



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

# Hybrid assembly of an agricultural slurry virome reveals a diverse and stable community with the potential to alter the metabolism and virulence of veterinary pathogens

### Citation for published version:

Cook, R, Hooton, S, Trivedi, U, King, L, Dodd, CER, Hobman, JL, Stekel, DJ, Jones, MA & Millard, AD 2021, 'Hybrid assembly of an agricultural slurry virome reveals a diverse and stable community with the potential to alter the metabolism and virulence of veterinary pathogens', *Microbiome*, vol. 9, no. 1, 65.  
<https://doi.org/10.1186/s40168-021-01010-3>

### Digital Object Identifier (DOI):

[10.1186/s40168-021-01010-3](https://doi.org/10.1186/s40168-021-01010-3)

### Link:

[Link to publication record in Edinburgh Research Explorer](#)

### Document Version:

Publisher's PDF, also known as Version of record

### Published In:

Microbiome

### General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.




RESEARCH

Open Access



# Hybrid assembly of an agricultural slurry virome reveals a diverse and stable community with the potential to alter the metabolism and virulence of veterinary pathogens

Ryan Cook<sup>1</sup>, Steve Hooton<sup>2</sup>, Urmi Trivedi<sup>3</sup>, Liz King<sup>2</sup>, Christine E. R. Dodd<sup>2</sup>, Jon L. Hobman<sup>2</sup>, Dov J. Stekel<sup>2</sup>, Michael A. Jones<sup>1\*</sup> and Andrew D. Millard<sup>4\*</sup> 

## Abstract

**Background:** Viruses are the most abundant biological entities on Earth, known to be crucial components of microbial ecosystems. However, there is little information on the viral community within agricultural waste. There are currently ~2.7 million dairy cattle in the UK producing 7–8% of their own bodyweight in manure daily, and 28 million tonnes annually. To avoid pollution of UK freshwaters, manure must be stored and spread in accordance with guidelines set by DEFRA. Manures are used as fertiliser, and widely spread over crop fields, yet little is known about their microbial composition. We analysed the virome of agricultural slurry over a 5-month period using short and long-read sequencing.

**Results:** Hybrid sequencing uncovered more high-quality viral genomes than long or short-reads alone; yielding 7682 vOTUs, 174 of which were complete viral genomes. The slurry virome was highly diverse and dominated by lytic bacteriophage, the majority of which represent novel genera (~98%). Despite constant influx and efflux of slurry, the composition and diversity of the slurry virome was extremely stable over time, with 55% of vOTUs detected in all samples over a 5-month period. Functional annotation revealed a diverse and abundant range of auxiliary metabolic genes and novel features present in the community, including the agriculturally relevant virulence factor VapE, which was widely distributed across different phage genera that were predicted to infect several hosts. Furthermore, we identified an abundance of phage-encoded diversity-generating retroelements, which were previously thought to be rare on lytic viral genomes. Additionally, we identified a group of crAssphages, including lineages that were previously thought only to be found in the human gut.

(Continued on next page)

\* Correspondence: [michael.a.jones@nottingham.ac.uk](mailto:michael.a.jones@nottingham.ac.uk); [adm39@le.ac.uk](mailto:adm39@le.ac.uk)

<sup>1</sup>School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington Campus, College Road, Loughborough, Leicestershire LE12 5RD, UK

<sup>4</sup>Dept Genetics and Genome Biology, University of Leicester, University Road, Leicester, Leicestershire LE1 7RH, UK

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

**Conclusions:** The cattle slurry virome is complex, diverse and dominated by novel genera, many of which are not recovered using long or short-reads alone. Phages were found to encode a wide range of AMGs that are not constrained to particular groups or predicted hosts, including virulence determinants and putative ARGs. The application of agricultural slurry to land may therefore be a driver of bacterial virulence and antimicrobial resistance in the environment.

**Keywords:** Phages, Viromics, VapE, Diversity-generating retroelements, crAssphage, PromethION, Slurry, Dairy

## Background

Bacteriophages, or simply phages are recognised as the most abundant biological entities on the planet [1] and drive bacterial evolution through predator-prey dynamics [2, 3], and horizontal gene transfer [4]. In all systems where phages have been studied in detail, they have significant ecological roles [5–7]. The contribution of phages to microbial communities has arguably been most extensively studied in the oceans [8–12] where, in addition to releasing large quantities of organic carbon and other nutrients through lysing bacteria, marine phages are thought to contribute to biogeochemical cycles by augmenting host metabolism with auxiliary metabolic genes (AMGs) [12–15]. Since their initial discovery, AMGs have been identified in diverse environments, including the ocean and soils [10, 16]. The putative functions of AMGs are wide-ranging with the potential to alter photosynthesis, carbon metabolism, sulphur metabolism, nitrogen uptake and complex carbohydrate metabolism [11–13, 16–21].

In addition to augmenting host metabolism, phages can contribute to bacterial virulence through phage conversion via the carriage of virulence factors and toxins [22–27]. Phages have also been implicated in the transfer of antimicrobial resistance genes (ARGs) [28]; however, the study into the importance of phages in the transfer of ARGs has reached polarising conclusions [29, 30]. Despite the vital and complex contributions of phages to microbial ecology, there is a lack of knowledge about their roles in agricultural slurry.

Manure is an unavoidable by-product from the farming of livestock. There are ~2.7 million dairy cattle in the UK, with ~1.8 million in milking herds [31]. A fully grown milking cow produces 7–8% of their own body-weight as manure per day [32], leading to an estimated 28.31 million tonnes of manure produced by UK dairy cattle in 2010 alone [33]. These wastes are rich in nitrates and phosphates, making them valuable as a source of organic fertiliser, with an average value of £78 per cow per year [34]. However, agricultural wastes can be an environmental pollutant. Inadequate storage and agricultural run-off may lead to an increased biological oxygen demand (BOD) of freshwaters, leading to algal blooms and eutrophication [35–38]. Areas particularly at risk of nitrate pollution of ground or surface waters are classified as nitrate vulnerable zones (NVZs), and these

constitute 55% of land in England [39]. For this reason, the application of organic fertilisers to fields in the UK is strictly controlled and can only be applied during certain times of the year [40]. Thus, there is the requirement to store vast volumes of slurry for several months.

To produce slurry, solids are separated from manure using apparatus such as a screw press. The liquid fraction forms the basis of slurry, which is stored in a tank or lagoon, where it is mixed with water and other agricultural wastes before its application as fertiliser. Despite being widely used as a fertiliser, the composition of the virome within slurry is poorly studied. Culture-based approaches have been used to study phages infecting specific bacteria such as *Escherichia coli* [41–43], but total viral diversity within cattle slurry remains largely unexplored.

Short-read viromics has transformed our understanding of phages in other systems, allowing an overview of the abundance and diversity of phages [8, 9, 12, 44] and AMGs found within their genomes [12, 13, 16]. The power of viromics is exemplified by the study of crAssphage, which was first discovered in viromes in 2014 [45] and has subsequently been found to be the most abundant phage in the human gut and has recently been brought into culture [45–47]. However, the use of short-reads is not without limitations. Phages that contain genomic islands and/or have high micro-diversity, such as phages of the ubiquitous *Pelagibacterales* [48, 49], can cause genome fragmentation during assembly [50–53]. The development of long-read sequencing technologies—most notably Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT)—offer a solution to such issues. The longer reads are potentially able to span the length of entire phage genomes, overcoming assembly issues resulting from repeat regions and low coverage [50–52]. The cost of longer reads is a higher error rate, which can lead to inaccurate CDS prediction [54, 55].

Recently, a Long-Read Linker-Amplified Shotgun Library (LASL) approach was developed that combines LASL library preparation with ONT MinION sequencing [56]. This approach overcame both the requirement for high DNA input for MinION sequencing and associated assembly issues with short-read sequencing. The resulting assembly increased both the number and completeness of phage genomes compared to short-read assemblies [56]. An alternative approach that has utilised

long-read sequencing used the ONT GridION platform to obtain entire phage genomes using an amplification-free approach on high molecular weight DNA [57]. While this approach recovered over 1000 high-quality viral genomes that could not be recovered from short-reads alone, it requires large amounts of input DNA [57], that may be a limiting factor of many environments.

The aim of this work was to utilise viral metagenomics to investigate the diversity, community structure and ecological roles of viruses within dairy cattle slurry that is spread on agricultural land as an organic fertiliser.

## Methods

### DNA extraction and sequencing

DNA from the viral fraction was extracted from 10 ml of slurry as previously described [58]. Briefly, slurry was mixed with PBS buffer and centrifuged, prior to filtration to remove bacteria. Viral particles were concentrated using an Amicon column (Sigma-Aldrich) and DNA was extracted using a standard phenol-chloroform extraction. For short-read sequencing on un-amplified DNA, Illumina sequencing was carried out on NovaSeq using 2 × 150 library. For long read sequencing, DNA from four viral samples was pooled and subject to amplification with Illustra Ready-To-Go Genomphi V3 DNA amplification kit (GE, Healthcare) following the manufacturer's instructions. Post amplification DNA was de-branched with S1 nuclease (Thermo Fisher Scientific), following the manufacturer's instructions and cleaned using a DNA Clean and Concentrator column (Zymo Research). Sequencing was carried out by Edinburgh Genomics, with size selection of DNA to remove DNA < 5 kb prior to running on single PromethION flow cell. Reads were base called with guppy v2.3.35.

### Assembly and quality control

Illumina virome reads were trimmed with Trimmomatic v0.36 [59] using the following settings; PE illuminaclip, 2:30:10 leading:15 trailing:15 slidingwindow:4:20 minlen:50. Reads from the five samples were co-assembled with MEGAHIT v1.1.2 [60] using the settings; --k-min 21 --k-max 149 --k-step 24. Long-reads were assembled with flye v2.6-g0d65569, reads were mapped back against the assembly with Minimap2 v2.14-r892-dirty [61] to produce BAM files and initially corrected with marginPolish v1.0.0 with 'allParams.np.ecoli.json'. Bacterial contamination and virus-like particle (VLP) enrichment was assessed with ViromeQC v1.0 [62].

### Identifying viral operational taxonomic units

To identify viral contigs, a number of filtering steps were applied. All contigs ≥ 10 kb and circular contigs < 10 kb [53] were processed using MASH v2.0 [63] separately

against the RefSeq70 database [64] and a publicly available database of phage genomes (March 2020;  $P = 0.01$ ). If the closest RefSeq70 hit was to a phage/virus, the contig was included as a viral operational taxonomic unit (vOTU). Failing this, if the contig obtained a closer hit to the phage database than RefSeq70, the contig was included as a vOTU. Remaining contigs were included as vOTUs if they satisfied at least two of the following criteria; 1: VIBRANT v1.0.1 indicated sequence is viral [65], 2: obtained adjusted  $p$  value ≤ 0.05 from DeepVirFinder v1.0 [66], 3: 30% of ORFs on the contig obtained a hit to a prokaryotic virus orthologous group (pVOG) [67] using Hmmscan v3.1b2 (-E 0.001) [68]. However, circular contigs < 10 kb only had to satisfy either criteria 1 or 3, as DeepVirFinder scores for these contigs were inconsistent.

### Prophage analysis

A set of prophage sequences was identified from bacterial metagenomes from the same tank. These were filtered as above, however contigs < 10 kb were not included even if circular. To determine which prophage vOTUs could be detected in the free viral fraction, Illumina virome reads were mapped to vOTUs using Bbmap v38.69 [69] at 90% minimum ID and the ambiguous=all flag, and PromethION reads were mapped to prophage vOTUs using Minimap2 v2.14-r892-dirty [61] with parameters '-a -x map-ont'. vOTUs were deemed as present in the free viral fraction if they obtained ≥ 1x coverage across ≥ 75% of contig length in at least one sample [53]. To determine the ends of prophages, differential coverage obtained by mapping the Illumina virome reads was investigated. Median coverage of the whole prophage was calculated and compared to median coverage across a 500 bp sliding window (Supplementary Tables 6 & 7). If the 500 bp window had a depth of coverage ≥ 2x standard deviations lower than the median coverage of the whole prophage, this was considered a break in coverage and used to infer the ends of the prophage. An example is provided in supplementary Figure 1.

