CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Metagenomic approaches to assess bacteriophages in various environmental niches
Authors	Hayes, Stephen;Mahony, Jennifer;Nauta, Arjen;van Sinderen, Douwe
Publication date	2017-05-24
Original Citation	Hayes, S., Mahony, J., Nauta, A. and van Sinderen, D. (2017) 'Metagenomic Approaches to Assess Bacteriophages in Various Environmental Niches', Viruses, 9, 127 (22pp). doi: 10.3390/v9060127
Type of publication	Article (peer-reviewed)
Link to publisher's version	http://www.mdpi.com/1999-4915/9/6/127 - 10.3390/v9060127
Rights	© 2017 by the Authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/). - https://creativecommons.org/licenses/by/4.0/
Download date	2026-11-17 03:49:26
Item downloaded from	https://hdl.handle.net/10468/4104

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

viruses

MDPI

Review

Metagenomic Approaches to Assess Bacteriophages in Various Environmental Niches

Stephen Hayes [1], Jennifer Mahony [1,2], Arjen Nauta [3] and Douwe van Sinderen [1,2,*]

[1] School of Microbiology, University College Cork, Cork T12 YT20, Ireland; stephen.hayes@umail.ucc.ie (S.H.); j.mahony@ucc.ie (J.M.)
[2] APC Microbiome Institute, University College Cork, Cork T12 YT20, Ireland
[3] Friesland Campina, Amersfoort 3800 BN, The Netherlands; Arjen.Nauta@frieslandcampina.com
* Correspondence: d.vansinderen@ucc.ie; Tel.: +353-21-4901365

Academic Editors: Tessa E. F. Quax, Matthias G. Fischer and Laurent Debarbieux
Received: 31 March 2017; Accepted: 19 May 2017; Published: 24 May 2017

Abstract: Bacteriophages are ubiquitous and numerous parasites of bacteria and play a critical evolutionary role in virtually every ecosystem, yet our understanding of the extent of the diversity and role of phages remains inadequate for many ecological niches, particularly in cases in which the host is unculturable. During the past 15 years, the emergence of the field of viral metagenomics has drastically enhanced our ability to analyse the so-called viral 'dark matter' of the biosphere. Here, we review the evolution of viral metagenomic methodologies, as well as providing an overview of some of the most significant applications and findings in this field of research.

Keywords: virome; phage; marine; microbiota

1. Introduction

Viruses infecting bacteria, or bacterio(phages), are presumed to represent the most abundant biological entities on earth [1]. They exist wherever bacterial life is found [2], with millions of phages found in every drop of seawater [3], and the human gut is estimated to contain between 10^8 and 10^9 virus-like particles (VLPs) per gram of faeces [4,5], of which many are undoubtedly phages. Since their initial discovery in 1915 by Frederick Twort, followed by the realisation in 1917 by Felix d'Herelle that phages had the potential to kill bacteria [6], phages have been at the cutting edge of molecular biology research, both as model systems and as potential biological tools for the manipulation of bacterial genomes [7]. Numerous studies have explored the role of phages in various ecosystems, and it has thus become apparent that phages exert their influence across every aspect of life. For example, cholera toxin, the causative agent of many symptoms of cholera, is encoded by a temperate bacteriophage or prophage [8], as are the virulence factors in various pathogens causing bacterially-derived food poisoning and diphtheria [9,10].

Given their ubiquity in the environment, it is perhaps unsurprising that bacterial evolution is to a large degree driven by phages [11,12], facilitated by recombination with or the integration of prophages [13] or by evolutionary responses to evade lytic phage infection [14]. Thus, bacteriophages are a core part of the ecosystem, modulating microbial nutrient cycles, community structure, and long-term evolution [15–17]. Despite this, phages are often marginalised, if not completely ignored, in many studies, and this omission may thus result in conclusions that ignore a crucial part of the picture [18]. Therefore, even though phage genomes are orders of magnitude smaller than bacteria and thus easy to sequence, less than 2700 double-stranded DNA (dsDNA) virus and retrovirus genomes are deposited in NCBI (National Center for Biotechnology Information) databases, compared to almost 90,000 prokaryotic genomes (as of February 2017). Consequently, a full understanding of the scope of virus-host interactions is lacking [19], and current estimates state that the field of virology has possibly

explored less than 1% of the extant viral diversity [20]. However, with in silico, culture-independent techniques becoming available in recent decades, this is beginning to change. In this review, we assess the range of methods employed in viral metagenome analyses, while we also outline the limitations and advances in the field. Furthermore, we will discuss the range of environmental niches where such approaches have been successfully implemented to improve our understanding of viral communities and their perceived impact on microbial landscapes.

2. Culture-Dependent vs. Culture-Independent Methods of Bacteriophage Study

For over a century, culture-based methods have been the primary approach to detect and isolate phages [20]. This involves cultivating hosts and isolating individual phages from plaques using methods such as double agar assays. However, such approaches carry significant limitations due to the recalcitrance of the majority of microbes to cultivation under laboratory conditions [21]. It is currently predicted that up to 99% of microorganisms are unculturable [22,23]. Moreover, even in the presence of a culturable host, phage identification and isolation may remain difficult [24], as not all phages are capable of forming plaques and there is evidence that many successfully infect in the absence of discernible plaque formation [25,26]. Further factors such as pseudolysogeny (the stalled development of a bacteriophage in a host cell) [27] and differences between laboratory-based assays and phage–host interactions in nature [28], with many phages for example requiring their host to be in a specific growth phase to successfully infect [29], only serve to increase the difficulty of studying phages. Thus, while a wealth of information has been gained from these methods, it is clear that culture-independent methods are vital to further our understanding of the diversity of phages and their role and impact in various environmental niches.

Culture-Independent Methods of Bacteriophage Study

The first culture-independent methods arose in the 1980s as an approach to characterise microbial diversity via DNA sequencing [30–32]. These early techniques utilized 16S ribosomal DNA (rDNA) as a common marker gene since it is universally present in bacteria and archaea [33–35]. However, the limitations of such approaches, combined with the lack of universally conserved markers in phages [36], has led to the development of alternative technologies for the culture-independent study of bacteriophages, including randomly amplified polymorphic DNA (RAPD) PCR [37], flow cytometry [38], electron microscopy [3,39,40], single virus genomics [41], and viral tagging [42] (Table 1). Studies employing electron microscopy showed that viruses are far more abundant in the oceans than previously predicted and by inference in several other ecological niches. These techniques testify to the rapidly increasing and evolving range of approaches to the culture-independent study of previously inaccessible viral communities. Furthermore, it highlights that their use either individually or as part of a complementary approach can provide insights into the population composition and genetic diversity of environmental viral samples. However, by far the most successful method that has arisen for the culture-independent study of viral communities is metagenomic analysis.

Table 1. Selection of culture-independent methods for the study of bacteriophages.

Method	Description	Limitations
Gene marker-based studies [43–47]	Utilise marker genes, ranging from major capsid proteins to photosynthesis related genes, to study the diversity of viruses in a sample.	Lack of universal viral gene limits the focus of studies to particular phage genera [48]; cannot provide quantitative analysis [24].
Randomly Amplified Polymorphic DNA (RAPD) PCR [37]	Uses short, random primers to amplify fragments of environmental DNA of assorted sizes. Provides a rapid, rudimentary comparison of viral diversity.	Limited inferences possible; difficult to reproduce results due to high sensitivity of the technique to reaction conditions [49].
Electron microscopy [3,39,40]	Allows enumeration of uncultured viruses, particularly in marine samples. Accuracy and speed improved by epifluorescent microscopy [50–53].	Limited to observation of morphologies and rough estimates of quantity of viral particles; no sequence data generated.

Table 1. *Cont.*

Method	Description	Limitations
Flow Cytometry [38,54,55]	Rapid enumeration of viral particles in a sample via their staining with highly fluorescent nucleic acid dyes followed by counting via flow cytometry.	Limited to estimations of quantity; no sequence data generated or morphology information.
Single virus genomics [41]	Enables isolation and complete genome sequencing of single viral particles. Involves sorting of single viruses by flow cytometry, followed by genome amplification via multiple displacement amplification (MDA) and whole genome sequencing.	Does not provide community-wide view of viral population.
Viral Tagging [42,56]	Allows study of phage–host interactions by fluorescently labelling phages and using them to 'tag' their host. Phages inject labelled genomes into their host, rendering the bacteria fluorescent. Potential hosts are then sorted via fluorescence-activated cell sorting (FACS).	Requires a culturable host, extensive optimisation required for each new host [57].

3. Metagenomics

The term metagenomics was first coined in 1998 [58] and is defined as the direct sequencing and analysis of all genetic material recovered from an environmental sample [59]. There are two primary approaches used for the metagenomic study of uncultured microbial populations; shotgun metagenomics, which involves the sequencing of the entire nucleic acid compliment of a sample [60], and marker gene amplification metagenomics, typically using the 16S ribosomal RNA gene [61]. The optimal method to be used varies according to the goals of a study and the resources available. Full shotgun metagenomics is far more costly and time-consuming but, due to the unrestricted sequencing of all genomes in the sample, will provide far more information, while 16S metagenomics is largely restricted to taxonomic composition of the bacterial/archaeal population of the sample but is much more rapid and less costly [62]. Thus, using these techniques, both qualitative and quantitative analysis of uncultured microbial communities has now become possible [63]. This versatility, coupled with the emergence in the new millennium of widely available high-throughput sequencing (HTS), has resulted in metagenomics becoming the most effective and comprehensive approach for the genomic analysis of uncultured microbial populations [24]. As a result, the number of published metagenomic studies has risen explosively from 11 in 2002 to over 10,000 in 2017 (using metagenomics as a search term in the PubMed search engine), while there are now more than 8000 metagenomic datasets (of which over 4600 are publicly available) deposited in the Integrated Microbial Genomes and Microbiomes (IMG/M) system, a public comparative resource for both sequenced genomes and metagenomic datasets [64]. Far from simply producing massive volumes of sequencing data, these metagenomic studies are continually yielding novel information such as genomic linkage between function and phylogeny and evolutionary profiles of community function and structure [59], as well as facilitating the discovery of novel genes and enzymes [65–68].

4. Viral Metagenomic Sample to Sequence Pipeline

The growing success of microbial metagenomics, coupled with the increasing awareness of the vital role of viruses in nature, has resulted in attention quickly turning towards the application of metagenomics to the field of virology. The challenges of applying metagenomics to viral samples are many: the lack of universal marker genes, meaning full shotgun metagenomics is required; contamination with bacterial DNA, which is far more abundant than viral, and the consequential difficulties in separating these during sequence analysis [69]; the vast diversity of virus types in nature, making the isolation, sequencing, and assembly of an unbiased viral metagenome extremely difficult [70]; and a lack of viral sequence in databases, making comparative genomics of limited value, to name but a few. However, viruses also possess a number of characteristics favourable to metagenomic analysis. For example, their small size permits the effective removal of cellular debris by centrifugation and/or filtration, allowing rapid purification of phages in a diverse array of samples.

The characteristic buoyant densities of viruses facilitate their selective purification via cesium chloride gradient [71] (although it must always be remembered that this will select against VLPs outside the densities examined, biasing the resulting population [72]). Additionally, the comparatively small genome size of viruses is well suited to sequencing techniques, provided nucleic acids of sufficient quality and quantity can be isolated [18]. Thus, as metagenomic techniques have advanced, the field of viral metagenomics has expanded significantly (Figure 1). Since the first application of viral metagenomics to uncultured marine samples in 2002 [73], virome (the nucleic acid complement of all viruses in a population) studies have been applied to a wide range of environments and locations including freshwater, seawater, soil, industrial fermentations, and the guts of humans and many other organisms. Due to the diversity of environments, there is no single, one-fits-all method that can be employed, with protocols requiring sample/source-specific adaptations. However, studies generally involve a number of key steps: (i) viral particle purification; (ii) nucleic acid extraction, (iii) high-throughput sequencing of purified viral nucleic acids, and (iv) bioinformatic analysis and interpretation of sequence data. While the application of metagenomics to the study of phages is the primary focus of this review, phages are generally isolated from a sample in congruence with eukaryotic viruses, except if the desired phages could be separated from eukaryotic viruses in the early stages of a metagenome isolation using methods such as density gradients. Indeed, most virome studies aim to characterise all viruses, only distinguishing phages during the final analysis of sequence data. Thus, the following viral metagenomic pipeline provides a broad overview of techniques for the generation of viral metagenomes, which is equally applicable to both phages and eukaryotic viruses.

