

Review article: the human intestinal virome in health and disease

S. R. Carding^{1,2}  | N. Davis¹ | L. Hoyles³ 

¹Norwich Medical School, University of East Anglia, Norwich, UK

²The Gut Health and Food Safety Research Programme, The Quadram Institute, Norwich Research Park, Norwich, UK

³Department of Surgery and Cancer, Imperial College London, London, UK

Correspondence

Prof. SR Carding, The Gut Health and Food Safety Research Programme, The Quadram Institute, Norwich Research Park Norwich, UK.

Email: simon.carding@quadram.ac.uk

Funding information

This study was funded in part by an institutional grant from the Biotechnology and Biological Sciences Research Council, grant number BB/J004529/1 (SRC), by an MRC Intermediate Research Fellowship in Data Science awarded to LH, and by a medical student research bursary from Invest In ME charity awarded to ND. This work used the computing resources of the UK MEDical BIOinformatics partnership—aggregation, integration, visualisation and analysis of large, complex data (UK MED-BIO), which is supported by the Medical Research Council, grant number MR/L01632X/1

Summary

Background: The human virome consists of animal-cell viruses causing transient infections, bacteriophage (phage) predators of bacteria and archaea, endogenous retroviruses and viruses causing persistent and latent infections. High-throughput, inexpensive, sensitive sequencing methods and metagenomics now make it possible to study the contribution dsDNA, ssDNA and RNA virus-like particles make to the human virome, and in particular the intestinal virome.

Aim: To review and evaluate the pioneering studies that have attempted to characterise the human virome and generated an increased interest in understanding how the intestinal virome might contribute to maintaining health, and the pathogenesis of chronic diseases.

Methods: Relevant virome-related articles were selected for review following extensive language- and date-unrestricted, electronic searches of the literature.

Results: The human intestinal virome is personalised and stable, and dominated by phages. It develops soon after birth in parallel with prokaryotic communities of the microbiota, becoming established during the first few years of life. By infecting specific populations of bacteria, phages can alter microbiota structure by killing host cells or altering their phenotype, enabling phages to contribute to maintaining intestinal homeostasis or microbial imbalance (dysbiosis), and the development of chronic infectious and autoimmune diseases including HIV infection and Crohn's disease, respectively.

Conclusions: Our understanding of the intestinal virome is fragmented and requires standardised methods for virus isolation and sequencing to provide a more complete picture of the virome, which is key to explaining the basis of virome-disease associations, and how enteric viruses can contribute to disease aetiologies and be rationalised as targets for interventions.

The Handling Editor for this article was Professor Jonathan Rhodes, and this commissioned review was accepted for publication after full peer-review.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2017 The Authors. *Alimentary Pharmacology & Therapeutics* published by John Wiley & Sons Ltd

1 | THE HUMAN VIROME

Viruses are thought to be the most abundant and diverse entities on Earth, numbering as many as 10^{31} virus-like particles (VLPs),¹ though this paradigm is likely to change in light of information from large-scale sequencing surveys of marine environments and improved analytical tools.² With the advent of new, sequence-based technologies that do not rely on the ability to isolate viruses for their identification, it is now possible to define and characterise viruses in different environmental samples in greater detail than ever before, which has resulted in an increased interest in the role the viral assemblage of the human gut microbiota plays in health and disease. The following reviews our current knowledge on the human intestinal virome.

A virome comprises all the nucleic acids (ssDNA, dsDNA, ssRNA and dsRNA) belonging to the VLPs associated with a particular ecosystem. The human virome is a genetically complex component of the microbiome, with the blood, nose, skin, conjunctiva, mouth, vagina, lungs and gastrointestinal (GI) tract harbouring their own distinct virus assemblages (Table S1). The genetic content of VLPs comprising bacteriophages (phages) that infect bacteria and archaea

and, to a much lesser extent, human-, plant-, amoebae- and animal-infecting viruses found along the GI tract constitute the human intestinal virome (Figure 1).

2 | BACTERIOPHAGES

Phages can be lytic or lysogenic. Lytic (virulent) phages bind to specific host-cell receptors, then penetrate and infect their host cell to hijack its replication and translation machinery to produce virions; once sufficient virions have accumulated (tens to thousands, depending on the phage), the lytic enzymes they produce cause the host cell to lyse, releasing the virions into the surrounding environment where they can infect new host cells. Lytic phages can have narrow or broad host ranges, infecting only one strain of a prokaryote or multiple species of closely related prokaryotes. These entities have been used in the past to treat infections, and have recently been revisited as alternatives to antibiotic therapies for a range of infections and as means of controlling food-contaminating bacteria such as *Listeria* spp.³ Their gene products are also of interest in biotechnological and medical applications.⁴ Lysogenic

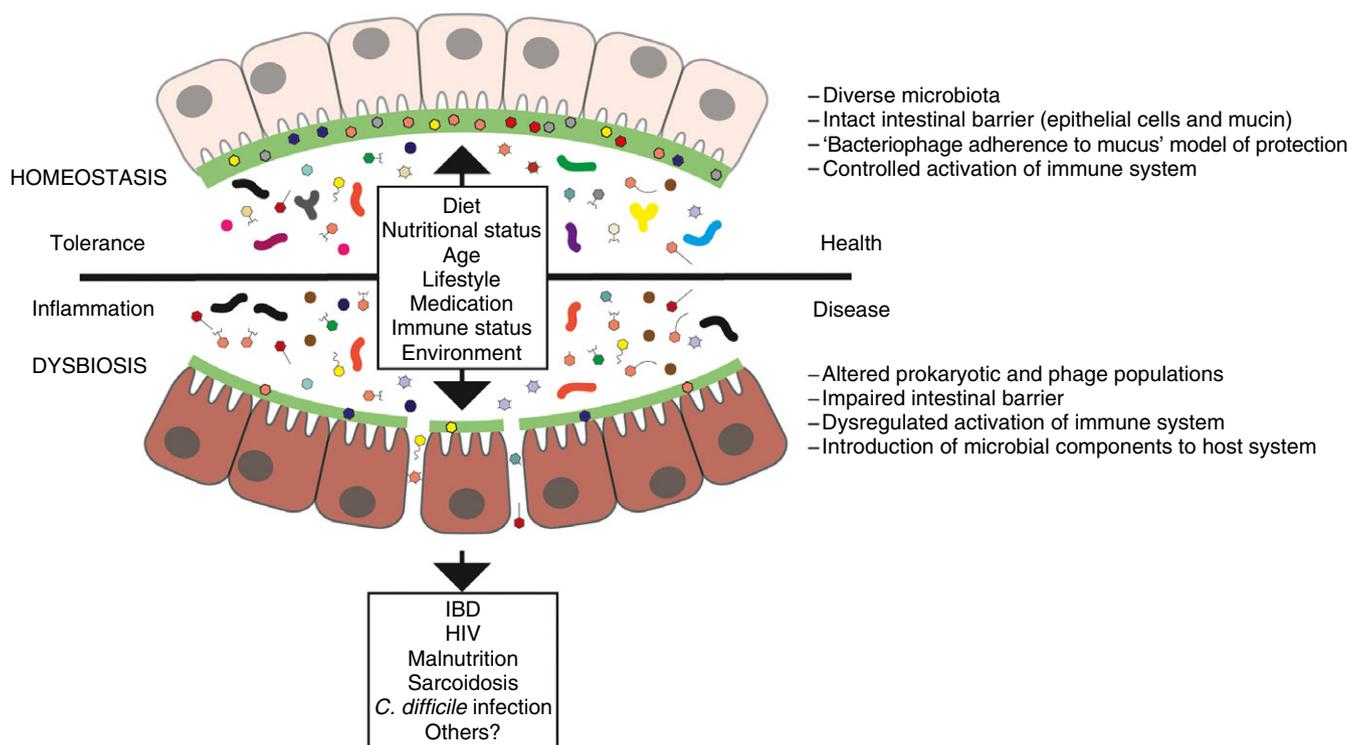


FIGURE 1 The mammalian intestinal virome comprises viruses that infect eukaryotic and prokaryotic cells. It is established soon after birth and is dominated by viruses that infect bacteria (ie, phages). The virome establishes a mutualistic relationship with eukaryotes/prokaryotes, contributing to intestinal homeostasis by influencing microbial ecology and host immunity. Composition of the virome is influenced by numerous factors that affect viruses directly (infection) or change host-cell populations (eg, antibiotics, diet). Members of the virome may contribute to the pathogenesis of certain diseases via microbial host lysis leading to dysbiosis, infection of epithelial cells, and/or translocation of the compromised or damaged mucosal barrier to gain access to underlying tissues and immune cells, leading to immune activation. Dysbiosis can be defined as a microbial imbalance or any changes to the composition of resident microbial communities relative to the community found in healthy individuals.^{101–103} Virome association with certain disease states is characterised by changes in diversity, and predominance of specific virotypes (eg, members of the order *Caudovirales* in IBD)

(temperate) phages do not kill their host but instead integrate into their host's genome without interfering with its replication, incorporating the phage into its genome as a prophage that is transmitted to its progeny at each cell division. Lysogenic phages can be converted to a lytic cycle in response to environmental stressors (eg, antibiotics).

Phage-host interactions influence host and viral evolution. The ability of phages to transfer genes from one prokaryotic host to another can lead to increased diversification of viral species, and increased antibiotic resistance and/or induction of toxins or virulence factors in prokaryotes.⁵ Some phages alter the antigenicity of their hosts by producing enzymes that modify the O-antigen component of lipopolysaccharides.⁶ Modification of surface structures of prokaryotes has the potential to affect microbial interactions with the human host, and influence niche specialisation within the GI tract.⁷ Presence of clustered regularly interspaced short palindromic repeats (CRISPRs) confers upon prokaryotes resistance to phage infection and contributes to prokaryotic adaptive immunity. Analyses of metagenomic sequence data provide detailed information on phage-host and phage-phage competition within the human faecal microbiome, implying CRISPR spacers are actively and continuously acquired by prokaryotes in response to the presence of phages in the GI tract.⁸ The potential effects of such phage-host interactions on microbiota composition/function or host health are unknown.

3 | CHARACTERISING THE HUMAN INTESTINAL VIROME

Viruses do not encode universally conserved genes such as the 16S or 18S rRNA genes of prokaryotes and eukaryotes, respectively, and are genetically highly diverse. Consequently, it is not possible to use metataxonomic approaches such as 16S rRNA gene sequencing to characterise VLPs within ecosystems. Traditionally, classical approaches—mainly microscopy and cultivation—have been relied upon to characterise VLPs in the human gut.

