Virus Dynamics and their Interactions with Microbial Communities and Ecosystem Functions in Engineered Systems.

A thesis submitted to Newcastle University in partial fulfilment of the requirements for the degree of Doctor of Philosophy in the Faculty of Science, Agriculture and Engineering.

Author: Mathew Robert Brown

Supervisor: Professor Tom Curtis

Co-Supervisor: Dr Russell Davenport

2019

Declaration

I hereby certify that this work is my own, except where otherwise acknowledged, and that it has not been submitted for a degree at this, or any other, institution.

Mathew Robert Brown

Abstract

Climate change, population growth and increasingly strict environmental regulation means the global water industry is currently facing an unprecedented coincidence of challenges (Palmer, 2010). Better microbial ecology could significantly contribute, since explicitly engineering and maintaining efficient and functionally stable microbial communities would allow existing assets to be optimised and their robustness improved. Given its role in natural systems viral infection could be an important, yet overlooked, factor. Here we attempt to address this lacuna, particularly within activated sludge systems.

To facilitate this process we developed, optimised and validated a flow cytometry method, allowing rapid (relative to other methods), accurate and highly reproducible quantification of total free viruses in activated sludge samples (mixed liquor (ML)). Its use spatially identified viruses are highly abundant, with concentrations ranging from 0.59 - 5.14 × 10^9 viruses mL^{-1} across 25 activated sludge plants.

Subsequently we applied this method to ML collected from one full- and twelve replicate lab-scale activated sludge systems respectively. At both scales viruses in the ML were shown to be both abundant and temporally/spatiotemporally dynamic, thus ever present across activated sludge systems. Through statistical inference they were shown to be associated (positively) with total host (bacterial) abundance, with microbial community structure and with a systems function (the removal of COD and NH_4^+-N from influent wastewaters), whilst exogenous factors, particularly those involved in adsorption processes, played an important role in their dynamics.

Evidence of predator-prey dynamics between a subset of measured viruses and a key functional group (ammonia oxidising bacteria (AOB)) within the full-scale system is also presented, whilst a detailed examination of all garnered abundances highlights the relative abundance of viruses, as reported in marine systems, declined with increasing host density. Finally preliminary metagenomic data shows wastewater viromes are largely phylogenetically and functionally uncharacterised, yet relative abundances of known viruses vary throughout the wastewater treatment stream.

Considering the evidence presented viruses appear to play a more central role in the dynamics of activated sludge systems than hitherto realised and thus should be considered more frequently when assessing the key factors governing bacterial abundance, community composition and functional stability.

Table of Contents

List of Figures

CHAPTER 1

CHAPTER 2

CHAPTER 3

List of Tables

APPENDIX VI

APPENDIX VII

CHAPTER 1

INTRODUCTION

1.1. Background

Biological wastewater treatment, in the form of aerobic (activated sludge, trickling filter, etc.) and anaerobic (upflow anaerobic sludge blanket, expanded granular sludge bed etc.) processes, is the largest application of biotechnology in the world (Wang *et al*., 2012a, see Appendix I for an introduction to wastewater treatment). By concentrating and managing microbial communities the beneficial activities of naturally occurring microorganisms are harnessed and accelerated, thus enabling the degradation of oxygen-depleting organics, transformation of toxic substances and removal of nutrients and pathogens from wastewaters (Wells *et al*., 2011; Wang *et al*., 2012a). Despite the successful and widespread application of these globally important processes the underpinning microbial communities have, until recently, been difficult to study. Little is known about the complex dynamics of the microbial populations interacting in bioreactors and how these dynamic interactions affect a systems performance and functional stability (Valentin-Vargas *et al*., 2012). Thus our fundamental knowledge and understanding of how such processes actually work is lacking.

Current practice in wastewater treatment plant design was established in the 1960's when Downing *et al*. (1964) and Lawrence and McCarty (1970) developed a theory combining simple mass balance concepts with Monod Kinetics (Curtis and Sloan, 2006). They consider waste biodegradation as a problem in chemical thermodynamics, where groups of microorganisms are seen as omnipresent catalysts. All that an engineer need do is ensure the correct environmental and chemical conditions prevail for a function to be fulfilled. Such tools however, do not always predict the engineered reality, thus engineers are never sure if they have the optimal microbial community for the desired function, if they can establish new functions or if they can restore functions that have been lost (Curtis and Sloan, 2006). Moreover the inherent robustness of a system, that is its ability to resist inhibition or shock loading, is simply unknown, thus failure in such systems is common, unpredictable and often inexplicable (Curtis and Sloan, 2006). These deficiencies arise because whilst we can predict the size of a system we cannot predict its composition or community structure, the two things that ultimately define its operational characteristics (Curtis and Sloan, 2006).

Hitherto our inability to do this has been tolerable; however the global water industry is currently facing an unprecedented coincidence of challenges (Palmer, 2010). Climate change and its mitigation is requiring water companies to minimise their carbon footprint, increasing energy, and therefore operating, costs, which is undermining their profit margins (Palmer, 2010). Moreover rapid and localised population growth, coupled with increasingly strict environmental regulation,

is requiring additional and more advanced levels of wastewater treatment, thus necessitating further investment in and development of the capacity and capability of a water company's asset base (Palmer, 2010; Shoener *et al*., 2014). These issues are compounded by the long life span of capital assets, varying from 15 - 30 years for mechanical and electrical assets and 30 – 60 years (or more) for civil assets, as it may take a generation to replace them with more energy efficient formats (Palmer, 2010). Consequently explicitly engineering and managing the best microbial community is of increasing strategic importance, by doing so existing assets can be optimised, in terms of treatment quality, energy efficiency and the functions required, and robustness improved. In order to achieve this and facilitate rational improvements in system design and operation a firm understanding of the microbial ecology of wastewater treatment bioreactors (WWTB's) is essential.

The advent and application of a plethora of new culture independent molecular tools namely: quantitative polymerase chain reaction (qPCR), fluorescence in situ hybridisation (FISH), terminal restriction fragment length polymorphism (T-RFLP) and more recently pyrosequencing holds promise for providing new insights into this "black box". Using such techniques wastewater engineers have elucidated the microbiology of important wastewater treatment processes, including nitrogen transformations (Schmidt *et al*., 2003), biological phosphorous removal (Seviour *et al*., 2003) and anaerobic degradation networks (Talbot *et al*., 2008). Moreover much progress has been made in ascertaining the microbial basis for filamentous bulking (Martins *et al*., 2004) and foaming (Blackall *et al*., 1996; Davenport *et al*., 2000) in activated sludge processes. However perhaps the most exciting and significant advancement facilitated by these techniques is the ability to now link a systems performance and function with the dynamics of its microbial community.

There is a growing body of evidence linking community composition and diversity with the function (or performance) and stability of a WWTB, a concept borrowed from microbial ecology (Cook *et al*., 2006; Figuerol and Erijman, 2010; Saikaly and Oerthe, 2011; Hernandez-Raquet *et al*., 2013). The function and functional stability of natural systems has been proposed to correlate with species richness, the number of species, and evenness, the relative abundance of a species (Naeem and Li, 1997; Stirling and Wilsey, 2001; Tilman *et al*., 1997; Wittebolle *et al*., 2009a). Species richness promotes increased productivity (or treatment performance) through resource (oxygen depleting organics, toxic substances, nutrients) partitioning and complimentary functional traits, thus more of the overall resource is consumed (Cardinale *et al*., 2002; Bell *et al*., 2005; Latta *et al*., 2011). Moreover there is an increased probability of a species with a large effect

on the ecosystem function (process performance) being present (Bell *et al*., 2005). Species evenness increases functional redundancy and thus increases resistance to a perturbation (inhibition or shock loading), in that it ensures the presence of a reservoir of species able to perform the same ecological function (treatment process) (Briones and Raskin, 2003; Wittebolle *et al*., 2009a). Thus to break new ground in the treatment of wastewater one could argue answering the fundamental question of what dictates community composition and diversity in such systems is paramount (Pholchan *et al*., 2013).

Ample inspiration for those seeking such an answer can be found in classical ecological literature, since the rules governing engineered and natural systems should be alike as the microbe is unaware of the distinction (Pholchan *et al*., 2013). With respect to community assembly one can borrow from two opposing, but not mutually exclusive, perspectives (Pholchan *et al*., 2013). Classically microbial community composition is thought to be shaped by deterministic factors such as competition and niche differentiation, where spatial and temporal environmental heterogeneity dictates (Tilman, 2004; Ofiteru *et al*., 2010). Accordingly numerous studies have revealed that certain environmental/operational conditions exert a distinctive selective pressure, indicating the important contribution of niche based mechanisms on community assembly and dynamics in WWTB's (Van der Gast *et al*., 2004, 2008; Dytczak *et al*., 2008; Wittebolle *et al*., 2009b; Wells *et al*., 2009, 2011; Huang *et al*., 2010; Falk and Wuertz, 2010;).

However not all temporal variations in community composition are related to operational and/or environmental parameters. The finding that replicated communities undergo erratic changes in time and the fact that distinct communities develop even when great care is taken to operate bioreactors in parallel, under identical environmental conditions has led to an opposing perspective: neutral theory (Fernandez *et al*., 2000; Kaewpipat and Grady, 2002; Gentile *et al*., 2007; Beecroft *et al*., 2012). In neutral theory the formation of a community is viewed as a stochastic process, in which all species are ecologically equivalent and colonise an environment due to an amalgamation of random, continuous processes, birth, death, dispersal and speciation (Bell, 2001; Hubbell, 2001). Despite its apparent simplicity and small number of parameters neutral theory has been successfully observed in and applied to microbial communities (Sloan *et al*., 2006, 2007; Woodcock *et al*., 2007). In fact Ofiteru *et al*. (2010) showed that the population dynamics of an activated sludge community in California were consistent with neutral community assembly, with chance and random immigration playing an important and predictable role. For two AOB and two heterotrophic taxa they were able to explain 23% and 20% respectively of the variance in the time series data of abundance using purely neutral processes (Ofiteru *et al*. 2010).

Intuitively however we know that community assembly and development is not as black and white; in reality the two perspectives are not mutually exclusive because birth, death and immigration are unavoidable features of microbial life whose rates are affected by niche differentiation (Gravel *et al*., 2006; Pholchan *et al*., 2013). Indeed neutral and deterministic components have been shown to operate simultaneously in the community assembly of activated sludge flocs (Ayarza *et al*., 2010; Ayarza and Erijman, 2011), activated sludge bioreactors (van der Gast *et al*., 2008; Valentin-Vargas *et al*., 2012) and microbial electrolysis cells (Zhou *et al*., 2013). Likewise when Ofiteru *et al*. (2010) supplemented neutral processes with environmental influence, by giving an advantage to the reproduction (or birth) rate of individual taxa, a larger percentage of the AOB and heterotroph variability through time could be explained, 37% and 28% respectively.

Whilst microbial ecologists and wastewater engineers have only recently attempted to explain community assembly by incorporating both components (Tilman, 2004; Gravel *et al*., 2006; Haegeman and Loreau, 2011), what is striking, at least when looking at the data of Ofiteru *et al*. (2010), is the substantial amount of unexplained variation. The authors suggest this could be attributable to unmeasured environmental factors, a nonlinear relationship between environment and advantage or substantial measurement error due to methodological limitations (Ofiteru *et al*., 2010). Whilst these are conceivable and will undoubtedly account for some of the discrepancy one could argue that a vital driver of microbial community assembly and development is still not accounted for in such neutral-niche models, thus a major portion of what dictates function and functional stability in WWTB's is still unknown. One can again look to ecological literature to identify additional drivers of diversity in natural systems, for these should again be applicable in an engineered setting.

When looking at natural microbial communities what becomes apparent is the presence of an additional layer of complexity, in that bacteria and/or archaea are subject to strong predation pressure from viruses and, to a lesser extent, protozoan bacteriovores (e.g. ciliates and flagellates) (McMahon *et al*., 2007). Subsequently bacterial fitness is measured not only by their adaptation to available resources, environmental conditions (the niche) and stochastic processes (the neutral) but also by their adaptation to the biotic environment (Rodriguez-Valera *et al*., 2009). Bacteriophage (viruses whom attack bacteria) predation is thus a major cause of bacterial mortality and consequently is thought to be a key driver in the functional structure (composition), functional stability and metabolic characteristics and activity (carbon and nutrient cycling) of naturally occurring bacterial communities (Rodriguez-Valera *et al*., 2009; Winter *et al*., 2010; Breitbart, 2012). It seems plausible that such processes are at work in engineered biological

systems and thus their effect on community composition and diversity and subsequently on function and functional stability could be important.

1.2 Viruses in Wastewater Treatment Processes

Viruses, including bacteriophage, (see Appendix II for an introduction to bacteriophage) are long known to be found in WWTB's (Ewert and Paynter, 1980; Ogata *et al*., 1980), however only very recently have estimates of abundance sparked interest in their potential role in the ecology of such systems. A trend perhaps analogous to, although slightly lagging, phage research in natural, chiefly aquatic systems (Weinbauer, 2004; Breitbart, 2012). Otawa *et al*. (2007) applied epifluorescence microscopy (EFM) to estimate total viral abundance in 18 full scale activated sludge bioreactors, reporting concentrations of 10^8 – 10^9 virus like particles (VLP) mL^{-1}. Similar values were reported for an activated sludge plant in Singapore (Wu and Liu, 2009) and for anaerobic digesters treating waste activated sludge (Wu and Liu, 2009), brewery waste (Park *et al*., 2007) and acetate (Chien *et al*., 2013) respectively. These studies highlight the fact that viral abundance in WWTB's is amongst the highest, if not the highest studied in any system (Wommack and Colwell, 2000).

Phage diversity has also been estimated in WWTB's. Using pulsed field and field inversion gel electrophoresis (PFGE, FIGE), which separate phage populations based on genome size, the composition of phage communities has been shown to vary between activated sludge plants and over time in a lab scale bioreactor (Otawa *et al*., 2007), as well as between different stages of a full scale treatment process (Park *et al*., 2007; Wu and Liu, 2009). Though we can infer from these studies that phage community composition is variable across time and space only highly abundant genomes are detected using these approaches and phages with similar genome sizes are not separated. More recently Tamaki *et al*. (2012) applied a metagenomic approach to investigate viral diversity in all four stages (influent, activated sludge, anaerobic digester and effluent) of a typical wastewater treatment plant. The study revealed a high degree of novelty in the diversity of the viral communities, a finding recently corroborated in a suite of anaerobic digesters (Calusinska *et al*., 2016), and highlighted < 5% of metagenomic sequences are similar to those present in other environments. Despite this the diversity estimates for the four viromes was moderate with 423 – 560 species recognised, a number comparable to that found in freshwater environments but lower than that found in oceans and soils (Tamaki *et al*., 2012). Interestingly a high number of viral genotypes (> 82%) were shared across the four viromes, although the abundance of individual species was variable.

1.2.1. The Role of Bacteriophages in Wastewater Treatment: Interactions with Microbial Community composition

1.2.1.1 Kill the Winner

Whilst bacteriophages are indisputably a dynamic and variable component of WWTB's their impact on the microbial community and subsequently treatment performance of such systems is still inadequately understood. A commonly accepted model to describe phage-host interactions in microbial ecology is that of killing the winner (KtW), an extension of classical Lotka-Volterra predator prey dynamics to the microbial world (Thingstad and Lignell, 1997; Thingstad, 2000). In KtW the assumption is that bacterial communities contain two populations competing for the same limited resource (e.g. phosphate): competition specialists, that use the resource for growth and reproduction, and defence specialists, that utilise it to counteract stress (e.g. predation, unfavourable environmental conditions, or competitive ability) (Winter *et al*., 2010; Breitbart, 2012). In the absence of predators the competition specialists would dominate, sequestering all of the limited resource (Breitbart, 2012). Intuitively one can see that if predation pressure results in the selective loss of competition specialists and prevents complete resource sequestering by this population then a share will become available to the defence specialists (Winter *et al*., 2010). The consequence of this phage control and resultant trade-offs (competitive vs. defensive ability) is coexistence, or in a different guise, maintenance, of a communities diversity and diversification (Weinbauer and Rassoulzadegan, 2004; Brockhurst *et al*., 2006; Hewson and Fuhrman, 2006, 2007).

One possible prediction of this model is the periodic rise and fall of specific bacterial hosts and their corresponding phages in a classical predator prey fashion, resulting in the cycling of the most dominant or active taxa (Fig. 1.1 A, Rodriguez-Brito *et al*., 2010; Shapiro and Kushmaro, 2011). Another implication is that some of the most active bacteria (competition specialists) may be rare in ambient communities due to high levels of phage pressure (Fig. 1.1 D, Harcombe and Bull, 2005; Suttle, 2007; Winter *et al*., 2010), conversely the most dominant bacteria may constitute defence specialists that are inferior competitors for resources but gain a selective advantage as a result of phage predation (Fig. 1.1 B, Bouvier and del Giorgio, 2007). It is noteworthy that KtW predicts that under highly productive conditions, such as those in a WWTB, predation is the major regulatory mechanism for community composition, whereas in environments with low productivity competition drives community composition (Winter *et al*., 2010).

Patterns consistent with KtW have recently been demonstrated in a full scale membrane bioreactor treating industrial wastewater, where PFU counts on several bacterial strains appeared to oscillate in correlation with their suggested hosts (Shapiro *et al*., 2010). A similar pattern was reported for a phage and its host inoculated into a laboratory scale activated sludge reactor (Lee *et al*., 2007), whilst the change in dominant strains of a full-scale system was attributed to strain specific phage predation (Ogata *et al*., 1980). Successive and significant shifts in the species composition of an anaerobic digesters bacterial community could also be attributed to cyclic phage predation (Zumstein *et al*., 2000), whilst fluctuating phage abundances in full and lab scale systems is also consistent with KtW dynamics (Hantula *et al*., 1991; Otawa *et al*., 2007).

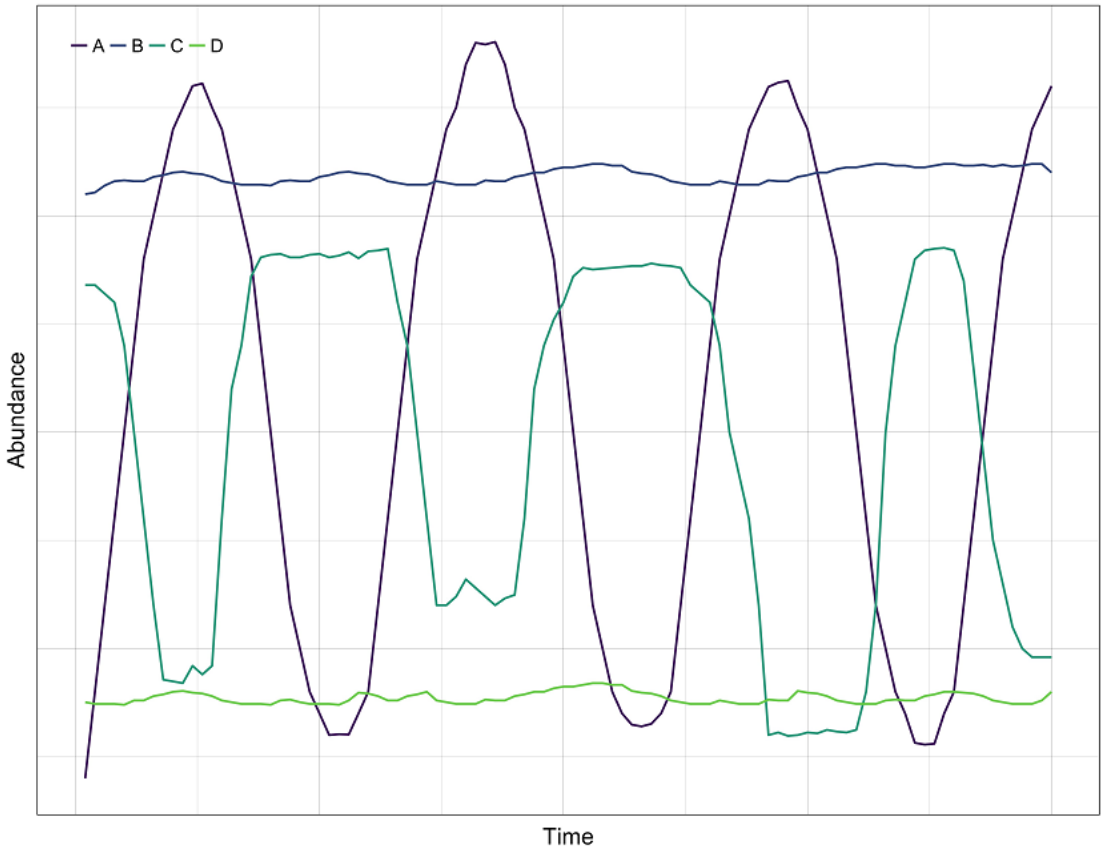

Figure 1. 1. Possible scenarios for temporal changes in host abundance as a consequence of viral infection, (**A** – **D**) described in the text.

1.2.1.2 Antagonistic Coevolution

An alternative mechanism describing phage host interactions is antagonistic coevolution (ACE), which can be viewed as a continuum (Fig. 1.1 C, Hall *et al*., 2011). At one extreme arms race dynamics (ARD) cause continual reciprocal evolution of host resistance and, to counter act this, phage infectivity (Buckling and Rainey, 2002; Forde *et al*., 2008; Gomez and Buckling, 2011), imposing directional selection for hosts and phages with broader resistance and infectivity ranges

respectively. At the other extreme fluctuating selection dynamics (FSD), in which there is no directional change in the evolution of resistance and infectivity ranges just inherent differences across genotypes, is governed by negative frequency-dependent selection, favouring hosts and phages that are resistant and infectious to the most frequently occurring phages and hosts respectively (Hall *et al*., 2011; Betts *et al*., 2014). Thus ACE, whether through ARD or FSD, is believed to play a critical role in phage host dynamics, speciation, coexistence and ultimately a system's biodiversity and function (Buckling and Rainey, 2002; Gomez and Buckling, 2011). Interestingly, due to genetic (phage) and metabolic (host) constraints on mutation (Lenski and Levin, 1985; Bohannan and Lenski, 2000; Hall *et al*., 2011), ARD has been proposed to predominate during initial stages of coevolution, when mutations are largely cost-free, but then give way to FSD, when mutations impose costs (Hall *et al*., 2011).

Possible evidence for such processes was reported for two geographically remote Enhanced Biological Phosphorus Removal (EBPR) bioreactors enriched with *Candidatus* Accumulibacter phosphatis (CAP), a phosphate accumulating organism (PAO) (Kunin *et al*., 2008). Both CAP communities were consistently exposed to viral activity and were dominated by one genetically similar strain, sharing 95% nucleotide identity across most of the genome. One striking difference was the variability in extracellular polymeric substance (EPS) expression and clustered regularly interspaced short palindromic repeat (CRISPR) elements, both of which serve as defence mechanisms against phage predation (Hughes *et al*., 1998; Sorek *et al*., 2008). The authors concluded that the differences were a result of recent ACE dynamics, were by rapid acquisition and substitution of EPS gene cassettes and CRISPR elements was a bacterial response to local phage predation pressure (Kunin *et al*., 2008). A similar explanation, i.e. local evolution of resistance, could be applied to two identical EBPR CAP dominated bioreactors run by He *et al*. (2010), where a random shift in the dominant CAP clade (Clade IA to Clade IIA) occurred within one reactor without any detectable environmental change. Conversely the evolution of phage infectivity, through the incorporation and expression of anti-CRISPR genes (Bondy-Denomy *et al*., 2012; Bondy-Denomy *et al*., 2015), could have occurred, thereby incurring a fitness cost on Clade IA and giving a competitive advantage to Clade IIA. Such explanations could also be applicable to the 6 identical denitrifying bioreactors of Gentile *et al*. (2007), where the community structure in two diverged from the others, chiefly due to the increased abundance of one bacterial species. The demonstration of more than one infection cycle for phages targeting a single host could similarly be evidence of such dynamics in a wastewater environment, with several weeks or months separating each cycle (Huntula *et al*., 1991; Shapiro and Kushmaro, 2010).

1.2.1.3 Horizontal Gene Transfer

Bacteriophages can also impact bacterial diversity via direct (by transduction) or indirect (by transformation) horizontal gene transfer (HGT) (Jacquet *et al*., 2010; Breitbart, 2012). Transformation consists of the assimilation and incorporation of free or extracellular DNA (eDNA) by bacteria, thus the lytic action of bacteriophages may stimulate this process and considerably enhance the reservoir of genetic information available (Jacquet *et al*., 2010; Breitbart, 2012). Transduction involves a portion of the genetic material of a host cell being packaged by a phage and transferred to a new host during infection (Weinbauer, 2004; Jacquet *et al*., 2010). Both are followed by the subsequent expression of the transferred genetic traits in the recipient host. Thus HGT can contribute to genetic variation in host populations (alteration of metabolic properties, increased fitness, resistance to stress etc.), driving genetic evolution and consequently influencing population dynamics (Replicon *et al*., 1995; Chiura, 1997; Brussow *et al*., 2004; Weinbauer, 2004; Jacquet *et al*., 2010).

The high concentration of eDNA in WWTB's and its importance to floc and granule formation would imply the opportunity for transformation in such systems is high, however the direct effect of phage induced transformation in WWTB's is difficult to quantify and currently unknown (Dominiak *et al*., 2011). In contrast the presence of transducing phage's in WWTB's was demonstrated by Sander and Schmieger (2001) and the potential for gene transduction highlighted by Parsley *et al*. (2010a), who detected diverse antimicrobial resistant genes in a metagenomic survey of both bacterial and viral communities from an activated sludge system. Colomer-Lluch *et al*. (2011) demonstrated through transfection, a non-viral transfer of nucleic acids, that antibiotic resistant genes originating from raw sewage phage isolates were able to confer resistance to a recipient bacterial strain, suggesting transduction processes could also do so. More recently estimates of the extent of phage mediated HGT in a municipal wastewater treatment plant was reported by Del Casale *et al*. (2011a, 2011b). Using bacterial 16S rRNA genes transduced from bacteriophages the authors reported that transduction occurs temporally in several dominant bacterial groups (proetobacteria, Firmicutes and Rhodococcus, Del Casale *et al*., 2011a, 2011b).

A further consideration to the importance of transduction is the presence of polyvalent bacteriophages, phages with a wide host range, in WWTB's. Several studies have demonstrated polyvalent phage isolates in raw sewage (Namura *et al*., 2008; Synnott *et al*., 2009), bench scale (Khan *et al*., 2002a, 2002b) and full-scale activated sludge (Huntula *et al*., 1991; Thomas *et al*., 2002) and membrane bioreactors (Shapiro and Kushmaro, 2010) respectively. Such phages are

typically able to infect several strains of the same species, however some polyvalent phage isolates have been found to infect hosts from different bacterial classes (Jensen *et al*., 1998; Sullivan *et al*., 2003) as well as bacterial isolates with different gram staining (Khan *et al*., 2002a, 2002b). Thus Fard *et al*. (2011) were able to demonstrate the transfer of genes between different taxa as a result of transducing polyvalent bacteriophages.

Whilst transduction is irrefutably present in WWTB's until recently it was considered a rare event, occurring once every 10^7 – 10^9 phage infections and thus having minimal impact on a systems bacterial community (Muniesa *et al*., 2013). A recent study, however, indicates that transduction may occur at a frequency several orders of magnitude greater than previously thought and in a wide range of bacteria (Kenzaka *et al*., 2010). With these frequencies, the presence of transducing polyvalent bacteriophages and the sheer abundance of phages and hosts found in WWTB's gene transfer by transduction could take place an exceptional number of times per second (Muniesa *et al*., 2013). Consequently transducing phages could mediate the mobilisation and transfer of genetic material within WWTB's and in nature to a far greater extent than previously thought.

1.2.1.4 Lysogeny

Lysogenic, or temperate, bacteriophages, phages that remain dormant within the host genome (known as a prophage) and replicate with it until a lytic cycle is induced (see Appendix II for more information), are also likely drivers of microbial dynamics within WWTB's, since they can provide bacteria with virulence and fitness factors affecting cell metabolism, bacterial adhesion, colonisation, immunity and antibiotic resistance (Wagner and Waldor, 2002). Currently percent lysogeny, the percentage of bacterial cells in a community containing an inducible phage genome, is unknown in WWTB's, however it is speculated that more than 80% of bacterial strains contain prophages (Canchaya *et al*., 2003). With this in mind and considering lysogeny may confer an ecological advantage in WWTB's, since prophages are retained within the system and free phage particles are constantly washed out with process effluents, the lysogenic life cycle has the potential to be extremely influential on bacterial populations in such systems (Shapiro and Kushmaro, 2011). Its true influence however is still to be determined.

1.2.2 The Role of Bacteriophages in Wastewater Treatment: Interactions with food web processes and biogeochemical cycles

As agents of mortality phage's also have the ability to influence biochemical cycling of nutrients, elements and both the flux and character of carbon, since the lysis of host cells not only releases progeny virus particles but also cell debris (Jacquet *et al*. 2010; Winter *et al*., 2010). This cell debris

is made up of a cocktail of dissolved molecules (sugars, proteins, peptides, amino acids, and nucleic acids), macro and micronutrients (e.g. phosphorous, nitrates, iron etc.) plus colloids and cell fragments, typically defined as dissolved organic matter (DOM) (Fuhrman, 1999; Jacquet *et al*., 2010; Winter *et al*., 2010). This DOM is then rapidly or eventually recycled back into the food web and becomes available for bacteria to consume. In oceans this so called viral shunt is suggested to account for ~20 – 30% of the daily carbon pool (Wilhelm and Suttle, 1999; Middelboe and Jorgensen, 2006; Suttle, 2007) and significantly increase dissolved levels of phosphorous, iron, selenium and organic nitrogen (Middelboe *et al*., 1996; Gobler *et al*., 1997; Wilhelm and Suttle, 1999; Poorvin *et al*., 2004; Shelford *et al*., 2012). The net effect of this, illustrated by Fuhrman (1992) using a theoretical model, is that a system with 50% bacterial mortality from viruses, compared to a system with no viruses yet containing identical food webs, has 27% more bacterial respiration and production. In essence phage activity via the viral shunt has the potential to help maintain higher levels of biomass and productivity within a system.

Possible evidence for such a process within WWTB'S was reported by Rosenberg *et al*. (2010) for a two-stage bioreactor treating oil refinery drainage water. The authors observed higher than expected organic carbon removal, 90% compared with the theoretically expected 69%, across the two-stage system. Virus counts indicated the presence of phage predation and pointed to a potential viral shunt type process, increasing bacterial productivity and thus organic carbon removal to levels greater then theoretically expected. Further evidence supporting viral shunt processes within WWTB's is lacking, however the increased DOM, particulate matter degradation and biogas production observed in the anaerobic digestion of ultrasound pre-treated, compared with untreated, waste activated sludge perhaps also emphasises their potential contribution in WWTB's (Braguglia *et al*., 2012a, 2012b). The increase in DOM likely maintains a higher bacterial/archaeal abundance and thus productivity; like the viral shunt, contributing to increased anaerobic digestion efficiency.

1.2.3 The Role of Bacteriophages in Wastewater Treatment: Can they influence process performance?

Although modest, literary evidence would imply bacteriophages can directly affect the microbial community in WWTB's, however there is little indication of their role in functional performance (other than that aforementioned). It could be argued, since phage predation can drive microbial diversification, that phage activity maintains, and to some extent creates, the functional redundancy that confers stability of performance and resilience to perturbations within WWTB's (Shapiro and Kushmaro, 2011). This functional redundancy then buffers the loss of key species

through phage predation, masking their true effect on a systems performance. One could also speculate that the bacterial/archaeal communities within WWTB's, due to the rich resource environment, are dominated by defence specialists (Winter *et al*., 2010), hence phage predation has little effect on the most abundant and active taxa and thus system performance. Both notions however are yet to be tested.

Perhaps an easier question to answer is the affect phages have on specific metabolic processes, such as nitrification, de-nitrification, phosphate accumulation and methanogenesis, within WWTB's, since such processes are often dictated by a small number of distinct, functional microbes (Shapiro and Kushmaro, 2011). Possible evidence for the involvement of bacteriophages in the reduction of phosphorus (P) removal from a CAP enriched laboratory EBPR reactor was reported by Barr *et al*. (2010). To investigate further 4 additional EBPR reactors were inoculated with active biomass and then 2 of them spiked with supernatant from the original, failing reactor. Both "infected" reactors, when compared to the controls, exhibited negative P removal, strongly indicative of P release through cell lysis, and elevated phage counts, thus supporting the original hypothesis that phage activity was responsible for reduced P removal.

Prophage induction, following copper (CU), cyanide and ciprofloxacin addition, has also been reported as a possible contributor to the reduced performance of a laboratory scale EBPR reactor (Motlagh *et al*., 2015). When compared to a control reactor a significant increase in phage abundance, coupled with a significant decrease in the most dominant CAP clades and P removal efficiency, was observed after spiking 0.5mg/l CU, 500µg/L cyanide and 0.05 – 0.4 µg/L ciprofloxacin respectively. Once induced it was also shown, through isolation and infection of fresh biomass, that the now free phage, following the lytic infection cycle, could also decrease P uptake rates and removal efficiency when compared to uninfected biomass.

The lysogenic-lytic switch was also witnessed by Choi *et al*., (2010) in a pure culture of *Nitrosospira multiformis* 25196, a member of genus *Nitrosospira* which are active AOB in nitrifying activated sludge plants (Siripong and Rittmann, 2007; Whang *et al*., 2009), as a result of heavy metal (chromium) exposure and various other physico-chemical stress factors (pH, temperature and potassium cyanide). Increased phage counts, coupled with deterioration in ammonia oxidation, could be attributed to high and low pH levels (5 and 8), low concentrations of chromium (0.002 – 0.1mM) and cyanide (0015 – 0.15mM) and higher than normal (35^{o}C) temperatures. Foaming, biomass bulking and decreased process performance in WWTB's are known outcomes of sudden temperature shifts (Morgan-Sagastume and Allen, 2003, 2005a, 2005b; Nadarajah *et al*., 2007), heavy metal surges (Stasinakis *et al*., 2003; Hu *et al*., 2004; You *et al*., 2009), wastewater toxicity

(Kim *et al*., 2011; Han *et al*., 2014) and pH shocks (Lü *et al*., 2008; Gao *et al*., 2010; Law *et al*., 2011). Thus massive prophage induction, caused by such stress factors, followed by the sudden lysis of a proportion of a WWTB's microbial community could also cause such failures and thus can't be ruled out as a possible, unexplored cause (Shapiro and Kushmaro, 2011).

Indeed lytic bacteriophage have been shown to influence filamentous organisms associated with biomass bulking (Kotay *et al*., 2011) and sludge foaming (Khairnar *et al*., 2014; Liu *et al*., 2015) within activated sludge systems, albeit as a potential method of control. Kotay *et al*. (2011), using a lytic phage isolated from mixed liquor, showed biomass bulking caused by *Haliscomenobacter hydrossis* was significantly reduced following spiking of the isolated phage, whilst Liu *et al*. (2015) observed suppression in the abundance of a number of foaming associated filamentous *Gordonia* strains inoculated into mixed liquor following addition of isolated, infective bacteriophages.

1.3 Insights from Natural Systems

In reviewing recent work concerning phage ecology in WWTB's it is apparent that our knowledge and understanding of phage dynamics in engineered systems, and their potential role in controlling bacterial abundance, community composition and ecosystem function, is fundamentally lacking. This is particularly apparent when work in WWTB is compared to natural, chiefly marine, systems, where viruses are generally accepted as ecologically important components of the environment (Breitbart, 2012). Great insights thus can be gained by adopting approaches used in these environments, as well as learning from their shortfalls and future perspectives.

1.3.1 Expanding the Spatial and Temporal Resolution of Viral Studies

In the late 1980's and throughout the 1990s marine viral ecology underwent a transformation following the development of culture independent enumeration techniques, namely transmission electron microscopy (TEM, Bergh *et al*., 1989), EFM (Hara *et al*., 1991; Hennes and Suttle, 1995) and flow cytometry (FCM, Marie *et al*., 1999; Brussaard, 2004), and thus the realisation that viruses were highly abundant throughout the global oceans (e.g. Jiang and Paul, 1994; Weinbauer *et al*., 1995; Bratbak *et al*., 1996; Li and Dickie, 2001). Thus the initial focus of this new era in marine viral ecology was the spatiotemporal dynamics of total virus numbers, viral production, the virus to bacteria/microbe ratio and their role in bacterial mortality (Brum and Sullivan, 2015). More recently the emergence of metagenomic based studies has made clear the incredible diversity of marine viruses (Breitbart *et al*., 2002; Breitbart *et al*., 2004) and that the composition

of viral assemblages changes in different geographic regions (Venter *et al*., 2004; Angly *et al*. 2006) and at different depths (Hurwitz and Sullivan, 2013).

Whilst such studies underpin our understanding of virus dynamics within marine systems it is increasingly apparent that spatial and temporal studies are rather limited, a consequence of limited "ship time" and other logistical constraints, and that these snapshots of viral abundance and community composition are inadequate for describing the microbial ecology of marine systems (Hewson *et al*., 2006; Breitbart, 2012; Brum and Sullivan, 2015; Wigington et al., 2016), especially considering we lack an understanding of the degree of spatiotemporal variability that may exist. As such recent reviews in marine viral ecology have emphasised the need for functionally relevant time series data that is undertaken at appropriate temporal and spatial scales (Breitbart, 2012; Brum and Sullivan, 2015; Wigington *et al*., 2016). The incorporation of exogenous factors, typically sampled infrequently and using limited analytical techniques, into such studies has also been emphasised as a priority (Wigington *et al*., 2016), since variations across time and space have the ability to influence the life history traits of viruses and their hosts, e.g. growth, resistance and infectivity, and thus their abundance and diversity.

1.3.2 Identifying True Viral Diversity

Double stranded DNA (dsDNA) phages were generally thought to account for the majority of marine viruses (Wommack and Coldwell, 2000; Breitbart *et al*., 2004; Weinbauer, 2004). However the recent discovery of single stranded DNA (ssDNA) (Angly *et al*., 2006; Tucker *et al*., 2011; Labonte and Suttle, 2013) and RNA (Culley *et al*., 2006; Lang *et al*., 2009; Gustavsen *et al*., 2014) viruses in metagenomic studies, as well as in culture (Nagasaki, 2008), suggests a subset of the marine viral community has been overlooked. Indeed the abundance of ssDNA and RNA viruses in the ocean is currently unclear. Recent work indicates RNA viruses may be highly abundant (Steward *et al*., 2013), yet EFM and FCM based counts, believed to underestimate both viral types due to poor staining of their small genomes (Brussaard *et al*., 2000; Tomaro and Nagasaki, 2007; Holmfeldt *et al*., 2012), typically exceed corresponding TEM (which doesn't rely on nucleic acid staining) based counts, suggesting their contribution to the total viral pool may be small (Hennes and Suttle, 1995; Weinbauer and Suttle, 1997; Marie *et al*., 1999; Bettarel *et al*., 2000).

Likewise the diversity of marine ssDNA and RNA viruses is unknown. ssDNA viruses were initially overlooked in diversity studies because their small genomes prevent inclusion in PFGE and FIGE analysis (Wommack *et al*., 1999; Steward *et al*., 2000), whilst their small particle size may have caused them to be lost during concentration procedures (Breitbart, 2012). Fundamentally

however, early metagenomic surveys solely targeted dsDNA viruses (e.g. Breitbart *et al*., 2004). Whilst the application of metagenomics to ssDNA and RNA viruses more recently has provided useful information about genome structure and diversity (Angly *et al*., 2006; Culley et al., 2006; Lang *et al*., 2009; Tucker *et al*., 2011; Labonte and Suttle, 2013; Gustavsen *et al*., 2014), the use of reverse transcription PCR and multiple displacement amplification, with their inherent biases (Culley *et al*., 2010; Kim and Bae, 2011), means the data is non-quantitative. Marine viral ecologists have thus called for a universal pipeline in which viruses of all nucleic acid types can be simultaneously and quantitatively studied, as well as techniques to accurately determine the contribution of each viral type to the total viral pool (Breitbart, 2012; Brum and Sullivan, 2015).

1.3.3 Elucidating who infects whom

Despite the apparent importance of virus-host interactions on bacterial community composition and ecosystem function knowledge in marine systems has been largely bottlenecked by cultivation and technical limitations (Dang and Sullivan, 2014). Only a fraction (<1%) of nature's microbes grow under laboratory conditions (Rappé and Giovannoni, 2003) and few of the 50 known bacterial phyla have cultured phages. The cultured representatives are dominated by *Cyanobacteria*, *Proteobacteria* and *Bacteroidetes* (Deng and Sullivan, 2014). Whilst such model systems are invaluable in testing experimental hypotheses and represent the gold standard for developing a mechanistic understanding of phage-host dynamics, such approaches are likely unable to map the immense network of such interactions in natural systems (Dang and Sullivan, 2014). Consequently a suite of cultivation-independent methods have emerged to study virus-host interactions in nature. These include viral tagging (Deng *et al*., 2012a, 2014), phageFISH (Allers *et al*., 2013), microfluidic digital PCR (Tadmor *et al*., 2011) and viral genome identification through single cell and fosmid genomic sequence mining (Mizuno *et al*., 2013; Rodriguez-Valera *et al*., 2014; Roux *et al*., 2014a; Labonté *et al*., 2015). It will be necessary to utilise these methods, individually or in combination, to unravel the complex nature of phage-host interactions in natural and engineered samples.

1.4 Aims and Objectives

Better microbial ecology could make a contribution to tackling the unprecedented coincidence of challenges the global water industry is currently facing. However, the inability of niche and neutral components to adequately explain shifts in community composition and functional failures suggests that a better understanding of phage ecology in WWTB's is a necessity. Especially since our current understanding is modest and suffers from the same methodical limitations and

shortfalls as that undertaken within marine systems. Therefore the aim of this research is to expand our knowledge in this exciting and potentially important arena by addressing the following objectives:

- Develop a high throughput flow cytometric method to quantify total viral abundance in activated sludge samples, the most important biological wastewater treatment process.
- Monitor total virus abundance in a full scale nitrifying activated sludge plant, describe their dynamics and assess their potential role in temporal bacterial dynamics and plant functions.
- Monitor total virus abundance in replicate, lab scale nitrifying activated sludge reactors, describe their spatiotemporal dynamics and synchronicity and asses their potential role in spatiotemporal bacterial dynamics and plant functions.
- Characterise the community structure and metabolic diversity of viruses throughout the wastewater treatment stream.

CHAPTER 2

FLOW CYTOMETRIC QUANTIFICATION OF VIRUSES IN ACTIVATED SLUDGE

This chapter, in part, was published as:

Brown, M. R., S. Camézuli, R. J. Davenport, E. Petelenz-Kurdziel, L. Øvreås and T. P. Curtis (2015). "Flow cytometric quantification of viruses in activated sludge." Water Research **68**(0): 414-422.

Viruses may play a critical role in the microbial dynamics of activated sludge systems; however the difficulty of their quantification makes long term and large scale studies costly, timely and challenging. Thus a flow cytometric protocol was optimised and employed to determine virus abundance in activated sludge samples. The best flow cytometry signatures and highest virus counts were obtained by separating the indigenous floc-associated viruses using Tween 80 and sodium pyrophosphate, diluting the sample with Tris-EDTA and staining with SYBR Green II. Using the optimised protocol viral concentrations from 25 activated sludge plants were determined, with average concentrations of 2.35×10^9 mL^{-1} observed. Direct counts by transmission electron microscopy were highly correlated with flow cytometric counts (p = < 0.05 and R^2=0.77), with concentrations from both quantification methods comparable at the order of magnitude level. The high counting efficiency, ease of preparation and rapidity and reproducibility of analysis makes flow cytometric quantification of viruses in activated sludge ideal for routine investigation and thus invaluable in unravelling the complexity of phage host interactions in such systems.

2.1. Introduction

Bacteria are an integral part of activated sludge (AS) processes; dozens, perhaps hundreds, of different species play key roles in nutrient removal and the transformation and mineralisation of organic matter (Shapiro and Kushmaro, 2011). Consequently factors controlling bacterial abundance, diversity and activity are central to understanding, developing and predicting the behaviour of such processes. Among these factors, top down control through viral lysis could have an important role. Bacteriophages (viruses that infect bacteria) are the most abundant and diverse biological entities on earth, typically in the order of 10^7 mL^{-1} in most studied ecosystems, and are known to continuously regulate microbial ecology and activity by affecting carbon and nutrient fluxes, food web dynamics and microbial diversity and diversification (Suttle, 2007; Shapiro and Kushmaro, 2011). Whilst viruses, including bacteriophages, are known to be found at high abundance and diversity in AS (10^8 – 10^9 virus like particles (VLP) mL^{-1}), they have proven difficult and time consuming to study (Otawa *et al.*, 2007; Tamaki *et al.*, 2012). Consequently our knowledge and understanding of phage ecology in AS processes, and their potential influence on these globally important processes, is limited.

Traditionally, viruses have been enumerated by culture based methods (Adams, 1959; Havelaar and Hogeboom, 1983; Kott, 1966) or by transmission electron microscopy (TEM) (Torrella and Morita, 1979; Bergh *et al*., 1989). The former is selective for host-specific infectious viruses, thus counts only represent a small fraction of the total population. Whilst the latter, though providing information on phage shape and size, is expensive, time consuming and lacks precision (Weinbauer, 2004). Over the past two decades the introduction of highly sensitive fluorescent nucleic acid-specific dyes (for example SYBR Green I, DAPI, and YOPRO-1) in combination with epifluorescence microscopy (EFM) has significantly improved the detection and quantification of viruses in aquatic ecosystems (Brussaard, 2004; Brussaard *et al*., 2010). EFM is considerably quicker, incurs lower costs and thus allows for a greater throughput of samples compared with TEM. With the introduction of flow cytometric detection and enumeration of free viruses, again in combination with sensitive nucleic acid-specific dyes, the sensitivity of detection, accuracy and precision of quantification and the speed of analysis has further improved. Consequently flow cytometry (FCM) has become the method of choice for quantifying viruses in aquatic samples (Brussaard *et al*., 2010). Despite this, virus abundance in AS has only been determined using TEM or EFM and not FCM, though the literature in this area is still modest (Ewert and Paynter, 1980; Otawa *et al*., 2007; Wu and Liu, 2009).

The aim of this chapter was to critically describe a rapid FCM protocol to enumerate planktonic and floc-associated extracellular viruses in AS, to evaluate the protocol against that of Brussaard *et al*. (2010) and a TEM based approach, and to present virus abundance data from 25 AS plants.

2.2. Materials and Methods

2.2.1. Protocol Optimisation

2.2.1.1. Samples

AS samples were collected from a nitrifying domestic wastewater treatment plant (WWTP) in Tudhoe Mill, Durham, United Kingdom (UK), in March 2013. Samples were collected in polypropylene containers, stored at ~4^{o}C during transit and fixed within 2 hours, as previously reported by Brussaard *et al.* (2004; 2010). Briefly 1 mL aliquots of each sample were transferred to 2 mL cryovials and fixed at a final concentration of 0.5% Glutaraldehyde for 15-30 minutes at 4^{o}C in the dark. After fixation aliquots were flash frozen in liquid nitrogen and stored at -80^{o}C. Samples were thawed at room temperature and mixed via manual shaking for 10 seconds prior to pre-treatment. Once established optimal pre-treatments were used in subsequent experiments.

2.2.1.2. Pre-treatments for Dislodgment of Floc Bound Viruses

2.2.1.2.1. Chemical Treatment

Four dispersants, the surfactants - polyoxyethylene-sorbitan monooleate (Tween 80, Sigma) and Triton X-100 (TX, Sigma), and the ionic dispersants - sodium pyrophosphate (SP, Sigma) and sodium cholate (SC, Sigma), were tested separately and in combination at various concentrations as a sample pre-treatment for virus dislodgment from AS flocs (1 and 5% for Tween 80 and TX, 5 and 10mM for SP and 0.1 and 1% for SC). Thus once thawed the dispersants were added to samples and incubated for 15 minutes in the dark at room temperature. All dispersants, with the exception of Tween 80, were autoclaved prior to use. Each treatment was analysed in triplicate, with a paired control (dispersant free samples) per replicate.

2.2.1.2.2. Physical Treatment

The effect of ultrasound treatment, in combination with chemical treatment, on virus dislodgment was tested using a sonicating water bath (Decon FS200b; 120W; 40 KHz), with 1 mL samples being run for 1, 2, 3, 5 and 8 minutes. Sonication was interrupted for 30 seconds every minute, during which time the samples were shaken manually (Danovaro et al., 2001). Each treatment was analysed in triplicate, with a paired control (samples without sonication) per replicate.

2.2.1.3. Extracellular DNA Interference

In order to eliminate the uncertainties in virus counting due to extracellular DNA (eDNA) a nuclease treatment was tested, since viral nucleic acids will generally be protected from DNase degradation by their protein capsids and sometimes by a lipid envelope (Allander *et al*., 2001; Breitbart and Rohwer, 2005). DNase I (Qiagen, UK), at concentrations of 1500 U μL^{-1} and 1.5 U μL^{-1}, was added to samples and incubated for 15 minutes in the dark at room temperature. Each treatment, in addition to a DNase free sample (control), was analysed in triplicate.

2.2.1.4. Staining Optimisation

SYBR Green I (SG I), SYBR Green II (SG II) and SYBR Gold (SG), used to stain double stranded DNA (dsDNA), single stranded DNA (ssDNA) and RNA and dsDNA, ssDNA and RNA respectively, were tested separately, to count specific virus communities, and in combination, to achieve the greatest total count, at various dilutions (0.5×10^{-4} and 1×10^{-4} of each stains stock solution respectively). To further optimise the staining procedure incubation temperatures of 75, 80 and 85^{o}C were also investigated. All stain combinations and incubation temperatures were analysed in triplicate.

2.2.2. Fluorescent Staining and FCM Analysis

AS samples were diluted with TE-buffer (10 mM Tris-HCl 1 mM EDTA; pH 8.0) to achieve an event rate between 200 and 800 viruses s^{-1} and avoid coincidence (i.e., two or more viruses and/or particles being simultaneously within the sensing zone). To achieve this five 1 mL dilutions (1/500, 1/750, 1/1000, 1/1250 and 1/1500) were prepared per replicate. Diluted samples were then stained using either Brussaard *et al's*.(2010) protocol, 10 µL of 0.02 µm filtered SYBR Green I (0.5×10^{-4} dilution of the commercial stock) for 10 minutes in the dark at 80°C, or variations of this regarding staining optimisation. Sample dilutions were analysed in triplicate using a FACScan flow cytometer (Becton Dickinson, California) equipped with a 15-mW 488-nm air-cooled argon-ion laser and a standard filter setup. The trigger was set on green fluorescence (GFL). Highly diluted and well-mixed yellow-green fluorescent microspheres (FluoSpheres, 1.0 µm diameter; Invitrogen, Molecular Probes; F8823) were added as an internal reference to all samples. Readings were collected in logarithmic mode (at least 5,000 events per sample) and analysed with FlowJo v10.0.7r2 (FlowJo LLC, Oregon). Data was collected using GFL/side scatter (SSC) dot plots and specified gates taken from Brussaard *et al*. (2010), V1, V2 and V3 which correspond to viruses of differing fluorescence intensity (total count = V1+V2+V3). This enabled optimal distinction between stained viruses and other microbial cells and/or background noise, thus filtration to remove such particles wasn't required. Blanks, consisting of TE-buffer and autoclaved 0.2-µm-filtered sample, were pre-treated and analysed identically to samples, further facilitating the correction of virus counts for noise.

2.2.3. Virus Recovery Efficiency

AS samples, collected from Tudhoe Mill WWTP, were seeded with the dsDNA coliphage T4 (NCIMB, UK) and left for 15 minutes prior to fixing. Triplicate samples, with and without the seeded T4 coliphage, were then prepared and analysed following both the optimised protocol and that of Brussaard *et al.* (2010). The seeded T4 abundance was determined by FCM ($0.91 \pm 0.04 \times 10^9$ VLP mL^{-1}) and, for comparison, by plaque assay ($1.0 \pm 0.17 \times 10^9$ Viruses mL^{-1}). Briefly 20 µL of the host isolate *Escherichia coli* was suspended in 8 mL of sterile sloppy agar (0.5% agar in nutrient broth medium) together with 20 µL of filter-fertilised (0.2 µm) T4 coliphage culture. The sloppy agar was then poured over a pre-warmed (37°C) nutrient agar plate and incubated for 2 days at 37°C. Plates were checked after 24 and 48 hours for plaque formation. The FCM seeded concentration of 0.91×10^9 was used for calculations.

2.2.4. Virus abundance at a suite of AS WWTP's

AS samples were collected from 25 domestic WWTP's situated within the North East of England, UK, in April and May 2013 (see Table 2.1 for plant configurations/characteristics). Triplicate samples were collected, fixed and then analysed using the optimised protocol. The mixed liquor (volatile) suspended solids (MLSS/MLVSS) were determined according to Standard Methods (APHA, 1989).

Table 2. 1. WWTP details, AS process configurations and sample dates.

WWTP	Process Configuration	Aeration	Wastewater Type	TEP*	Sampling Date
Amble	SBR	Fine bubble	Municipal	16607	02.05.13
Aycliffe	Conventional	Jet	Municipal/Industrial	61106	04.04.13
Berwick	Conventional	Surface	Municipal	15537	21.05.13
Billingham	Conventional	Fine bubble	Municipal	35293	15.05.13
Blyth	SBR	Fine bubble	Municipal	37859	13.05.13
Bowsden	Oxidation Ditch	Surface	Municipal	250	21.05.13
Bran Sands	Conventional	Jet	Municipal/Industrial	391142	20.05.13
Branxton	Oxidation Ditch	Surface	Municipal	250	21.05.13
Broomhaugh	Oxidation Ditch	Surface	Municipal	7095	11.04.13
Browney	Conventional	Jet	Municipal	21586	24.05.13
Cambois	Conventional	Fine bubble	Municipal	28655	13.05.13
Cramlington	Conventional	Surface	Municipal/Industrial	45309	05.06.13
Haggerston	Oxidation Ditch	Medium bubble	Municipal	2040	21.05.13
Hendon	Conventional	Fine bubble	Municipal	229108	09.04.13
Hexham	Conventional	Surface	Municipal	29714	11.04.13
Hordon	Conventional	Fine bubble	Municipal	100299	09.04.13
Howdon	Oxidation Ditch	Fine bubble	Municipal/Industrial	947811	13.04.13
Marske	SBR	Fine bubble	Municipal	93556	09.04.13
Newbiggin	Conventional	Fine bubble	Municipal	38487	13.05.13
Seaham	Conventional	Fine bubble	Municipal	23595	15.05.13
Seahouses	SBR	Jet	Municipal	11213	02.05.13
Seaton Carew	Conventional	Fine bubble	Municipal	120222	09.04.13
Sedgeletch	Conventional	Fine bubble	Municipal	51152	04.04.13
Tudhoe Mill	Conventional	Fine bubble	Municipal	22493	30.04.13
Washington	Oxidation Ditch	Surface	Municipal/Industrial	74916	09.05.13

*Total equivalent population (TEP) served by the plant. SBR: Sequencing batch reactor.

2.2.5. Comparison of FCM and TEM counts

FCM AS viral counts obtained from 7 of the WWTP's were compared with TEM counts. For TEM analyses the preconcentration procedure (i.e. ultracentrifugation) typically used was omitted, since the number of viruses in AS was expected to be very high (Otawa *et al*., 2007; Wu and Liu, 2009). 1 mL of pre-treated sample was diluted with 1 mL of deionised water, mixed and then 2 µL was spotted onto a 200 mesh Formvar coated copper grid and air dried at room temperature. Unrinsed grids were negatively stained with 2% uranyl acetate for 1 minute. Excess stain was drained off with a pointed piece of glass fibre filter paper and grids were then left to dry at room temperature for 24 hours. Observations were made using a Philips CM 100 compustage transmission electron microscope, operating at 100 kV. Duplicate grids were prepared for each

sample, with 30 fields of view (FOV), determined as a sufficient sample size as described by Davenport and Curtis (2004), examined per grid at a magnification of 13,500.

2.2.6. Statistical Analyses

All statistical analysis was undertaken in RStudio (v. 1.0.143, R Core Team, 2017) using R version 3.4.0 (R Core Team, 2017). The Shapiro-Wilk Test ($p > 0.05$, shapiro.test, "stats" v. 3.4.0, R Core Team, 2017) and the Bartlett Test ($p > 0.05$, bartlett.test, "stats" v. 3.4.0, R Core Team, 2017) were utilised to test normality and homogeneity of variance respectively, unless stated otherwise.

2.2.6.1. Protocol Optimisation

Virus abundance after each treatment was compared and analysed for significant differences using ANOVA with Tukey's pairwise comparisons (TukeyHSD, "stats" v. 3.4.0, R Core Team, 2017). Prior to analysis data was checked for normality and homogeneity of variance as described in 2.2.10.

2.2.6.2. Agreement between FCM and TEM Enumeration

Agreement between the two methods was assessed using linear regression and Bland Altman analysis using the functions .lm and bland.altman.stats in packages "stats" ("stats" v. 3.4.0, R Core Team, 2017) and "BlandAltmanLeh" (v. 0.3.1, Lehnert, 2015) respectively. Linear models were checked visually for homoscedasticity, linearity and residual autocorrelation, whilst model residuals were checked for normality as in 2.2.10 (Norman and Streiner, 2008; Zuur *et al*., 2010; Ghasemi and Zahediasl, 2012). For Bland Altman analysis calculated differences between FCM and TEM virus abundances were checked for normality as described in 2.2.10 so that 95% confidence intervals (CI's) could be estimated.

2.3. Results

2.3.1. Optimisation of Protocol for AS Virus Enumeration by FCM

AS samples incubated with dispersants displayed higher virus counts than untreated samples (Fig. 2.1 A). The most effective treatment, based on the largest increase in virus abundance from its paired control, was Tween 80 (5%) and SP (10mM); $1.52 \pm 0.19 \times 10^9$ VLP mL^{-1} treated and $1.15 \pm 0.12 \times 10^9$ VLP mL^{-1} control (ANOVA: $P < 0.05$).

Sonication had no statistically significant effect in four pairwise comparisons (1, 2, 3 and 5 minutes) with unsonicated samples (ANOVA: $P > 0.05$), whilst 8 minutes had a non-significant negative effect on virus counts (ANOVA: $P > 0.05$, Fig. 2.1 B).

Virus counts obtained from DNase treated samples gave contrasting results (Fig. 2.2 A). Samples treated with 1500U μL^{-1} gave significantly lower counts than those in untreated samples (1.31 ± 0.14 × 10^9 and 1.95 ± 0.05 × 10^9 VLP mL^{-1} respectively, ANOVA: $P < 0.05$), a percentage decrease of 32.7%, whilst samples treated with 1.5U μL^{-1} showed no significant difference to those in untreated samples (1.92 ± 0.1 × 10^9 and 1.95 ± 0.05 × 10^9 VLP mL^{-1} respectively, ANOVA: $P > 0.05$, Fig. 2.2 A).

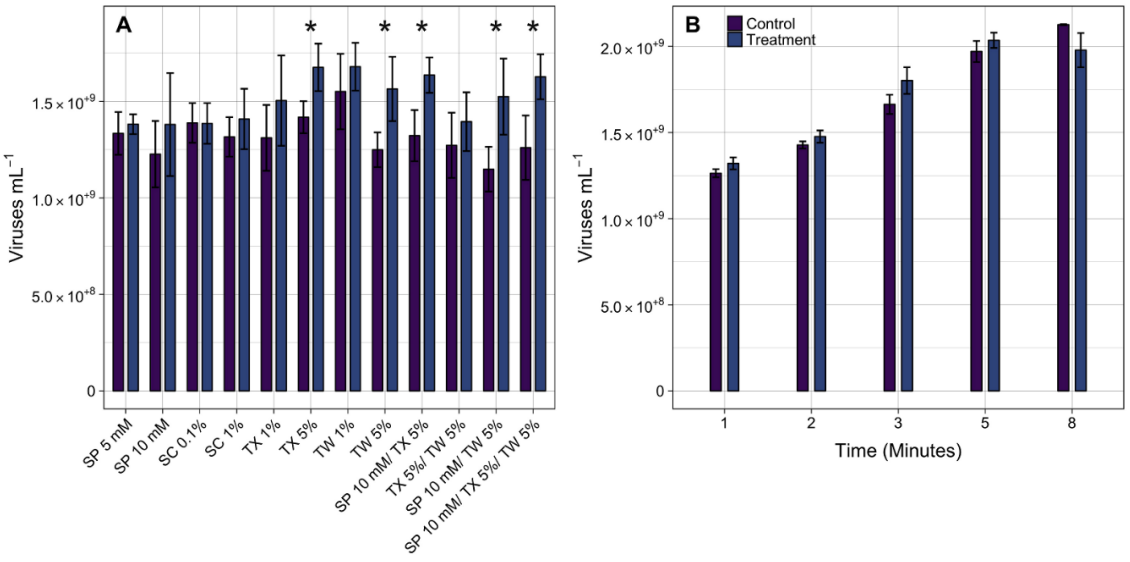

Figure 2. 1. Effect of dispersants (**A**) and sonication time (**B**) on FCM virus abundance. Main bars indicate mean virus abundance across triplicates, whilst error bars indicate standard deviation across triplicates. * Significantly different from controls at the 0.05 level.

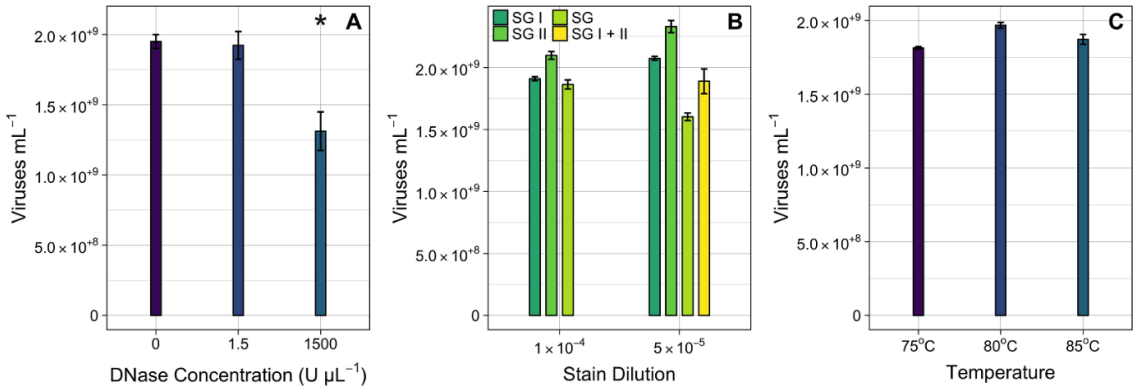

Figure 2. 2. Effect of DNase treatment (**A**), stain type and dilution (**B**) and incubation temperature (**C**) on FCM virus abundance. Main bars indicate mean virus abundance across triplicates, whilst error bars indicate standard deviation across triplicates. * Significantly different from other treatments at the 0.05 level.

The highest virus count was achieved using SG II at a dilution of 0.5 × 10^{-4} (2.3 ± 0.05 × 10^9 VLP mL^{-1}, Fig. 2.2 B), although counts were not significantly higher than those obtained using SG I, SG or SG I + II (ANOVA: $P > 0.05$). No large difference in GFL or SSC single was detected between the three stains, thus distinguishing between dsDNA, ssDNA or RNA viruses was not possible (Fig. 2.3). The original incubation temperature of 80°C gave the highest counts (1.97 ± 0.02 × 10^9 VLP mL^{-1}),

they were not however significantly greater than those obtained at 75 and 85°C (1.82 ± 0.01 × 10^9 and 1.87 ± 0.03 × 10^9 VLP mL^{-1} respectively, ANOVA: $P > 0.05$, Fig. 2.2 C).

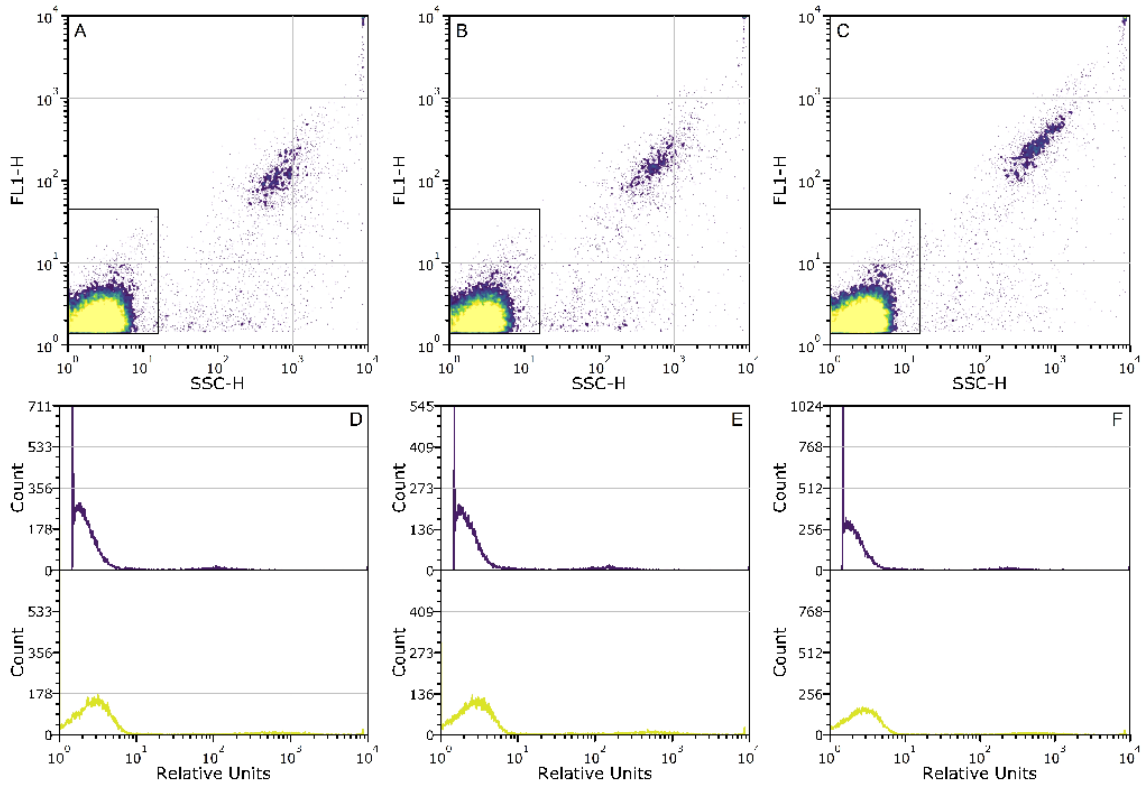

Figure 2. 3. FCM density plots (**A** – **C**) and histograms (**D** – **F**) of AS samples taken from Tudhoe Mill WWTP stained with SG I (**A** and **D**), SG (**B** and **E**) and SG II (**C** and **F**), all at a 0.5 × 10-4 dilution of commercial stock. All events plotted, sample dilution 1:1000, purple and yellow lines (**D** – **F**) are GFL (FL1-H) and SSC (SSC-H) signals respectively. Total virus gate (**A** – **C**) taken from Brussaard *et al*. (2010).

2.3.2. Virus Recovery and Enumeration Efficiency

The efficiency of virus detachment and staining for both protocols was tested by estimating the recovery of the T4 virus from seeded samples, as well as total virus recovery. The recovery efficiency of the seeded T4 coliphage varied between the two protocols, with the optimised protocol presented here (Fig. 2.4) recovering 102 ± 2.7% compared to that of Brussaard *et al*. (2010), which recovered 85.4 ± 2.1% (0.93 ± 0.02 × 10^9 mL^{-1} and 0.78 ± 0.02 × 10^9 mL^{-1} of the seeded 0.91 × 10^9 VLP mL^{-1} respectively). Total virus recovery also varied, the optimised protocol recovered 1.07 ± 0.03 × 10^9 VLP mL^{-1} compared with 0.87 ± 0.02 × 10^9 VLP mL^{-1} recovered by that of Brussaard *et al*. (2010), an increase of 22.9%.