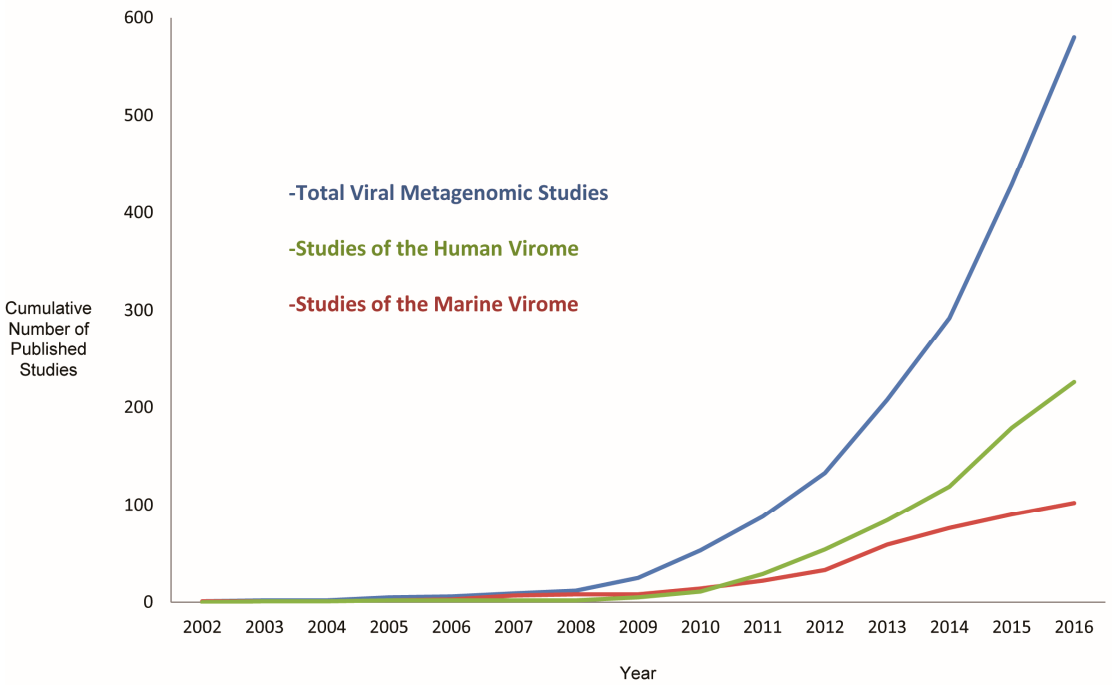


Figure 1. Graph illustrating the large increase in the publication of viral metagenomic studies from the initial study of Breitbart et al. [73] in 2002 to the end of 2016. The total cumulative number of studies is represented in blue. The number of metagenomic studies of the human virome is represented in green, and studies of the marine virome in red. The number of studies in each case was determined via Pubmed search.

4.1. Viral Particle Extraction

The purification of VLPs from an environmental sample is the initial step of any virome project and arguably the most critical. Techniques must strive to produce VLP samples that both qualitatively and quantitatively represent the diversity of the population, thus allowing the linkage

of metagenomic observations to the original population [24]. The basic steps indicated below can be adapted to suit the requirements of the sample of interest: (i) recovery of VLPs within the sample, (ii) VLP purification and concentration, and (iii) optional final purification via cesium chloride gradient [24,74–78]. The extraction and purification of bacteriophages from the human gut for viral metagenomic analysis has recently been optimised [74], and the extraction protocols used in this study are outlined in Figure 2. By optimising the basic extraction steps previously mentioned, this study succeeded in greatly increasing the number of phage particles and the quantity of viral DNA obtained and produced two optimised extraction protocols, which are easily modifiable and thus in principle applicable to almost any environmental sample [74].

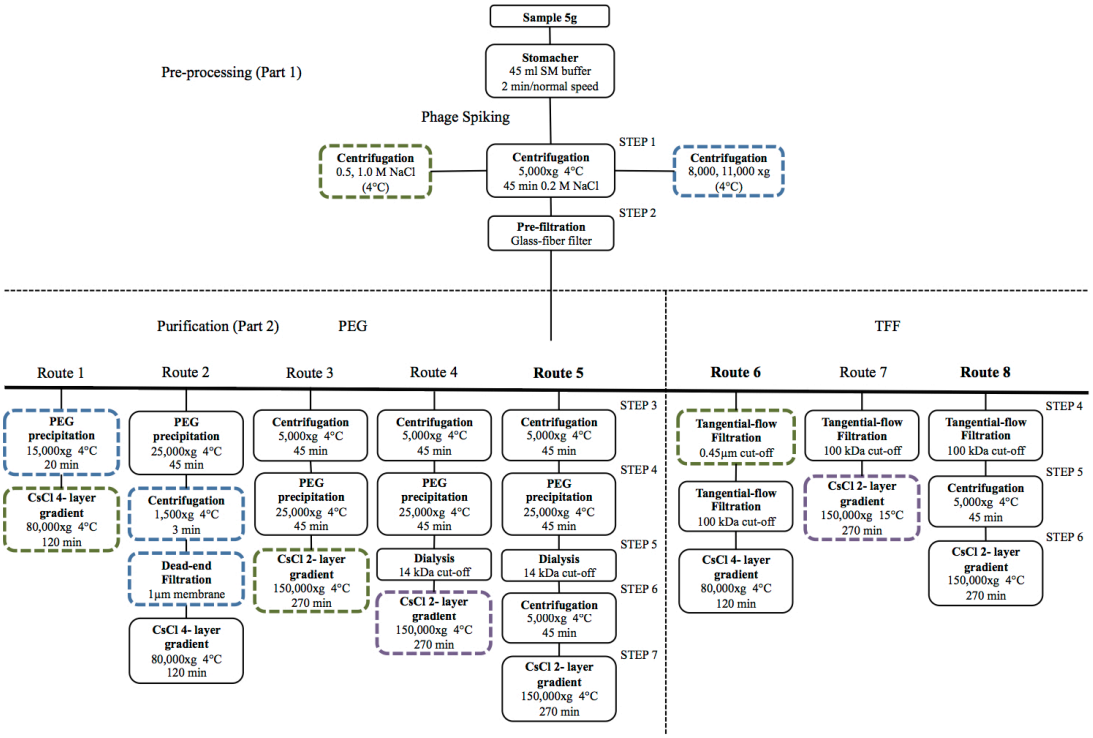


Figure 2. Optimization of the extraction of phages from the human gut for viral metagenomic analysis. In this study, samples were 'spiked' with a set titre of known phages, and the efficiency of recovery of these phages was monitored throughout a number of extraction protocols. Part 1 (pre-processing) involves the suspension/dissolution of phages from the samples and the removal of large particles. Part 2 (phage purification) then removed lower molecular weight impurities and microbial cells. Boxes with non-continuous borders represent steps that were deemed unsuitable for phage extraction, either due to large losses of spiked phages or impure samples. Green bordered boxes represent steps that resulted in impure samples, as assessed by visual inspection and transmission electron microscopy (TEM). Blue bordered boxes indicate steps at which >50% of spiked phages were lost. Purple bordered boxes signify steps that failed to remove microbial contamination. Two main purification routes were optimized polyethylene glycol (PEG) precipitation and tangential flow filtration (TFF), and routes were diverted into new extraction routes (i.e., from route 1 to route 2, etc.) until the highest recovery of spiked phages was reached. This optimisation resulted in the purification of greatly increased numbers of phage particles and much higher quantities of DNA in comparison to previous protocols [74]. Reprinted with permission from Castro-Mejía et al. [74].

Beyond the general approaches, specialised methods catering to particular environments have been developed such as the flocculation, filtration, and resuspension (FFR) method, which uses an iron-based flocculation and subsequent resuspension of virus-containing precipitates and which boasts a 90% recovery rate of marine viral particles [79,80].

4.2. Nucleic Acid Extraction

The next step in the preparation of a sample is the extraction and purification of its viral nucleic acid content. This step must yield nucleic acids of sufficient purity and concentration for downstream library construction and sequencing [59]. Before this step, a DNase treatment is generally performed to eliminate contaminating cellular DNA, the presence of which can drown out the sequences recovered from viral genomes, while also creating difficulties in downstream analysis [69]. However, due to the diverse range of virus types, the nature of this treatment will be highly dependent on the aims of the researcher and the focus of the study. Nucleic acid extraction is then generally performed using one of a range of commercially available kits [76], although it is important to note that studies in bacteria suggest that the choice of kit can have an influence on the community structure produced by HTS [81], which could in turn result in inaccurate or biased conclusions being drawn [82]. After viral DNA extraction, PCR analysis employing 16S/18S rRNA gene-based primers may be useful to assess the presence of bacterial host DNA, offering a semi-quantitative measure of microbial contamination [83], and quantitative PCR (qPCR) can be used to provide a more accurate estimate of host contaminants.

4.3. Library Preparation and Sequencing

The advent of next generation sequencing (NGS) has resulted in the emergence of rapid, affordable, and high throughput methods such as Illumina MiSeq and Ion Torrent sequencing [84], making metagenomic sequencing far more accessible. In preparation for sequencing, libraries are prepared from the isolated viral DNA, following fragmentation to suitable lengths for the specific sequencing platform(s) to be used whilst minimizing sample loss and preventing the introduction of bias [84]. Varying fragmentation techniques are used depending on the sequencing platform, with an in-depth summary of suitable methods and their advantages and disadvantages discussed in [85]. Fragmentation frequently results in single-stranded DNA (ssDNA) ends, which require repair in preparation for the next step in sequencing, dsDNA adaptor ligation. Once the adaptors are ligated, they serve as primer sites during the sequencing reaction and may also contain a barcode to allow the sequencing of pooled libraries in a single run.

Amplification is commonly one of the final steps in library preparation before sequencing, creating ca. 1000 copies of the DNA to be read by the sequencer [85]. This step is especially crucial to virome studies as, though viruses are much more abundant than their hosts, their genomes are orders of magnitude smaller than those of microbes [24]. The earliest metagenomic amplification technique was linker-amplified shotgun libraries (LASLs) [25]. LASLs consist of randomly sheared cDNA fragments ligated to known adapter sequences for PCR amplification, which are then cloned into plasmid vectors and Sanger sequenced [73]. Whole genome amplification using methods such as multiple displacement amplification (MDA), which utilises the polymerase of Φ29, were developed to improve throughput using NGS platforms [25]. However, these systems have been shown to not only introduce a systematic bias related to the preferential amplification of single-stranded and circular DNA templates, which is particularly relevant in the case of viral metagenomics [86], but also non-predictable and random biases [87]. These combined biases skew the taxonomic representation of a community, resulting in non-quantitative metagenomes [88,89], thereby preventing comparative analyses.

Improved library preparation methods have been developed (Table 2), including linear amplification for deep sequencing (LADS) [90] and the Nextera fragmentation and adapter ligation methods [91]. Nextera is an extremely rapid two-step method in which the simultaneous fragmentation and tagging of genomic DNA using modified transposition in a process designated as 'tagmentation' [91] is followed by a reduced cycle PCR to add adaptors. However, Nextera has been demonstrated to introduce amplification bias, preferentially amplifying genomic regions of low GC content [91]. Additionally, Nextera requires a minimum of 50 ng of starting material, which can be difficult for certain virome samples [24], although the newer iteration, Nextera XT, has reduced this requirement to 1 ng, with one study of a microbial metagenome even lowering this to 50 pg of input DNA [92]. The LADS protocol, developed for Illumina Sequencing, has emerged as an attempt to

circumvent GC bias by replacing the PCR step with a transcription step [90]. However, LADS requires significant expertise and is particularly labour-intensive [25]. Despite recent advances, an optimization of the original linker amplification (LA) method [93] by Duhaime et al. in 2012 [94], may provide the most quantitative, next-generation-sequence-ready DNA and can be adapted for use on the Illumina, 454 or Ion Torrent sequencing platforms [24].