Based on transmission electron microscopy (TEM), mucosal samples contain $\sim 1.2 \times 10^9$ VLPs/biopsy.⁹ VLPs have been detected in caecal contents at 10^6 /mL using TEM¹⁰ with faeces harbouring 10^8 – 10^9 VLPs/g wet weight.^{10,11} In all GI contents examined to date by microscopy, the overwhelming majority of VLPs have been phages of the order *Caudovirales*, with the human GI tract estimated to harbour 10^{15} phages in total.^{3,9,10,12} The order *Caudovirales* encompasses most known phages (Figure 2), comprising the families *Myoviridae* (nonenveloped, head-tail structure, contractile, dsDNA genomes of 30–250 kb), *Siphoviridae* (nonenveloped, head-tail structure, noncontractile, dsDNA genomes of ~ 50 kb) and *Podoviridae* (nonenveloped, head-short tail structure, dsDNA genomes of 40–42 kb). Every individual harbours a morphologically unique phage population.^{10,12}

For many years, the isolation of gut-associated phages was confined to samples derived from sewage or water sources, with the

presence of phages considered indicative of nonspecific (animal, human) faecal contamination. Recently attention has turned to isolating phages directly from human microbiota samples (Table 1). While in its infancy, this work holds great promise for improving our knowledge of the gut virome, will facilitate metagenomic studies and aid in vitro and in vivo studies investigating the influence of phages on prokaryote populations in the human gut. Significant improvements to isolation methods are needed to allow the efficient and reliable isolation and characterisation of phages infecting fastidious anaerobic prokaryotes. In addition, no single isolation approach is applicable to all prokaryotes, and based on evidence from *Bacteroides* spp. it is likely gut-associated phages have very narrow host ranges.⁷

Study of viruses in clinical settings will be facilitated by methods such as the Virome Capture Sequencing platform for Vertebrate viruses (VirCapSeq-VERT), a system that uses ~ 2 million probes targeting the genomes of 207 viruses known to infect humans and other vertebrates to capture viral nucleic acids present in blood and tissue homogenates.¹³

4 | METAGENOMICS AND THE HUMAN INTESTINAL VIROME

The only reliable molecular method currently available for routine surveys of the human virome is metagenomics. Metagenomics is a culture-independent, molecular-based approach that allows functional and sequence-based analyses of the collective microbial genomes contained in an environmental sample, providing a powerful approach for exploring the ecology of complex microbial communities.¹⁴ It has been used to examine prokaryotic communities associated with the human gut microbiome in health and disease,^{15–18} and to examine viromes associated with different regions of the human body (Table S1).

A protocol involving homogenisation of faeces in buffer, centrifugation (to remove cell debris), tangential flow filtration (TFF) (to concentrate large-volume samples and isolate VLPs), ultracentrifugation and metagenomic reconstruction was used to characterise the first human faecal virome.¹⁹ However, recent improvements in recovery methods give the potential to characterise human-associated VLP assemblages at the molecular level in greater detail than ever before (Table 2). These methods have yet to be compared with one another directly, and the issue of which DNA extraction method or kit is best for use with gut virome samples has not been addressed. To avoid the problems that have beset studies of the prokaryotic component of the gut microbiome with respect to reproducibility, contaminants and comparable approaches across studies and datasets, it is important the virome community adopts robust, standardised methods at the earliest opportunity.

The first faecal metagenome was generated from the faeces of a 33-year-old healthy male using a clone-based strategy.¹⁹ It was predicted to comprise $\sim 1,200$ virotypes, with the most abundant virus making up 4% of the total. Siphophage sequences made up the

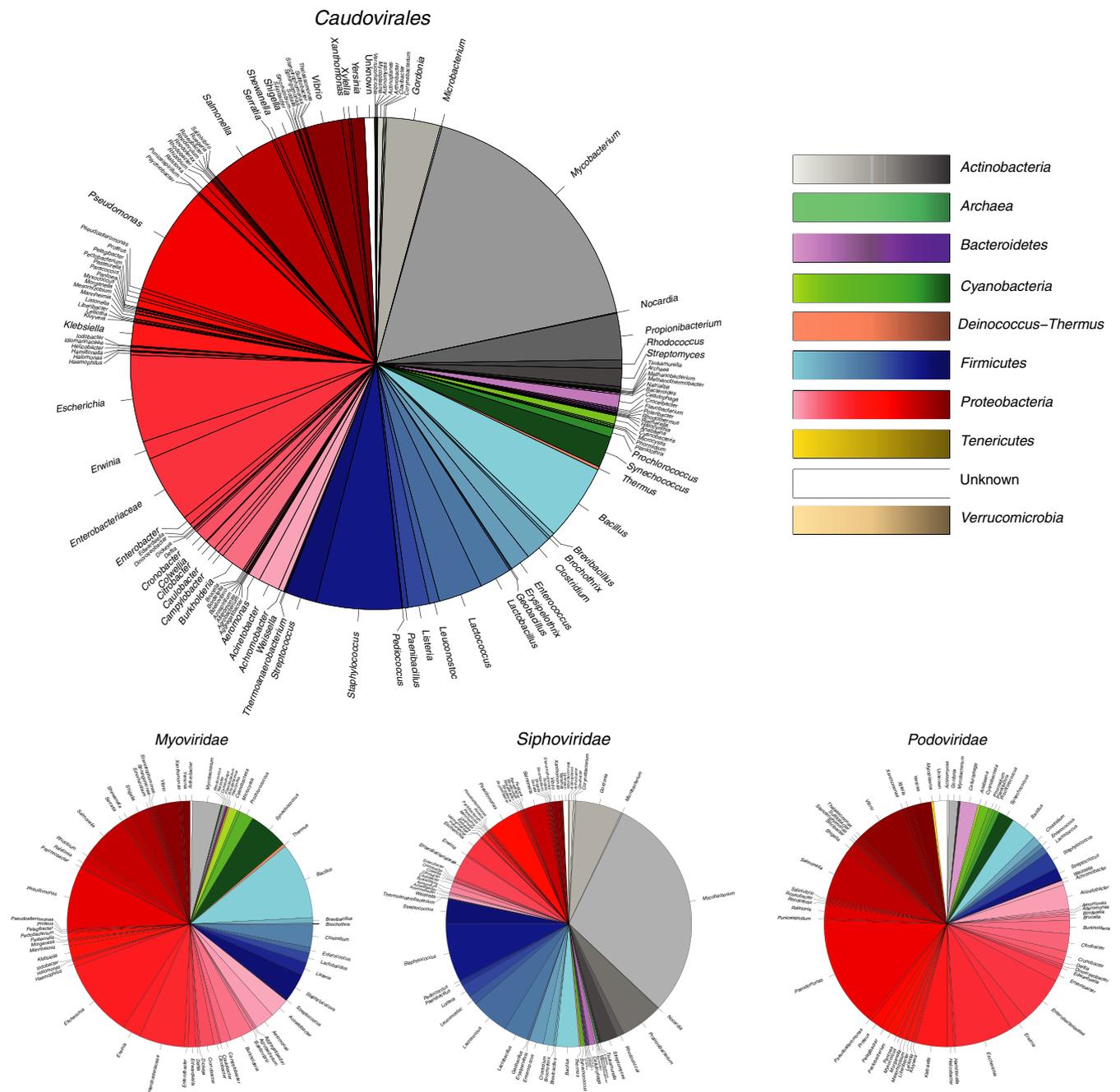


FIGURE 2 Breakdown of complete genomes of members of the order *Caudovirales* (dsDNA viruses, no RNA stage) available from NCBI Genome on November 21, 2016 ($n = 1943$). The number of publicly available *Caudovirales* genomes had only increased to 2044 by July 6, 2017. Genera of prokaryotes (with some families and higher taxa) infected by phages are shown around the chart; colours correspond to phyla (kingdom in the case of the *Archaea*) of prokaryotes infected by different phages. Aggregated data are shown for each of the three families of the order *Caudovirales* in the large plot. *Myoviridae*, $n = 538$ complete genomes; *Siphoviridae*, $n = 1063$ complete genomes; *Podoviridae*, $n = 342$ complete genomes. Larger versions of the pie charts shown in this figure are available in Supporting Information

majority of recognisable viruses, with 59% of the sequences not related to anything within public databases. Sequences with significant matches to myophages, podophages and microphages were also detected. The most common sequence matches were with phages associated with *Listeria monocytogenes*, *Burkholderia thailandensis* and *Lactococcus lactis*. Prophages within bacterial genomes also made a large contribution to the recognisable viruses identified.¹⁹ Numerous

sequences with similarity to phages that infect Gram-positive bacteria were also detected, reflecting results from studies on faecal bacterial populations that have shown *Firmicutes* and, to a lesser extent, *Actinobacteria* are among the most predominant bacteria in human faeces. This suggested the most abundant phage groups in a given environment reflect the kinds of microbes found in that environment.¹

TABLE 1 Phages recently isolated from the human gut microbiota

Host bacterium	Source of host bacterium	Phage(s); source	Lytic/lysogenic
<i>Klebsiella pneumoniae</i> L4-FAA5 ⁸²	Caecal effluent	KLPN1 ^a ; caecal effluent	Lytic
<i>Clostridium difficile</i> HMC114 ⁸³	Faeces	Phage tail-like particle; <i>Clostridium difficile</i> HMC114	Lysogenic
<i>Escherichia coli</i> DPC6051 ⁸⁴	Unknown	φAPCEc01, φAPCEc02, φAPCEc03; faeces	Lytic
<i>Proteus mirabilis</i> CEMTC 73 ⁸⁵	Faeces of a patient after bariatric surgery	PM16 ^a ; faeces	Lytic

^aPhage isolated from same sample as host strain.

5 | MINING TOTAL COMMUNITY METAGENOMES TO REVEAL VIRUS DIVERSITY

With the exceptions of the *Proteobacteria*, mycobacteria and some lactic acid bacteria, very few phages infecting prokaryotes have been isolated and sequenced (Figure 2). Between 4% and 22% of DNA in total community metagenomes can be derived from VLPs.^{20–22} This allows potential recovery of phage sequences not well represented in virome studies, though still excludes RNA viruses from detailed analyses.