### Hybrid assembly composition

Illumina reads were mapped to PromethION vOTUs using Minimap2 v2.14-r892-dirty [61] and the contigs were polished using Pilon v1.22 [70]. The PromethION vOTUs underwent multiple rounds of polishing until changes to the sequence were no longer made, or the same change was swapped back and forth between rounds of polishing. The Illumina vOTUs, hybrid vOTUs and prophage vOTUs (only those detected in the viral fraction) were de-replicated at 95% average nucleotide identity (ANI) over 80% genome length using ClusterGenomes v5.1 [71] to produce a final set of vOTUs, hereby referred to as the Final Virome. To determine

assembly quality, CheckV v0.5.0 [72] was used. As this pipeline was released after the analysis in this work was performed, this was performed post-analysis.

### Alpha diversity and population dynamics

To estimate relative abundance, Illumina reads were mapped to vOTUs using Bbmap v38.69 [69] at 90% minimum ID and the `ambiguous=all` flag. vOTUs were deemed as present in a sample if they obtained  $\geq 1\times$  coverage across  $\geq 75\%$  of contig length [53]. The number of reads mapped to present vOTUs were normalised to reads mapped per million. Relative abundance values were analysed using Phyloseq v1.26.1 [73] in R v3.5.1 [74] to calculate diversity statistics.

Statistical testing of similarity of vOTU profiles between samples was carried out using DirtyGenes [75]. We used the randomization option with 5000 simulations rather than chi-squared because of the small number of samples, but resampling from the null hypothesis Dirichlet distribution because there are no replicated libraries; the updated code has been uploaded to GitHub (<https://github.com/LMShaw/DirtyGenes>). The analysis was repeated using both the preferred cut-off of minimum 1% abundance in at least one sample and also with minimum abundance at 0.5% in at least one sample. This is because with a 1% cut-off only seven vOTUs were included (plus an 'other' category binning all remaining lower abundance vOTUs) which we did not consider to be sufficiently representative; with 0.5%, 22 vOTUs were included (plus an 'other' category).

### Functional annotation

Final Virome vOTUs were annotated using Prokka v1.12 [76] with a custom database created from phage genomes downloaded at the time (March, 2020) [77], and ORFs were compared to profile HMMs of pVOGs [67] using Hmmscan v3.1b2 (-E 0.001) [68]. Final Virome vOTU ORFs were clustered at 90% ID over 90% contig length using CD-HIT v4.6 [78] to reduce redundancy. The resultant proteins were submitted to eggNOG-mapper v2.0 [79] with default parameters, and the output was manually inspected to identify AMG of interest. Translated ORFs identified as carbohydrate-active enzymes (CAZymes) by eggNOG were submitted to the dbCAN2 meta-server for CAZyme identification using the HMMER method to confirm their identity [16, 80].

### Diversity-generating retroelement analysis

vOTUs found to encode a putative reverse transcriptase were processed using MetaCCST [81] to identify potential diversity-generating retroelements (DGRs). To identify hypervariable regions in the target gene of DGRs, reads from each sample were individually mapped to vOTUs using Bbmap v38.69 [69] at 95% minimum ID

with the `ambiguous=all` flag. Resultant bam files were processed with Samtools v1.10 [82] to produce a mpileup file. Variants were called using VarScan v2.3 [83] mpileup2snp command with parameters '`--min-coverage 10 --min-avg-qual-30`'. The percentage of SNP sites per gene were calculated for both DGR target gene(s) and all other genes on the vOTU, in order to identify if the DGR target gene(s) contained more SNP sites than on average across the vOTU.

### Taxonomy and predicted host

Final Virome vOTUs were clustered using vConTACT2 v0.9.13 [84] with parameters; `--db 'ProkaryoticViralRefSeq85-Merged' --pcs-mode MCL --vcs-mode Cluster-ONE`. A set of publicly available phage genome sequences (7527), that had been deduplicated at 95% identity with dedupe.sh v36.20 [69], were included. The resultant network was visualised using Cytoscape v3.7.1 [85].

To determine if any previously known phage genomes were present in slurry viromes, reads were mapped to a dataset of publicly a set of publicly available phage genome sequences (March, 2020; 11,030), that had been deduplicated at 95% identity with dedupe.sh v36.20 [69]. Illumina reads were mapped using Bbmap v38.69 [69] at 90% minimum ID [53] and the `ambiguous=all` flag. PromethION reads were mapped using Minimap2 v2.14-r892-dirty [61] with parameters '`-a -x map-ont`'. Phages were deemed as present if they obtained  $\geq 1\times$  coverage across  $\geq 75\%$  of sequence length [53].

Putative hosts for viral vOTUs were predicted with WiSH v1.0 [86] using a database of 9620 bacterial genomes. A *p* value cut-off of 0.05 was used. Taxonomy for the predicted hosts was obtained using the R [74] package Taxonomizr v0.5.3 [87].

### Lifestyle prediction

To determine which Final Virome vOTUs were temperate, ORFs were compared to a custom set of 29 profile HMMs for transposase, integrase, excisionase, resolvase and recombinase proteins downloaded from Pfam (PF07508, PF00589, PF01609, PF03184, PF02914, PF01797, PF04986, PF00665, PF07825, PF00239, PF13009, PF16795, PF01526, PF03400, PF01610, PF03050, PF04693, PF07592, PF12762, PF13359, PF13586, PF13610, PF13612, PF13701, PF13737, PF13751, PF13808, PF13843 and PF13358) [88] using Hmmscan v3.1b2 [68] with the `--cut_ga` flag. Any vOTUs with an ORF which obtained a hit were classified as temperate.

### Positive selection analysis

Final Virome vOTUs which obtained  $\geq 15\times$  median coverage across  $\geq 75\%$  of contig length in every sample (excluding PHI75) were included in variant analysis.



Briefly, reads were mapped onto the contigs using Bbmap v38.69 [69] at 95% minimum ID with the ambiguous=all flag, and a sorted indexed BAM file was produced. Snippy v4.4.5 [89] was used to call variants with parameters '--mapqual 0 --mincov 10'. For genes which contained at least one single nucleotide polymorphism (SNP) or multiple nucleotide polymorphism (MNP), natural selection (pN/pS) was calculated using a method adapted from Gregory et al. [9]. In this method, adjacent SNPs were linked as MNPs by Snippy.

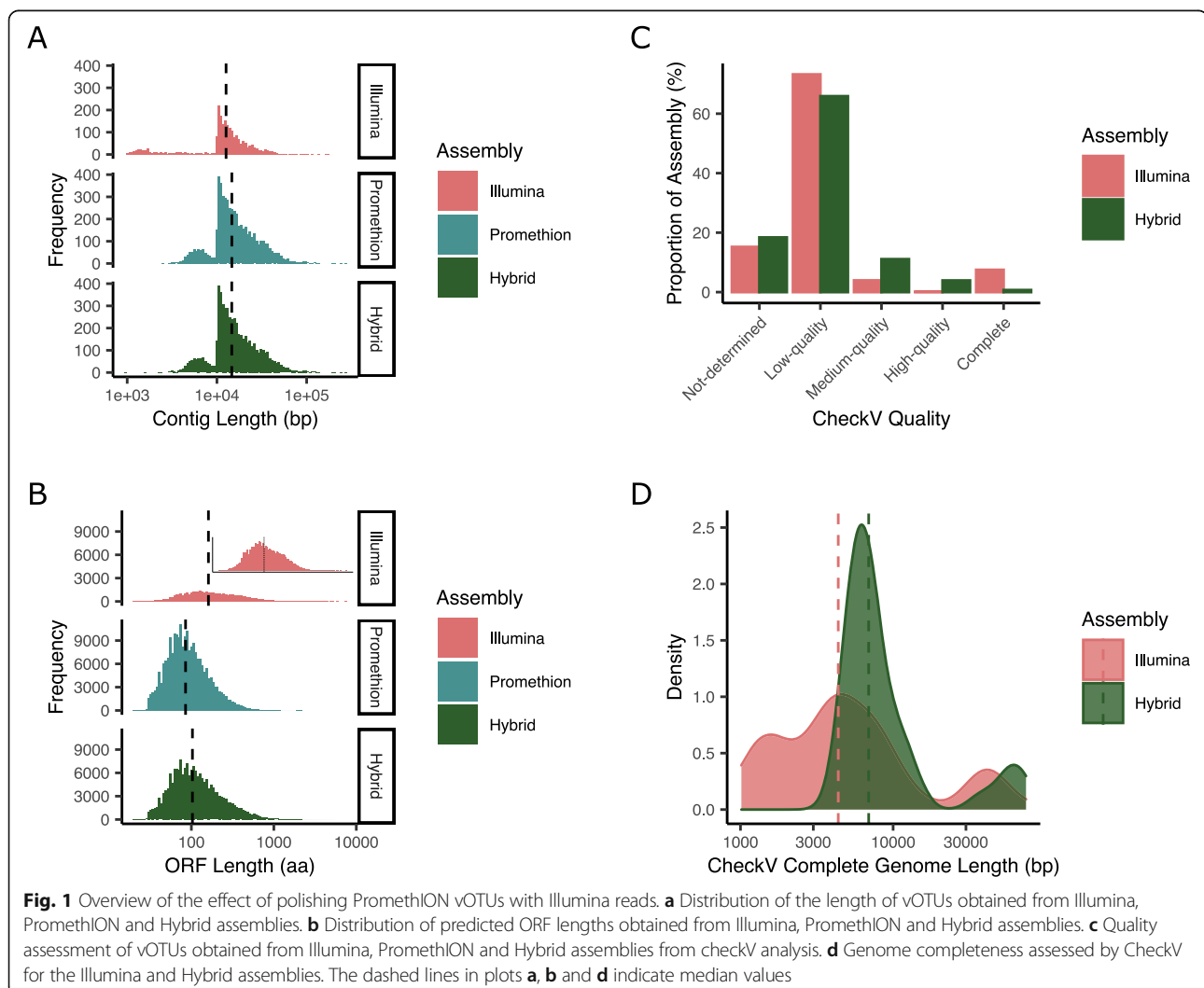
## Results

The farm in this study is a high-performance dairy farm in the East Midlands, UK with ~200 milking cattle. It houses a three million litre capacity slurry tank and an additional seven million litre lagoon to house overflow from the tank. The tank receives daily influent from the dairy farm including faeces, urine, washwater, footbath and waste milk through a slurry handling and general

farm drainage system. Slurry solids are separated using a bed-press and solids are stored in a muck heap. The slurry tank and muck heap are open to the elements and the slurry tank also receives further influent from rainwater, muck heap run-off, and potentially from wildlife. The tank is emptied to ~10% of its maximum volume every ~6 weeks and the slurry is applied on fields as fertiliser.

## Comparison of short- and long-read assemblies

Five samples were collected from the slurry tank over a five-month period (07/06/2017–10/10/2017) (Supplementary Table 1) with Illumina libraries prepared from each sample. Initial analysis of the five samples sequencing data using viromeQC [62] indicated that one sample (PHI75) had high levels of bacterial contamination (Supplementary Table 1). Sample PHI75 was excluded from further analysis, with remaining DNA from the



