Viral dark matter in the gut virome of elderly humans

Stephen R. Stockdale*[,1,2], Feargal J. Ryan*[,1], Angela McCann[1], Marion Dalmasso[1,†], Paul R. Ross[1,2,3], Colin Hill[1,3,‡]

[1]APC Microbiome Ireland, University College Cork, Co. Cork, Ireland

[2]Teagasc Food Research Centre, Moorepark, Fermoy, Co. Cork, Ireland

[3]School of Microbiology, University College Cork, Cork, Ireland

†Present address: Normandie University, UNICAEN, ABTE, 14000 Caen, France

*These authors contributed equally to this work.

Keywords: Human Virome; Human Microbiome; Bacteriophage; Elderly Adults

‡Corresponding author: Colin Hill

c.hill@ucc.ie

Abstract

The human virome is an area of increasing interest with relation to human health and disease. It has been demonstrated to alter in concert with the bacterial microbiome in early life and was also found to be different in patients with certain diseases such as inflammatory bowel disease. However, all virome analyses are hampered by a lack of annotated representative database sequences, often referred to as the 'viral dark matter'. Here we provide the first description of the gut DNA virome in elderly individuals (>65 years old) as well as the description of novel bacteriophages not present in current reference databases. Diversity analysis comparing elderly persons from different residence locations (community living vs long term care facilities) did not reveal any difference in their virome diversity profiles despite the reported differences at the bacteriome level. An abundance of M*icroviridae* of the subfamily *Gokushovirinae* were present in the faeces of elderly individuals. Several novel members of the order *Caudovirales* were also characterized and annotated. Assignment of host bacteria to detected viral genomes was attempted using a combination of CRISPR spacers, tRNA genes and a probabilistic approach. Further characterization of the viral dark matter is necessary for developing tools and expanding databases to study the human virome. This study focused on the virome of an aging human cohort with the goal of illuminating part of the viral dark matter.

Introduction

The human microbiome has been an area of growing research in the past decade and many diseases have associated microbial alterations which researchers hope will lead to diagnostics, disease sub-typing or even the identification of the pathology for certain conditions. Faecal Microbiota Transplantations (FMT) have been a highly effective treatment for recurrent *Clostridium difficile* infections and to date is the best example of a microbiome-based therapy [1,2]. However, microbiome research is often hampered by our lack of knowledge about many of the resident microbes. The Human Microbiome Project (HMP) has tackled this problem, firstly by providing a reference set of genomes of bacteria inhabiting the body and secondly by providing in-depth shotgun metagenomic data [3,4]. Despite this, a recent analysis from the expanded HMP found that many abundant members of the gut bacteriome diverged considerably from the closest available reference and certain clades entirely lacked reference genomes [5]. This lack of database representatives is amplified for the human virome, given the comparatively small number of studies and lack of dedicated projects such as the HMP. Thus, studies to date have found human viromes to be composed predominantly of novel viral sequences, referred to as “viral dark matter” [6,7].

Despite this, research on the human virome has proceeded by utilizing a combination of database-dependent and *de novo* assembly based approaches [8]. Early work examining viral communities examined uncultured marine and human faecal communities [9,10]; however, these studies were conducted at a far lower sequencing depth than that currently available with modern technologies. More recently, Norman *et al.* found an expansion of bacteriophage *Caudovirales* richness is associated with IBD [11] and Lim *et al.* found that the first 2 years of life is associated with a contraction of bacteriophage diversity matched by an expansion of bacteria

[12]. Manrique *et al.* identified a set of core bacteriophages present in the gut microbiome of healthy individuals [13]. McCann and colleagues (2018) found differences in the virome diversity of 1 year old infants born by spontaneous vaginal delivery or caesarean section [14]. The lack of database representation for members of the gut virome is epitomized by the discovery of the sequence of CrAssphage by Dutilh *et al.* (2014) and also showed that it is the most abundant virus in the human gut [15]. This work was expanded on by Yutin *et al.* (2017) who showed that crAssphage is just one member of an expansive bacteriophage family [16].

It is becoming increasingly recognized that viral taxonomy will need to incorporate viruses which are not yet cultured and for which there is only sequence data available, such as crAssphage. However, the lack of a universal marker gene in viruses significantly complicates sequence based taxonomy. The International Committee on Taxonomy of Viruses (ICTV) in 2017 released a consensus statement describing a proposed classification pipeline for incorporating viruses for which there is only sequence data available into available taxonomy [17]. However, this framework has yet to be fully implemented and detailed methodologies are currently not available.

The ELDERMET project has examined the microbial composition of elderly citizens (>65 years old) in Ireland. This has focused on an examination of the bacterial component and has found that the microbiome of the elderly is capable of predicting residence location (community vs long term care facilities) and identified that this trend was heavily related to diet [18]. To date, however, no study has examined the virome of elderly subjects, with several studies focusing on infants [12,14,19], disease control cases [11] or healthy adults [13]. Thus, we sought to examine the diversity and composition of the DNA virome in an elderly cohort, which included individuals present both in the community and in long term care in efforts to characterize viral dark matter.

Methods

Selection of faecal samples

Faecal samples for all elderly individuals were collected with informed consent as part of the ELDERMET study [18]. This study was approved by the Clinical Research Ethics Committee of Cork Teaching Hospitals. From the elderly faecal samples available, samples were chosen in order to have: (i) 10 community and 10 long term care representatives, (ii) sufficient faecal material for DNA virome extractions and (iii) a mixture of high and low bacterial 16S Shannon diversity samples within and between the cohorts stratified by residence so as to represent the mixture of individuals captured by the ELDERMET study. For details related to samples chosen in this study, see Supplementary Table 1.

Preparation and sequencing of DNA faecal viromes

The DNA faecal viromes from elderly subjects were prepared using the protocol described by McCann *et al.* (2018). Briefly, faecal material was suspended in 1:20 (w/v) of SM buffer with large faecal particulates and bacterial cells removed by centrifugation and 0.45 µm pore diameter filtration. Unprotected DNA and RNA was removed by DNase and RNase treatment, respectively, before heat inactivating these enzymes. Subsequently, lysis and release of virion-protected DNA was performed using guanidine thiocyanate. Viral DNA was randomly amplified using Illustra GenomiPhi V2 kit (GE Healthcare) multiple displacement amplification. Purified amplified viral DNA samples were prepared for 300bp paired-end read metagenomic sequencing on an Illumina MiSeq platform (Teagasc Moorepark, Cork) using a Nextera-XT library preparation kit (Illumina) as described by the manufacturer.

Bioinformatics analysis

The quality of the raw reads was visualized with FastQC v0.11.3. Nextera adapters were removed with Cutadapt v1.9.1 [20] followed by read trimming and filtering with Trimmomatic v0.36 [21] to ensure a minimum length of 60 bps, maximum length of 150 bps, and a sliding window that cuts a read once the average quality in a window size of 4 falls below a Phred score of 30. Sequencing reads aligning to the human genome release GRCh38.p7 were removed using Kraken v0.10.5 [22]. A summary of the number and quality of paired-end reads following sequencing is available in Supplementary Table 2. Levels of bacterial contamination were estimated by classifying reads with SortMeRNA v2.0 [23] against the SILVA database and by aligning reads against the cpn60db [24] with bowtie2 in end-to-end alignment mode [25]. Reads were then assembled with the metaSPAdes assembler [26], as per the findings of Roux *et al.* 2017 [27].

Virome sequence reads were classified into known viral orders and families using the Kaiju metagenomic classifier [28] and the NCBI non-redundant protein database (March 2nd 2018; [29]). This classifier was chosen as it utilizes a 6 frame translation approach to classify sequences on the basis of amino acid homology, and thus is more sensitive to more distant relatedness. Raw reads were deposited in the NCBI under BioProject PRJNA385126. The accession for each individual sample is listed in Supplementary Table 1.

Detection of viral contigs

To further ensure there is no bacterial contamination following the assembly of contigs, viral contigs were detected using VirSorter [30]. Predicted viral contigs were annotated using VIGA (pre-print [31]). A table of the VirSorter positive viruses and their properties is summarized in Supplementary Table 3. The similarity of elderly viruses against publically deposited sequences of the NCBI nr database is available in Supplementary Table 4. Taxonomic classification of viral contigs to 'Order' and 'Family' levels was performed using DemoVir (pre-print [TBA]), with results summarized in Supplementary Table 5.

Statistical analyses

All statistical analyses were performed in R v3.3.0 [32]. Alpha diversity metrics including Chao1 richness and Shannon index were computed with PhyloSeq v1.16.2 [33] and plotted with ggplot2 v2.2.1 [34]. Between-group differences in alpha diversity were tested with a Mann-Whitney test (also known as a two sample Wilcoxon test). Unweighted Bray-Curtis distance was used as input for a Principal Coordinate Analysis (PCoA) as performed by the pcoa function in the ape package v4.1. Adonis tests were performed using the vegan package v2.4.3 [35] in R to test community level differences.