2.3.3. Virus Abundance in Full Scale Activated Sludge WWTP's

Virus abundance in 25 AS plants ranged from 0.59 ± 0.04 × 10^9 VLP mL^{-1} (Bowsden) to 5.14 ± 0.37 × 10^9 VLP mL^{-1} (Howdon), with a mean concentration of 2.35 × 10^9 VLP mL^{-1} (Table 2.2). The concentration of viruses per gram (dry) of MLSS ranged from 2.64 ± 0.10 × 10^{11} (Brand Sands) to 28.11 ± 3.15 × 10^{11} (Washington), with a mean concentration of 9.59 × 10^{11}. 93.8% ± 2.4% of

viruses found across all plants were those associated with the V1 subpopulation, with the V2 and V3 subpopulations making up 6.3 ± 2.5% and 0.3 ± 0.1% respectively (Fig. 2.5). No clear relationship was apparent between MLSS and virus concentrations.

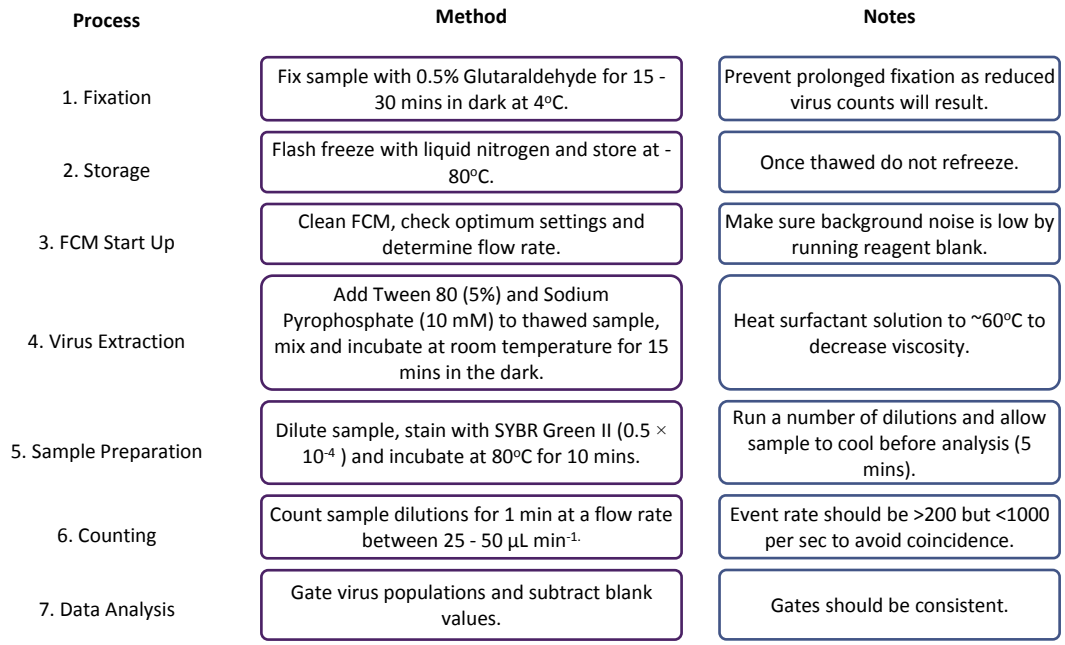

Figure 2. 4. Diagram of the optimised protocol based on our findings, including processes, methodology and critical notes.

Table 2. 2. Concentration of viruses from 25 activated sludge plants in the North East of England, UK.

WWTP	Virus concentration[a] (10^9 mL^{-1})	MLSS (g L^{-1})	Virus concentration[b] (10^{11} g $^{-1}$)
Amble	3.25 ± 0.16	3.1	10.48 ± 0.51
Aycliffe	1.81 ± 0.15	1.9	9.51 ± 0.79
Berwick	1.21 ± 0.09	2.09	5.78 ± 0.43
Billingham	1.89 ± 0.10	1.36	13.92 ± 0.73
Blyth	3.40 ± 0.18	4.58	7.43 ± 0.39
Bowsden	0.59 ± 0.04	2.13	2.76 ± 0.21
Bran Sands	2.96 ± 0.11	11.21	2.64 ± 0.10
Branxton	1.05 ± 0.07	2.02	5.21 ± 0.34
Broomhaugh	1.35 ± 0.15	4.01	3.38 ± 0.39
Browney	0.72 ± 0.08	1.87	3.83 ± 0.41
Cambois	2.23 ± 0.07	2.88	7.75 ± 0.23
Cramlington	3.54 ± 0.23	1.38	25.65 ± 1.66
Haggerston	1.23 ± 0.05	2.86	4.30 ± 0.19
Hendon	3.25 ± 0.13	3.08	10.55 ± 0.44
Hexham	2.51 ± 0.13	2.77	9.05 ± 0.45
Hordon	2.23 ± 0.21	2.16	10.33 ± 0.97
Howdon	5.14 ± 0.37	2.19	23.46 ± 1.71
Marske	3.60 ± 0.21	2.94	12.25 ± 0.71
Newbiggin	2.88 ± 0.34	4.23	6.82 ± 0.79
Seaham	2.41 ± 0.19	3.22	7.49 ± 0.59
Seahouses	1.00 ± 0.04	1.54	6.54 ± 0.25
Seaton Carew	2.65 ± 0.11	3.1	8.56 ± 0.35
Sedgeletch	1.12 ± 0.06	3.04	3.68 ± 0.21
Tudhoe Mill	2.70 ± 0.47	2.64	10.21 ± 1.79
Washington	3.98 ± 0.45	1.415	28.11 ± 3.15

[a] concentrations determined by FCM using the optimised protocol, ± denotes standard deviation between triplicate samples. [b] values calculated from virus concentration per millilitre and the MLSS data.

2.3.4. FCM vs. TEM

The best-fit linear regression model of FCM and TEM viral counts yielded a highly significant, positive correlation coefficient and high coefficient of determination ($P < 0.05$, $R^2 = 0.76$, Fig. 2.6 A). The 95% CI's for the slope coefficient (β), which was also significant ($P < 0.05$), and the intercept (α) additionally include 1 and 0 respectively, implying virus abundances obtained by both methods were comparable. Nonetheless FCM estimates were 2.7 times higher, on average, than those given by TEM, with Bland Altman analysis identifying a bias estimate, mean difference between FCM and TEM viral counts, of 1.70×10^9 VLP mL^{-1} (Fig. 2.6 B). Differences between the two methods were however consistent across the measurement range (90% of the mean of FCM and TEM counts, Fig. 2.6 C), despite analysis of raw counts suggesting the contrary (Fig. 2.6 B). Thus FCM does appear a suitable method for viral enumeration in AS, particularly when one considers the average coefficient of variation between replicates was 6.67 for FCM and 17.25 for TEM.

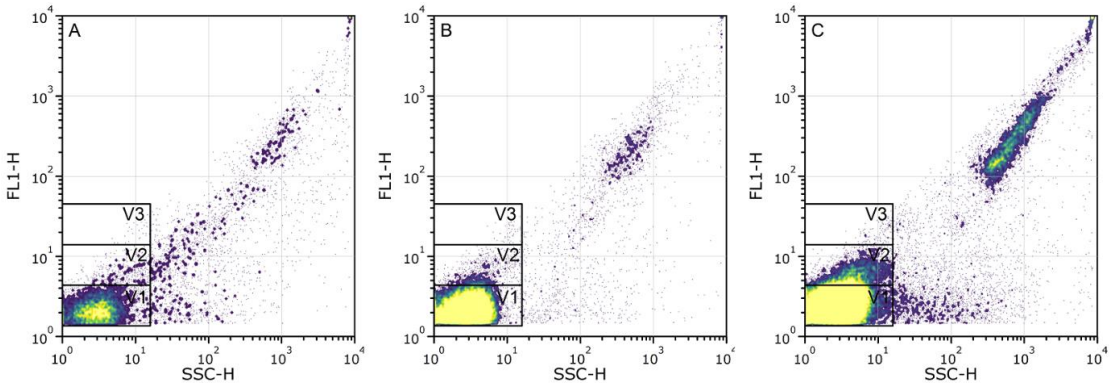

Figure 2. 5. Flow cytometry density plots of AS samples taken from Bowsden (**A**), Sea Houses (**B**) and Howdon (**C**) WWTP's following pre-treatment and processing as described in the optimised protocol All events plotted, sample dilution 1:1000. V1, V2 and V3 gates taken from Brussaard *et al.* (2010).

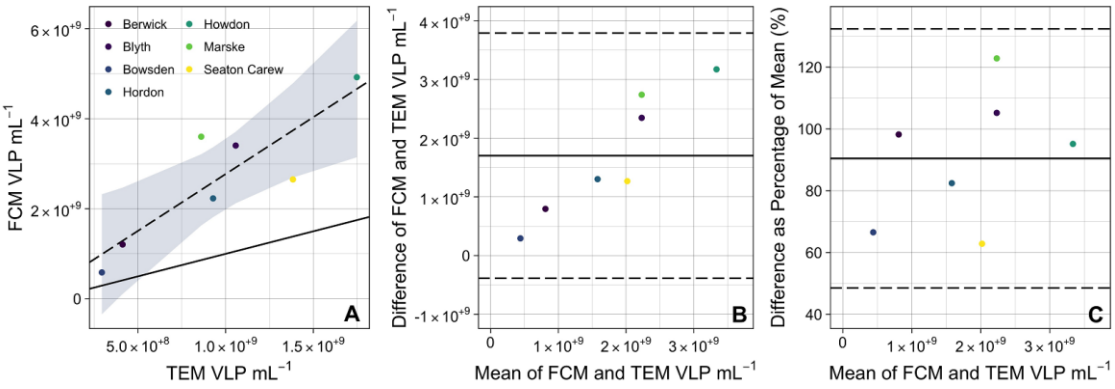

Figure 2. 6. Statistical comparisons of FCM and TEM viral counts. FCM Vs TEM viral counts (**A**), solid line represents the theoretical slope of a 1:1 relationship and the dashed line, and associated shaded area, represent the best fit regression line and 95% CI's respectively (n = 7, Shapiro-Wilk Test P = 0.98, $R^2 = 0.76$, $\alpha = 2.49 \times 10^8$ (CI = $-1.50 - 2.00 \times 10^9$) and $\beta = 2.52$* (CI = -0.88 – 4.16)). Bland Altman plot of mean FCM and TEM viral counts plotted against their differences (**B**) and their differences as a percentage of the mean abundance (**C**), solid and dashed black lines represent the mean difference and associated 95% CI's respectively (n = 7, Shapiro-Wilk Test P = 0.68 and 0.81 respectively). * Significant at the 0.05 level.

2.4. Discussion

We present a rapid, accurate and sensitive flow cytometric method specifically optimised for enumerating total planktonic and floc-associated extracellular viruses in AS. This constitutes an improvement in the study of AS viral communities because FCM is faster and less dependent on the operator than EFM and TEM. The performance of FCM virus quantification, is however strongly affected by AS sample pre-treatment, optimisation of the staining procedure and the presence of false positives, i.e. the staining of DNA associated with membrane-derived vesicles (MVs), gene transfer agents (GTAs) and eDNA (Forterre *et al*., 2013).

The very different effects of sample pre-treatment emphasises the importance of selecting appropriate techniques to enable accurate virus quantification in AS samples. Dispersants have previously been used successfully as an eluent for dislodging viruses from sludge (Wu and Liu, 2009) and marine (Danovaro *et al*., 2001; Danavaro and Middelboe, 2010) and freshwater (Duhamel and Jacquet, 2006) sediments. This survey confirms these findings: the addition of SP (10mM) in combination with Tween 80 (5%) producing the highest and most accurate counts. Sonication has also been used to dislodge viruses from marine (Danovaro *et al*., 2001; Danavaro and Middelboe, 2010) and freshwater sediments (Duhamel and Jacquet, 2006), soils (Williamson *et al*., 2003) and anaerobic digester sludge (Wu and Liu, 2009), with optimum sonication times of 30 seconds (Otawa *et al*., 2007) and 1 minute (Wu and Liu, 2009) reported for AS samples. No significant effect was observed in this study. Possible explanations could be the more powerful equipment (120W compared to 10W and 100W), smaller sample size (1 mL compared to 10 mL and 50 mL) or greater dislodgment efficiency of the chemical pre-treatment used in this study. Disruption of viral protein capsids or lipid envelopes is thought to occur with enhanced sonication times (Wu and Liu, 2009) and may have lowered counts in this study.

FCM enumeration of free viruses requires working close to the limits of staining methodology and the detection limit of a flow cytometer, thus the intensity of the GFL and/or SSC signal is crucial for optimising such protocols. SG I is commonly used for counting pelagic marine viruses (Marie *et al*., 1999; Brussaard, 2004) by FCM, however in some instances SG and SG II have provided increased and more reproducible counts (Chen *et al*., 2001; Duhamel and Jacquet, 2006; Tomaru and Nagasaki, 2007). Our results suggest SG II at a dilution of 0.5×10^{-4} provides the most accurate enumeration of total free viruses in AS. SG II has a strong affinity to RNA and thus a greater ability, when compared to SG I and SG, to stain small genome sized RNA viruses, which could explain the small increase in counts recorded. However the total counts and GFL/SSC signals observed would imply that all three dyes have a very similar ability to stain dsDNA, ssDNA and RNA viruses

respectively, a finding also reported by Brussaard *et al*. (2000) and Brussaard (2004). Consequently distinguishing between these virus populations is not possible with the method presented, as it is apparent that the total count obtained encompasses all three.

Another important factor when trying to increase GFL is the incubation temperature, as heat treatment affects the permeability of the viral capsid and denatures the nucleic acid, thereby improving staining efficiency (Brussaard, 2004). An incubation temperature of 80°C is most commonly used for FCM enumeration of pelagic marine viruses (Marie *et al*., 1999; Brussaard, 2004), however incubation at room temperature and 75°C has been shown to provide increased and more reproducible counts in marine samples (Tomaru and Nagasaki, 2007) and freshwater sediments (Duhamel and Jacquet, 2006). Our results suggest an incubation temperature of 80°C provide the most accurate enumeration of AS viruses.

The significant (P = < 0.05) linear relationship between FCM and TEM counts, with its non-significant intercept, and the consistent bias estimate obtained by Bland Altman analysis suggests that these two methods are evaluating the same virus particles. However the FCM counts were typically 2.7 times higher than corresponding TEM values. A direct comparison of FCM and TEM has never previously been undertaken. However direct comparisons of EFM and TEM for marine and freshwater environments suggest a similar discrepancy between the fluorescent and direct counts with Hennes and Suttle (1995), Weinbauer and Suttle (1997) and Noble and Fuhrman (1998) reporting differentials of 2.3, 1.5 and 1.3 respectively.

Discrepancies could result from the presence of false positives, eDNA, GTAs and MVs, causing FCM to overestimate virus abundance, a growing concern in natural environments (Forterre *et al*., 2013). Treatment with DNase has previously been used to eliminate or reduce such an outcome, although no significant difference in EFM virus counts was observed by Otawa *et al*. (2007) and Wu and Liu (2009) between treated and untreated AS samples. Our results gave contrasting results, with the more concentrated DNase samples showing a significant decline in virus counts and the less concentrated samples showing little affect. The sensitivity of viruses to DNase has been demonstrated previously (Jiang and Paul, 1995; Bettarel *et al*., 2000), it is probable that at the higher DNase concentrations true viruses were degraded and thus counts reduced. MVs produced by Proteobacteria, which dominate AS communities (Wagner *et al*., 2002), and some hyperthermophilic archaea, as well eDNA adsorbed to cell debris or mineral surfaces, are also known to produce false EFM and FCM positives even after DNase treatment (Nielsen *et al*., 2007; Soler *et al*., 2008; Zhao *et al*., 2010). As there are no good methods to rapidly discriminate between viruses, GTA'S, MV'S and eDNA and DNase is ineffective in their removal and can degrade

the viruses of interest, it is recommended that an FCM (or EFM) count be controlled for the presence or absence of such false positives by a TEM count in a selection of samples, as done in this study.

Discrepancies might also be caused by particulate matter and detritus obscuring the virus particles during TEM counts (Hennes and Suttle, 1995; Betteral *et al*., 2000). It was difficult in this study to find areas of the TEM grids devoid of such particles, though sufficient clear FOV were found to make an accurate count. The underestimation of values and significant greater variability of the TEM method in comparison to FCM may also be explained by the high magnifications used and the potential loss of viruses during the staining procedure (Betteral *et al*., 2000).

The abundance of viruses in AS, as determined from 25 WWTP's, is of the order $10^8 – 10^9$ mL^{-1}, similar to results reported by Otawa *et al*. (2007) and Wu and Liu (2009) (2.35×10^9 mL^{-1} compared with 1.1×10^9 and 1.19×10^9 mL^{-1} respectively). The concentration of viruses per gram (dry) of MLSS was also within the same order of magnitude, $10^{11} – 10^{12}$ g^{-1}, across all three studies. The majority of viruses found were those associated with the low and medium fluorescence intensity V1 and V2 virus subpopulations, thought to be bacteriophages of the smallest class (30 – 60 nm in size) (Marie *et al*., 1999; Brussaard *et al*., 2010). Whilst V1 viruses are thought to be smaller in size than V2 viruses, true size estimates are not viable since the GFL and SSC signals are not related to genome size or virus size or shape (Marie *et al*., 1999; Brussaard, 2000).

The concentration of viruses in AS is thus amongst the highest of all systems studied to date. In marine environments concentrations range from 10^4 and 10^8 mL^{-1} (Wommack and Colwell, 2000), in freshwater ecosystems the highest virus abundance to date is 9.6×10^9 mL^{-1} (Hennes and Suttle, 1995) whilst in marine and freshwater sediments virus abundance ranges from $0.03 – 11.7 \times 10^9$ g^{-1} (Danovaro *et al*., 2002). In such environments viruses are proposed to continuously regulate microbial activity and ecology, including carbon and nutrient fluxes, food web dynamics and microbial diversity and diversification (Weinbauer, 2004). Given the high concentrations found in this study and the apparent dominance of bacteriophages it is speculated that viruses are active and dynamic in AS processes and could, in theory, influence microbial activity and ecology, thus directly affecting system performance and functional stability. The availability of a rapid quantification method will facilitate in testing this hypothesis.

2.5. Conclusions

- The results show that the optimised protocol presented is an accurate and highly reproducible method for enumerating total free viruses in AS and thus is ideal for routine investigation.
- FCM counts were highly correlated with TEM based counts and results were comparable to previously published EFM counts.
- The major advantage of FCM over TEM and EFM is its high throughput, removing a key obstacle to undertaking detailed spatial and temporal studies of virus dynamics in AS systems. Such studies are a fundamental prerequisite to understanding their possible impact on a systems bacterial population and thus performance and functional stability.

2.6. Acknowledgments

This work was carried out as part of a Frontiers Engineering research project (NUFEB, http://research.ncl.ac.uk/nufeb/) awarded and funded by the Engineering and Physical Sciences Research Council (EP/H012133/1). I'd like to thank Northumbrian Water who permitted collection of wastewater samples and provided plant operational data. Also Tracey Davey and Kathryn White from the electron microscopy research unit at Newcastle University who assisted with electron microscopy procedures. Finally Stella Camézuli who assisted with laboratory procedures and Elzbieta Petelenz-Kurdziel and Lise Øvreås of Bergen University who facilitated an exploratory investigation into whether AS samples could be run on a FCM.

CHAPTER 3

COUPLED VIRUS-BACTERIA INTERACTIONS AND ECOSYSTEM FUNCTION IN AN ENGINEERED MICROBIAL SYSTEM

Viruses in natural environments are thought to control bacterial abundance, affect community composition and influence ecosystem function. Yet their dynamics have seldom been measured or modelled in engineered systems, where loss of function (plant failures) is common and unpredictable, or in any system for substantial periods at functionally relevant time scales. Thus virus abundance, in conjunction with total and ammonia oxidising bacterial abundances, bacterial community profiles, and a suite of environmental and operational parameters, was monitored weekly for two years in a full-scale activated sludge plant. Mixed liquor virus abundance fluctuated over an order of magnitude ($3.18 \times 10^8 – 3.41 \times 10^9$ virus's mL^{-1}) and was shown statistically to interact with both total and ammonia oxidising bacterial abundance, influence or be influenced by community composition, and negatively affect ecosystem function (effluent concentrations of COD and NH_4^+- N). This suggests viruses play a more central role in the dynamics of activated sludge systems than hitherto realised and might be considered one of the key factors regulating bacterial abundance, community structure and functional stability. The quantitative association of viruses with physically credible abiotic factors gives credence to these findings but also emphasises the role that exogenous factors play in virus dynamics.

3.1. Introduction

Viruses are agents of mortality (Wommack and Colwell, 2000), nutrient regeneration (Middelboe and Jørgensen, 2006; Haaber and Middleboe, 2009; Shelford *et al*., 2012) and horizontal gene transfer (Lindell *et al*., 2004; Sullivan *et al*., 2006) and therefore key drivers of bacterial abundance, activity and community composition in natural environments, as well ecosystem function (Rodriguez-Valera *et al*., 2009; Winter *et al*., 2010; Breitbart, 2012; Liu *et al*., 2015). Yet their dynamics have seldom been monitored in activated sludge; an engineered microbial ecosystem utilised globally to degrade oxygen-depleting organics, transform toxic substances and remove nutrients from wastewaters, where loss of function is observed frequently, unpredictably and often inexplicably (Curtis and Sloan, 2006).

Recent advances in molecular methods and the adoption of ecological approaches to engineered systems have shed some light on the complex mechanisms driving the microbial communities in, and the performance and functional stability of, activated sludge systems. Deterministic (e.g. reactor design, process configuration, operational and environmental conditions) and stochastic processes (e.g. microbial birth, death, immigration and speciation), alone and in combination, are two such classes of mechanism (van der Gast *et al*., 2008; Wells *et al*., 2009, 2011; Ayarza *et al*., 2010; Ofiteru *et al*., 2010; Ayarza and Erijman, 2011; Valentin-Vargas *et al*., 2012). Both, however

only partially explain the variation seen in the microbial composition and performance of such systems (Ofiteru *et al*., 2010).

Given its role in natural environments viral infection could be another important mechanism, especially considering its implication in host abundance fluctuations and functional instability (Lee *et al*., 2007; Barr *et al*., 2010; Motlagh *et al*., 2015), evidence of predator-prey type oscillations (Lee *et al*., 2007; Shapiro *et al*., 2010) and the sheer abundance of viruses (Otawa *et al*., 2007; Wu and Liu, 2009; Brown *et al*., 2015 (Chapter 2)) within activated sludge systems. To date however, viral dynamics have been largely overlooked.

The absence of available methods to link viruses to their host's (Brum *et al*., 2015; Dang and Sullivan, 2014; Brum and Sullivan, 2015) is undoubtedly a contributing factor. However, even total abundance methods, which sparked a transformation in virus ecology in the 1980s (Bergh *et al*., 1989) and 1990's (Hara *et al*., 1991; Marie *et al*., 1999), have only found limited application in activated sludge systems: being used to compare different activated sludge plants (Wu and Liu, 2009; Brown *et al*., 2015 (Chapter 2)), or across very modest time scales (Otawa *et al*., 2007). This is in contrast to the numerous studies that have yielded great insights into and underpin our understanding of virus ecology in marine (e.g. Jiang and Paul, 1994; Weinbauer *et al*., 1995; Bratbak *et al*., 1996; Li and Dickie, 2001) and freshwater (e.g. Hennes and Simon, 1995; Hofer and Sommaruga, 2001; Bettarel *et al*., 2004; Jacquet *et al*., 2005) environments. Although even here the spatial and temporal scale of such studies are typically modest; a consequence of limited "ship time" and other logistical constraints. Thus multiyear studies at functionally relevant temporal scales, and which incorporate time varying exogenous factors, have recently been called for in viral ecology (Breitbart, 2012; Brum and Sullivan, 2015; Wigington *et al*., 2016).

To this end we utilised a recently adapted flow cytometric method (Brown *et al*., 2015 (Chapter 2)) to measure virus abundance weekly for two years in a full scale nitrifying activated sludge plant. Such sampling frequency and longevity was considered foremost logistically feasible, but also biologically and functionally relevant considering nitrifying activated sludge systems maintain solid retention times (SRT) > 7 days (the average period of time biomass remains within a system, Tchobanoglous et al., 2003). By doing so the relationship between virus abundance and the dynamics of total and ammonia oxidising bacteria (AOB), as well as community structure, was evaluated, whilst the influence of exogenous factors (environmental and operational parameters) was also determined. Thus factors affecting virus proliferation and their potential role in the microbial ecology, and potentially the performance (removal of COD and NH_4^+- N) and functional stability, of activated sludge systems were elucidated.

3.2. Materials and Methods

3.2.1. Sample Collection

Mixed liquor (ML) grab samples were collected from the aeration basin (3600 m^3) of a conventional nitrifying domestic wastewater treatment plant (WWTP, 6751 m^3 day^{-1}), situated in the North East of England, United Kingdom, on a weekly basis for a period of two years from June 2011 to May 2013 (104 weeks). Samples were collected in 50 mL polypropylene containers and transported to the lab on ice for immediate processing. Concurrent primary settled sewage (influent) and effluent samples were also collected, in addition to a number of operational variables.

3.2.2. Analytical Methods

For all samples (influent, ML and effluent) suspended/volatile suspended solids (SS/VSS), soluble chemical oxygen demand (COD_s) and soluble ammonium (NH_4^+-N) were determined according to Standard Methods (APHA, 1998) and using Merck COD and NH_4^+-N test kits (VWR, UK) respectively. Anion concentrations, including nitrate, nitrite, sulphate and phosphate, were determined using high performance Ion Chromatography (Dionex ICS-1000 with AS40 auto sampler); samples were filtered through a 0.2µm polyethersulfone membrane prior to analysis. Influent trace metals, including cadmium, zinc, lead and copper, were measured by inductively coupled plasma optical emission spectroscopy (ICP-OES) (Vista MPX axial ICP-OES, Varian, UK), as described by Martin *et al*. (1994). Samples were acidified on collection to pH < 2, digested and then filtered through a 0.45µm polyethersulfone membrane prior to analysis. Finally temperature, dissolved oxygen (DO) and pH within the aeration basin were measured in real time using *in situ* probes, influent flow rate was determined using a Parshall flume, and sludge age was obtained from plant operators.

3.2.3. Molecular Methods

3.2.3.1. Flow Cytometry

For virus enumeration 1 mL sub-samples of influent, ML and effluent were taken, transferred into 2 mL cryovials and fixed at a final concentration of 0.5% glutaraldehyde for 15-30 minutes at 4°C in the dark. Samples were then flash frozen in liquid nitrogen and stored at -80 °C. After defrosting, samples were pre-treated and analysed in triplicate as described by Brown *et al*. (2015 (Chapter 2)) using a FACScan flow cytometer (Becton Dickinson, USA) equipped with a 15-mW 488-nm air-cooled argon-ion laser and a standard filter setup.

3.2.3.2. DNA Extraction

DNA was extracted from 250 µL of ML and from 15 mL of influent, the latter being centrifuged at 3392 × *g* for 15 minutes and the supernatant removed down to a working volume of 250 µL. Cell wall disruption was then carried out using the FastDNA SPIN Kit for soil (MP Biomedicals, USA), thus 244.5 µL of sodium phosphate buffer and 30.5 µL of MT buffer was added to samples and the mixture transferred to Lysing Matrix E tubes. Samples were then lysed at 6.5 ms^{-1} for 30 seconds in a FastPrep instrument (MP Biomedicals, USA) and centrifuged at 14000 × *g* for 15 minutes. DNA from 250 µL of the supernatant was then purified using a MagNA Pure LC 2.0 (Roche, UK) and the MagNA Pure LC DNA Isolation Kit III.

3.2.3.3. Illumina Sequencing

Sample preparation for Illumina sequencing generally followed the protocol of Caporaso *et al*. (2012). Thus the V4 region of the bacterial 16S rRNA gene was amplified using primers 515F (aatgatacggcgaccaccgagatctacactatggtattgtGTGCCAGCMGCCGCGGTAA; adapter, primer pad and primer linker in small letters; and specific primer sequence in capital letters) and 806R (caagcagaagacggcatacgagat**barcode**agtcagtcagccGGACTACHVGGGTWTCTAAT), the latter was barcoded with a 12-base error-correcting Golay code to facilitate sample multiplexing. Each sample was amplified in duplicate, pooled and then cleaned using a MinElute 96 UF Purification Kit as per the manufacturer's instructions (Qiagen Ltd., West Sussex, UK). PCR reactions consisted of: 2 µl DNA extract, 20 µl 5 Prime Hot Master Mix (VWR, Lutterworth, UK), 1 µl each of forward and reverse primer (10 µM final concentration), and 26 µl molecular-grade water. Reactions were denatured at 94 °C for 3 minutes, with amplification proceeding for 35 cycles at 94 °C for 45 seconds (further denaturing), 50 °C for 60 seconds (annealing) and 72 °C for 90 seconds (extension); a final extension at 72 °C for 10 minutes was added to ensure complete amplification. The concentration and purity of pooled amplicons was assessed using a Quant-iT PicoGreen dsDNA assay kit (Life Technologies, Paisley, UK). A composite sample for sequencing was created by combining all samples in equimolar amounts. The composite sample was cleaned twice using Agencourt AMPure XP beads (Beckman Coulter, High Wycombe, UK) according to the manufacturer's instructions, with fragment selection undertaken using E-Gel 2% agarose gels (Life Technologies, Paisley, UK). The size-selected fragment was then purified using the QIAquick PCR purification kit (Qiagen Ltd., West Sussex, UK) and quantified using the Quant-iT PicoGreen dsDNA assay kit, to ensure the sample contained enough DNA for sequencing analysis. Sequencing was carried out on the Illumina MiSeq personal sequencer at the Centre for Genomic Research, University of Liverpool. A total of 9.5 million reads were obtained.

Raw reads were processed using the DADA2 (v. 1.4, Callahan *et al*., 2016) pipeline, specifically following the workflow for Big Data (https://benjjneb.github.io/dada2/bigdata.html), and using R version 3.4.0 (R Core Team, 2017). Forward and reverse read pairs were trimmed and filtered (minimum length 200 nucleotides, EEmax < 2 expected errors). Amplicon sequence variants (ASV's) were independently inferred in each sample from forward and reverse reads using the run-specific error rates, and then joined using the ~mergePairs~ function. Chimeric ASVs were inferred and identified using ~removeBimeraDenovo~ and removed. This resulted in 2756 final ASVs (per sample average = 20750, min = 8708 and max = 63995) which were taxonomically classified against the SILVA database (v.128, Quast *et al*., 2013) using the DADA2 implementation of the RDP's naive Bayesian classifier (Wang *et al*., 2007).

3.2.3.4. qPCR

Quantification of total bacteria and AOB was carried out using qPCR and amplification of the 16S rRNA gene and the ammonia monooxygenase (*amoA*) gene respectively. Samples were amplified in triplicate on a CFX96 Real-Time PCR Detection System (Bio-Rad, UK) using the primer sets 338F (Muyzer *et al*., 1993) and 1046R (Huber *et al*., 2007) for total bacteria and amoA-1F* (Stephen *et al*., 1999) and amoA-2R (Rotthauwe *et al*., 1997) for AOB. qPCR reactions contained 3 μL of template DNA (sample DNA, standard DNA or molecular grade water (negative control)), 0.5 μL of forward and reverse primer (10 ρmoles per μL), 5 μl of SsoFast EvaGreen supermix (Bio-Rad, UK) and 1 μl of molecular-grade water. Reaction conditions were: 1 cycle at 98 ^{o}C for 3 minutes, followed by 40 cycles consisting of 98 ^{o}C for 5 seconds and 60 ^{o}C (338F/1046R) or 56 ^{o}C (amoA-1F/amoA-2R) for 5 seconds. Purified circular plasmids containing the target gene were used as standards and run in triplicate for each qPCR reaction. Efficiencies for all qPCR reactions ranged between 90-110% and had a $R^2 \geq 0.99$. Gene copy numbers per unit volume were converted to cell numbers per unit volume using accompanying sequence data for the 16S rRNA gene (described in Appendix III) and assuming each AOB cell contained 2 copies of the amoA gene (McTavish *et al*., 1993; Norton *et al*., 2002).

3.2.4. Statistical Analysis

All statistical analysis, unless otherwise stated, was undertaken in RStudio (v. 1.0.143, R Core Team, 2017) using R version 3.4.0 (R Core Team, 2017).

3.2.4.1. Virus – Biotic/Abiotic Interactions

Virus – biotic/abiotic interactions were assessed by multivariate generalised least squares (GLS) regression (gls, "nlme" v. 3.1-131, Bates *et al*., 2017), with models describing ML virus, total bacteria and AOB abundance produced. All measured biotic and abiotic parameters, unless otherwise stated (see footnotes of Table III.2 – III.4), were initially used as covariates (K_i = ~59, n = 102), with multivariate ordinary least squares (OLS) regression (lm, "stats" v. 3.4.0, R Core Team, 2017) followed by bidirectional elimination (stepAIC, "MASS" v. 7.3-47, Venables and Ripley, 2002) based on Bayesian information criterion (BIC) used for model simplification and selection. OLS models were then rerun as GLS to allow incorporation of a correlation structure (corCAR1(form = ~ Week), accounting for slight variations in sampling frequency (approximately weekly) and temporal dependence (Zuur *et al*., 2010). To guarantee parsimony and adherence to the assumptions of linear regression manual backward elimination and forward selection based on BIC was subsequently undertaken on all GLS models, which were fit by maximum likelihood.

Models were checked visually for linearity (Fig. III.4 – III.6, A and B), homoscedasticity (Fig. III.4. – III.6, B and C), residual autocorrelation (acf and pacf, "stats" v. v. 3.4.0, R Core Team, 2017, Fig. III.4. – III.6, E and F), and normality (qqPlot, "car" v. 2.1-4, Fox et al., 2016, Fig. III.4 – III.6, D), with the Anderson-Darling Test (P > 0.05, ad.test, "nortest" v. 1.0-4, Gross and Ligges, 2015) additionally used to confirm the latter. Collinearity amongst explanatory variables was assessed by variance inflation factor's (VIF's, vif, "car" v. 2.1-4, Fox *et al*., 2016), with variables contributing to VIF's > 3 removed based on statistical significance (P-value) until all fell below this threshold (Zuur *et al*., 2010). To aid in adherence to these assumptions variables were either transformed ($\log_{10}$ for virus, total bacteria and AOB abundances (biotic) and natural log for COD, NH_4^+- N, nitrate, nitrite and phosphate terms (abiotic)) or standardised to mean 0 variance 1 (all other environmental and operational parameters (abiotic)). Finally ANOVA (anova, "stats" v. v. 3.4.0, R Core Team, 2017) was used to test the statistical significance of each model's correlation structure, whilst a pseudo-R^2 was calculated to ascertain a representation of the variance in ML abundances explained by each of the models respectively (r.squaredLR, "MuMin" v. 1.15.6, Bartoń, 2016).

3.2.4.2. Virus – Community Structure Interactions

Canonical correspondence analysis (CCA) was performed to assess the response of the bacterial communities to abiotic/biotic conditions (cca, "vegan" v. 2.4-3, Oksanen *et al*., 2017), with all measured biotic and abiotic parameters, transformed as in 2.4.1, initially used as explanatory variables (K_i = ~59, n = 102) unless otherwise stated (see caption of Fig. 3.3). Automated

bidirectional selection, based on Akaike information criterion (AIC) and Monte Carlo permutation tests (999 permutations), was then used for model simplification and thus identification of the most significant explanatory variables (step, test="perm", "stats" v. v. 3.4.0, R Core Team, 2017). Collinearity was assessed by VIF's (vif.cca, "vegan" v. 2.4-3, Oksanen *et al*., 2017), with variables contributing to VIF's > 3 being removed as in 2.4.1 (Zuur *et al*., 2010). Finally the statistical significance of the CCA model (constrained components), its axes and the marginal effects of each explanatory variable were assessed using permutation tests (999 permutations, anova.cca, "vegan" v. 2.4-3, Oksanen *et al*., 2017). Note CCA was chosen over redundancy analysis as unimodal approaches are better suited to relative abundances and the presence of zeros (Ramette, 2007).

Local similarity analysis (LSA, Ruan *et al*., 2006) was utilised to observe correlations between the 50 most abundant OTU's (relative abundances) and ML virus abundance. Its use over more traditional approaches (e.g. Pearson and Spearman's rank correlation) was justified given it assesses local (short periods of time) and time-delayed associations, as well as those non-lagged and occurring across the whole sampling period (Xia *et al*., 2011; 2013). Analysis was undertaken using eLSA v 1.0.2 (Xia *et al*., 2011; 2013) in Python v 2.7. A maximum time delay of three was utilised (delayLimit = 3), *P*-values were calculated by permutation tests (1000, *P*-valueMethod = perm), the required precision of *P*-values was set at 1/1000 (precision = 1000) and data was rank-normalised and z-transformed (normMethod = robustZ, Ruan *et al*., 2006; Xia *et al*., 2011; 2013). Multiple hypothesis correction was undertaken using *Q*-values (Storey, 2002). The LSA output was then visualised as an association network in Cytoscape v 3.6.0 (Shannon *et al*., 2003), only correlations with a *P*-value ≤ 0.05, a LSA score (LS) ≥ 0.3 and a *Q*-value ≤ 0.01 were examined. Calculation of the peak (max relative abundance) to average (mean relative abundance) ratio (PAR) for examined OTU's allowed assessment of their persistence within the ML, three arbitrarily defined ecological categories facilitated this process; persistent (PAR ≤ 5), intermittent (PAR > 5 < 10) and transient (PAR ≥ 10).

The influence of virus abundance on alpha diversity was assessed by calculation of Spearman's rank correlation coefficients (cor.test, "stats" v. 3.4.0, R Core Team, 2017) between ML and effluent virus abundance and D_1 (exponential of Shannon diversity) and D_2 (inverse of Simpson diversity) Hills diversities (diversity, "vegan" v. 2.4-3, Oksanen *et al*., 2017), which better represent rare and common taxa respectively (Vuono *et al*., 2015). Its use, over Pearson correlation, was justified since all variable combinations, transformed as in 3.2.4.1, were not bivariate normal (roystonTest, "MVN" v. 4.0.2, Korkmaz *et al*., 2014). Bonferroni corrections were applied to all calculated correlations.

3.2.4.3. Virus – Community Function Interactions

Virus-community function interactions were assessed by calculation of Spearman's rank correlation coefficients between ML and effluent virus abundance and ML and effluent COD and NH_4^+-N concentrations respectively, performed, justified and corrected as in 3.2.4.1. Structured equation modelling (SEM), which is well suited to studying hypotheses about multiple processes operating in a system (Grace *et al*., 2010), was also conducted to infer hypothesised causal links between virus abundance and effluent COD and NH_4^+-N concentrations (sem, "lavaan" v. 0.5-23.1097, Rosseel, 2012). *A priori* models were constructed based on literature and theory, improved upon using statistical associations identified by regression analysis (section 3.2.4.1) and, to guarantee parsimony, simplified by removing nonsignificant indicators and pathways ($P < 0.1$). Model fits were assessed using the χ^2 test ($P > 0.05$), the root square mean error of approximation (RMSEA, < 0.06), the root mean square residual (RMR, < 0.08), the comparative fit index (CFI, > 0.95) and AIC respectively (Hu and Bentler, 1999; Grace *et al*., 2010), with R^2 values obtained for each dependent matrix (Grace *et al*., 2010).

Prior to SEM analysis all bivariate relationships were checked for nonlinearity (cor.test, "stats" v. 3.4.0, R Core Team, 2017) and, through evaluation of skewness and kurtosis values, multivariate normality (mardiaTest, "MVN" v. 4.0.2, Korkmaz *et al*., 2014). To aid in adherence to these assumptions variables were transformed (log_{10} and natural log for biotic and abiotic respectively) and subsequently standardised to mean 0 variance 1, whilst SEM models, to counter minor normality departures, were estimated using a robust maximum likelihood approach (estimator = "MLM", "lavaan" v. 0.5-23.1097, Rosseel, 2012) and Satorra-Bentler scaled fit statistics (Curran *et al*., 1996).

3.3. Results

3.3.1. Bioreactor Performance and Abiotic Conditions

Influent flow rates were highly variable over the two-year study period (Table III.1 and Fig. III.1), leading to fluctuating hydraulic and solids retention times (10.6 ± 3.1 hours and 11.1 ± 3.1 days respectively, Fig. III.1) and reactor biomass concentrations (2.7 ± 0.5 g L^{-1} MLSS and 2.01 ± 0.4 g L^{-1} MLVSS, Fig. III.2). Reactor temperature varied on a seasonal basis, with a summer maximum and winter minimum of 17.9 °C (August 2012) and 6.3 °C (January 2013) recorded respectively (Fig. III.1). DO concentrations were moderately variable (2.5 ± 0.8 mg L^{-1}) and pH was maintained within a narrow range (6.6 ± 0.2) through flow dependent dosing of sodium hydroxide (Fig. III.1).

The discharge consent, a legislated maximum final effluent concentration, of COD_s (125 mg L^{-1}) was achieved 96.1% of the time, thus mean removal efficiency (82.1 ± 18.1 %) and effluent concentrations (32 ± 28.5 mg L^{-1}) were relatively stable (Fig. III.2 B). Nitrification was less efficient and more unstable (mean NH_4^+-N removal of 75.8 ± 27.6 %); accordingly effluent NH_4^+-N and nitrite concentrations were highly variable (0 - 76.4 and 0 - 5.3 mg L^{-1}) and the discharge consent limit (5 mg L^{-1} ammonia) was adhered to only 57.8% of the time (Fig. III.2). Certainly nitrite accumulation, thus incomplete nitrification, was periodically evident, particularly from day 500 onwards (Fig. III.2 E). Of the trace metals monitored in the influent calcium, potassium and magnesium were present at the highest concentrations (mean values of 47.4 ± 8.7, 13.1 ± 3.1 and 11 ± 2.8 mg L^{-1} respectively), whilst cadmium and arsenic were present at the lowest concentrations (mean values of 3.9 ± 11.3 and 9.3 ± 7.9 µg L^{-1} respectively, Fig. III.3). All monitored trace metals, however, were routinely present in the influent (Table III.1 and Fig. III.3).

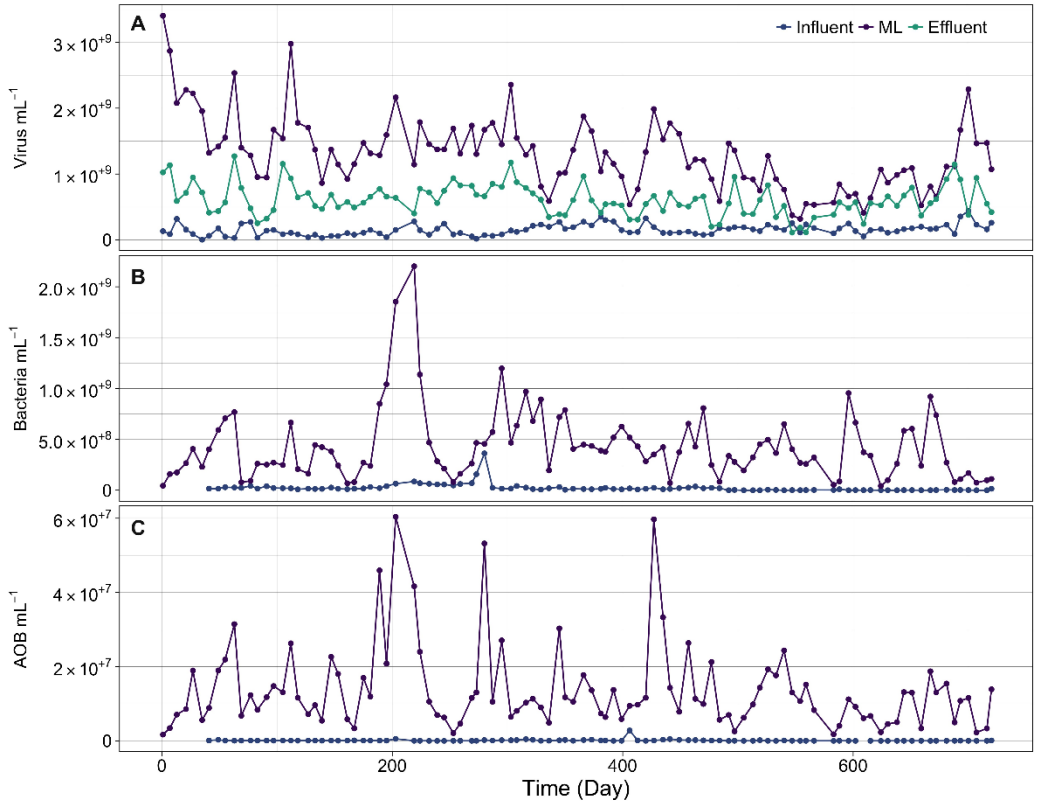

Figure 3. 1. The abundance of viruses (**A**), total bacteria (**B**) and AOB (**C**) in the influent, ML and effluent (viruses only) of full scale WWTP.

3.3.2. Temporal Abundance Dynamics of Viruses and Bacteria

ML virus abundance varied by roughly an order of magnitude (3.18×10^8 - 3.41×10^9 viruses mL^{-1}) throughout the two years, with temporal fluctuations evident (Fig. 3.1 A). Virus abundance was greater than bacterial abundance; averaging $1.31 \pm 0.57 \times 10^9$ viruses mL^{-1} compared with $4.30 \pm 3.48 \times 10^8$ bacteria mL^{-1} and $1.36 \pm 1.13 \times 10^7$ AOB mL^{-1}. The abundance of total bacteria and AOB

also varied across orders of magnitude ($4.29 \times 10^7 – 2.20 \times 10^9$ bacteria mL^{-1} and 1.69×10^6 - 6.04×10^7 AOB mL^{-1}), with fluctuations generally, but not exclusively, occurring concurrently (Fig. 3.1 B and C). The mean ML virus to total bacteria ratio (VBR) was 6.15, although this varied greatly (0.52 – 76.86).

The abundance of all three communities within the influent was orders of magnitude lower (Fig. 3.1), with mean abundances of $1.52 \pm 0.83 \times 10^8$ viruses mL^{-1}, $2.35 \pm 4.24 \times 10^7$ bacteria mL^{-1} and $1.14 \pm 2.98 \times 10^5$ AOB mL^{-1} recorded respectively. Synchronous, order of magnitude abundance fluctuations, analogous to those in the ML, were again evident (1.79×10^6 - 4.38×10^8 viruses mL^{-1}, $2.81 \times 10^5 – 23.62 \times 10^8$ bacteria mL^{-1} and 2.1×10^3 - 2.8×10^6 AOB mL^{-1}, Fig. 3.1). Effluent virus abundance ($6.1 \pm 2.39 \times 10^8$ viruses mL^{-1}) was also highly variable (1.1×10^8 - 1.27×10^9 viruses mL^{-1}, Fig. 3.1 A) and often mimicked abundance fluctuations in the ML. Indeed ML and effluent virus abundance were positively correlated ($P < 0.001$, Fig. 3.1), suggesting ML viruses are lost in effluent wastewater.

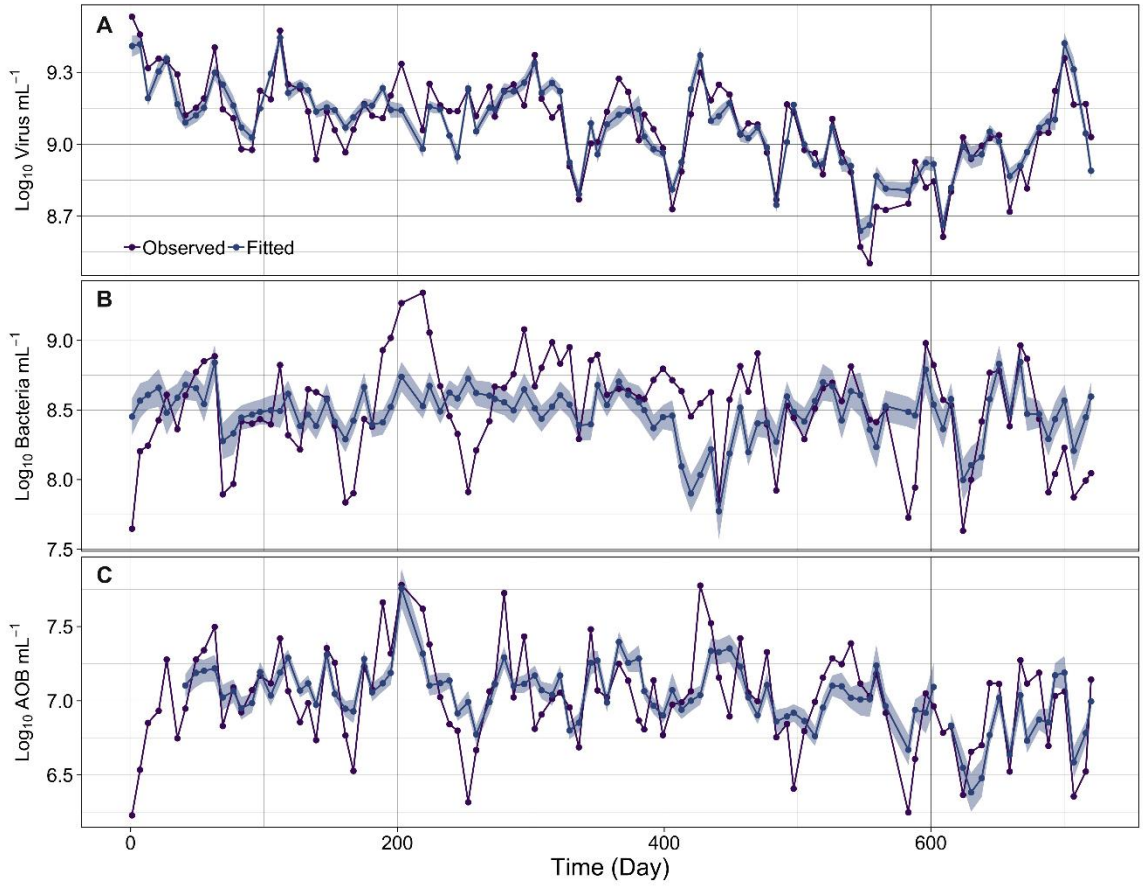

Figure 3. 2. Log_{10} ML virus (**A**), bacteria (**B**) and AOB (**C**) abundance from the full scale WWTP (observed) and respective multivariate GLS regression models (fitted). Shaded area represents 95% confidence intervals for the regression models. *n* = 102 (**A** and **B**) and 95 (**C**) respectively, the latter due to missingness in influent AOB abundance.

3.3.3. Virus Interactions with Biotic and Abiotic Conditions

Fitted values from the ML virus GLS regression model agreed well with observed abundances (Fig. 3.2 A and Fig. III.4 A), with explanatory variables explaining 83% of abundance variations. The model identified strong positive associations with influent virus abundance ($P < 0.05$), influent NH_4^+- N ($P < 0.01$), phosphate ($P < 0.01$) and sulphur ($P < 0.001$) concentrations, ML AOB abundance ($P < 0.05$) and ML pH ($P < 0.05$). In contrast influent magnesium concentrations ($P < 0.001$) and ML nitrate ($P < 0.001$), nitrite ($P < 0.01$) and sulphate ($P < 0.001$) concentrations were highly negatively associated (Table III.2).

Conversely the fitted values from the ML total bacteria and AOB GLS regression models showed some disparity from observed abundances (Fig. 3.2, III.5 A and III.6 A), with explanatory variables explaining 53% and 47% of variations in abundances respectively. Both models did however identify strong positive associations with ML and, for the latter, effluent virus abundance ($P < 0.01$, $P < 0.001$ and $P < 0.01$ respectively, Table III.3 and III.4). Other explanatory variables significantly contributing to each model are summarised in Appendix III (section III.2.2.1 and III.2.2.2 respectively).

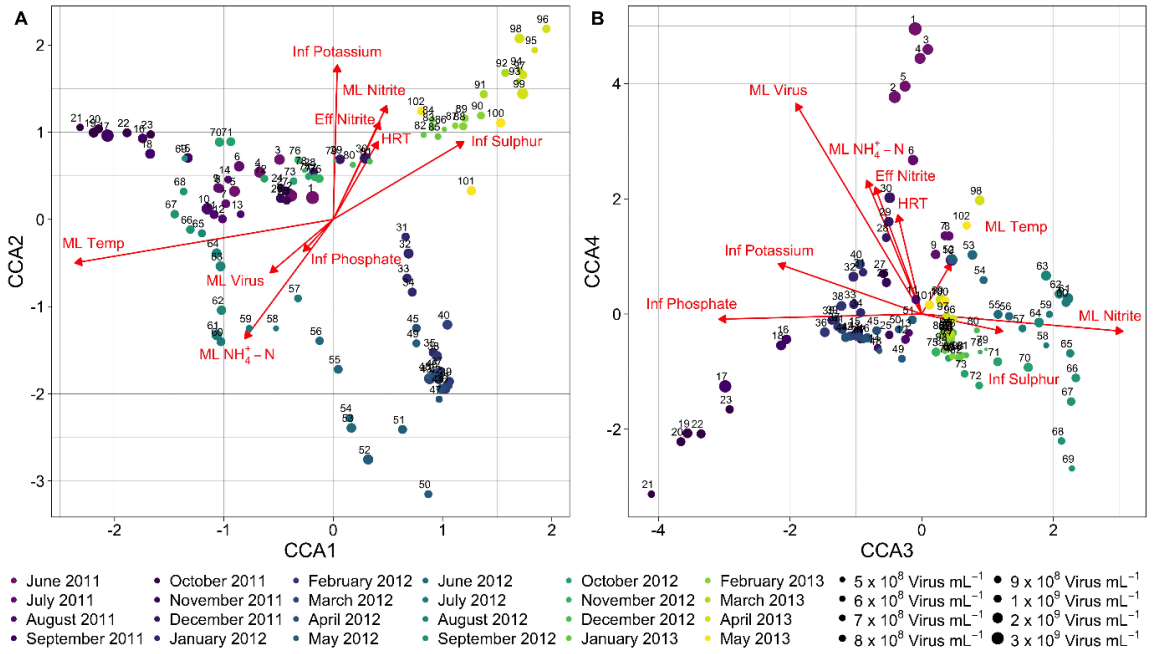

Figure 3. 3. CCA ordination of temporal ML bacterial community dynamics from the full scale WWTP and associated virus and abiotic variables, axes 1 and 2 (**A**) and 3 and 4 (**B**) respectively. Points and associated numbers represent individual samples, coloured and sized based on the month and year of collection and the samples virus concentration mL^{-1} respectively. Proximity of samples indicates similarities in community composition. Arrows indicate the direction of increase in each explanatory variable, their length indicates the strength of correlation with each axis and the angle between arrows indicates the approximate degree to which explanatory variables are correlated. Inf = influent, Eff = effluent.

3.3.4. Virus Interactions with Bacteria Community Structure

The CCA ordination explained the majority of observed variance in taxa-environment associations (77.6%, Table III.5), however it only encompassed a minority of the variance in the taxa data (21.1%, Table III.5). All four axes displayed high taxa-environment correlations (Table III.5), indicating strong associations between taxa, viruses and abiotic conditions (Fig. 3.3).

ML temperature was the most significant variable to the ordination ($P < 0.001$), followed by influent potassium ($P < 0.001$) and ML virus abundance ($P < 0.001$) respectively (Fig. 3.3 and Table III.5). ML Nitrite ($P < 0.001$) and influent phosphate ($P < 0.001$) and sulphur ($P < 0.001$) were also seemingly important, as were HRT ($P < 0.01$), ML NH_4^+- N ($P < 0.01$) and effluent nitrite ($P < 0.01$) but to a lesser degree (Fig. 3.3 and Table III.5). Although not included in the final CCA ordination influent virus abundance was found to significantly influence the ordination if manually added ($P < 0.1$), although weakly.

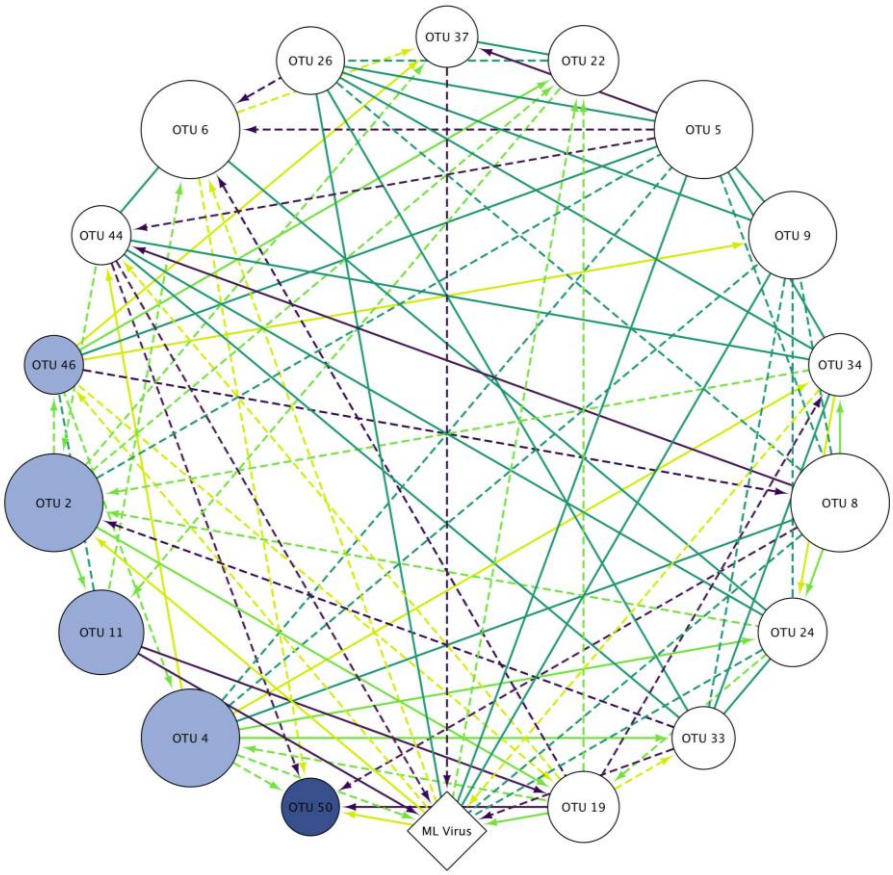

Figure 3. 4. Network visualisation of LSA associations between all OTU's from the top 50 most abundant that were associated with ML virus abundance (LS ≥ 0.3, $P \leq 0.05$ and Q ≤ 0.01) at the full scale WWTP. Circular nodes indicate bacterial taxa, coloured and sized based on the OTU's PAR (≤ 5 = white, > 5 < 10 = light blue and ≥ 10 = dark blue) and mean relative abundance respectively. For reference OTU 2 and OTU 50 had mean relative abundances of 4.68%, the highest, and 0.36%, the lowest. The diamond shaped node indicates ML virus abundance. Solid and dashed lines indicate positive and negative associations respectively, whilst a lines colour indicates the correlations delay (teal = 0, purple = 1, yellow = 2 and green = 3). Arrows point to the delayed, or trailing, node in delayed associations.

ML virus abundance was found to be significantly associated (LS > 0.3, $P \leq 0.05$ and $Q \leq 0.01$) with 17 of the top 50 OTU's, of which 6 were in the top 10 (Fig. 3.4). Of these associations 5 had no time lag, 2 negative (OTU's 8 and 24) and 3 positive (OTU's 5, 9 and 26). Indeed this group of OTU's were highly interconnected at time T, OTU's 5, 9 and 26 being positively associated and OTU pairs 8 and 5, 8 and 9, 8 and 26 and 24 and 9 being negatively associated respectively (Fig. 3.4). All 5 were also highly persistent in the ML (PAR ≤ 5, Fig. 3.4). Of the remaining 12 identified associations 4 had a time delay of 1 (Fig. 3.4), 1 positive (OTU 11) and 3 negative (OTU's 33, 37 and 44), 5 had a time delay of 2, 2 positive (OTU's 2 and 50) and 3 negative (OTU's 5, 34 and 46 and 50), and 3 had a time delay of 3, 1 positive (OTU 19) and 2 negative (OTU's 4 and 22). 5 of these OTU's, 4 of which were present intermittently (PAR > 5 < 10) or transiently (PAR ≥ 10) within the ML (Fig. 3.4), were identified as bulking, filamentous genera, including *Microthrix* (OTU's 2, 11 and 37), Thiothrix (OTU 50) and *Tetrasphaera* (OTU 4). Other noteworthy genera associated with ML virus abundance included *Zoogloea*, *Acidovorax*, *Defluviimonas* and *Dechloromonas*, OTU's 5, 19, 24 and 26 respectively.

ML virus abundance was not statistically associated with either Hill's diversity indices (Fig. 3.5 A), although effluent virus abundance was weakly negatively correlated with D_1 (rare taxa, $P < 0.05$, Fig. 3.5 A) if Bonferroni corrections are disregarded.

3.3.5. Virus interactions with community function

ML and effluent virus abundance were highly positively correlated with both ML COD ($P < 0.001$ and $P < 0.01$) and NH_4^+- N ($P < 0.001$ and $P < 0.05$) concentrations, and weakly positively correlated with effluent NH_4^+- N concentrations ($P < 0.1$ and $P < 0.1$, Fig. 3.5 B).

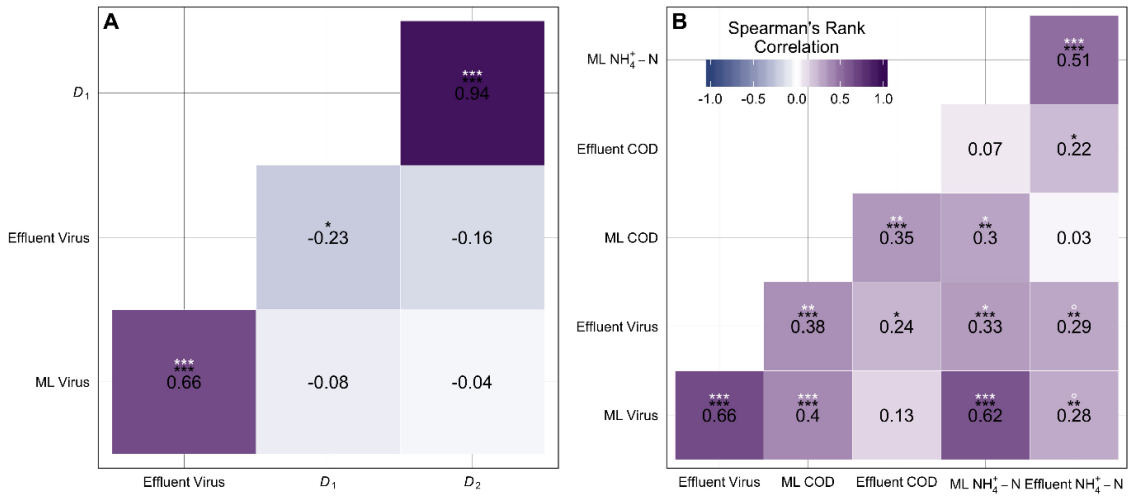

Figure 3. 5. Spearman's rank correlation coefficients describing virus – community structure (**A**) and virus – community function (**B**) interactions. Black and white * values are standard and Bonferroni corrected respectively. Royston's Multivariate Normality Test: *P* >

0.05 for all correlations and n = 102 (**A** and **B**). D_1 = exponential of Shannon diversity and D_2 inverse of Simpson diversity. ° P < 0.1, * P < 0.05, ** P < 0.01, *** P < 0.001.

The SEM analysis also identified significant positive associations, although non-directional, between ML virus abundance and both ML NH_4^+- N (P < 0.05, Fig 3.6 A) and COD (P < 0.01, Fig 3.6 B) concentrations, which were directly influencing, and being influenced by, corresponding effluent (P < 0.001 and P < 0.01, Fig 3.6) and influent concentrations (P < 0.001 and P < 0.05, Fig 3.6) respectively. Effluent virus abundance was also none directionally positively associated with effluent NH_4^+- N (P < 0.05, Fig 3.6 A) and COD (P < 0.01, Fig 3.6 B) concentrations, whilst being directly positively influenced by ML virus abundance (P < 0.001 and P < 0.001, Fig 3.6). ML virus abundance (P < 0.001, Fig 3.6 A), as well as ML AOB abundance (P < 0.1, Fig 3.6 A), was also directly positively influenced by influent concentrations of NH_4^+- N. No significant associations were found between ML AOB/bacteria abundance and ML and effluent NH_4^+- N/COD concentrations and virus abundance (Fig 3.6), the only exception being the negative association found between ML AOB and effluent virus abundance (P < 0.01, Fig 3.6 A). Overall the SEM analysis was able to explain 28% and 12% of effluent NH_4^+- N and COD concentrations respectively.

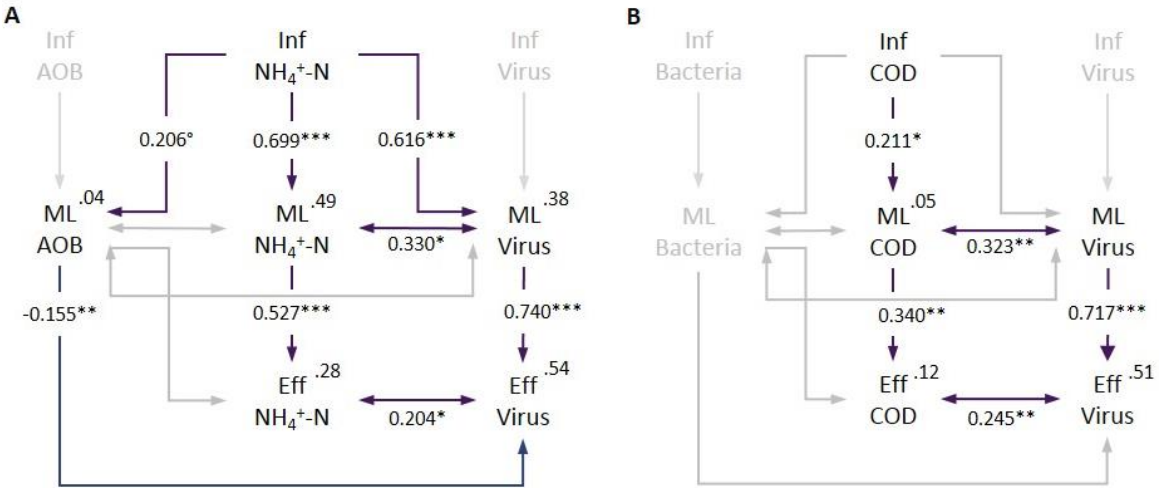

Figure 3. 6. SEM analysis describing viral interaction with NH_4^+- N (**A**) and COD (**B**) concentrations throughout the treatment stream at t WWTP. Standardised path coefficients and their significance are given along path arrows, which are coloured based on sign and significance (purple = positive, blue = negative, grey = non-significant (removed from final model)). Double-headed arrows indicate associations lacking a clear direction of causality, whilst the explained variance (R^2) of each endogenous variable is proximal to its label. n = 102, χ^2 = 7.79 (**A**, P = 0.352) and 3.64 (**B**, P = 0.603), RMSEA = 0.033 (**A**) and 0.000 (**B**), RMR = 0.044 (**A**) and 0.052 (**B**) and CFI = 0.997 (**A**) and 1.000 (**B**). ° P < 0.1, * P < 0.05, ** P < 0.01, *** P < 0.001. Inf = influent, Eff = effluent.

3.4. Discussion

Negative density dependent viral selection of hosts is increasingly considered in efforts to describe bacterial mortality, diversity and function in natural environments (Suttle, 2007; Brussard *et al.*, 2008; Rohwer and Thurber, 2009), a paradigm unanswered, until now, in an engineered setting, where such interactions could cause functional failures with environmental and financial consequences. The complex nature of ML microbial communities, coupled with the difficulty of

obtaining direct evidence of virus-host interactions (Dang and Sullivan, 2014; Brum and Sullivan, 2015), made this a challenging prospect, however if viruses significantly influence bacterial dynamics then this should be visible at the total abundance level using statistical inferences. Here we show this is indeed the case.