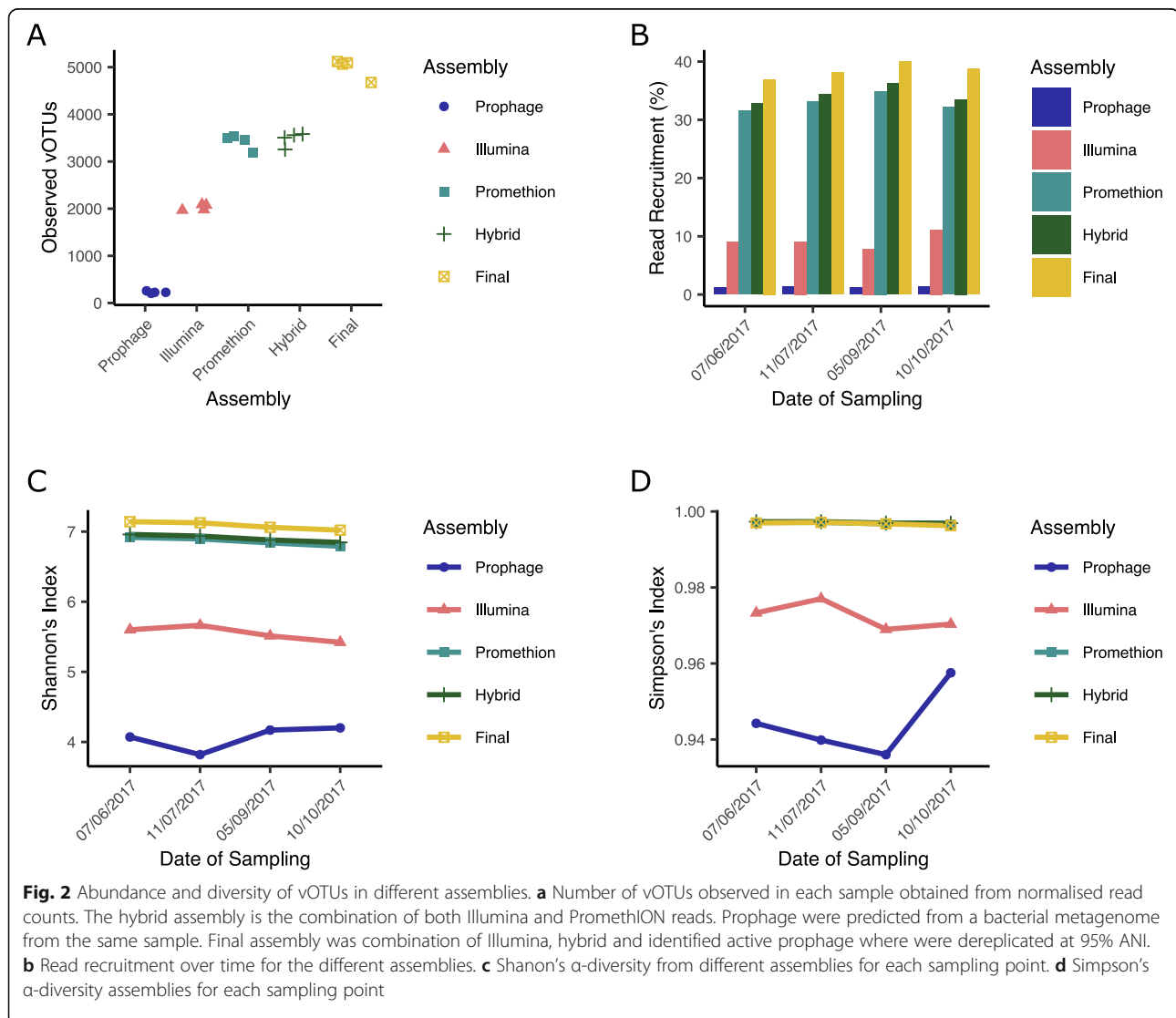
other four samples pooled, amplified and sequenced by PromethION sequencing.

Assembly was carried out with just Illumina or PromethION reads, resulting in 1844 and 4954 vOTUs  $\geq 10$  kb respectively. The PromethION assembly resulted in an increase in the median contig size from 12,648 to 14,658 compared to the Illumina only assembly (Fig. 1a). The number of predicted genes per kb was also higher in the PromethION assembly. The increased error rate of Nanopore sequencing compared to Illumina sequencing is known to result in truncated gene calls [54, 55]. To alleviate this, PromethION contigs were polished with Illumina reads, creating a hybrid assembly and resulting in a decrease in the number of genes per kb from 2.059 (median length: 85 aa) to 1.706 (median length: 103 aa; Fig. 1b).

As whole genome amplification was used to gain sufficient material for PromethION sequencing, all diversity

statistics and relative abundance data was determined from Illumina reads only. The percentage of reads that could be recruited to each different assembly was assessed. Both the PromethION (32.663%) and hybrid (33.976%) assemblies recruited more reads than the Illumina assembly (9.048%; Fig. 2b). The median number of observed vOTUs per sample was higher in the PromethION (3,483) and hybrid (3,532) assemblies than that of the Illumina assembly (2028; Fig. 2a). The predicted Shannon and Simpson diversity indices increased in the hybrid (Shannon: 6.909; Simpson: 0.997) and PromethION (Shannon: 6.867; Simpson: 0.997) assemblies compared to the Illumina assembly (Shannon: 5.557; Simpson: 0.972; Fig. 2c, d).

To determine the completeness and quality of the identified viral contigs, CheckV [72] was used. The hybrid assembly contained a lower proportion of low-quality genomes (65.886%), and a higher proportion of



medium and high-quality (15.015%) genomes than the Illumina assembly (low-quality: 73.217%; medium and high-quality: 4.083%; Fig. 1c). Conversely, the Illumina assembly contained more predicted complete genomes than the hybrid assembly (Illumina: 167; hybrid: 40). This may be due to the size selection of PromethION sequencing for longer reads, reflected in the longer average length of the complete genomes obtained from hybrid assembly (Fig. 1d).

To fully understand the diversity of phages within the slurry tank, we also investigated the presence of prophage elements in the bacterial fraction. A total of 2892 putative prophages were predicted, of which only 407 could be detected in the free phage fraction by read mapping. We combined the predicted 407 active prophages, with the Illumina and hybrid assemblies. Redundancy was removed using `cluster_phages_genomes.pl` [71], resulting in 7682 vOTUs. Having established the most comprehensive DNA virome possible, the data was further analysed.

### Characterisation of the slurry virome

The percentage of reads that could be recruited from each sample varied from 36.943% (PHI73; 07/06/2017; Fig. 2b) to 39.996% (PHI76; 05/09/2017; Fig. 2b). Across the five-month sampling period, the Shannon's index alpha diversity estimates only varied from 7.02 (PHI77; 10/10/2017) to 7.141 (PHI73; 07/06/2017), suggesting a stable and diverse virome across seasons (Fig. 2c, d). Although diverse, the virome remained stable across all sampling points with 55% (4,256) of 7682 vOTUs found in all samples, and only 477 (~6%) of vOTUs unique to any one sampling point. Furthermore, testing with DirtyGenes [75] found no significant difference between the vOTU abundance profiles of the samples ( $p = 0.1142$  with 1% cut-off;  $p = 0.863$  with 0.5% cut-off). To determine if the stability in macro-diversity was mirrored by changes in micro-diversity, we assessed which predicted phage genes were under positive selection ( $pN/pS > 1$ ). Our analysis showed 1610/210,997 genes (0.763%) to be under positive selection in at least one sample (Supplementary Table 2). From these, putative function could be assigned to 388 translated genes. The most common predicted functions were related to phage tail (30), and phage structure (24).

To give a broader overview of the type of viruses present in the sample, pVOGs were used to infer the taxonomic classification of each vOTU. Of the vOTUs that contained proteins that matched the pVOG databases [67], 91% were associated with the order *Caudovirales*, 2.17% associated with non-tailed viruses and the remainder not classified. Approximately 10% (710) of vOTUs were identified as temperate, suggesting that the community is dominated by lytic phages of the order

*Caudovirales*. The abundance of temperate vOTUs was constant across samples, ranging from 5.605% (PHI76; 05/09/2017) to 8.866% (PHI77; 10/10/2017), further demonstrating the stability of the system across time.

In order to identify the species of phages present within the slurry, all vOTUs were compared against all known phages (March, 2020) using MASH [63], with an average nucleotide identity (ANI) of >95% as currently defined as a cut-off for phage species [90]. Only vOTUs *ctg5042* and *ctg217* with similarity to *Mycoplasma bacteriophage L2* (accession BL2CG) and *Streptococcus phage Javan630* (accession MK448997) respectively were detected. Furthermore, no vOTUs were similar to any phages that have previously been isolated from this system [41–43]. Thus, the vast majority of vOTUs represent novel phage species.

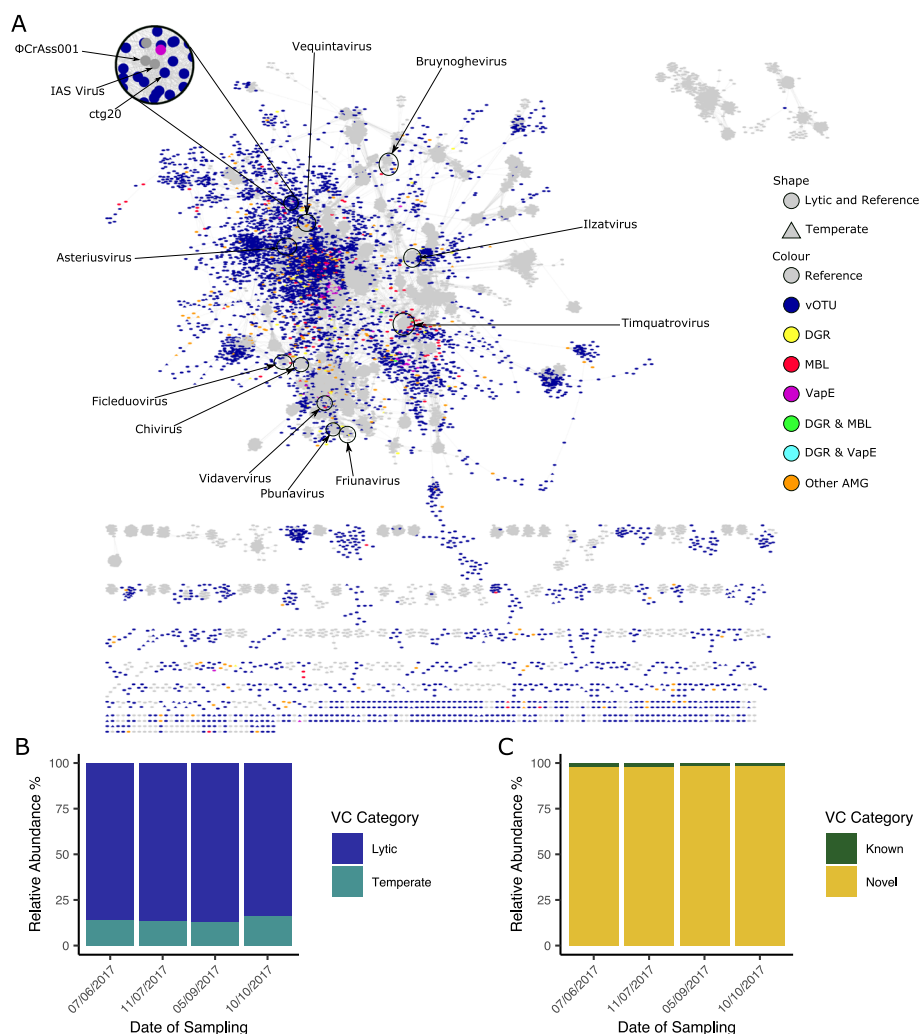
To gain an understanding of the composition at higher taxonomic levels, vConTACT2 [84] was run. Only 217 (2.825%) vOTUs clustered with a reference genome, indicating they are related at the genus level (Fig. 3a). Notably, 18 vOTUs formed a cluster with *ΦCrAss001* (accession MH675552) and phage IAS (accession KJ003983), with *ctg20* appearing to be a near-complete phage genome (~99 kb; Fig. 4b). The other 7465 vOTUs clustered only with other vOTUs (3369; 43.856%) or were singletons (4096; 53.319%), indicating 5242 putative new genera. These new genera comprised 98.037% of phages across all samples, suggesting this system is dominated by novel viruses (Fig. 3b). Working on the assumption that if a vOTU within a viral cluster (VC) was identified as temperate all other vOTUs in the cluster are, the relative abundance of temperate phages was predicted. This ranged from 13.09% (PHI76; 05/09/2017) to 16.249% (PHI77; 10/10/2017), further demonstrating the dominance of lytic viruses and stability of the system over time (Fig. 3c).

Hosts were predicted for 3189 vOTUs and the system was found to be dominated by phages predicted to infect bacteria belonging to *Firmicutes* and *Bacteroidetes*, the most dominant phyla found in the cow gut [91–93]. The proportions of host-specific abundances appeared stable across all time points (Supplementary Figure 2).