Table 2. Required nucleic acid quantities, advantages, and drawbacks of being commonly employed for virome library preparation.

Method	Nucleic Acid Quantity	Advantages	Drawbacks
Multiple displacement amplification (MDA) [95]	1–100 ng	Rapid and high-throughput	Introduces both predictable and stochastic biases
Linear amplification for deep sequencing (LADS) [90]	3–40 ng	Low levels of bias introduced, resulting in near-quantitative metagenomes	Low throughput, requires significant expertise
Linker amplified library construction [94]	>10 pg	Remains the most quantitatively accurate method, requires minimal nucleic acid input	Low throughput, requires significant expertise
Nextera XT (Illumina)	50 pg	Rapid, combines fragmentation and tagging of DNA into single 5 min 'tagmentation' step	Slight sequence-dependent biases at low nucleic acid input levels [92]

Improved library preparation techniques are still emerging, including the previously mentioned Nextera XT, multiple annealing and looping-based amplification cycles (MALBAC), and NuGEN's Mondrian microfluidic workstation used in conjunction with the NuGEN Ovation library prep kit. MALBAC reduces amplification bias and increases coverage by utilising a semi-linear amplification method [96], while the Mondrian approach automates much of the library preparation protocol. The impact of these three methods, along with template quantity, on the metagenomic output obtained from a mock microbial community from as little as 1 pg of DNA has recently been assessed [97]. It was found that template quantity in all three methods had a significant impact on the revealed community composition, highlighting that unbiased amplification techniques are beyond current capabilities and represent an area that will need further improvement [97]. However, the ability to perform metagenomic sequencing with reduced DNA quantities (as compared to previous protocols) represents an especially relevant advance in the field of viral metagenomics.

Once a DNA library has been prepared, sequencing is performed, primarily utilising one of the three main NGS platforms; Illumina (currently the most popular [62]), Roche 454 (discontinued in 2016), and Ion Torrent PGM (for a review on sequencing technologies, see [98]). Following sequencing, quality control steps are performed as recently reviewed [84] to ensure that the sequence data is ready for analysis. These quality controls include ensuring the adequate sequencing coverage of each sample and the removal of rare reads, which may represent sequencing errors or contamination and thus cause overestimation of the protein richness of a virome [80].

4.4. Analysis

While continuing advances in sequencing technology have opened up a multitude of opportunities in the field of viral metagenomics, the enormous amounts of data produced by NGS has also resulted in a major challenge; quality analysis and the processing of sequence data. The primary constraining factor regarding the effective bioinformatic analysis of viromes is the previously discussed absence of universal gene markers in viral genomes, meaning that the detection of viral reads is limited to their alignment against reference viral sequence databases. As of February 2017, the NCBI virus genome

database contained just over 2000 phage genomes, with nearly half of those derived from just four genera of bacteria, which have been studied in detail due to their clinical relevancy (*Mycobacterium*, *Enterobacteria*, *Pseudomonas*, and *Staphylococcus* [84]). It is widely accepted that virtually all bacteria suffer from phage predation; however, just eight of the 29 known bacterial phyla with cultured isolates have sequenced phage representatives [99]. Therefore, it is clear that vast data exists that is inaccessible by current comparative analysis. The lack of sequence identity typically results in viral metagenomes containing between 60% and 99% of sequences, which possess no significant homology to sequences in current databases [89]. However, this high proportion of 'unknown' sequences also presents the greatest opportunity, with a treasure trove of uncharacterised sequences.

The taxonomic composition of a sequenced viral metagenome is often analysed by the alignment of the virome against reference databases using BLAST (Basic Local Alignment Search Tool), or BLAST-based programs such as MG-RAST (MetaGenomic-Rapid Annotation using Subsystem Technology) [100], MetaPhyler [101], or CARMA [102], which is a time-consuming process. Alternatively, rapid fast *k*-mer algorithms may be applied to reduce the time associated with such analyses [84,89]. These algorithms are more than 50× faster than alignment-based approaches [103] and are incorporated in programs including CLARK (CLAssifier based on Reduced *K*-mers) [104] and USEARCH [105]. However, these more rapid methods typically require heavy computing power (>128 Gb of Random Access Memmory (RAM)) [104] and are thus restricted to comparisons against reference databases such as Refseq, which, given the modest proportion of viral representation in these databases, limits the level of annotation possible. For example, Hurwitz et al. (2013) compared the annotation of the Pacific Ocean Virome (POV) dataset, originally obtained employing BLASTx alignments against all known proteins at the genus and family level [106] against taxonomic data obtained via a re-annotation using CLARK [84]. The authors found that 1.12% and 0.96% of reads matched regions in bacterial and viral genomes, respectively, using CLARK, in contrast to 4.01% and 6.87% matches against bacterial and viral proteins using BLASTx. This example serves to highlight that there is currently a persistent trade-off in the field of viral metagenomics between speed and accuracy.

4.5. Bioinformatic Tools for Viral Metagenome Analysis

In order to circumvent the large demands for computer processing power of these methods, a number of online resources and tools have become available, making metagenomic analysis more accessible to the uninitiated. These computational pipelines have been designed for the purpose of analysing the composition of metagenomic datasets; in the case of viromes, this means that the abundance and types of viruses present in a sample can be defined. These include virome-specific programs such as VIROME (Viral Informatics Resource for Metagenome Exploration) [107], Metavir [108], and VMGAP (Viral MetaGenome Annotation Pipeline) [109] and also more 'generalist' pipelines including those previously mentioned programs incorporating BLAST based analysis. These pipelines are generally utilizing ORF (Open reading frame)-finding algorithms, which predict coding sequences followed by subsequent comparison with protein databases. A recent study by Tangherlini et al. [110] involved an in-depth comparison of these tools for the analysis of the taxonomic composition of both simulated and actual benthic deep-sea viral metagenomes. This study confirmed translated BLAST (tBLASTx) as the most reliable tool for the accurate analysis of viral diversity, followed by the Metavir tool. Furthermore, the authors highlight that, as with all steps in the viral metagenome process, the choice of bioinformatic tool can significantly influence the obtained findings and derived conclusions [110].

In addition to these tools, all based upon sequence comparisons to reference databases, numerous similarity-independent methods have arisen in order to circumvent the lack of sequence similarity in current databases [111]. The primary tool designed for this purpose remains PHACCS (Phage Communities from Contig Spectrum), which provides estimates of the richness, evenness, and abundance of the most abundant viruses in a viral metagenome [112], based on the principle that the most abundant virotypes (taxonomic classification based on a percentage identity threshold rather than phylogenetic markers) will more likely be assembled into large contigs [111].

Other reference-independent tools include MaxiPhi [113], which analyses inter-sample diversity between two samples, and crAss [114], which facilitates the simultaneous cross-assembly of all samples in a data set.

These tools offer just a sample of those available, and the range of bioinformatic tools ready for use in the analysis of viral metagenomes has recently been reviewed [115,116]. Moreover, new tools are continually emerging, such as VirSorter [117] and MetaPhinder [118], both designed for the detection of viral sequences in metagenomic data; VirusSeeker, released in early 2017 (mainly focused on eukaryotic viruses, though it does incorporate bacteriophage analysis in the pipeline [119]); and the iVirus community resource, which provides access to a range of viral metagenomic tools and datasets [120]. Thus, as methods improve, the discrepancies and biases introduced by these programs will hopefully be overcome.

5. Current and Potential Areas of Interest for Viral Metagenomics

By applying the workflow outlined in Section 4 to the sample of interest, it is theoretically possible to perform viral metagenome analysis on virtually any sample. Indeed, a plethora of studies have already been performed on an array of environments, and some of the dominant niche areas are discussed below.

5.1. Marine Viral Metagenomics

Since the pioneering study of Breitbart et al. in 2002 [73], marine phage genomics has been at the forefront in the field of viral metagenomics. Oceans cover over 70% of the Earth's surface, produce over half of the oxygen in the atmosphere, and absorb the most carbon dioxide from it. Driving these energy cycles are the marine microbes, which constitute more than 90% of the living biomass in the sea. Considering that viruses kill roughly 20% of this biomass each day [16], driving the so-called 'viral shunt' (the role of viruses in the transformation of living biomass into dissolved organic matter) [57], it is clear that marine bacteriophages play a critical role in the biosphere. Consequently, a decade after the above-mentioned first marine virome study, a large-scale, consortia-driven study [106] of the marine virome culminated in the development of the previously mentioned POV.

The POV dataset consists of 32 near-quantitative dsDNA viromes collected from samples at various locations, depths, and seasons and has led to many novel insights [57]. To organise the 'unknown' sequences, which comprised the vast majority of viral sequences obtained, protein clusters (PCs) were formed using an approach originally applied to microbial metagenomics [121]. Protein clusters are groups of ORFs grouped together by sequence similarity [122], and this resulted in the creation of over 450,000 PCs to aid in the mapping of future viromes [106]. In addition, by performing rarefaction analysis of the PCs obtained from different locations/depths/seasons, the community diversity of the viral populations was compared. From this, it was demonstrated that the functional capacity of viral communities varied according to depth, season, and distance from the shore (Figure 3), with the greatest richness (not to be confused with abundance, richness simply measures the number of species in a community without considering the population size of each) observed in areas near to the shore and in the aphotic zone of the ocean during winter [106]. Furthermore, it was found that core PCs (those present in every sample) were enriched in the photic zone relative to the aphotic zone, possibly indicating unidirectional genetic exchange from the surface to deep oceans, most probably facilitated by viral particles descending rapidly in aggregated biomass via vertical flux [123]. Finally, the rarefaction curves of POV PCs lowered the predicted number of marine PCs to 0.5–1.3 million and the global virome to 3.9 million PCs [124], far lower than the two billion distinct viral proteins previously predicted [1].

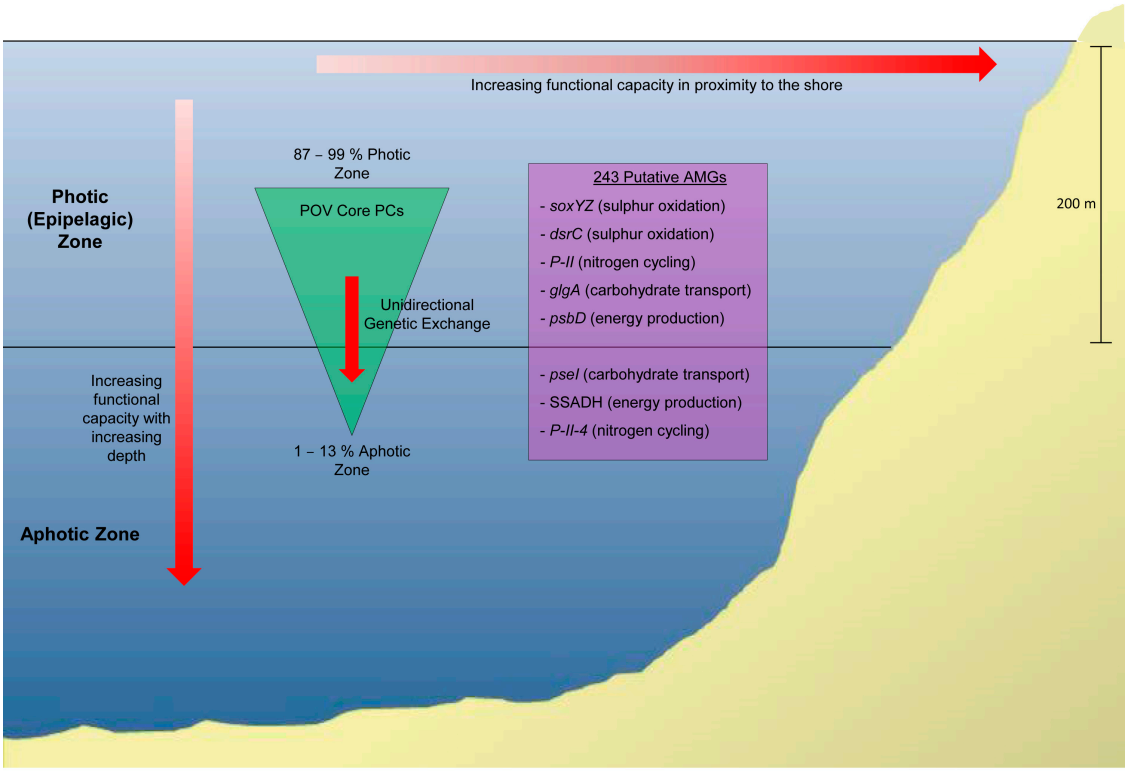


Figure 3. Functional richness of the ocean virome as determined by the Pacific Ocean Virome (POV) [106] and the Global Ocean Virome (GOV) [125] datasets. Analysis of functional diversity revealed that the functional richness of viral communities decreased in surface communities as distance from the coast increased and increased in the ocean as depth increased. It was also found that core PCs (protein clusters) of the POV (i.e., those present in all samples) are enriched in the photic zone relative to the aphotic zone, indicating unidirectional genetic exchange from surface waters to the deep ocean. Additionally, 243 putative auxiliary metabolic genes (AMGs) were identified, examples of which can be seen for each zone.

