DOCTOR OF PHILOSOPHY

Community-scale viral ecology of agricultural soils and implications for biosolids disposal, wastewater treatment and public health

Hillary, Luke

Award date:
2022

Awarding institution:
Bangor University

Community-scale viral ecology of agricultural soils and implications for biosolids disposal, wastewater treatment and public health

Luke Hillary

August 2022

PRIFYSGOL
BANGOR
UNIVERSITY

A thesis submitted to Bangor University
in candidature for the degree
Philosophiae Doctor

School of Natural Sciences
Bangor University, Deiniol Road, Bangor, LL57 2UW

Declaration

I hereby declare that this thesis is the results of my own investigations, except where otherwise stated. All other sources are acknowledged by bibliographic references. This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree unless, as agreed by the University, for approved dual awards. I confirm that I am submitting this work with the agreement of my Supervisor(s).

Yr wyf drwy hyn yn datgan mai canlyniad fy ymchwil fy hun yw'r thesis hwn, ac eithrio lle nodir yn wahanol. Caiff ffynonellau eraill eu cydnabod gan droednodiadau yn rhoi cyfeiriadau eglur. Nid yw sylwedd y gwaith hwn wedi cael ei dderbyn o'r blaen ar gyfer unrhyw radd, ac nid yw'n cael ei gyflwyno ar yr un pryd mewn ymgeisiaeth am unrhyw radd oni bai ei fod, fel y cytunwyd gan y Brifysgol, am gymwysterau deuol cymeradwy. Rwy'n cadarnhau fy mod yn cyflwyno'r gwaith hwn gyda chytundeb fy Ngoruchwyliwr (Goruchwylwyr)

Thesis Summary

In the UK, 70% of solid waste from wastewater treatment processes is recycled to land as agricultural fertiliser. As wastewater treatment works collect viruses shed in faeces and urine of the population in their catchment areas, these biosolids could potentially contaminate agricultural land with various human, plant and animal pathogens. Biosolids also represent a valuable source of nutrients for sustainable food production, however the impact of their use on the soil virus community, and the ability of biosolids-associated viruses to persist, has yet to be explored at a community scale.

The aims of this thesis were to assess the impact of biosolids amendment on the soil virus community, assess the persistence of biosolids-associated viruses over different timescales, and to develop techniques that expand the characterisation of terrestrial virus communities. This thesis reviews our current knowledge of terrestrial viral ecology, the techniques used to investigate viruses in the environment and the production and use of wastewater derived biosolids (Chapter 1), before examining the long-term impact of biosolids amendment on the soil virus community over 25 years and comparing this with soils that had historically been amended with biosolids 19 years previously (Chapter 2). The study found no difference in overall virus community diversity, and a 100-fold reduction in the relative abundance of biosolids-associated viruses between long-term and historically amended soils. To build on this understanding, the impact of a single amendment of biosolids on the soil viral community was assessed under controlled conditions (Chapter 3). This revealed that at the point of amendment, biosolids import substantial quantities of viruses into the soil virus community, causing a reduction in overall diversity but after one year, biosolids-associated viruses had reduced eight-fold in their relative abundance.

Neither experiment detected any human pathogens. However, many pathogenic viruses are RNA based, and no studies of soil RNA viruses existed at the start of this research. To fill this knowledge gap, a feasibility study on the application of viromics to studying soil RNA viruses was carried out on an altitudinal productivity gradient of grassland soils (Chapter 4). This study identified 3,462 novel soil RNA viruses and demonstrated that RNA virus communities contain viruses of a wide range of host organisms, whereas soil DNA viral ecology studies are often dominated by viruses of bacteria.

The emergence of SARS-CoV-2 and subsequent COVID-19 pandemic highlighted the impact that viruses continue to have on public health. To meet this emerging threat, a study was conducted on the use of wastewater based epidemiology to monitor the prevalence and diversity of SARS-CoV-2 within wastewater treatment plant catchments (Chapter 5). Quantities of SARS-CoV-2 RNA in influent wastewater correlated with local clinical cases and deaths from COVID-19, and the genetic diversity of SARS-CoV-2 in wastewa-

ter mirrors that observed in clinical testing. This work also formed the basis for national surveillance programmes within the UK.

This body of work clearly demonstrates how viral ecology can develop our understanding of the diversity of soil ecosystems and the potential roles that viruses can play in their function via their influence on host organisms. It has also demonstrated how viral ecology can be used to assess the impact of land-management on soil virus communities, and monitor the spread of viral pathogens. Future work will continue to build on this understanding by examining how active virus infections integrate into microbial and macrobiological soil community dynamics, and expand our knowledge of non-DNA bacteriophage viruses in these environments. Finally, the knowledge gained will also aid in improving global soil health, food security and reducing the burden on society of COVID-19 and future pandemic diseases.

Publications and Policy Impact

Publications

Chapter 4 has been published in ISME Communications:

Hillary, L.S. et al., 2022. RNA-viromics reveals diverse communities of soil RNA viruses with the potential to affect grassland ecosystems across multiple trophic levels. ISME Commun. 2, 34. doi: 10.1038/s43705-022-00110-x

Chapter 5 has been published in Water Research:

Hillary, L. S. et al., 2021. Monitoring SARS-CoV-2 in municipal wastewater to evaluate the success of lockdown measures for controlling COVID-19 in the UK. Water Res. 200, 117214. doi: 10.1016/j.watres.2021.117214.

In addition to the primary research chapters above, I have contributed to the following review and method articles during the course of the PhD:

Jones, D. L. et al., 2020. Shedding of SARS-CoV-2 in feces and urine and its potential role in person-to-person transmission and the environment-based spread of COVID-19. Sci. Total Environ. 749, 141364.

Farkas, K. et al., 2020. Viral indicators for tracking domestic wastewater contamination in the aquatic environment. Water Res. 181, 115926

Farkas, K., Mannion, F., Hillary, L. S., Malham, S. K. & Walker, D. I., 2020. Emerging technologies for the rapid detection of enteric viruses in the aquatic environment. Curr. Opin. Environ. Sci. Heal. doi:10.1016/j.coesh.2020.01.007.

Farkas, K., Hillary, L. S., Malham, S. K., McDonald, J. E. & Jones, D. L., 2020. Wastewater and public health: the potential of wastewater surveillance for monitoring COVID-19. Curr. Opin. Environ. Sci. Heal. 17, 14–20.

Farkas, K. et al., 2021. Concentration and Quantification of SARS-CoV-2 RNA in Wastewater Using Polyethylene Glycol-Based Concentration and qRT-PCR. Methods Protoc. 4, 17.

Policy

The results from this thesis have been communicated with industrial sponsor Welsh Water, and will contribute to their future planning and management of biosolids production and usage. The project described in Chapter 5 led to the development of both the Welsh and English national SARS-CoV-2 wastewater monitoring programmes that continue to inform the response of the UK and Welsh governments to the Covid-19 pandemic. As part of my involvement in this, I aided the establishment of the Welsh Environmental Wastewater Analysis and Surveillance for Health (WEWASH) programme by contributing to method and logistical developments. I led the nomination of WEWASH for an Institute of Water Innovation award, with the project placing first in Wales, and third overall within the UK.

Outreach

Results from this study were shared through various outreach activities including interviews with the Microbiology Society and Times Radio, and presentations to external organisations, including the University of the Third Age, the Institute of Water, and Welsh Water. During the PhD I participated in the education outreach programme I'm A Scientist's Environment Zone, and again, by invitation, in their Pandemic Preparation event in partnership with the Royal Institute Christmas Lectures.

Acknowledgements

I would like to start by thanking my supervisors Prof. James McDonald and Prof. Davey Jones for their support, both professionally and personally, over the past 4+ years. Their guidance, encouragement, feedback and support through a period of unprecedented global challenge was essential in the production of this work.

This thesis would not have been possible without the input from multiple informal advisors. What Kata Farkas doesn't know about qPCR is not worth knowing, and I will always be grateful for her reminders that whatever can go wrong, will go wrong, this is how to deal with it when it does, and just occasionally, sometimes, things can actually go right! Evelien Andrienssens' support in the grant application, data analysis and publication of Chapter 4 was critical to getting this chapter funded and completed and her knowledge of viral taxonomy and bioinformatics has been invaluable.

Speaking of funding, I owe much gratitude to my funders, Welsh Water, and NERC, through the STARS CDT. Welsh Water have been incredibly supportive with the logistics of sample collection/ transport and also financially backing much of the work in Chapter 3 and their input was also essential for the work in Chapter 5. In particular, I would like to thank Gemma Hall and Faye Ward for acting as my points of contact and the numerous wastewater treatment plant operators who were involved in the sampling efforts of this thesis.

The Soils Training And Research Studentships Centre for Doctoral Training (STARS CDT) has been like an extended family during this PhD. The opportunities provided by this training-intensive, cohort-centric, multi-institution PhD programme has been invaluable. Special thanks should also go to the STARS management board, Phil Haygarth and STARS administrator Olivia Lawrenson for their passionate commitment to providing the best possible PhD experience under some significantly challenging circumstances related to the ongoing COVID-19 pandemic.

I would like to extend my thanks to the entirety of the Microbial Ecology Group who I've had the privilege of working with during my time at Bangor. I would like to specifically mention David Fidler for chatting climbing, stats and coding and Beth Pettifor, with whom I often set the world (or at least Bangor) to rights and who kept me (mostly) on the straight and narrow during what has been on occasion, a turbulent time. I would also like to expand my thanks to the whole of the Molecular Ecology and Evolution at Bangor (MEEB) team, a great set of collaborators, and particularly to Sarah Coates, for being a huge source of encouragement as I was getting this thesis over the line.

Finally, I would like to thank my family for instilling a curiosity and love of learning in me that has been the source of my motivation throughout the PhD and will drive me forward into the next phase of my scientific career.

Contents

List of Figures

List of Tables

1

Introduction

1.1. Foreword

Wastewater harbours a range of enteric human pathogenic viruses such as norovirus, human adenovirus and sapovirus that can persist through the wastewater treatment process (Adriaenssens et al., 2021; Bibby and Peccia, 2013; Farkas et al., 2018a, 2018b; Hewitt et al., 2011; Meschke and Sobsey, 1998). This creates both a source of information on viral outbreaks at a community scale, and a concern that pathogenic viruses could be spread through the environment via discharge of raw sewage and through ineffective treatment of liquid and solid effluents from municipal wastewater treatment plants (WWTPs). The UK produces 1.8 million tonnes of wastewater sludge, or biosolids, each year with 70% being applied to agricultural land (Kelessidis and Stasinakis, 2012). Microbiological safety monitoring of these sludges is limited in the UK to the enumeration of faecal bacteria such as *Escherichia coli* and *Salmonella spp.* (Assured Biosolids Limited, 2020) and consequently, no routine evaluation of the efficacy of sewage treatment on the removal of viruses currently exists. In addition to the transfer of human pathogenic viruses to soil, application of biosolids to agricultural land also has the capacity to affect the intrinsic soil virus community. Soils are fundamental to terrestrial ecosystems, providing a complex environment that supports both green and brown food webs. In addition, good soil health is essential to safeguarding food security and other ecosystem services in an increasingly overpopulated world.

This thesis aims to assess the persistence of biosolids-associated viruses and the viral risk to human and soil health from the use of biosolids as an agricultural fertiliser. The introductory chapter first examines the range of viral diversity, their broad phylogeny, origins and evolution, before considering the current state of our knowledge of terrestrial

viral ecology. It is followed by an examination of the workflow used for molecular viral ecology, and how the methods used are affected by the scientific questions being examined. Our understanding of viruses within the wastewater treatment process is then considered, alongside the impact of the COVID-19 pandemic on the research field. Finally, the future challenges of this area of research are outlined with the aims and objectives of the thesis.

1.2. Viruses and the virosphere

1.2.1. Viral classification and taxonomy

Viruses are the most numerate and diverse biological entities on Earth, and exert significant influence on global biogeochemical processes (Breitbart et al., 2004; Emerson et al., 2018). This scale of diversity has a substantial impact on how viral ecology is studied, and so to fully appreciate the significance and complexity of viruses and subviral agents, we must first address the question *"what is a virus?"*. Life can be divided into two separate "empires" of reproducers (cellular life) and replicators (viruses and other mobile genetic elements, or MGEs) (Koonin, 2010). True viruses have traditionally been defined as nucleic acid genomes surrounded by protective protein coats, or capsids, that replicate, rather than reproduce, by infecting a host cell and hijacking its metabolism to manufacture additional virus particles, or virions. However, as our understanding of the diversity of viruses and other MGEs has developed, it has been necessary to adjust that definition (Koonin et al., 2021). This increased knowledge is reflected in the most recent International Committee on Taxonomy of Viruses (ICTV) working definition of what viruses are, and how they relate to other MGEs:

> "Viruses *sensu stricto* are defined operationally by the ICTV as a type of MGE that encode at least one protein that is a major component of the virion encasing the nucleic acid of the respective MGE and therefore the gene encoding the major virion protein itself; or MGEs that are clearly demonstrable to be members of a line of evolutionary descent of such major virion protein-encoding entities. Any monophyletic group of MGEs that originates from a virion protein-encoding ancestor should be classified as a group of viruses." (Kuhn et al., 2020)

Along with the definition of what a virus is, the concept of the virosphere, the total range of possible forms that viruses may take, lying within a broader replicator space has emerged and can be further subdivided into the orthovirosphere, containing true viruses, and perivirosphere, containing virus-like replicators:

- " 'orthovirosphere': the part of the virosphere represented by MGEs considered to be "viruses" per the new operational definition;

- 'perivirosphere': the part of the virosphere represented by MGEs that functionally resemble viruses but do not fulfill the operational definition of viruses (e.g., satellite nucleic acids, viroids, prokaryotic gene transfer agents (GTAs), and MGEs currently classified in taxon Polydnaviridae)."(Kuhn et al., 2020)

The viruses found within the orthovirosphere can be further subdivided, and this was traditionally done according to the Baltimore classification system, based on genome structure (Baltimore, 1971, see Fig. 1.1). Viral genomes can be formed from DNA or RNA, and comprise single stranded (ss) or double stranded (ds) molecules. In ssRNA viruses, the direction of sense can be used to divide further. Positive sense (+) nucleic acids have sequences that match the mRNA precursors of viral proteins. Negative sense (-) nucleic acids must first undergo an RNA dependent transcription step before protein translation. Double stranded genomes are often ambisense (+/-) and carry both positive and negative sense elements. Reverse-transcribing viruses (Baltimore Class VI and VII) encode reverse transcriptase genes which synthesise copies of genomic DNA using an RNA intermediate (Krupovic et al., 2022). Class VI viruses possess ssRNA genomes that are reverse-transcribed to DNA and often integrate into host genomes as part of their replication cycle. Class VII form dsDNA genomes, but replicate via an RNA intermediate (Nassal and Schaller, 1993).

Although the Baltimore classification creates morphological distinctions between the different classes of viruses within the orthovirosphere, the recent explosion in high throughput sequencing data and resultant increased knowledge of macro-scale viral phylogeny has revealed extensive viral diversity that does not neatly map onto the seven class system (Koonin et al., 2020). A new viral megataxonomy has subsequently been adopted (see Fig. 1.2), that divides the orthovirosphere into six realms based on both genomic material and the phylogeny of their major capsid protein (MCP) or other viral hallmark gene:

- *Riboviria* - The vast majority of RNA viruses, e.g. SARS-CoV-2 (Phan, 2020), and retroviruses, e.g. HIV (Rambaut et al., 2001). This group also includes some MGEs that have evolved from RNA viruses but lost the structural genes from their genomes, e.g. *Narnaviridae* (Hillman and Cai, 2013).

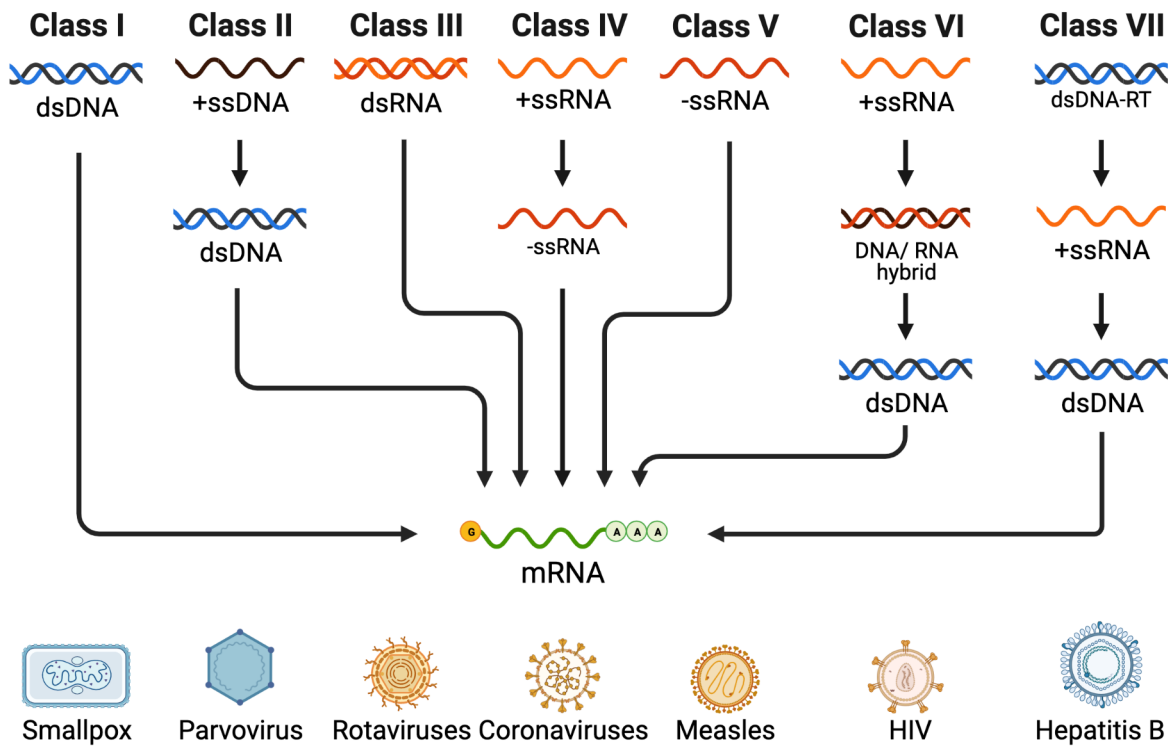


Figure 1.1.: The Baltimore classification system of viruses, based on the nucleic acids that form the viral genome, with examples of each class. Note that many class IV viruses are also capable of directly translating their +ssRNA genomes into protein.

- *Ribozyviria* - Negative sense, circular single stranded RNA (-ssRNA) viruses related to hepatitis delta virus 1. These viruses may have originated from viroids by gaining their MCP, however they are dependent on co-infection alongside helper virurses for replication (Koonin et al., 2021).

- *Monodnaviria* - Predominantly single stranded DNA (ssDNA) viruses encoding a HUH superfamily endonuclease. These are thought to have evolved multiple times from bacterial or archaeal plasmids (Krupovic, 2013).

- *Duplodnaviria* - Double stranded DNA (dsDNA) viruses possessing a HK97-fold MCP, including the class *Caudoviricetes* of tailed bacteriophages and herpesviruses (Koonin et al., 2020).

- *Varidnaviria* - tailless dsDNA viruses encoding a double jelly roll (DJR) MCP. They include nucleo-cytoplasmic large DNA viruses (NCLDVs) and the clinically/ economically important animal pathogens adenoviruses and poxviruses (Woo et al.,

2021).

- *Adnaviria* - filamentous archaeophages that lack an identifiable replication machinery within their genomes, but contain an α-helical MCP that converts viral DNA from B-form to the more compact A-form DNA (Krupovic et al., 2021).

In addition to the six realms of the orthovirosphere, there are some notable groups occupying the perivirosphere:

- Viroids - infectious circular non-coding RNAs that rely on host cell machinery or helper viruses for replication and do not form encapsidated virions (Dickson, 2018).

- Satellite RNAs - non-coding RNAs similar to viroids but are dependent on helper viruses for encapsidation (Palukaitis, 2016).

- Satellite DNAs - circular ssDNA molecules with similar replication cycles to satellite RNAs, sometimes non-coding (Adams et al., 2017).

- Viriforms - virus-derived MGEs that have been utilised by their hosts for cellular functions (ICTV, 2021).

Beyond the virosphere, but still within replicator space, lie other non-viral replicators, such as plasmids, DNA transposons, and conjugative transposons, otherwise known as integrative conjugative elements. The boundaries between the orthovirosphere, perivirosphere and this wider replicator space remain porous, as some MGEs blur these boundaries by having features of two different regions, while others may have belonged to one region in their evolutionary past, but crossed a threshold to another through gain or loss of function (Koonin et al., 2021). Virus-derived elements continue to exist beyond replicator space in the form of gene transfer agents (GTAs). These *Caudoviricetes* derived structures are encoded within bacterial genomes and facilitate horizontal gene transfer of random genome fragments, but do not possess mechanisms to ensure the transfer of genes required for replication (Lang et al., 2012).

All viruses require a host reproducer to facilitate their replication and the distribution of susceptible hosts is uneven across different viral realms (see Fig. 1.2), with few known

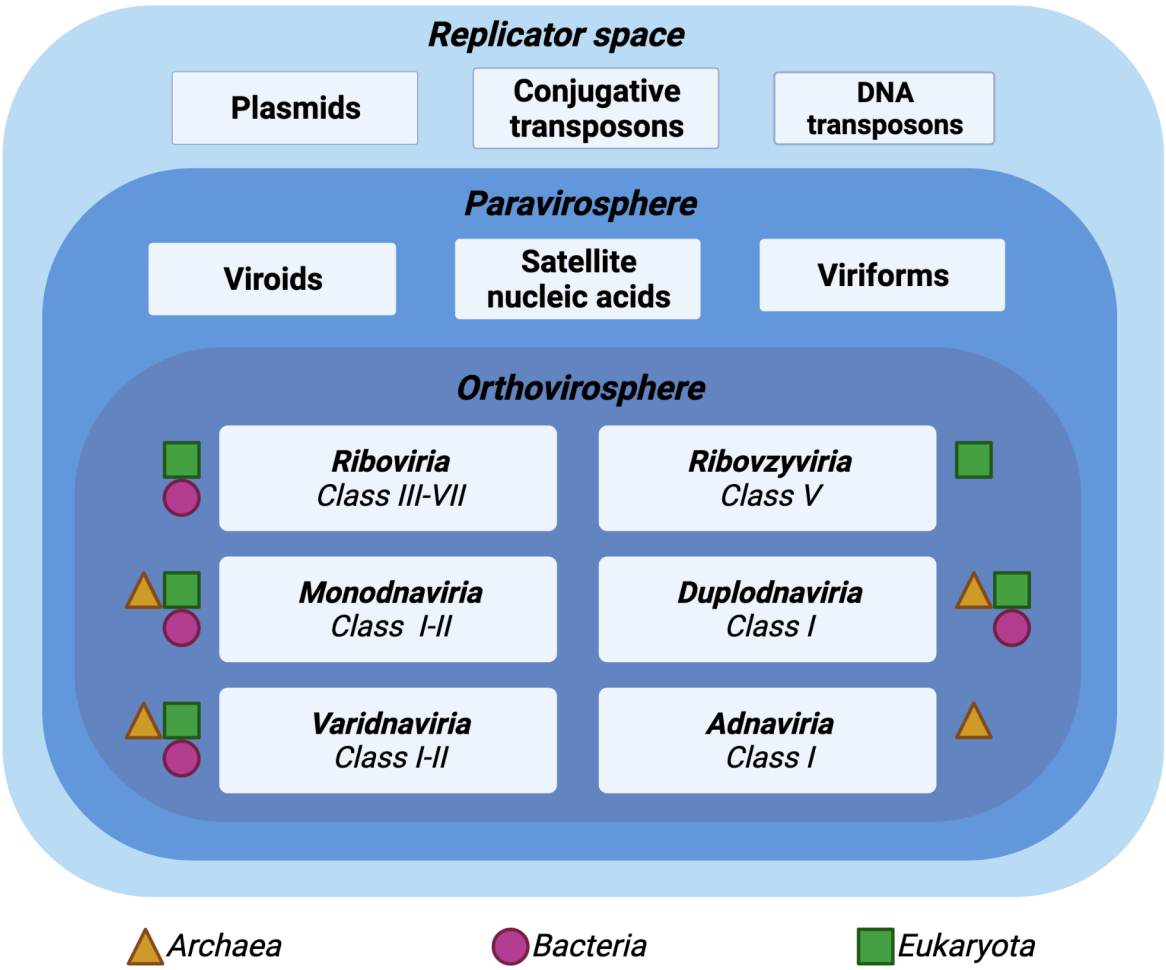


Figure 1.2.: The virosphere model of viruses and their relationship with other replicators. Identities of different replicators are in bold, with Baltimore system classes of virus mapped onto the six virus realms of the orthovirosphere and known potential hosts indicated by symbols. Figure adapted from Krupovic *et al.* (2021)

bacterial, and no archaeal *Riboviria* (Koonin et al., 2020). There are several possible explanations for this that are not mutually exclusive, and each could apply to different viral realms. RNA viruses have been demonstrated to be major components of ocean viral communities but are significantly understudied, compared to their DNA counterparts (Steward et al., 2013). With the exception of *Adnaviria*, animal and plant hosts have known viruses from each realm which may more of a reflection of viruses' significant importance to public health and food security. In contrast, fungal viruses are almost exclusively from the *Riboviria* realm, with some *Geminiviridae* ssDNA viruses from the *Monodnaviria*, and often share evolutionary lineages with plant viruses at lower taxonomic levels (Roossinck, 2019). These observed high-level taxonomic differences in virus-host relationships may be a function of evolutionary change/ biological susceptibility of differ-

ent hosts, or may be due to biases in global research focus. As some realms share more distantly related hosts than others, it is also possible that some viral realms originating after divergence from a common host ancestor and loss of extensive lineages of viruses has also contributed to the observable imbalance in host distribution.

1.2.2. Viral structural diversity

As well as genetic diversity, viruses show substantial morphological diversity (see Fig. 1.3). As discussed above, most true viruses are capable of producing extracellular particles, called virions, which share a common core structure of a nucleic acid genome surrounded by a protein capsid. The capsid is formed by multiple repeats of viral structural proteins called protomers. The repeating unit, which can comprise more than one protomer, is referred to as a capsomere. These protein coats serve to protect the nucleic acids inside and to mediate infection of host cells. Virion capsid diameters range in size from the 17 nm porcine circovirus to the aptly named 440 nm *Megavirus chilensis* (Allan and Ellis, 2000; Arslan et al., 2011). The nucleocytoplasmic large DNA viruses (NCLDVs) also rival cellular life in their complexity. Whilst they are superficially similar to icosahedral viruses, they have a modified vertex or 'stargate' through which their large genomes are able to pass into host cells (Kuznetsov et al., 2010). As an example of genomic complexity, the NCLDV *Mimivirus* genome contains numerous accessory genes involved in DNA repair, metabolic pathways and protein translation, including four tRNA synthetases previously only found in cellular genomes (Raoult et al., 2004).

Viral nucleocapsid structure can be divided into two categories: polyhedral and helical (see Fig. 1.3a). Helical viruses such as tobacco mosaic virus show a helically repeating pattern of structural proteins surrounding the core genome whereas polyhedral viruses are sphere-like shapes. Both helical and polyhedral capsids can be surrounded by a lipid envelope formed from host cell or organelle membranes. These lipid envelopes are often embedded with viral glycoproteins that mediate infection of new host cells, e.g. hepatitis C virus (Lyu et al., 2015). Tailed bacteriophages of the *Duplodnaviria* realm pictured in Fig. 1.3c contain polyhedral and helical elements with protein fibres that bind to cell receptors, triggering the injection of genomic material into the host cell (Ofir and Sorek, 2018). The mechanism of tailed phages have been intensively studied as a model nano-

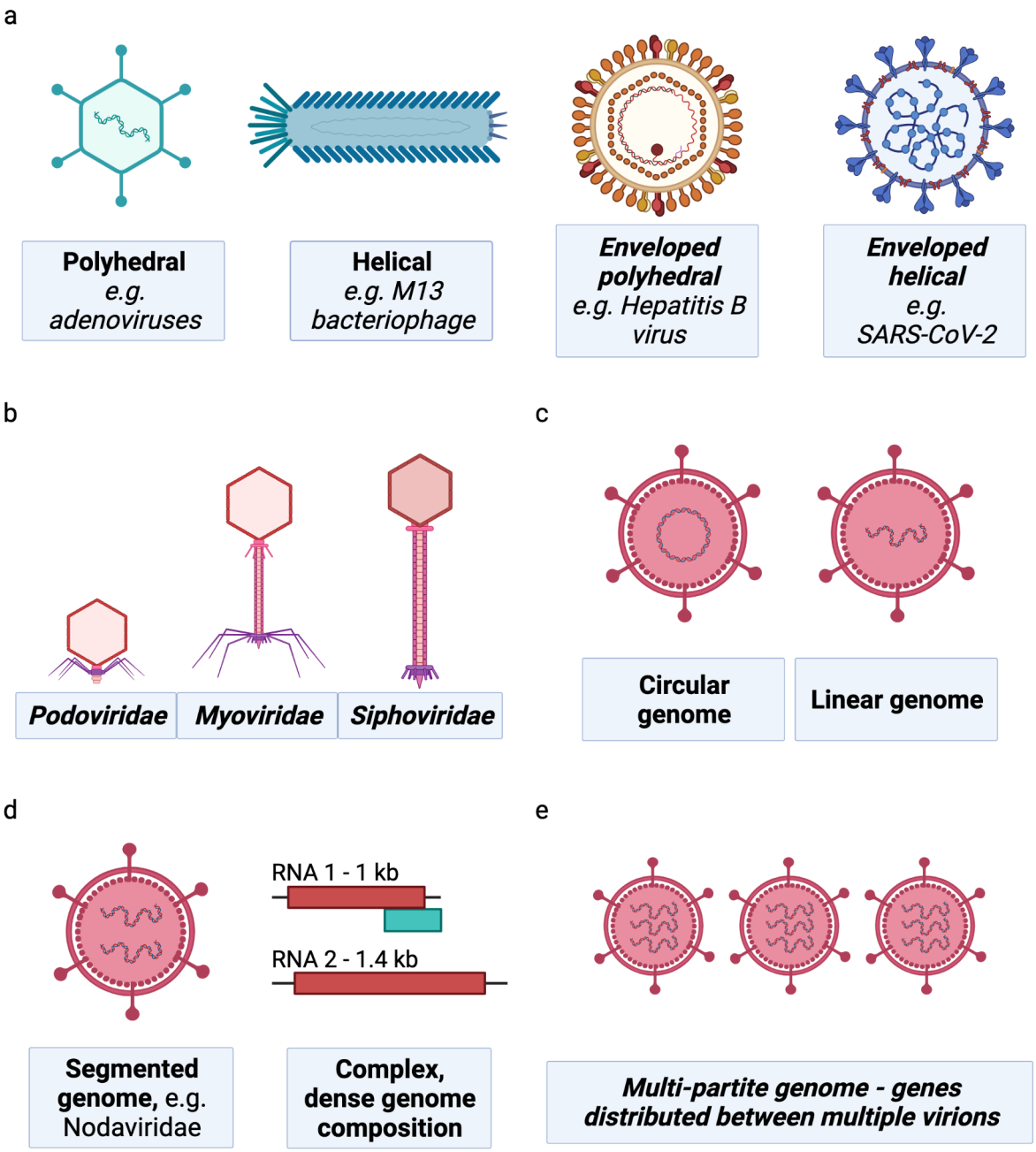


Figure 1.3.: Structural diversity of viruses. Viral particles, or virions can be (a) polyhedral or helical, with or without a lipid envelope. (b) More complex structures exist within the *Caudoviricetes* class. As well as physical structural variation, there is substantial variation in viruses' genomic structure (c, d, e)

machine with potential applications in drug delivery (Lam and Steinmetz, 2018).

In addition to morphological diversity, viruses also display extensive diversity in the structures of their genomes beyond that seen in the Baltimore classification scheme. Viruses can posess circular or linear genomes (Fig. 1.3c) and these can be found in one or more segment (Fig. 1.3d). In addition to segmentation, multipartite viruses can package their genomes across multiple virion particles which require simultaneous infection, sometimes in different host tissues, to enable viral replication (Fig. 1.3e) (Lucía-Sanz and Manrubia, 2017).

1.2.3. Virus replication strategies

The dependency of viruses on cellular organisms for replication typically necessitates an intracellular and extracellular phase of a virus's replication cycle, but there are notable exceptions: some plant and fungal viruses are obligately intracellular and transmit vertically through transovarial passage or hyphal fusion (Hillman and Cai, 2013; Jia et al., 2018). While in the intracellular phase, the viral nucleic acids become detached from the viral coat proteins and, in the case of viruses with a DNA phase of their replication cycle, can become integrated into the host genome.

Table 1.1.: Viral lifecycle strategies (reviewed in Howard-Varona et al, (2017)

Lifecycle strategy	Description
Lysis	Continuous production of virions resulting in cell lysis.
Lysogeny	Stable infection with suppression of virion production. The viral genome is equally distributed during host cell division and often, but not always, integrated into the host cell genome.
Pseudolysogeny	An unstable infection where the viral genome is maintained in the host cell but neither virion production or genome replication occurs. Viral genomes are divided asymmetrically during cell division.
Chronic infection	Continuous low-level production, sometimes with release of virions without triggering cell lysis.

A viral infection can have four outcomes: lysogenic dormancy, unstable pseudolysogeny, cell lysis or chronic virion release, which are summarised in Table 1.1 (Howard-Varona et al., 2017). Temperate viruses will become lysogenic by synchronising the replication of their genome with that of the host cell and surpressing the production of intact virion particles, becoming dormant in the process. Some viruses additionally possess integrase genes have the ability to integrate their genomes into the host genome, forming a provirus. The switch between lysogenic and lytic replication pathways is highly regulated and can respond to external stimuli (Dou et al., 2018). The detection of temperate phages capably of lysogeny can be performed in three ways:

- Detection of proviruses in metagenomic datasets - there are several specialised tools for identifying the boundaries between host and viral sequences within a host genome, e.g. Prophage Hunter (Song et al., 2019).
- Identification of viral hallmark genes required for lysogeny within viral genomes, such as integrase genes, e.g. with CheckV or VIBRANT (Kieft et al., 2020; Nayfach et al., 2020).
- Treating samples with chemicals or UV light to induce virion production (Raya and H'bert, 2009).

In addition to lysis or lysogeny, some viruses are capable of chronic infections affecting only a portion of the total host population, e.g. ϕCrAss001 (Shkoporov et al., 2018). Chronic infections can also occur without triggering host cell lysis where new virion particles bud off from host cell membranes (Welsch et al., 2007).

1.2.4. Viral ecology of terrestrial ecosystems

The global virosphere of 10^{31} viruses represents the greatest reservoir of genetic diversity on the planet (Breitbart and Rohwer, 2005). For example, marine sediments contain on average 10^4 viral genotypes per kg (Breitbart et al., 2004). Whilst the majority of these are thought to infect bacteria, viruses infect most known unicellular and multicellular organisms. The occurrence of the same viral species in diverse biomes suggest that some viruses are capable of moving between different habitats (Breitbart and Rohwer, 2005). A comparison of Arctic and Antarctic viral communities showed broad overlap between polar

freshwater environments that were distinct from marine viromes, suggesting viral exchange between the two poles (Aguirre de Carcer et al., 2015). Clustering viral metagenome datasets often groups samples taken from similar environments together (Han et al., 2017) and an extensive analysis of oceanic DNA viromes revealed the division of the global ocean viral community into five distinct ecological zones (Gregory et al., 2019). However, whilst the role of viruses in marine environments has been studied in detail, viruses in terrestrial ecosystems are less well understood (Williamson et al., 2017).

Fig. 1.4 describes how a terrestrial virus community is influenced by different factors, either directly by impacting the viral community itself, or indirectly by affecting the host community. In this model, there is no distinction made between microbial hosts and macrobiological hosts, e.g. insects, plants, etc, as the principles remain the same for both cases. A distinction can, and should, be made when examining organism-associated viral communities, e.g. gut viromes, as these harbour both viruses of the macro-organism providing the habitat, and viruses of its microbiome. Viromics studies and those using counts of Virus-Like-Particles (VLPs) indicate that pH, temperature and soil moisture content are key determinants of viral and microbial community structure and abundance (Adriaenssens et al., 2017; Bi et al., 2021; Wu et al., 2021). Seasonal fluctuations can also influence viral community dynamics with virus-like-particle abundance increasing between February to September in US prairie and cropland soils (Cornell et al., 2021). This study also utilised structural equation modelling to demonstrate the effects of management and environmental factors with both viral and bacterial abundances. In addition to crop cover and tillage, organic manure has been shown to increase both VLP abundance and α-diversity in amended soils (Chen et al., 2014; Chen et al., 2021) but it is unclear whether this effect is a direct result of the import of viruses within the manure or an indirect effect of increased microbial host activity.

As well as being influenced by biotic and abiotic environmental factors, viruses are able to exert an influence on these factors, both directly and through their hosts. A key example of this is the 'viral shunt' that affects nutrient availability in aquatic ecosystems. The viral shunt creates a short-circuit of marine trophic cascades, allowing nutrients to be recycled within microbial communities and it is believed that virus mediated cell lysis is responsible for the daily turnover of 20-75% of marine bacterial life (Fuhrman, 1999; Tsai

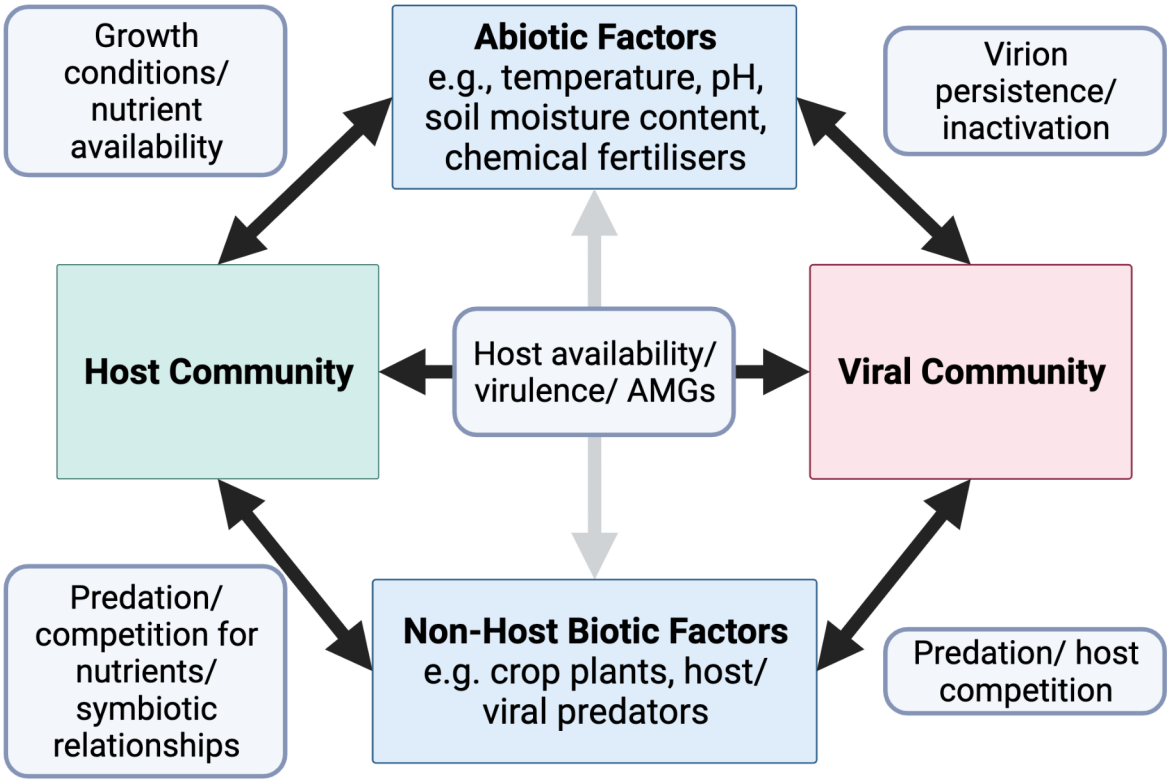


Figure 1.4.: A conceptual model of the interactions between a viral community, host community and other elements of a terrestrial environment. Influences between the viral community, host community and other biotic/ abiotic factors can be direct or indirect, and occur in multiple directions. For example, viruses may confer evolutionary advantages on a host that is capable of growing at higher temperatures, raising the temperature of an anaerobic digester, and so impacting the persistence of extracellular virus particles

et al., 2016; Wilhelm and Suttle, 1999). The combined effect of virus mediated cell lysis is thought to release 0.37 – 0.63 gigatonnes of carbon per year and plays an essential role in the microbial food web of nutrient poor deep-sea ecosystems (Danovaro et al., 2008). At present, the existence/ significance of a viral shunt in other ecosystems is unknown. It is hypothesised that the increased incidence of lysogeny, potentially due reduced connectivity in terrestrial ecosystems, may reduce this impact in soils (Adriaenssens et al., 2017). As the effect of the viral shunt is influenced by the availability of alternative nutrient sources, it is likely to be greater in more nutrient poor soils (Schmidt et al., 2011). The lysogenic fraction of soil viruses was observed to be higher in subsurface than surface soils (Liang et al., 2019) and so any terrestrial viral shunt is more likely to be more closely related to plant nutrient uptake, than to the release of nutrients in nutrient starved conditions at greater depths, in contrast to the the marine viral shunt.

In addition to direct effects on host abundance through cell lysis, viruses can also augment their hosts' cellular functions through the presence of auxiliary metabolic genes (AMGs). These are functional genes found in viral genomes that give the host additional capacity in, e.g. carbon metabolism, but are not directly required for viral replication. Viral AMGs have been shown to have the potential to influence carbon, nitrogen, sulphur and phosphorus metabolism and so their role in global-scale nutrient flow is of significant interest (Emerson et al., 2018; Monier et al., 2017; Roux et al., 2016a; Zeng and Chisholm, 2012).

One class of AMGs that has received significant interest are antimicrobial resistance genes (ARGs). ARG diversity has been shown to increase in soils amended with sewage sludge (Chen et al., 2021) but significant care must be taken when analysing viromics data for ARGs and AMGs more broadly. Enault *et al.* (2016) found that low similarity thresholds can result in the overestimation of ARG content in viral genomes, potentially leading to the role of virus-mediated horizontal gene transfer in the spread of ARGs being overstated. Additionally, viruses possess genes involved in their primary replication that may be misclassified as AMGs if they resemble proteins used in cellular metabolism (Pratama et al., 2021). For example, bacteriophages may utilise carbohydrate binding proteins and peptidases for host cell entry (Shaffer et al., 2020). Specialist bioinformatics tools such as DRAM-v and VIBRANT have been developed for the identification of AMGs however their outputs may require additional manual curation which makes medium or large scale analysis of AMG evolution challenging (Kieft et al., 2020; Pratama et al., 2021; Shaffer et al., 2020).

Identifying the above patterns in viral metagenome data is a significant computational challenge as a substantial proportion of viromics sequencing data remains taxonomically and functionally unassigned. Reanalysis of public databases with improved tools for viral detection often generates new insights on a global scale (Graham et al., 2019; Gregory et al., 2019; Paez-Espino et al., 2016). The drawback of such analyses and the computational tools that underpin them is that they often target dsDNA bacteriophages, and so the large-scale ecological role of viruses of archaea, fungi and unicellular/ multicellular eukaryotes, and the role of ssDNA and RNA viruses remains underexplored.

Whilst mining the viral 'dark matter' in public databases can detect patterns of viral diversity on global scales, it can also reveal single groups of previously unknown viruses that significantly enhance our understanding of the viral world. Cross-Assembly bacteriophages, or crAssphages are highly abundant in the human gut virome and were discovered through cross-assembly of multiple human virome datasets (Dutilh et al., 2014). CrAssphage has subsequently been adapted as a faecal indicator marker for sewage/ human faecal contamination in a variety of environments (Ahmed et al., 2018; Crank et al., 2019; Farkas et al., 2019).

1.3. Methods in molecular viral ecology

The impact of the extensive viral diversity discussed above is such that no one technique can adequately survey the viral community within wastewater-based or terrestrial environments. Virion morphology, degree of lysogeny and genome structure can all impact a virus's detectability in studies of viral abundance/ community composition. This next section examines the workflow commonly applied to the study of viral ecology in terrestrial and wastewater-based environments. Typically, molecular viral ecology employs one of the following five techniques:

1. Size selection of virus-like-particles (VLPs) and sequencing of concentrated/ purified samples, otherwise known as viromics
2. Bulk metagenomic sequencing of environmental samples
3. Single-cell sequencing of either host cells or VLPs
4. Large-scale meta-analysis of pre-existing sequencing datasets
5. PCR-based quantification of known viruses

In addition to the above techniques, additional enrichment steps such as integrated cell culture quantitative polymerase chain reaction (ICC-qPCR) assays, capsid integrity assays and long-read sequencing can be used to further augment the information gained from sequencing/ qPCR based studies. This continually developing toolkit for detecting enteric and biosolids-associated viruses in the environment has recently been reviewed by Farkas *et al.* (2020b)[1] and this section expands upon this review to consider sample processing

[1]Farkas, K., Mannion, F., **Hillary, L.S.**, Malham, S.K. & Walker, D.I. Emerging technologies for the

and the techniques of viromics and qPCR in greater detail.

1.3.1. Extraction of nucleic acids

The first stage of any environmental virology study is the extraction of viral nucleic acids from the sample matrix and the optimum protocol will depend on the downstream analytical techniques to be used. qPCR-based analysis of highly abundant viruses and bulk metagenomics/ metatranscriptomics of soil/ biosolids often utilise direct extractions of DNA and/ or RNA. However viromics and qPCR of less abundant viruses, and studies of aquatic systems often first use additional elution and/or concentration steps prior to nucleic acid extraction to improve sensitivity (see Fig. 1.5).

Viruses readily adhere to soil and sediment particulates and the degree of adherence is heavily affected by soil characteristics and viral species (Bitton, 1975; Meschke and Sobsey, 1998). In order to isolate viral nucleic acids, any VLPs adhering to soil, or biosolids particulates need to be desorbed. This is achieved by agitating the soil in a desorption buffer. The purpose of desorption buffers is to stabilise virion structural integrity during physical disruption and provide a negatively charged competitor ion to displace VLPs and/or to disrupt the ionic bonds between them and soil particles. Several buffers are reported in the literature, but the most commonly used is a potassium citrate (KC) solution buffered by sodium potassium phosphate. This was originally used in a study by Williamson *et al.* (2003) who compared the effects of KC, 10% beef extract, 250 mM glycine buffer and 10 mM sodium pyrophosphate, finding that KC buffer eluted the highest average number of VLPs across two different soil types. This buffer was compared to saline magnesium (SM), phosphate buffered saline (PBS) and sodium pyrophosphate (SP) buffers by Narr et al. (2017) where SM buffer gave 10-fold higher VLP counts. Trubl *et al.* (2016) compared KC and SP buffers to an amended potassium citrate buffer (KC buffer with 5 mM EDTA, 150 mM $MgSO_4$) increased VLP yields by 100 fold across three different three carbon rich soil types when compared to KC and SP.

Following addition of a suitable buffer, the soil suspensions are physically agitated to desorb VLPs from soil particles. Again, there are several options available, namely shaking

rapid detection of enteric viruses in the aquatic environment. Curr. Opin. Environ. Sci. Heal. (2020)

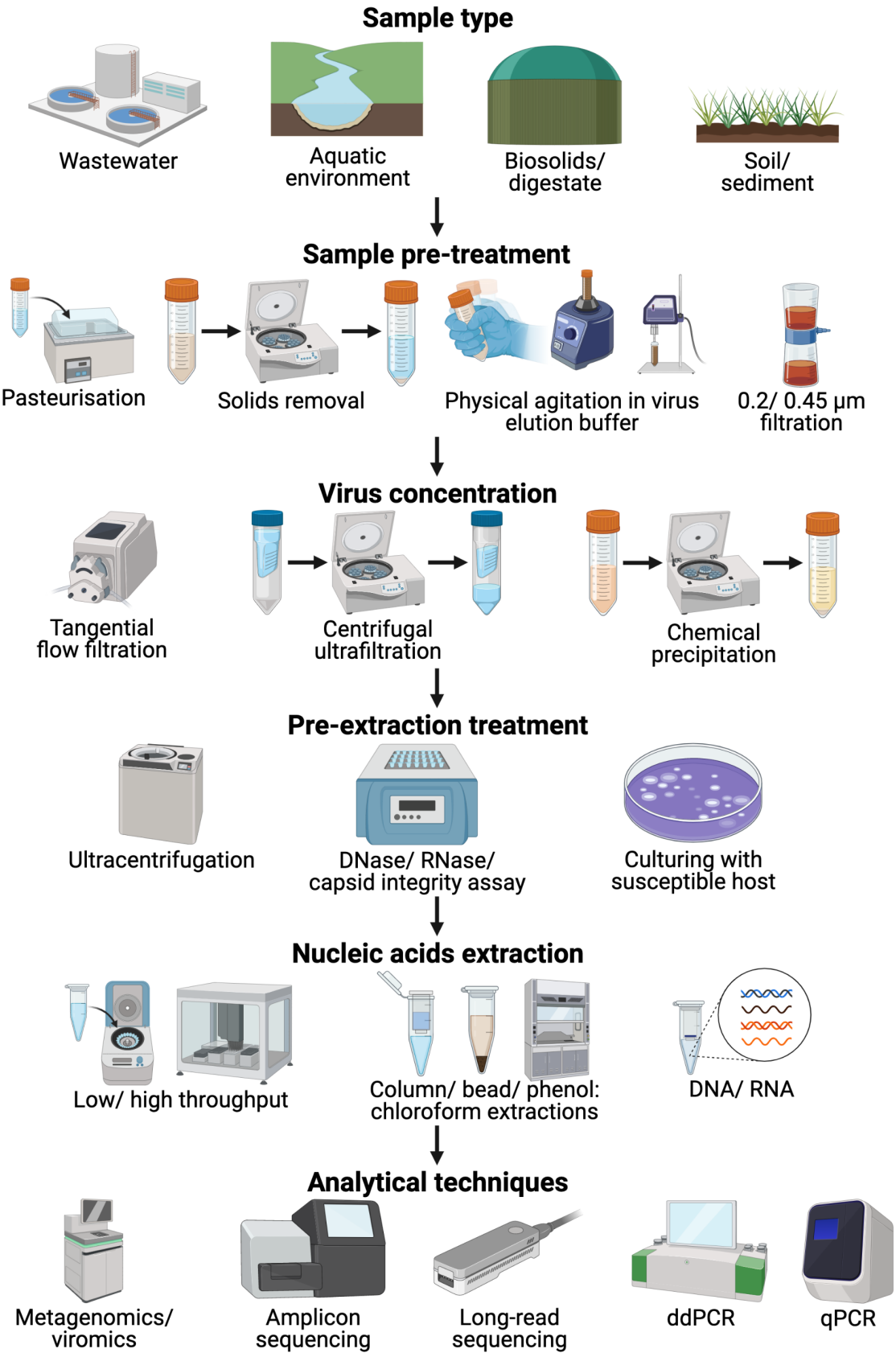


Figure 1.5.: Methods of sample processing in viral ecology (continues on next page...)

Figure 1.5.: (...) Environmental samples are often pre-treated to inactivate pathogens, remove solids or desorb VLPs from solid particulates. These can then be concentrated, further purified, or treated, and then undergo nucleic acid extraction prior to analysis. The choice of which techniques to apply is highly dependent on the research questions being asked and the downstream analytical techniques.

(manual or mechanical), sonication, bead-beating or vortexing. The latter three were also compared by Trubl *et al.* (2016) who found no significant differences between disruption methods. As sonicator power settings, bead types, duration, etc of these methods can vary substantially, it is important to take note of the exact settings used when deciding on an extraction method. It is also important to note that extraction comparison studies typically consider higher VLP counts as 'better', however as Williamson *et al.* (2003) highlight, this is not necessarily true. It is widely recognised that not all VLPs, as detected by EFM or flow cytometry, are actual viruses (Trubl et al., 2020) and so optimising extraction conditions using this measure may not produce the optimum extraction protocol for analysis by viromics or qPCR. For example, the optimum extraction conditions for VLP counts can differ from those for recovering viable phages and some buffer conditions produce background fluorescence that prevents VLP enumeration (Williamson et al., 2003). Mock communities of representative viruses could be used and extraction efficiencies compared via qPCR or amplicon sequencing, however this would add additional complexity and cost. This approach may also lead to a method optimised towards the mock community rather than the whole viral community of the environment to be examined. Ultimately, the decision on which method to use depends as much on the study specifics, research questions being asked and availability of equipment as it does on the results of optimisation comparisons. It is also possible that the relative influence of desorption buffer and physical disruption method will vary by soil type and the optimum approach across different soil types can vary (Williamson et al., 2013). The VLP count, and therefore the extractable viral DNA and community composition, of a specific soil is a function of the soil type and the extraction efficiency of the chosen method and this needs to be kept in mind when analysing data from different soil types.

Following agitation, the VLP containing solution is separated from suspended solid particulates and microbial cells by low speed centrifugation, and filtration of the supernatant.

This filtration step also has the effect of removing any remaining free micro-organisms larger than the filter pore size, with the chosen pore size typically being 0.2 or 0.45 μm. The chosen pore size is a compromise, as the 0.45 μm filter will allow more small microorganisms through, but has the potential to allow more VLPs to be recovered. In mock communities of saliva viromes, a 0.2 μm filter has been shown to bias virome composition against larger viruses, but this study did not assess the impact on the reduction of microbial contamination and both pore sizes have been used for soil viromics studies (Parras-Moltó et al., 2018).

Once passed through the 0.2 or 0.45 μm pore size filter, the VLP solution must then be concentrated. This is achieved either through chemical flocculation, or using <100 kDa pore size filtration. High molecular weight polyethylene glycol (PEG) is often used for concentrating VLPs in soil viromics studies and $FeCl_3$ has been used in aquatic viromics (John et al., 2011). Other possibilities include skimmed milk powder flocculation and various charged membrane adsorption-desorption techniques. Hjelmsø *et al.* (2017) compared PEG, skimmed milk, monolithic adsorption filtration and glass wool filtration for extracting viruses from raw sewage. Their findings showed that sewage viromes from PEG precipitated concentrates gave a significantly higher percentage of viral reads when compared to glass wool filtration and skimmed milk precipitation but no difference in Chao1 species richness. Small pore size filtration uses pore sizes up to 100 kDa in either centrifugal or tangential flow filtration (TFF) setups. Several membrane materials are available and when compared, PES was found to allow more viral particles to permeate than polyvinylidene fluoride (PVDF) while regenerated cellulose (RCL) demonstrated higher recovery than PES with aquatic samples (Cai et al., 2015). Viral recovery from centrifugal concentrators can be increased by pre-treating them with a blocking agent such as bovine serum albumin (BSA) and by post-concentration sonication of the membranes (Deng et al., 2012), although the efficacy of BSA treatment can vary by sample type (Trubl et al., 2016).

The concentrated VLP solution can be further purified by density gradient ultracentrifugation using CsCl or iodixanol gradients, or sucrose cushions. These have the advantage of more effective removal of cellular debris and other contaminants but CsCl gradients in particular can restrict the diversity of detected viruses as this procedure can compromise

viral capsid integrity (Kleiner et al., 2015).

Finally, exogenous DNA must be removed by DNase digestion as the VLP concentrate will also contain unprotected free environmental DNA. This treatment has been shown to significantly improve the number of viral contigs from agricultural soil viromes, without altering the ecological patterns when compared to untreated viromes (Sorensen et al., 2021). The duration and enzyme concentration used should first be optimised to ensure complete digestion of unprotected DNA as the conditions required can vary by sample type (Thurber et al., 2009). Whilst these reactions are typically performed at 37°C for 1 hour, some studies have used overnight at 4°C, as soil phages will rarely be exposed to and may not be adapted to this high a temperature in their natural environment (Emerson et al., 2018).

1.3.2. Sequencing library preparation, sequencing and data processing

Amplification of purified DNA can be required for viromics as the amount of DNA recovered from environmental samples is often too low for PCR free library preparation. The amplification method chosen can introduce substantial bias into the results that the investigator must be aware of. Linker Amplified Shotgun Libraries (LASL) were used in earlier metagenomics studies and can require as little as 1 pg of starting material (Duhaime et al., 2012), but this method has been shown to be biased towards amplification of dsDNA and so is unreliable at detection of ssDNA viruses. Multiple Displacement Amplification (MDA) can be used as an alternative but has been shown to be substantially biased towards small circular DNA fragments (Kim and Bae, 2011). This can be advantageous if the investigator wishes to target ssDNA viruses in particular as the effect is increased if the initial denaturing step is eliminated.

The bias between ssDNA and dsDNA amplification can be reduced by diluting DNA prior to MDA (Brinkman et al., 2018), however this would need to be optimised before deep sequencing. Random Priming Sequence Independent Single Primer Amplification (RPSISPA) has been used specifically in soil viromics (Adriaenssens et al., 2017) but shows bias to more common fragments (Karlsson et al., 2013). Alternatively, ssDNA and ds-

DNA can be separated for independent library preparation, either through hydroxyapatite-based separation (Andrews-Pfannkoch et al., 2010) or by selective purification of ssDNA and digestion of ssDNA in dsDNA samples (Yoshida et al., 2018). Adaptase-Linker Amplification (A-LA) has been proposed as a solution to library construction bias, as it has been shown to be broadly unbiased towards single and double stranded DNA by denaturing dsDNA prior to the adapter ligation step (Roux et al., 2016b).

Whilst the *de facto* standard sequencing technology is provided by Illumina, long-read data is increasingly being used to generate scaffolds onto which short reads can be assembled in order to generate more contiguous assemblies. This has the advantage of being able to improve assembly by spanning regions of low complexity and detect viral populations with greater microdiversity than can be achieved using short read data alone (Warwick-Dugdale et al., 2019; Zablocki et al., 2021).

1.3.3. Post sequencing quality control and assembly of viromes

Processing of sequencing data can be divided into four steps:

1. Trimming and filtering of raw reads
2. Deduplication of replicates
3. Error correction
4. Assembly

Firstly, raw sequencing reads are trimmed to adapter sequences and poor quality bases found towards the end of reads (Jo et al., 2020). This is typically performed using known sequences of adapters and a sliding quality window using quality scores provided for each base within an fastq file. Additionally, reads that are below a certain length threshold, have a low mean quality score, or contain unresolved bases, i.e. Ns, can also be removed. This process will inevitably generate orphan reads from paired read data where the forward or reverse read is rejected, but the paired read is retained.

The library preparation process will generate duplicated reads during PCR amplification and so identical reads should be removed as this can artificially inflate abundances (Solonenko et al., 2013). Depending on the study system, it is sometimes advisable to filter

out host reads or those from known contaminants. This can be done by mapping reads to the genomes of known contaminants, or by including a negative control in the sample processing design. In environmental samples, it is possible to introduce bias by removing reads from temperate bacteriophages that may be present in bacterial genomes in reference databases. As sequencing introduces known errors, error correction can also improve the recovery of viral genomes. This is particularly important when using long read data, but is also beneficial when using short read data (Roux et al., 2017).

Once reads have been trimmed and filtered, they are typically assembled into contigs which can represent whole viral genomes. Multiple assembly programs are available and most rely on multi-kmer de Bruijn graph assembly strategies. Several studies have compared assembly algorithms and regularly place MEGAHIT and SPAdes as the best performing, with SPAdes requiring significantly more computational resources (Roux et al., 2017; Sutton et al., 2019).

Once assembled, contigs must be identified as viral or non-viral. Many tools are available for this and are typically based on either machine learning or sequence similarity (Ho et al., 2021). No one strategy has emerged as dominant and so it is not uncommon to combine the results from multiple tools and apply additional checks using tools such as CheckV to further improve accuracy and remove contamination from host sequences (Nayfach et al., 2020). As single contigs often represent whole viral genomes, contig binning is often not performed on viromics data but may be necessary for large DNA viruses (Schulz et al., 2022). Segmented and multipartite viruses present a particular challenge due to the often small size of their genome segments and the lack of a single contig representing the entire viral genome. This may be overcome by identifying conserved regulatory sequences and comparing the phylogeny of different segments, but this overlooks the detection of viral reassortment (Varsani et al., 2018).

Once viral contigs are identified, they can be analysed in similar ways to microbial metagenomics data using tools that assign taxonomy, e.g. vConTACT v.2.0 (Bin Jang et al., 2019), annotate functional genes, e.g. DRAMv (Shaffer et al., 2020) or analyse ecological patterns (Gregory et al., 2022).

1.4. Quantitative PCR-based detection and wastewater based epidemiology

1.4.1. Polymerase Chain Reaction based quantification of viruses in wastewater and biosolids

Whilst viromics and bulk metagenomics provide a top-down overview of a viral community, the relatively low sample throughput and high per-sample financial costs preclude the use of these techniques for regular high temporal resolution monitoring of specific human pathogenic viruses and other faecal indicator species (Farkas et al., 2020b). The polymerase chain reaction (PCR) synthesises multiple copies of a DNA sequence specified by a pair of short primer sequences chosen by the user and by combining this with the creation of a fluorescent signal that is dependent on the copy number of the PCR product, it is possible to quantify the number of copies of this sequence in the original sample. A reverse transcription (RT) step to generate cDNA from RNA templates can be added before PCR to also detect copy numbers of RNA viruses. This can either be carried out as a separate reaction, allowing the possibility of finer optimisation (two step RT-qPCR) or immediately prior to qPCR in the same reaction tube (one step RT-qPCR), which can increase sensitivity by removing the need for sample dilution. The abbreviation RT has also been used to refer to real-time as a synonym of quantitative PCR.

For qPCR, two dye-based methods are applied; intercalating dye and probe-based methods (see Fig. 1.6). Intercalating dyes such as SYBR Green are non-specific and will ultimately detect the formation of any dsDNA product and fail to distinguish between targeted amplification, the formation of primer dimers and non-specific amplification, sometimes necessitating a subsequent melt-curve analysis to confirm specificity. Low level fluorescence can also be caused by SYBR Green binding non-specifically to the walls of multi-well plates, necessitating a melt-curve analysis to verify specific amplification (Zipper, 2004). The advantage here is that once optimised, intercalation based qPCR assays are easier and cheaper to use than probe-based methods. Four forms of probe-based qPCR assay are currently available: Taqman, LightCycler, LUX and Molecular Beacons (reviewed by VanGuilder *et al.*, 2008). Probe-based methods are often more expensive due to the need for both primers and a fluorescently labelled probe oligonucleotide, but they also show enhanced specificity, as the fluorescent signal is directly linked to the in-

corporation of the probe into amplified DNA molecules during each amplification cycle. Probe-based assays can also be multiplexed for the detection of multiple targets in a single reaction by combining sets of primers/ probes labelled with different fluorophores however extensive optimisation and compatible annealing temperatures of different assays are required (Hawkins and Guest, 2017).

(RT)qPCR has been adopted for the monitoring of multiple human pathogenic viruses in wastewater, soil and wastewater-contaminated environments (Hewitt et al., 2011). The considerations for obtaining nucleic acids suitable for analysis by qPCR and (RT)qPCR are similar to those for viromics, however additional control steps are more commonly used, such as positive and extraction controls, qPCR no template controls and increased biological and technical replication (Ahmed et al., 2020c; Borchardt et al., 2021; Bustin et al., 2009).

Provided the target gene sequence is conserved, qPCR and (RT)qPCR can be used rapidly enumerate known viruses in an environmental sample. Of all the techniques detailed here, qPCR shows the highest sensitivity (detection limits are often close to 1 target molecule per reaction volume) and selectivity. However, as it is based on amplification of nucleic acids, care must be taken to remove PCR inhibitors from samples prior to analysis. Furthermore, the technique cannot distinguish between infective and non-infective virus particles. Coupling qPCR with DNase/ RNase treatments and affinity based pull-down assays can select for intact viral particles capable of binding host proteins required for infection and act as a surrogate for infectivity models (Leifels et al., 2021).

1.4.2. SARS-CoV-2, COVID-19 and the emergence of national-scale WBE

The emergence of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and resulting Coronavirus disease 2019 (COVID-19) saw an explosion of the utilisation of (RT)qPCR based monitoring of viral prevalence and a significant expansion in the use of wastewater based epidemiology (WBE) to track the progression of this global pandemic. SARS-CoV-2 is readily detectible in the faeces and urine of 40% and 2% of cases respectively, however the likelihood of transmission via the faecal-oral route or from wastewater-

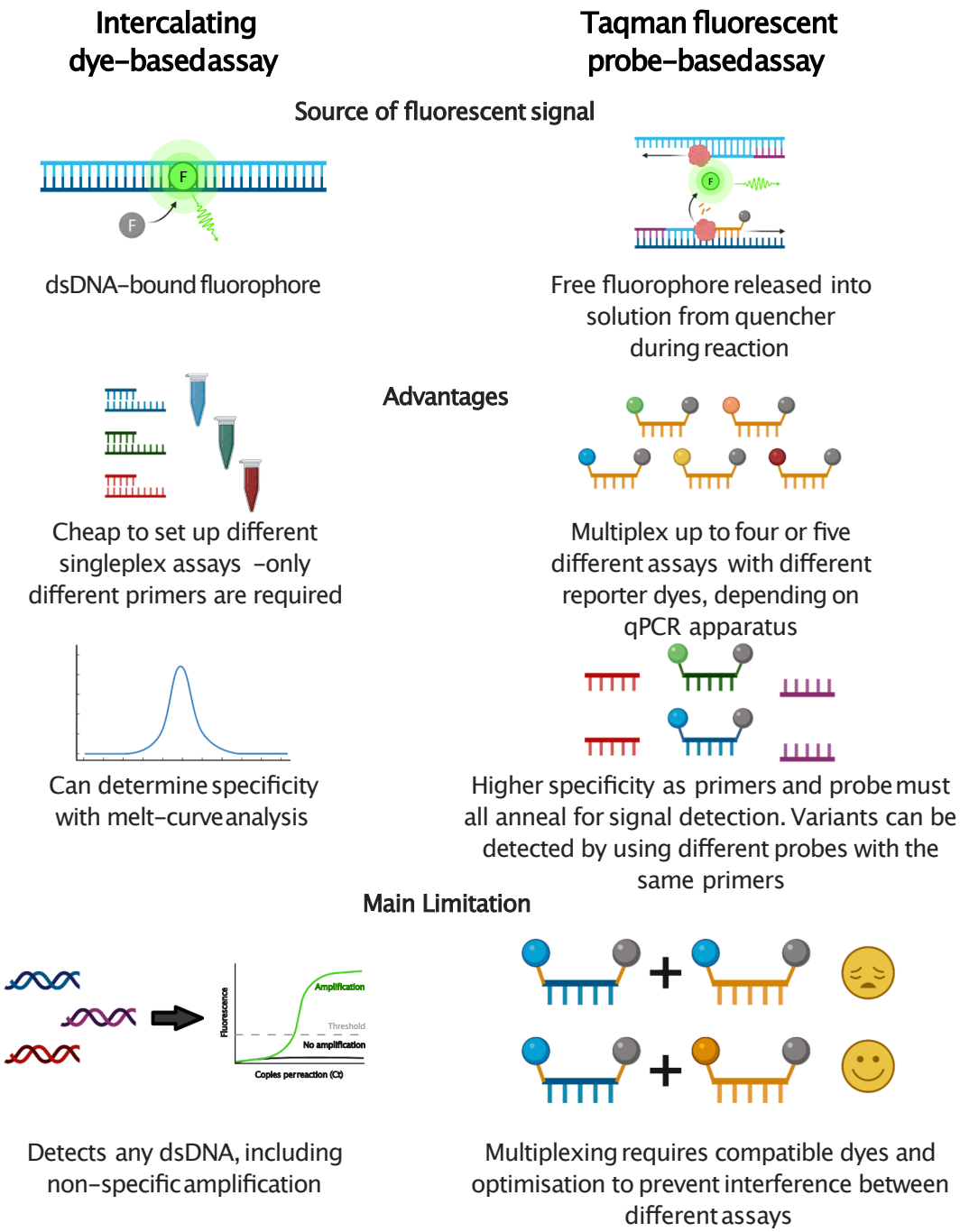


Figure 1.6.: Comparison of intercalating dye Sybr-based and fluorescent probe Taqman-based qPCR assays. Probe based qPCR and RT-qPCR assays are more specific than Sybr dye based assays due to the addition of an additional fluorescently labelled oligonucleotide that is incorporated into amplicons during the PCR reaction, triggering the release of the fluorophore, which forms the basis of quantification. Several assays can be combined in a multiplexed reaction but this requires careful optimisation.

contamination is extremely low (Jones et al., 2020)[2]. The amount of SARS-CoV-2 genome copies present in wastewater can then be quantified and give an indicator of the amount of virus circulating within the wastewater treatment plant catchment area (Farkas et al., 2020a)[3]. This information can then be used to inform public health policy which will in turn reduce the amount of virus circulating in the population (see Fig. 1.7).

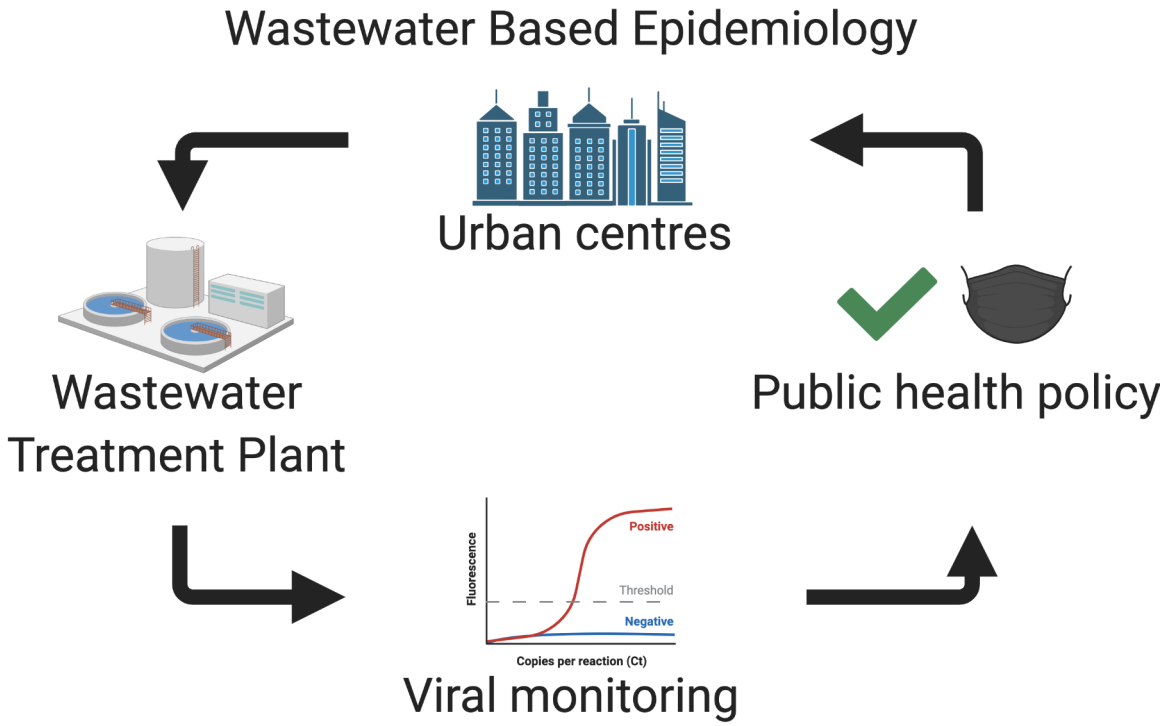


Figure 1.7.: Diagram of the application of wastewater based epidemiology. Municipal wastewater containing viruses of interest shed by the local population flows into a wastewater treatment plant where they can be sampled and detected. This data informs public health policy which impacts the prevalence of the viral infection within the local population. This change can then be observed in the quantity of virus detected in the wastewater.

Since the beginning of the COVID-19 pandemic, national scale WBE programs have been established in several countries around the world and this global research effort has rapidly developed our understanding of the factors affecting the quantification of viruses in wastewater and biosolids (Pecson et al., 2021). In addition to using WBE to monitor the quantity of SARS-CoV-2 in local populations, tiled amplicon sequencing of SARS-

[2]Jones, D.L., Baluja, M.Q., Graham, D.W. Corbishley, A., McDonald, J.E. Malham, S.K., **Hillary, L.S.**,Connor, T.R., Gaze, W.H., Moura, I.B., Wilcox, M.H., & Farkas, K. Shedding of SARS-CoV-2 in feces and urine and its potential role in person-to-person transmission and the environment-based spread of COVID-19. Sci. Tot. Env. (2020)

[3]Farkas, K., **Hillary, L.S.**, Malham, S.K., McDonald, J.E. & Jones, D.L., Wastewater and public health: the potential of wastewater surveillance for monitoring COVID-19. Curr. Opin. Environ. Sci. Heal. (2020)

CoV-2 genomes found within wastewater has been used to detect different variants that may impact the progression of the pandemic (Lin et al., 2022).

1.5. Monitoring the removal of viruses from the wastewater treatment process

The faecal-oral route is a common mode of transmission for human enteric viral infections, and therefore, domestic sewage and the by-products of its treatment represents a potential environmental reservoir of human pathogenic viruses (Fong and Lipp, 2005). In addition to using WBE for the monitoring of viruses circulating in local populations, it is necessary to monitor the efficacy of wastewater treatment processes in removing or inactivating human pathogenic viruses prior to releasing treated effluent into the environment. Wastewater treatment involves three or four stages (Sonune and Ghate, 2004):

1. Preliminary treatment - removal of grit and other large material.
2. Primary treatment - settling of organic and inorganic solids.
3. Secondary treatment - aerobic biological treatment to further remove organic material.
4. Tertiary treatment (not consistently employed, depending on operational need) - chemical or physical treatments designed to further remove or inactivate microbial or viral contaminants prior to discharging treated effluent into the environment.