### Identification of CrAss-like phages in the slurry virome

The appearance of a cluster of 18 vOTUs that are similar to crAssphage was surprising given the discovery and abundance of crAssphage in human gut viromes [45–47, 94]. To further investigate this, phylogenies based on the method of Guerin et al. were used [47] for 15 vOTUs that contained the specific marker genes. All vOTUs formed part of the previously proposed genus VI [47], including the near complete phage (*ctg20*; Fig. 4a; Supplementary Figure 3). Furthermore, the crAssphages identified from slurry did not form a single





**Fig. 3** Taxonomic analysis of vOTUs. **a** vConTACT2 network analysis of vOTUs from this study and a database of phage genomes extracted from Genbank. The presence of selected viral accessory metabolic genes within viral clusters (VCs) is marked by different colours. **b** Abundance of viral clusters that contained  $\geq 1$  previously known viral genome (known) or no previously known viral genomes (novel). **c** Abundance of viral clusters that contained  $\geq 1$  vOTU predicted to be temperate (temperate) or none (lytic)

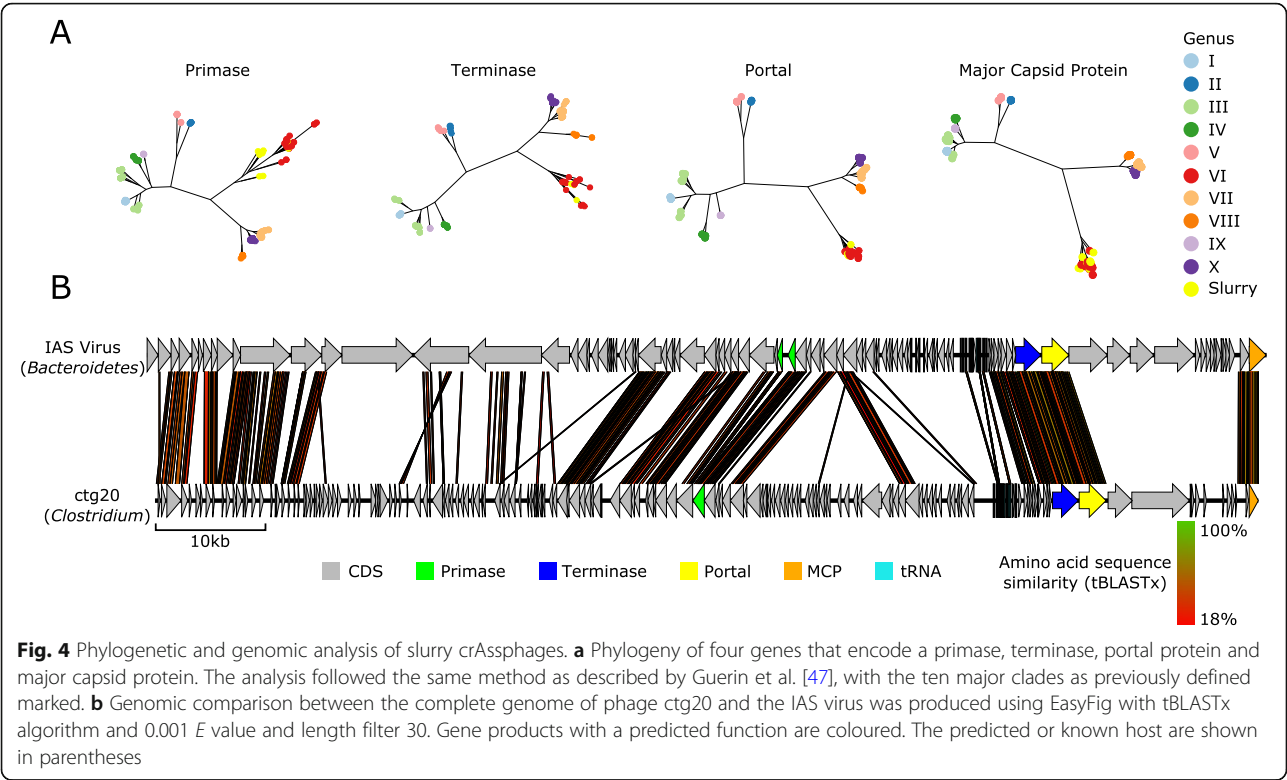
monophyletic clade. Instead, they were interspersed with human crAssphages, with some slurry crAssphages more closely related to human crAssphages than other slurry crAssphages (Fig. 4a; Supplementary Figure 3). Genome comparison of ctg20 and phage IAS from genus VI identified synteny in genome architecture between the phages, yet there are clearly several areas of divergence (Fig. 4b). The predicted host of ctg20 was *Clostridium*, which contrasts to the *Bacteroides* and *Bacteroidetes* that other crAssphages have been demonstrated or predicted to infect respectively [46, 95].

#### Abundance and diversity of auxiliary metabolic genes

In order to understand the role phages might have on the metabolic function of their hosts, function was assigned to proteins using eggNOG [79]. Out of 210,997

predicted proteins, only 48,819 (23.137%) could be assigned a putative function. The most abundant clusters of orthologous groups (COG) categories [96] were those associated with viral lifestyle; notably replication, recombination and repair, cell wall/membrane/envelope biogenesis, transcription and nucleotide transport and metabolism (Supplementary Figure 4).

In addition to this, a number of putative AMGs were identified, including putative ARGs, CAZymes, assimilatory sulfate reduction (ASR) genes, MazG, VapE and Zot (Supplementary Table 3). These AMGs were found to be abundant and not constrained to particular set of phages or hosts they infect (Fig. 3a; Supplementary Table 4). For instance, carbohydrate-active enzymes were identified on 91 vOTUs across 77 putative viral genera, with 41 vOTUs predicted to infect bacteria

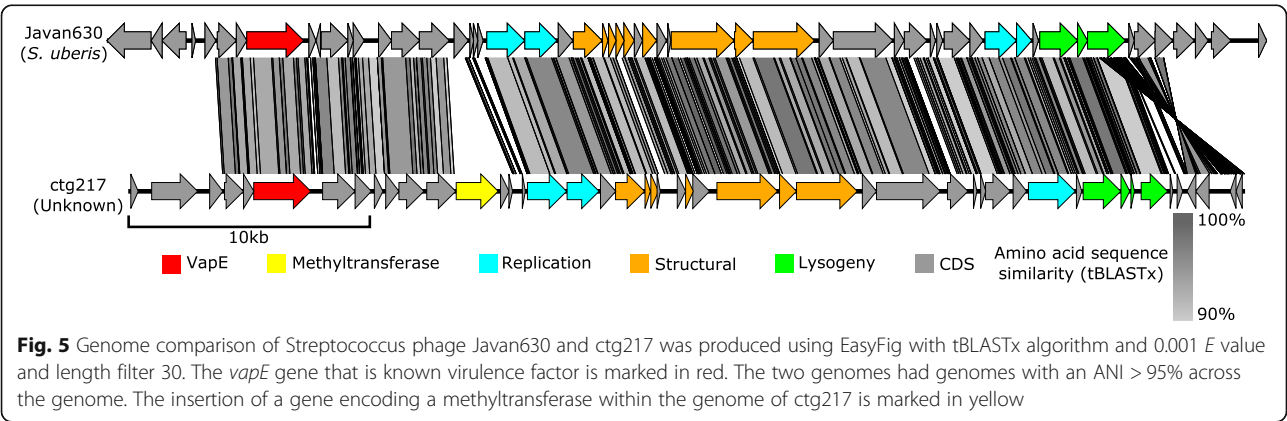


spanning 21 families (Supplementary Table 4), and genes involved in the sulphur cycle were identified on 148 vOTUs across 138 putative phage genera, with 42 vOTUs predicted to infect bacteria spanning 19 families (Supplementary Table 4).

**Abundance of virulence-associated proteins**

Genes encoding Zot were identified on 36 vOTUs across 33 putative genera, predicted to infect five different families of bacteria (Supplementary Table 4). The bacterial virulence factor VapE which is widespread in the

agricultural pathogens *Streptococcus* and *Dichelobacter* was also detected [97–99]. Recently, it has been demonstrated that deletions of prophage encoded *vapE* in *Streptococcus* have decreased growth rate in serum compared to wild type strains [100]. VapE homologues were found on 82 vOTUs (~ 1%) across 65 clusters, including 10 high-quality genomes (Fig. 3a). Bacterial hosts could be predicted for 17 vOTUs and spanned 10 families of bacteria (Supplementary Table 4). One vOTU (ctg217) shared ~ 95% ANI with the prophage Javan630 (accession MK448997) [100]. Genome comparison between ctg217



and Javan630 revealed highly conserved genomes, with insertion of a gene encoding a putative methyltransferase in ctg217 being the largest single difference (Fig. 5).

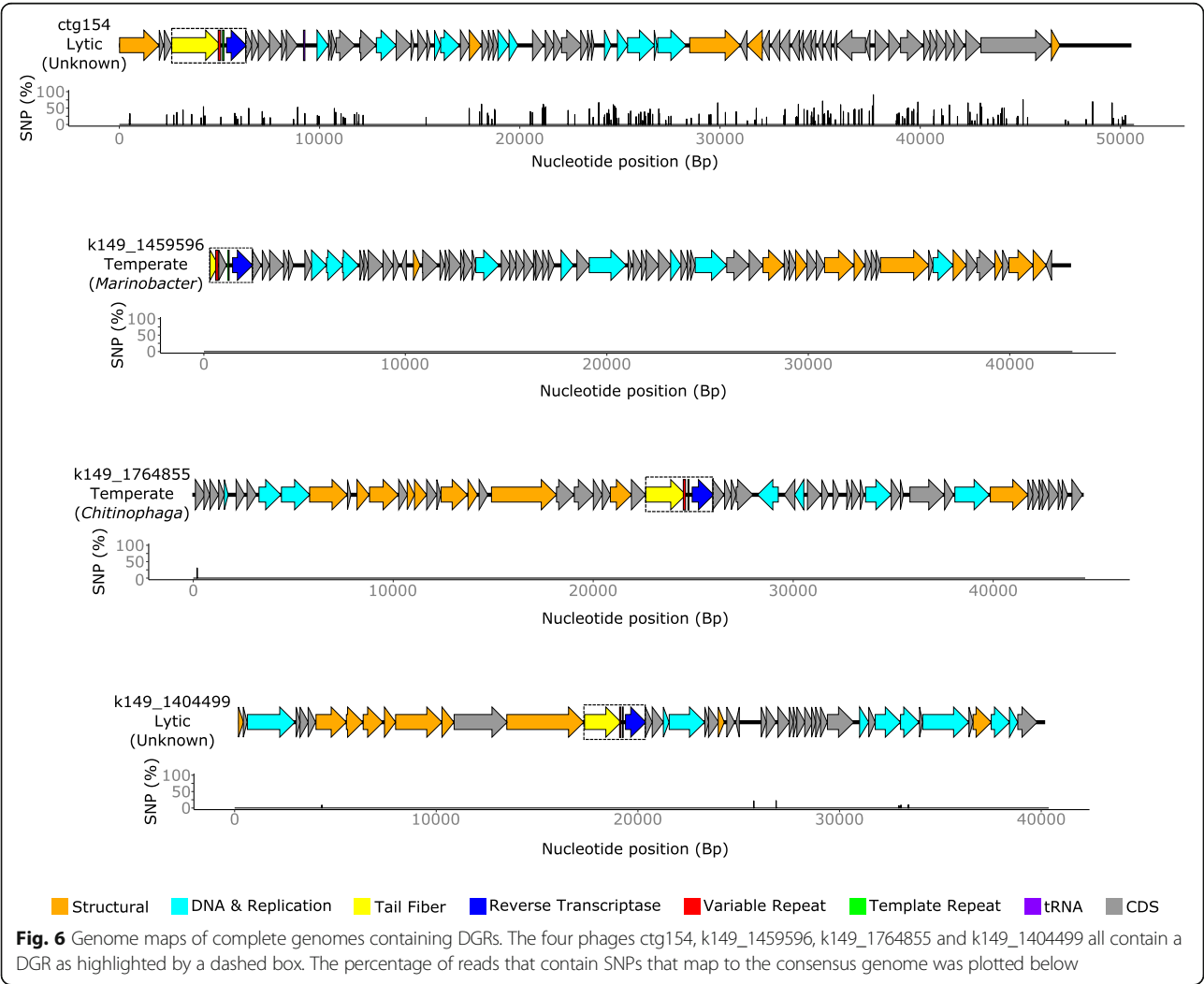
Detection of putative antimicrobial resistance genes

Putative metallo-beta-lactamases (MBLs) were identified on 146 vOTUs across 116 putative genera, with 60 vOTUs predicted to infect bacterial hosts that spanned 23 families (Supplementary Table 4). Although low in sequence similarity, structural modelling with Phyre2 [101] found many of these sequences to have the same predicted structure as the novel *bla*<sub>PNGM-1</sub> beta-lactamase (100% confidence over 99% coverage) [102]. Furthermore, these sequences contained conserved zinc-binding motifs characteristic of subclass B3 MBLs [102]. Phylogenetic analysis of putative phage MBLs, along with representative bacterial MBLs and a known phage-encoded *bla*<sub>HRVM-1</sub> [103], showed some clustered with previously characterised bacterial MBLs and others with a characterised phage *bla*<sub>HRVM-1</sub> (Supplementary Figure

5). In addition to MBLs, two putative multidrug efflux pumps were identified on two vOTUs predicted to infect two different bacterial genera (Supplementary Table 4).