We were able to demonstrate ML virus abundance corresponded with ML AOB abundance and was positively associated with both ML total bacteria and AOB abundance respectively. Given viruses are obligate parasites and require hosts to replicate this makes intuitive sense, yet advocates, for the first time, coupled virus-bacteria dynamics at the total abundance level within ML, with greater bacterial densities plausibly increasing encounter rates, successful infections and thus virus abundance. We additionally sought and identified statistical associations between ML viruses and ML bacterial community shifts and the dynamics of specific OTU's, suggesting virus-host interactions, as observed previously, yet rarely, in an engineered setting (Huntula *et al*., 1991; Lee *et al*., 2007; Barr *et al*., 2010; Shapiro *et al*., 2010), appear to be occurring, particularly between viruses and highly abundant OTU's.

Given such taxa, along with AOB, likely dictate community function it follows that viral predation may be involved in observed functional failures, i.e. adversely affecting the removal of COD and NH_4^+- N from influent wastewater. It was perhaps expected that such associations, although not unequivocal, were demonstrated. ML virus abundance and concentrations of COD and NH_4^+- N were positively associated and were noteworthy predictors of respective abundances and concentrations in the effluent, where these relationships persisted. Such associations imply that increased host lysis in the ML at high virus densities, coupled with the subsequent release of intracellular carbon and nutrients, were a potential cause, as observed previously in lab scale engineered systems removing phosphorous (Barr *et al*., 2010; Motlagh *et al*., 2015). The observations here however are a first indication that viruses appear to influence the function of full-scale, nitrifying activated sludge systems.

Conversely the virus-community function interactions demonstrated here, may imply that the nutritional or metabolic status of bacterial hosts is critical to lytic infection and viral proliferation (Proctor *et al*., 1993; Middelboe, 2000; Weinbauer, 2004), especially in light of the corresponding associations found between resource availability (NH_4^+- N, nitrite, phosphate and sulphur concentrations) and ML virus, total bacterial and AOB abundance and bacterial community structure, as shown elsewhere (Hewson *et al*., 2003; Øvreås *et al*., 2003; Williamson and Paul, 2004; Motegi and Nagata, 2007; Sandaa *et al*., 2009). Thus viruses in the ML could be responding to, not instigating, resource fluctuations and subsequent functional failures, although their

increased activity could exacerbate a systems ability to cope with greater, or shock, loads. Furthermore the concurrent, community wide abundance and structural fluctuations may also suggest such resource driven, deterministic triggers are filtered through bacterial hosts, and thus viruses are responding to, not driving, changes in the ML bacterial community. Certainly the association between nitrite and ML virus abundance and the ML bacterial community here and elsewhere (Wells *et al*., 2009, 2011) supports such a concept, given it can be inhibitory to bacterial metabolism (Philips and Verstraete, 2001; Vadivelu *et al*., 2006; Zhou *et al*., 2011) and thus halt viral replication, reducing counts as observed.

Nevertheless the observed concurrent and/or sequential shifts in resources and both host and viral communities respectively is further, indirect evidence of viral-host interaction within full-scale activated sludge systems. Moreover, although direct evidence is lacking, it could be argued that such associations are in principle agreement with Killing the Winner" (KtW, Thingstad, 2000; Winter *et al*., 2010) and/or fluctuating selection dynamics (FSD, Hall *et al*., 2011; Avrani *et al*., 2012). Both principles are underpinned by a metabolic cost of host resistance, thus under such dynamics resources drive host and subsequently viral community fluctuations, as observed. They also predict viral interaction with, and as a consequence oscillatory dynamics in, the most abundant bacterial hosts. It is thus striking that 6 of the top 10 (17 of the top 50) OTU's were significantly associated with ML virus abundance, moreover the majority of these OTU's, as predicted under steady-state KtW dynamics (Thingstad, 2000), varied in their relative abundance but remained relatively abundant. Exceptions to this norm were *Microthrix*, *Thiothrix* and *Tetrasphaera* genera, filamentous bacteria known to periodically bloom and cause activated sludge bulking (a process disturbance, Martins *et al*., 2004; Guo and Zhang, 2012). During such events these organisms can dominate ML bacterial communities (relative abundance > 10% here and elsewhere (Griffin and Wells, 2017)), thus a viral response, as observed, would be expected under KtW and/or FSD. Particularly considering lytic viruses have previously been shown to influence filamentous organisms associated with bulking events (Kotay *et al*., 2011).

Viral dispersal was also determined as an important system parameter, with viruses in the influent correlating with ML virus abundance and bacterial community structure (weakly). Over 82% of virus genotypes are shared across influent and ML viromes (Tamaki *et al*., 2012), suggesting successful mass immigration is plausible and perhaps vital to virus proliferation in activated sludge systems. The influx of new viruses increases population densities, replication rates and beneficial gene abundance, increasing advantageous mutations (Miralles *et al*., 1999; Morgan *et al*., 2005, 2007; Ching *et al*., 2013) that can provide evolutionary advantages to viruses in their phage-host

co-evolutionary arms race (Buckling and Rainey, 2002). Such migration effects are perhaps even more important when one considers that ML viruses are continually washed out with process effluents (hence the correlation between ML and effluent virus abundance), whilst bacteria are maintained/augmented through sludge recirculation.

Finally the demonstration that ML virus abundance was associated with abiotic factors supports recent calls in viral ecology for their inclusion in viral studies (Wigington *et al*., 2016). The regulation of surface charge and the promotion or inhibition of electrostatic interactions appeared highly important to ML virus abundance, with ML pH and sulphate and influent magnesium all playing a significant role. At low pH and high ionic and cationic strength, the net negative surface charge of viruses (Michen and Graule, 2010; Nap *et al*., 2014), bacteria (Klausen *et al*., 2004) and activated sludge flocs (Cousin and Ganczarczak, 1999; Liao *et al*., 2002; Wilén *et al*., 2003; Tixier *et al*., 2003) decreases, collectively reducing repulsive electrostatic forces and thus increasing virus adsorption (Schaldach *et al*., 2006; Pham *et al*., 2009) and microcolony and floc compactness (Cousin and Ganczarczak, 1999; Liao *et al*., 2002; Klausen *et al*., 2004). Conceivably, such interactions increase viral floc enmeshment and thus EPS/colloidal masking of bacterial receptors (Kunin *et al*., 2008), reducing virus proliferation and, as observed, their abundance.

On the whole the statistically inferred relationships, including those beyond the scope of this paper (those identified by the ML total bacteria and AOB GLS models, as well those variables not discussed but contributing to the CCA ordination), have plausible physical, biological and chemical explanations. Those observed between ML virus abundance and ML fluoride and nitrate, which lack obvious mechanisms, and that between ML AOB and effluent NH_4^+- N, since a negative relationship would be expected, are perhaps the only anomalous associations. This may suggest that despite being the most comprehensive time series study of virus and microbial dynamics to date, sample frequency and the number of exogenous factors measured was still insufficient to capture the true behaviour of the system.

In summary, this study demonstrates that viruses appear to play a more central role in the dynamics of activated sludge systems than hitherto realised and should be considered more frequently when assessing the key factors governing bacterial abundance, community composition and functional stability. To gain further insight into virus dynamics in engineered systems a survey encompassing increased sample frequency, i.e. over more functionally relevant temporal scales, is perhaps required, whilst the identification and enumeration of active viruses and their hosts is paramount. Those taxa found to be associated with ML virus abundance here may be a good basis for investigation in this regard.

3.5. Acknowledgements

This work was carried out as part of a Frontiers Engineering research project (NUFEB, http://research.ncl.ac.uk/nufeb/) awarded and funded by the Engineering and Physical Sciences Research Council (EP/H012133/1). I'd like to thank Northumbrian Water who permitted collection of wastewater samples from the full scale plant, Mary Lunn of Oxford University for guidance on statistical analysis and Joana Baptista and Ben Allen of Newcastle University who undertook sample preparation for illumina sequencing and subsequent bioinformatics analysis respectively.

CHAPTER 4

VIRUS - BACTERIA INTERACTIONS, SYNCHRONICITY AND ECOSYSTEM FUNCTION IN REPLICATE ENGINEERED MICROBIAL SYSTEMS

Viruses are ubiquitous, abundant and a key component of natural environments, influencing bacterial host abundance, community composition and ecosystem functions. Assessment of their role in the microbial ecology of engineered systems has however been precluded by a paucity of data, a trend analogous to the examination of biotic and abiotic drivers of natural virus abundance variations. Yet evaluation of the former fundamentally contributes to the latter, given the physical partitioning of engineered systems into homogeneous ecological islands with well-defined chemical and physical conditions. Hence virus abundance, in conjunction with total and ammonia oxidising bacterial abundances, bacterial community profiles, and a suite of environmental and operational parameters, was monitored every other day for ~200 days in 12 replicate lab-scale activated sludge systems. We demonstrate that mixed liquor virus abundances were spatiotemporally variable and asynchronous, yet statistically associated with total bacterial abundances, the composition of bacterial communities, bacterial diversity and a number of previously identified abiotic factors (Chapter 3, including effluent concentrations of COD and NH_4^+-N). Such findings corroborate previous findings (Chapter 3) and imply viruses may play a role in, or respond to, bacterial community divergences, thus are more central in the dynamics of activated sludge system than hitherto realised.

4.1. Introduction

Activated sludge systems, the most frequently used form of biological wastewater treatment (Seviour *et al*., 2010), are reliant on dozens, perhaps hundreds, of different bacterial species to come together and form a microbial community consistently capable of removing organics and nutrients from wastewaters. Thus the function, functional stability and robustness of such systems is ultimately dictated by the mechanisms involved in the assembly and maintenance of their microbial communities. Our knowledge of these processes, despite global application, is however poor, thus functional failures are observed frequently, unpredictably and often inexplicably (Curtis and Sloan, 2006).

In response, contemporary wastewater microbiology has elucidated two classes of community assembly mechanism at work in engineered biological systems, niched-based deterministic factors, such as environmental and operational conditions, and stochastic processes, including microbial birth, death, immigration and speciation (van der Gast *et al*., 2008; Wells *et al*., 2009, 2011; Ayarza *et al*., 2010; Ofiteru *et al*., 2010; Ayarza and Erijman, 2011; Valentin-Vargas *et al*., 2012). Indeed the literature has utilised replicated or geographically localised biological reactors to evaluate the generality of such principles, with synchronous (Falk *et al*., 2007; Vanwonterghem *et al*., 2014; Griffin and Wells, 2017) and asynchronous (Kaewpipat and Grady, 2002; Fernandez *et al*., 2000;

Gentile *et al*., 2007; Beecroft *et al*., 2012; Zhou *et al*., 2013) microbial communities invoking niche-based deterministic factors and neutral processes as drivers of observed behaviour respectively.

Considering recent findings (Lee *et al*., 2007; Barr *et al*., 2010; Shapiro *et al*., 2010; Motlagh *et al*., 2015, Chapter 3) such an approach could help further elucidate the role viruses play in the dynamics of engineered systems. Moreover, given the physical partitioning of wastewater treatment plants into homogeneous ecological islands with well-defined and monitored chemical and physical conditions and functions (Daims *et al*., 2006), such systems offer a fertile testing ground for answering recent calls in viral ecology; that is, the need for greater efforts in understanding temporal variations in total virus abundance and how microbial abundances and exogenous factors relate (Breitbart, 2012; Brum and Sullivan, 2015; Wigington *et al*., 2016). In this regard lab based systems are particularly pertinent. Their accessibility facilitates high frequency temporal sampling, whilst their amenability to manipulation and replication enables the generality of specific deterministic drivers of virus abundance to be tested.

Accordingly the primary objective of this work was to gain a better understanding of total virus abundance dynamics in activated sludge systems. Thus their abundance and relationship with total bacterial and ammonia oxidising bacterial (AOB) abundance, bacterial community structure and a suite of exogenous factors, including functional stability (COD and NH_4^+-N removal), was assessed at high temporal frequency in 12 replicate lab scale activated sludge systems. Temperature, known to influence microbial and virus communities in activated sludge (e.g. Wells *et al*., 2009, 2011; Ofiteru *et al*., 2010) and aquatic environments respectively (Wommack and Colwell, 2000; Weinbauer, 2004), was manipulated in 6 of these systems to instigate community divergences. Thus temporal population synchrony was quantified to identify repeatable viral and bacterial community responses to abiotic or biotic triggers, alone or reciprocally. Synchrony was thus examined across multiple components, including individual total abundances, bacterial community profiles and abiotic conditions.

4.2. Materials and Methods

4.2.1. Reactor Set Up

Twelve replicate continuous flow stirred-tank reactors (CSTR's), with a working volume of 950 mL ± 26 mL, were operated simultaneously for 204 days. At day 0 reactors were seeded with activated sludge from a nitrifying domestic wastewater treatment plant (WWTP; Tudhoe Mill, Durham, United Kingdom (UK)) at a mixed liquor suspended solids (MLSS) concentration of ~2 g L^{-1}. Settled sewage (influent), collected weekly and stored at 4°C, was continuously fed to each CSTR to

achieve a hydraulic and solids retention time (HRT/SRT) of ~4 days. Dissolved oxygen (DO) concentrations were maintained above 4 mg L^{-1}, temperature was controlled using two cooling incubators (MIR554, Panasonic, UK) and each CSTR was stirred at 200 rpm using magnetic stirrers (Maxdrive 1 Eco, 2 Mag, Germany) and mixing bars (40 mm × 8 mm ø). The temperature, pH and dissolved oxygen (DO) concentration within each CSTR was monitored in real time using individual probes (Type K Thermocouples (RS, UK), F-635 Fermprobes and D-140 Oxyprobes (Broadley James, UK) respectively) and associated transmitters (MXD70, Broadley James, UK), data loggers (PICO, UK) and computer software (PicoLog 5.21.5, PICO, UK), with readings taken every 10 minutes.

4.2.2. CSTR Operational Conditions

All CSTR's were initially operated under steady state conditions, where possible, for a period of 72 days, allowing for acclimatisation. Thus temperature and DO concentrations were maintained at 14.5°C and above 4 mg L^{-1} respectively, whilst pH was left to stabilise naturally. In order to maintain conditions of an open, system real wastewater was utilised and therefore influent characteristics, including soluble chemical oxygen demand (COD_s), ammonium (NH_4^+-N) and phosphate concentrations, varied on a weekly basis.

Once stable conditions were achieved 6 CSTR's (from herein called Test (T)) were subjected to a 132 day temperature sine wave mimicking annual variation in the UK; thus temperature's ranged from 8°C to 21°C. The remaining 6 CSTR's (from herein called Control (C)) were exposed to a constant temperature of 14.5°C. All other operational and environmental conditions were maintained identically across both sets of CSTR's

4.2.3. Sample Collection and Storage

Every other day 40 mL of mixed liquor (ML) was collected from each CSTR into sterile 50mL centrifuge tubes and divided into 3 sub samples. 15 mL of each sample was stored at 4°C for viral analysis, which was undertaken within 1 hour of sampling. A second 15mL was stored at -20°C until subsequent molecular analysis, with the remaining 10 mL stored at -80°C as an archived sample. Concurrent influent and effluent samples, as well as weekly influent samples, were also collected. Analysis was performed on samples collected every other day unless otherwise stated.

4.2.4. Analytical Methods

Influent and effluent COD_s and NH_4^+-N concentrations were determined using Merck COD and NH_4-N test kits (VWR, UK) respectively, whilst anion concentrations, which included nitrate, nitrite, sulphate and phosphate, were determined using high performance Ion Chromatography (Dionex

ICS-1000 with AS40 auto sampler). Samples were filtered through a 0.2 µm polyethersulfone membrane prior to analysis. Influent and ML suspended/volatile suspended solids (SS/VSS) were determined according to Standard Methods (APHA, 1998). Influent trace metals, including cadmium, zinc, lead and copper, were measured weekly by inductively coupled plasma optical emission spectroscopy (ICP-OES) (Vista MPX axial ICP-OES, Varian, UK), as described by Martin *et al*. (1994). Samples were acidified on collection to pH < 2, digested and then filtered through a 0.45 µm polyethersulfone membrane prior to analysis. Finally temperature, DO and pH within each CSTR was measured as described in 2.1, with influent pH additionally measured manually (Jenway 3310 pH meter).

4.2.5. Molecular Methods

4.2.5.1. Flow Cytometry

For virus enumeration, 1mL sub-samples of influent (every week) and ML were taken, transferred into 2 mL cryovials and fixed at a final concentration of 0.5% Glutaraldehyde for 15-30 minutes at 4^{o}C in the dark. Samples were then flash frozen in liquid nitrogen and stored at -80^{o}C. After defrosting samples were pre-treated and analysed in triplicate as described by Brown *et al*. (2015 (Chapter 2)) using a FACScan flow cytometer (Becton Dickinson, USA) equipped with a 15-mW 488-nm air-cooled argon-ion laser and a standard filter setup.

4.2.5.2. DNA Extraction

DNA was extracted from 15 mL of influent (every week) and ML respectively, both were centrifuged at 3392 × *g* for 15 minutes and the supernatant removed down to a working volume of 250 µL. Cell wall disruption was then carried out using the FastDNA SPIN Kit for soil (MP Biomedicals, USA), thus 244.5 µL of sodium phosphate buffer and 30.5 µL of MT buffer were added to samples and the mixture transferred to Lysing Matrix E tubes. Samples were then lysed at 6.5 ms^{-1} for 30 seconds in a FastPrep instrument (MP Biomedicals, USA) and centrifuged at 14000 × *g* for 15 minutes. DNA from 250 µL of the supernatant was then purified using a MagNA Pure LC 2.0 (Roche, UK) and the MagNA Pure LC DNA Isolation Kit III

4.2.5.3. Illumina Sequencing

Sample preparation for Illumina sequencing generally followed the protocol of Caporaso *et al*. (2012). Thus the V4 region of the bacterial 16S rRNA gene was amplified using primers 515F (aatgatacggcgaccaccgagatctacactatggtattgtGTGCCAGCMGCCGCGGTAA; adapter, primer pad and primer linker in small letters; and specific primer sequence in capital letters) and 806R

(caagcagaagacggcatacgagat**barcode**agtcagtcagccGGACTACHVGGGTWTCTAAT), the latter was barcoded with a 12-base error-correcting Golay code to facilitate sample multiplexing. Each sample was amplified in duplicate, pooled and then cleaned using a MinElute 96 UF Purification Kit as per the manufacturer's instructions (Qiagen Ltd., West Sussex, UK). PCR reactions consisted of: 2 µl DNA extract, 20 µl 5 Prime Hot Master Mix (VWR, Lutterworth, UK), 1 µl each of forward and of reverse primer (10 µM final concentration), and 26 µl Molecular-grade water. Reactions were denatured at 94 °C for 3 mins, with amplification proceeding for 35 cycles at 94 °C for 45 s (further denaturing), 50 °C for 60 s (annealing) and 72 °C for 90 s (extension); a final extension at 72 °C for 10 min was added to ensure complete amplification. The concentration and purity of pooled amplicons was assessed using a Quant-iT PicoGreen dsDNA assay kit (Life Technologies, Paisley, UK). A composite sample for sequencing was created by combining all samples in equimolar amounts. The composite sample was cleaned twice using Agencourt AMPure XP beads (Beckman Coulter, High Wycombe, UK) according to the manufacturer's instructions, with fragment selection undertaken using E-Gel 2% agarose gels (Life Technologies, Paisley, UK). The size-selected fragment was then purified using the QIAquick PCR purification kit (Qiagen Ltd., West Sussex, UK) and quantified using the Quant-iT PicoGreen dsDNA assay kit, to ensure the sample contained enough DNA for sequencing analysis. Sequencing was carried out on the Illumina MiSeq personal sequencer at the Centre for Genomic Research, University of Liverpool. A total of 9.5 million reads were obtained.

Raw reads were processed using the DADA2 (v. 1.4, Callahan *et al*., 2016) pipeline, specifically following the workflow for Big Data (https://benjjneb.github.io/dada2/bigdata.html), and using R version 3.4.0 (R Core Team, 2017). Forward and reverse read pairs were trimmed and filtered (minimum length 200 nucleotides, EEmax < 2 expected errors). Amplicon sequence variants (ASV's) were independently inferred in each sample from forward and reverse reads using the run-specific error rates, and then joined using the ~mergePairs~ function. Chimeric ASVs were inferred and identified using ~removeBimeraDenovo~ and removed. This resulted in 11253 final ASVs (per sample average = 21105, min = 2126 and max = 70897) which were taxonomically classified against the SILVA database (v.128, Quast *et al*., 2013) using the DADA2 implementation of the RDP's naive Bayesian classifier (Wang *et al*., 2007).

4.2.5.4. qPCR

Quantification of total bacteria and AOB was carried out using qPCR and amplification of the 16S rRNA gene and the ammonia monooxygenase (*amoA*) gene respectively. Samples were amplified in triplicate on a CFX96 Real-Time PCR Detection System (Bio-Rad, UK) using the primer sets 338F

(Muyzer *et al*., 1993) and 1046R (Huber *et al*., 2007) for total bacteria and amoA-1F* (Stephen *et al*., 1999) and amoA-2R (Rotthauwe *et al*., 1997) for AOB. qPCR reactions contained 3 μL of template DNA (sample DNA, standard DNA or molecular grade water (negative control)), 0.5 μL of forward and reverse primer (10 ρmoles per μL), 5 μl of SsoFast EvaGreen supermix (Bio-Rad, UK) and 1 μl of Molecular-grade water. Reaction conditions were: 1 cycle at 98°C for 3 min, followed by 40 cycles consisting of 98°C for 5 s and 60°C (338F/1046R) or 56°C (amoA-1F/amoA-2R) for 5 s. Purified circular plasmids containing the target gene were used as standards and run in triplicate for each qPCR reaction. Efficiencies for all qPCR reactions ranged between 90-110% and had a $R^2 \geq 0.99$. Gene copy numbers per unit volume were converted to cell numbers per unit volume using accompanying sequence data for the 16S rRNA gene (described in Appendix III) and assuming each AOB cell contained 2 copies of the amoA gene (McTavish *et al*., 1993; Norton *et al*., 2002).

4.2.6. Statistical Analysis

All statistical analysis, unless otherwise stated, was undertaken in RStudio (v. 1.0.143, R Core Team, 2017) using R version 3.4.0 (R Core Team, 2017).

4.2.6.1. Comparison of CSTR's Abiotic Conditions

The performance of and abiotic conditions within each CSTR were compared and analysed for significant differences using the Kruskal-Wallis test ($P > 0.05$, kruskal.test, "stats" v. 3.3.2, R Core Team, 2017) and, if significant, the Dunn test with Bonferroni corrections ($P > 0.05$, dunnTest, "FSA" v. 0.8.12, Ogle, 2017). To justify non-parametric analysis data was checked for normality and homogeneity of variance using the Anderson-Darling Test ($P > 0.05$, ad.test, "nortest" v. 1.0-4, Gross and Ligges, 2015) and the Bartlett Test ($P > 0.05$, bartlett.test, "stats" v. 3.3.2, R Core Team, 2017) respectively.

The synchrony of functional components, defined here as synchronous changes in effluent concentrations of COD and NH_4^+-N respectively in separate CSTR's, was calculated using spearman's rank correlation (cor.test, "stats" v. 3.4.0, R Core Team, 2017). For each component, *N*, spearman's correlation $S_{i,j}^N$ of its concentration across time in each CSTR pair (i, j) was calculated. Global ($\bar{S}$), among test ($\overline{S_T}$) and among control ($\overline{S_C}$) synchrony values for each component were then calculated by averaging $S_{i,j}^N$ across all, test and control CSTR pairs respectively. Owing to compositional effects (Aitchison, 1982) and inherent autocorrelation in time-series data (Liebhold *et al*., 2004) correlation coefficients will not necessarily be zero for uncorrelated components.

Finally standardised Euclidean distances were generated for all possible pairs of samples using abiotic measurements collected from the mixed liquor and effluent of each CSTR. This distance matrix was then used to generate analysis of similarity (ANOSIM) and analysis of variance (anosim and adonis respectively with 999 permutations, "vegan" v. 2.4-3, Oksanen *et al*., 2017) statistics, allowing assessment of the similarity in the abiotic conditions across all, between test and control, among test, among control and between all CSTR pairs. ANOSIM and Adonis generate an R and R^2 statistic respectively, the magnitude of R, which ranges between 0 and 1, and R^2 indicates the degree of separation between groups of samples, with smaller and larger values indicating lower and greater separation respectively. For pairwise analysis between all CSTR's Bonferroni corrections were applied to R and R^2 P-Values.

4.2.6.2. Comparison of CSTR's Biotic Conditions

Significant differences between, and the synchronicity of, each CSTR's biotic conditions was analysed as in 4.2.6.1 using total virus, bacteria and AOB abundance and alpha diversity (α) indices D_1 (exponential of Shannon diversity) and D_2 (inverse of Simpson diversity) (diversity, "vegan" v. 2.4-3, Oksanen *et al*., 2017), which better represent rare and common taxa respectively (Vuono *et al*., 2015). Correspondence analysis (CA) was performed to assess the temporal development of bacterial community composition in each CSTR (cca, "vegan" v. 2.4-3, Oksanen *et al*., 2017), the degree of association, or concordance, between each CSTR's microbial community was then assessed using procrustean matrix superimposition on the first 3 CA axes (protest, "vegan" v. 2.4-3, Oksanen *et al*., 2017). This generates an m^2 statistic, the sum of squared residuals between scaled and rotated configurations of each ordination solution, which varies between 0 and 1, with smaller values indicating stronger concordance.

The Bray-Curtis dissimilarity coefficient (β_{BC}) was also calculated for all possible pairs of samples using microbial community data, allowing the similarity in microbial communities across all, between test and control, among test and among control CSTR pairs to be assessed using ANOSIM and Adonis as in 4.2.6.1. To further asses temporal shifts in the microbial communities of each CSTR a pairwise β_{BC} and Sorenson's dissimilarity index (β_{sor}) was calculated between each successive sample (beta.pair, "betapart" v.1.4-1, Baselga *et al*., 2017). The latter was partitioned to allow assessment of taxa incidence variations caused by species replacement ($\beta_{sor\text{-}tur}$) or nestedness ($\beta_{sor\text{-}nes}$), i.e. taxa at t are replaced with different taxa at t+1 or taxa at t+1 are a subset of taxa at t (beta.pair, "betapart" v.1.4-1, Baselga *et al*., 2017). The synchronicity of these components across CSTR's was assessed as in 4.2.6.1.

4.2.6.3. Virus – Total Abundance Biotic and Abiotic Interactions

Virus total abundance biotic and abiotic interactions were assessed by generation of a multivariate linear mixed effects model, fit by maximum likelihood (lme, "nlme" v. 3.1-131, Bates *et al*., 2017), describing ML virus abundance across all 12 CSTR's. All measured biotic and abiotic parameters, unless otherwise stated (see footnotes of Table IV.16), were initially used as covariates (K_i = ~51, n = 847), with multivariate ordinary least squares (OLS) regression (lm, "stats" v. 3.4.0, R Core Team, 2017) followed by bidirectional elimination (stepAIC, "MASS" v. 7.3-47, Venables and Ripley, 2002) based on Bayesian information criterion (BIC) used for model simplification and selection. OLS models were then rerun as lme to allow incorporation of a correlation structure (corCAR1(form = ~ Day) and CSTR as a random effect, helping to account for temporal and spatial dependence and slight variations between CSTR's (Zuur *et al*., 2010). To guarantee parsimony and adherence to the assumptions of regression manual backward elimination and forward selection based on BIC was subsequently undertaken.

The model was checked visually for linearity (Fig. IV.7 A and B), homoscedasticity (Fig IV.7 B and C), residual autocorrelation (acf and pacf, "stats" v. v. 3.4.0, R Core Team, 2017, Fig IV.7 F and G), and normality of fixed and random effects (qqPlot, "car" v. 2.1-4, Fox et al., 2016, Fig. IV. 7D and E), with the Anderson-Darling Test ($p > 0.05$, ad.test, "nortest" v. 1.0-4, Gross and Ligges, 2015) additionally used to confirm the latter. Collinearity amongst explanatory variables was assessed by variance inflation factor's (VIF's, vif, "car" v. 2.1-4, Fox *et al*., 2016), with variables contributing to VIF's > 3 removed based on statistical significance (p-value) until all fell below this threshold (Zuur *et al*., 2010). To aid in adherence to these assumptions variables were either transformed (log_{10} for virus, total bacteria and AOB abundances (biotic) and natural log for COD, NH_4^+- N, nitrate, nitrite and phosphate terms (abiotic)) or standardised to mean 0 variance 1 (all other environmental and operational parameters (abiotic)). Finally ANOVA (anova, "stats" v. v. 3.4.0, R Core Team, 2017) was used to test the statistical significance of the models correlation structure, whilst a pseudo-R^2 was calculated to ascertain a representation of the variance in ML virus abundance explained by the model (r.squaredGLMM, "MuMin" v. 1.15.6, Barton, 2016).

4.2.6.4. Virus – Community Structure Interactions

Canonical correspondence analysis (CCA) was performed to assess the response of bacterial communities in all 12 CSTR's to abiotic and biotic conditions (cca, "vegan" v. 2.4-3, Oksanen *et al*., 2017), with all measured biotic and abiotic parameters, transformed as in 2.6.3, initially used as explanatory variables (K_i = ~51, n = 847) unless otherwise stated (see caption of Fig.4). Automated

bidirectional selection, based on Akaike information criterion (AIC) and Monte Carlo permutation tests (999 permutations), was then used for model simplification and thus identification of the most significant explanatory variables (step, test="perm", "stats" v. v. 3.4.0, R Core Team, 2017). Collinearity was assessed by VIF's (vif.cca, "vegan" v. 2.4-3, Oksanen *et al*., 2017), with variables contributing to VIF's > 3 being removed as in 4.2.6.3 (Zuur *et al*., 2010). Finally the statistical significance of the CCA model (constrained components), its axes and the marginal effects of each explanatory variable were assessed using permutation tests (999 permutations, anova.cca, "vegan" v. 2.4-3, Oksanen *et al*., 2017). Note CCA was chosen over redundancy analysis as unimodal approaches are better suited to relative abundances and the presence of zeros (Ramette, 2007).

Local similarity analysis (LSA, Ruan *et al*., 2006) was utilised to observe correlations between ML virus abundance and the 50 most abundant OTU's in each set of 6 replicate CSTR's. Analysis was undertaken using eLSA v 1.0.2 (Xia *et al*., 2011; 2013) in Python v 2.7, its use over more traditional approaches (e.g. Pearson and Spearman's rank correlation) was justified given it assesses local (short periods of time) and time-delayed associations across replicates, as well as those non-lagged and occurring across the whole sampling period (Xia *et al*., 2011; 2013). A maximum time delay of three was utilised (delayLimit = 3), *P*-values were calculated by permutation tests (1000, *P*-valueMethod = perm), the required precision of *P*-values was set at 1/1000 (precision = 1000), the number of replicates for each time point was set at six (repNum = 6), replicate data was summarised as an absolute deviation weighted median (trasnFunc = MAD) and raw data was rank-normalised and z-transformed (normMethod = robustZ, Ruan *et al*., 2006; Xia *et al*., 2011; 2013). Multiple hypothesis correction was undertaken using *Q*-values (Storey, 2002). The LSA output was then visualised as an association network in Cytoscape v 3.6.0 (Shannon *et al*., 2003), only correlations with a *P*-value ≤ 0.05, a LSA score (LS) ≥ 0.3 and a *Q*-value ≤ 0.05 were examined. Calculation of the peak (max relative abundance) to average (mean relative abundance) ratio (PAR) for examined OTU's allowed assessment of their persistence within the ML, three arbitrarily defined ecological categories facilitated this process; persistent (PAR ≤ 5), intermittent (PAR > 5 < 10) and transient (PAR ≥ 10).

The influence of virus abundance on both α and β diversity was assessed by calculation of spearman's rank correlation coefficients (ρ, cor.test, "stats" v. 3.4.0, R Core Team, 2017) between ML and influent virus abundance at t and D_1 and D_2 diversity indices at t and $\beta_{BC,}$ $\beta_{sor\text{-}tur}$ and $\beta_{sor\text{-}nes}$ between t and t+1 respectively. Its use, over Pearson correlation, was justified since all variable combinations, transformed as in 4.2.6.3, were not bivariate normal (roystonTest, "MVN" v. 4.0.2,

Korkmaz *et al.*, 2014). Bonferroni corrections were applied to all correlations, whilst for correlations involving β diversity sample 102 for ML and influent virus abundance was disregarded.

4.2.6.5. Virus – Community Function Interactions

Virus-community function interactions were assessed by calculation of spearman's rank correlation coefficients between ML and influent virus abundance and ML and effluent COD and NH_4^+-N concentrations respectively, performed, justified and corrected as in 4.2.6.4.

4.3. Results

4.3.1. Functional performance and Abiotic Conditions

4.3.1.1. Acclimatisation (day 0 - 72)

Throughout acclimatisation all CSTR's consistently reduced influent COD_s to compliant levels (< 125 mg L^{-1}, UK discharge consent), with effluent COD_s concentrations not differing significantly amongst CSTR's ($P > 0.05$, Fig. IV.2, Table IV.1, IV.2 and IV.4). Nitrification was less efficient and more unstable (Fig. IV.3), with effluent concentrations of NH_4^+- N, nitrate and nitrite varying temporally, which was significant for the latter, between CSTR's ($P > 0.05$, $P > 0.05$ and $P < 0.05$ respectively, Fig. IV.3, Table IV.1, IV.2 and IV.4). Certainly nitrite accumulation, thus incomplete nitrification, was periodically evident throughout start up (C1, 2 and 3 and T1, 3, 4, 5 and 6), with stable nitrification and compliant effluent (< 5mg L^{-1} ammonia, UK discharge consent), only achieved across all CSTR's on day 52 (Fig. IV.3). Biomass concentrations dropped from seeded concentrations of ~2 g L^{-1} SS, rapidly and consistently in all CSTR's ($P > 0.05$ for SS and VSS), to an operating concentration of ~0.1 g L^{-1} SS (~0.1 g L^{-1} VSS) on day 30 (Table IV.2 and IV.4).

Despite a heavily controlled environment, influent flow rates, thus the HRT/SRT of CSTR's, temporally fluctuated, with significant differences between CSTR's ($P < 0.05$ and $P < 0.05$ respectively, Table IV.2 and IV.4). Subtle temporal shifts in ML temperature and DO concentrations similarly caused significant spatial discrepancies ($P < 0.05$ and $P < 0.05$ respectively, Fig. IV.1, Table IV.2 and IV.4), whilst more pronounced temporal variations in ML pH, which varied from 4.31 – 8.15 across all CSTR's, also resulted in significant differences between CSTR's ($P < 0.05$, Fig. IV.1, Table IV.2 and IV.4). In contrast effluent (ML) phosphate, sulphate, fluoride and chloride concentrations, which were temporally stable, were significantly similar between all CSTR's ($P > 0.05$ for all, Fig. IV.2, Table IV.2 and IV.4).

4.3.1.2. Temperature variation (day 72 - 204)

During this period all CSTR's consistently reduced influent COD_s to compliant levels, although effluent concentrations were significantly different across CSTR's ($P < 0.05$, Fig. IV.2, Table IV.3 and IV.5, seemingly caused by T1). Despite effluent concentrations of NH_4^+-N, nitrate and nitrite varying temporally and, for the former, significantly ($P < 0.05$, $P > 0.05$ and $P > 0.05$ respectively) between CSTR's, nitrification was also generally stable (Fig. IV.3, Table IV.3 and IV.5), although both nitrite accumulation and noncompliant effluent NH_4^+-N concentrations were periodically evident (particularly from day 186 and in T4, Fig. IV.3). Biomass concentrations averaged 0.102 ± 0.070 g L^{-1} SS (0.092 ± 0.063 g L^{-1} VSS) across all CSTR's and were largely temporally stable (Fig. IV.1), though concentrations in C1 were statistically different from those in other CSTR's ($P > 0.05$ for SS and VSS, Table IV.3 and IV.5).

Influent flow rates, and thus the HRT/SRT of CSTR's, temporally fluctuated, with significant differences between CSTR's ($P < 0.05$ and $P < 0.05$ respectively, Table IV.3 and IV.5) like those during acclimatisation. Temporal and significant spatial temperature differences were also found ($P < 0.05$, Fig. IV.1, Table IV.3 and IV.5), though mean temperatures across Control and Test CSTR's were statistically comparable ($P > 0.05$, 14.51 ± 0.195°C and 14.50 ± 4.365°C respectively) due to the mirrored nature of the temperature sinewave. DO concentrations in all CSTR's averaged 7.94 ± 1.36 mg L^{-1} (i.e. above the 4 mg L^{-1} target) but showed temporal and significant spatial discrepancies ($P < 0.05$, Table IV.3 and IV.5). pH similarly varied temporally (Fig. IV.1), with divergences in C3 and C6 causing significant differences between CSTR's ($P < 0.05$, Table IV.3 and IV.5). In contrast, effluent (ML) phosphate, sulphate, fluoride and chloride concentrations were temporally stable and significantly similar across all CSTR's ($P > 0.05$ for all, Fig. IV.2, Table IV.3 and IV.5), like they were during acclimatisation.

4.3.2. Spatiotemporal Dynamics of Virus, Bacterial and AOB Abundances

4.3.2.1. Baseline ML Abundances (day 62 – 72)

ML total virus, bacterial and AOB abundances were determined from day 62 onwards, providing base line abundances for all CSTR's prior to temperature variation (day 72). ML virus abundances decreased from seeded concentrations of 3.41×10^9 viruses mL^{-1} to an average concentration of $1.79 \pm 1.12 \times 10^8$ viruses mL^{-1} on day 62 for all CSTR's, although concentrations ranged from 2.85×10^7 viruses mL^{-1} in T1 to 3.64×10^8 viruses mL^{-1} in C3 (Fig. 4.1). During this 10 day period, ML viruses were temporally dynamic and found to be statistically different between CSTR's ($P < 0.05$, Fig. 4.1, Table IV.7 and IV.8), driven by differences in C5 and T2.

ML bacterial and AOB abundances similarly decreased from seeded concentrations of 4.43×10^7 bacteria mL^{-1} and 1.69×10^6 AOB mL^{-1}, to $4.43 \pm 3.31 \times 10^7$ bacteria mL^{-1} and $1.01 \pm 1.02 \times 10^5$ AOB mL^{-1} across all CSTR's on day 62. No significant differences in either bacterial or AOB abundances were found between CSTR's ($P > 0.05$ and $P > 0.05$ respectively, Fig. 4.1, Table IV.7 and IV.8) despite slight temporal variations across the 10 days.

4.3.2.2. ML Abundances during Temperature Variation (day 72 - 204)

ML virus abundance continued to be temporally and spatially dynamic within and between CSTR's respectively (Fig. 4.1, Table IV.7 and IV.8), averaging $0.94 \pm 0.54 \times 10^8$ viruses mL^{-1} and ranging across an order of magnitude between CSTR's from 1.63×10^7 viruses mL^{-1} in T6 to 3.57×10^8 viruses mL^{-1} in T3 (Fig. 4.1). Consequently significant differences were found between individual CSTR's (Fig. 4.1, Table IV.7 and IV.8), although abundances were statistically comparable between Control and Test reactors ($P > 0.05$).

ML bacterial and AOB abundances were similarly temporally and spatially dynamic, averaging $1.80 \pm 3.07 \times 10^7$ bacteria mL^{-1} and $1.33 \pm 2.53 \times 10^5$ AOB mL^{-1} across all CSTR's (Fig. 4.1, Table IV.7 and IV.8). However significant discrepancies between CSTR's were only found for ML AOB ($P < 0.05$) and not ML bacteria ($P > 0.05$, Fig. 4.1, Table IV.7 and IV.8), whilst no significant difference was found between Control and Test reactors for either community ($P > 0.05$ and $P > 0.05$ respectively).

4.3.3. Spatiotemporal Interactions

4.3.3.1. Virus Interactions with Total Biotic and Abiotic Conditions

Fitted values from the ML virus linear mixed-effects model showed some disparity from observed abundances (Fig. IV.4 A), with explanatory variables accounting for 30% of abundance variations across time and between CSTR's. The model did, however, identify strong positive associations with ML total bacterial abundance ($P < 0.001$), MLSS ($P < 0.001$), ML pH ($P < 0.01$) and effluent chloride concentrations ($P < 0.01$, Table IV.9). In contrast influent calcium ($P < 0.05$) and effluent nitrate ($P < 0.01$) concentrations, as well the HRT of CSTR's ($P < 0.001$), were highly negatively associated (Table IV.9).

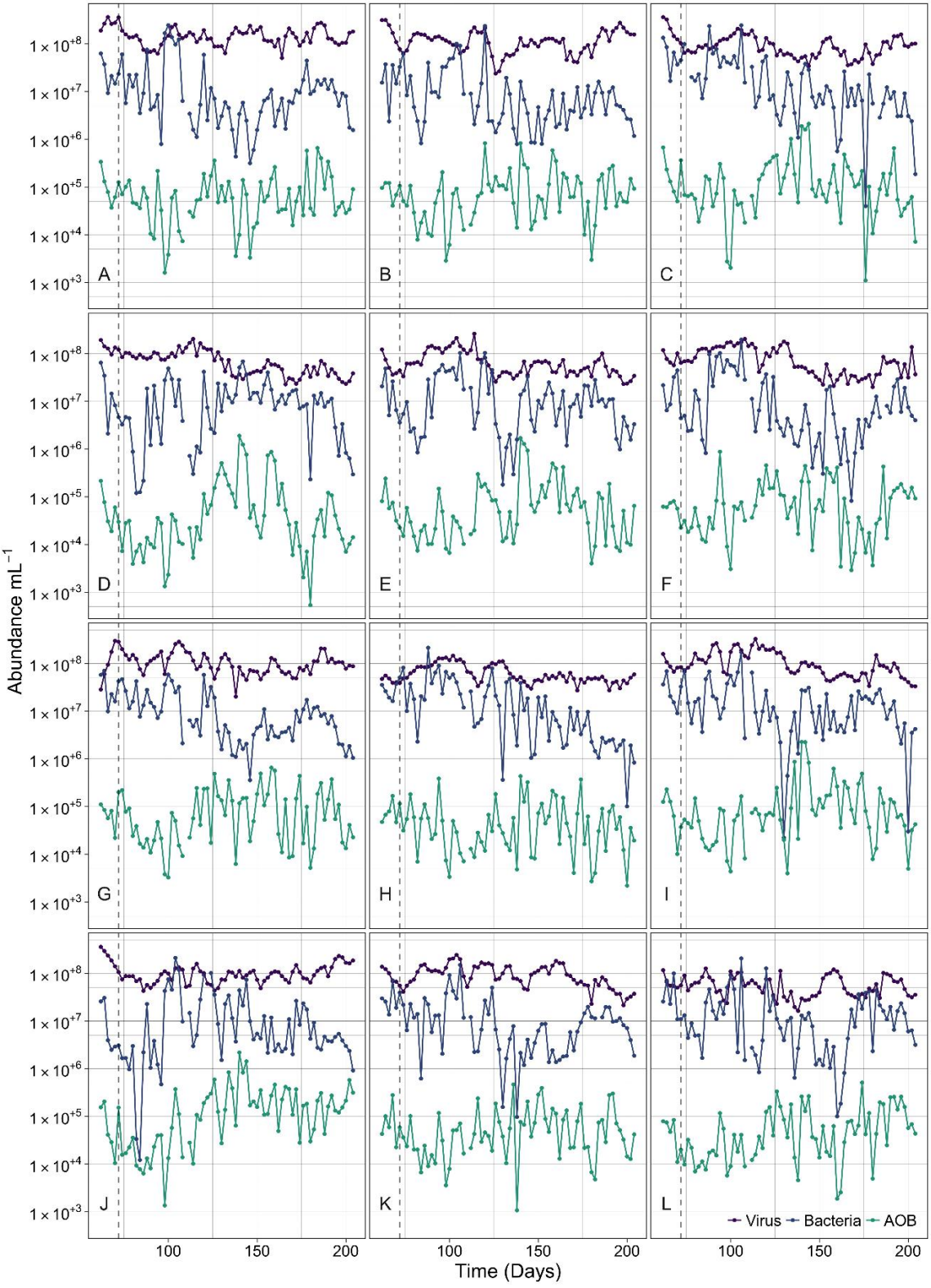

Figure 4. 1. Temporal variation in the biotic conditions within each CSTR over the 204-day study. Dashed grey line represents the end of acclimatisation. (**A** – **F**) Control 1 – 6, (**G** – **L**) Test 1 – 6.

4.3.3.2. Virus Interactions with Bacteria Community Structure

The CCA ordination explained the majority of observed variance in taxa-environment associations (64.5%, Table IV.10). All four axes displayed high taxa-environment correlations (Table IV.10),

indicating strong associations between taxa, viruses and abiotic conditions (Fig. 4.2). ML pH (P < 0.001) and temperature (P < 0.001) were the most significant variables to the ordination followed by influent silicon (P < 0.001) and ML DO (P < 0.001) respectively (Fig. 4.2 and Table IV.10). ML (P < 0.001) and influent (P < 0.001) virus abundance were also highly significant, whilst other notable variables included influent potassium (P < 0.001) and effluent chloride (P < 0.001) and phosphate (P < 0.001, Fig. 4.2 and Table IV.10).

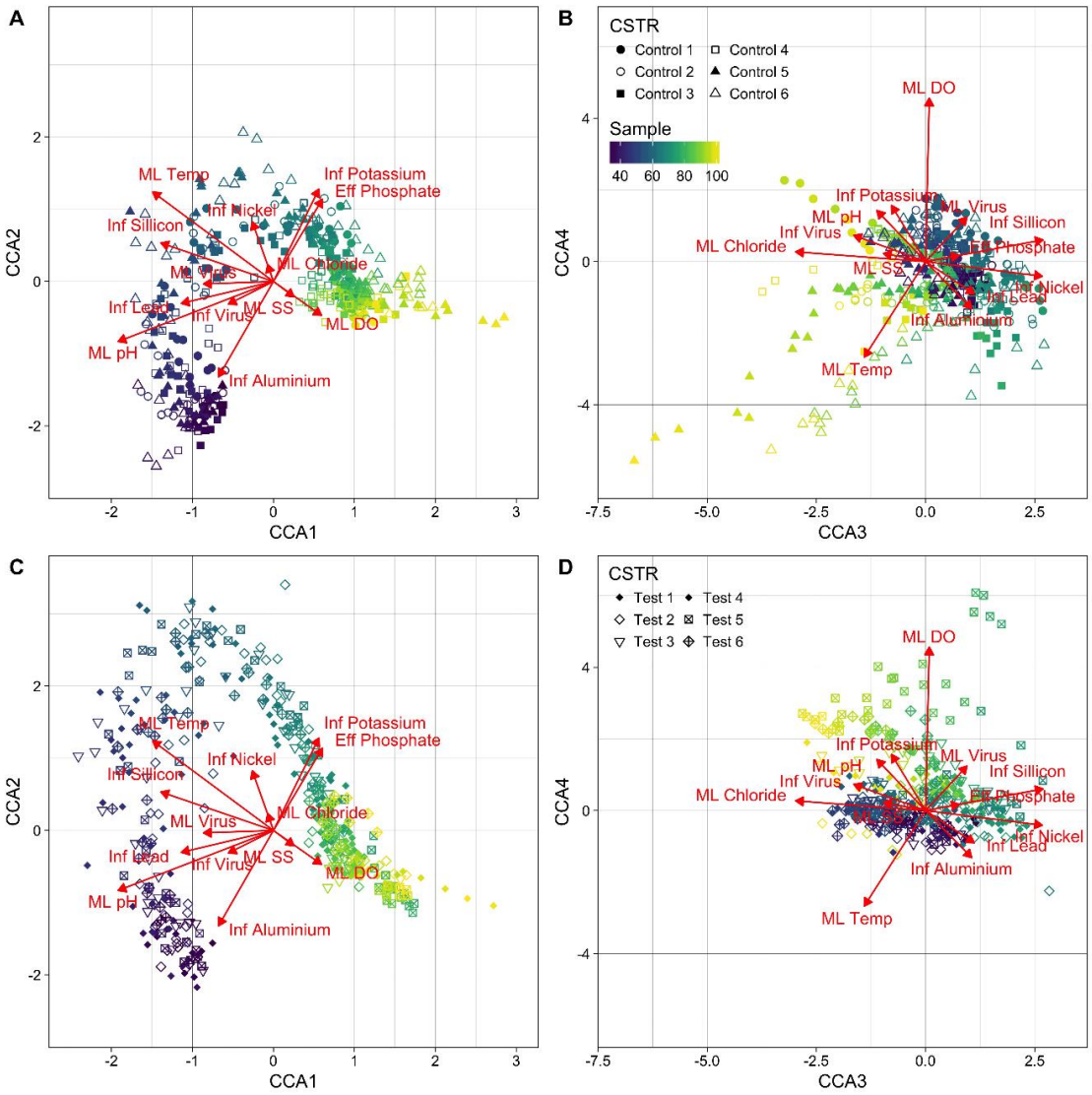

Figure 4. 2. CCA ordination of spatiotemporal ML bacterial community dynamics and associated biotic and abiotic variables within each CSTR (**A** and **B** Controls, **C** and **D** Tests) over the 204-day study, axes 1 and 2 (**A** and **C**) and 3 and 4 (**B** and **D**) respectively. Points and associated numbers represent individual samples, coloured and shaped based on sample and CSTR. Proximity of samples indicates similarities in community composition. Arrows indicate the direction of increase in each explanatory variable, their length indicates the strength of correlation with each axis and the angle between arrows indicates the approximate degree to which explanatory variables are correlated. Inf = influent, Eff = effluent.

ML virus abundance was found to be significantly associated (LS > 0.3, $P \leq 0.05$ and $Q \leq 0.05$) with 8 of the top 50 OTU's in control CSTR's (Fig. 4.3), whilst in test CSTR's 9 of the top 50 were associated. In the controls 2 of these associations had no time lag (Fig. 4.3 A), both were negative

(OTU's 3306 and 1585), 5 had a time delay of 2, 4 positive (OTU's 1585, 4187, 2315 and 8856) and 1 negative (OTU 10048), and 1 had a positive time delay of 3 (OTU 10044). The majority of these OTU's were present intermittently (PAR > 5 < 10) or transiently (PAR ≥ 10) within control CSTR's (Fig. 4.3 A), whilst noteworthy genera included *Nannocytis* (OTU 2315) and *Haliscomenobacter* (OTU 4187). Of the 9 associations identified in the test CSTR's 2 had no time lag, 1 positive (OTU 1303) and 1 negative (OTU 7650), 1 had a positive time delay of 2 (OTU 7705) and 6 had a time delay of 3, 4 positive (OTU's 6072, 9456, 9507 and 10046) and 2 negative (OTU's 8248 and 4004). Again the majority of these OTU's were present intermittently (PAR > 5 < 10) or transiently (PAR ≥ 10, Fig. 4.3 B), whilst noteworthy genera included *Afipia* (OTU 9456), *Aquabacterium* (OTU 7709) and *Phenylobacterium* (OTU 9507). No OTU was found to be significantly associated (LS > 0.3, $P \leq 0.05$ and $Q \leq 0.05$) with ML virus abundance in both control and test CSTR's.

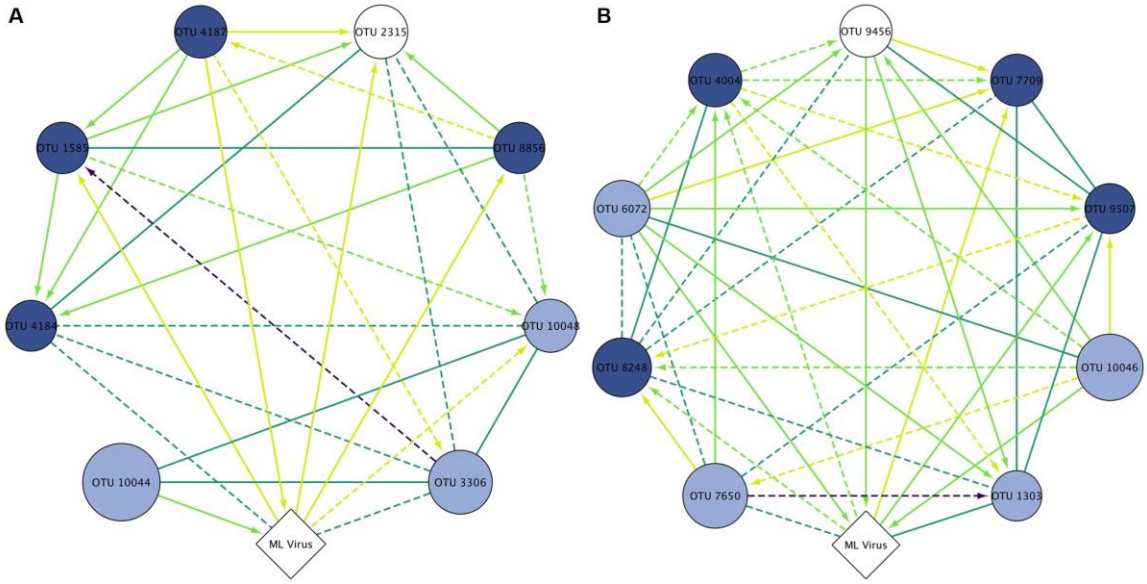

Figure 4. 3. . Network visualisation of LSA associations between all OTU's from the top 50 most abundant that were associated with ML virus abundance (LS ≥ 0.3, P ≤ 0.05 and Q ≤ 0.05) in control (**A**) and test (**B**) CSTR's respectively. Circular nodes indicate bacterial taxa, coloured and sized based on the OTU's PAR (≤ 5 = white, > 5 < 10 = light blue and ≥ 10 = dark blue) and mean relative abundance respectively. For reference OTU 10044 and OTU 4184 had mean relative abundances of 1.56%, the highest, and 0.35%, the lowest. The diamond shaped node indicates ML virus abundance. Solid and dashed lines indicate positive and negative associations respectively, whilst a lines colour indicates the correlations delay (teal = 0, purple = 1, yellow = 2 and green = 3). Arrows point to the delayed, or trailing, node in delayed associations.

Corroboratory correlations were found between ML virus abundance at time t and shifts in β_{BC}, $\beta_{sor\text{-}nes}$ and $\beta_{sor\text{-}tur}$ between t and t+1 ($P < 0.05$, 0.01 and 0.1 respectively, Fig. 4.4 B), whilst ML and influent virus abundance was also highly positively correlated with both Hill's diversity indices at time t ($P < 0.001$ for D_1 and D_2 and $P < 0.001$ and $P < 0.05$ for D_1 and D_2 respectively, Fig. 4.4 A).

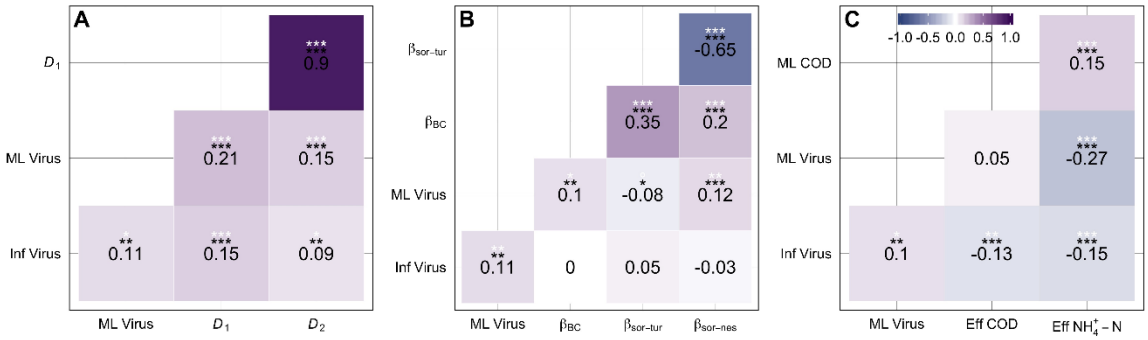

Figure 4. 4. Spearman's rank correlation coefficients describing virus – community structure (**A** and **B**) and virus – community function (**C**) interactions. *n* = 847 (A), 835 (**B**) and 864 (**C**). Black and white ⍰ values are standard and Bonferroni corrected respectively. Royston's Multivariate Normality Test: P > 0.05 for all correlations. Inf = Influent and Eff = effluent. ° P < 0.1, * P < 0.05, ** P < 0.01, *** P < 0.001.

4.3.3.3. Virus Interactions with Community Function

ML virus abundance was highly negatively correlated with effluent NH_4^+- N ($P < 0.001$, Fig. 4.4 C) concentrations, as was influent virus abundance which was also significantly negatively correlated with effluent concentrations of COD ($P < 0.001$ and $P < 0.01$ for NH_4^+- N and COD respectively, Fig. 4.4 C).

4.3.4. Similarity and Synchronicity of CSTR's (day 62 - 204)

Given ML total virus, bacterial and AOB abundances and bacterial community profiles were only determined from day 62, to allow for acclimatisation, CSTR synchrony and similarity was determined from this point forwards.

4.3.4.1. Abiotic Conditions and Function

ML and effluent abiotic conditions, based on ANOSIM and Adonis test statistics, were very weakly different, thus similar, across all CSTR's (ANOSIM $R = 0.084$, $P < 0.001$ and Adonis $R^2 = 0.004$, $P < 0.001$), with no significant differences evident between any CSTR pairs (Table IV.11). Among control (ANOSIM $R = 0.078$, $P < 0.001$ and Adonis $R^2 = 0.028$, $P < 0.001$) and among test (ANOSIM $R = 0.037$, $P < 0.001$ and Adonis $R^2 = 0.006$, $P < 0.05$) CSTR similarity was also high, as was that between control and test CSTR's respectively (ANOSIM $R = 0.060$, $P < 0.001$ and Adonis $R^2 = 0.008$, $P < 0.001$). Correspondingly functional synchrony was evident across all CSTR's, although synchrony was generally higher among control and among test CSTRs (Fig. IV.5 A and Table IV.12, COD_s $\bar{S} = 0.63 \pm 0.10$, $\overline{S_C} = 0.71 \pm 0.06$ and $\overline{S_T} = 0.67 \pm 0.06$, NH_4^+-N $\bar{S} = 0.67 \pm 0.08$, $\overline{S_C} = 0.65 \pm 0.07$ and $\overline{S_T} = 0.73 \pm 0.04$).

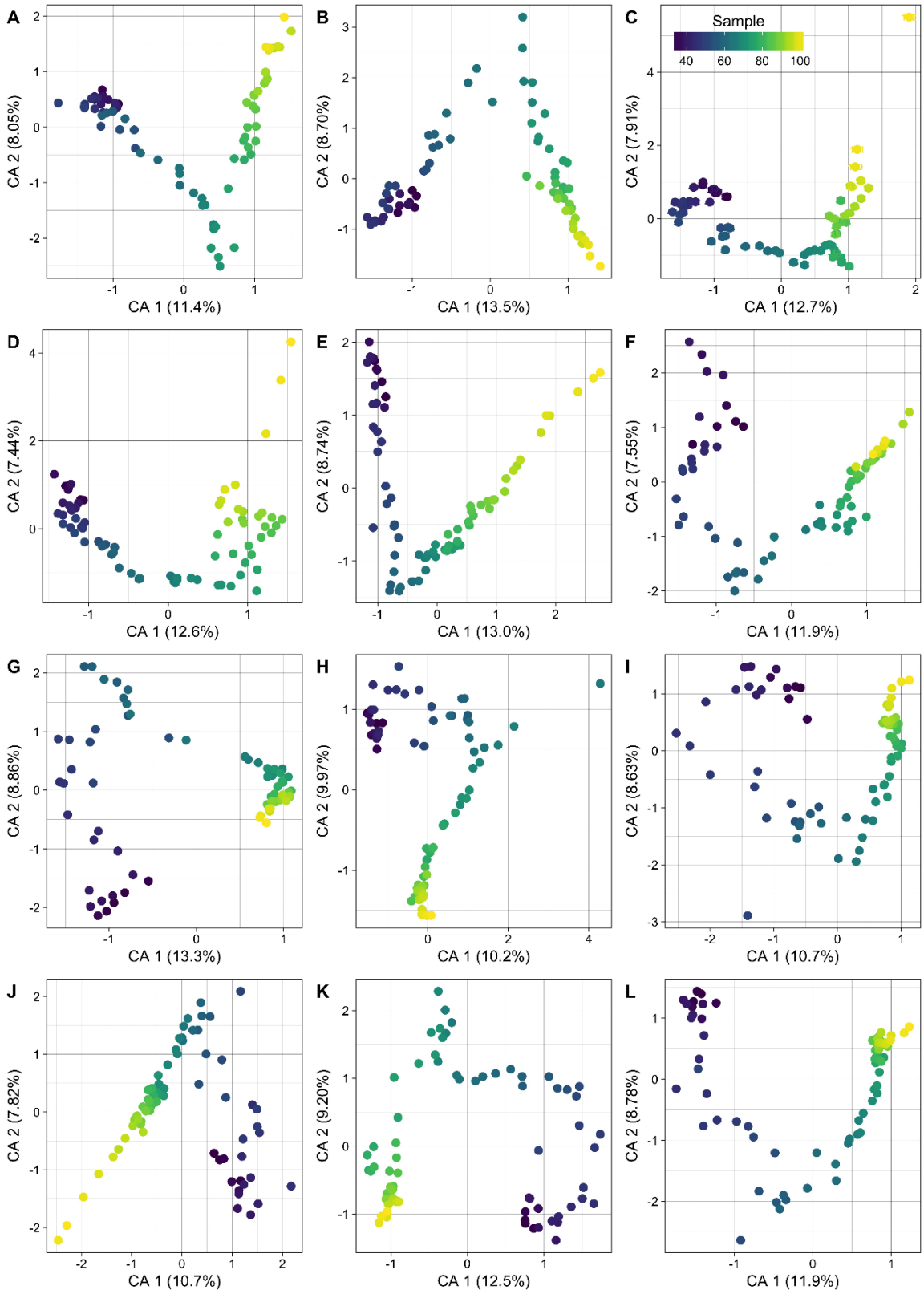

Figure 4. 5. Temporal variation in the microbial communities within each CSTR over the 204-day study, structured by CA. (**A – F**) Control 1 – 6, (**G – L**) Test 1 – 6.

4.3.4.2. ML Total Abundances

Synchrony in total abundance was variable across all three communities within the ML (Fig. IV.5 B, Table IV.13). Viruses, particularly within test CSTR's, appeared predominantly asynchronous, (Fig. IV.5 B, $\bar{S}$ = 0.25 ± 0.26, $\overline{S_C}$ = 0.31 ± 0.24 and $\overline{S_T}$ = 0.16 ± 0.28), though some CSTR pairs, namely C4-C5, C4-C6, C4-T3, C6-T2 and C6-T3, did show some degree of synchrony (Table IV.13). Both total bacteria and AOB were more synchronous across CSTR pairs (Fig. IV.5 B, Table IV.13), with little difference seen between overall synchronicity and that among control and test CSTR's respectively (Fig. IV.4 B, bacteria $\bar{S}$ = 0.48 ± 0.17, $\overline{S_C}$ = 0.50 ± 0.18 and $\overline{S_T}$ = 0.51 ± 0.15, AOB $\bar{S}$ = 0.53 ± 0.10, $\overline{S_C}$ = 0.54 ± 0.12 and $\overline{S_T}$ = 0.54 ± 0.13).

4.3.4.3. Bacterial Community Dynamics

The microbial communities in each CSTR were temporally dynamic and spatially different across all CSTR's (Fig. 4.5, Table IV.14, ANOSIM R = 0.313, P < 0.001 and Adonis R^2 = 0.049, P < 0.001), although compositions were more similar between (ANOSIM R = 0.173, P < 0.001 and Adonis R^2 = 0.043, P < 0.001) and among control (ANOSIM R = 0.213, P < 0.001 and Adonis R^2 = 0.035, P < 0.001) and test CSTR's (ANOSIM R = 0.293, P < 0.001 and Adonis R^2 = 0.026, P < 0.001) respectively than between the majority of CSTR pairs (Table IV.14). Correspondingly temporal bacterial community concordance was variable across CSTR's (Table IV.15), with mean m^2 values of 0.40 ± 0.14, 0.34 ± 0.13 and 0.44 ± 0.13 observed for overall, among controls and among tests respectively. Thus temporal concordance was greater in control CSTR's, although some CSTR pairs, regardless of abiotic conditions, showed greater concordance then others (Table IV.15).

Calculated synchrony coefficients for β_{BC}, $\beta_{sor\text{-}tur}$ and $\beta_{sor\text{-}nes}$ across all CSTR's further emphasised the spatial variability in microbial community shifts (Fig. IV.5 D, Fig. IV.6 and Table IV.16), with all three showing a lack of synchrony ($\beta_{BC}\,\bar{S}$ = 0.21 ±0.16, $\beta_{sor\text{-}tur}\,\bar{S}$ = 0.09 ± 0.13 and $\beta_{sor\text{-}nes}\,\bar{S}$ = 0.06 ± 0.12). Although it is evident that shifts in taxa incidence across all, among control and among test CSTR's was caused by both species turnover (mean ratio $\beta_{sor\text{-}tur}$/ β_{sor} = 0.79, 0.80 and 0.78) and nestedness (mean ratio $\beta_{sor\text{-}nes}$/ β_{sor} = 0.21, 0.20 and 0.22) respectively (Fig. IV.6).

Diversity across all CSTR's was also spatiotemporally variable (Fig. IV.7) and lacked synchrony (D_1 $\bar{S}$ = 0.38 ± 0.16, $D_2\,\bar{S}$ = 0.21 ± 0.19, Fig. IV.5 C and Table IV.17), although synchrony among controls was greater than among test CSTR's ($D_1\,\overline{S_C}$ = 0.51 ± 0.17, $D_2\,\overline{S_C}$ = 0.32 ± 0.16 and $D_1\,\overline{S_T}$ = 0.32 ± 0.14, $D_2\,\overline{S_T}$ = 0.25 ± 0.15, Table IV.17). Unsurprisingly D_1 and D_2 diversity indices, which averaged 158. 82 ± 62.03 and 71.00 ± 34.58 respectively across all CSTR's, were found to be significantly

different between ($P < 0.001$, $D_1 = 173.64$ and 144.04 and $D_2 = 79.38$ and 62.65 for controls and tests respectively) and among control and test CSTR's respectively ($P < 0.001$, Table IV.7 and IV.8).

4.4. Discussion

Efforts to unravel the complex mechanisms underpinning total virus abundance dynamics, i.e. their relationship with microbial abundance and exogenous factors, have, until now, been met in chiefly natural, open systems. We argued in the introduction that engineered systems, where viruses could contribute to functional failures with environmental and financial consequences, offer an ideal test bed for their elucidation, particularly at the lab-scale. Using high frequency, replicated spatiotemporal observations, coupled with statistical inferences, we have demonstrated this is indeed the case. Viruses were linked to bacterial community structure, bacterial diversity, total bacterial abundance and a number of physically credible and previously reported deterministic factors, including ecosystem functions. This is the first finding that such interactions appear over time in multiple lab-scale activated sludge reactors and corroborates previous findings from a full-scale plant (Chapter 3), lending credence to the ubiquitous nature of such relationships.

Spatiotemporal fluctuations in ML virus and total bacterial abundance were statistically concomitant, corroborating previous temporal findings (Chapter 3) and implying coupled virus - bacteria dynamics are homogenous across activated sludge systems. Whilst such a finding makes intuitive sense, since viruses depend on the infection of host cells to produce progeny, the lack of spatiotemporal synchrony in ML virus abundance, particularly when compared to that of their hosts, suggests their distribution is only partially explained by host abundance, as recently alluded to (Wigington *et al.*, 2016).

The greater bacterial community concordance, compositional similarity and diversity among the controls, coupled with more synchronous viral abundances, implies the structure of host assemblages may also play an important role in the development and divergence of viral communities. Certainly the association of ML virus abundance with spatiotemporal variations in bacterial assemblages and differing OTU's among control and test CSTR's supports such inferences. As does the corresponding associations found between select deterministic factors (ML pH, ML SS, ML chloride and influent cations) and both ML virus abundance and bacterial community structure, since fluctuations in such parameters appear to result in concurrent shifts in both host and viral communities respectively. Thus as previously reported (Chapter 3) viruses seem significantly intertwined in the community composition of activated sludge systems, as would be

expected under "Killing the Winner" (KtW, Thingstad, 2000; Winter *et al*., 2010) and/or fluctuating selection dynamics (FSD, Hall *et al*., 2011; Avrani *et al*., 2012).