Viral host prediction

Several approaches were undertaken to try and identify host bacteria for viral sequences. Firstly, all viral genomes were queried against a database of CRISPR spacers [36]. Secondly, viral encoded tRNA sequences were detected using ARAGORN v1.2.36 [37] and the closest bacterial homologue detected through a BLAST query against the NCBI nt database. Finally, the most likely host bacterium of elderly viral contigs was calculated using WIsH ('Who IS the Host') that employs a probabilistic approach [38]. An in-house custom database of complete phage genome sequences was built to test the accuracy of WIsH by combining the European Nucleotide Archive (ENA) phage genomes database (2,010 viral sequences; May 2015) with phage sequences obtained from NCBI by the lab of Andrew Millard (8,761 viral sequences; March 2018; http://millardlab.org). A custom database of bacterial sequences was compiled by combining sequences from the ENA bacterial genomes database (3,316 bacterial sequences; May 2015) with the NCBI RefSeq database (2,477 bacterial sequences; October 2017). Redundancy was removed from these databases by removing the shorter sequence if it aligned within the larger sequence with >95% identity across 90% of its length. In addition, all sequences which contained ambiguous 'N' nucleotides, or any bacterial 'genome' sequence ≤500kb, was removed. The finalized list and accession details of phage and bacteria custom databases used during this work is available in Supplementary Tables 6 and 7, respectively. The accuracy of WIsH phage-host prediction program, using the custom built phage database against the custom bacterial databases, was calculated by comparing the bacterial genus textual descriptions. The number of matches for the phage's known host compared to the predicted host was at minimum 35%, with this accuracy expected to increase with a detailed manual curation of the downloaded bacterial and viral databases. The complete results for the WIsH accuracy estimation is reported in Supplementary Table 8. Subsequently, WIsH was applied to the DNA viruses detected in faeces of elderly individuals using the custom built database of bacteria as potential hosts. The most likely predicted host bacterium for viruses detected in the faeces of elderly individuals is available in Supplementary Table 9.

Phylogeny of phage proteins

The phylogeny of putative elderly-associated *Microviridae* viral contigs was conducted on their predicted capsid protein sequence as follows. A text search using the term 'Microviridae' was performed against the NCBI Genome and Nucleotide web-resource (March 2018) resulting in 1,289 sequences, the accession of which were downloaded using NCBI Batch Entrez. All sequences smaller than 3kb were removed, resulting in 771 *Microviridae* sequences. These sequences formed an NCBI *Microviridae* database for subsequent analyses.The protein encoded sequences of 47 elderly-associated *Microviridae* and the NCBI *Microviridae* database sequences were predicted using Prodigal v2.6.3 with the 'meta' option enabled for small contig sequences. Subsequently, a protein BLAST (E-value 1E-05) of the *Microviridae* capsid protein F (Pfam seed sequences of PF02305) was performed to identify capsid protein sequences. Resultant BLAST hits were sorted by bitscore and only the top hit per genome was retained. Next, the predicted capsid proteins of elderly-associated *Microviridae* were BLAST against the custom NCBI *Microviridae* database, with only the top resulting BLAST hit for each of the queried elderly-associated *Microviridae* retained. All 95 of the predicted *Microviridae* capsid sequences were aligned using Muscle v3.8.31 [39], with phylogeny determined using PhyML v20131022 with 1000x bootstraps using a JTT substitution model [40]. Phylogenetic relationships were visualized using FigTree v1.4.3. The *Microviridae* subfamily taxonomic classifications for the NCBI *Microviridae* sequences were extracted from their GenBank files.

A phylogenetic tree of the predicted *Caudovirales* phages was performed in a similar manner to the described *Microviridae* phylogeny, except the terminase protein was employed as a genetic marker (Pfam seed sequences of PF04466) with related terminase sequences from the created custom phage database included (Supplementary Table 6). Phage taxonomy was inferred by their DemoVir predicted family classifications and additionally, groups are highlighted similar hosts as predicted by WIsH.

The average nucleotide identity between phage genomes and taxonomic groups was calculated using pyani [41]. Input sequences were aligned using MUMmer (ANIm method) with the ‘maxmatch’ option enabled.

Viral-encoded antibiotic resistance

In order to assess if the viruses present in the faeces of the studied elderly individuals encode antibiotic resistance genes, encoded proteins were predicted using Prodigal v2.6.3 with the ‘meta’ option enabled and subsequently screened against the Comprehensive Antibiotic Resistance Database (CARD) database [42] using an E-value threshold of 1E-05 (Supplementary Table 10). Subsequently, the viral DfrE-related sequence yielding the top BLAST hit against the CARD database was compared with other thymidylate synthases downloaded from Pfam (seed sequences PF00303). Briefly, all thymidylate synthase sequences were aligned using Muscle v3.8.31, and their phylogeny was determined using PhyML v20131022 using a JTT substitution model with default 20 bootstraps. The elderly viral encoded DfrE-related sequence and its closest Pfam thymidylate synthase sequence, TYSY_SYMTH, were aligned using Muscle and visualized using JalView [43]. The coordinate of the highlighted conserved cysteine residue of thymidylate synthases was provided by Pfam.

Results

Viral-like-particles purified from faecal samples were sequenced on an Illumina MiSeq to generate a median of 1.3 million read pairs per sample. Following quality control, the remaining paired end reads were classified into known viral groups using the Kaiju classifier [28] against the nr database at NCBI (Figure 1). However, even with this approach, a median of 72% of reads per sample remained unclassified. Those reads which were assigned to a known viral group were primarily into the viral order *Caudovirales* and family *Microviridae*. The latter in particular were found to be most abundant identifiable viral group in many of the samples sequenced here, although this may be due to the use of Multiple Displacement Amplification (MDA) which has been shown to distort abundance of ssDNA viruses [44].

Due to the large number of unassigned viruses, reads were assembled using MetaSPAdes [26]. In order to avoid contamination with bacterial sequences, only those contigs predicted as viral by VirSorter [30] were considered for further analysis. This resulted in a total of 205 contigs, ranging in size from 1,353 bases to 118,143 bases (Supplementary Table 3). The distribution of read coverage and size of VirSorter-detected viruses is shown in Figure 2. Again, the impact of MDA is evident by the extremely high coverage of smaller viruses, presumably ssDNA phage. Paired end reads were aligned back to this contig set using bowtie2 in end-to-end mode which recruited a median of 70.42% reads per sample. Only 100 contigs showed significant nucleotide homology to any sequence in the NCBI nt database (BLASTn cut offs: maximum E-value 1E-10 and alignment length of 500 bases) and in many cases these alignments were only a small fraction of the assembled sequence (Supplementary table 4). Over half of these sequences were identified as circular by VirSorter, thus making it likely that the contig set described above is primarily composed of complete viral genomes which are not currently present in reference databases.

Analysis of bacterial composition in the ELDERMET cohort previously found differences in both alpha and beta diversity between those elderly living in long term care and in the community [18]. However, when we repeated this analyses using the above contig set as a reference we detected no difference by alpha or beta diversity in tested metrics (Figure 1A and

1B). It is worth noting that we only investigated the DNA portion of the virome and the amplification of DNA before sequencing is predicted to skew diversity estimates. In addition, samples were chosen with prior knowledge of bacterial 16S diversity metrics in order to capture a range of sample types within the ELDERMET cohort.

A search of the putative elderly-associated viruses against a CRISPR spacer database resulted in 4 matches, supporting the characterization of these contigs as mobile genetic elements (Supplementary Table 3). There were also tRNA genes identified in a further 9 viruses providing an indication of host range for these contigs. The top results for the phage-host prediction program WIsH linked the elderly-associated viruses to *Paenibacillus*, *Clostridium*, *Bacteroides*, *Lachnoclostridium*, *Bacillus* and *Sphingobacterium* (Supplementary Figure 1).

A total of 47 predicted *Microviridae* viral sequences were detected in the faeces of the elderly individuals. In order to determine the relatedness of these viruses, phylogenetic analysis of their capsid protein sequences was performed (Figure 3). By interspersing characterized *Microviridae* sequences within the phylogenetic tree, it is possible to observe that the majority of elderly-associated *Microviridae* are members of the *Gokushovirinae* subfamily with only a single sequence clustering with members of the *Microviridae Bullavirinae* subfamily. The average nucleotide identity of all *Microviridae* sequences was calculated and showed the genomes of *Bullavirinae* subfamily members (both elderly and NCBI sequences) have >70% average nucleotide identity (Supplementary Figure 2). However, the majority of *Gokushovirinae* subfamily genome sequences share <70% average nucleotide identity, demonstrating there is extensive uncharacterized diversity of *Microviridae* viruses associated with human faeces and there is also a need to revise the taxonomy of *Gokushovirinae* using a sequence-based approach.