Identification of diversity-generating retroelements

In addition to AMGs, we also identified 202 vOTUs that carry genes encoding a reverse transcriptase. Although dsDNA phages are known to have genes that encode for a reverse transcriptase as part of diversity-generating retroelement (DGR) and the mechanism understood [104], they are rarely reported. To determine if the identified genes encoding a reverse transcriptase were part of a DGR, MetaCCST [81] was used to identify such elements. Of the 202 vOTUs carrying a reverse transcriptase gene, 82 were predicted to be part of a DGR, which accounts for ~1% of vOTUs in the virome. In comparison, we calculated the number of DGRs that can be identified in publicly available phage genomes (12,354 unique genomes -March 2020) to be 0.178% (22 genomes).



For vOTUS where a complete DGR system (template repeat, variable repeat, reverse transcriptase and target gene) could be identified, the most commonly predicted function of the target gene was a tail fibre. The distribution of DGRs across 74 viral clusters and 15 families of predicted host bacteria (Supplementary Table 4) suggest that this is not a feature that is unique to a particular VC of phages or hosts they infect (Fig. 3a).

DGRs were predicted to occur on four phages that were deemed high-quality complete genomes (Fig. 6). These phage genomes varied in size from 40.3 to 52.07 kb, with two genomes containing putative integrases (k149\_1459596 and k149\_1764855), suggesting they are temperate, with the other two likely lytic phages (ctg154 and k149\_1404499). Interestingly, phage k149\_1459596 could not be detected between 07/06/2017 and 05/09/2017 but was the most abundant vOTU on 10/10/2017, representing over 3% of the viral population at that time. As vCONTACT2 [84] analysis was unable to classify the phages, phylogenetic analysis was carried out with gene encoding TerL to identify the closest known relatives (Supplementary Figure 6). Phage k149\_1459596 closest relative was *Vibrio* phage Rostov 7 (accession MK575466) and member of the *Myoviridae*, whilst the closest known members of the three others phages are all members of the *Siphoviridae*.

We hypothesised that the widespread distribution of DGRs would reflect widespread tropism switching in these phages, and that hypervariable DGR target genes could be detected. To investigate this, we examined variants per gene and calculated which genes were under positive selection. For the 69 DGR containing vOTUs in which a target gene could be identified, 22 of these contained a higher proportion of SNP sites in the DGR target gene(s) than the average proportion of SNP sites for non-DGR target genes on that given vOTU. One of which, a predicted phage tail protein (ctg187\_00023), was predicted to be under positive selection. Thus, many of the DGR target genes were more variable than other genes on a given vOTU (Fig. 6).

## Discussion

### Assembly comparison

Comparison of assemblies between both short-read and long-read based sequencing methods revealed significant differences in the distribution of viral contigs and the median gene length. As has been found previously, the use of long-reads alone causes problems in gene calling due to higher error rates [54]. We therefore used short-reads to polish the long-read assembly and alleviate these issues [56]. In contrast to previous methods that used LASLs combined with ONT MinION sequencing [56], we utilised whole

genome amplification followed by size selection for PromethION sequencing.

In using MDA for production of PromethION libraries, a bias in the amplification of ssDNA phage most likely occurred due to well established preference for ssDNA using this method [105]. A size selection of fragments was applied prior to promethION sequencing that would likely remove some of these smaller ssDNA genomes. However, there was a peak in contigs of 4–5 kb length in the PromethION assembly, indicative of ssDNA genomes. Given the known MDA bias, we only utilised Illumina libraries (no MDA amplification) for determining the abundance of contigs and estimates of diversity. Comparison of diversity statistics on Illumina, PromethION and hybrid assemblies suggest Illumina only assemblies may underestimate the diversity within a sample, whereas diversity estimates even on un-corrected PromethION assemblies is closer to that of hybrid assemblies. We also observed a number of smaller genomes that were obtained from Illumina only assemblies and were not present in the PromethION assembly. This likely results as part of the selection process for high molecular weight DNA (HMW) for PromethION sequencing that would exclude some small phage genomes. Therefore, whilst long-reads improved assembly statistics, the use of long-reads alone may result in exclusion of smaller phage genomes if size selection is included (as we did) and may introduce a bias of increased ssDNA genomes.

To provide the most comprehensive set of viral contigs, we included 230 predicted prophages derived from bacterial metagenomes that could be detected in the free viral fraction but were not assembled from virome reads, thus providing a more comprehensive set of viral contigs.

### Virome composition

Comparison of diversity across the period of five months revealed a highly diverse and stable virome across time. Initially, this may be somewhat surprising given the dynamics of the slurry tank, which has constant inflow from animal waste, farm effluent and rainwater, and is emptied leaving only ~10% of the tank volume every ~6 weeks. We reason that most viruses in the slurry tank will originate from cow faeces, as this is the most dominant input of the tank. Host prediction suggested the virome was dominated by viruses predicted to infect bacteria belonging to *Firmicutes* and *Bacteroidetes*, which are the two most abundant bacterial phyla in the cow rumen and gut [91–93]. To date, there has been limited study into the dairy cow gut virome and its dynamics over time. However, there is a parallel with the human gut virome which is known to be temporally stable despite constant influx and efflux [106–108], and its composition influenced by environmental factors including diet [109–111]. Assuming most viruses in the slurry tank are derived from cow faeces, the controlled



environment and diet of dairy cattle results in a temporally stable virome.

Our positive selection analyses found the most common genes to be under positive selection were those involved in bacterial attachment and adsorption. We reasoned that these findings, in conjunction with the extreme stability in macro-diversity, fit with the Royal Family model of phage-host dynamics [5]. This model suggests that dominant phages are optimised to their specific ecological niche, and in the event of bacterial resistance to infection, a highly similar phage will fill that niche. Changes in community composition over time would therefore be reflected in fine-scale diversity changes, and macro-diversity would be relatively unchanged [5]. Instead of population crashes, phages may overcome bacterial resistance through positive selection of genes involved in attachment and adsorption, and are potentially accelerating the variation of these genes with DGRs.

#### Diversity-generating retroelements

DGRs were first discovered in the phage BPP-1 (accession AY029185) where the reverse transcriptase, in combination with terminal repeat, produces an error-prone cDNA that is then stably incorporated into the tail fibre [104]. This hypervariable region mediates the host switching of BPP-1 across different *Bordetella* species [104]. Very few DGRs have been found in cultured phage isolates since, with only two DGRs found in two temperate vibriophages [112, 113]. We expanded this to 22 phages (0.178%) by searching publicly available phage genomes. Whilst not common in phage genomes, DGRs have been identified in bacterial genomes, with phage associated genes often localised next to the DGRs [113]. A recent analysis of ~32,000 prophages was able to identify a further 74 DGRs in what are thought to be active prophages from diverse bacterial phyla [112]. Within this study, we were able to predict a further 82 DGRs on phage genomes, four of which are thought to be complete. Two of these complete phage genomes are thought to be lytic. In fact, the majority of DGR-containing contigs in this study are thought to be lytic, thus demonstrating that DGRs on phage are far more common than previously found and also observed widely on lytic phages, which has not previously been observed.

Given the prevalence of DGRs, we expected to find evidence of widespread phage tropism switching by occurrence of SNPs in DGR target genes as others have done [112]. Whilst SNPs could be identified in DGR target genes supporting this, many other areas in the same phage genome contained similar levels of variation. This is likely a result of multiple evolutionary pressures and mechanisms that are exerted on a phage genome, with DGRs only one such mechanism of creating variation.

#### CrAss-like phages

Currently, crAss-like phages are classified into four subfamilies and ten genera [47], and found in a variety of environments including human waste [45–47], primate faeces [114], dog faeces [115] and termite guts [95]. Here, we identified a further 18 crAss-like phages, including a near complete genome that belongs to the proposed genus VI [47]. Genus VI is part of the *Betacrassvirinae* subfamily and currently only includes other crAss-like phages occurring within the human gut, including IAS virus that is highly abundant in HIV-1 infected individuals [116]. Thus, we have expanded the environments genus VI crAss-like phages are found in to include non-human hosts. The exact source of these phages is unknown due to the number of possible inputs of the slurry tank. However, the most likely reservoir is from cows, as this is the most abundant input. Unlike its human counterpart IAS virus, which can account for 90% of viral DNA in human faeces [45], crAss-like phages in the slurry tank were only found at low levels (~0.065%).

Phylogenetic analysis clearly demonstrated that human and slurry tank crAss-like phages share a common ancestor, with genetic exchange between them. The direction and route of this exchange is unclear. It may be linked to modern practices of using slurry on arable land used to produce product consumed by humans. Alternatively, it may be transferred from humans to cows via the use of biosolids derived from human waste that are applied to crops that serve as animal feed [117].

#### Auxiliary metabolic genes

We identified a vast array of diverse and abundant AMGs in dairy farm slurry including putative ARGs, CAZymes, ASR genes, MazG, VapE and Zot. Whilst these have all been identified before in viromes from different environments [16, 29, 30, 100, 118–122], this is the first time they have been identified in slurry. The presence of different AMGs is likely a reflection of the unique composition of slurry that has a very high water content combined with organic matter. CAZymes were detected, which have previously been identified in viromes from mangrove soils and the cow rumen where they are thought to participate in the decomposition of organic carbon and boost host energy production during phage infection [16, 123]. Given the high cellulose and hemicellulose content of slurry [124], they likely act in a similar manner within slurry to boost energy for phage replication. As well as involvement in the cycling of carbon, it also appears phage derived genes are involved in sulphur cycling within slurry. Sulfate-reducing bacteria (SRB) are active in animal wastes [125, 126], and sulfate may therefore be limiting within the tank. The ASR



pathway makes sulphur available for incorporation into newly synthesised molecules, such as L-cysteine and L-methionine [127], so the presence of phage encoded ASR genes on both lytic and temperate phages may overcome a metabolic bottleneck in amino acid synthesis. Alternatively, the newly synthesised ASR pathway products may be degraded for energy via the TCA cycle [128].

The AMG *mazG*, that is widespread within marine phages, in particular cyanophages [119, 129, 130], was also found to be abundant. The cyanophage MazG protein was originally hypothesised as a modulator of the host stringent response by altering intracellular levels of (p)ppGpp [131, 132]. However, more recent work found this not to be the case [119]. The identification in a slurry tank suggests this gene is not limited to marine environments and is widespread in different phage types, although its precise role remains to be elucidated.

#### Antibiotic resistance genes

There is ongoing debate as to the importance of phages in the transfer of ARGs [29, 30]. We identified ARGs on ~2% of vOTUs; accounting for ~0.082% of total predicted phage genes from assembled viral contigs. The predicted ARGs were dominated by putative MBLs that contain core motifs and structural similarity with the known bacterial and phage MBLs *bla*<sub>PNGM-1</sub> [102] and *bla*<sub>HRVM-1</sub> [103] respectively. Thus, are likely functionally active, although this remains to be proven. Our estimate of the abundance of ARGs in slurry is lower than earlier reports from other environments that predict an upper estimate of ~0.45% of genes in viromes are ARGs [133, 134]. However, some of these studies have used unassembled reads to estimate abundance [133, 134], whereas we only counted ARGs on contigs that had passed stringent filtering. Our prediction of ~0.082% is similar to more recent estimates of 0.001% to 0.1% in viromes from six different environments that also used assembled viromes [30], suggesting that phages might be an important reservoir of ARGs in slurry.

#### Virulence-associated proteins

The virulence genes *zot* and *vapE* were found to abundant and carried by several vOTUs that were predicted to infect a range of bacterial hosts. The role of *zot* has been well studied in *Vibrio cholerae* and has previously been reported in a range of *Vibrio* and *Campylobacter* prophages [120, 121, 135, 136]. Here, we found *zot* homologues in phages with predicted hosts other than *Vibrio* and *Campylobacter*, further expanding the diversity of phages that carry these genes.