Human gut-specific *Bacteroides* phage φB124-14 was isolated from municipal waters and infects *Bacteroides fragilis* GB-124 and a limited range of other *B. fragilis* strains.^{7,23} Sequences with homology to this phage were found in faecal metagenomes from MetaHIT (104/124) and Japan (3/13), and human gut viromes (2/16). Sequences associated with this phage were not found in American gut-associated datasets (metagenome or virome), suggesting geographical variation in distribution of gut-associated phages.⁷ Using a similar metagenomic mining approach but with major capsid protein VP1 specific for the ssDNA phage family *Microviridae*, members of the subfamilies *Gokushovirinae*, *Alpnavirinae* and *Pichovirinae* were detected in faecal metagenomes.²⁴

By using phage genome signature-based recovery with metagenomes, several (predominantly temperate) potential gut-specific phages infecting members of the order *Bacteroidales* were identified.²¹ This alignment-free approach was able to resolve phage sequences not readily detected by conventional alignment-driven approaches. The main limitation of this method is the lack of available genome sequences from isolated phages, as “driver sequences” are required to provide baseline information upon which inferences can be made regarding sequence data and host phylogeny (only four such “driver sequences” were required in the *Bacteroidales* study).²¹ Using co-occurrence profiling, crAssphage was identified in publicly available metagenomes, and by CRISPR analysis and co-occurrence profiling predicted to infect *Bacteroides* or *Prevotella* spp.²² This phage, with a 97,065 bp genome encoding 80 proteins, is predicted to make up to 24%–90% of all reads in faecal viromes, and 1.7% of reads in all human faecal metagenomes (up to 22% in the HMP dataset). Long-range PCR and Sanger sequencing were used to confirm the presence of the genome in a faecal sample. The role of this phage in the human gut is

unknown, and a representative of it remains to be isolated and purified.

Novel human-associated viruses present in the gut virome can also be detected using viral metagenomics. Analyses of diarrhoea samples from Dutch patients with unexplained acute gastroenteritis led to the identification of novel members of the families *Anelloviridae*, *Picobirnaviridae*, *Herpesviridae* and *Picornaviridae*.²⁵ The role of these viruses in gastroenteritis is still to be established. It is also important to remember in exploratory (and other) studies that detecting virus nucleic acids in samples gives no indication of whether the viruses infect human or other cells.²⁶

6 | INDIVIDUAL INTESTINAL VIROMES ARE UNIQUE, BUT SOME PHAGES ARE SHARED AMONG INDIVIDUALS

A longitudinal study (three time points over 1 year) of four pairs of adult monozygotic twins and their mothers demonstrated that, while the faecal prokaryote communities of related individuals are very similar, each individual harbours a unique faecal virome.²⁷ Intrapersonal diversity of the virome was low, with >95% of viruses retained over a 1-year period, with a few temperate phages dominating the viromes. This contrasts with Lotka-Volterra predator-prey relationships, in which cyclic changes in phage and prokaryote abundance would be expected, or with kill-the-winner dynamics, in which detectable changes in phage-host pairs would be detected. The majority (81%) of sequenced reads did not match any known viruses, and only 3.2 ± 2.8% reads could be assigned functions using KEGG. The predicted hosts of the identifiable phages were members of the phyla *Bacteroidetes* and *Firmicutes*, in accordance with the predominant bacteria detected in the samples by 16S rRNA gene sequencing. The ratio of virotypes to bacterial phylotypes was predicted to be ~1:1.

Assembly of sequence data from four deep-sequenced DNA viromes (8.1 Gb sequence data in total) from two healthy individuals led to recovery of 4301 contigs, representing 72 complete phage genomes plus potentially complete and partial phage genomes.²⁸ A minority (1703) of the contigs and 63/72 complete genomes contained at least one ORF with homology to a phage protein. Only 314 of all contigs contained homologues that allowed their assignment to a higher phage taxon, with order *Caudovirales*

TABLE 2 Methods used for recovering VLPs from intestinal/faecal contents

Sample(s)	Comments
Caecal and faecal ¹⁰	Filtration and polyethylene glycol (PEG) precipitation to concentrate VLPs, with and without CsCl centrifugation 0.45 µm filters instead of 0.22 µm doubled the number of VLPs and the quantity of DNA recovered from samples Used pulsed-field gel electrophoresis to demonstrate differences in viromes among individuals
Faecal ⁸⁶	ViSeq (VIDISCA enrichment of VLPs/conversion of RNA to cDNA/SLIM) Documented HIV-1 and overall viral content of faecal samples of late-stage AIDS patients
Faecal ¹²	Microfiltration-free PEG and TFF methods optimised Spiked samples with low numbers of c2 (<i>Siphoviridae</i>), φ29 (<i>Podoviridae</i>), and T4 (<i>Myoviridae</i>) phages to determine phage recovery PEG more effective than TFF, yielding 16 times more phage particles and 68 times more phage DNA per volume Increased recovery of <i>Caudovirales</i> at the expense of eukaryotic viruses
Mock virome containing 9 phages/viruses ⁶⁶	NetoVIR (novel enrichment technique of viromes) Homogenisation with glass beads led to destruction of viruses Increasing centrifugation times at 17,000 g led to reductions in virus recovery, possibly due to presence of aggregates in the sample Filtering through 0.2, 0.45 or 0.8 µm filters led to significant losses of mimivirus; 0.2 µm filter led to significant loss of Herpesvirus Chloroform treatment led to losses of enveloped viruses (coronavirus and mimivirus) and nonenveloped viruses (rotavirus and polyomavirus) Demonstrated effect of increasing number of PCR cycles on recovery of viruses in the mock community (ie, lower yields of some viruses as number of cycles increased) No data presented on NetoVir when used on gut/faecal samples
Faecal ⁸⁷	Filtered (0.2 µm) sample subjected to fluorescence-activated cell sorting (FACS) Only looked at a small fraction of the faecal virome, but different viral fractions could theoretically be separated using FACS Allowed direct sequencing of viral DNA without the need for whole-genome amplification Avoided contamination with bacteria Led to improved assembly of viral genomes
Artificial community of 25 viruses ⁸⁸	Mock community contained dsDNA (<i>Adenoviridae</i> , <i>Herpesviridae</i>), dsRNA (<i>Reoviridae</i>), and ssRNA (<i>Astroviridae</i> , <i>Caliciviridae</i> , <i>Coronaviridae</i> , <i>Picornaviridae</i> , <i>Orthomyxoviridae</i> , <i>Paramyxoviridae</i>) viruses Filtration and nuclease digestion had little effects on overall results Pre-amplification of nucleic acids prior to sequencing led to more biased genome coverage with near-duplicate reads accumulating over the same region, decreasing genome coverage ScriptSeq > TruSeq > Nextera for virus detection and genome coverage Maximum of 22/25 viruses recovered in sequenced viromes
Artificial intestinal microbiota (six phages, two bacteria) ⁶⁷	CsCl centrifugation most efficient for removing host DNA, but discriminated against some phages and was not reproducible; not suitable for quantitative studies Did not process PEG-precipitated samples for sequencing Recommended using filtration + DNase, or dithiothreitol treatment + filtration + DNase to collect VLPs, but did not test on human faeces Cautioned about over-interpreting data: abundances of individual phages within samples deviated by more than one order of magnitude from the actual input number of phage particles

predominating. Using a low-stringency approach (ie, recruitment of a single sequence read to any of the 4301 contigs) to detect genomes in samples, 17% of the contigs were found to be present in both individuals in at least one sample and 7% were present in all four samples. Sequence reads from ~20% of 62 faecal viromes (generated from healthy cohorts in three American cities and Cambridge, UK²⁹) mapped to 160 contigs. A consortium of 155 phages contributed to the “healthy gut phageome” (HGP), though only 23 phages were present in >50% of samples (“core” phages, including crAssphage and nine other complete genomes), and 132 were present in 20%-50% of samples (“common” phages); 1679 were present in 2%-19% (“low overlap” phages). The HGP represented only 4% of the total phage community, with the expectation that the HGP will increase in size by analysing additional healthy individuals using deep sequencing.²⁸

Although it has been proposed the HGP may play a role in maintaining and possibly restoring a dysbiotic microbiota,³⁰ how this might be achieved, and the role and mechanisms by which shared communities of phages contribute to gut health, remains a matter of debate.

7 | INTESTINAL VIROME VARIATION AND STABILITY

Interindividual variation among and intra-personal stability of faecal viromes was confirmed in six healthy individuals enrolled in a dietary intervention study (two high-fat/low-fibre diet; three low-fat/high-fibre diet, one ad-lib diet).²⁰ Greater sequencing depth was achieved compared

with earlier studies (336 Mb vs 70 Mb across all samples,^{19,27} with 98% of sequence reads having no known hits in databases. Greater sequencing depth allowed assembly of contigs (7175 of at least 500 bp in length), with 87% of reads contributing to these. PHACCS (PHAge Communities from Contig Spectrum)-predicted median species richness of the samples was 44 (range 19–785), in agreement with Reyes et al²⁷ (median of 35, range 10–984). Tailed phages predominated (*Siphoviridae*, 18% of the total; *Myoviridae*, 10%; *Podoviridae*, 4.8%; *Microviridae*, 0.9%; other families, 0.4%). The majority of contigs (55%) shared no significant amino acid homology with any phage family, while 11% had homology to multiple phage families. No eukaryotic viruses were detected. Minor convergence of viromes was seen when individuals were fed the same diet, though each individual retained a unique virome. Induction of lysis may have contributed to changes in phage abundance during the dietary intervention, though this has not been demonstrated empirically. In contrast to total faecal metagenomes, which were enriched in genes for synthesis (amino acid, carbohydrate precursors), translational machinery and cell component biogenesis, VLPs were enriched for genes associated with replication, recombination, repair and unknown functions; consistent with viruses recruiting host-cell machinery for their propagation. Findings regarding prophages were not in accord with previous studies (they did far more in-depth analyses than previous studies): 2814 contigs had significant amino acid homology with an ACLAME prophage sequence, 41% were assigned to *Firmicutes*, 0.9% were assigned to *Bacteroidetes*, 16% were assigned to the *Proteobacteria*.²⁰ This pattern did not match abundance of bacteria within samples, leading to the suggestion prevalence of temperate phages in faeces may differ among different phyla with the caveat that conclusions were dependent on phage sequences available in public databases at the time of the study. Although it is acknowledged overall individual faecal viromes are dissimilar, analyses of combined virome^{20,27} and total faecal metagenome data suggest the existence of a nontrivial common reservoir of phages shared among individuals in different geographical regions.⁸