In addition to the above treatment of liquid wastewater, the settled solids, or sludge, are further treated via a combination of three processes (Kelessidis and Stasinakis, 2012):

- Stabilisation - aerobic or anaerobic digestion reduces pathogen and biomass content and can be used to generate energy through the production of biogas.
- Conditioning (pre or post stabilisation) - further reduction of pathogen content or inorganic compounds, alteration of physical and chemical properties through addition of lime, polymers, and thermal hydrolysis treatment, which can also enhance biogas production from post-conditioning anaerobic digestion.
- Dewatering - removal of water via industrial centrifuges, belt presses or other means prior to disposal.

Finally, the treated sludge, often referred to as biosolids, is disposed of via incineration, landfill, or use as agricultural fertiliser. The balance between these different methods is heavily dependent on national regulatory frameworks, historical practices and available infrastructure (Kelessidis and Stasinakis, 2012). Within the UK, 70% of treated sludge is reused as agricultural fertiliser and its management is governed by a variety of regulations and voluntary codes drawn together to form the Biosolids Assurance Scheme (BAS) (Assured Biosolids Limited, 2020). This standard stipulates how sludges should be treated, transported, stored and applied to agricultural land. There are three identified standards of sludge: untreated, conventionally treated and enhanced treated. Although both are voluntary, the BAS and preceding ADAS Safe Sludge Matrix (Agricultural Development and Advisory Service, 2001) have effectively banned the use of untreated sludge since 2005. The differences between conventional and enhanced treatments are summarised in Table 1.2 and Table 1.3.

Table 1.2.: Comparison of micro-organism reduction requirements for conventional and enhanced sludge treatments within the UK (Assured Biosolids Limited, 2020).

Micro-organism Reduction		
Parameter	**Conventional Treatment**	**Enhanced Treatment**
E. coli destruction	2 $\log_{10}$	6 $\log_{10}$
E. coli Maximum Allowable Concentration	100,000/ gram of dry solids	1,000/ gram of dry solids
Salmonella spp.	N/A	Free on completion of treatment process

Table 1.3.: Summary of sludge use regulations in the UK (Agricultural Development and Advisory Service, 2001).

Application Regulations		
Crop	**Conventional Treatment**	**Enhanced Treatment**
Fruit	Not permitted	10 month harvest interval

Application Regulations		
Salads	Not permitted (30 month harvest interval after prior application)	10 month harvest interval
Vegetables	Not permitted (30 month harvest interval after prior application)	10 month harvest interval
Horticulture	Not permitted	10 month harvest interval
Grazed animal feed	Deep injection/ ploughed down only (3 week grazing/ harvest interval, extended to 2 months grazing interval for pigs)	3 week grazing/ harvest interval, extended to 2 months grazing interval for pigs
Harvested animal feed	No grazing in season of application (3 week grazing/ harvest interval, extended to 2 months grazing interval for pigs)	3 week grazing/ harvest interval, extended to 2 months grazing interval for pigs
Sludge cannot be applied between March and August ahead of fodder crops to be used before winter frosts		

Addition of biosolids to soil can potentially affect the intrinsic soil virus community by changing soil structure, chemical composition and host availability (see Fig. 1.4). It can also introduce biosolids-associated viruses that can compete for hosts or remain inert, decaying over time. As human faeces is a major constituent of biosolids, there is potential for human pathogenic viruses to be incorporated into amended soils, before coming into contact with potential hosts. Key viruses that can be transmitted via the faecal-oral route, and therefore may be present in biosolids include (Rodríguez-Lázaro et al., 2012):

- *Adenoviridae* (specifically strains 3, 7, 40 and 41)

- *Caliciviridae* (e.g. norovirus)
- *Picornaviridae* (coxsackievirus, echoviruses, enteroviruses and hepatitis A virus)
- *Reoviridae* (reoviruses and rotaviruses)
- *Hepeviridae* (Hepatitis E virus)

Adenoviridae are non-enveloped dsDNA viruses. The family is divided into five genera with human adenovirus species limited to the genus *Mastadenovirus* (Lefkowitz et al., 2018). Within this genus, the seven human adenovirus species can cause respiratory and urinary infections, conjunctivitis and gastroenteritis (Ghebremedhin, 2014).

Caliciviridae are small non-enveloped(+)ssRNA viruses. The family contains five genera of which norovirus and sapovirus are capable of causing gastroenteritis (Lefkowitz et al., 2018). Feline calicivirus of the genus *Vesivirus* is sometimes used as a surrogate virus for human norovirus. Whilst infection models of norovirus are limited, murine norovirus is commonly used as a surrogate virus which can be grown in culture.

Picornaviridae are a large family of +ssRNA viruses capable of causing a range of diseases including gastroenteritis (enteroviruses - Melnick, 1984) and hepatitis (*Hepatovirus* - Martin and Lemon, 2002).

Reoviridae are dsRNA viruses and are divided into two subfamilies: *Sedorevirinae* and *Spinareovirinae*. The genus *Rotavirus* is a common cause of gastroenteritis in infants (Bernstein, 2009).

Orthohepevirus and Piscihepevirus form the *Hepeviridae* family of (+)ssRNA viruses (Lefkowitz et al., 2018). Hepatitis E virus (HEV) infections can be both acute and chronic and affect multiple organs. It can be transmitted via the faecal-oral route and through zoonotic transfer from pigs.

Whilst the above viral families are the main causes of gastroenteritis, many other human pathogens can be found in sewage sludge and the wider environment. Viromics studies of raw and treated sewage sludge can detect various non-enteric human pathogenic viruses in

mesophilic anaerobic digester effluent (Bibby and Peccia, 2013). Non-pathogenic human-gut associated viruses such as crAssphage and pepper mild mottled virus (PMMoV) can be utilised as indicator viruses and surrogates for human pathogens as they are often present at higher concentrations and so are more easily detectable in environmental samples (Farkas et al., 2020c)[4].

Human pathogenic viruses are known to persist through the sludge treatment process and retain their infectivity (Wong et al., 2010). Once these biosolids are applied to land, viral persistence in terrestrial environments can be influenced by various factors and the possibility remains that they can return to the food and water production systems via a number of routes. For human pathogenic viruses to pose a threat to public health via the use of biosolids, they must retain infectivity at clinically relevant concentrations during these transport and transmission processes. In contrast to soil-borne bacteriophages, there is no mechanism for human pathogenic viral reproduction in these environments and so the persistence of virion presence and infectivity is of key importance.

1.6. Current challenges and aims of the thesis

The virus removal efficacy of a WWTP is highly dependent on the source of the influent wastewater, the treatment process, and the virus being examined. The persistence of contaminating viruses in soils post-amendment is also dependent on a variety of factors, e.g. soil type, pH, temperature, moisture content. The effects of these factors are well characterised for aquatic environments, such as rivers and groundwater, and for some specific individual known pathogens in soils and biosolids. However, our knowledge of the persistence of biosolids-associated viruses at a community scale is extremely limited.

Whilst the behaviour of individual pathogens and their surrogates in wastewater, biosolids and soil is well characterised, the quantity and diversity of biosolids-associated viruses being imported into amended soils, and their persistence over different timescales is largely unknown. The effects on the intrinsic soil viral community of biosolids amendment at a community scale is also unexplored. Current tools for investigating these questions are

[4]Farkas, K., Walker, D.I., Adriaenssens, E.M., McDonald, J.E., **Hillary, L.S.**, Malham, S.K. & Jones, D.L. Viral indicators for tracking domestic wastewater contamination in the aquatic environment. Water Res. (2020) doi:10.1016/j.watres.2020.115926.

limited to DNA viruses, and often focus on dsDNA bacteriophages. Whilst the regulatory restrictions on the use of biosolids on agricultural land provides a buffer time for human pathogens to degrade, the possibility that these viral particles could leach into watercourses, be taken up by crop plants, or persist beyond this period remains. The effects of environmental factors are highly virus-specific and our ability to investigate them is limited by the availability of molecular and culture-based techniques for viruses of interest.

The overarching aim of this thesis is to assess the effects of biosolids amendment on soil-borne virus communities and further develop techniques to investigate their diversity. This produces several key research questions (also see Fig. 1.8):

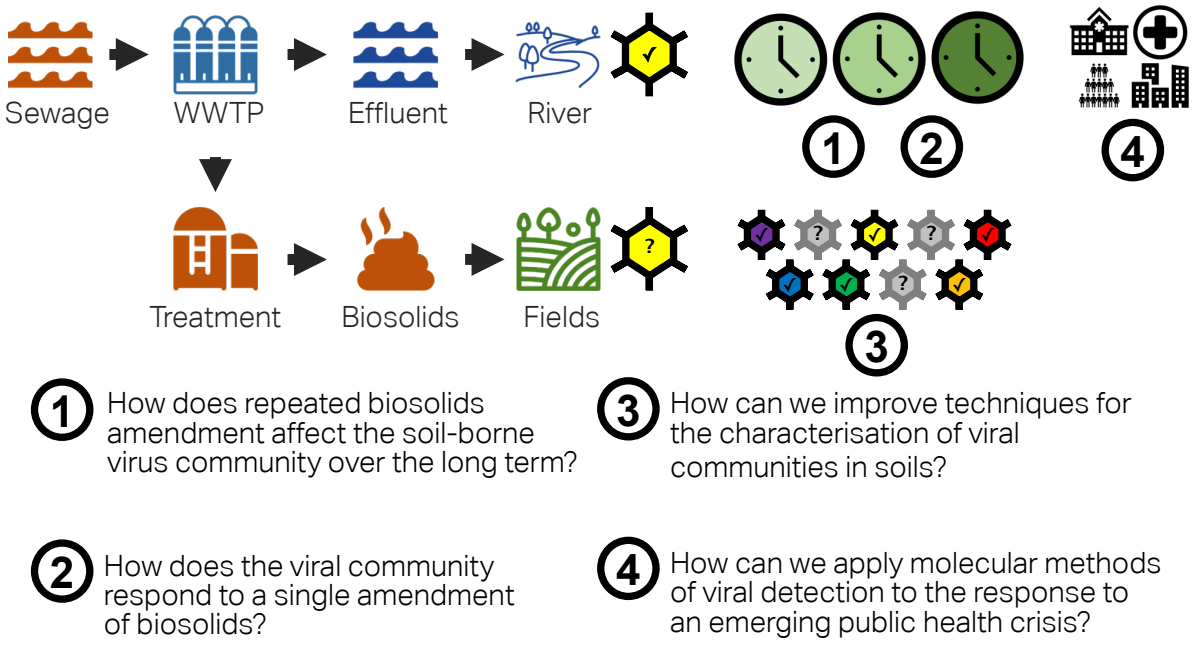


Figure 1.8.: Diagramatic respresentation of the research questions addressed in this thesis. The flow of viruses through the wastewater treatment process has been well characterised, but the flow to land via biosolids is poorly understood. This work examines how biosolids amendment affects the soil virus community over different timescales (research questions 1 and 2), how we can improve our understanding of terrestrial viral ecology (research question 3) and how this knowledge can inform policy decisions (research question 4).

1) How does repeated biosolids amendment affect the soil-borne virus community over the long term? Our current understanding of terrestrial viral ecology is limited, and no studies have examined the impact of repeated or historical biosolids amendment on the soil virus community using long-term field trials. Chapter 2 addresses

this research question by applying viromics to the characterisation of the DNA virus communities in soils from control, historically (1994-1997) and long-term (1994-2018) biosolids amended soils and compares these to the virome of biosolids used in their amendment. In order to expand the diversity of viruses detected and partially address research question 3 below, the A-LA library preparation technique was used to amplify both viral dsDNA and ssDNA with reduced bias.

2) How does the viral community respond to a single amendment of biosolids? Whilst Chapter 2 addresses the long-term effects, the question of the impact of biosolids amendment on the soil virus community over the short-term has also not been previously addressed. Chapter 3 addresses this by examining shifts in the DNA soil virus communities in conventional biosolids amended and unamended soil microcosms at the time of amendment and after one year. Controlled conditions were used to establish a baseline wost-case-scenario in terms of viral persistence and to reduce the impact of site-specific variables such as seasonal climatic variation or choice of crop.

3) How can we improve techniques for the characterisation of viral communities in soils? DNA terrestrial viral ecology has expanded rapidly, however the many human, animal and plant pathogens of interest are RNA viruses, and at the start of this research, no studies of soil RNA viral ecology had been published. Chapter 4 addresses this by applying viromics to the characterisation of soil RNA viral communities along an altitudinal productivity gradient of peat, managed grassland and coastal soils. This chapter acts as a proof of concept of terrestrial RNA viromics and demonstrates how it could potentially be applied to characterising the impact of land management on viruses beyond dsDNA bacteriophages and how they can influence inter-kingdom interactions across terrestrial biomes.

4) How can we apply molecular methods of viral detection to the response to an emerging public health crisis? The COVID-19 pandemic has had a profound effect on global society, creating significant pressures on national healthcare systems and unprecedented international scientific efforts to understand and curb the spread of SARS-CoV-2. Prior to the declaration of a global pandemic by the WHO, it was evident that clinical testing would be an limitation on estimates of viral spread and an urgent need

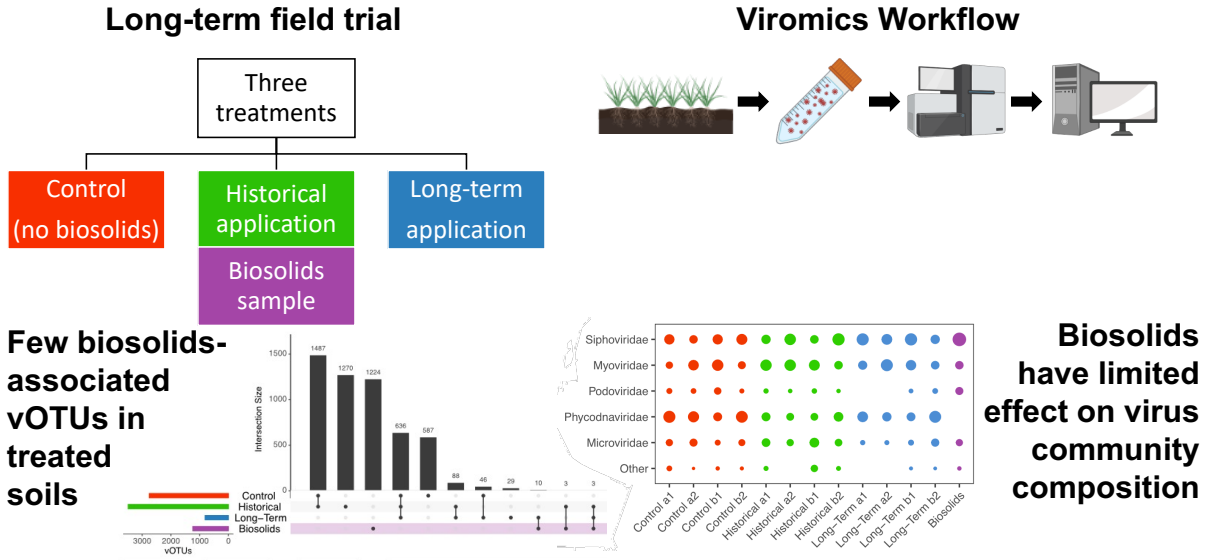
emerged for scalable alternatives. Chapter 5 applies both qPCR and targeted genome sequencing to the monitoring of SARS-CoV-2 prevalence during the early stages of the COVID-19 pandemic. Data from this study informed public health decisions by both the Welsh and UK governments and aided the establishment of both the Welsh and English national SARS-CoV-2 WBE monitoring programs.

Each subsequent chapter addresses these research questions in turn before the overall findings of the work are summarised and synthesised and placed in the context of the wider literature (Chapter 6). Their strengths and weaknesses are discussed, followed by their policy implications and potential directions for future work.

2

Long-term effects of biosolid amendment on the soil viral community

2.1. Graphical abstract



2.2. Abstract

Biosolids are commonly used as agricultural fertiliser across the world and could therefore be a source of viral pathogen contamination in the food production chain. Little is known about the long-term persistence of biosolids-associated viruses in amended soil or their impact on the naturally occurring soil viral community. We used viromics to characterise the DNA virus communities in soils from control, historically (1994-1997) and long-term (1994-2018) biosolids amended soils. We found that biosolids-associated viruses formed 0.9% ±0.44 (mean ±s.d.) of the soil viral community in plots receiving long-term repeated biosolids amendment, compared to 0.009% ±0.009 in historically treated soils, with no biosolids-associated viruses detected in the control soils. All soil viral communities were predominantly composed of *Siphoviridae*, *Myoviridae*, *Microviridae*, *Podoviridae* and *Phycodnaviridae*. Although soil viral communities differed between treatments, no statistically significant differences were observable in overall α-diversity. 1048 auxiliary metabolic genes were identified with a wide range potential functional capacity shown to exist in each viral community, whilst antimicrobial resistance genes (ARGs) were extremely rare, with only four identified across the whole study. No biosolids-associated ARGs were identified in any soil virome. These results suggest that biosolids amendment has the potential to import biosolids-associated viruses into soil, but they do not establish themselves as substantial long-term constituents of the soil viral community. This implies, that unlike chemical contaminants such as heavy metals and microplastics, the possible threat from biosolids-associated viruses to crop production and human health is potentially short-lived.

2.3. Keywords

biosolids, virome, soil, phycodnaviridae, AMGs

2.4. Introduction

Biosolids, the solid byproducts of the wastewater treatment process, are often utilised as a source of macro- and micro-nutrients and as a soil-quality improver in agriculture (Sharma et al., 2017), but can contain contaminants potentially hazardous to human health, such as heavy metals, microplastics, antimicrobial resistance genes, pathogenic microorganisms and viruses (Chen et al., 2021; Crossman et al., 2020; Sloan et al., 1997; Xie et al., 2016). The annual risk of viral infection to humans from correctly applied anaerobically digested biosolids has been estimated at < 1:10,000 (Gerba et al., 2002) however this may be higher for more stable viral pathogens (Viau et al., 2011). For example, human adenovirus (HAdV) is regularly detected in treated biosolids (Viau and Peccia, 2009) and can persist for >180 days in biosolids amended soils (Schwarz et al., 2014). Whilst the decay and dissemination of individual pathogenic viruses and their surrogates in biosolids and soil have been well studied (Roberts et al., 2016; Schwarz et al., 2014), the long-term effects of biosolids amendment on the soil viral community is poorly characterised.

The soil viral community can be impacted directly by biosolids amendment through the import of biosolids-associated viruses, and indirectly by altering soil physical and chemical properties and microbial host availability. Long-term use of biosolids as an agricultural fertiliser can increase soil organic carbon, potential mineralisable nitrogen and microbial biomass (Ippolito et al., 2021). Biosolids amendment has also been shown to reduce arbuscular mycorrhizal fungi and bacterial abundance and cause significant shifts in soil microbial community structure (Hu et al., 2019; Schlatter et al., 2019), potentially altering the availability of microbial hosts for intrinsic soil viruses. Biosolids amendment can have residual effects on soil properties years after the original treatment, although the effect on microbial community can be difficult to distinguish (Reardon and Wuest, 2016).

The intrinsic soil viral community have been shown to have the potential to influence carbon, nitrogen, sulphur and phosphorus metabolism (Emerson et al., 2018; Monier et al., 2017; Roux et al., 2016a; Zeng and Chisholm, 2012). Viromics and other sequencing based methods can be used for monitoring the presence and genetic diversity of viruses in soils, biosolids and environmental waters impacted by the discharge of treated wastewater (Adriaenssens et al., 2021; Bibby and Peccia, 2013; Santos-Medellin et al., 2021). Biosolids

amended soils were shown to have increased viral antimicrobial resistance genes (ARGs) (Chen et al., 2021), however significant caution is needed in the analysis of such data to avoid false positives or negatives (Enault et al., 2016). As with any form of gene annotation, BLAST search parameters are a compromise between precision and sensitivity: Enault et al. (2016) found that exploratory search parameters could yield high numbers of false positives that showed no *in vitro* antimicrobial activity, whilst stricter settings could exclude known ARGs. In addition to the direct impact of importing biosolids-associated viruses into soil, the long-term and historical impacts of biosolids amendment on naturally occuring soil viral community remains poorly understood.

In this study, we used viromics to assess the historical and cumulative effects of biosolids amendment on DNA virus communities of agricultural soil under historical and long-term biosolids amendment. We aimed to address (1) how biosolids amendment impacts the soil viral community, (2) how long-term amendment compares to historical treatment, and (3) whether biosolids-associated viruses can persist in soils over extended timespans.

2.5. Materials and methods

2.5.1. Site description, sample collection and processing

Soil samples were collected from a long-term field trial experiment in Woburn, Bedfordshire, UK, examining the effects of biosolids on soil health (Gibbs et al., 2006). Soil properties and the biosolids application parameters are described in Supplementary Table A.1 and A.2. The experimental design consisted of three randomised blocks of 6 x 8 m plots receiving various treatments, of which three were sampled (see Fig. 2.1):

- Control - no biosolids amendment
- Historical biosolids amendment – high application rate (designed to raise heavy metal levels to the UK limit after 4 years of application, 1994-1998)
- Long-term biosolids amendment - moderate application rate (approximately 1/25th of the total mass of sludge applied annually, 1995-2018)

Biosolids were incorporated to a depth of 25 cm using a Celli spading machine and wheat grown annually for the duration of the experiment. N, P, K fertiliser was added according

to requirements assessed by soil analyses over the duration of the experiment (Gibbs et al., 2006).

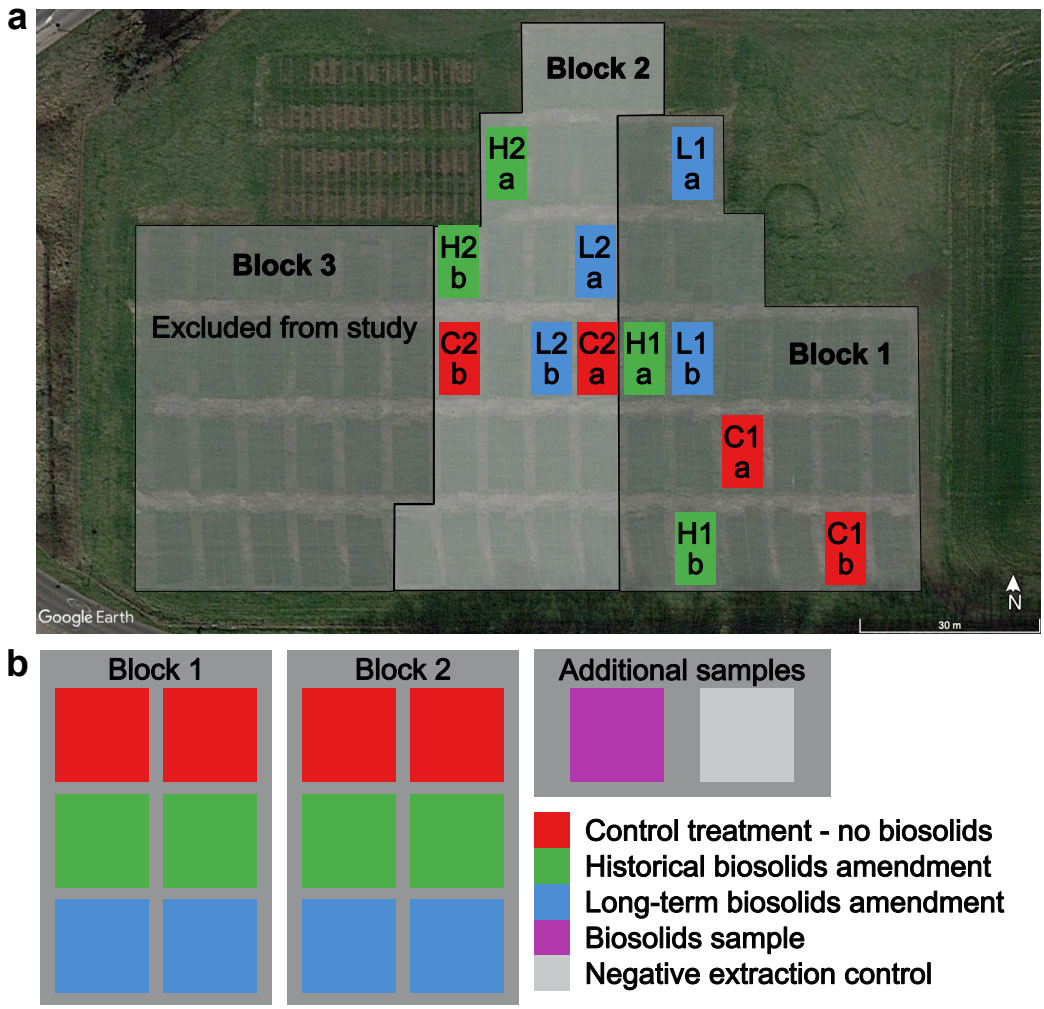


Figure 2.1.: Three treatments were sampled from a long-term field trial investigating the impact of heavy metals on soil health. The experiment was divided into three blocks (a). Block 3 was excluded from this study as it was under a different crop. Plot codes represent treatment - C/ H/ L, block - 1/2 and independent replicate - a/ b. A total of 14 samples were analysed in this study (b): 12 soil samples, one sample of biosolids used to amend the long-term plots, and one negative extraction control.

Each treatment was sampled on 22nd November 2017 from duplicate plots within two blocks (a third block was removed from this study due to a change in cropping). Six evenly distributed cores (20 cm depth) were extracted using a 2.5 cm diameter half moon auger and stored separately on ice before transport back to Bangor University. Between the sampling of each plot, the auger was sequentially wiped clean with 1% Virkon disinfectant, deionized water and 70% industrial methylated spirit. A dummy core was taken from the middle of each plot and discarded before collecting samples for analysis. Cores from each

plot were sieved to 2 mm and bulked by combining equal masses followed by manual homogenisation. Homogenised soil from each plot was then frozen in 10 x 10 g aliquots at -80°C until viral DNA extraction (see Fig. 2.2 a).

2.5.2. Virus-like-particle DNA extraction and sequencing library preparation

Virus-like-particle DNA was extracted from 12 soil samples, one biosolids sample, and one negative extraction control sample of 30 mL molecular biology grade water. Three 10 g aliquots of material per sample were each thawed and resuspended in 30 mL of potassium citrate buffer (90 mL total volume, per litre: 10 g potassium citrate, 1.44 g $Na_2HPO_4.7H_2O$, 0.24 g KH_2PO_4, pH 7.0) and sonicated (3 cycles of 30 s manual shaking followed by 1 minute sonication on ice with a Soniprep 150 (MSE) and exponential probe at an amplitude of 7 μm amplitude). Following shaking and sonication, the samples were then shaken on ice at 300 rpm for 30 minutes (see Fig. 2.2 b). The soil suspension was centrifuged for 30 minutes at 3,000 x g at 4 °C. Supernatants from the three aliquots per plot were recombined and vacuum filtered through 0.22 μm PES membrane Stericup filters (Millipore) and concentrated using 50 kDa MWCO Amicon Ultra centrifugal ultrafiltration devices (Millipore). Buffer exchange was performed by two rounds of 1:10 dilution with 10 mM Tris-HCl, 10 mM $MgSO_4$, 150 mM NaCl, 0.5 mM $CaCl_2$, pH 7.5) and samples were concentrated to a final volume of <300 μL (see Fig. 2.2 c). Samples were centrifuged at 10,000 x g for 1 minute to pellet aggregated particulates and the supernatants removed for further processing. Exogenous DNA/ RNA was digested by incubation with 125U/ mL Turbo DNase (Thermo Fisher) and 100 μg/ mL RNase I (Promega) for 1 hour at 37 °C followed by inactivation with addition of EDTA (final concentration 20 mM) and heating to 70 °C for 10 minutes.

Viral DNA was extracted and purified with a PowerViral DNA/ RNA kit (Qiagen) according to the manufacturer's instructions with an elution buffer incubation time of 5 minutes. DNA concentrations were quantified using the Qubit DNA HS assay kit and Qubit 2 fluorometer (Invitrogen) and the presence of contaminants checked using a Nanodrop 1000 spectrophotometer (ThermoFisher). Sequencing libraries were prepared using the Accel NGS 1S+ kit (Swift Biosciences) >40 bp fragment protocol to allow the sequencing of

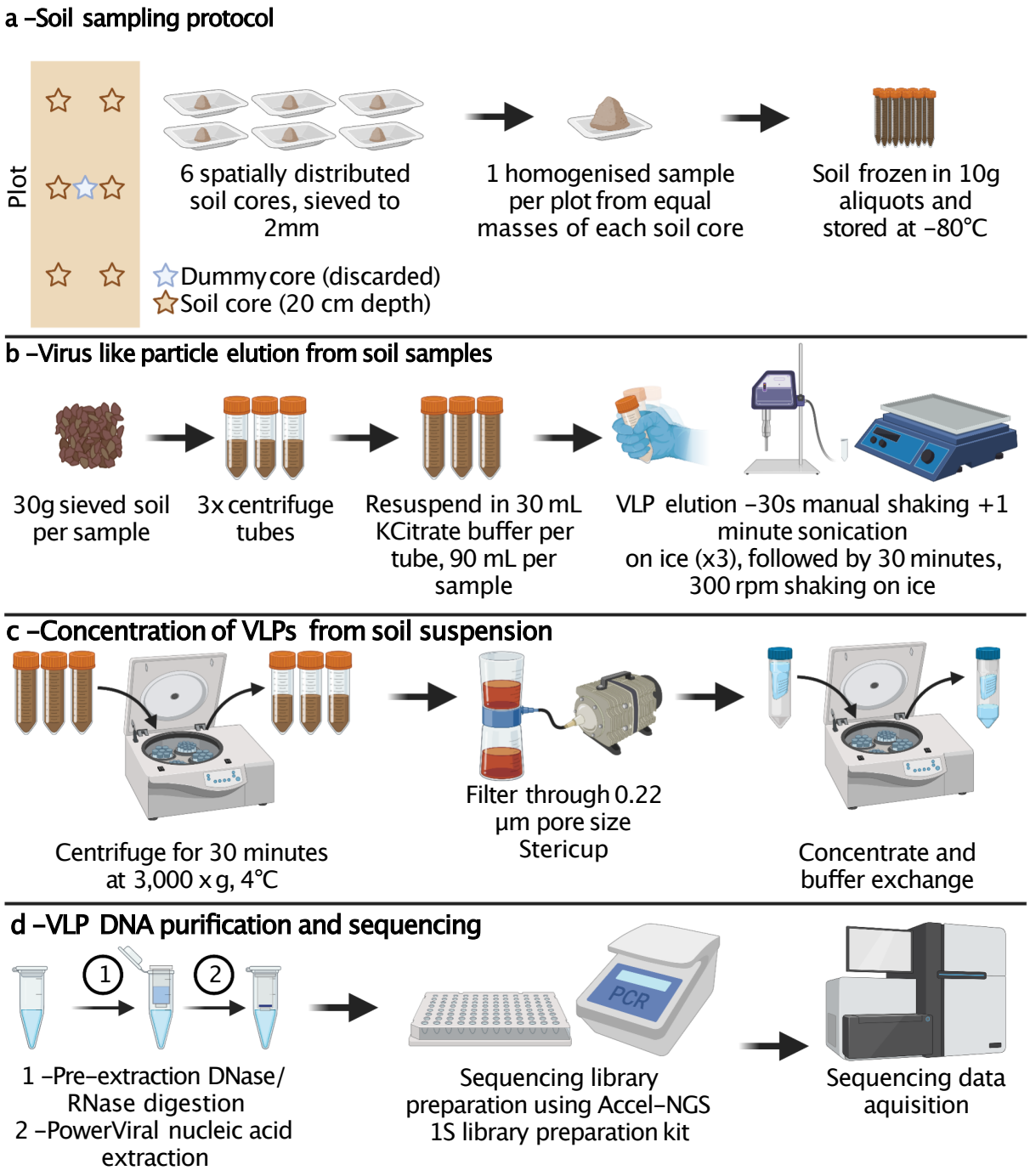


Figure 2.2.: Sample processing and VLP DNA extraction. Six spatially distributed cores were taken per plot, sieved to 2 mm, homogenised in equal quantities (approximately 100g of total soil per plot) and frozen in 10 g aliquots (a). VLPs were eluted from soil particulates (b), concentrated (c) and VLP DNA used to generate a total of 14 sequencing libraries (12 soil samples, 1 biosolids sample and 1 negative extraction control sample).

both ssDNA and dsDNA viruses with reduced bias (Roux et al., 2016b). A total of 14 libraries (12 soil libraries, one biosolids library and one negative extraction control library) were pooled and sequenced (150 bp paired end) on one lane of an Illumina HiSeq 4000 at the University of Liverpool Centre for Genomics Research (see Fig. 2.2 d). Initial demultiplexing and quality control removed Illumina adapters using Cutadapt version 1.2.1 (Martin, 2011) with option -O 3 and Sickle version 1.200 (Joshi and Fass, 2011) with a minimum quality score of 20.

2.5.3. Sequencing data processing and vOTU identification and classification

The raw read QC, assembly and viral contig identification pipeline is summarised in Fig. 2.3. Raw sequencing reads were assessed with FastQC v0.11.8 and reports collated and visualised with MultiQC v1.10.1 (Andrews, 2010; Ewels et al., 2016). Reads were trimmed and filtered using Prinseq-lite v0.20.4 (-min_len 35 -min_gc 5 -max_gc 95 -min_qual_mean 25 -trim_left 10 -trim_right 10) (Schmieder and Edwards, 2011), error corrected with Tadpole (mode=correct ecc=t prefilter=2), and deduplicated with Clumpify (dedupe subs=0 passes=2) from the BBTools package (v38.49: sourceforge.net/projects/bbmap/). All processed reads were then combined and co-assembled with MEGAHIT 1.1.3 (–presents meta-large) (Li et al., 2015). Assembled contigs >1,500 bp in length were processed with three viral sequence identification tools: Deep Virfinder v1.0 (score >= 0.7 AND p value < 0.05), VirSorter 2 v2.1 (Score >= 0.5) and VIBRANT v1.2.0 (>= 4 hallmark genes) (Guo et al., 2021; Kieft et al., 2020; Ren et al., 2020). Contigs passing these thresholds with a length >= 5,000 bp or with a length >= 1,500 bp and identified as circular by either VirSorter 2 or Vibrant were processed with CheckV (database v1.0) (Nayfach et al., 2020) and contigs retained if they met the following criteria:

- Category 1 contigs - CheckV identified viral genes > 0
- Category 2 contigs - CheckV viral genes = 0 **and**:
  - CheckV host genes = 0 **or**
  - DeepVirFinder score >= 0.9 **and** DeepVirFinder p-value <0.05 **or**
  - VirSorter 2 score >= 0.95 **or** VirSorter 2 hallmark genes >= 2 **or**

- Vibrant quality score = High or Medium

Sequencing reads were rarefied by subsampling paired reads to the size of the smallest non-negative extraction control library using seqtk v1.3 (https://github.com/lh3/seqtk). Rarefied reads were mapped back to vOTUs using bbmap (vslow = t, minid = 0.9, part of BBTools v38.49) at 90% minimum identity and contigs with a horizontal coverage above a cutoff threshold of 70% genome length in any given sample were regarded as present within that sample. vOTUs with a horizontal coverage over 70% from reads in the negative extraction control library were excluded from further analyses.

Processing of the outputs of DeepVirFinder, VirSorter 2, VIBRANT and CheckV was performed by a custom R script (LT-viral_contig_curation.Rmd) made available in this chapter's GitHub repository (https://github.com/LSHillary/LtDnaVirome).

Taxonomic assignment and host prediction was carried out using VPF-Class (membership ratio and confidence score 0.2, or 0.3 and 0.5 respectively) (Pons et al., 2021). If more than one taxonomic assignment could be made at this threshold, then the vOTU was assigned as inconclusive. All contigs >1,500 bp were also compared to the Refseq viral database protein database (v210) using DIAMOND BLASTX v2.0.8 (–ultra-sensitive –evalue 0.00001 –max-target-seqs 15) and taxonomy assigned using MEGAN Community Edition 6.21.17 (longReads lowest common ancestor algorithm) (Buchfink et al., 2014; Huson et al., 2016).

Analysis of virus lifestyle was based on circular contigs identified as lytic or lysogenic by VIBRANT. Auxillary metabolic genes were predicted by VIBRANT using default settings. Identification of antibiotic resistance genes was performed using DIAMOND v2.0.8 to run a BLASTx search (e-value = 0.001, minimum score = 70) of vOTUs against the SARG v2.0 database (Yin et al., 2018).

2.5.4. Ecological data analysis

Statistical analyses were performed with R (script LtDnaVirome_Figure_Generation.R). α-diversity metrics were calculated using the Vegan package. Differences in α-diversity, viral lifestyle and relative abundances of biosolids-associated viruses were tested by first

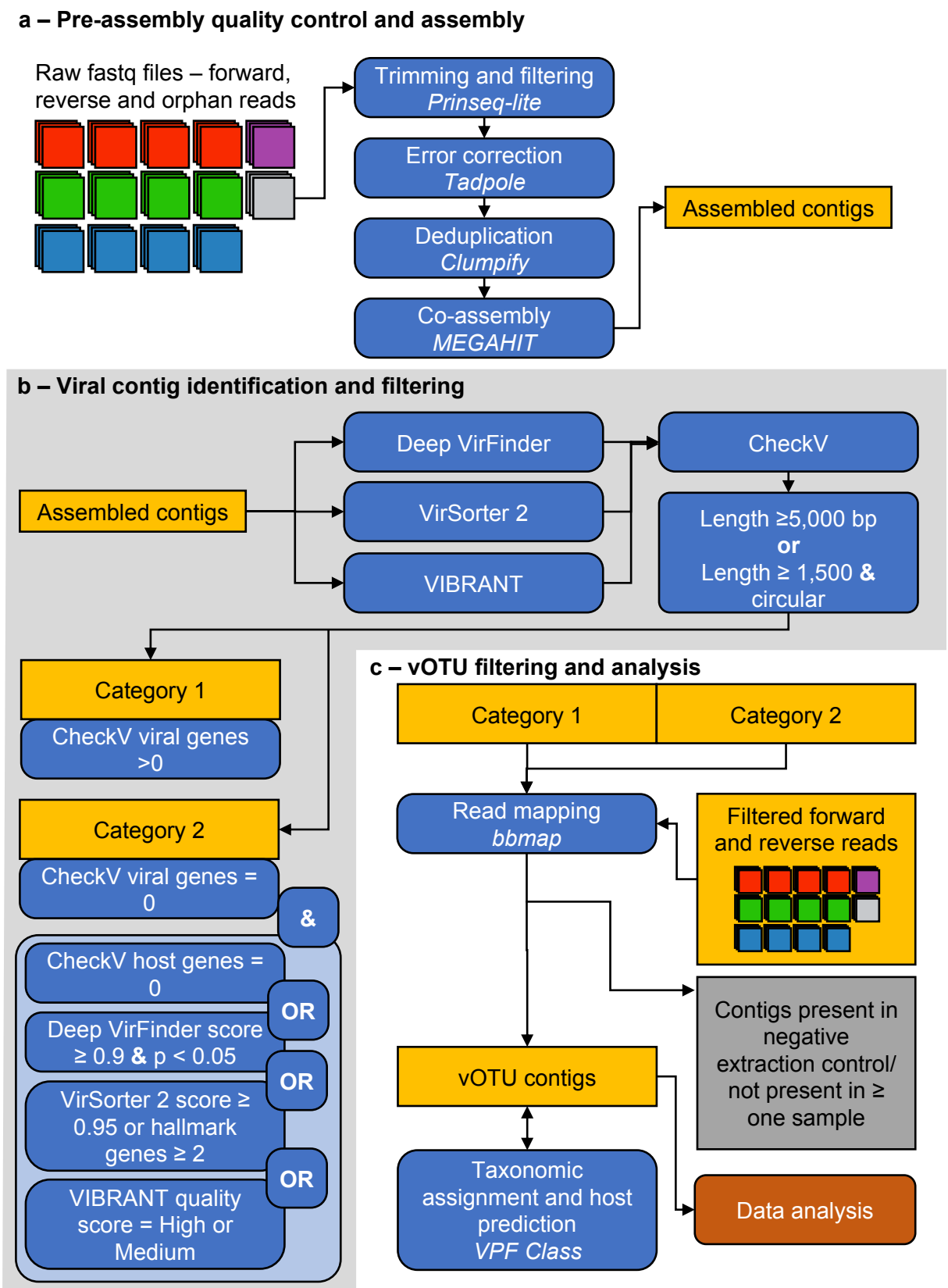


Figure 2.3.: Bioinformatics workflow of how raw sequencing reads underwent quality control and were assembled into contigs (a), how viral contigs were identified (b) and how vOTUs were further screened and taxonomy/ host prediction assigned (c).

generating generalised linear models or beta regression models and analysis of deviance tests using the function Anova in R. Descriptions of each GLM used in this study are included in Supplementary Table A.5. β-diversity non-metric multidimensional scaling was performed on a Bray-Curtis distance matrix using the function metaMDS. PERMANOVA tests were used to assess differences between treatments and blocks.

2.6. Results and discussion

2.6.1. Identification and classification of viral contigs

In this work, we characterised the effect of long-term (1994-2018 - 25 years) and historical (1994-1997 - 4 years) biosolids amendment on the soil DNA virus community (Fig. 2.4a). Virus-like particle DNA was extracted from soil samples of four plots of unamended control, historically biosolids amended and long-term biosolids amendment treatments, arranged in two blocks, each containing two plots of each treatment. The biosolids material used to amend treated plots and a negative extraction control of molecular biology grade water were also analysed, making a total of 14 samples. DNA from purified, DNase treated VLPs was used to prepare 14 virome sequencing libraries for high-throughput Illumina sequencing. Filtered sequencing reads were co-assembled and contigs >1500 bp were used in further analysis. The co-assembled contigs were first screened for putative viral sequences by DeepVirFinder, VirSorter and VIBRANT (Guo et al., 2021; Kieft et al., 2020; Ren et al., 2020). Each tool identified approximately one third of all putative viral contigs with 69% initially identified by two or more viral identification tools (Fig. 2.4b). Contigs underwent further quality control using CheckV (Nayfach et al., 2020) or by application of stringent cutoff thresholds for DeepVirFinder, VirSorter2 and VIBRANT (Category 1 and Category 2 viral contigs, seem methods and Fig. 2.4c).

Quality filtered reads from individual libraries were rarefied by randomly subsampling without replacement and mapped to the collection of viral contigs. A vOTU was determined as present within a sample if 70% of the length of the contig achieved >1.0x coverage (Fig. 2.4 d). Contigs were excluded from further analysis if they were present in the negative extraction control or not present in at least one sample (Fig. 2.4d). Relative abundances were calculated by normalising fragments per kilobase million (FPKM)

counts by library size (CPM - counts per million). 5,385 contigs were identified as present in at least one sample and are subsequently referred to as viral operational taxonomic units (vOTUs). Virome summary statistics are provided in Supplementary Fig. A.1.

Taxonomic assignment and host prediction was carried out using VPF-Class with a membership ratio and confidence score of 0.2 (Pons et al., 2021). 29% of vOTUs could be unambiguously assigned at the genus level, with a further 16% assigned as inconclusive, sharing genes from multiple genera (Fig. 2.5). The remaining 55% could not be assigned and this could be due to a combination of lack of predicted genes with known taxonomy and viral sequences that are too divergent from those described in the VPF-Class database. At the family level, 51% of vOTUs were unassigned, whereas 30% were inconclusive and 19% unambiguously assigned. Host prediction was also carried out by VPF Class, with a membership ratio of 0.3 and confidence score of 0.5. Host genus could be assigend for 9% of vOTUs (plus 1% inconclusively).

2.6.2. Biosolids amendment has limited effects on soil viral community structure

The majority of vOTUs in both the control and long-term treated plots were shared with other soil treatments (see Fig. 2.6a) with historically treated plots containing a high number (n = 1270) of vOTUs unique to this treatment. The number of vOTUs shared between treatments varied, with 96%, 64% and 79% of all vOTUs found within a treatment being shared with at least one other soil treatment (long-term, historical and control treatments, respectively). The number of vOTUs shared between replicates also varied, with 46%, 55% and 62% of vOTUs appearing in a single replicate within the long-term, historical and control treatments, respectively. A common core virome of 636 vOTUs (15%) was shared between all three soil treatments. The reduced number of total and shared contigs in the long-term treated viromes are likely to be linked to shorter vOTU contig length and a lower percentage of reads mapping to vOTUs in these viromes, suggesting that these libraries were more fragmented (see Supplementary Fig. A.1). Despite this, the long-term treated plots shared 13 vOTUs with the biosolids virome, whilst the historically treated virome shared six, despite having more total detected vOTUs. No vOTUs from the biosolids virome were detected in the control treatment soils, and no

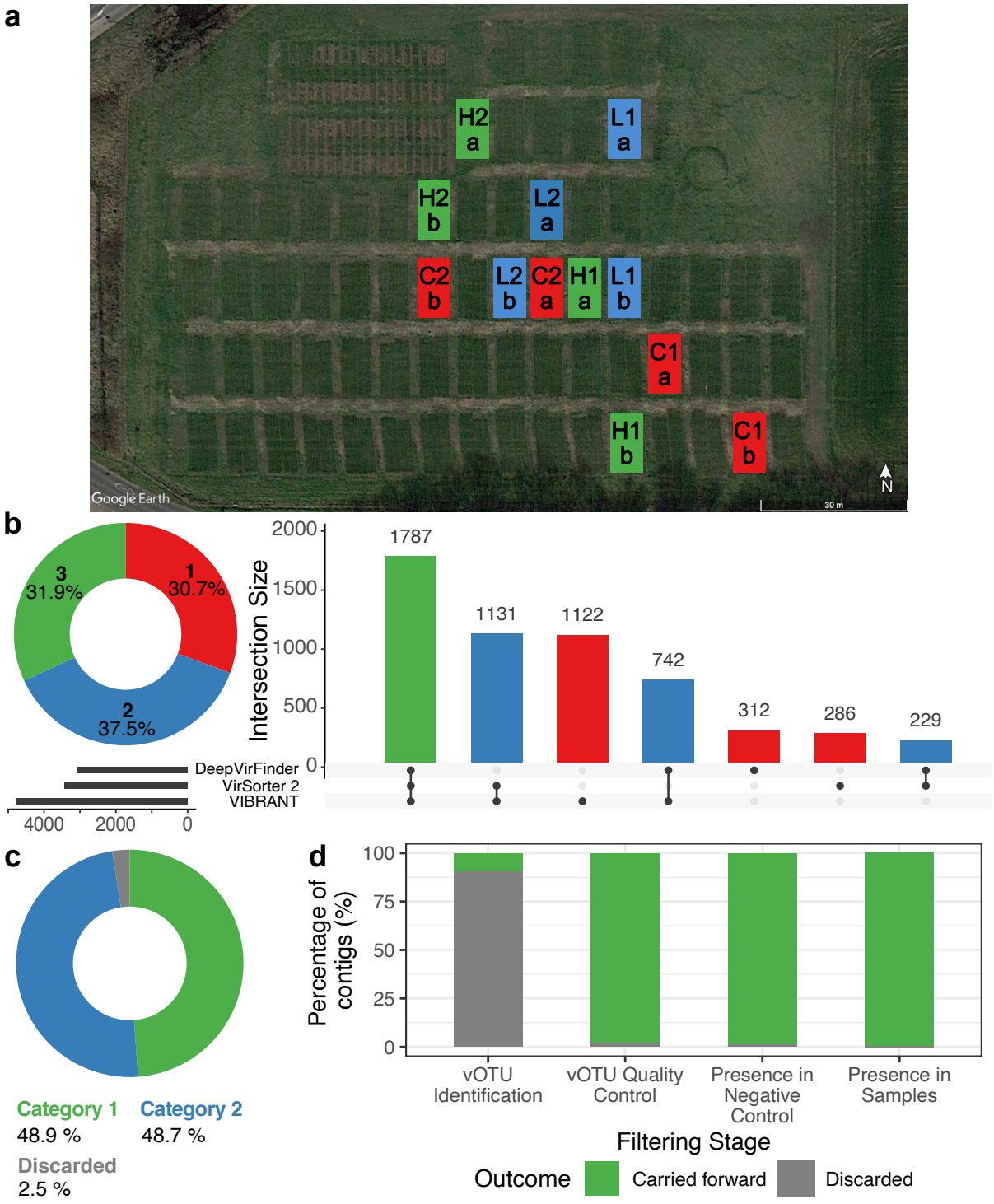


Figure 2.4.: Experimental design (a) and outcome of viral contig identification. Plots are coded by treatement (C = control, H = historical, L = long-term), block (1 or 2) and within block replicate (a or b). Putative viral contigs were identified using DeepVirFinder, VirSorter 2 or VIBRANT, with 69.4% of viral contigs identified by two or more tools (b). The contig identification UpSet plot (b) shows that VIBRANT identified the most contigs, with DeepVirFinder and VirSorter 2 identifying similar quantities. When passing contigs through CheckV or more stringent tests, 48.9% of contigs contained at least one viral gene identified by CheckV (c), 48.7% of viral contigs with no CheckV viral genes, but passing strict thresholds from the other three tools, and only 2.5% of contigs discared at this stage. (continues on next page...)

Figure 2.4.: (...) Although <12% of contigs were predicted to be viral, >95% of those passed subsequent additional checks.

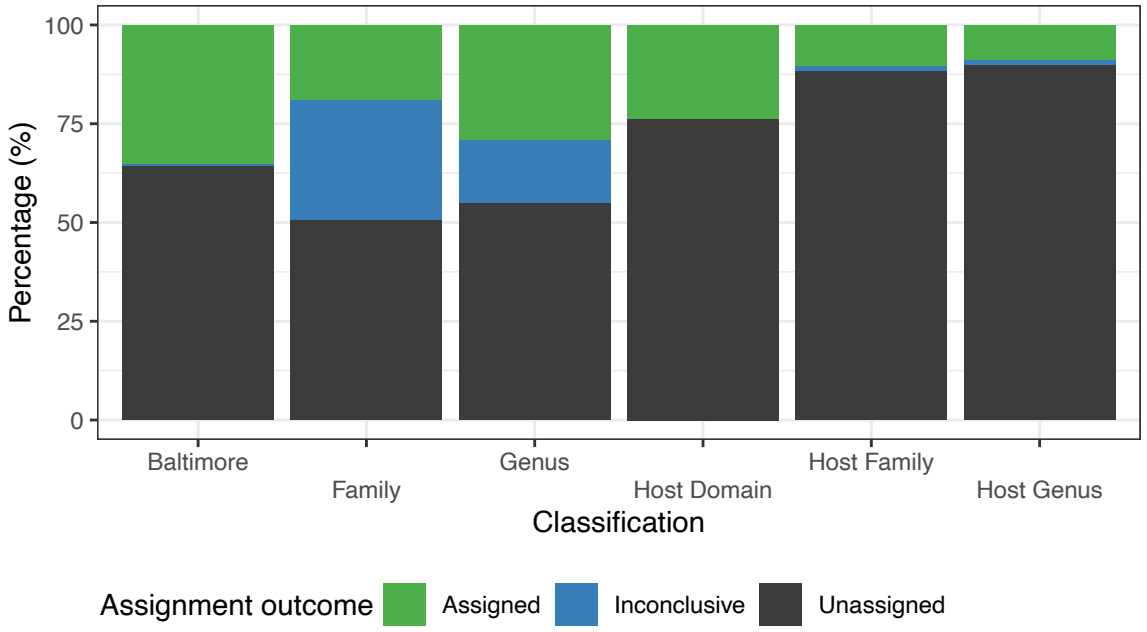


Figure 2.5.: Taxonomic and host assignments of viral contigs identified in all samples. vOTUs were assigned taxonomy at Baltimore (equivalent to kingdom, but superseded by recent changes to ICTV viral taxonomy), family and genus level by VPF Class using a membership ratio and confidence score of 0.2. Host prediction was also performed with VPF Class, at domain, family and genus level, using a membership ratio of 0.3 and confidence score of 0.5. In each case, if more than one taxonomic classification could be made at that level, the vOTU was assigned "inconclusive". The majority of vOTUs were unassigned both taxonomy and host prediction at each taxonomic level.

biosolids-associated vOTUs were detected at a higher relative abundance in soil viromes compared to the biosolids virome (see Supplementary Table A.4 and Supplementary Figs. A.4 and A.5).

Generalised linear models (GLMs) and beta regression models were used to statistically evaluate apparent differences in α-diversity metrics Fig. 2.6b, c and d). Significant differences were found in vOTU richness (ANOVA, treatment -$\chi 2$ = 14.2, p = 0.000821 but not block - $\chi 2$ = 0.0711, p = 0.790), and were not found in Simpson or Shannon indexes (ANOVA, p > 0.05, see Table A.5). Although fewer vOTUs were recovered from the long-term plots compared to the control and historical plots, Fig. 2.6c and d show that the soil viral communities are expectedly highly diverse, with Shannon or Simpson

indexes of similar magnitude to previous soil viromics studies (Chen et al., 2021; Lee et al., 2021). This also indicates that biosolids amendment does not alter the overall diversity of the soil viral community. In comparison, the effects of biosolids amendment on microbial communities is mixed, with some studies reporting no change (Zerzghi et al., 2010), whilst others report increases in key soil microbial taxa (Price et al., 2021).

Analysis of β-diversity was performed via non-parametric multidimensional scaling (NMDS) and differences between communities assessed via PERMANOVA (Fig. 2.6 e). Assigned viral families were fitted to the ordination plot to identify which viral families were driving differences in communities. *Microviridae* and *Podoviridae* were determined as significant drivers of vOTU-level diferences between sites ($p = 0.007$ and 0.012), with the directions of the fitted vectors indicating that *Podoviridae* separated control block 1 viromes from the other samples, *Microviridae* separating the majority of historically treated and control block 2 viromes from the others. Despite no observable difference in α-diversity, both treatment and blocking significantly impact virus community composition, suggesting that the magnitude of shift due to biosolids amendment is relatively small (PERMANOVA, treatment - $R^2 = 0.299$, $p = 0.0053$, block - $R^2 = 0.209$, $p = 0.0049$).

2.6.3. Biosold-associated viruses form minor components of soil viral communities

Fig. 2.7 shows the proportion and relative abundance of biosolids-associated vOTUs in the three treated soils. Whilst the number of biosolids-associated vOTUs present in the long-term treated plots in Fig. 2.5 (a) is approximately twice that of the historical plots, this difference increased to 15-fold once adjusted in Fig. 2.7 (a) for the total number of recovered vOTUs in each sample. In addition, the relative abundance of biosolids-associated vOTUs in historically amended soils is 100-fold lower when compared to long-term amended soils (see Fig. 2.7 (b)). These results suggest that whilst biosolids-associated vOTUs can potentially persist in soils for extended periods of time (>20 years), they become a very minor component of the overall soil viral community. Interestingly, the small number of biosolids-associated vOTUs form a substantial percentage of the viral communities in the long-term plots. This suggests that biosolids represent a significant source of viruses, but

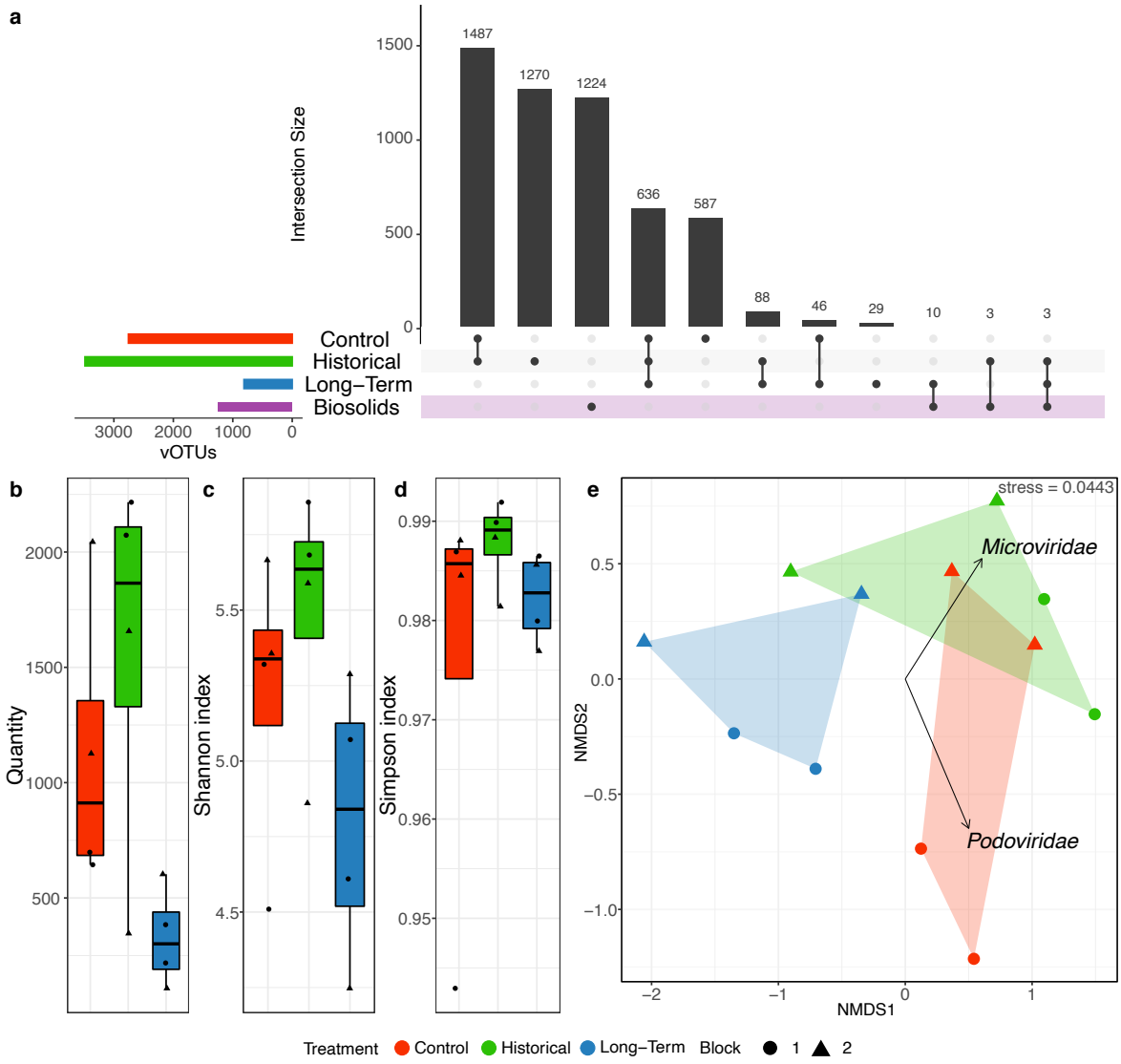


Figure 2.6.: Ecological analysis of vOTU diversity across treatments. An UpSet plot was generated (a) to highlight soil-borne and biosolids-associated vOTUs shared between treatments. The majority of vOTUs were either shared between Control (no amendment) and Histroical (1994-1997) treatmnets, with far fewer vOTUs detected in Long-Term treated plots and the biosolids virome. A core soil virome of 636 vOTUs were detected in all three soil treatments. Of the biosolids-associated vOTUs, 10 were shared with the Long-Term treatment plots, 3 with Historical plots and 3 between with both. No biosolids-associated vOTUs were detected in any Control treatment plot. α-diversity metrics (b-d) were calculated from number of vOTUs present (b) or CPM values (c-d). Although fewer vOTUs were detected in long-term treated plots, no significant differences were detected in Shannon index or Simpson index. β-diversity was analysed by NMDS (Bray-Curtis) ordination of vOTUs discovered in this study (e) and displayed according to treatment and experimental blocking. Summed abundances of viral families were fitted to the NMDS co-ordinates (e) and families with a p-value <0.05 displayed as vectors. *Podoviridae* and *Microviridae* appear to drive differences largely between block 1 (circles) and block 2 (triangles), and between the Long-Term treatment and the Control/ Historical treatments. Both treatment and block produced significant differences between viral communities (PERMANOVA, treatment - $R^2 = 0.299$, $p = 0.0053$, block - $R^2 = 0.209$, $p = 0.0049$).

do not become a permanent feature of the soil viral community.

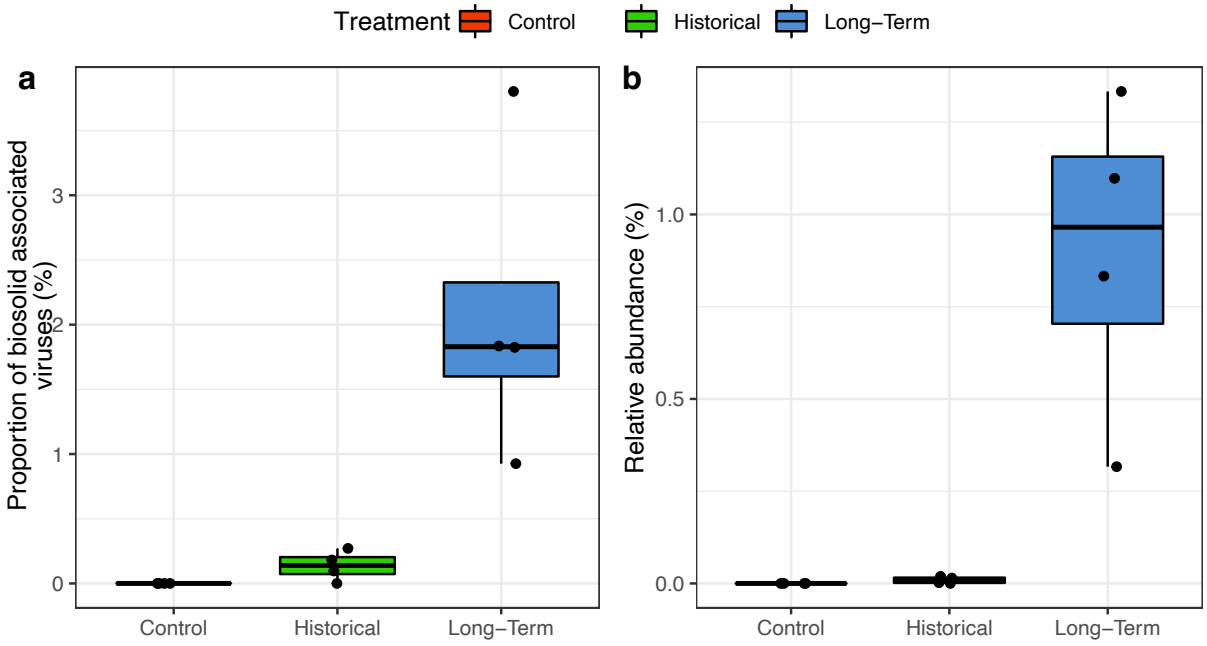


Figure 2.7.: Biosolids-associated vOTUs as a proportion of the total number of vOTUs within each sample (a) and their relative abundance (b) in long-term, historical and control treatment soils. Biosolids-associated viruses formed 2.1% ±1.2 (mean ±s.d.) of all vOTUs identified in long-term amended plots compared to 0.14% ±0.12 of all vOTUs identified in the historical treatment plots. This difference is magnified further when considering relative abundance, with biosolids-amended vOTUs forming 0.9% ±0.44 of the soil viral community in long-term amended plots, compared to 0.009% ±0.009 in historically treated plots. Treatment was found to have a statistically significant effect in both measures (GLM, quasibinomial distribution, ANOVA, $p < 0.0001$).

2.6.4. vOTU taxonomic assignment and host-prediction demonstrates the presence of viruses of a wide range of potential hosts in biosolids amended and unamended soils

vOTU taxonomic assignment and host prediction was carried out using VPF-class (Fig. 2.8). The main families of virus identified were bacteriophages of the *Caudoviricetes* class (*Siphoviridae*, *Myoviridae* and *Podovidae*), the ssDNA bacteriophages *Microviridae* and the viruses of eukaryotic algae, *Phycodnaviridae* (Koonin et al., 2020; Van Etten et al., 2002).

Podoviridae are dsDNA bacteriophages and typically ubiquitous in viral ecology studies. A specific group of *Podovidae* of interest are the crAss-like phages, which form a common

component of the human gut viral community (Edwards et al., 2019). To identify specific human pathogens, all contigs >=1,500 bp in length were compared to the refseq viral database (BLASTX, e-value cutoff of 10^{-5}) although none were identified in any virome. 43 contigs were identified as possible matches to the genome of crAssphage ϕ-001, of which 6 contigs also passed the viral contig identification filtering. All 6 contigs were found within the biosolids virome sample, but not in any of the soil samples. These results indicate that crAss-like phages are relatively stable within biosolids and can be potentially used as long-term markers of faecal contamination. However, future work seeking to apply this would most likely need to apply more sensitive methods such as qPCR (Farkas et al., 2020c).

Microviridae are small ssDNA viruses of bacteria that can be found in a variety of ecosystems, including bog, fen and agricultural soils, as well as human saliva and faeces (Han et al., 2017; Quaiser et al., 2015). Their detection here demonstrates the utility, and importance, of using library preparation techniques that can amplify both ssDNA and dsDNA with minimal bias when investigating DNA viral communities. *Phycodnaviridae* were present only in soil viromes. These spherical dsDNA viruses typically infect freshwater algae, however they have been observed previously in soil (Adriaenssens et al., 2017).

A variety of other viral families were detected in soil and biosolids samples in low relative abundances and inconsistently between sample replicates. Fig. 2.8 b shows that the rare viral families present in these samples include viruses of archaea, arthropods, and vertebrates (e.g. *Baculoviridae*, *Parvoviridae*). The *Fuselloviridae*, *Lipothrixviridae* and *Pleolipoviridae* have to date, only been described in extreme environments (Blake et al., 2004; Pietilä et al., 2016; Prangishvili and Krupovic, 2012). Their presence in temperate soil viromes may indicate distant relatives, or potential misclassification, and interpretation of the presence of rare taxa should be performed with caution. Whilst this dataset does not support a more extensive analysis of their role in soil ecology, the examination of the ecological functions of rare viral families should be a focus for future research.

The dominant predicted hosts across all treatments were *Actinobacteria*, *Bacteroidetes*, *Firmicutes* or *Proteobacteria*, all of which are ubiquitous in soil environments (Manuel et al., 2018), demonstrating a similar trend to acidic soils observed by Lee et al. (2022).

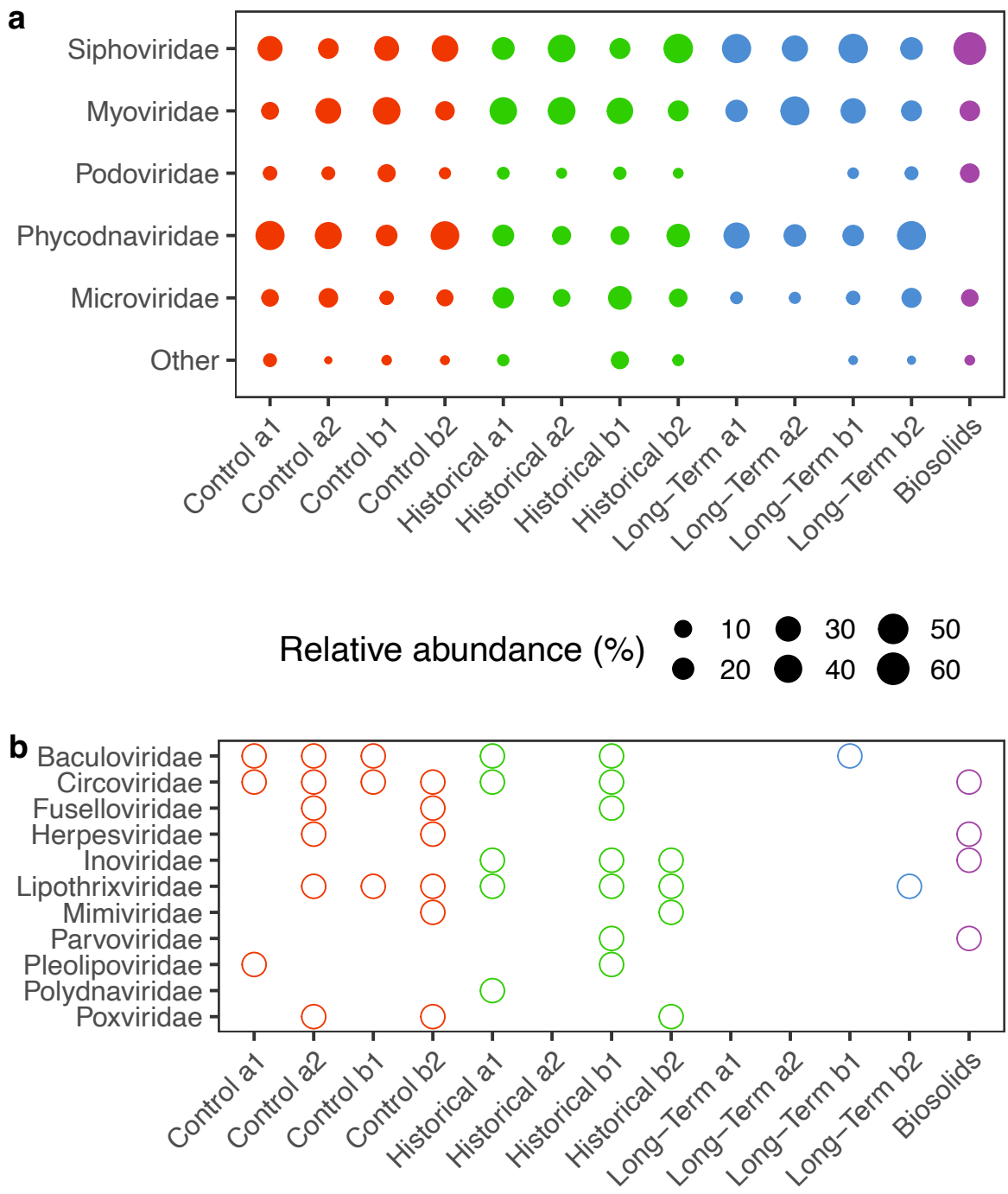


Figure 2.8.: Relative abundances of common viral families (a), coloured by treatment type and sized by the percentage of CPM values of all vOTUs with assigned taxonomy at the family level (see Methods) and presence of rare viral families (b) by treatment.

Interestingly in this study, no bacteriophages of the common soil bacterial phyla *Acidobacteria*, *Chloroflexi*, *Planctomycetes* or *Verrucomicrobia* were identified in any soil virome. A sizable minority of vOTUs (3-26%) have predicted hosts belonging to the *Chlamidiae* (see Fig. 2.9). Although these are obligate, intracellular parasites, they are still commonly found in terrestrial environments, and particularly abundant in soils (Collingro et al., 2020). The harsh sample processing methods required for eluting viruses from soil particulates may have resulted in cell lysis, releasing these bacteriophages into the buffer solution during virion desorption.

It is notable that viruses of both *Archaea* and *Eukarya* were detected in this dataset, although at relatively rare abundances and some caution should be used in interpreting their presence as host prediction, particularly of *Eukarya*, can vary in accuracy (Pons et al., 2021). The combination of this inaccuracy, reliance on marker genes with established phylogeny and limited numbers of vOTUs in this study with predicted hosts may be causal contributors to the high level of variation between replicates that was observed, including no predicted hosts in long-term amendment replicate a2. Nevertheless, the presence of archaeophages, *Phycodnaviridae*, and *Baculoviridae* demonstrate that soil viral ecology of terrestrial archaea and eukaryotes warrants further investigation as they may play an important, and as yet undercharacterised, role in soil microbial and macro-biological ecology.

2.6.5. Auxillary metabolic genes, antimicrobial resistance genes and vOTU lifestyle

VIBRANT was used to identify 1048 auxillary metabolic genes (AMGs) carried by 909 vOTUs, with no discernable pattern observable between treatments, given the number of AMGs identifiable (Fig. 2.10 a). No statistically significant difference in the number of AMGs per vOTU was found between treatments or block (Beta regression, $p > 0.05$). Enrichment of specific forms of metabolism was examined by analysing the number of AMGs from different pathways, normalised by total vOTUs per virome using NMDS (Fig. 2.10 b), with treatment, but not block, exhibiting a significant difference (PERMANOVA, p(treatment) = 0.0382, p(block) = 0.546). The individual plots from the long-term and historical treatments that are separated from the cluster in the centre of the NMDS plot

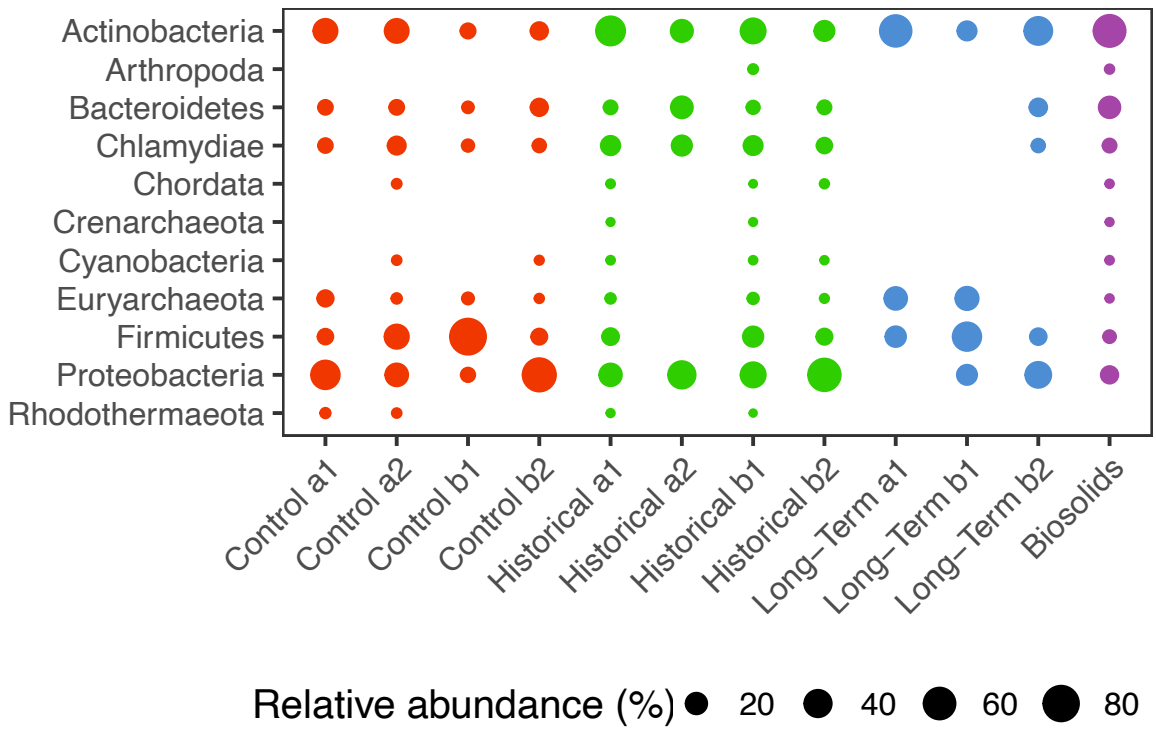


Figure 2.9.: Relative abundances of host prediction from VPF Class, coloured by treatment, sized by the percentage of summed CPM values of all vOTUs with hosts predicted at the family level.

are those with the lowest number of vOTUs, suggesting that the observed differences in AMG distribution are a function of vOTU recovery rather than the effects of biosolids amendment. It is therefore likely that biosolids amendment does not substantially alter viral community functional capacity as well as overall diversity previously discussed, with similar results having been observed for pH (Lee et al., 2022).