A similar observation was found for the virulence factor *vapE*, which has previously been found in several agricultural pathogens including *Streptococcus* and *Dichelobacter* [97–99]. *VapE* encoded on prophage elements is known to enhance the virulence of *Streptococcus* and is widespread on *Streptococcus* prophages [100]. Whilst the role of *vapE* in virulence has been established, previous work did not demonstrate the mobility of these prophage-like elements. Here, we identified a high quality near-complete phage genome (ctg217) which was remarkably similar to the *vapE* encoding prophage Javan630. Phage Javan630 was originally identified as a prophage within a mastitis causing strain of *Streptococcus uberis* isolated from a dairy cow some 15 years earlier on a dairy farm ~100 mi away [100]. The identification of ctg217 in the free viral fraction indicates that a close relative of phage Javan630 is an active prophage. Along with the numerous other phages encoding *vapE* found in the free virome, it suggests that phage is active in mediating the transfer of *vapE*. The horizontal transfer of *vapE* is of particular concern in the dairy environment where mastitis causing pathogens *Strep. uberis*, *Strep. agalactiae* and *Strep. dysgalactiae* are found [137–139]. Any increase in virulence of these pathogens is detrimental to the dairy industry as it affects both animal welfare and economic viability [140]. *Streptococcus* infections result in mastitic milk, which cannot be sold and is often disposed of into slurry tanks. The continual detection of phages containing *vapE* in slurry suggests a likely continual input, given the regular emptying of the tank. The exact source of phages containing *vapE* cannot be ascertained but is likely cow faeces or mastitic milk. It remains to be determined if the use of slurry as an organic fertiliser contributes to the spread of phage encoded virulence factors and toxins. However, their abundance and presence suggests it is worthy of further investigation.

#### Conclusions

We have demonstrated that using a hybrid approach produces a more complete virome assembly than using short or long-reads alone. Whilst short-reads may underestimate the total viral diversity of a given environment, estimates from long-reads alone were far closer to the hybrid values than short-reads. The use of low input amplified genomic DNA allows the technique to be applied to previously sequenced metagenomes without need for further DNA extraction. We provide a comprehensive analysis of the slurry virome, demonstrating that the virome contains a diverse and stable viral community dominated by lytic viruses of novel genera. Functional annotation revealed a diverse and abundant range of AMGs including virulence factors, toxins and antibiotic resistance genes, suggesting that phages may play

a significant role in mediating the transfer of these genes and augmenting both the virulence and antibiotic resistance of their hosts.

### Abbreviations

(p)ppGpp: Guanosine tetraphosphate or guanosine pentaphosphate; AMG: Auxiliary metabolic gene; ANI: Average nucleotide identity; ARG: Antimicrobial resistance gene; ASR: Assimilatory sulfate reduction; BOD: Biological oxygen demand; CAZYme: Carbohydrate-active enzyme; CDS: Coding sequence; COG: Clusters of orthologous groups; DEFRA: Department for Environment, Food and Rural Affairs; DGR: Diversity-generating retroelement; eggNOG: Evolutionary genealogy of genes: Non-supervised Orthologous Groups; HGT: Horizontal gene transfer; HMW: High molecular weight; LASL: Linker-amplified shotgun library; MAG: Metagenome assembled genome; MazG: Nucleoside triphosphate pyrophosphohydrolase; MBL: Metallo-beta-lactamase; MNP: Multiple nucleotide polymorphism; ONT: Oxford Nanopore Technologies; ORF: Open reading frame; PAPS: 3'-Phosphoadenosine-5'-phosphosulfate; pVOG: Prokaryotic virus orthologous groups; SNP: Single nucleotide polymorphism; SRB: Sulfate-reducing bacteria; TerL: Terminase large subunit; VapE: Virulence-associated protein E; VC: Viral cluster; VLP: Virus-like particle; vOTU: Viral operational taxonomic unit; Zot: *Zona occludens* toxin

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40168-021-01010-3>.

**Additional file 1: Supplementary Table 1.** ViromeQC enrichment scores of Illumina viromes. **Supplementary Table 2.** Genes found to be under positive selection and their predicted function. **Supplementary Table 3.** Predicted functions of putative phage-encoded AMGs. **Supplementary Table 4.** Predicted host taxa for vOTUs. **Supplementary Table 5.** Relative abundance of vOTUs in each sample, alongside vCONTACT2 cluster and predicted lifestyle. **Supplementary Table 6.** Mapping statistics for active prophage vOTUs that were used to infer the ends of prophage sequences. **Supplementary Table 7.** Mapping statistics for active prophage vOTUs that were used to infer the ends of prophage sequences for which at least one end could be predicted.

**Additional file 2: Supplementary Figure 1.** Representative figure for the identification of prophage ends. Reads were mapped against vOTU k87\_12210044 at 95 % identity threshold, the median coverage was calculated for 500 bp windows with the cutoff value calculated as median coverage minus (2 \* standard deviations of median coverage) and plotted in orange. In this particular example, only one end was predicted. **Supplementary Figure 2.** Predicted hosts of viral contigs at the phylum level. Predicted hosts were obtained using WISH. The relative abundance of phages predicted to infect different hosts was calculated by stringent mapping of reads to each viral contig as normalising for contig length and sequencing depth as described in materials and methods. **Supplementary Figure 3.** Phylogeny of crAss-like vOTUs based upon the method of Guerin et al. [47]. Phylogeny of four genes that encode a primase, terminase, portal protein and major capsid protein. The analysis followed the same method as described by Guerin et al. [47], with the ten major clades as previously defined marked. Bootstrap values >70% are marked by a circle. **Supplementary Figure 4.** Functional classification of viral proteins into COG categories by eggNOG mapping. **Supplementary Figure 5.** Phylogeny of putative metallo-β-lactamases. The phylogeny was built on the alignment of the amino acid sequences that were aligned by MAFFT. A WAG model of evolution was used in IQ-TREE with 1000 bootstraps. Putative MBLs identified in the slurry tank are marked in orange, along with a previously experimentally validated phage-encoded MBL (yellow). Bacterial subclass B1 (green), B2 (blue), B3 (red) MBLs are also marked. Bootstrap values >70% are marked by a circle. **Supplementary Figure 6.** Phylogeny of phage genomes that contain a complete DGR. Phylogeny was constructed from the amino acid sequence of TerL protein that were aligned in MAFFT and phylogeny constructed with IQ-TREE with a WAG model of evolution and 1000 bootstraps. Bootstrap values >70% are marked by a circle. Different

viral families are differentiated by the coloured ring around the outside of the tree.

### Acknowledgements

Not applicable.

### Authors' contributions

R.C, M.J, A.M, J.H, C.D and D.S conceived the study. S.H and E.K carried out experiments and collected the data. R.C, S.H, U.T and A.M carried out the bioinformatic analysis. R.C, M.J and A.M drafted the manuscript. All authors approved and contributed to the final manuscript.

### Funding

R.C is supported by a scholarship from the Medical Research Foundation National PhD Training Programme in Antimicrobial Resistance Research (MRF-145-0004-TPG-AVISO). J.H, S.H, A.M, D.S, C.D, M. J are funded by NERC AMR-EVALFARMS (NE/N019881/1). PromethION sequencing was funded by a NERC-NBAF award to AM SH. A.M is funded by MRC (MR/L015080/1 and MR/T030062/1). Bioinformatics analysis was carried out on infrastructure provided by MRC-CLIMB (MR/L015080/1).

### Availability of data and materials

Reads and assemblies from Illumina and PromethION virome sequencing were submitted to the ENA under the study PRJEB38990. Annotated vOTUS included within the analysis, along with sequence alignments and trees from phylogenetic analysis are provided via Figshare (10.25392/leicester.data.13061633)

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington Campus, College Road, Loughborough, Leicestershire LE12 5RD, UK. <sup>2</sup>School of Biosciences, University of Nottingham, Sutton Bonington Campus, College Road, Loughborough, Leicestershire LE12 5RD, UK. <sup>3</sup>Edinburgh Genomics, School of Biological Sciences, University of Edinburgh, Charlotte Auerbach Road, Edinburgh EH9 3FL, UK. <sup>4</sup>Dept Genetics and Genome Biology, University of Leicester, University Road, Leicester, Leicestershire LE1 7RH, UK.

Received: 28 October 2020 Accepted: 2 February 2021

Published online: 20 March 2021

### References

- Cobián Güemes AG, Youle M, Cantú VA, Felts B, Nulton J, Rohwer F. Viruses as winners in the game of life. *Annu Rev Virol.* 2016;3:197–214 Annual Reviews.
- Bohannan BJM, Lenski RE. Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage. *Ecol Lett.* 2000;3:362–77. John Wiley & Sons, Ltd.
- Buckling A, Rainey PB. Antagonistic coevolution between a bacterium and a bacteriophage. *Proc R Soc B Biol Sci.* 2002;269:931–6 The Royal Society.
- Canchaya C, Fournous G, Chibani-Chennoufi S, Dillmann ML, Brüssow H. Phage as agents of lateral gene transfer. *Curr Opin Microbiol.* 2003;6:417–24 Elsevier Ltd.
- Breitbart M, Bonnain C, Malki K, Sawaya NA. Phage puppet masters of the marine microbial realm. *Nat Microbiol.* 2018;3:754–66. Nature Publishing Group.
- Clokier MR, Millard AD, Letarov AV, Heaphy S. Phages in nature. *Bacteriophage.* 2011;1:31–45 Landes Bioscience.