Because of the small size of virus genomes compared with prokaryote genomes, it is relatively easy to achieve greater sequencing depth than total metagenome studies. Deep sequencing of faecal viromes (48 Gb of virome DNA across 12 individuals at three different time points over a 2-month period; ~1.3 Gb of sequence per virome) allowed empirical assessment of hypervariation driven by a unique reverse-transcriptase (RT)-based mechanism in faecal viromes.³¹ Between 573 and 3390 contigs longer than 1 kb were assembled from the donors' viromes. The majority of predicted genes (72%) within these contigs encoded unknown proteins, but as with previous studies the majority of sequences were associated with DNA phages. Only one donor harboured a eukaryotic DNA virus—*Human papilloma virus* type 6b. This is typical of faecal viromes of healthy humans, in which recovery of DNA eukaryotic viruses is rare.^{19,20,27,32,33} “Diversity-generating retroelements” (DGRs), RT-based systems that introduce mutations at adenines in specific repeated sequences,³³ were identified in the viromes of 11/12 subjects studied. Six predicted hypervariable genes encoded Ig-superfamily β -sandwich domains that have been associated with

viral evasion of the immune system³¹ and binding of phages to host mucin glycoproteins.³⁴ Phage enrichment within the mucosa is thought to protect the host from bacterial infections by limiting the presence of mucosal bacteria (“bacteriophage adherence to mucus” model of protection), providing nonhost-derived immunity operating throughout the GI tract.³⁴

8 | EVOLUTION OF THE INTESTINAL VIROME, AND THE INFLUENCE OF ANTIBIOTICS

In-depth analysis of one healthy adult's faecal virome over 2.5 years (56 Gb of DNA sequenced over 24 samples taken at 16 time points) demonstrated its rapid evolution.³³ The majority (80%) of 478 assembled contigs persisted over the duration of the study, and accumulated sequence variation over time. The stability of phage nucleotide sequences was family-dependent. Temperate phages had the lowest substitution rate, whereas *Microviridae* had high substitution rates ($>10^{-5}$ per nt per day), with identification of new viral species over the study period; numerous new viral species may emerge in the human gut over the course of a lifetime. Evolution of DGRs and CRISPRs in phage contigs was also observed. Therefore, multiple mechanisms contribute to phage sequence variation within the human faecal virome. Antibiotic therapy influences viromes, not by affecting VLPs directly but by affecting their prokaryote hosts.³⁵ For example, quinolones have been shown to induce Shiga-toxin-encoding prophages by *Escherichia coli* in mice, leading to toxin production and death of animals.³⁶ In a longitudinal study involving four individuals on various broad-spectrum, long-term intravenous antibiotic therapies for *Staphylococcus* infections, oral viromes were affected more than faecal viromes by antibiotic treatment, with an increase in representation of papillomaviruses in the oral virome of antibiotic-treated patients.³⁵ A nonsignificant increase in antibiotic-resistance genes was seen in the faecal viromes of the patients. It was suggested sampling of individuals should be done at more regular intervals to get a true picture of the effect of antibiotics on viromes. It should also be noted the study was hampered by confounders relating to age of patients compared with controls and the patients being on different treatment regimes.

Willner et al³⁷ first suggested the sharing of VLPs among individuals in their study of sputum samples from individuals with lung disease (Table S1). Later work on oral and faecal viromes of 20 individuals over a 6-month period confirmed this, demonstrating phages were persistent in these viromes and readily shared among related and unrelated members of the same household.³⁸ Sharing of phages encoding virulence or antibiotic-resistance genes has implications for shaping of microbiomes of those in close contact with one another, and warrants further study. With respect to antibiotic-resistance genes, their presence may have been over-estimated in human viromes^{11,20,27,31,35,37} as these entities are thought to be rarely encoded in phages.³⁹ Consequently, care should be taken when analysing data: proper *in silico* checks with conservative cut-offs are

required to refine functional assignments and avoid over-interpretation of data.³⁹ Induction of prophages encoded in prokaryote genomes during antibiotic treatment appears to be more relevant to gut microbial ecology—with an increased level of generalised transduction and low level of transfer of antibiotic-resistance genes.³⁹

9 | ssDNA VIRUSES IN THE INTESTINAL VIROME

To date, only one study has specifically examined the contribution of ssDNA viruses to the human faecal virome.¹¹ *Microviridae* were found to represent the majority (27%-49%) of known phage sequences in samples from five healthy South Koreans. This overestimation of *Microviridae* abundance was due to preferential amplification of small circular genomes by the phi29 polymerase used to amplify the ssDNA. Corrected abundances suggested the *Microviridae* represent 3%-9% of known viruses in human faeces. Sequence analyses of a major capsid protein of the *Microviridae* and gut bacteria predicted *Prevotella* and *Bacteroides* to be the hosts of a number of these viruses, sharing high sequence similarity with prophage-like sequences encoded by these bacteria. Newer protocols for amplification and sequencing of ssDNA from environmental samples will improve characterisation of this group of viruses within the human gut.⁴⁰⁻⁴²

10 | RNA PHAGES IN FAECES OF HEALTHY ADULTS

The GI tract of mammals is one of the natural habitats of RNA phages, with special relationships appearing to exist between specific RNA phage groups and their hosts.⁴³ Coliphage RNA viral assemblages in the faeces of cows, foxes, elephants, pigs, horses, birds and humans and found group II and III RNA phages were predominant in human faeces, while group I and II RNA phages were predominant in the faeces of other animals. Of the 597 human faecal samples examined, 76.5% had fewer than 10 pfu/mL sample and 14.6% had between 10 and 100 pfu/mL sample (20%, w/v, faecal homogenates).

Metagenomics was used to examine RNA viral assemblages in faecal samples obtained from two healthy adults: two samples were collected from one donor 6 months apart, with one sample obtained from the other donor.⁴⁴ Three shotgun viral cDNA libraries were generated, with a total of 36,769 viral sequences obtained; 91.5% of these had significant similarity to sequences in the GenBank database, with three-quarters being most similar to viruses. Most (25,040/25,779 known) sequences were similar to plant-pathogenic RNA viruses. Forty-two viral species (2 animal, 35 plant, 1 yeast and 4 phage) were identified in the faecal RNA viral libraries, with *Pepper mild mottle virus* (PMMV), a well-characterised plant-pathogenic virus that infects all species of the genus *Capsicum*, being the most abundant. Using RT-PCR, PMMV was detected in food and faecal (10^6 -

10^9 virions/g dry weight faeces) samples from three different individuals. Prevalence of PMMV in the human population was also demonstrated, being detected in faecal samples from North American and Singaporean (both 6/9 positive) individuals. In addition to PMMV, 24 of the 35 plant-pathogenic viruses were associated with consumable products such as cereals, fruits, tobacco and vegetables, suggesting a large proportion of faecal RNA viruses come from exogenous sources. Support for this hypothesis came from the detection of PMMV in some processed pepper-containing foods. The amount of virus detected in faeces exceeded that detected in the ingested foodstuffs, suggesting plant RNA viruses replicate in the human GI tract; however, no data were presented to support this hypothesis. Faeces-derived PMMV could infect wax pepper plants, suggesting humans and other animals could act as transmission vectors for plant viruses through their digestive tracts. A second study⁴⁵ confirmed the presence of PMMV, and other plant RNA viruses, in human faeces. Studies into the ability of plant viruses to replicate in the human gut (in epithelial and microbial cells) should be initiated.^{44,45}

The low numbers of phage sequences retrieved (corresponding to *Haemophilus influenza* phage HP2, Roseophage SIO1, phage L cro and *Shigella flexneri* phage V⁴⁴ and enterobacteria phage phiK and phage phiV10⁴⁵) reflected earlier findings⁴³ in relation to the low numbers of RNA phages recovered from human faeces.

11 | ESTABLISHMENT OF THE FAECAL VIROME

Microscopic examination of the meconium (ie, first faecal sample) of a newborn infant did not detect any VLPs.³² By the end of the first week of life, $\sim 10^8$ VLPs/g wet weight faeces could be detected, with this number seen as an underestimation of the true number due to high background fluorescence and clumping of the sample. A viral metagenome was generated from the faeces of the 1-week-old infant. Similar to adult DNA viral metagenomes, siphophage and prophage sequences predominated. Eight viral genotypes were predicted to be present in the infant's faecal virome, with the most abundant viral genotype contributing 43.6% to the total. The low diversity of the infant faecal virome correlated well with the known low diversity of the faecal bacterial assemblage in the first week of life. Sequences with significant similarity to podophages and myophages were also detected. More than 50% of the significant hits were to phages that infect Gram-positive bacteria (*Lactobacillus*, *Enterococcus*, *Bacillus*, *Lactococcus*, *Streptococcus* and *Staphylococcus*), with hits to phages that infect Gram-negative bacteria (*Bacteroides* and *Listonella*) also identified. Microarray analyses of viromes obtained from faecal samples at weeks 1 and 2 of life with probes constructed from the week 1 sequencing library showed the temporal nature of the infant gut virome, with >50% sequences that were at high abundance at week 1 absent at week 2, and the most abundant sequences at week 2 being of medium abundance at week 1. No sequences were present at high levels at week 1 or 2,

demonstrating turnover of the most abundant viral sequences over a short period of time.

Primers designed based on the sequences of the three largest contigs generated from sequencing library amplified sequences corresponding to a *Streptococcus pyogenes* phage terminase (large subunit), a *Bacteroides thetaiotaomicron* integrase and *Bacillus thuringiensis* phage MZTP02. A portion of the sequence amplified by the *B. thetaiotaomicron* integrase primer set was identical to a viral sequence retrieved in the first adult faecal virome study.³² VLP DNA was isolated from the breast and formula milk fed to the infant, faecal samples obtained at 1, 2, 5 and 14 weeks of age, and from the faeces (two samples taken a year apart) of an unrelated adult. All three VLP sequences were present in the first 3 months of the infant's life and did not originate from breast or formula milk. The *B. thetaiotaomicron* integrase was detected in all faecal (infant and adult) samples examined, while the other two VLP sequences were detected only in the infant faecal samples. It is possible the assay used in this study was not sensitive enough to detect VLP sequences in milk samples, as one would assume that populations would be in very low numbers in these substrates. This point seems especially relevant given >25% of the sequences identified in the faecal viral assemblage were most similar to those of phages infecting lactic acid bacteria known to be abundant in breast milk. It is not known from where the initial infant gut virome originates, although it has been suggested dietary, environmental and maternal sources may be sources of inocula.³² In addition, primary infant faecal viruses may originate from induction of prophages from primary bacterial colonisers.³²