The POV also provided the first in-depth, large-scale survey of the presence of Auxiliary Metabolic Genes (AMGs) in marine viruses. AMGs are host-derived genes present in phage genomes, and their encoded proteins are suggested to augment the metabolism of infected hosts at key metabolic bottlenecks, facilitating and enhancing the production of new viral particles [126]. Sequence data from the POV greatly advanced the repertoire of known AMGs and demonstrated that AMGs are far more diverse than previously thought [127]. Virtually all genes encoding functions that are required for carbon metabolism were found in the virus population, as well as genes involved in amino acid production and energy production and genes involved in other cellular functions such as motility and transport [123]. Thus, the POV has provided a wealth of information and continues to be consolidated and expanded by emerging studies.

A number of other large-scale investigations have been performed [57], and in 2016 the Global Ocean Virome (GOV) was released [125]. The GOV is comprised of 104 virome datasets (both surface- and deep-ocean) sampled during the Tara Oceans and Malaspina research expeditions [128,129], providing a global map of abundant, dsDNA marine viruses. In this study, viral contigs were clustered based on co-occurrence and nucleotide signature to form 'viral populations', roughly equivalent to viral species. Using this method, 15,280 viral populations were identified, and rarefaction analyses indicated that, as a result of the GOV, sampling of dsDNA viral communities in the epipelagic zone, or surface layer, of the ocean is now nearing completion [125]. Virus populations, along with the publicly available phages and archaeal viruses, were then categorised into Viral Clusters (VCs), roughly

corresponding to viral genera [130]. This produced 1259 VCs, of which 658 were exclusive to the GOV, and 209 others were contained GOV sequences, indicating a doubling of known phage and archaeal virus genera. Subsequently, the global abundance of each VC was analysed, and it was observed that only 38 of 867 VCs were abundant in more than one station. Interestingly, of the 38 'abundant' VCs, only 20 contained previously known (either from viral isolates or environmental sequencing) viral sequences, while 18 were completely unreported. Thus it is clear that, even in the case of some of the most globally abundant viruses, much is still unknown about them.

The GOV consortia then went further, and sought to link this viral data to microbial hosts. The linking of viral sequence data to host strains has long been a major challenge in the field of viral metagenomics [57], although a number of new methods have emerged. Such methods include physical methods such as 'viral tagging' [42] and sequence-based methods utilising similarity approaches. Similarity based approaches include similarity searches between viral and microbial genomes [131,132]; linking viral genomes and their host via clustered regularly interspaced short palindromic repeats (CRISPR) spacers [133]; and comparing viral and host genome nucleotide signatures [130]. These methods were applied to the GOV, which led to host-range predictions for 392 VCs [125].

Finally, the prevalence of AMGs in the sequence data was investigated, and 243 putative AMGs were identified, of which only 95 were previously known [123]. These AMGs included genes with reputed roles in sulphur and nitrogen cycling, with analysis revealing many genes vital to these pathways in epipelagic viruses [125].

The POV and GOV datasets are not definitive representatives of the marine virome due to the limitations and inherent biases of current viral metagenomic techniques and equipment. The above-mentioned studies were dominated by the examination of dsDNA viruses, thus leaving ssDNA viruses underrepresented, an issue which is currently one of the most pressing areas of concern in the field of viral metagenomics [70,134]. However, the POV, GOV, and many other studies of the marine virome have unquestionably demonstrated the potential of viral metagenome analysis of environmental samples, and exponentially expanded our knowledge of viral communities in the oceans while providing critical reference frameworks for future studies.

5.2. Human Viral Metagenomics

Human metagenome studies have assessed the complex microbial communities associated with the mouth, gastrointestinal tract, lungs, and skin, among other areas. Additionally, human studies have attempted to define correlations between microbial composition and the subject's age, health status/disease states and lifestyle. Among these, the human gut is undoubtedly the most extensively characterised and was thoroughly reviewed recently [135]. While such studies provide an array of data pertaining to the bacterial landscapes in these niches, it is imperative to simultaneously assess the role of phages in shaping these landscapes. In view of this, a plethora of recent virome studies of a variety of human surfaces/organs and the impact of different demographic and environmental conditions have been undertaken and will be discussed herein.

5.2.1. Virome of the Skin and Oral Cavity

The epidermidis of human skin serves as the primary protective barrier to the external environment and the surface and natural crevices of the skin are host to a variety of microbes. A recent study highlighted that, consistent with other studies, the total metagenome of skin in healthy individuals exhibits temporal stability, while the virome is subject to much greater variability [136]. Bacterial species including *Propionibacterium*, *Staphylococcus*, and *Corynebacterium* were identified in the subjects' skin in this study, while the viruses recovered through virome analysis of isolated VLPs were dominantly either unclassified or identified as *Staphylococcus* phages. The presence of high levels of seemingly novel phage sequences in this niche exemplifies the high degree of viral dark matter that is present and highlights that current knowledge on the skin virome represents merely the 'tip of the iceberg' of the genetic diversity of phages in the biosphere. The majority of phages recovered from

this environment are dsDNA viruses, and this apparent dominance is likely owing to the temperate nature of the majority of these isolates. Most metagenome studies employ Illumina-based sequencing technologies as they produce high sequence coverage, and they are relatively inexpensive. However, the application of PacBio SMRT (single molecule real time) sequencing was recently applied to produce 'finished quality' genomes of skin metagenome samples and resulted in the identification of a previously uncharacterised *Corynebacterium simulans* phage–host genome combination [137]. To assess the potential diversity of the strains of this species in the sample, HiSeq datasets were simultaneously generated and revealed a dominant strain within the population. Therefore, while high throughput sequencing technologies are the method of choice, long-read technologies/hybrid approaches may be useful in the evaluation and reconstruction of complex microbial and viral communities.

It is reported that the human oral cavity is host to six billion bacteria and up to 35 times as many phages [138]. In congruence with the findings in the skin environment, the viromes of the oral cavity are suggested to primarily contain temperate phages, with 10% of viral reads possessing integrase homologues in a study of the saliva of five healthy subjects [139]. Viral reads with homology to viruses of *Veillonella*, *Streptococcus*, and *Megasphaera* were identified in this study, and, in addition to the identification of lysogeny-related functions, virulence factors were also abundant, which are asserted to present a reservoir of pathogenic potential/conversion in the human oral environment. It is predicted that the viral communities in the oral cavity are persistent, which is likely a reflection of the stability of host bacterial communities in this environment [138]. Furthermore, phages that exist in such complex communities encounter significant competition, which places them under a high level of evolutionary pressure to adapt to their host if the latter is changing and to ensure the success of the host so as to secure their own persistence.

5.2.2. Virome of the Lungs

The human lung microbiome has been the focus of several recent studies with implications for understanding and advancing treatment of a variety of pulmonary (and other) diseases. The relationship between the microbiome of the upper and lower respiratory airways is unclear and represents an area of growing interest in the research community. Furthermore, the impact of diseases such as cystic fibrosis (CF) and human immunodeficiency virus (HIV) infection on the lung microbiome has recently received justified research scrutiny [140–142]. In contrast to the array of studies on the lung microbiome, relatively few studies have looked at the lung virome, although the value of such studies has recently been highlighted [143]. Therefore, it is likely that an increasing number of lung virome studies will be undertaken in the near future. Viral metagenomic analysis of samples from the respiratory tract of lung transplant patients identified an increase in the relative abundance and complex populations of human-associated anelloviruses, which are small circular, non-enveloped, ssDNA viruses [144]. In a recent study of the metagenome and virome of the respiratory tracts of CF patients, viral reads displaying sequence relatedness to phages infecting typical CF pathogens including *Streptococcus*, *Burkholderia*, *Mycobacterium*, *Enterobacteria*, and *Pseudomonas* were identified [145]. Surprisingly, the phage profiles between distinct patients were quite similar, while, in contrast, greater diversity was observed among the microbial communities. This suggests that the recovered phage sequences had multiplied on persistent host bacteria and were thus stable, while, in contrast, transient microbial populations were less conserved and their infecting phages (if present) would presumably be less abundant than the persistent colonisers in CF patients. This finding is remarkable among virome analysis studies and highlights the condition/environment-specific nature of viromes, particularly when considering disease states.

5.2.3. Virome of the Human Gut

The microbiota of the human intestinal tract is undeniably one of the most intensely studied microbiomes and is estimated to be constituted by up to 10^{14} bacterial cells [146]. Studies assessing the effect of age [147–149], geographical location [150], diet [151,152], and health status [153–155],

among other factors, on the human gut microbiota have been undertaken. Similar to other fields of metagenomics, there is an increasing awareness that it is essential to evaluate the presence and potential impact of (pro)phages on the microbiome; therefore, there is a growing number of studies in this area. Recently, the prophage-related sequences identified on the genomes of 38 bifidobacterial strains were extracted and used to identify the presence of related sequences in the metagenomic data of infants [155]. *Bifidobacterium* sequences were identified in the data relating to 20 of the 173 infants, and (pro)phage sequences largely with *Siphoviridae* genomic organisation were additionally identified with relatedness to phages of *Actinobacteria*, *Firmicutes*, and Gram-negative bacteria. The virome of the human gut is influenced by the health status of the individual, and patients with Crohn's disease and ulcerative colitis were shown to exhibit 'abnormal' viromes, with a much greater level of phages belonging to the *Caudovirales* family [156]. The viromes were also condition-specific, highlighting the need for data pertaining to individual disease states. Furthermore, while the gut virome of individuals over time may not change dramatically in most cases, it is highly individual-specific [157,158]. The individuality and condition-specific nature of the gut virome and its sensitivity to environmental factors consolidate the requirement for extensive and representative sampling to provide robust data.

One of the most intriguing findings from the many human gut virome studies was the identification of the so-called 'crAssphage' [159]. This phage and its related sequences are highly abundant among the viromes of all human faecal metagenome datasets [160]. This phage, which was until very recently unknown to exist, represents just one example of the necessity for the examination of what currently constitutes viral dark matter and highlights the possibility of defining and understanding the vast array of phage–host interactions that may exist and influence our gut microbial landscapes.

5.3. Potential Applications

Microbial metagenomics has long been recognised as possessing the capability to significantly influence industrial biotechnology [161] from functional screening of metagenomic sequences for novel enzymes [162] to increasing our understanding of the dynamics of various biological processes such as food fermentations, thus facilitating improved efficiency, product quality, and profitability [63]. The field of viral metagenomics possesses similar potential, although its applications to this end remain understudied and very much in their infancy.