It is important to note that these data do not verify that these genes are actively expressed in the viruses' hosts, nor that the biosolids-associated viruses present within the long-term treatment soils are able to actively replicate. Stable isotope probing has previously been used to examine virus-host dynamics in actively replicating soil microbial communities (Trubl et al., 2021) and may in the future be used for assessing how land management practices can influence community dynamics, as well as diversity and functional capacity, alongside metatranscriptomics.

Additionally, antimicrobial resistance (AMR) genes were searched for in all vOTUs, identifying four vOTUs containing putative AMR genes (see Table 2.1). These were primarily

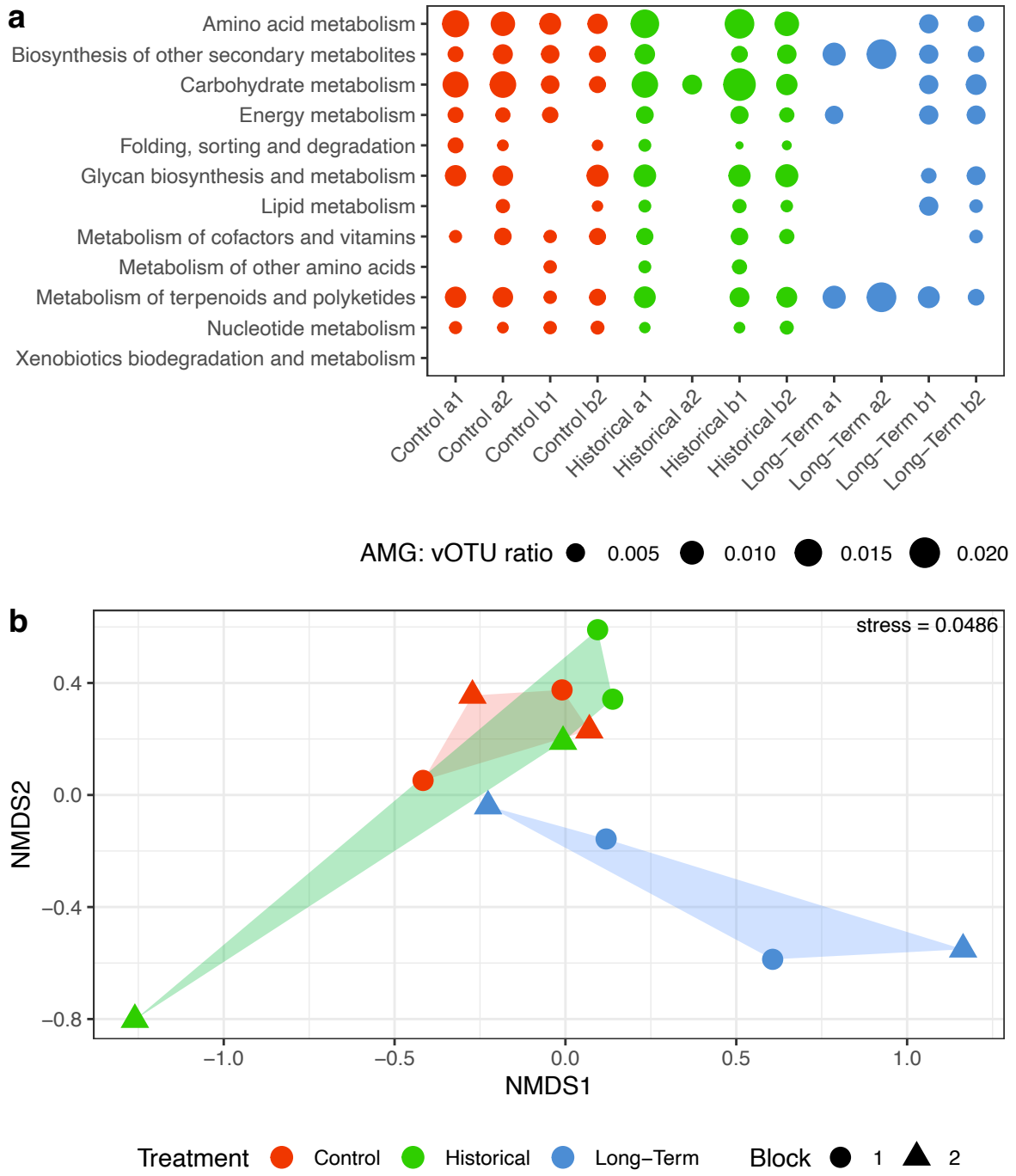


Figure 2.10.: (a) AMG:vOTU ratios of different metabolisms within Control, Historical and Long-Term treated plots with size being proportional to the AMG:vOTU ratio. (b) NMDS plot (Bray-Curtis) of the same data, showing metabolisms that cause significant differences between samples. Treatment caused a significant difference between samples whereas experimental block did not (PERMANOVA, p(treatment) = 0.0382, p(block) = 0.546). Note, the outlier sample Historical a2, which showed substantially fewer AMGs than the other replicates within that treatment

located in biosolids or historical samples, with one ARG carrying vOTU also located in a Control treatment sample. None of the biosolids-associated ARG-carrying vOTUs were also found in a soil virome. The small number of AMR genes identified in this study is similar to results of Chen et al. (2021), who identified a total of 16 ARGs across eight different treatments, indicating that the presence of virus-associated ARGs in soil viral communities is a rare occurence.

Table 2.1.: Identity of ARG containing vOTUs

Contig	Locations	ARG class
k127_490966	Biosolids	vanD gene cluster
k127_2835375	Historical a1, b1 and b2	bifunctional UDP-4-amino-4-deoxy-L-arabinose formyltransferase/UDP-glucuronic acid oxidase ArnA (polymyxin resistance)
k127_3375170	Biosolids	vanD gene cluster
k127_3820138	Historical a1 and Control a2	Dihydrofolate reductase type 2

VIBRANT was used to predict lysogenic viruses from those vOTUs forming circular contigs, demonstrating that >99% of assignable viruses were lytic (see Fig. 2.11), with no statistically significant difference between treatment or block. This further underlines that biosolids amendment has minimal long-term impact on soil viral communities. The high percentage of lytic viruses is likely due to a combination two factors: the need to identify integrase genes to classify vOTUs as lysogenic (Kieft et al., 2020) and that soil/ biosolids viromes are generated by eluting, purifying and sequencing intact pre-existing virus like particles from sample matrices (Trubl et al., 2019). The strategy of pairing viromes and unfiltered metagenomes, or inducing bacteriophage production and lysis with mitomicin c are two strategies that can increase the diversity of recoverable viral sequences and so allow a better understanding of the diversity of soil viruses and their functional potential.

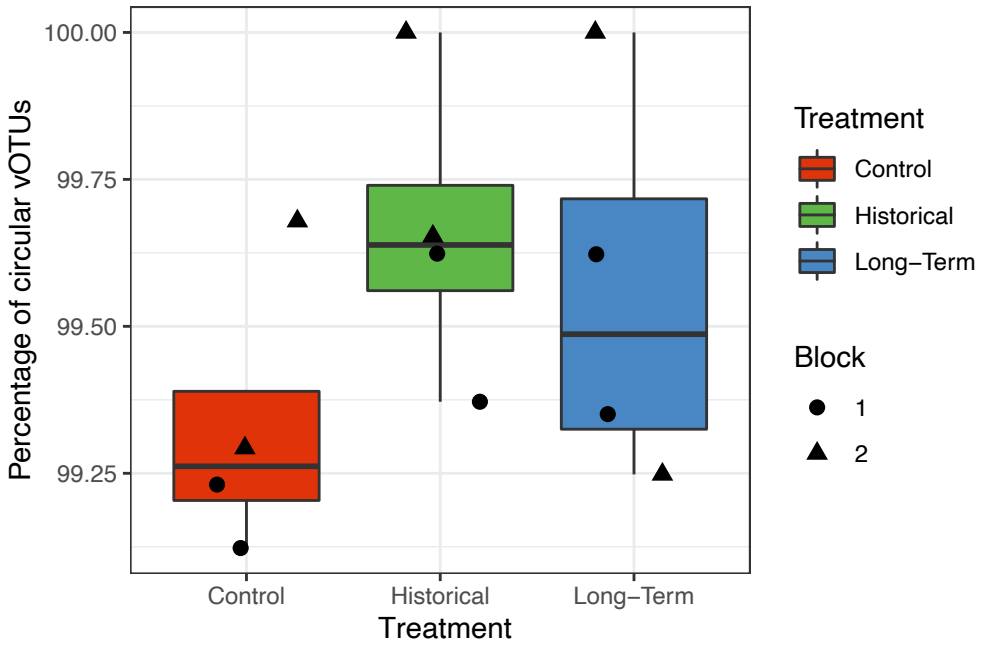


Figure 2.11.: Comparison of the percentage of circular vOTUs as identified as lytic by VIBRANT. No statistically significant differences were found between treatments or blocks (binomial GLM, ANOVA, $p = 0.351$ (treatment), $p = 0.314$ (block))

2.7. Conclusions

The aims of this study were to examine (1) how biosolids amendment impacts the soil viral community, (2) how long-term amendment compares to historical treatment, and (3) whether biosolids-associated viruses can persist in soils over extended timespans. Long-term amendment of soil with biosolids ws shown to have a direct effect on the soil viral community by introducing biosolids-associated viruses into the soil environment that can persist over a timescale of months to potentially years. The amount of biosolids-associated viruses present in historically treated soils were lower in both terms of number of biosolids-associated vOTUs (15-fold) and their relative abundance within the soil virome (100-fold). This suggests that whilst the use of biosolids as a fertiliser does impact the soil viral community, both directly and indirectly, those effects diminish over time. Biosolids amendment has limited to no significant effect on the overall diversity of soil viral communities and their functional capacity, but further work is required to examine their effects on host range, metabolic activity and community dynamics. The role of *Phycodnaviridae* and viruses of eukaryotes more broadly within soil microbial ecology warrants further investigation. The lack of human pathogenic viruses such as human adenovirus suggests that the risk to human health from biosolids-associated viruses is minimal,

however follow-up studies using qPCR, and a comparison between different standards of biosolids would be needed to further understand this.

Biosolids are used globally for their beneficial effects on soil health and enhancement of a variety of ecosystem services (Trimmer et al., 2019). Whilst this study demonstrates biosolids amendment has limited long-term effects on soil viral communities, and that previous work has estimated the public health risk from biosolids amendment to be low (Gerba et al., 2002), the potential for inadequately treated biosolids to transfer human pathogens to land still remains (Bibby and Peccia, 2013). The role and accumulation of other substances potentially hazardous to health such as antimicrobial resistance genes in bacteria, heavy metals and microplastics remain areas of concern when utilising biosolids as an agricultural nutrient source as part of the circular economy (Crossman et al., 2020; Ju et al., 2016; Venegas et al., 2021).

2.8. Data and code availability

Sequencing read files and vOTU contigs analysed in this study will be deposited with the NCBI and made publicly available on publication in a peer reviewed journal. Code related to this chapter will be made publicly available via GitHub at https://github.com /LSHillary/LtDnaVirome upon publication.

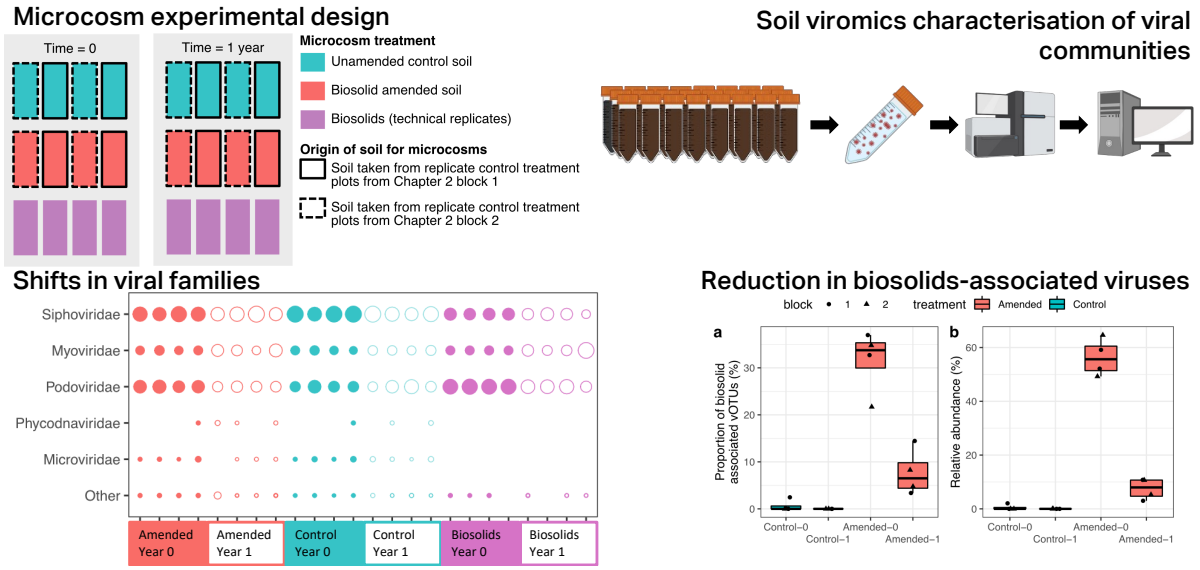
2.9. Acknowledgements

LSH was supported by a Soils Training and Research Studentship (STARS) grant from the Biotechnology and Biological Sciences Research Council (BBSRC) and Natural Environment Research Council (NE/M009106/1). Additional funding was provided by Welsh Water. DNA sequencing was carried out by the Centre for Genomics Research at the University of Liverpool with special thanks to John Kenny and Charlotte Nelson for assistance with library preparation. Data analysis utilised high performance computing resources from Supercomputing Wales. The authors would like to thank Javier Hernandez for assistance with sample collection, and David Fidler, Marine Cambon, Jim Downie, Dave Chadwick and Robert Griffiths for comments/ advice on data analysis and visualisation.

3

Impact of a single biosolids application on soil virus communities

3.1. Graphical abstract



3.2. Abstract

Anaerobically digested biosolids are frequently applied to agricultural land as a source of nutrients, but can also be a source of viral pathogens harmful to human, animal and plant health. Although the persistence of specific viruses in biosolids and in soil has been well studied, little is known about how this practice affects the soil viral community, the number and diversity of biosolids-associated viruses imported during amendment and their persistence post-treatment. We used viromics to characterise shifts in the DNA soil virus communities in conventional biosolids amended and unamended soil microcosms at the time of amendment and after one year. The viromes of 24 microcosms of biosolids, soil, and biosolids amended soil were characterised by high-throughput sequencing of virus-like-particle associated DNA (4 replicates x 3 sample types x 2 time points - at the start of the experiment and after 1 year incubation at 10°C). We found that at the start of the experiment, biosolids-associated viruses formed 56% of the amended soil viral community, decreasing to 7.4% after one year. The addition of biosolids-associated viruses also resulted in shifts in auxillary metabolic gene profiles and the introduction of virus-carried antimicrobial-resistance genes. All soil viral communities were predominantly composed of *Siphoviridae*, *Myoviridae*, *Microviridae* and *Podoviridae*. Host prediction indicates an increase in relative abundance of bacteriophages of *Bacteroidetes* and a corresponding reduction of bacteriophages of *Actinobacteria* due to biosolids amendment. This effect and others from biosolids amendment are substantially reduced after one year, suggesting that the majority of effects on the soil viral community are short-lived. As pressures of food security from increasing populations rise, effective monitoring of the biosafety of biosolids will become increasingly important. With molecular techniques becoming more widely available, assessing biosolids for their viral content could become integrated into existing monitoring frameworks.

3.3. Keywords

biosolids, virome, soil, viral persistence, soil viral ecology, sewage sludge

3.4. Introduction

The persistence of biosolid-associated viruses is impacted by a range of soil properties, with temperature and adherence to soil particulates being the most important (Hurst et al., 1980). Chapter 2 demonstrated that whilst viruses can persist for over 10 years, their relative abundance diminishes significantly when compared to soils receiving annual biosolids amendment. Although the soil viral community remains largely unchanged by repeated biosolid amendment, this does not address how soil viral communities can be impacted by a single treatment of biosolids.

The viral community of anaerobically digested biosolids produced by a sludge treatment works is comprised of two components: viruses from outside of the sludge treatment works that have persisted through the wastewater and sludge treatment processes, and those associated with the treatment process itself, i.e. viruses of micro-organisms involved in anaerobic sludge digestion. The use of biosolids as agricultural fertiliser therefore poses a number of potential ecological risks: the dissemination of viral pathogens, disruption to naturally occurring soil viral and microbial communities and the dissemination of bacteriophages carrying antimicrobial resistance genes (ARGs).

Biosolids for use as agricultural fertilisers can be treated in a variety of ways (Sonune and Ghate, 2004), but viral persistence, and therefore pathogen dissemination, is dependent on viral species and treatment technology (Mocé-Llivina et al., 2003). Similarly, once biosolids have been applied to land, the local environments that viruses experience will affect viral persistence. Experiments examining the persistence of biosolid-associated viruses in soil have typically relied on detection of culturable human pathogens, or surrogate bacteriophages (Oliveira et al., 2019; e.g. Straub et al., 1992). Whilst this has the benefits of evaluating infectivity, it is limited to the detection of viruses with established *in vitro* culture techniques. Molecular techniques such as qPCR and high-throughput sequencing can also be used to detect the presence of viral pathogens (Farkas et al., 2020c; Roberts et al., 2016). From a community perspective, the extent of changes in microbial communities caused by biosolid amendment are varied (Price et al., 2021; Schlatter et al., 2019; Zerzghi et al., 2010), however a community-based approach is yet to be taken to assessing viral persistence and community shifts in the soil viral community triggered by single amendments of biosolids.

Bacteriophage associated ARGs have been previously detected in a range of environments, and are more commonly associated with environments where antimicrobial concentrations are high, such as pig faeces and untreated sewage sludges (Lekunberri et al., 2017). However, soil-borne bacteriophages have been demonstrated to possess a wide range of ARGs (Anand et al., 2016; Joseph et al., 2015), and so source tracking of bacteriophage-associated ARGs is critical to understanding their origin, and implications for land management policy and risk analyses.

Current UK regulations on the use of biosolids on agricultural land place time-limits on grazing and harvest intervals after amendment between 3 weeks and 30 months, depending on crop type and whether biosolids are treated to a conventional or enhanced standard (Agricultural Development and Advisory Service, 2001). In this chapter, a medium-term microcosm experiment was established to examine the shifts in soil viral community and persistence of biosolids-associated viruses from a community-level perspective after one year. 50 g microcosms of soil taken from control plots analysed in Chapter 2 were amended with the equivalent of 7 dry tonnes per hectare of conventionally treated biosolids, a typical application rate for arable soil, although this can vary substantially, depending on national/ local limits and soil properties (LeBlanc et al., 2008). Amended and unamended soil microcosms were either frozen at -80°C or stored at 10°C, the approximate UK mean soil temperature (Busby, 2015), for one year. Viromics was used to characterise the viral communities of each microcosm and the effects of treatment, time and soil experimental block were examined. We aimed to answer (1) what is the diversity of biosolids-associated viruses imported into the soil viral community after amendment, (2) what is the level of biosolids-associated virus persistence after 1 year under controlled conditions and (3) how does biosolids-amendment affect the soil viral community over this time period.

3.5. Materials and methods

3.5.1. Experimental design

Soil samples were collected from four plots, arranged into two blocks, with no history of biosolids amendment from a long-term experiment in Woburn, Bedfordshire, UK examining the effects of biosolids derived on soil health (Gibbs et al., 2006). Soil properties are

described in Supplementary Table A.1 and Table A.2.

Wheat was grown annually on the plots with N, P and K fertilisers added as required, as indicated by soil testing. Six evenly distributed cores (20 cm depth) were extracted using a 2.5 cm diameter half moon auger and stored separately on ice before transport back to Bangor University. Between the sampling of each plot, the auger was sequentially wiped clean with 1% Virkon, deionized water and 70% industrial methylated spirit. A dummy core was taken from the middle of each plot and discarded before collecting samples for analysis. Each core was be sieved to 2 mm and equal masses of the six cores combined, manually homogenised and frozen in aliquots at -80°C until analysed. Additional 100 cm^3 cores were taken from the centre of each plot, weighed and dried overnight at 105 °C and used to calculate bulk density.

Mesophilic anaerobically digested biosolids treated to the UK conventional standard were sourced from a single wastewater treatment site. Replicate microcosms of 50 g soil, biosolids or biosolids-amended soil were stored in light-fast 50 ml centriguge tubes and sealed with Parafilm to permit gas exchange. Biosolids were incorporated into amended soil microcosms at an equivalent rate of 7 tonnes per hectare, assuming incorporation down to a depth of 20 cm.

Time=0 microcosms were frozen at -80 °C until extraction. Time = 1 year samples were sealed with Parafilm to permit gas exchange and placed inside a plastic container sealed with Parafilm, containing 1 cm depth of water to increase humidity and reduce evaporation and stored in the dark at 10 °C for 12 months. Samples were weighed every 4 weeks to monitor evaporation and an equivalent mass of sterile deionised water was added to replace any mass lost between weigh-ins. Parafilm seals were replaced after each weigh-in. At the end of the incubation period, samples were frozen at -80 °C until DNA extraction.

3.5.2. Virus-like-particle DNA extraction and sequencing library preparation

Soil/ biosolids samples were resuspended in a 1:3 w/v ratio with amended potassium citrate buffer (10% potassium citrate, 1% phosphate buffered saline, 150 mM $MgSO_4$, 50 g with a total of 150 mL buffer per sample, divided between four centrifuge tubes - see Fig. 3.1) and agitated in 3 rounds of 30 s manual shaking followed by 1 minute vortexing. Samples were then shaken for 15 minutes on an orbital shaker at 300 rpm at 4 °C and centrifuged at 4,600 x g for 30 minutes at 4 °C. Supernatants were filtered through 0.22 μm PES membrane filters (Millipore) and concentrated using 50 kDa MWCO Amicon Ultra centrifugal ultrafiltration devices (Millipore) to a volume of <250 μL. Concentrates were DNase treated with 10U/ 100 μL of DNaseI (Invitrogen) at room temperature for 2 hours with samples gently mixed through inversion after 1 hour. DNase was inactivated with the addition of 100 μL per mL of 50 mM EDTA and incubating at 65 °C for 10 minutes. DNA was extracted using a DNeasy PowerSoil Pro DNA extraction kit (Qiagen) with minor modifications to the manufacturer's protocol: after the addition of solution CD1, samples were lysed by two rounds of vortexing for 3-4 seconds and incubating at 70 °C for 5 minutes. VLP DNA was also extracted from a negative control sample of 50 mL of molecular biology grade water.

DNA concentrations were quantified using the Qubit HS assay kit (ThermoFisher) and the presence of contaminants assessed using a Nanodrop 1000 spectrophotometer (ThermoFisher). A total of 25 sequencing libraries (see Fig. 3.1) were prepared by staff at the University of Liverpool Centre for Genomics Research using the NEBNext Ultra II FS library preparation kit. Libraries were pooled and sequenced on one lane of an S1 chip on an Illumina NovaSeq 6000. Initial demultiplexing and quality control performed by CGR removed Illumina adapters using Cutadapt version 1.2.1 (Martin, 2011) with option -O 3 and Sickle version 1.200 (Joshi and Fass, 2011) with a minimum quality score of 20.

3.5.3. Sequencing data processing and viral contig identification

Sequencing data were broadly processed as described in Chapter 2 (see Fig. 2.3). Raw sequencing reads were filtered with bbduk (ftl=3 maq=25 minlen=35), and error corrected with tadpole (mode=correct ecc=t prefilter=2). PCR duplicates were removed using

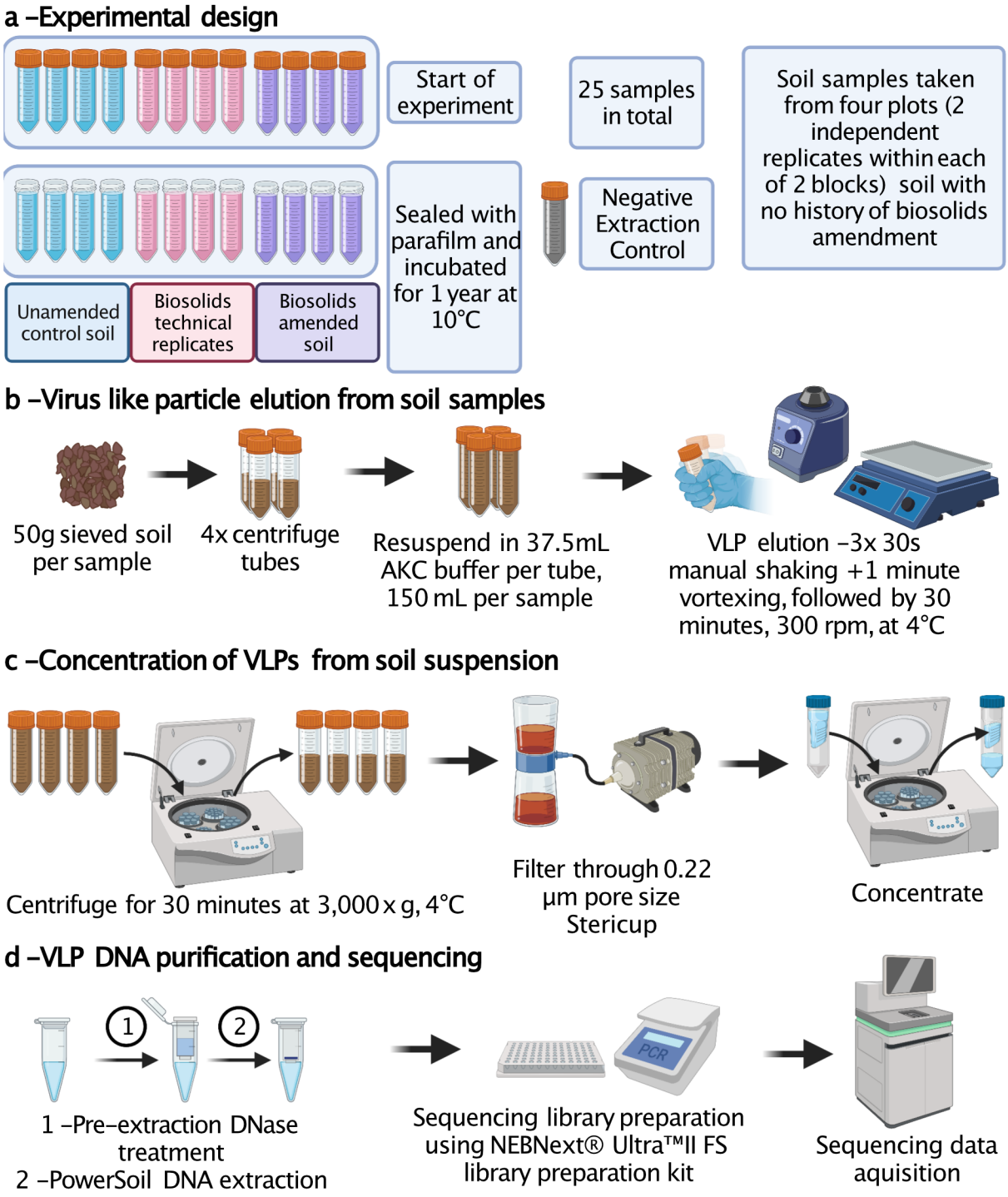


Figure 3.1.: Experimental design and sample processing. Unamended soil, biosolids amended soil, and biosolids microcosms were either frozen at -80°C or stored in darkness in light-fast 50 mL centrifuge tubes at 10°C, sealed with parafilm (a). The viromes of a total of 25 samples (4 replicates x 3 sample types x 2 time points + 1 negative extraction control) were characterised. Virus-like particles were eluted (b) and concentrated (c) and DNA extracted and used to prepare 25 libraries for sequencing on one lane of a an S1 chip on an Illumina NovaSeq 6000.

clumpify (dedupe subs=0 passes=2). All programs are from the BBTools package (v38.49: sourceforge.net/projects/bbmap/). All processed reads from each library were then co-assembled using MEGAHIT 1.1.3 (–presets meta-large) (Li et al., 2015).

Assembled contigs >1,000 bp in length were processed with three viral sequence identification tools: Deep Virfinder v1.0 (score >= 0.7 AND p value < 0.05), VirSorter 2 v2.1 (Score >= 0.5) and VIBRANT v1.2.0 (>= 4 hallmark genes) (Guo et al., 2021; Kieft et al., 2020; Ren et al., 2020). Contigs passing these thresholds with a length >= 5,000 bp or with a length >= 1,500 bp and identified as circular by either VirSorter 2 or Vibrant were processed with CheckV (database v1.0) (Nayfach et al., 2020) and contigs retained if they met the following criteria:

- CheckV identified viral genes > 0
- CheckV viral genes = 0 **and**:
  - CheckV host genes = 0 **or**
  - DeepVirFinder score >= 0.9 **and** DeepVirFinder p-value <0.05 **or**
  - VirSorter 2 score >= 0.95 **or** VirSorter 2 hallmark genes >= 2 **or**
  - Vibrant quality score = High or Medium

Reads were rarefied by subsampling paired reads to the size of the smallest non-negative control library using seqtk v1.3 (https://github.com/lh3/seqtk). Rarefied reads were mapped back to vOTUs using bbmap (vslow = t, minid = 0.9, part of BBTools v38.49) and contigs with coverage above a cutoff threshold of >70% genome length in any given sample were regarded as present within that sample. vOTUs with horizontal genome coverage of >70% from mapped reads from the negative control sample were excluded from further analyses.

Processing of the outputs of DeepVirFinder, VirSorter 2, VIBRANT and CheckV was performed by a custom R script (ST-viral_contig_curation.Rmd) made available in this chapter's GitHub repository (https://github.com/LSHillary/StDnaVirome).

Taxonomic assignment and host prediction was carried out using VPF-Class (Pons et al., 2021). A membership ratio and confidence score of 0.2 was used for taxonomic assignment and for host prediction, membership ratio = 0.3 and confidence score = 0.5. If more

than one assignment could be made at this threshold, then the vOTU was assigned as inconclusive.

Accessory metabolic genes were predicted using VIBRANT (default settings). Identification of antibiotic resistance genes was performed using DIAMOND v2.0.8 to run a BLASTx search (e-value = 0.001, minimum score = 70) of vOTUs against the SARG v2.0 database (Buchfink et al., 2014; Yin et al., 2018).

3.5.4. Ecological data analysis

Statistical analysis of the results of short-read mapping and taxonomic assignment was performed using a custom R script (ST_figure_generation.R, see GitHub repository). vOTUs were deemed within a sample present if >70% of the contig length was covered >1-fold from an individual library. Contigs with >70% coverage in the negative control library were removed from further analysis. The relative abundancies of vOTUs present in each sample were calculated by converting Fragments Per Kilobase Million (FPKM) values provided by bbwrap to Counts per Million using the fpkm2tpm function from the R package RNAontherBENCH (Germain et al., 2016).

Differences in α-diversity and proportion and relative abundance of biosolid-associated viruses were tested for by first generating generalised linear models (GLMs) or beta regression models and applying analysis of deviance tests using the function Anova in R (details in Supplementary Table B.3). Post-hoc Tukey tests were performed using the package emmeans.

β-diversity in viral community structure and AMG enrichment profiles were analysed by non-metric multidimensional scaling (NMDS) on a Bray-Curtis distance matrix using the R function metaMDS. PERMANOVA tests were used to assess differences between time, treatment and blocks. Differences in AMG profiles were similarly assessed in the same way.

Distributions of viral families were analysed by summing CPM values and converting to percentages.

3.6. Results and discussion

3.6.1. Identification and classification of viral contigs and comparison to the long-term effects of biosolids amendment

In this work, we characterised shifts in the soil viral community of soil microcosms amended with conventionally treated biosolids and compared these to unamended soil and biosolids microcosms after incubation at 10°C for 1 year. DNA extracted from virus-like-particles was purified from a total of 24 microcosms (four soil sampling replicates taken from two replicate plots within two blocks of control treatments featured in Chapter 2, two treatments, two time points, plus four biosolids microcosm replicates, two time points) plus one negative extraction control (Fig. 3.2a). Soil samples from each plot were treated separately to preserve the variation in soil viral community from field-scale heterogeneity. Raw sequencing reads were filtered for quality and co-assembled. A combination of VirSorter2 (Guo et al., 2021), VIBRANT (Kieft et al., 2020), DeepVirFinder (Ren et al., 2020) and CheckV (Nayfach et al., 2020) were used (See Materials and Methods) to identify viral contigs to which rarefied read-pairs were mapped and vOTU presence determined using a threshold of 70% coverage horizontal coverage. 13,567 viral operational taxonomic units (vOTUs) were present in at least one sample. Virome summary statistics are provided in Supplementary Fig. B.1.

85.2% of viral contigs were identified by two or more classification tools. This is higher than in Chapter 2 and is potentially due to a combination of greater read depth, a less fragmented assembly in this dataset and differences in library preparation kit. Libraries from Chapter 3 were prepared with the NEBNext Ultra II FS library preparation kit, due to COVID-19 related disruption to sequencing centre capacity, whilst Chapter 2 sequencing libraries were prepared with the Accel NGS 1S+ kit (Swift Biosciences) which is specifically designed to sequence both ssDNA and dsDNA. Library preparation techniques are known to impact viral detection of different viral realms (Roux et al., 2016b) and the longer L50 from this dataset (42 vs 15 kbp) would increase the number of longer dsDNA contigs with sufficient gene content for VirSorter 2 and VIBRANT to detect as viral, whereas DeepVirFinder is known to be more sensitive when screening shorter contigs (Ho et al., 2021). Similar shifts in contig detection are observed when more stringent cutoffs (Fig. 3.2 c), with 67% (vs 49%) being classified as viral by CheckV, although a

greater proportion were discarded at this stage (5% vs 2.5%), possibly due to increased detection of host genes.

Fig. 3.2 d shows an increased proportion of contigs rejected at the stage of vOTU identification. This may be due to changes in VLP purification made to reduce the amount of fragmentation, leading to changes in VLP (i.e., using vortexing rather than sonication, and a lower concentration of DNase during non-VLP DNA digestion). Similarly to Chapter 2, few contigs were eliminated at the vOTU QC, negative control or presence in sample stages (Fig. 3.2 d).

VPF-Class was used to taxonomically classify and predict the hosts of vOTUs at different taxonomic levels, at a reduced percentage than in Chapter 2 (see Fig. 3.3). This is potentially due to factors discussed above leading to improved recovery and identification of distantly related viruses, however VPF-Class similarly relies on gene-level taxonomic classification to classify contigs, suggesting that the hypothesis that more longer contigs leads to greater classification may not hold true for taxonomic classification. Family-level host-prediction was similar between studies (8.8% vs 10%).

3.6.2. Biosolids amendment introduces large numbers of viruses into soil viral communities

Figure 3.4 shows the number of vOTUs shared between different treatments. The largest intersection is those vOTUs only found in control microcosms. Examination of the UpSet plot of individual control microcosms (see Supplementary Figure B.2) reveals that this is due to a large number of vOTUs (2938) unique to a single sample. The second largest intersection represents the core virome shared across all soil mesocosms (highlighted in blue). In total, 978 of 2215 (44%) biosolids-associated vOTUs were also found in soil viromes, with 724 (74%) of these shared with amended soil microcosms at the start of the experiment but not after one year (compared to 221 - 23% shared by biosolids and both time periods, and 33 - 3% shared with amended soil microcosms after one year but not at the start).

It is interesting to note that 51 biosolids-associated vOTUs were also shared with control

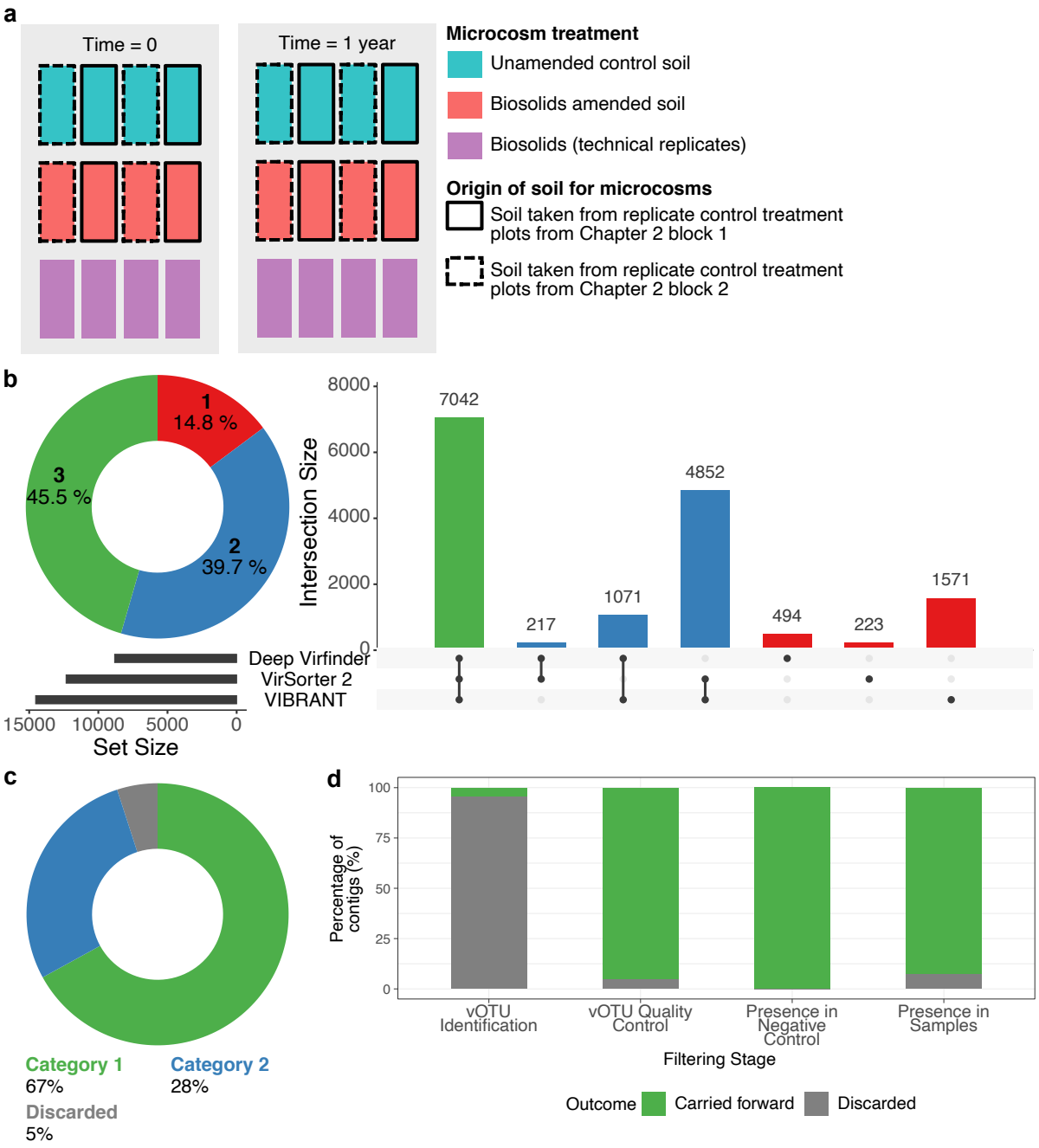


Figure 3.2.: The viromes of 24 microcosms were characterised across three sample types: soil, biosolids-amended soil and biosolids (a). Soil samples were taken from four unamended control plots featured in Chapter 2. An additional negative extraction control of PCR-grade water (not shown) was sequenced for the detection of contaminants. Four contigs were identified as present in the negative control and excluded from subsequent analysis. The majority of viral contigs were identified by at least two tools (b). Further screening (c and d) removed minor quantities of contigs based on contig screening, presence in the negative control library and confrimed presence in individual samples using a horizontal genome coverage threshold of 70 % (d).

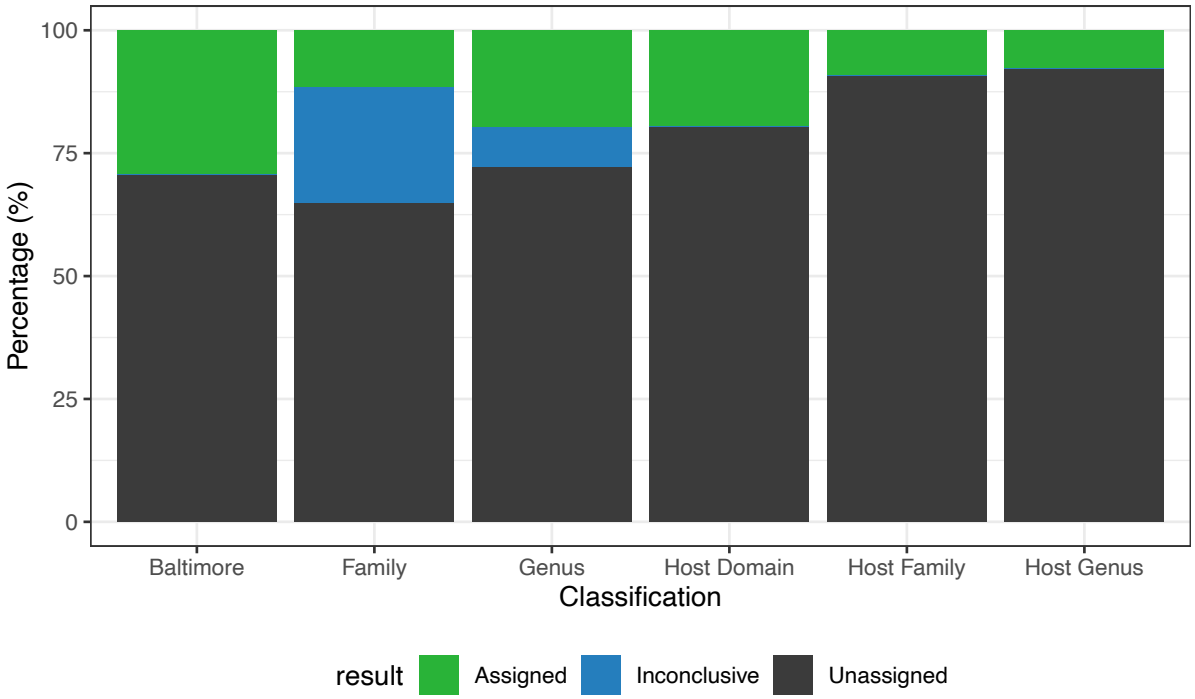


Figure 3.3.: VPF-Class was used for vOTU classification (membership ratio and confidence score of 0.2) and host prediction (membership ratio = 0.3 and confidence score = 0.5) at different taxonomic levels. A lower percentage of vOTUs could be taxonomically classified (11% vs 19% at family-level), although host prediction was marginally higher (8.8% vs 10% at family-level) when compared to results from Chapter 2.

microcosms at the start of the experiment, however these were all located within one microcosm (b1c0 - replicate b, block 1, control treatment, time = 0), which contained over twice the number of vOTUs than other control replicates. These vOTUs formed 2.5% of identified vOTUs with a total relative abundance of 2%, and 48 of 51 having an individual relative abundance of <0.1%. Comparison of the ratios of CPM values between vOTUs in sample b1c0 and other samples revealed that all 51 vOTUs were present at higher relative abundances in biosolids or biosolids amended soil microcosms (see Supplementary Fig. B.3), suggesting that their detection may be false positives, contamination, or rare cases of vOTUs existing in both environments.

α-diversity metrics were calculated from numbers of vOTUs present (richness) or relative abundance measures in counts per million - CPM (see Fig. 3.5 a-c). Generalised linear models were built and evaluated by analysis of deviance to test for the significance of treatment, time and block (see supplementary table Table B.3). Treatment and time were not significant ($X^2 = 0.86$, $p = 0.35$ and $X^2 = 3.4$, $p = 0.67$ respectively) whilst block was significant($X^2 = 7.6$, $p = 0.0057$), although this is likely to have been heavily influenced by two outliers within the Amended-1 and Control-0 viromes. Treatment and block effects on Shannon index were significant ($X^2 = 7.2$, $p = 0.0075$ and $X^2 = 6.4$, $p = 0.011$ respectively) whilst time was not ($X^2 = 0.47$, $p = 0.49$). Simpson Index showed significant effects from both treatment ($X^2 = 15.8$, $p < 0.0001$) and time ($X^2 = 7.3$, $p = 0.007$) but not block ($X^2 = 1.7$, $p = 0.19$). Pairwise comparisons of treatment and time contrasts (Tukey tests) showed that Simpson Index was significantly different between the start and end of the experiment in amended but not control treatments ($p = 0.037$ and $p = 0.063$ respectively). Amended and control treatments were significantly different from each other at the start and end of the experiment ($p = 0.0006$ and 0.0023, respectively) although the amended microcosms at the end of the experiment had no significant difference with control microcosms at the start ($p = 0.74$). No significant pairwise comparisons were found between Shannon index values or number of vOTUs ($p > 0.05$).

The decrease in Simpson Index due to treatment at the start of the experiment is likely to be due to the introduction of a limited number of highly abundant biosolids-associated vOTUs, as Simpson Index is heavily influenced by the most abundant species (DeJong,

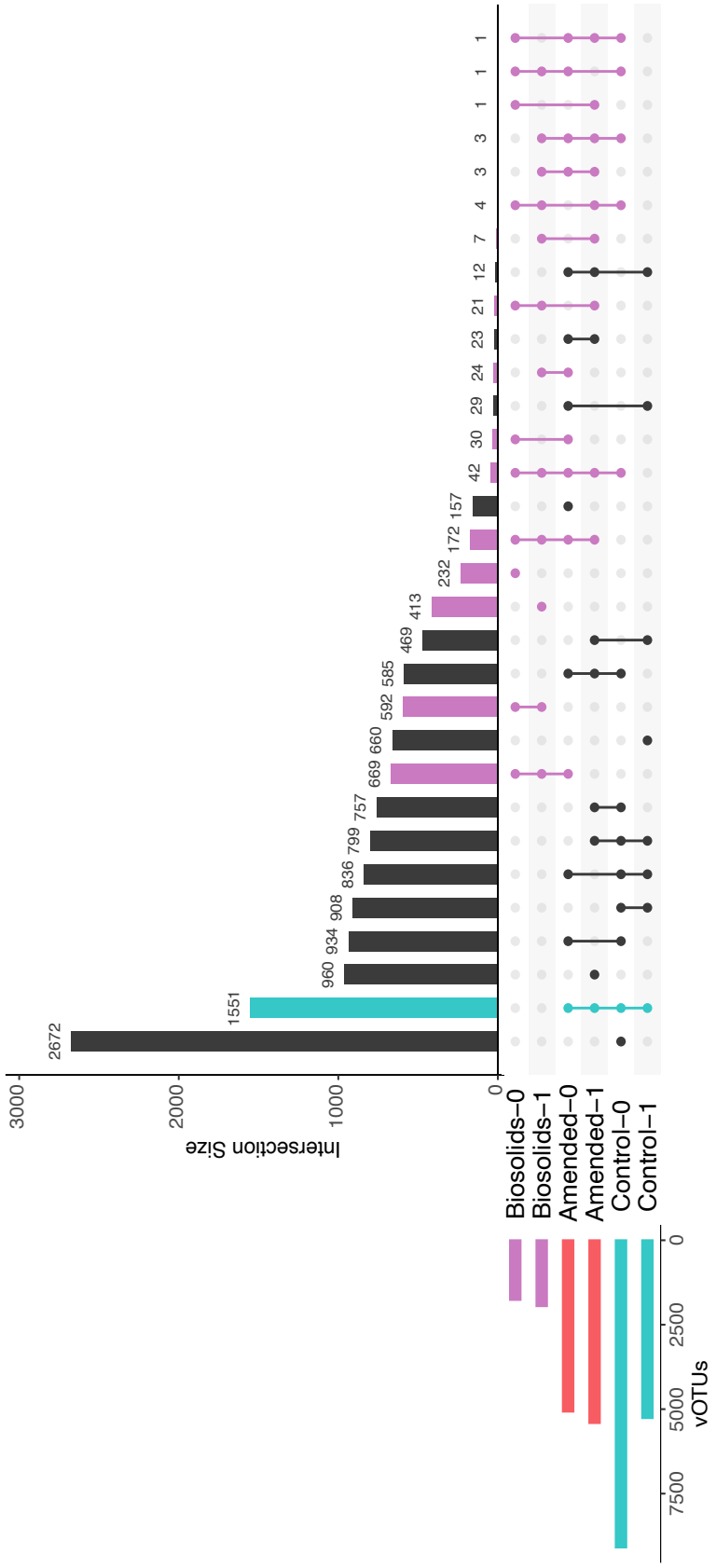


Figure 3.4.: UpSet plot of the number of vOTUs shared by each sample type. The core soil virome is highlighted in blue whilst biosolids-associated vOTUs are highlighted in purple.

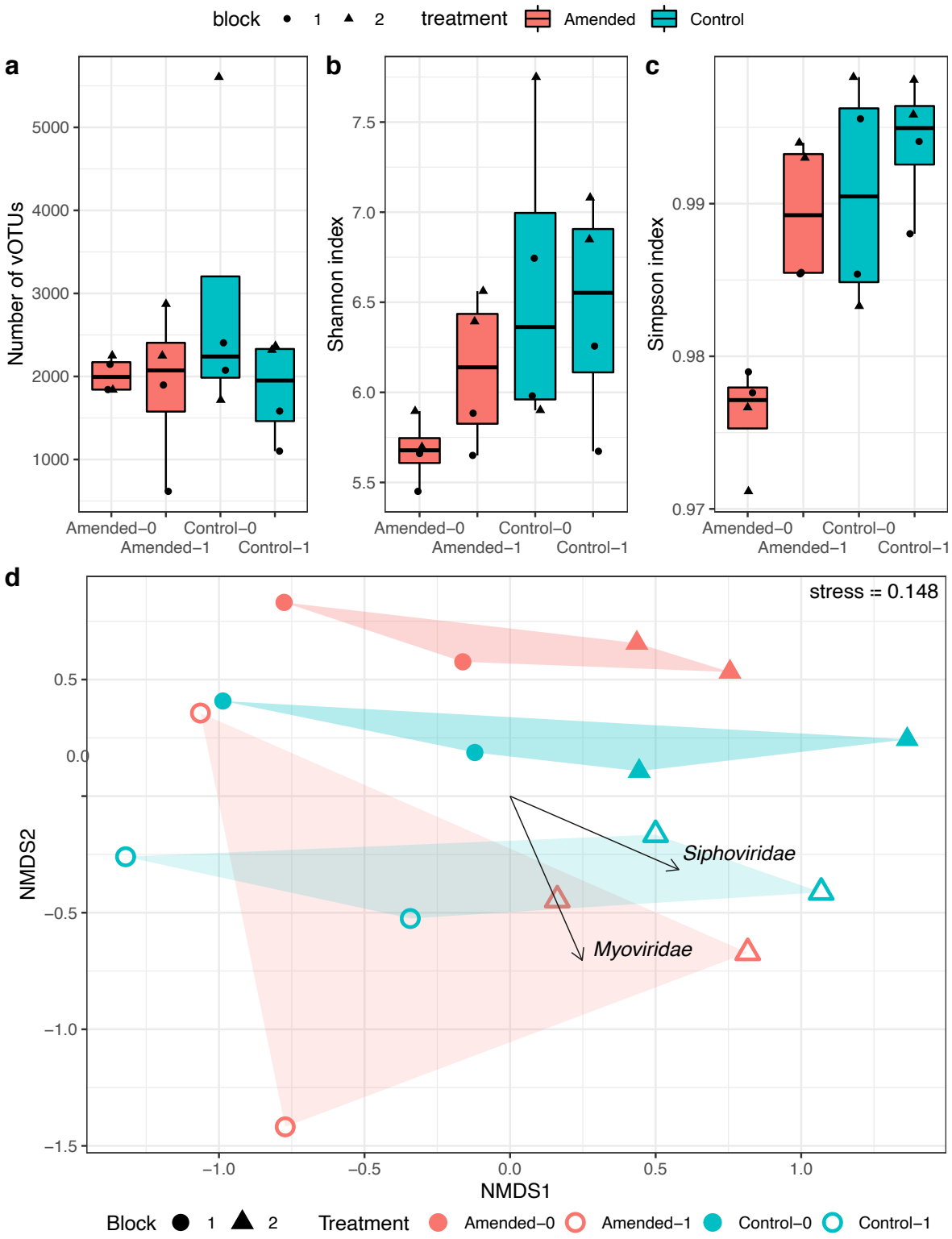


Figure 3.5.: α-diversity of amended and control microcosms at the start of the experiment and after one year (a - number of vOTUs, b - Shannon index, c - Simpson index). Panel d shows β-diversity of amended and control microcosms at the start of the experiment and after one year, displayed as an NMDS plot. Differences between communities were tested for using PERMANOVA (treatment - $R^2 = 0.12954$, $p = 0.0054$, block - $R^2 = 0.20$, $p = 0.0001$, time - $R^2 = 0.12$, $p = 0.013$, interaction effects $p>0.05$).

1975). The effect of the dominance of biosolids-associated vOTUs in amended soil microcosms reduces over time. Statistically, this effect is still present after one year, however it is difficult to say whether this is real or artifactual, due to similarities between Amended-1 and Control-0 sample types. The lack of significant differences between Shannon Index values indicates that overall community diversity remains unaltered by time or treatment. The range of Shannon and Simpson index values are towards the high end of the range previously reported (Chen et al., 2021; Lee et al., 2022) but higher than those reported in Chapter 2, most likely due to increased sequencing depth recovering a higher number of vOTUs.

Shifts in β-diversity were examined by non-metric multidimensional scaling (see Fig. 3.5 d) and statistically significant differences between populations tested for by PERMANOVA. Overall, treatment, time and block generated statistically significant differences, however there were no significant interaction effects between the three factors (see Table 3.1). PERMANOVA tests were repeated for time = 0 years and time = 1 year. At the start of the experiment, treatment and block caused significant differences in virome community structure but after one year, block remained significant, but treatment did not. Similarly to the overall analysis, no significant interaction effects existed ($p > 0.05$). These effects can also be seen graphically in Fig. 3.5 d, where there is clear separation of the amended and control communities at the start of the experiment (solid symbols) but substantial overlap after one year (open symbols). The 0-year communities are both distinctly separate from the 1-year communities, regardless of treatment, and they clearly separate by block (circles = block 1, triangles = block 2), with this effect being consistent, regardless of treatment and time.

Table 3.1.: PERMANOVA results for differences in overall β-diversity and at the start and end of the experiment

Factor	All soil microcosms	Soil microcosms at t=0	Soil microcosms at t=1 year
Treatment	$R^2 = 0.130$ p = 0.0054	$R^2 = 0.335$ p = 0.0038	$R^2 = 0.164$ p = 0.171
Block	$R^2 = 0.204$ p = 0.0001	$R^2 = 0.256$ p = 0.0465	$R^2 = 0.281$ p = 0.0082

Factor	All soil microcosms	Soil microcosms at t=0	Soil microcosms at t=1 year
Time	$R^2 = 0.117$ p = 0.0133		

This further demonstrates that whilst biosolids amendment has a substantial effect on the soil viral community at the time of amendment, overall community structure returns to one similar to unamended soil after one year, but remains influenced by spatial heterogeneity. The impact of spatial heterogeneity has been similarly observed in other studies, suggesting that soil viral communities are highly variable at the vOTU level (Santos-Medellin et al., 2021). It is also possible that this spatial variation is an artefact of sampling intensity, as many viromics studies who report species accumulation curves indicate undersaturation (Adriaenssens et al., 2017; Lee et al., 2022; Santos-Medellin et al., 2021). A similar trend can be seen in this study (see Supplementary Fig. B.7), indicating that increased sequencing effort, or analysing data at a higher taxonomic level may improve our understanding of spatial heterogeneity in soil viral ecology, however these solutions rely on the reduction of sequencing costs, or improvements in viral taxonomic classification algorithms.

3.6.3. Caudoviricetes dominate the viromes of soils and biosolids

The majority of viruses detected in all three treatments belong to the *Caudoviricetes* class of bacteriophages (*Siphoviridae*, *Myoviridae* and *Podoviridae* - see Fig. 3.6. Compared to Chapter 2, the relative abundance of *Phycodnaviridae* and *Microviridae* are substantially lower and these families are not detected in any biosolids samples (purple). This may be due to changes in the library preparation and sample processing protocols between Chapters 2 and 3. To identify specific human pathogens, all contigs >=1,500 bp in length were compared to the refseq viral database (BLASTX, e-value cutoff of 10^{-5}) with no pathogenic viruses identified.

Rare families include viruses of archaea, amoeba, arthropods and potentially vertebrates. It is also interesting to note the presence of *Lavidaviridae* as the virions of these virophages

would be capable of being detected by viromics using a size selection of 0.22 μm, but their nucleocytoplasmic large DNA virus (NCLDV) host *Mimiviridae*, with a typical diameter of 0.3-0.7 μm, would be excluded (Claverie and Abergel, 2018; Fischer, 2021).

Another family of NCLDVs, the familiy *Marseilleviridae*, possess virion diameters of approximately 0.25 μm, may be detected by viromics (Colson et al., 2013). In this study, a single, yet abundant 11Kb *Marseilleviridae* vOTU (k127_1331608) was detected in the majority of biosolids and all biosolids-amended soil microcosms at the start of the experiment, but sporadically after one year. Although relative abundance of this vOTU was low (<1%), horizontal coverage ranged from 70-100% in samples where it was identified as present, with average vertical coverage ranging from 2.0-13.2 fold, demonstrating that the vOTU's presence is likely to be genuine. However, the average genome length of *Marseilleviridae* is 300 Kb (Aherfi et al., 2014), implying that this contig is either a small fragment of a much larger unassembled viral genome, or has been potentially misassigned. This case highlights that, given the low number of individual contigs detected, caution should be used in avoiding the over-interpretation of the presence of rare taxa, due to possible inaccuracies in vOTU taxonomic assignment. All rare taxa displayed in Fig. 3.6 b had relative abundancies of <1%, with the exception of *Papillomaviridae* in sample a1b1 (replicate a, block 1, amended, 1-year) at 5.9% of total assigned CPM values. Further work is needed to more closely examine the presence, and ecological role of non-bacteriophage soil viruses, but similarly to the issue of spatial heterogeneity mentioned above, this will be dependent on improved sequencing technology and increased sensitivity of virus classification algorithms in detecting non-bacteriophage viruses.

3.6.4. Host-prediction reveals the contribution of biosolids-associated *Bacteroidetes* viruses to viral community shifts

Similarly to Chapter 2, the dominant host phyla are *Actinobacteria*, *Bacteroidetes*, *Firmicutes* and *Proteobacteria* (see Fig. 3.7). However, a pattern can be observed in that viruses of *Bacteroidetes* are highly abundant in the biosolids microcosms (purple) and amended soil microcosms at the start of the experiment (red closed circles). The converse pattern is observable in viruses of *Actinobacteria*, with their relative abundance being lower in microcosms with more biosolids-associated viruses. Both *Actinobacteria* and *Bacteroidetes*

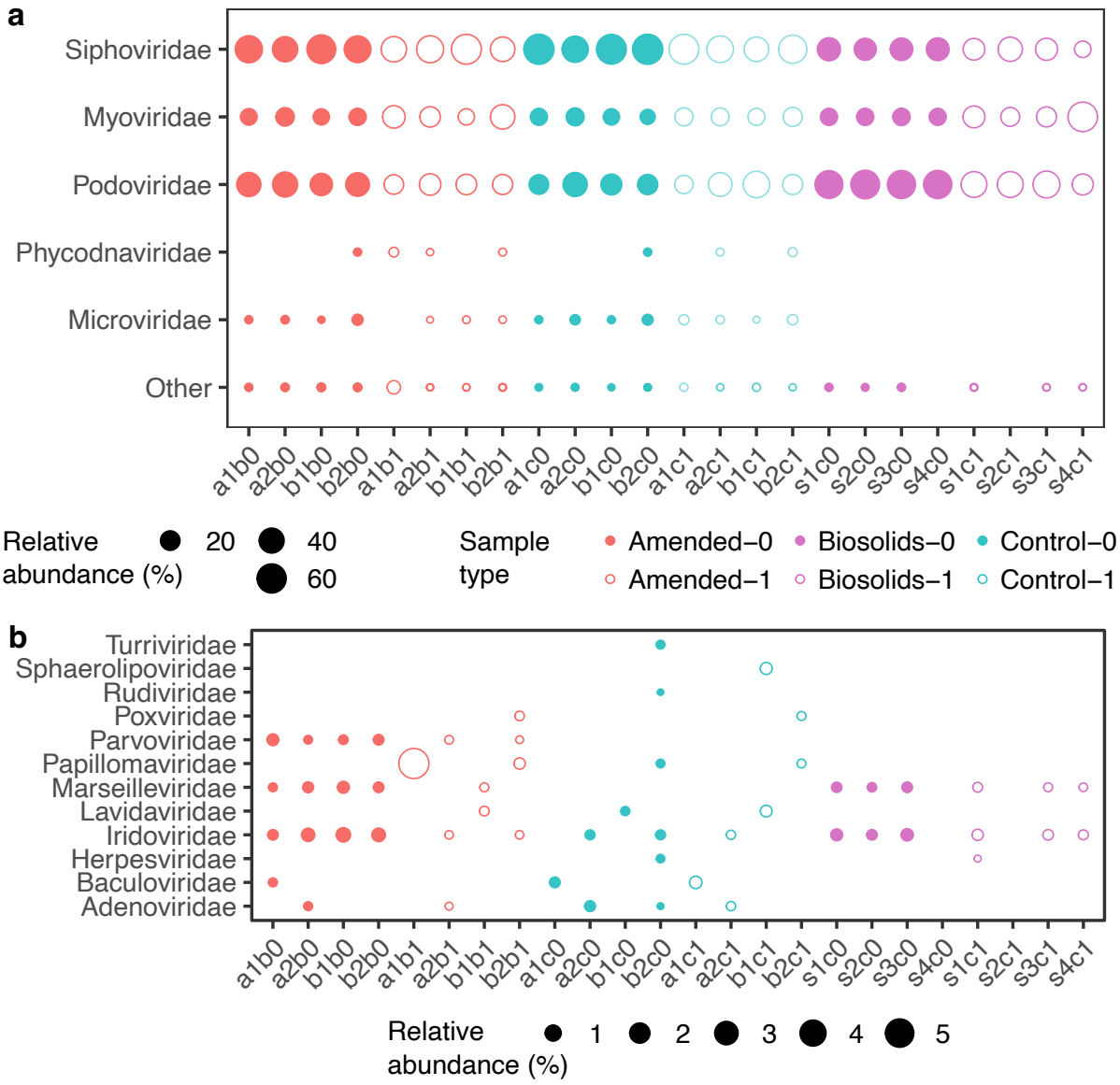


Figure 3.6.: Relative abundance of family level taxonomically assigned vOTUs divided into those with high relative abundance anod those of interest from Chapter 2 (a) and rare taxa (b). Relative abundance is calculated as the percentage of total CPM values of taxonomically assigned vOTUs. Closed circles represent microcosms at the start of the experiment and open circles represent microcosms after one year. Colour indicates treatment.

are key saprophytic components of the soil microbiome (Bhatti et al., 2017; Larsbrink et al., 2020) and so a shift in the balance of bacteriophages of these two phyla could cause changes to soil carbon and other nutrient cycling. This would likely require cross-species host ranges as for the introduced phages to have an impact on the soil microbiome, they would have to be able to infect, replicate within, and kill soil-borne hosts. Instead, the proportion of these bacteriophages returns to levels similar to those in control soils, suggesting that any changes are short-lived, and that a more likely scenario is that introduced phages decay over time as they are unlikely to be able to find a susceptible host within which to replicate.

A number of other rarer hosts were identified, including both *Crenarchaeota* and *Euryarchaeota*, as well as several phyla of eukaryotes. Although all samples indicate the presence of viruses of *Chordata*, these viruses were taxonomically assigned as non-vertebrate infecting NCLDVs and VPF-Class can face challenges when assigning hosts to eukaryote-infecting viruses (Pons et al., 2021), therefore some of the eykaryotic host-assignments should be treated with some caution.

3.6.5. Biosolids-associated vOTUs reduce substantially reduce in virome proportion and relative abundance after one year

3040 vOTUs from the biosolids viromes were also detected in the amended soil viromes. These biosolids-associated vOTUs formed 31.5% ±4.67 of the total number of vOTUs within the amended microcosms at the start of the experiment (Fig. 3.8 a). This quantity reduced 4-fold to 7.71% ±1.70 after one year. A similar pattern can be observed in the relative abundance of biosolids-associated vOTUs, where they form 56.3% ±0.914 of the total virome in the amended soil microcosms at the start of the experiment and this reduced 7.6-fold to 7.43% ±0.759 after one year (Fig. 3.8 b).

Decay rates of individual viral species are known to vary substantially (Schwarz et al., 2014) and maximum persistences of pathogenic viruses estimated at 3-6 months (Gerba and Smith, 2005). The data from this study indicate that this may be an underestimate, however it is important to emphasise that detection of biosolids-associated viruses via molecular methods is not indicative of their infectivity (Wong et al., 2010). Further use

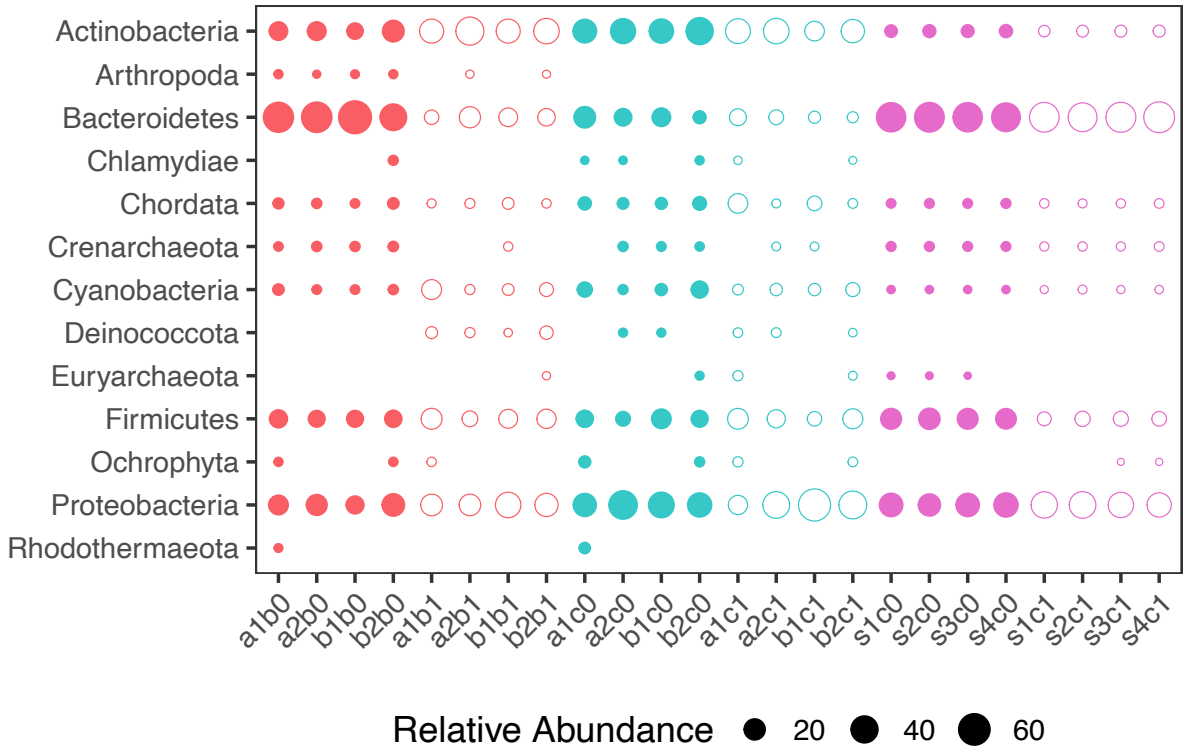


Figure 3.7.: Relative abundance of vOTUs with predicted hosts summarised at the family level. Relative abundance is calculated as the percentage of total CPM values of vOTUs with predicted hosts. Closed circles represent microcosms at the start of the experiment and open circles represent microcosms after one year. Colour indicates treatment.

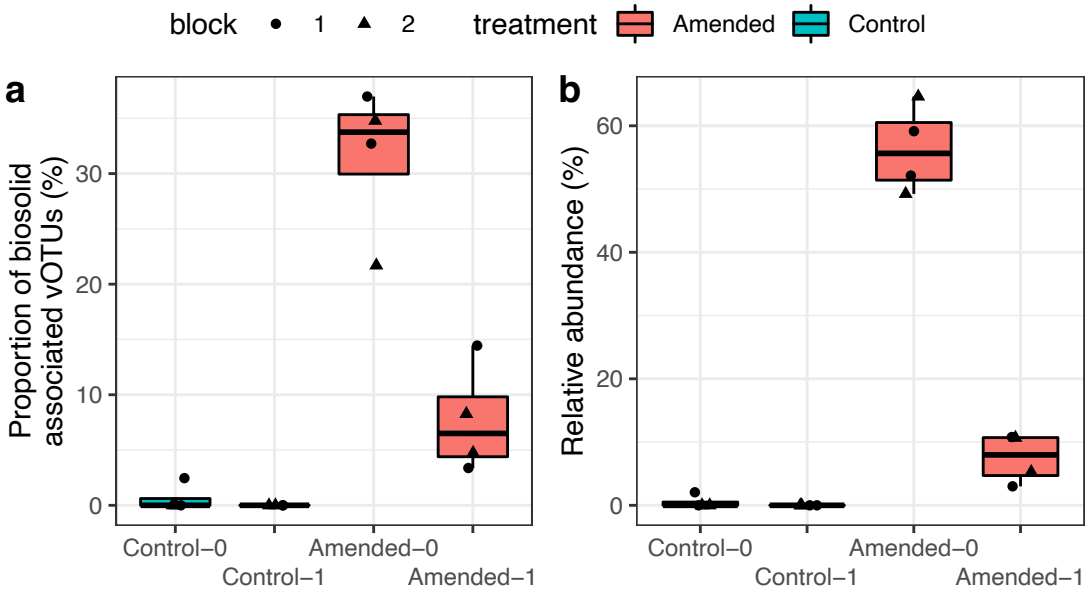


Figure 3.8.: (a) Proportion of biosolids-associated vOTUs in amended and control control treatment viromes at the start of the experiment and after one year. Treatment but not block was found to have a statistically significant effect for amended viromes (Beta regression, analysis of deviance - $X^2 = 38.0$, $p < 0.0001$ and $X^2 = 1.20$, $p = 0.274$ respectively) (b) Relative abundance of biosolids-associated vOTUs in amended and control treatment viromes at the start of the experiment and after one year. Treatment, but not block, had a statistically significant effect (Beta regression, analysis of deviance - $X^2 = 104.3$, $p < 0.0001$ and $X^2 = 0.21$, $p = 0.647$, respectively).

of assays that couple infectivity assays to sensitive molecular techniques will be able to assess the impact of time, land management strategies and environmental conditions on the reduction in viral risk from the use of biosolids as agricultural fertilisers (Farkas et al., 2020b).

3.6.6. Auxiliary metabolic genes indicate diverse soil microcosm viral community functional capacity

5,335 auxiliary metabolic genes (AMGs) were identified in 4,527 vOTUs by VIBRANT covering a large variety of pathways (see Fig. 3.9 a). The most common AMGs belonged to cofactor/ vitamin, and carbohydrate metabolisms, suggesting that some soil viruses may enhance their microbial hosts' ability to harness carbon sources and utilise cofactors required for enzyme activity. Comparison of different treatments (Fig. 3.9 b) indicates a high degree of overlap between soil microcosms, with the biosolids technical replicates separating from the majority of samples in the top left region of the ordination plot, and the Amended-0 treatment located between the two. A single Amended-1 replicate lies at the bottom left corner of the ordination plot (a1b1), with their location potentially being heavily influenced by a lower diversity of metabolisms being detected.

Bacteriophages are known to have the potential to influence microbiome functional capacity in carbon, nitrogen, sulphur and phosphorus metabolism (Han et al., 2022; Monier et al., 2017; Roux et al., 2016a; Trubl et al., 2018). Whilst the results indicate that biosolids-amendment does influence the AMG profile, and therefore the functional capacity of soil microcosm viral communities, it is not possible to say whether these genes are being actively expressed, or if microbial community function has been impacted. Given the reduction in biosolids-associated viruses over the course of the experiment, it is unlikely that biosolids-associated AMGs are utilised by their microbial hosts in the soil environment. The shift in the amended soil microcosm viral community towards an unamended-like structure, indicates that any indirect effects of biosolids amendment on viral community functional capacity are also likely to be short-lived.