The family level taxonomic prediction of *Caudovirales* phages was performed using DemoVir (pre-print [TBA]; Supplementary Table 5). Subsequently, a phylogenetic comparison of the terminase protein of *Caudovirales* phages detected in the faeces of elder individuals was performed (Figure 4). The phylogeny of terminase sequences results in phages of the families *Myoviridae*, *Podoviridae* and *Siphoviridae* intermingled throughout the phylogenetic tree. Only one group of *Myoviridae* phages, predicted to infect members of the order *Clostridiales* (*Clostridium* and *Lachnoclostridium*), are widespread amongst elderly individuals. Average nucleotide identity and genome comparisons of the putative *Clostridiales Myoviridae* phages were performed, highlighting the genomic variations within this cluster (Supplementary Figures 3 & 4). The terminase phylogenetic analysis also identified a smaller cluster of putative *Sphingobacterium*-infecting *Myoviridae* phages amongst three individuals. Visual genome comparisons show these phages are highly related, but not clonal (Supplementary Figure 5).

Recently, there has been significant interest in characterizing the mobilization of antibiotic resistance genes, with mixed reports about phage involvement [45]. Therefore, we investigated whether antibiotic resistance genes were associated with the viruses detected in the faeces of elder individuals. A total of 73 BLAST hits were obtained against the 205 elderly viral contigs below the chosen cut-off (E-value <1E-05; Supplementary Table 10). However, a manual examination of these BLAST hits highlighted the majority had small alignment lengths, low percentage identities and high numbers of mismatches. Therefore, these results were considered insignificant and not pursued.

Of interest were the BLAST hits of phage protein encoding sequences against *dfrE* of *Enterococcus faecalis* within the CARD database (two E-values < 1E-34) (Figure 5; Supplementary Table 10).A literature search for DfrE of *E. faecalis* identified it encodes a thymidylate synthase ([46]; accession AF028811.1). The putative DfrE of phage contigEM298_T0.NODE_5 was compared to the thymidylate synthase Pfam seed sequences (PF00303). DfrE of EM298_T0.NODE_5 clusters with other thymidylate synthases (Supplementary Figure 6A). A comparison of DfrE of EM298_T0.NODE_5 with the closest related sequence, TYSY_SYMTH, showed that both sequences are conserved at the predicted catalytic cysteine ([47]; Supplementary Figure 6B).

Discussion

There remains a vast amount of uncharacterized sequence diversity associated with the viral fraction of the human microbiota, referred to as 'viral dark matter'. While there are an increasing number of metagenomic studies characterizing viruses and phages associated with humans, most microbiota research is still focused on correlating presence/absence with health/disease and the next big challenge is moving towards determining causation [48]. Altered phage populations have been observed in various human diseased states [13]; in addition, restoration of viral and phage populations through filtered, bacteria-free faecal transplants have had initial success in treating recurrent *Clostridium difficile* infections [49].

With each additional phage metagenomic study identifying and characterizing more of the unknown viruses and phages present in the human microbiota, the viral dark matter, future studies will become more informative. Current metagenomic studies of phages are hampered by the vast amounts of sequence diversity and lack of database representatives. This is epitomized by crAssphage, the most abundant virus of the human microbiota, which was only identified in 2014 by Dutilh and colleagues [15]. CrAssphage, which was lacking a database homologue when it was first discovered, was identified through *de novo* assembly approaches. Subsequently, researchers have used crAssphage to identify related phage sequences and propose a taxonomic structure to these abundant viruses ([16]; pre-print [50]). Therefore, in the absence of an all-encompassing curated viral database, researchers must characterize viral populations through both database dependent methods and also through database independent approaches to identify novel sequences.

With an aging human population, understanding the various facets of the human microbiota through multi-omic approaches will be important in designing strategies to prolong health. A previous examination of elderly gut microbiotas in Ireland demonstrated that the bacterial 16S rRNA composition differentiated individuals based on residence location [18]. This study initiated an examination of the human faecal virome associated with elderly individuals (>65 years of age). Following metagenomic sequencing of the elderly viromes, we observed no differences in the viral diversity between residential cohorts when queried against known viral sequences. In addition, no diversity variations were observed between elderly viromes with younger, average-aged healthy controls from the study of Norman and colleagues ([11]; data not shown). However, this is not to say that differences do not exist. It must be noted that samples were chosen for this study with prior 16S rRNA compositional knowledge and were not randomized. In addition, amplification of DNA prior to sequencing may have masked some of the diversity differences between the elderly-virome cohorts. Such biases should be removed from subsequent efforts to characterize the virome of aging individuals.

One of the challenges of viral metagenomic studies is identifying the host organism. Several methods have been applied in this study to try and infer potential host information to novel viral sequences: (i) CRISPR spacer sequences, (ii) tRNA sequences and (iii) a probabilistic approach. Finding matches between specific bacterial CRISPR spacer sequences and viral contigs is database dependent and CRISPR spacer sequences are highly strain dependent. However, a match between a CRISPR spacer sequence and phage is a strong indication of a true biological interaction.Within this study, the bacteria *Bacteroides vulgatus* and *Odoribacter splanchnicus* encoded CRISPR spacer sequences which closely matched elderly associated viruses.

Several phages encode their own tRNA sequences to optimize their host range and replication [51]. While tRNA sequences are often associated with mobile genetic elements, they are functionally conserved and their relatedness to bacterial homologues can be used to infer a potential host relationship. The tRNA sequences associated with elderly viruses had close homologues to human gut bacteria such as *E. coli* and *B. vulgatus*. However, one of the closest bacterial homologues to a phage-encoded tRNA was *Borrelia garinii*, a causative agent of Lyme disease. While there is no data available to suggest this elderly individual had Lyme disease, this prediction is most likely an artifact of selecting the top BLAST hit for a divergent phage tRNA encoded sequence from uncharacterized phages of the human viral dark matter.

Finally, to predict potential hosts, the probability of a phage-host pair was calculated using the program WIsH, which is suggested to be at its most accurate when applied to small phage genomes without tRNA sequences as these phages would be more dependent on their host's replication machinery [38]. While WIsH is able to infer the most probable host for all of the 205 elderly-associated viruses from the bacteria supplied, in only 1of the 9 phages which encode a tRNA did the WIsH host prediction match the predicted tRNA host prediction. In addition, there were occasions where the WIsH predicted a host that was not detectable in the 16S rRNA sequencing results of the same sample. Thus, without laboratory isolation of a phage-host pair, all *in silico* host predictions should be treated extremely cautiously.

A phylogenetic analysis of the *Microviridae* phages present in the faeces of elderly individuals was performed. Taxonomically, the majority of elderly faecal *Microviridae* phages are members of the *Gokushovirinae* subfamily, with only a single sequence clustering with *Bullavirinae* subfamily members. *Gokushovirinae* have previously been detected in the faeces of humans and wild chimpanzees [52,53], with *Gokushovirinae* phages predicted to infect obligate parasitic bacteria, such as *Chlamydiae* and *Bdellovibrio*, and *Spiroplasma* [54]. While the *Gokushovirinae* subfamily contains three genera [55], the average nucleotide identities of the *Gokushovirinae* detected in elderly faeces do not result in three distinct clusters, while all *Bullavirinae* sequences formed a single cluster by average nucleotide identity. Therefore, there is a significant amount of uncharacterized *Microviridae* diversity associated with human faeces.

A comparison of all the detected elderly-associated phage terminase sequences was performed to assess the diversity of *Caudovirales* phages. Interestingly, phages of the families *Myoviridae*, *Podoviridae* and *Siphoviridae* (of the order *Caudovirales*) did not cluster together, supporting the need for a sequence based taxonomic scheme rather than categorizing phages by morphology. As supporting evidence, two distinct clusters of phage terminase sequences were present in our phylogenetic analysis, composed of three or more sequences, and the respective phages putatively infect similar host bacteria (*Clostridiales* and *Sphingobacterium*). Thus, despite conserved morphologies, divergence of *Caudovirales* phage sequences appears to be driven by the targeted host organism. Additionally, during the terminase phylogenetic analysis, all of the most-closely related terminase sequences identified from a custom database of 3,134 phages were included in our dendrogram. However, only 12 related terminase sequences were recruited for the resultant terminase tree. This result exemplifies the vast amounts of viral dark matter within the human gut virome still not characterized and deposited in a public database.

