genomics. Virus Res. *244*, 181–193.

Laforest-Lapointe, I., and Arrieta, M.-C. (2018). Microbial eukaryotes: a missing link in gut microbiome studies. mSystems *3*, e00201–e00217.

Latino, L., Midoux, C., Hauck, Y., Vergnaud, G., and Pourcel, C. (2016). Pseudolysogeny and sequential mutations build multiresistance to virulent bacteriophages in Pseudomonas aeruginosa. Microbiology *162*, 748–763.

Lepage, P., Colombet, J., Marteau, P., Sime-Ngando, T., Doré, J., and Leclerc, M. (2008). Dysbiosis in inflammatory bowel disease: a role for bacteriophages? Gut *57*, 424–425.

Liang, Y.Y., Zhang, W., Tong, Y.G., and Chen, S.P. (2016). crAssphage is not associated with diarrhoea and has high genetic diversity. Epidemiol. Infect. *144*, 3549–3553.

Lim, E.S., Zhou, Y., Zhao, G., Bauer, I.K., Droit, L., Ndao, I.M., Warner, B.B., Tarr, P.I., Wang, D., and Holtz, L.R. (2015). Early life dynamics of the human gut virome and bacterial microbiome in infants. Nat. Med. *21*, 1228–1234.

Lima-Mendez, G., Van Helden, J., Toussaint, A., and Leplae, R. (2008). Reticulate representation of evolutionary and functional relationships between phage genomes. Mol. Biol. Evol. *25*, 762–777.

Lima-Mendez, G., Toussaint, A., and Leplae, R. (2011). A modular view of the bacteriophage genomic space: identification of host and lifestyle marker modules. Res. Microbiol. *162*, 737–746.

Lourenço, M., De Sordi, L., and Debarbieux, L. (2018). The diversity of bacterial lifestyles hampers bacteriophage tenacity. Viruses *10*, 327.

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. USA *113*, 10400–10405.

Manrique, P., Dills, M., and Young, M.J. (2017). The human gut phage community and its implications for health and disease. Viruses *9*, 141.

Martinez-Castillo, A., Quirós, P., Navarro, F., Miró, E., and Muniesa, M. (2013). Shiga toxin 2-encoding bacteriophages in human fecal samples from healthy individuals. Appl. Environ. Microbiol. *79*, 4862–4868.

Martinez-Hernandez, F., Fornas, O., Lluesma Gomez, M., Bolduc, B., de la Cruz Peña, M.J., Martínez, J.M., Anton, J., Gasol, J.M., Rosselli, R., Rodriguez-Valera, F., et al. (2017). Single-virus genomics reveals hidden cosmopolitan and abundant viruses. Nat. Commun. *8*, 15892.

Maura, D., Morello, E., du Merle, L., Bomme, P., Le Bouguénec, C., and Debarbieux, L. (2012). Intestinal colonization by enteroaggregative Escherichia coli supports long-term bacteriophage replication in mice. Environ. Microbiol. *14*, 1844–1854.

Mavrich, T.N., Casey, E., Oliveira, J., Bottacini, F., James, K., Franz, C.M.A.P., Lugli, G.A., Neve, H., Ventura, M., Hatfull, G.F., et al. (2018). Characterization and induction of prophages in human gut-associated Bifidobacterium hosts. Sci. Rep. *8*, 12772.

McCann, A., Ryan, F.J., Stockdale, S.R., Dalmasso, M., Blake, T., Ryan, C.A., Stanton, C., Mills, S., Ross, P.R., and Hill, C. (2018). Viromes of one year old

infants reveal the impact of birth mode on microbiome diversity. PeerJ *6*, e4694.

Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S.A., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2011). The human gut virome: inter-individual variation and dynamic response to diet. Genome Res. *21*, 1616–1625.

Minot, S., Grunberg, S., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2012). Hypervariable loci in the human gut virome. Proc. Natl. Acad. Sci. USA *109*, 3962–3966.

Minot, S., Bryson, A., Chehoud, C., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2013). Rapid evolution of the human gut virome. Proc. Natl. Acad. Sci. USA *110*, 12450–12455.

Modi, S.R., Lee, H.H., Spina, C.S., and Collins, J.J. (2013). Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. Nature *499*, 219–222.

Monaco, C.L., Gootenberg, D.B., Zhao, G., Handley, S.A., Ghebremichael, M.S., Lim, E.S., Lankowski, A., Baldridge, M.T., Wilen, C.B., Flagg, M., et al. (2016). Altered virome and bacterial microbiome in human immunodeficiency virus-associated acquired immunodeficiency syndrome. Cell Host Microbe *19*, 311–322.

Norman, J.M., Handley, S.A., Baldridge, M.T., Droit, L., Liu, C.Y., Keller, B.C., Kambal, A., Monaco, C.L., Zhao, G., Fleshner, P., et al. (2015). Disease-specific alterations in the enteric virome in inflammatory bowel disease. Cell *160*, 447–460.