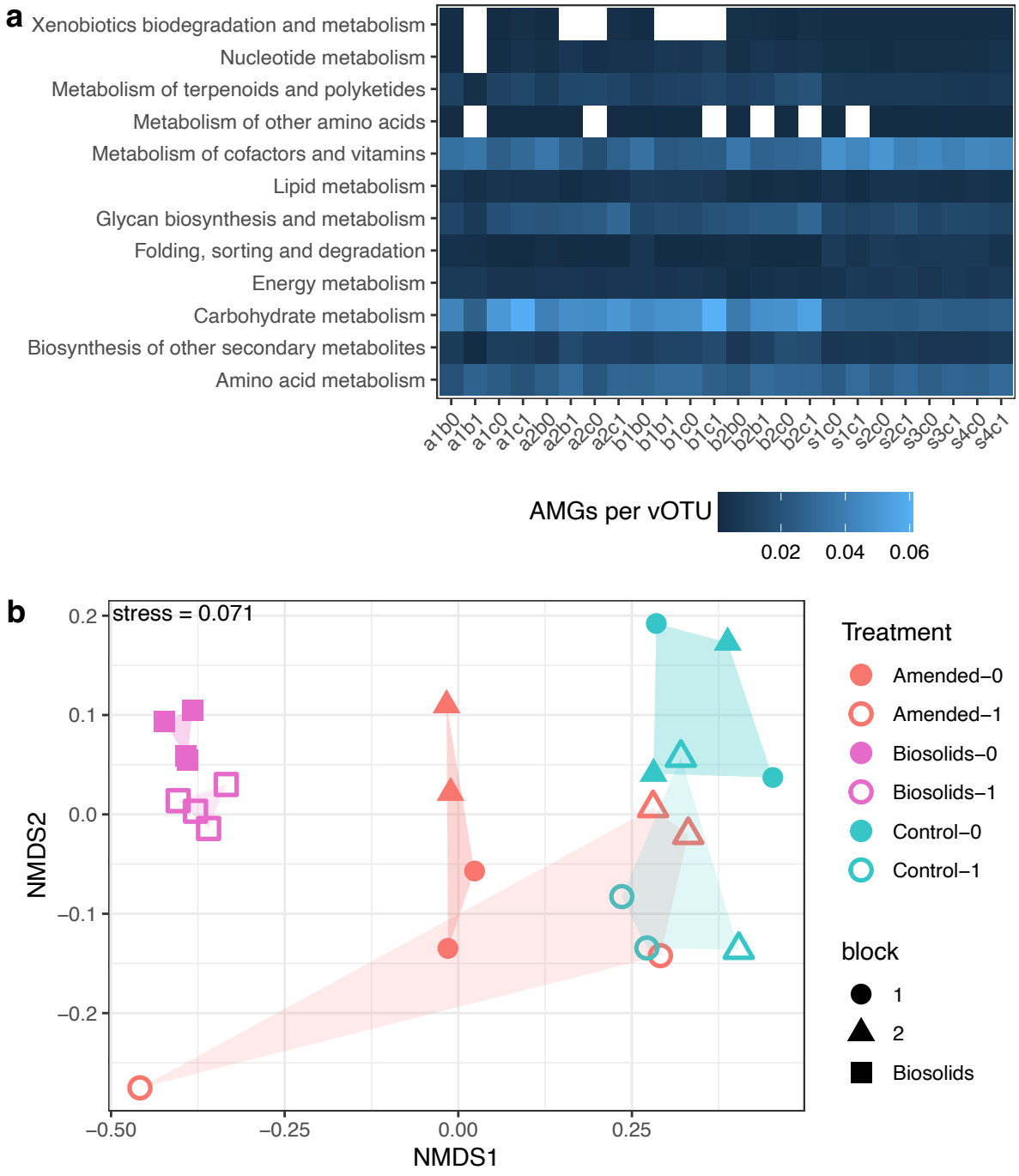


Figure 3.9.: (a) heatmap and (b) NMDS plot (Bray-Curtis) of the number of auxillary metabolic genes identified, normalised by the number of vOTUs for each virome.

3.6.7. Anti-microbial resistance genes are low in abundance and naturally occur in amended and unamended soil microcosms

A total of 14 vOTUs were found to carry putative antimicrobial resistance genes (ARGs - see Fig. 3.10) of which nine were found in soil microcosms, and eight were found in biosolids microcosms. Whilst Fig. 3.10 suggests that ARG-carrying vOTUs decreased over the course of the experiment, this trend could be a random occurence, as a single additional vOTU could substantially alter this pattern. Differences caused by treatment and time in the number, proportion of vOTUs and relative abundance of AMR genes were not significant ($p > 0.05$, GLM (quasibinomial distribution) + analysis of deviance).

Similarly, the presence of biosolids-associated ARG-carrying vOTUs only in amended soil microcosms at the start of the experiment suggests that biosolids amendment can import ARGs of viral origin, but further work would be needed to demonstrate that these genes are genuine ARGs. Whilst viral ARGs have been previously detected in biosolids-amended field trials (Chen et al., 2021), this study did not explore if those ARGs were biosolids-associated. This study demonstrates that viral ARGs could be a naturally occurring component of the soil viral communty's functional capacity. Further *in vitro* work would be needed to demonstrate that the genes detected here are produce antibiotic resistance phenotypes because, as discussed in Chapter 2, the detection thresholds used are a trade-off between sensitivity and precision (Enault et al., 2016).

3.6.8. Conclusions

Community-level soil viral ecology remains a substantially underexplored and neglected area of interest (Pratama and Elsas, 2018), particularly when using viromics-based studies to impact land-management policy. Our understanding of the viral community dynamics of the wastewater treatment process are more advanced, but often (necessarily) focus on the dissemination of viruses hazardous to human, plant and animal health, particularly through the discharge and use of treated wastewater effluent (Adriaenssens et al., 2021; Bačnik et al., 2020). Biosolids, the solid byproduct of the wastewater treatment process, have the potential to disseminate viruses from the wastewater treatment process to land, where they could re-enter the food-production system, although this risk has been estimated to be <1:10,000 per year if solid waste is treated rigorously (Gerba et al.,

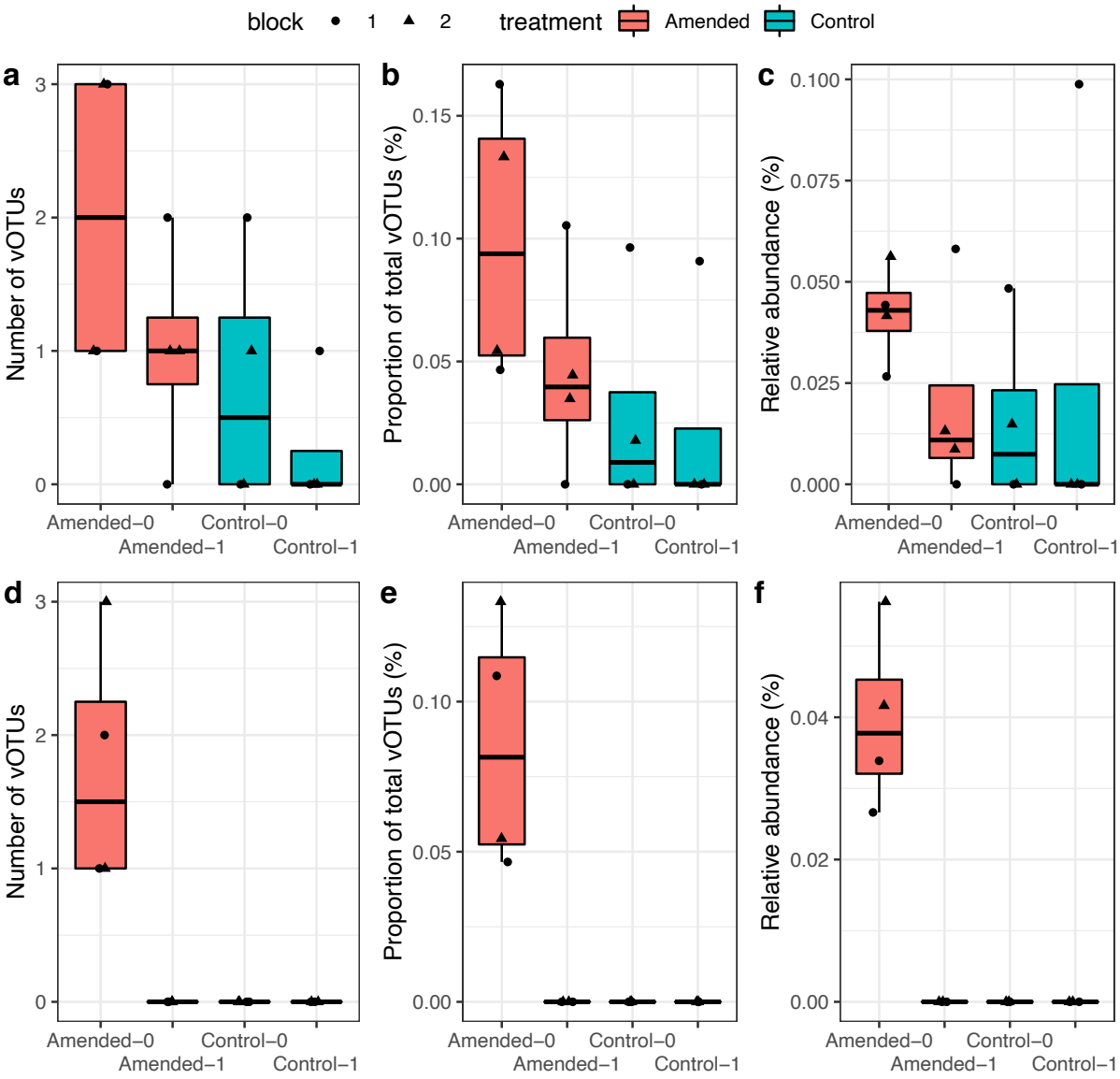


Figure 3.10.: (a-c) Number, proportion of total vOTUs and relative abundance of all ARG-carrying vOTUs. (d-f) Number, proportion of total vOTUs and relative abundance of biosolids-associated ARG-carrying vOTUs. These were only found in Amended-0 microcosms.

2002).

This study characterised the viromes of biosolids amended soil microcosms to monitor the decay of biosolids-associated viruses and shifts in the soil viral community over one year, and compared these results to unamended control soils. We aimed to examine (1) the diversity of biosolids-associated viruses imported into the soil viral community after amendment, (2) the level of persistence of biosolids-associated virus after 1 year under controlled conditions and (3) the effects of biosolids amendment on the soil viral community over this time period.

Biosolids amendment introduces a large quantity of viruses to the soil virome, but the majority of these viruses decay after one year under controlled conditions. The remaining viruses form a 7.6-fold reduced relative abundance of the total soil viral community (from 56% to 7.4%). The majority of vOTUs that could be taxonomically assigned belonged to the *Caudoviricetes* class of bacteriophages. No difference in soil viral community α- or β-diversity was observable after 1 year, indicating a limited impact on the overall soil viral community structure.

Soil viral communities are recognised as a natural reservoir of ARGs, with biosolids amendment enriching the ARG content of purified bacterial but not viral communities and biosolids-associated bacteriophages can enhance antimicrobial resistance in host bacteria through specific or generalised transduction (Joseph et al., 2015). The presence of antimicrobial resistance genes was demonstrated to be relatively rare within the soil viral communities in this study, with biosolids-associated viral ARGs only occurring in amended soil microcosms at the start of the experiment. Biosolids amendment is known to increase the quantity and diversity of ARGs (Xie et al., 2016) but this study suggests that bacteriophages may play a minor, and potentially short-lived role in their dissemination. Further work is needed to understand the role of biosolids-associated bacteriophages in the dissemination of ARGs in the environment, as the current evidence is unclear on how biosolids compare to other organic manures, and the causative nature of increased ARG presence after amendment.

This study represents a baseline assessment of the impact of a single amendment of conventionally treated biosolids on the soil viral community and current policy on the monitoring

of soils pre and post biosolids amendment varies globally, with reduced restrictions on usage of biosolids treated to higher standards (Assured Biosolids Limited, 2020). Similarly to Chapter 2, no human pathogens were identified in the treated biosolids or any of the soil microcosms. Environmental monitoring is increasingly making use of environmental DNA for the surveying of wildlife biology, wastewater based epidemiology and soil health (Deiner et al., 2017; Fierer et al., 2021; Lorenzo and Picó, 2019) and there is significant potential to also apply these approaches for the detection and assessment of the impacts of organic manure usage and broader land management practices on the introduction and persistence of micro-organisms, viruses and other genetic markers that could be used to identify the source of wastewater/ biosolids contamination.

3.7. Data and code availability

Sequencing read files and vOTU contigs analysed in this study will be deposited with the NCBI and made publicly available on publication in a peer reviewed journal. Code related to this chapter will be made publicly available via GitHub at https://github.com /LSHillary/StDnaVirome upon publication.

3.8. Acknowledgements

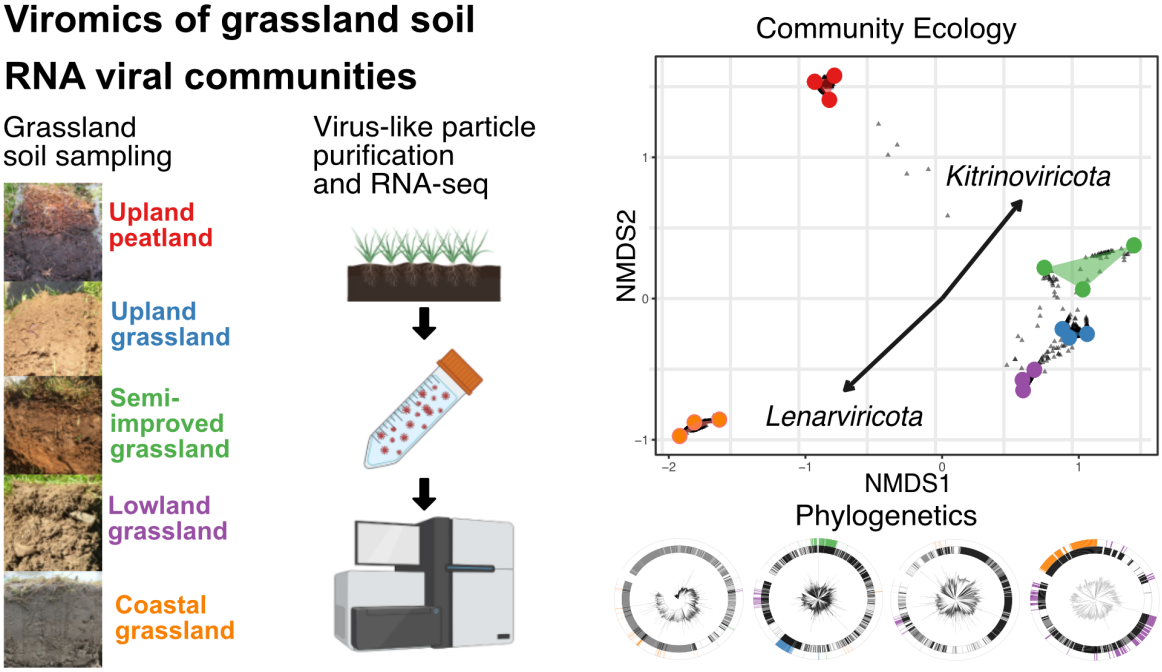
LSH was supported by a Soils Training and Research Studentship (STARS) grant from the Biotechnology and Biological Sciences Research Council (BBSRC) and Natural Environment Research Council (NE/M009106/1) and by Welsh Water. DNA sequencing was carried out by the Centre for Genomics Research at the University of Liverpool. Data analysis utilised high performance computing resources from Supercomputing Wales. The authors would like to thank Javier Hernandez for assistance with sample collection, Welsh Water for providing additional funding and samples of conventionally treated biosolids for use in this work, and David Fidler, Marine Cambon, Jim Downie, Dave Chadwick and Robert Griffiths for comments/ advice on data analysis and visualisation.

4

Diverse soil RNA viral communities have the potential to influence grassland ecosystems across multiple trophic levels

---

4.1. Graphical abstract



4.2. Foreward

This chapter has been published in ISME Communications:

> Hillary, L.S. et al., 2022. RNA-viromics reveals diverse communities of soil RNA viruses with the potential to affect grassland ecosystems across multiple trophic levels. ISME Commun. 2, 34. doi: 10.1038/s43705-022-00110-x

LSH, EMA, DLJ and JMD conceived the study and acquired funding. LSH carried out the sample collection and viral RNA extraction and led the data analysis and preparation

of the initial draft of the manuscript. All authors contributed to the final version of the article.

4.3. Abstract

The distribution and diversity of RNA viruses in soil ecosystems are largely unknown, despite their significant impact on public health, ecosystem functions, and food security. Here, we characterise soil RNA viral communities along an altitudinal productivity gradient of peat, managed grassland and coastal soils. We identified 3,462 viral contigs in RNA viromes from purified virus-like-particles in five soil-types and assessed their spatial distribution, phylogenetic diversity and potential host ranges. Soil types exhibited minimal similarity in viral community composition, but with >10-fold more viral contigs shared between managed grassland soils when compared with peat or coastal soils. Phylogenetic analyses predicted soil RNA viral communities are formed from viruses of bacteria, plants, fungi, vertebrates and invertebrates, with only 12% of viral contigs belonging to the bacteria-infecting *Leviviricetes* class. 11% of viral contigs were found to be most closely related to members of the *Ourmiavirus* genus, suggesting that members of this clade of plant viruses may be far more widely distributed and diverse than previously thought. These results contrast with soil DNA viromes which are typically dominated by bacteriophages. RNA viral communities therefore have the potential to exert influence on inter-kingdom interactions across terrestrial biomes.

4.4. Introduction

Viruses are the most common and diverse biological entities on Earth (Paez-Espino et al., 2016) and can exert significant influence on their hosts. In addition to their ecological functions, viruses are key influencers of public health and food security, causing 47% and 44% of plant and human emerging infectious diseases, respectively (Anderson et al., 2004; Taylor et al., 2001). The current COVID-19 pandemic highlights the critical importance of understanding the role of viruses in the environment, and how natural and anthropogenic ecosystems can function as sources of novel zoonotic infections. Grassland ecosystems form 30-40% (White et al., 2000) of total land cover and provide essential ecosystem services, including food production, flood mitigation and carbon storage (White et al., 2000; Zhao et al., 2020). Within these, and other terrestrial ecosystems, DNA viruses are known to play essential roles in microbial community dynamics and carbon biogeochemical cycling (Adriaenssens et al., 2017; Emerson et al., 2018; Jin et al., 2019; Trubl et al., 2018; Zablocki et al., 2017), yet the role of viruses within these critical ecosystems remains undercharacterised (Williamson et al., 2017) and in particular, our knowledge of soil RNA viruses is significantly limited (Starr et al., 2019; Wu et al., 2021). To date, soil viral ecology has focused almost exclusively on DNA viruses of bacteria and archaea. In contrast, marine DNA and RNA viruses have been characterised on an ocean-wide scale (Hurwitz and Sullivan, 2013), and the significant level of diversity observed suggests that the global virome could be the largest reservoir of genetic diversity on the planet (Breitbart et al., 2018).

The vast majority of known RNA viruses lie within the realm *Riboviria* and possess a universally conserved RNA dependent RNA polymerase (RdRP) gene (Wolf et al., 2018). This gene can be used to identify viral RNA genomes and genome fragments from large scale metatranscriptome datasets. A number of recent studies have used this strategy to dramatically increase the number of known RNA viral sequences (Callanan et al., 2020; Shi et al., 2016; Starr et al., 2019) allowing the construction of a broad global viral taxonomy (Koonin et al., 2020). Difficulties in generating and analysing environmental RNA viral sequencing data remain, largely due to experimental challenges of extracting sufficient viral RNA from environmental samples, and in computationally identifying RNA viral genome fragments in large metatranscriptome datasets (Cobbin et al., 2021).

The detection of RNA viral genome fragments in soils can be enhanced by enriching, concentrating and purifying virus like particles (VLPs) from the soil matrix. Viromics uses size selection to selectively enrich for VLPs in environmental samples, ensuring that they represent a greater proportion of the data obtained from high throughput sequencing (Trubl et al., 2020). This can significantly improve the quality and quantity of viral genomes recovered from soils over bulk-soil metagenomes and metatranscriptomes (Santos-Medellin et al., 2021). Viral RNA for use in viromics studies can be readily extracted from water, sewage and sediments (Adriaenssens et al., 2018; Bibby and Peccia, 2013; Culley, 2018) and it is possible to detect RNA viral sequences in bulk soil and rhizosphere metatranscriptomes (Starr et al., 2019; Wu et al., 2021) but to date, and to the best of our knowledge, there has been no published attempt to apply viromics to the study of RNA viruses in soil.

Here, we use viromics to characterise the soil RNA viral communities of five contrasting soil types along a typical temperate oceanic grassland altitudinal productivity gradient (Withers et al., 2020). We identified RdRP containing viral contigs and examined their distribution across different soil types at both a viral contig and phylum level. We then used phylogenetic analyses to place these viral contigs within phylogenetic trees of known viruses, and compared them to viral contigs detected by a previous mesocosm bulk soil metatranscriptomics study (Starr et al., 2019). Our findings demonstrate that soils represent a significant reservoir of viral diversity that have the potential to impact not just the soil microbial community, but also across multi-kingdom host ranges.

4.5. Materials and methods

4.5.1. Field site description, soil sampling and processing

Five sites along an altitudinal gradient at Henfaes Research Centre, Abergyngregyn, Wales were sampled on 31st October 2018. Three adjacent 5×5 m plots were marked out at each site and approximately 2 kg of soil was extracted from each site between 0-10 cm depth using a 3 cm diameter screw auger with evenly spaced sampling within each grid. The augur was cleaned with 1% Virkon disinfectant and a dummy soil core taken and discarded outside of the sampling area prior to sampling each plot. Soil from each plot was sieved

separately to 2 mm and stored in 100 g aliquots at -80 °C prior to RNA extraction.

4.5.2. Viral RNA enrichment and extraction

Virus-like particle extraction was based on protocols developed by Trubl *et al.* (2016) and Adriaenssens *et al.* (2018). A total of 16 samples, three per site and one 100 mL PCR-grade water negative control extraction were processed separately. 100 g of soil per sample was thawed and evenly divided into eight 50 mL centrifuge tubes (12.5 g of soil per tube, hereon referred to as subsamples). Each subsample was suspended in 37.5 mL of amended potassium citrate buffer (1% potassium citrate, 10% phosphate buffered saline (PBS), 5 mM ethylenediaminetetraacetic acid (EDTA), and 150 mM magnesium sulphate ($MgSO_4$), 300 mL total volume per sample). Each subsample was subjected to 30 seconds manual shaking followed by 60 seconds vortexing at maximum speed. After physical disruption, subsamples were placed on ice on an orbital shaker and shaken at 300 rpm for 30 minutes and then centrifuged for 30 minutes at 3,000 × g, 4 °C. Supernatants were removed to new centrifuge tubes and polyethylene glycol, (PEG - 6,000 MW) and sodium chloride (NaCl) were added to 15% (w/v) and 2% (w/v) respectively to precipitate VLPs overnight at 4 °C. Precipitates were recovered by centrifuging tubes for 80 minutes at 2,500 × g, 4 °C and discarding the supernatants. The eight subsample pellets from each 100 g soil sample were recombined by resuspending them in a total volume of 10 mL of Tris buffer (10 mM Tris-HCl, 10 mM $MgSO_4$, 150 mM NaCl, pH 7.5). Recombined samples were filtered through sterile polyethersulfone 0.22 μm pore size syringe filters and concentrated to <600 μL using Amicon Ultra-15 centrifugal filter units (50 kDa MWCO, Merck) prior to RNA extraction.

All RNA extraction protocols were used according to the manufacturer's instructions except where specified. Nucleic acids were extracted using the AllPrep PowerViral DNA/RNA extraction kit (Qiagen) with the addition of 10 μL/ mL 2-β-mercaptoethanol. Co-purified DNA was DNase digested using the Turbo DNA Free kit (Thermo Fisher) using two sequential 30-minute incubations at 37 °C, each using 1 U of Turbo DNase. DNase was inactivated and removed using the supplied DNase inactivation resin and RNA was further purified using the RNA min-elute kit (Qiagen). Unlike the DNase equivalent commonly used in DNA virome protocols, no pre-extraction RNase treatment

was performed as this has previously been suggested to be detrimental to RNA viral recovery (Adriaenssens et al., 2018).

4.5.3. Library preparation, sequencing and initial short read QC

Sequencing libraries were prepared using total RNA without mRNA isolation or rRNA depletion using the NEBNext Ultra Directional RNA Library Prep Kit (New England Biolabs) by the Centre for Genomics Research (CGR), University of Liverpool. Initial fragmentation, denaturation and priming for cDNA synthesis was performed with an incubation time of 7 minutes at 94°C and random primers. In total, 17 libraries were prepared using unique dual indexes and pooled: 15 soil virome samples from five sites, plus one extraction negative control and one library construction negative control of PCR-grade water which were processed alongside the samples and during sequencing library production. A volume of each negative control library equal to the largest volume from a soil virome sample was added to the final pool. Libraries were pooled and sequenced (150 bp paired end) on one lane of a HiSeq 4000.

4.5.4. RNA virome data analysis

Initial demultiplexing and quality control performed by CGR removed Illumina adapters using Cutadapt version 1.2.1 (Martin, 2011) with option -O 3 and Sickle version 1.200 (Joshi and Fass, 2011) with a minimum quality score of 20. Libraries were further filtered by removing reads with a read length <35 bp, a GC percentage of <5% or >95%, or a mean quality score of <25, using Prinseq-lite v0.20.4 (Schmieder and Edwards, 2011). Ribosomal reads were removed using SortMeRNA v3.0.3 (Kopylova et al., 2012) using default parameters. Reads from each library were pooled, error corrected using tadpole.sh (mode=correct ecc=t prefilter=2) and deduplicated with clumpify.sh (dedupe subs=0 passes=2) from the BBTools package (v37.76: sourceforge.net/projects/bbmap/). Reads from all libraries were co-assembled using MEGAHIT 1.1.3 (Li et al., 2015) (--k-min 27, --k-max 127, --k-step 10, --min-count 1).

4.5.5. Identification and abundance of viral sequences

Assembled contigs were compared to the NCBI nr complete database (downloaded on 27th November 2019) using Diamond BLASTx (Buchfink et al., 2014) (--sensitive, --max-target-seqs 15, --evalue 0.00001) and taxonomic assignments made using MEGAN v6 (Huson et al., 2016). All contigs with hits matching cellular organisms or dsDNA/ ssDNA viruses were excluded from subsequent analysis.

HMMs used in RdRP detection were generated from alignments previously published by Wolf *et al.* (2018). Protein coding genes in contigs >300 bp in length were predicted using Prodigal v2.6.3 (-p meta) (Hyatt et al., 2010) and searched for RdRP genes using HMMSearch (Mistry et al., 2013). Hits with E-values <0.001 and scores >50 were clustered with CD-Hit (Li and Godzik, 2006) to 95% average nucleotide identity across 85% alignment fraction (Roux et al., 2018). Each contig was assigned a broad taxonomic classification based on HMMsearch results. Contigs with hits from more than one RdRP phylum were assigned to the classification with the lowest E-value.

Reads were mapped to contigs using BBwrap (vslow=t minid=0.9 - https://sourceforge.net/projects/bbmap/) and contigs with any mapped reads from either of the two negative control libraries were excluded from further analysis. Viral contigss with a horizontal genome coverage of >50% were determined as present for each sample. Any viral contigs with coverage of <50% had its abundance reset to zero. Fragments per kilobase million (FPKM) values calculated by BBwrap were converted to Counts Per Million (CPM) using the fpkm2tpm function from the R package RNAontheBENCH (Germain et al., 2016).

4.5.6. Ecological data analysis

Community analysis was performed using R and the Vegan package (Oksanen et al., 2019). UpSet plots were produced of contigs shared between sampling sites based on the combined collection of viral contigs identified as present at each site. Separate UpSet plots for contigs shared between sampling replicates are contained within Supplementary Fig. C.5. A table of CPM values for each viral contig in each sampling replicate for each site was used to calculate α-diversity metrics (richness, Simpson and Shannon diversity indexes).

The same table was used to generate a β-diversity distance matrix (Bray-Curtiss) using the function metaMDS within the Vegan package. Statistically significant differences between sampling sites were tested for using Kruskall-Wallis (α-diversity metrics) and PERMANOVA (β-diversity). Co-ordinates for individual viral contigs were taken from expanded scores based on a Wisconsin (square root) transformation of the CPM value matrix. Summed CPM values for each phylum were fitted to the NMDS ordination plot using the function env_fit and phyla with a p-value of <0.05 displayed as vectors. Data visualisation was performed using the packages ggplot2 and upsetR (Conway et al., 2017; Wickham, 2016).

4.5.7. Phylogenetic analysis

RdRP sequences from contigs produced by Starr *et al.* (2019) were identified and processed by the same methods described above. These were pooled with those identified by this study and those identified by Wolf *et al.* (2018). Sequences were then aligned using MAFFT v7.427 (Katoh, 2002) (–retree 2 –maxiterate 2) and trees generated with FastTree v2.1.11 (Price et al., 2010) (-wag -spr 4 -mlacc 2 -pseudo -slownni). Trees were visualised with iToL (Letunic and Bork, 2019) and annotated with the aid of table2itol (https://github.com/mgoeker/table2itol).

4.6. Results and discussion

4.6.1. Viromics reveals extensive diversity in soil RNA viral communities

In this work, we characterised the soil RNA viromes of five contrasting soil types along a typical temperate oceanic grassland altitudinal productivity gradient (Figure 4.1 a-d, further described in Supplementary Table C.1) (Withers et al., 2020). Raw reads were filtered for quality and rRNA contamination. The percentage of rRNA reads in each library was highly variable (0.5-95% of total reads) but did not impact the amount or percentage of reads mapping to the collection of assembled viral contigs (Spearman rank correlation, $p = 0.667$ and $p = 0.611$ respectively, (see Materials and Methods section, summary statistics on rRNA read removal and read mapping are provided in Supplemen-

tary Fig. C.1). Future viromics work would benefit from rRNA removal, assuming the yield of purified viral RNA is sufficient for sequencing library construction. Pre-viral lysis RNase digestion has been used previously to achieve this but this can also remove substantial quantities of viral RNA as well (Adriaenssens et al., 2018). Filtered sequencing reads from all libraries were co-assembled and contigs >300 bp were used in further analysis. Genes were predicted by Prodigal (Hyatt et al., 2010) and searched for the RNA viral hallmark gene RNA dependent RNA polymerase (RdRP) using HMMER (Mistry et al., 2013) and five Hidden Markov Models (HMMs) built from multiple sequence alignments of the RdRPs from the five major *Riboviria* RNA viral phyla (Wolf et al., 2018).

A total of 3,471 contigs containing putative viral RdRP genes were taken forward for further analysis. 16 additional contigs were excluded where one or more read mapped to the contig from either the negative extraction, or negative control libraries, of which one had a horizontal coverage of >50%, but only in the negative extraction control. Although a clustering step on the co-assembled contigs was performed at 95% identity over 85% of the contig length, matching the thresholds established for demarking dsDNA viral species (Roux et al., 2018), all clusters from this dataset contained a single viral contig. As boundaries for RNA viral operational taxonomic units are yet to be established, the term "viral contig" has been used in place of vOTU for the purposes of this study. Read mapping was used to identify 3,462 viral contigs present within a sample where the horizontal genome coverage was 50% (Fig. 4.1e). Read mapping and contig coverage statistics are provided in Supplementary Figs. C.1-C.4.

An UpSet plot of viral contigs shared between sites (Fig. 4.1 e) shows that few were common between sites (0.79-32% per site) with the managed grassland sites showing the most similarity. 97-99% of viral contigs shared by managed grassland sites were shared with at least one other managed grassland site (see Supplementary Fig. C.5 for the distribution of contigs shared between replicates of each site). The coastal grassland site shared the least viral contigs with any other site (4 in total), whilst the upland peatland site, although markedly different, shared more viral contigs in common with managed grassland sites it was geographically closer to (7 with the upland-grassland site, 14 overall). This could reflect similarities between those habitats, or result from viral particles being transferred between these habitats by ground/ surface water runoff. As

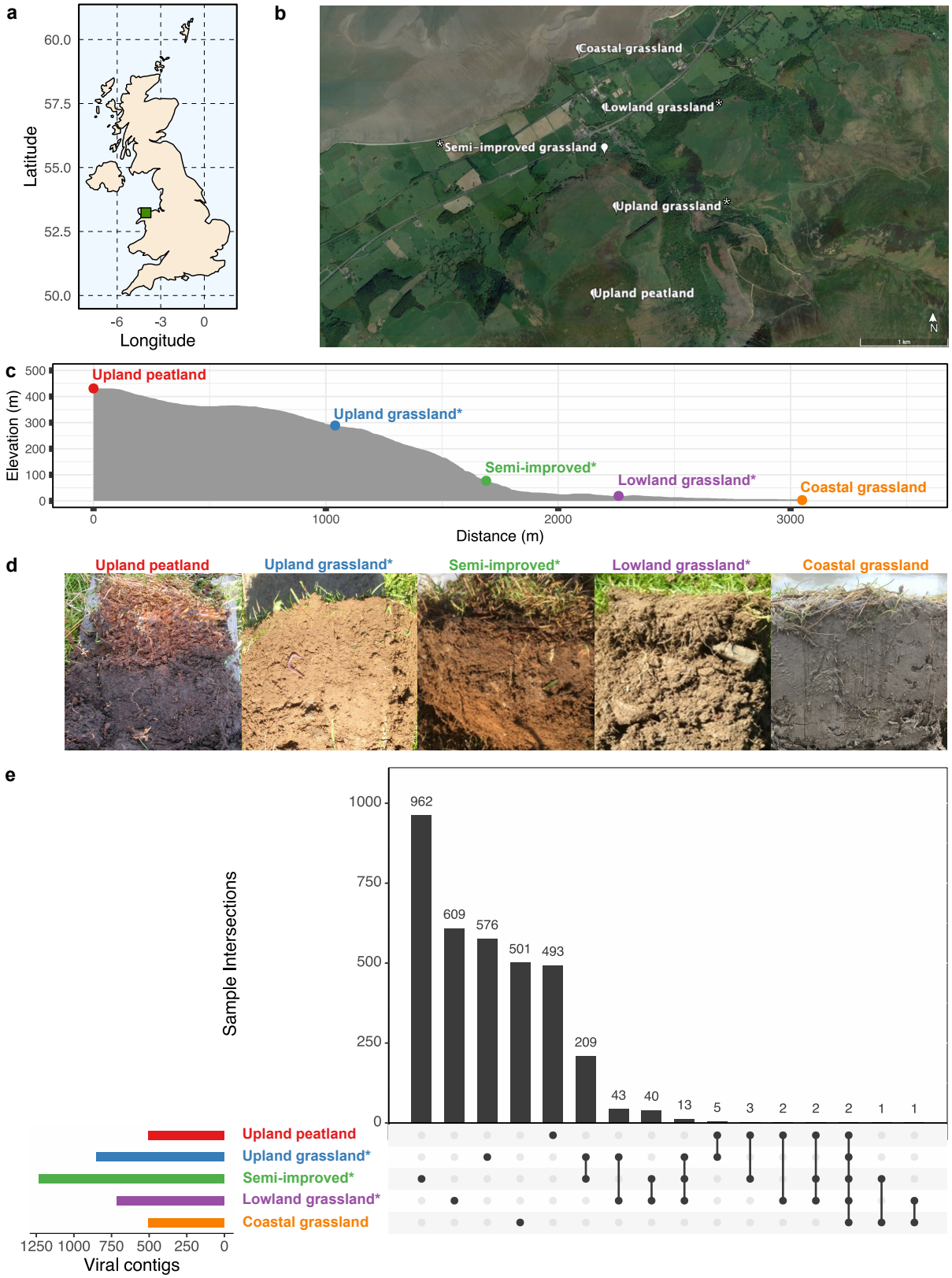


Figure 4.1.: Soil samples were taken along an altitudinal primary productivity gradient in North Wales, UK (a). Sampling sites included upland peatland, (*) three forms of grassland under different management regimes (unimproved upland, semi-improved and improved lowland grassland) and unmanaged coastal grassland (b - made using Google Earth Pro). Elevation varied by 400 m along the transect (c). Site co-ordinates and soil descriptions can be found in Supplementary Table 1. (continues on next page...)

Figure 4.1.: (...) Images of soils from each site can be seen in (d) from left to right: upland peatland, upland grassland, semi-improved grassland, lowland grassland and coastal grassland. An Upset plot of the distribution of identified viral contigs (e) demonstrates how whilst the majority of viral contigs are found solely at each site, the managed grassland sites share more viral contigs in common than with the upland peat or coastal grassland sites, with the coastal grassland site being almost completely unique.

the horizontal coverage threshold used to determine viral contig presence can influence the sensitivity and precision of detection (Roux et al., 2017), the same analysis was repeated with 25%, 75% and 95% horizontal genome coverage and the same pattern of higher overlap between managed grassland sites than with upland peatland and unmanaged coastal grassland sites was repeated (see Supplementary Fig. C.2).

Relative abundance was calculated using mapped reads normalised by contig length and library size (CPM – counts per million, see Materials and Methods) for viral contigs identified as present in each sample. No significant differences in α-diversity were found between the five sites, with Simpson diversity index ranging between 0.93 and 0.99, indicating that all sites are highly diverse (Fig. 4.2 a). Although no overall difference in richness was observed, the semi-improved grassland site showed substantially higher range in viral contig relative abundance than the other four sites, possibly due to increased site heterogeneity.

β-diversity was by non-metric multidimensional scaling (NMDS - Fig. 4.2 b) and phylum-level CPM values fitted to the ordination plot. Each site is separate and significantly distinct (PERMANOVA, $R^2 = 0.738$, $P < 10^{-4}$). Dense collections of viral contigs (grey triangles) are located between the sampling replicates of each site, with samples from managed grassland sites positioning closer together than to samples from upland peatland or unmanaged coastal grassland sites. Figures 4.1e and 4.2 b also show the lack of a clear core soil RNA virome at the viral contig level, and that a combination of soil type, plant coverage and land management may be determining factors of soil RNA viral abundance and diversity.

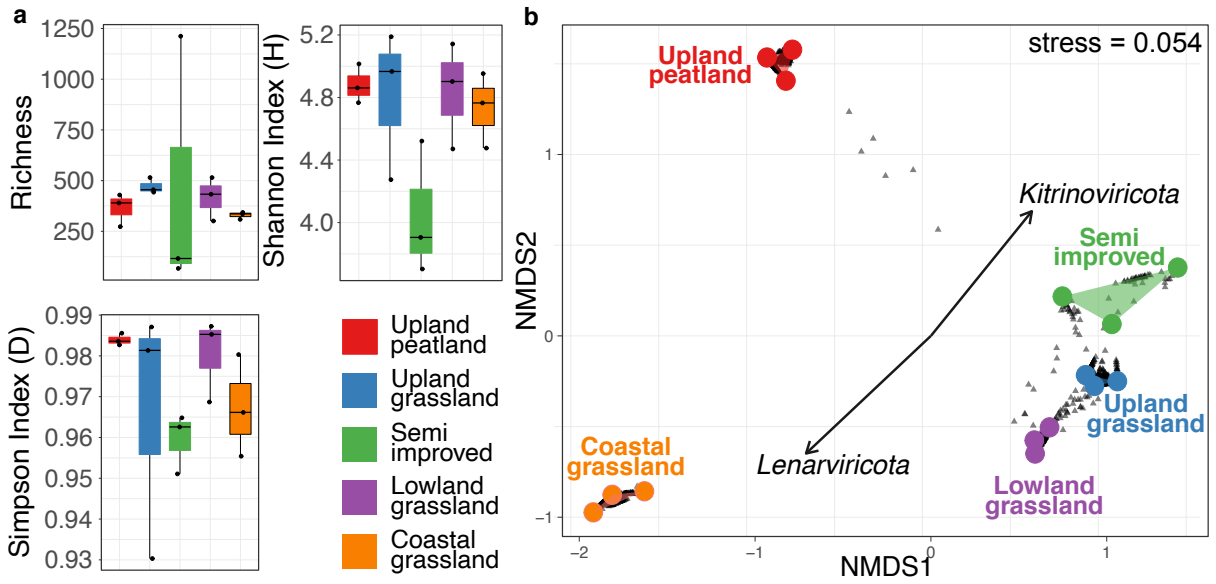


Figure 4.2.: (a) α- diversity metrics and (b) β-diversity NMDS ordination of viral contig relative abundance in 5 contrasting soil types along an altitudinal primary productivity gradient. No statistically significant differences between sites were found in richness (Kruskall-Wallis, $p = 0.369$), Simpson index (Kruskall-Wallis, $p = 0.138$) or Shannon index (Kruskall-Wallis, $p = 0.264$) but a significant difference in β-diversity (Bray-Curtis) was found (PERMANOVA, $R^2 = 0.738$, $P < 10^{-4}$). Managed grassland sites cluster closely together in the NMDS ordination whilst peatland and coastal RNA viromes are clearly separated. Whilst a small number of viral contigs (grey triangles) can be seen to be shared between upland-peat and grassland sites, the coastal grassland site is distinctly separate. Phylum-level CPM values were fitted to the ordination plot to determine which phyla were driving community difference, with *Leviviricota* and *Kitrinoviricota* showing significant effects ($p < 0.05$).

4.6.2. Habitat affects phylum level RNA viral community structure

To explore broader similarities shared between sites, contigs containing RdRP genes were classified based on the broad phylogenetic scheme constructed by Wolf et *al* (2018). This divides the *Riboviria* realm into five phyla, based on RdRP amino acid multiple sequence alignments: positive-sense single stranded *Lenarviricota*, *Pisuviricota* (including some double-stranded RNA viruses) and *Kitronoviricota*, double-stranded *Duplornaviricota* and negative-sense single stranded *Negarnaviricota*, which form a clade located within the *Duplornaviricota* (see Fig. 4.3 a) (Koonin et al., 2020; Wolf et al., 2018). The proportion of *Lenarviricota* members increases, whilst the proportion of *Kitrinoviricota* decreases when comparing lowland and upland sites (Fig. 4.3 b) and these two phyla are significant drivers of differences in β-diversity as indicated in Fig. 4.2b. As the lowland sites are also closer to the coastline (Fig. 4.1 b) and other soil characteristics co-vary with altitude, it

is difficult to identify the environmental drivers of this difference. As the lowland sites are also closer to the coastline (Fig. 4.1 b) and other soil characteristics co-vary with altitude, it is difficult to identify the environmental drivers of this difference. In contrast, the relative abundance of *Pisuviricota* stays broadly similar between each site.

The RNA viromes are all heavily dominated by positive-sense single-stranded RNA (+ssRNA) viruses, with the double-stranded RNA (dsRNA) *Duplornaviricota* far fewer in relative abundance and mostly observed in the semi-improved and coastal grassland samples. Only four negative-sense *Negarnaviricota* RNA (-ssRNA) viral contigs were identified in the whole study and these were exclusively found in the managed grassland sites.

4.6.3. Phylogenetic analyses reveal expanded fine-scale RNA diversity

To explore the phylogeny of the viruses discovered in this study further, protein alignments of RdRP genes for viruses in this study, reference viruses from Wolf *et al.* (2018) and sequences from a recently published bulk soil and leaf litter metatranscriptomics study by Starr et al. (2019) were used to generate phylogenetic trees (Fig. 4.4, more detailed trees are found in Supplementary Fig.s C.6-C.10). Many viruses found in this study appear as blocks of closely related viruses containing few reference sequences, similar to the observations of Starr et al.(2019) In other regions of the phylogenetic trees, e.g. *Pisuviricota and Kitrinoviricota* (Fig. 4.4b and c), novel viruses are fewer in number and evenly distributed across the known RdRP phylogeny.

The phylogenetic tree for the phylum *Lenarviricota* (Fig. 4.4 a) can be divided into three sections containing reference viruses belonging to the class *Leviviricetes*, the family *Narnaviridae* and related mitoviruses, and the genus *Ourmiavirus*, found within the newly reclassified family *Botourmiaviridae* (Ayllón et al., 2020; Koonin et al., 2020). Similarly to the work of Starr *et al.* (2019), this study has detected a large number of potential leviviruses (416 in total, see Fig. 4.4 a, top right quadrant, and Supplementary Fig. C.6). Isolated members of the class *Leviviricetes* predominantly infect *Proteobacteria* and their known diversity has recently been significantly expanded (Callanan et al., 2020; Krishnamurthy et al., 2016). We found comparatively few *Narnaviridae* (21 in total) and this may

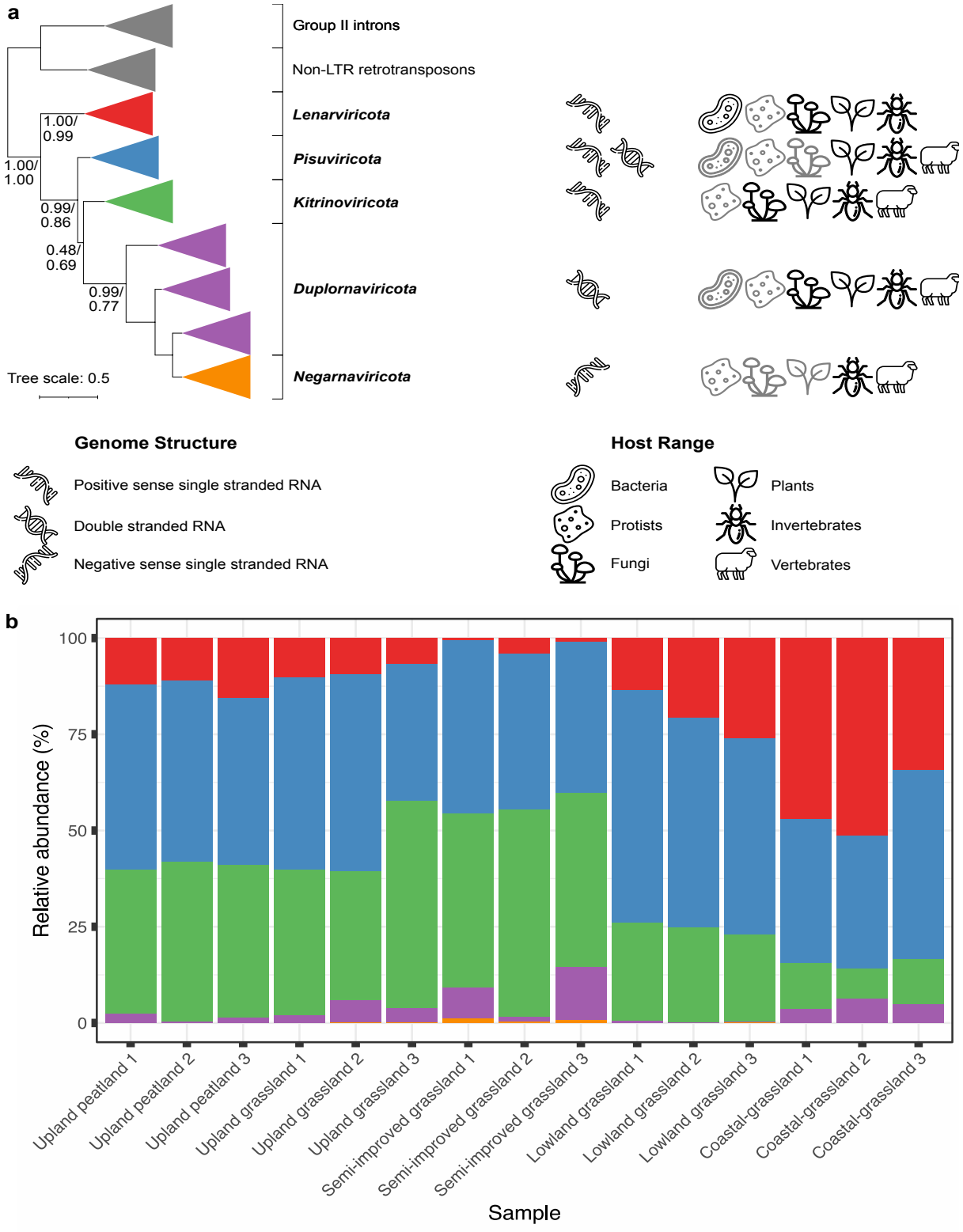


Figure 4.3.: Phylogeny, genome structure and host range of RNA viruses within the *Riboviria* realm, based on RdRp multiple sequence alignments (adapted from Wolf et al. 2018). Grey host icons indicate limited known numbers of viruses infecting these hosts. (continues on next page...)

Figure 4.3.: (...) Viral contigs from 3 independent samples of 5 contrasting soil types along an altitudinal primary productivity gradient were classified into one of these five phyla and relative abundances calculated from CPM values of mapped reads (b). *Pisuviricota* remain broadly similar between samples whilst *Lenarviricota* and *Kitrinoviricota* increase and decrease in proportion respectively moving from upland to coastal sampling sites.

be linked to their structure: *Narnaviridae* members are capsidless +ssRNA viruses that encode no structural proteins and are obligately intracellular (Hillman and Cai, 2013). As viromics approaches isolate intact virions, narnaviruses would most likely not be enriched using this technique. No pre-lysis RNase digestion was performed in this study and those few examples detected here may have been released into the environment from soil processing causing damage to host cells. In contrast, the encapsidated genus *Ourmiavirs*, within the family *Botourmiaviridae,* comprises plant pathogens with segmented genomes of three ssRNA molecules, each carrying genes for a RdRP, movement protein or capsid protein. 377 ourmia-like viral contigs were identified and are almost exclusively found in managed grassland or upland peat sites (Supplementary Fig. C.11), suggesting that this genus of plant viruses may form a larger, more diverse and undercharacterised clade of grassland plant viruses within the *Botourmiaviridae* family. The recovery of complete segmented viral genomes from metagenomics or viromics datasets is particularly challenging due to their segmented nature. Previous studies have used co-occurrence of viral contigs in other publicly available datasets (Obbard et al., 2020) or sequence homology to known viral species (Xu et al., 2020), however the lack of suitable publicly available datasets and extensive horizontal gene transfer can hamper these efforts. Unlike other members of the Botourmiaviridae family, viruses of the genus *Ourmiavirus* are known to possess coat proteins that show similarity with highly disparate viruses spanning multiple phyla (Wolf et al., 2018) and with only three classified species of this genus known, reconstructing their full genomes and classifying novel ourmia-like viruses is particularly challenging. However, their presence here in such high quantities (11% of all detected viral contigs) suggests that they could potentially play an important, but as yet unknown role in grassland ecology. Although viruses are often thought of in terms of pathogenicity, some form persistent mutualistic relationships with their hosts (Roossinck, 2011) whilst others have been shown to trigger hypovirulent phenotypes in normally pathogenic plant fungi and are usable as biocontrol agents (Milgroom and Cortesi, 2004), creating the prospect that

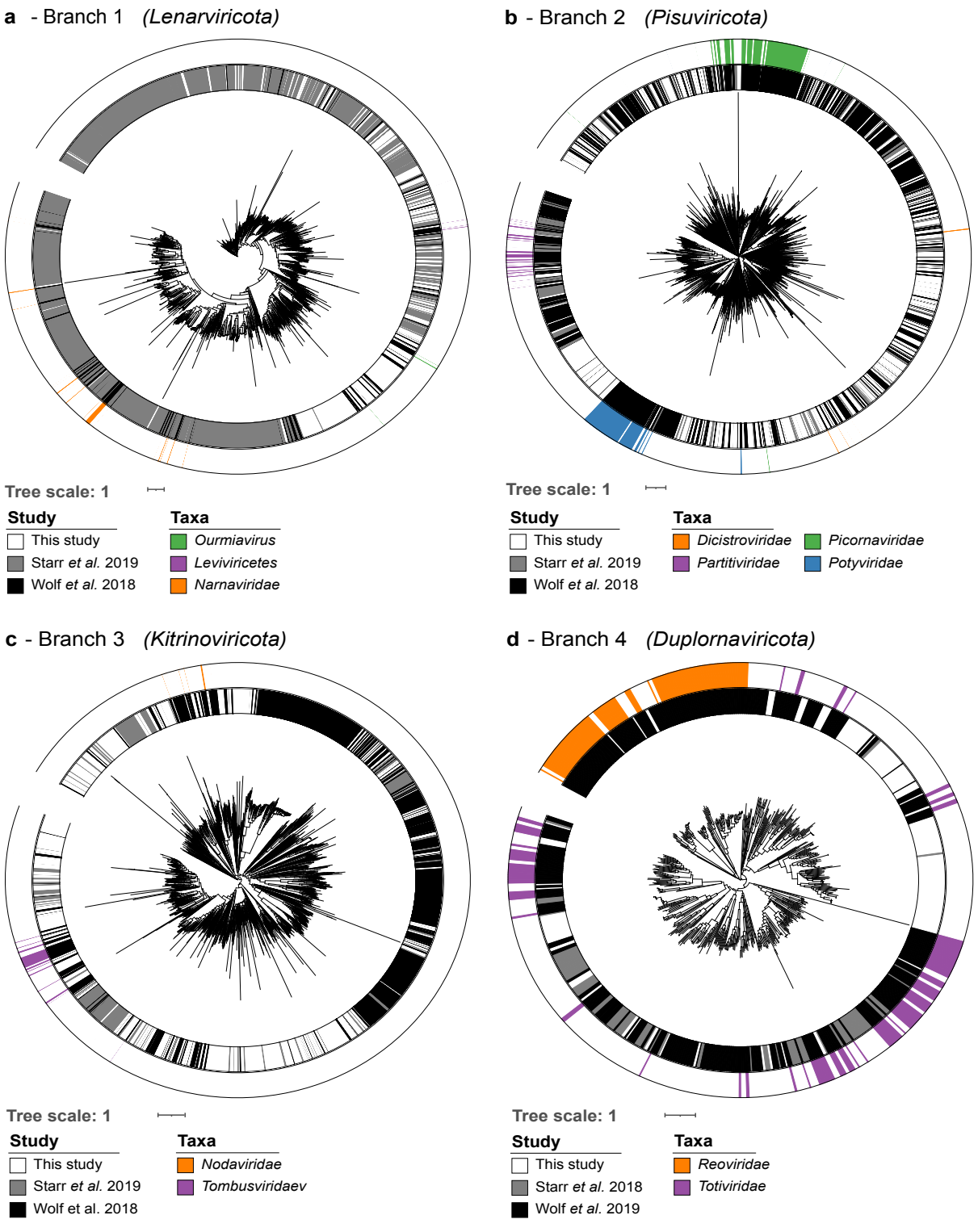


Figure 4.4.: Phylogenetic trees of RdRP genes based on protein multiple sequence alignments. Sequences found across all soil sites from this study (inner ring, white) were aligned with those used to construct the RNA global taxonomy (Wolf et al. (2018) inner ring, black) and another soil study (Starr et al. (2019) inner ring, grey) (continues on next page...)

Figure 4.4.: (...) Global RNA phylogeny is divided into the five proposed RNA viral phyla (a) *Lenarviricota*, (b) *Pisuviricota* (the picornavirus supergroup), (c) *Kitrinoviricota*, (d) *Duplornaviricota*, and *Negarnavirota* (not featured). RdRP genes with established phylogeny of interest are highlighted separately in each panel.

soil viruses may present opportunities for agricultural biotechnology applications.

Pisuviricota (Fig. 4.4b), was the most highly represented RNA virus phylum in this study, comprising 40% of identified viral contigs. This is a highly divergent group of viruses with a broad host range and so reliably identifying the specific host for individual viruses is challenging. Members of the family *Picornaviridae* infect vertebrates and often cause economically important infections (Zell et al., 2017). Relatively few potential *Picornaviridae* were found in this study; however those that were found occupied branches containing various bovine enteric viruses and could be derived from fertiliser manure or sheep dung. Although the separate fields of soil, plant and animal viromics are well established, there are few, if any, studies that consider these separate environments together in order to understand the flow of viruses between them at a community ecology scale.

Of particular interest here are the members of the family *Dicistroviridae* (Fig. 4.4b, bottom right, orange). These arthropod-infecting viruses can range from commensal to lethal disease-causing pathogens with significant economic consequences (Valles et al., 2017). The dicistro-like viruses found in this study observable in the enlarged tree in Supplementary Fig. 4.4a divide into 4 clades: the reference viruses found within the first clade infect crustaceans and the two out of three viral contigs from this study were found in either two or all three coastal samples, the other found in one semi-improved site. The other three clades contain insect-paralysis causing reference-viruses, suggesting that soils may harbour arthropod viruses capable of acutely affecting local mesofauna populations. As soil mesofauna are critical to multiple soil functions (Barrios, 2007), the diversity and ecological roles of arthropod and other invertebrate infecting viruses in soil ecosystems warrants further investigation.

In addition to +ssRNA viruses, viruses belonging to the bisegmented dsRNA *Partitiviridae* family are found within this group, and are capable of infecting plants, fungi and protozoa (see Supplementary Fig. Fig. C.14). Few viral contigs were found within this

group, most likely because, as with *Narnaviridae*, members of the *Partitiviridae* are transmitted exclusively via intracellular mechanisms during spore formation in fungi or ovule/ pollen production in plants (Vainio et al., 2018). The partiti-like viruses found in this study may have been released from plant and fungal tissue in the soil during the extraction process or be present as free virions in the extracellular environment. Although this has not been demonstrated empirically here, it is possible to speculate that infection from normally obligate intracellular viruses could occur when mechanical damage occurs to plants and fungi in soils containing infectious and intact virions.

The viral contigs placed within the phylym *Kitrinoviricota* are distributed throughout the phylogenetic tree of known RNA viruses and are highly numerate, representing 32.4% of identified viral contigs (Fig. 4.4c). Of particular interest here are the three divergent clades in the bottom left quadrant, with the one on the far left containing many known members of the family *Tombusviridae* (blue). These viruses have a wide host range, including plants, protists, invertebrates and vertebrates. The nodaviruses (Fig. 4.4c, orange) divide into two categories: alpha-nodaviruses, predominantly isolated from insects but featuring a wide host range under laboratory conditions, and beta-nodaviruses, infecting fish (Yong et al., 2017). The noda-like viral contigs identified in this dataset were relatively evenly distributed between the managed grassland and upland peat sites, but none were detected in the coastal grassland (Supplementary Fig. C.16 and Fig. C.16).

The phylum *Duplornaviricota* contain the majority of known dsRNA viruses and comparatively few were detected. *Totiviridae* members (Fig. 4.4d - purple) infect fungi, protozoa, vertebrates and invertebrates (Wolf et al., 2018). viral contigs identified in this study were predominantly found to cluster with isolates that infect animalia (right hand side) but some could be found with the fungi-associated *Totiviridae* (left hand side). Very few viral contigs were found amongst possible *Reoviridae* (Fig. 4.4d – orange), with one contig (k127_2512471) showing 97% nucleotide sequence similarity to human rotavirus A (EF554115), found in samples coastal grassland-1 and semi-improved grassland-2.

Only four -ssRNA viral contigs were found in this study and poorly aligned with other known *Negarnaviricota*, which form a clade located within the *Duplornaviricota* (see Fig. 4.3 a). -ssRNA virus structure may inhibit detection by viromics: they are almost ex-

clusively lipid-enveloped, occasionally lacking nucleocapsid proteins (Schmitt and Lamb, 2004), and the harsh extraction protocol may lead to virion disruption and loss of viral RNA. This may also be due to the underrepresentation of plant and soil-dwelling arthropod *Negarnaviricota* within nucleotide databases hindering their detection. Examples of this clade are significantly biased towards vertebrate pathogens (Käfer et al., 2019), however, increased use of high-throughput sequencing has rapidly expanded our breadth of knowledge of *Negarnaviricota* in plants (Randles et al., 2020).

4.7. Conclusions

Using an altitudinal primary productivity gradient as a source of soils with contrasting ecological properties for RNA virome analysis, this study is the first to apply a direct viromics approach to examine the *in-situ* soil RNA viral community of soil ecosystems. We detected 3,462 viral contigs across five sample sites, and observed site-specific variation in viral contig relative abundance. The viral contigs we detected are predicted to be from viruses of a range of hosts, including fungi, bacteria, vertebrates, invertebrates and plants. Therefore, RNA viruses have the potential to influence the grassland soil ecosystem at multiple trophic levels. From a technical standpoint, further development of both wet-lab and bioinformatics techniques are needed to further improve the detection and study of soil RNA viruses. Many RNA viruses have segmented and multipartite genomes, complicating the recovery of full RNA viral genomes from metatranscriptomics and metaviromics datasets. This study found comparatively fewer putative mycoviruses compared to a previous study (Starr et al., 2019) examining RdRP containing contigs in soil metatranscriptomics data. This may be due in part to the different structural characteristics and methods of dispersal used by viruses infecting fungi. While it has been shown that viromics outperforms metagenomics in the recovery of DNA viral genomes (Santos-Medellin et al., 2021), the lack of capsid production in key clades of mycoviruses requires consideration when developing future soil RNA viral ecology methodologies. Use of paired metagenomes, meta-transcriptomes and DNA/ RNA viromics will potentially overcome this difference in detection between RNA viruses and further our understanding of how virus-host interactions and actively replicating viruses influence soil macro- and microbiology.

Whilst environmental DNA viromes are typically dominated by viruses of the prokaryote-infecting class *Caudoviricetes*, the balance in RNA viromes is heavily skewed towards eukaryotic viruses (Wolf et al., 2020). There are multiple explanations for this discrepancy. Many current DNA virus discovery tools are tuned to detect prokaryotic viruses and so may not detect distantly related eukaryotic viruses, however BLAST-based studies report similar biases towards prokaryotic DNA viruses (Adriaenssens et al., 2017; Mahmoud and Jose, 2017). Reference databases of RNA viral sequences are also biased towards viruses of eukaryotes (Cobbin et al., 2021) and so HMM-based search strategies may be more sensitive to these clades due to biases in the underlying HMMs they are based on. These discrepancies could also be due to evolutionary bottlenecks creating a genuine difference in the number of viruses of each domain of cellular life found in terrestrial and aquatic environments. The development of specialist tools for detecting novel eukaryotic DNA viruses and/ or prokaryotic RNA viruses and further exploration of the RNA viral communities of different ecosystems will aid in assessing the true extent of the overall RNA virosphere.

The impact that soilborne RNA viruses have on their host organisms has only just started to be explored, and future work is needed to establish the many influences they may have on global terrestrial ecosystems. Grassland soil bacterial communities show clear responses to the effects of climate change that are mediated by plant-soil-microbial interactions (Koyama et al., 2018) and viruses have the potential to influence soil nutrient cycling through host metabolic reprogramming (Trubl et al., 2018) and their effects on soil microbial community dynamics (Starr et al., 2019) similarly to marine viral communities (Hurwitz and Sullivan, 2013). Our work demonstrates that RNA viral communities are heavily influenced by location, with upland peatland and unmanaged coastal grassland soils sharing very few viral contigs with managed grassland ecosystems and also showing broad differences at the phylum level. Soilborne RNA viruses identified in this study potentially infect hosts across a wide range of trophic levels and can therefore influence soil ecosystems at a variety of scales. Linking these effects of soilborne RNA virus-host interactions with naturally occurring and anthropogenic environmental processes, will be critical in developing a complete picture of how soil ecosystems respond to environmental change.

4.8. Author Contributions

LSH, EMA, DLJ and JMD conceived the study and acquired funding. LSH carried out the sample collection and viral RNA extraction and led the data analysis and preparation of the initial draft of the manuscript. All authors contributed to the final version of the article.

4.9. Data and Code Availability

Post-sequencing centre QC reads are available from the European Nucleotide Archive (BioProject accession number PRJEB45714). Assembled viral contigs were deposited at DDBJ/ENA/GenBank under the accession JAKNTR000000000. The version described in this paper is version JAKNTR010000000. The parent BioProject accession number for all sequencing data is PRJNA804556. RdRP protein sequences, custom R scripts used for data analysis, required input files, multiple sequence alignments and phylogenetic trees are available from Github (https://github.com/LSHillary/ RnaSoilVirome).

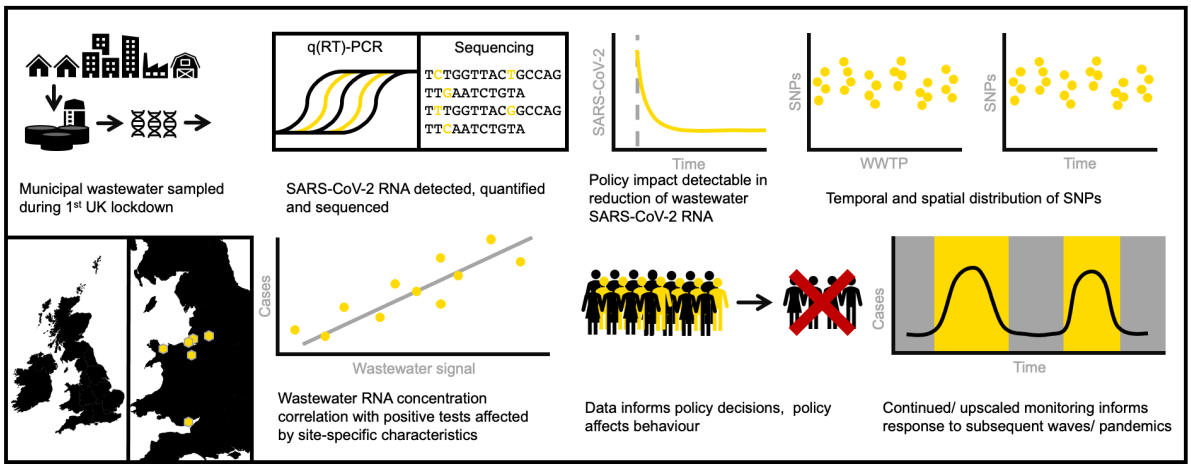
4.10. Acknowledgements

This work was supported by funding from the NERC Biomolecular Analysis Facility pilot project competition (project NBAF1158). LSH was supported by a Soils Training and Research Studentship (STARS) grant from the Biotechnology and Biological Sciences Research Council (BBSRC) and Natural Environment Research Council (NE/M009106/1). EMA was funded by the BBSRC Institute Strategic Programme Gut Microbes and Health BB/R012490/1 and its constituent projects BBS/E/F/000PR10353 and BBS/E/F/000PR10356. RNA sequencing library preparation and data acquisition was carried out by the Centre for Genomics Research at the University of Liverpool. Data analysis utilised high performance computing resources from Supercomputing Wales. The authors would like to thank David Fidler for assistance during sample collection, Mike Grimwade-Mann, Sam Morley and Tom Regan for assistance during sample processing, and Dave Chadwick and Robert Griffiths for comments/ advice on data analysis and visualisation.

5

Monitoring SARS-CoV-2 in municipal wastewater to evaluate the success of lockdown measures for controlling COVID-19 in the UK

5.1. Graphical Abstract



5.2. Foreward

The work described in this chapter was conducted in response to the emerging developing and continuing COVID-19 pandemic and has been published as:

> Hillary, L. S. et al. (2021) 'Monitoring SARS-CoV-2 in municipal wastewater to evaluate the success of lockdown measures for controlling COVID-19 in the UK', Water Research, 200. doi: 10.1016/j.watres.2021.117214.

DLJ, LSH, KF, SKM and JEM conceived the project. LSH, JT, MAD and KF undertook the experimental work. LSH and KF undertook the processing and analysis of the q(RT-)PCR data. KHM, TB and SP undertook the processing and analysis of the sequencing data. LSH, KF and DLJ led the data interpretation and writing of the manuscript. All other authors contributed to the final draft of the article.

This chapter is dedicated to the memory of those who have passed, and those who sadly will pass in the future from COVID-19.

5.3. Abstract

SARS-CoV-2 and the resulting COVID-19 pandemic represents one of the greatest recent threats to human health, wellbeing and economic growth. Wastewater-based epidemiology (WBE) of human viruses can be a useful tool for population-scale monitoring of SARS-CoV-2 prevalence and epidemiology to help prevent further spread of the disease, particularly within urban centres. Here, we present a longitudinal analysis (March–July 2020) of SARS-CoV-2 RNA prevalence in sewage across six major urban centres in the UK (total population equivalent 3 million) by q(RT-)PCR and viral genome sequencing. Our results demonstrate that levels of SARS-CoV-2 RNA generally correlated with the abundance of clinical cases recorded within the community in large urban centres, with a marked decline in SARS-CoV-2 RNA abundance following the implementation of lockdown measures. The strength of this association was weaker in areas with lower confirmed COVID-19 case numbers. Further, sequence analysis of SARS-CoV-2 from wastewater suggested that multiple genetically distinct clusters were co-circulating in the local populations covered by our sample sites, and that the genetic variants observed in wastewater reflected similar SNPs observed in contemporaneous samples from cases tested in clinical diagnostic laboratories. We demonstrate how WBE can be used for both community-level detection and tracking of SARS-CoV-2 and other virus' prevalence, and can inform public health policy decisions. Although, greater understanding of the factors that affect SARS-CoV-2 RNA concentration in wastewater are needed for the full integration of WBE data into outbreak surveillance. In conclusion, our results lend support to the use of routine WBE for monitoring of SARS-CoV-2 and other human pathogenic viruses circulating in the population and assessment of the effectiveness of disease control measures.

5.4. Introduction

The emergence of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), and the resulting global Coronavirus disease 2019 (COVID-19) pandemic has had disastrous socioeconomic and political consequences worldwide (Chakraborty and Maity, 2020). This led to the World Health Organisation (WHO) declaring the COVID-19 pandemic a global health emergency (WHO, 2020). In response to this, many countries implemented a

range of mitigation strategies to reduce the spread of disease, including social distancing, restricted movement, use of personal protective equipment, contact tracing, shielding of vulnerable populations, local or national lockdowns, and community mass testing (Cirrincione et al., 2020; Iacobucci, 2020). These measures are of particular importance in urbanised areas where the spread of disease is most likely (Zhang and Schwartz, 2020). These measures proved to be largely effective at reducing the first wave of COVID-19, albeit not completely eliminating infections (Goscé et al., 2020; Jarvis et al., 2020). The occurrence of subsequent waves of COVID-19 is of significant concern, as countries seek to learn from the effectiveness of the mitigation measures used during the first wave of infection (Aleta et al., 2020).

A large proportion of SARS-CoV-2 infections are asymptomatic or result in only a mild infection (Nishiura et al., 2020). When symptoms do become apparent, this typically occurs 3–7 days after infection (Arons et al., 2020) and severity can vary widely across different sectors of society, disproportionately affecting the elderly (Wang et al., 2020). Evidence points towards the fact that individuals can transmit the virus unknowingly prior to developing symptoms. Furthermore, a- and pre-symptomatic individuals pose challenges to surveillance efforts to accurately estimate the presence and extent of infection in the community. In a more practical sense, both asymptomatic and pre-symptomatic individuals also pose a major threat to public health as they can unknowingly spread the virus to more vulnerable groups (He et al., 2020).

Although mass community testing has been instigated in many countries to estimate the prevalence of COVID-19 in the population, this is costly and the demand for tests frequently exceeds the capacity of testing facilities (Barasa et al., 2020). Focussing testing solely on symptomatic cases may also fail to capture asymptomatic and pre-symptomatic infections, and may focus on populations such as those who are hospitalised, meaning that surveillance is unavailable for the wider community. In some cases, it can also be difficult to obtain nasopharyngeal swabs from high-risk parts of the community due to a range of physical, logistical or cultural issues. Wastewater-based epidemiology (WBE) detects genome fragments of SARS-CoV-2 shed in faeces and urine, and represents an alternative strategy to monitor the levels of virus circulating at population-level scales (Farkas et al., 2020b; Kitajima et al., 2010; Polo et al., 2020). WBE approaches have previously been

successful in evaluating the prevalence of other viral diseases (e.g. polio, norovirus) and also for tracking the use of illicit substances, pharmaceuticals and exposure to xenobiotics (Castiglioni et al., 2014; Ozawa et al., 2019; Zuccato et al., 2008). Monitoring viruses in wastewater also allows an evaluation of the potential risk posed by the discharge of treated and untreated wastewater into the wider environment. Overall, WBE may represent a cost-effective method for determining viral prevalence at the population-level, and has been used to monitor SARS-CoV-2 in a range of countries (Supplementary Table D.1).

Despite the simplicity of the approach, the quantitative recovery of viruses and viral nucleic acids from wastewater is notoriously difficult (Farkas et al., 2018a). For example, virus concentrations in wastewater can be heavily influenced by (i) dilution by rainfall and industrial inputs, (ii) the presence of compounds that may degrade the virus (e.g. detergents, pH, salt), (iii) the presence of substances that physically protect the virus (e.g. faecal matter), (iv) loss of viral RNA during long transit times through the wastewater network due to decay and sorption, (v) variable shedding rates in the community, and (vi) inhibitory substances in the wastewater that may interfere with quantitative (reverse transcription)- PCR (q(RT-)PCR) reactions (Polo et al., 2020). In addition to these factors, the protocols used to concentrate and purify viral nucleic acids from wastewater samples can have substantial impacts on recovery, leading to underestimation of the quantities of the virus present in the wastewater system. Consequently, there is a need to better understand the factors that influence observable levels of SARS-CoV-2 in wastewater to allow validation of the approach for surveillance purposes.

Large-scale efforts to monitor changes in the SARS-CoV-2 genome and track its circulation at national and global scales have largely relied on the analysis of high-throughput sequencing of the SARS-CoV-2 genome in symptomatic individuals (Islam et al., 2020; Meredith et al., 2020; Plessis et al., 2021). As retrospective screening of respiratory samples has detected asymptomatic cases of COVID-19 (Meredith et al., 2020), it suggests that lineages may appear in wastewater samples prior to observation in clinical cases. Because wastewater aggregates samples from across a community/area, sequencing of SARS-CoV-2 RNA recovered from wastewater is likely to contain multiple lineages and so analysis of this data also has the potential to assess the proportions of different lineages circulating in the wider population. This potentially enables the identification of lineages

that are known to be present and early warning of new lineages not previously observed in a catchment.