Antibiotic resistance is a significant concern in long term health care facility. A search for potential antibiotic resistance genes associated with elderly gut viral contigs resulted in numerous hits which were not considered significant, but there were several strong hits against *dfrE* of *Enterococcus faecalis*. An examination of the literature for DfrE thymidylate synthase demonstrated it conferred resistance to trimethoprim in *E. coli* when cloned into a high copy plasmid [46]. However, the majority of the literature surrounding phage encoding thymidylate synthases characterize this enzyme as important in DNA synthesis [56] with recent literature showing these enzymes are present in phages which synthesize modified DNA bases during replication [57]. Therefore, it is not clear if the elderly-associated faecal viruses examined in this study confer trimethoprim resistance to their host bacteria; however, any potential trimethoprim resistance conferred through these phages is likely a secondary consequence of the phage's natural replication strategy.

Conclusion

Understanding and manipulating the human microbiome is important in treating various conditions and providing an overall improvement in quality of life. However, the first steps towards this ultimate goal is a better understanding of the constituent members of the human microbiota. In this study, we focused on the viral populations of elderly individuals. To our knowledge there have been no studies characterizing phage communities associated with elderly (>65 years old) individuals, a demographic which is increasingly important as improvements in healthcare result in longer lives. While no differences were observed in the diversity of viral populations between elderly individuals by residential location in this study, or observed against younger healthy adults, there is proposed to be an overall gradual decline in the microbiota of aging individuals. However, making conclusions about differences in viral populations and presence/absence of specific viruses with health/disease is strongly dependent on available databases, yet there are significant amounts of unknown sequences detected in the viral fraction of the human microbiome and elderly individuals. Therefore, further characterization of the ‘viral dark matter’ will facilitate a better understanding of the complete human microbiota. This will hopefully result in a better understanding of host-microbiota interactions that will hopefully lead to future therapeutic interventions and diagnostics to ultimately improve human health.

Acknowledgements

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under Grant Numbers SFI/12/RC/2273 and SFI/14/SP APC/B3032, with researchers supported from a grant from Janssen Biotech, Inc.

Author Contributions

SRS and MD performed the laboratory work. SRS, FJR and AM conducted the bioinformatic analysis. SRS and FJR wrote the paper and generated the figures. PR and CH secured the funding and wrote the paper. All authors contributed to the analysis of the data.

Conflicts of Interest

The authors declare no conflict of interest.

Data deposition

Raw reads are accessible through NCBI BioProject PRJNA385126. While contigs are being submitted to GenBank, they are currently available at:

https://figshare.com/articles/Elderly_virome_contigs/6729095

References

1. Kassam, Z.; Lee, C. H.; Yuan, Y.; Hunt, R. H. Fecal Microbiota Transplantation for Clostridium difficile Infection: Systematic Review and Meta-Analysis. *Am. J. Gastroenterol.* **2013**, *108*, 500–508, doi:10.1038/ajg.2013.59.
2. Kelly, C. R.; Ihunnah, C.; Fischer, M.; Khoruts, A.; Surawicz, C.; Afzali, A.; Aroniadis, O.; Barto, A.; Borody, T.; Giovanelli, A.; Gordon, S.; Gluck, M.; Hohmann, E. L.; Kao, D.; Kao, J. Y.; McQuillen, D. P.; Mellow, M.; Rank, K. M.; Rao, K.; Ray, A.; Schwartz, M. A.; Singh, N.; Stollman, N.; Suskind, D. L.; Vindigni, S. M.; Youngster, I.; Brandt, L. Fecal Microbiota Transplant for Treatment of Clostridium difficile Infection in Immunocompromised Patients. *Am. J. Gastroenterol.* **2014**, *109*, 1065–1071, doi:10.1038/ajg.2014.133.
3. The Human Microbiome Project Consortium Structure, function and diversity of the healthy human microbiome. *Nature* **2012**, *486*, 207–214, doi:10.1038/nature11234.
4. Lloyd-Price, J.; Abu-Ali, G.; Huttenhower, C. The healthy human microbiome. *Genome Med.* **2016**, *8*, doi:10.1186/s13073-016-0307-y.
5. Lloyd-Price, J.; Mahurkar, A.; Rahnavard, G.; Crabtree, J.; Orvis, J.; Hall, A. B.; Brady, A.; Creasy, H. H.; McCracken, C.; Giglio, M. G.; McDonald, D.; Franzosa, E. A.; Knight, R.; White, O.; Huttenhower, C. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* **2017**, *550*, 61–66, doi:10.1038/nature23889.
6. 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*, doi:10.7554/eLife.08490.
7. Krishnamurthy, S. R.; Wang, D. Origins and challenges of viral dark matter. *Virus Res.* **2017**, *239*, 136–142, doi:10.1016/j.virusres.2017.02.002.
8. Carding, S. R.; Davis, N.; Hoyles, L. Review article: the human intestinal virome in health and disease. *Aliment. Pharmacol. Ther.* **2017**, *46*, 800–815, doi:10.1111/apt.14280.
9. 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.* **2002**, *99*, 14250–14255, doi:10.1073/pnas.202488399.
10. Breitbart, M.; Hewson, I.; Felts, B.; Mahaffy, J. M.; Nulton, J.; Salamon, P.; Rohwer, F. Metagenomic Analyses of an Uncultured Viral Community from Human Feces. *J. Bacteriol.* **2003**, *185*, 6220–6223, doi:10.1128/JB.185.20.6220-6223.2003.
11. 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.; Stappenbeck, T. S.; McGovern, D. P. B.; Keshavarzian, A.; Mutlu, E. A.; Sauk, J.; Gevers, D.; Xavier, R. J.; Wang, D.; Parkes, M.; Virgin, H. W. Disease-Specific

Alterations in the Enteric Virome in Inflammatory Bowel Disease. *Cell* **2015**, *160*, 447–460, doi:10.1016/j.cell.2015.01.002.

12. Lim, E. S.; Zhou, Y.; Zhao, G.; Bauer, I. K.; Droit, L.; Ndao, I. M.; Warner, B. B.; Tarr, P. I.; Wang, D.; Holtz, L. R. Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nat. Med.* **2015**, *21*, 1228.
13. 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.* **2016**, *113*, 10400–10405, doi:10.1073/pnas.1601060113.
14. McCann, A.; Ryan, F. J.; Stockdale, S. R.; Dalmasso, M.; Blake, T.; Ryan, C. A.; Stanton, C.; Mills, S.; Ross, P. R.; Hill, C. Viromes of one year old infants reveal the impact of birth mode on microbiome diversity. *PeerJ* **2018**, *6*, e4694, doi:10.7717/peerj.4694.
15. Dutilh, B. E.; Cassman, N.; McNair, K.; Sanchez, S. E.; Silva, G. G. Z.; Boling, L.; Barr, J. J.; Speth, D. R.; Seguritan, V.; Aziz, R. K.; Felts, B.; Dinsdale, E. A.; Mokili, J. L.; Edwards, R. A. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat. Commun.* **2014**, *5*, 4498.
16. Yutin, N.; Makarova, K. S.; Gussow, A. B.; Krupovic, M.; Segall, A.; Edwards, R. A.; Koonin, E. V. Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut. *Nat. Microbiol.* **2018**, *3*, 38–46, doi:10.1038/s41564-017-0053-y.
17. 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.; Hull, R.; King, A. M. Q.; Koonin, E. V.; Krupovic, M.; Kuhn, J. H.; Lefkowitz, E. J.; Nibert, M. L.; Orton, R.; Roossinck, M. J.; Sabanadzovic, S.; Sullivan, M. B.; Suttle, C. A.; Tesh, R. B.; van der Vlugt, R. A.; Varsani, A.; Zerbini, F. M. Virus taxonomy in the age of metagenomics. *Nat. Rev. Microbiol.* **2017**, *15*, 161.
18. Claesson, M. J.; Jeffery, I. B.; Conde, S.; Power, S. E.; O'Connor, E. M.; Cusack, S.; Harris, H. M. B.; Coakley, M.; Lakshminarayanan, B.; O'Sullivan, O.; Fitzgerald, G. F.; Deane, J.; O'Connor, M.; Harnedy, N.; O'Connor, K.; O'Mahony, D.; van Sinderen, D.; Wallace, M.; Brennan, L.; Stanton, C.; Marchesi, J. R.; Fitzgerald, A. P.; Shanahan, F.; Hill, C.; Ross, R. P.; O'Toole, P. W. Gut microbiota composition correlates with diet and health in the elderly. *Nature* **2012**, *488*, 178.
19. Reyes, A.; Blanton, L. V.; Cao, S.; Zhao, G.; Manary, M.; Trehan, I.; Smith, M. I.; Wang, D.; Virgin, H. W.; Rohwer, F.; Gordon, J. I. Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. *Proc. Natl. Acad. Sci.* **2015**, *112*, 11941–11946, doi:10.1073/pnas.1514285112.
20. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **2011**, *17*, 10, doi:10.14806/ej.17.1.200.
21. Bolger, A. M.; Lohse, M.; Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120, doi:10.1093/bioinformatics/btu170.
22. Wood, D. E.; Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **2014**, *15*, R46, doi:10.1186/gb-2014-15-3-r46.
23. Kopylova, E.; Noé, L.; Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **2012**, *28*, 3211–3217, doi:10.1093/bioinformatics/bts611.
24. Hill, J. E.; Penny, S. L.; Crowell, K. G.; Goh, S. H.; Hemmingsen, S. M. cpnDB: a chaperonin sequence database. *Genome Res.* **2004**, *14*, 1669–1675, doi:10.1101/gr.2649204.
25. Langmead, B.; Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359, doi:10.1038/nmeth.1923.
26. Nurk, S.; Meleshko, D.; Korobeynikov, A.; Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **2017**, *27*, 824–834, doi:10.1101/gr.213959.116.
27. Roux, S.; Emerson, J. B.; Eloe-Fadrosh, E. A.; Sullivan, M. B. Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ* **2017**, *5*, e3817, doi:10.7717/peerj.3817.