7. Sutton TDS, Hill C. Gut bacteriophage: Current understanding and challenges. *Front Endocrinol (Lausanne)*. 2019;10:784 Frontiers Media S.A.
8. Paez-Espino D, Elie-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, et al. Uncovering Earth's virome. *Nature*. 2016;536:425–30 Nature Publishing Group.
9. Gregory AC, Zayed AA, Conceição-Neto N, Temperton B, Bolduc B, Alberti A, et al. Marine DNA viral macro- and microdiversity from Pole to Pole. *Cell*. 2019;177:1109–1123.e14.
10. Hurwitz BL, U'Ren JM. Viral metabolic reprogramming in marine ecosystems. *Curr Opin Microbiol*. 2016;31:161–8 Elsevier Ltd.
11. Yooshep S, Sutton G, Rusch DB, Halpern AL, Williamson SJ, Remington K, et al. The Sorcerer II global ocean sampling expedition: expanding the universe of protein families. *PLoS Biol*. 2007;5:0432–66 Public Library of Science.
12. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature*. 2016;537:689–93 Nature Publishing Group.
13. Anantharaman K, Duhaime MB, Breier JA, Wendt KA, Toner BM, Dick GJ. Sulfur oxidation genes in diverse deep-sea viruses. *Science (80- )*. 2014;344:757–60 American Association for the Advancement of Science.
14. Zhang R, Wei W, Cai L. The fate and biogeochemical cycling of viral elements. *Nat Rev Microbiol*. 2014;12:850–1 Nature Publishing Group.
15. York A. Marine microbiology: Algal virus boosts nitrogen uptake in the ocean. *Nat Rev Microbiol*. 2017;15:573 Nature Publishing Group.
16. Jin M, Guo X, Zhang R, Qu W, Gao B, Zeng R. Diversities and potential biogeochemical impacts of mangrove soil viruses. *Microbiome*. 2019;7:1–15 BioMed Central Ltd.
17. Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brulc JM, et al. Functional metagenomic profiling of nine biomes. *Nature*. 2008;452:629–32 Nature Publishing Group.
18. Sharon I, Battchikova N, Aro EM, Giglione C, Meinel T, Glaser F, et al. Comparative metagenomics of microbial traits within oceanic viral communities. *ISME J*. 2011;5:1178–90 Nature Publishing Group.
19. Hurwitz BL, Hallam SJ, Sullivan MB. Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biol*. 2013;14:R123 BioMed Central.
20. Hurwitz BL, Brum JR, Sullivan MB. Depth-stratified functional and taxonomic niche specialization in the “core” and “flexible” Pacific Ocean Virome. *ISME J*. 2015;9:472–84 Nature Publishing Group.
21. Monier A, Chambouvet A, Milner DS, Attah V, Terrado R, Lovejoy C, et al. Host-derived viral transporter protein for nitrogen uptake in infected marine phytoplankton. *Proc Natl Acad Sci U S A*. 2017;114:E7489–98 National Academy of Sciences.
22. Freeman VJ. Studies on the virulence of bacteriophage-infected strains of *Corynebacterium diphtheriae*. *J Bacteriol*. 1951;61:675–88 American Society for Microbiology (ASM).
23. Eklund MW, Poysky FT, Meyers JA, Pelroy GA. Interspecies conversion of clostridium botulinum type C to Clostridium novyi type A by bacteriophage. *Science (80- )*. 1974;186:456–8 Science.
24. Waldor MK, Mekalanos JJ. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science (80- )*. 1996;272:1910–3 American Association for the Advancement of Science (AAAS).
25. Wagner PL, Livny J, Neely MN, Acheson DWK, Friedman DI, Waldor MK. Bacteriophage control of Shiga toxin 1 production and release by *Escherichia coli*. *Mol Microbiol*. 2002;44:957–70 John Wiley & Sons, Ltd.
26. Khalil RKS, Skinner C, Patfield S, He X. Phage-mediated Shiga toxin (Stx) horizontal gene transfer and expression in non-Shiga toxigenic *Enterobacter* and *Escherichia coli* strains. *Pathog Dis*. 2016;74:1–11.
27. Fortier LC, Sekulovic O. Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence*. 2013;4:354–65 Taylor and Francis Inc.
28. Balcázar JL. Implications of bacteriophages on the acquisition and spread of antibiotic resistance in the environment. *Int Microbiol*. 2020;23:475–9. Springer.
29. Enault F, Briet A, Bouteille L, Roux S, Sullivan MB, Petit MA. Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. *ISME J*. 2017;11:237–47 Nature Publishing Group.
30. Debroas D, Siguret C. Viruses as key reservoirs of antibiotic resistance genes in the environment. *ISME J*. 2019;13:2856–67 Nature Publishing Group.
31. AHDB. UK and EU cow numbers [Internet]. 2018 [cited 2020 Jun 19]. Available from: <https://ahdb.org.uk/dairy/uk-and-eu-cow-numbers>
32. Font-Palma C. Methods for the Treatment of Cattle Manure—A Review. *C MDPI AG*. 2019;5:27.
33. Smith KA, Williams AG. Production and management of cattle manure in the UK and implications for land application practice. Nicholson F, editor. *Soil Use Manag*. 2016;32:73–82 Blackwell Publishing Ltd.
34. AHDB. Cost effective slurry storage strategies [Internet]. Available from: <https://dairy.ahdb.org.uk/resources-library/technical-information/health-welfare/cost-effective-slurry-storage-strategies/#.XvCQompKjwd>.
35. De Vries JW, Groenestein CM, De Boer IJM. Environmental consequences of processing manure to produce mineral fertilizer and bio-energy. *J Environ Manage*. 2012;102:173–83.
36. Prapasongsa T, Christensen P, Schmidt JH, Thrane M. LCA of comprehensive pig manure management incorporating integrated technology systems. *J Clean Prod*. 2010;18:1413–22.
37. Sandars DL, Audsley E, Cañete C, Cumby TR, Scotford IM, Williams AG. Environmental benefits of livestock manure management practices and technology by life cycle assessment. *Biosyst Eng*. 2003;84:267–81.
38. Thomassen MA, van Calker KJ, Smits MCJ, Iepema GL, de Boer IJM. Life cycle assessment of conventional and organic milk production in the Netherlands. *Agric Syst*. 2008;96:95–107.
39. UK Government. Nitrate Vulnerable Zones (NVZs). *Eur Comm Nitrates Dir*. 2013;1996 [cited 2020 Jun 19]. Available from: <https://www.gov.uk/government/collections/nitrate-vulnerable-zones>.
40. UK Government. Use organic manures and manufactured fertilisers on farmland [Internet]. [cited 2020 Jul 14]. Available from: <https://www.gov.uk/government/publications/nitrates-and-phosphates-plan-organic-fertiliser-and-manufactured-fertiliser-use/use-organic-manures-and-manufactured-fertilisers-on-farmland>.
41. Besler I, Sazinas P, Harrison C, Gannon L, Redgwell T, Michniewski S, et al. Genome sequence and characterization of Coliphage vB\_Eco\_SLUR29. *PHAGE*. 2020;1:38–44 Mary Ann Liebert Inc.
42. Sazinas P, Redgwell T, Rihtman B, Grigonyte A, Michniewski S, Scanlan DJ, et al. Comparative genomics of bacteriophage of the genus Seuratvirus. *Genome Biol Evol*. 2018;10:72–6 Oxford University Press.
43. Smith R, O'Hara M, Hobman JL, Millard AD. Draft genome sequences of 14 *Escherichia coli* phages isolated from cattle slurry. *Genome Announc*. 2015;3:e01364–15 American Society for Microbiology.
44. Brum JR, Cesar Ignacio-Espinoza J, Roux S, Doulcier G, Acinas SG, Alberti A, et al. Patterns and ecological drivers of ocean viral communities. *Science*. 2015;348:1261498. American Association for the Advancement of Science.
45. Dutilh BE, Cassman N, McNair K, Sanchez SE, Silva GGG, Boling L, et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat Commun*. 2014;5:4498.
46. Shkoporov AN, Khokhlova EV, Fitzgerald CB, Stockdale SR, Draper LA, Ross RP, et al. ΦCrAss001 represents the most abundant bacteriophage family in the human gut and infects *Bacteroides intestinalis*. *Nat Commun*. 2018;9:1–8 Nature Publishing Group.
47. Guerin E, Shkoporov A, Stockdale SR, Clooney AG, Ryan FJ, Sutton TDS, et al. Biology and taxonomy of crAss-like bacteriophages, the most abundant virus in the human gut. *Cell Host Microbe*. 2018;24:653–664.e6 Cell Press.
48. Zhao Y, Temperton B, Thrash JC, Schwalbach MS, Vergin KL, Landry ZC, et al. Abundant SAR11 viruses in the ocean. *Nature*. 2013;494:357–60 Nature Publishing Group.
49. Martinez-Hernandez F, Fornas Ö, Lluésma Gomez M, Garcia-Heredia I, Maestre-Carballea L, López-Pérez M, et al. Single-cell genomics uncover Pelagibacter as the putative host of the extremely abundant uncultured 37-F6 viral population in the ocean. *ISME J*. 2019;13:232–6 2018/09/18. Nature Publishing Group UK.
50. Olson ND, Treangen TJ, Hill CM, Cepeda-Espinoza V, Ghurye J, Koren S, et al. Metagenomic assembly through the lens of validation: recent advances in assessing and improving the quality of genomes assembled from metagenomes. *Brief Bioinform*. 2019;20:1140–50 Oxford University Press.
51. Temperton B, Giovannoni SJ. Metagenomics: microbial diversity through a scratched lens. *Curr Opin Microbiol*. 2012;15:605–12. Elsevier Current Trends.
52. Mizuno CM, Ghai R, Rodriguez-Valera F. Evidence for metaviromic islands in marine phages. *Front Microbiol*. 2014;5:27. Frontiers Research Foundation.
53. Roux S, Emerson JB, Elie-Fadrosh EA, Sullivan MB. Benchmarking viromics: An in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ*. 2017;2017:e3817.
54. Watson M, Warr A. Errors in long-read assemblies can critically affect protein prediction. *Nat Biotechnol*. 2019;37:124–6.
55. Buck D, Weirather JL, de Cesare M, Wang Y, Piazza P, Sebastiano V, et al. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore

- Technologies and their applications to transcriptome analysis. *F1000Research*. 2017;6:100. Faculty of 1000 Ltd.
56. Warwick-Dugdale J, Solonenko N, Moore K, Chittick L, Gregory AC, Allen MJ, et al. Long-read viral metagenomics captures abundant and microdiverse viral populations and their niche-defining genomic islands. *PeerJ*. 2019;7:e6800.
57. Beaulaurier J, Luo E, Eppley JM, Den Uyl P, Dai X, Burger A, et al. Assembly-free single-molecule sequencing recovers complete virus genomes from natural microbial communities. *Genome Res*. 2020;30:437–46 Cold Spring Harbor Laboratory Press.
58. Sazinas P, Michniewski S, Rihtman B, Redgwell T, Grigonyte A, Brett A, et al. Metagenomics of the viral community in three cattle slurry samples. *Microbiol Resour Announc*. 2019;8:e01442–18 Am Soc Microbiol.
59. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
60. Li D, Luo R, Liu CM, Leung CM, Ting HF, Sadakane K, et al. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods*. 2016;102:3–11 Academic Press Inc.
61. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34:3094–100.
62. Zolfo M, Pinto F, Asnicar F, Manghi P, Tett A, Bushman FD, et al. Detecting contamination in viromes using ViromeQC. *Nat Biotechnol*. 2019;37:1408–12 Nature Research.
63. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biol*. 2016;17:132 BioMed Central Ltd.
64. O'Leary NA, Wright MW, Brister JR, Ciufu S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 2016;44:D733–45.
65. Kieft K, Zhou Z, Anantharaman K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome*. 2020;8:90 Cold Spring Harbor Laboratory.
66. Ren J, Song K, Deng C, Ahlgren NA, Fuhrman JA, Li Y, et al. Identifying viruses from metagenomic data using deep learning. *Quant Biol*. 2020;8:64–77 Higher Education Press.
67. Graziotin AL, Koonin EV, Kristensen DM. Prokaryotic Virus Orthologous Groups (pVOGs): A resource for comparative genomics and protein family annotation. *Nucleic Acids Res*. 2017;45:D491–8 Oxford University Press.
68. HMMER [Internet]. [cited 2020 May 29]. Available from: <http://hmmer.org/>.
69. Bushnell B. BBMap download | SourceForge.net [Internet]. 2013 [cited 2020 May 29]. Available from: <https://sourceforge.net/projects/bbmap/>.
70. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*. 2014;9:e112963.
71. GitHub - simroux/ClusterGenomes: archive for ClusterGenomes scripts [Internet]. [cited 2020 May 29]. Available from: <https://github.com/simroux/ClusterGenomes>.
72. Nayfach S, Pedro Camargo A, Eloe-Fadrosh E, Roux S. CheckV: assessing the quality of metagenome-assembled viral genomes. *bioRxiv*. 2020; 2020.05.06.081778. Cold Spring Harbor Laboratory. Available from: <https://doi.org/10.1101/2020.05.06.081778>.
73. McMurdie PJ, Holmes S. Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. Watson M, editor. *PLoS One*. 2013;8:e61217 Public Library of Science.
74. Team RC. R: A language and environment for statistical computing [Internet]. Vienna: R Foundation for Statistical Computing; 2018. Available from: <https://www.r-project.org/>.
75. Shaw LM, Blanchard A, Chen Q, An X, Davies P, Töttemeyer S, et al. DirtyGenes: testing for significant changes in gene or bacterial population compositions from a small number of samples. *Sci Rep*. 2019;9:1–10 Nature Publishing Group.
76. Seemann T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30:2068–9.
77. Michniewski S, Redgwell T, Grigonyte A, Rihtman B, Aguilo-Ferretjans M, Christie-Oleza J, et al. Riding the wave of genomics to investigate aquatic coliphage diversity and activity. *Environ Microbiol*. 2019;21:2112–28 2019/04/04. John Wiley & Sons, Inc.
78. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. 2012;28:3150–2 2012/10/11. Oxford University Press.
79. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res*. 2018;47:D309–14.
80. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res*. 2018;46:W95–101.
81. Yan F, Yu X, Duan Z, Lu J, Jia B, Qiao Y, et al. Discovery and characterization of the evolution, variation and functions of diversity-generating retroelements using thousands of genomes and metagenomes. *BMC Genomics*. 2019;20:595 BioMed Central Ltd.
82. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078–9 2009/06/08. Oxford University Press.
83. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012;22:568–76 2012/02/02. Cold Spring Harbor Laboratory Press.
84. Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat Biotechnol*. 2019;37:632–9 Nature Publishing Group.
85. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13:2498–504 Cold Spring Harbor Laboratory Press.
86. Galiez C, Siebert M, Enault F, Vincent J, Söding J. WISH: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. *Bioinformatics*. 2017;33:3113–4.
87. Sherrill-Mix S. taxonomizr: Functions to Work with NCBI Accessions and Taxonomy. R Packag version 051 [Internet]. 2018 [cited 2020 May 29]. Available from: <https://cran.r-project.org/web/packages/taxonomizr/>.
88. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. *Nucleic Acids Res*. 2019;47:D427–32 Oxford University Press.
89. Seemann T. snippy: Rapid haploid variant calling and core genome alignment [Internet]. 2015. [cited 2020 May 29]. Available from: <https://github.com/tseemann/snippy>.
90. Adriaenssens EM, Rodney BJ. How to name and classify your phage: an informal guide. *Viruses*. 2017;9:1–9.
91. Kim M, Wells JE. A meta-analysis of bacterial diversity in the feces of cattle. *Curr Microbiol*. 2016;72:145–51 Springer New York LLC.
92. Delgado B, Bach A, Guasch I, González C, Elcso G, Pryce JE, et al. Whole rumen metagenome sequencing allows classifying and predicting feed efficiency and intake levels in cattle. *Sci Rep*. 2019;9:1–13 Nature Publishing Group.
93. Li F, Hitch TCA, Chen Y, Creevey CJ, Guan LL. Comparative metagenomic and metatranscriptomic analyses reveal the breed effect on the rumen microbiome and its associations with feed efficiency in beef cattle 06 Biological Sciences 0604 Genetics 06 Biological Sciences 0605 Microbiology. *Microbiome*. 2019;7:6 BioMed Central Ltd.
94. Shkoporov AN, Clooney AG, Sutton TDS, Ryan FJ, Daly KM, Nolan JA, et al. The human gut virome is highly diverse, stable, and individual specific. *Cell Host Microbe*. 2019;26:527–541.e5 Elsevier Inc.
95. Yutin N, Makarova KS, Gussow AB, Krupovic M, Segall A, Edwards RA, et al. Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut. *Nat Microbiol*. 2018;3:38–46 2017/11/13.
96. Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res*. 2000;28:33–6 Oxford University Press.
97. Billington SJ, Johnston JL, Rood JL. Virulence regions and virulence factors of the ovine footrot pathogen, *Dichelobacter nodosus*. *FEMS Microbiol Lett*. 1996;145:147–56.
98. Bloomfield GA, Whittle G, McDonagh MB, Katz ME, Cheetham BF. Analysis of sequences flanking the vap regions of *Dichelobacter nodosus*: evidence for multiple integration events, a killer system, and a new genetic element. *Microbiology*. 1997;143:553–62 Microbiology Society.
99. Ji X, Sun Y, Liu J, Zhu L, Guo X, Lang X, et al. A novel virulence-associated protein, vapE, in *Streptococcus suis* serotype 2. *Mol Med Rep*. 2016;13:2871–7 Spandidos Publications.