Novel enzyme discovery is chief among these applications. Bacteriophages have long been a rich source of enzymes [7], but the majority of the most commonly used enzymes in laboratories are still derived from a handful of cultivable phages such as T4, T7, λ, and Φ29 [163]. These phages have yielded crucial enzymes such as T4 DNA ligase, utilized in virtually all laboratory ligations during cloning, and numerous DNA polymerases [164]. However, considering the abundance of phage sequences that continue to be isolated either via culture-dependent or metagenomic methods, as well as the genetic diversity and enzyme-richness of these sequences, it is clear that the functional potential of bacteriophages remains greatly under-exploited. Functional viral metagenomics has the potential to rectify this, facilitating the discovery of technologically useful enzymes. For example, a single 2 Mb bacterial genome possesses a single DNA polymerase 1 gene, while up to 20–40 *pol1* genes can be found in 2 Mb of viral metagenomic sequence [163]. Aside from the identification of suitable hosts for the expression of potential enzymes, however, the primary challenge preventing the realisation of this potential is the difficulty in annotating viral metagenomes [164]. Nonetheless, methods are continually improving, and recent studies have identified numerous thermostable DNA polymerases from viral communities in hot springs in Yellowstone National Park (YNP), California, and Nevada [164,165].

The discovery of novel enzymes is not limited to replication-related genes. Bacteriophage lysins are highly evolved enzymes produced by phages which cleave the bacterial cell wall during the final stages of the lytic cycle to facilitate the release of phage particles. Due to their function, attention has turned to these enzymes and their potential use as novel antimicrobials [166], including their application as food preservatives and as therapeutic agents against human pathogens [167], where the host-specific activity of phage lysins prevents non-target negative effects in addition to circumventing

antibiotic resistance. As a result, attention has turned towards the application of functional viral metagenomics to the discovery of novel phage lysins [168].

These examples offer a snapshot of the potential of functional viral metagenomics to serve as a platform to unlock the wealth of useful enzymes that is undoubtedly present in the vast viral sequence space. Indeed, as the annotation of viral sequences continues to improve, the discovery of novel enzymes will increase considerably.

6. Conclusions and Future Perspectives

The continual improvement of technology and techniques to minimise the introduction of biases and the skewing of produced population structures is the primary challenge facing the field of viral metagenomics. The challenges remain many and varied, but as the methods approach a level of quantitative rigour capable of producing faithful representations of environmental viral communities, viral metagenomics can transition from a tool of observation and description to a means of prediction and application. These advances will also increase confidence in the validity of viral genomes identified purely through metagenomic sequencing, leading to the acceptance of these sequences as bona fide viruses and their inclusion in formal ICTV (International Committee on Viral Taxonomy) viral taxonomy, a process about which discussion has already begun [169]. The identification of the widespread existence and abundance of crAssphage in the human gut indicates the existence of previously unknown and uncharacterised viral entities and highlights the wealth of undiscovered data that may exist. Viral metagenomics is poised to vastly increase our knowledge of viral dark matter and to further elucidate the fundamental role viruses play in every aspect of the biosphere.

Acknowledgments: S. Hayes is the recipient of an Irish Research Council Enterprise Partnership Scheme postgraduate scholarship; J. Mahony is the recipient of a Starting Investigator Research Grant funded by Science Foundation Ireland (SFI) (Ref. No. 15/SIRG/3430); D. van Sinderen is the recipient of an SFI Investigator award (Ref. No.13/IA/1953).

Author Contributions: J.M. and D.v.S. were involved in the design and layout of the review; S.H. and J.M. prepared the manuscript; and A.N. and D.v.S. were involved in reviewing and editing the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rohwer, F. Global phage diversity. *Cell* **2003**, *113*, 141. [CrossRef]
2. Suttle, C.A. Viruses in the sea. *Nature* **2005**, *437*, 356–361. [CrossRef] [PubMed]
3. Bergh, Ø.; Børsheim, K.Y.; Bratbak, G.; Heldal, M. High abundance of viruses found in aquatic environments. *Nature* **1989**, *340*, 467–468. [CrossRef] [PubMed]
4. Hoyles, L.; McCartney, A.L.; Neve, H.; Gibson, G.R.; Sanderson, J.D.; Heller, K.J.; van Sinderen, D. Characterization of virus-like particles associated with the human faecal and caecal microbiota. *Res. Microbiol.* **2014**, *165*, 803–812. [CrossRef] [PubMed]
5. Kim, M.-S.; Park, E.-J.; Roh, S.W.; Bae, J.-W. Diversity and abundance of single-stranded DNA viruses in human feces. *Appl. Environ. Microbiol.* **2011**, *77*, 8062–8070. [CrossRef] [PubMed]
6. Clokie, M.R.; Millard, A.D.; Letarov, A.V.; Heaphy, S. Phages in nature. *Bacteriophage* **2011**, *1*, 31–45. [CrossRef] [PubMed]
7. McGrath, S.; Fitzgerald, G.F.; van Sinderen, D. The impact of bacteriophage genomics. *Curr. Opin. Biotechnol.* **2004**, *15*, 94–99. [CrossRef] [PubMed]
8. Waldor, M.K.; Mekalanos, J.J. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **1996**, *272*, 1910–1914. [CrossRef] [PubMed]
9. Waldor, M.K.; Friedman, D.I. Phage regulatory circuits and virulence gene expression. *Curr. Opin. Microbiol.* **2005**, *8*, 459–465. [CrossRef] [PubMed]
10. Brüssow, H.; Hendrix, R.W. Phage genomics: Small is beautiful. *Cell* **2002**, *108*, 13–16. [CrossRef]
11. Gómez, P.; Buckling, A. Bacteria-phage antagonistic coevolution in soil. *Science* **2011**, *332*, 106–109. [CrossRef] [PubMed]

12. Pal, C.; Maciá, M.D.; Oliver, A.; Schachar, I.; Buckling, A. Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature* **2007**, *450*, 1079–1081. [CrossRef] [PubMed]
13. Canchaya, C.; Fournous, G.; Chibani-Chennoufi, S.; Dillmann, M.-L.; Brüssow, H. Phage as agents of lateral gene transfer. *Curr. Opin. Microbiol.* **2003**, *6*, 417–424. [CrossRef]
14. Labrie, S.J.; Samson, J.E.; Moineau, S. Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* **2010**, *8*, 317–327. [CrossRef] [PubMed]
15. Rodriguez-Valera, F.; Martin-Cuadrado, A.-B.; Rodriguez-Brito, B.; Pašić, L.; Thingstad, T.F.; Rohwer, F.; Mira, A. Explaining microbial population genomics through phage predation. *Nat. Rev. Microbiol.* **2009**, *7*, 828–836. [CrossRef] [PubMed]
16. Suttle, C.A. Marine viruses—Major players in the global ecosystem. *Nat. Rev. Microbiol.* **2007**, *5*, 801–812. [CrossRef] [PubMed]
17. Brum, J.R.; Morris, J.J.; Décima, M.; Stukel, M.R. Mortality in the oceans: Causes and consequences. *Assoc. Sci. Limnol. Oceanogr.* **2014**, 16–48. [CrossRef]
18. Rohwer, F.; Youle, M. Consider something viral in your research. *Nat. Rev. Microbiol.* **2011**, *9*, 308–309. [CrossRef]
19. Paez-Espino, D.; Eloe-Fadrosh, E.A.; Pavlopoulos, G.A.; Thomas, A.D.; Huntemann, M.; Mikhailova, N.; Rubin, E.; Ivanova, N.N.; Kyrpides, N.C. Uncovering Earth's virome. *Nature* **2016**, *536*, 425–430. [CrossRef] [PubMed]
20. Mokili, J.L.; Rohwer, F.; Dutilh, B.E. Metagenomics and future perspectives in virus discovery. *Curr. Opin. Virol.* **2012**, *2*, 63–77. [CrossRef] [PubMed]
21. Amann, R.I.; Ludwig, W.; Schleifer, K.-H. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* **1995**, *59*, 143–169. [PubMed]
22. Connon, S.A.; Giovannoni, S.J. High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Appl. Environ. Microbiol.* **2002**, *68*, 3878–3885. [CrossRef] [PubMed]
23. Rappe, M.S.; Giovannoni, S.J. The uncultured microbial majority. *Annu. Rev. Microbiol.* **2003**, *57*, 369–394. [CrossRef] [PubMed]
24. Duhaime, M.B.; Sullivan, M.B. Ocean viruses: Rigorously evaluating the metagenomic sample-to-sequence pipeline. *Virology* **2012**, *434*, 181–186. [CrossRef] [PubMed]
25. Willner, D.; Hugenholtz, P. From deep sequencing to viral tagging: Recent advances in viral metagenomics. *Bioessays* **2013**, *35*, 436–442. [CrossRef] [PubMed]
26. Chen, J.; Novick, R.P. Phage-mediated intergeneric transfer of toxin genes. *Science* **2009**, *323*, 139–141. [CrossRef] [PubMed]
27. Łoś, M.; Węgrzyn, G. Pseudolysogeny. *Adv. Virus Res.* **2011**, *82*, 339–349.
28. Bryan, D.; El-Shibiny, A.; Hobbs, Z.; Porter, J.; Kutter, E.M. Bacteriophage T4 Infection of Stationary Phase *E. coli*: Life after Log from a Phage Perspective. *Front. Microbiol.* **2016**, *7*, 1391. [CrossRef] [PubMed]
29. Chibani-Chennoufi, S.; Bruttin, A.; Dillmann, M.-L.; Brüssow, H. Phage-host interaction: An ecological perspective. *J. Bacteriol.* **2004**, *186*, 3677–3686. [CrossRef] [PubMed]
30. Barns, S.M.; Fundyga, R.E.; Jeffries, M.W.; Pace, N.R. Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 1609–1613. [CrossRef] [PubMed]
31. Hugenholtz, P.; Pace, N.R. Identifying microbial diversity in the natural environment: A molecular phylogenetic approach. *Trends Biotechnol.* **1996**, *14*, 190–197. [CrossRef]
32. Olsen, G.J.; Lane, D.J.; Giovannoni, S.J.; Pace, N.R.; Stahl, D.A. Microbial ecology and evolution: A ribosomal RNA approach. *Annu. Rev. Microbiol.* **1986**, *40*, 337–365. [CrossRef] [PubMed]
33. Liu, W.-T.; Marsh, T.L.; Cheng, H.; Forney, L.J. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl. Environ. Microbiol.* **1997**, *63*, 4516–4522. [PubMed]
34. Giovannoni, S.J.; Mullins, T.D.; Field, K.G. Microbial diversity in oceanic systems: rRNA approaches to the study of unculturable microbes. In *Molecular Ecology of Aquatic Microbes*; Springer: Berlin/Heidelberg, Germany, 1995; pp. 217–248.
35. Schmidt, T.M.; DeLong, E.; Pace, N. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J. Bacteriol.* **1991**, *173*, 4371–4378. [CrossRef] [PubMed]
36. Edwards, R.A.; Rohwer, F. Viral metagenomics. *Nat. Rev. Microbiol.* **2005**, *3*, 504–510. [CrossRef] [PubMed]