Here, we present a 3.5-month longitudinal analysis of SARS-CoV-2 RNA prevalence and genetic diversity across six different urban centres during the imposition and gradual lifting of the first national lockdown period in the UK (March-July 2020). The aims of this study were to (i) investigate the use of WBE for tracking SARS- CoV-2 after the implementation of national lockdown measures at six urban centres of varying size within the UK, (ii) determine the influence of environmental factors (e.g. flow) on levels of SARS- CoV-2 RNA and a human faecal marker DNA virus (crAssphage) in wastewater, (iii) investigate the impact of wastewater treatment on the removal of SARS-CoV-2 RNA from wastewater, and (iv) assess the utility of WBE in understanding SARS-CoV-2 genetic variation through high-throughput sequencing.

5.5. Materials and methods

All laboratory procedures were carried out in line with Public Health England/ Public Health Wales advice on the handling of samples suspected of containing SARS-CoV-2.

5.5.1. Sampling sites and wastewater sampling

Untreated influent and treated effluent wastewater was collected from six wastewater treatment plants (WWTPs) located in Wales and Northwest England. The WWTPs served urban areas in the local authority areas of Gwynedd, Cardiff, Liverpool, Manchester, the Wirral and Wrexham, with a total combined population equivalent of ~3 million people (Supplementary Fig. D.1). Untreated wastewater influent from the six WWTPs was sampled on a weekly basis between March and July 2020. Samples were collected in polypropylene bottles as single grab samples with the exception of the Wirral site, which was collected as a 24 h composite sample using an autosampler. Grab samples were collected on weekdays between 08.00 and 09.00 a.m. to ensure temporal comparability, and treated effluent was also collected periodically at the same time as influent. Samples were transported on either the same day, or overnight on ice, to the laboratory, stored at 4 °C and processed within 24 h of receipt. Aliquots of wastewater samples (1.5 ml) were

also frozen in polypropylene vials at −80 °C for subsequent physicochemical analyses and extraction of pre-concentration viral nucleic acids.

5.5.2. Wastewater physicochemical analyses

Wastewater samples were pasteurised before physicochemical analysis by heating to 60 °C for 90 min. Wastewater ammonium concentrations were determined colorimetrically using the salicylic acid procedure of Mulvaney (1996). Nitrate was determined colorimetrically using the vanadate procedure of Miranda et al. (2001) while molybdate-reactive phosphate (MRP) was determined according to Murphy and Riley (1962). All analysis was performed in a 96-well plate format using a PowerWave XS Microplate Spectrophotometer (BioTek Instruments Inc., Winooski, VT). Wastewater electrical conductivity (EC) was measured using a Jenway 4520 conductivity metre and pH with a Hanna 209 pH metre (Hanna Instruments Ltd., Leighton Buzzard, UK).

5.5.3. Wastewater concentration and nucleic acid extraction

Duplicate samples of 50–100 mL of unpasteurised wastewater influent underwent centrifugation (10,000 g , 30 min, 4 °C) and the supernatant and pellet retained. Supernatants were concentrated to 500 μL using Centriprep 50 kDa MWCO centrifugal concentrators (Merck KGaA, Germany). For wastewater effluent samples (see Supplementary Table 5), 1–2 L of each effluent was initially concentrated using tangential flow ultrafiltration with a 100 kDa PES membrane (Spectrumlabs, USA) as previously described (Farkas et al., 2018c), followed by secondary concentration using Centriprep concentrators as described above.

Selected wastewater concentrates, centrifugation pellets and unconcentrated wastewater samples were spiked with approximately 4×10^5 genome copies (gc) of murine norovirus (MNV) as a viral RNA extraction control. Positive and negative nucleic acid control extractions of nuclease-free water with or without the same quantity of MNV spike-in were used to quantify MNV recovery by q(RT-)PCR and to check for cross-contamination during the nucleic acid extraction process or q(RT-)PCR assay setup (described below). The MNV was cultured in BV2 cells in Dulbecco's modified Eagle's minimum essential

medium supple- mented with 2% foetal bovine serum (FBS) at 37 °C in 5% CO2 for two days. Viruses were harvested by three cycles of freeze-thawing (−20 °C/ +37 °C) followed by centrifugation and 100 ×dilution of the supernatant in phosphate-buffered saline pH 7.4. Aliquots of MNV stock were stored at –80 °C until use. The MNV and BV2 tissue stocks were kindly provided by Prof Ian Goodfellow (University of Cambridge, UK).

Nucleic acids were extracted using the NucliSENS MiniMag Nucleic Acid Purification System (BioMérieux SA, Marcy-l'Étoile, France) according to the manufacturer's protocol as described elsewhere (Farkas et al., 2021) and eluted in a final volume of 50 (last week of March 2020) or 100 μL (April-July 2020) of elution buffer. Extracted nucleic acids were stored at –80 °C prior to q(RT-)PCR quantification. The nucleic acid extractions and q(RT-)PCR assay preparation were carried out in separate laboratories inside class II microbiological safety cabinets to minimise the risk of contamination.

5.5.4. q(RT-)PCR and qPCR assays

The q(RT-)PCR assays were carried out in a QuantStudio® Flex 6 Real-Time PCR System (Applied Biosystems, USA) using primers, probes and reaction conditions described in Supplementary Table 2. SARS-CoV-2 N1 and MNV RNA were quantified using a duplex q(RT-)PCR assay or in triplex with SARS-CoV-2 E gene, as described in Farkas et al. (2021) . The 25 μL reaction mix contained 1 ×RNA Ultrasense Reaction Mix with 1 μL RNA Ultrasense Enzyme Mix (Invitrogen, USA), 12.5 pmol of the forward and the reverse primers, 6.25 pmol of the probe/ probes, 0.1 ×ROX reference dye, 1.25 μmg bovine serum albumin (BSA) and 2–5 μL of the extracted wastewater RNA, molecular grade water as a negative control or virus standards. Initially, 5 μL of extracted RNA was tested for wastewater samples. If the MNV recovery was lower than 1%, samples were retested with 2 μL sample/reaction to assess inhibition of the q(RT-)PCR assay, however this was found to be detrimental to assay sensitivity. All data-points used in the analysis came from assays of 5 μL of extracted nucleic acids.

CrAssphage was used as a marker of human faecal abundance/loading in the wastewater (Farkas et al., 2021; Stachler et al., 2018). CrAssphage DNA was quantified using a singleplex qPCR as described previously (Farkas et al., 2019). The 20 μL reaction mix

contained 1 ×KAPA Probe Force qPCR mix (KAPA Biosystems, USA) with 10 pmol of the forward, 10 pmol of the reverse primers, 5 pmol of the probe, 1 μg bovine serum albumin, and 2 μL and 4 μL of the concentrated and original wastewater nucleic acid extracts or controls. A serial dilution of DNA standards within the range of 10^5–10^0 gc μL^{-1} was used for quantification. For SARS-CoV-2, commercially available circular plasmids carrying the N gene or E gene were used (Integrated DNA Technologies Inc., Coralville, IA). Plasmid DNA concentrations were halved when setting up serial dilutions to account for ssRNA producing half the fluorescence signal of dsDNA at the same concentration. For MNV and crAssphage, custom-made, single-stranded oligo DNA sequences carrying the target region were used (Life Technologies, USA). Negative controls (molecular grade water) were included in each run. All samples, standards and controls were run in duplicate and the mean value for each extraction replicate used for further analysis. The limit of detection (LoD) and limit of quantification (LoQ) of the triplex q(RT-)PCR assays were determined previously (Farkas et al., 2021) by running wastewater samples spiked with low concentrations of SARS-CoV-2 (1–150 gc μL^{-1} N1 CDC and 1–200 gc μL^{-1} E Sarbeco) and MNV RNA (1–80 gc μL^{-1}) in ten replicates. The q(RT-PCR) assay LoD (the lowest concentration where all replicates were positive) were 1.7, 3.8 and 3.1 gc μL^{-1} for the N gene, E gene and MNV, respectively. The LoQ (the lowest concentration where the coefficient of variance was below 0.25) were 11.8, 25.1 and 32.1 gc μL^{-1} for the N gene, E gene and MNV, respectively.

5.5.5. q(RT-)PCR data analysis and visualisation

Data were analysed using QuantStudio^{TM} Real-Time PCR Software, version 1.3 (Applied Biosystems, USA). The baseline (cycle threshold; Ct) was manually adjusted after each run, when necessary. Viral concentrations were expressed as mean gc 100 ml^{-1} wastewater calculated from two q(RT-)PCR duplicates of two extraction duplicates (n = 4) per sampling time-point. Statistical analyses and data visualisation was performed in R v4.0.2 (R Core Team, 2013; Wickham, 2016). Supplementary Table Table D.3 contains a full list of packages used in the data analysis.

5.5.6. SARS-CoV-2 RNA amplicon sequencing and data processing

RNA from 84 extraction duplicates from 42 time-points, plus no-template negative controls, were treated with DNase, and used to generate cDNA (NEB Luna Script). Subsequently, SARS-CoV-2 cDNA underwent PCR amplification using V3 nCov-2019 primers (ARTIC) generating 400 bp amplicons tiling the viral genome (Tyson et al., 2020). Amplicon generation was followed by sequencing library construction (NEB Ultra II DNA), with equimolar pooling of samples and quantification. Final library size was assessed on a Bioanalyser high sensitivity DNA chip, and DNA concentration determined by Qubit double-stranded DNA high sensitivity assay, and then by qPCR using the Illumina Library Quantification Kit from Kapa (KK4854) on a Roche Light Cycler LC480II according to the manufacturer's instructions. Libraries were sequenced on an Illumina MiSeq generating 2 $\times$250 bp paired end reads. An average of ca 291,000 reads (ca 146 Mbp) per sample were mapped using bwa-mem against the SARS-CoV-2 genome reference (MN908947.3) within the ncov2019-artic-nf v3 pipeline (https://github.com/connor-lab/ncov2019-artic-nf). SNPs and indels were identified using Varscan v2.4.4 with default settings and summary statistics for coverage and diversity were generated in R v4.0.2 (R Core Team, 2013; Wickham, 2016). Sites were filtered to remove SNPs and indels with a coverage of less than 50$\times$ and a variant frequency of less than 10% per sample. The number of SNP and indel sites were calculated per sample.

The relationship between SNP and indel site frequency and the proportion of the genome with coverage at greater than 50$\times$ coverage and the log 10 gc mL^{-1} were examined with Spearman's correlations. An index of SNP plus indel frequency per sample was calculated by taking the number of SNP and indel sites and dividing by the proportion of the genome with coverage at greater than 50 reads. A mean SNP and indel frequency index were then calculated per pair of wastewater samples to examine the effect of the number of positive tests in the previous 7 days in the local authority area, sample date and WWTP site on the number of SNPs and indels discovered, using a general linear model using the 'glm' function and type II ANOVA using the R package 'car'. A Spearman's correlation was used to examine the relationship between the index of SNP and indel frequency and the log population equivalent served by each wastewater treatment plant. Variants at SNP and indel sites were compared to those recorded in clinical samples using the 'cov_glue_snp_lineage' function from R package 'sars2pack'.

5.6. Results and discussion

5.6.1. Study description and q(RT-)PCR assay development

We monitored the SARS-CoV-2 RNA concentration in influent wastewater at six wastewater treatment plants (WWTPs) using q(RT-)PCR over a period of 3.5 months during the imposition and gradual lifting of the first UK-wide lockdown, and compared these data to the numbers of positive clinical tests and deaths reported by the Office for National Statistics (ONS), UK Government and Public Health Wales for lower tier local authority areas within which the WWTPs were located (HM Government, 2020; Office for National Statistics, 2020; Public Health Wales, 2020). WWTPs represent a range in size (population equivalents from 40 thousand to 1.1 million) and spatial distribution (see Supplementary Fig. D.1) and all implemented combined stormwater, domestic and trade wastewater collection. Influent wastewater grab samples were collected at the same time each week with the exception of The Wirral WWTP which was sampled from a 24 h composite autosampler. Limits of detection (LoD) and quantification (LoQ) were determined as described in Farkas *et al.* (2021).

Results for SARS-CoV-2 RNA concentrations from q(RT-)PCR quantification are displayed as unadjusted mean genome copies (gc 100 ml^{-1}) of wastewater rather than normalised by crAssphage concentrations as factors such as extraction efficiency can vary depending on the virus used (Medema et al., 2020). Although studies suggest that 24 h composite sampling is more representative than grab sampling, it has been shown that grab samples are accurate to within an order of magnitude (Ahmed et al., 2021a; Curtis et al., 2020). Further, our previous work has shown limited diurnal variability, particularly in large wastewater catchments where transit times can be up to 24 h and where large amounts of mixing occurs within the network (Farkas et al., 2018b). Transit times may also influence observable virus quantities due to degradation of viral nucleic acids as they pass through the sewage system; however, SARS-CoV-2 RNA has been shown to be relatively stable in wastewater under environmental conditions, with a T_{90} of 24 or 28 days at 15 or 4 °C (Ahmed et al., 2020b).

We compared mean SARS-CoV-2 RNA concentrations to daily flow and influent wastewater chemistry but found no statistically significant correlations (see Supplementary Table D.4). The highly abundant bacteriophage crAssphage was used as a human faecal marker.

No correlation was found between crAssphage and SARS- CoV-2 nucleic acid concentrations (Spearman, p = 0.834). No effect on crAssphage concentration was observable from sampling week (Kruskal-Wallis, p = 0.904), but a significant effect was found between crAssphage concentration and WWTP site (Kruskal-Wallis, p = 0.0175). These data indicate that faecal loading was constant throughout the study period and that different WWTPs have different balances of human waste and industrial/ other domestic wastewater sources.

5.6.2. Temporal trends in SARS-CoV-2 RNA concentration in wastewater and comparison to COVID-19 epidemiology

For each WWTP, 64% ±6.8 q(RT-)PCR tests (mean ±standard error (SEM), sites = 6, n = 90) detected SARS-CoV-2 in influent wastewater above the LoD, with SARS-CoV-2 RNA concentrations in wastewater influent having quantities above the LoQ in 28.9% ±2.2 of samples (see Supplementary Fig. D.2). No sites showed SARS-CoV-2 concentrations in WWTP effluent above the LoQ and only one above the LoD Wrexham, 19/05/20, n = 22, see Supplementary D.5). Fig. 5.1 a shows a drop in wastewater SARS-CoV-2 RNA concentration, new positive clinical tests and COVID-19 related deaths following the imposition of the UK-wide lockdown beginning in late March 2020. A number of spikes in clinical cases can be observed without corresponding spikes in wastewater, e.g. Wrexham in late June. These can occur due to surge testing following local workplace-related outbreaks and changes in testing eligibility during the study, highlighting the inherent difficulties in comparing wastewater loads to positive tests when testing is both limited and non-random.

WWTPs in Manchester, Liverpool and the Wirral showed strong correlations between SARS-CoV-2 RNA concentration and daily positive tests (Fig. 5.1 b and Supplementary Fig. D.3). Negative correlations were also observed between viral concentrations in all sites and time following the implementation of national lockdown, except Cardiff, indicating these measures lowered the prevalence of the virus in local populations. The Cardiff, Gwynedd and Wrexham WWTPs did not show the same trends between viral RNA concentrations and tests/ deaths, potentially due to several different factors such as water chemistry or lower, broader peaks in SARS-CoV-2 prevalence. Gwynedd is also a popular

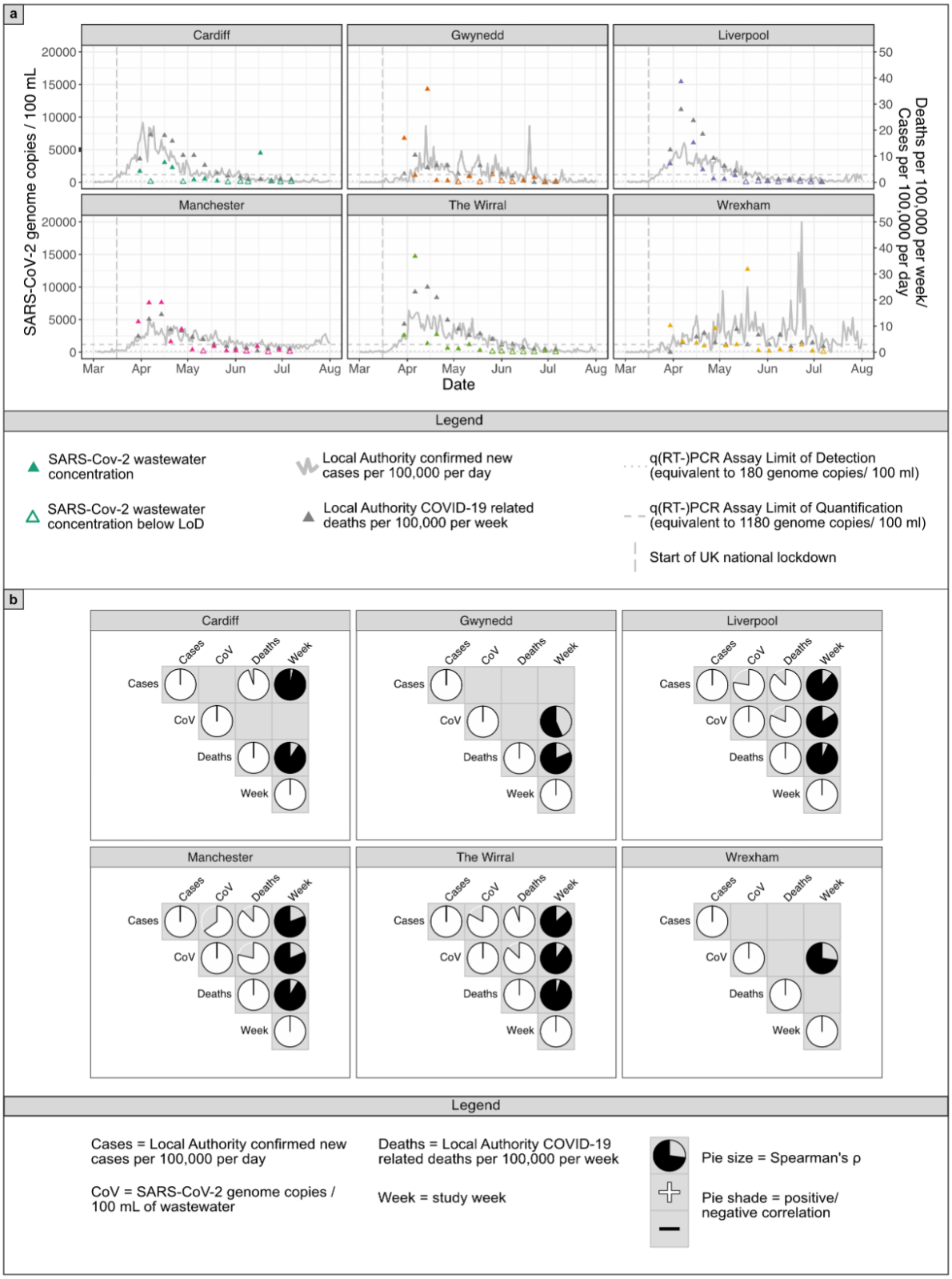


Figure 5.1.: (a)Temporal trend of the recorded number of COVID-19 infections and deaths at six urban centres in the UK and the corresponding levels of SARS-CoV-2 in wastewater. The coloured triangles represent levels of SARS-CoV-2 in influent wastewater, with open triangles being below LoD. Grey triangles represent the (continues on next page...)

Figure 5.1.: (...) number of COVID-19 reported deaths and the solid line represents the number of COVID-19 cases reported in each study region. The dashed and dotted horizontal lines represent the assay LoQ (scaled to 1180 genome copies/ 100ml) and LoD (180 genome copies/ 100 ml) respectively, scaled for a sample volume of 100 mL. The dashed vertical line represents the imposition of UK-wide lockdown measures. (b) Correlation of SARS-CoV-2 RNA concentration (CoV) in influent wastewater with COVID-19 related cases and deaths at six urban centres in the UK. Pie charts represent Spearman correlation ρ where $p < 0.05$ with fullness indicating degree of correlation and colour representing positive (white) or negative (black) correlations.

holiday destination and sees regular weekend influxes of holiday makers from other parts of the UK, which could affect WWTP SARS-CoV-2 concentrations either positively (through visits from asymptomatic/ pre-symptomatic individuals) or negatively (through people commuting from rural areas outside of the WWTP catchment area). Additional factors such as transit time within the sewage network, catchment flow dynamics, and differences between local authority reporting areas for positive tests and WWTP sewershed coverage could affect viral RNA recovery. In contrast to the Gwynedd site, the Wirral site showed the strongest correlation between SARS-CoV-2 RNA concentrations and the number of positive clinical tests/ COVID-19 related deaths, and is of a size between that of the Wrexham and Gwynedd WWTPs (see Supplementary Fig. D.1), suggesting that the use of 24-hour composite sampling may improve the correlation between SARS-CoV-2 wastewater quantification and local clinical cases.

Further exploration of site-specific factors and improved access to higher resolution spatial distributions of positive test locations is required to improve the accuracy of WBE in predicting COVID-19 prevalence amongst local populations as part of national monitoring programmes. Previous studies have corrected SARS-CoV-2 RNA concentration for WWTP flow (Gonzalez et al., 2020), and adjusted cases or positive tests for differences between local authority populations and WWTP catchment areas (Medema et al., 2020). Statistically, we found no benefit of correcting for these factors on Spearman correlation coefficients between WWTP SARS-CoV-2 RNA concentration and positive tests/ COVID-19 related deaths (see Supplementary Fig. D.3), however due to differences between WWTP sites and sewersheds, we would caution against making extensive quantitative comparisons between sites.

Our data confirm that SARS-CoV-2 RNA is readily detectable in wastewater influent across a range of concentrations from $< 1.2 \times 10^3$ (<LoQ) to the highest recorded concentration of 1.5×10^4 gc 100 mL^{-1} . This highlights how site-specific factors, concentration and quantification protocols, and sampling strategies can complicate quantitative comparisons between WWTPs within the same study, and when making comparisons to other international studies. There is a need to standardise SARS-CoV-2 wastewater quantification and take WWTP site identity into account when expanding WWTP monitoring programmes to national and international scales (Chik et al., 2021; Pecson et al., 2021). Nonetheless, this study demonstrates the longitudinal benefit of using WBE to monitor viral prevalence and the impact of public health interven- tions, particularly in the early stages of a novel disease outbreak.

5.6.3. Effect of window size/ offset on correlations

Due to shedding of SARS-CoV-2 from asymptomatic and pre-symptomatic individuals, a key driver of WBE research is the potential to detect upcoming spikes in infection in wastewater before increase in positive clinical tests. Consequently, several studies have used modelling approaches to assess if the wastewater concentration of SARS-CoV-2 preceded new spikes in clinical cases of COVID-19 (Ahmed et al., 2021b; D'Aoust et al., 2021). However, this is challenging due to variabilities in the point of an infection cycle at which a person gets tested, the severity and duration of symptoms, and the variability in viral shedding. The effect of varying the difference between the number of days between wastewater sampling and testing date and the number of days over which to sum the number of positive tests on the correlation between wastewater SARS-CoV-2 concentrations and cases was examined (Fig. 5.2). If only considering daily clinical testing data, the SARS-CoV-2 wastewater RNA concentration leads testing data by 2–4 days but this can be extended by approximately 1 day by using a rolling sum of positive clinical test cases over a series of days leading up to the clinical testing date being considered. It should be noted that the overall effect of varying these parameters is not large in that the correlation coefficients stay between 0.8 and 0.9 over a range of permutations.

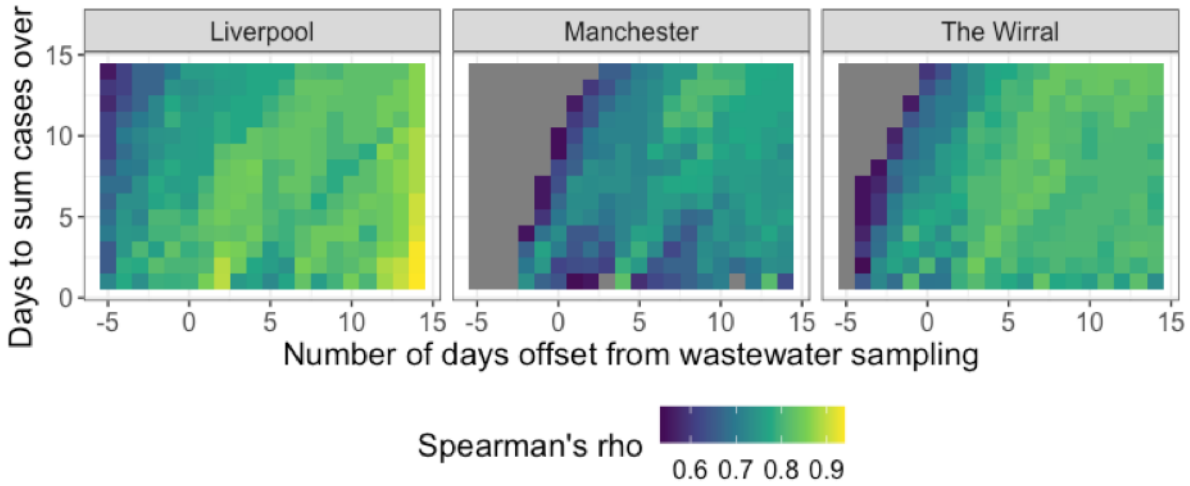


Figure 5.2.: Effects of varying the number of days between wastewater sampling date and clinical testing date (x axis) and the number of days over which to sum cases over (y axis) on the strength of correlation between wastewater SARS-CoV-2 concentration and local authority positive tests. Quantities are shown where a false discovery rate corrected p-value was below 0.05.

5.6.4. Sequencing detects mutations in the SARS-CoV-2 genome comparable to those observable in clinical cases

WBE can also be used to monitor the genetic diversity SARS-CoV-2 circulating in the wider population. To this end, SARS-CoV-2 RNA was amplified using the ARTIC protocol primers in both extraction duplicates, where at least one of which showed q(RT-)PCR amplification were sequenced. In these samples, between 25 and 75% of the SARS-CoV-2 genome was recovered (Fig. 5.3a), with coverage randomly distributed across the genome (Fig. 5.3b). This included samples that showed no amplification (8.3%) or amplification below the LoD (3.6%) of the N1 q(RT-)PCR assay (n = 84), suggesting that multi-locus amplicon sequencing based monitoring of wastewater for WBE may be of significant use in the early stages of future viral outbreaks. The proportion of the genome sequenced positively correlated with the amount of template (Spearman's ρ= 0.376, p = 0.0004, Fig. 5.3c).

In total, 702 unique SNP sites and 267 indels were detectable across the 84 samples after filtering to remove sites with less than 50 reads and a variant frequency within a sample of less than 10%. The number of SNPs found correlated positively with the proportion of the genome that was sequenced (Spearman's ρ= 0.581, p < 0.0001; Fig. 5.3d).

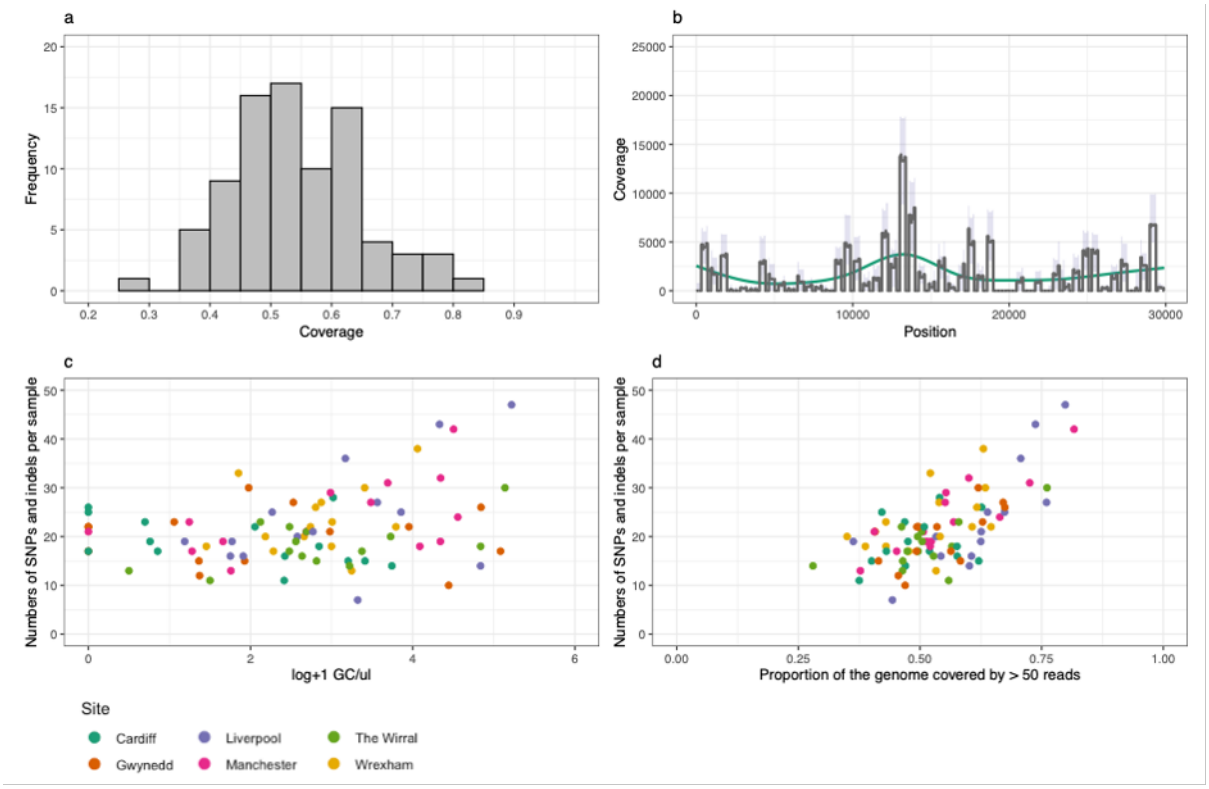


Figure 5.3.: Coverage of the SARS-CoV-genome from reads recovered from wastewater samples. a) Frequency of the proportion of the genome sequenced at 50× depth or greater. b) Coverage across the genome, median plotted in dark grey, interquartile ranges in purple and a smoothed GAM spline in green. c) Proportion of the genome sequenced relative to the estimated number of genome copies estimated from (RT)-qPCR. Note that sequence was obtained in several samples where the (RT)-qPCR for this locus was negative, reflecting the ability of the protocol to sequence genomes of low copy number. d) The number of SNP and indel sites detected relative to the proportion of the genome that was sequenced at 50× or higher.

Preliminary modelling suggests that the rate of positive tests in the source population and sampling week did not affect the mean number of SNPs and indels controlled for genome coverage ($p > 0.05$; Fig. 5.4a and b), but a reduced model suggested that there was heterogeneity amongst sites ($X^2 = 11.57$, df = 5, $p = 0.041$; Fig. 5.4c). The index of SNP plus indel frequency was not related to log population equivalent served by each wastewater treatment plant (Spearman's $\rho = 0.251$, $p = 0.251$; Fig. 5.4d). This is explained by the presence of multiple viral lineages present within the sample, corresponding to the diverse infections in the population represented in the wastewater sample. A substantial fraction of the detected SNPs has previously been identified in clinical samples across the UK, and has the potential to be informative for distinguishing viral lineages (Supplementary Table D.6).

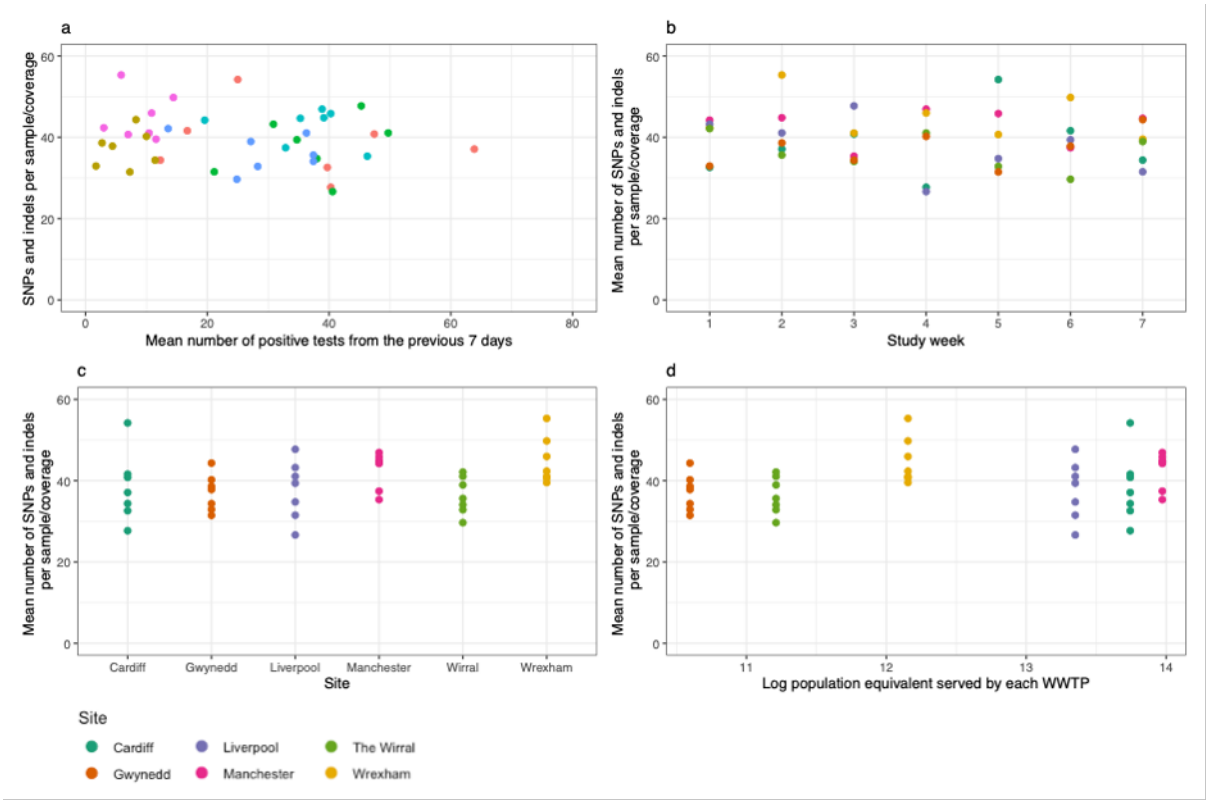


Figure 5.4.: Comparison of the mean number of SNP/ INDELs sites divided by genome coverage to (a) positive tests in the previous 7 days in the local authority, (b) sample date, (c) WWTP site and (d) log10 population equivalent.

Multiple SARS-CoV-2 lineages can be present within a single wastewater sample. Samples have the potential to contain viruses from both symptomatic and asymptomatic individuals within the community, as SARS-CoV-2 has been detected in the faeces of both asymptomatic and symptomatic individuals (Jones et al., 2020; Tang et al., 2020). Previ-

ous studies have sequenced SARS-CoV-2 genomes from wastewater (Ahmed et al., 2020a; Izquierdo-Lara et al., 2021; Martin et al., 2020; Nemudryi et al., 2020). We have shown not only that viral genome sequences can be recovered from wastewater samples, but that they exhibit substantial diversity across dozens of samples. Sequencing the genomes therefore has the potential to assess the diversity of viral infections in the wastewater catchment population and to identify emerging genetic variants before they are seen in clinical samples. In support of this, preliminary analysis suggests that the detected SNPs were consistent with those detected previously in clinical samples (see Supplementary Table D.6). However, because the SNPs from wastewater samples are not phased across the genome, and because the genome coverage is imperfect, assigning viral lineages to samples will require a bespoke statistical framework to be developed.

5.6.5. Use of wastewater-based epidemiology in COVID-19 and future pathogen surveillance

Attempting to quantitatively link observed viral RNA concentrations to detectable cases is challenging (Medema et al., 2020). Many assumptions need to be made regarding the persistence of SARS-CoV-2 in wastewater, quantities of the virus shed in faeces and the influence of water chemistry (Ahmed et al., 2020a).

Sample processing methodology can also be a substantial source of variability. Concentration method, qPCR assay design and inter-lab variation can create variation in detectable SARS-CoV-2 RNA quantities (Pecson et al., 2021; Westhaus et al., 2021). Use of appropriate process controls is necessary to monitor the effects of these factors when making intra- and inter-laboratory comparisons. Choice of process control is complex as a closely related surrogate virus should be used where available and further global collaboration and co-ordination is required to widen access to WBE technologies (Polo et al., 2020). In addition to this, the effects of SARS-CoV-2 on global supply chains and the need to perform WBE at scale create additional pressures where sub-optimal protocols may become necessary in the future to achieve testing scale desired for national monitoring programs.

Despite the possible sources of variability mentioned above, we have demonstrated that

WBE is suitable for quantitatively track- ing the course of the early stages of the SARS-CoV-2 pandemic and the effects of public health interventions, even in the early stages of a novel outbreak, where lack of surge capacity prevents optimal sampling. We highlight how tiled primer array sequencing complements q(RT-)PCR based detection of SARS-CoV-2 and enhances the sensitivity and usefulness of WBE in detecting the presence of novel mutations in the SARS-CoV-2 genome. Early detection of viral pathogens by q(RT-)PCR requires a suitable assay and routine monitoring of WWTPs however alternative technologies such as viral metagenomics may be more suited to initial detection of emerging and unknown pathogens (Farkas et al., 2020b). Our results suggest that viral amplicon sequencing could be more sensitive than q(RT-)PCR for detection of known pathogens. In future, monitoring could be targeted towards ports of entry and major metropolitan centres to maximise the likelihood of detection (Medema et al., 2020).

5.6.6. Conclusions

- Our results demonstrate that levels of SARS-CoV-2 RNA in wastewater generally correlated well with the abundance of clinical COVID-19 cases recorded within the community in large urban centres.

- At the population level, wastewater-based epidemiology was used to confirm the success of lockdown measures (i.e. restricted movement and human-to-human contact) implemented at the national scale to control the transmission of SARS-CoV-2.

- The genetic diversity of SARS-CoV-2 from wastewater suggests that multiple genetically distinct clusters were co-circulating in the local populations, and that the genetic variants observed in wastewater reflect similar SNPs observed in samples from nasopharyngeal swabs taken contemporaneously at clinical testing centres.

- A greater understanding of the factors that affect SARS-CoV-2 RNA quantification in wastewater is still required to enable the full integration of wastewater-based epidemiology data into wider outbreak surveillance programmes.

- Our results lend support to the use of routine wastewater- based epidemiology to monitor SARS-CoV-2 and other human pathogenic viruses circulating in the popu-

lation and to assess the effectiveness of disease control measures.

5.6.7. Data availability

q(RT-)PCR and chemical data recorded in this study is avail- able as supplementary information and from the Environmental Information Data Centre (EIDC, www.eidc.ceh.uk). DOI: 10.5285/ce40e62a21ae45b9ba5b031639a504f7. Sequencing read files analysed in this study can be accessed from the European Nucleotide Archive (project PRJEB42191).

6

Discussion and Future Research

6.1. Introduction

An increasing global population and the pressures it puts on the natural environment, food and water security, and public health makes the need for increased understanding of the emergence and spread of pathogens ever more important. Plant pathogens cause an estimated annual loss in yield of over $220 billion (Savary et al., 2019) and the COVID-19 pandemic serves as a warning of the potential impact of emerging viral diseases, which form 47% and 44% of plant and human emerging infectious diseases, respectively (Anderson et al., 2004; Taylor et al., 2001). Population increases will put pressure on municipal wastewater treatment infrastructure, but the drive towards a greener circular economy creates an opportunity and need for increased use of wastewater byproducts as agricultural fertilisers, due to their high nutrient content (Sharma et al., 2017), and the increased generation of biogas in their production (Di Fraia et al., 2018). As biosolids are derived from human waste, their effective treatment and an understanding of the movement of viruses through our wastewater treatment systems and potential re-entry into the food chain via the application of biosolids to land is critical for managing these risks (Gerba et al., 2002). Whilst the movement of viruses through the wastewater treatment process and release into downstream watercourses has been well characterised(Adriaenssens et al., 2021; Farkas et al., 2020b), the spread of viruses through the use of biosolids as an agricultural fertiliser and how this affects intrinsic soil virus communities is barely understood.

The overall aims of this thesis were to address the following knowledge gaps:

- How does repeated biosolids amendment affect the soil-borne virus community over

the long term?

- How does the viral community respond to a single amendment of biosolids?
- How can we improve techniques for the characterisation of viral communities in soils?
- How can we apply molecular methods of viral detection to the response to an emerging public health crisis?

The previous chapters of this thesis reviewed the current state of the fields of viral ecology and production of biosolids and their use, examined the role of viral ecology in assessing the impact of biosolids amendment on the soil viral community across different temporal scales, described the development of the application of viromics for the study of RNA viruses in soils, and detailed how viral molecular ecology techniques can be applied to public health monitoring of the COVID-19 pandemic during its early phases and beyond. The purpose of this chapter is to summarise and synthesise the findings of these studies and place them in the context of the wider literature. The methodological strengths and weaknesses of these studies will be discussed and finally, the policy implications of these findings and potential directions for future work will be examined.

6.2. Synthesis of findings

6.2.1. Application of viromics for monitoring biosolids amendment of soil

Chapters 2 and 3 demonstrated the utility of taking viromics beyond fundamental viral ecology to examine more applied questions of land management, with regards to using biosolids as an agricultural fertiliser. They examined the levels of persistence of biosolids-associated viruses and the impact that biosolids amendment has on the intrinsic soil virus community over different timescales. Chapter 2 demonstrated that biosolids-associated viruses can persist in soils for years, and potentially decades, although their relative abundance within the overall soil virus community is minimal, at approximately 1%. However, annual biosolids application does raise the possibility of a build-up of persistent viruses that are resistant to degradation in the soil environment, leading to the establishment of a form of steady state whereby decaying viruses are replaced during annual biosolids

applications. The detection of crAssphage-like sequences in Chapter 2 suggests that this group of viruses may represent a tool for long-term monitoring of wastewater-derived viral contamination of the environment. However as crAssphage has also been detected in cat faeces (Y. Li et al., 2021), source tracking of detected contamination to wastewater and biosolids rather than other sources remains an issue. The long-term experiment examined in Chapter 2 amended all plots with chemical N, P, K fertiliser to maximise wheat yield, and so the impacts observed from biosolids amendment was largely independent of the availability of nutrients for the microbial host community. As a result, the differences in the viral communities under different treatments are likely to be solely due to biosolids amendment.

Chapter 3 similarly indicated that after a single amendment of biosolids, the relative abundance of biosolids-associated viruses decays 7.6-fold after one year under controlled conditions. Shifts in α-diversity and predicted host ranges present at the start of the experiment in amended soils returned to levels similar to those observed in control soils after one year. This experiment excluded factors such as soil-type, plowing depth, mesofaunal activity and seasonal climatic effects that could enhance viral degradation and therefore these results most-likely represent a baseline level of residual biosolids-associated viral decay. As no human pathogens were identified in either Chapter 2 or 3, the focus of future viral ecology research on the interaction of biosolids-associated viruses with soil viral communities will also need to consider soil microbial health, and microbial virus: host interactions. As described in Chapter 1, virus communities can be impacted directly by the introduction of viruses from within the biosolids material, but also indirectly through changes in host competition and availability of nutrients to sustain their growth, and soil chemical and physical properties. The effects of spatial variation (Santos-Medellin et al., 2021), temporal changes (Cornell et al., 2021), pH (Lee et al., 2022) and organic fertilisers (Chen et al., 2021) on soil viral communities have been examined and future work will most likely expand to examine how other land management practices impact the soil viral community and integrate this into an understanding of overall soil health.

To date, the role of viruses in soil health has been largely overlooked (Pratama and Elsas, 2018) but it is a rapidly expanding field. Future research will need to integrate viruses into microbial soil processes and examine their mechanistic interactions, as well as their

functional capacity. Achieving this aim is challenging as despite advances in microbial cultivation techniques (Dror et al., 2020), the majority of micro-organisms remain unculturable (Bodor et al., 2020). As a result, few soil-borne virus: host systems exist in which mechanistic interactions can be investigated. Trubl et al. (2021) recently used stable isotope probing to demonstrate substantial differences between active virus and bacterial host communities and the more dormant unlabelled background communities and the role of auxiliary metabolic genes in carbon metabolism within arctic peat soils. A significant knowledge gap currently exists in the role of similarly actively replicating viruses within agricultural soil systems, and it may be that this knowledge could be exploited to enhance soil health from an applied land management perspective via agricultural phage therapy (Dy et al., 2018; Svircev et al., 2018). Currently, such viral ecology studies are predominantly focused on dsDNA bacteriophages, but as Chapters 2 and 3 have demonstrated, viruses of *Archaea*, eukaryotic micro-organisms and higher eukaryotes all have a role in soil microbial and macrobiological ecology, which is even more understudied.

In summary, these two chapters have demonstrated that repeated usage of biosolids as agricultural fertiliser under field conditions, and their single usage under controlled conditions have limited effects on the intrinsic soil viral community. Both studies demonstrated that AMR genes are rare occurrences within bacteriophage genomes and that the distinction between biosolids-associated and soil-associated AMR-containing bacteriophages needs to be made. Auxillary metabolic genes of soil-borne bacteriophages are diverse, covering a wide range of metabolic activities, but an expansion of the techniques used to investigate soil viral ecology will be needed to interpret the mechanisms of how viral community functional capacity impacts host functional activity. The studies have also highlighted the existence of non-dsDNA bacteriophages as part of the soil viral community and that future work should examine their role within soil microbial community dynamics, and the impact of other land-management practices on soil microbial health.

6.2.2. Using viromics to increase the range of detectible viruses

Whilst Chapters 2 and 3 demonstrated the impact of biosolids amendment on the DNA viral community, at the start of the project, no published studies had examined the role of RNA viruses in soil viral ecology. Chapter 4 represents the first study to apply viromics to

the study of soil-borne RNA viruses, revealing that potential host ranges of RNA viruses are distinctly different from DNA viruses, with the majority of RNA viruses infecting eukaryotes. At present, it is unclear if this is a genuine biological difference, or due to a lack of sensitivity in the computational detection of RNA bacteriophages, caused by their underrepresentation in reference databases. A recent large-scale meta-analysis of environmental metatranscriptomes increased the number of known RNA viruses five-fold, but their results still suggest that RNA bacteriophages are in the minority when compared to eukaryotic viruses, both in terms of number of known viral species and in numbers of viruses within different ecological habitats (Callanan et al., 2020; Neri et al., 2022).

In addition to the study from Chapter 4, two other studies have examined the diversity of RNA viruses in soil (Starr et al., 2019; Wu et al., 2021), but no other study has employed methods of virus-like-particle (VLP) enrichment. The results from this study show that viromics most likely selectively enriches certain portions of the virosphere, due to differences in viral lifestyle and virion structure. If soil DNA viral ecology is in its infancy, compared to soil microbial ecology, then soil RNA viral ecology is embryonic. Future work will likely need to address the following key challenges:

- How do we make RNA virus enrichment more efficient?
- How do we improve RNA virus detection?
- How do we further improve host-virus relationship prediction?
- How do we manage the increasing scale and complexity of our datasets?

To answer the first question, methodological development studies are greatly needed, in particular, the depletion of rRNA reads, that can substantially reduce read depth, and impact RNA virus detection sensitivity. The use of paired metagenomes and viromes has been successfully employed to capture a greater diversity of environmental DNA viruses (Santos-Medellin et al., 2021) and the same is even more likely to be true of RNA viruses too.

Improving enrichment and increased sequencing depth will also impact the second issue of improving detection. This will require an expansion of reference databases on which to base sequence homology searches and build machine learning models of viral diversity that can be used for virus detection. A number of large-scale meta-analyses have recently

dramatically increased the number of known RNA viruses (Callanan et al., 2020; Neri et al., 2022) and a major challenge in the future will be unifying these collections of viral sequences into a single, usable database, that can in turn, be used to update the reference databases used by virus identification and classification programs such as DeepVirfinder, VirSorter and vContact (Bin Jang et al., 2019; Guo et al., 2021; Ren et al., 2020). The basic assumptions of these tools will also need to be re-evaluated for RNA viruses, as the guided detection tools such as VirSorter and vContact rely on sequence similarity, contig length, shared genes, etc that hold true for dsDNA bacteriophages, but may not be the same for RNA viruses, eukaryotic viruses or ssDNA viruses.

Fundamental questions such as "what is the definition of an RNA viral species" will need to be addressed by the ICTV, as currently, different RNA species boundaries are demarcated by different marker genes, sometimes carried on different segments of RNA viral genomes. A unified species definition based on RdRP sequence similarity would significantly aid computational analysis. However, this approach could be a double-edged sword, as RNA viral evolution has been demonstrated to be highly modular, with viruses recombining RdRP, capsid and other genes. In addition to this, there is also evidence of motif swapping within the RdRP gene, making the analysis of RNA viral sequencing data much more demanding when applying tools often originally developed for cellular sequence analysis (Neri et al., 2022).

6.2.3. Viral molecular ecology and public health

Although wastewater-based epidemiology (WBE) has been used prior to 2019/ 2020 to monitor the quantity and diversity of viruses within human populations (Farkas et al., 2020a), the COVID-19 pandemic saw a rapid expansion in the application of qPCR and sequencing-based methods to track SARS-CoV-2 in the environment, and use this data to inform public health decisions aimed at reducing the spread of the virus. The study described in Chapter 5 demonstrated the utility of applying WBE to monitor SARS-CoV-2 prevalence in local populations at a time of national, and global crisis, with similar studies conducted in various countries across the world (Ahmed et al., 2020a; Haramoto et al., 2020; La Rosa et al., 2020; Medema et al., 2020; Westhaus et al., 2021). Since the publication of this study, the project has developed into the Wales Environmental

Wastewater Analysis & Surveillance for Health (WEWASH) programme, a collaboration between Bangor and Cardiff Universities, Welsh Water and Public Health Wales aiming to apply WBE to the monitoring of SARS-CoV-2 and other viruses on a national scale, with similar programmes established in England, Scotland and across the world. Substantial efforts have also been made in improving the reliability of WBE data (Ahmed et al., 2020c; Fitzgerald et al., 2021) and have centred on the following themes:

- Improving recovery of viral nucleic acids in the context of large-scale monitoring programs (LaTurner et al., 2021)
- Developing strategies that capture representative samples of the local populations (Bivins et al., 2021)
- Estimating uncertainty in modelling case prevalence from WBE data (X. Li et al., 2021)
- Expanding analyses to examine viral diversity as well as quantity (Lin et al., 2022)
- Additional monitoring of wastewater treatment plant effluent to assess their impacts on downstream bathing waters, and within sewersheds to increase data resolution (Adriaenssens et al., 2021; Reeves et al., 2021)

At the time of writing, mass testing is due to come to an end whilst all legal restrictions intended on limiting the spread of SARS-CoV-2 in the UK are planned to be lifted, all of which is happening against a backdrop of rapidly increasing cases. This creates a situation where WBE will become critical to understanding the levels of SARS-CoV-2 circulating in local populations, similar to the early stages of the pandemic when mass community testing was unavailable. Now that the national infrastructure for mass wastewater testing has been established, and industry - academia - government partnerships built, it is crucial that this capability is not lost. There is an opportunity to utilise this network to understand the viral diversity of our wastewater at a scale not previously possible. There is also the possibility of applying WBE to the monitoring of other materials such as antimicrobial resistance genes (Bürgmann et al., 2018), microplastics (Sun et al., 2019), legal and illicit drugs (Kuloglu Genc et al., 2021), and pharmaceuticals (Oliveira et al., 2020) at the national scale.

6.3. Methodological considerations

All four experimental chapters of this thesis have pushed the technical boundaries of their fields. This section seeks to evaluate the strengths and weaknesses of their approaches. Viromics sample processing and library preparation techniques are known to introduce biases into DNA virome characterisation, particularly with regard to ssDNA viruses (Hjelmso et al., 2017; Roux et al., 2016b), and the same is potentially true of dsRNA viruses as some library preparation techniques do not include denaturation steps prior to random hexamer priming in cDNA synthesis. Chapter 2 sought to overcome this, by using Adaptase-Linker Amplification library construction that would capture both single-stranded and double-stranded DNA in the form of the Accel-1S library preparation kit from Swift Biosciences (now acquired by Integrated DNA Technologies). This has the advantage of accepting inputs of DNA as low as 10 pg but adds a 10 bp linker sequence to each end of a sequencing read, which will be sequenced at the 5' end, and also the 3' end in cases where DNA is severely fragmented, reducing the amount of sequencing information available. Unfortunately, due to the impact of the COVID-19 pandemic on sequencing centre capacity to cater to non-standard library preparation techniques, it was not possible to use the same technology for Chapter 3, and so the NEBNext Ultra II FS library preparation kit was used instead. Whilst this similarly uses linker-amplification, it is not explicitly designed to target ssDNA. This, changes in extraction protocols to reduce the amount of DNA fragmentation, and increased sequencing depth may be sources of the variation in taxonomic profiles observed between the two studies. Soil samples were collected in the same manner, approximately 13 months apart so temporal variation could also be a cause of the observed differences between the control treatments of Chapters 2 and 3, however this is unlikely to be so significant at the family level. Viral ecology could potentially benefit from examining other fields that analyse atypical nucleic acid samples, such as that of ancient/ historical DNA. Here, significant effort has been applied to optimising library preparation, due to low input quantities and the influence that DNA damage can have on sequencing quality and techniques could potentially be adapted to the sequencing of viral DNA (Kapp et al., 2021; Wales et al., 2015).

Long-read data and hybrid assembly also has the potential to further improve our understanding of viral diversity, by improving detection of viral microdiversity (Warwick-Dugdale et al., 2019; Zablocki et al., 2021). These techniques have been applied to

the characterisation of cattle slurry, revealing the extensive presence of virulence factors, AMR genes and crAssphages (Cook et al., 2021). Although the results from Chapters 2 and 3 and other works (Chen et al., 2021) indicated a limited presence of bacteriophage-associated AMR genes in biosolids, it would be interesting to examine how long-read data and hybrid assembly, with its improved ability to generate longer contigs, could enhance our understanding of the functional potential of both soil and biosolids viral communities.

Future experimental strategies may include a multitude of sequencing techniques, including size-selected and bulk metagenomics, metatranscriptomics, long-read data, prophage induction, single cell and single virion genomics, in order to fully capture the viral and microbial diversity present and integrate this data into an understanding of viral ecology (Sommers et al., 2021). Applying these techniques in parallel still remains financially challenging, however as sequencing costs continue to decrease, it may in the not-so-distant future, become a viable approach to apply all the techniques listed above, fully optimised for viral sequence recovery, to examining the role of viruses in diverse ecosystems. This increased diversity of sequencing datasets will also need to be supported by advances in bioinformatics techniques to cope with the sheer volume of information, and also reduce the computational barriers to analysing such datasets.

6.4. Concluding remarks

In addressing these knowledge gaps, this thesis has demonstrated that:

- Repeated biosolids amendment has limited effects on the soil-borne virus community structure or functional capacity.
- A single amendment of biosolids introduces large quantities of biosolids-associated viruses into the soil virus community, but the majority of these are not detectable after one year under controlled conditions. Changes in predicted host profiles suggest an increase in relative abundance of viruses of *Bacteroidetes* and a corresponding decrease in viruses of *Actinobacteria* in amended soil viral communities. These return to proportions similar to control soil viral communities after one year.
- Viromics can be used to detect RNA viruses in soil, expanding their known diversity,

particularly of Ourmia-like viruses, but virion morphology is likely to influence the range of viruses detected by this technique.

- qPCR and amplicon sequencing of viruses in wastewater can be practically applied to monitor changes in viral prevalence and diversity in local populations during the early stages of an emerging pandemic.

Future work has and will continue to build on these results by developing our understanding of the mechanisms that affect viral community dynamics in agricultural soils, how they impact their hosts, crop yield and our food and water security. We will increasingly be able to utilise this knowledge to impact and improve the circular economy, by increasing our reuse of waste products such as biosolids, food production, through the use of bacteriophages to augment soil productivity and public health, by improving our responsiveness to emerging diseases. Ultimately, these changes will contribute to the world meeting the challenges of an increasing global population in a sustainable way.

A

Supplemental information for Chapter 2

A.1. Supplementary tables

Table A.1.: Soil properties and biosolid application parameters

Soil property	**Control**	**Historical**	**Long-term**
pH	5.73 (0.26)	5.78 (0.20)	5.79 (0.09)
%N	0.10 (0.01)	0.16 (0.02)	0.13 (0.01)
%C	1.25 (0.16)	1.76 (0.15)	1.52 (0.16)
Soil microbial biomass C (mg kg $^{-1}$)*	192.1 (19.75)	265.1 (14.21)	189.8 (29.66)
P (mg/ kg)	15.0 (0.6)	16.2 (1.2)	16.0 (0.8)
Cd (mg/ kg)	0.19 (0.03)	0.29 (0.04)	0.23 (0.05)
Cu (mg/ kg)	13.2 (1.2)	36.9 (3.3)	21.0 (1.7)
Zn (mg/ kg)	57.6 (4.4)	70.0 (5.2)	59.5 (4.0)
Biosolid total input (t ha^{-1})**	-	146	138
Biosolid total organic carbon input (t ha^{-1})**	-	62	58

* measured in 2005

** samples were collected prior to the 2017 application

Table A.2.: Biosolid sampling location and properties

Biosolid property	
Sampling location	Banbury
pH	7.1
Total N (%)	5.22
Organic C (%)	42.2

Table A.3.: Settings of short read processing and assembly programs used in Chapter 2

**Process	**Program**	Settings**
Trimming/	Prinseq lite	-min_len 35 -min_gc 5 -max_gc 95
filtering	v0.20.4	-min_qual_mean 25 -trim_left 10 -trim_right 10
Error	Tadpole	mode=correct ecc=t prefilter=2
correction	(BBTools v38.49)	
Deduplication	Clumpify	dedupe subs=0 passes=2
	(BBTools v38.49)	
Assembly	MEGAHIT	–presents meta-large
	v1.1.3	

Table A.4.: Biosolids-associated vOTUs identified in soil viral communities

vOTU	**Length**	**Family**	**Host family**
k127_878067	4463	NA	NA
k127_739563	2675	NA	NA
k127_584835	2385	NA	NA
k127_554424	2065	NA	NA
k127_3373087	27357	inconclusive	NA
k127_3205173	1537	NA	NA
k127_270925	6321	inconclusive	Mycobacteriaceae
k127_245266	6675	inconclusive	NA
k127_240699	1545	NA	NA
k127_2384432	2065	NA	NA
k127_2009273	3257	NA	NA
k127_190657	3561	NA	NA
k127_1673294	12156	inconclusive	NA

Table A.5.: Generalised linear models and beta regression models used in this chapter

Index/ property	**Treatment X^2**	**Treatment p**	**Block X^2**	**Block p**	**GLM family**	**Formula**
Richness	14.2101	0.000821	0.0711	0.789780	Negative binomial (link = log)	richness ~ treatment + block
Shannon Index	3.9574	0.1382	0.0007	0.9793	Gamma (link = log)	shannon ~ treatment + block
Simpson Index	2.1907	0.3344	0.1836	0.6683	Beta	simpson ~ treatment + block
Number of biosolids-associated vOTUs in amended plots	98.457	< 2.2e-16	6.928	0.008486	Quasibinomial	proportion/100 ~ treatment + block
Relative abundance of biosolids-associated vOTUs in amended plots	80.411	< 2.2e-16	8.224	0.004134	Quasibinomial	relative abundance/100 ~ treatment + block

Index/ property	**Treatment X^2**	**Treatment p**	**Block X^2**	**Block p**	**GLM family**	**Formula**
Percentage of circular vOTUs identified as lysogenic	2.09495	0.3508	0.99537	0.3184	Binomial	Proportion of ysogenic vOTUs ~ treatment + block

A.2. Supplementary figures

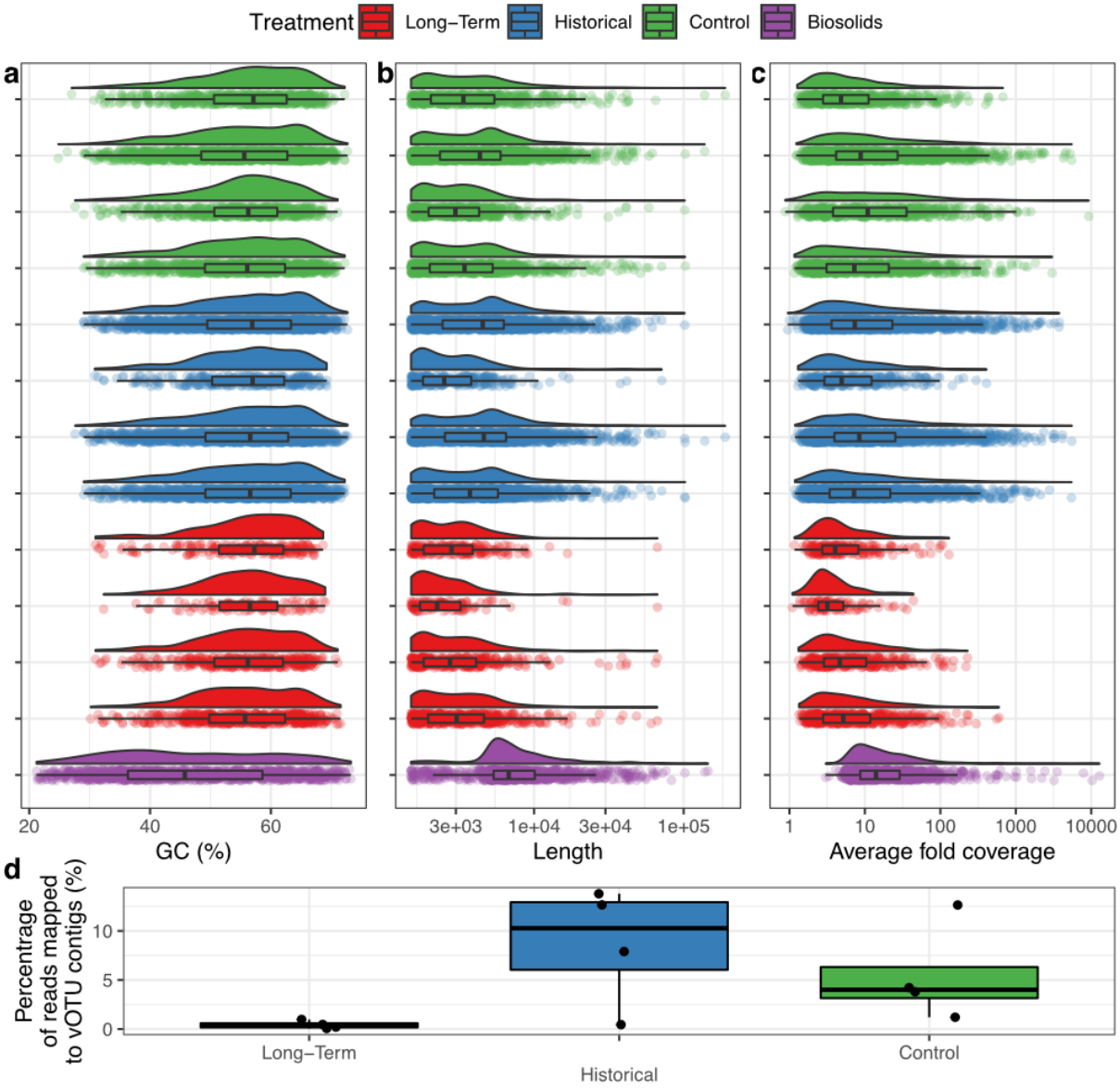


Figure A.1.: vOTU contig summary statistics, including (a) %GC content, (b) length, (c) vertical coverage and (d) Percentage of rarefied reads mapping to vOTU contigs. A lower percentage of reads were mapped to vOTU contigs in the long-term treatment libraries than the historical or control treatments, however this difference is marginally insignificant (Kruskal-Wallis, $p = 0.055$). Both contig length and average fold coverage were statistically significantly shorter/ lower in the long-term treatment compared to the control and historical treatments (Kruskal-Wallis, $p < 0.0001$.

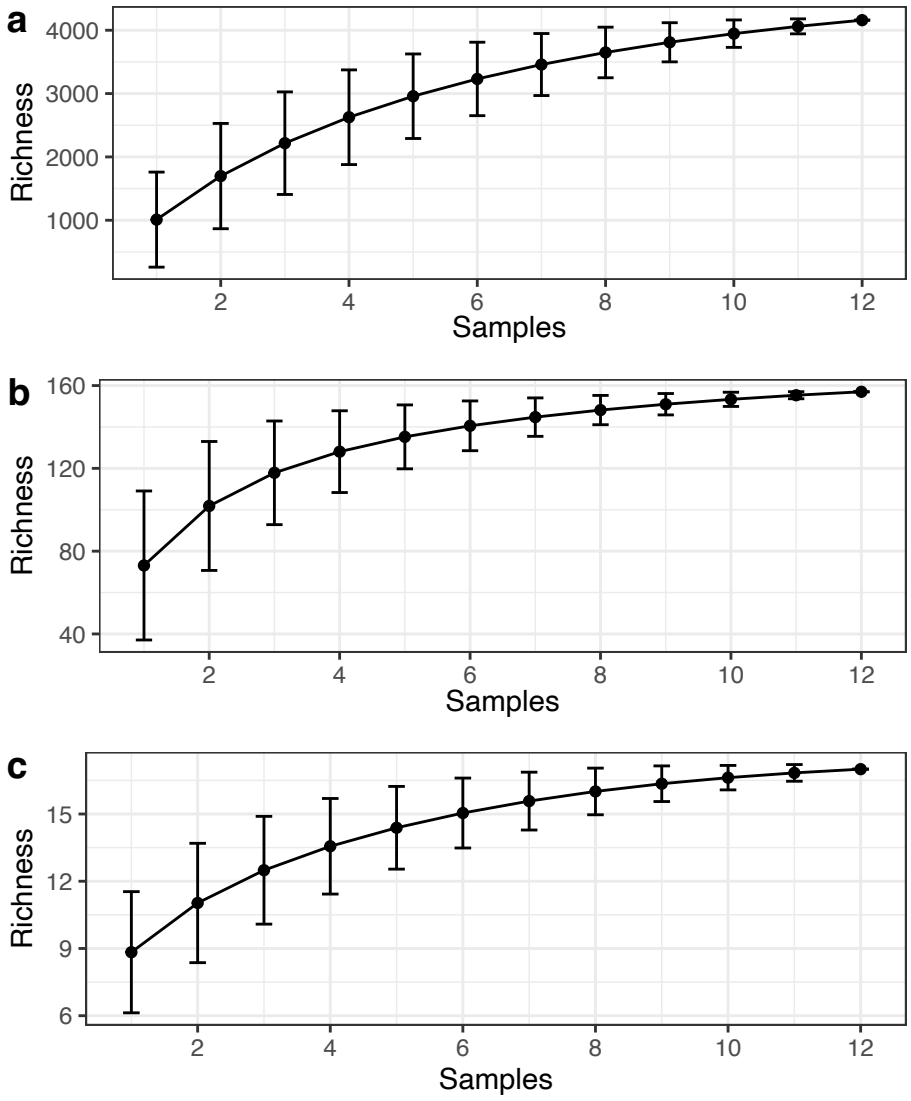


Figure A.2.: vOTU accumulation curves at (a) vOTU, (b) genus and (c) family level

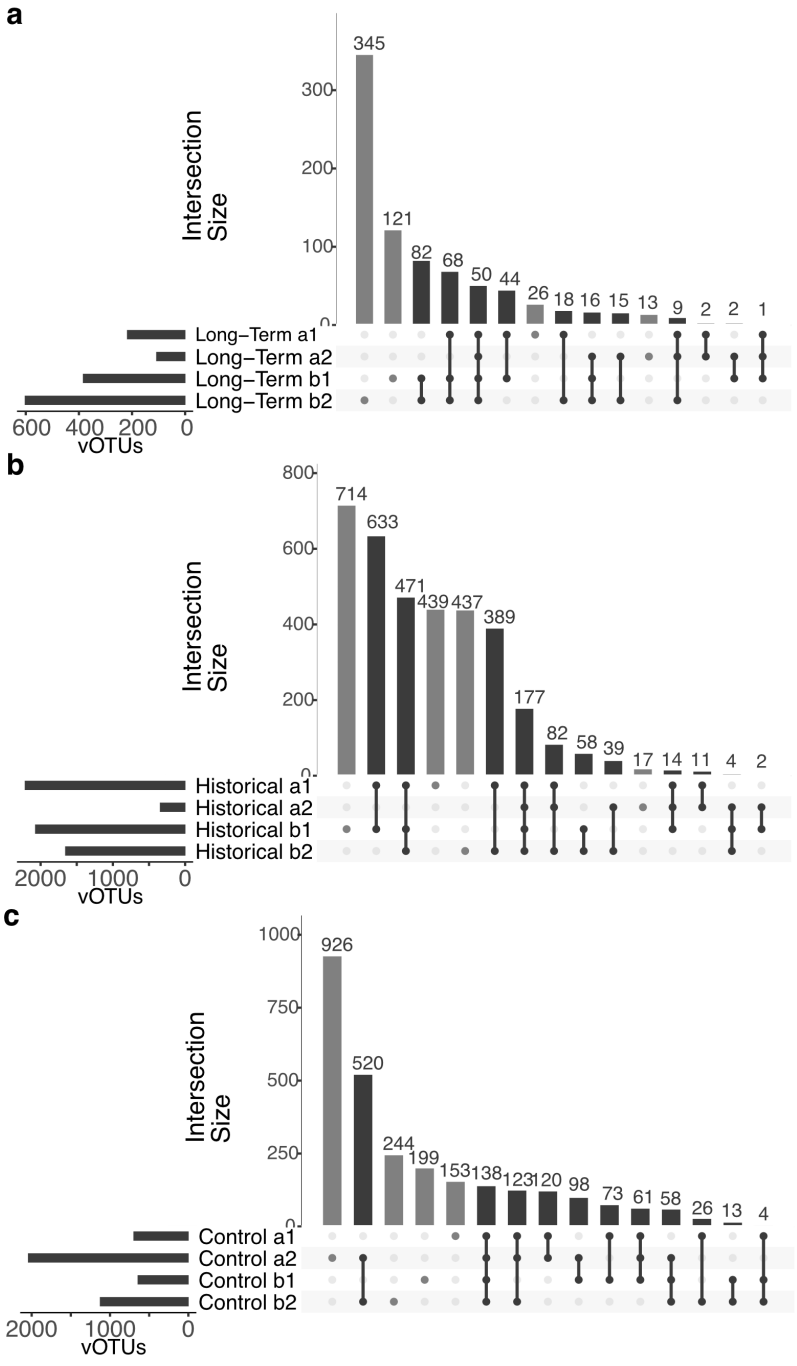


Figure A.3.: Contigs shared between replicate plots in (a) long-term biosolid amendment, (b) historical biosolid amendment and (c) control (unamended) treatments. The number of vOTUs unique to single plots are highlighted in medium-grey.

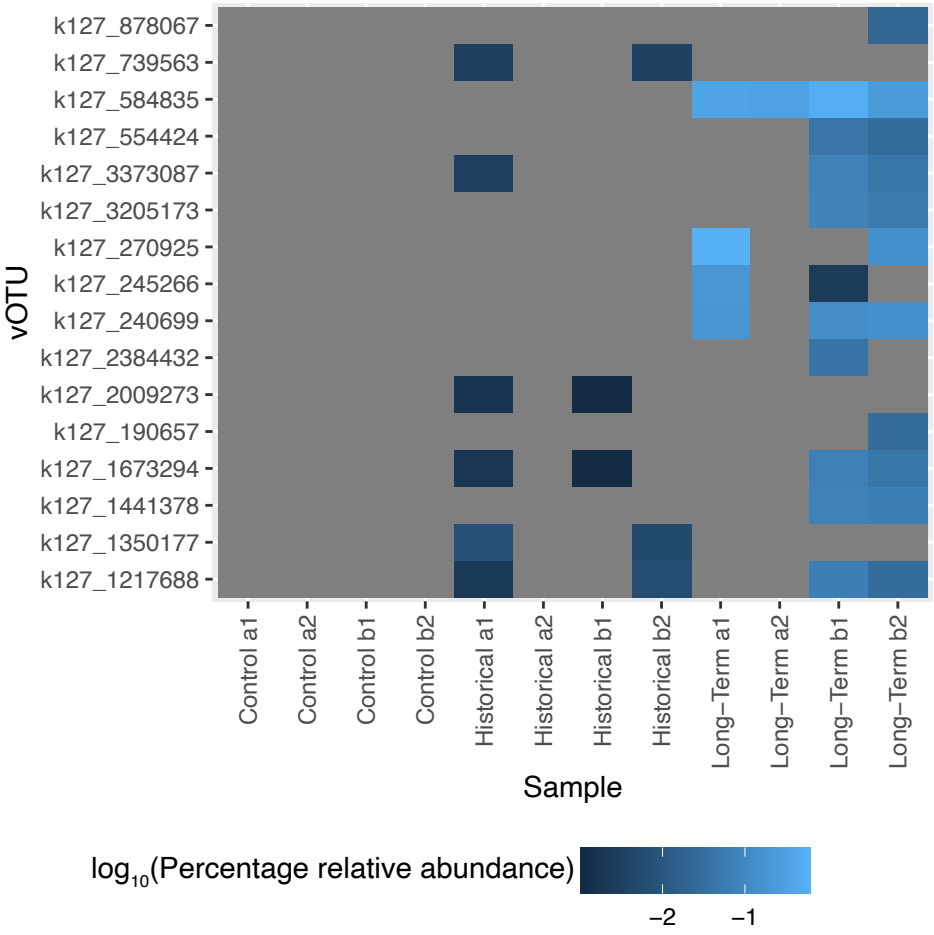


Figure A.4.: Heatmap of biosolids-associated vOTU percentage relative abundance in each soil sample. Relative abundance was calculated from CPM values (see Methods) and displayed on a log_10 scale.