100. Rezaei Javan R, Ramos-Sevillano E, Akter A, Brown J, Brueggemann AB. Prophages and satellite prophages are widespread in *Streptococcus* and may play a role in pneumococcal pathogenesis. *Nat Commun*. 2019;10:1–14 Nature Publishing Group.
101. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015;10: 845–58 Nature Publishing Group.
102. Park KS, Kim TY, Kim JH, Lee JH, Jeon JH, Karim AM, et al. PNGM-1, a novel subclass B3 metallo- $\beta$ -lactamase from a deep-sea sediment metagenome. *J Glob Antimicrob Resist*. 2018;14:302–5 Elsevier Ltd.
103. Moon K, Jeon JH, Kang I, Park KS, Lee K, Cha CJ, et al. Freshwater viral metagenome reveals novel and functional phage-borne antibiotic resistance genes. *Microbiome*. 2020;8:75 NLM (Medline).
104. Liu M, Deora R, Doulatov SR, Gingery M, Eiserling FA, Preston A, et al. Reverse transcriptase-mediated tropism switching in Bordetella bacteriophage. *Science* (80- ). 2002;295:2091–4.
105. Roux S, Solonenko NE, Dang VT, Poulos BT, Schwenck SM, Goldsmith DB, et al. Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. *PeerJ*. 2016;4:e2777 Available from: <https://pubmed.ncbi.nlm.nih.gov/28003936>. PeerJ Inc.
106. Garmaeva S, Sinha T, Kurilshikov A, Fu J, Wijmenga C, Zhernakova A. Studying the gut virome in the metagenomic era: challenges and perspectives. *BMC Biol*. 2019;17:1–14 BioMed Central Ltd.
107. Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, et al. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature*. 2010;466:334–8 NIH Public Access.
108. Minot S, Bryson A, Chehoud C, Wu GD, Lewis JD, Bushman FD. Rapid evolution of the human gut virome. *Proc Natl Acad Sci U S A*. 2013;110: 12450–5 2013/07/08. National Academy of Sciences.
109. Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, et al. The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Res*. 2011;21:1616–25.
110. Lim ES, Zhou Y, Zhao G, Bauer IK, Droit L, Ndao IM, et al. Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nat Med*. 2015;21:1228–34 Nature Publishing Group.
111. Moreno-Gallego JL, Chou SP, Di Rienzi SC, Goodrich JK, Spector TD, Bell JT, et al. Virome diversity correlates with intestinal microbiome diversity in adult monozygotic twins. *Cell Host Microbe*. 2019;25:261–272.e5 Cell Press.
112. Benler S, Cobián-Güemes AG, McNair K, Hung SH, Levi K, Edwards R, et al. A diversity-generating retroelement encoded by a globally ubiquitous Bacteroides phage 06 Biological Sciences 0605 Microbiology. *Microbiome*. 2018;6:1–10.
113. Wu L, Gingery M, Abebe M, Arambula D, Czornyj E, Handa S, et al. Diversity-generating retroelements: natural variation, classification and evolution inferred from a large-scale genomic survey. *Nucleic Acids Res*. 2018;46:11–24.
114. Edwards RA, Vega AA, Norman HM, Ohaeri M, Levi K, Dinsdale EA, et al. Global phylogeography and ancient evolution of the widespread human gut virus crAssphage. *Nat Microbiol*. 2019;4:1727–36.
115. Cuscó A, Salas A, Torre C, Francino O. Shallow metagenomics with Nanopore sequencing in canine fecal microbiota improved bacterial taxonomy and identified an uncultured CrAssphage. *bioRxiv*. 2019:1–12 Available from: <https://doi.org/10.1101/585067>. Cold Spring Harbor Laboratory.
116. Oude Munnink BB, Canuti M, Deijis M, de Vries M, Jebbink MF, Rebers S, et al. Unexplained diarrhoea in HIV-1 infected individuals. *BMC Infect Dis*. 2014;14:22 BioMed Central.
117. Biosolids Assurance Scheme. ABOUT BIOSOLIDS : assured biosolids [Internet]. 2020 [cited 2020 Jul 22]. Available from: <https://assuredbiosolids.co.uk/about-biosolids/>.
118. Gao SM, Schippers A, Chen N, Yuan Y, Zhang MM, Li Q, et al. Depth-related variability in viral communities in highly stratified sulfidic mine tailings. *Microbiome*. 2020;8:89 BioMed Central.
119. Rihman B, Bowman-Grahl S, Millard A, Corrigan RM, Clokie MRJ, Scanlan DJ. Cyanophage MazG is a pyrophosphohydrolase but unable to hydrolyse magic spot nucleotides. *Environ Microbiol Rep*. 2019;11:448–55 Wiley-Blackwell.
120. Liu F, Lee H, Lan R, Zhang L. Zonula occludens toxins and their prophages in *Campylobacter* species. *Gut Pathog*. 2016;8:43 BioMed Central Ltd.
121. Castillo D, Pérez-Reytor D, Plaza N, Ramírez-Araya S, Blondel CJ, Corsini G, et al. Exploring the genomic traits of non-toxicogenic *Vibrio parahaemolyticus* strains isolated in southern Chile. *Front Microbiol*. 2018;9:161 Frontiers Media S.A.
122. Romero P, Croucher NJ, Hiller NL, Hu FZ, Ehrlich GD, Bentley SD, et al. Comparative genomic analysis of ten *Streptococcus pneumoniae* temperate bacteriophages. *J Bacteriol*. 2009;191:4854–62 American Society for Microbiology (ASM).
123. Anderson CL, Sullivan MB, Fernando SC. Dietary energy drives the dynamic response of bovine rumen viral communities. *Microbiome*. 2017;5:155 Available from: <https://pubmed.ncbi.nlm.nih.gov/29179741>. BioMed Central.
124. Chen S, Liao W, Liu C, Wen Z, Kincaid RL, Harrison JH, et al. Value-Added Chemicals from Animal Manure. *Pacific Northwest Natl Lab*. 2003;PNNL-14495:1–142.
125. Cook KL, Whitehead TR, Spence C, Cotta MA. Evaluation of the sulfate-reducing bacterial population associated with stored swine slurry. *Anaerobe*. 2008;14:172–80 Academic Press.
126. St-Pierre B, Wright ADG. Implications from distinct sulfate-reducing bacteria populations between cattle manure and digestate in the elucidation of H<sub>2</sub>S production during anaerobic digestion of animal slurry. *Appl Microbiol Biotechnol*. 2017;101:5543–56 [cited 2020 Jul 22]. Available from: <https://link.springer.com/article/10.1007/s00253-017-8261-1>. Springer Verlag.
127. Rückert C. Sulfate reduction in microorganisms—recent advances and biotechnological applications. *Curr Opin Microbiol*. 2016;33:140–6. Elsevier Ltd.
128. Howard-Varona C, Lindback MM, Bastien GE, Solonenko N, Zayed AA, Jang HB, et al. Phage-specific metabolic reprogramming of virocells. *ISME J*. 2020; 14:881–95 [cited 2020 Aug 9]. Springer Nature. Available from: <https://doi.org/10.1038/s41396-019-0580-z>.
129. Sullivan MB, Huang KH, Ignacio-Espinoza JC, Berlin AM, Kelly L, Weigle PR, et al. Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ Microbiol*. 2010;12:3035–56 Blackwell Publishing Ltd.
130. Millard AD, Zwirgmaier K, Downey MJ, Mann NH, Scanlan DJ. Comparative genomics of marine cyanomyoviruses reveals the widespread occurrence of *Synechococcus* host genes localized to a hyperplastic region: implications for mechanisms of cyanophage evolution. *Environ Microbiol*. 2009;11:2370–87 John Wiley & Sons, Ltd.
131. Clokie MRJ, Mann NH. Marine cyanophages and light. *Environ Microbiol*. 2006;8:2074–82 [Internet]. John Wiley & Sons, Ltd. Available from: <https://doi.org/10.1111/j.1462-2920.2006.01171.x>.
132. Clokie MRJ, Millard AD, Mann NH. T4 genes in the marine ecosystem: studies of the T4-like cyanophages and their role in marine ecology. *Virology*. 2010;7:291 Available from: <https://pubmed.ncbi.nlm.nih.gov/21029435>. BioMed Central.
133. Balcazar JL. Bacteriophages as vehicles for antibiotic resistance genes in the environment. *PLoS Pathog*. 2014;10:e1004219 Available from: <https://pubmed.ncbi.nlm.nih.gov/25078987>. Public Library of Science.
134. Lekunberri I, Subirats J, Borrego CM, Balcazar JL. Exploring the contribution of bacteriophages to antibiotic resistance. *Environ Pollut*. 2017;220:981–4 Elsevier Ltd.
135. Koonin EV. The second cholera toxin, Zot, and its plasmid-encoded and phage-encoded homologues constitute a group of putative ATPases with an altered purine NTP-binding motif. *FEBS Lett*. 1992;312:3–6 No longer published by Elsevier.
136. Schmidt E, Kelly SM, van der Walle CF. Tight junction modulation and biochemical characterisation of the zonula occludens toxin C-and N-termini. *FEBS Lett*. 2007;581:2974–80 No longer published by Elsevier.
137. Zadoks RN, Middleton JR, McDougall S, Katholm J, Schukken YH. Molecular epidemiology of mastitis pathogens of dairy cattle and comparative relevance to humans. *J Mammary Gland Biol Neoplasia*. 2011;16:357–72 Springer.
138. Whist AC, Østerås O, Sølverød L. *Streptococcus dysgalactiae* isolates at calving and lactation performance within the same lactation. *J Dairy Sci*. 2007;90:766–78 American Dairy Science Association.
139. Keefe GP. *Streptococcus agalactiae* mastitis: a review. *Can Vet J*. 1997;38: 429–37 Canadian Veterinary Medical Association.
140. Ruegg PL, Petersson-Wolfe CS. Mastitis in dairy cows [Internet]. *Vet Clin North Am Food Anim Pract*. 2018;ix–x [cited 2020 Jun 8]. Available from: <https://dairy.ahdb.org.uk/technical-information/animal-health-welfare/mastitis/Xt5XWZ5Kjwc>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.