Through such processes viruses are thought to mediate bacterial competition in a negative density dependant manner, resulting in compositional population fluctuations and greater diversity due to augmented resource partitioning and/or the prevention of competitive exclusion (Avrani *et al*., 2012). The positive association of ML virus abundance with greater taxa abundance variations (β_{BC}) and community diversity (D_1 and D_2 respectively) is thus is in principle agreement with both concepts. Moreover their respective negative and positive associations with incidence based taxa turnover ($\beta_{sor\text{-}tur}$) and nestedness ($\beta_{sor\text{-}nes}$) could be similarly indicative of KtW and/or FSD, given both principles can lessen taxa extinction and dictate a more orderly, or selective, temporal disaggregation of bacterial assemblages (White and Pickett, 1985; Bloch *et al*., 2007; Petsch *et al*., 2015). Indeed virus-host infection networks are believed to be inherently nested in structure (Flores *et al*., 2011; Jover *et al*., 2013, 2016; Korytowski and Smith, 2015), which could help explain observed associations. Furthermore such findings imply viruses could account for, or contribute to, the general lack of observed microbial community concordance. Especially considering the stochastic nature of bacteria-phage coevolution, subsequent local adaptation (Paterson *et al*., 2010) and the fact their abundances were spatiotemporally asynchronous and statistically associated with differing intermittent and transient OTU's among control and test CSTR's.

The association of influent virus abundance with spatiotemporal community shifts, diversity (D_1 and D_2 respectively) and ML virus abundance additionally identifies predator dispersal as a potential source of system divergence, as observed in less complex systems (Brockhurst *et al*., 2007; Morgan *et al*., 2007; Vogwill *et al*., 2008). The influx of new viruses increases population densities, replication rates and the provision of novel genetic variation, increasing the chance of advantageous mutations (Miralles *et al*., 1999; Morgan *et al*., 2005, 2007; Ching *et al*., 2013) that provide viruses an evolutionary advantage in their co-evolutionary arms race (Buckling and Rainey, 2002). Consequent virus (Vos *et al*., 2009; Gomez and Buckling, 2011; Koskella *et al*., 2011) and, in response, host adaptation (Bolotin *et al*., 2004; Kunin *et al*., 2008; Rho *et al*., 2012) can thus give rise to divergent local communities, especially given the stochasticity of dispersal and co-evolutionary trajectories (Buckling and Rainey, 2002; Paterson *et al*., 2010). Interestingly temperature can also alter the rate and type of coevolution (Duncan *et al*., 2016; Gorter *et al*., 2016) and thus, although speculative, could account for the greater community divergence in test CSTR's, presumably by selecting for qualitatively different resistance and infectivity mechanisms at different temperatures (Forde *et al*., 2008; Lopez Pascua *et al*., 2012).

Indeed shifts in ML bacterial community structure were statistically with temperature, as reported previously in full scale activated sludge systems (Wells *et al*., 2009; Griffin and Wells, 2017). Whilst this was not the case for ML virus abundances deterministic factors, namely ML pH, MLSS, effluent nitrate, effluent chloride, influent calcium and HRT, did play a significant role in their proliferation. Such associations corroborate previous findings (Chapter 3) and extend them spatially, yet, with the exception of ML pH and HRT, spatiotemporally they were significantly similar and thus unlikely sources of divergent viral and bacterial communities. Subtle differences in pH and HRT however, given both can influence virus proliferation through the regulation of surface charge (Schaldach *et al*., 2006; Pham *et al*., 2009; Michen and Graule, 2010; Nap *et al*., 2014) and resource provision respectively (Proctor *et al*., 1993; Middelboe, 2000; Weinbauer, 2004), could certainly account for observed virus abundance asynchrony, alone or in combination. Whilst undoubtedly a contributing factor it is striking that differences in both were equally observed across all systems, yet some degree of viral abundance synchrony, although weak, was evident amongst controls. As such observed deterministic triggers may be filtered through, and thus dampened by, bacterial hosts, allowing for greater than expected bacterial community and subsequently viral synchrony.

In contrast to all facets of the microbial assemblages, community function i.e. the removal of COD and NH_4^+-N, was considerably synchronous across all systems, corroborating previous findings that divergent engineered microbial communities can be functionally similar (Kaewpipat and Grady, 2002; Beecroft *et al*., 2012). However, conversely to Brown *et al*. (Chapter 3) ML virus abundance was negatively associated with effluent NH_4^+-N concentrations and not significantly associated with effluent COD concentrations, whilst influent virus abundance was negatively associated with both. It could be argued the positive affect of virus abundance on bacterial community diversity here could increase functional redundancy and resource partitioning (Shapiro and Kushmaro, 2011), thus increasing functional stability and thereby amplifying removal. However the extended HRT of 48 hours here, compared to ~10 hours in Brown *et al*. (Chapter 3), may also mask the true effect of viruses on community function.

Overall the spatiotemporal associations identified between virus abundance and total bacterial abundance, bacterial community composition and deterministic factors corroborate previous temporal findings (Chapter 3). Even those beyond the scope of this paper, those variables not discussed but contributing to the CCA ordination, add weight to our findings, considering significant effects on activated sludge bacterial communities have previously been reported (Wells *et al*., 2009, 2011; Huang *et al*., 2010; Wang et al., 2012a; Griffin and Wells, 2017). Interestingly despite greater sampling frequency the explained variation in virus abundance was far lower than

that explained by Brown *et al*. (Chapter 3). Thus whilst universal deterministic drivers in virus abundance clearly exist, i.e. those abiotic parameters identified in both studies, the spatial heterogeneity of bacterial communities, even when the utmost care is taken in running replicate systems, means virus communities become locally adapted and thus similarly heterogeneous spatiotemporally. Alternatively, despite being the most comprehensive spatiotemporal study of virus and microbial dynamics to date, the number of measured exogenous factors was still insufficient to capture the true behaviour of the systems.

In summary viruses appear an active and dynamic component of activated sludge systems. Whilst differences in abundance were spatiotemporally apparent, thus synchrony was lacking, clear interactions with bacterial community structure were evident, including associations with total host abundance, abundant OTU's, community composition and community-wide diversity metrics. As such viruses could have contributed, or subsequently responded, to observed bacterial community divergences, thus appear to play a more central role in the dynamics of activated sludge systems than hitherto realised. Across all systems virus proliferation was also linked to deterministic factors with plausible and previously reported mechanistic drivers, adding credibility to our findings but also emphasising the important role they play in virus dynamics in activated sludge, and potentially natural, systems. Whilst our results highlight the apparent ubiquitous nature of virus-host interactions in activated sludge systems the identification and enumeration of active viruses and their hosts, although a challenging prospect, is paramount in achieving a greater understanding of such dynamics and the role they play in the microbial ecology of these globally important systems.

4.5. Acknowledgements

This work was carried out as part of a Frontiers Engineering research project (NUFEB, http://research.ncl.ac.uk/nufeb/) awarded and funded by the Engineering and Physical Sciences Research Council (EP/H012133/1). I'd like to thank Northumbrian Water who permitted collection of wastewater samples from the full scale plant, Mary Lunn of Oxford University for guidance on statistical analysis and Joana Baptista and Ben Allen of Newcastle University who undertook sample preparation for illumina sequencing and subsequent bioinformatics analysis respectively.

CHAPTER 5

EVIDENCE OF PREDATOR-PREY DYNAMICS BETWEEN BACTERIOPHAGE AND AMMONIA OXIDISING BACTERIA IN AN ENGINEERED MICROBIAL SYSTEM

The interactions of bacteria and their viruses (bacteriophage) are virtually ubiquitous within microbial communities and lead to population fluctuations in both hosts and virus's that should, in principle, be describable by the Lotka-Volterra equations. Yet such predator-prey dynamics have seldom been observed in nature. A modest reworking of the Lotka-Volterra equations enabled us to explicitly test for signatures of such dynamics between viruses and ammonia oxidising bacteria (AOB) within the mixed liquor of a full-scale activated sludge plant, dynamics which were undetectable by simple observation of raw abundance data. Rates of viral induced AOB mortality, virus replication and AOB growth were subsequently estimated, with viral predation typically killing half the AOB each day. This is the first estimated predation rate of a specific functional group in the environment and implies viruses play a more important role in regulating bacterial abundance in activated sludge systems than previously thought, as well as providing quantitative evidence of their role in the ecology of less abundant functional organisms.

5.1. Introduction

Bacteria-phage interactions are ubiquitous in nature and considered central to the ecology, evolution and functioning of both natural and engineered microbial ecosystems. Indeed phage are thought to mediate bacterial competition in a negative density dependent manner, as described by "Killing the Winner" (KtW, Thingstad, 2000; Winter *et al*., 2010) and/or fluctuating selection dynamics (Hall *et al*., 2011; Avrani *et al*., 2012), and thus influence the diversity within and among bacterial strains, species and communities (e.g. Brockhurst *et al*., 2004; Harcombe and Bull, 2005; Brockhurst *et al*., 2006). Yet quantitative empirical support for these powerful concepts in natural communities, both describable, in principle, by the Lotka-Volterra equations (Volterra, 1926; Lotka, 1934), remains elusive (reviewed by Winter *et al*., 2010; Avrani *et al*., 2012).

It has been argued that strain/genotypic antagonistic coevolution dampens (e.g. Middleboe *et al*., 2009; Rodriguez-Brito *et al*., 2010) or alters (see Cortez and Weitz, 2014) typical Lotka-Volterra signals, however obtaining a quantitative time series of predator and prey abundances needed to observe such dynamics is problematic (see Brum and Sullivan, 2015). Viruses are typically enumerated as total numbers and thus distinguishing one "predator" from another is not feasible (Brown *et al*., 2015) without recourse to a plaque assay (e.g. Middleboe *et al*., 2009; Shapiro *et al*., 2010), which has well documented limitations for bacterial and, consequently, viral quantification (e.g. Rappé and Giovannoni, 2003). Moreover molecular alternatives are of variable precision and accuracy (Baptista *et al*., 2014), typically targeting a specific conserved gene which may not distinguish between bacterial taxa with sufficient resolution. The refinement of KtW host groups

from species to strains (Thingstad *et al*., 2014; 2015) lends weight to these fears and suggests predator prey relationships may be even harder to detect than previously thought.

Nonetheless we sought evidence of Lotka-Volterra dynamics in the mixed liquor of a full-scale activated sludge system using weekly total virus and AOB abundance data obtained by flow cytometry (Brown *et al*., 2015(Chapter 2)) and qPCR (Baptista *el al*., 2014). Whilst there was no obvious predator prey relationship between viral numbers and AOB abundance (Fig. 5.1 A and B), which typically account for less than 3% of all bacteria in activated sludge systems (Coskuner *et al*., 2005), a modest reworking of the Lotka-Volterra equations enabled us to quantitatively test for viral predation and to estimate viral lysis and replication rates respectively.

5.2. Materials and Methods

5.2.1. Sample Collection

Mixed liquor (ML) grab samples were collected from the aeration basin (3600 m^3) of a conventional nitrifying domestic wastewater (6751 m^3 day^{-1}) treatment plant, situated in the North East of England, United Kingdom (UK), on a weekly basis for a period of two years from June 2011 to May 2013 (104 weeks (Chapter 3)). Samples were collected in 50 mL polypropylene containers and transported to the lab on ice for immediate processing.

5.2.2. Flow Cytometry Analysis

For virus enumeration 1 mL sub-samples of ML were taken, transferred into 2 mL cryovials and fixed at a final concentration of 0.5% Glutaraldehyde for 15-30 minutes at 4^{o}C in the dark. Samples were then flash frozen in liquid nitrogen and stored at -80^{o}C. After defrosting samples were pre-treated and analysed in triplicate as described by Brown *et al*. (2015 (Chapter 2)) using a FACScan flow cytometer (Becton Dickinson, USA) equipped with a 15-mW 488-nm air-cooled argon-ion laser and a standard filter setup.

5.2.3. DNA Extraction qPCR

DNA was extracted from 250 µL of ML. Cell wall disruption was carried out using the FastDNA SPIN Kit for soil (MP Biomedicals, USA), thus 244.5 µL of sodium phosphate buffer and 30.5 µL of MT buffer was added to samples and the mixture transferred to Lysing Matrix E tubes. Samples were then lysed at 6.5 ms-1 for 30 seconds in a FastPrep instrument (MP Biomedicals, USA) and centrifuged at 14000 × g for 15 minutes. DNA from 250 µL of the supernatant was then purified using a MagNA Pure LC 2.0 (Roche, UK) and the MagNA Pure LC DNA Isolation Kit III.

Quantification of AOB was carried out using qPCR and amplification of the ammonia monooxygenase (amoA) gene (Baptista *et al*., 2014). Samples were amplified in triplicate on a CFX96 Real-Time PCR Detection System (Bio-Rad, UK) using the primer sets amoA-1F* (Stephen *et al*., 1999) and amoA-2R (Rotthauwe *et al*., 1997). qPCR reactions contained 3 µL of template DNA (sample DNA, standard DNA or molecular grade water (negative control)), 0.5 µL of forward and reverse primer (10 ρmoles per µL), 5 µL of SsoFast EvaGreen supermix (Bio-Rad, UK) and 1 µL of Molecular-grade water. Reaction conditions were: 1 cycle at 98°C for 3 minutes, followed by 40 cycles consisting of 98°C for 5 seconds and 56°C for 5 seconds. Purified circular plasmids containing the target gene were used as standards and run in triplicate for each qPCR reaction. Efficiencies for all qPCR reactions ranged between 90-110% and had a $R^2 \geq 0.99$. Gene copy numbers per unit volume were converted to cell numbers per unit volume assuming each AOB cell contained 2 copies of the amoA gene (McTavish *et al*., 1993; Norton *et al*., 2002).

5.2.4. Statistical Analysis

All statistical analysis was performed in RStudio (v. 1.0.143, R Core Team, 2017) using R version 3.4.0 (R Core team, 2017). Standard major axis (SMA) regression was undertaken using the function lmodel2 in package lmodel2 (v. 1.7-2, Legendre, 2014), with the second difference of natural log transformed response variables ($\Delta^2 \ln(x)$ and $\Delta^2 \ln(V)$), the rate of change of the rate of change (e.g. $(x_{t+1} - x_t) - (x_t - x_{t-1})$), being regressed on the first difference of unlogged explanatory variables (ΔV and Δx), the rate of change (e.g. $V_t - V_{t-1}$). Note the range of ΔV and Δx was adjusted by dividing by 10^9 and 10^8 respectively. Prior to regression analysis response and explanatory variables were checked for outliers, bivariate normality and Pearson correlation using the functions aq.plot, roystonTest and cor.test in packages mvoutlier (v. 2.0.8, Filzmoser and Gschwandtner, 2017), MVN (v. 4.0.2, Korkmaz *et al*., 2014) and stats (v. 3.4.0, R Core Team, 2017) respectively. Note all outliers, based on an alpha value of 0.05, were removed prior to analysis, 0 for $\Delta^2 \ln(x) \sim \Delta V$ and 9 for $\Delta^2 \ln(V) \sim \Delta x$.

5.2.5. Theory

The Lotka-Volterra predator-prey model is given by the equations,

$$\frac{dx}{dt} = \mu x - \varphi x V, \qquad Eq. 5.1$$

$$\frac{dV}{dt} = \delta x V - mV, \qquad Eq. 5.2$$

where, in this case, V is viruses mL^{-1}, x is AOB mL^{-1}, μ is the AOB growth rate and φ, δ and m are the virus lysis, replication and mortality rate respectively.

Rearranging *Eq. 5.1* and *Eq. 5.2* and taking their first derivative by applying the chain rule (see Appendix V, *Eq. V1 – V6*) gives,

$$\frac{d\ln(x)}{dt} = \varphi\left(\frac{\mu}{\varphi} - V\right), \qquad Eq.5.3$$

$$\frac{d\ln(V)}{dt} = -\delta\left(\frac{m}{\delta} - x\right). \qquad Eq.5.4$$

Taking the second derivatives gives,

$$\frac{d^2\ln(x)}{dt^2} = -\varphi\frac{dV}{dt}\ , \qquad Eq.5.5$$

$$\frac{d^2\ln(V)}{dt^2} = \delta\frac{dx}{dt}\ . \qquad Eq.5.6$$

Thus if we have *n* measurements of x and V that are equally spaced in time, x_i and V_i, then we can estimate the first and second derivatives at sampling points using finite difference approximations (see Appendix V): $(\Delta^2\ln(V))_i \approx \frac{d^2\ln(V_i)}{dt^2}$; $(\Delta V)_i \approx \frac{d(V_i)}{dt}$; $(\Delta^2\ln(x))_i \approx \frac{d^2\ln(x_i)}{dt^2}$; $(\Delta x)_i \approx \frac{d(x_i)}{dt}$, $i = 2 \ldots n-1$. If predator-prey dynamics are observed then $(\Delta^2\ln(x))_i$ will be negatively, linearly associated with $(\Delta V)_i$ (*Eq.* 5.5), $(\Delta^2\ln(V))_i$ will be positively, linearly associated with $(\Delta x)_i$(*Eq.* 5.6) and the slopes will approximate - φ and δ respectively. Performing standard major axis (SMA) regression on these variables, which incorporates measurement error, allows assessment of these predictions, which, considering criticism of the Lotka-Volterra models unrealistic assumptions, would hold true for more complex, and arguably more realistic models (Appendix V, section V.1.2).

5.3. Results and Discussion

As predicted, a highly significant, negative association was found between $\Delta^2\ln(x)$ and ΔV, with constant terms approximating to 0 ($P < 0.01$, Table 5.1 and Fig. 5.1 C). The slope implies φ = -2.95 × 10^{-9} week^{-1} or -4.21 × 10^{-10} day^{-1} and proposes a viral induced AOB mortality rate (φV) of -0.504 day^{-1} (Appendix V, section V.1.3, φV_{min} = -0.134 day^{-1}, φV_{max} = -1.436 day^{-1}), this is the first estimated predation rate of a specific functional group from any natural environment. Accordingly viruses kill ~50% of AOB per day, a value consistent with those in marine environments (Noble and Fuhrman, 2000), and thus must be considered a major source of bacterial mortality in activated

sludge systems. Moreover μ was conceivably ~0.6 day^{-1} (*Eq. V10*), a value consistent with AOB μ_{max} values obtained using real wastewaters (0.7 ± 0.4 day^{-1}, Munz *et al.*, 2011). Thus at φV_{max} mortality likely exceeds μ_{max} and could contribute to nitrification failures.

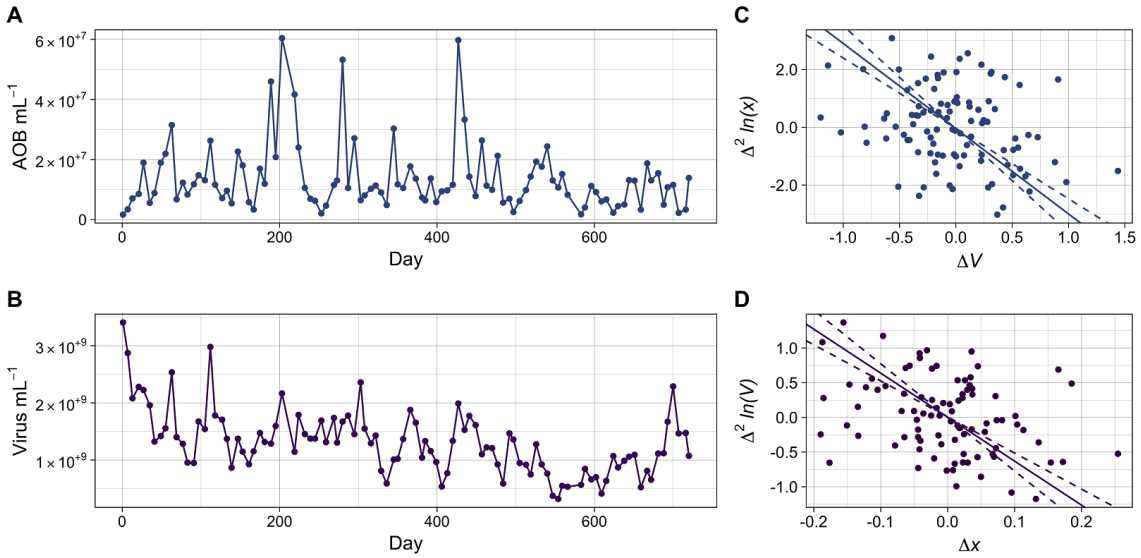

Figure 5. 1. Temporal dynamics of ML AOB (**A**) and Virus (**B**) abundance from the full scale wastewater treatment plant and $\Delta^2 ln(x) \sim \Delta V$ (**C**) $\Delta^2 ln(V) \sim \Delta x$ (**D**). Coloured solid and dashed lines (**C** and **D**) represent slope estimates and their 95% confidence intervals respectively for each SMA regression analysis. Note Δx (**C**) and ΔV (**D**) were rescaled by dividing by 10^9 and 10^8 respectively.

Table 5. 1. SMA regression analysis of differenced abundance data.

Coefficient	Estimate	Estimate 95% CI (min, max)	Slope	Slope 95% CI (min, max)	Angle (°)
$\Delta^2 \ln(x) \sim \Delta V$ [a]					
Intercept	-0.050	-0.062, -0.040	-2.951	-3.575, -2.436	-71.279
$\Delta^2 \ln(V) \sim \Delta x$ [b]					
Intercept	0.003	-0.000, 0.005	-6.338	-7.729, -4.230	-81.034

[a] n = 100. P-Values: 2-tailed = 0.0067, 1-tailed = 0.0033. R^2 = 0.07. Royston's Multivariate Normality Test: P = 0.66. Pearson Correlation Test: Estimate = -0.2696, P = 0.0067. [b] n = 91. P -Values: 2-tailed = 0.0022, 1-tailed = 0.0011. R^2 = 0.10. Royston's Multivariate Normality Test: P = 0.41. Pearson Correlation Test: Estimate = -0.3171, P = 0.0022. CI = Confidence Intervals.

The interpretation of δ was more problematic since the association between $\Delta^2 \ln(V)$ and Δx, though significant with constant terms approximating to 0, was negative, thus had "the wrong sign" (P < 0.01, Table 5.1 and Fig. 5.1 D). A value for δ of -6.34 × 10^{-8} week^{-1} or -9.05 × 10^{-9} day^{-1} implies virus-AOB encounters cause significant viral loss and, since both terms in *Eq. 5.2* are negative, that viruses are in permanent decline. As viruses are not in permanent decline this implies our simple model is too simple. Certainly viruses are also produced by other bacteria at a rate η (days $^{-1}$) and are present in influent wastewater at a concentration of V_{inf} mL^{-1}, which is washed in at the dilution rate (θ days $^{-1}$, reciprocal of the hydraulic retention time). Additionally the negative value of δ could reflect the fact that viral infection is a two-step process, viruses adsorb to and then lyse AOB cells at rate A (mL^{-1} min^{-1}) and φ respectively, producing progeny viruses determined by a burst size (β). Thus *Eq. 5.2* becomes,

$$\frac{dV}{dt} = \theta V_{inf} + \eta V + \beta\varphi xV - AxV - mV\ . \qquad Eq.5.7$$

Therefore

$$\delta = \beta\varphi - A\ , \qquad Eq.5.8$$

and

$$A = \beta\varphi - \delta\ . \qquad Eq.5.9$$

Thus δ will be negative if A is greater than the product of β (100, assuming genome sizes of 2.8 million and 40000 base pairs for AOB and AOB phage respectively (Weitz *et al*., 2015)) and φ (4.21×10^{-10} day^{-1}). Accordingly we estimate A to be 3.3×10^{-8} mL day^{-1} or 2.29×10^{-11} mL min^{-1}, a value well below the theoretical maximum of ~10^{-7} mL min^{-1} (Weitz *et al*., 2015) or values measured under laboratory conditions (Ellis and Delbruck, 1939). This might explain a negative δ and is intuitively reasonable since most of the viruses we measure, and assume to be adsorbing to AOB, will be incapable of infecting that organism.

In summary the data and analysis presented implies AOB and a subset of viruses interact in a predator-prey type manner within activated sludge systems, with viruses significantly affecting AOB net growth. AOB diversity is low in such systems, with one taxon constituting the majority of the AOB biomass. Thus viral predation of the most abundant AOB taxon could account for well-known nitrification instabilities in activated sludge. Moreover such findings lends credence to the notion that bacteriophage may play a significant role in shaping the microbial ecology, function and functional stability within these globally important systems. We caution that these findings are based on statistical associations, of which one had "the wrong sign", and thus their basis is not yet certain. Clearly, deeper mechanistic research is needed to determine the accuracy and details of these statistical observations. This endeavour will be strengthened by the ongoing revolution in viral metagenomics and single cell analysis (Dang and Sullivan, 2014; Brum and Sullivan, 2015).

5.4. Acknowledgements

This work was carried out as part of a Frontiers Engineering research project (NUFEB, http://research.ncl.ac.uk/nufeb/) awarded and funded by the Engineering and Physical Sciences Research Council (EP/H012133/1). I'd like to thank Northumbrian Water who permitted collection of wastewater samples and Mary Lunn of Oxford University for reworking the Lotka-Volterra equations which ultimately enabled me to gain insights into predator prey dynamics within

activated sludge systems. Finally I’d like to thank Ben Allen and Dana Ofiteru for valuable comments.

CHAPTER 6

A WASTEWATER PERSPECTIVE ON VIRAL AND MICROBIAL ABUNDANCES AND VIRUS-MICROBE RATIOS

Recent re-examination of virus and bacterial densities in marine systems has highlighted substantial variation in their consensus 10:1 ratio and identified abundances actually increase non-linearly with each other, i.e. viruses are relatively less abundant at high host densities. Here we compile 1044 bacterial and virus abundance estimates from 14 distinct engineered systems and corroborate these findings. Yet we do so at host and viral densities up to two orders of magnitude greater than previously examined and at high frequency temporal scales. Given these findings we argue engineered systems could be ideal environments for exploring the underlying mechanisms of this emerging paradigm in viral ecology.

6.1 Introduction

In the late 1980's culture independent methods sparked a transformation in viral ecology (Bergh *et al*., 1989) after elucidating virus densities were orders of magnitude greater in aquatic environments than culture based estimates. Subsequently spatiotemporal quantification of virus abundance, alongside microbial hosts, became central to efforts to characterise the scope of viral influences on ecosystems and their functions (reviewed in Wommack and Coldwell, 2000; Weinbauer, 2004). Such studies have yielded great insights into, and underpin, our understanding of virus ecology in marine environments. We now understand that bacterial and viral communities are active, interconnected and critical components of microbial ecosystem production and nutrient recycling (Suttle, 2007; Brussard *et al*., 2008; Rohwer and Thurber, 2009).

A consensus has emerged over the past 25 years: viruses are typically an order of magnitude more abundant than their microbial hosts, approximately 10^7 and 10^6 per mL^{-1} respectively in marine systems (Wommack and Coldwell, 2000; Weinbauer, 2004). Thus, although large variations have been observed, the virus to microbe ratio (VMR) is presumed to equal 10 (Wigington *et al*., 2016), implying a linear association between host and virus abundance. Recently however this consensus has been re-examined. Using complimentary data and analysis, Knowles *et al*. (2016) and Wigington *et al*. (2016) found that VMR's are poorly described by a 10:1, or indeed any fixed, ratio, instead they tend to decrease with host abundance. Although systems with higher microbial densities were found to have correspondingly higher viruses in total, per microbe there were actually fewer.

Wigington *et al*. (2016) propose exogenous factors, host diversity and lysogeny as potential reasons for this observed trend and emphasise increased temporal studies would aid in understanding its implication. Knowles *et al*. (2016) go one step further and offer the "Piggyback-the-Winner" (PtW) hypothesis as a mechanistic explanation, an extension of the commonly

accepted "Killing the Winner" hypothesis (KtW, Thingstad, 2000; Winter *et al*., 2010), which describes virus-host interactions (KtW). They argue, contrary to previous findings (Jiang and Paul, 1994; Maurice *et al*., 2010; Payet and Suttle, 2013; Brum *et al*., 2015b), that temperateness is favoured at high host density, thus more microbes, fewer viruses.

Given the high host (Chapter 3 and 4) and virus (Otawa *et al*., 2007; Wu and Liu, 2009; Brown *et al*., 2015 (Chapter 2, 3 and 4)) densities in activated sludge we sought evidence of such a trend in the influent and mixed liquor of one full- and twelve lab-scale wastewater treatment systems using total virus and bacterial abundance data obtained by flow cytometry (Brown *et al*., 2015 (Chapter 2)) and qPCR respectively.

6.2 Materials and Methods

6.2.1 Sample Collection

Influent and mixed liquor (ML) grab samples were collected from the aeration basin (3600 m^3) of a conventional nitrifying domestic wastewater (6751 m^3 day^{-1}) treatment plant, situated in the North East of England, United Kingdom (UK), on a weekly basis for a period of two years from June 2011 to May 2013 (104 weeks (Chapter 3)). Similarly mixed liquor (ML) grab samples were collected from twelve lab-scale reactors, described in detail elsewhere (Chapter 4), every other day for a period of 142 days. In both instances samples were collected in sterile 50 mL polypropylene containers and transported/stored on ice/at 4°C until further processing. Suspended solids (SS) of all samples was determined according to Standard Methods (APHA, 1989).

6.2.2 Flow Cytometry Analysis

For virus enumeration 1 mL sub-samples of influent and ML were taken, transferred into 2 mL cryovials and fixed at a final concentration of 0.5% Glutaraldehyde for 15-30 minutes at 4°C in the dark. Samples were then flash frozen in liquid nitrogen and stored at -80°C. After defrosting samples were pre-treated and analysed in triplicate as described by Brown *et al*. (2015 (Chapter 2)) using a FACScan flow cytometer (Becton Dickinson, USA) equipped with a 15-mW 488-nm air-cooled argon-ion laser and a standard filter setup.

6.2.3 DNA Extraction qPCR

DNA was extracted from 250 µL of full-scale ML and from 15 mL of influent and lab-scale ML, the latter two being centrifuged at 3392 × *g* for 15 minutes and the supernatant removed down to a working volume of 250 µL. Cell wall disruption was then carried out using the FastDNA SPIN Kit for soil (MP Biomedicals, USA), thus 244.5 µL of sodium phosphate buffer and 30.5 µL of MT buffer

was added to samples and the mixture transferred to Lysing Matrix E tubes. Samples were then lysed at 6.5 ms^{-1} for 30 seconds in a FastPrep instrument (MP Biomedicals, USA) and centrifuged at 14000 × *g* for 15 minutes. DNA from 250 µL of the supernatant was then purified using a MagNA Pure LC 2.0 (Roche, UK) and the MagNA Pure LC DNA Isolation Kit III.

Quantification of total bacteria was carried out using qPCR and amplification of the 16S-rRNA gene. Samples were amplified in triplicate on a CFX96 Real-Time PCR Detection System (Bio-Rad, UK) using the primer sets 338F (Muyzer *et al*., 1993) and 1046R (Huber *et al*., 2007). qPCR reactions contained 3 µL of template DNA (sample DNA, standard DNA or molecular grade water (negative control)), 0.5 µL of forward and reverse primer (10 ρmoles per µL), 5 µl of SsoFast EvaGreen supermix (Bio-Rad, UK) and 1 µl of molecular-grade water. Reaction conditions were: 1 cycle at 98 ^{o}C for 3 minutes, followed by 40 cycles consisting of 98 ^{o}C for 5 seconds and 60 ^{o}C for 5 seconds. Purified circular plasmids containing the target gene were used as standards and run in triplicate for each qPCR reaction. Efficiencies for all qPCR reactions ranged between 90-110% and had a R^2 ≥ 0.99. Gene copy numbers per unit volume were converted to cell numbers per unit volume using accompanying sequence data for the 16S rRNA gene (described in Appendix III).

6.2.4 Statistical Analysis

All statistical analysis was performed in RStudio (v. 1.0.143, R Core Team, 2017) using R version 3.4.0 (R Core team, 2017). Ordinary least squares regression was undertaken using the function lm in package stats (v. 3.4.0, R Core Team, 2017), with $\log_{10}$ virus abundance and $\log_{10}$ VMR's being regressed on $\log_{10}$ bacteria abundance for specific, all lab scale and all systems respectively (the latter only for $\log_{10}$ VMR's). Given the log-log nature of the analysis all models represent possible power-law fits, whereby the slope (β) and intercept (α) of fitted lines denote the power-law exponent best describing the relationship and the logarithmically transformed pre-factor respectively (α_1 and α_0 in Wigington *et al*. (2016)).

6.3 Results and Discussion

At the full-scale wastewater treatment plant, 95% of bacterial abundances ranged from 6.13×10^7 to 1.17×10^9 per mL^{-1} and 4.86×10^5 to 8.09×10^7 per mL^{-1} in the ML and influent respectively, whilst 95% of virus abundances ranged from 5.26×10^8 to 2.33×10^9 per mL^{-1} of ML and 2.97×10^7 to 3.41×10^8 per mL^{-1} of influent (Fig. 6.1 A). Bacteria and virus abundances were also high in the lab-scale systems, with 95% of counts ranging from 3.21×10^5 to 9.98×10^7 per mL^{-1} and 2.65×10^7 to 2.56×10^8 per mL^{-1} respectively (Fig. 6.1 A). Thus host and virus concentrations were generally 10 to 100 times higher than those examined by both Wigington *et al*. (2016) and Knowles

et al. (2016) and, for comparison, 10 to 1000 times higher than those found in marine and freshwater sediments (Table VI.1, Danovaro *et al*., 2008a, b).

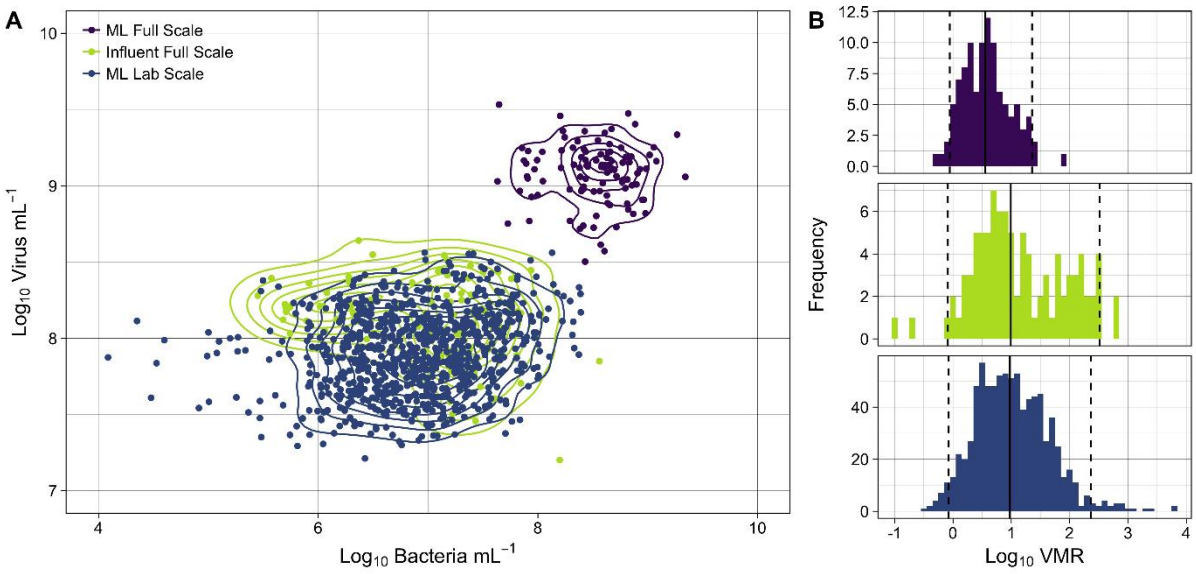

Figure 6. 1. Variation in virus and bacterial abundance (**A**) and VMR's (**B**). Solid and dashed lines (**B**) denote median and central 95% range values respectively for each system.

Median VMR's of 3.59, 9.83 and 9.49 were observed in full-scale ML and influent and lab-scale ML respectively (all twelve systems), thus, like Wigington *et al*. (2016), the consensus 10:1 ratio does accurately represent the median VMR in wastewater environments with lower host and virus densities (Fig. 6.1 B). However substantial variation in VMR's was evident at both full- and lab-scale (Fig. 6.1 B and 6.2 B), with 95% of this variation lying between 0.88 and 22.65, 0.82 and 328.48 and 0.85 and 232.01 in full-scale ML and influent, and lab-scale ML (all twelve systems) respectively. Accordingly at high host densities wastewater VMR's appear more akin to those in marine sediments (Danovaro *et al*., 2008b), whilst at low host densities they transition to those observed in the water column (Wigington *et al*., 2016; Knowles *et al*., 2016; Lara *et al*., 2017).

Virus abundance was significantly, positively associated with bacterial abundance in only 6 of 14 cases when specific systems were individually analysed (Fig. 6.2 A, VI.1 and Table VI.2), with a median β of 0.05 observed. When all lab scale or all systems were considered, both of which resulted in significant, positive relationships, overall β values of 0.08 and 0.31 were observed respectively, thus slopes excluded, and were entirely < 1, in all instances (Fig. 6.2 A and Table VI.2). Such a finding, coupled with the inherent significant decline in VMR's with increasing host density (Fig. 6.2 B and Table VI.3), emphasises that virus abundance increases disproportionately with bacterial abundance in wastewater environments. It also corroborates the findings of Wigington

et al. (2016) and Knowles *et al*. (2016) but additionally confirms the continuation of this relationship at high host and virus densities and across functionally relevant temporal scales.

The findings here are thus in agreement with PtW. Under such dynamics high host densities suppress lytic infection and density-dependent growth due to an increase in the prevalence of lysogeny, reflecting the benefits of lysogeny to hosts (Paul, 2008; Bondy-Denomy *et al*., 2016; van Houte *et al*., 2016; Dedrick *et al*., 2017) but contradicting the consensus that lysogeny is inversely related to host density (Jiang and Paul, 1994; Maurice *et al*., 2010; Payet and Suttle, 2013; Brum *et al*., 2015b). Indeed the hypothesis has been refuted (Weitz *et al*., 2017) and defended (Knowles and Rohwer, 2017) since its proposition and additional, inconclusive, evidence sought (Knowles *et al*., 2017). Yet given peak marine VMR's are evident at ~10^6 bacteria mL^{-1}, as here (Fig. 6.2 B), and host densities rarely breach 10^8 per mL (Wigington *et al*., 2016; Knowles *et al*., 2016; Lara *et al*., 2017) it could be argued marine host densities, and similarly those in freshwater, sediments and soils (Knowles *et al*., 2017), are typically insufficient to trigger community wide lysogeny (Knowles *et al*., 2017), thus lytic infection predominates. In contrast, considering host densities are typically 10^8 – 10^9 per mL and can breach 10^{10} per mL (data not shown), lysogeny may prevail more frequently in an engineered setting.

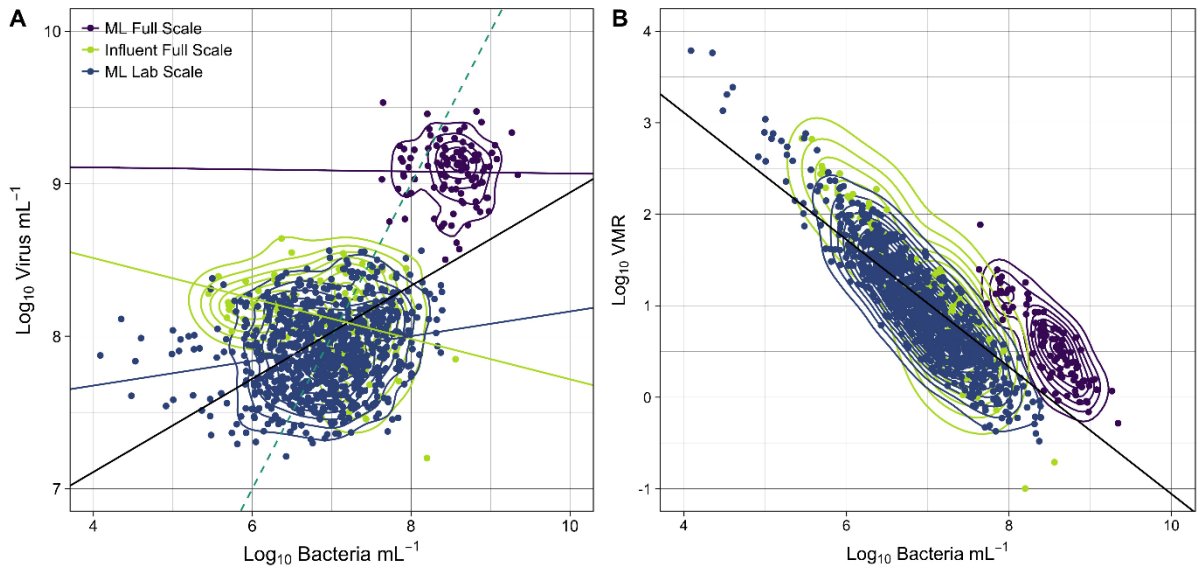

Figure 6. 2. (**A**) Variation in virus and bacterial abundance and VMR's (**B**) as a function of bacterial abundance Solid coloured lines (**A**) denote best fit linear regression for respective systems, black solid lines (**A** and **B**) denote best fit linear regression for all systems and dashed light blue line (**A**) depicts a 10:1 linear relationship. ML = Mixed Liquor.

In light of inconclusive evidence for PtW, Knowles *et al*. (2017) go on to suggest environmental conditions may drive lysogeny, whilst Wigington *et al*. (2016) similarly promote such factors as a potential source of variation in VMR's. Indeed prophage induction through physico-chemical stress has previously be shown in activated sludge systems (Choi *et al*., 2010; Motlagh *et al*., 2015).

Moreover, when environmental and operational variables are utilised to extend the simple models used here, a much larger amount of variation in virus and bacterial abundance is explained, whilst, in general, their association with each other also increases (Chapter 3 and 4). Interestingly some variables, namely concentrations of phosphate, sulphate, sulphur and fluoride, influence the abundance of both communities (Chapter 3). Thus variation in abundances could systematically be driven by such factors individually or mutually, influencing the inferred relationship between virus and bacterial abundances, VMR's and the lytic-lysogenic switch.

In summary the data and analysis presented highlights viruses in wastewater environments are relatively less abundant at high host abundance, thus VMR's decrease with increasing host density. Such findings corroborate those of Wigington *et al*. (2016) and Knowles *et al*. (2016) but also emphasise that this association continues temporally and at greater host densities than previously examined. Whilst no attempt is made to elucidate the mechanism behind such a dynamic the high host density and evidence of lysogeny within engineered systems, coupled with their highly controlled and monitored environments, make them fertile testing grounds for exploring what now appears to be the nascent area of viral ecology.

6.4 Acknowledgements

This work was carried out as part of a Frontiers Engineering research project (NUFEB, http://research.ncl.ac.uk/nufeb/) awarded and funded by the Engineering and Physical Sciences Research Council (EP/H012133/1). I'd like to thank Northumbrian Water who permitted collection of wastewater samples from the full scale plant and Mary Lunn of Oxford University for guidance on statistical analysis.

CHAPTER 7

PRELIMINEARY METAGENOMIC CHARACTERISATION OF WASTEWATER VIRUSES

7.1 Introduction

Viruses play an active role in the ecology of natural environments, influencing the composition, diversity and function (Rodriguez-Valera *et al*., 2009; Winter *et al*., 2010; Breitbart, 2012; Liu *et al*., 2015) of bacterial populations through predation (Wommack and Colwell, 2000), nutrient regeneration (Middelboe and Jørgensen, 2006; Haaber and Middleboe, 2009; Shelford *et al*., 2012) and horizontal gene transfer (Lindell *et al*., 2004; Sullivan *et al*., 2006). Recent examination of their dynamics in biological wastewater treatment systems, where viruses are known to be highly abundant (Otawa *et al*., 2007; Wu and Liu, 2009; Brown *et al*., 2015 (Chapter 2)), would suggest their influence is analogous, with viruses being implicated in host abundance fluctuations and functional instability at both the total abundance (Barr *et al*., 2010; Motlagh *et al*., 2015, Chapter 3, 4 and 5) and strain level (Lee *et al*., 2007; Shapiro *et al*., 2010; Zhang *et al*., 2017) respectively.

Despite the emerging importance of viruses in the dynamics of such systems very little is known about their diversity and function, or how this changes through a typical wastewater treatment stream. Based on genome size viral communities have been shown to vary between different stages of a treatment process (Park *et al*., 2007; Wu and Liu, 2009), between different systems and temporally within the same system (Otawa *et al*., 2007), whilst more contemporary metagenomic approaches have identified highly novel viromes in activated sludge (Parsley *et al*., 2010b; Tamaki *et al*., 2012) and anaerobic digesters (Tamaki *et al*., 2012; Calusinska *et al*., 2016) respectively. Only Tamaki *et al*. (2012) simultaneously characterised influent and effluent viromes and identified over 82% of viral genotypes are shared amongst the different treatment stages, although the relative abundance of known viruses did fluctuate. No study to date however has assessed the presence and diversity of temperate viruses in engineered biological systems.

Viruses can either be classified as lytic or temperate. Upon host infection lytic viruses rapidly replicate culminating in cell lysis, whilst their temperate counterparts enter a symbiosis with host cells, forming a lysogen, and lay dormant as a prophage within host chromosomes until induced, triggering lytic replication and cell lysis. Our understanding of what triggers this lysogenic-lytic switch is, however, poor (Knowles *et al*., 2017). Although recent re-examination of virus and host abundances in marine systems (Wigington *et al*., 2016; Knowles *et al*., 2016), where viruses were found to be relatively less abundant at high host densities, has led to the proposition that temperateness, contrary to previous findings (Jiang and Paul, 1994; Maurice *et al*., 2010; Payet and Suttle, 2013; Brum *et al*., 2015b), is favoured at high host density, the so called "Piggyback-the-Winner" (PtW) hypothesis (Knowles *et al*., 2016). As engineered biological systems harbour

high bacterial numbers (Chapter 3 and 4) and viruses are relatively less abundant at high host densities (Chapter 6) temperate viruses could predominate, particularly considering physico-chemical prophage induction has previously been observed (Choi *et al*., 2010; Motlagh *et al*., 2015).

Consequently using a metagenomic approach we characterised and compared the phylogenetic and functional profiles of free (influent, mixed liquor and effluent) and temperate (mixed liquor only) DNA viral communities throughout the treatment stream of an activated sludge plant, the most frequently used and important form of biological wastewater treatment (Seviour *et al*., 2010). Given previous findings we hypothesised ML and effluent viral communities would differ substantially from those in the influent, giving further indication that viruses interact with, thus have a role in the dynamics of, bacterial hosts and thus activated sludge system performance. Moreover, assuming PtW to be true, we theorised free and temperate ML viral communities would vary in their composition and function.

7.2 Materials and Methods

7.2.1 Sample Collection

Primary settled sewage (influent), mixed liquor (ML) and effluent grab samples were collected from a conventional nitrifying domestic wastewater (6751 m^3 day^{-1}) treatment plant, situated at Tudhoe Mill (~22,500 people), Durham, United Kingdom, on the 1st June, 2016. Samples were collected in 50 mL polypropylene containers and transported to the lab on ice for immediate processing.

7.2.2 Sample Processing

To dislodge viruses and promote floc disaggregation the ionic dispersant sodium pyrophosphate (SP, Sigma, UK) was added to ML at a final concentration of 10mM (Brown *et al*., 2015 (Chapter 2)) and then incubated in the dark for 15 minutes (200 rpm, 37^{o}C). ML was subsequently centrifuged at 3392 × *g* for 5 minutes and the supernatant passed through a 0.2 µM sterile syringe filter, this was made up to a working volume of 2 mL using autoclaved, 0.02 µM filter-sterilised distilled water (ML free viruses). Concurrent samples of influent and effluent were prepared as described for ML, excluding SP addition, incubation and centrifugation (influent and effluent free viruses).

7.2.3 Prophage Induction

The ML pellet was re-suspended in 2 mL of autoclaved, 0.02 µM filter-sterilised distilled water and broken up through continued pipetting. Prophages were then chemically induced using Norfloxacin at a concentration of 1 µg mL^{-1} (Sigma, UK) for 1 hour (200 rpm, 37°C, Tariq *et al*., 2015). Subsequently viral lysates were centrifuged at 3392 × *g* for 5 minutes and the supernatant passed through a 0.2 µM sterile syringe filter, this was then made up to a working volume as in 7.2.2 (ML temperate viruses).

7.2.4 Molecular Methods

7.2.4.1 Viral DNA Isolation

Viral DNA was isolated following the protocol of Tariq *et al*. (2015). Briefly bacterial chromosomal DNA in each sample was attenuated using 2 µL of TURBO DNAse and 2 µL of RNAse Cocktail (Life Technologies Ltd, UK) respectively, prior to incubation at 37°C for 30 minutes and then heat inactivation at 65°C for 15 minutes with 0.5M EDTA (Sigma, UK). Viral DNA was then purified according to the manufacturer's instructions using a NORGEN Phage DNA Isolation Kit (Geneflow Ltd, UK), inclusive of the optional addition of proteinase K (Sigma, UK). Viral DNA was checked for bacterial contamination, of which there was little, using 2% agarose gels following PCR and amplification of the 16S-rRNA gene using primer sets 338F (Muyzer *et al*., 1993) and 1046R (Huber *et al*., 2007). Finally viral DNA was quantified using a Qubit quantitation assay (Life Technologies Ltd, UK) to ensure samples contained enough DNA for sequencing.

7.2.4.2 Illumina Sequencing

The Illumina Nextera XT (Illumina, UK) library preparation kit was used to prepare and multiplex isolated viral DNA for next generation sequencing on an Ilumina MiSeq at the NU-OMICS facility, Northumbria University, UK. A 2 × 250 cycle V2 kit was used for loading and running the DNA samples, which were diluted to 0.2 ng μL^{-1} prior to normalization and pooling. Paired end sequencing reads where provided as FASTQ files and subject to downstream analysis.

7.2.5 Bioinformatics

7.2.5.1 Virome taxonomic and functional profiling

Raw reads were initially assigned taxonomically and functionally by comparison with the NCBI nucleotide (nt) and SEED databases using Kraken (default settings, Wood and Salzberg, 2014) and MG-RAST (default settings, Keegan *et al*., 2016) respectively.

7.2.5.2 Sequence Assembly

Raw reads were filtered for sequences matching the Nextera XT adapters and then quality trimmed using Trimmomatic (v.0.36, Bolger *et al*., 2014), with a quality sliding window of minimum thread score of 25, across 4 consecutive bases, and a minimum read length of 150 bp. Individual viromes were then assembled using SPAdes (v.3.9.1, Bankevich *et al*., 2012), specifying the "--meta" flag to indicate metagenomic assemblies. Assembled contigs were then screened for predicted genes using PROKKA (v.1.12, default settings, Seemann, 2014) and subsequently taxonomically assigned using BLASTP (*E*-value ≤ 10^{-10}, 50% similarity cut-off value) against the NCBI non-redundant (nr) database.

7.2.5.3 Mining Assemblies for Viral-like Genomes

Per-sample assembly graphs were examined in Bandage (Wick *et al*., 2015) providing an initial assessment of the quality of assembled genomes. A novel approach was then utilised to enrich virus-like contigs in each viromes assemblies (see supplementary methods for rationale behind this approach (Appendix VII)). Previously obtained putative proteins were screened for likely virus-specific homologies using HMMER (v. 3.1b2, http://hmmer.org/)) by comparison with Pfam (v. 31) HMM database, which was filtered to only include entries containing one or more of the following descriptions: "integrase", "excisonase", "phage", "virus", "viral" and "capsid". Proteins matching those in the filtered database were then reported and the source contig and subsequent assembly graph node identified. All assembly graphs were then parsed and nodes identified as viral in origin, and all connected nodes, extracted, enabling all viral associated contigs to be amalgamated into collections of individual virus sub-graphs (Fig. VII. 1 B, thought to be genome fragments of specific and/or closely related viruses).

Mapping of raw reads onto all individual virus sub-graph sequences enabled the recovery of "virus only" reads, which were subsequently taxonomically and functionally assigned by comparison with the SEED nr database using MG-RAST (Keegan *et al*., 2016).

7.2.5.4 Preliminary Comparisons of Wastewater Viromes

To determine the similarity of all wastewater viromes, a reciprocal BLAST approach was employed. All genes with 100% match across multiple viromes were considered to be equivalent, the relative abundance of these genes was then determined per virome using raw reads mapped to individual viral sub-graphs. Using these relative abundances the Bray-Curtis (β_{BC}) dissimilarity coefficient was calculated for all possible pairs of samples (beta.pair.abund, "betapart" v.1.4-1, Baselga *et al*., 2017), this was also performed on functionally assigned "viral-only" reads. All statistical analysis

was undertaken in RStudio (v. 1.0.143, R Core Team, 2017) using R version 3.4.0 (R Core Team, 2017).

7.3 Results and Discussion

Due to time constraints the presented results are very preliminary, further work will be needed to answer the hypotheses stated in the introduction (discussed in section 7.4 and Chapter 9).

7.3.1 General Characteristics of the Wastewater Viromes

Across the 4 metagenomic libraries ~ 9 million raw paired-end reads (~ 5 million quality trimmed reads) were generated, ranging from 1.6 – 2.9 (0.8 – 1.8) million depending on a samples origin (Table VII.1). Thus despite low sample volumes (2 mL), which negated the need for concentration procedures and therefore reduced associated bias (John *et al*., 2011; Hurwitz *et al*., 2013), libraries were comparable to, and tenfold greater than, those garnered from anaerobic digesters (50 mL, tangential flow filtration, ultracentrifugation, Nextera XT, Ilumina MiSeq (Calusinska *et al*., 2016)) and a wastewater treatment stream (2 – 10 L, tangential flow filtration, caesium chloride gradient and ultracentrifugation, shotgun 454 (Tamaki *et al*., 2012)) respectively. Of the raw reads on average > 90% could not be taxonomically assigned using the NCBI nt database (Table 7.1), with < 1% being assigned to known viruses. Indeed the majority of characterised sequences were assigned to bacteria (Table 7.1), as reported previously in engineered (Tamaki *et al*., 2012; Calusinska *et al*., 2016) and natural systems (e.g. Angly *et al*., 2006) respectively.

7.3.2 Phylogenetic and Functional Profiles of Nucleotide Sequences

Taxonomic analysis of assigned viral sequences were classified into 18 – 33 viral families depending on a samples origin and processing (Table VII.2). Double-stranded DNA (dsDNA) viruses dominated all wastewater viromes (Table VII.2), with tailed phages of the order *Caudovirales* constituting 71.6% of all identified free viruses and thus corroborating previous findings in engineered systems (Parsley *et al*., 2010b; Cantalupo *et al*., 2011; Tamaki *et al*., 2012; Calusinska *et al*., 2016). Indeed *Caudovirales* represent over 95% of all known dsDNA viruses (Ackermann and Prangishvili, 2012). Within this order the most prevalent family across free viromes was *Myoviridae* followed by *Siphoviridae* and *Podoviridae* respectively, although the relative abundance of these three families did change markedly in the differing wastewater viromes (Table 7.1), as observed previously (Tamaki *et al*., 2012). In contrast the temperate ML virome was dominated by unclassified dsDNA viruses (Table 7.1) of which all belonged to genus *Pandoravirus*, giant viruses known to predate Eukaryotes. Other dominant dsDNA viral families across all viromes were *Polydnaviridae*, *Phycodnaviridae* and *Mimiviridae* (Table 7.1), all of which infect Eukarya

respectively. Indeed 31 of the 37 identified viral family's had eukaryotic hosts, consisting of 19 dsDNA, 3 ssDNA and 9 ssRNA viruses respectively (Table VII.1).

Across all wastewater viromes 5784 "known" viral species were identified from assigned viral sequences, yet only 42, of which 37 were bacteriophages, were present in all viromes (Table VII.3). The most abundant phage species across all free viromes was *Aeromonas* phage vB_AsaM-56 (Table VII.4), given its host is synonymous with aquatic environments, particularly in summer, this is perhaps to be expected (Janda *et al*., 2010). Yet its abundance, along with that of *Enterobacteria* phage T4 and an unclassified Phix174 microvirus, decreased substantially throughout the treatment stream, thus implying respective hosts are inactive within ML (Table VII.4). In contrast *Pseudomonas* virus PaMx74, *Pseudomona*s virus Yua and *Burkholderia* virus Bcep22 more than doubled in the ML and effluent, corroborating previous findings (Tamaki *et al*., 2012) and agreeing with the presence of arguably active hosts in the ML (Wang *et al*., 2012b; Saunders *et al*., 2015; Chapter 3 (data not shown)). Interestingly the ML temperate virome was dominated by Eukarya infecting viral species (Table VII.4), although this was likely skewed by the amount of assigned viral sequences classified as unknown dsDNA viruses (Table 7.1).

Table 7. 1. Summary of metagenomic libraries and the composition of known wastewater viromes determined by similarity to known nucleotide sequences at the family level.

	Wastewater Viromes			
	Inf Free	ML Free	ML Temp	Eff Free
Functional Assignment				
Unknown (%)	48.35	44.56	57.93	36.99
Known (%)	51.42	55.33	41.69	63.00
Phylogenetic Assignment				
Unknown (%)	90	94	94	99
Known (%)	10	6.0	6.0	1.0
Biological Classification				
Virus (%)	0.3	0.1	1.0	0.09
Virus Classification				
Caudovirales (Siphoviridae)	18.83	37.14	1.59	47.83
Caudovirales (Myoviridae)	54.17	27.55	1.88	24.60
Caudovirales (Podoviridae)	4.73	5.56	2.29	12.24
Unclassified Caudovirales	0.02	0.12	0.01	0.12
Phycodnaviridae	0.14	0.97	2.15	0.55
Polydnaviridae	0.17	3.09	14.48	0.20
Mimiviridae	0.19	0.28	0.26	0.16
Unclassified dsDNA viruses	0.00	0.71	57.34	0.73
Other*	21.76	24.56	20.01	13.58
Bacteria (%)	9.0	5.0	4.9	0.9
Archaea (%)	0.4	0.4	0.1	0.01
Eukarya (%)	0.0	0.0	0.0	0.0

Phylogenetic assignment of nucleotide sequences was determined by Kraken (Wood and Salzberg, 2014) using the NCBI NT database. Inf = influent, Eff = effluent, Temp = temperate viruses. *Includes ssDNA, ssRNA and unclassified viruses.

In contrast to phylogenetic assignment functionality was assigned to > 50% of raw sequences across all wastewater viromes (Table 7.1) using a subsystems approach (Overbeek *et al*., 2005). Yet on average < 6% of annotated proteins fell into the category "phages, prophages, transposable elements, plasmids" (Fig. 7.1 A and Table VII.5), with the largest proportion found in the influent

virome. Other abundant categories in this virome were “clustering-based subsystems” (unknown function) “carbohydrates” and “amino acids and derivatives”, whilst abundant categories in the other three viromes included “motility and chemotaxis”, “sulphur metabolism” and “cell wall and capsule” (Fig. 7.1 A and Table VII.5). Thus subtle differences in each viromes functional, and indeed phylogenetic, profiles were apparent.

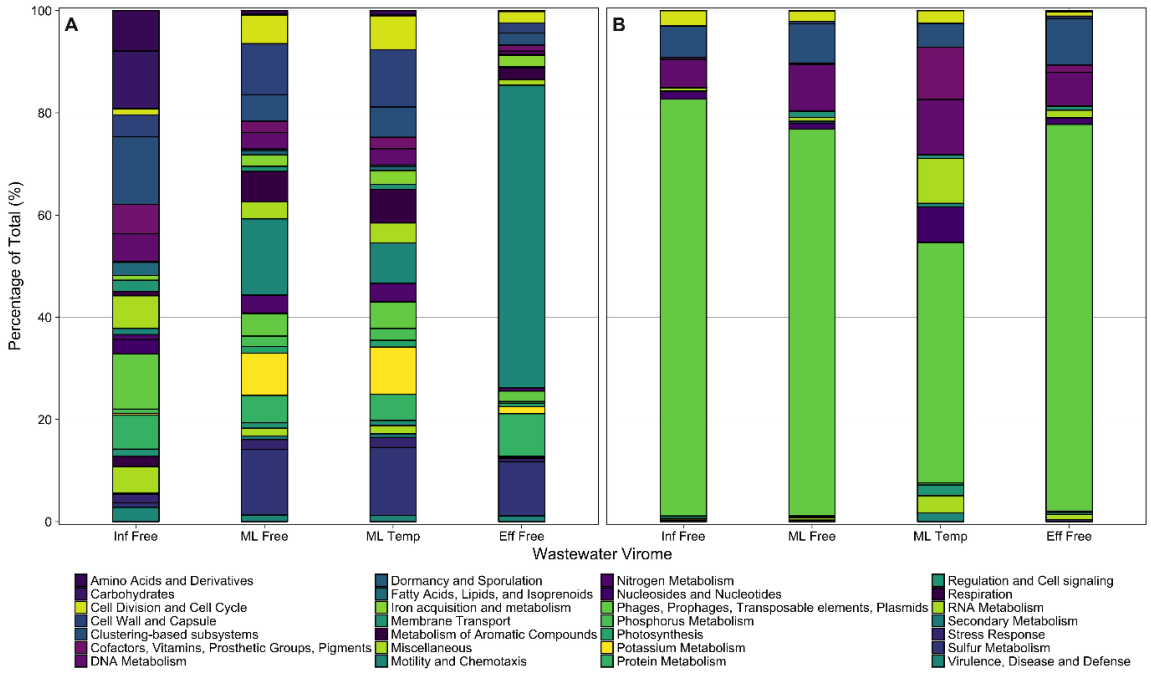

Figure 7. 1. Functional profiles of each wastewater virome determined from raw (**A**) and virus-only (**B**) nucleotide sequences.

7.3.3 Phylogenetic Profiles of Assembled Protein Sequences

Individual assemblies of trimmed reads for each wastewater virome yielded > 45000 contigs and, following extraction of opening reading frames (ORF’s), identification of > 450000 putative genes, of which by far the greatest number were assembled and extracted from reads and contigs in the effluent free virome (Table VII.1). Taxonomic analysis of these protein sequences identified 11 viral families across all wastewater viromes (Table VII.6). Order *Caudovirales* again dominated and constituted 69% of all identified viruses (Table 7.2). Within this order *Myoviridae* and *Siphoviridae* were the most abundant depending on a samples origin, followed by *Podoviridae* respectively. When compared to abundances garnered from nucleotide sequences subtle differences in each family’s relative contribution was apparent in the free viromes, yet in the temperate ML virome, particularly for *Myoviridae*, their contributions increased greatly (Table 7.1 and 7.2). Other classified viral families were all nucleocytoplasmic large DNA viruses (NCLDV) from the proposed order *Megavirales* (Colson *et al.*, 2013) and included *Phycodnaviridae*, *Mimiviridae* and *Marseilleviridae* (Table VII.6). Notably the giant viruses *Klosneuvirus*, *Indivirus*, *Catovirus* and

Hokovirus from *Mimiviridae*'s subfamily *Klosneuvirinae*, originally discovered in ML metagenomic data (Schulz *et al*., 2017), were identified in ML free and temp and effluent free viromes respectively. Across all viromes 25% of protein sequences identified as viral in origin could not be classified to existing viral orders or families.

7.3.4 Mined Viral-like Genomes

From the > 450000 identified protein sequences > 35000 were identified as homologous viral proteins across all wastewater viromes, the vast majority of which were present in the effluent free virome (Table VII.I). In total this enabled 284 (> 20k bp) individual virus sub-graphs to be extracted, putatively representing fragments of (249), and complete (35), viral genomes respectively (Table VII.1). Subsequent functional assignment of garnered "virus-only" reads highlighted that the majority were known (> 75%) and, as expected, fell into the category "phages, prophages, transposable elements, plasmids" across all viromes (> 68%, Fig. 7.1 B and Table VII.1). Taxonomic assignment was less successful, classifying < 5% of "virus-only" reads from all viromes as known viruses. Thus whilst the innovative approach described appears to extract viral-like genomes successfully the majority are uncharacterised in extant databases, highlighting, as alluded to previously (Parsley *et al*., 2010b; Cantalupo *et al*., 2011; Tamaki *et al*., 2012; Calusinska *et al*., 2016), the novel diversity in wastewater viromes.

Table 7. 2. Summary of the composition of wastewater viromes determined by similarity to known protein sequences at the family level.

	Wastewater Virome			
	Inf Free	ML Free	ML Temp	Eff Free
Phylogenetic assignment				
Unknown (%)	90.7	91.5	92.4	92.2
Known (%)	9.3	8.5	7.6	7.8
Biological Classification				
Virus (%)	22	14	40	17
Virus Classification				
Caudovirales (Siphoviridae)	36.97	34.52	7.28	20.19
Caudovirales (Myoviridae)	26.26	14.17	55.93	39.08
Caudovirales (Podoviridae)	14.21	13.76	3.26	8.85
Unclassified Caudovirales	0.00	0.00	0.00	0.00
Phycodnaviridae	5.30	7.47	1.23	1.56
Mimiviridae	0.00	0.18	0.00	0.19
Unclassified dsDNA viruses	0.00	0.00	0.00	0.00
Other	17.25	29.91	32.29	30.13
Bacteria (%)	77	85	59	83
Archaea (%)	0.2	0.6	1	0.6
Eukarya (%)	0.1	0.5	0.01	0.3

Phylogenetic assignment of protein sequences was determined using BLASTP against the NCBI nr database. Inf = influent, Eff = effluent, Temp = temperate viruses.

7.3.5 Comparison of Wastewater Viromes

β_{BC} coefficients generated for the relative abundance of genes across viromes indicated that ML and effluent free viromes were most similar, which was also apparent when functionally assigned

"viral-only" reads were considered (Table VII. 8). Influent free and ML temp were the next most similar in both instances, whilst both of these viromes were evidently dissimilar to both the ML and effluent free viromes respectively.

7.4 Preliminary Conclusions and Recommendations

Given the incomplete analysis presented drawing any valid conclusions is problematic. However we have shown that sufficient reads to assemble viral metagenomic libraries can be garnered from only 2 mL of wastewater, precluding the need for concentration procedures which can introduce bias. Across all wastewater viromes, whether determined from raw reads or predicted protein sequences, dsDNA viruses of the order *Caudovirales* were found to be dominant, with *Myoviridae*, *Siphoviridae* and *Podoviridae* being the most abundant families. The relative abundances of these families, as well as known viral genera and species, was found to change markedly throughout the wastewater treatment stream, as well as between free and temperate viromes respectively. Yet, given that < 10% of raw and/or predicted protein sequences could be taxonomically assigned, such conclusions should be taken with caution. Equally whilst β_{BC} coefficients identified clear similarities and differences between wastewater viromes greater predicted genes and "viral-only" sequences were found in the effluent and ML free viromes, thus likely biasing results. Although this could be indicative of greater and different viruses in these two viromes, as indicated by abundances (Chapter 3).

Accordingly to fully assess the phylogenetic and functional viral diversity throughout a wastewater treatment stream, and how it changes, assessment of the uncharacterised reads, that is > 90% of those generated, is required. The novel approach adopted here, which seemingly was able to identify and extract 284 novel viral-like genomes, offers promise in this regard, yet requires refinement (discussed in Chapter 9). In hindsight greater characterisation of these reads could have been achieved through co-assembly, wherein reads from all samples are combined and assembled simultaneously. Such an approach would likely increase our ability to extract individual virus sub-graphs, thus identify more novel viral-like genomes. Moreover the relative abundance of these viral-like genomes in each sample could be obtained, thus whilst they would remain uncharacterised their dynamics throughout a treatment stream could be described. Estimations of α and β diversity, using CatchAll (Allen *et al*., 2013) for example, would also be possible. Undoubtedly such analysis would thus improve upon the initial characterisation of wastewater viruses presented here.

7.5 Acknowledgements

This work was carried out as part of a Frontiers Engineering research project (NUFEB, http://research.ncl.ac.uk/nufeb/) awarded and funded by the Engineering and Physical Sciences Research Council (EP/H012133/1). I'd like to thank Northumbrian Water who permitted collection of wastewater samples and Ben Allen of Newcastle University who undertook incorporated bioinformatics analysis.