A more-extensive metagenomic study examining the infant (DNA and RNA) virome in four pairs of twins over the first 24 months of life (sampling at 0, 3, 6, 9, 12, 18 and 24 months) identified eukaryotic DNA viruses from month 3 onwards (ssDNA—anelloviruses, circoviruses, geminiviruses, nanoviruses, parvoviruses; dsDNA—adenoviruses, polyomaviruses), but only anelloviruses, circoviruses and geminiviruses were detected in at least one twin of each pair between months 3 and 24 of the study.⁴⁶ Anelloviruses, associated with host immune status, were most prevalent from 3 months of age and highly divergent from known anelloviruses, with their abundance peaking at 6–12 months of age. The same anelloviruses were detected in faecal samples of the same infant collected 12 months apart, suggesting a persistent or stable source of recurring infection.⁴⁶ It was speculated the increase in anelloviruses between 6 and 12 months was the result of lowered immune status associated with the decrease in maternal IgG, though no data were provided to support this. Eukaryotic viral (DNA) richness was low at 3 months but increased thereafter, indicating environmental exposure contributes to the establishment of this component of the virome. Eukaryotic RNA viruses (ssRNA—alphaflexiviruses, astroviruses, calciviruses, picornaviruses, tombamoviruses, virgaviruses; dsRNA—chrysoviruses, picobirnaviruses) were detected less frequently across the study period, but related twins harboured the same strains of parechovirus and enterovirus. Phages were detected at high abundance in all infants across the study period, with the majority of reads belonging

to the order *Caudovirales* (*Siphoviridae* > *Myoviridae* > *Podoviridae*), *Inoviridae* (ssDNA), *Microviridae* (ssDNA), *Corticoviridae* and *Tectiviridae*. No RNA phages were detected in any of the faecal samples. Phage richness and diversity were greatest at month 0 and decreased as the infants aged, and the relative abundance of *Caudovirales* (high at 0 months, low at 24 months) was inversely correlated with the abundance of *Microviridae*. crAssphage²² was detected in one sample of one infant, suggesting this phage is not acquired early in life. Of note, reads associated with the family *Lipothrixviridae*, an archaeal virus family, were detected in the faecal virome of all twins at various points across the study period. Each infant had a unique faecal virome, though the viromes of related infants were more similar to one another than nonrelated infants after controlling for age. Bacterial species richness and diversity increased with age, being predominated by *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, *Verucomicrobia*, *Proteobacteria* and *Fusobacteria* at 24 months. An inverse correlation of phage and bacterial richness (high phage–low bacteria to low phage–high bacteria) was seen from month 0 to 24 months. The infant faecal virome is suggested, therefore, to display a reversed predator–prey cycle: ie, limited prey diversity controls predator abundance.⁴⁶ The initially high phage population is unsustainable because of low numbers of bacteria colonising the GI tract. Consequently, the phage assemblage shrinks in size and diversity, relieving pressure on the bacterial community and allowing it to establish and colonise the gut.

12 | INFLUENCE OF ENVIRONMENTAL FACTORS ON DEVELOPMENT OF THE INFANT VIROME

Differences in birth mode (Caesarean vs vaginal delivery) and diet (breast- vs formula-feeding; age at which weaning began) were confounders in the study of Lim et al.⁴⁶ Differences in the faecal microbiota of breast- and formula-fed infants are well documented, as are the effects of weaning.⁴⁷ It is also known that 16S rRNA gene-based sequencing studies do not always give accurate representation of *Bifidobacterium* populations in the infant faecal microbiota.^{48,49} Therefore, shotgun metagenome studies examining the infant microbiome and virome are required in carefully controlled populations to study the acquisition and development of the infant GI virome.

The influence of malnutrition on development of the infant faecal virome was investigated during the first 30 months of life in eight pairs of healthy twins and 12 pairs of twins discordant for severe acute malnutrition (six pairs with one twin healthy, one with marasmus; six pairs with one twin healthy, one with kwashiorkor).⁵⁰ As with the study of Lim et al.⁴⁶ *Anelloviridae* were prevalent in early life, but diminished at 15–18 months of age. In addition, *Microviridae* increased in abundance with age. *Siphoviridae* were the most abundant VLPs during 0–10 months, but then slowly decreased. This demonstrated that, as with the prokaryotic component of the faecal microbiota, age affects virome composition in early life. Discordant

twins could be distinguished from healthy twins based on phage populations and the presence of higher levels of eukaryotic viruses of the families *Anelloviridae* and *Circoviridae*.

13 | THE INTESTINAL VIROME, IMMUNE RESPONSE, AND DISEASE

Alterations in the intestinal virome characterised by reduced diversity and altered intestinal barrier function have been associated with infectious and autoimmune diseases in addition to metabolic disorders and cardiovascular disease (Figure 1). A potential pathogenic mechanism involves translocation of viruses present in the gut microbiota across the intestinal epithelial barrier, particularly in situations of altered and increased permeability that exert immunomodulatory effects on the host.⁵¹

HIV-1 and rotavirus are known to escape the gut and disseminate to peripheral tissues via components of the immune system.⁵¹ Enterocytes are able to recognise specific phage peptides and transport them across the mucosal barrier.⁵¹ HIV patients are affected by translocation, and have associated systemic inflammation.⁵² Changes to the faecal virome and bacterial microbiome have been seen in HIV patients, with increased numbers of adenovirus-associated sequences and increased bacterial diversity and richness associated with peripheral CD4 T-cell counts of <200; phage populations are not detectably affected by HIV disease status or anti-retroviral therapies,⁵² but plasma samples from HIV patients contain gut microbiota-derived nucleic acids.⁵³ Phages have been detected in blood samples of healthy humans, while mycobacteriophages have been found in the blood of patients with sarcoidosis, and Crohn's disease (CD), and in healthy individuals.^{51,54} Circulating phages have also been detected in blood samples after oral administration of antibiotics to patients with bacterial infections, whereas no phages could be detected in their blood prior to antibiotic therapy.⁵¹ It is not known if these phages were of GI origin. The presence of phages in the peripheral blood has been termed "phagemia".⁵¹ However, whether the blood contains its own "baseline" VLP population to which the GI virome can contribute in disease is unknown.

Human-associated viruses and intact phages and/or their nucleic acids are immunogenic. Newborns and children vaccinated with phage (ϕ X147) formed antibodies against the phage within 1 week of exposure.⁵⁵ Tumour-specific phages can be used to treat disease by initiating infiltration of neutrophilic granulocytes in a toll-like-receptor (TLR)-dependent manner.⁵⁶ Mono-association of germ-free mice with *Murine norovirus* (positive strand RNA virus) was sufficient to replicate the beneficial effects of commensal bacteria, restoring small-intestinal morphology and lymphocyte function without overt inflammation or disease.⁵⁷ Presence of innate lymphoid cells was reduced in the virus-treated mice compared with controls. Inoculation of antibiotic-treated, wild-type mice with *Murine norovirus* for 10 days increased villi width, Paneth cell granules, intestinal T cells and IFN- γ expression,

producing comparable but nonidentical results to when antibiotic-treated mice were inoculated with strains of commensal bacteria. Intestinal gene expression was affected by *Murine norovirus* inoculation, influencing lymphoid cell development and immune responses mostly related to the anti-viral type I interferon (IFN-I) response; IFN-I signalling was necessary for virus-mediated improvements in the *Murine norovirus*-infected mice. Genes associated with development and function of haematopoietic cells were also affected. Inoculation of antibiotic-treated mice with *Murine norovirus* protected the mice from *Citrobacter rodentium* superinfection. The protective effects of the virome may explain recent findings with respect to treatment of patients with *Clostridium difficile* infections with sterile faecal filtrates: based on the published protocol, the faecal filtrates used in the study are likely to have contained phages in addition to the "bacterial debris, proteins, antimicrobial compounds, metabolic products, and oligonucleotides/DNA" reported.⁵⁸ Support for a role for phages in the success of faecal microbial transplants (FMTs) to treat *C. difficile* infections comes from a pilot study by Zuo et al⁵⁹: when donors had higher diversity (richness) of *Caudovirales* than recipients, FMT was successful; in instances where recipient virome richness was greater than that of the donor, over half of the recipients experienced disease recurrence after FMT. Large cohort studies that characterise donor and recipient viromes are required to validate these findings.⁵⁹

14 | THE INTESTINAL VIROME AND INFLAMMATORY BOWEL DISEASE

The most compelling and direct evidence for enteric viruses playing a causal role in chronic GI inflammation, however, comes from studies in mice.⁶⁰ Mice with a mutation in the autophagy gene *Atg16L1*, which in humans predisposes to CD, develop normally and are symptom-free. However, enteric infection by norovirus resulted in the manifestation of the disease, although solely having the virus or the susceptible allele alone did not produce the disease, demonstrating the importance of virus-susceptibility gene interactions in disease progression. Using a dextran sulphate sodium (DSS)-induced rodent model of colitis, importance of the gut virome to GI immune homeostasis was demonstrated.⁶¹ In mice administered a cocktail of anti-viral drugs followed by DSS, the animals treated with anti-virals had much more severe colitis than the animals with an intact virome, along with greater weight loss and reduction in colon length. Protective effects of the virome were mediated by stimulation of both TLR3 and TLR7, but not individually. Administration of inactivated rotavirus partially restored the virome-associated protective effect to the anti-viral-treated mice. TLR3/TLR7-deficient mice were more sensitive to colitis than their wild-type counterparts. A link to human inflammatory bowel disease (IBD) is the observation that patients carrying mutations in both TLR3 and TLR7 had higher rates of hospitalisation compared with IBD patients without mutations.⁶¹

15 | THE INTESTINAL VIROME IN HUMAN IBD

A link between the gut virome and CD was first suggested when it was found CD patients ($n = 19$) had significantly more VLPs (phages) in colonic biopsies than healthy individuals ($n = 10$) (2.9×10^9 vs 1.2×10^8 VLPs/biopsy), and for the CD patients ulcerated mucosa had significantly fewer VLPs than nonulcerated mucosa (2.1×10^9 vs 4.1×10^9 VLPs/biopsy).⁹ More recent, small-scale studies have been undertaken to characterise the gut virome of IBD.^{29,62–64} A pilot metagenomic study of ileal and colonic contents of six CD patients and ileal samples from six non-IBD controls showed CD DNA-based viromes had higher phage abundance than control samples.⁶² However, the ways in which virome data from healthy controls ($n = 8$) and patients with ileocolic CD ($n = 11$) were analysed were found to contribute to interpretation of results.⁶³ Prophages represented the most hits in the CD samples when unassembled reads were compared, but were higher in the control group when the assembled data were compared. Analysis based on assembled data showed fewer differences between the CD and control group. Nonrarefied data were used to generate estimators of species richness: the viromes of CD patients were less diverse than the healthy patients, but showed greater heterogeneity across samples. The CD and control samples could not be fully separated into two groups based on VLP composition and abundance.