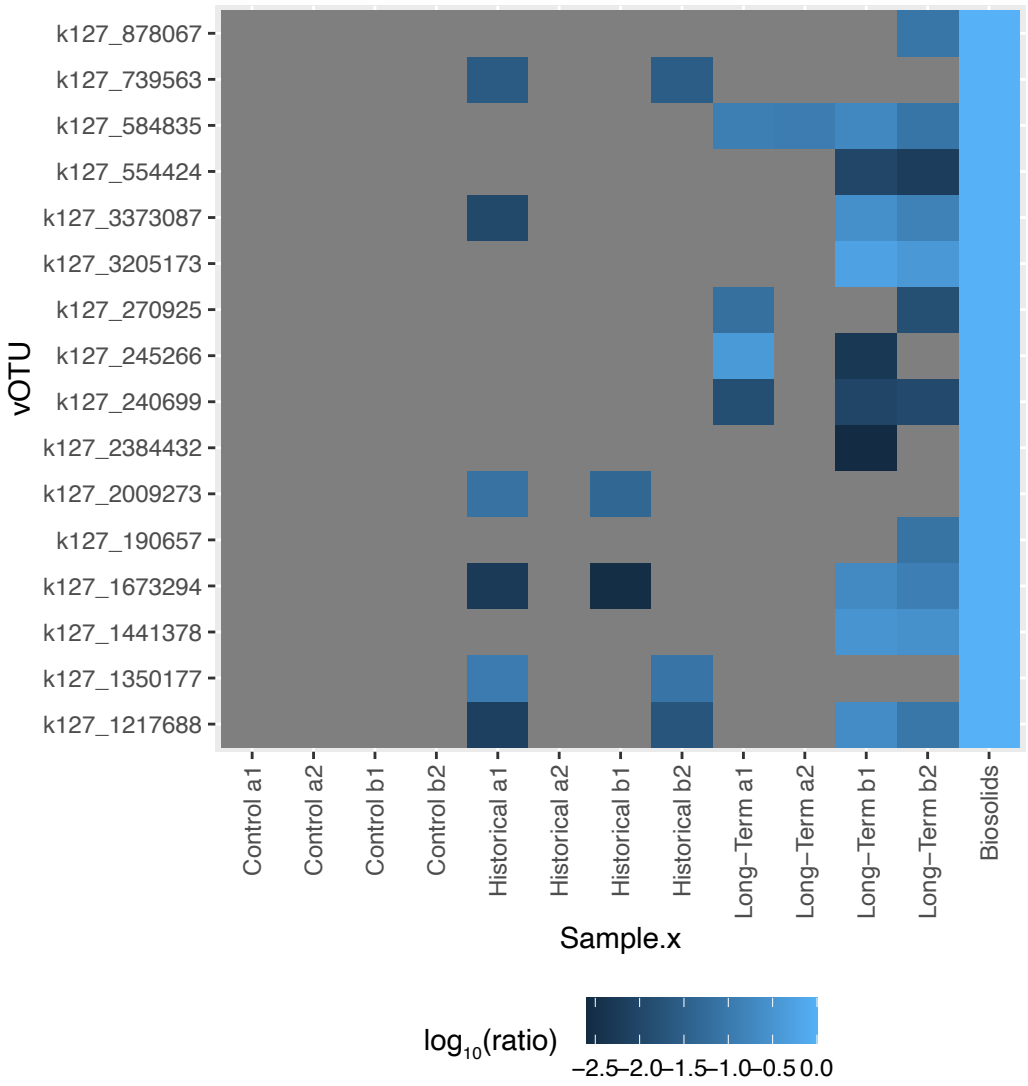


Figure A.5.: Heatmap of biosolids-associated vOTU relative abundance ratios between biosolids and soil viromes. Ratios were calculated by normalising CPM values by the CPM value for each contig in the biosolids virome and displayed on a log_10 scale.

B

Supplementary Information for Chapter 3

B.1. Supplementary tables

Table B.1.: Soil properties

Soil Property	
pH	5.73 (0.26)
%N	0.10 (0.01)
%C	1.25 (0.16)
Soil organic carbon (%)	1.12 (0.00)
Soil microbial biomass C (mg kg $^{-1}$)*	192.1 (19.75)
P (mg/ Kg)	15.0 (0.6)
Cd (mg/ Kg)	0.19 (0.03)
Cu (mg/ Kg)	13.2 (1.2)
Zn (mg/ Kg)	57.6 (4.4)
Biosolid total input (t ha^{-1})**	8

Table B.2.: Short read processing and assembly

Process	**Program**	**Settings**
Filtering	bbduk	ftl=3 maq=25 minlen=35
Error correction	tadpole	mode=correct ecc=t prefilter=2
Deduplication	clumpify	dedupe subs=0 passes=2
Assembly	MEGAHIT	–presets meta-large
Mapping	bbwrap	vslow=t minid=0.9

Table B.3.: Generalised linear models and beta regression models used in this chapter

Metric	**Formula**	**Family**
Number of vOTUs	response ~ treatment + time + block	Negative binomial
Shannon Index	response ~ treatment + time + block	Gaussian (link = log)
Simpson Index	response ~ treatment + time + block	Beta regression
Difference in biosolids-associated vOTUs as a proportion of total vOTUs in amended microcosms	response ~ time + block	Beta regression
Difference in total biosolids-associated vOTU relative abundance in amended microcosms	response ~ time + block	Beta regression

B.2. Supplementary figures

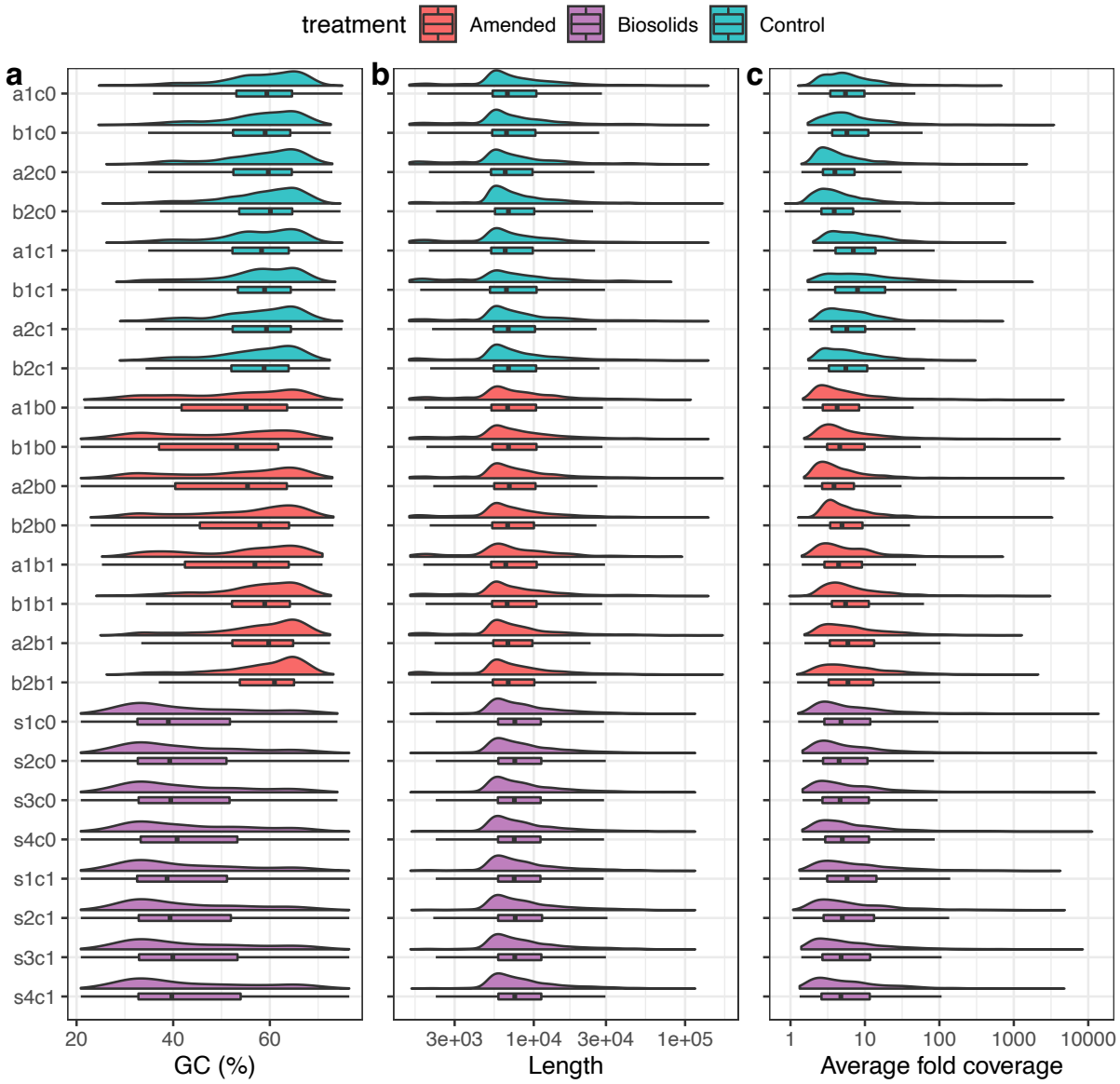


Figure B.1.: vOTU contig summary statistics, including (a) %GC content, (b) length, (c) vertical coverage and (d) Percentage of rarefied reads mapping to vOTU contigs.

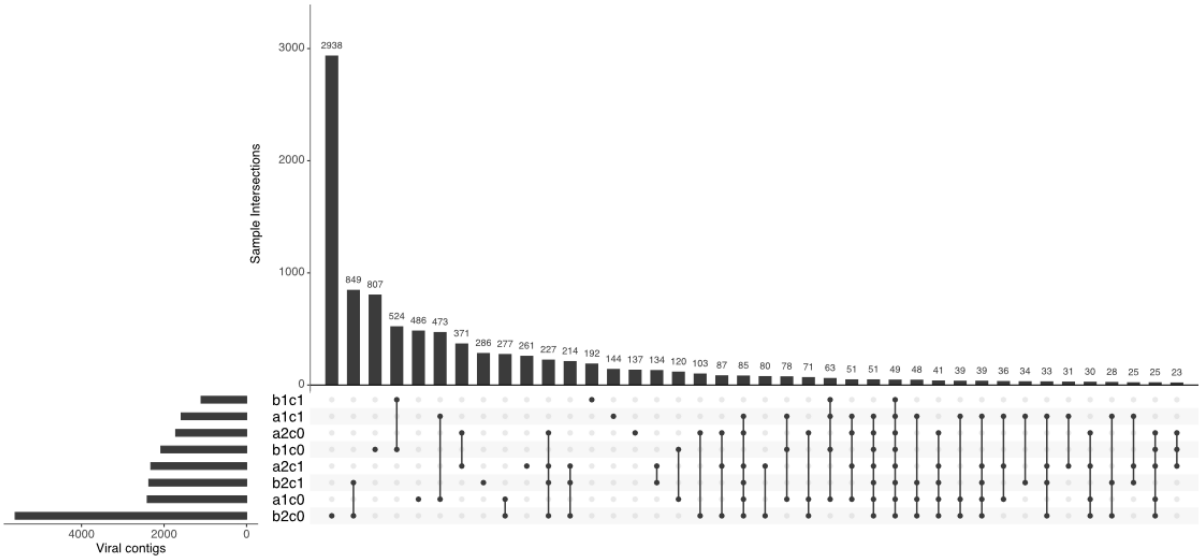


Figure B.2.: UpSet plot of vOTUs in individual control soil microcosms: first letter - a/b = replicate a or b; first digit - 1/2 = block 1 or 2; second digit - 0/1 = year 0 or year 1. Some intersections with single members not displayed.

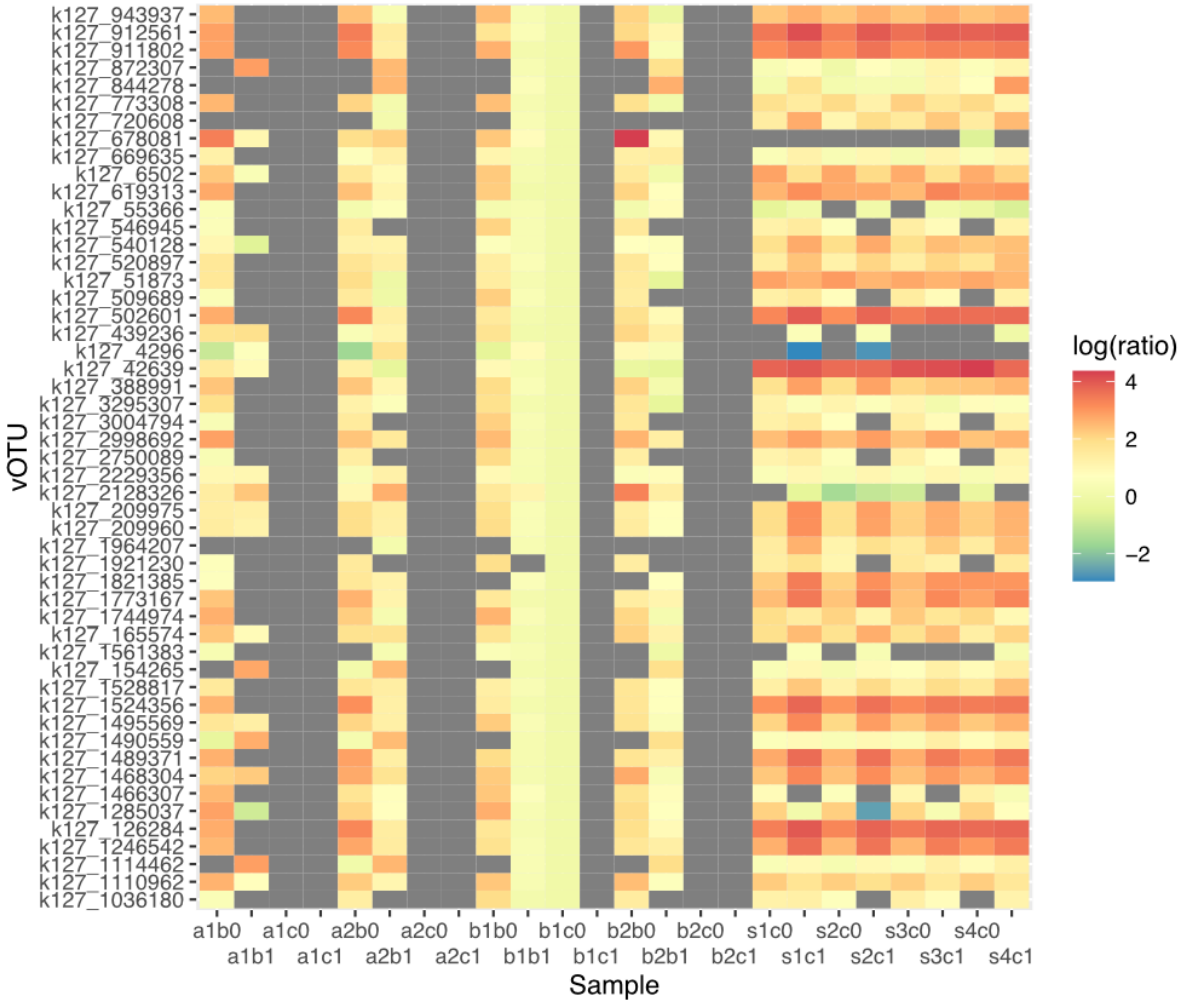


Figure B.3.: Relative abundance ratios of possible biosolids-associated contaminant vOTUs in sample b1c0. Colours represent the log ratio of CPM values (relative abundance) for each vOTU between the sample on the x-axis and sample b1c0. Only 2% of combinations had ratios between 0-1, and no vOTUs present in any other control sample. This suggests that their presence in sample b1c0 is due to false positives, contamination or genuine presence in biosolids-amended and soil environments.

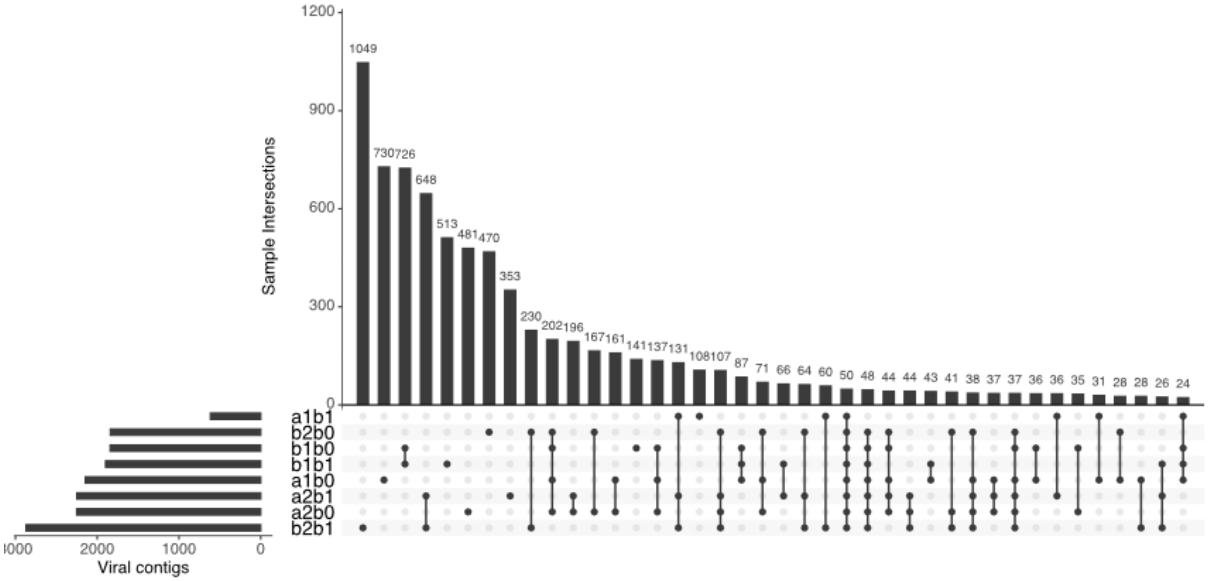


Figure B.4.: UpSet plot of vOTUs in individual biosolid-amended soil microcosms: first letter - a/b = replicate a or b; first digit - 1/2 = block 1 or 2; second digit - 0/1 = year 0 or year 1. Some intersections with single members not displayed.

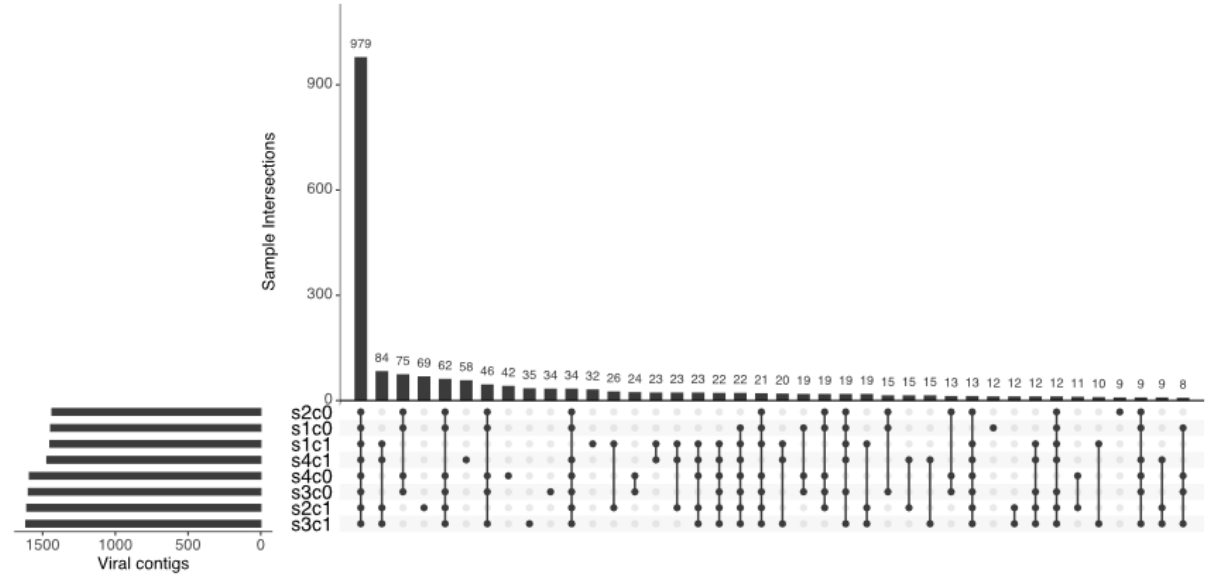


Figure B.5.: UpSet plot of vOTUs in individual biosolids microcosms: first digit - 1/2/3/4 = replicate 1, 2, 3 or 4; second digit - 0/1 = year 0 or year 1. Some intersections with single members not displayed.

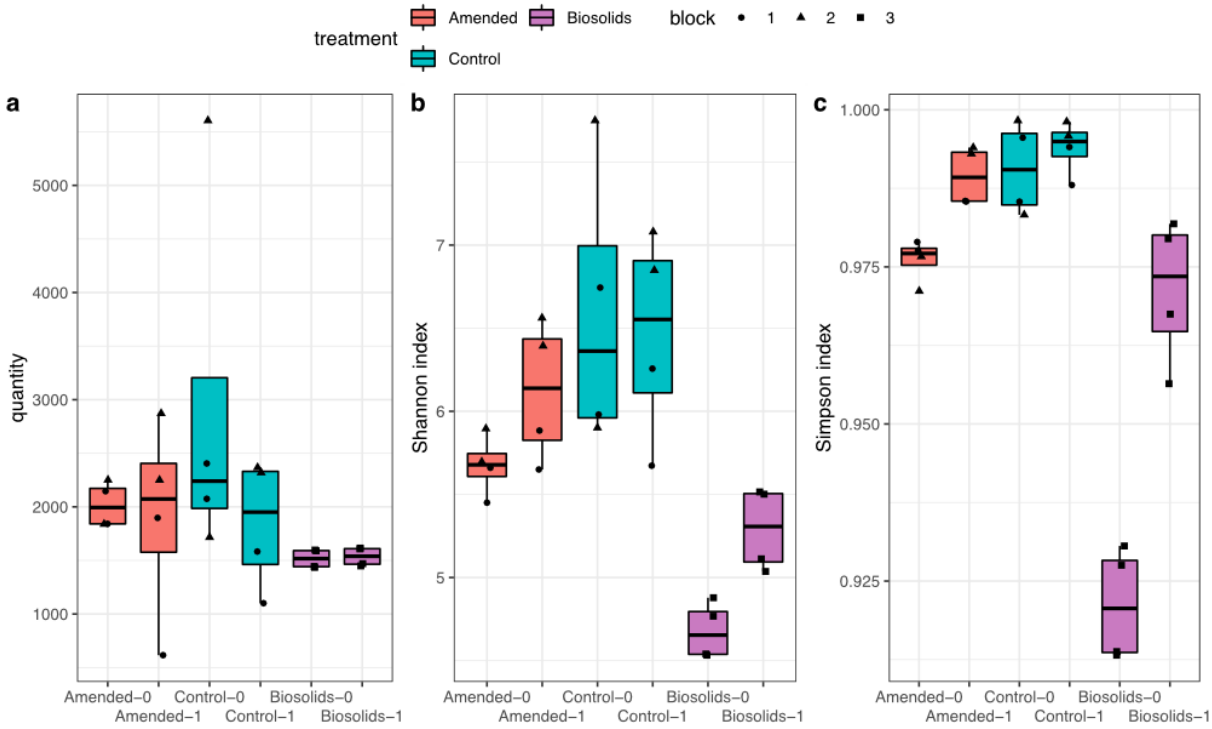


Figure B.6.: α-diversity of biosolids and amended and control soil microcosms at the start of the experiment and after one year (a - number of vOTUs, b - Shannon index, c - Simpson index).

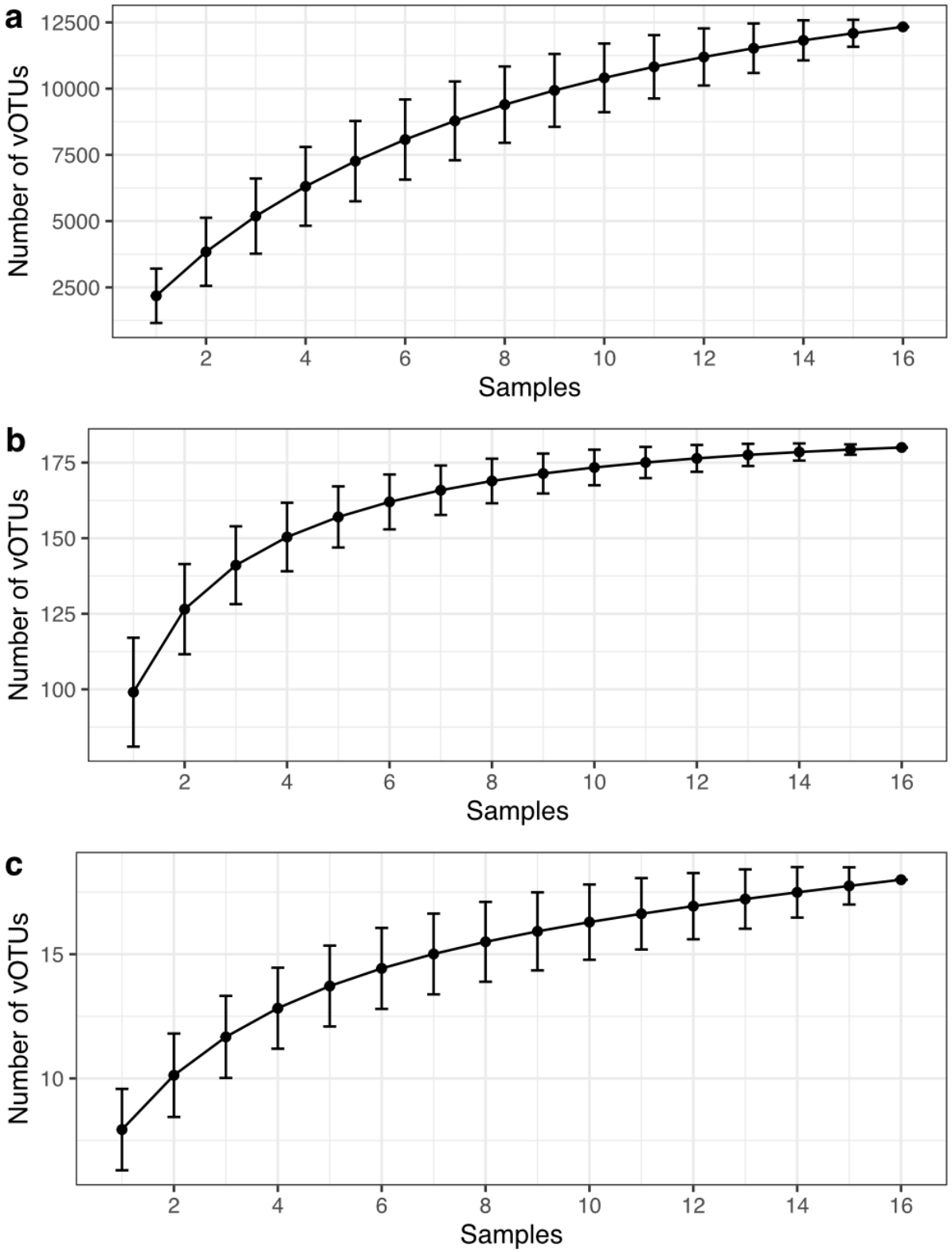


Figure B.7.: vOTU accumulation curves at (a) vOTU, (b) genus and (c) family level

C

Supplementary information for Chapter 4

C.1. Supplementary tables

Table C.1.: Sampling site descriptions

	Site	Location	Soil classification	Soil description	Elevation (m asl)	pH	Electrical conductivity (m asl)	Total carbon (%)
A	Upland peatland	53° 13' 1.22" N 4° 1' 8.78" W	Non-calcaric lithosol	Very acid upland soil with a wet highly organic topsoil	431	4.27	39	29.1
B	Upland grassland	53°13'33.00" N	Typical podzolic brown soil	Freely draining acid loamy soil over rock	289	5.89	30	11.3

	Site	Location	Soil classification	Soil description	Elevation (m asl)	pH	Electrical conductivity (m asl)	Total carbon (%)
		4° 0’ 54.86” W						
C	Semi-improved grassland	53° 13’ 55.24” N 4° 1’ 2.22” W	Typical podzolic brown soil	Freely draining slightly acid loamy soil	77	4.61	27	11.2
D	Lowland grassland	53° 14’ 10.98” N 4° 1’ 1.74” W	Typical orthic brown soil	Sandy clay loam, freely draining sheep-grazed soil	19	5.78	42	3.62
E	Coastal grassland	53° 14’ 34.17” N 4° 1’ 18.78” W	Saline alluvial gley soil	Silt-textured, poorly draining soil with periodic tidal inundation	3	8.03	1810	2.89

Soils were classified according to Avery (1990). The major properties of the sites and soils are shown in Table C.1 above, while a general description of the catena sequence is provided in Farrell et al. (2014), Shaw et al (2014) and Withers et al. (2020). The altitudinal gradient represents a primary productivity gradient with more intensive agricultural production at Site D which receives regular fertiliser (N, P and K) and lime applications. The mean annual temperature at the bottom and top sites was 10.2 and 7.3 °C respectively, while the gradient in annual rainfall was 1065 to 1690 mm, respectively. All sites had a different vegetation cover (all dominated by grasses) and were grazed by Welsh mountain sheep (*Ovis aries* L.). Soil pH and electrical conductivity were measured in 1:2.5 (w/v) soil-to-distilled water extracts using standard electrodes. Total C and N were determined on a TruSpec CN analyser (Leco Corp., St Joseph, MI). Site E is a soil developed on recent marine deposits and were classified according to Avery (1990). The major properties of the sites and soils are shown in Table S1 above, while a general contains CaCO3 from shell deposits. Its pedogenic age is ca. 500 years. All other sites have a pedogenic age of ca. 10,000 years. Site A is developed on a rhyolite parent material, Sites B and C on Ordovician age schist and shale, and Site D on mixed glacial till (rhyolite, mudstone, slate, shale, microdiorite). Sites D and E rarely undergo freezing, while Sites A-C experience periodic freezing in winter with winter snow cover often present at Site A. The vegetation at Site A comprises *Festuca ovina* L., *Juncus effusus* L. and *Trichophorum cespitosum* (L.) Hartman. The vegetation at Site B is dominated by *Agrostis canina* L., *Agrostis capillaris* L., *Anthoxanthum odoratum* L. and *Potentilla erecta* (L.) Rauschel. The vegetation at Site C is dominated by *Festuca ovina* L. and *Pteridium aquilinum* L.. Site D is dominated by *Lolium perenne* L. and *Trifolium repens* L. while the vegetation at site E is dominated by *Plantago maritima*, *Festuca* sp. and *Salicornia europaea*. Grazing intensity decreases with altitude due to the decline in primary productivity. All soils are free draining, with the exception of Site E which periodically experiences coastal inundation at spring tides (i.e. leading to the high EC values observed in Table 1) and has anaerobic features (Fe^{3+}/ Fe^{2+} mottles and FeS production). The texture of the mineral soils is as follows: Sites B to D, sandy clay loam, Site E silty sand. The humification status of the peat at Site A is H5 on the von Post scale indicating a moderate degree of humification (Ekono, 1981). Sites B and D have earthworms present, although these are much more abundant at Site D where intensive bioturbation has led to the development of a crumb structure and

Eutric Cambisol horizon structure (Ah, Bw, C). The structure at Site C is described as granular while at Site E it is classified as massive, lacking macropores. The bulk density of the soils ranged from 0.45 g cm^{-3} at Site A to 1.15 g cm^{-3} at Site E with no site showing signs of compaction that would inhibit root growth.

C.2. Supplementary figures

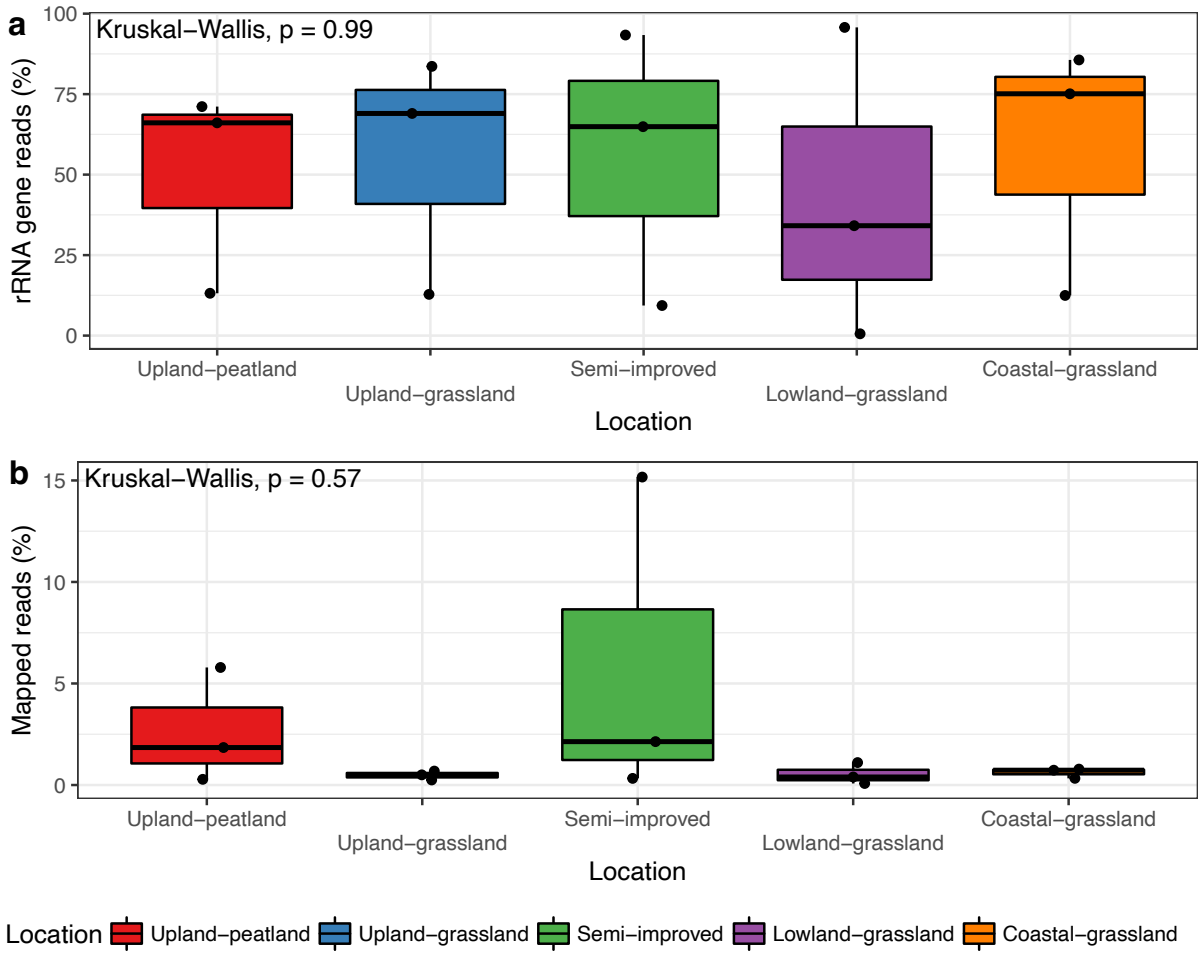


Figure C.1.: Total raw reads (a), percentage of rRNA reads (b), and total (c) and percentage (d) read pairs mapped to viral contigs at each location. No significant overall effect from location was observed on these metrics (Kruskal-Wallis) and no significant pairwise interactions were observed (all adjusted p-values >0.2). No correlations were found between raw read pairs and rRNA reads removed and the percentage of reads mapping to viral contigs (Spearman rank correlation, p = 0.119 and p = 0.611 respectively).

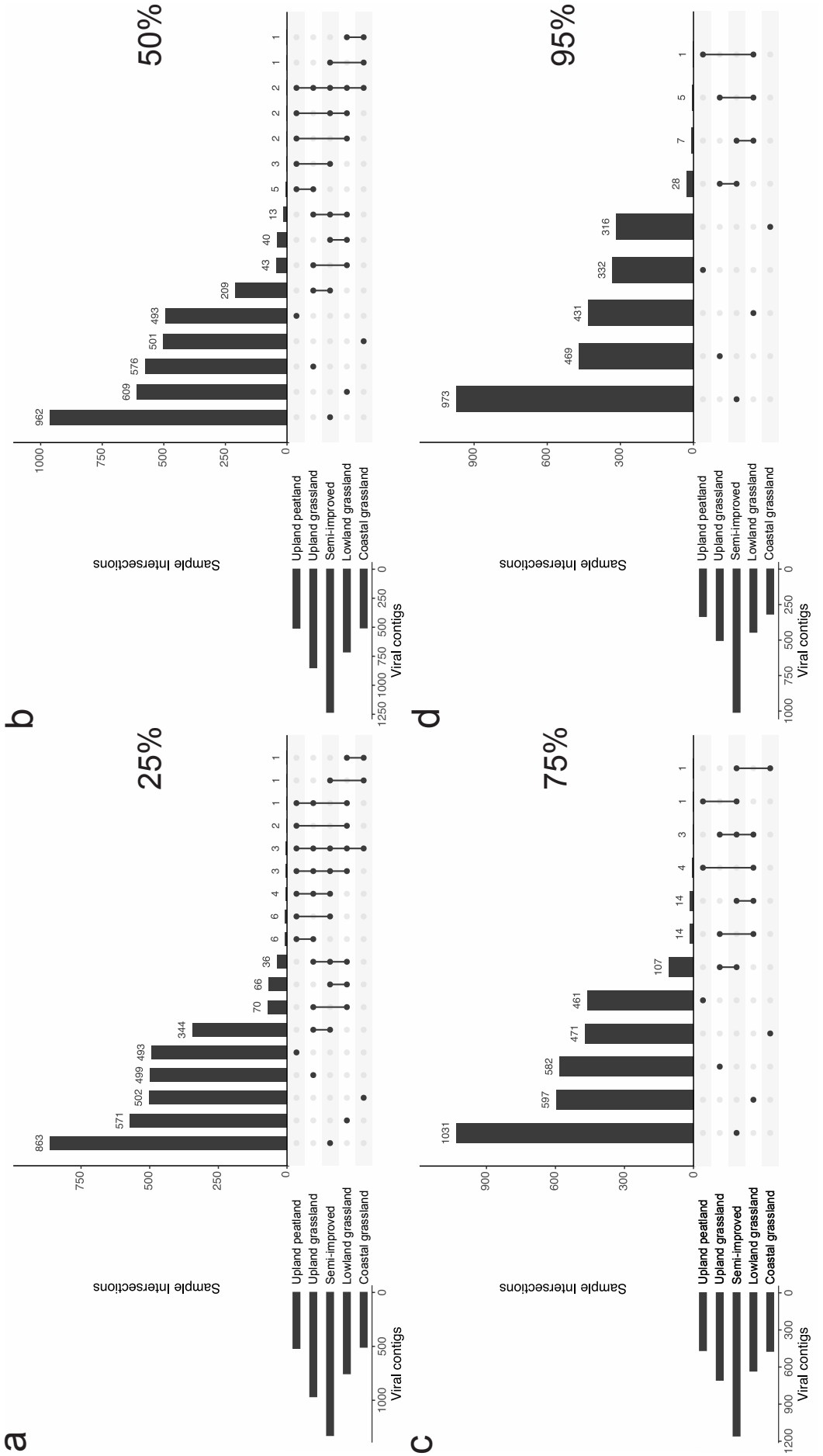


Figure C.2.: Comparison of co-occurring vOTUs at (a) 25%, (b) 50%, (c) 75% and (d) 95% horizontal genome coverage thresholds. Similar conclusions on vOTU sharing between sampling sites when varying the alignment fraction required to judge a vOTU present within a sample. vOTUs are most commonly found within one site with the grassland sites sharing more common vOTUs than other habitats. The costal grassland site consistently shared the least number of vOTUs with the other habitats.

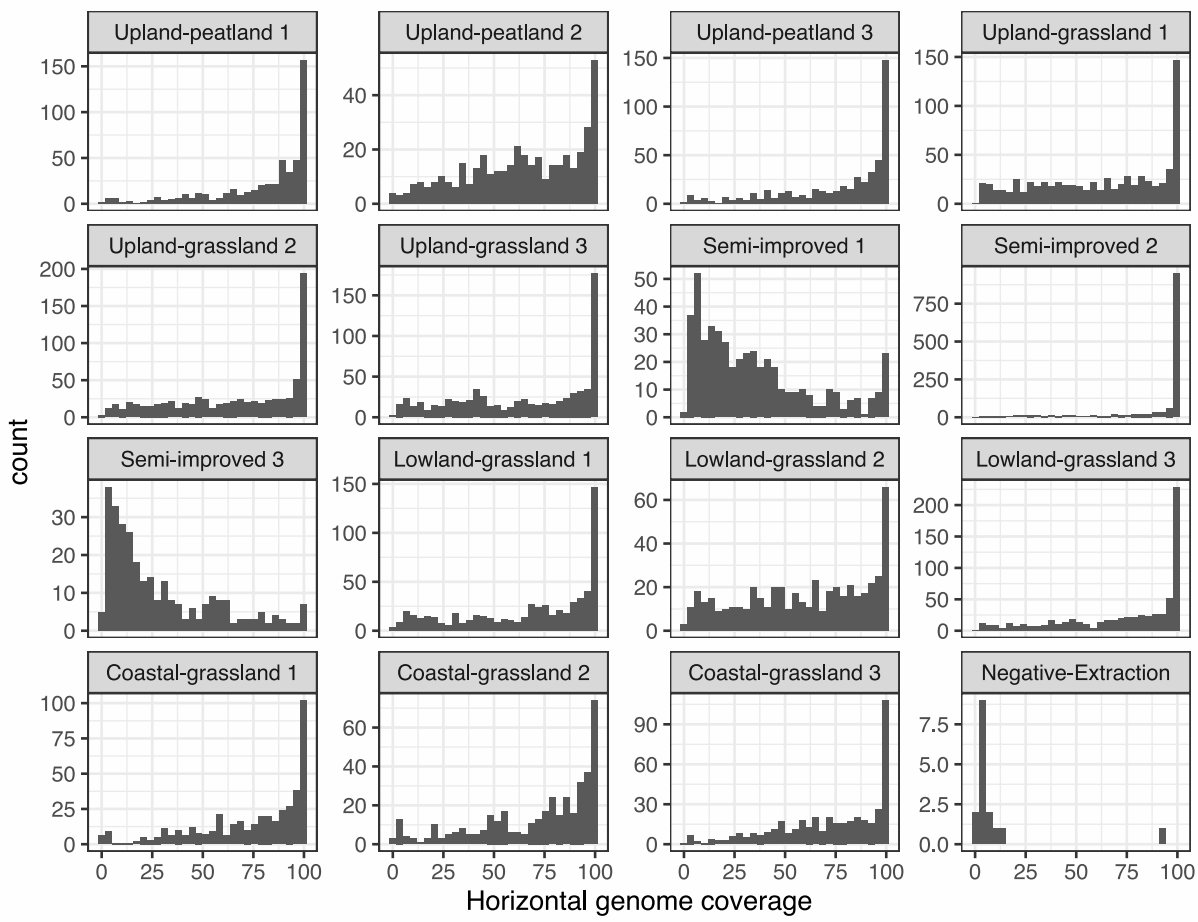


Figure C.3.: Histograms of horizontal genome coverage for all contigs where coverage was 0%. Some samples show the majority of viruses with coverages close to 100% (e.g. Upland grassland samples), whilst others show a range of values. Other samples show an increase in the number of genomes from 50-100%, whilst remaining relatively flat, or decreasing from 0-50% (e.g. Semi-improved 1). A horizontal genome coverage threshold of 50% was chosen as a compromise between preventing false-positives from short contigs being covered by one read and false negatives from longer viral contigs that may be present at low abundances in some samples.

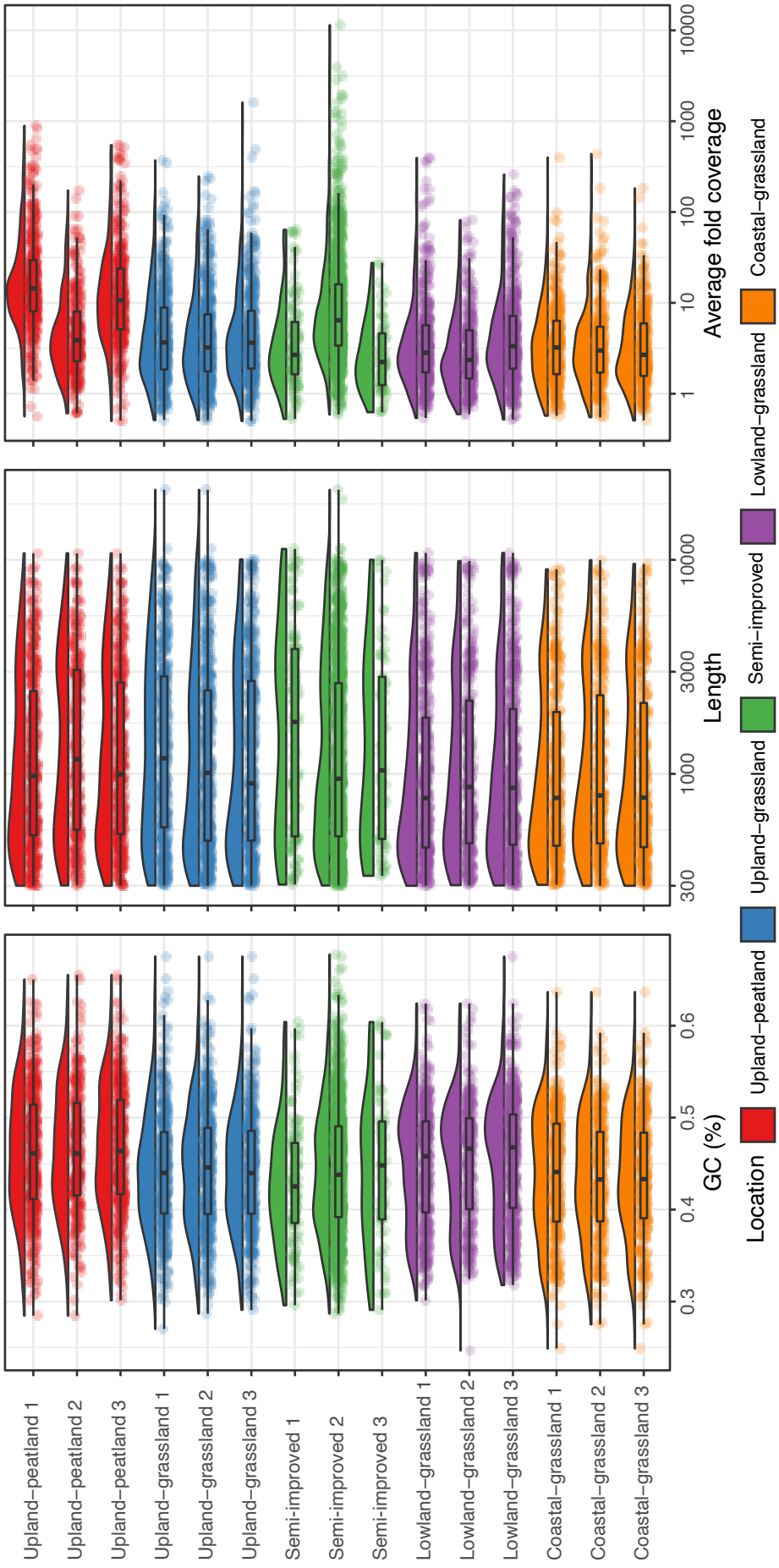


Figure C.4.: GC content, length and vertical average fold coverage of viral contigs with 50% horizontal coverage within each sample. Note that plots for length and average fold coverage are displayed on a log scale.

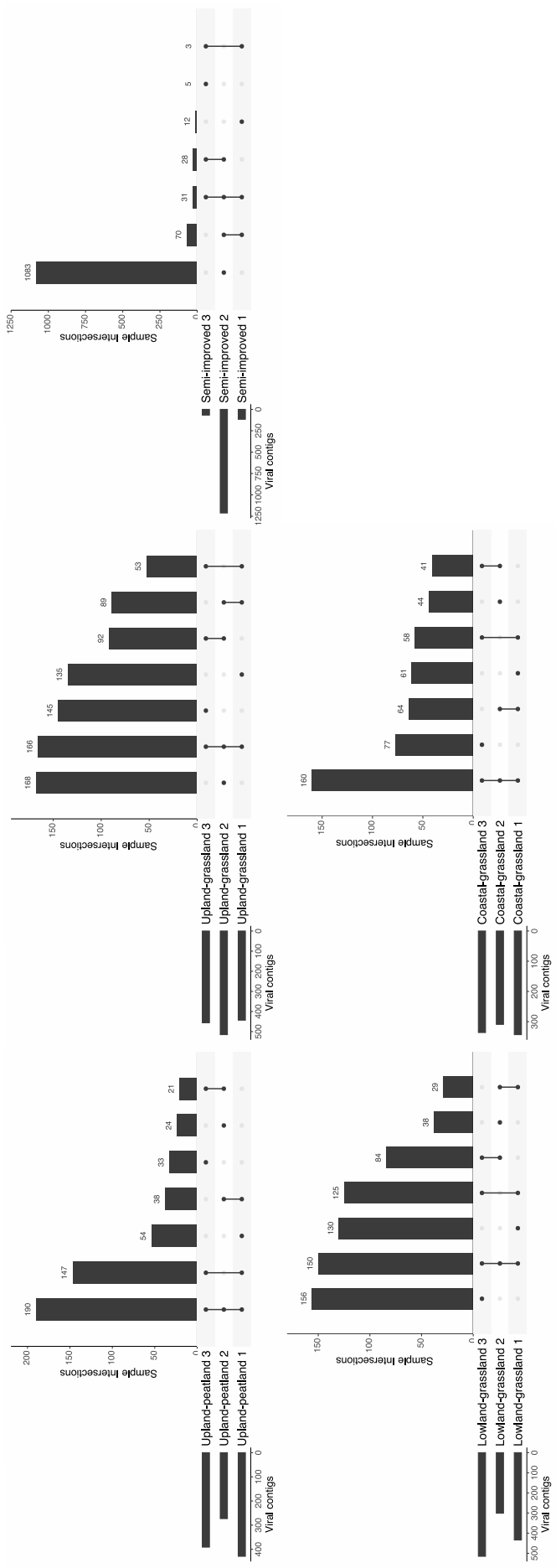


Figure C.5.: UpSet plots of viral contigs shared between sampling replicates. The intersection between all three replicates is the largest, or second largest at all sampling sites except for the semi-improved site, where the majority of viral contigs were identified in replicate 2. Although replicate 2 had fewer post-rRNA read removal, more reads aligned to viral contigs suggesting local heterogeneity in the quantity of VLPs at this site. All other samples had comparable numbers of viral contigs per sample.

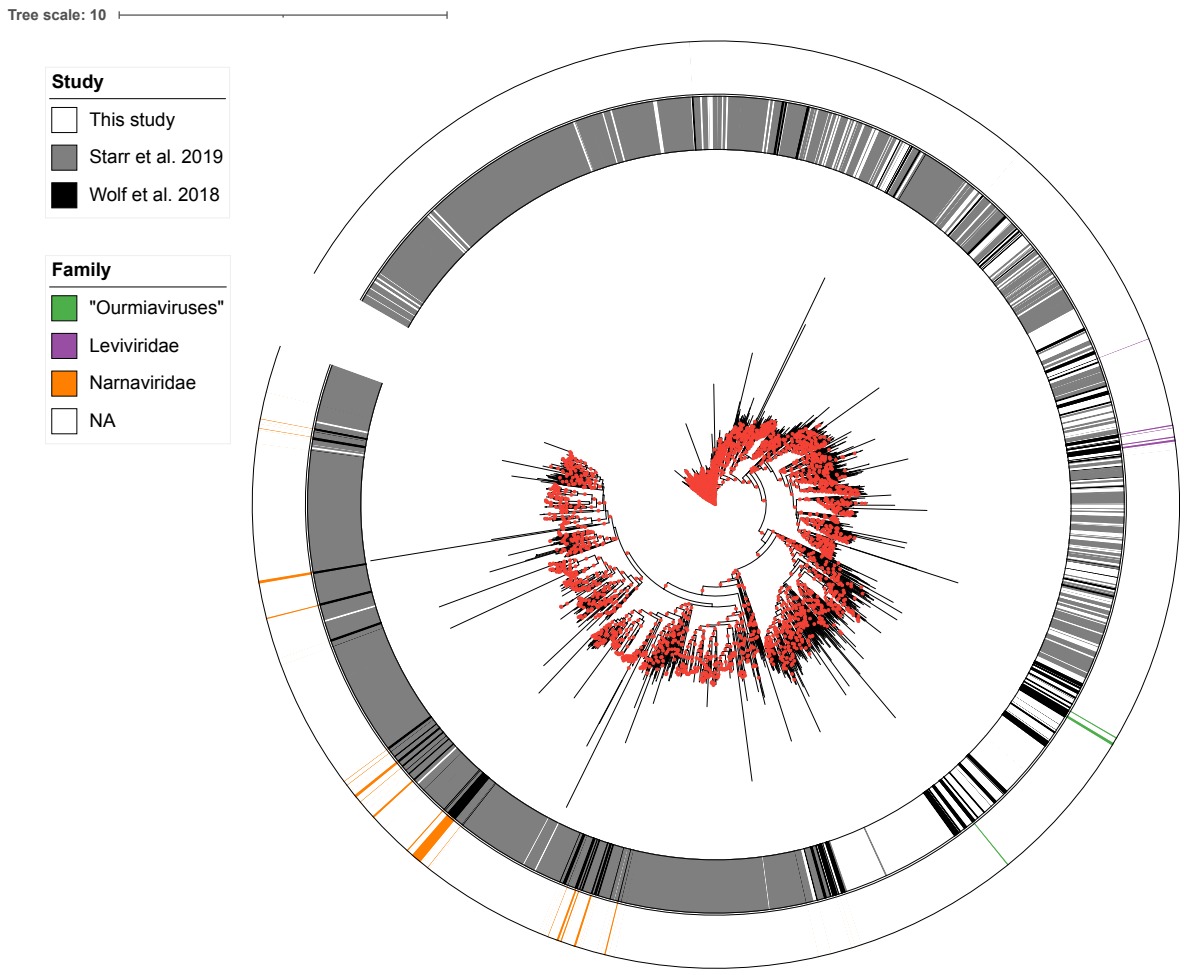


Figure C.6.: Enlarged *Lenarviricota* phylogenetic tree with branch support. Branches with branch support ≥0.6 are indicated by red circles.

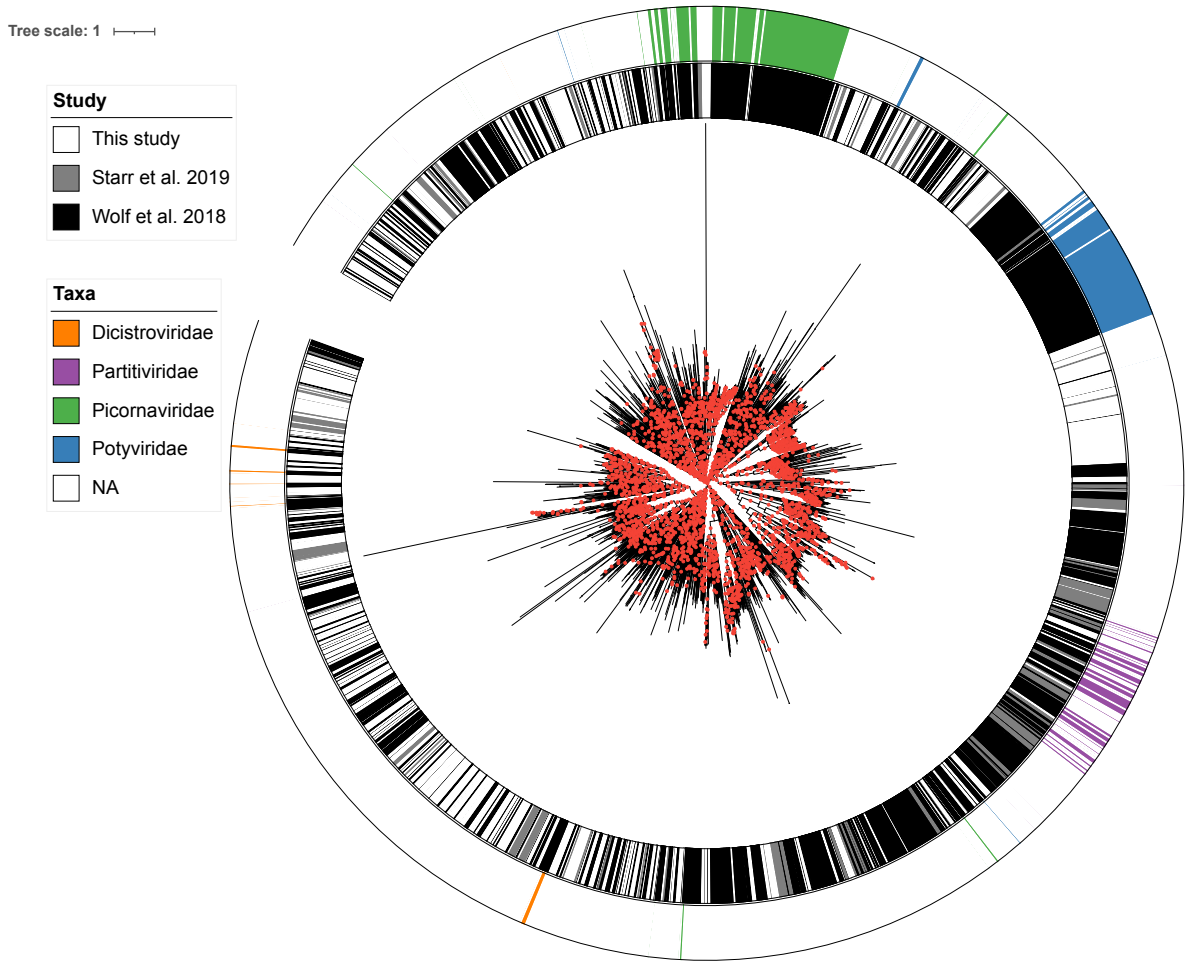


Figure C.7.: Enlarged *Pisuviricota* phylogenetic tree with branch support. Branches with branch support ≥0.6 are indicated by red circles.

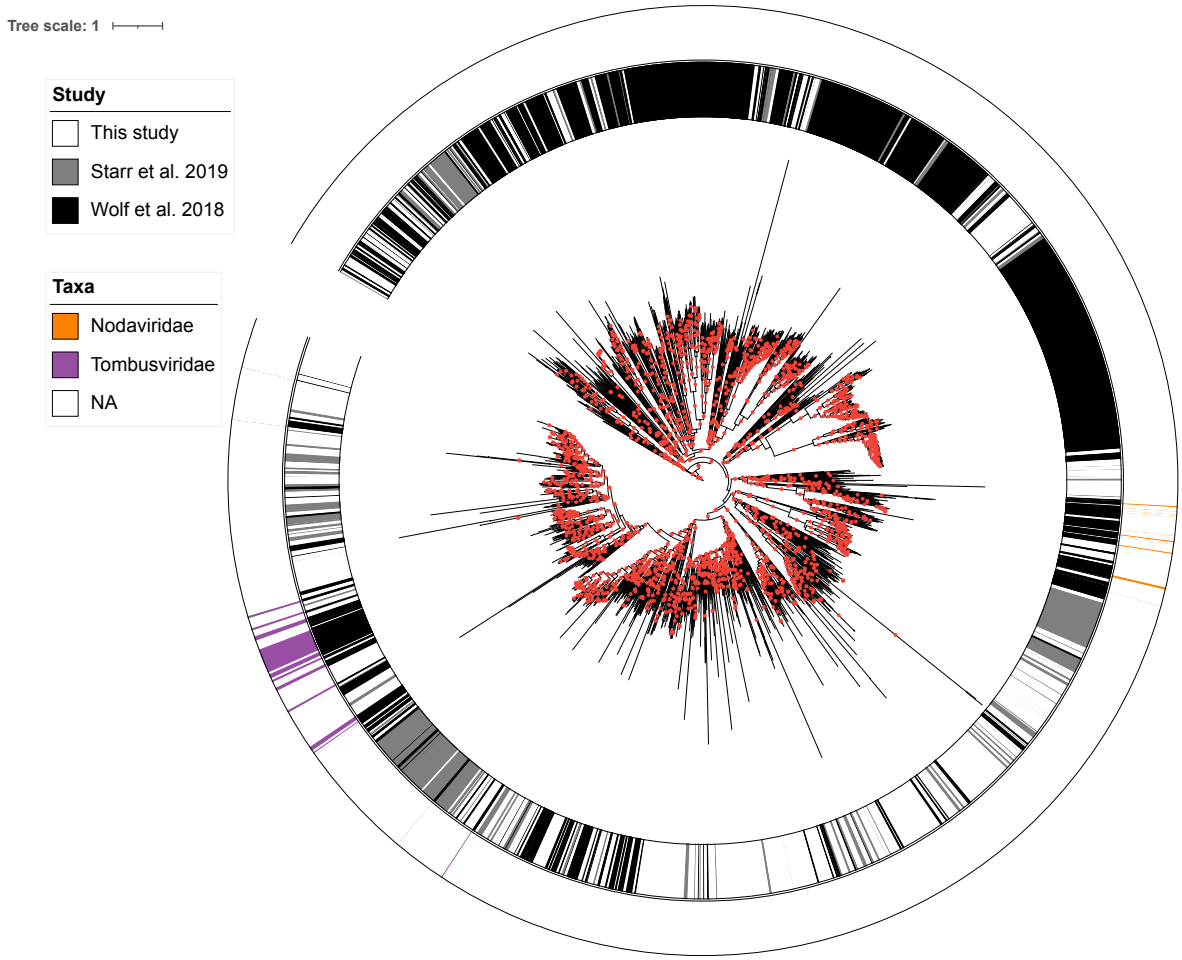


Figure C.8.: Enlarged *Kitrinoviricota* phylogenetic tree with branch support. Branches with branch support ≥0.6 are indicated by red circles.

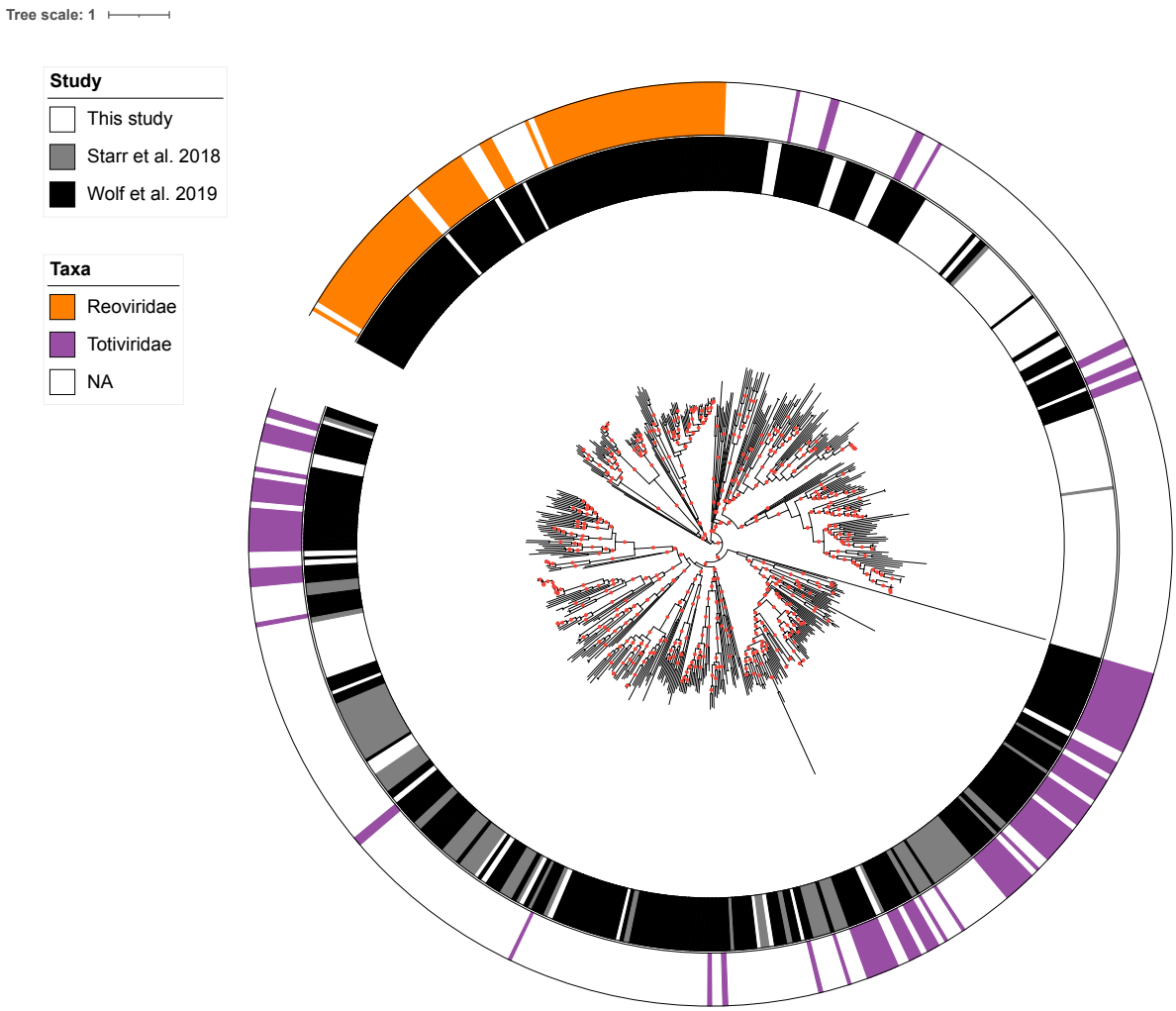


Figure C.9.: Enlarged *Duplornaviricota* phylogenetic tree with branch support. Branches with branch support ≥0.6 are indicated by red circles.

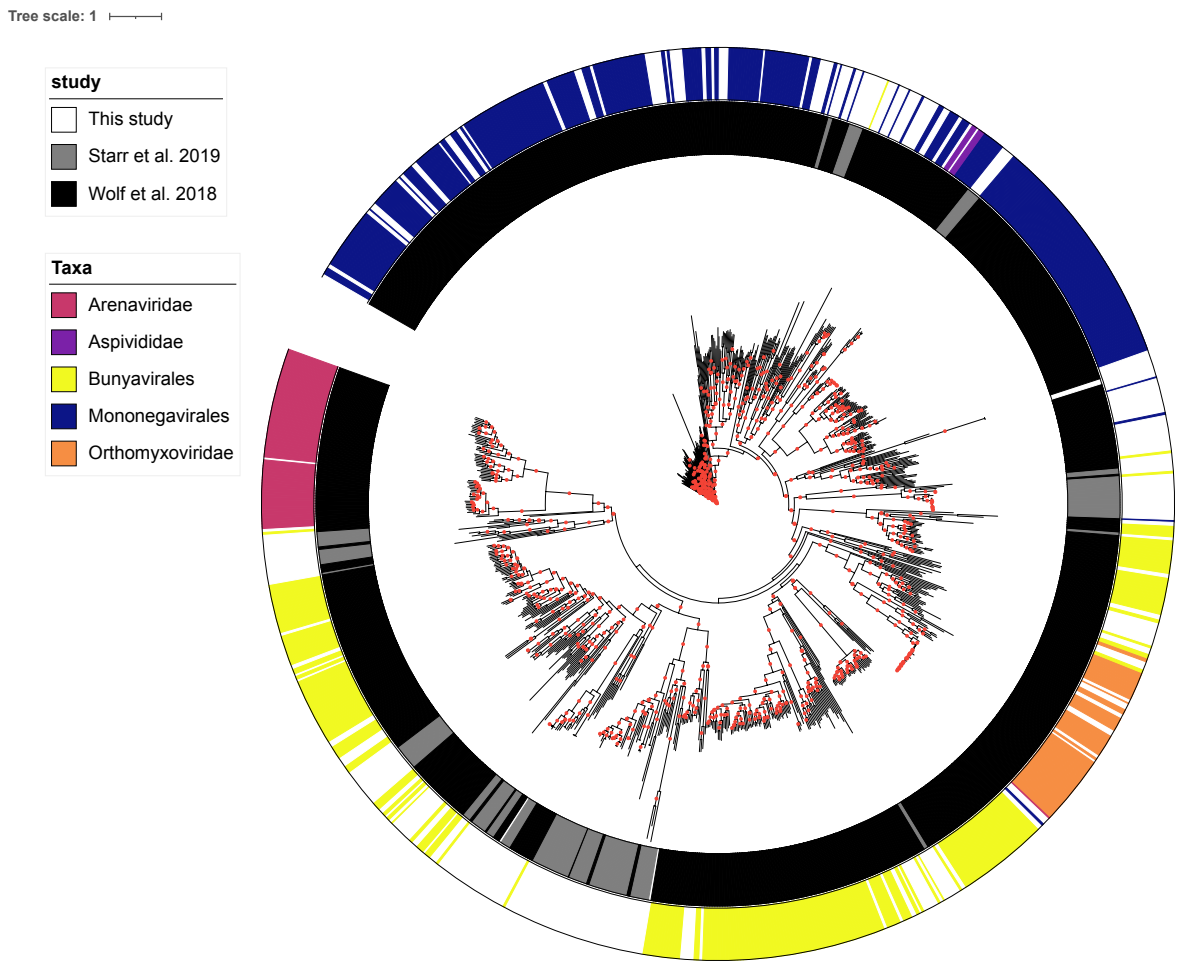


Figure C.10.: Enlarged *Negarnaviricota* phylogenetic tree with branch support. Branches with branch support ≥0.6 are indicated by red circles.

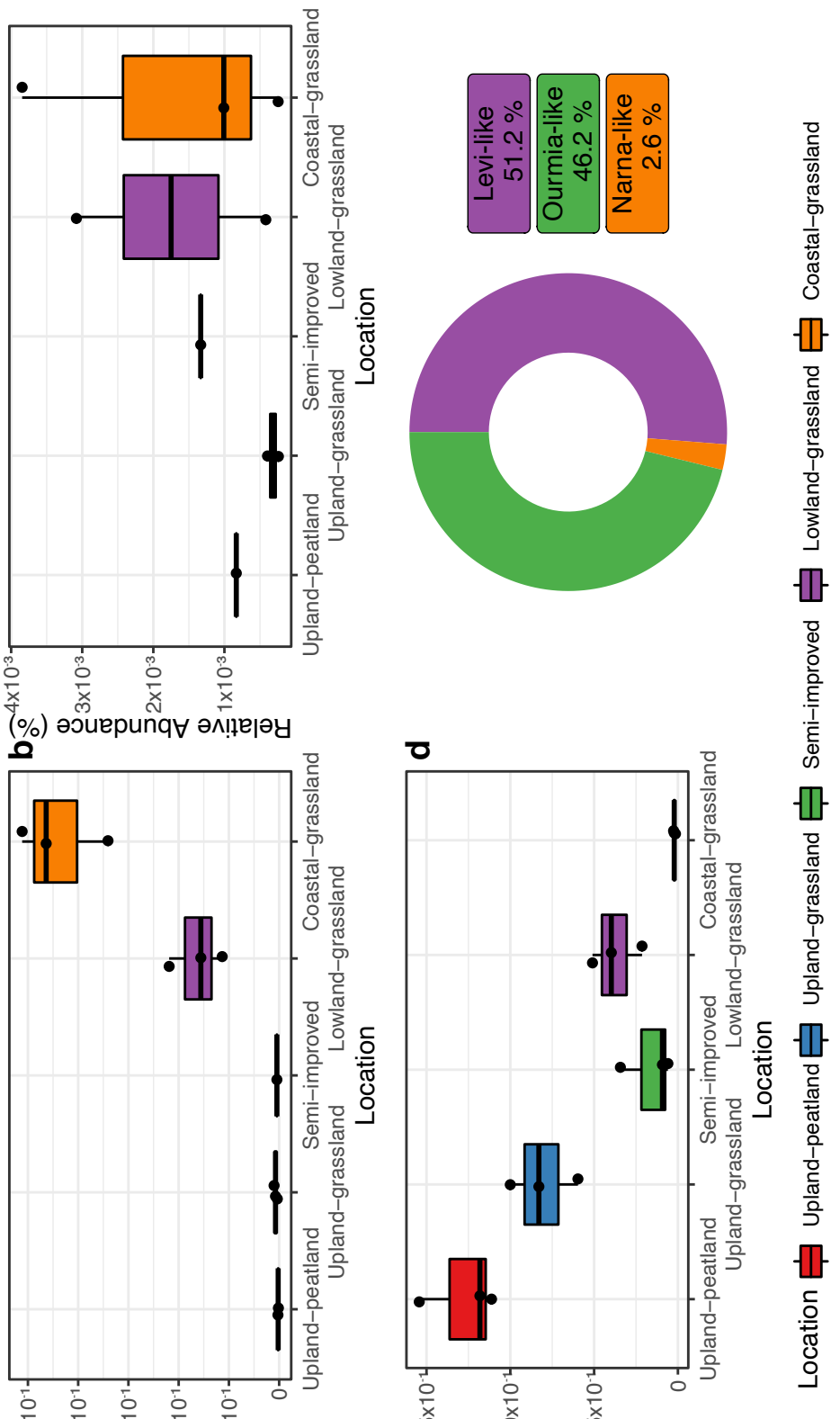


Figure C.11.: Boxplots of relative abundance of Levi-like (a), Narna-like (b) and Ourmia-like (c) viruses within the five soil-types sampled in this study. Location had a significant effect on the relative abundance of Levi-like and Ourmia-like viruses but not on Narna-like viruses (Kruskall-Wallis, $X^2 = 10.3$, 3.38, 13.2 and $p = 0.0357$, 0.497, 0.0102 respectively). (d) Doughnut plot of percentages of viral contigs from this study identified as Levi-like, Ourmia-like and Narna-like viruses.