37. Winget, D.M.; Wommack, K.E. Randomly amplified polymorphic DNA PCR as a tool for assessment of marine viral richness. *Appl. Environ. Microbiol.* **2008**, *74*, 2612–2618. [CrossRef] [PubMed]
38. Brussaard, C.P. Optimization of procedures for counting viruses by flow cytometry. *Appl. Environ. Microbiol.* **2004**, *70*, 1506–1513. [CrossRef] [PubMed]
39. Børsheim, K.; Bratbak, G.; Heldal, M. Enumeration and biomass estimation of planktonic bacteria and viruses by transmission electron microscopy. *Appl. Environ. Microbiol.* **1990**, *56*, 352–356. [PubMed]
40. Bratbak, G.; Heldal, M. Total count of viruses in aquatic environments. In *Handbook of Methods in Aquatic Microbial Ecology*; Lewis Publishers: Boca Raton, FL, USA, 1993; pp. 135–138.
41. Allen, L.Z.; Ishoey, T.; Novotny, M.A.; McLean, J.S.; Lasken, R.S.; Williamson, S.J. Single virus genomics: A new tool for virus discovery. *PLoS ONE* **2011**, *6*, e17722. [CrossRef] [PubMed]
42. Deng, L.; Gregory, A.; Yilmaz, S.; Poulos, B.T.; Hugenholtz, P.; Sullivan, M.B. Contrasting life strategies of viruses that infect photo-and heterotrophic bacteria, as revealed by viral tagging. *MBio* **2012**, *3*, e00373-12. [CrossRef] [PubMed]
43. Breitbart, M.; Miyake, J.H.; Rohwer, F. Global distribution of nearly identical phage-encoded DNA sequences. *FEMS Microbiol. Lett.* **2004**, *236*, 249–256. [CrossRef] [PubMed]
44. Sullivan, M.B.; Lindell, D.; Lee, J.A.; Thompson, L.R.; Bielawski, J.P.; Chisholm, S.W. Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLoS Biol.* **2006**, *4*, e234. [CrossRef] [PubMed]
45. Sharon, I.; Tzahor, S.; Williamson, S.; Shmoish, M.; Man-Aharonovich, D.; Rusch, D.B.; Yooseph, S.; Zeidner, G.; Golden, S.S.; Mackey, S.R.; et al. Viral photosynthetic reaction center genes and transcripts in the marine environment. *ISME J.* **2007**, *1*, 492–501. [CrossRef] [PubMed]
46. Chenard, C.; Suttle, C. Phylogenetic diversity of sequences of cyanophage photosynthetic gene *psbA* in marine and freshwaters. *Appl. Environ. Microbiol.* **2008**, *74*, 5317–5324. [CrossRef] [PubMed]
47. Comeau, A.M.; Krisch, H.M. The capsid of the T4 phage superfamily: The evolution, diversity, and structure of some of the most prevalent proteins in the biosphere. *Mol. Biol. Evol.* **2008**, *25*, 1321–1332. [CrossRef] [PubMed]
48. Sullivan, M.B. Viromes, not gene markers, for studying double-stranded DNA virus communities. *J. Virol.* **2015**, *89*, 2459–2461. [CrossRef] [PubMed]
49. Hadrys, H.; Balick, M.; Schierwater, B. Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. *Mol. Ecol.* **1992**, *1*, 55–63. [CrossRef] [PubMed]
50. Hara, S.; Terauchi, K.; Koike, I. Abundance of viruses in marine waters: Assessment by epifluorescence and transmission electron microscopy. *Appl. Environ. Microbiol.* **1991**, *57*, 2731–2734. [PubMed]
51. Weinbauer, M.; Suttle, C. Comparison of epifluorescence and transmission electron microscopy for counting viruses in natural marine waters. *Aquat. Microb. Ecol.* **1997**, *13*, 225–232. [CrossRef]
52. Noble, R.T.; Fuhrman, J.A. Use of SYBR Green I for rapid epifluorescence counts of marine viruses and bacteria. *Aquat. Microb. Ecol.* **1998**, *14*, 113–118. [CrossRef]
53. Wen, K.; Ortmann, A.C.; Suttle, C.A. Accurate estimation of viral abundance by epifluorescence microscopy. *Appl. Environ. Microbiol.* **2004**, *70*, 3862–3867. [CrossRef] [PubMed]
54. Marie, D.; Brussaard, C.P.; Thyrhaug, R.; Bratbak, G.; Vaulot, D. Enumeration of marine viruses in culture and natural samples by flow cytometry. *Appl. Environ. Microbiol.* **1999**, *65*, 45–52. [PubMed]
55. Brussaard, C.P.; Marie, D.; Bratbak, G. Flow cytometric detection of viruses. *J. Virol. Methods* **2000**, *85*, 175–182. [CrossRef]
56. Ohno, S.; Okano, H.; Tanji, Y.; Ohashi, A.; Watanabe, K.; Takai, K.; Imachi, H. A method for evaluating the host range of bacteriophages using phages fluorescently labeled with 5-ethynyl-2′-deoxyuridine (EdU). *Appl. Microbiol. Biotechnol.* **2012**, *95*, 777–788. [CrossRef] [PubMed]
57. Brum, J.R.; Sullivan, M.B. Rising to the challenge: Accelerated pace of discovery transforms marine virology. *Nat. Rev. Microbiol.* **2015**, *13*, 147–159. [CrossRef] [PubMed]
58. Handelsman, J.; Rondon, M.R.; Brady, S.F.; Clardy, J.; Goodman, R.M. Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. *Chem. Biol.* **1998**, *5*, R245–R249. [CrossRef]
59. Thomas, T.; Gilbert, J.; Meyer, F. Metagenomics-a guide from sampling to data analysis. *Microb. Inform. Exp.* **2012**, *2*, 3. [CrossRef] [PubMed]

60. Jovel, J.; Patterson, J.; Wang, W.; Hotte, N.; O'Keefe, S.; Mitchel, T.; Perry, T.; Kao, D.; Mason, A.L.; Madsen, K.L.; et al. Characterization of the gut microbiome using 16S or shotgun metagenomics. *Front. Microbiol.* **2016**, *7*, 459. [CrossRef] [PubMed]
61. Handelsman, J. Metagenetics: Spending our inheritance on the future. *Microb. Biotechnol.* **2009**, *2*, 138–139. [CrossRef] [PubMed]
62. Oulas, A.; Pavloudi, C.; Polymenakou, P.; Pavlopoulos, G.A.; Papanikolaou, N.; Kotoulas, G.; Arvanitidis, C.; Iliopoulos, I. Metagenomics: Tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. *Bioinform. Biol. Insights* **2015**, *9*, 75–88. [PubMed]
63. De Filippis, F.; Parente, E.; Ercolini, D. Metagenomics insights into food fermentations. *Microb. Biotechnol.* **2016**, *10*, 91–102. [CrossRef] [PubMed]
64. Markowitz, V.M.; Chen, I.-M.A.; Palaniappan, K.; Chu, K.; Szeto, E.; Grechkin, Y.; Ratner, A.; Jacob, B.; Huang, J.; Williams, P.; et al. IMG: The integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* **2012**, *40*, D115–D122. [CrossRef] [PubMed]
65. Culligan, E.P.; Sleator, R.D.; Marchesi, J.R.; Hill, C. Metagenomics and novel gene discovery: Promise and potential for novel therapeutics. *Virulence* **2014**, *5*, 399–412. [CrossRef] [PubMed]
66. Thies, S.; Rausch, S.C.; Kovacic, F.; Schmidt-Thaler, A.; Wilhelm, S.; Rosenau, F.; Daniel, R.; Streit, W.; Pietruszka, J.; Jaeger, K.-E. Metagenomic discovery of novel enzymes and biosurfactants in a slaughterhouse biofilm microbial community. *Sci. Rep.* **2016**, *6*, 27035. [CrossRef] [PubMed]
67. Uchiyama, T.; Miyazaki, K. Functional metagenomics for enzyme discovery: Challenges to efficient screening. *Curr. Opin. Biotechnol.* **2009**, *20*, 616–622. [CrossRef] [PubMed]
68. Handelsman, J. Metagenomics: Application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 669–685. [CrossRef] [PubMed]
69. Roux, S.; Krupovic, M.; Debroas, D.; Forterre, P.; Enault, F. Assessment of viral community functional potential from viral metagenomes may be hampered by contamination with cellular sequences. *Open Biol.* **2013**, *3*, 130160. [CrossRef] [PubMed]
70. Labonté, J.M.; Suttle, C.A. Previously unknown and highly divergent ssDNA viruses populate the oceans. *ISME J.* **2013**, *7*, 2169–2177. [CrossRef] [PubMed]
71. Van Regenmortel, M.H.; Fauquet, C.M.; Bishop, D.H.; Carstens, E.; Estes, M.; Lemon, S.; Maniloff, J.; Mayo, M.; McGeoch, D.; Pringle, C. Virus taxonomy: Classification and nomenclature of viruses. In *Seventh Report of the International Committee on Taxonomy of Viruses*; Academic Press: Cambridge, MA, USA, 2000.
72. Kleiner, M.; Hooper, L.V.; Duerkop, B.A. Evaluation of methods to purify virus-like particles for metagenomic sequencing of intestinal viromes. *BMC Genom.* **2015**, *16*, 7. [CrossRef] [PubMed]
73. Breitbart, M.; Salamon, P.; Andresen, B.; Mahaffy, J.M.; Segall, A.M.; Mead, D.; Azam, F.; Rohwer, F. Genomic analysis of uncultured marine viral communities. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14250–14255. [CrossRef] [PubMed]
74. Castro-Mejía, J.L.; Muhammed, M.K.; Kot, W.; Neve, H.; Franz, C.M.; Hansen, L.H.; Vogensen, F.K.; Nielsen, D.S. Optimizing protocols for extraction of bacteriophages prior to metagenomic analyses of phage communities in the human gut. *Microbiome* **2015**, *3*, 64. [CrossRef] [PubMed]
75. Thurber, R.V.; Haynes, M.; Breitbart, M.; Wegley, L.; Rohwer, F. Laboratory procedures to generate viral metagenomes. *Nat. Protoc.* **2009**, *4*, 470–483. [CrossRef] [PubMed]
76. Iker, B.C.; Bright, K.R.; Pepper, I.L.; Gerba, C.P.; Kitajima, M. Evaluation of commercial kits for the extraction and purification of viral nucleic acids from environmental and fecal samples. *J. Virol. Methods* **2013**, *191*, 24–30. [CrossRef] [PubMed]
77. Hall, R.J.; Wang, J.; Todd, A.K.; Bissielo, A.B.; Yen, S.; Strydom, H.; Moore, N.E.; Ren, X.; Huang, Q.S.; Carter, P.E.; et al. Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. *J. Virol. Methods* **2014**, *195*, 194–204. [CrossRef] [PubMed]
78. Pelzek, A.J.; Schuch, R.; Schmitz, J.E.; Fischetti, V.A. Isolation of bacteriophages from environmental sources, and creation and functional screening of phage DNA libraries. *Curr. Protoc. Essent. Lab. Tech.* **2013**, *7*, 13.3.1–13.3.35.
79. John, S.G.; Mendez, C.B.; Deng, L.; Poulos, B.; Kauffman, A.K.M.; Kern, S.; Brum, J.; Polz, M.F.; Boyle, E.A.; Sullivan, M.B. A simple and efficient method for concentration of ocean viruses by chemical flocculation. *Environ. Microbiol. Rep.* **2011**, *3*, 195–202. [CrossRef] [PubMed]