28. Menzel, P.; Ng, K. L.; Krogh, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* **2016**, *7*, 11257.
29. Wheeler, D. L.; Barrett, T.; Benson, D. A.; Bryant, S. H.; Canese, K.; Chetvernin, V.; Church, D. M.; DiCuccio, M.; Edgar, R.; Federhen, S.; Geer, L. Y.; Kapustin, Y.; Khovayko, O.; Landsman, D.; Lipman, D. J.; Madden, T. L.; Maglott, D. R.; Ostell, J.; Miller, V.; Pruitt, K. D.; Schuler, G. D.; Sequeira, E.; Sherry, S. T.; Sirotkin, K.; Souvorov, A.; Starchenko, G.; Tatusov, R. L.; Tatusova, T. A.; Wagner, L.; Yaschenko, E. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2007**, *35*, D5-12, doi:10.1093/nar/gkl1031.
30. Roux, S.; Enault, F.; Hurwitz, B. L.; Sullivan, M. B. VirSorter: mining viral signal from microbial genomic data. *PeerJ* **2015**, *3*, e985, doi:10.7717/peerj.985.
31. González-Tortuero, E.; Sutton, T. D.; Velayudhan, V.; Shkoporov, A. N.; Draper, L. A.; Stockdale, S. R.; Ross, R. P.; Hill, C. VIGA: a sensitive, precise and automatic de novo VIral Genome Annotator. **2018**, doi:10.1101/277509.
32. Becker, R. A.; Chambers, J. M.; Wilks, A. R. The new S language. A programming environment for data analysis and graphics. Wadsworth & Brooks. **1988**.
33. McMurdie, P. J.; Holmes, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE* **2013**, *8*, e61217, doi:10.1371/journal.pone.0061217.
34. Wickham, H. *Ggplot2: elegant graphics for data analysis*; Use R!; Springer: New York, 2009; ISBN 978-0-387-98140-6.
35. Oksanen, J.; Kindt, R.; Legendre, P.; O'Hara, B. VEGAN: R package for community ecology. **2006**.
36. Grissa, I.; Vergnaud, G.; Pourcel, C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **2007**, *8*, 172, doi:10.1186/1471-2105-8-172.
37. Laslett, D.; Canback, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* **2004**, *32*, 11–16, doi:10.1093/nar/gkh152.
38. 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*, 3113–3114, doi:10.1093/bioinformatics/btx383.
39. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797, doi:10.1093/nar/gkh340.
40. Guindon, S.; Delsuc, F.; Dufayard, J.-F.; Gascuel, O. Estimating Maximum Likelihood Phylogenies with PhyML. In *Bioinformatics for DNA Sequence Analysis*; Posada, D., Ed.; Humana Press: Totowa, NJ, 2009; Vol. 537, pp. 113–137 ISBN 978-1-58829-910-9.
41. Pritchard, L.; Glover, R. H.; Humphris, S.; Elphinstone, J. G.; Toth, I. K. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal. Methods* **2016**, *8*, 12–24, doi:10.1039/C5AY02550H.
42. McArthur, A. G.; Waglechner, N.; Nizam, F.; Yan, A.; Azad, M. A.; Baylay, A. J.; Bhullar, K.; Canova, M. J.; De Pascale, G.; Ejim, L.; Kalan, L.; King, A. M.; Koteva, K.; Morar, M.; Mulvey, M. R.; O'Brien, J. S.; Pawlowski, A. C.; Piddock, L. J. V.; Spanogiannopoulos, P.; Sutherland, A. D.; Tang, I.; Taylor, P. L.; Thaker, M.; Wang, W.; Yan, M.; Yu, T.; Wright, G. D. The Comprehensive Antibiotic Resistance Database. *Antimicrob. Agents Chemother.* **2013**, *57*, 3348–3357, doi:10.1128/AAC.00419-13.
43. Waterhouse, A. M.; Procter, J. B.; Martin, D. M. A.; Clamp, M.; Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, *25*, 1189–1191, doi:10.1093/bioinformatics/btp033.
44. 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, doi:10.7717/peerj.2777.

45. Enault, F.; Briet, A.; Bouteille, L.; Roux, S.; Sullivan, M. B.; Petit, M.-A. Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. *Isme J.* **2016**, *11*, 237.
46. Coque, T. M.; Singh, K. V.; Weinstock, G. M.; Murray, B. E. Characterization of dihydrofolate reductase genes from trimethoprim-susceptible and trimethoprim-resistant strains of Enterococcus faecalis. *Antimicrob. Agents Chemother.* **1999**, *43*, 141–147.
47. Carreras, C. W.; Santi, D. V. The catalytic mechanism and structure of thymidylate synthase. *Annu. Rev. Biochem.* **1995**, *64*, 721–762, doi:10.1146/annurev.bi.64.070195.003445.
48. Surana, N. K.; Kasper, D. L. Moving beyond microbiome-wide associations to causal microbe identification. *Nature* **2017**, doi:10.1038/nature25019.
49. Ott, S. J.; Waetzig, G. H.; Rehman, A.; Moltzau-Anderson, J.; Bharti, R.; Grasis, J. A.; Cassidy, L.; Tholey, A.; Fickenscher, H.; Seegert, D.; Rosenstiel, P.; Schreiber, S. Efficacy of Sterile Fecal Filtrate Transfer for Treating Patients With Clostridium difficile Infection. *Gastroenterology* **2017**, *152*, 799-811.e7, doi:10.1053/j.gastro.2016.11.010.
50. Guerin, E.; Shkoporov, A.; Stockdale, S.; Clooney, A. G.; Ryan, F. J.; Sutton, T. D. S.; Draper, L. A.; Gonzalez-Tortuero, E.; Ross, R. P.; Hill, C. Biology and taxonomy of crAss-like bacteriophages, the most abundant virus in the human gut. **2018**, doi:10.1101/295642.
51. Wilson, J. H. Function of the bacteriophage T4 transfer RNA's. *J. Mol. Biol.* **1973**, *74*, 753–757, doi:10.1016/0022-2836(73)90065-X.
52. Guo, L.; Hua, X.; Zhang, W.; Yang, S.; Shen, Q.; Hu, H.; Li, J.; Liu, Z.; Wang, X.; Wang, H.; Zhou, C.; Cui, L. Viral metagenomics analysis of feces from coronary heart disease patients reveals the genetic diversity of the Microviridae. *Virol. Sin.* **2017**, *32*, 130–138, doi:10.1007/s12250-016-3896-0.
53. Walters, M.; Bawuro, M.; Christopher, A.; Knight, A.; Kraberger, S.; Stainton, D.; Chapman, H.; Varsani, A. Novel Single-Stranded DNA Virus Genomes Recovered from Chimpanzee Feces Sampled from the Mambilla Plateau in Nigeria. *Genome Announc.* **2017**, *5*, e01715-16, doi:10.1128/genomeA.01715-16.
54. Roux, S.; Krupovic, M.; Poulet, A.; Debroas, D.; Enault, F. Evolution and Diversity of the Microviridae Viral Family through a Collection of 81 New Complete Genomes Assembled from Virome Reads. *PLoS ONE* **2012**, *7*, e40418, doi:10.1371/journal.pone.0040418.
55. Carstens, E. B. Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2009). *Arch. Virol.* **2010**, *155*, 133–146, doi:10.1007/s00705-009-0547-x.
56. Finer-Moore, J. S.; Maley, G. F.; Maley, F.; Montfort, W. R.; Stroud, R. M. Crystal structure of thymidylate synthase from T4 phage: component of a deoxynucleoside triphosphate-synthesizing complex. *Biochemistry (Mosc.)* **1994**, *33*, 15459–15468.
57. Kropinski, A.; Turner, D.; Nash, J.; Ackermann, H.-W.; Lingohr, E.; Warren, R.; Ehrlich, K.; Ehrlich, M. The Sequence of Two Bacteriophages with Hypermodified Bases Reveals Novel Phage-Host Interactions. *Viruses* **2018**, *10*, 217, doi:10.3390/v10050217.