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P.A. (2017). metaSPAdes: a new versatile metagenomic assembler. Genome Res. *27*, 824–834.

O'Hara, A.M., and Shanahan, F. (2006). The gut flora as a forgotten organ. EMBO Rep. *7*, 688–693.

Ogilvie, L.A., Caplin, J., Dedi, C., Diston, D., Cheek, E., Bowler, L., Taylor, H., Ebdon, J., and Jones, B.V. (2012). Comparative (meta)genomic analysis and ecological profiling of human gut-specific bacteriophage φB124-14. PLoS One *7*, e35053.

Ott, S.J., Waetzig, G.H., Rehman, A., Moltzau-Anderson, J., Bharti, R., Grasis, J.A., Cassidy, L., Tholey, A., Fickenscher, H., Seegert, D., et al. (2017). Efficacy of sterile fecal filtrate transfer for treating patients with clostridium difficile infection. Gastroenterology *152*, 799–811.e7.

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.

Paez-Espino, D., Chen, I.A., Palaniappan, K., Ratner, A., Chu, K., Szeto, E., Pillay, M., Huang, J., Markowitz, V.M., Nielsen, T., et al. (2017). IMG/VR: a database of cultured and uncultured DNA Viruses and retroviruses. Nucleic Acids Res. *45* (D1), D457–D465.

Paradis, E., and Schliep, K. (2018). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. https://doi.org/10.1093/bioinformatics/bty633.

Prangishvili, D., Bamford, D.H., Forterre, P., Iranzo, J., Koonin, E.V., and Krupovic, M. (2017). The enigmatic archaeal virosphere. Nat. Rev. Microbiol. *15*, 724–739.

Puig, M., and Gironés, R. (1999). Genomic structure of phage B40-8 of Bacteroides fragilis. Microbiology *145*, 1661–1670.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., et al.; MetaHIT Consortium (2010). A human gut microbial gene catalogue established by metagenomic sequencing. Nature *464*, 59–65.

Rajilić-Stojanović, M., and de Vos, W.M. (2014). The first 1000 cultured species of the human gastrointestinal microbiota. FEMS Microbiol. Rev. *38*, 996–1047.

Reyes, A., Haynes, M., Hanson, N., Angly, F.E., Heath, A.C., Rohwer, F., and Gordon, J.I. (2010). Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature *466*, 334–338.

Reyes, A., Wu, M., McNulty, N.P., Rohwer, F.L., and Gordon, J.I. (2013). Gnotobiotic mouse model of phage-bacterial host dynamics in the human gut. Proc. Natl. Acad. Sci. USA *110*, 20236–20241.

Reyes, A., Blanton, L.V., Cao, S., Zhao, G., Manary, M., Trehan, I., Smith, M.I., Wang, D., Virgin, H.W., Rohwer, F., and Gordon, J.I. (2015). Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. Proc. Natl. Acad. Sci. USA *112*, 11941–11946.

Roux, S., Krupovic, M., Poulet, A., Debroas, D., and Enault, F. (2012). Evolution and diversity of the Microviridae viral family through a collection of 81 new complete genomes assembled from virome reads. PLoS One *7*, e40418.

Roux, S., Krupovic, M., Debroas, D., Forterre, P., and Enault, F. (2013). Assessment of viral community functional potential from viral metagenomes may be hampered by contamination with cellular sequences. Open Biol. *3*, 130160.

Roux, S., Tournayre, J., Mahul, A., Debroas, D., and Enault, F. (2014). Metavir 2: new tools for viral metagenome comparison and assembled virome analysis. BMC Bioinformatics *15*, 76.

Roux, S., Hallam, S.J., Woyke, T., and Sullivan, M.B. (2015a). Viral dark matter and virus-host interactions resolved from publicly available microbial genomes. Elife *4*, e08490.

Roux, S., Enault, F., Hurwitz, B.L., and Sullivan, M.B. (2015b). VirSorter: mining viral signal from microbial genomic data. PeerJ *3*, e985.

Roux, S., Solonenko, N.E., Dang, V.T., Poulos, B.T., Schwenck, S.M., Goldsmith, D.B., Coleman, M.L., Breitbart, M., and Sullivan, M.B. (2016). Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. PeerJ *4*, e2777.

Roux, S., Emerson, J.B., Eloe-Fadrosh, E.A., and Sullivan, M.B. (2017). Benchmarking Viromics: An In Silico Evaluation of Metagenome-Enabled Estimates of Viral Community Composition and Diversity (PeerJ Inc.).

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics *30*, 2068–2069.

Seguritan, V., Alves, N., Jr., Arnoult, M., Raymond, A., Lorimer, D., Burgin, A.B., Jr., Salamon, P., and Segall, A.M. (2012). Artificial neural networks trained to detect viral and phage structural proteins. PLoS Comput. Biol. *8*, e1002657.