CHAPTER 8

GENERAL DISCUSSION AND CONCLUDING REMARKS

8.1 Rationale

Climate change, population growth and increasingly strict environmental regulation means the global water industry is currently facing an unprecedented coincidence of challenges (Palmer, 2010). Better microbial ecology could significantly contribute, since explicitly engineering and maintaining efficient and functionally stable microbial communities would allow existing assets to be optimised and their robustness improved. Indeed contemporary wastewater microbiology has elucidated both niche and neutral components as factors in microbial community shifts and functional failures in wastewater treatment bioreactors, however both mechanisms inadequately explain all of such events. Given its role in natural systems viral infection could be another important factor, yet viruses have seldom been monitored in engineered systems. Here we attempted to address this lacuna, particularly within activated sludge systems.

8.2 Synopsis of results

Culture independent methods sparked a transformation in viral ecology in the late 1980s (Bergh *et al*., 1989) and 1990's (Hara *et al*., 1991; Marie *et al*., 1999) after revealing virus densities in aquatic environments were orders of magnitude greater than culture based estimates. Subsequently their quantification, alongside microbial cells, became central in efforts to characterise the scope of viral influences on ecosystems and their functions (reviewed in Wommack and Coldwell, 2000; Weinbauer, 2004). Contrary to natural systems however such methods found limited application in an engineered setting: EFM being used spatially to compare different activated sludge plants (Wu and Liu, 2009) or across very modest temporal scales (Otawa *et al*., 2007). Thus to facilitate our endeavours an FCM protocol, the method of choice in marine systems (Brussaard *et al*., 2010), was initially developed, optimised and validated, allowing rapid (relative to other methods), accurate and highly reproducible quantification of total free viruses in activated sludge samples (Chapter 2). Using the FCM protocol viruses were found to be highly abundant across 25 activated sludge plants, with concentrations ranging from 0.59 - 5.14 × 10^9 viruses mL^{-1} corroborating previous findings (Otawa *et al*., 2007; Wu and Liu, 2009) and highlighting virus abundance in activated sludge is amongst the highest of all systems studied to date (Wommack and Coldwell, 2000; Weinbauer, 2004).

The optimised protocol was subsequently utilised to assess virus abundance temporally (Chapter 3) and spatiotemporally (Chapter 4) in full- and replicate lab-scale activated sludge systems respectively, the first two comprehensive studies of virus abundance dynamics in activated sludge systems. Whilst principally mimicking studies in marine (e.g. Jiang and Paul, 1994; Weinbauer *et*

al., 1995; Bratbak *et al*., 1996; Li and Dickie, 2001) and freshwater (e.g. Hennes and Simon, 1995; Hofer and Sommaruga, 2001; Bettarel *et al*., 2004; Jacquet *et al*., 2005) environments the two studies were undertaken at functionally relevant temporal scales and incorporated a suite of time varying exogenous factors, answering recent calls in viral ecology (Breitbart, 2012; Brum and Sullivan, 2015; Wigington *et al*., 2016).

At the full-scale (Chapter 3) viruses in the mixed liquor (ML) were shown to be both abundant and temporally dynamic, a finding corroborated and extended spatially at the lab-scale (Chapter 3). Thus viruses appear both temporally and spatially dynamic, yet ever present across activated sludge systems. Accordingly, given viruses are obligate parasites, ML virus abundance was positively associated with total host (bacterial) abundance at both full- and lab-scales, i.e. temporally and spatiotemporally respectively. Such findings are a first in an engineered setting and imply coupled virus-bacteria dynamics are homogenous across activated sludge systems, yet without the measurement of accompanying exogenous factors they may have remained undetected (Chapter 6). Indeed the relative decline in virus abundance with increasing host density (Chapter 6) across wastewater systems corroborates recent re-examination of marine datasets (Wigington *et al*., 2016; Knowles *et al*., 2016) and may imply lysogeny is the dominant infection strategy in activated sludge systems, although clarification through further work is needed (see Chapter 9).

ML virus abundance, at both full- and lab-scales respectively, was also linked spatiotemporally to microbial community structure, another novel finding which implies viruses may play a significant role in shaping, or are influenced by, activated sludge bacterial dynamics, as predicted by KtW and/or FSD. Indeed numerous lines of evidence exist for the presence of these powerful concepts in activated sludge systems. At the full-scale the corresponding associations found between resource availability (NH_4^+- N, phosphate and sulphur concentrations) and both ML virus abundance and bacterial community structure is in principle agreement (Winter *et al*., 2010; Avrani *et al*., 2012; Pascua *et al*., 2014), whilst at the lab-scale virus abundance was positively associated with host α and β diversity respectively. However evidence of predator-prey (Lotka-Volterra) dynamics between a subset of measured viruses and a key functional group (AOB) at the full-scale, dynamics which fundamentally underpin both concepts, is the most significant and principally important (Chapter 5). The estimated predation rate, akin to those from marine environments (Noble and Fuhrman, 2000), suggests viruses kill ~50% of AOB per day, thus if corroboratory evidence can be found (discussed in Chapter 9) such a finding might single handily change the way viruses are viewed in an engineered setting.

Equally the significant associations found between virus abundance and community function (effluent concentrations of COD and NH_4^+-N), although contradictory results were found at full and lab scale respectively, suggests viruses may play a role in the performance of activated sludge plants, thus viruses warrant greater exploration in this regard (discussed in Chapter 9).

Exogenous factors clearly played an important role in virus (Chapter 3 and 4) and host (Chapter 3) abundance and host community dynamics (Chapter 3 and 4), vindicating recent calls in viral ecology (Wigington *et al*., 2016). ML virus abundance across both full and lab scales was heavily influenced by pH and cation concentrations, thus environmental variables involved in the regulation of surface charge and electrostatic interactions, i.e. adsorption processes, are evidently central to virus proliferation in activated sludge, and potentially natural, systems. Nutrients availability, as in marine systems (Hewson *et al*., 2003; Øvreås *et al*., 2003; Williamson and Paul, 2004; Motegi and Nagata, 2007; Sandaa *et al*., 2009), was also evidently important to ML virus abundance, shown directly at the full scale (NH_4^+- N, phosphate and sulphur concentrations) and indirectly, since the HRT of CSTR's controls its provision, at the lab scale. The corresponding associations found between certain deterministic parameters and ML virus and bacterial abundance (Chapter 3) and ML community composition (Chapter 3 and 4) additional supports they are coupled, as aforementioned, but also suggests deterministic triggers may be filtered through bacterial hosts and that variation in densities could systematically be driven by such factors. Interestingly nitrate concentrations at both full and lab scale were negatively associated with virus abundance, the lack of a known mechanistic reason for this warrants further study (Chapter 9).

Finally wastewater viromes are evidently largely phylogenetically and functionally uncharacterised, although relative abundances of known dominant families and species vary throughout the wastewater treatment stream and between free and temperate viromes (Chapter 7). Thus initial analysis implies free viruses in the ML and effluent differ to those in the influent and to temperate ML viruses, although future work is warranted to substantiate such claims. Refinement of the unique method introduced here, which identified 284 novel viral-like genomes, would certainly aid in this regard (Chapter 7).

Overall the various studies incorporated here emphasise viruses appear spatiotemporally active and dynamic and potentially highly diverse in activated sludge, interacting with hosts at a number of levels and influencing the performance of these globally important systems. Certainly such findings substantially contribute to our understanding of virus dynamics in engineered and natural systems. Thus further exploration of their dynamics in such systems is warranted (Chapter 9).

8.3 Caveats

The goal of this body of work was to ascertain the potential role virus's play in the microbial ecology and function of engineered, principally activated sludge, systems, given their importance in natural environments. Preferably one would assess individual phage-host interactions in answering this question, however functional organisms of interest, thus their associated viruses, are typically unculturable, rendering culture dependent methods, i.e. plaque assays, unusable (discussed in Chapter 1). Culture independent methods are emerging yet their use in real world situations is currently limited (discussed in Chapter 1), thus the total abundance methods which underpin our understanding of viral ecology in marine systems seemed an obvious starting point, especially considering their limited application in an engineered setting. However, whilst the analysis undertaken elucidated some interesting biotic and abiotic associations, ultimately the analysis is at too coarser level to gain a true picture of virus-host/virus-community function interactions, particularly since viruses are now thought to influence bacterial communities at the strain level (Rodriguez-Brito *et al*., 2010; Thingstad *et al*., 2014; 2015).

The complex nature of the microbial communities inherent to these systems, or in any system for that matter, also meant we were reliant on statistical inference to summarise and describe observed spatiotemporal patterns in and between total abundances, community composition and ecosystem functions. This has inherent problems. Firstly interactions may be asymmetric yet are depicted as symmetric, i.e. A influences B negatively but B influences A positively, yet the correlation of A with B and B with A is identical. Secondly the direction of interaction may fluctuate yet correlation requires constancy, i.e. A influences B positively but in certain situations A influences B negatively. Consequently certain identified associations maybe spurious, whilst others may have been missed. Moreover whilst we tried to back up detected associations with literary evidence and thus partially substantiate perceived causal relationships we cannot guarantee this to be the case, accordingly all relationships should be viewed as a foundation for future work exploring the true underlying ecological processes (Chapter 9).

Finally the methods utilised in this study are not error-free. Quantitative real-time PCR (qPCR), although highly precise and sensitive across an extensive linear range (Suzuki *et al*., 2000; Klein, 2002), suffers from many of the biases associated with PCR; only measuring gene copy number (not cell numbers) and its dependence on factors other than the amplification reaction e.g. sample preparation, DNA extraction, standard quality, choice of target gene and amplification primers and probes (Klein, 2002; Yu *et al*., 2005; Lim *et al*., 2008). Thus whilst all efforts were made to reduce these errors and accurately convert gene copy numbers per mL to bacteria per mL

respectively (described in Appendix III), total bacterial and AOB counts may have suffered from bias. Although the method adopted for AOB quantification has been comprehensively validated (Baptista *et al*., 2014). Likewise, since PCR was used in their generation, the microbial community data produced using 16s rRNA sequencing may have also suffered. There is also uncertainty errors associated with the determination of physicochemical environmental parameters. Overall however the errors in this study are not perceived to be any greater than those experienced in similar studies.

8.4 Conclusions

- FCM is suitable for quantifying total virus abundance in activated sludge samples when accompanied by appropriate sample pre-treatment.
- Viruses in the ML were highly abundant and temporally dynamic across activated sludge systems.
- ML Viruses were positively associated with total bacterial abundance spatiotemporally across activated sludge systems.
- Viruses were associated with shifts in the composition of ML bacterial community's and the dynamics of specific, highly abundant OTU's.
- ML Viruses were spatiotemporally associated with abiotic parameters with underlying mechanistic reasons.
- Influent, ML and effluent Viruses were spatiotemporally associated with system function.
- Evidence of predator-prey dynamics was evident between a subset of viruses and AOB within the ML of a full scale activated sludge plant.
- ML viruses were relatively less abundant at high host densities, following the emerging paradigm in viral ecology.
- The majority of viruses are phylogenetically and functionally uncharacterised in influent, ML and effluent wastewaters, thus have novel diversity. Yet known viral families and species evidently differ throughout the treatment stream, with potentially greater diversity apparent in ML and effluent viromes.

CHAPTER 9

FUTURE WORK

Demonstration that viruses are abundant, spatially and temporally dynamic and statistically associated with bacteria at the total abundance, community, functional and OTU level has led to the conclusion that viruses may play a major role in the dynamics of activated sludge systems, particularly considering they were also implicated in a systems functional stability. In establishing such a role we have proposed a number of hypotheses and generated a vast amount of data which could form the basis of further work. Moreover, and perhaps more importantly given the recent explosion of novel tools, technologies and theories, we have access to 1000's of frozen (-80^{o}C), archived samples from both the full (Chapter 3) and lab scale studies (Chapter 4), of which the former now extends to 5 years of weekly samples.

Given the coarse nature of the analysis undertaken, i.e. the use of total virus abundance, identifying whom infects whom is paramount in obtaining a greater understanding of viral dynamics and virus-host interactions in engineered systems, a situation analogous to that in natural environments (Breitbart, 2012; Brum and Sullivan, 2015). The theoretical approach adopted here (Chapter 5) could help in this regard, since it offers a means of identifying predator-prey type dynamics in complex, natural microbial communities and thus those bacterial taxa potentially under predation. Indeed the qPCR data here could be used to convert proportional bacterial abundances to actual (Chapter 3 and 4) and thus facilitate a theory-based investigation of predator-prey dynamics across thousands of taxa at both full and lab scales. Those taxa with comparable associations to those identified between ammonia oxidising bacteria (AOB) and mixed liquor (ML) virus abundance (Chapter 5) would be prime candidates for hosts actively interacting with viruses. Moreover analogous associations utilising sequence based AOB (*Nitrosomonas*) counts from both full- and lab-scale studies would help corroborate previous findings (Chapter 5). Alternatively, or in addition, those bacterial taxa found to be associated with ML virus abundance using local similarity analysis (LSA, Chapter 3 and 4) may also be potential hosts for future work, supplementary approaches for computing such correlation networks could help corroborate and extend such findings (e.g. CoNet (Faust *et al*., 2012), the miximal information coefficient (MIC, Reshef *et al*., 2011), MENA (Zhou *et al*., 2011; Deng *et al*., 2012b) and SparCC (Friedman and Alm, 2012)).

Whilst this identifies potential bacterial hosts of interest, alongside known functional organisms such as AOB, the metagenomic data produced here (Chapter 7) also provides a source of potential viral targets, particularly those novel viruses extracted using the presented novel technique found solely, or at greater abundance, in the ML and effluent, that is those presumed to be actively interacting with hosts. Indeed computational analysis (reviewed by Edwards *et al*., 2016) of these

viromes, alongside FISH-FACS generated, AOB enriched metagenomes from the same wastewater treatment plant (Fig. 8.1 D, Bell *et al*., unpublished, modified from Haroon *et al*., 2013), could offer a first pass at identifying AOB infecting viruses. Single-cell genomics-based analysis of the AOB metagenomes, as described by Labonté *et al*. (2015), may equally detect AOB infecting viruses and identify modes of virus-AOB interaction. Accordingly temporal extension of both the ML virus (and effluent) and AOB enriched metagenomic data sets using archived samples, as well as the generation of standard metagenomic data, would aid in such approaches, as would the generation of RNA viromes. Moreover the former and the latter, coupled with the already generated 16s sequence data (Chapter 3 and 4), may enable statistical approaches (e.g. LSA, CoNet, MIC etc.) to further identify potential virus-host pairs (e.g. Zhang *et al*., 2017).

Whilst such *in silico* analysis offers prioritisation of virus/host candidates for further experimental analysis, ultimately identifying whom infects whom requires *in vivo* experiments. The garnered host and virus sequence data (Chapter 3, 4 and 7), or that generated in the future, is thus principally important, since available culture independent methods, such as viral tagging (Deng *et al*., 2012, 2014), phageFISH (Allers *et al*., 2013), microfluidic digital PCR (Tadmor *et al*., 2011) and more recently the polony method (Baran *et al*., 2018), require such knowledge. Of those available the polony method (Baran *et al*., 2018), for which I attended a workshop in February 2015, perhaps offers the greatest potential in linking viruses to their hosts, particularly since archived samples could be utilised. Using viral sequences of interest specific PCR primers and fluorescently labelled probes could be designed and utilised in this solid-phase PCR amplification method (Fig. 9.1 A), whereby polonies, amplification spheres of fluorescently labelled viral DNA, can be visualised and quantified on polyacrylamide gels (Fig. 9.1 B). The use of general 16s probes and primers through duplex PCR and differing fluorophores enables simultaneous visualisation of potential hosts (Fig. 9.1 B), with colocalised polonies indicating an infected bacterial host. These colocalised polonies can then be picked, sequenced and thus the host identified.

The development of FISH-FACS for sorting AOB from complex ML samples (Bell *et al*., unpublished), coupled with viral tagging (Deng *et al*., 2012a, 2014), also offers a potential opportunity for identifying AOB infecting viruses, or by modifying the FISH-FACS method viruses infecting any host of interest (Fig. 9.1 C). Here the viral fraction of a live sample is stained with a generic DNA stain (e.g. SYBR Green I /II or SYBR Gold) and then reintroduced and incubated alongside host cells, some of which, through adsorption and subsequent infection, become tagged with fluorescently labelled viruses. Host cells of interest are then fluorescently labelled and sorted into three fractions using FISH-FACS; host of interest with tagged viruses, other hosts with tagged viruses

and untagged hosts (Fig. 9.1 C). The former fraction can be sequenced and infectious viruses of interest identified, sequencing the second fraction would additionally identify potential host-virus pairs for future work. Validation of both the polony method and viral tagged FISH-FACS could be achieved using previously isolated and culturable virus-host pairs.

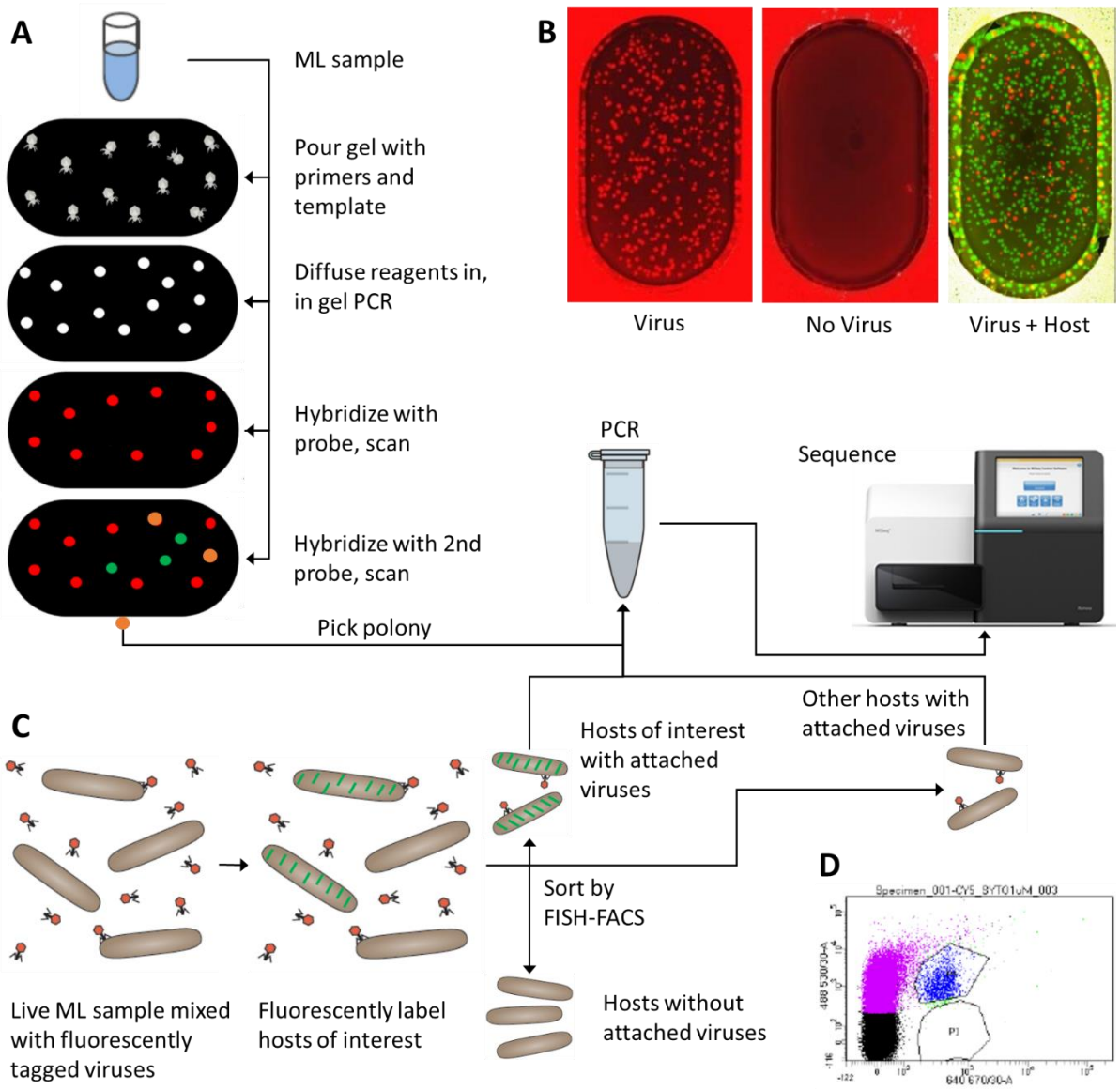

Figure 9. 1. (**A**) Proposed polony method for identifying hosts of known viruses from ML samples (modified from Baran *et al*., 2018). (**B**) Polony formation for a known virus, a no virus control and a virus and hosts (modified from Baran *et* al., 2018 using workshop protocols). (**C**) Proposed viral tagged FISH-FACS method for identifying viruses of known hosts from ML samples (modified from Brum and Sullivan, 2015). (**D**) Example cytogram of FISH-FACS sorted AOB cells (Blue, Bell *et al*., unpublished).

Once relevant virus-host pairs have been identified the polony method, or simpler approaches such as qPCR and/or droplet digital PCR, could be utilised to temporally track virus-host dynamics using the archived samples. Such work would facilitate assessment and development of existing evolutionary/ecological models describing virus-host interactions (e.g. kill the winner and fluctuating selection), it would enable viruses to be incorporated into new predictive ecological

models and, from an engineered systems perspective, it may permit their inclusion in biokinetic models predicting a systems performance (e.g. the activated sludge (Henze *et al*., 2000) and/or the anaerobic digestion models (Batstone *et al*., 2002)). As such elucidation of the viruses infecting functional organisms is a priority if the latter, as well as a better understanding of the role viruses play in the functional dynamics of engineered systems, is to be achieved.

Elucidating drivers of the lysogenic-lytic switch is another area of importance across all of viral ecology, particularly in light of recent findings that viruses are relatively less abundant at high host densities (Wigington *et al*., 2016; Knowles *et al*., 2016, Chapter 6) and the consequent proposition of the contentious "Piggyback-the-Winner" hypothesis (PtW, Knowles *et al*., 2016; Weitz *et al*., 2017; Knowles and Rohwer, 2017; Knowles *et al*., 2017). As previuously argued (Chapter 6) engineered systems, given their comparatively high host densities, evidence of lysogeny (Choi *et al*., 2010; Motlagh *et al*., 2015) and their highly controlled and monitored environments, offer fertile testing grounds in this regard. Indeed frequent monitoring of total virus and host abundance in a simple dilution experiment using ML could confirm or discredit PtW, if correct the lytic cycle should predominate at higher dilutions and thus increase viral abundance in a step-wise manner, or perhaps once host abundance drops below a critical threshold. Simultaneous metagenomic analysis of both the viral and cellular fractions from each dilution would allow assessment of the presence or absence of "hallmark" genes associated with lysogeny and, potentially, an increase or decrease in identifiable prophages (Labonté *et al*., 2015), further corroborating or discrediting PtW. Quantification of free viruses in ML post and pre chemical induction may also help identify the presence, or absence, of lysogenic viruses in ML, although the robustness of this method as a measure of lysogeny has recently been questioned (Knowles *et al*., 2017).

Although not previously discussed those factors influencing virus abundance at both full (Chapter 3) and lab scales (Chapter 4), particularly those influencing adsorption process, could be prime candidates for abiotic drivers of the lysogenic-lytic switch. Manipulation of such variables in simple, well monitored replicate microcosm experiments, where virus and host abundance and metagenomes (also 16s for hosts) can be determined/generated at high temporal frequency, would aid in understanding their role, whether that be related to lysogeny or not. Given its role at both full (Chapter 3) and lab scales (Chapter 4) similar experiments and analysis could also help elucidate the role of virus immigration, whereby a wastewater viral concentrate, generated using tangential flow filtration, is spiked at known, varying concentrations into a number of replicate ML microcosms.

Another issue across viral ecology, as experienced here (Chapter 7), is that the majority of generated viromes lack similarity to reference sequences in extant databases, thus the inferential power of metagenomic data is limited (Brum and Sullivan, 2015). As such refinement of the approach to identify novel viral-like genomes presented here is pertinent (Chapter 7). Certainly defining and curating a robust set of representative viral homologies will be essential for maximising the recovery of individual virus sub-graphs (viral-like genomes), whilst detecting and incorporating ORFans, coding sequences with no homologs, may also pay dividends (Barrientos-Somarribas *et al*., 2018). Comparison against existing techniques would also be warranted (e.g. Roux *et al*., 2014; Li *et al*., 2016; Barrientos-Somarribas *et al*., 2018), whilst extending its use to existing, hitherto untapped datasets, given the comparatively small-scale investigation here extracted > 280 novel viral-like genomes, could unearth further complete or nearly complete novel viral genomes. Accordingly following refinement and validation its use will hopefully enable new insight into viral dynamics in engineered and natural systems alike.

Other potential areas of future work lie in the application of the developed, or proposed, methods in other engineered systems. Indeed the role of viruses in anaerobic digesters has recently been explored (Zhang *et al*., 2017) and modelled (Louca and Doebeli, 2017), with members of the viral community being linked to host taxa and system performance (Zhang *et al*., 2017) and implicated in high taxonomic turnover and functional stability (Louca and Doebeli, 2017). Alike AOB and nitrification the process is dependent on a functional group (methanogens) low in diversity, thus system performance could be impacted by viral infection, warranting investigation. Similarly evidence implying viruses can control biofilm growth and alter biofilm diversity, architecture and function (Sutherland *et al*., 2004) warrants exploration of their dynamics in attached growth systems such as trickling (wastewater treatment) and slow sand filters (drinking water treatment). In summary a vast amount of future work, although challenging, could be undertaken in this emerging and exciting nexus of viral ecology and wastewater/water microbial ecology.

CHAPTER 10

REFERENCES

Ackermann, H. W. and D. Prangishvili (2012). "Prokaryote viruses studied by electron microscopy." Arch Virol **157**(10): 1843-1849.

Adams, M. H. 1959. Bacteriophages. Interscience Publishers, New York.

Aitchison, J. (1982). "The Statistical Analysis of Compositional Data." Journal of the Royal Statistical Society. Series B (Methodological) **44**(2): 139-177.

Allander, T., S. U. Emerson, R. E. Engle, R. H. Purcell and J. Bukh (2001). "A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species." Proc Natl Acad Sci U S A **98**(20): 11609-11614.

Allers, E., C. Moraru, M. B. Duhaime, E. Beneze, N. Solonenko, J. Barrero-Canosa, R. Amann and M. B. Sullivan (2013). "Single-cell and population level viral infection dynamics revealed by phageFISH, a method to visualize intracellular and free viruses." Environ Microbiol **15**(8): 2306-2318.

Angly, F. E., B. Felts, M. Breitbart, P. Salamon, R. A. Edwards, C. Carlson, A. M. Chan, M. Haynes, S. Kelley, H. Liu, J. M. Mahaffy, J. E. Mueller, J. Nulton, R. Olson, R. Parsons, S. Rayhawk, C. A. Suttle and F. Rohwer (2006). "The marine viromes of four oceanic regions." Plos Biology **4**(11): 2121-2131.

Angly, F. E., P. G. Dennis, A. Skarshewski, I. Vanwonterghem, P. Hugenholtz and G. W. Tyson (2014). "CopyRighter: a rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction." Microbiome **2**(1): 11.

APHA (1998). Standard Methods for the Examination of Water and Wastewater, APHA.

Ardern, E. and W. T. Lockett (1914). "Experiments on the oxidation of sewage without the aid of filters." Journal of the Society of Chemical Industry **33**(10): 523-539.

Avrani, S., D. A. Schwartz and D. Lindell (2012). "Virus-host swinging party in the oceans: Incorporating biological complexity into paradigms of antagonistic coexistence." Mob Genet Elements **2**(2): 88-95.

Avrani, S., O. Wurtzel, I. Sharon, R. Sorek and D. Lindell (2011). "Genomic island variability facilitates Prochlorococcus–virus coexistence." Nature **474**: 604.

Ayarza, J. M. and L. Erijman (2011). "Balance of Neutral and Deterministic Components in the Dynamics of Activated Sludge Floc Assembly." Microbial Ecology **61**(3): 486-495.

Ayarza, J. M., L. D. Guerrero and L. Erijman (2010). "Nonrandom Assembly of Bacterial Populations in Activated Sludge Flocs." Microbial Ecology **59**(3): 436-444.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev and P. A. Pevzner (2012). "SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing." Journal of Computational Biology **19**(5): 455-477.

Baptista, J. D., M. Lunn, R. J. Davenport, D. L. Swan, L. F. Read, M. R. Brown, C. Morais and T. P. Curtis (2014). "Agreement between amoA gene-specific quantitative PCR and fluorescence in situ hybridization in the measurement of ammonia-oxidizing bacteria in activated sludge." Appl Environ Microbiol **80**(19): 5901-5910.

Baran, N., S. Goldin, I. Maidanik and D. Lindell (2018). "Quantification of diverse virus populations in the environment using the polony method." Nature Microbiology **3**(1): 62-72.

Barr, J. J., F. R. Slater, T. Fukushima and P. L. Bond (2010). "Evidence for bacteriophage activity causing community and performance changes in a phosphorus-removal activated sludge." Fems Microbiology Ecology **74**(3): 631-642.

Barrientos-Somarribas, M., D. N. Messina, C. Pou, F. Lysholm, A. Bjerkner, T. Allander, B. Andersson and E. L. L. Sonnhammer (2018). "Discovering viral genomes in human metagenomic data by predicting unknown protein families." Scientific Reports **8**(1): 28.

Bartoń, K (2016). MuMIn: Multi-Model Inference. R package version 1.15.6. https://CRAN.R-project.org/package=MuMIn

Baselga, A., D. Orme, S. Villeger, J. De Bortoli and F. Leprieur (2017). betapart: Partitioning Beta Diversity into Turnover and Nestedness Components. R package version 1.4-1. https://CRAN.R-project.org/package=betapart

Bates, P.J., D, DebRoy S, Sarkar D and R Core Team (2017). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-131, https://CRAN.R-project.org/package=nlme.

Batstone, D. J., J. Keller, I. Angelidaki, S. V. Kalyuzhnyi, S. G. Pavlostathis, A. Rozzi, W. T. Sanders, H. Siegrist and V. A. Vavilin (2002). "The IWA Anaerobic Digestion Model No 1 (ADM1)." Water Sci Technol **45**(10): 65-73.

Beecroft, N. J., F. Zhao, J. R. Varcoe, R. C. T. Slade, A. E. Thumser and C. Avignone-Rossa (2012). "Dynamic changes in the microbial community composition in microbial fuel cells fed with sucrose." Applied Microbiology and Biotechnology **93**(1): 423-437.

Bell, G. (2005). "The co-distribution of species in relation to the neutral theory of community ecology." Ecology **86**(7): 1757-1770.

Benson, D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell and E. W. Sayers (2009). "GenBank." Nucleic Acids Res **37**(Database issue): D26-31.

Bergh, O., K. Y. Borsheim, G. Bratbak and M. Heldal (1989). "High abundance of viruses found in aquatic environments." Nature **340**(6233): 467-468.

Bernhard Lehnert (2015). BlandAltmanLeh: Plots (Slightly Extended) Bland-Altman Plots. R package version 0.3.1. URL https://CRAN.R-project.org/package=BlandAltmanLeh

Bettarel, Y., T. Sime-Ngando, C. Amblard and J. Dolan (2004). "Viral activity in two contrasting lake ecosystems." Applied and Environmental Microbiology **70**(5): 2941-2951.

Betts, A., O. Kaltz and M. E. Hochberg (2014). "Contrasted coevolutionary dynamics between a bacterial pathogen and its bacteriophages." Proceedings of the National Academy of Sciences **111**(30): 11109-11114.

Blackall, L. L., E. M. Seviour, D. Bradford, H. M. Stratton, M. A. Cunningham, P. Hugenholtz and R. J. Seviour (1996). "Towards understanding the taxonomy of some of the filamentous bacteria causing bulking and foaming in activated sludge plants." Water Science and Technology **34**(5–6): 137-144.

Bloch, C. P., C. L. Higgins and M. R. Willig (2007). "Effects of large-scale disturbance on metacommunity structure of terrestrial gastropods: temporal trends in nestedness." Oikos **116**(3): 395-406.

Bohannan, B. J. M. and R. E. Lenski (2000). "Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage." Ecology Letters **3**(4): 362-377.

Bolger, A. M., M. Lohse and B. Usadel (2014). "Trimmomatic: a flexible trimmer for Illumina sequence data." Bioinformatics **30**(15): 2114-2120.

Bolotin, A., B. Quinquis, P. Renault, A. Sorokin, S. D. Ehrlich, S. Kulakauskas, A. Lapidus, E. Goltsman, M. Mazur, G. D. Pusch, M. Fonstein, R. Overbeek, N. Kyprides, B. Purnelle, D. Prozzi, K. Ngui, D. Masuy, F. Hancy, S. Burteau, M. Boutry, J. Delcour, A. Goffeau and P. Hols (2004). "Complete

sequence and comparative genome analysis of the dairy bacterium Streptococcus thermophilus." Nature Biotechnology **22**(12): 1554-1558.

Bondy-Denomy, J., A. Pawluk, K. L. Maxwell and A. R. Davidson (2012). "Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system." Nature **493**: 429.

Bondy-Denomy, J., B. Garcia, S. Strum, M. Du, M. F. Rollins, Y. Hidalgo-Reyes, B. Wiedenheft, K. L. Maxwell and A. R. Davidson (2015). "Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins." Nature **526**(7571): 136-139.

Bondy-Denomy, J., J. Qian, E. R. Westra, A. Buckling, D. S. Guttman, A. R. Davidson and K. L. Maxwell (2016). "Prophages mediate defense against phage infection through diverse mechanisms." The Isme Journal **10**: 2854.

Bouvier, T. and P. A. del Giorgio (2007). "Key role of selective viral-induced mortality in determining marine bacterial community composition." Environmental Microbiology **9**(2): 287-297.

Braguglia, C. M., M. C. Gagliano and S. Rossetti (2012a). "High frequency ultrasound pretreatment for sludge anaerobic digestion: Effect on floc structure and microbial population." Bioresource Technology **110**(0): 43-49.

Braguglia, C. M., M. C. Gagliano, A. Gallipoli and S. Rossetti (2012b). "Enhanced anaerobic digestion performances: effect of sludge ultrasound pre-treatment and role of the microbial population." Environmental Engineering and Management Journal **11**(10): 1803-1810.

Bratbak, G., M. Heldal, T. F. Thingstad and P. Tuomi (1996). "Dynamics of virus abundance in coastal seawater." FEMS Microbiology Ecology **19**(4): 263-269.

Breitbart, M. (2012). Marine Viruses: Truth or Dare. Annual Review of Marine Science, Vol 4. C. A. G. S. J. Carlson. **4:** 425-448.

Breitbart, M. and F. Rohwer (2005). "Method for discovering novel DNA viruses in blood using viral particle selection and shotgun sequencing." Biotechniques **39**(5): 729-736.

Breitbart, M., B. Felts, S. Kelley, J. M. Mahaffy, J. Nulton, P. Salamon and F. Rohwer (2004). "Diversity and population structure of a near-shore marine-sediment viral community." Proceedings of the Royal Society of London Series B-Biological Sciences **271**(1539): 565-574.

Breitbart, M., P. Salamon, B. Andresen, J. M. Mahaffy, A. M. Segall, D. Mead, F. Azam and F. Rohwer (2002). "Genomic analysis of uncultured marine viral communities." Proceedings of the National Academy of Sciences of the United States of America **99**(22): 14250-14255.

Briones, A. and L. Raskin (2003). "Diversity and dynamics of microbial communities in engineered environments and their implications for process stability." Curr Opin Biotechnol **14**(3): 270-276.

Brockhurst, M. A., A. Fenton, B. Roulston and P. B. Rainey (2006). The impact of phages on interspecific competition in experimental populations of bacteria. BMC Ecology. **6:** 19.

Brockhurst, M. A., Awells. D. Morgan, A. Fenton and A. Buckling (2007). "Experimental coevolution with bacteria and phage The Pseudomonas fluorescens - Phi 2 model system." Infection Genetics and Evolution **7**(4): 547-552.

Brockhurst, M. A., P. B. Rainey and A. Buckling (2004). "The effect of spatial heterogeneity and parasites on the evolution of host diversity." Proceedings. Biological sciences **271**(1534): 107-111.

Brown, M. R., S. Camézuli, R. J. Davenport, E. Petelenz-Kurdziel, L. Øvreås and T. P. Curtis (2015). "Flow cytometric quantification of viruses in activated sludge." Water Research **68**(0): 414-422.

Brum, J. R. and M. B. Sullivan (2015a). "Rising to the challenge: accelerated pace of discovery transforms marine virology." Nat Rev Micro **13**(3): 147-159.

Brum, J. R., B. L. Hurwitz, O. Schofield, H. W. Ducklow and M. B. Sullivan (2015b). "Seasonal time bombs: dominant temperate viruses affect Southern Ocean microbial dynamics." The Isme Journal **10**: 437.

Brum, J. R., J. C. Ignacio-Espinoza, S. Roux, G. Doulcier, S. G. Acinas, A. Alberti, S. Chaffron, C. Cruaud, C. de Vargas, J. M. Gasol, G. Gorsky, A. C. Gregory, L. Guidi, P. Hingamp, D. Iudicone, F. Not, H. Ogata, S. Pesant, B. T. Poulos, S. M. Schwenck, S. Speich, C. Dimier, S. Kandels-Lewis, M. Picheral, S. Searson, P. Bork, C. Bowler, S. Sunagawa, P. Wincker, E. Karsenti and M. B. Sullivan (2015c). "Patterns and ecological drivers of ocean viral communities." Science **348**(6237).

Brussaard, C. P. D. (2004). "Optimization of Procedures for Counting Viruses by Flow Cytometry." Applied and Environmental Microbiology **70**(3): 1506-1513.

Brussaard, C. P. D., D. Marie and G. Bratbak (2000). "Flow cytometric detection of viruses." Journal of Virological Methods **85**(1–2): 175-182.

Brussaard, C. P. D., J. P. Payet, C. Winter and M. G. Weinbauer (2010). Quantification of aquatic viruses by flow cytometry. Manual of Aquatic Viral Ecology. S. W. Wilhelm, M. G. Weinbauer and C. A. Suttle, ASLO**:** 102–109.

Brussaard, C. P. D., S. W. Wilhelm, F. Thingstad, M. G. Weinbauer, G. Bratbak, M. Heldal, S. A. Kimmance, M. Middelboe, K. Nagasaki, J. H. Paul, D. C. Schroeder, C. A. Suttle, D. Vaqué and K. E. Wommack (2008). "Global-scale processes with a nanoscale drive: the role of marine viruses." The Isme Journal **2**: 575.

Brussow, H., C. Canchaya and W. D. Hardt (2004). "Phages and the evolution of bacterial pathogens: From genomic rearrangements to lysogenic conversion." Microbiology and Molecular Biology Reviews **68**(3): 560-+.

Buckling, A. and P. B. Rainey (2002). "Antagonistic coevolution between a bacterium and a bacteriophage." Proc Biol Sci **269**(1494): 931-936.

Calendar, R. L. and S. T. Abedon (2005). The Bacteriophages, Oxford University Press, USA.

Calusinska, M., M. Marynowska, X. Goux, E. Lentzen and P. Delfosse (2016). "Analysis of dsDNA and RNA viromes in methanogenic digesters reveals novel viral genetic diversity." Environmental Microbiology **18**(4): 1162-1175.

Campbell, A. (1961). "Conditions for the Existence of Bacteriophage." Evolution **15**(2): 153-165.

Canchaya, C., C. Proux, G. Fournous, A. Bruttin and H. Brüssow (2003). "Prophage Genomics." Microbiology and Molecular Biology Reviews **67**(2): 238-276.

Cantalupo, P. G., B. Calgua, G. Zhao, A. Hundesa, A. D. Wier, J. P. Katz, M. Grabe, R. W. Hendrix, R. Girones, D. Wang and J. M. Pipas (2011). "Raw Sewage Harbors Diverse Viral Populations." Mbio **2**(5).

Caporaso, J. G., C. L. Lauber, W. A. Walters, D. Berg-Lyons, J. Huntley, N. Fierer, S. M. Owens, J. Betley, L. Fraser, M. Bauer, N. Gormley, J. A. Gilbert, G. Smith and R. Knight (2012). "Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms." Isme j **6**(8): 1621-1624.

Cardinale, B. J., M. A. Palmer and S. L. Collins (2002). "Species diversity enhances ecosystem functioning through interspecific facilitation." Nature **415**(6870): 426-429.

Casjens, S. (2003). "Prophages and bacterial genomics: what have we learned so far?" Molecular Microbiology **49**(2): 277-300.

Chen, F., J. R. Lu, B. J. Binder, Y. C. Liu and R. E. Hodson (2001). "Application of digital image analysis and flow cytometry to enumerate marine viruses stained with SYBR gold." Applied and Environmental Microbiology **67**(2): 539-545.

Chien, I. C., J. S. Meschke, H. L. Gough and J. F. Ferguson (2013). "Characterization of Persistent Virus-Like Particles in Two Acetate-Fed Methanogenic Reactors." PLOS ONE **8**(11): e81040.

Ching, J., S. A. Musheyev, D. Chowdhury, J. A. Kim, Y. Choi and J. J. Dennehy (2013). "Migration Enhances Adaptation In Bacteriophage Populations Evolving In Ecological Sinks." Evolution **67**(1): 10-17.

Chiura, H. X. (1997). "Generalized gene transfer by virus-like particles from marine bacteria." Aquatic Microbial Ecology **13**(1): 75-83.

Choi, J. D., S. M. Kotay and R. Goel (2010). "Various physico-chemical stress factors cause prophage induction in Nitrosospira multiformis 25196-an ammonia oxidizing bacteria." Water Research **44**(15): 4550-4558.

Colomer-Lluch, M., J. Jofre and M. Muniesa (2011). "Antibiotic Resistance Genes in the Bacteriophage DNA Fraction of Environmental Samples." PLoS ONE **6**(3): e17549.

Colson, P., X. De Lamballerie, N. Yutin, S. Asgari, Y. Bigot, D. K. Bideshi, X. W. Cheng, B. A. Federici, J. L. Van Etten, E. V. Koonin, B. La Scola and D. Raoult (2013). ""Megavirales", a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses." Arch Virol **158**(12): 2517-2521.

Cook, K. L., J. L. Garland, A. C. Layton, H. M. Dionisi, L. H. Levine and G. S. Sayler (2006). "Effect of Microbial Species Richness on Community Stability and Community Function in a Model Plant-Based Wastewater Processing System." Microbial Ecology **52**(4): 725-737.

Cortez, M. H. and J. S. Weitz (2014). "Coevolution can reverse predator–prey cycles." Proceedings of the National Academy of Sciences **111**(20): 7486-7491.

Coskuner, G., S. J. Ballinger, R. J. Davenport, R. L. Pickering, R. Solera, I. M. Head and T. P. Curtis (2005). "Agreement between Theory and Measurement in Quantification of Ammonia-Oxidizing Bacteria." Applied and Environmental Microbiology **71**(10): 6325-6334.

Cousin, C. P. and J. J. Ganczarczyk (1999). "Effect of Calcium Ion Concentration on the Structure of Activated Sludge Flocs." Environmental Technology **20**(11): 1129-1138.

Culley, A. I., A. S. Lang and C. A. Suttle (2006). "Metagenomic Analysis of Coastal RNA Virus Communities." Science **312**(5781): 1795-1798.

Curran, P. J., S. G. West and J. F. Finch (1996). "The Robustness of Test Statistics to Nonnormality and Specification Error in Confirmatory Factor Analysis." Psychological Methods **1**(1): 16-29.

Curtis, T. P. and W. T. Sloan (2006). "Towards the design of diversity: stochastic models for community assembly in wastewater treatment plants." Water Science and Technology **54**(1): 227-236.

Daims, H., M. W. Taylor and M. Wagner (2006). "Wastewater treatment: a model system for microbial ecology." Trends in Biotechnology **24**(11): 483-489.

Dang, V. T. and M. B. Sullivan (2014). "Emerging methods to study bacteriophage infection at the single-cell level." Front Microbiol **5**: 724.

Danovaro, R. and M. Middelboe (2010). Separation of free virus particles from sediments in aquatic systems. Manual of Aquatic Viral Ecology. S. W. Wilhelm, W. M. G. and C. Suttle, ASLO**:** 75 - 81.

Danovaro, R., A. Dell'Anno, C. Corinaldesi, M. Magagnini, R. Noble, C. Tamburini and M. Weinbauer (2008a). "Major viral impact on the functioning of benthic deep-sea ecosystems." Nature **454**: 1084.

Danovaro, R., A. Dell'anno, A. Trucco, M. Serresi and S. Vanucci (2001). "Determination of virus abundance in marine sediments." Applied and Environmental Microbiology **67**(3): 1384-1387.

Danovaro, R., C. Corinaldesi, M. Filippini, U. R. Fischer, M. O. Gessner, S. Jacquet, M. Magagnini and B. Velimirov (2008b). "Viriobenthos in freshwater and marine sediments: a review." Freshwater Biology **53**(6): 1186-1213.

Danovaro, R., E. Manini and A. Dell'Anno (2002). "Higher Abundance of Bacteria than of Viruses in Deep Mediterranean Sediments." Applied and Environmental Microbiology **68**(3): 1468-1472.

Davenport, R. J., T. P. Curtis, M. Goodfellow, F. M. Stainsby and M. Bingley (2000). "Quantitative Use of Fluorescent In Situ Hybridization To Examine Relationships between Mycolic Acid-Containing Actinomycetes and Foaming in Activated Sludge Plants." Applied and Environmental Microbiology **66**(3): 1158-1166.

Davenport, R.J. and T.P. Curtis (2004). Quantitative fluorescence in situ hybridisation (FISH): statistical methods for valid cell counting. Molecular Microbial Ecology Manual. Kowalchuk, G.A., F.J. Bruijn, I.M. Head, A.J. Van der Zijpp and J.D. van Elsas, Kluwer Academic Publishers: 1487 – 1516.

Dedrick, R. M., D. Jacobs-Sera, C. A. Guerrero Bustamante, R. A. Garlena, T. N. Mavrich, W. H. Pope, J. C. C. Reyes, D. A. Russell, T. Adair, R. Alvey, J. A. Bonilla, J. S. Bricker, B. R. Brown, D. Byrnes, S. G. Cresawn, W. B. Davis, L. A. Dickson, N. P. Edgington, A. M. Findley, U. Golebiewska, J. H. Grose, C. F. Hayes, L. E. Hughes, K. W. Hutchison, S. Isern, A. A. Johnson, M. A. Kenna, K. K. Klyczek, C. M. Mageeney, S. F. Michael, S. D. Molloy, M. T. Montgomery, J. Neitzel, S. T. Page, M. C. Pizzorno, M. K. Poxleitner, C. A. Rinehart, C. J. Robinson, M. R. Rubin, J. N. Teyim, E. Vazquez, V. C. Ware, J. Washington and G. F. Hatfull (2017). "Prophage-mediated defense against viral attack and viral counter-defense." Nature microbiology **2**: 16251-16251.

Del Casale, A., P. V. Flanagan, M. J. Larkin, C. C. R. Allen and L. A. Kulakov (2011a). "Analysis of transduction in wastewater bacterial populations by targeting the phage-derived 16S rRNA gene sequences." Fems Microbiology Ecology **76**(1): 100-108.

Del Casale, A., P. V. Flanagan, M. J. Larkin, C. C. R. Allen and L. A. Kulakov (2011b). "Extent and Variation of Phage-Borne Bacterial 16S rRNA Gene Sequences in Wastewater Environments." Applied and Environmental Microbiology **77**(15): 5529-5532.

Deng, L., A. Gregory, S. Yilmaz, B. T. Poulos, P. Hugenholtz and M. B. Sullivan (2012a). "Contrasting Life Strategies of Viruses that Infect Photo- and Heterotrophic Bacteria, as Revealed by Viral Tagging." mBio **3**(6).

Deng, L., J. C. Ignacio-Espinoza, A. C. Gregory, B. T. Poulos, J. S. Weitz, P. Hugenholtz and M. B. Sullivan (2014). "Viral tagging reveals discrete populations in Synechococcus viral genome sequence space." Nature **513**: 242.

Deng, Y., Y.-H. Jiang, Y. Yang, Z. He, F. Luo and J. Zhou (2012b). "Molecular ecological network analyses." BMC Bioinformatics **13**(1): 113.

Dominiak, D. M., J. L. Nielsen and P. H. Nielsen (2011). "Extracellular DNA is abundant and important for microcolony strength in mixed microbial biofilms." Environ Microbiol **13**(3): 710-721.

Downing, A.L., H.A. Painter and G. Knowles (1964). “Nitrification in the activated sludge process”. J. Inst. Sewage Purif **2**: 130–158.

Duncan, A. B., E. Dusi, F. Jacob, J. Ramsayer, M. E. Hochberg and O. Kaltz (2017). "Hot spots become cold spots: coevolution in variable temperature environments." Journal of Evolutionary Biology **30**(1): 55-65.

Dytczak, M. A., K. L. Londry and J. A. Oleszkiewicz (2008). "Activated sludge operational regime has significant impact on the type of nitrifying community and its nitrification rates." Water Research **42**(8–9): 2320-2328.

Edwards, R. A., K. McNair, K. Faust, J. Raes and B. E. Dutilh (2016). "Computational approaches to predict bacteriophage-host relationships." FEMS Microbiol Rev **40**(2): 258-272.

Ellis, E. L. and M. Delbrück (1939). "The growth of bacteriophage." The Journal of General Physiology **22**(3): 365-384.

Ewert, D. L. and M. J. B. Paynter (1980). "Enumeration of bacteriophages and host bacteria in sewage and the activated-sludge treatment process." Applied and Environmental Microbiology **39**(3): 576-583.

Falk, M. W. and S. Wuertz (2010). "Effects of the toxin 3-chloroaniline at low concentrations on microbial community dynamics and membrane bioreactor performance." Water Research **44**(17): 5109-5115.

Falk, M. W., K.-G. Song, M. G. Matiasek and S. Wuertz (2009). "Microbial community dynamics in replicate membrane bioreactors – Natural reproducible fluctuations." Water Research **43**(3): 842-852.

Faust, K., J. F. Sathirapongsasuti, J. Izard, N. Segata, D. Gevers, J. Raes and C. Huttenhower (2012). "Microbial Co-occurrence Relationships in the Human Microbiome." PLOS Computational Biology **8**(7): e1002606.

Fernandez, A. S., S. A. Hashsham, S. L. Dollhopf, L. Raskin, O. Glagoleva, F. B. Dazzo, R. F. Hickey, C. S. Criddle and J. M. Tiedje (2000). "Flexible Community Structure Correlates with Stable Community Function in Methanogenic Bioreactor Communities Perturbed by Glucose." Applied and Environmental Microbiology **66**(9): 4058-4067.

Figuerola, E. L. M. and L. Erijman (2010). "Diversity of nitrifying bacteria in a full-scale petroleum refinery wastewater treatment plant experiencing unstable nitrification." Journal of Hazardous Materials **181**(1-3): 281-288.

Filzmoser, P. and M. Gschwandtner (2017). mvoutlier: Multivariate Outlier Detection Based on Robust Methods. R package version 2.0.8. https://CRAN.R-project.org/package=mvoutlier.

Flores, C. O., J. R. Meyer, S. Valverde, L. Farr and J. S. Weitz (2011). "Statistical structure of host–phage interactions." Proceedings of the National Academy of Sciences **108**(28): E288.

Forde, S. E., J. N. Thompson, R. D. Holt and B. J. M. Bohannan (2008). "Coevolution drives temporal changes in fitness and diversity across environments in a bacteria–bacteriophage interaction." Evolution **62**(8): 1830-1839.

Forterre, P., N. Soler, M. Krupovic, E. Marguet and H. W. Ackermann (2013). "Fake virus particles generated by fluorescence microscopy." Trends in Microbiology **21**(1): 1-5.

Fox, J. and S. Weisberg (2011). An {R} Companion to Applied Regression, Second Edition. Thousand Oaks CA: Sage. URL: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Friedman, J. and E. J. Alm (2012). "Inferring Correlation Networks from Genomic Survey Data." PLOS Computational Biology **8**(9): e1002687.

Fuhrman, J. (1992). Bacterioplankton Roles in Cycling of Organic Matter: The Microbial Food Web. Primary Productivity and Biogeochemical Cycles in the Sea. P. G. Falkowski, A. D. Woodhead and K. Vivirito. Boston, MA, Springer US**:** 361-383.

Fuhrman, J. A. (1999). "Marine viruses and their biogeochemical and ecological effects." Nature **399**: 541.

Gao, W. J. J., H. J. Lin, K. T. Leung and B. Q. Liao (2010). "Influence of elevated pH shocks on the performance of a submerged anaerobic membrane bioreactor." Process Biochemistry **45**(8): 1279-1287.

Gentile, M. E., J. L. Nyman and C. S. Criddle (2007). "Correlation of patterns of denitrification instability in replicated bioreactor communities with shifts in the relative abundance and the denitrification patterns of specific populations." Isme Journal **1**(8): 714-728.

Ghasemi, A. and S. Zahediasl (2012). "Normality Tests for Statistical Analysis: A Guide for Non-Statisticians." International Journal of Endocrinology and Metabolism **10**(2): 486-489.

Gobler, C. J., D. A. Hutchins, N. S. Fisher, E. M. Cosper and S. A. Sanudo-Wilhelmy (1997). "Release and bioavailability of C, N, P, Se, and Fe following viral lysis of a marine chrysophyte." Limnology and Oceanography **42**(7): 1492-1504.

Gómez, P. and A. Buckling (2011). "Bacteria-Phage Antagonistic Coevolution in Soil." Science **332**(6025): 106-109.

Gorter, F. A., P. D. Scanlan and A. Buckling (2016). "Adaptation to abiotic conditions drives local adaptation in bacteria and viruses coevolving in heterogeneous environments." Biology Letters **12**(2).

Grace, J. B., T. M. Anderson, H. Olff and S. M. Scheiner (2010). "On the specification of structural equation models for ecological systems." Ecological Monographs **80**(1): 67-87.

Gravel, D., C. D. Canham, M. Beaudet and C. Messier (2006). "Reconciling niche and neutrality: the continuum hypothesis." Ecology Letters **9**(4): 399-409.

Griffin, J. S. and G. F. Wells (2017). "Regional synchrony in full-scale activated sludge bioreactors due to deterministic microbial community assembly." Isme Journal **11**(2): 500-511.

Grigoriev, I. V., H. Nordberg, I. Shabalov, A. Aerts, M. Cantor, D. Goodstein, A. Kuo, S. Minovitsky, R. Nikitin, R. A. Ohm, R. Otillar, A. Poliakov, I. Ratnere, R. Riley, T. Smirnova, D. Rokhsar and I. Dubchak (2012). "The Genome Portal of the Department of Energy Joint Genome Institute." Nucleic Acids Research **40**(D1): D26-D32.

Guo, F. and T. Zhang (2012). "Profiling bulking and foaming bacteria in activated sludge by high throughput sequencing." Water Res **46**(8): 2772-2782.

Gustavsen, J. A., D. M. Winget, X. Tian and C. A. Suttle (2014). "High temporal and spatial diversity in marine RNA viruses implies that they have an important role in mortality and structuring plankton communities." Frontiers in Microbiology **5**(703).

Haaber, J. and M. Middelboe (2009). "Viral lysis of Phaeocystis pouchetii: Implications for algal population dynamics and heterotrophic C, N and P cycling." ISME J **3**(4): 430-441.

Haegeman, B. and M. Loreau (2011). "A mathematical synthesis of niche and neutral theories in community ecology." Journal of Theoretical Biology **269**(1): 150-165.

Hall, A. R., P. D. Scanlan, A. D. Morgan and A. Buckling (2011). "Host-parasite coevolutionary arms races give way to fluctuating selection." Ecol Lett **14**(7): 635-642.

Han, Y., X. Jin, Y. Wang, Y. Liu and X. Chen (2014). "Inhibitory effect of cyanide on nitrification process and its eliminating method in a suspended activated sludge process." Environmental Science and Pollution Research **21**(4): 2706-2713.

Hantula, J., A. Kurki, P. Vuoriranta and D. H. Bamford (1991). "Ecology of bacteriophages infecting activated sludge bacteria." Applied and Environmental Microbiology **57**(8): 2147-2151.

Hara, S., K. Terauchi and I. Koike (1991). "Abundance of viruses in marine waters - assessment by epifluorescence and transmission electron-microscopy." Applied and Environmental Microbiology **57**(9): 2731-2734.

Harcombe, W. R. and J. J. Bull (2005). "Impact of phages on two-species bacterial communities." Applied and Environmental Microbiology **71**(9): 5254-5259.

Haroon, M. F., C. T. Skennerton, J. A. Steen, N. Lachner, P. Hugenholtz and G. W. Tyson (2013). "In-solution fluorescence in situ hybridization and fluorescence-activated cell sorting for single cell and population genome recovery." Methods Enzymol **531**: 3-19.

Havelaar, A. H. and W. M. Hogeboom (1983). "Factors affecting the enumeration of coliphages in sewage and sewage-polluted waters." Antonie van Leeuwenhoek **49**(4-5): 387-397.

Hennes, K. P. and C. A. Suttle (1995). "Direct counts of viruses in natural-waters and laboratory cultures by epifluorescence microscopy." Limnology and Oceanography **40**(6): 1050-1055.

Hennes, K. P., C. A. Suttle and A. M. Chan (1995). "Fluorescently labelled virus probes show that natural virus populations can control the structure of marine microbial communities." Applied and Environmental Microbiology **61**(10): 3623-3627.

Henze, M., W. Gujer, T. Mino and M. C. M. van Loosdrecht (2000). Activated sludge models ASM1, ASM2, ASM2d and ASM3, IWA Publishing Company.

Hernandez-Raquet, G., E. Durand, F. Braun, C. Cravo-Laureau and J.-J. Godon (2013). "Impact of microbial diversity depletion on xenobiotic degradation by sewage-activated sludge." Environmental Microbiology Reports **5**(4): 588-594.

Hewson, I. and J. A. Fuhrman (2006). "Viral impacts upon marine bacterioplankton assemblage structure." Journal of the Marine Biological Association of the United Kingdom **86**(3): 577-589.

Hewson, I. and J. A. Fuhrman (2007). "Covariation of viral parameters with bacterial assemblage richness and diversity in the water column and sediments." Deep-Sea Research Part I-Oceanographic Research Papers **54**(5): 811-830.

Hewson, I., D. M. Winget, K. E. Williamson, J. A. Fuhrman and K. E. Wommack (2006). "Viral and bacterial assemblage covariance in oligotrophic waters of the West Florida Shelf (Gulf of Mexico)." Journal of the Marine Biological Association of the United Kingdom **86**(3): 591-603.

Hewson, I., G. A. Vargo and J. A. Fuhrman (2003). "Bacterial Diversity in Shallow Oligotrophic Marine Benthos and Overlying Waters: Effects of Virus Infection, Containment, and Nutrient Enrichment." Microbial Ecology **46**(3): 322-336.

Hofer, J. S. and R. Sommaruga (2001). "Seasonal dynamics of viruses in an alpine lake: importance of filamentous forms." Aquatic Microbial Ecology **26**(1): 1-11.

Holmfeldt, K., D. Odic, M. B. Sullivan, M. Middelboe and L. Riemann (2012). "Cultivated Single-Stranded DNA Phages That Infect Marine Bacteroidetes Prove Difficult To Detect with DNA-Binding Stains." Applied and Environmental Microbiology **78**(3): 892-894.

Hu, L. t. and P. M. Bentler (1999). "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives." Structural Equation Modeling: A Multidisciplinary Journal **6**(1): 1-55.

Hu, Z., K. Chandran, D. Grasso and B. F. Smets (2004). "Comparison of nitrification inhibition by metals in batch and continuous flow reactors." Water Research **38**(18): 3949-3959.

Huang, Z., P. B. Gedalanga, P. Asvapathanagul and B. H. Olson (2010). "Influence of physicochemical and operational parameters on Nitrobacter and Nitrospira communities in an aerobic activated sludge bioreactor." Water Research **44**(15): 4351-4358.

Hubbell, S. P. (2001). The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32), Princeton University Press.

Huber, J. A., D. Mark Welch, H. G. Morrison, S. M. Huse, P. R. Neal, D. A. Butterfield and M. L. Sogin (2007). "Microbial population structures in the deep marine biosphere." Science **318**(5847): 97-100.

Hughes, K. A., I. W. Sutherland and M. V. Jones (1998). "Biofilm susceptibility to bacteriophage attack: the role of phage-borne polysaccharide depolymerase." Microbiology **144 (Pt 11)**: 3039-3047.

Hurwitz, B. L. and M. B. Sullivan (2013). "The Pacific Ocean Virome (POV): A Marine Viral Metagenomic Dataset and Associated Protein Clusters for Quantitative Viral Ecology." PLOS ONE **8**(2): e57355.

Hurwitz, B. L., L. Deng, B. T. Poulos and M. B. Sullivan (2013). "Evaluation of methods to concentrate and purify ocean virus communities through comparative, replicated metagenomics." Environmental Microbiology **15**(5): 1428-1440.

Jacquet, S., H. Havskum, T. F. Thingstad and D. Vaulot (2002). "Effects of inorganic and organic nutrient addition on a coastal microbial community (Isefjord, Denmark)." Marine Ecology Progress Series **228**: 3-14.

Jacquet, S., I. Domaizon, S. Personnic, A. S. P. Ram, M. Hedal, S. Duhamel and T. Sime-Ngando (2005). "Estimates of protozoan- and viral-mediated mortality of bacterioplankton in Lake Bourget (France)." Freshwater Biology **50**(4): 627-645.

Jacquet, S., T. Miki, R. Noble, P. Peduzzi and S. Wilhelm (2010). "Viruses in aquatic ecosystems: important advancements of the last 20 years and prospects for the future in the field of microbial oceanography and limnology." Advances in Oceanography and Limnology **1**(1): 97-141.

Janda, J. M. and S. L. Abbott (2010). "The Genus Aeromonas: Taxonomy, Pathogenicity, and Infection." Clinical Microbiology Reviews **23**(1): 35-73.

Jensen, E. C., H. S. Schrader, B. Rieland, T. L. Thompson, K. W. Lee, K. W. Nickerson and T. A. Kokjohn (1998). "Prevalence of Broad-Host-Range Lytic Bacteriophages of Sphaerotilus natans, Escherichia coli, and Pseudomonas aeruginosa." Applied and Environmental Microbiology **64**(2): 575-580.

Jiang, S. C. and J. H. Paul (1994). "Seasonal and diel abundance of viruses and occurrence of lysogeny/bacteriocinogeny in the marine-environment." Marine Ecology Progress Series **104**(1-2): 163-172.

Jiang, S. C. and J. H. Paul (1995). "Viral contribution to dissolved DNA in the marine environment as determined by differential centrifugation and kingdom probing." Applied and Environmental Microbiology **61**(1): 317-325.

John, S. G., C. B. Mendez, L. Deng, B. Poulos, A. K. M. Kauffman, S. Kern, J. Brum, M. F. Polz, E. A. Boyle and M. B. Sullivan (2011). "A simple and efficient method for concentration of ocean viruses by chemical flocculation." Environmental Microbiology Reports **3**(2): 195-202.

Jover, L. F., J. Romberg and J. S. Weitz (2016). "Inferring phage–bacteria infection networks from time-series data." Royal Society Open Science **3**(11).

Jover, L. F., M. H. Cortez and J. S. Weitz (2013). "Mechanisms of multi-strain coexistence in host–phage systems with nested infection networks." Journal of Theoretical Biology **332**: 65-77.

Juergen Gross and Uwe Ligges (2015). nortest: Tests for Normality. R package version 1.0-4. https://CRAN.R-project.org/package=nortest

Kaewpipat, K. and C. P. L. Grady (2002). "Microbial population dynamics in laboratory-scale activated sludge reactors." Water Science and Technology **46**(1-2): 19-27.

Keegan, K. P., E. M. Glass and F. Meyer (2016). "MG-RAST, a Metagenomics Service for Analysis of Microbial Community Structure and Function." Methods Mol Biol **1399**: 207-233.

Kenzaka, T., K. Tani and M. Nasu (2010). "High-frequency phage-mediated gene transfer in freshwater environments determined at single-cell level." Isme j **4**(5): 648-659.

Khairnar, K., P. Pal, R. H. Chandekar and W. N. Paunikar (2014). "Isolation and characterization of bacteriophages infecting nocardioforms in wastewater treatment plant". *Biotechnol Res Int*.

Khan, M. A., H. Satoh, H. Katayama, F. Kurisu and T. Mino (2002a). "Bacteriophages isolated from activated sludge processes and their polyvalency." Water Research **36**(13): 3364-3370.

Khan, M. A., H. Satoh, T. Mino, H. Katayama, F. Kurisu and T. Matsuo (2002b). "Bacteriophage-host interaction in the enhanced biological phosphate removing activated sludge system." Water Science and Technology **46**(1-2): 39-43.

Kim, K-H. and J-W. Bae (2011). "Amplification Methods Bias Metagenomic Libraries of Uncultured Single-Stranded and Double-Stranded DNA Viruses." Applied and Environmental Microbiology **77**(21): 7663-7668.

Kim, Y. M., D. S. Lee, C. Park, D. Park and J. M. Park (2011). "Effects of free cyanide on microbial communities and biological carbon and nitrogen removal performance in the industrial activated sludge process." Water Research **45**(3): 1267-1279.

Klausen, M. M., T. R. Thomsen, J. L. Nielsen, L. H. Mikkelsen and P. H. Nielsen (2004). "Variations in microcolony strength of probe-defined bacteria in activated sludge flocs." FEMS Microbiol Ecol **50**(2): 123-132.

Klein, D. (2002). "Quantification using real-time PCR technology: applications and limitations." Trends in Molecular Medicine **8**(6): 257-260.

Knowles, B. and F. Rohwer (2017). "Knowles & Rohwer reply." Nature **549**: E3.

Knowles, B., B. Bailey, L. Boling, M. Breitbart, A. Cobián-Güemes, J. del Campo, R. Edwards, B. Felts, J. Grasis, A. F. Haas, P. Katira, L. W. Kelly, A. Luque, J. Nulton, L. Paul, G. Peters, N. Robinett, S. Sandin, A. Segall, C. Silveira, M. Youle and F. Rohwer (2017). "Variability and host density independence in inductions-based estimates of environmental lysogeny." Nature Microbiology **2**: 17064.

Knowles, B., C. B. Silveira, B. A. Bailey, K. Barott, V. A. Cantu, A. G. Cobián-Güemes, F. H. Coutinho, E. A. Dinsdale, B. Felts, K. A. Furby, E. E. George, K. T. Green, G. B. Gregoracci, A. F. Haas, J. M.

Haggerty, E. R. Hester, N. Hisakawa, L. W. Kelly, Y. W. Lim, M. Little, A. Luque, T. McDole-Somera, K. McNair, L. S. de Oliveira, S. D. Quistad, N. L. Robinett, E. Sala, P. Salamon, S. E. Sanchez, S. Sandin, G. G. Z. Silva, J. Smith, C. Sullivan, C. Thompson, M. J. A. Vermeij, M. Youle, C. Young, B. Zgliczynski, R. Brainard, R. A. Edwards, J. Nulton, F. Thompson and F. Rohwer (2016). "Lytic to temperate switching of viral communities." Nature **531**: 466.

Korkmaz, S., D. Goksuluk and G. Zararsiz (2014). "MVN: An R Package for Assessing Multivariate Normality." The R Journal 6(2):151-162.

Korytowski, D. A. and H. L. Smith (2015). "How nested and monogamous infection networks in host-phage communities come to be." Theoretical Ecology **8**(1): 111-120.

Koskella, B., J. N. Thompson, G. M. Preston and A. Buckling (2011). "Local Biotic Environment Shapes the Spatial Scale of Bacteriophage Adaptation to Bacteria." American Naturalist **177**(4): 440-451.

Kotay, S. M., T. Datta, J. Choi and R. Goel (2011). "Biocontrol of biomass bulking caused by Haliscomenobacter hydrossis using a newly isolated lytic bacteriophage." Water Res **45**(2): 694-704.