Figure Legends:

Figure 1. Virome diversity of community versus long-term residential care elderly individuals living in Ireland. **(A)** Shannon diversity comparison of elderly-associated viromes. **(B)** Bray-curtis PCoA plot comparing the virome composition of elderly cohorts. **(C)** The percentage log relative abundance differences of specific viral taxa detected in elderly faeces.

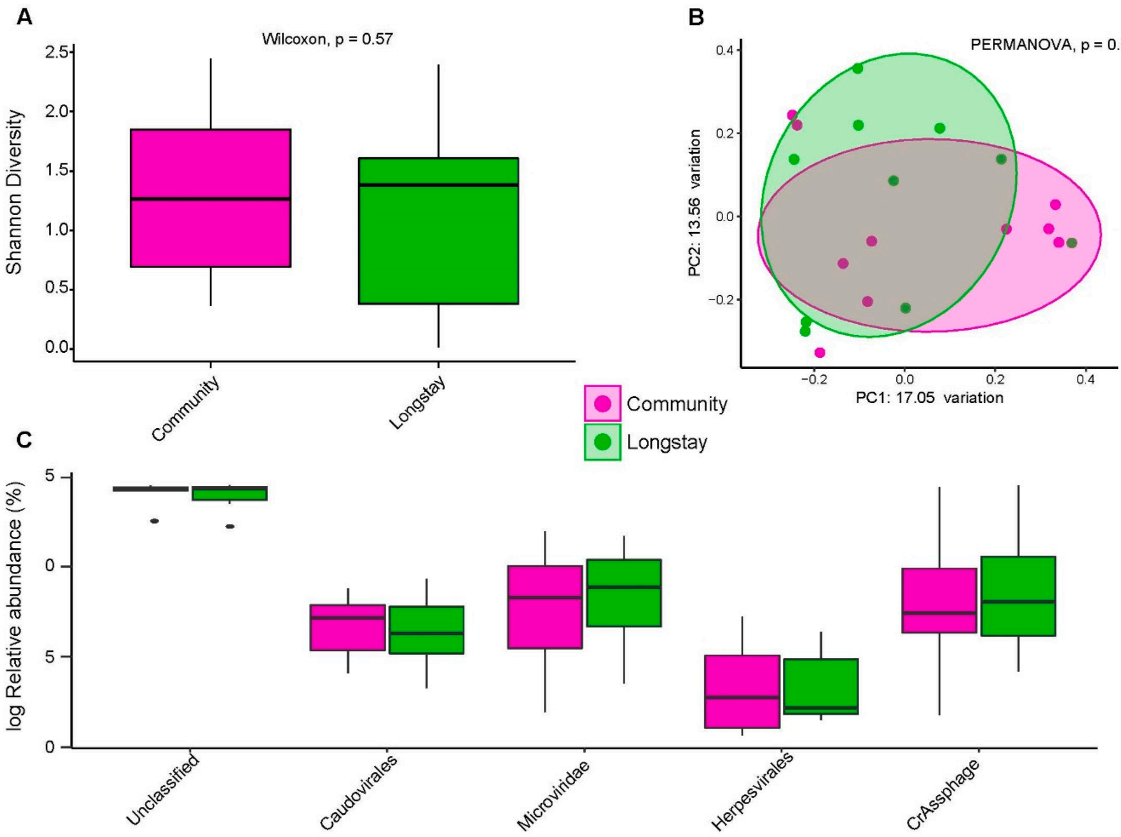
Figure 2. VirSorter detected viruses present in the faeces of elderly individuals, highlighting viral size distribution relative to sequencing coverage. Numerous small circular contigs (blue) were characterized as *Microviridae*, while most viral contigs share no nucleotide homology to known viruses (diamonds).

Figure 3. Dendrogram demonstrating the diversity of *Microviridae* capsid protein-encoded sequences detected in the faeces of elderly individuals. Sequences in red were downloaded from NCBI 'genomes' database.

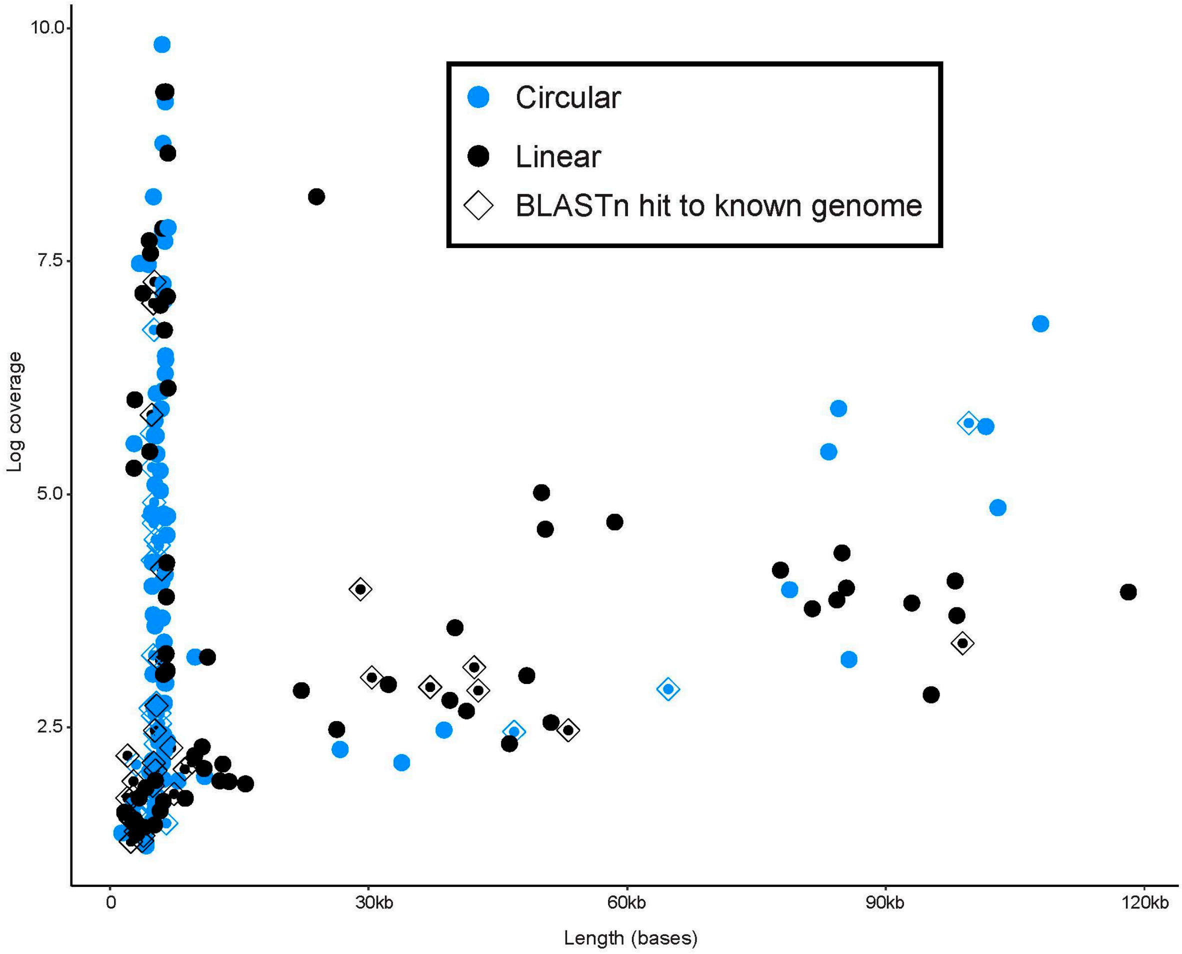
Figure 4. Phylogenetic comparison of the encoded terminase protein-encoded sequences present in the viromes of elderly subjects. Sequences in red were downloaded from NCBI 'genomes' database. The assignment of a putative taxonomic rank to *Caudovirales* phages was performed using DemoVir. The lack of relatedness between phages with a predicted similar morphology highlights the need to move towards a sequence based taxonomic system.

Figure 5. Assessment of antibiotic resistance genes associated with viruses detected in the faeces of elderly individuals. (A) The number of BLASTp hits between the proteins encoded by elderly-associated viruses (n = 205) versus the Comprehensive Antibiotic Resistance Database (CARD). Hits are grouped together based on the potential antibiotic resistance mechanism. (B) Graphical representation of E-values returned for viral hits against CARD, with only 4 strong E-value hits. (C) A non-redundant list of the viral hits against specific antibiotic resistance mechanisms, with all strong hits against DfrE. A literature investigation into viral *dfrE* sequences indicate it is most probably involved in the phage's normal life cycle (see Discussion).

Figure 1.