80. Hurwitz, B.L.; Deng, L.; Poulos, B.T.; Sullivan, M.B. Evaluation of methods to concentrate and purify ocean virus communities through comparative, replicated metagenomics. *Environ. Microbiol.* **2013**, *15*, 1428–1440. [CrossRef] [PubMed]
81. Wesolowska-Andersen, A.; Bahl, M.I.; Carvalho, V.; Kristiansen, K.; Sicheritz-Pontén, T.; Gupta, R.; Licht, T.R. Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis. *Microbiome* **2014**, *2*, 19. [CrossRef] [PubMed]
82. Sachsenröder, J.; Twardziok, S.; Hammerl, J.A.; Janczyk, P.; Wrede, P.; Hertwig, S.; Johne, R. Simultaneous identification of DNA and RNA viruses present in pig faeces using process-controlled deep sequencing. *PLoS ONE* **2012**, *7*, e34631. [CrossRef] [PubMed]
83. Salter, S.J.; Cox, M.J.; Turek, E.M.; Calus, S.T.; Cookson, W.O.; Moffatt, M.F.; Turner, P.; Parkhill, J.; Loman, N.J.; Walker, A.W. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* **2014**, *12*, 87. [CrossRef] [PubMed]
84. Hurwitz, B.L.; U'ren, J.M.; Youens-Clark, K. Computational prospecting the great viral unknown. *FEMS Microbiol. Lett.* **2016**, *363*, fnw077. [CrossRef] [PubMed]
85. Solonenko, S.A.; Sullivan, M.B. Preparation of metagenomic libraries from naturally occurring marine viruses. *Methods Enzyml.* **2013**, *531*, 143–160.
86. Kim, K.-H.; Bae, J.-W. Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. *Appl. Environ. Microbiol.* **2011**, *77*, 7663–7668. [CrossRef] [PubMed]
87. Woyke, T.; Xie, G.; Copeland, A.; Gonzalez, J.M.; Han, C.; Kiss, H.; Saw, J.H.; Senin, P.; Yang, C.; Chatterji, S.; et al. Assembling the marine metagenome, one cell at a time. *PLoS ONE* **2009**, *4*, e5299. [CrossRef] [PubMed]
88. Yilmaz, S.; Allgaier, M.; Hugenholtz, P. Multiple displacement amplification compromises quantitative analysis of metagenomes. *Nat. Methods* **2010**, *7*, 943–944. [CrossRef] [PubMed]
89. Bikel, S.; Valdez-Lara, A.; Cornejo-Granados, F.; Rico, K.; Canizales-Quinteros, S.; Soberón, X.; Del Pozo-Yauner, L.; Ochoa-Leyva, A. Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: Towards a systems-level understanding of human microbiome. *Comput. Struct. Biotechnol. J.* **2015**, *13*, 390–401. [CrossRef] [PubMed]
90. Hoeijmakers, W.A.; Bártfai, R.; Françoijs, K.-J.; Stunnenberg, H.G. Linear amplification for deep sequencing. *Nat. Protoc.* **2011**, *6*, 1026–1036. [CrossRef] [PubMed]
91. Marine, R.; Polson, S.W.; Ravel, J.; Hatfull, G.; Russell, D.; Sullivan, M.; Syed, F.; Dumas, M.; Wommack, K.E. Evaluation of a transposase protocol for rapid generation of shotgun high-throughput sequencing libraries from nanogram quantities of DNA. *Appl. Environ. Microbiol.* **2011**, *77*, 8071–8079. [CrossRef] [PubMed]
92. Chafee, M.; Maignien, L.; Simmons, S.L. The effects of variable sample biomass on comparative metagenomics. *Environ. Microbiol.* **2015**, *17*, 2239–2253. [CrossRef] [PubMed]
93. Henn, M.R.; Sullivan, M.B.; Stange-Thomann, N.; Osburne, M.S.; Berlin, A.M.; Kelly, L.; Yandava, C.; Kodira, C.; Zeng, Q.; Weiand, M.; et al. Analysis of high-throughput sequencing and annotation strategies for phage genomes. *PLoS ONE* **2010**, *5*, e9083. [CrossRef] [PubMed]
94. Duhaime, M.B.; Deng, L.; Poulos, B.T.; Sullivan, M.B. Towards quantitative metagenomics of wild viruses and other ultra-low concentration DNA samples: A rigorous assessment and optimization of the linker amplification method. *Environ. Microbiol.* **2012**, *14*, 2526–2537. [CrossRef] [PubMed]
95. Dean, F.B.; Hosono, S.; Fang, L.; Wu, X.; Faruqi, A.F.; Bray-Ward, P.; Sun, Z.; Zong, Q.; Du, Y.; Du, J.; et al. Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 5261–5266. [CrossRef] [PubMed]
96. Zong, C.; Lu, S.; Chapman, A.R.; Xie, X.S. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* **2012**, *338*, 1622–1626. [CrossRef] [PubMed]
97. Bowers, R.M.; Clum, A.; Tice, H.; Lim, J.; Singh, K.; Ciobanu, D.; Ngan, C.Y.; Cheng, J.-F.; Tringe, S.G.; Woyke, T. Impact of library preparation protocols and template quantity on the metagenomic reconstruction of a mock microbial community. *BMC Genom.* **2015**, *16*, 856. [CrossRef] [PubMed]
98. Kelleher, P.; Murphy, J.; Mahony, J.; Van Sinderen, D. Next-generation sequencing as an approach to dairy starter selection. *Dairy Sci. Technol.* **2015**, *95*, 545–568. [CrossRef] [PubMed]
99. Bibby, K. Improved bacteriophage genome data is necessary for integrating viral and bacterial ecology. *Microb. Ecol.* **2014**, *67*, 242–244. [CrossRef] [PubMed]

100. Glass, E.M.; Wilkening, J.; Wilke, A.; Antonopoulos, D.; Meyer, F. Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harb. Protoc.* **2010**, *2010*, pdb.prot5368. [CrossRef] [PubMed]
101. Liu, B.; Gibbons, T.; Ghodsi, M.; Treangen, T.; Pop, M. Accurate and fast estimation of taxonomic profiles from metagenomic shotgun sequences. *BMC Genom.* **2011**, *12*, S4. [CrossRef] [PubMed]
102. Gerlach, W.; Jünemann, S.; Tille, F.; Goesmann, A.; Stoye, J. WebCARMA: A web application for the functional and taxonomic classification of unassembled metagenomic reads. *BMC Bioinform.* **2009**, *10*, 430. [CrossRef] [PubMed]
103. Hurwitz, B.L.; Westveld, A.H.; Brum, J.R.; Sullivan, M.B. Modeling ecological drivers in marine viral communities using comparative metagenomics and network analyses. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 10714–10719. [CrossRef] [PubMed]
104. Ounit, R.; Wanamaker, S.; Close, T.J.; Lonardi, S. CLARK: Fast and accurate classification of metagenomic and genomic sequences using discriminative *k*-mers. *BMC Genom.* **2015**, *16*, 236. [CrossRef] [PubMed]
105. Edgar, R.C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **2010**, *26*, 2460–2461. [CrossRef] [PubMed]
106. Hurwitz, B.L.; Sullivan, M.B. The Pacific Ocean Virome (POV): A marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. *PLoS ONE* **2013**, *8*, e57355. [CrossRef] [PubMed]
107. Wommack, K.E.; Bhavsar, J.; Polson, S.W.; Chen, J.; Dumas, M.; Srinivasiah, S.; Furman, M.; Jamindar, S.; Nasko, D.J. VIROME: A standard operating procedure for analysis of viral metagenome sequences. *Stand. Genom. Sci.* **2012**, *6*, 427–439. [CrossRef] [PubMed]
108. Roux, S.; Tournayre, J.; Mahul, A.; Debroas, D.; Enault, F. Metavir 2: New tools for viral metagenome comparison and assembled virome analysis. *BMC Bioinform.* **2014**, *15*, 76. [CrossRef] [PubMed]
109. Lorenzi, H.A.; Hoover, J.; Inman, J.; Safford, T.; Murphy, S.; Kagan, L.; Williamson, S.J. The Viral MetaGenome Annotation Pipeline (VMGAP): An automated tool for the functional annotation of viral Metagenomic shotgun sequencing data. *Stand. Genom. Sci.* **2011**, *4*, 418–429. [CrossRef] [PubMed]
110. Tangherlini, M.; Dell'Anno, A.; Allen, L.Z.; Riccioni, G.; Corinaldesi, C. Assessing viral taxonomic composition in benthic marine ecosystems: Reliability and efficiency of different bioinformatic tools for viral metagenomic analyses. *Sci. Rep.* **2016**, *6*, 28428. [CrossRef] [PubMed]
111. Reyes, A.; Semenkovich, N.P.; Whiteson, K.; Rohwer, F.; Gordon, J.I. Going viral: Next-generation sequencing applied to phage populations in the human gut. *Nat. Rev. Microbiol.* **2012**, *10*, 607–617. [CrossRef] [PubMed]
112. Wommack, K.E.; Nasko, D.J.; Chopyk, J.; Sakowski, E.G. Counts and sequences, observations that continue to change our understanding of viruses in nature. *J. Microbiol.* **2015**, *53*, 181–192. [CrossRef] [PubMed]
113. Angly, F.E.; Felts, B.; Breitbart, M.; Salamon, P.; Edwards, R.A.; Carlson, C.; Chan, A.M.; Haynes, M.; Kelley, S.; Liu, H.; et al. The marine viromes of four oceanic regions. *PLoS Biol.* **2006**, *4*, e368. [CrossRef] [PubMed]
114. Dutilh, B.E.; Schmieder, R.; Nulton, J.; Felts, B.; Salamon, P.; Edwards, R.A.; Mokili, J.L. Reference-independent comparative metagenomics using cross-assembly: CrAss. *Bioinformatics* **2012**, *28*, 3225–3231. [CrossRef] [PubMed]
115. Rose, R.; Constantinides, B.; Tapinos, A.; Robertson, D.L.; Prosperi, M. Challenges in the analysis of viral metagenomes. *Virus Evol.* **2016**, *2*, vew022. [CrossRef]
116. Sharma, D.; Priyadarshini, P.; Vrati, S. Unraveling the web of viroinformatics: Computational tools and databases in virus research. *J. Virol.* **2015**, *89*, 1489–1501. [CrossRef] [PubMed]
117. Roux, S.; Enault, F.; Hurwitz, B.L.; Sullivan, M.B. VirSorter: Mining viral signal from microbial genomic data. *PeerJ* **2015**, *3*, e985. [CrossRef] [PubMed]
118. Jurtz, V.I.; Villarroel, J.; Lund, O.; Larsen, M.V.; Nielsen, M. MetaPhinder—Identifying Bacteriophage Sequences in Metagenomic Data Sets. *PLoS ONE* **2016**, *11*, e0163111. [CrossRef] [PubMed]
119. Zhao, G.; Wu, G.; Lim, E.S.; Droit, L.; Krishnamurthy, S.; Barouch, D.H.; Virgin, H.W.; Wang, D. VirusSeeker, a computational pipeline for virus discovery and virome composition analysis. *Virology* **2017**, *503*, 21–30. [CrossRef] [PubMed]
120. Bolduc, B.; Youens-Clark, K.; Roux, S.; Hurwitz, B.L.; Sullivan, M.B. iVirus: Facilitating new insights in viral ecology with software and community data sets imbedded in a cyberinfrastructure. *ISME J.* **2016**, *11*, 7–14. [CrossRef] [PubMed]