Kott, Y. (1966). "Estimation of low numbers of Escherichia coli bacteriophage by use of the most probable number method." Appl Microbiol **14**(2): 141-144.

Kunin, V., S. He, F. Warnecke, S. B. Peterson, H. Garcia Martin, M. Haynes, N. Ivanova, L. L. Blackall, M. Breitbart, F. Rohwer, K. D. McMahon and P. Hugenholtz (2008). "A bacterial metapopulation adapts locally to phage predation despite global dispersal." Genome Res **18**(2): 293-297.

Kutter, E. and A. Sulakvelidze (2004). Bacteriophages: Biology and Applications, Taylor & Francis.

Labonté, J. M. and C. A. Suttle (2013). "Previously unknown and highly divergent ssDNA viruses populate the oceans." The Isme Journal **7**: 2169.

Labonte, J. M., B. K. Swan, B. Poulos, H. W. Luo, S. Koren, S. J. Hallam, M. B. Sullivan, T. Woyke, K. E. Wommack and R. Stepanauskas (2015). "Single-cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton." Isme Journal **9**(11): 2386-2399.

Lang, A. S., M. L. Rise, A. I. Culley and G. F. Steward (2009). "RNA viruses in the sea." FEMS Microbiology Reviews **33**(2): 295-323.

Lara, E., D. Vaqué, E. L. Sà, J. A. Boras, A. Gomes, E. Borrull, C. Díez-Vives, E. Teira, M. C. Pernice, F. C. Garcia, I. Forn, Y. M. Castillo, A. Peiró, G. Salazar, X. A. G. Morán, R. Massana, T. S. Catalá, G.

M. Luna, S. Agustí, M. Estrada, J. M. Gasol and C. M. Duarte (2017). "Unveiling the role and life strategies of viruses from the surface to the dark ocean." Science Advances **3**(9).

Larsen, A., T. Castberg, R. A. Sandaa, C. P. D. Brussaard, J. Egge, M. Heldal, A. Paulino, R. Thyrhaug, E. J. van Hannen and G. Bratbak (2001). "Population dynamics and diversity of phytoplankton, bacteria and viruses in a seawater enclosure." Marine Ecology-Progress Series **221**: 47-57.

Latta, L. C., M. Baker, T. Crowl, J. J. Parnell, B. Weimer, D. B. DeWald and M. E. Pfrender (2011). "Species and genotype diversity drive community and ecosystem properties in experimental microcosms." Evolutionary Ecology **25**(5): 1107-1125.

Law, Y., P. Lant and Z. Yuan (2011). "The effect of pH on N2O production under aerobic conditions in a partial nitritation system." Water Research **45**(18): 5934-5944.

Lawrence, A. W. and P. L. McCarty (1970). "Unified basis for biological treatment design and operation." Journal of the Sanitary Engineering Division **96**(3): 757-778.

Lee, S. H., K. Otawa, M. Onuki, H. Satoh and T. Mino (2007). "Population dynamics of phage-host system of Microlunatus phosphovorus indigenous in activated sludge." Journal of Microbiology and Biotechnology **17**(10): 1704-1707.

Legendre, P. (2014). lmodel2: Model II Regression. R package version 1.7-2. URL https://CRAN.R-project.org/package=lmodel2.

Li, W. K. W. and P. M. Dickie (2001). "Monitoring phytoplankton, bacterioplankton, and virioplankton in a coastal inlet (Bedford Basin) by flow cytometry." Cytometry **44**(3): 236-246.

Li, Y., H. Wang, K. Nie, C. Zhang, Y. Zhang, J. Wang, P. Niu and X. Ma (2016). "VIP: an integrated pipeline for metagenomics of virus identification and discovery." Scientific Reports **6**: 23774.

Liao, B. Q., D. G. Allen, G. G. Leppard, I. G. Droppo and S. N. Liss (2002). "Interparticle interactions affecting the stability of sludge flocs." J Colloid Interface Sci **249**(2): 372-380.

Liebhold, A., Koenig, W., & Bjørnstad, O. (2004). "Spatial Synchrony in Population Dynamics." Annual Review of Ecology, Evolution, and Systematics, ***35*,** 467-490.

Lim, J., H. Do, S. G. Shin and S. Hwang (2008). "Primer and probe sets for group-specific quantification of the genera Nitrosomonas and Nitrosospira using real-time PCR." Biotechnol Bioeng **99**(6): 1374-1383.

Lindell, D., M. B. Sullivan, Z. I. Johnson, A. C. Tolonen, F. Rohwer and S. W. Chisholm (2004). "Transfer of photosynthesis genes to and from Prochlorococcus viruses." Proc Natl Acad Sci U S A **101**(30): 11013-11018.

Liu, H., X. Yuan, J. Xu, P. J. Harrison, L. He and K. Yin (2015). "Effects of viruses on bacterial functions under contrasting nutritional conditions for four species of bacteria isolated from Hong Kong waters." Scientific Reports **5**: 14217.

Liu, M., J. J. Gill, R. Young and E. J. Summer (2015). "Bacteriophages of wastewater foaming-associated filamentous Gordonia reduce host levels in raw activated sludge." Scientific Reports **5**: 13754.

Lopez Pascua, L., S. Gandon and A. Buckling (2012). "Abiotic heterogeneity drives parasite local adaptation in coevolving bacteria and phages." Journal of Evolutionary Biology **25**(1): 187-195.

Lotka, A. J. (1934). Théorie analytique des associations biologiques. Paris, Hermann.

Louca, S. and M. Doebeli (2017). "Taxonomic variability and functional stability in microbial communities infected by phages." Environmental Microbiology **19**(10): 3863-3878.

Lü, F., P. J. He, L. M. Shao and D. J. Lee (2008). "Stress of pH and acetate on product formation of fermenting polysaccharide-rich organic waste." Biochemical Engineering Journal **39**(1): 97-104.

Marie, D., C. P. D. Brussaard, R. Thyrhaug, G. Bratbak and D. Vaulot (1999). "Enumeration of marine viruses in culture and natural samples by flow cytometry." Applied and Environmental Microbiology **65**(1): 45-52.

Markowitz, V. M., I. M. A. Chen, K. Palaniappan, K. Chu, E. Szeto, Y. Grechkin, A. Ratner, B. Jacob, J. Huang, P. Williams, M. Huntemann, I. Anderson, K. Mavromatis, N. N. Ivanova and N. C. Kyrpides (2012). "IMG: the integrated microbial genomes database and comparative analysis system." Nucleic Acids Research **40**(D1): D115-D122.

Martin, T.D., J.T. Creed and C.A. Brockhoff (1994). Sample preparation procedure for spectrochemcial determination of total recoverable elements: U.S. Environmental Protection Agency Report, Revision 2.8, EMMC Version, 12 p.

Martins, A. M. P., K. Pagilla, J. J. Heijnen and M. C. M. van Loosdrecht (2004). "Filamentous bulking sludge—a critical review." Water Research **38**(4): 793-817.

Maurice, C. F., T. Bouvier, J. Comte, F. Guillemette and P. A. del Giorgio (2010). "Seasonal variations of phage life strategies and bacterial physiological states in three northern temperate lakes." Environmental Microbiology **12**(3): 628-641.

Mazaheri Nezhad Fard, R., M. D. Barton and M. W. Heuzenroeder (2011). "Bacteriophage-mediated transduction of antibiotic resistance in enterococci." Lett Appl Microbiol **52**(6): 559-564.

McDaniel, L. D., E. Young, J. Delaney, F. Ruhnau, K. B. Ritchie and J. H. Paul (2010). "High Frequency of Horizontal Gene Transfer in the Oceans." Science **330**(6000): 50.

McMahon, K. D., H. G. Martin and P. Hugenholtz (2007). "Integrating ecology into biotechnology." Current Opinion in Biotechnology **18**(3): 287-292.

McTavish, H., J. A. Fuchs and A. B. Hooper (1993). "Sequence of the gene coding for ammonia monooxygenase in Nitrosomonas europaea." Journal of Bacteriology **175**(8): 2436-2444.

Michael, S. R. a. and J. G. Stephen (2003). "The Uncultured Microbial Majority." Annual Review of Microbiology **57**(1): 369-394.

Michen, B. and T. Graule (2010). "Isoelectric points of viruses." J Appl Microbiol **109**(2): 388-397.

Middelboe, M. (2000). "Bacterial growth rate and marine virus-host dynamics." Microbial Ecology **40**(2): 114-124.

Middelboe, M. and N. O. G. Jorgensen (2006). "Viral lysis of bacteria: an important source of dissolved amino acids and cell wall compounds." Journal of the Marine Biological Association of the United Kingdom **86**(3): 605-612.

Middelboe, M., K. Holmfeldt, L. Riemann, O. Nybroe and J. Haaber (2009). "Bacteriophages drive strain diversification in a marine Flavobacterium: implications for phage resistance and physiological properties." Environ Microbiol **11**(8): 1971-1982.

Middelboe, M., N. O. G. Jorgensen and N. Kroer (1996). "Effects of viruses on nutrient turnover and growth efficiency of noninfected marine bacterioplankton." Applied and Environmental Microbiology **62**(6): 1991-1997.

Miralles, R., A. Moya and S. F. Elena (1999). "Effect of population patchiness and migration rates on the adaptation and divergence of vesicular stomatitis virus quasispecies populations." Journal of General Virology **80**: 2051-2059.

Mizuno, C. M., F. Rodriguez-Valera, N. E. Kimes and R. Ghai (2013). "Expanding the Marine Virosphere Using Metagenomics." PLoS Genetics **9**(12): e1003987.

Morgan, A. D., S. Gandon and A. Buckling (2005). "The effect of migration on local adaptation in a coevolving host-parasite system." Nature **437**(7056): 253-256.

Morgan, A., M. Brockhurst, L. Lopez-Pascua, C. Pal and A. Buckling (2007). "Differential impact of simultaneous migration on coevolving hosts and parasites." BMC Evolutionary Biology **7**(1): 1.

Morgan-Sagastume, F. and D. G. Allen (2003). "Effects of temperature transient conditions on aerobic biological treatment of wastewater." Water Research **37**(15): 3590-3601.

Morgan-Sagastume, F. and D. G. Allen (2005a). "Physicochemical properties and stability of activated sludge flocs under temperature upshifts from 30 to 45 °C." Journal of Colloid and Interface Science **281**(1): 136-145.

Morgan-Sagastume, F. and D. Grant Allen (2005b). "Activated sludge deflocculation under temperature upshifts from 30 to 45 °C." Water Research **39**(6): 1061-1074.

Motegi, C. and T. Nagata (2007). "Enhancement of viral production by addition of nitrogen or nitrogen plus carbon in subtropical surface waters of the South Pacific." Aquatic Microbial Ecology **48**(1): 27-34.

Motlagh, A. M., A. S. Bhattacharjee and R. Goel (2015). "Microbiological study of bacteriophage induction in the presence of chemical stress factors in enhanced biological phosphorus removal (EBPR)." Water Res **81**: 1-14.

Muniesa, M., M. Colomer-Lluch and J. Jofre (2013). "Potential impact of environmental bacteriophages in spreading antibiotic resistance genes." Future Microbiol **8**(6): 739-751.

Munz, G., C. Lubello and J. A. Oleszkiewicz (2011). "Factors affecting the growth rates of ammonium and nitrite oxidizing bacteria." Chemosphere **83**(5): 720-725.

Muyzer, G., E. C. de Waal and A. G. Uitterlinden (1993). "Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA." Applied and Environmental Microbiology **59**(3): 695-700.

Nadarajah, N., D. Grant Allen and R. R. Fulthorpe (2007). "Effects of transient temperature conditions on the divergence of activated sludge bacterial community structure and function." Water Research **41**(12): 2563-2571.

Naeem, S. and S. Li (1997). "Biodiversity enhances ecosystem reliability." Nature **390**(6659): 507-509.

Nagasaki, K. (2008). "Dinoflagellates, diatoms, and their viruses." Journal of Microbiology **46**(3): 235-243.

Namura, M., T. Hijikata, K. Miyanaga and Y. Tanji (2008). "Detection of Escherichia coli with fluorescent labeled phages that have a broad host range to E. coli in sewage water." Biotechnol Prog **24**(2): 481-486.

Nap, Rikkert J., Anže L. Božič, I. Szleifer and R. Podgornik (2014). "The Role of Solution Conditions in the Bacteriophage PP7 Capsid Charge Regulation." Biophysical Journal **107**(8): 1970-1979.

Nielsen, K. M., P. J. Johnsen, D. Bensasson and D. Daffonchio (2007). "Release and persistence of extracellular DNA in the environment." Environmental Biosafety Research **6**(1-2): 37-53.

Noble, R. T. and J. A. Fuhrman (1998). "Use of SYBR Green I for rapid epifluorescence counts of marine viruses and bacteria." Aquatic Microbial Ecology **14**(2): 113-118.

Nordberg, H., M. Cantor, S. Dusheyko, S. Hua, A. Poliakov, I. Shabalov, T. Smirnova, I. V. Grigoriev and I. Dubchak (2014). "The genome portal of the Department of Energy Joint Genome Institute: 2014 updates." Nucleic Acids Res **42**(Database issue): D26-31.

Nordgren, J., A. Matussek, A. Mattsson, L. Svensson and P.-E. Lindgren (2009). "Prevalence of norovirus and factors influencing virus concentrations during one year in a full-scale wastewater treatment plant." Water Research **43**(4): 1117-1125.

Norman, G. R. and D. L. Streiner (2008). Biostatistics: The Bare Essentials, B.C. Decker.

Norton, J., J. Alzerreca, Y. Suwa and M. Klotz (2002). "Diversity of ammonia monooxygenase operon in autotrophic ammonia-oxidizing bacteria." Archives of Microbiology **177**(2): 139-149.

Ofiteru, I. D., M. Bellucci, C. Picioreanu, V. Lavric and T. P. Curtis (2015). "Multi-scale modelling of bioreactor-separator system for wastewater treatment with two-dimensional activated sludge floc dynamics." Water Research **50**: 382-395.

Ofiteru, I. D., M. Lunn, T. P. Curtis, G. F. Wells, C. S. Criddle, C. A. Francis and W. T. Sloan (2010). "Combined niche and neutral effects in a microbial wastewater treatment community." Proceedings of the National Academy of Sciences of the United States of America **107**(35): 15345-15350.

Ogata, S., H. Miyamoto and S. Hayashida (1980). "An investigation of the influence of bacteriophages on the bacterial-flora and purification powers of activated-sludge." Journal of General and Applied Microbiology **26**(2): 97-108.

Ogle, D.H. (2017). FSA: Fisheries Stock Analysis. R package version 0.8.17.

Oksanen, J., F. Guillaume Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. Henry, H. Stevens, E. Szoecs and H. Wagner (2017). vegan: Community Ecology Package. R package version R package version 2.4-4. https://CRAN.R-project.org/package=vegan

Otawa, K., S. H. Lee, A. Yamazoe, M. Onuki, H. Satoh and T. Mino (2007). "Abundance, diversity, and dynamics of viruses on microorganisms in activated sludge processes." Microbial Ecology **53**(1): 143-152.

Overbeek, R., T. Begley, R. M. Butler, J. V. Choudhuri, H.-Y. Chuang, M. Cohoon, V. de Crécy-Lagard, N. Diaz, T. Disz, R. Edwards, M. Fonstein, E. D. Frank, S. Gerdes, E. M. Glass, A. Goesmann, A. Hanson, D. Iwata-Reuyl, R. Jensen, N. Jamshidi, L. Krause, M. Kubal, N. Larsen, B. Linke, A. C. McHardy, F. Meyer, H. Neuweger, G. Olsen, R. Olson, A. Osterman, V. Portnoy, G. D. Pusch, D. A. Rodionov, C. Rückert, J. Steiner, R. Stevens, I. Thiele, O. Vassieva, Y. Ye, O. Zagnitko and V. Vonstein (2005). "The Subsystems Approach to Genome Annotation and its Use in the Project to Annotate 1000 Genomes." Nucleic Acids Research **33**(17): 5691-5702.

Øvreås, L., D. Bourne, R. A. Sandaa, E. O. Casamayor, S. Benlloch, V. Goddard, G. Smerdon, M. Heldal and T. F. Thingstad (2003). "Response of bacterial and viral communities to nutrient manipulations in seawater mesocosms." Aquatic Microbial Ecology **31**(2): 109-121.

Palmer, S. J. (2010). "Future challenges to asset investment in the UK water industry: the wastewater asset investment risk mitigation offered by minimising principal operating cost risks." Journal of Water and Climate Change **1**(1): 17-35.

Park, M. O., H. Ikenaga and K. Watanabe (2007). "Phage diversity in a methanogenic digester." Microbial Ecology **53**(1): 98-103.

Parsley, L. C., E. J. Consuegra, K. S. Kakirde, A. M. Land, W. F. Harper and M. R. Liles (2010a). "Identification of Diverse Antimicrobial Resistance Determinants Carried on Bacterial, Plasmid, or Viral Metagenomes from an Activated Sludge Microbial Assemblage." Applied and Environmental Microbiology **76**(11): 3753-3757.

Parsley, L. C., E. J. Consuegra, S. J. Thomas, J. Bhavsar, A. M. Land, N. N. Bhuiyan, M. A. Mazher, R. J. Waters, K. E. Wommack, W. F. Harper, Jr. and M. R. Liles (2010b). "Census of the Viral Metagenome within an Activated Sludge Microbial Assemblage." Applied and Environmental Microbiology **76**(8): 2673-2677.

Paterson, S., T. Vogwill, A. Buckling, R. Benmayor, A. J. Spiers, N. R. Thomson, M. Quail, F. Smith, D. Walker, B. Libberton, A. Fenton, N. Hall and M. A. Brockhurst (2010). "Antagonistic coevolution accelerates molecular evolution." Nature **464**(7286): 275-U154.

Paul, J. H. (2008). "Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas?" The Isme Journal **2**: 579.

Payet, J. P. and C. A. Suttle (2013). "To kill or not to kill: The balance between lytic and lysogenic viral infection is driven by trophic status." Limnology and Oceanography **58**(2): 465-474.

Petsch, D. K., G. D. Pinha, J. D. Dias and A. M. Takeda (2015). "Temporal nestedness in Chironomidae and the importance of environmental and spatial factors in species rarity." Hydrobiologia **745**(1): 181-193.

Pham, M., E. A. Mintz and T. H. Nguyen (2009). "Deposition kinetics of bacteriophage MS2 to natural organic matter: Role of divalent cations." Journal of Colloid and Interface Science **338**(1): 1-9.

Philips, S. and W. Verstraete (2001). "Effect of repeated addition of nitrite to semi-continuous activated sludge reactors." Bioresour Technol **80**(1): 73-82.

Pholchan, M. K., J. d. C. Baptista, R. J. Davenport, W. T. Sloan and T. P. Curtis (2013). "Microbial community assembly, theory and rare functions." Frontiers in microbiology **4**: 68-68.

Poorvin, L., J. M. Rinta-Kanto, D. A. Hutchins and S. W. Wilhelm (2004). "Viral release of iron and its bioavailability to marine plankton." Limnology and Oceanography **49**(5): 1734-1741.

Proctor, L. M., A. Okubo and J. A. Fuhrman (1993). "Calibrating estimates of phage-induced mortality in marine bacteria: Ultrastructural studies of marine bacteriophage development from one-step growth experiments." Microb Ecol **25**(2): 161-182.

Props, R., F.-M. Kerckhof, P. Rubbens, J. De Vrieze, E. Hernandez Sanabria, W. Waegeman, P. Monsieurs, F. Hammes and N. Boon (2017). "Absolute quantification of microbial taxon abundances." ISME J **11**(2): 584-587.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Ramette, A. (2007). "Multivariate analyses in microbial ecology." FEMS Microbiology Ecology **62**(2): 142-160.

Rappe, M. S. and S. J. Giovannoni (2003). "The uncultured microbial majority." Annu Rev Microbiol **57**: 369-394.

Replicon, J., A. Frankfater and R. V. Miller (1995). "A continuous-culture model to examine factors that affect transduction among pseudomonas-aeruginosa strains in fresh-water environments." Applied and Environmental Microbiology **61**(9): 3359-3366.

Reshef, D. N., Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher and P. C. Sabeti (2011). "Detecting novel associations in large data sets." Science **334**(6062): 1518-1524.

Rho, M., Y. W. Wu, H. X. Tang, T. G. Doak and Y. Z. Ye (2012). "Diverse CRISPRs Evolving in Human Microbiomes." Plos Genetics **8**(6).

Richard, E. L. and R. L. Bruce (1985). "Constraints on the Coevolution of Bacteria and Virulent Phage: A Model, Some Experiments, and Predictions for Natural Communities." The American Naturalist **125**(4): 585-602.

Rodriguez-Brito, B., L. Li, L. Wegley, M. Furlan, F. Angly, M. Breitbart, J. Buchanan, C. Desnues, E. Dinsdale, R. Edwards, B. Felts, M. Haynes, H. Liu, D. Lipson, J. Mahaffy, A. B. Martin-Cuadrado, A. Mira, J. Nulton, L. Pasic, S. Rayhawk, J. Rodriguez-Mueller, F. Rodriguez-Valera, P. Salamon, S. Srinagesh, T. F. Thingstad, T. Tran, R. V. Thurber, D. Willner, M. Youle and F. Rohwer (2010). "Viral and microbial community dynamics in four aquatic environments." ISME J **4**(6): 739-751.

Rodriguez-Valera, F., A.-B. Martin-Cuadrado, B. Rodriguez-Brito, L. Pasic, T. F. Thingstad, F. Rohwer and A. Mira (2009). "OPINION Explaining microbial population genomics through phage predation." Nature Reviews Microbiology **7**(11): 828-836.

Rodriguez-Valera, F., C. M. Mizuno and R. Ghai (2014). "Tales from a thousand and one phages." Bacteriophage **4**(2): e28265.

Rohwer, F. and R. V. Thurber (2009). "Viruses manipulate the marine environment." Nature **459**: 207.

Rosenberg, E., G. Bittan-Banin, G. Sharon, A. Shon, G. Hershko, I. Levy and E. Z. Ron (2010). "The phage-driven microbial loop in petroleum bioremediation." Microbial Biotechnology **3**(4): 467-472.

Rotthauwe, J. H., K. P. Witzel and W. Liesack (1997). "The ammonia monooxygenase structural gene amoA as a functional marker: Molecular fine-scale analysis of natural ammonia-oxidizing populations." Applied and Environmental Microbiology **63**(12): 4704-4712.

Roux, S., A. K. Hawley, M. Torres Beltran, M. Scofield, P. Schwientek, R. Stepanauskas, T. Woyke, S. J. Hallam and M. B. Sullivan (2014a) "Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta-genomics." eLife **3**, e03125 DOI: 10.7554/elife.03125.

Roux, S., J. Tournayre, A. Mahul, D. Debroas and F. Enault (2014b). "Metavir 2: new tools for viral metagenome comparison and assembled virome analysis." BMC Bioinformatics **15**(1): 76.

Ruan, Q., D. Dutta, M. S. Schwalbach, J. A. Steele, J. A. Fuhrman and F. Sun (2006). "Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors." Bioinformatics **22**(20): 2532-2538.

Sabour, P. M. and M. Griffiths (2010). Bacteriophages in the Control of Food- And Waterborne Pathogens, ASM Press.

Saikaly, P. E. and D. B. Oerther (2011). "Diversity of Dominant Bacterial Taxa in Activated Sludge Promotes Functional Resistance following Toxic Shock Loading." Microbial Ecology **61**(3): 557-567.

Sandaa, R. A., L. Gomez-Consarnau, J. Pinhassi, L. Riemann, A. Malits, M. G. Weinbauer, J. M. Gasol and T. F. Thingstad (2009). "Viral control of bacterial biodiversity - evidence from a nutrient-enriched marine mesocosm experiment." Environmental Microbiology **11**(10): 2585-2597.

Sander, M. and H. Schmieger (2001). "Method for host-independent detection of generalized transducing bacteriophages in natural habitats." Applied and Environmental Microbiology **67**(4): 1490-1493.

Saunders, A. M., M. Albertsen, J. Vollertsen and P. H. Nielsen (2015). "The activated sludge ecosystem contains a core community of abundant organisms." The Isme Journal **10**: 11.

Sayers, E. W., T. Barrett, D. A. Benson, S. H. Bryant, K. Canese, V. Chetvernin, D. M. Church, M. DiCuccio, R. Edgar, S. Federhen, M. Feolo, L. Y. Geer, W. Helmberg, Y. Kapustin, D. Landsman, D. J. Lipman, T. L. Madden, D. R. Maglott, V. Miller, I. Mizrachi, J. Ostell, K. D. Pruitt, G. D. Schuler, E. Sequeira, S. T. Sherry, M. Shumway, K. Sirotkin, A. Souvorov, G. Starchenko, T. A. Tatusova, L.

Wagner, E. Yaschenko and J. Ye (2009). "Database resources of the National Center for Biotechnology Information." Nucleic Acids Res **37**(Database issue): D5-15.

Schaldach, C. M., W. L. Bourcier, H. F. Shaw, B. E. Viani and W. D. Wilson (2006). "The influence of ionic strength on the interaction of viruses with charged surfaces under environmental conditions." J Colloid Interface Sci **294**(1): 1-10.

Schmidt, I., O. Sliekers, M. Schmid, E. Bock, J. Fuerst, J. G. Kuenen, M. S. Jetten and M. Strous (2003). "New concepts of microbial treatment processes for the nitrogen removal in wastewater." FEMS Microbiol Rev **27**(4): 481-492.

Schulz, F., N. Yutin, N. N. Ivanova, D. R. Ortega, T. K. Lee, J. Vierheilig, H. Daims, M. Horn, M. Wagner, G. J. Jensen, N. C. Kyrpides, E. V. Koonin and T. Woyke (2017). "Giant viruses with an expanded complement of translation system components." Science **356**(6333): 82-85.

Schwalbach, M. S., I. Hewson and J. A. Fuhrman (2004). "Viral effects on bacterial community composition in marine plankton microcosms." Aquatic Microbial Ecology **34**(2): 117-127.

Seemann, T. (2014). "Prokka: rapid prokaryotic genome annotation." Bioinformatics **30**(14): 2068-2069.

Seviour, R. J., T. Mino and M. Onuki (2003). "The microbiology of biological phosphorus removal in activated sludge systems." FEMS Microbiology Reviews **27**(1): 99-127.

Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker (2003). "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks." Genome Research **13**(11): 2498-2504.

Shapiro, O. H. and A. Kushmaro (2011). "Bacteriophage ecology in environmental biotechnology processes." Current Opinion in Biotechnology **22**(3): 449-455.

Shapiro, O. H., A. Kushmaro and A. Brenner (2010). "Bacteriophage predation regulates microbial abundance and diversity in a full-scale bioreactor treating industrial wastewater." Isme Journal **4**(3): 327-336.

Shelford, E. J., M. Middelboe, E. F. Moller and C. A. Suttle (2012). "Virus-driven nitrogen cycling enhances phytoplankton growth." Aquatic Microbial Ecology **66**(1): 41-46.

Shoener, B. D., I. M. Bradley, R. D. Cusick and J. S. Guest (2014). "Energy positive domestic wastewater treatment: the roles of anaerobic and phototrophic technologies." Environmental Science: Processes & Impacts **16**(6): 1204-1222.

Siripong, S. and B. E. Rittmann (2007). "Diversity study of nitrifying bacteria in full-scale municipal wastewater treatment plants." Water Research **41**(5): 1110-1120.

Sloan, W. T., M. Lunn, S. Woodcock, I. M. Head, S. Nee and T. P. Curtis (2006). "Quantifying the roles of immigration and chance in shaping prokaryote community structure." Environmental Microbiology **8**(4): 732-740.

Sloan, W. T., S. Woodcock, M. Lunn, I. M. Head and T. P. Curtis (2007). "Modeling taxa-abundance distributions in microbial communities using environmental sequence data." Microbial Ecology **53**(3): 443-455.

Soler, N., E. Marguet, J.-M. Verbavatz and P. Forterre (2008). "Virus-like vesicles and extracellular DNA produced by hyperthermophilic archaea of the order Thermococcales." Research in Microbiology **159**(5): 390-399.

Sorek, R., V. Kunin and P. Hugenholtz (2008). "CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea." Nat Rev Microbiol **6**(3): 181-186.

Stasinakis, A. S., N. S. Thomaidis, D. Mamais, E. C. Papanikolaou, A. Tsakon and T. D. Lekkas (2003). "Effects of chromium (VI) addition on the activated sludge process." Water Research **37**(9): 2140-2148.

Stephen, J. R., Y. J. Chang, S. J. Macnaughton, G. A. Kowalchuk, K. T. Leung, C. A. Flemming and D. C. White (1999). "Effect of toxic metals on indigenous soil beta-subgroup proteobacterium ammonia oxidizer community structure and protection against toxicity by inoculated metal-resistant bacteria." Appl Environ Microbiol **65**(1): 95-101.

Steward, G. F., A. I. Culley, J. A. Mueller, E. M. Wood-Charlson, M. Belcaid and G. Poisson (2013). "Are we missing half of the viruses in the ocean?" Isme Journal **7**(3): 672-679.

Stirling, G. and B. Wilsey (2001). "Empirical Relationships between Species Richness, Evenness, and Proportional Diversity." Am Nat **158**(3): 286-299.

Storey, J. D. and R. Tibshirani (2003). "Statistical significance for genomewide studies." Proceedings of the National Academy of Sciences **100**(16): 9440-9445.

Sullivan, M. B., D. Lindell, J. A. Lee, L. R. Thompson, J. P. Bielawski and S. W. Chisholm (2006). "Prevalence and Evolution of Core Photosystem II Genes in Marine Cyanobacterial Viruses and Their Hosts." PLoS Biol **4**(8): e234.

Sullivan, M. B., J. B. Waterbury and S. W. Chisholm (2003). "Cyanophages infecting the oceanic cyanobacterium Prochlorococcus." Nature **424**: 1047.

Sutherland, I. W., K. A. Hughes, L. C. Skillman and K. Tait (2004). "The interaction of phage and biofilms." FEMS Microbiology Letters **232**(1): 1-6.

Suttle, C. A. (2007). "Marine viruses - major players in the global ecosystem." Nature Reviews Microbiology **5**(10): 801-812.

Suzuki, M. T., L. T. Taylor and E. F. DeLong (2000). "Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5'-nuclease assays." Appl Environ Microbiol **66**(11): 4605-4614.

Synnott, A. J., Y. Kuang, M. Kurimoto, K. Yamamichi, H. Iwano and Y. Tanji (2009). "Isolation from Sewage Influent and Characterization of Novel Staphylococcus aureus Bacteriophages with Wide Host Ranges and Potent Lytic Capabilities." Applied and Environmental Microbiology **75**(13): 4483-4490.

Tadmor, A. D., E. A. Ottesen, J. R. Leadbetter and R. Phillips (2011). "Probing Individual Environmental Bacteria for Viruses by Using Microfluidic Digital PCR." Science **333**(6038): 58-62.

Talbot, G., E. Topp, M. F. Palin and D. I. Massé (2008). "Evaluation of molecular methods used for establishing the interactions and functions of microorganisms in anaerobic bioreactors." Water Research **42**(3): 513-537.

Tamaki, H., R. Zhang, F. E. Angly, S. Nakamura, P.-Y. Hong, T. Yasunaga, Y. Kamagata and W.-T. Liu (2012). "Metagenomic analysis of DNA viruses in a wastewater treatment plant in tropical climate." Environmental Microbiology **14**(2): 441-452.

Tariq, M. A., F. L. C. Everest, L. A. Cowley, A. De Soyza, G. S. Holt, S. H. Bridge, A. Perry, J. D. Perry, S. J. Bourke, S. P. Cummings, C. V. Lanyon, J. J. Barr and D. L. Smith (2015). "A metagenomic approach to characterize temperate bacteriophage populations from Cystic Fibrosis and non-Cystic Fibrosis bronchiectasis patients." Frontiers in Microbiology **6**(97).

Tchobanoglous, G., F. L. Burton, H. D. Stensel, Metcalf and I. Eddy (2003). Wastewater Engineering: Treatment and Reuse, McGraw-Hill Education.

Thingstad, T. F. (2000). "Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems." Limnology and Oceanography **45**(6): 1320-1328.

Thingstad, T. F. and R. Lignell (1997). "Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand." Aquatic Microbial Ecology **13**(1): 19-27.

Thingstad, T. F., B. Pree, J. Giske and S. Våge (2015). "What difference does it make if viruses are strain-, rather than species-specific?" Frontiers in Microbiology **6**(320).

Thingstad, T. F., S. Våge, J. E. Storesund, R.-A. Sandaa and J. Giske (2014). "A theoretical analysis of how strain-specific viruses can control microbial species diversity." Proceedings of the National Academy of Sciences **111**(21): 7813-7818.

Thomas, J. A., J. A. Soddell and D. I. Kurtboke (2002). "Fighting foam with phages?" Water Sci Technol **46**(1-2): 511-518.

Tilman, D. (2004). "Niche tradeoffs, neutrality, and community structure: A stochastic theory of resource competition, invasion, and community assembly." Proceedings of the National Academy of Sciences of the United States of America **101**(30): 10854-10861.

Tilman, D., J. Knops, D. Wedin, P. Reich, M. Ritchie and E. Siemann (1997). "The influence of functional diversity and composition on ecosystem processes." Science **277**(5330): 1300-1302.

Tixier, N., G. Guibaud and M. Baudu (2003). "Effect of pH and ionic environment changes on interparticle interactions affecting activated sludge flocs: a rheological approach." Environ Technol **24**(8): 971-978.

Tomaru, Y. and K. Nagasaki (2007). "Flow cytometric detection and enumeration of DNA and RNA viruses infecting marine eukaryotic microalgae." Journal of Oceanography **63**(2): 215-221.

Torrella, F. and R. Y. Morita (1979). "Evidence by electron micrographs for a high incidence of bacteriophage particles in the waters of Yaquina Bay, oregon: ecological and taxonomical implications." Appl Environ Microbiol **37**(4): 774-778.

Tucker, K. P., R. Parsons, E. M. Symonds and M. Breitbart (2011). "Diversity and distribution of single-stranded DNA phages in the North Atlantic Ocean." Isme j **5**(5): 822-830.

Vadivelu, V. M., J. Keller and Z. Yuan (2006). "Effect of free ammonia and free nitrous acid concentration on the anabolic and catabolic processes of an enriched Nitrosomonas culture." Biotechnology and Bioengineering **95**(5): 830-839.

Valentín-Vargas, A., G. Toro-Labrador and A. A. Massol-Deyá (2012). "Bacterial Community Dynamics in Full-Scale Activated Sludge Bioreactors: Operational and Ecological Factors Driving Community Assembly and Performance." Plos One **7**(8).

van der Gast, C. J., A. S. Whiteley and I. P. Thompson (2004). "Temporal dynamics and degradation activity of an bacterial inoculum for treating waste metal-working fluid." Environ Microbiol **6**(3): 254-263.

van der Gast, C. J., D. Ager and A. K. Lilley (2008). "Temporal scaling of bacterial taxa is influenced by both stochastic and deterministic ecological factors." Environmental Microbiology **10**(6): 1411-1418.

van Houte, S., A. Buckling and E. R. Westra (2016). "Evolutionary Ecology of Prokaryotic Immune Mechanisms." Microbiology and Molecular Biology Reviews **80**(3): 745-763.

Vanwonterghem, I., P. D. Jensen, P. G. Dennis, P. Hugenholtz, K. Rabaey and G. W. Tyson (2014). "Deterministic processes guide long-term synchronised population dynamics in replicate anaerobic digesters." Isme Journal **8**(10): 2015-2028.

Venables, W. N. and B. D. Ripley (2002). Modern applied statistics with S. New York, Springer.

Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y.-H. Rogers and H. O. Smith (2004). "Environmental Genome Shotgun Sequencing of the Sargasso Sea." Science **304**(5667): 66-74.

Vogwill, T., A. Fenton and M. A. Brockhurst (2008). "The impact of parasite dispersal on antagonistic host–parasite coevolution." Journal of Evolutionary Biology **21**(5): 1252-1258.

Volterra, V. (1926). Variazioni e fluttuazioni del numero d'individui in specie animali conviventi. Città di Castello.

Vos, M., P. J. Birkett, E. Birch, R. I. Griffiths and A. Buckling (2009). "Local Adaptation of Bacteriophages to Their Bacterial Hosts in Soil." Science **325**(5942): 833-833.

Vuono, D. C., J. Benecke, J. Henkel, W. C. Navidi, T. Y. Cath, J. Munakata-Marr, J. R. Spear and J. E. Drewes (2014). "Disturbance and temporal partitioning of the activated sludge metacommunity." The Isme Journal **9**: 425.

Wagner, P. L. and M. K. Waldor (2002). "Bacteriophage control of bacterial virulence." Infect Immun **70**(8): 3985-3993.

Wang, X., M. Hu, Y. Xia, X. Wen and K. Ding (2012b). "Pyrosequencing Analysis of Bacterial Diversity in 14 Wastewater Treatment Systems in China." Applied and Environmental Microbiology **78**(19): 7042-7047.

Wang, X., X. Wen, Y. Xia, M. Hu, F. Zhao and K. Ding (2012a). "Ammonia Oxidizing Bacteria Community Dynamics in a Pilot-Scale Wastewater Treatment Plant." Plos One **7**(4).

Weinbauer, M. G. (2004). "Ecology of prokaryotic viruses." Fems Microbiology Reviews **28**(2): 127-181.

Weinbauer, M. G. and C. A. Suttle (1997). "Comparison of epifluorescence and transmission electron microscopy for counting viruses in natural marine waters." Aquatic Microbial Ecology **13**(3): 225-232.

Weinbauer, M. G. and F. Rassoulzadegan (2004). "Are viruses driving microbial diversification and diversity?" Environmental Microbiology **6**(1): 1-11.

Weinbauer, M. G., D. Fuks, S. Puskaric and P. Peduzzi (1995). "Diel, seasonal, and depth-related variability of viruses and dissolved DNA in the northern Adriatic Sea." Microb Ecol **30**(1): 25-41.

Weitz, J. S., C. A. Stock, S. W. Wilhelm, L. Bourouiba, M. L. Coleman, A. Buchan, M. J. Follows, J. A. Fuhrman, L. F. Jover, J. T. Lennon, M. Middelboe, D. L. Sonderegger, C. A. Suttle, B. P. Taylor, T. Frede Thingstad, W. H. Wilson and K. Eric Wommack (2015). "A multitrophic model to quantify the effects of marine viruses on microbial food webs and ecosystem processes." Isme j **9**(6): 1352-1364.

Weitz, J. S., S. J. Beckett, J. R. Brum, B. B. Cael and J. Dushoff (2017). "Lysis, lysogeny and virus–microbe ratios." Nature **549**: E1.

Wells, G. F., H. D. Park, B. Eggleston, C. A. Francis and C. S. Criddle (2011). "Fine-scale bacterial community dynamics and the taxa-time relationship within a full-scale activated sludge bioreactor." Water Research **45**(17): 5476-5488.

Wells, G. F., H. D. Park, C. H. Yeung, B. Eggleston, C. A. Francis and C. S. Criddle (2009). "Ammonia-oxidizing communities in a highly aerated full-scale activated sludge bioreactor: betaproteobacterial dynamics and low relative abundance of Crenarchaea." Environmental Microbiology **11**(9): 2310-2328.

Whang, L.-M., I. C. Chien, S.-L. Yuan and Y.-J. Wu (2009). "Nitrifying community structures and nitrification performance of full-scale municipal and swine wastewater treatment plants." Chemosphere **75**(2): 234-242.

White, P. S. and S. T. A. Pickett (1985). "Natural disturbance and patch dynamics: an introduction." The ecology of natural disturbance and patch dynamics: 3-13.

Wick, R. R., M. B. Schultz, J. Zobel and K. E. Holt (2015). "Bandage: interactive visualization of de novo genome assemblies." Bioinformatics **31**(20): 3350-3352.

Wigington, C. H., D. Sonderegger, C. P. D. Brussaard, A. Buchan, J. F. Finke, J. A. Fuhrman, J. T. Lennon, M. Middelboe, C. A. Suttle, C. Stock, W. H. Wilson, K. E. Wommack, S. W. Wilhelm and J. S. Weitz (2016). "Re-examination of the relationship between marine virus and microbial cell abundances." Nature Microbiology **1**: 15024.

Wilén, B.-M., B. Jin and P. Lant (2003). "The influence of key chemical constituents in activated sludge on surface and flocculating properties." Water Research **37**(9): 2127-2139.

Wilhelm, S. W. and C. A. Suttle (1999). "Viruses and Nutrient Cycles in the Sea - Viruses play critical roles in the structure and function of aquatic food webs." Bioscience **49**(10): 781-788.

Wilkinson, D. J. (2011). Stochastic Modelling for Systems Biology, Second Edition, CRC Press.

Williamson, S. J. and J. H. Paul (2004). "Nutrient stimulation of lytic phage production in bacterial populations of the Gulf of Mexico." Aquatic Microbial Ecology **36**(1): 9-17.

Winter, C., A. Smit, G. J. Herndl and M. G. Weinbauer (2004). "Impact of virioplankton on archaeal and bacterial community richness as assessed in seawater batch cultures." Applied and Environmental Microbiology **70**(2): 804-813.

Winter, C., T. Bouvier, M. G. Weinbauer and T. F. Thingstad (2010). "Trade-Offs between Competition and Defense Specialists among Unicellular Planktonic Organisms: the "Killing the Winner" Hypothesis Revisited." Microbiology and Molecular Biology Reviews **74**(1): 42-57.

Withey, S., E. Cartmell, L. M. Avery and T. Stephenson (2005). "Bacteriophages - potential for application in wastewater treatment processes." Science of the Total Environment **339**(1-3): 1-18.

Wittebolle, L., M. Marzorati, L. Clement, A. Balloi, D. Daffonchio, K. Heylen, P. De Vos, W. Verstraete and N. Boon (2009a). "Initial community evenness favours functionality under selective stress." Nature **458**(7238): 623-626.

Wittebolle, L., N. Van Vooren, W. Verstraete and N. Boon (2009b). "High reproducibility of ammonia-oxidizing bacterial communities in parallel sequential batch reactors." Journal of Applied Microbiology **107**(2): 385-394.

Wommack, K. E. and R. R. Colwell (2000). "Virioplankton: viruses in aquatic ecosystems." Microbiol Mol Biol Rev **64**(1): 69-114.

Wommack, K. E., J. Ravel, R. T. Hill, J. S. Chun and R. R. Colwell (1999). "Population dynamics of Chesapeake bay virioplankton: Total-community analysis by pulsed-field gel electrophoresis." Applied and Environmental Microbiology **65**(1): 231-240.

Wood, D. E. and S. L. Salzberg (2014). "Kraken: ultrafast metagenomic sequence classification using exact alignments." Genome Biology **15**(3): R46.

Woodcock, S., C. J. van der Gast, T. Bell, M. Lunn, T. P. Curtis, I. M. Head and W. T. Sloan (2007). "Neutral assembly of bacterial communities." Fems Microbiology Ecology **62**(2): 171-180.

Wu, Q. L. and W. T. Liu (2009). "Determination of Virus abundance, diversity and distribution in a municipal wastewater treatment plant." Water Research **43**(4): 1101-1109.

Xia, L. C., D. Ai, J. Cram, J. A. Fuhrman and F. Sun (2013). "Efficient statistical significance approximation for local similarity analysis of high-throughput time series data." Bioinformatics **29**(2): 230-237.

Xia, L. C., J. A. Steele, J. A. Cram, Z. G. Cardon, S. L. Simmons, J. J. Vallino, J. A. Fuhrman and F. Sun (2011). "Extended local similarity analysis (eLSA) of microbial community and other time series data with replicates." BMC Systems Biology **5**(2): S15.

You, S.-J., Y.-P. Tsai and R.-Y. Huang (2009). "Effect of heavy metals on nitrification performance in different activated sludge processes." Journal of Hazardous Materials **165**(1–3): 987-994.

Yu, Y., C. Lee, J. Kim and S. Hwang (2005). "Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction." Biotechnology and Bioengineering **89**(6): 670-679.

Yves Rosseel (2012). lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48(2), 1-36. URL http://www.jstatsoft.org/v48/i02/.

Zeileis, A. and T. Hothorn (2002). Diagnostic Checking in Regression Relationships. R News 2(3), 7-10. URL http://CRAN.R-project.org/doc/Rnews/

Zhang, J., Q. Gao, Q. Zhang, T. Wang, H. Yue, L. Wu, J. Shi, Z. Qin, J. Zhou, J. Zuo and Y. Yang (2017). "Bacteriophage–prokaryote dynamics and interaction within anaerobic digestion processes across time and space." Microbiome **5**(1): 57.

Zhao, Y. L., K. Wang, H. W. Ackermann, R. U. Halden, N. Z. Jiao and F. Chen (2010). "Searching for a "Hidden" Prophage in a Marine Bacterium." Applied and Environmental Microbiology **76**(2): 589-595.

Zhou, J., W. Liu, Y. Deng, Y.-H. Jiang, K. Xue, Z. He, J. D. Van Nostrand, L. Wu, Y. Yang and A. Wang (2013). "Stochastic Assembly Leads to Alternative Communities with Distinct Functions in a Bioreactor Microbial Community." mBio **4**(2).

Zhou, J., Y. Deng, F. Luo, Z. He and Y. Yang (2011). "Phylogenetic Molecular Ecological Network of Soil Microbial Communities in Response to Elevated CO2." mBio **2**(4).

Zhou, Y., A. Oehmen, M. Lim, V. Vadivelu and W. J. Ng (2011). "The role of nitrite and free nitrous acid (FNA) in wastewater treatment plants." Water Research **45**(15): 4672-4682.

Zumstein, E., R. Moletta and J. J. Godon (2000). "Examination of two years of community dynamics in an anaerobic bioreactor using fluorescence polymerase chain reaction (PCR) single-strand conformation polymorphism analysis." Environmental Microbiology **2**(1): 69-78.

Zuur, A. F., E. N. Ieno and C. S. Elphick (2010). "A protocol for data exploration to avoid common statistical problems." Methods in Ecology and Evolution **1**(1): 3-14.

Wang, Q., G. M. Garrity, J. M. Tiedje and J. R. Cole (2007). "Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy." Applied and Environmental Microbiology **73**(16): 5261-5267.

Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies and F. O. Glöckner (2013). "The SILVA ribosomal RNA gene database project: improved data processing and web-based tools." Nucleic Acids Research **41**(D1): D590-D596.

Callahan, B. J., P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson and S. P. Holmes (2016). "DADA2: High-resolution sample inference from Illumina amplicon data." Nature Methods **13**: 581.

APPENDIX I

AN INTRODUCTION TO WASTEWATER TREATMENT

I.1. Introduction

Municipal and industrial wastewaters are characterised by varying amounts of oxygen-depleting organics, inorganic particles, toxic substances, heavy metals, nutrients, pharmaceuticals and pathogens, thus prior to its reuse or release into receiving waters treatment is required. This is typically achieved through combining physical, chemical and biological processes, termed preliminary, primary, secondary and tertiary treatment respectively in conventional systems. In the UK the level of treatment, thus the number and type of processes needed, is defined by legislation described in the Urban Wastewater Directive (91/271/EEC), which ensures that public health and the environment is protected.

Preliminary treatment involves removing gross solids, such as rags, floatables, grit and grease, that could damage or interfere with subsequent process equipment, reduce treatment reliability and effectiveness and/or cause effluent (treated wastewater) contamination. This is typically achieved using coarse and/or fine screens and grit chambers (Fig. I.1). Subsequently primary sedimentation, using large, mechanically cleaned circular or rectangular tanks, is implemented to remove a substantial portion, 50 – 70%, of readily settleable organic solids, as well as remaining floatables. Chemical addition prior to primary sedimentation, termed pre-precipitation, facilitates the concurrent removal of phosphorous and some heavy metals with primary solids. Chemical processes, such as oxidation and precipitation, are similarly used within secondary treatment to remove the majority of remaining organics and nutrients, although biological systems, discussed in greater detail in section I.2, are more widespread. Finally tertiary processes, including filtration, oxidation, precipitation and disinfection, are implemented when residual contamination exists.

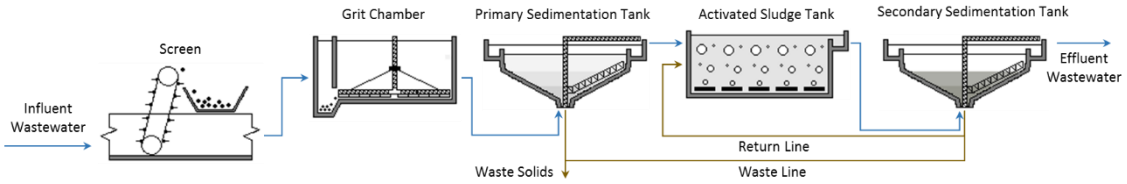

Figure I. 1. Process flow diagram of a conventional, biological wastewater treatment system.

I.2. Biological Secondary Treatment

Secondary treatment is dominated by biological systems which utilise microorganisms to biodegrade organics, capture and incorporate colloids, transform and remove nutrients and, in some instances, degrade trace constituents and compounds. These systems can be divided into two main categories: attached growth processes, such as trickling filters, biological towers and rotating biological contactors, and suspended growth processes, such as activated sludge,

anaerobic digestion and aerated lagoons. The most common of all these systems, regardless of category, is activated sludge, which, as a consequence, is the system under study here.

I.3. The Activated Sludge Process

The activated sludge process, discovered by Ardern and Lockett (1914), typically consists of an aeration tank, a secondary sedimentation tank and a solids recycling and waste line respectively (Fig. I.1). The aeration tank is a suspended-growth reactor in which microorganisms, harnessed to consume and oxidise input organics and nutrients, are grown in suspended aggregates, or flocs (Fig. I.2), due to continuous aeration and/or mechanical mixing. The slurry of treated wastewater and flocs, termed mixed liquor or activated sludge, then passes to the sedimentation tank where flocs settle, producing a clear effluent at the surface, which can be discharged or tertiary treated, and a "thickened" return activated sludge (RAS) at the base, which can be returned to the aeration tank or wasted (Fig. I.1). Capturing and recycling flocs in this way augments microbial biomass within the aeration tank and reduces the hydraulic retention time (HRT) required to treat influent wastewaters, whilst wasting RAS enables the solids retention time (SRT) and thus, due to varying growth rates, the presence and absence of certain bacteria to be controlled.

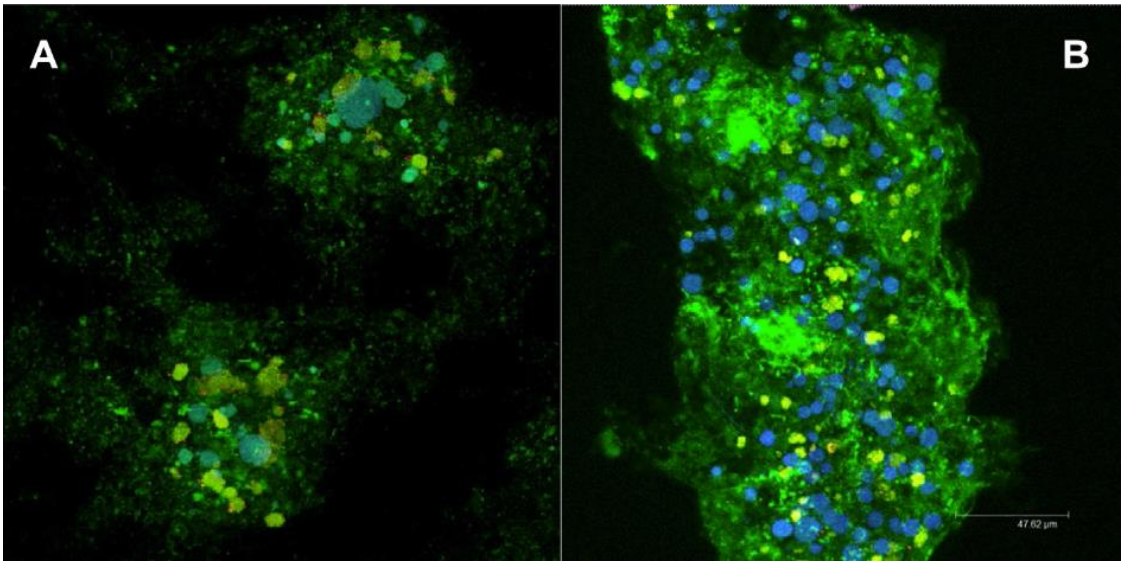

Figure I. 2. Full (**A**) and lab (**B**) scale activated sludge flocs observed by confocal laser scanning microscopy after fluorescence in situ hybridization (taken from Ofiteru *et al.*, 2015). Green - heterotrophic bacteria; blue - ammonia oxidizing bacteria; yellow - nitrite oxidizing bacteria.

Mixed liquor itself consists of a complex and varied community of microorganisms, with bacteria, protozoa, rotifers, fungi and bacteriophage all thought to be present. The most abundant of these are heterotrophic bacteria, which, alongside extracellular polymers, inert particulates, non-biodegradable organics and water, are the main constituent of activated sludge flocs (Fig. I.2). As primary consumers of organic waste heterotrophs are also heavily important to the active sludge

process, converting it to carbon dioxide, water and new bacterial cells. Autotrophs, particularly ammonia and nitrite oxidisers, are also important to floc structure (Fig. I.2) and the activated sludge process, since they oxidise ammonia in a two-step process to nitrite and then nitrate. It is these two groups of organism, and their growth kinetics, that form the basis of activated sludge process design.

APPENDIX II

AN INTRODUCTION TO BACTERIOPHAGES

I.1. What is a phage?!

Viruses are a group of small biological entities, typically 30 – 60nm consisting of a nucleic acid genome (single or double stranded RNA or DNA) encapsulated in a protein or lipoprotein coat, or capsid (Kutter and Sulakvelidze, 2004; Weinbauer, 2004; Withey *et al*., 2005). Bacterial viruses, coined "bacteriophages" (phages) by D'Herelle in 1917, are those viruses that infect prokaryotes and are split into 13 families (Weinbauer, 2004). In general phages, or a specific group of phages making up 95% of all those discussed in the literature and almost all those thought to be significant in the environments discussed in this paper, namely tailed phages, are often characterised by a head, or capsid, and a tail, held together by a protein connector (Sabour and Griffiths, 2010). Phages, like all viruses, are obligate intracellular parasites with no intrinsic metabolism, thus they require the metabolic machinery of a host cell to survive and reproduce (Withey *et al*., 2005).

I.2. The Phage Infection Cycle

Initial contact with a host cell typically occurs through a diffusion mediated extracellular search and subsequent chance collision between a phage and a bacterium, thus contact is often dependent on host concentrations (Kutter and Sulakvelidze, 2004). Phage adsorption follows and occurs in two steps (Weinbauer, 2004). Firstly reversible binding is undertaken, where a phage recognizes a susceptible host through the interaction of attachment sites on specialised adsorption structures, such as tail fibres or spikes, and host surface molecules (Kutter and Sulakvelidze, 2004; Weinbauer, 2004). This positions the phage correctly on the host cell surface and triggers a structural rearrangement in the tail, allowing irreversible attachment, the second adsorption step (Sabour and Griffiths, 2010). Here a different adsorption structure protein binds to the host bacterium; this is again mediated by specific receptors on the surface of the host, including carbohydrate, protein and lipopolysaccharide molecules (Marks and Sharp, 2000; Weinbauer, 2004). Many phages require a cluster of one specific type of molecule present in high concentrations to properly position adsorption structures for attachment (Kutter and Sulakvelidze, 2004).

Succeeding irreversible adhesion the phage genome must cross two or three major bacterial barriers, the outer membrane, the peptidoglycan layer and the inner membrane, to initiate infection (Sabour and Griffiths, 2010). To do this phages employ strategies that vary with morphology, but in general an enzymatic mechanism exists within the tail tip for penetrating the aforementioned membranes (Kutter and Sulakvelidze, 2004). Once penetrated the protein connecting the head and tail of the phage changes shape, disabling the blocking mechanism stopping the premature extrusion of genomic material from the capsid (Sabour and Griffiths,

2010). Consequently the DNA is drawn rapidly into the hosts cytoplasm by cellular energetics, often involving available ATP, a membrane potential or enzyme action, although the mechanisms are highly varied (Kutter and Sulakvelidze, 2004; Sabour and Griffiths, 2010).

After internalization of the bacteriophage genome the phase of infection begins, this can be divided into an eclipse period and a period of phage-progeny maturation (Calendar and Abedon, 2005). The length and timing of each is dependent on a phages life cycle (discussed below), thus the eclipse period is either prevegetative in the sense of immediately preceding phage-progeny maturation, or is temporarily or greatly extended (Calendar and Abedon, 2005). The eclipse period begins with recognition of strong phage promoters by host RNA polymerase, leading to the transcription of immediate early genes, the products of which protect the phage genome and restructure the host cell appropriately for the needs of the phage (Kutter and Sulakvelidze, 2004). This may involve the redirection or inhibition of various macromolecular processes, such as transcription, translation or replication, as well as single enzymatic functions (Kutter and Sulakvelidze, 2004; Sabour and Griffiths, 2010). Once optimal metabolic conditions are established middle genes are then transcribed, producing products that synthesise new phage DNA, followed by a set of late genes, that encode components of the actual phage particle (Kutter and Sulakvelidze, 2004).

The period of phage-progeny maturation, or morphogenesis, follows, where the replicated, concatemerized phage genomes are packed into new phage heads or capsids, a highly regulated process that can be briefly split into three major stages (Sabour and Griffiths, 2010). Firstly the phage head, an icosahedral protein shell known as the prohead or procapsid at this stage, is assembled around scaffolding proteins (Kutter and Sulakvelidze, 2004; Sabour and Griffiths, 2010). Located at one vertex of the prohead is the portal complex, the starting point for head assembly (Kutter and Sulakvelidze, 2004). In the next phase terminase, a DNA packing enzyme which docks to the portal complex, translocates the phage genome into the procapsid through ATP hydrolysis, transforming it into a mature capsid (Sabour and Griffiths, 2010). Following DNA packing the tail or adsorption structure, formed separately, is attached to the portal complex, completing the replication of new phage particles (Kutter and Sulakvelidze, 2004). The number of phages produced during a single cycle of infection, the burst size, varies between 50 and 200 new phage particles (Withey *et al*., 2005).

At the end of the replication cycle and completing the phase of infection, newly formed phage particles, progeny, exit the host cell in search of new prey. To do this bacteriophages have developed two basic strategies. Filamentous phages, as a result of their unique morphogenesis,

continuously extrude their progeny across a host's cell wall without causing death (Sabour and Griffiths, 2010). However most tailed phages exit host cells via fatal lysis, using specific lysis proteins. During the phase of infection a small hydrophobic membrane spanning protein, named Holin, accumulates and oligomerizes in the cytoplasmic membrane of the host (Sabour and Griffiths, 2010). Simultaneously a soluble and active endolysin, a phage encoded muralytic enzyme, also accrues. At a genetically predetermined time the holins forms a membrane lesion, permeabilizing the inner membrane and allowing the cognate endolysin contact with the peptidoglycan layer (Sabour and Griffiths, 2010). The endolysin degrades the peptidoglycan layer until the cell can no longer withstand the internal osmotic pressure, causing it to burst and thus release phage progeny into extracellular space (Sabour and Griffiths, 2010). The extracellular stage ends with the death of a phage or a new infection (Weinbauer, 2004).

I.3. Phage Life Cycles

The life cycle of most bacteriophages can be divided into two major groups: the lytic cycle and the lysogenic cycle, although such classification is probably a simplification of the diversity of phage life cycles that exist (Weinbauer, 2004). During the lytic cycle, the lytic or virulent phage immediately redirects the host metabolism towards the production of new phage virions, which are released upon cell lysis within minutes to hours of initial infection (Weinbauer, 2004). Thus during the infection process the eclipse period immediately proceeds the period of phage-progeny maturation. In contrast during the lysogenic cycle temperate phages, although they are able to propagate lytically like virulent phages, act more subtly and establish a stable relationship with their host cell, thus greatly extending the eclipse period (Casjens, 2003; Sabour and Griffiths, 2010). In this state the phage genome is replicated in concert with the host's chromosome during host doubling, and virus genes that are detrimental to the host are not expressed (Casjens, 2003; Jacquet *et al*., 2010). During this association phage DNA, now called a prophage, is usually physically integrated into one of the native replicons of the host, or in some cases exists as a circular or linear plasmid in the bacterial cytoplasm (Casjens, 2003). The integrated prophage remains dormant inside the cell until induction (the triggering of the lytic cycle), at which point prophage genes required for lytic growth are expressed and progeny are produced and released through cell lysis (Jacquet *et al*., 2010). Induction can happen spontaneously and randomly in a small fraction of bacteria that harbour a given prophage, or specific environmental signals can cause simultaneous induction of a particular prophage in many cells (Casjens, 2003).

APPENDIX III

SUPPLEMENTARY INFORMATION - COUPLED VIRUS-BACTERIA INTERACTIONS AND ECOSYSTEM FUNCTION IN AN ENGINEERED MICROBIAL SYSTEM

III.1. Supplementary Methods

While it is common to simply assume constant per-taxa rRNA copy numbers (usually 2), it is recognized (Angly *et al*., 2014) that the value may vary quite considerably between taxa of different lineages, and that simply assuming a constant copy number can lead to significant misinterpretation of results (Angly *et al*., 2014; Props *et al*., 2017). For the current study, it is important that the qPCR measurements of total bacteria accurately reflect the true individual proportions, despite shifts in community composition. Thus, we corrected the qPCR counts using inferred copy numbers based on the taxomonic composition of each sample. The method closely follows that employed by Angly *et al*. (2014) and their CopyRighter software. Unfortunately, due to the relative age of CopyRighter and its choice of taxonomic database, it was necessary to re-implement the code, utilizing up-to-date SILVA (v. 128, Quast *et al*., 2013), IMG (Markowitz *et al*., 2012) and NCBI taxonomy tree (Benson *et al*., 2009; Sayers *et al*., 2009) databases (2017/08/18).

Briefly, the complete Bacterial and Archael IMG database was retrieved from the JGI Genome Portal (Grigoriev *et al*., 2012; Nordberg *et al*., 2014) and filtered to include only those entries classified as "Finished" genomes. These finished genomes were then used to provide a "first pass" 16s copy number annotation for a copy of the NCBI taxonomy tree, for those leaf nodes with a 1 to 1 exact match to the corresponding IMG taxon id.

Thus the starting point is a full taxonomic tree, for which some of the leaf nodes (species) have a "true" 16s copy number assignment. The copy number of each parent node (initially at the Genus rank) is then determined from the mean of the assigned child taxa, and any unassigned children are also given this mean value. The process is then repeated, but at successively higher taxonomic ranks (Family, Order, Class and finally Phyla).

At the end of the process, all nodes in the taxonomy tree have been assigned an inferred 16s copy number, which is subsequently used to correct per-sample 16s taxonomic abundances from the 16s amplicon data. The weighted mean of these corrected CN values was then used to correct the qPCR results.

III.2. Supplementary Results

III.2.1. Bioreactor performance and abiotic conditions

Table III. 1. Summary of bioreactor performance and operational conditions over the 2 year study period.

Characteristic	Average	Range
Influent Flow Rate (m^3 day^{-1})	8860.0 ± 2854.5	5365.4 - 17392.3
Hydraulic Retention Time (hours)	10.6 ± 3.1	4.97 - 29.43
Solids Retention Time (days)	11.13 ± 3.09	5.84 - 28.32
ML Temperature (°C)	12.6 ± 3	6.3 - 17.9
ML DO (mg L^{-1})	2.5 ± 0.8	1.09 - 8.34
ML pH	6.56 ± 0.22	5.97 - 7.04
Influent SS (g L^{-1})	0.13 ± 0.11	0.02 - 1.02
ML SS (g L^{-1})	2.73 ± 0.47	0.97 - 3.65
Effluent SS (g L^{-1})	0.03 ± 0.03	0 - 0.11
Influent VSS (g L^{-1})	0.11 ± 0.06	0 - 0.51
ML VSS (g L^{-1})	2.01 ± 0.4	0.69 - 3.4
Effluent VSS (g L^{-1})	0.02 ± 0.02	0 - 0.11
Influent COD (mg L^{-1})	223.2 ± 156.3	29 - 1172
ML COD (mg L^{-1})	41.28 ± 35.35	13 - 254
Effluent COD (mg L^{-1})	31.97 ± 28.51	2 - 154
Influent NH4 - N (mg L^{-1})	32.1 ± 16.2	5.8 - 99
ML NH4 - N (mg L^{-1})	14.06 ± 15,21	1.4 - 66.2
Effluent NH4 - N (mg L^{-1})	8.4 ± 14	0 - 76.4
Influent Chloride (mg L^{-1})	81.1 ± 37.7	0.3 - 320.3
ML Chloride (mg L^{-1})	88.02 ± 47.77	32.39 - 355.31
Effluent Chloride (mg L^{-1})	89.8 ± 64.2	26 - 461.4
Influent Fluoride (mg L^{-1})	2.99 ± 2.8	0.15 - 18.6
ML Fluoride (mg L^{-1})	0.62 ± 0.83	0 - 5.16
Effluent Fluoride (mg L^{-1})	0.61 ± 1.1	0 - 7.5
Influent Nitrate (mg L^{-1})	3.8 ± 5.6	0 - 25.3
ML Nitrate (mg L^{-1})	17.93 ± 17.66	0 - 63.39
Effluent Nitrate (mg L^{-1})	44.5 ± 21.2	0.7 - 96.7
Influent Nitrite (mg L^{-1})	0.34 ± 0.74	0 - 4.9
ML Nitrite (mg L^{-1})	0.7 ± 1.05	0 - 5.64
Effluent Nitrite (mg L^{-1})	0.47 ± 0.97	0 - 5.3
Influent Phosphate (mg L^{-1})	12.8 ± 5.5	0.81 - 28.2
ML Phosphate (mg L^{-1})	8.52 ± 9.22	0 - 68.98
Effluent Phosphate (mg L^{-1})	4.7 ± 3.9	0 - 24.5
Influent Sulphate (mg L^{-1})	68.5 ± 20.1	0 - 117.9
ML Sulphate (mg L^{-1})	91.52 ± 9.22	0 - 147.64
Effluent Sulphate (mg L^{-1})	96.6 ± 29.5	0 - 160.6
Influent Aluminium (µg L^{-1})	352.4 ± 142.5	0 - 756.2
Influent Arsenic (µg L^{-1})	9.3 ± 7.9	0 - 28.3
Influent Cadmium (µg L^{-1})	3.1 ± 11.3	0 - 91.9
Influent Calcium (mg L^{-1})	47.4 ± 8.7	31.3 - 70.2
Influent Chromium (µg L^{-1})	15.8 ± 8.5	5.8 - 57.8
Influent Copper (µg L^{-1})	63.3 ± 47.4	9.8 - 303.8
Influent Iron (mg L^{-1})	3.6 ± 2.2	0 - 11.7
Influent Lead (µg L^{-1})	11.8 ± 9.4	0 - 48.2
Influent Magnesium (mg L^{-1})	11 ± 2.8	6.4 - 17.9
Influent Manganese (µg L^{-1})	175.3 ± 66.9	59.9 - 537.4
Influent Nickel (µg L^{-1})	18.7 ± 16.7	6.4 - 135.8
Influent Potassium (mg L^{-1})	13.1 ± 3.1	4.6 - 20.9
Influent Silicone (mg L^{-1})	3.65 ± 1.21	0.01 - 6.8
Influent Sodium (mg L^{-1})	51.3 ± 14.5	27.5 - 97.8
Influent Sulphur (mg L^{-1})	22.6 ± 4.3	10.6 - 36.6
Influent Zinc µg L^{-1})	97.6 ± 43.8	38.5 - 428.1

±denotes standard deviation across the 2 year sampling period.