A
Wilcoxon, p = 0.57
Shannon Diversity
2.5
2.0
1.5
1.0
0.5
0.0
Community
Longstay
B
PERMANOVA, p = 0.57
PC2: 13.56 variation
0.4
0.2
0.0
−0.2
−0.2
0.0
0.2
0.4
PC1: 17.05 variation
● Community
● Longstay
C
log Relative abundance (%)
5
0
5
0
Unclassified
Caudovirales
Microviridae
Herpesvirales
CrAssphage

621 **Figure 2.**



622

Figure 3.

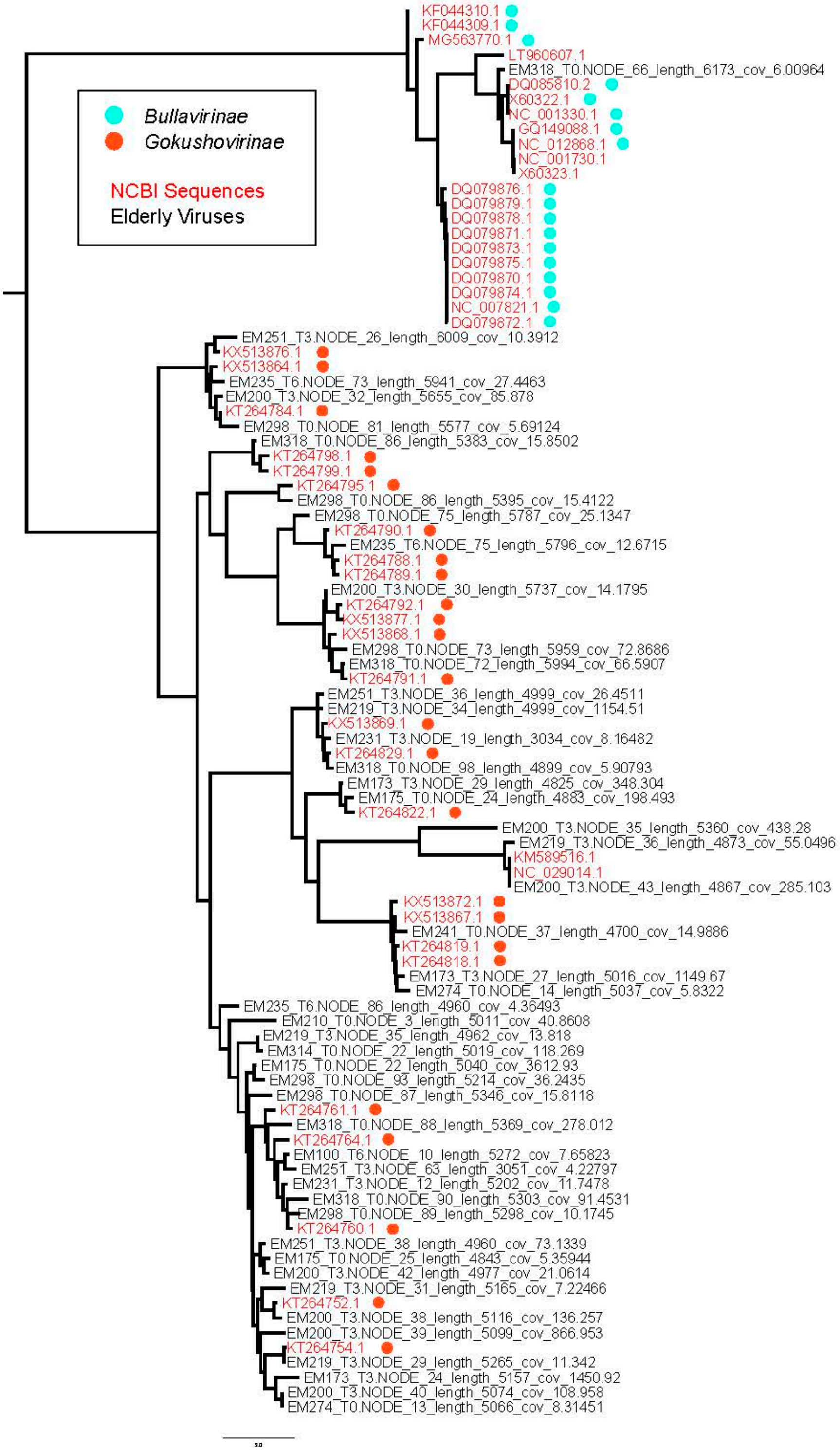

● *Bullavirinae*
● *Gokushovirinae*
NCBI Sequences
Elderly Viruses
KF044310.1
KF044309.1
MG563770.1
LT960607.1
EM318_T0.NODE_66_length_6173_cov_6.00964
DQ085810.2
X60322.1
NC_001330.1
GQ149088.1
NC_012868.1
NC_001730.1
X60323.1
DQ079876.1
DQ079879.1
DQ079878.1
DQ079871.1
DQ079873.1
DQ079875.1
DQ079870.1
DQ079874.1
NC_007821.1
DQ079872.1
EM251_T3.NODE_26_length_6009_cov_10.3912
KX513876.1
KX513864.1
EM235_T6.NODE_73_length_5941_cov_27.4463
EM200_T3.NODE_32_length_5655_cov_85.878
KT264784.1
EM298_T0.NODE_81_length_5577_cov_5.69124
EM318_T0.NODE_86_length_5383_cov_15.8502
KT264798.1
KT264799.1
KT264795.1
EM298_T0.NODE_86_length_5395_cov_15.4122
EM298_T0.NODE_75_length_5787_cov_25.1347
KT264790.1
EM235_T6.NODE_75_length_5796_cov_12.6715
KT264788.1
KT264789.1
EM200_T3.NODE_30_length_5737_cov_14.1795
KT264792.1
KX513877.1
KX513868.1
EM298_T0.NODE_73_length_5959_cov_72.8686
EM318_T0.NODE_72_length_5994_cov_66.5907
KT264791.1
EM251_T3.NODE_36_length_4999_cov_26.4511
EM219_T3.NODE_34_length_4999_cov_1154.51
KX513869.1
EM231_T3.NODE_19_length_3034_cov_8.16482
KT264829.1
EM318_T0.NODE_98_length_4899_cov_5.90793
EM173_T3.NODE_29_length_4825_cov_348.304
EM175_T0.NODE_24_length_4883_cov_198.493
KT264822.1
EM200_T3.NODE_35_length_5360_cov_438.28
EM219_T3.NODE_36_length_4873_cov_55.0496
KM589516.1
NC_029014.1
EM200_T3.NODE_43_length_4867_cov_285.103
KX513872.1
KX513867.1
EM241_T0.NODE_37_length_4700_cov_14.9886
KT264819.1
KT264818.1
EM173_T3.NODE_27_length_5016_cov_1149.67
EM274_T0.NODE_14_length_5037_cov_5.8322
EM235_T6.NODE_86_length_4960_cov_4.36493
EM210_T0.NODE_3_length_5011_cov_40.8608
EM219_T3.NODE_35_length_4962_cov_13.818
EM314_T0.NODE_22_length_5019_cov_118.269
EM175_T0.NODE_22_length_5040_cov_3612.93
EM298_T0.NODE_93_length_5214_cov_36.2435
EM298_T0.NODE_87_length_5346_cov_15.8118
KT264761.1
EM318_T0.NODE_88_length_5369_cov_278.012
KT264764.1
EM100_T6.NODE_10_length_5272_cov_7.65823
EM251_T3.NODE_63_length_3051_cov_4.22797
EM231_T3.NODE_12_length_5202_cov_11.7478
EM318_T0.NODE_90_length_5303_cov_91.4531
EM298_T0.NODE_89_length_5298_cov_10.1745
KT264760.1
EM251_T3.NODE_38_length_4960_cov_73.1339
EM175_T0.NODE_25_length_4843_cov_5.35944
EM200_T3.NODE_42_length_4977_cov_21.0614
EM219_T3.NODE_31_length_5165_cov_7.22466
KT264752.1
EM200_T3.NODE_38_length_5116_cov_136.257
EM200_T3.NODE_39_length_5099_cov_866.953
KT264754.1
EM219_T3.NODE_29_length_5265_cov_11.342
EM173_T3.NODE_24_length_5157_cov_1450.92
EM200_T3.NODE_40_length_5074_cov_108.958
EM274_T0.NODE_13_length_5066_cov_8.31451

Figure 4.