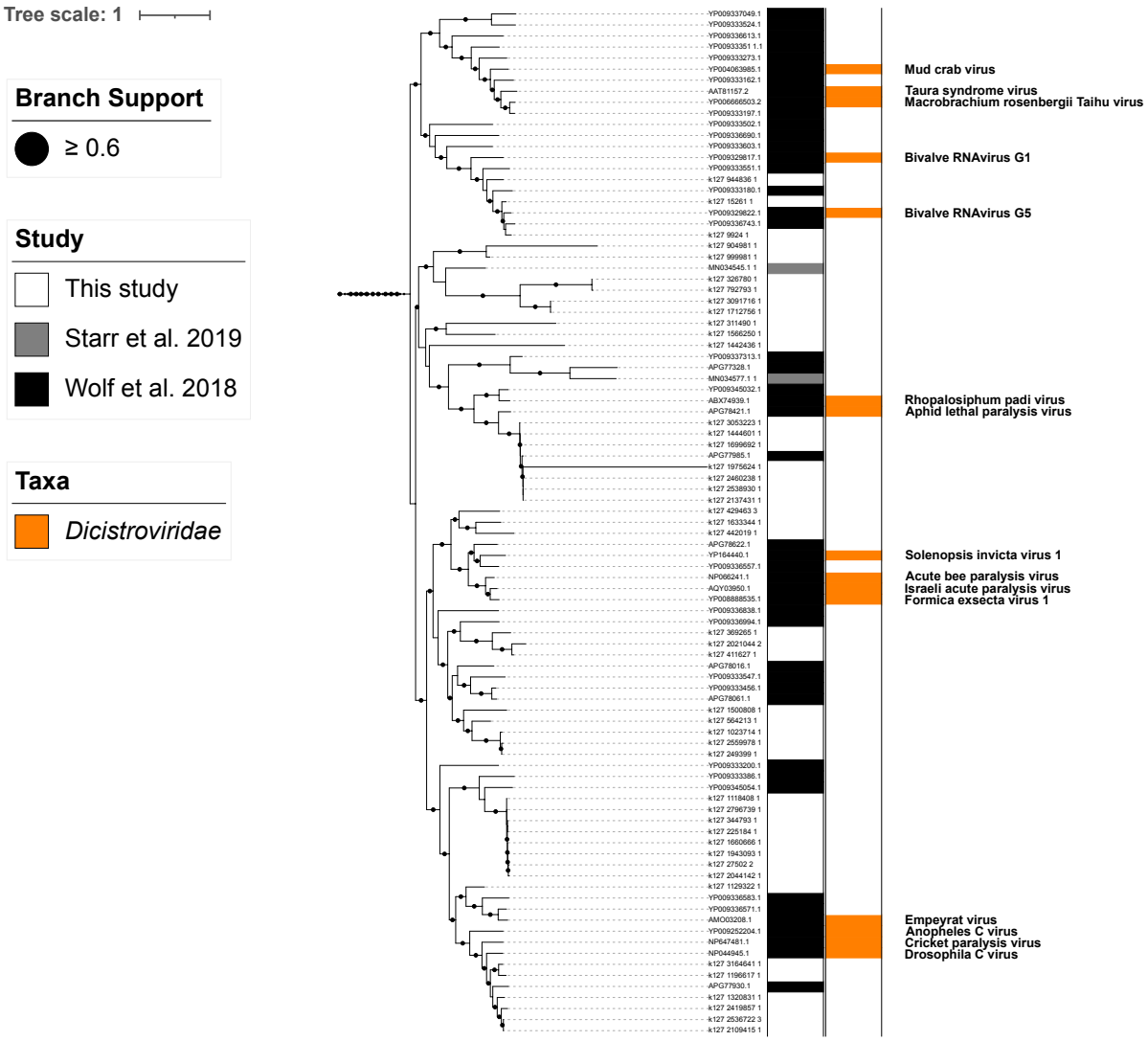


Figure C.12.: Pruned phylogenetic tree of putative dicistro-like viruses

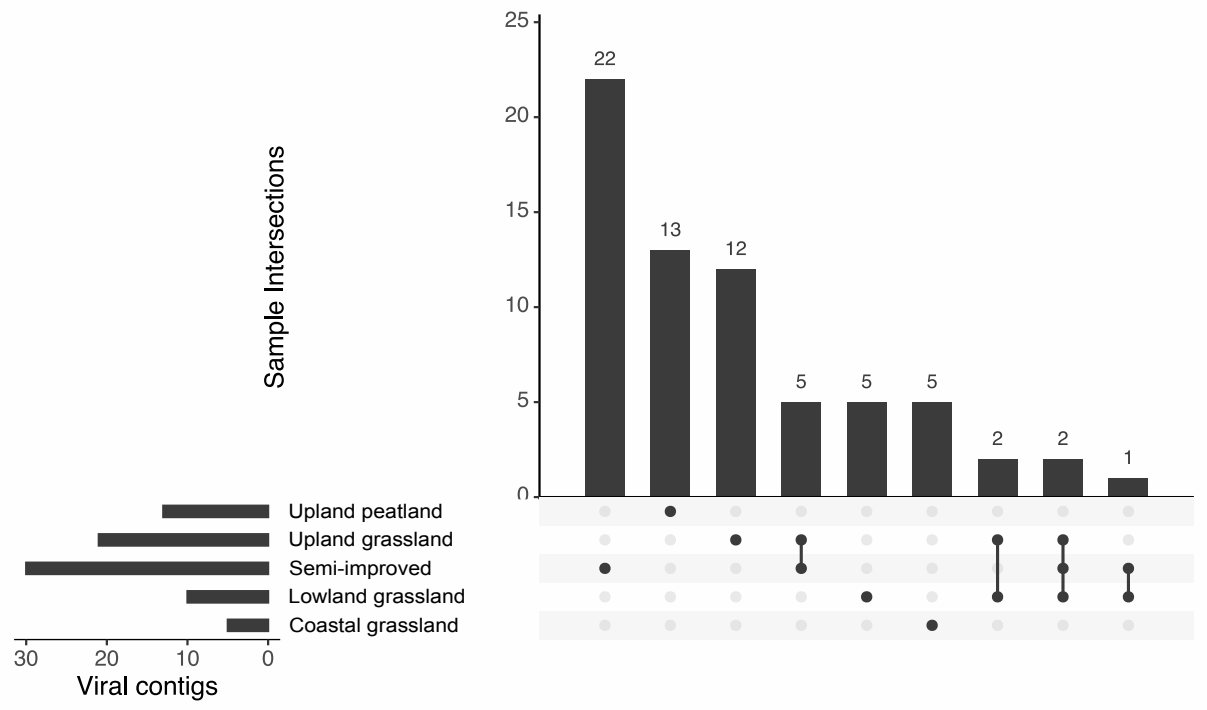


Figure C.13.: Site distribution of putative dicistro-like viruses

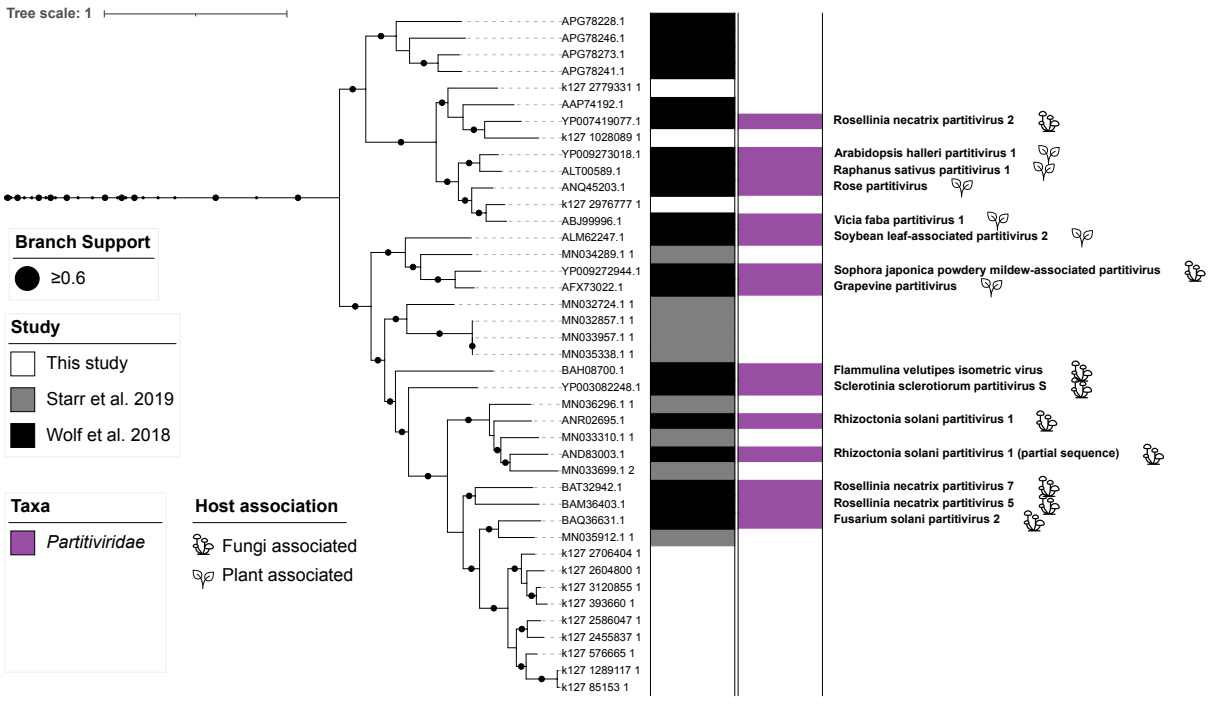


Figure C.14.: Pruned phylogenetic tree of putative partiti-like viruses. The majority of partiti-like vOTUs identified in this study are relatively closely related to *Fusarium solani* partitivirus 2 (indicated by short branch lengths at the bottom of the figure).

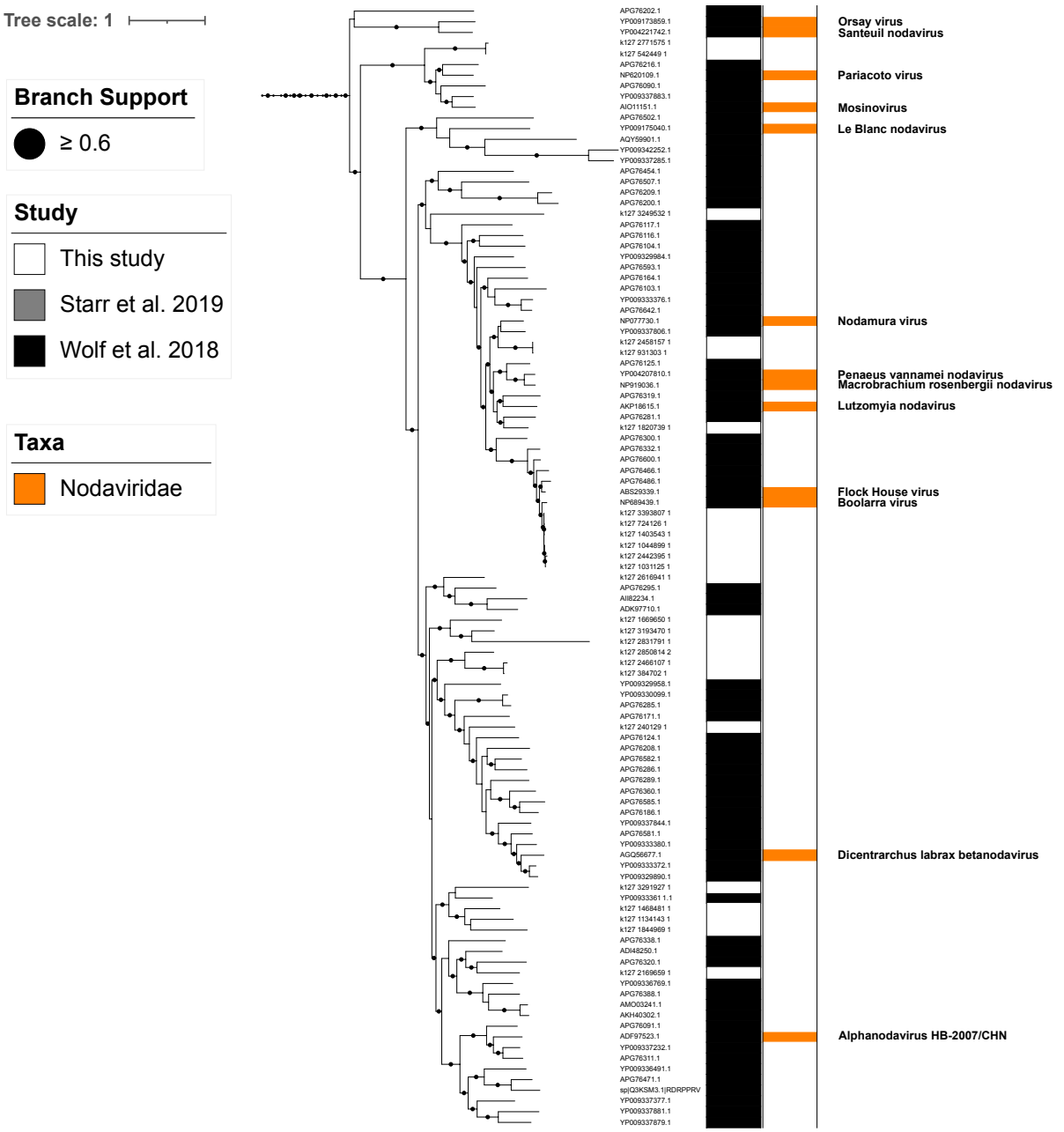


Figure C.15.: Enlarged phylogenetic tree of noda-like viruses found in this study clustering near reference *Nodaviridae* RdRP sequences.

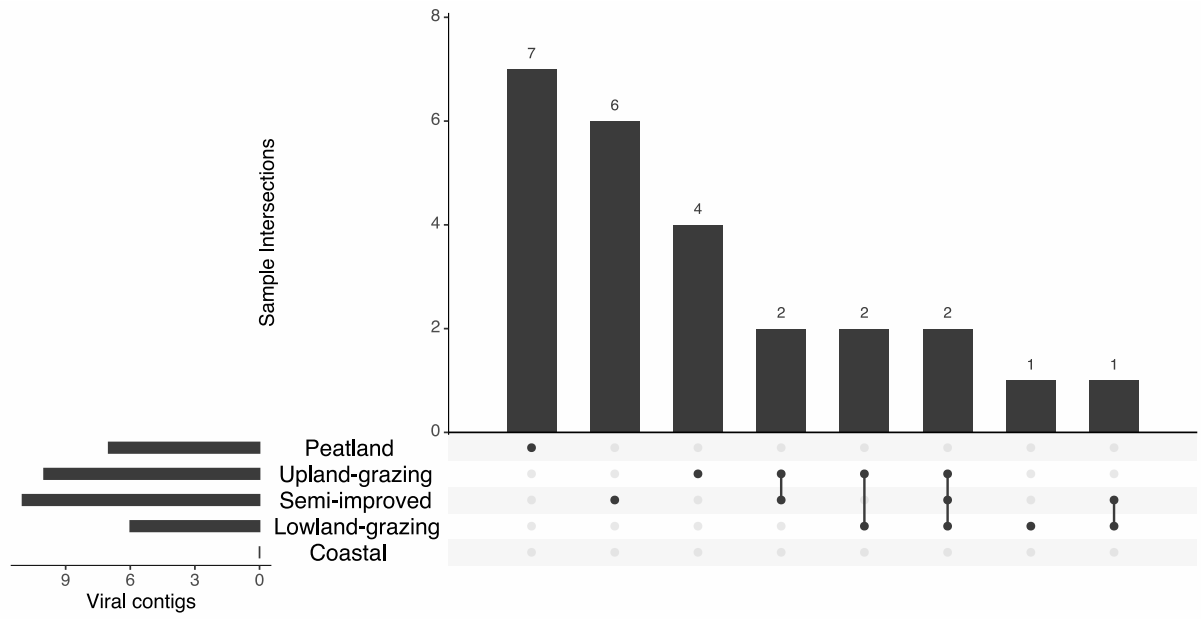


Figure C.16.: Site distribution of noda-like viruses found in this study clustering near reference *Nodaviridae* RdRP sequences. Viral sequences were predominantly found in upland areas, with no noda-like viruses found in the coastal grassland site.

D

Supplemental Information for Chapter 5

D.1. Supplementary results

D.1.1. Comparison of N1 CDC and E Sarbeco SARS-CoV-2 q(RT-)PCR assays

Significant correlation was found between SARS-CoV-2 RNA quantified by the N1 CDC and E Sarbeco gene markers in the same samples (Spearman's $\rho = 0.56$, $p < 0.0001$), however the LoD and LoQ of the E Sarbeco marker were both 2.1-fold higher than the CDC N1 assay and three times less likely to detect SARS-CoV-2 in wastewater samples see Supplementary D.2. Westhaus et al. (2021) similarly demonstrated varying sensitivity and specificity for SARS-CoV-2 in commonly used q(RT)-PCR assays and so further comparison, optimisation and standardisation is required when expanding monitoring programs and making international comparisons.

D.1.2. Detection of SARS-CoV-2 in WWTP influent suspended solids and effluent

SARS-CoV-2 RNA concentrations were also determined for the pellet from the initial centrifugation step in the first three weeks of the sampling programme. Only three samples (n = 18) produced quantities above the LoD and consequently, only results for SARS-CoV-2 in wastewater supernatants were considered in further analysis. It should be noted that SARS-CoV-2 has been detected in the solid phase in other studies (e.g. primary thickened sludge), however, the quantity of pelletable solids can be highly variable between samples and between treatment sites (Peccia et al., 2020; Westhaus et al., 2021).

Effluent samples were collected as detailed in Supplementary D.5 but similarly to sus-

pended solids, above one sample (Wrexham, 19/05/20) had detectable quantities over the LoD.

D.2. Supplementary tables

Table D.1.: Studies reporting SARS-CoV-2 RNA. p/a = presence/ absence, ct = Ct values only.

Site	Peak (gc/ 100 mL)	Reference
Published articles in peer-reviewed journals		
Netherlands (various)	2.2 × 105	(Medema et al., 2020)
England/ Wales (various)	1.5 × 104	This study
USA (Montana)	105	(Nemudryi et al., 2020)
Italy (Milan/ Rome)	5.6 × 103	(La Rosa et al., 2021)
Australia (Brisbane)	1.2 × 101	(Ahmed et al., 2020a)
India (Gujarat)	3.5 × 101	(Kumar et al., 2020)
USA (Louisiana)	7.5 × 102	(Sherchan et al., 2020)
Spain (Mercia)	ct	[Randazzo et al. (2020)
Japan (Yamanashi Prefecture)	8.2 × 103	(Haramoto et al., 2020)
France (Montpellier)	8 × 104	(Trottier et al., 2020)
Brazil (Rio de Janeiro)	ct	(Prado et al., 2020)
Germany (various)	2 × 103	(Westhaus et al., 2021)
USA (Virginia)	104	(Gonzalez et al., 2020)
Reports hosted on preprint servers		
France (Paris)	106	(Wurtzer et al., 2020)
India (Jaipur)	pa	(Arora et al., 2020)
Israel (various)	ct	(Bar Or et al., 2020)
Japan (Ishikawa and Toyama)	4.4 × 103	(Hata et al., 2020)
Spain (Ourense)	ct	(Balboa et al., 2020)
Spain (Barcelona)	104	(Chavarria-Miró et al., 2020)
Turkey (Istanbul)	1.8 × 103	(Kocamemi et al., 2020)
USA (Massachusetts)	2.4 × 104	(Wu et al., 2020)
USA (New York State)	1.2 × 104	(Green et al., 2020)

Table D.2.: q(RT-)PCR and qPCR assay parameters

Primer/Probe	**Sequence (5'-3')**	**Reference**	**q(RT-)PCR parameters**
SARS-CoV-2 (N1)	GACCCCAAAATCAGCGAAAT	(CDC, 2020)	55 °C – 60 min
Forward primer			95 °C – 5 min
SARS-CoV-2 (N1)	TCTGGTTACTGCCAGTTGAATCTG		45 cycles:
Reverse primer			95 °C – 15 s
SARS-CoV-2 (N1)	[FAM]ACCCCGCATTACGTTTGGTGGACC[MGB]		60 °C – 1 min
Probe*			65 °C – 1 min
SARS-CoV-2 (E)	ACAGGTACGTTAATAGTTAATAGCGT	(Corman et al., 2020)	
Forward primer			
SARS-CoV-2 (E)	ATATTGCAGCAGTACGCACACA		
Reverse primer			
SARS-CoV-2 (E)			
Probe*	[VIC]ACACTAGCCATCCTTACTGCGCTTCG[QSY]		
MNV	CCGCAGGAACGCTCAGCAG	(Kitajima et al., 2010)	
Forward primer			
MNV	GGYTGAATGGGGACGGCCTG		
Reverse primer			
MNV	[ABY]ATGAGTGATGGCGCA[QSY]		

Primer/Probe	Sequence (5'-3')	Reference	q(RT-)PCR parameters
Probe*			
CrAssphage_Q56	CAGAAGTACAAACTCCTAAAAAACGTAGAG	(Stachler et al., 2017)	98°C – 5 min
Forward primer			40 cycles:
CrAssphage_Q56	GATGACCAATAAACAAGCCATTAGC		95°C – 15 s
Reverse primer			60°C – 1 min
CrAssphage_Q56	[FAM]AATAACGATTTACGTGATGTAAC[TAMRA]		
Probe			

*Quencher was modified to be compatible with QuantStudio environment.

Table D.3.: R packages used in this work

Package Name	**Reference**
corrplot	(Wei and Simko, 2017)
cowplot	(Wilke, 2020)
data.table	(Dowle and Srinivasan, 2020)
FSA	(Ogle et al., 2020)
ggpubr	(Kassambara, 2020)
ggrepel	(Slowikowski, 2020)
Hmisc	(Harrell, 2020)
plotrix	(Lemon, 2006)
rnaturalearth	(South, 2017)
rworldmap	(South, 2011)
sf	(Pebesma, 2018)
tidyverse	(Wickham et al., 2019)
zoo	(Zeileis and Grothendieck, 2005)

Table D.4.: Comparison of SARS-CoV-2 with water quality parameters

Water quality indicator	Site effect (Kruskal-Wallis *p*-value)	Spearman's correlation with SARS-CoV-2 wastewater concentration (*p*-value)
Daily flow/ population equivalent	$<$ 2.2e-16	0.1108
NH_4^+	4.402e-05	0.8238
MRP	0.0006494	0.1462
pH	0.004882	0.8141
EC	7.178e-08	0.5206
NO_3^-	0.003202	0.06433

Table D.5.: Genome copies of SARS-CoV-2 in effluent (ND = no detection). All values except one were below the LoD (1.7 gc/μl) and all were below the LoQ (11.8 gc/μl).

Site	Sampling Date	Effluent mean SARS-CoV-2 concentration (gc/ μL of RNA extract)
Gwynedd	4/5/20	0.51125
Liverpool	11/5/20	ND
Manchester	11/5/20	ND
The Wirral	11/5/20	ND
Gwynedd	18/05/20	ND
Liverpool	18/05/20	ND
Manchester	18/05/29	0.4995
The Wirral	18/05/20	0.492
Wrexham	19/05/20	3.416
Cardiff	27/05/20	1.1365
Liverpool	26/05/20	0.687
Manchester	26/05/20	0.857
The Wirral	26/05/20	ND
Cardiff	4/6/20	0.08925
Liverpool	1/6/20	ND
Manchester	1/6/20	0.405
The Wirral	1/6/20	0.1385
Wrexham	2/6/20	1.4115
Liverpool	8/6/20	0.0965
Manchester	8/6/20	ND
The Wirral	8/6/20	ND
Wrexham	9/6/20	0.198

Table D.6.: Number of unique SNP/INDEL sites per sample. Of those SNP/INDELs we report the number and percentage of sites that match the locations of SNP/INDELs found from clinical samples and have the expected variant recorded.

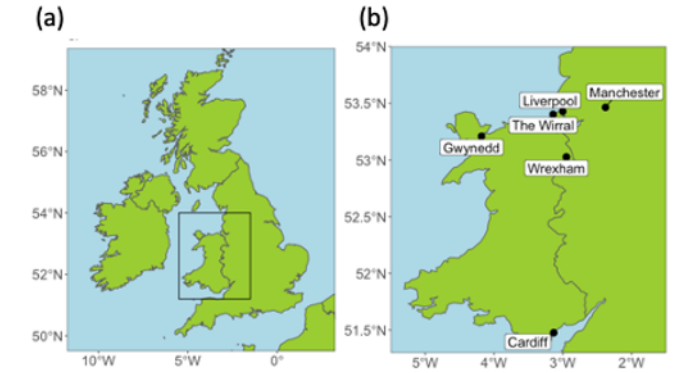
Sample	Number of unique SNP/INDEL sites	Number of sites that match locations in clinical samples	Number of sites that match the expected SNP/INDEL in clinical samples	Percentage of sites that match locations in clinical samples	Percentage of matching sites that match expected base/INDEL from clinical samples
C1WK1	15	11	5	73.33 4	5.45
C2WK1	16	8	5	50.0 62	.50
D1WK1	42	18	13	42.86 7	2.22
D2WK1	19	8	8	42.11 1	00.00
F1WK1	13	8	8	61.54 1	00.00
F2WK1	37	20	14	54.05 7	0.00
L1WK1	27	15	15	55.56 1	00.00
L2WK1	35	19	11	54.29 5	7.89
M1WK1	17	11	8	64.71 7	2.73
M2WK1	14	8	6	57.14 7	5.00
T1WK1	22	14	9	63.64 6	4.29
T2WK1	10	6	5	60.00 8	3.33
C1WK2	26	14	9	53.85 6	4.29
C2WK2	17	9	9	52.94 1	00.00
D1WK2	19	11	7	57.89 6	3.64
D2WK2	32	20	16	62.50 8	0.00
F1WK2	20	13	7	65.00 5	3.85
F2WK2	23	11	8	47.83 7	2.73
L1WK2	14	9	7	64.29 7	7.78
L2WK2	47	27	22	57.45 8	1.48
M1WK2	29	15	9	51.72 6	0.00

Sample	Number of unique SNP/INDEL sites	Number of sites that match locations in clinical samples	Number of sites that match the expected SNP/INDEL in clinical samples	Percentage of sites that match locations in clinical samples	Percentage of matching sites that match expected base/INDEL from clinical samples
M2WK2	18	8	5	44.44 6	2.50
T1WK2	15	11	5	73.33 4	5.45
T2WK2	21	15	10	71.43 6	6.67
C1WK3	14	10	8	71.43 8	0.00
C2WK3	28	13	8	46.43 6	1.54
D1WK3	24	13	12	54.17 9	2.31
D2WK3	18	10	6	55.56 6	0.00
F1WK3	16	10	7	62.50 7	0.00
F2WK3	18	9	5	50.00 5	5.56
L1WK3	25	17	11	68.00 6	4.71
L2WK3	43	24	16	55.81 6	6.67
M1WK3	17	4	4	23.53 1	00.00
M2WK3	15	8	6	53.33 7	5.00
T1WK3	17	9	6	52.94 6	6.67
T2WK3	26	16	14	61.54 8	7.50
C1WK4	18	10	7	55.56 7	0.00
C2WK4	15	12	8	80.00 6	6.67
D1WK4	21	11	8	52.38 7	2.73
D2WK4	29	18	15	62.07 8	3.33
F1WK4	27	15	13	55.56 8	6.67
F2WK4	18	10	10	55.56 1	00.00
L1WK4	7	3	3	42.86 1	00.00
L2WK4	20	8	5	40.00 6	2.50
M1WK4	20	11	10	55.00 9	0.91

Sample	Number of unique SNP/INDEL sites	Number of sites that match locations in clinical samples	Number of sites that match the expected SNP/INDEL in clinical samples	Percentage of sites that match locations in clinical samples	Percentage of matching sites that match expected base/INDEL from clinical samples
M2WK4	21	13	8	61.90 6	1.54
T1WK4	15	9	5	60.00 5	5.56
T2WK4	22	15	11	68.18 7	3.33
C1WK5	23	12	10	52.17 8	3.33
C2WK5	25	14	10	56.00 7	1.43
D1WK5	27	14	12	51.85 8	5.71
D2WK5	31	16	11	51.61 6	8.75
F1WK5	30	13	9	43.33 6	9.23
F2WK5	22	12	9	54.55 7	5.00
L1WK5	25	16	14	64.00 8	7.50
L2WK5	19	11	10	57.89 9	0.91
M1WK5	19	12	8	63.16 6	6.67
M2WK5	13	10	8	76.92 8	0.00
T1WK5	23	12	10	52.17 8	3.33
T2WK5	12	4	3	33.33 7	5.00
C1WK6	22	7	5	31.82 7	1.43
C2WK6	19	12	10	63.16 8	3.33
D1WK6	13	6	5	46.15 8	3.33
D2WK6	23	13	8	56.52 6	1.54
F1WK6	22	6	5	27.27 8	3.33
F2WK6	33	16	15	48.48 9	3.75
L1WK6	19	11	9	57.89 8	1.82
L2WK6	15	7	6	46.67 8	5.71
M1WK6	23	13	12	56.52 9	2.31

Sample	Number of unique SNP/INDEL sites	Number of sites that match locations in clinical samples	Number of sites that match the expected SNP/INDEL in clinical samples	Percentage of sites that match locations in clinical samples	Percentage of matching sites that match expected base/INDEL from clinical samples
M2WK6	11	7	5	63.64 7	1.43
T1WK6	22	10	9	45.45 9	0.00
T2WK6	17	9	7	52.94 7	7.78
C1WK7	11	6	6	54.55 1	00.00
C2WK7	17	7	6	41.18 8	5.71
D1WK7	17	8	6	47.06 7	5.00
D2WK7	21	8	6	38.10 7	5.00
F1WK7	20	12	10	60.00 8	3.33
F2WK7	26	12	9	46.15 7	5.00
L1WK7	16	13	12	81.25 9	2.31
L2WK7	21	14	13	66.67 9	2.86
M1WK7	16	12	8	75.00 6	6.67
M2WK7	22	12	10	54.55 8	3.33
T1WK7	30	16	11	53.33 6	8.75
T2WK7	27	10	7	37.04 7	0.00
F1WK13	23	15	11	65.22 7	3.33
F2WK13	43	27	20	62.79 7	4.07
G1WK13	15	11	8	73.33 7	2.73
G2WK13	11	5	5	45.45 1	00.00
H1WK13	8	4	4	50.00 1	00.00
H2WK13	16	11	9	68.75 8	1.82
T1WK13	14	5	4	35.71 8	0.00
T2WK13	7	3	3	42.86 1	00.00

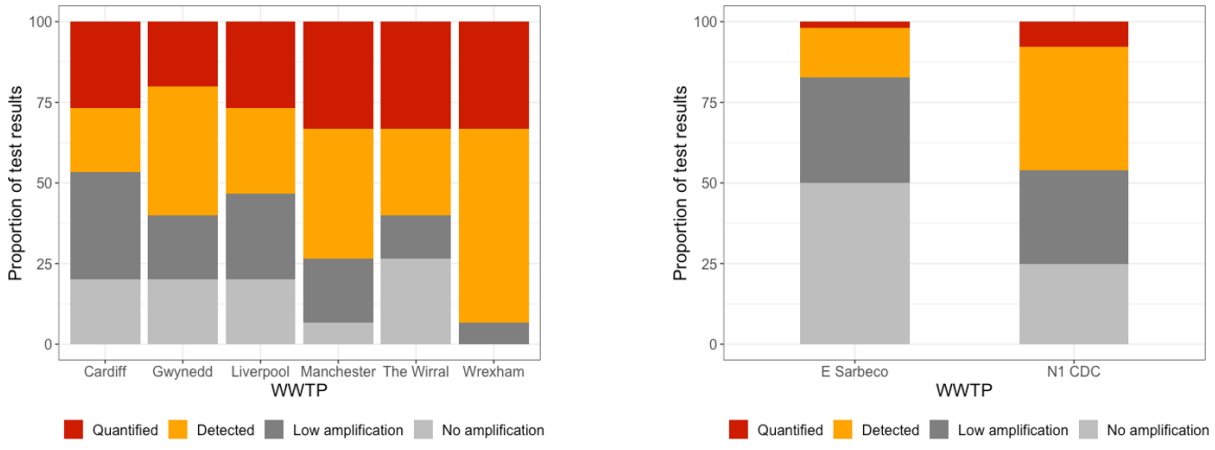
D.3. Supplementary figures



(c)

Location	Treatment	Population Equivalent
Cardiff	Sequential batch reactor	930,000
Gwynedd	Activated sludge plant	40,000
Liverpool	Activated sludge plant	628,000
Manchester	Activated sludge plant with tertiary treatment	1,170,000
The Wirral	Activated sludge plant with tertiary treatment	74,000
Wrexham	Oxidation ditch and biological filter bed	190,000

Figure D.1.: General (a) and specific (b) locations of wastewater treatment sites surveyed in this study and (c) the equivalent population sizes served. All WWTPs combine domestic, trade and stormwater



(a) N1 CDC marker

(b) E Sarbecco marker

Figure D.2.: Proportion of tests that were above LoQ and LoD for (a) SARS-CoV-2 N1-gene q(RT-)PCR assay split by site and (b) samples assayed with both N1 and E gene markers.

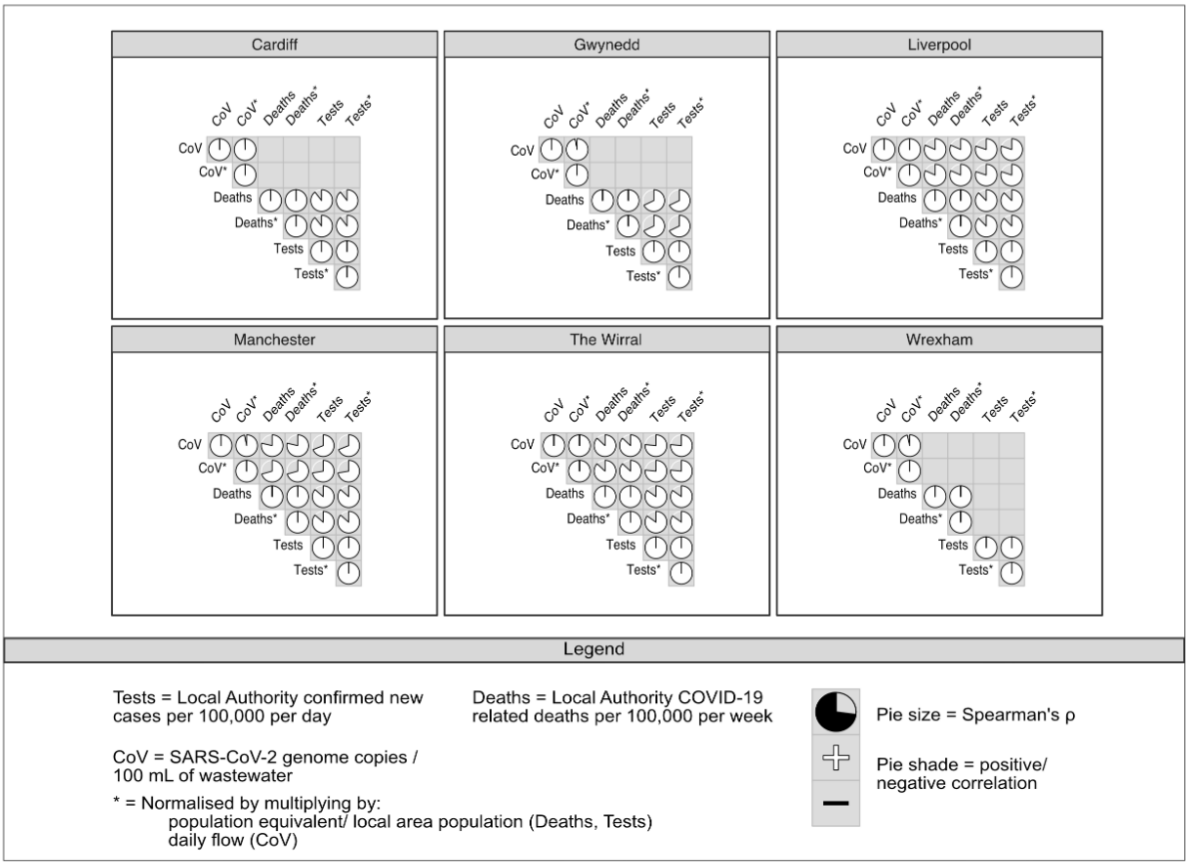


Figure D.3.: Correlations of SARS-CoV-2 genome copies/ 100 ml of wastewater with local authority daily positive tests and COVID-19 related deaths per 100,000. SARS-CoV-2 wastewater concentrations were also normalised by daily flow (*) and tests/ cases adjusted to take account of differences between sewershed population equivalents and local authority populations. These corrections had no substantial effect on correlations with only Manchester seeing a slight decrease in correlation between SARS-CoV-2 wastewater concentrations and tests/ deaths when corrected for the population size mismatch.

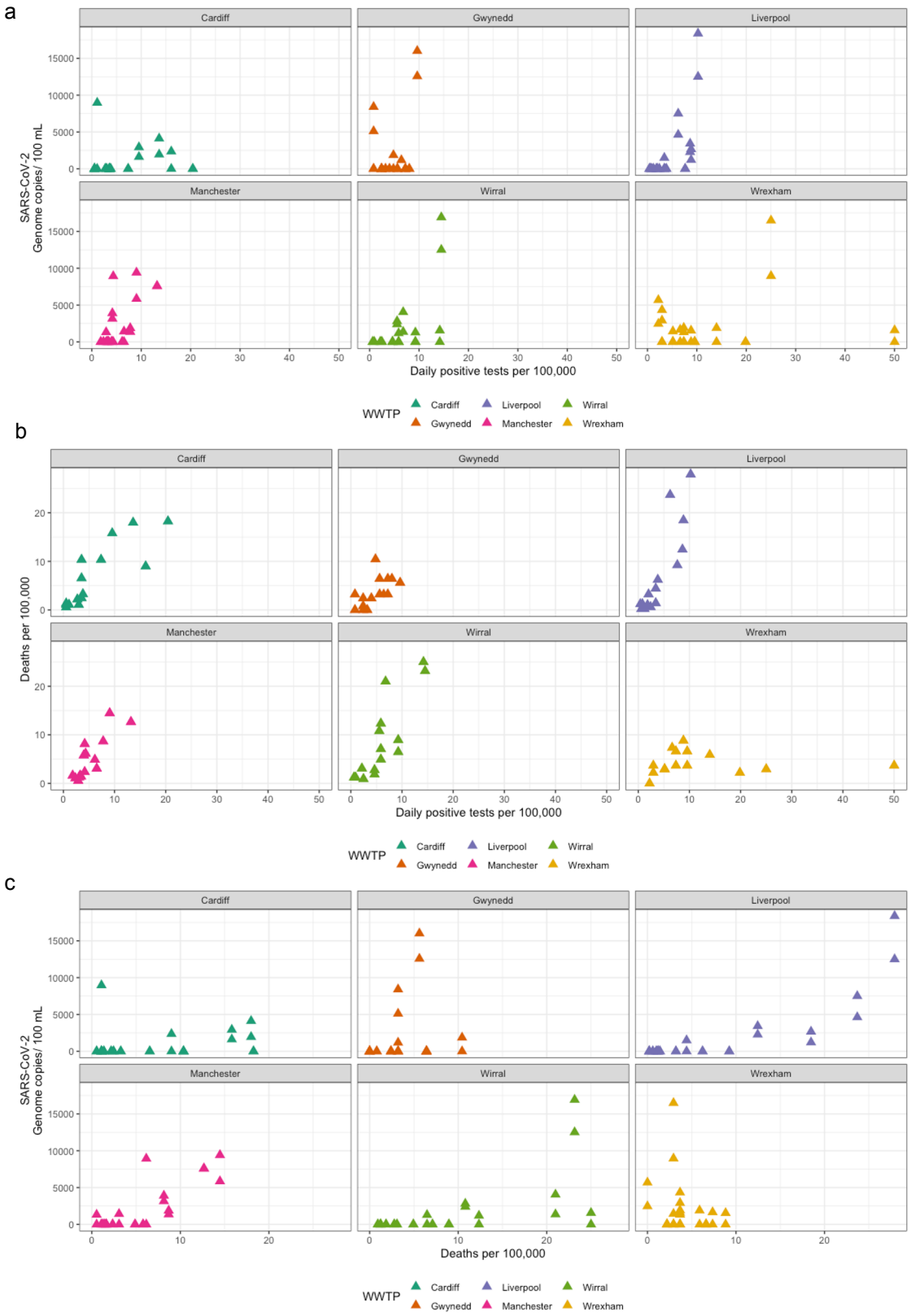


Figure D.4.: Comparisons of SARS-CoV-2 RNA wastewater concentration with daily positive tests (a), COVID-19 related deaths (b) and between tests and deaths (c).

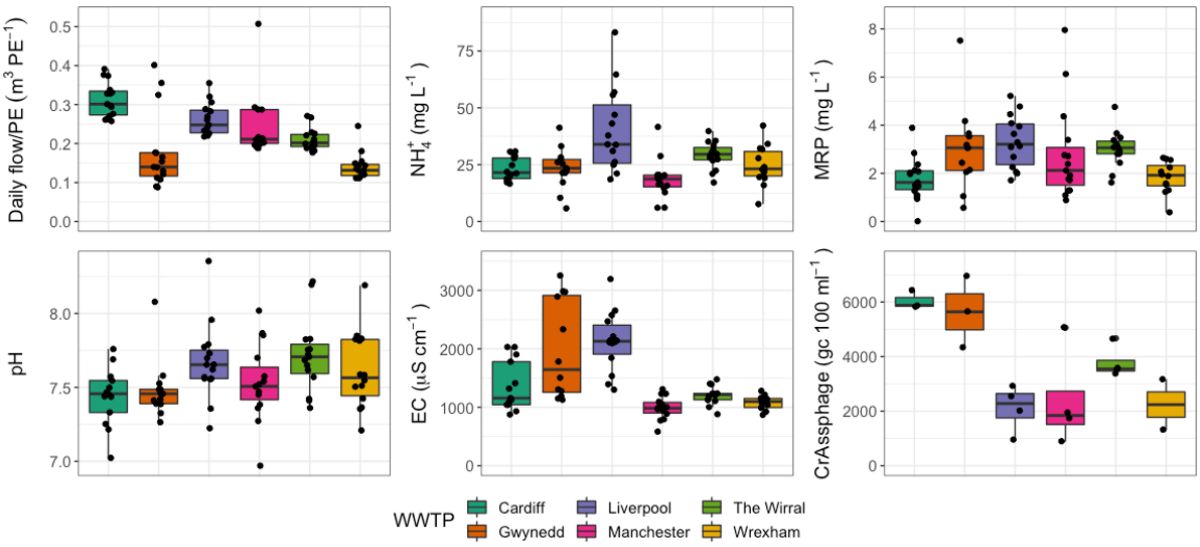


Figure D.5.: Site-specific variation in daily wastewater flow-rate (normalised by population equivalent), chemical indicators (NH_4^+, molybdate-reactive phosphate (MRP), pH, Electrical Conductivity (EC)) and a marker virus for human faecal loading (crAssphage) at six urban wastewater treatment facilities over the course of the study. Boxes are bounded on the first and third quartiles; horizontal lines denote medians. Black dots are outliers beyond the whiskers, which denote 1.5× the interquartile range.

References

Adams, M.J., Lefkowitz, E.J., King, A.M.Q., Harrach, B., Harrison, R.L., Knowles, N.J., Kropinski, A.M., Krupovic, M., Kuhn, J.H., Mushegian, A.R., Nibert, M., Sabanadzovic, S., Sanfaçon, H., Siddell, S.G., Simmonds, P., Varsani, A., Zerbini, F.M., Gorbalenya, A.E., Davison, A.J., 2017. Changes to taxonomy and the international code of virus classification and nomenclature ratified by the international committee on taxonomy of viruses (2017). Archives of Virology 162, 2505–2538. https://doi.org/10.1007/s00705-017-3358-5

Adriaenssens, E.M., Farkas, K., Harrison, C., Jones, D.L., Allison, H.E., McCarthy, A.J., 2018. Viromic Analysis of Wastewater Input to a River Catchment Reveals a Diverse Assemblage of RNA Viruses. mSystems 3, e00025–18. https://doi.org/10.1128/mSystems.00025-18

Adriaenssens, E.M., Farkas, K., McDonald, J.E., Jones, D.L., Allison, H.E., McCarthy, A.J., 2021. Tracing the fate of wastewater viruses reveals catchment-scale virome diversity and connectivity. Water Research 203, 117568. https://doi.org/10.1016/J.WATRES.2021.117568

Adriaenssens, E.M., Kramer, R., Van Goethem, M.W., Makhalanyane, T.P., Hogg, I., Cowan, D.A., 2017. Environmental drivers of viral community composition in antarctic soils identified by viromics. Microbiome 5, 83. https://doi.org/10.1186/s40168-017-0301-7

Agricultural Development and Advisory Service, 2001. The safe sludge matrix, 3rd ed. Wolverhampton.

Aguirre de Carcer, D., Lopez-Bueno, A., Pearce, D.A., Alcami, A., 2015. Biodiversity and distribution of polar freshwater DNA viruses. Science Advances 1, e1400127–e1400127. https://doi.org/10.1126/sciadv.1400127

Aherfi, S., La Scola, B., Pagnier, I., Raoult, D., Colson, P., 2014. The expanding family

marseilleviridae. Virology 466-467, 27–37. https://doi.org/10.1016/j.virol.2014.07.014

Ahmed, W., Angel, N., Edson, J., Bibby, K., Bivins, A., O'Brien, J.W., Choi, P.M., Kitajima, M., Simpson, S.L., Li, J., Tscharke, B., Verhagen, R., Smith, W.J.M., Zaugg, J., Dierens, L., Hugenholtz, P., Thomas, K.V., Mueller, J.F., 2020a. First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community. Science of The Total Environment 728, 138764. https://doi.org/10.1016/j.scitotenv.2020.138764

Ahmed, W., Bertsch, P.M., Bibby, K., Haramoto, E., Hewitt, J., Huygens, F., Gyawali, P., Korajkic, A., Riddell, S., Sherchan, S.P., Simpson, S.L., Sirikanchana, K., Symonds, E.M., Verhagen, R., Vasan, S.S., Kitajima, M., Bivins, A., 2020b. Decay of SARS-CoV-2 and surrogate murine hepatitis virus RNA in untreated wastewater to inform application in wastewater-based epidemiology. Environmental Research 191, 110092. https://doi.org/10.1016/j.envres.2020.110092

Ahmed, W., Bivins, A., Bertsch, P.M., Bibby, K., Choi, P.M., Farkas, K., Gyawali, P., Hamilton, K.A., Haramoto, E., Kitajima, M., Simpson, S.L., Tandukar, S., Thomas, K.V., Mueller, J.F., 2020c. Surveillance of SARS-CoV-2 RNA in wastewater: Methods optimization and quality control are crucial for generating reliable public health information. Current Opinion in Environmental Science & Health 17, 82–93. https://doi.org/10.1016/j.coesh.2020.09.003

Ahmed, W., Bivins, A., Bertsch, P.M., Bibby, K., Gyawali, P., Sherchan, S.P., Simpson, S.L., Thomas, K.V., Verhagen, R., Kitajima, M., Mueller, J.F., Korajkic, A., 2021a. Intraday variability of indicator and pathogenic viruses in 1-h and 24-h composite wastewater samples: Implications for wastewater-based epidemiology. Environmental Research 193, 110531. https://doi.org/10.1016/j.envres.2020.110531

Ahmed, W., Payyappat, S., Cassidy, M., Besley, C., Power, K., 2018. Novel crAssphage marker genes ascertain sewage pollution in a recreational lake receiving urban stormwater runoff. Water Research 145, 769–778. https://doi.org/10.1016/J.WATRES.2018.08.049

Ahmed, W., Tscharke, B., Bertsch, P.M., Bibby, K., Bivins, A., Choi, P., Clarke, L., Dwyer, J., Edson, J., Nguyen, T.M.H., O'Brien, J.W., Simpson, S.L., Sherman, P., Thomas, K.V., Verhagen, R., Zaugg, J., Mueller, J.F., 2021b. SARS-CoV-2 RNA

monitoring in wastewater as a potential early warning system for COVID-19 transmission in the community: A temporal case study. Science of The Total Environment 761, 144216. https://doi.org/10.1016/j.scitotenv.2020.144216

Aleta, A., Martín-Corral, D., Pastore y Piontti, A., Ajelli, M., Litvinova, M., Chinazzi, M., Dean, N.E., Halloran, M.E., Longini Jr, I.M., Merler, S., Pentland, A., Vespignani, A., Moro, E., Moreno, Y., 2020. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. Nature Human Behaviour 4, 964–971. https://doi.org/10.1038/s41562-020-0931-9

Allan, G.M., Ellis, J.A., 2000. Porcine circoviruses: A review. Journal of Veterinary Diagnostic Investigation 12, 3–14.

Anand, T., Bera, B.C., Vaid, R.K., Barua, S., Riyesh, T., Virmani, N., Hussain, M., Singh, R.K., Tripathi, B.N., 2016. Abundance of antibiotic resistance genes in environmental bacteriophages. Journal of General Virology 97, 3458–3466. https://doi.org/https://doi.org/10.1099/jgv.0.000639

Anderson, P.K., Cunningham, A.A., Patel, N.G., Morales, F.J., Epstein, P.R., Daszak, P., 2004. Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. Trends in Ecology & Evolution 19, 535–544. https://doi.org/10.1016/j.tree.2004.07.021

Andrews, S., 2010. FastQC: A quality control tool for high throughput sequence data [online]. Available online at: Http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Andrews-Pfannkoch, C., Fadrosh, D.W., Thorpe, J., Williamson, S.J., 2010. Hydroxyapatite-mediated separation of double-stranded DNA, single-stranded DNA, and RNA genomes from natural viral assemblages. Applied and Environmental Microbiology 76, 5039–5045. https://doi.org/10.1128/AEM.00204-10

Arons, M.M., Hatfield, K.M., Reddy, S.C., Kimball, A., James, A., Jacobs, J.R., Taylor, J., Spicer, K., Bardossy, A.C., Oakley, L.P., Tanwar, S., Dyal, J.W., Harney, J., Chisty, Z., Bell, J.M., Methner, M., Paul, P., Carlson, C.M., McLaughlin, H.P., Thornburg, N., Tong, S., Tamin, A., Tao, Y., Uehara, A., Harcourt, J., Clark, S., Brostrom-Smith, C., Page, L.C., Kay, M., Lewis, J., Montgomery, P., Stone, N.D., Clark, T.A., Honein, M.A., Duchin, J.S., Jernigan, J.A., 2020. Presymptomatic SARS-CoV-2 infections and transmission in a skilled nursing facility. New England Journal of Medicine 382, 2081–2090. https://doi.org/10.1056/NEJMoa2008457

Arora, S., Nag, A., Sethi, J., Rajvanshi, J., Saxena, S., Shrivastava, S.K., Gupta, A.B., 2020. Sewage surveillance for the presence of SARS-CoV-2 genome as a useful wastewater based epidemiology (WBE) tracking tool in India. medRxiv 2020.06.18.20135277. https://doi.org/10.1101/2020.06.18.20135277

Arslan, D., Legendre, M., Seltzer, V., Abergel, C., Claverie, J.-M., 2011. Distant mimivirus relative with a larger genome highlights the fundamental features of megaviridae. Proceedings of the National Academy of Sciences 108, 17486–17491. https://doi.org/10.1073/pnas.1110889108

Assured Biosolids Limited, 2020. Biosolids assurance scheme - the scheme standard (issue 5).

Avery, B.W., 1990. Soils of the British Isles. CAB International, Wallingford, UK.

Ayllón, M.A., Turina, M., Xie, J., Nerva, L., Marzano, S.Y.L., Donaire, L., Jiang, D., 2020. ICTV Virus Taxonomy Profile: Botourmiaviridae. Journal of General Virology 101, 454–455. https://doi.org/10.1099/jgv.0.001409

Bačnik, K., Kutnjak, D., Pecman, A., Mehle, N., Tušek Žnidarič, M., Gutiérrez Aguirre, I., Ravnikar, M., 2020. Viromics and infectivity analysis reveal the release of infective plant viruses from wastewater into the environment. Water Research 177, 115628. https://doi.org/10.1016/j.watres.2020.115628

Balboa, S., Mauricio-Iglesias, M., Rodríguez, S., Martínez-Lamas, L., Vasallo, F.J., Regueiro, B., Lema, J.M., 2020. The fate of SARS-CoV-2 in wastewater treatment plants points out the sludge line as a suitable spot for incidence monitoring. medRxiv. https://doi.org/10.1101/2020.05.25.20112706

Baltimore, D., 1971. Expression of animal virus genomes. Bacteriological Reviews 35, 235–241.

Bar Or, I., Yaniv, K., Shagan, M., Ozer, E., Erster, O., Mendelson, E., Mannasse, B., Shirazi, R., Kramarsky-Winter, E., Nir, O., Abu-Ali, H., Ronen, Z., Rinott, E., Lewis, Y.E., Friedler, E.F., Paitan, Y., Bitkover, E., Berchenko, Y., Kushmaro, A., 2020. Regressing SARS-CoV-2 sewage measurements onto COVID-19 burden in the population: A proof-of-concept for quantitative environmental surveillance. medRxiv. https://doi.org/10.1101/2020.04.26.20073569

Barasa, E.W., Ouma, P.O., Okiro, E.A., 2020. Assessing the hospital surge capacity of the Kenyan health system in the face of the COVID-19 pandemic. PLOS ONE 15,

e0236308. https://doi.org/10.1371/journal.pone.0236308

Barrios, E., 2007. Soil biota, ecosystem services and land productivity. Ecological Economics 64, 269–285. https://doi.org/10.1016/j.ecolecon.2007.03.004

Bernstein, D.I., 2009. Rotavirus overview. The pediatric Infectious Disease Journal 28, S50–S53. https://doi.org/10.1097/INF.0b013e3181967bee

Bhatti, A.A., Haq, S., Bhat, R.A., 2017. Actinomycetes benefaction role in soil and plant health. Microbial Pathogenesis 111, 458–467. https://doi.org/10.1016/j.micpath.2017.09.036

Bi, L., Yu, D.-T., Du, S., Zhang, L.-M., Zhang, L.-Y., Wu, C.-F., Xiong, C., Han, L.-L., He, J.-Z., 2021. Diversity and potential biogeochemical impacts of viruses in bulk and rhizosphere soils. Environmental Microbiology 23, 588–599. https://doi.org/10.1111/1462-2920.15010

Bibby, K., Peccia, J., 2013. Identification of viral pathogen diversity in sewage sludge by metagenome analysis. Environmental Science and Technology 47, 1945–1951. https://doi.org/10.1021/es305181x

Bin Jang, H., Bolduc, B., Zablocki, O., Kuhn, J.H., Roux, S., Adriaenssens, E.M., Brister, J.R., Kropinski, A.M., Krupovic, M., Lavigne, R., Turner, D., Sullivan, M.B., 2019. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. Nature Biotechnology 37, 632–639.

Bitton, G., 1975. Adsorption of viruses onto surfaces in soil and water. Water Research 9, 473–484. https://doi.org/10.1016/0043-1354(75)90071-8

Bivins, A., North, D., Wu, Z., Shaffer, M., Ahmed, W., Bibby, K., 2021. Within- and between-day variability of SARS-CoV-2 RNA in municipal wastewater during periods of varying COVID-19 prevalence and positivity. ACS ES&T Water 1, 2097–2108. https://doi.org/10.1021/acsestwater.1c00178

Blake, W., Kenneth, S., Francisco, R., Deborah, W., Anne-Kathrin, G., Luisa, Z., Jamie, S., Trevor, D., Mark, Y., 2004. Comparative genomic analysis of hyperthermophilic archaeal fuselloviridae viruses. Journal of Virology 78, 1954–1961. https://doi.org/10.1128/JVI.78.4.1954-1961.2004

Bodor, A., Bounedjoum, N., Vincze, G.E., Erdeiné Kis, Á., Laczi, K., Bende, G., Szilágyi, Á., Kovács, T., Perei, K., Rákhely, G., 2020. Challenges of unculturable bacteria: Environmental perspectives. Reviews in Environmental Science and Bio/Technology

19, 1–22. https://doi.org/10.1007/s11157-020-09522-4

Borchardt, M.A., Boehm, A.B., Salit, M., Spencer, S.K., Wigginton, K.R., Noble, R.T., 2021. The Environmental Microbiology Minimum Information (EMMI) guidelines: qPCR and dPCR quality and reporting for environmental microbiology. Environmental Science & Technology 55, 10210–10223. https://doi.org/10.1021/ACS.EST.1C01767

Breitbart, M., Bonnain, C., Malki, K., Sawaya, N.A., 2018. Phage puppet masters of the marine microbial realm. Nature Microbiology 3, 754–766. https://doi.org/10.1038/s41564-018-0166-y

Breitbart, M., Felts, B., Kelley, S., Mahaffy, J.M., Nulton, J., Salamon, P., Rohwer, F., 2004. Diversity and population structure of a near-shore marine-sediment viral community. Proceedings of the Royal Society B: Biological Sciences 271, 565–574. https://doi.org/10.1098/rspb.2003.2628

Breitbart, M., Rohwer, F., 2005. Here a virus, there a virus, everywhere the same virus? Trends in Microbiology 13, 278–284. https://doi.org/10.1016/j.tim.2005.04.003

Brinkman, N.E., Villegas, E.N., Garland, J.L., Keely, S.P., 2018. Reducing inherent biases introduced during DNA viral metagenome analyses of municipal wastewater. PLOS ONE 13, 1–23. https://doi.org/10.1371/journal.pone.0195350

Buchfink, B., Xie, C., Huson, D.H., 2014. Fast and sensitive protein alignment using DIAMOND. https://doi.org/10.1038/nmeth.3176

Bürgmann, H., Frigon, D., H Gaze, W., M Manaia, C., Pruden, A., Singer, A.C., F Smets, B., Zhang, T., 2018. Water and sanitation: An essential battlefront in the war on antimicrobial resistance. FEMS Microbiol Ecology 94, fiy101. https://doi.org/10.1093/femsec/fiy101

Busby, J., 2015. UK shallow ground temperatures for ground coupled heat exchangers. Quarterly Journal of Engineering Geology and Hydrogeology 48, 248–260. https://doi.org/10.1144/qjegh2015-077

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J., Wittwer, C.T., 2009. The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. Clinical Chemistry 55, 611–622. https://doi.org/10.1373/CLINCHEM.2008.112797

Cai, L., Yang, Y., Jiao, N., Zhang, R., 2015. Evaluation of tangential flow filtration for the concentration and separation of bacteria and viruses in contrasting marine environments. PLOS ONE 10, e0136741. https://doi.org/10.1371/journal.pone.0136741

Callanan, J., Stockdale, S.R., Shkoporov, A., Draper, L.A., Ross, R.P., Hill, C., 2020. Expansion of known ssRNA phage genomes: From tens to over a thousand. Science Advances 6, eaay5981. https://doi.org/10.1126/sciadv.aay5981

Castiglioni, S., Thomas, K.V., Kasprzyk-Hordern, B., Vandam, L., Griffiths, P., 2014. Testing wastewater to detect illicit drugs: State of the art, potential and research needs. Science of The Total Environment 487, 613–620. https://doi.org/10.1016/j.scitotenv.2013.10.034

CDC, 2020. Research Use Only 2019-Novel Coronavirus (2019-nCoV) Real-time RT-PCR Primers and Probes.

Chakraborty, I., Maity, P., 2020. COVID-19 outbreak: Migration, effects on society, global environment and prevention. Science of The Total Environment 728, 138882. https://doi.org/10.1016/j.scitotenv.2020.138882

Chavarria-Miró, G., Anfruns-Estrada, E., Guix, S., Paraira, M., Galofré, B., Sáanchez, G., Pintó, R., Bosch, A., 2020. Sentinel surveillance of SARS-CoV-2 in wastewater anticipates the occurrence of COVID-19 cases. medRxiv. https://doi.org/10.1101/2020.06.13.20129627

Chen, L., Xun, W., Sun, L., Zhang, N., Shen, Q., Zhang, R., 2014. Effect of different long-term fertilization regimes on the viral community in an agricultural soil of Southern China. European Journal of Soil Biology 62, 121–126. https://doi.org/10.1016/J.EJSOBI.2014.03.006

Chen, M.-L., An, X.-L., Liao, H., Yang, K., Su, J.-Q., Zhu, Y.-G., 2021. Viral Community and Virus-Associated Antibiotic Resistance Genes in Soils Amended with Organic Fertilizers. Environmental Science & Technology acs.est.1c03847. https://doi.org/10.1021/ACS.EST.1C03847

Chik, A.H.S., Glier, M.B., Servos, M., Mangat, C.S., Pang, X.-L., Qiu, Y., D'Aoust, P.M., Burnet, J.-B., Delatolla, R., Dorner, S., Geng, Q., Giesy, J.P., McKay, R.M., Mulvey, M.R., Prystajecky, N., Srikanthan, N., Xie, Y., Conant, B., Hrudey, S.E., 2021. Comparison of approaches to quantify SARS-CoV-2 in wastewater using RT-

qPCR: Results and implications from a collaborative inter-laboratory study in Canada. Journal of Environmental Sciences 107, 218–229. https://doi.org/10.1016/j.jes.2021.01.029

Cirrincione, L., Plescia, F., Ledda, C., Rapisarda, V., Martorana, D., Moldovan, R.E., Theodoridou, K., Cannizzaro, E., 2020. COVID-19 Pandemic: Prevention and protection measures to be adopted at the workplace. Sustainability 12, 3603. https://doi.org/10.3390/su12093603

Claverie, J.-M., Abergel, C., 2018. Mimiviridae: An expanding family of highly diverse large dsDNA viruses infecting a wide phylogenetic range of aquatic eukaryotes. Viruses 10, 506.

Cobbin, J.C.A., Charon, J., Harvey, E., Holmes, E.C., Mahar, J.E., 2021. Current challenges to virus discovery by meta-transcriptomics. Current Opinion in Virology 51, 48–55. https://doi.org/10.1016/j.coviro.2021.09.007

Collingro, A., Köstlbacher, S., Horn, M., 2020. Chlamydiae in the environment. Trends in Microbiology 28, 877–888. https://doi.org/10.1016/j.tim.2020.05.020

Colson, P., Pagnier, I., Yoosuf, N., Fournous, G., La Scola, B., Raoult, D., 2013. "Marseilleviridae," a new family of giant viruses infecting amoebae. Archives of Virology 158, 915–920. https://doi.org/10.1007/s00705-012-1537-y

Conway, J.R., Lex, A., Gehlenborg, N., 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics 33, 2938–2940. https://doi.org/10.1093/bioinformatics/btx364

Cook, R., Hooton, S., Trivedi, U., King, L., Dodd, C.E.R., Hobman, J.L., Stekel, D.J., Jones, M.A., Millard, A.D., 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 9, 65. https://doi.org/10.1186/s40168-021-01010-3

Corman, V.M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D.K., Bleicker, T., Brünink, S., Schneider, J., Schmidt, M.L., Mulders, D.G., Haagmans, B.L., Veer, B. van der, Brink, S. van den, Wijsman, L., Goderski, G., Romette, J.-L., Ellis, J., Zambon, M., Peiris, M., Goossens, H., Reusken, C., Koopmans, M.P., Drosten, C., 2020. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Eurosurveillance 25, 2000045. https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000

045

Cornell, C.R., Zhang, Y., Van Nostrand, J.D., Wagle, P., Xiao, X., Zhou, J., 2021. Temporal changes of virus-like particle abundance and metagenomic comparison of viral communities in cropland and prairie soils. mSphere 6, e01160–20. https://doi.org/10.1128/MSPHERE.01160-20

Crank, K., Petersen, S., Bibby, K., 2019. Quantitative microbial risk assessment of swimming in sewage impacted waters using crAssphage and pepper mild mottle virus in a customizable model. Environmental Science and Technology Letters 6, 571–577. https://doi.org/10.1021/acs.estlett.9b00468

Crossman, J., Hurley, R.R., Futter, M., Nizzetto, L., 2020. Transfer and transport of microplastics from biosolids to agricultural soils and the wider environment. Science of The Total Environment 724, 138334. https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.138334

Culley, A., 2018. New insight into the RNA aquatic virosphere via viromics. Virus Research 244, 84–89. https://doi.org/10.1016/j.virusres.2017.11.008

Curtis, K., Keeling, D., Yetka, K., Larson, A., Gonzalez, R., 2020. Wastewater SARS-CoV-2 Concentration and Loading Variability from Grab and 24-Hour Composite Samples. medRxiv 2020.07.10.20150607. https://doi.org/10.1101/2020.07.10.20150607

D'Aoust, P.M., Graber, T.E., Mercier, E., Montpetit, D., Alexandrov, I., Neault, N., Baig, A.T., Mayne, J., Zhang, X., Alain, T., Servos, M.R., Srikanthan, N., MacKenzie, M., Figeys, D., Manuel, D., Jüni, P., MacKenzie, A.E., Delatolla, R., 2021. Catching a resurgence: Increase in SARS-CoV-2 viral RNA identified in wastewater 48 h before COVID-19 clinical tests and 96 h before hospitalizations. Science of The Total Environment 770, 145319. https://doi.org/10.1016/j.scitotenv.2021.145319

Danovaro, R., Dell'Anno, A., Corinaldesi, C., Magagnini, M., Noble, R., Tamburini, C., Weinbauer, M., 2008. Major viral impact on the functioning of benthic deep-sea ecosystems. Nature 454, 1084–1087. https://doi.org/10.1038/nature07268

Deiner, K., Bik, H.M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D.M., Vere, N. de, Pfrender, M.E., Bernatchez, L., 2017. Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Molecular Ecology 26, 5872–5895. https://doi.org/10.1111/mec.1435

0

DeJong, T.M., 1975. A comparison of three diversity indices based on their components of richness and evenness. Oikos 26, 222–227. https://doi.org/10.2307/3543712

Deng, L., Gregory, A., Yilmaz, S., Poulos, B.T., Hugenholtz, P., Sullivan, M.B., 2012.]Contrasting life strategies of viruses that infect photo- and heterotrophic bacteria, as revealed by viral tagging. mBio 3, e00373–12. https://doi.org/10.1128/mBio.00373-12

Di Fraia, S., Figaj, R.D., Massarotti, N., Vanoli, L., 2018. An integrated system for sewage sludge drying through solar energy and a combined heat and power unit fuelled by biogas. Energy Conversion and Management 171, 587–603. https://doi.org/10.1016/j.enconman.2018.06.018

Dickson, E., 2018. Viroids: Infectious RNA in Plants, in: Nucleic Acids in Plants. CRC Press, pp. 153–193. https://doi.org/10.4324/9781351075084-7

Dou, C., Xiong, J., Gu, Y., Yin, K., Wang, J., Hu, Y., Zhou, D., Fu, X., Qi, S., Zhu, X., Yao, S., Xu, H., Nie, C., Liang, Z., Yang, S., Wei, Y., Cheng, W., 2018. Structural and functional insights into the regulation of the lysis–lysogeny decision in viral communities. Nature Microbiology 3, 1285–1294. https://doi.org/10.1038/s41564-018-0259-7

Dror, B., Jurkevitch, E., Cytryn, E., 2020. State-of-the-art methodologies to identify antimicrobial secondary metabolites in soil bacterial communities-a review. Soil Biology and Biochemistry 147, 107838. https://doi.org/10.1016/j.soilbio.2020.107838

Duhaime, M.B., Deng, L., Poulos, B.T., Sullivan, M.B., 2012. Towards quantitative metagenomics of wild viruses and other ultra-low concentration DNA samples: a rigorous assessment and optimization of the linker amplification method. Environmental Microbiology 14, 2526–2537. https://doi.org/10.1111/j.1462-2920.2012.02791.x

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., 2014. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nature Communications 5, 1–11. https://doi.org/10.1038/ncomms5498

Dy, R.L., Rigano, L.A., Fineran, P.C., 2018. Phage-based biocontrol strategies and their application in agriculture and aquaculture. Biochemical Society Transactions 46, 1605–

1613. https://doi.org/10.1042/BST20180178

Edwards, R.A., Vega, A.A., Norman, H.M., Ohaeri, M., Levi, K., Dinsdale, E.A., Cinek, O., Aziz, R.K., McNair, K., Barr, J.J., Bibby, K., Brouns, S.J.J., Cazares, A., Jonge, P.A. de, Desnues, C., Díaz Muñoz, S.L., Fineran, P.C., Kurilshikov, A., Lavigne, R., Mazankova, K., McCarthy, D.T., Nobrega, F.L., Reyes Muñoz, A., Tapia, G., Trefault, N., Tyakht, A.V., Vinuesa, P., Wagemans, J., Zhernakova, A., Aarestrup, F.M., Ahmadov, G., Alassaf, A., Anton, J., Asangba, A., Billings, E.K., Cantu, V.A., Carlton, J.M., Cazares, D., Cho, G.-S., Condeff, T., Cortés, P., Cranfield, M., Cuevas, D.A., De la Iglesia, R., Decewicz, P., Doane, M.P., Dominy, N.J., Dziewit, L., Elwasila, B.M., Eren, A.M., Franz, C., Fu, J., Garcia-Aljaro, C., Ghedin, E., Gulino, K.M., Haggerty, J.M., Head, S.R., Hendriksen, R.S., Hill, C., Hyöty, H., Ilina, E.N., Irwin, M.T., Jeffries, T.C., Jofre, J., Junge, R.E., Kelley, S.T., Khan Mirzaei, M., Kowalewski, M., Kumaresan, D., Leigh, S.R., Lipson, D., Lisitsyna, E.S., Llagostera, M., Maritz, J.M., Marr, L.C., McCann, A., Molshanski-Mor, S., Monteiro, S., Moreira-Grez, B., Morris, M., Mugisha, L., Muniesa, M., Neve, H., Nguyen, N., Nigro, O.D., Nilsson, A.S., O'Connell, T., Odeh, R., Oliver, A., Piuri, M., Prussin II, A.J., Qimron, U., Quan, Z.-X., Rainetova, P., Ramírez-Rojas, A., Raya, R., Reasor, K., Rice, G.A.O., Rossi, A., Santos, R., Shimashita, J., Stachler, E.N., Stene, L.C., Strain, R., Stumpf, R., Torres, P.J., Twaddle, A., Ugochi Ibekwe, M., Villagra, N., Wandro, S., White, B., Whiteley, A., Whiteson, K.L., Wijmenga, C., Zambrano, M.M., Zschach, H., Dutilh, B.E., 2019. Global phylogeography and ancient evolution of the widespread human gut virus crAssphage. Nature Microbiology 4, 1727–1736. https://doi.org/10.1038/s41564-019-0494-6

Ekono, 1981. Report on energy use of peat, in: Report of the United Nations Conference on New and Renewable Sources of Energy. United Nations, New York.

Emerson, J.B., Roux, S., Brum, J.R., Bolduc, B., Woodcroft, B.J., Jang, H.B., Singleton, C.M., Solden, L.M., Naas, A.E., Boyd, J.A., Hodgkins, S.B., Wilson, R.M., Trubl, G., Li, C., Frolking, S., Pope, P.B., Wrighton, K.C., Crill, P.M., Chanton, J.P., Saleska, S.R., Tyson, G.W., Rich, V.I., Sullivan, M.B., 2018. Host-linked soil viral ecology along a permafrost thaw gradient. Nature Microbiology 3, 870–880. https://doi.org/10.1038/s41564-018-0190-y

Enault, F., Briet, A., Bouteille, L., Roux, S., Sullivan, M.B., Petit, M.-A., 2016. Phages

rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. The ISME Journal 2017 11:1 11, 237–247. https://doi.org/10.1038/ismej.2016.90

Ewels, P., Magnusson, M., Lundin, S., Käller, M., 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32, 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Farkas, K., Adriaenssens, E.M., Walker, D.I., McDonald, J.E., Malham, S.K., Jones, D.L., 2019. Critical evaluation of crAssphage as a molecular marker for human-derived wastewater contamination in the aquatic environment. Food and Environmental Virology 11, 113–119. https://doi.org/10.1007/s12560-019-09369-1

Farkas, K., Cooper, D.M., McDonald, J.E., Malham, S.K., Rougemont, A. de, Jones, D.L., 2018a. Seasonal and spatial dynamics of enteric viruses in wastewater and in riverine and estuarine receiving waters. Science of The Total Environment 634, 1174–1183. https://doi.org/10.1016/j.scitotenv.2018.04.038

Farkas, K., Hillary, L.S., Malham, S.K., McDonald, James E., Jones, D.L., 2020a. Wastewater and public health: The potential of wastewater surveillance for monitoring COVID-19. Current Opinion in Environmental Science & Health 17, 14–20. https://doi.org/10.1016/j.coesh.2020.06.001

Farkas, K., Hillary, L.S., Thorpe, J., Walker, D.I., Lowther, J.A., McDonald, J.E., Malham, S.K., Jones, D.L., 2021. Concentration and quantification of SARS-CoV-2 RNA in wastewater using polyethylene glycol-based concentration and qRT-PCR. Methods and Protocols 4, 17. https://doi.org/10.3390/mps4010017

Farkas, K., Mannion, F., Hillary, L.S., Malham, S.K., Walker, D.I., 2020b. Emerging technologies for the rapid detection of enteric viruses in the aquatic environment. Current Opinion in Environmental Science & Health 16, 1–6. https://doi.org/10.1016/j.coesh.2020.01.007

Farkas, K., Marshall, M., Cooper, D., McDonald, J.E., Malham, S.K., Peters, D.E., Maloney, J.D., Jones, D.L., 2018b. Seasonal and diurnal surveillance of treated and untreated wastewater for human enteric viruses. Environmental Science and Pollution Research 25, 33391–33401. https://doi.org/10.1007/s11356-018-3261-y

Farkas, K., McDonald, J.E., Malham, S.K., Jones, D.L., 2018c. Two-step concentration of complex water samples for the detection of viruses. Methods and Protocols 1, 35. https://doi.org/10.3390/mps1030035

Farkas, K., Walker, D.I., Adriaenssens, E.M., McDonald, James E., Hillary, L.S., Malham, S.K., Jones, D.L., 2020c. Viral indicators for tracking domestic wastewater contamination in the aquatic environment. Water Research 181, 115926.