A later DNA-based study showed differences in CD patients related to disease status (newly diagnosed, active onset, active pre-surgery) and therapy.⁶⁵ Similar to their previous study, individual variability and sample origin had a greater effect on the virome than CD, although more over-represented viruses were found in the CD viromes than the healthy, and newly diagnosed patients had higher diversity in faecal and biopsy viromes than those with active disease. Patients on steroids and/or immunosuppressors had lower diversity than untreated patients, while those on immunosuppressors only had lower diversity than those on combination therapy or steroids only. The study benefitted from recruiting CD and ulcerative colitis (UC) patients and their family members in the UK (Cambridge) and USA (Los Angeles), thereby controlling for household factors that may influence the microbiome.²⁹ An increase in phage-associated richness (predominantly *Caudovirales* and *Microviridae*) was seen in IBD faecal viromes compared with those of controls. Initial findings were confirmed using two independent and geographically distinct US (Boston, Chicago) patient cohorts with matched controls.²⁹ It was suggested decreases in bacterial richness concomitant with increases in phage richness may be due to predator-prey dynamics, but it was unclear how these contribute to the pathogenesis of IBD. Also proposed was the virome as a target for therapeutic modulation.²⁹ Although sharing some VLPs, CD and UC viromes could be separated from one another and harboured unique phages. Phages of *Lactococcus*, *Lactobacillus*, *Clostridium*, *Enterococcus* and *Streptococcus* were significantly associated with disease. Similar to other studies,^{27,38} there was also conservation of VLPs among members of the same household. Manrique et al²⁸ found the HGP was perturbed in

the UC ($n = 66$) and CD ($n = 36$) viromes of Norman et al²⁹: healthy subjects harboured on average 62% of the 23 core phages, while this number was significantly reduced (37% and 30%, respectively) in UC and CD patients.

Examination of RNA viruses present in colonic tissue from 10 IBD (6 CD, 4 UC) and five healthy individuals demonstrated IBD patients carried more mammalian viruses and human pathogens than healthy individuals; *Adenoviridae* were present in the IBD samples but absent from the healthy controls.⁶⁴ HERV (human endogenous retroviruses) were found in the IBD viromes; expression of HERV proteins has been linked with inflammatory conditions. HERV was present at significantly higher levels in IBD patients co-infected with herpesviruses. The significance of this association with respect to IBD is unknown, but infection with other viruses may trigger expression of HERV.⁶⁴

16 | CHALLENGES FACING STUDIES OF THE HUMAN INTESTINAL VIROME

Metagenomic studies of the virome are not without their limitations or challenges. Many of these are related to isolation and processing methods. The majority of studies published to date rely on the use of 0.2 μm filters, but it is known these halve the number of viruses detected compared with use of 0.45 μm filters and will not detect giant viruses.^{10,66} Consequently, we do not have a true picture of the diversity of the human virome.

16.1 | Sequencing depth

Characterisation of total faecal metagenomes has, to date, relied upon generating ~5 Gb of sequence data per sample to allow full characterisation of microbiomes.^{15–18} With the exception of a recent study that generated 5–8 Gb sequence data per sample,⁵⁹ the most extensive surveys of virome diversity have sequenced only a fifth of this depth per sample, with many relying on much lower sequencing depths for characterisation of viromes. With the exception of the four viromes generated by Manrique et al²⁸ all viromes generated to date have relied upon PCR amplification or multiple displacement amplification (MDA) to generate sufficient DNA for sequencing. These methods introduce bias into sequencing libraries. For example, when using the WTA2 kit it is recommended to use the Nextera XT library preparation kit to avoid cluster-calling problems during Illumina sequencing. However, the Nextera kit causes under-representation of ends of viral genomes because of the way this kit works.⁶⁶ CsCl centrifugation has been shown to be the most efficient method for removing host DNA, but it discriminates against some phages and is not reproducible; consequently, this method is not suitable for quantitative studies.⁶⁷

16.2 | Phage isolation

Methods of recovery of faecal phages have improved recently (Table 2), so it should be feasible to generate PCR- or MDA-free,

deep-sequenced viromes to allow greater characterisation of the human gut virome. The question will then become how many sequence data are required to adequately characterise the virome. Current sequencing depths and paucity of available virus genome sequences (Figure 2) may mean phages that infect different bacterial hosts may be indistinguishable.²⁰ Low sequencing depth also means low coverage of VLP genomes within samples, making it difficult to differentiate between spurious alignments and real virus sequences and potentially leading to overestimation of VLPs in samples when relying on automated read alignments.⁶⁸ Bioinformatics tools are required that allow differentiation of authentic sequences from artefacts.

16.3 | Isolation-kit contaminants

Few studies of the human virome have considered the effects of contaminants in reagents and nucleic acid extraction kits on viral metagenomes (Table S1), though these are being increasingly recognised in prokaryote-focused studies.^{26,69,70} Nucleases, proteases and polymerases used in studies are produced in protein expression systems, and sequences associated with these expression vectors have been detected in virome studies.⁷⁰ Columns used to purify DNA

may introduce parvovirus-like, circoviruses/densoviruses and iridoviruses sequences into samples, and it has been suggested their use should be avoided in virome studies.^{26,70} Efforts should be made to prevent cross-contamination of samples when processing samples for nucleic acid extraction and sequencing.^{26,70} Low-biomass samples obtained from non-GI sites will be reliant on PCR-based approaches for the foreseeable future, and these are most likely to be affected by contaminants in reagents. Therefore, appropriate negative controls should be included in sequencing studies, to allow the identification of sample-associated sequences rather than those from contaminants.

16.4 | Combining virus isolation and metagenomics

Those studies targeting only the VLPs of the microbiota will not capture prophage diversity within samples, and will contain little information to allow studies of phage-host interactions or host ranges, limiting in-depth analyses of the overall ecology of the gut microbiome.²¹ Studying total community metagenomes in conjunction with VLP-derived metagenomes will provide a more-complete picture of virus-host interactions in the human gut.²¹ The greatest challenge to integrated studies of the whole gut microbiome is

TABLE 3 Available tools for analysis of viruses in total community and viral metagenomes

Name	Description	Website/reference(s)
PHACCS ⁸⁹	Assesses the biodiversity of viromes; requires knowledge of virus genome sizes; desktop app	https://sourceforge.net/projects/phaccs/
ACLAME ⁹⁰	Classification of mobile genetic elements; web-based	http://aclame.ulb.ac.be
MetaVir 2 ^{91,92}	Annotates viral metagenomes (raw reads or assembled contigs); web-based	http://metavir-meb.univ-bpclermont.fr
VIROME ⁹³	Classification of predicted open-reading frames (ORFs) from viral metagenomes; each ORF placed into one of 9 sequence categories; functional classification; web-based	http://virome.dbi.udel.edu
Vanator ⁹⁴	Perl pipeline using a number of alignment, assembly and analysis tools to assess metagenomic data derived from Illumina data; virus discovery; desktop app	https://sourceforge.net/projects/vanator-cvr/
VirSorter ⁹⁵	Detects virus signals in single-cell amplified genomes of uncultivated organisms or genomic fragments assembled from metagenomic sequencing; desktop app	https://github.com/simroux/VirSorter
ViromeScan ⁹⁶	Characterises taxonomy of eukaryotic viruses directly from raw reads; desktop app	https://sourceforge.net/projects/viromescan/
VIP ⁹⁷	Identification of eukaryotic viruses (pathogens, influenza virus) from metagenomes; desktop app	https://github.com/keylabivdc/VIP
VirusDetect ⁹⁸	Analyses small RNA datasets for known and novel viruses; web-based and desktop app	http://bioinfo.bti.cornell.edu/tool/VirusDetect/
PHASTER ⁹⁹	PHAge Search Tool—Enhanced Release, rapid identification and annotation of prophage sequences within bacterial genomes, plasmids and metagenomic sequences	http://phaster.ca/
VirusSeeker ¹⁰⁰	BLAST-based sequence analysis pipeline for eukaryotic and prokaryotic virus discovery and virome composition; desktop app	https://wupathlabs.wustl.edu/virusseeker/
VirHostMatcher ⁷⁴	Computes Oligonucleotide Frequency (ONF) scores between viral and host sequences in total community metagenomes, and visualises results; higher predictive accuracy at class and phylum levels	https://github.com/jessieren/VirHostMatcher
VirFinder ⁷³	k-mer frequency-based, machine-learning method for virus contig identification in total community metagenomes	https://github.com/jessieren/VirFinder

the lack of available phage sequences from isolated viruses (Figure 2). Because most nucleotide sequences obtained from viral metagenomes are novel, BLASTN analyses are generally uninformative.⁷¹ It has been stated that isolating “just a few phage genomes from novel environments will greatly increase our understanding of viral diversity in these environments”.⁷² Recently, OligoNucleotide Frequency profiling, in conjunction with abundance profiling and homology methods, has been proposed as a means of improving identification of microbial hosts for novel viruses from total community metagenomes^{73,74}; deriving full benefit from this approach relies on classical isolation and genomic characterisation of viruses and their hosts from environmental samples. Therefore, it is of utmost importance that isolation of phages from the GI tract and their genomic characterisation are at the forefront of future research: incorporation of ecological studies into those exploiting lytic phages or phage-encoded gene products for biotechnological or medical applications would be of great benefit to the research community.

16.5 | Environmental confounders

As with the prokaryote component of the microbiome, the genetic makeup of an individual's virome is influenced by diet, nutrition status, health, socioeconomic group, geographical location, age, lifestyle, season and medication.^{7,20,29,50,75} Immune status, and viral and human genetics will also affect interactions of the virome with the host.^{29,52,75} Virome studies have so far focused mainly on small human cohorts from restricted geographical locations, and to a large extent have focussed on characterising phage populations. As of July 2017, there are ~2040 complete genome sequences available for the *Caudovirales*, the main order of phages (Figure 2). Our coverage of eukaryote-associated infectious viruses is equally poor, mostly because we do not (1) know how to propagate known infectious agents and (2) there are a number of medical conditions speculated to be linked with viral infections (eg, type I diabetes, chronic fatigue syndrome, obesity) but for which no infectious agent has been found. It has been estimated over half of human-infecting viruses remain to be discovered.⁷⁶ Numerous conditions in which patients suffer unexplained symptoms may also be virus-associated. Therefore, there is a need to sample a wide range of body fluids and tissues from individuals globally to better understand the role of viruses in human health and disease.⁷⁵ Propagation and characterisation of prokaryote- and eukaryote-infecting viruses will require huge efforts to fully characterise extant viral populations, and there is still a wealth of viral diversity (pathogenic and nonpathogenic) to be discovered within human viromes. Those in areas with poor sanitation and crowded living conditions, travellers, children and the immunocompromised are expected to have higher viral loads, at least in terms of potentially pathogenic viruses,⁷⁵ while pre-term infants, breast- and formula-fed infants and the elderly are likely to have very different viromes compared with adults based on the general information we have about the composition of the human gut microbiome throughout life.^{27,50}