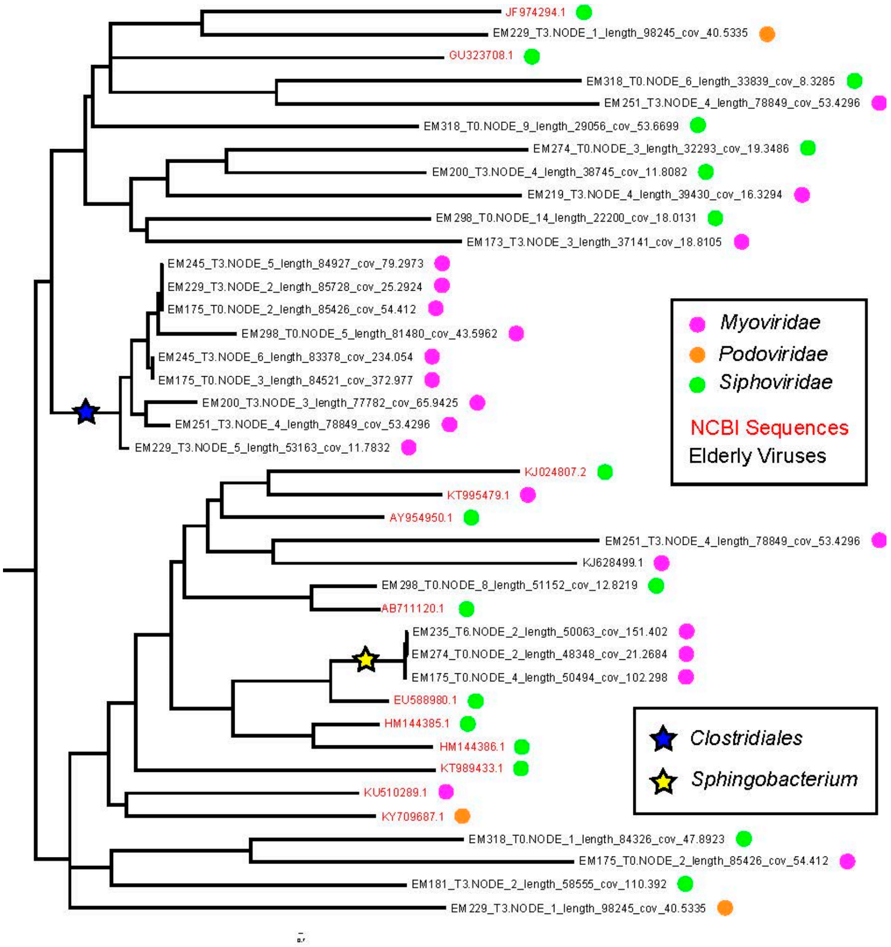
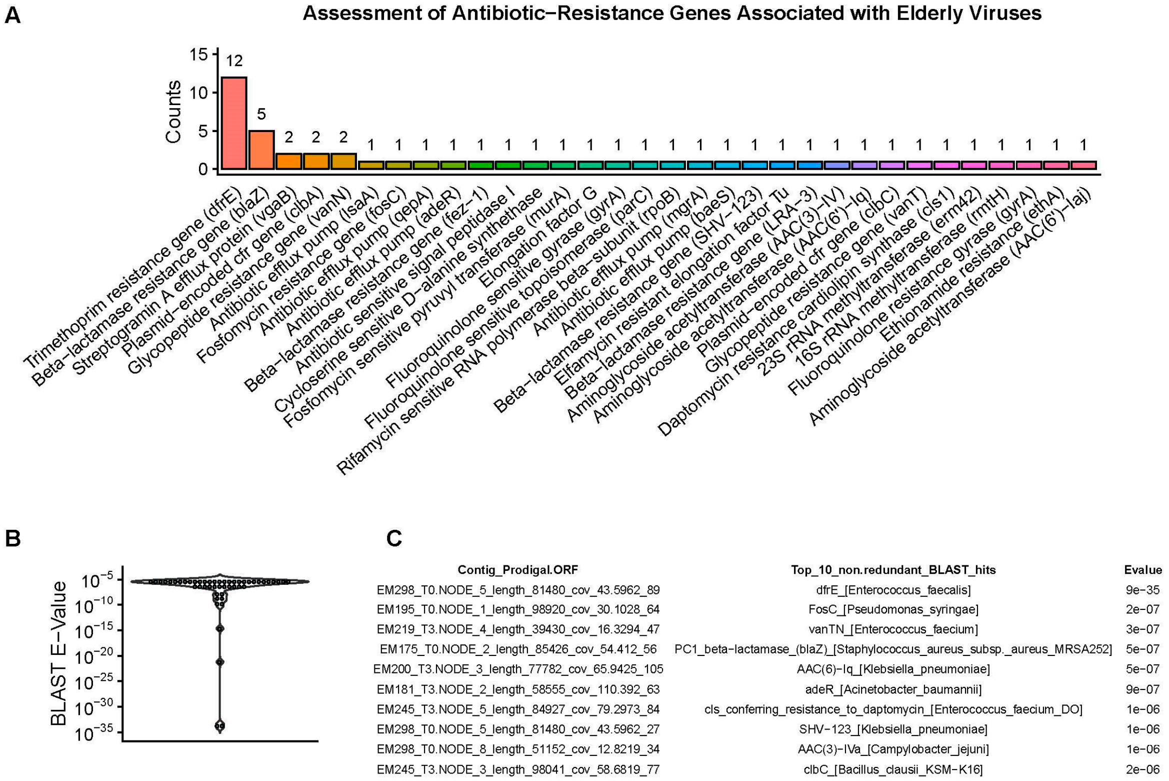

JF974294.1
EM229_T3.NODE_1_length_98245_cov_40.5335
GU323708.1
EM318_T0.NODE_6_length_33839_cov_8.3285
EM251_T3.NODE_4_length_78849_cov_53.4296
EM318_T0.NODE_9_length_29056_cov_53.6699
EM274_T0.NODE_3_length_32293_cov_19.3486
EM200_T3.NODE_4_length_38745_cov_11.8082
EM219_T3.NODE_4_length_39430_cov_16.3294
EM298_T0.NODE_14_length_22200_cov_18.0131
EM173_T3.NODE_3_length_37141_cov_18.8105
EM245_T3.NODE_5_length_84927_cov_79.2973
EM229_T3.NODE_2_length_85728_cov_25.2924
EM175_T0.NODE_2_length_85426_cov_54.412
EM298_T0.NODE_5_length_81480_cov_43.5962
EM245_T3.NODE_6_length_83378_cov_234.054
EM175_T0.NODE_3_length_84521_cov_372.977
EM200_T3.NODE_3_length_77782_cov_65.9425
EM251_T3.NODE_4_length_78849_cov_53.4296
EM229_T3.NODE_5_length_53163_cov_11.7832
KJ024807.2
KT995479.1
AY954950.1
EM251_T3.NODE_4_length_78849_cov_53.4296
KJ628499.1
EM298_T0.NODE_8_length_51152_cov_12.8219
AB711120.1
EM235_T6.NODE_2_length_50063_cov_151.402
EM274_T0.NODE_2_length_48348_cov_21.2684
EM175_T0.NODE_4_length_50494_cov_102.298
EU588980.1
HM144385.1
HM144386.1
KT989433.1
KU510289.1
KY709687.1
EM318_T0.NODE_1_length_84326_cov_47.8923
EM175_T0.NODE_2_length_85426_cov_54.412
EM181_T3.NODE_2_length_58555_cov_110.392
EM229_T3.NODE_1_length_98245_cov_40.5335
● *Myoviridae*
● *Podoviridae*
● *Siphoviridae*
NCBI Sequences
Elderly Viruses
★ *Clostridiales*
★ *Sphingobacterium*
0.7

Figure 5.



Contig_Prodigal.ORF	Top_10_non.redundant_BLAST_hits	Evalue
EM298_T0.NODE_5_length_81480_cov_43.5962_89	dfrE_[Enterococcus_faecalis]	9e−35
EM195_T0.NODE_1_length_98920_cov_30.1028_64	FosC_[Pseudomonas_syringae]	2e−07
EM219_T3.NODE_4_length_39430_cov_16.3294_47	vanTN_[Enterococcus_faecium]	3e−07
EM175_T0.NODE_2_length_85426_cov_54.412_56	PC1_beta−lactamase_(blaZ)_[Staphylococcus_aureus_subsp._aureus_MRSA252]	5e−07
EM200_T3.NODE_3_length_77782_cov_65.9425_105	AAC(6)−Iq_[Klebsiella_pneumoniae]	5e−07
EM181_T3.NODE_2_length_58555_cov_110.392_63	adeR_[Acinetobacter_baumannii]	9e−07
EM245_T3.NODE_5_length_84927_cov_79.2973_84	cls_conferring_resistance_to_daptomycin_[Enterococcus_faecium_DO]	1e−06
EM298_T0.NODE_5_length_81480_cov_43.5962_27	SHV−123_[Klebsiella_pneumoniae]	1e−06
EM298_T0.NODE_8_length_51152_cov_12.8219_34	AAC(3)−IVa_[Campylobacter_jejuni]	1e−06
EM245_T3.NODE_3_length_98041_cov_58.6819_77	clbC_[Bacillus_clausii_KSM−K16]	2e−06