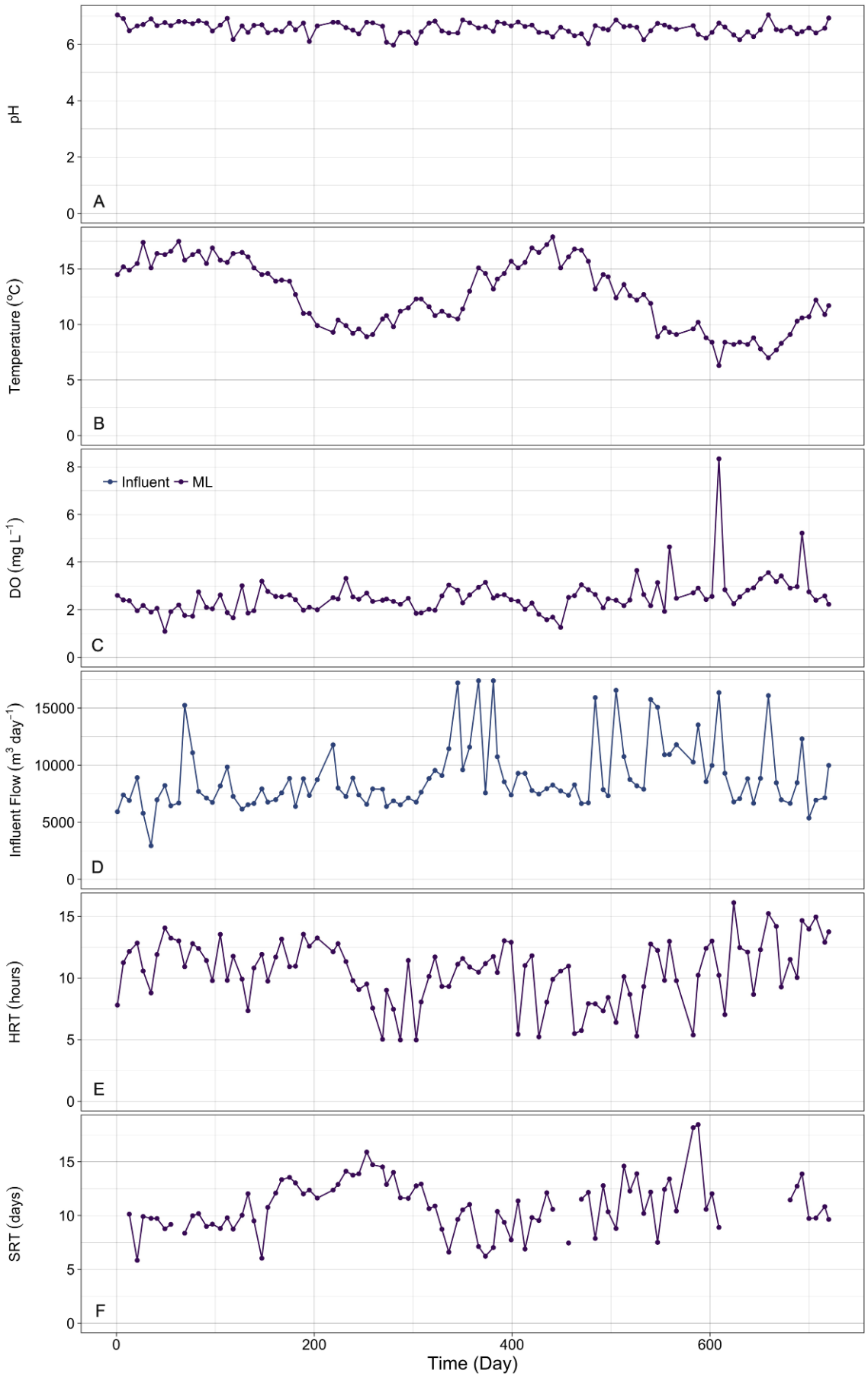

Figure III. 1. Variations in plant operational and environmental parameters during the 2 year study period.

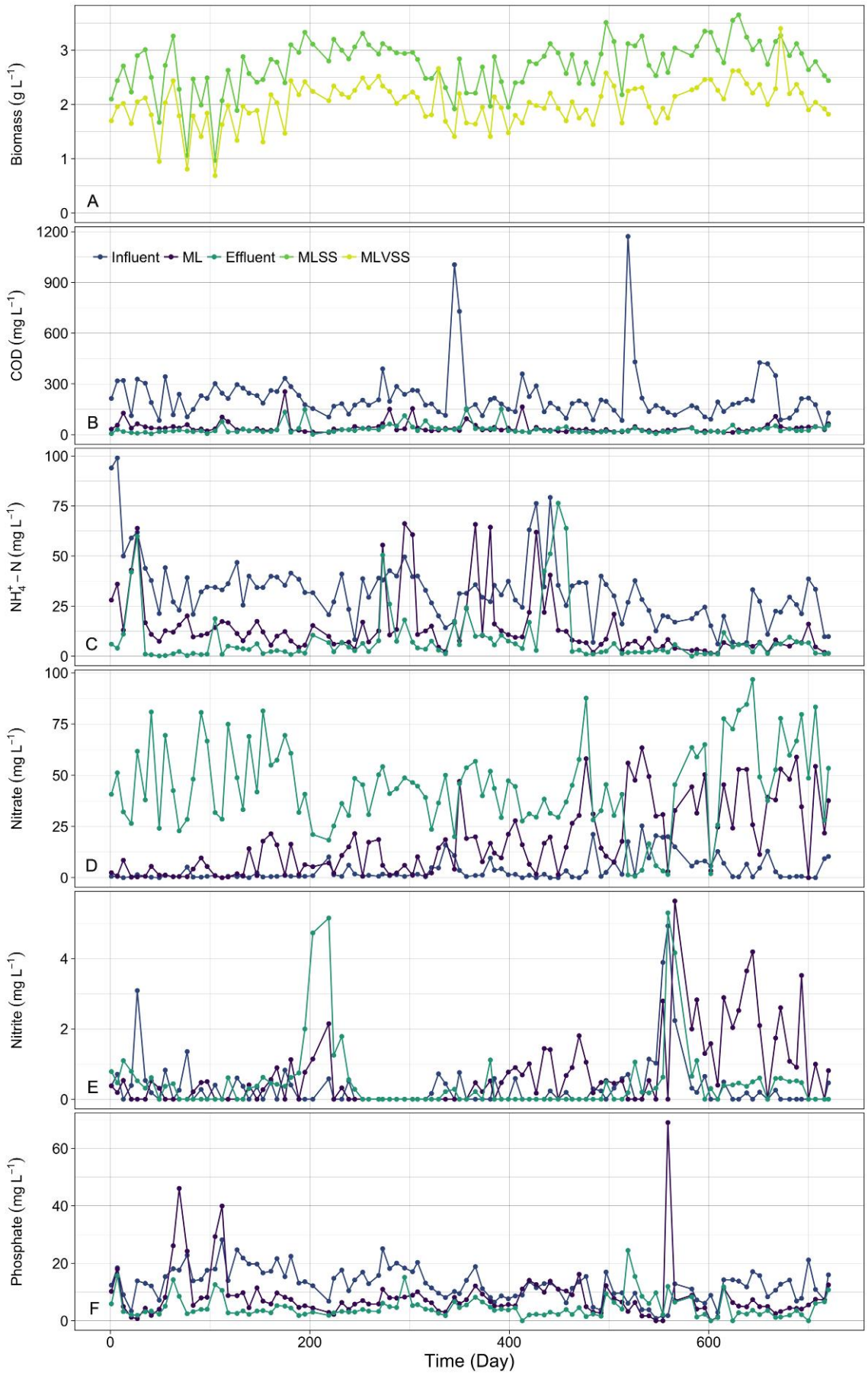

Figure III. 2. Variations in measured operational and environmental parameters during the 2 year study period.

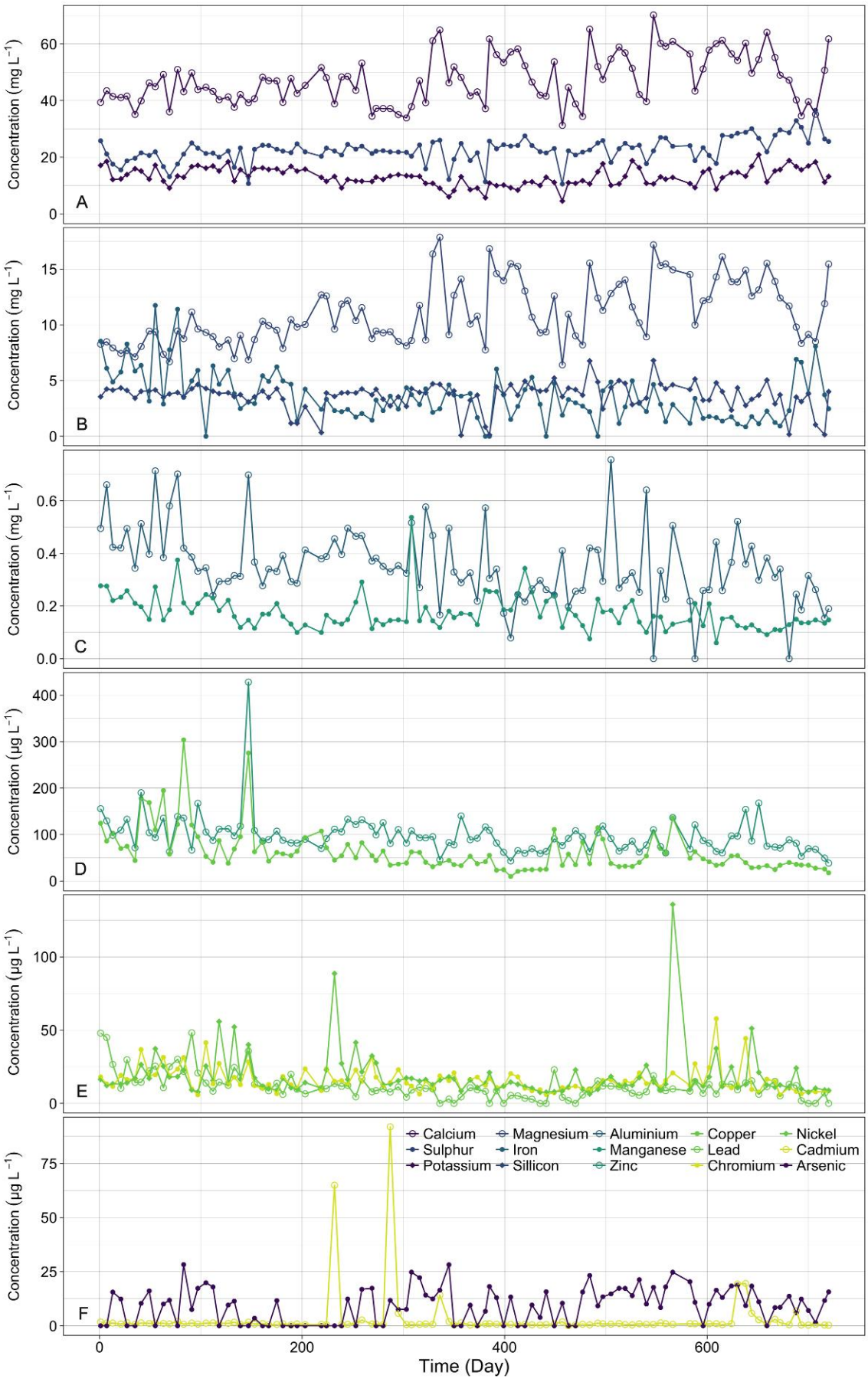

Figure III. 3. Variations in influent trace metal concentrations during the 2 year study period.

III.2.2. Virus interactions with biotic and abiotic conditions

Table III. 2. ML virus GLS regression model.

Correlation Structure	Estimate (Phi)	Phi 95% CI's (min, max)	*P*-Value$^{\Psi}$	
Continuous AR(1) ~Week	0.2701805	0.01, 0.56	0.0242	*

Coefficient	Estimate	SE	t-Value	VIF	95% CI's (min, max)	*P*-Value	
Intercept	7.658220	0.3269756	23.421379	NA	7.01, 8.31	< 0 x 10^{-4}	***
Influent Magnesium	-0.111927	0.0157274	-7.116712	2.66	-0.14, -0.08	< 0 x 10^{-4}	***
Influent NH_4^+-N	0.082932	0.0245058	3.384159	1.94	0.03, 0.13	0.0011	**
Influent Phosphate	0.051780	0.0192322	2.692371	1.34	0.01, 0.09	0.0085	**
Influent Sulphur	0.084989	0.0139979	6.071562	2.39	0.06, 0.11	< 0 x 10^{-4}	***
Influent Virus	0.070807	0.0301316	2.349925	1.15	0.01, 0.13	0.0210	*
ML AOB	0.068876	0.0324239	2.124228	1.30	0.01, 0.13	0.0364	*
ML Fluoride	0.024139	0.0098516	2.450294	1.15	0.01, 0.04	0.0162	*
ML Nitrate	-0.031683	0.0086946	-3.643961	2.05	-0.05, -0.01	0.0004	***
ML Nitrite	-0.028170	0.0092757	-3.037017	1.67	-0.05, -0.01	0.0031	**
ML pH	0.024975	0.0104911	2.380580	1.42	0.00, 0.05	0.0194	*
ML Sulphate	-0.054635	0.0113081	-4.831522	1,51	-0.08, -0.03	< 0 x 10^{-4}	***

All measured biological, operational and environmental variables included as covariates except Effluent Virus (prior ML community), K_i = 57, n = 102, R^2 = 0.83. Anderson-Darling Test P = 0.6435. All VIF scores < 3. SE = standard error. CI = confidence interval. $^{\Psi}$ ANOVA *P*-Value comparing models with identical fixed effects but with and without correlation structure. ° P < 0.1, * P < 0.05, ** P < 0.01, *** P < 0.001.

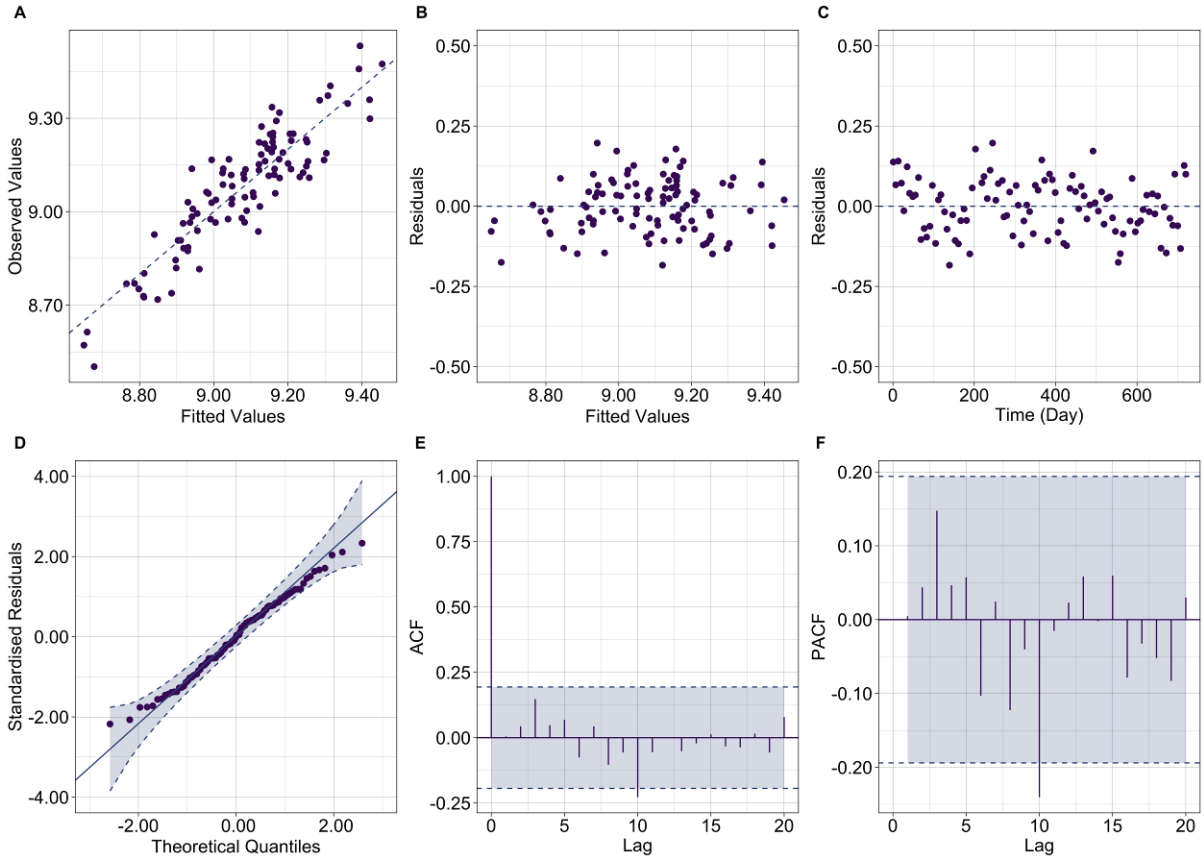

Figure III. 4. Diagnostic plots assessing linearity (**A**, **B**), homoscedasticity (**B**, **C**), residual normality (**D**) and residual independence (**E**, **F**) of the ML virus GLS regression model.

III.2.2.1. ML total bacteria GLS model

Table III. 3. ML total bacteria GLS regression model.

Correlation Structure	Estimate (Phi)	Phi 95% CI's (min, max)	*P*-Value$^{\Psi}$	
Continuous AR(1) ~Week	0.6923042	0.511, 0.692	$< 1 \times 10^{-4}$	***

Coefficient	Estimate	SE	t-Value	VIF	95% CI's (min, max)	*P*-Value	
Intercept	2.9063594	1.9755254	1.471	NA	-1.02, 6.83	0.1447	
Influent Calcium	0.0583154	0.0324226	1.799	1.37	-0.01, 0.12	0.0754	°
Influent Fluoride	-0.0945430	0.0304022	-3.110	1.17	-0.16, -0.03	0.0025	**
Influent Iron	-0.0559659	0.0266322	-2.210	1.14	-0.11, -0.00	0.0383	*
Influent Sulphur	-0.0770958	0.0285081	-2.704	1.30	-0.14, -0.02	0.0082	**
ML Chloride	0.0661114	0.0240350	2.751	1.04	0.02, 0.11	0.0072	**
ML Phosphate	-0.1117296	0.0447715	-2.496	1.75	-0.20, -0.02	0.0144	*
ML Virus	0.6147413	0.2188601	2.809	1.43	0.18, 1.05	0.0061	**
Effluent Phosphate	0.1489278	0.0466262	3.194	1.67	0.06, 0.24	0.0019	**
Effluent Sulphate	0.1228601	0.0285461	4.304	1.22	0.07, 0.18	$< 0 \times 10^{-4}$	***

All measured biological, operational and environmental variables included as covariates except ML AOB (subset of ML Bacteria), K_i = 57, n = 102, R^2 = 0.53. Anderson-Darling Test P = 0.3025. All VIF scores < 3. SE = standard error. CI = confidence interval. $^{\Psi}$ ANOVA *P*-Value comparing models with identical fixed effects but with and without correlation structure. ° $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

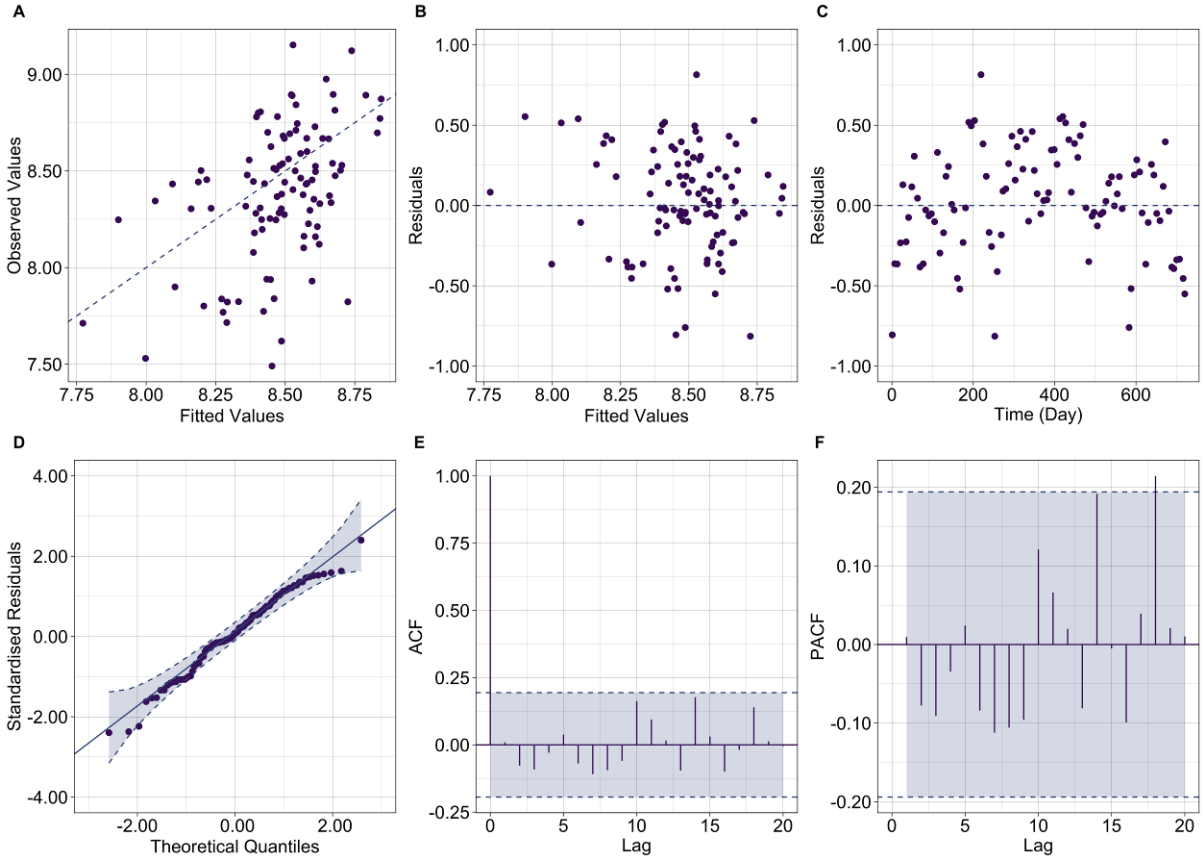

Figure III. 5. Diagnostic plots assessing linearity (**A**, **B**), homoscedasticity (**B**, **C**), residual normality (**D**) and residual independence (**E**, **F**) of the ML bacteria GLS regression model.

The ML total bacteria GLS model also identified strong positive associations with influent calcium ($P < 0.1$), ML chloride ($P < 0.01$) and effluent sulphate ($P < 0.001$) and phosphate ($P < 0.01$) concentrations, whilst influent Fluoride ($P < 0.01$), iron ($P < 0.05$) and sulphur ($P < 0.01$) and ML phosphate ($P < 0.05$) concentrations were negatively associated (Table III.3).

III.2.2.2. ML AOB GLS model

Table III. 4. ML AOB GLS regression model.

Correlation Structure	Estimate (Phi)	Phi 95% CI's (min, max)	*P*-Value[ψ]		
Continuous AR(1) ~Week	0.220039	0.20, 0.24	0.019	*	
Coefficient	Estimate	SE	t-Value	VIF	95% CI's (min, max)	*P*-Value	
Intercept	2.7946612	1.5482351	1.805063	NA	-0.28, 5.87	0.0746	°
Influent AOB	0.1302831	0.0487218	2.674021	1.34	0.03, 0.23	0.0090	**
Influent Sulphate	0.0958521	0.0300576	3.188945	1.43	0.04, 0.16	0.0020	***
Influent Sulphur	-0.0829007	0.0326164	-2.541688	1.56	-0.15, -0.02	0.0128	*
ML Virus	0.9233023	0.2199362	4.198046	1.87	0.49, 1.36	0.0001	***
Effluent NH_4^+-N	0.0577049	0.0249816	2.309898	1.16	0.01, 0.11	0.0233	*
Effluent Chloride	0.0622500	0.0283608	2.194933	1.15	0.01, 0.12	0.0309	*
Effluent Nitrite	0.0699123	0.0256705	2.723451	1.06	0.02, 0.12	0.0078	**
Effluent Virus	-0.5384314	0.1920804	-2.803156	1.79	-0.92, -0.16	0.0063	**

All measured biological, operational and environmental variables included as covariates except ML Bacteria (encompasses AOB), Influent AOB included as statistically significant, K_l = 59, n = 95, R^2 = 0.47. Anderson-Darling Test P = 0.7136. All VIF scores < 3. SE = standard error. CI = confidence interval. [ψ] ANOVA *P*-Value comparing models with identical fixed effects but with and without correlation structure. ° $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

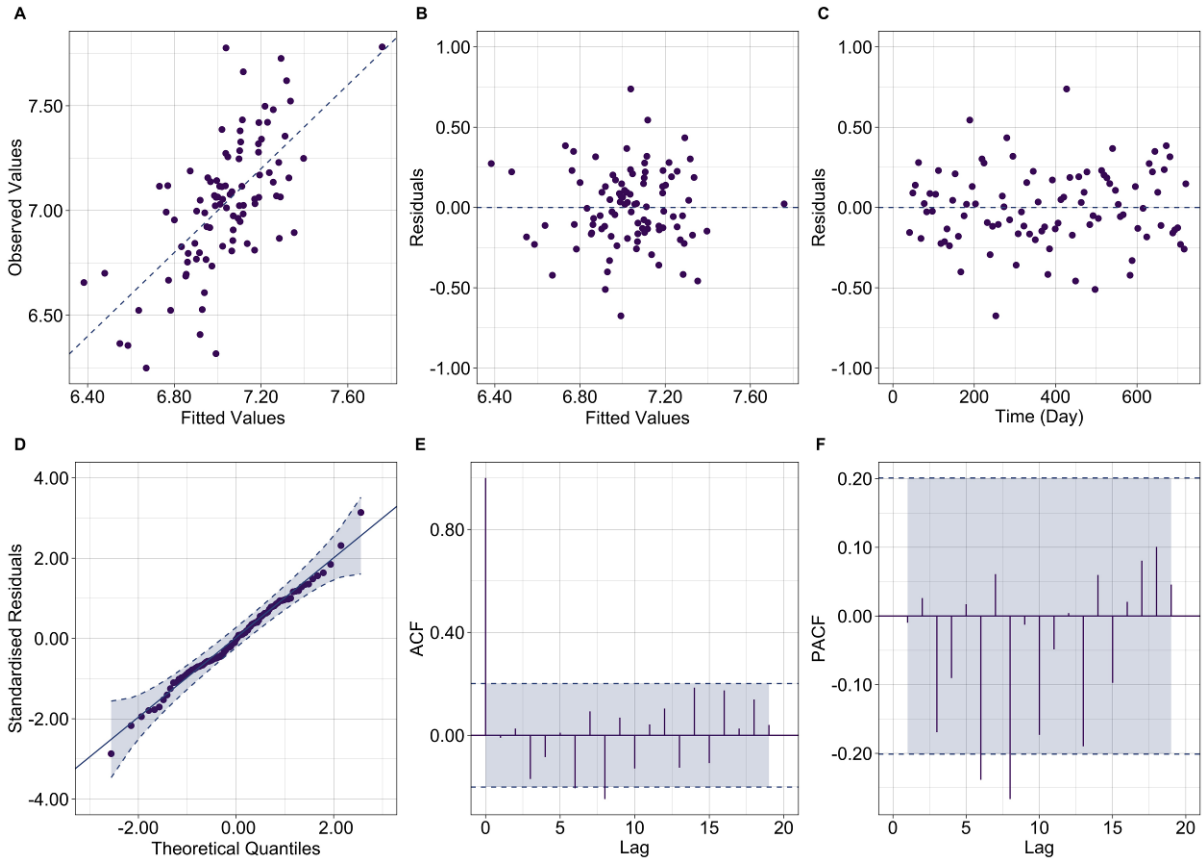

Figure III. 6. Diagnostic plots assessing linearity (**A**, **B**), homoscedasticity (**B**, **C**), residual normality (**D**) and residual independence (**E**, **F**) of the ML AOB GLS regression model.

The ML AOB GLS model also identified strong positive associations with influent AOB abundance ($P < 0.01$) and concentrations of influent sulphate ($P < 0.001$) and effluent NH_4^+-N ($P < 0.05$), chloride ($P < 0.05$) and nitrite ($P < 0.01$), whilst influent sulphur ($P < 0.05$) concentrations were negatively associated (Table III.4).

III.2.3. Virus interactions with bacteria community structure

Table III. 5. Summary of CCA analysis.

	CCA Axis			
	1	2	3	4
Eigenvalue	0.30019	0.16720	0.14967	0.08197
Species-Environment Correlation	0.9187647	0.8177232	0.7518984	0.7376913
Cumulative Percentage Variance				
Of Species Data	9.041	14.076	18.584	21.053
Of Species-Environment Relationship	33.310	51.860	68.460	77.557
Permutation Test (Axis)[ψ]				
F ratio	11.4163	6.3588	5.6922	3.1174
P-Value	< 0.001***	< 0.001***	< 0.001***	0.004 **

	Explanatory Variable								
	ML Temperature	Influent Potassium	ML Virus	ML Nitrite	Influent Phosphate	Influent Sulphur	HRT	ML NH_4^+-N	Effluent Nitrite
Permutation Test (Marginal Effect)									
F ratio	7.8360	3.9854	2.8313	3.0122	3.0684	2.2224	2.1521	2.0749	1.8994
P-Value	< 0.001***	< 0.001***	< 0.001***	< 0.001***	< 0.001***	< 0.002**	0.002**	0.005**	0.009**
Intraset Correlation Coefficient									
Axis 1	-0.91649	0.01446	-0.22353	0.18997	-0.10557	0.46368	0.15925	-0.31352	0.16563
Axis 2	-0.19340	0.68680	-0.2376	0.50620	-0.14210	0.34480	0.34630	-0.53070	0.43280
Axis 3	0.08593	-0.42061	-0.36894	0.58938	-0.59410	0.23713	-0.07004	-0.16102	-0.13638
Axis 4	0.17000	0.16787	0.70477	-0.05776	-0.01740	-0.05791	0.33206	0.44728	0.42471
VIF	1.75	2.02	2.61	1.53	1.70	2.04	1.11	1.98	1.58

All measured biological, operational and environmental variables included as explanatory variables except ML Bacteria and AOB, *n* = 102. Global permutation test F ratio = 3.801, *P*-Value <0.001***. All VIF scores < 3. [ψ] CCA axis 5 and 6 also significant, *P* < 0.01 and *P* < 0.05 respectively. * *P* < 0.05, ** *P* < 0.01, *** *P* < 0.001.

APPENDIX IV

SUPPLEMENTARY INFORMATION - VIRUS - BACTERIA INTERACTIONS, SYNCHRONICITY AND ECOSYSTEM FUNCTION IN REPLICATE ENGINEERED MICROBIAL SYSTEMS

IV.1. Supplementary Results

Table IV. 1. Summary of influent characteristics.

Characteristic	Average	Range
Influent pH	6.87 ± 0.20	6.48 - 7.48
Influent SS (g L^{-1})	0.13 ± 0.07	0.00 - 0.67
Influent VSS (g L^{-1})	0.12 ± 0.06	0.00 - 0.51
Influent COD (mg L^{-1})	218.45 ± 78.38	59.00 - 441.00
Influent NH4 - N (mg L^{-1})	38.46 ± 10.23	10.10 - 69.00
Influent Chloride (mg L^{-1})	69.12 ± 16.44	0.51 - 125.85
Influent Fluoride (mg L^{-1})	3.80 ± 2.25	0.26 - 14.41
Influent Nitrate (mg L^{-1})	1.29 ± 6.75	0.00 - 68.86
Influent Nitrite (mg L^{-1})	0.21 ± 0.53	0.00 - 3.89
Influent Phosphate (mg L^{-1})	18.96 ± 8.54	3.35 - 72.80
Influent Sulphate (mg L^{-1})	67.08 ± 18.02	29.83 - 154.58
Influent Aluminium (µg L^{-1})	415.1 ± 132.4	239.9 - 714
Influent Arsenic (µg L^{-1})	6.8 ± 8.0	0 - 28.3
Influent Cadmium (µg L^{-1})	1.02 ± 0.38	0.37 – 2.05
Influent Calcium (mg L^{-1})	42.8 ± 3.8	35.1 – 50.9
Influent Chromium (µg L^{-1})	18.3 ± 8.7	5.8 – 41.5
Influent Copper (µg L^{-1})	101.4 ± 64.6	38.3 - 303.8
Influent Iron (mg L^{-1})	5.30 ± 2.54	0 - 11.7
Influent Lead (µg L^{-1})	20.2 ± 11.2	6.2 - 48.2
Influent Magnesium (mg L^{-1})	8.66 ± 1.12	6.7 – 11.2
Influent Manganese (µg L^{-1})	195.9 ± 57.6	99.5 – 375
Influent Nickel (µg L^{-1})	20.1 ± 162.5	7.8 – 55.9
Influent Potassium (mg L^{-1})	14.9 ± 2.21	9.14 – 18.5
Influent Silicone (mg L^{-1})	3.68 ± 0.78	1.16 – 4.65
Influent Sulphur (mg L^{-1})	20.7 ± 3.43	10.7 – 25.8
Influent Zinc µg L^{-1})	119.2 ± 61.5	63.2 - 428.1

±denotes standard deviation, *n* = 102.

Table IV. 2. Summary of Control CSTR's performance and operational parameters during acclimatisation (day 0 - 72).

	Control 1					Control 2					Control 3					Control 4					Control 5					Control 6				
Characteristic	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig
HRT/SRT (Days)*	4.48	0.62	3.92	7.88	T7	4.55	0.60	4.03	6.69	T7	5.24	3.18	4.03	22.37	T1 T7	5.70	4.14	3.20	27.89	C6 T1 - T3 T6	4.48	0.47	3.41	6.16	T7	4.43	0.53	3.42	6.13	C4
Influent Flow (mL Hour)*	9.01	0.80	5.06	10.17	T2	8.80	0.92	5.90	9.81	T2	8.42	1.63	1.77	9.82	T2	8.30	2.00	1.46	12.71	T2	9.14	0.86	6.58	11.90		9.09	0.91	6.50	11.63	
ML Temp (°C)*	14.42	0.41	14.17	16.74	C4 - C6 T1 - T4, T6	14.59	0.43	14.15	16.93	C5, C6 T1 - T3	14.59	0.36	14.35	16.59	C4 - C6 T1, T4, T6	14.74	0.33	14.36	16.54	C1, C3 T1, T5	14.75	0.35	14.50	16.66	C1 - C3 T1, T5	14.84	0.36	14.60	16.81	C1 - C3 T3 - T6
ML pH*	6.20	0.74	4.97	7.26		6.14	0.67	5.01	7.12		5.86	0.74	4.31	7.10		6.20	0.81	4.91	7.37		5.94	0.77	4.77	7.09		5.89	0.73	4.61	6.90	
ML DO (mg L^{-1})*	9.17	0.56	6.53	9.65	C3, C5, C6 T2. T3	8.87	0.56	7.12	9.47	C3 T2. T4	8.19	0.71	5.23	8.97	C3, C5, C6 T2. T3	8.93	0.60	7.74	9.63	C3 T2. T4	8.40	0.93	6.73	9.42	C1 T4. T6	8.21	1.15	6.21	9.42	C1 T4. T6
Effluent COD (mg L^{-1})	25.54	9.73	4.00	49.00		31.06	14.53	12.00	82.00		27.46	10.89	8.00	56.00		28.43	13.27	2.00	60.00		28.49	16.44	2.00	106.00		29.94	16.12	10.00	102.00	
Effluent NH_4^+-N (mg L^{-1})	4.82	6.33	0.10	25.00		5.96	7.68	0.30	39.50		6.48	9.52	0.30	51.00		6.30	9.85	0.10	46.00		9.12	14.98	0.30	78.00		7.32	12.24	0.20	64.60	
ML SS (g L^{-1})	0.21	0.40	0.00	2.01		0.23	0.41	0.01	1.90		0.31	0.45	0.06	2.03		0.20	0.36	0.00	1.79		0.21	0.36	0.00	1.85		0.21	0.38	0.00	1.86	
ML VSS (g L^{-1})	0.17	0.29	0.00	1.52		0.19	0.32	0.00	1.54		0.25	0.35	0.00	1.63		0.17	0.26	0.00	1.32		0.17	0.27	0.00	1.41		0.17	0.27	0.00	1.37	
Effluent Nitrate (mg L^{-1})	170.16	36.74	90.96	235.40		162.28	39.93	68.67	243.94		164.93	34.00	111.50	230.10		174.41	39.06	110.00	312.90		174.04	36.78	61.48	235.24		181.20	44.34	115.60	368.50	
Effluent Nitrite (mg L^{-1})*	2.58	6.53	0.00	26.35		4.84	11.51	0.00	54.10		2.45	5.89	0.00	26.24		0.49	0.81	0.00	2.44		0.49	0.84	0.00	3.34		0.50	1.31	0.00	5.84	
Effluent Sulphate (mg L^{-1})	93.00	49.07	58.14	328.30		93.06	39.40	55.15	224.31		96.56	45.76	57.02	280.63		95.89	45.31	55.22	303.21		91.82	43.61	55.37	310.63		96.21	55.14	19.51	321.21	
Effluent Phosphate (mg L^{-1})	16.87	3.95	9.16	26.62		17.19	3.83	8.70	23.46		16.93	3.84	8.36	23.37		17.42	4.02	8.68	29.10		16.91	3.50	8.86	22.56		17.56	3.33	9.28	25.44	
Effluent Fluoride (mg L^{-1})	0.47	0.29	0.30	1.53		0.48	0.29	0.29	1.46		0.52	0.31	0.28	1.46		0.59	0.43	0.30	2.34		0.56	0.47	0.29	2.76		0.56	0.41	0.29	2.03	
Effluent Chloride (mg L^{-1})	83.56	10.36	65.70	129.62		82.13	7.90	63.09	95.56		80.68	7.96	50.54	90.88		83.94	10.98	54.35	128.87		81.39	9.62	41.87	96.61		84.98	13.60	66.42	151.20	

n = 36, SD = standard deviation, min and max = minimum and maximum value. * Significantly different between CSTR's and from corresponding CSTR (Sig) at the 0.05 level. Anderson-Darling and Bartlett Test $P < 0.05$ for all comparisons.

Table IV. 3. Summary of Control CSTR's performance and operational parameters following acclimatisation (day 72 - 204).

	Control 1					Control 2					Control 3					Control 4					Control 5					Control 6				
Characteristic	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig
HRT/SRT (Days)*	4.92	1.18	3.70	10.31	T7 T4	4.77	0.87	3.96	8.59	T7 T4	4.73	0.84	3.80	8.99	T7 T4	5.16	2.32	3.99	19.98	T7 T4	5.45	4.63	3.95	39.19	C3 T7 T4	4.72	0.98	3.76	9.25	T7 T4
Influent Flow (mL Hour)*	8.44	1.36	3.84	10.71	T2. T4	8.42	1.22	4.25	10.16	T7 - T4	8.35	1.07	4.40	10.10	C5 T2. T3. T5. T6	8.48	1.58	1.98	10.12	T2. T4	8.51	1.62	1.03	9.98	T2. T4	8.61	0.93	4.82	10.18	T7 - T4
ML Temp (°C)*	14.42	0.22	13.99	14.92	C4	14.36	0.13	14.00	14.62	C4, C6	14.36	0.11	13.95	14.54	C4, C6	14.70	0.06	14.47	14.84	C1 - C3 T4	14.55	0.11	14.15	14.87		14.69	0.10	14.24	14.86	C1 - C3 T4
ML pH*	5.97	0.98	4.44	7.51	C3, C6	5.89	1.03	4.42	7.68		5.41	1.04	3.92	7.22	C1, C4 T3. T6	5.94	0.98	4.52	7.64	C3	5.53	1.03	4.23	7.31		5.44	1.13	4.13	7.61	C1 T6
ML DO (mg L^{-1})*	9.22	0.38	8.35	9.80	C2 - C6 T2, T4 - T6	7.15	0.61	6.01	8.30	C1 T1, T3 - T6	6.89	0.82	5.48	8.41	C1, C4 T1. T3 - T6	7.92	1.23	6.44	9.89	C1, C3 T1 - T3	7.27	0.61	5.33	8.28	C1 T1, T3 - T6	7.40	1.02	5.17	9.72	C1 T1, T3, T5
Effluent COD (mg L^{-1})*	19.45	5.32	5.00	34.00	T1	20.63	6.23	6.00	37.00	T1	20.58	5.95	5.00	33.00	T1	19.61	7.37	6.00	51.00	T1	20.82	5.50	8.00	36.00	T1	20.37	6.84	9.00	42.00	T1
Effluent NH_4^+-N (mg L^{-1})*	2.07	1.98	0.10	11.80		2.26	2.42	0.20	16.50		2.32	1.47	0.20	7.20		2.56	2.10	0.00	11.40		2.68	2.14	0.00	14.30		2.77	2.33	0.40	15.90	
ML SS (g L^{-1})	0.08	0.03	0.00	0.17	C3 T3. T4	0.09	0.04	0.01	0.22		0.11	0.06	0.00	0.31		0.10	0.07	0.00	0.28		0.10	0.06	0.02	0.39		0.11	0.10	0.01	0.72	
ML VSS (g L^{-1})	0.07	0.03	0.00	0.17	T3. T4	0.08	0.04	0.01	0.20		0.10	0.05	0.00	0.30		0.09	0.05	0.00	0.20		0.09	0.05	0.02	0.37		0.10	0.09	0.01	0.58	
Effluent Nitrate (mg L^{-1})	153.48	24.50	74.59	200.74		155.31	23.59	81.43	191.20		154.37	27.76	83.28	198.28		159.24	26.76	86.17	234.60		162.30	23.20	102.08	197.27		160.21	27.15	82.70	219.99	
Effluent Nitrite (mg L^{-1})	0.08	0.38	0.00	2.84		0.06	0.24	0.00	1.46		0.08	0.26	0.00	1.45		0.06	0.19	0.00	1.34		0.06	0.23	0.00	1.31		0.10	0.33	0.00	2.31	
Effluent Sulphate (mg L^{-1})	68.31	9.08	46.30	94.59		67.68	8.73	40.26	89.31		68.58	10.64	37.96	108.30		68.70	12.86	40.71	125.24		68.75	10.87	48.23	120.72		70.38	14.30	45.38	133.10	
Effluent Phosphate (mg L^{-1})	23.09	4.48	14.64	32.89		22.99	4.23	13.22	30.39		23.04	4.50	12.86	31.98		23.40	5.14	2.12	33.70		23.75	4.36	16.29	32.86		23.62	4.73	14.53	37.92	
Effluent Fluoride (mg L^{-1})	0.26	0.14	0.00	0.89		0.26	0.42	0.00	3.56		0.22	0.12	0.00	0.50		0.25	0.22	0.00	1.53		0.22	0.14	0.00	0.71		0.26	0.19	0.00	1.20	
Effluent Chloride (mg L^{-1})	68.78	10.98	47.03	99.42		68.55	10.79	46.30	93.77		68.27	13.76	36.33	129.21		68.69	11.40	36.10	94.57		68.53	10.36	46.18	92.70		69.07	11.64	39.73	93.58	

n = 66, SD = standard deviation, min and max = minimum and maximum value. * Significantly different between CSTR's and from corresponding CSTR (Sig) at the 0.05 level. Anderson-Darling and Bartlett Test $P < 0.05$ for all comparisons.

Table IV. 4. Summary of Test CSTR's performance and operational parameters during acclimatisation (day 0 - 72).

CSTR	Test 1					Test 2					Test 3					Test 4					Test 5					Test 6				
Characteristic	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig
HRT/SRT (Days)*	4.41	1.38	0.00	9.96	C3, C4 T4	4.25	0.62	3.67	7.03	C1 - C5 T4. T5	4.31	0.35	3.46	5.18	C4	4.60	0.82	3.38	8.85	T1 T7	4.51	0.63	3.57	7.01	T7	4.53	1.48	0.00	10.27	C4
Influent Flow (mL Hour)*	8.77	1.95	0.00	10.16		9.55	1.01	5.69	10.91	C1 - C4 T4	9.20	0.75	7.60	11.40		8.72	1.04	4.44	11.61	T2	9.15	1.03	5.80	11.40		8.74	1.98	0.00	10.02	
ML Temp (°C)*	14.92	0.33	14.75	16.79	C1 - C5 T3 - T6	14.78	0.35	14.53	16.76	C1 - C3 T3, T5	14.59	0.39	14.11	16.49	C6 T1, T2, T5	14.70	0.30	14.45	16.36	C1, C6 T1. T5	14.29	0.40	13.92	16.44	C2 - C6 T1 - T4, T6	14.69	0.35	14.17	16.53	C1, C6 T1. T5
ML pH*	6.36	0.65	5.04	7.14		6.32	0.70	5.24	7.51		6.06	0.68	4.82	7.18		6.26	0.78	4.91	8.15		6.07	0.72	4.84	7.10		6.40	0.78	4.99	7.49	
ML DO (mg L^{-1})*	9.03	0.45	6.68	9.42	C3 T2. T4	8.13	0.65	6.01	8.93	C1, C2, C4 T1. T4. T6	8.54	0.54	6.71	9.23	C1 T4. T6	9.52	0.44	8.10	10.05	C2 - C6 T1 - T3. T6	8.64	0.80	6.99	9.70	T4. T6	9.28	0.52	8.05	9.96	C3, C5, C6 T2. T3. T5
Effluent COD (mg L^{-1})	31.23	8.10	14.00	48.00		28.66	10.60	2.00	62.00		32.60	13.69	16.00	76.00		33.00	15.57	8.00	80.00		30.29	13.87	4.00	74.00		29.91	13.13	10.00	84.00	
Effluent NH_4^+-N (mg L^{-1})	7.75	10.53	0.10	52.30		6.44	10.21	0.00	56.70		10.17	18.49	0.10	99.00		10.34	22.38	0.00	117.00		9.22	19.84	0.20	108.00		11.04	19.70	0.00	109.00	
Effluent SS (g L^{-1})	0.23	0.38	0.00	1.99		0.24	0.36	0.01	1.84		0.31	0.48	0.01	2.38		0.21	0.35	0.00	1.74		0.23	0.37	0.00	1.71		0.23	0.44	0.00	2.29	
ML SS (g L^{-1})	0.20	0.30	0.00	1.56		0.20	0.27	0.01	1.39		0.25	0.37	0.01	1.82		0.16	0.26	0.00	1.32		0.20	0.28	0.00	1.31		0.20	0.34	0.00	1.77	
ML VSS (g L^{-1})	166.36	34.31	103.65	232.97		177.17	34.56	112.30	246.29		170.02	40.75	75.22	233.20		149.95	50.47	37.04	231.92		177.17	35.69	107.10	280.46		169.18	33.62	97.60	231.19	
Effluent Nitrite (mg L^{-1})*	0.34	0.87	0.00	4.96		0.85	2.57	0.00	11.66		7.71	16.40	0.00	63.59		2.16	4.38	0.00	16.06		1.06	1.70	0.00	7.57		1.01	1.69	0.00	7.57	
Effluent Sulphate (mg L^{-1})	90.66	36.06	56.51	253.07		94.97	48.26	54.14	287.87		92.14	53.47	54.83	366.03		97.59	58.99	54.72	317.26		97.49	41.94	56.91	233.93		83.73	25.34	54.55	197.16	
Effluent Phosphate (mg L^{-1})	17.31	4.68	8.85	35.91		17.94	6.76	8.79	51.71		17.82	5.93	9.47	44.02		16.76	4.16	7.75	32.85		17.52	3.60	8.58	26.20		16.71	3.78	8.53	23.58	
Effluent Fluoride (mg L^{-1})	0.46	0.19	0.28	1.10		0.50	0.33	0.28	1.74		0.46	0.30	0.27	1.80		0.54	0.42	0.29	2.22		0.52	0.28	0.31	1.56		0.41	0.18	0.28	1.37	
Effluent Chloride (mg L^{-1})	83.23	7.71	70.28	99.17		80.26	13.81	15.11	99.62		79.86	11.81	46.06	110.14		81.74	7.71	69.37	101.32		81.77	15.52	14.16	117.14		83.13	8.46	68.63	106.46	

n = 36, SD = standard deviation, min and max = minimum and maximum value. * Significantly different between CSTR's and from corresponding CSTR (Sig) at the 0.05 level. Anderson-Darling and Bartlett Test $P < 0.05$ for all comparisons.

Table IV. 5. Summary of Test CSTR's performance and operational parameters following acclimatisation (day 72 - 204).

CSTR	Test 1					Test 2					Test 3					Test 4					Test 5					Test 6				
Characteristic	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig
HRT/SRT (Days)*	4.88	1.21	4.01	10.97	T7 T4	4.60	1.22	3.90	9.52	C1 - C5 T1. T4. T6	4.80	1.42	3.97	11.78	T4	5.09	1.25	3.72	11.63	C1 - C5 T1. T2. T3. T5. T	4.68	0.93	4.06	9.26	T4	5.16	3.18	3.92	26.15	T7 T4
Influent Flow (mL Hour)*	8.36	1.22	3.61	9.88	T2. T3. T5. T6	9.07	1.19	4.20	9.95	C1 - C6 T1, T4	8.81	1.34	3.35	10.41	C2, C3, C6 T1, T4	7.98	1.22	3.38	10.55	C1, C2, C4-C6 T2, T3, T5, T6	8.81	1.17	4.27	10.02	C2, C3 T1, T4	8.81	1.17	4.27	10.02	C2, C3 T1, T4
ML Temp (°C)*	14.82	4.40	8.62	21.11		14.74	4.32	8.65	21.01		14.48	4.32	8.42	20.80		14.60	4.40	8.56	20.99		13.98	4.38	7.85	20.27	C4, C6	14.41	4.28	8.29	20.57	
ML pH*	5.66	1.05	4.26	7.51		5.64	0.87	4.37	7.21		5.90	0.85	4.49	7.49	C3	5.76	0.77	4.58	7.23		5.74	0.87	4.43	7.37		5.94	0.97	4.34	7.50	C3, C6
ML DO (mg L^{-1})*	8.72	0.78	7.50	10.08	C2 - C6 T7	6.96	0.72	5.65	8.54	C1, C4 T1, T3 - T6	8.59	0.81	6.50	10.21	C2 - C6 T7	8.14	1.10	6.31	10.57	C1, C2, C3, C5 T7	8.54	1.80	4.98	10.83	C1-C3, C5, C6 T7	8.49	2.25	4.91	11.62	C1, C2, C3, C5 T7
Effluent COD (mg L^{-1})*	27.18	13.11	10.00	83.00	C1 - C6 T4, T6	21.81	7.32	1.00	49.00		22.64	6.63	6.00	39.00		20.13	6.50	10.00	47.00	T1	21.97	6.48	7.00	38.00		20.15	5.41	10.00	33.00	
Effluent NH_4^+-N (mg L^{-1})*	2.73	2.09	0.40	9.50		2.57	2.15	0.30	9.90		3.02	2.55	0.20	15.30		2.65	2.08	0.20	9.50		3.16	2.59	0.20	14.10		3.46	4.26	0.00	30.30	
ML SS (g L^{-1})	0.09	0.06	0.01	0.29		0.11	0.06	0.00	0.29		0.12	0.07	0.01	0.44	C1	0.12	0.12	0.01	0.91	C1	0.08	0.05	0.01	0.24		0.11	0.07	0.01	0.48	
ML VSS (g L^{-1})	0.08	0.05	0.01	0.26		0.10	0.05	0.00	0.28		0.10	0.06	0.01	0.37	C1	0.11	0.10	0.01	0.77	C1	0.09	0.07	0.01	0.41		0.09	0.07	0.01	0.41	
Effluent Nitrate (mg L^{-1})	162.48	28.02	94.90	223.62		165.97	24.42	100.10	210.44		162.69	28.69	80.49	254.60		161.86	24.89	85.26	197.89		159.55	28.67	81.42	201.72		158.32	25.80	91.97	200.65	
Effluent Nitrite (mg L^{-1})	0.08	0.34	0.00	2.24		0.12	0.32	0.00	1.57		0.06	0.22	0.00	1.30		0.18	0.54	0.00	3.74		0.05	0.24	0.00	1.65		0.03	0.14	0.00	1.02	
Effluent Sulphate (mg L^{-1})	69.41	14.80	41.00	134.15		68.93	11.42	38.79	98.86		67.90	12.96	40.67	141.19		69.79	19.57	40.92	210.13		67.79	7.75	45.50	85.66		67.37	9.66	42.09	91.57	
Effluent Phosphate (mg L^{-1})	24.16	4.77	16.50	40.88		24.09	4.08	17.07	35.41		24.05	3.90	17.28	31.75		24.39	4.25	16.36	33.06		23.62	4.37	15.76	32.35		23.50	4.57	14.26	32.62	
Effluent Fluoride (mg L^{-1})	0.35	0.47	0.00	3.72		0.31	0.50	0.00	3.97		0.26	0.22	0.00	1.59		0.25	0.20	0.00	1.32		0.23	0.13	0.00	0.61		0.29	0.50	0.00	4.15	
Effluent Chloride (mg L^{-1})	70.61	14.56	45.15	146.11		68.59	11.17	41.97	91.48		68.22	10.90	47.13	92.94		68.02	9.69	45.65	91.82		67.55	10.71	39.35	93.18		68.31	11.56	42.85	90.47	

n = 66, SD = standard deviation, min and max = minimum and maximum value. * Significantly different between CSTR's and from corresponding CSTR (Sig) at the 0.05 level. Anderson-Darling and Bartlett Test $P < 0.05$ for all comparisons.

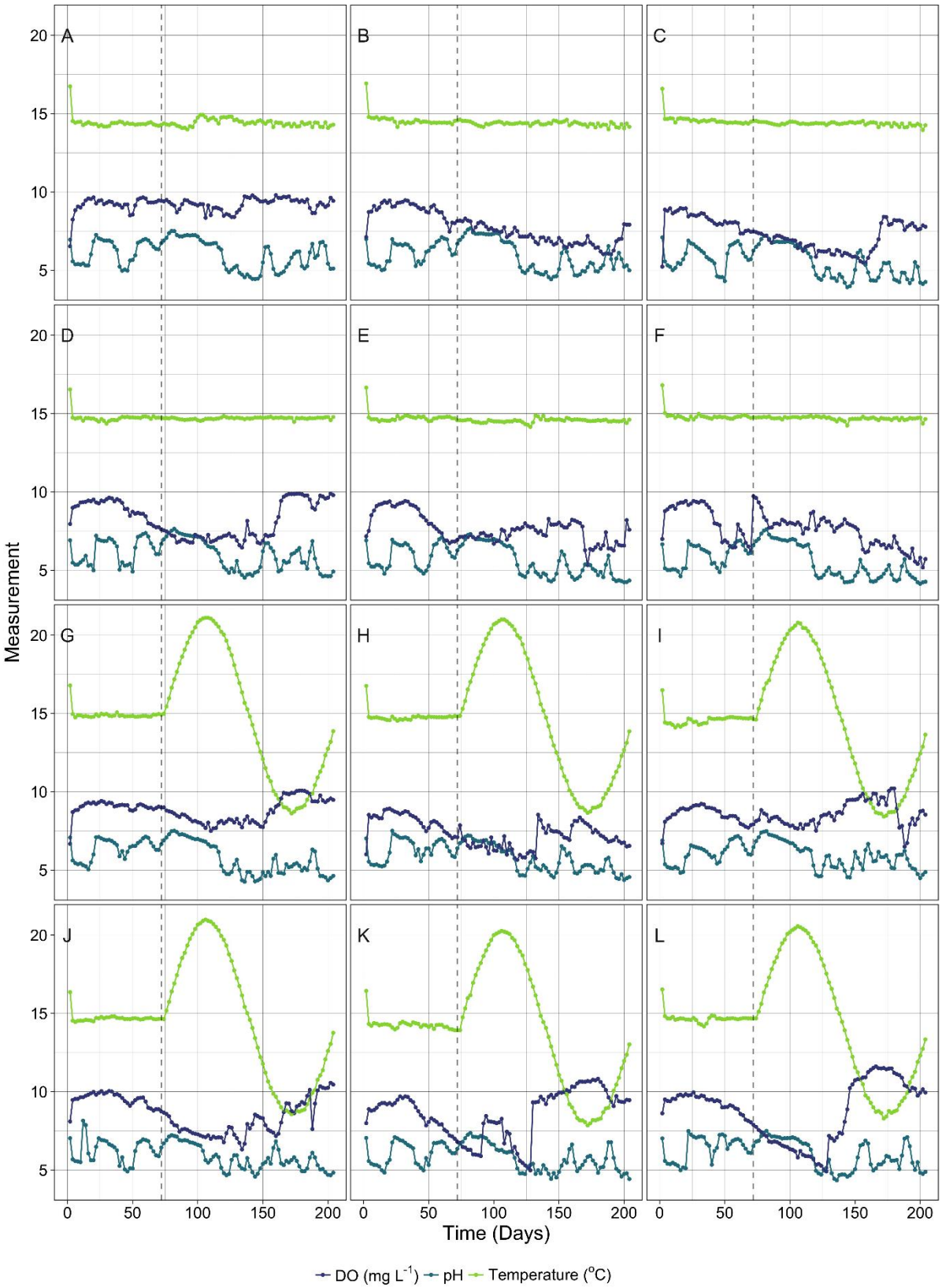

Figure IV. 1. Variation in operational and environmental parameters in all CSTR's over the 204 day study. Dashed line represents the end of acclimatisation. (**A** – **F**) Control 1 – 6, (**G** – **L**) Test 1 – 6.

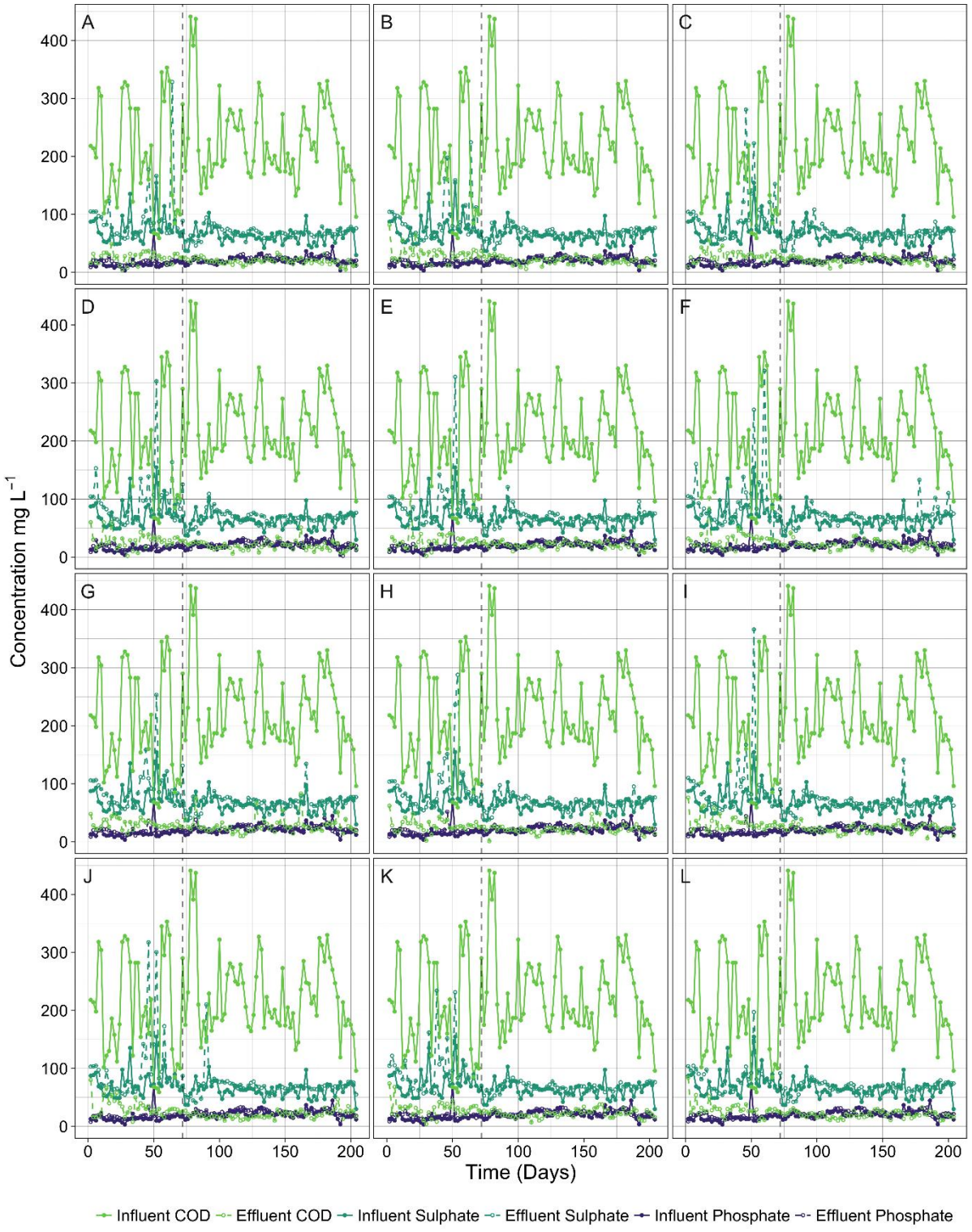

Figure IV. 2. Variation in operational and environmental parameters in all CSTR's over the 204 day study. Dashed line represents the end of acclimatisation. (**A** – **F**) Control 1 – 6. (**G** – **L**) Test 1 – 6.

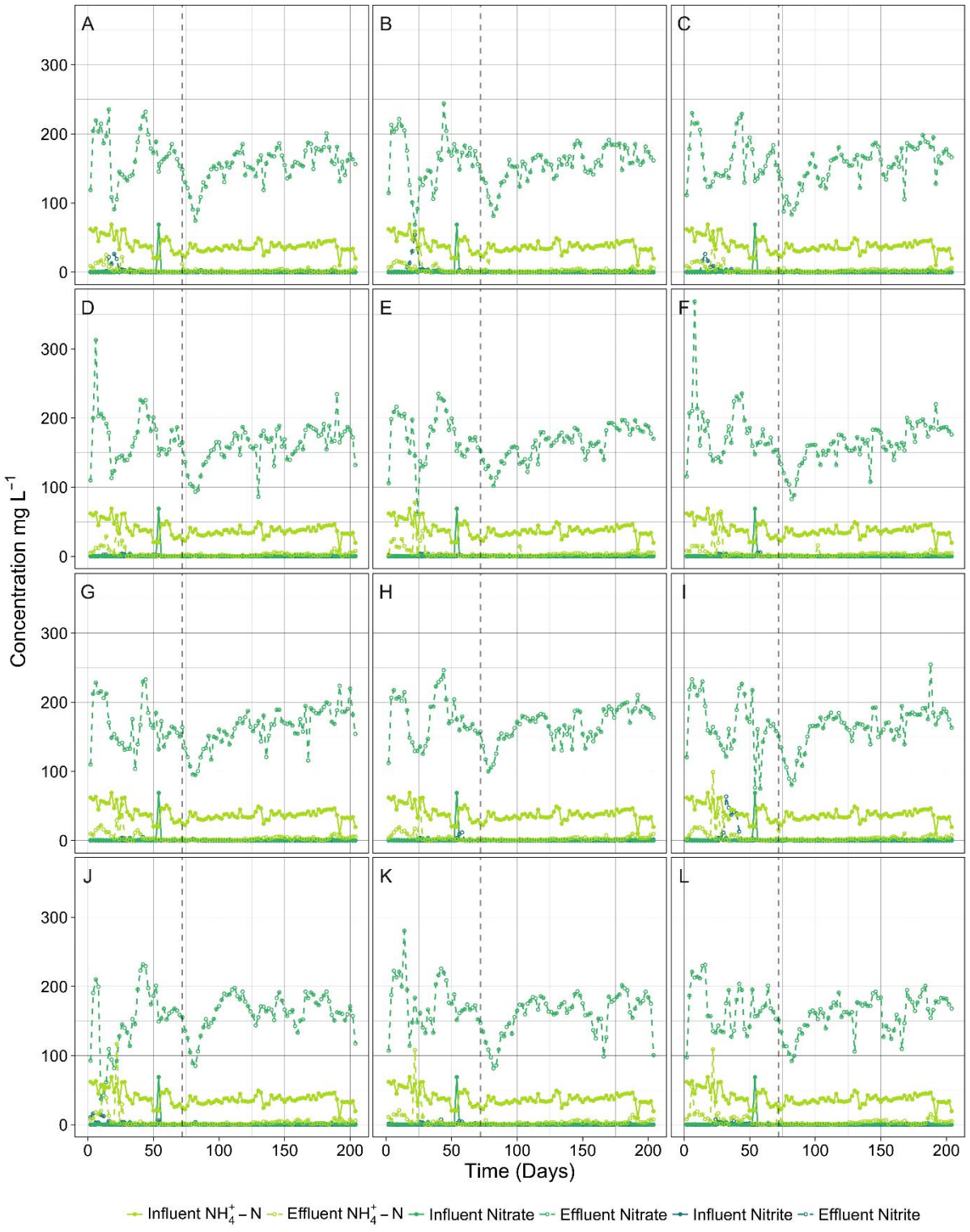

Figure IV. 3. Variation in operational and environmental parameters in all CSTR's over the 204 day study. Dashed line represents the end of acclimatisation. (**A** – **F**) Control 1 – 6. (**G** – **L**) Test 1 – 6.

Table IV. 6. Summary of influent biotic conditions.

Biotic Parameter	Mean	SD	Min	Max
Influent Viruses × 10^8 mL^{-1}	0.98	0.65	0.29	2.72
Influent Bacteria × 10^7 mL^{-1}	2.27	1.09	0.96	6.49
Influent AOB × 10^5 mL^{-1}	0.81	0.57	0.53	5.43

±denotes standard deviation across 204 days, 102 samples

Table IV. 7. Summary of biotic conditions within Control CSTR's (day 62 - 204).

CSTR	Control 1					Control 2					Control 3					Control 4					Control 5					Control 6				
Day 62 - 70																														
Abundance	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig
ML Viruses × 10^8 mL^{-1}*	2.72	0.62	1.90	3.64	C5 T7	2.31	0.91	1.09	3.17	T7	2.33	1.07	1.24	3.64	T7	1.39	0.35	0.96	1.93		0.65	0.36	0.36	1.22	C5	0.84	0.26	0.57	1.18	
ML Bacteria × 10^7 mL^{-1}	2.91	2.17	0.93	6.31		2.11	1.50	0.24	3.73		6.79	4.58	1.67	13.47		2.49	2.55	0.21	6.50		2.15	1.78	0.51	4.90		2.28	1.61	0.65	4.48	
ML AOB × 10^5 mL^{-1}	1.29	1.21	0.37	3.36		0.90	0.35	0.40	1.22		2.33	2.56	0.50	6.74		0.80	0.79	0.19	2.15		0.98	0.83	0.32	2.42		0.67	0.10	0.56	0.81	
Day 72 - 204																														
ML Viruses × 10^8 mL^{-1}*	1.47	0.59	0.50	3.57	C3 - C6 T7 - T4	1.17	0.54	0.24	2.77	C4 - C6 T2, T6	0.91	0.41	0.34	2.07	C1 T2, T6	0.70	0.40	0.23	2.04	C1, C2 T1, T2 - T5	0.71	0.47	0.23	2.61	C1, C2 T1, T2 - T5	0.82	0.52	0.20	2.04	C1, C2 T1
ML Bacteria × 10^7 mL^{-1}Ψ	2.16	4.32	0.03	24.61		1.62	3.39	0.08	23.70		2.55	4.38	0.00	24.42		1.26	1.40	0.01	6.87		1.67	2.18	0.02	10.37		1.82	3.24	0.01	19.69	
ML AOB × 10^5 mL^{-1}*Ψ	0.96	1.25	0.02	6.59		1.17	1.61	0.03	8.27		2.31	4.07	0.01	21.16	C5 T2, T5, T6	1.41	3.21	0.01	19.02	T4	1.33	2.85	0.04	17.07	C5	1.11	1.48	0.03	8.79	
Day 62 - 204																														
D_1	191.9	58.6	95.9	318.1	C5 T2, T3, T5, T6	171.5	38.9	59.2	252.0	T5, T6	168.7	54.4	56.9	282.7	T5, T6	179.7	78.2	67.1	418.8	T5, T6	152.2	62.0	38.6	299.9	C1 T5	177.7	56.5	70.5	317.9	T5, T6
D_2	89.9	32.7	24.4	157.1	C5 T2, T3, T5, T6	77.3	26.9	12.5	131.4	T5, T6	79.4	29.8	15.6	149.2	T5, T6	79.1	38.8	23.0	227.6	T5, T6	67.7	31.1	11.9	142.1	C1 T5	82.9	34.7	26.9	160.2	T5, T6

n = 5 (day 62 – 72) and 67 (72 – 204), SD = standard deviation, min and max = minimum and maximum value. * Significantly different between CSTR's and from corresponding CSTR (Sig) at the 0.05 level. Anderson-Darling and Bartlett Test $P < 0.05$ for all comparisons. Ψ n = 66 due to missingness in bacteria and AOB data, except for ML bacteria in C6 (n = 65) and C3 (n = 64).