Interactions with pets, insects and wild animals will also influence the composition of the human virome. From a pathogen perspective, bush-workers, abattoir workers and individuals exposed to insect bites have been highlighted as likely sources of new human-infecting viruses.⁷⁵

With the advent of high-throughput, inexpensive, sensitive sequencing methods, it has become possible to study the contribution of dsDNA, ssDNA and RNA VLPs to the human virome. Gene transfer agents, which appear to represent defective phages, have not been examined in the context of the human microbiome or virome, but may also make a contribution to its genetic content.^{77,78} Similarly, exosomes—extracellular vesicles of 40–100 nm in diameter consisting of proteins, lipids, miRNA, mRNA and DNA—derived from cells lining the GI epithelium are likely to contribute to nucleic acids found in viromes.^{79,80} Outer-membrane vesicles, the equivalent (in size and composition) of eukaryotic exosomes for Gram-negative bacteria, may also contribute to what we currently call the virome, as these vesicles have recently been reported to contain DNA as well as RNA.⁸¹

16.6 | Virome-specific bioinformatics tools

Whereas 16S rRNA gene sequence data are readily processed and analysed using packages such as QIIME and Mothur, no such tools exist for the analyses of sequence data derived from host-associated viromes.³⁵ Tools exist for functional annotation of viromes and estimating viral diversity (Table 3). No easy-to-use pipeline that takes raw reads, strips out host DNA, looks for bacterial contaminants then assigns taxonomy and functionality to prokaryote and eukaryote viruses within samples exists, though efforts are being made to generate such tools. These will need to take into account the presence of human endogenous retroviruses, which comprise ~8% of the human genome and still possess the ability to encode retrovirus polymerase and envelope protein and can be reactivated by exogenous retroviruses such as HIV.⁵³

ACKNOWLEDGEMENTS

The authors would like to thank Sam Carding for creating Figure 1.

Declaration of personal interest: None.

AUTHORSHIP

Guarantor of the article: Simon R. Carding.

Author contributions: SRC takes responsibility for the integrity of the work as a whole, from inception to published article. LH and SRC designed the research study. LH and ND collected and analysed data. SRC and LH wrote the paper. All authors approve the final version of the manuscript.

ORCID

S. R. Carding  <http://orcid.org/0000-0002-2383-9701>

L. Hoyles  <http://orcid.org/0000-0002-6418-342X>

REFERENCES

- Breitbart M, Rohwer F. Here a virus, there a virus, everywhere the same virus? *Trends Microbiol.* 2005;13:278-284.
- Sullivan MB, Weitz JS, Wilhelm S. Viral ecology comes of age. *Environ Microbiol Rep.* 2017;9:33-35.
- Dalmasso M, Hill C, Ross RP. Exploiting gut bacteriophages for human health. *Trends Microbiol.* 2014;22:399-405.
- Keary R, Sanz-Gaitero M, van Raaij MJ, et al. Characterization of a bacteriophage-derived murein peptidase for elimination of antibiotic-resistant *Staphylococcus aureus*. *Curr Protein Pept Sci.* 2016;17:183-190.
- Weinbauer MG. Ecology of prokaryotic viruses. *FEMS Microbiol Rev.* 2004;28:127-181.
- Rakhuba DV, Kolomiets EI, Dey ES, Novik GI. Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell. *Pol J Microbiol.* 2010;59:145-155.
- Ogilvie LA, Caplin J, Dedi C, et al. Comparative (meta)genomic analysis and ecological profiling of human gut-specific bacteriophage ϕ B124-14. *PLoS ONE.* 2012;7:e35053.
- Stern A, Mick E, Tirosh I, Sagy O, Sorek R. CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome. *Genome Res.* 2012;22:1985-1994.
- Lepage P, Colombet J, Marteau P, Sime-Ngando T, Doré J, Leclerc M. Dysbiosis in inflammatory bowel disease: a role for bacteriophages? *Gut.* 2008;57:424-425.
- Hoyles L, McCartney AL, Neve H, et al. Characterization of virus-like particles associated with the human faecal and caecal microbiota. *Res Microbiol.* 2014;165:803-812.
- Kim M-S, Park E-J, Roh SW, Bae J-W. Diversity and abundance of single-stranded DNA viruses in human feces. *Appl Environ Microbiol.* 2011;77:8062-8070.
- Castro-Mejía JL, Muhammed MK, Kot W, et al. Optimizing protocols for extraction of bacteriophages prior to metagenomic analyses of phage communities in the human gut. *Microbiome.* 2015;3:64.
- Briese T, Kapoor A, Mishra N, et al. Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *mBio.* 2015;6:e01491-01415.
- Riesenfeld CS, Schloss PD, Handelsman J. Metagenomics: genomic analysis of microbial communities. *Annu Rev Genet.* 2004;38:525-552.
- Le Chatelier E, Nielsen T, Qin J, et al. Richness of human gut microbiome correlates with metabolic markers. *Nature.* 2013;500:541-546.
- Cotillard A, Kennedy SP, Kong LC, et al. Dietary intervention impact on gut microbial gene richness. *Nature.* 2013;500:585-588.
- Qin N, Yang F, Li A, et al. Alterations of the human gut microbiome in liver cirrhosis. *Nature.* 2014;513:59-64.
- Xie H, Guo R, Zhong H, et al. Shotgun metagenomics of 250 adult twins reveals genetic and environmental impacts on the gut microbiome. *Cell Syst.* 2016;3:572-584.e3.
- Breitbart M, Hewson I, Felts B, et al. Metagenomic analyses of an uncultured viral community from human feces. *J Bacteriol.* 2003;185:6220-6223.
- Minot S, Sinha R, Chen J, et al. The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res.* 2011;21:1616-1625.
- Ogilvie LA, Bowler LD, Caplin J, et al. Genome signature-based dissection of human gut metagenomes to extract subliminal viral sequences. *Nat Commun.* 2013;4:2420.
- Dutilh BE, Cassman N, McNair K, et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat Commun.* 2014;5:4498.
- Ebdon J, Muniesa M, Taylor H. The application of a recently isolated strain of *Bacteroides* (GB-124) to identify human sources of faecal pollution in a temperate river catchment. *Water Res.* 2007;41:3683-3690.
- Roux S, Krupovic M, Poulet A, Debroas D, Enault F. Evolution and diversity of the *Microviridae* viral family through a collection of 81 new complete genomes assembled from virome reads. *PLoS ONE.* 2012;7:e40418.
- Smits SL, Schapendonk CME, van Beek J, et al. New viruses in idiopathic human diarrhea cases, the Netherlands. *Emerg Infect Dis.* 2014;20:1218-1222.
- Sauvage V, Laperche S, Cheval J, et al. Viral metagenomics applied to blood donors and recipients at high risk for blood-borne infections. *Blood Transfus Trasfus Sanguie.* 2016;14:400-407.
- Reyes A, Haynes M, Hanson N, et al. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature.* 2010;466:334-338.
- Manrique P, Bolduc B, Walk ST, van der Oost J, de Vos WM, Young MJ. Healthy human gut phageome. *Proc Natl Acad Sci USA.* 2016;113:10400-10405.
- Norman JM, Handley SA, Baldrige MT, et al. Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell.* 2015;160:447-460.
- Manrique P, Dills M, Young MJ. The human gut phage community and its implications for health and disease. *Viruses.* 2017;9:141.
- Minot S, Grunberg S, Wu GD, Lewis JD, Bushman FD. Hypervariable loci in the human gut virome. *Proc Natl Acad Sci USA.* 2012;109:3962-3966.
- Breitbart M, Haynes M, Kelley S, et al. Viral diversity and dynamics in an infant gut. *Res Microbiol.* 2008;159:367-373.
- Minot S, Bryson A, Chehoud C, Wu GD, Lewis JD, Bushman FD. Rapid evolution of the human gut virome. *Proc Natl Acad Sci U S A.* 2013;110:12450-12455.
- Barr JJ, Auro R, Furlan M, et al. Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proc Natl Acad Sci U S A.* 2013;110:10771-10776.
- Abeles SR, Ly M, Santiago-Rodriguez TM, Pride DT. Effects of long term antibiotic therapy on human oral and fecal viromes. *PLoS ONE.* 2015;10:e0134941.
- Zhang X, McDaniel AD, Wolf LE, Keusch GT, Waldor MK, Acheson DW. Quinolone antibiotics induce Shiga toxin-encoding bacteriophages, toxin production, and death in mice. *J Infect Dis.* 2000;181:664-670.
- Willner D, Furlan M, Haynes M, et al. Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PLoS ONE.* 2009;4:e7370.
- Ly M, Jones MB, Abeles SR, et al. Transmission of viruses via our microbiomes. *Microbiome.* 2016;4:64.
- Enault F, Briet A, Bouteille L, Roux S, Sullivan MB, Petit M-A. Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. *ISME J.* 2017;11:237-247.
- Gansauge M-T, Meyer M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat Protoc.* 2013;8:737-748.
- Zhong X, Guidoni B, Jacas L, Jacquet S. Structure and diversity of ssDNA *Microviridae* viruses in two peri-alpine lakes (Annecy and Bourget, France). *Res Microbiol.* 2015;166:644-654.
- Roux S, Solonenko NE, Dang VT, et al. Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. *PeerJ.* 2016;4:e2777.
- Osawa S, Furuse K, Watanabe I. Distribution of ribonucleic acid coliphages in animals. *Appl Environ Microbiol.* 1981;41:164-168.
- Zhang T, Breitbart M, Lee WH, et al. RNA viral community in human feces: prevalence of plant pathogenic viruses. *PLoS Biol.* 2006;4:e3.
- Nakamura S, Yang C-S, Sakon N, et al. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an