Farrell, M., Macdonald, L.M., Hill, P.W., Wanniarachchi, S.D., Farrar, J., Bardgett, R.D., Jones, D.L., 2014. Amino acid dynamics across a grassland altitudinal gradient. Soil Biology and Biochemistry 76, 179–182. https://doi.org/10.1016/j.soilbio.2014.05.015

Fierer, N., Wood, S.A., Bueno de Mesquita, C.P., 2021. How microbes can, and cannot, be used to assess soil health. Soil Biology and Biochemistry 153, 108111. https://doi.org/10.1016/j.soilbio.2020.108111

Fischer, M.G., 2021. The virophage family lavidaviridae. Current Issues in Molecular Biology 40, 1–24.

Fitzgerald, S.F., Rossi, G., Low, A.S., McAteer, S.P., O'Keefe, B., Findlay, D., Cameron, G.J., Pollard, P., Singleton, P.T.R., Ponton, G., Singer, A.C., Farkas, K., Jones, D., Graham, D.W., Quintela-Baluja, M., Tait-Burkard, C., Gally, D.L., Kao, R., Corbishley, A., 2021. Site specific relationships between COVID-19 cases and SARS-CoV-2 viral load in wastewater treatment plant influent. Environmental Science & Technology 55, 15276–15286. https://doi.org/10.1021/acs.est.1c05029

Fong, T.-T., Lipp, E.K., 2005. Enteric viruses of humans and animals in aquatic environments: Health risks, detection, and potential water quality assessment tools. Microbiology and Molecular Biology Reviews 69, 357–371. https://doi.org/10.1128/MMBR.69.2.357-371.2005

Fuhrman, J.A., 1999. Marine viruses and their biogeochemical and ecological effects. Nature 399, 541–548. https://doi.org/10.1038/21119

Gerba, C.P., Pepper, I.L., Whitehead, I., L. F., 2002. A risk assessment of emerging pathogens of concern in the land application of biosolids. Water Science and Technology 46, 225–230. https://doi.org/10.2166/wst.2002.0338

Gerba, C.P., Smith, J.E., 2005. Sources of pathogenic microorganisms and their fate during land application of wastes. Journal of Environmental Quality 34, 42–48. https://doi.org/https://doi.org/10.2134/jeq2005.0042a

Germain, P.-L., Vitriolo, A., Adamo, A., Laise, P., Das, V., Testa, G., 2016. RNAontheBENCH: computational and empirical resources for benchmarking RNAseq quan-

tification and differential expression methods. Nucleic Acids Research 44, 5054–5067. https://doi.org/10.1093/nar/gkw448

Ghebremedhin, B., 2014. Human adenovirus: Viral pathogen with increasing importance. European Journal of Microbiology and Immunology 4, 26–33. https://doi.org/10.1556/EuJMI.4.2014.1.2

Gibbs, P.A., Chambers, B.J., Chaudri, A.M., McGrath, S.P., Carlton-Smith, C.H., Bacon, J.R., Campbell, C.D., Aitken, M.N., 2006. Initial results from a long-term, multi-site field study of the effects on soil fertility and microbial activity of sludge cakes containing heavy metals. Soil Use and Management 22, 11–21. https://doi.org/10.1111/j.1475-2743.2006.00003.x

Gonzalez, R., Curtis, K., Bivins, A., Bibby, K., Weir, M.H., Yetka, K., Thompson, H., Keeling, D., Mitchell, J., Gonzalez, D., 2020. COVID-19 surveillance in Southeastern Virginia using wastewater-based epidemiology. Water Research 186, 116296. https://doi.org/10.1016/j.watres.2020.116296

Goscé, L., Phillips, P.A., Spinola, P., Gupta, D.R.K., Abubakar, P.I., 2020. Modelling SARS-COV2 spread in London: Approaches to lift the lockdown. The Journal of Infection 81, 260–265. https://doi.org/10.1016/j.jinf.2020.05.037

Graham, E.B., Paez-Espino, D., Brislawn, C., Neches, R.Y., Hofmockel, K.S., Wu, R., Kyrpides, N.C., Jansson, J.K., Mcdermott, J.E., 2019. Untapped viral diversity in global soil metagenomes. https://doi.org/10.1101/583997

Green, H., Wilder, M., Middleton, F.A., Collins, M., Fenty, A., Gentile, K., Kmush, B., Zeng, T., Larsen, D.A., 2020. Quantification of SARS-CoV-2 and cross-assembly phage (crAssphage) from wastewater to monitor coronavirus transmission within communities. medRxiv. https://doi.org/10.1101/2020.05.21.20109181

Gregory, A.C., Gerhardt, K., Zhong, Z.-P., Bolduc, B., Temperton, B., Konstantinidis, K.T., Sullivan, M.B., 2022. MetaPop: A pipeline for macro- and microdiversity analyses and visualization of microbial and viral metagenome-derived populations. Microbiome 10, 49. https://doi.org/10.1101/2020.11.01.363960

Gregory, A.C., Zayed, A.A., Conceição-Neto, N., Temperton, B., Bolduc, B., Alberti, A., Ardyna, M., Arkhipova, K., Carmichael, M., Cruaud, C., Dimier, C., Domínguez-Huerta, G., Ferland, J., Kandels, S., Liu, Y., Marec, C., Pesant, S., Picheral, M., Pisarev, S., Poulain, J., Tremblay, J.É., Vik, D., Acinas, S.G., Babin, M., Bork, P.,

Boss, E., Bowler, C., Cochrane, G., Vargas, C. de, Follows, M., Gorsky, G., Grimsley, N., Guidi, L., Hingamp, P., Iudicone, D., Jaillon, O., Kandels-Lewis, S., Karp-Boss, L., Karsenti, E., Not, F., Ogata, H., Poulton, N., Raes, J., Sardet, C., Speich, S., Stemmann, L., Sullivan, M.B., Sunagawa, S., Wincker, P., Culley, A.I., Dutilh, B.E., Roux, S., 2019. Marine DNA viral macro- and microdiversity from pole to pole. Cell 177, 1109–1123.e14. https://doi.org/10.1016/J.CELL.2019.03.040

Guo, J., Bolduc, B., Zayed, A.A., Varsani, A., Dominguez-Huerta, G., Delmont, T.O., Pratama, A.A., Gazitúa, M.C., Vik, D., Sullivan, M.B., Roux, S., 2021. VirSorter2: A multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. Microbiome 9, 37. https://doi.org/10.1186/s40168-020-00990-y

Han, L.-L., Yu, D.-T., Bi, L., Du, S., Silveira, C., Cobián Güemes, A.G., Zhang, L.-M., He, J.-Z., Rohwer, F., 2022. Distribution of soil viruses across china and their potential role in phosphorous metabolism. Environmental Microbiome 17, 6. https://doi.org/10.1186/s40793-022-00401-9

Han, L.-L., Yu, D.-T., Zhang, L.-M., Shen, J.-P., He, J.-Z., 2017. Genetic and functional diversity of ubiquitous DNA viruses in selected chinese agricultural soils. Scientific Reports 7, 45142. https://doi.org/10.1038/srep45142

Haramoto, E., Malla, B., Thakali, O., Kitajima, M., 2020. First environmental surveillance for the presence of SARS-CoV-2 RNA in wastewater and river water in Japan. Science of The Total Environment 737, 140405. https://doi.org/10.1016/j.scitotenv.2020.140405

Hata, A., Honda, R., Hara-Yamamura, H., Meuchi, Y., 2020. Detection of SARS-CoV-2 in wastewater in Japan by multiple molecular assays-implication for wastewater-based epidemiology (WBE). medRxiv. https://doi.org/10.1101/2020.06.09.20126417

Hawkins, S.F.C., Guest, P.C., 2017. Multiplex analyses using real-time quantitative PCR, in: Guest, P.C. (Ed.), Multiplex Biomarker Techniques: Methods and Applications. Springer New York, New York, NY, pp. 125–133. https://doi.org/10.1007/978-1-4939-6730-8_8

He, X., Lau, E.H.Y., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y.C., Wong, J.Y., Guan, Y., Tan, X., Mo, X., Chen, Y., Liao, B., Chen, W., Hu, F., Zhang, Q., Zhong, M., Wu, Y., Zhao, L., Zhang, F., Cowling, B.J., Li, F., Leung, G.M., 2020. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nature Medicine 26,

672–675. https://doi.org/10.1038/s41591-020-0869-5

Hewitt, J., Leonard, M., Greening, G.E., Lewis, G.D., 2011. Influence of wastewater treatment process and the population size on human virus profiles in wastewater. Water Research 45, 6267–6276. https://doi.org/10.1016/J.WATRES.2011.09.029

Hillman, B.I., Cai, G., 2013. The family Narnaviridae. Simplest of RNA viruses 86, 149–176. https://doi.org/10.1016/B978-0-12-394315-6.00006-4

Hjelmso, M.H., Hellmer, M., Fernandez-Cassi, X., Timoneda, N.N., Lukjancenko, O., Seidel, M., Elsaesser, D., Aarestrup, F.M., Lofstrom, C., Bofill-Mas, S.S., Abril, J.F., Girones, R., Schultz, A.C., Hjelmsø, M.H., Hellmér, M., Fernandez-Cassi, X., Timoneda, N.N., Lukjancenko, O., Seidel, M., Elsässer, D., Aarestrup, F.M., Löfström, C., Bofill-Mas, S.S., Abril, J.F., Girones, R., Schultz, A.C., 2017. Evaluation of methods for the concentration and extraction of viruses from sewage in the context of metagenomic sequencing. PLOS ONE 12, 1–17. https://doi.org/10.1371/journal.pone.0170199

HM Government, 2020. Coronavirus (COVID-19) in the UK - Cases Legacy CSV download.

Ho, S.F.S., Millard, A.D., Schaik, W. van, 2021. Comprehensive benchmarking of tools to identify phages in metagenomic shotgun sequencing data. bioRxiv. https://doi.org/10.1101/2021.04.12.438782

Howard-Varona, C., Hargreaves, K.R., Abedon, S.T., Sullivan, M.B., 2017. Lysogeny in nature: Mechanisms, impact and ecology of temperate phages why study lysogeny? ISME Journal 11, 1511–1520. https://doi.org/10.1038/ismej.2017.16

Hu, Y., Pang, S., Yang, J., Zhao, X., Cao, J., 2019. Changes in soil microbial community structure following amendment of biosolids for seven years. Environmental Pollutants and Bioavailability 31, 24–31. https://doi.org/10.1080/26395940.2019.1569478

Hurst, C.J., Gerba C, P., Cech, I., 1980. Effects of environmental variables and soil characteristics on virus survival in soil. Applied and Environmental Microbiology 40, 1067–1079. https://doi.org/10.1128/aem.40.6.1067-1079.1980

Hurwitz, B.L., Sullivan, M.B., 2013. The Pacific Ocean Virome (POV): A marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. PLoS ONE 8, e57355. https://doi.org/10.1371/journal.pone.0057355

Huson, D.H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh,

H.-J., Tappu, R., 2016. MEGAN Community Edition - interactive exploration and analysis of large-scale microbiome sequencing data. PLOS Computational Biology 12, e1004957. https://doi.org/10.1371/journal.pcbi.1004957

Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11, 119. https://doi.org/10.1186/1471-2105-11-119

Iacobucci, G., 2020. Covid-19: UK lockdown is "crucial" to saving lives, say doctors and scientists. BMJ m1204. https://doi.org/10.1136/bmj.m1204

ICTV, 2021. The international code of virus classification and nomenclature of ICTV. https://doi.org/10.1016/b978-0-12-249951-7.50020-1

Ippolito, J.A., Ducey, T.F., Diaz, K., Barbarick, K.A., 2021. Long-term biosolids land application influences soil health. Science of The Total Environment 791, 148344. https://doi.org/https://doi.org/10.1016/j.scitotenv.2021.148344

Islam, M.R., Hoque, M.N., Rahman, M.S., Alam, A.S.M.R.U., Akther, M., Puspo, J.A., Akter, S., Sultana, M., Crandall, K.A., Hossain, M.A., 2020. Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity. Scientific Reports 10, 14004. https://doi.org/10.1038/s41598-020-70812-6

Izquierdo-Lara, R., Elsinga, G., Heijnen, L., Munnink, B.B.O., Schapendonk, C.M.E., Nieuwenhuijse, D., Kon, M., Lu, L., Aarestrup, F.M., Lycett, S., Medema, G., Koopmans, M.P.G., Graaf, M. de, 2021. Monitoring SARS-CoV-2 circulation and diversity through community wastewater sequencing, the Netherlands and Belgium. Emerging Infectious Diseases 27, 1405–1415. https://doi.org/10.3201/eid2705.204410

Jarvis, C.I., Van Zandvoort, K., Gimma, A., Prem, K., Auzenbergs, M., O'Reilly, K., Medley, G., Emery, J.C., Houben, R.M.G.J., Davies, N., Nightingale, E.S., Flasche, S., Jombart, T., Hellewell, J., Abbott, S., Munday, J.D., Bosse, N.I., Funk, S., Sun, F., Endo, A., Rosello, A., Procter, S.R., Kucharski, A.J., Russell, T.W., Knight, G., Gibbs, H., Leclerc, Q., Quilty, B.J., Diamond, C., Liu, Y., Jit, M., Clifford, S., Pearson, C.A.B., Eggo, R.M., Deol, A.K., Klepac, P., Rubin, G.J., Edmunds, W.J., CMMID COVID-19 working group, 2020. Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. BMC Medicine 18, 124. https://doi.org/10.1186/s12916-020-01597-8

Jia, D., Chen, Q., Mao, Q., Zhang, X., Wu, W., Chen, H., Yu, X., Wang, Z., Wei, T.,

2018. Vector mediated transmission of persistently transmitted plant viruses. Current Opinion in Virology 28, 127–132. https://doi.org/10.1016/j.coviro.2017.12.004

Jin, M., Guo, X., Zhang, R., Qu, W., Gao, B., Zeng, R., 2019. Diversities and potential biogeochemical impacts of mangrove soil viruses. Microbiome 7, 58. https://doi.org/10.1186/s40168-019-0675-9

Jo, J., Oh, J., Park, C., 2020. Microbial community analysis using high-throughput sequencing technology: A beginner's guide for microbiologists. Journal of Microbiology 58, 176–192. https://doi.org/10.1007/s12275-020-9525-5

John, S.G., Mendez, C.B., Deng, L., Poulos, B., Kauffman, A.K.M., Kern, S., Brum, J., Polz, M.F., Boyle, E.A., Sullivan, M.B., 2011. A simple and efficient method for concentration of ocean viruses by chemical flocculation. Environmental Microbiology Reports 3, 195–202. https://doi.org/10.1111/j.1758-2229.2010.00208.x

Jones, D.L., Baluja, M.Q., Graham, D.W., Corbishley, A., McDonald, J.E., Malham, S.K., Hillary, L.S., Connor, T.R., Gaze, W.H., Moura, I.B., Wilcox, M.H., Farkas, K., 2020. Shedding of SARS-CoV-2 in feces and urine and its potential role in person-to-person transmission and the environment-based spread of COVID-19. Science of The Total Environment 749, 141364. https://doi.org/10.1016/j.scitotenv.2020.141364

Joseph, R., Edward, T., Schaffner D., W., 2015. Abundance of antibiotic resistance genes in bacteriophage following soil fertilization with dairy manure or municipal biosolids, and evidence for potential transduction. Applied and Environmental Microbiology 81, 7905–7913. https://doi.org/10.1128/AEM.02363-15

Joshi, N., Fass, J., 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files.

Ju, F., Li, B., Ma, L., Wang, Y., Huang, D., Zhang, T., 2016. Antibiotic resistance genes and human bacterial pathogens: Co-occurrence, removal, and enrichment in municipal sewage sludge digesters. Water Research 91, 1–10. https://doi.org/10.1016/j.watres.2015.11.071

Käfer, S., Paraskevopoulou, S., Zirkel, F., Wieseke, N., Donath, A., Petersen, M., Jones, T.C., Liu, S., Zhou, X., Middendorf, M., Junglen, S., Misof, B., Drosten, C., 2019. Re-assessing the diversity of negative strand RNA viruses in insects. PLoS Pathogens 15, e1008224. https://doi.org/10.1371/journal.ppat.1008224

Kapp, J.D., Green, R.E., Shapiro, B., 2021. A fast and efficient single-stranded genomic

library preparation method optimized for ancient DNA. Journal of Heredity 112, 241–249. https://doi.org/10.1093/jhered/esab012

Karlsson, O.E., Belák, S., Granberg, F., 2013. The effect of preprocessing by sequence-independent, single-primer amplification (SISPA) on metagenomic detection of viruses. Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science 11, S227–S234. https://doi.org/10.1089/bsp.2013.0008

Katoh, K., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30, 3059–3066. https://doi.org/10.1093/nar/gkf436

Kelessidis, A., Stasinakis, A.S., 2012. Comparative study of the methods used for treatment and final disposal of sewage sludge in European countries. Waste Management 32, 1186–1195. https://doi.org/10.1016/J.WASMAN.2012.01.012

Kieft, K., Zhou, Z., Anantharaman, K., 2020. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. Microbiome 8, 1–23. https://doi.org/10.1186/S40168-020-00867-0

Kim, K.H., Bae, J.W., 2011. Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. Applied and Environmental Microbiology 77, 7663–7668. https://doi.org/10.1128/AEM.00289-11

Kitajima, M., Oka, T., Takagi, H., Tohya, Y., Katayama, H., Takeda, N., Katayama, K., 2010. Development and application of a broadly reactive real-time reverse transcription-PCR assay for detection of murine noroviruses. Journal of Virological Methods 169, 269–273. https://doi.org/10.1016/j.jviromet.2010.07.018

Kleiner, M., Hooper, L.V., Duerkop, B.A., 2015. Evaluation of methods to purify virus-like particles for metagenomic sequencing of intestinal viromes. BMC Genomics 16, 1–15. https://doi.org/10.1186/s12864-014-1207-4

Kocamemi, B.A., Kurt, H., Hacioglu, S., Yarali, C., Saatci, A.M., Pakdemirli, B., 2020. First data-set on SARS-CoV-2 detection for Istanbul wastewaters in Turkey. medRxiv 2020.05.03.20089417. https://doi.org/10.1101/2020.05.03.20089417

Koonin, E.V., 2010. The two empires and three domains of life in the postgenomic age. Nature Eductation 3, 27.

Koonin, E.V., Dolja, V.V., Krupovic, M., Kuhn, J.H., 2021. Viruses defined by the

position of the virosphere within the replicator space. Molecular Biology Reviews 85, e00193–20. https://doi.org/10.1128/MMBR.00193-20

Koonin, E.V., Dolja, V.V., Krupovic, M., Varsani, A., Wolf, Y.I., Yutin, N., Zerbini, F.M., Kuhn, J.H., 2020. Global organization and proposed megataxonomy of the virus world. Microbiology and Molecular Biology Reviews 84, e00061–19.

Kopylova, E., Noé, L., Touzet, H., 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28, 3211–3217. https://doi.org/10.1093/bioinformatics/bts611

Koyama, A., Steinweg, J.M., Haddix, M.L., Dukes, J.S., Wallenstein, M.D., 2018. Soil bacterial community responses to altered precipitation and temperature regimes in an old field grassland are mediated by plants. FEMS Microbiology Ecology 94. https://doi.org/10.1093/femsec/fix156

Krishnamurthy, S.R., Janowski, A.B., Zhao, G., Barouch, D., Wang, D., 2016. Hyperexpansion of RNA Bacteriophage Diversity. PLOS Biology 14, e1002409. https://doi.org/10.1371/journal.pbio.1002409

Krupovic, M., 2013. Networks of evolutionary interactions underlying the polyphyletic origin of ssDNA viruses. Current Opinion in Virology 3, 578–586. https://doi.org/10.1016/j.coviro.2013.06.010

Krupovic, Ma., Blomberg, Jonas, Coffin John, M., Dasgupta, and F., Indranil, Geering Andrew, D., Gifford, R., Harrach, B., Hull, R., Johnson, W., Kreuze Jan, F., Lindemann, D., Llorens, C., Lockhart, B., Mayer, J., Muller, E., Olszewski Neil, E., Pappu Hanu, R., Pooggin Mikhail, M., Richert-Pöggeler Katja, R., Sabanadzovic, S., Sanfaçon, H., Schoelz James, E., Seal, S., Stavolone, L., Stoye Jonathan, P., Teycheney, P.-Y., Tristem, M., Koonin Eugene, V., Kuhn Jens, H., Sandri-Goldin Rozanne, M., 2022. Ortervirales: New virus order unifying five families of reverse-transcribing viruses. Journal of Virology 92, e00515–18. https://doi.org/10.1128/JVI.00515-18

Krupovic, M., Kuhn, J.H., Wang, F., Baquero, D.P., Dolja, V.V., Egelman, E.H., Prangishvili, D., Koonin, E.V., 2021. Adnaviria : A new realm for archaeal filamentous viruses with linear A-form double-stranded DNA genomes. Journal of Virology 95. https://doi.org/10.1128/JVI.00673-21

Kuhn, J., Dolja, V., Krupovic, M., Adriaenssens, E., Di Serio, F., Dutilh, B., Flores, R., Harrach, B., Mushegian, A., Owens, B., Randles, J., Rubino, L., Sabanadzovic, S.,

Simmonds, P., Varsani, A., Zerbini, M., Koonin, E., 2020. Expand , amend , and emend the International Code of Virus Classification and Nomenclature (ICVCN ; " the Code ") and the Statutes to clearly define the remit of the ICTV. ICTV.

Kuloglu Genc, M., Mercan, S., Yayla, M., Tekin Bulbul, T., Adioren, C., Simsek, S.Z., Asicioglu, F., 2021. Monitoring geographical differences in illicit drugs, alcohol, and tobacco consumption via wastewater-based epidemiology: Six major cities in Turkey. Science of The Total Environment 797, 149156. https://doi.org/10.1016/j.scitotenv.2021.149156

Kumar, M., Patel, A.K., Shah, A.V., Raval, J., Rajpara, N., Joshi, M., Joshi, C.G., 2020. First proof of the capability of wastewater surveillance for COVID-19 in India through detection of genetic material of SARS-CoV-2. Science of The Total Environment 746, 141326. https://doi.org/10.1016/j.scitotenv.2020.141326

Kuznetsov, Y.G., Xiao, C., Sun, S., Raoult, D., Rossmann, M., McPherson, A., 2010. Atomic force microscopy investigation of the giant mimivirus. Virology 404, 127–137. https://doi.org/10.1016/j.virol.2010.05.007

La Rosa, G., Iaconelli, M., Mancini, P., Bonanno Ferraro, G., Veneri, C., Bonadonna, L., Lucentini, L., Suffredini, E., 2020. First detection of SARS-CoV-2 in untreated wastewaters in Italy. Science of The Total Environment 736, 139652. https://doi.org/10.1016/j.scitotenv.2020.139652

La Rosa, G., Mancini, P., Bonanno Ferraro, G., Veneri, C., Iaconelli, M., Bonadonna, L., Lucentini, L., Suffredini, E., 2021. SARS-CoV-2 has been circulating in northern Italy since December 2019: Evidence from environmental monitoring. Science of The Total Environment 750, 141711. https://doi.org/10.1016/j.scitotenv.2020.141711

Lam, P., Steinmetz, N.F., 2018. Plant viral and bacteriophage delivery of nucleic acid therapeutics. Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology 10, e1487. https://doi.org/10.1002/wnan.1487

Lang, A.S., Zhaxybayeva, O., Beatty, J.T., 2012. Gene transfer agents: phage-like elements of genetic exchange. Nature Reviews Microbiology 10, 472–482. https://doi.org/10.1038/nrmicro2802

Larsbrink, J., McKee, L.S., Gadd, G.M., Sariaslani, S., 2020. Chapter two - bacteroidetes bacteria in the soil: Glycan acquisition, enzyme secretion, and gliding motility, in: Advances in Applied Microbiology. Academic Press, pp. 63–98. https://doi.org/10.1

016/bs.aambs.2019.11.001

LaTurner, Z.W., Zong, D.M., Kalvapalle, P., Gamas, K.R., Terwilliger, A., Crosby, T., Ali, P., Avadhanula, V., Santos, H.H., Weesner, K., Hopkins, L., Piedra, P.A., Maresso, A.W., Stadler, L.B., 2021. Evaluating recovery, cost, and throughput of different concentration methods for SARS-CoV-2 wastewater-based epidemiology. Water Research 197, 117043. https://doi.org/10.1016/j.watres.2021.117043

LeBlanc, R.J., Richard, R.P., Beecher, N., 2008. Global atlas of excreta, wastewater sludge, and biosolids management: Moving forward the sustainable and welcome uses of a global resource. United Nations Human Settlement Programme, Nairobi. https://doi.org/10.2175/193864709793846402

Lee, S., Sorensen, J.W., Walker, R.L., Emerson, J.B., Nicol, G.W., Hazard, C., 2022. Soil pH influences the structure of virus communities at local and global scales. Soil Biology and Biochemistry 166, 108569. https://doi.org/10.1016/j.soilbio.2022.108569

Lee, S., Sorensen, J.W., Walker, R.L., Emerson, J.B., Nicol, G.W., Hazard, C., 2021. Soil viral communities are structured by pH at local and global scales. bioRxiv. https://doi.org/10.1101/2021.10.20.465127

Lefkowitz, E.J., Dempsey, D.M., Hendrickson, R.C., Orton, R.J., Siddell, S.G., Smith, D.B., 2018. Virus taxonomy: The database of the international committee on taxonomy of viruses (ICTV). Nucleic Acids Research 46, D708–D717. https://doi.org/10.1093/nar/gkx932

Leifels, M., Cheng, D., Sozzi, E., Shoults, D.C., Wuertz, S., Mongkolsuk, S., Sirikanchana, K., 2021. Capsid integrity quantitative PCR to determine virus infectivity in environmental and food applications – A systematic review. Water Research X 11, 100080. https://doi.org/10.1016/J.WROA.2020.100080

Lekunberri, I., Subirats, J., Borrego, C.M., Balcázar, J.L., 2017. Exploring the contribution of bacteriophages to antibiotic resistance. Environmental Pollution 220, 981–984. https://doi.org/10.1016/j.envpol.2016.11.059

Letunic, I., Bork, P., 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Research 47, W256–W259. https://doi.org/10.1093/nar/gkz239

Li, D., Liu, C.-M., Luo, R., Sadakane, K., Lam, T.-W., 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de

Bruijn graph. Bioinformatics 31, 1674–1676. https://doi.org/10.1093/bioinformatics/btv033

Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Li, X., Zhang, S., Shi, J., Luby, S.P., Jiang, G., 2021. Uncertainties in estimating SARS-CoV-2 prevalence by wastewater-based epidemiology. Chemical Engineering Journal 415, 129039. https://doi.org/10.1016/j.cej.2021.129039

Li, Y., Gordon, E., Shean, R.C., Idle, A., Deng, X., Greninger, A.L., Delwart, E., 2021. CrAssphage and its bacterial host in cat feces. Scientific Reports 11, 815. https://doi.org/10.1038/s41598-020-80076-9

Liang, X., Wagner, R.E., Zhuang, J., DeBruyn, J.M., Wilhelm, S.W., Liu, F., Yang, L., Staton, M.E., Sherfy, A.C., Radosevich, M., 2019. Viral abundance and diversity vary with depth in a southeastern United States agricultural ultisol. Soil Biology and Biochemistry 137, 107546. https://doi.org/10.1016/j.soilbio.2019.107546

Lin, X., Glier, M., Kuchinski, K., Mierlo, T.R.-V., McVea, D., Tyson, J.R., Prystajecky, N., Ziels, R.M., 2022. Assessing multiplex tiling PCR sequencing approaches for detecting genomic variants of SARS-CoV-2 in municipal wastewater. mSystems 6, e01068–21. https://doi.org/10.1128/mSystems.01068-21

Lorenzo, M., Picó, Y., 2019. Wastewater-based epidemiology: Current status and future prospects. Current Opinion in Environmental Science & Health 9, 77–84. https://doi.org/10.1016/j.coesh.2019.05.007

Lucía-Sanz, A., Manrubia, S., 2017. Multipartite viruses: Adaptive trick or evolutionary treat? npj Systems Biology and Applications 3, 1–11. https://doi.org/10.1038/s41540-017-0035-y

Lyu, J., Imachi, H., Fukunaga, K., Yoshimoto, T., Zhang, H., Murao, K., 2015. Roles of lipoprotein receptors in the entry of hepatitis c virus. World journal of hepatology 7, 2535–42. https://doi.org/10.4254/wjh.v7.i24.2535

Mahmoud, H., Jose, L., 2017. Phage and nucleocytoplasmic large viral sequences dominate coral viromes from the arabian gulf. Frontiers in Microbiology 8, 2063.

Manuel, D.-B., Oliverio Angela, M., Brewer Tess, E., Alberto, B.-G., Eldridge David, J., Bardgett Richard, D., Maestre Fernando, T., Singh Brajesh, K., Noah, F., 2018.

A global atlas of the dominant bacteria found in soil. Science 359, 320–325. https://doi.org/10.1126/science.aap9516

Martin, A., Lemon, S.M., 2002. The molecular biology of hepatitis a virus, in: Ou, J.-H.J. (Ed.), Hepatitis Viruses. Springer US, Boston, MA, pp. 23–50. https://doi.org/10.1007/978-1-4615-0881-6_2

Martin, J., Klapsa, D., Wilton, T., Zambon, M., Bentley, E., Bujaki, E., Fritzsche, M., Mate, R., Majumdar, M., 2020. Tracking SARS-CoV-2 in sewage: Evidence of changes in virus variant predominance during COVID-19 pandemic. Viruses 12, 1144. https://doi.org/10.3390/v12101144

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10–12. https://doi.org/10.14806/ej.17.1.200

Medema, G., Heijnen, L., Elsinga, G., Italiaander, R., Brouwer, A., 2020. Presence of SARS-Coronavirus-2 RNA in sewage and correlation with reported COVID-19 prevalence in the early stage of the epidemic in The Netherlands. Environmental Science & Technology Letters 7, 511–516. https://doi.org/10.1021/acs.estlett.0c00357

Melnick, J.L., 1984. Enteroviruses, in: Evans, A.S. (Ed.), Viral Infections of Humans: Epidemiology and Control. Springer US, Boston, MA, pp. 187–251. https://doi.org/10.1007/978-1-4684-4727-9_9

Meredith, L.W., Hamilton, W.L., Warne, B., Houldcroft, C.J., Hosmillo, M., Jahun, A.S., Curran, M.D., Parmar, S., Caller, L.G., Caddy, S.L., Khokhar, F.A., Yakovleva, A., Hall, G., Feltwell, T., Forrest, S., Sridhar, S., Weekes, M.P., Baker, S., Brown, N., Moore, E., Popay, A., Roddick, I., Reacher, M., Gouliouris, T., Peacock, S.J., Dougan, G., Török, M.E., Goodfellow, I., 2020. Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: A prospective genomic surveillance study. The Lancet Infectious Diseases 20, 1263–1272. https://doi.org/10.1016/S1473-3099(20)30562-4

Meschke, J.S., Sobsey, M.D., 1998. Comparative adsorption of Norwalk virus, poliovirus 1 and F+ RNA coliphage MS2 to soils suspended in treated wastewater. Water Science and Technology 38, 187–189. https://doi.org/10.1016/S0273-1223(98)00823-3

Milgroom, M.G., Cortesi, P., 2004. Biological control of chestnut blight with hypovirulence: A critical analysis. Annual Review of Phytopathology 42, 311–338. https://doi.org/10.1146/annurev.phyto.42.040803.140325

Miranda, K.M., Espey, M.G., Wink, D.A., 2001. A Rapid, Simple Spectrophotometric Method for Simultaneous Detection of Nitrate and Nitrite. Nitric Oxide 5, 62–71. https://doi.org/10.1006/niox.2000.0319

Mistry, J., Finn, R.D., Eddy, S.R., Bateman, A., Punta, M., 2013. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Research 41, e121–e121. https://doi.org/10.1093/nar/gkt263

Mocé-Llivina, L., Muniesa, M., Pimenta-Vale, H., Lucena, F., Jofre, J., 2003. Survival of bacterial indicator species and bacteriophages after thermal treatment of sludge and sewage. Applied and Environmental Microbiology 69, 1452–1456. https://doi.org/10.1128/AEM.69.3.1452–1456.2003

Monier, A., Chambouvet, A., Milner, D.S., Attah, V., Terrado, R., Lovejoy, C., Moreau, H., Santoro, A.E., Derelle, É., Richards, T.A., 2017. Host-derived viral transporter protein for nitrogen uptake in infected marine phytoplankton. Proceedings of the National Academy of Sciences 114, E7489–E7498. https://doi.org/10.1073/PNAS.1708097114

Mulvaney, R.L., 1996. Nitrogen - Inorganic Forms, in: Methods of Soil Analysis, Part 3. Chemical Methods (D.L. Sparks Ed.). SSSA, Madison, WI, USA, pp. 1123–1184.

Murphy, J., Riley, J.P., 1962. A modified single solution method for the determination of phosphate in natural waters. Analytica Chimica Acta 27, 31–36. https://doi.org/10.1016/S0003-2670(00)88444-5

Narr, A., Nawaz, A., Wick, L.Y., Harms, H., Chatzinotas, A., 2017. Soil viral communities vary temporally and along a land use transect as revealed by virus-like particle counting and a modified community fingerprinting approach (fRAPD). Frontiers in Microbiology 8, 1975. https://doi.org/10.3389/fmicb.2017.01975

Nassal, M., Schaller, H., 1993. Hepatitis b virus replication. Trends in Microbiology 1, 221–228. https://doi.org/10.1016/0966-842X(93)90136-F

Nayfach, S., Camargo, A.P., Schulz, F., Eloe-Fadrosh, E., Roux, S., Kyrpides, N.C., 2020. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nature Biotechnology 39, 578–585. https://doi.org/10.1038/s41587-020-00774-7

Nemudryi, A., Nemudraia, A., Wiegand, T., Surya, K., Buyukyoruk, M., Cicha, C., Vanderwood, K.K., Wilkinson, R., Wiedenheft, B., 2020. Temporal detection and phylogenetic assessment of SARS-CoV-2 in municipal wastewater. Cell Reports Medicine

1, 100098. https://doi.org/10.1016/j.xcrm.2020.100098

Neri, U., Wolf, Y.I., Roux, S., Camargo, A.P., Lee, B., Kazlauskas, D., Chen, I.M., Ivanova, N., Allen, L.Z., Paez-Espino, D., Bryant, D.A., Bhaya, D., Consortium, R.N.A.V.D., Krupovic, M., Dolja, V.V., Kyrpides, N.C., Koonin, E.V., Gophna, U., 2022. A five-fold expansion of the global RNA virome reveals multiple new clades of RNA bacteriophages. bioRxiv 2022.02.15.480533. https://doi.org/10.1101/2022.02.15.480533

Nishiura, H., Kobayashi, T., Miyama, T., Suzuki, A., Jung, S., Hayashi, K., Kinoshita, R., Yang, Y., Yuan, B., Akhmetzhanov, A.R., Linton, N.M., 2020. Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). International Journal of Infectious Diseases 94, 154–155. https://doi.org/10.1016/j.ijid.2020.03.020

Obbard, D.J., Shi, M., Roberts, K.E., Longdon, B., Dennis, A.B., 2020. A new lineage of segmented RNA viruses infecting animals. Virus Evolution 6, 61. https://doi.org/10.1093/ve/vez061

Office for National Statistics, 2020. Deaths registered weekly in England and Wales, provisional.

Ofir, G., Sorek, R., 2018. Contemporary phage biology: From classic models to new insights. Cell 172, 1260–1270. https://doi.org/10.1016/j.cell.2017.10.045

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., 2019. vegan: Community Ecology Package.

Oliveira, F.C., Faria, M.F. de, Bertoncini, E.I., Sato, M.I.Z., Hachich, E.M., Guerrini, I.A., Passos, J.R.S., James, J.N., Harrison, R.B., Feitoza, T.G., Chiaradia, J.J., Abreu-Junior, C.H., Firme de Moraes, L.P., 2019. Persistence of fecal contamination indicators and pathogens in class b biosolids applied to sugarcane fields. Journal of Environmental Quality 48, 526–530. https://doi.org/10.2134/jeq2018.07.0270

Oliveira, M. de, Frihling, B.E.F., Velasques, J., Filho, F.J.C.M., Cavalheri, P.S., Migliolo, L., 2020. Pharmaceuticals residues and xenobiotics contaminants: Occurrence, analytical techniques and sustainable alternatives for wastewater treatment. Science of The Total Environment 705, 135568. https://doi.org/10.1016/j.scitotenv.2019.135568

Ozawa, H., Yoshida, H., Usuku, S., 2019. Environmental surveillance can dynamically track ecological changes in enteroviruses. Applied and Environmental Microbiology 85,

e01604–19, /aem/85/24/AEM.01604–19.atom. https://doi.org/10.1128/AEM.01604-19

Paez-Espino, D., Eloe-Fadrosh, E.A., Pavlopoulos, G.A., Thomas, A.D., Huntemann, M., Mikhailova, N., Rubin, E., Ivanova, N.N., Kyrpides, N.C., 2016. Uncovering Earth's virome. Nature 536, 425–430. https://doi.org/10.1038/nature19094

Palukaitis, P., 2016. Satellite RNAs and satellite viruses. Molecular Plant-Microbe Interactions 29, 181–186. https://doi.org/10.1094/MPMI-10-15-0232-FI

Parras-Moltó, M., Rodríguez-Galet, A., Suárez-Rodríguez, P., López-Bueno, A., 2018. Evaluation of bias induced by viral enrichment and random amplification protocols in metagenomic surveys of saliva DNA viruses. Microbiome 6, 119. https://doi.org/10.1186/s40168-018-0507-3

Peccia, J., Zulli, A., Brackney, D.E., Grubaugh, N.D., Kaplan, E.H., Casanovas-Massana, A., Ko, A.I., Malik, A.A., Wang, D., Wang, M., Warren, J.L., Weinberger, D.M., Omer, S.B., 2020. SARS-CoV-2 RNA concentrations in primary municipal sewage sludge as a leading indicator of COVID-19 outbreak dynamics. medRxiv 2020.05.19.20105999. https://doi.org/10.1101/2020.05.19.20105999

Pecson, B.M., Darby, E., Haas, C.N., Amha, Y.M., Bartolo, M., Danielson, R., Dearborn, Y., Giovanni, G.D., Ferguson, C., Fevig, S., Gaddis, E., Gray, D., Lukasik, G., Mull, B., Olivas, L., Olivieri, A., Qu, Y., Consortium, S.-C.I., 2021. Reproducibility and sensitivity of 36 methods to quantify the SARS-CoV-2 genetic signal in raw wastewater: Findings from an interlaboratory methods evaluation in the U.S. Environmental Science: Water Research & Technology. https://doi.org/10.1039/D0EW00946F

Phan, T., 2020. Novel coronavirus: From discovery to clinical diagnostics. Infection, Genetics and Evolution 79, 104211. https://doi.org/10.1016/j.meegid.2020.104211

Pietilä, M.K., Roine, E., Sencilo, A., Bamford, D.H., Oksanen, H.M., 2016. Pleolipoviridae, a newly proposed family comprising archaeal pleomorphic viruses with single-stranded or double-stranded DNA genomes. Archives of Virology 161, 249–256. https://doi.org/10.1007/s00705-015-2613-x

Plessis, L. du, McCrone, J.T., Zarebski, A.E., Hill, V., Ruis, C., Gutierrez, B., Raghwani, J., Ashworth, J., Colquhoun, R., Connor, T.R., Faria, N.R., Jackson, B., Loman, N.J., O'Toole, Á., Nicholls, S.M., Parag, K.V., Scher, E., Vasylyeva, T.I., Volz, E.M., Watts, A., Bogoch, I.I., Khan, K., Consortium†, COVID-19 Genomics UK (COG-UK),

Aanensen, D.M., Kraemer, M.U.G., Rambaut, A., Pybus, O.G., 2021. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. Science 371, 708–712. https://doi.org/10.1126/science.abf2946

Polo, D., Quintela-Baluja, M., Corbishley, A., Jones, D.L., Singer, A.C., Graham, D.W., Romalde, J.L., 2020. Making waves: Wastewater-based epidemiology for COVID-19 – approaches and challenges for surveillance and prediction. Water Research 186, 116404. https://doi.org/10.1016/j.watres.2020.116404

Pons, J.C., Paez-Espino, D., Riera, G., Ivanova, N., Kyrpides, N.C., Llabrés, M., 2021. VPF-Class: taxonomic assignment and host prediction of uncultivated viruses based on viral protein families. Bioinformatics 37, 1805–1813. https://doi.org/10.1093/bioinformatics/btab026

Prado, T., Fumian, T.M., Mannarino, C.F., Maranhão, A.G., Siqueira, M.M., Miagostovich, M.P., Prado, T., Fumian, T.M., Mannarino, C.F., Maranhão, A.G., Siqueira, M.M., Miagostovich, M.P., 2020. Preliminary results of SARS-CoV-2 detection in sewerage system in Niterói municipality, Rio de Janeiro, Brazil. Memórias do Instituto Oswaldo Cruz 115. https://doi.org/10.1590/0074-02760200196

Prangishvili, D., Krupovic, M., 2012. A new proposed taxon for double-stranded DNA viruses, the order "ligamenvirales." Archives of Virology 157, 791–795. https://doi.org/10.1007/s00705-012-1229-7

Pratama, A.A., Bolduc, B., Zayed, A.A., Zhong, Z.-P., Guo, J., Vik, D.R., Gazitúa, M.C., Wainaina, J.M., Roux, S., Sullivan, M.B., 2021. Expanding standards in viromics: in silico evaluation of dsDNA viral genome identification, classification, and auxiliary metabolic gene curation. PeerJ 9, e11447. https://doi.org/10.7717/peerj.11447

Pratama, A.A., Elsas, J.D. van, 2018. The "neglected" soil virome - potential role and impact. Trends in Microbiology 26, 649–662. https://doi.org/10.1016/j.tim.2017.12.004

Price, G.W., Langille, M.G.I., Yurgel, S.N., 2021. Microbial co-occurrence network analysis of soils receiving short- and long-term applications of alkaline treated biosolids. Science of The Total Environment 751, 141687. https://doi.org/10.1016/j.scitotenv.2020.141687

Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2 - Approximately maximum-likelihood trees for large alignments. PLoS ONE 5, e9490. https://doi.org/10.1371/

journal.pone.0009490

Public Health Wales, 2020. Public Health Wales Rapid COVID-19 surveillance.

Quaiser, A., Dufresne, A., Ballaud, F., Roux, S., Zivanovic, Y., Colombet, J., Sime-Ngando, T., Francez, A.-J., 2015. Diversity and comparative genomics of microviridae in sphagnum-dominated peatlands. Frontiers in Microbiology 6, 375. https://doi.org/10.3389/fmicb.2015.00375

R Core Team, 2013. R: A language and environment for statistical computing.

Rambaut, A., Robertson, D.L., Pybus, O.G., Peeters, M., Holmes, E.C., 2001. Phylogeny and the origin of HIV-1. Nature 410, 1047–1048. https://doi.org/10.1038/35074179

Randazzo, W., Truchado, P., Cuevas-Ferrando, E., Simón, P., Allende, A., Sánchez, G., 2020. SARS-CoV-2 RNA in wastewater anticipated COVID-19 occurrence in a low prevalence area. Water Research 181, 115942. https://doi.org/10.1016/j.watres.2020.115942

Randles, J.W., Fiallo-Olivé, E., Xu, Y., Candresse, T., Bejerman, N., Debat, H., Dietzgen, R.G., 2020. The plant negative-sense RNA virosphere: Virus discovery through new eyes 11, 2303. https://doi.org/10.3389/fmicb.2020.588427

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., Claverie, J.-M., 2004. The 1.2-megabase genome sequence of mimivirus. Science 306, 1344–50. https://doi.org/10.1126/science.1101485

Raya, R.R., H'bert, E.M., 2009. Isolation of Phage via Induction of Lysogens, in: Methods in Molecular Biology (Clifton, n.j.). Humana Press, pp. 23–32. https://doi.org/10.1007/978-1-60327-164-6_3

Reardon, C.L., Wuest, S.B., 2016. Soil amendments yield persisting effects on the microbial communities–a 7-year study. Applied Soil Ecology 101, 107–116. https://doi.org/10.1016/j.apsoil.2015.12.013

Reeves, K., Liebig, J., Feula, A., Saldi, T., Lasda, E., Johnson, W., Lilienfeld, J., Maggi, J., Pulley, K., Wilkerson, P.J., Real, B., Zak, G., Davis, J., Fink, M., Gonzales, P., Hager, C., Ozeroff, C., Tat, K., Alkire, M., Butler, C., Coe, E., Darby, J., Freeman, N., Heuer, H., Jones, J.R., Karr, M., Key, S., Maxwell, K., Nelson, L., Saldana, E., Shea, R., Salveson, L., Tomlinson, K., Vargas-Barriga, J., Vigil, B., Brisson, G., Parker, R., Leinwand, L.A., Bjorkman, K., Mansfeldt, C., 2021. High-resolution within-sewer SARS-CoV-2 surveillance facilitates informed intervention. Water Re-

search 204, 117613. https://doi.org/10.1016/j.watres.2021.117613

Ren, J., Song, K., Deng, C., Ahlgren, N.A., Fuhrman, J.A., Li, Y., Xie, X., Poplin, R., Sun, F., 2020. Identifying viruses from metagenomic data using deep learning. Quantitative Biology 8, 64–77. https://doi.org/10.1007/s40484-019-0187-4

Roberts, B.N., Bailey, R.H., McLaughlin, M.R., Brooks, J.P., 2016. Decay rates of zoonotic pathogens and viral surrogates in soils amended with biosolids and manures and comparison of qPCR and culture derived rates. Science of The Total Environment 573, 671–679. https://doi.org/10.1016/j.scitotenv.2016.08.088

Rodríguez-Lázaro, D., Cook, N., Ruggeri, F.M., Sellwood, J., Nasser, A., Nascimento, M.S.J., D'Agostino, M., Santos, R., Saiz, J.C., Rzeżutka, A., Bosch, A., Gironés, R., Carducci, A., Muscillo, M., Kovač, K., Diez-Valcarce, M., Vantarakis, A., Bonsdorff, C.-H. von, Roda Husman, A.M. de, Hernández, M., Poel, W.H.M. van der, 2012. Virus hazards from food, water and other contaminated environments. FEMS Microbiology Reviews 36, 786–814. https://doi.org/10.1111/j.1574-6976.2011.00306.x

Roossinck, M.J., 2019. Evolutionary and ecological links between plant and fungal viruses. New Phytologist 221, 86–92. https://doi.org/10.1111/nph.15364

Roossinck, M.J., 2011. The good viruses: Viral mutualistic symbioses. Nature Reviews Microbiology 9, 99–108. https://doi.org/10.1038/nrmicro2491

Roux, S., Adriaenssens, E.M., Dutilh, B.E., Koonin, E.V., Kropinski, A.M., Krupovic, M., Kuhn, J.H., Lavigne, R., Brister, J.R., Varsani, A., Amid, C., Aziz, R.K., Bordenstein, S.R., Bork, P., Breitbart, M., Cochrane, G.R., Daly, R.A., Desnues, C., Duhaime, M.B., Emerson, J.B., Enault, F., Fuhrman, J.A., Hingamp, P., Hugenholtz, P., Hurwitz, B.L., Ivanova, N.N., Labonté, J.M., Lee, K.-B., Malmstrom, R.R., Martinez-Garcia, M., Mizrachi, I.K., Ogata, H., Páez-Espino, D., Petit, M.-A., Putonti, C., Rattei, T., Reyes, A., Rodriguez-Valera, F., Rosario, K., Schriml, L., Schulz, F., Steward, G.F., Sullivan, M.B., Sunagawa, S., Suttle, C.A., Temperton, B., Tringe, S.G., Thurber, R.V., Webster, N.S., Whiteson, K.L., Wilhelm, S.W., Wommack, K.E., Woyke, T., Wrighton, K.C., Yilmaz, P., Yoshida, T., Young, M.J., Yutin, N., Allen, L.Z., Kyrpides, N.C., Eloe-Fadrosh, E.A., 2018. Minimum Information about an Uncultivated Virus Genome (MIUViG). Nature Biotechnology 37, 29–37. https://doi.org/10.1038/nbt.4306

Roux, S., Brum, J.R., Dutilh, B.E., Sunagawa, S., Duhaime, M.B., Loy, A., Poulos,

B.T., Solonenko, N., Lara, E., Poulain, J., Pesant, S., Kandels-Lewis, S., Dimier, C., Picheral, M., Searson, S., Cruaud, C., Alberti, A., Duarte, C.M., Gasol, J.M., Vaqué, D., Bork, P., Acinas, S.G., Wincker, P., Sullivan, M.B., 2016a. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. Nature 537, 689–693. https://doi.org/10.1038/nature19366

Roux, S., Emerson, J.B., Eloe-Fadrosh, E.A., Sullivan, M.B., 2017. Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. PeerJ 5, e3817. https://doi.org/10.7717/peerj.3817

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., 2016b. Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. PeerJ 4, e2777. https://doi.org/10.7717/peerj.2777

Santos-Medellin, C., Zinke, L.A., Horst, A.M. ter, Gelardi, D.L., Parikh, S.J., Emerson, J.B., 2021. Viromes outperform total metagenomes in revealing the spatiotemporal patterns of agricultural soil viral communities. The ISME Journal 15, 1956–1970. https://doi.org/10.1038/s41396-021-00897-y

Savary, S., Willocquet, L., Pethybridge, S.J., Esker, P., McRoberts, N., Nelson, A., 2019. The global burden of pathogens and pests on major food crops. Nature Ecology & Evolution 3, 430–439. https://doi.org/10.1038/s41559-018-0793-y

Schlatter, D.C., Paul, N.C., Shah, D.H., Schillinger, W.F., Bary, A.I., Sharratt, B., Paulitz, T.C., 2019. Biosolids and tillage practices influence soil bacterial communities in dryland wheat. Microbial Ecology 78, 737–752. https://doi.org/10.1007/s00248-019-01339-1

Schmidt, M.W.I., Torn, M.S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I.A., Kleber, M., Kögel-Knabner, I., Lehmann, J., Manning, D.A.C., Nannipieri, P., Rasse, D.P., Weiner, S., Trumbore, S.E., 2011. Persistence of soil organic matter as an ecosystem property. Nature 478. https://doi.org/10.1038/nature10386

Schmieder, R., Edwards, R., 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27, 863–864. https://doi.org/10.1093/bioinformatics/btr026

Schmitt, A.P., Lamb, R.A., 2004. Escaping from the Cell: Assembly and Budding of Negative-Strand RNA Viruses, in: Kawaoka, Y. (Ed.), Biology of Negative Strand

RNA Viruses: The Power of Reverse Genetics. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 145–196. https://doi.org/10.1007/978-3-662-06099-5_5

Schulz, F., Andreani, J., Francis, R., Boudjemaa, H., Bou Khalil, Jacques Yaacoub, Lee, J., La Scola, B., Woyke, T., 2022. Advantages and limits of metagenomic assembly and binning of a giant virus. mSystems 5, e00048–20. https://doi.org/10.1128/mSystems.00048-20

Schwarz, K.R., Sidhu, J.P.S., Pritchard, D.L., Li, Y., Toze, S., 2014. Decay of enteric microorganisms in biosolids-amended soil under wheat (triticum aestivum) cultivation. Water Research 59, 185–197. https://doi.org/10.1016/j.watres.2014.03.037

Shaffer, M., Borton, M.A., McGivern, B.B., Zayed, A.A., La Rosa, S.L., Solden, L.M., Liu, P., Narrowe, A.B., Rodríguez-Ramos, J., Bolduc, B., Gazitúa, M.C., Daly, R.A., Smith, G.J., Vik, D.R., Pope, P.B., Sullivan, M.B., Roux, S., Wrighton, K.C., 2020. DRAM for distilling microbial metabolism to automate the curation of microbiome function. Nucleic Acids Research 48, 8883–8900. https://doi.org/10.1093/nar/gkaa621

Sharma, B., Sarkar, A., Singh, P., Singh, R.P., 2017. Agricultural utilization of biosolids: A review on potential effects on soil and plant grown. Waste Management 64, 117–132. https://doi.org/https://doi.org/10.1016/j.wasman.2017.03.002

Shaw, R., Williams, A.P., Jones, D.L., 2014. Assessing soil nitrogen availability using microdialysis-derived diffusive flux measurements. Soil Science Society of America Journal 78, 1797–1803.

Sherchan, S.P., Shahin, S., Ward, L.M., Tandukar, S., Aw, T.G., Schmitz, B., Ahmed, W., Kitajima, M., 2020. First detection of SARS-CoV-2 RNA in wastewater in North America: A study in Louisiana, USA. Science of The Total Environment 743, 140621. https://doi.org/10.1016/j.scitotenv.2020.140621

Shi, M., Lin, X.D., Tian, J.H., Chen, L.J., Chen, X., Li, C.X., Qin, X.C., Li, J., Cao, J.P., Eden, J.S., Buchmann, J., Wang, W., Xu, J., Holmes, E.C., Zhang, Y.Z., 2016. Redefining the invertebrate RNA virosphere. Nature 540, 539–543. https://doi.org/10.1038/nature20167

Shkoporov, A.N., Khokhlova, E.V., Fitzgerald, C.B., Stockdale, S.R., Draper, L.A., Ross, R.P., Hill, C., 2018. ΦCrAss001 represents the most abundant bacteriophage family in the human gut and infects Bacteroides intestinalis. Nature Communications 2018

9:1 9, 1–8. https://doi.org/10.1038/s41467-018-07225-7

Sloan, J.J., Dowdy, R.H., Dolan, M.S., Linden, D.R., 1997. Long-term effects of biosolids applications on heavy metal bioavailability in agricultural soils. Journal of Environmental Quality 26, 966–974. https://doi.org/https://doi.org/10.2134/jeq1997.00472425002600040006x

Solonenko, S.A., Ignacio-Espinoza, J., Alberti, A., Cruaud, C., Hallam, S., Konstantinidis, K., Tyson, G., Wincker, P., Sullivan, M.B., 2013. Sequencing platform and library preparation choices impact viral metagenomes. BMC Genomics 14, 320. https://doi.org/10.1186/1471-2164-14-320

Sommers, P., Chatterjee, A., Varsani, A., Trubl, G., 2021. Integrating viral metagenomics into an ecological framework. Annual Review of Virology 8, 133–158. https://doi.org/10.1146/annurev-virology-010421-053015

Song, W., Sun, H.-X., Zhang, C., Cheng, L., Peng, Y., Deng, Z., Wang, D., Wang, Y., Hu, M., Liu, W., Yang, H., Shen, Y., Li, J., You, L., Xiao, M., 2019. Prophage Hunter: an integrative hunting tool for active prophages. Nucleic Acids Research 47, W74–W80. https://doi.org/10.1093/NAR/GKZ380

Sonune, A., Ghate, R., 2004. Developments in wastewater treatment methods. Desalination 167, 55–63. https://doi.org/10.1016/J.DESAL.2004.06.113

Sorensen, J.W., Zinke, L.A., Horst, A.M. ter, Santos-Medellín, C., Schroeder, A., Emerson, J.B., 2021. DNase treatment improves viral enrichment in agricultural soil viromes. mSystems 6, e00614–21. https://doi.org/10.1128/mSystems.00614-21

Stachler, E., Akyon, B., Carvalho, N.A. de, Ference, C., Bibby, K., 2018. Correlation of crAssphage qPCR markers with culturable and molecular indicators of human fecal pollution in an impacted urban watershed. Environmental Science & Technology 52, 7505–7512. https://doi.org/10.1021/acs.est.8b00638

Stachler, E., Kelty, C., Sivaganesan, M., Li, X., Bibby, K., Shanks, O.C., 2017. Quantitative CrAssphage PCR assays for human fecal pollution measurement. Environmental Science & Technology 51, 9146–9154. https://doi.org/10.1021/acs.est.7b02703

Starr, E.P., Nuccio, E.E., Pett-Ridge, J., Banfield, J.F., Firestone, M.K., 2019. Metatranscriptomic reconstruction reveals RNA viruses with the potential to shape carbon cycling in soil. Proceedings of the National Academy of Sciences 116, 25900–25908. https://doi.org/10.1073/pnas.1908291116

Steward, G.F., Culley, A.I., Mueller, J.A., Wood-Charlson, E.M., Belcaid, M., Poisson, G., 2013. Are we missing half of the viruses in the ocean? The ISME Journal 7, 672–679. https://doi.org/10.1038/ismej.2012.121

Straub, T.M., Pepper, I.L., Gerba, C.P., 1992. Persistence of viruses in desert soils amended with anaerobically digested sewage sludge. Applied and Environmental Microbiology 58, 636–641.

Sun, J., Dai, X., Wang, Q., Loosdrecht, M.C.M. van, Ni, B.-J., 2019. Microplastics in wastewater treatment plants: Detection, occurrence and removal. Water Research 152, 21–37. https://doi.org/10.1016/j.watres.2018.12.050

Sutton, T.D.S., Clooney, A.G., Ryan, F.J., Ross, R.P., Hill, C., 2019. Choice of assembly software has a critical impact on virome characterisation. Microbiome 7, 12. https://doi.org/10.1186/s40168-019-0626-5

Svircev, A., Roach, D., Castle, A., 2018. Framing the future with bacteriophages in agriculture. Viruses 10. https://doi.org/10.3390/v10050218

Tang, A., Tong, Z., Wang, H., Dai, Y., Li, K., Liu, J., Wu, W., Yuan, C., Yu, M., Li, P., Yan, J., 2020. Detection of novel coronavirus by RT-PCR in stool specimen from asymptomatic child, china. Emerging Infectious Diseases 26, 1337–1339. https://doi.org/10.3201/eid2606.200301

Taylor, L.H., Latham, S.M., Woolhouse, M.E.J., 2001. Risk factors for human disease emergence. Philosophical Transactions of the Royal Society B: Biological Sciences 356, 983–989. https://doi.org/10.1098/rstb.2001.0888

Thurber, R.V., Haynes, M., Breitbart, M., Wegley, L., Rohwer, F., 2009. Laboratory procedures to generate viral metagenomes. Nature Protocols 4, 470–483. https://doi.org/10.1038/nprot.2009.10

Trimmer, J.T., Miller, D.C., Guest, J.S., 2019. Resource recovery from sanitation to enhance ecosystem services. Nature Sustainability 2, 681–690. https://doi.org/10.1038/s41893-019-0313-3

Trottier, J., Darques, R., Ait Mouheb, N., Partiot, E., Bakhache, W., Deffieu, M.S., Gaudin, R., 2020. Post-lockdown detection of SARS-CoV-2 RNA in the wastewater of Montpellier, France. One Health 10, 100157. https://doi.org/10.1016/j.onehlt.2020.100157

Trubl, G., Hyman, P., Roux, S., Abedon, S.T., 2020. Coming-of-age characterization

of soil viruses: A user's guide to virus isolation, detection within metagenomes, and viromics. Soil Systems 4, 1–34. https://doi.org/10.3390/soilsystems4020023

Trubl, G., Jang, H.B., Roux, S., Emerson, J.B., Solonenko, N., Vik, D.R., Solden, L., Ellenbogen, J., Runyon, A.T., Bolduc, B., Woodcroft, B.J., Saleska, S.R., Tyson, G.W., Wrighton, K.C., Sullivan, M.B., Rich, V.I., 2018. Soil viruses are underexplored players in ecosystem carbon processing. mSystems 3, e00076–18. https://doi.org/10.1128/mSystems.00076-18

Trubl, G., Kimbrel, J.A., Liquet-Gonzalez, J., Nuccio, E.E., Weber, P.K., Pett-Ridge, J., Jansson, J.K., Waldrop, M.P., Blazewicz, S.J., 2021. Active virus-host interactions at sub-freezing temperatures in arctic peat soil. Microbiome 9, 208. https://doi.org/10.1186/s40168-021-01154-2

Trubl, G., Roux, S., Solonenko, N., Li, Y.-F., Bolduc, B., Rodríguez-Ramos, J., Eloe-Fadrosh, E.A., Rich, V.I., Sullivan, M.B., Anesio, A., 2019. Towards optimized viral metagenomes for double-stranded and single-stranded DNA viruses from challenging soils. PeerJ 7, e7265. https://doi.org/10.7717/peerj.7265

Trubl, G., Solonenko, N., Chittick, L., Solonenko, S.A., Rich, V.I., Sullivan, M.B., 2016. Optimization of viral resuspension methods for carbon-rich soils along a permafrost thaw gradient. PeerJ 4, e1999. https://doi.org/10.7717/peerj.1999

Tsai, A.-Y., Gong, G.-C., Chien, &., Chao, F., 2016. Contribution of viral lysis and nanoflagellate grazing to bacterial mortality at surface waters and deeper depths in the coastal ecosystem of subtropical western pacific. https://doi.org/10.1007/s12237-016-0098-9

Tyson, J.R., James, P., Stoddart, D., Sparks, N., Wickenhagen, A., Hall, G., Choi, J.H., Lapointe, H., Kamelian, K., Smith, A.D., Prystajecky, N., Goodfellow, I., Wilson, S.J., Harrigan, R., Snutch, T.P., Loman, N.J., Quick, J., 2020. Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. bioRxiv : the preprint server for biology 2020.09.04.283077. https://doi.org/10.1101/2020.09.04.283077

Vainio, E.J., Chiba, S., Ghabrial, S.A., Maiss, E., Roossinck, M., Sabanadzovic, S., Suzuki, N., Xie, J., Nibert, M., 2018. ICTV Virus Taxonomy Profile: Partitiviridae. Journal of General Virology 99, 17–18. https://doi.org/10.1099/jgv.0.000985

Valles, S.M., Chen, Y., Firth, A.E., Guérin, D.M.A., Hashimoto, Y., Herrero, S., Miranda,

J.R. de, Ryabov, E., 2017. ICTV virus taxonomy profile: Dicistroviridae. Journal of General Virology 98, 355–356. https://doi.org/10.1099/jgv.0.000756

Van Etten, J.L., Graves, M.V., Müller, D.G., Boland, W., Delaroque, N., 2002. Phycodnaviridae- large DNA algal viruses. Archives of Virology 147, 1479–1516. https://doi.org/10.1007/s00705-002-0822-6

VanGuilder, H., Vrana, K., Freeman, W., 2008. Twenty-five years of quantitative PCR for gene expression analysis. BioTechniques 44 Supplem, 619–626. https://doi.org/10.2144/000112776

Varsani, A., Lefeuvre, P., Roumagnac, P., Martin, D., 2018. Notes on recombination and reassortment in multipartite/segmented viruses. Current Opinion in Virology 33, 156–166. https://doi.org/10.1016/j.coviro.2018.08.013

Venegas, M., Leiva, A.M., Reyes-Contreras, C., Neumann, P., Piña, B., Vidal, G., 2021. Presence and fate of micropollutants during anaerobic digestion of sewage and their implications for the circular economy: A short review. Journal of Environmental Chemical Engineering 9, 104931. https://doi.org/10.1016/j.jece.2020.104931

Viau, E., Bibby, K., Paez-Rubio, T., Peccia, J., 2011. Toward a consensus view on the infectious risks associated with land application of sewage sludge. Environmental Science & Technology 45, 5459–5469. https://doi.org/10.1021/es200566f

Viau, E., Peccia, J., 2009. Survey of wastewater indicators and human pathogen genomes in biosolids produced by Class A and Class B stabilization treatments. Applied and Environmental Microbiology 75, 164–174. https://doi.org/10.1128/AEM.01331-08

Wales, N., Carøe, C., Sandoval-Velasco, M., Gamba, C., Barnett, R., Samaniego, J.A., Ramos Madrigal, J., Orlando, L., Gilbert, M.T.P., 2015. New insights on single-stranded versus double-stranded DNA library preparation for ancient DNA. BioTechniques 59, 368–371. https://doi.org/10.2144/000114364

Wang, L., He, W., Yu, X., Hu, D., Bao, M., Liu, H., Zhou, J., Jiang, H., 2020. Coronavirus disease 2019 in elderly patients: Characteristics and prognostic factors based on 4-week follow-up. Journal of Infection 80, 639–645. https://doi.org/10.1016/j.jinf.2020.03.019

Warwick-Dugdale, J., Solonenko, N., Moore, K., Chittick, L., Gregory, A.C., Allen, M.J., Sullivan, M.B., Temperton, B., 2019. Long-read viral metagenomics captures abundant and microdiverse viral populations and their niche-defining genomic islands.

PeerJ 7. https://doi.org/10.7717/PEERJ.6800

Welsch, S., Müller, B., Kräusslich, H.-G., 2007. More than one door - budding of enveloped viruses through cellular membranes. FEBS Letters 581, 2089–2097. https://doi.org/10.1016/j.febslet.2007.03.060

Westhaus, S., Weber, F.-A., Schiwy, S., Linnemann, V., Brinkmann, M., Widera, M., Greve, C., Janke, A., Hollert, H., Wintgens, T., Ciesek, S., 2021. Detection of SARS-CoV-2 in raw and treated wastewater in Germany – Suitability for COVID-19 surveillance and potential transmission risks. Science of The Total Environment 751, 141750. https://doi.org/10.1016/j.scitotenv.2020.141750

White, R., Murray, S., Rohweder, M., 2000. Pilot Analysis of Global Ecosystems: Grassland Ecosystems. World Resources Institute, Washington, DC.

WHO, 2020. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020.

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wilhelm, S.W., Suttle, C.A., 1999. Viruses and nutrient cycles in the sea. BioScience 49, 781–788. https://doi.org/10.2307/1313569

Williamson, K.E., Corzo, K.A., Drissi, C.L., Buckingham, J.M., Thompson, C.P., Helton, R.R., 2013. Estimates of viral abundance in soils are strongly influenced by extraction and enumeration methods. Biology and Fertility of Soils 49, 857–869. https://doi.org/10.1007/s00374-013-0780-z

Williamson, K.E., Fuhrmann, J.J., Wommack, K.E., Radosevich, M., 2017. Viruses in soil ecosystems: An unknown quantity within an unexplored territory. Annual Review of Virology 4, 201–219. https://doi.org/10.1146/annurev-virology-101416-041639

Williamson, K.E., Wommack, K.E., Radosevich, M., 2003. Sampling natural viral communities from soil for culture-independent analyses. Applied and Environmental Microbiology 69, 6628–6633. https://doi.org/10.1128/AEM.69.11.6628-6633.2003

Withers, E., Hill, P.W., Chadwick, D.R., Jones, D.L., 2020. Use of untargeted metabolomics for assessing soil quality and microbial function. Soil Biology and Biochemistry 143, 107758. https://doi.org/10.1016/j.soilbio.2020.107758

Wolf, Y.I., Kazlauskas, D., Iranzo, J., Lucía-Sanz, A., Kuhn, J.H., Krupovic, M., Dolja, V.V., Koonin, E.V., 2018. Origins and evolution of the global RNA virome. mBio 9,

e02329–18. https://doi.org/10.1128/mBio.02329-18

Wolf, Y.I., Silas, S., Wang, Y., Wu, S., Bocek, M., Kazlauskas, D., Krupovic, M., Fire, A., Dolja, V.V., Koonin, E.V., 2020. Doubling of the known set of RNA viruses by metagenomic analysis of an aquatic virome. Nature Microbiology 5, 1262–1270. https://doi.org/10.1038/s41564-020-0755-4

Wong, K., Onan Brandon, M., Xagoraraki, I., 2010. Quantification of enteric viruses, pathogen indicators, and salmonella bacteria in class b anaerobically digested biosolids by culture and molecular methods. Applied and Environmental Microbiology 76, 6441–6448. https://doi.org/10.1128/AEM.02685-09

Woo, A.C., Gaia, M., Guglielmini, J., Da Cunha, V., Forterre, P., 2021. Phylogeny of the varidnaviria morphogenesis module: Congruence and incongruence with the tree of life and viral taxonomy. Frontiers in Microbiology 12, 1708. https://doi.org/10.3389/fmicb.2021.704052

Wu, F., Xiao, A., Zhang, J., Gu, X., Lee, W.L., Kauffman, K., Hanage, W., Matus, M., Ghaeli, N., Endo, N., Duvallet, C., Moniz, K., Erickson, T., Chai, P., Thompson, J., Alm, E., 2020. SARS-CoV-2 titers in wastewater are higher than expected from clinically confirmed cases. medRxiv 2020.04.05.20051540. https://doi.org/10.1101/2020.04.05.20051540

Wu, R., Davison, M.R., Gao, Y., Nicora, C.D., Mcdermott, J.E., Burnum-Johnson, K.E., Hofmockel, K.S., Jansson, J.K., 2021. Moisture modulates soil reservoirs of active DNA and RNA viruses. Communications Biology 2021 4:1 4, 1–11. https://doi.org/10.1038/s42003-021-02514-2

Wurtzer, S., Marechal, V., Mouchel, J.-M., Maday, Y., Teyssou, R., Richard, E., Almayrac, J.L., Moulin, L., 2020. Evaluation of lockdown impact on SARS-CoV-2 dynamics through viral genome quantification in Paris wastewaters. medRxiv. https://doi.org/10.1101/2020.04.12.20062679

Xie, W.-Y., McGrath, S.P., Su, J.-Q., Hirsch, P.R., Clark, I.M., Shen, Q., Zhu, Y.-G., Zhao, F.-J., 2016. Long-term impact of field applications of sewage sludge on soil antibiotic resistome. Environmental Science & Technology 50, 12602–12611. https://doi.org/10.1021/acs.est.6b02138

Xu, X., Bei, J., Xuan, Y., Chen, J., Chen, D., Barker, S.C., Kelava, S., Zhang, X., Gao, S., Chen, Z., 2020. Full-length genome sequence of segmented RNA virus from ticks

was obtained using small RNA sequencing data. BMC Genomics 21, 1–8. https://doi.org/10.1186/s12864-020-07060-5

Yin, X., Jiang, X.-T., Chai, B., Li, L., Yang, Y., Cole, J.R., Tiedje, J.M., Zhang, T., 2018. ARGs-OAP v2.0 with an expanded SARG database and Hidden Markov Models for enhancement characterization and quantification of antibiotic resistance genes in environmental metagenomes. Bioinformatics 34, 2263–2270. https://doi.org/10.1093/bioinformatics/bty053

Yong, C.Y., Yeap, S.K., Omar, A.R., Tan, W.S., 2017. Advances in the study of nodavirus. PeerJ 2017, e3841. https://doi.org/10.7717/peerj.3841

Yoshida, M., Mochizuki, T., Urayama, S.I., Yoshida-Takashima, Y., Nishi, S., Hirai, M., Nomaki, H., Takaki, Y., Nunoura, T., Takai, K., 2018. Quantitative viral community DNA analysis reveals the dominance of single-stranded DNA viruses in offshore upper bathyal sediment from Tohoku, Japan. Frontiers in Microbiology 9. https://doi.org/10.3389/fmicb.2018.00075

Zablocki, O., Adriaenssens, E.M., Frossard, A., Seely, M., Ramond, J.-B., Cowan, D., 2017. Metaviromes of extracellular soil viruses along a namib desert aridity gradient. Genome Announcements 5, e01470–16. https://doi.org/10.1128/genomeA.01470-16

Zablocki, O., Michelsen, M., Burris, M., Solonenko, N., Warwick-Dugdale, J., Ghosh, R., Pett-Ridge, J., Sullivan, M.B., Temperton, B., 2021. VirION2: a short- and long-read sequencing and informatics workflow to study the genomic diversity of viruses in nature. PeerJ 9, e11088. https://doi.org/10.7717/PEERJ.11088

Zell, R., Delwart, E., Gorbalenya, A.E., Hovi, T., King, A.M.Q., Knowles, N.J., Lindberg, A.M., Pallansch, M.A., Palmenberg, A.C., Reuter, G., Simmonds, P., Skern, T., Stanway, G., Yamashita, T., 2017. ICTV virus taxonomy profile: Picornaviridae. Journal of General Virology 98, 2421–2422. https://doi.org/10.1099/jgv.0.000911

Zeng, Q., Chisholm, S.W., 2012. Marine viruses exploit their host's two-component regulatory system in response to resource limitation. Current Biology 22, 124–128. https://doi.org/10.1016/J.CUB.2011.11.055

Zerzghi, H., Brooks, J.P., Gerba, C.P., Pepper, I.L., 2010. Influence of long-term land application of class b biosolids on soil bacterial diversity. Journal of Applied Microbiology 109, 698–706. https://doi.org/10.1111/j.1365-2672.2010.04698.x

Zhang, C.H., Schwartz, G.G., 2020. Spatial disparities in coronavirus incidence and

mortality in the United States: An ecological analysis as of May 2020. The Journal of Rural Health 36, 433–445. https://doi.org/10.1111/jrh.12476

Zhao, Y., Liu, Z., Wu, J., 2020. Grassland ecosystem services: a systematic review of research advances and future directions. Landscape Ecology 35, 793–814. https://doi.org/10.1007/s10980-020-00980-3

Zipper, H., 2004. Investigations on DNA intercalation and surface binding by SYBR green i, its structure determination and methodological implications. Nucleic Acids Research 32, e103–e103. https://doi.org/10.1093/nar/gnh101

Zuccato, E., Chiabrando, C., Castiglioni, S., Bagnati, R., Fanelli, R., 2008. Estimating community drug abuse by wastewater analysis. Environmental Health Perspectives 116, 1027–1032. https://doi.org/10.1289/ehp.11022