Table IV. 8. Summary of biotic conditions within Test CSTR's (day 62 - 204).

CSTR	Test 1					Test 2					Test 3					Test 4					Test 5					Test 6				
Day 62 - 70																														
Abundance	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig	Mean	SD	Min	Max	Sig
ML Viruses × 10^8 mL^{-1}*	1.31	1.08	0.28	2.98		0.46	0.07	0.39	0.56	C1 - C3 T4	1.00	0.36	0.67	1.59		2.43	0.86	1.40	3.58		0.99	0.32	0.67	1.38		0.68	0.27	0.50	1.16	
ML Bacteria × 10^7 mL^{-1}	3.55	2.72	0.99	7.10			0.94	1.63	3.73		3.01	2.41	0.90	6.91		1.31	1.38	0.25	3.05		3.18	2.28	1.36	7.11		4.60	3.78	1.09	9.99	
ML AOB × 10^5 mL^{-1}	0.71	0.33	0.22	1.11			0.50	0.47	1.67		1.12	0.83	0.10	2.31		0.88	0.86	0.11	2.05		1.01	1.04	0.23	2.79		0.59	0.31	0.11	0.84	
Day 72 - 204																														
ML Viruses × 10^8 mL^{-1}*	1.11	0.57	0.20	2.85	C4 - C6 T2, T6	0.65	0.30	0.27	1.46	C1 - C3 T3 - T5	1.12	0.70	0.33	3.27	C1, C4, C5 T2, T6	1.00	0.40	0.43	2.31	C1, C4, C5 T2, T6	1.00	0.51	0.21	2.43	C1, C4, C5 T3	0.58	0.29	0.16	1.26	C1, C2, C3 T1, T3, - T5
ML Bacteria × 10^7 mL^{-1}Ψ	1.23	1.43	0.04	5.86		2.18	3.28	0.01	21.42		1.93	2.58	0.00	17.17		1.92	3.49	0.00	21.02		1.44	2.26	0.01	14.99		1.87	3.17	0.01	20.62	
ML AOB × 10^5 mL^{-1}*Ψ	1.23	1.49	0.03	6.56		0.70	0.94	0.02	4.34	C3 T4	1.86	4.07	0.04	22.66		2.26	3.46	0.01	21.82	C3 T2, T5	0.77	0.95	0.01	4.66	C3 T4	0.80	0.98	0.02	5.12	T4
Day 62 - 204																														
D_1	153.7	49.6	55.5	280.3	T6	156.2	78.9	19.6	340.7	C1 T6	144.2	47.6	44.0	299.2	C1 T6	168.7	56.5	21.1	291.2	T5, T6	131.8	59.3	47.4	278.5	C1:C6 T4	109.7	47.7	40.1	291.6	C1:C4, C6 T1:T4
D_2	72.3	32.4	15.4	147.2	T6	69.7	42.0	5.8	173.3	C1 T6	61.4	26.8	8.2	161.9	C1 T6	76.5	33.5	4.8	168.1	T5, T6	54.5	33.0	12.5	136.3	C1:C6 T4	41.6	21.4	14.1	126.7	C1:C4, C6 T1:T4

n = 5 (day 62 – 72) and 67 (72 – 204), SD = standard deviation, min and max = minimum and maximum value. * Significantly different between CSTR's and from corresponding CSTR (Sig) at the 0.05 level. Anderson-Darling and Bartlett Test $P < 0.05$ for all comparisons. Ψ n = 66 due to missingness in bacteria and AOB data, except for ML bacteria in T4 and T5 (n = 65).

Table IV. 9. ML virus linear mixed-effects model.

Correlation Structure	Estimate (Phi)	95% CI's (min, max)	*P*-Value$^{\Psi}$			
Continuous AR(1) ~Day|Reactor	0.8752334	0.85, 0.90	< 0.0001	***		
Random Effect ~1|Reactor	Estimate (Intercept)	95% CI's (min, max)	Residual	95% CI's (min, max)		
Reactor	0.1044857	0.06, 0.19	0.216438	0.20, 0.24		
Coefficient	Estimate	SE	t-Value	VIF	95% CI's (min, max)	*P*-Value
Intercept	8.699633	0.3359819	25.893250	NA	8.05, 9.36	< 0 x 10^{-4} ***
Influent Calcium	-0.014279	0.0061565	-2.319392	1.01	-0.03, -0.00	0.0206 *
ML SS	0.014676	0.0043358	3.384794	1.01	0.01, 0.02	0.0007 ***
ML Bacteria	0.030969	0.0086346	3.586611	1.02	0.01, 0.05	0.0004 ***
Effluent Nitrate	-0.196696	0.0658093	-2.988878	2.63	-0.33, -0.07	0.0029 **
ML pH	0.065139	0.0125068	5.208276	1.14	0.04, 0.09	< 0 x 10^{-4} ***
HRT	-0.015949	0.0039751	-4.012202	1.01	-0.02, -0.01	0.0001 ***
Effluent Chloride	0.036196	0.0111755	3.238831	2.42	0.01, 0.06	0.0012 **

All measured biological, operational and environmental variables included as covariates, $K_i = 51$, $n = 847$, $R^2 = 0.30$. Anderson-Darling Test p = 0.05081 for fixed effects and p = 0.7303 for random effect. All VIF scores < 3. SE = standard error. CI = confidence interval. $^{\Psi}$ ANOVA *P*-Value comparing models with identical fixed and random effects but with and without correlation structure. ° $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

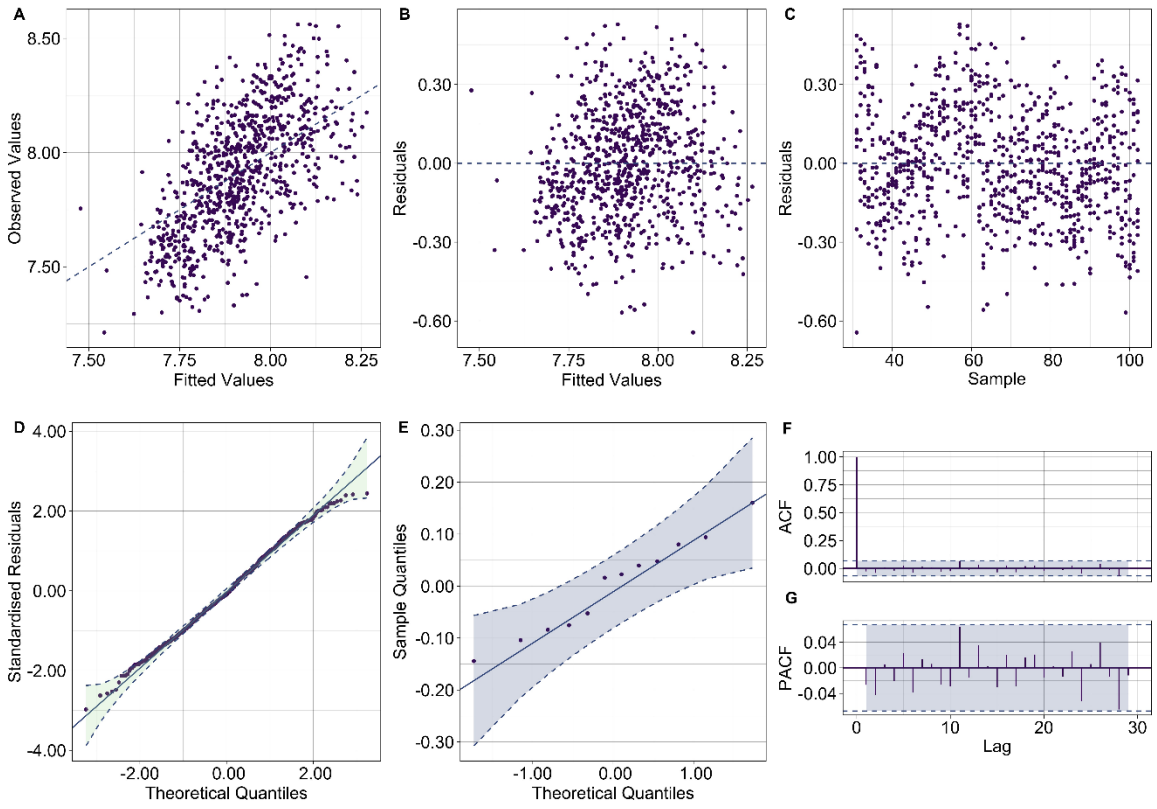

Figure IV. 4. Diagnostic plots assessing linearity (**A**, **B**), homoscedasticity (**B**, **C**), residual normality (**D**, **E**) and residual independence (**F**, **G**) of the ML virus linear mixed-effects model.

Table IV. 10. Summary of CCA analysis.

	CCA Axis			
	1	2	3	4
Eigenvalue	0.33740	0.18679	0.14861	0.11414
Species-Environment Correlation	0.9225150	0.8773796	0.8096894	0.7780930
Cumulative Percentage Variance				
Of Species-Environment Relationship	27.66	42.97	55.15	64.50
Permutation Test (Axis)[Ψ]				
F ratio	21.4672	11.8833	9.4548	7.2617
P-Value	< 0.001***	< 0.001***	< 0.001***	< 0.001***

	Explanatory Variable							
	ML pH	ML Temperature	Influent Silicon	ML DO	Influent Potassium	ML Virus	Effluent Chloride	Influent Virus
Permutation Test (Marginal Effect)								
F ratio	6.4902	7.7511	5.3143	5.7537	4.1553	5.2868	3.3403	4.2046
P-Value	< 0.001***	< 0.001***	< 0.001***	< 0.001***	< 0.001***	< 0.001***	< 0.001***	< 0.001***
Intraset Correlation Coefficient								
Axis 1	-0.82154	-0.63896	-0.59534	0.25438	0.240788	-0.36994	-0.03044	-0.2383
Axis 2	-0.35612	0.531503	0.22845	-0.20089	0.548216	-0.01439	0.100166	-0.13418
Axis 3	-0.12849	-0.23138	0.442727	0.013952	-0.18362	0.156203	-0.48985	-0.27078
Axis 4	0.257979	-0.43915	0.099615	0.750875	0.234694	0.204397	0.044868	0.120185
VIF	2.48	1.80	1.73	1.43	1.81	1.32	1.45	2.31

All measured biological, operational and environmental variables included as explanatory variables except ML Bacteria and AOB, n = 847. Global permutation test F ratio = 5.5443, P -Value <0.001***. Influent manganese and magnesium and effluent nitrate removed due to VIF scores > 3. [Ψ] CCA axis 5 and 6 also significant, $P < 0.01$ and $P < 0.05$ respectively. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Table IV.10. Continued.

	Explanatory Variable					
	Influent Lead	Influent Aluminium	Influent Nickel	ML SS	Effluent Phosphate	Influent Copper
Permutation Test (Marginal Effect)						
F ratio	3.0639	2.9808	4.5291	2.5676	2.0890	2.6844
P-Value	< 0.001***	< 0.001***	< 0.001***	< 0.001***	< 0.001***	< 0.001***
Intraset Correlation Coefficient						
Axis 1	-0.48691	-0.29001	-0.11167	0.108932	0.259385	-0.36607
Axis 2	-0.12983	-0.56522	0.35392	-0.09087	0.488573	-0.27086
Axis 3	0.183579	0.173983	0.439977	-0.15871	0.126702	0.129729
Axis 4	-0.1484	-0.21664	-0.06761	0.03712	0.028471	-0.13441
VIF	1.90	2.63	1.68	1.05	1.48	2.73

Table IV. 11. Abiotic similarity statistics for each CSTR pair.

CSTR	C1	C2	C3	C4	C5	C6	T1	T2	T3	T4	T5	T6
Control 1	0.000	0.091	0.132	0.038	0.080	0.094	0.026	0.108	0.022	0.034	0.022	0.013
Control 2	0.221	0.000	0.013	0.014	0.010	0.011	0.050	0.015	0.042	0.028	0.035	0.027
Control 3	0.290	0.006	0.000	0.035	0.009	0.010	0.070	0.019	0.070	0.043	0.056	0.047
Control 4	0.010	0.006	0.273	0.000	0.014	0.018	0.025	0.030	0.015	0.007	0.016	0.007
Control 5	0.080	0.027	0.059	0.177	0.000	0.005	0.042	0.012	0.036	0.021	0.030	0.024
Control 6	0.033	-0.004	0.064	0.031	0.116	0.000	0.045	0.010	0.038	0.025	0.030	0.026
Test 1	0.073	0.096	0.059	0.126	0.033	0.108	0.000	0.057	0.013	0.018	0.017	0.014
Test 2	0.223	0.151	0.056	0.093	0.107	0.131	0.014	0.000	0.045	0.045	0.040	0.034
Test 3	0.000	0.172	0.039	0.012	0.072	0.157	0.068	0.148	0.000	0.016	0.003	0.003
Test 4	0.009	0.057	0.101	0.075	0.019	0.042	0.002	0.172	0.121	0.000	0.016	0.011
Test 5	0.038	0.147	0.078	0.107	0.078	0.132	0.014	0.056	0.026	0.022	0.000	0.004
Test 6	0.179	0.110	0.129	0.097	0.018	0.098	0.128	0.154	0.083	0.032	-0.004	0.000

ANOSIM R statistic and Adonis R^2 statistic shown in lower left and upper right portion of table respectively, Bonferroni corrected $P > 0.05$ for all CSTR pairs for both analyses. n = 72.

Table IV. 12. Functional synchrony of each CSTR pair.

CSTR	C1	C2	C3	C4	C5	C6	T1	T2	T3	T4	T5	T6
Control 1	1	0.61	0.62	0.64	0.53	0.61	0.58	0.49	0.64	0.58	0.52	0.55
Control 2	0.69	1	0.69	0.67	0.58	0.64	0.51	0.64	0.61	0.58	0.57	0.59
Control 3	0.80	0.77	1	0.72	0.57	0.66	0.61	0.6	0.63	0.58	0.65	0.68
Control 4	0.71	0.75	0.74	1	0.68	0.78	0.70	0.68	0.71	0.69	0.71	0.78
Control 5	0.67	0.63	0.67	0.8	1	0.77	0.69	0.74	0.71	0.66	0.70	0.60
Control 6	0.68	0.62	0.69	0.73	0.68	1	0.75	0.80	0.75	0.72	0.71	0.77
Test 1	0.51	0.38	0.51	0.55	0.52	0.57	1	0.69	0.75	0.79	0.74	0.76
Test 2	0.49	0.47	0.59	0.56	0.58	0.58	0.55	1	0.74	0.7	0.73	0.67
Test 3	0.58	0.57	0.62	0.70	0.67	0.66	0.67	0.78	1	0.78	0.75	0.75
Test 4	0.38	0.36	0.52	0.51	0.43	0.44	0.67	0.65	0.68	1	0.68	0.75
Test 5	0.55	0.63	0.72	0.71	0.68	0.64	0.62	0.75	0.69	0.69	1	0.69
Test 6	0.61	0.60	0.65	0.73	0.66	0.64	0.66	0.63	0.70	0.58	0.76	1

Synchronicity coefficients for effluent CODs and NH_4^+ - N, lower left and upper right portion of table respectively. *n* = 72. Synchrony coefficients > 0.60 are coloured light blue.

Table IV. 13. Biotic synchrony of each CSTR pair.

CSTR	C1	C2	C3	C4	C5	C6	T1	T2	T3	T4	T5	T6	C1	C2	C3	C4	C5	C6	T1	T2	T3	T4	T5	T6
							A												B					
Control 1	1	0.66	0.53	0.21	0.45	0.56	0.76	0.42	0.58	0.30	0.82	0.61	1											
Control 2	0.09	1	0.77	0.37	0.60	0.63	0.57	0.58	0.55	0.34	0.54	0.53	0.60	1										
Control 3	0.14	0.56	1	0.33	0.60	0.6	0.63	0.70	0.67	0.23	0.53	0.5	0.55	0.73	1									
Control 4	0.11	0.21	0.59	1	0.45	0.1	0.11	0.27	0.29	0.52	0.04	0.28	0.41	0.67	0.64	1								
Control 5	0.06	0.29	0.47	0.57	1	0.6	0.43	0.46	0.57	0.48	0.43	0.42	0.41	0.64	0.59	0.54	1							
Control 6	-0.10	0.06	0.50	0.64	0.46	1	0.52	0.46	0.38	0.24	0.56	0.44	0.32	0.60	0.42	0.54	0.46	1						
Test 1	0.13	0.13	0.45	0.48	0.25	0.49	1	0.66	0.66	0.23	0.78	0.56	0.63	0.68	0.55	0.54	0.49	0.52	1					
Test 2	-0.26	-0.15	0.24	0.46	0.47	0.71	0.31	1	0.57	0.35	0.5	0.44	0.54	0.61	0.60	0.42	0.52	0.43	0.60	1				
Test 3	0.03	0.11	0.28	0.69	0.60	0.62	0.31	0.43	1	0.43	0.61	0.57	0.33	0.59	0.54	0.50	0.68	0.47	0.59	0.59	1			
Test 4	0.33	0.45	0.27	-0.09	-0.13	-0.16	0.11	-0.42	-0.25	1	0.33	0.38	0.33	0.64	0.50	0.55	0.54	0.45	0.46	0.29	0.60	1		
Test 5	-0.12	0.02	0.23	0.48	0.57	0.32	0.05	0.54	0.41	-0.20	1	0.64	0.44	0.60	0.51	0.50	0.60	0.53	0.74	0.67	0.76	0.48	1	
Test 6	0.00	0.34	0.54	0.45	0.34	0.19	0.32	0.19	0.23	-0.04	0.36	1	0.42	0.62	0.52	0.37	0.40	0.45	0.48	0.37	0.43	0.55	0.50	1

Synchronicity coefficients for ML virus (*n* = 72), bacteria (*n* = 66) and AOB (*n* = 71) abundance, lower left, upper right (**A**) and lower left (**B**) portion of table respectively. Synchrony coefficients > 0.60 are coloured light blue.

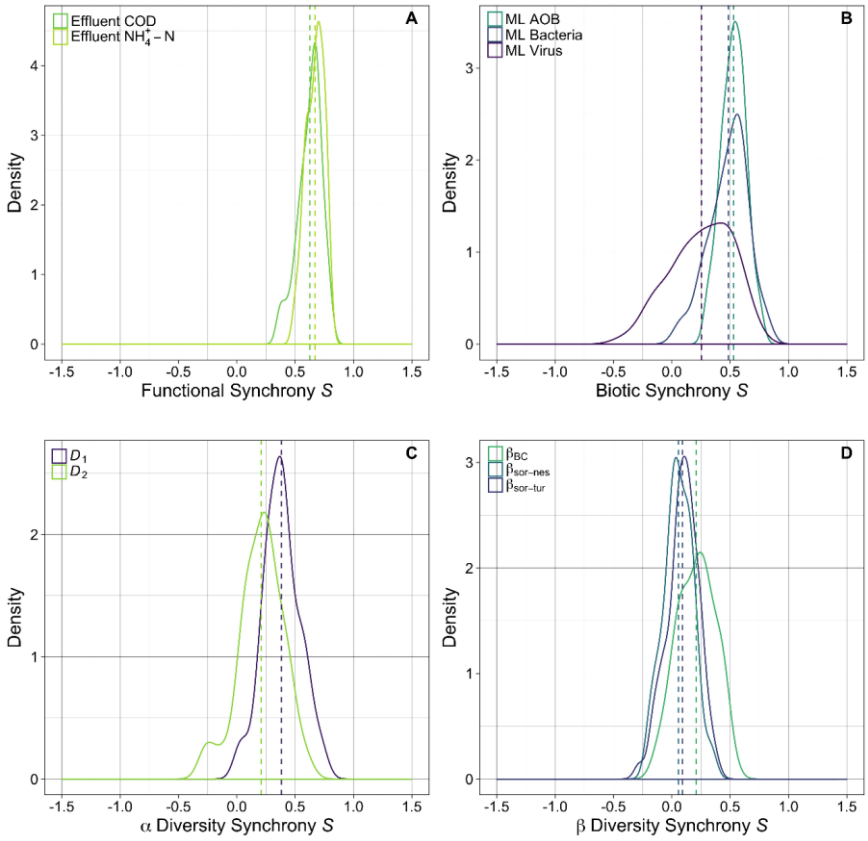

Figure IV. 5. Functional (**A**), total abundance (**B**), α diversity (**C**) and β diversity (**D**) synchrony across all CSTR's from day 62 onwards. Dashed lines represent overall mean synchrony coefficients for each component.

Table IV. 14. Microbial community similarity statistics for each CSTR pair.

CSTR	C1	C2	C3	C4	C5	C6	T1	T2	T3	T4	T5	T6
Control 1	0.000	0.065	0.046	0.054	0.109	0.072	0.105	0.117	0.125	0.059	0.156	0.232
Control 2	0.174	0.000	0.076	0.066	0.087	0.067	0.090	0.099	0.103	0.083	0.125	0.194
Control 3	0.101	0.177	0.000	0.069	0.124	0.082	0.119	0.132	0.143	0.061	0.180	0.256
Control 4	0.210	0.241	0.373	0.000	0.076	0.069	0.096	0.099	0.098	0.072	0.125	0.192
Control 5	0.150	0.191	0.294	0.463	0.000	0.056	0.074	0.071	0.078	0.126	0.074	0.136
Control 6	0.173	0.143	0.402	0.300	0.149	0.000	0.064	0.077	0.089	0.085	0.100	0.170
Test 1	0.197	0.338	0.287	0.208	0.191	0.579	0.000	0.055	0.052	0.109	0.071	0.111
Test 2	0.358	0.258	0.179	0.259	0.399	0.415	0.184	0.000	0.052	0.124	0.062	0.093
Test 3	0.243	0.360	0.214	0.119	0.243	0.646	0.148	0.648	0.000	0.126	0.061	0.091
Test 4	0.393	0.287	0.128	0.115	0.339	0.401	0.139	0.853	0.557	0.000	0.165	0.236
Test 5	0.208	0.200	0.420	0.139	0.392	0.187	0.592	0.590	0.321	0.250	0.000	0.094
Test 6	0.218	0.162	0.304	0.223	0.404	0.318	0.794	0.419	0.252	0.771	0.253	0.000

ANOSIM *R* statistic and Adonis R^2 statistic shown in lower left and upper right portion of table respectively, Bonferroni corrected $P > 0.05$ for all CSTR pairs for both analyses.

Table IV. 15. Community concordance among CSTR's.

CSTR	C1	C2	C3	C4	C5	C6	T1	T2	T3	T4	T5	T6
Control 1	0.00											
Control 2	0.39	0.00										
Control 3	0.29	0.32	0.00									
Control 4	0.16	0.43	0.17	0.00								
Control 5	0.49	0.19	0.31	0.49	0.00							
Control 6	0.46	0.20	0.42	0.53	0.24	0.00						
Test 1	0.39	0.25	0.19	0.31	0.23	0.31	0.00					
Test 2	0.62	0.31	0.53	0.61	0.45	0.43	0.46	0.00				
Test 3	0.63	0.38	0.51	0.59	0.45	0.47	0.54	0.46	0.00			
Test 4	0.62	0.45	0.48	0.54	0.41	0.50	0.50	0.56	0.14	0.00		
Test 5	0.44	0.16	0.41	0.51	0.25	0.21	0.27	0.32	0.46	0.55	0.00	
Test 6	0.55	0.28	0.42	0.53	0.25	0.35	0.36	0.47	0.58	0.58	0.31	0.00

m^2 values from concordance analysis, $P < 0.001$ for all CSTR pairs.

Table IV. 16. β diversity synchrony of each CSTR pair.

CSTR	C1	C2	C3	C4	C5	C6	T1	T2	T3	T4	T5	T6	C1	C2	C3	C4	C5	C6	T1	T2	T3	T4	T5	T6
							A												B					
Control 1	1	0.07	0.11	0.06	0.1	0.03	0.15	0.25	0.22	0.16	0.2	0.31	1											
Control 2	-0.02	1	0.14	-0.18	0.07	0.15	0.11	0.24	0.24	0.03	0.14	-0.08	0.3	1										
Control 3	-0.1	0.28	1	-0.09	-0.16	-0.06	0.13	0.1	-0.12	0.08	0.18	-0.02	0.2	0.23	1									
Control 4	0.22	0.12	0.05	1	0.13	0.06	0.06	-0.15	0.05	0.2	-0.03	0.15	0.14	0.14	0.24	1								
Control 5	0.07	0	-0.04	0.16	1	0.38	0.13	0.15	0.25	-0.06	0.04	0.24	0.22	0.46	0.3	0.25	1							
Control 6	-0.18	0.03	-0.07	0.19	0.35	1	0.05	0.25	-0.04	-0.28	0.03	0.1	0.06	0.29	0.29	-0.03	0.42	1						
Test 1	0.17	0	-0.02	0.06	0.05	-0.15	1	0.18	0.31	-0.13	0.33	0.22	0.08	0.08	0.07	-0.02	0.03	-0.11	1					
Test 2	-0.09	0.01	0.21	0.04	0	0.16	-0.2	1	0.26	0.01	0.06	0.11	0.41	0.34	0.02	0.01	0.09	0.19	0.18	1				
Test 3	0.21	0.16	-0.16	0.1	0.21	-0.06	0.11	0.05	1	0.01	0.18	0.18	0.34	0.39	0.08	0.12	0.18	0.04	0.34	0.41	1			
Test 4	0.03	0.06	0.14	0.13	0	-0.12	-0.1	0.04	-0.13	1	0.11	-0.07	0.13	0.24	-0.07	0.33	0.11	-0.06	0.26	0.33	0.45	1		
Test 5	0.12	0.05	0.11	0.16	0.13	0.14	-0.01	-0.06	0.02	-0.02	1	0.04	0.21	0.44	0.03	0.24	0.44	0.27	0.36	0.31	0.45	0.53	1	
Test 6	0.22	-0.19	0.05	0.19	0.33	0.16	0.11	0.02	0.06	0.12	-0.02	1	0.11	0.22	0.03	0.27	0.09	-0.14	0.26	0.19	0.4	0.46	0.18	1

Synchronicity coefficients for $\beta_{sor\text{-}nes}$, $\beta_{sor\text{-}tur}$ and β_{BC}, lower left, upper right (**A**) and lower left (**B**) portion of table respectively. Synchrony coefficients > 0.60 are coloured light blue.

Table IV. 17. α diversity synchrony of each CSTR pair.

CSTR	C1	C2	C3	C4	C5	C6	T1	T2	T3	T4	T5	T6
Control 1	1	0.52	0.71	0.69	0.59	0.26	0.02	0.28	0.38	0.33	0.54	0.39
Control 2	0.27	1	0.62	0.56	0.4	0.37	0.27	0.39	0.23	0.39	0.32	0.28
Control 3	0.47	0.3	1	0.74	0.59	0.5	0.36	0.57	0.43	0.38	0.5	0.42
Control 4	0.41	0.13	0.61	1	0.31	0.21	0.04	0.39	0.32	0.16	0.34	0.27
Control 5	0.44	0.15	0.53	0.18	1	0.58	0.36	0.42	0.4	0.43	0.59	0.51
Control 6	0.17	0.23	0.38	0.07	0.46	1	0.42	0.47	0.21	0.24	0.23	0.36
Test 1	-0.26	0.28	0.06	-0.25	0.07	0.32	1	0.62	0.3	0.32	0.24	0.17
Test 2	-0.13	0.29	0.34	0.18	0.23	0.35	0.55	1	0.43	0.28	0.28	0.05
Test 3	0.04	0.25	0.06	-0.13	0.18	0.09	0.43	0.39	1	0.5	0.42	0.38
Test 4	-0.01	0.22	0	-0.27	0.23	0.14	0.23	0.08	0.44	1	0.4	0.27
Test 5	0.28	0.07	0.39	0.19	0.42	0.09	0.12	0.21	0.27	0.21	1	0.19
Test 6	0.28	0.24	0.33	-0.01	0.41	0.25	0.12	0.09	0.29	0.26	0.07	1

Synchronicity coefficients for D_1 and D_2 diversity indices, upper right and lower left portion of table respectively. Synchrony coefficients > 0.60 are coloured light blue.

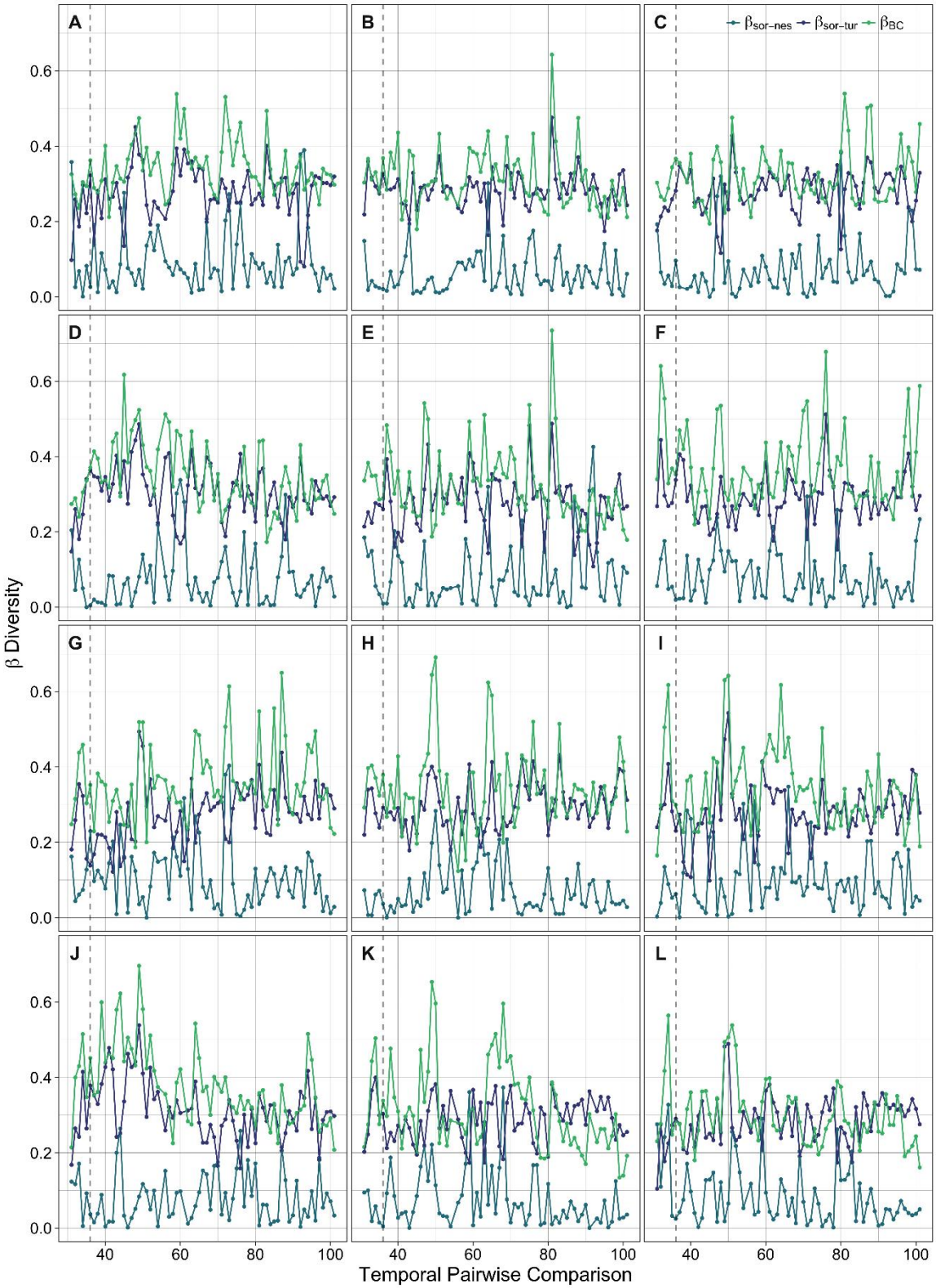

Figure IV. 6. Temporal β diversity trajectories across all CSTR's from day 62 onwards. Dashed line represents the end of acclimatisation. (**A – F**) Control 1 – 6. (**G – L**) Test 1 – 6.

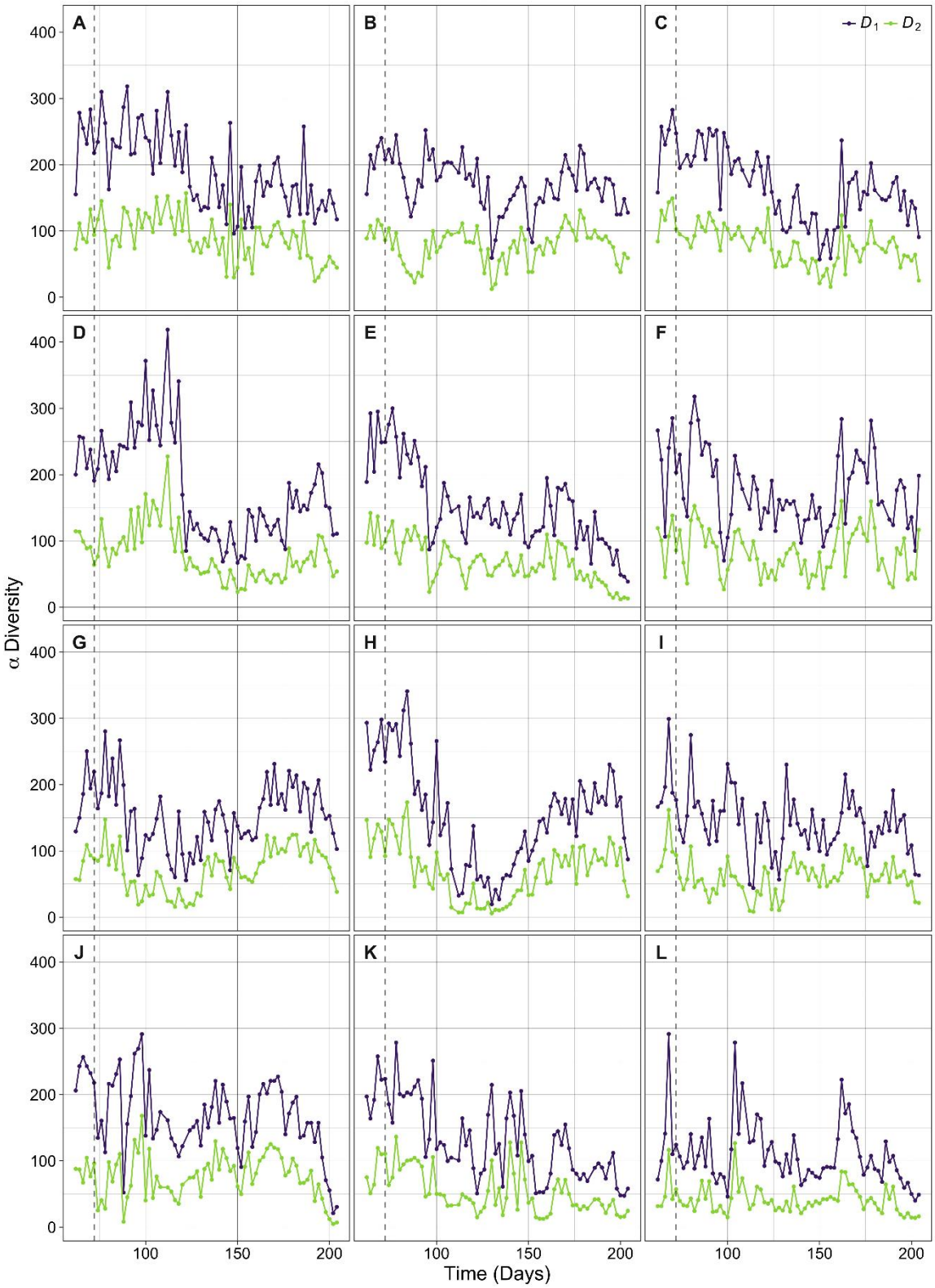

Figure IV. 7. Temporal diversity trajectories across all CSTR's from day 62 onwards. Dashed line represents the end of acclimatisation. (**A – F**) Control 1 – 6. (**G – L**) Test 1 – 6.

APPENDIX V

SUPPLEMENTARY INFORMATION - EVIDENCE OF PREDATOR-PREY DYNAMICS BETWEEN BACTERIOPHAGE AND AMMONIA OXIDISING BACTERIA IN AN ENGINEERED MICROBIAL SYSTEM

V.1. Supplementary Theory

V.1.1. Extension of LV Equations

Rearranging *Eq. 5.1* and taking its first derivative by applying the chain rule gives,

$$\frac{dx}{dt} = x(\mu - \varphi V)\,, \qquad Eq.V1$$

$$\frac{1}{x}\frac{dx}{dt} = \varphi\left(\frac{\mu}{\varphi} - V\right)\,, \qquad Eq.V2$$

$$\frac{df(x(t))}{dt} = \frac{df}{dx}\frac{dx}{dt}\,, \qquad Eq.V3$$

$$\frac{d\,\ln x}{dt} = \frac{d\,\ln(x)}{dx}\frac{dx}{dt}\,. \qquad Eq.V4$$

However

$$\frac{d\,\ln(x)}{dx} = \frac{1}{x}\,, \qquad Eq.V5$$

and so

$$\frac{d\,\ln x}{dt} = \frac{d\,\ln(x)}{dx}\frac{dx}{dt} = \frac{1}{x}\frac{dx}{dt}\,. \qquad Eq.V6$$

Thus the first derivative of *Eq. V2* is,

$$\frac{d\,\ln(x)}{dt} = \varphi\left(\frac{\mu}{\varphi} - V\right). \qquad Eq.5.3$$

Similarly *Eq. 5.2* can be rearranged and its first derivative taken,

$$\frac{d\,\ln(V)}{dt} = -\delta\left(\frac{m}{\delta} - x\right). \qquad Eq.5.4$$

V.1.2. More complex models

More complex models are conceivable to take account of bacterial and viral loss (Campbell, 1961). However as such expressions would be a constant multiplied by prey or virus abundance these terms

would disappear by the second differentiation. For example if α were the bacterial growth rate (time^{-1}) and *mcrt* were the mean cell retention time then,

$$\frac{dx}{dt} = \alpha x - \varphi x V - \frac{1}{mcrt} x \,. \qquad Eq.V7$$

The first and second differentials would be,

$$\frac{d\,\ln(x)}{dt} = \varphi\left(\frac{\alpha}{\varphi} - \frac{1}{\varphi mcrt} - V\right), \qquad Eq.V8$$

$$\frac{d^2\,\ln(x)}{dt} = -\varphi\frac{dV}{dt}\,. \qquad Eq.V9$$

Which is equal to *Eq. 5.5* and thus the prediction still stands.

Moreover the time delay between prey infection and death introduced in Campbell's (1961) model may be ignored if the time steps are larger than the time taken for cell lysis, a likely situation here.

V.1.3. AOB Mortality and Growth

Using the slope of the SMA regression analysis (Table. 5.1) we can obtain φ, -2.95 × 10^{-9} week^{-1}, and δ, -6.34 × 10^{-8} week^{-1}, virus lysis and replication rates respectively. Thus using φ and the geometric mean of V (1.195×10^9 viruses mL^{-1}) we can obtain the typical mortality rate of AOB due to virus predation (φV), as well as minimum (φV_{min}) and maximum (φV_{max}) rates using minimum and maximum values of V (0.318×10^9 and 3.407×10^9 viruses mL^{-1} respectively). Moreover negating ($\neg$) φV and considering a mcrt of 11.1 ± 3.1 days (determined across the two years) μ can be estimated for AOB (*Eq. V10*).

$$\mu = \frac{1}{mcrt} + \neg\varphi V. \qquad Eq.V10$$

APPENDIX VI

SUPPLEMENTARY INFORMATION - A WASTEWATER PERSPECTIVE ON VIRAL AND MICROBIAL ABUNDANCES AND VIRUS-MICROBE RATIOS

VI.1. Supplementary Results

Table VI. 1. Virus and bacterial densities per g of suspended solids within 95% quantiles.

	Median (10^{11} g $^{-1}$)	Minimum (10^{10} g $^{-1}$)	Maximum (10^{11} g $^{-1}$)
Virus			
ML FS[a]	4.77	6.17	13.19
Inf FS[b]	0.52	1.00	1.53
ML CSTR[c]	9.53	20.4	77.97
Bacteria			
ML FS[a]	1.30	2.21	3.99
Inf FS[b]	0.06	0.02	0.38
ML CSTR[c]	1.02	0.70	13.81

Values calculated from virus and bacterial abundance per millilitre and the SS data (not shown). *n* = 102[a], 95[b], 847[c]. ML = Mixed Liquor, Inf = Influent, FS = Full Scale, CSTR = Continuously Stirred Tank Reactor

Table VI. 2. Linear regression models of virus and bacterial abundance.

System	Intercept	Intercept 95% CI (min, max)	Estimate	Estimate 95% CI (min, max)	*P*-value	
ML FS[a]	9.134343	8.21, 10.06	-0.006707	-0.12, 0.10	0.903	
Inf FS[b]	9.04257	8.49, 9.59	-0.13249	-0.21, -0.05	0.001	**
ML C1[c]	7.85153	7.35, 8.35	0.04384	-0.03, 0.12	0.233	
ML C2[c]	7.66341	6.96, 8.36	0.05506	-0.05, 0.16	0.285	
ML C3[e]	7.21249	6.67, 7.76	0.10413	0.93, 0.18	0.009	**
ML C4[c]	8.00629	7.40, 8.62	-0.03010	-0.12, 0.06	0.505	
ML C5[c]	6.40573	5.81, 7.00	0.19819	0.11, 0.28	$< 1.77 \times 10^{-5}$	***
ML C6[d]	6.29583	6.96, 8.36	0.22476	-0.05, 0.16	$< 1.53 \times 10^{-6}$	***
ML T1[c]	7.03787	6.33, 7.75	0.13913	0.04, 0.24	0.009	**
ML T2[c]	6.78831	6.37, 7.20	0.14042	0.08, 0.20	$< 1.15 \times 10^{-5}$	***
ML T3[c]	7.2470	6.61, 7.88	0.1035	0.01, 0.19	0.025	*
ML T4[d]	7.86073	7.42, 8.30	0.02011	-0.04, 0.08	0.535	
ML T5[d]	7.76092	7.09, 8.43	0.02641	-0.07, 0.12	0.592	
ML T6[c]	7.682407	7.11, 8.25	0.005262	-0.08, 0.09	0.898	
ML CSTR[f]	7.35230	7.16, 7.54	0.08106	0.05, 0.11	$< 8.82 \times 10^{-9}$	***
All[g]	5.88830	5.67, 6.08	0.30540	0.28, 0.33	$< 2.2 \times 10^{-16}$	***

n = 102[a], 95[b], 91[c], 70[d], 69[e], 847[f] and 1044[g]. CI = Confidence Intervals, ML = Mixed Liquor, Inf = Influent, FS = Full Scale, CSTR = Continuously Stirred Tank Reactor, C = Control, T = Test. ° $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Table VI. 3. Linear regression model of VMR's and bacterial abundance.

System	Intercept	Intercept 95% CI (min, max)	Estimate	Estimate 95% CI (min, max)	*P*-value	
All	5.88830	-0.72, -0.67	-0.69460	5.69, 6.09	$< 2.2 \times 10^{-16}$	***

n = 1044. CI = Confidence Intervals. ° $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

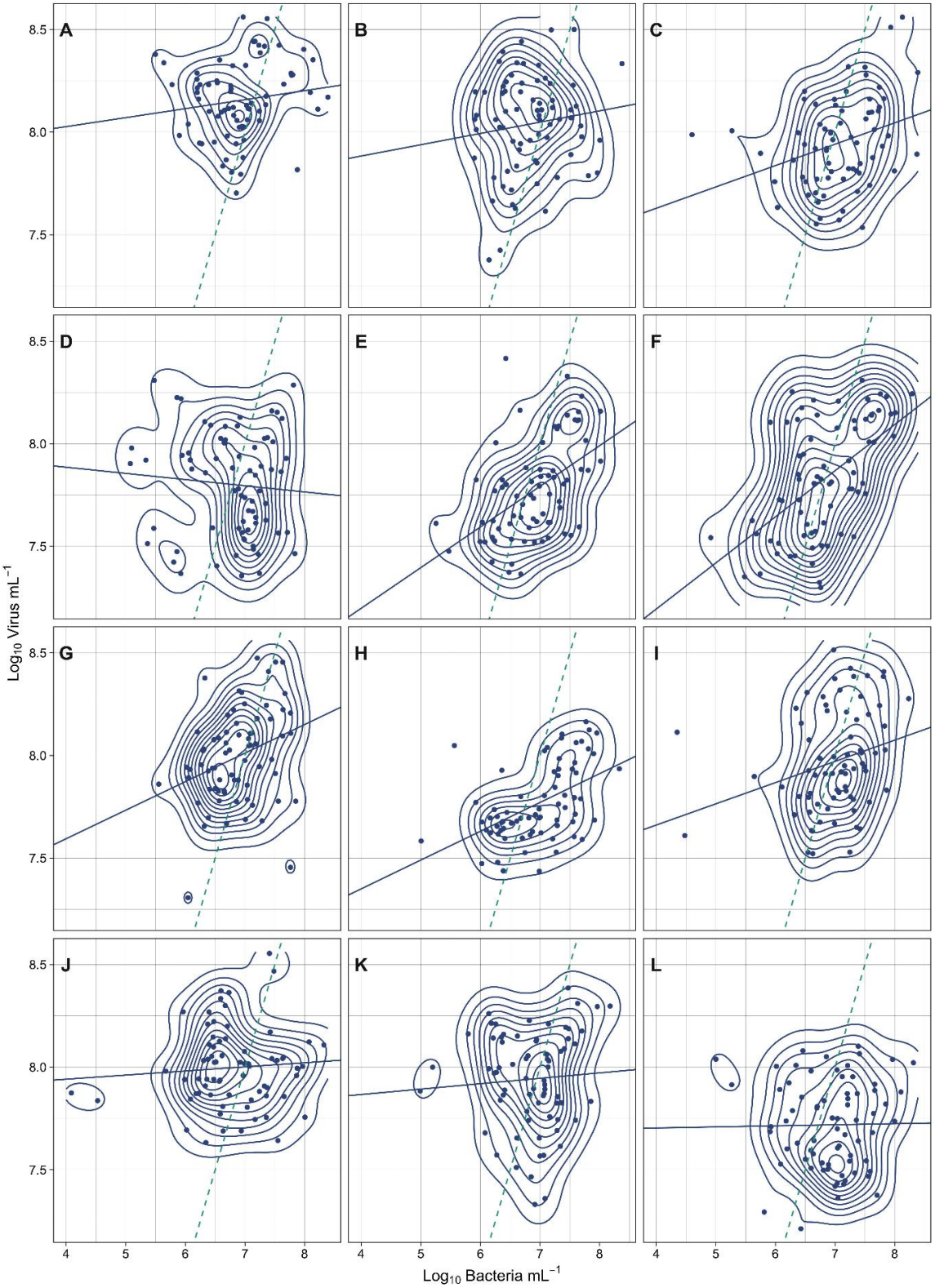

Figure VI. 1. Virus-bacteria relationships in all 12 (**A – L**) lab scale systems as determined by individual linear regression models. Blue solid lines denote best fit linear regression for each system and dashed light blue line depicts a 10:1 linear relationship.

APPENDIX VII

SUPPLEMENTARY INFORMATION - PRELIMINARY METAGENOMIC CHARACTERISATION OF WASTEWATER VIRUSES

VII.1. Supplementary Methods

To the best of our knowledge all current assembly-based metagenomic analysis methods, whether on viral enriched or bacterial samples respectively, rely heavily on post processing of assembled contigs. Yet for viral metagenomes (viromes) in particular such an approach discards a hugely rich source of additional information, namely the intermediate assembly graph structure. As an example two circular, virus like chromosomes (~ 38 kb each) discovered in the effluent free virome were isolated in Bandage (Fig. VII. 1). The left chromosome (Fig. VII. 1 A) is presented as a single entry in the kmer graph and thus constitutes a single contig in the final assembly. In contrast the right chromosome (Fig. VII. 1 B), which is clearly circular and complete (barring assembly errors), would represent two long (~ 15 kb each) and four short (~ 2 kb each) unrelated contigs. This arises because the four short contigs, small regions of heterogeneity in the chromosome (top and bottom Fig. VII. 1 B), represent two possible branches that the contig building algorithm has to choose between, as they are similar in depth (reads mapping to each pathway) it halts building and splits up the "single" chromsosome.

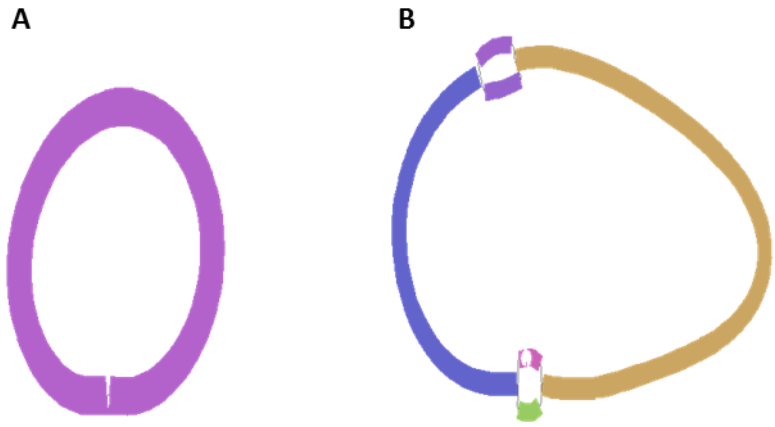

Figure VII. 1. Assembly graph of two viral like chromosomes discovered in the effluent free virome, here both the single contig (**A**) and all the viral associated contigs (**B**) would be kept as two separate individual virus sub-graphs for future analysis.

Such a network of related contigs (Fig. VII. 1 B) is likely formed by the presence and co-assembly of two highly related viruses, which only differ in nucleotide sequence at two points in the genome. Thus, the assembler produces the correct structures to fully describe an associated virus chromosome at an intermediate step, but is unable to fully capitalize on this information to produce a complete circular genome. We attempted to take advantage of the "correct" graph structure to enrich the virus-like content of the per-sample assemblies, thus both the left and right chromosome would be considered as individual virus sub-graphs and used in future analysis.

VII.2. Supplementary Results

Table VII. 1. Summary of assembly statistics.

	Wastewater Virome			
	Inf Free	ML Free	ML Temp	Eff Free
Reads				
Raw	2092974	2373697	1644812	2875027
Quality Trimmed	1334137	1386106	832684	1803162
Contigs				
Total	308	6055	439	41512
Max Length	67102	86159	206488	197766
Mean Length	4791	2370	3840	1432
n50	13388	4811	6563	2033
Proteins				
Total	9906	92577	11311	348743
Homologous Viral Proteins	924	7860	864	27274
Individual Virus sub-graphs*				
Circular	2	11	3	19
Total	18	60	14	192
Reads Mapped to ISV's$^{\Psi}$	80704	353731	61800	770372

*inclusive only of those sub-graph sequences > 20k bp. $^{\Psi}$ i.e. "virus-only" reads. Inf = influent, Eff = effluent, Temp = temperate viruses, ISV's = Individual Virus sub-graphs.

Table VII. 2. Composition of wastewater viromes determined by similarity to known nucleotide sequences at the family level.

Viral Type	Viral Family	Host	Wastewater Virome			
			Inf Free	ML Free	ML Temp	Eff Free
dsDNA	*Adenoviridae*	Eukarya	0.00	0.16	0.04	0.04
	Alloherpesviridae	Eukarya	0.03	1.37	4.34	0.00
	Ascoviridae	Eukarya	0.02	0.16	0.19	0.00
	Baculoviridae	Eukarya	0.07	0.66	1.58	0.08
	Chrysoviridae	Eukarya	0.00	0.06	0.00	0.00
	Herpesviridae	Eukarya	0.09	3.41	2.36	0.43
	Iridoviridae	Eukarya	0.05	0.00	0.13	0.00
	Marseilleviridae	Eukarya	0.00	0.03	0.00	0.00
	Mimiviridae	Eukarya	0.19	0.28	0.26	0.16
	Myoviridae	Bacteria, Archaea	54.17	27.55	1.88	24.60
	Nudiviridae	Eukarya	0.05	0.00	0.37	0.00
	Papillomaviridae	Eukarya	0.00	0.00	0.03	0.00
	Partitiviridae	Eukarya	0.64	1.16	0.46	2.78
	Phycodnaviridae	Eukarya	0.14	0.97	2.15	0.55
	Pithoviridae	Eukarya	0.00	0.06	0.14	0.00
	Podoviridae	Bacteria	4.73	5.56	2.29	12.24
	Polydnaviridae	Eukarya	0.17	3.09	14.48	0.20
	Polyomaviridae	Eukarya	0.00	0.03	0.01	0.00
	Poxviridae	Eukarya	0.14	0.56	0.97	0.20
	Reoviridae	Eukarya	0.00	0.00	0.03	0.00
	Rudiviridae	Archaea	0.02	0.00	0.00	0.00
	Siphoviridae	Bacteria, Archaea	18.83	37.14	1.59	47.83
	Tectiviridae	Bacteria	0.02	0.00	0.00	0.00
	Totiviridae	Eukarya	0.02	0.19	0.05	0.04
ssDNA	*Anelloviridae*	Eukarya	0.00	0.00	0.01	0.00
	Circoviridae	Eukarya	0.00	0.22	0.02	0.00
	Genomoviridae	Eukarya	0.00	0.03	0.01	0.00
	Microviridae	Bacteria	4.58	0.66	0.14	2.11
ssRNA	*Arenaviridae*	Eukarya	0.00	0.00	0.01	0.00
	Astroviridae	Eukarya	0.00	0.06	0.03	0.08
	Bromoviridae	Eukarya	0.05	0.56	0.94	0.35
	Flaviviridae	Eukarya	0.60	0.97	0.18	2.35
	Hytrosaviridae	Eukarya	0.00	0.00	0.02	0.00
	Potyviridae	Eukarya	0.02	0.06	0.02	0.08
	Solinviviridae	Eukarya	0.00	0.00	0.01	0.00
	Togaviridae	Eukarya	0.00	0.00	0.01	0.00
	Virgaviridae	Eukarya	0.02	0.06	0.01	0.08
Total Number of Families			22	26	33	18

Phylogenetic assignment of raw sequences was determined using Krakken (Wood and Salzberg, 2014) against the NCBI nt database. Inf = influent, Eff = effluent, Temp = temperate viruses.

Table VII. 3. Viral species present in all wastewater viromes and their relative abundance as determined by similarity to known nucleotide sequences.

Viral Order	Viral Family	Viral Species	Wastewater Virome			
			Inf Free	ML Free	ML Temp	Eff Free
Caudovirales	*Myoviridae*	Acinetobacter virus AP22	0.26	0.84	0.06	4.58
		Aeromonas phage vB_AsaM-56	43.12	22.93	0.62	10.36
		Begomovirus[Ψ]	0.07	0.09	0.01	0.20
		Caulobacter phage Cr30	0.12	0.25	0.14	2.15
		Enterobacteria phage phi92	0.45	0.03	0.10	1.02
		Enterobacteria phage T4	3.72	0.47	0.12	0.55
		Iodobacteriophage phiPLPE	0.14	0.53	0.08	1.13
		Klebsiella phage JD001	0.31	0.31	0.02	0.43
		Spo1virus[Ψ]	0.52	0.37	0.19	1.76
	Podoviridae	Bacillus phage Stitch	0.38	0.47	0.17	1.41
		Bordetella virus BPP1	0.10	0.62	0.03	0.82
		Burkholderia virus Bcep22	0.17	1.06	0.08	1.49
		Phikmvvirus[Ψ]	0.24	0.28	0.01	0.20
		Rhodoferax phage P26218	0.21	0.31	0.03	0.27
		Salmonella virus 9NA	0.15	0.84	0.09	1.29
		unclassified N4likevirus	0.65	0.28	0.01	2.19
	Siphoviridae	Arthrobacter virus Mudcat	0.02	0.09	0.01	0.04
		Azospirillum phage Cd	0.02	0.16	0.01	0.39
		Cellulophaga phage phi10:1	0.19	0.06	0.02	0.55
		Enterococcus phage IME_EF3	0.05	0.16	0.05	0.51
		Escherichia virus K1g	0.53	0.72	0.01	0.08
		Flavobacterium phage 11b	0.05	0.03	0.02	0.86
		Gordonia phage Kita	0.03	0.06	0.01	0.55
		Gordonia phage Wizard	0.09	0.47	0.05	1.02
		Mycobacterium phage Dante	0.05	0.28	0.05	0.74
		Mycobacterium phage Keshu	0.02	0.16	0.01	0.31
		Mycobacterium phage PattyP	0.02	0.19	0.01	0.20
		Mycobacterium virus Brujita	0.05	0.37	0.03	0.59
		Mycobacterium virus Che9d	0.07	0.69	0.02	1.29
		Polaribacter virus P12002S	0.17	0.19	0.15	2.27
		Pseudomonas phage PS-1	0.17	0.09	0.05	0.63
		Pseudomonas virus MP1412	0.84	2.00	0.01	1.25
		Pseudomonas virus PaMx28	0.28	1.37	0.09	1.41
		Pseudomonas virus PaMx74	1.58	4.47	0.25	5.87
		Pseudomonas virus Yua	1.03	2.72	0.03	2.42
		Stenotrophomonas phage S1	0.05	0.06	0.02	0.04
		Verrucomicrobia phage P8625	0.05	0.56	0.04	0.20
Herpesvirales	*Herpesviridae*	Simplexvirus[Ψ]	0.02	0.59	0.79	0.08
		unassigned Betaherpesvirinae	0.02	0.22	0.09	0.12
Microviridae	*Bullavirinae*	unclassified Phix174microvirus	4.58	0.66	0.14	2.07
ssRNA	*Bromoviridae*	Spring beauty latent virus	0.03	0.25	0.02	0.27
	Flaviviridae	Hepacivirus C	0.59	0.97	0.17	2.35
Total of "known" Virome			61.20	47.30	3.85	55.92

Phylogenetic assignment of raw sequences was determined using Krakken (Wood and Salzberg, 2014) against the NCBI nt database. Inf = influent, Eff = effluent, Temp = temperate viruses. [Ψ]only classified to genus level.

Table VII. 4. Most abundant viral species in each wastewater virome and their relative abundance as determined by similarity to known nucleotide sequences.

Viral Order	Viral Family	Viral Species	Host	Wastewater Virome			
				Inf Free	ML Free	ML Temp	Eff Free
Caudovirales	Myoviridae	Aeromonas phage vB_AsaM-56	Gammaproteobacteria	43.12	22.93	0.62	10.36
		Enterobacteria phage T4	Gammaproteobacteria	3.72	0.47	0.12	0.55
		Acinetobacter virus AP22	Gammaproteobacteria	0.26	0.84	0.06	4.58
		Caulobacter phage Cr30	Alphaproteobacteria	0.12	0.25	0.14	2.15
		Spo1virus[Ψ]	Bacilli	0.52	0.37	0.19	1.76
	Podoviridae	unclassified N4likevirus	Gammaproteobacteria	0.65	0.28	0.01	2.19
		Burkholderia virus Bcep22	Betaproteobacteria	0.17	1.06	0.08	1.49
		T7virus	Gammaproteobacteria	0.09	0.28	1.74	0.00
	Siphoviridae	Pseudomonas virus PaMx74	Gammaproteobacteria	1.58	4.47	0.25	5.87
		Pseudomonas virus Yua	Gammaproteobacteria	1.03	2.72	0.03	2.42
		Pseudomonas virus MP1412	Gammaproteobacteria	0.84	2.00	0.01	1.25
		Pseudomonas virus PAE1	Gammaproteobacteria	0.71	1.06	0.00	0.94
		Lactococcus phage 1706	Bacilli	0.55	0.00	0.00	0.00
		Pseudomonas virus PaMx28	Gammaproteobacteria	0.28	1.37	0.09	1.41
		Polaribacter virus P12002S	Flavobacteria	0.17	0.19	0.15	2.27
Herpesvirales	Alloherpesviridae	Ictalurid herpesvirus 1	Eukaryote	0.00	0.12	2.84	0.00
		Cyprinid herpesvirus 3	Eukaryote	0.02	0.37	0.69	0.00
		Cyprinid herpesvirus 1	Eukaryote	0.02	0.59	0.63	0.00
	Herpesviridae	Roseolovirus[Ψ]	Eukaryote	0.00	1.00	0.12	0.04
		Simplexvirus[Ψ]	Eukaryote	0.02	0.59	0.79	0.08
		Cytomegalovirus[Ψ]	Eukaryote	0.02	0.03	0.59	0.00
		Proboscivirus[Ψ]	Eukaryote	0.00	0.12	0.52	0.00
Microviridae	Bullavirinae	unclassified Phix174microvirus	Gammaproteobacteria	4.58	0.66	0.14	2.07
Phycodnaviridae	Prymnesiovirus	Phaeocystis globosa virus	Eukaryote	0.00	0.62	1.82	0.12
ssRNA	Bromoviridae	Tomato aspermy virus	Eukaryote	0.02	0.28	0.91	0.00
	Flaviviridae	Hepacivirus C	Eukaryote	0.59	0.97	0.17	2.35

Phylogenetic assignment of raw sequences was determined using Krakken (Wood and Salzberg, 2014) against the NCBI nt database. Inf = influent, Eff = effluent, Temp = temperate viruses. [Ψ]only classified to genus level. Top 10 most abundant species in each virome are coloured light blue.

Table VII. 5. Functional composition of wastewater viromes.

Protein Feature	Wastewater Virome			
	Inf Free	ML Free	ML Temp	Eff Free
Amino Acids and Derivatives	7.88	0.63	0.79	0.13
Carbohydrates	11.29	0.21	0.22	0.05
Cell Division and Cell Cycle	1.20	5.61	6.63	2.23
Cell Wall and Capsule	4.31	10.00	11.21	1.97
Clustering-based subsystems	13.20	5.18	5.92	2.37
Cofactors, Vitamins, Prosthetic Groups, Pigments	5.78	2.23	2.25	1.17
DNA Metabolism	5.39	3.19	3.17	0.59
Dormancy and Sporulation	0.19	0.35	0.33	0.06
Fatty Acids, Lipids, and Isoprenoids	2.56	0.81	0.76	0.22
Iron acquisition and metabolism	0.94	2.25	2.69	2.15
Membrane Transport	2.21	1.01	1.00	0.23
Metabolism of Aromatic Compounds	0.84	5.96	6.58	2.33
Miscellaneous	6.38	3.32	3.93	1.08
Motility and Chemotaxis	1.23	14.93	7.89	59.25
Nitrogen Metabolism	0.97	3.47	3.54	0.59
Nucleosides and Nucleotides	2.78	0.13	0.13	0.03
Phages, Prophages, Transposable elements, Plasmids	10.81	4.41	5.15	2.03
Phosphorus Metabolism	0.82	2.06	2.33	0.33
Photosynthesis	0.05	1.27	1.31	0.69
Potassium Metabolism	0.33	8.27	9.26	1.37
Protein Metabolism	6.64	5.35	5.10	8.33
Regulation and Cell signaling	1.39	1.06	0.91	0.13
Respiration	2.02	0.07	0.09	0.01
RNA Metabolism	5.18	1.45	1.62	0.19
Secondary Metabolism	0.22	0.71	0.73	0.15
Stress Response	1.75	1.94	1.96	0.59
Sulphur Metabolism	0.84	12.84	13.27	10.62
Virulence, Disease and Defense	2.80	1.29	1.22	1.12

Functional assignment of raw sequences was determined using MG-RAST (Keegan *et al.*, 2016) against the SEED nr database. Inf = influent, Eff = effluent, Temp = temperate viruses. Top 5 most abundant functional categories in each virome are light blue.

Table VII. 6. Composition of wastewater viromes determined by similarity to known protein sequences at the family level.

Viral Type	Viral Family	Host	Wastewater Virome			
			Inf Free	ML Free	ML Temp	Eff Free
dsDNA	*Circoviridae*	Eukarya	0.000	0.032	0.000	0.000
	Lavidaviridae	Viruses	0.000	0.000	0.000	0.001
	Marseilleviridae	Eukarya	0.230	0.000	1.362	0.692
	Mimiviridae	Eukarya	0.000	0.179	0.000	0.192
	Myoviridae	Bacteria, Archaea	26.256	14.169	55.933	39.077
	Phycodnaviridae	Eukarya	5.304	7.469	1.233	1.559
	Pithoviridae	Eukarya	0.000	0.000	0.000	0.005
	Podoviridae	Bacteria	14.213	13.760	3.265	8.855
	Siphoviridae	Bacteria, Archaea	36.973	34.515	7.280	20.188
	Tectiviridae	Eukarya	0.000	0.000	0.000	0.003
ssDNA	*Iridoviridae*	Eukarya	0.000	0.000	0.000	0.006
Total Number of Families			5	6	5	10

Phylogenetic assignment of sequences was determined using BLASTP against the NCBI nr database. Inf = influent, Eff = effluent, Temp = temperate viruses.

Table VII. 7. Functional composition of garnered "virus-only" reads.

Protein Feature	Wastewater Virome			
	Inf Free	ML Free	ML Temp	Eff Free
Amino Acids and Derivatives	0.00	0.02	0.00	0.00
Carbohydrates	0.00	0.11	0.00	0.21
Cell Division and Cell Cycle	2.98	1.99	2.48	0.86
Cell Wall and Capsule	0.00	0.46	0.03	0.48
Clustering-based subsystems	6.17	7.67	4.66	9.10
Cofactors, Vitamins, Prosthetic Groups, Pigments	0.38	0.27	10.21	1.43
DNA Metabolism	5.48	9.14	10.81	6.54
Dormancy and Sporulation	0.00	0.00	0.00	0.00
Fatty Acids, Lipids, and Isoprenoids	0.00	0.08	0.00	0.13
Iron acquisition and metabolism	0.00	0.00	0.00	0.00
Membrane Transport	0.10	1.14	0.73	0.72
Metabolism of Aromatic Compounds	0.00	0.00	0.00	0.01
Miscellaneous	0.56	0.77	8.80	1.47
Motility and Chemotaxis	0.18	0.35	0.70	0.09
Nitrogen Metabolism	0.00	0.00	0.00	0.00
Nucleosides and Nucleotides	1.48	1.20	6.95	1.20
Phages, Prophages, Transposable elements, Plasmids	81.57	75.68	47.07	75.68
Phosphorus Metabolism	0.00	0.00	0.00	0.02
Photosynthesis	0.00	0.00	0.00	0.00
Potassium Metabolism	0.00	0.00	0.00	0.00
Protein Metabolism	0.00	0.15	0.40	0.23
Regulation and Cell signaling	0.49	0.15	2.08	0.42
Respiration	0.00	0.00	0.00	0.00
RNA Metabolism	0.31	0.52	3.39	1.02
Secondary Metabolism	0.00	0.00	0.00	0.00
Stress Response	0.22	0.15	0.00	0.32
Sulphur Metabolism	0.00	0.02	0.00	0.01
Virulence, Disease and Defense	0.09	0.14	1.71	0.06

Functional assignment of "virus-only" sequences was determined using MG-RAST (Keegan *et al.*, 2016) against the SEED nr database. Inf = influent, Eff = effluent, Temp = temperate viruses. Top 3 most abundant functional categories in each virome are light blue.

Table VII. 8. Bray-Curtis coefficients describing the similarity of wastewater viromes.

	Inf Free	ML Free	ML Temp	Eff Free
Inf Free	0	0.78	0.66	0.88
ML Free	0.54	0	0.93	0.48
ML Temp	0.32	0.41	0	0.95
Eff Free	0.59	0.28	0.47	0

Bray-Curtis coefficients for predicted genes and functionally assigned "viral-only" reads across all viromes, lower left and upper right portion of table respectively.