Sender, R., Fuchs, S., and Milo, R. (2016). Revised estimates for the number of human and bacteria cells in the body. PLoS Biol. *14*, e1002533.

Shkoporov, A., Khokhlova, E.V., Fitzgerald, C.B., Stockdale, S.R., Draper, L.A., Ross, R.P., and Hill, C. (2018a). ΦCrAss001, a member of the most abundant bacteriophage family in the human gut, infects Bacteroides. Nat. Commun. *9*, 4781.

Shkoporov, A.N., Ryan, F.J., Draper, L.A., Forde, A., Stockdale, S.R., Daly, K.M., McDonnell, S.A., Nolan, J.A., Sutton, T.D.S., Dalmasso, M., et al. (2018b). Reproducible protocols for metagenomic analysis of human faecal phageomes. Microbiome *6*, 68.

Silveira, C.B., and Rohwer, F.L. (2016). Piggyback-the-Winner in host-associated microbial communities. NPJ Biofilms Microbiomes *2*, 16010.

Siringan, P., Connerton, P.L., Cummings, N.J., and Connerton, I.F. (2014). Alternative bacteriophage life cycles: the carrier state of Campylobacter jejuni. Open Biol. *4*, 130200.

Stockdale, S., Ryan, F., McCann, A., Dalmasso, M., Ross, P., and Hill, C. (2018). Viral Dark Matter in the Gut Virome of Elderly Humans. Preprints. https://doi.org/10.20944/preprints201807.0101.v1.

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

Sutton, T.D.S., Clooney, A.G., Ryan, F.J., Ross, R.P., and Hill, C. (2019). Choice of assembly software has a critical impact on virome characterisation. bioRxiv. https://doi.org/10.1101/479105.

Thingstad, T.F., Bratbak, G., and Heldal, M. (2008). Aquatic phage ecology. In Bacteriophage Ecology, Population Growth, Evolution, and Impact of Bacterial Viruses, S.T. Abedon, ed. (Cambridge University Press), pp. 251–280.

Vandeputte, D., Kathagen, G., D'hoe, K., Vieira-Silva, S., Valles-Colomer, M., Sabino, J., Wang, J., Tito, R.Y., De Commer, L., Darzi, Y., et al. (2017). Quantitative microbiome profiling links gut community variation to microbial load. Nature *551*, 507–511.

Wampach, L., Heintz-Buschart, A., Hogan, A., Muller, E.E.L., Narayanasamy, S., Laczny, C.C., Hugerth, L.W., Bindl, L., Bottu, J., Andersson, A.F., et al. (2017). Colonization and succession within the human gut microbiome by archaea, bacteria, and microeukaryotes during the first year of life. Front. Microbiol. *8*, 738.

Warwick-Dugdale, J., Solonenko, N., Moore, K., Chittick, L., Gregory, A.C., Allen, M.J., Sullivan, M.B., and Temperton, B. (2018). Long-read metagenomics reveals cryptic and abundant marine viruses. bioRxiv. https://doi.org/10.1101/345041.

Weinbauer, M.G. (2004). Ecology of prokaryotic viruses. FEMS Microbiol. Rev. *28*, 127–181.

Weiss, M., Denou, E., Bruttin, A., Serra-Moreno, R., Dillmann, M.-L., and Brüssow, H. (2009). In vivo replication of T4 and T7 bacteriophages in germ-free mice colonized with Escherichia coli. Virology *393*, 16–23.

Wigington, C.H., Sonderegger, D., Brussaard, C.P.D., Buchan, A., Finke, J.F., Fuhrman, J.A., Lennon, J.T., Middelboe, M., Suttle, C.A., Stock, C., et al. (2016). Re-examination of the relationship between marine virus and microbial cell abundances. Nat. Microbiol. *1*, 15024.

Wommack, K.E., and Colwell, R.R. (2000). Virioplankton: viruses in aquatic ecosystems. Microbiol. Mol. Biol. Rev. *64*, 69–114.

Yap, M.L., and Rossmann, M.G. (2014). Structure and function of bacteriophage T4. Future Microbiol. *9*, 1319–1327.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., et al. (2012). Human gut microbiome viewed across age and geography. Nature *486*, 222–227.

Yutin, N., Makarova, K.S., Gussow, A.B., Krupovic, M., Segall, A., Edwards, R.A., and Koonin, E.V. (2018). Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut. Nat. Microbiol. *3*, 38–46.

Zhang, T., Breitbart, M., Lee, W.H., Run, J.-Q., Wei, C.L., Soh, S.W.L., Hibberd, M.L., Liu, E.T., Rohwer, F., and Ruan, Y. (2006). RNA viral community in human feces: prevalence of plant pathogenic viruses. PLoS Biol. *4*, e3.