- unbiased high-throughput sequencing approach. *PLoS ONE*. 2009;4:e4219.
46. Lim ES, Zhou Y, Zhao G, et al. Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nat Med*. 2015;21:1228-1234.
 47. Roger LC, McCartney AL. Longitudinal investigation of the faecal microbiota of healthy full-term infants using fluorescence *in situ* hybridization and denaturing gradient gel electrophoresis. *Microbiology*. 2010;156:3317-3328.
 48. Sim K, Cox MJ, Wopereis H, et al. Improved detection of bifidobacteria with optimised 16S rRNA-gene based pyrosequencing. *PLoS ONE*. 2012;7:e32543.
 49. Walker AW, Martin JC, Scott P, Parkhill J, Flint HJ, Scott KP. 16S rRNA gene-based profiling of the human infant gut microbiota is strongly influenced by sample processing and PCR primer choice. *Microbiome*. 2015;3:26.
 50. Reyes A, Blanton LV, Cao S, et al. Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. *Proc Natl Acad Sci U S A*. 2015;112:11941-11946.
 51. Górski A, Wazna E, Dabrowska B-W, Dabrowska K, Swiata-Jeleń K, Miedzybrodzki R. Bacteriophage translocation. *FEMS Immunol Med Microbiol*. 2006;46:313-319.
 52. Monaco CL, Gootenberg DB, Zhao G, et al. Altered virome and bacterial microbiome in human immunodeficiency virus-associated acquired immunodeficiency syndrome. *Cell Host Microbe*. 2016;19:311-322.
 53. Li S-K, Leung RK-K, Guo H-X, et al. Detection and identification of plasma bacterial and viral elements in HIV/AIDS patients in comparison to healthy adults. *Clin Microbiol Infect*. 2012;18:1126-1133.
 54. Breitbart M, Rohwer F. Method for discovering novel DNA viruses in blood using viral particle selection and shotgun sequencing. *Biotechniques*. 2005;39:729-736.
 55. Uhr JW, Dancis J, Franklin EC, Finkelstein MS, Lewis EW. The antibody response to bacteriophage phi-X 174 in newborn premature infants. *J Clin Invest*. 1962;41:1509-1513.
 56. Eriksson F, Tsagozis P, Lundberg K, et al. Tumor-specific bacteriophages induce tumor destruction through activation of tumor-associated macrophages. *J Immunol*. 2009;182:3105-3111.
 57. Kernbauer E, Ding Y, Cadwell K. An enteric virus can replace the beneficial function of commensal bacteria. *Nature*. 2014;516:94-98.
 58. Ott SJ, Waetzig GH, Rehman A, et al. Efficacy of sterile fecal filtrate transfer for treating patients with *Clostridium difficile* infection. *Gastroenterology*. 2017;152:799-811.e7.
 59. Zuo T, Wong SH, Lam K, et al. Bacteriophage transfer during faecal microbiota transplantation in *Clostridium difficile* infection is associated with treatment outcome. *Gut*. 2017; <https://doi.org/10.1136/gutjnl-2017-313952>
 60. Cadwell K, Patel KK, Maloney NS, et al. Virus-plus-susceptibility gene interaction determines Crohn's disease gene *Atg16L1* phenotypes in intestine. *Cell*. 2010;141:1135-1145.
 61. Yang J-Y, Kim M-S, Kim E, et al. Enteric viruses ameliorate gut inflammation via Toll-like receptor 3 and Toll-like receptor 7-mediated interferon- β production. *Immunity*. 2016;44:889-900.
 62. Wagner J, Maksimovic J, Farries G, et al. Bacteriophages in gut samples from pediatric Crohn's disease patients: metagenomic analysis using 454 pyrosequencing. *Inflamm Bowel Dis*. 2013;19:1598-1608.
 63. Pérez-Brocá V, García-López R, Vázquez-Castellanos JF, et al. Study of the viral and microbial communities associated with Crohn's disease: a metagenomic approach. *Clin Transl Gastroenterol*. 2013;4:e36.
 64. Wang W, Jovel J, Halloran B, et al. Metagenomic analysis of microbiome in colon tissue from subjects with inflammatory bowel diseases reveals interplay of viruses and bacteria. *Inflamm Bowel Dis*. 2015;21:1419-1427.
 65. Pérez-Brocá V, García-López R, Nos P, Beltrán B, Moret I, Moya A. Metagenomic analysis of Crohn's disease patients identifies changes in the virome and microbiome related to disease status and therapy, and detects potential interactions and biomarkers. *Inflamm Bowel Dis*. 2015;21:2515-2532.
 66. Conceição-Neto N, Zeller M, Lefrère H, et al. Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. *Sci Rep*. 2015;5:16532.
 67. Kleiner M, Hooper LV, Duerkop BA. Evaluation of methods to purify virus-like particles for metagenomic sequencing of intestinal viromes. *BMC Genom*. 2015;16:7.
 68. Abbas AA, Diamond JM, Chehoud C, et al. The perioperative lung transplant virome: Torque teno viruses are elevated in donor lungs and show divergent dynamics in primary graft dysfunction. *Am J Transplant*. 2016;17:1313-1324.
 69. Salter SJ, Cox MJ, Turek EM, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol*. 2014;12:87.
 70. Zoll J, Rahamat-Langendoen J, Ahout I, et al. Direct multiplexed whole genome sequencing of respiratory tract samples reveals full viral genomic information. *J Clin Virol*. 2015;66:6-11.
 71. Cann AJ, Fandrich SE, Heaphy S. Analysis of the virus population present in equine faeces indicates the presence of hundreds of uncharacterized virus genomes. *Virus Genes*. 2005;30:151-156.
 72. Angly FE, Felts B, Breitbart M, et al. The marine viromes of four oceanic regions. *PLoS Biol*. 2006;4:e368.
 73. Ren J, Ahlgren NA, Lu YY, Fuhrman JA, Sun F. VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome*. 2017;5:69.
 74. Ahlgren NA, Ren J, Lu YY, Fuhrman JA, Sun F. Alignment-free d_2^* oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. *Nucleic Acids Res*. 2017;45:39-53.
 75. Delwart E. A roadmap to the human virome. *PLoS Pathog*. 2013;9:e1003146.
 76. Woolhouse M, Scott F, Hudson Z, Howey R, Chase-Topping M. Human viruses: discovery and emergence. *Philos Trans R Soc Lond B Biol Sci*. 2012;367:2864-2871.
 77. Stanton TB. Prophage-like gene transfer agents-novel mechanisms of gene exchange for *Methanococcus*, *Desulfovibrio*, *Brachyspira*, and *Rhodobacter* species. *Anaerobe*. 2007;13:43-49.
 78. Brown-Jaque M, Calero-Cáceres W, Muniesa M. Transfer of antibiotic-resistance genes via phage-related mobile elements. *Plasmid*. 2015;79:1-7.
 79. Koga Y, Yasunaga M, Moriya Y, et al. Exosome can prevent RNase from degrading microRNA in feces. *J Gastrointest Oncol*. 2011;2:215-222.
 80. Edgar JR. Q&A: What are exosomes, exactly? *BMC Biol*. 2016;14:46.
 81. Sjöström AE, Sandblad L, Uhlin BE, Wai SN. Membrane vesicle-mediated release of bacterial RNA. *Sci Rep*. 2015;5:15329.
 82. Hoyles L, Murphy J, Neve H, et al. *Klebsiella pneumoniae* subsp. *pneumoniae*-bacteriophage combination from the caecal effluent of a healthy woman. *PeerJ*. 2015;3:e1061.
 83. Sangster W, Hegarty JP, Stewart DB. Phage tail-like particles kill *Clostridium difficile* and represent an alternative to conventional antibiotics. *Surgery*. 2015;157:96-103.
 84. Dalmasso M, Strain R, Neve H, et al. Three new *Escherichia coli* phages from the human gut show promising potential for phage therapy. *PLoS ONE*. 2016;11:e0156773.
 85. Morozova V, Kozlova Y, Shedko E, et al. Lytic bacteriophage PM16 specific for *Proteus mirabilis*: a novel member of the genus *Phikmvirus*. *Arch Virol*. 2016;161:2457-2472.

86. Cotten M, Oude Munnink B, Canuti M, et al. Full genome virus detection in fecal samples using sensitive nucleic acid preparation, deep sequencing, and a novel iterative sequence classification algorithm. *PLoS ONE*. 2014;9:e93269.
87. Džunková M, D'Auria G, Moya A. Direct sequencing of human gut virome fractions obtained by flow cytometry. *Front Microbiol*. 2015;6:955.
88. Li L, Deng X, Mee ET, et al. Comparing viral metagenomics methods using a highly multiplexed human viral pathogens reagent. *J Virol Methods*. 2015;213:139-146.
89. Angly F, Rodriguez-Brito B, Bangor D, et al. PHACCS, an online tool for estimating the structure and diversity of uncultured viral communities using metagenomic information. *BMC Bioinformatics*. 2005;6:41.
90. Lepiae R, Lima-Mendez G, Toussaint A. ACLAME: a CLAssification of Mobile genetic Elements, update 2010. *Nucleic Acids Res*. 2010;38(Database issue):D57-D61.
91. Roux S, Faubladier M, Mahul A, et al. Metavir: a web server dedicated to virome analysis. *Bioinformatics*. 2011;27:3074-3075.
92. Roux S, Tournayre J, Mahul A, Debroas D, Enault F. Metavir 2: new tools for viral metagenome comparison and assembled virome analysis. *BMC Bioinformatics*. 2014;15:76.
93. Wommack KE, Bhavsar J, Polson SW, et al. VIROME: a standard operating procedure for analysis of viral metagenome sequences. *Stand Genomic Sci*. 2012;6:427-439.
94. Jarrett RF, Gallagher A, Gatherer D. Molecular methods of virus detection in lymphoma. *Methods Mol Biol*. 2013;971:277-293.
95. Roux S, Enault F, Hurwitz BL, Sullivan MB. VirSorter: mining viral signal from microbial genomic data. *PeerJ*. 2015;3:e985.
96. Rampelli S, Soverini M, Turroni S, et al. ViromeScan: a new tool for metagenomic viral community profiling. *BMC Genom*. 2016;17:165.
97. Li Y, Wang H, Nie K, et al. VIP: an integrated pipeline for metagenomics of virus identification and discovery. *Sci Rep*. 2016;6:23774.
98. Zheng Y, Gao S, Padmanabhan C, et al. VirusDetect: An automated pipeline for efficient virus discovery using deep sequencing of small RNAs. *Virology*. 2017;500:130-138.
99. Arndt D, Grant JR, Marcu A, et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res*. 2016;44:W16-W21.
100. Zhao G, Wu G, Lim ES, et al. VirusSeeker, a computational pipeline for virus discovery and virome composition analysis. *Virology*. 2017;503:21-30.
101. Petersen C, Round JL. Defining dysbiosis and its influence on host immunity and disease. *Cell Microbiol*. 2014;16:1024-1033.
102. Tamboli CP, Neut C, Desreumaux P, Colombel JF. Dysbiosis in inflammatory bowel disease. *Gut*. 2004;53:1-4.
103. Moos WH, Faller DV, Harpp DN, et al. Microbiota and neurological disorders: a gut feeling. *BioResearch Open Access*. 2016;5:137-145.

SUPPORTING INFORMATION

Additional Supporting Information will be found online in the supporting information tab for this article.

How to cite this article: Carding SR, Davis N, Hoyles L. Review article: the human intestinal virome in health and disease. *Aliment Pharmacol Ther*. 2017;00:1-16. <https://doi.org/10.1111/apt.14280>