121. Yooseph, S.; Sutton, G.; Rusch, D.B.; Halpern, A.L.; Williamson, S.J.; Remington, K.; Eisen, J.A.; Heidelberg, K.B.; Manning, G.; Li, W.; et al. The Sorcerer II Global Ocean Sampling expedition: Expanding the universe of protein families. *PLoS Biol.* **2007**, *5*, e16. [CrossRef] [PubMed]
122. ONeill, K.; Klimke, W.; Tatusova, T. *Protein Clusters: A Collection of Proteins Grouped by Sequence Similarity and Function*; Protein Clusters Help; National Center for Biotechnology Information: Bethesda, MD, USA, 2007. Available online: https://www.ncbi.nlm.nih.gov/books/NBK3797/ (accessed on 30 March 2017).
123. Hurwitz, B.L.; Brum, J.R.; Sullivan, M.B. Depth-stratified functional and taxonomic niche specialization in the 'core'and 'flexible' Pacific Ocean Virome. *ISME J.* **2015**, *9*, 472–484. [CrossRef] [PubMed]
124. Ignacio-Espinoza, J.C.; Solonenko, S.A.; Sullivan, M.B. The global virome: Not as big as we thought? *Curr. Opin. Virol.* **2013**, *3*, 566–571. [CrossRef] [PubMed]
125. Roux, S.; Brum, J.R.; Dutilh, B.E.; Sunagawa, S.; Duhaime, M.B.; Loy, A.; Poulos, B.T.; Solonenko, N.; Lara, E.; Poulain, J.; et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* **2016**, *537*, 689–693. [CrossRef] [PubMed]
126. Breitbart, M.; Thompson, L.R.; Suttle, C.A.; Sullivan, M. Exploring the vast diversity of marine viruses. *Oceanography* **2007**, *20*, 135–139. [CrossRef]
127. Hurwitz, B.L.; Hallam, S.J.; Sullivan, M.B. Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biol.* **2013**, *14*, R123. [CrossRef] [PubMed]
128. Karsenti, E.; Acinas, S.G.; Bork, P.; Bowler, C.; De Vargas, C.; Raes, J.; Sullivan, M.; Arendt, D.; Benzoni, F.; Claverie, J.-M.; et al. A holistic approach to marine eco-systems biology. *PLoS Biol.* **2011**, *9*, e1001177. [CrossRef] [PubMed]
129. Duarte, C.M. Seafaring in the 21st century: The Malaspina 2010 Circumnavigation Expedition. *Limnol. Oceanogr. Bull.* **2015**, *24*, 11–14. [CrossRef]
130. Roux, S.; Hallam, S.J.; Woyke, T.; Sullivan, M.B. Viral dark matter and virus–host interactions resolved from publicly available microbial genomes. *eLife* **2015**, *4*, e08490. [CrossRef] [PubMed]
131. Mizuno, C.M.; Rodriguez-Valera, F.; Kimes, N.E.; Ghai, R. Expanding the marine virosphere using metagenomics. *PLoS Genet.* **2013**, *9*, e1003987. [CrossRef] [PubMed]
132. Edwards, R.A.; McNair, K.; Faust, K.; Raes, J.; Dutilh, B.E. Computational approaches to predict bacteriophage-host relationships. *FEMS Microbiol. Rev.* **2016**, *40*, 258–272. [CrossRef] [PubMed]
133. Andersson, A.F.; Banfield, J.F. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* **2008**, *320*, 1047–1050. [CrossRef] [PubMed]
134. Roux, S.; Solonenko, N.E.; Dang, V.T.; Poulos, B.T.; Schwenck, S.M.; Goldsmith, D.B.; Coleman, M.L.; Breitbart, M.; Sullivan, M.B. Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. *PeerJ* **2016**, *4*, e2777. [CrossRef] [PubMed]
135. Wang, W.L.; Xu, S.Y.; Ren, Z.G.; Tao, L.; Jiang, J.W.; Zheng, S.S. Application of metagenomics in the human gut microbiome. *World J. Gastroenterol.* **2015**, *21*, 803–814. [CrossRef] [PubMed]
136. Hannigan, G.D.; Meisel, J.S.; Tyldsley, A.S.; Zheng, Q.; Hodkinson, B.P.; SanMiguel, A.J.; Minot, S.; Bushman, F.D.; Grice, E.A. The human skin double-stranded DNA virome: Topographical and temporal diversity, genetic enrichment, and dynamic associations with the host microbiome. *MBio* **2015**, *6*, e01578-15. [CrossRef] [PubMed]
137. Tsai, Y.C.; Conlan, S.; Deming, C.; Segre, J.A.; Kong, H.H.; Korlach, J.; Oh, J. Resolving the complexity of human skin metagenomes using single-molecule sequencing. *MBio* **2016**, *7*, e01948-15. [CrossRef] [PubMed]
138. Edlund, A.; Santiago-Rodriguez, T.M.; Boehm, T.K.; Pride, D.T. Bacteriophage and their potential roles in the human oral cavity. *J. Oral Microbiol.* **2015**, *7*, 27423. [CrossRef] [PubMed]
139. Pride, D.T.; Salzman, J.; Haynes, M.; Rohwer, F.; Davis-Long, C.; White, R.A., III; Loomer, P.; Armitage, G.C.; Relman, D.A. Evidence of a robust resident bacteriophage population revealed through analysis of the human salivary virome. *ISME J.* **2012**, *6*, 915–926. [CrossRef] [PubMed]
140. Boutin, S.; Graeber, S.Y.; Weitnauer, M.; Panitz, J.; Stahl, M.; Clausznitzer, D.; Kaderali, L.; Einarsson, G.; Tunney, M.M.; Elborn, J.S.; et al. Comparison of microbiomes from different niches of upper and lower airways in children and adolescents with cystic fibrosis. *PLoS ONE* **2015**, *10*, e0116029. [CrossRef] [PubMed]
141. Moran Losada, P.; Chouvarine, P.; Dorda, M.; Hedtfeld, S.; Mielke, S.; Schulz, A.; Wiehlmann, L.; Tummler, B. The cystic fibrosis lower airways microbial metagenome. *ERJ Open Res.* **2016**, *2*, 00096-2015. [CrossRef] [PubMed]

142. Cui, L.; Lucht, L.; Tipton, L.; Rogers, M.B.; Fitch, A.; Kessinger, C.; Camp, D.; Kingsley, L.; Leo, N.; Greenblatt, R.M.; et al. Topographic diversity of the respiratory tract mycobiome and alteration in HIV and lung disease. *Am. J. Respir. Crit. Care Med.* **2015**, *191*, 932–942. [CrossRef] [PubMed]
143. Mitchell, A.B.; Oliver, B.G.; Glanville, A.R. Translational aspects of the human respiratory virome. *Am. J. Respir. Crit. Care Med.* **2016**, *194*, 1458–1464. [CrossRef] [PubMed]
144. Young, J.C.; Chehoud, C.; Bittinger, K.; Bailey, A.; Diamond, J.M.; Cantu, E.; Haas, A.R.; Abbas, A.; Frye, L.; Christie, J.D.; et al. Viral metagenomics reveal blooms of anelloviruses in the respiratory tract of lung transplant recipients. *Am. J. Transplant.* **2015**, *15*, 200–209. [CrossRef] [PubMed]
145. Lim, Y.W.; Schmieder, R.; Haynes, M.; Willner, D.; Furlan, M.; Youle, M.; Abbott, K.; Edwards, R.; Evangelista, J.; Conrad, D.; et al. Metagenomics and metatranscriptomics: Windows on CF-associated viral and microbial communities. *J. Cyst. Fibros.* **2013**, *12*, 154–164. [CrossRef] [PubMed]
146. Duerkop, B.A.; Hooper, L.V. Resident viruses and their interactions with the immune system. *Nat. Immunol.* **2013**, *14*, 654–659. [CrossRef] [PubMed]
147. O'Toole, P.W.; Jeffery, I.B. Gut microbiota and aging. *Science* **2015**, *350*, 1214–1215. [CrossRef] [PubMed]
148. Jeffery, I.B.; Lynch, D.B.; O'Toole, P.W. Composition and temporal stability of the gut microbiota in older persons. *ISME J.* **2016**, *10*, 170–182. [CrossRef] [PubMed]
149. Lynch, D.B.; Jeffery, I.B.; O'Toole, P.W. The role of the microbiota in ageing: Current state and perspectives. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2015**, *7*, 131–138. [CrossRef] [PubMed]
150. Mancabelli, L.; Milani, C.; Lugli, G.A.; Turroni, F.; Ferrario, C.; van Sinderen, D.; Ventura, M. Meta-analysis of the human gut microbiome from urbanized and pre-agricultural populations. *Environ. Microbiol.* **2017**, *19*, 1379–1390. [CrossRef] [PubMed]
151. Milani, C.; Ferrario, C.; Turroni, F.; Duranti, S.; Mangifesta, M.; van Sinderen, D.; Ventura, M. The human gut microbiota and its interactive connections to diet. *J. Hum. Nutr. Diet.* **2016**, *29*, 539–546. [CrossRef] [PubMed]
152. Lynch, D.B.; Jeffery, I.B.; Cusack, S.; O'Connor, E.M.; O'Toole, P.W. Diet-microbiota-health interactions in older subjects: Implications for healthy aging. *Interdiscip. Top. Gerontol.* **2015**, *40*, 141–154. [PubMed]
153. Joyce, S.A.; Gahan, C.G. Disease-associated changes in bile acid profiles and links to altered gut microbiota. *Dig. Dis.* **2017**, *35*, 169–177. [CrossRef] [PubMed]
154. Milani, C.; Ticinesi, A.; Gerritsen, J.; Nouvenne, A.; Lugli, G.A.; Mancabelli, L.; Turroni, F.; Duranti, S.; Mangifesta, M.; Viappiani, A.; et al. Gut microbiota composition and *Clostridium difficile* infection in hospitalized elderly individuals: A metagenomic study. *Sci. Rep.* **2016**, *6*, 25945. [CrossRef] [PubMed]
155. Lugli, G.A.; Milani, C.; Turroni, F.; Tremblay, D.; Ferrario, C.; Mancabelli, L.; Duranti, S.; Ward, D.V.; Ossiprandi, M.C.; Moineau, S.; et al. Prophages of the genus *Bifidobacterium* as modulating agents of the infant gut microbiota. *Environ. Microbiol.* **2016**, *18*, 2196–2213. [CrossRef] [PubMed]
156. 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. Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* **2015**, *160*, 447–460. [CrossRef] [PubMed]
157. Minot, S.; Sinha, R.; Chen, J.; Li, H.; Keilbaugh, S.A.; Wu, G.D.; Lewis, J.D.; Bushman, F.D. The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Res.* **2011**, *21*, 1616–1625. [CrossRef] [PubMed]
158. Reyes, A.; Haynes, M.; Hanson, N.; Angly, F.E.; Heath, A.C.; Rohwer, F.; Gordon, J.I. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **2010**, *466*, 334–338. [CrossRef] [PubMed]
159. Dutilh, B.E.; Cassman, N.; McNair, K.; Sanchez, S.E.; Silva, G.G.; Boling, L.; Barr, J.J.; Speth, D.R.; Seguritan, V.; Aziz, R.K.; et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat. Commun.* **2014**, *5*, 4498. [CrossRef] [PubMed]
160. Manrique, P.; Bolduc, B.; Walk, S.T.; van der Oost, J.; de Vos, W.M.; Young, M.J. Healthy human gut phageome. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 10400–10405. [CrossRef] [PubMed]
161. Lorenz, P.; Eck, J. Metagenomics and industrial applications. *Nat. Rev. Microbiol.* **2005**, *3*, 510–516. [CrossRef] [PubMed]
162. Coughlan, L.M.; Cotter, P.D.; Hill, C.; Alvarez-Ordóñez, A. Biotechnological applications of functional metagenomics in the food and pharmaceutical industries. *Front. Microbiol.* **2015**, *6*, 672. [CrossRef] [PubMed]
163. Schoenfeld, T.; Liles, M.; Wommack, K.E.; Polson, S.W.; Godiska, R.; Mead, D. Functional viral metagenomics and the next generation of molecular tools. *Trends Microbiol.* **2010**, *18*, 20–29. [CrossRef] [PubMed]

164. Schoenfeld, T.W.; Moser, M.J.; Mead, D. Functional Viral Metagenomics and the Development of New Enzymes for DNA and RNA Amplification and Sequencing. In *Encyclopedia of Metagenomics: Genes, Genomes and Metagenomes: Basics, Methods, Databases and Tools*; Springer International Publishing: New York, NY, USA, 2015; pp. 198–218.
165. Moser, M.J.; DiFrancesco, R.A.; Gowda, K.; Klingele, A.J.; Sugar, D.R.; Stocki, S.; Mead, D.A.; Schoenfeld, T.W. Thermostable DNA polymerase from a viral metagenome is a potent RT-PCR enzyme. *PLoS ONE* **2012**, *7*, e38371. [CrossRef] [PubMed]
166. Schmelcher, M.; Donovan, D.M.; Loessner, M.J. Bacteriophage endolysins as novel antimicrobials. *Future Microbiol.* **2012**, *7*, 1147–1171. [CrossRef] [PubMed]
167. Fischetti, V.A. Bacteriophage lysins as effective antibacterials. *Curr. Opin. Microbiol.* **2008**, *11*, 393–400. [CrossRef] [PubMed]
168. Schmitz, J.E.; Schuch, R.; Fischetti, V.A. Identifying active phage lysins through functional viral metagenomics. *Appl. Environ. Microbiol.* **2010**, *76*, 7181–7187. [CrossRef] [PubMed]
169. Simmonds, P.; Adams, M.J.; Benkő, M.; Breitbart, M.; Brister, J.R.; Carstens, E.B.; Davison, A.J.; Delwart, E.; Gorbalenya, A.E.; Harrach, B.; et al. Consensus statement: Virus taxonomy in the age of metagenomics. *Nat. Rev. Microbiol.* **2017**, *15*, 161–168. [CrossRef] [PubMed]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).